Modelling epidemics on networks with static and time-varying topologies

vorgelegt von
M.Sc.
Andreas Koher

von der Fakultät II - Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
Dr. rer. nat.

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Dr. Michael Lehmann
Gutachter: Dr. habil. Philipp Hövel
Gutachterin: Prof. Dr. Kathy Lüdge
Tag der wissenschaftlichen Aussprache: 03.07.2019

Berlin 2019

Zusammenfassung

Ziel dieser Arbeit ist es, epidemiologische Modelle zur Beschreibung infektiöser Krankheiten zu entwickeln. Dabei integrieren wir Mobilitätsdaten, die mathematisch durch Netzwerke mit statischen oder zeitlich veränderlichen Topologien beschrieben werden können. Als zentrales Ergebniss dieser Arbeit präsentieren wir drei Krankheitsmodelle, die eine effiziente Analyse komplexer Probleme gestatten.

Zuerst stellen wir im Kontext globaler Epidemien ein Netzwerk-basiertes Maß vor und zwar die *random-walk effective distance* (RWED), mit dem die erwartete Ankunftszeit einer Krankheit abgeschätzt werden kann. Dazu gehen wir vom globalen Mobilitätsnetzwerk aus, in dem Knoten und Kanten urbane Regionen abstrahieren, die durch den globalen Flugverkehr verbunden sind. Die RWED berücksichtigt, entgegen früherer Methoden, alle potentiellen Übertragungswege vom Ausbruchsort bis zum Zielknoten und ermöglicht somit präzisere Vorhersagen zum Verlauf einer Pandemie.

Als nächstes entwickeln wir einen Matrix-basierten Algorithmus, der zeitlich veränderliche Topologien mit generischen Krankheitsmodellen integriert. Durch boolesche Verknüpfungen leiten wir eine Erreichbarkeitsmatrix ab, mit der die Vulnerabilität und das Schadenspotential einzelner Knoten effizient untersucht werden können.

Schließlich stellen wir mit dem Kontakt-basierten Modell einen versatilen Ansatz vor, in dem temporär auftretende Kanten statt Knoten zentrale Elemente eines dynamischen Systems sind. Dieser Perspektivwechsel ermöglicht präzisere Vorhersagen bei Infektionskrankheiten mit dauerhafter Immunität. Insbesonders leiten wir ein analytisches Kriterium ab, mit dem das Risiko eines globalen Krankheitsausbruchs effizient abgeschätzt werden kann.

Die vorgestellten Modelle integrieren Elemente der statistischen Physik, nichtlinearen Dynamik und der linearen Algebra, um spezifische Fragestellungen der Epidemiologie zu beleuchten. Analytische Vorhersagen vergleichen wir mit numerischen Modellen auf zahlreichen empirischen Mobilitätsnetzwerken, um die Qualität und Grenzen unserer Methoden zu evaluieren.

Abstract

We aim to develop epidemiological models that describe and predict the evolution of infectious diseases. To this end, we integrate mobility data, represented by networks with static and time-varying topologies. As a central result of this work, we present three disease models that facilitate an efficient analysis of complex epidemiological problems.

First, in the context of global epidemics, we present a network-based measure, termed random-walk effective distance (RWED), as a proxy for the expected infection arrival time. Here, we use the global mobility network where nodes and edges represent urban regions and air traffic connections, respectively. The RWED integrates all potential transmission paths from the outbreak location to the target node and thus improves predictions about the course of a pandemic.

Next, we are developing a matrix-based algorithm that integrates time-varying topologies with generic disease models. Based on Boolean algebra, we derive a reachability matrix in order to assess the vulnerability and damage potential of individual nodes in a temporal network.

Finally, we present the so-called contact-based model as a versatile approach in which temporal edges instead of nodes form central elements of a dynamic system. This change of perspective improves previous, individual-based approaches for infectious diseases with permanent immunity. In particular, we derive an analytical criterium to determine the critical disease parameters that separate local and global outbreaks.

The presented models integrate elements of statistical physics, nonlinear dynamics, and linear algebra to provide quantitative insights into epidemiological problems. We compare these analytical predictions with numerical models on numerous empirical mobility networks in order to evaluate the quality and limitations of our analytical approaches.

Table of Contents

Title Page i

Zusammenfassung iii

Abstract v

1 Introduction 1

2 Theory 9
- 2.1 Compartemental disease models . . . 9
  - 2.1.1 Dynamic mean-field model . . . 11
- 2.2 Complex network analysis . . . 15
  - 2.2.1 Network analysis of pig-trade in Germany . . . 19
- 2.3 Dynamics on networks . . . 26
  - 2.3.1 Random-walk on static graphs . . . 26
  - 2.3.2 Meta-population model . . . 30
  - 2.3.3 Individual-based model . . . 34
- 2.4 Summary . . . 39

3 Effective Distances for Global Epidemic Outbreaks 41
- 3.1 Introduction . . . 41
- 3.2 The global air-traffic network . . . 44
- 3.3 The meta-population model . . . 45
- 3.4 A framework for effective distances . . . 47
  - 3.4.1 Two populations. . . . 49
  - 3.4.2 A chain of populations. . . . 52
  - 3.4.3 Tree graph. . . . 54
  - 3.4.4 Parallel paths. . . . 57
  - 3.4.5 Multiple path effective distance. . . . 59
  - 3.4.6 Dominant path effective distance. . . . 62
- 3.5 The random walk effective distance . . . 65
- 3.6 Computation of the RWED . . . 67
- 3.7 Numerical comparison . . . 70
- 3.8 Summary . . . 74

4 Matrix based approach to epidemic modeling 77
4.1 Introduction . . . 77
4.2 Materials and Methods . . . 79
4.3 Derivation of the epidemiological model . . . 80
4.3.1 State Vector Formalism . . . 80
4.3.2 Tracking All Initial Conditions: Matrix Formalism . . . 82
4.4 Application to Empirical Contact Networks . . . 83
4.4.1 Social Contacts Network . . . 83
4.4.2 Sexual Contacts network . . . 85
4.4.3 Livestock-Trade Network . . . 85
4.4.4 Critical Infectious Period . . . 86
4.5 Conclusion . . . 88

5 Contact-based modelling 91
5.1 Introduction . . . 91
5.2 Conceptual framework . . . 93
5.3 Dynamic equations . . . 94
5.3.1 Individual-based model . . . 95
5.3.2 Contact-based model . . . 95
5.4 Epidemic threshold . . . 97
5.5 Application . . . 100
5.5.1 Numerical analysis of the mean-field dynamics . . . 101
5.5.2 Epidemic threshold . . . 108
5.5.3 Application to German cattle trade. . . . 108
5.6 Conclusion . . . 110

6 Conclusion and outlook 113

Appendix A Alternative derivation of the effective distance 117

Appendix B Additional material to Chap. 5 121
B.1 Weighted networks & heterogeneous parameters . . . 121
B.2 German cattle trade network . . . 123
B.3 Additional numerical results . . . 125

Appendix C Continuous-time model 133
C.1 Derivation of the continuous time limit . . . 133
C.2 Spectral properties of the continuous time model . . . 135
C.3 Average over random graphs and the EBC model . . . 138
C.4 Relation to message-passing . . . 141
C.5 General epidemic models on temporal networks . . . 143

References **149**

List of Figures **165**

List of Tables **167**

Abbreviations **169**

1

Introduction

The past two centuries witnessed enormous advances in the struggle against communicable diseases, which had previously been the leading cause of death [1]. Especially after the successful eradication of smallpox in 1977, optimism prevailed that other highly contagious diseases such as diphtheria, tetanus and measles would soon disappear and that medicine would ultimately succeed in the fight against infectious diseases [1]. As a consequence, both professional and public opinion were caught by surprise when new diseases such as HIV appeared, new forms of influenza turned into global epidemics, and antibiotic-resistant strains raised major concerns. Expressed in numbers, approximately 15 million people (over 25 %) die each year from communicable diseases [2]. This mainly affects developing countries and especially children: about three million children die annually from malaria and diarrhoea [2]. Despite medical progress, infectious diseases are also the second most important cause of death in industrial nations [2], where some of the key factors contributing to this observation are increasing global mobility and urbanization [3–5]. This concerning development, stresses the need for new epidemiological models that assess risks, predict epidemic patterns and help to evaluate mitigation strategies.

Models of epidemics. The foundation of modern theoretical epidemiology was established at the beginning of the 20th century, mainly by health physicians such as Ross, Hamer, McKendrick, and Kermack who introduced the *compartment model* [6–8]. This approach separates individuals within a population into epidemic categories or compartments, depending on their health status such as susceptible, infected and, recovered. Since the early years, development in the field of mathematical epidemiology has accelerated, not least due to the seminal works of Bailey [9], Anderson & May [10] and Hethocote [11]. Modern models include stochasticity [12–15], non-Markovian dynamics [16–20], demographic structures, vaccinations, disease vectors and, quarantine (see references in [11] for details). Thus, the field of research ranges from simple explanatory models that reveal

hidden patterns and reproduce fundamental observations to elaborate numerical models that provide realistic predictions [21].

In recent years we witnessed a second "golden age" [22] of epidemiological modelling. The driving forces behind this development are increasing computing power and an unprecedented amount of mobility data. The combination of both allows scientists to simulate the behaviour of entire populations at the level of single individuals [23–29] and thus to advise policy makers by means of quantitative models.

Moreover, many methods targeting human infectious diseases are increasingly used in veterinary medicine. Here too, the growing livestock trade poses new risks for epidemic outbreaks. A prominent example is the tremendous loss due to foot-and-mouth disease in the UK in 2001 [30]. In order to prevent future outbreaks, it was decided in 2001 by EU legislation to record animal trade. This new source of time-resolved mobility data allows researchers to develop new surveillance and vaccination strategies, to trace the origin of an outbreak and assess the risk of potential outbreaks [31–33].

Network science. All data-driven models rely on information about human mobility patterns or, food and livestock trade. A mathematically meaningful representation of this relational data is a network in which nodes and edges represent individuals (or isolated populations) and their contact structure, respectively [34–38]. In the absence of empirical data, early graph theory focused mostly on random networks until the end of the 20th century. Notably, Erdős and Rényi laid the mathematical foundations in the 1950s and 1960s by studying the structure and percolation properties of dense random graphs [39–41]. Independently, well-known sociologists such as Milgram, Granovetter, Zachary, Freeman, Wasserman, and Faust developed central concepts based on empirical surveys [42–46]. Many network measures that were introduced to rank the importance of single individuals in a society remain cornerstones of modern network theory [36].

The new era of Big Data allows researchers today to overcome the limitations of random graphs and small sample sizes. Leveraging the extensive use of smartphones, the social web and, the internet of things, network scientists explore today the behavior of millions of individuals [47–49]. New sensor technologies record interactions between humans, allowing possible infection pathways to be studied at the micro-level [50–52].

This "data deluge" [53] also changes our understanding of social and economic networks. Recently, numerous studies demonstrated that empirical networks exhibit highly heterogeneous structures - often as a result of self-organizing dynamic processes [36, 54–62]. These patterns differ significantly from regular grids and simple random graphs studied earlier. Indeed, empirical networks are often hierarchically organized, with a few *hubs* accounting for the majority of contacts [54]. Typical features include also community structures within which nodes are closely linked [63–67], as well as a significant degree of clustering, i.e., a large

number of closed triangles that characterizes social interactions [42, 68]. Finally, an essential signature of heterogeneous network structures are heavy-tailed and skewed distributions that extend over several orders of magnitude [69].

An important observation of the past years is that contacts change over time, often on a similar time scale as the dynamics taking place on them. However, methods from classic network analysis do not take into account the temporal resolution of contact data resulting in a factitious transitivity as demonstrated in Fig. 1.1 [70–72]. In order to consider the temporal dimension, Lentz et al. developed a concept known as unfolding accessibility [71], whereby a similar approach was already presented in [73, 74]. The basic idea is that information is only transmitted if edges appear in a time-respecting order. The authors' merit is above all to have formalized this intuitive idea in an efficient matrix algorithm, which we will integrate with epidemiological models in Chap. 4 of this thesis.

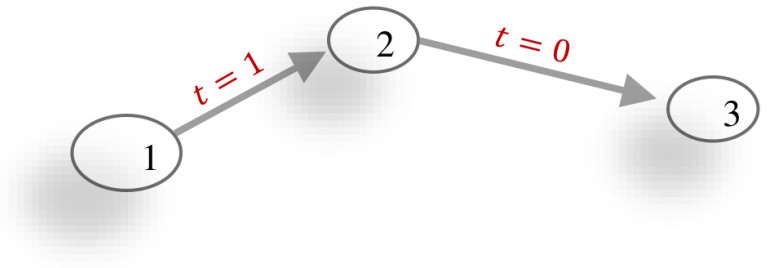

Figure 1.1: Transitivity is not assured in temporal networks [72]. Here, links $1 \to 2$ and $2 \to 3$ exist, but the temporal order ($1 \to 2$ at time $t = 1$ and $2 \to 3$ at time $t = 0$) prevents information to spread from node 1 to 3.

The topological and temporal heterogeneity of empirical networks has a direct impact on the spread of infectious diseases. Scale-free networks, for instance, are on the one hand particularly suitable for targeted vaccination strategies that focus on highly connected nodes, i.e., hubs [75, 76]. On the other hand, this prevalent type of a graph is highly susceptible to global outbreaks, once a hub is infected [77]. Moreover, the simple example in Fig. 1.1 demonstrates that the temporal sequence of edges determines whether or not an infection is possible in principle. Numerous consequences of this complex interplay between network dynamics and disease spreading have been documented in recent years [78–84].

As a consequence, the abundance of (time-resolved) mobility data leads to promising opportunities as well as new challenges for epidemiologists. New mathematical approaches are required to combine classical models of disease dynamics with heterogeneous contact structures that evolve in time. Thereby, the difficulty is to reduce the complexity of the problem because, even for simple epidemiological models a complete stochastic description of the system using a master equation is rarely feasible [85]. In recent years, this challenge led to a large number of new analytic approaches, some of which we present here as examples that inspired the present work.

Data-driven modelling. One of the cornerstones of network-based disease models is the so-called individual-based approach. It is a drastic simplification of the exact description using a master equation, because it assumes that the epidemiological states of neighboring nodes are statistically independent. As a consequence, one can define a set of dynamic equations for the marginal probability to find a node in a given disease state [15, 33, 86–91]. This method is widely employed, because it offers an intuitive and analytically tractable approach to integrate the underlying contact network. As a particularly important result, we mention that the largest eigenvalue of the adjacency matrix, which represents the topology of the network, determines the critical disease parameters that separate local and global outbreaks [33, 88]. This result is particularly relevant for public health institutions as it facilitates a rapid risk assessment and we will present the mathematical details in Chap. 2.

From the numerous extensions of the individual-based approach, *pair approximations* are particularly worth mentioning [92]. The main idea is to include two-point marginals, i.e., the probability for the joint state of adjacent nodes, into the set of dynamic equations. Then, a mean-field approximation becomes necessary on the level of three-point marginals, for which many different approaches exist [93–97].

Another noteworthy approach is Gleeson's general model of binary state dynamics [98], which is based on the work of Marceau et al. [99]. Using an approximate master equation, the authors determine the state of a selected node with high accuracy for a variety of different contagion models. Here too, dynamic correlations between adjacent nodes are explicitly included (see also Lindquist et al. [100] for a similar model).

Finally, Karrer & Newman substantially improved previous models of unidirectional diseases, such as the generic susceptible-infected-recovered model, using the message-passing framework [101]. This approach dates back to the computer scientist Pearl [102], who formulated an exact inference algorithm on trees. Based on earlier works, Karrer & Newman derived an integro-differential equation as a model for non-Markovian disease dynamics and improved previous estimates of the critical disease parameters [103]. A crucial conceptual difference to earlier works is that edges instead of nodes appear as central elements of the model. This approach influenced considerably further research [17, 104–106] and serves as an essential starting point to our *contact-based* framework in Chap. 5.

So far, we have presented models that consider a node as an individual characterized by a discrete disease state. In contrast, another class of *reaction-diffusion models* exists that identifies nodes with populations. Within each population, individuals interact randomly with each other (hence the term *homogeneous mixing*) and thus an epidemic develops locally according to the classical compartmental model [4, 107, 108]. Additionally, individuals move along edges and thus contribute to the disease propagation across the network [109–112].

We will leverage this appoach in Chap. 3 to model global epidemics, where nodes represent urban regions connected by air traffic routes [3–5, 107, 108, 113–115].

Finally, we mention a third class of models that surprise in the context of epidemiological models, namely random walks. This simplest of all dynamic processes has a diverse history [116–118]. The first works date back to the "Gambler's ruin" problem analyzed by Fermat, Pascal, Bernoulli, and others [119]. Since then, this model has been used to describe, among others, gases [120], stock prices [121, 122] or, movement of bacteria [123, 124]. Further applications include ranking methods, like the famous PageRank algorithm [125, 126] and the analysis of network structures [127–130]. In Chap. 3, we will show how this generic model of a diffusive process can help to answer complex questions in epidemiology.

This brief introduction shows that recent years have seen a rapid development of new epidemiological models that integrate methods from various disciplines such as Bayesian inference, statistical physics, nonlinear dynamics, and linear algebra. Driven by the availability of mobility data, these integrative models support policy-makers and public health institutions in an increasingly interconnected world.

Open challenges. Despite many remarkable advances in recent years, significant challenges remain unsolved or inadequately addressed. In order to estimate the expected outbreak size or infection arrival time, for instance, one must either resort to time-consuming probabilistic Monte-Carlo simulations or integrate high-dimensional models. Instead, first results indicate that simple network-based quantities can serve as robust and efficient proxies, many of which still need to be explored [5]. Time-varying topologies, in particular, are lacking basic methods that have been applied successfully with static graphs. These include efficient, matrix-based approaches to evaluate the risk and vulnerability of single nodes. Also, a conceptionally simple extension to the over-simplistic individual-based framework that integrates time-varying topologies is still missing.

This thesis attempts to address some of the challenges mentioned above. To this end, the following questions will guide our work:

- When does a disease arrive in a distant population (Chap. 3)?
- How can we determine vulnerable and high-risk nodes, i.e. nodes that are most influential with the respect to spreading processes (Chap. 4, Chap. 5)?
- What are the critical epidemic parameters that lead to global outbreaks (Chap. 4, Chap. 5)?

To answer these question the thesis combines models of epidemic spreading with complex network theory. In the present work we focus on empirical networks, i.e. mobility data that determines the coupling between dynamical elements such as individuals or populations, respectively.

The thesis is structured as follows: In Chap. 2, we introduce the main mathematical concepts. In particular, we present relevant epidemiological models, as well as fundamentals from network theory that serve as a basis for our contribution in the main part. In Chap. 3, we derive a network-based distance measure as a proxy for the expected infection arrival time in a global epidemic outbreak. Next, we focus on time-varying topologies and present in Chap. 4 a compartmental model that integrates network and disease dynamics in an efficient matrix-based framework. In Chap. 5, we propose a contact-based model as opposed to an individual-based approach and thus improve previous estimates. In particular, we derive a novel spectral criterion for the critical disease parameters, which marks the transition from local to global outbreaks. Finally, we conclude the work with a summary and outlook in Chap. 6. As a reference for frequently used variables we refer to Tab. 1.1.

Table 1.1: Notation of frequently used variables and parameters.

Chap. 2	
S, I, R	Susceptibele, infected, recovered compartment
$M_i^X(t)$	Number of individuals in population i and compartment $X \in \{\mathrm{S}, \mathrm{I}, \mathrm{R}\}$ at time t
M_i	Size of population i
$\rho_i^X(t)$	Fraction of individuals in population i and compartment $X \in \{\mathrm{S}, \mathrm{I}, \mathrm{R}\}$ at time t
β	Per capita infection rate / per time step infection probability
μ	Recovery rate / per time step recovery probability
α	Force of infection
λ	Early growth rate
R_0	Basic reproductive ratio
R_∞	Final outbreak size
G	Static graph
$\mathcal{G}$	Temporal network / time-ordered sequence of static graphs
$\mathcal{N}$	Set of nodes
N	Number of nodes
$\mathcal{E}$	Set of edges
E	Number of edges
F_{ij}	Edge weight from node i to j
k_i	Degree of node i (undirected network)
F_i	Strength of node i (undirected network)
k_i^{tot}, k_i^{out}, k_i^{in}	Total, out- in-degree of node i (directed networks)
F_i^{tot}, F_i^{out}, F_i^{in}	Total, out- in-strength of node i (directed networks)
$\langle k \rangle$	Mean degree
$\langle F \rangle$	Mean strength
$\boldsymbol{A}$	Adjacency matrix (binary)
$\bar{\boldsymbol{A}}$	Time-averaged adjacency matrix
$\mathcal{A}$	time-ordered sequence of adjacency matrices
$\boldsymbol{F}$	Adjacency matrix (weighted)
ω	walk
π	path
$\boldsymbol{R}$	Reachability matrix (static)
$\mathcal{R}$	Reachability matrix (temporal)
$\boldsymbol{T}$	(Random walk) transition matrix
T_{ij}	Transition probability from node i to j
$p_i(n)$	Probability that a random-walker is in node i after n jumps
$\boldsymbol{p}(n)$	Vector with elements $p_i(n)$ for all $i \in \mathcal{N}$
$\boldsymbol{p}^*$	Equilibrium distribution of a random-walk
$\phi(\tau)$	PDF for a CTRW to jump after a waiting time τ
γ_{ij}	Transition rate from node i to j
γ_i	Exit rate from node i
γ	Universal exit rate / jump rate of a random-walker
$\boldsymbol{L}$, $\boldsymbol{L'}$	(Combinatorial) Laplace matrix, random-walk Laplacian
$\Omega_i^X(t)$	Resulting flux of individuals in state X to and from node i
$\Theta(x)$	Heaviside function
ξ_i	Local invasion threshold of node i
$P_i^X(t)$	Probability $\Pr[X_i(t) = X]$ to find node i in state X at time t
$\lambda^{(1)}(.)$	Largest eigenvalue / spectral radius of a matrix
$\boldsymbol{M}$	Infection propagator matrix
$\mathbb{1}$	Identity matrix

Continued on next page

Additional variables of Chap. 3	
Φ	Total passenger numbers per day
c_v	Coefficient of variation
r	Pearson's correlation coefficient
t^*	Intection arrival time
$G(t)$	Gumbel distribution
$G^{(1)}$	First moment of the Gumbel distribution
$\langle t_{uv} \rangle$	Expected infection arrival time $u \to v$ (simulation)
χ_{uv}	Estimated infection arrival time $u \to v$ (analysis)
$\chi(\pi)$	Estimated transmission time along a path π
$\mathcal{P}_{uv}$	Set of paths from node u to v
$\mathcal{W}_{uv}$	Set of walks from node u to v
γ_e	Euler-Mascheroni number
δ	System parameter
D_{uv}^{MP}	Multiple-path effective distance from node *source* to v (MPED)
D_{uv}^{DP}	Dominant-path effective distance (DPED)
D_{uv}^{RW}	Random-walk effective distance (RWED)
$P(\pi)$, $P(\omega)$	Probability to realize a path π or, a walk ω
$P_{uv}(n)$	Probability to walk from node u to v in n jumps
$\tilde{P}_{uv}(c)$	Laplace transformed probability
$H_{uv}(n)$	First passage probability from node u to v
$\boldsymbol{Z}$	Fundamental matrix
$\mathrm{diag}(\boldsymbol{Z})$	Matrix with Z_{ii} elements on the diagonal and zero otherwise
Additional variables of Chap. 4	
τ	Infectious period in number of time steps
$\boldsymbol{i}(t)$	Boolean state vector indicating infected individuals at time t
$\boldsymbol{r}(t)$	Boolean state vector indicating recovered individuals at time t
$\boldsymbol{I}(t)$	Boolean state matrix indicating infected individuals for all outbreak locations
$\boldsymbol{R}(t)$	Boolean state matrix indicating recovered individuals
$\boldsymbol{C}(t)$	Boolean state matrix indicating individuals that have been infected up to time t
$\boldsymbol{J}(t)$	Boolean state matrix indicating newly infected individuals at time t
$\rho(\boldsymbol{X})$	Density of matrix $\boldsymbol{X}$
Additional variables of Chap. 5	
$\mathcal{C}$	Set of contacts
C	Total number of contacts
z_l	Probability that node l is initially susceptible, i.e., $z_l \equiv P_l^S(0)$
$\Phi_l(t)$	Probability that l does not contract the disease up to time t
$\theta_{k\to l}(t)$	Probability that cavity node l has not contracted the disease from k
$I_{k\to l}(t)$	Probability that k is infected and has not transmitted the disease to cavity node l
$S_{k\to l}(t)$	Probability that k is susceptible with l in cavity state
$a_{k\to l}(t)$	Indicates if a contact (k,l) is present at time t
$\boldsymbol{a}(t)$	Vector with elements $a_{k\to l}(t)$ for all edges (k,l)
$\mathcal{N}_k \setminus l$	Set of nodes adjacent to k without l
$\delta_{k\to l}(t)$	Probability that k is not infected with l in cavity state and $\delta \ll 1$
$\mathrm{diag}(\boldsymbol{x})$	Diagonal matrix with vector $\boldsymbol{x}$ defining diagonal elements
$\boldsymbol{B}$	Non-backtracking matrx

2

Theory

Abstract. In this chapter, we present the mathematical framework used to describe the dynamics of infectious diseases. To this end, we introduce the compartment model with homogeneous mixing at the beginning of Sec. 2.1 and derive basic results for the *susceptible-infected-recovered* (SIR) model such as the epidemic threshold. In Sec. 2.2 we then give an overview on relevant concepts of modern network analysis and illustrate selected methods on the example of the German pig-trade network. Finally, we integrate dynamical models and graph theory in Sec. 2.3. In particular, we discuss random-walks on networks, the meta-population model and the SIR contact-process.

2.1 Compartemental disease models

Infectious diseases display a broad spectrum of transmission pathways and clinical symptoms with distinct temporal variations. Despite the differences many phenomena appear qualitatively similar across a wide range of different diseases. These include the sudden transition from local to global outbreaks with decreasing time to recovery [131] as well as universal spreading patterns on a global scale [5]. In order to study generic features of epidemic dyanmics, we will explore in this thesis *phenomenological models* that abstract from details of a specific disease.

On the first level of abstraction, one can divide infectious diseases into those that (i) lead to permanent immunity, such as influenza, measles, chicken pox and diseases that (ii) allow multiple reinfections, e.g., tuberculosis, syphilis, and parasites. This thesis focuses on the first category, i.e., non-recurring infectious diseases. To this end, we follow the introduction in [22] and briefly review the most generic framework, namely the SIR model with homogeneous mixing.

The model separates a population according to the disease status into classes of susceptible, infected, and recovered, denoted S, I and R, respectively. The latter category is characterised by lifelong immunity. Moreover, we assume a

constant population size M and neglect demographic effects. Then, we analyze the transitions $S \to I$ and $I \to R$ as follows:

The recovery process $I \to R$ is also called spontaneous transition because no interaction with another individual is necessary. In reality, many diseases show a roughly constant duration of infection, e.g., about a week for flu-like infections. Due to stochastic fluctuations, a realistic distribution of the infectious period features a well-pronounced peak close to the average value with strongly non-exponential tails [132]. In Chap. 4 we will start from the assumption of a fixed infection duration and model the recovery process using a delay term. The application of delay differential equations to modeling non-exponential recovery dynamics dates back to the early Kermack-McKendrick work [8], which is rarely mentioned in this context today.

Another popular approach to model the recovery process, assumes a constant transition rate μ. In other words, an infected individual recovers within an infinitesimally short period $[t, t + \mathrm{d}t]$ with a constant probability $\mu \mathrm{d}t$. This approach leads to exponentially distributed infection durations with a mean at μ^{-1}. The assumption of a constant recovery rate is in many cases less realistic than a constant infectious period. However, it allows researchers to trade accuracy against a simplified mathematical treatment in terms of ordinary differential equations (ODE) and hence permits further analytical insights. We will leverage this so-called Markovian approach in Chap. 3 and Chap. 5.

The second transition $S \to I$, i.e., the infection process is considerably more complicated due to the complex interaction patterns between susceptible and infected individuals. At this point, mobility data and contact networks will play an essential role, but first, we will look at the simplest, so-called *homogeneous mixing* approximation [10]. Here, we assume that contacts change instantaneously with no underlying structure or correlation. Hence, the probability for an individual to get in contact with an infected equals the corresponding fraction $\rho^I = M^I/M$ of the population, thereby we denote with $M^I(t)$ and M, the number of infected at time t and the total population size, respectively. We further assume that a susceptible individual contracts the disease upon contact with a constant rate β. Hence, every member of the population undergoes the transition $S \to I$ at a constant rate α, also known as the *force of infection*:

$$\alpha = \beta \rho^I. \tag{2.1}$$

Alternative definitions of α are possible depending on the transmission mechanism. For example, a common approach to account for the mean degree $\langle k \rangle$ of a population is to scale the infection rate with the average number of contacts, i.e., $\alpha = \langle k \rangle \beta \rho^I$. In other cases, α scales with the population size [10], integrates higher order network properties such as clustering [133] or accounts for recurrent mobility patterns [134].

The diagram in Fig. 2.1 summarizes our Markovian model of an SIR-type infection with the homogeneous mixing assumption. It represents the flow of

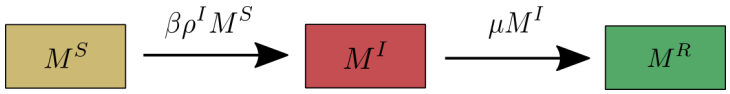

Figure 2.1: Flow chart of the SIR model with transitions between the following compartments: (S) - susceptible, (I) - infected (R) - recovered or immune.

individuals from the susceptible compartment to the infected and finally to the recovered. Here, we denote with $M^X(t)$ the number of individuals in a compartment $X \in \{S, I, R\}$ at the time t.

2.1.1 Dynamic mean-field model

Our Markovian model with homogeneous mixing, in short *mass-action* (MA) model, as depicted schematicly in Fig. 2.1 translates directly into the following set of coupled ordinary differential equations (ODE) for the number of individuals M^X in each compartment $X \in \{S, I, R\}$:

$$\frac{\mathrm{d}M^S(t)}{\mathrm{d}t} = -\beta\rho^I(t)M^S(t) \tag{2.2a}$$

$$\frac{\mathrm{d}M^I(t)}{\mathrm{d}t} = \beta\rho^I(t)M^S(t) - \mu M^I(t) \tag{2.2b}$$

The assumption of a constant population size, i.e., $M = M^S + M^I + M^R$, closes the set of dynamic equations. Finally, we normalize Eqs. (2.2) with the population size M and denoting the corresponding density with $\rho^X = M^X/M$, we obtain:

$$\frac{\mathrm{d}\rho^S(t)}{\mathrm{d}t} = -\beta\rho^I(t)\rho^S(t) \tag{2.3a}$$

$$\frac{\mathrm{d}\rho^I(t)}{\mathrm{d}t} = \beta\rho^I(t)\rho^S(t) - \mu\rho^I(t) \tag{2.3b}$$

together with the constraint $1 = \rho^S + \rho^I + \rho^R$. The MA model offers a simple and intuitive approach to analyze an SIR type epidemic. In Fig. 2.2, we give an example of a typical time evolution.

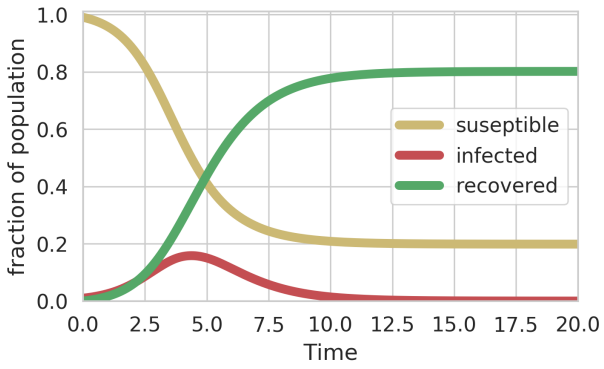

Figure 2.2: Time evolution of an SIR epidemic. The recovery and infection rate are set to $\beta = 1$ and $\mu = 2$, respectively.

Linear stability theory. In the following, we derive basic results of the MA model that are also characteristic to complex, network-based models as we will see later in the chapter. First, we study the fixed point

$$(\rho_*^S,\, \rho_*^I,\, \rho_*^R) = (1,\, 0,\, 0) \tag{2.4}$$

that corresponds to a fully susceptible population by means of a linear stability analysis. To this end, we consider a small fraction of initially infected individuals $\rho^I(0) \ll 1$ and expand Eqs. (2.3) to the first order in ρ^I:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho^I \simeq \beta\rho^I - \mu\rho^I \tag{2.5a}$$

$$= (\beta - \mu)\rho^I. \tag{2.5b}$$

The solution to Eq. (2.5b) is an exponential with growth rate $\lambda = \beta - \mu$:

$$\rho^I(t) = \rho^I(0)e^{\lambda t} \tag{2.6}$$

We note that a small perturbation dies out and hence, the fixed point remains stable if $\lambda < 0$. In other words, an epidemic outbreak remains local if the disease parameters satisfy:

$$\frac{\beta}{\mu} < 1. \tag{2.7}$$

The critical condition $\beta/\mu = 1$ defines a one-dimensional manifold in the configuration space. Here, the fixed point in Eq. (2.4) changes stability in a transcritical bifurcation [135]. From Eq. (2.6), we expect an early exponential growth in the number of infected if $\beta/\mu > 1$. Eventually, the dynamics reaches a maximum prevalence level after a characteristic time before the epidemic starts to die out (see Fig. 2.2). Thus, for $t \to \infty$, all infected individuals will eventually enter the absorbing state R. The fraction of recovered individuals at the end of an epidemic defines the outbreak size $\rho^R(\infty) \equiv R_\infty$, which is a central quantity in mathematical epidemiology. The new asymptotically stable fixed point

$$(\rho_*^S,\, \rho_*^I,\, \rho_*^R) = (1 - R_\infty,\, 0,\, R_\infty). \tag{2.8}$$

indicates that a considerable fraction of the population has contracted the disease.

In order to calculate R_∞ within the MA model, we derive the following ODE from the dynamic equations in Eqs. (2.3):

$$\frac{\mathrm{d}\rho^S}{\mathrm{d}\rho^R} = \frac{\mathrm{d}\rho^S}{\mathrm{d}t} \cdot \left(\frac{\mathrm{d}\rho^R}{\mathrm{d}t}\right)^{-1} \tag{2.9a}$$

$$= -\frac{\beta}{\mu}\rho^S. \tag{2.9b}$$

After separating the variables, we integrate from $t = 0$ to $t = \infty$ and use the initial condition $\rho^S(0) = 1$:

$$\int_0^{1-R_\infty} \frac{\mathrm{d}\rho^S}{\rho^S} = \frac{\beta}{\mu} \int_0^{R_\infty} \mathrm{d}\rho^R. \tag{2.10}$$

With this, we obtain the transcendental equation

$$R_\infty = 1 - \exp\left(-\frac{\beta}{\mu} R_\infty\right). \tag{2.11}$$

The trivial solution $R_\infty = 0$, which solves Eq. (2.11), is stable for parameter values β and μ that satisfy

$$\frac{\mathrm{d}}{\mathrm{d}R_\infty} \left[1 - \exp\left(-\frac{\beta}{\mu} R_\infty\right)\right]\Big|_{R_\infty=0} < 0 \Leftrightarrow \frac{\beta}{\mu} < 0. \tag{2.12}$$

For $\beta/\mu > 0$, however, the trivial solution becomes unstable and a stable non-trivial fixed point $R_\infty > 0$ appears. The latter can be determined by iteration with arbitrary accuracy. The condition in Eq. (2.12) is identical to Eq. (2.7) from the linear stability analysis of the dynamical model in Eqs. (2.3). In Fig. 2.3, we present the solution to Eq. (2.11) and observe the transition from the subcritical to the supercritical parameter region.

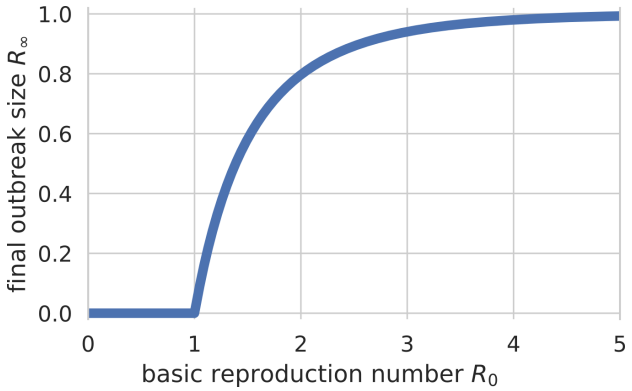

Figure 2.3: Final outbreak size R_∞ according to Eq. (2.11) of the deterministic MA model.

The strength of the MA model lies in its ability to describe key characteristics of an SIR-type epidemic within an intuitive framework. In this section, we focused in particular on the transition from local to global epidemics. However, the model suffers from well-known weaknesses, three of which we want to highlight: (i) We started from a homogeneous mixing assumption where all individuals can get in contact with each other. In reality, however, interactions are sparse and strongly heterogeneous. As a consequence, highly connected individuals, so-called superspreaders, accelerate the early spread of the infection [136]. (ii) We assumed implicitly that contact durations are infinitesimally short. However, static networks or graphs that evolve at the same time scale as the dynamic process provide a better approximation in many cases. In a static environment, for instance, it is possible that most contacts of an initially infected individual have already contracted the disease thus limiting the ability of the latter to cause new infections.

In the literature, this effect is also called local depletion of susceptible individuals [137]. (iii) The deterministic model neglects stochastic fluctuations that may cause, for instance, an early extinction of a supercritical infection. In other words, the exponential growth condition $\beta/\mu > 1$ is not sufficient to observe a global outbreak. Dynamic equations, such as the MA model discussed here, are only meaningful if the prevalence is sufficiently large and thus fluctuations become negligible. Epidemiological models that include stochastic effects exist, for example, within the framework of branching processes [138] or as approximations of the master equation [100, 139]. Methods from percolation theory, which we will discuss later, can also add valuable insights to understand stochastic effects, particularly for SIR-type infections.

In connection with the critical transition from local to global epidemics as shown in Fig. 2.3, it is important to introduce the *basic reproduction number* R_0. This central quantity of mathematical epidemiology indicates the mean number of secondary infections caused by a typical infected individual in a fully susceptible population [10]. This definition clearly shows that $R_0 = 1$ marks the transition from the sub- to the supercritical parameter range: For $R_0 < 1$ an infected individual leads to less than one secondary infection, and thus an outbreak dies out eventually. However, if $R_0 > 1$, then a multiplicative growth process sets in that facilitates a global epidemic. The basic reproduction number integrates both, the disease parameters and the underlying contact structure. Therefore the functional form of R_0 depends on the details of a given epidemiological and topological model.

For the MA SIR model, we can determine R_0 from the following consideration: The average time to recovery equals μ^{-1}. Then, at any given time $t \in [0, \mu^{-1})$ during the infection period, the considered individual is in contact with a randomly selected partner for an infinitesimal short moment $\mathrm{d}t$. The latter is likely to be susceptible because we assume an early stage of the epidemic and thus the susceptible node contracts the disease with probability $\beta \mathrm{d}t$. Integrating over the entire infection period, we obtain the expected number of secondary infections:

$$R_0 = \int_0^{\mu^{-1}} \beta \mathrm{d}t \tag{2.13}$$

$$= \beta/\mu. \tag{2.14}$$

We already derived the dimensionless parameter β/μ in Eq. (2.7) from the early-growth rate λ and it is not surprising that different approaches lead to the same critical value [140]. However, it must be added that the *Malthusian parameter* λ and the basic reproduction number R_0 are fundamentally different. While the former averages the population and determines the early growth at a fixed time, the latter is a time-averaged quantity that refers to a single representative individual [137].

Percolation theory. As mentioned earlier, methods of non-equilibrium statistical physics can help to formulate epidemiological models that include stochastic effects [141, 142]. In particular, we will now highlight some parallels between the transcritical bifurcation that we observed in Fig. 2.3 and percolation theory. The latter is concerned with phase transitions of systems that are out of thermal equilibrium. A textbook example is the sudden magnetization of a ferromagnet with increasing strength of an external magnetic field [143]. This example belongs to the class of continuous phase transitions, which also include the abrupt occurrence of an epidemic as the infection rate increases [144, 145]. Continuous phase transitions are characterized by an order parameter such as the outbreak size R_∞ (see Fig. 2.3). Infinitely large systems move from a sub- to a super-critical phase with $R_\infty = 0$ and $R_\infty > 0$, respectively, as the control parameter, here the basic reproduction number R_0, increases. Among others, an important insight from percolation theory is the existence of universal exponents that characterize the phase transition for a broad class of models.

For example, contrary to the representation in Fig. 2.3 the outbreak size R_∞ increases according to a power law close to the percolation threshold $R_0^{\text{crit}} = 1$, i.e., $R_\infty \sim |R_0 - R_0^{\text{crit}}|^{\eta_1}$ with some universal exponent η_1 [146]. Of practical importance for the present work is the behavior of local epidemics. Here, percolation theory predicts that the mean outbreak size diverges close to R_0^{crit} according to $R_\infty \sim |R_0 - R_0^{\text{crit}}|^{-\eta_2}$, where $\eta_2 > 0$ denotes another universal exponent. We will use this divergent behavior in Chap. 5 to determine numerically the percolation threshold.

2.2 Complex network analysis

Realistic contact and mobility structures are sparse, heterogeneous and can often be considered static compared to the time scale of the spreading process. Thus they differ distinctly from the MA model that we discussed in the previous section. Moreover, numerous topological features exist with a nonlinear effect on the outcome of an epidemic. A natural representation of complex interaction patterns on a finite set of elements are networks, the analysis of which plays an ever-increasing role in the mathematical modeling of epidemic outbreaks as a vast amount of data becomes publicly available.

This section provides a brief introduction with a focus on empirical networks. First, we introduce the basic terminology and concepts for networks with static and time-varying topologies. Then, we present and apply selected methods to the German pig-trade network [147]. In doing so, we specifically address features of the aggregated graph such as component structure and centrality measures, as well as characteristics of the temporal network such as the distribution of waiting times and time-respecting accessibility. These results are based on our publication [72].

Static topologies. A static network or graph $G = (\mathcal{N}, \mathcal{E})$, is defined by a set of nodes (vertices) and edges (links), denoted by $\mathcal{N}$ and $\mathcal{E}$, respectively. Each edge is a tuple $(i, j) \in \mathcal{E} \subset \mathcal{N} \times \mathcal{N}$ and represents a relation between adjacent nodes i and j. An edge can be considered directed or undirected, depending on the context such as trade relations or human friendships, respectively. A useful and widespread representation of a network is the adjacency matrix $\boldsymbol{A}$:

$$(\boldsymbol{A})_{i,j} = \begin{cases} 1 & \text{, if } (i,j) \in \mathcal{E} \\ 0 & \text{, otherwise.} \end{cases} \tag{2.15}$$

Entries on the diagonal represent self-loops, i.e., nodes with edges that point back to themselves. The sum $\sum_{j=1}^{N} A_{i,j}$ over all entries in row i is called out-degree k_i^{out} and indicates the number of nodes that can be reached from i through one edge. Similarly, the column sum $\sum_{j=1}^{N} A_{i,j}$ is called in-degree k_i^{in}. For undirected graphs, the adjacency matrix $\boldsymbol{A}$ is symmetric because each entry $A_{i,j} = 1$ implies an edge in the opposite direction, i.e., $A_{j,i} = 1$. Therefore the row sum is equal to the column sum and one simply speaks of the degree k_i of the node i:

$$k_i = \sum_{j=1}^{N} A_{i,j}. \tag{2.16}$$

Powers of the adjacency matrix provide information about *walks*, i.e., a sequence of edges

$$\omega(s \to t) = [(u_0, v_0), (u_1, v_1), ..., (u_{n-1}, v_{n-1})] \tag{2.17}$$

where $(u_i, v_i) \in \mathcal{E}$ for all $i = 0, ..., n-1$. Also, $s \equiv u_0$ and $t \equiv v_{n-1}$ denote the first and last node, respectively. Hence, one also speaks of st-walks. Consider for example the element $(\boldsymbol{A}^3)_{s,t}$: It gives the number of walks that lead from s to t after three steps, whereby a walk can visit a node multiple times. Diagonal elements indicate the number of walks that return to the first node, so-called closed walks. In our example, $(\boldsymbol{A}^3)_{s,s}$ indicates the number of triangles originating in node s. In a social context, triangles are a central measure for the analysis of friendship networks [68, 148].

An absorbing walk is a special case when the target node t is visited only once. If every node appears exactly once in the sequence, we speak of a *path* π. For many applications the term *shortest path* is central, and denotes the (not neccessarily unique) path that minimizes the number of edges. As an example, networks are frequently characterised by the average shortest path between all pairs of nodes and the so-called *diameter*, i.e., the largest shortest path.

Powers of the adjacency matrices also lead us to the concept of accessibility. To this end, we first define the accessibility matrix $\boldsymbol{R}$ with binary elements

$$(\boldsymbol{R})_{s,t} = \begin{cases} 1, & \text{if a path } s \to t \text{ exists} \\ 0, & \text{otherwise} \end{cases} \tag{2.18}$$

For a static network with N nodes, we can calculate $\boldsymbol{R}$ as follows:

$$\boldsymbol{R} \sim \sum_{n=1}^{N} \boldsymbol{A}^n. \tag{2.19}$$

The symbol "$\sim$" indicates that we are not interested in the exact value of an element $R_{s,t}$, which provides the number of st-walks. Instead, we set a positive element to one and thus treat $\boldsymbol{R}$ as a Boolean matrix [149]. We will present the details of Boolean arithmetics in Chap. 4.

For weighted networks, we define an edge by a triple $(i, j, F_{i,j})$ with $w \in \mathbb{R}^+$. The weight $F_{i,j}$ can be interpreted as the intensity or frequency of a contact. We denote the corresponding weighted adjacency matrix with $\boldsymbol{F}$, where we replace positive (binary) entries in Eq. (2.15) with the corresponding edge weight:

$$(\boldsymbol{F})_{i,j} = \begin{cases} F_{i,j} & \text{, if } (i, j, F_{i,j}) \in \mathcal{E} \\ 0 & \text{, otherwise.} \end{cases} \tag{2.20}$$

Hence, the i-th row sum is called *strength* F_i of node i or correspodingly the in- / out- / total-strength for directed networks:

$$F_i = \sum_{j=1}^{N} F_{i,j} \tag{2.21}$$

Important to mention in the context of weigted graphs is the (weighted) shortest path from s to t. It minimizes the sum of edge weights along all st-paths. A particularly efficient method for finding the shortest path is the Dijkstra-algorithm [150]. It will be a central tool in Chap. 3. and hence we will leave the details to the main chapter.

Time-varying topologies. The temporal resolution of contact and mobility data has recently allowed researchers to study the dynamics of networks and the influence of time-varying topologies on propagation processes [151–154]. *Temporal networks* or time-varying graphs are a natural model for interconnected systems where relationships fluctuate in time.

In this thesis, we focus on temporal networks, sampled at equidistant time intervals. Hence, we can represent the time-evolving graph $\mathcal{G}$ by a sequence

$$\mathcal{G} = [G(0), G(1), ...G(T-1)] \tag{2.22}$$

of T sample networks $G(t)$ with $t = 0, 1, ..., T-1$. Each sample represents the state of the system at a given time t, i.e., the configuration of active links. Furthermore, each graph of the sequence can be either directed or undirected, and can potentially include edge weights. An obvious representation of the temporal network is a

sequence of adjacency matrices

$$\mathcal{A} = [\boldsymbol{A}(0), \boldsymbol{A}(1), ...\boldsymbol{A}(T-1)] \tag{2.23}$$

and correspondingly $\mathcal{F}$ for weighted temporal graphs.

If we ignore the temporal dimension and consider all edges as concurrent in time, we obtain the *(time-)aggregated graph*. To this end, one usually takes the sum over all adjacency matrices:

$$\boldsymbol{A} = \sum_{t=0}^{T-1} \boldsymbol{A}(t) \tag{2.24}$$

Here, $\boldsymbol{A}$ can be considered either binarized, i.e., $(\boldsymbol{A})_{i,j} \in [0, 1]$ or otherwise, values indicate the number of contacts within the observation time.

There are numerous alternative methods for constructing a static graph from a temporary network [83, 155]. These approaches attempt to minimize the information loss due to the aggregation process. However, in no case is it possible to fully capture the complex temporal correlations that characterize empirical networks with time-varying topologies. Nor does a static representation succeed in reproducing causal contact patterns.

To explain the latter, we consider the example in Fig. 2.4. The aggregated graph suggests a connected network, i.e., for each pair of nodes, there is a path connecting both. However, the time-resolved representation shows that an epidemic beginning in node three, for instance, cannot reach node four through a sequence of temporal edges. In order to account for the correct timing of edges, necessary to reach a distant node, one defines a *time-respecting walk* (or similarly a *time-respecting path*) as a sequence

$$[(u_0, v_0, t_0), (u_1, v_1, t_1), ..., (u_{n-1}, v_{n-1}, t_{n-1})]. \tag{2.25}$$

Here, each temporal edge $(u_i, v_i, t_i) \in \mathcal{E}(t)$ appears in a chronological order, i.e.

$$t_0 \leq t_1 \leq ... \leq t_{n-1}. \tag{2.26}$$

Many methods from static network theory rely on paths to determine the importance (also called centrality) of a node [156, 157]. Due to the causality condition in Eq. (2.26), however, corresponding measures for time-varying topologies are frequently missing. On the other hand, the concept of accessibility remains the same and thus may help to generalize methods from classical network theory: In analogy to the static case (see Eq. (2.18)), we define a time-respecting accessibility matrix $\mathcal{R}$ with

$$(\mathcal{R})_{i,j} = \begin{cases} 1, & \text{if a time-respecting path } i \to j \text{ exists} \\ 0, & \text{else.} \end{cases} \tag{2.27}$$

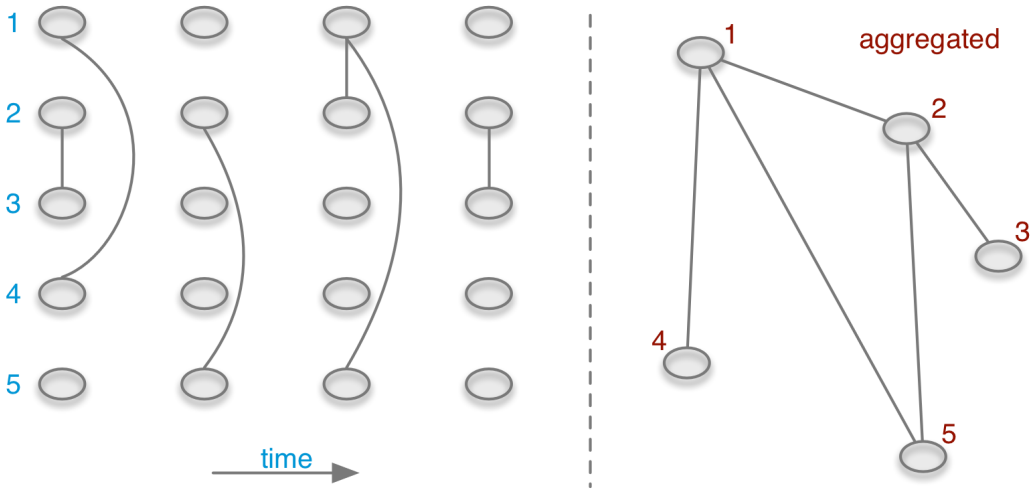

Figure 2.4: A temporal network (left) and the corresponding time-aggregated graph (right).

For a temporal network $\mathcal{G}$ defined by T samples, we can calculate $\mathcal{R}$ as follows [71, 73]:

$$\mathcal{R} \sim \prod_{t=0}^{T-1} [\mathbb{1} + \boldsymbol{A}(t)] \tag{2.28}$$

Here $\mathbb{1}$ is the $N \times N$ identity matrix. This formalism, which is termed *unfolding accessibility*, will provide the basis for a matrix-based disease model in Chap. 4 of this thesis.

In Fig. 2.5, we present the temporal accessibility network derived from the example in Fig. 2.4. In comparing the accessibility graph to the time-aggregated counterpart, we find that every time-respecting path exists in the static network but, not necessarily the other way around. Consequently, a simulated disease outbreak on the time-aggregated graph can reach more nodes and thus overestimate the actual outbreak size. This observation demonstrates once again the importance of temporal networks for realistic disease models.

In the following section, we will present selected methods for analyzing static and time-varying graphs. For illustration, we apply these measures to the German pig-trade network, which is considered to be a major transmission route for animal-related diseases like classical swine fever [158]. The results of the following section are based on our publication in [159].

2.2.1 Network analysis of pig-trade in Germany

The dataset under consideration represents an excerpt of the national German livestock database HI-Tier [147], which has been established according to EU legislation. It comprises the movement of pigs within 1461 days from 2011-01-01 to 2014-12-31 with a time resolution of one day. From this, we derive a network in which nodes and directed edges represent livestock premises and trade connections. Furthermore, we consider the trade volume as additional edge weights. Basic information about the network is listed in Tab. 2.1.

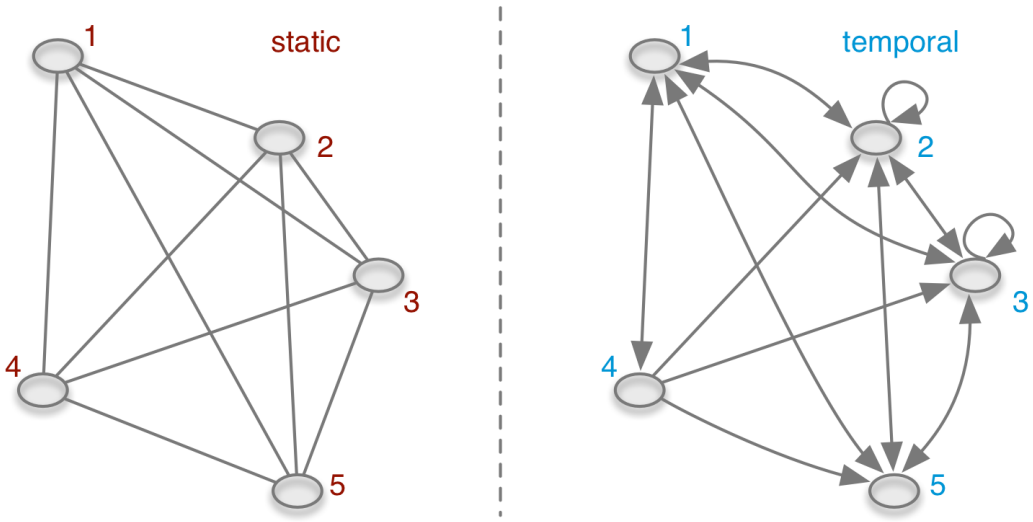

Figure 2.5: Comparison between time-respecting and static accessibility graph. All nodes in the static accessibility graph (left) contain self-loops (not shown). Note that the temporal accessibility graph is directed although the underlying temporal network is undirected.

Number of nodes N	97,980
Number of edges E	315,333
Edge density E/N^2	3.2×10^{-5}
Size of GSCC	28%
Size of GWCC	99%
Diameter	18
Average shortest path length	5.5

Table 2.1: Standard network properties of the static German pig-trade network. The giant strongly connected component (GSCC) as well as the giant weakly connected component (GWCC) are schematically shown in Fig. 2.6. Diameter and shortest path length are computed for the GSCC.

2.2.1.1 Static network analysis

We begin with an analysis of the time-aggregated and unweighted trade network. Hence, an edge indicates at least one recorded animal movement during the observation time between the corresponding premises.

Large scale structure - component analysis. Regardless of the disease model, the spread of an epidemic is limited by the component structure. Therefore, we will examine components of the time-aggregated network and emphasize that directionality of trade plays an essential role.

We begin, however, by neglecting the directionality of edges, which allows for a first (though very coarse) structural analysis. Given an undirected network G we define a connected component (CC) as a subgraph $G_0 = (\mathcal{N}_0, \mathcal{E}_0)$ in which any two nodes i and j with $i, j \in \mathcal{N}_0 \subseteq \mathcal{N}$ are connected through a path.

Many undirected empirical networks frequently exhibit a *giant connected component*. If the original network is directed, as in our case, one also speaks of a *giant weakly connected component* (GWCC). The GWCC of the pig-trade network

contains the majority of all nodes (99%), and hence the remaining disconnected components are of negligible size. The occurrence of a giant component has been observed and investigated in a variety of systems [54, 160, 161]. Indeed, it reflects a universal phenomenon, which is the subject of percolation theory. The latter suggests that a dominant component appears if the number of edges exceeds a critical threshold.

If we include the directionality of edges, then we obtain a far more diverse picture, which we depict schematically in Fig. 2.6 (see also [160]). By analogy with the GWCC, we define a *giant strongly connected component* (GSCC) through the largest subset of nodes $\mathcal{N}_0 \subseteq \mathcal{N}$, where every pair $i, j \in \mathcal{N}_0$ is connected through a path (considering the directionality of edges, see red box in Fig. 2.6). Note that a path from i to j does not imply the opposite direction, i.e., from j to i. In the data set under consideration, approximately 1/4 of all nodes belong to the GSCC. This subgraph forms the functional backbone in the sense that it provides global connectivity. From an epidemiological point of view, its existence facilitates the spread of a global epidemic.

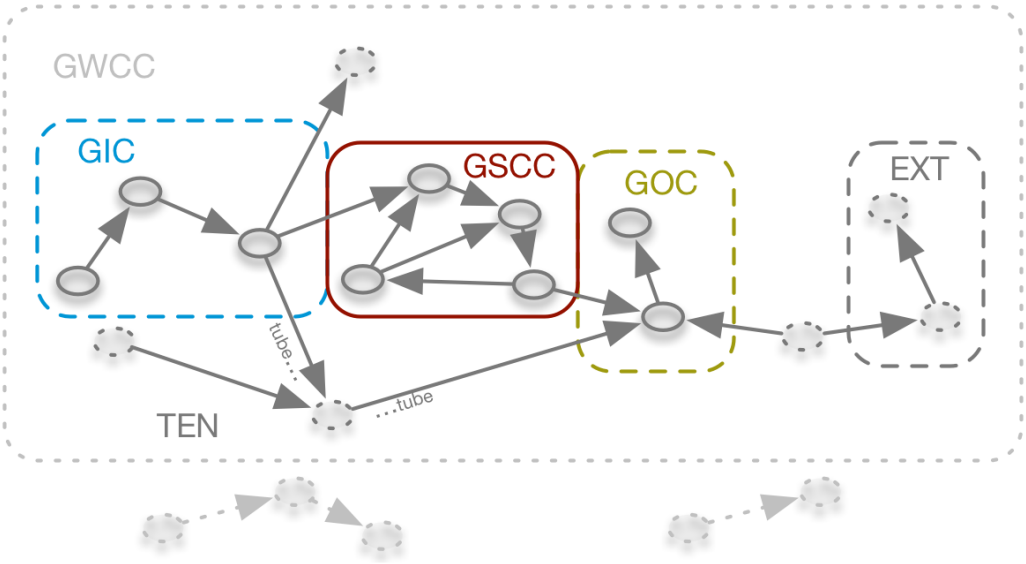

Figure 2.6: Schematic component structure of directed networks. The giant strongly connected component (GSCC) forms the functional core of the network (red box). Nodes of the GSCC and the giant in-component (GIC, dashed blue box) have a high spreading potential, whereas all other nodes (GOC—giant out component, dashed yellow box; TEN—tendril; EXT—external nodes, grey dashed box) lead to local outbreaks. The giant weekly connected component (GWCC) is given by the grey dotted box.

Furthermore, all nodes with access to the GSCC are called *giant in-component* (GIC). Similarly, the *giant out-component* (GOC) contains all nodes accessible from the GSCC. Of lesser importance are *tendrils* (TEN) - Nodes accessible from the GIC or those reaching the GOC but, that are not part of the GSCC. Finally the remaining nodes of the GWCC are called *external nodes*. (EXT).

In Fig. 2.7, we present the distribution of nodes and edges, categorized according to the component structure. Particularly significant from an epidemiological point

of view is the fact that more than half of all nodes belong either to the GIC or to the GSCC. These nodes can in principle cause a global epidemic.

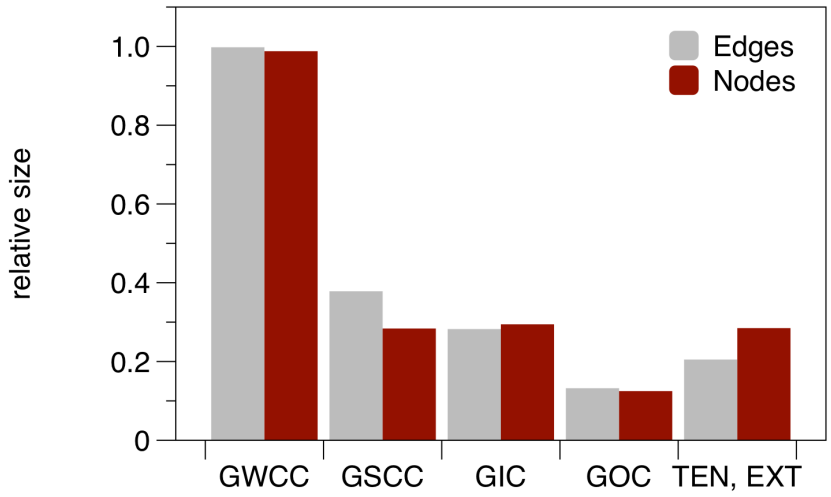

Figure 2.7: Component structure of the pig-trade network. Sizes are normalized to the total number of nodes or edges, respectively, in the network.

In order to quantify the spreading potential of a single node $i \in \mathcal{N}$, we consider the number of nodes accessible from i via a (directed) path. This measure, called *range*, provides a simple, model-independent risk assessment. The accessibility matrix $\boldsymbol{R}$ from Eq. (2.19) provides an efficient approach to calculating the range for all pairs of nodes. Here, a positive entry in row i indicates if a node is accessible from i. Hence, the row sum gives the total number of reachable nodes:

$$\text{range}(i) = \sum_{j=1}^{N} R_{i,j} \tag{2.29}$$

The accessibility statistics in Fig. 2.8 shows a clear bimodal structure: (i) nodes with a high infection potential (about 58%, that is 56,700 nodes) that are either part of the GIC or GSCC. (ii) nodes that can infect only a negligible part of the network (maximum 70 nodes).

Small scale structure - centrality. In order to implement effective control strategies to mitigate the impact of a disease, it is necessary to assess the risk posed by individual nodes. For this purpose, a variety of *centrality measures* have been proposed including our previously defined *range* (see Eq. (2.29)) [162]. The simplest network measure, though, is the *degree* of a node. For directed networks, we differentiate between out-, in- and total-degree. The degree distribution found in empirical networks is typically heavy-tailed and in case of the pig-trade network the out-degree can be asymptotically approximated by a power law $p_k \sim k^{-\mu}$, with $\mu \approx 2.7$ (see Fig. 2.9). The broad distribution indicates a large number of well-connected nodes, that can potentially accelerate the early spread of an outbreak. These hubs are first candidates for targeted surveillance strategies [32].

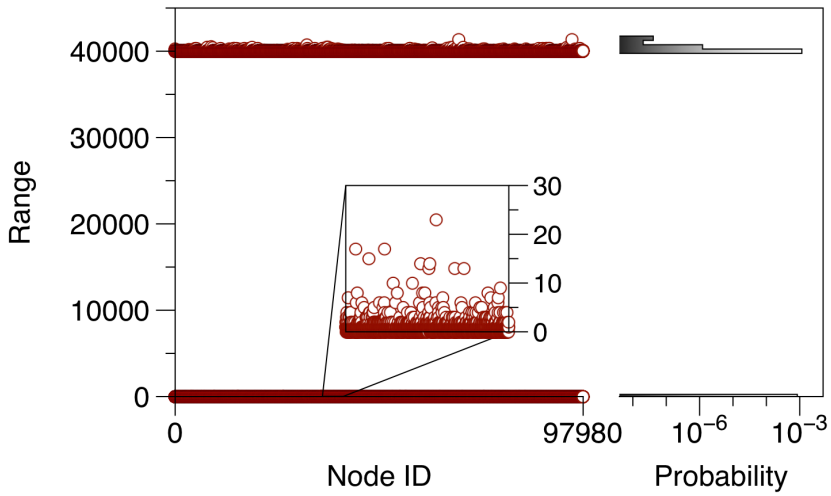

Figure 2.8: Range for every node in the pig-trade network. About 58% of the nodes have access to a significant part of the network - approximately 40,000 nodes. These nodes can therefore lead to global outbreaks, affecting almost half the network (assuming static links). The remaining nodes have a short range and may only lead to local outbreaks (with a maximum range of 70 nodes). The right panel shows the corresponding histogram.

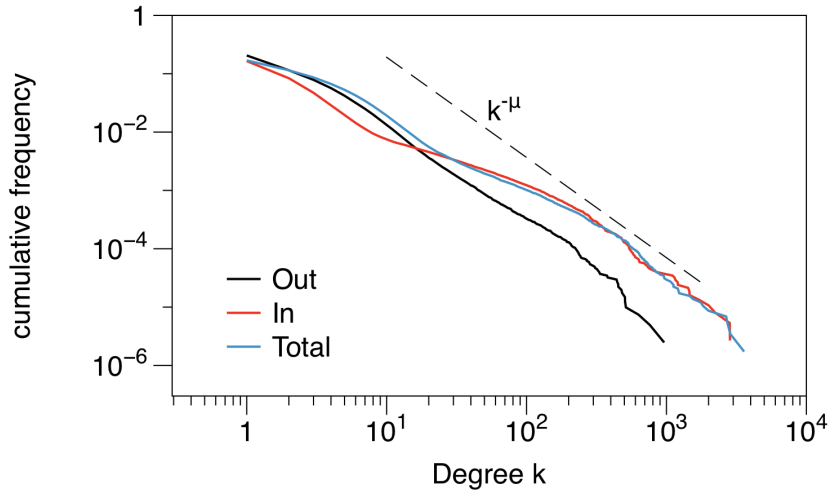

Figure 2.9: Degree distribution of the German pig-trade network. The figure shows the complementary cumulative distribution to minimize fluctuations. We used a maximum likelihood approach [69] to fit a power law of the form $p_k \sim k^{-\mu}$ with $\mu \approx 2.7$ to the out-degree distribution (black).

Weighted network analysis. The number of animals traded, i.e., the edge weight, plays an important role in the spread of infectious diseases. From a modelling point of view, the edge weight can, for example, define the probability to transmit a disease from an infected node to its susceptible neighbour. Similar to the degree, we find in Fig. 2.10 a broad distribution of in- and out-strength, i.e., the total number of animals that entered and left a node, respectively, within the observation period. Nodes that dominate the distribution in Fig. 2.10 are not necessarily hubs in terms of the degree and hence, this additional information allows to refine surveillance strategies.

The additional topological information also changes the shortest path statistics. We find that the mean shortest path length is 9.7 and the diameter increases

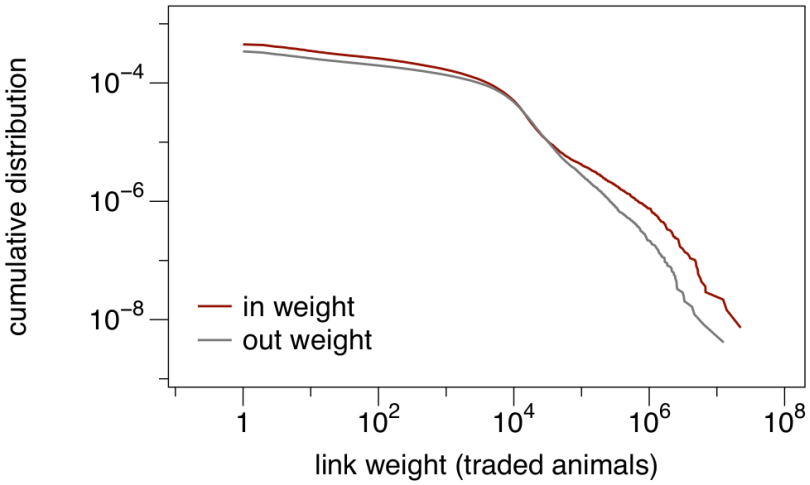

Figure 2.10: Strength distribution of the German pig-trade network. The figure shows the complementary cumulative distribution for the in- and out-strength to minimize fluctuations.

to 30. Thus both measures are twice as large compared to the unweighted case (see Tab. 2.1). Since highly contagious diseases (such as classical swine fever, Aujeszky's disease, foot-and-mouth disease) spread predominantly along shortest distances [114, 115], this observation emphasises again the importance of edge weights for epidemiological models.

2.2.1.2 Temporal network analysis.

In the previous section, we aggregated the temporal dimension over the full observation period. With this, however, we lost relevant information for realistic disease models. Next, we include temporal information and discuss some characteristic features.

In the following we consider the animal trade network as a time-varying graph

$$\mathcal{G} = [G(0), G(1), ..., G(T-1)] \tag{2.30}$$

with the observation time of $T = 1461$ days and a temporal resolution of $\Delta t = 1$ day. A time-stamped edge $(i, j, t) \in G(t)$ indicates the transport of animals from i to j at time t.

Waiting time distribution. The activation time of an edge, as well as the inter-activation time, are central characteristics of temporal networks. In our case, the latter, i.e., the inter-activity or waiting time, is particularly relevant. Fig. 2.11 demonstrates a heterogeneous distribution typical for empirical data with a cut-off at the observation time of $T = 1461$ days. The broad distribution refutes a typical assumption on edge weights in static graphs. To explain this, we consider a temporal network that evolves much faster than the spreading process, i.e., the *annealed limit*. Here, the topology can be considered as static and edge weights are given by the mean number of contacts within a certain time, in other words,

the contact rate. However, a constant rate assumes that the underlying temporal network is Markovian, i.e., generated from a poisson-process. In this case, we would expect exponentially distrubuted waiting times, which clealy contradict our observations in Fig. 2.11.

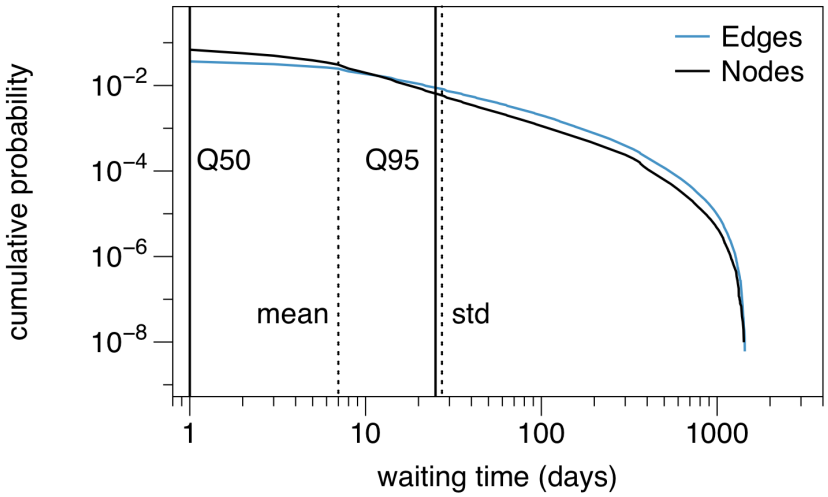

Figure 2.11: Distribution of waiting times (also called inter-activity times) for nodes and edges. Dashed lines show mean and standard deviation, respectively. Solid lines show median and 95% quantile of the node waiting time distribution.

Temporal accessibility. We have already discussed in the previous section leading to Eq. (2.27) how to translate the concept of accessibility from static to temporal networks. In analogy to the static case, we use the time-respecting accessibility matrix $\mathcal{R}$ (see Eq. (2.28)) to calculate the range of a node i:

$$\text{range}(i) = \sum_{j=1}^{N} \mathcal{R}_{i,j} \tag{2.31}$$

In Fig. 2.12, we observe that the time-respecting range deviates significantly from our previous observation on the aggregated network (see Fig. 2.8). There exists still a bimodal structure, but now we find values for the entire spectrum. This result allows for a more differentiated risk assessment of individual nodes. A further improvement of this centrality measure will be the topic of Chap. 4 in the main part of this thesis.

To summarize, we introduced in this section the basic concepts of modern network analysis. In particular, we addressed conceptual differences between static and time-varying topologies and presented selected methods for the analysis of empirical data. As an example of a temporal network, we studied the German pig-trade and highlighted among others, that the mass-action assumption of fast and randomly changing links deviates strongly from empirical observations. It is, therefore, essential to integrate complex networks into the classical epidemiological models, which wil be the subject of Sec. 2.3.3.

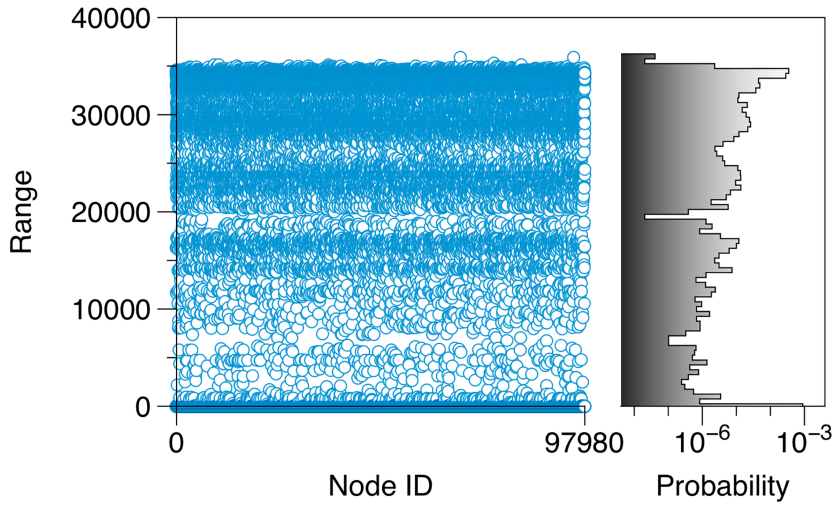

Figure 2.12: The individual range for every node in the temporal network. The right panel shows the corresponding histogram on a log scale. In contrast to the static case (see Fig. 2.8) the values cover the whole spectrum.

2.3 Dynamics on networks

So far, we discussed in Sec. 2.1 the spread of diseases in a population with homogeneous mixing and analyzed in Sec. 2.2 empirical contact and mobility structures using modern network theory. In this section, we will combine dynamical and complex topologies leading to a novel class of data-driven models. To this end, we start in Sec. 2.3.1 with a generic diffusion model, namely the random-walker on a static network. In Sec. 2.3.2, we connect the diffusive transport of individuals along the edges of a network with a local SIR dynamic to the class of *meta-population models*. Finally, in Sec. 2.3.3, we consider each node as an epidemiological unit and investigate the transition from local to global outbreaks in what is known as the *individual-based model*.

2.3.1 Random-walk on static graphs

We begin with what is probably the simplest and yet one of the most fundamental dynamic models: a random-walk on a graph. This is a stochastic process whose future state is conditioned only on the current position, i.e., independent of previously visited nodes. Starting from this *Markovian assumption*, a vast number of models has been proposed to describe phenomena that range from the diffusive transport in semiconductors [163] to the human mobility patterns of global traveling [3].

In this section, we first introduce the discrete-time random-walk (DTRW), where the number of jumps parameterizes time. Then, we generalize to the continuous-time random-walk (CTRW) and focus on exponentially distributed waiting times.

2.3.1.1 Discrete-time random-walks

A DTRW on a finite graph $G = (\mathcal{N}, \mathcal{E})$ with N nodes is also called *finite Markov chain* or more humorously a *drunkard's walk*. We assume that the underlying network is undirected and weighted. Eventhough, a generalization to directed graphs is possible, we restrict ourselves to the undirected case in order to focus on the basic concepts. We also assume that our graph is connected so that a walker can reach every node starting from any other position in the network.

We now focus on a single walker that moves between adjacent nodes at discrete time steps. The transition probability $T_{i,j}$ from node i to j is proportional to the edge weight $F_{i,j}$. The latter is an element of the weighted adjacency matrix $\boldsymbol{F}$ (see Eq. (2.20)). From the normalization condition $\sum_{j=1^N} T_{i,j} = 1$ that the walker chooses a node $j \in \mathcal{N}_i$ from its neighborhood $\mathcal{N}_i$, we determine the transition probability

$$T_{i,j} = \frac{F_{i,j}}{\sum_{j=1}^{N} F_{i,j}}. \tag{2.32}$$

The denominator is identical to the previously defined strength F_i (see Eq. (2.21)). Next, we interpret $T_{i,j}$ as an element of the *transition matrix* $\boldsymbol{T}$ with the row sum

$$\sum_{j=1}^{N} T_{i,j} = 1 \qquad \forall i \in \mathcal{N} \tag{2.33}$$

reflecting the normalization condition.

As a next step, we turn to an ensemble of random-walkers and denote with $p_i(n)$ the probability to find the walker after n steps on the node i. For notational convenience, we also introduce the N dimensional row vector $\boldsymbol{p}(n)$ with elements $p_i(n)$ for all nodes $i \in \mathcal{N}$. The conservation of probability hence requires

$$\sum_{i=1}^{N} p_i(n) = 1 \qquad \forall n \in \mathcal{N}. \tag{2.34}$$

We can now use the transition probabilities $T_{i,j}$ from Eq. (2.32) to construct a dynamic equation for the probability $p_i(n)$:

$$p_i(n+1) = \sum_{j=1}^{N} T_{i,j} p_j(n) \tag{2.35}$$

Now, we integrate this set of N coupled equations into a matrix-based representation using the row vector $\boldsymbol{p}(n)$ and the transition matrix $\boldsymbol{T}$:

$$\boldsymbol{p}(n+1) = \boldsymbol{p}(n)\boldsymbol{T}. \tag{2.36}$$

By applying Eq. (2.36) repeatedly to the initial condition $\boldsymbol{p}(0)$, we finally obtain the explicit solution

$$\boldsymbol{p}(n) = \boldsymbol{p}(0)\boldsymbol{T}^n. \tag{2.37}$$

Stationary state. As an important application, we determine the equilibrium distribution $\boldsymbol{p}^* = \lim_{n\to\infty} \boldsymbol{p}(n)$. To this end, we substitute $\boldsymbol{p}^* = \boldsymbol{p}(n) = \boldsymbol{p}(n+1)$ in Eq. (2.36) and obtain the eigenvalue equation

$$\boldsymbol{p}^* = \boldsymbol{p}^* \boldsymbol{T} \tag{2.38}$$

to the dominant eigenvalue $\lambda_0 = 1$. The equilibrium distribution $\boldsymbol{p}^*$ is thus equal to the (left-sided) dominant eigenvector. Direct substitution confirms the following solution:

$$p_i^* = \frac{F_i}{\sum_{j=1}^{N} F_j}. \tag{2.39}$$

Hence, the probability to find the walker in the late-time limit at node i scales with the node strength F_i.

We approach the stationary distribution once again from another perspective: A central principle of statistical physics is the temporal reversibility of a process in thermodynamic equilibrium. Formally, the detailed balance relation reflects this property:

$$p_i(n) T_{i,j} = p_j(n) T_{j,i}, \quad \forall n = 1, 2, ... \tag{2.40}$$

In words, the probability flow along the edge $i \to j$ offsets the opposite flow $j \to i$ at any time step n (assuming an undirected network). Note, that we do not make any assumptions on the probabilities $p_i(n)$ at this point. From normalization property in Eq. (2.33) and the detailed balance relation (see Eq. (2.40)) we obtain

$$p_i(n) = p_i(n) \sum_{j=1}^{N} T_{i,j} \tag{2.41}$$

$$= \sum_{j=1}^{N} p_i(n) T_{i,j} \tag{2.42}$$

$$= \sum_{j=1}^{N} p_j(n) T_{j,i}. \tag{2.43}$$

Using a matrix notation, we can express Eq. (2.43) as $\boldsymbol{p}(n) = \boldsymbol{p}(n)\boldsymbol{T}$ for all time steps $n = 0, 1, ...$. In comparing this result with Eq. (2.38) we find that a distribution $p_i(n)$ that satisfies the detailed balance Eq. (2.40) describes the stationary state of a random-walk on an undirected graph. In other words, the stationaŕy distribution $p_i(n) \equiv p_i^*$ for all nodes $i \in \mathcal{N}$ follows from the principle of reversibility.

2.3.1.2 Continuous-time random-walks

In order to adequately describe processes that take place in nature, we must regard time as a continuous variable. To this end, we generalize our results of the previous section, where the number of steps n parametrizes time and introduce instead the CTRW [164].

In a CTRW, the walker jumps from node i to j after some waiting time τ. This is a random value drawn from the *probability density function* (PDF) $\phi(\tau)$. In the special case $\phi(\tau) = \delta(\tau - \tau_0)$ the waiting time is deterministically set to $\tau_0 > 0$ and thus we effectively recover the discrete-time process from the previous section. In general, however, any disorder in waiting times will complicate the mathematical description. Numerous classes of functions for $\phi(\tau)$ have been studied in the past 50 years including mostly exponential and heavy-tailed distributions [116, 165, 166]. The latter seems to be especially realistic for many processes in nature. In particular, empirical studies of human activity suggest that waiting times follow a broad distribution [153, 167]. Consistent with these observations is also our temporal analysis of inter-activity times in the German pig-trade (see Fig. 2.11).

In the following, however, we will focus on a Poisson process and thus assume exponentially distributed waiting times:

$$\phi(t) = \gamma \exp(-\gamma t). \tag{2.44}$$

This common approach greatly simplifies mathematical modeling and although realistic distributions are non-exponential, accurate disease models have been proposed based on Eq. (2.44) [3–5].

With Eq. (2.44) we assume that a walker leaves node i at the constant rate γ. The probability flow from i to the adjacent node j is thus equal to $\gamma T_{i,j}$. Let us denote with $p_i(t)$ the probability to find the walker in node i at time t. In order to write down a dynamic equation for $p_i(t)$, we balance the probability flows leading into and out of a node, which gives the *master equation*:

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t} = \gamma \sum_{j=1}^{N} [p_j(t) T_{j,i} - p_i(t) T_{i,j}] \tag{2.45}$$

Next, we turn this linear system of ODEs into a matrix-based representation. Simmilar to the discrete-time case in Eq. (2.36), we define the row vector $\boldsymbol{p}(t)$ with elements $p_i(t)$ for all nodes $i \in \mathcal{N}$. The right side of Eq. (2.45) translates into a linear operator acting on the state vector $\boldsymbol{p}(t)$. Using the transition matrix $\boldsymbol{T}$ and the identity matrix $\mathbb{1}$ we obtain:

$$\frac{\mathrm{d}\boldsymbol{p}(t)}{\mathrm{d}t} = \gamma \boldsymbol{p}(t)[-\mathbb{1} + \boldsymbol{T}] \tag{2.46a}$$

$$= -\gamma \boldsymbol{p}(t) \boldsymbol{D}^{-1} \boldsymbol{L} \tag{2.46b}$$

Here $\boldsymbol{D} = \mathrm{diag}(F_1, F_2, ...F_N)$ is a diagonal matrix with the node strength F_i (see Eq. (2.21)) in the i-th row and column. In addition, we call $\boldsymbol{L}$ the *(combinatorial) Laplacian* of the network, which plays a central role in many diffusion processes [168]:

$$L_{i,j} = \begin{cases} F_i & \text{, if } i = j \\ -F_{i,j} & \text{, otherwise} \end{cases} \tag{2.47}$$

Because of the normalization $\sum_{i=1}^{N} p_i(t) = 1$ the CTRW evolves according to the normalized *random-walk Laplacian* $\boldsymbol{L}'$:

$$\boldsymbol{L}' = \boldsymbol{D}^{-1}\boldsymbol{L}. \tag{2.48}$$

The CTRW is an example of a *subordination process* [165]. This means (among others) that the timing of a jump and the choice of the adjacent node are independent random processes. Therefore, if we look at the trajectory of the CTRW on the network and ignore the temporal component, we obtain the same statistics as with the DTRW. In particular, the equilibrium distribution in the limit $t \to \infty$ is identical to Eq. (2.39). We confirm our intuition by substituting the equilibrium condition $\mathrm{d}\boldsymbol{p}(t)/\mathrm{d}t = 0$ into Eq. (2.46a):

$$0 = \boldsymbol{p}^*(-\mathbb{1} + \boldsymbol{T}). \tag{2.49}$$

From Eq. (2.49), we can directly read the stationary condition $\boldsymbol{p}^* = \boldsymbol{p}^*\boldsymbol{T}$ and thus we arrive at equilibrium distribution in Eq. (2.39).

The CTRW model is commonly applied to model the global mobility of human travel. In the following section we will connect the diffusive exchange of individuals between urban areas with the SIR model from Sec. 2.1.

2.3.2 Meta-population model

We are now prepared to study the first of two models that integrates networks and disease spreading, namely the meta-population model. Here, each node is a population of individuals that are separated into epidemiological compartments. Assuming homogeneous mixing inside every node, we can use a mass-action approximation to describe the local disease dynamics. Additionally, individuals move diffusively between adjacent populations and thus contribute to the global spread of the infection. Hence, the meta-population model allows to investigate the complex interaction between infection dynamics and diffusive transport.

As a particular example, we focus on the *Rvachev-Longini model* [113], in which each node represents a densely populated area around an airport and air traffic connections indicate the flow of individuals in a global air traffic network. As for the local disease dynamics, we choose the SIR model as introduced in Sec. 2.1.

In this section, we start from a stochastic description of the model. Assuming given passenger numbers as well as population sizes, we derive the dynamic mean-field equations for the evolution of a global SIR-type epidemic. We then rewrite the model such that it does no longer depend explicitly on the population sizes. Finally, we motivate a heuristic modification that accounts for stochastic effects at the early stage of a local outbreak.

2.3.2.1 Stochastic model

Assuming the SIR model, we separate individuals within a population according to their health status into susceptible (S), infected (I) and recovered (R) individuals.

The latter category is characterized by a lifelong immunity as discussed in Sec. 2.1. Mass-action type of equations determine the flow of individuals between the different compartments within a local population as illustrated schematically in Fig. 2.1. In addition to the classical compartment model, however, individuals move between adjacent populations at a given rate. This flow of individuals is determined by the air traffic network $G = (\mathcal{N}, \mathcal{E})$ and represented conveniently by the weighted adjacency matrix $\boldsymbol{F}$. The daily passenger flux from node i to j gives the corresponding elements $F_{i,j}$. Furthermore, we assume a reciprocal flux $F_{i,j} = F_{j,i}$ for all edges $(i, j) \in \mathcal{E}$ and hence an undirected graph. The following reaction kinetics summarize the transitions between different epidemiological compartments:

$$I_i + S_i \xrightarrow{\beta} 2I_i \tag{2.50a}$$

$$I_i \xrightarrow{\mu} R_i \tag{2.50b}$$

$$X_i \xrightarrow{F_{i,j}/M_i} X_j. \tag{2.50c}$$

Here, S_i, I_i and R_i, respectively, refer to the epidemiological state of an individual in population i with $i \in \mathcal{N}$. The first transition (see Eq. (2.50a)) describes the infection of a susceptible individual upon contact with a carrier, where β is the mean per-capita infection rate (in units of d^{-1}). Once infected, the transition to the recovered state occurs according to Eq. (2.50b) with an average rate μ (in units of d^{-1}). Finally, an individual of the compartment $X \in \{S, I, R\}$ can move from i to the adjacent population j at a rate $F_{i,j}/M_i$ (see Eq. (2.50c)), where M_i denotes the total number of individuals in population i.

In Sec. 2.1, we introduced the basic reproduction number $R_0 = \beta/\mu$ as a central parameter of mathematical epidemiology. It indicates the expected number of secondary infections caused by a typical individual in a completely susceptible population. In this thesis, we only consider super-critical dynamics defined by $\beta > \mu$ or $R_0 > 1$, respectively. Hence a small number of initially infected nodes within a single population is likely to spread the disease, first locally within a node, and then cause a global outbreak across the network of populations.

2.3.2.2 Mean-field model

If the local prevalence level, i.e., the number of infected individuals in a population i, is sufficiently high, we can neglect fluctuations and determine the mean number of susceptible, infected and recovered individuals, denoted by M_i^S, M_i^I and M_i^R, respectively, using a deterministic mean-field approximation. To this end, we

extend the MA-model in Eqs. (2.2) as follows:

$$\frac{\mathrm{d}M_i^S}{\mathrm{d}t} = -\beta\rho_i^I M_i^S + \Omega_i^S \tag{2.51a}$$

$$\frac{\mathrm{d}M_i^I}{\mathrm{d}t} = \beta\rho_i^I M_i^S - \mu M_i^I + \Omega_i^I \tag{2.51b}$$

$$\frac{\mathrm{d}M_i^R}{\mathrm{d}t} = \mu M_i^I + \Omega_i^R. \tag{2.51c}$$

Here, $\rho_i^I = M_i^I/M_i$ denotes the fraction of infected in population i. The set of coupled ODEs in Eqs. (2.51) adds a mobility term Ω_i^X with $X \in \{S, I, R\}$ to the classical mean-field model as introduced in Eqs. (2.2). The additional contribution to M_i^X, i.e., the number of individuals in compartment X, results from the migration of individuals across populations. Assuming diffusive transport, Ω takes the form of a balance equation similar to the right side of the master equation in Eq. (2.45):

$$\Omega_i^X = \sum_{j\neq i}\left(\frac{F_{j,i}}{M_j}M_j^X - \frac{F_{i,j}}{M_i}M_i^X\right) \tag{2.52}$$

From the reciprocal passenger fluxes $F_{i,j} = F_{j,i}$ for all edges $(i,j) \in \mathcal{E}$, it follows immediately that the total flow into a population equals the out-directed flow and hence the total population size M_i remains constant. The total flow defines the strength $F_i = \sum_{j=1}^N F_{i,j}$ of population i as introduced earlier in Eq. (2.21). Moreover, the fraction of individuals leaving the population determines a population dependent exit rate

$$\gamma_i = \frac{F_i}{M_i}. \tag{2.53}$$

In other words, a randomly chosen individual leaves the population i at a constant node-dependent rate γ_i. With this, we can express the transition rate $F_{i,j}/M_i$ across the edge $i \to j$ as follows:

$$\frac{F_{i,j}}{M_i} = \gamma_i\frac{F_{i,j}}{F_i} \tag{2.54a}$$

$$= \gamma_i T_{i,j} \tag{2.54b}$$

Here, we introduced the conditional probability $T_{i,j}$ that an individual, leaving the population i, chooses the adjacent node j. It is identical to the transition probability of a random-walker on an undirected graph as defined in Eq. (2.32).

In analogy to the previous calculations leading to Eq. (2.54b) and using the symmetry $F_{i,j} = F_{j,i}$, we obtain the reciprocal rate

$$\frac{F_{j,i}}{M_j} = \gamma_i\frac{F_{j,i}}{F_i} \tag{2.55a}$$

$$= \gamma_i\frac{F_{i,j}}{F_i} \tag{2.55b}$$

$$= \gamma_i T_{i,j}\frac{M_i}{M_j} \tag{2.55c}$$

Substituting Eqs. (2.54b) and (2.55c) into the balance relation Eq. (2.52), we obtain

$$\Omega_i^X = \gamma_i \sum_{j \neq i} \left(\frac{T_{i,j} M_i}{M_j} M_j^X - T_{i,j} M_i^X \right) \tag{2.56a}$$

$$= \gamma_i M_i \sum_{j \neq i} T_{i,j} \left(\frac{M_j^X}{M_j} - \frac{M_i^X}{M_i} \right). \tag{2.56b}$$

In order to write down a mean-field model that is not explicitly dependent on population sizes, we normalize Eq. (2.51) as well as the mobility term in Eq. (2.56b) with M_i and thus arrive at the following mean-field model:

$$\frac{\mathrm{d}\rho_i^S}{\mathrm{d}t} = -\beta \rho_i^I \rho_i^S + \gamma_i \sum_{j \neq i} T_{i,j} (\rho_j^S - \rho_i^S) \tag{2.57a}$$

$$\frac{\mathrm{d}\rho_i^I}{\mathrm{d}t} = \beta \rho_i^I \rho_i^S - \mu \rho_i^I + \gamma_i \sum_{j \neq i} T_{i,j} (\rho_j^I - \rho_i^I) \tag{2.57b}$$

$$\frac{\mathrm{d}\rho_i^I}{\mathrm{d}t} = \mu \rho_i^I + \gamma_i \sum_{j \neq i} T_{i,j} (\rho_j^R - \rho_i^R) \tag{2.57c}$$

The fraction of individuals $\rho_i^X = M_i^X / M_i$ can be interpreted as a probability that a randomly selected individual from population i belongs to the compartment $X \in \{S, I, R\}$.

2.3.2.3 Invasion threshold

As mentioned earlier, we can justify the deterministic mean-field model in Eqs. (2.57) as soon as the disease prevalence in a population is sufficiently high. In this case, we can neglect fluctuations around the expected value. However, stochastic effects dominate if the number of infected is small, i.e., in particular if we determine the arrival time of the first infected individual in Chap. 3. In this case, the limits of the mean-field approximation become particularly apparent: According to Eqs. (2.57), a small fraction of infected individuals spreads immediately across the entire network and thus initiates the local dynamics in every population. One possibility to avoid this artifact is to return to the stochastic description of the system. However, these *agent-based models* are numerically complex, and therefore, simple heuristics have been introduced instead [5, 113, 114]. Here, we follow the approach of Brockmann & Helbing [5] and add a Heaviside function $\Theta(x)$ to the force of infection, thus acting as an invasion threshold. Hence, we modify Eqs. (2.57) as follows:

$$\beta \rho_i^I \to \beta \Theta(\rho_i^I / \xi_i) \rho_i^I. \tag{2.58}$$

Due to the local invasion threshold, a node remains susceptible unless the local prevalence ρ_i^I (as a fraction of the population size) exceeds a threshold value ξ_i. Formally, we obtain $\Theta(\rho_i^I / \xi_i) = 1$ for $\rho_i^I > \xi_i$ and $\Theta(\rho_i^I / \xi_i) = 0$, otherwise. The authors in [5] chose a uniform parameter $\xi_i \equiv 10^{-6}$ and argued that results are

robust for a wide range of values. In this work, however, we set $\xi_i = 1/M_i$ in order to avoid a somewhat arbitrary choice. Instead, we require that at least one infected individual is present in the population, i.e., a fraction $1/M_i$ on average, in order to initiate the outbreak. The modified and final system of equations reads:

$$\frac{\mathrm{d}\rho_i^S}{\mathrm{d}t} = -\beta\Theta(\rho_i^I M_i)\rho_i^I\rho_i^S + \gamma_i\sum_{j\neq i} T_{i,j}(\rho_j^S - \rho_i^S) \tag{2.59a}$$

$$\frac{\mathrm{d}\rho_i^I}{\mathrm{d}t} = \beta\Theta(\rho_i^I M_i)\rho_i^I\rho_i^S - \mu\rho_i^I + \gamma_i\sum_{j\neq i} T_{i,j}(\rho_j^I - \rho_i^I) \tag{2.59b}$$

$$\frac{\mathrm{d}\rho_i^R}{\mathrm{d}t} = \mu\rho_i^I + \gamma_i\sum_{j\neq i} T_{i,j}(\rho_j^R - \rho_i^R). \tag{2.59c}$$

We summarize our effort to derive a first, network-based epidemiological model as follows: Starting from the stochastic description of a meta-population SIR model in Eqs. (2.50), we finally proposed a set of mean-field equations in Eqs. (2.59). This reaction-diffusion type of model assumes homogeneous mixing inside each population and diffusive coupling between nodes. In our particular model of global epidemic spreading, we further assume that passenger fluxes and population sizes are provided, which translate into relative fluxes and node-dependent exit rates. Moreover, our mean-field model is deterministic and thus neglects stochastic effects. However, in order to account for variations at the early stage of an outbreak, we introduced a heuristically motivated invasion threshold. The meta-population model in Eqs. (2.59) will serve as a numerical reference model for our analytical approach in Chap. 3.

Unlike the reaction-diffusion type of model discussed here, we will look at an example from a *contact process* in the next section. There, each node is a single individual that can contract the infection from an adjacent node.

2.3.3 Individual-based model

Unlike the dynamics of global diseases, many contagion processes require to model nodes in a network as individual units. These include the spread of sexually transmitted diseases through a network of romantic encounters [151], computer viruses in a technological web [169], and information diffusion such as tweets in a social network [130, 170]. A highly versatile approach to study these phenomena is the individual-based (IB) model. In this section, we will first introduce the IB model on networks with time-varying topologies and then study the early phase of an outbreak. In particular, we estimate the epidemic threshold from the largest eigenvalue of the Jacobian matrix, which captures both, the temporal and topological information of the graph.

2.3.3.1 Derivation of the dynamical model

We consider a temporal network $\mathcal{G}$ with N nodes, represented by a sequence of unweighted and symmetric adjacency matrices $[\boldsymbol{A}(0), \boldsymbol{A}(1), ..., \boldsymbol{A}(T-1)]$. A node

$i \in \mathcal{N}$ constitutes an individual characterized by an epidemiological state $X_i(t)$. The latter is a random variable, where $X_i(t) = X$ means that node i is in state $X \in S, I, R$ at time t. Moreover, we denote with $\Pr[X_i(t) = X] = P_i^X(t)$ the corresponding probability to find node i is in state X at time t.

Essentially, the stochastic process is a Markov chain with 3^N possible states [14, 15]. Once, the high dimensional transition operator that maps the state at time t to $t+1$ has been defined, one can determine the joint probability, i.e., the simultaneous epidemic state of all nodes in the network:

$$\Pr[X_1(t) = X_1, X_2(t) = X_2, ..., X_N(t) = X_N]. \tag{2.60}$$

In principle, the Markov-chain approach gives an accurate model on the level of individual nodes. The number of equations, however, grows exponentially with the network size N, and in order to reduce complexity for large graphs, different approximations exist. The simplest way is to consider the one-point marginals, i.e., the probability $P_i^X(t)$ to find node i in state $X \in \{S, I, R\}$ directly as variables in a set of $3 \times N$ coupled dynamic equations [15, 33, 86–91]. We will refer to this approach as the IB model, though it is sometimes also called N-intertwined [15] or quenched mean-field model[22, 171].

Other techniques, such as the moment-closure approximation [133, 137, 172, 173], consider also two-point marginals, i.e., the joint probability

$$\Pr[X_i(t) = X_i, X_j(t) = X_j] \text{ with } i, j \in \mathcal{E} \tag{2.61}$$

as variables in an extended dynamical system. The additional information allows to account for correlations between the epidemic states of adjacent nodes. Here, recent research focuses on different closure techniques to approximate higher-order marginals and thereby include local network properties such as the clustering coefficient [93, 133]. In this thesis, however, we will focus on the basic IB model to prepare the ground for the *contact-based model* in Chap. 5.

Unlike the previous sections, we will consider a discrete-time version of the SIR model. For convenience, we reuse the notation and denote with β and μ the *per time step probabilities* to transmit a disease upon contact and to recover, respectively. In the following, we will go through the details of the model and derive the dynamic equations. To this end, we begin with the infection process.

First, we assume that only direct neighbors can contract the disease within one timestep, i.e., we do not account for secondary infections. Consequently, node i contracts the infection from j only if (i) both are in contact at the time as indicated by the adjacency matrix element $A_{j,i}(t)$, (ii) node j is infected with probability $P_j^I(t)$ and, (iii) j passes the pathogen to i with probability β. Hence, the overall probability to contract the virus from j is $\beta A_{j,i} P_j^I$. Then, the product $\prod_{j=1}^{N}[1 - \beta A_{j,i}(t) P_j^I(t)]$ gives the probability to receive no infection in the interval $[t, t+1)$ from any potential neighbor. With this, we determine the marginal probability $P_i^S(t+1)$ for node i to be susceptible at time $t+1$: It is the joint

probability to be susceptible in the previous time step t and not contract the infection in $[t, t+1)$. In the IB model, the joint probability factorizes and we obtain

$$P_i^S(t+1) = P_i^S(t) \prod_{j \in \mathcal{N}} [1 - \beta A_{j,i}(t) P_j^I(t)]. \tag{2.62}$$

Again, the crucial simplification in Eq. (2.62) is to treat the epidemic states of i and its neighbors as mutually independent such that joint probabiliies factorize:

$$\Pr[X_i(t) = X_i, X_j(t) = X_j] = P_i^X(t) P_j^X(t) \quad \forall i, j \in \mathcal{E}. \tag{2.63}$$

This assumption is similar to classical mean-field approaches in statistical physics. Note, however, that the full topological information is still present through the elements $A_{j,i}$ of the adjacency matrix. The literature sometimes refers to this approximation as neglecting dynamic correlations [174].

We further assume in this thesis that infection transmission and recovery are independent processes. Hence, the marginal probability P_i^I evolves according to two contributions: (i) Infected nodes recovers with a constant probability μ, and thus P_i^I decreases by μP_i^I at every time step. (ii) The probability P_i^I increases by $\Delta P_i^S(t) = P_i^S(t) - P_i^S(t+1)$ due to the probability flow from the susceptible to the infected compartment. Combining both contributions leads to

$$P_i^I(t+1) = (1-\mu) P_i^I(t) + \Delta P_i^S(t). \tag{2.64}$$

From Eq. (2.62) we obtian the probability $\Delta P_i^S(t)$ that node i is newly infected:

$$\Delta P_i^S(t) = P_i^S(t) \left\{ 1 - \prod_{j=1}^{N} [1 - \beta A_{j,i}(t) P_j^I(t)] \right\}. \tag{2.65}$$

The remaining marginal probability $P_i^R(t)$ to find node i in the recovered state follows from the conservation condition $P_i^S(t) + P_i^I(t) + P_i^R(t) = 1$ for all $i \in \mathcal{N}$. For convenience, we restate the system of coupled dynamic equations that constitutes the IB model for temporal networks:

$$P_i^S(t+1) = P_i^S(t) \prod_{j \in \mathcal{N}} [1 - \beta A_{j,i}(t) P_j^I(t)] \tag{2.66a}$$

$$P_i^I(t+1) = (1-\mu) P_i^I(t) + P_i^S(t) \left\{ 1 - \prod_{j=1}^{N} [1 - \beta A_{j,i}(t) P_j^I(t)] \right\} \tag{2.66b}$$

$$P_i^R(t+1) = 1 - P_i^S(t) + P_i^R(t). \tag{2.66c}$$

The authors in [175] studied the percolation-type of transition from local to global outbreaks using the IB SIR model in Eqs. (2.66). The result is identical to earlier works for the recurrent SIS model [33, 176–178] and we will summarize their main steps in the following section.

2.3.3.2 Derivation of the epidemic threshold

We study the early phase of the SIR dynamics akin to the linear stability analysis of the classical compartment model in Sec. 2.1, which eventually led to the condition $\beta/\mu > 1$ for global global epidemic outbreaks. Here, we note that

$$(P_i^S, P_i^I, P_i^R) = (1., 0, 0) \qquad \forall i \in \mathcal{N} \tag{2.67}$$

is a stable fixed point of Eqs. (2.66). Next, we evaluate the stability of a fully susceptible population after a pertubation with a small fraction of infected individuals at $t = 0$:

$$P_i^I(0) = 1 - P_i^S(0) = \epsilon_i(0), \text{ with } \epsilon_i(0) \ll 1 \qquad \forall i \in \mathcal{N}. \tag{2.68}$$

To this end, we expand Eq. (2.66b) to the leading order in ϵ_i:

$$\epsilon_i(t+1) = (1-\mu)\epsilon_i(t) + [1-\epsilon_i(t)] \left\{ 1 - \prod_{j=1}^{N} [1 - \beta A_{j,i}(t)\epsilon_j(t)] \right\} \tag{2.69a}$$

$$= (1-\mu)\epsilon_i(t) + [1-\epsilon_i(t)] \sum_{j=1}^{N} \beta A_{j,i}(t)\epsilon_j(t) \tag{2.69b}$$

$$= (1-\mu)\epsilon_i(t) + \beta \sum_{j=1}^{N} A_{j,i}(t)\epsilon_j(t). \tag{2.69c}$$

We observe that the original dynamics in Eqs. (2.66) decouples and reduces to a set of N linear equations in Eq. (2.69c). Next, we turn to a matrix-based representation by introducing the N dimensional row vector $\boldsymbol{\epsilon}(t)$ with elements ϵ_i and the adjacency matrix $\boldsymbol{A}(t)$ for $t = 0, 1, .., T-1$. With this, the sum $\sum_{j=1}^{N} A_{j,i}(t)\epsilon_j(t)$ in Eq. (2.69c) turns into the product $\boldsymbol{\epsilon A}$ and with this, we get

$$\boldsymbol{\epsilon}(t+1) = (1-\mu)\boldsymbol{\epsilon}(t) + \beta\boldsymbol{\epsilon A}(t) \tag{2.70a}$$

$$= \boldsymbol{\epsilon}\left[(1-\mu)\mathbb{1} + \beta\boldsymbol{A}(t)\right]. \tag{2.70b}$$

Here, we denote with $\mathbb{1}$ the $N \times N$ dimensional identity matrix. Applying Eq. (2.70b) repeatedly on the initial pertubation $\boldsymbol{\epsilon}(0)$, we obtain the explicit solution at the final observation time T:

$$\boldsymbol{\epsilon}(T) = \boldsymbol{\epsilon}(0)\boldsymbol{M}. \tag{2.71}$$

Here, $\boldsymbol{M}$ is the *system matrix* or *infection propagator*, respectively:

$$\boldsymbol{M} = [(1-\mu)\mathbb{1} + \beta\boldsymbol{A}(0)] \cdot [(1-\mu)\mathbb{1} + \beta\boldsymbol{A}(1)] \cdot ... \cdot [(1-\mu)\mathbb{1} + \beta\boldsymbol{A}(T-1)] \tag{2.72a}$$

$$= \prod_{t=0}^{T-1} \left[(1-\mu)\mathbb{1} + \beta\boldsymbol{A}(t)\right]. \tag{2.72b}$$

In order to evaluate the asymptotic stability of the disease-free fixed point, we assume a periodic boundary condition in time, i.e., $\boldsymbol{A}(t) = \boldsymbol{A}(t+T)$. With this, the dominant eigenvalue $\rho(\boldsymbol{M})$ or equally, the spectral radius of the system matrix $\boldsymbol{M}$, determines the long term behavior: The system is asymptotically stable if

$$\rho\left(\prod_{t=0}^{T-1}\left[(1-\mu)\,\mathbb{1} + \beta\boldsymbol{A}(t)\right]\right) < 1. \tag{2.73}$$

In other words, for any set of parameters (β, μ) that satisfies Eq. (2.73), we expect that the outbreak size remains small with respect to the total number of nodes.

We add a short note on the properties of the largest eigenvalue. Assuming that the aggregate network, as represented by the sum of adjacency matrices $\sum_{t=0}^{T-1}\boldsymbol{A}(t)$, is connected, an arbitrary node can eventually reach any other in the network (potentially after going through multiple cycles of the temporal network). Hence, the reachability graph (see Eq. (2.27)) and equally, the graph represented by the system matrix in Eq. (2.72) are strongly connected. Then, the Perron-Frobenius theorem guarantees a positive (real-valued) and simple dominant eigenvalue.

We close the section with a short discussion on two limiting cases. First, we study a disease that evolves on a much faster time-scale than the temporal network. In this *quenched limit*, the disease evolves effectively on a static graph represented by the first adjacency matrix of the temporal sequence, i.e. $A \equiv A(0)$. Again, we restrict the analysis to the connected component such that the Perron-Frobenius theorem applies. The corresponding infection propagator reads

$$\boldsymbol{M}_{\text{fast}} = \prod_{t=0}^{T-1}\left[(1-\mu)\mathbb{1} + \beta\boldsymbol{A}\right] \tag{2.74a}$$

$$= \left[(1-\mu)\mathbb{1} + \beta\boldsymbol{A}\right]^T. \tag{2.74b}$$

In Eq. (2.74b), we find a product of T identical matrices and from basic linear algebra, we find that the corresponding spectral radius simplifies to

$$\rho(\boldsymbol{M}_{\text{fast}}) = \rho\left(\left[(1-\mu)\mathbb{1} + \beta\boldsymbol{A}\right]^T\right) \tag{2.75a}$$

$$= \rho\left[(1-\mu)\mathbb{1} + \beta\boldsymbol{A}\right]^T \tag{2.75b}$$

$$= \left[(1-\mu) + \beta\rho(\boldsymbol{A})\right]^T. \tag{2.75c}$$

The condition $\rho(\boldsymbol{M}_{\text{fast}}) = 1$ hence determines the critical parameters

$$\left(\frac{\beta}{\mu}\right)_{\text{crit, fast}} = \frac{1}{\rho(\boldsymbol{A})}. \tag{2.76}$$

The static limit in Eq. (2.76) is well-known in the literature [86, 88, 179] and admits a direct comparison to our previous analysis of the classical compartmental model: In Sec. 2.1, we derived the condition $(\beta/\mu)_{\text{crit, MA}} = 1$ for the epidemic

threshold of the MA model. Because of $\rho(A) > 1$, we conclude that

$$\left(\frac{\beta}{\mu}\right)_{\text{crit, fast}} < \left(\frac{\beta}{\mu}\right)_{\text{crit, MA}}.$$

In other words, due to the homogeneous mixing assumption, the MA model overestimates the critical parameter $(\beta/\mu)_{\text{crit}}$.

Similarly, we can study the limit of a slowly evolving diseases, also known as the annealed limit in statistical physics. In this case, we assume that the epidemic parameters β and μ are sufficiently small, such that at most one infection or recovery event takes place during one time-step. We then expand the propagator in Eq. (2.72) to the first order in β and μ:

$$\boldsymbol{M}_{\text{slow}} = (1 - T\mu)\mathbb{1} + \beta \sum_{t=0}^{T-1} \boldsymbol{A}(t). \tag{2.77}$$

Denoting with $\bar{\boldsymbol{A}} = 1/T \sum_{t=0}^{T-1} \boldsymbol{A}(t)$ the time-averaged adjacency matrix, we obtain the following criticality condition

$$\left(\frac{\beta}{\mu}\right)_{\text{crit, slow}} = \frac{1}{\rho(\bar{\boldsymbol{A}})}. \tag{2.78}$$

With this final result, we conclude our analysis of the IB model. After introducing the details of the model, we derived the set of dynamic equations in Eqs. (2.66). Then, we studied the stability of the disease-free fixed point corresponding to a fully susceptible population. This led to the conclusion that a small outbreak remains confined if the spectral radius of the infection propagator matrix is below one (see Eq. (2.73)). The spectral condition captures both, the topological and temporal information of the underlying network. However, it remains yet unclear to what extend the limitations of the IB model impact the epidemic threshold. Therefore, we will compare in Chap. 5 of this thesis Monte-Carlo simulations with analytical estimations and propose a contact-based approach to overcome the inherent limitations of the IB model.

2.4 Summary

In this chapter, our aim was to provide the reader the basic analytical tools and concepts relevant to the main chapters of the thesis. To this end, we first introduced the generic SIR model in Sec. 2.1 assuming homogeneous-mixing. With this generic model we studied typical features of a non-recurrent epidemic outbreak. In particular, the analysis of the early and late-time behavior, respectively, motivated the concept of the epidemic threshold, i.e. a critical set of parameters that mark the transition from local to global epidemic outbreaks. Then, in Sec. 2.2, we turned our attention to modern network theory. Here, we introduced selected methods to study empirical mobility and contact data. The example of the German pig-trade network revealed structural and temporal features that are frequently observed

across a wide range of networks such as broad distributions for the number of contacts and heterogeneous edge activation times. The particularly important concept of accessibility as introduced in Eq. (2.28) will serve as a basis for a novel matrix-based epidemiological model in Chap. 4.

Overall, our analysis of the pig-trade network suggests that the homogeneous-mixing assumption of the MA model deviates strongly from empirical observations. Hence, it is essential to integrate empirical contact and mobility data into epidemiological models. To this end, we first studied in Sec. 2.3 the random-walk process as a generic model for diffusion on graphs. Integrating local SIR dynamics with diffusive transport, we then introduced the first of two data-driven epidemic models in Sec. 2.1, namely the meta-population model. We identified nodes and edges of a mobility graph with well-mixed populations and airtraffic connections, respectively. This reaction-diffusion type of model allows to study global epidemic outbreaks and will thus serve as a reference model in Chap. 3. The second class of epidemic models identifies nodes with single epidemiological units that can contract the infection through time-varying contacts with their neighbors. In that context, we studied the IB model as a commonly used example and derived the epidemic threshold from a linear stability analysis. The IB model suffers, however, from limitations that arise from the ad hoc factorization of joint probabilities. In order to improve the IB framework, we introduce in Chap. 5 the contact-based model and derive a structurally similar threshold criterium.

3

Effective Distances for Global Epidemic Outbreaks

Abstract. We study the expected arrival time of the first infected individual in a global epidemic outbreak. To this end, we analyse a meta-population model, where airports and flight connections form nodes and edges of the global air-traffic network, respectively. Building on the work of Gautreau et al. (2008) and Brockmann & Helbing (2013), we propose the random-walk effective distance (RWED), which improves previous results by taking into account all potential transmission routes. From the theory of absorbing random-walks we derive a computationally efficient approach to calculate effective distances between all pairs of nodes simultaneously from a single matrix inversion operation. Finally, we compare the predictive power with numerical simulations on several synthetic and empirical datasets. Our results suggest that the RWED can be used as a reliable and computationally efficient approach to estimate the expected infection arrival time.

3.1 Introduction

In recent years, the public debate repeatedly focused on the risk of global epidemic outbreaks. The spread of SARS (2003) [180], H1N1 (2009) [181], MERS-CoV (2013) [182], Ebola (2014) [183] and most recently, ZIKV (2016) [184] affected millions of people across the world and revealed the vulnerability of modern society. Among the factors contributing to the rapid spread of the disease are increasing urbanisation as well as growing mobility between densely populated areas [3, 185, 186].

With decreasing prices for airline tickets, every major city becomes accessible within a couple of days (see Fig.3.1). Also, a seemingly endless number of details

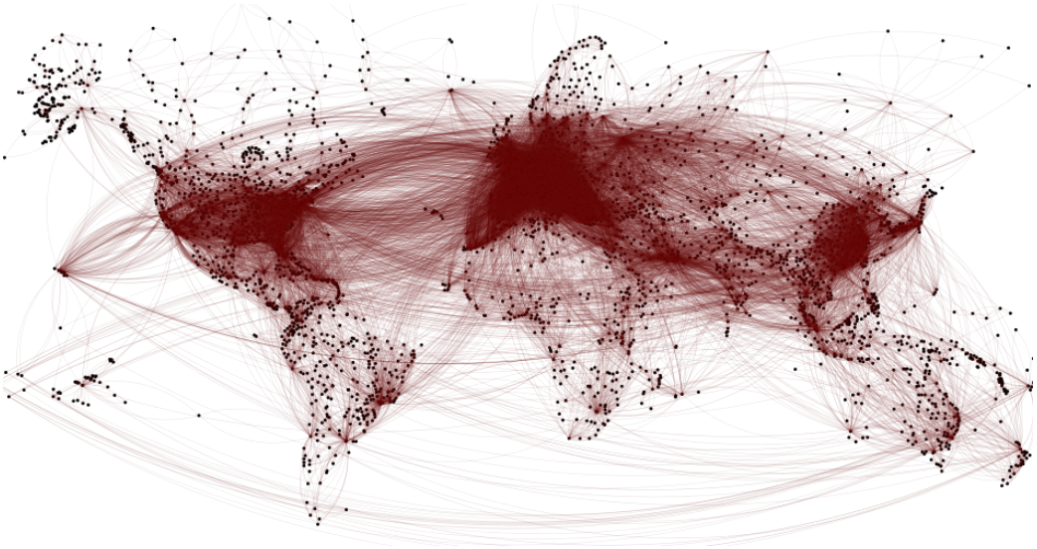

Figure 3.1: Geographically embedded global air-traffic network. A total of 4069 airports and 26,691 undirected flight connections are represented as nodes and edges, respectively, in a geographically embedded network. For detailed information see Tab. 3.1.

and unforeseeable events appears to affect the spreading dynamics. Hence it seems unlikely to predict and contain the future course of an outbreak based on simple mathematical models.

Instead, research on public health relies often on complex and computationally demanding Monte-Carlo simulations. These can account for estimated population sizes and mobility data down to the level of inner-city traffic, as well as complex disease models [187]. Good predictions in the past have advocated this approach, but similar to a weather forecast, the underlying mechanism remains often unclear [24, 180, 188, 189].

According to [190], the success of different numerical approaches indicates, however, that possibly few factors are decisive. One explanation is that stochastic effects average out with a sufficiently high prevalence and only a few parameters dominate the dynamics on a global scale. Therefore, research has focused recently on the importance of global air-traffic as the dominant transmission route [3, 191] (see Fig. 3.1).

Based on the topology of flight connections, simple graph-theoretic approaches have been proposed to approximate the complex nature of global epidemic spreading [114]. In particular, the so-called *effective distance* [5] has received significant attention as a conceptually simple and yet reliable proxy for infection arrival times.

Information on when the first infected person arrives at a remote location is central to public health institutions. It allows to (i) anticipate likely paths for epidemic spreading, (ii) evaluate counter measures and, (iii) determine retrospectively the outbreak origin.

The first theoretical work towards effective distances [114] begins with a probabilistic analysis of the meta-population model as introduced in Sec. 2.3.2. The authors in [114] highlight the essential role of competing paths for epidemic

spreading. However, in empirical graphs such as the global air-traffic network, the large number of parallel routes by which a disease can potentially reach a distant population poses a computational limit. Therefore, the authors eventually simplified their approach and estimated infection arrival times based on a single dominant transmission path.

Independently, the authors in [5] introduced a non linear transformation of edge weights from heuristic arguments. The central idea is to rescale passenger fluxes such that a high travel probability translates into a short distance. Then, the weighted shortest path from the outbreak location to a distant node, termed *effective distance*, can be used as a proxy to estimate the infection arrival time. We will show later how this result can be derived from a special case of [114]. As a relevant contribution to the field, the authors in [5] compare analytic estimations with outbreak data from the SARS (2003) and H1N1 (2009) epidemic. They demonstrate that effective distances can predict reliably complex spreading dynamics using only the network topology. Moreover, the authors propose an intuitive approach to solve the inverse problem, i.e. to determine the outbreak location from a given snapshot.

Until recently, the analysis has been restricted to the dominant transmission path only [5, 114]. In order to improve the predictive power and generalize the scope of effective distances, we will present in this chapter an approach that accounts efficiently for all possible transmission paths. The key idea is to make a Markovian approximation and thus estimate the infection arrival time with the first-passage probability of a random walk. Extensive numerical simulations on the global air-traffic network as well as additional empirical and synthetic graphs demonstrate that our *random-walk effective distance* (RWED) outperforms previous approaches [5, 114]. Moreover, the accuracy of the RWED hints at a potential correspondence between complex contagion dynamics and a simple diffusive process. This surprising result opens up a promising research direction, where well-established tools from random-walk theory could enter theoretical epidemiology.

We structure this chapter as follows: After this introduction, we present the key characteristics of the global air-traffic network in Sec. 3.2. Then, we extent the previously introduced meta-population model in Sec. 3.3, which is the theoretical framework for both, the numerical simulations and the analytical estimations. In Sec. 3.4 we introduce the *multiple path effective distance* (MPED) and *dominant path effective distance* (DPED) based on the works of [5, 114]. This chapter examines in depth the two main publications that precede our contribution. Thereby, we also discuss the necessary assumptions and evaluate their limits. Based on the previous analysis, we introduce the RWED as our main result in Sec. 3.5. Leveraging the theory of Markov chains, we derive a computationally efficient approach to calculate the RWED in Sec. 3.6 and compare the analytical approach in Sec. 3.7 with simulated infection arrival times on the global air-traffic network

as well as additional synthetic and empirical datasets. Finally, we summarize and discuss our results in Sec. 3.8.

This chapter and in particular Sec. 3.5 is based on our publication in [115]. The derivation and comparison with previous methods in Sec. 3.4, however, go beyond the published work. Also, we improved the computation of the RWED in Sec. 3.6 considerably from N matrix inversions, where N is the number of nodes, to only a single inversion operation in order to obtain the distances between all pairs of nodes.

3.2 The global air-traffic network

We base our work on the global air-traffic network, provided by the authors of [5]. The dataset contains 4,069 airports with geographic locations and the number of seats on commercial air services for 2007. We merge multiple flight connections between any two airports and sum up the number of seats. Then, we assume that these flights operate at full capacity for economic reasons, although the exact number of passengers is not necessary to determine the arrival time of an epidemic. We now interpret the global air-traffic as a directed graph with airports and flight connections as nodes and edges, respectively. Also, we consider as edge-weights F_{ij} the estimated passenger flux per day between any two airports i and j. These weights occur highly symmetric, i.e. $F_{ij} \approx F_{ji}$ and for convenience we will, therefore, consider an undirected graph as suggested in [5]. To this end, we replace edge-weights with the sum of its reciprocal parts:

$$F_{ij} \to F_{ij} + F_{ji}. \tag{3.1}$$

With this, we obtain a connected graph with 4,069 nodes and 26,691 undirected edges. The total passenger volume sums up to

$$\Phi = \sum_{i,j=1}^{N} F_{ij} = 8.91 \cdot 10^6 \text{ passengers / day}. \tag{3.2}$$

In Fig. 3.2 A.2, we give a give a circular-layout representation of the air-traffic network from Fig. 3.1. Here, we bundle edges on multiple levels using a community detection algorithm [66] and color nodes according to their geographic location (see Fig. 3.2 A.1). This representation reveals a hierarchical community structure, where airports (mostly from the same region) form well-connected sub-graphs on multiple scales. The observation reflects our intuition that local airports connect to local hubs, which in turn serve a regional hub and so on.

The average number of passengers per flight connection is just $\langle F \rangle \approx 167$ persons per day. However, a high coefficient of variation $c_v(F) = 2.45$ (standard deviation relative to the mean), indicates that passenger numbers are broadly distributed (see Fig. 3.2 B.3). Similarly, we also find broad distributions of degree and strength, i.e., the total passenger flow as shown in Fig. 3.2 B.1 and Fig. 3.2 B.2,

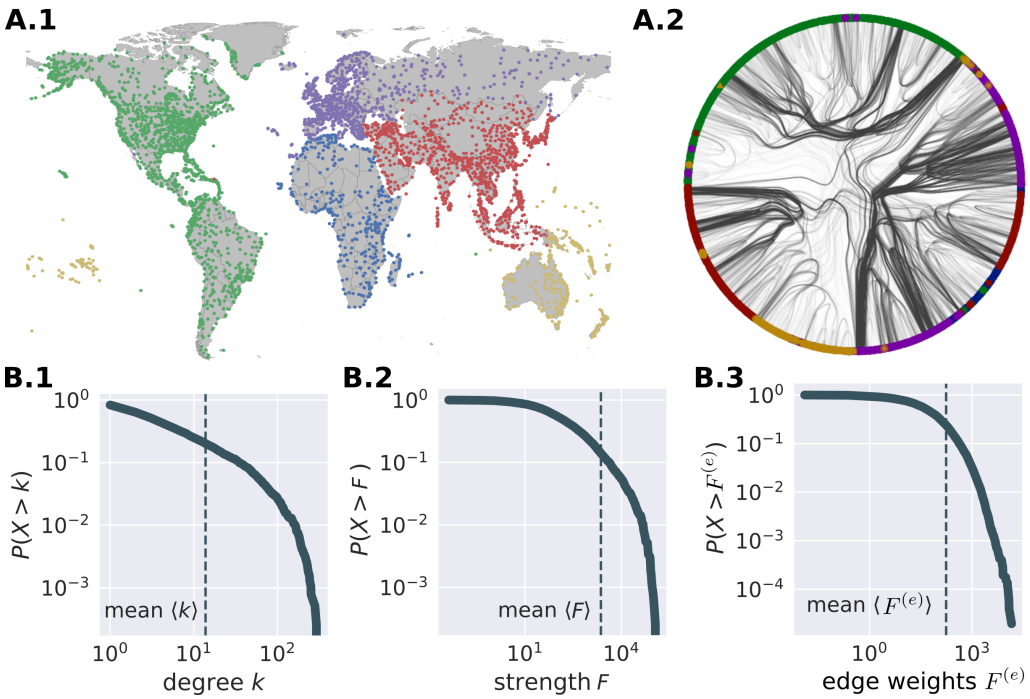

Figure 3.2: Properties of the symmetrized global air-traffic network. **A.1**: Geographic distribution of airports with node colors according to their geographic region. **A.2**: Circular hierarchy layout of the air-traffic network. Edges bundling reflects hierarchical communities found by the nested stochastic block model [66]. Nodes color as in A.1. **B.1** - **B.3**: Degree, strength and edge weight (passenger flux) distribution. The vertical line indicates the corresponding mean value.

respectively. We summarize further network characteristics in Tab. 3.1 and refer to [3, 192–194] for additional information.

Nodes N	4,069
Edges E	26,691
Link density $2E/N^2$	$3.6 \cdot 10^{-3}$
Total traffic flux Φ [d^{-1}]	$8.91 \cdot 10^6$
Mean flux per edge $\langle F^{(e)} \rangle$ [d^{-1}]	167
Rel. deviation $c_v(F^{(e)})$	2.45
mean degree $\langle k \rangle$	13.81
Rel. deviation $c_v(k)$	2.10
Mean strength $\langle F \rangle$ [d^{-1}]	$2.31 \cdot 10^3$
Rel. deviation $c_v(F)$	3.66

Table 3.1: Statistical properties of the global air-traffic network.

3.3 The meta-population model

Mean-field model. We base our work on a deterministic reaction-diffusion model as introduced in Sec. 2.3.2. Originally developed to study the Hong Kong flu epidemic, the *Rvachev-Longini model* [113] model allows to investigate the complex interaction between infection dynamics and mobility. For this purpose, we

associate nodes in the global air-traffic network $G = (\mathcal{N}, \mathcal{E})$ with densely populated areas that are considered *well-mixed*. In other words, these urban areas have no internal structure, and thus all individuals could potentially come into contact with each other. For the local disease dynamics, we focus on the compartmental SIR model, which divides the population into susceptible (S), infected (I) and recovered (R) individuals. Mass action type of equations determine the flow of individuals between the different epidemic compartments [8, 10] with β and μ denoting the per-capita infection rate and the recovery rate, respectively. Moreover, individuals leave their population with a node-dependent exit rate γ_i for all $i \in \mathcal{N}$. The mean-field model as introduced in Sec. 2.3.2 reads:

$$\frac{\mathrm{d}\rho_i^S}{\mathrm{d}t} = -\beta\Theta(\rho_i^I M_i)\rho_i^I\rho_i^S + \gamma_i \sum_{j\neq i} T_{i,j}(\rho_j^S - \rho_i^S) \tag{3.3a}$$

$$\frac{\mathrm{d}\rho_i^I}{\mathrm{d}t} = \beta\Theta(\rho_i^I M_i)\rho_i^I\rho_i^S - \mu\rho_i^I + \gamma_i \sum_{j\neq i} T_{i,j}(\rho_j^I - \rho_i^I) \tag{3.3b}$$

$$\frac{\mathrm{d}\rho_i^R}{\mathrm{d}t} = \mu\rho_i^I + \gamma_i \sum_{j\neq i} T_{i,j}(\rho_j^R - \rho_i^R). \tag{3.3c}$$

Here, we denote with ρ_i^X the fraction of individuals from population i that are in state $X \in \{S, I, R\}$. The population size is given by M_i and $T_{i,j}$ denotes the probability that an individual, leaving population i chooses the adjacent node j (see also Eq. (2.32)):

$$T_{i,j} = \frac{F_{i,j}}{\sum_{j=1}^N F_{i,j}}. \tag{3.4}$$

The authors in [114] estimated the population sizes M_i with data from census surveys. However, the *effective* population size belonging to the catchment area of an airport does not necessarily match census data as the authors in [3] argued. For this reason, we follow the approach of [3, 5] instead and estimate M_i from the passenger flux $F_i = \sum_{j=1}^N F_{i,j}$. To this end, we assume a global, i.e., node-independent exit rate $\gamma_i \equiv \gamma$ for all $i \in \mathcal{N}$ and solve Eq. (2.53) for M_i:

$$M_i = \frac{F_i}{\gamma}. \tag{3.5}$$

We determine the parameter γ from the fraction of the total population $\Gamma = \sum_{i=1}^N M_i$ that leaves the population:

$$\gamma = \Gamma/\Phi. \tag{3.6}$$

Assuming that the total population Γ participating in the air-traffic is about 2-6 billion people and with the passenger flux Φ from Tab. 3.1, we estimate the exit rate γ to lie within the following range (in units of days^{-1}):

$$1.48 \cdot 10^{-3} < \gamma < 4.45 \cdot 10^{-3} \tag{3.7}$$

This estimate agrees with a previous analyzis in [5]. Moreover, we note that our work aims to predict *relative* infection arrival times and not absolute values. For this reason, the exact value of γ is irrelevant in this context. We also point out that the mobility rate γ is two orders of magnitude smaller than the typical infection and recovery rate, β and μ, respectively. This observation will become important for the analytical approximation in Sec. 3.4. The modified and final system of equations reads:

$$\frac{\mathrm{d}\rho_i^S}{\mathrm{d}t} = -\beta\Theta(\rho_i^I F_i/\gamma)\rho_i^I\rho_i^S + \gamma\sum_{j\neq i} T_{i,j}(\rho_j^S - \rho_i^S) \tag{3.8a}$$

$$\frac{\mathrm{d}\rho_i^I}{\mathrm{d}t} = \beta\Theta(\rho_i^I F_i/\gamma)\rho_i^I\rho_i^S - \mu\rho_i^I + \gamma\sum_{j\neq i} T_{i,j}(\rho_j^I - \rho_i^I) \tag{3.8b}$$

$$\frac{\mathrm{d}\rho_i^R}{\mathrm{d}t} = \mu\rho_i^I + \gamma\sum_{j\neq i} T_{i,j}(\rho_j^R - \rho_i^R). \tag{3.8c}$$

In Fig. 3.3, we simulate an outbreak that starts near São Paulo Guarulhos International Airport (GRU). The subfigures in Fig. 3.3 A show how the infection propagates first locally in South America, then infected individuals arrive at global hubs and, finally the epidemic spreads to regional airports across the world. Moreover, we observe that the global prevalence peaks after around 200 days (see Fig. 3.3 B top). From the local prevalence curves in Fig. 3.3 B (bottom), we determine the infection arrival time at individual airports. To this end, we consider a node infected as soon as the prevalence exceeds one infected individual, i.e., $\rho_i^I > 1/$ for a given population i.

Finally, we evaluate in Fig. 3.3 C the geographic distance from the outbreak location as a proxy for infection arrival times. For many applications, including source reconstruction it is sufficient to determine *relative* infection arrival times instead of absolute values. Hence, the quality of our measure can be conveniently evaluated by looking at Pearson'scorrelation coefficient r, excluding the source node. From the best linear fit, as indicated by the black line in Fig. 3.3 C, we obtain a relatively low correlation of $r = 0.48$ between the geographic distance and simulated infection arrival times. This conclusion motivates our goal to find a network-based measure that improves predictions of infection arrival times.

3.4 A framework for effective distances

We discussed in Fig. 3.3 that geographic distance is of limited use as a proxy for infection arrival times. Indeed, we can achieve a considerably higher correlation by comparing arrival times with the topological distance as depicted in Fig. 3.4 A. This observation applies in particular to large airports. We define the topological distance as the number of steps along the shortest path from the outbreak location to the target node. However, like most empirical graphs the air-traffic network shows a small world property, i.e., the typical distance is much smaller than the number of nodes. Here, the maximum distance from São Paulo Guarulhos is nine

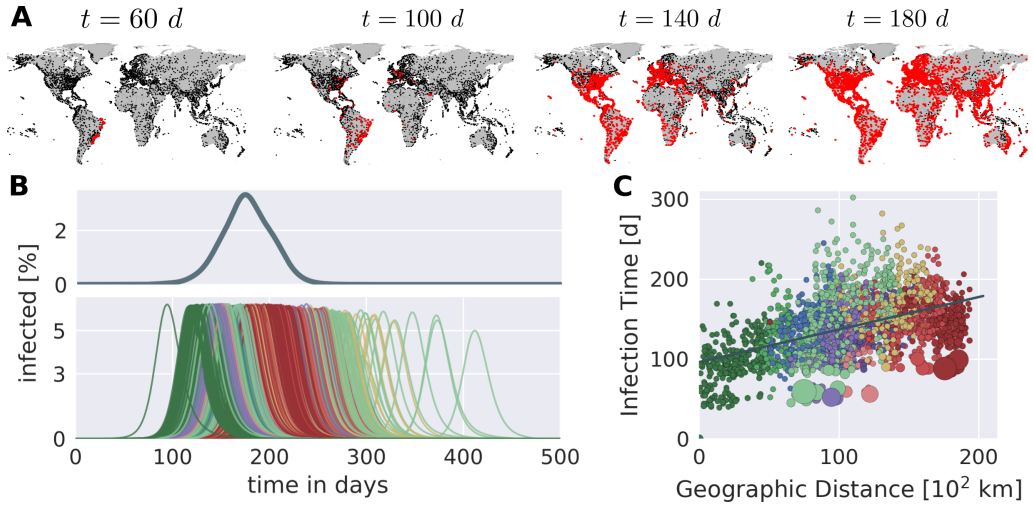

Figure 3.3: Outbreak simulation based on the global air-traffic network (see Fig. 3.1). Numerical integration of Eq. (3.8) with São Paulo-Guarulhos International Airport (GRU) in Brazil as outbreak location. The parameters are $\mu = 0.285$ d^{-1}, $R_0 = 1.5$ and $\gamma = 0.0028$ d^{-1}. **A**: Geographic distribution of infected nodes, i.e., populations with at least one infected individual for four different snapshots. **B** Top: Global prevalence as a fraction of the total population. Bottom: Local prevalence, i.e., fraction of infected for each population with colors according to region (see Fig. 3.2). **C**: Correlation between infection arrival time and geographic distance. Marker sizes scale with population size M_i for all $i \in \mathcal{N}$ and colors represent the region. From the best linear fit (black line) we determine the Pearson's correlation coefficient $r = 0.48$.

steps and the majority of nodes is reachable within a few connections. Therefore, the predictive power of the topological distance is relatively limited.

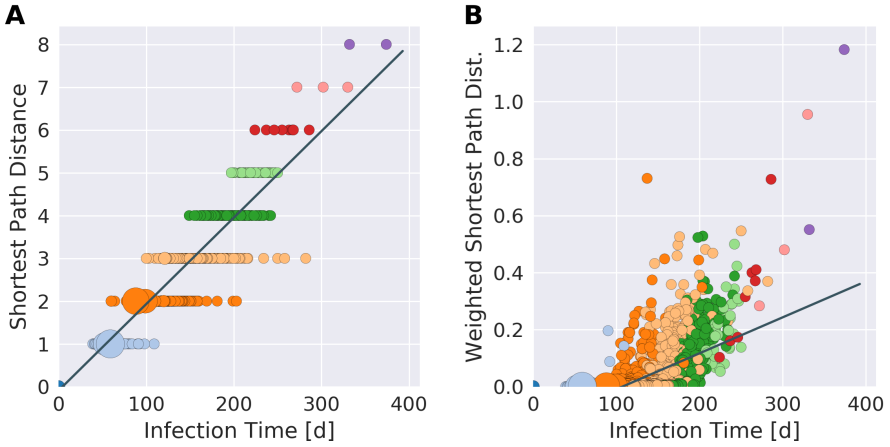

Figure 3.4: Correlation between simulated infection arrival times with simple graph metrics. **A**: The topological shortest path distance, i.e., the number of steps along the shortest path from the outbreak location (GRU), to the remaining nodes. Pearson's correlation coefficient $r = 0.90$ from linear regression (black line). **B**: Weighted shortest path distance with inverse passenger numbers as edge weights. From the linear fit (black line) we find $r = 0.62$. Marker size scales with population size and colors represent the topological distance, i.e., the shortest path distance, to the outbreak origin. Parameters as in Fig. 3.3.

One can try to improve the prediction by taking into account edge-weights, too. Intuitively, we expect that a high passenger flux translates into a shorter infection

transmission time. Therefore, a simple and obvious approach is to calculate the weighted shortest path distance with inverse edge-weights as a proxy for infection arrival times. In Fig. 3.4 B, we confirm the expected trend but nonetheless, the correlation leaves room for improvement.

Instead of examining more arbitrary network measures, we follow the approach of [114] and begin our analysis from first principles, i.e., we study the transmission along a single edge. Building on this, we increase the topological complexity gradually until we introduce the concept of effective distances. Thereby, we adapt and re-derive the main steps of the analysis in [114] for two reasons. First, we aim to clarify all assumptions that our work relies on and discuss their accuracy. In particular, we find limitations for when the analytical approach fails, and these limits apply not only to the previous work in [5, 114] but also to our main results, i.e., the RWED that will be discussed in Sec. 3.5. The second reason is that we deviate from [114] by (i) incorporating the assumption of a global exit rate as suggested in [3, 5], and (ii) using a deterministic mean-field model (see Eqs. (3.8)) instead of the stochastic approach in [114]. The assumption of a global exit rate allows to derive population sizes from passenger flows (see Eq. (3.5)) and simplifies the subsequent results. We label this assumption (A0):

(A0) Individuals move to an adjacent node at a constant rate γ, independent of their current location.

Since this section is closely related to the previous analyzis in [114], we state our scientific contributions as follows:

- We further simplify the analytical approximation from [114] such that the results depend only on two factors: (i) the topology of the network and (ii) a single system parameter.
- We reveal and discuss the similarities of two independent works, namely [114] and [5].
- We derive the multiple-path effective distance (MPED) from [114] as a generalization of the previously introduced dominant-path effective distance (DPED) [5].
- We interpret the transmission route of an infection as the realization of a random-walk. This central idea forms the basis for the random-walk effective distance (RWED) in Sec. 3.5

In the following we discuss a number of examples with increasing topological complexity and we begin with the simplest conceivable topology.

3.4.1 Two populations.

We start with two populations of size $M_1 = M_2 = M$ that are connected through an undirected edge as depicted in Fig. 3.5 A. The first node contains a small fraction of infected individuals and the symmetrical flow $F_{1,2} = F_{2,1} = F$ permits the

transport of pathogens to the second (susceptible) population. In order to estimate the time until the first individual arrives, we adopt the following assumptions from [114]:

(A1) The number of passengers per day is much smaller than the population size, i.e., $F \ll M$.

(A2) The early growth rate is much larger than the rate at which individuals leave a population, i.e., $\lambda \ll \gamma$.

(A3) The prevalence level $\rho^I(t^*)$ at time t^* when the first infected arrives at the adjacent node is small compared to the population size M (see Fig. 3.5 C).

Assumption (A1) is equivalent to requiring $\gamma \ll 1$ and appears justified according to our estimation in Eq. (3.7). In (A2), we limit the scope of our analysis to highly contagious diseases. However, we will demonstrate in Sec. 3.7 a high accuracy of our results for a broad range of epidemic parameters. From assumptions (A1) and (A2) we can expect a sufficiently high prevalence level $1/M \ll \rho^I(t^*)$ at time t^* when the infection hits the adjacent node such that we can neglect stochastic effects. At the same time, we assume $\rho^I(t^*) \ll 1$ according to (A3) and thus approximate the early stage of the epidemic with an exponential function $\rho^I(t) \approx \rho^I(0)\exp(-\lambda t)$. Here, the growth rate $\lambda = \beta - \mu$ of an early SIR dynamic results from the linearization around the disease-free state and under the condition $R_0 = \beta/\mu > 1$. Indeed, the low prevalence level $\rho^I(t^*)$, marked by a vertical line in Fig. 3.5 B justifies the exponential ansatz.

With the assumptions (A1)–(A3), the authors in [114] derive a Gumbel distribution $\widehat{G}(t)$, where $\widehat{G}(t)\mathrm{d}t$ estimates the probability that the infection enters the adjacent population within a small time interval $[t, t+\mathrm{d}t]$:

$$\widehat{G}(t) = \gamma \exp\left(\lambda t - \frac{F}{M\lambda} e^{\lambda t}\right) \tag{3.9a}$$

$$= \gamma \exp\left(\lambda t - \frac{\gamma}{\lambda} e^{\lambda t}\right) \tag{3.9b}$$

Here, we use the initial condition $\rho^I(0) = 1$ and the global exit rate γ from Eq. (3.6). Note that $\widehat{G}(t)$ is also defined for negative values of t and in order to avoid potentially negative infection arrival times that would violate causality, the authors in [114] introduced in Eq. (3.9b) the Heaviside function $\Theta(t)$. With this, we obtain the truncated distribution

$$G(t) = \gamma \exp\left(\lambda t - \frac{\gamma}{\lambda} e^{\lambda t}\right) \Theta(t). \tag{3.10}$$

Under assumption (A1) and (A2) the probability $\int_{-\infty}^{0} t\, \widehat{G}(t)\, \mathrm{d}t$ for negative arrival times is, however, negligible as we will observe later.

The French mathematician Gumbel introduced the distribution $G(t)$ in 1935 [195, 196] to investigate the largest and smallest values from a series of samples.

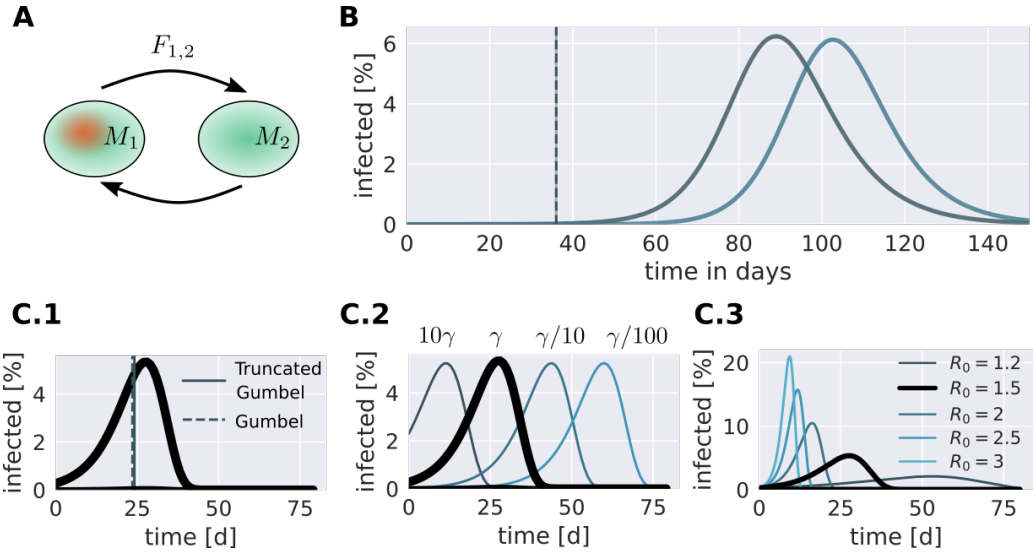

Figure 3.5: **A**: Meta-population model with two nodes of size $M_1 = M_2 = M$ and a symmetrical flow $F_{1,2} = F_{2,1} = F$. An epidemic outbreak begins in the first (left) node , whereas the second is initially susceptible. **B**: Local prevalence level relative to the population size. The vertical line indicates the time t^* when we consider the second node infected, i.e. $\rho^I(t^*) = 1/M$. Parameters as in Fig. 3.3. **C.1**: Gumbel distribution with same parameters as before. The overlapping vertical lines indicate the first moment of $\widehat{G}(t)$ (see Eq. (3.12e)) and the corresponding value of the truncated Gumbel distribution $G(t)$ (see Eq. (3.10)). **C.2**, **C.3**: Gumbel distribution for different values of γ and R_0, respectively. The latter determines the infection rate $\beta = R_0\mu$ for a fixed recovery rate μ as before.

Since then, the distribution has often been used to model rare and extreme events in various fields such as meteorology, seismology, biology, and finance [197–199].

In general, the so-called left-sided Gumbel distribution describes the smallest value $\min_{1\leq i\leq m} Y_i$ from a set $\{Y_i\}$ of m samples. Similar to the central limit theorem, it is not necessary to know the exact distribution from which the samples Y_i were drawn. Instead, the extremal distribution converges asymptotically for a large class of distributions to the canonical form

$$\lim_{m\to\infty} P\left(\min_{1\leq i\leq m} Y_i = x\right) = e^{z-e^z} \text{ , where } z = \frac{x-\mu}{\sigma} \tag{3.11}$$

Here, the parameters $\mu = \log(\lambda/\gamma)/\lambda$ and $\sigma = 1/\lambda$ determine the position and shape, respectively. In order to illustrate the impact of the mobility and growth rate on the shape of the distribution, we plot $G(t)$ in Fig. 3.5 C.2 and Fig. 3.5 C.3 for different values of γ and λ, respectively.

The first moment $\widehat{G}^{(1)}$ of the Gumbel distribution $\widehat{G}(t)$ (see Eq. (3.9b)) is well-known and we thus obtain the following expression for the expected transmission

time $\langle t \rangle$ of an infection:

$$\langle t \rangle \approx G^{(1)}(t) \tag{3.12a}$$

$$= \int_0^\infty t\, G(t)\, \mathrm{d}t \tag{3.12b}$$

$$\approx \int_{-\infty}^\infty t\, \widehat{G}(t)\, \mathrm{d}t \tag{3.12c}$$

$$= \widehat{G}^{(1)} \tag{3.12d}$$

$$= \frac{1}{\lambda} \log\left(\frac{\lambda e^{-\gamma_e}}{\gamma}\right). \tag{3.12e}$$

In the fist step (see Eq. (3.12a)), we approximate the simulated infection arrival time $\langle t \rangle$ with the first moment $G^{(1)}$ of the truncated Gumbel distribution $G(t)$. Then, we extend the integration in Eq. (3.12b) formally to the negative domain assuming that $\int_{-\infty}^0 t\widehat{G}(t)\, \mathrm{d}t$ is negligible. Finally, we take the well-known expression for the first moment $\widehat{G}^{(1)}$ of the Gumbel distribution to arrive at Eq. (3.12e). Here, we denote with $\gamma_e \simeq 0.57721$ the *Euler-Mascheroni number* [114].

In Fig. 3.5 C.1, we compare the first moment of $\widehat{G}(t)$ (see Eq. (3.12e)), which includes the negative domain, with the expectation value from the truncated distribution $G(t)$ (see Eq. (3.10)). With the parameters from Fig. 3.3, the relative difference is only 2.5%. Hence, the integral over the negative domain $\int_{-\infty}^0 t\widehat{G}(t)\, \mathrm{d}t$ is indeed neglible, which justifies our assumptions (A1) and (A2). Note, however, that a ten-fold increase in γ as shown in Fig. 3.5 C.2 violates (A2) considerably and we will return to the limits of this assumption in the discussion of parallel transmission paths.

For convenience, we introduce the dimensionless quantity δ and refer to it as the system parameter in the remaining chapter:

$$\delta = \log\left(\frac{\lambda e^{-\gamma_e}}{\gamma}\right). \tag{3.13}$$

With this important definition, we conclude our analysis of the two population model and proceed with a more complex system.

3.4.2 A chain of populations.

Next, we look at a chain of populations $i = 0, \ldots, N-1$ with symmetric passenger flows $F_{i,i+1} = F_{i+1,i}$ between adjacent nodes i and $i+1$. Here, we initiate a disease outbreak in the first population, i.e. at node $u = 0$, and determine the arrival time at a distant node v. The difficulty is that the flux of infected individuals to the adjacent population increases exponentially and these individuals can then propagate to the second neighbor and so forth. Thus, the infection arrival time in

population v depends not only on the adjacent node $v-1$ but, on all populations of the preceding chain.

However, following assumption (A2), the local disease dynamic dominates the epidemic growth rate compared to the in-flux of already infected individuals. With this, the authors in [114] demonstrated numerically that a good approximation is possible under the following additional assumption:

(A4) Once, a population i is infected, we treat the local disease dynamic as if the node was isolated, i.e., we ignore the in-flux of additional infected individuals from adjacent populations $j \in \mathcal{N}_i$.

An equivalent statement that emphasises the Markovian character goes as follows: After seeding the disease, the outbreak evolves independently of previously infected nodes. In Fig. 3.6 A.1, we consider a sequence of infections along a chain with homogeneous edge weights. Following assumption (A4), we expect that the infection time between adjacent nodes remains constant in the course of an epidemic and indeed, numerical simulations in Fig. 3.6 A.2 confirm the accuracy of our premise.

With assumption (A4), we can approximate the arrival time statistics from population i to $i+1$ with a Gumbel distribution similar to Eq. (3.10):

$$G_{i,i+1}(t) = \gamma T_{i,i+1} \exp\left(\lambda t - \frac{F_{i,i+1}}{\lambda} e^{\lambda t}\right) \Theta(t) \tag{3.14a}$$

$$= \gamma T_{i,i+1} \exp\left(\lambda t - \frac{\gamma T_{i,i+1}}{\lambda} e^{\lambda t}\right) \Theta(t). \tag{3.14b}$$

In contrast to [114], we derived Eq. (3.14b) under the assumption of a global mobility rate γ (see Eq. (3.5)). Moreover, we used the conditional probability T_{ij} that an individual leaving the population i will choose the neighboring node $i+1$ as defined in Eq. (3.4). The corresponding first moment $G^{(1)}_{i,i+1}$ approximates the mean transmission time $\langle t_{i,i+1} \rangle$ from i to $i+1$ and is given by

$$G^{(1)}_{i,i+1} = \frac{1}{\lambda} \log\left(\frac{\lambda e^{-\gamma_e}}{\gamma T_{i,i+1}}\right). \tag{3.15}$$

Next, we consider the transmission time from the first to the last population with index $u = 0$ and $v = N-1$, respectively. Also, we differentiate between the simulated and estimated arrival time, denoted by $\langle t_{uv} \rangle$ and χ_{uv}, respectively. Using assumption (A4), we obtain χ_{uv} from the sum over the first moments $G^{(1)}_{i,i+1}$

for all consecutive edges $i \to i+1$:

$$\lambda\langle t_{uv}\rangle \approx \lambda\chi_{uv} \tag{3.16a}$$

$$= \sum_{i=0}^{N-1} G^{(1)}_{i,i+1} \tag{3.16b}$$

$$= \sum_{i=0}^{N-1} \log\left[\frac{\lambda e^{-\gamma_e}}{\gamma T_{i,i+1}}\right] \tag{3.16c}$$

$$= N\log\left(\frac{\lambda e^{-\gamma_e}}{\gamma}\right) - \sum_{i=0}^{N-1}\log\left(T_{i,i+1}\right) \tag{3.16d}$$

$$= N\delta - \log\left(\prod_{i=0}^{N-1} T_{i,i+1}\right). \tag{3.16e}$$

In Eq. (3.16c), we replaced the first moments $G^{(1)}_{i,i+1}$ with Eq. (3.15) and separate the contributions into two factors: (i) We integrate the epidemic growth and the mobility rate $\lambda = \beta - \mu$ and γ, respectively, into the system parameter δ according to Eq. (3.13). (ii) A separate product aggregates the conditional probabilities $T_{i,i+1}$ (see Eq. (3.16e)). This term adds the topology to the estimation of the infection arrival time.

In Fig. 3.6 B we study an outbreak on a chain with heterogeneous edge-weights. Here, we draw edge weights from the air-traffic network and normalize them such that the mobility rate γ remains the same as in Fig. 3.3. Despite the heterogeneity, we find in Fig. 3.6 B.3 a close-to-constant prevalence level relative to the population size. Moreover, Fig. 3.6 B.2 confirms that we are able to estimate the infection arrival time (marked by vertical lines in Fig. 3.6 B.3) with high accuracy. The slope deviates slightly due to the simplifying assumption (A4) but, we emphasize again that our aim is to optimize the correlation and not the prediction of absolute times.

3.4.3 Tree graph.

From the transition probabilities $T_{i,i+1}$ in Eq. (3.15) it becomes obvious that transmission times between an arbitrary pair of adjacent populations i and $i+1$ are not symmetric in general, even as we assume undirected passenger flows. In order to illustrate this idea, we look at a star graph in Fig. 3.7 A. Here, the transition probability from a hub to any peripheral node scales with the system size, i.e., $T_{\text{hub}\to} = 1/N$. Therefore, we expect that the corresponding transmission time $\langle t_{\text{hub}\to}\rangle$ increases logarithimally with N while in the opposite direction, the time to infection $\langle t_{\to\text{hub}}\rangle$ remains constant because the probability $T_{\to\text{hub}} = 1$ is

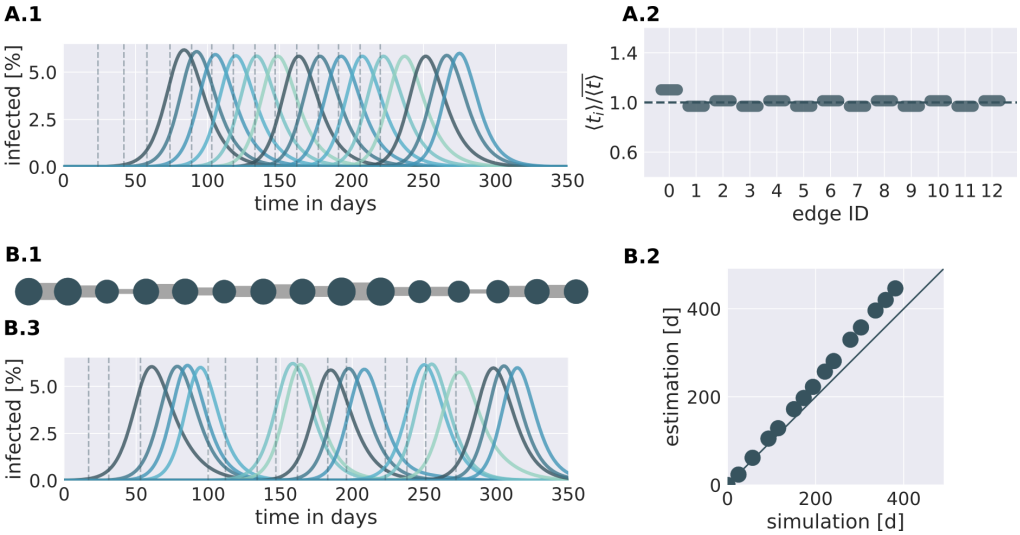

Figure 3.6: Disease arrival times for a chain topology, where the first node is initially infected. **A.1**: Relative prevalence curves for all nodes along a homogeneous chain with disease parameters as in Fig. 3.3. Vertical lines mark simulated infection arrival time. **A.2**: Infection time between adjacent nodes relative to the average value. **B.1**: Heterogeneous chain with edge weights sampled from the air-traffic network (see Fig. 3.2 B.3). We normalize the average node strength $\langle F \rangle$ as in Tab. 3.1 remaining parameters before. **B.2**: Local prevalence curves and simulated infection arrival times marked by vertical lines for the heterogeneous chain. **B.3**: Correlation between simulated and estimated infection arrival times (according to Eq. (3.16e)) for the heterogeneous chain. .

independent of the system size:

$$\lambda \langle t_{\text{hub}\rightarrow} \rangle \approx \delta - \log(1/N) \tag{3.17a}$$

$$\lambda \langle t_{\rightarrow \text{hub}} \rangle \approx \delta \tag{3.17b}$$

Indeed, the simulation in Fig. 3.7 A.2 confirms the expected behavior and we find a close correlation between the simulated and estimated results in Fig. 3.7 A.3.

A fundamental contribution of our work is to identify T_{ij} in Eq. (3.16e) with the transition probability of a canonical random walk. We formulated the definition of T_{ij} in Eq. (3.4) and referred to individuals that leave population i and arrive in adjacent node j. Here, we consider an abstract random process instead, where the infected population as a whole defines the random walker. One can think of it as a propagation front that is defined by the last node in a chain of infections. This out-most infected population, suppose it is node i, seeds a new outbreak in one of the adjacent susceptible nodes and hence the propagation front moves from i to j, similar to a random walker, with transition probability T_{ij}.

From the perspective of a random-walk process, we can interpret the product of transition probabilities in Eq. (3.16e) as a sampling probability for a Markov chain that moves straight from node u to v, i.e., without returning to previously infected populations. We call this sequence of jumps where every node is visited at most once a uv-path. In other words, the sampling probability $P(\pi)$ to realize a given path $\pi = [(u, i_1), (i_1, i_2), \dots (i_{N-1}, v)]$ with a random walk equals the product of

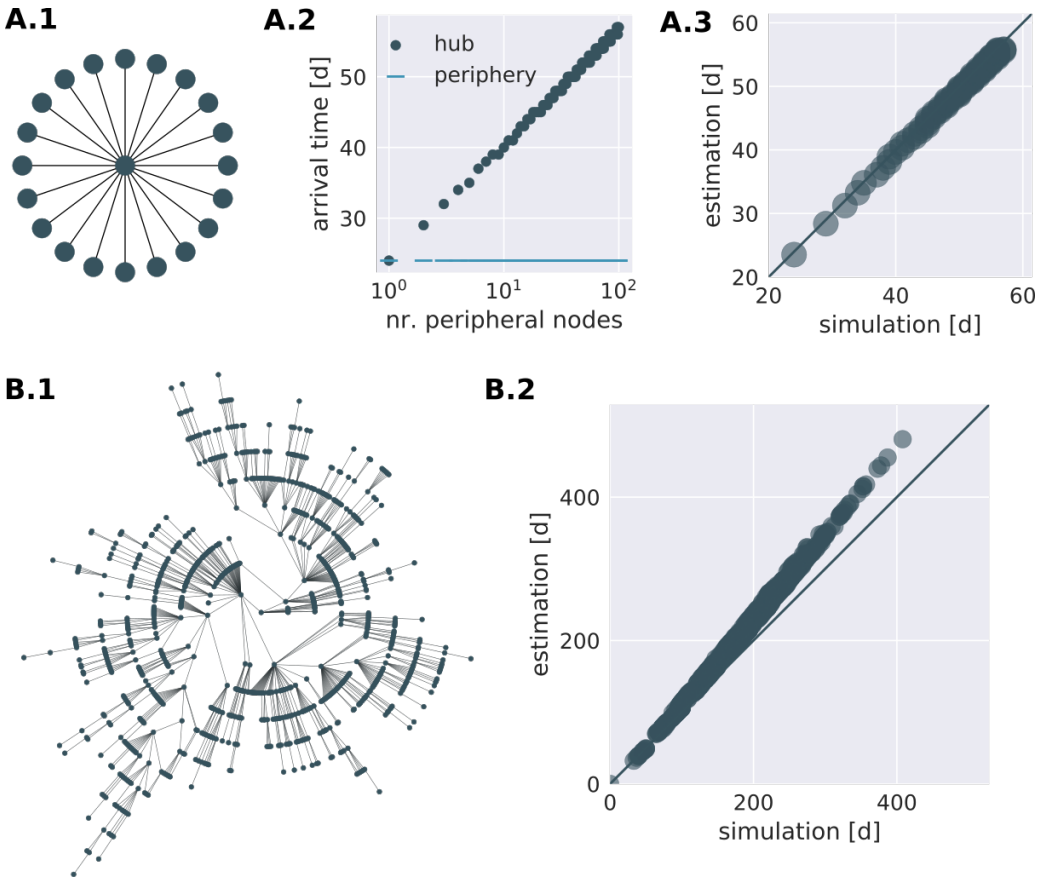

Figure 3.7: **A.1**: Hub-periphery network with homogeneous edge weights. **A.2**: Disease transmission from the hub to a given peripheral node (dark circles) and vice versa (horizontal lines) for an increasing network size (semi-logarithmic). **A.3**: Correlation with estimated arrival times according to Eq. (3.17). **B.1**: Random tree graph with heterogeneous edge weights sampled from the air-traffic network (see Fig. 3.2B.3) and with normalized node strength $\langle F \rangle$ according to Tab. 3.1. **B.2**: Correlation with estimated infection arrival times according to Eq. (3.19). . Parameters as in Fig. 3.3.

the transition probabilities along π:

$$P(\pi) = \prod_{(i,j)\in\pi} T_{ij}. \tag{3.18}$$

With these general considerations we return to Eq. (3.16e), i.e., to our propagation process along a chain. Here, we identify the sampling probability $P(\pi)$ for the uv-path π that starts in u and arrives in v. Also, we denote the estimated transmission time and path length with $\chi(\pi) \equiv \chi_{uv}$ and $n(\pi)$, respectively. With these notational changes we re-formulate Eq. (3.16e):

$$\lambda\chi(\pi) = n(\pi)\delta - \log[P(\pi)]. \tag{3.19}$$

The outbreak dynamics on an arbitrary tree as depicted in Fig. 3.7 B is closely related to the case of a chain graph because, there exists only a single uv-path π that leads from the outbreak origin, i.e., the central node in Fig. 3.7 B.1 to any other node v. In order to evaluate a more realistic scenario, we add heterogeneous edge

weights, sampled from the air-traffic network and normalize them such that the mobility parameter γ is the same as in Fig. 3.3. Indeed, we confirm in Fig. 3.7 B.2 a high correlation between the estimated transmission according to Eq. (3.19) and a simulated epidemic. Although the transition to random walk theory at this point appears rather cosmetic and formal, we will benefit from the correspondence as we proceed to complex topologies.

3.4.4 Parallel paths.

An essential observation with complex topologies is that a large number of paths can connect any two nodes. In the context of contact networks, the effect of parallel infection paths on arrival times has already been studied [200]. Also for a stochastic meta-population model, the authors in [114] demonstrated that a growing number of parallel transmission routes alters the expected infection arrival time significantly.

In order to improve our intuitive understanding, we first return to a hub-periphery network with $N+1$ nodes and homogeneous edge weights as depicted in Fig. 3.8 A.1. Unlike the previous example in Fig. 3.7 A, we consider now the case where all peripheral nodes are initially infected. For this example with N infected neighbours, we can determine the probability density function $G_N(t)$ for a disease transmission to the hub at time t as follows:

$$G_N(t|\gamma,\lambda) = N\left[1-\int_0^t G(t')\,\mathrm{d}t'\right]^{N-1} G(t) \tag{3.20a}$$

$$= N\left[1-1+\exp\left(-\frac{\gamma}{\lambda}e^{\lambda t}\right)\right]^{N-1}\gamma\exp\left(\lambda t-\frac{\gamma}{\lambda}e^{\lambda t}\right) \tag{3.20b}$$

$$= N\gamma\exp\left(\lambda t-\frac{N\gamma}{\lambda}e^{\lambda t}\right) \tag{3.20c}$$

$$= G(t|N\gamma,\lambda). \tag{3.20d}$$

In Eq. (3.20a), we calculate the probability that $N-1$ nodes fail to transmit up to time t each with probability $1-\int_0^t G(t')\,\mathrm{d}t' = \exp(-\gamma/\lambda e^{\lambda t})$. The remaining peripheral node succeeds with probability $G(t)\mathrm{d}t$, where the Gumbel distribution $G(t)$ has been derived in Eq. (3.10). The pre-factor N in Eq. (3.20a) accounts for N independent and identical transmission paths that can lead to the infection of the hub. According to Eq. (3.20d), it turns out that $G_N(t|\gamma,\lambda)$ is a Gumbel distribution, where the mobility rate γ scales with the number of peripheral nodes N, hence the notation $G(t|N\gamma,\lambda)$. Therefore, we can immediately formulate an analytical estimation for the infection arrival time from N infected neighbors to a central hub:

$$\lambda\,\langle t_{\to\mathrm{hub}}\rangle \approx \log\left(\frac{\lambda e^{-\gamma_e}}{N\gamma}\right). \tag{3.21}$$

Indeed, we confirm in Fig. 3.8 A.3 the accuracy of Eq. (3.21) with respect to numerical simulations for sufficiently small γ. Note, however, that a large number of infected neighbors may violate assumption (A2), i.e. $N\gamma \ll \lambda$. This can be observed in Fig. 3.8 A.2. Here, the negative domain of the Gumbel distribution (see Eq. (3.9b)) can significantly alter the expected arrival time in Eq. (3.21). Therefore, we expect that our analysis underestimates the infection arrival time at a hub in the air-traffic network.

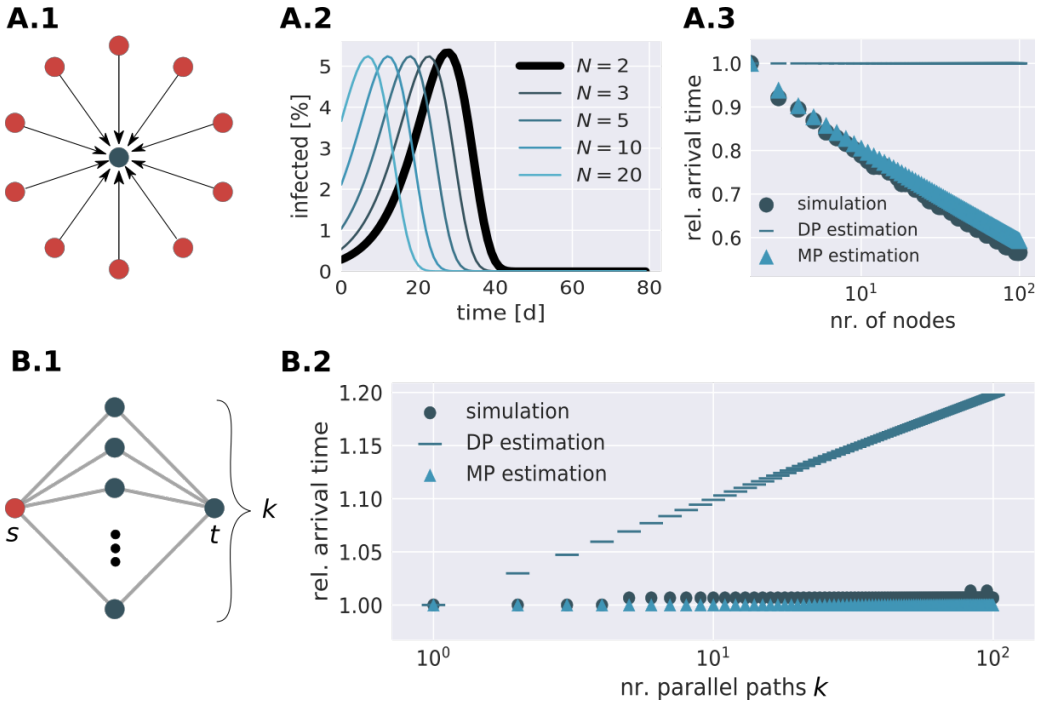

Figure 3.8: **A.1**: Hub-periphery graph with homogeneous edge weights and all peripheral nodes infected initially. **A.2**: Gumbel distribution according to Eq. (3.20d) for different numbers of nodes N (see panel A.1.). **A.3**: Simulated and estimated infection arrival times for an increasing number of nodes (see panel A.1). We compare the multiple path (MP) estimation according to $G_N(t)$ (see Eq. (3.21)) with the dominant path (DP) approximation $G(t)$ (see Eq. (3.12e)) that accounts only for a single transmission edge. **B.1**: Topology with k independent transmission routes from node u (initially infected) to v. **B.2**: Simulated and estimated infection arrival times for an increasing number of parallel paths. The multiple path (MP) estimation according to Eq. (3.24) accounts for competing infections, whereas the dominant path (DP) estimation (see Eq. (3.25)) focuses on a single transmission path. Parameters as in Fig. 3.3 except for the exit rate, which we reduced to $\gamma = 10^{-5}$ in order to satisfy assumption (A2).

Next, we turn to a topology with k parallel propagation paths $\pi_1, \dots \pi_k$, all of which begin at node u and eventually converge into v (see Fig. 3.8 B.1). We determine the propagation along each isolated path π_i according to Eq. (3.19). Then, we re-use the symbol χ_{uv} to denote the estimated infection arrival time, including all paths $\pi \in \mathcal{P}_{uv}$, where $\mathcal{P}_{uv}$ denotes the set of uv-paths that lead from u to v. The authors in [114] expressed χ_{uv} in terms of individual transmission times $\chi(\pi_i)$ along each path π_i. For this purpose they derived the following addition

theorem under the previously made assumptions (A1)-(A4):

$$e^{\lambda \chi_{uv}} = \sum_{\pi \in \mathcal{P}_{uv}} e^{\lambda \chi(\pi)}. \tag{3.22}$$

Then, the authors in [114] demonstrated numerically that Eq. (3.22) provides a good approximation of the actual arrival time, i.e., $\lambda \langle t_{uv} \rangle \approx \chi_{uv}$. Within our framework that includes the additional assumption (A0) of a global mobility rate γ, we thus obtain the following estimation:

$$\lambda \langle t \rangle \approx -\log \left[\sum_{\pi \in \mathcal{P}_{uv}} e^{-\lambda \chi(\pi)} \right] \tag{3.23a}$$

$$= -\log \left[\sum_{\pi \in \mathcal{P}_{uv}} e^{-n(\pi)\delta + \log[P(\pi)]} \right] \tag{3.23b}$$

$$= -\log \left[\sum_{\pi \in \mathcal{P}_{uv}} P(\pi) e^{-n(\pi)\delta} \right] \tag{3.23c}$$

In Fig. 3.8 B.2 we confirm the accuracy of Eq. (3.23c) for an increasing number of parallel paths as depicted in Fig. 3.8 B.1. Note, however, that the accuracy depends crucially on a sufficiently small mobility rate γ as mentioned earlier for Fig. 3.8 A.3.

3.4.5 Multiple path effective distance.

In a complex topology, such as the air-traffic network, we usually find a large set of paths $\mathcal{P}_{uv}$ that connect pairs of nodes u and v. Unlike independent paths that we considered in Fig. 3.8 B, it may now happen that different transmission routes cross each other and thus introduce correlations. A precise analysis, within the previously made assumptions, requires to disassemble the network into elementary units, i.e., chains and parallel paths as depicted in Fig. 3.6 B.1 and Fig. 3.8 B.1, respectively. This approach is similar to a complex electric circuit, where a careful analysis requires to consider seperate components with serial and parallel elements. For large empirical networks, this analysis becomes, however, exceedingly demanding and instead, we introduce the final assumption (A5) to simplify the calculation:

(A5) We consider all paths from the outbreak location to a given target node as independent transmission routes.

In other words, following the authors in [114], we assume that Eq. (3.23c) applies to arbitrary topologies. By ignoring correlations between transmission paths, we underestimate the actual transmission time to distant nodes. The extent to which assumption (A5) introduces an additional error depends on the topology. However, we will show in Sec. 3.7 that a surprisingly good prediction is possible for a broad class of graphs, which justifies assumption (A5).

We already mentioned in the introduction to this chapter that we aim to find a network measure that *correlates* with the mean infection arrival time. Hence, we

do not value the exact moment, because in practice the absolute time depends on too many unknown parameters. Instead, the authors in [5] argue that a linear fit after the first disease transmissions have been detected gives a more robust prediction. Following this consideration we denote the dimensionless quantity on the right side of Eq. (3.23c) the *multiple-path effective distance* (MPED) from node u to v:

$$D_{uv}^{\mathrm{MP}} = -\log\left[\sum_{\pi\in\mathcal{P}_{uv}} P(\pi)e^{-n(\pi)\delta}\right]. \tag{3.24}$$

With this measure, we have found a conceptually attractive measure for the expected transmission time $\langle t_{uv}\rangle$ of a disease from a given outbreak location u to a remote node v of a complex network. In Eq. (3.24), we weigh the probability $P(\pi)$ to realize a certain uv-path π with a factor that decreases exponentially with the corresponding path length $n(\pi)$. Furthermore, the dimensionless system parameter δ determines the importance of short transmission paths. This quantity is strictly increasing with the growth rate $\lambda = \beta - \mu$ and decreases with the global mobility parameter γ (see Eq. (3.13)). In other words, a rapidly spreading disease will tend to run along the topologically shortest path, whereas alternative propagation paths may become important as the mobility rate increases.

We summarize the most important characteristics of the MPED:

- We derived the MPED from a detailed analysis of the meta-population model.
- The MPED includes all potential transmission paths and thus integrates the complex topology of the network.
- Following our intuition, we assign a short (effective) distance from u to v if a large number of probable and short transmission routes exists.
- The dimensionless paramter δ integrates the epidemic growth rate $\lambda = \beta - \mu$ and the mobility rate γ.
- With an increasing value of δ, fewer transmission paths dominate the sum in Eq. (3.24). These routes are both, likely realizations of a random walk and topologically short.

The difficulty in using the MPED (see Eq. (3.24)) is the exceedingly vast number of uv-paths that lead from the outbreak location u to a target node v. A direct, but computationally expensive way is to iterate a depth-first search, i.e., we go through each path π, assign a value $P(\pi)e^{-n(\pi)\delta}$, and add it to the sum in Eq. (3.24). The time complexity to find a single path scales linearly with the number of nodes N and edges E, i.e. $\mathcal{O}(V + E)$. However in the worst-case of a complete graph, $N!$ paths exist [201].

A better approach leverages the fact that the contribution of a single path π to the sum in Eq. (3.24) decreases exponentially with its length $n(\pi)$. It is therefore advantageous to first find the topologically shortest connections using a

depth-first search. In a second step, one can consider longer paths incrementally according to the algorithm of [202]. This approach leads to a time complexity of $\mathcal{O}(KN^2 \log(N))$, where K denotes a cut-off for the maximum number of paths.

The simplest approximation, however, is to limit the sum in Eq. (3.24) to the dominant contribution. This idea finally leads to the *dominant path effective distance* (DPED), which we will focus on later.

In Fig. 3.9, we illustrate the importance of competing transmission paths and computational difficulties associated with them. As an example, we estimate the transmission time from GRU to Bratislava (BTS) in Fig. 3.9 A. This choice is instructive because the dominant transmission route of length $n = 3$ (via London Heathrow (LHR) and Dublin (DUB)) differs from the topologically shortest route of length $n = 2$ (via Lamu (LAU) in Kenya). Here the significance of the probability $P(\pi)$ compared to the path length $n(\pi)$ becomes particularly evident.

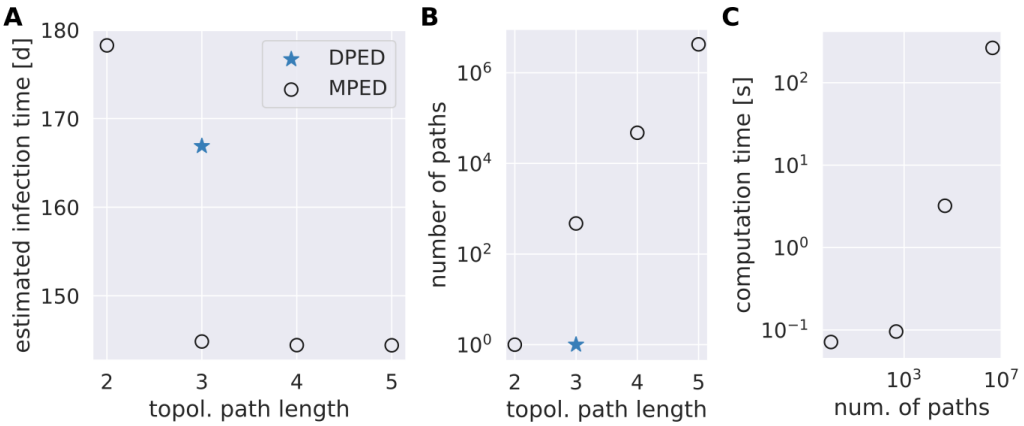

Figure 3.9: Impact of multiple competing transmission paths. **A**: Estimated infection arrival time from GRU to Bratislava (BTS) using the same parameters as in Fig. 3.3. We compare the DPED from Eq. (3.25) (blue star) with the MPED from Eq. (3.24) (circles). Thereby we consider an increasing number of path lengths $n = 2, \ldots, 5$. Note that the dominant path is of length $n = 3$, whereas the shortest path has length $n = 2$. **B**: Number of paths from GRU to BTS for different path lengths. **C**: Computational time on a commercial laptop for an increasing number of paths. . Parameters as in Fig. 3.3.

This example also demonstrates how the estimated transmission time declines with an increase in the number of possible paths. In Fig. 3.9 A, we calculate the MPED using Eq. (3.24) first, by considering only the shortest topological path via LAU. In the second step, we add 473 paths of length three, including the dominant transmission path via LHR and DUB. By accounting for these competing paths, we reduce the estimated transmission time by about 25% from 179 to 144 days. The deviation to the dominant transmission path is smaller, but still more than 11 days or 15%.

The number of paths from GRU to BTS increases exponentially according to Fig. 3.9 B and so does the computational time (see Fig. 3.9 C). Also, the contribution of larger path lengths appears negligible. Therefore, one could suppose that the topologically shortest paths together with those that have one more connection give a sufficiently accurate estimation. Such a rule of thumb

cannot, however, be generalized because, the saturation in Fig. 3.9 A depends on the topology and the selected source and target node. One can find a pair of nodes, where greater path lengths have to be included, whereas for yet another choice paths of length $n = 2$ can no longer be calculated in a reasonable time.

This example shows, on the one hand, the necessity to consider competing transmission paths and on the other hand the difficulty in dealing with the MPED.

3.4.6 Dominant path effective distance.

In order to provide a computationally efficient method to estimate infection arrival times, the authors in [114] suggest to consider only the dominant transmission routes. Moreover, the authors in [5] argue that the heterogeneity of the air-traffic network regarding the degree and weight distribution (see Fig. 3.2) justifies this simplification. In this case, the argument goes, it is highly probable to find a path π^* that dominates the sum in Eq. (3.24). We call this approximation the *dominant path effective distance* (DPED):

$$D_{uv}^{\mathrm{DP}} = n(\pi^*)\delta - \log\left[P(\pi^*)\right]. \tag{3.25}$$

With Eq. (3.25) we can now determine the shortest path tree that leads from the outbreak location to the remaining nodes in the network along the dominant transmission routes. In Fig. 3.10, we visualize this idea for a single potential outbreak location. To this end, we rearrange the air-traffic network and provide an ego-centric view from the perspective of GRU. The distance from the central node to the remaining airports is proportional to the DPED according to Eq. (3.25). Note also that the angular position is arbitrary.

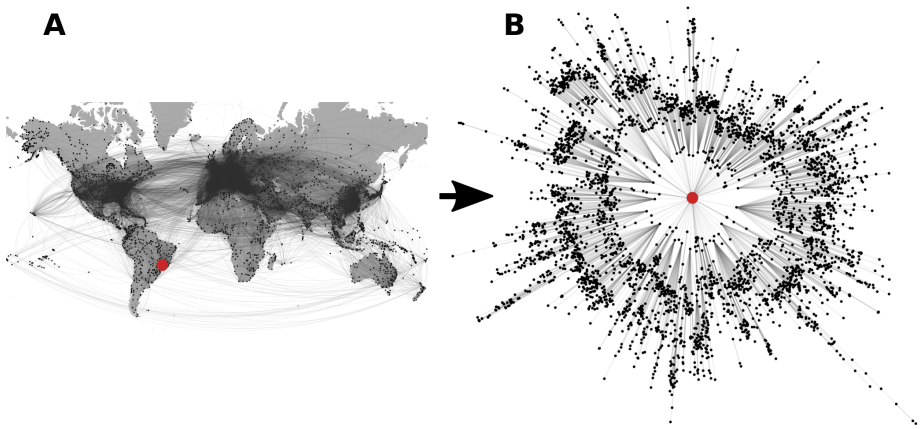

Figure 3.10: Example of an ego-centric visualization. **A**: Global air-traffic network as in Fig. 3.1, with São Paulo-Guarulhos (GRU) marked in red. **B**: Shortest path tree with distances from the center, i.e. GRU (marked in red), according to the DPED (see Eq. (3.25)). Edges mark the dominant transmission routes.

In [5] the authors introduced independently the special case of Eq. (3.25) with $\delta = 1$ and coined the term *effective distance*. Their heuristic considerations, as opposed to the model-based derivation in [114] and our previous analysis, are as follows:

- A high (conditional) probability $T_{ij} = F_{ij}/F_i$ to move from i to j should correspond to a low effective distance and vice versa.
- If there is only a single outgoing edge from i to j, such that $T_{ij} = 1$, then we define $D_{ij}^{\mathrm{DP}} = 1$.
- The effective distance for a sequence of edges is supposed to be additive, whereas the probabilities T_{ij} should behave multiplicatively.

The authors in [5] defined the following metric that meets the requirements:

$$d_{ij} = (1 - \log T_{ij}) > 0 \text{ for all edges } i \to j. \tag{3.26}$$

The so-defined metric between adjacent nodes is not symmetric, i.e. $d_{ij} \neq d_{ji}$ in general, even if the graph is undirected. Thus, Eq. (3.26) is not a distance measure in a strict sense. Indeed, we observed asymmetric infection arrival times already for the hub-periphery networks in Fig. 3.8 A. Here, the transmission from a central node, i.e. a the hub, to a given peripheral node takes longer than the other way round.

Next, the authors in [5] assign a value $\chi(\pi) = \sum_{(i,j)\in\pi} d_{ij}$ to each uv-path π that connects a given pair of nodes u and v. This corresponds to our earlier considerations about propagation along a chain that led to Eq. (3.19). Finally, the authors define the (dominant path) effective distance D_{uv}^{DP} from the outbreak location u to a distant node v as the minimum over all $\chi(\pi)$, i.e.

$$D_{uv}^{\mathrm{DP}} = \min_{\pi} \chi(\pi). \tag{3.27}$$

Numerically, the DPED can be determined efficiently using a breadth-first search. To this end, we replace all edge weights F_{ij} with the distance d_{ij} as defined by Eq. (3.26). Then, the weighted shortest distance from u to v on the rescaled graph determines the DPED as defined by Eq. (3.25) and Eq. (3.27), respectively. For a given outbreak location u, we can calculate the DPED to all remaining nodes with time complexity $\mathcal{O}(E + V \log V)$ using, e.g., the well-known Dijkstra algorithm [150]. The computational advantage and simple implementation make the DPED a highly attractive tool for the analysis of large graphs such as the air-traffic network.

In Fig. 3.11 we study again a global outbreak seeded in GRU with the same parameters as in Fig. 3.3. The subfigures in Fig. 3.11 A visualize the dynamics for four different time steps using an ego-centric perspective. Unlike the geographic representation (see Fig. 3.3 A), one can easily observe a concentric pattern that begins at the origin (GRU) and propagates wave-like to the remaining node.

In Fig. 3.11 B we sketch the inverse problem to find the outbreak location from a given snapshot. The authors in [5] suggest an intuitive and numerically efficient approach motivated by the concentric spreading pattern observed in Fig. 3.11 A: Here, we select four representative airports. For each, we arrange the remaining nodes in a distance according to the DPED and mark the nodes that are already

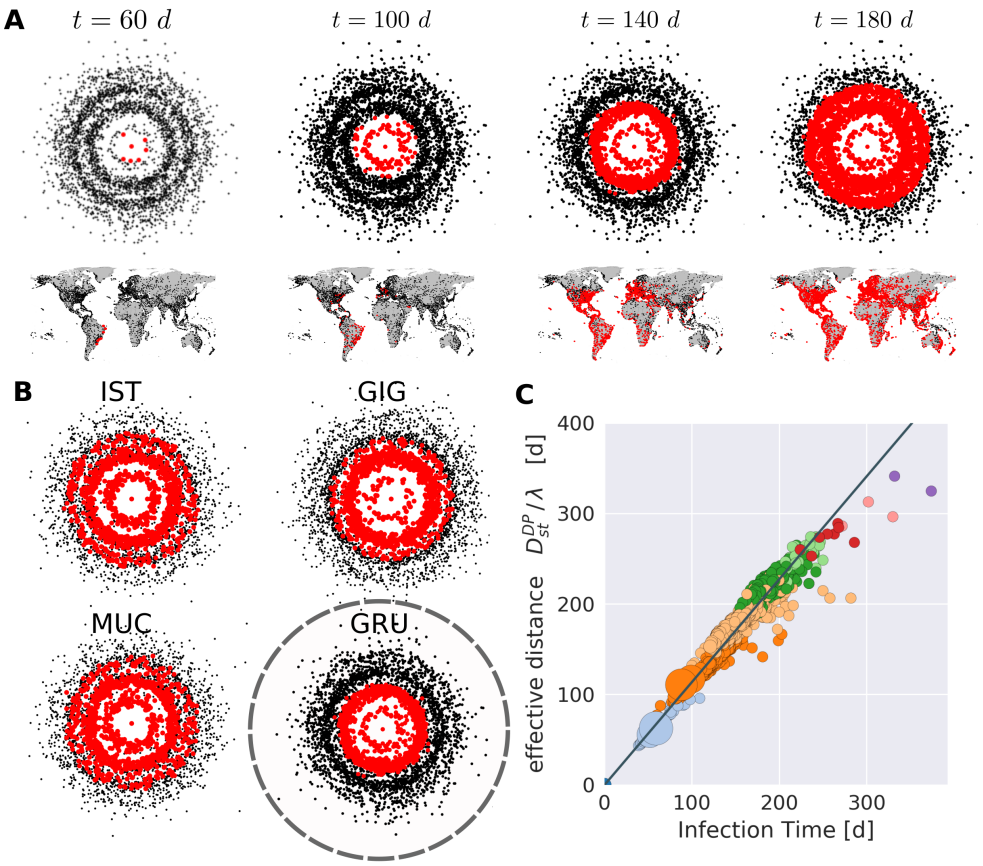

Figure 3.11: Applications of the DPED. **A**: Four snapshots of the simulated outbreak as in Fig. 3.3 using a ego-centric visualisation. The DPED (see Eq. (3.25)) determines the distance to the outbreak location, i.e. the GRU. The angular position is chosen randomly. Infected populations according to the simulation are marked red. **B**: Different ego-centric visualisations of the same outbreak simulation at time $v = 140$ days. A dashed cicle marks the true outbreak origin, i.e. GRU (see also Fig. 3.11 A at time $v = 140$ days). **C**: Correlation between simulated infection arrival times and estimated values according to the DPED (see Eq. (3.25)). Node size scales with population size and colors indicate the topological distance to the outbreak location (see Fig. 3.4). Best linear fit (black line) determines Pearson's correlation coefficient $r = 0.97$. . Parameters as in Fig. 3.3.

infected. Now, we identify the correct outbreak location using the following criteria: (i) The average distance of all infected nodes to the centre is minimal as well as (ii) the variance around the average value. Intuitively speaking, we select the node that gives the most compact spreading pattern, in this case, GRU.

The ability to predict the epidemic as well as the inverse problem to find the outbreak origin depend crucially on the high correlation between the simulated and estimated infection arrival times. In Fig. 3.11, we demonstrate for GRU that the DPED can be used as a good proxy for simulated values and indeed for real epidemic outbreaks as demonstrated in [5].

To conclude this section, we summerize our main results as follows. We adopted the work of [114] and derived two network-based measures as a proxy for infection arrival times. The MPED in Eq. (3.24) is conceptually appealing but, can be only applied to small networks due to the large number of potential transmission routes. As a computational efficient alternative, we motivated the DPED in Eq. (3.25).

It is, however, an approximation (see Fig. 3.9 A) that applies only to strongly heterogeneous topologies and a sufficiently fast disease propagation. Therefore, we conclude that a computationally efficient method that can account for alternative transmission paths is still an open problem. In the following section, we address this issue and derive the *random-walk effective distance* and thus extend the scope of the DPED.

3.5 The random walk effective distance

We began our analysis in Sec. 3.4 with assumption (A0), i.e., the existence of a global mobility parameter γ. With this, we derived the MPED in Eq. (3.24) and the DPED in Eq. (3.25) as a proxy for expected infection arrival times from a detailed analysis adopted from [114]. The main contributing factors to both distance measures are: (i) the probability $P(\pi)$ of realising a path π that connects the outbreak location to a remote node using a random-walk (see Eq. (3.18)) and (ii) the corresponding path length $n(\pi)$.

The probability $P(\pi)$ hints at a potential correspondence between a diffusive process and epidemic spreading. Indeed, we noted in the derivation of Eq. (3.19) that we can interpret epidemic transmission routes as self-avoiding random-walks with a bias towards short and probable paths. In this section, we will, therefore, leverage this correspondence and improve analytical estimations of the mean infection arrival time by applying the theory of Markov chains.

A self-avoiding random walk is a correlated random process, where an abstract agent, i.e. the random walker, moves between adjacent nodes. After n jumps, the random walker chooses an edge $i \to j$ from the set of all out-directed edges with step-dependent transition probability:

$$T_{ij}(n) = \begin{cases} \frac{F_{ij}}{\sum_{j' \in \mathcal{N}_i(n)} F_{ij'}} & \text{if } j \in \mathcal{N}_i(n) \\ 0 & \text{otherwise.} \end{cases} \tag{3.28}$$

Here, we denote with $\mathcal{N}_i(n)$ the set of nodes adjacent to i that have not been visited during the past n jumps. It reflects the requirement to visit every node at most once, which leads to long-range correlations. These, in turn, limit the applicability of the MPED in Eq. (3.24) to small networks.

Instead of considering only the dominant contribution as in [5, 114], we approximate the MPED by neglecting the correlations in Eq. (3.28). In other words, we consider a Markovian random-process that can visit nodes repeatedly. Formally, we replace the sum over all uv-paths $\pi \in \mathcal{P}_{uv}$ in Eq. (3.24) with a sum over absorbing uv-walks $\omega \in \mathcal{W}_{uv}$. An absorbing uv-walk ω leads from the outbreak location u to a distant target node v, and can pass every node *except for* v multiple times. Hence, our only restriction is that a walk ends as soon as it reaches the target node v, i.e., the absorbing state. We rephrase our main assumption as follows:

(A6) Every random walk ω from the outbreak location u to an absorbing state v represents a potential transmission path of an epidemic outbreak.

The Markovian approximation (A6) leads to our definition of the *random walk effective distance*:

$$D_{uv}^{\mathrm{RW}} = -\log\left[\sum_{\omega\in\mathcal{W}_{uv}} P(\omega)e^{-n(\omega)\delta}\right]. \tag{3.29}$$

Unlike (self-avoiding) paths, walks can return to a node multiple times and thus form loops. Intuitively speaking, a disease can re-infect a previously infected population as if it was entirely susceptible. This consequence seems undesirable, in particular, for the SIR epidemic model, which leads to permanent immunity. However, we will demonstrate in Sec. 3.7 that the error introduced as a result of our Markovian assumption (A6) is negligible for realistic disease parameters. The accuracy of our approach is a consequence of the additional number of steps that are required to realize a loop. In this case, the associated factor $e^{-n(\omega)\delta}$ in Eq. (3.29), leads to an exponentially dampened contribution as compared to a loop-free realisation. Since simple paths form a subset of all possible walks, our newly proposed approach gives a lower bound to Eq. (3.24) (see Fig. 3.12). Similarly, the DPED gives an upper bound and hence we find for all pairs of nodes u and v the following relation:

$$D_{uv}^{\mathrm{RW}} \leq D_{uv}^{\mathrm{MP}} \leq D_{uv}^{\mathrm{DP}} \tag{3.30}$$

We find the equality $D_{uv}^{\mathrm{MP}} = D_{uv}^{\mathrm{DP}}$ in case of a unique uv-path, e.g. in a tree topology as depicted in Fig. 3.7 B.1. The first equality $D_{uv}^{\mathrm{RW}} = D_{uv}^{\mathrm{MP}}$ requires additionally non-reciprocal edges, such that a random walker cannot return to previously visited nodes.

It is instructive to rearrange the sum in Eq. (3.29) such that all walks of the same length are grouped together. All contributions to a particular walk length n can now be associated with the first passage or hitting probability $H_{uv}(n)$:

$$D_{uv}^{\mathrm{RW}} = -\log\left[\sum_{n=0}^{\infty} H_{uv}(n)e^{-\delta n}\right]. \tag{3.31}$$

Our new quantity $H_{uv}(n)$ denotes the probability that a random walker, starting in u, arrives in v for the first time after exactly n steps. The hitting probability has been studied in a variety of different contexts that involve memoryless fluctuations, ranging from chemical reaction, electric circuits to the triggering of stock options [165, 203–205].

From a conceptual point of view it is interesting to note that the path length n can interpreted as a random variable with a corresponding probability distribution $H_{uv}(n)$. Then $\tilde{H}_{uv}(c) = \sum_{n=0}^{\infty} H_{uv}(n)e^{-cn}$ denotes the discrete Laplace transform of $H_{uv}(n)$. By definition, we can formally identify the RWED with the cumulant

generating function $\log \tilde{H}_{uv}(\delta)$ evaluated at $\delta = -\log(\lambda/\beta) + \gamma$ (see Eq. (3.13)):

$$D_{uv}^{\mathrm{RW}} = -\log \tilde{H}_{uv}(\delta) \tag{3.32a}$$

$$\tilde{H}_{uv}(\delta) = \sum_{n=0}^{\infty} H_{uv}(n) e^{-\delta n}. \tag{3.32b}$$

This surprising result highlights the close relationship between the expected infection arrival time and the cumulant generating function of the hitting probability, i.e. $\langle t_{uv} \rangle \propto \log \tilde{H}_{uv}(\delta)$. Also, it establishes a link between the complex reaction-diffusion dynamics in Eqs. (3.8) on the one hand and a generic random walk process on the other.

The seemingly minor change in convention from paths (see Eq. (3.24)) to walks (see Eq. (3.29)) defining the MPED and the RWED, respectively, appears to complicate the computation at first glance: Instead of summing over an exponentially growing number of paths we propose to consider an infinite number of walks. The main advantage, however, is that we can treat epidemic spreading as a memory-less jump process and thus leverage the theory of random-walks to derive an efficient algorithm for Eq. (3.29).

3.6 Computation of the RWED

From a computational perspective, the crucial advantage of our Markovian approximation is that we can now calculate the hitting probability $H_{uv}(n)$ and hence Eq. (3.29) efficiently using a matrix representation. For that purpose, we first introduce the transition matrix $\boldsymbol{T}$ with elements T_{ij} as defined in Eq. (3.4). Then, the probability P_{uv} that a random walker starting in node u arrives after exactly n steps at node v is determined by the n-th power of the transition matrix:

$$P_{uv} = [\boldsymbol{T}^n]_{uv}\,. \tag{3.33}$$

As a next step, we return to the hitting probability $H_{uv}(n)$, which we defined in Eq. (3.31) via *absorbing* random-walks. Here, we use an equivalent definition using the following recursion rule [165, 168]:

$$P_{uv}(n) = \sum_{k=0}^{n} H_{uv}(k) P_{vv}(n-k). \tag{3.34}$$

The idea is to determine the probability that a walker starting in u, arrives in v after exactly n steps, though not necessarily for the first time. In this case, the walker may have first reached node v at an earlier time step k with (hitting) probability $H_{uv}(k)$ and then returned to the target node v with the probability $P_{vv}(n-k)$ in the remaining $n-k$ steps. For $n = 0$ and $v \neq u$, we use the convention $H_{uv}(0) = 0$ and $P_{vv}(0) = 1$. Note also, that we are not considering here the initial condition $H_{uu}(0)$ with $v = u$ because, the estimated spreading time

from the outbreak location u to itself is not well-defined. Instead, we will discuss this case separately, later.

To solve the recursive relation in Eq. (3.34) for H_{uv}, we use the discrete Laplace transform (similar to the generating function formalism). For that, we multiply Eq. (3.34) with e^{cn} and sum over n:

$$\sum_{n=0}^{\infty} P_{uv}(n)e^{cn} = \sum_{n=0}^{\infty} e^{cn} \sum_{k=1}^{n} H_{uv}(k)P_{vv}(n-k) \tag{3.35a}$$

$$= \sum_{k=0}^{\infty} H_{uv}(k)e^{ck} \sum_{n=k}^{\infty} P_{vv}(n-k)e^{c(n-k)} \tag{3.35b}$$

$$= \sum_{k=0}^{\infty} H_{uv}(k)e^{ck} \sum_{k'=0}^{\infty} P_{vv}(k')e^{ck'}. \tag{3.35c}$$

In Eq. (3.35b) we swapped the order of summation and renamed in Eq. (3.35c) the index to $k' = n - k$. Next we identify the Laplace transformed probability distributions $\tilde{P}_{uv}(c) = \sum_{k=0}^{\infty} P_{uv}(n)e^{ck}$ and $\tilde{H}_{uv}(c) = \sum_{k=0}^{\infty} H_{uv}(n)e^{ck}$:

$$\tilde{P}_{uv}(c) = \tilde{H}_{uv}(c)\tilde{P}_{vv}(c). \tag{3.36}$$

The steps from Eqs.(3.35) to Eq. (3.36) reveal a central property of the Laplace transformation, namely that a convolution of probabilities turns to a product in Laplace space.

We remember from Eq. (3.32a) that the logarithm of the Laplace-transformed hitting probability $\tilde{H}_{uv}(c)$ equals the RWED when evaluated at $c \equiv \delta$. We can calculate the system parameter δ from the (early) epidemic growth and mobility rate $\lambda = \beta - \mu$ and γ, respectively (see Eq. (3.13)). With Eq. (3.36), we can now determine the RWED as follows:

$$D_{uv}^{\mathrm{RW}} = -\log \tilde{H}_{uv}(\delta) \tag{3.37a}$$

$$= -\log\left(\frac{\tilde{P}_{uv}(\delta)}{\tilde{P}_{vv}(\delta)}\right). \tag{3.37b}$$

In order to calculate efficiently the RWED for all (u, v) pairs we vectorize Eq. (3.37) in the next step. To this end, we use Eq. (3.33) to construct a matrix $\tilde{\boldsymbol{P}}(\delta)$, where each element $\tilde{P}_{uv}(\delta)$ is a Laplace-transform of the sampling probability P_{uv} evaluated at $c = \delta$:

$$\tilde{\boldsymbol{P}}(\delta) = \sum_{n=0}^{\infty} \boldsymbol{T}^n e^{-\delta n} \tag{3.38a}$$

$$= \sum_{n=0}^{\infty} \left[\boldsymbol{T}e^{-\delta}\right]^n \tag{3.38b}$$

Now, it is important to note that matrix $\boldsymbol{T}e^{-\delta}$ is sub-stochastic for $\delta > 0$. In other words, all matrix elements are non-negative and because of the factor $e^{-\delta} < 1$

the row sum satisfies the inequality $\sum_t T_{uv} e^{-\delta} < 1$ for every row $u = 1..N$. From this property it follows that the dominant eigenvalue is smaller than 1, so that the limit of the geometric sum in Eq. (3.38b) exists and equals the *fundamental matrix* $\tilde{\boldsymbol{P}}(\delta) = \boldsymbol{Z}$ [206] with:

$$\boldsymbol{Z} = \left[\mathbb{1} - \boldsymbol{T} e^{-\delta}\right]^{-1} \tag{3.39}$$

Here, $\mathbb{1}$ denotes a $N \times N$ identity matrix. Next we define the diagonal matrix

$$\mathrm{Diag}(\boldsymbol{Z})^{-1} = \begin{cases} 1/Z_{vv} & \text{for diagonal elements } v = 1, \ldots, N \\ 0 & \text{else} \end{cases} \tag{3.40}$$

and thus reformulate the fraction $\tilde{P}_{uv}/\tilde{P}_{vv}$ from Eq. (3.37b) in terms of the fundamental matrix $\boldsymbol{Z}$:

$$\frac{\tilde{P}_{uv}(\delta)}{\tilde{P}_{vv}(\delta)} = \left[\boldsymbol{Z} \cdot \mathrm{Diag}(\boldsymbol{Z})^{-1}\right]_{uv} \tag{3.41}$$

Finally, we define the logarithm as an element-wise operation and thus obtain the RWED from Eq. (3.37) for any (u, v) pair as an element of the matrix

$$\boldsymbol{D}^{\mathrm{RW}} = -\log\left[\boldsymbol{Z} \cdot \mathrm{Diag}(\boldsymbol{Z})^{-1}\right]. \tag{3.42}$$

With Eq. (3.42), we finally derived an efficient method to obtain the RWED for all (u, v) pairs. From this vectorized form, we can estimate the mean transmission time $\langle t_{uv} \rangle$ for every outbreak location u and target node v. The pseudo-code in Alg. 1 summarizes the numerical steps.

Algorithm 1 Calculation of the RWED.

Require: Undirected connected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with $\mathcal{N}$, $\mathcal{E}$ and $\{F_{ij}\}$ denoting the set of nodes, edge and edge weights, respectively. The infection (β), recovery (μ) and mobility rate (γ), determine the system parameter (δ) according to Eq. (3.13).

1: **procedure** ALL_PAIRS_RWED($\{F_{ij}\}, \delta$)

2: $\boldsymbol{T} \leftarrow \begin{cases} F_{ij} / \sum_{j'} F_{ij'} & \text{if direct flight } i \to j \text{ exists} \\ 0 & \text{else} \end{cases}$ ▷ transition matrix (see Eq. (3.4))

3: $\boldsymbol{Z} \leftarrow \left[\mathbb{1} - \boldsymbol{T} e^{-\delta}\right]^{-1}$ ▷ fundamental matrix (see Eq. (3.39))

4: $\boldsymbol{D}^{\mathrm{RW}} \leftarrow \mathbb{0}$ ▷ Initialize RWED matrix with zeros

5: **for** $(u, v) \in \mathcal{E}$ **do**

$D^{\mathrm{RW}}_{uv} \leftarrow \log\left[\frac{Z_{uv}}{Z_{vv}}\right]$ ▷ see Eq. (3.42)

6: **return** $\boldsymbol{D}^{\mathrm{RW}}$

Step 3 in Alg. 1 requires a single matrix inversion to determine the fundamental matrix $\boldsymbol{Z}$. This operation poses the computational bottleneck. Using the Coppersmith-Winograd algorithm, we find a time complexity that scales with $\mathcal{O}(N^{2.4})$. Therefore, the numerical effort to compute all effective distances is only

slightly above the DPED with $\mathcal{O}(N^2 \log(N))$, where we account for the dominant transmission path only. This minor increase in complexity allows us to estimate the arrival time of an epidemic with significantly higher accuracy, as we will examine in the following section. We also note, that the computation of the RWED in this section improves the derivation proposed in our publication [115] that required N matrix inversion operations.

3.7 Numerical comparison

We begin the numerical section by reviewing the illustrative example in Fig. 3.9. In addition to the DPED and the MPED (see Eq. (3.25) and Eq. (3.24), respectively), we use the newly introduced RWED from Eq. (3.42) in Fig. 3.12 A to estimate the infection arrival time from GRU to BTS. This particular example shows a characteristic observation, typical for most pairs of source and target nodes: As we increase the number of paths entering the calculation of the MPED, the estimated infection arrival time approaches the RWED. This trend becomes particularly apparent when considering the relative deviation to the RWED in Fig 3.12 B on a semi-logarithmic scale. Hence, the RWED gives a tight lower bound to the MPED, which justifies our Markovian assumption (A6), i.e., to consider transmission paths where nodes can be infected multiple times.

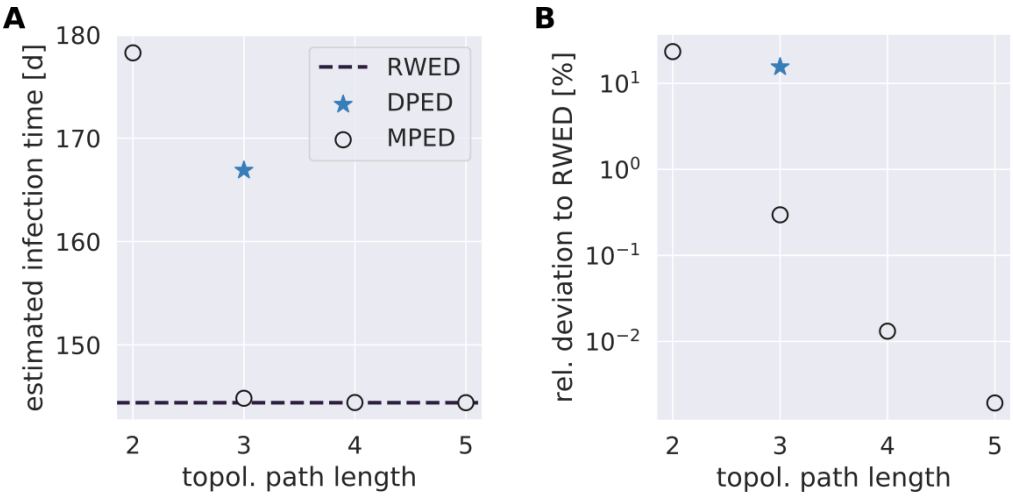

Figure 3.12: Impact of competing transmission paths for an exemplary transmission from GRU to BTS as in Fig. 3.12. **A**: Estimated infection arrival times using the DPED (blue star, Eq. (3.25)), the MPED (circles, Eq. (3.24)), and the RWED (dashed line, Eq. (3.42)). We estimate the MPED using a cutoff $n = 2, \ldots, 5$ for the maximum topological path length. The dominant transmission path has a topological length of $n = 3$. **B**: Deviation of the DPED and MPED from the RWED in percent and semi-log scale. Parameters as in Fig. 3.3.

More importantly, however, is to compare the analytical estimations with numerically determined results. To this end, we integrate the meta-population model from Eq. (3.8) and thus simulate an outbreak that starts in GRU with the same parameters as in Fig. 3.3. The numerically determined infection arrival times to the remaining nodes, serve as reference values in Fig. 3.13. Here, we compare

the correlation between the simulated and estimated infection arrival times using the RWED from Eq. (3.42). The size of an airport in terms of passenger numbers determines the marker size and from this we observe that global hubs are among the first infected nodes. Moreover, the infection arrival time is clearly correlated with the topological distance as we observed already in Fig. 3.4 A. From the best fit (black line), we determine Pearson's correlation coefficient to $r = 0.99$ and thus improve the previous result using the DPED where we found $r = 0.97$ in Fig. 3.11 C.

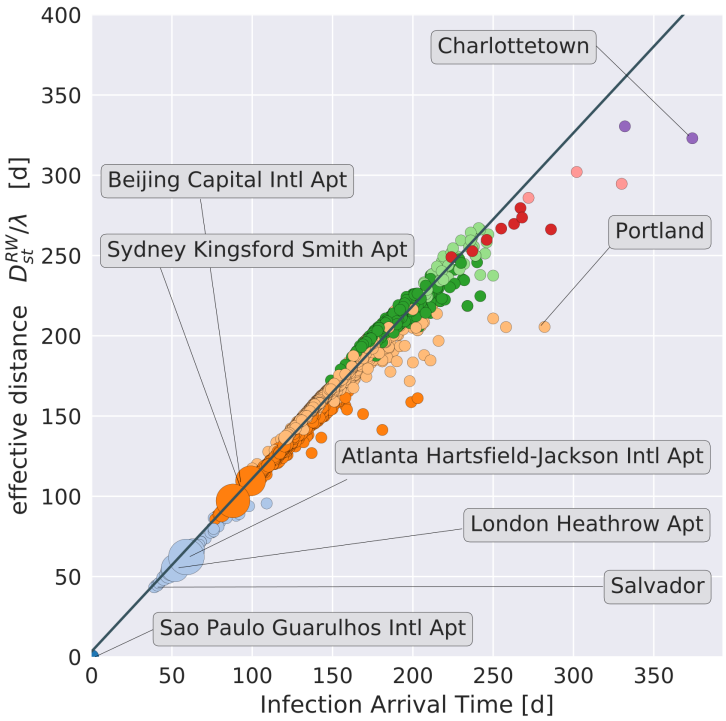

Figure 3.13: Correlation between simulated and estimated infection arrival times using the RWED $D^{\mathrm{RW}}_{uv}/\lambda$ from Eq. (3.42). Color indicate the topological distance to the outbreak origin (GRU), and marker size scales with the population size. From the best fit (black line), we determine the Pearson's correlation coefficient $r = 0.99$. Parameters as in Fig. 3.3.

Next, we compare the performance of our newly proposed RWED with the well-established DPED for different outbreak locations and parameters. To this end, we rely on Pearson'scorrelation coefficient r as a measure of quality. A value close to one indicates a high positive correlation between the numerical values and their analytical estimates. As an overall quality measure we propose to look at the distribution of Pearson'scoefficients for every initially infected population. The histogram in Fig. 3.14 A gives a comparison between the previously introduced DPED and the RWED, colored in blue and red, respectively. The distribution corresponding to the RWED is clearly shifted towards higher values in r as compared to the DPED. Therefore, the comparison shows that a measure which takes into account all walks as possible transmission paths outperforms the

previously introduced approach which relies only on the dominant transmission route.

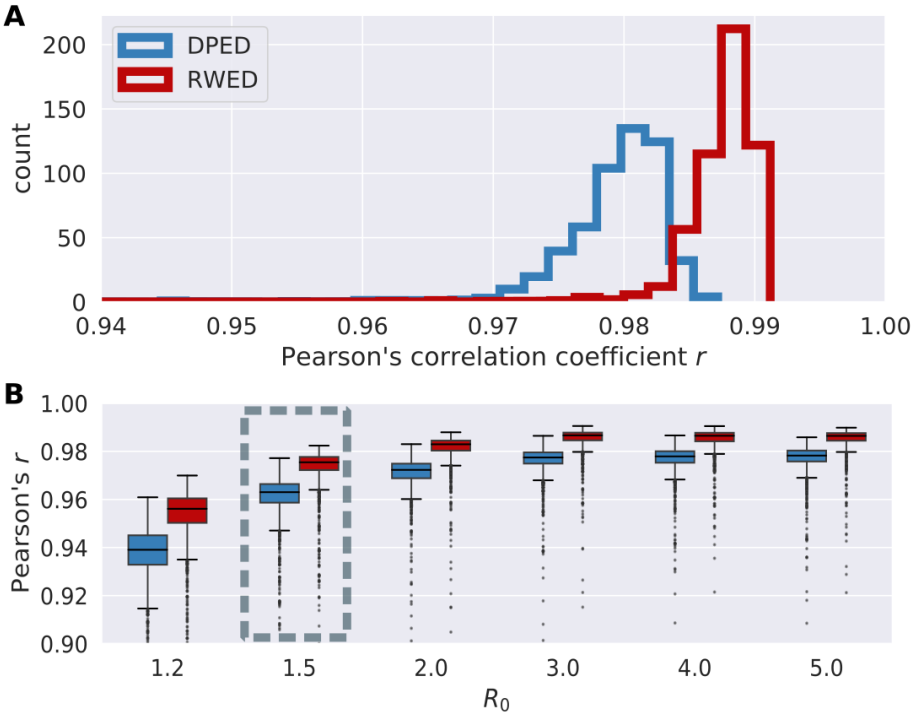

Figure 3.14: Correlation analysis for different parameters. **A**: Distribution of Pearson's correlation coefficient r for different outbreak locations. We compare the RWED and DPED colored in red and blue, respectively (see also Fig. 3.13 and Fig. 3.11 C with GRU as outbreak location). Simulation parameters as in Fig. 3.3. **B**: Boxplot for different values of $R_0 = \beta/\mu$ with $\mu = 0.285$ as before. Boxes mark first and third quartiles and horizontal lines indicate the median. Whiskers extend to 1.5 times the inter-quartile range while dots mark outliers. Color code as in (panel A). Dashed rectangle marks the boxplot corresponding to (panel A).

We also investigate in Fig. 3.14 B the impact of the basic reproduction number $R_0 = \beta/\mu$ on the performance. Given a fixed recovery rate μ, higher values of R_0 increase the early growth rate $\lambda = \beta - \mu = (R_0 - 1)\mu$ and thus the velocity with which the disease spreads across the network. The box-plot in Fig. 3.14 B indicates that correlation tends to increase with higher values of R_0. Also, the difference between the RWED and the DPED decreases, which, in part, is a trivial consequence of correlations approach the upper limit of $r = 1$. However, the difference remains significant, even for fast spreading epidemics.

We can explain the tendencies observed in Fig. 3.14 B as follows: The system parameter δ increases with R_0 and thus paths with a short topological length $n(\pi)$ dominate the contribution to the RWED according to Eq. (3.29). Hence, the RWED reduces to the DPED in the limit of large δ. Also, for sufficiently large values of R_0 the epidemic spreads predominantly along the shortest path tree as visualized in Fig. 3.10 B. The high accuracy of our approach for tree topologies, as demonstrated in Fig. 3.7, thus explains the high correlation.

Finally, we also compare in Fig. 3.15 the performance on various topologies and demonstrate the correlation between estimated and simulated infection arrival

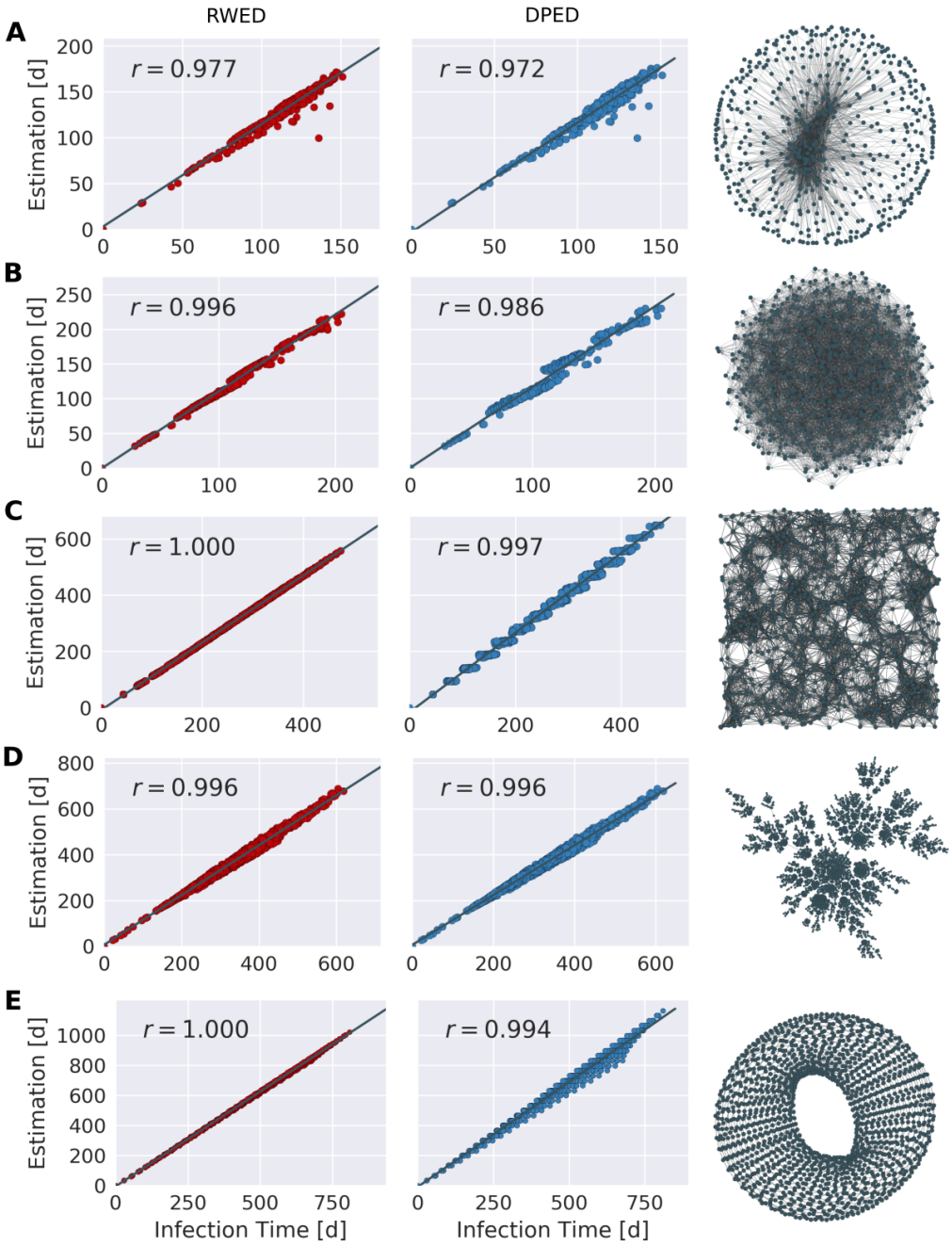

Figure 3.15: Correlation between simulated and estimated infection arrival times for different graphs and a randomly chosen outbreak location. Estimation according to the RWED (left column, red) and DPED (middle column, blue). Best fit (black line) gives Pearson's correlation coefficient r (see value in each subfigure). Right column shows the corresponding graph. **A**: U.S. 500 largest airport network [4]. **B**: Erdős-Réyni graph, randomisation of (A). **C**: Geometric graph with 1000 nodes and connectivity radius of 0.5. **D**: Barabási-Albert network with 5000 nodes and connectivity parameter $m = 1$. Edge weights are drawn from the global air-traffic network. **E** Lattice with 50×30 nodes and periodic boundary condition.. Parameters as in Fig. 3.3.

times for a randomly chosen outbreak location. As the first example in Fig. 3.15 A, we take the weighted network of 500 largest airports from [4] and obtain a similarly high correlation as in Fig. 3.14. We also find robust results for several artificial networks, such as the Erdős-Réyni graph [39] in Fig. 3.15 B, which we obtain by randomizing the edges of the previous graph in Fig. 3.15 A. Another widely used class of graphs are random geometric networks. For the single realisation in Fig. 3.15 C, we place 1000 nodes randomly on a 1×1 square and connect them if the Euclidean distance is smaller than 0.1. Next, we also look at a Barabái-Albert network [207] with 5000 nodes in Fig. 3.15 C. We construct the graph dynamically by adding a single node at a time and connecting it to m other nodes, chosen with a probability proportional to their degree. With $m = 1$ in Fig. 3.15 D, we thus obtain a tree graph with a broad distribution of degrees. Finally, we compare the DPED and RWED on a regular 50×30 lattice with periodic boundary conditions in Fig. 3.15 E.

Our numerical results demonstrate that our newly proposed RWED is an accurate proxy for the mean infection arrival time. By taking into account competing paths for epidemics, this network-based measure outperforms consistently the previously introduced DPED, which relies only on the dominant transmission route. Our results are robust for different parameters and several empirical as well as synthetic networks.

3.8 Summary

In this chapter, we have studied the spread of global diseases using a metapopulation model, where airports and flight connections form nodes and edges, respectively, of a global mobility network. This type of reaction-diffusion model connects susceptible-infected-recovered spreading dynamics within well-mixed populations and diffusive transport between them. It serves as a numerical reference model and forms the basis of our analytical work, in which we determine the expected time until the first infected individual reaches a distant population. In Sec. 3.4, we adopted the work of [114] and derived the dominant and multiple path effective distance, denoted DPED and MPED, respectively. These network-based measures can be used as proxies for the mean infection arrival time. The detailed analysis in Sec. 3.4 served multiple purposes:

(i) It is a critical review of the two main publications that provide the basis for further contributions. In examining the two independent works [5, 114] under the same assumptions, we provided a deeper understanding of both approaches. On the one hand, we gave a model-based derivation of effective distances in addition to the heuristic definition in [5]. On the other hand, we simplified the main results from [114].

(ii) In Sec. 3.4, we also outlined and motivated the underlying assumptions that are necessary to introduce the random-walk effective distance (RWED) in the next section. To this end, we analysed simple topologies with increasing

complexity. Thereby, we demonstrated the limits and accuracy of our assumptions.

(iii) Finally, we motivated the need to improve existing methods. Here, we noted that the MPED in Eq. (3.24) is only applicable to small networks due to its computational complexity. Furthermore, the DPED, though highly efficient, is an approximation that applies only to strongly heterogeneous topologies and a sufficiently fast disease propagation.

To overcome the limitations of current approaches, we proposed in Sec. 3.5 the RWED. Unlike the DPED that ignores all but the dominant transmission route, the RWED accounts for an arbitrary large number of competing paths. Leveraging the theory of Markov chains in Sec. 3.6, we proposed a computationally efficient method to determine the RWED between all pairs of nodes from a single matrix inversion operation. Finally, we compared in Sec. 3.7 the analytical estimations with numerically determined infection arrival times. Thereby, we demonstrated that the RWED outperforms the DPED consistently for all potential outbreak locations and a wide range of parameters. We also explored additional empirical and artificial networks with very different properties. In all cases, it is beneficial to account for competing transmission routes using the RWED.

Conceptually interesting, this work has shown that the complex dynamics of a meta-population model can be related to a generic diffusion model thus bridging the gap between random-walk theory and contagion dynamics. In particular, we found that the infection arrival time can be approximated by the cumulant generating function of the first-passage probability. This interpretation points to an even deeper relationship with statistical physics, where many extensive quantities such as particle number, energy and entropy (to name only a few) are related to cumulants of random variables.

The relation to an abstract underlying mechanism is subject of our current research and could potentially extend the scope of the RWED another important class of epidemic spreading models, i.e., the *contact processes*. We present the first steps towards a general theory of effective distances in Appendix A. Also, we expect that similar, network-based measures can be derived for multi-layer networks and graphs with time-varying topologies.

4

Matrix based approach to epidemic modeling

Abstract. We extend the concept of accessibility in temporal networks to model infections with a finite infectious period such as the susceptible-infected-recovered (SIR) model. This approach is entirely based on elementary matrix operations and unifies the disease and network dynamics within one algebraic framework. We demonstrate the potential of this formalism for three examples of networks with high temporal resolution: networks of social contacts, sexual contacts, and livestock-trade. Our investigations provide a new methodological framework that can be used, for instance, to estimate the epidemic threshold, a quantity that determines disease parameters, for which a large-scale outbreak can be expected.

4.1 Introduction

In the preceeding Chap. 3, we studied epidemic outbreaks based on the global airtraffic network. This example of a mobility graph can be considered as static with respect to the disease dynamics. Ignoring the temporal variability of contacts can, however, lead to severe deviations. In particular, a major drawback arises from the fact that the number of paths relevant for the disease spreading can be strongly overestimated as we demonstrated in our analysis of the German pig trade network (see Sec. 2.2.1). The timing of links between nodes matters, especially when the network evolves on a similar time scale as the spreading dynamics, which led to an increasing interest in the interplay between disease and network dynamics [72, 79, 104, 159, 208, 209]. To this end, numerous studies emphasized in recent years the importance of contact data with high temporal resolution, so called *temporal networks* to improve existing models of epidemic spreading [151–154, 210, 211]. However, many methods from static network theory cannot be generelized

to time-varying topologies and therefore, it remains, to a large extend, an open problem to integrate epidemiological models with time-resolved data.

The challenge is to consider only transmission paths, with links that appear in the correct temporal order and a natural way to include the causal sequence of edges into the analysis is to consider accessibility graphs (see Fig. 2.5). Here, a directed edge between two nodes appears only if one is accessible by the other through a time-respecting path. In other words, an accessibility graph provides information whether a temporal path connects two nodes within a certain time period (see Eq. (2.27)). In order to calculate the corresponding accessibility matrix, we introduced in Eq. (2.28) an analytical approach termed *unfolding accessibility* [71, 73]. This framework is based on a list of adjacency matrices, whose elements refer to the snapshots of the temporal network under consideration. From an epidemiological perspective, the approach models a susceptible-infected (SI) type of disease, i.e. an infection without recovery. Here, each contact leads to newly infected neighbours in a deterministic way, thus, serving as a worst-case scenario.

In this chapter, we generalize the concept of accessibility and introduce a fixed recovery time. This allows us to include diseases with a finite infection period, and we will consider the widely used susceptible-infected-recovered (SIR) model. As in the previous work [71], we base our formalism on elementary matrix operations with Boolean arithmetic, which allows us to integrate the disease and network dynamics in an unified framework.

From the modelling perspective our formalism gives us the advantage that we take into account the full temporal and topological complexity of the interaction patterns, without restricting ourselves to limiting assumptions such as Markovian dynamics [212], pair approximations [93], individual-based [86] and degree-based mean-field theory [136]. This is a consequence of our approach to use time-ordered products of exact adjacency matrices, which preserve the causal appearances of links. Furthermore, the concept of accessibility matrices allows us to track the individual infection status. Thus, no dynamic approximation on the number of contacts between susceptible and infected, which are crucial for disease transmission, is needed.

The remaining four sections of the chapter are structured as follows. First, we introduce the terminology and main concepts in Sec. 4.2. Then, we derive the epidemiological model in Sec. 4.3. Thereby, we begin with a state vector formalism, which we will then generalize to an entirely matrix-based approach. In order to demonstrate the our formalism, we analyse in Sec. 4.4 disease dynamics on different empirical networks: networks of social contacts, sexual contacts, and livestock-trade. As a particularly important application, we also estimate the critical infectious period that indicates a transition from minor to global outbreaks. Finally, we summarize our results in Sec. 4.5.

This chapter is based on our publication [72].

4.2 Materials and Methods

Unlike the previous Chap. 3, we focus now on a *discrete time* susceptible-infected-recovered (SIR) model, where nodes of a contact network form *single epidemiological units*. In other words, a node can be either in the susceptible, infected or, recovered state denoted with S, I and R, respectively. Moreover, a susceptible node can contract the disease with a constant (per time-step) probability β from an infected neighbor.

Another important difference to the previous meta-population model in Chap. 3, we will consider a non-Markovian recovery process. Our motivation is that many diseases lead to a roughly constant duration of infection [18] and therefore, our simplified model assumes that an infected node recovers deterministically after τ time steps. With this, we can write down the following dynamic system for the fraction $\rho^X(t)$ of individuals in compartment $X \in \{S, I, R\}$:

$$\rho^S(t+1) = \rho^S(t) - \beta\rho^{SI}(t) \tag{4.1a}$$

$$\rho^I(t+1) = \rho^I(t) + \beta\rho^{SI}(t) - \rho^I(t-\tau) \tag{4.1b}$$

$$\rho^S(t+1) = \rho^R(t) + \rho^I(t-\tau). \tag{4.1c}$$

Here, the interaction term $\rho^{SI}(t)$ denotes the expected number of new incidences as a fraction of the population size. Furthermore, we note that Eqs. (4.1) are subject to the conservation condition $\rho^S(t)+\rho^I(t)+\rho^R(t) = 1$ for all times $t = 0, \ldots, T-1$. Starting from this set of equations, a number of approaches have been proposed to approximate the crucial term $\rho^{SI}(t)$, which reflects the fraction of contacts between susceptible and infected individuals and thus includes the underlying topological structure [213]. In this chapter we will introduce a similar scheme and identify the critical contacts $\rho^{SI}(t)$ directly from the temporal network. Equations (4.1a) to (4.1c), which describe the widely used SIR model, serve as a reference before elaborating on our matrix based alternative.

The interactions in a static network are typically represented by an adjacency matrix $\boldsymbol{A}$ with a positive entry $A_{ij} = 1$ if node j is connected to i and $A_{ij} = 0$ otherwise. A contact between two nodes appears as a link in the network at the given time and can be either undirected, which is naturally the case for proximity data [151, 214] or, directed if trade is considered for example [215]. In this chapter we will allow for directed links and therefore we will assume asymmetric adjacency matrices in general.

A temporal network can be defined in a variety of ways, depending on the underlying contact data and its purpose [211]. We adapt a widely used approach, which assumes that the temporal network evolves in equidistant time steps, given by the sampling resolution of the data. It allows us to define the temporal network conveniently through a sequence of adjacency matrices

$$\mathcal{A} = [\boldsymbol{A}(0), \boldsymbol{A}(1), \ldots, \boldsymbol{A}(T-1)]. \tag{4.2}$$

Here, each element $\boldsymbol{A}(t)$ describes a snapshot of interactions at time $t = 0, \ldots, T-1$, with T being the observation time (see Sec. 2.2 for details). This is a valid approximation as long as the time to form a link, e.g. the latency in case of transport connections, can be neglected compared to the sampling time [211]. Furthermore, we will use the highest resolution available throughout the chapter in order to minimize potential errors. This happens due to the fact, that causal appearances of links can be violated if snapshots are partially aggregated. As a result, potential outbreaks may be systematically overestimated [71].

If we assume that the pathogen can only be passed to the nearest neighbours within every snapshot, then a temporal network allows us to infer the fraction of critical contacts ρ^{SI} (Eq. (4.1b)) directly from the disease status of every node. The assumption can be justified if the temporal resolution prohibits consecutive interactions within one time step and is an approximation in the case of partially aggregated data. Therefore, we can use the time-depending topology in order to determine the disease dynamics directly.

4.3 Derivation of the epidemiological model

Our analysis extends the concept of accessibility as introduced in the Sec. 2.2. Here, we introduced a computationally efficient approach to calculate the accessibility matrix of a temporal network. Elements of this matrix indicate if a temporal path exists that leads from the outbreak location to a given target node. From an epidemiological perspective, the target node can thus potentially contract a disease. However, this approach does not account for a limited infectious period, which significantly alters the dynamic. Therefore, we will extend in this section the concept of accessibility and begin for didactic reasons with a state vector formalism, before we introduce a matrix-based approach akin to the *unfolding accessibility* in Eq. 2.28.

4.3.1 State Vector Formalism

We begin with a definition of two boolean state vectors $\boldsymbol{i}(t)$ and $\boldsymbol{r}(t)$, with elements indicating if the corresponding nodes are infected or recovered at time t, respectively:

$$[\boldsymbol{i}]_k(t) = \begin{cases} 1, & \text{node } k \text{ is infected at time } t \\ 0, & \text{otherwise} \end{cases} \tag{4.3a}$$

$$[\boldsymbol{r}]_k(t) = \begin{cases} 1, & \text{node } k \text{ is recovered at time } t \\ 0, & \text{otherwise.} \end{cases} \tag{4.3b}$$

Following the binary nature of the state vectors, we will only use Boolean arithmetic [216] without explicitly indicating it. Hence, element-wise addition and scalar multiplication are replaced by the logical "or" and "and", respectively. Matrix multiplication is handled similarly through the definition $(\boldsymbol{X} \cdot \boldsymbol{Y})_{ij} = \sum_k X_{ik} Y_{kj}$

and element-wise multiplication $\circ$ is defined by $(\boldsymbol{X} \circ \boldsymbol{Y})_{ij} = X_{ij}Y_{ij}$. Finally, we introduce the element-wise negation $\neg$ of a vector or a matrix, i.e. $(\neg \boldsymbol{X})_{ij} = 0$ if $X_{ij} = 1$ and $(\neg \boldsymbol{X})_{ij} = 1$ otherwise.

Furthermore, we will consider the worst-case scenario, where the disease is transmitted upon contact, i.e. $\beta = 1$ in Eq. (4.1b). It is a convenient way to separate topological and probabilistic fluctuations in a temporal network and to concentrate on the first. This assumption, is not crucial and we will sketch now one possible generalization to an arbitrary $\beta \leq 1$.

If an infected node does not transmit the disease to its susceptible neighbour with a probability $1 - \beta$, we can just as well neglect the link with the same probability. This idea can be formalized by a stochastic operator $\Omega(\boldsymbol{A})$, which acts element-wise on the adjacency matrix $\boldsymbol{A}$: For $A_{ij} = 0$ we define $\Omega(A_{ij}) = 0$ and for positive entries $A_{ij} = 1$ we choose $\Omega(A_{ij}) = 1$ with probability $\beta \leq 1$ and $\Omega(A_{ij}) = 0$ otherwise. In order to obtain one realization of the stochastic process, the original sequence of adjacency matrices $\mathcal{A}$ (see Eq. (4.2)) would have to be replaced by the thinned out modification

$$\widetilde{\mathcal{A}} = [\Omega(\boldsymbol{A}(0)), \Omega(\boldsymbol{A}(1)), \ldots, \Omega(\boldsymbol{A}(T-1))]. \tag{4.4}$$

This approach inherits from the simplified model with $\beta = 1$ the advantage to account for dynamic and topological correlations. A potentially large number of realizations increases, however, the computational effort and mixes topological with probabilistic effects. In order to focus on the former we will therefore exclude stochastic fluctuations and choose $\beta = 1$. Using the explicit contact structure given by the set of static snapshots $\mathcal{A}$ (see Eq. (4.2)), we propose the following model for the SIR dynamics:

$$\boldsymbol{i}(t+1) = [\boldsymbol{A}(t)\boldsymbol{i}(t) + \boldsymbol{i}(t)] \circ [\neg\, \boldsymbol{r}(t+1)] \tag{4.5a}$$

$$\boldsymbol{r}(t+1) = \sum_{t'=0}^{t+1-\tau} \boldsymbol{i}(t'). \tag{4.5b}$$

The number of newly infected individuals is given by the product $\boldsymbol{A}(t)\boldsymbol{i}(t)$ in Eq. (4.5a), where $\boldsymbol{A}(t)$ is the $N \times N$ dimensional adjacency matrix of the temporal network at time $t = 0, 1, \ldots, T-1$. A node enters the recovered state $I \to R$ (Eq. (4.5b)) if it has been infectious for τ time steps. Therefore this model, contrary to the meta-population model in Chap. 3, is not Markovian, i.e. the state of the system depends not only on the previous, but on the last τ time steps. This is a frequently used simplification [131, 144], which leads to a very efficient implementation and is arguably a better approximation to realistic diseases than a Markovian ansatz. Finally, the element-wise operation $\circ[\neg\, \boldsymbol{r}(t+1)]$ guarantees that a recovered node cannot be reinfected or pass a disease.

The system of equations (4.5) describes the disease dynamics, which starts with arbitrary initial conditions $\boldsymbol{i}(0)$ and $\boldsymbol{r}(0)$. From a conceptual and practical point of view, it is advisable to generalize the state vector to a matrix formalism

that accounts for different initial conditions at the same time. This will be subject of the next section.

4.3.2 Tracking All Initial Conditions: Matrix Formalism

It is often important to evaluate the role of single nodes in a contact network with respect to their influence on the outbreak dynamics. For example, one could look for nodes with a high temporal out-component (virulence) [70] or nodes with a high probability of catching a disease (vulnerability) [31]. This consideration motivates us to choose the initial conditions as $\boldsymbol{i}(0) \equiv \boldsymbol{e}^{(k)}$ and $\boldsymbol{r}(0) = \boldsymbol{0}$ indicating that only the node k is infected and all other nodes are susceptible. Here, $\boldsymbol{e}^{(k)}$ denotes the k-th column of the identity matrix, where $(\boldsymbol{e}^{(k)})_i = 1$, if $i = k$ and 0 otherwise. Instead of repeatedly considering Eqs. (4.5) for every $\boldsymbol{e}^{(k)}$ with $k = 1, \ldots, N$, it will prove useful to stack all initial vectors into $N \times N$ matrices:

$$\boldsymbol{I}(t=0) = \left[\boldsymbol{e}^{(1)}, \boldsymbol{e}^{(2)}, \ldots, \boldsymbol{e}^{(N)}\right] = \mathbb{1}_N \tag{4.6a}$$

$$\boldsymbol{R}(t=0) = [\boldsymbol{0}, \boldsymbol{0}, \ldots, \boldsymbol{0}] = \mathbb{0}_N. \tag{4.6b}$$

This approach allows us to convert the state-vector description into a matrix-based formalism, which enables us to consider all homogeneously sampled, initial conditions simultaneously:

$$\boldsymbol{I}(t+1) = [\boldsymbol{A}(t)\boldsymbol{I}(t) + \boldsymbol{I}(t)] \circ [\neg\, \boldsymbol{R}(t+1)] \tag{4.7a}$$

$$\boldsymbol{R}(t+1) = \sum_{t'=0}^{t+1-\tau} \boldsymbol{I}(t'), \tag{4.7b}$$

The kth column of the matrix $\boldsymbol{I}(t)$ indicates the nodes that are infected at time t if the epidemic starts from the initial node k. As we are working with Boolean arithmetic, we only have binary entries in $\boldsymbol{I}(t)$. Therefore, we can regard it as an adjacency matrix of an accessibility graph: A directed edge appears only if a time-respecting path exists that connects the initial node with the target and satisfies the epidemiological constraints: An infection can only be passed on within a fixed time window of size τ and a node cannot be reinfected.

In the special case, if the infectious period τ exceeds the observation time T, we are effectively dealing with an SI-type of infection. Then, we do not account for recovered nodes and thus all operations including the matrix R can be ignored. Note that the simplified expression of Eq. (4.7a) $\boldsymbol{I}(t+1) = \boldsymbol{A}(t)\boldsymbol{I}(t) + \boldsymbol{I}(t)$ can be solved directly

$$\boldsymbol{I}(t+1) = \prod_{t'=0}^{T-1} [\boldsymbol{A}(t') + \mathbb{1}], \tag{4.8}$$

which recovers a result that was independently presented in [71, 73] and that we introduced already in Eq. (2.28). The formalism described by Eq. (4.8) was termed *unfolding accessibility* as the corresponding accessibility graph accumulates edges, which point from the initial seed to all nodes accessible via a time-respecting

path within the given time frame. From a graph-theoretical point of view, adding a unity matrix to an adjacency matrix, i.e. $\boldsymbol{A}(t') + \mathbb{1}$ in Eq. (4.8), introduces a self loop to every node. It means that a path from a node back to itself is always possible. This circumstance mimics an infinite recovery time for disease propagation since a node can always be the source of a new transmission path. The unfolding accessibility itself is a special case of an earlier work [73], which is a temporal generalization of the Katz centrality [217] and therefore accounts for walks that visit a node repeatedly.

The matrix $\boldsymbol{I}(t)$ of Eq. (4.7a) provides a direct mapping from the seed of infection to the infected individuals. Therefore, we will also refer to it as the prevalence matrix. Given $\boldsymbol{I}(t)$, we can derive the incidence matrix $\boldsymbol{J}(t)$, which links the source nodes to the newly infected ones. Similarly, we introduce the cumulative incidence matrix $\boldsymbol{C}(t)$, which maps the seed of infection to all nodes that have been infected up to the observation time t.

$$\boldsymbol{J}(t+1) = \boldsymbol{I}(t+1) \circ \neg \boldsymbol{I}(t) \tag{4.9a}$$

$$\boldsymbol{C}(t+1) = \sum_{t'=0}^{t+1} \boldsymbol{I}(t'). \tag{4.9b}$$

In order to visualize the results we will use the density ρ of a matrix. It is defined by the ratio between the number of non-zero elements and the squared size of the matrix N^2. Note that the three scalar values $\rho(\boldsymbol{I}(t))$, $\rho(\boldsymbol{J}(t))$ and $\rho(\boldsymbol{C}(t))$ reflect the mean prevalence, incidence and cumulative incidence, respectively. That is, we automatically average over the ensemble of all homogeneously sampled initial conditions or in other words, we implicitly consider every node as a seed of infection. Similarly, we define ρ_k to be the ratio between the number of non-zero elements in the k-th column and the network size N. Thus, ρ_k is the fraction of infected individuals for one particular source node k.

From the topology of the three accessibility graphs, which we have defined in this section, we can directly analyse the dynamics of a disease. To demonstrate the feasibility of the developed matrix formalism, we will apply our approach in the next section to three real-world networks.

4.4 Application to Empirical Contact Networks

4.4.1 Social Contacts Network

As a first example, we consider a social contact graph [214] that has been recorded during a three-days conference. Each of the 113 participants corresponds to a node and as soon as a proximity sensor detected a face-to-face contact over a period of 20 s an edge was generated. As a result one obtains a sequence of adjacency matrices, which refers to the snapshots of the interaction dynamics with a temporal resolution of 20 s. The data on social contacts allows to analyse the spread of

airborne diseases as well as the propagation of information. With our model, we will hereby focus on the number of potential transmission paths.

Fig. 4.1 depicts an SIR process on the social contacts graph with a fixed recovery time of 20 hours. This choice is motivated by the characteristics of the dataset, i.e. a night break of several hours. Starting from one seed of infection, we observe a quick rise in prevalence $\rho(\boldsymbol{I}(t))$ (Fig. 4.1A) within only a few hours. This observation can be explained by a high number of changing interactions, which is the purpose of conferences. After around 21 hours the prevalence drops just as quickly, whereas one can observe two rather stable phases in between. These are reflected by plateaus in the cumulative incidence $\rho(\boldsymbol{C}(t))$ (Fig. 4.1B , dashed curve). They correspond to periods of vanishing incidence $\rho(\boldsymbol{J}(t))$ (Fig. 4.1B, blue bars) and are due to the fact that no interactions have been recorded during the night. The spreading process disappears before the third day of the conference. Despite the short infection time, about 85% of all nodes will be affected on average (table 4.1).

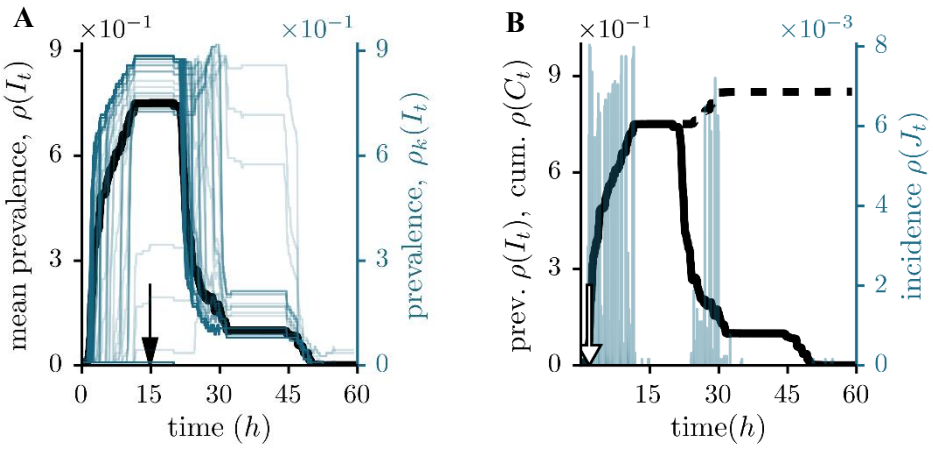

Figure 4.1: Prevalence, incidence and cumulative incidence for the social contacts network. (A) Comparison between the individual single source outbreaks (blue, right axes) and the corresponding, averaged prevalence (black, left axes). The infectious period is fixed at $\tau = 20$h. The arrow at $t = 14.8$h indicates the maximum averaged prevalence. (B) Mean prevalence $\rho(\boldsymbol{I}(t))$ (solid curves), incidence $\rho(\boldsymbol{J}(t))$ (blue bars, right scale) and cumulative incidence $\rho(\boldsymbol{C}(t))$ (dashed curves). Here, the arrow points at the maximum averaged incidence ($t = 1.8$h).

	$\rho\left(\boldsymbol{C}(T)^{\text{temporal}}\right)$	$\rho\left(\boldsymbol{C}(T)^{\text{static}}\right)$
social interactions	0.85	1.0
sexual contacts	0.027	0.88
livestock-trade	0.0048	0.15

Table 4.1: Comparison between the temporal networks and the time aggregated networks. We list the average fraction of individuals, which have been infected up to the observation time for the temporal networks ($\rho(\boldsymbol{C}(T)^{\text{temporal}})$) and the corresponding time aggregated graph ($\rho(\boldsymbol{C}(T)^{\text{static}})$). We use the same recovery times as in Figs. 4.1-4.3.

In order to compare the result we ran an equivalent experiment on the time aggregated network. We used the same disease parameter and assumed that the infection can spread only to the nearest neighbours within one time step. As a result, every node would be finally infected, independent of the seed of infection (table 4.1). The difference to the temporal network amounts to only 17 nodes on average. Therefore, with regard to the total impact of an outbreak a temporal representation of the data might not be necessary.

4.4.2 Sexual Contacts network

The result is rather different, if we analyse the dynamics on a sexual contacts network from a Brazilian escort website [151] (see Fig. 4.2). 16,730 participants have volunteered to record their sexual interactions online. The data spans the period from September 2002 to October 2008 and given the context of this dataset, allows to study the dynamics of sexually transmitted disease. We follow the numerical analysis of [218] and choose 91 days as the infectious period τ. It reflects the most contagious period after an HIV-1 infection, which is followed by a chronical stage with low transmission probability [219]. Furthermore, we ignore the first 1000 days in order to avoid transient effects during the early stage of the growing network [218].

We notice that the mean prevalence peaks after 305 days (Fig. 4.2, panel A), which is followed by a very slow decline. Moreover, we observe in Fig. 4.2B a considerable incidence throughout the observation period. The highest number of new infections is reached after 245 days, which is again a characteristic time that is much longer than the infectious period. Unlike the previous example, the temporal characteristics of the sexual contacts graph lead to a prolonged disease outbreak. This feature is mainly a consequence of the fact that the network is steadily growing [151] and an increasing number of interaction is recorded. The mean cumulative incidence shows that 2.7% of all nodes are affected at the end of the observation time. A comparison with the static network reveals that the total impact would be overestimated by more than one order of magnitude if we used aggregated data (table 4.1).

4.4.3 Livestock-Trade Network

Finally, we apply our formalism to an excerpt of the national German livestock database *HI-Tier* (WWW.HI-TIER.DE), which has been established according to EU legislation. It comprises the movement of pig in a period of 200 days from 2011-01-01 to 2011-07-20. With a time resolution of one day more than $9 \cdot 10^5$ trade transactions (directed links) have been recorded between 70,286 agricultural premises and traders (nodes), respectively. Livestock-trade is considered to be a major transmission route for animal-related diseases like classical swine fever [158]. In a potential outbreak, a detected node would be isolated from the trade network and therefore we can consider the infectious period to be mainly determined by

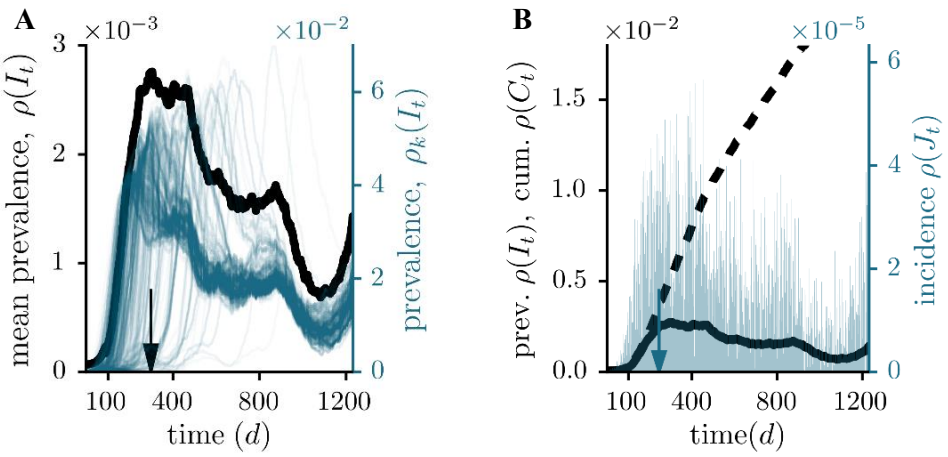

Figure 4.2: Prevalence, incidence and cumulative incidence for the sexual contacts network. (A) Comparison between the individual single source outbreaks (blue, right axes) and the corresponding, averaged prevalence (black, left axes). The infectious period is fixed at $\tau = 91$ d. The arrow indicates the maximum averaged prevalence. (B) Mean prevalence $\rho(\boldsymbol{I}(t))$ (solid curves), incidence $\rho(\boldsymbol{J}(t))$ (blue bars, right scale) and cumulative incidence $\rho(\boldsymbol{C}(t))$ (dashed curves). Here, the arrow points at the maximum averaged incidence.

the detection time. In our example shown in Fig. 4.3 we assume a detection time of 14 days.

Similar to the sexual network, we observe a broad evolution of the epidemic with a characteristic time scale of the mean prevalence (64 days) and incidence (60 days), respectively. We note again that both values are much larger than the infectious period, which indicates a slow mixing within the network. At the end of the observation, a fraction of 0.5% (around 340 nodes) would be affected. The corresponding static analysis would yield a result, which is again more than one order of magnitude higher (table 4.1).

These findings confirm previous observations [71, 218] that time-respecting paths can lead to a considerable improvement compared to static network analysis. It is of crucial importance to take into account the temporal sequence of the links in contact networks, especially, when the infection dynamics takes place on similar time scales as the network evolution.

4.4.4 Critical Infectious Period

In order to further analyse the impact of the infectious period τ, we examine the fraction of nodes, which have been infected once during the observation time, i.e. $\boldsymbol{C}(T)$. Again, we operate under the assumption that one node is infected at the beginning of the observation. Our formalism given by Eqs. (4.7) and (4.9b) allows us to calculate efficiently the cumulative incidence $\boldsymbol{C}(T)$ at the end of the observation time T for every initial condition in dependence on the disease parameter τ. This is depicted in Fig. 4.4 for the three considered contact networks.

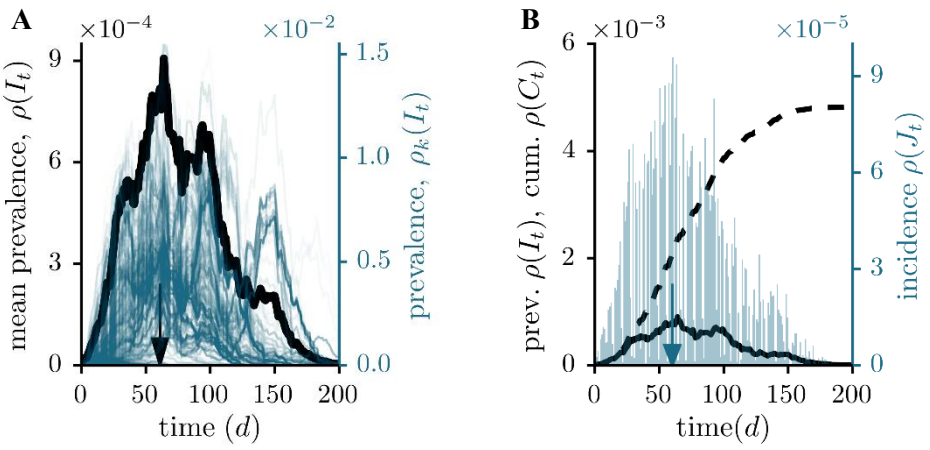

Figure 4.3: Prevalence, incidence and cumulative incidence for the livestock-trade network. (A) Comparison between the individual single source outbreaks (blue, right axes) and the corresponding, averaged prevalence (black, left axes). The infectious period is fixed at $\tau = 14$ d. The arrow indicates the maximum averaged prevalence. (B) Mean prevalence $\rho(\boldsymbol{I}(t))$ (solid curves), incidence $\rho(\boldsymbol{J}(t))$ (blue bars, right scale) and cumulative incidence $\rho(\boldsymbol{C}(t))$ (dashed curves). Here, the arrow points at the maximum averaged incidence.

Focusing on the small-τ range, we notice a transition, where the impact of the epidemic changes qualitatively from local to global outbreaks. This behaviour is well known in the literature [33, 86, 136] and indicates the existence of a critical infectious period τ_c or an epidemic threshold under the assumption that the transmission probability is one. It has been shown that the epidemic threshold for SIR infections is related to properties known from percolation theory [131, 144, 220, 221]. Here, a giant connected component emerges if the fraction of links exceeds a certain threshold [40, 145]. We estimate the critical transition point through a linear regression. This approach is motivated by the effective medium theory, which predicts a linear behaviour above the critical value [222]. The result gives an upper bound to the typical transition point. Thus, it can be considered a loose statistical criterion for the point, beyond which a considerable fraction of the network can become infected.

The social network is highly dynamic as observed before and therefore it does not surprise that the critical infectious period is as low as 1.2 hours (Fig. 4.4A). For the sexual contact network, we find that a disease with an infectious period above 48 days can affect a considerable fraction of the network (Fig. 4.4B). Finally, we observe in the pig trade data a critical transition at $\tau_c = 10.8$ days (Fig. 4.4C). This finding may suggest for example that the detection time of any outbreak in the livestock production should be below 11 days in order to minimize the impact of a disease.

In Sec. 2.3.3, we studied the percolation-like transition from local to global outbreaks within the *individual-based* model and following [33] we derived a spectral criterium to determine the critical parameter. This approach is based on the mean

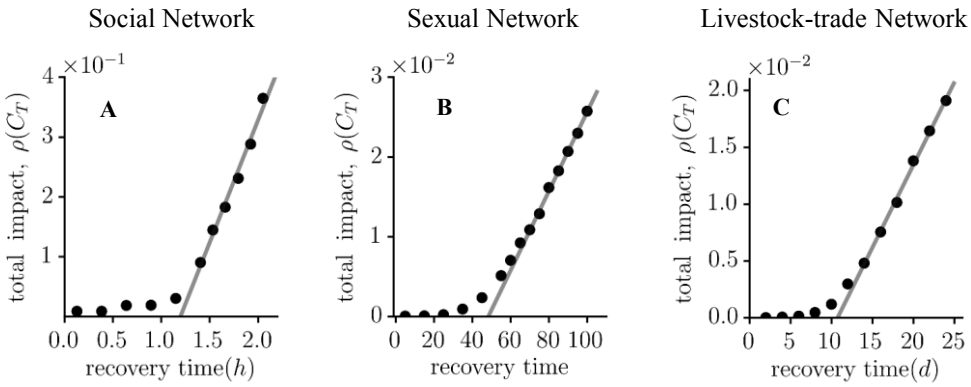

Figure 4.4: Estimating the critical infectious period for the social (A), the sexual (B) and the livestock-trade network (C). Fraction of nodes, which have been infected up to the observation time as a function of the infectious period τ. The grey line is a linear regression through the last 6 data points and the zero crossing gives a rough estimate of the critical infectious period τ_c. We found $\tau_c = 1.20 \pm 0.05$ hours, $\tau_c = 48 \pm 2$ days and $\tau_c = 10.8 \pm 0.3$ days for the social, the sexual and the livestock-trade network, respectively. The uncertainties are calculated from the least-squares fit.

field assumption that dynamic correlations can be ignored which means that the disease status of a node is independent of its nearest neighbour states. Though unrealistic for some networks [223], this approximation offers a convinient way to predict the epidemic threshold. On the other hand, the matrix-based approach, which we have presented in this chapter takes into account all time and topology dependent correlations. It therefore extends existing approaches and allows to deal with networks, where these correlations are crucial.

4.5 Conclusion

We have introduced a formalism to calculate paths for infections in temporal networks, which is based on elementary operations from linear algebra and Boolean arithmetic. In particular, we have extended the concept of unfolding accessibility to include finite infectious periods. Thereby, we have focused on the susceptible-infected-recovered (SIR) model with a fixed infectious periods τ and a transmission probability $\beta = 1$. Apart from the disease parameter τ, the proposed formalism is based entirely on the adjacency matrices of the temporal network, which refer to snapshots of the underlying contact network. Thus, we have been able to deal with the disease dynamics and the temporal network in one unifying framework, taking into account both temporal and topological correlations.

In order to apply our framework to epidemiology, we have derived three accessibility matrices that allow to compute simultaneously the prevalence, incidence and cumulative incidence for different initial conditions. The density of each matrix hence returns the mean value, averaged over all realizations.

From a computational point of view, the time complexity is dominated by the matrix multiplication $\boldsymbol{A}(t)\boldsymbol{I}(t)$ and the element-wise multiplication in $\boldsymbol{A}(t)\boldsymbol{I}(t) \circ (\neg\boldsymbol{R}(t+1))$ (see Eq. (4.7a)). Both can be implemented efficiently for sparse Boolean matrices. Note however, that the negated term $\neg\boldsymbol{R}(t+1)$ becomes dense for a small fraction of recovered nodes. Similarly, we find that well mixing networks, such as the social contacts graph, can lead to a dense prevalence matrix $\boldsymbol{I}(t)$ and hence slowdown the performance. Furthermore, we find that the storage complexity grows linearly with the recovery time and quadratically with the number of nodes in the case of a dense accessibility matrix.

We applied the formalism to three real-world networks: conference interactions, sexual contacts, and livestock-trade. In each case, we have presented an analysis based on the average prevalence, incidence, and cumulative incidence and recovered relations between the dynamics of the disease and the features of underlying contact graph. We have also analysed the total fraction of nodes, which have been infected at some point during the observation period depending on the infectious period τ. By considering the small-τ limit, we have observed a critical value τ_c, above which a disease can affect a considerable fraction of the network. This value can be related to the well known basic reproduction number R_0, which has been applied only recently to temporal networks, using an alternative method [33].

5

Contact-based modelling

Abstract. We present a contact-based model to study the spreading of epidemics by means of extending the dynamic message passing approach to temporal networks. The shift in perspective from node- to edge-centric quantities allows to accurately model Markovian susceptible-infected-recovered outbreaks on time-varying trees, i.e., temporal networks with a loop-free underlying topology. On arbitrary graphs, the proposed contact-based model incorporates potential structural and temporal heterogeneities of the underlying contact network and improves analytic estimations with respect to the individual-based (node-centric) approach at a low computational and conceptual cost. Within this new framework, we derive an analytical expression for the epidemic threshold on temporal networks and demonstrate the feasibility on empirical data.

5.1 Introduction

A fundamental challenge to mathematical epidemiologists is the accurate determination of the critical parameters that separate local and global epidemic outbreaks, i.e., the epidemic threshold as introduced in Sec. 2.1. There, we studied the early Kermack-McKendrick model [8] that separates a population according to the disease status into compartments of susceptible, infected, and recovered individuals with mass action type of equations to determine the transitions between them. A wide range of improvements has been proposed since, including the impact of stochasticity [13–15], non-Markovian dynamics [16–20] and, notably, heterogeneity in the contact structure [30, 94, 136, 224–226].

In recent years, the availability of mobility and contact data with high temporal resolution, so called *temporal networks*, offers another opportunity to improve analytical predictions [151–154, 210, 211]. To this end, we introduced in the previous Chapt. 4 a matrix-based approach that integrates the SIR model with time-varying topologies. Leveraging the concept of accessibility, this framework

offers an accurate and computationally efficient method to study outbreak dynamics. These advantages come, however, at a cost: In order to avoid a potentially large ensemble average, we focused on the infection probability $\beta = 1$, thus effectively removing stochastic effects. Moreover, we determined the epidemic threshold in Sec. 4.4.4 from a linear regression, which requires to calculate the outbreak size for a range of parameters. In order to propose a simpler and yet accurate method, we compare in this chapter two alternative probabilistic approaches.

A simplistic ansatz, to model the state $X_i(t)$ of a single nodes i in a network, takes the corresponding probabilities $P_i^X(t)$ directly as variables in a set of coupled dynamic equations [15, 33, 86–91]. We will refer to this approach as the *individual-based* (IB) model. In Sec. 5.3.1, we derived the set of dynamic equations (see Eq. (5.1)) and following [33], we arrived at a spectral criterion (see Eq. (5.14)) that determines the critical epidemic parameters. However intuitive, the analytic predictability suffers from the simplifying assumption that epidemic states of adjacent nodes are independent. As a consequence, the IB model can significantly overestimate the outbreak risk and therefore, a more accurate approach is required that inherets, however, the simplicity of the IB model.

Recently, a change from a node to an edge-centric perspective has been discussed within different frameworks in order ot overcome the inherent limitation of the IB model. These approaches include branching processes [138], message passing [18, 105], belief propagation [227] and, the edge-based compartmental model [104]. So far, however, edge-centric models are mostly limited to static topologies. It thus remains an open challenge to account simultaneously for topological and temporal properties of the underlying contact data and hence improve current predictions of the epidemic threshold [33, 175–178, 228].

In this chapter, we generalize the dynamic message passing approach from [105] for discrete-time Markovian *susceptible-infected-recovered* (SIR) spreading to time-evolving networks and derive the epidemic threshold within this new framework. The proposed model takes an edge-centric perspective, because the relevant dynamic equations are based on the set of edges. Furthermore, the framework integrates the complete temporal and topological information of the underlying network into the epidemic model. We will refer to our approach as *contact-based* (CB) and compare numeric predictions with the widely used IB model that takes a node-centric perspective. Within the CB framework we then derive a new analytic expression of the epidemic threshold for temporal networks and show that the edge-centric approach improves existing results [15, 33, 86, 87, 176, 177] at a low conceptual and numerical cost. Although both modelling frameworks can in principle account for contact weights that indicate the strength of a connection, we will focus on unweighted networks for simplicity and refer to the Appendix B.1 for an extension of the model. The CB and IB models have been implemented in Python with the source code available on Github [229].

The remainder of this chapter is structured as follows: First, we summarize the conceptual framework in Sec. 5.2 and formulate in Sec. 5.3 the dynamic equations

of the IB and CB models. Then, we derive the epidemic threshold for temporal networks within the CB framework in Sec. 5.4. We compare the edge- and node-centric approaches against Monte-Carlo (MC) simulations in Sec. 5.5 and close with a discussion in Sec. 5.6. Appendix B includes an extension to weighted contacts and heterogeneous epidemiological parameters in Sec. B.1 as well as a network analysis of the German cattle trade data Sec. B.2. Further results and applications of the CB model are summerized in Sec. B.3.

This chapter is based on our publication in [230]. The extension of the CB model to continuous time in Appendix C is part of a yet unpublished work.

5.2 Conceptual framework

As in the previous Chapt. 4, we consider a temporal network $\mathcal{G} = [G(0), G(1), ..., G(T-1)]$ with N nodes and T snapshots. A node $l \in \mathcal{N}$ represents an individual that is either susceptible, infected or, recovered at a given time t with a corresponding probability $P_l^S(t)$, $P_l^I(t)$ and, $P_l^I(t)$, respectively. Emphasizing the important difference between temporal and static elements, we refer to *contacts* as time-stamped links $(t, k, l) \in \mathcal{C} \subset \mathcal{T} \times \mathcal{N} \times \mathcal{N}$. Here, we denote with $\mathcal{N}$, $\mathcal{T}$ and, $\mathcal{C}$ the set of nodes, time stamps and, contacts, respectively. We further assume that every contact is of a constant duration and equal to the sampling time of the temporal network. With *edges*, we refer to the corresponding static elements in the time-aggregated network. In other words, an edge $(k, l) \in \mathcal{E} \subset \mathcal{N} \times \mathcal{N}$ exists if and only if at least one (temporal) contact is recorded between k and l. Here, we denote with $\mathcal{E}$ the set of edges. Moreover, we will assume directed edges throughout the chapter and represent an undirected contact as two reciprocal contacts otherwise. Following the convention in [18], we denote with $k \to l$ a directed edge from k to l, thereby also indicating edge-based quantities. For further details on static and temporal network theory, we refer to Sec. 2.2.

As the stochastic process, we assume a discrete-time SIR model. Here, a susceptible node that is in contact with an infected neighbor, contracts the disease with a constant and uniform (per time-step) probability β. Furthermore, we treat the transmission event from multiple infected neighbors as independent and similarly, we interpret potential (integer) edge-weights as independent infection attempts. Also, we do not account for secondary infections within one time step, i.e., only direct neighbors can be affected. Once infected, the individual recovers with a uniform and constant probability μ independently of the infection process and acquires henceforth a permanent immunity.

Concerning the contact data, we will focus our numerical analysis first on a face-to-face interaction network between 100 conference participants [50]. This so called proximity graph has a resolution of 20s and the observation time is limited to the first 24h. If necessary, we extend the data set with a periodic boundary condition in time. The time-resolved contacts allow to study the spread of airborne

diseases as well as the propagation of ideas and rumors. The data is available on SOCIOPATTERNS.ORG and using the source code on [229], results of this chapter can be easily reproduced.

As an illustrative example, we present in Fig. 5.1 the over-time probability that a selected node in the proximity graph is either susceptible (yellow), infected (red) or, recovered (gray). The results are derived from 10^4 Monte-Carlo simulations with the same initially infected node. The trajectories reflect the bursty activity of the underlying temporal network [50] within the first 12h and the subsequent inactive night time.

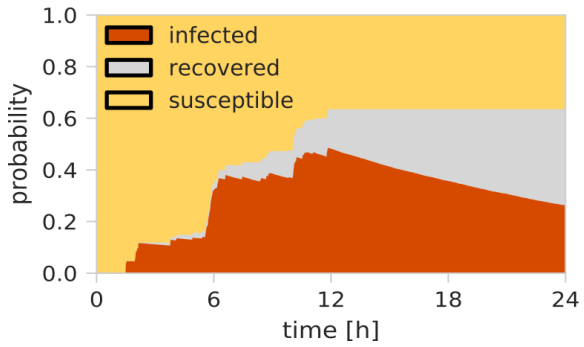

Figure 5.1: Illustrative examples of a simulated epidemic outbreak from a single initially infected node. Colors give probability that another arbitrarily selected node is in the susceptible (yellow), infected (red) or recovered (grey) state, respectively. Simulation parameters: $\mu = 2.85 \cdot 10^{-4}$, $\beta = 100\mu$, 10^4 MC realizations

As a second source of data with direct relevance to public health, we consider an excerpt of the national German livestock database HI-Tier (WWW.HI-TIER.DE). This temporal network comprises the movement of cattle between farms in Germany for the year 2010 with daily resolution. Within the observation window of 365 days more than 3 million transactions have been recorded between over 180,000 farms and traders, respectively. For more details on the graph see Appendix B.2. Cattle trade is considered an important transmission route for livestock-related diseases such as foot-and-mouth disease (FMD), which broke out in Great Britain in 2001 with estimated costs of 8 billion British Pounds [21]. Therefore, the analysis of the corresponding spatio-temporal graphs is highly relevant to public health institutions.

5.3 Dynamic equations

In this section, we will present the mathematical framework to model the stochastic SIR process as outlined in the introduction and Sec. 5.2. Our a main focus is the CB model but, in order to facilitate a direct comparison between the node and the edge-based approach, we will begin with a short overview of the IB model and refer to Sec. 5.3.1 for details.

5.3.1 Individual-based model

In the IB model, we focus on the marginal probabilities $P_l^S(t)$, $P_l^I(t)$ and, $P_l^I(t)$ that node $l \in \mathcal{N}$ is in state $X\{S, I, R\}$ at time t. These variables enter directly a set of $3 \times N$ coupled dynamic equations that we derived in Sec. 2.3.3:

$$P_i^S(t+1) = P_i^S(t) \prod_{j \in \mathcal{N}} [1 - \beta a_{i \to j}(t) P_j^I(t)] \tag{5.1a}$$

$$P_i^I(t+1) = (1-\mu) P_i^I(t) + P_i^S(t) \left\{ 1 - \prod_{j=1}^{N} [1 - \beta a_{i \to j}(t) P_j^I(t)] \right\} \tag{5.1b}$$

$$P_i^R(t+1) = 1 - P_i^S(t) + P_i^R(t). \tag{5.1c}$$

Here, we introduce the indicator function $a_{k \to l}(t)$ with $a_{k \to l}(t) = 1$ if a (directed) contact from k to l exists at time t and $a_{k \to l}(t) = 0$ otherwise. In Sec. 5.3.1, we interpreted $a_{i \to j}(t)$ as an element of the time-dependent adjacency matrix $\boldsymbol{A}(t)$. In this chapter, however, we will refer to it as an edge-based quantity as indicated by the slightly different notation. Moreover, we will assign a probability $z_l = P_l^S(0)$ that node l is initially susceptible as well as $P_l^I(0) = 1 - z_l$ and $R_l^I(0) = 0$ throughout the chapter.

Though intuitive and in many cases sufficient from a modelling perspective, the limits of the IB model are difficult to estimate due to the ad-hoc factorization of the joint probability in Eq. (5.1a). Even for the simplest network with two nodes connected by an undirected static edge, the IB approach can deviate significantly from the expected outcome as illustrated in [231]. In their example recovery is neglected for simplicity and only the first node is infected initially with some probability $0 < z_1 \leq 1$. Counter-intuitively, the probabilities to find each node in the infected state converge to $P_1^I(\infty) = P_2^I(\infty) = 1$ according to the IB model, independent of the initial condition z_1. This is because integrating Eqs. (5.1) admits a probability flux from the outbreak location to the adjacent node and back to its origin again. This mutual re-infection, coined *echo chamber effect* in [231], appears because we neglect the fact that the probability P_2^I, to find the second node in the infected state, is conditioned on the state of the first node and thus the factorization in Eq. (5.1a) is not justifiable.

In an arbitrary network an initially infected node leads to a cascade of secondary infections within which all marginal probabilities are highly correlated. An accurate model excludes these previously infected nodes from those that can potentially contract the infection in the future. We will discuss in the next section how a shift from a node-centric to an edge-centric view can take into account some such dependencies.

5.3.2 Contact-based model

We begin with a slightly different approach to the marginal probability $P_l^S(t)$. First, we note that l is susceptible at time t, if it was susceptible initially (with probability $S_l(0) = z_l$) and has not contracted the infection from any of its neighbors up to

time t. We assign the probability $\Phi_l(t)$ to the latter statement. Thus without introducing any approximation at this stage, we can write

$$P_l^S(t) = z_l \Phi_l(t). \tag{5.2}$$

In order to determine $\Phi_l(t)$, we make the assumption that the underlying time-aggregated graph is a tree (ignoring directionality). Then, different branches originating in node l are independent as long as l remains susceptible and thus $\Phi_l(t)$ factorizes. However, if node l contracts a disease from a neighbor k with some probability and passes it on to another node k' then the corresponding probabilities P_k^I and $P_{k'}^I$ are clearly correlated. A simple solution that allows different branches to nonetheless be treated as independent is to prevent a probability flow through the root node in the first place. From a graph-theoretic perspective, this corresponds to the (virtual) removal of all out-directed contacts from the root node. This approach does not modify the dynamics of the node under consideration, because it can still contract the disease and once infected, the recovery process is independent of the topology. The idea reduces, however, considerably the amount of bookkeeping that would otherwise be necessary, if we accounted for the correlations directly. The singular node l is said to be a cavity node or in the cavity state [18, 105], a concept closely related to the test-node assumption [104] and the idea of cut-vertices [85]. With this, we can factorize $\Psi_l(t)$ and thus obtain

$$P_l^S(t) = z_l \prod_{k \in \mathcal{N}_l} \theta_{k \to l}(t). \tag{5.3}$$

Here, we introduced the probability $\theta_{k \to l}(t)$ that no disease has been transmitted from node k to the cavity node l up to time t.

The change in perspective towards an edge-centric analysis introduces new auxiliary dynamic quantities such as $\theta_{k \to l}(t)$. These are defined on the set of edges $\mathcal{E}$ of the time-aggregated network and thus the number of dynamic variables scales with E, the number of edges.

In order to obtain a system of dynamic equations, we focus on our first edge-centric variable $\theta_{k \to l}$. Initially, no disease was transmitted such that $\theta_{k \to l}(0) = 1$ for all edges $(k, l) \in \mathcal{E}$. Henceforth, the dynamic quantity reduces only (i) upon a temporal contact, indicated by $a_{k \to l}(t)$ and (ii) if the adjacent node k is infected without having transmitted the disease earlier to the cavity node l - we denote the corresponding probability with $I_{k \to l}(t)$. Hence, the out-flow of probability is given by $\beta a_{k \to l}(t) I_{k \to l}(t)$, leading to our first dynamic equation

$$\theta_{k \to l}(t+1) = \theta_{k \to l}(t) - \beta a_{k \to l}(t) I_{k \to l}(t). \tag{5.4}$$

Next, the probability $I_{k \to l}(t)$ that node k is infective at time t and has not yet passed the disease to the cavity node l evolves according to three contributions: (i) It decreases with the recovery probability μ and (ii) with the probability β to infect its target node upon a temporal contact. These processes are independent

and may contribute simultaneously with the joint probability $\beta\mu$. (iii) $I_{k\to l}(t)$ increases with the probability $\Delta S_{k\to l}(t) = S_{k\to l}(t) - S_{k\to l}(t+1)$ that k is newly infected by at least one of its incident neighbors *excluding* the cavity node l. In sum and with the initial condition $I_{k\to l}(0) = 1 - z_k$ these contributions lead to:

$$I_{k\to l}(t+1) = (1-\mu)[1-\beta a_{k\to l}(t)]I_{k\to l}(t) + \Delta S_{k\to l}(t). \tag{5.5}$$

Finally, we consider the probability $S_{k\to l}(t)$ that node k, adjacent to the cavity node l, is susceptible. Since k is not affected by the state of l, it stays susceptible if it does not contract the disease from any of its remaining, incident neighbors $j \in \mathcal{N}_k \setminus l$. It has been shown in [232] that the corresponding probability $\Phi_{k\to l}(t) = \prod_{j\in\mathcal{N}_k\setminus l}\theta_{j\to k}(t)$ factorizes and thus similar to Eq. (5.2), we find $S_{k\to l}(t) = z_l\Phi_{k\to l}(t)$ or equivalently

$$S_{k\to l}(t+1) = z_k \prod_{j\in\mathcal{N}_k\setminus l}\theta_{j\to k}(t+1). \tag{5.6}$$

The disease progression in the CB framework is fully characterized by Eqs. (5.4) and (5.5), a set of $2L$ coupled equations. Equation (5.6) is introduced here for convenience only and can be substituted into Eq. (5.5). Next, we return to the node-centric quantities. To this end, we note that $P_l^S(t)$ has been already determined in Eq. (5.3). The remaining marginals P_l^I and R_l are equivalent to the IB model and given by the conservation condition Eq. (5.7b), and Eq. (5.7c), respectively. Hence, we arrive at the following node-centric equations:

$$P_l^S(t+1) = z_l \prod_{k\in\mathcal{N}_l}\theta_{k\to l}(t+1) \tag{5.7a}$$

$$P_l^I(t+1) = 1 - P_l^S(t+1) - P_l^R(t+1) \tag{5.7b}$$

$$P_l^R(t+1) = P_l^R(t) + \mu P_l^I(t). \tag{5.7c}$$

The CB model is exact for temporal networks, whose undirected time-aggregated version is a tree-graph and therefore loop-free. Thus, the change of perspective allows us to take full account of dynamic correlations on tree topologies. Most realistic networks, however, contain a large number of loops such as triangles in social graphs, where two friends are likely to have many more friends in common. Here the CB model nevertheless appears to be “unreasonably effective" (cf. [233]) and improves predictions significantly with respect to the IB approach as we will see in Sec. 5.5. For further extensions to the model that include heterogeneous infection and recovery probabilities as well as weighted contacts see Appendix B.1.

5.4 Epidemic threshold

The condition defining the epidemic thresholds can be derived by examining small perturbations around the disease-free state. If such perturbations die out then any outbreak remains local, but if the perturbation grows then a global epidemic may

occur. For that, we consider a linearization of the dynamic Eqs. (5.4) - (5.6), which will give rise to a criticality condition, determining the *epidemic threshold*. We begin with the ansatz $\theta_{k\to l}(t) = 1 - \delta_{k\to l}(t)$ and $z_l = 1 - \epsilon_l$, where $\delta_{k\to l}(t), \epsilon_l \ll 1$ are small perturbations around the disease-free state for all nodes l and edges (k, l). Thus, Eq. (5.4) becomes:

$$\delta_{k\to l}(t+1) = \delta_{k\to l}(t) + \beta a_{k\to l}(t) I_{k\to l}(t). \tag{5.8}$$

In Eq. (5.6) we keep the linear terms of the Taylor expansion, which transforms the product into a corresponding sum:

$$\begin{aligned} S_{k\to l}(t+1) &= (1-\epsilon_k) \prod_{j\in\mathcal{N}_k\setminus l} [1 - \delta_{j\to k}(t+1)] && \text{(5.9a)} \\ &\approx 1 - \epsilon_k - \sum_{j\in\mathcal{N}_k\setminus l} \delta_{j\to k}(t+1) && \text{(5.9b)} \\ &= S_{k\to l}(t) + \beta \sum_{j\in\mathcal{N}_k\setminus l} a_{j\to k}(t) I_{j\to k}(t). && \text{(5.9c)} \end{aligned}$$

In Eq. (5.9b) we substituted the dynamic Eq. (5.8) and identified $S_{k\to l}(t)$ in the next step. From the resulting Eq. (5.9c) we can read the linearized form of $\Delta S_{k\to l}$, which allows us to decouple the dynamic equations for $I_{k\to l}$ (see Eq. (5.5)):

$$I_{k\to l}(t+1) \approx (1-\mu)[1 - \beta a_{k\to l}(t)] I_{k\to l}(t) + \beta \sum_{j\in\mathcal{N}_k\setminus l} a_{j\to k}(t) I_{j\to k}(t). \tag{5.10}$$

Next, we rewrite the remaining set of E coupled dynamic equations in a compact, matrix-based formulation and therefore introduce the vectors $\boldsymbol{I}(t)$ and $\boldsymbol{a}(t)$ with elements $I_{k\to l}(t)$ and $a_{k\to l}(t)$, respectively. To this end, we also express the linear operation $\sum_{j\in\mathcal{N}_k\setminus l} a_{j\to k}(t)$ in Eq. (5.10), which acts on the elements $I_{k\to l}(t)$ of the state vector, through the temporal unweighted non-backtracking matrix $\boldsymbol{B}(t)$:

$$B_{kl,jk'}(t) = \begin{cases} 1, & \text{if } k' = k,\ j \neq l,\ (t, j, k') \in \mathcal{C}, \text{ and } (k, l) \in \mathcal{E} \\ 0, & \text{otherwise.} \end{cases} \tag{5.11}$$

In other words, $B_{kl,jk'}(t) = 1$ if the contact (t, j, k') at time t is incident on the edge (k, l) (implying $k' = k$), and additionally $j \neq l$. Otherwise we have $B_{kl,jk'}(t) = 0$. It is only the non-backtracking property $j \neq l$ that sets $\boldsymbol{B}$ apart from the adjacency matrix of the ordinary line-graph. For temporal networks a subtle distinction has to be made between the first and the second index of the $L \times L$ dimensional matrix $\boldsymbol{B}$: The first corresponds to an out-directed (*static*) edge $(k, l) \in \mathcal{E}$ of the underlying aggregated network and can be interpreted as a potential contact in the future. The second, however, is an incident (*temporal*) contact $(t, j, k') \in \mathcal{C}$ from node j to k' at time t. We also introduce the diagonal matrix $\mathrm{diag}(\mathbf{1} - \beta\boldsymbol{a}(t))$, with diagonal elements given by the vector $\mathbf{1} - \beta\boldsymbol{a}(t)$. Here, we denote with $\mathbf{1}$ the

vector of all ones. With these definitions, we rewrite Eq. (5.10) to

$$\boldsymbol{I}(t+1) = \left[(1-\mu)\,\mathrm{diag}(\mathbf{1} - \beta\boldsymbol{a}(t)) + \beta\boldsymbol{B}(t)\right]\boldsymbol{I}(t). \tag{5.12}$$

The explicit solution to the state vector $\boldsymbol{I}(T)$ at final observation time T is formally given by $\boldsymbol{I}(T) = \boldsymbol{M}(\beta,\mu)\boldsymbol{I}(0)$, where the so-called *infection propagator* $\boldsymbol{M}$ [175] is introduced for notational convenience:

$$\boldsymbol{M}(\beta,\mu) = \prod_{t=0}^{T-1}\left[(1-\mu)\,\mathrm{diag}(\mathbf{1} - \beta\boldsymbol{a}(t)) + \beta\boldsymbol{B}(t)\right]. \tag{5.13}$$

In order to evaluate the asymptotic behavior, we assume a periodic boundary condition in time, i.e., $\boldsymbol{B}(t) = \boldsymbol{B}(t+T)$. This allows us to assess the vulnerability of the temporal network through the spectral radius, i.e., the dominant eigenvalue λ_1 of the propagator $\boldsymbol{M}$. In particular, we find that a SIR-type outbreak is asymptotically stable under small perturbations, i.e., remains confined to a small set of nodes, as long as the dominant eigenvalue satisfies $\lambda_1[\boldsymbol{M}(\beta,\mu)] < 1$. Thus, the phase transition is given by the criticality condition

$$1 = \lambda_1\left(\prod_{t=0}^{T-1}\left[(1-\mu)\,\mathrm{diag}(\mathbf{1} - \beta\boldsymbol{a}(t)) + \beta\boldsymbol{B}(t)\right]\right). \tag{5.14}$$

Note that for irreducible and non-negative matrices the largest eigenvalue is simple and positive according to the Perron-Frobenius theorem [206]. Assuming $0 \leq \beta, \mu < 1$, a sufficient condition for temporal networks is to restrict contacts to the *giant strongly connected component* (GSCC) of the underlying time-aggregated graph. In Sec. 5.5.2 we will fix the recovery probability μ and determine the critical infection probability β_{crit} as the root of $f(\beta) = 1 - \lambda_1[\boldsymbol{M}(\beta,\mu)]$ for different empirical networks.

We conclude this section with a discussion on the static network limit. In the so-called quenched regime, the disease evolves on a much faster time scale than the dynamic topology and thus operates on an effectively static network with $\boldsymbol{B}(t) \equiv \boldsymbol{B}(0) \equiv \boldsymbol{B}$ and $\boldsymbol{a}(t) \equiv \mathbf{1}$ for all times t. As in the temporal analysis, we restrict the network to the GSCC so that the Perron-Frobenius theorem [206] applies. In this limit the dynamic equations (5.4) - (5.6) reduce to the *dynamic message passing* formulation in [105]. Moreover, Eq. (5.13) becomes now a product $\prod_{t=0}^{T-1}\boldsymbol{M}_{\mathrm{fast}}(\beta,\mu) = [\boldsymbol{M}_{\mathrm{fast}}(\beta,\mu)]^T$ of T identical, single time step propagators

$$\boldsymbol{M}_{\mathrm{fast}}(\beta,\mu) = (1-\mu)(1-\beta)\mathbb{1} + \beta\boldsymbol{B},. \tag{5.15}$$

Here, $\mathbb{1} = \mathrm{diag}(\mathbf{1})$ denotes the identity matrix.

The spectral radius in Eq. (5.14) factorizes to $\lambda_1[\boldsymbol{M}_{\mathrm{fast}}(\beta,\mu)^T] = \lambda_1[\boldsymbol{M}_{\mathrm{fast}}(\beta,\mu)]^T$, and it follows that the criticality condition Eq. (5.14) reduces to $\lambda_1[\boldsymbol{M}_{\mathrm{fast}}(\beta,\mu)] = 1$. Furthermore, we find from basic linear algebra that $\lambda_1[\boldsymbol{M}_{\mathrm{fast}}(\beta,\mu)] = (1-\mu)(1-\beta) + \beta\lambda_1(\boldsymbol{B})$ and hence we obtain the corresponding

static threshold condition

$$\left(\frac{\beta}{\beta+\mu-\beta\mu}\right)_{\text{crit, fast}} = \frac{1}{\lambda_1(\boldsymbol{B})}. \tag{5.16}$$

The result in Eq. (5.16) deviates from the continuous-time result in [18, 103]. In the derivation presented here, the term $\beta\mu$ in Eq. (5.16) accounts for the simultaneous events when a node infects a neighbor and recovers within the same time step.

In contrast to the quenched regime, one can also consider the so-called annealed limit. Then, parameters β and μ are sufficiently small such that no more than one infection or recovery event can take place within the observation time. Therefore, we expand the infection propagator to the first order in β and μ and obtain:

$$\boldsymbol{M}_{\text{slow}}(\beta,\mu) = (1-T\mu)\mathbb{1} - T\beta\,\text{diag}(\bar{\boldsymbol{a}}) + T\beta\bar{\boldsymbol{B}}. \tag{5.17}$$

Here, $\bar{\boldsymbol{a}} = 1/T\sum_t \boldsymbol{a}(t)$ and $\bar{\boldsymbol{B}} = 1/T\sum_t \boldsymbol{B}(t)$ denote the corresponding time averaged quantities. It is insightful to evaluate simple bounds for the set of parameters $(\beta,\mu)_{\text{crit, slow}}$ that satisfies the threshold condition $\lambda_1(\boldsymbol{M}_{\text{slow}}) = 1$ in the annealed limit. With $1/T \leq \bar{\boldsymbol{a}} \leq 1$ for all elements in $\bar{\boldsymbol{a}}$ we thus find:

$$\left(\frac{\beta}{\beta+\mu}\right)_{\text{crit, slow}} \leq \frac{1}{\lambda_1(\bar{\boldsymbol{B}})} \leq \left(\frac{\beta}{\beta/T+\mu}\right)_{\text{crit, slow}}. \tag{5.18}$$

Assuming the upper bound in Eq. (5.18) overestimates the outbreak risk and can be considered a conservative choice from an epidemiological perspective. This limit is realized for a temporal network where every edge appears exactly once within the observation time, hence $\bar{\boldsymbol{a}} = 1/T$. The lower bound in Eq. (5.18) is exact in case of a static network, thus $\bar{\boldsymbol{a}} = 1$, and corresponds to the continuous-time result in [103]. However, this limit underestimates the outbreak risk and therefore we conclude with a note of caution when applying results from static network theory directly to time-varying topologies.

5.5 Application

A big advantage common to both the node-centric IB and edge-centric CB modeling framework is a significant reduction in computational complexity compared to MC simulation. The CB model requires iteration through all edges at every time step and thus the time complexity scales with $\mathcal{O}(ET)$. The IB formulation and a single MC realization require $\mathcal{O}(\bar{C}T)$, where $\bar{C}$ denotes the average number of active contacts, which can be significantly smaller than E. Stochastic MC simulations on the other hand require a large number of realizations in order to provide reliable statistics. The computational disadvantage of MC simulations becomes even more apparent when we consider a complex quantity such as the epidemic threshold, which requires multiple ensemble averages for different sets of epidemic parameters in order to fit the critical infection probability (see Sec. 5.5.2). Equally important however, is the accuracy of our analytic approach. Therefore, we will compare in

this section estimations from the IB and CB mean-field model with MC simulations using empirical data as introduced in Sec. 5.2.

5.5.1 Numerical analysis of the mean-field dynamics

Indivual node trajectories. We begin with an analysis on the level of individual nodes. In Fig. 5.2, we show the cumulative infection probability for a small number of example nodes from the conference data set given the same outbreak location. The selection is intended to present qualitatively different trajectories, also demonstrating that deviations between the two models vary considerably. The MC result (blue curve) in Fig. 5.2A corresponds to the introductory example in Fig. 5.1. Here, a comparison with the analytic estimation shows that the CB approach leads to a substantial improvement to the IB model. Also in Fig. 5.2B-D, the trajectories are erratic, as they reflect the sudden changes in the underlying topology, highly individual and yet well approximated by the CB model. For all nodes in the network, we found that the CB model gives a closer upper bound to MC simulations because, unlike the IB framework, it accounts for dynamic correlations between nearest neighbor states.

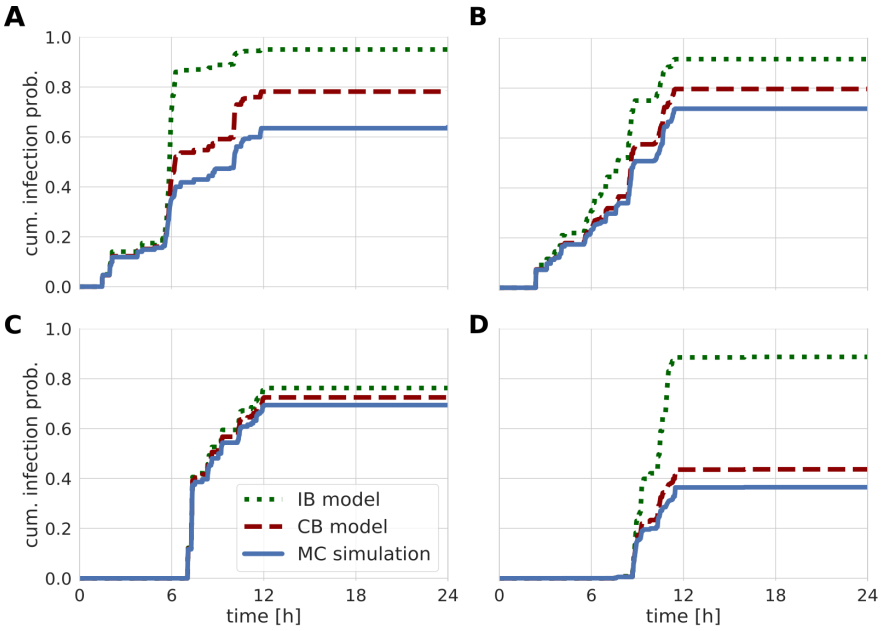

Figure 5.2: Epidemic trajectories for four exemplary individual nodes. We compare the cumulative infection probability from MC simulations (blue line) with estimations from the CB model (red dashed line) and the IB approach (green dotted line). Simulated results are averaged over 10^4 MC realizations with the same outbreak location and disease parameters as in Fig. 5.1.

From the detailed, node-level infection trajectory we can estimate the infection arrival time from a given outbreak location to all remaining nodes. For that purpose, we extend the contact sequence periodically in time until the infection probabilities are negligible. Then, we derive the infection arrival time to a single node from the corresponding cumulative infection probability (see Fig. 5.2) as

follows: (i) The discrete derivative of the cumulative infection probability gives the probability distribution to contract the infection at a given time step. (ii) The expectation value of probability distribution gives the mean infection arrival time at the a single node, corresponding to a scatter point in Fig. 5.3A. Here, we compare the expected values from MC simulations with the estimated infection arrival times given by the CB and IB model, respectively. In a perfect prediction the scattered values would lie on the diagonal but, as the contact network is far from a tree-like structure, the models estimate infection arrival times smaller than the observed values. The comparison between the IB and CB framework in Fig. 5.3B shows a considerably smaller relative deviation of CB estimations from the corresponding MC simulations for the given set of disease parameters and outbreak location.

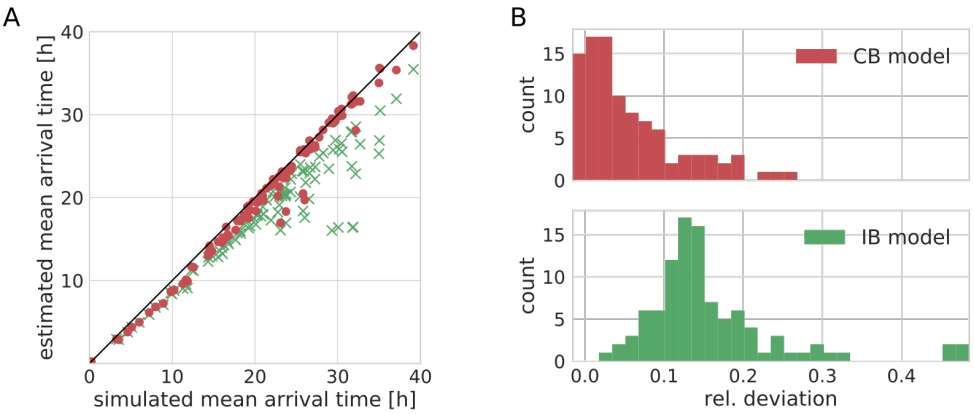

Figure 5.3: **A**: Comparison between simulated and estimated mean infection arrival times. We extend the data set periodically in time until the outbreak dies out. The discrete derivative of the cumulative infection probabilities (see Fig. 5.2) yields the infection arrival probabilities of which we take the average value for every node. Results according to the CB and IB model are visualized as red circles and green crosses, respectively. The epidemic starts from the same outbreak origin and disease parameters as in Fig. 5.1. **B**: Histogram over the relative deviation from the simulated infection arrival times. The numerical values are averaged over 10^5 realizations.

Another application focuses on the vulnerability of nodes with respect to a given outbreak location. Again, we assume an infinite time horizon and compare the cumulative probability that a node has been infected in the limit $t \to \infty$. As before, we find a good correlation between simulations and the estimated vulnerability in Fig. 5.4, whereas the CB model consistently outperforms the IB approach and overestimates the expected values surprisingly little given that the underlying aggregated network is fairly dense (the average degree is $\langle k \rangle \approx 19$) and far from being tree-like.

Impact of the outbreak location. Dynamic mean-field models such as the IB and CB framework provide realistic expectation values only if stochastic fluctuations are negligible. In order to illustrate the limitations, we study epidemic outbreaks for three different initially infected nodes in Fig. 5.5 A, B, and C,

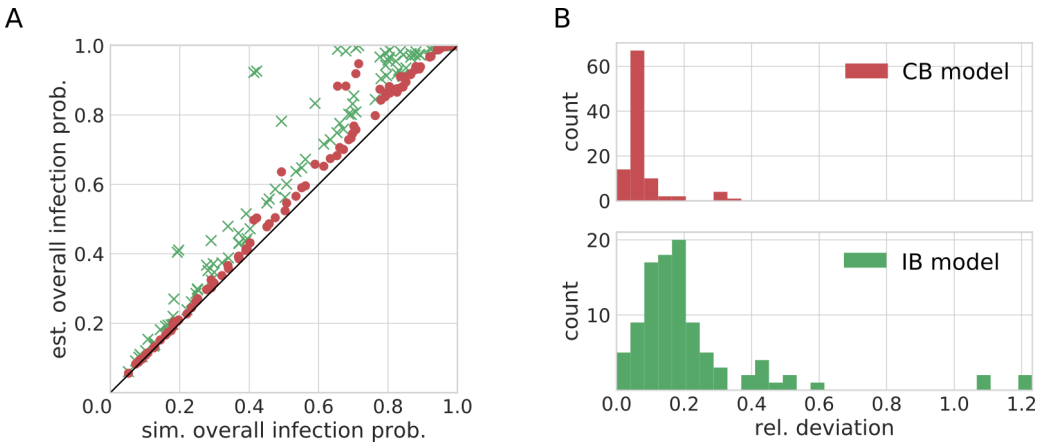

Figure 5.4: **A**: Comparison between simulated and estimated vulnerability. We compute the cumulative infection probability in the limit $t \to \infty$, also denoted as vulnerability. The comparison with CB and IB estimations visualized by red circles and green crosses, respectively. Each value corresponds to the vulnerability of a node given the same outbreak location and disease parameters as in Fig. 5.1. **B**: Relative deviation of the estimated values with respect to MC simulations. The numerical values are averaged over 10^5 realizations

respectively. The left column gives the time-resolved distribution of the outbreak size, and the right column presents the final distribution at the end of the three day observation period. For the ensemble average (blue line), we consider only realizations with more than 20 infected nodes overall. This threshold separates outbreaks that die out early due to stochastic fluctuations and thus permits a direct comparison with estimations from to the IB and CB framework in green and red, respectively.

We chose the outbreak locations such that the degree of stochasticity increases from top to bottom. In Fig. 5.5A.1, we find a narrow distribution around the ensemble average, which is well approximated by the mean-field models. Minor outbreaks due to early extinctions are well separated in Fig. 5.5A.2 from large epidemics. In Fig. 5.5B, the initially infected node leads to realizations with considerably stronger fluctuations and in Fig. 5.5C it is barely possible to separate early extinctions at all. Additionally, we observe a second source of stochastic variation, namely the time at which a disease escalates and hence evolves into a global epidemic. As a consequence, early outbreak sizes may be overestimated significantly before the analytic trajectory approaches the expectation value again (see Fig. 5.5C.1).

Remarkably, the performance of both mean-field models varies significantly with the outbreak location, even for R_0 well above the epidemic threshold. At the late phase of an outbreak, however, the mean-field models provide good approximations and consistently with Fig. 5.2, we find that the CB model outperforms the IB approach.

Large number of initially infected nodes. For some applications, we may be interested in the trajectory of a global epidemic, averaged over outbreak locations. A sufficiently large number of initially infected nodes would then avoid

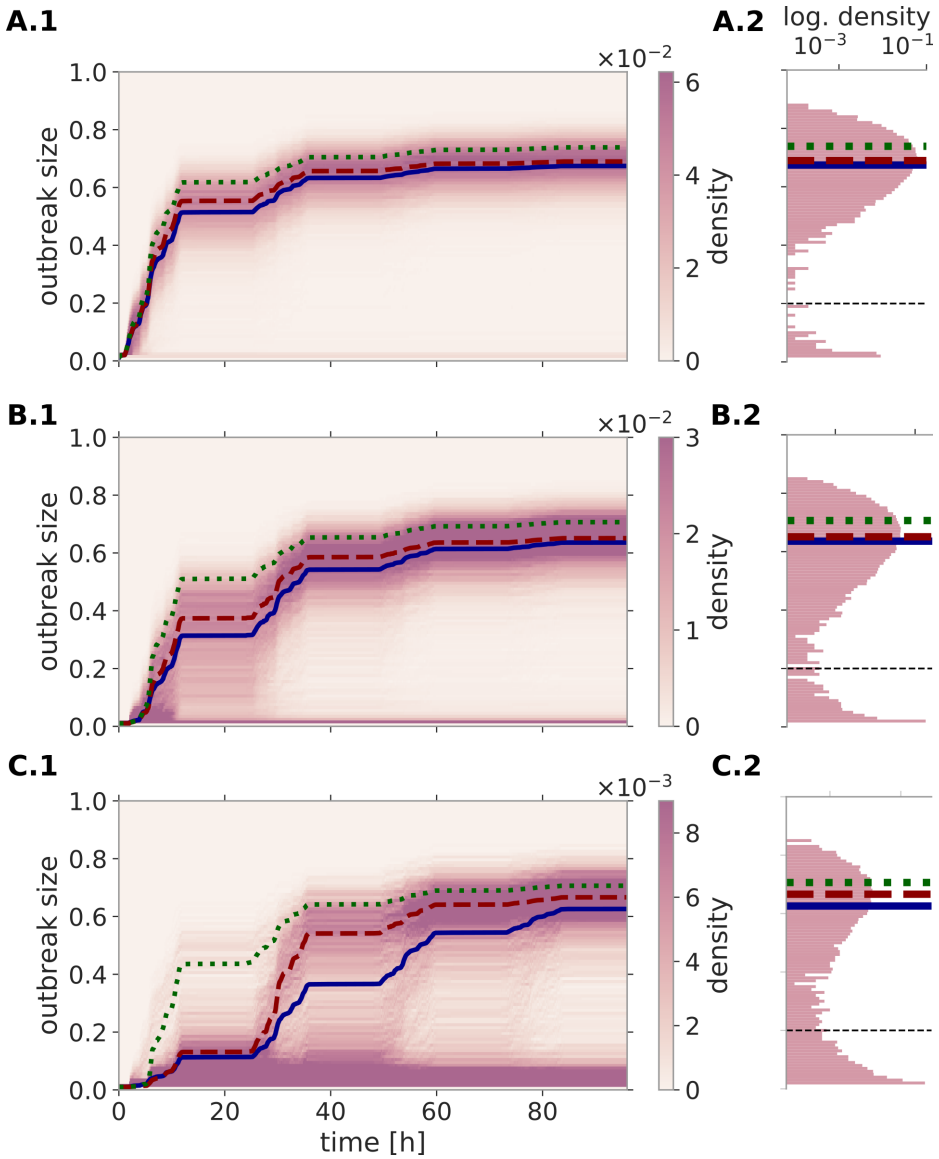

Figure 5.5: Left column: time-resolved and normalized outbreak size distributions for three different initially infected nodes. Epidemic parameters as in Fig. 5.1 and 10^4 MC realizations. The expectation value (blue line) assumes a cut-off at 20% of the population size (black dashed line). The CB and IB model are presented as a red dashed and green dotted line, respectively. Right column: Outbreak size distribution at the final observation time in logarithmic scale with corresponding expectation values.

complications with the early outbreak phase [104, 234]. In this case, we adjust the MC simulations such that every node is infected independently with a given probability $1 - z_l = 1 - z, \quad \forall\, l \in \mathcal{N}$ at $t = 0$. As for the analytic approach we only need to set a corresponding homogeneous initial condition and thus the computational complexity remains the same as in the previous case of one initially infected node.

In Fig. 5.6 (left column) we observe a narrow, time-dependent distribution of cumulatively infected nodes around the mean value for three different infection probabilities. Without applying any additional threshold, we find a close agreement between the averaged trajectory, the CB and IB model in all cases. In contrast to

Fig. 5.5, we observe in Fig. 5.6 (right column) only one peak in the distribution, due to the large number of initially infected nodes.

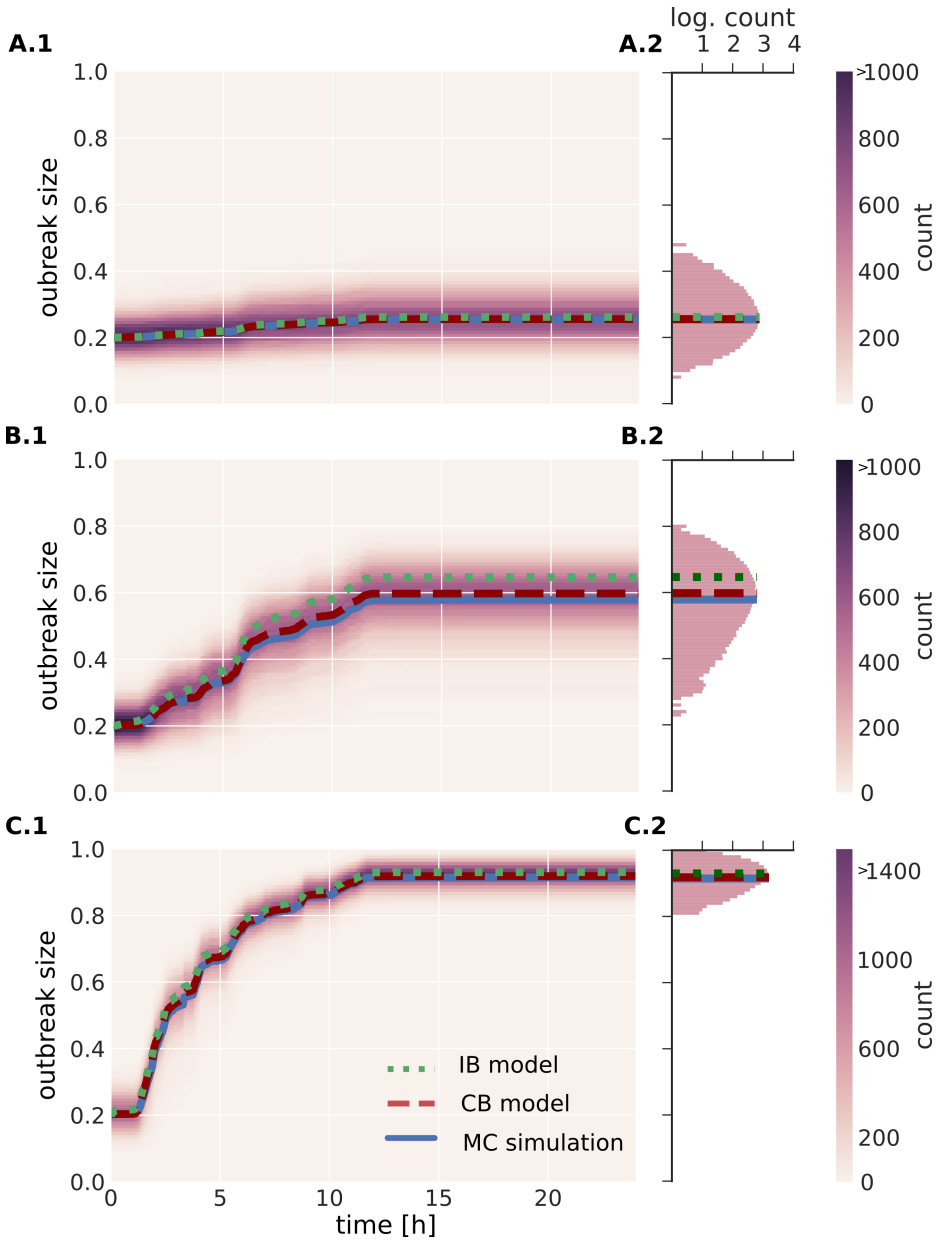

Figure 5.6: Cumulative infection probability with a large fraction of initially infected nodes for three different values of β and the same outbreak location as in Fig. 5.5A. Left column (panels A.1 - C.1): Time-evolving distribution (linear scale) of cumulatively infected individuals for infection probabilities $\beta = 0.01, 0.02$, and 0.1, respectively. We average over outbreak locations with 20% of the network initially infected at random. The mean outbreak size (blue line), averaged over 10^5 realizations and with a standard deviation below 10^{-4}, can thus be compared to the CB and IB model (red dashed and green dotted line, respectively) with no threshold applied. Right column (panels A.2 - C.2): Final distribution (logarithmic scale) together with the averaged values.

One potential application is to calculate the vulnerability of a node indepent of the outbreak location as discussed in [235]. Here, the vulnerability is defined as the probability that a given node is eventually infected by a disease that started somewhere in the network. The value can be used to rank nodes in order to

prioritize surveillance or vaccination measures to the nodes that are most likely to contract the disease when resources are limited. In Fig. 5.7 every curve represents the vulnerability of one node as a function of the infection probability β. These results are derived from the CB model, given an initial infection probability of $1 - z = 0.2$. The individual colors correspond to the degree of the node in the underlying time-aggregated graph and serve as a guide to the eye. Interestingly, we find that the ranking, as estimated by the CB model, may change with increasing infection probabilities β as can be seen from the highlighted curve in Fig. 5.7. This effect has been observed earlier in the context of static networks [235] and indicates that network properties alone are often not sufficient to rank nodes as they do not take into account details of the dynamic system.

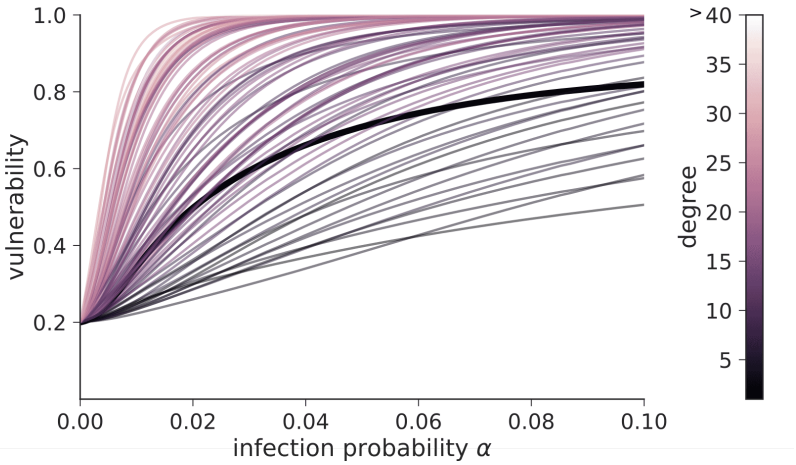

Figure 5.7: Vulnerability as a function of the infection probability β estimated from the CB model. Each curve represents the vulnerability of a node, i.e., the probability to contract the infection from a set of randomly chosen outbreak locations. Here, we estimate the vulnerability according to the CB model. Starting from an initial infection probability of $1 - z = 0.2$ we propagate the infection over time until convergence. We stop when the largest increase in vulnerability after 24h falls below 10^{-3}. The colors indicate the degree of each node in the underlying time-aggregated graph. Moreover, the behavior of one selected node is highlighted.

Final outbreak size. Another source of stochasticity, apart from the outbreak location (see Fig. 5.5), is the choice of disease parameters β and μ, respectively. For that, we focus on the final outbreak size, averaged over all (single node) outbreak locations. The distribution as a function of the infection probability β (see Fig. 5.8) shows a percolation-like transition from localized spreading to epidemics that affect a considerable fraction of the network. We apply the same threshold as in Fig. 5.5 for a direct comparison between the averaged outbreak size and the mean-field models for $\beta > 0.02$. Here, we find that the difference between the expected size and the CB estimation is close to negligible, whereas the IB model consistently overestimates the expected value.

A comparison at low values of the infection probability β becomes unreliable as stochasticity impedes a reasonable distinction between minor and global outbreaks. In order to illustrate the effect, we present in Fig. 5.9 the outbreak size distribution

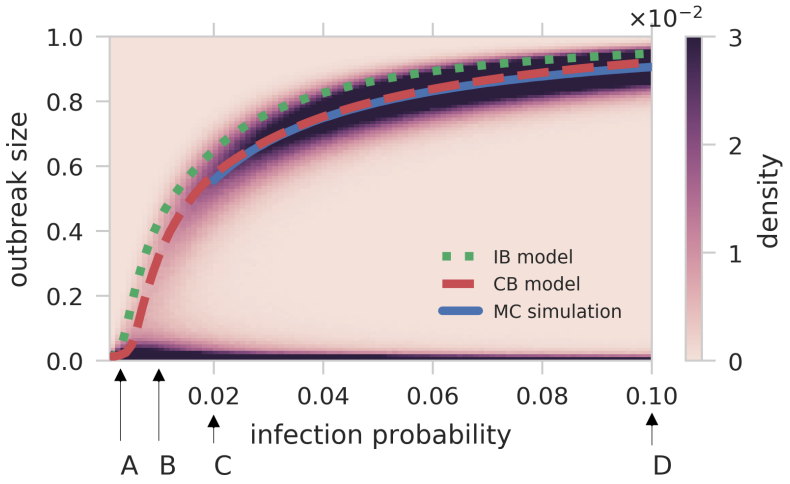

Figure 5.8: Distribution of final outbreak sizes as a function of the infection probability β. We perform 10^3 MC realizations for every value of β, each starting with one randomly chosen outbreak location. For $\beta > 0.02$, we show the averaged final outbreak size (blue line) with a cut-off at 20% of the population size. Estimations from the CB and IB model are presented as red dashed and green dotted lines, respectively. Labeled arrows at the bottom mark infection probabilities that correspond to Fig. 5.9 A,B,C and D, respectively.

for different values of β as marked by the arrows in Fig. 5.8. This representation highlights the transition from the sub- to the super-critical parameter domain: The unimodal distribution in Fig. 5.9A characterizes localized outbreaks, whereas the bimodal distribution in Fig. 5.9D clearly separates early extinctions and global epidemics. Next, we focus on the critical infection probability that marks the transition.

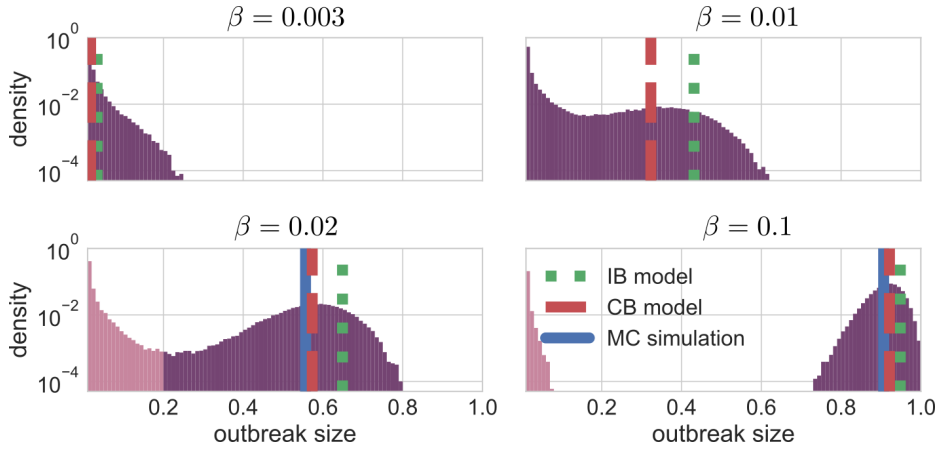

Figure 5.9: Distribution of final outbreak sizes for $\beta = 0.003$, 0.01, 0.02, and 0.1, respectively. For $\beta = 0.02$ and 0.1 we mark part of the distribution with outbreak sizes below the given threshold of 20% by a lighter color tone and neglect this contribution to the averaged value (blue vertical line). The expected outbreak size from MC simulations and the estimations from the CB and IB model are plotted as blue, red dashed and green dotted vertical lines, respectively.

5.5.2 Epidemic threshold

In Fig. 5.10A, we present the region of small β from Fig. 5.8, in order to focus on the transition from localized outbreaks to the sudden emergence of global epidemics. We determine the critical infection probability β_{crit} (vertical blue line) from the maximum of the relative standard deviation [175], also known as coefficient of variation (see blue line in Fig. 5.10B):

$$c_v = \frac{\sqrt{\langle R_\infty^2 \rangle - \langle R_\infty \rangle^2}}{\langle R_\infty \rangle}. \tag{5.19}$$

Here, we denote with $\langle R_\infty \rangle$ and $\langle R_\infty^2 \rangle$ the first and second moment of the outbreak size distribution. The coefficient of variation captures the intuition that fluctuations dominate the outbreak size distribution close to the transition. Indeed, c_v diverges at the critical point for infinitely large networks, indicating a second-order phase transition [145].

Analytically, we determine β_{crit} from the spectral criterion in Eq. (5.14) for the CB model and similarly within the IB framework [175, 236]. The comparison in Fig. 5.10 shows that the IB and CB model, marked by a red dashed and green dotted line, respectively, underestimate the critical infection probability from MC simulations (blue line) and thus overestimate the outbreak risk. Consistent with our previous results, we can state that a shift from a node- to an edge-centric framework improves the analytic prediction. In Appendix B.3, we present similar results for different values of the recovery probability μ. Next, we continue with a realistic application of the epidemic threshold to the German cattle trade network.

5.5.3 Application to German cattle trade.

We now consider a completely different data set, where the system size is large and contacts are sparse over time. Our example is a cattle trade network, where the movements of animals between farms in Germany are recorded on a daily basis. Next, we isolate the trade within each federal state of Germany as visualized in Fig. 5.11 and restrict trade to the GSCC of the underlying aggregated graph. Disregarding the smallest networks (those with less than 27 nodes), we thus obtain 12 time-varying graphs with sizes varying from 254 to 27,863 nodes and highly heterogeneous topological and temporal features (see Appendix B.2).

As in the previous section, we assume that premises can be either susceptible, infected or, recovered and trade events facilitates the transmission of a disease. Unlike before however, we take into account the number of traded animals during each transaction, i.e., the weight $w_{k\to l}$ of a (temporal) contact from node k to l. To this end, we modify the infection propagator in Eq. (5.13) and replace β by $1-(1-\beta)^{w_{k\to l}}$ (see Appendix B.1 for more information). In a potential outbreak, we assume that an infected node is detected with a constant probability μ each day after which it would be isolated and thus removed from the network. As a

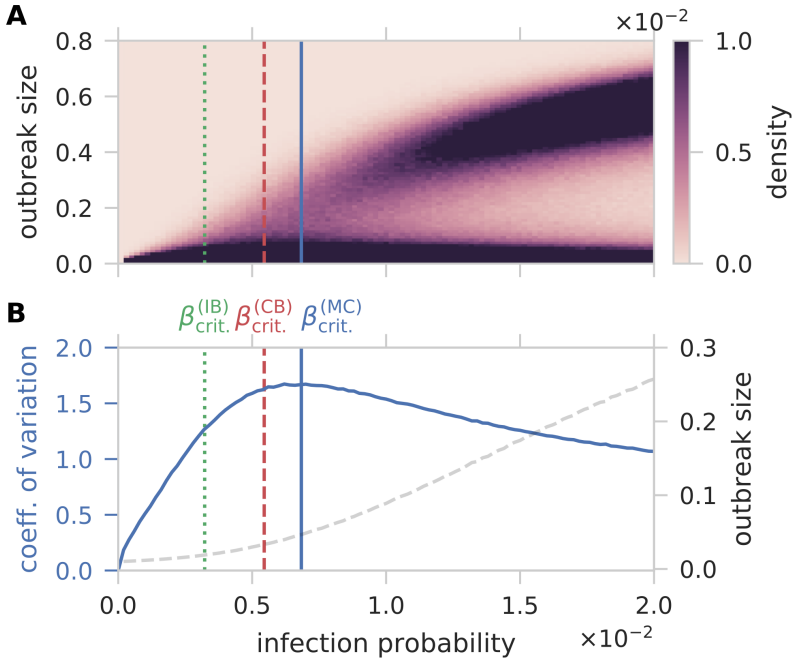

Figure 5.10: Estimation of the critical infection probability β_{crit}. **A**: Outbreak size distribution as in Fig. 5.8 for small values of β. The vertical blue, red dashed and green dotted line mark the critical value according to MC simulations ($\beta_{\text{crit.}}^{(\text{MC})}$), the CB model ($\beta_{\text{crit.}}^{(\text{CB})}$) and IB approach ($\beta_{\text{crit.}}^{(\text{IB})}$), respectively. **B**: From the distribution in A, we derive the coefficient of variation (blue line, left axis) and the mean outbreak size (grey dashed line, right axis).

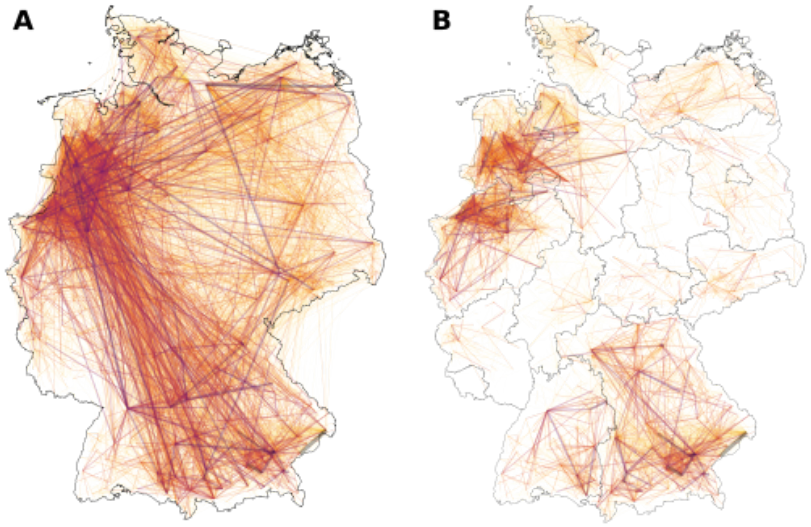

Figure 5.11: **A**: Cattle trade within Germany. Weighted edges correspond to directed trade relations within the year 2010, whereas the color indicates the accumulated number of traded animals. **B**: Cattle trade within the federal states of Germany. We confine the underlying time-aggregated graph to the GSCC and visualize here only edges with a flux of at least 50 animals.

consequence highly infectious diseases such as FMD can be modelled as SIR-type epidemics [21].

In Fig. 5.12, we compare the critical infection probability similar to Fig. 5.10 for six selected federal states with different transition characteristics. The critical value

derived from MC simulations varies between $\beta_{\text{crit.}} = 0.018$ (BY) and $\beta_{\text{crit.}} = 1.0$ (SN). The latter indicates that outbreaks remain localized for every choice of β due to sparse intra-state trade.

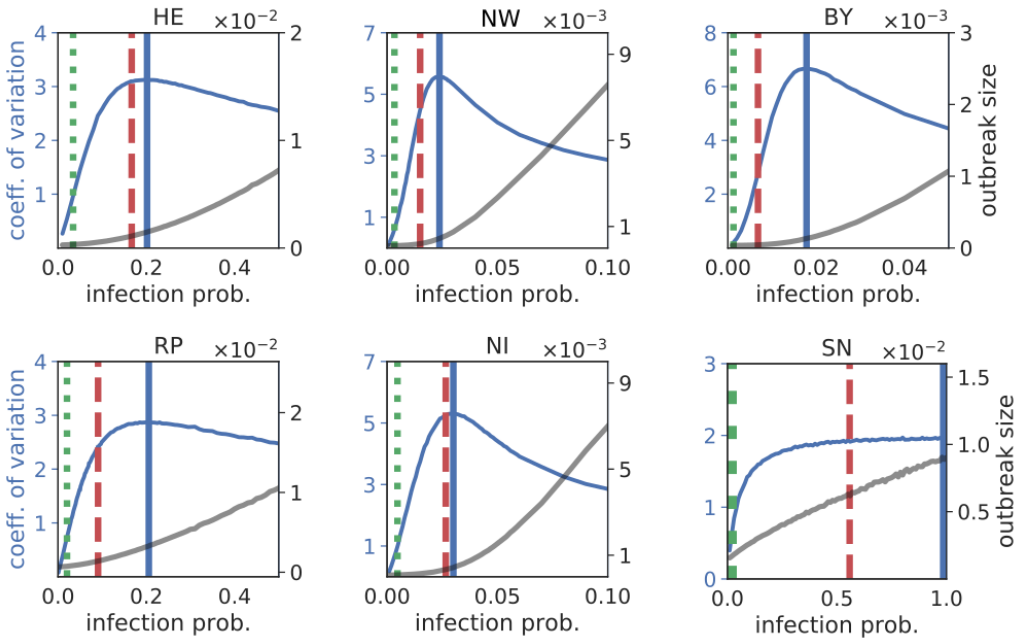

Figure 5.12: Detailed threshold analysis for six selected federal states with $\mu = 1/28$ (cf. Tab. B.1). Simulated mean outbreak size (grey line, right axis) and coefficient of variation (blue line, left axis) averaged over all initially infected nodes. The critical infection probability from MC simulations, the IB and CB models are presented as vertical blue, green dotted and red dashed lines, respectively.

As a potential application to public health institutions, we present in Fig. 5.13A the spatial variation of the epidemic risk in terms of β_{crit}. The quantitive comparison in Fig. 5.13B demonstrates that spectral methods provide a lower bound with a varying degree of accuracy depending on the network details. Despite their heterogeneity in size and activity, we find for all networks that the CB model outperforms the IB approach. The detailed results for all states as well as a similar analysis for $\mu^{-1} = 120$ are available in Appendix B.3.

5.6 Conclusion

In this chapter we have presented the contact-based (CB) model for epidemic SIR spreading on temporal networks as a conceptually similar framework to the widely used individual-based (IB) approach. Derived from the message-passing framework [18, 105] it inherits its accuracy on loop-free topologies and improves analytic estimations with respect to the IB approach for arbitrary time-evolving graphs. Moreover, the focus on edge-based quantities that are updated in discrete time steps allows a seamless integration of temporal interactions. Structually similar to the node-centric IB model, the proposed CB approach poses a low conceptual barrier and admits application on large graphs.

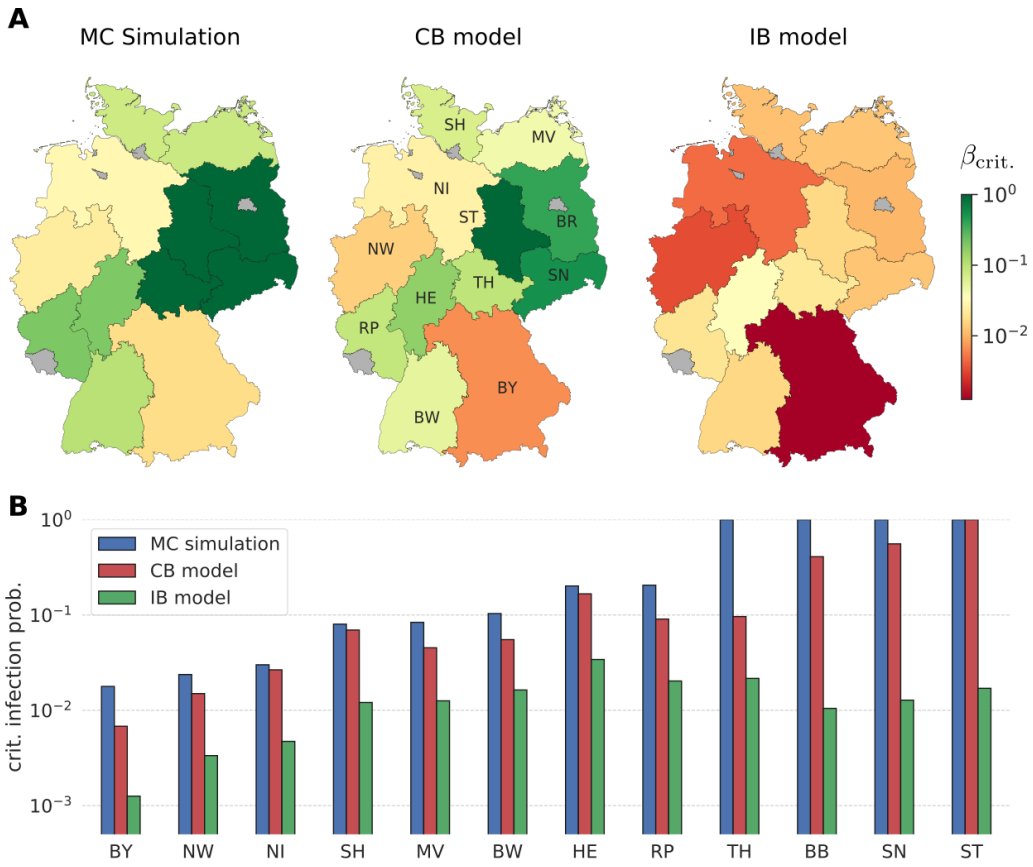

Figure 5.13: **A**: Spatial variation of epidemic risk due to cattle trade in Germany for $\mu = 1/28$. Federal states are colored according to the critical infection probability β_{crit} as determined from MC simulations, the CB and IB model, respectively (see Fig. 5.12 and appendix Fig. B.6 for details). The city states Berlin (B), Hamburg (HH) and Bremen (HB) as well as Saarland (SL) are excluded due to the small network size (see appendix Tab. B.1). **B**: Critical infection probability β_{crit} in logarithmic scale, sorted from high (left) to low risk (right). Results from MC simulations, the CB and IB model are presented as groups of blue, red, and green bars, respectively. Disclaimer: A realistic vulnerability analysis requires, for instance, heterogeneous recovery probabilities and complex counter measures (see Appendix B.2 for details).

Importantly, the accuracy of the CB model improves existing approximations of the epidemic threshold, which is a crucial risk measure for public health institutions. To this end, we have studied the largest eigenvalue of the infection propagator matrix, which determines the disease propagation in the low prevalence limit and takes into account the full temporal and topological information up to the observation time. The largest eigenvalue can be easily found through repeated matrix multiplications, i.e., the so-called power method. Without relying on extensive Monte-Carlo (MC) simulations and a subsequent parameter fit, the critical value can thus be estimated with efficient, vectorizable tools from linear algebra that are available for most high-level programming languages.

In the application section, we have focused first on a social contact-graph that can be used to analyze the propagation of airborne diseases as well as the spread of information. Our comparison between MC simulations and analytic estimations from the CB and IB model followed a bottom-up approach: We looked at (i) epidemic trajectories of individual nodes and their applications, (ii) the outbreak

size given the same initially infected node, (iii) the impact of a large number of initially infected nodes and, (iv) the percolation like transition of the outbreak size for a range of infection probabilities. In all cases, the CB model provides a closer upper bound to MC simulations than the widely used IB model. All results based on the conference data set can be reproduced using the PYTHON code provided in [229]

As a particularly important application, we then compared analytic estimations of the critical infection probability with extensive MC simulations. To this end, we included a case study of livestock trade within 12 federal states in Germany with highly heterogeneous characteristics in terms size, density and temporal activity. Consistently, we found that the CB model improves the previous lower bound of the IB framework at a low conceptual and computational cost.

Many excellent results have already been derived within the IB framework for empirical networks and in the context of random graphs (see [137] for a recent review) that can further improve the CB model. We therefore expect that the conceptual simplicity of the CB framework allows to integrate features such as non-Markovianity [17], stochastic effects [237], and estimations of uncertainty [106] that are important to realistic disease models on temporal networks. Also, first steps towards higher order models that go beyond the tree-graph assumption have been proposed in the context of percolation theory [238] and diffusive transport [239, 240] and we expect these improvement to also be applicable to the CB model.

6

Conclusion and outlook

This thesis aimed to contribute to the renewed global effort to mitigate the impact of communicable diseases. The causes of epidemic outbreaks are manifold and range from a lack of medical care in developing countries to vaccination resistance in industrial nations. Furthermore, climatic changes, as well as ever more intensifying livestock production, contribute to the emergence of new diseases. Finally, increasing mobility and urbanization facilitate a global spread of pathogens with a velocity not anticipated in the mid-20th century.

In order to understand the complex interplay of all factors and consequently to develop effective mitigation strategies, new integrative models are required that extend classical methods of theoretical epidemiology and integrate the intricate interaction patterns within our society. To this end, the rapid digitalization of numerous processes and the associated amount of personalized information offers the opportunity for a new class of data-driven models that could lead to highly targeted intervention schemes.

In this thesis, we have developed three integrative frameworks that combine well-established epidemic models with empirical mobility and contact data that potentially evolve in time. We will now revisit a series of questions that have guided our research from the beginning and thereby summarize the results of the main chapters.

When does a disease arrive in a distant population? A key challenge to public health institutions is to estimate the transmission time of global epidemics from the initial outbreak until the first infected individual arrives in a remote area. To this end, we defined densely populated regions and airline connections as nodes and edges in a global mobility graph and proposed a new network-based measure as a proxy for the expected arrival infection time.

We structured our analysis into two parts. In a first step, we examined two preceding works under the same assumptions and thus identified similarities and limits of their models. Thereby, we studied topologies of increasing complexity,

which finally motivated two network-based measures as potential proxies for the expected arrival time of a disease: (i) The multiple-path effective distance (MPED) includes all possible propagation paths and offers a conceptionally appealing framework. However, the computational complexity increases exponentially with the network size, and therefore, applications are limited to small networks only. (ii) The dominant-path effective distance (DPED) estimates the expected arrival time from the most probable transmission path. Although this method is highly efficient, it can lead to significant deviations due to the focus on a single transmission path.

In a second step, we proposed the random-walk effective distance (RWED) to improve previous approximations. This approach integrates all potential transmission paths in order to estimate the arrival time. Moreover, we leveraged methods of Markov chain theory to formulate an efficient matrix-based algorithm. Through extensive numerical simulations, we then demonstrated that the RWED improves consistently in comparison with the DPED. Furthermore, our analysis hints at an underlying relationship between complex disease dynamics and a generic diffusion process that will be subject to future research.

How can we determine the vulnerable and spreading potential of individual nodes? Closely related to this question is the problem of identifying central nodes in a network, which represents one of the earliest challenges in network theory. Despite numerous publications and countless network measures, few methods integrate epidemic models with time-varying topologies, even though the transmission path of a disease depends crucially on the timing of the corresponding edges.

In order to integrate the generic susceptible-infected-recovered model with temporal networks, we extend the concept of unfolding accessibility and introduce a disease-dependent infection duration in Chap. 4. This formalism derives the accessibility matrix from a sequence of time-ordered adjacency matrices. From this, the vulnerability or spreading potential of an individual node follows from the corresponding row or column sum, respectively. In contrast to earlier methods, we take full account of dynamic correlations, and at the same time, our matrix-based formulation facilitates a computationally efficient implementation that considers all potential outbreak locations simultaneously.

What are the critical epidemic parameters that lead to global outbreaks? Since the beginning of the 20th century, theoretical epidemiologists studied the abrupt transition from local to global epidemics for different disease parameters. In addition to the biological properties of the pathogen, the spreading potential depends crucially on the topological and temporal contact patterns between individuals.

In order to combine temporal networks into the estimation of the epidemic threshold, we presented the contact-based (CB) model in Chap. 5. It improves the widely used individual-based (IB) approach through a shift in perspective

from node-centric to edge-based quantities. This conceptual change allows to account for correlations between the epidemic states of adjacent nodes and thus improves the accuracy of current predictions. From the non-linear dynamic model, we have then derived a spectral condition, which determines the critical disease parameters that mark the transition from local to global outbreaks. This method integrates the complete temporal and topological information available and can be implemented efficiently using vectorizable algorithms. Therefore, our spectral criterion is a promising tool for public health institutions to evaluate the risk of global outbreaks.

To conclude this summary, we take another look back at the main results of this thesis. With the increasing mobility and interconnectedness of our society, the risk of fast-spreading epidemics poses a threat to public health. To meet this growing challenge, new data-driven models have to integrate the temporal and topological complexity of interactions. Our contribution to this research focuses on three models that address specific problems of public health institutions and, hopefully, help to improve future mitigation strategies.

Outlook. The analysis of global epidemics in Chap. 3 revealed a high correlation between the RWED and the expected infection arrival time. This network measure is based on the first passage probability, which in turn is a central quantity in the theory of diffusive transport. This relationship between spreading and diffusion suggests that the RWED can be derived not only from a meta-population model but also entirely within a random-walk framework. To this end, the theory of randomized shortest paths provides a promising starting point. Here, the solution of an optimization problem defines a biased random walk that favors the shortest connection between two given nodes on the one hand but also allows for deviations on the other. This perspective could include the broad class of contact processes and thus provide a more general approach to study infection arrival times. We demonstrate the first steps towards a new framework for effective distances in Appendix A.

Moreover, it would be interesting to address a limitation of the RWED. So far, the measure also includes non-physical paths where nodes are visited several times. This inaccuracy is reminiscent of the echo-chamber effect of the individual-based model as discussed in Chap. 5. Similar to our solution using a contact-based approach, we can change to an edge-based perspective and define a non-backtracking effective distance: The edge-centric formulation would exclude the previously visited population from the set of reachable nodes. This idea reflects the intuition that an infected population cannot seed a new outbreak in the adjacent node, which caused the infection in the first place.

Concerning the contact-based (CB) model from Chap. 5, future studies could include a continuous-time formulation. This approach would correspond to the underlying (real) process and thus represents an obvious extension. First results in

Appendix C show that a conceptually similar formulation is possible and we expect that a comparable improvement over the time-continuous IB model is possible.

Moreover, in Chap. 5, we have studied only sufficiently large epidemics, where stochastic effects can be neglected. An attractive extension is to account for early extinctions due to stochastic fluctuations. A promising approach from static network theory maps the late time behavior of an epidemic to a percolation problem and thus, identifies minor outbreaks with disconnected components, i.e., subgraphs that are not part of the giant strongly connected component.

Finally, a central approximation of the CB model is that we assume a constant recovery rate. This results in an exponentially distributed infection period, which often deviates from realistic statistics. Again, first data-driven models with a non-Markovian recovery process have been proposed, but a generalization to time-varying topologies remains an open problem. In Appendix C, we present the first step towards a general epidemic model on temporal networks, which will be part of our ongoing research.

A

Alternative derivation of the effective distance

In Chap. 3, we derived the effective distance from a detailed analysis of a stochastic meta-population model. The results suggest, however, a more general derivation within the framework of random-walks. To this end, we will leverage in this appendix the theory of randomized shortest paths [241–243], which defines the sampling probability of a biased random walk through an optimization problem. Our results are part of an ongoing and yet unpublished research project.

A biased random walk is both, directed as it minimizes the weighted distance from the outbreak location u to a target node v and exploratory. The idea is to admit deviations from the shortest path to some degree in order account for parallel (suboptimal) connections. Intuitively, this interplay between randomness and a directed behaviour can be also used to characterize contagion dynamics: On the one hand every neighbor of an infected node can contract the disease (exploratory) while on the other hand an epidemic front propagates predominantly along shortest or, in general the most likely route [5]. This general correspondence in mind, we will re-derive the *random-walk effective distance* (RWED) D_{uv}^{RW} from Chap. 3:

$$D_{uv}^{\mathrm{RW}} = -\log\left(\sum_{\omega\in\mathcal{W}_{uv}} P^{\mathrm{ref}}(\omega)e^{-n(\omega)\delta}\right). \tag{A.1}$$

Here, $P^{\mathrm{ref}}(\omega)$ denotes the probability to sample a walk ω of length n. We will refer to $P^{\mathrm{ref}}(\omega)$ as a reference probability; hence the superscript 'ref'.

The res

As in the main Chap. 3, we assume a directed graph or network, $G = (\mathcal{V}, \mathcal{E})$ that is strongly connected with a set $\mathcal{N}$ of N nodes and a set $\mathcal{E}$ of weighted edges. An edge between node i and j is denoted by $i \to j$ or (i, j). Furthermore, we represent the graph through an adjacency matrix $\boldsymbol{F}$ with elements $F_{ij} > 0$ if a

link $i \to j$ exists and $F_{ij} = 0$, otherwise. Then, the transition probability from a node i to a neighbor j is proportional to the edge weight (see also Eq. (3.28)):

$$T_{ij} = \frac{F_{ij}}{\sum_{j \in \mathcal{N}_i} F_{ij}} \tag{A.2}$$

In addition, we assume that a transition cost $c_{ij} \geq 0$ is associated to each link $i \to j$ of the graph G. In the general derivation [243] all c_{ij} can be chosen arbitrarily and independently of the adjacency matrix. In the case of an epidemic spreading on a social graph, we will identify the cost later with the expected infection time for a meta-population model or, with the transmissibility [131] for a contact process. In order to focus on the main idea and to be consistent with Chap. 3, we assume in this appendix a homogeneous value $c_{ij} \equiv c$ for all edges $i \to j$.

Next, we remind that a uv-walk from a source node u to a target node v is a finite sequence of jumps to adjacent nodes, where every node can be visited multiple times. Moreover, we call $\mathcal{W}_{uv}$ the (infinite) set of uv-walks and denote with $C(\omega) = cn$ the total cost of a walk ω, i.e., the sum of transition costs along ω, while the length n of a walk is defined by the number of jumps. Moreover, the probability $P^{\text{ref}}(\omega)$ to sample a uv-walk ω is equal to the product of transition probabilities (see Eq. (3.18)):

$$P^{\text{ref}}(\omega) = \prod_{(i,j) \in \omega} T_{ij}. \tag{A.3}$$

We continue now with the sampling probability $P(\omega)$ for a *biased* random walk that corresponds to our SIR spreading process (not to be confused with P^{ref}). Following our intuition that a contagion dynamics is both, directed and exploratory we will choose $P(\omega)$ such that the probability distribution minimizes an objective function $\mathcal{F}(P)$, which is sometimes referred to as free energy (FE) [242]:

$$\mathcal{F}(P) = \overline{C}(P) - S(P \mid\mid P^{\text{ref}}), \tag{A.4}$$

In analogy a thermodynamic system (albeit not in a strict sense), the FE functional balances the internal energy, here the expected cost $\overline{C}$, and the (relative) Entropy S:

$$\overline{C}(P) = c \sum_{\omega \in \mathcal{W}_{uv}} n(\omega) P(\omega) \tag{A.5}$$

$$S(P \mid\mid P^{\text{ref}}) = - \sum_{\omega \in \mathcal{W}_{uv}} P(\omega) \log \left(\frac{P(\omega)}{P^{\text{ref}}(\omega)} \right) \tag{A.6}$$

The transition cost c in Eq. (A.4) plays a similar role as the temperature in a thermodynamic system: It determines how close the system is to the state of greatest possible entropy. It is instructive to discuss two limiting cases for the cost c, associated to step along one edge: (i) In the limit $c \to 0$, we maximize solely the entropy S relative to the reference model. In this case the optimal stochastic process is given by the random walk as defined in Eq. (A.3) and hence $P^{\text{FE}}(\omega) = P^{\text{ref}}(\omega)$. (ii) The opposite limit $c \to \infty$ leads to a shortest path, or in

general, a lowest cost problem. Hence, the probability distribution $P^{\mathrm{FE}}(\omega)$ that minimizes Eq. (A.4) has both, a directed and diffusive component to it.

In addition to Eq. (A.4), we require that the optimal sampling probability is normalized and thus, our constrained optimization problem reads

$$P^{\mathrm{FE}} = \arg\min_{P} \mathcal{F}(P) \tag{A.7}$$

$$\text{subject to} \sum_{\omega \in \mathcal{W}_{uv}} P(\omega) = 1 \tag{A.8}$$

The solution to Eq. (A.8) can be obtained using Lagrange multipliers and is given by the Boltzmann distribution [241]:

$$P^{\mathrm{FE}}(\omega) = \frac{1}{Z_{uv}} P^{\mathrm{ref}}(\omega) \exp[-c\, n(\omega)] \tag{A.9}$$

$$Z_{uv} = \sum_{\omega \in \mathcal{W}_{uv}} P^{\mathrm{ref}}(\omega) \exp[-c\, n(\omega)]. \tag{A.10}$$

The normalization factor Z_{uv} is known as the partition function and plays a central role in statistical mechanics. As the probability $P^{\mathrm{FE}}(\omega)$ reduces exponentially with the length $n(\omega)$, the random-walk that is defined by Eq. (A.9), and Eq. (A.10) is biased towards a short distance from u to v.

It is instructive to note, that the same result can be obtained using the maximum entropy principle [244] or by minimizing the expected cost subject to a fixed entropy spread S_0:

$$P^{\mathrm{FE}} = \arg\min_{P} \overline{C}(P) \tag{A.11}$$

$$\text{subject to} \begin{cases} S(P \mid\mid P^{\mathrm{ref}}) = S_0 \\ \sum_{\omega \in \mathcal{W}_{uv}} P^{\mathrm{FE}} = 1. \end{cases} \tag{A.12}$$

Next, we determine the free energy $\mathcal{F}(P^{\mathrm{FE}})$ using the optimal probability distribution P^{FE} to sample a biased random walk that starts in u and ends in v. For that we expand the relative entropy $S(P^{\mathrm{FE}} \mid\mid P^{\mathrm{ref}})$ according to:

$$S(P^{\mathrm{FE}} \mid\mid P^{\mathrm{ref}}) = -\sum_{\omega \in \mathcal{W}_{uv}} P^{\mathrm{FE}}(\omega) \log\left(\frac{P^{\mathrm{FE}}(\omega)}{P^{\mathrm{ref}}(\omega)}\right) \tag{A.13a}$$

$$= -\sum_{\omega \in \mathcal{W}_{uv}} P^{\mathrm{FE}}(\omega) \log\left(\frac{P^{\mathrm{ref}}(\omega) \exp[-c\, n(\omega)]}{Z_{uv}}\right)$$

$$+ \sum_{\omega \in \mathcal{W}_{uv}} P^{\mathrm{FE}}(\omega) \log\left(P^{\mathrm{ref}}(\omega)\right) \tag{A.13b}$$

$$= -\sum_{\omega \in \mathcal{W}_{uv}} P^{\mathrm{FE}}(\omega) \log\left(P^{\mathrm{ref}}(\omega)\right) + \sum_{\omega \in \mathcal{W}_{uv}} P^{\mathrm{FE}}(\omega) c\, n(\omega) + \log\left(Z_{uv}\right)$$

$$+ \sum_{\omega \in \mathcal{W}_{uv}} P^{\mathrm{FE}}(\omega) \log\left(P^{\mathrm{ref}}(\omega)\right) \tag{A.13c}$$

$$= \overline{C}(P^{\mathrm{FE}}) + \log\left(Z_{uv}\right) \tag{A.13d}$$

Using the definition of the free energy $\mathcal{F}$ from Eq. (A.4), we recover a well-known result from statistical physics:

$$\mathcal{F}(P^{\mathrm{FE}}) = -\log\left(Z_{uv}\right) \tag{A.14}$$

With the definition of the partition function Z_{uv} in Eq. (A.10), we can identify the free energy $\mathcal{F}$ with the random-walk effective distance in Eq. (A.1). In a meta-population model, the cost c can be identified with the expected infection time δ. The latter has been determined in [114] in terms of the mobility and epidemic parameters and denotes the expected time until the first infected individual moves from a population with a small fraction of initially infected to a fully susceptible, adjacent population. Assuming that a population population size proportional to the in/out-flux of individuals, the expected infection time δ is constant for all pairs of nodes.

In order to see how Eq. (A.14) can be applied to a contact network, we return to the partition function in Eq. (A.10). Every element $P^{\mathrm{ref}}(\omega)\exp[-c\,n(\omega)]$ of the sum can be interpreted as a probability to sample an 'evaporating' random walk [245]. This notion refers to a random walk that can disappear after each step with some probability. In terms of a contagion process, we identify the corresponding survival propability $\exp(-c)$ with the transmissibility T, i.e. the probability to infect a given neighbor before recovery. This observation defines the cost c in terms of the epidemic parameters [131]:

$$c \equiv T \tag{A.15a}$$

$$= -\log\left(\frac{\beta}{\beta+\mu}\right) \tag{A.15b}$$

With this, we can interpret $P^{\mathrm{FE}}(\omega)$ as the probability to sample a walk ω from the outbreak location u to node v, where every node succeeds to infect its neighbor before recovering.

We summerize our findings as follows. Instead of the meta-population analysis in Chap. 3, we derived the RWED within the theory of randomized shortest paths. Our main contribution is to identify the cost of a single jump c with the system parameter δ that we derived from the first moment of a Gumbel distribution in Eq. (3.13). The conceptional advantage of the more abstract derivation is that we can now extend the effective distance to the large class of contact processes by identifing the cost c with the transmissibility T. The results of this appendix present an outlook to an ongoing research project, where we develop a general framework for effective distances.

B

Additional material to Chap. 5

This appendix summerizes additional material that is directly related to the (discrete-time) contact-based model in Chap. 5. In particular, we present in Sec. B.1 a generalization of the dynamic equations and the epidemic threshold to weighted networks as well we heterogeneous infection and recovery probabilities. In Sec. B.2, we briefly analyze the German cattle trade network and discuss systematic errors of the spatial risk analysis from Chap. 5. Finally, we provide in Sec. B.3 additional numerical results from Monte-Carlo simulations and evaluate the accuracy of the contact-based and individual-based model.

The results in this appendix are based on our publication in [230]

B.1 Weighted networks & heterogeneous parameters

In order to improve the predictive power of a network model it is often required to take into account additional information. In the main article, we focused on the temporal dimension. However, another important piece of information is the weight of a contact. The interpretation of weight can range from passenger numbers in the global air traffic network to the impedance in a network of electric components. The distribution of weights in static as well as temporal empirical networks, shows often a broad tail [193] and as such the averaged edge weight can become meaningless due to large fluctuations. It is therefore often required to account for heterogeneous edge weights explicitly in epidemiological models. However, depending on the interpretation, weights may enter the model in different ways. Typically, a time-dependent edge weight $w_{k\to l}(t)$ is considered similar to the conductivity between two nodes k and l in an electric circuit. Translated to an epidemiological context, we would thus scale the infection probability linearly, i.e., $\beta a_{k\to l}(t)$ becomes $\beta w_{k\to l}(t)$ in a weighted network. Another approach, popular in the context of random walks and disease spreading, is to interpret an integer-valued weight $w_{k\to l}(t)$ as a number of parallel and unweighted edges that connect

k with l [33, 90]. From an epidemiological viewpoint, this idea would translate to $w_{k\to l}$ independent attempts to transmit the disease at time t. Here, the infection probability $\beta a_{k\to l}(t)$ becomes $1-(1-\beta)^{w_{k\to l}(t)}$ in the weighted case. In the main article we apply the latter interpretation to calculate the epidemic risk in the context of livestock trade (see Fig. 5.13). Here, weights correspond to the number of animals traded, each of which can infect the target population independently. For small probabilities $\beta \ll 1$ the adjusted infection probability simplifies to $1-(1-\beta)^{w_{k\to l}(t)} \approx \beta w_{k\to l}(t)$.

A second source of heterogeneity, which is commonly considered, includes heterogeneous infection and recovery probabilities, denoted as $\beta_{k\to l}$ and μ_k, respectively. With these modifications the dynamic equations Eq. (5.4)-(5.6) from Chap. 5 translate to

$$\theta_{k\to l}(t+1) = \theta_{k\to l}(t) - \Psi_{k\to l}(t) I_{k\to l}(t) \tag{B.1a}$$

$$S_{k\to l}(t+1) = z_k \prod_{j\in\mathcal{N}_k\setminus l} \theta_{j\to k}(t+1) \tag{B.1b}$$

$$I_{k\to l}(t+1) = (1-\mu_k)[1-\Psi_{k\to l}(t)] I_{k\to l}(t) + \Delta S_{k\to l}(t). \tag{B.1c}$$

Here, $\Psi_{k\to l}(t)$ denotes the probability that k infects l at time t given that the former is infected and has not yet transmitted the disease. For weighted networks we can choose $\Psi_{k\to l}(t) = \beta_{k\to l} w_{k\to l}(t)$ or $\Psi_{k\to l}(t) = 1-(1-\beta_{k\to l})^{w_{k\to l(t)}}$ as discussed above. The linearization of Eq. (B.1a)-(B.1c) around the disease-free state leads to

$$\boldsymbol{I}(t+1) = \mathrm{diag}[(\mathbf{1}-\boldsymbol{\mu})\circ(\mathbf{1}-\boldsymbol{\Psi}(t))]\boldsymbol{I}(t) + \boldsymbol{B}_\beta(t)\boldsymbol{I}(t). \tag{B.2}$$

Here, the circle denotes the elementwise product. Moreover, the E-dimensional vectors $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}(t)$ integrate the node- and edge-dependent values μ_k and $\Psi_{k\to l}(t)$, respectively. We also generalize the temporal non-backtracking matrix $\boldsymbol{B}_\beta(t)$ from Eq. (5.11) to the weighted:

$$[\boldsymbol{B}_\beta(t)]_{kl,jk'} = \begin{cases} \Psi_{j\to k'}(t), \text{ if } k'=k,\ j\neq l \text{ and } (k,l)\in\mathcal{E} \\ 0, \text{ otherwise.} \end{cases} \tag{B.3}$$

Note that the additional constraint $(t,j,k')\in\mathcal{C}$ from definition in Eq. (5.11) enters indirectly through $\Psi_{j\to k'}(t)$.

The largest eigenvalue λ_1 of the infection propagator determines the asymptotic stability for small perturbations around the disease-free state. Accounting for heterogeneity in β, μ and contact-weights, the criticality condition Eq. (5.14) from Chap. 5 reads:

$$1 = \lambda_1 \left(\prod_{t=0}^{T-1} \{\mathrm{diag}[(\mathbf{1}-\boldsymbol{\mu})\circ(\mathbf{1}-\boldsymbol{\Psi}(t))] + \boldsymbol{B}_\beta(t)\} \right). \tag{B.4}$$

Assuming $\beta \equiv \beta_{k\to l}$ for all edges $k \to l$, we can determine from Eq. (B.4) the critical (homogeneous) infection probability β_{crit} given a weighted, temporal network with heterogeneous recovery probabilities μ_k. Similarly, one can assume $\mu \equiv \mu_k$ for all nodes k and thus derive the critical (homogeneous) value μ_{crit} with heterogeneity in the infection probability $\beta_{k\to l}$.

B.2 German cattle trade network

The system of traceability of cattle in the EU requires that each animal is identified with ear-tags and that each movement, birth or death event has to be reported within 7 days of the event to the national livestock database. We consider an excerpt of the national German livestock database HI-Tier (WWW.HI-TIER.DE) for the year 2010. The database is administered by the Bavarian State Ministry for Agriculture and Forestry on behalf of the German Federal States. It records 3.2 million animal movements with a total of 13.4 million traded animals between 183,454 premises, such as farms, pastures, slaughter houses and traders within the observation window. Location of each animal holding was provided at the resolution of the municipality. We consider each trade event between two premises a temporal contact and we identify an edge if at least one contact has been recorded. The distribution of edges is highly heterogeneous in terms of geography, degree and weight. In Fig. B.1A we observe clusters of trade activity mostly within and between North Rhine-Westphalia (NW), Lower Saxony (NI), Baden-Württemberg (BW) and Bavaria (BY). The number of trading partners, i.e., the node degree, is broadly distributed as demonstrated in Fig. B.1B. Here, we differentiate between in-, out- and total degree. Similarly we find a broad distribution of edge weights in Fig. B.1C, i.e., the number of traded animals along a given edge.

The geographic distribution of nodes in Fig. B.2A shows dense regions in the north-west and south-east including the above mentioned federal states North Rhine-Westphalia (NW), Lower Saxony (NI), Baden-Württemberg (BW) and Bavaria (BY). Here, we also find the largest premises in terms of total traded animals: In Fig. B.2C color and size indicate the node strength, i.e., the aggregated trade volume. The heterogeneous distribution of strength becomes also apparent in Fig. B.2D where in-, out- and total strength is analyzed separately. Finally, we observe in Fig. B.2B the net flux, i.e., the difference between in- and out-directed trade volume. We display only nodes with at least 500 traded animals in Fig. B.2B and Fig. B.2C.

From a temporal perspective, we find that trade fluctuates between 10^2 and 10^4 active nodes, i.e., farms with at least one trade event on a given day, whereas minima appear regularly on the weekends (see Fig. B.3A). The weekly pattern is also apparent with a view to the inter-activation time distribution, i.e., the time interval between two successive trade events for a given node (see Fig. B.3). Here, we find a broad distribution of activity with peaks around 7, 14, and 21 days.

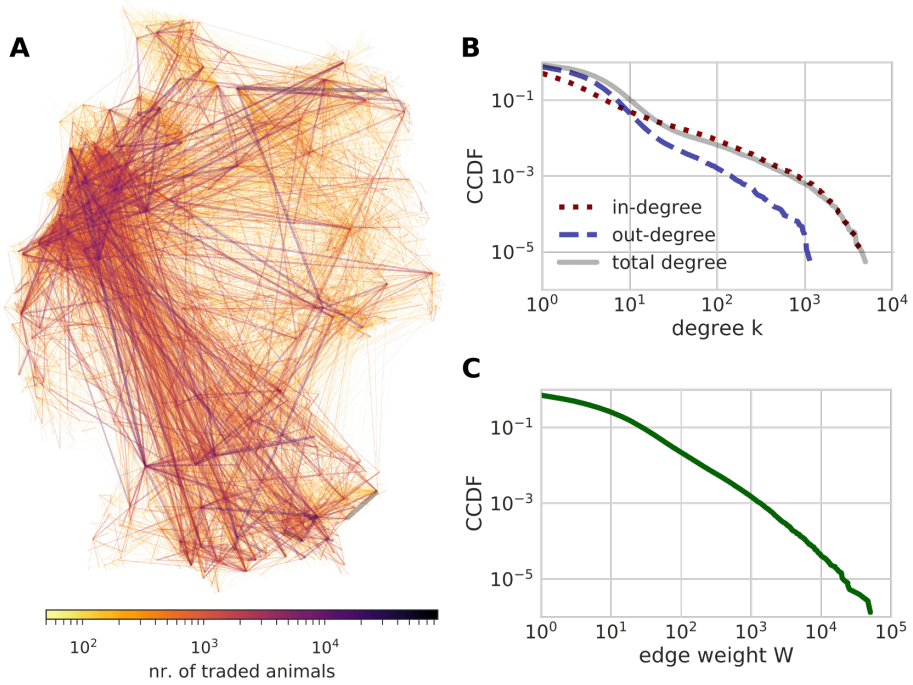

Figure B.1: Degree and weight analysis of cattle trade in the year 2910. Trade data is aggregated over time, i.e., the year 2010. **A**: Visualization of edges using geo-location with at least 50 animals traded along each link. The color and edge width indicates the aggregated edge weight. **B**: Complementary cumulative distribution function (CCDF) of in-, out- and total degree. **C**: CCDF of the edge weight distribution.

The geographic risk analysis in Fig. 5.13 requires us to separate the network into sub-graphs that correspond to the intra-state trade (see Fig. 5.11). The largest eigenvalue of the corresponding infection propagator allows us to evaluate the outbreak risk within a federal state due to the local movement of infected animals. In Tab. B.1 we list the names of all 16 federal states of Germany together with the corresponding ISO-abbreviation and basic statistics: The number of nodes, (static) edges and (temporal) contacts in the GSCC. The city states Berlin, Hamburg and Bremen as well as Saarland, which is particularly small state in terms of nodes, are marked with an Asterisk and are not considered for risk analysis in Fig. 5.13.

Separating the trade network into subgraphs as visualized in Fig. 5.11B inevitably reduces the outbreak risk as the neglected cross-border edges would otherwise facilitate the disease transmission. In Fig. B.4A we find that a considerable fraction of trade is directed across federal states and has thus been removed. This applies in particular for the federal states NI, NW and BW. Similarly, we find that the ratio between intra-state and in-directed trade lies between 0.6 (NW) and 0.9 (BY). Thus, we conclude that a considerable fraction of trade across borders is being neglected in the geographic risk analysis in Fig. 5.13.

It is also important to stress that we use the same parameter μ across all federal states and thus assume a uniform detection probability. In reality, federal states with a large number of premises tend to enforce stricter hygiene and intervention standards so that the actual epidemic risk for states such as Bavaria (BY) and North Rhine-Westphalia (NW) is much lower. A realistic evaluation for public

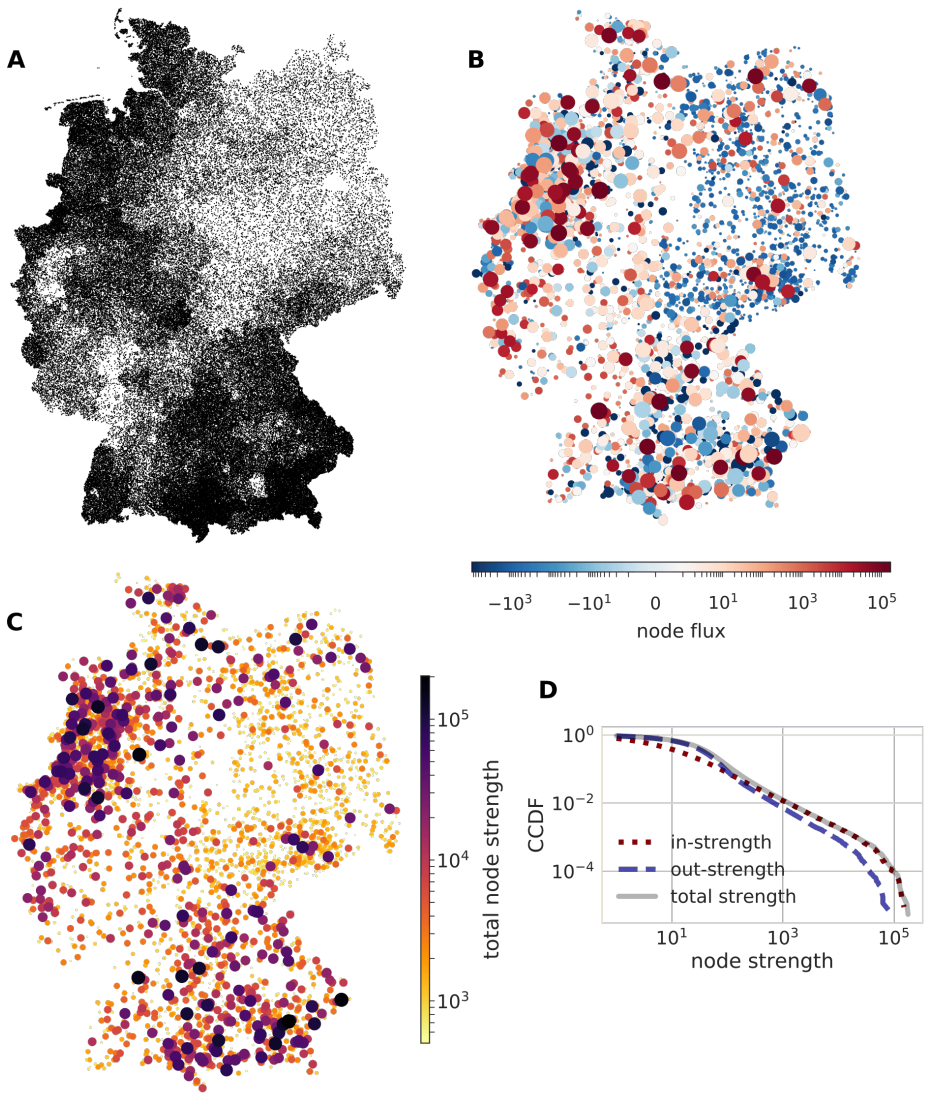

Figure B.2: Node strength and flux analysis of cattle trade in the year 2010. **A**: Geographic location of all 183,454 premises that have been recorded. **B**: Color indicates net animal flux, i.e., difference between in- and out-directed animals flux. Node size corresponds to the sum of both (see subfigure C). **C**: Color and size indicate the total node strength, i.e., the sum of incoming and leaving animals. **D**: Complementary cumulative distribution function (CCDF) of the in-, out- and total strength.

health must therefore include heterogeneous recovery (detection) probabilities on the level of states or individual nodes as discussed in [246] and Appendix B.1 as well as a complex disease response that includes trade restrictions, increased awareness and higher bio-security.

B.3 Additional numerical results

The analysis of the conference data set in Chap. 5 was limited to a single value of the recovery probability with $\mu = 2.85 \cdot 10^{-4}$. This choice corresponds to an expected infectious period of about 19.5 hours. In addition to the analysis of Chap. 5, we present in Fig. B.5 similar results for different values of μ. The

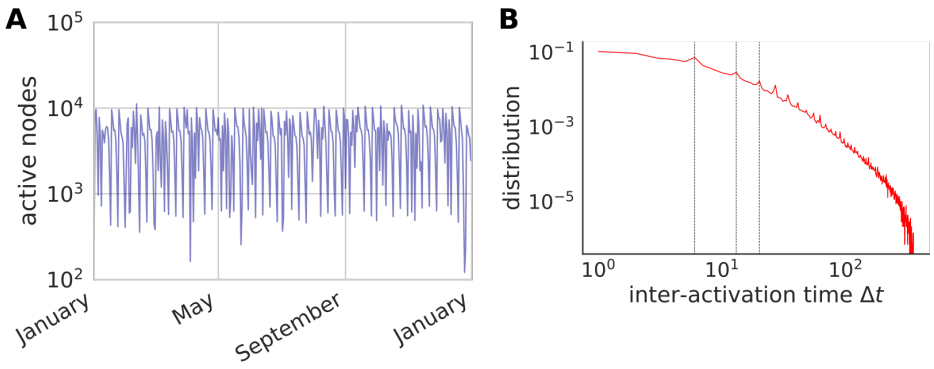

Figure B.3: Temporal analysis cattle trade in the year 2010. **A** Number of premises that trade at least once on a given day. **B**: Inter-event time distribution, i.e., the interval between subsequent trade events of an arbitrary farm. Vertical lines indicate 7, 14 and 21 days.

Federal state	ISO code	nodes	edges	contacts
Schleswig-Holstein	SH	3570	13541	49748
Hamburg*	HH	2	2	8
Lower Saxony	NI	12838	61044	272579
Bremen*	HB	4	6	21
North Rhine-Westphalia	NW	9826	42835	209677
Hesse	HE	3622	12498	51287
Rhineland-Palatinate	RP	1526	5386	29184
Baden-Württemberg	BW	9168	34434	150171
Bavaria	BY	27863	128596	550047
Saarland*	SL	26	52	343
Berlin*	BE	0	0	0
Brandenburg	BB	715	2144	11535
Mecklenburg-Vorpommern	MV	844	2852	21864
Saxony	SN	690	1935	12369
Saxony-Anhalt	ST	254	714	3422
Thuringia	TH	344	957	5361

Table B.1: Federal states of Germany and the corresponding abbreviation together with basic network statistics: number of nodes, number of (static) edges, number of (temporal) contacts. In all cases we confined the network to the GSCC. We mark the smallest networks, i.e., the city states Berlin, Hamburg, and Bremen as well as Saarland with an asterisk.

left, middle and right column correspond to Fig. 5.8, Fig. 5.10A, and Fig. 5.10B, respectively. In all cases, the CB model gives a closer bound to MC simulations as compared to the IB approach.

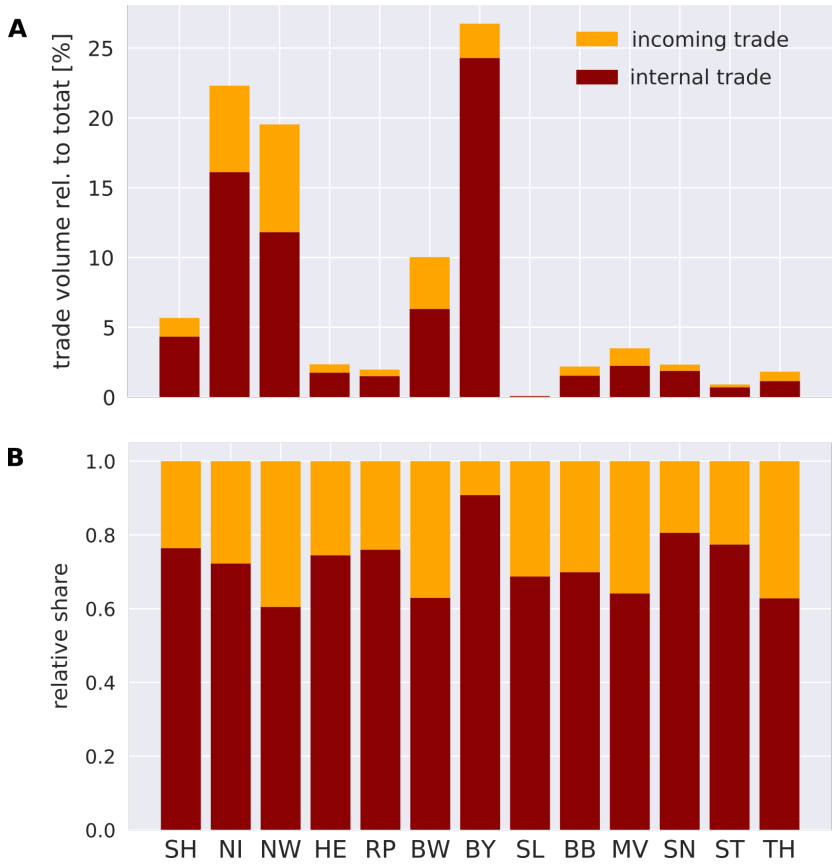

Figure B.4: Error estimation after restricting trade to the federal states. **A**: Intra-state trade in terms of trade volume relative to total trade as red bars. Yellow bars on top indicate fraction of in-coming animals. All bars together sum up to 100%. **B**: Sum of internal and in-coming trade volume is normalized for each federal state. See Tab. B.1 for list of states and their abbreviations.

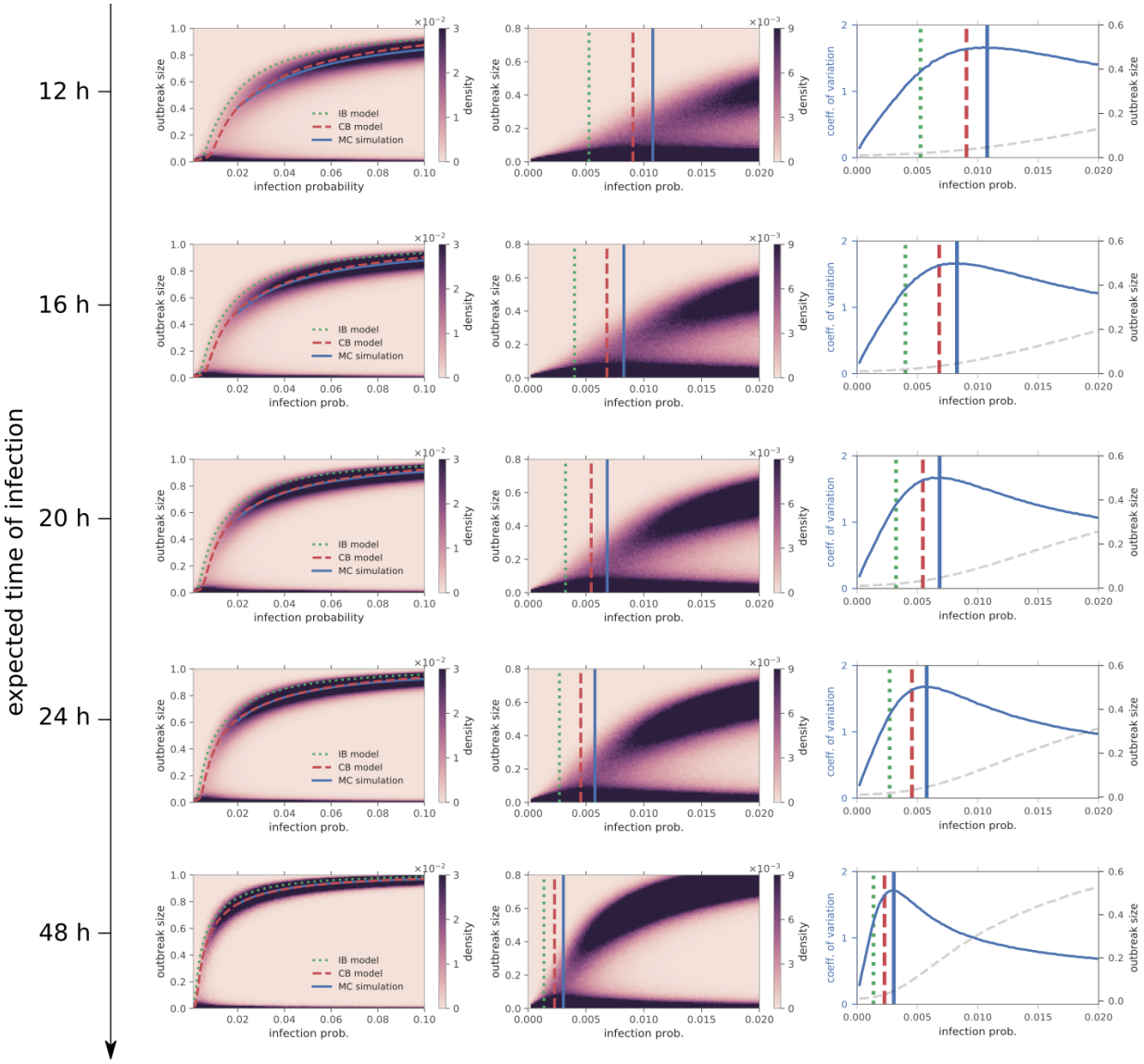

Figure B.5: Comparison between MC simulations and the mean-field models for different values of μ. Every row assumes a fixed recovery probability μ with decreasing values from top to bottom: $\mu = 4.63 \cdot 10^{-4}$, $\mu = 3.47 \cdot 10^{-4}$, $\mu = 2.78 \cdot 10^{-4}$, $\mu = 2.31 \cdot 10^{-4}$, and $\mu = 1.16 \cdot 10^{-4}$. These values correspond to an expected infection period of 12, 16, 20, 24, and 48 hours. Left column, middle column and right column correspond to Fig. 5.8, Fig. 5.10A, and Fig. 5.10B, respectively, of Chap. 5.

Next, we complete the analysis in Fig. 5.12 of Chap. 5 on the epidemic threshold for the example of the animal trade network. Again, we assume $\mu = 1/28$ and provide in Fig. B.6 a detailed analysis of all federal states, excluding the city states Berlin, Hamburg, and Bremen. The left and middle column in Fig. B.6, provide a similar analysis to Fig. 5.10A, and Fig. 5.10B of Chap. 5, respectively. In other words, we present the distribution of outbreak sizes for different values of the infection probability β (left column) from which we derive the coefficient of variation c_v (blue line, middle column). The right column presents values of c_v that are close to the peak and a quadratic fit (green line, right column) that determines the numerical estimation of the critical infection probability (blue vertical line). This value can be compared to spectral estimations from the mean-field models. In agreement with previous results, we find that the criticality condition in Eq. (5.14) of the CB model (see Chap. 5) improves previous results of the IB approach.

Finally, we provide an additional analysis of the epidemic threshold for the cattle trade data with $\mu = 1/120$. Results in Fig. B.7 are akin to our previous analysis in Fig. B.6, except for Saarland (SL). Here, the spectral condition in Eq. (5.14) of the CB model predicts that every outbreak remains localized, i.e. $\beta_{\text{crit.}}^{\text{CB}} = 1$, whereas MC simulations suggest a transition to global epidemics, hence $\beta_{\text{crit.}}^{\text{MC}} < 1$. We attribute the inconsistency to the small size of the network (26 nodes). The spectral approach assumes implicitly an infinitely large network, which is clearly violated in this case.

We summarize the results for $\mu = 1/120$ in Fig. B.8, akin to Fig. 5.13 of Chap. 5. The risk map in Fig. B.8A visualizes the spatial variability of the outbreak risk and each group of blue, red and green bars in Fig. B.8B provide a quantitative comparison between MC results, the CB model and the IB approach, respectively.

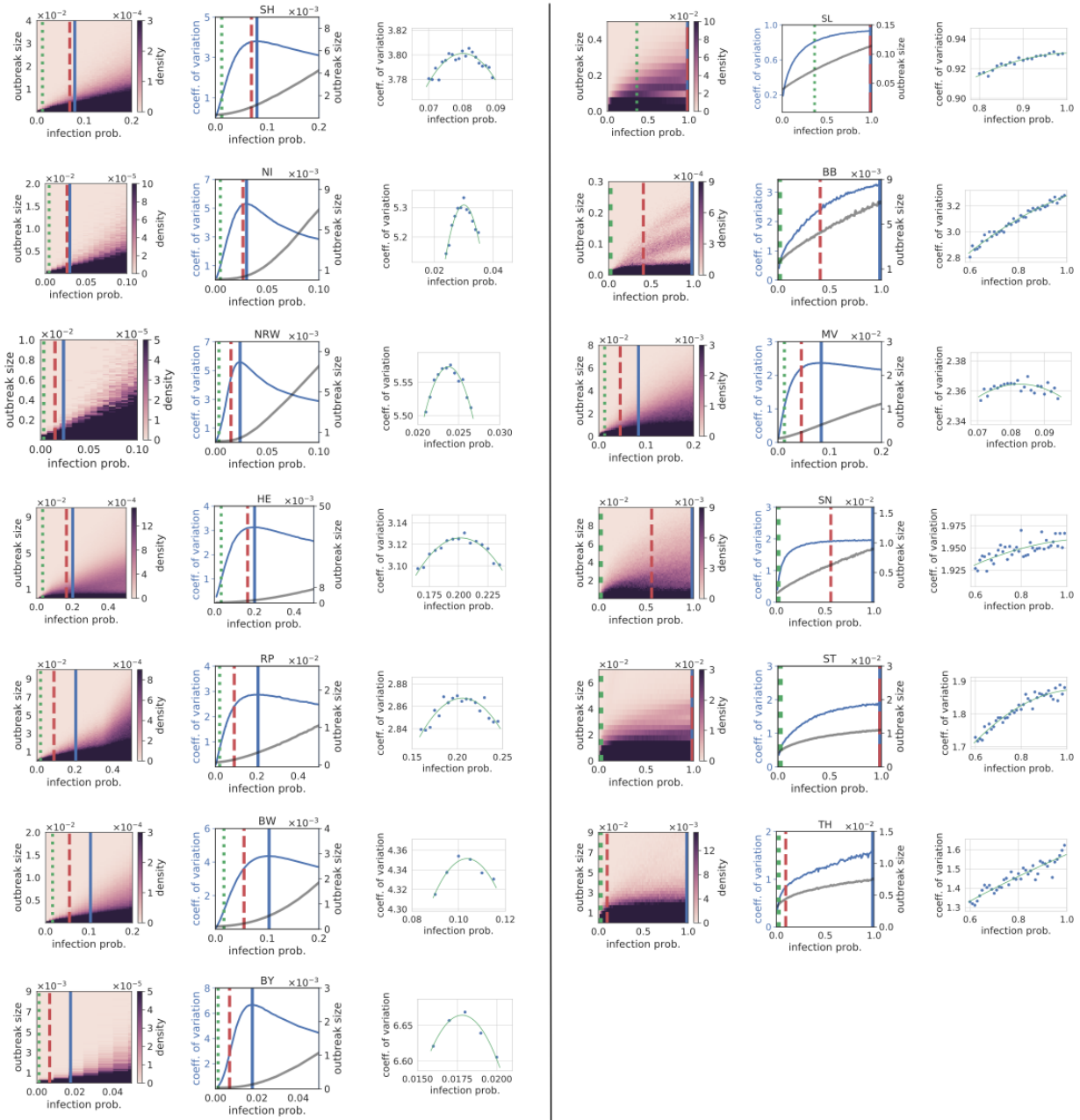

Figure B.6: Detailed analysis of the epidemic threshold for the cattle trade network with $\mu = 1/28$. The central black line separates the figure into two panels. Every row (in a panel) provides results for a single federal state in Germany, excluding the city states Berlin, Bremen, and Hamburg. Left column: outbreak size distribution as a function of the infection probability β. Vertical blue, red and green lines mark the critical infection probability according to MC simulations, the CB, and IB model, respectively. Middle column: Coefficient of variation c_v (blue line, left axis), mean outbreak size (gray dashed line, right axis). Right column: Selected values of c_v for infection probabilities close to the critical value. Quadratic fit (green line) estimates the maximum and hence $\beta_{\text{crit.}}$.

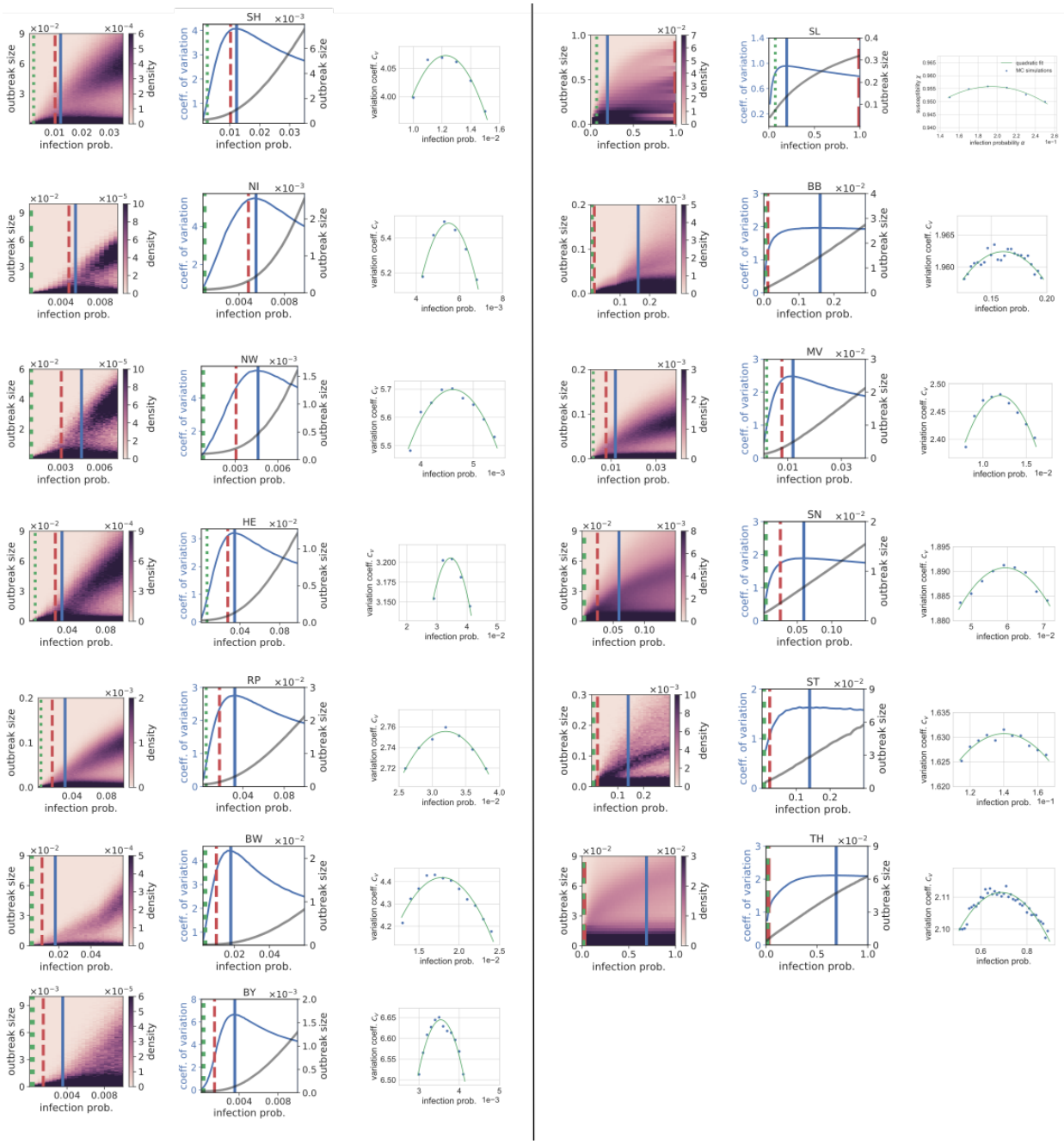

Figure B.7: Detailed analysis of the epidemic threshold for the cattle trade network with $\mu = 1/120$. The analysis is akin to Fig. B.6.

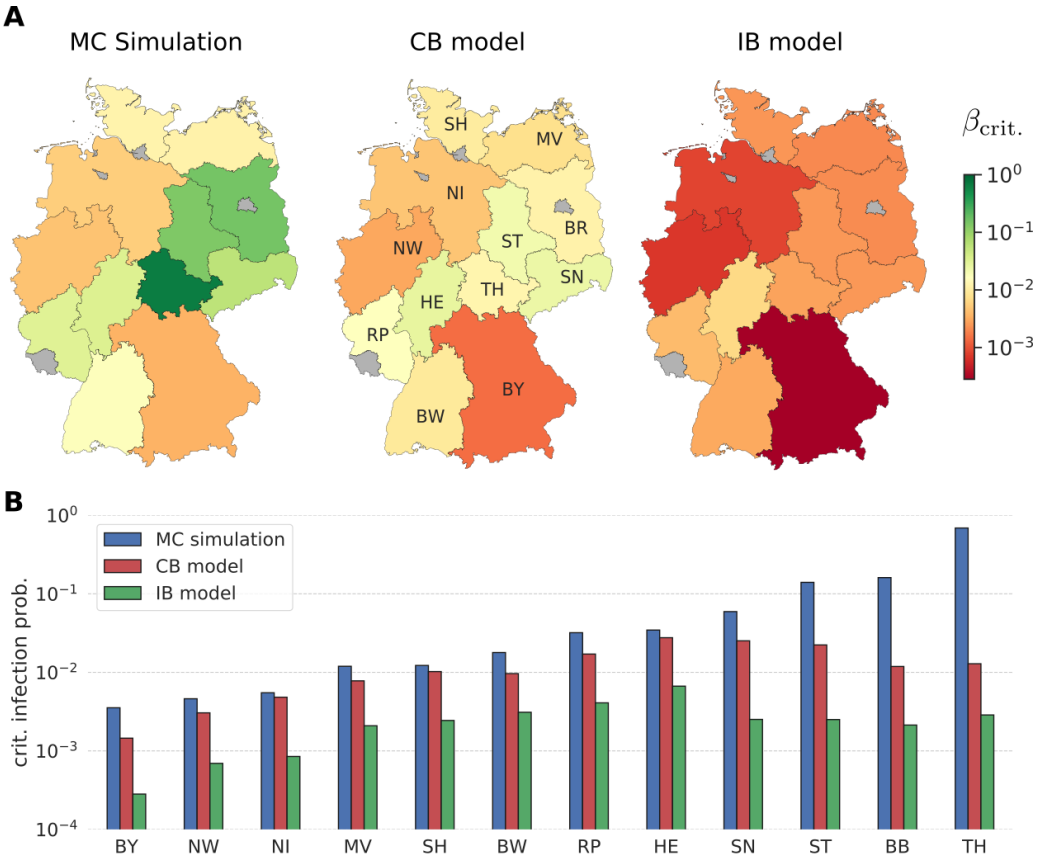

Figure B.8: **A**: Spatial variation of epidemic risk for $\mu = 1/120$. The critical infection probability β_{crit} determines the color of the federal states (see Fig. B.7 for details). The visualization is akin to Fig. 5.13 of Chap. 5.

C

Continuous-time model

In this appendix, we will focus on a continuous-time contact based model akin to Chap. 5. We will derive first derive the dynamic equations in Sec. C.1, then we determine the epidemic threshold from a stability analysis of the disease-free fix point in Sec. C.2. Moreover, we will link the continuous-time results to existing models and in particular to the *edge-based compartmental model* and *message-passing* in Sec. C.3 and Sec. C.4, respectively. Finally, we sketch a first attempt to provide a general model for epidemics on temporal network in Sec. C.5, which assumes a non-Markovian infection and recovery process.

The results in this appendix represent an ongoing and yet unpublished research project.

C.1 Derivation of the continuous time limit

We derive the continuous time limit starting from the discrete-time dynamics Eq. (5.4)-(5.6). To this end, we first assume a small but finite sampling interval Δt and replace the infection and recovery probabilities with their corresponding (continuous-time) rates $\beta \to \beta\Delta t$ and $\mu \to \mu\Delta t$, respectively. Moreover, we focus on the open interval between two successive network snapshots first, in order to avoid complications with discontinuous jumps. Our first dynamic Eq. (5.4) rewritten in terms of the sampling interval Δt reads

$$\theta_{k\to l}(t+\Delta t) = \theta(t)_{k\to l} - \beta\Delta t a_{k\to l}(t) I_{k\to l}(t). \tag{C.1}$$

In the continuous time limit $\Delta t \to 0$, Eq. (C.1) leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta(t)_{k\to l} = -\beta a_{k\to l}(t) I_{k\to l}(t). \tag{C.2}$$

Next, we focus on S and rewrite Eq. (5.6) in terms of the sampling interval:

$$S_{k\to l}(t+\Delta t) = z_k \prod_{j\in\mathcal{N}_k\setminus l} \theta_{j\to k}(t+\Delta t) \tag{C.3a}$$

$$S_{k\to l}(t+\Delta t) = z_k \prod_{j\in\mathcal{N}_k\setminus l} \left[\theta_{j\to k}(t) - \beta\Delta t a_{k\to l}(t) I_{j\to k}(t)\right]. \tag{C.3b}$$

In the second step (Eq. (C.3b)) we applied the definition of $\theta_{j\to k}(t+\Delta t)$ from Eq. (5.4). For a sufficiently small sampling interval Δt, such that $\theta_{j\to k}(t) \gg \beta\Delta t a_{k\to l}(t) I_{k\to l}(t)$ we can linearize Eq. (C.3b) to obtain:

$$S_{k\to l}(t+\Delta t) = z_k \prod_{j\in\mathcal{N}_k\setminus l} \theta_{j\to k}(t) \cdot \left[1 - \sum_{j'\in\mathcal{N}_k\setminus l} \frac{\beta\Delta t a_{k\to l}(t) I_{j'\to k}(t)}{\theta_{j'\to k}(t)}\right] \tag{C.4a}$$

$$S_{k\to l}(t+\Delta t) = S_{k\to l}(t)\left[1 - \beta\Delta t \sum_{j'\in\mathcal{N}_k\setminus l} a_{k\to l}(t)\frac{I_{j'\to k}(t)}{\theta_{j'\to k}(t)}\right] \tag{C.4b}$$

In Eq. (C.4b) we inserted the definition of $S_{k\to l}(t)$ from Eq. (5.6) and this leads directly to our second continuous-time dynamic equation

$$\frac{\mathrm{d}}{\mathrm{d}t} S_{k\to l}(t) = -\beta S_{k\to l}(t) \sum_{j\in\mathcal{N}_k\setminus l} a_{k\to l}(t)\frac{I_{j\to k}(t)}{\theta_{j\to k}(t)}. \tag{C.5}$$

The quotient $I_{j\to k}(t)/\theta_{j\to k}(t)$ can be interpreted as the conditional probability that j is infected *given* that cavity node k has not yet contracted the disease from j. It is worth noting that Eq. (C.5) is well defined because we start from the initial condition $\theta_{j\to k}(t) = 1$ for all edges $k \leftarrow j$ and Eq. (C.9) asserts positivity for $\theta_{j\to k}(t)$ for all finite observation times t. The remaining discrete-time Eq. (5.5), can be immediately written down in terms of a difference quotient $\Delta X(t) = [X(t+\Delta t) - X(t)]/\Delta t$:

$$\frac{\Delta I_{k\to l}(t)}{\Delta t} = (-\mu - \beta a_{k\to l}(t) + \mu\beta\Delta t a_{k\to l}(t)) I_{k\to l}(t) - \frac{\Delta S_{k\to l}(t)}{\Delta t} \tag{C.6}$$

In the continuous time limit, the higher order term $\mu\beta\Delta t a_{k\to l}(t)$ disapears and making use of our previous result Eq. (C.5) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} I_{k\to l}(t) = [-\mu - \beta a_{k\to l}(t)] I_{k\to l}(t) - \frac{\mathrm{d}}{\mathrm{d}t} S_{k\to l}(t). \tag{C.7}$$

At last it is also instructive to formulate the dynamic equation for $R_{k\to l}(t)$, i.e., the probability that k has recovered at time t without transmitting the disease to cavity node l. The $R_{k\to l}(t)$ can only increase over time and the corresponding in-flux at time t is given by (i) the probability $I_{k\to l}(t)$ that node k is in state I and has not infected it's neighbor l as well as (ii) the probability $\mu\Delta t$ to recover within Δt. With this, we obtain:

$$R_{k\to l}(t+\Delta t) = R_{k\to l}(t) + \mu\Delta t I_{k\to l}(t). \tag{C.8}$$

The corresponding continuous-time equation reads

$$\frac{\mathrm{d}}{\mathrm{d}t}R_{k\to l}(t) = \mu I_{k\to l}(t). \tag{C.9}$$

Unlike the discrete-time model, it is now obvious that the dynamic Eq. (C.2), (C.5), (C.7), and (C.9) satisfy the conservation condition

$$\theta_{k\to l}(t) = S_{k\to l}(t) + I_{k\to l}(t) + R_{k\to l}(t) \tag{C.10}$$

at every time t. Moreover, we can rescale time according to $\mu t \to t$ and rewrite the continuous time contact-based model in terms of the dimensionless epidemic parameter $\gamma = \beta/\mu$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta_{k\to l}(t) = -\gamma a_{k\to l}(t) I_{k\to l}(t) \tag{C.11a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}S_{k\to l}(t) = -\gamma S_{k\to l}(t) \sum_{j\in\mathcal{N}_k\setminus l} a_{j\to k}(t)\frac{I_{j\to k}(t)}{\theta_{j\to k}(t)} \tag{C.11b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}I_{k\to l}(t) = -[1+\gamma a_{k\to l}(t)] I_{k\to l}(t) - \frac{\mathrm{d}}{\mathrm{d}t}S_{k\to l}(t) \tag{C.11c}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}R_{k\to l}(t) = I_{k\to l}(t) \tag{C.11d}$$

Similar to the derivation in Sec. 5.3 in the main article, we then return to node-centric quantities, i.e., the probability that a given node l is susceptible, infected or recovered. The time-continuous equivalent formulation to Eq. (5.7a)-(5.7c) reads

$$P_l^S(t) = z_l \prod_{k\in\mathcal{N}_l} \theta_{k\to l}(t) \tag{C.12a}$$

$$P_l^I(t) = 1 - P_l^S(t) - P_l^I(t) \tag{C.12b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}P_l^I(t) = P_l^I(t). \tag{C.12c}$$

Finally, we choose initial conditions as in Sec. 5.3 of the main article, i.e.,

$$P_l^S(0),\, P_l^I(0),\, P_l^R(0) = z_l,\, 1-z_l,\, 0 \tag{C.13a}$$

$$\theta_{k\to l},\, S_{k\to l}(0),\, I_{k\to l}(0),\, R_{k\to l}(0) = 0,\, z_l,\, 1-z_l,\, 0 \tag{C.13b}$$

for all nodes $l \in \mathcal{N}$ and edges $(k,l) \in \mathcal{E}$.

C.2 Spectral properties of the continuous time model

Similar to the discrete time analysis in Sec. 5.3.2, we determine the spectral properties by linearizing Eqs. (C.11) around the disease free state, i.e., we assume $\theta_{k\to l}(t) = 1 - \delta_{k\to l}(t)$, where $\delta_{k\to l}(t) \ll 1$ as well as $I_{k\to l} \ll 1$. With this, we

linearize Eq. (C.11b) and obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}S_{k\to l}(t) = -\gamma\left[1 - \sum_{j\in\mathcal{N}_k\setminus l}\delta_{j\to k}(t)\right]\cdot\sum_{j\in\mathcal{N}_k\setminus l}a_{k\to l}(t)\frac{I_{j\to k}(t)}{1-\delta_{k\to l}(t)} \tag{C.14a}$$

$$= -\gamma\sum_{j\in\mathcal{N}_k\setminus l}a_{k\to l}(t)I_{j\to k}(t) \tag{C.14b}$$

In Eq. (C.14b) we kept only linear terms in $\delta_{k\to l}(t)$ and $I_{j\to k}(t)$. This results allows us to decouple the set of dynamic equations and express the conservation condition Eq. (C.11c) only in terms of $I_{k\to l}(t)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}I_{k\to l}(t) = [-1-\gamma a_{k\to l}(t)]I_{k\to l}(t) + \gamma\sum_{j\in\mathcal{N}_k\setminus l}a_{k\to l}(t)I_{j\to k}(t). \tag{C.15}$$

Similar to the discrete-time derivation in Sec. 5.4 we vectorize Eq. (C.15) using the time-dependent non-backtracking operator $\boldsymbol{B}(t)$, the diagonal matrix $\mathrm{diag}(\boldsymbol{a}(t))$ with elements $a_{k\to l}(t)$ on the diagonal and the unit matrix $\mathbb{1}$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{I}(t) = \left[-\mathbb{1} - \gamma\mathrm{diag}(\boldsymbol{a}(t)) + \gamma\boldsymbol{B}(t)\right]\boldsymbol{I}(t). \tag{C.16}$$

Structurally Eq. (C.16) corresponds to the recursive (discrete-time) Eq. (5.12) from Chap. 5. The only difference is that the higher order term $\beta a_{k\to l}\mu$ does not appear because, the simultaneous event of infection and recovery does not need to be accounted for in the continuous time formulation.

Within the open interval $[t_n, t_{n+1})$ where the boundaries t_n and t_{n+1}, respectively, mark subsequent change points of the network topology, we integrating Eq. (C.16) and obtain

$$\boldsymbol{I}(t_{n+1}) = \boldsymbol{M}_n(\gamma)\boldsymbol{I}(t_n) \tag{C.17a}$$

$$\boldsymbol{M}_n(\gamma) = \exp\left(\int_{t_n}^{t_{n+1}}\mathrm{d}\tau[-\mathbb{1} - \gamma\mathrm{diag}(\boldsymbol{a}(\tau)) + \gamma\boldsymbol{B}(\tau)]\right). \tag{C.17b}$$

Using the initial condition $\boldsymbol{I}(0)$, we can formally state the explicit solution as follows:

$$\boldsymbol{I}(T) = \prod_{n=0}^{N-1}\boldsymbol{M}_n(\gamma)\boldsymbol{I}(0). \tag{C.18}$$

Here, N is the total number of discrete changing points of the network topology. Following [247] we can state the propagator $\boldsymbol{M}(\gamma) = \prod_{n=0}^{N-1}\boldsymbol{M}_n(\gamma)$ in a compact notation using Dyson's time ordering operator $\mathcal{T}\boldsymbol{B}(\tau_1)\boldsymbol{B}(\tau_2) = \boldsymbol{B}(\tau_1)\boldsymbol{B}(\tau_2)\Theta(\tau_1 - \tau_2) + \boldsymbol{B}(\tau_2)\boldsymbol{B}(\tau_1)\Theta(\tau_2-\tau_1)$, where $\Theta(x)$ denotes the Heaviside function:

$$\boldsymbol{M}(\gamma) = \mathcal{T}\exp\left(\int_0^t\mathrm{d}\tau\left[-\mathbb{1} - \gamma\mathrm{diag}(\boldsymbol{a}(\tau)) + \gamma\boldsymbol{B}(\tau)\right]\right) \tag{C.19}$$

The continuous-time propagator in Eq. (C.19) is akin to our discrete-time matrix in Eq. (5.13) of Chap. 5. However, here we assumed a continuous-time SIR process on a temporal network where transition events between subsequent network

configurations occur at arbitrary time points. Any small initial perturbation will decrease exponentially if the largest eigenvalue, i.e., the spectral radius satisfies $\lambda_1[\boldsymbol{M}(\gamma)] < 1$. This result corresponds to [247] where the epidemic propagator $\boldsymbol{M}(\gamma)$ has been derived within the individual-based framework and reads

$$\boldsymbol{M}(\gamma) = \mathcal{T} \exp\left(\int_0^t \mathrm{d}\tau \left[-\mathbb{1} + \gamma \boldsymbol{A}(\tau)\right]\right) \tag{C.20}$$

In Eq. (C.20) $\boldsymbol{A}(\tau)$ is the time dependent adjacency matrix and $\mathbb{1}$ is a unit matrix with the same dimension.

In many cases the temporal network was sampled with equidistant time steps Δt as in the data set [50] that is used in the main article. Here, we can simplify the propagator Eq. (C.20) and obtain a formulation that corresponds to an individual-based result, first derived in [178]:

$$\boldsymbol{M}(\gamma) = \prod_{n=0}^{N-1} \exp\left(\Delta t \left[-\mathbb{1} - \gamma \mathrm{diag}(\boldsymbol{a}(n)) + \gamma \boldsymbol{B}(t_n)\right]\right) \tag{C.21}$$

In the special case of a static underlying topology, we identify $\mathrm{diag}(\boldsymbol{a}(t)) \equiv \mathbb{1}$ and $\boldsymbol{B}(t) \equiv \boldsymbol{B}$, similar to our discussion at the end of Sec. 5.4. The linearized dynamic Eq. (C.16) simplifies to

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{I}(t) = \left[-(1+\gamma)\mathbb{1} + \gamma \boldsymbol{B}\right] \boldsymbol{I}(t) \tag{C.22}$$

Finally, Eq. (C.22) is asymptotically stable if the largest eigenvalue λ_1 of the infection operator $\boldsymbol{M}(\gamma) = (1+\gamma)\mathbb{1} + \gamma \boldsymbol{B}$ is negativ. Hence, we recover the continuous time threshold as previously derived within the more general message-passing framework on static networks [18, 103]:

$$\frac{\gamma}{\gamma + 1} = \frac{1}{\lambda_1(\boldsymbol{B})}. \tag{C.23}$$

For non-Markovian infection and recovery processes the generalized criticality condition reads $T = 1/\lambda_1(\boldsymbol{B})$, where the transmissibility T is given by

$$T = \int_0^\infty s(\tau) \left(\int_\tau^\infty r(\tau')\mathrm{d}\tau'\right) \mathrm{d}\tau \tag{C.24}$$

Intuitively, T can be interpreted as the probability that a newly infected node transmits the disease to a given neighbor prior to recovery [18, 131]. Within this general formulation $s(\tau)\mathrm{d}\tau$ is the probability that an infected node passes the disease to a neighbor within a time interval $[\tau, \tau + \mathrm{d}\tau]$ after contracting the infection. Similarly, we define the probability $r(\tau)\mathrm{d}\tau$ that a node recovers in the interval $[\tau, \tau + \mathrm{d}\tau]$ after it has been infected. For a constant infection and recovery rate, i.e., for a Markovian dynamics as analyzed in the main article, we find $s(\tau) = \beta \exp(-\beta\tau)$ and $s(\tau) = \mu \exp(-\mu\tau)$. This particularly simple and widely studied choice then leads to $T = \beta/(\beta + \mu)$ and thus to Eq. (C.23).

For temporal networks we cannot separate in general the transmissibility T from the network topology in order to find a similarly elegant results like Eq. (C.23). The reason is the complex interplay between timing of contacts and the disease dynamics and as a consequence the transmissibility is edge- and time-dependent even in the simple Markovian case.

C.3 Average over random graphs and the EBC model

An equally important research branch in theoretical epidemiology focuses not on a single realization of a graph but on an ensemble of random networks. In what follows, we will indicate ensemble averaged quantities with $\langle\cdot\rangle$. Arguably, the most prominent class of random networks is termed *configuration model* [248, 249] and the idea goes as follows: Suppose we want to study the effect of the degree distribution on the spread of infections [131]. To begin with, we have to create an ensemble of networks with the same degree distribution and that are otherwise maximally random. This can be done according to the Molloy-Reed algorithm [248], which begins with a set of N vertices. To each node we assign a number of k (undirected) *stubs*, i.e., edges with no target node that are drawn independently from the given degree distribution $p(k)$. In the next step, we connect two randomly chosen stubs which then form a proper edge between the corresponding nodes. The step is repeated until no more stubs are available and if initially the number of stubs were found to be odd then the we would replace one node repeatedly until the sum is even.

The configuration model has been studied extensively (see [36, 137] for an overview) and forms one of the cornerstones of modern network theory. A particularly accurate and compact model of epidemic spreading on this class of random networks has been proposed in [104] and is termed *edge-based compartmental model* (EBCM). The original work focused not only on the classical configuration model but, also on different classes of random, temporal networks. Since then several extensions have been proposed, such as non-Markovian recovery dynamics [17] and arbitrary initial conditions [234]. Moreover, studies in [250, 251] investigated links to other existing models such as *pair-approximations* [93], the *effective degree model* [100] and the *message-passing* [18]. In this section, we will discuss how the contact-based approach relates to the EBC model.

In what follows, we will derive the EBCM formulation [234] starting from the CB model for static random networks generated by the configuration model and assuming unweighted and undirected edges as in Chap. 3. Thereby we follow closely the approach outlined in [18] and start with the following crucial observation: The state of a given edge $k \to l$ in a single realization of a graph displays a characteristic trajectory in state space, i.e., a time-dependent curve given by the triple

$$[\theta_{k\to l}(t), S_{k\to l}(t), I_{k\to l}(t), R_{k\to l}(t)]\,. \tag{C.25}$$

As we perform an average over the ensemble of graphs, this edge will assume every position within a network. As a consequence the averaged state trajectory is identical to the one that we would obtain, if we had started with a different edge initially and then performed the average. In other words, it is sufficient to determine the ensemble averaged probabilities for *one representative edge*:

$$\theta(t) \equiv \langle \theta_{k\to l}(t) \rangle \tag{C.26a}$$
$$\psi^S(t) \equiv \langle S_{k\to l}(t) \rangle \tag{C.26b}$$
$$\psi^I(t) \equiv \langle I_{k\to l}(t) \rangle \tag{C.26c}$$
$$\psi^R(t) \equiv \langle R_{k\to l}(t) \rangle. \tag{C.26d}$$

Next, we derive the dynamic equation for the ensemble averaged probability $\psi^S(t)$ that a randomly chosen neighbor of the cavity node is susceptible at time t (cf. Eq. (5.6)):

$$\psi^S(t) \equiv \langle S_{k\to l}(t) \rangle \tag{C.27a}$$
$$= \left\langle z_k \prod_{j\in\mathcal{N}_k\setminus l} \theta_{j\to k}(t) \right\rangle. \tag{C.27b}$$

A crucial property of large networks that are generated by the configuration model is they are locally tree-like in the sense that the average length of the smallest cycle diverges with increasing network size. Hence, we can assume in the thermodynamic limit that different branches emerging from k can be treated independently. The average over the product thus equals the product over averages. Moreover, we remember that ensemble averaged dynamic quantities are equal for all edges and in particular $\theta_{j\to k}(t) \equiv \theta(t)$. With this the product in Eq. (C.27b) simplifies to $[\theta(t)]^{k_e}$, where k_e is the average number of next neirest neighbors, or equally, the *excess degree*. From a given degree distribution p_n in the configuration model, we can derive the excess degree distribution q_n according to $q_n = (n+1)p_{n+1}/k$ [249], where $k = \langle n \rangle$ denotes the average degree. Finally, we make use of the corresponding generating function $G_1(x) = \sum_n q_n x^n$ and thus rewrite Eq. (C.27b) to

$$\psi^S(t) = \langle z_k \rangle \left\langle \prod_{j\in\mathcal{N}_k\setminus l} \theta_{j\to k}(t) \right\rangle \tag{C.28a}$$
$$= z \sum_{n=0}^{N} q_n [\theta(t)]^n \tag{C.28b}$$
$$= z G_1(\theta(t)). \tag{C.28c}$$

Here, $z = \langle z_k \rangle$ denotes the ensemble averaged probability that node k is initially susceptible. Similarly, we derive the remaining dynamic equations from the ensemble average of Eq. (5.4) and Eq. (5.6). With the shorthand notation $\Delta\psi^S(t) =$

$\psi^S(t) - \psi^S(t+1)$, we obtain:

$$\theta(t+1) = \theta(t) - \beta I(t) \tag{C.29a}$$
$$\psi^S(t+1) = zG_1(\theta(t+1)) \tag{C.29b}$$
$$\psi^I(t+1) = (1-\mu)(1-\beta)\psi^I(t) + \Delta\psi^S(t) \tag{C.29c}$$
$$\psi^R(t+1) = \psi^R(t) + \mu\psi^I(t). \tag{C.29d}$$

Following the main article, we choose the following initial condition:

$$\theta(0), \psi^S(0), \psi^I(0), \psi^R(0) = 1, z, 0, 0. \tag{C.30}$$

Next, we take the continuous time limit as described in Sec. C.1 in order to arrive at the EBC model [234]. For static random networks, Eq. (C.29a)-(C.29d) translate to the following set of coupled ODEs, where we drop the time dependence of the dynamic quantities for simplicity:

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta = -\beta\psi^I \tag{C.31a}$$
$$\psi^S = zG_1(\theta) \tag{C.31b}$$
$$\psi^I = \theta - \psi^S - \psi^R \tag{C.31c}$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\phi_R = \mu\psi^I. \tag{C.31d}$$

We can now substitute ψ^I in Eq. (C.31a) with (C.31d) and thus arrive at the following relation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\psi^R = -\frac{\mu}{\beta}\frac{\mathrm{d}}{\mathrm{d}t}\theta \tag{C.32a}$$
$$\psi^R = \mu(1-\theta)/\beta. \tag{C.32b}$$

Here, we integrated Eq. (C.32a) using the initial condition $\theta(0) = 1$ and $\psi^R(0) = 1$. Together with Eq. (C.31a)-(C.31c), we finally recover the ECB model as formulated in [234]:

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta = -\beta\theta + \beta z G_1(\theta) - \mu(1-\theta) \tag{C.33}$$

From $\theta(t)$ we can directly infer the probability $P^S(t)$ that our randomly chosen cavity node is susceptible. For that we observe that a node is susceptible if was susceptible initially with probability z and has not contracted the disease from any of it's n neighbors. The corresponding probability reads $z[\theta(t)]^n$, where n is now drawn from the degree distribution p_n. We denote the corresponding generating function with $G_0(x) = \sum_n p_n x^n$ and thus the average over all n leads to Eq. (C.34a). From Eq. (C.12b) and (C.12c) we can directly formulate the remaining equations

for the corresponding probabilities $P^I(t)$ and $P^R(t)$, respectively:

$$P^S = zG_0(\theta) \tag{C.34a}$$

$$P^I = 1 - P^S - P^R \tag{C.34b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}P^R = \mu P^I. \tag{C.34c}$$

We close the section with a linear stability analysis of Eq. (C.33). Similar to the main article, we start with a small initial perturbation: $\theta = 1 - \delta$ and $z = 1 - \epsilon$ with $\delta, \epsilon \ll 1$. We then expand the generating function $G_1(1-\delta)$ to the first order in δ:

$$G_1(1-\delta) = \sum_n p_n (1-\delta)^n \tag{C.35a}$$

$$= 1 + \langle n \rangle_q (1-\delta) + \mathcal{O}(\delta^2) \tag{C.35b}$$

Here, we used two properties of the generating function, namely $G_1(1) = \sum_n q_n = 1$ and $G_1'(1) = \sum_n n q_n = \langle n \rangle_q$, where $\langle n \rangle_q = k_e$ denotes the mean excess degree. With this, the linearized Eq. (C.33) around the disease-free stable fix-point reads:

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta = (\beta + \beta\langle n \rangle_q + \mu)\delta \tag{C.36}$$

From Eq. (C.36) we can easily see that a transition occurs from local to global outbreaks if $\beta + \beta\langle n \rangle_q + \mu > 0$. Commonly, $\langle n \rangle_q$ is expressed in terms of the degree. For that we use $\langle n \rangle_q = \sum_n n q_n$ and $\langle n \rangle = \sum_n n p_n$ together with $q_n = (n-1)p_n/\langle n \rangle$ [249] to find $\langle n \rangle_q = (\langle n^2 \rangle - \langle n \rangle)/\langle n \rangle$ and thus we recover the well-known criticality condition from [131, 252]:

$$\frac{\beta}{\beta + \mu} = \frac{\langle n^2 \rangle - \langle n \rangle}{\langle n \rangle} \tag{C.37}$$

This relation is akin to the static network limit in Eq. (5.18) of Chap. 5 and the continuous-time result in Eq. (C.23), where we studied a single realization of a graph and hence expressed the right hand side of Eq. (C.37) through the spectral radius $\lambda_1(\boldsymbol{B})$ of the non-backtracking matrix $\boldsymbol{B}$.

C.4 Relation to message-passing

In the seminal work of Karrer et al. [18], the authors proposed a general model for SIR spreading processes on sparse networks with non-Markovian infection and recovery dynamics. The integro-differential formulation in [18] is a foundation of our CB model from Chap. 5 and therefore, we will discuss in this section the relation to the message-passing approach. For that we assume a static and finite graph G and, as a first step, transform the dynamic equations in Eq. (C.2), (C.5), (C.7) and, (C.9) that define the continuous-time CB model to an integro-differential

equation. To this end, we notice first that Eq. (C.7) is of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}I_{k\to l}(t) = -\lambda_{k\to l}(t)I_{k\to l}(t) - \frac{\mathrm{d}}{\mathrm{d}t}S_{k\to l}(t). \tag{C.38}$$

For notational convenience, we use the short-hand notation $\lambda_{k\to l}(\tau) = \mu + \beta a_{k\to l}(\tau)$ and $\Lambda_{k\to l}(t, t_k) = \exp[-\int_{t_k}^{t} \lambda_{k\to l}(\tau)\mathrm{d}\tau]$. The former denotes the probability that node k recovers or infects the cavity node l within the time interval $[\tau, \tau + \mathrm{d}t)$ after contracting the disease and the latter corresponds to the probability that no such event took place between the time of infection and the observation time t_k and t, respectively. Here, we denote the absolute and relative time after infection with t and τ, respectively. Together with the initial condition $I_{k\to l}(0) = 1 - z_k$ the solution to the differential equation is given by

$$I_{k\to l}(t) = (1 - z_k)\Lambda_{k\to l}(t, 0) + \int_0^t \left[-\frac{\mathrm{d}}{\mathrm{d}t_k}S_{k\to l}(t_k)\right] \Lambda_{k\to l}(t, t_k)\mathrm{d}t_k. \tag{C.39}$$

In words Eq. (C.39) states that node k has contracted the disease but, not infected it's neighbor l by absolute time t if (i) node k was infected initially but, has neither recovered nor passed the infection or (ii) it was susceptible initially, contracted the disease at time t_k and has then neither recovered nor infected it's neighbor up to the observation time t.

Next, we integrate Eq. (C.2) and using the initial condition $\theta_{k\to l}(0) = 1$ we get

$$1 - \theta_{k\to l}(t) = \int_0^t \beta a_{k\to l}(t')I_{k\to l}(t')\mathrm{d}t' \tag{C.40a}$$

$$1 - \theta_{k\to l}(t) = (1 - z_k)\int_0^t \mathrm{d}t' f_{k\to l}(t'|0) + \tag{C.40b}$$
$$+ \int_0^t \int_0^{t'} \mathrm{d}t'\mathrm{d}t_k \left[-\frac{\mathrm{d}}{\mathrm{d}t_k}S_{k\to l}(t_k)\right] f_{k\to l}(t'|t_k)$$

In Eq. (C.40b) we used $I_{k\to l}(t)$ from Eq. (C.39) and we also introduced the transmission probability $f(t'|t_k) = \beta a_{k\to l}(t')\Lambda_{k\to l}(t', i_i)$: Given that node k contracted the infection at absolute time t_k, $f(t'|t_k)$ gives the probability that the same node passes the disease to it's neighbor l at absolute time t'. In the context of static networks the quantity $\int_{t_k}^{\infty} f_{k\to l}(t'|t_k)\mathrm{d}t'$ is frequently referred to as *transmissibility* and plays a crucial role in linking epidemic spreading to a percolation process [18, 131]. Note that the transmissibility can be smaller than one as node k might recover before passing on the infection and for temporal networks, unlike the static case, the value is edge- and time-dependent.

In order to arrive at the message passing formulation [18], we reformulate Eq. (C.40b) for a static underlying topology. With $a_{k\to l}(t) \equiv 1$ the transmission probability $f_{k\to l}(t|t_k) \to f(\tau)$ depends only on the relative time $\tau = t - t_k$ after

infection and becomes an identical function for all edges $k \to l$. With this we get

$$1 - \theta_{k\to l}(t) = (1 - z_k) \int_0^t \mathrm{d}\tau f(\tau) + \int_0^t \int_0^\tau \mathrm{d}\tau \mathrm{d}\tau_i f(\tau - \tau_i) \left[-\frac{\mathrm{d}}{\mathrm{d}\tau_i} S_{k\to l}(\tau_i) \right] \tag{C.41a}$$

With $f(\tau) = \beta \exp[-\gamma\tau]$ and $\gamma = \beta + \mu$ we integrate the second term in Eq. (C.41a) by parts to obtain

$$\beta \int_0^t \mathrm{d}\tau \exp[-\gamma\tau] \int_0^\tau \mathrm{d}\tau_i \exp[\gamma\tau_i] \left[-\frac{\mathrm{d}}{\mathrm{d}\tau_i} S_{k\to l}(\tau_i) \right] = \tag{C.42a}$$

$$- \beta \int_0^t \mathrm{d}\tau S(\tau) + z \int_0^t \mathrm{d}\tau \beta \exp[-\gamma\tau] +$$

$$+ \beta \int_0^t \mathrm{d}\tau \exp[-\gamma\tau] \int_0^\tau \mathrm{d}\tau_i \exp[\gamma\tau_i] S_{k\to l}(\tau_i)$$

$$= z \int_0^t \mathrm{d}\tau f(\tau) + \int_0^t \mathrm{d}\tau f(t - \tau) S_{k\to l}(\tau) \tag{C.42b}$$

In a final step we substitute $t - \tau \to \tau'$ in the second term and feed the result back into Eq. (C.41a). Hence, we arrive at the message passing formulation from [18]

$$\theta_{k\to l}(t) = 1 - \int_0^t \mathrm{d}\tau f(t)(1 - S_{k\to l}(t - \tau)) \tag{C.43}$$

With this we have linked the continuous-time CB model with a previously introduced message passing framework for general epidemic models in the case of a static underlying graph. The correspondence depends crucially on the particular choice of the transmissibility $f(\tau)$, i.e. exponentially decaying probability distributions from which we draw the recovery time and the time until the first infectious contact. Based on this derivation, we discuss a non-Markovian extension to our model in Sec. C.5.

C.5 General epidemic models on temporal networks

Much of the past epidemiologic research focused on Markovian dynamics that assume constant rates at which node becomes infected upon contact or recovers. As a consequence the duration of an infection is exponentially distributed and thus individuals recover most likely shortly after contracting the disease. This behaviour contradicts most real diseases where the duration time is typically peaked around a characteristic time, say a week, for instance.

Even though the Markovian assumption is convenient from a mathematical perspective, recent studies have shown that non-exponential distributions can significantly alter the expected trajectory of an outbreak [17, 18, 20]. So far few approaches exist that account for non-Markovian properties on finite static networks [18] and graph ensembles [16, 17]. Given the recent interest on graphs with time-vaying topologies, we propose in this section a general epidemic model

on temporal network that reduces to the continuous-time CB model in the case of Markovian dynamics and to the general MP formulation [18] for static graphs.

We start with the intermediate result Eq. (C.40b) from Sec. C.4 and differentiating with respect to t gives:

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta_{k\to l}(t) = (1-z_k)f_{k\to l}(t,0) + z\int_0^{t_k} \mathrm{d}t_k f_{k\to l}(t,t_k)\left[-\frac{\mathrm{d}}{\mathrm{d}t_k}\prod_{j\in\mathcal{N}_k\setminus l}\theta_{j\to k}(t_k)\right]. \tag{C.44}$$

In words, Eq. (C.44) states that the probability density $\dot{\theta}_{k\to l}(t)$ that node l contracts a disease from its neighbor k within the time interval $[t, t+\mathrm{d}t)$ evolves according to two processes: (1) Either k was initially infected and passes the disease at the observation time t or (2) it contracted the infection itself at time $t_k \in (0,t)$ before passing it to its neighbor l. Note that the derivative in the second term is negative and gives the probability that k is newly infected by one of its incident neighbors (except the cavity node l) within the interval $[t_k, t_k+\mathrm{d}t)$.

Unlike Sec. C.4 where we introduced the probability of transmission $f_{k\to l}(t|t_k)$ as a short-hand notation, we will start here from a general, i.e. potentially non-Markovian formulation:

$$f_{k\to l}(t,t_k) = \beta(t-t_k)a_{k\to l}(t)\Lambda_{k\to l}(t,t_k) \tag{C.45a}$$

$$\Lambda_{k\to l}(t,t_k) = \exp\left\{-\int_{t_k}^{t}\left[\beta(t-t_k)a_{k\to l}(t)+\mu(t-t_k)\right]\right\} \tag{C.45b}$$

Here, $\beta(t-t_k)$ and $\mu(t-t_k)$ are so-called hazard rates. The former, gives the probability that a node, already infected for a time $\tau = t-t_k$, transmits the disease to it's neighbor upon contact. Similarly $\mu(t-t_k)$ denotes the probability to recovers within a time interval $[\tau, \tau+\mathrm{d}t)$ after infection. Moreover, $\Lambda_{k\to l}(t,t_k)$ can be considered a survival probability in that it gives the probability that node k, infected at absolute time t_k, neither recovers nor does it infect its neighbor l up to the observation time t. The overall probability that l contracts the infection from k prior to it's recovery, i.e. the *transmissibility* is given by $\int_{t_k}^{\infty} f_{k\to l}(t,t_k)\mathrm{d}t$ and unlike the static case in [18] it is edge- and time-dependent. Similar to the considerations in [253] and Sec. B.1, one could extend the model further to allow for heterogeneous hazard rates across the population. However, we will focus on the homogeneous case here for simplicity.

Using the new definition Eq. (C.45a) of the transmission probability for general SIR-type epidemics, we expand the dynamic Eq. (C.44) similar to the set of

Eq. (C.2), (C.5) and (C.7) for the Markovian case in Sec. C.1:

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta_{k\to l}(t) = \int_0^t \mathrm{d}t_k \beta(t-t_k) a_{k\to l}(t) \mathcal{I}_{k\to l}(t,t_k) \tag{C.46a}$$

$$\mathcal{I}_{k\to l}(t,t_k) = (1-z)\Lambda_{k\to l}(t,0)\delta(t_k) + \tag{C.46b}$$
$$+ \Lambda_{k\to l}(t,t_k)\left[-\frac{\mathrm{d}}{\mathrm{d}t_k} S_{k\to l}(t_k)\right]$$

$$\frac{\mathrm{d}}{\mathrm{d}t} S_{k\to l}(t) = S_{k\to l}(t) \sum_{j\in\mathcal{N}_k\setminus l} \frac{\dot{\theta}_{k\to l}(t)}{\theta_{k\to l}(t)} \tag{C.46c}$$

Here, we introduced the probability density $\mathcal{I}_{k\to l}(t,t_k)$ that a node k, infected at time t_k is still infectious at the observation time t and has not yet passed the disease to it's neighbor, i.e. the cavity node l. Note that integrating Eq. (C.46b) would lead to Eq. (C.39).

Finally, we return to node-based observables and we start with the observation that node l is susceptible at time t if it has not contracted the disease from any of it's neighbors. Similar to the CB model (Eq. (5.7a)), the corresponding probability is therefore given by

$$P_l^S(t) = z_l \prod_{k\in\mathcal{N}_l} \theta_{k\to l}(t) \tag{C.47}$$

The probability to find node l in the infected state evolves according to three contributions: (1) As node l leaves the susceptible state it enters the infected compartment and thus the (non-negative) flux $-\dot{S}_l(t)$ contributes directly to $\dot{I}_l(t)$. (2) Given that node l was initially infected with probability $1-z_l$, it may recover within the interval $[t, t+\mathrm{d}t]$ with probability $r(t) = \mu(t)\Lambda_r(t)$. Here, $\Lambda_r(\tau) = \int_0^\tau \mu(\tau')\mathrm{d}\tau'$ is the survival probability with respect to the recovery process, i.e. the probability that a node has not yet recovered after being infected for a time period τ. (3) Node l may contract the disease at absolute time t_l, stay infected for a period $\tau = t - t_l$ and recover at last. The corresponding out-flux is given by $\int_0^t [-\dot{S}_l(t_l)] r(t-t_l)\mathrm{d}t_l$. Putting these contributions together we find

$$\frac{\mathrm{d}}{\mathrm{d}t} P_l^I(t) = -\frac{\mathrm{d}}{\mathrm{d}t} P_l^S(t) - (1-z_l) r(t) + \int_0^t \frac{\mathrm{d}}{\mathrm{d}t_l} P_l^S(t_l) r(t-t_l)\mathrm{d}t_l \tag{C.48}$$

Together with the conservation condition $P_l^I(t) = 1 - P_l^S(t) - P_l^I(t)$ and consistent initial conditions as in the main article, we have thus presented a non-Markovian extention to the widely used SIR model on temporal networks.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. habil. Philipp Hövel for his endless support and constant dedication to our projects. Over a period of 4 years, Philipp has always managed to maintain a motivating and friendly atmosphere in our group, which turned the PhD period into a wonderful part of my life. During that time, I have learned a great deal - from managing time to formulating a publication - and I will be grateful to him for this in many years to come. Moreover, Philipp facilitated constantly scientific exchanges and collaborations with guest researchers, without which this work would not have been possible and finally, I am particularly thankful for endless hours of scientific and general discussions with him that shaped this thesis and gave me constant suppport.

This thesis is a result of multiple collaborations and therefore, I am deeply grateful to my coauthors. In partucular, I would like to thank Prof. Sokolov, who taught me theoretical physics, supervised my final thesis and contrinuted fundamentally to my first two publications during the PhD. I hope that also in the future there will be again opportunities for inspiring conversations in the haze of his pipe. My final project, i.e., Chap. 5 of this thesis, would have not been possible without Prof. Gleeson's support. His hospitality during a 6 month visit to the University of Limerick and our regular discussions turned this scientific exchange into my most productive period. Finally, I am also thankful to my coauthors Hartmut Lentz and Flavio Iannelli.

My friends who read the thesis, corrected an uncountable number of spelling errors and gave their valuable, though sometimes very unadorned critique. To this end, I would like to mention Teresa, Gert, Jason, Robert, Fabi, and Philipp (Ostendorf). A huge "thank you" goes also Teresa, who reminded me regularly to enjoy the PhD instead of focusing only on the next deadline. This PhD period would not have been such an amazing time without my friends and office mates Jason and Gert as well as my colleagues Fabi, Robert, and Stefan. I also acknowledge inspiring discussion with our French collaborateurs from the group of Vittoria Colizza and in particular, Eugenio Valdano.

Finally, I am full of gratitude to my parents, who were always proud of the work especially at times when I myself was not convinced of the results.

References

1. Schlipköter, U. & Flahault, A. Communicable Diseases: Achievements and Challenges for Public Health. *Publ. Health Rev.* **32,** 90–119 (June 2010).
2. Organization, W. H. *The World Health Report* 2004.
3. Hufnagel, L., Brockmann, D. & Geisel, T. Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci.* **101,** 15124 (2004).
4. Colizza, V., Pastor-Satorras, R. & Vespignani, A. Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.* **3,** 276 (2007).
5. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342,** 1337–1342 (2013).
6. Hamer, W. H. Epidemic disease in England. *Lancet* **1,** 733–739 (1906).
7. Ross, R. *The prevention of malaria* (New York, E.P. Dutton, 1910).
8. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. A* **115,** 700–721 (1927).
9. Bailey, N. T. J. *The mathematical theory of epidemics* (Griffin, 1957).
10. Anderson, R. H. & May, R. M. *Infectious diseases of humans: dynamics and control.* (Oxford Univ. Press, Oxford and New York, 1992).
11. Hethcote, H. W. The mathematics of infectious diseases. *SIAM review* **42,** 599–653 (2000).
12. Van Kampen, N. G. *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
13. Bailey, N. T. J. *The Mathematical Theory of Infectious Diseases and Its Applications* (Griffin, London, 1975).
14. Simon, P. L., Taylor, M. & Kiss, I. Z. Exact epidemic models on graphs using graph-automorphism driven lumping. *J. Math. Biol.* **62,** 479–508 (Apr. 2011).
15. Van Mieghem, P., Omic, J. & Kooij, R. Virus Spread in Networks. *IEEE/ACM Trans. Netw.* **17,** 1–14 (2009).
16. Kiss, I. Z., Röst, G. & Vizi, Z. Generalization of pairwise models to non-Markovian epidemics on networks. *Phys. Rev. Lett.* **115,** 078701 (2015).

17. Sherborne, N. *et al.* Mean-field models for non-Markovian epidemics on networks. *J. Math. Biol.* **76,** 755–778 (Feb. 2018).

18. Karrer, B. & Newman, M. E. J. Message passing approach for general epidemic models. *Phys. Rev. E* **82,** 016101 (1 July 2010).

19. Gonçalves, S., Abramson, G. & Gomes, M. F. C. Oscillations in SIRS model with distributed delays. *Eur. Phys. J. B* **81,** 363 (2011).

20. Van Mieghem, P. & van de Bovenkamp, R. Non-Markovian Infection Spread Dramatically Alters the Susceptible-Infected-Susceptible Epidemic Threshold in Networks. *Phys. Rev. Lett.* **110,** 108701 (10 Mar. 2013).

21. Keeling, M. J. & Rohani, P. *Modeling infectious diseases in humans and animals* (Princeton University Press, 2008).

22. Pastor-Satorras, R. *et al.* Epidemic processes in complex networks. *Rev. Mod. Phys.* **87,** 925–979 (3 Aug. 2015).

23. Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.* (2009).

24. Eubank, S. *et al.* Modelling disease outbreaks in realistic urban social networks. *Nature* **429,** 180–184 (2004).

25. Ferguson, N. M. *et al.* Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437,** 209–214 (2005).

26. Halloran, M. E. *et al.* Modeling targeted layered containment of an influenza pandemic in the United States. *Proc. Natl. Acad. Sci.* **105,** 4639–4644 (2008).

27. Chao, D. L. *et al.* FluTE, a Publicly Available Stochastic Influenza Epidemic Simulation Model. *PLOS Comput. Biol.* **6,** 1–8 (2010).

28. Longini, I. M. *et al.* Containing Pandemic Influenza at the Source. *Science* **309,** 1083–1087 (2005).

29. Merler, S., Ajelli, M. & Pugliese. Determinants of the Spatiotemporal Dynamics of the 2009 H1N1 Pandemic in Europe: Implications for Real-Time Modelling. *PLOS Comput. Biol.* **7,** 1–13 (2011).

30. Keeling, M. J. & Eames, K. T. D. Networks and epidemic models. *Journal of The Royal Society Interface* **2,** 295–307 (2005).

31. Bajardi, P. *et al.* Optimizing surveillance for livestock disease spreading through animal movements. *J. Royal Soc. Interface* **9,** 2814–2825 (2012).

32. Schirdewahn, F. *et al.* in *Temporal Network Epidemiology. Theoretical Biology* (eds Masuda, N. & Holme, P.) 215–240 (Springer, Singapore, 2017).

33. Valdano, E. *et al.* Analytical computation of the epidemic threshold on temporal networks. *Phys. Rev. X* **5,** 021005 (2015).

34. Butts, C. T. Revisiting the Foundations of Network Analysis. *Science* **325,** 414–416 (2009).

35. Barabási, A. .-.-L. Network science. *Philos. Trans. R. Soc. London, Ser. A* **371,** 20120375 (2013).

36. Newman, M. E. J. *Networks: an introduction* (Oxford University Press, Inc., New York, 2010).

37. Vespignani, A. Predicting the Behavior of Techno-Social Systems. *Science* **325,** 425–428 (2009).

38. Vespignani, A. Modelling dynamical processes in complex socio-technical systems. *Nat. Phys.* **8,** 32–39 (2012).

39. Erdös, P. & Rényi, A. On random graphs. *Publ. Math. Debrecen* **6,** 290–297 (1959).

40. Erdös, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **5,** 17–61 (1960).

41. Bollobás, B. *Random graphs* 2nd ed. (Cambridge University Press, 2001).

42. Milgram, S. The Small-World Problem. *Psychology Today* **106,** 61–67 (May 1967).

43. Granovetter, M. in *Social networks* 347–367 (Elsevier, 1977).

44. Zachary, W. W. An information flow model for conflict and fission in small groups. *Journal of anthropological research* **33,** 452–473 (1977).

45. Freeman, L. C. Centrality in social networks conceptual clarification. *Social networks* **1,** 215–239 (1978).

46. Wasserman, S. & Faust, K. *Social network analysis: Methods and applications* (Cambridge university press, 1994).

47. Michel, J. .-.-B. *et al.* Quantitative analysis of culture using millions of digitized books. *science* **331,** 176–182 (2011).

48. Leskovec, J. & Krevl, A. *SNAP Datasets: Stanford Large Network Dataset Collection* 2015.

49. Eksombatchai, C. *et al. Pixie: A System for Recommending 3 Billion Items to 200 Million Users in Real-Time* in *Proceedings of the 2018 World Wide Web Conference* (International World Wide Web Conferences Steering Committee, Lyon, France, 2018), 1775–1784.

50. Isella, L. *et al.* What's in a Crowd? Analysis of Face-to-Face Behavioral Networks. *J. Theor. Biol.* **271,** 166–180 (2011).

51. Lentz, H. H. K. *Paths for epidemics in static and temporal networks* PhD thesis (Humboldt-Universität zu Berlin, 2013).

52. Cattuto, C. *et al.* Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLOS ONE* **5,** 1–9 (2010).

53. Valdano, E. *Analyse quantitative de la vulnérabilité des réseaux temporels aux maladies infectieuses* PhD thesis (UPMC, Sorbonne Universités, 2015).

54. Albert, R. & Barabási, A. .-.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74,** 47–97 (Jan. 2002).

55. Baronchelli, A. *et al.* Networks in Cognitive Science. *Trends Cogn. Sci.* **17,** 348–360 (2013).

56. Boccaletti, S. *et al.* Complex networks: Structure and dynamics. *Phys. Rep.* **424,** 175–308 (2006).

57. Caldarelli, G. *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford Univ. Press, New York, 2007).

58. Cohen, R. & Havlin, S. *Complex Networks: Structure, Robustness and Function* (Cambridge Univ Pr, 2010).

59. Costa, L. d. F. *et al.* Characterization of complex networks: A survey of measurements. *Adv. Phys.* **56,** 167–242 (2007).

60. Dorogovtsev, S. N. & Mendes, J. F. F. Evolution of networks. *Adv. Phys.* **51,** 1079 (2002).

61. Dorogovtsev, S. N. & Mendes, J. F. F. *Evolution of networks: From biological nets to the Internet and WWW* (OUP Oxford, 2013).

62. Newman, M. E. J. The Structure and Function of Complex Networks. *SIAM Review* **45,** 167–256 (2003).

63. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69,** 026113 (2004).

64. Palla, G. *et al.* Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435,** 814–818 (2005).

65. Fortunato, S. Community detection in graphs. *Phys. Rep.* **486,** 75–174 (2010).

66. Peixoto, T. P. Hierarchical block structures and high-resolution model selection in large networks. *Phys. Rev. X* **4,** 011047 (1 Mar. 2014).

67. Rosvall, M. & Bergstrom, C. T. An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences* **104,** 7327–7331 (2007).

68. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393,** 440–442 (June 1998).

69. Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-Law Distributions in Empirical Data. *SIAM Rev.* **51,** 661–703 (2009).

70. Konschake, M. *et al.* On the Robustness of In- and Out-Components in a Temporal Graph. *PLOS ONE* **8,** e55223 (Feb. 2013).

71. Lentz, H. H. K., Selhorst, T. & Sokolov, I. M. Unfolding Accessibility Provides a Macroscopic Approach to Temporal Networks. *Phys. Rev. Lett.* **110,** 118701 (Mar. 2013).

72. Koher, A. *et al.* Infections on Temporal Networks - A Matrix-Based Approach. *PLOS ONE* **11,** e0151209 (Apr. 2016).

73. Grindrod, P. *et al.* Communicability across evolving networks. *Phys. Rev. E* **83,** 046120 (4 Apr. 2011).

74. Grindrod, P. & Higham, D. J. A Matrix Iteration for Dynamic Network Summaries. *SIAM Rev.* **55,** 118–128 (2013).

75. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86,** 3200 (2001).

76. Pastor-Satorras, R. & Vespignani, A. Immunization of complex networks. *Phys. Rev. E* **65,** 036104 (2002).

77. Moreno, Y., Pastor-Satorras, R. & Vespignani, A. Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B* **26,** 521–529 (2002).

78. Ferreri, L. *et al.* Interplay of network dynamics and heterogeneity of ties on spreading dynamics. *Phys. Rev. E* **90,** 012812 (1 July 2014).

79. Gross, T., D'Lima, C. J. D. & Blasius, B. Epidemic Dynamics on an Adaptive Network. *Phys. Rev. Lett.* **96,** 208701 (May 2006).

80. Vazquez, A. *et al.* Impact of Non-Poissonian Activity Patterns on Spreading Processes. *Phys. Rev. Lett.* **98,** 158702 (15 Apr. 2007).

81. Iribarren, J. L. & Moro, E. Impact of Human Activity Patterns on the Dynamics of Information Diffusion. *Phys. Rev. Lett.* **103,** 038702 (2009).

82. Karsai, M. *et al.* Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E* **83,** 025102 (2 Feb. 2011).

83. Miritello, G., Moro, E. & Lara, R. Dynamical strength of social ties in information spreading. *Phys. Rev. E* **83,** 045102 (4 Apr. 2011).

84. Rocha, L. E. C. & Blondel, V. D. Bursts of Vertex Activation and Epidemics in Evolving Networks. *PLoS Comput Biol* **9,** e1002974 (2013).

85. Kiss, I. Z. *et al.* Exact deterministic representation of Markovian SIR epidemics on networks with and without loops. *J. Math. Biol.* **70,** 437–464 (Feb. 2015).

86. Wang, Y. *et al. Epidemic spreading in real networks: an eigenvalue viewpoint* in *Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on* (Oct. 2003), 25–34.

87. Rocha, L. E. C. & Masuda, N. Individual-based approach to epidemic processes on arbitrary dynamic contact networks. *Sci. Rep.* **6,** 31456 (2016).

88. Chakrabarti, D. *et al.* Epidemic Thresholds in Real Networks. *ACM Trans. Inf. Syst. Secur.* **10,** 1:1–1:26 (Jan. 2008).

89. Ganesh, A., Massoulié, L. & Towsley, D. *The effect of network topology on the spread of epidemics* in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE* **2** (2005), 1455–1466.

90. Gómez, S. *et al.* Discrete-time Markov chain approach to contact-based disease spreading in complex networks. *Europhys. Lett.* **89,** 38009 (2010).

91. Youssef, M. & Scoglio, C. An individual-based approach to SIR epidemics in contact networks. *J. Theor. Biol.* **283,** 136–144 (2011).

92. ben-Avraham, D. & Köhler, J. Mean-field (n,m)-cluster approximation for lattice models. *Phys. Rev. A* **45,** 8358–8370 (12 June 1992).

93. Eames, K. T. D. & Keeling, M. J. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc. Natl. Acad. Sci.* **99,** 13330–13335 (2002).

94. Kiss, I. Z., Miller, J. C. & Simon, P. L. *Mathematics of Epidemics on Networks* (Springer Int. Publ., 2017).

95. Mata, A. S. & Ferreira, S. C. Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks. *Europhys. Lett.* **103,** 48003 (Aug. 2013).

96. Cator, E. & Van Mieghem, P. Second-order mean-field susceptible-infected-susceptible epidemic threshold. *Phys. Rev. E* **85,** 056111 (5 May 2012).

97. Demirel, G. *et al.* Moment-closure approximations for discrete adaptive networks. *Physica D* **267.** Evolving Dynamical Networks, 68–80 (2014).

98. Gleeson, J. P. Binary-State Dynamics on Complex Networks: Pair Approximation and Beyond. *Phys. Rev. X* **3,** 021004 (2 Apr. 2013).

99. Marceau, V. *et al.* Adaptive networks: Coevolution of disease and topology. *Phys. Rev. E* **82,** 036116 (3 Sept. 2010).

100. Lindquist, J. *et al.* Effective degree network disease models. *J. Math. Biol.* **62,** 143–164 (Feb. 2011).

101. Karnatak, R. *et al.* Nature of the phase-flip transition in the synchronized approach to amplitude death. *Phys. Rev. E* **82,** 046219 (2010).

102. Pearl, J. *Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach* in *Proceedings of the Second AAAI Conference on Artificial Intelligence* (AAAI Press, Pittsburgh, Pennsylvania, 1982), 133–136.

103. Karrer, B., Newman, M. E. J. & Zdeborová, L. Percolation on Sparse Networks. *Phys. Rev. Lett.* **113,** 208702 (20 Nov. 2014).

104. Miller, J. C., Slim, A. C. & Volz, E. M. Edge-based compartmental modelling for infectious disease spread. *J. Royal Soc. Interface* **9,** 890–906 (2012).

105. Lokhov, A. Y. *et al.* Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Phys. Rev. E* **90,** 012801 (2014).

106. Wilkinson, R. R., Ball, F. G. & Sharkey, K. J. The relationships between message passing, pairwise, Kermack–McKendrick and stochastic SIR epidemic models. *J. Math. Biol.* **75,** 1563–1590 (Dec. 2017).

107. Colizza, V. & Vespignani, A. Invasion Threshold in Heterogeneous Metapopulation Networks. *Phys. Rev. Lett.* **99,** 148701 (2007).

108. Colizza, V. & Vespignani, A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations. *J. Theor. Biol.* **251,** 450–67 (Apr. 2008).

109. Gallos, L. K. & Argyrakis, P. Absence of Kinetic Effects in Reaction-Diffusion Processes in Scale-Free Networks. *Phys. Rev. Lett.* **92,** 138301 (13 Apr. 2004).

110. Keeling, M. J. & Rohani, P. Estimating spatial coupling in epidemiological systems: a mechanistic approach. *Ecol. Lett.* **5,** 20–29 (2002).

111. Sattenspiel, L. & Dietz, K. A structured epidemic model incorporating geographic mobility among regions. *Math. Biosci.* **128,** 71–91 (1995).

112. Watts, D. J. *et al.* Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl. Acad. Sci.* **102,** 11157–11162 (2005).

113. Rvachev, L. A. & Longini, I. M. A mathematical model for the global spread of influenza. *Math. Biosci.* **75,** 3–22 (1985).

114. Gautreau, A., Barrat, A. & Barthélemy, M. Global disease spread: Statistics and estimation of arrival times. *J. Theor. Biol.* **251,** 509–522 (2008).

115. Iannelli, F. *et al.* Effective Distances for Epidemics Spreading on Complex Networks. *Phys. Rev. E* **95,** 012313 (Jan. 2017).

116. Hughes, B. D. *Random walks and random environments. Vol. 1.* (Oxford University Press, Oxford, 1995).

117. Aldous, D. & Fill, J. A. *Reversible Markov chains and random walks on graphs, unfinished monograph* 2002.

118. Feller, W. *An introduction to probability theory and its applications* (John Wiley & Sons, New York, 2008).

119. Ore, O. Pascal and the Invention of Probability Theory. *Am. Math. Mon.* **67,** 409–419 (1960).

120. Einstein, A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.* **322,** 549–560 (1905).

121. Campbell, J. Y., Lo, A. W. & MacKinlay, A. C. *The econometrics of financial markets* (Princeton University Press, Princeton, NJ, USA, 1996).

122. Mantegna, R. N. & Stanley, H. E. *Econophysics: An introduction* (Cambridge University Press, Cambridge, 1999).

123. Codling, E. A., Plank, M. J. & Benhamou, S. Random walks in biology. *J. Royal Soc. Interface* **5,** 813–34 (2008).

124. Okubo, A. & Levin, S. A. *Diffusion and Ecological Problems: Modern Perspectives* (Springer, NY, 2001).

125. Brin, S. & Page, L. *The Anatomy of a Large-Scale Hypertextual Web Search Engine* in *WWW 1998* (1998).

126. Gleich, D. F. PageRank Beyond the Web. *SIAM Rev.* **57,** 321–363 (2015).

127. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. of the Natl. Acad. Sci.* **105,** 1118–1123 (2008).

128. Djurdjevac, N. *et al.* Random Walks on Complex Modular Networks. *JNAIAM* **6,** 29–50 (2011).

129. Lorenz, P. *et al.* in *Complex Networks & Their Applications VI.* (eds Cherifi, C. *et al.*) 401–413 (Springer, Cham, 2018).

130. Lorenz-Spreen, P. *et al.* Tracking online topics over time: understanding dynamic hashtag communities. *Comput. Soc. Netw.* **5,** 9 (2018).

131. Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E* **66,** 016128 (July 2002).

132. Blythe, S. P. & Anderson, R. M. Distributed Incubation and Infectious Periods in Models of the Transmission Dynamics of the Human Immunodeficiency Virus (HIV). *Math. Med. Biol.* **5,** 1–19 (1988).

133. Keeling, M. J. The effects of local spatial structure on epidemiological invasions. *Proc. Roy. Soc. B* **266,** 859–867 (1999).

134. Belik, V., Geisel, T. & Brockmann, D. Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases. *Phys. Rev. X* **1,** 011001 (1 Aug. 2011).

135. Strogatz, S. H. *Nonlinear Dynamics and Chaos* (Westview Press, Cambridge, MA, 1994).

136. Pastor-Satorras, R. & Vespignani, A. Epidemic Spreading in Scale-Free Networks. *Phys. Rev. Lett.* **86,** 3200–3203 (2001).

137. Kiss, I. Z., Miller, J. C. & Simon, P. L. *Mathematics of epidemics on networks: From Exact to Approximate Models* (Springer, 2017).

138. Gleeson, J. P. *et al.* Competition-Induced Criticality in a Model of Meme Popularity. *Phys. Rev. Lett.* **112,** 048701 (4 Jan. 2014).

139. Gleeson, J. P. High-Accuracy Approximation of Binary-State Dynamics on Networks. *Phys. Rev. Lett.* **107,** 068701 (6 Aug. 2011).

140. Diekmann, O., Heesterbeek, J. A. P. & Roberts, M. G. The construction of next-generation matrices for compartmental epidemic models. *J. Royal Soc. Interface,* rsif200903 (2009).

141. Henkel, M., Hinrichsen, H. & LÃŒbeck, S. *Non-Equilibrium Phase Transitions: Absorbing Phase Transitions* (Springer, Netherlands, 2008).

142. Marro, J. & Dickman, R. *Nonequilibrium Phase Transitions in Lattice Models* (Cambridge University Press, 1999).

143. Nolting, W. *Grundkurs Theoretische Physik 4/2: Thermodynamik* (Springer Spektrum, 2016).

144. Grassberger, P. On the critical behavior of the general epidemic process and dynamical percolation. *Math. Biosci.* **63,** 157–172 (1983).

145. Stauffer, D. & Aharony, A. *Introduction to percolation theory* (CRC press, 1994).

146. Yeomans, J. M. *Statistical mechanics of phase transitions* (Oxford University Press, Oxford, 1992).

147. Bayerisches Staatsministerium für Ernährung, L. u. F. (*Herkunftssicherungs und Informationssystem für Tiere [website]* Available from: `www.hi-tier.de`. 2016.

148. Holland, P. W. & Leinhardt, S. Transitivity in Structural Models of Small Groups. *Comparative Group Studies* **2,** 107–124 (1971).

149. Warshall, S. A Theorem on Boolean Matrices. *J. ACM* **9,** 11–12 (Jan. 1962).

150. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math.* **1,** 269–271 (1959).

151. Rocha, L. E. C., Liljeros, F. & Holme, P. Information dynamics shape the sexual networks of Internet-mediated prostitution. *Proc. Natl. Acad. Sci.* **107,** 5706–5711 (Mar. 2010).

152. Eagle, N., Pentland, A. & Lazer, D. Inferring friendship network structure by using mobile phone data. *PNAS* **106,** 15274 (2009).

153. Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.* **519,** 97–125 (Mar. 2012).

154. Stopczynski, A. *et al.* Measuring Large-Scale Social Networks with High Resolution. *PLOS ONE* **9,** e95978 (2014).

155. Scholtes, I., Wider, N. & Garas, A. Higher-order aggregate networks in the analysis of temporal networks: path structures and centralities. *Eur. Phys. J. B* **89,** 61 (Mar. 2016).

156. Borgatti, S. P. Centrality and network flow. *Social networks* **27,** 55–71 (2005).

157. Borgatti, S. P. & Everett, M. G. A Graph-theoretic perspective on centrality. *Soc. Netw.* **28,** 466–484 (2006).

158. Fritzemeier, J. *et al.* Epidemiology of classical swine fever in Germany in the 1990s. *Vet. Microbiol.* **77,** 29–41 (2000).

159. Lentz, H. H. K. *et al.* Livestock disease spread through animal movements: a static and temporal network analysis of pig trade in Germany. *PLOS ONE* **11,** e0155196 (May 2016).

160. Dorogovtsev, S. N., Mendes, J. F. F. & Samukhin, A. N. Giant strongly connected component of directed networks. *Phys. Rev. E* **64,** 025101 (2 July 2001).

161. Newman, M. E. J. The structure and function of complex networks. *SIAM Review* **45,** 167–256 (2003).

162. Konschake, M. *et al.* On the Robustness of In- and Out-Components in a Temporal Network. *PLOS ONE* **8,** e55223 (2013).

163. Metzler, R. & Klafter, J. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **339,** 1–77 (2000).

164. Montroll, E. W. & Weiss, G. H. Random Walks on Lattices. II. *J. Math. Phys.* **6,** 167–181 (1965).

165. Sokolov, I. M. & Klafter, J. *First Steps in Random Walks* (Oxford University Press, 2011).

166. Kutner, R. & Masoliver, J. The continuous time random walk, still trendy: fifty-year history, state of art and outlook. *Eur. Phys. J. B* **90,** 50 (Mar. 2017).

167. Vazquez, A. *et al.* Modeling bursts and heavy tails in human dynamics. *Phys. Rev. E* **73,** 036127 (2006).

168. Masuda, N., Porter, M. A. & Lambiotte, R. Random walks and diffusion on networks. *Phys. Rep.* **716-717** (2017).

169. Van Mieghem, P. *Graph Spectra for Complex Networks* (Cambridge University Press, New York, 2011).

170. Lorenz-Spreen, P. *et al.* Accelerating dynamics of collective attention. *Nat. Commun.* in print (2019).

171. Gómez-Gardeñes, J. *et al.* From modular to centralized organization of synchronization in functional areas of the cat cerebral cortex. *PLOS ONE* **5,** e12313 (2010).

172. Rattana, P. *et al.* A Class of Pairwise Models for Epidemic Dynamics on Weighted Networks. *Bull. Math. Biol.* **75,** 466–490 (Mar. 2013).

173. Sharkey, K. J. Deterministic epidemiological models at the individual level. *J. Math. Biol.* **57,** 311–331 (Sept. 2008).

174. Cai, C. .-.-R. *et al.* Solving the Dynamic Correlation Problem of the Susceptible-Infected-Susceptible Model on Networks. *Phys. Rev. Lett.* **116,** 258301 (25 June 2016).

175. Valdano, E., Poletto, C. & Colizza, V. Infection propagator approach to compute epidemic thresholds on temporal networks: impact of immunity and of limited temporal resolution. *Eur. Phys. J. B* **88,** 341 (Dec. 2015).

176. Bokharaie, V. S., Mason, O. & Wirth, F. *Spread of epidemics in time-dependent network* in *Proc. 19th Int. Symp. Math. Theory Netw. Syst. (MTNS'10)* (Budapest, Hungary, July 2010), 1717–1719.

177. Prakash, B. A. *et al. Virus Propagation on Time-Varying Networks: Theory and Immunization Algorithms* in *Machine Learning and Knowledge Discovery in Databases* (eds Balcázar, J. L. *et al.*) (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010), 99–114.

178. Speidel, L. *et al.* Temporal interactions facilitate endemicity in the susceptible-infected-susceptible epidemic model. *New J. Phys.* **18,** 073013 (2016).

179. Wang, W. *et al.* Predicting the epidemic threshold of the susceptible-infected-recovered model. *Scientific Reports* **6** (2015).

180. Colizza, V. *et al.* Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study. *BMC Medicine* **5,** 34 (Nov. 2007).

181. Yang, Y. *et al.* The Transmissibility and Control of Pandemic Influenza A (H1N1) Virus. *Science* **326,** 729–733 (2009).

182. Poletto, C. *et al.* Assessment of the Middle East respiratory syndrome coronavirus (MERS-CoV) epidemic in the Middle East and risk of international spread using a novel maximum likelihood analysis approach. *Eurosurveillance* **19** (June 2014).

183. Poletto, C. *et al.* Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. *Eurosurveillance* **19** (2014).

184. Ferguson, N. M. *et al.* Countering the Zika epidemic in Latin America. *Science* **353,** 353–354 (2016).

185. Moore, M., Gould, P. & Keary, B. S. Global urbanization and impact on health. *Int. J. Hyg. Environ. Health* **206,** 269–278 (2003).

186. Cooper, B. S. *et al.* Delaying the International Spread of Pandemic Influenza. *PLOS Medicine* **3** (2006).

187. Van den Broeck, W. *et al.* The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infect. Dis.* **11,** 37 (2011).

188. Ballesteros, S., Vergu, E. & Cazelles, B. Influenza A gradual and epochal evolution: insights from simple models. *PLOS ONE* **4,** e7426 (2009).

189. Tizzoni, M. *et al.* Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Medicine* **10,** 165 (Dec. 2012).

190. Borjigin, J. *et al.* Surge of neurophysiological coherence and connectivity in the dying brain. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 14432 (Aug. 2013).

191. Colizza, V. *et al.* The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 2015–2020 (2006).

192. Woolley-Meza, O. *et al.* Complexity in human transportation networks: a comparative analysis of worldwide air transportation and global cargo-ship movements. *Eur. Phys. J. B* **84,** 589–600 (Dec. 2011).

193. Barrat, A. *et al.* The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 3747–3752 (2004).

194. Barrat, A., Barthélemy, M. & Vespignani, A. The effects of spatial constraints on the evolution of weighted complex networks. *J. Stat. Mech.* **2005,** P05003 (2005).

195. Gumbel, E. J. Les valeurs extrêmes des distributions statistiques. fr. *Ann. Inst. H. Poincare Anal.* **5,** 115–158 (1935).

196. Gumbel, E. J. The Return Period of Flood Flows. *Ann. Math. Statist.* **12,** 163–190 (1941).

197. Gilli, M. & Këllezi, E. An Application of Extreme Value Theory for Measuring Financial Risk. *Comp. Econ.* **27,** 207–228 (May 2006).

198. Burke, E. J., Perry, R. H. J. & Brown, S. J. An extreme value analysis of UK drought and projections of change in the future. *Journal of Hydrology* **388,** 131–143 (2010).

199. Cornell, C. A. Engineering seismic risk analysis. *Bull. Seismol. Soc. Am.* **58,** 1583–1606 (1968).

200. Crépey, P., Alvarez, F. P. & Barthélemy, M. Epidemic variability in complex networks. *Phys. Rev. E* **73,** 046131 (4 Apr. 2006).

201. Sedgewick, R. *Algorithms in C, Part 5: Graph Algorithms, Third Edition* Third (Addison-Wesley Professional, 2001).

202. Yen, J. Y. Finding the K Shortest Loopless Paths in a Network. *Manag. Sci.* **17,** 712–716 (1971).

203. Redner, S. *A Guide to First-Passage Processes* (Cambridge University Press, 2001).

204. Iyer-Biswas, S. & Zilman, A. in, 261–306 (John Wiley & Sons, Ltd, 2016).

205. Metzler, R., Oshanin, G. & Redner, S. *First-passage phenomena and their applications* (World Scientific, Singapore, 2014).

206. Meyer, C. D. *Matrix Analysis and Applied Linear Algebra* (SIAM, 2000).

207. Barabási, A. .-.-L. & Albert, R. Emergence of Scaling in Random Networks. *Science* **286,** 509 (Oct. 1999).

208. Delvenne, J. .-.-C., Lambiotte, R. & Rocha, L. E. C. Diffusion on networked systems is a question of time or structure. *Nat. Commun.* **6,** 7366 (2015).

209. Darbon, A. *et al.* Network-based assessment of the vulnerability of Italian regions to bovine brucellosis. *Prev. Vet. Med.* **158,** 25–34 (2018).

210. Sekara, V., Stopczynski, A. & Lehmann, S. Fundamental structures of dynamic social networks. *Proc. Natl. Acad. Sci. USA* **113,** 9977–9982 (2016).

211. Casteigts, A. *et al.* Time-varying graphs and dynamic networks. *Int. J. Parallel Emergent Distributed Syst.* **27,** 387–408 (2012).

212. Van Mieghem, P., Sahneh, F. D. & Scoglio, C. *Exact Markovian SIR and SIS epidemics on networks and an upper bound for the epidemic threshold* (IEEE, 2014).

213. Danon, L. *et al.* Networks and the epidemiology of infectious disease. *Interdiscip. Perspect. Infect. Dis.* **2011,** 1 (2011).

214. Stehlé, J. *et al.* Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC medicine* **9,** 87 (2011).

215. Kiss, I. Z., Green, D. M. & Kao, R. R. The network of sheep movements within Great Britain: network properties and their implications for infectious disease spread. *J. R. Soc. Interface* **3,** 669–677 (2006).

216. Luce, R. D. A Note on Boolean Matrix Theory. *Proceedings of the American Mathematical Society* **3,** 382–388 (1952).

217. Katz, L. A new status index derived from sociometric analysis. English. *Psychometrika* **18,** 39–43 (1953).

218. Rocha, L. E. C., Liljeros, F. & Holme, P. Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts. *PLoS Comput Biol* **7,** e1001109 (Mar. 2011).

219. Hollingsworth, T. D., Anderson, R. M. & Fraser, C. HIV-1 Transmission, by Stage of Infection. *The Journal of Infectious Diseases* **198,** 687–693 (2008).

220. Sander, L. M. *et al.* Percolation on heterogeneous networks as a model for epidemics. *Math. Biosci.* **180,** 293–305 (2002).

221. Sander, L. M., Warren, C. P. & Sokolov, I. M. Epidemics, disorder, and percolation. *Physica A* **325.** Stochastic Systems: From Randomness to Complexity, 1–8 (2003).

222. Kirkpatrick, S. Percolation and Conduction. *Rev. Mod. Phys.* **45,** 574–588 (4 Oct. 1973).

223. Gleeson, J. P. *et al.* Accuracy of mean-field theory for dynamics on real-world networks. *Phys. Rev. E* **85,** 026106 (2012).

224. May, R. M. & Lloyd, A. Infection dynamics on scale-free networks. *Phys. Rev. E* **64,** 066112 (2001).

225. Newman, M. E. J., Barabási, A. .-.-L. & Watts, D. J. *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, USA, 2006).

226. Porter, M. A. & Gleeson, J. P. *Dynamical Systems on Networks* (Springer Int. Publ., 2016).

227. Altarelli, F. *et al.* Bayesian Inference of Epidemics on Networks via Belief Propagation. *Phys. Rev. Lett.* **112,** 118701 (11 Mar. 2014).

228. Valdano, E. *et al.* GRAZE: A generic and standardized Python package for the analysis of livestock trade data. *in preparation* (2018).

229. Koher, A. *Source code of 'Contact-based model for epidemic spreading on temporal networks'* `github.com/andreaskoher/Contact_Based_Epidemiology.git`. 2018.

230. Koher, A. *et al. Contact-based model for epidemic spreading on temporal networks* 2018.

231. Shrestha, M., Scarpino, S. V. & Moore, C. Message-passing approach for recurrent-state epidemic models on networks. *Phys. Rev. E* **92,** 022821 (2 2015).

232. Lokhov, A. Y. *Dynamic cavity method and problems on graphs* Theses (Université Paris Sud - Paris XI, Nov. 2014).

233. Melnik, S. *et al.* The unreasonable effectiveness of tree-based theory for networks with clustering. *Phys. Rev. E* **83,** 036112 (3 Mar. 2011).

234. Miller, J. C. Epidemics on Networks with Large Initial Conditions or Changing Structure. *PLOS ONE* **9,** 1–9 (2014).

235. Rogers, T. Assessing node risk and vulnerability in epidemics on networks. *Europhys. Lett.* **109,** 28005 (2015).

236. Valdano, E. *et al.* Predicting Epidemic Risk from Past Temporal Contact Data. *PLOS Comput. Biol.* **11,** e1004152 (2015).

237. Valdez, L. D., Macri, P. A. & Braunstein, L. A. Temporal Percolation of the Susceptible Network in an Epidemic Spreading. *PLOS ONE* **7,** 1–5 (2012).

238. Radicchi, F. & Castellano, C. Beyond the locally treelike approximation for percolation on real networks. *Phys. Rev. E* **93,** 030302 (3 Mar. 2016).

239. Scholtes, I. *et al.* Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nat. Commun.* **5,** 5024 (2014).

240. Lambiotte, R., Rosvall, M. & Scholtes, I. Understanding Complex Systems: From Networks to Optimal Higher-Order Models. *arXive* **abs/1806.05977** (2018).

241. Yen, L. *et al. A Family of Dissimilarity Measures Between Nodes Generalizing Both the Shortest-path and the Commute-time Distances* in *Proc. 14th ACM SIGKDD* (ACM, Las Vegas, Nevada, USA, 2008), 785–793.

242. Kivimäki, I., Shimbo, M. & Saerens, M. Developments in the theory of randomized shortest paths with a comparison of graph node distances. *Physica A* **393,** 600–616 (2014).

243. Françoisse, K. *et al.* A bag-of-paths framework for network data analysis. *Neural Netw.* **90,** 90–111 (2017).

244. Jaynes, E. T. Information Theory and Statistical Mechanics. *Phys. Rev.* **106,** 620–630 (4 May 1957).

245. Fouss, F. *et al.* An experimental investigation of kernels on graphs for collaborative recommendation and semisupervised classification. *Neural Networks* **31,** 53–72 (2012).

246. Darbon, A. *et al. Disease persistence on temporal contact networks accounting for heterogeneous infectious periods* preprint. 2018.

247. Valdano, E. *et al.* Epidemic Threshold in Continuous-Time Evolving Networks. *Phys. Rev. Lett.* **120,** 068302 (Nov. 2018).

248. Molloy, M. & Reed, B. A critical point for random graphs with a given degree sequence. *Random Structures Algorithms* **6,** 161–180 (1995).

249. Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64,** 026118 (July 2001).

250. Miller, J. C. & Volz, E. M. Model hierarchies in edge-based compartmental modeling for infectious disease spread. *J. Math. Biol.* **67,** 869–899 (Oct. 2013).

251. Miller, J. C. & Kiss, I. Z. Epidemic Spread in Networks: Existing Methods and Current Challenges. *Math. Model. Nat. Phenom.* **9,** 4–42 (2014).

252. Miller, J. C. Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. *Phys. Rev. E* **76,** 010101 (1 July 2007).

253. Wilkinson, R. R. & Sharkey, K. J. Message passing and moment closure for susceptible-infected-recovered epidemics on finite networks. *Phys. Rev. E* **89,** 022808 (2 Feb. 2014).

List of Figures

1.1 Transitivity is not assured in temporal networks. 3

2.1 Flow chart of the SIR model. 11
2.2 Time evolution of an SIR epidemic. 11
2.3 Final outbreak size R_∞. 13
2.4 A temporal network and the time-aggregated graph. 19
2.5 Comparison between time-respecting and static accessibility graph. 20
2.6 Schematic component structure of directed networks. 21
2.7 Component structure of the pig-trade network. 22
2.8 Range for every node in the pig-trade network. 23
2.9 Degree distribution of the German pig-trade network. 23
2.10 Strength distribution of the weighted German pig-trade network. . 24
2.11 Distribution of waiting times. 25
2.12 The individual range for every node in the temporal network. . . . 26

3.1 Geographically embedded global air-traffic network. 42
3.2 Properties of the symmetrized global air-traffic network. 45
3.3 Outbreak simulation based on the global air-traffic network. 48
3.4 Correlation between simulated infection arrival times and geographic location. 48
3.5 Meta-population model with two nodes. 51
3.6 Infection arrival times for a chain topology. 55
3.7 Infection arrival times on a tree graph. 56
3.8 Parallel infection transmission routes. 58
3.9 Impact of multiple competing transmission paths. 61
3.10 Example of an ego-centric visualization. 62
3.11 Applications of the DPED. 64
3.12 Impact of competing transmission paths including the RWED. . . 70
3.13 Correlation between simulated infection arrival times and the RWED. 71
3.14 Correlation analysis for different parameters. 72
3.15 Correlation analysis for different topologies. 73

4.1 Prevalence, incidence and cumulative incidence for the social contacts network. 84

4.2 Prevalence, incidence and cumulative incidence for the sexual contacts network. 86
4.3 Prevalence, incidence and cumulative incidence for the livestock-trade network. 87
4.4 Estimating the critical infectious period for the social (A), the sexual (B) and the livestock-trade network (C). 88

5.1 Illustrative examples of a simulated epidemic outbreak from a single initially infected node. 94
5.2 Epidemic trajectories for four exemplary individual nodes. 101
5.3 Comparison between simulated and estimated mean infection arrival times. 102
5.4 Comparison between simulated and estimated vulnerability. 103
5.5 Time-resolved and normalized outbreak size distributions for three different initially infected nodes. 104
5.6 Cumulative infection probability with a large fraction of initially infected nodes. 105
5.7 Vulnerability as a function of the infection probability β estimated from the CB model. 106
5.8 Distribution of final outbreak sizes as a function of the infection probability. 107
5.9 Distribution of final outbreak sizes for $\beta = 0.003$, 0.01, 0.02, and 0.1, respectively. 107
5.10 Estimation of the critical infection probability β_{crit} 109
5.11 German cattle trade network. 109
5.12 Detailed threshold analysis for six selected federal states with $\mu = 1/28$. 110
5.13 Spatial variation of epidemic risk due to cattle trade in Germany for $\mu = 1/28$. 111

B.1 Degree and weight analysis of cattle trade in the year 2010. 124
B.2 Node strength and flux analysis of cattle trade in the year 2010. . 125
B.3 Temporal analysis cattle trade in the year 2010. 126
B.4 Error estimation after restricting trade to the federal states. 127
B.5 Comparison between MC simulations and the mean-field models for different values of μ. 128
B.6 Detailed analysis of the epidemic threshold for the cattle trade network with $\mu = 1/28$. 130
B.7 Detailed analysis of the epidemic threshold for the cattle trade network with $\mu = 1/120$. 131
B.8 Spatial variation of epidemic risk for $\mu = 1/120$. 132

List of Tables

1.1 Notation of frequently used variables 7

2.1 Standard network properties of the static German pig-trade network. 20

3.1 Statistical properties of the global air-traffic network. 45

4.1 Comparison between the temporal networks and the time aggregated networks. 84

B.1 Federal states of Germany and the corresponding abbreviation together with basic network statistics. 126

Abbreviations

CC Connected component. 12

CTRW Continuous-time random-walk. 17

DTRW Discrete-time random-walk. 17

EXT External nodes. 13

GIC Giant in-component. 12

GOC Giant out-component. 12

GSCC Giant strongly connected component. 12

GWCC Giant weakly connected component. 12

MA Mass-action. 3

ODE Ordinary differential equation. 3

SIR Susceptible-infected-recovered model. 1

TEN Tendrils. 13