Spreading processes in complex networks and systems

Ma, L.

DOI
10.4233/uuid:0a27d29e-604e-4122-b646-0d79b7e9fbc7

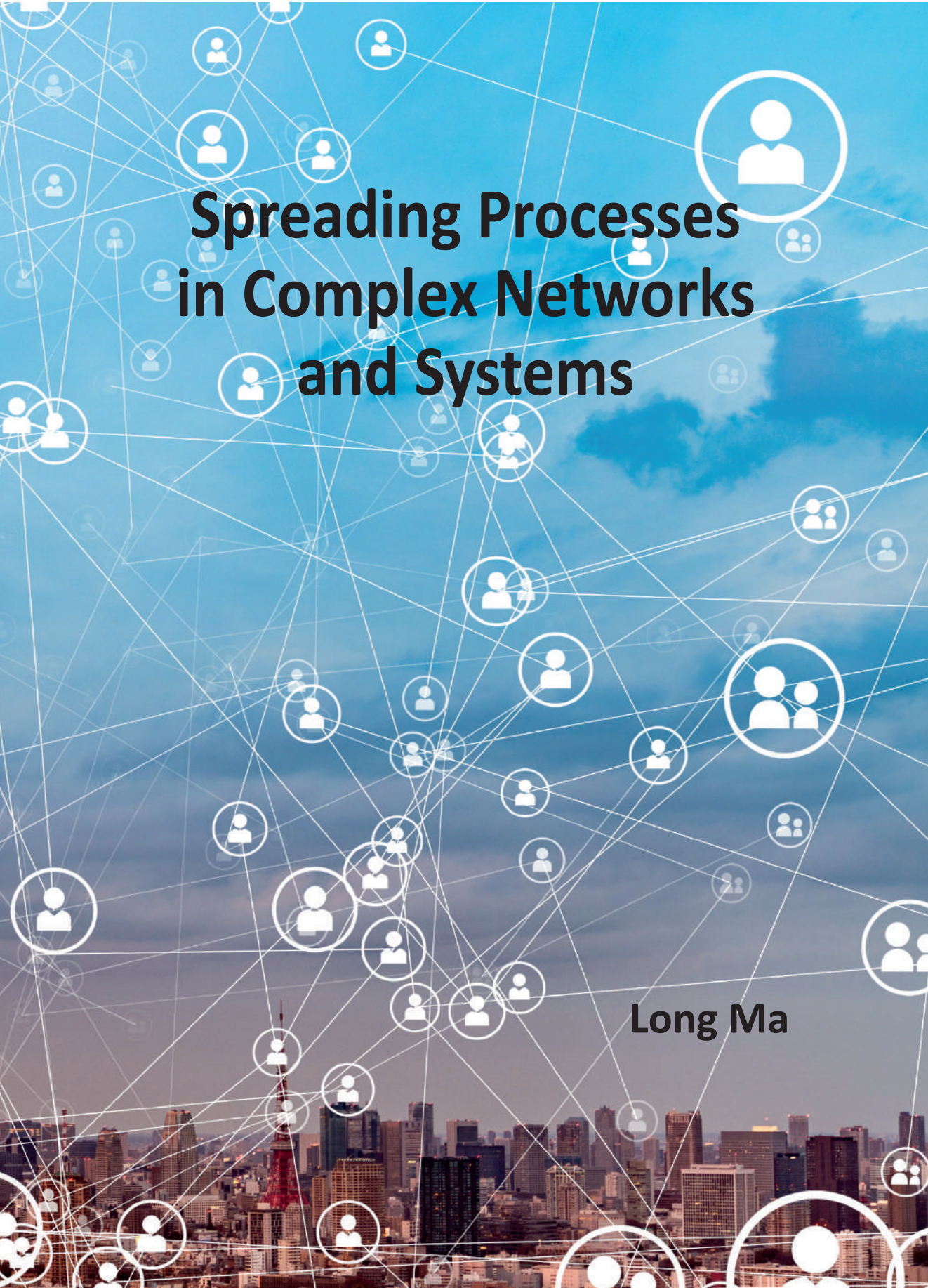
Publication date
2022

Document Version
Final published version

Citation (APA)
Ma, L. (2022). *Spreading processes in complex networks and systems*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:0a27d29e-604e-4122-b646-0d79b7e9fbc7

Spreading Processes in Complex Networks and Systems

Long Ma



SPREADING PROCESSES IN COMPLEX NETWORKS AND SYSTEMS

SPREADING PROCESSES IN COMPLEX NETWORKS AND SYSTEMS

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
Friday 21 January 2022 at 10:00 o'clock

by

Long MA

Master of Science in System Theory
Beijing Normal University, China
born in Shandong, China.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof.dr.ir. P.F.A. Van Mieghem,	Delft University of Technology, promotor
Dr. M.A. Kitsak,	Delft University of Technology, copromotor
Independent members:	
Prof.dr. A. Hanjalic	Delft University of Technology
Prof.dr.ir. R.E. Kooij	Delft University of Technology
Prof.dr. A. Pugliese	University of Trento, Italy
Dr. J.L.A. Dubbeldam	Delft University of Technology
Prof.dr.ir. G.J.T. Leus	Delft University of Technology, reserve member
Other member:	
Dr. J. Sanders	Eindhoven University of Technology

Keywords: Complex Networks, Epidemics, COVID-19 Pandemic, Inferring Network Properties, Reporting Delays, Forecast, Spectral Clustering, Error Accumulation, Quantum Circuits, Markov Chains

Printed by: Ipskamp Printing

Cover by: Long Ma

Email: l.ma-2@tudelft.nl

ISBN 978-94-6421-621-9

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

To my family

CONTENTS

SUMMARY

Complex systems are made up of many interconnected components. The interactions between components further produce emerging complex collective behaviors. In most cases, a complex system can be represented as a network in which the nodes represent the elements and the connected links illustrate the interactions. The spreading process is one of the most important dynamics in complex networks and systems. Since 2000, scientific findings related to spreading models and complex networks are experiencing a boom. Quantum computers are also complex systems. The development of quantum computing is to bring qualitative change to many fields by solving some problems faster than classical computers. However, during quantum computing, errors are spreading in the quantum circuits and accumulated with time.

The first part of this dissertation focuses on the epidemic spreading in complex networks and systems. Dynamical processes running on different networks behave differently, making reconstructing the underlying network from dynamical observations possible. In Chapter 2, we focus on inferring network properties from the dynamics of a susceptible-infected-susceptible (SIS) epidemic. Different from the previous works that based on the complete infection data of each node, we investigate what network properties can be inferred only based on the epidemic prevalence, i.e., the average fraction of infected nodes.

The recent Coronavirus disease 2019 (COVID-19) outbreak, which is caused by Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2), asks for a precise understanding of the transmission mechanisms of this virus among people. Different compartmental epidemic models are applied or proposed to fit and forecast the pandemic. In Chapter 3, we discover significant reporting delays in COVID-19 data, especially the daily recovered data. It is essential to consider the reporting delays in forecasting the COVID-19 pandemic. Another main topic about the COVID-19 pandemic is how to efficiently control and suppress the pandemic. Although many strategies have been applied to flatten the epidemic curves, it is still challenging to efficiently reduce mortality. Chapter 4 uses a two-population Susceptible-Infected-Removed (SIR) model to investigate the COVID-19 spreading when the connection between elderly and non-elderly individuals is diminished due to the high mortality risk of older people. The efficient strategies to reduce mortality are further discussed in this chapter.

The Markovian SIS epidemic on complete graphs can be exactly reduced to the birth-and-death process. However, the birth-and-death approximation for random graphs, e.g., Erdős–Rényi (ER) random networks, could lead to relatively large errors. In Chapter 5, to lessen the approximation error, we propose a spectral clustering SIS approximation (SCSA), which combines the spectral clustering of the Markov graph and the birth-and-death approximation.

The second part of this dissertation, as shown in Chapter 6, presents an exact study of an error accumulation model for multi-qubit quantum circuits using techniques from the fields of probability theory and operations research. By modeling the error accumulation process in quantum circuits using two coupled Markov chains, we can capture a weak form of time-dependency between errors in the past and future analytically. By relating the error probabilities to so-called hitting times, the presented approach can calculate probability distributions for a wide class of error measures precisely, deal with multi-qubit scenarios, and allow for fairly generic error distributions: for example, it can handle errors that are gate- as well as time-dependent. The availability of an analytical expression for the accumulation of errors also allows us to proceed with second-tier optimization methods: we illustrate how simulated annealing can be combined with our formulae to tailor-make circuits with lower error accumulation rates based on the error distributions of one's experiment.

SAMENVATTING

Complexe systemen zijn opgebouwd uit vele onderling verbonden componenten. De interacties tussen de componenten leiden tot complexe collectieve gedragingen. In de meeste gevallen kan een complex systeem worden voorgesteld als een netwerk waarin de knopen de elementen voorstellen en de verbonden links de interacties illustreren. Het spreidingsproces is een van de belangrijkste dynamieken in complexe systemen. Sinds 2000 beleven wetenschappelijke bevindingen in verband met verspreidingsmodellen en complexe netwerken een hausse. Kwantumcomputers zijn ook complexe systemen. De ontwikkeling van quantumcomputing moet op vele gebieden een kwalitatieve verandering teweegbrengen door sommige problemen sneller op te lossen dan klassieke computers. Echter, tijdens kwantumcomputing verspreiden fouten zich in de kwantumcircuits en stapelen zich op met de tijd.

Het eerste deel van dit proefschrift richt zich op de epidemieverspreiding in complexe netwerken en systemen. Dynamische processen die op verschillende netwerken draaien gedragen zich verschillend, waardoor reconstructie van het onderliggende netwerk uit dynamische waarnemingen mogelijk is. In hoofdstuk 2 richten we ons op het afleiden van netwerkeigenschappen uit de dynamica van een susceptible-infected-susceptible (SIS) epidemie. We onderzoeken welke netwerkeigenschappen kunnen worden afgeleid op basis van de epidemieprevalentie, d.w.z. de gemiddelde fractie van geïnfecteerde knooppunten.

De recente uitbraak van Coronavirusziekte 2019 (COVID-19), die veroorzaakt wordt door Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2), vraagt om een nauwkeurig begrip van de transmissiemechanismen van dit virus onder mensen. Verschillende compartimentele epidemische modellen worden toegepast of voorgesteld om de pandemie in te passen en te voorspellen. In hoofdstuk 3 ontdekken we aanzienlijke vertragingen in de rapportage van COVID-19 gegevens, met name de dagelijks herstelde gegevens. Het is essentieel om de vertragingen in de rapportage in overweging te nemen bij het voorspellen van de COVID-19 pandemie. Een ander belangrijk onderwerp in verband met de COVID-19 pandemie is hoe de pandemie efficiënt te controleren en te onderdrukken. Hoewel veel strategieën zijn toegepast om de epidemiecurven af te vlakken, is het nog steeds een uitdaging om de mortaliteit terug te dringen. Hoofdstuk 4 gebruikt een twee-populatie Susceptible-Infected-Removed (SIR) model om de verspreiding van COVID-19 te onderzoeken wanneer de verbinding tussen oudere en niet-oudere indivi-

duen verminderd is door het hoge sterfterisico van ouderen. De efficiënte strategieën om de mortaliteit te verminderen worden verder besproken in dit hoofdstuk.

De Markoviaanse SIS-epidemie op volledige grafiek kan exact worden herleid tot het geboorte-en-doodsproces. De geboorte-en-dood benadering voor willekeurige grafiek, b.v. Erdős-Rényi (ER) willekeurige netwerken, kan echter tot relatief grote fouten leiden. In Hoofdstuk 5, om de benaderingsfout te verminderen, stellen we een spectrale clustering SIS benadering (SCSA) voor, die de spectrale clustering van de Markov-grafiek en de geboorte-en-dood benadering combineert.

Het tweede deel van dit proefschrift, zoals weergegeven in Hoofdstuk 6, presenteert een exacte studie van een foutenaccumulatiemodel voor multi-qubit kwantumcircuits met behulp van technieken uit de domeinen van de waarschijnlijkheidsrekening en besliskunde. Door het foutaccumulatieproces in kwantumschakelingen te modelleren met behulp van twee gekoppelde Markovketens, kunnen we een zwakke vorm van tijdsafhankelijkheid tussen fouten in het verleden en in de toekomst analytisch vastleggen. Door de foutkansen te relateren aan de zogenaamde inslagtijden, kan de voorgestelde benadering de kansverdelingen voor een brede klasse van foutmaten nauwkeurig berekenen, multi-qubit scenario's behandelen, en vrij generieke foutverdelingen toestaan: zo kan zij bijvoorbeeld fouten behandelen die zowel gate- als tijdsafhankelijk zijn. De beschikbaarheid van een analytische uitdrukking voor de accumulatie van fouten stelt ons ook in staat om verder te gaan met tweederangs optimalisatiemethoden: we illustreren hoe gesimuleerde annealing gecombineerd kan worden met onze formules om circuits op maat te maken met lagere accumulatie van fouten op basis van de foutverdelingen van iemands experiment.

1

INTRODUCTION

Spreading processes are ubiquitous in the real world, including the spread of disease among people or animals, the spread of computer viruses in Email networks, the spread of news and rumors in social media, the cascading failure of power grids, etc. The most basic way to model the spreading processes in complex systems are compartment models [1]. Many complex systems are represented as networks, where nodes represent individuals and links indicate the interactions between nodes [2]. Quantum computers are also complex systems. During quantum computing, errors spread in the quantum circuits and are accumulated with time [3].

This dissertation focuses on the Markov processes in which the future state is solely determined by the current state. The Markovian SIS epidemic in complex networks and the Markovian error accumulation process in quantum circuits are all continuous-time or discrete-time Markov chains [4]. The Markov theory can further be applied to study these Markovian spreading processes in complex networks and systems.

1.1. COMPARTMENTAL EPIDEMIC MODELS

The Susceptible-Infectious-Recovered-Dead (SIRD) model [5, 2, 6, 7, 8] and their extensions are widely applied to describe the dynamics of the COVID-19 pandemic. The SIRD model consists of four compartments [9, 10, 11]: the fraction of susceptible individuals (S), the fraction of actively infected individuals (I), the fraction of recovered individuals (R) and the fraction of deceased individuals (D). There are transition rates between compartments. Specifically, the transition rate from S to I is βSI, where β denotes the infection

rate; the transition rates from I to R and D are respectively $\gamma_r I$ and $\gamma_d I$, where γ_r denotes the recovery rate and γ_d denotes the deceased rate. For the COVID-19 pandemic, more realistic compartments, e.g., the quarantine state [12] and the asymptomatic infected state [13], are further considered.

The reporting delay of cases should be eliminated before fitting the real daily infection, recovery and death data by epidemic models. The reporting delays for real epidemic data occur due to many reasons, e.g., the patient delay, the diagnosis delay and the public health authorities (PHA) delay [14]. The reporting delays of infections are found to be significant for many infectious diseases [14, 15]. Specifically, the reporting delays of infections for some diseases, e. g., Hepatitis B, Shigellosis and Salmonella, can take on average 2-3 weeks [14, 16]. The cases for some diseases, e. g., Hepatitis A, Measles and Mumps, are usually reported after eight days [14]. The median diagnosis delay for Malaria is around four days [17]. Research on the Middle East Respiratory Syndrome CoronaVirus (MERS-CoV), a kind of virus similar to SARS-CoV-2, found a time difference between the symptom onset and confirmation around four days [18]. Recent research on COVID-19 reveals significant reporting delays of infections for China [19, 20, 21, 22, 23, 24], Italy [25], Germany [26, 27], Singapore [28], the USA [29] and the UK [30].

1.2. SIS EPIDEMICS IN COMPLEX NETWORKS

In Markovian Susceptible-Infected-Susceptible (SIS) epidemic processes on networks [31, 32, 33, 34], both the infection process per link and the curing process are Poisson processes with rates β and δ. We can apply the continuous-time Markov chain to exactly describe the SIS epidemic processes on networks. The *network infection state* at any time t in the Markov chain can be represented by a $N \times 1$ vector with elements $x_k \in \{0, 1\}$ for $k \in \{1, 2, \cdots, N\}$, where k is a label of a node and x_k is the state of node k. Furthermore, $x_k = 0$ if node k is susceptible to be infected and $x_k = 1$ if node k is infected. Since each node has two possible states, there are 2^N possible network infection states in this Markov chain. It is infeasible to calculate the exact SIS prevalence due to the huge state space for large networks. Hence approximation methods with much lower complexity are required.

Network reconstruction is to infer the underlying networks based on the observations of an epidemic outbreak [35]. Nowadays, most network-reconstruction related papers focus on reconstructing the underlying graphs by measuring the time-dependent dynamical state of each node [36, 37, 38, 39, 40, 41, 42, 43, 44]. With the complete dynamics of each node, the network may be approximately reconstructed by different heuristic algorithms, e.g., the Bayesian methods [45, 46], the conflict-based method [34], statistical inference based method [47] and the compressed sensing or lasso methods [35]. Apart from reconstructing simple networks, there are many works on the reconstruction of the

stochastic temporal networks [48], multilayer networks [49], weighted networks [50] and directed networks [51]. All of the above methods are based on the data from all or at least most nodes, but in real scenarios, individual-level observations of spreading are hard to obtain while most of the epidemic data are population-level [52, 53].

1.3. ERROR ACCUMULATION IN QUANTUM CIRCUITS

The development of a quantum computer is expected to revolutionize computing by being able to solve hard computational problems faster than any classical computer [3, 54, 55]. However, present-day state-of-the-art quantum computers are prone to errors in their calculations due to physical effects such as unwanted qubit–qubit interactions, qubit crosstalk, and state leakage [56]. Minor errors can be corrected, but error correction methods will still be overwhelmed once too many errors occur [57, 58, 59]. Quantum circuits with different numbers of qubits and circuit depths have been designed to implement algorithms more reliably [60] and the susceptibility of a circuit to the accumulation of errors remains an important evaluation criterion.

1.4. RESEARCH QUESTIONS

This dissertation aims to study the application of techniques from epidemics, complex networks and probability theory to resolve problems about the actual spreading processes in complex networks and systems. The main questions considered in this dissertation are:

Chapter 2: What network properties can be inferred only based on the epidemic prevalence data?

Chapter 3: Are there significant reporting delays in COVID-19 data? If so, to what extent can the reporting delays affect the forecast accuracy?

Chapter 4: What's the effect of the reduction of connections between young and old people due to the high mortality rate for the elderly on the epidemic spreading? How can we efficiently reduce mortality in the COVID-19 pandemic?

Chapter 5: The birth-and-death approximation is a method to approximate the Markovian SIS epidemics on networks. How can we further improve the approximation accuracy? What's the relationship between the size of reduced state space and the approximation accuracy?

Chapter 6: How to exactly model the accumulation of Markovian errors in multi-qubit quantum computations? How to design new quantum circuits to lower the rate of error accumulation?

1

1.5. CHAPTER OVERVIEW

There are seven chapters in this thesis. The key chapters can be split into two parts. Part I (Chapters 2-5) focuses on the epidemic spreading in complex networks and systems. Part II (Chapters 6) studies the error accumulation in quantum circuits.

In Chapter 2, we focus on the problem of inferring network properties from the dynamics of an SIS epidemic and we assume that only the epidemic prevalence curve, i.e., the average fraction of infected nodes, is given. This chapter first studies what network properties can be inferred if we know the network type. Next, a simulated annealing link-rewiring algorithm, called SARA, is proposed to obtain an optimized network whose prevalence is close to the benchmark. The output of the algorithm is applied to classify the network types.

In Chapter 3, by analyzing the epidemic models and the COVID-19-related infection, recovery and death data, we discover significant reporting delays in reported data, especially the recovery data. The reporting delays are found to be different for different countries. We further demonstrate that accounting for reporting delays can substantially improve the accuracy to forecast the COVID-19 incidence.

In Chapter 4, we find that reducing connections between the young and old populations can delay the death curve but cannot reduce the final mortality. We propose a merged Susceptible-Infectious-Removed (SIR) model, which advises elderly individuals to interact less with their non-elderly connections at the initial stage but interact more with their non-elderly relationships later, to better reduce mortality.

The Markovian SIS epidemics in a complex network with N nodes can be exactly described by the continuous-time Markov chain with 2^N network infection states. In Chapter 5, we propose a spectral clustering SIS approximation (SCSA) method, which combines the spectral clustering of the infinitesimal generator matrix (also known as the transition rate matrix) and the birth-and-death approximation, to reduces the huge 2^N state space of the Markov chain to a smaller number of states. We discover that the relationship between the approximation error ϵ and the number of clusters r in the spectral clustering roughly obeys $\epsilon \sim r^{-\alpha}$, where $\alpha \in (\frac{1}{4}, \frac{4}{5})$. The exponent α tends to be larger if the network has a higher link density.

In Chapter 6, we investigate a classical model for the accumulation of errors in multi-qubit quantum computations. By modeling the error process in a quantum computation using two coupled Markov chains, we are able to capture a weak form of time-dependency between errors in the past and future. By subsequently using techniques from the field of discrete probability theory, we calculate the probability that error quantities such as the fidelity and trace distance exceed a threshold analytically. Besides this, we study a model describing continuous errors accumulating in a single qubit. Finally, taking inspiration

from the field of operations research, we illustrate how our expressions can be used to decide how many gates one can apply before too many errors accumulate with high probability and how one can lower the rate of error accumulation in existing circuits through simulated annealing.

2

INFERRING NETWORK PROPERTIES BASED ON THE EPIDEMIC PREVALENCE

Dynamical processes running on different networks behave differently, which makes the reconstruction of the underlying network from dynamical observations possible. However, to what level of detail the network properties can be determined from incomplete measurements of the dynamical process is still an open question. In this chapter, we focus on the problem of inferring the properties of the underlying network from the dynamics of an susceptible-infected-susceptible (SIS) epidemic and we assume that only a time series of the epidemic prevalence, i.e., the average fraction of infected nodes, is given. We find that some of the network metrics, namely those that are sensitive to the epidemic prevalence, can be roughly inferred if the network type is known. A simulated annealing link-rewiring algorithm, called SARA, is proposed to obtain an optimized network whose prevalence is close to the benchmark. The output of the algorithm is applied to classify the network types.

2.1. INTRODUCTION

Graphs are the underlying structures of many systems and many dynamic processes on those systems can be modeled by a spreading process on their underlying graphs [2, 61, 62]. The difference in the underlying graphs may lead to contrasting dissimilar behavior of the process. One well-known result is that the mean-field epidemic threshold of the spreading process vanishes with the size of the scale-free network [63, 64], while the threshold of a sparse homogeneous network is non-zero. Another key difference is that a near-threshold spreading process is localized just above the threshold in a heterogeneous network, but delocalized in a homogeneous network [65, 66]. Moreover, the autocorrelation of the infection state of each node in a regular graph is irrelevant to the curing rate in the steady state [67]. In a real scenario, reviewing of the spreading data of cholera in London in 1854 under the susceptible-infected-susceptible (SIS) model indicates that the trajectory of the prevalence reflecting network properties supporting the hypotheses that the Broad Street pump was the source of the cholera outbreak and that cholera does not spread via the air [68]. Since the dynamics of different networks behave differently, the inverse question raises: "How much can we deduce about the underlying contact network by measuring the dynamics on the network?" The inverse question is meaningful when the direct measurement of the underlying graph is unavailable. For example, a disease control agency usually has the statistics of disease infection, but the underlying graph bearing the spreading of the disease is generally unknown.

Much work on the inverse problem exists [69, 70]. Most of the papers focus on reconstructing the underlying graphs by measuring the time-dependent dynamical state of each node [36, 37, 38, 39, 40, 41, 42, 43, 44]. With the complete dynamics of each node, the network may be approximately reconstructed by different heuristic algorithms, e.g., the Bayesian methods [45, 46], the conflict-based method [34], statistical inference based method [47] and the compressed sensing or lasso methods [35]. Different networked dynamical processes have been studied, such as the evolutionary game model [71, 72], the SIS model [35] and the Ising model [47]. Apart from reconstructing simple networks, there are many works on the reconstruction of the stochastic temporal networks [48], multilayer networks [49], weighted networks [50] and directed networks [51].

All of the above methods are based on the data from all or at least most nodes, but in real scenarios, individual-level observations of spreading are hard to obtain while most of the epidemic data are population-level [52, 53]. Motivated by the incompleteness of realistic situations, we study how much about the underlying network can be deduced with incomplete measurements. We assume that only the prevalence, which is the average fraction of infected nodes in the network is measured, but not the infection state of each node. Under this setting, network reconstruction does not seem possible, but inferring

some network properties may be possible, in particular, when additional information apart from the prevalence is available. In this chapter, we confine ourselves to four types of classical network models: the scale-free configuration (SF) graphs [73], the Barabási-Albert (BA) graphs [74], the Erdős-Rényi (ER) random graphs [75] and the Watts-Strogatz (WS) small-world graphs [76]. The network size N of these networks considered in this chapter is not larger than 2000. Additionally, we focus on the SIS epidemic process on networks, which is one of the basic models resembling the dynamics of many networked systems and assume that the infection and curing rate of the SIS process are known. Under our assumptions, part of the network properties can be inferred, provided that the network type is additionally given. Furthermore, the network type among the four above-mentioned graphs can be identified, given the network size N and the average degree $E[D]$, which is also emphasized by recent work from a different approach [77]: the ER, regular and BA graphs are distinguished by the epidemic prevalence.

This chapter is organized as following: In Section 2.2, we first briefly review the SIS process on networks. We further evaluate the correlation between the network metric difference and the corresponding SIS prevalence difference given the network type. A high correlation implies that, if an estimated network, whose prevalence is close to the benchmark prevalence, can be found, then the metric of this estimated network may be also close to the metric of the benchmark network. We further verify the possibility of estimating the network metrics, whose differences are highly correlated with the prevalence difference. In Section 2.3, we propose a simulated annealing link-rewiring algorithm (SARA) to find a possible network whose prevalence is close to the benchmark. The output of the algorithm is applied to classify the network types. In Section 2.4, we test the performance of SARA by inferring the structure of small networks and by forecasting the future trend of the prevalence. Finally, we conclude in Section 2.5.

2.2. CORRELATIONS BETWEEN THE SIS PREVALENCE AND NETWORK METRICS

2.2.1. PRELIMINARIES

We consider the SIS process on an unweighed, undirected network without self-loops. In the network, all the nodes are divided into two compartments: infected nodes and susceptible (healthy) nodes. An infected node can infect each healthy neighbor with rate β and the infected node can be cured spontaneously with rate δ, both as Poisson processes. If we denote the infection state of node i at time t by a Bernoulli random variable $X_i(t)$, with $X_i(t) = 1$ being infected and $X_i(t) = 0$ being healthy, the exact SIS

process of node i in an N-node network is governed by the following equation [31],

$$\frac{dE[X_i(t)]}{dt} = E\left[-\delta X_i(t) + [1 - X_i(t)]\,\beta \sum_{k=1}^{N} a_{ki} X_k(t)\right], \tag{2.1}$$

where $a_{ki} \in \{0,1\}$ is the element of the adjacency matrix A of the network. In the brackets of the right-hand side of (2.1), the first term represents the curing process and the second term represents the infection process. If the effective infection rate $\tau \triangleq \beta/\delta$ is above an epidemic threshold, then the infection can persist in the network; below the threshold, the epidemic dies out exponentially fast for sufficiently long time [78]. The endemic phase and all-healthy phase are identified by the time-dependent prevalence $y(t) = \frac{1}{N}\sum_{i=1}^{N} E[X_i(t)]$. in this chapter, the SIS prevalence is generated by an event-driven simulation based on the Gillespie algorithm [79, 80, 81].

Two different networks may produce a similar prevalence, and thus we need to understand which network properties are important factors in the SIS process. If the SIS prevalence is sensitive to a specific network metric, then the prevalence generated by two networks with different values of this metric may be distinct. Assume that we have a benchmark network with a metric M_b and an estimated network with the metric M_e. If the time series of the prevalence on the benchmark and estimated networks are $\{y_b(i\Delta t)\}_{i=0,\ldots,T-1}$ and $\{y_e(i\Delta t)\}_{i=0,\ldots,T-1}$, respectively, then their correlation can be evaluated by computing the prevalence difference

$$\mathscr{D}_p \triangleq \frac{1}{T}\sum_{i=0}^{T-1} \left| y_e(i\Delta t) - y_b(i\Delta t) \right| \tag{2.2}$$

and the metric difference

$$\mathscr{D}_G \triangleq |M_e - M_b|. \tag{2.3}$$

If we have n corresponding realizations of the differences $(\mathscr{D}_{pi}, \mathscr{D}_{Gi})$ for $i = 1,\ldots,n$, then we can compute their correlation by the Pearson correlation coefficient [4, p. 26],

$$\rho(\mathscr{D}_p, \mathscr{D}_G) \triangleq \frac{\sum_{i=1}^{n} (\mathscr{D}_{pi} - \overline{\mathscr{D}_p})(\mathscr{D}_{Gi} - \overline{\mathscr{D}_G})}{\sqrt{\sum_{i=1}^{n} (\mathscr{D}_{pi} - \overline{\mathscr{D}_p})^2}\sqrt{\sum_{i=1}^{n} (\mathscr{D}_{Gi} - \overline{\mathscr{D}_G})^2}}. \tag{2.4}$$

Only if $\rho(\mathscr{D}_p, \mathscr{D}_G)$ approaches one, then the metric M and the prevalence $y(t)$ are highly correlated, which indicates that inferring the metric from the prevalence may be possible.

2.2.2. EVALUATED NETWORK METRICS

The graph metrics considered in this section are shown in Tab. 2.1.

The assortativity ρ_D, which is the degree correlation between connected nodes [82], can

N	Network size (the number of nodes)
$E[D]$	Average degree
$E[D^2]$	Second moment of degree
$d_{\max}$	Largest degree
$E[H]$	Average shortest path length (the average hop-count)
$E[1/H]$	Global efficiency
λ_1	Spectral radius (the largest eigenvalue of the adjacency matrix)
μ_{N-1}	Algebraic connectivity (the second smallest eigenvalue of the Laplacian matrix)
ρ_D	Assortativity
C_G	Average clustering coefficient

Table 2.1: Graph metrics

be calculated as

$$\rho_D = \frac{L^{-1}\sum_i j_i d_i - [L^{-1}\sum_i \frac{1}{2}(j_i + d_i)]^2}{L^{-1}\sum_i \frac{1}{2}(j_i^2 + d_i^2) - [L^{-1}\sum_i \frac{1}{2}(j_i + d_i)]^2}, \tag{2.5}$$

where j_i and d_i are the degrees of the nodes at the ends of the ith link, with $i = 1, \cdots, L$, and L is the number of links.

The average clustering coefficient C_G, which is the probability that the node pairs with same neighbors are also connected, can be computed as

$$C_G = \frac{1}{N}\sum_{i=1}^{N} C_i = \frac{1}{N}\sum_{i=1}^{N} \frac{2\blacktriangle_i}{d_i(d_i-1)},$$

where $\blacktriangle_i$ is the number of triangles containing node i.

Some of the above metrics can be strongly correlated with the prevalence $y(t)$. For example, the epidemic threshold τ_c^{HMF} derived from the heterogeneous mean-field (HMF) approach [2] is

$$\tau_c^{\text{HMF}} = \frac{E[D]}{E[D^2]},$$

where D is the degree of a randomly selected node and the epidemic threshold $\tau_c^{(1)}$ derived from NIMFA [83] is

$$\tau_c^{(1)} = \frac{1}{\lambda_1}.$$

Many graph metrics can also be bounded. For example, the average degree follows $E[D] \leqslant \lambda_1$ in connected graphs [84] and the largest eigenvalue of the Laplacian matrix $\mu_1 \geqslant \frac{N}{N-1} d_{\max}$, while the algebraic connectivity is $\mu_{N-1} \leqslant d_{\min}$.

2.2.3. CORRELATION ANALYSIS

For any pair of networks, the prevalence difference $\mathscr{D}_p$ and the metric difference $\mathscr{D}_G$ can be calculated based on (2.2) and (2.3). For each network metric, we calculate the correlations via (2.4) between a set of metric differences $\mathscr{D}_G$ and their corresponding prevalence differences $\mathscr{D}_p$ on four network models: the SF graphs [73], the BA graphs [74], the ER random graphs [75] and the WS small-world graphs [76]. Specifically, the SF graphs are generated by the configuration model [73, 85] and the degree exponent parameter γ is uniformly at random chosen in the interval $[2.5, 3.0]$ in this chapter.

Specifically, we first randomly generate the four kinds of networks each with 100 realizations. The network sizes N and the average degrees $E[D]$ are chosen uniformly at random in the interval $[1000, 2000]$ and $[4, 12]$, respectively. The effective infection rate is set as $\tau = 3.0$, which is above the epidemic threshold of every network realization. Two kinds of initial state are chosen: $y_0 = 0.2$ or $y_0 = 1.0$, which means that 20% of the nodes are randomly chosen to be infected or all nodes are infected initially. For each network and initial state, a corresponding time series of the prevalence is obtained by averaging over 100 realizations of the SIS simulation. We mark the prevalence difference $\mathscr{D}_p$ under initial condition y_0 as $\mathscr{D}_p(y_0)$. We further denote the metric difference $\mathscr{D}_G$ for one specific metric as $\mathscr{D}_G(\text{metric})$. All metrics shown in Section 2.2.2 are considered and the Pearson correlation coefficients $\rho\left(\mathscr{D}_p(y_0), \mathscr{D}_G(\text{metric})\right)$ are calculated by Eq. (2.4). The sample size of each correlation coefficient is $\binom{100}{2} = 4950$. Table 2.2 and Tab. 2.3 indicate that there are generally strong correlations between the difference of the prevalence $\mathscr{D}_p$ and the differences of the average degree $E[D]$, the second moment of degree $E[D^2]$, the average shortest path length $E[H]$, the global efficiency $E[1/H]$ and the spectral radius λ_1. A strong positive correlation indicates that the metric between two networks with the same network type can be similar if they have similar prevalence curves. However, there are relatively weak correlations between the difference of the prevalence $\mathscr{D}_p$ and the differences of the network size N, the largest degree $d_{\max}$, the algebraic connectivity μ_{N-1}, the assortativity ρ_D and the average clustering coefficient C_G. Moreover, the initial state has very slight influence on the correlations.

To summarize, if the type of the underlying graph is given, then inferring the network properties, whose differences $\mathscr{D}_G$ are highly correlated to the difference of the prevalence $\mathscr{D}_p$, is possible. A straightforward method is randomly generating the network realizations by the corresponding network model and selecting the one realization produces minimum prevalence difference $\mathscr{D}_p$.

$\rho\left(\mathscr{D}_p(y_0),\mathscr{D}_G(\text{metric})\right)$	$\mathscr{D}_G(E[D])$	$\mathscr{D}_G(E[D^2])$	$\mathscr{D}_G(\lambda_1)$	$\mathscr{D}_G(E[H])$	$\mathscr{D}_G\left(E\left[\frac{1}{H}\right]\right)$
ER graphs, $\mathscr{D}_p(y_0=0.2)$	0.941	0.856	0.940	0.953	0.939
WS graphs, $\mathscr{D}_p(y_0=0.2)$	0.877	0.826	0.921	0.952	0.958
BA graphs, $\mathscr{D}_p(y_0=0.2)$	0.940	0.838	0.871	0.952	0.945
SF graphs, $\mathscr{D}_p(y_0=0.2)$	0.944	0.612	0.561	0.861	0.823
ER graphs, $\mathscr{D}_p(y_0=1.0)$	0.947	0.866	0.944	0.948	0.932
WS graphs, $\mathscr{D}_p(y_0=1.0)$	0.905	0.818	0.927	0.952	0.954
BA graphs, $\mathscr{D}_p(y_0=1.0)$	0.945	0.856	0.908	0.954	0.948
SF graphs, $\mathscr{D}_p(y_0=1.0)$	0.948	0.631	0.459	0.792	0.783

Table 2.2: Metrics with strong positive correlations

$\rho\left(\mathscr{D}_p(y_0),\mathscr{D}_G(\text{metric})\right)$	$\mathscr{D}_G(d_{\max})$	$\mathscr{D}_G(C_G)$	$\mathscr{D}_G(\mu_{N-1})$	$\mathscr{D}_G(\rho_D)$	$\mathscr{D}_G(N)$
ER graphs, $\mathscr{D}_p(y_0=0.2)$	0.821	0.477	0.490	−0.014	−0.059
WS graphs, $\mathscr{D}_p(y_0=0.2)$	0.805	−0.036	−0.002	0.624	−0.012
BA graphs, $\mathscr{D}_p(y_0=0.2)$	0.386	0.358	0.854	0.595	−0.031
SF graphs, $\mathscr{D}_p(y_0=0.2)$	0.398	0.182	0.657	0.013	−0.038
ER graphs, $\mathscr{D}_p(y_0=1.0)$	0.856	0.525	0.524	0.082	−0.018
WS graphs, $\mathscr{D}_p(y_0=1.0)$	0.807	−0.031	0.081	0.666	−0.039
BA graphs, $\mathscr{D}_p(y_0=1.0)$	0.284	0.410	0.813	0.535	−0.003
SF graphs, $\mathscr{D}_p(y_0=1.0)$	0.247	0.100	0.659	0.006	0.034

Table 2.3: Metrics with weak positive correlations

2.2.4. INFERRING NETWORK METRICS GIVEN THE NETWORK TYPE

We further try to infer the network metrics based on the prevalence from a single realization of the SIS process given the network type. Specifically, for each network type, we first generate 1000 benchmark networks whose network sizes N and average degrees $E[D]$ are chosen uniformly at random in the interval $[200,500]$ and $[4,8]$, respectively. For each benchmark network, one corresponding benchmark prevalence is generated from only one realization of the SIS process.

We then try to estimate the network metrics of each benchmark network as follows. For each benchmark, 1000 networks with the same network type as the benchmark network are generated. The network sizes N and average degrees $E[D]$ of the generated networks are also chosen uniformly at random in the interval $[200,500]$ and $[4,8]$, respectively. The network with the smallest prevalence difference $\mathscr{D}_p$ to the benchmark is selected as the estimated network. The metrics of this estimated network are regarded as the estimated metrics of the benchmark network.

We measure the performance of the metric inference under the mean absolute error (MAE) and the mean squared error (MSE). The MAE and MSE for n underlying graphs is given by

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |M_{ei} - M_{bi}| \tag{2.6}$$

and

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (M_{ei} - M_{bi})^2, \tag{2.7}$$

where M_{ei} and M_{bi} denote the estimated and real metrics of the benchmark network G_i, $i = 1,2,\cdots,n$.

Tables in the supplementary material A.1 show MAE and MSE of each network metric for different network types (the ER random graphs, the WS small-world graphs, the BA graphs and the SF graphs). For the treatment group, we calculate MAE and MSE of each metrics which are estimated by selecting the network whose prevalence is closest to the benchmark. For the control group, we calculate MAE and MSE of each metrics which are estimated by randomly generating a network whose network sizes N and average degrees $E[D]$ are chosen uniformly at random in the interval $[200,500]$ and $[4,8]$. For the network metrics whose differences are closely correlated with the prevalence difference $\mathscr{D}_p$, i.e., the average degree $E[D]$, the second moment of degree $E[D^2]$, the average shortest path length $E[H]$, the global efficiency $E[1/H]$ and the spectral radius λ_1, their MAE and MSE of the treatment group are much less than those of the control group, which indicates that these metrics can be roughly deduced based on the prevalence given the network type. However, for the network metrics whose differences are weakly correlated with the prevalence difference $\mathscr{D}_p$, i.e., the network size N, the largest degree $d_{\max}$, the algebraic

connectivity μ_{N-1}, the assortativity ρ_D and the average clustering coefficient C_G, their MAE and MSE of the treatment group are close to those of the control group.

2.3. DISTINCTION BETWEEN NETWORK TYPES

In this section, we try to distinguish the type of the underlying network given the time series of the prevalence $\{y_b(i\Delta t)\}_{i=0,\dots,T-1}$, the network size N, the number of links L and the effective infection rate τ. We propose a simulated annealing link-rewiring algorithm (SARA) to optimize a network whose prevalence can be close to the input prevalence benchmark and the performance difference between different rewiring mechanisms in SARA can be applied to identify the graph type.

2.3.1. SIMULATED ANNEALING LINK-REWIRING ALGORITHM (SARA)

The basic principle of SARA is that the links of an estimated network are continually rewired based on different rewiring methods to minimize the prevalence difference D_p between the optimized network and the benchmark network.

The algorithm operates iteratively and a random network is initialized. In each iteration, the network will be renewed by rewiring the links of partial nodes. A new corresponding time series of the prevalence $\{y_e(i\Delta t)\}_{i=0,\dots,T-1}$ can be generated by simulating the SIS process on the network and its difference D_p to the benchmark time series of the prevalence $\{y_b(i\Delta t)\}_{i=0,\dots,T-1}$ is calculated. If the difference D_p decreases, then the rewired network will be accepted. If D_p increases, then the rewired network is accepted with an acceptance probability p and rejected with rejection probability $1-p$ to prevent local optima. Moreover, a stable final converging result is obtained provided that the acceptance probability p decreases with the iterations. The final result of this algorithm is an estimated graph, whose corresponding prevalence is almost the same as the benchmark prevalence. Inspired by the generation processes of ER graphs and BA scale-free graphs, we consider two different rewiring methods: randomly connecting (RC) and preferential attachment (PA). In RC, the selected nodes are rewired uniformly at random to the rest of the nodes in the network, and in PA, the selected nodes are rewired to a node with probability proportional to the node's degree. The pseudo-code of SARA is shown in Algorithm 5.

2.3.2. DISTINCTION BETWEEN THE NETWORK TYPES

We try to distinguish four kinds of graphs (the SF graphs, the BA graphs, the ER random graphs and the WS small-world graphs) based on the optimized prevalence curves generated by SARA. The experiment and the results are as follows. For each network model, we generate 100 network realizations with $N = 1000$ nodes and $L = 4000$ links as the

Algorithm 1: Pseudo-code of the simulated annealing link-rewiring algorithm (SARA)

Input : $\{y(i\Delta t)\}_{i=0,\ldots,T-1}$, N, L, τ, initial temperature V_{tmp}, cooling rate $0 < r < 1$ and step length S_N

Output : Estimated network G_e, final prevalence difference $\mathscr{D}_p$

An initial network is chosen uniformly at random from the set of all networks with N nodes and L links.

for *iteration bound* **do**

  Uniformly randomly choose $N_c = \mathrm{round}(S_N \times \mathscr{D}_p)$ nodes.

  Delete all links of each chosen node and then rewire these links to new neighbors.

  If we randomly choose new neighbors without preference (RC), the probability p_i that the rewired link is connected to a neighbor i is $p_i = 1/(N - N_c)$, where node i belongs to the $N - N_c$ uncollected nodes.

  If we rewire links based on preferential attachment mechanism (PA) , the probability p_i that the rewired link is connected to a neighbor i is $p_i = d_i / \sum_j d_j$, where d_i is the degree of node i in residual network and the sum is made over all unselected nodes.

  If n isolated nodes appear after the rewiring process in step 5 or step 6, we remove n links uniformly at random and rewire them to the isolated nodes based on the RC or PA mechanism, respectively. This step continues until there is no isolated node in the network.

  Simulate the SIS process on the new network and calculate the prevalence difference $\mathscr{D}_2$ to the benchmark.

  if $\mathscr{D}_2 < \mathscr{D}$ **then**

    $\mathscr{D} \longleftarrow \mathscr{D}_2$; $G \longleftarrow G_2$;

  else if $\mathrm{Exp}(-(\mathscr{D}_2 - \mathscr{D})/V_{\mathrm{tmp}}) > \mathrm{random}(0,1)$ **then**

    $\mathscr{D} \longleftarrow \mathscr{D}_2$; $G \longleftarrow G_2$;

  end

  $V_{\mathrm{tmp}} = r \times V_{\mathrm{tmp}}$;

end

benchmark networks. For the SF graphs, the degree exponent γ ranges in the interval $\gamma \in [2.5, 3.0]$. For the SW graphs, the rewiring probability $p_r \in [0.5, 1.0]$. The corresponding time series of the prevalence are obtained by averaging 10 realizations with effective infection rate $\tau = 1$, which is above the epidemic threshold for benchmark networks. For each benchmark graph realization and corresponding prevalence, we apply SARA with RC and PA mechanisms separately and obtain two corresponding prevalence differences $\mathscr{D}_{\mathrm{RC}}$ and $\mathscr{D}_{\mathrm{PA}}$ from the final output of the optimization, respectively. The performance difference between these two rewiring mechanisms provides a possibility of identifying the types of underlying graphs by applying different rewiring methods in SARA. We then try to classify the networks by the difference value $\mathscr{D}_{\mathrm{RC}} - \mathscr{D}_{\mathrm{PA}}$. Figure 2.1 shows that these four kinds of networks can be almost exactly classified by the difference value $\mathscr{D}_{\mathrm{RC}} - \mathscr{D}_{\mathrm{PA}}$. Indeed, $\mathscr{D}_{\mathrm{RC}} > \mathscr{D}_{\mathrm{PA}}$ for almost all SF and BA graphs while $\mathscr{D}_{\mathrm{RC}} < \mathscr{D}_{\mathrm{PA}}$ for almost all ER and SW graphs as shown in Fig. 2.1a. We examine the classification performance by the receivers operating characteristic (ROC) curve, which is a curve of the True Positive Rate (TPR)

$$R_{\mathrm{TPR}}(d) = \frac{N_{\mathrm{TP}}(d)}{N_{\mathrm{TP}}(d) + N_{\mathrm{FN}}(d)}$$

against the False Positive Rate (FPR)

$$R_{\mathrm{FPR}}(d) = \frac{N_{\mathrm{FP}}(d)}{N_{\mathrm{FP}}(d) + N_{\mathrm{TN}}(d)},$$

where d is the threshold of the difference value $\mathscr{D}_{\mathrm{RC}} - \mathscr{D}_{\mathrm{PA}}$, $N_{\mathrm{TP}}(d)$ is the number of true positives of $\mathscr{D}_{\mathrm{RC}} - \mathscr{D}_{\mathrm{PA}} > d$, $N_{\mathrm{FP}}(d)$ is the number of false positives of $\mathscr{D}_{\mathrm{RC}} - \mathscr{D}_{\mathrm{PA}} > d$. The denominators $N_{\mathrm{TP}}(d) + N_{\mathrm{FN}}(d)$ and $N_{\mathrm{FP}}(d) + N_{\mathrm{TN}}(d)$ are the number of real positives and real negatives of $\mathscr{D}_{\mathrm{RC}} - \mathscr{D}_{\mathrm{PA}} > d$, respectively.

The area under the ROC curve (AUC) depicts the accuracy of classification. If AUC = 1, then the classification is perfect. In Figure 2.1b and Fig. 2.1d, the ROC curves of the difference value $\mathscr{D}_{\mathrm{RC}} - \mathscr{D}_{\mathrm{PA}}$ between any two kinds of networks show that these networks can be distinguished almost exactly.

2.4. ESTIMATING THE TOPOLOGY OF SMALL NETWORKS AND PREVALENCE

2.4.1. THE NETWORK OUTPUT OF SARA: AN EXAMPLE

In this section, we test the feasibility of approximately reconstructing small graphs from the prevalence. We show example output of SARA under the benchmark of a small tree network and a small wheel network. In SARA, the initialized networks are chosen uniformly at random from all networks with the same number of nodes and links as the benchmark networks. The rewiring methods are selected to be the one with a smaller

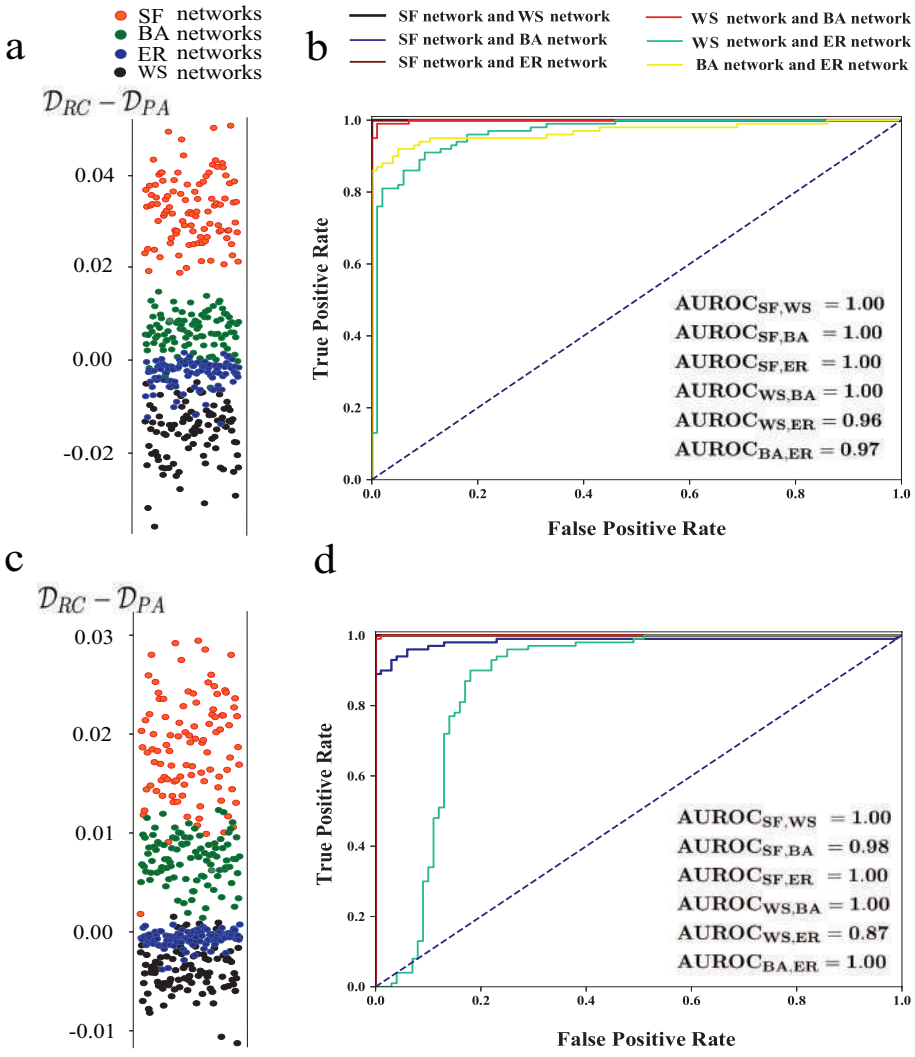


Figure 2.1: The classification of network types. a) and b): The results for the initial state $y_0 = 0.2$. c) and d): The results for the initial state $y_0 = 1.0$. The time series of the prevalence are obtained by averaging over 10 realizations.

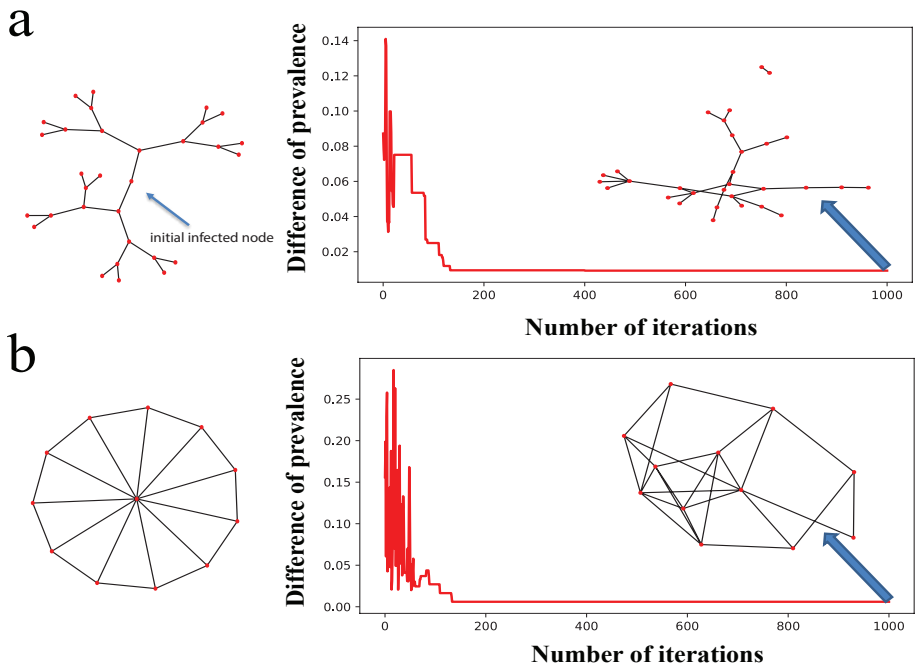


Figure 2.2: The reconstruction of a tree network and a wheel network. The left parts are the benchmark underlying network and the right parts are the estimated networks. The curves are the difference of prevalence against the number of iterations. The difference of prevalence is already small when the number of iterations is around 150. The prevalence curves are obtained by averaging 500 realizations and only the central node is infected initially.

difference of the prevalence in the output. As shown in Fig. 2.2, the main features of the benchmark networks are captured fairly well by the final output of SARA.

2.4.2. FORECAST THE FUTURE TRENDS OF EPIDEMIC PREVALENCE

Any benchmark prevalence from either homogeneous or heterogeneous networks can be fitted well by SARA. Therefore, we can further analyze the feasibility of predicting the future prevalence evolution by fitting the few initial prevalence observations.

We fit only the initial part (10%) of the time series of the prevalence $\{y(i\Delta t)\}_{i=0,\ldots,\lfloor T/10\rfloor-1}$ generated by four different benchmark networks and compare the whole prevalence output of the algorithm with the benchmark prevalence. RC rewiring is applied for ER and WS graphs, and PA rewiring is applied for BA and SF graphs. As shown in Fig. 2.3a about the ER and WS graphs, the estimated prevalence (dashed curves) are close to the benchmark (solid curves). However, as shown in Fig. 2.3b, the prediction is inaccurate for BA and SF

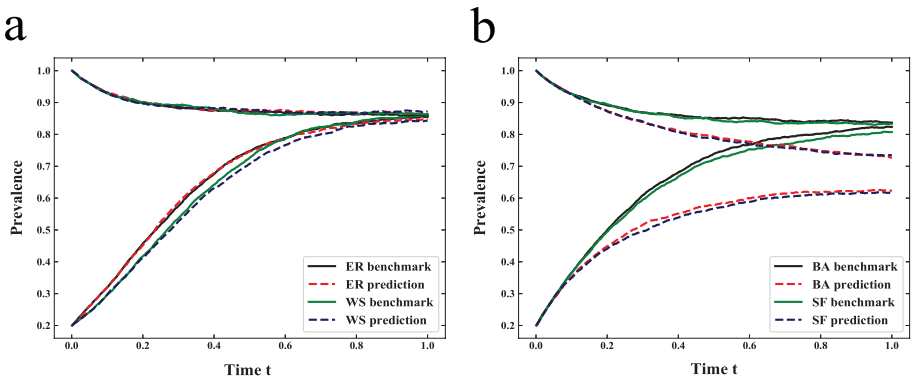


Figure 2.3: Forecast the future trend of epidemic prevalence. We fit the curves before $t = 0.1$ and forecast the rest prevalence. a) The results about ER graphs and WS graphs. b) The results about BA graphs and SF graphs. Two kinds of initial states are chosen: $y_0 = 0.2$ and $y_0 = 1.0$. The time series of the prevalence are the mean of 10 propagation.

graphs.

2.5. CONCLUSION

We study the feasibility of inferring properties of the underlying graphs based on the SIS prevalence. Pearson's correlations (2.4) between the differences of prevalence and the network metrics are evaluated. Given network type, the difference of the epidemic prevalence is highly related to the differences of some network metrics, such as the average degree $E[D]$, the second moment of degree $E[D^2]$, the average shortest path length $E[H]$, the global efficiency $E[1/H]$ and the spectral radius λ_1. If the network type is known, then these metrics can be roughly estimated by finding a network whose prevalence curve is close to the benchmark. To distinguish the network type, we further propose an algorithm SARA, which can find a network whose epidemic prevalence is close to the benchmark. Given network size and the number of links, four network types (the SF graphs, the BA graphs, the ER random graphs and the WS small-world graphs) can be classified by different rewiring methods combined with SARA. Visually, the output network of SARA captures the features of small benchmark networks well. Finally, we show that it is possible to predict the later prevalence from the initial stage prevalence for homogeneous networks.

The epidemic prevalence in the SIS model resembles the population-level observations. Population-level observations lose details of nodal infection but may still provide information about the underlying network. In real scenarios, the population-level ob-

servations are available for many different infectious diseases, such as influenza, Ebola virus disease, Zika virus disease, etc. Disease control agencies may take advantage of the population-level observations to understand the detailed spreading pattern, further forecast the outbreaks more accurately and control the diseases more efficiently. For example, a small diameter of the network inferred by the population-level observations implies that modern transportation plays a role; a large clustering coefficient means that spreading is effectively exploring a community or geographical area; using the initial stage prevalence, it is possible to approximately reconstruct the small-size local network containing the initial infections.

This chapter considers the average of multiple epidemic outbreaks. However, the prevalence of each epidemic outbreak may contain more information on the network properties, which could be further studied in future works.

3

ACCOUNTING FOR COVID-19 REPORTING DELAYS TO ENHANCE THE ACCURACY OF EPIDEMIC FORECASTS

COVID-19 is a global pandemic that has directly affected 220 millions of individuals worldwide. While the rapid research and development of COVID-19 vaccines and their timely deployment is expected to mitigate the pandemic in the near future, the planning of the COVID-19 containment measures and vaccine distribution strategies is currently of utmost importance. All these activities are impossible without accurate data reports of COVID-19 clinical cases. Through the statistical analysis of COVID-19 data reports in several countries we identify discrepancies in correlations among the infected, recovered and deceased time series. We explain the observed discrepancies by the existence of reporting delays in data and devise a statistical framework to infer delay mechanisms. We demonstrate that the accounting for reporting delays can substantially improve the accuracy to forecast the COVID-19 incidence. We anticipate that our findings and the developed statistical framework will prove useful in forecasts of not only COVID-19 but also other viral infections.

3.1. INTRODUCTION

The recent outbreak of the Coronavirus disease 2019 (COVID-19), which is caused by Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2), asks for a precise understanding of the transmission mechanisms of this virus among people [86, 87, 88, 89, 90]. The Susceptible-Infectious-Recovered-Dead (SIRD) model is one of the most basic epidemic models to describe the COVID-19 spreading [7, 6, 2, 9]. There are many variations on the basic SIRD model [91, 92, 93, 94, 95], e.g., considering the exposed state [5, 2], the time-varying infection rate [96] and the underlying traffic network [7]. Recent works attempt to forecast how many people will be infected or deceased in the future by fitting the history data via epidemic models [89, 93, 97, 98, 99, 100, 101, 102, 103].

A prominent topic in epidemiology is the reporting delays of cases [30, 104] defined as the time difference between two events: 1. an individual was infected, recovered or deceased; 2. the infected, recovered or deceased case was reported. The reporting delays occur due to many reasons, e.g., the patient delay, the diagnosis delay and the public health authorities (PHA) delay [14]. The reporting delays of infections are found to be significant for many infectious diseases [14, 15]. Specifically, the reporting delays of infections for some diseases, e. g., Hepatitis B, Shigellosis and Salmonella, can take on average 2-3 weeks [14, 16]. The cases for some diseases, e. g., Hepatitis A, Measles and Mumps, are usually reported after eight days [14]. The median diagnosis delay for Malaria is around four days [17]. Research on the Middle East Respiratory Syndrome CoronaVirus (MERS-CoV), a kind of virus similar to SARS-CoV-2, found a time difference between the symptom onset and confirmation around four days [18]. Recent researches on COVID-19 reveal significant reporting delays of infections for China [19, 20, 21, 22, 23, 24], Italy [25], Germany [26, 27], Singapore [28], the USA [29] and the UK [30]. These reporting delays are fitted by the negative binomial distribution [25, 27], the log-normal distribution [20], the Weibull distribution [27] and the gamma distribution [30, 23, 28]. Dehning *et al.* [26] and Mitze *et al.* [105] considered the reporting delay of infections in the data-driven modeling of the COVID-19 pandemic for Germany. No reporting delay distributions for the recovered and deceased individuals have been considered in COVID-19 modeling and forecasts.

This chapter focuses on reporting delays of infections, recoveries and deaths in COVID-19 data and the effect of reporting delays on epidemic forecasts. We first discover that, to some extent, the infection, recovery and death data disagree with many commonly-used compartmental epidemic models. There is a time shift for the real time series and we hypothesize that the reporting delays could be an essential factor that cannot be neglected. We model the reporting delays of infections, recoveries and deaths and further explain the inconsistency between the COVID-19 data and the compartmental epidemic model

by considering these reporting delays. Next, we propose a correlation-based method to infer the reporting delay information for different countries. Significant reporting delays of recoveries are found in many countries such as Italy and Spain. One possible reason is that the discharge criteria for the recovered patients are usually strict due to the risk of a relatively long SARS-CoV-2 virus shedding after remission of symptoms [106, 107]. Finally, we analyze the impact of reporting delays on COVID-19 forecasts. The results reveal that the forecast accuracy can be significantly improved by considering the reporting delays.

Nomenclature. The data one can directly obtain is the reported data with delays. The delays work on daily new cases [26], which are different from the compartments in epidemic models. To prevent confusion among symbols, we provide the critical symbols in Table 3.1.

Fractions of compartments	Fractions of new cases in each day
Y: fractions of cases	ΔY: fractions of new cases
$\widetilde{Y}$: fractions of reported cases	$\Delta \widetilde{Y}$: fractions of reported new cases
$\hat{Y}$: fractions of predicted cases	$\Delta \hat{Y}$: fractions of predicted new cases

Table 3.1: This work considers the fractions, which are the proportions of the number of cases to total population. The left notations are fractions applied in the compartmental epidemic models and the right notations are fractions applied in the reporting delays. The fraction Y can denote the fraction of infectious cases I, the fraction of recovered cases R or the fraction of deceased cases D.

3.2. EVIDENCE FOR REPORTING DELAYS IN EPIDEMIOLOGICAL DATA

We begin the exposition by considering daily reports on the number of infected, recovered and deceased individuals in Spain, whose time series are relatively smooth compared with the epidemiological data of many other countries. We observe in Fig. 3.1(A) that both the fraction of reported new deceased cases $\Delta\widetilde{D}$ and the fraction of reported infectious cases $\tilde{I}$ peaked within the second month of the COVID-19 pandemic. The times, at which these peaks occurred, are substantially different: the fraction of reported new deceased cases $\Delta\widetilde{D}$ reached its peak on April 1, 2020, while the largest fraction of infectious individuals $\tilde{I}$ was reported 22 days later on April 23, 2020. This observation is not specific to Spain: the peak of the fraction of reported new deceased cases precedes that of the fraction of reported infectious cases by more than one week not only for Spain but also for many countries, including China and Turkey, Fig. B.1. We make similar observations for the fraction of reported new recoveries $\Delta\widetilde{R}$, which also exhibit the peak after that of the fraction of reported new deaths $\Delta\widetilde{D}$, Fig. 3.1(A) and Fig. B.1. When plotted as a function

of fraction of daily new deceased cases $\Delta\widetilde{D}$, fractions of reported infectious cases $\tilde{I}$ and daily recovered cases $\Delta\widetilde{R}$ form loop patterns, see Fig. 3.1(B),(C) and Fig. B.2.

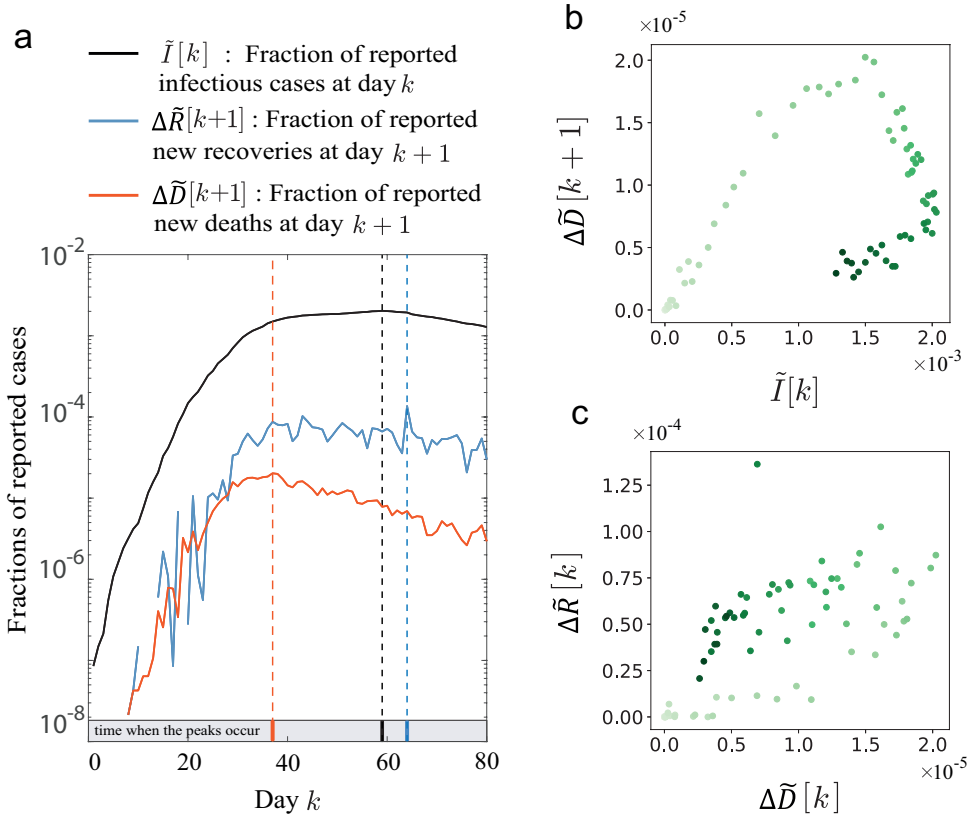


Figure 3.1: Time series of the fractions of reported infectious cases $\tilde{I}[k]$, reported new recovered cases $\Delta\widetilde{R}[k+1]$ and reported new deceased cases $\Delta\widetilde{D}[k+1]$ for Spain (panel a). The peak locations (marked with the vertical dashed lines) of $\tilde{I}[k]$, $\Delta\widetilde{R}[k+1]$ and $\Delta\widetilde{D}[k+1]$ are remarkably different. Panels b and c are data points between two of the three fractions. Symbol color (from light to dark green) show the day k changing from day 0 to day 80. Loop patterns instead of straight lines are observed. Data points can evolve in a counter-clockwise or a clockwise direction, indicating which data is more delayed than another. The first day $k = 0$ is February 25, 2020.

The observed patterns carry the evidence of reporting delays. Indeed, recent observations [108, 109, 110, 111] indicate the recovery rate γ_r and the fatality rate γ_d are almost not time-varying during the first wave of the pandemic, suggesting that the peaks of the observed fraction of infectious individuals should coincide with those of the daily recovery and deceased fractions for the first wave of the COVID-19 pandemic, contradicting Fig. 3.1(A). To explain the formation of the loop patterns in Fig. 3.1(B,C), we employ the SIRD discrete time compartmental model [5, 2]. Within the SIRD model, the population

is split into four compartments: susceptible (S), infectious (I), recovered(R), deceased (D). Compartment S denotes the fraction of susceptible individuals who can be infected by the infectious individuals. Compartment I denotes the fraction of individuals who have been infected but haven't recovered or deceased. The compartments R and D are respectively the fractions of individuals who have recovered or deceased. The SIRD model assumes that recovered individuals become immune and cannot be infected by the virus in the future. Discrete-time transitions between the compartments are governed by

$$\begin{aligned} I[k+1] - I[k] &= \beta I[k] S[k] - (\gamma_r + \gamma_d) I[k], \\ \Delta R[k+1] &\equiv R[k+1] - R[k] = \gamma_r I[k], \\ \Delta D[k+1] &\equiv D[k+1] - D[k] = \gamma_d I[k], \\ \Delta I[k+1] &\equiv I[k+1] - I[k] + \Delta R[k+1] + \Delta D[k+1], \end{aligned} \tag{3.1}$$

where β, γ_r and γ_d are SIRD model parameters quantifying, respectively, the infection rate, the recovery rate and the deceased rate.

We use the SIRD model to simulate COVID-19 spreading without and with reporting delays, observing in the latter case peak shifts and loop patterns similar to those in Fig. 3.1, see Figs. B.3-Fig. B.11. Intuitively, the observed loop patterns are the result of the effective time shift between two non-monotonous time series. The upper part of the loop in Fig. 3.1b is due to the fact that initially $\Delta D[k]$ data grows faster than infectious data $I[k]$. The lower part of the loop is observed when $\Delta D[k]$ is decreasing after experiencing its peak, while $I[k]$ is still increasing. The shape of the observed loop patterns depends on the effective delay between the datasets and the exact shape of each data function.

STATISTICAL FRAMEWORK FOR UNCOVERING REPORTING DELAYS

To uncover reported delays we employ the following null model. For brevity, we present the rigorous derivation of the model in Appendix B.2 and only provide a simple argument here. Within the model, each individual i is endowed with Y_i random time of getting infected (recovered or deceased) and T_i random time of delay report, resulting in the random reported time as $K_i = Y_i + T_i$. Assuming that the reporting delays T_i are independent of Y_i, we obtain

$$\Pr[K_i = k] = \sum_{y=0}^{\infty} \sum_{m=0}^{\infty} \Pr[Y_i = y] \Pr[T_i = m] \delta_{y+m,k},$$

where δ is the Kronecker delta function. Assuming that delay times are independent and identically distributed, $\Pr[T_i = m] \leftarrow \Pr[T = m]$ for $m = 0, 1, 2, \cdots$, we obtain

$$\Pr[K_i = k] = \sum_{m=0}^{k} \Pr[Y_i = k - m] \Pr[T = m].$$

By ignoring the correlations between individual epidemic events, $\Delta\widetilde{Y}[k] = \frac{1}{N}\sum_i \Pr[K_i = k]$, we obtain

$$\Delta\widetilde{Y}[k] = \sum_{m=0}^{k} \Pr[T = m]\Delta Y[k-m], \tag{3.2}$$

where $Y = \{I, R, D\}$ and $T = \{T_I, T_R, T_D\}$. Equations [3.1] and [3.2] form the statistical framework for uncovering reporting delays and improving epidemic forecasts.

3

3.3. UNCOVERING REPORTING DELAYS

To uncover reporting delays and reconstruct the original epidemic data $Y = \{I, R, D\}$ from the reported data $\widetilde{Y} = \{\widetilde{I}, \widetilde{R}, \widetilde{D}\}$, we aim to find the delay distribution $\Pr[T = m]$ for days $m = 0, 1, 2, \cdots$ based the correlations among the fractions of new recoveries ΔR, new deaths ΔD and infectious cases I in Eq. [3.1].

While this inference problem can be formulated and solved in a non-parametric way, to simplify the exposition we solve the problem parametrically. We assume that the shape of delay distributions $\Pr[T = m]$ are known but their parameters $\boldsymbol{\kappa}$ are not. We consider three families of two-parameter discrete-time distributions that are appropriate for skewed data [112, 113]: the negative binomial distribution, the Neyman type A distribution and the Pólya-Aeppli distribution, see 3.5.1. These distributions are related to well-known one-parameter distributions. The logarithmic, geometric and Poisson distributions are, for instance, all special cases of the negative binomial distribution [112]. For the same mean and variance, the shapes of the negative binomial distribution and the Neyman type A distribution are usually very different and the Pólya-Aeppli distribution is in between of these two distributions [114]. For brevity, we only investigate the Pólya-Aeppli distribution in the main text and present the results for the other two distributions in Figs. B.3-Fig. B.11. There are in total 6 parameters in the delay distributions to be optimized. For the Pólya-Aeppli distributions, the parameter vector $\boldsymbol{\kappa} = (\lambda_I, \theta_I, \lambda_R, \theta_R, \lambda_D, \theta_D)^\intercal$, where λ_Y and θ_Y are parameters of the Pólya-Aeppli distribution describing the reporting delays of data $Y = \{I, R, D\}$.

We then aim to determine parameters $\boldsymbol{\kappa}$ of the distributions using the information that the true data ΔR, ΔD and I are proportional to each other. If we use a candidate parameter vector $\bar{\boldsymbol{\kappa}}$ which is far different from the true parameters $\boldsymbol{\kappa}$, the Pearson correlation coefficients between the inferred data $\Delta\bar{R}$, $\Delta\bar{D}$ and $\bar{I}$ which are obtained by solving Eq. [3.2] will be significantly smaller than 1. We thus infer the delay distributions by maximizing the product of three pairwise Pearson correlation coefficients

$$\mathscr{O}(\widetilde{Y}, \bar{\boldsymbol{\kappa}}) \equiv \rho(\Delta\bar{R}, \Delta\bar{D})\rho(\bar{I}, \Delta\bar{R})\rho(\bar{I}, \Delta\bar{D}). \tag{3.3}$$

We apply the random search [115] to determine $\bar{\boldsymbol{\kappa}}$ maximizing correlations $\mathscr{O}(\widetilde{Y}, \bar{\boldsymbol{\kappa}})$. Here we conduct a large set of independent random iterations. In each iteration, we uniformly

at random select the delay parameters in vector $\bar{\boldsymbol{\kappa}}$. We use the selected delay parameters inferred $\Delta\bar{R}$, $\Delta\bar{D}$ and $\bar{I}$ and to compute the corresponding $\mathcal{O}(\widetilde{Y},\bar{\boldsymbol{\kappa}})$. After conducting all iterations, we select the delay parameters maximizing $\mathcal{O}(\widetilde{Y},\bar{\boldsymbol{\kappa}})$, see 3.5.2.

To test the accuracy of the random search, we conducted a series of experiments using synthetic data. We first generate synthetic curves of the daily cases $\Delta I[k]$, $\Delta R[k]$ and $\Delta D[k]$ at day $k = 1,2,\cdots$ that close to the real data for Spain. Second, we uniformly randomly select the mean delays for deaths $E[T_D] \in [0,5]$ days, the mean differences $E[T_I] - E[T_D] \in [0,20]$ days, $E[T_R] - E[T_D] \in [0,20]$ days, parameters $\theta_D \in [0,1]$, $\theta_I \in [0,1]$ and $\theta_R \in [0,1]$ to generate the synthetic delay parameters $\boldsymbol{\kappa}$. We consider the situation that the reporting delays for deaths are relatively short due to the diagnosis of deceased cases is easier than the diagnosis of the infected or recovered cases. Besides, the real data of reporting delays for the UK also show that the expected reporting delays for infections are longer than the delays for deaths [30, 116]. We generate 50 different delay parameter vectors $\boldsymbol{\kappa}$ and add these reporting delays to the synthetic curves of the daily cases $\Delta I[k]$, $\Delta R[k]$ and $\Delta D[k]$ by (3.2). Finally, we infer the delay information by the random search. Figure 3.2(A) reveals that the inference errors decrease as a function of the number of iteration stops reaching their minimal values after 10^7 iterations. Figure 3.2(B) shows that the true and inferred values are close and most errors are within 2 days. Figure 3.2(C) further indicates that, compared with the reported curves, the inferred curves are significantly closer to the data with no reporting delays.

After testing the inference algorithm on synthetic data, we move on to uncover reporting delays in 8 regions of interest. Table 3.2 summarizes uncovered delays in 8 regions located in Europe, Asia and the Middle East. Although the inferred delay distributions are significantly different for different countries, The reporting delays for deceased cases are always the shortest. As what previous works have discovered, there are significant reporting delays of infections. We also discover that there are relatively long reporting delays of recoveries for Italy, Spain, Turkey and China, which haven't studied before. Specifically, the mean reporting delays for recoveries $E[T_R]$ are more than two weeks in these countries. On the contrary, there are relatively small expected delay $E[T_R]$ for Denmark, Romania and Germany. We also note that the standard deviations of reporting delays tend to vary across different regions. The standard deviations for the delays in recovered data, σ_R, for Germany and Romania are larger than the other regions, indicating that some recovered events for Germany and Romania are reported with extremely long delays. The standard deviations for the delays in infected data, σ_I, for Italy and Spain are smaller than the other regions, revealing that the infection cases in Italy or Spain are reported with almost same delays. Table 3.2 also shows the maximum objective function $\mathcal{O}(\widetilde{Y},\bar{\boldsymbol{\kappa}})$ by the random research. The maximum objective function $\mathcal{O}(\widetilde{Y},\bar{\boldsymbol{\kappa}})$ for some regions, e.g., Wuhan, Hubei, Romania and Germany, are relatively low, indicating that the inferred reporting delays

for these regions could be inaccurate. One possible reason is that the reporting delay distributions for some regions are not constant over time.

Figure 3.3 shows the inferred results for Spain. Compared with the reported data as shown in Fig. 3.1, after removing the reporting delays, peaks of the three curves coinciding with one another and the loop patterns are turned in good correlations.

3.4. ACCOUNTING FOR REPORTING DELAYS IMPROVES EPIDEMIC FORECASTS

Our statistical framework allows one not only to uncover the reporting delays in the epidemiological data but also allows to improve the accuracy of epidemic forecasts. To demonstrate its utility in epidemic forecasting we design the following two experiments. Experiment 1 aims to forecast the epidemic data in the testing set without accounting for reporting delays, and serves as a basis for Experiment 2, which uncovers reporting delays prior to forecasting epidemic data. In both experiments we split the reported epidemic data in two parts, which we call the training and the testing sets, respectively, see Fig. 3.4a.

In Experiment 1, we fit the training set with the SIRD epidemic model, obtaining model parameters β, $\gamma_r + \gamma_d$ and $I[0]$, where $I[0]$ denotes the percentage of initial infected cases. We then use the SIRD model with the obtained parameters to forecast the epidemic data, and compare it with that in the testing set.

In Experiment 2, we first use the reported data $\Delta\widetilde{I}$ in the training set to infer the parameters of the delay distributions $\bar{\boldsymbol{\kappa}}$. We rely on the inferred delay parameters $\bar{\boldsymbol{\kappa}}$ and to strip the data in the testing set of the delays. We then use these parameters in (3.2) to uncover the original epidemic data, which we call $\Delta\bar{I}$. At the next step, we fit the $\Delta\bar{I}$ data with the SIRD model, obtaining β, $\gamma_r + \gamma_d$ and $I[0]$ parameters. We rely on the obtained SIRD parameters to forecast the epidemic data $\Delta\hat{I}$. The last step is to compare $\Delta\hat{I}$ with the reported data in the testing set. The caveat here is that $\Delta\widetilde{I}$ in the testing set contains the reporting delays while $\Delta\hat{I}$ does not. Therefore, our last step is to add the inferred reporting delays to the $\Delta\hat{I}$ by (3.2) so that the forecast results in Experiment 2 can be fairly compared with the forecast results in Experiment 1. See 3.5.3 for technical details.

Figure 3.4 summarizes the results of epidemic forecasts for the 8 considered regions. To compare the results of the two experiments we start with the epidemic data of Spain. We plot the original reported data $\Delta\widetilde{I}$ as well as the forecasts $\Delta\hat{I}$ produced by experiments 1 and 2 for two different starting dates, Figs. 3.4(B,C), respectively. As seen from both plots, the forecasts accounting for reporting delays, experiment 2, are more accurate than the baseline forecasts in experiment 1.

We quantify the accuracy of forecasts in both experiments using the root mean square error $\mathrm{RMSE}(X,Y) = \sqrt{\frac{1}{n}\sum_{k=0}^{n-1}(X[k]-Y[k])^2}$, where n is the length of the pre-

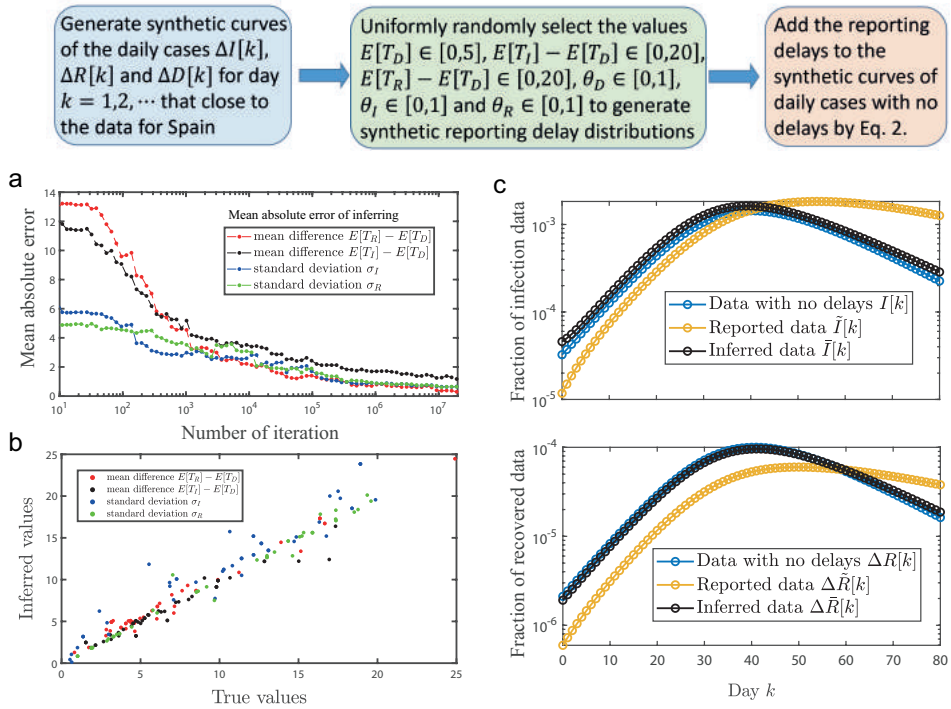


Figure 3.2: Uncovering reporting delays in synthetic data. The schematics illustrates the process to generate the synthetic reporting delay distributions and the synthetic daily reported cases. We generate synthetic data that close to the real data for Spain. We further add different reporting delays to the synthetic curves and try to infer these delays of death, infections and recoveries by our inference method. The synthetic reporting delays we considered hold that $E[T_D]$ is smaller than $E[T_I]$ and $E[T_R]$ for the reason that the diagnosis of deceased cases is easier than the diagnosis of the infected and recovered cases and the real data also indicates that the delays for deaths are shorter [30, 116]. a, The mean absolute inference errors as the function of the number of iteration steps in the search. Note that all inference errors decrease as a function of the number of iteration stops reaching their minimal values after 10^7 iterations. Panel b reveal the relationship between the true values and the inferred values. Panel c further indicates that, compared with the reported curves, the inferred curves are significantly closer to the data with no reporting delays (blue curves).

Regions	$E[T_R] - E[T_D]$	$E[T_I] - E[T_D]$	σ_I	σ_R	$\rho(\Delta\bar{R}, \Delta\bar{D})$	$\rho(\bar{I}, \Delta\bar{R})$	$\rho(\bar{I}, \Delta\bar{D})$	$\mathcal{O}(\tilde{Y}, \tilde{\boldsymbol{\kappa}})$
Italy	28.52	6.59	5.08	27.79	0.91	0.94	0.99	0.84
Spain	22.66	6.36	8.63	28.39	0.99	0.99	1.00	0.98
Wuhan	14.80	20.64	51.67	16.13	0.78	0.89	0.91	0.63
Turkey	14.13	24.85	43.89	12.47	0.91	0.95	0.98	0.85
Hubei	9.96	21.07	78.62	10.89	0.85	0.90	0.92	0.71
Romania	3.30	19.05	43.08	90.12	0.81	0.89	0.97	0.70
Germany	2.72	16.55	39.25	106.89	0.87	0.88	0.99	0.76
Denmark	0.17	25.07	36.90	2.71	0.83	0.93	0.94	0.72

Table 3.2: Inferred delay distributions for infections, recoveries and deaths. This table show the inferred mean differences of delay distributions $E[T_R] - E[T_D]$, $E[T_I] - E[T_D]$, standard deviations of delay distributions σ_I and σ_R and the final Pearson correlation coefficients. It shows that the reporting delays of infections and recoveries are longer than the reporting delays of deaths in all these regions. The inferred mean differences and standard deviations are significantly different for different regions.

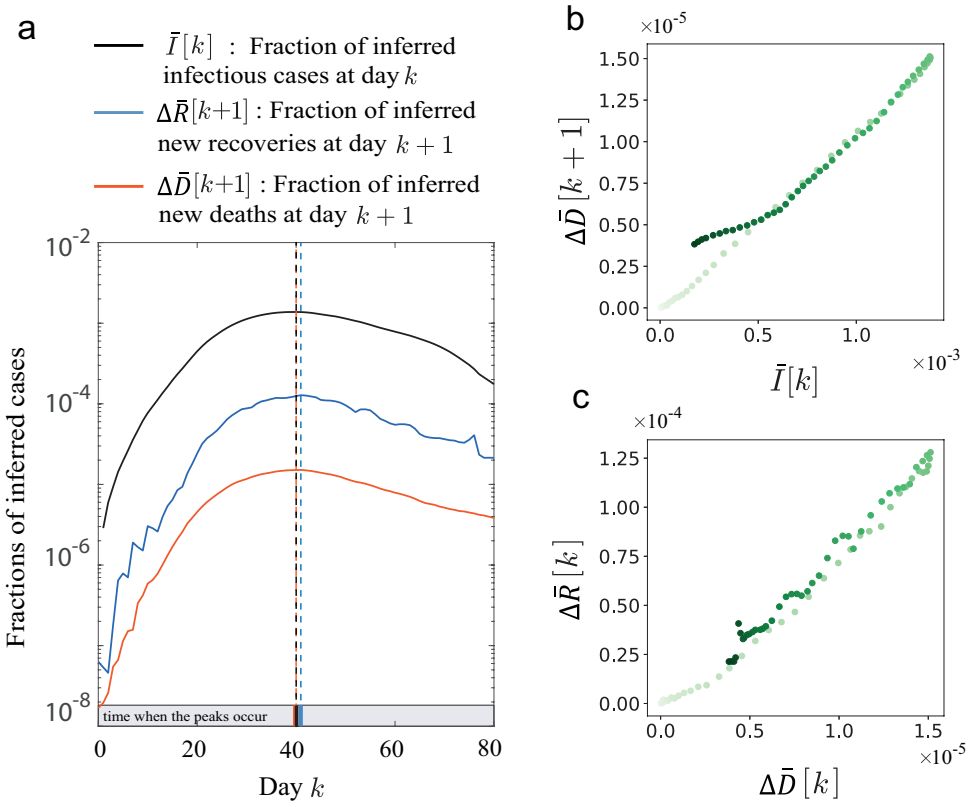


Figure 3.3: Inferred delay data for Spain. Panel a shows the time series of the revised fractions $\bar{I}[k]$, $\Delta\bar{R}[k+1]$ and $\Delta\bar{D}[k+1]$ for Spain. Panels b and c are data points between two of the three revised fractions. Loop patterns are revised to be roughly straight lines.

diction time window, Figs. 3.4(D) and 3.4(E). By comparing the average ratios of RMSE errors, $RMSE2/RMSE1$ in Fig. 3.4(E), we find the benefits of accounting for reporting delays vary depending on the region of interest. While nearly two-fold improvement in the forecast accuracy is observed for Denmark, there is almost no forecast improvement for Chinese provinces of Wuhan and Hubei.

We apply the above experiments on real data for the 8 regions or countries mentioned in Table 3.2. We apply the inferred reporting delays for infections in Experiment 2 to better forecast the daily infection data $\Delta\tilde{I}$. Figure 3.4(D) shows that, compared with the Experiment 1 (considering the reporting delays), the negligence of the reporting delays (Experiment 2) results in significantly larger forecast errors for Spain. However, not all forecast results are as good as Fig. 3.4(D). For example, as shown in Fig. 3.4(E), the improvement of forecast accuracy for Germany data is not significant. The forecast results for all 8 regions are shown in Figs.B.12-B.14. It indicates that the improvements of the forecast accuracy for Spain, Italy, Turkey and Denmark are more significant than the other regions. Panel h shows the mean of the ratio RMSE2/RMSE1 for all 8 countries. The forecast accuracy for some countries, e.g., Spain, can improve by around 40% by considering the effect of reporting delays. The improvement of forecast accuracy for some countries, e.g., Germany, however, is not significant.

There could be three reasons to explain the performance difference among regions: first, the SIRD model is too basic and thus cannot well describe the epidemic processes in some regions; second, the reporting delays for some regions are longer than the others; third, the inferred reporting delays in some regions are significantly time-varying. To support our claims, we show Fig. B.15 and Fig. B.16. It shows that the better the SIRD model fit, the better is the forecast, see Fig. B.15d. Moreover, there are some correlations between the mean reporting delays $E[T_I]$ and the ratio $RMSE2/RMSE1$, see Fig. S16b, indicating that the bigger is the reporting delays, the worse we do, with the exception of Denmark. Besides, we also observe that the closer the objective function $\mathcal{O}(\tilde{Y}, \bar{\boldsymbol{\kappa}})$ to 1, the smaller is the forecast error, see Fig. 3.4(G).

3.5. MATERIALS AND METHODS

3.5.1. TYPES OF DISTRIBUTIONS

We assume that probability distributions for infections, recoveries and deaths are the same functional. To determine the family of reporting delay distributions that best suit our data, we consider three different two-parameter discrete distributions [114] below:
(I) *Negative binomial distribution.* The probabilities that a deceased, infected or recovered

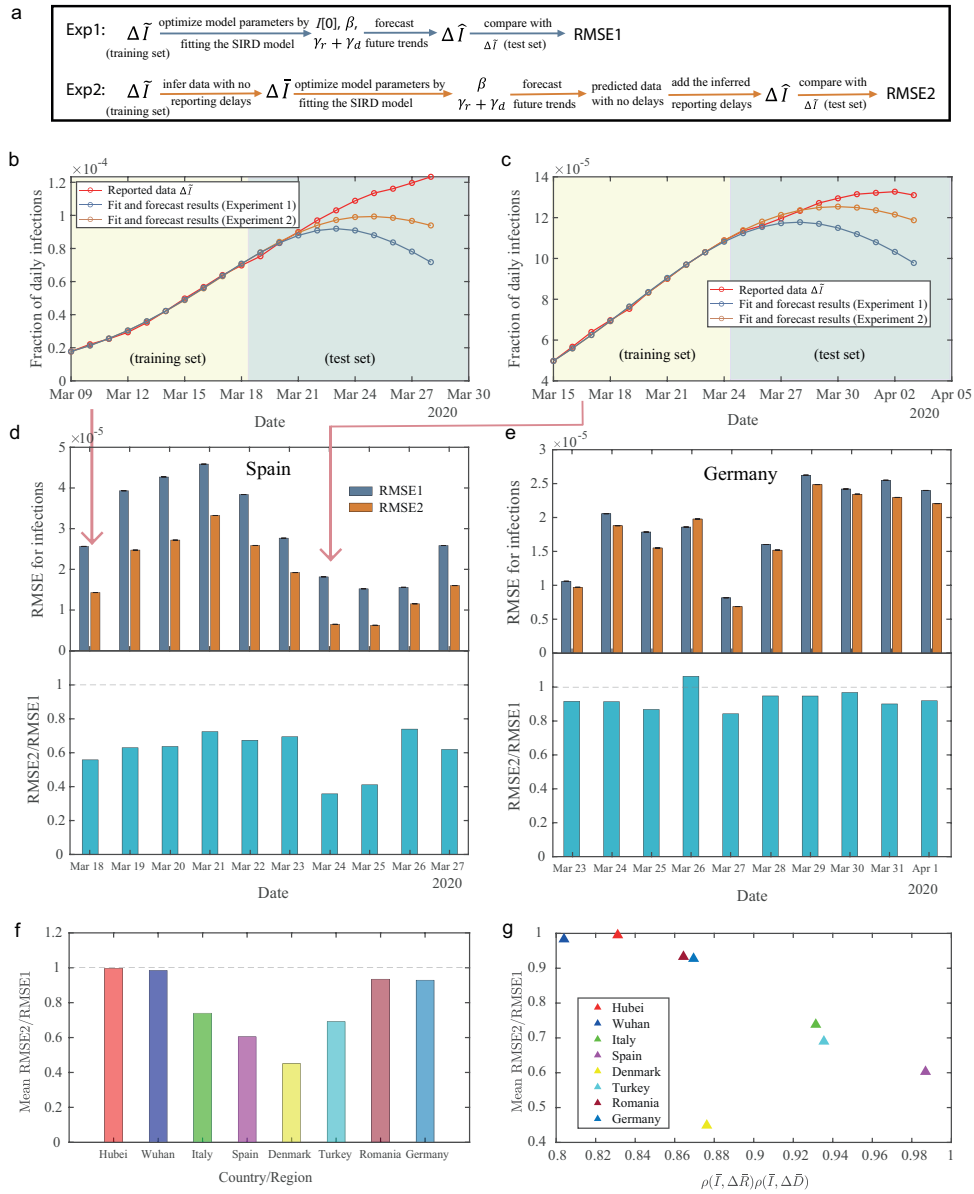


Figure 3.4: Forecast the COVID-19 pandemic for 8 regions by the SIRD model considering and ignoring the reporting delays for a forecast horizon of 10 days.

individual is reported after $m \in \mathbb{N}$ days are

$$\Pr[T = m] = \binom{m+r-1}{m}(1-p)^m p^r. \tag{3.4}$$

The negative binomial distribution with parameters $r > 0$ and $p \in [0,1]$ has mean value $E[T] = r(1-p)/p$ and variance $Var[T] = r(1-p)/p^2$.

(II) *Pólya-Aeppli distribution (also called the geometric Poisson distribution).* The probabilities that a deceased, infected or recovered individual is reported after $m \in \mathbb{N}$ days are

$$\Pr[T = m] = \begin{cases} \sum_{j=1}^{m} e^{-\lambda}\frac{\lambda^j}{j!}(1-\theta)^{m-j}\theta^j\binom{m-1}{j-1}, & m > 0 \\ e^{-\lambda}, & m = 0 \end{cases}. \tag{3.5}$$

The Pólya-Aeppli distribution with parameters $\lambda > 0$ and $\theta \in [0,1]$ has mean value $E[T] = \lambda/\theta$ and variance $Var[T] = \lambda(2-\theta)/\theta^2$.

(III) *Neyman type A distribution.* The probabilities that a deceased, infected or recovered individual is reported after $m \in \mathbb{N}$ days are,

$$\Pr[T = m] = \frac{\mu^m e^{-\xi}}{m!}\sum_{j=0}^{\infty}\frac{(\xi e^{-\mu})^j}{j!}j^m. \tag{3.6}$$

The Neyman type A distribution with parameters $\xi > 0$ and $\mu > 0$ has mean value $E[T] = \xi\mu$ and variance $Var[T] = \xi\mu(1+\mu)$.

3.5.2. MECHANISM TO INFER THE REPORTING DELAYS

We aim to infer the reporting delay information based on the reported data. (3.2) can be rewritten in the matrix form as

$$\Delta\widetilde{Y} = \Psi\Delta Y, \tag{3.7}$$

where the matrix $\Psi \in [0,1]^{n\times n}$ are element-wise defined by

$$\Psi_{i,j} \triangleq \begin{cases} \Pr[T = i-j] & \text{if} \quad i \geq j, \\ 0 & \text{otherwise.} \end{cases},$$

the column vectors for reported data and real data are

$$\Delta\widetilde{Y} = \left(\Delta\widetilde{Y}[0], \cdots, \Delta\widetilde{Y}[n]\right),\ \Delta Y = \left(\Delta Y[0], \cdots, \Delta Y[n]\right).$$

The data with no reporting delays can be deduced given the reported data and the reporting delay distributions by

$$\Delta Y = \Psi^{-1}\Delta\widetilde{Y}. \tag{3.8}$$

If the reporting delay distributions are unknown, we can guess a reporting delay matrix $\bar{\Psi}$ and obtain an inferred data by

$$\Delta\bar{Y} = \bar{\Psi}^{-1}\Delta\widetilde{Y}.$$

The inferred data $\Delta\bar{Y}$ are expected to be close to the original data ΔY if the guessed delay matrix $\bar{\Psi}$ is close to the true matrix Ψ. We infer the reporting delay information by maximizing the following objective function:

$$\begin{aligned} &\arg\max_{\boldsymbol{x}} \quad \mathscr{O}(\widetilde{Y}, \bar{\boldsymbol{\kappa}}) \equiv \rho(\Delta\bar{R}, \Delta\bar{D})\rho(\bar{I}, \Delta\bar{R})\rho(\bar{I}, \Delta\bar{D}) \\ &\text{s.t.} \qquad \min(\Delta\bar{I}[k], \Delta\bar{R}[k], \Delta\bar{D}[k], \bar{I}[k]) \geq 0, \qquad \text{for} \quad k = 0, 1, \cdots \end{aligned}$$

where the 6×1 vector $\boldsymbol{\kappa}$ denotes the parameters for delay distributions of infections, recoveries and deaths. In the random search to optimize the objective function, the mean reporting delays $E[T_I] = \lambda_I/\theta_I$, $E[T_R] = \lambda_R/\theta_R$ and $E[T_D] = \lambda_D/\theta_D$ are uniformly randomly chosen in range $[0, 30]$. The parameters θ_D, θ_R and θ_I are uniformly randomly chosen in range $[0, 1]$. Then the parameters λ_D, λ_R and λ_I are also determined.

3.5.3. FORECAST THE FUTURE PREVALENCE CONSIDERING THE REPORTING DELAYS

If we neglect the reporting delays, one can fit the prevalence data directly by the epidemic models, e.g., the SIRD model, and extend the model curve to forecast the future trends. If we consider the reporting delays, we forecast the future trend by more steps. We first modify the reported data $\Delta\widetilde{Y}$ to the data with no delays ΔY given the inferred or real reporting delay distributions by (3.8). We further optimize the model parameters by fitting the data ΔY based on the epidemic model. Next, we extend the model curve to forecast the future prevalence. Finally, we convert the forecast data with no delays back to predicted data with delays $\Delta\hat{Y}$ by (3.7). We fit the data of previous 10 days (training set) and forecast the prevalence of future 10 days (test set). We apply the grid search [117] to optimize the parameters in the SIRD model. Specifically, we respectively consider 100 candidate values for the infection rate β, the curing rate γ_r and the ratio γ_d/γ_r ranging from 0.01 to 1. The loss function we applied is still the RMSE $\sqrt{\frac{1}{n}\sum_{k=0}^{n-1}(\Delta\hat{I}[k] - \Delta\tilde{I}[k])^2}$. Details of the forecast method can be found in the SI Appendix G.

3.6. CONCLUSION

In this chapter, we discover that there are loops and shifts in reported infection, recovery and death data, indicating that there are reporting delays in COVID-19 data and the delays for infection, recovery and death data are different. We propose a method to infer

the reporting delays. This chapter discovered for the first time that there are significant reporting delay in recovery data. Finally, It indicates that accounting for reporting delays improves epidemic forecasts.

4

TWO-POPULATION SIR MODEL AND STRATEGIES TO REDUCE MORTALITY IN PANDEMICS

Despite many studies on the transmission mechanism of the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), it remains still challenging to efficiently reduce mortality. In this chapter, we apply a two-population Susceptible-Infected-Removed (SIR) model to investigate the spread of the coronavirus SARS-CoV-2 when contacts between elderly and non-elderly individuals are reduced due to the high mortality risk of elderly people. We discover that the reduction of connections between two populations can delay the death curve but cannot well reduce the final mortality. We propose a merged SIR model, which advises elderly individuals to interact less with their non-elderly connections at the initial stage but interact more with their non-elderly relationships later, to reduce the final mortality. Finally, immunizing elderly hub individuals can also significantly decrease mortality.

4.1. INTRODUCTION

In many countries, the first wave of the Coronavirus disease 2019 (COVID-19) appeared in early 2020. In the summer of 2020, the spread of COVID-19 was significantly reduced due to strict restrictions [118] and weather effects [119]. At that time, the majority of the population and politicians were hoping for the end of the COVID-19 pandemic. At the beginning of autumn, students went back to school, which marked the beginning of the second wave. However, rising infections were not taken seriously because infections were mainly among the young population and no significant hospitalization and deaths were observed [120]. Simultaneously, the high decease rate and self-preservation have caused that many elderly individuals reduced their contact with young people [121]. In October 2020, the hospitalization rates in many countries started increasing and the second COVID-19 wave was born. At the beginning of 2021, more contagious mutations of the coronavirus marked the third wave in many countries [122]. Even though there are COVID-19 vaccines, the distribution in many countries is painfully slow. Moreover, SARS-CoV-2 viral mutations lead to uncertainty about the effectiveness of recent vaccines. The third wave might not be the last COVID-19 pandemic and efficient strategies to reduce mortality will remain on the agenda.

The Susceptible-Infected-Removed (SIR) model [96, 123] and its variations are commonly applied to describe the COVID-19 pandemic [124, 125, 2, 126, 127] and to forecast the number of infected and deceased cases in a population [7, 6, 128, 129, 130]. The ratio of deceased elderly cases to deceased non-elderly cases each day is expected to be constant over time in classic epidemic models but is time-varying in reality. Recent works start to consider the age-structured SIR model to describe the COVID-19 pandemic more realistically [131, 132, 133, 134, 135]. The age-structured SIR model divides the whole population into several age groups and the infection rates are age-dependent. Real data reveal that the elderly infected had a 30- to 100- fold higher risk of dying than younger individuals in many European countries [136]. Here, elderly and non-elderly individuals are respectively defined as individuals who are < 65 years old and ≥ 65 years old [136]. The elderly population accounts for a proportion of around 20% in many European countries [137]. Since the main difference in the COVID-19 pandemic is between elderly and non-elderly individuals, we construct a two-population SIR model [138] as follows:

1. There are two sub-populations: non-elderly and elderly individuals uniformly distributed over the social contact network. The virus spreading in a region is likely to start from non-elderly individuals because the virus can be carried into a community from other areas by commuters [139] and most commuters are non-elderly individuals.

2. There are four infection rates between and within non-elderly and elderly individ-

uals. We believe that the highest infection rate is among elderly people. Elderly individuals are advised to a kind of self-isolation to protect themselves [140]. Staying in relative isolation from non-elderly people could be feasible, but some strong connections among elderly individuals, e.g., couples and people in the same nursing home, cannot be cut off. Conversely, the ties among elderly individuals will be stronger when their connections with non-elderly individuals are significantly reduced. The second highest infection rate is among the non-elderly population. The inter-group infection rates are the smallest since elderly individuals are afraid of being infected by non-elderly individuals. The infection rates between non-elderly and elderly individuals are low, but not zero, as elderly people still depend on younger people one way or another.

This chapter first investigates the features of fatality curves in the two-population SIR model when the connection between two populations is reduced. It shows that non-elderly deceased cases are prone to occur at the initial stage and most elderly deceased patients appear more often at a later stage. The difference in infection probability between non-elderly and elderly individuals is significant when the inter-population infection rates are low and the infection rate among elderly individuals is slightly above the epidemic threshold. The final mortality, however, cannot be reduced by only limiting the connection between two populations. Moreover, reducing the infection rate among non-elderly individuals, e.g., closing schools, can also not efficiently reduce mortality. In this chapter, we propose a merged SIR model to reduce the final mortality significantly. There are two stages in the merged SIR model: in the first stage, the model is the same as the two-population SIR model of Magal *et al.* [138] and in the second stage, the merged SIR model reduces to the standard SIR model. The physical meaning of the merged SIR model is that elder people are advised to reduce their connections with non-elderly individuals at the beginning of the pandemic and interact more with non-elderly individuals later. The merged SIR model benefits the mortality reduction since many recovered non-elderly people can protect the susceptible elderly individuals.

Compartmental epidemic models assume that social contact networks are homogeneous with an infinite network size, but the actual network size is finite and the degree distributions of many real social networks follow a power law [141] with an exponent $\gamma \in [2,3]$. We thus simulate the two-population SIR model on scale-free networks with a realistic network size to investigate the effect of network topology on the reduction of mortality. By comparing the simulation results of the two-population SIR model for the scale-free network and the Erdős–Rényi random network [142], the epidemic spreading in the heterogeneous network is much faster due to the star (or super spreader) effect. Reducing the connections between elderly and non-elderly individuals cannot

decrease mortality in the compartmental model, but can reduce the mortality in the two-population SIR epidemic on complex networks. The merged SIR model is the best strategy to efficiently mitigate mortality. Finally, we illustrate that mortality can be efficiently reduced by only immunizing rare elderly hub individuals.

4.2. TWO-POPULATION SIR MODEL

The two-population SIR model was first proposed by Magal *et al.* [138]. Similar models, that incorporate the underlying contact graph, are the networked SIR model proposed by Youssef and Scoglio [143], that was later entirely generalized to GEMF in [144]. Our work here applies the two-population SIR model to a realistic scenario related to the COVID-19 pandemic, systematically analyzes the death-related curve features, explores the effect of restrictions on mortality reduction and proposes an improved model to reduce the final mortality.

Suppose that the elderly and non-elderly populations are well-mixed and large enough, then the fractions of susceptible individuals $S(t)$, infectious individuals $I(t)$ and removed (recovered or deceased) individuals $R(t)$ at time t are reasonably well modeled by the following well-known differential equations:

$$\begin{cases} \dfrac{dS(t)}{dt} = -\text{diag}(S(t))BI(t), \\ \dfrac{dI(t)}{dt} = \text{diag}(S(t))BI(t) - EI(t), \\ \dfrac{dR(t)}{dt} = EI(t), \end{cases} \tag{4.1}$$

where the vectors of fractions $S(t)$, $I(t)$ and $R(t)$ are respectively,

$$S(t) = \begin{pmatrix} S_n(t) \\ S_e(t) \end{pmatrix}, \quad I(t) = \begin{pmatrix} I_n(t) \\ I_e(t) \end{pmatrix}, \quad R(t) = \begin{pmatrix} R_n(t) \\ R_e(t) \end{pmatrix}, \tag{4.2}$$

and the matrices E (removed rates) and B (infection rates) are respectively

$$E = \begin{pmatrix} \delta_n & 0 \\ 0 & \delta_e \end{pmatrix}, \quad B = \begin{pmatrix} \beta_{nn} & \beta_{ne} \\ \beta_{en} & \beta_{ee} \end{pmatrix}, \tag{4.3}$$

where β_{ne} denotes the infection rate from elderly infectious individuals to non-elderly susceptible individuals, β_{en} denotes the infection rate from non-elderly infectious individuals to elderly susceptible individuals, β_{nn} denotes the infection rate among non-elderly individuals, β_{ee} denotes the infection rate among elderly individuals, δ_n denotes the removed rate for non-elderly infectious individuals and δ_e denotes the removed rate for elderly infectious individuals. To simplify, we let the infection rates between two

populations be equal, $\beta_{ne} = \beta_{en} = \epsilon\beta_{nn}$, and thus the matrix B can be rewritten as

$$B = \beta_{nn} \begin{pmatrix} 1 & \epsilon \\ \epsilon & \kappa \end{pmatrix}. \tag{4.4}$$

For the COVID-19 pandemic, it holds that $\epsilon \ll 1$ and $\kappa \geq 1$. Furthermore, we have that the non-elderly fractions $S_n(t) + I_n(t) + R_n(t) = N_n$ and the elderly fractions $S_e(t) + I_e(t) + R_e(t) = N_e$, where N_n denotes the fraction of non-elderly population and N_e denotes the fraction of elderly population. The two-population SIR model assumes that the total population is unchanged and thus $N_n + N_e = 1$. We denote the initial state by $v[0] = (S_n[0], I_n[0], R_n[0], S_e[0], I_e[0], R_e[0])$. A schematic depiction of the two-population SIR model is shown in Fig. 4.1. The infectious individuals will turn to be immune with a recovery rate (ξ_n for non-elderly individuals and ξ_e for elderly individuals) or deceased with a fatality rate (η_n for non-elderly individuals and η_e for elderly individuals). It holds that the removed rates $\delta_n = \eta_n + \xi_n$ and $\delta_e = \eta_e + \xi_e$. This chapter focuses on the fractions of *new* deceased non-elderly and elderly cases that are $\eta_n I_n(t)$ and $\eta_e I_e(t)$, respectively. We are also interested in the fractions of deceased non-elderly and elderly cases that are $D_n(t) = \eta_n R_n(t)/\delta_n$ and $D_e(t) = \eta_e R_e(t)/\delta_e$.

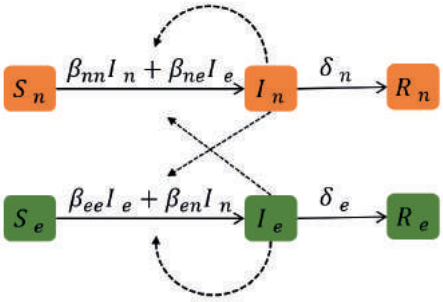


Figure 4.1: Schematic depiction of the two-population SIR model. There are two populations in the model: non-elderly individuals (highlighted in orange) and elderly individuals (highlighted in green). There are four different infection rates β_{nn}, β_{ne}, β_{en} and β_{ee} between and among the populations and two different removed rates δ_n and δ_e for each population.

By numerical solving Equations (4.1), we analyze the effect of infection rates on the following four death-related curve features,

1. maximum of $\eta_n I_n(t)$ and $\eta_e I_e(t)$: $\max_{t\geq 0} \eta_n I_n(t)$ and $\max_{t\geq 0} \eta_e I_e(t)$,
2. time points at which the maximum of $\eta_n I_n(t)$ and $\eta_e I_e(t)$ occur: $\operatorname{argmax}_{t\geq 0} \eta_n I_n(t)$ and $\operatorname{argmax}_{t\geq 0} \eta_e I_e(t)$,

3. time difference between two arguments of the maxima:
$\operatorname{argmax}_{t\geq 0} \eta_e I_e(t) - \operatorname{argmax}_{t\geq 0} \eta_n I_n(t)$,
4. fractions of final deceased non-elderly cases $D_n(\infty)$ and elderly cases $D_e(\infty)$.

In this chapter, we set the fraction of non-elderly individuals as $N_n = 0.8$, the fraction of elderly individuals as $N_e = 0.2$ and the removed rates as $\delta_n = \delta_e = 0.1$. The fatality rates for non-elderly and elderly infections are set to be $\eta_n = 0.0001$ and $\eta_e = 0.01$, respectively. The initial state is set as $\nu[0] = (0.7999, 0.0001, 0, 0.2, 0, 0)$. These parameters are set based on real data. The elderly population makes up around 20% of the whole population in many European countries [137]. Elderly people who were infected had 30- to 100- fold higher risk of dying than younger people in several European countries [136]. The time to recovery or death is on average around 10 days [145]. We also investigate various parameter settings and find that the changing of these parameters has little effect on the main conclusions drawn in this chapter.

There are three parameters in matrix (4.4), which are β_{nn}, ϵ and κ. We first set $\epsilon = 0.001$ and $\kappa = 4$ and study the effect of the infection rate β_{nn} on death-related curves. Figure 4.2 reveals that both the non-elderly related curves and elderly related curves are significantly affected by the parameter β_{nn}. The time difference $\operatorname{argmax}_{t\geq 0} \eta_e I_e(t) - \operatorname{argmax}_{t\geq 0} \eta_n I_n(t)$ is positive and increases with the infection rate β_{nn} deceasing. The final non-elderly deceased fraction $D_n(\infty)$ and elderly deceased fraction $D_e(\infty)$ increase with the infection rate β_{nn}.

We further set parameters $\beta_{nn} = 0.15$ and $\kappa = 4$ and study the effect of parameter ϵ on death-related curves. Figure 4.3 shows the death-related fractions with different parameter ϵ. It indicates that the parameter ϵ has almost no effect on non-elderly related curves. The final deceased fractions are little affected by the parameter ϵ. The effect of smaller ϵ is approximately to delay the elderly related curves and there will be larger time difference $\operatorname{argmax}_{t\geq 0} \eta_e I_e(t) - \operatorname{argmax}_{t\geq 0} \eta_n I_n(t)$ when ϵ is smaller.

We finally set parameters $\beta_{nn} = 0.15$ and $\epsilon = 0.001$ and study the effect of parameter κ on death-related curves. Figure 4.3 shows the death-related fractions with different parameters κ. The parameter κ has little effect on non-elderly curves but has large impact on elderly related curves. The time difference $\operatorname{argmax}_{t\geq 0} \eta_e I_e(t) - \operatorname{argmax}_{t\geq 0} \eta_n I_n(t)$ is the largest when the parameter $\kappa = 3.5$ in three considered parameters κ. The final elderly deceased fraction $D_e(\infty)$ increases as the parameter κ.

To better understand the effect of parameters β_{nn} and κ on death-related curves, we plot the heatmaps as shown in Fig. 4.5. It indicates that there are large time difference $\operatorname{argmax}_{t\geq 0} \eta_e I_e(t) - \operatorname{argmax}_{t\geq 0} \eta_n I_n(t)$ when the infection rate β_{ee} is around the epidemic threshold. Specifically, suppose that the infection rate between two populations $\beta_{en} \to 0$,

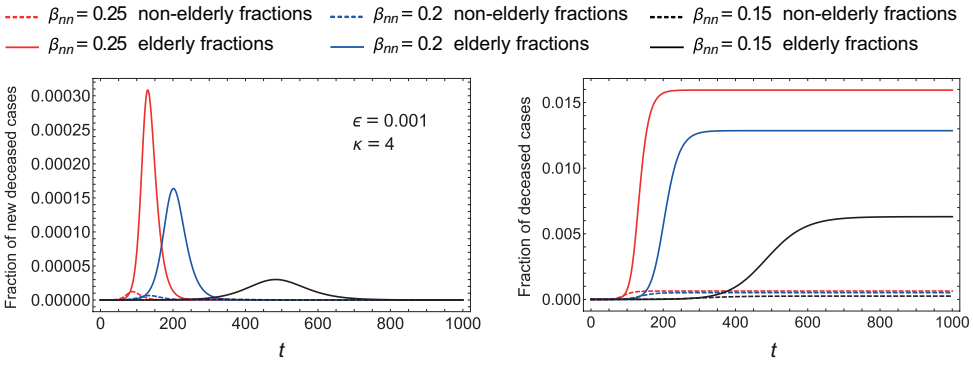


Figure 4.2: Death-related curves of the two-population SIR model with different infection rates β_{nn}. The left figure is about the fractions of new deceased cases $\eta_n I_n(t)$ for non-elderly individuals (dashed curves) and $\eta_e I_e(t)$ for elderly individuals (solid curves). The right figure is about the fractions of deceased cases $D_n(t)$ for non-elderly individuals (dashed curves) and $D_e(t)$ for elderly individuals (solid curves). Parameters ϵ and κ are set as 0.001 and 4, respectively. Three different infection rates β_{nn} are considered: $\beta_{nn} = 0.25$ (red curves), $\beta_{nn} = 0.2$ (blue curves) and $\beta_{nn} = 0.15$ (black curves).

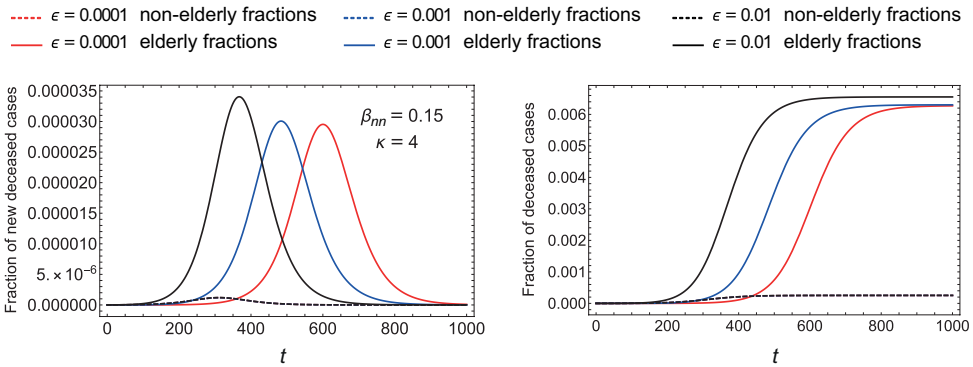


Figure 4.3: Death-related curves of the two-population SIR model with different parameters ϵ. Parameters β_{nn} and κ are set as 0.15 and 4, respectively. Three different parameters ϵ are considered: $\epsilon = 0.0001$ (red curves), $\epsilon = 0.001$ (blue curves) and $\epsilon = 0.01$ (black curves).

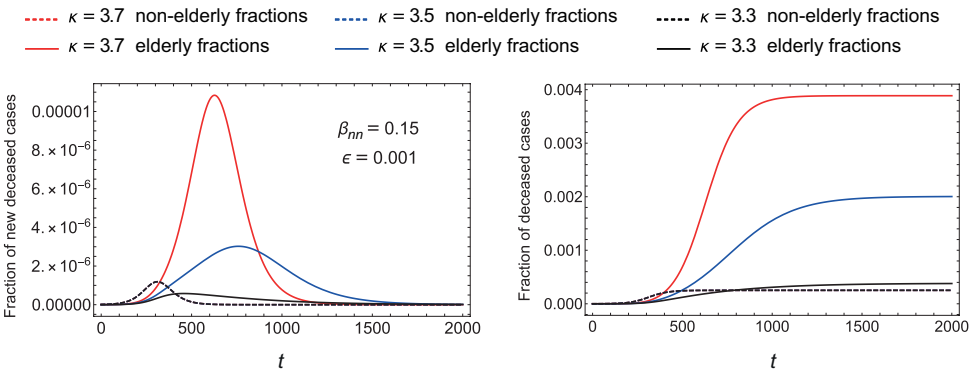


Figure 4.4: Death-related curves of the two-population SIR model with different parameters κ. Parameters β_{nn} and ϵ are set as 0.15 and 0.001, respectively. Three different parameters κ are considered: $\kappa = 3.7$ (red curves), $\kappa = 3.5$ (blue curves) and $\kappa = 3.3$ (black curves).

the epidemic threshold for elderly individuals is $\beta_{ee} = \delta_e / N_e$ (shown as the black curves in Fig. 4.5). The mortality cannot be significantly reduced by only reducing the infection rate among non-elderly individuals β_{nn}, e.g., closing schools, given that the infection rate β_{ee} is above the epidemic threshold. The only efficient way to well reduce the mortality in the two-population SIR model is to keep the infection rate β_{ee} among elderly individuals below the epidemic threshold.

In conclusion, we observe the following interesting curve properties: 1) the death-related curves for non-elderly individuals $\eta_n I_n(t)$ are mainly affected by the infection rate β_{nn}, 2) the time difference $\underset{t\geq 0}{\operatorname{argmax}} \eta_e I_e(t) - \underset{t\geq 0}{\operatorname{argmax}} \eta_n I_n(t)$ will be large if the inter-population infection rates β_{ne} and β_{en} are small and the infection rate β_{ee} is slightly above the epidemic threshold, 3) the fraction of eventually deceased cases $D_n(\infty) + D_e(\infty)$ will be small if the infection rate among elderly individuals $\beta_{ee} < \delta_e / N_e$, 4) only reducing the infection rates among non-elderly individuals cannot significantly reduce mortality.

The above observations are theoretically explained in Appendix C.1.

Although mortality can be reduced best by reducing the infection rates among elderly individuals β_{ee}, this strategy is not realistic since elderly people necessitate a sufficient amount of social interaction. This chapter discusses possible strategies to reduce mortality considering the social needs of all the people. Elderly people reduce their social connections with non-elderly individuals and increase their interactions with elderly relationships. Thus their interaction frequency [146], which is the total number of social interactions per unit time, is unchanged. We study the effect of reducing connections between elderly and non-elderly individuals on mortality reduction by comparing the mortality in the standard SIR model and the two-population SIR model. To keep the

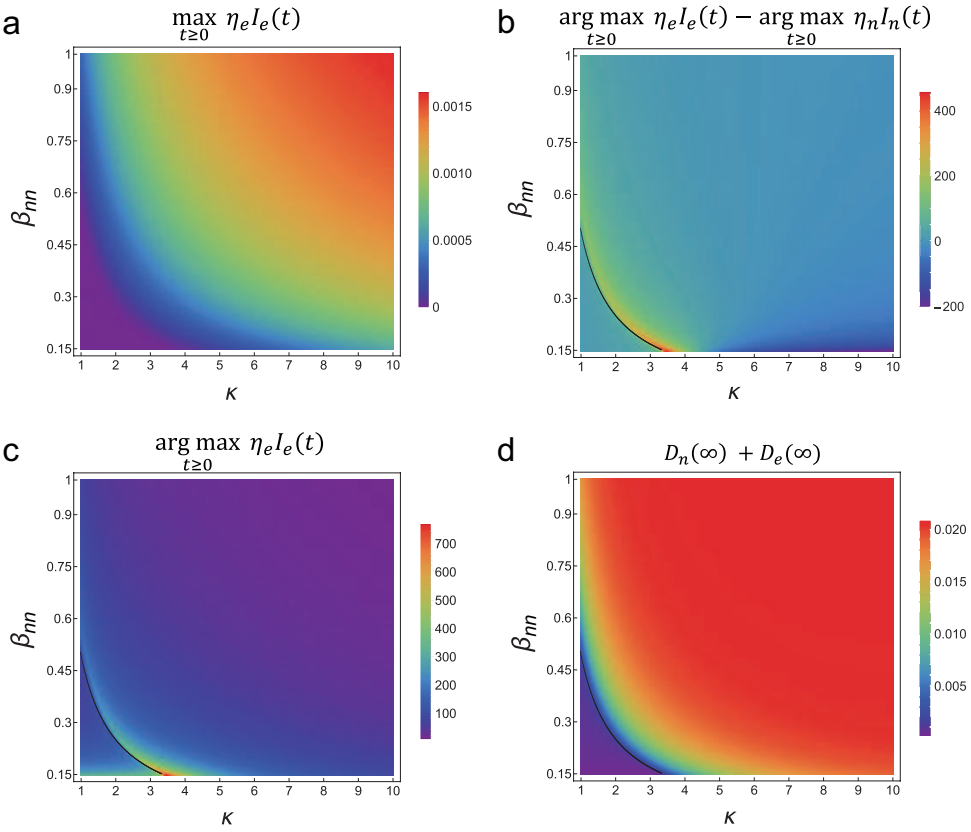


Figure 4.5: Curve features for the two-population SIR model with different parameters β_{nn} and κ. The parameter ϵ is set to be $\epsilon = 0.001$. The black curves show the parameters when the infection rate β_{ee} is at the epidemic threshold δ_e / N_e. The time difference $\underset{t\geq 0}{\text{argmax}}\, \eta_e I_e(t) - \underset{t\geq 0}{\text{argmax}}\, \eta_n I_n(t)$ will be large when the infection rate β_{ee} is slightly above the epidemic threshold. The fraction of total deceased individuals will be small when the the infection rate $\beta_{ee} < \delta_e / N_e$.

interaction frequency in the standard SIR model and the two-population SIR model to be at the same level, the equivalent infection rate in the standard SIR model is

$$\beta = \beta_{nn} N_n^2 + \beta_{ee} N_e^2 + (\beta_{ne} + \beta_{en}) N_n N_e. \tag{4.5}$$

It holds that $\beta = \beta_{nn} N_n = \beta_{ee} N_e$ when $\beta_{ne} \to 0$, $\beta_{en} \to 0$ and $\beta_{nn} N_n = \beta_{ee} N_e$. Figure 4.7a indicates that the fractions of the final deceased individuals for the standard SIR model and the two-population SIR model are the same. The effect of reducing the connection between elderly and non-elderly groups is only to delay the deceased curve, but not to effectively reduce mortality.

4.3. MERGED SIR MODEL TO REDUCE MORTALITY

To effectively reduce mortality, we propose a merged SIR model in which the epidemic spreading follows the two-population SIR model in the first stage and follows the standard SIR model in the second stage. The illustration of the merged SIR model is shown in Fig. 4.6. The reduction of the connection between two populations can delay the pandemic among elderly people. The reconnection of these two populations further protect elderly people due to the herd immunity effect of recovered non-elderly individuals. Figure 4.7 shows that the merged SIR model can significantly reduce the final deceased fractions and there is the best switch time point to minimize the final mortality. Heatmaps in Fig. 4.8 show the effect of parameters β_{nn} and ϵ on the best switch time point and reduced rate of the final mortality. The reduced rate of the final mortality is defined as

$$\frac{D_e(\infty) + D_n(\infty) - \widetilde{D}_e(\infty) - \widetilde{D}_n(\infty)}{D_e(\infty) + D_n(\infty)}, \tag{4.6}$$

where $D_e(\infty)$ and $D_n(\infty)$ are respectively the elderly and non-elderly mortality for the two-population SIR model and $\widetilde{D}_e(\infty)$ and $\widetilde{D}_n(\infty)$ are respectively the elderly and non-elderly mortality for the merged SIR model. Figure 4.8 reveals that the first stage (reducing the connection between non-elderly and elderly people) should take a longer time if parameters β_{nn} and ϵ are smaller. Besides, the final mortality can be reduced more significantly for smaller parameters β_{nn} and ϵ.

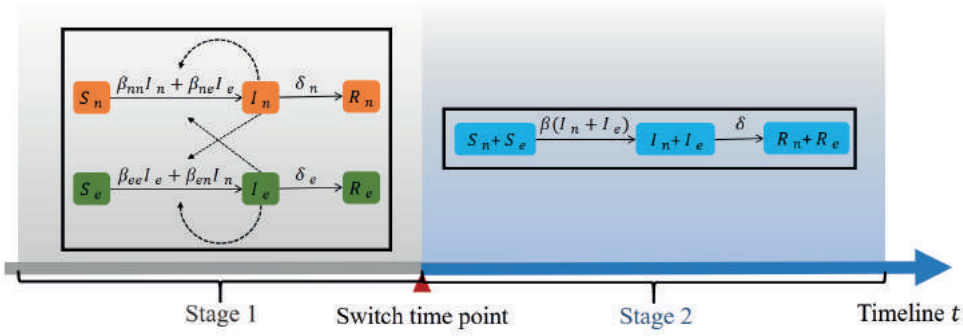


Figure 4.6: Schematic depiction of the merged SIR model. This model has two stages: the first stage follows the two-population SIR model and the second stage follows the standard SIR model. The physical meaning of this model is to reduce the connection between the elderly and non-elderly populations initially and reconnect these two populations after many non-elderly infected individuals have been recovered. There is a switch time point between these two stages.

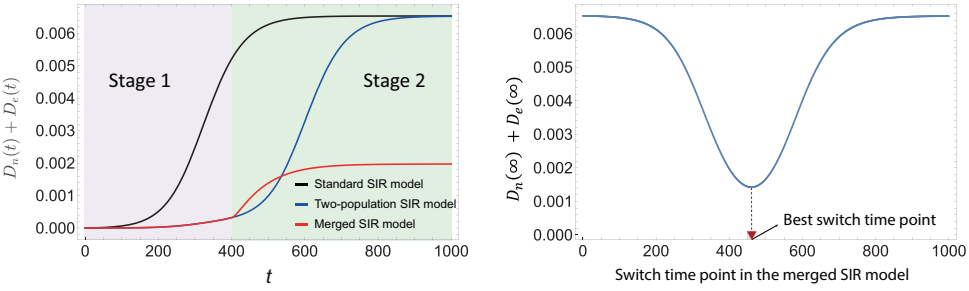


Figure 4.7: The fractions of deceased individuals in the standard SIR model, the two-population SIR model and the merged SIR model. The parameters for the two-population SIR model are $\beta_{nn} = 0.15$, $\epsilon = 0.0001$ and $\kappa = 4$. We set the infection rate $\beta = 0.12$ for the standard SIR model to keep the interaction frequency the same among models. The left figure reveals that the two-population SIR model cannot, but the merged SIR model can efficiently reduce mortality. The right figure shows the final deceased fractions with different switch time points, indicating that there is the best switch time point to minimize the final mortality.

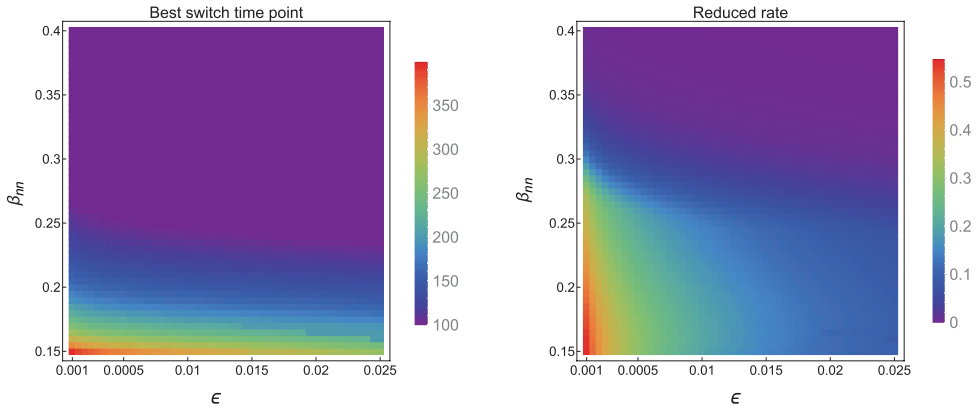


Figure 4.8: The best switch time point and reduced rate of the final mortality. We choose different parameters β_{nn} and ϵ and a fixed parameter $\kappa = 4$ for the two-population SIR model. The best switch time point will be larger and the final mortality will be reduced more significantly if parameters β_{nn} and ϵ are smaller.

4.4. TWO-POPULATION SIR EPIDEMIC ON LARGE COMPLEX NETWORKS

We apply the Monte Carlo method [147] to simulate the two-group SIR epidemic on complex networks. In this chapter, we consider large networks with network size $N = 10^5$ generated by the configuration model [148] and the simulation starts from 100 non-elderly infected individuals. We first compare the simulation results on the scale-free network and the Erdős–Rényi random network to analyze the effect of network heterogeneity on epidemic curves. The network size N and mean degree $E[D]$ of the Erdős–Rényi random network are the same as the scale-free network. Figure 4.9a and Fig. 4.9b indicate that the epidemic spreading in the scale-free network is much faster than the spreading in the Erdős–Rényi random network due to the super spreaders. Figure 4.9c and Fig. 4.9d illustrate that the epidemic spreads quicker when the mean degree $E[D]$ is higher.

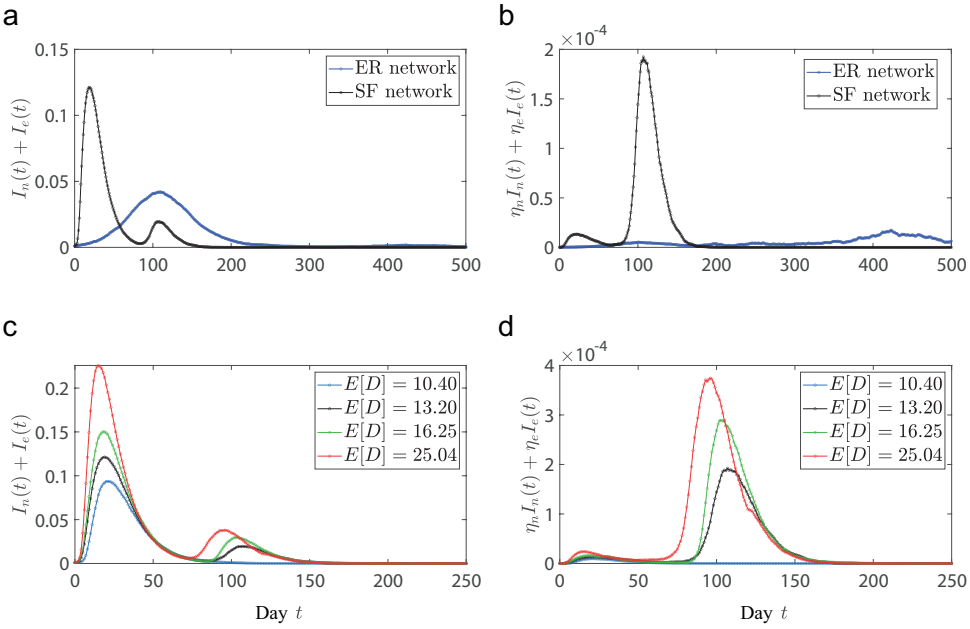


Figure 4.9: Fractions of infectious cases $I_n(t) + I_e(t)$ and daily deceased cases $\eta_n I_n(t) + \eta_e I_e(t)$ for the two-population SIR epidemic on the scale-free network and the Erdős–Rényi network with the network size $N = 10^5$. The infection parameters are set to be $\beta_{nn} = 0.015$, $\kappa = 4$ and $\epsilon = 0.001$. The spreading in the scale-free network is much faster than the spreading in the Erdős–Rényi network. Figures c and d show the effect of mean degree $E[D]$ of the scale-free network on the two-population SIR epidemic. With the increase of the scale-free networks' link density, there are more individuals infected and deceased. The exponent in the scale free networks is set as $\gamma = 2.5$. The minimum degree of the scale-free network is set to be 5.

We simulate the standard SIR model, the two-population SIR model and the merged SIR model on the scale-free network as shown in Fig. 4.10. Different from the results for the compartmental models as demonstrated in Fig. 4.7, for the epidemic spreading on complex networks, the final mortality for the two-population SIR model is lower than the standard SIR model since a part of susceptible elderly people can be protected by their recovered non-elderly relationships. This type of local immunity, which differs from the herd immunity, can only be observed in the epidemic spreading on networks. The merged SIR model is the best strategy to reduce mortality.

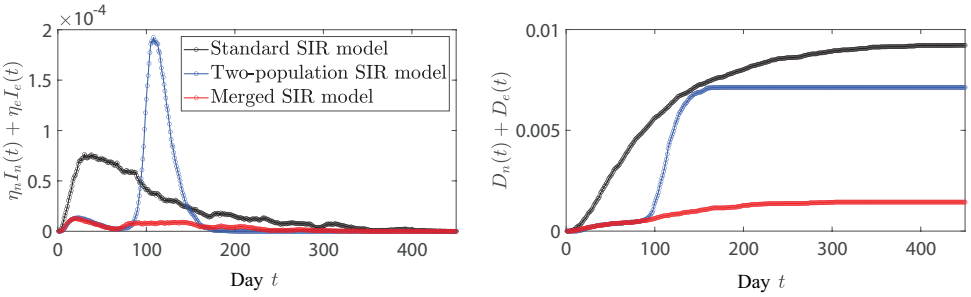


Figure 4.10: Fractions of daily deceased cases $\eta_n I_n(t)+\eta_e I_e(t)$ and deceased cases $D_n(t)+D_e(t)$ for the standard SIR epidemic, the two-population SIR epidemic and the merged SIR epidemic on the scale-free network with the network size $N = 10^5$. The infection rate in the standard SIR model is $\beta = 0.012$. The infection rates in the two-population SIR model are the same as Fig. 4.9. Different from the result in Fig. 4.7, the final deceased fraction for the two-population SIR model is lower than the standard SIR model. The final deceased fraction for the merged SIR model is the smallest, indicating that the merged SIR model is the best strategy to reduce the mortality.

Given that there have been COVID-19 vaccines but the vaccine is still insufficient, it is valuable to study the strategy to reduce mortality by immunizing specific population. There are rare elderly hub individuals in social networks, e.g., the priests, which are the virus's primary route of transmission from non-elderly to elderly people. Figure 4.11a and Fig. 4.11b reveal that the final mortality can be significantly reduced by only immunize 20 elderly hub individuals in 10^5 population assuming that the vaccines are 100% effective. In reality, the COVID-19 vaccine efficacy cannot reach 100% and thus we analyze the situation when the vaccines are 80% effective. Figure 4.11c and Fig. 4.11d illustrate that more elderly hub individuals require to be immunized to reduce mortality efficiently.

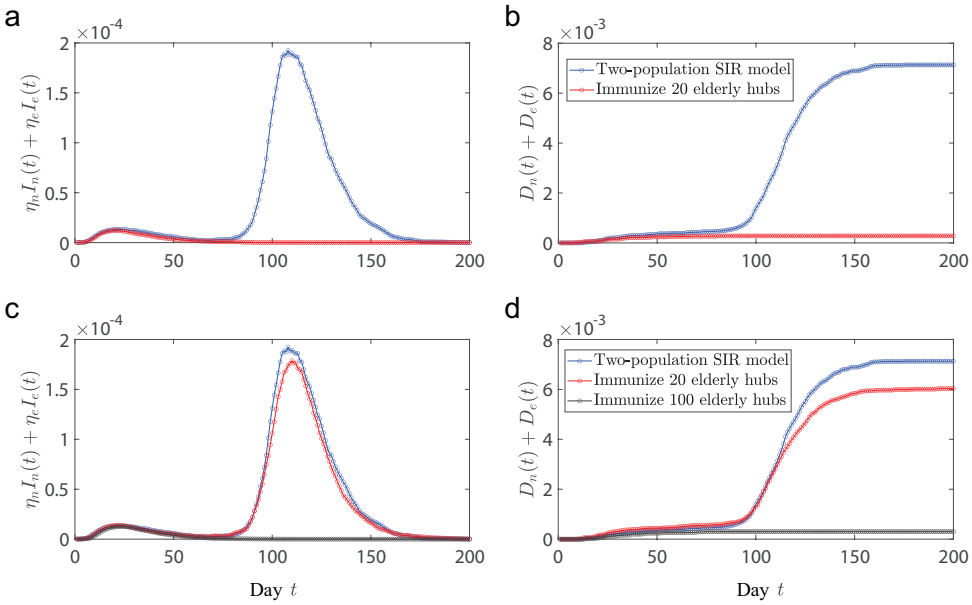


Figure 4.11: Effect of immunizing rare elderly hub individuals on reducing the final mortality. Figures a and b respectively show the fractions of daily deceased cases $\eta_n I_n(t) + \eta_e I_e(t)$ and deceased cases $D_n(t) + D_e(t)$ for the two-population SIR epidemic with and without immunizing elderly hub individuals. We immunize 20 elderly individuals with the largest degree in the simulation of the two-population SIR model on the scale-free network with 10^5 population assuming that the vaccines are 100% effective. The mortality can be significantly reduced by immunizing such rare specific individuals. Figures c and d show the fractions when the vaccines are 80% effective. It requires more vaccine doses to effectively reduce the mortality if the vaccines are less effective.

4.5. CONCLUSION

Since early 2020, scientists have found that COVID-19 is substantially more dangerous for the elderly. Elderly people's interactions with their non-elderly relationships are reduced to lower the risk of being infected. This chapter applies the two-population SIR model to describe the COVID-19 pandemic when the connections between elderly and non-elderly individuals are significantly reduced. We analyze how the reduction of connections between two populations can affect the COVID-19 pandemic, especially the mortality. It reveals that severing ties between two populations can postpone the pandemic but not effectively cut mortality. We further find that reconnecting two populations at an appropriate time can significantly lessen the final mortality. Assuming that rare vaccines are available, this study recommends immunizing elderly hub individuals first to better decrease mortality.

5

APPROXIMATION OF SIS EPIDEMICS ON NETWORKS

Markovian Susceptible-Infected-Susceptible (SIS) epidemics on a network with N nodes can be described by a continuous-time Markov chain, that consists of 2^N network infection states. The Markovian SIS epidemics on complete graphs can be exactly described by the birth-and-death process. The birth-and-death approximation for random networks, e.g., Erdős–Rényi (ER) random networks, however, leads to errors. The approximation errors for some networks are significantly larger than other networks and thus the SIS epidemics on these networks require a higher order of approximation to ensure a relatively high accuracy. To lessen the birth-and-death approximation error and study how many reduced states are required to make sure the approximation error smaller than a threshold based on the network properties, we propose a spectral clustering SIS approximation (SCSA) method, which combines the spectral clustering of the huge Markov graph and the birth-and-death approximation, to reduces the huge 2^N state space of the Markov chain to a smaller number of states. We discover that the relationship between the approximation error ϵ and the number of clusters r roughly obeys $\epsilon \sim r^{-\alpha}$, where $\alpha \in (\frac{1}{4}, \frac{4}{5})$. The exponent α tends to be larger if the network has a higher link density. The approximation errors are usually larger for networks with larger network size N. For Watts-Strogatz (WS) small-world networks, the approximation error ϵ deceases faster with the number of clusters r when the rewiring probability p is higher.

5.1. INTRODUCTION

The Susceptible-Infected-Susceptible (SIS) model [2] describes the spread of an infectious disease on a network G. Each of the N nodes in the network is either susceptible (S) or infectious (I) at any time t. If the infection and curing processes are independent Poisson processes, then the SIS model can be represented as a continuous-time Markov chain with 2^N network infection states [149, 150, 31]. A network infection state represents which nodes in the networks are infected and which nodes are susceptible. Any random graph with more than $N > 20$ nodes renders an exact analysis infeasible due to the huge number of network infection states.

Researchers have attempted to apply exact reduction methods on the huge Markov chain based on the automorphisms of the underlying graph [150, 151, 152, 153, 154, 155]. Unfortunately, few networks can be reduced significantly, e.g. the complete graph, the star graph and the household graph [156, 150]. For other graphs, e.g., Erdős–Rényi (ER) random networks, the number of automorphisms is very small compared to the size of the network, which makes the exact method intractable. As an alternative, various mean-field methods have been developed with a much lower complexity than the original Markov chain with 2^N states. A prominent example is the N-Intertwined mean field approximation (NIMFA) which approximates the SIS prevalence by assuming that the states of any two neighboring nodes are statistically independent [149, 157]. The NIMFA method can be extended to consider dynamical correlations between pairs of adjacent nodes [158]. Alternatively, the Heterogeneous Mean-Field (HMF) approximation was proposed by Pastor-Satorras and Vespignani [159], which assumes that all nodes with the same degree are statistically equivalent. Like the individual-based mean-field approaches, the degree-based mean-field theory was also extended to higher orders such as the second-order pairwise approximation [160, 161, 162, 163]. The first-order mean-field methods are less accurate for sparse networks [144], small networks [164] and spreading parameters around the epidemic threshold [165]. Moreover, all mean-field approaches fail to describe the eventual convergence to the absorbing state (all-healthy state). A completely different approach is the birth-and-death approximation [166], which approximates the 2^N-sized Markov chain with a birth-and-death process with $N+1$ states. Each state in the birth-and-death process represents a certain number of infected nodes. The transition rates in the birth-and-death approximation are estimated based on the original Markov chain [166]. The advantage of the birth-and-death approximation is that the absorbing state is reserved and the prevalence after the metastable state is well approximated.

The birth-and-death process can exactly describe the Markovian SIS epidemics on complete graphs, but leads to errors in the approximation of the SIS epidemics on random graphs. As what we have mentioned in the above paragraph, in the birth-and-death

approximation, all network infection states with the same number of infected nodes are grouped into one reduced state. But this degree of approximations may lead to significant errors for some random networks and thus require a kind of higher order approximation method to achieve a higher accuracy.

In the continuous-time SIS Markov processes, the state transition rate diagram with 2^N network infection states and transition rates between any two states can be seen as a huge graph with directed weighted links and 2^N nodes. Spectral clustering is one computationally efficient method based on the spectral decomposition of the infinitesimal generator matrix to approximately partition the graph into clusters so that the links between clusters have low weights and the links within clusters have high weights.

In this chapter, we propose the spectral clustering SIS approximation (SCSA) that combines the spectral clustering with the birth-and-death approximation. Specifically, we partition the 2^N network infection states into r clusters by the spectral clustering of the infinitesimal generator matrix (also known as the transition rate matrix) Q. Different from the birth-and-death approximation in which the network infection states with the same number of infected nodes are grouped into one state, the SCSA groups the network infection states with the same number of infected nodes and the same cluster into one state. The number of clusters r can be any integer value between 1 and 2^N. If the number of clusters $r = 1$, then the approximation reduces to the birth-and-death approximation [166]. For $r > 1$, the resulting method more accurately approximates the original SIS process, at the cost of more reduced states. If $r = 2^N$, we recover the original SIS process. We measure the approximation error between the approximated prevalence and the exact prevalence for different numbers of clusters r. We show that the relationship between the approximation error ϵ and the number of clusters r roughly obeys $\epsilon \sim r^{-\alpha}$, where $\alpha \in (\frac{1}{4}, \frac{4}{5})$. The exponent α tends to be larger if the network has higher link density. Besides, for Watts-Strogatz (WS) small-world networks, the approximation error decreases faster for networks with larger rewiring probability p.

This chapter is structured as follows. We compare the prevalence of the SIS process with the birth-and-death approximation and N-Intertwined mean field approximation (NIMFA) in Section 5.2. Although the birth-and-death approximation has performance advantages for the prevalence after the metastable state, there are still large errors compared with the exact SIS prevalence for some random networks. Section 5.3 introduces the spectral clustering on the infinitesimal generator matrix that could be used to group the states. In Section 5.4, we propose the spectral clustering SIS approximation (SCSA), which combined the spectral clustering information with the birth-and-death approximation, to better approximate the prevalence of the SIS network epidemic. We apply the SCSA in Section 5.5 to SIS epidemics on ER random networks [142] and WS small-world networks [76]. We further analyze the relationship between the approximation error ϵ and the

number of clusters r. The impact of the network size N, the link density, the rewiring probability p of the WS small-world networks and the effective infection rate τ are also analyzed. Finally, we conclude in Section 5.6.

5.2. PREVALENCE CURVES OF THE SIS PROCESS, THE BIRTH-AND-DEATH APPROXIMATION AND THE NIMFA MODEL

The Susceptible-Infected-Susceptible (SIS) network epidemic model describes the spread of an infectious disease on a network with N nodes and L links. We use the adjacency matrix A with elements a_{ij} to indicate if there is a connection between nodes i and j ($a_{ij} = 1$) or not ($a_{ij} = 0$). Each node in the network at time t can be either in state 0 (susceptible) or state 1 (infected). The SIS process consists of two independent processes. Infected nodes can recover from the disease, which is modelled as a Poisson process with a curing rate δ. Simultaneously, infected nodes may infect their susceptible neighbours with an infection rate β. The effective infection rate τ is defined as the ratio $\tau = \beta/\delta$ of the infection rate β to the curing rate δ. We set the curing rate $\delta = 1$ in this chapter.

At any time t, each node i in the network is either susceptible (0) or infected (1). We apply the Bernoulli random variable $X_i(t) \in \{0,1\}$ to denote the infection state of node i at time t: $X_i(t) = 0$ when node i is susceptible at time t, otherwise $X_i(t) = 1$. We represent the current infection state of all nodes in the network as an $N \times 1$ vector x, where each element $x_k \in \{0,1\}$. Using the representation from Van Mieghem [31], we represent the state vector x by an N-digits binary number $(x_N x_{N-1} \cdots x_2 x_1)$ and the corresponding decimal value is

$$i = \sum_{k=1}^{N} x_k(i) 2^{k-1}. \tag{5.1}$$

The transition rate q_{ij} specifies the rate to go from state i to state j, and can be computed as [149, 32, 167]

$$q_{ij} = \begin{cases} \delta & \text{if} \begin{cases} j = i - 2^{m-1}; m = 1,2,\cdots,N \\ \text{and } \ x_m(i) = 1 \end{cases} \\ \beta \sum_{k=1}^{N} a_{mk} x_k(i) & \text{if} \begin{cases} j = i + 2^{m-1}; m = 1,2,\cdots,N \\ \text{and } \ x_m(i) = 0 \end{cases} \\ -\sum_{k=1;k\neq j}^{2^N} q_{kj} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

The set of all transition rates q_{ij} forms the $2^N \times 2^N$ infinitesimal generator Q. The probability state vector at time t can be calculated by $s^\mathsf{T}(t) = s^\mathsf{T}(0)e^{Qt}$, where the i-th element

of the $N \times 1$ column vector $s(t)$ is the joint probability $s_i(t) = \Pr[X_1(t) = x_1(i), X_2(t) = x_2(i), \cdots, X_N(t) = x_N(i)]$. The prevalence, which denotes the average fraction of the infected nodes, at time t equals

$$y(t) = \frac{1}{N} s^{\mathsf{T}}(0) e^{Qt} \xi, \tag{5.3}$$

where ξ denotes the column vector of the number of infected nodes for each state. Specifically, the vector ξ can be calculated by $\xi = Mu$, where $u = (1, 1, \cdots, 1)$ is the all-one vector and the $2^N \times N$ matrix M includes all states in binary notation, but bit-revised:

$$M = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

Figure 5.1 shows the prevalence curves generated by the birth-and-death approximation ($r = 1$) and the exact SIS network epidemic ($r = 2^N$). We randomly generate the ER random networks and the WS small-world networks with the average degree $E[D] = 4$ and the network sizes $N = 6$, $N = 8$ or $N = 10$. The rewiring probability of the WS small-world networks is $p = 0.1$. The epidemic threshold $\tau_c^{(1)} = 1/\lambda_1$, where λ_1 denotes the largest eigenvalue of the network, derived from NIMFA [2], is the lower bound of the actual epidemic threshold τ_c. We considered four different effective infection rates: $\tau = \tau_c^{(1)}$, $\tau = 3\tau_c^{(1)}$, $\tau = 5\tau_c^{(1)}$ and $\tau = 10\tau_c^{(1)}$. Figure 5.1 indicates that the overall trends can be captured by the birth-and-death approximation ($r = 1$). However, the approximation errors are relatively large for some networks. To investigate the advantages of the birth-and-death approximation compared with the mean-field approximation, we analyse the prevalence curves of the NIMFA model, the exact SIS model and the birth-and-death approximation as shown in Fig. 5.2. The networks and parameters considered are the same as Fig. 5.1c and Fig. 5.1f. Figure 5.2 indicates that the mean squared error ϵ between the NIMFA prevalence and the exact prevalence are much larger than the error between the birth-and-death approximated prevalence and the exact prevalence. This result is consistent with previous works which show that mean-field approximations work poorly for sparse and/or small networks [168]. Moreover, mean-field approximations cannot capture the prevalence trend after the metastable state. The gap between the exact prevalence and the birth-and-death approximated prevalence can be further reduced by considering more clusters.

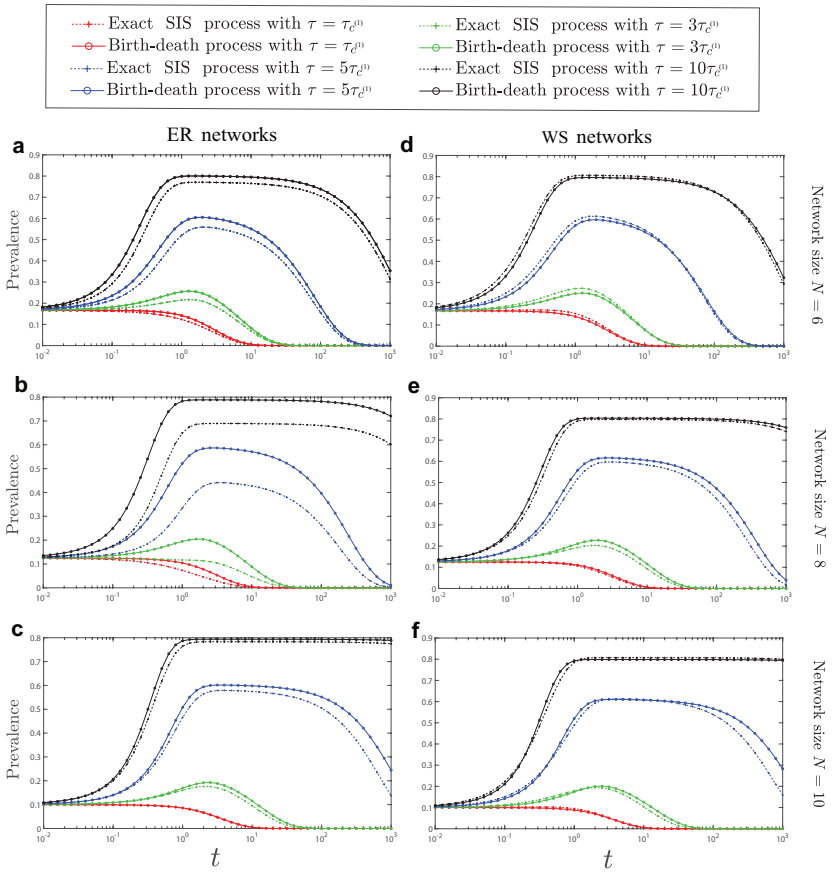


Figure 5.1: The comparison between the exact SIS prevalence and the birth-and-death approximated prevalence. We consider ER random networks and WS small-world networks with average degree $E[D] = 4$ and network sizes $N = 6$, $N = 8$ and $N = 10$. The rewiring probability for the WS small-world networks is $p = 0.1$. Four different effective infection rates are considered: $\tau = \tau_c^{(1)}$, $\tau = 3\tau_c^{(1)}$, $\tau = 5\tau_c^{(1)}$ and $\tau = 10\tau_c^{(1)}$. The curing rate equals $\delta = 1$. All prevalence curves in this paper have a horizontal logarithmic scale.

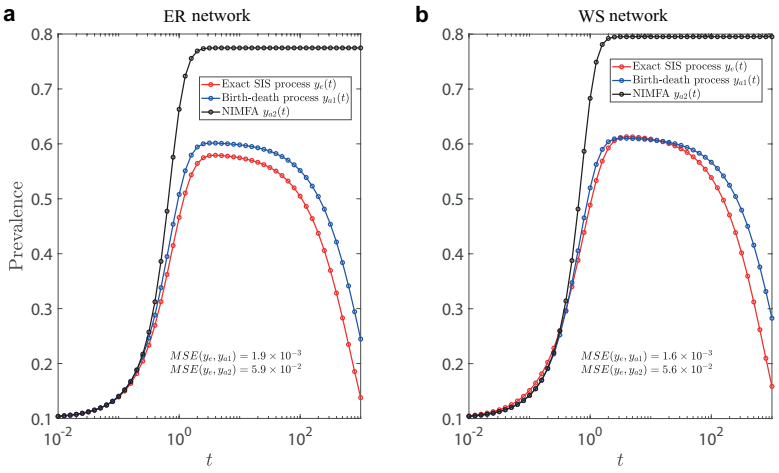


Figure 5.2: The prevalence of the exact SIS model, the birth-and-death approximation and the NIMFA model for the ER random network (left) and the WS small-world network (right) with the network size $N = 10$. The effective infection rate is $\tau = 5\tau_c^{(1)}$ and the curing rate equals $\delta = 1$.

5

5.3. SPECTRAL CLUSTERING ON THE INFINITESIMAL GENERATOR MATRIX

We aim for a dimensionality reduction on the $2^N \times 2^N$ infinitesimal generator matrix Q of the Markov chain for the SIS processs by spectral clustering. However, the matrix Q is, in general, asymmetric so that its eigenvalues and eigenvectors can be complex. A heuristic method to apply the spectral clustering on the asymmetric matrix Q is to construct a corresponding symmetric matrix $Q+Q^T$, but the clustering results may not be optimized since a set of different asymmetric matrices may associate with the same constructed symmetric matrix [169, 170, 171]. Recent research reveals that, for any irreducible stochastic matrix, the clustering behavior can still be captured by only considering the real part of the eigenvalues and eigenvectors if the real part of the eigenvalues is much greater than the imaginary part. To guarantee that the infinitesimal generator matrix Q is irreducible [31, p. 447], we assume that each node i can be infected spontaneously with a small self-infection rate $\varepsilon = 10^{-10}$. In Appendix D.1, we show that the eigenvalues of the infinitesimal generator Q have a dominant real part, which means that the imaginary part is usually too small and thus can be ignored. Based on the above analysis, in this paper, we only consider the real part of the eigenvalues and eigenvectors in the spectral clustering. The spectral-based clustering consists of two steps:

1. find the r eigenvectors corresponding to the r smallest eigenvalues (the real part)

of Laplacian matrix $-Q$;

2. partition the 2^N states into r clusters by the k-means algorithm [172].

Although k-means algorithm is not the only way to cluster the 2^N states in step 2, researches usually use k-means algorithm here mainly due to its advantage in computational efficiency.

Figure 5.3 illustrates the results of spectral clustering on the infinitesimal generator matrix Q for the SIS process on an example network with 5 nodes. In the state transition rate diagrams of Fig. 5.3, the network infection states in the same column have the same number of infected nodes. The network infection states that belong to different clusters are marked with different colors. With the number of clusters $r = 1$, there is only one cluster and thus all network infection states are with the same color. With the number of clusters $r = 2$, in this example, only the all healthy state (absorbing state) in one and all others in another cluster. With the number of clusters $r = 3$, the clustering results turn to be complicated, but we can still see some relation between clusters and the underlying network. For example, the network infection states 3 and 5 are in the same cluster. In the network infection state 3, the nodes 1 and 2 in the network are infected. In the network infection state 5, the nodes 1 and 3 are infected. These two network infection states are symmetric from the view of the underlying network.

5

5.4. SPECTRAL CLUSTERING SIS APPROXIMATION

The states in the same cluster and with the same number of infected nodes are aggregated into one state. Figure 5.4 shows an example of the partition into $r = 3$ clusters. The states in the m-th partition all have ξ_m infected nodes and are all in the C_m-th cluster. We denote the set of states in the m-th partition as S_m. We further denote the number of partitions as $\widetilde{N}$ and the upper bound of the partition number[1] $\widetilde{N}$ is $(N-1)r+2$. The partitioned infinitesimal generator matrix $\widetilde{Q}$ is

$$\widetilde{Q}_{m,n} = \begin{cases} \frac{1}{|S_m|}\sum_{i\in S_m}\sum_{j\in S_n} Q_{ij} & \text{if} \quad |\xi_m - \xi_n| = 1, \quad S_m \neq \emptyset \quad \text{and} \quad S_n \neq \emptyset \\ -\sum_{z=1;z\neq j}^{\widetilde{N}} \widetilde{Q}_{kj} & \text{if} \quad m = n \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

The probability partition vector is $\tilde{s}(0) = \left(\sum_{i\in S_1} s_i(0), \sum_{i\in S_2} s_i(0), \cdots, \sum_{i\in S_{\widetilde{N}}} s_i(0)\right)$ at the initial time $t = 0$ and the vector of the number of the infected nodes for the partitions

[1] For a network with N nodes, the possible number of infected nodes is any integer ranging from 0 to N. There are $N+1$ columns in the state transition rate diagram as shown in Fig. 5.3 if we let each column show the states with the same number of infected nodes. Apart from the all healthy state 0 and the all-infected state 2^N-1, the states in each other column can be split into a maximum of r partitions. Thus the maximum number of partitions is $(N-1)r+2$.

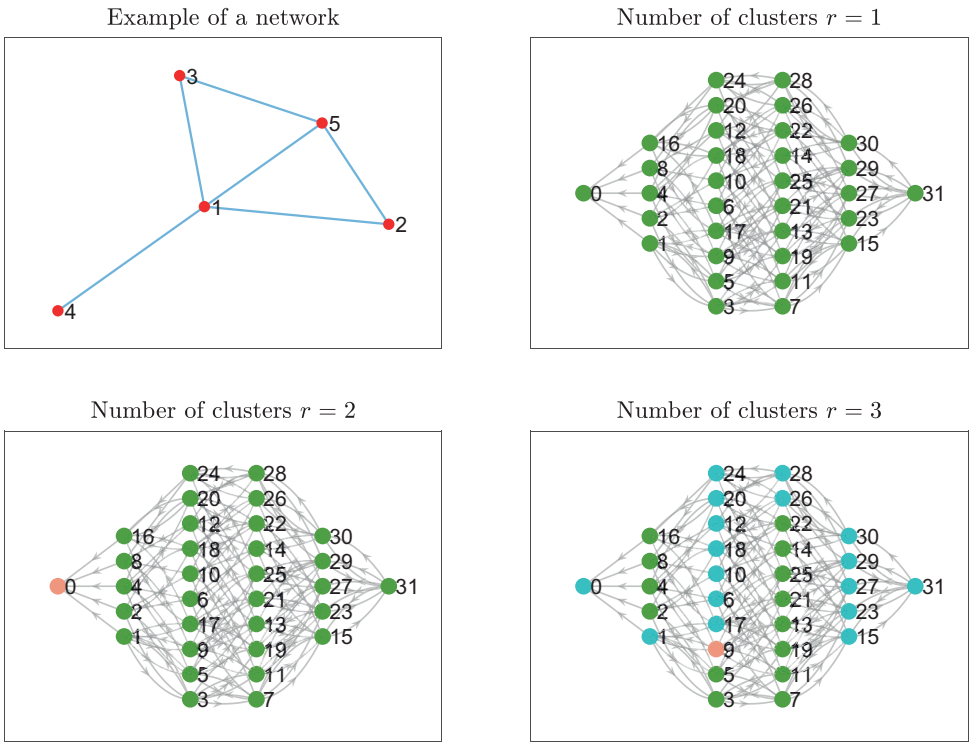


Figure 5.3: Example of the spectral clustering on the infinitesimal generator matrix Q for the SIS process on a network with 5 nodes. Each state can be represented by the N-digits binary number $x_5x_4x_3x_2x_1$ and the corresponding integer value follows from Eq. (5.1). The binary digit x_i denotes whether node i is susceptible ($x_i = 0$) or infected ($x_i = 1$). For example, state 0, which can also be represented as state 00000, indicates that all 5 nodes are susceptible. The effective infection rate is $\tau = 5\tau_c^{(1)}$, where $\tau_c^{(1)}$ is the epidemic threshold derived by NIMFA. The curing rate equals $\delta = 1$.

is $\tilde{\xi} = (\xi_1, \xi_2, \cdots, \xi_{\tilde{N}})$. The probability partition vector at time t is calculated by $\tilde{s}^{\mathsf{T}}(0)e^{\tilde{Q}t}$. The m-th element of the vector $\tilde{s}(t)$ denotes the probability that the state is in the m-th partition at time t. The approximated prevalence at the time t equals

$$\tilde{y}(t) = \frac{1}{N}\tilde{s}^{\mathsf{T}}(0)e^{\tilde{Q}t}\tilde{\xi}. \tag{5.5}$$

By adjusting the number of clusters, one can tune the results from the birth-and-death approximated prevalence ($r = 1$) to the exact SIS prevalence ($r = 2^N$). The pseudocode of the approach is shown in Appendix D.2. Compared with the birth-and-death approximation, the spectral clustering method with the number of clusters $r > 1$ reduces the error ϵ, but also increases the size of the partitioned infinitesimal generator $\tilde{Q}$.

There are an immense amount of possible way and the spectral clustering is just one possible way to partition the 2^N network infection states. The performance of the spectral clustering in the approximation of SIS network epidemics compared with the other clustering results is still unknown. To access the quality of SCSA, we compare the SCSA results with a random clustering method, which places each state in a uniformly random cluster. As shown in Fig. D.3 of Appendix D.4, with the same size of the partitioned state space, the approximate accuracy of SCSA is significantly higher than the random clustering benchmark. Fig. D.4 of Appendix D.4 further indicates that the SCSA result is close to the best approximation accuracy one can obtained if we keep the size of the reduced state space to be the same.

Another question is the approximation performance of the spectral clustering without considering the birth-and-death restriction. In other words, why do we combine the spectral clustering with the birth-and-death approximation rather than directly group the network infection states in the same cluster? To answer this question, we show the SIS approximation results without the birth-and-death restriction in Appendix D.3. It shows that the prevalence before and after the metastable state will be significantly different from the exact prevalence if we do not consider the birth-and-death restriction.

5.5. APPROXIMATION ERROR

We measure the approximation error between the approximated prevalence $\tilde{y}(t)$ and the exact prevalence $y(t)$ at time t using the Mean Squared Error $\epsilon(t)$. We separate the time interval from $t = 10^{-2}$ to $t = 10^3$ into $N_t = 51$ logarithmically spaced time points, because the main dynamics of the SIS process takes place over multiple orders to magnitude [173]. The Mean Squared Error ϵ is then computed as

$$\epsilon = \frac{1}{N_t}\sum_{t=0}^{N_t-1}\big(y(t) - \tilde{y}(t)\big)^2 \tag{5.6}$$

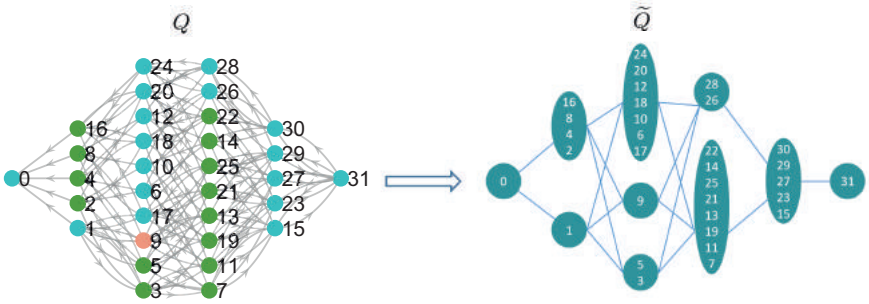


Figure 5.4: Diagram of the spectral clustering SIS approximation (SCSA). The first step of SCSA is clustering the infinitesimal generator matrix Q and the second step is to get the partitions by considering the number of infections. The left graph shows the infinitesimal generator matrix Q and the right graph shows the partitioned infinitesimal generator matrix $\widetilde{Q}$. The states in the same cluster and with the same number of infections are partitioned into one state.

The spectral clustering approximation method (SCSA) is evaluated by simulations of the spectral clustering approximation. We compare the approximation results with the exact SIS epidemic on *connected* ER random networks and WS small-world networks for small network sizes N, due to the 2^N computing complexity of the exact Markov chain.

We study how fast the approximation error decreases by increasing the number of clusters. Figure 5.5 and 5.6 depict the mean squared error ϵ between the approximated prevalence and the exact prevalence for the ER random networks and the WS small-world networks with different link densities. Relatively large errors may occur when applying the birth-and-death approximation ($r = 1$). The errors can be efficiently reduced by considering more clusters. The insets in Fig. 5.5 and 5.6 indicate that the relationship between the error ϵ and and the number of clusters r roughly obeys $\epsilon \sim r^{-\alpha}$. The figures also indicate that the approximation errors are larger for sparse networks. Similar results can be discovered for networks with other link densities, which are shown in Appendix D.5. We also notice that the approximation error for $r = 2$ is generally similar to the approximation error for $r = 1$. Usually, the clustering algorithm with $r = 2$ takes the all-healthy state 0 in one cluster and the other states in the other cluster, which is also depicted in Figure 5.3. Thus the partitions for the number of clusters $r = 2$ and the partitions for the number of clusters $r = 1$ are the same, because we also assemble states with a different number of infected nodes in a different group.

In practice, the spectral clustering approximation will be of little significance if plenty of clusters are considered, because a large amount of clusters implies a large state space. Thus we focus on the approximation performance for the number of clusters $1 \leq r \leq N$. We analyze the performance of the spectral clustering approximation method for different

link densities. Figure 5.7 and 5.8 show the error ϵ between the approximated prevalence and the exact prevalence with different network sizes and indicate that the approximation error will be larger for the network with larger network size N. We further analyze the approximation performance for the underlying networks with different rewiring probability p for the WS small-world networks. The rewiring probability p is an important parameter which varies the WS small-world network between a regular network ($p = 0$) and the network close to the ER random network ($p = 1$). The larger the rewiring probability p, the more random the network will be. The results in Figure 5.9 reveal that, for the WS small-world networks with lower rewiring probability p, the approximation errors tend to be larger for the birth-and-death approximation ($r = 1$) but reduce faster with the increase of the number of clusters r.

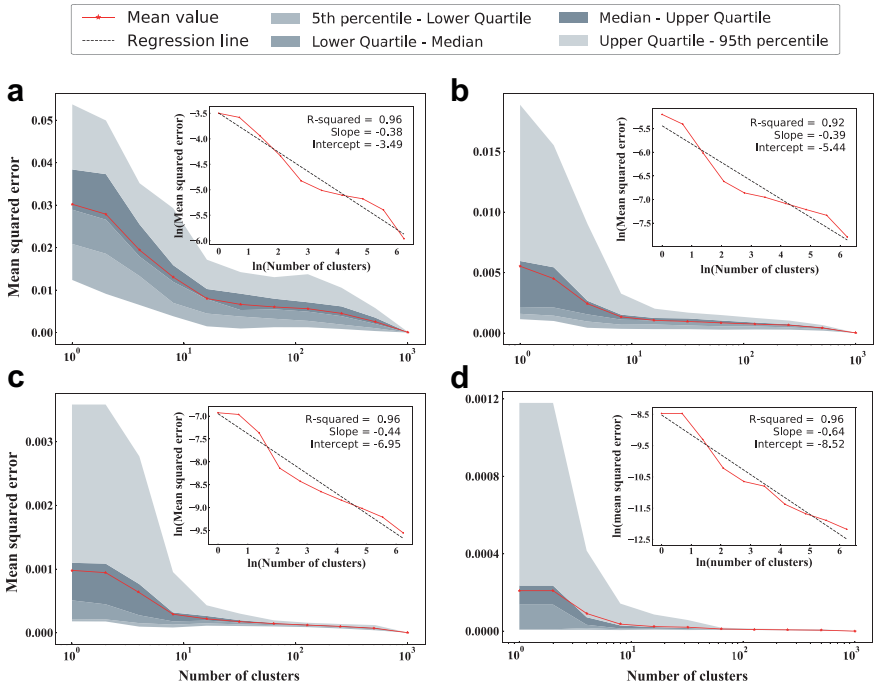


Figure 5.5: Mean squared error ϵ between the approximated prevalence and the exact prevalence for the ER random networks with network size $N = 10$ and different number of links: $L = 9$ (a), $L = 19$ (b), $L = 29$ (c), $L = 39$ (d). Note that all plots have a horizontal logarithmic scale. The mean and percentiles of the error are obtained by considering 1000 randomly generated networks. The effective infection rate $\tau = 5\tau_c^{(1)}$. The curing rate equals $\delta = 1$. The mean and percentiles of error are obtained by considering 100 randomly generated networks.

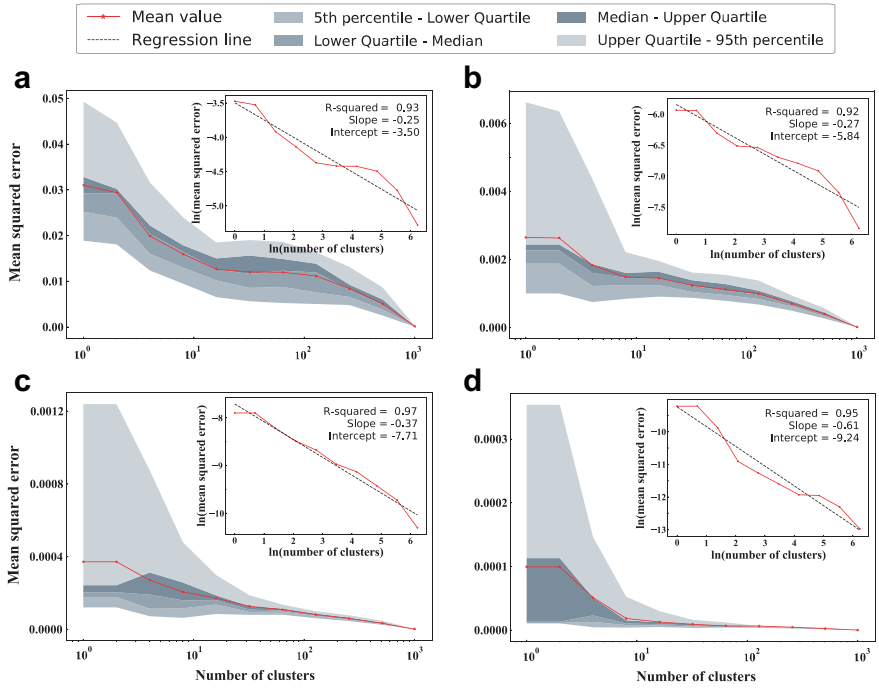


Figure 5.6: Mean squared error ϵ between the approximated prevalence and the exact prevalence for the WS small-world networks with network size $N = 10$ and different average degree: $E[D] = 2$ (a), $E[D] = 4$ (b), $E[D] = 6$ (c), $E[D] = 8$ (d). Note that all plots have a horizontal logarithmic scale. The rewiring probability for the WS small-world networks is $p = 0.1$. The effective infection rate is $\tau = 5\tau_c^{(1)}$. The curing rate equals $\delta = 1$. The mean and percentiles of the error are obtained by considering 1000 randomly generated networks.

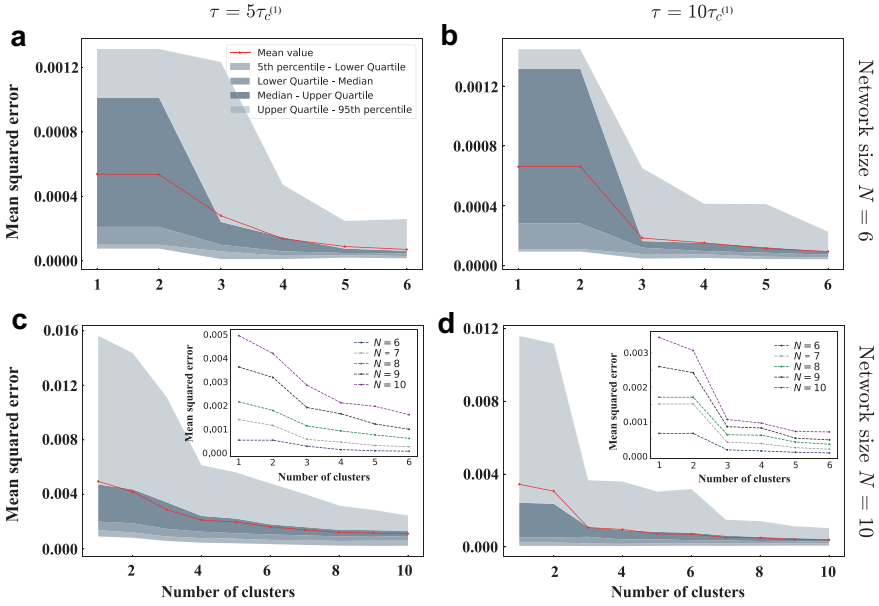


Figure 5.7: Mean squared error ϵ between the approximated prevalence and the exact prevalence for the ER random networks with different network sizes. The average degree is $E[D] = 4$. The mean and percentiles of the error are obtained by considering 1000 randomly generated networks. The curing rate equals $\delta = 1$.

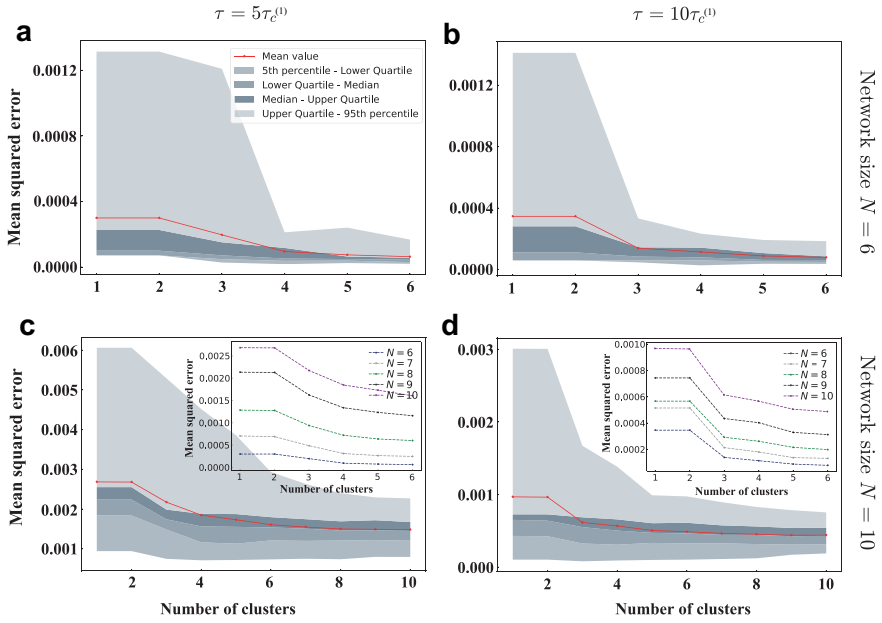


Figure 5.8: Mean squared error ϵ between the approximated prevalence and the exact prevalence for the WS small-world networks with different network sizes. The average degree is $E[D] = 4$. The rewiring probability for the WS small-world networks is $p = 0.1$. The mean and percentiles of the error are obtained by considering 1000 randomly generated networks. The curing rate equals $\delta = 1$.

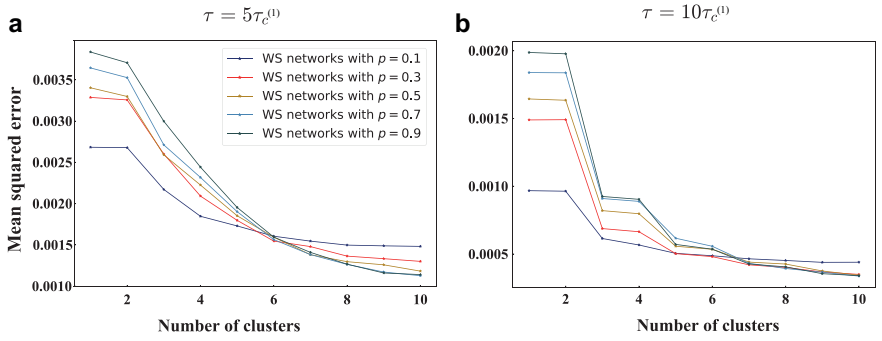


Figure 5.9: The approximation performance of SCSA for the WS small-world networks with different rewiring probability p. The average degree is $E[D] = 4$. The mean error ϵ is obtained by considering 1000 randomly generated networks. The curing rate equals $\delta = 1$.

5.6. CONCLUSION

This chapter proposes the SCSA method to approximate the SIS network epidemic by combining the spectral clustering and the birth-and-death approximation. We discover that the relationship between the approximation error ϵ and the number r of clusters roughly obeys $\epsilon \sim r^{-\alpha}$, where $\alpha \in (\frac{1}{4}, \frac{4}{5})$. The exponent α tends to be larger if the network has higher link density. Besides, for WS small-world networks, the approximation error decreases faster for networks with larger rewiring probability p. These rules can be applied to decide how many reduced states are required to make sure the approximation error is smaller than a threshold for the SIS spreading on a specific network. Using SCSA, the effect of more network properties on the approximation accuracy can be studied in future works. The spectral clustering approximation method can also be applied to other epidemiological models, like the Susceptible-Infected-Removed (SIR) model. Then the infinitesimal generator will contain 3^N states, which is even more challenging than the SIS model.

6

ERROR ACCUMULATION IN QUANTUM CIRCUITS

We study a classical model for the accumulation of errors in multi-qubit quantum computations. By modeling the error process in a quantum computation using two coupled Markov chains, we are able to capture a weak form of time-dependency between errors in the past and future. By subsequently using techniques from the field of discrete probability theory, we calculate the probability that error quantities such as the fidelity and trace distance exceed a threshold analytically. The formulae cover fairly generic error distributions, cover multi-qubit scenarios and are applicable to e.g. the randomized benchmarking protocol. To combat the numerical challenge that may occur when evaluating our expressions, we additionally provide an analytical bound on the error probabilities that is of lower numerical complexity. Besides this, we study a model describing continuous errors accumulating in a single qubit. Finally, taking inspiration from the field of operations research, we illustrate how our expressions can be used to decide how many gates one can apply before too many errors accumulate with high probability and how one can lower the rate of error accumulation in existing circuits through simulated annealing.

6.1. INTRODUCTION

We study Markov chains that provide a model for the accumulation of errors in quantum circuits. Different types of errors [174] that can occur and are included in our model are e.g. Pauli channels [3], Clifford channels [175, 176], depolarizing channels [3] and small rotational errors [177, 178]. If the random occurrence of such errors only depends on the last state of the quantum mechanical system, then the probability that error quantities such as the fidelity and trace distance accumulate beyond a threshold can be related to different hitting time distributions of two coupled Markov chains [179]. These hitting time distributions are then calculated analytically using techniques from probability theory and operations research.

Error accumulation models that share similarities with the Markov chains under consideration here can primarily be found in the literature on randomized benchmarking [180]. From the modeling point of view, the dynamical description of error accumulation that we adopt is shared in [181, 182, 183, 184]. These articles however do not explicitly tie the statistics of error accumulation to a hitting time analysis of a coupled Markov chain. Furthermore, while Markovianity assumptions on noise are common [185], the explicit mention of an underlying random walk is restricted to a few papers only [182, 186]. From the analysis point of view, research on randomized benchmarking has predominantly focused on generalizing expressions for the expected fidelity over time. For example, the expected decay rates of the fidelity are analyzed for cases of randomized benchmarking with restricted gate sets [187], Gaussian noise with time-correlations [188], gate-dependent noise [184] and leakage errors [189]; and the expected loss rate of a protocol related to randomized benchmarking is calculated in [190, 181, 191, 184, 192]. In this chapter, we focus instead on the probability distributions of both the error and maximum error in the Markov chain model – which capture the statistics in more detail than an expectation – for arbitrary distance measures and in random as well as nonrandom quantum circuits. Finally, [181, 182, 184, 189] resort to perturbation or approximate analyses (via e.g. Taylor expansions and independence or decorrelation assumptions) to characterize the fidelity, whereas here we provide the exact, closed-form expressions for the distributions using the theory of Markov chains.

To be precise: this chapter first studies a model for discrete Markovian error accumulation in a multi-qubit quantum circuit. We suppose for simplicity that both the quantum gates and errors belong to a finite unitary group $\mathcal{G}_n \subseteq \mathcal{U}(2^n)$, where $\mathcal{U}(2^n)$ is the unitary group for n qubits. The group $\mathcal{G}_n$ can e.g. be the generalized Pauli group (i.e., the discrete Heisenberg–Weyl group), or the Clifford group. By modeling the quantum computation with and without errors as two coupled Markov chains living on the state space consisting of pairs of elements from these groups, we are able to capture a

weak form of time-dependency within the process of error accumulation. To see this, critically note that the assumption of a Markov property does not imply that the past and the future in the quantum computation are independent given any information concerning the present [179]. We must also note that while the individual elements of our two-dimensional Markov chain belong to a group, the two-dimensional Markov chain itself, here, is generally not a random walk on a group. Lastly, our Markov chain model works for an arbitrary number of qubits. These model features are all relevant to the topic of error modeling in quantum computing and since the Markov property is satisfied in randomized benchmarking, the model has immediate application. The method is generic in the sense that any measure of distance between two pure quantum states may be used to quantify the error and that it allows for a wide range of error distributions. The method can handle nonuniform, gate- and time-dependent errors. Concretely, for arbitrary measure of distance and a wide range of error distributions, we will calculate (i) the expected error at time t, (ii) the probability that an error is larger than a threshold δ at time t and (iii) the probability that the error has *ever* been larger than a threshold δ before time t and we do so both for random and nonrandom circuits.

In addition to studying a model for discrete Markovian error accumulation in quantum circuits, we also briefly study a random walk model on the three-dimensional sphere [193]. This model is commonly used to describe the average dephasing of a single qubit (or spin) [194]. Using this model, we characterize the distribution and expectation of the trace distance measuring the error that is accumulated over time. These derivations are, essentially, refinements that provide information about the higher-order statistics of the error accumulation in a single qubit.

The approach taken in this chapter is a hybrid between classical probability theory and quantum information theory. This hybridization allows us to do quite detailed calculations, but not every quantum channel will satisfy the necessary assumptions such as Markovianity of the error distribution. On the other hand, in cases where one introduces their own source of randomness (such as in randomized benchmarking), the assumptions are met naturally. It should furthermore be noted that the numerical complexity of the exact expressions we provide is high for large quantum circuits. The precise difficulty of evaluating our expressions depends on the particulars of the quantum circuit one looks at. For practical purposes, we therefore also provide an explicit bound on the maximum error probability that is of lower numerical complexity. Furthermore, we also discuss a reduction in complexity that occurs when starting a quantum computation from a stabilizer state: the coupled Markov chain's state space then reduces in size. Reference [183] is relevant to mention here, because similar to our observations, these authors also note the generally high computational complexity of error analysis in quantum circuits. The issue is approached in [183] differently and in fact combinatorially by converting circuits

into directed graphs, tracing so-called fault-paths through these graphs and therewith estimating the success rates of circuits.

To illustrate and substantiate our theoretical results, we provide detailed discussions of further numerical experiments that we ran with a quantum simulator purpose-built for this research. Experiments include: the application of our formulae to randomized benchmarking; a comparison between simulated results on the accumulation of errors on a single qubit and our explicit formulae, as well as to a traditional method that calculates the evolution of the trace distance when repeating a depolarization channel; the effect of gate-dependent error distributions on the accumulation of errors in a one- and two-dimensional quantum circuit; and lowering the misclassification probability in a circuit that implements the Deutsch–Jozsa Algorithm for one classical bit.

Finally, we use the expressions that describe how likely it is that errors accumulate to answer two operational questions that will help advance the domain of practical quantum computing [195]. First, we calculate and bound analytically how many quantum gates $t^{\star}_{\delta,\gamma}$ one can apply before an error measure of your choice exceeds a threshold δ with a probability above γ. This information is useful for deciding how often a quantum computer should perform repairs on qubits and is particularly opportune at this moment since quantum gates fail $O(0.1\text{–}1\%)$ of the time [195]. Related but different ideas can be found in [174, §2.3], where the accumulation of bit-flips and rotations on a repetition code is studied and a time to failure is derived and in [196, §V], where an upper bound on the number of necessary measurements for a randomized benchmarking protocol is derived. Second, using techniques from optimization, we design a simulated annealing method that improves existing circuits by swapping out gate pairs to achieve lower rates of error accumulation. There is related literature where the aim is to reduce the circuit depth [197, 198, 199], but an explicit expression for error accumulation has not yet been leveraged in the same way. Moreover, we also discuss conditions under which this tailor-made method is guaranteed to find the best possible circuit. Both of these excursions illustrate how the availability of an analytical expression for the accumulation of errors allows us to proceed with second-tier optimization methods to facilitate quantum computers in the long-term. We further offer an additional proof-of-concept that simulated annealing algorithms can reduce error accumulation rates in existing quantum circuits when taking error distributions into account: we illustrate that the misclassification probability in a circuit that implements the Deutsch–Jozsa Algorithm for one classical bit [200, 201] can be lowered by over 40%. In this proof of concept we have chosen an example error distribution that is gate-dependent and moreover one that is such that *not* applying a gate gives the lowest error rate in this model; applying a single-qubit gate results in a medium error rate; and applying a two-qubit gate gives the largest probability that an error may occur.

This chapter is structured as follows. In Section 6.2, we give the model aspects pertaining to the quantum computation (gates, error dynamics and error measures) and we introduce the coupled Markov chain to describe error accumulation. In Section 6.3, we provide the relation between the probability of error and the hitting time distributions and we derive the error distributions as well as its bound. We also calculate the higher-order statistics of an error accumulation model for a single qubit that undergoes (continuous) random phase kicks and depolarization. In Section 6.4, we illustrate our theoretical results by comparing to numerical results of a quantum simulator we wrote for this chapter. In Section 6.5, we discuss the simulated annealing scheme. Finally, in Section 6.6, we conclude with ideas for future research.

6.2. MODEL AND COUPLED MARKOV CHAIN

6.2.1. GATES AND ERRORS IN QUANTUM COMPUTING

It is generally difficult to describe large quantum systems on a classical computer for the reason that the state space required increases exponentially in size with the number of qubits [202]. However, the stabilizer formalism is an efficient tool to analyze such complex systems [203]. Moreover, the stabilizer formalism covers many paradoxes in quantum mechanics [204], including the Greenberger–Horne–Zeilinger (GHZ) experiment [205], dense quantum coding [206] and quantum teleportation [207]. Specifically, the stabilizer circuits are the smallest class of quantum circuits that consist of the following four gates: $\omega = \mathrm{e}^{\mathrm{i}\pi/4}$, $H = (1/\sqrt{2})\big((1,1);(1,-1)\big)$, $S = \big((1,0);(0,\mathrm{i})\big)$, and $Z_c = \big((1,0,0,0);(0,1,0,0);(0,0,1,0);(0,0,0,-1)\big)$. These four gates are closed under the operations of tensor product and composition [208]. As a consequence of the Gottesman–Knill theorem, stabilizer circuits can be efficiently simulated on a classical computer [209].

Unitary stabilizer circuits are also known as the Clifford circuits; the Clifford group $\mathscr{C}_n$ can be defined as follows. First: let $P \triangleq \{I, X, Y, Z\}$ denote the Pauli matrices, so $I = ((1,0);(0,1))$, $X = ((0,1);(1,0))$, $Y = ((0,-i);(i,0))$, and $Z = ((1,0);(0,-1))$, and let $P_n \triangleq \{\sigma_1 \otimes \cdots \otimes \sigma_n \mid \sigma_i \in P\}$ denote the Pauli matrices on n qubits. The Pauli matrices are commonly used to model errors that can occur due to the interactions of the qubit with its environment [210]. In the case of a single qubit, the matrix I represents that there is no error, the matrix X that there is a bit-flip error, the matrix Z that there is a phase-flip error and the matrix Y that there are both a bit-flip and a phase-flip error. The multi-qubit case interpretations follow analogously. Second: let $P_n^* = P_n / I^{\otimes n}$. We now define the Clifford group on n qubits by $\mathscr{C}_n \triangleq \{U \in \mathscr{U}(2^n) \mid \sigma \in \pm P_n^* \Rightarrow U\sigma U^\dagger \in \pm P_n^*\}/\mathscr{U}(1)$.

The fact that $\mathscr{C}_n$ is a group can be verified by checking the two necessary properties (see Appendix E.1). The Clifford group on n qubits is finite [211], and we will ignore the global phase throughout this chapter for convenience; its size is then $|\mathscr{C}_n| =$

$2^{n^2+2n}\prod_{i=1}^{n}\left(4^i-1\right)$. Moreover, for a single qubit, a representation for the Clifford group $\mathscr{C}_1 = \{C_1, C_2, \cdots, C_{24}\}$ can then be enumerated and its elements are for example shown in [180] and [182].

6.2.2. DYNAMICS OF ERROR ACCUMULATION

Suppose that we had a faultless, perfect quantum computer. Then a faultless quantum mechanical state ρ_t at time t could be calculated under a gate sequence $\mathscr{U}_\tau = \{U_1, \ldots, U_\tau\}$ from the initial state $\rho_0 \triangleq |\psi_0\rangle\langle\psi_0|$. Here $\tau < \infty$ denotes the sequence length and $t \in \{0, 1, \cdots, \tau\}$ enumerates the intermediate steps. On the other hand, with an imperfect quantum computer, a possibly faulty quantum mechanical state σ_t at time t would be calculated under both $\mathscr{U}_t$ and some (unknown) noise sequence $\mathscr{E}_t = \{\Lambda_1, \ldots, \Lambda_t\}$ starting from an initial state $\sigma_0 \triangleq |\Psi_0\rangle\langle\Psi_0|$ possibly different from ρ_0. We define the set of all pure states for n qubits as $\mathscr{S}^n$ and consider the situation that $|\psi_0\rangle, |\Psi_0\rangle \in \mathscr{S}^n$.

To be precise, define for the faultless quantum computation

$$\rho_t \triangleq |\psi_t\rangle\langle\psi_t| = U_t|\psi_{t-1}\rangle\langle\psi_{t-1}|U_t^\dagger \tag{6.1}$$

for times $t = 1, 2, \ldots, \tau$. Let $X_t \triangleq U_t U_{t-1}\cdots U_1$ be shorthand notation such that $\rho_t = X_t \rho_0 X_t^\dagger$. For the possibly faulty quantum computation, define

$$\sigma_t \triangleq |\Psi_t\rangle\langle\Psi_t| = \Lambda_t U_t|\Psi_{t-1}\rangle\langle\Psi_{t-1}|U_t^\dagger\Lambda_t^\dagger$$

for times $t = 1, 2, \ldots, \tau$, respectively. Introduce also the shorthand notation $Y_t \triangleq \Lambda_t U_t \Lambda_{t-1} U_{t-1}\cdots\Lambda_1 U_1$ such that $\sigma_t = Y_t \sigma_0 Y_t^\dagger$. The analysis in this chapter can immediately be extended to the case where errors (also) precede the gate. The error accumulation process is also illustrated in Figure 6.1.

a) Faultless computation:

$$\rho_0 \xrightarrow{U_1} \rho_1 \xrightarrow{U_2} \cdots \xrightarrow{U_{\tau-1}} \rho_{\tau-1} \xrightarrow{U_\tau} \rho_\tau$$

b) Potentially faulty computation:

$$\sigma_0 \xrightarrow{\Lambda_1 U_1} \sigma_1 \xrightarrow{\Lambda_2 U_2} \cdots \xrightarrow{\Lambda_{\tau-1} U_{\tau-1}} \sigma_{\tau-1} \xrightarrow{\Lambda_\tau U_\tau} \sigma_\tau$$

Figure 6.1: Schematic depiction of the coupled quantum mechanical states ρ_t and σ_t for times $t = 0, 1, \cdots, \tau$. a) Faultless computation. The state ρ_t is calculated based on a gate sequence $\mathscr{U}_t = \{U_1, \ldots, U_t\}$ from the initial state ρ_0. b) Potentially faulty computation. The state σ_t is calculated using *the same* gate sequence $\mathscr{U}_t = \{U_1, \ldots, U_t\}$ and an additional error sequence $\mathscr{E}_t = \{\Lambda_1, \ldots, \Lambda_t\}$. The final state σ_τ can depart from the faultless state ρ_τ because of errors.

6.2.3. DISTANCE MEASURES FOR QUANTUM ERRORS

The error can be quantified by any measure of distance between the faultless quantum-mechanical state ρ_t and the possibly faulty quantum-mechanical state σ_t for steps $t = 0, 1, \ldots, \tau$. For example, we can use the fidelity $F_t \triangleq \mathrm{Tr}\sqrt{{\rho_t}^{1/2}\sigma_t{\rho_t}^{1/2}}$ [3], or the Schatten d-norm [212] defined by

$$D_t \triangleq \|\sigma_t - \rho_t\|_d = \tfrac{1}{2}\mathrm{Tr}\left[\left\{(\sigma_t - \rho_t)^\dagger(\sigma_t - \rho_t)\right\}^{\frac{d}{2}}\right]^{\frac{1}{d}} \tag{6.2}$$

for any $d \in [1, \infty)$. The Schatten d–norm reduces to the trace distance for $d = 1$, the Frobenius norm for $d = 2$ and the spectral norm for $d = \infty$. In the case of one qubit, the trace distance between quantum-mechanical states ρ_t and σ_t equals half of the Euclidean distance between ρ_t and σ_t when representing them on the Bloch sphere [3]. It is well known that the trace distance is invariant under unitary transformations [3]; a fact that we leverage in Section 6.3.

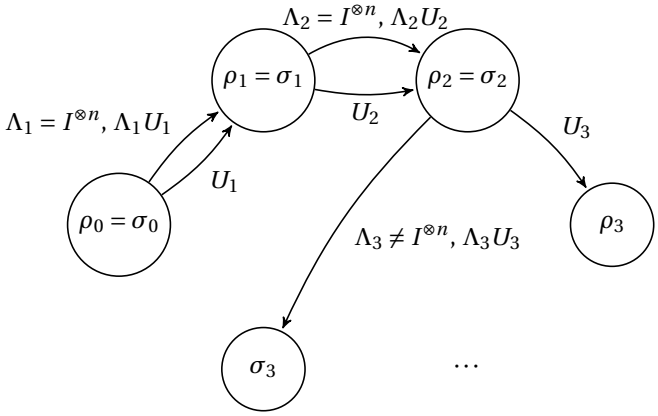


Figure 6.2: Coupled chain describing the quantum circuit with errors. In this depiction, we start from *the same* initial state for simplicity. Here an error $\Lambda_3 \neq I^{\otimes n}$ occurs as the third gate is applied. Note that the coupled chain ρ_t, σ_t separates.

In this chapter, we are going to analyze the statistical properties of some arbitrary distance measure (one may choose) between the quantum mechanical states ρ_t and σ_t for times $t = 0, 1, \ldots, \tau$. For illustration, we will state the results in terms of the Schatten d–norm and so are after its expectation $\mathbb{E}[D_t]$, as well as its probabilities $\mathbb{P}[D_t \leq \delta]$, $\mathbb{P}[\max_{0\leq s\leq t} D_s \leq \delta]$. Throughout this chapter, the operator $\mathbb{P}$ and thus also $\mathbb{E}$ are with respect to a sufficiently rich probability space $(\Omega, \mathbb{P}, \mathscr{F})$ that each time can describe the Markov chain being considered. As we show in Appendix E.2, in case of the trace distance ($d = 1$), these probabilities can then be related to the corresponding probabilities for the

fidelity. With $d = 1$, it holds that $\mathbb{P}[F_t \geq 1-\varepsilon] \geq \mathbb{P}[D_t \leq \varepsilon]$ for all $t \geq 0$. Furthermore,

$$\mathbb{P}[\min_{0\leq s\leq t} F_s \geq 1-\varepsilon] \geq \mathbb{P}[\max_{0\leq s\leq t} D_s \leq \varepsilon]. \tag{6.3}$$

6.3. ERROR ACCUMULATION

6.3.1. DISCRETE, RANDOM ERROR ACCUMULATION (MULTI-QUBIT CASE)

Following the model described in Section 6.2 and illustrated in Figure 6.1 and Figure 6.2, we define the gate pairs $Z_t \triangleq (X_t, Y_t)$ for $t = 0, 1, 2, \ldots, \tau$ and suppose that $Z_0 = z_0$ with probability one where $z_0 = (x_0, y_0)$ is deterministic and given a priori. Note in particular that if the initial state is prepared without error, then $\rho_0 = \sigma_0$ and consequently $z_0 = (I^{\otimes n}, I^{\otimes n})$. If on the other hand the initial state is e.g. prepared incorrectly as $y_0|\psi_0\rangle$ instead of $|\psi_0\rangle$, then $z_0 = (I^{\otimes n}, y_0)$.

THE CASE OF RANDOM CIRCUITS

We consider first the scenario that each next gate is selected randomly and independently from everything but the last system state. This assumption is satisfied in e.g. the randomized benchmarking protocol [180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190]. The probabilities $\mathbb{P}_{z_0}[D_t > \delta]$ and $\mathbb{P}_{z_0}[\max_{0\leq s\leq t} D_s \leq \delta]$ can then be calculated once the initial states $|\psi_0\rangle$, $|\Psi_0\rangle$ and the *transition matrix* are known. Here, the subscript z_0 reminds us of the initial state the Markov chain is started from.

Let the transition matrix of the Markov chain $\{Z_t\}_{t\geq 0}$ be denoted element-wise by $P_{z,w} \triangleq \mathbb{P}[Z_{t+1} = w | Z_t = z]$ for $z = (x, y), w = (u, v) \in \mathcal{G}_n^2$. The transition matrix satisfies $P \in [0,1]^{|\mathcal{G}_n|^2 \times |\mathcal{G}_n|^2}$ and the elements of each of its rows sum to one. Let

$$P^{(t)}_{z_0,w} \triangleq \mathbb{P}[Z_t = w | Z_0 = z_0] = (P^t)_{z_0,w} \tag{6.4}$$

stand in for the probability that the process is at state w at time t starting from $Z_0 = z_0$. Note that the second equality follows from the Markov property [179].

Example 1: Consider the situation that the error depends on the last gate. The transition probability $P_{z,w}$ for $z = (x, y), w = (u, v) \in \mathcal{G}_n^2$ can then be calculated as follows. For the faultless computation, a gate $U = ux^{-1}$ that transfers the density matrix $x\rho_0 x^\dagger$ to $u\rho_0 u^\dagger$ is randomly chosen according to a gate probability vector κ. For the possibly faulty computation, an error that transfers the density matrix $y\sigma_0 y^\dagger$ to $v\sigma_0 v^\dagger$, after the gate $U = ux^{-1}$, is $\Lambda = vy^{-1}xu^{-1}$. Let $\zeta(\Lambda = vy^{-1}xu^{-1}|ux^{-1})$ denote the probability that the error $\Lambda = vy^{-1}xu^{-1}$ occurs given that the gate $U = ux^{-1}$ just occurred. The transition matrix then satisfies $\mathbb{P}[Z_{t+1} = w | Z_t = z] = \kappa(U = ux^{-1})\zeta(\Lambda = vy^{-1}xu^{-1}|ux^{-1})$ component-wise.

Example 2: If we assume that errors and gates are independently generated, then the transition matrix satisfies $\mathbb{P}[Z_{t+1} = w | Z_t = z] = \kappa(U = ux^{-1})\zeta(\Lambda = vy^{-1}xu^{-1})$ component-wise.

We are now after the probability that the distance D_t is larger than a threshold δ. We define thereto the set of δ-*bad gate pairs* by

$$\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta} \triangleq \{(x,y) \in \mathscr{G}_n^2 | \|x\rho_0 x^\dagger - y\sigma_0 y^\dagger\|_d > \delta\} \tag{6.5}$$

for $|\psi_0\rangle, |\Psi_0\rangle \in \mathscr{S}^n, \delta \geq 0$, as well as the *hitting time* of any set $\mathscr{A} \subseteq \mathscr{G}_n^2$ by

$$T_{\mathscr{A}} \triangleq \inf\{t \geq 0 | Z_t \in \mathscr{A}\} \tag{6.6}$$

with the convention that $\inf\phi = \infty$. Note that $T_{\mathscr{A}} \in \mathbb{N}_0 \cup \{\infty\}$ and that it is random. With definitions (6.5), (6.6), we have the convenient representation

$$\mathbb{P}_{z_0}[\max_{0\leq s\leq t} D_s \leq \delta] = 1 - \mathbb{P}_{z_0}[\max_{0\leq s\leq t} D_s > \delta] = 1 - \mathbb{P}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}} \leq t] \tag{6.7}$$

for this homogeneous Markov chain. As a consequence of (6.7), the analysis comes down to an analysis of the hitting time distribution for this coupled Markov chain (Figure 6.3).

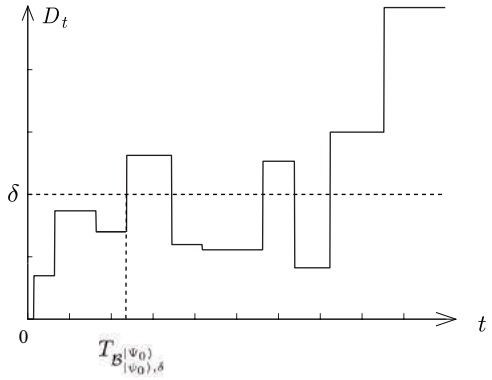


Figure 6.3: Schematic diagram of the hitting time $T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}}$.

Results. Define the matrix $B^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta} \in [0,1]^{|\mathscr{G}_n|^2 \times |\mathscr{G}_n|^2}$ element-wise by

$$\left(B^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}\right)_{z,w} \triangleq \begin{cases} P_{z,w} & \text{if} \quad w \notin \mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}, \\ 0 & \text{otherwise.} \end{cases} \tag{6.8}$$

Let the initial state vector be denoted by e_{z_0}, a $|\mathscr{G}_n|^2 \times 1$ vector with just the z_0-th element 1 and the others 0. Also let $1_{\mathscr{A}}$ denote the $|\mathscr{G}_n|^2 \times 1$ vector with ones in every coordinate corresponding to an element in the set $\mathscr{A}$. Let the transpose of an arbitrary matrix A be denoted by A^{T} and defined element-wise $(A^{\mathrm{T}})_{i,j} = A_{j,i}$. Finally, we define a $|\mathscr{G}_n|^2 \times 1$ vector $d^{|\Psi_0\rangle}_{|\psi_0\rangle} = \left(\|x\rho_0 x^\dagger - y\sigma_0 y^\dagger\|_d\right)_{(x,y)\in\mathscr{G}_n^2}$ enumerating all possible Schatten d-norm distances. We now state our first result:

Proposition 1 (Error accumulation in random circuits)**.** *For any $z_0 \in \mathscr{G}_n^2$, $\delta \geq 0$, $t = 0, 1, \ldots, \tau < \infty$: the expected error is given by*

$$\mathbb{E}_{z_0}[D_t] = e_{z_0}^{\mathrm{T}} P^t d_{|\psi_0\rangle}^{|\Psi_0\rangle}. \tag{6.9}$$

Similarly, the probability of error is given by

$$\mathbb{P}_{z_0}[D_t > \delta] = e_{z_0}^{\mathrm{T}} P^t 1_{\mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}}, \tag{6.10}$$

and is nonincreasing in δ. Furthermore; if $z_0 \notin \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}$, the probability of maximum error is given by

$$\mathbb{P}_{z_0}[\max_{0\leq s\leq t} D_s > \delta] = \sum_{s=1}^{t} e_{z_0}^{\mathrm{T}} \big(B_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}\big)^{s-1} \big(P - B_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}\big) 1_{\mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}}, \tag{6.11}$$

and otherwise it equals one. Lastly, (6.11) *is nonincreasing in δ and nondecreasing in t.*

The probability in (6.11) is a more stringent error measure than e.g. (6.10) is. The event $\{\max_{0\leq s\leq t} D_s < \delta\}$ implies after all that the error D_t has always been below the threshold δ up to and including at time t. The expected error $\mathbb{E}_{z_0}[D_t]$ and probability $\mathbb{P}_{z_0}[D_t > \delta]$ only concern the error *at* time t. Additionally, (6.11) allows us to calculate the maximum number of gates that can be performed. That is, $\mathbb{P}_{z_0}[\max_{0\leq s\leq t} D_s > \delta] \leq \gamma$ as long as

$$t \leq t^\star_{\delta,\gamma} \triangleq \max\big\{t \in \mathbb{N}_0 \big| \mathbb{P}_{z_0}[\max_{0\leq s\leq t} D_s > \delta] \leq \gamma\big\}. \tag{6.12}$$

In words: at most $t^\star_{\delta,\gamma}$ gates can be applied before an accumulated error of size at least δ occurred with probability at least γ.

Proof of (6.10). It follows from (6.5), mutual exclusivity and (6.4) that

$$\mathbb{P}_{z_0}[D_t > \delta] = \mathbb{P}_{z_0}[Z_t \in \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}] = \sum_{w\in\mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}} \mathbb{P}_{z_0}[Z_t = w] = \sum_{w\in\mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}} (P^t)_{z_0,w}. \tag{6.13}$$

The right-hand side equals (6.10) in matrix notation. To obtain the expression for the expectation, directly apply the definition of expectation for a discrete random variable:

$$\mathbb{E}_{z_0}[D_t] = \sum_{(x,y)\in\mathscr{G}_n^2} \|x\rho_0 x^\dagger - y\sigma_0 y^\dagger\|_D \mathbb{P}_{z_0}[Z_t = (x,y)]. \tag{6.14}$$

Using (6.4) and the definition of $d_{|\psi_0\rangle}^{|\Psi_0\rangle}$, this gives the result.

Proof of (6.11). If $z_0 \in \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}$, then $\mathbb{P}_{z_0}[T_{\mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}} = 0] = 1$. If $z_0 \notin \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}$, then use (6.8) to write

$$\begin{aligned}
\mathbb{P}_{z_0}[T_{\mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}} = s] &= \mathbb{P}_{z_0}[Z_1 \notin \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}, \ldots, Z_{s-1} \notin \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}, Z_s \in \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}] \\
&= \sum_{z_1\notin\mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}} \cdots \sum_{z_{s-1}\notin\mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}} \sum_{z_s\in\mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}} \mathbb{P}_{z_0}[Z_1 = z_1, \ldots, Z_s = z_s] \\
&= e_{z_0}^{\mathrm{T}} \big(B_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}\big)^{s-1} \big(P - B_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}\big) 1_{\mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}}
\end{aligned} \tag{6.15}$$

in matrix notation. The result follows after summing (6.15) for $s = 0, 1, \ldots, t-1$ by mutual exclusivity.

Note finally that for arbitrary $\delta_2 > \delta_1$, we have that $\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta_2} \subseteq \mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta_1}$. As a consequence,

$$\mathbb{P}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta_2}} \leq t] \leq \mathbb{P}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta_1}} \leq t]. \tag{6.16}$$

This establishes that $\mathbb{P}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}} \leq t]$ is nonincreasing in δ. By positivity of the summands, $\mathbb{P}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}} \leq t]$ is nondecreasing in t. □

Lower bound. For general $\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}$, the explicit calculation of (6.11) can be numerically intensive. It is however possible to provide a lower bound of lower numerical complexity via the expected hitting time of the set $\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}$.

Lemma 2 (Lower bound for random circuits)**.** *For any set $\mathscr{A} \subseteq \mathscr{G}_n^2$, the expected hitting times of a homogeneous Markov chain are the solutions to the linear system of equations $\mathbb{E}_z[T_{\mathscr{A}}] = 0$ for $z \in \mathscr{A}$, $\mathbb{E}_z[T_{\mathscr{A}}] = 1 + \sum_{w \notin \mathscr{A}} P_{z,w}\mathbb{E}_w[T_{\mathscr{A}}]$ for $z \notin \mathscr{A}$. Furthermore; for any $z_0 \in \mathscr{G}_n^2$, $\delta \geq 0$, $t = 0, 1, \ldots, \tau < \infty$:*

$$\mathbb{P}_{z_0}[\max_{0\leq s\leq t} D_s > \delta] \geq 0 \vee \Big(1 - \frac{\mathbb{E}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}}]}{t+1}\Big). \tag{6.17}$$

Here $a \vee b \triangleq \max\{a, b\}$.

Proof of (6.17). The first part is a standard result, see e.g. [31, p. 202]. The second part follows from Markov's inequality, i.e.,

$$\mathbb{P}_{z_0}[\max_{0\leq s\leq t} D_s \leq \delta] = \mathbb{P}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}} > t] \leq \frac{\mathbb{E}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}}]}{t+1}. \tag{6.18}$$

That is it. □

As a consequence of Lemma 2, $\mathbb{P}_{z_0}[\max_{0\leq s\leq t} D_s > \delta] \geq \gamma$ when $t \geq \mathbb{E}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}}]/(1-\gamma) - 1$, and in particular $\mathbb{P}_{z_0}[\max_{0\leq s\leq t} D_s > 0] > 0$ when $t \geq \mathbb{E}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,0}}]$. The values in the right-hand sides are thus upper bounds to the number of gates $t^\star_{\delta,\gamma}$ one can apply before δ error has occurred with probability γ:

$$t^\star_{\delta,\gamma} \leq \mathbb{E}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,0}}] \wedge \Big(\frac{\mathbb{E}_{z_0}[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}}]}{1-\gamma} - 1\Big) \tag{6.19}$$

for $\delta \geq 0, \gamma \in [0, 1]$. Here, $a \wedge b \triangleq \min\{a, b\}$.

Limitations of the method: types of quantum noise channels. The approach taken in this chapter is a hybrid between classical probability theory and quantum information theory. The results of this chapter are therefore not applicable to all quantum channels and it is important that we signal you the limitations.

As an illustrative example, consider the elementary circuit of depth $\tau = 1$ with $n = 1$ qubit, in which the one gate is restricted to the Clifford group $\{C_1, \ldots, C_{24}\}$, say. For such an elementary circuit, this chapter describes a classical stochastic process that chooses one of twenty-four quantum noise channel $\mathscr{F}^{(1)}, \ldots \mathscr{F}^{(24)}$ say according to some arbitrary classical probability distribution $\{p_i(\rho)\}$, i.e.,

$$\rho_0 \to \rho_1 = \mathscr{F}(\rho_0) = \begin{cases} \mathscr{F}^{(1)}(\rho_0) = C_1 \rho_0 C_1^\dagger & \text{w.p. } p_1(\rho_0), \\ \mathscr{F}^{(2)}(\rho_0) = C_2 \rho_0 C_2^\dagger & \text{w.p. } p_2(\rho_0), \\ \ldots & \\ \mathscr{F}^{(24)}(\rho_0) = C_{24} \rho_0 C_{24}^\dagger & \text{w.p. } p_{24}(\rho_0). \end{cases} \tag{6.20}$$

Here, the classical probability distribution $\{p_i(\rho)\}$ may be chosen arbitrarily and depend on the initial quantum state ρ_0 as indicated. For this elementary quantum circuit of depth $\tau = 1$ with $n = 1$ qubit, (6.20) characterizes the set of stochastic processes covered by our results in its entirety.

For example, Proposition 1 cannot be applied to the deterministic process

$$\rho_0 \to \rho_1 = \left\{ \mathscr{E}^{(1)}(\rho_0) = (1-p)\rho_0 + pY\rho_0 Y^\dagger \text{ w.p. } 1, \right. \tag{6.21}$$

nor to the deterministic process

$$\rho_0 \to \rho_1 = \left\{ \mathscr{E}^{(2)}(\rho_0) = (1-p)\rho_0 + \tfrac{p}{2} U\rho_0 U^\dagger + \tfrac{p}{2} U^\dagger \rho_0 U \text{ w.p. } 1. \right. \tag{6.22}$$

Here, $p \in (0,1)$ can be chosen arbitrarily and $U = \mathrm{e}^{-\mathrm{i}\pi Y/4}$ is a Clifford gate. The reason is that $\left(\mathscr{F}^{(1)} \neq \mathscr{F}^{(2)} \neq \cdots \neq \mathscr{F}^{(24)}\right) \neq \left(\mathscr{E}^{(1)} = \mathscr{E}^{(2)}\right)$ by the unitary freedom in the operator-sum representation [3, Thm. 8.2]. A meticulous reader will now note that the example quantum channels $\mathscr{E}^{(1)}$, $\mathscr{E}^{(2)}$ are however *averages* of two particular stochastic processes $\mathscr{F}$. That is: if $p_I = 1-p, p_Y = p$, then $\mathscr{E}^{(1)}(\rho) = \mathbb{E}[\mathscr{F}(\rho)]$; or if $p_I = 1-p, p_U = p_{U^\dagger} = \frac{p}{2}$, then $\mathscr{E}^{(2)}(\rho) = \mathbb{E}[\mathscr{F}(\rho)]$.

An alternative way to understand what is going on, is to consider that we are describing the time-evolution of a density matrix and that a density matrix expresses a subjective state of knowledge. The classical model described in this paper assumes that your best description of the system at each intermediate time step is a pure state and this is not the case in quantum channels $\mathscr{E}^{(1)}$, $\mathscr{E}^{(2)}$. Your best description of the system at each intermediate time step is a pure state e.g. in randomized benchmarking when you are drawing classical random variables to randomly choose a quantum gate *and* are being

informed of their outcomes. Note finally that the expectation and probability operators in this paper are with respect to a classical stochastic process that drives a random choice of quantum gates and that quantum measurements are thus not being modeled.

On how to construct the P matrix. Both Proposition 1 and Lemma 2 rely on constructing the P matrix. For illustrative purposes, we have written a script that will generate a valid P matrix after a user inputs a vector describing (gate-dependent) error probabilities. The code is publicly available on TU/e's GitLab server at `https://gitlab.tue.nl/20061069/markov-chains-for-error-accumulation-in-quantum-circuits`. Additionally, we discuss an example in Appendix E.3 for which we construct the P matrix as well as evaluate the lower bound in (6.11).

An average over the trajectories of the Markov chain. It is noteworthy that the results in Proposition 1 are averages over all noise trajectories that can be generated by the Markov chain. Consider e.g. (6.10), which reads in matrix notation:

$$\mathbb{P}_{z_0}[D_t > \delta] = e_{z_0}^T P^t 1_{\mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}} = \sum_{w \in \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}} (P^t)_{z_0,w}. \tag{6.23}$$

Expanding the matrix power, the right-hand side equals

$$\underbrace{\sum_{z_1 \in \mathscr{G}_n^2} \sum_{z_2 \in \mathscr{G}_n^2} \cdots \sum_{z_{t-1} \in \mathscr{G}_n^2} \sum_{w \in \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}}}_{\text{Term I}} \underbrace{P_{z_0,z_1} P_{z_1,z_2} \cdots P_{z_{t-1},w}}_{\text{Term II}} \tag{6.24}$$

$$= \mathbb{E}_{z_0}[\mathbb{1}[Z_1 \in \mathscr{G}_n^2, \ldots, Z_{t-1} \in \mathscr{G}_n^2, Z_t \in \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}]]. \tag{6.25}$$

Here, Term I enumerates all possible length-t trajectories of the Markov chain that start at some state $z_0 \in \mathscr{G}_n^2$ and end at any state $w \in \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}$. Term II is the probability that the specific trajectory $z_0 \to z_1 \to z_2 \to \cdots \to z_{t-1} \to w$ occurs in this Markov chain. Consequently, (6.25) is the expectation (average) of the random variable $\mathbb{1}[Z_1 \in \mathscr{G}_n^2, \ldots, Z_{t-1} \in \mathscr{G}_n^2, Z_t \in \mathscr{B}_{|\psi_0\rangle,\delta}^{|\Psi_0\rangle}]$ as indicated.

THE CASE OF NONRANDOM CIRCUITS

Suppose that the gate sequence $\mathscr{U}_\tau = \{U_1, ..., U_\tau\}$ is fixed *a priori* and that it is not generated randomly. Because the gate sequence is nonrandom, we have now that the faultless state $\rho_t = X_t \rho_0 X_t^\dagger$ is deterministic for times $t = 0, 1, \ldots, \tau$. On the other hand the potentially faulty state $\sigma_t = Y_t \rho_0 Y_t^\dagger$ is still (possibly) random.

We can now use a lower dimensional Markov chain to represent the system. To be precise: we will now describe the process $\{Y_t\}_{t \geq 0}$ (and consequently $\{\sigma_t\}_{t \geq 0}$) as an *inhomogeneous Markov chain*. Its transition matrices will now be time-dependent and

given element-wise by $Q_{y,v}(t) = \mathbb{P}[Y_{t+1} = v | Y_t = y]$ for $y, v \in \mathscr{G}_n$, $t \in \{0, 1, \ldots, \tau - 1\}$. Letting $Q^{(t)}_{y,v} \triangleq \mathbb{P}[Y_t = v | Y_0 = y]$ stand in for the probability that the process $\{Y_t\}_{t \geq 0}$ is at state v at time t starting from y, we have by the Markov property [179] that

$$Q^{(t)}_{y,v} = \big(\prod_{s=1}^{t} Q(s)\big)_{y,v} \quad \text{for} \quad y, v \in \mathscr{G}_n. \tag{6.26}$$

Note that the Markov chain modeled here is inhomogeneous, which is different from Section 6.3.1. In particular, the time-dependent transition matrix $Q(t)$ here cannot be expressed in terms of a power P^t of a transition matrix P on the same state space as in Section 6.3.1.

Example 3: Consider the situation that the probability that an error occurs depends on which gate was applied last. If we assume that $\mathbb{P}[\Lambda_{t+1} = \lambda | Y_t = y] = \zeta_{y,U_{t+1}}(\lambda)$ are given distributions for $y \in \mathscr{G}_n$, $t \in \{0, 1, \cdots, \tau - 1\}$ on $\lambda \in \mathscr{G}_n$, we can alternatively write the elements of the transition matrices as

$$\begin{aligned} Q_{y,v}(t) &= \mathbb{P}[Y_{t+1} = v | Y_t = y] \\ &= \sum_{\lambda \in \mathscr{G}_n} \mathbb{P}[Y_{t+1} = v | Y_t = y, \Lambda_{t+1} = \lambda] \mathbb{P}[\Lambda_{t+1} = \lambda | Y_t = y] \\ &= \sum_{\lambda \in \mathscr{G}_n} \mathbb{1}[\lambda U_{t+1} y \rho_0 y^\dagger U^\dagger_{t+1} \lambda^\dagger = v \rho_0 v^\dagger] \zeta_{y,U_{t+1}}(\lambda). \end{aligned} \tag{6.27}$$

Here, we have used the law of total probability.

Example 4: If errors occur independently and with probability $\mathbb{P}[\Lambda_{t+1} = \lambda] = \zeta(\lambda)$, then
$Q_{y,v}(t) = \sum_{\lambda \in \mathscr{G}_n} \mathbb{1}[\lambda U_{t+1} y \rho_0 y^\dagger {U_{t+1}}^\dagger \lambda^\dagger = v \rho_0 v^\dagger] \zeta(\lambda)$.

Results. Now define the sets of (δ, t)-bad gate pairs by $\mathscr{B}^{|\Psi_0\rangle, t}_{|\psi_0\rangle, \delta} \triangleq \big\{x \in \mathscr{U}_n \big| \|\rho_t - x \sigma_0 x^\dagger\|_d > \delta\big\}$ for $|\psi_0\rangle, |\Psi_0\rangle \in \mathscr{S}^n$, $t \in \{0, 1, \ldots, \tau\}$, $\delta \geq 0$. Also define the matrices $B^{|\Psi_0\rangle, t}_{|\psi_0\rangle, \delta} \in [0,1]^{|\mathscr{G}_n| \times |\mathscr{G}_n|}$ element-wise by

$$\big(B^{|\Psi_0\rangle, t}_{|\psi_0\rangle, \delta}\big)_{y,v} \triangleq \begin{cases} Q_{y,v}(t) & \text{if} \quad v \notin \mathscr{B}^{|\Psi_0\rangle, t}_{|\psi_0\rangle, \delta}, \\ 0 & \text{otherwise}, \end{cases} \tag{6.28}$$

for $t = 0, 1, \ldots, \tau$. Recall the notation introduced above Proposition 1. Similarly enumerate in the vector d_{ρ_t} the Schatten d-norms between any of the possibles states of σ_t and the faultless state ρ_t. We state our second result:

Proposition 3 (Error accumulation in nonrandom circuits)**.** *For any $y_0 \in \mathscr{G}_n$, $\delta \geq 0$, $t = 0, 1, \ldots, \tau < \infty$: the expected error is given by $\mathbb{E}_{y_0}[D_t] = e^{\mathrm{T}}_{y_0}\big(\prod_{k=1}^{t} Q(k)\big) d_{\rho_t}$. Similarly, the distribution of error is given by*

$$\mathbb{P}_{y_0}[D_t > \delta] = e^{\mathrm{T}}_{y_0}\big(\prod_{k=1}^{t} Q(k)\big) 1_{\mathscr{B}^{|\Psi_0\rangle, t}_{|\psi_0\rangle, \delta}}. \tag{6.29}$$

Furthermore; if $y_0 \notin \mathscr{B}^{|\Psi_0\rangle,0}_{|\psi_0\rangle,\delta}$, the distribution of maximum error is given by

$$\mathbb{P}_{y_0}[\max_{0\le s\le t} D_s > \delta] = \sum_{s=0}^{t-1}\Big(e_{y_0}^{\mathrm{T}}\big(\prod_{r=0}^{s} B^{|\Psi_0\rangle,r}_{|\psi_0\rangle,\delta}\big)\times\big(Q(s+1)-B^{|\Psi_0\rangle,s+1}_{|\psi_0\rangle,\delta}\big)1_{\mathscr{B}^{|\Psi_0\rangle,s+1}_{|\psi_0\rangle,\delta}}\Big), \tag{6.30}$$

and otherwise it equals one.

Proof of (6.29). From $\mathscr{B}^{|\Psi_0\rangle,t}_{|\psi_0\rangle,\delta}$'s definition and mutual exclusivity it follows immediately that

$$\mathbb{P}_{y_0}[D_t > \delta] = \mathbb{P}_{y_0}[Y_t \in \mathscr{B}^{|\Psi_0\rangle,t}_{|\psi_0\rangle,\delta}] = \sum_{v\in\mathscr{B}^{|\Psi_0\rangle,t}_{|\psi_0\rangle,\delta}} \mathbb{P}_{y_0}[Y_t = v] \tag{6.31}$$

for $|\psi_0\rangle, |\Psi_0\rangle \in \mathscr{S}^n, \delta \ge 0$. Using (6.26) and continuing from (6.31), we obtain

$$\mathbb{P}_{y_0}[D_t > \delta] = \sum_{v\in\mathscr{B}^{|\Psi_0\rangle,t}_{|\psi_0\rangle,\delta}} e_{y_0}^{\mathrm{T}}\big(\prod_{k=1}^{t} Q(k)\big)_{y,v}. \tag{6.32}$$

This simplifies to (6.29) in matrix notation. To obtain the expression for the expectation, apply the same arguments as were used for Proposition 1, but use (6.26) instead.

Proof of (6.30). We can again explicitly calculate the result using a hitting time analysis, but the expressions expand due to the time-dependency of $\mathscr{B}^{|\Psi_0\rangle,t}_{|\psi_0\rangle,\delta}$. If $y_0 \in \mathscr{B}^{|\Psi_0\rangle,0}_{|\psi_0\rangle,\delta}$, then $\mathbb{P}_{y_0}[\max_{0\le r\le s} D_r > \delta] = 1$. Otherwise

$$\mathbb{P}_{y_0}[\{\max_{0\le r\le s-1} D_r \le \delta\}\cap\{D_s > \delta\}] \tag{6.33}$$

$$= \mathbb{P}_{y_0}[Y_1 \notin \mathscr{B}^{|\Psi_0\rangle,1}_{|\psi_0\rangle,\delta}, \dots, Y_{s-1} \notin \mathscr{B}^{|\Psi_0\rangle,s-1}_{|\psi_0\rangle,\delta}, Y_s \in \mathscr{B}^{|\Psi_0\rangle,s}_{|\psi_0\rangle,\delta}] \tag{6.34}$$

$$= \sum_{y_1\notin\mathscr{B}^{|\Psi_0\rangle,1}_{|\psi_0\rangle,\delta}} \cdots \sum_{y_s\in\mathscr{B}^{|\Psi_0\rangle,s}_{|\psi_0\rangle,\delta}} \mathbb{P}_{y_0}[Y_1 = y_1, \dots, Y_s = y_s] \tag{6.35}$$

$$= \sum_{y_1\notin\mathscr{B}^{|\Psi_0\rangle,1}_{|\psi_0\rangle,\delta}} \cdots \sum_{y_{s-1}\in\mathscr{B}^{|\Psi_0\rangle,s-1}_{|\psi_0\rangle,\delta}} \sum_{y_s\in\mathscr{B}^{|\Psi_0\rangle,s}_{|\psi_0\rangle,\delta}} \prod_{r=0}^{s-1} Q_{y_r,y_{r+1}}(r).$$

Recalling (6.28), we can equivalently write (6.35) in matrix notation as

$$\mathbb{P}_{y_0}[\{\max_{0\le r\le s-1} D_r \le \delta\}\cap\{D_s > \delta\}] = e_{y_0}^{\mathrm{T}}\big(\prod_{r=1}^{s-1} B^{|\Psi_0\rangle,r}_{|\psi_0\rangle,\delta}\big)\big(Q(s)-B^{|\Psi_0\rangle,s}_{|\psi_0\rangle,\delta}\big)1_{\mathscr{B}^{|\Psi_0\rangle,s}_{|\psi_0\rangle,\delta}}. \tag{6.36}$$

Summing (6.36) over $s = 0, 1, \dots, t-1$ completes the proof by mutual exclusivity. □

On how to construct the Q matrix. The script that we created that can generate example P matrices, can also generate valid Q matrices after the user inputs a vector describing (gate-dependent) error probabilities. Recall that this code is available on TU/e's GitLab server here: https://gitlab.tue.nl/20061069/markov-chains-for-error-accumulation-in-quantum-circuits.

STATE SPACE REDUCTION IN STABILIZER CIRCUITS

The set of stabilizer gates [213] for a state $|\psi\rangle$ is defined as the set of gates $\mathscr{M} \in \mathscr{G}_n \setminus I^{\otimes n}$ that satisfy $\mathscr{M}|\psi\rangle = e^{i\gamma}|\psi\rangle$ for some $\gamma \in \mathbb{R}$. Since $e^{i\gamma}$ is a global phase that cannot be observed, $\mathscr{M}|\psi\rangle = e^{i\gamma}|\psi\rangle$ can also be understood as part of an equivalence class $\mathscr{M}|\psi\rangle \equiv |\psi\rangle$. The state $|\psi\rangle$ in $\mathscr{M}|\psi\rangle \equiv |\psi\rangle$ is called the *stabilizer state* [214]. For one qubit and in case of the Pauli group, examples include $|0\rangle$, $|1\rangle$ and $|\pm\rangle = (1/2)(|0\rangle \pm |1\rangle)$. Remark 4 shows that there exist 2^n stabilizer states for any gate $\mathscr{M} \in \mathscr{G}_n \setminus I^{\otimes n}$. Its proof is relegated to Appendix E.4.

Remark 4. *For any gate $\mathscr{M} \in \mathscr{G}_n \setminus I^{\otimes n}$ there are 2^n states $|\psi_0\rangle$ that satisfy $\mathscr{M}|\psi_0\rangle = e^{i\gamma}|\psi_0\rangle$ for some $\gamma \in \mathbb{R}$.*

The advantage of starting a quantum circuit from a stabilizer state is that the state space is smaller. It moreover can be proved that, under the assumptions of Section 6.2, when starting initially from a stabilizer state, all states reached during the quantum computation will themselves be stabilizer states. Define the set of *reachable density matrices* from an initial state $|\psi_0\rangle \in \mathscr{S}^n$, by

$$\mathscr{R}_{|\psi_0\rangle} \triangleq \{g|\psi_0\rangle \,|\, g \in \mathscr{G}_n\}. \tag{6.37}$$

The exact number of reachable states can be calculated by the method in Appendix E.7. Taking the Clifford group gates on two qubits as an example, the number of gates $|\mathscr{C}_2| = 11520$. However, there are just 60 reachable states if the initial state is $|00\rangle$. The proof of Remark 5 can be found in Appendix E.5.

Remark 5. *Given a gate $\mathscr{M} \in \mathscr{G}_n \setminus I^{\otimes n}$ and a state $|\psi_0\rangle \in \mathscr{S}_n$ such that $\mathscr{M}|\psi_0\rangle = e^{i\gamma}|\psi_0\rangle$ for some $\gamma \in \mathbb{R}$, then for any state $|\psi_1\rangle \in \mathscr{R}_{|\psi_0\rangle}$ there exists an $\mathscr{H} \in \mathscr{G}_n \setminus I^{\otimes n}$ such that $\mathscr{H}|\psi_1\rangle = e^{i\gamma}|\psi_1\rangle$.*

A consequence of Remark 5 is namely that for any reachable state $|\Psi\rangle$ there are at least two different gates $\mathscr{M}_i, \mathscr{M}_j \in \mathscr{G}_n$ whose corresponding states $\mathscr{M}_i|\psi_0\rangle$ and $\mathscr{M}_j|\psi_0\rangle$ are equivalent (up to a phase) to same state $|\Psi\rangle$, since $\mathscr{M}_i|\psi_0\rangle \equiv \mathscr{M}_j|\psi_0\rangle \equiv |\Psi\rangle$ if we let $|\Psi\rangle = \mathscr{M}_i|\psi_0\rangle$ and $\mathscr{M}_j = \mathscr{H}\mathscr{M}_i$. The number of reachable states $|\mathscr{R}_{|\psi_0\rangle}|$ is thus upper bounded by $1/2|\mathscr{G}_n|$ when starting from a stabilizer state.

6.3.2. CONTINUOUS, RANDOM ERROR ACCUMULATION (ONE-QUBIT CASE)

In this section, we analyze the case where a single qubit:

1. receives a random perturbation on the Bloch sphere after each s-th unitary gate according to a continuous distribution $p_s(\alpha)$ and
2. depolarizes to the completely depolarized state $I/2$ with probability $q \in [0,1]$ after each unitary gate,

by considering it an absorbing random walk on the Bloch sphere. The key point leveraged here is that the trace distance is invariant under rotations. Hence a sufficiently symmetric random walk distribution will give the error probabilities.

Model. Let R_0 be an initial point on the Bloch sphere. Every time a unitary quantum gate is applied, the qubit is rotated and receives a small perturbation. This results in a random walk $\{R_t\}_{t\geq 0}$ on the Bloch sphere for as long as the qubit has not depolarized. Because the trace distance is invariant under rotations and since the rotations are applied both to ρ_t and σ_t, we can ignore the rotations. We let ν denote the random time at which the qubit depolarizes. With the usual independence assumptions, $\nu \sim \text{Geometric}(q)$.

Define $\mu_t(r)$ for $t < \nu$ as the probability that the random walk is in a solid angle Ω about r (in spherical coordinates) conditional on the qubit not having depolarized yet. That is,

$$\mathbb{P}[R_t \in \mathscr{S} | \nu > t] \triangleq \int_{\mathscr{S}} \mu_t(r) \mathrm{d}\Omega(r). \tag{6.38}$$

We assume without loss of generality that $R_0 = \hat{z}$. From [193], the initial distribution is then given by

$$\mu_0 = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} P_n(\cos\theta). \tag{6.39}$$

Here, the $P_n(\cdot)$ denote the Legendre polynomials. Also introduce the shorthand notation

$$\Lambda_{n,t} \triangleq \prod_{s=1}^{t} \int_0^{\pi} P_n(\cos\alpha) \mathrm{d}p_s(\alpha) \tag{6.40}$$

for convenience. As we will see in Proposition 6 in a moment, these constants will turn out to be the coefficients of an expansion for the expected trace distance (see (6.42)). Recall that here, $p_s(\alpha)$ denotes the probability measure of the angular distance for the random walk on the Bloch sphere at time t (see (i) above). In particular: if $p_t(\alpha) = \delta(\alpha)$ for all $t \geq 0$ meaning that each step is taken into a random direction but exactly of angular length α, then $\Lambda_{n,t} = (P_n(\cos\alpha))^t$. From [193], it follows that after t unitary quantum gates have been applied without depolarization having occurred,

$$\mu_t = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \Lambda_{n,t} P_n(\cos\theta). \tag{6.41}$$

Results. In this section we specify D_t as the trace distance. We are now in position to state our findings:

Proposition 6 (Single qubit)**.** *For $0 \leq \delta \leq 1$, $t \in \mathbb{N}_+$: the expected trace distance satisfies*

$$\mathbb{E}[D_t] = \tfrac{1}{2} - (1-q)^t \Big(\tfrac{1}{2} + 2 \sum_{n=0}^{\infty} \frac{\Lambda_{n,t}}{(2n-1)(2n+3)} \Big). \tag{6.42}$$

The probability of the trace distance is given by

$$\mathbb{P}[D_t \leq \delta] = \mathbb{1}[\tfrac{1}{2} \in [0,\delta]]\big(1-(1-q)^t\big) \tag{6.43}$$

$$+ (1-q)^t \sum_{n=0}^{\infty} (2n+1)\Lambda_{n,t} \sum_{r=1}^{n+1} (-1)^{r+1}\delta^{2r} C_{r-1} \binom{n+r-1}{2(r-1)}. \tag{6.44}$$

Here, the C_r denote the Catalan numbers. Alternative forms include:

$$\mathbb{P}[D_t \leq \delta | \nu > t] = \delta^2 \sum_{n=0}^{\infty} (2n+1)\Lambda_{n,t} {}_2F_1(-n, n+1, 2; \delta^2), \quad \text{and} \tag{6.45}$$

$$\mathbb{P}[D_t \leq \delta | \nu > t] = \delta^2 \sum_{n=0}^{\infty} (2n+1)\Lambda_{n,t} \frac{n!}{(2)_n} P_n^{(1,-1)}(1-2\delta^2) \tag{6.46}$$

with ${}_2F_1(a,b,c;z)$ the Hypergeometric function, $(\cdot)_n$ the Pochhammer symbol and $P_n^{(\alpha,\beta)}(x)$ the Jacobi polynomials. Finally; the probability of maximum trace distance is lower bounded by

$$\mathbb{P}[\max_{0\leq s\leq t} D_s \leq \delta | \nu > t] \geq 0 \vee \Big(1 - t + \delta^2 \sum_{s=1}^{t} \sum_{n=0}^{\infty} (2n+1)\Lambda_{n,s} \frac{n!}{(2)_n} P_n^{(1,-1)}(1-2\delta^2)\Big). \tag{6.47}$$

6

Proof of (6.42). By the law of total expectation, we have

$$\mathbb{E}[D_t] = \mathbb{E}[D_t | \nu > t]\mathbb{P}[\nu > t] + \mathbb{E}[D_t | \nu \leq t]\mathbb{P}[\nu \leq t].$$

Since $\nu \sim \text{Geometric}(q)$, we have that

$$\mathbb{P}[\nu > t] = 1 - \mathbb{P}[\nu \leq t] = (1-q)^t.$$

Note additionally that $D_t = 1/2$ whenever $t \geq \nu$. Therefore

$$\mathbb{E}[D_t] = \mathbb{E}[D_t | \nu > t](1-q)^t + \tfrac{1}{2}\big(1-(1-q)^t\big) = \tfrac{1}{2} + \big(\mathbb{E}[D_t | t < \nu] - \tfrac{1}{2}\big)(1-q)^t.$$

We now calculate $\mathbb{E}[D_t | \nu > t]$ using (6.41) and the Bloch sphere representation:

$$\mathbb{E}[D_t | \nu > t] = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \Lambda_{n,t} \int_0^{\pi} 2\pi \sin\theta \sin\frac{\theta}{2} P_n(\cos\theta) \mathrm{d}\theta \tag{6.48}$$

$$= \sum_{n=0}^{\infty} \frac{2n+1}{2} \Lambda_{n,t} \int_{-1}^{1} \sqrt{\frac{1-x}{2}} P_n(x) \mathrm{d}x. \tag{6.49}$$

Also recall two facts about the Legendre polynomials: the recurrence relation in [215] states that

$$P_n(x) = \frac{1}{2n+1}\big(P'_{n+1}(x) - P'_{n-1}(x)\big), \tag{6.50}$$

and Rodrigues formula [216, (8.6.18)] states that

$$P_n(x) = \frac{1}{2^n n!} \frac{\mathrm{d}^n}{\mathrm{d}x^n} (x^2-1)^n. \tag{6.51}$$

Using (6.50), (6.51) and integration by parts, we then obtain

$$\int_{-1}^{1}\sqrt{\frac{1-x}{2}}P_n(x)\mathrm{d}x=\frac{1}{2n+1}\Big(-\int_{-1}^{1}\frac{P_{n+1}(x)}{2\sqrt{2-2x}}\,\mathrm{d}x+\int_{-1}^{1}\frac{P_{n-1}(x)}{2\sqrt{2-2x}}\,\mathrm{d}x\Big). \tag{6.52}$$

We have by [217, (12.4)] that the generating function of the Legendre polynomials is given by

$$\sum_{m=0}^{\infty}P_m(x)s^m=\frac{1}{\sqrt{1-2xs+s^2}}. \tag{6.53}$$

Based on (6.53) with $t=1$ and the orthogonality of Legendre polynomials,

$$\int_{-1}^{1}\frac{P_n(x)}{\sqrt{2-2x}}\,\mathrm{d}x=\int_{-1}^{1}P_n(x)\sum_{m=0}^{\infty}P_m(x)\,\mathrm{d}x=\sum_{m=0}^{\infty}\int_{-1}^{1}P_n(x)P_m(x)\,\mathrm{d}x=\frac{2}{2n+1}. \tag{6.54}$$

Here, we have used Lebesgue's dominated convergence theorem with $|P_n(x)|\leq 1\ \forall n$. Therefore, continuing from (6.49) using (6.52) and (6.54),

$$\begin{aligned}\mathbb{E}[D_t|v>t]&=\sum_{n=0}^{\infty}\frac{2n+1}{2}\Lambda_{n,t}\Big(-\int_{-1}^{1}\frac{P_{n+1}(x)}{2\sqrt{2-2x}}\,\mathrm{d}x+\int_{-1}^{1}\frac{P_{n-1}(x)}{2\sqrt{2-2x}}\,\mathrm{d}x\Big)\\&=\sum_{n=0}^{\infty}\frac{2n+1}{2}\Lambda_{n,t}\frac{-4}{(2n-1)(2n+1)(2n+3)}.\end{aligned} \tag{6.55}$$

Simplifying gives the result.

Proof of (6.44). Similar to above we have by the law of total probability that

$$\mathbb{P}[a\leq D_t\leq b]=\mathbb{P}[a\leq D_t\leq b|v\leq t]\mathbb{P}[v\leq t]+\mathbb{P}[a\leq D_t\leq b|v>t]\mathbb{P}[v>t],$$

and we note now that $\mathbb{P}[a\leq D_t\leq b|v\leq t]=\mathbb{1}[\frac{1}{2}\in[a,b]]$. Therefore

$$\mathbb{P}[a\leq D_t\leq b]=\mathbb{1}[\tfrac{1}{2}\in[a,b]]\big(1-(1-q)^t\big)+\mathbb{P}[a\leq D_t\leq b|v>t](1-q)^t. \tag{6.56}$$

We now calculate $\mathbb{P}[a\leq D_t\leq b|v>t]$; again using (6.41). Let $0\leq a\leq b\leq 1$. From the equivalence of the events

$$\big\{a\leq D_t\leq b\big\}=\big\{2\arcsin(a)\leq\Theta_t\leq 2\arcsin(b)\big\},$$

where Θ_t denotes the polar angle of R_t, it follows that

$$\mathbb{P}\left[a\leq D_t\leq b\right]=\big(1-(1-q)^t\big)\mathbb{1}[\tfrac{1}{2}\in[a,b]] \tag{6.57}$$

$$+(1-q)^t\sum_{n=0}^{\infty}\frac{2n+1}{4\pi}\Lambda_{n,t}\int_{2\arcsin a}^{2\arcsin b}2\pi\sin\theta P_n(\cos\theta)\mathrm{d}\theta. \tag{6.58}$$

Now let $0\leq\delta\leq 1$. Continuing from (6.58), since $\cos(2\arcsin\delta)=1-2\delta^2$ for $\delta\in[0,1]$ and letting $\cos\theta=x$,

$$\mathbb{P}\left[D_t\leq\delta|v>t\right]=\sum_{n=0}^{\infty}\frac{2n+1}{4\pi}\Lambda_{n,t}\int_{0}^{2\arcsin\delta}2\pi\sin\theta P_n(\cos\theta)\mathrm{d}\theta \tag{6.59}$$

$$=\sum_{n=0}^{\infty}\frac{2n+1}{2}\Lambda_{n,t}\int_{1-2\delta^2}^{1}P_n(x)\mathrm{d}x. \tag{6.60}$$

By the explicit representation of Rodrigues' formula [216, (8.6.18)],

$$\mathbb{P}\left[D_t \leq \delta | \nu > t\right] = \sum_{n=0}^{\infty} \frac{2n+1}{2} \Lambda_{n,t} \int_{1-2\delta^2}^{1} \sum_{k=0}^{n} \binom{n}{k} \binom{n+k}{k} \left(\frac{x-1}{2}\right)^k \mathrm{d}x \tag{6.61}$$

$$= \sum_{n=0}^{\infty} (2n+1) \Lambda_{n,t} \sum_{k=0}^{n} \binom{n}{k} \binom{n+k}{k} \frac{(-1)^k}{k+1} \delta^{2(k+1)}. \tag{6.62}$$

Finally, let $r = k+1$, such that

$$\mathbb{P}\left[D_t \leq \delta | \nu > t\right] = \sum_{n=0}^{\infty} (2n+1) \Lambda_{n,t} \sum_{r=1}^{n+1} \binom{n}{r-1} \binom{n+r-1}{r-1} \frac{(-1)^{r-1}}{r} \delta^{2r} \tag{6.63}$$

$$= \sum_{n=0}^{\infty} (2n+1) \Lambda_{n,t} \sum_{r=1}^{n+1} (-1)^{r-1} \delta^{2r} C_{r-1} \binom{n+r-1}{2(r-1)}.$$

Proof of (6.47). This follows directly after applying De Morgan's law and Boole's inequality, i.e.,

$$\mathbb{P}[\max_{0 \leq s \leq t} D_s \leq \delta | \nu > t] = \mathbb{P}\left[\bigcap_{s=0}^{t} \left\{D_s \leq \delta\right\} \middle| \nu > t\right] \tag{6.64}$$

$$= \mathbb{P}\left[\left(\bigcup_{s=0}^{t} \{D_s > \delta\}\right)^c \middle| \nu > t\right] = 1 - \mathbb{P}\left[\bigcup_{s=0}^{t} \{D_s > \delta\} \middle| \nu > t\right]$$

$$\geq 1 - \sum_{s=0}^{t} \mathbb{P}[D_s > \delta | \nu > t] = 1 - t + \sum_{s=0}^{t} \mathbb{P}[D_s \leq \delta | \nu > t]. \tag{6.65}$$

6.4. SIMULATIONS

In this section, we investigate and validate our results numerically. This section also serves to illustrate the models. We also compare our results to the following traditional error calculation and fitting method.

Fit method using just a depolarizing channel First, one readily calculates the expected trace distance of a depolarizing quantum channel [3, p. 378]

$$\rho_0 \to \rho_1 = \mathscr{E}(\rho_0) = \frac{\mu}{2} I + (1-\mu)\rho_0 \quad \text{w.p. 1} \tag{6.66}$$

when repeated $t \in \mathbb{N}_+$ times as a function of its decay parameter $\mu \in [0,1]$. To see how, note that after t applications of this depolarizing channel, the quantum state would be $\mathscr{E}^t(\rho) = \frac{1}{2}(1-(1-\mu)^t) I + (1-\mu)^t \rho$ w.p. one. The trace distance after t depolarizing channels is thus

$$D_t = \tfrac{1}{2}(1-(1-\mu)^t) \quad \text{w.p. 1.} \tag{6.67}$$

Next, one fits (6.67) to experimental or numerical data using e.g. the method of least squares. This curve follows the data as well as it can (but not necessarily perfect) and the corresponding fit parameter μ^{fit} is returned.

It is insightful to consider the difference between (6.66), (6.67) and the result in Proposition 6. Proposition 6 namely models a different type of error channel, specifically one in which the qubit can depolarize at each step according to a classical probability $\mu \in [0,1]$. Substituting $\Lambda_{n,t} = 0$ for all n, t so that the random perturbations of the model in Section 6.3.2 are neglected and only depolarization is included, this model tells us that $\mathscr{E}^t(\rho)$ equals either ρ w.p. $(1-\mu)^t$ or $I/2$ w.p. $1-(1-\mu)^t$. Consequently, under this model,

$$D_t = \begin{cases} 0 & \text{w.p. } (1-\mu)^t \\ \frac{1}{2} & \text{w.p. } 1-(1-\mu)^t, \end{cases} \qquad \mathbb{E}[D_t] = \tfrac{1}{2}(1-(1-\mu)^t). \tag{6.68}$$

6.4.1. ERROR ACCUMULATION IN RANDOMIZED BENCHMARKING

We will first consider error accumulation in single-qubit randomized benchmarking. In each randomized benchmarking simulation experiment, the initial state is set to $|1\rangle$ and subsequently $\tau - 1$ gates are selected one by one from the Clifford group $\mathscr{C}_1$ uniformly at random. Finally, based on the experimental setup in [180], we add a τ-th gate that transfers the state to $|0\rangle$ in the absence of errors. For simplicity we specify $d = 1$ and thus discuss the trace distance throughout this section.

PAULI AND CLIFFORD CHANNEL ERRORS

We consider two kinds of error models: Pauli channels and Clifford channels. For the Pauli channel model, let the probability that no noise occurs be $\mathbb{P}(\Lambda = I) = 1 - r$ and the probabilities of every noise type occurring be $\mathbb{P}(\Lambda = X) = \mathbb{P}(\Lambda = Y) = \mathbb{P}(\Lambda = Z) = r/3$, where $r \in [0,1]$. For the Clifford channel model, let the probability of no noise occurring be $\mathbb{P}(\Lambda = I) = 1 - r$ and the probabilities of every other gate type occurring equal $r/23$. In Figure 6.4 the parameter r is set to $1/100$. Two error thresholds δ are considered: $\delta = 1/10$ (a, c and d) and $\delta = 1/5$ (b). The insets show the influence of parameter r on the probability of error in (6.10) and the probability of maximum error in (6.11) at time $t = 100$. The results in Figure 6.4 illustrate the theoretical results for the probability of error (6.10), the expectation of the trace distance and the probability of maximum error (6.11) and their validity is supported by these simulations. Figure 6.4 also illustrates that different error models lead to different error accumulation behaviors. The two sample curves in Figure 6.4a and Figure 6.4b (the solidly drawn step functions) show the trace distance D_t between the faultless state ρ_t and the faulty state σ_t in two independent randomized benchmarking experiments. The dashed lines indicate our fits of (6.67) to the sample average of the numerical data. Note that the numerical sample average

of the trace distance in Figure 6.4a can be fitted perfectly – this is because under the present assumptions, the trace distance here is in fact geometrically distributed. The case depicted in Figure 6.4b is however different and does not satisfy a simple geometric distribution and we can see that the traditional fit method disagrees in the limit. This is because we are dealing with two different error models.

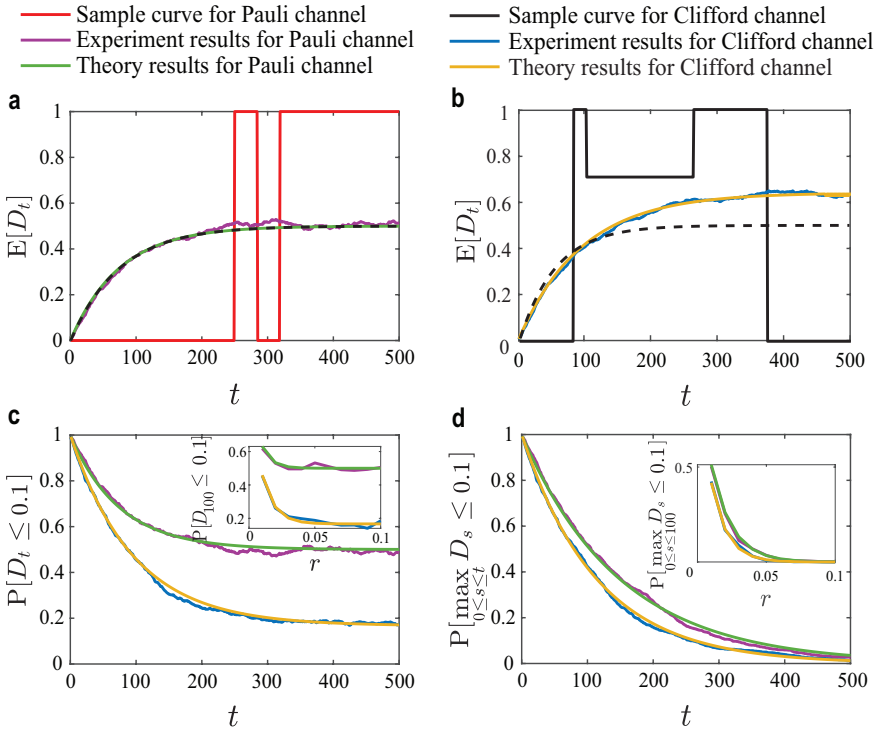


Figure 6.4: The error accumulation based on Pauli and Clifford channels in randomized benchmarking. Two error thresholds δ are considered, $\delta = 1/10$ (figures a, c, d) and $\delta = 1/5$ (figure b). The simulation results are calculated from 1000 independent randomized benchmarking experiments. The dashed, black curves are fits of (6.67) to the sample averages of the numerical data. The resulting fit parameters are $\mu^{\text{fit}} \approx 0.013$ (figure a) and $\mu^{\text{fit}} \approx 0.018$ (figure b), respectively.

INFLUENCE OF THE INITIAL STATE ON PAULI ERROR ACCUMULATION

In this section we consider the influence of the initial state on Pauli error accumulation. We ignore the last gate of randomized benchmarking for simplicity. Each gate is selected one by one from the Pauli group uniformly at random. The error model described above is considered again and the parameter r is set to $1/5$.

Figure 6.5 shows the state transition diagram for two different initial states: $|\zeta_0\rangle = \sqrt{7/10}|0\rangle + \sqrt{3/10}|1\rangle$ and $|\xi_0\rangle = \sqrt{4/5}|0\rangle + \sqrt{1/5}|1\rangle$. Recall that for the Pauli group of a single qubit, there are in total $|\mathscr{P}|^2 = 16$ state pairs, which correspond to the sixteen nodes depicted in Figure 6.5. More precisely, each of the nodes represents one of the 16 two-dimensional states $\{(I,I),(I,X),(I,Y),(I,Z),(X,I),...,(Z,Z)\}$. The initial state pair (ρ_0,σ_0), which here satisfies $\rho_0 = \sigma_0$, corresponds to state 1 in Figure 6.5. The bad state pairs that constitute $\mathscr{B}^{|\zeta_0\rangle}_{|\zeta_0\rangle,\delta}$ and $\mathscr{B}^{|\xi_0\rangle}_{|\xi_0\rangle,\delta}$, which have a trace distance over $\delta = 1/5$, are indicated in red. Each edge depicts the possibility of the two-dimensional Markov chain to jump between the two connected nodes. Note that the number of bad state pairs can be affected by the choice of initial state. Figure 6.6 shows the probability of maximum error in (6.11) and the maximum number of tolerant gates in (6.12) for the same two different initial states: $|\zeta_0\rangle$ (upper) and $|\xi_0\rangle$ (bottom). Figure 6.5 and Figure 6.6 illustrate too that the choice of initial state can affect the probability $\mathbb{P}[\max_{0\le s\le t} D_s > \delta]$ and the maximum number of tolerant gates $t^{\star}_{\delta,\gamma}$. Finally, when starting from the initial state $|\zeta_0\rangle$, in this simple case, (6.11) reduces to

$$\mathbb{P}[\max_{0\le s\le t} D_s > 1/5] = 1 - \left(1 - \tfrac{2}{3}r\right)^t,$$

while when starting from the initial state $|\xi_0\rangle$ we have

$$\mathbb{P}[\max_{0\le s\le t} D_s > 1/5] = 1 - (1-r)^t.$$

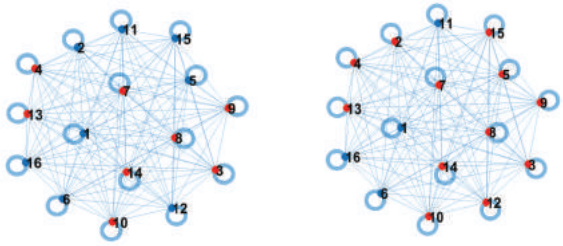


Figure 6.5: State transition diagram for different initial states: $|\zeta_0\rangle$ (left) and $|\xi_0\rangle$ (right) and the error threshold $\delta = 1/5$. The red nodes show the bad state pairs in $\mathscr{B}^{|\zeta_0\rangle}_{|\zeta_0\rangle,\delta}$ and $\mathscr{B}^{|\xi_0\rangle}_{|\xi_0\rangle,\delta}$, respectively, in which the trace distances are larger than δ.

6.4.2. ERROR ACCUMULATION IN NONRANDOM CIRCUITS

Here we illustrate error accumulation rates in two nonrandom circuits. The first is a periodical single-qubit circuit that repeats a Hadamard, Pauli-X, Pauli-Y and Pauli-Z gate $k = 25$ times and the second a two-qubit circuit that is repeated $k = 5$ times; see

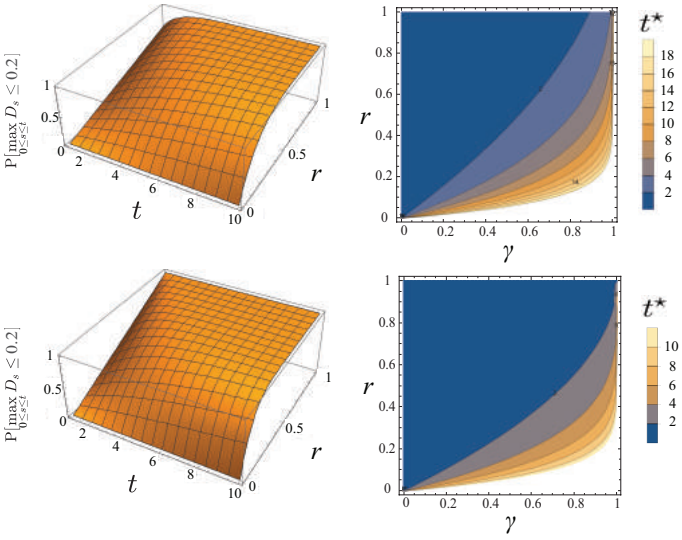


Figure 6.6: Pauli channel error accumulation on single-qubit randomized benchmarking when starting from different initial states: $|\zeta_0\rangle$ (top) and $|\xi_0\rangle$ (bottom). The error threshold is set to $\delta = 1/5$.

6

also Figure 6.7. Here the controlled-NOT gate CNOT $= \big((1,0,0,0);(0,1,0,0);(0,0,0,1);(0,0,1,0)\big)$. Consider also the following two error models in which the errors depend on the gates:

(i) For the single-qubit circuit, presume $\mathbb{P}[\Lambda = I] = 0.990, \mathbb{P}[\Lambda = Z] = 0.010$.

(ii) For the two-qubit circuit, when labeling the qubits by A and B, suppose

$$\mathbb{P}[\Lambda_A = I] = 0.990,\ \mathbb{P}[\Lambda_A = X] = 0.006, \mathbb{P}[\Lambda_A = Y] = 0.003,\ \mathbb{P}[\Lambda_A = Z] = 0.001;$$
$$\mathbb{P}[\Lambda_B = I] = 0.980,\ \mathbb{P}[\Lambda_B = X] = 0.002, \mathbb{P}[\Lambda_B = Y] = 0.014,\ \mathbb{P}[\Lambda_B = Z] = 0.004. \qquad (6.69)$$

In order to evaluate Proposition 3, we set the error threshold $\delta = 1/10$.

The theoretical and simulation results on the two circuits are shown in Figure 6.7. Note that the simulation curves almost coincide with the theoretical curves; the deviation is only due to numerical limits. Furthermore, because different gates influence error accumulation to different degrees, the periodical ladder shape occurs in Figure 6.7. Observe furthermore that this periodical ladder shape is not captured by the fit method that only takes into account the decay of t applications of a single depolarizing channel.

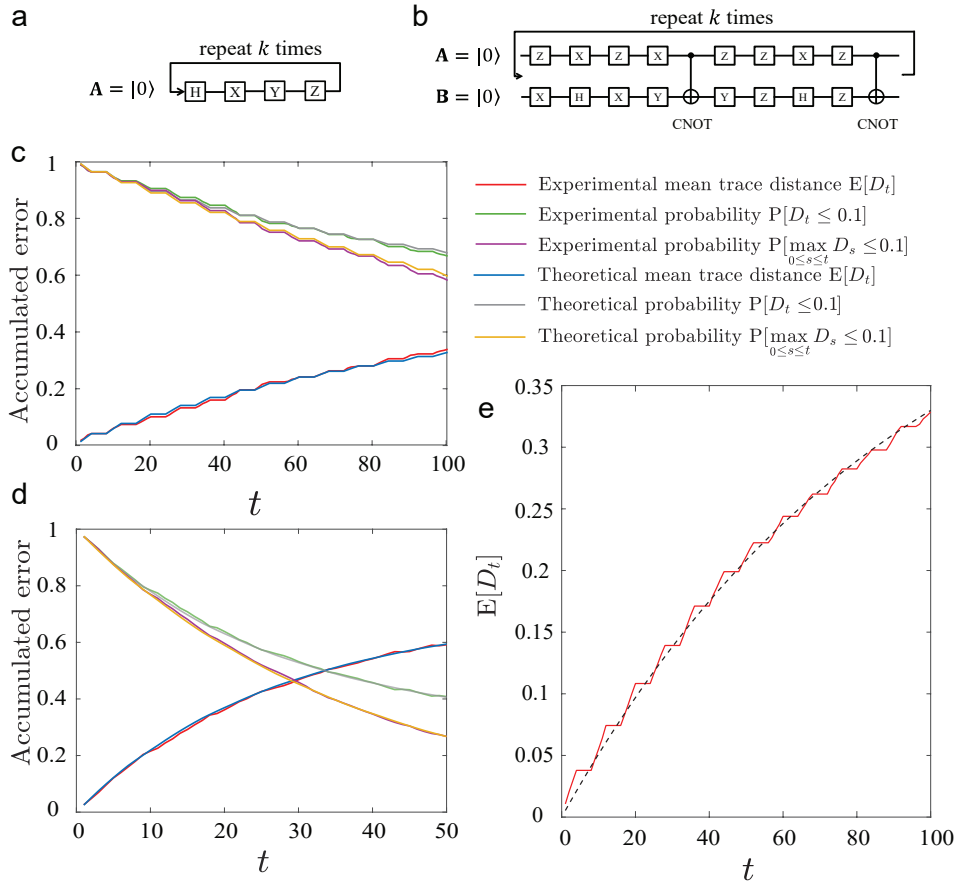


Figure 6.7: Theoretical and simulation results for error accumulation on a single-qubit circuit (figures a, c and e) and a two-qubit circuit (figures b and d). The numerical results are calculated from 2000 independent runs and almost indistinguishable from the formulae. The dashed, black curve in figure e is a fit of (6.67) to the data. The fit parameter is $\mu^{\text{fit}} \approx 0.011$.

6.4.3. CONTINUOUS, RANDOM ERROR ACCUMULATION IN A SINGLE QUBIT

We now simulate the accumulation of continuous errors without depolarization ($q = 0$) in a single qubit. Here, the noise is assumed to lead to a random walk on the Bloch sphere that takes steps of a fixed angle $\alpha = 1/10$ and therefore $p_t(\alpha) = \delta(\alpha)$. The threshold δ is set to be $1/10$. The theoretical mean trace distance $\mathbb{E}[D_t]$ and probability $\mathbb{P}[D_t \leq \delta]$ are calculated using (6.42) and (6.44). The theoretical results and simulations are shown in Figure 6.8. Note again that the traditional fit method disagrees at large t: this happens here because $\alpha \neq 0$.

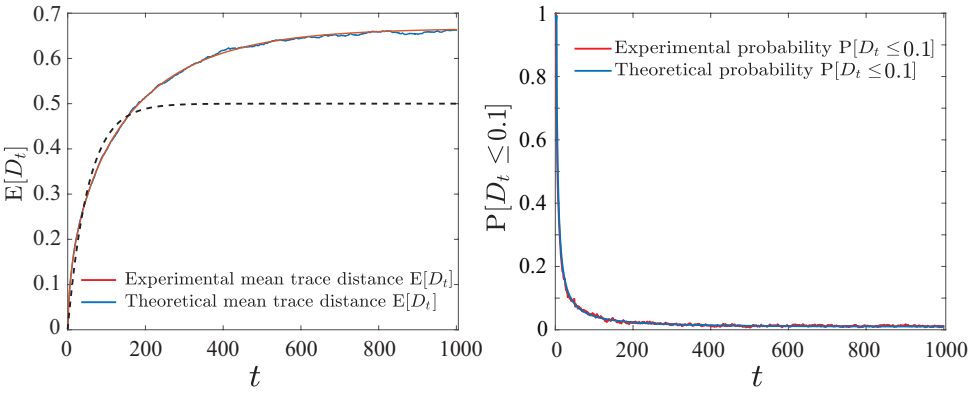


Figure 6.8: Continuous error accumulation in one qubit. The numerical results are from 2000 independent runs of our simulation. The dashed, black curve in the left figure is a fit of (6.67) to the data. The resulting fit parameter is $\mu^{\text{fit}} \approx 0.019$.

6.5. MINIMIZING ERRORS IN QUANTUM CIRCUIT THROUGH OPTIMIZATION

The rate at which errors accumulate may be different for different quantum circuits that can implement the same algorithm. Using techniques from optimization and (6.30), we can therefore search for the quantum circuit that has the lowest error rate accumulation while maintaining the same final state. To see this, suppose we are given a circuit $\mathscr{U}_\tau = \{U_1, U_2, \ldots, U_\tau\}$. For given ρ_0 this brings the quantum state to some quantum state ρ_τ. Other circuits may go to the same final state and have a lower probability of error at time τ. We will therefore aim to

$$\begin{aligned} &\underset{G_1,\ldots,G_\tau \in \mathscr{G}_n}{\text{minimize}} && u(\{G_1,\ldots,G_\tau\}) \\ &\text{subject to} && G_\tau \cdots G_1 = U_\tau \cdots U_1. \end{aligned} \tag{6.70}$$

Here, one can for example choose for the objective function $u(\cdot)$ the probability of error (6.29), or probability of maximum error (6.30). To solve (6.70), we design a simulated annealing algorithm in Section 6.5.1 to improve the quantum circuit.

The minimization problem in (6.70) is well-defined and has a few attractive features. For starters, the minimization problem automatically detects shorter circuits if the probability of error when applying the identity operator $I^{\otimes n}$ is relatively small. The optimum may then for example occur at a circuit of the form

$$G_\tau G_{\tau-1} G_{\tau-2} \cdots G_2 G_1 = I^{\otimes n} G_{\tau-1} I^{\otimes n} \cdots I^{\otimes n} G_1, \tag{6.71}$$

which effectively means that only the two gates $G_{\tau-1}G_1$ are applied consecutively. The identity operators in this solution essentially describe the passing of time. Now, critically, note that while the minimization problem does consider all shorter circuits of depth at most τ, this does not necessarily mean that the physical application of one specific group element $G \in \mathscr{G}_n$ is always the best. Concretely, in spite of the fact that any quantum circuit of the form $G_\tau \cdots G_1 = G \in \mathscr{G}_n$ performs the single group element $G \in \mathscr{G}_n$, it is not necessarily true that

$$u(\{G, I^{\otimes n}, \ldots, I^{\otimes n}\}) < u(\{G_1, \ldots, G_\tau\}). \tag{6.72}$$

The reason for this is that the error distribution on the direct group element G may be worse than using a circuit utilizing multiple other group elements. In other words, the optimal circuit need not always be the 'direct' circuit, but of course it can be. (In Section 6.5.2 we also consider the situation in which an experimentalist can only apply a subset $\mathscr{A} \subseteq \mathscr{G}_n$ that need not necessarily be a group and in such a case the direct group element G may not even be a viable solution to the experimentalist if $G \notin \mathscr{A}$.) Typically, the minimization problem will prefer shorter circuits if the probability of error when

applying the identity operator $I^{\otimes n}$ is relatively small and the error distributions of all gate distributions are relatively homogeneous.

6.5.1. SIMULATED ANNEALING

We will generate candidate circuits as follows. Let $\{G_1^{[\eta]},\dots,G_\tau^{[\eta]}\}$ denote the circuit at iteration η. Choose an index $I \in [\tau-1]$ uniformly at random, choose $G \in \mathscr{G}$ uniformly at random. Then set

$$G_i^{[\eta+1]} = \begin{cases} G & \text{if } i = I, \\ G_{I+1}^{[\eta]} G_I^{[\eta]} G^{\leftarrow} & \text{if } i = I+1, \\ G_i^{[\eta]} & \text{otherwise.} \end{cases} \tag{6.73}$$

Here, $G^{\leftarrow}$ denotes the (left) inverse group element, i.e., $G^{\leftarrow}G = I^{\otimes n}$. The construction thus ensures that

$$G_{I+1}^{[\eta+1]} G_I^{[\eta+1]} = \left(G_{I+1}^{[\eta]} G_I^{[\eta]} G^{\leftarrow}\right) G = G_{I+1}^{[\eta]} G_I^{[\eta]} \tag{6.74}$$

so that the circuit's intent does not change: $G_\tau^{[\eta+1]} \cdots G_1^{[\eta+1]} = G_\tau^{[\eta]} \cdots G^{[\eta]}$.

We will use the Metropolis algorithm. Let

$$E = \left\{\{G_1,\dots,G_\tau\} | G_\tau \cdots G_1 = U_\tau \cdots U_1\right\} \tag{6.75}$$

denote the set of all viable circuits. For two arbitrary circuits $i, j \in E$, let

$$\Delta(i,j) \triangleq \sum_{s=1}^{\tau-1} \mathbb{1}[i_s \neq j_s, i_{s+1} \neq j_{s+1}] \tag{6.76}$$

denote the number of consecutive gates that differ between both circuits. Under this construction, the *candidate-generator matrix* of the Metropolis algorithm is given by

$$q_{ij} = \begin{cases} \frac{1}{(\tau-1)|\mathscr{G}|} & \text{if } \Delta(i,j) \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{6.77}$$

Since the candidate-generator matrix is symmetric, this algorithm means that we set $\alpha_{i,j}(T) = \exp\left(-\frac{1}{T}\max\{0, u(j)-u(i)\}\right)$ as the *acceptance probability* of circuit j over i. Here $T \in (0,\infty)$ is a positive constant. Finally, we need a cooling schedule. Let $M \triangleq \sup_{\{i,j\in E|\Delta(i,j)\leq 1\}}\{u(j)-u(i)\}$. Based on [179], if we choose a cooling schedule $\{T_\eta\}_{\eta\geq 0}$ that satisfies $T_\eta \geq \frac{\tau M}{\ln \eta}$, then the Metropolis algorithm will converge to the set of global minima of the minimization problem in (6.70).

Lemma 7. *Algorithm 2 converges to the global minimizer of* (6.70) *whenever* $T_\eta \geq \tau M/\ln\eta$ *for* $\eta = 1,2,\cdots$.

Algorithm 2: Pseudo-code for the simulated annealing algorithm described in Section 6.5.1.

Input: A group $\mathscr{G}$, a circuit $\{U_1,\dots,U_\tau\}$ and number of iterations w
Output: A revised circuit $\{G_1^{[w]},\dots,G_\tau^{[w]}\}$

begin
  Initialize $\{G_1^{[0]},\dots,G_\tau^{[0]}\} = \{U_1,\dots,U_\tau\}$;
  for $\eta \leftarrow 1$ **to** w **do**
    Choose $I \in [\tau-1]$ uniformly at random;
    Choose $G \in \mathscr{G}$ uniformly at random;
    Set $J_I = G, J_{I+1} = G_{I+1}^{[\eta]} G_I^{[\eta]} G^{\leftarrow}, J_i = G_i^{[\eta]}\ \forall_{i\neq I,I+1}$;
    Choose $X \in [0,1]$ uniformly at random;
    if $X \leq \alpha_{G^{[\eta]},J}(T_\eta)$ **then**
      Set $G^{[\eta+1]} = J$;
    else
      Set $G^{[\eta+1]} = G^{[\eta]}$;
    end
  end
end

6.5.2. EXAMPLES

GATE-DEPENDENT ERROR MODEL

We are going to improve the one-qubit circuit in Figure 6.7 using Algorithm 2. The gates are limited to the Clifford group $\mathscr{C}_1$ and the errors will be limited to the Pauli channel. The error probabilities considered here are gate-dependent and can be found in Appendix E.6. The cooling schedule used here will be set as $T_\eta = C/\ln(\eta+1)$ and the algorithm's result when using $C = 0.004$ is shown in Figure 6.9. Figure 6.9 illustrates that the improved circuit can indeed lower the error accumulation rate. The circuit with the lowest error accumulation rate that was found is shown in Appendix E.9.

GATES IN A SUBSET OF ONE GROUP

The gates that are available in practice may be restricted to some subset $\mathscr{A} \subseteq \mathscr{G}$ not necessarily a group. Under such constraint, we could generate candidate circuits as follows: Let $\{G_1^{[\eta]},\dots,G_\tau^{[\eta]}\}$ denote the circuit at iteration η. In each iteration, two neighboring gates will be considered to be replaced by two other neighboring gates. There are $m \leq (\tau-1)$ neighboring gate pairs $(G_1^{[\eta]},G_2^{[\eta]}),\dots,(G_{m-1}^{[\eta]},G_m^{[\eta]})$ that can be replaced by two different neighboring gates. Choose an index $I \in [m-1]$ uniformly at random and replace $(G_I^{[\eta]},G_{I+1}^{[\eta]})$ by any gate pair from $\{(\tilde{G}_1,\tilde{G}_2) \in \mathscr{A}^2 \mid G_I^{[\eta]} G_{I+1}^{[\eta]} = \tilde{G}_1\tilde{G}_2\}$ uniformly at random. Pseudo-code for this modified algorithm can be found in Appendix E.8. It must be noted that this algorithm is not guaranteed to converge to the global minimizer of (6.70) (due to limiting the gates available); however, it may still find use in practical scenarios where

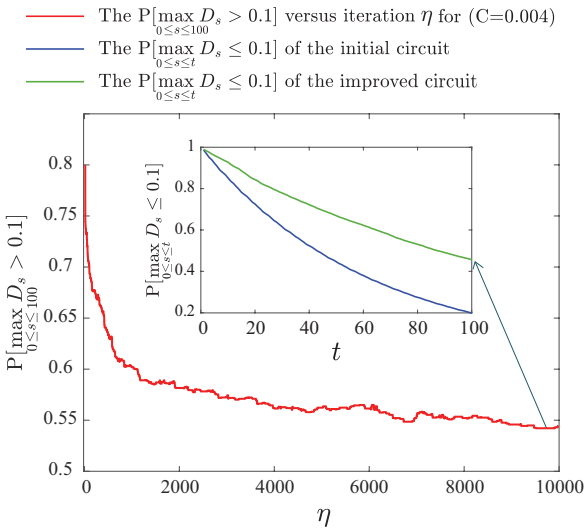


Figure 6.9: Circuit optimization when using Algorithm 2. The error probabilities are gate-dependent. Note that the probability of maximum error (6.30) decreases as the number of iterations η increases when using Algorithm 2 ($C = 0.004$). Here we started from the one-qubit circuit in Figure 6.7.

6

one only has access to a restricted set of gates.

We now aim to decrease the probability of maximum error (6.30) by changing the two-qubit circuit shown in Figure 6.7. The error model is the same as that in Section 6.4–B. The set of gates available for improving the circuit is here limited to $\{I, X, Y, Z, H, CNOT\}$. The result here for the two-qubit circuit is obtained by again using the cooling schedule $T_\eta = C/\ln(\eta+1)$ but now letting the parameter $C = 0.002$. Figure 6.10 shows that a more error-tolerant circuit can indeed be found using this simulated annealing algorithm. The improved circuit is shown in Appendix E.9.

DEUTSCH–JOZSA ALGORITHM

Let us give further proof of concept through the Deutsch–Jozsa Algorithm for one classical bit [200, 201]. This quantum algorithm determines if a function $f : \{0,1\} \to \{0,1\}$ is constant or balanced, i.e., if $f(0) = f(1)$ or $f(0) \neq f(1)$. It is typically implemented using the quantum circuit in Figure 6.11. If no errors occur in this quantum circuit, then the first qubit would measure $|0\rangle$ or $|1\rangle$ w.p. one if f constant or balanced, respectively. If errors occur in this quantum circuit, then there is a strictly positive probability that the first qubit measures $|1\rangle$ or $|0\rangle$ in spite of f being constant or balanced, respectively and thus for the algorithm to incorrectly output that f is constant or balanced. This *misclassification probability* v of the algorithm depends on the underlying error distributions and can be

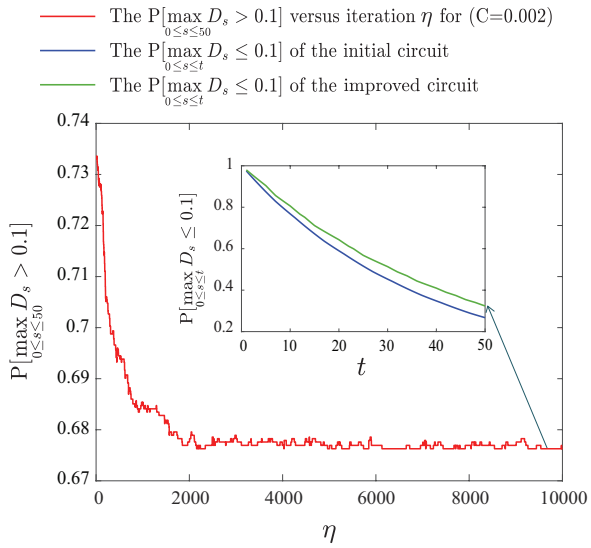


Figure 6.10: Circuit optimization when using Algorithm 6. The set of gates available is chosen limited to $\{I, X, Y, Z, H, CNOT\}$. Note that the probability of maximum error (6.30) decreases as the number of iterations η increases when using Algorithm 6 ($C = 0.002$). Here we started from the two-qubit circuit shown in Figure 6.7.

6

calculated by adapting (6.29)'s derivation.

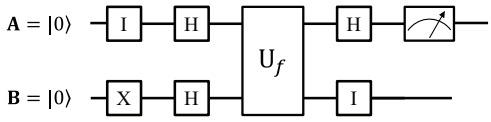


Figure 6.11: The Deutsch–Jozsa Algorithm for one classical bit in quantum circuit form.

We suppose now that errors occur according to a distribution in which two-qubit Clifford gates are more error prone than single-qubit gates, see Appendix E.10 for the details. We can then revise the quantum circuit in Figure 6.11 using a simulated annealing algorithm in Appendix E.11 that aims at minimizing (6.70) by randomly swapping out poor gate pairs for better gate pairs. This simulated annealing algorithm, like any other, is sensitive to the choice of *cooling schedule* [179], here set as $T_\eta = C\big(\gamma/\eta + (1-\gamma)/\ln(\eta+1)\big)$ with $C > 0$, $\gamma \in [0,1]$; the integer η indexes the iterations. Figure 6.12 shows the ratio $\Theta \triangleq \nu_{\text{original circuit}}/\nu_{\text{revised circuit}}$ as a function of C, γ for $f_a(x) = x$, $f_b(x) = 1-x$, $f_c(x) = 0$, $f_d(x) = 1$ where $x \in \{0,1\}$. Note that $\Theta \geq 1$ always, ≥ 1.60 commonly and sometimes even ≥ 2.20.

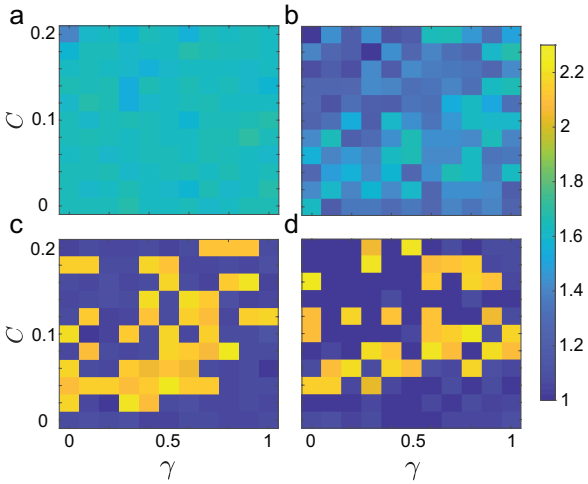


Figure 6.12: For every pair (C,γ) here, Θ was calculated using a Monte Carlo simulation with 10^5 independent repetitions for the best circuit found throughout $w = 10^3$ iterations of the annealing algorithm. This simulated annealing algorithm, like any other, is sensitive to the choice of *cooling schedule* [179], here set as $T_\eta = C\big(\gamma/\eta + (1-\gamma)/\ln(\eta+1)\big)$ with $C > 0$, $\gamma \in [0,1]$; the integer η indexes the iterations. Panels a-d respectively show the ratio $\Theta \triangleq \nu_{\text{original circuit}}/\nu_{\text{revised circuit}}$ as a function of C,γ for $f_a(x) = x$, $f_b(x) = 1-x$, $f_c(x) = 0$, $f_d(x) = 1$ where $x \in \{0,1\}$. $u(\cdot)$ was set to the misclassification probability for a, c; and to (6.30) for b, d.

6.6. CONCLUSION

In conclusion, we have proposed and studied a model for discrete Markovian error accumulation in a multi-qubit quantum computation, as well as a model describing continuous errors accumulating in a single qubit. By modeling the quantum computation with and without errors as two coupled Markov chains, we were able to capture a weak form of time-dependency, allow for fairly generic error distributions and describe multi-qubit systems. Furthermore, by using techniques from discrete probability theory, we could calculate the probability that error measures such as the fidelity and trace distance exceed a threshold analytically. To combat the numerical challenge that may occur when evaluating our expressions, we additionally provided an analytical bound on the error probabilities that is of lower numerical complexity. Finally, we showed how our expressions can be used to decide how many gates one can apply before too many errors accumulate with high probability and how one can lower the rate of error accumulation in existing circuits by using techniques from optimization.

7

CONCLUSION

7.1. MAIN CONTRIBUTIONS

This dissertation addresses the challenge of inferring network properties based on the population-based observations, the application of epidemic models and reporting delays in the forecast of the COVID-19 pandemic and reduce mortality, the way to approximate the Markovian SIS epidemic processes in complex networks and the method to model the error accumulation in quantum circuits and lower the rate of error accumulation in existing circuits through simulated annealing.

In Chapter 2, we conceptually evaluate the feasibility and limitation to deduce network properties based on prevalence data from the SIS model, which has an important guiding significance for future works on inferring network properties based on real prevalence data. Even if the full knowledge of epidemic states (the state time-series of each node) is available, the exact network reconstruction for large networks still seems infeasible since the maximum-likelihood network reconstruction for SIS processes is NP-hard. If we just have the prevalence, the rough network reconstruction is only workable for very small networks. For large networks, there can be different networks with very similar prevalence curves and the exact reconstruction is impossible. However, our results show that it is feasible to estimate some network properties for large networks. The network rewiring algorithm SARA is mainly to identify the network type. If the network type is known, one can generate multiple graphs with the same network type. The graph whose prevalence is close to the benchmark can be further selected. The metrics of the above-selected graph can thereby estimate the metrics of the underlying network that are sensitive to the

prevalence.

In Chapter 3, we find a conflict between the COVID-19 data and the compartmental epidemic models, indicating reporting delays in real data. We further modeled the reporting delays and proposed a correlation-based method to infer the reporting delays for different countries. Significant reporting delays of recoveries are discovered in real data for several countries. We finally forecast the pandemic trends by considering the reporting delays to improve accuracy.

In Chapter 4, we discover that reducing connections between two populations can delay the death curve but cannot reduce the final mortality. We propose a merged SIR model, which advises elderly individuals to interact less with their non-elderly connections at the initial stage but interact more with their non-elderly relationships later, to reduce the final mortality. Finally, immunizing elderly hub individuals can also significantly decrease mortality.

In Chapter 5, we propose a spectral clustering SIS approximation (SCSA) method, which combines the spectral clustering of the huge Markov graph and the birth-and-death approximation, to reduces the huge 2^N state space of the Markov chain to a smaller number of states. We discover that the relationship between the approximation error ϵ and the number of clusters c roughly obeys $\epsilon \sim c^{-\alpha}$, where $\alpha \in (\frac{1}{4}, \frac{4}{5})$. The exponent α tends to be larger if the network has a higher link density. Besides, the approximation error decreases faster for networks with higher randomness. These rules can be applied to roughly decide how much state space is required to make sure the approximation error is smaller than a threshold based on the features of a specific network.

The results in Chapter 6 cover (and go beyond the state-of-the-art analytical results for) the commonly-used randomized benchmarking protocol. Our bridging of techniques from probability theory and operations research to the domain of quantum computing looks to be a new angle. Finally, it opens up the exciting yet challenging avenue of "quantum operations research": a thus far undeveloped area that will be necessary to achieve practical quantum computing. Essentially, we can compute more and for longer with present technologies by combining operations research with quantum computing.

7.2. DIRECTIONS FOR FUTURE WORK

There are many interesting questions that can be further investigated based on this dissertation in future works:

– Chapter 2 attempts to infer the network properties only based on the prevalence data. In the future, additional known knowledge, e.g., population distribution, may be available and helps the inference of the network properties.

– Chapter 3 reveals that the reporting delays of infections and recoveries in many coun-

tries are significant, largely increasing the difference between the reported data and the epidemic models. It is promising that the forecast performance of many advanced model-based forecast methods, e.g., Kalman filter [218], the Network-Inference-Based Prediction Algorithm (NIPA) [7, 219], meta-population or agent-based network models [220], can all be improved by considering the reporting delays.

– Based on our results in Chapter 3 and previous studies, the death data is usually more timely than the infected and recovered data. However, there are more fluctuations in real death data since the number of daily deceased cases in many regions is very small. Thus the best way to forecast the pandemic is to apply all infection, recovery and death data and considering the reporting delays of each data.

– Our work in Chapter 3 suffers from some restrictions in real data. It would be valuable to collect more accurate and comprehensive real data about the reporting delays. Nowadays, the real published data about delays, especially the delays of recoveries, are rare. Although some countries published data about delays, the accuracy of these data is still questionable since it is difficult to accurately measure the time point that a person is infected or recovered in reality.

– Future works based on Chapter 4 may consider real contact networks, real population flow networks, multi-layer networks and time-varying networks in modeling the spreading processes.

– Chapter 5 studies the Markovian SIS epidemics on complex networks. However, in realistic the spreading processes could be non-Markovian. Thus it would be valuable to investigate whether the Poisson processes can well describe the realistic infection and curing processes.

– For large networks, SCSA cannot be applied because the computation time is too large. It would be valuable to study the way to reduce the computational complexity of SCSA in the future so that our analysis can be extended to many realistic networks. Here, we provide three intriguing ideas:

- Use fast spectral clustering [221] to reduce the computational complexity.
- Use the similarity-based algorithms based on the local structures, e.g., common neighbor, to cluster the infinitesimal generator matrix Q efficiently.
- Instead of directly clustering the infinitesimal generator matrix Q, find metrics related to the underlying network that correlated with the spectral clustering results. Gerrit and Luca [222] proposed a method to group the states based on counting nodes and counting edges. We further did related researches from the view of the cut-set [32, 223, 224] (see details in Appendix D.7).

– Chapter 5 mainly considered the error accumulation on finite state space. The accumulation of errors when using a universal gate set would need to be modeled using stochastic processes that live on infinite state spaces. Such an approach looks to be connected to the modeling of random walks on manifolds.
– The expressions in (6.11) and (6.30) are, essentially, generalized forms of a geometric distribution. For particular groups and error models, it may be that this expression is well-approximated by a standard geometric distribution (which would be of substantially lower numerical complexity). It would be interesting to investigate whether a reduction of (6.11) and (6.30) occurs, or whether an approximation can be found, for particular quantum systems.
– With that idea in mind, note that the hitting time of the set $\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,\delta}$ is naturally related to its size relative to the size of the group $\mathscr{G}_n$. As the number of qubits increases, both of these sets grow in size. Investigating the growth relation between these two sets for particular groups via techniques from analytical combinatorics [225] may reveal an asymptotic distributional law for the errors in quantum computations with many qubits.
– The availability of an analytical expression for the accumulation of errors allows us to proceed with second-tier optimization methods. For example, to achieve practical quantum computing in the near future, any quantum computer architecture would have some classical control mechanism that routinely takes operational decisions: which gate do we apply next, do we now apply an error correction procedure, etc. Each of these operations has its own cost associated with it, e.g., in the form of classical compute time or the loss of ancillary qubits. Using techniques from decision theory [226], we can weigh the long-term effects of different operations through the available analytical expressions and we could overall achieve more efficient computations in the future. Essentially, we could then compute more with fewer qubits.

A

APPENDIX TO CHAPTER 2

A.1. INFERRING NETWORK METRICS GIVEN THE NETWORK TYPE

Metric	Mean absolute error (MAE)		Mean squared error (MSE)	
	Treatment group	Control group	Treatment group	Control group
$E[D]$	0.381	1.34	0.239	2.69
$E[D^2]$	4.79	17.5	38.7	4.67×10^2
λ_1	0.372	1.31	0.219	2.55
$E[H]$	0.201	0.435	6.46×10^{-2}	0.297
$E[1/H]$	1.90×10^{-2}	3.73×10^{-2}	5.74×10^{-4}	2.17×10^{-3}
d_{max}	1.46	2.50	3.71	9.95
C_G	6.54×10^{-3}	8.31×10^{-3}	7.46×10^{-5}	1.12×10^{-4}
μ_{N-1}	0.177	0.312	6.54×10^{-2}	0.173
ρ_D	3.44×10^{-2}	3.46×10^{-2}	1.82×10^{-3}	1.89×10^{-3}
N	95.0	1.00×10^2	1.39×10^4	1.53×10^4

Table A.1: MAE and MSE of different network metrics when the benchmark networks are ER networks. The effective infection rate is set as $\tau = 1$ and the initial state $y_0 = 0.2$. The mean errors are obtained by averaging over 1000 ER benchmark networks.

A

Metric	Mean absolute error (MAE)		Mean squared error (MSE)	
	Treatment group	Control group	Treatment group	Control group
$E[D]$	0.434	1.36	0.458	2.97
$E[D^2]$	0.757	16.2	13.1	4.95×10^2
λ_1	9.12×10^{-2}	1.41	7.34×10^{-2}	3.50
$E[H]$	0.190	0.627	6.03×10^{-2}	0.625
$E[1/H]$	1.63×10^{-2}	4.62×10^{-2}	4.51×10^{-4}	3.32×10^{-3}
d_{max}	1.05	2.45	2.01	9.35
C_G	1.89×10^{-2}	2.35×10^{-2}	6.58×10^{-4}	9.60×10^{-4}
μ_{N-1}	8.63×10^{-2}	0.629	2.24×10^{-2}	0.703
ρ_D	3.42×10^{-2}	4.06×10^{-2}	1.88×10^{-3}	2.61×10^{-3}
N	91.2	1.04×10^2	1.31×10^4	1.59×10^4

Table A.2: MAE and MSE of different network metrics when the benchmark networks are WS networks. The parameter settings are the same as Tab. A.1.

Metric	Mean absolute error (MAE)		Mean squared error (MSE)	
	Treatment group	Control group	Treatment group	Control group
$E[D]$	3.60×10^{-2}	1.44	7.19×10^{-2}	3.72
$E[D^2]$	6.57	31.9	85.3	1.62×10^3
λ_1	0.820	2.36	1.09	8.46
$E[H]$	0.123	0.332	2.55×10^{-2}	0.177
$E[1/H]$	1.49×10^{-2}	3.44×10^{-2}	3.69×10^{-4}	1.83×10^{-3}
d_{max}	13.6	15.1	3.08×10^2	3.69×10^2
C_G	1.84×10^{-2}	1.98×10^{-2}	5.75×10^{-4}	6.15×10^{-4}
μ_{N-1}	9.14×10^{-2}	0.551	4.27×10^{-2}	0.520
ρ_D	2.39×10^{-2}	3.16×10^{-2}	9.53×10^{-4}	1.55×10^{-3}
N	97.3	1.01×10^2	1.47×10^4	1.55×10^4

Table A.3: MAE and MSE of different network metrics when the benchmark networks are BA networks. The parameter settings are the same as Tab. A.1.

Metric	Mean absolute error (MAE)		Mean squared error (MSE)	
	Treatment group	Control group	Treatment group	Control group
$E[D]$	0.151	0.300	0.216	2.86
$E[D^2]$	17.9	35.2	5.37×10^2	1.88×10^3
λ_1	1.83	2.73	5.17	11.2
$E[H]$	2.38×10^{-2}	0.137	3.61×10^{-2}	0.141
$E[1/H]$	1.70×10^{-2}	2.92×10^{-2}	4.51×10^{-4}	1.32×10^{-3}
d_{max}	18.7	23.6	5.46×10^2	8.52×10^2
C_G	2.34×10^{-2}	2.71×10^{-2}	8.57×10^{-4}	1.17×10^{-3}
μ_{N-1}	0.124	0.173	2.60×10^{-2}	4.90×10^{-2}
ρ_D	2.74×10^{-2}	3.02×10^{-2}	1.19×10^{-3}	1.42×10^{-3}
N	94.6	1.02×10^2	1.37×10^4	1.56×10^4

Table A.4: MAE and MSE of different network metrics when the benchmark networks are SF networks. The parameter settings are the same as Tab. A.1.

B

APPENDIX TO CHAPTER 3

B.1. QUALITATIVE EXPLANATION FOR THE FORMATION MECHANISM OF THE LOOP PATTERNS

The peak locations of the recovery data $\Delta\widetilde{R}[k+1]$ and infection data $\tilde{I}[k]$ cut the time series into three "phases", which are marked by different colors. We provide a qualitative explanation for the formation mechanism of the loop patterns based on these "phases" in Figure B.4c: (i) the fraction of new reported recoveries $\Delta\widetilde{R}[k+1]$ and reported active infections $\tilde{I}[k]$ are all growing with the day k, but the fraction of reported active infections $\tilde{I}[k]$ grows earlier than the fraction of new reported recoveries $\Delta\widetilde{R}[k+1]$, which leads to a small $\Delta\widetilde{R}[k+1]$ to $\tilde{I}[k]$ ratio at the beginning; (ii) the fraction of new recoveries $\Delta\widetilde{R}[k+1]$ is increasing but the fraction of reported active infections $\tilde{I}[k]$ is decreasing; (iii) the fraction of new reported recoveries $\Delta\widetilde{R}[k+1]$ and the fraction of reported active infections $\tilde{I}[k]$ are all decreasing, but the fraction of reported active infections $\tilde{I}[k]$ decreases earlier than the fraction of new reported recoveries $\Delta\widetilde{R}[k+1]$, which leads to a large $\Delta\widetilde{R}[k+1]$ to $\tilde{I}[k]$ ratio. These three "phases" lead to the blue points moving in a counter-clockwise direction and form the loop pattern in the end.

B.2. MODELING THE REPORTING DELAYS

Based on the above observations and inspired by the widespread reporting delays of infections [14], we hypothesize that reported data, not only the infections but also recoveries and deaths, are subject to delays. The fraction $Y[k]$, which can be the fraction of new

B

infections, recoveries or deaths, is contained in $[0,1]$ and measured per day. Hence, the analysis is in discrete day k. The basic observation lies in a time delay T between the new real value $\Delta Y[k]$ and the reported value with delays $\Delta \widetilde{Y}[k]$, which translates to

$$\Delta \widetilde{Y}[k] = \Delta Y[k-T]. \tag{B.1}$$

We further assume that the reporting delay T and the new value ΔY are independent. We argue that independence between T and ΔY is reasonable, although both random variables are weakly positively correlated. Indeed, the larger the fraction ΔY, the more people need to be checked and the longer the reporting may take. If the checking capacity is sufficiently large, we may assume approximate independence.

We implicitly assume that the new reported value $\Delta \widetilde{Y}[k]$ only differs from reality in the time delay T. The delay T must be non-negative, i.e. $T \geq 0$, because the real event $\Delta Y[k]$ occurs at discrete day k and its reporting occurs at $k+T \geq k$, which cannot be earlier than the day k. Moreover, the delay T is also an integer, else Eq. (B.1) demands us to take the integer value $[T]$ of $T = [T] + \langle T \rangle$, where the fractional part $0 \leq \langle T \rangle < 1$. We avoid that complication and consider T as a discrete random variable. If T is discrete, then all involved random variables are discrete. Before proceeding, we thus approximate $\Delta Y(t)$ at continuous time $k-1 < t \leq k$, by $\Delta Y[k] = \int_{k-1}^{k} \Delta Y(u)\,du$. Furthermore, the mean-field approximation only writes the equations for the *average* fraction of infected nodes and we refer to [31, p.454] for the relation between the Markov process and its mean-field approximation.

The basic observation in Eq. (B.1) contains the real fraction ΔY and the delay T. Hence, we use the law of total probability [31, p.23],

$$\Pr\left[\Delta Y[k-T] \leq y\right] = \sum_{m=0}^{\infty} \Pr\left[\Delta Y[k-T] \leq y \middle| T = m\right] \Pr[T = m],$$

If we assume that $\Delta Y[k]$ and T are independent (for any k), then

$$\Pr\left[\Delta Y[k-T] \leq y \middle| T = m\right] = \frac{\Pr\left[\{\Delta Y[k-T] \leq y\} \cap \{T = m\}\right]}{\Pr[T = m]} = \Pr\left[\Delta Y[k-m] \leq y\right].$$

Thus,

$$\Pr\left[\Delta Y[k-T] \leq y\right] = \sum_{m=0}^{\infty} \Pr\left[\Delta Y[k-m] \leq y\right] \Pr[T = m].$$

Since we confine to a mean-field analysis and are only interested in the *average* fractions, we better take the expectation operator $E[.]$ instead of the probability $\Pr[.]$ operator:

$$E[\Delta Y[k-T]] = \sum_{m=0}^{\infty} E[\Delta Y[k-m]] \Pr[T = m], \tag{B.2}$$

which is readily obtained from the former equation by using the definition of the mean (e.g. [31, (2.36)]) $E[\Delta Y[k-T]] = \int_0^1 \Pr\left[\Delta Y[k-T] > y\right] dy$. Finally, we simplify the notation as $\Delta\widetilde{Y}[k] = E\left[Y_{\text{rep}}[k]\right]$ and arrive at our approximative observation hypothesis for the average fraction of infected, recovered or deceased individuals

$$\Delta\widetilde{Y}[k] = \sum_{m=0}^{\infty} \Pr[T=m]\,\Delta Y[k-m]. \tag{B.3}$$

Let the day $k=0$ denote the day that the first infected individual appeared, which indicates that the fractions $\Delta D[k-m]=0$, $\Delta I[k-m]=0$ and $\Delta R[k-m]=0$ for all $m>k$. Under this situation, for new infections, recoveries and deaths, Eq. (B.3) reduces to

$$\Delta\widetilde{Y}[k] = \sum_{m=0}^{k} \Pr[T=m]\Delta Y[k-m], \tag{B.4}$$

where $\Pr[T=m]$ denotes the probability that the report of a case is delayed with m days.

B.3. REPORTING DELAYS FOLLOWING THE SAME DISTRIBUTION

Lemma 8. *The proportional relationships*

$$\Delta R[k+1] = \gamma_r I[k] \quad \textit{and} \quad \Delta D[k+1] = \gamma_d I[k], \tag{B.5}$$

will still hold after the reporting delays if the reporting delays T_D, T_I and T_R follow the same distribution, which means that the parameters of all three delay distributions are exactly the same.

Proof. If the reporting delays T_D, T_I and T_R follow the same distribution, which means that

$$\Pr[T_D=m] = \Pr[T_I=m] = \Pr[T_R=m] = \Pr[x=m] \quad \text{for} \quad m=1,2,\cdots \tag{B.6}$$

based on (B.4) and (B.6), the fraction of reported active infections leads to

$$\tilde{I}[k] = \sum_{\tau=0}^{k} (\Delta\widetilde{I}[\tau] - \Delta\widetilde{R}[\tau] - \Delta\widetilde{D}[\tau]) \tag{B.7}$$

$$= \sum_{\tau=0}^{k}\sum_{\eta=0}^{\tau-1} \Pr[x=\eta]\left(\Delta I[\tau-\eta] - \Delta R[\tau-\eta] - \Delta D[\tau-\eta]\right) \tag{B.8}$$

$$= \sum_{\eta=0}^{k-1} \Pr[x=\eta] I[k-\eta]. \tag{B.9}$$

If the spread of COVID-19 follows the SIRD model, we have that

$$\Delta R[k+1] = \gamma_r I[k] \quad \text{and} \quad \Delta D[k+1] = \gamma_d I[k]. \tag{B.10}$$

Combining (B.4), (B.6), (B.9) and (B.10), we have that

$$\Delta\widetilde{R}[k+1] = \sum_{\eta=0}^{k} \Pr[T_R = \eta]\Delta R[k-\eta] = \sum_{\eta=0}^{k} \Pr[x = \eta]\Delta R[k-\eta] \tag{B.11}$$

$$= \gamma_r \sum_{\eta=0}^{k-1} \Pr[x = \eta] I[k-\eta] = \gamma_r I[k], \tag{B.12}$$

$$\Delta\widetilde{D}[k+1] = \sum_{\eta=0}^{k} \Pr[T_D = \eta]\Delta D[k-\eta] = \sum_{\eta=0}^{k} \Pr[x = \eta]\Delta D[k-\eta] \tag{B.13}$$

$$= \gamma_d \sum_{\eta=0}^{k-1} \Pr[x = \eta] I[k-\eta] = \gamma_d I[k]. \tag{B.14}$$

□

B.4. SIMULATION RESULTS FOR DIFFERENT DELAY DISTRIBUTIONS

We assume that probability distributions for infections, recoveries and deaths are the same functional. To determine the family of reporting delay distributions that best suit our data, we consider three different two-parameter discrete distributions [114] below:
(I) *Negative binomial distribution.* The probabilities that a deceased, infected or recovered individual is reported after $m \in \mathbb{N}$ days are

$$\Pr[T = m] = \binom{m+r-1}{m}(1-p)^m p^r. \tag{B.15}$$

The negative binomial distribution with parameters $r > 0$ and $p \in [0,1]$ has mean value $E[T] = r(1-p)/p$ and variance $Var[T] = r(1-p)/p^2$.
(II) *Pólya-Aeppli distribution (also called the geometric Poisson distribution).* The probabilities that a deceased, infected or recovered individual is reported after $m \in \mathbb{N}$ days are

$$\Pr[T = m] = \begin{cases} \sum_{j=1}^{m} e^{-\lambda}\frac{\lambda^j}{j!}(1-\theta)^{m-j}\theta^j\binom{m-1}{j-1}, & m > 0 \\ e^{-\lambda}, & m = 0 \end{cases}. \tag{B.16}$$

The Pólya-Aeppli distribution with parameters $\lambda > 0$ and $\theta \in [0,1]$ has mean value $E[T] = \lambda/\theta$ and variance $Var[T] = \lambda(2-\theta)/\theta^2$.
(III) *Neyman type A distribution.* The probabilities that a deceased, infected or recovered individual is reported after $m \in \mathbb{N}$ days are,

$$\Pr[T = m] = \frac{\mu^m e^{-\xi}}{m!} \sum_{j=0}^{\infty} \frac{(\xi e^{-\mu})^j}{j!} j^m. \tag{B.17}$$

The Neyman type A distribution with parameters $\xi > 0$ and $\mu > 0$ has mean value $E[T] = \xi\mu$ and variance $Var[T] = \xi\mu(1+\mu)$.

Fig. B.3, Fig. B.4 and Fig. B.5 show the different delay distributions and corresponding time series and loop patterns.

B.5. REPORTING DELAYS ON SYNTHETIC DATA FROM DIFFERENT COMPARTMENTAL MODELS

The Susceptible-Infectious-Recovered-Dead (SIRD) model and their extensions are widely applied to describe the dynamics of the COVID-19 pandemic. These compartmental models are often described by the ordinary differential equations, which are valid in case of sufficiently large populations (the thermodynamic limit). Here we present four common-used compartmental models as follows.

1. The SIRD model [5, 2] is one of the simplest models to describe the outbreak of COVID-19. The SIRD model consists of four compartments: the fraction of susceptible individuals (S), the fraction of active infected individuals (I), the fraction of recovered individuals (R) and the fraction of deceased individuals (D). There are transition rates between compartments. Specifically, the transition rate from S to I is βSI, where β denotes the infection rate; the transition rates from I to R and D are respectively $\gamma_r I$ and $\gamma_d I$, where γ_r denotes the recovery rate and γ_d denotes the deceased rate. The SIRD model can be formulated as the following systems of ordinary differential equations (ODEs) [5, 2]:

$$\begin{cases} \dfrac{dI}{dt} = \beta IS - (\gamma_r + \gamma_d)I, \\ \dfrac{dR}{dt} = \gamma_r I, \\ \dfrac{dD}{dt} = \gamma_d I, \end{cases} \tag{B.18}$$

   and it holds that $S + I + R + D = 1$.

2. The infection rate β in the SIRD model can be time-varying due to the changes in humidity [119] and policies (e.g., the social distance, quarantine, city-wise lockdown [227] and wearing masks). The SIRD model [96] can then be extended by considering a time-varying infection rate $\beta(t)$. The SIRD model with a time-varying

infection rate $\beta(t)$ can be formulated as the following ODEs [96]:

$$\begin{cases} \dfrac{dI}{dt} = \beta(t)IS - (\gamma_r + \gamma_d)I, \\ \dfrac{dR}{dt} = \gamma_r I, \\ \dfrac{dD}{dt} = \gamma_d I, \\ S + I + R + D = 1. \end{cases} \tag{B.19}$$

3. The exposed state is also considered in some works. The SIRD model is then expanded to the Susceptible-Exposed-Infectious-Recovered-Dead (SEIRD) model [5, 2]. The Susceptible-Exposed-Infectious-Recovered-Dead (SEIRD) model can be formulated as the following ODEs [5, 2]:

$$\begin{cases} \dfrac{dE}{dt} = \beta IS - \sigma E, \\ \dfrac{dI}{dt} = \sigma E - (\gamma_r + \gamma_d)I, \\ \dfrac{dR}{dt} = \gamma_r I, \\ \dfrac{dD}{dt} = \gamma_d I, \\ S + E + I + R + D = 1, \end{cases} \tag{B.20}$$

where σ denotes the incubation rate.

4. By considering the population flow between regions, researchers propose the SIRD model on the human mobility network [7, 219]. The SIRD process on the human mobility network with N regions can be expressed as follows [7, 219],

$$\begin{cases} \dfrac{d\boldsymbol{I}}{dt} = \text{diag}(\boldsymbol{S})\mathbf{B}\boldsymbol{I} - \text{diag}(\boldsymbol{\gamma_r} + \boldsymbol{\gamma_d})\boldsymbol{I}, \\ \dfrac{d\boldsymbol{R}}{dt} = \text{diag}(\boldsymbol{\gamma_r})\boldsymbol{I}, \\ \dfrac{d\boldsymbol{D}}{dt} = \text{diag}(\boldsymbol{\gamma_d})\boldsymbol{I}, \\ \boldsymbol{S} + \boldsymbol{I} + \boldsymbol{R} + \boldsymbol{D} = \mathbf{u}, \end{cases} \tag{B.21}$$

where the elements of matrix $\mathbf{B}$ denote the transition rates between regions, which is given by

$$\mathbf{B} = \begin{pmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{N1} & \beta_{N2} & \dots & \beta_{NN} \end{pmatrix}, \tag{B.22}$$

the state vectors $\boldsymbol{S}$, $\boldsymbol{I}$, $\boldsymbol{R}$ and $\boldsymbol{D}$ are respectively

$$\boldsymbol{S} = (S_1[t], S_2[t], \cdots, S_N[t]), \quad \boldsymbol{I} = (I_1[t], I_2[t], \cdots, I_N[t]), \tag{B.23}$$

$$\boldsymbol{R} = (R_1[t], R_2[t], \cdots, R_N[t]), \quad \boldsymbol{D} = (D_1[t], D_2[t], \cdots, D_N[t]), \tag{B.24}$$

the recovered and mortality rate vectors $\boldsymbol{\gamma_r}$ and $\boldsymbol{\gamma_d}$ are respectively

$$\boldsymbol{\gamma_r} = (\gamma_{r,1}, \gamma_{r,2}, \cdots, \gamma_{r,N}), \quad \boldsymbol{\gamma_d} = (\gamma_{d,1}, \gamma_{d,2}, \cdots, \gamma_{d,N}), \tag{B.25}$$

the vector $\mathbf{u}$ denotes the all-one vector and diag$(*)$ denotes the diagonal matrix with elements of a vector $*$. The population-flow network $\mathbf{B}$ can also be time-varying due to the city-wise lockdown.

By analyzing the commonality in the above compartmental epidemic models mentioned above and under the discrete-time approximation, we have the following relationships between the fractions of ΔR, ΔD at the day $k+1$ and the fraction of I at day k:

$$\Delta R[k+1] = R[k+1] - R[k] = \gamma_r I[k], \tag{B.26}$$

and

$$\Delta D[k+1] = D[k+1] - D[k] = \gamma_d I[k], \tag{B.27}$$

where γ_r and γ_d respectively denote the recovered and mortality rates.

Fig. B.6-Fig. B.11 show the reported time series and loop patterns for different compartmental models and different delay distribution types. For the SIRD model and the SIRD model with a time-varying infection rate $\beta(t)$, the initial state is that 10^{-4} of all individuals are infected and the others are susceptible. For the SEIRD model, the initial state is that 10^{-4} of all individuals are infected, 10^{-2} of all individuals are exposed and the others are susceptible. Three different types of delay distributions are considered: the negative binomial distributions, the Neyman type A distributions and the Pólya-Aeppli distributions. Peak shifts and loop patterns are discovered in all compartmental models and delay distributions.

B.6. INFER DELAY INFORMATION USING SYNTHETIC AND REAL DATA

We do 10^7 random search on the parameters of delay distributions. Each time we uniformly random choose mean reporting delays $E[T_I]$, $E[T_R]$ and $E[T_D]$ in range $[0,30]$. For the negative binomial distributions, the parameters p_D, p_R and p_I are uniformly randomly chosen in range $[0,1]$. For the Pólya-Aeppli distributions, the parameters θ_D, θ_R and θ_I are also uniformly randomly chosen in range $[0,1]$. For the Neyman type A

Regions	λ_I	θ_I	λ_R	θ_R	λ_D	θ_D
Italy	2.8741	0.4193	2.0701	0.0719	0.1787	0.6750
Spain	1.0803	0.1632	1.2673	0.0553	0.1713	0.6574
Wuhan	0.4661	0.0186	2.6485	0.1377	0.0681	0.0154
Turkey	0.6803	0.0264	2.6537	0.1763	0.0135	0.0146
Hubei	0.1745	0.0075	2.2779	0.1866	0.0033	0.0015
Romania	0.3902	0.0204	0.0026	0.0008	0.0632	0.8985
Germany	0.4122	0.0230	0.0028	0.0007	0.1878	0.1371
Denmark	1.0399	0.0387	0.8382	0.4240	0.0081	0.0045

Table B.1: Inferred distribution parameters.

distributions, the parameters ξ_D, ξ_R and ξ_I are uniformly randomly chosen in range $[0, \sqrt{30}]$. The guessed data with no delays can be deduced given the reported data and the guessed distributions. The corresponding product $\rho(\Delta\bar{R}, \Delta\bar{D})\rho(\bar{I}, \Delta\bar{R})\rho(\bar{I}, \Delta\bar{D})$ can be further calculated.

Table B.2 shows the inferred mean differences and the maximal products $\rho(\Delta\bar{R}, \Delta\bar{D})\rho(\bar{I}, \Delta\bar{R})\rho(\bar{I}, \Delta\bar{D})$ for 8 countries or regions. The maximal products $\rho(\Delta\bar{R}, \Delta\bar{D})\rho(\bar{I}, \Delta\bar{R})\rho(\bar{I}, \Delta\bar{D})$ obtained based on the Pólya-Aeppli distributions are all larger than these values optimized based on the negative binomial or Neyman type A distributions, revealing that the Pólya-Aeppli distribution can describe real reporting delay distributions better.

B.7. FORECAST THE COVID-19 PANDEMIC

The COVID-19 pandemic is predicted based on the Algorithms 3 and 4. Algorithm 3 is to forecast the future infected fractions without considering the effect of reporting delays. Algorithm 4 is to forecast the future infected fractions considering the reporting delays.

B.8. DATA AVAILABILITY

The data sources of COVID-19 cases for each country/region are as follows:
1. China: National Health Commission of the People's republic of China (http://en.nhc.gov.cn/) and the health commission of every Chinese province. The data start from January 10, 2020 and end on March 15, 2020. We consider 31 cities including Wuhan and 30 other cities (in and out of Hubei province) that are close to Wuhan from the view of human mobility. The human mobility data are obtained via the Baidu map (https://qianxi.baidu.com/). There are 16 cities of these 31 cities located in Hubei

Country/Region	Negative binomial		Pólya-Aeppli		Neyman type A	
	$E[T_R]-E[T_D]$	Maximum	$E[T_R]-E[T_D]$	Maximum	$E[T_R]-E[T_D]$	Maximum
Italy	25.45	0.80	**28.52**	**0.84**	8.26	0.77
Spain	18.50	0.97	**22.66**	**0.98**	8.07	0.91
Wuhan	11.01	0.51	**14.80**	**0.63**	9.64	0.52
Turkey	14.29	0.83	**14.13**	**0.85**	8.00	0.73
Hubei	10.07	0.61	**9.96**	**0.71**	6.72	0.60
Romania	-0.18	0.69	**3.30**	**0.70**	0.01	0.69
Germany	-1.91	0.75	**2.72**	**0.76**	-3.07	0.73
Denmark	1.90	0.72	**0.17**	**0.72**	0.96	0.70

Table B.2: Inferred mean difference $E[T_R]-E[T_D]$ for different countries or regions. the maximum value denotes the maximal product of correlations $\rho(R,D)\rho(I,R)\rho(I,D)$ by 10^6 random search of the distribution parameters.

B

Algorithm 3: `Forecast the real pandemic without considering the reporting delays`

Input: fraction of daily reported infected cases $\Delta\tilde{I}[1],...,\Delta\tilde{I}[n]$.

Output: predicted fraction of infections $\Delta\hat{I}[n+1],...,\Delta\hat{I}[n+n_{\text{pred}}]$.

Smooth the data by Matlab toolbox *smoothdata*.

Set the initial value of the loss function $\Theta_s \leftarrow 1$; the initial infection rate $\beta_s \leftarrow 0.01$; the initial removed rate (the sum of recovered rate and deceased rate) $\gamma_s \leftarrow 0.01$; the initial infected fraction $I_s[0] \leftarrow 0.01$; **for** $\beta = 0.01, 0.02, ..., 1$ **do**

  for $\gamma = 0.01, 0.02, ..., 1$ **do**

    for $k = 0, 1, ..., 100$ **do**

$I[0] = 10^{-2-k/25}$;

Numerically solve the equations of SIR model based on the parameters $I[0]$, β and γ and obtain an infection curve $\Delta\mathscr{I}[1],...,\Delta\mathscr{I}[n+n_p]$.

Scale the simulated curve $\Delta\mathscr{I}[1],...,\Delta\mathscr{I}[n+n_p]$ to data $\Delta\hat{I}[1],...,\Delta\hat{I}[n+n_p]$ by letting $\Delta\hat{I}[i] = \Delta\mathscr{I}[i] \times \Delta\tilde{I}[1]/\Delta\mathscr{I}[1]$ for day $i-1, 2, \cdots, n+n_p$.

Calculate the lose function $\Theta = \sqrt{\frac{1}{n}\sum_{k=0}^{n-1}(\Delta\hat{I}[k] - \Delta\tilde{I}[k])^2}$. **if** $\Theta < \Theta_s$ **then**

$\Theta_s \leftarrow \Theta$; $I_s[0] \leftarrow I[0]$; $\beta_s \leftarrow \beta$; $\gamma_s \leftarrow \gamma$.

    end

  end

end

end

Obtain the best forecast results based on the optimized parameters $I_s[0]$, β_s and γ_s.

Algorithm 4: `Forecast the real pandemic considering the reporting delays`

Input: fraction of daily reported infected cases $\Delta\tilde{I}[1],...,\Delta\tilde{I}[n]$; fraction of daily reported deceased cases $\Delta\tilde{D}[1],...,\Delta\tilde{D}[n]$; fraction of daily reported recovered cases $\Delta\tilde{R}[1],...,\Delta\tilde{R}[n]$; prediction time n_{pred}.

Output: predicted fraction of infections $\Delta\hat{I}[n+1],...,\Delta\hat{I}[n+n_{\text{pred}}]$.

Smooth the data by Matlab toolbox *smoothdata*.

Infer delay distribution for infected cases T_I using the reported data $\Delta\tilde{I}$, $\Delta\tilde{R}$ and $\Delta\tilde{D}$.

Obtain the inferred data $\Delta\bar{I}[1],...,\Delta\bar{I}[n]$ by removing the effect of reporting delays.

Set the initial value of the loss function $\Theta_s \leftarrow 1$; the initial infection rate $\beta_s \leftarrow 0.01$; the initial removed rate (the sum of recovered rate and deceased rate) $\gamma_s \leftarrow 0.01$; the initial infected fraction $I_s[0] \leftarrow 0.01$; **for** $\beta = 0.01, 0.02, ..., 1$ **do**

  for $\gamma = 0.01, 0.02, ..., 1$ **do**

    for $k = 0, 1, ..., 100$ **do**

      $I[0] = 10^{-2-k/25}$;

      Numerically solve the equations of SIR model based on the parameters $I[0]$, β and γ and obtain an infection curve $\Delta\mathscr{I}[1],...,\Delta\mathscr{I}[n+n_p]$.

      Scale the simulated curve $\Delta\mathscr{I}[1],...,\Delta\mathscr{I}[n+n_p]$ to data $\Delta\hat{I}[1],...,\Delta\hat{I}[n+n_p]$ by letting $\Delta\hat{I}[i] = \Delta\mathscr{I}[i] \times \Delta\bar{I}[1]/\Delta\mathscr{I}[1]$ for day $i-1, 2, \cdots, n+n_p$.

      Calculate the lose function $\Theta = \sqrt{\frac{1}{n}\sum_{k=0}^{n-1}(\Delta\hat{I}[k] - \Delta\bar{I}[k])^2}$. **if** $\Theta < \Theta_s$ **then**

        $\Theta_s \leftarrow \Theta$; $I_s[0] \leftarrow I[0]$; $\beta_s \leftarrow \beta$; $\gamma_s \leftarrow \gamma$.

      end

    end

  end

end

Obtain the best forecast results based on the optimized parameters $I_s[0]$, β_s and γ_s.

Add the reporting delays on the forecast data.

province: Wuhan, Xiaogan, Huanggang, Ezhou, Xianning, Jingzhou, Huangshi, Xiangyang, Suizhou, Xiantao, Jingmen, Yichang, Shiyan, Tianmin, Enshi and Qianjiang. The rest 15 Chinese cities out of Hubei are listed as follows: Xinyang, Changsha, Shanghai, Beijing, Chongqing, Nanyang, Zhumadian, Zhengzhou, Jiujiang, Yueyang, Shenzhen, Guangzhou, Nanchang, Chengdu and Anqing.

2. Denmark: Statens Serum Institut (https://www.kl.dk). The data start from February 27, 2020 and end on May 16, 2020.

3. Turkey: Ministry of Health (Turkey) (https://covid19.saglik.gov.tr/). The data start from March 11, 2020 and end on May 16, 2020.

4. Italy: Dipartimento della Protezione Civile (http://www.salute.gov.it/portale/home.html). The data start from Feb 21, 2020 and end on May 4, 2020.

5. Spain: Ministry of Health (Spain) (https://www.mscbs.gob.es/en/home.htm). The data start from February 25, 2020 and end on May 15, 2020.

6. Germany: Robert Koch-Institut (RKI) (https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html) and new situation reports of the RKI (https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Gesamt.html). The data start from March 1, 2020 and end on May 19, 2020.

7. Romania: Ministry of Health (Romania) (http://www.ms.ro/comunicate/). The data start from February 26, 2020 and end on May 23, 2020.

8. South Korea: Korea Centers for Disease Control and Prevention (https://www.cdc.go.kr/). The data start from February 18, 2020 and end on April 8, 2020.

9. UK: The pandemic data are from the World Health Organization (https://covid19.who.int/). The reporting delay data for deaths are from the National Health Service (NHS) in England (https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-new-deaths/). The date we considered is the cases that were reported from Apr 2 to Apr 4, 2020. The mean and variance of the reporting delays for infections are obtained from the paper [30]. We generated the delay distribution using the mean and variance based on the Pólya-Aeppli distributions since Table 3.2 reveals that many real reporting delays prefer the Pólya-Aeppli distributions.

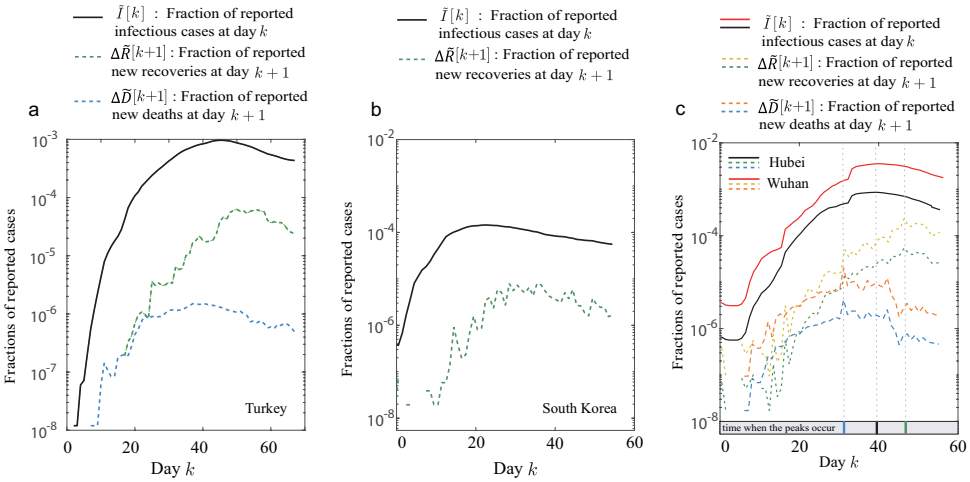


Figure B.1: Time series of the fractions of reported active infections $\tilde{I}[k]$, new reported recoveries $\Delta\widetilde{R}[k+1]$ and new reported deaths $\Delta\widetilde{D}[k+1]$ for Turkey (figure a), South Korea (figure b) and China (figure c). The fractions of new reported deaths for South Korea are too small to be seen effectively. The peak locations are remarkably different.

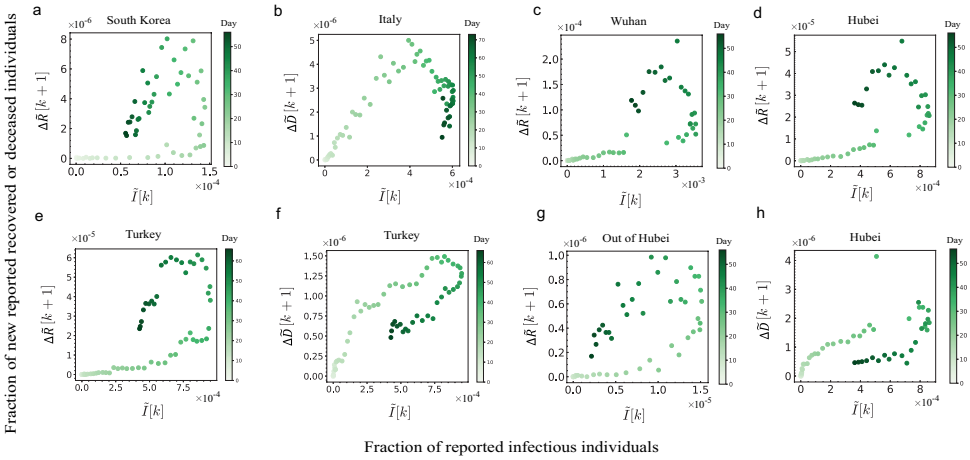


Figure B.2: Data points between the fraction of reported active infections $\tilde{I}[k]$ and the fraction of new reported recoveries $\Delta\widetilde{R}[k+1]$ or deaths $\Delta\widetilde{D}[k+1]$ for South Korea, Italy, Wuhan, Turkey, Hubei and Chinese cities out of Hubei. Loop patterns instead of straight lines are observed. Symbol color (from light to dark) show the day k changing from day 0 to day 57. Data points between $\tilde{I}[k]$ and $\Delta\widetilde{R}[k+1]$ evolve in a counter-clockwise direction, while data points between $\tilde{I}[k]$ and $\Delta\widetilde{D}[k+1]$ evolve in a clockwise direction.

B

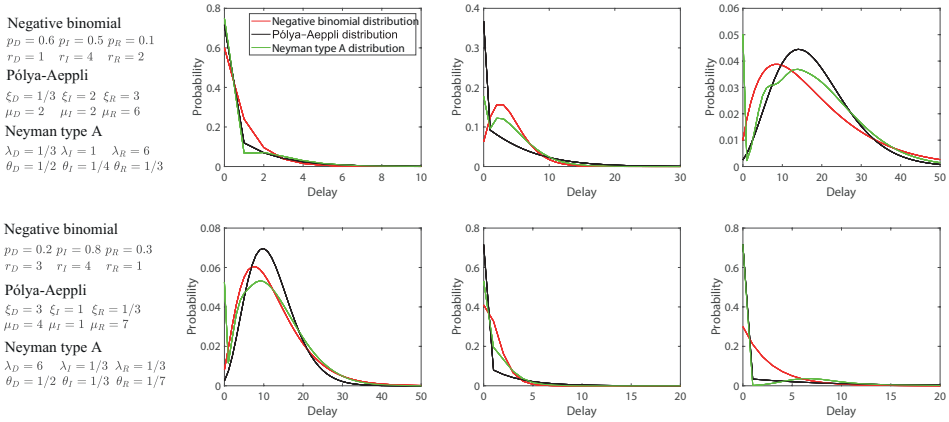


Figure B.3: Delay distributions considered in Fig. B.4. The upper figures from left to right are respectively delays of deceased, infected and recovered data corresponding to Fig. B.4a-c. The bottom figures are delays distributions for Fig. B.4d-f. The mean delays in each subplot are the same. The distribution curves from different types are usually very different.

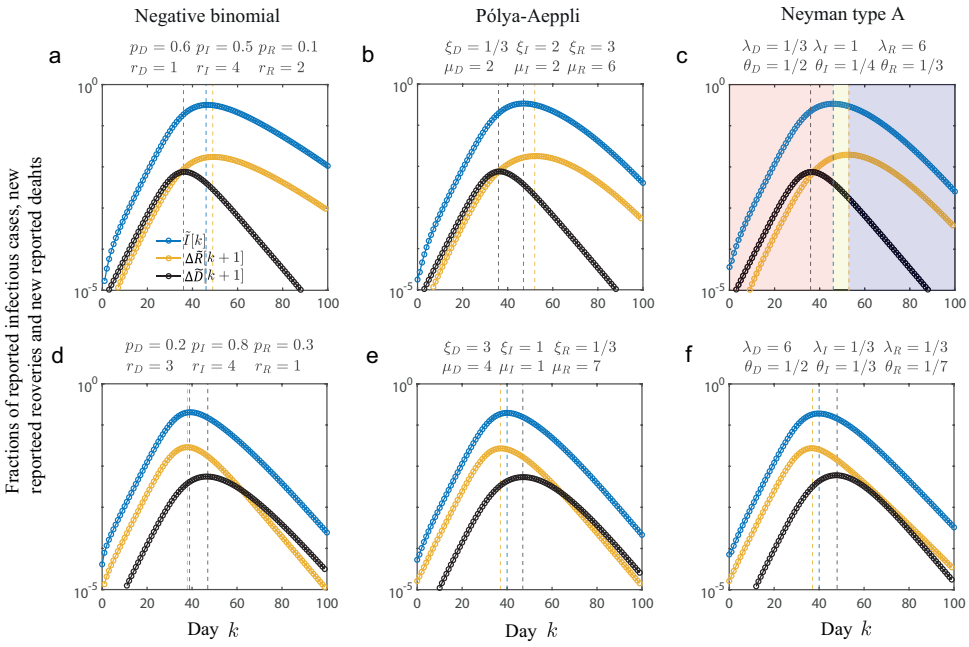


Figure B.4: Time series of the fractions of reported active infections $\tilde{I}[k]$, new reported recoveries $\Delta\widetilde{R}[k+1]$ and new reported deaths $\Delta\widetilde{D}[k+1]$. The synthetic data are generated based on the SIRD model and the delay distributions (the negative binomial distributions, the Pólya-Aeppli distributions and the Neyman type A distributions). The infection rate $\beta = 0.5$, the recovered rate $\gamma_r = 0.2$ and the mortality rate $\gamma_d = 0.05$. The dashed lines show the peak locations. Fractions smaller than 10^{-5} are ignored. The right distributions are delay distributions considered in the simulation.

B

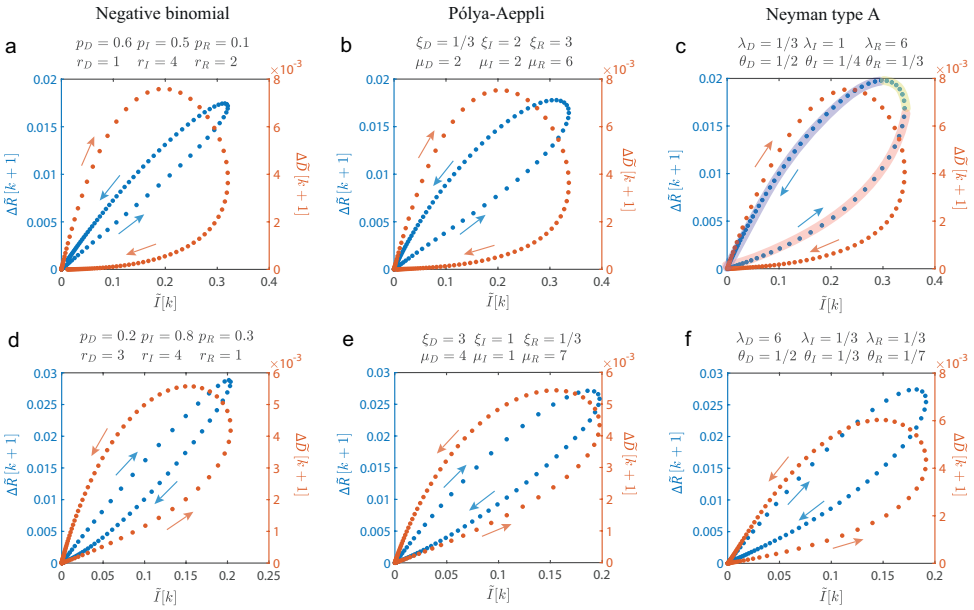


Figure B.5: Data points between the fraction of reported active infections $\widetilde{I}[k]$ and the fraction of new reported recoveries $\Delta\widetilde{R}[k+1]$ or deaths $\Delta\widetilde{D}[k+1]$. Loop patterns, instead of straight lines, which are expected for the data with no delays, are observed. The red and blue arrows show the evolution directions of the data points with time.

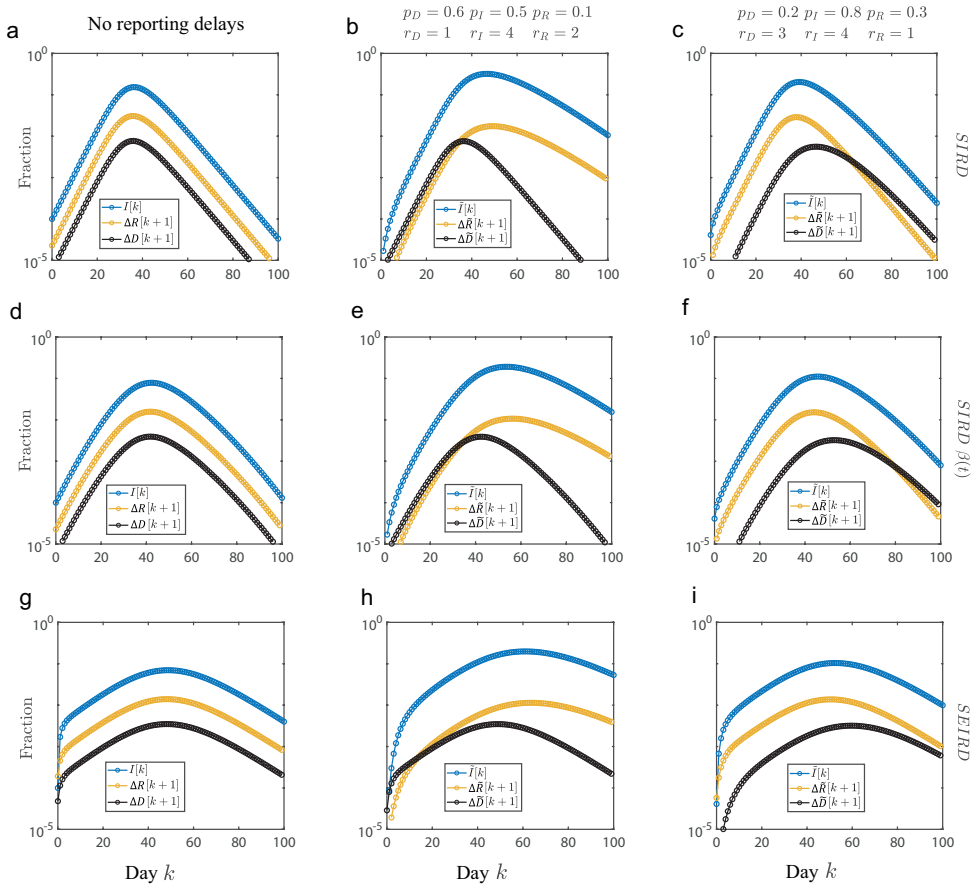


Figure B.6: Peak shift after reporting delays following the negative binomial distributions. Three different kinds of epidemic models are considered: the SIRD model, the SIRD model with a time-varying infection rate $\beta(t) = 0.5 - 0.003t$ and the SEIRD model. Figures a, d and g show the curves of real data $\Delta R[k+1]$, $\Delta D[k+1]$ and $I[k]$. The other figures show the curves after reporting delays and reveal that the peaks of curves are significantly shifted. Fractions smaller than 10^{-5} are ignored.

B

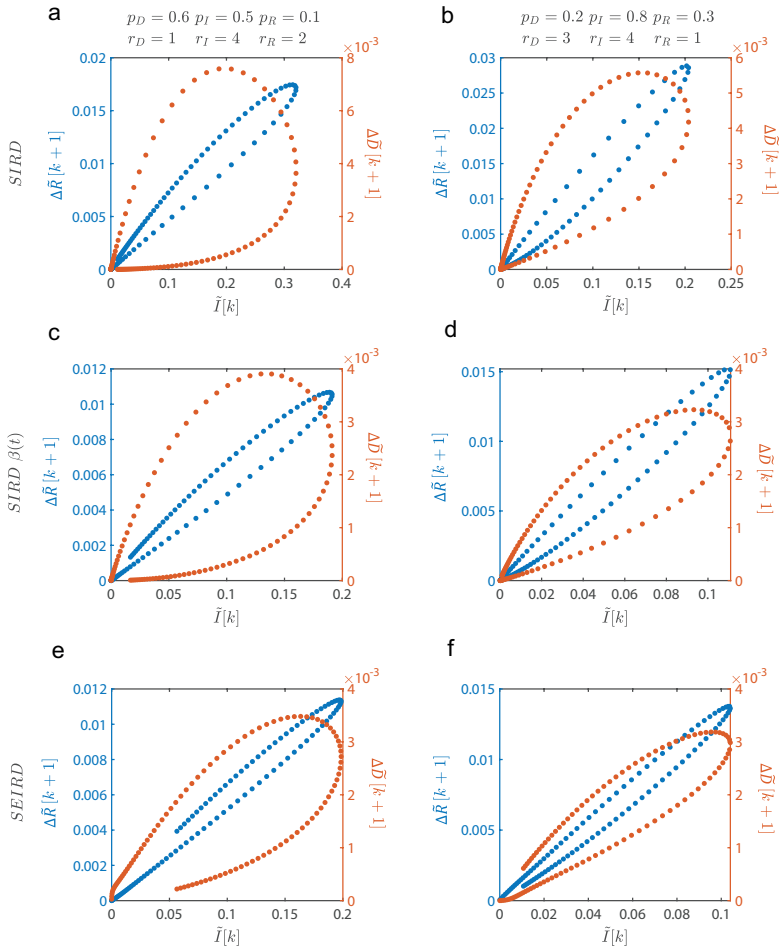


Figure B.7: The loop patterns for synthetic reported data for reporting delays following the negative binomial distributions. The shape and evolution direction (clockwise or counter-clockwise direction) of loop patterns can be different for different reporting delays and different epidemic models.

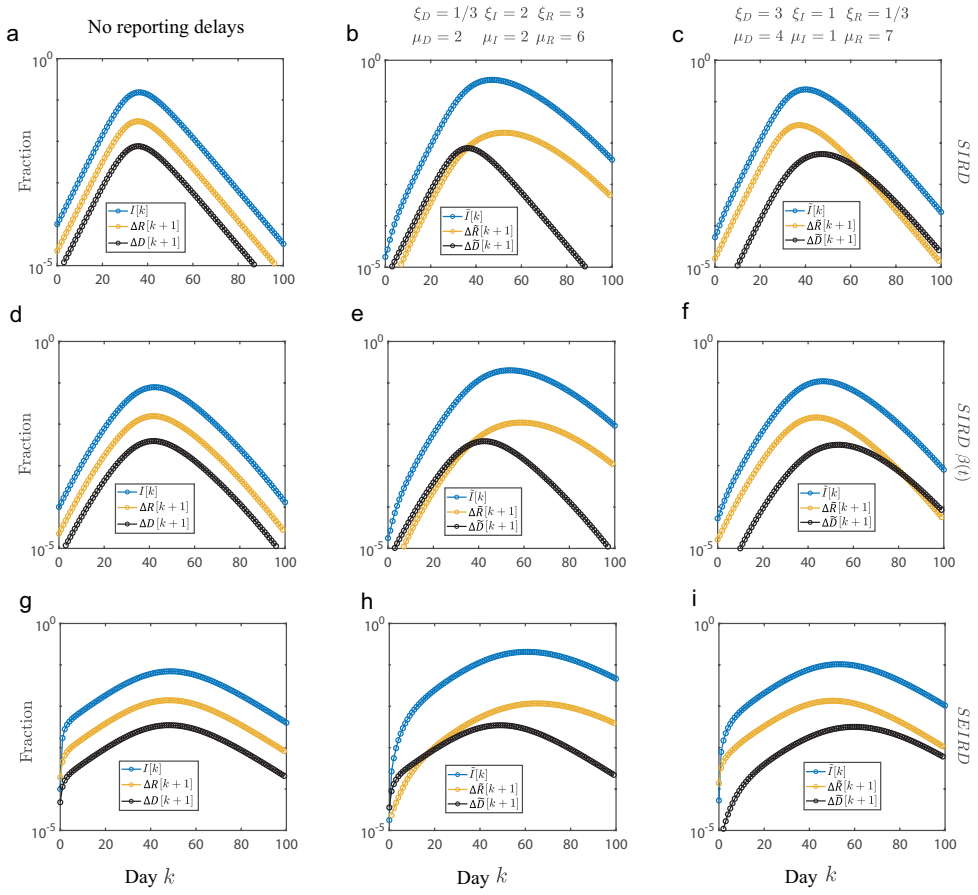


Figure B.8: Peak shift after reporting delays following the Pólya-Aeppli Distributions. Three different kinds of epidemic models are considered: the SIRD model, the SIRD model with time-varying infection rate $\beta(t) = 0.5 - 0.003t$ and the SEIRD model. Figures a, d and g show the curves of real data $\Delta R[k+1]$, $\Delta D[k+1]$ and $I[k]$. The other figures show the curves after reporting delays and reveal that the peaks of curves are significantly shifted. Fractions smaller than 10^{-5} are ignored.

B

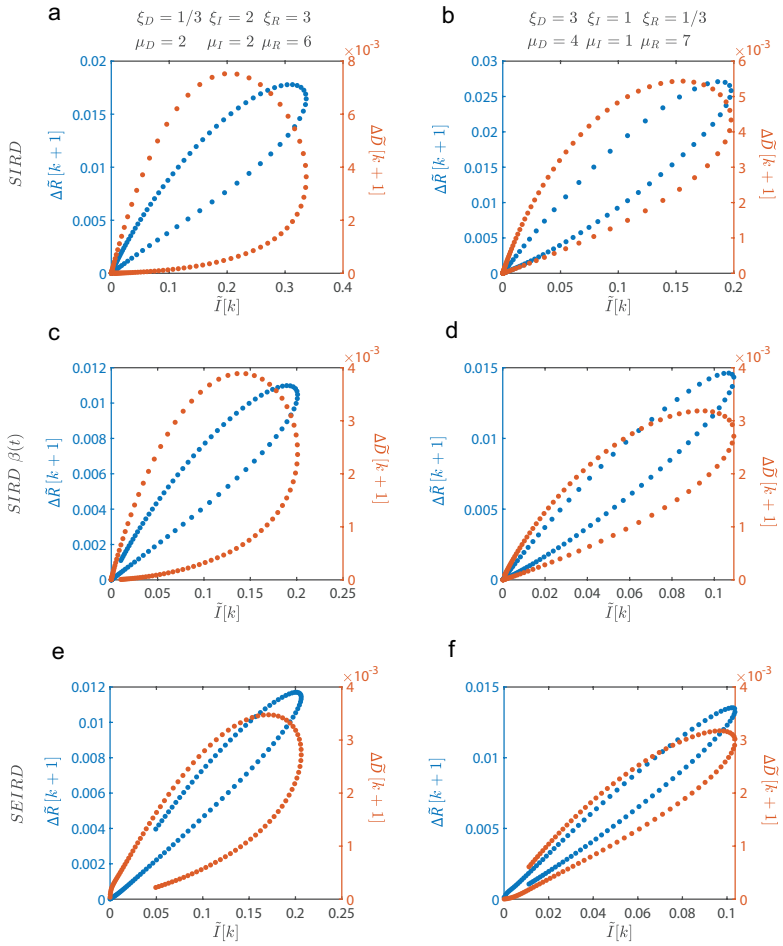


Figure B.9: The loop patterns for synthetic reported data for reporting delays following the Pólya-Aeppli distributions. The shape and evolution direction (clockwise or counter-clockwise direction) of loop patterns can be different for different reporting delays and different epidemic models.

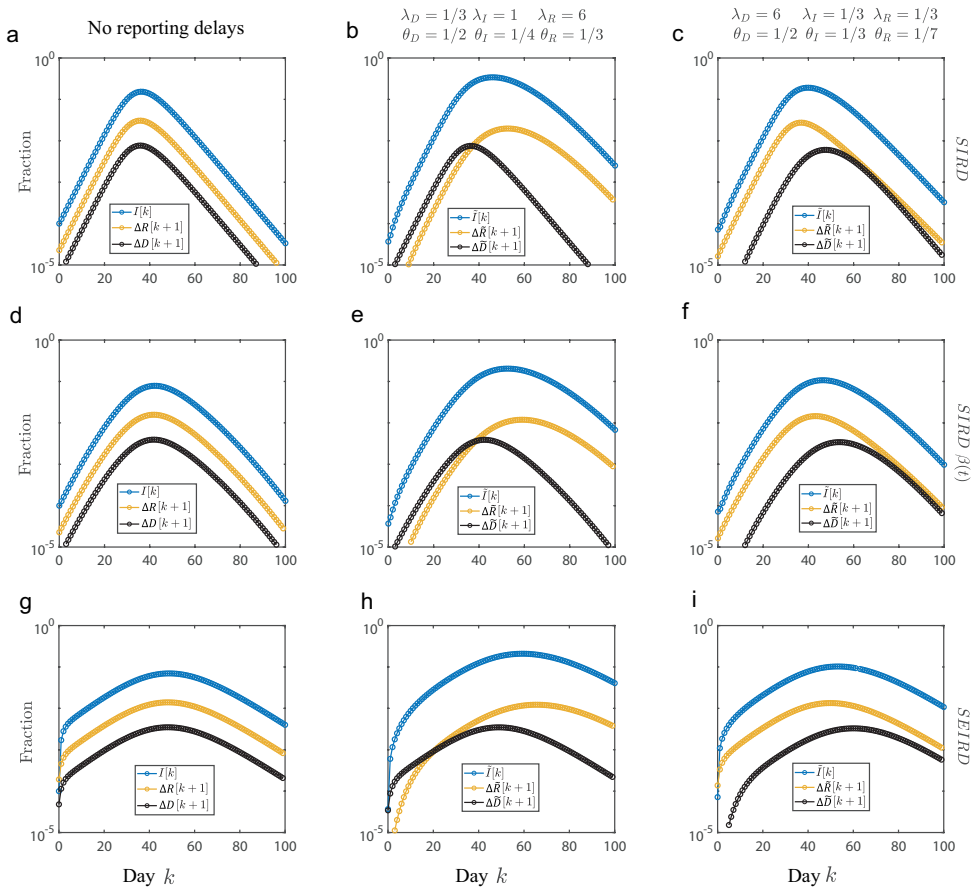


Figure B.10: Peak shift after reporting delays following the Neyman type A Distributions. Three different kinds of epidemic models are considered: the SIRD model, the SIRD model with time-varying infection rate $\beta(t) = 0.5 - 0.003t$ and the SEIRD model. Figures a, d and g show the curves of real data $\Delta R[k+1]$, $\Delta D[k+1]$ and $I[k]$. The other figures show the curves after reporting delays and reveal that the peaks of curves are significantly shifted. Fractions smaller than 10^{-5} are ignored.

B

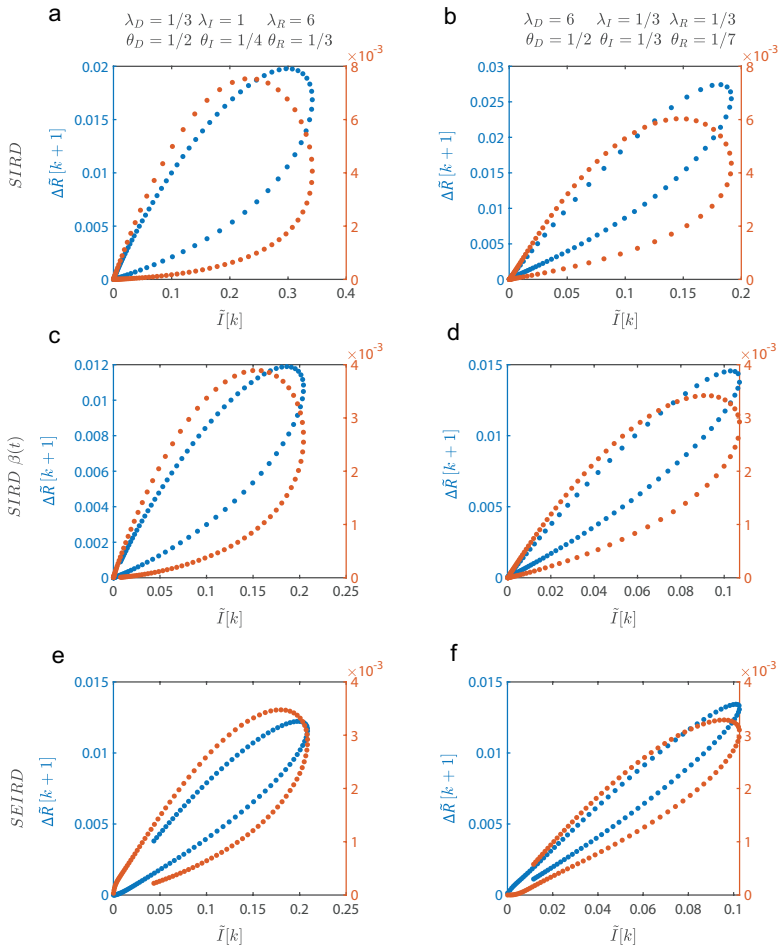


Figure B.11: The loop patterns for synthetic reported data for reporting delays following the Neyman type A distributions. The shape and evolution direction (clockwise or counter-clockwise direction) of loop patterns can be different for different reporting delays and different epidemic models.

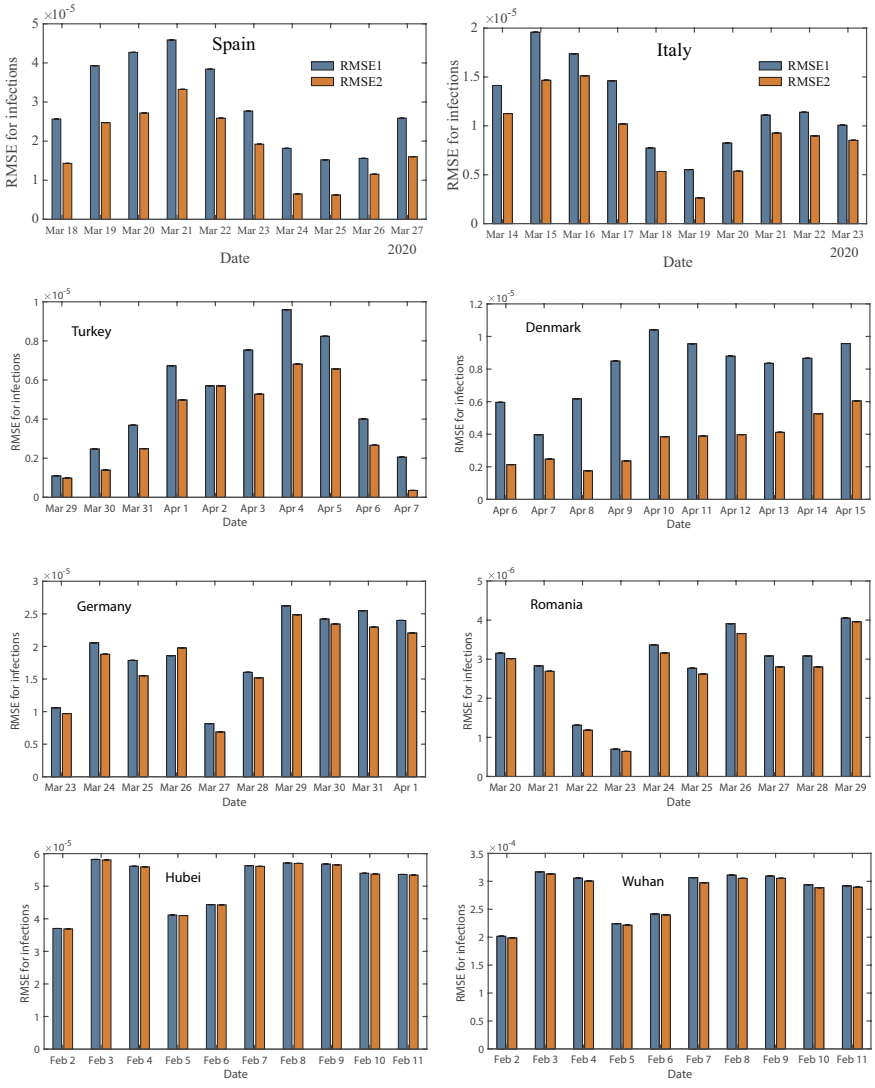


Figure B.12: Forecast the COVID-19 pandemic by the SIRD model considering and neglecting the the reporting delays. Each panel shows the RMSE between the forecast results and the real reported data when we forecast the future prevalence on different dates.

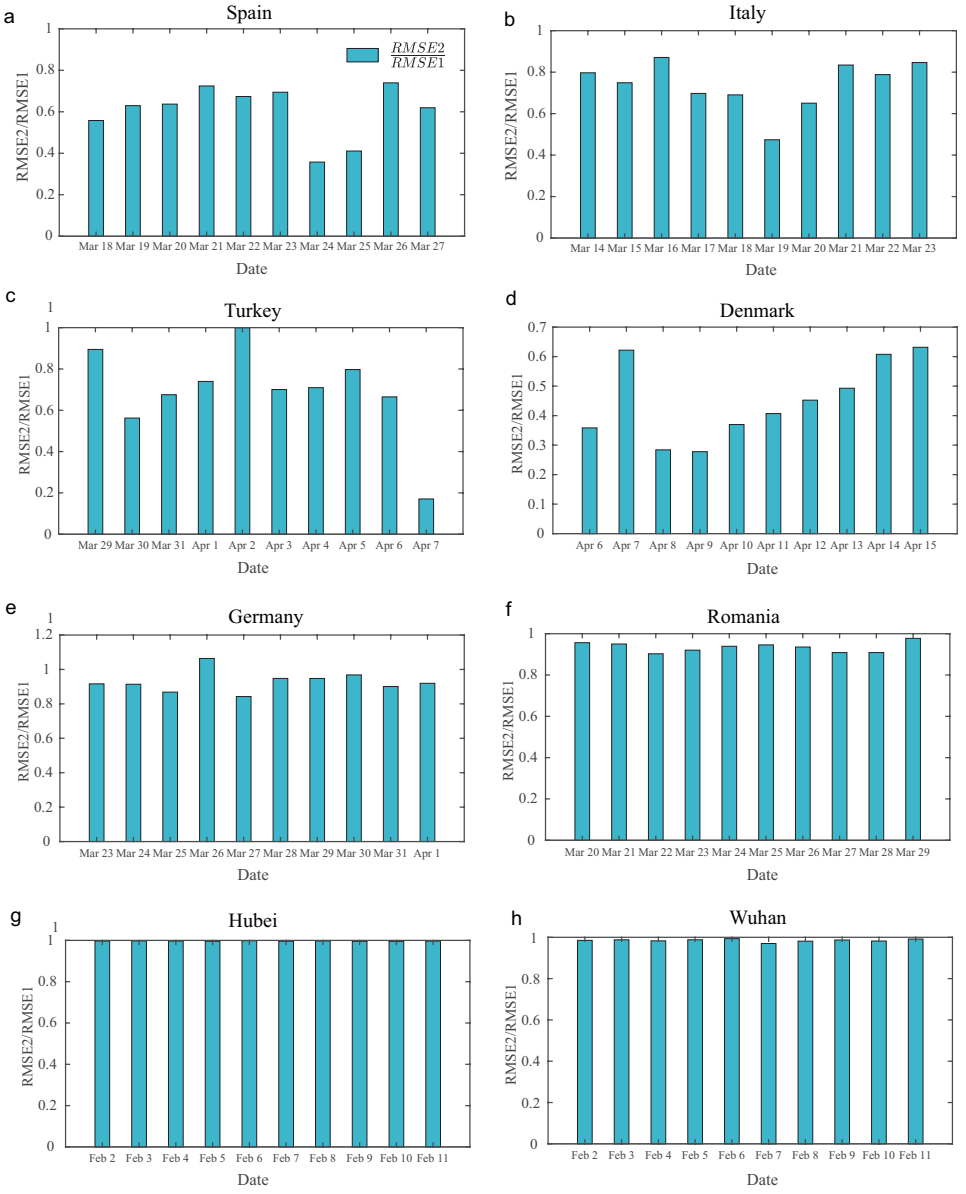


Figure B.13: The ratio between RMSE2 and RMSE1 for all 8 regions. The improvements of the forecast accuracy for Spain, Italy, Turkey and Denmark are more significant than the other regions.

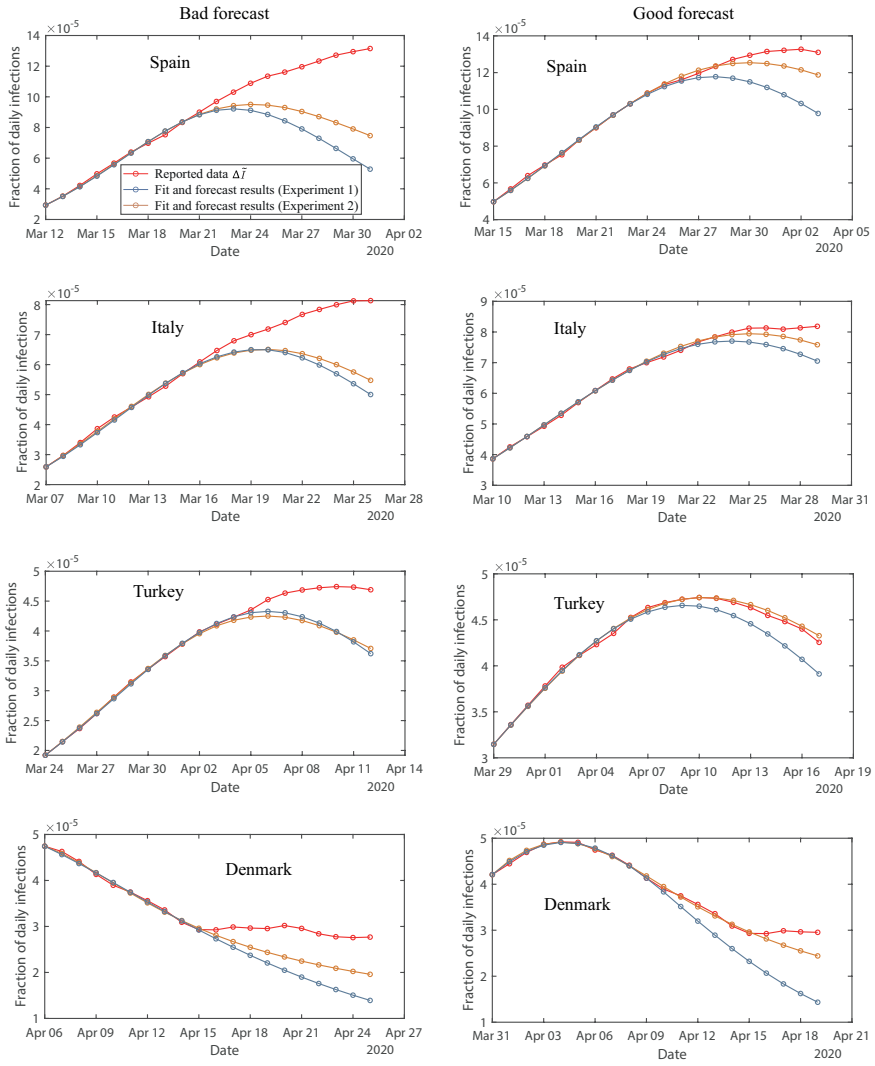


Figure B.14: Good and bad forecast results for each country. The improvements of forecast accuracy by considering the reporting delays are significant for some dates, but not significant for some other date.

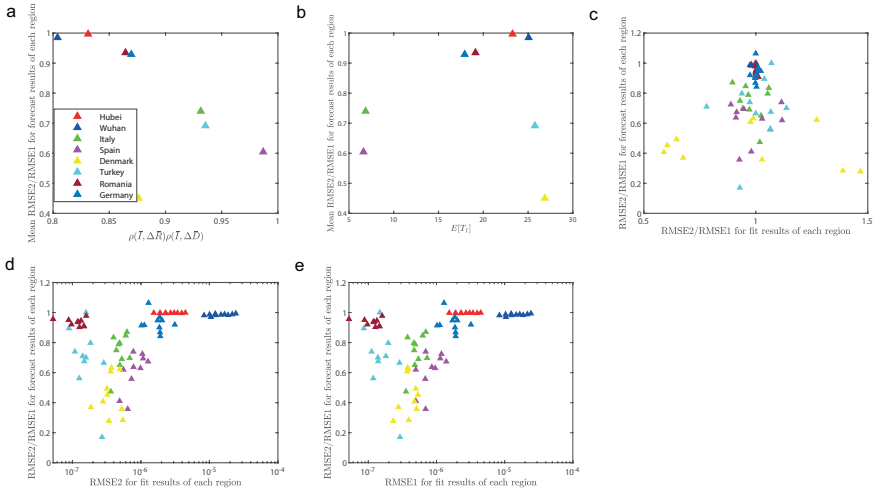


Figure B.15: Scatter plots between the ratio $RMSE2/RMSE1$ and $\rho(I,R)\rho(I,D)$ (a), $E[T_I]$(b), $EMSE2/RMSE1$(c), $RMSE2$(d) or $RMSE1$(e) for all 8 countries or regions. It indicates that there are some correlations between the mean reporting delays $E[T_I]$ and the ratio $RMSE2/RMSE1$, indicating that the bigger is the reporting delays, the worse we do, with the exception of Denmark. Besides, we also observe that the closer the objective function $\mathcal{O}(Y)$ to 1, the smaller is the forecast error.

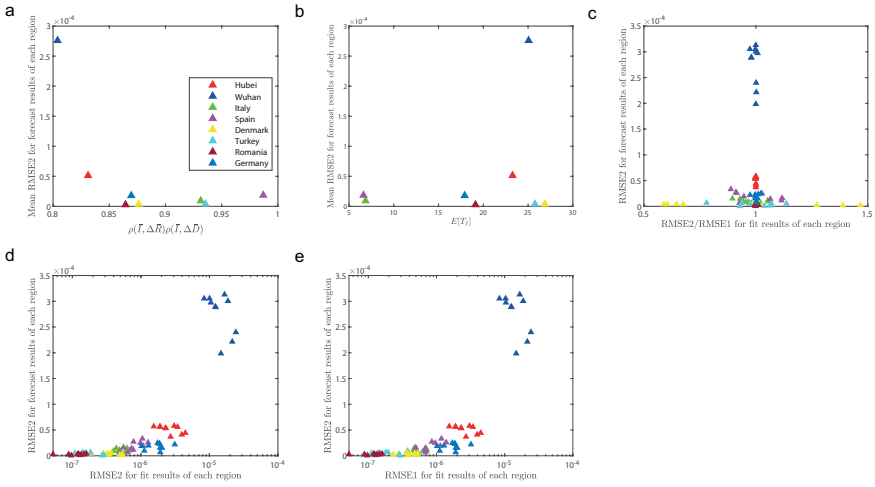


Figure B.16: Scatter plots between the error $RMSE2$ and $\rho(I,R)\rho(I,D)$ (a), $E[T_I]$(b), $EMSE2/RMSE1$(c), $RMSE2$(d) or $RMSE1$(e) for all 8 countries or regions. It shows that the better the SIRD model fit, the better is the forecast.

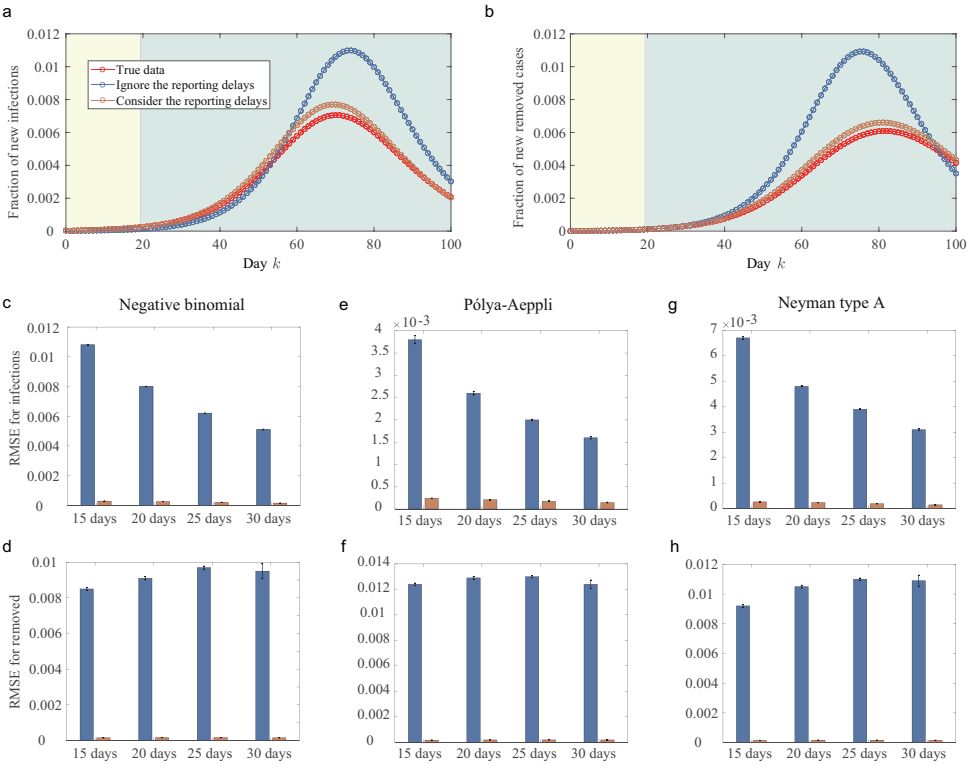


Figure B.17: We study the synthetic data. To get rid of the effect of the epidemic parameters, we generate 10^3 different synthetic data ΔY with the length of time series $n = 101$ by the SIRD model. For each time series, the infection rate β, the recovered rate γ_r and the mortality rate γ_d are uniformly randomly chosen in range $[0.5, 1]$, $[0.1, 0.4]$ and $[0, 0.1]$, respectively. The parameters of the reporting delay distributions are the same as Fig. B.4a-c. We fit the data for the first 15, 20, 25 or 30 days (the training dataset) and forecast the rest days' prevalence (the test dataset). In figures a and b, we fit the prevalence of the first 20 days (the training dataset) and forecast the prevalence of the other 81 days by considering and ignoring the reporting delays. It shows that the forecasting results turn to be much worse if we neglect the reporting delays. In figures c-h, we further separately fit the prevalence of the first 15, 20, 25 or 30 days and forecast the rest data. The forecast error is measured by RMSE. The figures show the mean RMSE of 10^3 independent experiments. The upper and bottom figures separately show the results for the fraction of reported new infections ΔI and the fraction of reported new removed (recovered or deceased) cases $\Delta R + \Delta D$. All results show that the negligence of the reporting delays results in a much greater error. The errorbar denotes the standard error (SE). There are mainly two possible reasons to explain why the improvement of forecast accuracy for the real data is less significant than the synthetic data: 1. the SIRD model is one of the most basic epidemic models and there are many more complicated epidemic models that could describe the COVID-19 pandemic more accurately; 2. the reporting delays for some regions can be not constant over time. A more realistic epidemic model and a deeper study of the reporting delays can be the future works to further improve the forecast accuracy by considering the reporting delays.

C

APPENDIX TO CHAPTER 4

C.1. THEORETICAL EXPLANATION OF THE DEATH-RELATED CURVE FEATURES

In Equations (4.1), we have that

$$\frac{dI_n(t)}{dt} = \beta_{nn} S_n(t) I_n(t) + \beta_{en} S_n(t) I_e(t) - \delta_n I_n(t) \tag{C.1}$$

$$= (\beta_{nn} I_n(t) + \beta_{en} I_e(t)) S_n(t) - \delta_n I_n(t). \tag{C.2}$$

Since non-elderly individuals are the majority in the whole population and the virus spreads from non-elderly individuals, it holds that $I_n(t) \gg I_e(t)$ at the initial stage of the spreading. Moreover, the infection rates satisfy $\beta_{nn} \gg \beta_{en}$ and thus we have $\beta_{nn} I_n(t) \gg \beta_{en} I_e(t)$, which indicates that elderly infections have little impact on the non-elderly susceptible individuals and the initial infection curve $I_n(t)$ for non-elderly individuals is close to the result in standard SIR model:

$$\begin{cases} \dfrac{dS_n(t)}{dt} = -\beta_{nn} S_n(t) I_n(t), \\ \dfrac{dI_n(t)}{dt} = \beta_{nn} S_n(t) I_n(t) - \delta_n I_n(t), \\ \dfrac{dR_n(t)}{dt} = \delta_n I_n(t). \end{cases} \tag{C.3}$$

This explains why curve features for non-elderly individuals are little affected by parameters ϵ and κ. When time $t \to 0$ and the inter-population infection rate $\beta_{ne} \to 0$, the fraction

for non-elderly infectious individuals is close to the exponential function

$$I_n(t) \approx I_n(0)e^{(\beta_{nn}S_n(0)-\delta_n)t}. \tag{C.4}$$

Substitute (C.4) into the equation $dI_e(t)/dt = \beta_{ne}S_e(t)I_n(t) + \beta_{ee}S_e(t)I_e(t) - \delta_e I_e(t)$ in (4.1) and we have that,

$$\begin{aligned}\frac{dI_e(t)}{dt} &\approx \beta_{ne}S_e(t)I_n(0)e^{(\beta_{nn}S_n(0)-\delta_n)t} + \beta_{ee}S_e(t)I_e(t) - \delta_e I_e(t)\\ &\approx \beta_{ne}S_e(0)I_n(0)e^{(\beta_{nn}S_n(0)-\delta_n)t} + (\beta_{ee}S_e(0) - \delta_e)I_e.\end{aligned} \tag{C.5}$$

We simplify the above equation by letting $a = \beta_{ne}S_e(0)I_n(0)$, $b = \beta_{ee}S_e(0) - \delta_e$ and $m = \beta_{nn}S_n(0) - \delta_n$:

$$\frac{dI_e}{dt} \approx ae^{mt} + bI_e(t). \tag{C.6}$$

By solving the above equation and combining the fact that $I_e(0) = 0$, the fraction of elderly infectious individuals when time $t \to 0$ is

$$I_e(t) \approx \frac{a(e^{mt} - e^{bt})}{m - b}. \tag{C.7}$$

This equation is the difference of two exponential functions, indicating that the initial curve for elderly individuals cannot be well described by an independent SIR model. Besides, at the initial stage of spreading, the growth rate of $I_e(t)$ decreases with the deceasing of β_{ne}. A slower growth of $I_e(t)$ at the initial stage will delay the further curve and peak position. Figure C.1 shows the values $\beta_{ee}I_e(t)$ and $\beta_{ne}I_n(t)$ in equation $dI_e(t)/dt = \beta_{ne}S_e(t)I_n(t) + \beta_{ee}S_e(t)I_e(t) - \delta_e I_e(t)$. It reveals that the value $\beta_{ne}I_n(t)$ dominates only at the very initial stage and the value $\beta_{ee}I_e(t)$ dominates the later stage. When the infection rate between two groups is relatively small, the later curve for elderly people will be close to the independent SIR model:

$$\begin{cases}\dfrac{dS_e(t)}{dt} = -\beta_{ee}S_e(t)I_e(t),\\ \dfrac{dI_e(t)}{dt} = \beta_{ee}S_e(t)I_e(t) - \delta_e I_e(t),\\ \dfrac{dR_e(t)}{dt} = \delta_n I_e(t).\end{cases} \tag{C.8}$$

Robert Schaback [228] proved that, for the independent SIR model, when the initial value $S_e(0) \approx N_e$ and $S_e(0)\beta_{ee} > \delta_e$, the upper bound of the peak position for the fraction of infectious individuals is

$$\frac{\delta_e}{I_e(0)\beta_{ee}} \log\left(\frac{S_e(0)\beta_{ee}}{\delta_e}\right). \tag{C.9}$$

The largest peak position can be obtained when β_{ee} is slightly larger than $\delta_e/N_e(0)$.

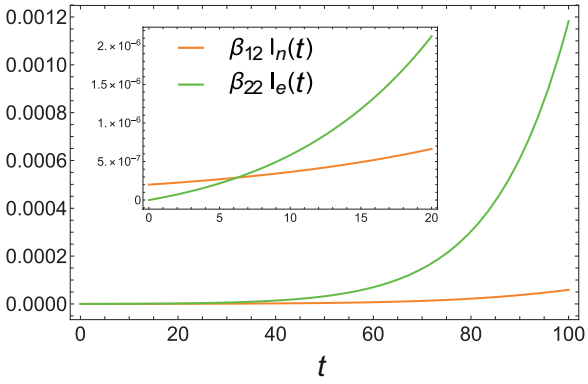


Figure C.1: The curves of $\beta_{ee} I_e(t)$ (green curves) and $\beta_{ne} I_n(t)$ (orange curves) for the two-population SIR model with parameters $\beta_{nn} = 0.2$, $\epsilon = 0.01$ and $\kappa = 4$. The value $\beta_{ee} I_e$ is much larger than $\beta_{ne} I_n$ after the very initial stage.

D

APPENDIX TO CHAPTER 5

D.1. THE ARGUMENT OF THE EIGENVALUES OF THE EXACT INFINITESIMAL GENERATOR MATRIX Q

We consider ER random networks with network size $N = 10$ and link probability $p = 0.1, 0.2, \cdots, 0.9$. For each graph, we choose the effective infection rate $\tau = x\tau_c^{(1)}$ where the NIMFA epidemic threshold $\tau_c^{(1)}$ is smaller than the true epidemic threshold τ_c. We consider $x = 0.1, 0.2, 0.3, \cdots, 10$ and let the curing rate $\delta = 1$. Figure D.1 show the argument of the eigenvalues of the infinitesimal generator matrix Q for the ER random networks with mean degree $E[D] = 0.9$, $E[D] = 4.5$ and $E[D] = 8.1$ and effective infection rates $\tau = 5\tau_c^{(1)}$ and $\tau = 10\tau_c^{(1)}$. It indicates that the real parts dominate the eigenvalues.

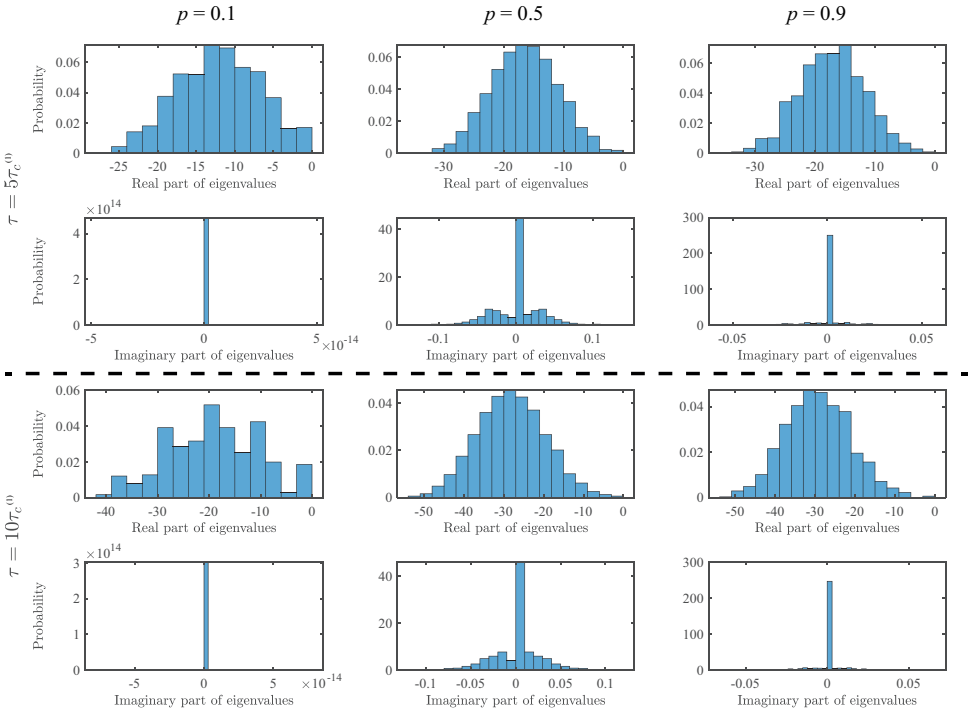


Figure D.1: Probability distributions of the real and imaginary parts of the eigenvalues of the infinitesimal generator matrix Q. In all cases, the imaginary parts are very small compared to the real parts.

D.2. PSEUDO-CODE OF SPECTRAL CLUSTERING SIS APPROXIMATION (SCSA)

Algorithm 5: Pseudo-code of spectral clustering SIS approximation (SCSA)

Input : Adjacency matrix A, infection rate β, curing rate δ, number of clusters r

Output: approximated prevalence $\tilde{y}(t)$

1 Add a small self-infection rate ϵ to the SIS process and compute the infinitesimal generator matrix Q.

2 Calculate the eigenvalues and eigenvectors of the weighted Laplacian matrix $-Q$.

3 Perform k-means clustering based on the real part of the eigenvectors corresponding to the r smallest eigenvalues.

4 Combine states with the same number of infected nodes and the same cluster into one combined state. Compute the transition rate between any two partitioned states using Eq. (5.4).

5 Generate the approximated prevalence by (5.5).

D.3. APPROXIMATION WITH AND WITHOUT THE BIRTH-AND-DEATH RESTRICTION

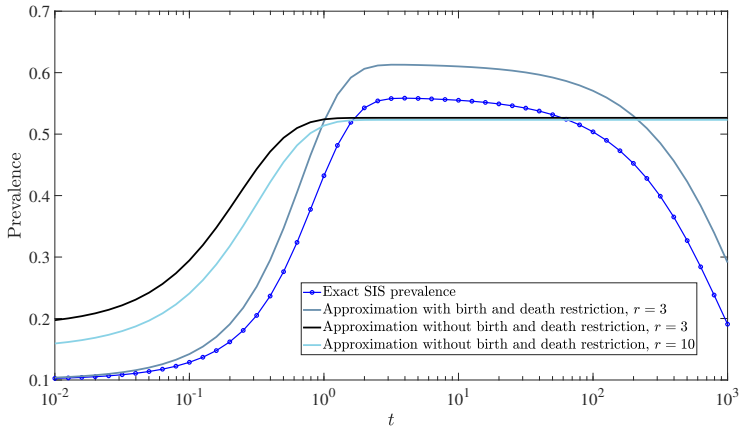


Figure D.2: The birth-and-death process restriction is important in SCSA, especially for the prevalence before and after the metastable state.

D.4. PERFORMANCE ANALYSIS OF SCSA

The spectral clustering method reduces the error ϵ, but also increases the size of the partitioned infinitesimal generator $\tilde{Q}$. To access the quality of SCSA, we compare the SCSA results with a random clustering method, which places each state in a uniform random cluster. As shown in Fig. D.3, with the same size of the partitioned state space, the approximate accuracy of SCSA is significantly higher than the random clustering benchmark.

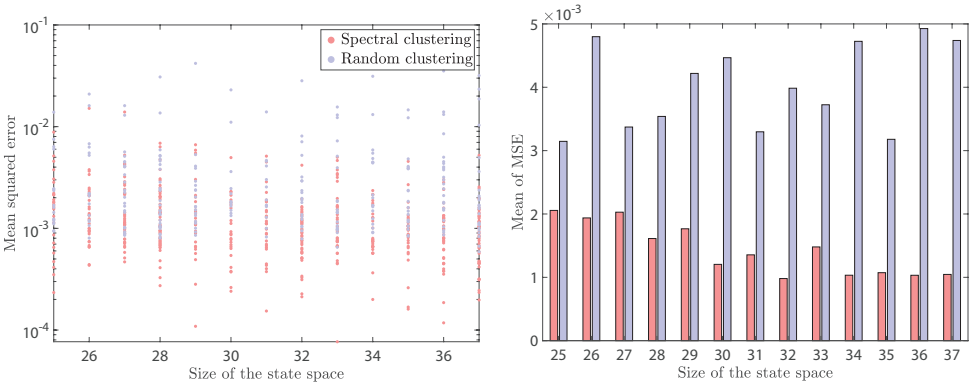


Figure D.3: A comparison of the spectral clustering SIS approximation with the random clustering approximation. The approximate accuracy of the spectral clustering based results is significantly higher than the random clustering based results.

The red line in Fig. D.4 shows the error ϵ of the spectral clustering SIS approximation when the size of the state space is 14. It is infeasible to go through all possible clustering results even for small graphs and thus we randomly select 10^5 possible clustering ways with the same size of the state space to check the approximation performance of SCSA.

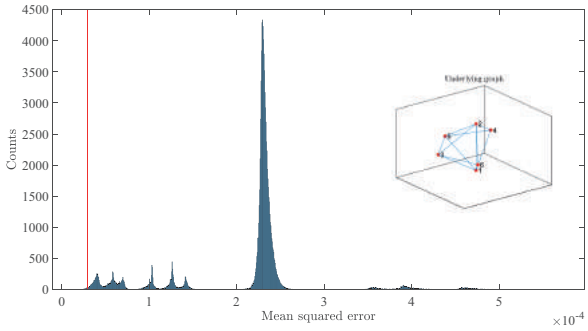


Figure D.4: Histogram of the mean squared error ϵ of 10^5 randomly selected clustering results. We focus on one graph with 6 nodes and 12 links as shown in the insert. The red line shows the error of the spectral clustering SIS approximation when the size of the state space is 14. We randomly select 10^5 possible clustering ways with the size of the state space equals 14. It reveals that the spectral clustering SIS approximation result is close to the upper bound of the approximation accuracy.

D.5. APPROXIMATION ERROR FOR NETWORKS WITH DIFFERENT LINK DENSITIES

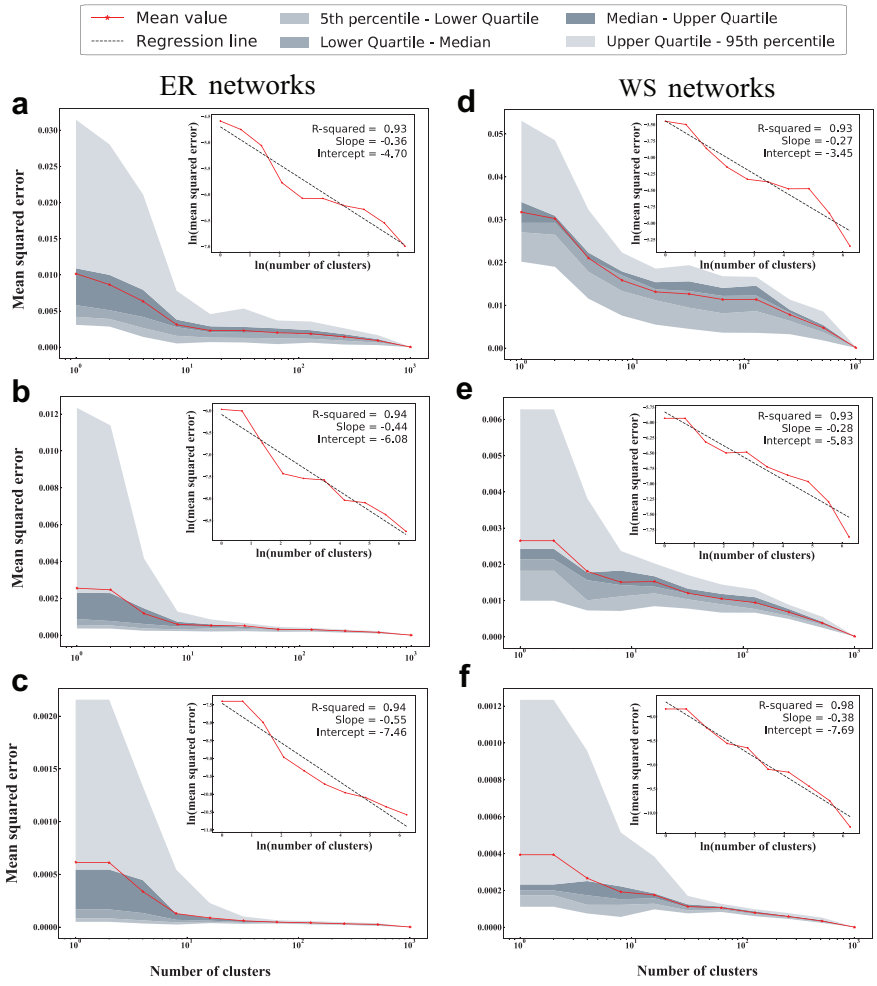


Figure D.5: Mean squared error ϵ between the approximated prevalence and the exact prevalence for the ER random networks with network size $N = 10$ and different number of links: $L = 14$ (a), $L = 24$ (b) and $L = 34$ (c) and for the WS networks with average degree $E[D] = 3$ (d), $E[D] = 5$ (e) and $E[D] = 7$ (f). Note that all plots have a horizontal logarithmic scale. The effective infection rate $\tau = 5\tau_c^{(1)}$. The mean and percentiles of the error are obtained by considering 1000 randomly generated networks. The curing rate equals $\delta = 1$.

D.6. APPROXIMATION ERROR FOR NETWORKS WITH DIFFERENT NETWORK SIZES

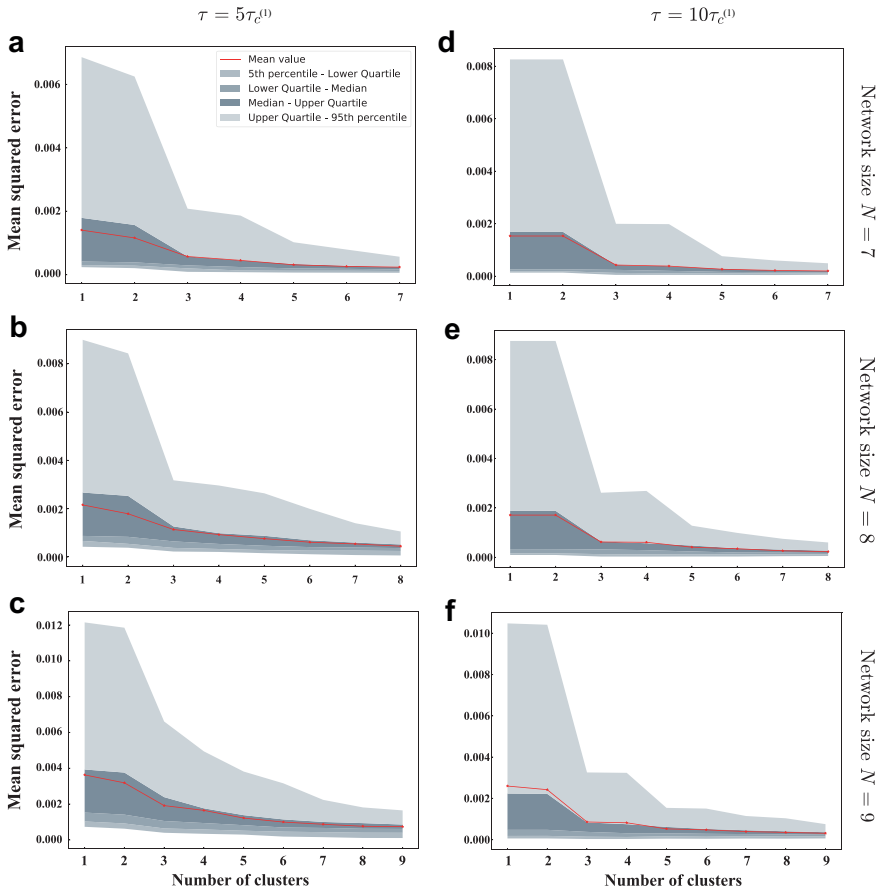


Figure D.6: Mean squared error ϵ between the approximated prevalence and the exact prevalence for the ER random networks with different network sizes. The average degree is $E[D] = 4$. The mean and percentiles of the error are obtained by considering 1000 randomly generated networks. The curing rate equals $\delta = 1$.

D.7. NUMBER OF INFECTIOUS LINKS

To understand why spectral clustering works well, we propose the *number of infectious links* that partially describes the accuracy improvements of the spectral clustering method. The number of infectious links is closely related to the cut-set, which is defined as the number of infectious links at time t, evolves and defines the prevalence [32, 223, 224]. To

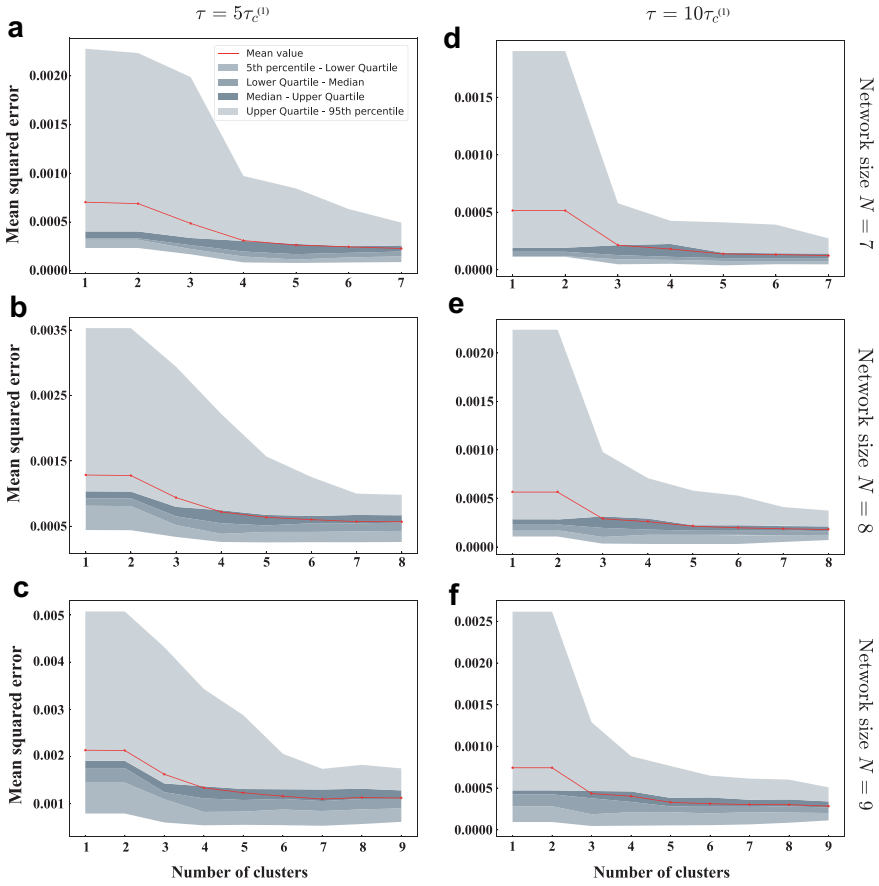


Figure D.7: Mean squared error ϵ between the approximated prevalence and the exact prevalence for the WS networks with different network sizes. The average degree is $E[D] = 4$. The mean and percentiles of the error are obtained by considering 1000 randomly generated networks. The curing rate equals $\delta = 1$.

calculate the number of infectious links. We consider two network infection states in the infinitesimal generator matrix Q with the same number of infections (in the same layer of the state transition rate diagram as shown in Fig. 5.3). We consider two link types: one is the link with a susceptible node at one end and an infected node at another end (name as S-I link); another is the link with infected nodes at both end (name as I-I link). The S-I link can be transferred into the I-I link by one time infection. There are different S-I links and I-I links in different states. For any two states, the number of infectious links is the defined as the number of links that are S-I links in one state but are I-I links in another state. To be specific, for networks infection states A and B, we consider any link that is S-I link in state A but is I-I link in state B, or that is I-I link in state A but is S-I link in state B. The diagram to calculate the number of infectious links is shown in Fig. D.8.

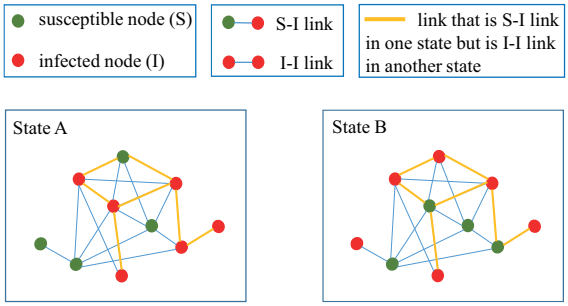


Figure D.8: Illustration of the number of infectious links between two network infection states A and B. The example network has 10 nodes and 20 links, with 4 susceptible nodes (green) and 6 infected nodes (red), but the location of the infected nodes is different for states A and B. We focus on two link types: the S-I links connect a susceptible node to an infected node and I-I links connect two infected nodes to each other. S-I links can be transferred into I-I links if an infection takes place over that link. Now, the *number of infectious links* between state A and B is the number of links that are S-I links in state A but I-I links in state B or links that are I-I links in state A but S-I links in state B.

We apply the receiver operating characteristic (ROC) curves to measure the performance of the number of infectious links to classify whether two states are in the same cluster. We consider 100 randomly generated ER networks with 10 nodes and 20 links. We randomly select 10 two states from the Markov chain with the same number of infected nodes and use the spectral clustering algorithm to determine if these states belong to the same cluster. The ROC curves of the number of infectious links are compared with the ROC curves of the common infected nodes between two states. The area under the receiver operating characteristic (AUROC) is a metric to evaluate classification performance. Figure D.9 indicates that the number of infectious links can better predict the spectral clustering results. The number of infectious links partially explained the spectral clustering results. Future more works could discover better metrics to obtain the clusters

that are close to the spectral clustering results, which can inspire more accurate heuristic approximation approaches of the SIS network epidemic.

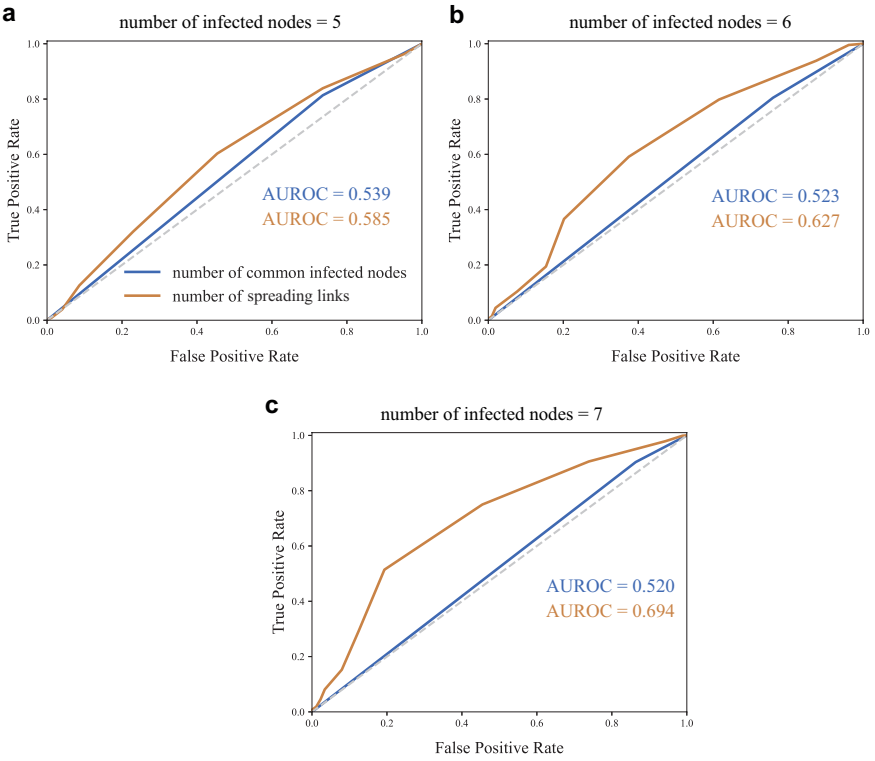


Figure D.9: Receiver operating characteristic (ROC) curves of the number of infectious links and the number of common infected nodes to predict whether two states are in the same cluster. The simulations are averaged over 100 randomly generated ER networks with 10 nodes and 20 links. Out of 10 nodes, we consider states with (a) 5 infected nodes, (b) 6 infected nodes and (c) 7 infected nodes. The effective infection rate is $\tau = 5\tau_c^{(1)}$, the curing rate $\delta = 1$ and the number of clusters $r = 5$.

E

APPENDIX TO CHAPTER 6

E.1. THE CLIFFORD GATES ARE A GROUP

The fact that $\mathscr{C}_n$ is a group can be verified by checking the necessary properties:

Binary operation. Suppose $A, B \in \mathscr{C}_n$. Thus for all $\sigma \in \pm P_n^*$, $A\sigma A^\dagger \in \pm P_n^*$ and $B\sigma B^\dagger \in \pm P_n^*$; and moreover $A(B\sigma B^\dagger)A^\dagger \in \pm P_n^*$. Let $\sigma \in \pm P_n^*$ be arb. and note that we have shown that $(AB)\sigma(AB)^\dagger \in \pm P_n^*$. Thus $AB \in \mathscr{C}_n$.

Associativity. This is for free because matrix multiplication is associative.

Identity. $I^{\otimes n} \in \mathscr{C}_n$ because it is unitary and for all $\sigma \in \pm P_n^*$, $I^{\otimes n}\sigma(I^{\otimes n})^\dagger = \sigma$.

Inverses. Suppose $C \in \mathscr{C}_n$, such that for any $\sigma \in \pm P_n^*$ we have that $C\sigma C^\dagger \in \pm P_n^*$. This implies that for any $\omega \in \pm P_n^*$, we can find a $\sigma \in \pm P_n^*$ such that $\omega = C\sigma C^\dagger$ (isomorpishm). Conclude that because C is unitary, $C^{-1}\omega(C^{-1})^\dagger = C^\dagger \omega C = C^\dagger C\sigma C^\dagger C = \sigma \in P_n^*$. Hence $C^{-1} \in \mathscr{C}_n$. □

E.2. RELATION BETWEEN THE ERROR PROBABILITIES WHEN USING THE TRACE DISTANCE AND FIDELITY

Let t be such that $0 \leq t \leq \tau$ and let $\omega \in \{D_t \leq \varepsilon\} = \{1 - D_t \geq 1 - \varepsilon\}$. By [3, (9.110)], we have that $1 - F_t \leq D_t \leq \sqrt{1 - F_t^2}$ for all $t \geq 0$. Consequentially $1 - D_t \leq F_t \leq \sqrt{1 - D_t^2}$ for all $t \geq 0$. On every such ω, we thus also have that $F_t \geq 1 - \varepsilon$. We have shown that $\{D_t \leq \varepsilon\} \subseteq \{F_t \geq 1 - \varepsilon\}$, which proves the first statement. For the second statement, we similarly note that $\{\min_{0 \leq s \leq t} F_s \geq 1 - \varepsilon\} \supseteq \{\min_{0 \leq s \leq t}(1 - D_s) \geq 1 - \varepsilon\} = \{\max_{0 \leq s \leq t} D_s \leq \varepsilon\}$. □

E.3. AN EXAMPLE EXPLICIT CALCULATION OF THE RESULTS IN PROPOSITION 6 AND LEMMA 2

In order to calculate the results of Proposition 6 or Lemma 2, one requires a transition matrix P. For the example of randomized benchmarking in Section 6.4.1 with Pauli channels $\{I, X, Y, Z\}$ only, when enumerating the two-dimensional states

$$\mathscr{G}_1^2 = \{(I,I),(I,X),(I,Y),(I,Z),(X,I),(X,X),...,(Z,Z)\} \tag{E.1}$$

lexicographically along both the rows and columns (so as indicated), the transition matrix P is represented as follows:

$$P = \begin{bmatrix}
\frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} \\
\frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} \\
\frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} \\
\frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} \\
\frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} \\
\frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} \\
\frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} \\
\frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} \\
\frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} \\
\frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} \\
\frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} \\
\frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} \\
\frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} \\
\frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} \\
\frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} \\
\frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} \\
\frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} \\
\frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} \\
\frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} \\
\frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} \\
\frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{r}{12} & \frac{1-r}{4}
\end{bmatrix} \tag{E.2}$$

Probability distribution of the maximum trace distance. Using (E.2), we can evaluate Proposition 1's results. When starting from the initial state $|\zeta_0\rangle$, (6.11) simplifies (after some algebra) to

$$\mathbb{P}[\max_{0\le s\le t} D_s > 1/5] = 1 - \left(1 - \tfrac{2}{3}r\right)^t. \tag{E.3}$$

Similarly when starting from the initial state $|\xi_0\rangle$, (6.11) leads to

$$\mathbb{P}[\max_{0\le s\le t} D_s > 1/5] = 1 - (1-r)^t. \tag{E.4}$$

Lower bound. Using (E.2), we can also evaluate the lower bound in Lemma 2. When the initial state $|\zeta_0\rangle = \sqrt{7/10}|0\rangle + \sqrt{3/10}|1\rangle$, the expected hitting time of $\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,1/5}$ turns out to be given by $\mathbb{E}_z[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,1/5}}] = 3/(2r)$. The lower bound in (6.17) therefore reads

$$\mathbb{P}_{z_0}[\max_{0\le s\le t} D_s > 1/5] \ge 0 \vee \left(1 - \frac{3}{2r(t+1)}\right). \tag{E.5}$$

Here $a \vee b \triangleq \max\{a, b\}$. Alternatively, when the initial state $|\xi_0\rangle = \sqrt{4/5}|0\rangle + \sqrt{1/5}|1\rangle$, the expected hitting time of $\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,1/5}$ can be calculated to be $\mathbb{E}_z[T_{\mathscr{B}^{|\Psi_0\rangle}_{|\psi_0\rangle,1/5}}] = 1/r$. The lower bound is thus given by

$$\mathbb{P}_{z_0}[\max_{0\leq s\leq t} D_s > 1/5] \geq 0 \vee \left(1 - \frac{1}{r(t+1)}\right). \tag{E.6}$$

Comparison of the probability distribution of the maximum trace distance to its lower bound The lower bounds and exact results with $r = 0.2$ are shown in Figure E.1.

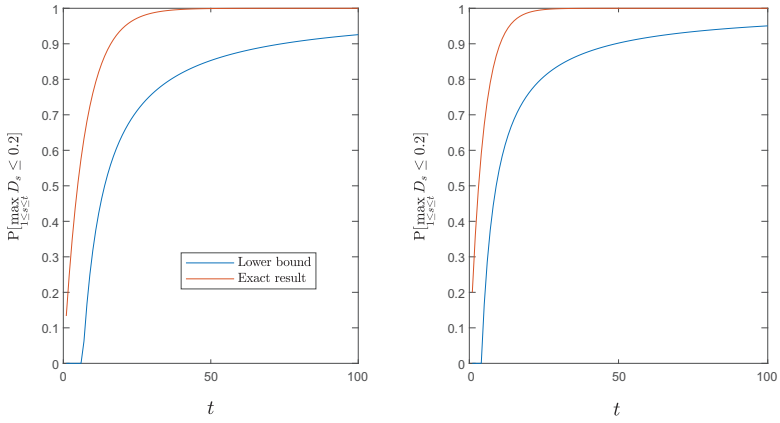


Figure E.1: Lower bounds and exact probabilities $\mathbb{P}[\max_{0\leq s\leq t} D_s > 1/5]$ with $r = 0.2$ for initial state $|\zeta_0\rangle$ (left) and $|\xi_0\rangle$ (right).

On how to construct a P or Q matrix. To assist you in constructing a transition matrix P or Q, which are needed for the results in Section 6.2, we have written an `R` script that can generate such matrices. The script generates a transition matrix when you a scenario with Pauli and Clifford channels and with error probabilities that are either dependent or independent of the gate: all you as user have to do, is to input a vector of (gate-dependent) error probabilities. The code of this script can be found at `https://gitlab.tue.nl/20061069/markov-chains-for-error-accumulation-in-quantum-circuits`. Additionally, for as long as the following public service remains available, the script can be tried out at `https://bevanschooten.shinyapps.io/qbiterrors/`.

E.4. NUMBER OF STABILIZER STATES FOR A GATE

For n qubits, any gate $\mathscr{M} \in \mathscr{G}_n \setminus I^{\otimes n}$ can be represented using a $2^n \times 2^n$ unitary matrix. Recall that any unitary matrix of finite size is unitarily diagonalizable since every unitary

matrix is normal [229]. A $2^n \times 2^n$ matrix that is diagonalizable must have a set of 2^n linearly independent eigenvectors [229].

The initial states $|\psi_0\rangle$ that can satisfy $\mathcal{M}|\psi_0\rangle = e^{i\gamma}|\psi_0\rangle$ are the eigenvectors of the matrix $\mathcal{M}$ with eigenvalue $\lambda = e^{i\gamma}$. For any unitary matrix A with eigenvalue λ and eigenvector v, $A^\dagger A = AA^\dagger = I$, $v^\dagger v = v^\dagger A^\dagger A v = \lambda^\dagger v^\dagger v \lambda = \lambda^\dagger \lambda v^\dagger v$. Also recall that any eigenvector $||v|| \neq 0$ by definition [229] and thus it always holds that $|\lambda| = 1$. So $\mathcal{M}|\psi_0\rangle = \lambda|\psi_0\rangle = e^{i\gamma}|\psi_0\rangle$. □

E.5. A STABILIZER STATE FOLLOWS AFTER A STABILIZER STATE

By assumption and the definition in (6.37), for any state $|\psi_1\rangle \in \mathcal{R}_{|\psi_0\rangle}$, $\exists \mathcal{Z} \in \mathcal{G}_n : |\psi_1\rangle = \mathcal{Z}|\psi_0\rangle$ since $\mathcal{G}_n$ is a group. we have furthermore that $\exists \mathcal{H} \in \mathcal{G}_n \setminus I^{\otimes n} : \mathcal{H}\mathcal{Z} = \mathcal{Z}\mathcal{M}$. Then $|\psi_1\rangle = \mathcal{Z}|\psi_0\rangle = e^{-i\gamma}\mathcal{Z}\mathcal{M}|\psi_0\rangle = e^{-i\gamma}\mathcal{H}\mathcal{Z}|\psi_0\rangle = e^{-i\gamma}\mathcal{H}|\psi_1\rangle$. So $\mathcal{H}|\psi_1\rangle = e^{i\gamma}|\psi_1\rangle$. □

E

E.6. GATE-DEPENDENT ERROR MODEL

In Table E.1, we provide the precise error probabilities used in Section 6.4. The specific values were simply randomly generated to result in an inhomogeneous example; we spent no time post-selecting these values.

E.7. METHOD TO FIND ALL REACHABLE STABILIZER STATES

All reachable stabilizer states can be found given the finite unitary group $\mathcal{G}_n$ of gates (and noise) and the initial stabilizer state $|\psi_0\rangle$. Given an initial stabilizer state $|\psi_0\rangle$, the reduced states can be found by the following steps. First list all gates (and noise) $\{\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_n\}$ in group $\mathcal{G}_n$. All reachable states are then $\{\mathcal{M}_1|\psi_0\rangle, \mathcal{M}_2|\psi_0\rangle, \cdots, \mathcal{M}_n|\psi_0\rangle\}$. At last, any two states $\mathcal{M}_i|\psi_0\rangle$ and $\mathcal{M}_j|\psi_0\rangle$ that satisfies $\mathcal{M}_i|\psi_0\rangle = e^{i\gamma}\mathcal{M}_j|\psi_0\rangle$ will fall into the same state.

E.8. PSEUDO-CODE FOR GATE-LIMITED SIMULATED ANNEALING

In Algorithm 6, we present the pseudo-code for the simulated annealing algorithm when restricting to a subset of available gates.

E.9. IMPROVED QUANTUM CIRCUITS

In Figure E.2, we present the circuits with the lowest error accumulation rates found by our implementations of the two simulated annealing algorithms.

Table E.1: The specific error probabilities. Here, $C_1, C_2, \ldots, C_{24}$ denote the single-qubit Clifford gates and refer specifically to the representation of these gates in [180] and [182].

Gate	$\mathbb{P}[\Lambda = I \mid C_i]$	$\mathbb{P}[\Lambda = X \mid C_i]$	$\mathbb{P}[\Lambda = Y \mid C_i]$	$\mathbb{P}[\Lambda = Z \mid C_i]$
C_1	0.990	$0.00\dot{3}$	$0.00\dot{3}$	$0.00\dot{3}$
C_2	0.965	$0.012\dot{3}$	$0.010\dot{3}$	$0.012\dot{3}$
C_3	0.983	$0.004\dot{3}$	$0.008\dot{3}$	$0.004\dot{3}$
C_4	0.977	$0.008\dot{3}$	$0.010\dot{3}$	$0.004\dot{3}$
C_5	0.969	$0.011\dot{3}$	$0.007\dot{3}$	$0.012\dot{3}$
C_6	0.984	$0.006\dot{3}$	$0.004\dot{3}$	$0.005\dot{3}$
C_7	0.979	$0.004\dot{3}$	$0.01\dot{3}$	$0.00\dot{3}$
C_8	0.987	$0.004\dot{3}$	$0.003\dot{3}$	$0.005\dot{3}$
C_9	0.979	$0.00\dot{3}$	$0.009\dot{3}$	$0.008\dot{3}$
C_{10}	0.985	$0.005\dot{3}$	$0.005\dot{3}$	$0.004\dot{3}$
C_{11}	0.980	$0.007\dot{3}$	$0.00\dot{3}$	$0.009\dot{3}$
C_{12}	0.975	$0.008\dot{3}$	$0.006\dot{3}$	$0.010\dot{3}$
C_{13}	0.974	$0.011\dot{3}$	$0.006\dot{3}$	$0.008\dot{3}$
C_{14}	0.975	$0.007\dot{3}$	$0.006\dot{3}$	$0.011\dot{3}$
C_{15}	0.972	$0.01\dot{3}$	$0.009\dot{3}$	$0.005\dot{3}$
C_{16}	0.980	$0.004\dot{3}$	$0.009\dot{3}$	$0.006\dot{3}$
C_{17}	0.979	$0.006\dot{3}$	$0.009\dot{3}$	$0.005\dot{3}$
C_{18}	0.982	$0.010\dot{3}$	$0.004\dot{3}$	$0.00\dot{3}$
C_{19}	0.977	$0.006\dot{3}$	$0.004\dot{3}$	$0.012\dot{3}$
C_{20}	0.976	$0.011\dot{3}$	$0.007\dot{3}$	$0.010\dot{3}$
C_{21}	0.975	$0.007\dot{3}$	$0.07\dot{3}$	$0.010\dot{3}$
C_{22}	0.967	$0.007\dot{3}$	$0.007\dot{3}$	$0.010\dot{3}$
C_{23}	0.974	$0.01\dot{3}$	$0.006\dot{3}$	$0.006\dot{3}$
C_{24}	0.978	$0.012\dot{3}$	$0.005\dot{3}$	$0.004\dot{3}$

Algorithm 6: Pseudo-code for gate-limited simulated annealing.

Input: A group $\mathscr{G}_n$, a set $\mathscr{A} \subseteq \mathscr{G}_n$, a circuit $\{U_1, \ldots, U_\tau\}$ and number of iterations w

Output: A revised circuit $\{G_1^{[w]}, \ldots, G_\tau^{[w]}\}$

begin
  Initialize $\{G_1^{[0]}, \ldots, G_\tau^{[0]}\} = \{U_1, \ldots, U_\tau\}$;
  for $\eta \leftarrow 1$ **to** w **do**
    Collect all m neighboring gates $\{(G_1^{[\eta]}, G_2^{[\eta]}), \ldots, (G_{m-1}^{[\eta]}, G_m^{[\eta]})\}$ with at least one replaceable candidate neighboring gates $\{G_w^{[\eta+1]} \in \mathscr{A}, G_{w+1}^{[\eta+1]} \in \mathscr{A}\}$;
    Choose $I \in [m-1]$ uniformly at random;
    Replace $(G_I^{[\eta]}, G_{I+1}^{[\eta]})$ by any gate pair in $\{(\tilde{G}_1, \tilde{G}_2) \in \mathscr{A}^2 \mid G_{I+1}^{[\eta]} G_I^{[\eta]} = \tilde{G}_2 \tilde{G}_1\}$ uniformly at random and then obtain the new circuit J;
    Choose $X \in [0, 1]$ uniformly at random;
    if $X \leq \alpha_{G^{[\eta]}, J}(T_\eta)$ **then**
      Set $G^{[\eta+1]} = J$;
    else
      Set $G^{[\eta+1]} = G^{[\eta]}$;
    end
  end
end

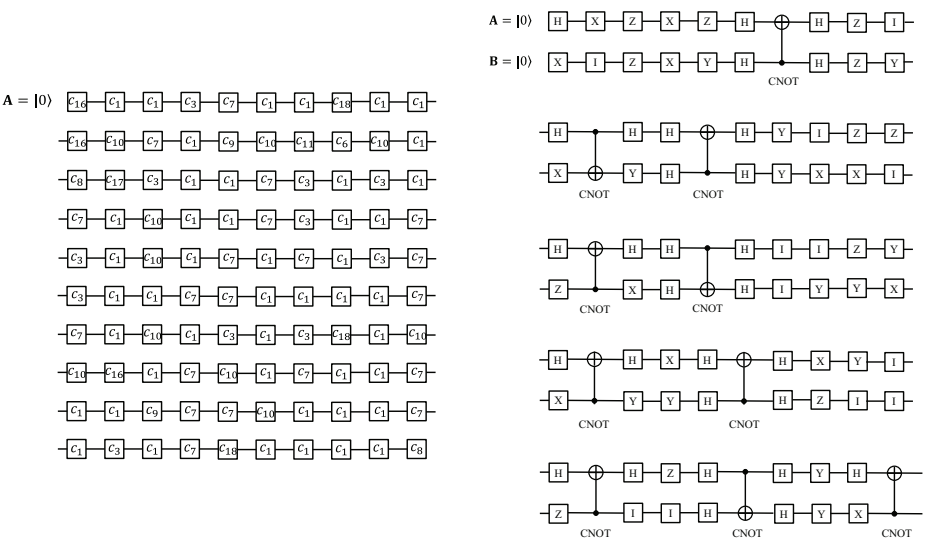


Figure E.2: (left) The entire improved one-qubit circuit with circuit length $\tau = 100$ obtained by Algorithm 2 ($C = 0.004$). (right) The entire improved two-qubit circuit with circuit length $\tau = 50$ obtained by Algorithm 6 ($C = 0.002$).

E.10. ERROR DISTRIBUTION

Recall that $Q_{y,v}(t) = \mathbb{P}[Y_{t+1} = v | Y_t = y]$. In the experiment of Figure 6.12, we assume a gate-dependent error model in which only single qubit errors occur that are (conditionally) i.i.d. on both qubits: that is

$$\mathbb{P}[\Lambda_{t+1} = \lambda | Y_t = y] = \begin{cases} \zeta_{U_{t+1}}(\lambda_1)\zeta_{U_{t+1}}(\lambda_2) & \text{if } \lambda = (\lambda_1, \lambda_2) \in \mathscr{C}_1^{\otimes 2} \\ 0 & \text{otherwise,} \end{cases} \tag{E.7}$$

for a set of distributions $\{\zeta_g | g \in \mathscr{C}_2\}$, say. It now follows from the law of total probability that

$$Q_{y,v}(t) = \sum_{(\lambda_1,\lambda_2)\in\mathscr{C}_1^{\otimes 2}} \mathbb{1}[(\lambda_1 \otimes \lambda_2) U_{t+1} y \rho y^\dagger U_{t+1}^\dagger (\lambda_1 \otimes \lambda_2)^\dagger = v \rho_0 v^\dagger] \zeta_{U_{t+1}}(\lambda_1)\zeta_{U_{t+1}}(\lambda_2). \tag{E.8}$$

Now, specifically, the error probabilities for e.g. the first qubit are set in the numerical experiment as shown in Table E.2. The error probabilities for the second qubit are set similarly so. Note that *not* applying a gate gives the lowest error rate; applying a single-qubit gate results in a medium error rate; and applying a two-qubit gate gives the largest probability that an error may occur.

Table E.2: The error probabilities for the first qubit as set in the numerical experiment that generates Figure 6.12. The error probabilities for the second qubit are set similarly so.

Case $g = I \times \mathscr{C}_1$:	Case $g \in (\mathscr{C}_1 \setminus I) \times \mathscr{C}_1$:	Case $g \in \mathscr{C}_2 \setminus \mathscr{C}_1^{\otimes 2}$:
$\zeta_g(\lambda_1) = \begin{cases} 0.990 & \text{if } \lambda_1 = I, \\ 0.006 & \text{if } \lambda_1 = X, \\ 0.003 & \text{if } \lambda_1 = Y, \\ 0.001 & \text{if } \lambda_1 = Z, \\ 0 & \text{otherwise.} \end{cases}$	$\zeta_g(\lambda_1) = \begin{cases} 0.950 & \text{if } \lambda_1 = I, \\ 0.030 & \text{if } \lambda_1 = X, \\ 0.015 & \text{if } \lambda_1 = Y, \\ 0.005 & \text{if } \lambda_1 = Z, \\ 0 & \text{otherwise.} \end{cases}$	$\zeta_g(\lambda_1) = \begin{cases} 0.900 & \text{if } \lambda_1 = I, \\ 0.060 & \text{if } \lambda_1 = X, \\ 0.030 & \text{if } \lambda_1 = Y, \\ 0.010 & \text{if } \lambda_1 = Z, \\ 0 & \text{otherwise.} \end{cases}$

E.11. PSEUDO-CODE FOR THE SIMULATED ANNEALING ALGORITHM THAT IMPROVES THE QUANTUM CIRCUIT THAT IMPLEMENTS DEUTSCH–JOZSA'S ALGORITHM

The algorithm that was used to generate the improved circuits for Figure 6.12 is shown in Algorithm 7.

Algorithm 7: Pseudo-code for the simulated annealing algorithm that improves the quantum circuit that implements Deutsch–Jozsa's Algorithm for one classical bit.

Input: A circuit $\{U_1,\ldots,U_\tau\}$ with $U_1,\ldots,U_\tau \in \mathscr{C}_2$ and number of iterations w
Output: A revised circuit $\{G_1^{[w]},\ldots,G_\tau^{[w]}\}$

begin
  Initialize $\{G_1^{[0]},\ldots,G_\tau^{[0]}\} = \{U_1,\ldots,U_\tau\}$;
  for $\eta \leftarrow 1$ **to** w **do**
    Choose $I \in [\tau - 1]$ uniformly at random;
    Choose $B \in \{-1,+1\}$ uniformly at random;
    Choose $G \in \mathscr{C}_1 \otimes \mathscr{C}_1$ uniformly at random;
    if $B = -1$ **then**
      Set $J_I = G, J_{I+1} = G_{I+1}^{[\eta]} G_I^{[\eta]} G^{\leftarrow}, J_i = G_i^{[\eta]}\ \forall_{i \neq I, I+1}$;
    else
      Set $J_{I+1} = G, J_I = G^{\leftarrow} G_{I+1}^{[\eta]} G_I^{[\eta]}, J_i = G_i^{[\eta]}\ \forall_{i \neq I, I+1}$;
    end
    Choose $X \in [0,1]$ uniformly at random;
    if $X \leq \exp\left(-(1/T_\eta)\max\{0, u(J) - u(G^{[\eta]})\}\right)$ **then**
      Set $G^{[\eta+1]} = J$;
    else
      Set $G^{[\eta+1]} = G^{[\eta]}$;
    end
  end
end

BIBLIOGRAPHY

[1] Linda JS Allen. "Some discrete-time SI, SIR, and SIS epidemic models". In: *Mathematical biosciences* 124.1 (1994), pp. 83–105.

[2] Romualdo Pastor-Satorras et al. "Epidemic processes in complex networks". In: *Reviews of modern physics* 87.3 (2015), p. 925.

[3] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. 10th. New York, NY, USA: Cambridge University Press, 2011. ISBN: 1107002176, 9781107002173.

[4] Piet Van Mieghem. *Performance analysis of complex networks and systems*. Cambridge: Cambridge University Press, 2014. ISBN: 978-1-107-05860-6.

[5] Herbert W Hethcote. "The mathematics of infectious diseases". In: *SIAM review* 42.4 (2000), pp. 599–653.

[6] Massimo A Achterberg et al. "Comparing the accuracy of several network-based COVID-19 prediction algorithms". In: *International journal of forecasting* (2020).

[7] Bastian Prasse et al. "Network-based prediction of the 2019-ncov epidemic outbreak in the chinese province hubei". In: *arXiv preprint arXiv:2002.04482* (2020).

[8] Bastian Prasse et al. "Network-inference-based prediction of the COVID-19 epidemic outbreak in the Chinese province Hubei". In: *Applied Network Science* 5.1 (2020), pp. 1–11.

[9] Ye Sun et al. "Spreading to localized targets in complex networks". In: *Scientific reports* 6.1 (2016), pp. 1–10.

[10] Long Ma, Maksim Kitsak, and Piet Van Mieghem. "Two-population SIR model and strategies to reduce mortality in pandemics". In: *arXiv preprint arXiv:2109.05964* (2021).

[11] Tian Gan and Long Ma. *Characterizing the Divergence Between Two Different Models for Fitting and Forecasting the COVID-19 Pandemic*. Tech. rep. EasyChair, 2021.

[12] Biao Tang et al. "The effectiveness of quarantine and isolation determine the trend of the COVID-19 epidemics in the final phase of the current outbreak in China". In: *International Journal of Infectious Diseases* 95 (2020), pp. 288–293.

[13] Rahul Subramanian, Qixin He, and Mercedes Pascual. “Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity”. In: *Proceedings of the National Academy of Sciences* 118.9 (2021).

[14] Axel Bonačić Marinović et al. “Quantifying reporting timeliness to improve outbreak control”. In: *Emerging infectious diseases* 21.2 (2015), p. 209.

[15] Corien Swaan et al. “Timeliness of notification systems for infectious diseases: A systematic literature review”. In: *PloS one* 13.6 (2018), e0198845.

[16] Centers for Disease Control, Prevention, et al. *Interpretation of Epidemic (Epi) Curves during Ongoing Outbreak Investigations*. 2013.

[17] Hamad Bastaki et al. “Time delays in the diagnosis and treatment of malaria in non-endemic countries: A systematic review”. In: *Travel medicine and infectious disease* 21 (2018), pp. 21–27.

[18] Anwar E. Ahmed. “Diagnostic delays in 537 symptomatic cases of Middle East respiratory syndrome coronavirus infection in Saudi Arabia”. In: *International Journal of Infectious Diseases* 62 (2017), pp. 47–51.

[19] Kaiyuan Sun, Jenny Chen, and Cécile Viboud. “Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study”. In: *The Lancet Digital Health* (2020).

[20] Natalie M. Linton et al. “Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data”. In: *Journal of clinical medicine* 9.2 (2020), p. 538.

[21] Stephen A. Lauer et al. “The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application”. In: *Annals of internal medicine* 172.9 (2020), pp. 577–582.

[22] Moritz U.G. Kraemer et al. “The effect of human mobility and control measures on the COVID-19 epidemic in China”. In: *Science* 368.6490 (2020), pp. 493–497.

[23] Kathy Leung et al. “First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment”. In: *The Lancet* (2020).

[24] Qianying Lin et al. “A conceptual model for the outbreak of Coronavirus disease 2019 (COVID-19) in Wuhan, China with individual reaction and governmental action”. In: *International journal of infectious diseases* (2020).

[25] Diletta Cereda et al. *The early phase of the COVID-19 outbreak in Lombardy, Italy*. 2020.

[26] Jonas Dehning et al. "Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions". In: *Science* (2020).

[27] Felix Guenther et al. "Nowcasting the COVID-19 pandemic in Bavaria". In: *medRxiv* (2020).

[28] Amna Tariq et al. "Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020". In: *BMC Medicine* 18 (2020), pp. 1–14.

[29] Jeffrey E. Harris. "Overcoming Reporting Delays Is Critical to Timely Epidemic Monitoring: The Case of COVID-19 in New York City". In: *medRxiv* (2020).

[30] Lorenzo Pellis et al. "Challenges in control of COVID-19: short doubling time and long delay to effect of interventions". In: *arXiv preprint arXiv:2004.00117* (2020).

[31] Piet Van Mieghem. *Performance analysis of complex networks and systems*. Cambridge University Press, 2014.

[32] Piet Van Mieghem and Eric Cator. "Epidemics in networks with nodal self-infection and the epidemic threshold". In: *Physical Review E* 86.1 (2012), p. 016116.

[33] Long Ma, Qiang Liu, and Piet Van Mieghem. "Inferring network properties based on the epidemic prevalence". In: *Applied Network Science* 4.1 (2019), pp. 1–13.

[34] Long Ma et al. "Efficient reconstruction of heterogeneous networks from time series via compressed sensing". In: *PloS one* 10.11 (2015), e0142837.

[35] Zhesi Shen et al. "Reconstructing propagation networks with natural diversity and identifying hidden sources". In: *Nature communications* 5 (2014).

[36] Srinivas Gorur Shandilya and Marc Timme. "Inferring network topology from complex dynamics". In: *New Journal of Physics* 13.1 (2011), p. 013004.

[37] Tyrus Berry et al. "Detecting connectivity changes in neuronal networks". In: *Journal of neuroscience methods* 209.2 (2012), pp. 388–397.

[38] Marc Timme and Jose Casadiego. "Revealing networks from dynamics: an introduction". In: *Journal of Physics A: Mathematical and Theoretical* 47.34 (2014), p. 343001.

[39] Mor Nitzan, Jose Casadiego, and Marc Timme. "Revealing physical interaction networks from statistics of collective dynamics". In: *Science advances* 3.2 (2017), e1600396.

[40] Bastian Prasse and Piet Van Mieghem. "Exact network reconstruction from complete SIS nodal state infection information seems infeasible". In: *IEEE Transactions on Network Science and Engineering* 6.4 (2018), pp. 748–759.

[41] Praneeth Netrapalli and Sujay Sanghavi. “Learning the Graph of Epidemic Cascades”. In: *SIGMETRICS Perform. Eval. Rev.* 40.1 (June 2012), pp. 211–222. ISSN: 0163-5999. DOI: 10.1145/2318857.2254783. URL: http://doi.acm.org/10.1145/2318857.2254783.

[42] Seth Myers and Jure Leskovec. “On the convexity of latent social network inference”. In: *Advances in neural information processing systems*. 2010, pp. 1741–1749.

[43] Emre Sefer and Carl Kingsford. “Convex risk minimization to infer networks from probabilistic diffusion data at multiple scales”. In: *Data engineering (ICDE), 2015 IEEE 31st international conference on*. IEEE. 2015, pp. 663–674.

[44] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. “Inferring networks of diffusion and influence”. In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2010, pp. 1019–1028.

[45] Karl J Friston. “Bayesian estimation of dynamical systems: an application to fMRI”. In: *NeuroImage* 16.2 (2002), pp. 513–530.

[46] Sinisa Pajevic and Dietmar Plenz. “Efficient network reconstruction from dynamical cascades identifies small-world topology of neuronal avalanches”. In: *PLoS computational biology* 5.1 (2009), e1000271.

[47] Chuang Ma et al. “Statistical inference approach to structural reconstruction of complex networks from binary time series”. In: *Physical Review E* 97.2 (2018), p. 022301.

[48] Xun Li and Xiang Li. “Reconstruction of stochastic temporal networks through diffusive arrival times”. In: *Nature communications* 8 (2017), p. 15729.

[49] Guofeng Mei et al. “Compressive-sensing-based structure identification for multilayer networks”. In: *IEEE transactions on cybernetics* 48.2 (2018), pp. 754–764.

[50] Emily SC Ching, Pik-Yin Lai, and CY Leung. “Reconstructing weighted networks from dynamics”. In: *Physical Review E* 91.3 (2015), p. 030801.

[51] Stefan Hempel et al. “Inner composition alignment for inferring directed networks from short time series”. In: *Physical review letters* 107.5 (2011), p. 054101.

[52] Jeffrey Shaman and Melvin Kohn. “Absolute humidity modulates influenza survival, transmission, and seasonality”. In: *Proceedings of the National Academy of Sciences* 106.9 (2009), pp. 3243–3248.

[53] Jeffrey Shaman et al. “Absolute humidity and the seasonal onset of influenza in the continental United States”. In: *PLoS biology* 8.2 (2010), e1000316.

[54] Long Ma and Jaron Sanders. “Markov chains and hitting times for error accumulation in quantum circuits”. In: *EAI International Conference on Performance Evaluation Methodologies and Tools*. Springer. 2021, pp. 36–55.

[55] Long Ma and Jaron Sanders. “Markov chains for error accumulation in quantum circuits”. In: *arXiv preprint arXiv:1909.04432* (2019).

[56] John Preskill. “Quantum computing: pro and con”. In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 454.1969 (1998), pp. 469–486.

[57] Daniel Gottesman. “Efficient fault tolerance”. In: *Nature* 540 (2016), p. 44.

[58] Julia Cramer et al. “Repeated quantum error correction on a continuously encoded qubit by real-time feedback”. In: *Nature communications* 7 (2016), p. 11526.

[59] Norbert M. Linke et al. “Fault-tolerant quantum error detection”. In: *Science advances* 3.10 (2017), e1701074.

[60] Austin G. Fowler and Lloyd C.L. Hollenberg. “Scalability of Shor’s algorithm with a limited set of rotation gates”. In: *Physical Review A* 70.3 (2004), p. 032329.

[61] Roy M Anderson, Robert M May, and B Anderson. *Infectious diseases of humans: dynamics and control*. Vol. 28. Wiley Online Library, 1992.

[62] T. E. Harris. “Contact Interactions on a Lattice”. In: *The Annals of Probability* 2.6 (1974), pp. 969–988. ISSN: 00911798. URL: http://www.jstor.org/stable/2959099.

[63] Romualdo Pastor-Satorras and Alessandro Vespignani. “Epidemic spreading in scale-free networks”. In: *Physical review letters* 86.14 (2001), p. 3200.

[64] Shirshendu Chatterjee and Rick Durrett. “CONTACT PROCESSES ON RANDOM GRAPHS WITH POWER LAW DEGREE DISTRIBUTIONS HAVE CRITICAL VALUE 0”. In: *The Annals of Probability* 37.6 (2009), pp. 2332–2356. ISSN: 00911798. URL: http://www.jstor.org/stable/27795079.

[65] Alexander V Goltsev et al. “Localization and spreading of diseases in complex networks”. In: *Physical review letters* 109.12 (2012), p. 128702.

[66] Q. Liu and P. Van Mieghem. “Network localization is unalterable by infections in bursts”. In: *IEEE Transactions on Network Science and Engineering* (2019), pp. 1–1. ISSN: 2327-4697. DOI: 10.1109/TNSE.2018.2889539.

[67] Qiang Liu and Piet Van Mieghem. “Autocorrelation of the susceptible-infected-susceptible process on networks”. In: *Physical Review E* 97.6 (2018), p. 062309.

E

[68] P. E. Paré et al. “Analysis, Estimation, and Validation of Discrete-Time Epidemic Processes”. In: *IEEE Transactions on Control Systems Technology* (2018), pp. 1–15. ISSN: 1063-6536. DOI: 10.1109/TCST.2018.2869369.

[69] Gonzalo Mateos et al. “Connecting the dots: Identifying network structure via graph signal processing”. In: *IEEE Signal Processing Magazine* 36.3 (2019), pp. 16–43.

[70] Xiaowen Dong et al. “Learning Graphs From Data: A Signal Representation Perspective”. In: *IEEE Signal Processing Magazine* 36.3 (2019), pp. 44–63.

[71] Xiao Han et al. “Robust reconstruction of complex networks from sparse data”. In: *Physical review letters* 114.2 (2015), p. 028701.

[72] Jingwen Li et al. “Universal data-based method for reconstructing complex networks with binary-state dynamics”. In: *Physical Review E* 95.3 (2017), p. 032303.

[73] K-I Goh, Byungnam Kahng, and Doochul Kim. “Universal behavior of load distribution in scale-free networks”. In: *Physical Review Letters* 87.27 (2001), p. 278701.

[74] Albert-László Barabási and Réka Albert. “Emergence of scaling in random networks”. In: *science* 286.5439 (1999), pp. 509–512.

[75] P. Erdős and A. Rényi. “On random graphs I”. In: *Publ. Math. Debrecen* 6 (1959), pp. 290–297.

[76] Duncan J. Watts and Steven H. Strogatz. “Collective dynamics of ‘small-world’networks”. In: *nature* 393.6684 (1998), p. 440.

[77] F Di Lauro et al. “Network Inference from Population-Level Observation of Epidemics”. In: *arXiv preprint arXiv:1906.10966* (2019).

[78] Qiang Liu and Piet Van Mieghem. “Evaluation of an analytic, approximate formula for the time-varying SIS prevalence in different networks”. In: *Physica A: Statistical Mechanics and its Applications* 471 (2017), pp. 325–336.

[79] Daniel T Gillespie. “Exact stochastic simulation of coupled chemical reactions”. In: *The journal of physical chemistry* 81.25 (1977), pp. 2340–2361.

[80] Qiang Liu and Piet Van Mieghem. “Evaluation of an analytic, approximate formula for the time-varying SIS prevalence in different networks”. In: *Physica A: Statistical Mechanics and its Applications* 471 (2017), pp. 325–336.

[81] Guillaume St-Onge et al. “Efficient sampling of spreading processes on complex networks using a composition and rejection algorithm”. In: *Computer Physics Communications* (2019).

[82] Mark E.J. Newman. “Assortative mixing in networks”. In: *Physical review letters* 89.20 (2002), p. 208701.

[83] P. Van Mieghem, J. Omic, and R. Kooij. "Virus Spread in Networks". In: *IEEE/ACM Transactions on Networking* 17.1 (Feb. 2009), pp. 1–14. ISSN: 1063-6692. DOI: 10.1109/TNET.2008.925623.

[84] Piet Van Mieghem. *Graph spectra for complex networks*. Cambridge University Press, 2010.

[85] Michele Catanzaro and Romualdo Pastor-Satorras. "Analytic solution of a static scale-free network model". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 44.2 (2005), pp. 241–248.

[86] Vincent J. Munster et al. "A novel coronavirus emerging in China—key questions for impact assessment". In: *New England Journal of Medicine* 382.8 (2020), pp. 692–694.

[87] Alexander F. Siegenfeld, Nassim N. Taleb, and Yaneer Bar-Yam. "Opinion: What models can and cannot tell us about COVID-19". In: *Proceedings of the National Academy of Sciences* (2020).

[88] Jasper Fuk-Woo Chan et al. "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster". In: *The Lancet* 395.10223 (2020), pp. 514–523.

[89] Jonathan M. Read et al. "Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions". In: *MedRxiv* (2020).

[90] Stefan Thurner, Peter Klimek, and Rudolf Hanel. "A network-based explanation of why most COVID-19 infection curves are linear". In: *Proceedings of the National Academy of Sciences* (2020).

[91] Qun Li et al. "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia". In: *New England Journal of Medicine* (2020).

[92] Ryad Ghanam, Edward L. Boone, and Abdel-Salam G. Abdel-Salam. "SEIRD Model for Qatar COVID-19 Outbreak: A Case Study". In: *arXiv preprint arXiv:2005.12777* (2020).

[93] Benjamin F. Maier and Dirk Brockmann. "Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China". In: *Science* 368.6492 (2020), pp. 742–746.

[94] Marino Gatto et al. "Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures". In: *Proceedings of the National Academy of Sciences* 117.19 (2020), pp. 10484–10491.

[95] Ernesto Estrada. "COVID-19 and SARS-CoV-2. Modeling the Present, Looking at the Future". In: *Physics Reports* (2020).

[96] Jesús Fernández-Villaverde and Charles I. Jones. *Estimating and Simulating a SIRD Model ofCOVID-19 for Many Countries, States, and Cities*. Tech. rep. National Bureau of Economic Research, 2020.

[97] Andrea L. Bertozzi et al. "The challenges of modeling and forecasting the spread of COVID-19". In: *Proceedings of the National Academy of Sciences* 117.29 (2020), pp. 16732–16738. ISSN: 0027-8424. DOI: 10.1073/pnas.2006520117. eprint: https://www.pnas.org/content/117/29/16732.full.pdf. URL: https://www.pnas.org/content/117/29/16732.

[98] Joseph T. Wu, Kathy Leung, and Gabriel M. Leung. "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study". In: *The Lancet* 395.10225 (2020), pp. 689–697.

[99] Wan Yang, Alicia Karspeck, and Jeffrey Shaman. "Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics". In: *PLoS computational biology* 10.4 (2014).

[100] Teresa K. Yamana, Sasikiran Kandula, and Jeffrey Shaman. "Individual versus superensemble forecasts of seasonal influenza outbreaks in the United States". In: *PLoS computational biology* 13.11 (2017), e1005801.

[101] Evan L. Ray and Nicholas G. Reich. "Prediction of infectious disease epidemics via weighted density ensembles". In: *PLoS computational biology* 14.2 (2018), e1005910.

[102] Vittoria Colizza et al. "The role of the airline transportation network in the prediction and predictability of global epidemics". In: *Proceedings of the National Academy of Sciences* 103.7 (2006), pp. 2015–2020.

[103] Matheus Henrique Dal Molin Ribeiro et al. "Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil". In: *Chaos, Solitons & Fractals* (2020), p. 109853.

[104] Daniel B. Larremore et al. "Test sensitivity is secondary to frequency and turnaround time for COVID-19 surveillance". In: *MedRxiv* (2020).

[105] Timo Mitze et al. "Face masks considerably reduce COVID-19 cases in Germany". In: *Proceedings of the National Academy of Sciences* 117.51 (2020), pp. 32293–32301. ISSN: 0027-8424. DOI: 10.1073/pnas.2015954117. eprint: https://www.pnas.org/content/117/51/32293.full.pdf. URL: https://www.pnas.org/content/117/51/32293.

[106] Camilla Rothe et al. “Transmission of 2019-nCoV infection from an asymptomatic contact in Germany”. In: *New England Journal of Medicine* 382.10 (2020), pp. 970–971.

[107] Francois-Xavier Lescure et al. “Clinical and virological data of the first cases of COVID-19 in Europe: a case series”. In: *The Lancet Infectious Diseases* (2020).

[108] Jun Chen et al. “Clinical progression of patients with COVID-19 in Shanghai, China”. In: *Journal of infection* 80.5 (2020), e1–e6.

[109] Irena Voinsky, Gabriele Baristaite, and David Gurwitz. “Effects of age and sex on recovery from COVID-19: Analysis of 5769 Israeli patients”. In: *Journal of Infection* 81.2 (2020), e102–e103.

[110] Christel Faes et al. “Time between symptom onset, hospitalisation and recovery or death: statistical analysis of Belgian COVID-19 patients”. In: *International journal of environmental research and public health* 17.20 (2020), p. 7560.

[111] Ewen M. Harrison, Annemarie Docherty, and Calum Semple. “COVID-19: time from symptom onset until death in UK hospitalised patients”. In: (2020). URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/928729/S0803_CO-CIN_-_Time_from_symptom_onset_until_death.pdf.

[112] G.H. Freeman. “Fitting two-parameter discrete distributions to many data sets with one common parameter”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 29.3 (1980), pp. 259–267.

[113] Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz. *Univariate discrete distributions*. Vol. 444. John Wiley & Sons, 2005.

[114] J. Keith Ord. *Families of frequency distributions*. Griffin, 1972.

[115] James Bergstra and Yoshua Bengio. “Random search for hyper-parameter optimization”. In: *Journal of machine learning research* 13.2 (2012).

[116] S. Seaman and D.D. Angelis. *Adjusting COVID-19 deaths to account for reporting delay*. Tech. rep. Technical report, MRC-Biostatistics Unit, 2020.

[117] P.M. Lerman. “Fitting segmented regression models by grid search”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 29.1 (1980), pp. 77–84.

[118] Rachel E. Jordan, Peymane Adab, and K.K. Cheng. *COVID-19: risk factors for severe disease and death*. 2020.

[119] Cory Merow and Mark C. Urban. “Seasonality and uncertainty in global COVID-19 growth rates”. In: *Proceedings of the National Academy of Sciences* (2020).

[120] Nicholas Reimann. *More Young People Are Dying Of Coronavirus In Florida, As State Shatters Death Record Tuesday*. 2020. URL: https://www.forbes.com/sites/nicholasreimann/2020/08/11/more-young-people-are-dying-of-coronavirus-in-florida-as-state-shatters-death-record-tuesday/?sh=7fc97d5d152e (visited on 08/11/2020).

[121] Richard Armitage and Laura B. Nellums. "COVID-19 and the consequences of isolating the elderly". In: *The Lancet Public Health* 5.5 (2020), e256.

[122] Kathy Leung et al. "Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020". In: *Eurosurveillance* 26.1 (2021), p. 2002106.

[123] Dylan H. Morris et al. "Optimal, near-optimal, and robust epidemic control". In: *Communications Physics* 4.1 (2021), pp. 1–8.

[124] S.T. Fahira et al. "The effect of social inequality on the growth of COVID-19 death case". In: *Journal of Physics: Conference Series*. Vol. 1722. 1. IOP Publishing. 2021, p. 012041.

[125] Ian Cooper, Argha Mondal, and Chris G. Antonopoulos. "A SIR model assumption for the spread of COVID-19 in different communities". In: *Chaos, Solitons & Fractals* 139 (2020), p. 110057.

[126] Davide Faranda and Tommaso Alberti. "Modeling the second wave of COVID-19 infections in France and Italy via a stochastic SEIR model". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30.11 (2020), p. 111101.

[127] Zeynep Ceylan. "Estimation of COVID-19 prevalence in Italy, Spain, and France". In: *Science of The Total Environment* 729 (2020), p. 138817.

[128] Armando G.M. Neves and Gustavo Guerrero. "Predicting the evolution of the COVID-19 epidemic with the A-SIR model: Lombardy, Italy and Sao Paulo state, Brazil". In: *Physica D: Nonlinear Phenomena* 413 (2020), p. 132693.

[129] Zifeng Yang et al. "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions". In: *Journal of Thoracic Disease* 12.3 (2020), p. 165.

[130] Xiang Zhou et al. "Forecasting the worldwide spread of COVID-19 based on logistic model and SEIR model". In: *medRxiv* (2020).

[131] Alex Arenas et al. "A mathematical model for the spatiotemporal epidemic spreading of COVID-19". In: *MedRxiv* (2020).

[132] Anca Rădulescu, Cassandra Williams, and Kieran Cavanagh. "Management strategies in a SEIR-type model of COVID-19 community spread". In: *Scientific reports* 10.1 (2020), pp. 1–16.

[133] Laura Di Domenico et al. "Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies". In: *BMC medicine* 18.1 (2020), pp. 1–13.

[134] Kiesha Prem et al. "The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study". In: *The Lancet Public Health* 5.5 (2020), e261–e270.

[135] Ze-Yu Zhao et al. "A five-compartment model of age-specific transmissibility of SARS-CoV-2". In: *Infectious diseases of poverty* 9.1 (2020), pp. 1–15.

[136] John P.A. Ioannidis, Cathrine Axfors, and Despina G. Contopoulos-Ioannidis. "Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters". In: *Environmental research* 188 (2020), p. 109890.

[137] Data Worldbank. *Population ages 65 and above (% of total population)*. 2019.

[138] Pierre Magal, Ousmane Seydi, and Glenn Webb. "Final size of an epidemic for a two-group SIR model". In: *SIAM Journal on Applied Mathematics* 76.5 (2016), pp. 2042–2059.

[139] Sen Pei et al. "Forecasting the spatial transmission of influenza in the United States". In: *Proceedings of the National Academy of Sciences* (2018), p. 201708856.

[140] André Hajek and Hans-Helmut König. "Social isolation and loneliness of older adults in times of the COVID-19 pandemic: can use of online social media sites and video chats assist in mitigating social isolation and loneliness?" In: *Gerontology* 67.1 (2021), pp. 121–124.

[141] Stephen Eubank et al. "Structural and algorithmic aspects of massive social networks". In: *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*. Citeseer. 2004, pp. 718–727.

[142] Paul Erdős and Alfréd Rényi. "On the evolution of random graphs". In: *Publ. Math. Inst. Hung. Acad. Sci* 5.1 (1960), pp. 17–60.

[143] Mina Youssef and Caterina Scoglio. "An individual-based approach to SIR epidemics in contact networks". In: *Journal of theoretical biology* 283.1 (2011), pp. 136–144.

[144] Faryad Darabi Sahneh, Caterina Scoglio, and Piet Van Mieghem. "Generalized epidemic mean-field model for spreading processes over multilayer complex networks". In: *IEEE/ACM Transactions on Networking* 21.5 (2013), pp. 1609–1620.

[145] Centers for Disease Control and Prevention. *Clinical questions about COVID-19: questions and answers*. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html (visited on 05/08/2020).

[146] Georg Simmel. *Über sociale Differenzierung: sociologische und psychologische Untersuchungen*. Vol. 10. Duncker & Humblot, 1890.

[147] David Juher, Jordi Ripoll, and Joan Saldaña. "Analysis and Monte Carlo simulations of a model for the spread of infectious diseases in heterogeneous metapopulations". In: *Physical Review E* 80.4 (2009), p. 041920.

[148] Mark E.J. Newman. "The structure and function of complex networks". In: *SIAM review* 45.2 (2003), pp. 167–256.

[149] Piet Van Mieghem, Jasmina Omic, and Robert Kooij. "Virus spread in networks". In: *IEEE/ACM Transactions On Networking* 17.1 (2008), pp. 1–14.

[150] Péter L. Simon, Michael Taylor, and Istvan Z. Kiss. "Exact epidemic models on graphs using graph-automorphism driven lumping". In: *Journal of mathematical biology* 62.4 (2011), pp. 479–508.

[151] Athina Economou, Antonio Gómez-Corral, and M. López-García. "A stochastic SIS epidemic model with heterogeneous contacts". In: *Physica A: Statistical Mechanics and its Applications* 421 (2015), pp. 78–97.

[152] Wayne Barrett, Amanda Francis, and Benjamin Webb. "Equitable decompositions of graphs with symmetries". In: *Linear Algebra and its Applications* 513 (2017), pp. 409–434.

[153] Amanda Francis et al. "Extensions and applications of equitable decompositions for graphs with symmetries". In: *Linear Algebra and its Applications* 532 (2017), pp. 432–462.

[154] István Z. Kiss, Joel C. Miller, Péter L. Simon, et al. "Mathematics of epidemics on networks". In: *Cham: Springer* 598 (2017).

[155] Jonathan A. Ward and Martín López-García. "Exact analysis of summary statistics for continuous-time discrete-state Markov processes on networks using graph-automorphism lumping". In: *Applied Network Science* 4.1 (2019), p. 108.

[156] E. Cator and P. Van Mieghem. "Susceptible-infected-susceptible epidemics on the complete graph and the star graph: Exact analysis". In: *Phys. Rev. E* 87 (1 Jan. 2013), p. 012811. DOI: 10.1103/PhysRevE.87.012811. URL: https://link.aps.org/doi/10.1103/PhysRevE.87.012811.

[157] P. Van Mieghem. "The N-intertwined SIS epidemic network model". In: *Computing* 93 (2011), pp. 147–169. DOI: 10.1007/s00607-011-0155-y. URL: https://doi.org/10.1007/s00607-011-0155-y.

[158] Eric Cator and Piet Van Mieghem. "Second-order mean-field susceptible-infected-susceptible epidemic threshold". In: *Physical review E* 85.5 (2012), p. 056111.

[159] Romualdo Pastor-Satorras and Alessandro Vespignani. "Epidemic dynamics and endemic states in complex networks". In: *Phys. Rev. E* 63 (6 May 2001), p. 066117. DOI: 10.1103/PhysRevE.63.066117. URL: https://link.aps.org/doi/10.1103/PhysRevE.63.066117.

[160] James P. Gleeson. "High-accuracy approximation of binary-state dynamics on networks". In: *Physical Review Letters* 107.6 (2011), p. 068701.

[161] Matthew J. Keeling. "The effects of local spatial structure on epidemiological invasions". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266.1421 (1999), pp. 859–867.

[162] Jennifer Lindquist et al. "Effective degree network disease models". In: *Journal of mathematical biology* 62.2 (2011), pp. 143–164.

[163] Matt J. Keeling et al. "Systematic approximations to susceptible-infectious-susceptible dynamics on networks". In: *PLoS computational biology* 12.12 (2016), e1005296.

[164] Cong Li, Ruud van de Bovenkamp, and Piet Van Mieghem. "Susceptible-infected-susceptible model: A comparison of N-intertwined and heterogeneous mean-field approximations". In: *Physical Review E* 86.2 (2012), p. 026116.

[165] Bastian Prasse and Piet Van Mieghem. "Time-dependent solution of the NIMFA equations around the epidemic threshold". In: *Journal of Mathematical Biology* 81.6 (2020), pp. 1299–1355.

[166] Karel Devriendt and Piet Van Mieghem. "Unified mean-field framework for susceptible-infected-susceptible epidemics on networks, based on graph partitioning and the isoperimetric inequality". In: *Physical Review E* 96.5 (2017), p. 052314.

[167] Piet Van Mieghem. "Decay towards the overall-healthy state in SIS epidemics on networks". In: *arXiv preprint arXiv:1310.3980* (2013).

[168] Sifat Afroj Moon. "Group-based general epidemic modeling for spreading processes on networks: GroupGEM". In: *IEEE Transactions on Network Science and Engineering* (2020).

[169] Fan Chung. "Laplacians and the Cheeger inequality for directed graphs". In: *Annals of Combinatorics* 9.1 (2005), pp. 1–19.

[170] Fragkiskos D. Malliaros and Michalis Vazirgiannis. “Clustering and community detection in directed networks: A survey”. In: *Physics Reports* 533.4 (2013), pp. 95–142.

[171] Yanhua Li and Zhi-Li Zhang. “Digraph laplacian and the degree of asymmetry”. In: *Internet Mathematics* 8.4 (2012), pp. 381–401.

[172] Kiri Wagstaff et al. “Constrained k-means clustering with background knowledge”. In: *Icml*. Vol. 1. 2001, pp. 577–584.

[173] M. A. Achterberg and P. Van Mieghem. “Spectrum of the metastable state in continuous-time Markovian epsilon-SIS epidemics on networks”. In: (2021). unpublished.

[174] Daniel Greenbaum and Zachary Dutton. “Modeling coherent errors in quantum error correction”. In: *Quantum Science and Technology* 3.1 (2017), p. 015007.

[175] Easwar Magesan et al. “Modeling quantum noise for efficient testing of fault-tolerant circuits”. In: *Physical Review A* 87.1 (2013), p. 012324.

[176] Mauricio Gutiérrez et al. “Approximation of realistic errors by Clifford channels and Pauli measurements”. In: *Physical Review A* 87.3 (2013), p. 030302.

[177] Sergey Bravyi et al. “Correcting coherent errors with surface codes”. In: *npj Quantum Information* 4.1 (2018), p. 55.

[178] Eric Huang, Andrew C. Doherty, and Steven Flammia. “Performance of quantum error correction with coherent errors”. In: *Physical Review A* 99.2 (2019), p. 022313.

[179] Pierre Brémaud. *Discrete probability models and methods*. Vol. 78. Springer, 2017.

[180] T. Xia et al. “Randomized benchmarking of single-qubit gates in a 2D array of neutral-atom qubits”. In: *Physical Review Letters* 114.10 (2015), p. 100503.

[181] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. “Scalable and robust randomized benchmarking of quantum processes”. In: *Physical review letters* 106.18 (2011), p. 180504.

[182] Harrison Ball et al. “Effect of noise correlations on randomized benchmarking”. In: *Physical Review A* 93.2 (2016), p. 022303.

[183] Smitha Janardan et al. “Analytical error analysis of Clifford gates by the fault-path tracer method”. In: *Quantum Information Processing* 15.8 (2016), pp. 3065–3079.

[184] Joel J Wallman. “Randomized benchmarking with gate-dependent noise”. In: *Quantum* 2 (2018), p. 47.

[185] Jeffrey M. Epstein et al. “Investigating the limits of randomized benchmarking protocols”. In: *Physical Review A* 89.6 (2014), p. 062321.

[186] Daniel Stilck França and A.K. Hashagen. “Approximate randomized benchmarking for finite groups”. In: *Journal of Physics A: Mathematical and Theoretical* 51.39 (2018), p. 395302.

[187] Winton G. Brown and Bryan Eastin. “Randomized benchmarking with restricted gate sets”. In: *Physical Review A* 97.6 (2018), p. 062323.

[188] Bryan H. Fong and Seth T. Merkel. “Randomized benchmarking, correlated noise, and ising models”. In: *arXiv preprint arXiv:1703.09747* (2017).

[189] Christopher J. Wood and Jay M. Gambetta. “Quantification and characterization of leakage errors”. In: *Physical Review A* 97.3 (2018), p. 032306.

[190] Joel J. Wallman, Marie Barnhill, and Joseph Emerson. “Robust characterization of loss rates”. In: *Physical Review Letters* 115.6 (2015), p. 060501.

[191] Timothy Proctor et al. “What randomized benchmarking actually measures”. In: *Physical review letters* 119.13 (2017), p. 130502.

[192] Arnaud Carignan-Dugas et al. “From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters”. In: *New Journal of Physics* 20.9 (2018), p. 092001.

[193] Paul Harry Roberts and Harold Douglas Ursell. “Random walk on a sphere and on a Riemannian manifold”. In: *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 252.1012 (1960), pp. 317–356.

[194] Henryk Gutmann. “Description and control of decoherence in quantum bit systems”. PhD thesis. lmu, 2005.

[195] Emanuel Knill. “Quantum computing with realistically noisy devices”. In: *Nature* 434.7029 (2005), p. 39.

[196] Robin Harper et al. “Statistical analysis of randomized benchmarking”. In: *Physical Review A* 99.5 (2019), p. 052350.

[197] Dmitri Maslov et al. “Quantum circuit simplification and level compaction”. In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 27.3 (2008), pp. 436–444.

[198] Vadym Kliuchnikov and Dmitri Maslov. “Optimization of Clifford circuits”. In: *Physical Review A* 88.5 (2013), p. 052307.

[199] Matthew Amy. “Formal Methods in Quantum Circuit Design”. In: (2019).

[200] David Deutsch and Richard Jozsa. “Rapid solution of problems by quantum computation”. In: *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* 439.1907 (1992), pp. 553–558.

[201] Richard Cleve et al. “Quantum algorithms revisited”. In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 454.1969 (1998), pp. 339–354.

[202] Nikolaj Moll et al. “Quantum optimization using variational algorithms on near-term quantum devices”. In: *Quantum Science and Technology* 3.3 (2018), p. 030503.

[203] Keisuke Fujii. “Stabilizer Formalism and Its Applications”. In: *Quantum Computation with Topological Codes*. Springer, 2015, pp. 24–55.

[204] Scott Aaronson and Daniel Gottesman. “Improved simulation of stabilizer circuits”. In: *Physical Review A* 70.5 (2004), p. 052328.

[205] Daniel M. Greenberger, Michael A. Horne, and Anton Zeilinger. “Going beyond Bell’s theorem”. In: *Bell’s theorem, quantum theory and conceptions of the universe*. Springer, 1989, pp. 69–72.

[206] Charles H. Bennett and Stephen J. Wiesner. “Communication via one- and two-particle operators on Einstein–Podolsky–Rosen states”. In: *Physical Review Letters* 69.20 (1992), p. 2881.

[207] Charles H. Bennett et al. “Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels”. In: *Physical Review Letters* 70.13 (1993), p. 1895.

[208] Peter Selinger. “Generators and relations for n-qubit Clifford operators”. In: *Logical Methods in Computer Science* 11 (2013).

[209] Daniel Gottesman. “The Heisenberg representation of quantum computers”. In: *arXiv preprint quant-ph/9807006* (1998).

[210] Mary Beth Ruskai. “Pauli exchange errors in quantum computation”. In: *Physical Review Letters* 85.1 (2000), p. 194.

[211] Robert Koenig and John A. Smolin. “How to efficiently select an arbitrary Clifford group element”. In: *Journal of Mathematical Physics* 55.12 (2014), p. 122202.

[212] Rajendra Bhatia. *Matrix analysis*. Vol. 169. Springer Science & Business Media, 2013.

[213] Daniel Gottesman. “Stabilizer codes and quantum error correction”. In: *arXiv preprint quant-ph/9705052* (1997).

[214] Héctor J. García, Igor L. Markov, and Andrew W. Cross. “On the geometry of stabilizer states”. In: *arXiv preprint arXiv:1711.07848* (2017).

[215] C.C. Grosjean. “Theory of recursive generation of systems of orthogonal polynomials: An illustrative example”. In: *Journal of Computational and Applied Mathematics* 12 (1985), pp. 299–318.

[216] Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Vol. 55. Courier Corporation, 1965.

[217] George B. Arfken and Hans J. Weber. *Mathematical methods for physicists*. AAPT, 1999.

[218] Qihui Yang et al. "Short-term forecasts and long-term mitigation evaluations for the COVID-19 epidemic in Hubei Province, China". In: *medRxiv* (2020).

[219] Massimo A. Achterberg et al. "Comparing the accuracy of several network-based COVID-19 prediction algorithms". In: *International Journal of Forecasting* (2020).

[220] Jiang Zhang et al. "Investigating time, strength, and duration of measures in controlling the spread of COVID-19 using a networked meta-population model". In: *Nonlinear Dynamics* 101.3 (2020), pp. 1789–1800.

[221] Donghui Yan, Ling Huang, and Michael I. Jordan. "Fast approximate spectral clustering". In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2009, pp. 907–916.

[222] Gerrit Großmann and Luca Bortolussi. "Reducing spreading processes on networks to Markov population models". In: *International Conference on Quantitative Evaluation of Systems*. Springer. 2019, pp. 292–309.

[223] Piet Van Mieghem. "Universality of the SIS prevalence in networks". In: *arXiv preprint arXiv:1612.01386* (2016).

[224] P Van Mieghem and K Devriendt. "An epidemic perspective on the cut size in networks". In: *Delft University of Technology, Report* 20180312 (2018).

[225] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. cambridge University press, 2009.

[226] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994.

[227] Yi-Cheng Chen et al. "A Time-dependent SIR model for COVID-19 with undetectable infected persons". In: *IEEE Transactions on Network Science and Engineering* (2020).

[228] Robert Schaback. "On COVID-19 modelling". In: *Jahresbericht der Deutschen Mathematiker-Vereinigung* 122.3 (2020), pp. 167–205.

[229] Steven Roman, S. Axler, and F.W. Gehring. *Advanced linear algebra*. Vol. 3. Springer, 2005, pp. 153–165.

ACKNOWLEDGEMENTS

First of all, I'd like to convey my gratitude to Prof. Piet Van Mieghem, my promoter. Thank you for providing me with the opportunity to pursue my doctorate in complex networks at TU Delft. I appreciate your suggestions for research topics, efforts and patience in training my soft skills, encouragement and support. Also, thank you for all the pleasant conversations and good times at the Friday drinks, lunches, BBQs and other events.

Second, I want to show my gratitude to my daily supervisor Dr. Maksim Kitsak. Maksim, you're the one who always encourages me to pursue my ideas, does brainstorms with me, gives me valuable comments and helps me to revise my works. I really enjoy working with you and thanks for your guidance and encouragement.

Next, I'd like to express my gratitude to Dr. Jaron Sanders, my other supervisor. I am grateful for the time you spent discussing ideas, assisting me in preparing talks and revising the paper and slides. You are always so thoughtful and kind. Thank you so much for your patience and support during this journey, Jaron.

Special thanks to all members and alumni members of the NAS group I met during my PhD studies: Albert Senen-Cerda, Dr. Bastian Prasse, Dr. Edgar van Boven, Dr. Eric Smeitink, Fenghua Wang, Gabriel Budel, Dr. Hale Çetinay İyicil, Dr. Hannah Bos, Ivan Jokić, Karel Devriendt, Maria Raftopoulou, Dr. Mattia Sensi, Massimo Achterberg, Dr. Marcus Märtens, Misa Taguchi, Peng Sun, Dr. Qiang Liu, Qingfeng Tong, Dr. Remco Litjens, Prof. Rob Kooij, Rogier Noldus, Dr. Zhidong He, who have made my PhD experience both enjoyable and memorable.

Last but not least, my gratitude goes to my beloved wife Ye, my mother Lian, my father Hongzeng and my brother Tao. Thanks for your unconditional love, encouragement and support. Life and research was difficult during the terrible COVID-19 pandemic. It is you who made my PhD research journey happy and enjoyable.

CURRICULUM VITÆ

Long MA

Long Ma obtained B.E. in Physics from Nankai University, Tianjin, China in 2014. During the undergraduate period, he joined in "The Plan of Top-Notch Students Training in Basic Disciplines". In 2012, he won the First Prize in China Undergraduate Physics Tournament (CUPT). From 2012 to 2013, he led an undergraduate students' academic fund project of Nankai University and finished it successfully. In 2013, he won the Outstanding Award in the Tenth "100 Projects" of Creative Research for the Undergraduates of Nankai University. After graduation, he joined the center for complexity science of Beijing Normal University as a master student. He has been involved in many projects on complex networks, especially spreading dynamics, evolutionary game dynamics, behavior experiments on networks and network reconstructions. In 2016, he won the National Scholarship. In 2017, he started to pursue his Ph.D. at the Network Architectures and Services Group under the supervision of Prof.dr.ir. Piet Van Mieghem, Dr. Maksim Kitsak and Dr. Jaron Sanders.

List of Publications

9. L. Ma, Q. Liu, P. Van Mieghem, *Inferring network properties based on the epidemic prevalence*, Applied Network Science **4(1)**, 93 (2019).

8. B. Prasse, M. A. Achterberg, L. Ma and P. Van Mieghem, *Network-inference-based prediction of the COVID-19 epidemic outbreak in the Chinese province Hubei*, Applied Network Science **5(1)**, 1-11 (2020).

7. M. A. Achterberg, B. Prasse, L. Ma, S. Trajanovski, M. Kitsak, P. Van Mieghem, *Comparing the Accuracy of Several Network-based COVID-19 Prediction Algorithms*, International Journal of Forecasting (2020)

6. T. Gan, L. Ma, *Characterizing the Divergence Between Two Different Models for Fitting and Forecasting the COVID-19 Pandemic*, French Regional Conference on Complex Systems (2021)

5. L. Ma, J. Sanders, *Markov chains and hitting times for error accumulation in quantum circuits*, 14th EAI International Conference on Performance Evaluation Methodologies and Tools (2021).

4. L. Ma, J. Sanders, *Markov chains and hitting times for error accumulation in quantum circuits (extended version)*, arXiv preprint arXiv:1909.04432 (2021).

3. L. Ma, M. Kitsak, P. Van Mieghem, *Two-population SIR model and strategies to reduce mortality in pandemics*, submitted (2021).

2. L. Ma, M. A. Achterberg, P. Van Mieghem, *Approximation of SIS epidemics on networks*, in preparation.

1. L. Ma, M. Kitsak, P. Van Mieghem, *Accounting for COVID-19 reporting delays to enhance the accuracy of epidemic forecasts*, in preparation.