Article

An Extension of the Susceptible–Infected Model and Its Application to the Analysis of Information Dissemination in Social Networks

Sergei Sidorov [1,*,†] , Alexey Faizliev [1,†] and Sophia Tikhonova [2,†]

1   Faculty of Mathematics and Mechanics, Saratov State University, 410012 Saratov, Russia; faizlievar@sgu.ru
2   Philosophy Department, Saratov State University, 410012 Saratov, Russia; segedasv@yandex.ru
*   Correspondence: sidorovsp@info.sgu.ru
†   These authors contributed equally to this work.

Abstract: Social media significantly influences business, politics, and society. Easy access and interaction among users allow information to spread rapidly across social networks. Understanding how information is disseminated through these new publishing methods is crucial for political and marketing purposes. However, modeling and predicting information diffusion is challenging due to the complex interactions between network users. This study proposes an analytical approach based on diffusion models to predict the number of social media users engaging in discussions on a topic. We develop a modified version of the susceptible–infected (SI) model that considers the heterogeneity of interactions between users in complex networks. Our model considers the network structure, abandons the assumption of homogeneous mixing, and focuses on information diffusion in scale-free networks. We provide explicit algorithms for modeling information propagation on different types of random graphs and real network structures. We compare our model with alternative approaches, both those considering network structure and those that do not. The accuracy of our model in predicting the number of informed nodes in simulated information diffusion networks demonstrates its effectiveness in describing and predicting information dissemination in social networks. This study highlights the potential of graph-based epidemic models in analyzing online discussion topics and understanding other phenomena spreading on social networks.

1. Introduction

In recent years, social media has come to dominate as a major source of rapid dissemination of information, gaining immense popularity among a huge number of online users. The influence of social media on business, politics, and society has becoming increasingly significant. Thanks to easy access and interaction among many users, information spreads epidemically across social networks. Understanding the mechanisms of information dissemination through these new publishing methods is important for political and marketing purposes. However, due to the high complexity of interactions between network users, it is still a significant challenge to adequately model the information dissemination processes and then accurately predict the information diffusion. This study proposes an analytical approach based on diffusion models to predict the number of social media users who engage in discussion on a topic. In this paper, we develop a modified version of the susceptible–infected (SI) model of block approximation for node degrees to predict the information diffusion process at all its stages. The modification considers the heterogeneity of interactions between social network users and is expressed by a system of nonlinear differential equations on complex networks. The proposed model differs from other similar models by three main features: it considers the network structure, it abandons the homogeneous mixing hypothesis, and it focuses on information diffusion in scale-free

networks. The paper provides explicit algorithms for modeling information propagation on various types of random graphs, as well as on real network structures. A comparison is also provided with alternative information dissemination approaches, both those that take into account network structure and those that do not. The number of informed nodes of the simulated information diffusion networks was compared with the number of informed nodes in a simulation-based approach to determine which information diffusion model provides the best fit. The high accuracy of the results showed that the proposed model is able to describe and predict the process of information dissemination in social networks. This study shows that graph-based epidemic models can expand their application to on-line discussion topics and can also help in understanding other phenomena spreading on social networks.

In recent years, attention to the study and analysis of the information diffusion in complex networks has increased significantly. One of the most popular areas of research is the modeling of the information diffusion in social and media networks. One important problem is to study the temporal dynamics for the total number of users who received a message on a specific topic. The curve describing the information diffusion in social networks is usually an S-shaped line as shown in [1]. The curve can be conditionally divided into three main sections, which correspond to three stages of the diffusion process. The process of information propagation in social networks can be described in terms of the epidemic diffusion.

At the first stage, only a small number of social network users have some specific information. They are usually called innovators. The innovators then distribute this information by posting it on their social media pages. The information becomes available to subscribers and, first of all, to their closest contacts. Then, those users newly infected with this information (topics, news, memes) continue to distribute information further on the social network via their contacts and subscribers. At this stage (outbreak), the curve is slowly growing.

At the second stage, the growth rate of the curve increases significantly, passing into the exponential mode. At this stage, there is the largest increase in the number of users who received the information. The increase in the growth speed of the information diffusion is high due to the large number of susceptible users who have not yet received this information. Moreover, by this point, a sufficient number of infected peaks have already appeared, among which there are super-spreaders.

Then, the curve goes into the so-called saturation mode, which implies a significant decrease in the speed of information diffusion, caused by the fact that most network users have already received this information and/or have lost interest in it.

Note that such processes also arise during the propagation of infectious diseases as well as the spread of innovations. For example, the growth of the mobile market in different countries is modeled as the innovation diffusion process using the logistic model in studies of [2–7].

Over the course of the past fifty years, many different diffusion models have been developed and studied, including the Bass model and the Gompertz model in the works of [8,9] and others. Scientists have widely used these models and their modifications to analyze various phenomena in [10–13]. An important step of a study is to compare several models to examine a specific dataset, identifying the advantages and disadvantages of each model (as in [14]). An overview of innovation diffusion models based on the agent approach can be found in the work of [15]. In recent years, one of the main topics in social sciences has been the study of the information diffusion in social networks and communities (see, e.g., the works of [16–20]).

Social networks have become a convenient object of study due to the provision of unprecedented amounts of data from network users. This has stimulated the emergence of many works on the information diffusion dynamics in social networks. Many previous studies of the information dissemination in social communities were based either on experimental analysis or on the solution of the rate equations describing the temporal dynamics of the diffusion process. In addition, some recent studies by [21,22] have de-

veloped a methodology for modeling information diffusion processes using heat transfer equations, reaction–diffusion equations, or hydrodynamic equations.

Unfortunately, the existing models focus mainly on modeling the information diffusion in the early stages. In our study, we primarily focus on creating a model that adequately describes the behavior not only at the outbreak stage but also at later periods of the information diffusion processes. In addition, previous studies have shown that it is necessary to take into account the network structure when modeling the diffusion process via differential equations. In this regard, this paper proposes the model that takes into account the network structure. Therefore, we reject the homogeneous mixing hypothesis and concentrate on the diffusion of information in scale-free networks.

The paper proposes a SI-type model on graphs which aims to describe information diffusion processes in complex social networks. The model takes into account the network structure. Moreover, the model abandons the homogeneous mixing hypothesis and concentrates on the information diffusion in scale-free networks. The proposed model is compared with other known models both on random graphs of two types (Erdős–Rényi, Barabási–Albert) and on real social networks. The agent-based simulation model of information diffusion is taken as a benchmark model. The empirical results show that the proposed model more adequately predicts behavior at all stages of the information diffusion process.

The work contains three sections. The classical SI information diffusion model is described in Section 2. The SI information diffusion model on graphs is presented in Section 3. Section 4 includes a description of the data and empirical results obtained using the considered models.

2. SI Model

To analyze the information diffusion, three frequently used logistic models are usually considered, the so-called susceptible–infected (SI), susceptible–infected–susceptible (SIS), and susceptible–infected–recovered (SIR) models, which elucidate the basic building blocks of information diffusion modeling [23]. The aim of all of these models is calculate an S-curve that reflects the information diffusion among specific groups of social media users.

The SI model is the most key of all segmental models used to describe the information diffusion in social networks. The SI model was first used by Griliches [24], who applied the logistic model to explain the widespread use of hybrid corn in the US. Refs. [25–28] considered this study as a model for using the logistic model in their works.

According to this model, social network users (who have participated in the discussion of a certain topic) can be classified into two categories: a group that starts disseminating information (for example, publishing articles or posts on a certain topic using unique hashtags) and a group that has not yet received information and has not yet participated in its distribution. The SI model of logistic growth is represented by the following differential equation:

$$\frac{dI(t)}{dt} = \beta \cdot \langle k \rangle \cdot \frac{S(t)I(t)}{N},$$
(1)

where

- N is the number of network users;
- $S(t)$ is the number of users who are susceptible to a given topic and have not yet received a news message and/or its derivative publications, i.e., not possessing the information at time t (susceptible);
- $I(t)$ is the number of users who received a news message and/or its derivative publications on a given topic at time t, i.e., received information and continued its dissemination (informed);
- β is the information diffusion probability, i.e., the probability that a network user who has received a message will be interested in its topic and will broadcast it or its derivative form to other network users, per unit of time;
- $\langle k \rangle$ is the average number of contacts of social network users.

$s = s(t) = S(t)/N$ and $i = i(t) = I(t)/N$ denote the shares of susceptible and informed users, respectively, at time t. Then, Equation (1) can be rewritten as follows:

$$\frac{di}{dt} = \beta\langle k\rangle si = \beta\langle k\rangle i(1-i),\qquad(2)$$

where the product $\beta\langle k\rangle$ is essentially the rate of the information diffusion.

The solution of this first-order differential equation with initial condition $i_0 = i(0)$ is

$$i = i(t) = \frac{i_0 e^{\beta\langle k\rangle t}}{1 - i_0 + i_0 e^{\beta\langle k\rangle t}}\qquad(3)$$

Equation (3) predicts that:

- At the beginning, the proportion of users who received the news message increases exponentially. Indeed, at an early stage, the informed user encounters only the receptive, so information can be easily disseminated.
- The characteristic time required to reach the share of $1/e$ (about 36%) of all susceptible individuals is

$$\tau = \frac{1}{\beta\langle k\rangle}\qquad(4)$$

Therefore, the value of τ is inversely proportional to the rate with which information is distributed among the network users. Thus, it follows from (4) that an increase in either the link density $\langle k\rangle$ or β increases the rate of diffusion information and reduces the characteristic time.

- Over time, a user who has received information (message) on a given topic broadcasts it further to a progressively smaller number of susceptible users. Consequently, the growth of i slows down for large t. The information diffusion ends when everyone is informed, i.e., when $i(t \to \infty) = 1$ and $s(t \to \infty) = 0$.

3. Information Diffusion on Networks

It should be noted that the models discussed above do not take into account the structure of the contact network in which the information is disseminated. The main assumption of such models is that any network user can infect any other user with some topic (homogeneous mixing hypothesis) and that all users have a comparable number of contacts (friends, subscribers). However, in real social networks, these assumptions are incorrect: a message posted by a social network user is available to their friends (subscribers); therefore, information is distributed through a complex network of contacts. In addition, such contact networks are usually scale-free, so the average number of connections (i.e., the average degree) in the networks is not enough to characterize their topology.

The failure of these assumptions has recently caused a fundamental revision of the basis for modeling the information dissemination in social networks. This change began with the work [29] that extended the basic epidemic diffusion models (which are easily portable to the case of information dissemination), taking into account the topological characteristics of the network. The information diffusion, as well as the best conditions for the emergence of super-spreaders, has been largely explored in network science via models of the spread of dangerous viruses and pathogens (as well as disinformation and gossip) in the paper [30]. Recently, Reference [31] has investigated the influence of network community structure on percolations that model the diffusion of an epidemic using the classical SIR epidemic model. The work [32,33] has concluded that intra-community diffusion is critically related to network density, and community structure is the most important factor for the spread of an epidemic, regardless of community size and shape. In addition, Reference [34] has studied an epidemic spread using the graph adjacency matrix.

Reference [35] presented a new SIR model on social networks for studying the spread of rumors which makes it possible to explicitly take into account the location of objects, as well as the connections and interactions between objects. Experiments in this model

have shown that the existence of a networked environment shortens the diffusion time of rumors. However, in the model, the authors neglected the connections of the network environment, which strongly influence the process of spreading rumors.

The study [36] improved the SIR model by using an integrated methodology to model the dissemination of opinions and ideas in web forums. This updated model was validated on a large dataset from a large retail company web forum as well as on a dataset from a general political discussion forum. The experimental results also showed that this updated SIR model works well with the spread of topics on web forums.

The work [37] studied the mechanism of information dissemination using heat transfer methods together with connections between network nodes to model the dissemination of information in social networks. The model proposed in [37] depends on network structures, since the proposed mechanisms are determined by the states and degrees of network nodes.

The study [21] presented a model of information dissemination based on the principles of the heat transfer process. It assumed that the diffusion mechanism of the model depends on how many times a particular node interacts with its neighbors. Experiments on a real social media dataset have demonstrated the effectiveness of the model. The authors also developed algorithms to find the top k nodes that were influenced by marketing information. The limitation of the model is that it does not take into account the friendships of users when disseminating information.

This study proposes SI-type models and their modifications to study the patterns of the information dissemination process on various types of random graphs, as well as on real social network data. We are interested in developing new models on graphs that could take into account the features of their topology.

3.1. SI Model on Graphs

To build the SI model on the network, we used Equation (2) of the classical SI model, which does not take into account the network structure. Note that in the process of information diffusion over the network, people with a large number of connections are more likely to come into contact with an informed person; therefore, they are more likely to obtain the information. Thus, the mathematical formalism must consider the degree of each node as an implicit variable. This can be achieved with the use of the degree block approximation, which distinguishes nodes based on their degrees and assumes that nodes with the same degree are statistically equivalent. Therefore,

$$i_k = \frac{I_k}{N_k} \tag{5}$$

denotes share of informed nodes with degree k among all N_k nodes with degree k in the network. The total share of informed nodes is equal to the sum of all such nodes in all blocks, i.e.,

$$i = \sum_k p_k i_k, \tag{6}$$

where p_k is the probability of choosing the node k.

Taking into account the notation, we can present the SI model for the block of nodes with degree k as follows:

$$\frac{di_k}{dt} = \beta(1 - i_k)k\theta_k. \tag{7}$$

Equation (7) has almost the same structure as the main equation of the classical SI model: the level of dissemination of information is proportional both to β and the share of not yet informed nodes with degree k, i.e., to $1 - i_k$. However, there are a few key differences:

- The average degree $\langle k \rangle$ in (2) is replaced by the actual degree k of each node in (7).
- The density function θ_k represents the proportion of informed neighbors of a susceptible node with degree k. Thus, θ_k is just the proportion of i_k nodes that are

informed. However, in a network environment, the proportion of informed nodes in the immediate vicinity of a node may depend on its degree k and time t.

- While (2) describes the behavior of the entire system with a single time-dependent equation, (7) is a system of $k_{\max}$ coupled equations, one equation for each degree k present in the network.

Ref. [29] studied the behavior of i_k in early time periods of diffusion process. In this work, we expand the formalism developed in [29] by considering the dissemination of information at all time stages (periods) of its diffusion over the network.

As discussed earlier, in order to calculate i_k, we must first determine θ_k. If the network is degree–degree dis-assortative, i.e., there are no degree correlations in the network, and the probability that a link leads from a node of degree k to a node of degree k' is independent of k and k'. Therefore, the probability that a randomly selected link of the network points to a node of degree k' is

$$\frac{k' p_{k'}}{\sum_k k p_k} = \frac{k' p_{k'}}{\langle k \rangle}. \tag{8}$$

It is quite natural that at least one link of each informed node is connected to another informed node that transmits information. In this regard, the authors of [29] assume that the number of links available for information transmission in the future periods is $k' - 1$. Then, the density function θ_k can be defined as follows:

$$\theta_k = \frac{\sum_{k'} (k' - 1) p_{k'} i_{k'}}{\langle k \rangle}. \tag{9}$$

It should be noted that such an assumption is rather rough and can take place only at early stages of information dissemination.

Now, we return to Equation (9). Note that in the absence of the degree–degree correlations, $\theta = \theta_k$ will not depend on k. Then, after differentiating (9), we get:

$$\frac{d\theta}{dt} = \sum_k \frac{(k - 1) p_k}{\langle k \rangle} \frac{d i_k}{dt}. \tag{10}$$

Then, it follows from (7) and (10) that

$$\frac{d\theta}{dt} = \sum_k \frac{k(k - 1) p_k}{\langle k \rangle} \beta (1 - i_k) \theta. \tag{11}$$

Obviously, at the early stages of information dissemination (for small t) the share of the informed nodes is much smaller than 1, and the multiplier $1 - i_k$ can be neglected. It is noted in [29] that in this case the share of informed nodes can be found explicitly. In our case, we cannot drop the $1 - i_k$ multiplier.

Thus, we get the following system of k_{max} coupled equations:

$$\begin{cases} \theta = \frac{\sum_k (k-1) p_k i_k}{\langle k \rangle}, \\ \frac{d i_k}{dt} = \beta (1 - i_k) k \theta, \ k = 1, \ldots, k_{\max}. \end{cases} \tag{12}$$

3.2. New Model

In this section, we propose a new model in which it is assumed that the number of links available to transmit information in the future will increase over time as the infectiousness of network users increases. This is in the contrast with the model of [29]. The justification for our assumption is that the more nodes become informed, the more opportunities (channels) arise for them to transmit information in the future. Then, the density function θ_k can be defined as follows:

$$\theta_k = \frac{\sum_{k'} k'(t, \beta) p_{k'} i_{k'}}{\langle k \rangle}. \tag{13}$$

We assume that the number of links $k'(t, \beta)$ available for information transmission in the future can be approximated by a function depending on the time t and the probability of information diffusion β as follows: $k'(t, \beta) = k'(1 - \beta^{c(\beta)t})$, where $c(\beta) = a\beta + b$ linearly depends on β. Coefficients a and b were estimated empirically on one of the graphs $\bar{a} = 2.8$, $\bar{b} = 0.06$.

Thus, we get the following system of $k_{\max}$ coupled equations:

$$\begin{cases} \theta = \frac{\sum_k k(1 - \beta^{c(\beta)t}) p_k i_k}{\langle k \rangle}, \\ \frac{di_k}{dt} = \beta(1 - I_k) k\theta, \ k = 1, \dots, k_{\max}. \end{cases} \tag{14}$$

The algorithm for solving these equations can be easily programmed.

The results of solving Equations (12) and (14), as well as Equation (2), are presented in Section 4 on various types of simulated graphs and real networks.

In contrast to existing models, the proposed model allows the diffusion of information to be modeled, both in the early stages and at later stages of information dissemination. In addition, the proposed model makes it possible to take into account the network structure when modeling differential equations. We abandon the homogeneous mixing hypothesis and concentrate on the diffusion of information on scale-free networks.

3.3. Agent-Based Simulation Model

In order to test the adequacy of the proposed SI models on graphs, this section will consider an approach that is essentially similar to the Monte Carlo method. We will simulate the information propagation on networks using agent-based modeling techniques. This approach involves studying information diffusion over a network, taking into account the behavior of nodes and their influence on neighboring vertices.

$G = (V, E)$ denotes the social network. The nodes of the graph (set V) are agents (or network users) who receive information and decide on its further distribution; the set of edges E are social links through which agents exchange information.

Graph nodes could be in several states. These states are the same as in the logistic growth models discussed earlier. For example, the active state I from the SI model indicates that the node is infected and may transmit an information to neighboring nodes. In the context of social networks, this means that the user posts a message on their page, which their friends or followers can see. We assume that the information dissemination may occur at discrete times. Initially, there is a set of informed graph nodes, for example, those that have received information from external sources. Their subscribers, by reposting their messages, thereby lead to a further dissemination of the information (i.e., activate neighboring nodes), changing the state of the neighboring node from S (susceptible) to I (informed). At the next step, the newly activated nodes are considered, and the neighboring nodes associated with them may be activated with probability p. The process continues until the diffusion is completed. The information propagation rules are determined according to the model of independent cascades: at each step, an activated node v has only one chance to activate an inactive neighbor u with probability p.

4. Testing Models on Random and Real Graphs

We tested the adequacy of our proposed new SI model on graphs. We also considered the predictive power of the classical SI model (which does not take into account the structure of the contact network) and the model on graphs proposed by [29] in their ability to predict the information diffusion on graphs. Simulations of agent-oriented modeling were taken as a benchmark. The proposed models were tested both on random graphs, such as the Erdős–Rényi (ER) graph and the Barabási–Albert (BA) graph, and on real social graphs. Different sizes of graphs were also considered.

For comparison, the following approximating differential Equations (models) were considered:

1. The classical SI model that does not take into account the network structure:

$$\frac{di}{dt} = \beta \langle k \rangle i (1 - i).$$

For brevity, we will call such a model an SI model.

2. The SI model on the network proposed by [29]:

$$\begin{cases} \theta = \frac{\sum_k (k-1) p_k i_k}{\langle k \rangle}, \\ \frac{di_k}{dt} = \beta (1 - i_k) k \theta, \ k = 1, \ldots, k_{\max}. \end{cases}$$

For brevity, we will call such a model SI on the network (netSI).

3. An approximating version of the SI model on graphs, taking into account the change in the number of information transmission channels over time:

$$\begin{cases} \theta = \frac{\sum_k k(1 - \beta^{c(\beta)t}) p_k i_k}{\langle k \rangle}, \\ \frac{di_k}{dt} = \beta (1 - I_k) k \theta, \ k = 1, \ldots, k_{\max}. \end{cases}$$

For brevity, we will call such a model SI on the network with approximation for the number of information transmission channels over time (netSIapprox).

4. An agent-based model. This approach involves simulating the process of information dissemination according to the SI model. For brevity, we will call such a model the "Benchmark". Ten simulations were run.

To compare models, two metrics were used: L_2-norm and L_∞ norm:

$$\| \cdot \|_2 = \sum_{i=1}^{n} (y_i - x_i)^2,$$

$$\| \cdot \|_\infty = \max_{1 \le i \le n} |y_i - x_i|.$$

Simulations of the agent-based model were averaged and taken as a benchmark for calculating the accuracy of the considered models in differential equations. Table 1 lists the characteristics of the considered real graphs for model testing.

Table 1. Characteristics of real graphs.

Graphs/Characteristics	Number of Nodes	Number of Edges	Density	Average Degree	Power Law Exponent γ
Twitch Social Networks (DE)	9498	153,138	0.003	32.25	2.01
Github-social (GS)	37,700	289,003	0.0004	15.33	2.4
Ego-Gplus (EG)	107,614	12,238,285	0.002	227.45	1.35
Large twitch (LT)	168,114	6,797,557	0.0005	80.87	2.23

As shown in Table 1, the considered real graphs are not similar. The only aspect that unites them is the power law of the distribution of degrees.

Table 2 gives estimates of model accuracy for random and real graphs of various sizes and densities.

Table 2. Model accuracy in L_2-norm and L_∞-norm.

Graph/Model	SI			netSI			netSIapprox		
	Random graphs, nodes = 10,000, density = 0.001								
β	0.001	0.005	0.02	0.001	0.005	0.02	0.001	0.005	0.02
ER	0.0032	0.0037	0.0059	0.0006	0.0009	0.0021	0.0047	0.004	0.0026
	0.129	0.138	0.179	0.064	0.073	0.113	0.165	0.154	0.128
BA	0.012	0.0096	0.00494	0.00008	0.00025	0.00194	0.0003	0.00065	0.0006
	0.257	0.242	0.178	0.023	0.041	0.104	0.042	0.057	0.054
	Random graphs, nodes = 25,000, density = 0.0008								
β	0.001	0.005	0.02	0.001	0.005	0.02	0.001	0.005	0.02
ER	0.00072	0.0008	0.0085	0.00014	0.0008	0.006	0.0009	0.0008	0.0005
	0.064	0.071	0.219	0.031	0.071	0.186	0.072	0.071	0.057
BA	0.0113	0.009	0.0022	0.00007	0.00039	0.00497	0.0002	0.00005	0.0001
	0.275	0.249	0.128	0.022	0.051	0.169	0.033	0.017	0.027
	Random graphs, nodes =100,000, density = 0.0004								
β	0.001	0.005	0.02	0.001	0.005	0.02	0.001	0.005	0.02
ER	0.00039	0.0022	0.021	0.00015	0.0016	0.018	0.00024	0.00009	0.0006
	0.048	0.117	0.335	0.032	0.101	0.318	0.04	0.019	0.06
BA	0.0124	0.0067	0.0012	0.00007	0.00146	0.0178	0.00002	0.00003	0.0003
	0.287	0.218	0.082	0.022	0.094	0.304	0.012	0.021	0.049
	Real graphs								
β	0.001	0.005	0.02	0.001	0.005	0.02	0.001	0.005	0.02
DE	0.0051	0.00452	0.0035	0.00005	0.00023	0.00156	0.00002	0.00001	0.000007
	0.341	0.305	0.189	0.045	0.098	0.234	0.025	0.01	0.013
GS	0.00894	0.00858	0.00747	0.00002	0.0001	0.00056	0.00006	0.00005	0.00003
	0.352	0.338	0.291	0.025	0.052	0.123	0.017	0.016	0.012
EG	0.0059	0.0068	0.012	0.0001	0.0006	0.0032	0.00004	0.00002	0.0002
	0.386	0.443	0.667	0.098	0.272	0.467	0.045	0.025	0.111
LT	0.004	0.00352	0.0045	0.00005	0.00042	0.00334	0.00001	0.000004	0.000008
	0.331	0.251	0.283	0.06	0.165	0.366	0.016	0.018	0.025

Note that the best results of our proposed model "SI approximation" are shown on real graphs. The root mean square error for various graphs and information dissemination probabilities does not exceed 0.0002. The "SI on the network" model also showed acceptable results, while the classical SI model turned out to be unsuitable for predicting the distribution of information on such graphs.

The "SI Approximation" model showed the best accuracy on Barabási–Albert random graphs, except for a small $\beta = 0.001$. In this case, the "SI on the network" model showed the best accuracy.

For the Erdős–Rényi graph, the "SI approximation" approach also yielded good results in information dissemination prediction, except for a graph of dimension 10,000 (density = 0.001). The models "SI on network" and "SI without network" showed good results for small β (0.005 and 0.001).

Diffusion trajectories of information for various models on social networks and random graphs are shown in Figures 1–6.

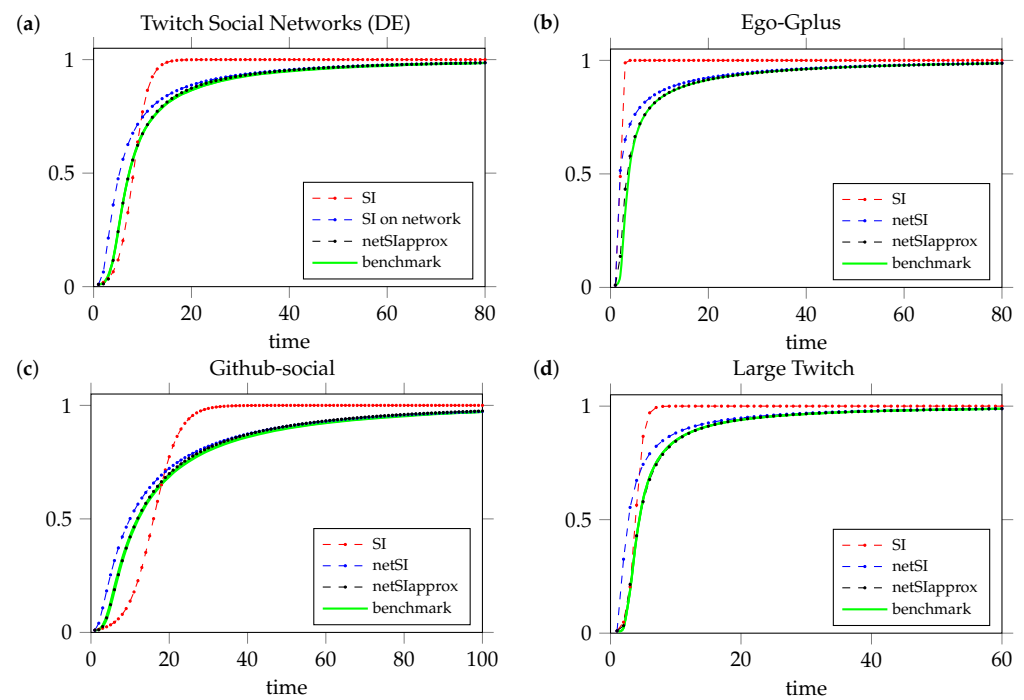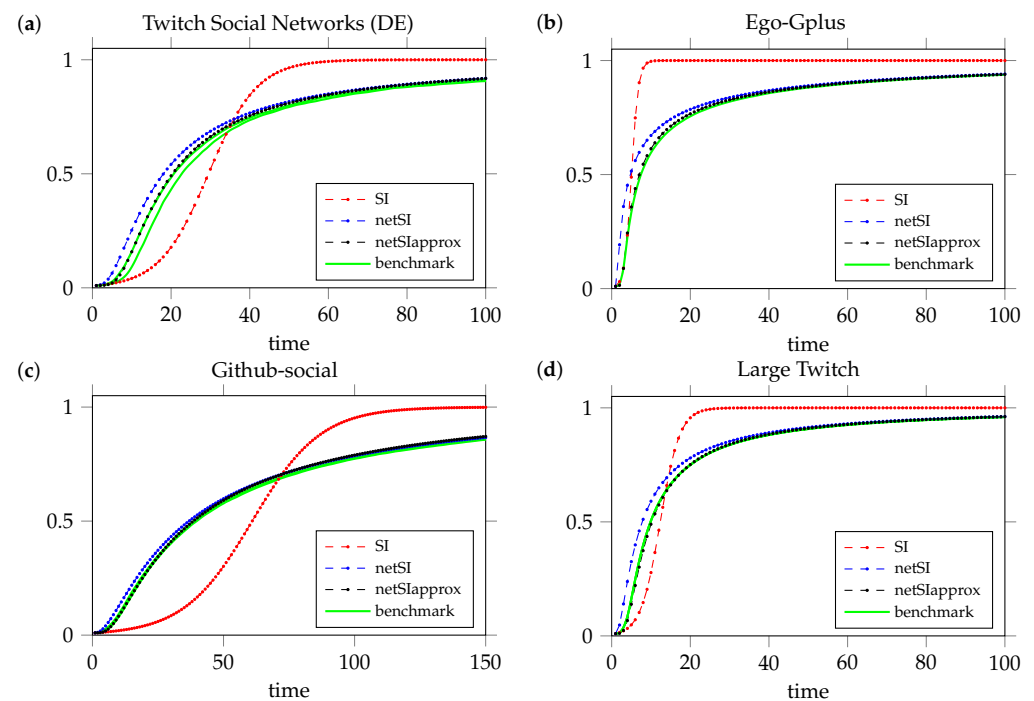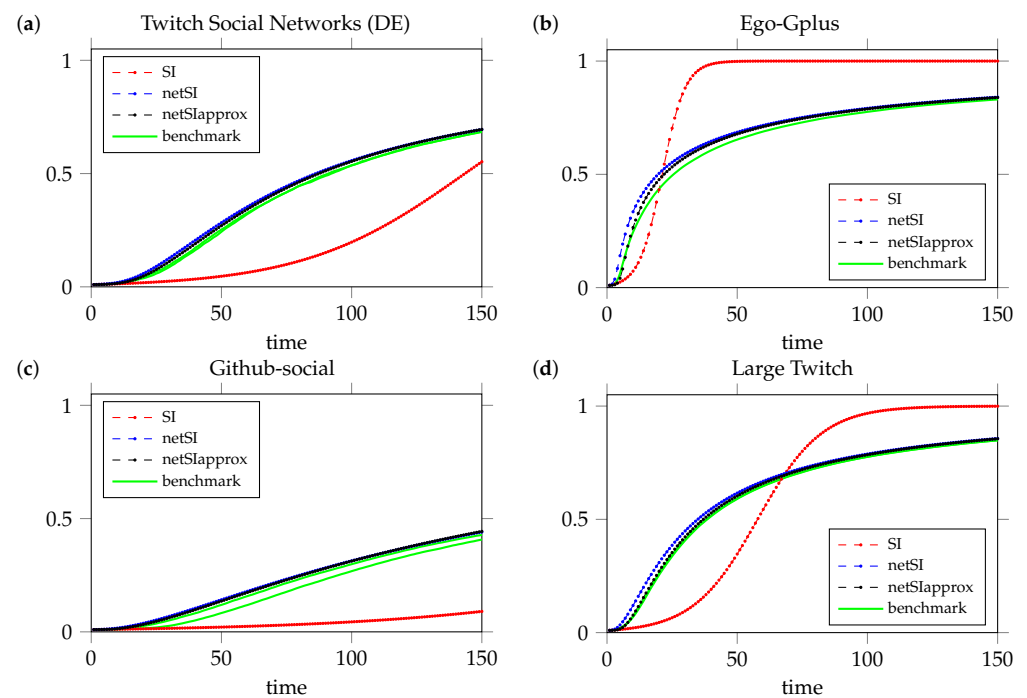**Figure 1.** Diffusion trajectories of information for different models on social networks: (**a**) Twitch Social Networks (DE), (**b**) Ego-Gplus, (**c**) Github-social, (**d**) Large twitch. Information diffusion probability $\beta = 0.02$.
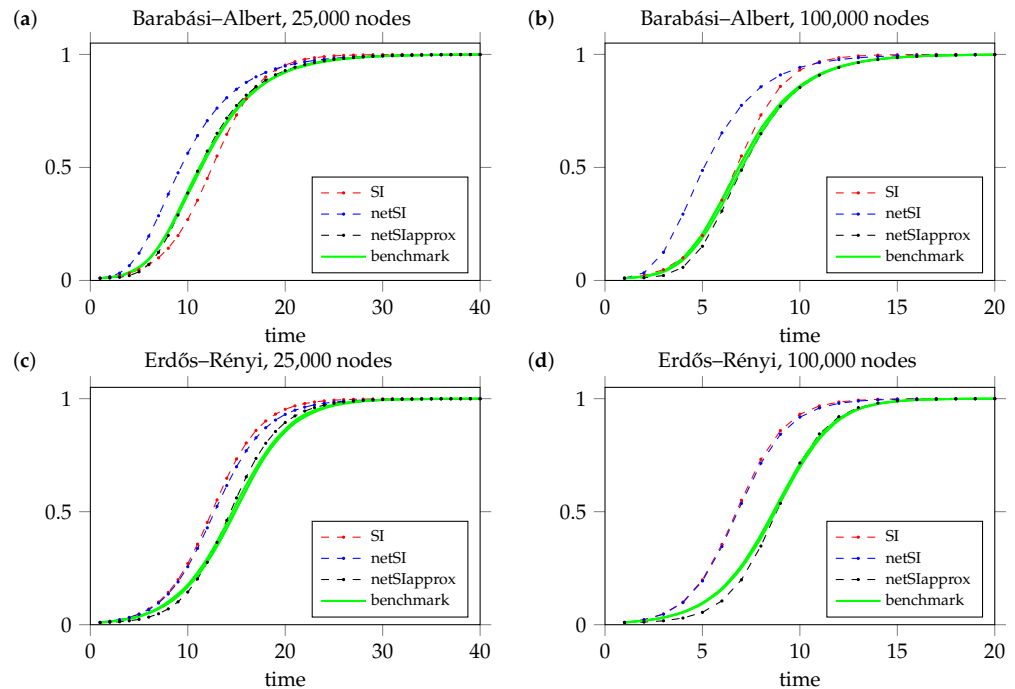

Figure 2. Diffusion trajectories of information for different models on social networks: (**a**) Twitch Social Networks (DE), (**b**) Ego-Gplus, (**c**) Github-social, (**d**) Large twitch. Information diffusion probability $\beta = 0.005$.

Figure 3. Diffusion trajectories of information for different models on social networks: (**a**) Twitch Social Networks (DE), (**b**) Ego-Gplus, (**c**) Github-social, (**d**) Large twitch. Information diffusion probability $\beta = 0.001$.
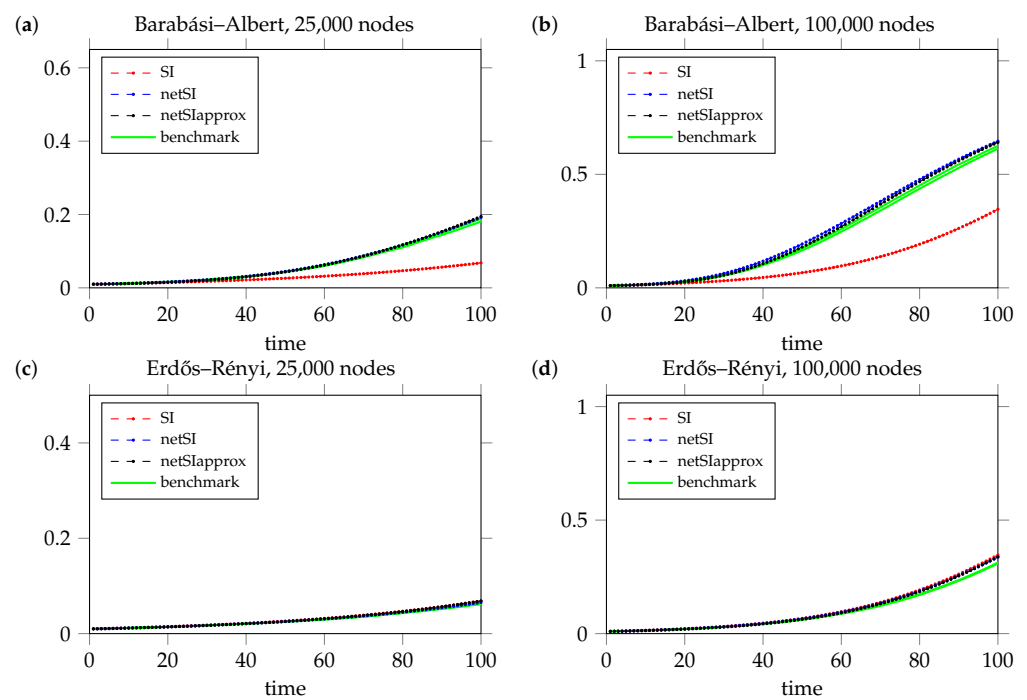

Figure 4. Diffusion trajectories of information for different models on random graphs: (**a**) Barabási–Albert, 25,000 nodes; (**b**) Barabási–Albert, 100,000 nodes; (**c**) Erdős–Rényi, 25,000 nodes; (**d**) Erdős–Rényi, 100,000 nodes. Information diffusion probability $\beta = 0.02$.

Figure 5. Diffusion trajectories of information for different models on random graphs: (**a**) Barabási–Albert, 25,000 nodes; (**b**) Barabási–Albert, 100,000 nodes; (**c**) Erdős–Rényi, 25,000 nodes; (**d**) Erdős–Rényi, 100,000 nodes. Information diffusion probability $\beta = 0.005$.

Figure 6. Diffusion trajectories of information for different models on random graphs: (**a**) Barabási–Albert, 25,000 nodes; (**b**) Barabási–Albert, 100,000 nodes; (**c**) Erdős–Rényi, 25,000 nodes; (**d**) Erdős–Rényi, 100,000 nodes. Information diffusion probability $\beta = 0.001$.

We tested and compared the dynamics of information dissemination according to the proposed model with other existing models both on generated random graphs and on real networks. Empirical results show that the proposed model gives adequate results regardless of the topology of random or real graphs. The model we propose is particularly good at describing the dissemination of information on graphs with a scale-free structure.

5. Conclusions and Discussion

Previous studies of diffusion on graph structures have typically focused on the spread of infections in the early stages of an epidemic. This is entirely justified because, if a cure has not yet been invented, the only way to change the course of the epidemic is to do so at an early stage, using various restrictive measures to slow its spread. In this regard, in order to make the right decision about the nature, timing, and scale of the epidemic, it is quite sufficient to estimate the number of people infected in early stages of the epidemic. From the point of view of mathematical modeling, this significantly simplifies the problem, since we assume that the proportion of infected people in the early stages is close to zero and it is possible to obtain an explicit solution.

In our study, we consider the dissemination of information, and therefore we are interested in all stages of this process. As a result, we use the block approximation approach and show that it successfully predicts the information propagation at all its stages. In addition, we proposed a new approach that assumes that the number of available information transmission channels increases over time as an increasing number of network users become infected. The rationale behind our assumption is that the more nodes are infected, the more opportunities (channels) they have to transmit information over the network. We described a function which approximates the number of channels available for transmitting information in the future, and we empirically estimated its parameters (depending on the time and probability of information dissemination). As a result, we received a system of simultaneous (coupled) equations, the solution of which was implemented in Python.

The proposed models were tested on various types of random graphs, as well as on real social networks and for different probabilities of information dissemination. It should be noted that

- All considered approaches showed good results on Erdős–Rényi graphs. Even the basic model (SI), which does not take into account the network structure, showed acceptable results. This is quite natural, since this type of random graph essentially embodies the homogeneous mixing hypothesis. However, real interactions between network users have a more complex structure.
- Secondly, the basic model (SI) turned out to be unsuitable for predicting the spread of information on Barabási–Albert graphs and real networks. This is quite expected since such graphs have a more complex structure and are scale-free, i.e., their degree distributions follow a power law. At the same time, the modified version of the block approximation model and especially the new approach showed good results in their ability to predict information diffusion on these types of graphs.
- In the modified version of the block approximation model, it was assumed that the number of channels available for transmitting information in the future will be one less than the degree of the vertex. In our opinion, this is a rather rough assumption for all stages of information dissemination, and this approach can be further improved.

The derivation of the SI graph model and our first analytical results open up many perspectives in modeling information propagation in complex networks. This paper represents the first stage of a research program on modeling in this vein. In particular, it is planned to implement the SIS and SIR models and their various modifications in network structures.

Abbreviations

The following abbreviations are used in this manuscript:

BA	Barabási–Albert
ER	Erdős–Rényi
SI	susceptible–infected
SIS	susceptible–infected–susceptible
SIR	susceptible–infected–recovered
netSI	SI on the network model (Section 3.1)
netSIapprox	SI on the network with approximation (Section 3.2)

References

1. Rogers, E. *Diffusion of Innovations*, 5th ed.; Free Press: New York, NY, USA, 2003.
2. Rouvinen, P. Diffusion of digital mobile telephony: Are developing countries different? *Telecommun. Policy* **2006**, *30*, 46–63. [CrossRef]
3. Vicente, M.R.; Lopez, A.J. Patterns of ICT diffusion across the European Union. *Econ. Lett.* **2006**, *93*, 45–51. [CrossRef]
4. Honoré, B. Diffusion of mobile telephony: Analysis of determinants in Cameroon. *Telecommun. Policy* **2019**, *43*, 287–298. [CrossRef]
5. Ahmad, M.; Almamri, A. Statistical models for mobile telephony growth in Oman. *Inf. Manag. Bus. Rev.* **2014**, *6*, 121–127. [CrossRef]
6. Baburin, V.; Zemtsov, S. *Diffussion of ICT-Products and "Five Russias"*; MPRA Paper 68926; University Library of Munich: Munich, Germany, 2014.
7. Guidolin, M.; Manfredi, P. Innovation diffusion processes: Concepts, models, and predictions. *Annu. Rev. Stat. Its Appl.* **2023**, *10*, 451–473. [CrossRef]
8. Bass, F.M. A new product growth for model consumer durables. *Manag. Sci.* **1969**, *15*, 215–227. [CrossRef]
9. Gompertz, B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. F. R. S. &c. By Benjamin Gompertz, Esq. F. R. S *Abstr. Pap. Print. Philos. Trans. R. Soc. Lond.* **1833**, *2*, 252–253. [CrossRef]
10. Bertotti, M.L.; Modanese, G. On the evaluation of the takeoff time and of the peak time for innovation diffusion on assortative networks. *Math. Comput. Model. Dyn. Syst.* **2019**, *25*, 482–498. [CrossRef]
11. Bahrami, S.; Atkin, B.; Landin, A. Innovation diffusion through standardization: A study of building ventilation products. *J. Eng. Technol. Manag.* **2019**, *54*, 56–66. [CrossRef]
12. Rakesh, K.; Anuj Kumar, S.; Kulbhushan, A. Dynamical analysis of an innovation diffusion model with evaluation period. *Bol. Soc. Parana. Mat.* **2020**, *38*, 87–104. [CrossRef]
13. Modanese, G. The network Bass model with behavioral compartments. *Stats* **2023**, *6*, 482–494. [CrossRef]
14. Kumar, R.N. Gillespie algorithm and diffusion approximation based on Monte Carlo simulation for innovation diffusion: A comparative study. *Monte Carlo Methods Appl.* **2019**, *25*, 209–215. [CrossRef]
15. Zhang, H.; Vorobeychik, Y. Empirically grounded agent-based models of innovation diffusion: A critical review. *Artif. Intell. Rev.* **2019**, *52*. [CrossRef]
16. Zheng, J.; Xu, M.; Cai, M.; Wang, Z.; Yang, M. Modeling group behavior to study innovation diffusion based on cognition and network: An analysis for garbage classification system in Shanghai, China. *Int. J. Environ. Res. Public Health* **2019**, *16*, 3349. [CrossRef] [PubMed]
17. Cramer, M.; Almeida, F.; Wendl, M.; Anderson, M.; Rautianinen, R. Innovation diffusion in an agricultural health center: Moving information to practice. *J. Agromed.* **2019**, *24*, 239–247. [CrossRef]
18. Yang, W.; Yu, X.; Zhang, B.; Huang, Z. Mapping the landscape of international technology diffusion (1994–2017): Network analysis of transnational patents. *J. Technol. Transf.* **2021**, *46*, 138–171. [CrossRef]
19. Akinyemi, O.; Harris, B.; Kawonga, M. Innovation diffusion: How homogenous networks influence the uptake of community-based injectable contraceptives. *BMC Public Health* **2019**, *19*, 1520. [CrossRef] [PubMed]

20. Boumaiza, A.; Abbar, S.; Mohandes, N.; Sanfilippo, A. Innovation diffusion for renewable energy technologies. In Proceedings of the 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018), Doha, Qatar, 10–12 April 2018; pp. 1–6. [CrossRef]

21. Doo, M.; Liu, L. Extracting top-k most influential nodes by activity analysis. In Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014), Redwood City, CA, USA, 13–15 August 2014; pp. 227–236. [CrossRef]

22. Hu, Y.; Song, J.; Chen, M. Modeling for information diffusion in online social networks via hydrodynamics. *IEEE Access* **2017**, *5*, 128–135. [CrossRef]

23. Bewley, R.; Fiebig, D.G. A flexible logistic growth model with applications in telecommunications. *Int. J. Forecast.* **1988**, *4*, 177–192. [CrossRef]

24. Griliches, Z. Hybrid corn: An exploration in the economics of technological change. *Econometrica* **1957**, *25*, 501–522. [CrossRef]

25. Frank, L.D. An analysis of the effect of the economic situation on modeling and forecasting the diffusion of wireless communications in Finland. *Technol. Forecast. Soc. Chang.* **2004**, *71*, 391–403. [CrossRef]

26. Gruber, H.; Verboven, F. The diffusion of mobile telecommunications services in the European Union. *Eur. Econ. Rev.* **2001**, *45*, 577–588. [CrossRef]

27. Lee, M.K.; Cho, Y. The diffusion of mobile telecommunications services in Korea. *Appl. Econ. Lett.* **2007**, *14*, 477–481. [CrossRef]

28. Liikanen, J.; Stoneman, P.; Toivanen, O. Intergenerational effects in the diffusion of new technology: The case of mobile phones. *Int. J. Ind. Organ.* **2004**, *22*, 1137–1154. [CrossRef]

29. Pastor-Satorras, R.; Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **2001**, *86*, 3200–3203. [CrossRef] [PubMed]

30. Zhang, X.; Zhang, Z.K.; Wang, W.; Hou, D.; Xu, J.; Ye, X.; Li, S. Multiplex network reconstruction for the coupled spatial diffusion of infodemic and pandemic of COVID-19. *Int. J. Digit. Earth* **2021**, *14*, 401–423. [CrossRef]

31. Berestycki, H.; Desjardins, B.; Weitz, J.; Oury, J.M. Epidemic modeling with heterogeneity and social diffusion. *J. Math. Biol.* **2023**, *86*. [CrossRef] [PubMed]

32. Eryarsoy, E.; Delen, D.; Davazdahemami, B.; Topuz, K. A novel diffusion-based model for estimating cases, and fatalities in epidemics: The case of COVID-19. *J. Bus. Res.* **2021**, *124*, 163–178. [CrossRef]

33. Dimarco, G.; Perthame, B.; Toscani, G.; Zanella, M. Kinetic models for epidemic dynamics with social heterogeneity. *J. Math. Biol.* **2021**, *83*, 1–32. [CrossRef]

34. Gómez, A.; Oliveira, G. New approaches to epidemic modeling on networks. *Sci. Rep.* **2023**, *13*, 468. [CrossRef] [PubMed]

35. Wang, J.; Wang, Y.Q. SIR rumor spreading model with network medium in complex social networks. *Chin. J. Phys.* **2015**, *53*, 1–21. [CrossRef]

36. Woo, J.; Chen, H. Epidemic model for information diffusion in web forums: Experiments in marketing exchange and political dialog. *SpringerPlus* **2016**, *5*, 66. [CrossRef] [PubMed]

37. Bao, H.; Chang, E.Y. AdHeat: An influence-based diffusion model for propagating hints to match ads. In Proceedings of the 19th International Conference on World Wide Web, New York, NY, USA, 26–30 April 2010; WWW '10; pp. 71–80. [CrossRef]