Studies of Rumor Spreading and Control using SIR models

Jerry Lingjie Mei*
Jingjing Tang†

May 18, 2018

Abstract

We are interested in studying social media dynamics using SIR model on a network. We introduced competitive SIR model with immunization and study the immunization threshold for the dynamics of two rumors spreading on a network. At the same time, we study the coupling strength of SIR metapopulation model with different population sizes and explore the optimal inoculation method for the rumor control.

1 Introduction

Online social networks have become one of the most important media of the information propagation among communities. Our study is triggered by two different social media phenomenon. The first one is that political and social rumors are rampant on social networks like Facebook and Twitter. Two rumors may represent entirely different social values or political standpoints, so it is worth telling which one of these two can be accepted by more users on the social media. On the other hand, social media platforms with different key features have different market competitiveness and participants. Although there exists the impact of multiple forms of communication adoption, we can assume that one of the social media platforms serve as an individual's main source of news and the main way of information spreading when receiving a certain type of news.

To explore the first phenomenon, we introduce the competitive SIR model with immunization which is a deviation from the standard SIR model. The model features two epidemics spreading on the network (red

*Class of 2020, Course 18 and 6-3, Massachusetts Institute of Technology, jerrymei@mit.edu

†SM60 Health Data Science program, T.H.Chan School of Public Health, Harvard, jtang@hsph.harvard.edu

and blue epidemics). Initially, both epidemics have exactly one random node infected, while other nodes stay susceptible to both epidemics. Each epidemic has a time step and transmissibility, and an infected node will infect its susceptible neighbor through an edge in such time step with given transmissibility, and then recovered from its epidemics. The two epidemics are considered competitive, in which once a node is infected or recovered from one epidemic, it will no longer be infected by the other epidemic. This model can also be equipped with immunization features, where a proportion of nodes are immune to both epidemics.

In real world scenario, this model can reflect that the two rumors have a noticeable difference in the rate of updating and levels of persuasion, creating a very complex dynamics. Immunization of users means they are neutral in such social or political debate or weary of spreading any news at all so that this node cannot be infected by either of the epidemics.

As for the second phenomenon, since the information transmission is predominantly a localized process, we focus on an SIR metapopulation model, which is a type of spatial model and is applicable to modeling many human diseases and also rumor spreading. In terms of rumor spreading through multiple social media platforms, the entire population is divided into distinct subpopulations. Each subpopulation has independent epidemiological dynamics together with limited interaction between the subpopulations. It is natural to think about commuters whose main active social platform is i for the most of time but temporarily changes to j. Besides, the infectious rates for different social media platforms are different and often have a positive relation to the proportion of time that a commuter spends away on another social media platform. Intuitively, we consider metapopulation models with multiple population sizes since the number of participants on different social media is very likely to be different. Starting from this assumption, we provide more complicated estimates for the 'coupling strengths' among populations to detect the impact of correlations.

Besides, we introduce targeted inoculation as a method to control the rumor spreading among and within the populations. Different from random inoculation, which performs well in homogeneous models, targeted inoculation is more suitable for our model to efficiently control rumor spreading in a more sensitive population since we have an approximately assortative transmission matrix.

2 Previous Work

The competitive SIR model without immunization havs been studied by Newman [1] as well as Karrer and Newman [4]. Let the time step of the two epidemics to infect the other one is t_r and t_b. The transmissibility

of two epidemics are T_r and T_b respectively, representing the probability that one epidemic will pass through a given edge. In Newman's work [1], he used a series of threshold measurements to determine the predominant epidemics

- Let the basic threshold be $\varphi_c = \frac{\mathbb{E}[D]}{\mathbb{E}[D^2]-\mathbb{E}[D]}$, then T_r and T_b has to be greater than ϕ_c for them to spread to a giant component.
- If an epidemic didn't go extinct, it will grow exponentially, and the ratio of exponential growth rate between the two epidemics is $\beta = \frac{\ln T_r/\phi_c}{\alpha \ln T_b/\phi_c}$. If $\beta > 1$, then the red epidemic will become the predominant epidemic on the network. On the other hand, if $\beta < 1$, then the blue epidemic will become the predominant epidemic on the network.
- There's also a possibility that the residual network after the first epidemic spread through is enough for the second epidemic to reach a giant component. This happens only when $\beta > 1$ and $T_r < \phi_x = \frac{1-u}{1-F(u)}$, where $F'(u) = \frac{1}{T_b}$ and the generating function for the excessive degree in the network is F.

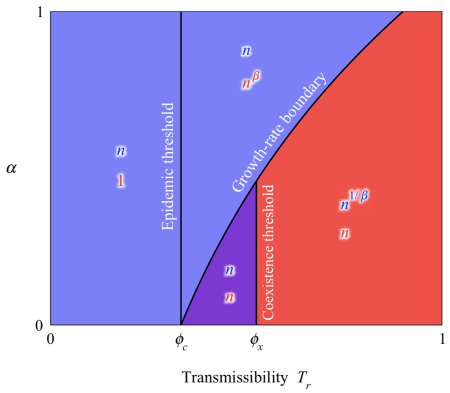


Figure 1: Newman's results[4]: The predominant epidemics under a specific T_b.

In terms of SIR metapopulation model, Keeling [2] developed a mathematical framework to approximate the behavior of stochastic metapopulation models and highlighted the effects of coupling. A variety of coupling forms have been constructed. Keeling and Rohani[3] showed that for two populations of equal size and equal epidemiological characteristics with the impact of commuters, the analytical relationship between the phenomenological and the mechanistic models is $\sigma = \frac{2\mu}{(1+\mu)^2}$ and the estimate of σ could be obtained by the correlation $C \approx \frac{\sigma}{\xi+\sigma}$ where ξ is a function of the particular disease parameters but does not depend upon population size.

3 Competitive SIR Model with Immunization

3.1 Model settings

We have the following model setting: Without loss of generality, we assume that the red epidemic has a higher rate of updating, or its time step is shorter than the blue epidemic. We denote α as the ratio between the time step of two epidemics: $\alpha = \frac{t_r}{t_b}$. For the immunization part, a node is immunized with probability $1 - \varphi$ independently of other nodes.

We also want to favor one rumor over another rumor: in the real world scenario, the rumor that needs to be supported is usually spreading with a low rate of update. Our question now turns into how much immunization in the network we need to mitigate the update rate advantage of the red rumor so that the blue rumor can become the predominant one.

3.2 Analysis

Given a graph with empirical degree distribution $p_k = \mathbb{P}[D = k]$, the excessive degree distribution is $q_k = \frac{(k+1)p_k}{\mathbb{E}[D]}$. Then the basic reproduction number for the red epidemic is

$$\begin{aligned} R_r &= T_r \sum_{k=0}^{\infty} \varphi k q_k = T_r \sum_{k=0}^{\infty} \varphi k(k-1)p_k \\ &= \varphi T_r \frac{\mathbb{E}[D^2] - \mathbb{E}[D]}{\mathbb{E}[D]} \end{aligned}$$

Let $\phi_c = \frac{\mathbb{E}[D]}{\mathbb{E}[D^2]-\mathbb{E}[D]}$, then $R_r = \varphi T_r/\phi_c$. Without the influence of the blue epidemic, the red epidemic spreads to a giant component of the nodes if $R_r > 1$, or $\varphi T_r > \phi_c$, and the same result applies for the blue epidemic.

What basic reproduction number means is every infected node will then infect another R_r number of nodes through its edges. As the time step of the red epidemics is t_a, as time t, the red epidemic have infected

$$N_r = O(R_r^{\frac{t}{t_r}}) = O\left(\left(\frac{\varphi T_r}{\phi_c}\right)^{\frac{t}{t_r}}\right)$$

And the time is will take the red epidemic to infect $O(n)$ of the population is

$$t_r \ln n / \ln \frac{\varphi T_r}{\phi_c}$$

The similar result holds for the blue epidemic as well. As shown in Newman's work [1] previously, as the blue epidemic has a longer time step, the blue epidemic is predominant iff it takes the blue epidemic to reach $O(n)$ nodes, that is to say:

$$t_r \ln n / \ln \frac{\varphi T_r}{\phi_c} > t_b \ln n / \ln \frac{\varphi T_b}{\phi_c}$$

The threshold that the inequality above holds true is when

$$\varphi_i = \phi_c \left(\frac{T_b^\alpha}{T_r} \right)^{\frac{1}{1-\alpha}}$$

We want to make sure that this immunization threshold is never too high to stay above 1, or too slow so that the blue epidemic will not spread to a giant component. We will split it into several cases:

a) If $\varphi_i > 1$, or $T_b^\alpha \phi_c^{1-\alpha} > T_r$, then there is no such threshold. The blue epidemic always stays predominant, and no immunization is required.

b) If $\varphi_i T_b < \phi_c$, or $T_b < T_r$, the transmission rate of the blue epidemic is not as large as the red one, then any kind of immunization will not help blue epidemic become predominant.

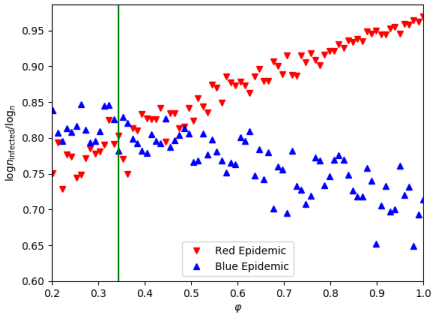
c) In the most typical cases where φ_i is a valid number, then if $\varphi < \varphi_i$ (extensive immunization), the blue epidemic will be predominant; if if $\varphi > \varphi_i$ (limited immunization), the red epidemic will be predominant.

3.3 Numerical results

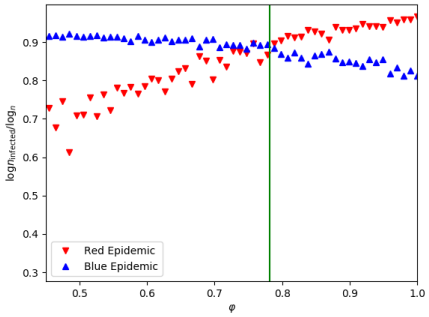
All the theoretical deductions above are based on the configuration model that generates the network until infinite size. To investigate how this model would perform in a finite size network, we ran some numerical tests on several simulated networks. For simulations we use three different graph generating methods with around the same number of nodes but different numbers of edges:

- A power-law graph generator with 7000 nodes and 90000 edges, following an empirical degree distribution of exponent 3.0.
- A small-world graph generator using Watts–Strogatz model on a 1 dimension lattice consisting of 7000 nodes. Every node on the lattice is connected to $k = 13$ of its neighbors on the left and on the right, with a rewiring probability of .25.

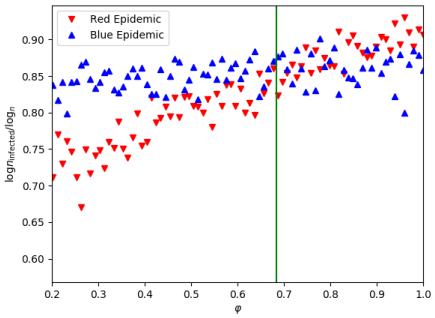
The size of the graph is designed to match the real-world network below. For each graph generator, we each choose three different parameter settings such that the threshold immunization rate is neither too small nor too high. We only collect the samples where both epidemics don't extinct. We run at least 50 iterations from each data point that is shown in the figure below:



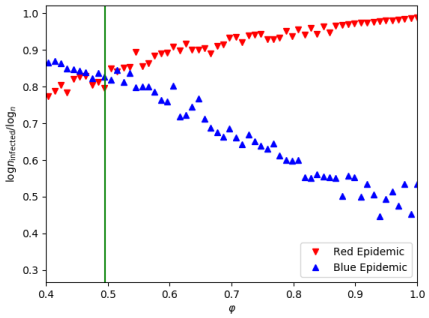
(a) Power law model network with $T_r = .1, T_b = .2, \alpha = .5$



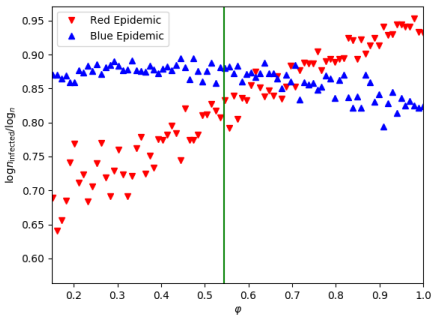
(b) Small-world model network with $T_r = .1, T_b = .2, \alpha = .5$



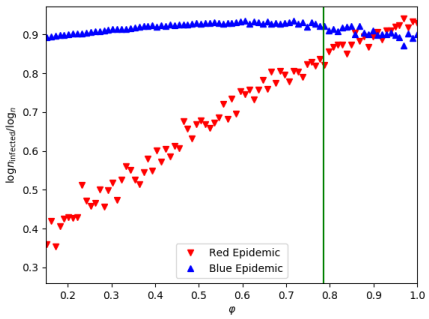
(c) Power law model network with $T_r = .1, T_b = .2, \alpha = .67$



(d) Small-world model network with $T_r = .1, T_b = .2, \alpha = .25$



(e) Power law model network with $T_r = .1, T_b = .3, \alpha = .5$



(f) Power law model network with $T_r = .1, T_b = .25, \alpha = .5$

The green line represents the immunization threshold we predicted, while red and blue triangles indicate the number of nodes ever recovered from two epidemics in a log scale. As you can see from the figures, our estimation of immunization threshold matches pretty well with the actual immunization threshold where the blue epidemic take dominance.

We also run some simulation on real-world data. Facebook has a large collection of pages that act as a portal for delivering messages and rumors. We use the network for Facebook pages in the "Government" category, with each node representing a page while links are likes among pages[5]. The network has a size of 7057 and 89455 edges.

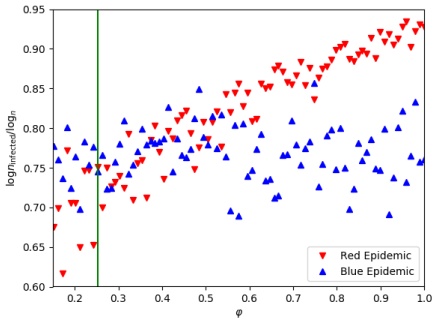


Figure 3: Government network with $T_r = .1, T_b = .2, \alpha = .5$

It turns out the estimation is still quite accurate as well even on some real-world networks

4 SIR Metapopulation Model with Inoculation

4.1 Model Settings and Analysis

The force of infection helps achieve the effect of a spatial structure when a number of populations are coupled together. For simplicity, our study only focuses on a system of two populations with different population sizes and infectious rates but the same recovery rates. We also assume that the birth and death rate which represents the introduction of new participants and their permanent leaving are equal($b = d$) so that social media environment is relatively stable. The spatial structure set of equations are given by:

$$\begin{aligned}
\frac{dS_1}{dt} &= bN_1 - \frac{\beta_{11}S_1I_1(1-\sigma) - \beta_{12}S_1I_2\sigma}{N_1} - dS_1 \\
\frac{dI_1}{dt} &= \frac{\beta_{11}S_1I_1(1-\sigma) - \beta_{12}S_1I_2\sigma}{N_1} - \gamma I_1 - dI_1 \\
\frac{dS_2}{dt} &= bN_2 - \frac{\beta_{22}S_2I_2(1-\sigma) - \beta_{21}S_2I_1\sigma}{N_2} - dS_2 \\
\frac{dI_2}{dt} &= \frac{\beta_{22}S_2I_2(1-\sigma) - \beta_{21}S_2I_1\sigma}{N_2} - \gamma I_2 - dI_2
\end{aligned} \tag{1}$$

where σ represents the coupling strength and is related to the movement rate between two social media platforms. Note that σ merely reflects how

much mixing exists between the two populations. Besides, β_{ij} are the infectious rates between two populations.

Notice that the movement patterns are determined by the population and are independent of the disease. Taking the movement of commuters into account, the equation above could be rewritten as following:

$$\frac{dS_{11}}{dt} = bN_{11} - \frac{\beta_{11}S_{11}(I_{11}+I_{21})}{N_{11}+N_{21}} - dS_{11} + rS_{12} - lS_{11}$$

where r is the rate at which individuals return their home platform, l is the rate at which individuals leave their home platform. However, we are more concerned about the ratio of time spent in the temporary to the main active platform which can be denoted as $\mu = \frac{l}{r}$. Since we now have two populations of different sizes, it is necessary to consider that the coupling strength is not a constant but related to the population sizes. This also leads to a sensible assumption that μ should be dependent on N_1 and N_2, that is $\mu_1 = \alpha\frac{N_2}{N_1+N_2}$ where α is a constant ratio.(The ratio of time spent in the current location should have a positive relation to the current population size) So that, the coupling terms which are defined as $\sigma = \frac{2\mu}{(1+\mu)^2}$ would converge to the asymptotic values:

$$\sigma_{11} \to \sigma_{21} = \frac{N_2}{(1+\mu_1)(1+\mu_2)}(\frac{\mu_2}{N_{11}+N_{21}} + \frac{\mu_1}{N_{12}+N_{22}})$$
$$\sigma_{12} \to \sigma_{22} = \frac{N_1}{(1+\mu_1)(1+\mu_2)}(\frac{\mu_2}{N_{11}+N_{21}} + \frac{\mu_1}{N_{12}+N_{22}})$$

Therefore, equation (1) should be rewritten as:

$$\begin{aligned}
\frac{dS_1}{dt} &= bN_1 - \frac{\beta_{11}S_1I_1(1-\sigma_{11}) - \beta_{12}S_1I_2\sigma_{12}}{N_1} - dS_1 \\
\frac{dI_1}{dt} &= \frac{\beta_{11}S_1I_1(1-\sigma_{11}) - \beta_{12}S_1I_2\sigma_{12}}{N_1} - \gamma I_1 - dI_1 \\
\frac{dS_2}{dt} &= bN_2 - \frac{\beta_{22}S_2I_2(1-\sigma_{22}) - \beta_{21}S_2I_1\sigma_{21}}{N_2} - dS_2 \\
\frac{dI_2}{dt} &= \frac{\beta_{22}S_2I_2(1-\sigma_{22}) - \beta_{21}S_2I_1\sigma_{21}}{N_2} - \gamma I_2 - dI_2
\end{aligned} \tag{2}$$

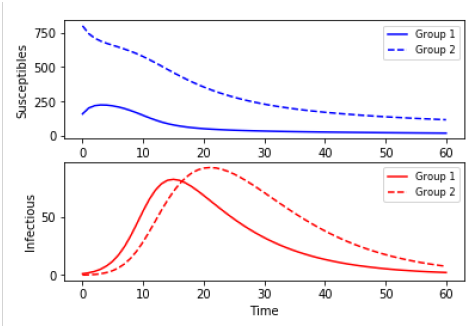
Theoretically, the records provided by social media applications are able to allow us to estimate the value of μ, r and therefore the corresponding valueσ. However, in reality, it is difficult to assess the level of movement between different platforms. The records provided by social media applications are not stratified sufficiently for us to obtain the potential mixing.

4.1.1 Correlations for Stochastic Model

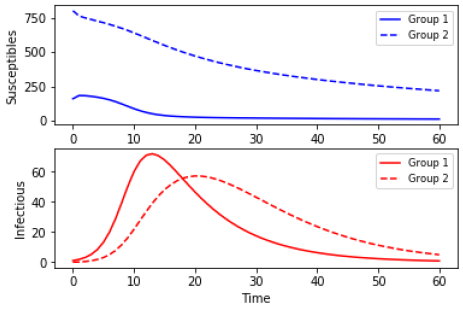
Using a stochastic model, where all events occur randomly, we assume that the distribution of population size is captured by the first few moments(means and variances, etc.) while the higher order terms only have

little influence on its statistical properties. This allows us to formulate the equations for the co-variance $\hat{C}$ using moment closure approximation[2]:

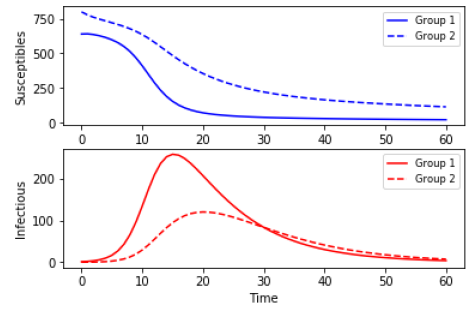
$$\begin{aligned}
\frac{dC_{I_2I_1}}{dt} &= 2 < \frac{\beta_{11}I_2S_1I_1(1-\sigma_{11}) - \beta_{12}I_2S_1I_2\sigma_{12}}{N_1} - \gamma I_2I_1 - dI_2I_1 > \\
&\quad - 2I_2 < \frac{\beta_{11}S_1I_1(1-\sigma_{11}) - \beta_{12}S_1I_2\sigma_{12}}{N_1} - \gamma I_1 - dI_1 > \\
&\quad = 2\frac{\beta_{11}(1-\sigma_{11})S_1C_{I_2I_1}}{N_1} + 2\frac{\beta_{12}\sigma_{12}S_1C_{I_2I_2}}{N_1} - 2\gamma C_{I_2I_1} - 2dC_{I_2I_1} + 2\frac{\beta_{12}I_2C_{S_1I_2}}{N_1} \\
\frac{dC_{I_1I_1}}{dt} &= 2 < \frac{\beta_{11}I_1S_1I_1(1-\sigma_{11}) - \beta_{12}I_1S_1I_2\sigma_{12}}{N_1} - \gamma I_1I_1 - dI_1I_1 > \\
&\quad - 2I_1 < \frac{\beta_{11}S_1I_1(1-\sigma_{11}) - \beta_{12}S_1I_2\sigma_{12}}{N_1} - \gamma I_1 - dI_1 > \\
&\quad = 2\frac{\beta_{11}(1-\sigma_{11})S_1C_{I_1I_1}}{N_1} + 2\frac{\beta_{12}\sigma_{12}S_1C_{I_1I_2}}{N_1} - 2\gamma C_{I_1I_1} - 2dC_{I_1I_1} + 2\frac{\beta_{12}I_1C_{S_1I_2}}{N_1}
\end{aligned}$$



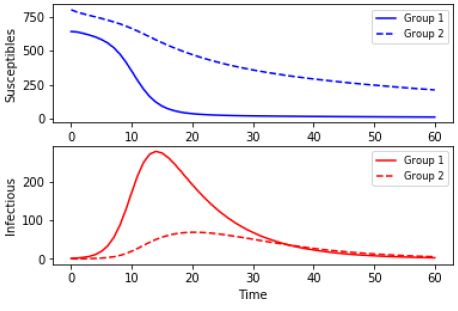
(a) Metapopulation Model with $N_1 = 200, N_2 = 1000, \mu = \frac{1}{6}$



(b) Metapopulation Model with $N_1 = 200, N_2 = 1000, \mu = \frac{1}{20}$



(c) Metapopulation Model with $N_1 = 800, N_2 = 1000, \mu = \frac{1}{6}$



(d) Metapopulation Model with $N_1 = 800, N_2 = 1000, \mu = \frac{1}{20}$

Here $< \cdot >$ corresponds to the long-term average. Symmetrically thinking, $\frac{dC_{I_1I_1}}{dt}$ and $\frac{dC_{I_2I_2}}{dt}$ are of similar forms. It is difficult to simplify and produce any rigorous results to this complicated problem but it can provide general insights together with the σ in a matrix structure. There

should be a symmetric structure of N_1 and N_2 which lead to further consideration of the magnitude of the two population size. Therefore, if the population sizes are of the same magnitude, the model is of similar form as that of a system with two populations with the same sizes. However, if the magnitude of the population sizes is vastly different, the correlations would not be approximately well enough.

4.1.2 Targeted Inoculation

Using targeted inoculation, the threshold which is calculated as $1 - \frac{1}{R_0}$ for a random inoculation could be further improved due to the structural nature of two populations with different epidemic dynamics. As mentioned above, it is sensible to consider that the infectious matrix β consists of two components: exposure and transmission potential. Intuitively, facing a certain type of news, a social media platform with greater exposure rate has greater transmission rate. Unfortunately, a rigorous solution to calculate the eigenvalues of the dynamic is impossible since the problem is very complicated. Instead, a numerical approach is provided with general insights. Notice that since transmission occurs frequently within a class but there is extremely limited transmission between class. We can consider the upper bound of the entire transmission matrix β^* which combine the impact of transmission within and among subpopulations, which means that for all $i, j = 1, 2, \beta^*_{ij} \geq \beta_{ij}$ and β^* is given by:

$$\beta^*_{ij} = \begin{cases} \beta_{ii} \geq 0 & \text{if i = j} \\ \delta & \text{otherwise} \end{cases}$$

where δ is a constant sufficiently near 0^+. Remember that the movement patterns of individuals are determined by the population, it is reasonable to assume that the total amount of movement is linearly related to the population size. However, after developing the equations based on proportion instead of individual behavior the impact of population size is ignorant. If the transmission rate among subpopulations is sufficiently low, we could approximate the dynamics without taking it into consideration. The inoculation control method is dominated by the change of transmission rate within subpopulations. The rumor spreading in one the population with higher transmission rate (and also higher risk) will be eradicated $p_1 \geq 1 - \frac{\gamma}{\beta^*_{11}}$ before spreading to another population. Therefore, it is optimal to inoculate each population at the eradication threshold for effective control effect.

5 Conclusion

By using the competitive SIR model with immunization, we have theoretically calculated the immunization threshold for a slower-updating epidemic to become the dominant epidemic, depending on the relations between time steps and transmission rates. The simulation result is accurate to the extent that it indicates what levels of immunization is needed in artificial networks and real-world networks. This turns out to be invaluable in studying the dynamics of rumors in a complex network.

In our study of the SIR metapopulation model with different population sizes, although relationship between the phenomenological coupling and the mechanistic movement rates cannot be numerically calculated, there exist asymptotic analytic relationship which is population size dependent. Equipped with detailed records in different social media platforms, stochastic simulations are used to calculate the correlation which contributes to the estimate of coupling σ. Based on this complicated metapopulation model, if the transmission rate among populations are ignorant, the most efficient control method is to inoculate at the eradication threshold for each population.

6 Future Work

For the competitive SIR model with immunization, it is simulated with infections happening only at the end of each time step. We will also need to try a slightly time-varying updating mechanism. In the spreading of a rumor, it updates slowly in the very first few hours, faster in the middle of a time step, and slowly again in the end. Two possible ways to simulate this time-varying infection process is to give a time-varying infectious rate, like a sinuous function with proper normalization, or just three infection rate that represents three stages of an infected node. The design should follow the constraint that the expected time that a node is infected through a particular edge (given the infection happens) is T_r or T_b.

For the numerical simulation, we have also tested on other graphs like Erdös-Rényi with reasonable parameters. It turns out our immunization threshold doesn't hold with such graph generating models, especially when the network size is sufficiently small. For further studies, we may also consider modeling competitive SIR processes without using the configuration model so that it matches more with the limited size and diameter of these case of networks. One potentially good method would be using ODEs to simulate the same process.

We may also study the immunization threshold for two epidemics to coexist, given that any immunization will not make the blue epidemic the predominant epidemic. In theory, this immunization threshold can be calculated by solving the equation $\frac{(1-u)F'(u)}{1-F(u)} = \frac{T_b}{T_r}$, then plugging u in $\varphi = \frac{1}{F'(u)T_b}$. However, the simulation such relations should be done on a

larger scale, possibly 10 times more in the size of the network and the iteration for each data point. Given enough computational power, we might verify this weaker threshold for immunization.

For the second phenomenon, we also notice that the delay between different epidemics when exploring the impact of interactions(levels of coupling) among populations have a strong relation to the population sizes as well as the movement mechanism. It could be an opportunity for advertisement. Individual-based model using ReedFrost to simulate may be a possible method to explore the inoculation which intentionally adds edges between certain pairs of nodes to minimize the time delay while restricting the amount of inoculation.

Additionally, the two phenomena exist simultaneously which means it should be more practical to combine the two types of model. The new competitive metapopulation SIR model is complicated with not only two epidemics but also multiple subpopulations with different transmission rate within and among the populations. The same, it is more reasonable to use an individual model like ReedFrost to simulate the dynamic so that for each individual, the change current social media platform possibly help to be exposed to another epidemic with a different infectious rate.

References

[1] B. Karrer and M. E. J. Newman. "Competing epidemics on complex networks". In: *pre* 84.3, 036106 (Sept. 2011), p. 036106. DOI: 10.1103/PhysRevE.84.036106. arXiv: 1105.3424 [physics.soc-ph].

[2] Matt J Keeling. "Metapopulation moments: coupling, stochasticity and persistence". In: *Journal of Animal Ecology* 69.5 (2000), pp. 725–736.

[3] Matt J Keeling and Pejman Rohani. "Estimating spatial coupling in epidemiological systems: a mechanistic approach". In: *Ecology Letters* 5.1 (2002), pp. 20–29.

[4] M. E. J. Newman. "Threshold Effects for Two Pathogens Spreading on a Network". In: *Physical Review Letters* 95.10, 108701 (Sept. 2005), p. 108701. DOI: 10.1103/PhysRevLett.95.108701. eprint: cond-mat/0502373.

[5] Benedek Rozemberczki et al. *GEMSEC: Graph Embedding with Self Clustering*. 2018. eprint: arXiv:1802.03997.