Modeling the Spread of Measles

Nima Seyedtalebi, Joseph Kaninberg, Seifalla Moustafa
Department of Computer Science
University of Kentucky
Lexington, KY 40506–0633

Abstract—In order to ascertain behavioral information related to a particular disease, it is integral that the propagation of said disease can be accurately modeled mathematically. Typically, this is accomplished by using a wealth of historical data related to any given disease. It is often the case that no one model can accurately simulate all behaviors of a disease realistically. As such, this paper will motivate and discuss a model specifically designed to simulate the propagation of Measles within an interaction based network. In doing so, the authors hope to provide and demonstrate a model that can be used to exclusively simulate measles behavior with a high degree of accuracy. Our model is a variation of [SEIR] model. We introduce the interaction graph based method in which nodes represent individuals and edges represent interactions.The transition from each state in the [SEIR] to another i.e E to I or I to R is determined by thresholds of each node. The authors consider a number of terms such as secondary infection rate, infection and incubation periods and the basic reproduction number R_0. By comparing the results of our model to the historical results we found that it works well and can be considered as a reference.

I. INTRODUCTION

The ability to accurately model the propagation of a disease within a given population is a key component in discerning relevant attributes related to the disease being propagated. The generalization of disease propagation simulation to reality hinges upon the ability to accurately model the phenomena within a simulated setting. Typically, this is accomplished by making use of historical data in order to ascertain unique aspects related to the simulated disease. Using these unique aspects, simulations can be run in order to model the general propagation trends of the disease. In most cases, the quickness of propagation and varying infection rates due to personal contact are important factors. Due to the importance of human interaction in disease propagation, interaction modeling is given a reasonable amount of attention. Often, this personal contact factor is simulated in one of two ways. The first is a stochastic method, while the remaining is an interaction network based method. In the stochastic simulation method, spreading occurs in a manner that is purely probability based, so a large number of simulations are needed to produce consistent results. In an interaction graph based method, individuals are modeled as vertices and sufficient interactions as edges. In some cases, hybrid models are produced in which both interaction methods are taken into account and intertwined.

The modeling of epidemic spreading within a community is an extremely well researched area in modern scientific inquiry. With the advent of modern vaccinations, certain diseases have all but died out and become of little to no concern to the general populace. Due to the rise of the anti-vaccination sentiment however, many previously dying diseases are beginning to propagate among human populations once more. With this sentiment becoming increasingly prominent, it would be wise to examine historical data relating to these previous deadly diseases. In particular, diseases such as measles have been given a 'second-wind' so to speak. As a result, the need to accurately simulate these diseases in order to help mitigate the potential harm done by these infections has become increasingly important. As such, models used to simulate these diseases need to be developed in order to help minimize the potential risks.

For modeling the spread of disease, it is common practice to use a pre-established model that applies to general propagation phenomena. For disease propagation, however, this method can be reasonably unreliable due to the variance in individual disease behaviors. As such, it is often appropriate to develop individual models that can be used to accurately model the behavior and propagation phenomena demonstrated by a single disease. With this in mind, this paper seeks to model the behavior and spreading of measles using a model specifically designed to simulate these aspects of the infection. This paper seeks to demonstrate the accuracy and viability of the proposed model by comparing the results to historical behavioral data related to measles.

II. RELATED WORK

In general, the modeling and simulation of diseases propagation is often implemented using an already existing model as a base. The most simplistic model that demonstrates propagation capabilities is the Susceptible-Infected (SI) model. In this model, individuals are either unexposed to an infection or infected. Due to the simplicity of the model, it is often a poor means by which to model the propagation of actual diseases. An improvement is made by the SIR model, which adds a recovered compartment that is used to account for individuals that have either survived or perished from the disease. Some models make a distinction between recovery and death, but most rarely differentiate between them. Often, these types of models work well for information diffusion modeling [1] rather than disease propagation modeling. As the primary goal of this paper is the accurate modeling of measles, unfitting models for this topic will be disregarded. Often, models that demonstrate interesting behavior on complex networks behave in a fashion that facilitates better accuracy in modeling disease propagation [2], [3]. An even better approximation model

than the SIR model for use in modeling disease propagation is described by Hethcote [2]. In particular, he presents a Susceptible-Exposed-Infected-Recovered (SEIR) model and a number of variations of the model due to changing disease behavioral phenomena. The SEIR model described by Hethcote has been compared to real world infection data through simulations, demonstrating a reasonable degree of accuracy [4]. The model forwarded later by this paper is a variation of Hethcote's SEIR model, except that the proposed model does not have a dynamically changing population size similar to the one presented in Hethcote's model [2].

Many studies have been published examining or analyzing the general propagation of diseases within a specific real-world population [5]–[9]. Using these findings, various models have been proposed with the intent of accurately modeling properties of measles described by related studies. Most of these models are reasonably successful, but there are still a number of areas in which improvement is needed. For example, the initial infection rate of measles is rarely considered in most models. It is also often the case that the number of individuals within a compartment- susceptible, exposed, infected, and recovered-are fit to a graph for the model, but an accurate simulation of similar time scaling is not also taken into account. Specifically, sometimes the graphs may demonstrate a best fit phenomena, but the timescales of the two graphs differ. As such, most models trade off accuracy in one area for greater accuracy in a targeted area. This often produces great results in the targeted area, but then the neglected behaviors typically stray reasonably far from historical data. The model proposed within this paper hopes to rectify this trade-off discrepancy due to parameter priority.

In defining most diseases mathematically, a number of attributes are generally well defined that can be used to describe the disease's behavior. The first of these is the secondary infection rate, which is the percentage of people that will become infected due to contact with an infected individual. For measles in particular, this value is about 0.9, meaning an infected individual will infect 90% of those they interact with on a daily basis [5]. For most measles outbreaks, the peak of infected individuals is typically within the range of 90-120 days [5].Two additional important terms to define would the period of infection and incubation. The period of incubation is the length of time an infected individual carries a disease without it manifesting, which for measles is 10-12 days on average [5]. This is equivalent to being within the exposed compartment in the SEIR model. Accordingly, the time infected before an individual recovers is typically 5-6 days on average for measles [5]. The last important term that needs defining is R_0, which is the average number of people typically infected directly by the first infected individual [10]. For measles, this value is 12-18 individuals on average [5]. For the purpose of this paper, these parameters should suffice in simulating the disease.

III. PROBLEM STATEMENT

The rise of social media has given people new ways to communicate and share ideas. In theory, anyone with a Twitter account can reach millions of people using only a computer with an internet connection. It is clear that online social networks have effects outside the digital world, but the extent and nature of these effects is less certain. The spread of beliefs via social media that vaccines are unsafe or harmful has received attention in the news, so we chose to model the spread of misinformation related to vaccines and how it might affect the spread of diseases. Measles is well-studied and effectively preventable, thus it is amenable to the kind of analysis we propose. In this paper, we combine an information diffusion model with an epidemic model to explore the effects of fact-checking on vaccine hesitancy.

We are interested in the effects of information diffusion and fact-checking on the prevalence of an infectious disease. We cannot test this directly so instead we propose a model that combines the information diffusion/fact-checking model from [1] and a similar network-based SEIR model. In our combined model, we will consider both a social network and several different populations simultaneously. Imagine superimposing a social network on a (geographical) map so each node in the social network is associated with the place where that person lives. Each person in the network has a physical location and might interact with people outside of the social network. This confluence is what we aim to explore.

The information diffusion model is based on a well-known compartmental model called SIS (susceptible-infective-susceptible). Compartmental models divide the population into separate compartments or groups based on status. The dynamics of the disease process are captured by the movement of people between the compartments.[2] The model introduced in [1] subdivides the infective (I) compartment into fact-checkers (F) and believers (B). These processes are considered:

- Spreading ($S \rightarrow B$, $S \rightarrow F$): susceptible person believes hoax or becomes fact-checker
- Verifying ($B \rightarrow F$): Believer becomes fact-checker, rejects the hoax, and spreads fact-checking behavior instead of misinformation
- Forgetting ($B \rightarrow S$, $F \rightarrow S$): Believer or fact-checker forgets about the hoax

The focus of this paper is on describing and testing the epidemic model. In addition to compartments S, I and R, we add an E compartment to represent people exposed to the virus but not yet contagious (infective). The population of interest is modeled by an undirected network where each node represents a person and each edge represents daily contact with another person sufficient to transmit the disease. For each step of the simulation, nodes with infective neighbors have a chance of α to become "exposed", which means they have "caught" the disease but are not yet able to infect others. Transitions from E to I and from I to R are determined by thresholds set for each node. These thresholds represent the number of days spent in the incubation period (compartment E) and infectious

period (compartment I) and are set based on epidemiological data. They are randomly chosen from a Gaussian distribution to reflect variation between individuals. When the number of days in the current compartment reaches the threshold, the node transitions to the next compartment (either E-I or I-R). Additionally, each node in the I compartment has a chance each timestep to move to R regardless of the threshold. This represents death due to the disease and is described by the parameter ζ.

The dynamic behavior of epidemic processes can be described by certain threshold values. The most important of these is called the basic reproduction number R_0. It is the number of people a single infective person is expected to infect in a completely susceptible population. It is known that if $R_0 < 1$, the disease will not invade the population.[2] Higher R_0 values correspond to more virulent infections. Measles is one of the most contagious diseases recorded with R_0 values ranging from 12 to 18.[5]

IV. Proposed Model

We propose a model that combine the information diffusion model [SIR]by adding new compartment E(exposed) to represent people exposed to virus but not yet infective. It will be [SEIR] model

S: susceptible E: exposed I : infected R: recovered

We define transition between compartment by a threshold transition (TT) and has a value 0 or 1:

$$\text{Threshold transition}(TT^I) = \begin{cases} 1 & \text{if } \tau_i^I \geq X_I \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Threshold transition}(TT^R) = \begin{cases} 1 & \text{if } \tau_i^R \geq X_R \\ 0 & \text{otherwise} \end{cases}$$

where τ_i is the period of node i stayed in the compartment and X is the threshold period the node i should remain until transition occur.The following threshold values are chosen from Gaussian distribution:

- Threshold incubation period from E-I [Mean 11, variance 2] days
- Threshold infection period from I-R [Mean 6, variance 1] days

Death rate from I- R due to disease 0.2%

A. Formulation

$$\text{Threshold transition from E-I}(TT^I) = \begin{cases} 1 & \text{if } \tau_i^I \geq X_I \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Threshold transition from I-R}(TT^R) = \begin{cases} 1 & \text{if } \tau_i^R \geq X_R \\ 0 & \text{otherwise} \end{cases}$$

where i is the period of node i stayed in the compartment and X is the threshold period the node i should remain until transition occur

	E	I	R
E	$1 - TT^I$	TT^I	0
I	0	$1 - TT^R$	TT^R
R	0	0	1

TABLE I: Transition Matrix

V. Experiment

We performed two sets of computational experiments to evaluate our model. In our first set of experiments, we compared the basic reproduction number R_0 of our model disease with estimates of R_0 for measles. R_0 acts as a threshold in most infection models: if $R_0 > 1$ the infection spreads and if $R_0 < 1$ it does not. It can be estimated from epidemiological data or from a mathematical model. It should not be viewed as a constant value that is fixed for a given disease. Different mathematical models can even provide conflicting estimates, and it is often impossible to measure directly in practice.[11] It is useful for confirming that our model behaves similarly to other epidemiological models and can serve as a rough measure of the infectiousness of a disease.

Since we are simulating each member of the population, we can measure R_0 directly - we can simply record the number of people infected by a single contagious person. The value will be different for each simulation because of the randomness in the model, so we found the average between thousands of different instances of the problem. For each simulation we generated a new contact network using either the Erdos-Renyi, Barabasi-Albert, or Watts-Strogatz random graph models. Then, we chose one node at random to initiate the epidemic. This node begins the simulation in the I compartment and is assigned a random number of days to remain, similar to nodes that enter compartment I during the course of the simulation. From there the simulation proceeds according to the rules of the model. We only run each simulation for a single timestep because we are only interested in the number of people infected by the initial person. We tested the effect of changing parameters in the random graph generators on the resulting R_0 values by varying these parameters and recording the results.

In the second set of experiments, we compared predictions made by our model with data from a real epidemic. The real-world data was taken from a case study [8] of a 2000 measles outbreak in Dublin, Ireland. We transcribed the data from a graph showing the weekly number of new cases reported and used this as our standard of comparison. We ran simulations with the same three types of random network as before and varied the parameters in the same way. The only difference was that we ran these simulations until either everyone was recovered or there were no more infective individuals. We averaged the results of 100 simulations for each combination of parameters we tested and compared the number of infectives each week to the case study data[1]. We used infective people instead of those in the incubation period because people only

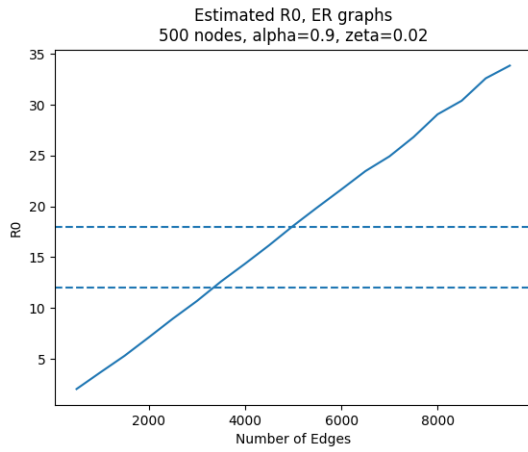---

[1]The case study data was reported in weeks

Fig. 1: Estimated R_0 values with an Erdos-Renyi contact network. Horizontal lines mark R_0 values in the range given for measles in the literature (12-18)

Fig. 2: Estimated R_0 values with a Barabasi-Albert contact network.

report the disease once they have symptoms, and with measles people are often contagious before the onset of symptoms.[5] Using this data we determined which random graph model and parameters produced the best fit for the case study data by computing the mean squared error (MSE) and choosing the parameters that minimized the MSE.

VI. RESULTS

Our R_0 values followed the pattern we expected. In our model, R_0 is determined by the mean degree of the contact network. This follows from the definition of R_0 as the average number of people infected by a single individual in a totally susceptible population. Since the contact network describes all of the sufficient contacts for each individual, a higher mean degree means more contacts on average. Note that the network topology does not affect the R_0 estimate - this is expected because for $R_0 > 1$, there is no difference in behavior between homogeneous and heterogeneous network topologies.[3] That is why figures 1, 2, and 3 look similar. Figure 4 is different because instead of varying the mean degree, we varied the rewiring probability in the Watts-Strogatz graph generator. Based on the data in figure 4, we concluded that the rewiring probability has little effect on R_0. All of the R_0 estimates were produced by averaging the results of runs on 1000 different random graphs.

We also evaluated our model against outbreak data taken from a case study.[8] We used mean squared error (MSE) to compare the predicted number of infective patients during each week with the actual number of new cases reported each week in the Dublin data. First, we tried varying the mean degree of the contact network for each of the topologies to see which topology fit the historical data the best. The outcome of these experiments is shown in figures 5, 6, and 7. From these, we determined that a lower mean degree than we had previously examined would produce the best results. We chose the Erdos-Renyi network for further optimization attempts.
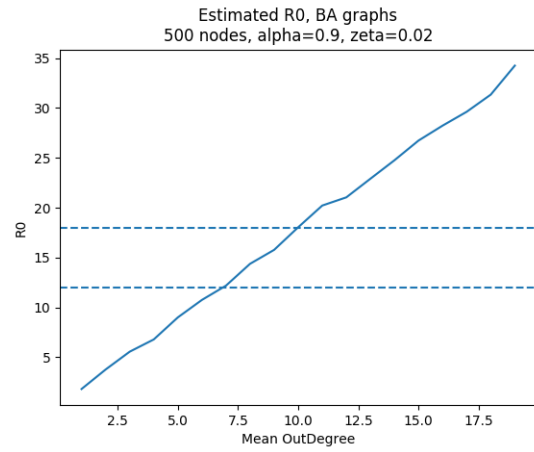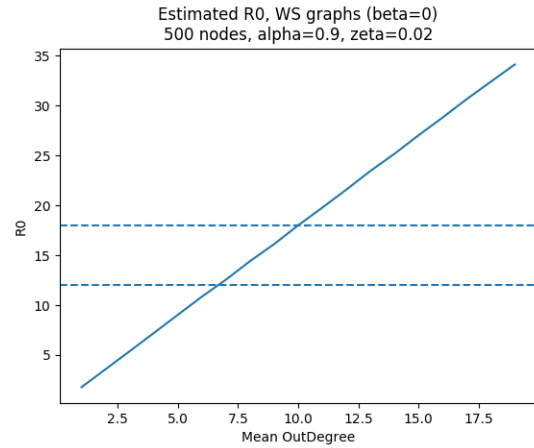

Fig. 3: Estimated R_0 values with a Watts-Strogatz contact network.

Using an ER model made it simple to generate graphs with a lower mean degree - we simply used fewer edges in the model. We again used the MSE to measure loss and varied the number of edges in the random graph model over a wide range and noted where the MSE reach a minimum. In this manner, we found that a $G(n, M)$ ER network with n nodes and $n\frac{22}{20}$ edges provides the best fit between our model and the historical data with a MSE of 831.164. This is comparable to the performance of some of the models in [4] Using these parameters, we ran a simulation for each 10000 different random graphs and averaged the results to produce figure 8.

On average, there were thirty-seven people still susceptible at the end of the epidemic. This was the first example of endemic behavior we saw in our data. A contact network with more than a single connected component would prevent the infection from reaching everyone in the population. This is actually more realistic than a contact network with a single connected component because we know that measles survives despite being so virulent. If everyone in the world were in the
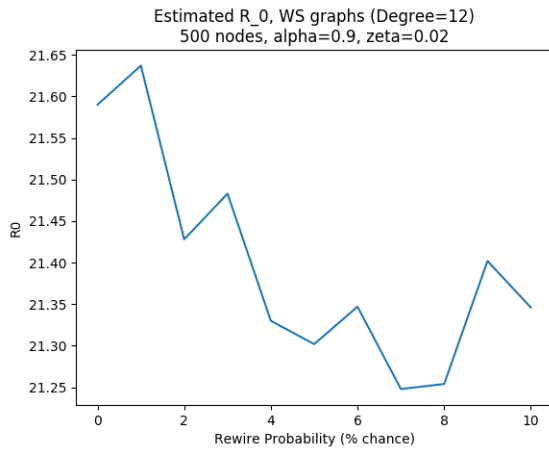
Fig. 4: Estimated R_0 values with respect to rewiring probability parameter for Watts-Strogatz contact network
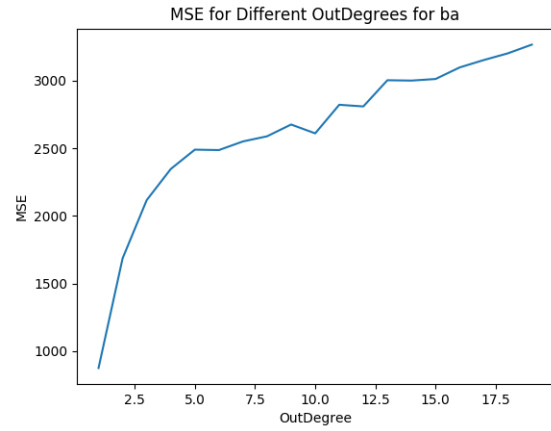

Fig. 6: MSE between model and historical data for different outdegrees, Barabasi-Albert contact network
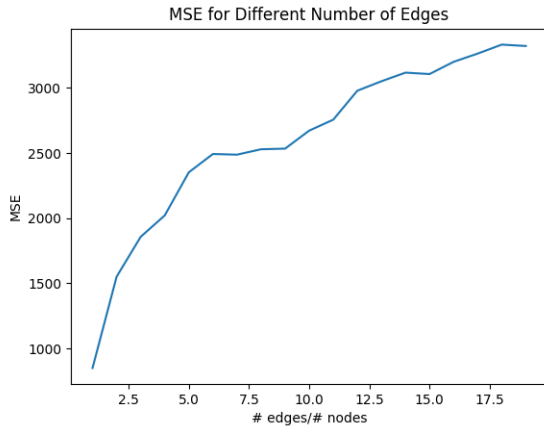

Fig. 5: MSE between model and historical data for different numbers of edges, Erdos-Renyi contact network
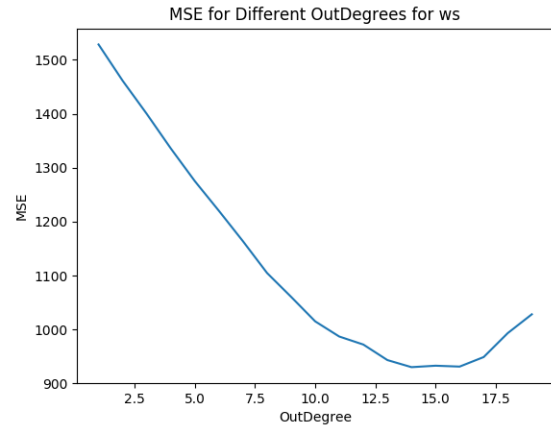

Fig. 7: MSE between model and historical data for different outdegrees, Watts-Strogatz contact network

same connected component of the contact network, measles could not survive in its current form because people become immune after infection.

The lower R_0 for the real-world data could potentially be explained by immunization. A fully immune person will never become infected, so they are effectively removed from the contact graph. Under the conditions of *herd immunity*, there are enough immunized people that the disease does not survive in the population. Under these conditions the contact graph is broken into many different connected components so the disease cannot propagate.[5] According to the case study, immunization rates in the affected area were around 66 percent[2], so a large portion of the population was not susceptible to the disease. We are ultimately interested in studying the effects of social networks on these contact networks and resulting epidemic processes.

[2] The national average was 76% at the time

VII. FUTURE WORK

For future consideration, the authors plan to improve the time-scaling accuracy of the model to a great degree. The current model demonstrates accurate modeling for the R0 phenomena and best-fit graphs, but the time scaling is not a factor that we heavily examine. For associated real-world data,[7]–[9] the peak of infection typically lies within the 90-120 day range due to the average incubation period of measles. In most of our initial simulations where R0 accuracy was given priority though, the average range was noticeably smaller than 90 days. Initially we suspected this could be due to the random graphs that we used to validate our simulations. This lead us to examine the possibility that this inaccuracy was due to the general size of the associated graphs. Increasing the population sizes made a significant difference in modeling for the WS network, but the simulated ER and BA interaction networks demonstrated similar total steps independent of the population size. As a result, different considerations will need
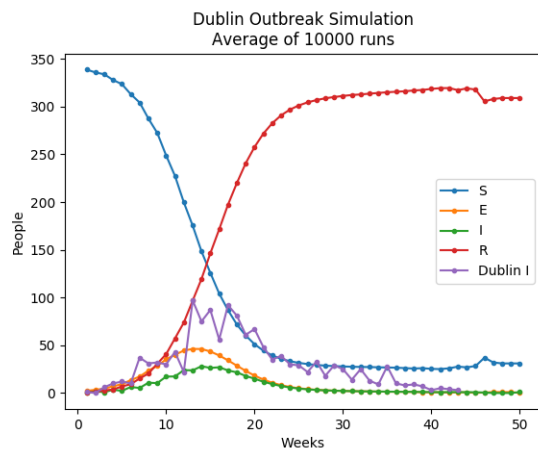
Fig. 8: Best model predictions with historical data

to be made so that the model will scale well to real-world data better for total timescale. Accordingly, the model will need to be properly tested against similar real-world outbreaks further in order to continue validating the demonstrated results. So while the model preforms accurately when compared to the expected historical r0 value, improvements must be made in order to demonstrate that the model accurately generalizes when modeling all of measles' real-world behavioral facets.

VIII. CONCLUSION

In this work we propose the SEIR model to describe the propagation of Measles within an interaction based network We estimate R_0 by measuring the mean degree of the contact network. We found that the lower mean degree values give the best results by using ER model to generate graphs.We consider the transition between compartments depends only on the threshold period the node should remain in that compartment.The threshold transition(TT) has a value 0 or 1. We ran the simulation for 10000 different graphs to produce the best fit figure for our SEIR model. Analysing our best fit figure we confirm that our model behave similarly to other epidemicological models and can serve to measure the propagation of the disease.

REFERENCES

[1] M. Tambuscio, G. Ruffo, A. Flammini, and F. Menczer, "Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks," in *WWW*, 2015.

[2] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, pp. 599–653, 2000.

[3] R. Pastor-Satorras, C. Castellano, P. V. Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Rev. Mod. Phys. 87, 925 (2015)*, Aug. 12, 2014. DOI: 10.1103/RevModPhys.87.925. arXiv: 1408.2701v2 [physics.soc-ph].

[4] J. R. Montalan, M. R. J. E. Estuar, K. Teknomo, and R. W. Gardon, "Measles metapopulation modeling using ideal flow of transportation networks," in *ICSIM 2019*, 2019.

[5] C. for Disease Control, *Epidemiology and Prevention of Vaccine-Preventable Disease*, ch. 13.

[6] *Distribution of measles cases by country and by month, 2011-2019*, https://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/Table_407_Web_Number_Measles_Cases_Month_2019-03-07.xlsx?ua=1.

[7] J. Jasem, K. Marof, A. Nawar, and K. M. Islam, "Epidemiological analysis of measles and evaluation of measles surveillance system performance in iraq, 2005–2010," *International Journal of Infectious Diseases*, e166–71, 2012.

[8] J. Mcbrien, J. Murphy, D. Gill, M. Cronin, C. O'Donovan, and M. T. Cafferkey, "Measles outbreak in dublin, 2000," *Pediatr Infect Dis Journal*, vol. 22, pp. 580–4, 2003.

[9] D. E. Sugerman, A. E. Barskey, M. G. Delea, I. R. Ortega-Sanchez, D. Bi, K. J. Ralston, P. A. Rota, K. Waters-Montijo, and C. W. LeBaron, "Measles outbreak in a highly vaccinated population, san diego, 2008: Role of the intentionally undervaccinated," *Pediatrics*, vol. 125, pp. 580–4, 2010.

[10] J. H. Jones, *Notes on r_0*, Stanford University, 2007.

[11] J. Li, D. Blakeley, and R. J. Smith, "The failure of r0," in *Comp. Math. Methods in Medicine*, 2011.