International Journal of Data Science and Analytics
https://doi.org/10.1007/s41060-022-00334-z

REGULAR PAPER

Exo-SIR: an epidemiological model to analyze the impact of exogenous spread of infection

Nirmal Kumar Sivaraman[1] · Manas Gaur[2] · Shivansh Baijal[1] · Sakthi Balan Muthiah[1] · Amit Sheth[2]

Received: 1 November 2021 / Accepted: 30 April 2022

Abstract
Epidemics like Covid-19 and Ebola have impacted people's lives significantly. The impact of mobility of people across the countries or states in the spread of epidemics has been significant. The spread of disease due to factors local to the population under consideration is termed the endogenous spread. The spread due to external factors like migration, mobility, etc., is called the exogenous spread. In this paper, we introduce the Exo-SIR model, an extension of the popular SIR model and a few variants of the model. The novelty in our model is that it captures both the exogenous and endogenous spread of the virus. First, we present an analytical study. Second, we simulate the Exo-SIR model with and without assuming contact network for the population. Third, we implement the Exo-SIR model on real datasets regarding Covid-19 and Ebola. We found that endogenous infection is influenced by exogenous infection. Furthermore, we found that the Exo-SIR model predicts the peak time better than the SIR model. Hence, the Exo-SIR model would be helpful for governments to plan policy interventions at the time of a pandemic.

Keywords Covid-19 · Ebola · Epidemic modeling · Compartment model · Exogenous infection · Endogenous infection · SIR · Exo-SIR

1 Introduction

An epidemic is a disease that spreads rapidly to a large number of people in a given population within a short period. Many epidemics occur in the world. Covid-19 and Ebola are recent prominent examples.

People have tried many methods to study epidemics. The susceptible, infected, and recovered model (SIR model) is considered as one of the seminal models of epidemics [1]. A recent work [2] gives a comprehensive review of the methods to model and analyze Covid-19. Out of these methods, the models relevant to our model are compartmental models. They are prominent methods used for the analysis and prediction of Covid-19 dynamics [2–4]. However, these works consider only infection from people to people from within the population and do not consider any external source of infection explicitly.

World Health Organization (WHO) has identified external transmission as one of the three modes of transmission [5]. According to WHO, the infection within the population is called as *Local transmission and community transmission*, and the infection external to the population is called as *Imported cases*. We call the infection from a source within the population and external to the population as the endogenous and exogenous spread of infection, respectively. Human migration is one of the prime reasons behind the exogenous spread of infection.

The governments can intervene to curb the spread of the disease by bringing in policies to stop human mobility [6]. However, the implementations of such intervention policies have a lot of challenges. Social disagreement is an example. Social disagreement means people do not abide by

✉ Nirmal Kumar Sivaraman
nirmal.sivaraman@lnmiit.ac.in

Manas Gaur
MGAUR@email.sc.edu

Shivansh Baijal
shivanshbaijal@gmail.com

Sakthi Balan Muthiah
sakthi.balan@lnmiit.ac.in

Amit Sheth
AMIT@sc.edu

[1] Department of Computer Science and Engineering, The LNM Institute of Information Technology, Jaipur, India

[2] AI Institute, University of South Carolina, Columbia, USA

Published online: 06 June 2022

the government's directives. The Tablighi Jamaat[1]. religious congregation that happened in India and the human mobility as a result of it is an example of social disagreement.

As a part of intervention, the Governments can restrict human mobility. However, they cannot completely prevent all such human mobility and migration. For example, in the Indian sub-continent, people migrate to metropolitan cities for work. Due to the risk of Covid-19 exposure in these overpopulated cities, people migrate back to their homes [7]. This is also known as reverse migration[2].

The government cannot deny one's right to go home. However, the government can allow necessary movement in a controlled manner. For example, when people move from one state to another, the state governments can issue passes for anyone who is allowed to travel to that state similar to what was practiced by the State of Kerala[3]. They can identify the incoming people and ensure that they correctly follow the procedures advised by the respective governments.

These movements will increase the exogenous spread of the infections compared to the ideal condition of sealed borders. To find the amount of infection during this movement and when the peak occurs, authorities need an explicit model that can predict infection through exogenous means [8]. Our model extends the SIR model and explicitly takes care of the amount of exogenous infection and endogenous infections. In the case of the spread of epidemics, even if there is a small increase in the number of infected people, the impact grows exponentially with time. Hence, it is important to consider the exogenous infections while studying the dynamics of epidemics [9]. This allows the governments to have pertinent information regarding the possible exogenous infections. This gives the government authorities time to prepare their medical resources accordingly.

There are challenges even if people do not migrate. For example, frontline workers like doctors and nurses are more frequently exposed to the virus than a common man. Correspondingly, we need to be able to model different rates of infections for different groups of people. The governments will have to make all the necessary safety equipment available to the frontline workers and monitor their health constantly to control the infection as a measure of intervention.

In this context, we address the following research questions that significantly modifies the current, well-studied SIR model by infusing external knowledge related to pandemic [10]:

1. How to quantify the exogenous spread of infection?
2. What is the interplay between the exogenous and endogenous spread of infection concerning the following:
   (a) In the presence of social disagreement.
   (b) In the presence of controlled migration.
   (c) In the presence of *n communities* that have a different rate of infection—e.g., frontline workers such as healthcare workers or hospitality workers.
3. What is the change in the peak position (the most significant number of people infected in a unit of time) in the presence of exogenous infection?
4. What is the change in the height of the peak in the presence of exogenous infection?

The following are our contributions in this work. We study the impact of external reasons of infections such as cross-border mobility on COVID infection by introducing a novel SIR-like compartmental model called Exo-SIR.

We study three variants of the model applicable for special scenarios like the presence of social disagreement, the presence of different groups that have a different amount of risk, and infectiousness like the frontline workers.

We analyze the interplay between endogenous and exogenous infections during the Covid-19 and Ebola pandemics in the following ways.

1. Analytically.
2. By simulating the Exo-SIR model with and without assuming contact network for the population.
3. By implementing the Exo-SIR model on real datasets regarding Covid-19 and Ebola.

We compare the predictions of Exo-SIR with the SIR model using real data on the recent spread of the Covid-19 in India and the USA and the spread of Ebola in Africa as the ground truth.

This paper is structured as follows. Section 2 discusses related works and preliminaries. Then, we formulate the Exo-SIR model by extending the SIR model and discuss the different variants of the model (in Sect. 3). We analyze our model by comparing it with the SIR model and study the behavior of the infected population in the presence and absence of exogenous infection (in Sect. 4). Then, we describe the simulation study where we simulated the SIR model and Exo-SIR model and compared them (in Sect. 5). Finally, we study the real data of Covid-19 and Ebola epidemics (in Sect. 6).

[1] https://en.wikipedia.org/wiki/Tablighi_Jamaat.

[2] https://www.epw.in/journal/2020/19/commentary/migration-and-reverse-migration-age-covid-19.html.

[3] https://www.news18.com/news/auto/covid-19-omicron-kerala-travel-guidelines-for-international-and-domestic-passengers-4525862.html.

2 Related works

Here, we discuss the works related to the idea of exogenous influence to the population under study.

The work in [11] considers exogenous infections for Malaria at China—Myanmar border. However, the model is not deterministic. In a deterministic model, individuals in the population are assigned to different subgroups or compartments, each representing a specific epidemic stage. Deterministic models often provide useful ways of gaining sufficient understanding about the dynamics of populations whenever they are large enough [12]. Also, the deterministic models are simpler and more popular [13,14]. A mobility-based SIR model [15] is a deterministic model. Our model is also deterministic.

2.1 Models of external influence on online social networks

Information diffusion in online social networks is similar to the way the virus spreads in a population [16]. There are a few recent works in the literature that attempt to model the external influence in information diffusion in online social networks [17]. Moreover, [17] and [18] propose information diffusion model on the network. These works assume that the information flows through an underlining network. Also, they consider links from other websites like the mainstream media as external sources of information. Internal diffusion is when the shared messages do not have any external links.

The work described in [17] uses very specific parameters like the following:

- Probability of any node receiving exposure at time t
- The random amount of time it takes an infected node to expose its neighbors
- How the probability of infection changes with each exposure
- The probability that a node I have received n exposures by time t

The work described in [18] traces the information cascade and thereby tries to reconstruct the underlying graph structure as much as possible. Also, they conclude that external influence has a bigger impact on the network when compared to the influence of social media influencers.

The model that is closest to our work is Yang et. al.'s model [19]. This model is an extends the SIR model (explained in Sect. 2.4) by including the external influence on the network. State transition diagram of the diffusion mechanisms of this model is given in Fig. 1.

This model is defined in the following way.

$$s + i + r = 1 \tag{1}$$

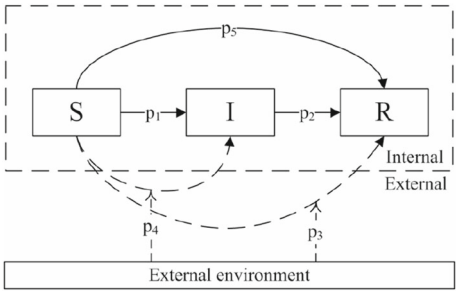

Fig. 1 State transition diagram of the diffusion mechanisms in Yang et al's model. Diagram taken from [19]

$$\frac{ds}{dt} = -p_1 ksi - ((1-p_1)p_3 + p_4)\theta s \tag{2}$$

$$\frac{ds}{dt} = -p_1 ksi - ((1-p_1)p_3 + p_4)\theta s - (1-p_1)p_5 ksi \tag{3}$$

$$\frac{di}{dt} = p_1 ksi + p_4 \theta s - p_2 i \tag{4}$$

$$\frac{dr}{dt} = p_2 i + (1-p_4)p_3 \theta s + (1-p_1)p_5 ksi \tag{5}$$

In Fig. 1, there are two possible transitions from the state S to I. One path is the normal endogenous path, and the second is due to external influence. These transitions have probabilities p_1 and $p4$, respectively. Similarly, there are two possible transitions from the state S to R—one through endogenous and the other through external influence. Their probabilities are p_5 and p_3, respectively. However, the transition from the state I to R is not affected by external influence(s).

Although the exogenous infection is modeled in Yang et al.'s model, it fails to capture the dynamics between endogenous and exogenous infections. This is because they do not differentiate between the infections due to exogenous factors from those due to endogenous factors.

2.2 Other studies of endogenous and exogenous information diffusion

The dual nature of message flow over the online social network is studied and verified in [20]. Here, the dual nature refers to the injection of exogenous opinions to the network and the endogenous influence-based dynamics. In [21], the authors propose a method for extracting the relative contributions of exogenous and endogenous contents. In [22], the authors postulate that the nature of the information plays a crucial role in the way it spreads through the network. They quantify two properties of the information—endogeneity and exogeneity. Endogeneity refers to its tendency to spread primarily through the connections between nodes, and exogeneity refers to its tendency to spread to the nodes, independently of the underlying network. In [23], the authors study the bursts that originate from endogenous and exogenous sources

and their temporal relationship with baseline fluctuations in the volume of tweets. The study reported in [24] classifies the bursts into endogenous and exogenous. According to this study, those bursts that reach the peak almost instantaneously after the diffusion starts and then go down slowly are exogenous bursts. Also, those bursts that gradually increase and slowly decrease are endogenous.

2.3 Compartmental models for Covid-19 modeling

Compartmental models are prominent methods that are used for the analysis and prediction of Covid-19 dynamics. The SIR model is one of the seminal compartmental models. Many compartmental models have come up recently to improve the SIR model. QSIR model [25,26] is an example in which they add an extra state to the standard SIR model that represents the number of people in Quarantine. SPCIRD model [27] adds three extra states—P, C, and D, where P represents the number of susceptible people who are partially controlled. Partially controlled people are those who can be considered as people not conforming to all the restrictions of the Quarantine. C represents the number of susceptible people who are controlled. Controlled people are those who can be considered as people conforming to all the restrictions of the Quarantine. D represents the number of people who died. Multiple epidemic wave model [28] as its name suggests models the multiple waves of infection that could occur. Time-dependent SIR model [29,30] considers the constants in the SIR model—beta and gamma to be varying with time. However, none of these models consider infections arising from outside the population, mostly due to the cross-border mobility of infected people. Hence, we introduce the Exo-SIR model to address this particular issue.

2.4 SIR model

This section briefly reviews the SIR epidemiological model to learn how epidemics spread through population. SIR is often used to study information diffusion by approximating the process of epidemic spread.

In this model, the population is classified into three—susceptible (who are prone to infection), infected (who contain the infection), and recovered (who do not have the infection and its associated symptoms). In the limit of sizeable total population N that does not change over time, the given equations model the dynamics of the spread [31]:

$$s(t) + i(t) + r(t) = 1 \tag{6}$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta si \tag{7}$$

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta si - \gamma i \tag{8}$$

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \gamma i \tag{9}$$

where the fraction of susceptible, infected, and recovered people at time t are represented by $s(t)$, $i(t)$, and $r(t)$, respectively. β is the rate of infection, and γ is the rate of recovery.

3 The model

In this section, we propose the Exo-SIR model. It differs from SIR model in the following ways. It classifies infected nodes into two different types—Infected from exogenous source and Infected from endogenous source. It also differentiates between the spread from endogenous and exogenous sources.

Susceptible nodes become infected with a certain probability called the rate of infection. This rate could be different for endogenous and exogenous infections. The nodes affected by endogenous and exogenous sources move into different states. We assume that susceptible nodes get infected from only one of these sources and never from both sources. Hence, even when some nodes are susceptible to endogenous and exogenous infection, they become infected by either an endogenous or an exogenous source but not both. The infected nodes recover with a certain probability called the recovery rate. These nodes move into the recovered state. The advantage of the Exo-SIR model compared to the SIR model is that we can observe the endogenous and exogenous diffusion separately.

We use the following notations:

S state of susceptible
I_x state of infected from exogenous source
I_e state of infected from endogenous source
R state of recovered
i_x Fraction of nodes that are infected from exogenous source
i_e Fraction of nodes that are infected from endogenous source
r Fraction of nodes that are recovered
β_x Rate at which the exogenous source infects the nodes
β_e Rate at which the nodes infect other nodes
γ Rate at which the nodes get recovered
We use the words infection, diffusion, and spread interchangeably according to the context.

The state transition diagram of the Exo-SIR model is given in Fig. 2.

We classify infected nodes into two different types—infected from exogenous source i_x and infected from endogenous source i_e.

$$i_e + i_x = i \tag{10}$$

We assume that the total population remains constant.

$$s + i + r = 1 \tag{11}$$

A fraction of the susceptible people s gets infected by exogenous sources, and another fraction of s gets infected by endogenous sources. For endogenous infection, the population that is infected plays a big role. Hence, we have

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta_x s - \beta_e s i \tag{12}$$

Increase in i_x is determined by the number of susceptible nodes and the decrease in i_x is determined by i_x. This gives

$$\frac{\mathrm{d}i_x}{\mathrm{d}t} = \beta_x s - \gamma i_x \tag{13}$$

Increase in i_e is determined by the number of susceptible nodes and the number of infected nodes and the decrease in i_e is determined by i_e. This gives

$$\frac{\mathrm{d}i_e}{\mathrm{d}t} = \beta_e s i - \gamma i_e \tag{14}$$

Increase in r is determined by the number of infected people in the network. This gives

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \gamma i \tag{15}$$

3.1 Variants of the model to address specific situations

In this section, we discuss how the Exo-SIR may be used in the different situations.

3.1.1 Exo-SIR model with social disagreement

This scenario occurs when people do not abide by the government's orders, for example, not wearing masks, not following social distancing, etc. As a result, more people contract the virus, and hence, the infectiousness of the disease will go up. This can be represented in the Exo-SIR model by increasing the β_e value.

3.1.2 Exo-SIR model with people migrating with the permission of the government

This scenario can be studied using the Exo-SIR model. Here, we assume that when the government allows people to travel, the government makes sure that these people are isolated and given treatment. Change in i_x is influenced by the action of the government that allowed people to travel across their border. Hence, planning and execution efficiency to minimize the impact are essential. This is captured in β_x. If the government efficiently contains the infection from these people, then the value of β_x goes down.

3.1.3 Exo-SIR model with multiple groups that have different risk of infection

This case may be depicted as shown in Fig. 3. In this case, there are n different groups of susceptible people with varying levels of infection risk. Hence, we add them up wherever we use s in the equations of the Exo-SIR model. Also, the value of each parameter is different for a different group of people. Hence, we have different values for each group of people for the parameters. Hence, there will be the summation of the n groups and parameters for each group. Figure 3 is the state diagram and the equations are given below.

$$i_e + i_x = i \tag{16}$$

$$s = \sum_{k=1}^{n} s_k \tag{17}$$

$$s + i + r = 1 \tag{18}$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\sum_{k=1}^{n} \beta_{xk} s_k - \sum_{k=1}^{n} \beta_{ek} s_k i \tag{19}$$

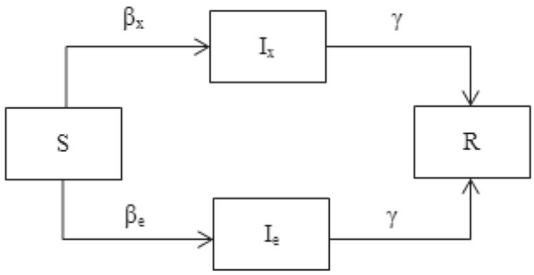

Fig. 2 State transition diagram of the nodes in the Exo-SIR model

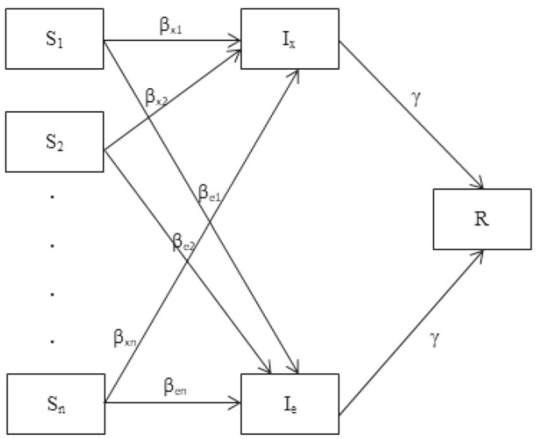

Fig. 3 State transition diagram of the model

$$\frac{\mathrm{d}i_x}{\mathrm{d}t} = \sum_{k=1}^{n} \beta_{xk} s_k - \gamma i_x \tag{20}$$

$$\frac{\mathrm{d}i_e}{\mathrm{d}t} = \sum_{k=1}^{n} \beta_{ek} s_k i - \gamma i_e \tag{21}$$

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \gamma i \tag{22}$$

4 Analysis

In this section, we compare our model with SIR model and analyze the dynamics of exogenous spread and endogenous spread.

4.1 Comparison with SIR model

Mirroring the rate of change of $s(t)$, $i(t)$, and $r(t)$ in the SIR model (Sect. 2), we find the expressions for the rate of change of $s(t)$, $i(t)$, and $r(t)$ for the Exo-SIR model.

Rate of change of s is given by

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta_x s - \beta_e s i \tag{23}$$

Rate of change of r is given by

$$\frac{dr}{dt} = \gamma i \tag{24}$$

Differentiating Eq. 10 with respect to time, we get

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{\mathrm{d}i_e}{\mathrm{d}t} + \frac{\mathrm{d}i_x}{\mathrm{d}t} \tag{25}$$

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta_e s i - \gamma i_e + \beta_x s - \gamma i_x \tag{26}$$

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta_e s i + \beta_x s - \gamma (i_x + i_e) \tag{27}$$

Applying Eqs. 10 on 27, we get

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta_e s (i_x + i_e) + \beta_x s - \gamma (i_x + i_e) \tag{28}$$

Here, even if we assume that there are no infected people in the beginning—i.e., $i_e = 0$ and $i_x = 0$, we get the following.

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta_x s \tag{29}$$

This shows that, unlike the SIR model, the Exo-SIR model explains how an infection starts spreading from the state where no one is infected. SIR model assumes that there is an initial outbreak size i_0. This means i_0 people are infected in the beginning and $i_0 > 0$ [32]. Our work addresses this limitation of the SIR model. Note that the Exo-SIR model would behave the same way as the SIR model if we assume that $i_x = 0 \; and \; \beta_x = 0$.

4.2 Dynamics of exogenous spread and endogenous spread

In this section, we find the relationship between the cumulative exogenous infections (i_x) and the daily endogenous infections ($\frac{\mathrm{d}i_e}{\mathrm{d}t}$).

Applying Eqs. 10 on 14, we get

$$\frac{\mathrm{d}i_e}{\mathrm{d}t} = \beta_e s (i_e + i_x) - \gamma i_e \tag{30}$$

$$\left.\frac{\mathrm{d}i_e}{\mathrm{d}t}\right|_{i_x > 0} = \beta_e s (i_e + i_x) - \gamma i_e \tag{31}$$

At $i_x = 0$,

$$\left.\frac{\mathrm{d}i_e}{\mathrm{d}t}\right|_{i_x = 0} = \beta_e s i_e - \gamma i_e \tag{32}$$

Since all $\beta_e, s, i_e, and \; \gamma$ are positive,

$$\left.\frac{\mathrm{d}i_e}{\mathrm{d}t}\right|_{i_x = 0} < \left.\frac{\mathrm{d}i_e}{\mathrm{d}t}\right|_{i_x > 0} \tag{33}$$

This shows that $\frac{\mathrm{d}i_e}{\mathrm{d}t}$ increases in the presence of i_x. In other words, this shows that the presence of exogenous diffusion causes endogenous diffusion to increase.

5 Simulation

We simulate the Exo-SIR model to determine its behavior for various scenarios that are represented by the different values of its parameters. We simulated the model in two ways:

One, by assuming no network (well-mixed population). In this scenario, a susceptible node can get infected from any of the infected nodes in the population under consideration.

Two, by assuming that the people network is a scale-free network. Within this network, the susceptible nodes can catch the infection from only those infected nodes, which they are connected to through an edge, i.e., their immediate neighbors. We chose scale-free network because there are pieces of evidence that the human disease network could be scale-free [33]. The results of these simulations are discussed in the following section.

5.1 Using scale-free network

The analysis presented in this section has been done considering a scale-free contact network for the population under

study, which is called Barabási-Albert network [34]. Under this scenario, the susceptible nodes can catch the infection from only those infected nodes, which they are connected to through an edge, i.e., their immediate neighbors. We have predicted the values for various combinations of β_x, β_e, and γ using the Exo-SIR model in the network mentioned above.

Next, we study the dependency of endogenous spread on the exogenous factors through simulation. The step-by-step methodology adopted to carry out the simulation and the analysis is given in Algorithm 1.

Algorithm 1 Algorithm to perform the simulations and analysis by assuming that the contact network in the population is scale-free

1: Initialize β_x, β_e, and γ with 3 different values, i.e., 0.1, 0.5, and 0.9. Henceforth, we have 27 different combination of these parameters.
2: For each of the combinations of β_x, β_e, and γ, iterate over steps 3 and 4 fifty times.
3: Setup a Barabási-Albert network of 1000000 nodes having an average node degree of 2 [34].
4: Simulate and predict the values of S, I_e, I_x, and R using the Exo-SIR model.
5: Extract the values of the height of the peak (we call it as peak value) and the time slice at which it occurs (we call it as peak tick), for both endogenous and exogenous peaks from each of the simulations.
6: Calculate the mean peak value and peak tick of exogenous and endogenous infections so that we have one value per combination of β_x, β_e, and γ.

In the above algorithm, we have carried out 50 simulations for each combination of the parameters and averaged it out to address the bias that might get introduced due to the network structure since the setting up of a network in step 3 in the above algorithm is random each time.

Sample simulation results are shown in Figs. 4 and 5. Figure 4 shows the SIR model's simulation results with no exogenous influence, and Fig. 5 shows the simulation results with exogenous influence. Here, we can see that when we consider exogenous factors, the peak of the distribution of the number of the infected population shows changes.

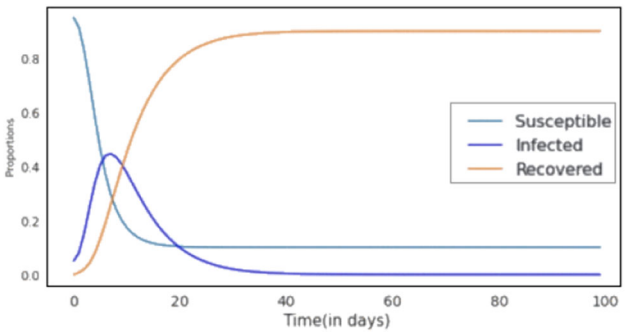

Fig. 4 plot of susceptible, infected, and recovered with no exogenous source

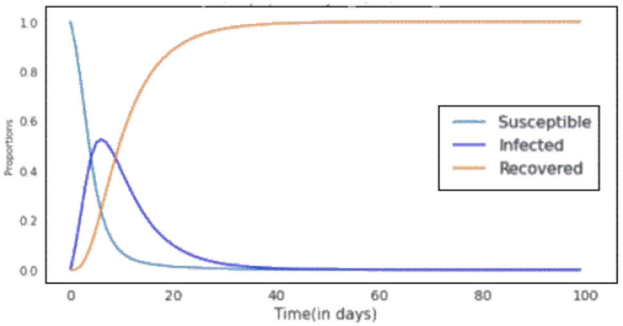

Fig. 5 plot of susceptible, infected, and recovered with exogenous source

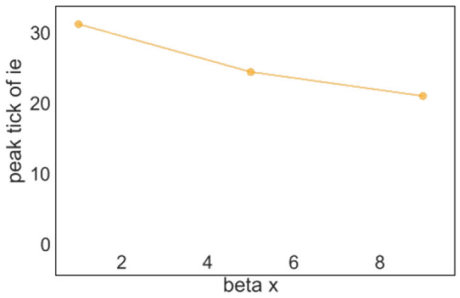

Fig. 6 impact of β_x on peak tick of i_e

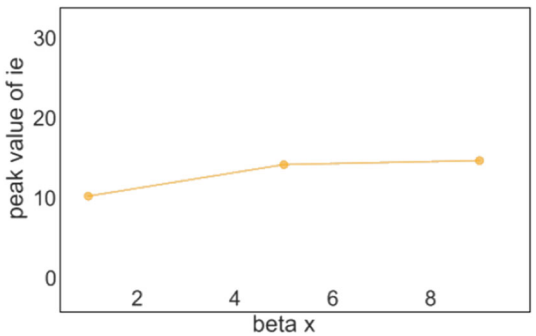

Fig. 7 impact of β_x on peak value of i_e

Figures 6 and 7 are a result of simulation and analysis done as described in Algorithm 1 and provide us with the following insights. Figure 6 shows that endogenous peak tick decreases with increase in β_x. Figure 7 shows that β_x (exogenous factors) influence the peak value of endogenous infections. The endogenous peak value increases with increase in β_x.

We can conclude that exogenous source and its infection impacts the endogenous spread in the network by advancing the peak and increasing the height of the peak.

5.2 With no network

In this section, we determine the relative effects of β_x, β_e, and γ on the endogenous peak statistically and measure the impact of β_x on endogenous infections, which is consis-

Table 1 Impact of β_e, β_x, and γ on $ln(\texttt{ie\_peak})$

	Coef	Std err	Confidence interval
β_e	0.6319	0.006	0.6204 to 0.6434
β_x	0.6319	0.006	0.6204 to 0.6434
γ	−0.4390	0.006	−0.4505 to −0.4274

tent with the results shown above. Here, we did not assume any network for our population, and the objective of these simulations was to determine the impact of β_x, β_e, and γ on endogenous peak value and peak tick (see Table 1). To achieve this, we took a sample of 27000 simulations and analyzed them as described in Algorithm 2.

Algorithm 2 Algorithm to perform the simulations and analysis by assuming no contact network in the population. **Note:** If we look at the differential equations, the system is not a linear one, but rather exponential. Therefore, we took natural log of the peak value as the dependent variable. Table 1 shows the impact of the above three independent variables on the dependent variable.

1: Initialize β_x, β_e and γ with 30 random values between 0 and 1.
2: Initialize the number of susceptible, infected (endogenous and exogenous), and recovered nodes experimentally as: N = 1000000.0, S_0 = 999996.0, I_{x_0} = 3.0, I_{e_0} = 1.0, and R_0 = 0.0.
3: For each of the 27000 combinations of β_x, β_e, and γ, with the above initial condition, predict the endogenous and exogenous peak value and peak tick using the Exo-SIR model.
4: Then, compute the natural logarithm of the peak value and scaled it between 0 and 1.
5: Finally, fit an OLS regression model with β_x, β_e and γ as the independent variables and the natural log of the peak value as the dependent variable and analyze the coefficients statistically.

The following inferences can be drawn from the results of regression analysis. The p value for all the three variables is less than 0.05. This means we would reject the null hypothesis and adopt the alternate hypothesis that the impact of all the three parameters on the peak endogenous infection's peak is statistically significant.

The adjusted R-squared value is maximum(0.70) when all the three parameters are considered while fitting the regression model. This means that we can better explain the variation in the dependent variable when considering all three, i.e., β_e, β_x, and γ. Removing any one of them would decrease the adjusted R-squared value. Also, the confidence interval of each parameter is mentioned in Table 1.

β_x impacts endogenous infections as much as β_e(the contribution of both is almost equal), which is an important observation. This means that exogenous factors also have a considerable impact on the endogenous infection, and ignoring the exogenous factors would not give an accurate estimate of the endogenous infections.

6 Analysis using real data

In this section, we describe the data and the analysis of the implementation of the SIR model and Exo-SIR model on the Covid-19 and Ebola epidemics.

6.1 Covid-19 infection in India

Covid-19 has caused large and persistent negative effects on the world economy[4]. India is one of the countries that are worst affected. There were many issues that made the spread of Covid-19 in India complicated. One of them was the migration of people from different parts of the country and abroad.

Many sub-events in India involved the migration of people. Examples are a celebrity coming to India from the UK and socializing at many places even after being tested positive for Covid-19[5]., laborers working in different states or other countries moving back to their native places [7], and large religious meetings with participation from many national and international locations.

A major sub-event was the Tablighi Jamaat religious congregation in Delhi from 1st March 2020 to 21st March 2020[6]. Over 9000 people from various states of India participated in this event[7]. Nearly 4300 cases have been reported that can be traced to the event[8]. As of 18th April 2020, 30% of the cases in India were due to this event[9]. The number of people from each state is widely deferred. Hence, the impact of the event was significantly different for different states. However, it is reasonable to state that the mobility of people is a causative phenomenon that changed the dynamics of the spread of the virus.

We apply the Exo-SIR model on a real dataset regarding the spread of the Covid-19 pandemic in the Indian states of Rajasthan, Tamil Nadu, and Kerala from 14th March, 2020 to 14th April, 2020. Exogenous spread dominates endogenous spread in Tamil Nadu, whereas the contrary is true in the case of Rajasthan. Both the endogenous and exogenous spread in Kerala have roughly the exact prevalence. The trends in the analytical study, results of the simulations, and the analysis of the real dataset are consistent.

We analyzed the data of three states in India, namely Tamil Nadu, Rajasthan, and Kerala. The reason for choosing these states is that $i_e \ll i_x$ in Tamil Nadu, $i_e \gg i_x$ in Rajasthan, and $i_e \approx i_x$ in Kerala.

[4] shorturl.at/qBZ05.
[5] shorturl.at/imBGK.
[6] https://en.wikipedia.org/wiki/2020_Tablighi_Jamaat_coronavirus_hotspot_in_Delhi.
[7] shorturl.at/qryKU.
[8] shorturl.at/myFQ2.
[9] shorturl.at/iyVY9.

We constructed our dataset from three different sources[10] for our analysis—covid19india.org.[11], the government website of the respective states for their press release to find the daily number of Tablighi cases and Wikipedia page on statewise daily data[12].

covid19india.org. is a publicly available volunteer-driven dataset of Covid-19 statistics in India[13]. There are multiple files in this dataset. One of which is called *raw data* that captures the anonymized details of the patients. In the raw data, the columns of interest for our study are DateAnnounced, DetectedState, and TypeOfTransmission.

Another file from covid19india.org is called *states_daily*. In this file the columns of interest are *states_daily/status*, *states_daily/kl*, *states_daily/rj*, *states_daily/tn*, and *states_daily/date*. *kl, rj* and *tn* are the codes used in this dataset for the states of Kerala, Rajasthan, and Tamil Nadu, respectively.

Here, status can have the following values: infected, recovered, and diseased. From these columns, we prepared the time series dataset for each state. The columns available in the dataset we created are daily confirmed, daily deceased, daily recovered, date, total confirmed, total deceased, totally recovered, and daily imported cases.

Another dataset that we used is the compilation of the press releases (news bulletins) from the states' governments under study. This is to get the daily number of cases due to a significant event that influenced the Covid-19 spread in India—Tablighi Jamaat religious congregation. Since there was no ready-made data available, we manually went through the press releases and collected the data.

Now, we discuss how the values in the dataset is mapped on to the variables in the Exo-SIR model. On a particular day, say day k, by rearranging and differentiating Eq. 11, we get the following.

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\left(\frac{\mathrm{d}i}{\mathrm{d}t} + \frac{\mathrm{d}r}{\mathrm{d}t}\right) \tag{34}$$

where $\frac{\mathrm{d}r}{\mathrm{d}t}$ is the sum of the numbers of the daily recovered and the daily deceased cases on day k and

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{\mathrm{d}i_e}{\mathrm{d}t} + \frac{\mathrm{d}i_x}{\mathrm{d}t} \tag{35}$$

where $\frac{\mathrm{d}i_e}{\mathrm{d}t}$ is the daily confirmed cases on day k and $\frac{\mathrm{d}i_x}{\mathrm{d}t}$ is the sum of daily imported cases on day k and the daily cases due to Tablighi event on day k.

[10] The code and all the data used in our experiments will be made openly available upon the acceptance of this paper

[11] www.covid19india.org.

[12] https://en.wikipedia.org/wiki/Statistics_of_the_COVID-19_pandemic_in_India.

[13] www.covid19india.org.

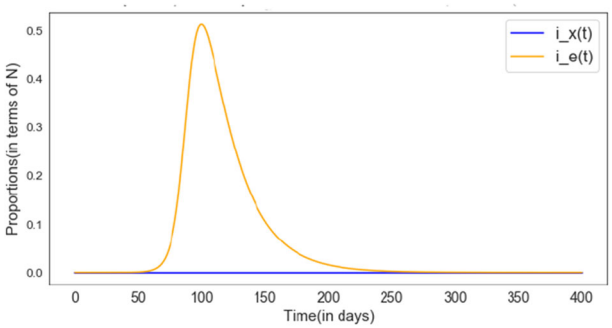

Fig. 8 I_e in the presence of i_x. The values of i_x are very small for the scale of this plot. Hence, it is plotted separately. Please refer Fig. 10

The initial values of s, i and r are found as follows.

$$s = 1 - \frac{\mathrm{d}(0)}{N} \tag{36}$$

where $d(0)$ is the daily confirmed on day 0 and N is the total population who are prone to the infection.

i is the total number of confirmed cases on day 0 and r is the sum of the total numbers of the deceased and the recovered cases on day 0.

Algorithm 3 Algorithm to plot the Exo-SIR model.

1: For each time slice, calculate the values of $\frac{di}{dt}$ and $\frac{dr}{dt}$ from the dataset.
2: Consider s as the susceptible people from the population of the state under study.
3: Calculate the cumulative values i and r.
4: Find γ, β_e and β_x using values of the time for which the data is available.
5: Run the Exo-SIR model with these values as the initial values and plot i_e in the presence of i_x and i_e in the absence of i_x

Next, we analyze the data from Tamil Nadu, Rajasthan, and Kerala. We compare the peak tick and peak value of the plot of i_e in the presence and absence of i_x. This would give information about the impact of i_x on i_e. For this purpose, we used Algorithm 3.

For the state of Tamil Nadu, the plots of I_e in the presence and absence of i_x are plotted in Figs. 8 and 9, respectively. I_x is plotted in Fig. 10. For the state of Rajasthan, the plots of I_e in the presence and absence of i_x are plotted in Figs. 11 and 12, respectively. For the state of Kerala, the plots of I_e in the presence and absence of i_x are plotted in Figs. 14 and 15, respectively. I_x is plotted in Fig. 13. In all these plots, we can see that i_x is very small compared to i_e. Yet, i_x is having an impact on i_e. I_x is plotted separately in Figs. 10, 13 and 16.

The peak tick and peak values corresponding to the I_e of the Exo-SIR model in the presence and absence of i_x for Tamil Nadu, Rajasthan, and Kerala are mentioned in Tables 2, 3, and 4, respectively. In all the tables, we can see that the peak value of i_e is different when the case of i_x is

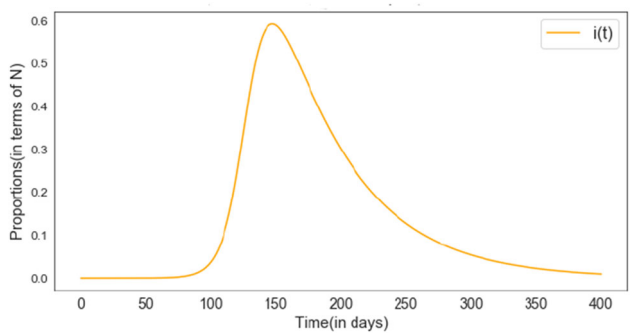

Fig. 9 I_e in the absence of i_x

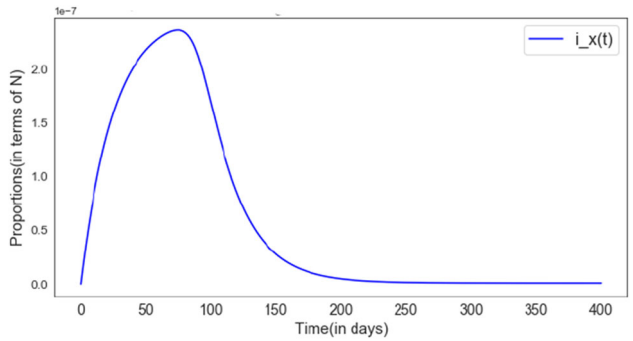

Fig. 10 i_x in Exo-SIR model. Please note that y axis is in the scale of 10^{-4}

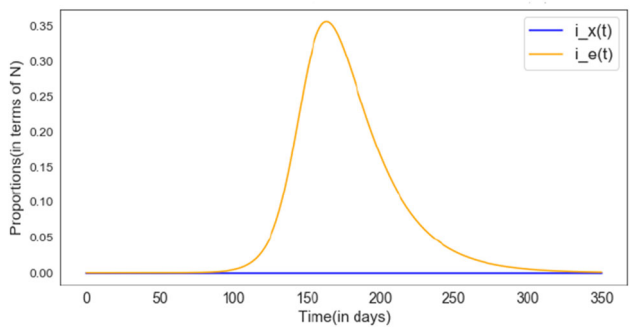

Fig. 11 I_e in the presence of i_x. The values of i_x are very small for the scale of this plot. Hence, it is plotted separately. Please refer Fig. 13

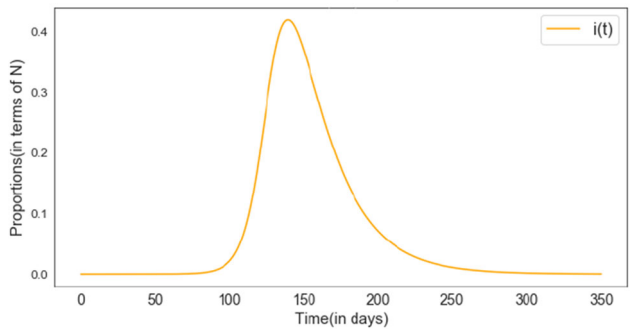

Fig. 12 I_e in the absence of i_x

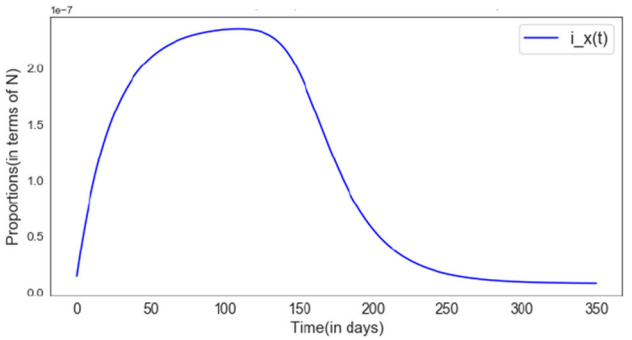

Fig. 13 i_x in Exo-SIR model. Please note that the y axis is in the scale of 10^{-6}

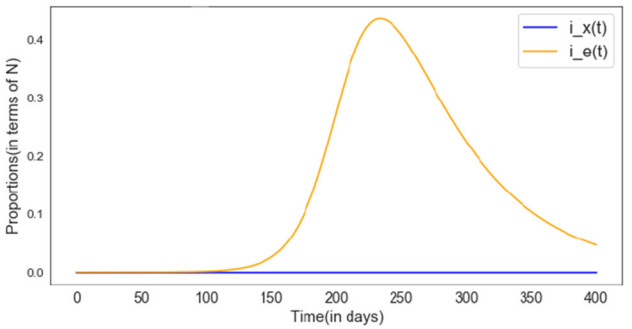

Fig. 14 I_e in the presence of i_x in the state of Kerala. The values of i_x are very small for the scale of this plot. Hence, it is plotted separately. Please refer Fig. 16

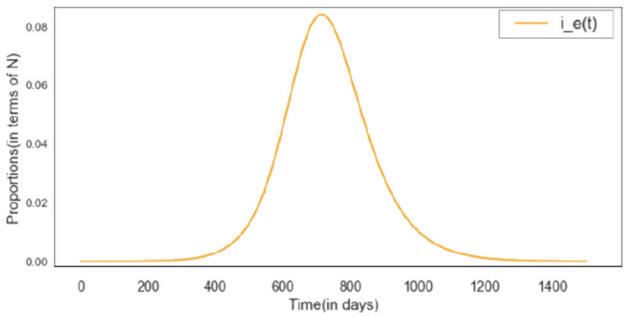

Fig. 15 I_e in the absence of i_x in the state of Kerala

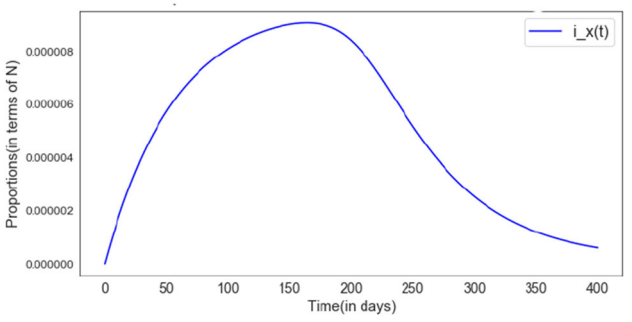

Fig. 16 i_x in Exo-SIR model. Please note that the y axis is in the scale of 10^{-5}

Table 2 Impact of I_x on I_e in the state of Tamil Nadu

	peak value	peak tick
I_e in the presence of i_x	0.1714	907 Days
I_e in the absence of i_x	0.1710	1351 Days

Table 3 Impact of I_x on I_e in the state of Rajasthan

	Peak value	Peak tick
I_e in the presence of i_x	0.3487077	143 Days
I_e in the absence of i_x	0.3486663	147 Days

Table 4 Impact of I_x on I_e in the state of Kerala

	Peak value	Peak tick
I_e in the presence of i_x	0.0842	608 Days
I_e of in the absence of i_x	0.0841	715 Days

present. Also, we can see that the peak tick of i_e is different for the instance when i_x is present.

Finally, we present the comparison of the predictions of Exo-SIR model and SIR model with the real data for the following cases:

1. Covid-19 in Kerala (Fig. 17)
2. Covid-19 in Tamil Nadu (Fig. 18)
3. Covid-19 in Rajasthan (Fig. 19)

Here, the peak values are scaled down as they are very high for both SIR and Exo-SIR predictions. This may be due to the fact that in both SIR and Exo-SIR models, we assume that each infected person is equally likely to infect all the susceptible people. In the real life, this is not true. However, we can see that in all the three cases (shown in Figs. 17, 18, and 19), the peak of the Exo-SIR model is closer to the peak of the real data.

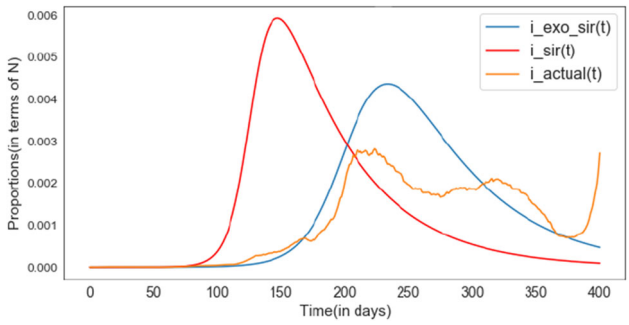

Fig. 17 Comparison of the predictions of Exo-SIR and SIR models with real data for Covid-19 in Kerala

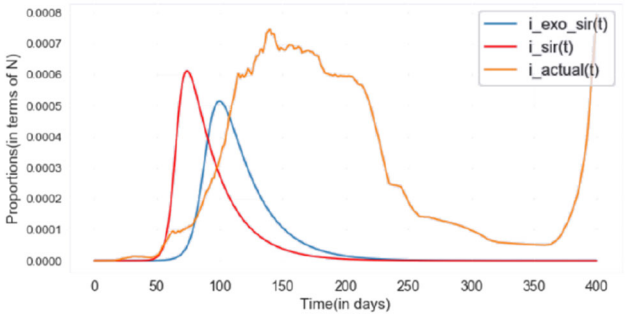

Fig. 18 Comparison of the predictions of Exo-SIR and SIR models with real data for Covid-19 in Tamil Nadu

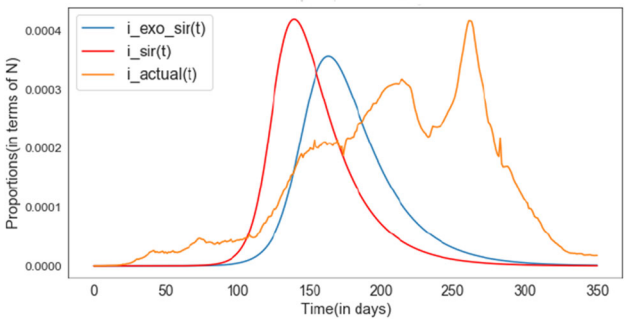

Fig. 19 Comparison of the predictions of Exo-SIR and SIR models with real data for Covid-19 in Rajasthan

6.2 Covid-19 infection in the USA

In this section, we discuss the analysis that we carried out on the data of Covid-19 infection in the USA.

We constructed our dataset from two different sources[14] for our analysis—kaggle.com and incoming tourists travel data for the USA from the CEIC database [15].

Now, we discuss how the values in the dataset is mapped on to the variables in the Exo-SIR model. We calculated the number of endogenous infections ($I_E(t)$) from the following equation.

$$I_E(t) = I_E(t-1) + \text{Daily}(t) - D(t-1) \tag{37}$$

where, Daily(t) is the daily new cases at the time slice t and $D(t-1)$ is the deaths from within the USA population at the time slice $t-1$.

We estimated infected tourists death number from endogenous deaths in the following way. First, we calculated γ from endogenous data by using equation

[14] The code and all the data used in our experiments will be made openly available upon the acceptance of this paper

[15] https://www.ceicdata.com/en/indicator/united-states/visitor-arrivals.

$$\gamma = \frac{\mathrm{d}r/\mathrm{d}t}{i} \tag{38}$$

Applied the same gamma to get the number of deaths from data of exogenous infections using the equation

$$r(t) = r(t-1) + \mathrm{d}r/\mathrm{d}t \tag{39}$$

where

$$\mathrm{d}r/\mathrm{d}t = \gamma * i(t-1) \tag{40}$$

Then, we calculated the number of exogenous infections ($I_X(t)$) by using the equation:

$$I_X(t) = I_X(t-1) \ + \mathrm{Daily}(t) - D(t) \tag{41}$$

where $\mathrm{Daily}(t)$ is the daily new tourist cases at the time slice t and $D(t)$ is the number of deaths at the time slice t

Then, we calculated the number of susceptible people by using the following equation:

$$S(t) = N - I_E^c(t) - I_X^c(t) \tag{42}$$

where $I_E^c(t)$ is the cumulative value of $I_E(t)$ and $I_X^c(t)$ is the cumulative value of $I_X(t)$.

Finally, we computed $\frac{\mathrm{d}(i_e)}{\mathrm{d}t}$, $\frac{\mathrm{d}(i_x)}{\mathrm{d}t}$, $\frac{\mathrm{d}(r)}{\mathrm{d}t}$ and $\frac{\mathrm{d}(s)}{\mathrm{d}t}$ values.

Next, we analyze the Covid-19 data from the USA by applying the Exo-SIR model. We compare the peak tick and peak value of the plot of i_e in the presence and absence of i_x. This would give information about the impact of i_x on i_e. For this purpose, we used Algorithm 3. The cases in the presence and absence of I_x are plotted in Figs. 20 and 21, respectively. I_x is plotted in Fig. 22.

In these plots, it can be observed that the peak and the height of the peak are different compared to the values in the absence of i_x The peak tick and peak values corresponding to the I_e of Exo-SIR model in the presence and absence of i_x are mentioned in Table 5.

Figure 23 shows the comparison of the predictions of Exo-SIR model and SIR model with the real data. Here, we can see

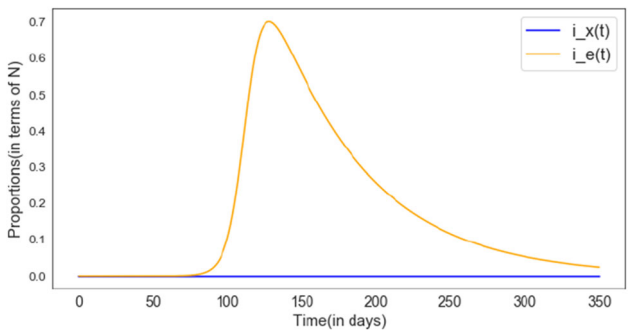

Fig. 20 i_e and i_x for Covid-19 in the USA

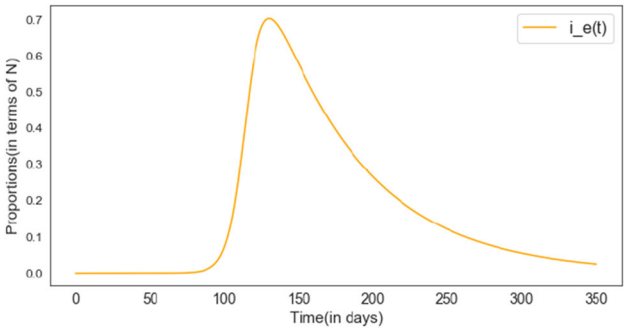

Fig. 21 i_e in the absence of i_x for Covid-19 in the USA

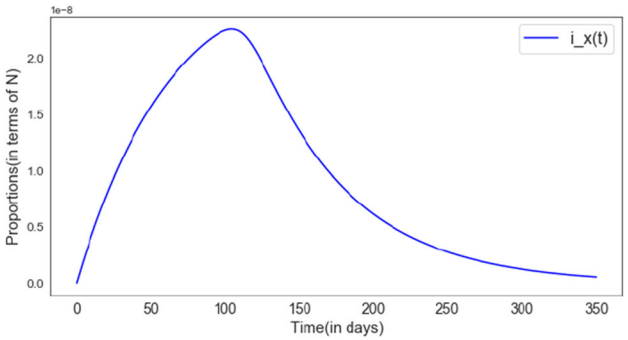

Fig. 22 i_x for Covid-19 in the USA

Table 5 Impact of I_x on I_e in Covid-19 in the USA

	Peak value	Peak tick
I_e in the presence of i_x	0.7	130 days
I_e in the absence of i_x	0.7	135 days

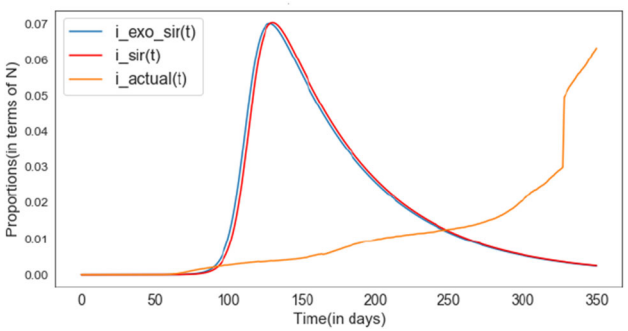

Fig. 23 Comparison of the predictions of Exo-SIR and SIR models with real data for Covid-19 in the USA

that the peaks in the SIR and Exo-SIR plots are of the same height and are coming more or less simultaneously. However, both of them are very different from the peak position in the real data.

6.3 Ebola infection in Guinea

Ebola, also known as EVD, was another severe, often fatal epidemic that hit the Western African countries from 2014 to 2016, particularly Guinea, Sierra Leone, and Liberia. Its fatality rate[16] varies from 25 to 90%. Like the case of Covid-19, there was migration of people from abroad, especially tourists traveling into these countries. The dataset regarding travel and tourism is publicly available[17].

We compared peak tick and peak value of the plot of i_e in the presence and absence of i_x, as per Algorithm 3. This gave us information and important insights on the impact of i_x on i_e.

We constructed our dataset from two different sources: kaggle.com and incoming tourists travel data for Guinea from UNWTO Dashboard[18].

Now, we discuss how the values in the dataset is mapped on to the variables in the Exo-SIR model. We calculated the number of endogenous infections ($I_E(t)$) from the following equation.

$$I_E(t) = I_E(t-1) + M(t) - D(t-1) \tag{43}$$

where $M(t)$ is the monthly new cases at the time slice t and $D(t-1)$ is the deaths from within the Guinea population at the time slice $t-1$.

We estimated infected tourists death number from endogenous deaths in the following way. First, we calculated γ from endogenous data by using the equation

$$\gamma = \frac{\mathrm{d}r/\mathrm{d}t}{i} \tag{44}$$

Then, we applied the same gamma to get the number of deaths from data of exogenous infections using the equation:

$$r(t) = r(t-1) + \mathrm{d}r/\mathrm{d}t \tag{45}$$

where

$$\mathrm{d}r/\mathrm{d}t = \gamma * i(t-1) \tag{46}$$

Then, we calculated number of exogenous infections ($I_X(t)$) by using the equation:

$$I_X(t) = I_X(t-1) + M(t) - D(t) \tag{47}$$

where $M(t)$ is the monthly new tourist cases at the time slice t and $D(t)$ is the number of deaths at the time slice t.

[16] https://www.who.int/health-topics/ebola.

[17] https://www.unwto.org/unwto-tourism-dashboard.

[18] https://www.unwto.org/seasonality.

Then, we calculated the number of susceptible people by using the following equation:

$$S(t) = N - I_E^c(t) - I_X^c(t) \tag{48}$$

where $I_E^c(t)$ is the cumulative value of $I_E(t)$ and $I_X^c(t)$ is the cumulative value of $I_X(t)$.

Finally, we computed $\frac{\mathrm{d}(i_e)}{\mathrm{d}t}$, $\frac{\mathrm{d}(i_x)}{\mathrm{d}t}$, $\frac{\mathrm{d}(r)}{\mathrm{d}t}$ and $\frac{\mathrm{d}(s)}{\mathrm{d}t}$ values.

Next, we analyze the data from Guinea. We compare the peak tick and peak value of the plot of i_e in the presence and absence of i_x. This would give information about the impact of i_x on i_e. For this purpose, we used Algorithm 3. The cases in the presence and absence of I_x are plotted in Figs. 24 and 25, respectively. I_x is shown in Fig. 26.

From Figs. 24, 25, and 26, the peak and the height of the peak are different compared to the values in the absence of i_x. The peak tick and values corresponding to I_e of the Exo-SIR in the presence and absence of i_x are mentioned in Table 6.

Figure 27 shows the comparison of the predictions of Exo-SIR model and SIR model with the real data. Here, we can see that the peak of Exo-SIR and SIR models is coming differently and they are coming far from the peak of the actual data.

6.4 Discussion

Both Covid-19 and Ebola satisfy our hypothesis that the endogenous spread changes in the presence of exogenous

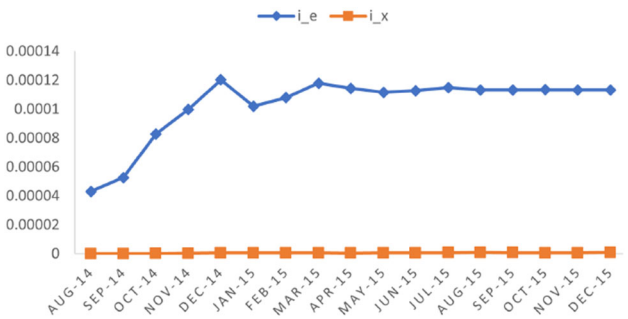

Fig. 24 i_e and i_x for Ebola

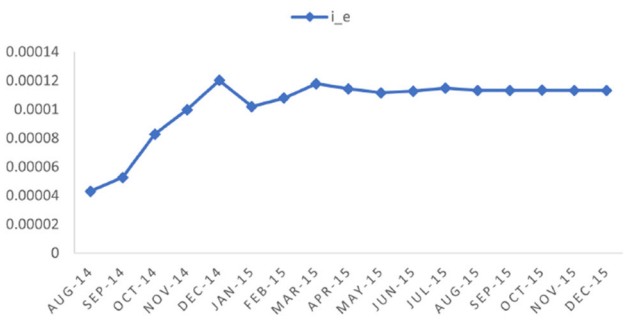

Fig. 25 i_e in the absence of i_x for Ebola

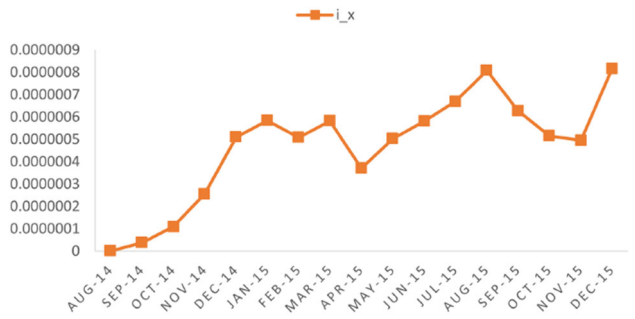

Fig. 26 i_x for Ebola

Table 6 Impact of I_x on I_e in Guinea

	Peak value	Peak tick
I_e in the presence of i_x	0.00012	5 Months
I_e in the absence of i_x	0.00012	6 Months

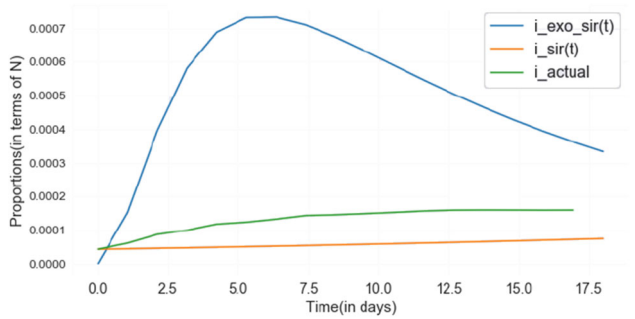

Fig. 27 Comparison of the predictions of Exo-SIR and SIR models with real data for Ebola in Guinea

spread. Also, the results in the case of Covid-19 infection in India show that the Exo-SIR model predicts the epidemic's peak tick better than the SIR model.

Covid-19 in the USA and Ebola in Guinea show less accurate predictions than Covid-19 in India. This may be because of the following reasons.

In these cases, we took the data from the beginning of the spread of the infection. As soon as the infections started growing, the governments began multiple interventions to curb the spread of the epidemics. If these efforts were successful, that would change the values of the constants that we calculated using the initial values. This will reflect in the curve of the real data primarily by delaying the peak and flattening the curve. This can be observed in the real data of Covid-19 in the USA and Ebola in Guinea. On the other hand, in the case of the data from India, we took the data when the migration of people after the Tablighi religious congregation happened. By this time, India was already on the alert, and the government had already intervened in the matter. Hence, our calculation of the constants was closer to the actual values.

We analyzed a sub-event in the case of Covid-19 in India, the Tablighi religious congregation, with many participants from almost all the states in India. The number of these people who traveled back to the states was considered I_x. The probability of these people being infected was very high as the event was a hot spot of the infection. However, in the case of Covid-19 in the USA and Ebola in Guinea, we considered the tourist arrival data as I_x. We made strong assumptions in these cases due to the unavailability of the daily inflow of the infected people to the population. In the case of Covid-19 in the USA, we calculated the external infection as the tourist arrival data multiplied by the total infection in the world. We normalized it by the total population of the world. In the case of Ebola in Guinea, we calculated the external infection as the tourist arrival data multiplied with the total population of the three countries where the infection was the most prevalent and normalized it by the world's total population. In these cases, the probability that all the people in the travel data are infected is comparatively less. This may be the reason for the difference. It is important to note that the SIR model performed equally bad in these cases. This also suggests that the issue might be with the data.

The peak value of the predictions of both SIR and Exo-SIR models was very high compared to the real values. The reason for this may be the following. In the case of SIR and Exo-SIR models, we assume that susceptible people are equally likely to get infected from each infected person in the population. This is not true in real life. In real life, people are likely to get infected only from those they contact. This number is much less than the assumption in both SIR and Exo-SIR models.

7 Conclusion

This study introduced the Exo-SIR model by extending the SIR model. Unlike the other epidemiological models, the Exo-SIR model differentiates between the endogenous and exogenous spread of virus/information. We studied the model in the following ways:

1. Analytical study
2. Simulation considering the presence of contact network of the population and assuming it to be a scale free network
3. Simulation without considering the presence of contact network
4. Implementation of the Exo-SIR model on real data about the spread of Covid-19 in India, Covid-19 in the USA, and the spread of Ebola in Guinea.

We found that all the four analyses mentioned here converge to the same result: the peak comes differently in time and size when the exogenous source is present. We studied the impact of exogenous infection on endogenous diffusion.

We found that exogenous diffusion impacts the endogenous spread of infection. If there are exogenous sources of infection, like in the case of Covid-19 or Ebola, then the Exo-SIR model is more appropriate to estimate the scenario better. This will help the government allocate its resources better as the endogenous and exogenous spread needs different sets of actions to stop them.

Limitations and Future works: We used the SIR model for comparison as it is simple and widely used. Other models like SEIR, SEYAR, etc., that could be used for a similar study. There is scope for introducing the external source of infection to these models like SEIR and SEYAR. Also, we have considered only one external source of infection. There may exist multiple external sources of infection like bats, pigs, birds, etc. Another possible scenario is the possible presence of multiple viruses. We propose to study these in the future.

Acknowledgements We acknowledge partial support from the National Science Foundation (NSF) award #2133842 "EAGER: Advancing Neuro-symbolic AI with Deep Knowledge-infused Learning". Any opinions, findings, and conclusions/recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF. We thank Melissa Nolan and Stella Self from Arnold School of Public Health, the University of South Carolina, for their valuable comments and helpful suggestions. In addition, we thank Ch V Radhasai Rupesh for helping in data collection.

References

1. Harko, T., Lobo, F.S., Mak, M.: Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. Appl. Math. Comput. **236**, 184–194 (2014)
2. Cao, L., Liu, Q., Hou, W.: COVID-19 modeling: a review. arXiv preprint arXiv:2104.12556 (2021)
3. Kumar, A., Gupta, P.K., Srivastava, A.: A review of modern technologies for tackling COVID-19 pandemic. Diabetes Metab. Syndr. Clin. Res. Rev. **14**(4), 569–573 (2020)
4. Kotwal, A., Yadav, A.K., Yadav, J., Kotwal, J., Khune, S.: Predictive models of COVID-19 in India: a rapid review. Med. J. Armed Forces India **76**(4), 377–386 (2020)
5. WHO: Critical preparedness, readiness and response actions for COVID-19: interim guidance, 4 November 2020. Technical documents (2020). https://apps.who.int/iris/handle/10665/336373
6. Asawa, P., Gaur, M., Roy, K., Sheth, A.: COVID-19 in Spain and India: comparing policy implications by analyzing epidemiological and social media data. arXiv preprint arXiv:2010.14628 (2020)
7. Rajan, S.I., Sivakumar, P., Srinivasan, A.: The COVID-19 pandemic and internal labour migration in India: a 'crisis of mobility'. Indian J. Labour Econ. **63**(4), 1021–1039 (2020)
8. Roy, K., Zhang, Q., Gaur, M., Sheth, A.: Knowledge infused policy gradients with upper confidence bound for relational bandits. In: Oliver, N., Pérez-Cruz, F., Kramer, S., Read, J., Lozano, J.A. (Eds.) Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 35–50. Springer, Cham (2021)
9. Roy, K., Zhang, Q., Gaur, M., Sheth, A.: Knowledge infused policy gradients for adaptive pandemic control. arXiv preprint arXiv:2102.06245 (2021)
10. Gaur, M., Kursuncu, U., Sheth, A., Wickramarachchi, R., Yadav, S.: Knowledge-infused deep learning. In: Proceedings of the 31st ACM Conference on Hypertext and Social Media, pp. 309–310 (2020)
11. Zhou, G., Sun, L., Xia, R., Duan, Y., Xu, J., Yang, H., Wang, Y., Lee, M.C., Xiang, Z., Yan, G., et al.: Clinical malaria along the China-Myanmar border, Yunnan Province, China, January 2011-August 2012. Emerg. Infect. Dis. **20**(4), 675 (2014)
12. Brauer, F., Castillo-Chavez, C.: Mathematical Models in Population Biology and Epidemiology. Texts in Applied Mathematics. Springer, New York (2011)
13. Tolles, J., Luong, T.: Modeling epidemics with compartmental models. Jama **323**(24), 2515–2516 (2020)
14. Walker, P.G., Whittaker, C., Watson, O.J., Baguelin, M., Winskill, P., Hamlet, A., Djafaara, B.A., Cucunubá, Z., Olivera Mesa, D., Green, W., et al.: The impact of COVID-19 and strategies for mitigation and suppression in low-and middle-income countries. Science **369**(6502), 413–422 (2020)
15. Goel, R., Bonnetain, L., Sharma, R., Furno, A.: Mobility-based SIR model for complex networks: with case study of COVID-19. Soc. Netw. Anal. Min. **11**(1), 1–18 (2021)
16. Kumar, P., Sinha, A.: Information diffusion modeling and analysis for socially interacting networks. Soc. Netw. Anal. Min. **11**(1), 1–18 (2021)
17. Myers, S.A., Zhu, C., Leskovec, J.: Information diffusion and external influence in networks. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 33–41 (2012)
18. Li, J., Xiong, J., Wang, X.: Measuring the external influence in information diffusion. In: 2015 16th IEEE International Conference on Mobile Data Management. IEEE vol. 2, pp. 92–97 (2015)
19. Yang, D., Liao, X., Wei, J., Chen, G., Cheng, X.: Modeling information diffusion with the external environment in social networks. J. Internet Technol. **20**(2), 369–377 (2019)
20. De, A., Bhattacharya, S., Ganguly, N.: Demarcating endogenous and exogenous opinion diffusion process on social networks. In: Proceedings of the 2018 World Wide Web Conference, pp. 549–558 (2018)
21. Fujita, K., Medvedev, A., Koyama, S., Lambiotte, R., Shinomoto, S.: Identifying exogenous and endogenous activity in social media. Phys. Rev. E **98**(5), 052304 (2018)
22. Agrawal, R., Potamias, M., Terzi, E.: Learning the nature of information in social networks. In: Sixth International AAAI Conference on Weblogs and Social Media (2012)
23. Oka, M., Hashimoto, Y., Ikegami, T.: Self-organization on social media: Endo-Exo bursts and baseline fluctuations. PLoS One **9**(10), 109293 (2014)
24. Crane, R., Sornette, D.: Robust dynamic classes revealed by measuring the response function of a social system. Proc. Natl. Acad. Sci. **105**(41), 15649–15653 (2008)
25. Hari Prasad, P.S.: COVID-19 disease spread modeling by QSIR method: the parameter optimal control approach. Clin. Epidemiol. Glob. Health **13**, 100934 (2022). https://doi.org/10.1016/j.cegh.2021.100934
26. Dandekar, R., Rackauckas, C., Barbastathis, G.: A machine learning-aided global diagnostic and comparative tool to assess effect of quarantine control in COVID-19 spread. Patterns **1**(9), 100145 (2020)
27. Zakary, O., Bidah, S., Rachik, M., Ferjouchia, H.: Mathematical model to estimate and predict the COVID-19 infections in Morocco: optimal control strategy. J. Appl. Math. **2020**, 1–13 (2020). https://doi.org/10.1155/2020/9813926
28. Kaxiras, E., Neofotistos, G.: Multiple epidemic wave model of the COVID-19 pandemic: modeling study. J. Med. Internet Res. **22**(7), 20912 (2020)

29. Chen, Y.-C., Lu, P.-E., Chang, C.-S., Liu, T.-H.: A time-dependent SIR model for COVID-19 with undetectable infected persons. IEEE Trans. Netw. Sci. Eng. **7**(4), 3279–3294 (2020)
30. Jung, S.Y., Jo, H., Son, H., Hwang, H.J.: Real-world implications of a rapidly responsive COVID-19 spread model with time-dependent parameters via deep learning: Model development and validation. J. Med. Internet Res. **22**(9), 19907 (2020)
31. Radcliffe, J.: The mathematical theory of infectious diseases and its applications. J. R. Stat. Soc. Ser. C (Appl. Stat.) **26**(1), 85–87 (1977)
32. Hethcote, H.W.: The mathematics of infectious diseases. SIAM Rev. **42**(4), 599–653 (2000)
33. Szabó, G.M.: Propagation and mitigation of epidemics in a scale-free network. arXiv preprint arXiv:2004.00067 (2020)
34. Barabási, A.-L.: Network science. Philos. Trans. R. Soci. A Math. Phys. Eng. Sci. **371**(1987), 20120375 (2013)