VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
SOFTWARE ENGINEERING STUDY PROGRAM

Epidemic Spread in Networks

Epidemijos plitimas tinkle

Master Thesis

Master student: Agnė Karbonskienė
Supervisor: prof. habil. dr. Mindaugas Bloznelis
Reviewer: dr. Tomas Plankis

Vilnius – 2024

Abstract

Most epidemic models with networks try to capture either spatial or social connections by embedding them into a graph. In this work, we show that these approaches can be combined into a single model. A stochastic simulation-based model concept was proposed by the supervisor prof. habil. dr. Mindaugas Bloznelis. The model encompasses actors moving in a spatial network called location graph. Furthermore, model also includes a social network, which determines an increased disease transmission chance between connected actors. Both networks are realized as random graphs. In addition, epidemic control measures are examined. Finally, we examine the possibility to apply this model to Covid-19 data in selected Lithuanian municipalities.

Key words: two-network epidemic model, SIR, random graph

Santrauka

Viena populiariausių epideminio modeliavimo strategijų yra SIR stadijų epidemijos modelis su tinklu, kuris nurodo erdvinius arba socialinius ryšius. Šiame darbe pristatomas modelis sujungiantis šias dvi idėjas. Darbo tema ir modelio idėja buvo pasiūlyta darbo vadovo prof. habil. dr. Mindaugo Bloznelio. Modelis yra pagrįstas atsitiktinėmis simuliacijomis, kurių metu individai juda erdviniame tinkle ir susitikę vienoje vietoje gali užkrėsti vieni kitus. Infekcinės ligos plitime taip pat atsižvelgiama į socialinius ryšius - vienas kitą pažįstantys individai lengviau perduoda infekciją, t.y., užsikrečia su didesne tikimybe. Abu, erdvinis ir socialinis tinklai, realizuojami kaip atsitiktiniai grafai. Toliau darbe nagrinėjamos epidemijos kontrolės priemonės. Galiausia, ištiriama galimybė pritaikyti šį modelį Covid-19 duomenims pasirinktose Lietuvos savivaldybėse.

Raktiniai žodžiai: dviejų tinklų epideminis modelis, SIR, atsitiktinis grafas

Contents

1 Introduction

Throughout human history there have been epidemics of diseases that affected a considerable number of people in the populations and often caused many deaths in some areas or all around the world. For example, it is estimated that the 14th century Black death epidemic has reduced European population by one-third [55]. Later in the 18th and 19th centuries plague outbreaks continued in North Africa, Egypt, Syria, Greece, China, and other countries resulting in more than ten million deaths. Due to epidemics having such devastating consequences for societies all over the world, the study of epidemics and diseases has been an important and relevant topic throughout history.

The initial study of epidemics mostly concerned with investigation of disease, its characteristics, and causes [11]. The study naturally progressed towards analysis of infectious disease data as records of numbers and causes of death have been collected and published. It was then possible to analyze various causes of death and estimate the risk of dying from various diseases. Further progress was made with the introduction of mathematical models for epidemic disease spread, which allowed to estimate the number of infected individuals in the population, the progression of the disease spread through time, and finally develop disease control measures, comparing and estimating their impact and effectiveness.

The topic of epidemic modelling has become especially relevant with the rise of the new Covid-19 epidemic. With modern technology and medicine, the course of the epidemic has been thoroughly recorded and there are more technological tools than ever available for the modelling of the disease. In the initial stages of the disease there were no vaccines or other pharmaceutical remedies, hence, epidemic modelling was a crucial tool that helped governments make decisions that could control and reduce disease spread.

The object of this master thesis is the discrete-time dynamic epidemic model with location and social networks (the topic of the master thesis and model concept were suggested by the supervisor prof. habil. dr. Mindaugas Bloznelis). In the current literature on epidemic models over networks, two approaches are prevalent: modeling epidemic spread via a social network or a geographical map. The overall

purpose of this research work is to show that these two modelling approaches can be combined into one epidemic model that is a more intuitive representation of the real world.

Aim of the thesis is to develop an epidemic model with location and social networks and investigate its dynamics.

Objectives for this work are:

1. Build a new epidemic model that combines location and social networks.
2. Analyze the model elements and their interactions.
3. Introduce infection control measures and assess their effectiveness.
4. Fit built model to real data and compare against other modelling approaches.

The structure of this model is distinct from other models described in literature. Epidemic models with networks are often implemented using a contact network where vertices of the network represent individuals and edges represent contacts between individuals through which disease can spread [30, 35, 39, 50]. On the other hand, the model explored in this work represents actors as independent agents. They move along a location graph, which represents the spatial aspect of the model, while the connections with other actors are defined by the social network. The disease is spread only when actors meet in the same location, which is in the same vertex of the location graph. Disease transmission chance increases for actors connected in the social network. This modelling approach more closely resembles real world disease spread dynamics. Actors just like people move in their environment and infect one another. People have closer contact with their acquaintances, which results in a higher disease transmission chance. This model is tailored to communicable diseases like Covid-19, influenza, chickenpox, etc. but it could be modified to apply to other types of diseases too.

The epidemic is modelled using the SIR model compartments. SIR is the most popular epidemic model due to its versatility and suitability for modelling a variety of different diseases. Even for diseases that could be modelled with more granular compartments, SIR is often used due to its good level of abstraction and simplification

of the disease stages [60, 43, 46, 33, 12].

In this work location and social graphs are synthetic networks, which are generated as random graphs. Based on literature on Zipf's law, we make an assumption that location network vertex degree sequence has a Pareto distribution [1]. For this reason, we use the configuration model that allows graph generation from a given vertex degree sequence. The social network is generated using a different method. We assume that actors form a connection in the social graph if they share common interests. For this reason, a bipartite graph is created connecting attribute vertices with actor vertices. Finally, an intersection graph based on the aforementioned bipartite graph is produced.

Afterwards we introduce epidemic control measures that aim to impede disease spread. We implement lockdown, social distancing, general preventative measures, and vaccination. The implementation approach of lockdown and social distancing are distinctive to this model in a way that only in a model with location and social networks is possible. We implement lockdown via removal of location network edges and social distancing via removal of edges from the social network.

This document is structured as follows - literature review section provides an overview of other research papers on the epidemic modelling topic. The methods section describes model design and model elements. Influence of model parameters section studies each of the model parameters and its interaction with model outcomes. The epidemic control section analyzes various epidemic control measures and their effectiveness. Comparison with the classical SIR model section provides a comparison between proposed and classical model. Finally, the data fitting section shows how proposed and classical model are fit to real Covid-19 data.

Acknowledgment. The master thesis topic, which consists of the idea for the model with location and social networks, and the concept of studying epidemic control measures by removing edges in location and social networks were proposed by the supervisor prof. habil. dr. Mindaugas Bloznelis.

Note. The terms *graph* and *network*; *node* and *vertex* are used interchangeably in this work.

2 Literature review

Epidemic modelling originated in the early 20th century. One of the earliest fundamental works were published by W. O. Kermack and A. G. McKendrick in 1927, later on in 1932, and 1933 [38]. The early epidemic models had three compartments: susceptible, infected, and recovered or dead. Their papers described a basic compartmental model characterized by a population of actors, multiple disease stages called compartments, and discrete or continuous time. In addition, model introduced a transmission or infection rate and recovery or death rate. Both rates are affected by the age of infection, which is the amount of time passed from becoming infected. The model assumed a stable population size and a homogeneous mixing of the individuals, meaning that the population would be randomly mixed to produce random contacts between individuals. Thus, each actor could have contact with any other actor in the population.

The area of epidemic modelling saw further development in the 1950s and 1960s. During this time, G. Macdonald has introduced the concept of reproduction number in his paper on malaria ([28] quoted by [47]). The reproduction number is the expected number of cases generated by an infectee in a population where all individuals are susceptible to infection. It was discovered that a threshold of one for the basic reproduction number determines if the infection becomes extinct or infection becomes an epidemic [11].

It was noted that Kermack's and McKendrick's assumption on homogeneous population mixing has defects in modelling the beginning of an epidemic when the fraction of infected individuals is small. The first work to improve the modelling of the beginning of an epidemic was by Metz, which introduced a stochastic branching process as a way to describe infection pattern in the initial stage of an epidemic ([51] as cited in [11]). In the branching process approach, the birth of an individual would represent the infection of a susceptible individual, and death would represent the recovery or death. The branching process was a good approximation of epidemic spread in the initial stages of the epidemic, and also the final ones [40]. In addition, the assumption of homogeneous population mixing was substituted by a new assumption stating that at the beginning of a disease there is a contact network described as a graph. Here the works of P. Erdös and A. Rényi in the study of

graphs were of great importance ([23, 24, 25] as cited by [11]). The contact network is a graph with vertices representing individuals and edges - contacts between individuals. The disease is transmitted along the edges of the graph from an infected to a susceptible individual.

The usage of networks in epidemiology has followed the development of network science in areas of social sciences and graph theory [34]. In 1951 Solomonoff and Rappaport were the first to relate epidemic models with networks ([66] as cited by [44]). In 1967, the work of Milgram in an area of social science led to the construction of a small-world social network and an idea of six degrees of separation ([67] quoted by [44]). The similarity between the spread of information via social network and the spread of infectious disease was utilized by Bass, Fisher, and Pry in their study of new product adoptions ([6, 7] as cited by [44]). Thus, further preparing a way for network-based diffusion models. In 1977 work by Leinhardt used network analysis as a tool to describe the spread of ideas and innovations in societies ([42] quoted by [34]). After that, many papers including Scott (1991), Wasserman and Faust (1994) have derived measures of the importance of an individual in a network ([63, 71] as cited by [34]). These measures varied from simple – like vertex degree, which is a number of connections of an individual, to complex – like betweenness centrality, which is a number of paths between other actors in which an individual is present.

Using principles laid in social sciences and graph theory many epidemiology papers analyzed the spread of HIV/AIDS through contact networks ([30, 35, 39, 50] as cited by [47]). Other papers developed epidemic models considering the spatial heterogeneity of the population [2, 3, 45, 49]. Later a variety of different networks were used for epidemic modelling: random networks [13, 53, 54], scale-free networks [75], multilevel networks [16, 72], and other [47].

The rise of the new epidemic of Covid-19 sparked an interest in the study of epidemic modelling. A plethora of papers were published on such topics as the spread of Covid-19 in networks, the impact of mobility on the spread of disease, vaccination strategies, lockdown, and social distancing strategies, etc. [46, 56, 12, 59, 33, 12].

2.1 Compartmental models

Compartmental models could be considered the origin of epidemic modelling. The main idea is that every individual in the population is assigned a compartment. Then rules are defined on how individuals migrate between different compartments. Compartmental models are mostly based on systems of ordinary differential equations, which are deterministic, but by adding a random component, they can also be based on stochastic processes formulated as discrete-time Markov chain, continuous-time Markov chain, and stochastic differential equations. Later it was realized that compartmental models have limitations in fully capturing spatial and differing social contact effects for disease spread. Currently, most popular epidemic models are not fully compartmental, but use the same ideas in conjunction with networks.

The SIR model over a network is by far the most popular epidemic model, which is described in papers by Ruget, Rossi, Pepler [60], Lev and Shmueli [43], Maheshwari and Albert [46], Szapudi [33], Pizzuti, Socievole et al. [12] to name a few. The model is characterized by $S(t)$ number of susceptible individuals, $I(t)$ infected individuals, and $R(t)$ immune or removed individuals [15]. Model assumes that population size is fixed, therefore, $N = S(t) + I(t) + R(t)$. The model has two parameters: β - an infection parameter indicating infection rate, and γ - a removal parameter, the rate at which infected individuals recover and become immune or die. Individuals start in the susceptible stage, then when infected flow to the infected stage, and lastly after some time individuals recover to the recovered stage.

Figure 1: The compartmental diagram for the SIR model.

The classical SIR model is defined by the following differential equations [38].

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta SI}{N}, \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I, \\ \frac{dR}{dt} &= \gamma I. \end{aligned} \tag{1}$$

Here S, I, and R denote the number or proportion of individuals that are in susceptible, infected, and recovered stages of infection accordingly. The γ parameter describes the length of time d spent by an individual in the infectious state. It is defined in such a way: $\gamma = 1/d$. The parameter β describes the average number of susceptible individuals one infected actor has contact with and infects.

Apart from SIR compartmental model there is an abundance of other compartmental models - SI, SIS, SEIRD, SEIRS to name a few. These models often expand on the classical model by introducing additional disease stages and appropriate transition parameters. For example, commonly added disease stages are E - exposed, describing the latent period of the infection, D - deceased stage for individuals that did not survive the disease, V - vaccinated, to introduce vaccination of susceptible individuals. Another common modification is allowing recovered individuals to return to susceptible stage, which is signified by additional letter "S" at the end of the name.

2.2 Networks in epidemic models

Traditionally epidemics were modelled using compartmental models. However, simple compartmental models did not account for spatial epidemic spread or non-homogenous social contacts in the population. Thus, with increasing technological possibilities, epidemic models are often combined with spatial or social networks creating a more sophisticated and improved model.

The most commonly compartmental model is enriched by an introduction of social network. Social networks focus on relationships among social entities, patterns, and implications of these relationships [71]. Nodes of social networks usually represent actors and edges relationships between individuals. In epidemic modelling, relationship links between actors are often channels for the transfer of infection from one individual to another.

Another type of network, which is less commonly used in epidemic modelling, is the spatial network. Spatial networks are types of networks where nodes and edges of the graph are embedded in space [5]. Nodes of the spatial network usually represent locations and edges have weights or costs corresponding to length or distance between vertices. The spatial networks are more often used in transportation and

mobility research area, they can often represent agent mobility based on mobile phone information or they can be a power grid network representing the size and development level of cities and connections between cities. When used in epidemic modelling, spatial networks are frequently recreated using some type of data like geographical information from maps, power grids or using movement information gathered from mobile phones usage, transportation routes, etc.

2.3 Modelling of epidemic prevention strategies and their effect on the spread of disease

Two interesting questions in epidemic modelling are how to prevent or stop disease spread and what is the effect of different prevention strategies. The answers to these questions are a particularly useful resource for decision making in the area of public health. From a modelling perspective, it is important to find optimal strategies for disease prevention that minimizes the number of infected individuals as well as does not overload the healthcare system, exceed the public budget, or overly restrict individual mobility, and economic activity.

Vaccination has been the most popular measure for the prevention of infection spread in epidemic models. Due to limited resources and often no possibility of vaccinating the entire population at once, the problem is then to find the best targeted vaccination strategy that allows to first vaccinate part of the population that is the most vulnerable or that has the highest potential to spread the disease. The majority of epidemiology literature relies on network topology for vaccination strategy. For example, vaccinating nodes with the highest degree [75], vaccinating nodes with the highest betweenness centrality score [61], vaccinating based on local information and friendship paradox called acquaintance strategies [43], etc.

Apart from vaccination, which is not always possible due to a long vaccine development period, even if vaccines are available there might be budget restrictions or shortage of vaccines, other disease prevention strategies are explored. Lockdown, social distancing, quarantine can be used to effectively mitigate, and control infection spread. Social distancing means reducing physical interactions, maintaining two meters distance from other people, closing offices, schools, and cancelling all group gatherings. A lockdown is government enforced closure of non-essential services or

activities, limiting travelling, and requesting people to stay at home. During a lockdown, only essential services are provided. Quarantine is the isolation of people that were exposed to the infection or suspected to be exposed.

3 Methods

Our model encompasses the following elements: location network, actors, social network, and compartmental model for defining infection stages and transitions between stages.

The process of constructing this model starts with the location graph, which is generated to resemble a city. The vertices of the graph represent various city locations, e.g., workplaces, shops, houses, etc., while the edges of the graph denote the connections between locations. The public places in the location graph are visited more frequently and by a wider variety of individuals as compared to residential locations. The difference in traffic is reproduced by having more connections leading to public places compared to residential areas. Then there are actors or city residents that occupy these locations. Actors can randomly move to neighboring locations or stay at the same place in the location graph. Furthermore, we construct a social network, which represents social links between individuals in the population. The vertices of the network represent actors and edges mark the social connections between actors. The last element of this system is the infection and its spread. The epidemic is modelled using the well-known and widely used SIR model. The model consists of three compartments: "S" susceptible, individuals who are not infected but could be infected during contact with a diseased individual; "I" infected, actors who are sick with the disease; "R" recovered - after some time previously sick actors heal from the disease and gain immunity to the infection. The classical SIR infection probability is split into two probabilities. Infection probability $0 < p_1 < 1$ is used when actors meet in location graph, however, they are not connected in the social network. Higher probability $0 < p_1 < p_2 < 1$ is used when actors meet in location graph, and they are connected in social network. This simulates having a higher probability of contracting a disease among people that have a social connection.

3.1 Location graph

In this section we discuss the generation method for location graph. First, we describe the chosen algorithm, relevant parameters, and assumptions, and then provide examples of generated location graph.

To generate a location graph, the configuration random graph model was used. The configuration model provides a way of generating a random graph with specified vertex degree sequence [4]. Thus, the most important part of the location graph construction is selecting an appropriate degree sequence.

We generate degrees as realizations of independent and identically distributed random variables. To select an appropriate random variable distribution, we look at literature. In 1913 it was first observed by Auerbach and later refined by Singer and Zipf that city size distribution follows Zipf's law [1], i.e., when all cities within a country are ranked based on their population size, it is observed that the largest city is twice as big as the second largest, and three times as big as the third largest, etc. [36]. This relationship is called Zipf's law and can be expressed as power law (with exponent between zero and two) or Pareto distribution with shape parameter being one. Although Zipf's law is not universal and there are cases when it is violated, for example, when the number of cities within country is too small like in Singapore, it is a good empirical rule for sufficiently big data sample that constitutes a whole unit, i.e. it is sufficiently interconnected and not strongly affected by other neighboring countries or regions. Also, Jiang et al. [36] observed that Zipf's law holds not only for cities within a country but also when comparing countries within continents. If we assume that Zipf's law extends to locations within a city in the same way as to countries within continents then we can argue that location distribution within a city holds the same Zipf's law. This allows us to conclude that empirically location size or degree distribution should follow Zipf's law or more generally power law.

The degree sequence elements are generated as realizations of independent random variables having a Pareto distribution. However, Pareto distribution generates non-integer values or values that are too large for the purpose of degree sequence. Thus, degree sequence is generated with modifications that ensure that sequence is fit to be degree sequence.

Another assumption we make is that real world location network is a connected graph. However, configuration model can generate graphs that are not connected. To ensure generated graph connectivity, first, shifted Pareto distribution is used where minimum generated vertex degree is higher or equal to three, which according to Theorem 4.25 (Connectivity of $CM_n(d)$) in [68] ensures graph connectivity with sufficiently high probability. Secondly, if the generated location graph is disconnected, which can happen with probability $\mathcal{O}(1/n)$, where n is number of vertices in the graph, the location graph is re-generated to a maximum of ten times. If after ten tries the generated graph is disconnected, we take the largest connected component and use it as a location graph.

The configuration model does not construct simple graphs, i.e., resulting graphs can have self-loops and multi-edges. These represent a very small share of the graph edges, and they are redundant in the context of location graph. The self-loops are equivalent to actors staying in the same vertex of location graph and multi-edges would translate to higher probability for actor to move to vertex with multi-edge connection. Hence, we discard them.

Degree sequence generation process:

1. **Generating**. The degree sequence elements are generated from shifted Pareto distribution.

   Let X be a random variable with Pareto distribution. The cumulative distribution function of X is

   $$F_X(x) = \begin{cases} 1 - x^{-a} & x \geq 1, \\ 0 & x < 1, \end{cases}$$

   and probability density function

   $$f_X(x) = \begin{cases} \frac{a}{x^{a+1}} & x \geq 1, \\ 0 & x < 1, \end{cases}$$

   where a is shape and $a > 0$. Hereinafter, we denote random variable that follows Pareto distribution as specified above with notation $X \sim \text{Pareto}(a)$.

Then the degree sequence is generated from the realization of $X + 2$.

2. **Resampling**. Since degree of the vertex is defined as the number of adjacent vertices [41], the degree of any vertex cannot exceed the total number of vertices in the graph. Thus, for undirected graph without loops with n vertices, the maximum degree of the vertex is $n - 1$. Any sequence elements exceeding $n - 1$ are resampled from the same Pareto distribution.

3. **Rounding**. Since vertex degree is a natural number, all sequence elements are rounded to the natural number.

4. **Sum is even**. The handshaking lemma [41] states that the sum of the degrees of the vertices V in a graph $G = (V, E)$ equals twice the number of edges E, that is $\sum_{v \in V} deg(v) = 2|E|$. This lemma implies that the sum of degrees is an even number. Generating sequences from a distribution may result in sequences where the sum of elements is not an even number. So, this modification ensures that the sum of sequence elements is an even number. In cases when the sum is not even, then one is added to the first element of the sequence ($y_0 = y_0 + 1$) and thus making the sum of sequence elements even.

5. **Is graphical**. The last step is done to assess if the generated sequence is indeed a vertex degree sequence and there exists a graph that can realize such degree sequence [27]. Sequences that cannot be realized as a degree sequence of a graph is called non-graphic and sequences that can be realized is called graphic. In this step we use Erdős–Gallai theorem 1 (see below) to ensure that generated sequence is graphical. If the generated sequence fails to be graphical, a new sequence is generated in its place starting from the first step.

Theorem 1 (Erdős–Gallai theorem [18]) *A non-increasing sequence $[d_i]_1^n$ of non-negative integers is a degree sequence if and only if $D = [d_i]_1^n$ is even and the inequality*

$$\sum_{i=1}^{k} d_i \leq k(k-1) + \sum_{i=k+1}^{n} min(d_i, k)$$

is satisfied for each integer k, $1 \leq k \leq n$.

Lastly, some examples of generated location graph are showed in the figure below.

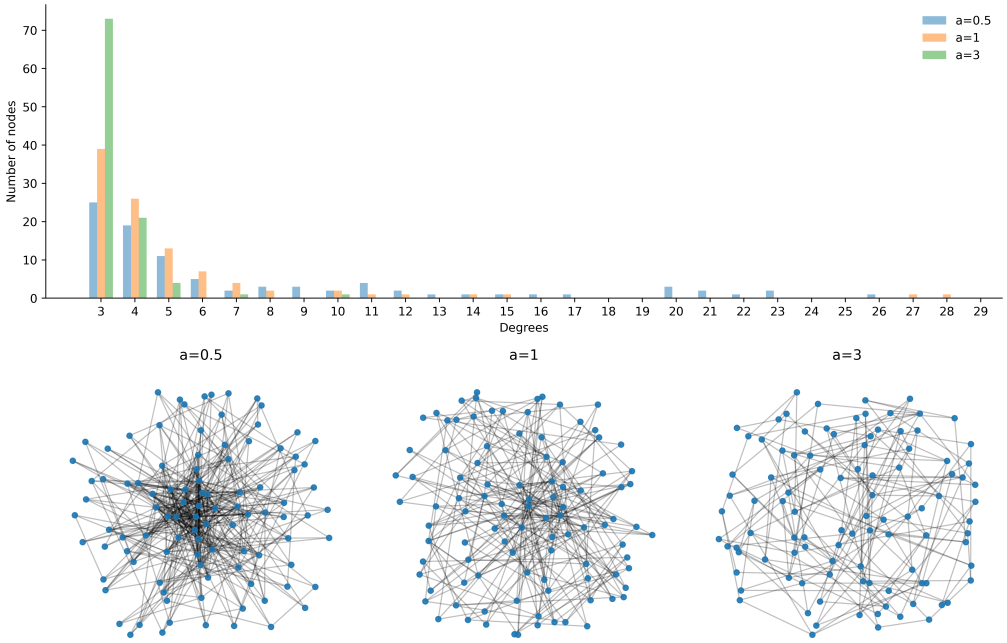

Figure 2: Location graphs with 100 vertices realized using different degree sequences.

It is apparent that lower Pareto parameters result in degree sequences with larger degree values and thus generating graphs with few strongly interconnected vertices. Whereas higher Pareto parameters result in most degrees being close to the minimum value of three, thus resulting in graphs that are less interconnected. In reality, different a value can represent different density cities, where high density cities would have a low a value, for example, $a = 0.5$, and cities with lower density would have higher a value like 3 or higher.

3.2 Social network

In this section we discuss the generation method for social graph. We describe the chosen algorithm, relevant parameters, and assumptions.

We construct a social network as a random intersection graph obtained from a random bipartite graph [8]. The vertices of the underlying bipartite graph can be divided into set $V = \{v_1, ..., v_n\}$, which contains vertices representing actors, and set $W = \{w_1, ..., w_m\}$, which contains vertices representing attributes. Attributes can be understood as common hobbies, common interests, or other activities like the same place of education, work, etc. that form connections between people. Actors and attributes are linked with probability based on actors' activity level and

attributes attractiveness. Then actors are connected in the social network if they have at least one hobby, interest, or activity in common.

The following steps describe the process of constructing social network in detail [8]:

1. **Constructing a random bipartite graph**

   A random bipartite graph $H = (V, W, E_h)$ with vertices consisting of n actors $V = \{v_1, ..., v_n\}$ and m attributes $W = \{w_1, ..., w_m\}$ is generated with the following steps.

   (a) **Generating weight sequences**.

   For each actor $v_i \in V$ generate a weight $x_i > 0$ and for each attribute $w_j \in W$ generate a weight $y_j > 0$. The weight of an actor represents its activity level, that is how likely an actor is to participate in many hobbies, have many interests, and activities. The weight of an attribute represents the attractiveness of an attribute to actors.

   (b) **Probability matrix**.

   The probability to connect actor v_i with attribute w_j is expressed as

   $$p_{ij} = min\left(\frac{x_i y_j}{\sqrt{nm}}, 1\right),$$

   where $x_i > 0$ is weight representing actor activity level, $y_j > 0$ is weight representing attribute attractiveness, n and m are the number of actors and attributes.

   (c) **Connecting actors with attributes**.

   With previously calculated probabilities we connect actors with attributes and obtain a realization of random bipartite graph H.

2. **Retrieving social network $G = (V, E)$ from bipartite graph H.**

   The social network G is retrieved from bipartite graph H by connecting all actors v_i and v_k that share the same attribute w_j in graph H.

It is apparent that the sequences of weights determine the geometry of the social network. From empirical studies it is observed that connections in the social network are following the power law. Thus, it is most intuitive to use Pareto distribution

to generate the weight sequences. However, differently from location graph, we will not use shifted Pareto distribution, since social network can be both connected and disconnected graph. Then the actor weight sequence $\{x_n\}$ is generated from realizations of $X \sim \text{Pareto}(\alpha)$ and attribute weight sequence $\{y_m\}$ from $Y \sim \text{Pareto}(\beta)$.

Lastly, it is important to use parameters $\alpha > 1$ and $\beta > 2$. Otherwise, generated social network can have some actors that are connected with all other actors in the network. This is unwanted graph structure considering that it is not realistic that few actors have the potential to infect all other actors when modelling epidemic spread.

Note that lower α and β values result in graph with more edges and higher values yield fewer connections (see Figure 3 for few examples).

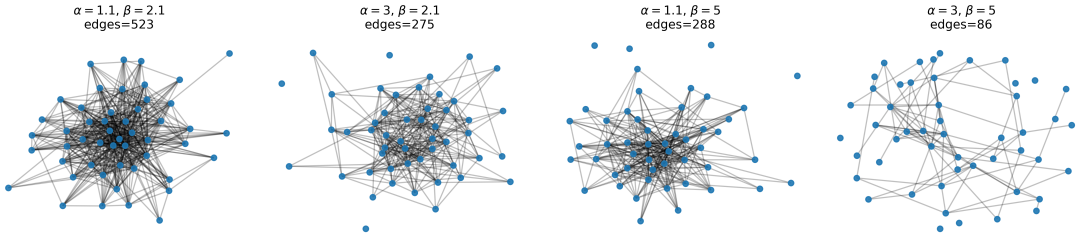

Figure 3: Social network with 50 vertices realized using different α and β values.

3.3 Model simulation

Once all model components are built, we can start simulating epidemic spread in networks. All model parameters are summarized in the table below.

Category	**Parameter**	**Description**
Location network	number of vertices	Number of vertices in a location network.
	a	Pareto parameter for generating degree sequence.
Social network	number of attributes	Number of attributes in an underlying bipartite graph.
	α	Pareto parameter for generating actor weight sequence.

	β	Pareto parameter for generating attribute weight sequence.
General	number of actors	Number of actors in the simulation.
	initial number of infected actors	Number of actors that are infected at the start of the simulation.
Simulation	p_1	Infection probability for one infected actor to transmit disease to susceptible actor in case when actors are not connected in the social network.
	p_2	Infection probability for one infected actor to transmit disease to susceptible actor in case when actors are connected in the social network.
	t_r	Time it takes to recover from infection.

Table 1: Model parameters.

Model initialization process:

1. Generate location graph.

2. Generate social graph.

3. For each actor, allocate infection stage and starting vertex in the location graph. Infected stage is randomly assigned to selected number of actors; the rest of the actors are assigned to susceptible stage. The starting location is selected randomly from all vertices in the location graph.

Model simulation process:

1. Spread infection. When infected actors meet susceptible actors in the same vertex of location graph, infected actors can spread the infection to susceptible individuals.

   Let susceptible actor a in location vertex v meet m_1 infected actors that are not connected in the social network, i.e., they are strangers to actor a, and m_2 infected actors that are connected in the social network, i.e., they are ac-

quaintances to actor a. Each of m_1 actors has the probability p_1 to infect a, while each of m_2 actors has the probability p_2. We assume that for actor a each disease transmission event from different infected individuals is independent. Then probability for actor a to get infected from m_1 strangers and m_2 acquaintances is[1]:

$$p_a := \mathbb{P}(\{\text{actor } a \text{ gets infected}\}) = 1 - \mathbb{P}(\{a \text{ does not get infected}\}) = 1 - \mathbb{P}(\{a \text{ does not get infected from strangers}\} \cap \{a \text{ does not get infected from acquaintances}\}) = 1 - (1 - p_1)^{m_1}(1 - p_2)^{m_2}$$

2. After t_r time moments counting from the time of infection, infected actors recover from infection and move to recovered stage.

3. All actors with equal probabilities move to neighboring vertices or stay at the current location in the location graph. For example, if actor a is in vertex v, which has a degree equal to 3 then actor a has 1/4 probability to stay in the current location and 1/4 probability to move to one of the neighboring locations.

The simulation process is repeated the desired number of times with each repetition time moment is increased by one.

The model simulation contains stochastic elements like random movement of actors, infection events, and even generation of graphs that represent location and social networks. Thus, the same initial parameters can yield varying simulation outcomes. To see a full range of model outcome possibilities we will perform Monte Carlo simulations. Monte Carlo simulations are used as it is not possible to use other approaches. The process is simple - select model inputs and then simulate the model with selected inputs many times. This provides an overview of all different model simulation outcomes and allows to estimate the mean and standard deviation of the results.

[1]The probability space here is constructed of all possible infection events for actor a.

4 Influence of model parameters

In this section we examine how each model parameter affects epidemic spread dynamics and suggest relevant value ranges for each parameter that results in more realistic simulation outcomes. The model has six relevant parameters, namely infection probabilities p_1, p_2, time to recovery from infection t_r, Pareto parameter a, which is used to generate degree sequence for location graph, Pareto parameters α and β, which are used to generate social network. We will measure the following aspects of the epidemic spread:

- Maximum number of active infection cases (also referred to as the peak of the infection, maximum active cases, maximum cases), i.e., the largest share of population that was in infected stage at the same time.
- Total infection cases (also referred to as total cases), i.e., the share of population that was infected with disease at any point of time during the simulation.
- Epidemic duration, i.e., the number of time moments from infection spread start till the last infected actor recovers.
- Basic reproduction number (R_0), which describes an expected number of individuals that get infected directly from one infected actor (secondary infections are not counted) where all other actors are susceptible to infection.

In the next subsections, we will analyze each parameter according to the four metrics listed above.

4.1 Location graph

The impact of location graph on simulation outcome is dependent not only on Pareto parameter a used to generate degree sequence but also on the population density. We denote the ratio of the number of actors to the number of location graph vertices as population density $d = \frac{N_{actors}}{N_{locations}}$.

The Figure 4 shows the relationship between parameter a used to generate degree sequence for location graph and epidemic spread with different population densities.

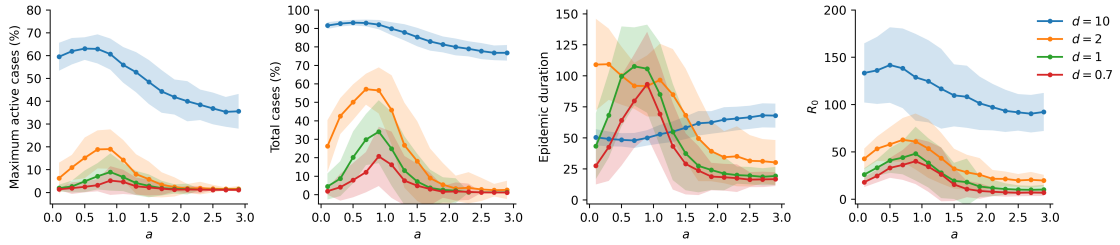

Figure 4: Relationship between a and epidemic spread with different population densities d. The remaining parameters are $p_1 = 0.01$, $p_2 = 0.1$, $t_r = 10$, 1000 actors, $\alpha = 1.1$, $\beta = 2.1$. The curve shows results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

Higher a values generate degree sequences with smaller degrees, which result in sparser location graph, which in turn makes it harder for actors to meet in the same vertex of the graph slowing down the infection spread. With sparser location graphs the peak of epidemic is smaller and the number of total infection cases is also lower. The R_0 is slightly lower for higher a values.

The population density is not a parameter of the model. However, it has strong influence over the epidemic spread as low enough density might stop epidemic spread. This happens because actors rarely meet other actors and have a low chance of spreading infection when there are too many locations or too few actors. As can be seen from Figure 4, with $d = 10$ (which equates to 1000 actors and 100 vertices) epidemic spreads fast and wide with peak from 30% to 70% and total infection cases between 70% to 95% depending on a value. Population density lower than 2 significantly slows down the infection spread with peak reaching not more than 25% of population and up to 60% total cases. The observed dynamic nicely corresponds to real world dynamics, for example, Covid-19 spread was faster and wider in high density areas like Hong Kong compared to low density ones [74].

The epidemic duration depends on Pareto parameter a and population density d. With high population density, epidemic duration is shortest when a value is low. With lower population density, the relation between a value and epidemic duration is non-monotonic. The epidemic duration is shortest with very low and very high a values and peak with a value at around 0.5 to 1.

The population density and location graph degree sequence have a strong impact on R_0 values. The calculation of R_0 often depends on the duration of infection

contagiousness (in our case it corresponds to the parameter t_r), the likelihood of spreading infection (in our model corresponds to p_1 and p_2 probabilities), and contact rate between individuals [17]. In our model the contact rate is influenced by the structure of the location graph and population density d. Hence, different values of location graph parameter a and population density d greatly affect the basic reproduction number (R_0).

4.2 Social graph parameters α and β

Pareto parameters α and β are used to generate actor and attribute weight sequences, which are then used to generate social network.

The Pareto parameter α is used to generate actor weight sequence that represents actors' activity level. Lower α values give higher weights and higher α values result in lower weights, which means that with lower α values actors create more connections with various attributes and thus have more connections with other actors sharing the same attribute. With higher α values actors form less connections with attributes and in turn with other actors. Hence, lower α means more connections in the social graph and faster and wider spread of the infection. With higher α values both maximum and total infection cases decrease.

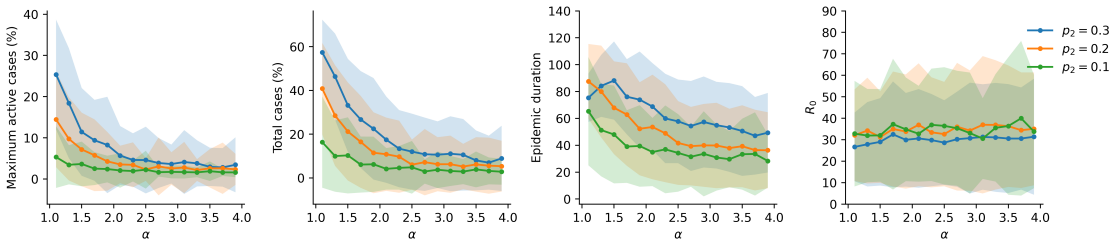

Figure 5: Relationship between α and epidemic spread with different p_2 probabilities. The remaining parameters are $p_1 = 0.01$, $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$, $\beta = 2.1$. The curves show results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

The β parameter has similar effect to α but it applies to the attributes rather than actors. Likewise lower β values give higher attribute weights making them more attractive to actors and resulting in more connections between different actors and attributes. Higher β values result in lower weights and less connections between actors and attributes. Thus, analogously lower β yields faster spread of infection,

higher peak values, and more total infection cases. Whereas with higher β values, infection spread is slower and less actors get infected.

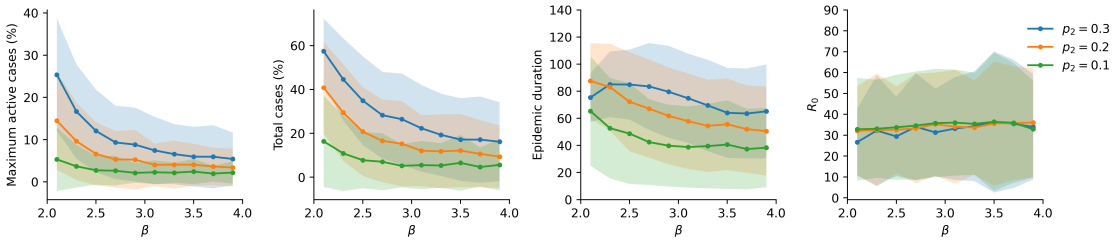

Figure 6: Relationship between β and epidemic spread with different p_2 probabilities. The remaining parameters are $p_1 = 0.01$, $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$, $\alpha = 1.1$. The curves show results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

Notice that social network has influence over epidemic spread only when p_2 probability is sufficiently high (see Figures 5 and 6). If p_2 is too low and too close to the value of p_1 (in our example when $p_2 = 0.1$) both α and β have no effect or only a slight effect on selected metrics. Similarly, when $p_1 = p_2$ social network is "turned off" and cannot affect epidemic spread. This is confirmed in Figure 7 where p_2 probability is set to p_1. The result is two horizontal lines signifying that social network (both α and β parameters) have no effect over the simulation outcomes.

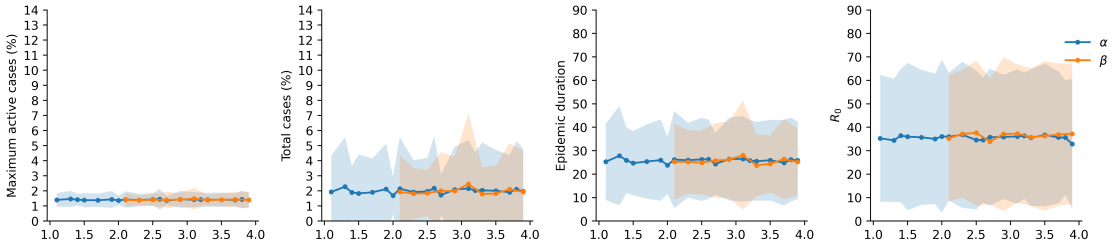

Figure 7: Relationship between α and β and epidemic spread when $p_1 = p_2$. The remaining parameters are $p_1 = 0.01$, $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$. The curves show results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

Lastly, regardless of what other parameters are selected, α and β have only a slight effect on epidemic duration and no effect on R_0.

4.3 Social graph structure

Another question to explore is what the influence of the social network structure on epidemic spread is. We hypothesize that epidemic spread mainly depends on the

average degree of the social network, whereas graph structure has weaker influence. To assess this hypothesis, we use five graphs with the same average degree, i.e., the same number of vertices and edges, but different structures. These graphs are then used as social networks in the model simulation. Then simulation outcomes in terms of peak (maximum share of active infection cases) and total infection cases are compared.

The first graph is social network used up till now. It is generated from an underlying bipartite graph with parameters α and β. The second graph is generated using a configuration model where degrees are drawn from Pareto distribution with shape a. This corresponds to the same algorithm that was used for location graph generation. The third graph is generated using a configuration model but with degree sequence sampled from the uniform distribution in the interval $[l, h)$. The fourth graph is a random d-regular graph where all vertices have the same degree d. The last graph is the Watts–Strogatz graph with parameters k - mean degree and p - the probability of rewiring each edge. The Watts–Strogatz degree distribution resembles a bell curve. All five graphs have significantly different degree distributions and global clustering coefficients (see Figure 8).

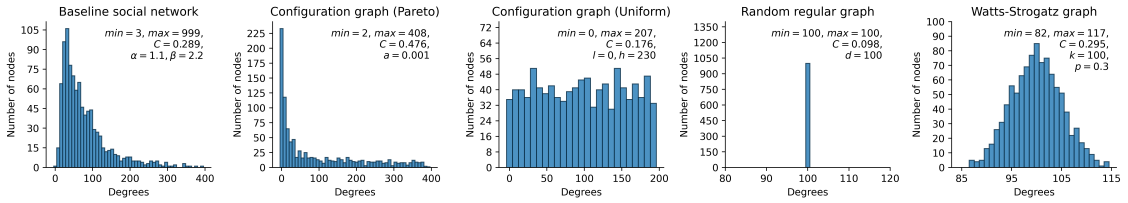

Figure 8: Degree distributions for five graphs with 100 average degree, 1000 vertices, 50000 edges. In the top right corner, minimum degree (min), maximum degree (max), and global clustering coefficient (C), and graph parameters are noted.

The simulation outcomes are summarized in the table below. It can be seen that when difference between p_1 and p_2 is smaller, social network has smaller overall influence on simulation outcomes and thus, social network structure is almost irrelevant with less than 1% difference in both maximum and total cases for all five graphs. When difference is larger, social network has stronger effect on simulation outcomes and hence, there are bigger differences in maximum and total cases between different graph structures.

Overall, the social network average degree has stronger influence over epidemic

spread as graph with average degree of 100 yields higher peak and more total infection cases compared to graphs with lower average degrees (for example, 50, 20, and 10). Although relative increase with higher average degree also depends on selected p_1 and p_2 probabilities.

Average degree	**Graph**	**Peak**, $p_1 = 0.01$, $p_2 = 0.2$	**Total**, $p_1 = 0.01$, $p_2 = 0.2$	**Peak**, $p_1 = 0.1$, $p_2 = 0.2$	**Total**, $p_1 = 0.1$, $p_2 = 0.2$
100	Baseline social network, $\alpha = 1.1, \beta = 2.2$	11.9%	35.8%	59.0%	94.3%
100	Configuration graph (Pareto), $a = 0.001$	11.9%	35.7%	58.9%	94.2%
100	Configuration graph (Uniform), $l = 0, h = 230$	9.5%	32.1%	58.8%	94.5%
100	Random regular graph, $d = 100$	7.6%	27.0%	59.0%	94.5%
100	Watts–Strogatz graph, $k = 100, p = 0.3$	7.4%	26.1%	59.0%	94.5%
50	Baseline social network, $\alpha = 1.6, \beta = 2.1$	4.6%	14.1%	56.4%	93.7%
50	Configuration graph (Pareto), $a = 0.19$	5.4%	16.5%	56.2%	93.6%
50	Configuration graph (Uniform), $l = 0, h = 110$	3.8%	11.8%	56.3%	93.6%
50	Random regular graph, $d = 50$	3.4%	10.6%	56.6%	93.8%
50	Watts–Strogatz graph, $k = 50, p = 0.3$	3.3%	10.3%	56.4%	93.6%
20	Baseline social network, $\alpha = 2.1, \beta = 2.3$	2.1%	5.1%	56.5%	93.5%
20	Configuration graph (Pareto), $a = 0.5$	2.7%	6.9%	56.6%	93.5%
20	Configuration graph (Uniform), $l = 0, h = 43$	2.0%	4.6%	56.4%	93.6%

20	Random regular graph, $d = 20$	1.9%	4.5%	56.2%	93.6%
20	Watts–Strogatz graph, $k = 20, p = 0.3$	1.8%	4.2%	56.5%	93.6%
10	Baseline social network, $\alpha = 2.8, \beta = 2.6$	1.8%	4.2%	56.2%	93.5%
10	Configuration graph (Pareto), $a = 0.8$	1.8%	4.5%	56.0%	93.6%
10	Configuration graph (Uniform), $l = 0, h = 21$	1.8%	3.9%	56.3%	93.5%
10	Random regular graph, $d = 10$	1.7%	3.7%	56.2%	93.5%
10	Watts–Strogatz graph, $k = 10, p = 0.3$	1.8%	4.2%	56.1%	93.4%

Table 2: Comparison between different graph structures and their effect on simulation results. Columns in table show graph average degree, peak - maximum active cases (%) with specified p_1 and p_2 values, total - total cases (%) with specified p_1 and p_2. Other simulation parameters are $p_1 = 0.01$, $p_2 = 0.2$, $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$. Simulation results are averaged over Monte Carlo simulations.

Simulation results confirm that social network structure has weaker effect on infection spread compared to the number of edges in the network. That is because actors walk randomly between vertices in the location graph. If random walk would be replaced with a more intelligent walking pattern within location graph, then social network structure would have stronger impact over simulation outcomes. For example, we could allocate a home vertex to each actor making sure that actors connected in social network get the same or close locations. Then random walk could be modified by limiting the actor's traveling distance from the home vertex and adding some preferential behavior to return to home vertex. This walking pattern would ensure that connected actors meet more often and spread the infection to friends, family members, and other connections in social network much more rapidly. With these modifications the structure of social network would have an integral role in epidemic spread and simulation outcomes.

4.4 Probability p_2

The p_2 probability corresponds to infection probability for actors that meet in the same location network vertex and are connected in the social network. The extent of p_2 influence over simulation results depends mostly on the selected social network, recovery time t_r, and probability p_1. Denser social networks yield higher epidemic spread with the same probability p_2 and very sparse networks (for example, when $\alpha = 5$ and $\beta = 5$) can switch off the effect of p_2, see Figure 9.

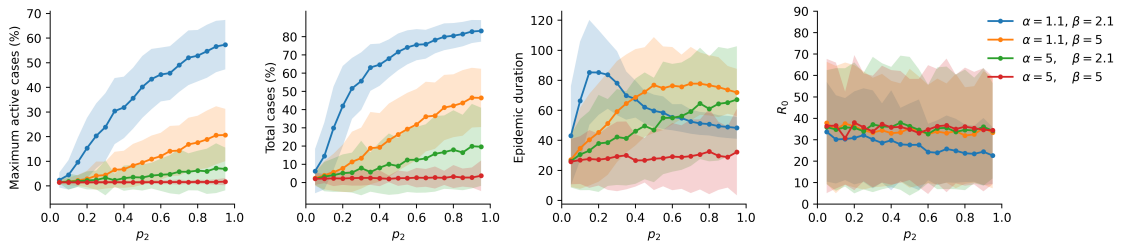

Figure 9: Relationship between p_2 and epidemic spread with different α and β. The remaining parameters are $p_1 = 0.01$, $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$. The curves show results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

Different sets of t_r and p_1 values also affect the influence of p_2. Figures 10 and 11 show that probability p_2 has the strongest effect on the maximum active infection cases with higher p_2 values resulting in rapid increase in the maximum infection cases. Depending on selected p_1 probability, total infection cases significantly increase with higher p_2 values or increase only slightly when approaching a maximum 100% of population.

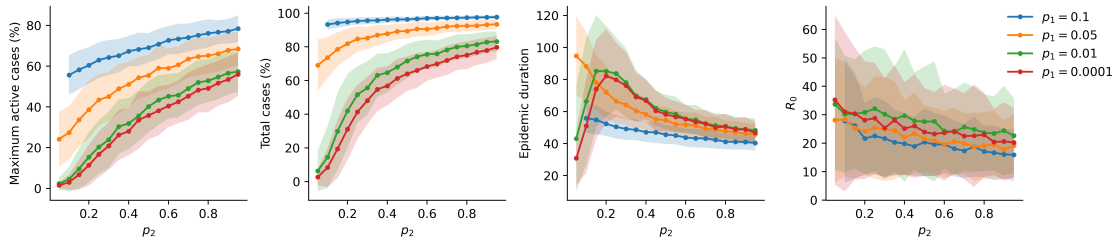

Figure 10: Relationship between p_2 and epidemic spread with different p_1. The remaining parameters are $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$, $\alpha = 1.1$, $\beta = 2.1$. The curves show results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

The epidemic duration is the shortest either with high p_2 values or very small values (see Figures 10 and 11), since higher p_2 probability result in faster spread of infection

and hence shorter total epidemic duration and very low p_2 probability result in infection spreading to small part of population and not spreading further, thus epidemic rapidly ends due to small share of population quickly getting infected and recovering.

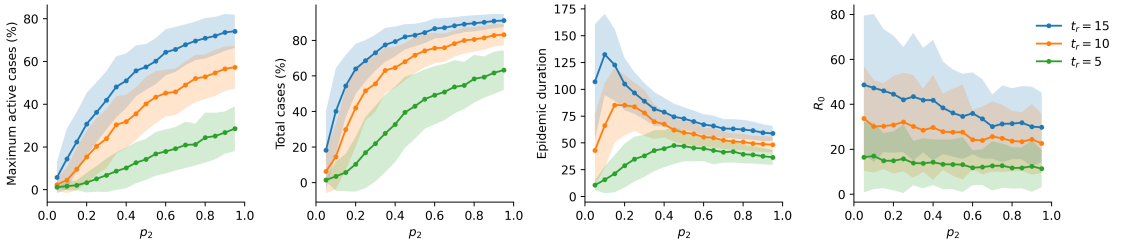

Figure 11: Relationship between p_2 and epidemic spread with different t_r. The remaining parameters are $p_1 = 0.01$, 1000 actors, 500 location graph vertices, $a = 1.5$, $\alpha = 1.1$, $\beta = 2.1$. The curves show results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

The p_2 has a slight effect on R_0. From Figure 11 it can be seen that R_0 is influenced more by selected recovery time t_r than different p_2 values.

4.5 Probability p_1

The p_1 probability corresponds to infection probability for actors that meet in the same location network vertex and are not connected in the social network. The influence of probability p_1 over the epidemic spread depends on the population density d and degree sequence of location graph, recovery time t_r, and probability p_2. Other model parameters have a coincidental effect over the simulation results. The effect of p_1 with different location graphs is showed in Figure below.

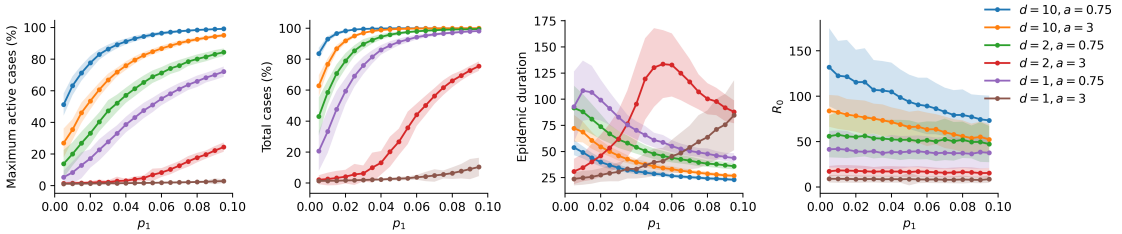

Figure 12: Relationship between p_1 probability and epidemic spread with different population densities d and degree sequence parameters a. The remaining parameters are $p_2 = 0.1$, $t_r = 10$, 1000 actors, $\alpha = 1.1$, $\beta = 2.1$. The curve shows results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

The effect of p_1 strongly depends on the selected location graph. Small and dense location graphs, for example, when $d = 10$ and $a = 0.75$, result in rapid and widespread epidemic where higher or lower p_1 probability can further accelerate or slow down epidemic spread. Given a dense location graph, the epidemic duration decreases with higher p_1 probability. Big and sparse location graphs, e.g., when $d = 1$ and $a = 3$, have the opposite effect where different p_1 values might have only marginal effect over the peak of infection and total cases due to sparseness and largeness of the location graph. The epidemic duration is then shortest with lower p_1 probability.

Furthermore, the effect of p_1 is different with different recovery times t_r (see Figure 13). Longer recovery time results in higher peak and more total infection cases and vice versa. The effect of increasing p_1 values result in a monotonic increase in maximum and total cases. The epidemic duration varies with different recovery times. Long recovery time yields shortest epidemic duration with high p_1 probability. Whereas short recovery time results in opposite effect where epidemic duration is the longest with high p_1.

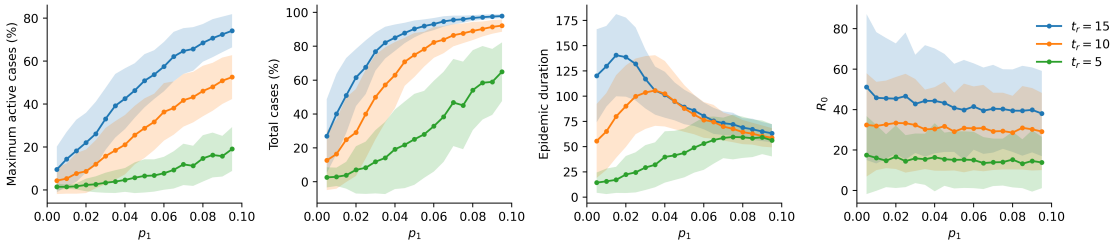

Figure 13: Relationship between p_1 probability and epidemic spread with different t_r. The remaining parameters are $p_2 = 0.1$, 1000 actors, 500 location graph vertices, $a = 1.5$, $\alpha = 1.1$, $\beta = 2.1$. The curve shows results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

Influence of probability p_1 is different with various p_2 values. Here similarly to recovery time, higher p_2 and p_1 values result in higher maximum and total cases. Whereas with lower p_2 values, the trend is preserved but with lower maximum value and total infection cases. The epidemic duration has similar non-monotonic trend for different p_2 probabilities. Epidemic duration is the shortest with very low and high p_1 values.

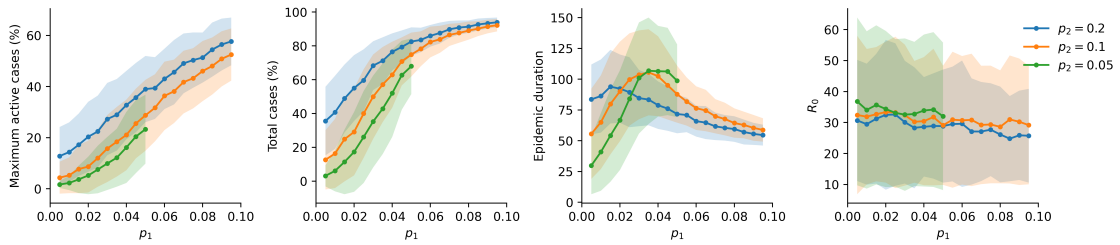

Figure 14: Relationship between p_1 probability and epidemic spread with different p_2. The remaining parameters are $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$, $\alpha = 1.1$, $\beta = 2.1$. The curve shows results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

All three Figures show that p_1 has slight effect on R_0. From Figure 13 can be seen that R_0 is influenced more by selected recovery time t_r than different p_1 values.

4.6 Recovery time t_r

The relationship between recovery time t_r and selected metrics is shown in the figure below.

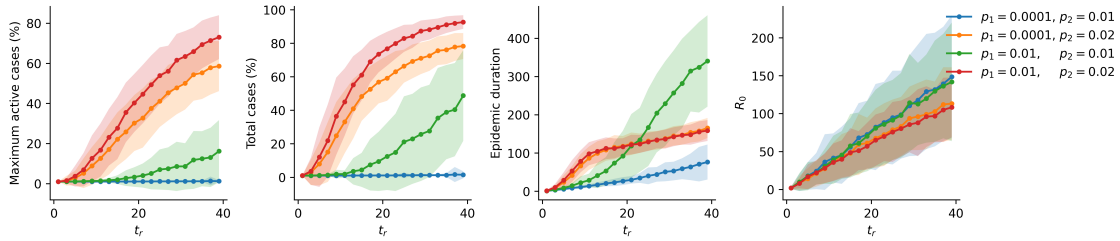

Figure 15: Relationship between t_r and epidemic spread with different p_1 and p_2 probabilities. The remaining parameters are 1000 actors, 500 location graph vertices, $a = 1.5$, $\alpha = 1.1$, $\beta = 2.1$. The curves show results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

It is clear that longer infection period results in larger peak of the infection and higher share of total infection cases. Similarly, with longer time to recover epidemic duration increases. Furthermore, recovery time together with population density has the strongest effect out of all other parameters on the R_0 even close to linear relationship with longer t_r resulting in higher R_0. This is due to the definition of R_0 - it is a number of susceptible actors one infectious individual can infect during the contagiousness period [17]. The contagiousness period in our model is t_r and setting different t_r value has a direct effect on R_0.

4.7 Parameter value ranges

We would like to choose such parameter values that result in a realistic epidemic spread. The realistic value range can be determined from empirical data of various epidemics. For example, in Lithuania the maximum number of active Covid-19 cases between municipalities never surpassed 13% of the population [20]. Similarly, in the United States maximum number of active cases did not exceed 5% of the population, for Germany - 5%, France - 10%, Italy - 5% [74]. The Institute for Health Metrics and Evaluation estimated that 77% of people globally have been infected with Covid-19 at least once [26].

There is significantly less data on other epidemics in history. The data on the number of active cases is not available for all other epidemics except for the Covid-19. However, the total infection cases are estimated for 1918 Spanish flu to be approximately 33% of the world's population (500 million people) [48], 2009 swine flu to around 11 to 21 percent of the global population (700 million to 1.4 billion) [37].

The R_0 can vary significantly between different diseases and estimation methods [17]. The estimates of R_0 for Covid-19 range from 0.4 to 12.58 between different studies. The overall pooled estimate of R_0 was estimated to be 2.66 [19]. For 1918 Spanish flu R_0 estimates range from 2.4 to 10.6 [70]. In some settings estimates were higher, for example, in confined "Devon" sailing boat setting R_0 was estimated to be 17. For 2009 swine flu, it was estimated that R_0 is between 1.3 and 2.3 [9]. For other diseases R_0 estimate ranges are influenza $1 - 1.5$, smallpox $5 - 6$, chickenpox $7 - 12$, measles $12 - 18$ [19].

We assume that real epidemics do not exceed approximately 20% active cases at any point of epidemic duration, 80% of total infection cases, and R_0 is between 0 and 30. Then we suggest the following value ranges for each parameter:

- The population density should not go below 0.7 as that stops the epidemic ($d > 0.7$).
- Pareto parameter a can be chosen from values above 0 ($a > 0$).
- Pareto parameters α and β should be $\alpha > 1$ and $\beta > 2$ to prevent actors from

having connections with all other actors in the social network.

- Recovery time t_r should be roughly below 20 to ensure that the peak of the infection does not exceed 20% ($0 < t_r < 20$).

- Infection probability p_1 should approximately range from 0 to 0.1 and be lower than p_2 to not exceed selected thresholds ($0 < p_1 < 0.1$ and $p_1 \leq p_2$).

- Infection probability p_2 should roughly go up to 0.5 to make sure that the peak of the infection is within the desired range ($p_1 \leq p_2 < 0.5$).

Using some combinations of parameters from the suggested range might still result in peculiar epidemic spread outcomes. Thus, parameters have to be reviewed and adjusted according to generated location and social graphs to get meaningful results.

5 Epidemic control

In this section we look at various epidemic control measures and their effectiveness. The effectiveness of epidemic control is measured against baseline scenario when disease is spread without any controls. The relevant metrics to compare are the size of infection peak and total infection cases.

5.1 Lockdown

One way to control the spread of disease is enforcing a lockdown, which restricts movement within communities, cities, or countries. Lockdown effect can be achieved by removing edges from the location graph. The removal of edges ensures that actors can no longer use those edges to move between vertices, thus limiting actors' movements.

We have implemented three methods to remove edges from location graph: random, from hubs, and based on edge betweenness centrality. Random removal of edges represents closure of randomly selected roads and pathways. This is a simple and straightforward approach that does not require any information about the location graph or disease spread. Removal of edges from hubs means that edges are removed from 10% of vertices with highest degrees[2]. This approach corresponds to closure

[2]If 10% of vertices with highest degrees do not have sufficient number of edges then more

of roads and pathways from strongly interconnected locations like city center, workplaces, etc. Lastly, removal based on edge betweenness centrality means that edges with highest betweenness centrality are removed first, which represent the closure of roads and pathways that are frequently used for travel like highways or arterial roads. Here betweenness centrality score for the edge $e \in E$ is calculated as the sum of the fraction of all-pairs shortest paths that pass through e [10]:

$$c_B(e) = \sum_{s,t \in V} \frac{\sigma(s,t|e)}{\sigma(s,t)},$$

where V is the set of vertices, E is the set of edges, $\sigma(s,t)$ is the number of shortest (s,t)-paths, and $\sigma(s,t|e)$ is the number of those paths passing through edge e.

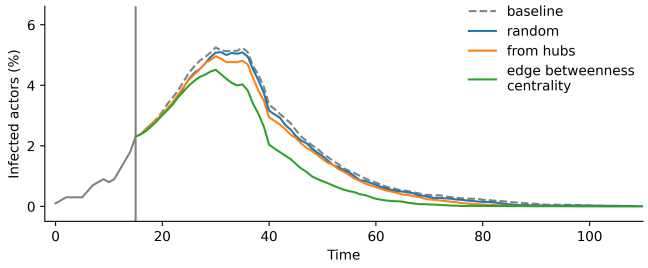

Figure 16: Simulated number of infected actors with and without lockdown control measures. The vertical line marks when lockdown was started. All lockdown approaches remove 10% of location graph edges. Model parameters are $p_1 = 0.01$, $p_2 = 0.1$, $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$, $\alpha = 1.1$, $\beta = 2.1$.

Figure 16 shows one realization example of a lockdown implementation. First, infection is spread without any controls until number of infected individuals reach a selected 2% threshold marked by vertical line. Then three different lockdown approaches are implemented. The baseline shows the continuation of epidemic spread without any controls, whereas blue, orange, and green lines show epidemic spread after randomly removing edges, removing edges from hubs, and removing edges based on edge betweenness centrality.

The overall effect of lockdown is showed in Figure 17. The baseline shows maximum infected actors and total infected actors when epidemic is spread without any controls. Three other curves show the three lockdown approaches. The effect of randomly removing location graph edges is twofold: removing less than approximately 30% of edges has almost no effect on simulation results as we can observe only mi-

vertices are added to remove selected number of edges.

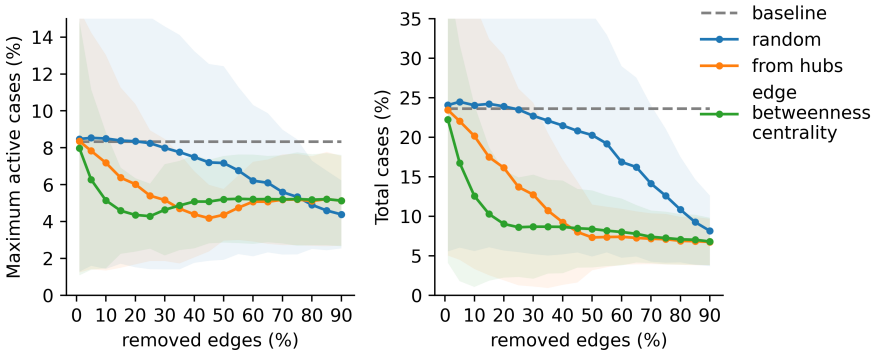

Figure 17: Relationship between lockdown and epidemic spread metrics. Model parameters are $p_1 = 0.01$, $p_2 = 0.1$, $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$, $\alpha = 1.1$, $\beta = 2.1$.

nor decrease or increase of infection spread, which is within standard deviation of the simulation results. Only when more than 30% of edges are removed epidemic starts to slow down. The second approach of removing edges from hubs is much more efficient as epidemic spread slows down after removing as few as 5% of edges. The third approach of removing edges based on betweenness centrality is the most efficient as the epidemic spread declines after removing only 1% of edges.

The effect on maximum active cases is not straightforward for edge removal from hubs and betweenness centrality approaches. The initial decrease in maximum value is followed by a minor increase and then stabilization at a general 51% level. This is a consequence of removing sufficient number of edges to divide location graph into smaller disconnected subgraphs. Actors that are in these subgraphs can no longer leave it, hence, disease spread is slightly accelerated in these "islands".

5.2 Social distancing

Social distancing is another infection spread prevention strategy. The World Health Organization (WHO) defines social distancing as keeping physical distance between people and avoiding gathering in large groups. Some examples of social distancing are keeping 1- or 2-meter distance between people, temporary closure of schools, workplaces, cancellation of festivals, sports events, etc. [73].

Similarly to lockdown, social distancing can be implemented by removing contacts between actors in the social network. The naive approach is to remove random edges from the social network. Removing edges from the social network means that people

are keeping distance from one another, hence, instead of having higher likelihood (p_2) to get infected from sick friends, family members or coworkers, they have a lower probability (p_1) to get infected.

Another implementation is removal of random attributes from the underlying bipartite graph. This approach also results in a reduction of contacts or edges in the social network. Contacts are removed not randomly but between actors that share the same removed attribute. In reality, removing attributes corresponds to closure of some common activities like schools, workplaces, sport clubs, and others.

We can also apply previous edge removal strategies like removing edges from hubs and removing edges based on edge betweenness centrality. However, these strategies do not have an intuitive interpretation for the social network.

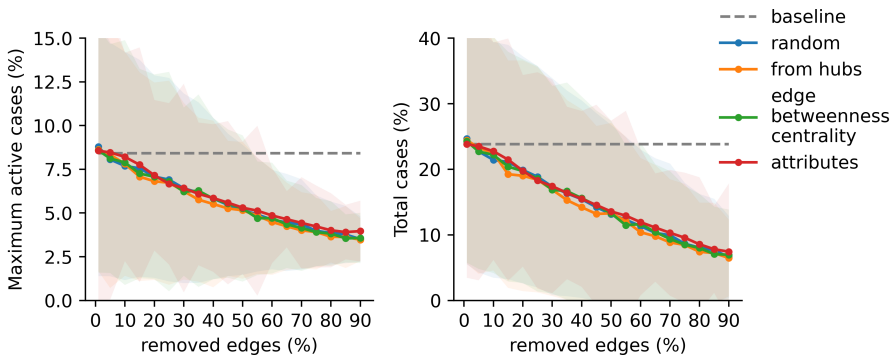

Figure 18: Relationship between social distancing and epidemic spread metrics. Model parameters are $p_1 = 0.01$, $p_2 = 0.1$, $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$, $\alpha = 1.1$, $\beta = 2.1$.

All four approaches have the same size effect for peak and total infection cases. That is because the number of edges in social network has stronger influence for disease spread compared to graph degree distribution. Hence, different edge removal strategies have the same effect as long as the same number of edges is removed.

5.3 General preventative measures

We understand the general preventative measures as a collection of personal protective measures and environmental measures. Personal protective measures are washing or sanitizing hands, wearing face masks, covering mouth when coughing or sneezing. Environmental measures are cleaning and disinfecting surfaces and objects,

ventilating rooms, etc. A collection of these preventative measures is implemented as scaling of infection probabilities, that is, following personal and environmental recommendations reduces both likelihoods to get infected from strangers (probability p_1) and from family members, friends, coworkers, etc. (probability p_2).

Figure 19 shows the effect of scaling infection probabilities. It can be observed that scaling probabilities is a very effective way to reduce epidemic spread. Reducing both p_1 and p_2 by 10% (equivalent to 0.9 factor) reduces peak from 8.6% to 7.5% and total cases from 24% to 21%.

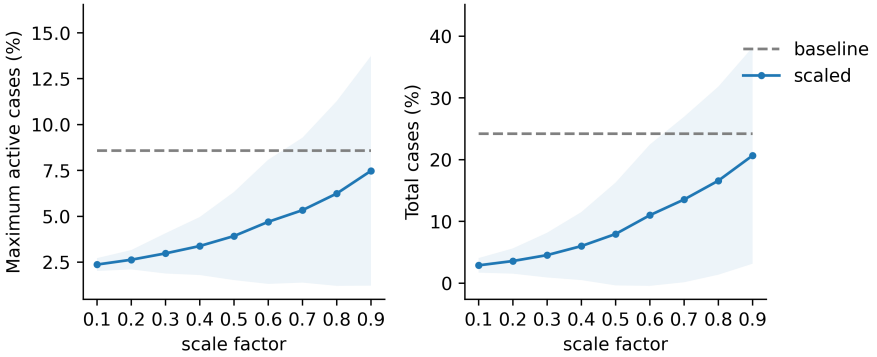

Figure 19: Relationship between general preventative measures and epidemic spread metrics. Model parameters are $p_1 = 0.01$, $p_2 = 0.1$, $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$, $\alpha = 1.1$, $\beta = 2.1$.

5.4 Vaccination

Vaccination has been the most popular measure for the prevention of infection spread. It is important to analyze the effect of vaccination on epidemic spread and identification of individuals that should be vaccinated first due to the high potential of disease spread.

Vaccination is implemented as an additional disease stage. Susceptible actors can be vaccinated and gain immunity to infection. After vaccination susceptible actors move to vaccination stage V and cannot be infected for the rest of the epidemic duration. The transitions between infection stages are showed in the figure below.

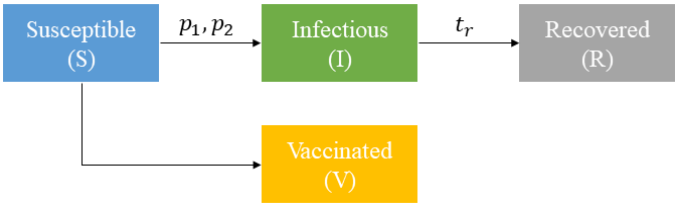

Figure 20: Diagram of susceptible, infectious, vaccinated, and recovered compartments of the model and the interactions of these compartments.

We explore the most popular vaccination strategies found in literature: random actors, highest degree actors, highest betweenness centrality actors. The simplest and naive approach is to vaccinate randomly selected individuals. The highest degree actor's strategy means vaccinating individuals with the highest degree in the social network, which corresponds to actors with the highest number of connections in the population. The highest betweenness centrality actor vaccination strategy means vaccinating individuals in order from the highest to lowest vertex betweenness centrality, which corresponds to vaccinating first actors that are like bridges in the social network. Betweenness centrality for vertex $v \in V$ is calculated in the following way [10]:

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)},$$

where V is the set of vertices, $\sigma(s,t)$ is the number of shortest (s,t)-paths, and $\sigma(s,t|v)$ is the number of those paths passing through vertex v other than s, t.

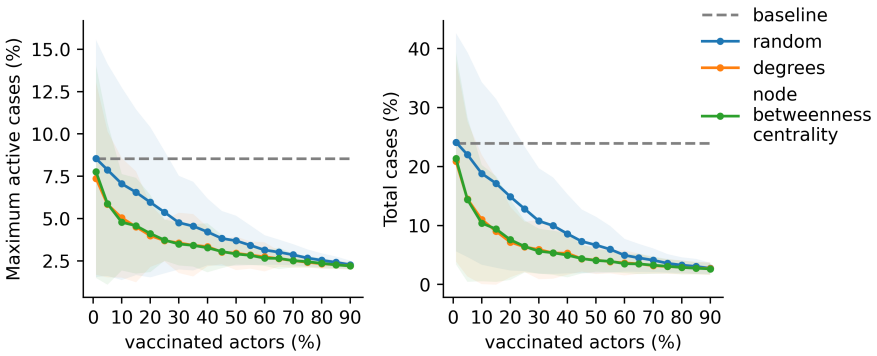

Figure 21: Relationship between vaccination and epidemic spread metrics. Model parameters are $p_1 = 0.01$, $p_2 = 0.1$, $t_r = 10$, 1000 actors, 500 location graph vertices, $a = 1.5$, $\alpha = 1.1$, $\beta = 2.1$.

In literature the most effective strategy is vaccinating actors with the highest be-

tweenness centrality score ([61, 62, 32, 14, 31] as cited in [43]). However, in our model vaccinating based on the highest degree and the highest betweenness centrality has the same effect because of the difference between the model structures. Articles often describe models with one network, where individuals are represented as nodes in a network, and the edges between the nodes represent the contacts along which an infection can spread. Since our model works differently, betweenness centrality approach is equivalently effective as the highest degree owing to the fact that vertices with the highest betweenness centrality are the ones with the highest number of shortest paths passing through them. This often coincides with the vertices with the highest degrees as they tend to have many shortest paths passing through them.

5.5 Comparison of control measures

In literature [21, 29, 52, 57, 58], comparison metrics for disease control measures are often model specific. Thus, to compare the efficiency of different control measures used in our model, we define a simple metric (inspired by statistical measures that compare areas under curves [64, 65]):

$$e = 1 - \frac{\sum_{i=1}^{n} c(x_i)}{bn} \left(\approx \frac{B - C}{B} \right),$$

where n is the number of data points in the measuring sample, $c(x_i)$ control measure outcome with control intensity x_i, and b is baseline simulation outcome. Metric e is roughly equal to ratio of difference between area under baseline curve (B) and area under the control measure outcome curve (C) to area under baseline curve (B).

Metric e represents the average efficiency of the control measure taking into account the intensity of control. For example, lockdown might achieve the same control efficiency as vaccination, however, it might require stronger control of movement compared to the share of actors that need to be vaccinated to achieve the same epidemic slow down effect.

Table 3 shows all epidemic control measures compared by metric e. For a reduction of epidemic peak, the most efficient approach is scaling infection probabilities followed by vaccination with betweenness centrality strategy. For reducing total infection

Control measure	**Approach**	e **for maximum**	e **for total**
Lockdown	random	0.166	0.200
Lockdown	from hubs	0.338	0.515
Lockdown	betweenness centrality	0.374	0.595
Social distancing	random	0.321	0.381
Social distancing	from hubs	0.340	0.402
Social distancing	betweenness centrality	0.327	0.385
Social distancing	attributes	0.328	0.396
General preventative measures	scale factor	**0.495**	**0.602**
Vaccination	random	0.366	0.454
Vaccination	degrees	0.434	0.546
Vaccination	betweenness centrality	0.446	0.554

Table 3: Comparison of epidemic control measures. Column "e for maximum" means that calculation of e was applied to the maximum number of active infection cases, and "e for total" means that calculation of e was applied to the total infection cases.

cases, the best control is also scaling of probabilities followed by lockdown with betweenness centrality strategy. Overall, the most efficient control strategy is scaling of infection probabilities followed by vaccination and lockdown. The least effective strategy is social distancing.

In general, the comparisons above are only theoretical. In a real-world epidemic scenario analysis to compare different disease control measures a comparable cost metric must be assigned to each of them. For example, reducing p_1 by 10% would cost much less than closing 10% of roads or vice versa. This introduces additional complexity and requires more data; thus, it was not feasible to do in the context of this thesis.

6 Comparison with the classical SIR model

In this section we compare our epidemic model with location and social networks against the classical SIR model characterized by a set of differential equations. For definition and description of the classical SIR model see subsection 2.1 Compartmental models and set of differential equations (1).

The comparison is carried out in the following way. First, we perform Monte Carlo simulations for the epidemic model with location and social networks. The proportions of susceptible, infected, and recovered individuals are calculated as mean

values of the Monte Carlo simulations. Then classical SIR model is fit to the outcome of the model with networks. The parameters are estimated using the least square method.

Figures 22 and 23 show two examples with comparison between classical and model with networks for proportion of actors in one of the three infection stages.

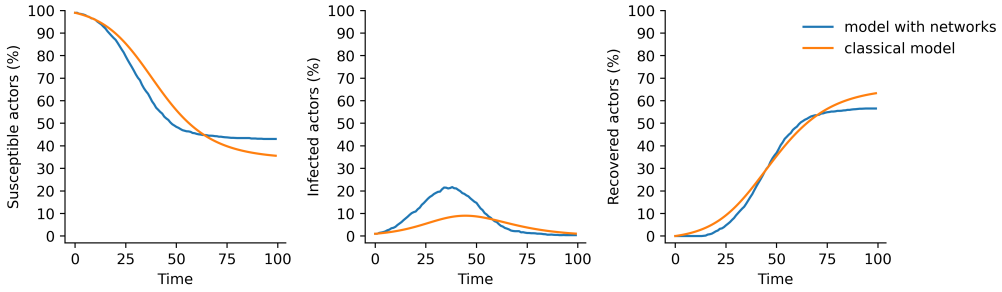

Figure 22: Comparison of proportion of susceptible, infected, and recovered individuals in classical SIR model and simulation-based model with two networks. The parameters used in model with networks are $p_1 = 0.01, p_2 = 0.2, t_r = 14$, 1000 actors starting simulation with 10 infected individuals.

When fitting classical SIR model β was estimated as 0.2109 and γ as 0.1306, which is equivalent to $d = 7.66$. The estimated time of infection is roughly half of the selected length of infection in the network model. It is difficult to compare β estimation as model with networks has two infection probabilities that have slightly different interpretation compared to β.

From figures it is visible that the set of differential equations manages to estimate the number of susceptible and recovered actors adequately. However, it cannot capture the number of infected individuals. The estimated infection spread speed is slower than what was simulated with networks as seen from steeper blue curve and flatter orange curve in Figure 22. Since the numbers of susceptible, infected, and recovered individuals are connected (in the form $N = S(t) + I(t) + R(t)$), it is natural that if classical SIR model cannot capture one of the disease compartments, in this example infected stage, then it would not fully capture the other compartments.

The second example results are showed in Figure 23 below.

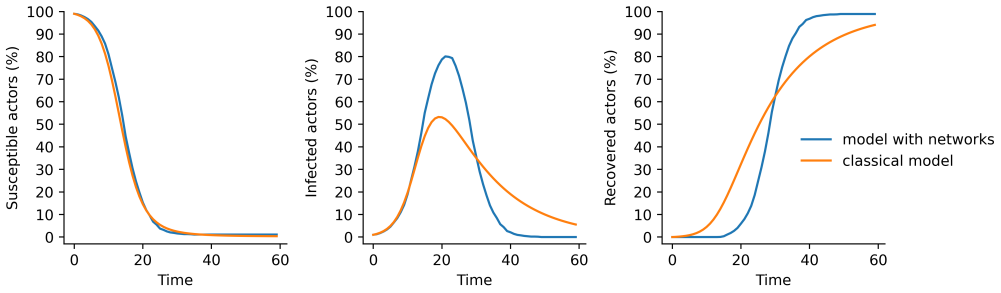

Figure 23: Comparison of proportion of susceptible, infected, and recovered individuals in classical SIR model and simulation-based model with two networks. The parameters used in model with networks are $p_1 = 0.1, p_2 = 0.2, t_r = 14$, 1000 actors starting simulation with 10 infected individuals.

Here β was estimated as 0.3897 and γ as 0.0658, which is equivalent to $d = 15.21$. The estimated γ is roughly the same as selected length of infection in the network model. In this example the proportion of susceptible individuals had a perfect match. Whereas the proportions of infected and recovered actors were matched inaccurately.

To evaluate the difference between the classical SIR model and our model with location and social networks we calculate normalized mean absolute error (NMAE). It is calculated as follows

$$NMAE = \frac{\sum_{t=1}^{n} |S_n(t) - S_c(t)|}{\sum_{t=1}^{n} S_c(t)} + \frac{\sum_{t=1}^{n} |I_n(t) - I_c(t)|}{\sum_{t=1}^{n} I_c(t)} + \frac{\sum_{t=1}^{n} |R_n(t) - R_c(t)|}{\sum_{t=1}^{n} R_c(t)},$$

where $S_n(t), I_n(t), R_n(t)$ are number of susceptible, infected, and recovered individuals at time t in the model with two networks, $S_c(t), I_c(t), R_c(t)$ is number of susceptible, infected, and recovered individuals at time t in the classical model, n is the number of time moments. The comparison results are showed in Figure 24.

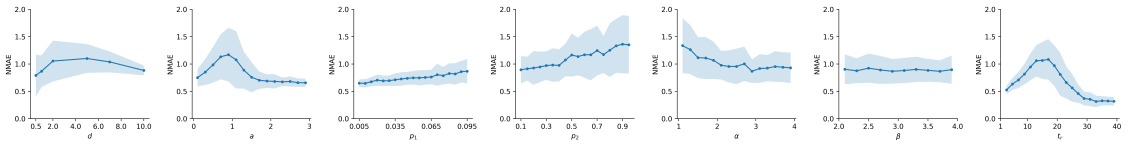

Figure 24: Comparison between the classical SIR model and simulation-based model with two networks in terms of NMAE. The NMAE is calculated for different model with networks parameter values starting from the left: population density d, location network parameter a, infection probabilities p_1 and p_2, social network parameters α and β, and recovery time t_r. The remaining parameters are $p_1 = 0.01$, $p_2 = 0.1$, $t_r = 10$, 1000 actors, 500 location graph vertices, $\alpha = 1.1$, $\beta = 2.1$. The curve shows results averaged over Monte Carlo simulations and colored range displays standard deviation from the mean value.

Overall, the set of differential equations manages to estimate one stage of infection quite well, mostly the number of susceptible or recovered individuals. However, after experimenting with different parameters, it seems that classical model cannot estimate all three stages of the infection in the same way as model with networks. This is further confirmed by the calculation of NMAE (see Figure 24), which is always above 0.3 for different values of model with networks parameters. Higher NMAE signifies a substantial difference between the number of susceptible, infected, and recovered individuals resulting from the classical model and model with two networks. This could be due to the fact that the SIR model characterized by the set of differential equations assumes homogeneous mixing of the population and thus creates a simplified view of the infected population. Whereas a model with location and social networks simulates individual behavior, which results in a more complex infection spread model.

7 Data fitting

In this section we describe how we fit our epidemic model to Covid-19 data, evaluate results and compare them against classical model and model with one network.

We obtained Covid-19 data for municipalities in Lithuania from Department of Statistics of Lithuania [20]. The Covid-19 data contains municipality population size and daily infection cases. Since SIR compartmental model can only describe one peak of epidemic, we select period from December 2021 to May 2022 that captures the highest peak of Covid-19 epidemic spread in Lithuania. We select two city municipalities - Vilnius and Šiauliai that visually have different disease spread dynamics. Other cities had very similar trends to the selected ones, for example, Kaunas had very similar trend to Vilnius, Klaipėda to Šiauliai etc.

The model parameters are estimated using the simplicial homology global optimization (shgo) algorithm with bounds. The shgo algorithm is a general-purpose optimization algorithm and it is suitable for any general class of low dimensional optimization problems [22]. It is the most applicable to global and derivative free optimization problems, especially science and engineering problems that are based on simulations or have a complex model structure. The algorithm requires only a black-box function input with an option to provide bounds, symmetry, or gradients.

It works best for problems with ten or fewer variables. This optimization algorithm is suitable for our model, since it is simulation based with quite a complex structure. In addition, we provide bounds to ensure that infection probabilities p_1 and p_2 are in the range $(0, 1)$.

For objective function we use the normalized mean squared error (NMSE):

$$NMSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} \hat{y}_i^{\,2}},$$

where y_i is observed value at time i, i.e., Covid-19 data, $\hat{y}_i$ is predicted value at time i, n is the number of time moments in the dataset.

We will compare model fit against model with one network and classical SIR model. For model with one network, we modify our model with two networks by removing social network and only keeping location network. The model with one network is fit to data in the same way as our model with two networks. The classical model is the SIR model defined by a set of differential equations described in subsection 2.1 Compartmental models and equations (1). The classical SIR model is fit using the Nelder-Mead algorithm and the same NMSE objective function.

Real Covid-19 data includes some epidemic control measures that are not taken into account in all three models. Hence, all data fitting results could be improved by introducing some epidemic control measures from section 5 Epidemic control. This would improve model precision but also would introduce additional complexity and require further parameter estimation.

To compare model fit we use previously described normalized mean squared error (NMSE), normalized mean absolute error (NMAE), bias, and maximum number of active infection cases (peak of the infection), i.e., the largest share of population that was in infected stage at the same time. The NMSE is widely used for model fitting and machine learning algorithms [69]. NMAE is another widely used measure. The difference between the two is that NMSE gives more importance to the most significant errors and hence it is more sensitive to outliers in the data, whereas NMAE gives the same importance to all errors. Thus, one big error can significantly deteriorate NMSE, whereas NMAE can be unaffected. We have also considered using mean absolute percentage error (MAPE), but it promotes very low estimation values

leading to consistent model underestimation.

We calculate NMAE and bias as follows:

$$NMAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{\sum_{i=1}^{n} \hat{y}_i},$$
$$bias = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i),$$

where y_i is observed value at time i, i.e., Covid-19 data, $\hat{y}_i$ is predicted value at time i, n is the number of time moments in the dataset.

7.1 Vilnius

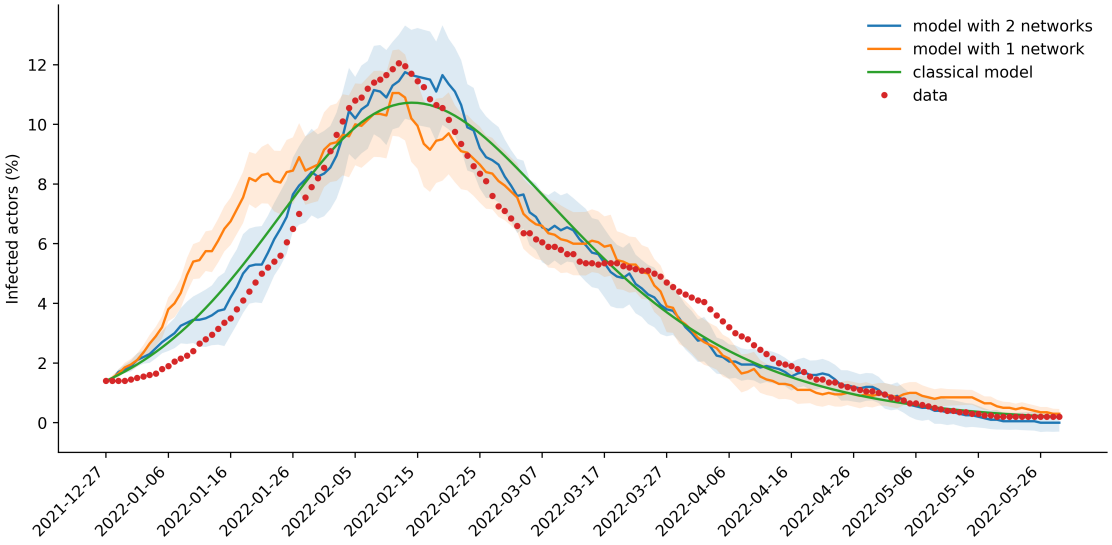

Figure 25: Model fitting to Covid-19 data for Vilnius city municipality. Blue curve - model with location and social graphs, orange curve - model with only location graph, green curve - the classical SIR model, red dots - Covid-19 data. The colored range displays standard deviation.

Classical model		**Model with one network**		**Model with two networks**		
β	γ	p_1	t_r	p_1	p_2	t_r
0.1698	0.1001	0.225	13	0.15	0.2996	13

Table 4: Estimated parameters.

Model	**NMSE**	**NMAE**	**bias**	**peak**
Covid-19 data	–	–	–	12.1%
Classical model	0.021	0.148	-2.955	10.7%
Model with one network	0.048	0.196	-5.981	11.1%
Model with two networks	**0.014**	**0.120**	**-2.286**	**11.8%**

Table 5: Model fit evaluation. NMSE is normalized mean squared error, NMAE - normalized mean absolute error, bias - the average error, peak - maximum number of active infection cases, i.e., the largest share of population that was in infected stage at the same time.

Figure 25 shows the results of proposed model with location and social networks, model with location network, and a classical SIR model, compared with the real data. Table 4 shows estimated parameters for all three models and Table 5 shows model fit measures.

Interestingly, all three models had similar estimated infection periods. Both models with networks had 13 days recovery time and classical SIR model had 10 days period ($d = 1/\gamma$). In model with only location network infection probability p_1 is naturally estimated at higher level than p_1 probability in model with two networks. This accounts for the removal of social network and possibility to be infected with the higher probability p_2. Classical model infection transmission parameter β cannot be directly compared to the other infection probabilities.

Model fitting results are quite similar between models. Our model with two networks has a slightly better fit compared to the other two as it has the lowest NMSE. The model with one network has the worst fit out of the three. Our model with two networks has the best fit due to the fact that the classical SIR model is not as good at capturing the peak of infection, whereas model with two networks captures the highest peak of the epidemic well. The model with one network has the worst fit in terms of NMSE and NMAE due to overestimation of infection cases in the initial phase of the epidemic.

7.2 Šiauliai

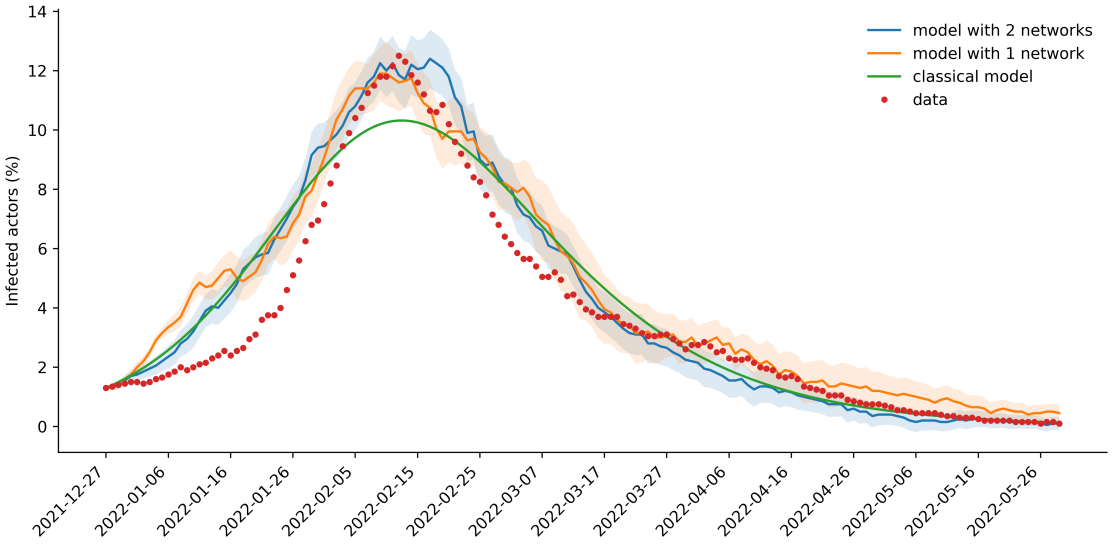

Figure 26: Model fitting to Covid-19 data for Šiauliai city municipality. Blue curve - model with location and social graphs, orange curve - model with only location graph, green curve - the classical SIR model, red dots - Covid-19 data. The colored range displays standard deviation.

Classical model		**Model with one network**		**Model with two networks**		
β	γ	p_1	t_r	p_1	p_2	t_r
0.1806	0.1077	0.225	13	0.075	0.3	13

Table 6: Estimated parameters.

Model	**NMSE**	**NMAE**	**bias**	**peak**
Covid-19 data	–	–	–	12.5%
Classical model	0.041	0.192	**-7.289**	10.3%
Model with one network	0.039	**0.182**	-15.065	11.9%
Model with two networks	**0.034**	0.185	-9.247	**12.4%**

Table 7: Model fit evaluation. NMSE is normalized mean squared error, NMAE - normalized mean absolute error, bias - the average error, peak - maximum number of active infection cases, i.e., the largest share of population that was in infected stage at the same time.

Figure 26 shows the comparison between proposed model with location and social networks, model with location network, classical SIR model and real data. Tables 6 and 7 show the estimated parameters and model fit measures for all three models.

Similarly to Vilnius case, the recovery time matches, and it is 13 days between model with two networks and one network. However, the classical model has shorter 9 days infection period. The p_1 probability in model with one network lies in between p_1 and p_2 values in model with two networks.

The model with location and social networks has slightly better fit based on NMSE. NMAE is lowest for the model with only location network, meaning that model with two networks is better at predicting average infected actor proportion, where model with one network is better at predicting the median infected actor proportion [69].

Like before, the classical model does not capture the peak of the infection and has a lower peak compared to data. Models with networks capture the peak well. However, all three models fail to capture the first few months of the epidemic spread by overshooting the number of infected actors.

To conclude, the new model with two networks is more realistic and manages to capture real Covid-19 data slightly better than the classical model. It can reflect more nuanced disease dynamics as compared to the classical SIR model. For example, it captures the peak of the infection more closely.

8 Conclusions

In this thesis, we have built the location and social networks, defined epidemic model simulation process, assessed the impact of various model parameters, introduced and assessed epidemic control measures, compared the new model with the classical SIR model, and fit the new model to the Covid-19 data for Vilnius and Šiauliai city municipalities.

The new model more closely reflects realistic epidemic characteristics and dynamics. For example, individuals are connected in the social network if they share common interests; individuals have a closer contact with their acquaintances and hence, have a higher chance to infect them. Furthermore, higher population density accelerates the infection spread resulting in higher peak and increased total cases.

The comparison with classical model and data fitting results show that the epidemic model with location and social networks is different from epidemic model with one

network and a classical SIR model defined by a set of differential equations. The novel approach combining two networks into one epidemic model reflects a more nuanced epidemic spread that cannot be captured by the classical model or model with only one network.

Concluding remarks:

1. The model with location and social networks is a more intuitive real-world representation. Although, some aspects of the model have a higher level of abstraction like the selected disease compartments (using susceptible-infectious-recovered disease stages) or the movement of actors (individuals moving along the location graph with random walk pattern). Nonetheless, the built model has sufficient level of detail and complexity to reflect real world epidemic spread and epidemic control measures.

2. The model simulations are computationally complex, and their outcomes can have high variance. In section 4 Influence of model parameters we show mean model simulation results and their standard deviation. It can be seen that for some of the infection characteristics standard deviation is higher. For example, the basic reproduction number (R_0) and epidemic duration results have higher variance.

3. The built model is substantially different from the classical SIR model characterized by a set of differential equations. This is because the classical SIR model is a simple representation of the disease spread in population, while a model with location and social networks is a more complex representation of disease spread.

4. For estimating real epidemic data, the model with location and social networks performs comparatively to the classical SIR model. In our analysis, the model with two networks showed slightly better fit to Covid-19 data compared to model with one network and classical model. One of the biggest obstacles experienced in data fitting was long computation times for simulation-based model with two networks.

5. The model could be used to study how location and social networks affect and

interact with disease spread characteristics. Also, it could be used to model real data and estimate disease progression through time. Lastly, due to model granularity and complexity, it could be used in scenario analysis. For example, it could be used when making projections about the disease spread and the effect of epidemic control measures.

In further study, the model could be improved with the following modifications to introduce additional dynamics:

1. Create a more complex location graph. One way to do that would be to improve the generation process of the random graph. For example, categorizing vertices in the location graph such as homes, supermarkets, workplaces, schools, social gathering places, etc. The vertex degree should correspond to the type of vertex, i.e., homes should have lower degrees compared to gathering places, schools, universities, workplaces. Another approach could be to base the location graph generation on real geographical data. For example, the location graph could be derived from city road and pathway maps.

2. A more intelligent walking pattern can be implemented. For example, actor movement can be modified by allocating a home vertex and limiting each actor's traveling distance from their home vertex.

3. More infection stages can be added. For example, exposed, infected without symptoms, quarantined, hospitalized, etc. Especially for modelling communicable diseases, the introduction of the exposed compartment would produce more realistic results.

4. Actors can be enriched with additional characteristics. For example, actors can be split by into age groups with differing infection probabilities depending on the age. Actors could also be categorized based on their behavior, like careful and reckless individuals.

5. For further research, the tradeoff between complexity of the model that improves model estimation results and computational complexity could be investigated. The current model is already quite complex and requires a comparatively long time to produce results. Further modifications that introduce

additional complexity would surely exacerbate computational complexity issues.

References

[1] Sidra Arshad, Shougeng Hu, and Badar Nadeem Ashraf. Zipf's law and city size distribution: A survey of the literature and future research agenda. *Physica A: Statistical Mechanics and its Applications*, 492:75–92, 2018.

[2] F. Ball. The threshold behaviour of epidemic models. *Journal of Applied Probability*, 20(2):227–241, 1983.

[3] F. Ball. A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. *Advances in Applied Probability*, 18(2):289–310, 1986.

[4] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016.

[5] Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1–3):1–101, 2011.

[6] F. M. Bass. A new product growth for model consumer durables. *Manage. Sci.*, 15:215–227, 1969.

[7] P. I. Bass and F. M. Bass. Diffusion of technology generations: A model of adoption and repeat sales. 2001.

[8] Mindaugas Bloznelis, Erhard Godehardt, Jerzy Jaworski, Valentas Kurauskas, and Katarzyna Rybarczyk. Recent progress in complex network analysis: Models of random intersection graphs. 48:69–78, 01 2015.

[9] P. Y. Boëlle, S. Ansart, A. Cori, and A. J. Valleron. Transmission parameters of the a/h1n1 (2009) influenza virus pandemic: a review. *Influenza and other respiratory viruses*, 5(5):306–316, 2011.

[10] Ulrik Brandes. On variants of shortest-path betweenness centrality and their

generic computation. *Social Networks*, 30(2):136–145, 2008.

[11] Fred Brauer. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling*, 2(2):113–127, 2017.

[12] Pizzuti C, Socievole A, Prasse B, and Van Mieghem P. Network-based prediction of covid-19 epidemic spreading in italy. *Appl Netw Sci*, 2020.

[13] Duncan S. Callaway, M. E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, 2000.

[14] Yiping Chen, Gerald Paul, Shlomo Havlin, Fredrik Liljeros, and H. Eugene Stanley. Finding a better immunization strategy. *Phys. Rev. Lett.*, 101:058701, Jul 2008.

[15] D. Daley and J. Gani. *Epidemic Modelling: An Introduction*. Cambridge University Press, 1999.

[16] F. Darabi Sahneh, C. Scoglio, and P. Van Mieghem. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Networking*, 21(5):1609–1620, 2013.

[17] P. L. Delamater, E. J. Street, T. F. Leslie, Y. T. Yang, and K. H. Jacobsen. Complexity of the basic reproduction number (r_0). *Emerging infectious diseases*, 25(1):1–4, 2019.

[18] A. Dharwadker and S. Pirzada. *Graph Theory*. Proceedings of the Institute of Mathematics. CreateSpace Independent Publishing Platform, 2011.

[19] B. Dhungel, M. S. Rahman, M. M. Rahman, A. K. C. Bhandari, P. M. Le, N. A. Biva, and S. Gilmour. Reliability of early estimates of the basic reproduction number of covid-19: A systematic review and meta-analysis. *International journal of environmental research and public health*, 19(18), 2022.

[20] Valstybės duomenų agentūra. Covid-19 statistika, 2024.

[21] Mohamed Elhia, Mostafa Rachik, and EL Habib Benlahmar. Optimal control

of an sir model with delay in state and control variables. *ISRN Biomathematics*, 2013:7 pages, 01 2013.

[22] Stefan Endres and Carl Sandrock. shgo documentation.

[23] P. Erdös and A. & Rényi. On random graphs. *Publicationes Mathematicae*, 1959.

[24] P. Erdös and A. & Rényi. On the evolution of random graphs. *Pub. Math. Inst. Hung. Acad. Science*, 1960.

[25] P. Erdös and A. & Rényi. On the strengths of connectedness of a random graph. *Acta Math. Scientiae Hung*, 1961.

[26] Institute for Health Metrics and Evaluation. Covid-19 results briefing, 2022.

[27] Piper Fowler-Wright. Degree sequences & the graph realisation problem, 2019.

[28] Macdonald G. The epidemiology and control of malaria. *Oxford University Press*, 1957.

[29] Elena Gubar, Vladislav Taynitskiy, Denis Fedyanin, and Ilya Petrov. Quarantine and vaccination in hierarchical epidemic model. *Mathematics*, 11(6), 2023.

[30] Sunetra Gupta, Roy M. Anderson, and Robert M. May. Networks of sexual contacts, aids. 3:807–818, 12 1989.

[31] Petter Holme, Beom Jun Kim, Chang No Yoon, and Seung Kee Han. Attack vulnerability of complex networks. *Phys. Rev. E*, 65:056109, May 2002.

[32] Laurent Hébert-Dufresne, Antoine Allard, Jean-Gabriel Young, and Louis J. Dubé. Global efficiency of local immunization on complex networks. *Scientific Reports*, 3:2171, 2013.

[33] Szapudi I. Heterogeneity in sir epidemics modeling: superspreaders and herd immunity. *Appl Netw Sci*, 2020.

[34] Keeling Matt J and Eames Ken T.D. Networks and epidemic models. *J. R.*

Soc. Interface, 2:295–307, 2005.

[35] John A. Jacquez, Carl P. Simon, James Koopman, Lisa Sattenspiel, and Timothy Perry. Modeling and analyzing hiv transmission: The effect of contact patterns. *Mathematical Biosciences*, 92 (2):119–99, 1988.

[36] Bin Jiang, Junjun Yin, and Qingling Liu. Zipf's law for all the natural cities around the world. *International Journal of Geographical Information Science*, 29, 02 2015.

[37] Heath Kelly, Heidi A. Peck, Karen L. Laurie, Peng Wu, Hiroshi Nishiura, and Benjamin J. Cowling. The age-specific cumulative incidence of infection with pandemic influenza h1n1 2009 was similar in various countries prior to vaccination. *PLOS ONE*, 6(8):1–9, 08 2011.

[38] William Ogilvy Kermack, A. G. McKendrick, and Gilbert Thomas Walker. 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.

[39] A.S. annd J.J. Potterat Klovdahl, D.E. Woodhouse, J.B. Muth, S.Q. Muth, and W.W. Darrow. Social networks and infectious disease: The colorado springs study. *Social Science & Medicine*, 38 (1):79–88, 1994.

[40] A. A. Lashari. Stochastic epidemics on random networks. *PhD dissertation, Department of Mathematics, Stockholm University*, 2019.

[41] Eric Lehman, F. Thomson Leighton, and Albert R Meyer. Mathematics for computer science. 2010.

[42] S. Leinhardt. *Social networks: a developing paradigm*. New York: Academic Press, 1977.

[43] Tomer Lev and Erez Shmueli. State-based targeted vaccination. *Applied Network Science*, 6(6), 2021.

[44] T.G. Lewis. *Network Science: Theory and Applications*. Wiley, 2009.

[45] Ira M. Longini. A mathematical model for predicting the geographic spread of new infectious agents. *Mathematical Biosciences*, 90 (1-2):367–83, 1988.

[46] P. Maheshwari and R. Albert. Network model and analysis of the spread of covid-19 with social distancing. *Appl Netw Sci*, 2020.

[47] Pandemic Science Maps. Epidemiology models: A history. 06 2020.

[48] M. Martini, V. Gazzaniga, N. L. Bragazzi, and I. Barberis. The spanish influenza pandemic: a lesson from history 100 years after 1918. *Journal of preventive medicine and hygiene, 60(1), E64–E67*, 2019.

[49] Robert M. May and Roy M. Anderson. Spatial heterogeneity and the design of immunization programs. *Mathematical Biosciences*, 72 (1):83–111, 1984.

[50] May Robert Mccredie and Anderson Roy Malcolm. The transmission dynamics of human immunodeficiency virus (hiv). *Phil. Trans. R. Soc. Lond.*, 1988.

[51] J. A. J. Metz. The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections. *Acta Biotheoretica*, 27:75–123, 1978.

[52] Erivelton G. Nepomuceno, Márcia L. C. Peixoto, Márcio J. Lacerda, Andriana S. L. O. Campanharo, Ricardo H. C. Takahashi, and Luis A. Aguirre. Application of optimal control of infectious diseases in a model-free scenario. *SN computer science*, 2(5), 2021.

[53] M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 2002.

[54] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64 (2), 2001.

[55] The Editors of Encyclopaedia. Encyclopedia britannica, plague.

[56] R. Patil, R. Dave, and H. et al. Patel. Assessing the interplay between travel patterns and sars-cov-2 outbreak in realistic urban setting. *Appl Netw Sci*, 2021.

[57] Alessandro Ramponi and Maria Elisabetta Tessitore. Optimal social and vaccination control in the svir epidemic model. *Mathematics*, 12(7), 2024.

[58] Aashiq Reza, Md Billah, and Sharmin Shanta. Effect of quarantine and vaccination in a pandemic situation: a mathematical modelling approach. *Journal of Mathematical Analysis and Modeling*, 2:77–87, 12 2021.

[59] S. Roy, R. Dutta, and P. Ghosh. Towards dynamic lockdown strategies controlling pandemic spread under healthcare resource budget. *Appl Netw Sci*, 2021.

[60] Anne-Sophie Ruget, Gianluigi Rossi, P. Theo Pepler, Gaël Beaunée, Christopher J. Banks, Jessica Enright, and Rowland R. Kao. Multi-species temporal network of livestock movements for disease spread. *Applied Network Science*, 6(15), 2021.

[61] Marcel Salathé and James H. Jones. Dynamics and control of diseases in networks with community structure. *PLOS Computational Biology*, 6(4):1–11, 04 2010.

[62] C. M. Schneider, T. Mihaljev, and H. J. Herrmann. Inverse targeting —an effective immunization strategy. *EPL (Europhysics Letters)*, 98(4):46002, May 2012.

[63] J. Scott. *Social network analysis: a handbook*. London: SAGE Publications, 1991.

[64] Denis A. Shah, Erick D. De Wolf, Pierce A. Paul, and Laurence V. Madden. Accuracy in the prediction of disease epidemics when ensembling simple but highly correlated models. *PLOS Computational Biology*, 17(3):1–23, 03 2021.

[65] Shima Shamkhali Chenar and Zhiqiang Deng. Development of genetic programming-based model for predicting oyster norovirus outbreak risks. *Water research*, 128:20–37, 10 2017.

[66] R. Solomonoff and A. Rapoport. Connectivity of random nets. *Bull. Math. Biophys*, 13:107–117, 1951.

[67] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 1967.

[68] Remco van der Hofstad. *Random Graphs and Complex Networks*, volume 2. Cambridge University Press. Expected publication date February 2024.

[69] Nicolas Vandeput. *Data Science for Supply Chain Forecasting*. 03 2021.

[70] Emilia Vynnycky, Amy Trindall, and Punam Mangtani. Estimates of the reproduction numbers of spanish influenza using morbidity data. *International Journal of Epidemiology*, 36(4):881–889, 05 2007.

[71] S. Wasserman and K. Faust. *Social Network Analysis*. Cambridge University Press, 1994.

[72] D. J. Watts, R. Muhamad, D. C. Medina, and P. S. Dodds. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences*, 102 (32), 2005.

[73] World Health Organization (WHO). Risk communication and community engagement guidance for physical and social distancing.

[74] Worldometer. Covid-19 statistics, 2024.

[75] Dezső Z and Barabási AL. Halting viruses in scale-free networks. *Phys Rev E*, 65(5), 2002.