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] and Amit Sheth[2]

[1]Department of Computer Science and Engineering, The LNM Institute of Information Technology, Jaipur, India.
[2]AI Institute, University of South Carolina, USA.

*Corresponding author(s). E-mail(s): nirmal.sivaraman@lnmiit.ac.in;
Contributing authors: MGAUR@email.sc.edu; shivanshbaijal@gmail.com; sakthi.balan@lnmiit.ac.in; AMIT@sc.edu;

Abstract

Epidemics like Covid-19 and Ebola have impacted people's lives greatly. The role of mobility of people across the countries or states in the spread of epidemics have been significant. Consider a population in a region, the spread of disease due to factors local to the region is termed as endogenous spread. If factors contributing to spread is external like migration, mobility, etc., it is termed as exogenous spread. In this paper, we introduce Exo-SIR model that is an extension of the popular SIR model. The novelty in our model is that it captures both the exogenous and endogenous spread of the virus. Also, we study a few variants of the model that are applicable for special scenarios like the presence of social disagreement, presence of different groups that has different amount of risk and infectiousness (e.g. front line workers). We analyze to find out the interplay between endogenous and exogenous infections during the Covid-19 and Ebola pandemics. First, we present an analytical study. Second, we simulate the Exo-SIR model without assuming any contact network for the population. Third, we simulate it by assuming that the contact network is a scale-free network. Fourth, we implement the Exo-SIR model on real datasets regarding Covid-19 and Ebola. We found that endogenous infection is influenced by exogenous infection. We found that Exo-SIR model predicts the time of the peak better than the SIR model. Hence, Exo-SIR model would be helpful for the 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 of time. There are many epidemics that occurred in the world. Covid-19 and Ebola are recent prominent examples.

Recently, people have tried many methods to study Covid-19. A recent work [1] gives a comprehensive review of the methods to model and analyze Covid-19. Out of these methods, the models that are relevant to our model are compartmental models. They are prominent methods used for analysis and prediction of Covid-19 dynamics. The susceptible, infected and recovered model (SIR model) is considered as one of the seminal models of epidemics [2]. 4% of all the literature on Covid-19 modeling uses the SIR model [1]. The SIR model describes the spread of the virus from people to people within the population under consideration. In another review paper [3], six out of the nine techniques to model the spread of the epidemic that are discussed are SIR based compartmental models. It is stated in another review paper [4] that 56% of the all the studies regarding Covid-19 modeling that are published between 21 March 2020 and 13 April 2020 use SIR model or its variants.

However, these works do not consider any external source of infection. 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 are called as *Local transmission and community transmission* and the infection external to the population are called as *Imported cases*. We call the infection from a source within the population and external to the population as endogenous and exogenous spread of infection, respectively. Human migration and mobility are prime reasons behind exogenous spread of infection.

The governments can intervene to curb the spread of the disease by bringing in policies to stop the human mobility. However, the implementations of any such steps by the governments have a lot of challenges. Social disagreement is an example. Social disagreement means people do not abide by 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 measure of intervention, the Governments should try to restrict human mobility. However, for all such human mobility and migration, the governments cannot say no completely. For example, in Indian sub-continent, people migrate to metropolitan cities for work. Due to risk of COVID-19 exposure in these over populated cities, people migrate back to their home [6]. This is also known as reverse migration[2].

The government cannot deny one's right to go home. However, the government can allow it in a controlled manner. For example, when people move from one state to

[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

another, the state governments can issue passes for anyone who wants to travel to that state. Through this they can identify the incoming people and ensure that they properly follow the procedures advised by the respective governments. This allows the states to have proper information regarding the possible exogenous infections. This also gives the state governments time to prepare their medical resources accordingly.

There are challenges even if people do not migrate. For example, the front line workers like doctors and nurses are exposed to the virus much more than a person who is sitting at home and doing her job online. For these groups of people, we have to assume a different rate of infection. The governments will have to make all the necessary safety equipment available to the front line workers and monitor their health constantly to control the infection as a measure of intervention.

In this context, the following are the research questions that we are going to address in this work that significantly modify the current, well situated SIR model:

1. How to quantify the exogenous spread of infection?
2. What is the interplay between exogenous and endogenous spread of infection with respect to 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 different rate of infection - front line workers like doctors and nurses.
3. What is the change in the position of the peak (the largest number of people infected in a unit of time) in the presence of exogenous infection?
4. What is the change in the size of the peak in the presence of exogenous infection?

This paper is structured as follows. Sections 2 and 3 describe the motivation and contributions respectively. Sections 4 and 5, discuss related works and preliminaries. Then, we formulate the Exo-SIR model by extending the SIR model (in Section 6). We discuss the different variants of the model (in Section 7). We compare our model with SIR model (in Section 8). We study the behavior of infected population in the presence and absence of exogenous infection analytically (in Section 9). Then we describe the simulation study where we simulated the SIR model and Exo-SIR model and compared them (in Section 10). Finally we study the real data of Covid-19 and Ebola epidemics (in Section 11). The workflow is given in Figure 1.

2 Motivation

The traditional compartmental models consider an initial outbreak and the propagation of the infection from one person to the other within a closed population to model the dynamics of epidemics. This closed population may be an administrative unit like a country, state, district, etc. However, infections may happen due to the mobility of infected people from outside to the population under consideration. We call these cases

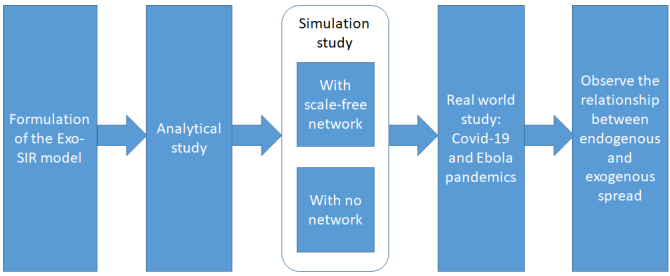

Fig. 1 The workflow of the study.

of infections as exogenous infections. During pandemics people movement cannot be fully stopped. Governments will have to admit influx of people at least in a controlled manner. For example, the Government may issue passes to people who wish to move into the country. In this case, the Government will get the exact data of the number of people coming into the population every day. These restricted movements will increase the exogenous spread of the infections. In order to find what will be the amount of infection during this movement and when the peak will occur authorities need an explicit model that can predict infection through the exogenous means. Our model extends the SIR model and explicitly takes care of the amount of exogenous infection in addition to the endogenous infections. This will help the authorities to prepare their medical resources accordingly. In case of the spread of epidemics, even if there is a small increase in the number of infected people, the impact grows exponentially with respect to time. Hence, considering the exogenous infections while studying the dynamics of epidemics is very important.

3 Contributions

The following are our contributions in this work:

- We study and address the impact of external reasons of infections such as cross-border mobility on COVID infection by introducing a novel SIR-like compartmental model.
- We study a few variants of the model that are applicable for special scenarios like the presence of social disagreement, presence of different groups that has different amount of risk and infectiousness like the front line workers.
- We analyze to find out 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 without assuming any contact network for the population.
  3. By simulating the Exo-SIR model by assuming that the contact network is a scale-free network.
  4. By implementing the Exo-SIR model on real datasets regarding Covid-19 and Ebola.

- We compare the predictions of our model and that of the SIR model with the real data on the recent spread of the Covid-19 in India and the USA and the spread of Ebola in Africa.

4 Related works

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

The work in [7] 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 stage of the epidemic. Deterministic models often provide useful ways of gaining sufficient understanding about the dynamics of populations whenever they are large enough [8]. Also, the deterministic models are simpler and more popular [9, 10]. Our model is a deterministic model.

4.1 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 [11]. There are a few recent works in the literature that attempt to model the external influence in information diffusion in online social networks [12]. Moreover, [12] and [13] propose information diffusion model on the network. In these works, they assume that the information flows through an underlining network. Also, they consider links from other web sites like the mainstream media as the external sources of information. The internal diffusion are those when the messages that are shared do not have any external links.

The work described in [12] uses very specific parameters like the following:

- probability of any node receiving an 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 the node i has received n exposures by time t

Whereas, the work described in [13] trace the cascade and reconstruct the graph as much as possible. Also, they conclude that external influence has bigger impact on the network when compared to the influence of the social media influencers.

The model that is closest to our work is Yang et al's model [14], which is an extension of the SIR model (explained in Section 5) to include the external influence to the network. State transition diagram of the diffusion mechanisms of this model is given in Figure 2.

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

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

$$-(1 - p_1)p_5 ksi$$

$$\frac{di}{dt} = p_1 ksi + p_4 \theta s - p_2 i \tag{2}$$

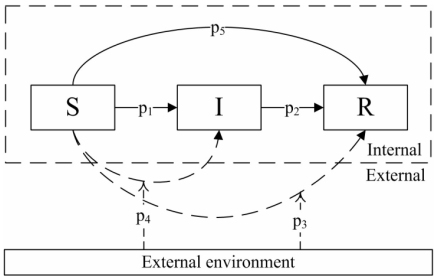

Fig. 2 State transition diagram of the diffusion mechanisms in Yang et al's model. Diagram taken from [14].

$$\frac{dr}{dt} = p_2 i + (1-p_4)p_3\theta s + (1-p_1)p_5 ksi \tag{3}$$

There are two possible transitions from the state S to I. One path is the normal endogenous path, and the second one 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 state I to R does not affect external influence(s).

Although the exogenous infection is modeled in Yang et al.'s model, it fails to capture the dynamics between the endogenous and exogenous infections.

4.2 Other studies of endogenous and exogenous information diffusion

The dual nature of message flow over the online social network is studied and verified in [15]. Here, the dual nature refers to the injection of exogenous opinions to the network and the endogenous influence-based dynamics. In [16], the authors propose a method for extracting the relative contributions of exogenous and endogenous contents. In [17], 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 [18], 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 [19] 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 burst. Also, those bursts that gradually increase and slowly decrease are endogenous.

4.3 Compartmental models for Covid-19 modeling

Compartmental models are prominent methods that are used for analysis and prediction of Covid-19 dynamics. The SIR model is one of the seminal compartmental models. There are many compartmental models that have come up recently to improve the SIR model. QSIR model [20, 21] 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 [22] 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 totally controlled. Totally 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 [23] as its name suggests models the multiple waves of infection that could occur. Time-dependent SIR model [24, 25] considers the constants in SIR model - beta and gamma to be varying with respect to time. However, none of these models take care of infections arising from outside the population mostly due to cross-border mobility of infected people. Hence, we introduce the Exo-SIR model to address this particular issue.

5 Preliminaries

In this section, we briefly review the SIR epidemiological model. It describes how epidemics spread through a population. This model is used for the study of information diffusion approximating that the way epidemics spread and the way information gets diffused in a population are similar.

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 models the dynamics of the spread [26]:

$$s(t) + i(t) + r(t) = 1 \tag{4}$$

$$\frac{ds}{dt} = -\beta si \tag{5}$$

$$\frac{di}{dt} = \beta si - \gamma i \tag{6}$$

$$\frac{dr}{dt} = \gamma i \tag{7}$$

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.

6 The Model

The Exo-SIR model is an extension of SIR model in which the nodes have Susceptible, Infected and Recovered states. 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 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 that got affected by endogenous source and exogenous sources move into different states. We assume that susceptible nodes get infected from only one of these sources and never from both

the sources. Hence, even when some nodes are susceptible to both endogenous and exogenous infection, they become infected by either an endogenous or an exogenous source. 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 infects 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 Figure 3.

We classify susceptible nodes into two different types – susceptible only to external infection and susceptible only to internal infection.

$$s_x + s_e = s \tag{8}$$

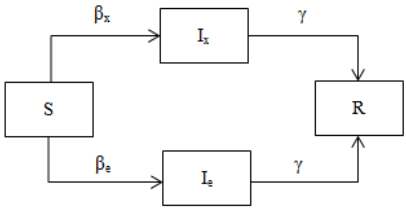

Fig. 3 State transition diagram of the nodes in the Exo-SIR model

We classify infected nodes into two different types – infected from exogenous source and infected from endogenous source.

$$i_e + i_x = i \tag{9}$$

We assume that the total population remains constant.

$$s + i + r = 1 \tag{10}$$

a fraction of the susceptible people (s) gets infected by exogenous source and another fraction of s gets infected by endogenous sources. For endogenous infection, the population that is infected plays a big role. Hence, we can say that:

$$\frac{ds}{dt} = -\beta_x s - \beta_e si \tag{11}$$

Change in i_x is influenced by all the susceptible nodes that are prone to the exogenous influence.

$$\frac{di_x}{dt} = \beta_x s - \gamma i_x \tag{12}$$

Change in i_e is influenced by all the susceptible nodes and infected nodes.

$$\frac{di_e}{dt} = \beta_e si - \gamma i_e \tag{13}$$

Change in r depends on the number of infected people in the network.

$$\frac{dr}{dt} = \gamma i \tag{14}$$

7 Variants of the model to address specific situations

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

7.1 Exo-SIR Model with social disagreement

This scenario occurs when people do not behave according to the way the Government orders. For example, not wearing masks, not following social distancing, etc. As a result, more people get in contact with the virus and hence the infectiousness of the disease will go up. This can be represented in the Exo-SIR model by reducing the β_e value.

7.2 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 make sure that these people are isolated and given treatment so that they do not spread the virus further. Change in i_x is influenced by the action of the government that allowed people to travel. Hence, the efficiency of the planning and execution to minimize the impact is important. This is captured in β_x. If the government is efficient in containing the infection from these people, then the value of β_x goes down.

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

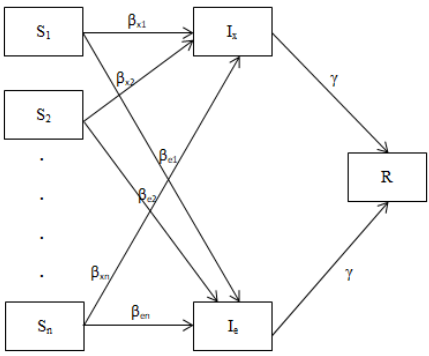

Fig. 4 State transition diagram of the model.

This case may be depicted as shown in Figure 4. In this case, there are n different groups of susceptible people that have different levels of risk of infection. 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 different group of people. Hence, for each group of people, we have different values for the parameters. Hence, there will be the summation of the n groups and parameters for each of the groups. The state diagram and equations are given below.

$$i_e + i_x = i \tag{15}$$

$$s = \sum_{k=1}^{n} s_k \tag{16}$$

$$s + i + r = 1 \tag{17}$$

$$\frac{ds}{dt} = -\sum_{k=1}^{n} \beta_{xk} s_k - \sum_{k=1}^{n} \beta_{ek} s_k i \tag{18}$$

$$\frac{di_x}{dt} = \sum_{k=1}^{n} \beta_{xk} s_k - \gamma i_x \tag{19}$$

$$\frac{di_e}{dt} = \sum_{k=1}^{n} \beta_{ek} s_k i - \gamma i_e \tag{20}$$

$$\frac{dr}{dt} = \gamma i \tag{21}$$

8 Comparison with SIR model

Mirroring the rate of change of s(t), i(t), and r(t) in the SIR model (Section 4), 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:

$$\frac{ds}{dt} = -\beta_x s - \beta_e si \tag{22}$$

Rate of Change of r:

$$\frac{dr}{dt} = \gamma i \tag{23}$$

Differentiating Eq 15 with respect to time, we get

$$\frac{di}{dt} = \frac{di_e}{dt} + \frac{di_x}{dt} \tag{24}$$

$$\frac{di}{dt} = \beta_e si - \gamma i_e + \beta_x s - \gamma i_x \tag{25}$$

$$\frac{di}{dt} = \beta_e si + \beta_x s - \gamma(i_x + i_e) \tag{26}$$

Applying Eq 15 on Eq 26, we get

$$\frac{di}{dt} = \beta_e s(i_x + i_e) + \beta_x s - \gamma(i_x + i_e) \tag{27}$$

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

$$\frac{di}{dt} = \beta_x s \tag{28}$$

This shows that unlike the SIR model, the Exo-SIR model models the situation when no one is infected initially. SIR model assumes that there is an initial outbreak size i_0. This means i_0 people are infected in the beginning and $i_0 > 0$ [27]. 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* $\beta_x = 0$.

9 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 ($\frac{di_e}{dt}$).

Applying Eq 15 on Eq 20, we get

$$\frac{di_e}{dt} = \beta_e s(i_e + i_x) - \gamma i_e \tag{29}$$

$$\left.\frac{di_e}{dt}\right|_{i_x>0} = \beta_e s(i_e + i_x) - \gamma i_e \quad (30)$$

At $i_x = 0$,

$$\left.\frac{di_e}{dt}\right|_{i_x=0} = \beta_e s i_e - \gamma i_e \quad (31)$$

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

$$\left.\frac{di_e}{dt}\right|_{i_x=0} < \left.\frac{di_e}{dt}\right|_{i_x>0} \quad (32)$$

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

10 Simulation

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

1. By assuming that the people network is 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
2. By assuming no network (well-mixed population). Contrary to the above, in this scenario, a susceptible node can get infected from any of the infected nodes in the population under consideration.

We chose Barabási-Albert network because there are pieces of evidence that the human disease network could be scale free [28]. The results of these are discussed in the following section.

10.1 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 [29]. 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.

10.1.1 Dynamics of exogenous spread on endogenous spread

In this section, we determine the changes in endogenous spread with the change in exogenous factors through simulation. The step by step methodology adopted to carry out the simulation, and the analysis is given in Algorithm 1.

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 structure of the network since the setting up of a network in step 3 in the above algorithm is random each time.

Sample simulation results are shown in Figures 5 and 6. Figure 5 shows the SIR model's simulation

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 [29].
4: Simulate and predict the values of S, I_e, I_x, and R using the Exo-SIR model.
5: Extract the values of endogenous peak and its time slice, exogenous peak, and time slice from each of the simulations.
6: Calculate the mean peak value and peak tick of Exo and Endo nodes so that we have one value per combination of β_x, β_e, and γ.

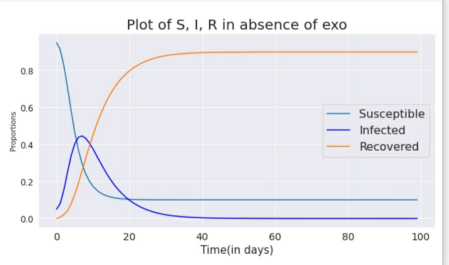

Fig. 5 plot of susceptible, infected and recovered with no exogenous source

results with no exogenous influence, and Figure 6 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 show changes.

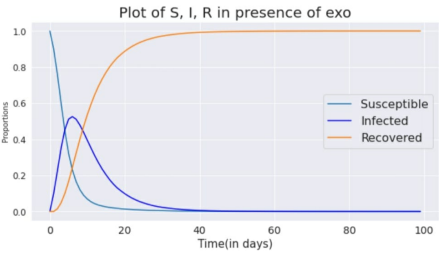

Fig. 6 plot of susceptible, infected and recovered with exogenous source

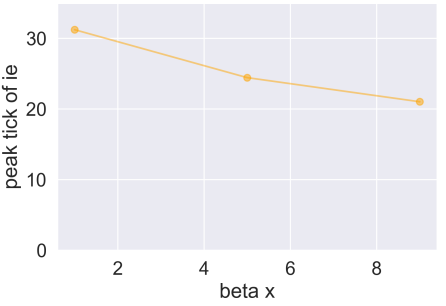

Fig. 7 impact of β_x on peak tick of i_e

Figures 7 and 8 are a result of simulation and analysis done as described in Algorithm 1 and provide us with the following insights:

- *Peak Value:* 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
- *Peak tick:* On the other hand, Figure 8 shows that endogenous peak tick decreases with increase in β_x

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

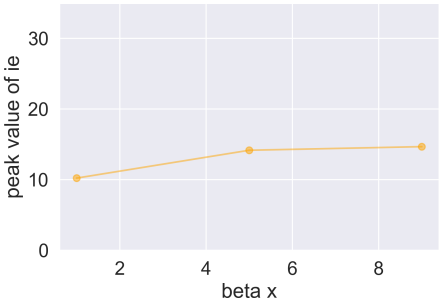

Fig. 8 impact of β_x on peak value of i_e

Table 1 Impact of β_e, β_x and γ on ln(`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

10.2 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 consistent 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.

The following inferences can be drawn from the results of Regression Analysis.

1. 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.
2. The adjusted R-squared value is maximum(0.70) when all the three parameters are considered

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 ln(`Ie_peak`) 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 uniformly.
2: Initialize the initial number of susceptible, infected(endo and exo) 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, we predicted the endogenous and exogenous peak value and peak tick using the Exo-SIR model.
4: Then, we computed the natural logarithm of the peak value and scaled it between 0 and 1.
5: Finally, we fitted an OLS Regression Model with β_x, β_e and γ as the independent variable and ln(`Ie_peak`) and the independent variable and analysed the coefficients statistically.

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.

3. 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.

11 Analysis using real data

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

11.1 Covid-19 infection in India

Covid-19 has caused large and persistent negative effects on the world economy[3]. 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.

There were many sub-events in India that involve 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[4], laborers working in different states or in other countries moving back to their native places [6] 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 1^{st} March 2020 to 21^{st} March 2020[5]. Over 9000 people from various states of India participated in this event[6]. Nearly 4300 cases have been reported that can be traced to the event[7]. As of 18^{th} April 2020, 30% of the cases in India were due to this event[8]. The number of people from each state 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 phenomena that changed the dynamics of the spread of the virus.

We apply the Exo-SIR model on real dataset regarding the spread of the Covid-19 pandemic in the Indian states of Rajasthan, Tamil Nadu, and Kerala from 14^{th} March, 2020 to 14^{th} April, 2020. Exogenous spread dominates endogenous spread in Tamil Nadu, whereas the contrary is true in the case of Rajasthan. In Kerala, both the endogenous and exogenous spread have roughly the same prevalence. The trends in the analytical study, results of the simulations, and the analysis of the real dataset are consistent. The summary of our approach is given in Figure 1. The details of the

[3] shorturl.at/qBZ05

[4] shorturl.at/imBGK

[5] https://en.wikipedia.org/wiki/2020_Tablighi_Jamaat_coronavirus_hotspot_in_Delhi

[6] shorturl.at/qryKU

[7] shorturl.at/myFQ2

[8] shorturl.at/iyVY9

dataset that we used to infer these relations are given in Section 11.1.1.

There are some works in the literature that model mobility of population like [30–32]. However, they do not differentiate between endogenous and exogenous mobility. Not modeling the external sources of infection could give rise to inaccurate estimation of the scenario. The compartmental models like SIR model [33, 34], assume that there are a finite number of infections in the closed community itself at the initial stage. With that assumption, it predicts the number of people infected at a later instant of time. However, the compartmental models do not differentiate the infections with respect to the sources of it. It is important as they need different kinds of intervention from the government.

We analyzed the data of three states in India – 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.

11.1.1 Datasets

The data pipeline is given in Figure 9. We constructed our dataset from three different sources[9] for our analysis – Covid19india.org, the government web site of Tamil Nadu for their press release to find the daily number of Tablighi cases and Wikipedia page on state wise daily data.

Covid19india.org is a publicly available volunteer-driven dataset of Covid-19 statistics in India[10]. There are multiple files in this dataset. One of which is called *raw data* that captures the anonymised details of the patients. In the raw data, the columns of interest for us are DateAnnounced, DetectedState and TypeOfTransmission.

Another file from same source 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 that we created are daily confirmed, daily deceased, daily recovered, date, total confirmed, total deceased, total recovered and daily imported cases.

Another dataset that we used is the compilation of the press releases (news bulletins) from the governments of the states under study. This is to get the daily number of cases due to a major sub event of Covid-19 spread in Inidia – Tablighi Jamaat religious congregation. We manually went through the press releases and collected the data.

11.1.2 Mapping

In this section, 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,

$$s(t) + i(t) + r(t) = 1 \qquad (33)$$

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

[10] www.covid19india.org

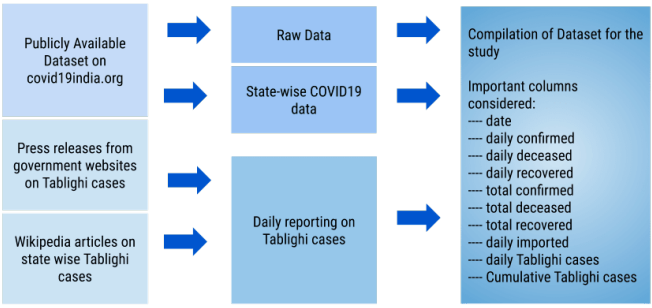

Fig. 9 The data pipeline for Covid-19 data

$$\frac{ds}{dt} = -(\frac{di}{dt} + \frac{dr}{dt}) \tag{34}$$

$\frac{di_e}{dt}$ is the daily confirmed cases on day k.

$\frac{di_x}{dt}$ is the sum of daily imported cases on day k and the daily cases due to Tablighi event on day k.

$$\frac{di}{dt} = \frac{di_e}{dt} + \frac{di_x}{dt}$$

$\frac{dr}{dt}$ is the sum of the numbers of the daily recovered and the daily deceased cases on day k.

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

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 and r is the sum of the total numbers of the deceased and the recovered cases.

11.1.3 Applying Exo-SIR model on real data of Covid-19

In this section, we analyze the data from the states of 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.

Algorithm 3 Algorithm to plot the Exo-SIR model.

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

For the state of Tamil Nadu, the plots of I_e in the presence and absence of i_x are plotted in Figure 10 and Figure 11 respectively. For the state of Rajasthan, the plots of I_e in the presence and absence of i_x are plotted in Figure 13 and Figure 14 respectively. For the state of Kerala, the plots of I_e in the presence and absence of i_x are plotted in Figure 16 and Figure 17, respectively. In all these plots, we can see that i_x is very small compared to i_e. Yet, i_x is having an

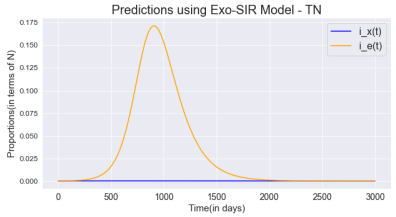

Fig. 10 I_e in the presence of i_x. The values of i_x are very small for the scale of this plot. Hence it is plotted separately. Please refer the Figure 12

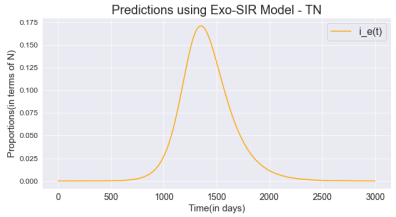

Fig. 11 I_e in the absence of i_x

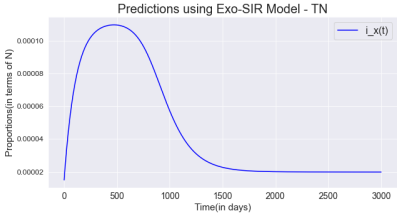

Fig. 12 i_x in Exo-SIR model. Please note that y axis is in the scale of 10^{-4}.

impact on i_e. I_x is plotted separately in Figure 12, Figure 15 and Figure 18.

Here, we assume that there is a constant movement of people into the population from outside that is represented as a fraction of the susceptible population. Figure 12, Figure 15 and Figure 18 follow this assumption. However, this assumption is not exactly true, it is only an approximation. We made this assumption for the simplicity of the analysis.

The peak tick and peak values corresponding to the I_e of Exo-SIR

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

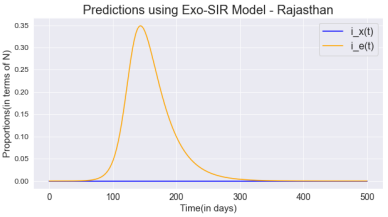

Fig. 13 I_e in the presence of i_x. The values of i_x are very small for the scale of this plot. Hence it is plotted separately. Please refer the Figure 15

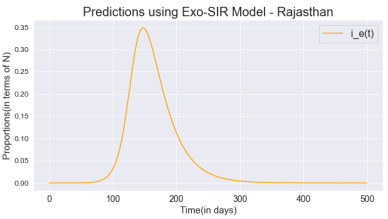

Fig. 14 I_e in the absence of i_x

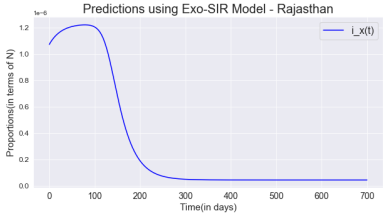

Fig. 15 i_x in Exo-SIR model. Please note that the y axis is in the scale of 10^{-6}.

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

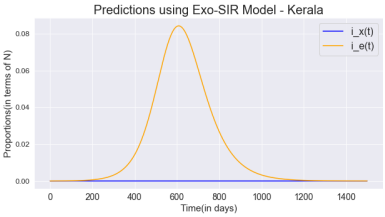

Fig. 16 I_e in the presence of i_x in the state of Kerala. The values of i_x are very small for the scale of this plot. Hence it is plotted separately. Please refer the Figure 18

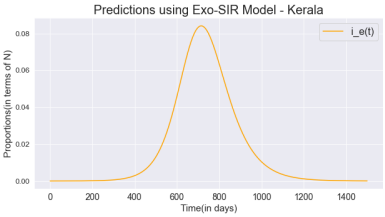

Fig. 17 I_e in the absence of i_x in the state of Kerala

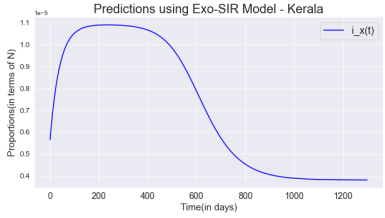

Fig. 18 i_x in Exo-SIR model. Please note that the y axis is in the scale of 10^{-5}.

model in the presence and absence of i_x for Tamil Nadu, Rajasthan and Kerala are mentioned in Table 2, Table 3 and Table 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 present. Also, we can see that the peak tick of i_e is different for the instance when i_x is present.

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

11.1.4 Comparison of Exo-SIR and SIR model with the real data

In this section, 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 (Figure 19)
2. Covid-19 in Tamil Nadu (Figure 20)
3. Covid-19 in Rajasthan (Figure 21)

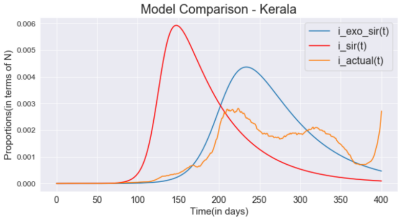

Fig. 19 Comparison of the predictions of Exo-SIR and SIR models with real data for Covid-19 in Kerala

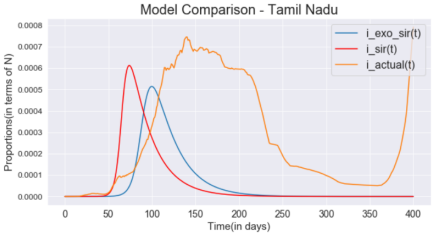

Fig. 20 Comparison of the predictions of Exo-SIR and SIR models with real data for Covid-19 in Tamil Nadu

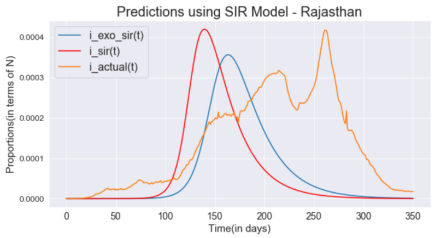

Fig. 21 Comparison of the predictions of Exo-SIR and SIR models with real data for Covid-19 in Rajasthan

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, the peak of the Exo-SIR model is closer to the peak of the real data.

11.2 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.

11.2.1 Dataset

The data pipeline is given in Figure 22. We constructed our dataset from two different sources[11] for our analysis – Kaggle.com and incoming tourists travel data for the USA from the CEIC database [12].

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

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

11.2.2 Mapping

We calculated the number of endogenous infections ($I_E(t)$) from the following equation.

$$I_E(t) = I_E(t-1) + Daily(t) - D(t-1)$$

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:

$$\gamma = \frac{dr/dt}{i}$$

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

$$r(t) = r(t-1) + dr/dt$$

where

$$dr/dt = \gamma * i(t-1)$$

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

$$I_X(t) = I_X(t-1) + Daily(t) - D(t)$$

where $Daily(t)$ is the daily new tourist cases at the time slice t and $D(t)$ is the number of Exo Deaths at the time slice t

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

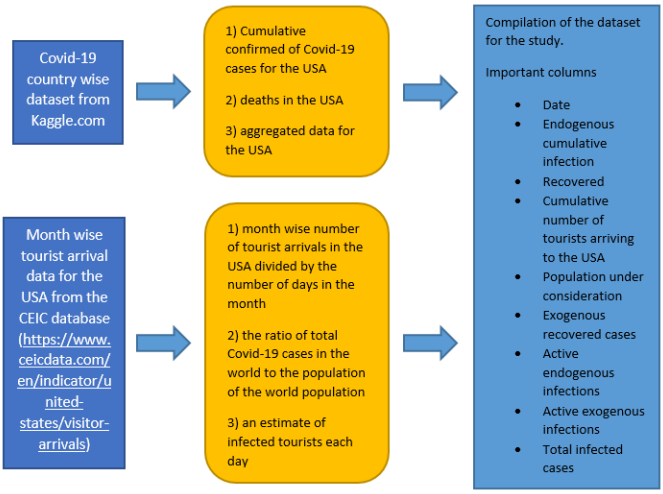

Fig. 22 The data pipeline for Covid-19 data in the USA

$$S(t) = N - I_E^c(t) - I_X^c(t)$$

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

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

11.2.3 Applying Exo-SIR model on real data of Covid-19 in the USA

In this section, 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 is plotted in Figure 23 and 24 respectively. I_x is plotted in the Figure 25.

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.

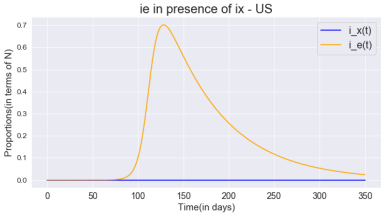

Fig. 23 i_e and i_x for Covid-19 in the USA

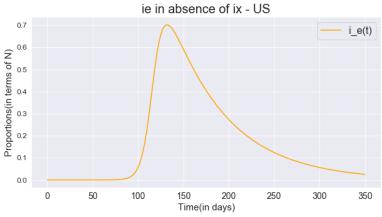

Fig. 24 i_e in the absence of i_x for Covid-19 in the USA

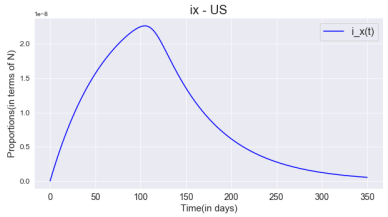

Fig. 25 i_x for Covid-19 in the USA

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

11.2.4 Comparison of Exo-SIR and SIR model with the real data

Figure 26 shows the comparison of the predictions of Exo-SIR model and SIR model with the real data. Here, we can see that the peaks in the SIR and Exo-SIR plots are of the same height and are coming more or less at the same time. However, both of them are very different from the position of the peak in the real data.

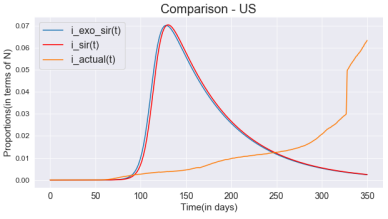

Fig. 26 Comparison of the predictions of Exo-SIR and SIR models with real data for Covid-19 in the USA

11.3 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[13] varies from 25% to 90%. Like the case of Covid-19, there was migration of people from abroad, especially there were tourists travelling into these countries. The dataset regarding travel and tourism is publicly available [14].

We compared the 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.

11.3.1 Dataset

The data pipeline is given in Figure 27. We constructed our dataset from two different sources[15] for our analysis – Kaggle.com and incoming tourists travel data for Guinea from UNWTO Dashboard [16].

11.3.2 Mapping

We calculated the number of endogenous infections ($I_E(t)$) from the following equation.

$$I_E(t) = I_E(t-1) + M(t) - D(t-1)$$

where, $M(t)$ is the monthly new cases at the time slice t and $D(t-1)$

[13] https://www.who.int/health-topics/ebola
[14] https://www.unwto.org/unwto-tourism-dashboard
[15] The code and all the data used in our experiments will be made openly available upon the acceptance of this paper
[16] https://www.unwto.org/seasonality

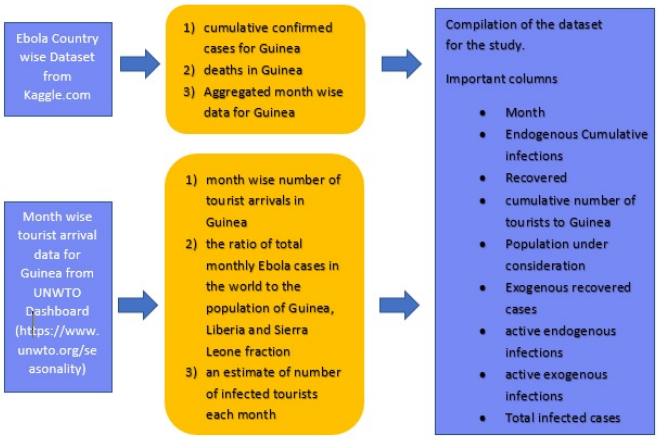

Fig. 27 The data pipeline for Ebola data

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

$$\gamma = \frac{dr/dt}{i}$$

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

$$r(t) = r(t-1) + dr/dt$$

where

$$dr/dt = \gamma * i(t-1)$$

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

$$I_X(t) = I_X(t-1) \ + M(t) - D(t)$$

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

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

$$S(t) = N - I_E^c(t) - I_X^c(t)$$

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

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

11.3.3 Applying Exo-SIR model on real data of Ebola infection in Guinea

In this section, 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 is plotted in Figure 28 and 29 respectively. I_x is plotted in the Figure 30.

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 6.

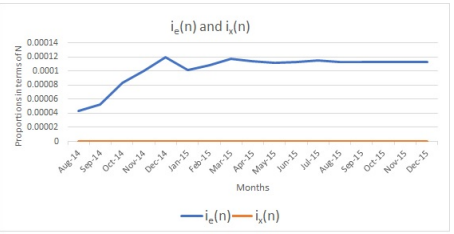

Fig. 28 i_e and i_x for Ebola

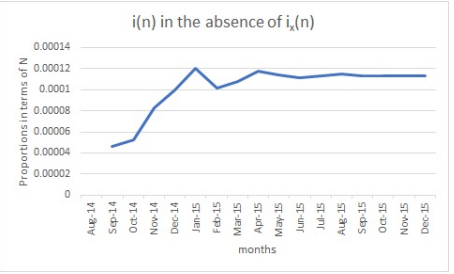

Fig. 29 i_e in the absence of i_x for Ebola

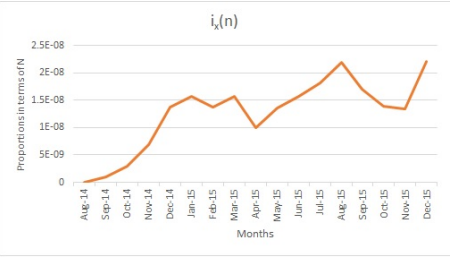

Fig. 30 i_x for Ebola

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

11.3.4 Comparison of Exo-SIR and SIR model with the real data

Figure 31 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 are coming differently and they are coming far from the peak of the actual data.

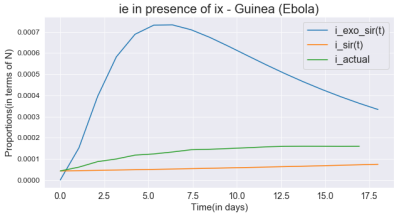

Fig. 31 Comparison of the predictions of Exo-SIR and SIR models with real data for Ebola in Guinea

11.4 Discussion

Both Covid-19 and Ebola satisfy our hypothesis that the endogenous spread changes in the presence of exogenous spread. Also, the results in the case of Covid-19 infection in India show that Exo-SIR model predicts the peak tick of the epidemic better compared to the SIR model.

Covid-19 in the USA and Ebola in Guinea show less accurate prediction compared to the 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 started 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 mostly 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 Thablighi religious congregation happened. By this time, India were already on the alert and the government had already intervened in the matter. Hence, our calculation of the constants were closer to the real values.

- We analyzed a sub event in case of Covid-19 in India, which is the Thablighi religious congregation with many participants from almost all the states in India. The number of these people who travelled back to the states were considered as I_x. The probability of these people being infected was very high as the event was a hot-spot of the infection. However, in 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 data of 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 with the total infection in the world and 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 total population of the world. In these cases, the probability of all the people in the travel data being infected is comparatively less. This may be the reason for the difference. It may also be noted that the SIR model also performed equally badly 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 were very high compared to the real values. The reason for this may be the following. In case of both SIR and Exo-SIR models, we assume that all the susceptible people are equally likely to get infected from each of the infected people in the population. This is not true in the real life. In real life, people are likely to get infected only from those that they come in contact with. This number is much less compared to the assumption in both SIR and Exo-SIR models.

12 Conclusion

In this study, we 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 contact network of the population to be a scale free network
3. Simulation without considering the 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 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 to allocate its resources in a better way as the endogenous and exogenous spread needs different sets of actions to stop them.

Limitations and Future works: We used SIR model for comparison as it is simple and widely used. There are other models like SEIR, SEYAR, etc. that could be used for 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. It is possible that there exists 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.

13 Acknowledgement

Amit Sheth and Manas Gaur are supported by 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, University of South Carolina for their valuable comments and helpful suggestions. We thank Ch V Radhasai Rupesh for helping in data collection.

References

[1] Cao, L., Liu, Q., Hou, W.: COVID-19 modeling: A review. arXiv preprint arXiv:2104.12556 (2021)

[2] 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. Applied Mathematics and Computation **236**, 184–194 (2014)

[3] Kumar, A., Gupta, P.K., Srivastava, A.: A review of modern technologies for tackling COVID-19 pandemic. Diabetes

& Metabolic Syndrome: Clinical Research & Reviews **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. Medical Journal Armed Forces India **76**(4), 377–386 (2020)

[5] World Health Organization and others, critical preparedness, readiness and response actions for covid-19: interim guidance, 4 november 2020. Technical report, World Health Organization (2020)

[6] Rajan, S.I., Sivakumar, P., Srinivasan, A.: The COVID-19 pandemic and internal labour migration in India: A 'crisis of mobility'. The Indian Journal of Labour Economics, 1–19 (2020)

[7] 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. Emerging infectious diseases **20**(4), 675 (2014)

[8] Brauer, F., Castillo-Chavez, C., Castillo-Chavez, C.: Mathematical models in population biology and epidemiology **2** (2012)

[9] Tolles, J., Luong, T.: Modeling Epidemics with compartmental models. Jama (2020)

[10] Walker, P.G., Whittaker, C., Watson, O.J., Baguelin, M., Winskill, P., Hamlet, A., Djafaara, B.A., Cucunubá, Z., Mesa, D.O., Green, W., et al.: The impact of COVID-19 and strategies for mitigation and suppression in low-and middle-income countries. Science (2020)

[11] Kumar, P., Sinha, A.: Information diffusion modeling and analysis for socially interacting networks. Social Network Analysis and Mining **11**(1), 1–18 (2021)

[12] 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)

[13] Li, J., Xiong, J., Wang, X.: Measuring the external influence in information diffusion. In: 2015 16th IEEE International Conference on Mobile Data Management, vol. 2, pp. 92–97 (2015). IEEE

[14] Yang, D., Liao, X., Wei, J., Chen, G., Cheng, X.: Modeling information diffusion with the external environment in social networks. Journal of Internet Technology **20**(2), 369–377 (2019)

[15] 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)

[16] Fujita, K., Medvedev, A., Koyama, S., Lambiotte, R., Shinomoto, S.: Identifying exogenous and endogenous activity in social media. Physical Review E **98**(5), 052304 (2018)

[17] 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)

[18] Oka, M., Hashimoto, Y., Ikegami, T.: Self-organization on social media: Endo-exo bursts and baseline fluctuations. PLoS One **9**(10), 109293 (2014)

[19] Crane, R., Sornette, D.: Robust dynamic classes revealed by measuring the response function of a social system. Proceedings of the National Academy of Sciences **105**(41), 15649–15653 (2008)

[20] subrahmanya hari Prasad Peri: COVID-19 disease spread modeling by QSIR method: The parameter optimal control approach. Clinical Epidemiology and Global Health, 100934 (2021). https://doi.org/10.1016/j.cegh.2021.100934

[21] 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)

[22] Zakary, O., Bidah, S., Rachik, M., Ferjouchia, H.: Mathematical model to estimate and predict the COVID-19 infections in morocco: Optimal control strategy. Journal of Applied Mathematics **2020** (2020)

[23] Kaxiras, E., Neofotistos, G.: Multiple epidemic wave model of the COVID-19 pandemic: modeling study. Journal of medical Internet research **22**(7), 20912 (2020)

[24] 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 Transactions on Network Science and Engineering **7**(4), 3279–3294 (2020)

[25] 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. Journal of medical Internet research **22**(9), 19907 (2020)

[26] Bailey, N., et al.: The mathematical theory of infectious diseases and its applications. The mathematical theory of infectious diseases and its applications. 2nd edition. (2nd ediition) (1975)

[27] Hethcote, H.W.: The mathematics of infectious diseases. SIAM review **42**(4), 599–653 (2000)

[28] Szabó, G.M.: Propagation and mitigation of epidemics in a scale-free network. arXiv

preprint arXiv:2004.00067 (2020)

[29] Barabási, A.-L.: Network science. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **371**(1987), 20120375 (2013)

[30] Wesolowski, A., Buckee, C.O., Engø-Monsen, K., Metcalf, C.J.E.: Connecting mobility to infectious diseases: the promise and limits of mobile phone data. The Journal of infectious diseases **214**(suppl_4), 414–420 (2016)

[31] Peixoto, P.S., Marcondes, D.R., Peixoto, C.M., Queiroz, L., Gouveia, R., Delgado, A., Oliva, S.M.: Potential dissemination of epidemics based on Brazilian mobile geolocation data. part i: Population dynamics and future spreading of infection in the states of Sao Paulo and Rio de Janeiro during the pandemic of COVID-19. medRxiv (2020)

[32] Wesolowski, A., Qureshi, T., Boni, M.F., Sundsøy, P.R., Johansson, M.A., Rasheed, S.B., Engø-Monsen, K., Buckee, C.O.: Impact of human mobility on the emergence of dengue epidemics in Pakistan. Proceedings of the National Academy of Sciences **112**(38), 11887–11892 (2015)

[33] Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character **115**(772), 700–721 (1927)

[34] Hethcote, H.W.: Three basic epidemiological models. Applied mathematical ecology, 119–144 (1989)