Safety Management and Human Factors, Vol. 227, 2026, 77–87
https://doi.org/10.54941/ahfe1007919

Misinformation Risk: Epidemiological and Social Models

Daniel Botelho, Marta Ribeiro, Maria Teresa Monteiro, and Senhorinha Teixeira

Department of Production and Systems & ALGORITMI Centre, University of Minho, Guimarães, Portugal

ABSTRACT

Understanding the mathematical dynamics of information propagation is critical for cognitive safety and crisis management. This study applies biological and sociolinguistic models to the diffusion of fake news in social media using the disinformation outbreak following the terrorist attack on Charlie Hebdo as case study. Biological models such as Susceptible-Infected-Recovered, Susceptible-Exposed-Infected-Recovered and sociolinguistic frameworks like Daley-Kendall and Maki-Thompson are fitted to real-world data using the nonlinear least squares method. The numerical results demonstrate that the Daley-Kendall model provides the most accurate fit to the observed data, outperforming the classical biological models. The findings indicate that the end of an explosive rumour is not governed by passive temporal recovery or incubation periods (as assumed in biological models) but by the mutual stifling. This suggests that in high-frequency social networks, information redundancy serves as the primary stabilizing mechanism. These insights propose that effective safety interventions should focus on accelerating network saturation to mitigate the spread of cognitive hazards.

Keywords: Risk modelling, Cognitive safety, Disinformation, Maki-Thompson model, Human factors

INTRODUCTION

The spread of fake news and rumours constitutes a threat to social and economic stability since it has the potential to incite public panic (Botelho et al., 2024). With the rise of social media and platforms such as Whatsapp and X, former Twitter, this dynamic has intensified. The digital age we currently live in allows for information, true or false, to proliferate like a virus resulting in an infodemic where information travels across the world independent of its veracity (Cinelli & Gesualdo, 2025).

This phenomenon has captured the attention of scientists due to its resemblance with the spread of infectious diseases, where transmission occurs through contact between individuals.

Despite its popularity, the direct application of classical epidemiological models such as Susceptible-Infected-Recovered (SIR) or Susceptible-Exposed-Infected-Recovered (SEIR) to information propagation faces substantial criticisms regarding its ontological and mechanical validity (Yee, 2025). While in biological models the transition from Infected to Recovered compartments generally occurs due to temporal processes (i.e., immunity or

Received February 2, 2026; Revised April 11, 2026; Accepted April 29, 2026; Available online July 20, 2026

death), the cessation of a rumour involves complex social mechanisms that have no direct biological parallel (Coletti et al., 2025; Yee, 2025).

Recent studies indicate that the end of a rumour actively depends on refutation mechanisms and information feedback, where the presence of sceptical individuals or those with critical capacity act as a barrier to the propagation. Unlike the passive recovery from a disease, the process of becoming a stifler can occur through contact with contradictory versions of the rumour, media influence, or confrontation with the truth, suggesting that purely biological models may fail by ignoring intentionality and the social of "truth" (Botelho et al., 2024; Yee, 2025).

This work aims to compare two distinct approaches to the mathematical modelling of fake news spread. On one hand, we will analyse the models based on the logic of biological recovery (SIR and its extensions SEIR and SEIZ) that try to adapt the epidemic dynamics introducing compartments such as Sceptics or Exposed to capture the hesitancy and refutal. On the other hand, we will also analyse classical social stifling models, specifically Daley-Kendall (DK) and Maki-Thompson (MT). These models differ by assuming that the spread actively ceases through social interactions: one spreader becomes a stifler only by contacting other who already know the rumour or that refute it, a mechanism fundamentally different from spontaneous biological recovery.

This analysis will allow to understand if including critical conscious mechanisms, absent on the biological simple models, is essential to describe the reality of disinformation (Cinelli & Gesualdo, 2025; Yee, 2025).

This paper is structured as follows: Section 2 outlines the theoretical background, presenting the mathematical formulations of the biological (SIR, SEIR) and sociolinguistic (Daley-Kendall, Maki-Thompson) models utilized. Section 3 describes the methodology and the dataset used, specifically the application of the nonlinear least-squares method to the Charlie Hebdo misinformation event. Section 4 presents the numerical results and the comparative analysis of the model fittings. Section 5 discusses the implications of these findings regarding information redundancy and network saturation. Finally, Section 6 summarizes the main conclusions.

THEORETICAL BACKGROUND

Mathematical modelling of information spread in social media has been consolidated as a fundamental tool to risk analysis in terms of cognitive security and crisis management (Comes et al., 2022). The central premise of this area, frequently designated by "infodemic" is that information spread (ie., rumours or disinformation) exhibits dynamic patterns similar to the spread of infectious diseases. Historically, this approach is based on compartmental models where the population is divided into mutually exclusive compartments such as Susceptible or Infected and the transition between them is governed by a system of Ordinary Differential Equations (ODEs) (Govindankutty & Gopalan, 2024).

However, the interpretation of these models is not purely biological. Among the earliest examples of sociolinguistically inspired mathematical models are the Maki-Thompson (MT) and Daley-Kendall (DK) frameworks, which were originally developed to describe rumour propagation driven by verbal social interactions (Daley & Kendall, 1965; Gani, 2000). In this context, the literature diverges when it comes to the exact mechanism that dictates the termination of an informational outbreak (Ferraz de Arruda et al., 2022). On one hand, the classical biological approach assumes that the end of the spread occurs through a process of temporal recovery (Govindankutty & Gopalan, 2024). On the other and, the sociolinguistic approach argues that human dynamics require active mechanisms of social interaction, specifically the phenomenon of suffocation (i.e., stifling) caused by information redundancy (Ferraz de Arruda et al., 2022; Gani, 2000).

SIR Model

The SIR model is a foundational compartmental framework in biological epidemiology. It divides the population into three exclusive compartments:

- Susceptible Individuals (S): those who have not yet encountered the disease;
- Infected Individuals (I): individuals infected and capable of transmitting the false information;
- Recovered Individuals (R): those who have Recovered and are immune to reinfection.

The dynamics between these compartments are governed by the following system of ODEs:

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{\beta SI}{N} \\ \dfrac{dl}{dt} = \dfrac{\beta SI}{N} - \gamma I \\ \dfrac{dR}{dt} = \gamma I \end{cases}$$

where β is the contact rate between susceptible and infected individuals, γ the recovery rate of infected individuals, N is the total number of individuals and t denotes time.

When applied to the context of information dissemination on social media, the SIR model's compartments acquire new interpretations. In this context, the "disease" is not a biological pathogen but rather false information (DeVerna et al., 2025; Govindankutty & Gopalan, 2024):

- Susceptible (S) individuals represent those who have not yet been exposed to the false claim or who have not adopted it.
- Infected (I) individuals are those who have been exposed to and believe the misinformation, actively sharing it within their social networks.

- Recovered (R) individuals are those who have ceased spreading the false information, either because they have become sceptical, fact-checked the claim, or simply lost interest due to information saturation.

This interpretation bridges epidemiological theory with the dynamics of digital information.

SEIR Model

The SEIR model extends the classical SIR framework by introducing an Exposed (E) compartment, representing individuals who have been infected but are not yet infectious. This intermediate step captures the latency period between exposure and active transmission.

The dynamics between these compartments are governed by the following system of ODEs:

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{\beta SI}{N} \\ \dfrac{dE}{dt} = \beta S\dfrac{I}{N} - \sigma\dfrac{E}{N} \\ \dfrac{dl}{dt} = \sigma\dfrac{E}{N} - \gamma I \\ \dfrac{dR}{dt} = \gamma\dfrac{I}{N} \end{cases}$$

where β is the contact rate between susceptible and exposed individuals, σ the rate of transition from Exposed to Infected and γ the recovery rate of infected individuals.

In the context of information diffusion, the Exposed compartment represents users who have encountered misinformation but have not yet started to shared it. This captures a critical social dynamic: People often require time to process, evaluate, and decide whether to amplify a false claim. Thus, the SEIR model provides a more nuanced representation of misinformation dynamics than SIR, acknowledging that exposure and belief (i.e., sharing) are distinct processes. Extended formulations such as the Susceptible-Exposed-Detected-Protected-Non-spreader-Recovered (SEDPNR) (Govindankutty & Gopalan, 2024) have been proposed to account for additional social mechanisms. The presence of “Doubters” (Individuals sceptical of the information) and “Restrained” (individuals who cease sharing due to loss of interest or fact-checking) are examples of new compartments that take into account real world dynamics making the models increasingly similar to reality.

DK Model

The Daley-Kendall (DK) model represents a shift from purely biological frameworks to social compartmental modelling. Rather than tracking disease susceptibility and infection, the DK model tracks the social dynamics of information propagation through populations. Similar to the SIR model, the population is divided into three states (Ferraz de Arruda et al., 2022):

- Ignorants (S_U) — those unaware of the rumour or misinformation;
- Spreaders (I_V) — those actively propagating the rumour;
- Stiflers (R_W) — those who have ceased spreading because they have encountered the information redundantly.

These model dynamics are governed by the following system of ODEs:

$$\begin{cases} \dfrac{dS_u}{dt} = -\beta S_u I_v \\ \dfrac{dI_v}{dt} \beta S_u I_v - \gamma I_v (I_V + R_w) \\ \dfrac{dR_w}{dt} = \gamma Iv(I_v + R_w) \end{cases}$$

where β denotes the rate at which spreaders contact ignorants and turn them into new spreaders, γ and represents the rate at which encounters between spreaders and already informed individuals (other spreaders or stiflers) lead to stifling due to information redundancy.

In the DK framework applied to social media, the "rumour" is understood as any piece of contested, false, or misleading information. Ignorants correspond to individuals unexposed to the claim; Spreaders are active sharers (the equivalent of the Infected in SIR model) and Stiflers represent individuals who have stopped amplifying the message.

In the DK model, the transition to the Stifler state occurs through a distinctly sociolinguistic mechanism (i.e., information redundancy). When a Spreader encounters either another Spreader or a Stifler, they transition to the Stifler state themselves, having learned that the information is already widely known or has been disputed (Ferraz de Arruda et al., 2022). This contrasts sharply with the biological SIR model where recovery is passive and temporal.

MT Model

The Maki-Thompson (MT) model introduces a critical refinement to the DK framework by making contact dynamics directional rather than undirected (Gani, 2000). In the MT model, a Spreader initiating contact with another individual (either Ignorant, Spreader, or a Stifler) causes the initial Spreader to become a Stifler. The contacted individual, however, may or may not transition state depending on their current status (Gani, 2000).

Let S_1, I_S, and denote, respectively, the numbers of Ignorants, Spreaders, and Stiflers at time . The model's dynamics are given by (Ferraz de Arruda et al., 2022):

$$\begin{cases} \dfrac{dS_i}{dt} = -\beta S_i I_s \\ \dfrac{dI_s}{dt} \beta S_i I_s - \gamma S_i (S_i + R) \\ \dfrac{dR}{dt} = \gamma S_i (S_i + R) \end{cases}$$

where β is the rate at which spreaders contact Ignorants and turn them into new Spreaders, and γ represents the rate at which directed contacts initiated by spreaders with already informed individuals (ignorants who remain unconvinced or stiflers) lead to stifling.

Similar to the DK model, in the context of misinformation, Ignorants correspond to users who have not yet encountered the false claim, Spreaders are users actively sharing it, and Stiflers are users who, after repeated exposure or debunking, decide to stop amplifying the misinformation. The directed contact structure of the MT model is particularly suitable for asymmetric online platforms, where information flows follow follower-following relationships rather than undirected contacts (DeVerna et al., 2025; Govindankutty & Gopalan, 2024).

The distinction between undirected (DK) and directed (MT) contact is particularly relevant to misinformation on asymmetric social media platforms such as X (former Twitter). Here, users can amplify a post (retweet), but the original poster and the re-sharer have asymmetric roles. The MT model captures this asymmetry: a user who originally shared misinformation may cease sharing once they see the claim is already circulating widely or has been debunked (i.e., they become a Stifler after "hearing" the redundant or contradictory information). This seemingly small change in contact directionality leads to substantially different dynamics, including phase transitions in how misinformation penetrates a population (Ferraz de Arruda et al., 2022).

METHODOLOGY AND DATA

The quality of a model ultimately hinges on how well it describes real world dynamics. In an information diffusion context, parameters such as the transmission rate and the recovery rate show us how quickly misinformation is adopted, propagated and eventually abandoned over time. In this section, the parameters of each model are fitted by aligning the model output to the observed data series using a nonlinear least-square approach. The optimisation problem consists in minimising the discrepancy between the number of active users sharing information and the infected population predicted by the model. Parameter fitting was carried out in MATLAB with the *lsqcurvefit* routine, which iteratively updates the parameter set until the simulated trajectory provides the best match to the empirical time series. This routine directly provides the Sum of Squared Residuals (SQR), which quantifies the final approximation error. Our empirical data corresponds to the Charlie Hebdo rumour, one of the real-world misinformation events documented in the PHEME (Derczynski et al., 2017) rumour information repository.

This dataset captures the activity on X, related to the terrorist attack on the Charlie Hebdo offices in Paris on January 7th of 2015. Post counts were aggregated into hourly intervals, yielding the total number of posts mentioning the rumour as a function in time. The observation window ranges from January 7th 2015, 08h00 to January 10th 2015, 07h00, covering approximately 72 hours and encompassing the full diffusion cycle of the rumour (i.e., from the initial surge following the attack through the peak of attention, to the subsequent decay as the story stabilised). The dataset is shown in Figure 1.

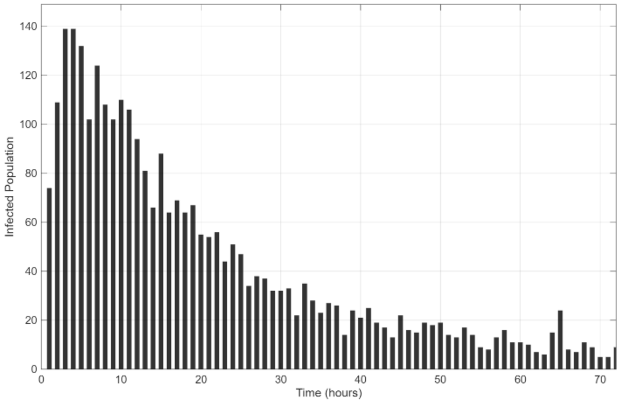


Figure 1: Number of posts after the Charlie Hebdo terrorist attack (Derczynski et al., 2017).

The dataset is used as the infected population curve in the epidemiological framework, representing the number of users actively engaging with the rumour at each hour. For modelling purposes, direct correspondence was assumed between social media activity and the number of infected individuals. Due to the scarcity of data regarding unique user identification in the available dataset, each post is treated as a proxy for one active spreader (1 post = 1 infected user). These values were derived directly from the dataset. The initial number of individuals (I_0) corresponds to the first observed value of the time series, and the total population ($N = 2996$) was set to the sum of all observations, assuming that the total number of individuals who participated in the spread remains constant. Table 1 shows the initial values of the different variables included in each model.

Table 1: Sample human systems integration test parameters (Folds et al., 2008).

Variable	SIR	SEIR	MT	DK
S_0	2921	2920	2921	2921
E_0	-	1	-	-
I_0	74	74	74	74
R_0	1	1	1	1

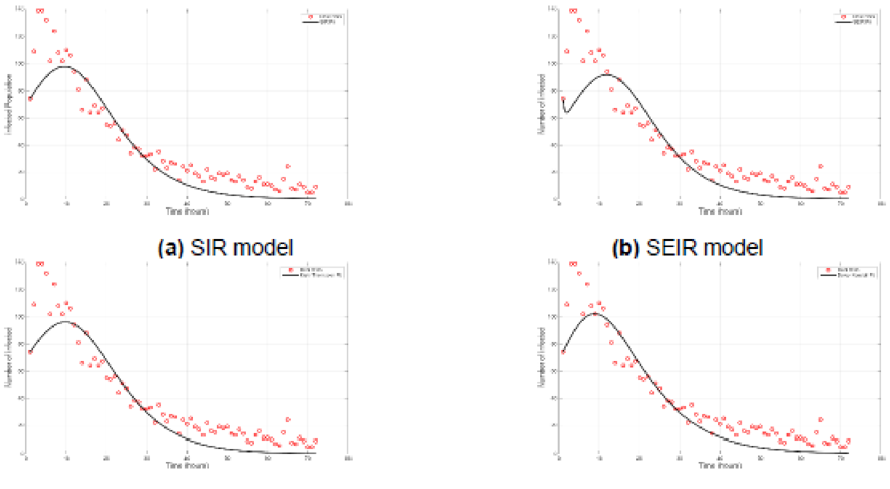
RESULTS

Comparing the goodness-of-fit metrics, the DK model demonstrated superior performance among all tested models, achieving the lowest Relative Error (RE) (0.252) and the lowest SQR (14231.9). The RE is defined as the norm of the residuals (the difference between observed data and model output) normalized by the norm of the observed data. This indicates that the DK model provided the most accurate representation of the fake news trajectory. The SIR model ranked second, with an RE of 0.272, followed closely by the MT model (RE = 0.278). The proximity of these values suggests that both biological recovery mechanism (SIR) and the unilateral social stifling mechanism (MT) offer comparable approximations for this specific dataset. The SEIR model yielded the poorest performance, presenting the highest RE (0.337) and SQR (2557.3) significantly diverging from the real data points when compared to the other three models. Table 2 presents the estimated optimal parameters (*β*, *γ*, σ), the RE and the SQR for each model.

Table 2: Comparison of optimized parameters and fitting performance of the models.

Model	Optimized Parameters	RE	SQR
SIR	$\beta = 0.526, \gamma = 0.449$	0.272	16572.5
SEIR	$\beta = 0.644, \sigma = 2.000, \gamma = 0.526$	0.337	25571.3
MT	$\beta = 0.1100, \gamma = 2.011$	0.278	17338.9
DK	$\beta = 0.1671, \gamma = 1.600$	0.252	14231.9

The numerical optimisation of the four mathematical models applied to the Charlie Hebdo dataset is also illustrated in Figure 2 - (a) to (d).



(a) SIR model

(b) SEIR model

(c) MT model

(d) DK model

Figure 2: Model fitting comparison to the Charlie Hebdo dataset using (a) SIR model, (b) SEIR model, (c) Maki-Thompson model and d) Daley-Kendall model.

The dataset exhibits a sharp, explosive growth followed by a rapid decay, characteristic of an explosive event. DK, SIR and MT models visually align with the peak of the infection curve and subsequent drop. However, the SIR and MT models produce the most similar curves. The SEIR model displays a noticeable delay in the initial phase. The inclusion of the incubation period (σ) forced the model to simulate a delayed start, which contradicts the immediate viral explosion observed in the real data, resulting in a poor fit.

DISCUSSION

The superior performance of the DK model is the most significant finding of this study. Unlike the MT model, where only one individual ceases propagation upon contact, the DK framework assumes a pairwise mutual cessation mechanism ($I + I + I = 2R$. In the context of the Charlie Hebdo attack, an explosive event, this mechanism reflects the real-world human behaviour of saturation. When the density of spreaders is high enough, the interactions become redundant for both parties involved. The fact that the DK model best minimised the error suggests that the end of the rumour was driven by an aggressive simultaneous loss of interest among interaction peers rather than a gradual (biological) or unilateral (MT) process. This confirms that in high-frequency social networks, redundancy is the primary inhibitor.

The poor performance of the SEIR model also showed us that latency is practically non-existent. In biological epidemics the Exposed compartment represents the incubation period where an individual is infected but not yet infectious. However, our results show that adding this state to information models is counterproductive. In social media platforms, the time gap between exposure (i.e., reading a post) and infection (i.e., sharing) is measured in seconds. The SEIR model's attempt to enforce a delay caused it to miss the explosive surge of the crisis.

A surprising result was the SIR model ranking in second, outperforming the social MT model. While sociologically less accurate, since it assumes information dies due to time and not interaction, the SIR model's mathematical simplicity allowed it to approximate the rapid decay effectively. This suggests that in explosive events like Charlie Hebdo, the complex social dynamic of stifling (where users stop sharing because they encounter redundancy) produces a data curve that is almost identical to the passive recovery found in biological models. Mathematically, this implies that the loss of interest driven by social interactions mimics a constant recovery rate , even though the underlying mechanisms are fundamentally different. However, relying on the SIR model for decision-making carries a risk since it implies that the crisis will end passively with time.

These findings shift the paradigm of mitigation strategies. If the most accurate model (DK) relies on mutual stifling to predict the end of a threat, then lawmakers should not focus solely on refuting false information. Instead, strategies should leverage network saturation. By rapidly disseminating verified facts to key nodes it is possible to trigger the DK stifling mechanism earlier and, thus, infecting less people. This study suggests that cognitive safety is better maintained by accelerating the lifecycle of the topic through the massive injection of verified information to the point of redundancy.

CONCLUSION

This study aimed to determine whether the end of a misinformation crisis is best characterized by temporal decay, similar to biological recovery, or by active social stifling mechanisms. The comparative analysis of the Charlie Hebdo dataset reveals that sociolinguistic models, specifically the DK framework, provide superior description of the explosive dynamics of fake news when compared to traditional epidemiological models.

These findings validate that in explosive events, the cessation of propagation is not a solely a process of spontaneous loss of interest but a phenomenon of network saturation. The success of the DK model confirms that mutual cessation (where interactions between spreaders lead to a simultaneous loss of interest due to redundancy) is the dominant suppressing factor in social media. Furthermore, the poor performance of the SEIR model highlights that latency periods are negligible in digital networks where exposure and infection are almost instantaneous. Thus, mitigation strategies should be active instead of passive, leveraging information redundancy.

ACKNOWLEDGMENT

This work has been supported by FCT – Fundacão para a Ciência e Tecnologia within the R&D Unit Project Scope UID/00319/2025 – Centro ALGORITMI (ALGORITMI/UM).

REFERENCES

Botelho, D., Monteiro, M. & Teixeira, S. (2024), The infodemic issue: Numerical modelling, in 'Optimization, Learning Algorithms and Applications 2024'. https://doi.org/10.1007/978-3-031-77426-321

Cinelli, M. & Gesualdo, F. (2025), 'Infodemic versus viral information spread: Key differences and open challenges', JMIR Infodemiology 5, e57455. https://doi.org/10.2196/57455

Coletti, C. F., Luiz, D. A. & Rada, A. (2025), 'The maki-thompson model with random awareness', arXiv preprint arXiv:2508.07099.

Comes, T., Yan, F. & Wijngaards, N. (2022), 'On the interplay of data and cognitive bias in crisis information management', Risk Analysis 42(3), 493–512. https://doi.org/10.1111/risa.13781

Daley, D. J. & Kendall, D. G. (1965), 'Epidemics and rumours', Nature 204, 1118. https://doi.org/10.1038/2041118a0

Derczynski, L., Bontcheva, K., Liakata, M., Procter, R., Wong Sak Hoi, G. & Zubiaga, A. (2017), 'Pheme rumour scheme dataset (July 2017)', Available at: https://figshare.com/articles/dataset/PHEME rumour scheme dataset july 2017/6392078. Accessed October 2025.

DeVerna, M. R. et al. (2025), 'Modeling the amplification of epidemic spread by individuals exposed to misinformation on social media', Nature Digital Medicine 1, 38. https://doi.org/10.1038/s44260-025-00038-y

Ferraz de Arruda, G., Rodrigues, F. A. & Moreno, Y. (2022), 'From subcritical behavior to a correlation-induced transition in rumor models', Nature Communications 13, 3203. https://doi.org/10.1038/s41467-022-30683-z

Gani, J. (2000), 'The maki–thompson rumour model: A detailed analysis', Environmental Modelling & Software 15(2), 155–161. https://doi.org/10.1016/S1364-8152(00)00029-3

Govindankutty, R. & Gopalan, M. (2024), 'Epidemic modeling for misinformation spread in digital networks through a social intelligence approach', Scientific Reports 14, 19100. https://doi.org/10.1038/s41598-024-69657-0

Yee, A. K. (2025), 'The limits of epidemiological models of misinformation', Synthese 206, 158. https://doi.org/10.1007/s11229-025-05246-6