h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Darmstadt University of Applied Sciences

Department of Mathematics
and Natural Sciences
& Computer Science

Graph-based risk modeling for infectious diseases in social systems

Thesis for obtaining the academic degree
Master of Science (M. Sc.)
in the program Data Science

submitted by

Maximilian Stäbler

Examiner:	Prof. Dr. Brucherseifer
Second Examiner:	Prof. Dr. Jahn

Grant date:	16. November 2020
Submission date:	03. May 2021

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder noch nicht veröffentlichten Quellen entnommen sind, sind als solche kenntlich gemacht. Die Zeichnungen oder Abbildungen in dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechenden Quellennachweis versehen. Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

Declaration on oath

I hereby certify that I have written this thesis independently and have not used any sources other than those listed in the bibliography. All passages that are taken verbatim or in spirit from published or as yet unpublished or not yet published sources are marked as such. The drawings or illustrations in this thesis have been made by myself or are provided with a corresponding source reference provided. This work has not been submitted in the same or similar form to any other examination authority.

Bisingen, 27. April 

Maximilian Stäbler

KURZFASSUNG

Die COVID-19-Pandemie hat das gesellschaftliche Leben der Menschen seit dem Ausbruch im Januar 2020 verändert. Der Ausbreitung der Krankheit durch Aerosole von infizierten Personen wurde von Regierungen weltweit entgegengewirkt. Unternehmen müssen interne Prozesse anpassen, was zu Kapazitätseinschränkungen führen kann. In dieser Arbeit wird ein Framework vorgestellt, mit welchem das Infektionsrisiko in einem sozialen Teilsystem, wie beispielsweise einem Unternehmen, bestimmt werden kann. Grundlage dafür sind Interaktionsdaten der Personen des Teilsystems. Das Interaktionsprotokoll dieser Arbeit wurde über sechs Wochen in einer Firma in Italien mit Hilfe von Bluetooth Sensoren aufgezeichnet. Unter Verwendung von arbeitsplatzspezifischen Reproduktionszahlen und Infektionszeiträumen aus der Literatur, werden systemspezifische Infektionsdynamik-Metriken für *SARS-CoV-2*, *SARS-CoV-2-B.1.1.7* und *Influenza* berechnet und verglichen. Die Ergebnisse zeigen, dass für das betrachtete soziale System *SARS-CoV-2-B.1.1.7* ein 2,6-fach höheres Infektionsrisiko pro sozialer Interaktion aufweist als die ursprüngliche Variante. Außerdem führt *SARS-CoV-2-B.1.1.7* zu 3,4-mal so vielen Sekundärfällen (920) wie *SARS-CoV-2* (270), wenn keine Gegenmaßnahmen ergriffen werden. Soziale Distanzierung erweist sich als wirksame Gegenmaßnahme für das betrachtete soziale Subsystem, welche eine Reduktion der Sekundärfälle der britischen Mutation auf 360 und für *SARS-CoV-2* auf 69 ermöglicht. Es konnte auch gezeigt werden, dass gezielte Gegenmaßnahmen, die auf topologischen Netzwerkeigenschaften für einen kleinen Teil der Individuen innerhalb des Systems basieren, die Anzahl an Ereignissen bei welchen eine Person viele weitere Personen auf einmal infiziert um 25% reduziert werden können, indem für 15% der Personen im betrachteten sozialen Untersystem soziale Distanzierungsmaßnahmen eingeführt werden (bezogen auf *SARS-CoV-2-B.1.1.7*). Das Framework kann für jede Infektionskrankheit verwendet werden, die durch soziale Interaktionen übertragen wird. Es ermöglicht Entscheidungsträgern, verschiedene Interventionen zu bewerten, soziale Strukturen besser zu verstehen und Individuen innerhalb des sozialen Subsystems zu identifizieren, die besonders gefährdet sind oder potentiell viele weitere Personen anstecken.

Schlagworte — COVID-19; Infection Model; Social Interaction; Social Network Analysis.

Abstract

The COVID-19 pandemic has changed people's social lives since the outbreak in January 2020. The transmission of the disease through aerosols from infected individuals has been counteracted by governments worldwide. Companies have to adapt internal processes, which can lead to reduced capacity. This thesis presents a framework for determining the risk of infection in a social subsystem, such as a company. The basis for this is interaction data of the persons of the subsystem. The interaction protocol of this work was recorded over six weeks in a company in Italy using Bluetooth sensors. System-specific infection dynamics metrics for *SARS-CoV-2*, *SARS-CoV-2-B.1.1.7*, and *Influenza* are calculated and compared using specific workplace reproduction numbers and infection periods reported in the literature. The results show that for the social system considered, *SARS-CoV-2-B.1.1.7* has a 2.6 times higher risk of infection per social interaction than the original variant. Moreover, *SARS-CoV-2-B.1.1.7* leads to 3.4 times as many secondary cases (920) as *SARS-CoV-2* (270) if no countermeasures are taken. Social distancing turns out to be an effective countermeasure for the social subsystem under consideration, allowing a reduction of secondary cases of the British mutation to 360 and for *SARS-CoV-2* to 69. It was also shown that targeted countermeasures based on topological network properties for a small fraction of individuals within the system can reduce the number of events in which an individual infects many more individuals at once by 25% by introducing social distancing measures for 15% of the individuals in the social subsystem under consideration (related to *SARS-CoV-2-B.1.1.7*). The framework can be used for any infectious disease transmitted through social interactions. It allows decision makers to evaluate different interventions, better understand social structures, and identify individuals within the social subsystem who are particularly at risk or transmitting infection.

Keywords — COVID-19; Risk model; Infection Model; Social Interaction; Social Network Analysis.

Contents

List of figures/tables

Figures

Tables

Code

Abbreviations

ABM	Agent-Based-Model
API	Application Programming Interface
DLR	Deutsches Zentrum für Luft- und Raumfahrt e.V. (German Aerospace Center)
HIV	Human-Immunodeficiency-Viruses
IQR	interquartile range
POI	Point of Interest
ResTAat	Infection Resilience by Targeted Action against Transmission
RKI	Robert-Koch-Institut
SARS	Severe-Acute-Respiratory-Syndrom
SARS-CoV-2	Severe-Acute-Respiratory-Syndrom-Coronavirus-2
SARS-CoV-2-B.1.1.7	Severe-Acute-Respiratory-Syndrom-Coronavirus-2 variant, first identified in the United Kingdom
SD	System dynamics
SNA	Social Network Analysis
SS	Super-Spreader
SSE	Super-Spreading-Event
UID	Unique identifier
WHO	World Health Organisation

1 Introduction

According to the World Health Organisation (WHO), diseases are becoming a current major international issue, especially infectious diseases that threaten health, economy and security [SV11]. As of December 2020, lower respiratory infections remained the world's most deadly communicable disease, ranked as the 4th leading cause of death. An example of a lower respiratory infection spreading rapidly worldwide is the Severe-Acute-Respiratory-Syndrom-Coronavirus-2 (SARS-CoV-2).
On December 31, 2019, the WHO China Country Office was informed of cases of pneumonia of unknown etiology detected in Wuhan, a metropolis of one million people in Hubei province. A novel coronavirus (SARS-CoV-2) was identified as the causative virus by Chinese authorities on January 7 2020. The original site of infection was the Wuhan wholesale fish and seafood market, from where the virus spread first to neighboring countries and then nearly around the world within a few weeks. The cumulative number of confirmed SARS-CoV-2 infections is more than 114.1 million worldwide by March 1, 2021. The number of coronavirus-related deaths rose to more than 2.5 million by that date [DDG20]. The SARS-CoV-2 pandemic has not only changed the social lives of people around the world [Cha+20a] [Sun+20] but also overwhelmed many health systems due to many life-threatening infections [And+20]. While 2020 was a challenging year, 2021 looks to be difficult with the emergence of multiple variants of SARS-CoV-2. The race to vaccinate the world will need to respond to the pathogen's constant evolution to evade immunity [Fon+21]. The WHO's growing concerns with epidemiological threats have two main reasons, according to a report [Org07] by the organization: First, migrations, increasing antimicrobial drug resistance, and health system failures continue to thwart implemented intervention plans. And secondly, the interconnectivity of people around the world. While problems such as antibiotic resistance or infrastructural problems in the health sector can only be solved in the long term and through structural changes, governments around the world have focused on containing the spread of the virus through general restrictions such as lockdowns and quarantine. Despite their efficacy, large-scale quarantine and population-wide lockdown strategies are far from optimal, and interventions at smaller scale, selectively targeting individuals at higher risk of spreading the disease, are more desirable [Cen+21].

Epidemiological models are often used by governments to help them decide on further measures to combat the virus. The aim of these models is to estimate the further course of infection, considering the measures taken [Fer+20a] [Hel+20] [RWC20] [ZC20].

In the current Corona pandemic, most of the models used by governments are mainly based on the *SIR* model introduced by *Kermack et al.* [KM27] or one of its further developments. *SIR* model (**S**usceptible-**I**nfected-**R**emoved model) is the term used in mathematical epidemiology to describe the spread of infectious diseases. Individuals in a given community are classified into one of three compartments, and constant rates describe the transfer of individuals between these groups. Models of the *SIR* type are widely accepted in epidemiology because of their simplicity and comprehensibility [Küh+20]. For example the model used by *Kühn et al.* [Küh+20] to simulate the SARS-CoV-2 outbreak in Germany is based on the *SIR* model. This model has been taken into account by the German Federal Government when deciding on further measures [RM21]. These models are used when making predictions about infection trajectories for large populations (social systems) with many different subsystems about which detailed information is not necessarily available [Fer+20b] [SPN20] [FF20]. An alternative to compartmental models is Agent-Based-Model (ABM). This framework allows to define behaviors at the individual and societal levels, describe the characteristics of the pathogen, and simulate the evolution of the infectious disease on a synthetic population [Eub+04]. However, this requires a great amount of detailed data about the population. Furthermore building, testing, and refining such models is time-consuming and rarely possible during deployment in the fight against an ongoing epidemic [Ven+18].
It is evident that different methods are currently being used to model the course of infection of the SARS-CoV-2 pandemic. The ever-changing threat posed by the virus through mutations means that some aspects of the pandemic may persist in the medium or long term [RWC20]. It is therefore necessary to make social and economic processes resilient in order to return to the normal life before the pandemic.
In addition to the health, environmental, and social challenges facing humanity, the coronavirus outbreak is disrupting the global economy. The lockdown measures and distance regulations imposed have interfered with industrial processes to such an extent that companies in various industries have had to close down for extended periods of time. Although many companies were able to maintain internal processes through home office, such a regulation is not possible in industries such as manufacturing, where employees must be

physically on site. For these companies it is necessary to find a solution with which the internal processes can be continued with minimal risk of infection to the workforce. An understanding of the transmission routes of infectious disease is therefore required.

According to recent research results by Robert-Koch-Institut (RKI) [Haa20] and *Zhang et al.* [Zha+20], respiratory ingestion of virus-containing liquid particles is seen as the main route of transmission for SARS-CoV-2. Hence, social interaction data, which describe the contact patterns of individual persons within the system under consideration, could be suited for modeling the infection course of SARS-CoV-2.
The dataset used in the thesis was collected through body-worn Bluetooth tags and include all social interactions between two individuals that occurred less than 1.5 meters apart and lasted longer than 15 seconds. The definition of these thresholds is consistent with the previously mentioned research by *Haas et al.* [Haa20] and *Zhang et al.* [Zha+20]. In addition, the work of *Jayaweera et al.* and *Kriegel et al.* provides evidence that SARS-CoV-2 has a realistic chance of surviving in air for a period of several hours [Jay+20]. According to the RKI, prolonged exposure in small or poorly ventilated spaces can increase aerosol transmission even beyond a distance of two meters, making the total duration of contact between two people a more important parameter than the exact distance during the interaction. For this reason the assumption is that the number of detected proximity contacts between two individuals is a surrogate for how long they have been in close proximity. This detailed dataset on individual person-to-person connections within a social subsystem was collected as part of the Infection Resilience by Targeted Action against Transmission (ResTAat) project at the Deutsches Zentrum für Luft- und Raumfahrt e.V. (German Aerospace Center) (DLR). It forms the basis for an investigations in a network graph, that maps all social interactions between different individuals. Within this graph, the individuals of the social subgroup under consideration are represented as nodes and the connections between these particular individuals are represented as edges between the various nodes [WF94] (an example of such a graph is shown in figure 1.1). Paremeters such as the frequency of social interaction can be stored as a value on the respective edge. What exactly is counted as social interaction must be defined in each research context. Valuable information can be derived from the topology of the resulting graph alone. For example, Figure 1.1 represents the correspondence between different scientists. In addition to the frequency of correspondence between researchers, which can be read from the thickness of the edge, the size of the individual nodes can also be used to

FIG. 1.1: Example of a social network graph depicting the collaboration of selected scientists between 1922 and 1930. In total, the network contains 887 nodes and 10363 edges. The thickness of the edges indicates how often the scientists collaborated with each other and the size of the node indicates how many colleagues each scientist collaborated with in total. The color indicates the membership of a scientific network (white = member, blue = no member). Figure according to *Grandjean* [Gra14]

identify which scientists have received or sent the most letters. Such simple information is very important for evaluating the significance of individual network nodes within the network and, depending on the context, provides approaches for further investigation. For example, if the graph in Figure 1.1 were examined in the context of an infectious disease and the edges represented personal meetings of scientists, the size of each node could be seen as a coarse indicator of the individual's risk of infection, as increasing numbers of social interactions increase the likelihood of infection of a disease transmitted through social interactions.

1.1 Aim of the thesis

This approach is taken up in the context of this thesis. The central research question of this thesis is how a protocol of social interactions within a complex social (sub)system can be used to calculate the risk of infection for an infectious disease that is transmissible through social interactions within that system. Countermeasures to contain the disease are then derived from this framework. Wearing masks and social distancing are considered as countermeasures. Both measures

are integrated into the framework. It is also investigated whether characteristics of contacts, such as duration of contacts between two individuals, provide information about the likelihood of infection and whether the topological properties of the resulting network graph allow inferences about infection dynamics. Using a company as an example, this could isolate individuals particularly at risk of infection before they either become infected themselves or infect others. Another goal is to generalize the framework proposed in this work so that it can be used to model other social interaction-transmitted diseases in other social subsystems. To enable such a general application of the model, individual social contacts and social subsystems need to be defined in the context of this work. According to Machens et al. [Mac+13], the establishment of countermeasures is in social subsystems is relevant to the containment and long-term control of socially transmitted diseases.

The findings on the course of infection and the countermeasures derived from them in this thesis relate exclusively to the social subsystem under consideration. Inference to other systems or scalability to, for example, the population of an entire country need to be validated. Likewise, site-specific parameters of social interaction such as ventilation or activity at the respective site are not taken into account in the context of this thesis for calculating the risk of infection.

1.2 Agenda

This thesis is divided into six chapters. After the introduction, Chapter 2 first presents the methods used in the field of epidemic modeling and then classifies them based on their application in research. After presenting the results of related research approaches, Chapter 3 introduces the dataset used with its specific characteristics. Based on these data and previous research in the field, the infection model is then derived. In addition, this section presents possible mitigation intervention options and their consideration in the model. Chapter 4 then describes all the experiments and simulations performed with the respective software implementations before all results are described, compared with the literature and discussed in Chapter 5. Finally, the results obtained are interpreted and, based on this, a conclusion is drawn for this thesis and further research approaches in Chapter 6.

2 Theoretical Background

The motivation of this chapter is to demonstrate the need for models to determine the risk of infection in a population or a specific part of that population. Since this thesis is focused on infectious diseases transmitted by proximity contacts, section 2.1 first defines the basic concepts and objectives of SNA. It is shown what SNA can be used for and which prerequisites have to be fulfilled. In section 2.2, different approaches for modeling a pandemic are presented. In particular, the use cases for each of these models and their advantages and disadvantages are described. Subsequently, the introduced models are discussed 2.3. Section 2.4 concludes the chapter by presenting similar research. In detail, approaches based on social interaction data within complex systems and used for modeling infectious diseases are considered here.

2.1 Social Network Analysis

Interest in social network graphs, such as those shown in Figure 1.1, has increased significantly over the past several years [Eir+18]. To keep terminologies and definitions consistent, this section gives an introduction to graph theory in a social context.

2.1.1 Definition of the subject

Nowadays the data generated from many real world applications are represented as a network of interconnected objects. The main objective is to extract more information than the traditional way of investigating independent objects. Of course, it increases the complexity of handling data as well. One of the major class of data networks is social networks. A social network can be constructed from relational data and can be defined as a set of social entities, such as people, groups, and organizations, with some relationships or interactions between them [Tab+18]. Examples of social networks are given in Table 2.1. Graph theory is the branch of mathematics devoted to the study of graphs and networks. Graphs and networks are defined by a set of vertices V and a set E of relations between the vertices. The simplest relation is an edge defined as a pair of vertices (a,b) with $a \in V$ and $b \in V$. A graph consisting of vertices and edges is called a simple graph or simply a graph. Another relation is an arc, defined as a pair of

Example	Application
Follower networks	Instagram, Facebook, Twitter etc.
Electronic interaction networks	E-Mails, Phone calls, Whatsapp, Snapchat etc.
Co-authorship networks	Science direct, Nature, NEJM etc.

TAB. 2.1: Examples of social networks. In all networks, individuals represent the nodes and the connection between them represents the respective relationship.

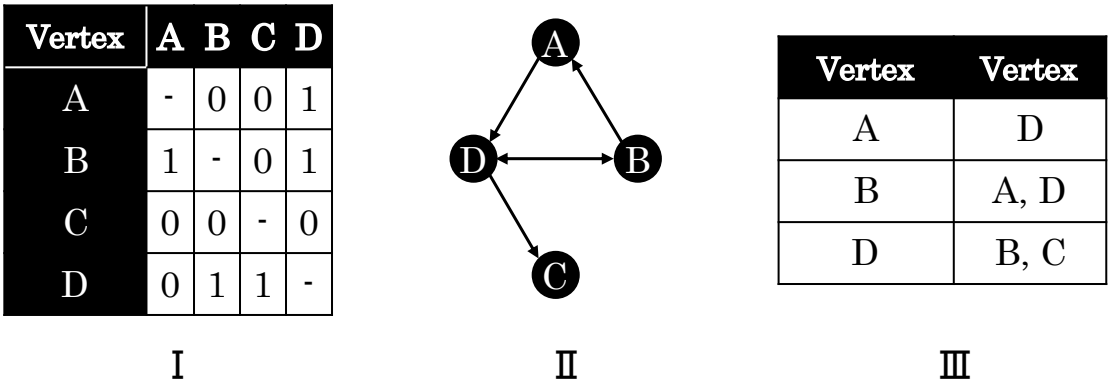

FIG. 2.1: A directed and unweighted graph G (*II*) represented by an adjacency matrix (*I*) and an adjacency list (*III*). Directed links are represented by an arrow and indicate from whom the link originates compared to undirected edges. Figure according to *Tabassum et al.* [Tab+18].

vertices with a direction $a \to b$ and $a \in V, b \in V$. A graph consisting of vertices and arcs is called a directed graph or digraph.

Figure 2.1 shows an example of a small directed graph. The edges of a directed graph can only be traversed in one direction and are often represented as arrows. Directed edges often are used to model asymmetric relations and relations such as "*depends on*", "*implies*", "*must be performed before*" or "*is better than*". Undirected graphs (with undirected edges), on the other hand, can only model symmetric relationships and relations such as "*are friends*", "*are neighbors*", or "*are connected*". Besides the visual representation of the graph Figure 2.1 *II*, two notation options are given in *I* and *III*. Figure part *I* shows the adjacency matrix of the graph. This matrix contains the information which nodes of the graph are connected by an edge. It has a row and a column for each node, resulting in an $n * n$ matrix for n vertices. The list in Figure 2.1 *III*, contains a set of all neighbors (in undirected graphs) or successors (in directed graphs) for each vertex in G.

A network is a graph with capacities assigned to the relationships between vertices. For example a weighted graph $G = (V, E)$ is attributed by a function w

that assigns a weight $w(e)$, typically $w(e) > 0$, to each edge $e \in E$. In a social network perspective this $w(e)$ could be a discrete or continuous attribute, for instance the total contact time or the number of emails written.
In the graphical representation of networks, the values of the weights w of the graph G are distinguished by line weight or value, line sign or line type [de 09]. Examples of value differentials in a social context include intensity, frequency, valence, or type of social relationship. The set of possible relationships per node is potentially infinite [MS06].

The social network perspective provides a set of methods for analyzing the structure of whole social entities as well as a variety of theories explaining the patterns observed in these structures. Analyses of this kind of graph structures are summarized under the term SNA. The focus of SNA investigation is on the relationship between the individual entities, rather than on the entities themselves. In fact, the goal of this technique is to examine both the contents and patterns of relationships in social networks in order to understand the relations among actors and the implications of these relationships [Ste19] [WF94]. It can be useful to identify local and global patterns, locate influential entities, and examine network dynamics. Therefore it is most valuable for characterizing population-level outcomes when there are relational features that play a role in the behavior of networked individuals [Ber05] [WF94] [LH07]. Edge weighting provides a way to fully capture the richness of the data [Les+09] and ensures that in a network with weighted edges other nodes take a central role than in a network with unweighted edges, even if both networks are topoloigically identical [OP09].

The mentioned „central role of a node“ by *Opsahl et al.* [OP09] within the network can be calculated by different SNA algorithms. However, each of these algorithms places value on different characteristics of the network and must therefore be interpreted in the overall context of the network or applications [Ryd+05].
In the following, A always represents the adjacency matrix of the network graph G and $A_{i,j}$ represents the number of contacts between Node i and j. Accordingly, the edge between i and j has the weight or affinity of $A_{i,j}$. As before, V notes the set of verticies of graph G and n corresponds to the number of rows in the adjacency matrix.

- **Degree Centrality**:
  It assigns an importance to each node within the network based on the number of connections of that node. Thus, 6 connections correspond to an importance of 6, etc. Consequently, the metric is the simplest measure of node connectivity and can be used to find highly connected people, people who are likely to have the most information, or people who can quickly connect to the broader network. For example, targeting these nodes with high degreee centrality may be an effective way to contain a pandemic if the underlying network does not have a distinct community structure [Llo+05]. If this distinct community structure is present it is not these highly interconnected nodes that are responsible for the majority of infections in the network, but the so-called *community bridges*, which may have fewer connections overall, but link several communities [Mor+96].

$$C_{Degree}(j) = \sum_{j=1}^{n} A_{i,j} \tag{2.1}$$

- **Betweeness Centrality**
  This metric takes these intra-community connections into account. The measure calculates the number of times a nodes is on the shortest path between other nodes, which gives an indication of the flow around the system. The betweenness centrality of a node v is given by the expression:

$$C_{Between}(v) = \sum_{i \neq j \neq v} \frac{\sigma_{ij}(v)}{\sigma_{ij}} \tag{2.2}$$

  with σ_{ij} as the total number of shortest paths from node i to node j and $\sigma_{ij}(v)$ as the number of those paths that pass through v.

- **Eigenvector Centrality**
  The third measure identifies the nodes that have influence on the entire network, not just those that are directly connected to it. Eigenvector centrality computes the centrality for a node based on the centrality of its neighbors. This metric also takes into account how well connected a node is and how many links its connections have through the network. This makes this measure particularly suitable for applications within a network, where a maximum effect should be achieved with minimal resources [CF10] [SJ10].

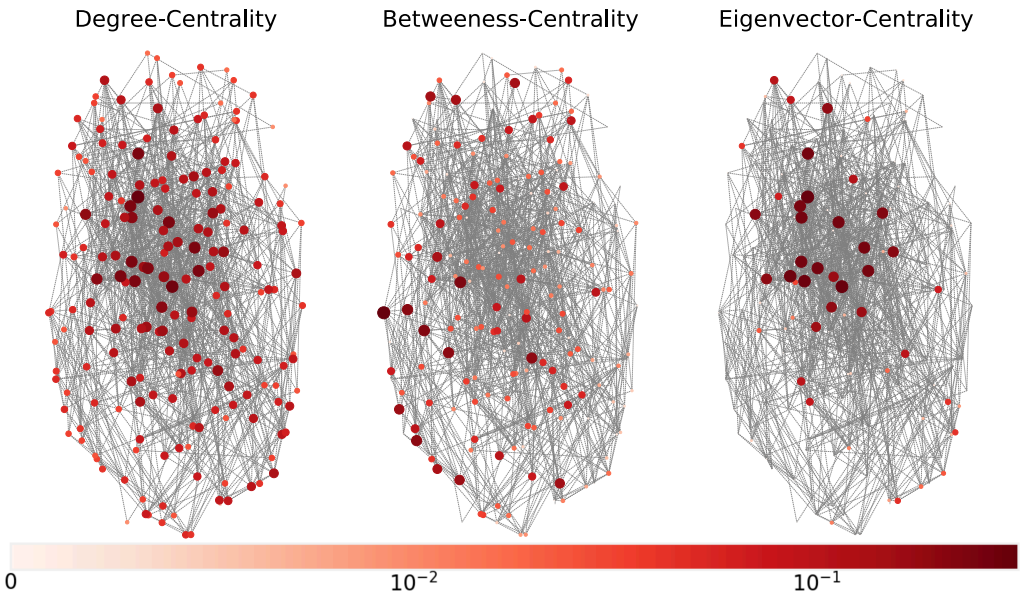

FIG. 2.2: Comparison of *Degree-Centrality*, *Bewteeness-Centrality*, and *Eigenvector-Centrality*. The arrangement of vertices is the same for all three graphs and high values mean "importance" in terms of the associated metric.

The relative centrality x for the two connected verticies i and j is calculated as follows:

$$x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j \in G} A_{i,j} * x_j \tag{2.3}$$

with $M(i)$ as a set of neighbor verticies of i and λ as a constant.

An overview of the three metrics presented is provided in Figure 2.2. The vertices layout is the same for all three graphs, showing that from left to right, fewer and fewer nodes have high values for each metric. *Degree-Cenrality* and *Eigenvector-Centrality* seem to consider partially similar nodes as important. *Betweeness-Centrality* marks different nodes as salient compared to the other two graphs. Therefore, several nodes in this network have a high degree, but only a few nodes have a high degree and influence on larger parts of the network.
These metrics provide a good basis for describing the properties of a network, but represent only a selection of the best-known topological properties.
Applications with real-world networks often involve developing their own metrics or adapting existing ones. The reason for this is that the available measurement methods or data sets can only capture the contact network that is really relevant for the application to a certain degree of detail [SJ10].

2.1.2 Communities in a social network

A network community is a subset of vertices $v \subset V$ with similar degree of relationship between the subset members, but dissimilar with members outside the subset. Social networks show significant community structure and social processes such as homophily and transitivity result in highly clustered and modular networks [SJ10]. The ability to capture this relationship between different groups or individuals allows graphs to represent social systems in detail [For10]. Such clusters or communities can be viewed as relatively independent compartments of a graph, playing a role similar to, the tissues or organs in the human body. For this reason, the methods to uncover and understand these important network (community) structures on multiple topological and temporal scales are of particular interest [Agg11]. Quality functions that quantify the goodness of a given network division into communities formalize the concept of communities. Some of these quality metrics are more common than others, such as *Normalized Cuts* [SM01] and *Modularity* [NG04], but none has achieved universal acceptance since no single metric is applicable in all circumstances.
It is known that modularity has a resolution limit and therefore is not able to detect small communities [For06]. However, being aware of these peculiarities, modularity can very well be considered a robust and useful measure that, according to *Görke et al.* [Gör+13], closely agrees with intuition on a wide range of real-world graphs. The metric is therefore assessed as appropriate for this work. *Modularity* is often used because of its independence from the number of clusters and has become an essential element of many clustering methods. In this algorithm, the farther the subgraph corresponding to each community is from a random subgraph (i.e., the null model), the better and more meaningful the discovered community structure is judged to be. According to *Newman et al.* [NG04] the *modularity* Q is defined as

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \Upsilon(c_i, c_j) \tag{2.4}$$

where m is the number of edges, A_{ij} is the adjacency matrix of G, k_i is the degree of i and

$$\Upsilon(c_i, c_j) := \begin{cases} 1, & \text{if } i \text{ and } j \text{ are in the same community} \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

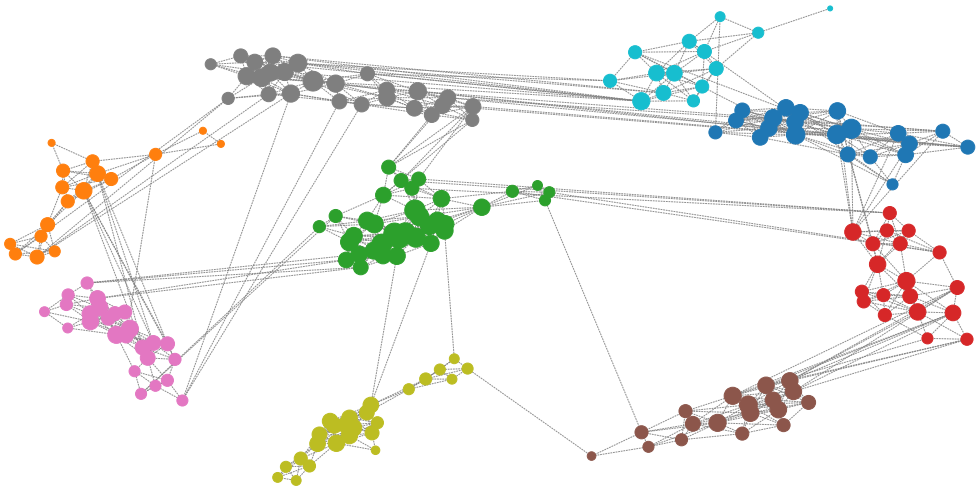

FIG. 2.3: Randomly generated undirected geometric graph $G = (V, E)$ with number of nodes (vertices) $|V| = 200$ and number of edges $|E| = 837$. Geometric graphs resemble human social networks in many ways. They often exhibit community structures, i.e., densely connected groups of nodes are formed. Here these communities are color coded. Corresponding graph to Fig 2.2

After a evaluation metrics for the classification of the graph G into different clusters (communities) has been introduced, an algorithm for the recognition of communities is now presented. This algorithm was developed by *Girvan and Newman* [NG04]. The *Girvan-Newman* method focuses on the concept of *betweenness*, which is a variable expressing the frequency of the participation of edges to a process. The measure is the number of shortest paths between all vertex pairs that run along the edge. It is an extension to edges of the popular concept of *site betweenness*, introduced by *Freeman* [Fre77] and expresses the importance of edges in processes like information spreading, where information usually flows through shortest paths [For10]. This metric is used to identify edges that connect different communities. It is therefore assumed that edges between communities have higher *Betweenness-Values* than edges within a community. These edges should then be capped in order to decompose the social network into its constituent communities.
The general form of the algorithm is as follows:

1. Compute *betweenness* score for all edges in the network using any measurement method.
2. Find edge with the highest score and remove it from the network.
3. Recalculate betweenness for all remaining edges.
4. Repeat process from step 2.

An example of the division of a graph into different communities (clusters) is shown in Figure 2.3. In total, the graph was divided into nine communities by *Girvan-Newman's method*. From a purely visual point of view, the division of the communities makes sense, since it appears that the nodes within a community are more strongly connected to each other than to the rest of the graph. The modularity for this example is 0.755. Comparatively, the modularity for the same graph where 20 of the 200 nodes were assigned to random communities is only 0.614. This comparison graph is shown in the Appendix A.1.

In this section, the fundamentals, metrics, and algorithms from the topic area of SNA that are important for this thesis were presented. These basics are important because network graphs and methods from SNA also lend themselves to modeling infectious diseases in social systems [BG11]. This possibility, along with other models for infectious disease modeling, will be addressed in Chapter 2.2.

2.2 Infection Models

Models for predicting the course of infections in complex systems (social systems), have been widely used in public health at least since the SARS-CoV-2 pandemic. According to *Luke et al.* [LS12], these systems consist of heterogeneous elements that interact with each other. Likewise, these systems have emergent properties that cannot be explained by individual elements and adapt to changing circumstances. Public health is beginning to use results from systems science studies to shape practice and policy, for example, in preparing for global pandemics. In studies of complex systems in a social context, three systems science methods have become established over the past several years:

1. *System dynamics (SD)*
2. *Social-Network-Analysis* (SNA) (cf. Section 2.1)
3. *Agent-based-modeling* (ABM)

Although there is some overlap, these three methods each approach the study of complex systems in different ways. According to *Osgood* [Osg07], each of these methods is appropriate with respect to certain properties of the social systems under consideration. Table 2.2 provides an overview. For example, ABM and SNA are both more appropriate for describing how individual actors in a

System property	SD	SNA	ABM
Model breadth	X		
Dynamic systems in real time	X		X
Interactions of individual actors		X	X
Complex relational structures		X	

TAB. 2.2: Primary strengths of each system science method according to [LS12] and [Osg07].

system interact with each other compared to SD, according to *Luke et al.* [LS12]. That said, SD, SNA, and ABM all have rich, multidisciplinary conceptual and technical histories. Recent developments in computer and modeling technology have further benefited the models.

2.2.1 Epidemiology basics

Basic epidemiological jargon

- R_0**:** The baseline reproduction number indicates the average number of people infected by an infectious person when no member of the population is immune to the pathogen (susceptible population).
- **Superspreading event** and **Superspreader:** In infectious disease epidemiology, a superspreading event is a sudden, "explosive" transmission event in which certain infected individuals, called superspreaders, infect an unusually large number of subsequent cases with a bacterial or viral pathogen, while most infected individuals infect few or no others. Thus, the number of people directly infected by a superspreader is significantly higher than the baseline replication number R_0. Thresholds need to be defined for each application.

Path of infection According to current research, the SARS-CoV-2 virus is transmitted by respiratory ingestion of virus-containing liquid particles transmitted [Haa20] [Zha+20]. Social interactions, such as conversations at close distance, hugging, or several people being in poorly ventilated rooms, are therefore the main routes of transmission. In detail, these particles are aerosols (i.e.small airborne particles and microdroplets with $\varnothing < 5\,\text{nm}$), which are also produced during normal breathing and talking. Activities such as singing or shouting significantly increase the emission of such particles [Lel+20]. In addition these

aerosols have, according to *Jayaweera et al.* [Jay+20] and *Kriegel et al.* [Kri20] a realistic chance of surviving and hovering in the air for an extended period of time. As a consequence, virus contamination of common objects or aerosolization of the virus in an enclosed space may occur.

Following factors play a crucial role in calculating the risk of an infection of a social interaction via the aerosol transmission route:

- Duration of contact
- Distance during contact
- Location of contact (within the subsystem)
- Environmental parameters of the contact location

2.2.2 System dynamics for infectious diseases

System Dynamics is a common modeling approach used to capture nonlinearity in complex systems. It is based on the premise that complex behaviors of a system (e.g., population prevalence of an infection) result from the interplay of feedback loops, stocks, and fluxes that all occur within the bounded system [Agg+20]. The approach focuses on modeling the relationships among the various key elements of each system and developing a top-down representation of the system as a whole. This is done using a series of "stock and flow" diagrams, where each stock represents the set of a particular entity and a flow represents the change in the set of a particular entity [RS08].

One type of system dynamics model commonly used in the field of epidemiology is the *SIR* model. The simple but basic *SIR* framework, was developed by *Kermack et al.* [KM27] as early as 1927 and is still used today to provide important insights into the evolution of a new epidemic in an idealized susceptible population with random mixing. The basic *SIR* model has three groups: *susceptible* (**S**), *infectious* (**I**), and *removed* (**R**). Each of these groups represents a population variable containing the number of individuals in the population in that infectious state. Thus, the sum $N = S + I + R$ is a collectively exhaustive representation of the entire population (N). It is parameterized by the infectious period $\frac{1}{\gamma}$, the baseline reproduction number R_0 (the number of secondary cases for each infection in a fully susceptible population), and the contact rate $\beta = \gamma * R_0$.

Figure 2.4 shows an example of the progression of an infectious disease within

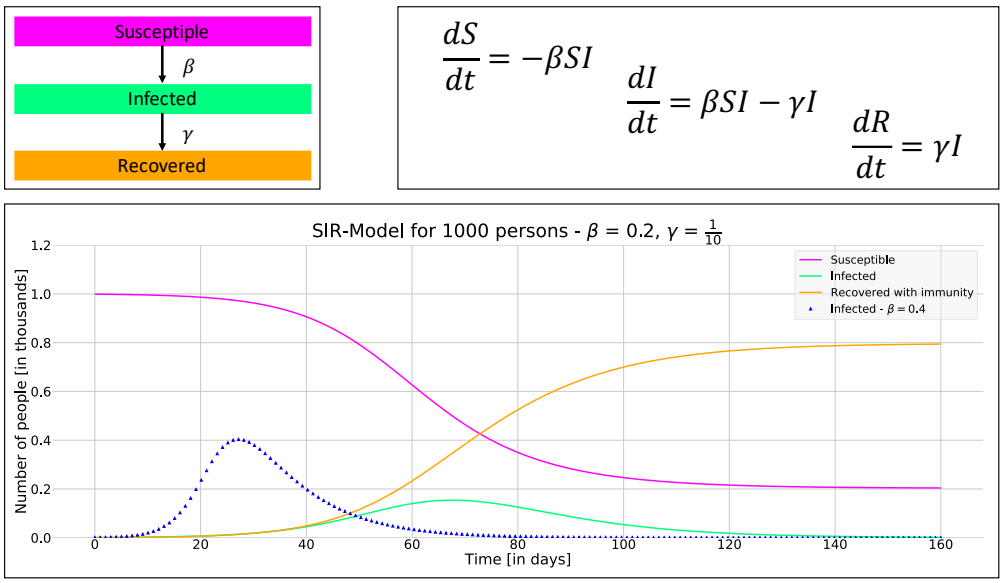

FIG. 2.4: *SIR* model overview with the three groups into which the entire population under consideration is divided in the upper left and the differential equations describing the transit of the persons between the groups in the upper right. The graph shows the number of persons per time step for each of the three groups. The total population is 1000 persons, the contact rate $\beta = 0.2$ and the infectious period $\gamma = \frac{1}{10}$. 160 days were simulated in total. Blue dashed line shows the curve for the case when β is raised to 0.4. This is for comparison only.

a population of 1000 people. It can be seen that the decrease in susceptible individuals and the increase in recovered (immune) individuals are somewhat opposite. This is explained by the green line, which accounts for the number of infected individuals per time point. Other interesting observations are that about 20% of people are still susceptible (i.e. have not been infected) and about 80% are immune to the disease after 160 days. At this point, no more people are infected, so no further infection can occur in this system. The blue line shows how drastically the progression of infected persons changes when β is changed from 0.2 to 0.4. The time period in which people are infected in the system is extremely shortened, which leads to a much higher maximum number of simultaneously infected people.

In recent years, various advancements of this model have been presented for different infectious diseases. For example, since most infectious diseases have a latent period between being infected and becoming infected, the *SEIR-model* with the *exposed* group (**E**) was presented. After infection, individuals migrate to this group at a rate $\beta * S * \frac{I}{N}$ and remain there for an average period of $\frac{1}{\sigma}$ before moving to group I. Again, because in many respiratory infections immunity after recovery is transient, and recovered individuals lose this immunity again, in a further development the *SEIRS* model was presented in which individuals

return to group S after an average period of protection of $\frac{1}{\omega}$.

With these and other compartments (groups) along with additional more complicated flows between them, including aspects such as birth, death, and age, more complex disease transmission scenarios can thus be modeled.

2.2.3 Social-Network-Analysis for infectious diseases

There is an extremely close relationship between epidemiology and network theory that dates back to the mid-1980s [MA87] [Klo85]. In most cases, the graphs used in this field are constructed in such a way that the individuals represent nodes and the edges between the individuals represent parameters such as contact, number of contacts, or some other value for a social interaction. In section 1.1, a similar example was given with the social contacts of scientists in Figure 1.1. Hence, it is the connections between individuals (or groups of individuals) that enable the spread of an infectious disease. These connections define a natural network from which, in turn, insights into epidemiological dynamics can be gained. Methods from the field of SNA are therefore used to characterize social networks and to draw conclusions about how network structures may influence the risk exposure of members of the network [ElS+12]. Social Network Analysis can be ideal for understanding social contagion as well as the influences of social interaction on population health. In particular, understanding the structure of the transmission network allows for better predictions of the likely distribution of infection and early growth of infection (post-invasion) and allows for simulation of the overall dynamics [Dan+11]. However, the interaction between networks and epidemiology goes even further. Because the network defines potential routes of transmission, knowledge of its structure can be used in the context of disease control. For example, contact tracing aims to identify likely transmission network links of known infected cases and thus treat or contain their contacts, reducing the spread of infection. Contact tracing is an effective public health measure because it uses the underlying transmission dynamics to target control measures rather than relying on a detailed understanding of the etiology of the infection [DH20]. Another example of the potential use of the social network graph is the identification of, with respect to infectious disease, peculiarly pivotal individuals or communities. These can be either particularly vulnerable due to their position within the network (for example, if they have a lot of contact with other individuals), or very critical to the entire network due to the fact that they have a lot of contact with many

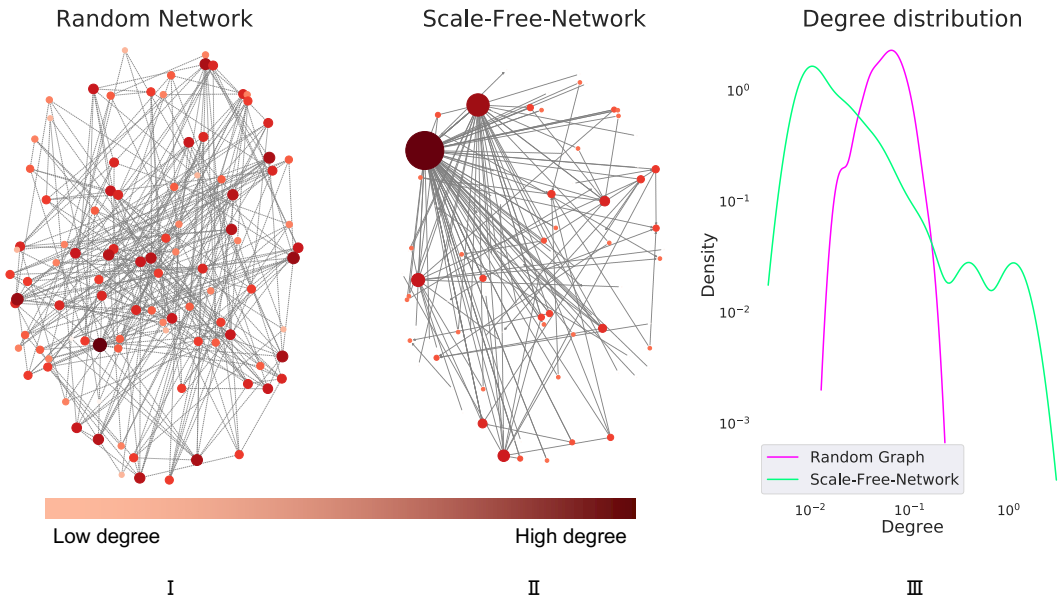

FIG. 2.5: Comparison of random and scale-free networks. (I) Example random network with 100 nodes and 300 links. All nodes have similar numbers of links. (II) Example scale-free network with 100 nodes and 300 links. Most nodes have few links, with a few nodes having many links. (III) Degree distributions for two classes of networks. Degree and associated color scale always refers to the respective graph.

different individuals (infecting many individuals) [ElS+12]. The latter gennate group of people in a network have the potential to become super-spreaders due to their position, by virtue of the definition given in 2.1.1. According to *Drosten* [DH20], the super-spreaders are the driving force behind the epidemic. In any case, the methods presented in chapter 2.1.1 and 2.1.2 can help to identify such parts of the network.

As an example, two graphs are shown in Figure 2.5, where in the random graph (I) each node has approximately the same degree. This means that in this graph each node has approximately the same number of connections to other nodes. However, small differences with respect to size and color still exist. A strong contrast is the Scale-Free graph (II). Scale-free graphs are complex networks whose number of connections per node is distributed according to a power law. That is, the fraction $P(k)$ of nodes in the network having k connections to other nodes goes for large values of k as

$$P(k) \sim k^{-\kappa} \tag{2.6}$$

where κ is a parameter whose value is typically in the range $2 < \kappa < 3$ [Onn+07]. This distribution can be clearly seen since most of the connections are dis-

tributed among a few nodes. In the context of infectious diseases, this could mean that very few people within the network caused the majority of infections. This is also emphasized by the degree distribution shown in Figure 2.5 (III).

2.2.4 Agent-Based-Modeling for infectious diseases

ABMs are stochastic computer simulations of simulated "agents" or individuals in simulated space, over simulated time [ElS+12]. They offer the possibility to describe complex behaviors by simulating each individual separately. A problem that is hard to describe globally can often be described locally on the level of the participating entities. With the help of a simulation the global behaviour can then be modeled. These models allow macro-level behavioral patterns to be developed from explicitly described behaviors, interactions, and movements of agents in their environment. Because the conceptualization and parameterization of the model is "bottom-up," these models are ideal for assessing emergence, or macro-level patterns that emerge from micro-level behavior [Bon02]. To achieve this complexity or heterogeneity must be added to the simple model. This can be done in several ways. Heterogeneity can be introduced into the system by considering more different individuals. In a pandemic simulation, one can start by adding an age distribution to the agents and changing the update rules depending on the age of the agent. Another important point where heterogeneity can be introduced is in the interactions. If well-mixed random interactions are assumed, social structure in behavioral models can be implemented. Often it can make a big difference whether agents act with random contacts or always with the same group of contacts [DBU12].

According to *Auchincloss et al.* [AD08] agent-based approaches are particularly appropriate when:

1. Individual agent behavior is complex, with learning and adaptation, feedback loops, and/or reciprocity.
2. When heterogeneous environments can affect agent behavior and interaction, and agents are not spatially or temporally fixed.
3. When interactions between agents are complex and nonlinear, and affect agent behavior.

Properties	Definition
Heterogeneous	Agents are allowed to differ from one another on important characteristics
Spatial	Agents are located in some explicitly define space
Interactive	Agents can interact locally with one another and their environments
Dynamic	Agents are assumed to have imperfect knowledge
Bounded rationality	Models are recursive, are allowed to change nonlinearly, and exhibit nonequilibrium

TAB. 2.3: Core properties that collectively underlie most agent-based models according to *Luke et al.* [LS12]

The resulting key properties of agents and agent-based models are listed in Table 2.3.

To implement these properties, agent-based modeling requires that the investigator explicitly describe and program agent characteristics and update rules during implementation. This includes the specification of agent characteristics and behaviors, as well as their changes over time (e.g., learning and adaptation). Agents can be nested in social networks that influence the degree and character of interaction between agents, and social interaction can be programmed to influence future behavior [ElS+12]. In addition, researchers can explicitly define the space in which agents are located over time and the influence of that space on agent behavior over time [DBU12]. ABMs are particularly well-suited for research concerned with understanding social processes, where agent behavior is a complex function of agent attributes and characteristics, environments, and interaction between agents over time. The main reason for this is that they maintain the centrality of the individual agent and its attributes, characteristics, and behaviors in the production of population-level phenomena [Ven+18]. Research has been able to integrate geographic information systems and social network information into agent-based models [Del+10] . In addition, ever-improving computational resources enable the use of extremely large sets of agents in simulations, including synthetic populations of entire communities or nations [Cue20] [CCW10].
These capabilities are another reason for the particular success of agent-based modeling in healthcare. Here, ABM are often used to study epidemics and infectious disease dynamics [Cue20] [Ven+18] [Cha+20b]. The goal here is to study disease transmission at different levels, from individual communities to global pandemics [Eps09]. Agent-based models of epidemics have helped move

epidemiology beyond the traditional *SIR* model and have demonstrated the importance of examining the role of social networks, transportation systems, local geography, and various behavioral responses to changing contexts in the spread of disease [Eps+08] [Eub+04] [YAE08].

Thus, ABM is focused on the individual characteristics and interactions in time and space. It also allows researchers to run multiple simulations under different model conditions to isolate the effects of specific conditions on the outcomes of interest.

2.3 Model discussion

To develop infectious disease prediction models, studies commonly combine elements from different systems methods to model the interaction between the behavior of individual agents with social networks [CMM10] [Klo+94] [Chr+05] and with the system dynamics of epidemics [SPN20]. As we are witnessing in the current SARS-CoV-2 pandemic, infectious diseases have significant public health, health care, macroeconomic, and societal implications. Many factors, including increasing antimicrobial resistance, increasing human interconnectedness, and changing human behavior, make prevention and control matters of national policy an international challenge. The availability of options to control and prevent the emergence, spread, or reemergence of pathogens warrants continuous evaluation using a variety of methods.

Applications of system dynamics in infectious diseases range from early studies that emphasized describing the dynamics of disease spread to more recent work that is more focused on testing potential impacts of infectious disease control strategies. Early examples of SD in infectious diseases, as applied to the *AIDS* epidemic, focused on describing the dynamics of the disease transmission process and characteristics of Human-Immunodeficiency-Viruses (HIV), such as incubation period [HS91]. Models have provided particularly meaningful results when data is available to allow a test of model validity [Fer+05].
With respect to SARS-CoV-2, the *SIR* model is often extended to a *SEIR* model that accounts for an additional exposed (E) stage in which individuals are infected but not yet contagious. Many additional parameters, including spatial heterogeneity [Küh+20], clustering [Luo+20], age heterogeneity [RWC20] [Ivo+20] [ZC20], and even meteorology [Jia+20] have been incorporated into the *SIR* and *SEIR* frameworks to increase the predictive power of these models. While these extensions increase the predictive power of the model based on

the respective research results, it remains unclear whether the human interaction network is sufficiently accounted for as the main transmission pathway. Thus, the *SIR* model and its partial hybrid advancements are well suited for inferring infection dynamics such as R_0, predicting the macroscopic dynamics of infections and deaths, and evaluating various nonpharmaceutical interventions aimed at containing the microscopic dynamics of person-to-person infections [RSW20]. However, individual social interactions are not represented. Neither at the local scale (i.e., egocentric networks) nor at the global scale (the topology of the resulting network). Criticizing this neglect, *Manzo et al.* [Man20] state that the widely used compartmental models in the current Corona pandemic can only lead to one type of intervention, i.e., interventions that indifferently affect large subsets of the population or even the entire population. In this regard, *Hermann et al.* [HS20] notes that the type of models used cannot evaluate targeted interventions that might surgically isolate specific individuals and/or cap specific human-to-human transmission pathways. It is important to note here that a major reason for using these models in the current pandemic is, first, the lack of data on individual interactions at, for example, the country population level and, second, the development of SNA or ABM models takes much more time that is often not available in the current situation.

The individual social interactions form the basis for studies using SNA. These aim to capture the complex interplay between individual behavior and social contexts at large scale. Social network analysis in the context of infectious disease focuses on characterizing social networks to draw inferences about how network structures may influence the risk exposure of network members. *Cauchemez et al.* [Cau+11], for example, were able to quantify how *Influenza* transmission is influenced by social networks following an *H1N1 (influenza)* outbreak that began in an elementary school and spread to a semi-rural community in Pennsylvania. Using social interaction data and SNA methods, the authors showed that sitting next to an ill person or being the playmate of an ill person did not significantly increase the risk of infection. However, the structuring of the school into classes and grades strongly influenced prevalence.
Such detailed information about infection progression is important for designing appropriate interventions. Complementing this, the results of *Christley et al.* [Chr+05] show that SNA measures such as *degree centrality* are good for predicting individuals' risk of infection. Because individuals near the center of a social network become infected earlier on average during the course of an outbreak than those on the periphery. Consequently, identification of more central individuals in populations can be used to inform surveillance and infection control

strategies and detect infection outbreaks early. *El Sayed et al.* [ElS+12] therefore concludes that network analysis can be ideal for understanding social contagion as well as the influences of social interaction on population health. However, network analysis requires network data, which may affect generalizability, and causal inference from current network analytic methods is limited.
According to *El Sayed et al.* [ElS+12], the main limitation of using SNA in epidemiology is the implicit trade-off between the use of network analytic techniques and the generalizability of network data. It is criticized that social network analysis requires data on the relationships between individuals in addition to data on the characteristics of individuals in networks. Traditional sampling methods that aim to improve the generalizability of studies by randomly sampling across environments are therefore not conducive to the use of network approaches because the data on relationships obtained from these methods are not of sufficient completeness or quality to support them. Cost and feasibility constraints therefore often force researchers to balance tradeoffs between the analytic advantages of social network approaches and the importance of generalizability when designing epidemiologic studies. Another problem in building models based on personal social contacts and information is privacy. This is therefore relevant for both social network creation and ABM. Collecting the required data deeply interferes with the privacy of individuals and deters many people. However, in an industry setting, the company may mandate the collection of such data by employees.

ABM can promote population-level inference from explicitly programmed micro-level rules in simulated populations over time and space [ElS+12]. Agent-based models have been used to simulate various infectious diseases [Dan+11]. *Lee et al.* [Lee+10c] modeled vaccine allocation policies in the face of an *H1N1* epidemic to examine priority recommendations for high-risk individuals versus highly infectious children when vaccines are scarce and to draw comparisons among outcomes such as seizure rates, hospitalizations, and total costs.
Members of the same research group [Lee+10b] identified problems with school closure strategies to control influenza outbreaks and found that short closures are counterproductive and that only longer closures provide the time needed to implement long-term effective vaccination programs.
Another study [Lee+10a] used influenza models to examine the impact of workplace *H1N1* vaccination strategies and found that programs targeted to larger firms were more efficient and effective than those distributed to a larger number of smaller workplaces. ABMs are useful for assessing health determinants at multiple levels of influence that, when combined with social interaction, can

Properties	ABM	SNA	SD
Model complexity	- -	0	+ +
Parameterization	- -	+	+
Required computing power	- -	0	+
Infection dynamics	+ +	0	+
Inferences for individuals	+ +	+	- -
Development time	-	+	+ +
Derivation of countermeasures	+ +	+	-
Suitable for large populations	+ +	0	+

TAB. 2.4: Comparison of methods for modeling infectious diseases in complex systems after own presentation. Scale from $--$ *worse than others* to $++$ *better than others* and 0 as *neutral*, always in comparison to the other two methods. Dependencies on, for example, data set size or available infrastructure are neglected.

contribute to population health. ABMs allow exploration of feedbacks and interactions between exposures and outcomes in the etiology of complex diseases. They can also provide opportunities for counterfactual simulations. However, appropriate implementation of ABMs requires a balance between mechanistic rigor and model parsimony, and the precision of the results of complex models is limited [Dan+11].

In this sense, the process of implementing the model should be adapted to the issues of interest to avoid unnecessary complexity. However, model fitting can be logically problematic, since a priori purposeful model fitting implies the exclusion of factors that should not have an obvious impact on the outcomes of interest. However, a central argument for agent-based approaches is the ability of these models to provide emergent phenomena based on the aggregation of complex micro-level processes to provide a macro-level overview. Compared to SNA and SD, building a model to model an infectious disease is the most costly and complex.

A summary of the three presented methods for modeling infectious diseases in complex systems, together with the evaluation of each strength and weakness is shown in Table 2.4. The evaluation is based on the scientific publications that were analyzed within the scope of this thesis. In each row of the table, the three methods are compared based on one criterion. The evaluation is therefore to be seen as a comparison.

ABMs are a powerful tool to model an infectious disease within a social system [ElS+12]. Due to a parameter-rich model and a detailed data basis, not only

can good statements be made about the infection dynamics, but also single individuals or groups of individuals can be indexed that play a major role in the infection dynamics. However, these advantages come at the price of very high demands on the modelers, the data and the available infrastructure. The development of such a model for large populations is theoretically possible, but practically, compromises often have to be made with respect to the level of detail in areas due to complicity. As can be seen in the table, SD performs well wherever ABM has weaknesses. Even with superficial population data, few parameters, and limited infrastructure, an infection model can be built in a short time to map infection dynamics. Of course, these advantages come at the expense of the model's informativeness or generalizability. It is also difficult to draw conclusions about individual persons or groups of persons and their respective role in the observed social system or influence on infection dynamics. A middle ground of these two methods is offered by SNA. Although more detailed data are required than for SD, it is not necessary to collect individual person characteristics, as is the case with ABM. For example, simple interaction protocols are sufficient to calculate the influence of individual persons on the rest of the network. The complexity, the required hardware resources and also the development time is also between ABM and SD. Depending on the application and the data, social networks in epidemiology with respect to different properties may represent a compromise between SD and ABM. Interaction data of individuals within a system are particularly suitable for social network creation and allow the detection of temporal structures [Onn+07].

2.4 Related Work

With respect to the SARS-CoV-2 virus, a variety of modeling techniques have been studied to date. Whether compartmental homogeneous mixture models, contact networks, or agent-based models: So far, most of these analyses have used simulated data [HKO21]. Yet studies have already shown that detailed models of social dynamics (microdynamics) are important for understanding dispersal. This finding has been addressed in several researches and how the use of a real contact network changes the understanding of infection dynamics, containment measures and infection risk [Cen+21] [Mac+13] [SSL16] [Mos+08] [Sat+20]. For these studies, dynamic social networks of densely connected populations (individuals) and their interactions (in close proximity) were used in the network of real person-to-person proximity. With this type of data, *Hambridge et al.* [HKO21] demonstrate, that while frequent testing can dramatically reduce

spread, this has limited impact when mask wearing and social distancing are not widespread. Moreover, even moderate levels of immunity can significantly reduce new infections, especially when combined with other interventions. The data used were recorded through configured smartphones and analyzed using a discrete-time, stochastic *SEIR* compartmental model. At each time step, individuals moved to the next compartment or remained in their current compartment. Because prolonged exposure has been shown to increase the risk of infection, the infection model is set up so that the number of interactions is the dertminate of virus spread.
Sattler et al. [Sat+20] have also recognized the potential of digital contact tracking, based on Bluetooth Low Energy (BLE), and the importance of exposure duration to efficiently contain and delay infectious disease outbreaks such as the current SARS-CoV-2 pandemic. A machine learning-based approach is proposed that can be used to reliably detect individuals who have spent enough time in close proximity to be at risk of infection. Specifically, the infection risk of each interaction between two people was classified into two groups (*high risk*, *low risk*) based on the duration and the respective distance.

Although infectious diseases are mainly transmitted through social interaction, recent research provides reason to believe that the location or environment of the social interaction is also important in creating a meaningful infection model [KH21]. For the risk of infection via aerosol particles in enclosed spaces, the inhaled dose is crucial which, according to *Kriegel et al.* depends on source strength (emission rate), breathing activity (source and receiver), aerosol concentration in the environment and the duration of stay in the environment. Source strength and breathing activity depend on the activity in the environment.
Buonanno et al. follow exactly this approach and calculate the risk of infection based on the characteristics of the place where the contact occurred. In detail, the aerosol exposure at a given location is calculated as a function of numerous parameters such as room size, air circulation / fresh air rate or activity (listening, singing, physically strenuous activities). The decisive factor for the risk is therefore when one was in the same room with whom and who was infected there and when. With the proposed approach, it was possible to model retroperspectively the high infection rate in two outbreaks in a restaurant, and during a choir rehearsal.
According to *Buonanno et al*, outbreaks with high case rates are not caused by a superspreader, but rather by the coexistence of conditions, including emission and exposure parameters, that lead to a superspreading event. With respect to the different risks of infection in different locations *Lelieveld et al.* [Lel+20] and *Kriegel*

Environment	Occupancy	Countermeasure	$R_s \leqslant 1$
Theater, Museum	30%	Mask	0.5
Public transport	∅	Mask	0.8
Shopping	10qm / person	Mask	1.1
Fitness center	30%	Mask	1.4
Multi-person office	20%	Mask	1.6
Restaurant	50%	-	2.3
School	50%	Mask	2.9
School	50%	-	3.4
Multi-person office	50%	-	8.0
School	100%	-	11.5

TAB. 2.5: SARS-CoV-2 contagions via aerosol particles. Evaluation of indoor spaces in terms of situational R_0 value with an infected person inside. In each case, x times the risk is given compared to a situational $R_s \leqslant 1$. The R_s value means the number of infected persons with an infected person present at the same time.

et al. [KH21] reach similar conclusions. A section of the infection risk at different locations according to *Kriegel et al.* is shown in Table 2.5.

A hybrid approach that considers both the different probabilities of infection at different locations (Point of Interest (POI)) and the social interaction patterns was presented by *Chang et al.* [Cha+20a]. Although the study aims to describe the risk of infection for income-dependent populations, this can be generalized to individual communities in a subpopulation or to individuals in that subpopulation without further problems. In this approach, infection risk is calculated using a metapopulation *SEIR* model that integrates POI characteristics (for example, type of location, ventilation, number of people per epoch, etc.) and a fine-grained dynamic mobility network for the ten largest metropolitan statistical areas in the United States.
Thus, both the residence time of individual nodes and the frequency of interaction of two nodes at a given location per epoch are considered. This approach enables analyses such as the identification of particularly infectious locations or the detection of individuals who are at increased risk due to frequent visits to the most exposed locations on the one hand, but who also serve as unavoidable virus transporters between two different locations on the other. In addition, research findings include evidence that a small minority of "superspreaders" POIs are responsible for a majority of infections and that limiting maximum occupancy at each POI is more effective than uniformly reducing mobility.

3 Model for infection probability per contact in a social network

As described in section 2.2.1 airborne infectious diseases such as *Influenza* or SARS-CoV-2 are transmitted via proximity contacts between individuals. The models presented in section 2.4 are each suitable for a specific application and each pursue different objectives. The available data basis is decisive for the development of a suitable model for determining the risk of infection. Proximity contacts take place in different contexts. At home between family members, in the public transportation, shopping centres and at school or workplaces. The focus in this thesis is on close contacts within an organizational unit. The characterization of proximity contacts is therefore a prerequisite to quantify the risk of infection. For this reason, this chapter will describe the available data (Section 3.1) and then the model approach for calculating the risk of infection between two persons (Section 3.2).

3.1 Interaction data within the ResTAat project

In detail, anonymous close contact data were collected in a manufacturing company in Italy. Each employee carried a button device (token) during working hours, which reports a distance alarm when the distance between two button devices (ergo two workers) is less than 1.5 meters for more than 15 seconds. These threshold values correspond to the standard configuration of the buttons used, but can be selected freely for each application. With reference to current research results on the transmission of SARS-CoV-2 [Jay+20] [KH21] [Zha+20], these values were considered plausible. The buttons communicated with gateway beacons via Bluetooth technology and calculated the distance to buttons in the vicinity via RSSI signal strength. A similar approach using modified smartphones was used by *Hambridge et al.* [HKO21]. Recorded near-contact alarms were sent to a backend server via gateways. These gateways were installed at various locations within the plant site. The exact physical location of the gateways is known. The used button hardware from the company *secufy* records the social interactions anonymously and in compliance with data protection regulations. The software backend used is from the company *safefactory*.

Figure 3.1 I shows the schematic structure of the infrastructure used and Figure 3.1 II shows the button devices used. The raw data extracted from the *safefactory* backend contains a total of 279445 near contact alarms between 621 workers. The

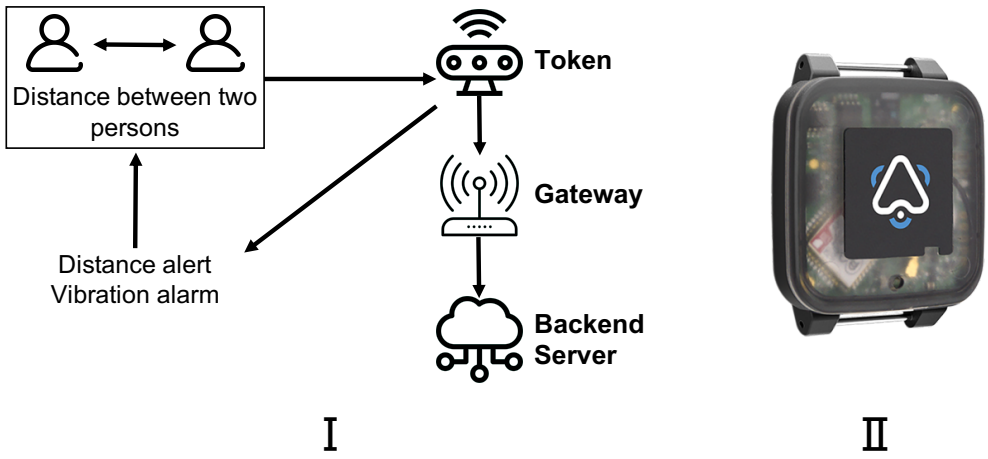

FIG. 3.1: Schematic of the hardware used to record the interaction data and a token. The battery-powered tokens are attached to the clothing of the various individuals and act autonomously. They give a haptic signal when the distance between them falls below the minimum distance. Illustration of the token was taken from page https://secufy-sos.com/pages/secufy-sos.

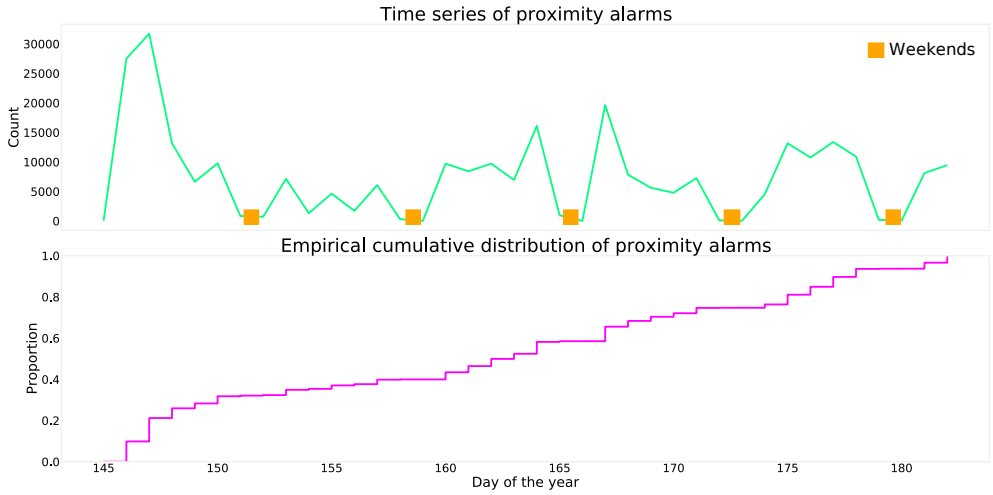

FIG. 3.2: All logged near contact alarms over time. Green plot shows the number of critical social interactions per day and the magenta plot shows the cumulative distribution. In both cases 44 days were considered.

time period considered is from May 24, 2020, 03:03:51 to June 22, 2020 22:44:19. Table 3.1 shows a small sample of the available unprocessed interaction data. Figure 3.2 shows the near-contact alarms over the considered period. It can be seen that the number of alarms per day varies. On weekdays, up to 30000 near contact alarms occurred, while on weekends the number of critical social interactions is almost zero. The cumulative distribution in the lower part shows that over the considered period the distribution of contact alarms is approximately constant, except for the sharp slope within the first five considered days. Since infected people are contagious only for a certain period of time, the fluctuations

id	deviceid	devicetime	beaconid	day	gateway
138404553	ISA_4_13	2020-05-24 03:03:51.037	ISA_4_43	145	bz2139
138404560	ISA_4_27	2020-05-24 03:03:52.533	ISA_4_49	145	bz2130
138404562	ISA_4_133	2020-05-24 03:03:52.550	ISA_4_492	145	bz2138
138404567	ISA_4_90	2020-05-24 03:03:54.045	ISA_4_30	145	bz2129
138404581	ISA_4_243	2020-05-24 03:03:58.594	ISA_4_55	145	bz2131

TAB. 3.1: Extract of the social interaction dataframe containing all logged proximity alerts within give time-range. *Deviceid* (primary) and *beaconid* (secondary) describe the tokens involved in the interaction.

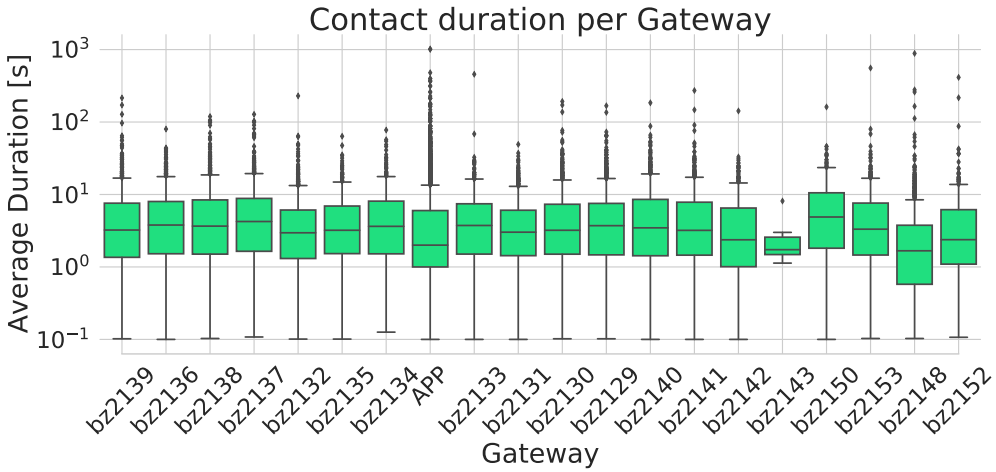

FIG. 3.3: Close contact duration based on the exact timestamps for each of the gateways used. All gateways show a similar distribution, only by the APP as gateways seems to show a larger amount of outliers. The global contact duration average is 6.09 seconds.

in the interaction patterns shown may introduce uncertainty in the calculation of the outbreak size. For example, a person who becomes infectious on Fridays and is infectious for three days may only infect people within the company on that same day. If this person becomes infectious on Tuesday, a larger number of secondary cases can be expected, since the infected person is present in the company for three days. The close contact alarm is triggered when persons are closer than 1.5 meters for more than 15 seconds. However, the tokens are programmed in such a way that other tokens in the vicinity are still detected and reported to the gateways. This data can be used to calculate the average time that people were closer than 1.5 meters. Figure 3.3 shows an overview of the average contact duration broken down for each gateway installed in the considered social subsystem. The global average contact duration is about 6 seconds and similar at all gateways. Thus, there is no gateway, or monitored location, within the company where people spend an above-average amount of

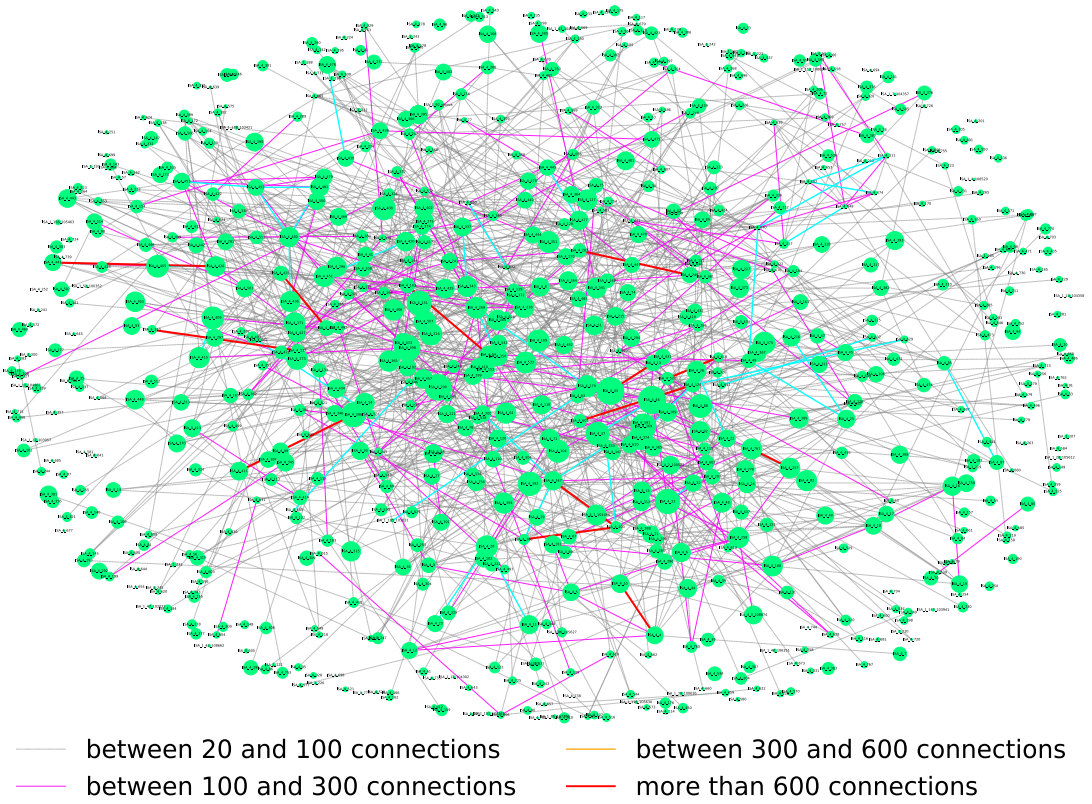

FIG. 3.4: Interaction network of the recorded data within the production company in Italy. Edges are only displayed if there are at least 20 close contact alarms between the two account points. The color of the edges describes the number of near contact alarms. The size of the nodes shows the eigenvector centrality of the respective node. The entire period over 44 days was included.

time. Since the average contact duration is below the threshold for alarms, no alarm was triggered in most close contact encounters. The distribution of social interactions across gateways is approximately even. Therefore, it is assumed that interactions are evenly distributed within the social system and that clustering does not occur at different locations on the site.

The data provides person-based interaction logs. As, apart from the number of contacts between two individuals (later also called agents), only the physical position of the transmitting gateway is known, epidemiological observation in social networks is possible on the basis of the criteria presented criteria in Tabular 2.4. An example of network topology visualization is described in Figure 3.4. All 279445 near contact alarms between the 621 workers within the 45 days considered are mapped. The displayed graph is undirected but has weighted edges. The color coding of the edge e between the nodes (i, j) indicates how many social interactions i and j had with each other within the time period. Also the size of the vertex v shows the *eigenvector centrality*, which is an indicator for the influence of v on the whole rest of the network. It can be seen that few edges

have a high weight, but multiple nodes have influence on larger parts of the network. Since the connection between agents is undirected, an alarm between agents A and B that originates from agent A will equally increase the weight of the edge between agent A and agent B, as an alarm that originates from agent B will.

$$A \to B = B \to A \tag{3.1}$$

3.2 Disease transmission via proximity contacts

The dataset described in the previous chapter provides a temporal network of social interactions for modeling the spread of airborne infectious diseases. In this context, a primary case is an infected individual, and a secondary case is an individual that became infected after contact with the primary case. Specifically, in the context of this work, the primary case is assumed to have been infected outside the system under consideration and therefore to have "imported" the disease. The number of potential cases depends on the time window considered. All persons who had contact with the primary case while it is infectious are potential cases. From the perspective of infectious disease transmission, the relevant time window is the infectious period, T, the time interval in which the infected individual transmits the disease. For *Influenza*, $T = 1$. For SARS-CoV-2, $T = 3$. For Severe-Acute-Respiratory-Syndrom-Coronavirus-2 variant, first identified in the United Kingdom (SARS-CoV-2-B.1.1.7), $T = 4$.

It is assumed that the number of detected close contacts between two individuals is a surrogate for how long they have been in close proximity. Disease transmission is then encoded in the probability of infection per contact p. Here, p is an effective parameter that depends on disease-related factors and the operational definition of close proximity. Such encoding allows the model to be used for different infectious diseases, such as *Influenza* or Severe-Acute-Respiratory-Syndrom (SARS), on the one hand, and to use different thresholds for triggering a close contact alarm (distance between individuals or contact duration) on the other. Once p is specified, the probability of disease transmission after n contacts with an infectious individual by the probability that the disease is transmitted in at least one of the contacts is defined as

$$P_n = 1 - (1 - p)^n. \tag{3.2}$$

Therefore, the probability of infection depends on the number of contacts of the primary case. Accordingly, the contagiosity T also has an effect on P_n as n depends on T.

3.2.1 Probability of disease transmission during the infectious period

An extract of the underlying data is shown in Table 3.1. The infectious disease transmission model takes as input the contact records for a workplace with n individuals over a period of m days. The $n * m$ matrix C_{ijd} therefore describes the number of close contact alerts between the pair of individuals (i,j) on day d for $i \in [1,...,n]$ and $d \in [1,...,m]$. Assuming that primary case i becomes infected on day d and remains infected for $d+T$, the transmission probability from i to j can be calculated considering formula 3.2 with

$$P_{i\to j,d}(p,C,T) = 1-(1-p)^{\sum_{k=d}^{d+T} C_{ijk}}, \tag{3.3}$$

where $i, j \in [1,\ldots,n]$ and $d \in [1,\ldots,m-T]$. To simplify subsequent implementation, the number of near-contact alarms between two individuals in the time interval $[d, d+T]$ becomes

$$D_{ijd}(T) = \sum_{k=d}^{d+T} C_{ijk} \tag{3.4}$$

$$P_{i\to j,d}(p,D,T) = 1-(1-p)^{D_{ijd}(T)}. \tag{3.5}$$

The result is a table of pairwise social interactions and the associated risk of infection per time-interval Δ for $\Delta \in [\Delta_0,...,\Delta_{m-T}]$. However, we will use the matrix C_{ijd} in the remainder of this section.
Since the number of infections is needed for the following risk calculation, each infection probability is converted into a binary variable *infection event* which describes whether an infection has occurred or not during the considered connection of two agents for period $d+T$.
Infection event $= 1$ for $Z \leqslant P_{i\to j,d}$ and *Infection event* $= 0$ for $Z > P_{i\to j,d}$ with Z as a random variable following the discrete uniform distribution between 0 and 1. Due to the dependence of $P_{i\to j,d}$ on $C_{ijd}(T)$, the probability of an infection event increases with the number of contacts of the considered agent pair. Summing up the infection events for $d \in [1,...,m-T]$ and taking the average of these values, the reproduction number R_M as a function of p, C, and T can be obtained.

$$R_M(p,C,T) = \frac{1}{n(m-T)} \sum_{ijd} P_{i\to j,d}(C,T) \tag{3.6}$$

Formula 3.6 shows the expected reproductive number as a function of disease transmission rate per contact for different infection periods T based on proximity alerts. By inverting this relationship, we can determine the transmission rate per contact p so that R_m is consistent with R_0. The transmission rate per contact is therefore the solution of the equation

$$R_M(p,C,T) = R_0, \tag{3.7}$$

given C,T and R_0 of the infectious disease.

3.2.2 Secondary cases and their distribution

The described dependence of R_M with the interactions of the agents and the infection probability per contact allows the calculation of the distribution of the secondary cases with regard to the mentioned parameters. Using the random variable Z presented in formula 3.6, the number of infection events over the time period d can be calculated for each interaction pair (i,j). Thus, to obtain the secondary cases, a Bernoulli test is performed for which holds:

$$X_{ijd} \in \{0,1\}, \tag{3.8}$$

with probability of success given by $P_{i\to j,d}$. The individual number of secondary cases for each individual i is obtained with

$$Y_{id} = \sum_{j} X_{ijd}, \tag{3.9}$$

for $d \in [0,...,m-T]$. The distribution of the number of secondary cases (k) is therefore

$$P_k = \frac{1}{n} * \left(\sum_{i=1}^{n} \delta_k \left(\sum_{d=0}^{m} Y_{id} \right) \right), \tag{3.10}$$

with

$$\delta_{k,x} := \begin{cases} 1, & \text{if } x = k \\ 0, & \text{otherwise.} \end{cases} \tag{3.11}$$

In addition, to prevent random extreme values and resulting erroneous conclusions, 10000 numerical simulations of P_k were performed.

With the distribution of secondary cases and the information contained therein, the rate of super-spreading events can be quantified. The number of secondary cases depends on the number of potential cases, which in turn depends on the temporary near contact network. According to the definition in chapter 2.2.1 a super spreading event occurs when the set threshold of secondary cases is exceeded. There is no general consensus in the literature about the number of required secondaries for such an event [Ada+20] [She+04] [LEK20] [Has+20] [WC20], but this threshold is set to 10 in the context of this thesis. The rate of super-spreading events (*S-index*) is therefore associated with the tail of the distribution of P_k. The weight of the tail therefore quantifies the *S-index*.

$$S\text{-}Index = \sum_{k \geqslant 10} P_k. \tag{3.12}$$

Because the *S-index* is determined by both the infectious disease and the temporal network of close contacts, different diseases as well as different social subsystems can be compared.

3.2.3 Model interventions

The modeling approach presented allows the establishment of countermeasures to contain the infectious disease. The WHO [Wor21] recommends wearing masks and keeping a greater physical distance from others during social interactions. Wearing masks lowers the emission of aerosols and thus decreases the probability of transmission,

$$p \rightarrow \mu * p, \tag{3.13}$$

with $0 < \mu < 1$, which leads to

$$P_n = 1 - (1 - p * \mu)^n. \tag{3.14}$$

Social distancing is modeled to the extent that some of the social interactions present in the interaction data are not considered. In detail, with binomial sampling

$$C^*_{ijd} \sim \text{Bin}\left(C_{ijd}; q; C_{ijd}\right), \tag{3.15}$$

given q as the probability that the proximity contact will take place after social distancing.

The graph approach, with its community analysis methods presented in Chapter 2.1.2, also allow the above countermeasures to be applied not only to the total population N of the social system under consideration, but also to the subset $S \subseteq T$. The altered risk of infection can thus be computed for both S and N.

4 Application of methology

In this chapter, the implementation of the individual subsections of the model from Section 3.2 is presented and the technical resources used are discussed in more detail. Subsequently, the infectious diseases *Influenza* and SARS-CoV-2 as well as the mutation from Great Britain SARS-CoV-2-B.1.1.7 are described and the characteristic parameters for the modeling are presented. Finally, the further software-side experimental setup and the performed modeling are described.

4.1 Development stack

The infection model developed in this thesis within the ResTAat project is used for different infectious diseases and different social systems. The implementation is therefore done in such a way that only the underlying interaction data and the infectious disease specific parameters need to be adjusted to perform calculations. An overview of the most important technologies and libraries used is given in Figure 4.1. All computations were performed on a Dell Precision

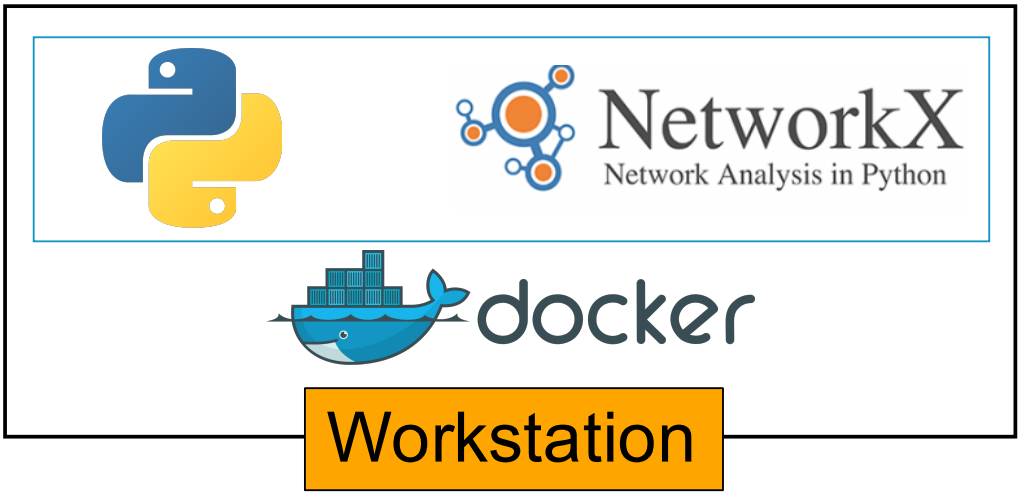

FIG. 4.1: Development stack for the infection model. The operating system on the Dell workstation is *Ubuntu 20.04 LTS*, *Docker* was used in version 19.03.13 and *Python* in version 3.8.3. The versions of all libraries used as well as the complete Docker container is described in appendix B.

7920 Tower workstation with 192 GB RAM and an Intel(R) Xeon(R) Platinum 8260 CPU. Depending on the amount of interaction data, the calculations can be computationally intensive. All applications have been programmed in *Python*.

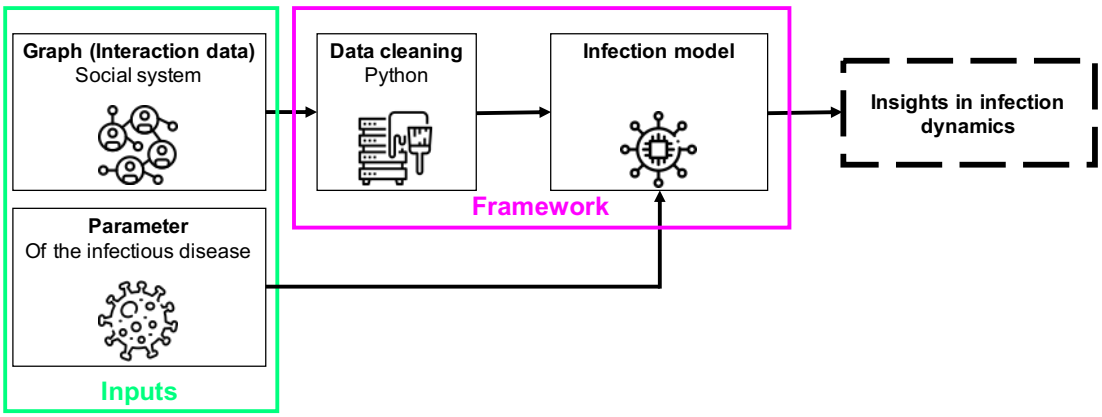

FIG. 4.2: Data processing diagram according to own representation..

To ensure the execution of the model in different hardware environments, all *Python* programs have been executed in *Docker* containers. The data basis was several *.csv* files which contain all logged social interactions. In addition to the libraries *Pandas* and *Numpy*, which are widely used in the field of data science to organize the data, all graph-related calculations were implemented with the library *Networkx*. This offers a seamless integration of *Pandas* DataFrames or *Numpy* arrays and facilitates the use of SNA algorithms.
The process implemented with these tools to determine the risk of infection is shown in Figure 4.2. The framework shown in magenta consists of a data preparation step and the infection model from section 3.2. Outlined in green are the required inputs. The output of the process is information about the infection dynamics within the social system under consideration. These can vary depending on the application and further processing of the data according to the infection model.

4.2 Data cleaning and data processing

Chapter 3.1 has already discussed how a close contact alarm is defined and the basic characteristics of the interaction data of the employees of the Italian production company. In this section, it is described how the raw data is processed and detailed characteristics and topological structures extracted. The individuals in the system under consideration carry identifiable tags. Via the gateways distributed in the enterprise (cf. 3.1), social interactions that exceed thresholds are sent to the backend. In detail, the interaction protocols were recorded in two different buildings of the company. The thresholds here are defined as 1.5 meters between tags and 15 seconds duration. A proximity alarm is therefore sent to the backend when two people are closer than 1.5 meters for more than

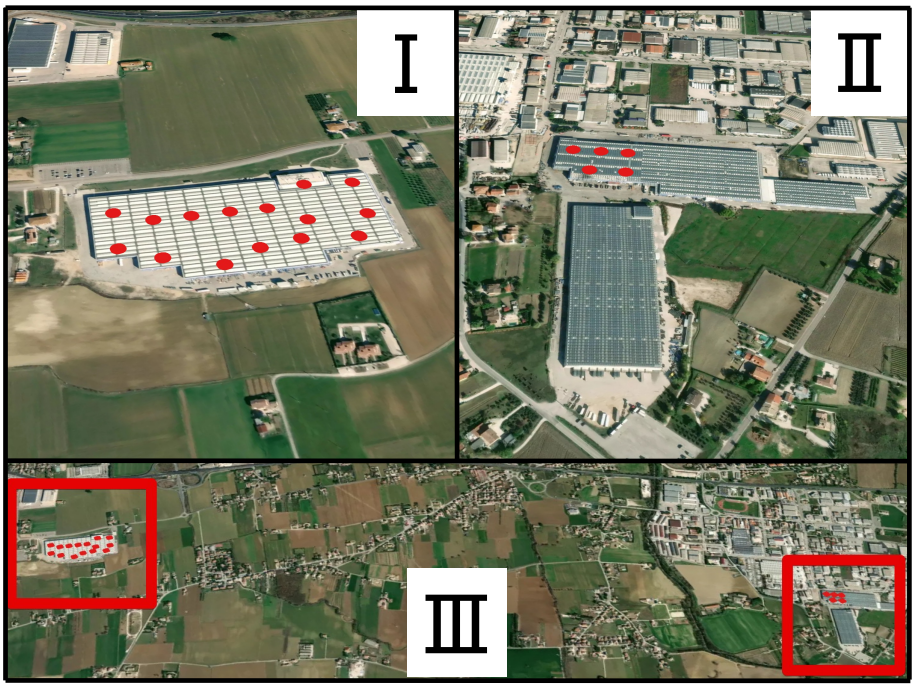

FIG. 4.3: Gateway distribution in the two buildings near the city of Bastia Umbra, Italy (III). The gateway distribution in I seems uniform over the whole area of the building. In image II the gateways are focused on the left part of the building.

15 seconds. An overview of the spatial distribution of the gateways and the two buildings is provided in Figure 4.3, where II and III show a close-up of the two buildings. Since the tags also communicate with the gateways without a second tag being in the vicinity, the raw data also contains entries that did not lead to a close contact alarm. These have to be filtered out if only the near contact alarms are considered. The data provided by *safefactory* was extracted from several databases and includes a total of four *.csv* files. These files must be linked together in a first step. If the data is provided via an Application Programming Interface (API), the same linking steps can be performed. All tags and all gateways contain Unique identifier (UID) numbers, which serve as links between the tables. The four tables can be described as follows:

- **proximity.csv**
  This table contains the complete communication protocol of all tags with all gateways. Both proximity alarms and individual communication between tag and gateway is contained. The following columns are included:
  - **id**: Identifier for the entry in the table.
  - **deviceid**: UID of the tag from which the near contact alarm or the connection to the gateway originates.

  - **devicetime**: Timestamp of the entry.
  - **beaconid**: UID of the connection partner. Contains a different tag UID for near contact alarms and a gateway UID for individual communication (no near contact alarm). This column is used for filtering proximity alarms.
  - **attributes**: Contains an attributsstring of the different hardware parameters and in case of a near contact alarm the UID of the gateway through which the alarm was sent to the backend.

- **beacons_tags_info.csv**
  This table contains the linkage of all UIDs that describe a unique tag like *ISA_4_192* and occur in the *deviceid* column of the *proximity.csv* table. Another column contains an attribute string that is not needed.

- **isa_beacons.csv**
  This table contains the linkage of all UIDs that describe a unique tag like *ISA_4_213* and occur in the *beaconid* column of the *proximity.csv* table. Another column contains an attribute string that is not needed.

- **controller_beacon_info.csv**
  This table contains the link of all UIDs that describe a unique gateway like *bz2131* and occur in the *beaconid* column of the *proximity.csv*. In addition, for each gateway the exact physical position of the gateway is described in the *Longitude* and *Latitude* columns.

Figure 4.4 shows the schematic structure of the tables and the performed joins. The following conditions apply:

1. The *deviceid* column from *proximity.csv* must always contain a tag. Gateways do not initiate a contact, but only appear as a receiver beacon in the *beaconid* column in *proximity.csv*.

2. The *beaconid* column in the *proximity.csv* table contains either a gateway or a tag. Therefore, to connect the UID from this column, the *isa_beacons.csv* and *controller_beacon_info.csv* tables must be joined in a first step.

3. Tables *isa_beacons.csv* and *beacons_tags_info.csv* overlap in some UIDs, but both also contain values that do not appear in the other table.

Through condition 3 the two tables *isa_beacons.csv* and *controller_beacon_info.csv* are concatenated in a first step. The resulting table contains all unique UID values that can occur in the *beaconid* column in the *proximity.csv* table and describe whether it is a tag or a gateway. Figure 4.4 shows an example for a gateway and for a tag in the *proximity.csv* table. This can be recognized either by the UID or by the value in the *attributes* column. In contrast to the tags, gateways do not have a parameter for *appid* in this column, which encodes the gateway through which the transmission was sent. In this example it is gateway *bz2139*.

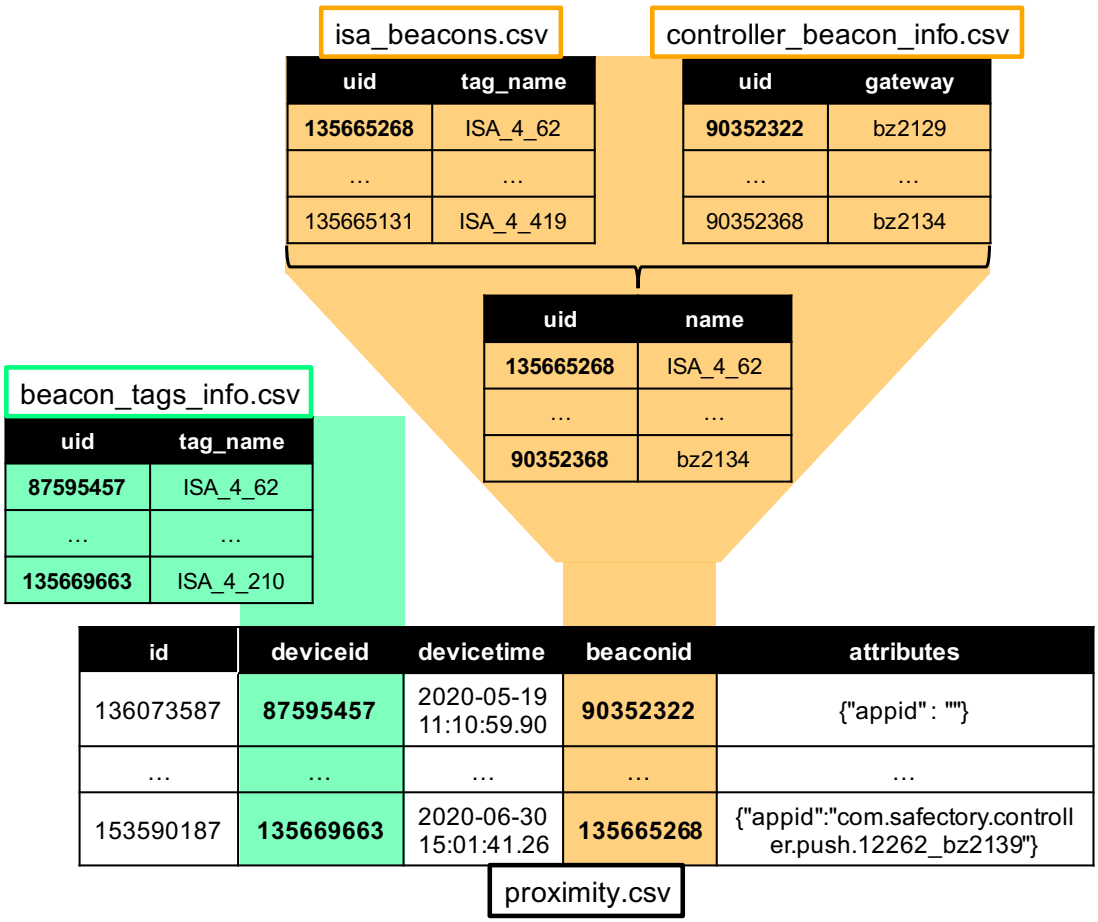

FIG. 4.4: Schematic representation of the table links. All values in columns *deviceid* and *beaconid* in table *proximity.csv* are replaced with the tag and gateway names from the three tables *controller_beacon_info.csv*, *isa_beacons.csv* and *beacons_tags_info.csv*.

Entries like the one with $id = 136073587$ from the *proximity* table are therefore not needed for the later infection model, because they do not represent a near contact alarm, but only a connection from tag to gateway. Therefore, these are filtered out. After filtering the table contains only social interactions that resulted in proximity alerts. These are now aggregated on a daily basis. Table

	amount	date	day	device	beacon
0	53	2020-05-24	145	ISA_4_13	ISA_4_43
1	347	2020-05-25	146	ISA_4_13	ISA_4_43
2	1	2020-05-25	146	ISA_4_133	ISA_4_38
3	2	2020-05-25	146	ISA_4_167	ISA_4_95
4	73	2020-05-25	146	ISA_4_167	ISA_4_143

TAB. 4.1: Pairwise display of proximity alarms caused by two agents each. Next to the number of contacts, the corresponding tag is shown.

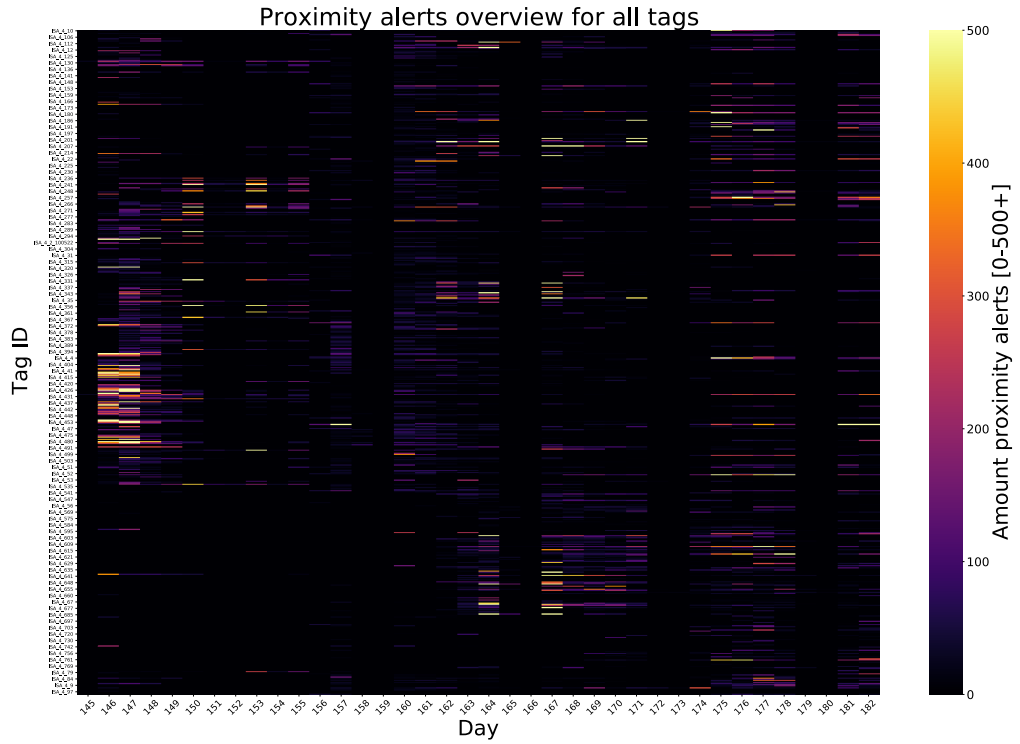

FIG. 4.5: Proximity alarms of the different tags. Y-axis represents only a selection of all unique tag ID. Color scale on the right side shows the number of near contacts per day and ranges from 0 to more than 500.

4.1 shows for each day which individuals triggered how many proximity alarms. For example, tag *ISA_4_13* and tag *ISA_4_43* triggered 53 alarms on 5/24/2020 and 347 on 5/25/2020. A visualization of the alarms triggered per day and tag is shown in Figure 4.5. It can be seen that most tags consistently report proximity alarms over the entire time period. Especially in the first days there is an accumulation of close contacts. Likewise, one can see the weekends when few proximity alarms are recorded. It appears that each tag has many proximity alarms on a few days and few on most days. No tag can be identified in this plot that has triggered an unusually large number of alarms. The data format from Table 4.1 forms the basis for the creation of the interaction graph, respectively

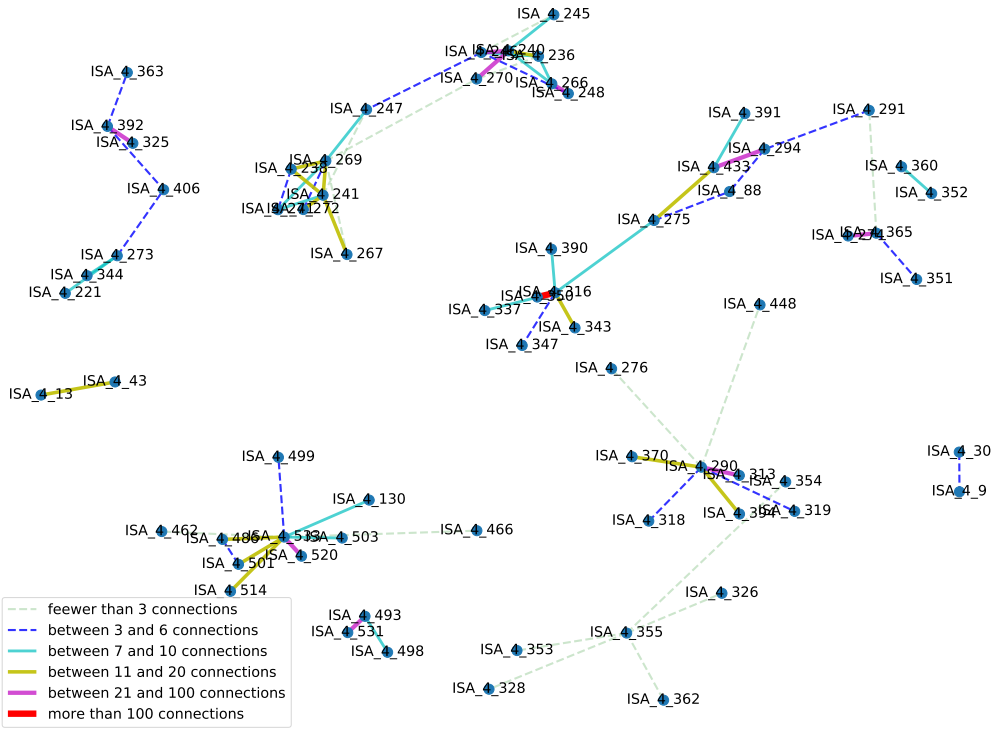

FIG. 4.6: Example of an interaction graph showing all connections between two agents with weighted edges. The value of this weight is based on the number of proximity alarms triggered by a pair of agents. The color coding is shown at the bottom left.

also for the infection model. Here, the weight of the undirected edges of the graph encode the number of social interactions of the respective vertex pair. In the sample visualization 4.6 for 05/30/2020, this weight is illustrated by the color and type of the connection. With respect to the number of proximity alarms, the day shown represents a below average day with 141 connections. Across all days, the average value for proximity alarms is 556 with a standard deviation of 477. However, the distribution of edge weights across vertices is a representative of the entire graph. In this regard, one can see that few connections between two agents have a value greater than 20 and that the graph is divided into several small groups of agents, most of which are not connected to each other. A temporary graph like this can be created for any time interval within the time period under consideration. The next chapter describes how the functions presented in chapter 3.2 are applied to this structure and how both the risk of infection per contact and the distribution of secondary cases are calculated.

4.3 Infection model implementation

The mathematical functions presented in Section 3.2 describe a general framework for calculating the risk of infection per contact based on the reproduction number of an infectious disease and for estimating the resulting secondary cases. Therefore, the implementation is done in a function with several input parameters, which allow both the characterization of the considered infectious disease and include the countermeasures described in function 3.15 and 3.13. The data basis is the network structure of proximity alarms per day with the structure described in Table 4.1. The inputs and outputs of the function are shown in the black box diagram 4.7. In detail, the inputs can be characterized as follows:

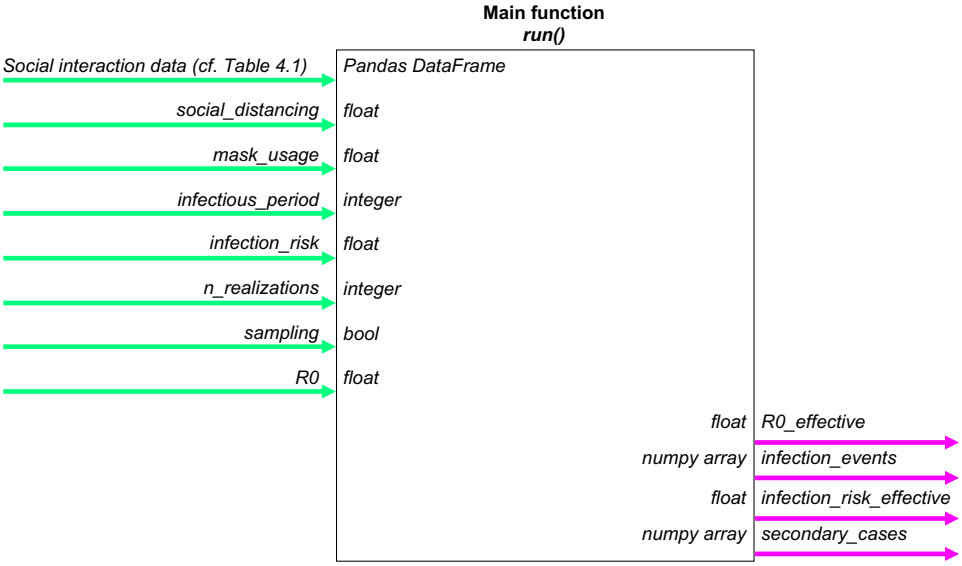

FIG. 4.7: *run()* function to determine the risk of infection per contact and calculate the secondary cases for an infectious disease. The input parameters for characterizing the infectious disease and the framework conditions for the simulation are shown in green. The outputs of the function are shown in magenta. Inside always the type of the parameter.

- **social_distancing:** Implementation of function 3.15. Describes the proportion of social interactions that are considered. Value between 0 and 1, where 1 means all interactions are considered and 0 no interactions are considered.
- **mask_usage:** Implementation of function 3.13. Scaling factor for risk of infection per contact. Value between 0 and 1, where 1 means that the risk of infection is unaffected and 0 means that there is no longer any risk of infection. Default value 1.

- **infectious_period:** Indicates how long infected persons are contagious. Any value greater than or equal to 1. Default value 3.
- **infectious_risk:** Risk of infection per contact. Default value 0.01.
- **n_realizations:** Number of simulations to generate the secondary cases. Default value 1.
- **sampling:** Bool value, specifies whether the infection risk is sampled using the passed R_0 or not. If *True*, then a value for R_0 suitable for the infectious disease must be passed. Default value *False*.
- **R0** Reproduction number for the considered infectious disease. Only used if sampling is *True*. Default value 1.

The output variables of the *run()* function can be described as follows:

- **R0_effective**: Calculated reproductive number based on the infection model and interaction data. Result of the function 3.6.
- **infection_events:** Returns the matrix X_{ijd} (cf. formula 3.9). This contains the information whether the social interaction of the couple (i,j) within the time period d led to an infection or not.
- **infection_risk_effective:** Returns the risk of infection per contact calculated from the disease-specific reproduction number R_0. If *sampling* is *False*, then the default infection risk 0.01 or the passed input infection risk is returned.
- **secondary_cases:** Distribution of secondary cases.

In a first step in the *run()* function, the whole considered time period M is divided into equal time intervals Δ using the *infectious_period* parameter T. As shown in Figure 4.8, the division of the time intervals is done on a daily basis and there are overlaps of the intervals if T is greater than 1. The size of the time intervals and the number of proximity alarms within these time intervals thus depends on M and on T with $[\Delta_0,...,\Delta_{m-T}]$ for $m \in M$.

For each Δ, a temporal network of social interactions is created, which represents all proximity alarms between two individuals within this interval. Importantly, the edges of the resulting undirected graph map the sum of all proximity alarms between a pair. The weighted adjacency matrices of the temporal graphs for $[\Delta_0,...,\Delta_{m-T}]$ are then combined and disaggregated on a per-day basis. The resulting matrix corresponds to D_{ijd} obtained in formula 3.4. The matrix is

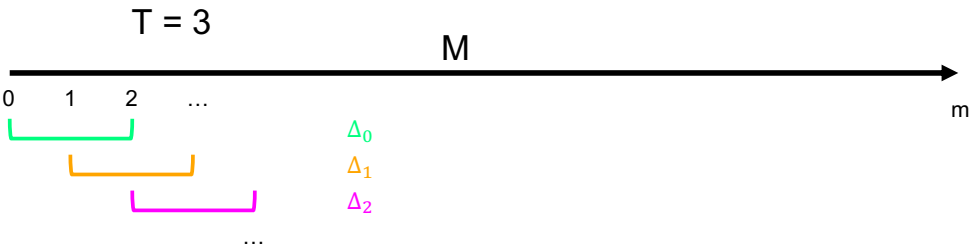

FIG. 4.8: Own representation of overlapping time intervals for total time period M and infection time period $T = 3$.

similar to Table 3.1, but contains pairwise proximity alarms multiple times due to the overlapping time intervals (depending on the infection parameter T).

Approximating the risk of infection per social interaction from a given reproduction number R_0, was implemented using the *fsolve* function from the *SciPy* library and can be computationally intensive depending on the amount of data. Specifically, the function *fsolve* calculates a reproduction number R_M based on the interaction data for different p and matches it with the given reproduction number R_0. This is until

$$R_M \overset{\pm 1.49 * 10^{-8}}{\approx} R_0. \tag{4.1}$$

Social Distancing was implemented using the sample function from the *Pandas* library and returns a random sample of elements from the proximity alarms. The parameter *social_distancing* specifies the fraction of the elements to be returned. Mask wearing is implemented by multiplying the risk of infection per contact p (irrespective if *sampling True* or *False*) by the *mask_usage* parameter.
The detailed implementation of the formulas 3.7, 3.13 and 3.15 is shown in code snippet B.1. These parameters are then used to calculate the risk of infection for each row in the matrix D_{ijd} according to formula 3.2. The *numpy* function *np.where* is used to decide whether an infection event has occurred or not. Here, the previously calculated infection risk for the pair under consideration is compared with the random variable also contained in the row of the matrix. If the random variable is smaller, an infection event occurred, if the random variable is larger, no infection event occurred.

Virus	R_0	R_0, workplace	Infectious period
2019 Influenza	1.7 [Bal+09]	0.34	1 [Cor+12]
SARS-CoV-2	2.4 [Li+20]	0.48	3 [Li+20]
SARS-CoV-2-B.1.1.7	4.8	1.1	4 [Kis+21]

TAB. 4.2: Overview of input parameters: Characteristics for infectious diseases, based on literature references in parentheses. The reproductive number of SARS-CoV-2-B.1.1.7 is reported as 1.74 [Gra+21] or 2.24 [Vol+21] times higher than that of SARS-CoV-2. Therefore, the average 2 between the two reports is used as factor.

The calculated values are entered in the D_{ijd} matrix. The comparison is repeated *n_realizations* times whereby two characteristics are calculated:

- **R0_effective:** Mean value of the sum of all infection events for each realization.
- **secondary_cases:** Mean values for secondary cases per day over all realizations.

R0_effective and *secondary_cases* are then returned together with the infection risk per contact p (*infection_risk_effective*) and the matrix D_{ijd} (*infection_events*). The latter contains the information about infections that have occurred and is the basis for further analyses.

4.4 Infectious disease parameters

In the context of this work, two different infectious diseases are considered - SARS-CoV-2 and *Influenza*. For SARS-CoV-2, both the variant identified at the beginning of 2020 as the trigger of the COVID-19 pandemic and the british variant11, SARS-CoV-2-B.1.1.7, are considered.
For a comparison, the characteristic reproduction number R_0 and the infectious period must be defined in each case in order to be able to determine the disease-specific infection risk per contact by sampling. Furthermore the recorded social interactions of persons take place during working hours. The proportion of contacts at work with respect to the total number of contacts of a person is given as 21% [Mos+08], 25% [FWC12], 16% [Fat+20] and 20% [ETd16]. Therefore, a value of 20% is assumed and consequently the absolute value of the reproduction numbers is also scaled 20% of the original value, which can be obtained in Table 4.2.

5 Evaluation of the results

This chapter presents the results of the infectious diseases studied. For this purpose, the infection probabilities per social interaction calculated with the framework presented in chapter 3 and the resulting secondary cases are presented first 5.1.1. Then, the structure of these secondary cases, the impact on the network and possible countermeasures are presented in 5.1.2. These results are then compared to the literature (5.2) and discussed (5.3) based on this.

5.1 Descriptive representation of results

5.1.1 Diseases

For each of the infectious diseases considered, SARS-CoV-2, *Influenza*, and SARS-CoV-2-B.1.1.7, the risk of infection per social interaction was calculated based on the respective reprodution number presented in Table 4.2. The results obtained are as follows:

Virus	R0, workplace	Infectious period	**transmission rate** p
2019 Influenza	0.34	1	**0.1342**
SARS-CoV-2	0.48	3	**0.0432**
SARS-CoV-2-B.1.1.7	1.1	4	**0.1128**

TAB. 5.1: Summary of transmission rate p per close contact for SARS-CoV-2, *Influenza*, and SARS-CoV-2-B.1.1.7 in relation to the respective reproduction number R_0 and infection period T.

Influenza has the highest risk of infection per social interaction among the three different diseases. This may sound counterintuitive, but this value can be explained by the shorter infection period that implies a higher transmission rate to generate the reproduction number. According to reports from the RKI [RKI21] in April 2021, the proportion of the Corona variant B.1.1.7 in Germany is now 72%. Therefore, the calculated ~ 2.6 times higher transmission rate of the British mutation is an important insight. The three transmission rates determine for each infectious disease the number of secondary cases resulting from the potentially infected persons. Figure 5.1 therefore shows the number of potentially infected individuals for each day (Figure Section I) and the number of secondary

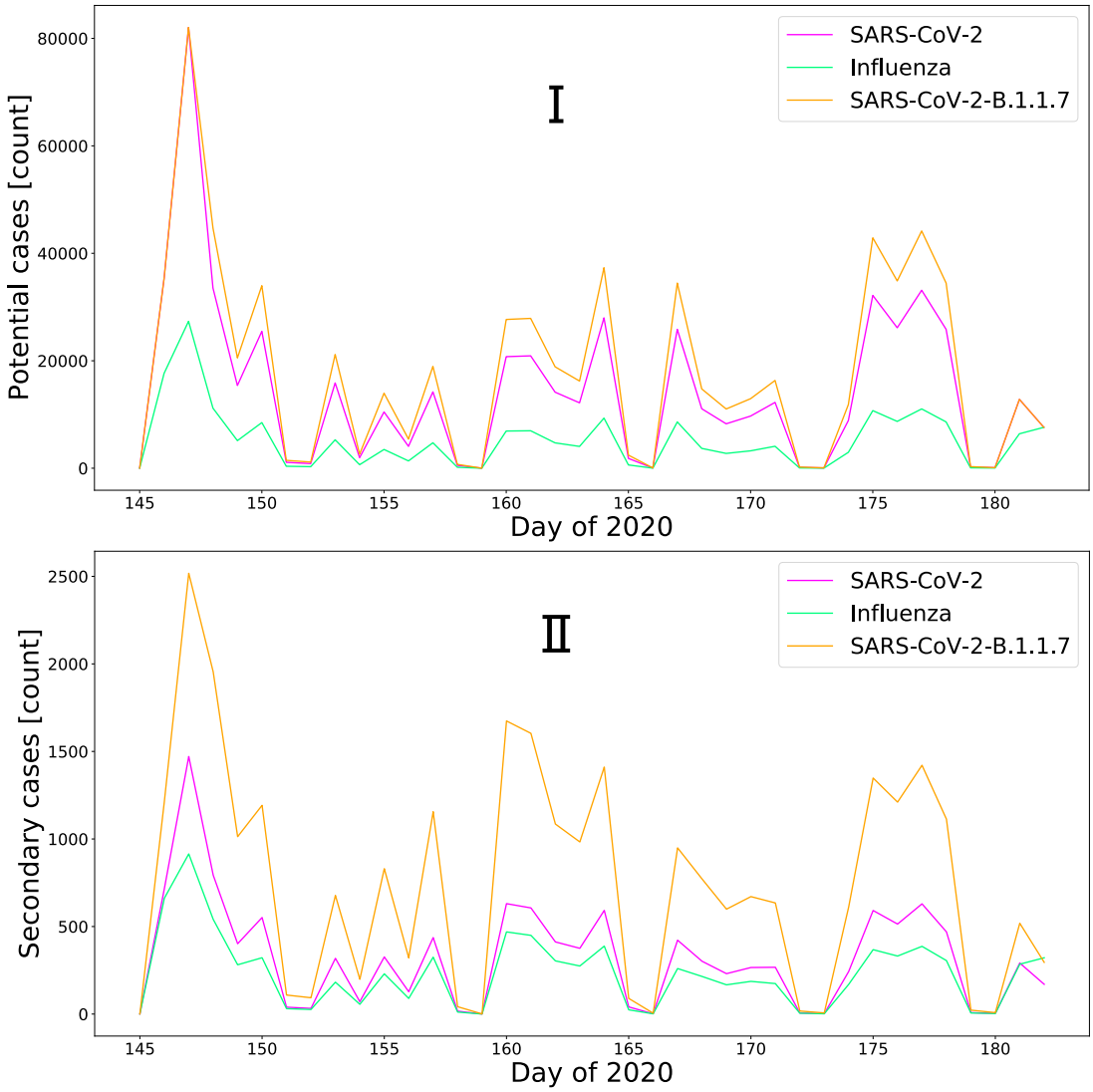

FIG. 5.1: Overview of potential and secondary infection cases. Both plotted over the period considered with the respective number per day on the Y-axis. Directly related to the number of social interactions per day (see Figure 3.2).

cases per day in Figure Section II. Each of the three transmission probabilities results in its own distribution of potentially infected persons and secondary cases. Because of the longer infection period, more people are at risk for SARS-CoV-2 and SARS-CoV-2-B.1.1.7 than for *Influenza*. The one-day longer infection period actually differentiates the UK mutation somewhat from SARS-CoV-2. The spike between days 145 and 150 in both parts of the figure is due to the higher than average proximity alarms during this period (see Figure 3.2). Also seen are the weekends marked in Figure 3.2 with very few potential cases. Intuitively, an average of 12648 potential cases per day sounds like a lot, but distributed among the 621 employees, this corresponds to approximately 5 minutes of close contact per employee per day. This is considered realistic for a production environment in the context of this thesis. In Figure Section II, a correlation of secondary cases

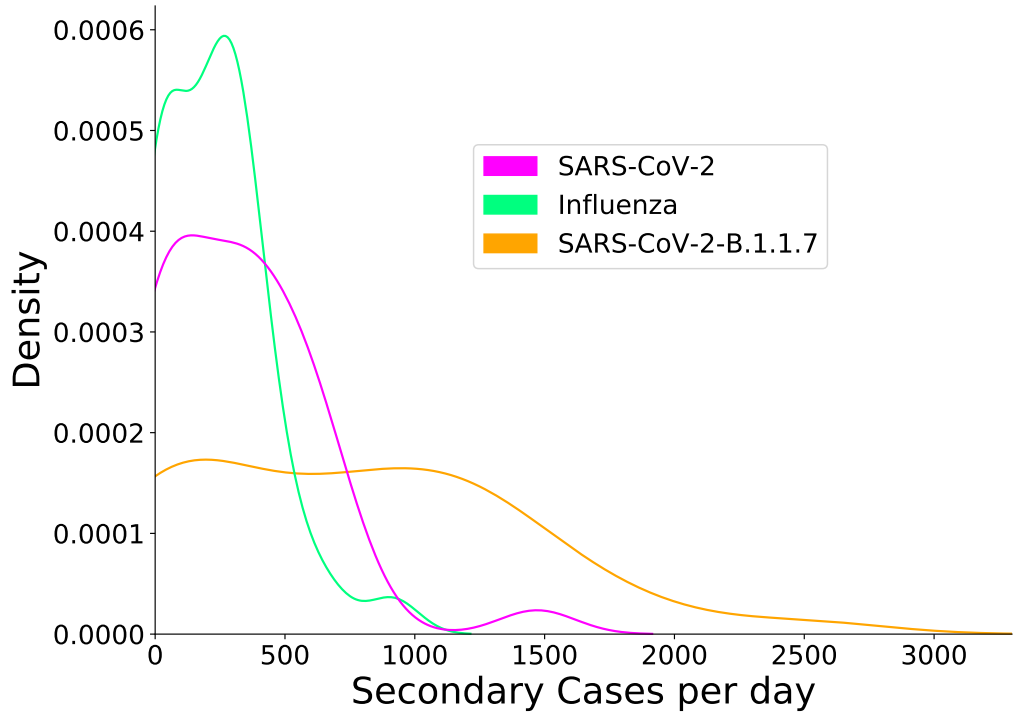

FIG. 5.2: Distribution of secondary cases per day for SARS-CoV-2, SARS-CoV-2-B.1.1.7 and *Influenza*.

to potential cases at risk can be seen, which is to be expected based on the function definitions. However, it is intriguing to note that the UK variant generates significantly more secondary cases than would be expected from Figure Section I. The significantly higher risk of infection identified earlier appears to be reflected here. Figure 5.2 shows that the distributions of secondary cases of the diseases under consideration have fat tails. This is consistent with reports [FF20] of these infectious diseases and means that a very large number of secondary cases were generated on very few days. It appears that SARS-CoV-2 has a slightly higher propensity to generate high secondary cases per day than *Influenza*. However, the difference between the British mutation and the other two diseases is significant. SARS-CoV-2-B.1.1.7 has a flatter distribution resulting in a higher average number of secondary cases per day. Also, there are higher maximum values in the edge of the distribution of SARS-CoV-2-B.1.1.7.

These maximum values above a certain threshold are referred to as Super-Spreading-Event (SSE) according to the definition in Chapter 2.2.1. To quantify these SSEs and to identify Super-Spreader (SS) in the network, a threshold of 10 caused secondary cases per day was set in Equation 3.12. Accordingly, the total number of these SSEs per day describes the *S-index* (Ψ). Therefore, the basis of calculation is no longer the secondary cases per days, but the secondary cases per person per day shown in Figure 5.3. Individuals infected with SARS-CoV-2 and

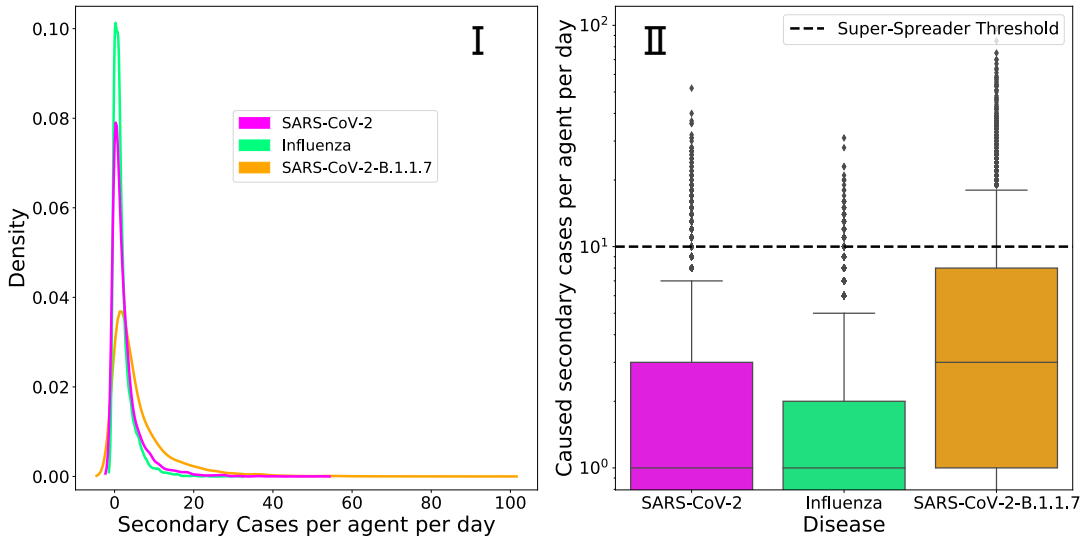

FIG. 5.3: Distribution of secondary cases per day and per agent for SARS-CoV-2, SARS-CoV-2-B.1.1.7 and *Influenza*. In II, the defined threshold of 10 secondary cases per person per day is marked with a dashed line. Also, the length of the whiskers is limited to a maximum of 1.5 times the interquartile range (IQR).

Influenza infect a similar average number of individuals per day (5.3 II). However, the *S-index* of SARS-CoV-2 is 270, more than twice the *S-index* of *Influenza* (107). The broader distribution of the number of secondary cases of the UK mutation means that many of mild outliers (within 1.5×IQR) have caused SSE. Compared with this, only a few outliers infected with *Influenza* and SARS-CoV-2 caused SSE.

These initial findings indicate that while individuals infected with SARS-CoV-2 within the subpopulation under consideration infect on average about the same number of individuals as individuals infected with *Influenza*, the *S-index* of SARS-CoV-2 is twice that of *Influenza*. The British mutation of SARS-CoV-2 is very different from *Influenza* and SARS-CoV-2 because many more people are infected and the *S-index* of this variant (920) is ~ 9 times that of *Influenza* and approximately 4.5 times that of SARS-CoV-2 (see Table 5.2). Another metric is $\overline{\Psi}$, which describes the average size of the SSE. This value, together with Ψ, provides information about whether there are many SSEs with smaller sizes or few SSEs with very many secondary cases. The infectious diseases considered in this report show similar values for $\overline{\Psi}$, but SARS-CoV-2-B.1.1.7 has a larger standard deviation (δ_Ψ). This suggests that although the UK mutation tends to have more SSE with a high number of secondary cases, the significantly higher absolute number of secondary cases of this disease compared with SARS-CoV-2 and *Influenza* is explained by the higher number of SSEs. Not by the severity of the SSEs.

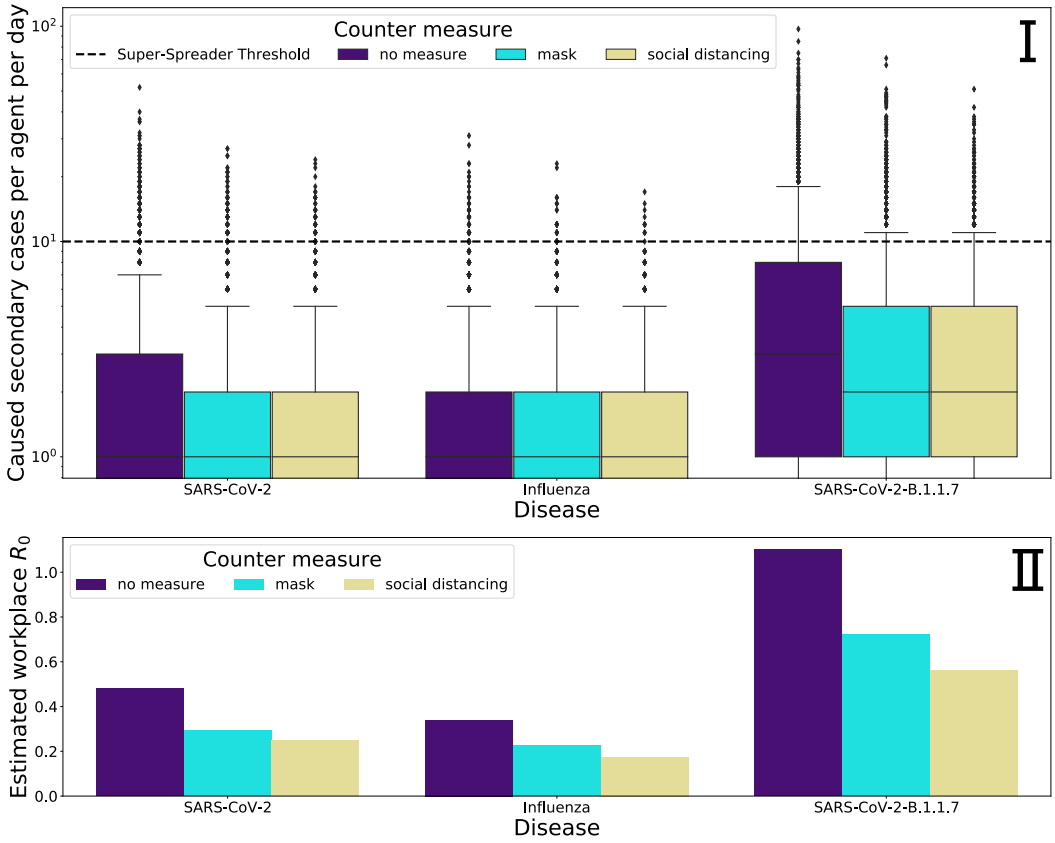

FIG. 5.4: Overview of the effectiveness of countermeasures. All results were calculated on the same interaction data. Part I describes the distributions of secondary cases caused per agent and day. The dashed line marks the set threshold for SSE. II shows the resulting reproduction numbers for the different infectious diseases as a function of the measures implemented. R_0 scaled to the workplace (20%).

5.1.2 Countermeasures

To address this, the effect of countermeasures presented in Formulas 3.13 and 3.15 was examined. Figure 5.4 shows the results. In figure part I, it can be observed that both mask wearing and social distancing lower the number of secondary cases caused per agent per day. It is noticeable that the mean value as well as the third quartile of the two measures taken have similar values for all three diseases. However, as shown in Table 5.2, the measures differ in the outliers, hence in the quantity (Ψ) and magnitude of SSE ($\overline{\Psi}$). For all diseases considered, social distancing leads to the lowest number of SSEs, with the lowest mean. A graphical representation of these distributions is shown in A.2. These differences in the tails of the distributions result in clearly different reproduction numbers of the measures, as seen in part II, although the distributions of the countermeasures in I are very similar. It appears that for the social sys-

	B.1.1.7			SARS-CoV-2			Influenza		
	-	M	SD	-	M	SD	-	M	SD
Ψ	920	540	360	270	124	69	107	36	12
$\overline{\Psi}$	18.86	16.73	15.51	15.51	13.64	12.48	13.09	12.61	11.5
σ_Ψ	10.47	8.27	6.76	6.09	3.82	3.62	3.74	3.08	1.31

TAB. 5.2: Overview of the *S-Index* (Ψ) with $\overline{\Psi}$ as average and σ_Ψ as standard deviation. - indicates no counter measures, *M* represents mask wearing and *SD* represents social distancing.

tem considered, social distancing is the better countermeasure and results in a lower reproduction number for all infectious diseases calculated and therefore a safer work environment for the workforce. Based on these results and the high proportion of the UK SARS-CoV-2 variant in total infections confirmed by recent research [RKI21], only social distancing is used as a countermeasure for further analyses in this thesis and only the british mutation is considered.

5.1.3 Influence of communities and individuals

With these prerequisites in place, the topological structure of the social interaction graph was now examined, agents critical to the system were identified, and targeted countermeasures for parts of the graph were evaluated.

It is suspected that the gateway positions drawn in Figure 4.3 influence the number of proximity alarms recorded. Figure 5.5 confirms this assumption and shows the distribution of alarms among the different gateways in both production halls. In production hall 1, the gateways at the edge of the hall seem to pick up fewer alarms than the more central gateways (Figure part I). One gateway stands out and picked up significantly more alarms than the others. However, for privacy reasons, no further information could be obtained about the reason for the contact clustering at this location in the production hall. Production hall 2 has fewer installed gateways and covers only about half of the total area (Figure part II). It is assumed that the unmonitored area is an area where no interactions can take place. For example, this area could be covered by a high bay or large production machinery.
One gateway in hall 2 has recorded a particularly high number of proximity alarms. Also in comparison to production hall 1 this gateway forms an extreme

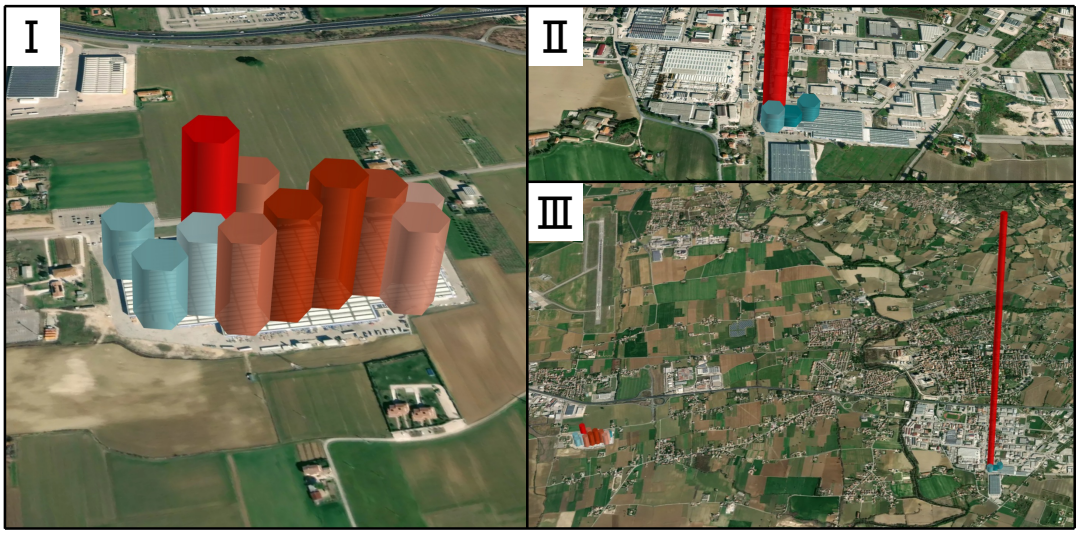

FIG. 5.5: Overview of proximity alarms per gateway. Total number is indicated by the size of the bars and the color. I shows production hall 1. II production hall 2 and III both halls in relation.

value (figure part III). Again for data protection reasons, the exact reason for this extreme value could not be determined at this point. The majority of the alarms were recorded within the first few days of the period under consideration (cf. 5.1). It cannot be ruled out that false measurements or incorrect operation of the Bluetooth tags by the workers occurred.

Apart from this extreme value in Hall 2, it can be concluded on the basis of the results obtained that there are more or less proximity alarms in some areas of the total production area. Since workers usually follow certain patterns in their daily work and normally always deal with a similar group of colleagues, one would expect the formation of a community structure. For verification, the interaction graph over the complete period was divided into different communities using the *Girvan-Newman* method presented in Chapter 2.1.2. The division with the highest calculated modularity of 0.45 is shown in Figure 5.6. The graph is divided into 9 communities. One unconfirmed conjecture would be that the nine communities represent different departments of the company or different shifts. The number of agents per community is evenly distributed between 29 and 115. The total degree of all agents within a community varies from 1 to 139. These results confirm the assumed community structure and show that some communities are more connected through more inter-community connections. Since according to *Saramaki et al.* [SM15] and *Golder et al.* [GM11] social interactions can unfold on many time scales, taking on structures and regularities ranging from changes every minute to annual rhythms, it makes sense to examine temporary structures as well. In the following, the previous findings on communities are

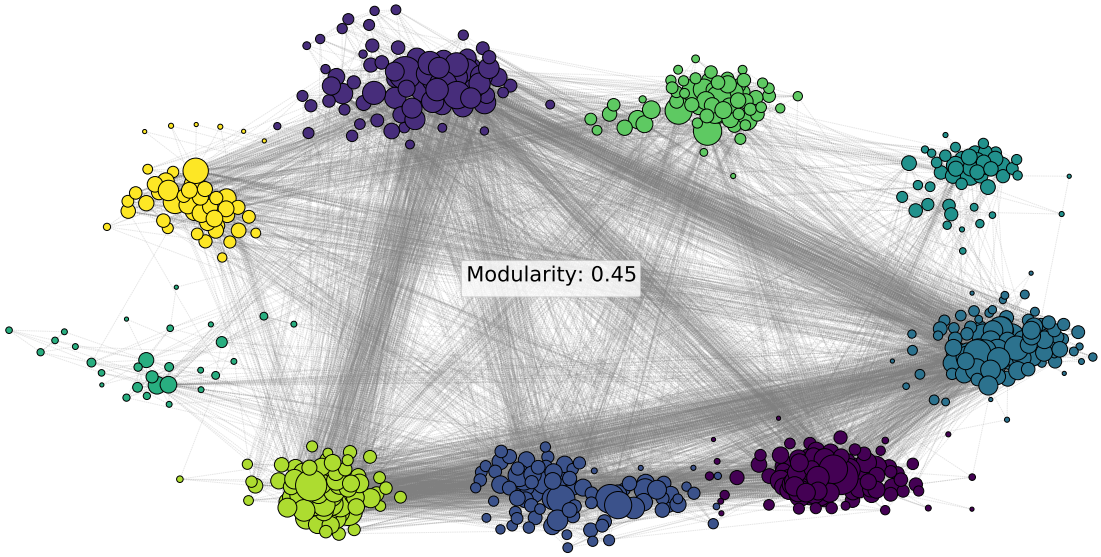

FIG. 5.6: Community structure over the entire period. The different communities are marked in color. The size of the individual vertices indicates the respective degree.

therefore repeated for different temporal intervals and all agents are classified using the SNA metrics from Section 2.1.1.

5.1.4 Temporal structures

Due to the structure of the interaction data, the amount of data and the industrial context, one week (7 days) is chosen as interval size for the investigation of the temporal structures. Figure 5.7 shows the distributions of the temporary graph into communities for each week between *May 24, 2020* and *June 30, 2020*. The weeks start on Mondays and end on Sundays. The structures shown are similar to that of the graph over the entire period, but differ in part in the number of communities which varies from 6 to 9. One can see from the edges drawn in gray that there are weeks on which the inter-community is more pronounced (weeks 22, 24, 26) than on other weeks. In the context of the social system under consideration, the hypothesis is put forward that the communities represent different departments that have more or less contact with each other. Due to the lack of personal information, this proposition cannot be verified.
It can also be observed that the modularity of each week is higher than the modularity for the whole graph (i.e. 5.6). This indicates that the division of the entire period into individual intervals separates the communities better from each other. It is interesting to note that smaller intervals, such as on a daily or hourly basis, resulted in worse modularities. Of course, the optimal interval

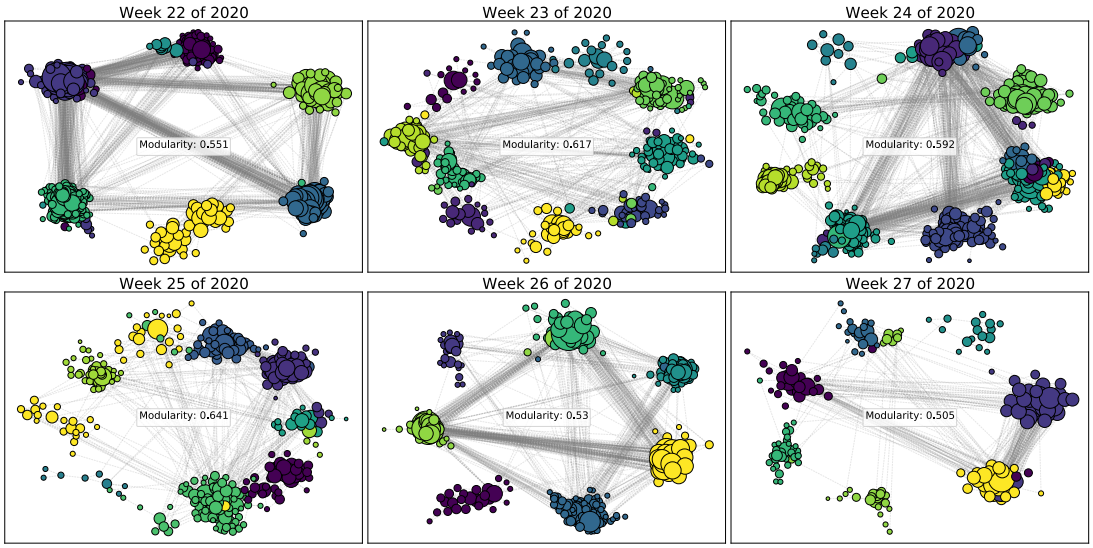

FIG. 5.7: Temporal community structures with the interval size of one week. Communities are again color coded as in 5.6 and the size of each point reflects the degree of different verticies.

size always depends on the data under consideration. Likewise, the community structure by definition depends on the social interactions of the agents within the community which is represented by the size of the respective circle. As previously shown, the risk of infection correlates with the number of contacts, which infers that individuals in a community have a different risk of infection. It follows that there are also differences between communities with respect to the risk of infection. Figure section I in 5.8 shows the different infection risks within communities. The presumed difference between communities with respect to members' risk of infection is confirmed. Because the number of communities per week varies, comparison over multiple weeks is difficult. However, it is noticeable that, for example, Community 7 has an above-average infection risk and Community 1 has a below-average infection risk. If we also look at the average number of individuals within a community (II), we see that Community 8 has the lowest average number of individuals, but generates the highest risk of infection in week 25. Similarly, Community 2 has the highest average number of individuals, but registers a lower than average risk of infection per week. These findings suggest that it is not the number of people per community that matters, but which people are in a community.

To investigate this, the "importance" of each agent within the temporal graph and the respective community was calculated for each time interval. In the context of infectious diseases, the importance means how at risk the person is or what kind of danger this person poses to others. Figure 5.9 shows the

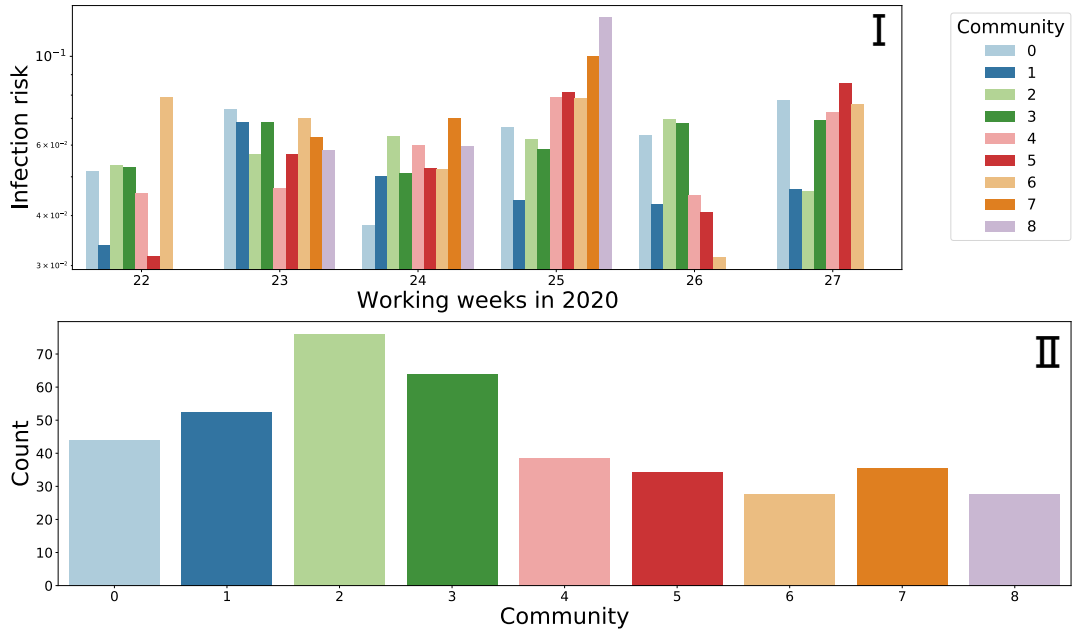

FIG. 5.8: Infection risk per community for each working week (I). Calculated based on the infection events per community arising from the near-contact alerts. II shows the average number of agents per community over the entire period. Color coding of the communities is the same in both parts of the figure.

Week 22

Agent	DC	IX	BC	IX	EC	IX
ISA_4_32	0,059	1	0,361	2	0,12	1
…	…	…	…	…	…	…
ISA_4_67	0,003	259	0,175	241	0,001	251

…

Week 27

Agent	DC	IX	BC	IX	EC	IX
ISA_4_87	0,037	15	0,231	28	0,11	5
…	…	…	…	…	…	…
ISA_4_13	0,034	150	0,075	100	0,016	130

Agent	DC	IX	BC	IX	EC	IX	week	sum
ISA_4_32	0,059	1	0,361	2	0,12	1	22	4
ISA_4_87	0,037	15	0,231	28	0,11	5	27	48
ISA_4_13	0,034	150	0,075	100	0,016	130	27	350
ISA_4_67	0,003	259	0,175	241	0,001	251	22	551

FIG. 5.9: SNA metric calculation for the temporal graphs. Color coded are *Degree Centrality* (DC - magenta), *Betweeness Centrality* (BC - orange) and *Eigenvector Centrality* (EC - sea green). For each metric, column *IX* also shows the ranking position per day. The lower this value, the more important the agent is for this metric. The final table on the right hand side concatenates all the weekly tables together. Column *sum* shows the sum of all *IX* of a row.

experimental setup implemented in *Python*, which is based on the SNA metrics *Degree Centrality*, *Betweeness Centrality* and *Eigenvector Centrality* presented in Chapter 2.1.1. Calculating the average placement per agent and week gives

	Agent ID	Occurrence	Avg. Importance
1	*ISA_4_369*	3	65
2	*ISA_4_392*	3	74.3
3	*ISA_4_760*	1	107
4	*ISA_4_433*	2	112.5
5	*ISA_4_43*	6	118

TAB. 5.3: Excerpt of the most important agents within the social system calculated from the three equally weighted SNA metrics *Degree Centrality*, *Betweeness Centrality*, and *Eigenvetcor Centrality*. The *Occurrence* column indicates in how many weeks the corresponding agent triggered at least one close contact. The entire table captures a ranking of each agent within the system.

the mentioned "importance". Table 5.3 shows the most important agents of the social system calculated using the three equally weighted metrics over the entire period. The average is calculated because not every agent triggers close contacts every week and is only calculated for the weeks in which the respective agent was active in the system. It can be seen that the two most important agents triggered at least 1 proximity alarm only in half of the periods considered. Also, agent *ISA_4_760* occurs only in one of the considered weeks, but in this one it has an important position within the temporal graph. Therefore, the interaction pattern seems to be more decisive than the pure frequency of interactions.
As a final investigation, it was evaluated whether measures that apply exclusively to the most important 50 agents in the system have significant impact on infection dynamics within the system. Such an approach has the advantage in the production environment that the existing processes are less affected by the measures, since most workers are not affected by the measures without ideally being exposed to a higher risk. Only the UK SARS-CoV-2-B.1.1.7 variant is considered and social distancing as a countermeasure. Figure 5.10 shows that targeted countermeasures applied to only about 15% of the agents within the social system can reduce the number of secondary cases. Figure part I shows that secondary cases per day for targeted actions are about halfway between social distancing and no actions. However, it is particularly interesting to see the right margin of the distributions in II. One can see that targeted measures for the most important agents according to SNA metrics, maximum values of secondary cases per day can be reduced by about 25%.

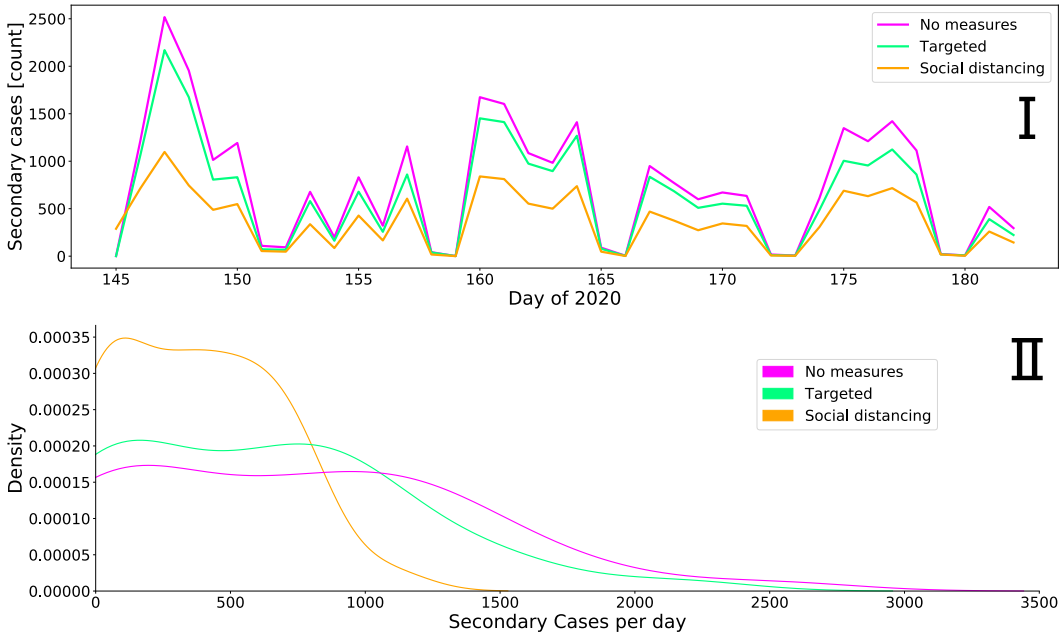

FIG. 5.10: Overview of the effectiveness of countermeasures for SARS-CoV-2-B.1.1.7. Figure I shows the effectiveness of targeted countermeasures compared with no countermeasures and social distancing with reference to secondary cases per day. Figure II shows the distribution of secondary cases for the same categories. This graph can be compared with 5.1 and 5.2.

5.2 Comparison to literature

To better place the results in the context of current research, they are compared with current research results. The calculated transmission probability for SARS-CoV-2 (0.0432) and the UK mutation (0.1128) in the workplace relate specifically to the underlying social system, but are within a range of values used and calculated in various papers $(0.04 - 0.2)$ [Tan+20] [CMA20] [Küh+20]. The approaches by *Kuhn et al.* [Küh+20], for example, divide individuals into age groups and assign each group its own risk of infection. Likewise, other work makes the risk of infection dependent on environmental parameters such as ventilation, number of people in a room, or temperature [BMS20] [Cha+20b]. Although only a uniform risk of infection was considered in this work, the finding that SARS-CoV-2-B.1.1.7 leads to more secondary cases due to the longer infection period is consistent with recent studies by *Kissler et al.* [Kis+21].
With respect to the containment of an infectious disease, to the best of our knowledge, no comparable work could be found comparing social distancing and mask wearing based on interaction data. The demonstrated advantage of social distancing compared to mask wearing therefore needs to be validated on further datasets. However, in social network analysis based epidemiology,

there are studies on the identification of individuals who have a large impact on the infection dynamics within the network. *Manzo et al.* [MR20] were able to show that targeted countermeasures can effectively stop infectious diseases through targeted interventions in a minority of highly networked individuals. Countermeasures were also applied to 15% of individuals. These individuals were selected in the same manner as the study in this paper using SNA metrics. *Salathe et al.* [SJ10] demonstrated that in networks with strong community structure, vaccination interventions targeting individuals that bridge communities are more effective than those targeting only strongly connected individuals. Although communities were identified in the present work, the countermeasures were not targeted to the connectivity indiviudals between them.

5.3 Discussion

This work highlights the importance of social subsystems, such as a workplace in this example, capturing the proximity contact patterns of its members. It also demonstrated how such anonymously collected data can be used to simulate the spread of infectious diseases and associated containment strategies. The design of the framework makes it possible to simulate the spread of a wide variety of infectious diseases transmitted by airborne particles. Only interaction data of the system under consideration and infection-specific parameters such as the reproduction number and the infection time are required for this. Important information about the disease such as the *S-index* or the risk of infection per contact can be estimated. For SARS-CoV-2, the transmission rate is estimated to be 0.0432 and the *S-index* is estimated to be 270. Assuming that the higher reproduction number for the British mutant is confirmed, the much higher infection risk of 0.1128 and the much higher *S-index* of 920 are alarming. However, it was not possible to determine whether these estimated parameters are specific to the social system under consideration or also apply to other subpopulations. This needs to be validated with further data sets. The same applies to the results of the simulated countermeasures. Here, social distancing was found to be more efficient than wearing masks. Reducing contacts by half through social distancing lowers the average *S-index* value to 69 for SARS-CoV-2 and to 360 for SARS-CoV-2-B.1.1.7. In a real social system, the combination of both measures is recommended and is also currently mandated by governments in many countries [Haa20].
Although the simplicity of the model allows for good comparability between

infectious diseases and the evaluation of these differences, it also limits the significance of the absolute resulting values. It has already been demonstrated that SARS-CoV-2 and the UK mutant have different infection dynamics depending on location. For example, social interaction indoors carries on average a significantly higher risk of infection than interaction outdoors. Currently, the presented framework does not take this difference into account. However, the used backend solution of the company *safefactory* offers the possibility to extract and include the exact locations of the interaction as well. It was also assumed in the present work that each individual has an equal risk of infection. This also limits the validity, since according to current studies [Luo+20] [Ada+20], people with lung diseases or older people have a higher risk of infection than healthy, young people. Including a prevalence for infection risk may be one way to address demographics and risk distribution.

Also shown was that the inclusion of network characteristics can lead to more efficient countermeasures. Social distancing applied only to the top 15% individuals identified by SNA metrics was able to reduce the maximum *S-index* values for SARS-CoV-2-B.1.1.7 by about 25%. This is an indication that when resources such as vaccine are scarce at the onset of a previously unknown infectious disease, targeted interventions can be a useful means of infection control. The metrics used for this are standard metrics for SNA and not specifically developed for epidemiology with social networks. Research results from *Salathé et al.* [SJ10] and *Christley et al.* [Chr+05] showed that specific disease metrics can be expected to yield even better results.
Also worth considering in this context is contact duration. Namely, if contact duration is inversely proportional to the number of contacts, then shorter average contact duration could be associated with lower risk of coronavirus infection and spread (assuming transmission probability is negatively correlated with contact duration).

The analysis of temporary structures within the interaction data has shown that the social behavior of employees leads to the formation of different communities. These communities have different infection risks. The exact reason for this could not be determined due to a lack of information about the individuals. However, the assumption is that the increased number of social interactions of different individuals or groups, is due to the tasks within the company. Although different interval sizes for the temporary interaction graphs were tested, whether the selected size of 7 days is the optimal size cannot be evaluated. This always depends on the investigation and the context. More fine-grained groups can

be observed with finer intervals and potentially promote further insights into the topology of the graph. Yet, analysis of the temporal structures within the interaction data was able to show that there is a different risk of infection in different groups of people (communities). In addition, it was shown that the communities within the considered company have different amounts of contact with each other. Eliminating the connections between communities could prove to be a worthwhile research approach to reduce the risk of infection.

6 Conclusion

In this chapter, the results of the work are summarized. Based on the findings, ideas for further work are presented. For this purpose, the research questions and their answers are first discussed in section 6.1. The achieved goals are presented and a compact overview of the results is provided. Finally, in section 6.2 suggestions for future work are presented. These represent, on the one hand, extensions of the framework and, on the other hand, considerations for further investigation possibilities of infection dynamics in social networks.

6.1 Summary of achieved results

1. **How can the risk of infection within a social system be determined on the basis of near-contact interaction data?**

In this work, a framework for modeling the spread of infection in a social system based on social interaction data was presented. Using this model and an infectious disease specific reproduction number R_0 and the infectious period T, the transmission rate of SARS-CoV-2, *Influenza* and the UK mutant of SARS-CoV-2 could be determined within the system under consideration. According to these results, the higher reproduction number of the British mutant compared to the other two diseases is not due to a higher risk of infection, but is primarily due to the longer infection period in this disease. With reference to the available interaction data, the infection dynamics of the three diseases could be simulated and the expected progression of secondary cases estimated. It was clearly seen that the British mutation leads to significantly more secondary cases and in addition the SSE in this disease reach significantly higher maximum values than both other diseases. The framework could also be kept general enough to be used for other airborne infectious diseases. The only condition is the same format of the data basis.

2. **How can information about infection dynamics in a social system be used to derive the effectiveness of different countermeasures?**

Simulations can be performed with the developed model. Different countermeasures lead to different numbers of secondary cases under the use of different parameter sets. Depending on the modeling objective, the distribution

of these secondary cases can be analyzed and the effectiveness of the different countermeasures can be inferred. As a countermeasure for the spread of infection in the system under consideration, social distancing proved to be effective, especially with regard to the reduction of the maximum values of SSE. Furthermore, the temporal structures of the graph were investigated by the simulations performed. With a one-week interval size, different communities of the graph could be identified which are stronger or weaker connected to each other and have internally different infection risks for the members. Further possible countermeasure are therefore the separation of the connections between the individual communities or countermeasures which affect only individual communities.

3. **Are there individuals or a group of individuals within the system who are particularly at risk of infection due to their contact patterns or who have a particularly negative impact on infection dynamics? If so, how can these be identified?**

It has been shown that SNA metrics such as *Degree Centrality*, *Betweeness Centrality* and *Eigenvetcor Centrality* can identify individuals important for the network structure and that countermeasures targeting this small fraction of individuals in the graph can be effective. Especially with respect to vaccination, this could be a promising approach to achieve a large impact with few resources and to slow down the infection dynamics. The identified nodes could also be used as an "early warning system" due to their intersectional position within the system and indicate a possible new infection. More fundamentally, quantifying the influence of single individuals on the entire network or the complete (temporary) graph, however, allows decision-makers to better understand what is happening in terms of infection within the company. Reactive as well as preventive decisions can be made. Ideally, this lowers the risk of infection and protects the members of the system from infection.

6.2 Continuing work

The presented framework calculates the risk of infection and thus the occurrence of secondary cases based on the number of social interactions. A research approach could therefore be to enrich the model with further parameters. This could include environmental parameters of the exact location of the interaction (similar to *Buonnano et al.* [BMS20]) or the inclusion of infection-specific temporary prevalences. It is expected that this will allow the risk of infection to be modeled more accurately and the behavioral patterns of individuals to be mapped even better. This would lead to a more accurate fit of the model to the social system under consideration. Furthermore, the previously described approach of community-based countermeasures offers possibilities to extend the infection model. An approach similar to that of *Salathé et al.* [SJ10] could be considered. In general, research findings by *Manzo et al.* [Man20] show that complex social networks explore sophisticated interventions that target specific categories or groups of individuals and are expected to have collective benefits. Identifying a new metric to determine the most important nodes within the network, comparable to the *Christley et al.* [Chr+05], also offered potential for further research. The combination of temporal structures of the graph and infection dynamics in particular offers an exciting field. Results in this direction could greatly improve the message about the utility of countermeasures in specifically considered social subsystems.
Another approach could be to apply the existing model to a different data set and compare the results. To enable this comparison it is important that the same hardware solution is used and the same parameters for the infectious diseases are used.

References

[AD08] A. H. Auchincloss & A. V. Diez Roux. "A New Tool for Epidemiology: The Usefulness of Dynamic-Agent Models in Understanding Place Effects on Health". In: *American Journal of Epidemiology* 168.1 (May 2008), pp. 1–8. ISSN: 0002-9262, 1476-6256. DOI: 10.1093/aje/kwn118.

[Ada+20] D. Adam; P. Wu; J. Wong; E. Lau; T. Tsang; S. Cauchemez; G. Leung & B. Cowling. *Clustering and Superspreading Potential of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infections in Hong Kong*. Preprint. In Review, May 2020. DOI: 10.21203/rs.3.rs-29548/v1.

[Agg+20] A. Aggarwal; A. Hwang; N. Jain & A. Goel. "Productionizing a COVID-19 System Dynamics Model for the Indian Government". In: (June 2020).

[Agg11] C. C. Aggarwal, ed. *Social Network Data Analytics*. eng. New York, NY: Springer, 2011. ISBN: 978-1-4419-8461-6 978-1-4419-8462-3.

[And+20] R. M. Anderson; H. Heesterbeek; D. Klinkenberg & T. D. Hollingsworth. "How Will Country-Based Mitigation Measures Influence the Course of the COVID-19 Epidemic?" en. In: *The Lancet* 395.10228 (Mar. 2020), pp. 931–934. ISSN: 01406736. DOI: 10.1016/S0140-6736(20)30567-5.

[Bal+09] D. Balcan; H. Hu; B. Goncalves; P. Bajardi; C. Poletto; J. J. Ramasco; D. Paolotti; N. Perra; M. Tizzoni; W. Van den Broeck; V. Colizza & A. Vespignani. "Seasonal Transmission Potential and Activity Peaks of the New Influenza A(H1N1): A Monte Carlo Likelihood Analysis Based on Human Mobility". en. In: *BMC Medicine* 7.1 (Dec. 2009), p. 45. ISSN: 1741-7015. DOI: 10.1186/1741-7015-7-45.

[Ber05] H. Bernard. "The Development of Social Network Analysis: A Study in the Sociology of Science, Linton c. Freeman. Empirical Press, Vancouver, BC (2004)". In: *Social Networks* 27 (Oct. 2005), pp. 377–384. DOI: 10.1016/j.socnet.2005.06.004.

[BG11] S. Blower & M.-H. Go. "The Importance of Including Dynamic Social Networks When Modeling Epidemics of Airborne Infections: Does Increasing Complexity Increase Accuracy?" en. In: *BMC Medicine* 9.1 (Dec. 2011), p. 88. ISSN: 1741-7015. DOI: 10.1186/1741-7015-9-88.

[BMS20] G. Buonanno; L. Morawska & L. Stabile. "Quantitative Assessment of the Risk of Airborne Transmission of SARS-CoV-2 Infection: Prospective and Retrospective Applications". en. In: *Environment International* 145 (Dec. 2020), p. 106112. ISSN: 01604120. DOI: 10.1016/j.envint.2020.106112.

[Bon02] E. Bonabeau. "Agent-Based Modeling: Methods and Techniques for Simulating Human Systems". In: *Proceedings of the National Academy of Sciences* 99.Supplement 3 (May 2002), pp. 7280–7287. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.082080899.

[Cau+11] S. Cauchemez; A. Bhattarai; T. L. Marchbanks; R. P. Fagan; S. Ostroff; N. M. Ferguson; D. Swerdlow; the Pennsylvania H1N1 working group; S. V. Sodha; M. E. Moll; F. J. Angulo; R. Palekar; W. R. Archer & L. Finelli. "Role of Social Networks in Shaping Disease Transmission during a Community Outbreak of 2009 H1N1 Pandemic Influenza". en. In: *Proceedings of the National Academy of Sciences* 108.7 (Feb. 2011), pp. 2825–2830. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1008895108.

[CCW10] J. Cajka; P. Cooley & W. Wheaton. *Attribute Assignment to a Synthetic Population in Support of Agent-Based Disease Modeling*. en. Tech. rep. Research Triangle Park, NC: RTI Press, Sept. 2010. DOI: 10.3768/rtipress.2010.mr.0019.1009.

[Cen+21] G. Cencetti; G. Santin; A. Longa; E. Pigani; A. Barrat; C. Cattuto; S. Lehmann; M. Salathé & B. Lepri. "Digital Proximity Tracing on Empirical Contact Networks for Pandemic Control". en. In: *Nature Communications* 12.1 (Dec. 2021), p. 1655. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21809-w.

[CF10] N. A. Christakis & J. H. Fowler. "Social Network Sensors for Early Detection of Contagious Outbreaks". en. In: *PLoS ONE* 5.9 (Sept. 2010). Ed. by O. Sporns, e12948. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0012948.

[Cha+20a] S. Chang; E. Pierson; P. W. Koh; J. Gerardin; B. Redbird; D. Grusky & J. Leskovec. "Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening". en. In: *Nature* (Nov. 2020). ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-020-2923-3.

[Cha+20b] S. L. Chang; N. Harding; C. Zachreson; O. M. Cliff & M. Prokopenko. "Modelling Transmission and Control of the COVID-19 Pandemic

in Australia". In: *Nature Communications* 11.1 (Dec. 2020), p. 5710. ISSN: 2041-1723. DOI: 10.1038/s41467-020-19393-6. arXiv: 2003.10218.

[Chr+05] R. M. Christley; G. L. Pinchbeck; R. G. Bowers; D. Clancy; N. P. French; R. Bennett & J. Turner. "Infection in Social Networks: Using Network Analysis to Identify High-Risk Individuals". In: *American Journal of Epidemiology* 162.10 (Sept. 2005), pp. 1024–1031. ISSN: 0002-9262. DOI: 10.1093/aje/kwi308. eprint: https://academic.oup.com/aje/article-pdf/162/10/1024/143045/kwi308.pdf.

[CMA20] I. Cooper; A. Mondal & C. G. Antonopoulos. "A SIR Model Assumption for the Spread of COVID-19 in Different Communities". en. In: *Chaos, Solitons & Fractals* 139 (Oct. 2020), p. 110057. ISSN: 09600779. DOI: 10.1016/j.chaos.2020.110057.

[CMM10] J. Chen; A. Marathe & M. Marathe. "Coevolution of Epidemics, Social Networks, and Individual Behavior: A Case Study". en. In: *Advances in Social Computing*. Ed. by D. Hutchison; T. Kanade; J. Kittler; J. M. Kleinberg; F. Mattern; J. C. Mitchell; M. Naor; O. Nierstrasz; C. Pandu Rangan; B. Steffen; M. Sudan; D. Terzopoulos; D. Tygar; M. Y. Vardi; G. Weikum; S.-K. Chai; J. J. Salerno & P. L. Mabry. Vol. 6007. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 218–227. ISBN: 978-3-642-12078-7 978-3-642-12079-4. DOI: 10.1007/978-3-642-12079-4_28.

[Cor+12] A. Cori; A. Valleron; F. Carrat; G. Scalia Tomba; G. Thomas & P. Boëlle. "Estimating Influenza Latency and Infectious Period Durations Using Viral Excretion Data". en. In: *Epidemics* 4.3 (Aug. 2012), pp. 132–138. ISSN: 17554365. DOI: 10.1016/j.epidem.2012.06.001.

[Cue20] E. Cuevas. "An Agent-Based Model to Evaluate the COVID-19 Transmission Risks in Facilities". en. In: *Computers in Biology and Medicine* 121 (June 2020), p. 103827. ISSN: 00104825. DOI: 10.1016/j.compbiomed.2020.103827.

[Dan+11] L. Danon; A. P. Ford; T. House; C. P. Jewell; M. J. Keeling; G. O. Roberts; J. V. Ross & M. C. Vernon. "Networks and the Epidemiology of Infectious Disease". en. In: *Interdisciplinary Perspectives on Infectious Diseases* 2011 (2011), pp. 1–28. ISSN: 1687-708X, 1687-7098. DOI: 10.1155/2011/284909.

[DBU12] C. Derksen; C. Branki & R. Unland. "A Framework for Agent-Based Simulations of Hybrid Energy Infrastructures". In: *2012 Federated*

Conference on Computer Science and Information Systems (FedCSIS). 2012, pp. 1293–1299.

[DDG20] E. Dong; H. Du & L. Gardner. "An Interactive Web-Based Dashboard to Track COVID-19 in Real Time". en. In: *The Lancet Infectious Diseases* 20.5 (May 2020), pp. 533–534. ISSN: 14733099. DOI: 10.1016/S1473-3099(20)30120-1.

[de 09] W. de Nooy. "Social Network Analysis, Graph Theoretical Approaches To". en. In: *Encyclopedia of Complexity and Systems Science*. Ed. by R. A. Meyers. New York, NY: Springer New York, 2009, pp. 8231–8245. ISBN: 978-0-387-75888-6 978-0-387-30440-3. DOI: 10.1007/978-0-387-30440-3_488.

[Del+10] S. A. Delre; W. Jager; T. H. A. Bijmolt & M. A. Janssen. "Will It Spread or Not? The Effects of Social Influences and Network Topology on Innovation Diffusion". en. In: *Journal of Product Innovation Management* 27.2 (Mar. 2010), pp. 267–282. ISSN: 07376782, 15405885. DOI: 10.1111/j.1540-5885.2010.00714.x.

[DH20] C. Drosten & K. Henning. "WISSENSCHAFTSREDAKTEURIN, NDR INFO". de. In: (May 2020), p. 10.

[Eir+18] M. Eirinaki; J. Gao; I. Varlamis & K. Tserpes. "Recommender Systems for Large-Scale Social Networks: A Review of Challenges and Solutions". In: *Future Generation Computer Systems* 78 (2018), pp. 413–418. ISSN: 0167-739X. DOI: 10.1016/j.future.2017.09.015.

[ElS+12] A. M. El-Sayed; P. Scarborough; L. Seemann & S. Galea. "Social Network Analysis and Agent-Based Modeling in Social Epidemiology". In: *Epidemiologic Perspectives & Innovations* 9.1 (2012), p. 1. ISSN: 1742-5573. DOI: 10.1186/1742-5573-9-1.

[Eps+08] J. M. Epstein; J. Parker; D. Cummings & R. A. Hammond. "Coupled Contagion Dynamics of Fear and Disease: Mathematical and Computational Explorations". en. In: *PLoS ONE* 3.12 (Dec. 2008). Ed. by A. P. Galvani, e3955. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0003955.

[Eps09] J. M. Epstein. "Modelling to Contain Pandemics". en. In: *Nature* 460.7256 (Aug. 2009), pp. 687–687. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/460687a.

[ETd16] C. H. Edwards; G. S. Tomba & B. F. de Blasio. "Influenza in Workplaces: Transmission, Workers' Adherence to Sick Leave Advice and European Sick Leave Recommendations". en. In: *The European Journal of Public Health* 26.3 (June 2016), pp. 478–485. ISSN: 1101-1262, 1464-360X. DOI: 10.1093/eurpub/ckw031.

[Eub+04] S. Eubank; H. Guclu; V. S. Anil Kumar; M. V. Marathe; A. Srinivasan; Z. Toroczkai & N. Wang. "Modelling Disease Outbreaks in Realistic Urban Social Networks". en. In: *Nature* 429.6988 (May 2004), pp. 180–184. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature02541.

[Fat+20] P. Fateh-Moghadam; L. Battisti; S. Molinaro; S. Fontanari; G. Dallago; N. Binkin & M. Zuccali. *Contact Tracing during Phase I of the COVID-19 Pandemic in the Province of Trento, Italy: Key Findings and Recommendations*. en. Preprint. Epidemiology, July 2020. DOI: 10.1101/2020.07.16.20127357.

[Fer+05] N. Ferguson; D. Cummings; S. Cauchemez; C. Fraser; S. Riley; A. Meeyai; S. Iamsirithaworn & D. Burke. "Strategies for Containing an Emerging Influenza Pandemic in Southeast Asia". In: *Nature* 437 (Jan. 2005), pp. 209–214.

[Fer+20a] N. Ferguson; D. Laydon; G. Nedjati Gilani; N. Imai; K. Ainslie; M. Baguelin; S. Bhatia; A. Boonyasiri; Z. Cucunuba Perez; G. Cuomo-Dannenburg; A. Dighe; I. Dorigatti; H. Fu; K. Gaythorpe; W. Green; A. Hamlet; W. Hinsley; L. Okell; S. Van Elsland; H. Thompson; R. Verity; E. Volz; H. Wang; Y. Wang; P. Walker; P. Winskill; C. Whittaker; C. Donnelly; S. Riley & A. Ghani. *Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand*. en. Tech. rep. Imperial College London, Mar. 2020. DOI: 10.25561/77482.

[Fer+20b] N. Ferguson; D. Laydon; G. Nedjati Gilani; N. Imai; K. Ainslie; M. Baguelin; S. Bhatia; A. Boonyasiri; Z. Cucunuba Perez; G. Cuomo-Dannenburg; A. Dighe; I. Dorigatti; H. Fu; K. Gaythorpe; W. Green; A. Hamlet; W. Hinsley; L. Okell; S. Van Elsland; H. Thompson; R. Verity; E. Volz; H. Wang; Y. Wang; P. Walker; P. Winskill; C. Whittaker; C. Donnelly; S. Riley & A. Ghani. *Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and*

Healthcare Demand. Tech. rep. Imperial College London, Mar. 2020. DOI: 10.25561/77482.

[FF20] M. Fukui & C. Furukawa. *Power Laws in Superspreading Events: Evidence from Coronavirus Outbreaks and Implications for SIR Models*. Preprint. Epidemiology, June 2020. DOI: 10.1101/2020.06.11.20128058.

[Fon+21] A. Fontanet; B. Autran; B. Lina; M. P. Kieny; S. S. A. Karim & D. Sridhar. "SARS-CoV-2 Variants and Ending the COVID-19 Pandemic". en. In: *The Lancet* (Feb. 2021), S0140673621003706. ISSN: 01406736. DOI: 10.1016/S0140-6736(21)00370-6.

[For06] S. Fortunato. "Resolution Limit in Community Detection". en. In: (July 2006), p. 6.

[For10] S. Fortunato. "Community Detection in Graphs". en. In: *Physics Reports* 486.3-5 (Feb. 2010), pp. 75–174. ISSN: 03701573. DOI: 10.1016/j.physrep.2009.11.002.

[Fre77] L. C. Freeman. "A Set of Measures of Centrality Based on Betweenness". In: *Sociometry* 40.1 (Mar. 1977), p. 35. ISSN: 00380431. DOI: 10.2307/3033543.

[FWC12] Y.-c. Fu; D.-W. Wang & J.-H. Chuang. "Representative Contact Diaries for Modeling the Spread of Infectious Diseases in Taiwan". en. In: *PLoS ONE* 7.10 (Oct. 2012). Ed. by P. Holme, e45113. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0045113.

[GM11] S. A. Golder & M. W. Macy. "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength across Diverse Cultures". In: *Science* 333.6051 (2011), pp. 1878–1881. ISSN: 0036-8075. DOI: 10.1126/science.1202775. eprint: https://science.sciencemag.org/content/333/6051/1878.full.pdf.

[Gör+13] R. Görke; P. Maillard; A. Schumm; C. Staudt & D. Wagner. "Dynamic Graph Clustering Combining Modularity and Smoothness". In: *ACM Journal of Experimental Algorithmics* 18 (Apr. 2013). ISSN: 1084-6654. DOI: 10.1145/2444016.2444021.

[Gra+21] F. Grabowski; G. Preibisch; S. Giziński; M. Kochańczyk & T. Lipniacki. "SARS-CoV-2 Variant of Concern 202012/01 Has about Twofold Replicative Advantage and Acquires Concerning Mutations". en. In: *Viruses* 13.3 (Mar. 2021), p. 392. ISSN: 1999-4915. DOI: 10.3390/v13030392.

[Gra14] M. Grandjean. “La Connaissance Est Un Réseau. Perspective Sur l’organisation Archivistique et Encyclopédique”. In: *Les cahiers du numérique* 10.3 (Sept. 2014), pp. 37–54. ISSN: 14693380. DOI: 10.3166/lcn.10.3.37-54.

[Haa20] W. Haas. “RKI - Navigation - Infektionsschutzmaßnahmen (Stand: 18.11.2020)”. In: (Nov. 20).

[Has+20] A. Hasan; H. Susanto; M. F. Kasim; N. Nuraini; B. Lestari; D. Triany & W. Widyastuti. “Superspreading in Early Transmissions of COVID-19 in Indonesia”. en. In: *Scientific Reports* 10.1 (Dec. 2020), p. 22386. ISSN: 2045-2322. DOI: 10.1038/s41598-020-79352-5.

[Hel+20] J. Hellewell; S. Abbott; A. Gimma; N. I. Bosse; C. I. Jarvis; T. W. Russell; J. D. Munday; A. J. Kucharski; W. J. Edmunds; S. Funk; R. M. Eggo; F. Sun; S. Flasche; B. J. Quilty; N. Davies; Y. Liu; S. Clifford; P. Klepac; M. Jit; C. Diamond; H. Gibbs & K. van Zandvoort. “Feasibility of Controlling COVID-19 Outbreaks by Isolation of Cases and Contacts”. en. In: *The Lancet Global Health* 8.4 (Apr. 2020), e488–e496. ISSN: 2214109X. DOI: 10.1016/S2214-109X(20)30074-7.

[HKO21] H. L. Hambridge; R. Kahn & J.-P. Onnela. *Examining SARS-COV-2 Interventions in Residential Colleges Using an Empirical Network*. en. Preprint. Infectious Diseases (except HIV/AIDS), Mar. 2021. DOI: 10.1101/2021.03.09.21253198.

[HS20] H. A. Herrmann & J.-M. Schwartz. “Why COVID-19 Models Should Incorporate the Network of Social Interactions”. In: *Physical Biology* 17.6 (Oct. 2020), p. 065008. ISSN: 1478-3975. DOI: 10.1088/1478-3975/aba8ec.

[HS91] J. B. Homer & C. L. St. Clair. “A Model of HIV Transmission through Needle Sharing”. en. In: *Interfaces* 21.3 (June 1991), pp. 26–49. ISSN: 0092-2102, 1526-551X. DOI: 10.1287/inte.21.3.26.

[Ivo+20] B. Ivorra; M. Ferrández; M. Vela-Pérez & A. Ramos. “Mathematical Modeling of the Spread of the Coronavirus Disease 2019 (COVID-19) Taking into Account the Undetected Infections. The Case of China”. en. In: *Communications in Nonlinear Science and Numerical Simulation* 88 (Sept. 2020), p. 105303. ISSN: 10075704. DOI: 10.1016/j.cnsns.2020.105303.

[Jay+20] M. Jayaweera; H. Perera; B. Gunawardana & J. Manatunge. "Transmission of COVID-19 Virus by Droplets and Aerosols: A Critical Review on the Unresolved Dichotomy". en. In: *Environmental Research* 188 (Sept. 2020), p. 109819. ISSN: 00139351. DOI: 10.1016/j.envres.2020.109819.

[Jia+20] J. Jia; J. Ding; S. Liu; G. Liao; J. Li; B. Duan; G. Wang & R. Zhang. "Modeling the Control of COVID-19: Impact of Policy Interventions and Meteorological Factors". In: *arXiv:2003.02985 [math, q-bio]* (Mar. 2020). arXiv: 2003.02985 [math, q-bio].

[KH21] M. Kriegel & A. Hartmann. "Covid-19 Ansteckung über Aerosolpartikel – vergleichende Bewertung von Innenräumen hinsichtlich des situationsbedingten R-Wertes". de. In: (Feb. 2021). DOI: 10.14279/DEPOSITONCE-11387.

[Kis+21] S. M. Kissler; J. R. Fauver; C. Mack; C. G. Tai; M. I. Breban; A. E. Watkins; R. M. Samant; D. J. Anderson; D. D. Ho; N. D. Grubaugh & Y. H. Grad. *Densely Sampled Viral Trajectories Suggest Longer Duration of Acute Infection with B.1.1.7 Variant Relative to Non-B.1.1.7 SARS-CoV-2*. en. Preprint. Epidemiology, Feb. 2021. DOI: 10.1101/2021.02.16.21251535.

[Klo+94] A. Klovdahl; J. Potterat; D. Woodhouse; J. Muth; S. Muth & W. Darrow. "Social Networks and Infectious Disease: The Colorado Springs Study". en. In: *Social Science & Medicine* 38.1 (Jan. 1994), pp. 79–88. ISSN: 02779536. DOI: 10.1016/0277-9536(94)90302-6.

[Klo85] A. S. Klovdahl. "Social Networks and the Spread of Infectious Diseases: The AIDS Example". en. In: *Social Science & Medicine* 21.11 (Jan. 1985), pp. 1203–1216. ISSN: 02779536. DOI: 10.1016/0277-9536(85)90269-2.

[KM27] W. O. Kermack & A. McKendrick. "A Contribution to the Mathematical Theory of Epidemics". en. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (Aug. 1927), pp. 700–721. ISSN: 0950-1207, 2053-9150. DOI: 10.1098/rspa.1927.0118.

[Kri20] M. Kriegel. "Anzahl der mit SARS-CoV-2 beladenen Partikel in der Raumluft und deren eingeatmete Menge, sowie die Bewertung des Infektionsrisikos, sich darüber mit Covid-19 anzustecken". de. In: (Oct. 2020). DOI: 10.14279/DEPOSITONCE-10655.3.

[Küh+20] M. J. Kühn; D. Abele; T. Mitra; W. Koslow; M. Abedi; K. Rack; M. Siggel; S. Khailaie; M. Klitz; S. Binder; L. Spataro; J. Gilg; J. Kleinert; M. Häberle; L. Plötzke; C. D. Spinner; M. Stecher; X. X. Zhu; A. Basermann & M. Meyer-Hermann. *Assessment of Effective Mitigation and Prediction of the Spread of SARS-CoV-2 in Germany Using Demographic Information and Spatial Resolution*. en. Preprint. Epidemiology, Dec. 2020. DOI: 10.1101/2020.12.18.20248509.

[Lee+10a] B. Y. Lee; S. T. Brown; P. C. Cooley; R. K. Zimmerman; W. D. Wheaton; S. M. Zimmer; J. J. Grefenstette; T.-M. Assi; T. J. Furphy; D. K. Wagener & D. S. Burke. "A Computer Simulation of Employee Vaccination to Mitigate an Influenza Epidemic". en. In: *American Journal of Preventive Medicine* 38.3 (Mar. 2010), pp. 247–257. ISSN: 07493797. DOI: 10.1016/j.amepre.2009.11.009.

[Lee+10b] B. Y. Lee; S. T. Brown; P. Cooley; M. A. Potter; W. D. Wheaton; R. E. Voorhees; S. Stebbins; J. J. Grefenstette; S. M. Zimmer; R. K. Zimmerman; T.-M. Assi; R. R. Bailey; D. K. Wagener & D. S. Burke. "Simulating School Closure Strategies to Mitigate an Influenza Epidemic". en. In: *Journal of Public Health Management and Practice* 16.3 (May 2010), pp. 252–261. ISSN: 1078-4659. DOI: 10.1097/PHH.0b013e3181ce594e.

[Lee+10c] B. Y. Lee; S. T. Brown; G. W. Korch; P. C. Cooley; R. K. Zimmerman; W. D. Wheaton; S. M. Zimmer; J. J. Grefenstette; R. R. Bailey; T.-M. Assi & D. S. Burke. "A Computer Simulation of Vaccine Prioritization, Allocation, and Rationing during the 2009 H1N1 Influenza Pandemic". en. In: *Vaccine* 28.31 (July 2010), pp. 4875–4879. ISSN: 0264410X. DOI: 10.1016/j.vaccine.2010.05.002.

[LEK20] Y. Liu; R. M. Eggo & A. J. Kucharski. "Secondary Attack Rate and Superspreading Events for SARS-CoV-2". en. In: *The Lancet* 395.10227 (Mar. 2020), e47. ISSN: 01406736. DOI: 10.1016/S0140-6736(20)30462-1.

[Lel+20] J. Lelieveld; F. Helleis; S. Borrmann; Y. Cheng; F. Drewnick; G. Haug; T. Klimach; J. Sciare; H. Su & U. Pöschl. "Model Calculations of Aerosol Transmission and Infection Risk of COVID-19 in Indoor Environments". en. In: *International Journal of Environmental Research and Public Health* 17.21 (Nov. 2020), p. 8114. ISSN: 1660-4601. DOI: 10.3390/ijerph17218114.

[Les+09] J. Leskovec; K. J. Lang; A. Dasgupta & M. W. Mahoney. "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters". In: *Internet Mathematics* 6.1 (Jan. 2009), pp. 29–123. ISSN: 1542-7951, 1944-9488. DOI: 10.1080/15427951.2009.10129177.

[LH07] D. A. Luke & J. K. Harris. "Network Analysis in Public Health: History, Methods, and Applications". In: *Annual Review of Public Health* 28.1 (Apr. 2007), pp. 69–93. ISSN: 0163-7525, 1545-2093. DOI: 10.1146/annurev.publhealth.28.021406.144132.

[Li+20] R. Li; S. Pei; B. Chen; Y. Song; T. Zhang; W. Yang & J. Shaman. "Substantial Undocumented Infection Facilitates the Rapid Dissemination of Novel Coronavirus (SARS-CoV-2)". en. In: *Science* 368.6490 (May 2020), pp. 489–493. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abb3221.

[Llo+05] J. O. Lloyd-Smith; S. J. Schreiber; P. E. Kopp & W. M. Getz. "Superspreading and the Effect of Individual Variation on Disease Emergence". en. In: *Nature* 438.7066 (Nov. 2005), pp. 355–359. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature04153.

[LS12] D. A. Luke & K. A. Stamatakis. "Systems Science Methods in Public Health: Dynamics, Networks, and Agents". en. In: *Annual Review of Public Health* 33.1 (Apr. 2012), pp. 357–376. ISSN: 0163-7525, 1545-2093. DOI: 10.1146/annurev-publhealth-031210-101222.

[Luo+20] X. Luo; S. Feng; J. Yang; X.-L. Peng; X. Cao; J. Zhang; M. Yao; H. Zhu; M. Y. Li; H. Wang & Z. Jin. *Analysis of Potential Risk of COVID-19 Infections in China Based on a Pairwise Epidemic Model*. Preprint. MATHEMATICS & COMPUTER SCIENCE, Feb. 2020. DOI: 10.20944/preprints202002.0398.v1.

[MA87] R. M. May & R. M. Anderson. "Transmission Dynamics of HIV Infection". en. In: *Nature* 326.6109 (Mar. 1987), pp. 137–142. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/326137a0.

[Mac+13] A. Machens; F. Gesualdo; C. Rizzo; A. E. Tozzi; A. Barrat & C. Cattuto. "An Infectious Disease Model on Empirical Networks of Human Contact: Bridging the Gap between Dynamic Network Data and Contact Matrices". en. In: *BMC Infectious Diseases* 13.1 (Dec. 2013), p. 185. ISSN: 1471-2334. DOI: 10.1186/1471-2334-13-185.

[Man20] G. Manzo. “Complex Social Networks Are Missing in the Dominant COVID-19 Epidemic Models”. en. In: *Sociologica* Vol 14 (May 2020), 31–49 Pages. DOI: 10.6092/ISSN.1971-8853/10839.

[Mor+96] M. Morris; C. Podhisita; M. Wawer & M. Handcock. “Bridge Populations in the Spread of HIV/AIDS in Thailand”. In: *AIDS (London, England)* 10.11 (Sept. 1996), pp. 1265–1271. ISSN: 0269-9370. DOI: 10.1097/00002030-199609000-00013.

[Mos+08] J. Mossong; N. Hens; M. Jit; P. Beutels; K. Auranen; R. Mikolajczyk; M. Massari; S. Salmaso; G. S. Tomba; J. Wallinga; J. Heijne; M. Sadkowska-Todys; M. Rosinska & W. J. Edmunds. “Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases”. en. In: *PLoS Medicine* 5.3 (Mar. 2008). Ed. by S. Riley, e74. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.0050074.

[MR20] G. Manzo & A. V. D. Rijt. “Halting SARS-CoV-2 by Targeting High-Contact Individuals”. In: *J. Artif. Soc. Soc. Simul.* 23 (2020).

[MS06] F. Martino & A. Spoto. “Social Network Analysis: A Brief Theoretical Review and Further Perspectives in the Study of Information Technology”. In: *PsychNology Journal* 4 (Jan. 2006), pp. 53–86.

[NG04] M. E. J. Newman & M. Girvan. “Finding and Evaluating Community Structure in Networks”. en. In: *Physical Review E* 69.2 (Feb. 2004), p. 026113. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.69.026113.

[Onn+07] J.-P. Onnela; J. Saramäki; J. Hyvönen; G. Szabó; D. Lazer; K. Kaski; J. Kertész & A.-L. Barabási. “Structure and Tie Strengths in Mobile Communication Networks”. en. In: *Proceedings of the National Academy of Sciences* 104.18 (May 2007), pp. 7332–7336. ISSN: 0027-8424. DOI: 10.1073/pnas.0610245104.

[OP09] T. Opsahl & P. Panzarasa. “Clustering in Weighted Networks”. In: *Social Networks* 31.2 (May 2009), pp. 155–163. ISSN: 03788733. DOI: 10.1016/j.socnet.2009.02.002.

[Org07] P. W. Organization, ed. *Global Public Health Security in the 21st Century: Global Public Health Security*. en. The World Health Report 2007. Genf: World Health Organization, 2007. ISBN: 978-92-4-156344-4.

[Osg07] N. Osgood. “Using Traditional and Agent Based Toolset for System Dynamics: Present Tradeoffs and Future Evolution”. In: (Jan. 2007).

[RKI21] RKI. “Täglicher Lagebericht des RKI zur Coronavirus-Krankheit-2019”. de. In: (Apr. 2021), p. 19.

[RM21] J. Römer & J. Merlot. *Coronavirus: Diese acht Fachleute beraten Bundesregierung und Länderchefs*. de. https://www.spiegel.de/wissenschaft/medizin/diese-sieben-fachleute-beraten-bundesregierung-und-laenderchefs-a-93abc4f5-cac1-4cbb-bc22-8d3b9c623b28. Feb. 2021.

[RS08] H. Rahmandad & J. Sterman. “Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models”. en. In: *Management Science* 54.5 (May 2008), pp. 998–1014. ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.1070.0787.

[RSW20] J. Rocklöv; H. Sjödin & A. Wilder-Smith. “COVID-19 Outbreak on the Diamond Princess Cruise Ship: Estimating the Epidemic Potential and Effectiveness of Public Health Countermeasures”. en. In: *Journal of Travel Medicine* 27.3 (May 2020), taaa030. ISSN: 1195-1982, 1708-8305. DOI: 10.1093/jtm/taaa030.

[RWC20] A. Radulescu; C. Williams & K. Cavanagh. “Management Strategies in a SEIR Model of COVID 19 Community Spread”. In: *arXiv:2003.11150 [physics, q-bio]* (Nov. 2020). arXiv: 2003.11150 [physics, q-bio].

[Ryd+05] J. J. Ryder; K. M. Webberley; M. Boots & R. J. Knell. “Measuring the Transmission Dynamics of a Sexually Transmitted Disease”. en. In: *Proceedings of the National Academy of Sciences* 102.42 (Oct. 2005), pp. 15140–15143. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0505139102.

[Sat+20] F. Sattler; J. Ma; P. Wagner; D. Neumann; M. Wenzel; R. Schäfer; W. Samek; K.-R. Müller & T. Wiegand. “Risk Estimation of SARS-CoV-2 Transmission from Bluetooth Low Energy Measurements”. en. In: *npj Digital Medicine* 3.1 (Dec. 2020), p. 129. ISSN: 2398-6352. DOI: 10.1038/s41746-020-00340-0.

[She+04] Z. Shen; F. Ning; W. Zhou; X. He; C. Lin; D. P. Chin; Z. Zhu & A. Schuchat. “Superspreading SARS Events, Beijing, 2003”. In: *Emerging Infectious Diseases* 10.2 (Feb. 2004), pp. 256–260. ISSN: 1080-6040, 1080-6059. DOI: 10.3201/eid1002.030732.

[SJ10] M. Salathé & J. H. Jones. “Dynamics and Control of Diseases in Networks with Community Structure”. en. In: *PLoS Computational*

Biology 6.4 (Apr. 2010). Ed. by C. Fraser, e1000736. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000736.

[SM01] R. V. Solé & M. Montoya. "Complexity and Fragility in Ecological Networks". en. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268.1480 (Oct. 2001), pp. 2039–2045. ISSN: 0962-8452, 1471-2954. DOI: 10.1098/rspb.2001.1767.

[SM15] J. Saramäki & E. Moro. "From Seconds to Months: An Overview of Multi-Scale Dynamics of Mobile Telephone Calls". en. In: *The European Physical Journal B* 88.6 (June 2015), p. 164. ISSN: 1434-6028, 1434-6036. DOI: 10.1140/epjb/e2015-60106-6.

[SPN20] A. Simha; R. V. Prasad & S. Narayana. "A Simple Stochastic SIR Model for COVID 19 Infection Dynamics for Karnataka: Learning from Europe". In: *arXiv:2003.11920 [math, q-bio]* (Apr. 2020). arXiv: 2003.11920 [math, q-bio].

[SSL16] V. Sekara; A. Stopczynski & S. Lehmann. "Fundamental Structures of Dynamic Social Networks". In: *Proceedings of the National Academy of Sciences* 113.36 (2016), pp. 9977–9982. ISSN: 0027-8424. DOI: 10.1073/pnas.1602803113. eprint: https://www.pnas.org/content/113/36/9977.full.pdf.

[Ste19] C. Stegbauer. "Granovetter (1973): The Strength of Weak Ties". In: *Schlüsselwerke Der Netzwerkforschung*. Ed. by B. Holzer & C. Stegbauer. Netzwerkforschung. Wiesbaden: Springer Fachmedien Wiesbaden, 2019, pp. 229–231. ISBN: 978-3-658-21741-9 978-3-658-21742-6. DOI: 10.1007/978-3-658-21742-6_52.

[Sun+20] J. Sun; W.-T. He; L. Wang; A. Lai; X. Ji; X. Zhai; G. Li; M. A. Suchard; J. Tian; J. Zhou; M. Veit & S. Su. "COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives". en. In: *Trends in Molecular Medicine* 26.5 (May 2020), pp. 483–495. ISSN: 14714914. DOI: 10.1016/j.molmed.2020.02.008.

[SV11] E. Stattner & N. Vidot. "Social Network Analysis in Epidemiology: Current Trends and Perspectives". In: *2011 FIFTH INTERNATIONAL CONFERENCE ON RESEARCH CHALLENGES IN INFORMATION SCIENCE*. Gosier, France: IEEE, May 2011, pp. 1–11. ISBN: 978-1-4244-8670-0. DOI: 10.1109/RCIS.2011.6006866.

[Tab+18] S. Tabassum; F. S. F. Pereira; S. Fernandes & J. Gama. "Social Network Analysis: An Overview". en. In: *WIREs Data Mining and Knowledge Discovery* 8.5 (Sept. 2018). ISSN: 1942-4787, 1942-4795. DOI: 10.1002/widm.1256.

[Tan+20] B. Tang; N. L. Bragazzi; Q. Li; S. Tang; Y. Xiao & J. Wu. "An Updated Estimation of the Risk of Transmission of the Novel Coronavirus (2019-nCov)". en. In: *Infectious Disease Modelling* 5 (2020), pp. 248–255. ISSN: 24680427. DOI: 10.1016/j.idm.2020.02.001.

[Ven+18] S. Venkatramanan; B. Lewis; J. Chen; D. Higdon; A. Vullikanti & M. Marathe. "Using Data-Driven Agent-Based Models for Forecasting Emerging Infectious Diseases". en. In: *Epidemics* 22 (Mar. 2018), pp. 43–49. ISSN: 17554365. DOI: 10.1016/j.epidem.2017.02.010.

[Vol+21] E. Volz; S. Mishra; M. Chand; J. C. Barrett; R. Johnson; L. Geidelberg; W. R. Hinsley; D. J. Laydon; G. Dabrera; Á. O'Toole; R. Amato; M. Ragonnet-Cronin; I. Harrison; B. Jackson; C. V. Ariani; O. Boyd; N. J. Loman; J. T. McCrone; S. Gonçalves; D. Jorgensen; R. Myers; V. Hill; D. K. Jackson; K. Gaythorpe; N. Groves; J. Sillitoe; D. P. Kwiatkowski; The COVID-19 Genomics UK (COG-UK) consortium; S. Flaxman; O. Ratmann; S. Bhatt; S. Hopkins; A. Gandy; A. Rambaut & N. M. Ferguson. *Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from Linking Epidemiological and Genetic Data*. en. Preprint. Infectious Diseases (except HIV/AIDS), Jan. 2021. DOI: 10.1101/2020.12.30.20249034.

[WC20] F. Wong & J. J. Collins. "Evidence That Coronavirus Superspreading Is Fat-Tailed". en. In: *Proceedings of the National Academy of Sciences* 117.47 (Nov. 2020), pp. 29416–29418. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2018490117.

[WF94] S. Wasserman & K. Faust. *Social Network Analysis: Methods and Applications*. First. Cambridge University Press, Nov. 1994. ISBN: 978-0-521-38707-1 978-0-521-38269-4 978-0-511-81547-8. DOI: 10.1017/CBO9780511815478.

[Wor21] H. World. "Coronavirus Disease (COVID-19) Advice for the Public". In: (Apr. 2021).

[YAE08] Y. Yang; P. Atkinson & D. Ettema. "Individual Space–Time Activity-Based Modelling of Infectious Disease Transmission within a City".

en. In: *Journal of The Royal Society Interface* 5.24 (July 2008), pp. 759–772. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2007.1218.

[ZC20] S. Zhao & H. Chen. "Modeling the Epidemic Dynamics and Control of COVID-19 Outbreak in China". en. In: *Quantitative Biology* 8.1 (Mar. 2020), pp. 11–19. ISSN: 2095-4689, 2095-4697. DOI: 10.1007/s40484-020-0199-0.

[Zha+20] R. Zhang; Y. Li; A. L. Zhang; Y. Wang & M. J. Molina. "Identifying Airborne Transmission as the Dominant Route for the Spread of COVID-19". en. In: *Proceedings of the National Academy of Sciences* 117.26 (June 2020), pp. 14857–14863. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2009637117.

Appendix

A Figures

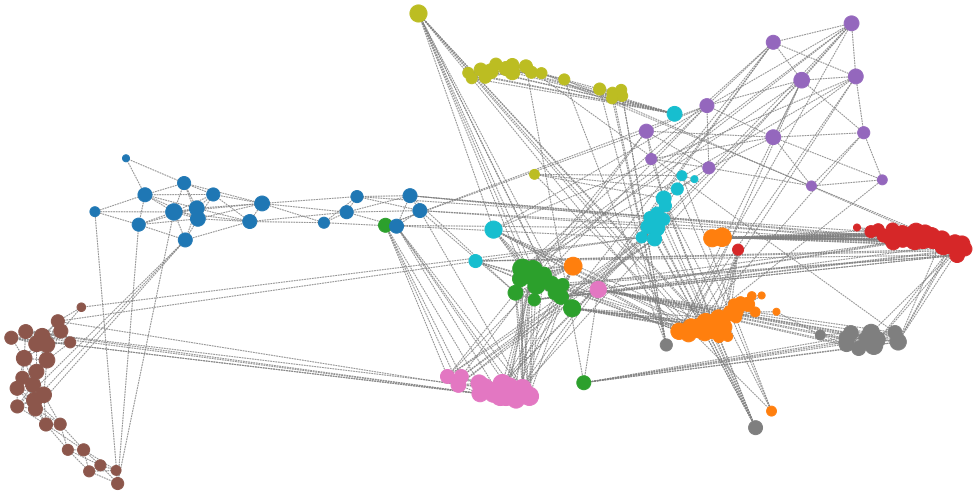

FIG. A.1: Randomly generated undirected geometric graph $G = (V, E)$ with number of nodes (verticies) $|V| = 200$ and number of edges $|E| = 837$. In this example, 20 nodes were assigned to a random community, which explains the unclean separation of these communities.

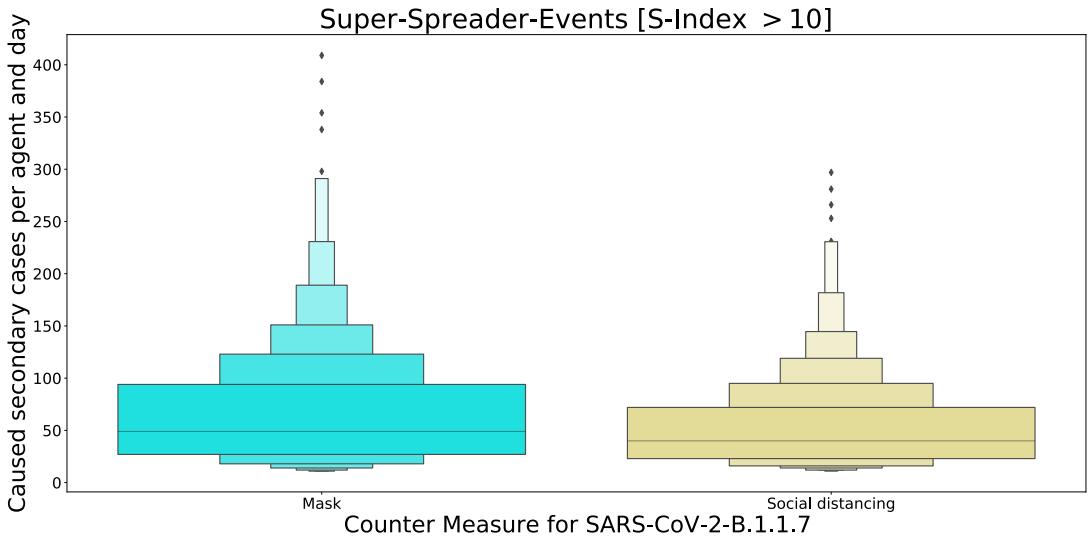

FIG. A.2: Distribution of SSE for the two countermeasures taken. *S-index* > 10 applies.

B Code

The model developed in this thesis was implemented with the software stack presented in chapter 4.1. All code used (*JupyterNotebooks*) is located on the SD memory card included with this thesis.

CODE B.1: Implement countermeasures and estimate system-specific reproduction numbers.

```

    # social distancing
    if social_distancing != 1:
        connections = connections.sample(frac=social_distancing)

    # Count amount of persons in social network
    agents_a = c_rate['primary'].unique()
    agents_b = c_rate['secondary'].unique()
    n_agents = len(list(set().union(agents_a, agents_b)))

    # Calculate infection_risk given a reproductive number
    def infec(contacts, infection_risk):
        return 1-(1-infection_risk)**contacts

    def reproductive_number(infection_risk):
        c_rate['infection_probability'] = c_rate['amount'].apply(lambda x:
↪infec(x, infection_risk))
        return c_rate['infection_probability'].sum()/(n_days*n_agents)

    # calculate infection_risk given a reproductive number
    if sampling:
        def match_reproductive_number(infection_risk):
            return reproductive_number(infection_risk) - Ro
        infection_risk = fsolve(match_reproductive_number, 0)[0]

    infection_risk_eff = infection_risk*mask_usage
```

C Data

The data used in this thesis were recorded as part of the ResTAat project of the DLR. For data protection reasons, the data may not be published. Requests to view or use the data should be addressed to Eva Brucherseifer (eva.brucherseifer@dlr.de) or Daniel Lichte (daniel.lichte@dlr.de).