

Predicting Information Diffusion on Social Media Using an Epidemiological Approach

Dajana Conte(✉), Samira Iscaro, and Beatrice Paternoster

Department of Mathematics, University of Salerno, Via Giovanni Paolo II, 132-84084 Fisciano, SA, Italy
{dajconte,siscaro,beapat}@unisa.it
https://www.dipmat.unisa.it/

Abstract. Recent decades have been characterized by deep changes in our way of communicating thanks to powerful new tools e.g., social networks, that allow a fast spread of information with very limited costs and control. In this paper, we will show the importance of analyzing information diffusion on social media in the first phases of the news spread, using a mathematical epidemiological approach. We will highlight the importance of predicting the evolution of the news trend over time before a loss of interest from social media users is observed. For this reason, firstly, we will analyze the characteristics of several kinds of mathematical models from literature to choose the most appropriate for our purposes, showing that a possible choice consists of a model of Ignorant - Spreader - Exposed - Skeptic (IESZ) type. Then, we will describe a possible strategy to compute optimized parameters for the chosen model starting from a dataset of real data. Finally, by exploiting several case studies regarding both true and fake news shared on X (Twitter) in the last years, we will show that the proposed strategy is truly applicable to reality.

Keywords: Information diffusion · Epidemiological models · IESZ model · Parameter estimation · Trust region methods

1 Introduction

The early years of the 21st century have been characterized by a deep change in communication means. New tools, such as content sharing sites, social networks, forum, blogs, and many other typologies of websites, have completely changed our social life, from the way of communicating to our way of establishing new relationships. This simplification in communications has led to a series of economic advantages, since the buying and selling of products has been made easier. However, also a rapid increase in the spread of false or only partially verified news has been observed. The latter phenomenon is also more worrying if it is considered that there exists a strict link between public emotions and common

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
O. Gervasi et al. (Eds.): ICCSA 2025 Workshops, LNCS 15887, pp. 208–224, 2026.
https://doi.org/10.1007/978-3-031-97589-9_15

opinions regarding particular events, as e.g. the outcomes of many social or political events or unusual situations, and the dissemination of information related to them [45,46]. For this reason, a proper way to analyze information diffusion is required. Several approaches are possible, from machine learning tools and statistical techniques [9,35], to the use of mathematical models based on systems of differential equations. Among these types of models, epidemiological ones can be considered. Generally based on a system of ordinary differential equations (ODE_S), these models describe the evolution in time of epidemics. In fact, their application to the field of the analysis of information diffusion is due to the possibility of considering a news, that is spread on a social network, in the same way as a virus that affects the human population. Just as a virus spreads through encounters between individuals, information spreads when a social media user reads a piece of news and decides to share it. In [26,32], D'Ambrosio et al. carried out an analysis of this kind considering two well known epidemiological models from literature, the Susceptible-Infectious-Recovered (SIR) and the Susceptible-Exposed-Infectious-Recovered ($SEIR$) model respectively, and linked their parameters to two indexes related to the grade of access to the web of a population and its ability to recognize fake news. In [7], Cardone et al. showed the importance of considering the age of individuals in the model formulation process, showing that the possibility that a news is spread is strictly related to the age of individuals who share it. Worthy of consideration is also the work carried out in [50] by Muhlmeyer et al. Here they show different possible mathematical models for studying the diffusion of information, presenting models suitable for the diffusion of fake news or highly debated news where two opposed groups of news spreaders are involved in the process of news sharing, or for situations in which there is first of all an initial interest in the news, then a phase when the news is in trend and popular, and a final phase in which there is a loss of interest in this news. Note that, since the scope of these kind of analyses is the description of real case studies, it is necessary to built a dataset by extracting data from a certain social network. However, this action is not trivial and can led to errors in the row data extracted. To take in consideration the noise in the data, stochastic models can be considered [50]. In particular, more details regarding stochastic differential equations can be found in [5,11,14,25,27–31,33,36]. Furthermore, the popularity of a news does not only depend on how many people are interested in it, but also by whom it is shared. If the person sharing the news in question is famous, the probability that this news will be seen by a large number of people is undoubtedly greater. For this reason, several studies related to the analysis of the information diffusion through mathematical models also require an accurate study of the main characteristics of the social network where the news is spread. In particular, these studies make use of graph theory and the possible use of centrality measures of a graph to determine the most influential users who can influence the outcome of the diffusion process of the considered news (see [4] for more details and references therein). Moreover, taking in consideration a possible spatial distribution of the social network users, also model based on partial differential equations (PDE_S) can be employed [55].

Researches in this field, however, is widespread. Epidemiological models, in fact, are characterized by positive solution, and by some invariant quantities, as for e.g. equilibrium points. For this reason, it is fundamental to apply proper numerical methods. For example, recent works in which a description of numerical methods able to preserve positivity and invariant quantities of ODE_S or PDE_S based models are [3,8,15,16,37–43]. Furthermore, in the specific case where models involving PDE_S are considered, it is important to know how to handle large-scale problems during the model resolution phase. In this sense, there are countless works in the literature in which numerical techniques for this kind of situations are presented. Among the most recent works, for example, we can mention [1,2,6,17,21,23,24,44,51].

According to the main features of the news, the most suitable typology of model should be considered. However, the only description of the evolution of the news popularity will not be useful in the long term. We are interested in predicting this evolution by observing the phenomenon just in the early phases of the news spread. An example of possible strategies to predict the evolution of news can be found in [10,20], where a strategy to predict the moment of maximum interest towards a news from social media users, starting from real data, is exploited. Certainly, one of the possible advantages of these studies is the possibility of blocking a fake news before it becomes popular. Moreover, also companies can benefit from these studies, since knowing the moment in which there is interest in their products can allow them to maximize earnings through a proper advertisement campaign. However, it is also true that news shared on a social network are not in trend for a long period of time, so in order to benefit from these kinds of analyses, it is also important to know in advance the possible total number of individuals that will share something related to the considered news after a short period of time since the start of the news spreading, and before the loss of interest towards it is observed. For this purposes, the aforementioned models are not suitable. Thus, in this work, we want to highlight the possible use of another mathematical model present in literature [47]: the Susceptible-Exposed-Infectious-Skeptic ($SEIZ$) model. To achieve this goal, starting from datasets of real data extracted from the social network X (Twitter), a parameter estimation phase (or data fitting phase), based on a trust-region method, will be employed.

The work will be organized as follows. *Section* 2, is dedicated to the description of the used mathematical model and its main features. *Section* 3 is dedicated to the experimental phase: in particular, we will describe both the strategy followed for the parameter estimation phase and the case studies being analyzed showing, to confirm the reliability of our strategies, the obtained numerical results. Finally, *Section* 4 will be dedicated to concluding remarks and future works.

2 Epidemiological Models for Information Diffusion

As above mentioned, a possible approach to analyze and possibly predict the spread out of information consists in the use of ODE_S based model of epidemi-

ological type. These models, in particular, are of compartmental type, i.e. they divide the population in classes or compartments, and study their evolution using ordinary differential equations [48]. Certainly according to the features of the epidemic being described, several classes can be considered, as for example:

- *Susceptibles* $S(t)$, i.e. the individuals that can be infected by the virus;
- *Infectious* $I(t)$, i.e. the individuals affected by the disease;
- *Exposed* $E(t)$, i.e. the individuals that have been exposed to the disease but that are not yet infectious;
- *Recovered* $R(t)$, i.e. the individuals that recovered from the disease.

Actually in the epidemiological context many other classes can be considered as hospitalized, individuals in quarantine, dead, vaccinated and so on. As aforementioned, news can be analyzed in the same way of diseases: however, it should be taken in consideration that the two contexts are different and that the meaning given to the above mentioned classes slightly changes. Thus, following the terminology of [50], we will refer to the following classes during the model formulation process. We will refer to:

- *Ignorants*, to describe the counterpart of the susceptible class of epidemics, i.e. all the individuals that can see a news and share it;
- *Spreaders*, to describe the counterpart of infectious class, i.e. the individuals that share the news;
- *Recovered* to describe individuals, that exactly as the Recovered of the epidemics, are no longer interested in the process being considered.

Moreover, in addition to these classes, new others can be introduced. For example, if the shared news is fake or related to an highly debated topic, it is reasonable to consider two classes of spreaders: one that shares the news and which is still addressed as the Spreader class and another class, the *Counter Spreader* class generally indicated with $C(t)$, of individuals that contrast the diffusion of news spread by spreaders by sharing news with the opposite content [50]. Using the above mentioned classes, the Ignorant-Spreader-Recovered (ISR) model [10,50] and the Ignorant-Spreader-Counter Spreader-Recovered (ISCR) model [50] can be formulated.

The ISR model can be formulated by considering that, when individuals of the ignorant class read a news shared by spreaders, they automatically believe in the news and decide to share it, while after a certain period of time, they will lose interest in the news, consequently transiting into the recovered class. These transitions, that happen respectively at rate β and γ and that are also influenced by the average number of connections κ among individuals, are summarized by the following graph (Fig. 1).

Specifically, the ISR model is formulated through the following system:

$$\begin{cases} I'(t) = -\beta\kappa I(t)S(t), \\ S'(t) = \beta\kappa I(t)S(t) - \gamma\kappa S(t)S(t), \\ R'(t) = \gamma\kappa S(t)S(t). \end{cases} \tag{1}$$

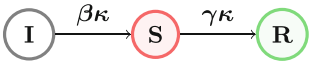

Fig. 1. Graph summarizes transitions among classes in the ISR model [50].

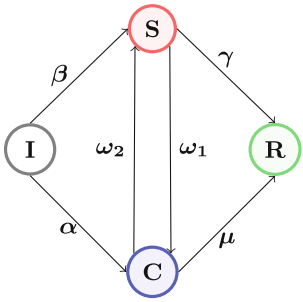

Fig. 2. Graph that shows possible transions among classes into the ISCR model [50].

Note that this is not the only way to formulate a model of ISR type for information diffusion. In fact, this model originally presented in [50] is specific for social media. Other possible formulations can be found also in [50] or in [32] and references therein.

On the contrary, the ISCR model that consider two opposed classes of individuals that share conflicting news items on the same topic, can be formulated as follows:

$$\begin{cases} I'(t) = -\beta\kappa I(t)S(t) - \alpha\kappa I(t)C(t), \\ S'(t) = \beta\kappa I(t)S(t) - (\omega_1 - \omega_2)\kappa S(t)C(t) - \gamma\kappa S(t)[S(t) + R(t)], \\ C'(t) = \alpha\kappa I(t)C(t) + (\omega_1 - \omega_2)\kappa S(t)C(t) - \mu\kappa C(t)[C(t) + R(t)], \\ R'(t) = \gamma\kappa S(t)[S(t) + R(t)] + \mu\kappa C(t)[C(t) + R(t)]. \end{cases} \tag{2}$$

Note that, in this case, apart from transitions from the ignorant class to the spreader or counter spreader class that respectively happen at rate β and α, and the corresponding transitions into the recovered class at rates γ and μ when a loss of interest in the news occurs, also other interactions are considered. In fact, parameters ω_1 and ω_2 that describe the transition rates from spreaders to counter spreaders class and vice versa have been introduced into the model formulation (Fig. 2).

Exhaustive examples of applications of these two models are given in [10,50]. As can be seen by results described there, both the two models take in consideration a possible loss of interest by spreaders, or by counter spreaders when present, towards the news. So, the function that represents the Spreaders, is firstly an increasing functions, that at a certain point reaches a peak, before decreasing.

Although results described in these works are good, in some situations it can be more advantageous to analyze the spread of news only in the early phases of their spread. In fact, since we are at the beginning of the spread, it is reasonable

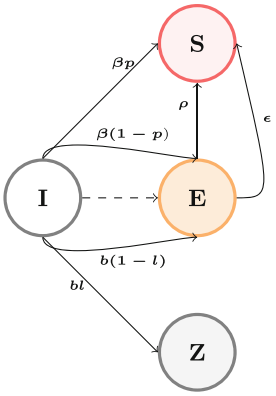

Fig. 3. Graph of the ISEZ model [47].

that the considered news is not well known. So, it could happen that not every social media user that reads it will immediately share something related to it. Moreover, it could also happen that some individuals will not show any particular reaction towards the news item. In this case, a more specific model can be adopted: the Ignorant-Exposed-Spreader-Skeptic (IESZ) model, also known in literature with the classical terminology of $SEIZ$ model. Described and analyzed in depth in [47] this model, born to describe information diffusion, provides a division of the population in four classes. Apart from Ignorant ($I(t)$) and Spreaders ($S(t)$) previously described, the classes of Exposed ($E(t)$) and Skeptic ($Z(t)$) are considered, and both refer to individuals who have seen a news but have not shared anything about it yet. However, in the model formulation, it is assumed that exposed believe in the news and so, sooner or later, will share something related to it, while skeptic have not showed any interest towards the news, choosing to ignore it. Note that, skeptic individuals seem to be similar to the recovered of the previous two models, however they have a completely different role. In fact, they have never had any reaction towards a certain news, while recovered believed in the news and shared it, before loosing interest. By indicating with N the total number of individuals of the population, computed as the sum of all the individuals of the four compartments in which it has been divided, the IESZ model reads as follows:

$$\begin{cases} I'(t) = -\beta I(t)\frac{S(t)}{N} - bI(t)\frac{Z(t)}{N}, \\ E'(t) = (1-p)\beta I(t)\frac{S(t)}{N} + (1-l)bI(t)\frac{Z(t)}{N} - \rho E(t)\frac{S(t)}{N} - \epsilon E(t), \\ S'(t) = p\beta I(t)\frac{S(t)}{N} + \rho E(t)\frac{S(t)}{N} + \epsilon E(t), \\ Z'(t) = lbI(t)\frac{Z(t)}{N}. \end{cases} \tag{3}$$

Transitions between different classes of the population are illustrated by the graph in Fig. 3.

As it can be seen from the model formulation in (3), a major number of parameters is involved. In particular, we take:

- β, the transition rate from the ignorant class to the spreader class: in this case, when individuals of the ignorant class read a news, spread by a spreader, they can transit into the spreader class with probability p, or into the exposed class, if they believe in the news but don't share anything, with probability $1 - p$.
- b, the transition rate from the ignorant class to the skeptic class. As it happens for transitions from the ignorant class to the spreader class, it can happen that ignorant individuals decide to ignore the news, transiting into the skeptic class with probability l, or they can be affected by it, transiting into the exposed class with probability $1 - l$. Note that, once in the exposed class, it is not possible to come back into the skeptic compartment anymore, since the only allowed transition is into the spreader class.
- ρ, the contact rate between an exposed individual and a spreader individual. This parameter is used to rule all the transitions from the exposed class to the spreader class, that happen when exposed individuals get in contact with news shared by spreaders, such that they directly transit into the spreader class.
- ϵ, the transition rate from the exposed class to the spreader class. Observe that, contrary to ρ, the part of exposed population that transits into the spreader class at rate ϵ is the one that transits there after a period of uncertainty relating to whether or not to believe the considered news. In this case, the exposed are not further influenced by news spread by spreaders.

In particular, in the IESZ model graph, there is a dotted line that describes the passage between classes of ignorant and exposed. This is due to the fact that individuals from the ignorant class can pass into the exposed class either because of the influence of a spreader or because, even if initially skeptical, they have then believed in the news.

Note that, from a modeling point of view, the above described IESZ model is suitable to analyze the initial phase of news spread, particularly if we want to predict the total number of individuals interested in the news after a short period of time since the beginning of its spread. Moreover, it should be noticed that, if we want to analyze real case studies, we need data to be used for the model validation phase. However, in this case, it is not possible to extract data regarding the initial number of exposed or skeptic, since it is not known who, of all the possible users interested by the news, has read it and decided to ignore it, or that has believed in it but without sharing anything yet. For this reason, a data fitting phase involving not only model parameters but also the initial conditions of the problem, is essential to obtain good results, as we will show in the next Section.

3 Parameter Estimation and Numerical Results

As previously anticipated, in order to analyze the process of information spread considering real case studies using the IESZ model, we need both to extract real data from a chosen social network and an adequate parameter estimation phase. Several techniques are available for these kinds of problems. For example, in [49], Physics Informed Neural Network (PINN) are used to optimize parameters involved into a model of SIR type (see also [54] for other possible uses of PINN). In [53] a parameter estimation phase is carried out solving a nonlinear least squares problem. Moreover, also in [47], a similar approach is presented. The authors, there, used the IESZ model to analyze the spread of a fake news, spread on X (Twitter) with the hashtag #DCBlackout, related to possible interruption of communications in the city of Washington D.C on June 1st, 2020 due to the Black Lives Matter movement. In particular, the IESZ model is employed using a parameter estimation strategy that allows to compute optimized parameters, such that the error between real data and the ones computed using the ODE_S solvers routine of MATLAB is minimum. However, in the case study presented in literature, all data are used for the fitting phase. So, in that case, it is possible just to describe the evolution of a news over time, but it is not possible to have predictions. For this reason, we follow two strategies. First of all we check that the IESZ model is suitable for describing information diffusion, employing it to describe the news trend of different case studies, other than the already mentioned #DCBlackout. Then, we will try to carry out the parameter estimation phase using just few data to compute optimized parameters, trying to make real predictions, and employing the remaining part of the dataset just to check if the outcomes are good or if they need to be improved. Note also that in our numerical experiments the IESZ model will be solved using the MATLAB ODE_S solver routine *ode45*, since we want to focus on the parameter estimation phase. Nevertheless, also other numerical methods can be used, as e.g. structure-preserving methods with good stability properties presented in [18,19,22,25] and reference therein.

3.1 Case Studies and Dataset Construction Process

For our purposes, firstly, we need to extract data from a social network and build datasets. Using tools as the *Twitter Developer Option* (www.developer.twitter.com) or the Python data scraper *Tweepy* (https://docs.tweepy.org/en/latest/), it has been possible to collect data from the social network X (Twitter) regarding news spread during the period 2020–2022. In particular, data extracted consist of qualitative information regarding a news spread with a certain hashtag: among these, the ID of users who shared a tweet, the date when they tweeted, the text of the tweet they shared. Then, in order to organize data in such a way that the acquired information can be used for our aims, we build our datasets, by modifying the algorithm construction process described in [10], where a similar process has been made to employ a model of ISR type, and organizing data in minutes.

The following case studies will be taken in considerations:

- #DCBlackout: the already mentioned fake news regarding possible trubles due to Black Lives Matter movement [47];
- #NoVaxDjokovic: fake news regarding a declaration of a "No-Vax" position by the serbian tennis player Novak Djokovic, after his refusal of getting vaccinated against the COVID-19 virus in order to play the Australian Open 2022 tennis tournament [12];
- #NovakDjokovic: true pieces of news spread on X with this hashtag after the 2022 edition of the Australian Open tennis tournament regarding the tennis player Novak Djokovic;
- #Macron: true news spread after Emmanuel Macron was reelected President of France in 2022. Note that tweets regarding this piece of news have also been collected by searching shared tweets containing the phrase *"Emmanuel Macron réélù Président de la République"*.

Observe that since the model we want to employ is of IESZ type, where we consider classes of individuals that have seen a news and have been possibly influenced by it but that have not shared anything yet, it is possible to determine numerically data only for the Spreader class. Data for all the other classes, especially at the starting time of the observation of the phenomenon, have to be optimized as well as the model parameters, as we will show in the next section.

3.2 Parameter Estimation

As anticipated, we will carry out the phase of numerical experiments twice, following two different strategies. In particular:

- *first strategy*: it consists of using the entire dataset for parameter estimation. In this way, the reliability of the IESZ model for describing different kinds of news will be tested. We will make simulations also using the #DCBlackout news. In fact, this piece of news has already been analyzed in [47]. However, tweets can be modified or removed. So, it is reasonable to think that our dataset is slightly different from the one used in [47], due to the fact that our data were extracted after a year since that work had been released.
- *second strategy*: it consists of using just a part of the dataset to compute optimized parameters. After that, solving again the model with them, the accuracy of our predictions will be checked.

Note that, during both the two numerical experiments, all the model parameters and the initial values for ignorants I_0, exposed E_0 and skeptics Z_0 will be optimized. The initial value for spreaders, on the contrary, is taken from the initial dataset, while for the population size, which is assumed as constant in the model, a simplifying assumption is made. We will assume it equal, during the entire period of the observation of the phenomenon, to the sum of the initial values of all the involved classes.

Regarding the parameter estimation process, it is carried out by solving a nonlinear least squares problem. In fact, we compute the set of optimized parameters as the value that minimizes the following function:

$$f(param) = \sum_{j=1}^{n} (S_j - S_{data}(t_j))^2, \tag{4}$$

where $param$ is a vector containing the parameters to be optimized, S_j is the number of spreaders at time t_j computed by solving the IESZ model, while $S_{data}(t_j)$ is the corresponding value for spreaders taken from the dataset. From an implementation point of view, the fitting phase will be carried out using the MATLAB built in function *lsqnonlin*. It solves nonlinear least squares problems allowing users to choose two kind of algorithms: the *Levemberg-Marquardt* algorithm and the *trust-region-reflective* algorithm. In particular, we will use the second one. This algorithm, in fact, permits to find the set of parameters that minimizes a function within a certain region of the space where parameters are defined (see [52] for more details). In this way, it allows to choose a lower and an upper bound for the parameters to be optimized, giving us the possibility to avoid meaningless physical situations in which negative parameters or population values are obtained.

3.3 Numerical Results

Table 1. Optimized parameters and initial conditions for the IESZ model, obtained when all data are used for the fitting phase.

News	β	b	p	l	ρ	ϵ	$I0$	$E0$	$Z0$
#DCBlackout	0.9451	0.36684	0.96331	0.96309	0.010493	0.031694	300000	2222	90775
#NoVaxDjokovic	0.26669	0.093984	0.7038	0.97469	0.01	0.01	186850	100	35786
#NovakDjokovic	0.10003	0.019876	0.012391	0.0123981	0.010017	0.01010	132240	809	215610
#Macron	0.18513	0.900036	0.29056	1	0.024875	0.025308	110910	2803	295630

Table 2. Optimized parameters and initial conditions for the IESZ model, obtained few data are used for the fitting phase.

News	β	b	p	l	ρ	ϵ	$I0$	$E0$	$Z0$
#DCBlackout	0.59597	0.66732	0.99979	0.94191	0.010003	0.043369	210310	1491	4013
#NoVaxDjokovic	0.33615	0.25249	0.011234	0.9665	0.11749	0.033041	153950	197	1000
#NovakDjokovic	0.13323	0.073073	0.011967	0.99924	0.5886	0.010044	144140	465	9197
#Macron	0.99999	0.93747	1	1	0.010427	0.017207	200580	1391	24163

Results regarding the first strategy are summarized in Table 1 and displayed in Fig. 4, while results regarding the second strategy are summarized in Table 2 and displayed in Fig. 5.

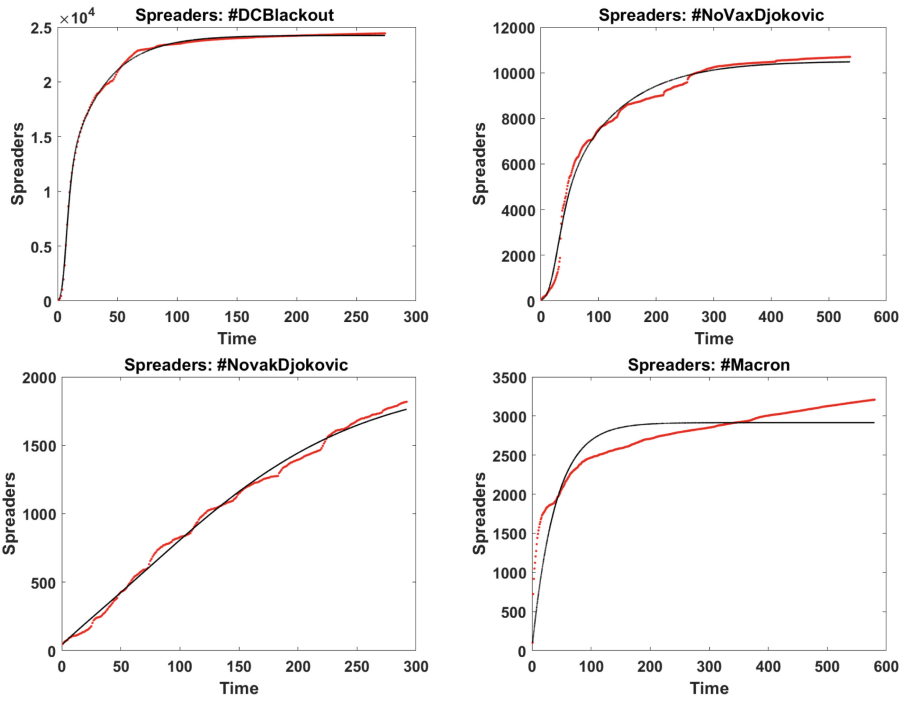

Fig. 4. Time evolution of news trend for all the case studies under consideration. Note •: data used for the fitting phase; **-**: estimated number of spreaders (Color figure online)

Observing results obtained using all data for the parameters estimation, it can be seen that the IESZ model is useful to describe the evolution of news spread over time, since the fashion of the curve of data extracted from X is similar, or in some cases even overlying, the one obtained solving the problem with the MATLAB ODE_S solver routine *ode45*. However, also the second strategy seem to be applicable. Even if there is an underestimation in the number of spreaders, due to oscillations and inaccuracies of the dataset caused by the fact that we are not analyzing the text of extracted tweets but we only consider the presence in them of the hashtag linked to the news, the estimated curve is similar to the real one. Note that in particular for fake news, as the one spread with hashtag #NoVaxDjokovic, the underestimation is more evident, despite the great number of data (the green $*$) used for the fitting phase. This is explainable by considering that this fake news is linked to a relevant figure of a worldwide followed sport. So, it is reasonable that in the first phases of the diffusion of the news, the number of spreaders rises quickly, thanks also to external influences as televisions or newspapers that publish news related to the figure of the tennis player, creating hype and increasing possible fake news regarding his opinion on vaccines. Nevertheless, in all the analyzed cases, the trend of the data is still respected.

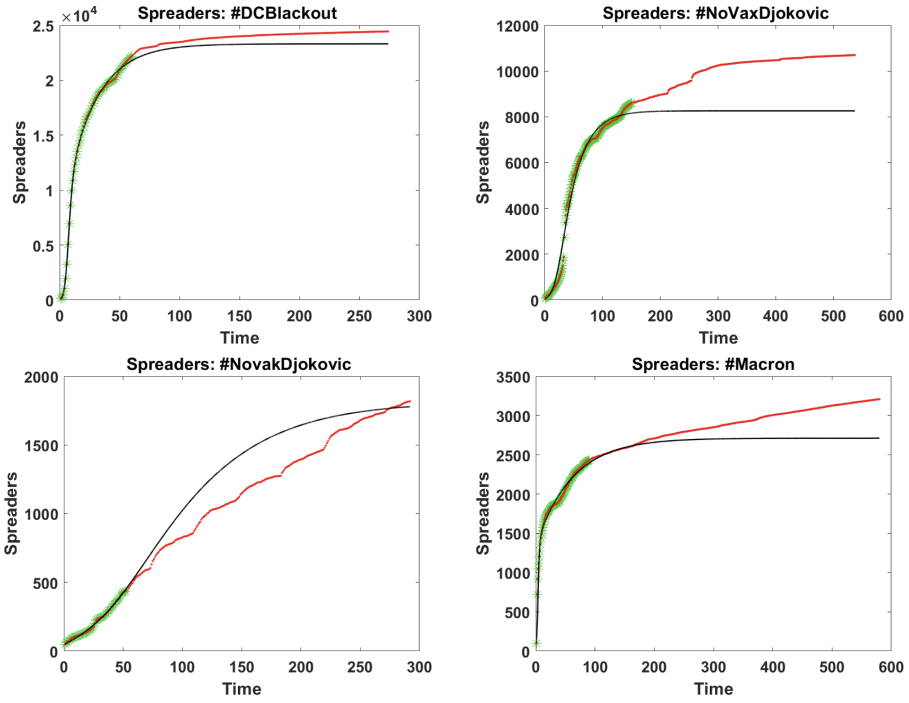

Fig. 5. Time evolution of news trend for all the case studies under consideration. Note ∗: data used for the fitting phase •: data used for predictions; -: estimated number of spreaders (Color figure online)

Table 3. Relative errors in norm two on spreaders obtained employing the second strategy for parameter estimation for all the news items

News	Error
#DCBlackout	0.0412
#NoVaxDjokovic	0.1831
#NovakDjokovic	0.1519
#Macron	0.1073

To have a quantitative representation of the performance of this second strategy, Table 3 can be considered, where for each of the news items, the value in norm two of the relative error computed on all the available data for spreaders is reported.

As it can be observed, this strategy can truly be used to describe the initial phase of a news spread, both in the case of true or fake news. This feature is not trivial: in fact, before the news is spread and analyzed by experts, we do not know if it is real or not, so at least in this initial phase of the spread, it is

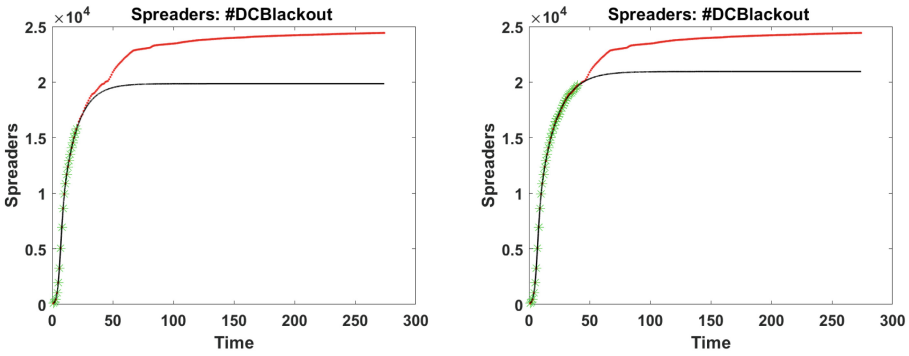

Fig. 6. Tests of parameter estimation carried out on the #DCBlackout news using data related to, respectively, 20 and 40 instants of time.

fundamental to model the phenomenon of dissemination of information by using an efficient model that can be adapted whatever type of news is considered.

Note, also, that in the case in which the second strategy is used, it becomes essential to determine the minimum number of data from the dataset to be employed for parameter estimation. For the news being analyzed, in fact, we use a number of data variable from 25 to 150 instants of time, corresponding to values from 10% to 30% of the dataset. Specifically, the minimum number of data that allows us to obtain good predictions, can be determined employing several approaches. In particular, we decided to avoid using a greater number of values for a simple reason. Since we are interested in predictions, it would not be very useful to optimize parameters using a very large part of the dataset. Moreover, in a real scenario, we are interested in predicting the evolution of a news before it is spread, so it is very probable that we will not have much time to observe it and extract data. Thus, starting from available dataset, we determined the number of data to be used for parameter estimation empirically. As example, consider Fig. 6, where taking into account the dataset related to the #DCBlackout news, two more tests of parameter estimation are exploited using respectively data related to 20 and 40 instants of time. In both these two cases, a stronger underestimation of the total number of spreaders can be observed in comparison to the results obtained using 60 instants of time, as showed in the first figure of Fig. 5.

4 Conclusions and Future Works

In this paper a possible use of a mathematical epidemiological model to describe information diffusion on social media has been analyzed. We have showed that the Ignorant-Exposed-Spreader-Skeptic model, when combined with a proper parameter estimation phase, can be applicable to describe the initial phase of the spread of a news, independently if true or false, before that a loss of interest of users towards it is observed.

Many research developments relating to this topic are possible. Certainly, an analysis of the type of tweets collected could lead to a better construction of the dataset and therefore to greater regularity in the data used for fitting. Furthermore, it would be interesting to carry out a description of the phenomenon in which the model presented here is used to describe what happens in the early stages of information diffusion, while a second model is used to describe any future developments. Note, also, that to solve the ODE_S system we employed the MATLAB routine *ode45*, because we want to focus on the parameter estimation strategy. However, several efficient numerical methods for this kind of systems are available in literature [13,18,22,34], so possible developments can consist in the implementation of the methods there described.

Acknowledgments. The authors are members of the research group GNCS-INdAM. This work has been supported by GNCS–INdAM projects and by the Italian Ministry of University and Research (MUR), through the PRIN PNRR 2022 project BAT-MEN (P20228C2PP, CUP: F53D23010020001).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aceto, L., Conte, D., Pagano, G.: On a generalization of time-accurate and highly-stable explicit operators for stiff problems. Appl. Numer. Math. **200**, 2–17 (2024). https://doi.org/10.1016/j.apnum.2023.04.001
2. Aceto, L., Conte, D., Pagano, G.: Modified TASE Runge-Kutta methods for integrating stiff differential equations. SIAM J. Sci. Comput. **47**(3), A1652–A1680 (2025). https://doi.org/10.1137/24M1667336
3. Amodio, P., Brugnano, L., Frasca-Caccia, G., Iavernaro, F.: Arbitrarily high-order energy-conserving methods for Hamiltonian problems with quadratic holonomic constraints. J. Comput. Math. **42**(4), 1145–1171 (2024). https://doi.org/10.4208/jcm.2301-m2022-0065
4. Benzi, M., Boito, P.: Matrix functions in network analysis. GAMM-Mitteilungen **43**(3), e202000012 (2020). https://doi.org/10.1002/gamm.202000012
5. Buckwar, E., D'Ambrosio, R.: Exponential mean-square stability properties of stochastic linear multistep methods. Adv. Comput. Math. **47**(4), 1–14 (2021). https://doi.org/10.1007/s10444-021-09879-2
6. Cardone, A., Conte, D., Paternoster, B.: Stability of two-step spline collocation methods for initial value problems for fractional differential equations. Commun. Nonlinear Sci. Numer. Simul. **115**, 106726 (2022). https://doi.org/10.1016/j.cnsns.2022.106726
7. Cardone, A., Diaz de Alba, P., Paternoster, B.: Analytical properties and numerical preservation of an age-group susceptible-infected-recovered model: application to the diffusion of information. J. Comput. Nonlinear Dyn. **19**(6), 061006 (2024). https://doi.org/10.1115/1.4065437
8. Cardone, A., Frasca-Caccia, G.: Numerical conservation laws of time fractional diffusion PDEs. Fract. Calc. Appl. Anal. **25**(4), 1459–1483 (2022). https://doi.org/10.1007/s13540-022-00059-7

9. Casillo, M., Colace, F., Conte, D., De Santo, M., Lombardi, M., Mottola, S., Santaniello, D.: A multi-feature Bayesian approach for fake news detection. In: Chellappan, S., Choo, K.-K.R., Phan, N.H. (eds.) CSoNet 2020. LNCS, vol. 12575, pp. 333–344. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-66046-8_27
10. Castiello, M., Conte, D., Iscaro, S.: Using epidemiological models to predict the spread of information on Twitter. Algorithms **16**(8), 391 (2023). https://doi.org/10.3390/a16080391
11. Citro, V., D'Ambrosio, R.: Long-term analysis of stochastic theta-methods for damped stochastic oscillators. Appl. Numer. Math. **150**, 18–26 (2020). https://doi.org/10.1016/j.apnum.2019.08.011
12. CNN (Cable News Network). https://edition.cnn.com/2022/01/20/sport/djokovic-serbia-national-hero-intl-spt/index.html. Accessed 1 Mar 2025
13. Conte, D., D'Ambrosio, R., Jackiewicz, Z.: Two-step Runge-Kutta methods with quadratic stability functions. J. Sci. Comput. **2**, 191–218 (2010). https://doi.org/10.1007/s10915-010-9378-x
14. Conte, D., D'Ambrosio, R., Paternoster, B.: Improved θ-methods for stochastic Volterra integral equations. Commun. Nonlinear Sci. Numer. Simul. **93**, 105528 (2021). https://doi.org/10.1016/j.cnsns.2020.105528
15. Conte, D., Frasca-Caccia, G.: Exponentially fitted methods that preserve conservation laws. Commun. Nonlinear Sci. Numer. Simul. **109**, 106334 (2022). https://doi.org/10.1016/j.cnsns.2022.106334
16. Conte, D., Frasca-Caccia, G.: Exponentially fitted methods with a local energy conservation law. Adv. Comput. Math. **49**(4), 49 (2023). https://doi.org/10.1007/s10444-023-10049-9
17. Conte, D., González-Pinto, S., Hernández-Abreu, D., Pagano, G.: On approximate matrix factorization and TASE W-methods for the time integration of parabolic partial differential equations. J. Sci. Comput. **100**(2), 34 (2024). https://doi.org/10.1007/s10915-024-02579-1
18. Conte, D., Guarino, N., Pagano, G., Paternoster, B.: Positivity-preserving and elementary stable nonstandard method for a COVID-19 SIR model. Dolomites Res. Notes Approx. **15**(5), 65–77 (2022). https://doi.org/10.14658/pupj-drna-2022-5-7
19. Conte, D., Guarino, N., Pagano, G., Paternoster, B.: On the advantages of nonstandard finite difference discretizations for differential problems. Numer. Anal. Appl. **15**(3), 219–235 (2022). https://doi.org/10.1134/S1995423922030041
20. Conte,D., Iscaro,S., Pagano,G., Paternoster,B.: Adapted numerical modeling for fake news diffusion on social networks. Submitted
21. Conte, D., Martin-Vaquero, J., Pagano, G., Paternoster, B.: Stability theory of TASE–Runge–Kutta methods with inexact Jacobian. SIAM J. Sci. Comput. **46**(6), A3628–A3657 (2024). https://doi.org/10.1137/24M1631869
22. Conte, D., Pagano, G., Paternoster, B.: Nonstandard finite differences numerical methods for a vegetation reaction–diffusion model. J. Comput. Appl. Math. **419**, 114790 (2023). https://doi.org/10.1016/j.cam.2022.114790
23. Conte, D., Pagano, G., Paternoster, B.: Two-step peer methods with equation-dependent coefficients. Comput. Appl. Math. **41**(4), 140 (2022). https://doi.org/10.1007/s40314-022-01844-z
24. Conte, D., Pagano, G., Paternoster, B.: Time-accurate and highly-stable explicit peer methods for stiff differential problems. Commun. Nonlinear Sci. Numer. Simul. **119**, 107136 (2023). https://doi.org/10.1016/j.cnsns.2023.107136
25. D'Ambrosio, R.: Numerical Approximation of Ordinary Differential Problems: From Deterministic to Stochastic Numerical Methods. Springer (2023). https://doi.org/10.1007/978-3-031-31343-1

26. D'Ambrosio, R., Díaz de Alba, P., Giordano, G., Paternoster, B.: A modified SEIR model: stiffness analysis and application to the diffusion of fake news. In: Gervasi, O., Murgante, B., Hendrix, E.M.T., Taniar, D., Apduhan, B.O. (eds.) ICCSA, 90-103, Springer International Publishing (2022). https://doi.org/10.1007/978-3-031-10522-7_7
27. D'Ambrosio, R., Di Giovacchino, S.: Mean-square contractivity of stochastic theta-methods. Comm. Nonlin. Sci. Numer. Simul. **96**, 105671 (2021). https://doi.org/10.1016/j.cnsns.2020.105671
28. D'Ambrosio, R., Di Giovacchino, S.: Nonlinear stability issues for stochastic Runge-Kutta methods. Comm. Nonlin. Sci. Numer. Simul. **94**, 105549 (2021). https://doi.org/10.1016/j.cnsns.2020.105549
29. D'Ambrosio, R., Di Giovacchino, S.: Strong backward error analysis of symplectic integrators for stochastic Hamiltonian systems. Appl. Math. Comput. **467**, 128488 (2024). https://doi.org/10.1016/j.amc.2023.128488
30. D'Ambrosio, R., Di Giovacchino, S.: Long-term analysis of stochastic Hamiltonian systems under time discretizations. SIAM J. Sci. Comput. **45**(2), A257–A288 (2023). https://doi.org/10.1137/21M1458612
31. D'Ambrosio, R., Di Giovacchino, S.: Numerical conservation issues for the stochastic Korteweg-de Vries equation. J. Comput. Appl. Math. **424**, 114967 (2023). https://doi.org/10.1016/j.cam.2022.114967
32. D'Ambrosio, R., Giordano, G., Mottola, S., Paternoster, B.: Stiffness analysis to predict the spread out of fake information. Future Internet **13**(9), 222 (2021). https://doi.org/10.3390/fi13090222
33. D'Ambrosio, R., Giordano, G., Paternoster, B., Ventola, A.: Perturbative analysis of stochastic Hamiltonian problems under time discretizations. Appl. Math. Lett. **120**, 107223 (2021). https://doi.org/10.1016/j.aml.2021.107223
34. D'Ambrosio, R., Izzo, G., Jackiewicz, Z.: Search for highly stable two-step Runge-Kutta methods for ODEs. Appl. Numer. Math. **62**(10), 1361–1379 (2012). https://doi.org/10.1016/j.apnum.2012.06.012
35. Giordano, G., Mottola, S., Paternoster, B.: A short review of some mathematical methods to detect fake news. Circuits Syst. Signal Process **14**, 255–265 (2020). https://doi.org/10.46300/9106.2020.14.37
36. D'Ambrosio, R., Paternoster, B.: Multivalue collocation methods free from order reduction. J. Comput. Appl. Math. **387**, 112515 (2021). https://doi.org/10.1016/j.cam.2019.112515
37. Frasca-Caccia, G.: Finite difference schemes with non polynomial local conservation laws. J. Comput. Appl. Math. **458**, 116330 (2025). https://doi.org/10.1016/j.cam.2024.116330
38. Frasca-Caccia, G., Hydon, P.E.: A new technique for preserving conservation laws. Found. Comput. Math. **22**(2), 477–506 (2022). https://doi.org/10.1007/s10208-021-09511-1
39. Frasca-Caccia, G., Hydon, P.E.: Simple bespoke preservation of two conservation laws. IMA J. Numer. Anal. **40**(2), 1294–1329 (2020). https://doi.org/10.1093/imanum/dry087
40. Frasca-Caccia, G., Hydon, P.E.: Numerical preservation of multiple local conservation laws. Appl. Math. Comput. **403**, 126203 (2021). https://doi.org/10.1016/j.amc.2021.126203
41. Frasca-Caccia, G., Hydon, P.E.: Locally conservative finite difference schemes for the modified KdV equation. J. Comput. Dyn. **6**(2), 307–323 (2019). https://doi.org/10.3934/jcd.2019015

42. Frasca-Caccia, G., Singh, P.: Optimal parameters for numerical solvers of PDEs. J. Sci. Comput. **97**(1), 11 (2023). https://doi.org/10.1007/s10915-023-02324-0
43. Frasca-Caccia, G., Valentino, C., Colace, F., Conte, D.: An overview of differential models for corrosion of cultural heritage artefacts. Math. Model. Nat. Phenom. **18**, 27 (2023). https://doi.org/10.1051/mmnp/2023031
44. González-Pinto, S., Hernández-Abreu, D., Pagano, G., Pérez-Rodríguez, S.: Generalized TASE-RK methods for stiff problems. Appl. Numer. Math. **188**, 129–145 (2023). https://doi.org/10.1016/j.apnum.2023.03.007
45. Lazer, D., et al.: The science of fake news. Science **359**(6380), 1094–1096 (2018). https://doi.org/10.1126/science.aao2998
46. Mahdikhani, M.: Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of Covid-19 pandemic. Int. J. Inf. Manage. Data Insights **2**(1), 100053 (2022). https://doi.org/10.1016/j.jjimei.2021.100053
47. Maleki, M., Mead, E., Arani, M., Agarwal, N.: Using an epidemiological model to study the spread of misinformation during the black lives matter movement. Preprint arXiv:2103.12191 (2021). https://doi.org/10.48550/arXiv.2103.12191
48. Martcheva, M.: An Introduction to Mathematical Epidemiology. TAM, vol. 61. Springer, Boston, MA (2015). https://doi.org/10.1007/978-1-4899-7612-3
49. Millevoi, C., Pasetto, D., Ferronato, M.: A physics-informed neural network approach for compartmental epidemiological models. PLoS Comput. Biol. **20**(9),e1012387 (2024). https://doi.org/10.1371/journal.pcbi.1012387
50. Muhlmeyer, M., Agarwal, S.: Information Spread in a Social Media Age: Modeling and Control. CRC Press (2021). https://doi.org/10.1201/9780429263842
51. Pagano, G.: Stabilized explicit peer methods with parallelism across the stages for stiff problems. Appl. Numer. Math. **207**, 156–173 (2025). https://doi.org/10.1016/j.apnum.2024.08.023
52. Quarteroni, A., Saleri, F., Gervasio, P.: Metodi numerici per problemi ai limiti stazionari ed evolutivi. In: Calcolo Scientifico. U, vol. 105, pp. 373–433. Springer, Milano (2017). https://doi.org/10.1007/978-88-470-3953-7_9
53. Treibert, S. M.: Mathematical Modelling and Nonstandard Schemes for the Corona Virus Pandemic. Springer Spektrum (2021). https://doi.org/10.1007/978-3-658-35932-4
54. Valentino, C., Pagano, G., Conte, D., Paternoster, B., Colace, F., Casillo, M.: Step-by-step time discrete physics-informed neural networks with application to a sustainability PDE model. Math. Comput. Simul. **230**, 541–558 (2025). https://doi.org/10.1016/j.matcom.2024.10.043
55. Wang, H., Wang, F., Xu, K.: Modeling information diffusion in online social networks with partial differential equations, vol. 7. Springer Nature (2020). https://doi.org/10.1007/978-3-030-38852-2