Department of Computer Science

Spreading and Epidemic Interventions

Effects of Network Structure and Dynamics

Abbas K. Rizi

Aalto University

DOCTORAL THESES

Aalto University publication series
DOCTORAL THESES 48/2024

Spreading and Epidemic Interventions

Effects of Network Structure and Dynamics

Abbas K. Rizi

A doctoral thesis completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall T2 of the school on 15 March 2024 at noon.

Aalto University
School of Science
Department of Computer Science
Complex Systems Group

Supervising professor
Prof. Mikko Kivelä, Assistant Professor at Aalto University, Finland

Preliminary examiners
Prof. Clara Stegehuis, Associate Professor at Twente University, Netherlands
Dr. Riccardo Gallotti, Head of the Complex Human Behaviour, FBK, Italy

Opponent
Prof. Nicola Perra, Reader in Applied Mathematics, Queen Mary University of London, UK

Aalto University publication series
DOCTORAL THESES 48/2024

ISBN 978-952-64-1704-2 (printed)
ISBN 978-952-64-1705-9 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-64-1705-9

Images: The front cover's illustration portrays a Plague Doctor engaging with a modern contact tracing app, a fusion of historical imagery and contemporary technology. It was crafted using advanced AI through a specific 'prompt' command.

Unigrafia Oy
Helsinki 2024

Finland

NORDIC SWAN ECOLABEL
Printed matter
4041-0619

Author
Abbas K. Rizi

Name of the doctoral thesis
Spreading and Epidemic Interventions

Publisher School of Science

Unit Department of Computer Science

Series Aalto University publication series DOCTORAL THESES 48/2024

Field of research Network Science, Computational Epidemiology, Complex Systems, Statistical Physics

Manuscript submitted 5 December 2023 **Date of the defence** 15 March 2024

Permission for public defence granted (date) 13 February 2024 **Language** English

☐ **Monograph** ☒ **Article thesis** ☐ **Essay thesis**

Abstract

The COVID-19 pandemic has highlighted the critical importance of understanding epidemic dynamics, particularly the significant gaps in our knowledge that need addressing to better prepare for future pandemics. This thesis delves into the intricacies of disease spread within complex human interaction networks, underlining the pivotal role of individual connectedness in influencing epidemic outcomes. By developing theoretical models inspired by real-world epidemiological data, this work provides a nuanced exploration of disease transmission dynamics across networked populations, emphasizing the heterogeneous, spatial, homophilic, and temporal characteristics inherent in human social structures.

A primary focus of this research is the investigation of intervention strategies, encompassing pharmaceutical measures, such as vaccination campaigns, and non-pharmaceutical interventions, including contact tracing techniques. These interventions are evaluated within more realistic network topologies, characterized by degree heterogeneity and group structures, to assess their effectiveness in mitigating epidemic spread. The thesis leverages mathematical and computational epidemiology to offer profound insights into optimizing intervention strategies within the complex web of human interactions, thereby contributing to the academic discourse and providing actionable intelligence for public health policy formulation and epidemic preparedness.

The avenues of research opened by this work offer deeper insights into the mechanisms of epidemic spread in social networks. By using stylized modeling, the study was able to delve into the non-trivial ways epidemics spread through social networks. This modeling approach simplified the real-world dynamics into more analytically tractable forms, allowing the researchers to capture the essence of contact network structures and their crucial role in transmitting infectious diseases. The primary objective of this study was to identify new pathways for academic exploration and offer valuable perspectives that can enhance public health policies and epidemic response strategies. Ultimately, this work seeks to contribute to a better understanding of epidemic dynamics by bridging knowledge gaps and fostering a more resilient response to public health challenges in the face of complex human interactions.

Keywords Complex Systems, Network Science, Spreading Phenomena, Computational Epidemiology, Digital Epidemiology, Complex Networks, Temporal Networks, Reachability, Phase Transitions, Percolation, Covid-19, Vaccination, Contact Tracing

ISBN (printed) 978-952-64-1704-2 **ISBN (pdf)** 978-952-64-1705-9

ISSN (printed) 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2024

Pages 183 **urn** http://urn.fi/URN:ISBN:978-952-64-1705-9

In loving memory of Grandpa,
Mohammad Karimi (1944 – 2023)

Preface

It is now time to recap this rather extensive journey. To all who made my research possible, I am deeply grateful, and on top of that endless list comes my supervisor, Mikko Kivelä. Working with Mikko has been one of the few things I always count myself lucky for. I value how he walked me through my doctorate with his serene demeanor and unwavering patience. Thanks to Mikko, I worked with many excellent people in and out of Aalto. I am profoundly thankful to all my coauthors, especially Jari Saramäki, Márton Karsai, Ali Faqeeh, David O'Sullivan, and Zahra Ghadiri. With special thanks to Arash Badie-Modiri and Takayuki Hiraoka, with whom I had the most fruitful conversations while preparing our work.

In my doctoral studies, I participated in the NordicMathCovid Project, benefiting greatly from insightful discussions with Lasse Leskelä, Tom Britton, Pieter Trapman, Arnoldo Frigessi, Gianpaolo Scalia Tomba, and Lasse Engbo Christiansen. My sincere gratitude goes to each of them. My appreciation also extends to Petter Holme. Although we never got to work on something together, I have always enjoyed conversing with him and reading his remarkable blog. I also want to devote some words of gratitude to Yasser Roudi. With Yasser, I have engaged in the most vital discussions about the pursuit of science over the past four years.

In the dreary hallways of Aalto's CS building, where conversations were sparse, I was fortunate to have kind-hearted colleagues like Vesa Vahermaa, Yan Xia, Zhiren Huang, Hasti Narimanzadeh and Tarmo Nurmi. Their presence added a touch of warmth to the workspace, making it a more pleasant environment. My sincere thanks go to them for making my time there enjoyable. I also wholeheartedly thank my Iranian friends in Finland, who made my stay more welcoming: Morteza Shiripour, Babak Nasrinpay, Minoo Zarsav, Amirreza Akbari, and Mehdi Saman Booy. On this note, I want to thank those who came into my life as colleagues but remained friends. If this is not nice, I do not know what is! Sergei Kozlukov, Pegah Pournajafi, Pourya Toranj Simin, Aymara Baumann, and Rouzbeh Hasheminezhad, you are all stars!

A list of appreciation is complete with acknowledging my family, particu-

larly my mom and dad, to whom I am most indebted. I owe a great deal of thanks to my late grandpa and my uncle, Saïd, for his generous financial support. I will never forget my uncle, Masoud, for his candid advice; “You know, don’t take it personal, but business is not your thing. Go all in with science; you’re pretty good at math."

Looking back now, there is something more that I need to document here. It may reveal the ambiance of my thesis. In the middle of my doctoral studies, something happened in my personal life that brought hope to my perspective and color to my life. It was a beautiful mess but did not last long. It radiated away all its energy, leaving a void that could never be filled. Most of my doctoral studies have been spent with no genuine smile on my face. I was just traveling through the dark like in a Franz Kafka story. There were nights if a close friend had opened his arms to me, I would have wept like a child. Oh man, I should thank Reza Heydari and Omid MnZ. Without our conversations, my nights could have been much longer and my life darker. I owe them big time. Anyway, life did not get any better for me whatsoever. With my grandpa’s passing, I remained a regular at the house of despair.

> (In my sleep I dreamed this poem)
> Someone I loved once gave me a box full of darkness.
> It took me years to understand that this, too, was a gift.
>
> Mary Oliver, “*Thirst: Poems*”, 2007

All these nightmares, however, had nothing to do with the fulfillment of doing research. If there was no formula to derive or code to implement, I could find no good reason to leave my bed in the morning. Science has been the axe for the frozen sea within me. In these years, whenever I wanted to seek shelter from the world or its indifferent population, I came to the temple of science. If this work is now done, it just means I have put too much effort into making it through. If it were another time that an air of melancholy did not surround me, I could do it with much more ease, less pain, and perhaps in a different class. What comes next is a pure result of my faith in doing science and some discipline I grew up with since my whole life was a sink or swim. Come what may!

Lauttasaari, February 13, 2024,

Abbas K. Rizi

Contents

Preface 7

Contents 9

List of Publications 11

Author's contributions 13

List of Figures 17

0. About This Work 19
- 0.0.1 Motivation behind the thesis . . . 19
- 0.0.2 Structure of the thesis . . . 21
- 0.0.3 Scope of the thesis . . . 21

1. Mathematical Epidemiology in One Sitting 23
- 1.1 Things to Consider When Modeling Epidemics . . . 24
  - 1.1.1 Continuous and Discrete Time Modeling . . . 26
- 1.2 Compartmental Models in Epidemiology . . . 26
  - 1.2.1 The SI Model . . . 27
  - 1.2.2 The SIR Model . . . 29
  - 1.2.3 The Basic Reproduction Number . . . 33
  - 1.2.4 The SIS Model . . . 34
  - 1.2.5 The SEIR Model . . . 35
  - 1.2.6 The Next-generation Method and R_0 . . . 37
  - 1.2.7 The Next-generation Matrix for Compartmental Models . . . 38
  - 1.2.8 Beyond the SIR Model . . . 40
- 1.3 Herd Immunity and Failures of Simple Models . . . 41
  - 1.3.1 Herd Immunity after Vaccination . . . 42
  - 1.3.2 Herd Immunity through Natural Infection . . . 43

2. Structured Populations and Networks 47
- 2.1 Random Graph Theory: Key Concepts and Definitions . . . 48

2.2 Erdős–Rényi Networks 50
2.2.1 Emergence of the Giant Component 52
2.2.2 Shortcomings of ER Networks in Modeling Contact Networks 54
2.3 The Configuration Model 56
2.3.1 Existence of the Giant Component 58
2.3.2 Size of Giant Component and Generating Functions 58
2.4 Random Graph Models For Networks with Group Structure 59
2.4.1 Bipartite Network 60
2.4.2 Random Clique Networks 60
2.4.3 Stochastic Block Models 61
2.4.4 Networks with Homophily 62
2.5 Spatial Random Graphs 63
2.5.1 Random Geometric Graphs 63
2.6 Lattices 64
2.7 Temporal Networks 64
2.8 Percolation Theory 67
2.8.1 Bond Percolation 67
2.8.2 Site Percolation 69
2.8.3 Percolation as a Critical Phenomenon 69
2.8.4 Directed Percolation 70

3. Epidemics on Networks **73**
3.1 Identifying Key Nodes in Networks 73
3.2 Modeling Epidemics on Networks 74
3.3 Mean-field Approximations 75
3.4 From the Contact Networks to Transmission Networks . . 75
3.5 Mapping Epidemics to Percolation 77
3.5.1 SIR Model and the Configuration Model 77
3.6 Spreading on Temporal Networks 78
3.6.1 Directed Percolation and Spreading Phenomena . 79
3.7 Homophily and Herd Immunity Threshold 80
3.8 Disease-induced Herd Immunity 82
3.9 Epidemic Spreading and Contact Tracing 84
3.9.1 Digital Contact Tracing 85
3.9.2 Contact Tracing and Social Groups 87

4. Conclusion **91**

References **93**

Publications **107**

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I T. Hiraoka, A. K. Rizi, J. Saramäki and M. Kivelä. Herd immunity and epidemic size in networks with vaccination homophily . *Physical Review E*, 105(5) L052301, May 2020.

II T. Hiraoka, A. K. Rizi, Z. Ghadiri, J. Saramäki and M. Kivelä. The strength and weakness of disease-induced herd immunity. *Presented at NetSci 2023 Conference*, pre-print: arXiv:2307.04700, Jul 2023.

III A. K. Rizi, L. A. Keating, J. P. Gleeson, David J.P. O'Sullivan and M. Kivelä. Effectiveness of Contact Tracing on Networks with Cliques. *Physical Review E*, 109, 024303, Feb 2024.

IV A. K. Rizi, A. Faqeeh, A. Badie-Modiri and M. Kivelä. Epidemic spreading and digital contact tracing: Effects of heterogeneous mixing and quarantine failures. *Physical Review E*, 105(4) 044313, April 2022.

V A. Badie-Modiri, A. K. Rizi, M. Karsai and M. Kivelä. Directed percolation in random temporal network. *Physical Review Research*, 4(7) L022047, May 2022.

VI A. Badie-Modiri, A. K. Rizi, M. Karsai and M. Kivelä. Directed percolation in random temporal network models with heterogeneities. *Physical Review E*, 105(17) 054313, May 2022.

Author's contributions

Publication I: "Herd immunity and epidemic size in networks with vaccination homophily "

A. K. Rizi played a pivotal role in the development of this work and in shaping the core aspects of the research alongside the primarily responsible author, T. Hiraoka. A. K. Rizi and T. Hiraoka jointly conducted the majority of simulations and mean-field calculations. A. K. Rizi participated in conceptualizing the idea, conducting an extensive literature review, and composing significant sections of the manuscript. M. Kivelä and J. Saramäki supervised the project. The refinement of the manuscript was a collaborative effort, with all project members contributing valuable input. J. Saramäki revised the final version of the manuscript.

Publication II: "The strength and weakness of disease-induced herd immunity"

A. K. Rizi played a crucial role in shaping the research's foundational aspects and its development. He was responsible for originating the research concept, conducting a comprehensive literature review, designing the initial simulations, and building the frameworks and necessary pipelines for the project. Moreover, A. K. Rizi contributed to writing key sections of the manuscripts. T. Hiraoka, the principal author, primarily handled the analytical calculations and led the project. Z. Ghadiri's extensive involvement included model implementation, measure design, and simulation execution. This project was done under the supervision of M. Kivelä and J. Saramäki. The refinement of the manuscript was a collective effort, with all members contributing.

Publication III: "Effectiveness of Contact Tracing on Networks with Cliques"

The primary author of this paper, A. K. Rizi, led the comprehensive development of the research, encompassing the initiation of the initial manuscript, software implementation, analytical and mean-field derivations, practical execution of experiments, literature review, and composition of the dedicated manuscript section. D. O'Sullivan contributed to the analytical calculations, simulation implementation, literature review, and manuscript writing. Additionally, D. O'Sullivan was actively involved in crafting relevant content within the manuscript. M. Kivelä played a significant role by contributing to the numerical calculations for the M matrix and providing a theoretical rationale for the results. L. Keating participated in the revision process of the paper. The project benefited from the oversight of D. O'Sullivan, J. Gleeson, and M. Kivelä. Notably, all project members actively participated in the rigorous revision process of the manuscript.

Publication IV: "Epidemic spreading and digital contact tracing: Effects of heterogeneous mixing and quarantine failures"

A. K. Rizi served as the primary author of this paper, having undertaken initial manuscript drafting, software implementation, and analytical calculations. A. K. Rizi and M. Kivelä had played integral roles in mean-field derivations, analytical outcomes, literature review, and the composition of the dedicated manuscript section. A. Badie-Modiri contributed to the compartmental model section of the article, encompassing model implementation, measure design, simulation execution, and the corresponding manuscript section. The overall project was conducted under the guidance of A. Faqeeh and M. Kivelä, with A. Faqeeh also contributed partially to software implementation. The refinement of the manuscript involved input from every project member.

Publication V: "Directed percolation in random temporal network"

A. Badie-Modiri led the paper's creation, undertaking initial manuscript drafting, software implementation, and experimental execution. A. K. Rizi and M. Kivelä contributed to mean-field derivations, analytical results, literature review, and manuscript composition for that section. The research idea was conceived by M. Karsai and M. Kivelä, building upon previous work. Manuscript refinement was done by all the authors, and project guidance was collectively overseen by M. Karsai and M. Kivelä.

Publication VI: "Directed percolation in random temporal network models with heterogeneities"

A. Badie-Modiri led the paper's creation, undertaking initial manuscript drafting, software implementation, and experimental execution. A. K. Rizi and M. Kivelä contributed to analytical results, literature review, and manuscript composition for that section. The research idea was conceived by M. Karsai and M. Kivelä, building upon previous work. Manuscript refinement was done by all the authors, and project guidance was collectively overseen by M. Karsai and M. Kivelä.

List of Figures

1.1 Logistic growth curve in the SI Epidemic Model 28
1.2 Evolution of compartments in SIR model 30
1.3 Flattening The Curve Strategy 31
1.4 Distribution of Recovery Time 33
1.5 Epidemic Phase Transition 34
1.6 Fraction of infected people in the SIS model 35
1.7 Epidemic Size in the SIS and SIR model 36
1.8 Incubation Period 37
1.9 Vaccination Homophily 43
1.10 Localized Immunity 44

2.1 Disconnected graph with two cliques 49
2.2 Phase transition in ER network 53
2.3 Divergence in the size of small components 54
2.4 r-regular c-clique networks 61
2.5 Visual representation of a temporal network 66
2.6 Percolation Phase Transition 68

3.1 Herd Immunity Threshold with Homophily 82
3.2 Degree Heterogeneity and Spatiality Space 83
3.3 Contact Tracing with Loops 87
3.4 Diffusion Patterns of a 3-clique 88
3.5 Phase diagrams for Clique Networks 90

0. About This Work

> The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work - that is correctly to describe phenomena from a reasonably wide area. Furthermore, it must satisfy certain esthetic criteria - that is, in relation to how much it describes, it must be rather simple.
>
> John von Neumann, "*Method in the Physical Sciences*", in The Unity of Knowledge (1955)

This chapter offers an overview of the thesis. This thesis includes a didactic introduction that precedes the six featured publications.

0.0.1 Motivation behind the thesis

The COVID-19 pandemic, a stark reminder of our vulnerability to infectious diseases, was neither the first nor will it be the last pandemic humanity confronts. This recent global health crisis underscores the urgent need for preparedness against future pandemics. Pandemics, by their very nature, can be devastatingly deadly, rapidly overwhelming healthcare systems and causing widespread social and economic disruptions. The unpredictable nature of viral mutations and the interconnectedness of our world only exacerbate these risks. Therefore, understanding the mechanisms of pandemic spread, including the role of social networks and human behavior, is crucial [164, 165, 158, 115]. This knowledge not only aids in developing effective intervention strategies [163] but also helps mitigate the severe consequences that unchecked pandemics can impose on

global health, economies, and societies [124]. Our future resilience lies in learning from past experiences, investing in public health infrastructure, advancing scientific research, and fostering global cooperation to manage and ultimately prevent the devastating impact of pandemics.

This research explores the intricate dynamics of how diseases spread through the complex networks of human societies [151], evaluating the effectiveness of different interventions within these interconnected populations [113, 112, 80, 78]. It emphasizes the critical role that the detailed connections between individuals play in influencing the spread and severity of epidemics [150]. The study introduces theoretical frameworks that better reflect certain aspects of real-world situations. A significant aspect of *network epidemiology* [53] examined in this work is how the structural and spatial characteristics of networks influence the effectiveness of herd immunity [92]. The focus is on deciphering the structure and dynamics of contact networks [177, 13, 12], particularly in understanding how specific health behaviors influence disease transmission [105, 93]. The research employs stylized models to demonstrate the significant impact of homophily in hesitancy in adopting Contact Tracing Apps and vaccine reluctance on achieving herd immunity [184, 105, 93]. The aim of these models is not to comprehensively represent the complexities of actual epidemics but to illustrate the potential effects of certain network structures and health behaviors, as well as non-pharmaceutical interventions like contact tracing, on the outcomes of epidemics.

Building on this foundation, the ultimate goal of this thesis is to highlight the importance of the contact network structure in modeling disease spreading and act as a guide for showing what kind of network structures can be crucial in building epidemic models. It's important to note that our approach to understanding the spread of infectious diseases follows a "modeling for insight" methodology. Our goal is to gain a deeper understanding of the underlying dynamics of these diseases, so we can provide insights into more effective public health strategies and interventions. I want to emphasize that our *spherical cow* models [30] are not intended to fit data directly. Instead, they aim to broaden the conceptual framework within which we understand and respond to epidemics. For readers keen on exploring a data-driven approach to modeling and predicting the global spread of infectious diseases using real-world data, I suggest diving into "Charting the Next Pandemic" by y Piontti et al. [206]. Our work contributes to academic discourse and provides a foundation for future empirical research in support of society's ongoing efforts to enhance resilience against infectious diseases.

0.0.2 Structure of the thesis

In Chapter 1, I present an introductory survey of Mathematical Epidemiology, focusing on compartmental models. The chapter delves into the historical context of translating epidemiological challenges into mathematical frameworks. Central to this exploration is the mass-action or fully-mixed population hypothesis, which enables the application of calculus tools to model population dynamics using differential equations. I comprehensively review essential epidemiological concepts and demonstrate their translation into mathematical terms. This chapter lays the groundwork for our understanding of how to formalize models that describe the spread of epidemics. I conclude with a discussion on the limitations of the models discussed, particularly the inadequacy of the fully-mixed population assumption for complex, structured human populations. I illustrate this point by examining vaccination strategies, showing how conventional infectious disease models fall short of accurately predicting vaccine quantities required for achieving herd immunity to ensure the safety of an entire population, including the unvaccinated.

In Chapter 2, I provide an accessible introduction to network science, outlining the essential concepts and methodologies that will be applied to more accurately model infectious diseases. I explore the critical aspects of human population structure that can significantly influence epidemic outcomes. The chapter navigates various random network models, their characteristics, and the phase transition phenomena observable in static and temporal network contexts.

Equipped with the insights and techniques from these two chapters, we embark on a detailed study of epidemic modeling on complex networks. In Chapter 3, I discuss how our foundational knowledge of spreading and diffusion processes can be applied to quantify epidemic behavior on networks. The chapter introduces various mean-field approximations and other modeling techniques. I emphasize that epidemic modeling on networks can reveal non-trivial dynamics, which are crucial for understanding effective epidemic control.

0.0.3 Scope of the thesis

A patient reader with a good background in science can follow the research papers presented in this work after reading the following three chapters. The theoretical minimum required for understanding them is a working knowledge of probability theory, calculus, and linear algebra. I also recommend Newman's book [151] as a network textbook and Diekmann et al. book [55] for a general understanding of mathematical epidemiology.

On another note, you should know that I grew up in the physics culture, and I like borrowing the tools and perspectives of that field and trying to

combine them with new ideas and techniques to tackle interdisciplinary problems. So, in terms of technicality and jargon, what follows in this work is not only written with the notation that is the best to the taste of a physicist but also the mathematical rigor that I have gone through is more physics-friendly than mathematics-friendly. A background in physics, with a focus on statistical mechanics, ideally equips you to engage with and understand the complexities of this work.

This is not a *theorem and proof* thesis, but I have mentioned resources where you can find more rigorous mathematical arguments. So, if you have a math background, you will have many exciting puzzles to think about after reading each chapter. I also invite you to delve into some proof to turn this work into an even better investment in your time. For a more mathematical approach to networks, I recommend referring to Random Graphs and Complex Networks by Van der Hofstad [194]. Depending on their particular field of study, computer scientists should also be capable of comprehending this work, albeit not as effortlessly as physicists. If you are more from an engineering background, you will find simulations in our papers very interesting. While this work leans heavily on mathematical modeling, it stands as an inviting challenge for epidemiologists keen on extending their professional skill set in this exciting direction. Moreover, they may find some of our models not precisely suitable to describe some particular phenomenon, as in this work, we have done our best to come up with reasonably simple mathematical models to explain the physics behind spreading phenomena and, more importantly, get the phenomenology right. Believe me, it is no less deserving of scrutiny!

In conclusion, writing this work has been a thoroughly enjoyable experience for me. Above all, I sincerely hope you find equal pleasure in reading it.

> Like most mathematicians, he takes the hopeful biologist to the edge of a pond, points out that a good swim will help his work, and then pushes him in and leaves him to drown.
>
> C. Elton, *1935 review on the mathematical ecology work of Lotka.*

> **What is physics?** To me, ... the central idea was that the world is understandable, that you should be able to take anything apart, understand the relationships between its constituents, do experiments, and on that basis be able to develop a quantitative understanding of its behavior.
>
> John Hopfield, 2018, *Now What?*

1. Mathematical Epidemiology in One Sitting

> I simply wish that, in a matter which so closely concerns the well-being of the human race, no decision shall be made without all knowledge which a little analysis and calculation can provide.
>
> D. Bernoulli, 1760, on smallpox inoculation

It is not easy to say precisely when and by whom the foundations of Mathematical Epidemiology were established. Here, by Mathematical Epidemiology, I primarily refer to a rigorous quantitative framework for studying how infectious diseases spread, predicting what might happen during an outbreak, and figuring out how to control it. One can argue that John Graunt, a businessman admitted to the Royal Society, lit the torch in the 17th century with a new approach to the analysis of causes of death in London [145]. Graunt did an excellent job of bookkeeping the so-called "Bills of Mortality" which were weekly records of numbers and causes of death in London parishes from 1592 to 1603 and tried to quantify causes of death in a systematic way [84]. Based on real data, he predicted the percentage of people who would live to different ages and their year-by-year life expectancy. Among his observations, he also noticed that in cities, more people died than in rural areas, and while more boys were born than girls, higher male mortality balanced it out, resulting in a nearly equal gender distribution in the population [171].

I admire Graunt's contribution to the field of epidemiology. His innovative analysis of the tables of health data (Bills of Mortality) is commendable, and from a technical perspective, his work can be recognized as seminal in *Quantitative* Epidemiology. However, when we consider the monumental significance of Daniel Bernoulli's 1766 paper, it becomes clear that this work forms the cornerstone of *Mathematical* Epidemiology. Graunt may have lit the torch, but Bernoulli's work led to a wildfire! Bernoulli introduced the first compartmental model of infectious disease in that work and used mathematical reasoning to argue for universal inoculation

against smallpox. Compartmental models in epidemiology allow us to strip unnecessary complicating factors away from disease dynamics to gain a reasonable understanding of such spreading phenomena. I will explain compartmental models and the effectiveness of interventions in a more pedagogical way in this section. For a more historical note on Bernoulli's epidemiological model, I recommend referring to Ref. [60]. Ref. [55] also serves as an excellent guide for those aiming to learn the formulation and analysis of mathematical models in infectious disease epidemiology.

> As a matter of fact all epidemiology, concerned as it is with variation of disease from time to time or from place to place, **must** be considered mathematically, if it is to be considered scientifically at all. And the mathematical method of treatment is really nothing but the application of careful reasoning to the problems at hand.
>
> Sir Ronald Ross, 1911, *The Prevention of Malaria*

1.1 Things to Consider When Modeling Epidemics

Let us begin with the epidemic in a closed population in a demographic steady state, with no history of a given infection or introduction of any intervention. We are also interested in spreading phenomena that take place in *naive* populations at particular time scales, often much smaller than fluctuations in population due to the birth or death of people out of epidemiological reasons. By naive, we mean every individual lacks immunity or immunologic memory to a disease and is susceptible to infection.

Imagine a population with only one infected person (seed of infection) at day zero. If an infected person can infect $R_0 > 1$ more susceptible people, this process would undergo a deterministic exponential growth such that the number of infected people after t generations, ε_t, would be given by the following geometric series:

$$\varepsilon_t = 1 + R_0 + R_0^2 + \ldots + R_0^t = \frac{R_0^{t+1} - 1}{R_0 - 1}. \tag{1.1}$$

We call ε_t the epidemic size after t generations. The first lesson from the mathematics of this equation is that if $R_0 < 1$, then in finite time, the disease spreading will die out. Compared to the population size, the epidemic size would be negligible even if we start with more than one infected person. For mathematical tractability, we model the epidemics in the thermodynamic limit, meaning that we assume the size of the population, a.k.a. the system size, N, is large enough so that we can follow the problem in the limit that $N \to \infty$. Therefore, the size of the

epidemic compared to the population size will be zero when $R_0 < 1$. Even if we consider that every person can, on average, infected R_0 susceptible people and R_0 is drawn from some distribution with finite first and second moments, one can easily show we again end up with a similar conclusion, this time through a stochastic growth know as Galton–Watson process [55].

However, since R_0, here, represents the expected number of people secondarily infected, which is an integer, the epidemic's size will increase monotonically without an epidemic threshold. According to this model, a massive outbreak comparable in scale to the population size will invariably occur swiftly.

Disease	**Basic Reproduction Number**
Measles	12–18
Chickenpox (Varicella)	10–12
Mumps	10–12
Rubella	6–7
Polio	5–7
Pertussis (Whooping Cough)	5.5
Smallpox	3.5–6.0
HIV/AIDS	2–5
COVID-19 (Ancestral Strain)	2.9 (2.4–3.4)
SARS	2–4
Diphtheria	2.6 (1.7–4.3)
Common Cold (e.g., Rhinovirus)	2–3
Mpox	2.1 (1.5–2.7)
Ebola (2014 Outbreak)	1.8 (1.4–1.8)
Influenza (Seasonal Strains)	1.3 (1.2–1.4)
Andes Hantavirus	1.2 (0.8–1.6)
Nipah Virus	0.5
MERS	0.5 (0.3–0.8)

Table 1.1. Basic Reproduction Numbers of Various Infectious Diseases [59, 54, 27].

Table 1.1 compares the average number of secondary infections, commonly referred to as the basic reproduction number, for various diseases. Assume a hypothetical scenario such that $R_0 = 3$, which is a reasonable number for some severe infectious diseases [59, 54] as it was estimated for some analysis during the COVID-19 pandemic [27]. Then, when $t = 10$, more than 88,000 people will be infected, and after 30 generations, it will reach more than 100 trillion people, much larger than the entire earth's

population. This astronomically big number does not make sense at all, as, in reality, many boundary conditions control such exponential growth. Obviously, there should be a cut-off; the process must cease when everyone is infected. We explicitly did not account for reinfection or recovery in this model. People may recover from the disease in a different time scale and gain immunity against reinfection. Therefore, in practice, every infected person cannot keep infecting R_0 more people as the spreading process unfolds. Not everyone has access to enough susceptible cases after the epidemic's early stages. Ultimately, the exponential growth should be saturated as we run out of susceptible people. On top of these, people change their social behavior [176] as they face a wave of infection, so either individuals on their own try to protect themselves from infection or some external effects such as seasonality or high-level social constructs such as governments will impose regulations and apply interventions to curb the epidemic. Therefore, for more realistic modeling, we need to consider all these scenarios, one at a time, and then try to consider them all together to the extent that we can handle the arising complexity of the problem.

1.1.1 Continuous and Discrete Time Modeling

Epidemic modeling on networks employs continuous and discrete time models, each with distinct advantages [161, 42, 198]. Continuous-time models, often formulated with non-linear differential equations, are helpful for uninterrupted changes but may not always accurately represent the epidemic spread [152]. On the other hand, discrete-time models, described by difference equations, align better with the discrete nature of epidemic data collection and are easier to implement in computer simulations. They may also offer a more granular view of the spread, accounting for specific time intervals and potentially capturing the spatial dependencies between individuals more accurately. Both methods are valuable, and we will use them in modeling epidemics on networks depending on our modeling requirements and objectives. In this section, we will continue with continuous-time compartmental models as they are beneficial for pedagogical purposes, leveraging calculus tools to elucidate the dynamics of disease spread. We will later show how we can formulate such models within discrete-time settings. In our publications, we utilize both continuous and discrete-time modeling approaches to capture the nuances of disease spread in networks.

1.2 Compartmental Models in Epidemiology

A convenient way to model epidemics is to divide people into groups, commonly known as compartments, and develop equations that govern how

one person leaves one and joins the other. Compartmental models provide a useful abstraction of disease dynamics by making it easier to track and predict the course of an epidemic, as the equations describe the rate at which people move from one compartment to another.

1.2.1 The SI Model

In the previous example, every individual was either susceptible or infected in a population of size, say, N and preferably $N \to \infty$. So, one can assign each individual to compartments with labels `S` and `I`, respectively, such that the size of the compartments respects the conservation of population size, $S + I = N$. Compartment `S` initially begins with $N - 1$ members, and people transition from that compartment to compartment `I` following Eq. 1.1. This model is an SI model, as there are only two compartments `S` and `I`, and people may progress between them following the order of labels in the title; `S` → `I`.

In Section 1.1.1, we explored how infectious diseases can be modeled in both continuous and discontinuous time frames. For simplicity, we will proceed with our epidemic modeling using a continuous-time approach. Additionally, we will address the issue of unbounded exponential growth identified in the earlier model. This will be achieved by employing the *fully mixed* or *mass-action* approximation [69]. Our method involves representing the interactions between different compartments using differential equations. Depending on our modeling perspective, the interactions between the compartments can be given through some rate equations or transition probabilities—the former leads to deterministic dynamics, and the latter leads to stochastic ones. So, given that each person, on average, interacts with β other randomly chosen people per unit time, we can rewrite the previous model with the following equations,

$$\begin{aligned} \frac{\mathrm{d}s}{\mathrm{d}t} &= -\beta s i, \quad \text{or/and} \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \beta s i\,, \end{aligned} \tag{1.2}$$

where s and i are the expected numbers of susceptible and infected individuals normalized by the population size N such that $s + i = 1$. Note that as they are average values, they may not be integers, in general, while the actual numbers of susceptible and infected people are always integers. If we were to repeat the epidemic dynamics given by Eq. 1.2 multiple times under the same conditions and then take the average, these would be the values we'd arrive at.

We can eliminate s from Eq. 1.2 as $s = 1 - i$ and easily arrive at the solution that yields a "logistic growth curve" for the fraction of infected

individuals given by

$$i(t) = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}}, \tag{1.3}$$

where $i_0 = i(0) \ll 1$ is the size of the seed of infection normalized by the population size. According to this equation, as depicted in Fig. 1.1, the expected number of infected individuals initially grows exponentially with the characteristic time $\frac{1}{\beta}$, and then saturates for large t, as $t \to \infty$.

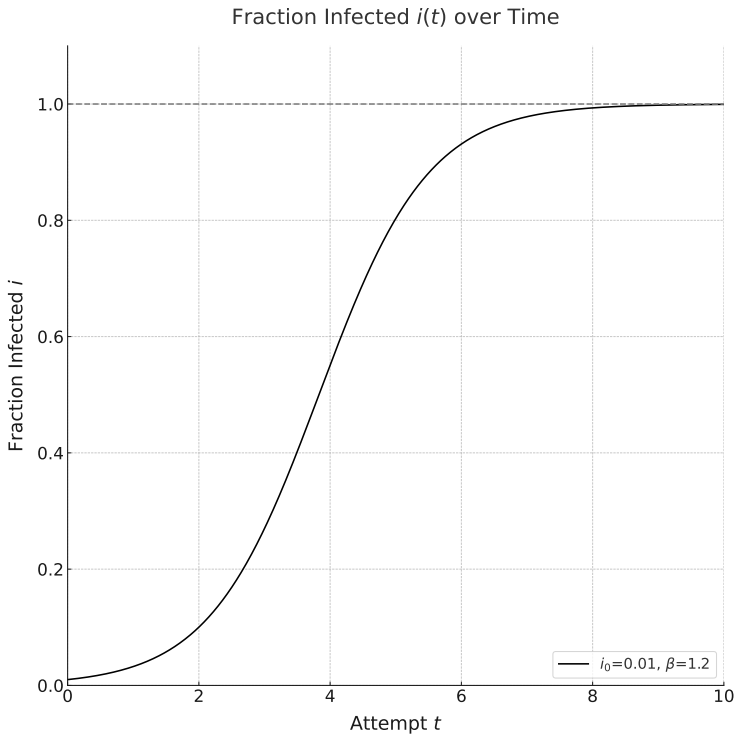

Figure 1.1. Logistic growth curve in the SI Epidemic Model as given by Eq. 1.2. The expected fraction of infected people within an SI model initially escalates exponentially. However, as the reservoir of susceptible individuals depletes over time, the growth rate reaches a saturation level, leading to a plateau in the curve as indicated by the dashed line.

As highlighted in Sec. 1.1, the SI model is overly simplistic for most practical applications. Despite its exponential growth, which aligns with many spreading phenomena in their early stages, it leads to the infection of the whole population when $i_0 > 0$, overestimating the final size of most outbreaks. The epidemic size may rapidly increase at the very early stages of an epidemic, but it slows down at higher rates due to natural immunity or facing interventions, and it almost never gets that close to the population size. Nevertheless, we can still learn from the SI model, as in the early stages of most epidemics, the number of infected people grows exponentially with the per-individual rate β similar to $i(t \approx 0) = i_0 e^{\beta t}$.

1.2.2 The SIR Model

A reasonable improvement to the SI model is to consider that people recover from the disease after some time and gain immunity against reinfection. So, we add a third compartment to our model, letting people leave the infected compartment with some rate and join the recovered compartment R, giving rise to the transition pattern S → I → R.

This three-state model is called the *susceptible–infected– recovered* or SIR model. You can find "SIR" in the literature as an acronym for different terms. Some may use *removed* instead or along with the recovered state as, in general, the person is now removed from interacting with others in the population. Regrettably, some diseases result in death rather than recovery, so the dead person is removed. For the sake of modeling, both of them can be considered in this model, and individuals in the R compartment would not take part in infecting other people. In concise terms, R is where people transition from the I compartment. In Sec. 1.3.1, we will demonstrate how vaccination can prevent individuals from contributing to the spread of infection, and in Sec. 3.9, we will explore how isolating or quarantining people can yield similar results. I can also stand for infected or infectious, as being infected does not necessarily imply being contagious. One may get infected but may never pass it on or do it after some period of time. However, from now on, everywhere we say infected, we mean the same thing as infectious, and we use them interchangeably unless otherwise noted.

To describe the dynamics of the SIR model, we use the same mass-action approximation and per-individual rate β as we used in Sec. 1.2.1, but this time, we let people leave the I compartment, with the constant average rate γ so that the fraction of recovered people, r, changes as $\frac{\mathrm{d}r}{\mathrm{d}t} = \gamma i$. Here, the conservation of population size implies $s + i + r = 1$, and the changes in each compartment would be summarized as:

$$\begin{aligned} \frac{\mathrm{d}s}{\mathrm{d}t} &= -\beta s i, \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \beta s i - \gamma i, \\ \frac{\mathrm{d}r}{\mathrm{d}t} &= \gamma i. \end{aligned} \tag{1.4}$$

There is no closed-form solution for the SIR model. However, a numerical solution is presented in Fig. 1.2.

Looking at Fig. 1.2, it is interesting to observe how the infection curve ascends, slows down, and then descends. This curve illustrates the real-time count of infected individuals, which is quite useful for a range of practical situations. For instance, in a city with limited hospital capacity, if the maximum of this curve surpasses what the hospitals can manage,

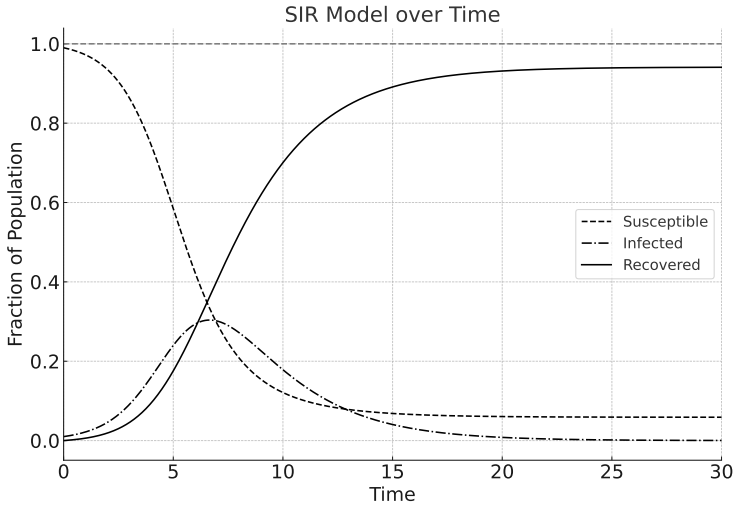

Figure 1.2. Evolution of the susceptible, infected, and recovered populations over time, based on the SIR model given. The model was integrated using parameters $\beta = 1.2$ (transmission rate) and $\gamma = 0.4$ (recovery rate). Initial conditions were set with 99% of the population as susceptible, 1% as infected, and 0% as recovered. The Infected curve is represented with a solid line for a clear distinction. The horizontal dashed line indicates the maximum value of 1 for the fraction of the population, which will always be larger than the fraction of recovered people.

things could get pretty challenging, especially with severe diseases in the mix. The "Flattening The Curve" strategy was all about promoting actions to lower this peak, aiming to ease the pressure on the healthcare system as the maximum number of individuals needing medical attention at any given time is minimized [135]. As shown in Fig. 1.3, this ensures the healthcare system is not overwhelmed. During the COVID-19 pandemic, key measures to achieve this were hand hygiene, wearing face masks, and practicing social distancing [136] to reduce the value for β.

Fig. 1.2 also shows that the value of r rises monotonically, yet it never gets to one. Similarly, the susceptible curve does not touch zero. This suggests that a portion of the population stays susceptible over time. The value of r reflects the total fraction of infected people up to time t, so the asymptotic value of r indicates the total size of the epidemic. From now on, we will always report the size of the epidemic as a fraction of the whole population.

We cannot solve the SIR model analytically, as it is not solvable by elementary functions in its standard form [89, 123]. However, we can study it at the $t \to \infty$ limit. Using equation Eq. 1.4, after some algebra, we can show the changes in r in a large population with a small seed of

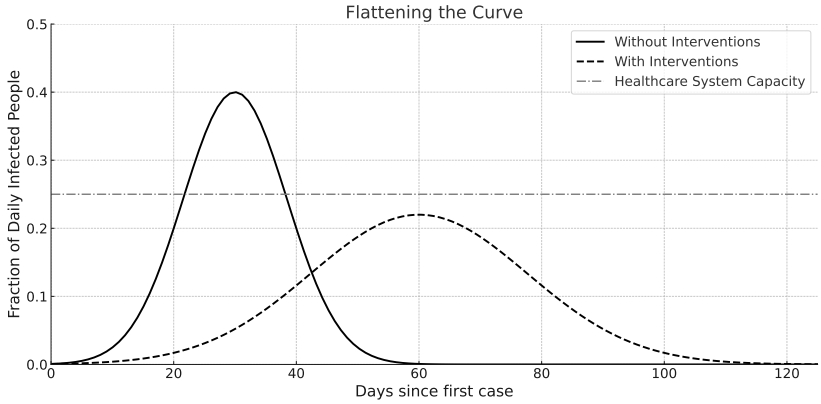

Figure 1.3. Time evolution of $i(t)$ for two different values of β. By adopting measures like hand washing, social distancing, and wearing face masks, the peak of active cases can be both lowered and postponed. This delay provides valuable time for healthcare systems to expand and handle the influx of patients more effectively. Consequently, this strategy of flattening the curve also offers an opportunity to enhance healthcare capabilities to accommodate the increased demand.

infection, can be given as [151]

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \gamma(1 - r - e^{-\beta r/\gamma}). \tag{1.5}$$

When $t \to \infty$, $\frac{\mathrm{d}r}{\mathrm{d}t} = 0$ which gives

$$r_\infty = 1 - e^{-\beta r_\infty/\gamma}, \tag{1.6}$$

where r_∞ is the final epidemic size. This equation is transcendental in r_∞, and one typically finds the solution graphically. They would also know that $r_\infty = 0$ is a trivial solution, but its stability as a fixed point of iteration depends on the value of β/γ. If $\beta \leqslant \gamma$, then $r_\infty = 0$ is a stable fixed point, which makes sense intuitively, as it suggests that if people recover from the disease faster than new infections occur, the disease will eventually die out exponentially fast. In simpler terms, in the long run, and the thermodynamic limit, the epidemic size will tend to be zero when recovery outpaces new infections. However, if $\beta > \gamma$, then $r = 0$ would be an unstable solution, and the final epidemic size would be between zero and one.

To analyze the stability of the solution $r = 0$ for Eq. 1.5, we can use linear stability analysis. This involves determining whether small perturbations around the equilibrium solution $r = 0$ grow or decay over time. The general approach is to linearize the differential equation around the equilibrium point. Let's denote a small perturbation around $r = 0$ by δ. When $r = \delta$ and δ is small, we can approximate the exponential term using the first term of its Taylor expansion, $e^{-\beta\delta/\gamma} \approx 1 - \beta\delta/\gamma$, since higher-order terms will be

negligibly small. Substituting this into the differential equation, we get:

$$\frac{\mathrm{d}\delta}{\mathrm{d}t} \approx \gamma(1-\delta-(1-\beta\delta/\gamma)) = \gamma\left(\frac{\beta\delta}{\gamma}-\delta\right) = \gamma\delta\left(\frac{\beta}{\gamma}-1\right).$$

So the linearized differential equation becomes:

$$\frac{\mathrm{d}\delta}{\mathrm{d}t} \approx \delta(\beta-\gamma).$$

From this, we can see that the sign of $\beta-\gamma$ determines the behavior of small perturbations δ: If $\beta>\gamma$, then $\beta-\gamma>0$ and any small perturbation δ will grow exponentially, which implies that the solution $r=0$ is unstable. If $\beta<\gamma$, then $\beta-\gamma<0$ and any small perturbation δ will decay exponentially, suggesting that the solution $r=0$ is stable. Therefore, when $\beta>\gamma$, the solution $r=0$ is unstable because small deviations from this point will increase over time rather than return to equilibrium.

The transition between having an epidemic and not having one occurs at a specific point where $\beta=\gamma$, known as **the epidemic threshold**. It's important to contrast this with the simpler SI model, where there is no such threshold. In the SI model, the disease spreads because once individuals are infected, they never recover. This fundamental difference means that the number of infected individuals in the SI model can never decrease as opposed to the SIR model, where there's a critical point where the dynamics change, leading to the possibility of disease containment. In other words, the SI can be considered as the special case of the SIR model with $\gamma=0$ so that β is larger than γ by design.

Given the value of γ, we can calculate for how long, τ, an infected person is likely to remain infected before recovery. The probability of recovering in any infinitesimal time window Δt is $\gamma\Delta t$, and the probability of not doing so is $1-\gamma\Delta t$ [151]. This means that the person is still infected after a total time τ with probability

$$\lim_{\Delta t\to\infty}(1-\gamma\Delta t)^{\tau/\Delta t} = e^{-\gamma\tau}. \tag{1.7}$$

So, the probability that the person stays infected for time τ and recovers $\mathrm{d}t$ later would be $e^{-\gamma\tau}\times\gamma\mathrm{d}t$ leading to an exponential distribution [151]

$$p(\tau) = \gamma e^{-\gamma\tau}, \tag{1.8}$$

with mean infectious time $1/\gamma$. The artifact of this model is that the distribution of times for which an individual remains infected is not that realistic. Based on Eq. 1.8, it is most likely for an individual to recover just after infection. However, this likelihood diminishes exponentially over time. Theoretically, a person could stay infected for a duration significantly longer than the average infectious period, denoted as $1/\gamma$. As depicted in Fig. 1.4, for some diseases, people remain infected for a typical range of

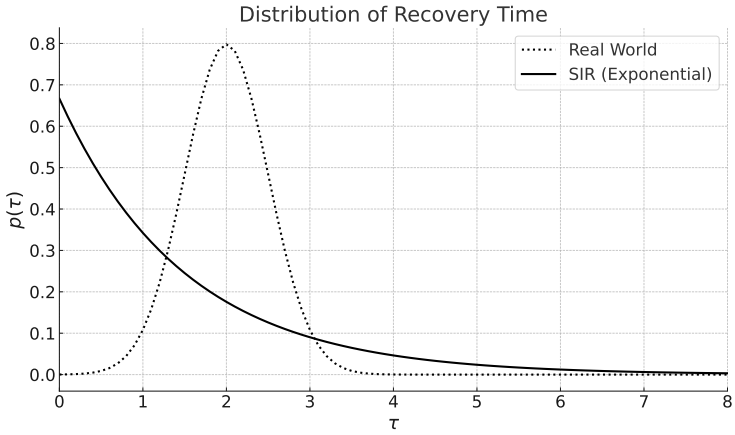

Figure 1.4. The SIR model suggests an individual has the highest likelihood of recovery immediately post-infection, with chances decreasing exponentially over time. However, there's typically a consistent recovery timeframe for many diseases, often peaking around a specific period, such as one or two weeks, as highlighted in the depicted distribution.

time, making its distribution narrowly peak around some expected value, like one or two weeks [152].

We can employ $p(\tau)$ to determine the average number of individuals that an infected person will infect before recovery, represented as R_0, in a naive population. Given that an individual stays infected for a duration τ, they would, on average, result in $\beta\tau$ new infections due to their contacts. By averaging over the distribution of τ, we can derive the average number R_0 [151]:

$$R_0 = \beta\gamma \int_0^\infty \tau e^{-\gamma\tau} \mathrm{d}t = \frac{\beta}{\gamma}. \tag{1.9}$$

We call R_0 the basic reproduction number as an indicator of how many new infections can happen after a previous one [57]. Equation 1.9 offers an alternative approach to deducing the epidemic threshold for the SIR model: the threshold is at $R_0 = 1$, consistent with our earlier findings based on the long-term behavior of the I compartment mimicked with Eq. 1.6.

1.2.3 The Basic Reproduction Number

In the deterministic approach, compartmental models are often described with ordinary differential equations, which, by design, have two fixed points: disease-free equilibrium and endemic equilibrium. The stability of these fixed points will be determined by evaluating the value of the bifurcation parameter R_0. Following Physics' jargon, when $R_0 < 1$, we say the epidemic spreading is in the sub-critical regime. When $R_0 = 1$, we are at the critical point with many exciting phenomena [81], which I will cover later in this thesis. An outbreak happens when $R_0 > 1$, and following

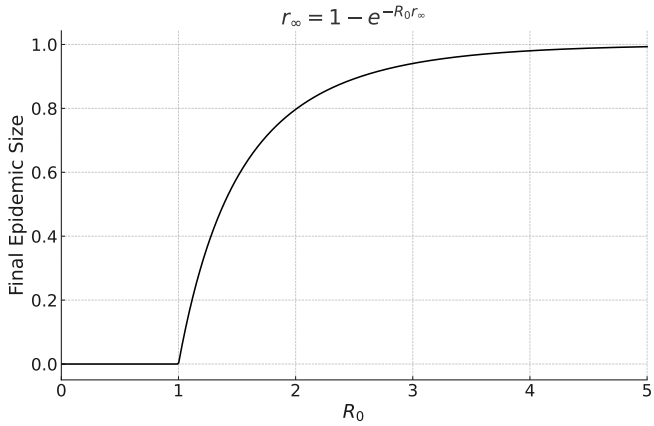

Figure 1.5. The epidemic can transition from a disease-free equilibrium state, with zero epidemic size, to an endemic equilibrium with a positive epidemic size. In the SIR model, the final epidemic size can be calculated using Eq. 1.6.

the same nomenclature, we call this regime super-critical, R_0 the control parameter and the epidemic size r_∞ the order parameter of the spreading phenomenon. In the thermodynamic limit, epidemic size vanishes in the sub-critical regime and scales with the population size in the super-critical regime.

In the SIR model, the final epidemic size can be given using Eq. 1.6 [151] such that

$$r_\infty = 1 - e^{-R_0 r_\infty}. \tag{1.10}$$

Fig. 1.5 shows the final epidemic size as a function of the basic reproduction number, and we witness a phase transition at $R_0 = 1$, commonly referred to as the epidemic threshold.

1.2.4 The SIS Model

Another extension to the SI model is to consider the possibility of not gaining immunity after infection and turning susceptible again. Making the transition between the compartments as $\mathtt{S} \to \mathtt{I} \to \mathtt{S}$. So, as opposed to the SIR model, we let *reinfection* happen in this two-state model. The governing equations in this model would be

$$\begin{aligned} \frac{\mathrm{d}s}{\mathrm{d}t} &= -\beta s i + \gamma i, \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \beta s i - \gamma i\,, \end{aligned} \tag{1.11}$$

with $s + i = 1$. This system of differential equations leads to:

$$i(t) = i_0 \frac{(\beta - \gamma) e^{(\beta-\gamma)t}}{\beta - \gamma + \beta i_0 e^{(\beta-\gamma)t}}\,, \tag{1.12}$$

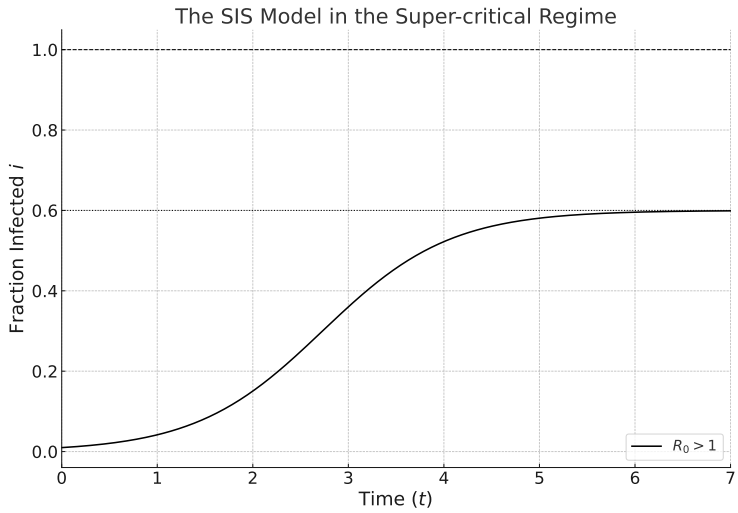

Figure 1.6. Evolution of fraction of infected people over time in the SIS model for $R_0 > 1$. i grows following a logistic curve but never saturates the whole population. Instead, it reaches a limiting value $1 - \frac{1}{R_0}$.

where i_0 is the size of the seed of infection. If $\beta < \gamma$, $i(t)$ would exponentially decay, and the disease will eventually die out. $\beta = \gamma$ will be the epidemic threshold, which can also be derived with the basic reproduction number arguments.

If $\beta > \gamma$, then this produces a logistic curve, similar to the SI model with a delicate feature that the epidemic ends up in a stable state where a steady fraction of the population, given by

$$i_\infty = \lim_{t\to\infty} i(t) = \frac{\beta - \gamma}{\beta} = 1 - \frac{1}{R_0}, \tag{1.13}$$

will be infected. Fig. 1.6 presents the time evolution of the fraction of infected people. Therefore, in the SIS model, the final size of the epidemic would always be $1/R_0$ away from saturating the whole population. This phenomenon is shown in Fig. 1.7 as it compares the dependence of the final size of the epidemic on R_0 in the SIS and SIR models, following Eq. 1.13 and Eq. 1.10, respectively.

1.2.5 The SEIR Model

In previous models, any exposed individual in the S compartment would most probably transition to the I compartment. As elaborated in Sec. 1.2.2, we often used "infected" and "infectious" interchangeably. This is because, in our model's context, an individual becomes contagious immediately upon exposure. So, there is no distinction between being infected and being infectious. Yet, when we examine diseases like COVID-19, we notice

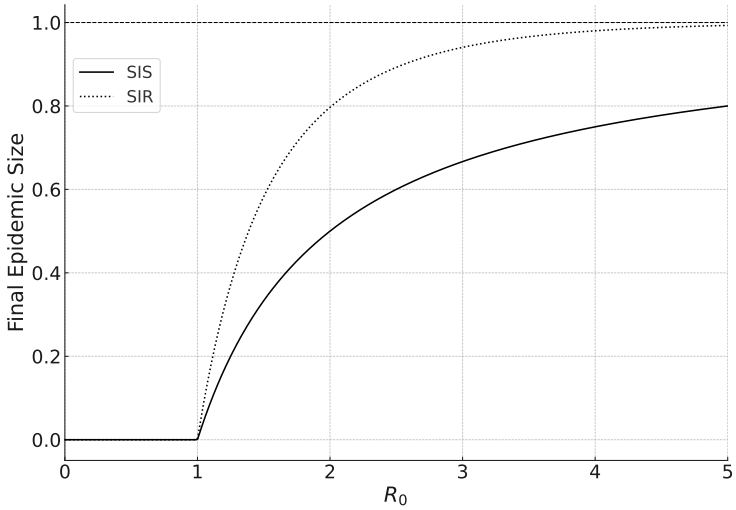

Figure 1.7. Comparison of the final epidemic size as a function of the basic reproduction number in the SIS and SIR model. In the sub-critical regime, where $R_0 < 1$, the final epidemic size is zero for both cases. In both dynamics, the entire population would never be infected; however, after the epidemic threshold, $R_0 = 1$, the SIR dynamics lead to a higher epidemic size.

that *exposed* individuals only begin to transmit the virus after a certain duration known as the incubation period, t_l, (sometimes referred to as the latent or latency period) [197]. For a more detailed understanding of infectious diseases, particularly when conceptualizing interventions, it would be instructive to contemplate a time profile similar to what is depicted in Fig. 1.8. To consider this latency, we can refine our model by introducing an intermediary `E` compartment between `S` and `I`. This modification results in the following transition sequence: $\mathtt{S} \to \mathtt{E} \to \mathtt{I} \to \mathtt{R}$.

Given that κ represents the average rate at which a latent individual transitions to being infectious, we can adapt the SIR model. Consequently, if the average fraction of individuals who have been exposed to the virus but are not yet infectious is e, the governing equations for the SEIR model can be expressed as:

$$\begin{aligned} \frac{\mathrm{d}s}{\mathrm{d}t} &= -\beta s i, \\ \frac{\mathrm{d}e}{\mathrm{d}t} &= \beta s i - \kappa e, \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \kappa e - \gamma i, \\ \frac{\mathrm{d}r}{\mathrm{d}t} &= \gamma i, \end{aligned} \tag{1.14}$$

with $s + e + i + r = 1$.

The SEIR model provides a more detailed insight into the spread of infec-

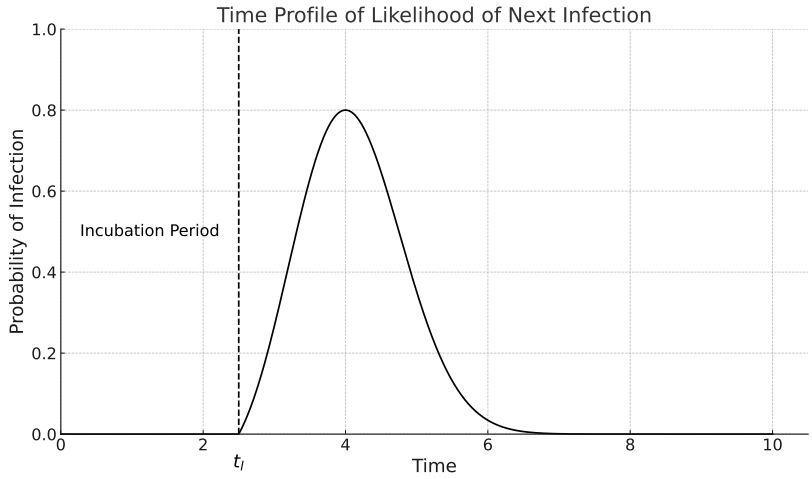

Figure 1.8. Infected individuals can start infecting their neighbors after some period known as the incubation period, t_l, denoting the interval from when an individual is first exposed to a pathogen to the onset of infectiousness or symptoms. For numerous infectious diseases, this time-frame captures the duration the pathogen needs to multiply sufficiently, eventually eliciting symptoms in the affected individual.

tious diseases, particularly when there is a clear latency period between exposure and the onset of infectiousness. However, within a reasonable range of parameter values, it does not differ fundamentally in terms of phenomenology compared to the SIR model.

1.2.6 The Next-generation Method and R_0

In structured models of epidemic spread encompassing multiple infection types, the computation of the basic reproduction number R_0 involves averaging the new infection rates across all types [56, 193, 40]. Consider a scenario like HIV transmission in a strictly heterosexual demographic, where infection types are distinctly categorized as male and female, with asymmetries in transmission rates. Here, M_{ij} represents the expected number of type i infections caused by an individual of type j in a wholly susceptible type i population. If f and m are the expected secondary infections in females and males, respectively, we can summarize the number of new infections after each type in a table $\mathbf{M}$, celebrated as *the next-generation matrix*, such that:

$$\mathbf{M} = \begin{pmatrix} 0 & f \\ m & 0 \end{pmatrix}. \tag{1.15}$$

As a non-negative matrix, $\mathbf{M}$ inherently possesses a principal eigenvalue, its spectral radius $\rho(\mathbf{M})$, which epitomizes the total averaged number of

new infections. Consequently, $R_0 = \rho(\mathbf{M}) = \sqrt{mf}$ which is also the geometric mean of the expected number of female and male secondary cases, encapsulating the epidemiological virulence of the pathogen in this structured population model. If secondary infections were gender-independent, with each individual capable of infecting R_0 others, then the anti-diagonal elements of the matrix $\mathbf{M}$, M_{12} and M_{21}, would be identical to R_0. As a result, the spectral radius would be $\rho(\mathbf{M}) = \sqrt{R_0^2} = R_0$.

1.2.7 The Next-generation Matrix for Compartmental Models

As we already have seen, compartmental models in epidemiology usually describe the dynamics of spreading using a set of coupled non-linear ordinary differential equations (ODE). This system of ODEs is constructed to have two fixed points: a disease-free equilibrium, an infection-free steady state where the final epidemic size is zero, and an endemic equilibrium, with a positive final epidemic size. The stability of these fixed points will be determined with the control parameter R_0, celebrated as the basic reproduction number in epidemic modeling. For a quick reminder, look at Fig. 1.7, which shows such a phase transition in the SIS and SIR model.

For a general model, with n compartments, $\mathtt{C}_1, \mathtt{C}_2, \cdots, \mathtt{C}_n$, the time evolution of the faction of the population in each compartment can be written as

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = f(c_1, c_2, \cdots, c_n), \tag{1.16}$$

where $f(\cdot)$ is a non-linear combination of other compartment sizes, respectively, and conventionally, $\mathtt{C}_1$ is set to be the susceptible compartment. R_0 is a threshold for the stability of the disease-free equilibrium such that its stability changes at $R_0 = 1$.

To determine R_0, we linearize the non-linear ODEs that describe the production of new infections and transitions among the infected individuals around the disease-free equilibrium. Epidemiological interpretation of this linearization is that R_0 encapsulates the potential outbreak initiated by an infected individual in a naive population, assuming the change in the susceptible population is negligible during the initial spread. Mathematically, R_0 can be articulated as the product of the near-disease-free equilibrium infection rate and the average infectious duration. It may be helpful to revisit the derivation of the SIR model presented in Eq. 1.9.

As matrices can represent linear ODE systems, we can average the expected number of new infections over all possible infected types with the next-generation matrix $\mathbf{M}$ [55]. This matrix is a linear positive operator that maps the current infection generation to the subsequent one. It's essential to highlight that this matrix operator primarily targets the number of infections, excluding other compartments. Iteratively applying this operator sheds light on the initial propagation of the infection across a

diverse population. Setting the spectral radius (Perron root), or the largest modulus of the eigenvalues of this operator to one, we can find the basic reproduction number [196, 57, 37].

$$R_0 = \rho(\mathbf{M}). \tag{1.17}$$

To clarify this framework, let's go through a rather complicated example. Assume a disease governed by SEIR dynamics similar to the one in Sec. 1.2.5 but with this consideration that new susceptible people arrive in that population with rate λ and any individual can leave it with rate μ. It can be a model for studying COVID-19 in a region without travel or mobility restrictions or in a closed population with a birth and death rate γ and μ, respectively. In this scenario, the disease dynamics would be updated as:

$$\begin{aligned}
\frac{\mathrm{d}s}{\mathrm{d}t} &= -\beta s i - \mu s + \lambda, \\
\frac{\mathrm{d}e}{\mathrm{d}t} &= \beta s i - (\kappa + \mu) e, \\
\frac{\mathrm{d}i}{\mathrm{d}t} &= \kappa e - (\gamma + \mu) i, \\
\frac{\mathrm{d}r}{\mathrm{d}t} &= \gamma i - \mu r.
\end{aligned} \tag{1.18}$$

We can represent the time evolution of this system with vector $\mathrm{x} = (s, e, i, r)$ where x_j represents the fraction of population in the j-th component corresponding to the SEIR order. Now, let $F_j(\mathrm{x})$ represent the rate of appearance of new infections in compartment j, considering only the infections that are newly emerging and excluding terms that depict the movement of infectious individuals between compartments [56, 193, 40]. If V_j^+ is the rate at which individuals move to compartment j through other means, and V_j^- is the rate of leaving the j-th compartment, then the difference $F_j(\mathrm{x}) - V_j(\mathrm{x})$ describes the rate of change, where $V_j(\mathrm{x}) = V_j^-(\mathrm{x}) - V_j^+(\mathrm{x})$.

From here, we can formulate matrices of partial derivatives of F and V such that

$$\begin{aligned}
F_{jk} &= \frac{\partial F_j(\mathrm{x}_{\mathrm{eq}})}{\partial \mathrm{x}_k}, \quad \text{and} \\
V_{jk} &= \frac{\partial V_j(\mathrm{x}_{\mathrm{eq}})}{\partial \mathrm{x}_k},
\end{aligned} \tag{1.19}$$

where $\mathrm{x}_{\mathrm{eq}} = (s_0, e_0, i_0, r_0) = (\lambda/\mu, 0, 0, 0)$ is the disease-free equilibrium. F is a non-negative matrix that represents the infection rates near the equilibrium, and V^{-1} represents the average duration of infectiousness. So, we can write the next-generation matrix as $\mathbf{M} = \mathbf{F}\mathbf{V}^{-1}$ [57, 56, 55, 193, 40]. Every element of the next-generation matrix, M_{jk}, gives the rate at which infected individuals in the compartment j produces new infections in

compartment k, times the average period an individual spends in a single visit to compartment j. For this example, we can have:

$$\mathbf{F} = \begin{pmatrix} 0 & \beta s_0 \\ 0 & 0 \end{pmatrix}, \quad \text{and} \\ \mathbf{V} = \begin{pmatrix} \mu+\kappa & 0 \\ -\kappa & \gamma+\mu \end{pmatrix}, \tag{1.20}$$

and therefore, the basic reproduction number is calculated as

$$R_0 = \rho(\mathbf{M}) = s_0 \frac{\beta\kappa}{(\mu+\kappa)(\mu+\gamma)}. \tag{1.21}$$

With the more challenging situations on the horizon, the next-generation matrix will come in handy. We plan to apply it to diverse situations later on.

1.2.8 Beyond the SIR Model

We can further extend the number of compartments to include more possibilities; for example, one may extend the SEIR model to account for the *waning immunity*. Waning immunity describes the gradual reduction of our body's protective response against disease over time, whether that protection was gained through natural infection or vaccination. As this immunity fades, there's a possibility of becoming susceptible to the disease again. So, the model can be extended to **SEIRS**. This phenomenon is a significant factor in the development and scheduling of vaccines. For certain diseases, periodic booster shots are administered to reinforce the body's immune response and ensure continued protection.

Compartmental models are the building blocks for epidemic modeling, allowing us to capture specific events of interest. However, this approach involves a balance between model complexity and its analytical tractability. With each additional compartment introduced, the model becomes subject to more mathematical constraints and wider implications. Consequently, the more compartments a model contains, the more challenging it becomes to rigorously follow and analyze its outcomes. This reflects the inherent complexity of translating the multifaceted nature of disease spread into a structured, mathematical framework. On the other hand, we eventually want to apply our model to the real world, and when dealing with data, we will face new challenges in inferring the model's parameters. Therefore, anyone who intends to work with more complex models should be concerned that without having enough justifiable data and reasons, it would be hard to shed light on any phenomenon. We are not the type of scientist who doesn't care about Occam's razor [181]; If we have two theories that both explain the observed facts, then we use the simplest until more evidence comes along.

SIR model shows how different factors influence the spread and eventual containment of an epidemic, making it a suitable choice for our analysis of the epidemic threshold and outbreak size, allowing for a comprehensive understanding of epidemic characteristics [207, 50, 2]. While the SEIR model and others offer more detail, especially regarding the latency period of an infection, they do not significantly alter the fundamental dynamics we aim to explore [150]. Therefore, SIR dynamics may sweep some epidemiological details under the rug, yet they can capture broader features of disease dynamics. These dynamics strike an ideal balance between realism and simplicity for epidemic problems we are interested in this work. Therefore, we will continue to use to this model and its different manifestations in the following chapters, leaving room for more complexity from the contact network structure. After all, the SIR model gets the phenomenology – macroscopic behaviors and trends – right, which is the most essential modeling point for a physicist to begin with.

> With four parameters I can fit an elephant,
> and with five I can make him wiggle his trunk.
>
> John von Neumann in conversation with
> Enrico Fermi [65]

1.3 Herd Immunity and Failures of Simple Models

In any population, individuals can acquire immunity to specific infections either through prior exposure to the pathogen, via prophylactic measures like vaccination, or a combination of both [122]. For instance, in Finland and other countries, an annual influenza vaccination initiative is launched at the beginning of autumn. This vaccination offers protection against influenza and shields from its secondary diseases, encompassing ear infections, bronchitis, pneumonia, myocardial infarction, and cerebral circulatory disorders [73]. The ultimate objective is to attain a population-level immunity; if a significant fraction of the population is immunized, it yields collective protection, often referred to as *herd immunity*. We should keep in mind that within each population, certain people either cannot receive the vaccine or opt not to for various reasons. Nevertheless, the principle of herd immunity ensures protection for the broader population, even when a disease has the potential to spread rapidly among the unvaccinated or those previously immune.

1.3.1 Herd Immunity after Vaccination

When vaccinating a population before the infection is introduced, we are trying to modify the basic reproduction number of the upcoming disease dynamics such that $R_0 < 1$ and, consequently, keeping the disease in its subcritical regime. But how much of a population should we vaccinate to reach herd immunity? Assume a population with no prior immunity that we can protect π_v fraction of them with perfect vaccines. What is the critical mass we need to vaccinate? For policy-makers, it is of vital significance to always estimate how many people they need to vaccinate as they have to assign enough budget for their vaccine campaigns. Moreover, they must not only consider the number of vaccines needed but also adapt their strategies. This includes prioritizing high-incidence regions and simultaneously vaccinating different age groups to effectively mitigate the pandemic's impact. Such dynamic strategies, grounded in scientific modeling, can significantly enhance the effectiveness of vaccine campaigns, especially in diverse and changing pandemic scenarios. Ref [142] evaluates the efficiency of various heuristic strategies for allocating COVID-19 vaccines, comparing them to strategies derived from optimal control theory.

In the paradigmatic SIR model of infectious disease in a fully mixed population, herd immunity is reached when the fraction π_v of the population that is immune to the disease through vaccination or previous infection is larger than

$$\pi_v^c = 1 - \frac{1}{R_0}. \tag{1.22}$$

The argument behind this is simple. According to the SIR model, every infected person will, on average, infect R_0 more people. If we vaccinate π_v fraction of people effectively, $1 - \pi_v$ fraction of people participate in the disease dynamics. Therefore, if $(1 - \pi_v)R_0 < 1$, then we reach herd immunity, and we can find the herd immunity threshold as given by Eq.1.22. Whenever we are referring to disease dynamics in the presence of interventions or not at the early stages of the epidemic, we will use the term *effective* reproduction number R_e instead of the basic reproduction number. In this case, $R_e = (1 - \pi_v)R_0$ and hence we need to keep R_e below one to prevent large outbreaks. Now, picture a scenario with $R_0 = 3$. Then, the herd immunity threshold would be roughly calculated as $1 - 1/3 \simeq 70\%$.

However, such a directive, though widespread, may oversimplify the complexities at play. First and foremost, Eq. 1.22 draws the epidemic threshold from the fully-mixed SIR model, which operates on the assumption that all individuals in a population mix uniformly, with interactions being random and uninfluenced by specific attributes, such as their vaccination status. When it comes to the real world, this generalization can be limiting as human–human interactions are far from being random. Moreover, in practice, vaccines never get distributed equally in a population. Think

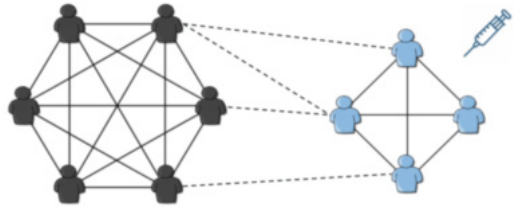

Figure 1.9. Vaccination patterns within a population reveal non-random connections between individuals. Social dynamics often lead to individuals interacting more with those who with the same vaccination status rather than vaccines being distributed uniformly at random across the population. Here, solid lines depict connections within the community, while dashed lines represent links between different communities.

of the correlation between an individual's vaccination status and that of the people they frequently interact with. For example, vaccinated people are more likely to interact with other vaccinated people in the population [31]. This propensity for like to associate with like is what we term as *homophily*. Fig. 1.9 demonstrates a situation with *vaccination homophily*. We use the term vaccination homophily to describe such patterns where interactions are more common within similar vaccination groups than between them. This phenomenon, rooted in the tendency of people with similar socio-demographic and behavioral characteristics to interact, leads to non-uniform vaccine adoption, as demonstrated in [38]. The interaction patterns between vaccinated and non-vaccinated individuals, shaped by homophily, crucially affect disease transmission dynamics. A study in British Columbia, Canada, involving 1 304 respondents, observed clear patterns of vaccine homophily, revealing its substantial role in epidemic growth and infection rates among different vaccination groups under varying conditions of vaccine efficacy [7].

We will later see that in some settings, herd immunity can be easily not achieved. As we show in Publication I, the presence of homophily considerably increases the critical vaccine coverage needed for herd immunity, and strong homophily can push the threshold entirely out of reach. Another artifact of vaccination homophily in our model is that the epidemic size monotonically increases as a function of homophily strength for a perfect vaccine. We will discuss the consequence of vaccination homophily in Sec. 3.7.

1.3.2 Herd Immunity through Natural Infection

As mentioned earlier, herd immunity can be gained through natural infection. For some diseases, recovered people stay immune to future infections. It is worth noting that herd immunity after natural immunity due to previous exposure varies in strength and nature compared to vaccine immunity.

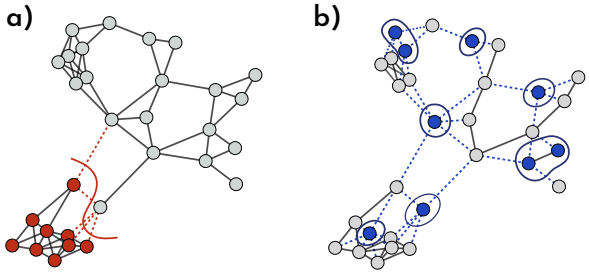

Figure 1.10. How natural immunity is localized in a population. (a) Nodes immunized due to previous exposure(s) shown with color red. The dashed lines and the red solid line indicate the interface between susceptible and immune nodes. (b) The same number of randomly immunized nodes in the same population, this time in blue, resulted in more interfaces.

Empirical observations in network science reveal that the number of connections per individual, known as *degrees*, varies considerably across a population [46, 36]. Such disparities in connectivity patterns highlight a departure from a normal distribution, underscoring the *heterogeneous* nature of real-world networks. When immunity is induced by natural infection, population heterogeneity may lead to a lower herd immunity threshold than expected under homogeneous mixing because the disease spreads among highly interactive individuals, known as *superspreaders*, at the early stage of the epidemic, resulting in more efficient immunization of these influential *hubs* [35]. On the other hand, as we show in Publication II, in addition to degree heterogeneity, the spatial and structural aspects of the population play a significant role here [22]. While an epidemic preferentially infects and removes people with more connections, strengthening the herd immunity effect, it is contiguous and localized in the population, weakening herd immunity. For a visual explanation, see Fig. 1.10 The implications of such localization on herd immunity are yet to be comprehensively explored.

Examining herd immunity within real-world scenarios, characterized by structural nuances like communities, core-periphery constructs, and household dynamics, can offer a deeper understanding of epidemic behavior in actual settings. Such endeavors can bridge the divide between theoretical frameworks and their real-world applications, aiding in formulating effective disease control and prevention strategies. We will delve into these complications in this work. We will see how, and in what ways, various interventions can be effective considering the networked structure of human populations. Meanwhile, you might explore Ref. [202] to get acquainted with the progress in epidemiological modeling, particularly how it incorporates vaccination, individual behaviors, and social structures

into understanding how diseases spread. This reference underscores the transition from simple to more sophisticated models that leverage statistical physics and digital information, shedding light on the complexities of disease transmission and the effects of vaccination. The research wraps up with suggestions for future investigations, making it crucial reading for those focused on disease modeling and public health initiatives.

> After many efforts, when names, definitions, observations and other sensory data are brought into contact and compared in depth, one juxtaposed with another, in the course of a scrutiny and an even-tempered but severe examination, at the end a light suddenly comes on, for whatever problem – our understanding, and a clarity of intelligence the effects of which express the limits of human power.
>
> Plato, *Letter VIII* [172]

2. Structured Populations and Networks

Traditional models in epidemiology often do not incorporate the complexity of social connections. As we saw, in the fully mixed populations of Chapter 1, there were no constraints on the interactions between the individuals beyond the characteristics of the disease's spread. In the compartmental models introduced in Sec. 1.2, an infected individual could infect any other susceptible individual which is not a realistic assumption for any physicist interested in modeling the propagation of a property (virus) in a physical system (society).

To better understand disease dynamics in a society, we need to develop realistic models for describing the population in which the epidemic is unfolding. First, we want to respect the *principle of locality* in the sense that people can only interact with their neighbors in some physical vicinity. For airborne diseases, the primary infectious agents are often viruses, bacteria, or fungi. These pathogens can be transmitted via activities such as breathing, talking, coughing, sneezing, and other actions that generate aerosol particles or droplets [4, 9, 127, 200, 85, 67]. Therefore, an airborne transmission can only happen through sharing some medium. Moreover, within a population of N people, the total number of connections between the individuals, L, is usually around the same number as the population size, N. In technical terms, we expect the population to be *sparse* such that the L scales with N, not N^2, for example, $L = \mathcal{O}(N)$ [132]. From this point on, space graphs will be our primary focus for network analysis, except where explicitly stated otherwise. Another essential feature we expect is that the number of connections, k, varies from person to person, such that the majority of people have almost the same number of connections, except for a few people with very high k, known as *hubs*, or in the context of spreading, *superspreaders* as they can infect much more people than a typical individual.

2.1 Random Graph Theory: Key Concepts and Definitions

With these minimum requirements, we can approximate a human population with a random graph G such that every node v in the graph represents an individual, and a link l between two people represents a connection or a pair-wise interaction between them [194]. An *undirected* graph is a type of graph where links lack direction, indicating bidirectional relationships. Within this framework, a *simple* graph is a specific form of an undirected graph characterized by two key constraints: it contains no loops (links connecting a node to itself) and no parallel links (multiple links between the same pair of vertices). These properties of a simple graph, with its non-repeating and non-self-looping links, make it an ideal model for scenarios like social networks, where edges represent unique and mutual connections between individuals.

The number of connections a node v has is called the *degree* of that node, k_v. The degree distribution p_k and other structural aspects of a network are typically referred to as network *topology*. For a graph with N nodes, and L links, the identify below holds:

$$\sum_{v=1}^{N} k_v = 2L = N\langle k\rangle, \tag{2.1}$$

where $\langle k\rangle$ is average degree of the nodes,

$$\langle k\rangle = \sum_{k} k p_k . \tag{2.2}$$

Over the past two decades, numerous random graph models have emerged to represent human populations, primarily within the realm of *Network Science* [151, 64]. These models are largely inspired by graph theory literature [58]. Consequently, we might use the terms "networks" and "graphs" interchangeably moving forward.

Fig. 1.9 displays a network with two distinct groups of nodes, representing vaccinated and unvaccinated *communities*, along with their interconnections. This network is termed a *connected* graph since there exists a *path* between any two nodes. Within the graph, any sequence where each successive pair of nodes is connected by a link is referred to as a *walk*. When this walk avoids revisiting nodes, it's termed a *path*. The length of a walk is determined by counting the number of links in its sequence. Furthermore, the *distance* between two nodes is defined as the length of the shortest path connecting them. The longest shortest path in a network is usually referred to as the *diameter* of a graph. The diameter represents the largest distance between any two connected nodes.

Figure 2.1 shows a *disconnected* network, as you can find a pair of nodes such that there is no path between them. Whenever there is no path between two nodes, we say that they are *infinitely* far from each other. The

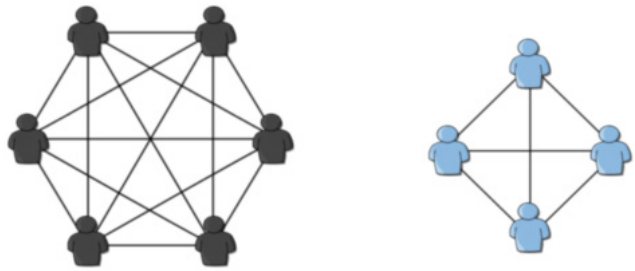

Figure 2.1. A disconnected network comprising two components: the larger left subset represents the giant component (GC) formed as a 6-clique, while the smaller right subset illustrates a 4-clique.

communities of the network in Fig. 2.1 are completely disconnected. A *subgraph* is a graph formed from a subset of the nodes and links of the original graph. A connected subgraph that is not part of any larger connected subgraph is called a *component*, and its size will be reported as the number of nodes it includes. Components can be used to partition a graph. We name the largest connected component of a graph as the *giant* component, [151, 29]. Typically, real-world networks feature a dominant giant component covering a significant portion of the entire network, wherein the majority of nodes are part of this component. A giant component's significance is underscored when its size is proportional to the overall network size. This scaling feature is integral to network infrastructure functionality. For instance, the internet's operational efficiency hinges on this network characteristic, and the rapid spread of infectious diseases can be attributed to the infection of the giant component in social networks.

In Fig. 2.1, we can see that each component has a very regular structure as every node is connected to any other node. We call such a structure a *complete* graph. Any subset of a graph that can be considered as an *induced* subgraph forming a complete graph with c nodes is called a c-clique. *Cliques* are very good candidates for representing groups in social networks. We will use them in our modeling in Publication III.

In a complete graph, every pair of neighbors of a node u is connected, making the graph maximally clustered. Generally, the amount of clustering, or *transitivity*, is typically measured by counting the number of closed loops. If a path uvw forms a loop of size three, we say the path is *closed* and nodes u, v, and w form a *closed triad* or triangle, as opposed to a *triple* that misses one of the edges of the triangle [203]. To address the level of transitivity of a network, we can define the *clustering coefficient* C of a network [151] as the ratio of the number of triangles Δ to the number of connected triples Λ

$$C = \frac{3\Delta}{\Lambda}, \tag{2.3}$$

where factor 3 counts for the different ways we can count paths of length two and three. This is a global property of a network, which can also be calculated by averaging the local clustering coefficient of all the nodes. We can define [204] a local clustering coefficient for a node u, C_u, as the number of connected pairs of neighbors of u over the number of pairs of neighbors of u. C_u represents the average probability that a pair of u's friends are friends of one another. Following Watts and Strogatz [204], we can also define the clustering coefficient for the network as the average of the local clustering coefficients for each node, $C_{\mathbf{WS}} = \langle C_u \rangle$.

We can represent the connections in a network through the *adjacency matrix* $\mathbf{A}$ such that $A_{uv} = 1$ when node u is connected to node v and $A_{uv} = 0$ otherwise. So, the matrix representation of the graph in Fig. 2.1 would be:

$$\mathbf{A} = \left(\begin{array}{cccccc|cccc} 0 & 1 & 1 & 1 & 1 & 1 & & & & \\ 1 & 0 & 1 & 1 & 1 & 1 & & & & \\ 1 & 1 & 0 & 1 & 1 & 1 & & \mathbf{0} & & \\ 1 & 1 & 1 & 0 & 1 & 1 & & & & \\ 1 & 1 & 1 & 1 & 0 & 1 & & & & \\ 1 & 1 & 1 & 1 & 1 & 0 & & & & \\ \hline & & & & & & 0 & 1 & 1 & 1 \\ & & & & & & 1 & 0 & 1 & 1 \\ & & \mathbf{0} & & & & 1 & 1 & 0 & 1 \\ & & & & & & 1 & 1 & 1 & 0 \end{array}\right) \tag{2.4}$$

where we have rearranged the rows and columns to show the block nature of the adjacency matrix for a network that has two components. $\mathbf{A}$ is a symmetric matrix, $A_{uv} = A_{vu}$ as we are not imposing any direction on the links in the network. The diagonal elements are also zero, $A_{uu} = 0$, as the network has no self-loop and it is a *simple* graph. A simple graph is a graph with no self-loops, and it does not have more than one link between any two nodes. The degree of node u can be calculated by summing over the specific rows or columns of matrix $\mathbf{A}$ such that $k_u = \sum_v A_{uv}$. The matrix representation of the graph serves as a gateway to the intriguing world of algebra with its fascinating machinery.

With these terms defined and agreed upon, we can now explore random graph models and what they offer.

2.2 Erdős–Rényi Networks

We begin with the simplest model that people usually refer to as *the* random graph, known as the *Erdős–Rényi* (ER) model [151], and will go through more detailed models later. Imagine we want to build a network

G with N nodes such that L distinct pairs of them are uniformly chosen at random and get connected. We will refer to this model as $G(N,L)$ [68]. We can do better and define the model as an ensemble of networks [151], such that the model $G(N,L)$ is the probability distribution $P(G)$ over all simple networks G with

$$P(G) = \frac{1}{\binom{\binom{N}{2}}{L}}, \tag{2.5}$$

as we are choosing L links from the $\binom{N}{2}$ possibilities. When we analyze property f of random graphs (like the number of connections each node has or how many steps it takes to get from one node to another), we are actually looking at the average of these properties across many different graphs generated by the same random model [151]. Since each graph produced by this process can be different, we focus on the average value in the ensemble of these graphs rather than on the specifics of any single graph. This approach helps us understand the general behavior of random graphs as a whole. Therefore, any property of random graphs, $f(G)$, is the average property of the ensemble [151]. The term *ensemble* comes from the community of physics. So, the ensemble average of a property f would be written as:

$$\langle f \rangle = \sum_G f(G)P(G). \tag{2.6}$$

For mathematical convenience in calculating the average values using Eq. 2.6, it would be better to use a slightly different mathematical definition than the one with a fixed number of links L given by Eq. 2.5 [188]. Instead, we connect every uniformly selected pair of nodes at random with independent probability p [151]. This will enable us to write Eq. 2.5 as

$$P(G) = p^L(1-p)^{\binom{N}{2}-L}. \tag{2.7}$$

We name this model $G(N,p)$. It is straightforward now to show [151] that the expected number of links in such an ensemble is:

$$\langle L \rangle = p\binom{N}{2} \simeq pN^2. \tag{2.8}$$

Moreover, the degree distribution, p_k, of a random graph made with $G(N,p)$ model follows a binomial distribution [151] as:

$$p_k = \binom{N-1}{2} p^k(1-p_k)^{N-1-k}. \tag{2.9}$$

For large enough network sizes, Eq. 2.9 can be approximated with a Poisson distribution. Therefore, in the thermodynamic limit, an ER network possesses a Poisson degree distribution with mean $\langle k \rangle$, such that

$$p_k = e^{-\langle k \rangle}\frac{\langle k \rangle^k}{k!}. \tag{2.10}$$

Due to the Poisson degree distribution, some scholars call this model *Poisson random graph* [151].

We can use this equation and find the average degree $\langle k \rangle$ in a network generated with this model as:

$$\langle k \rangle = \left\langle \frac{2L}{N} \right\rangle = \frac{2}{N} \langle L \rangle = p(N-1) \simeq pN, \tag{2.11}$$

as the mean degree in a network with L links is $2L/N$. The factor 2 is responsible for the contribution of the link in the degree of two nodes.

2.2.1 Emergence of the Giant Component

An interesting property of ER networks is how their giant component size changes as we vary p [29]. We will later in Sec. 2.8, Publication III and IV discuss how the size of the giant component will be related to evaluating the final outbreak size in an epidemic spreading on a network. Now, assume the probability that a node is not in the giant component is given by ϕ. This probability also represents the relative size of the giant component as $\Sigma = 1 - \phi$. A node u belongs to the giant component if all of its neighbors are connected to the giant component. Therefore, the probability that node u is not connected to the giant component via a neighbor v is $1 - p + p\phi$ [151]. Simply, $1 - p$ accounts for the probability of not connecting to node v, and $p\phi$ is the probability of the link between u and v existing but not leading to the giant component. Therefore, the total probability of not being connected to the giant component via any of the $N - 1$ other nodes in a large network can be written [151] as a self-consistent equation below

$$\phi = (1 - p + p\phi)^{N-1}. \tag{2.12}$$

We can rearrange the terms in this equation, use an approximation in the thermodynamic limit, and find that

$$\phi = \exp\big(-\langle k \rangle (1 - \phi)\big). \tag{2.13}$$

Eliminating ϕ in favor of the size of the giant component gives us

$$\Sigma = 1 - e^{-\langle k \rangle \Sigma}. \tag{2.14}$$

There is no closed-form solution for this equation, but we know that $\Sigma = 0$ is a stable solution when $\langle k \rangle < 1$. In cases where the *branching factor*, $\langle k \rangle$, is less than one, there is an exponential decay in this branching process [10], leading to its swift termination. When $\langle k \rangle > 1$, every node is, on average, connected to more than one other node, leading to a chain of connections. Therefore, we expect the size of the largest component to undergo a phase transition from constant size to extensive size at $\langle k \rangle = 1$. We can also show that in the thermodynamic limit, only one component spans the entire network in the super-critical regime [151].

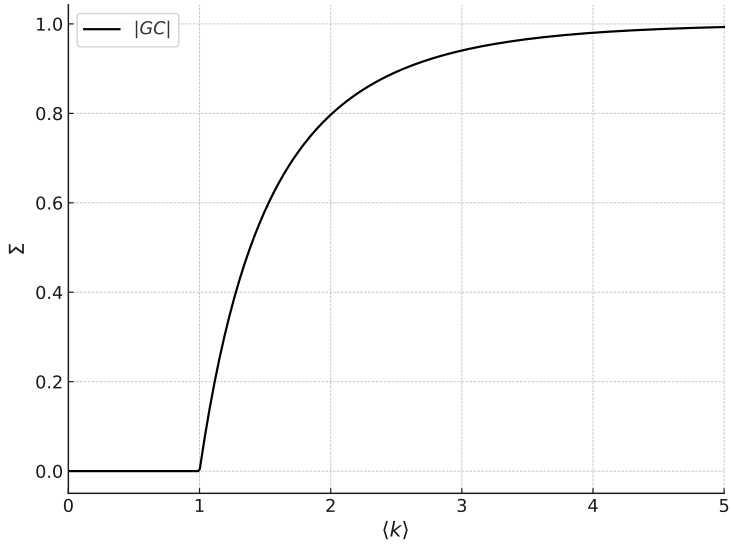

Figure 2.2. Phase transition in an ER network from disconnected to connected regime with an extensive giant component.

Figure 2.2 captures such a phase transition. Following Eq. 2.11, this transition happens at the critical writing probability $p_c = 1/N$. Interestingly, Eq. 2.14 has the same form as Eq. 1.6, and it manifests the same type of phrase transitions as depicted in Fig. 1.7.

We will use similar but more detailed and complicated arguments in all our publications to find the size of the giant component of a random graph under different conditions. Alternatively, we can show that the average size of the small components, any component other than the giant, vanishes in the super-critical regime as the average degree or p increases. The average size of a small component to which a randomly chosen node in an ER network belongs can be derived [151] as:

$$\sigma = \frac{1}{1 - \langle k \rangle + \langle k \rangle \Sigma}. \tag{2.15}$$

When $\langle k \rangle < 1$ and there is no giant component, this equation leads to $(1 - \langle k \rangle)^{-1}$ which diverges at the critical point. In the super-critical regime where $\langle k \rangle > 1$, we must solve for Σ and then evaluate σ. Fig. 2.3 depicts the behavior of the average size of small components in an ER network.

From the point of view of Critical Phenomena [183], we can think about σ as the *singular* property at the phase transition point. σ's behavior is analogous to *susceptibility* χ in the field of critical phenomena, which measures the response magnitude generated by a small external field disturbance [63]. In our case, a small external field disturbance can be adding a link uniformly at random in the network and measuring how the average size of its giant component changes. Even adding a single link

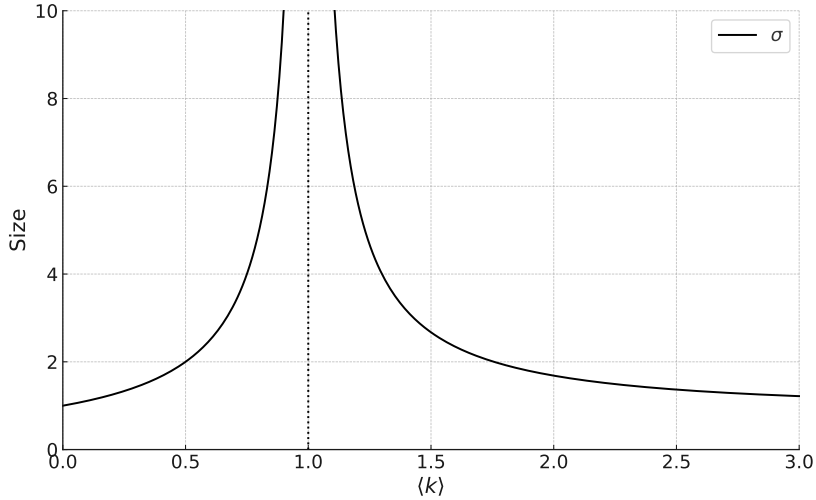

Figure 2.3. Average size of the small components in an ER network. The average size σ of the component to which a randomly chosen node in a small component belongs using Eq. 2.15.

can dramatically change the component average sizes at the critical point. Using the ER model, if we generate an ensemble of random graphs and look at the ensemble average of the giant component sizes, it will follow Eq. 2.14, and the fluctuations around this mean value would diverge at the critical point. Therefore, if we run a set of simulations and measure the coefficient of variation of the giant component sizes, the ratio of the standard deviation of giant component sizes to their ensemble average would behave the same as σ around the critical point.

2.2.2 Shortcomings of ER Networks in Modeling Contact Networks

We will examine in this section the limitations of the ER network in accurately modeling contact networks.

An ER network $G(N,p)$ comes with a single tuning parameter p. So, any property of a random graph of this ensemble depends on the *wiring probability* p. For example, the clustering coefficient of an ER network is

$$C = \frac{\langle k \rangle}{N-1} = p\,, \tag{2.16}$$

as the probability that any two nodes are neighbors is exactly the same across the network [151]. We can traverse the ER networks without spending much time going from one node to another. In technical words, the diameter of ER networks grows logarithmically with the network's size

[28, 151], such that

$$D = \frac{\ln N}{\ln \langle k \rangle}. \tag{2.17}$$

An intuitive way to understand the logarithmic growth of the diameter is that since the clustering of an ER network is given by p, the network can be approximated as a *tree*, a connected acyclic undirected graph that branches with $\langle k \rangle$. Therefore, in D steps, it should statistically cover all the N nodes such that $N = 1 + \langle k \rangle + \langle k \rangle^2 + \cdots + \langle k \rangle^D$. In the thermodynamic limit and when $\langle k \rangle > 1$, the dominant term in this series will be the last one, so $N \simeq \langle k \rangle^D$ which leads to Eq. 2.17. This result always holds even though we used a very crude approximation here. This *tree-like assumption* is a useful trick that allows us to ignore loops when we are calculating some network property locally [139].

ER networks come with a Poisson degree distribution, short (logarithmic) diameter, and low clustering. It is an improvement upon the fully-mixed population that is far from reality. By controlling p, we can always have a sparse network with a giant component, which is a necessary condition for the functionality of many networked systems. However, the implicit low clustering of this model makes it unrealistic to model social networks that exhibit transitivity, especially in the form of *triadic closure*, a tendency for two individuals with a mutual acquaintance to become connected themselves.

Sparse network candidates with high clustering can be the d-dimensional lattices that Condensed Matter physicists have used for decades to describe crystal structures. However, these networks usually possess so many symmetries or regularities that make them over-structured for us. Moreover, the diameter of a d-dimensional lattice scales algebraically with the space dimension as $N^{1/d}$, while we prefer shortcuts and logarithmically scaled diameter networks. Despite this, if we shuffle enough links in a lattice, we can add some randomness to make some long-range connections, making the diameter shorter while preserving the local structure to a reasonable extent. A random network with high clustering in which its diameter scales as $\ln N$ is called a *small-world* network. The celebrated Watts–Strogatz model features these rewirings and produces a small-world random graph [204].

The friendship paradox, first introduced by Feld (1991) [72], highlights a common reality in social networks where your friends are likely to have more friends than you. This concept is not only limited to the number of friends; it also applies to many other characteristics where your friends, on average, rank higher than you [66]. For example, your friends have more enemies than you do on average. If we uniformly at random sample nodes and count the links connected to them, we get an estimate over the average degree of a network. Suppose we sample links uniformly at random instead, follow them to one of their ends and look at the degree of

that node excluding the link we just came from. In that case, we will get an estimate over a network's average *excess degree* $\langle q \rangle$ [151]. In mathematical terms, the friendship paradox states that social networks have a larger average excess degree than an average degree.

The Erdős–Rényi (ER) networks with Poisson degree distributions fall short in regards to the "friendship paradox" primarily because of the uniformity of their connectivity. This uniformity means that ER networks lack the necessary variance in the number of connections to model the friendship paradox effectively, which is more accurately represented in networks with skewed degree distributions, such as scale-free networks [98].

As we will see in the next section, given the degree distribution p_k of a network that locally looks like a tree, the average excess degree can be derived from Eq. 2.23 in terms of the first and the second moments of the degree distribution such that

$$\langle q \rangle = \frac{\langle k^2 \rangle}{\langle k \rangle} - 1, \tag{2.18}$$

where $\langle k^m \rangle = \sum_k k^m p_k$ is the m-th moment of the degree distribution.

ER networks have a Poisson degree distribution, and consequently, the variance and mean of degrees are equal. So, $\langle k^2 \rangle = \langle k \rangle + \langle k \rangle^2$. Substituting this into Eq. 2.18 yields

$$\langle q \rangle = \langle k \rangle . \tag{2.19}$$

Therefore, in an ER network, your friends have the same number of friends as you. Sometimes ER networks are referred to as *homogeneous* random networks as degrees are highly centered around the mean value, and there is no chance of strong degree heterogeneity.

We are interested in models with large but finite variances in their degrees to better approximate the social networks. In practice, the degree distribution of networks should allow for the possibility of having nodes with high degrees. In ER models, the tail of the distribution falls dramatically and leaves no chance for the emergence of hubs. The next section shows that the configuration model network allows us to create a network with whatever distribution we want.

2.3 The Configuration Model

In this section, we explore the configuration model, an important tool in network science for constructing random networks. Unlike the ER model, the configuration model allows for more realistic representations of social networks by using a prescribed degree sequence. Each network created through this model is uniquely tailored based on its specific degree sequence, which dictates the number of connections each node should have.

This approach allows for flexible network structures, including uneven distributions of connections, providing a more accurate reflection of social network patterns than the homogenous ER model.

For a valid degree sequence represented by $k_1, k_2, \ldots, k_N$, when each of the N nodes has k_u stubs, a configuration model network can be constructed by connecting pairs of stubs uniformly at random [75]. In a configuration model network, the total probability of a connection between a pair of nodes u and v will be

$$p_{uv} = \frac{k_u k_v}{2L}, \tag{2.20}$$

where $2L = \sum_u k_u$ [151]. In the configuration model, while self-loops or multi-links might occur, their number becomes negligible in the thermodynamic limit where $N \to \infty$, especially when dealing with a degree sequence that ensures a finite average degree. If the variance of the degree sequence is finite, the infinitely large network that emerges from this model tends to be a simple graph without self-loops or multiple links between the same nodes. We can show the expected density of self-loops that may pop up in the network vanishes as $N \to \infty$ since

$$\frac{1}{N}\sum_u p_{uu} = \frac{1}{N}\sum_u \frac{k_u(k_u-1)}{4L} = \frac{1}{N}\frac{\langle k^2\rangle - \langle k\rangle}{2\langle k\rangle}. \tag{2.21}$$

The adjustment for p_{uu} is because the probability of a self-loop from node u to itself is $k_u(k_u-1)/4L$.

The expected number of common neighbors of node u and v, as a noteworthy characteristic, can be given as

$$n_{uv} = p_{uv}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle} = p_{uv}\langle q\rangle. \tag{2.22}$$

The last identity follows Eq. 2.18 for the average excess degree. Note that the excess degree distribution plays an important role in many calculations and can be derived as

$$Q_k = \frac{1}{\langle k\rangle}(k+1)p_k, \tag{2.23}$$

where $\langle k\rangle$ is averaged over the degree distribution p_k [151].

We can now calculate the clustering coefficient [151] in a configuration network model as

$$C = \sum_{k_u,k_v=0}^{\infty} Q_{k_u} Q_{k_v} p_{uv} = \frac{1}{N}\frac{\left[\langle k^2\rangle - \langle k\rangle\right]^2}{\langle k\rangle^3}. \tag{2.24}$$

Configuration model networks, the same as ER networks, have vanishingly small clustering in the thermodynamic limit, making them locally tree-like networks.

There are different versions of the configuration model with fewer constraints [194]. In various contexts, the term *soft* configuration model

is used to denote variants of the configuration model with relaxed constraints [157, 195]. For example, the Chung-Lu model [44, 45] is a soft configuration model which, instead of building a graph that satisfies a specific degree sequence, generates a random graph with given expected degrees $\bar{k}_1, \bar{k}_2, \ldots, \bar{k}_N$. The Chung-Lu model is prized for its ability to replicate complex network structures found in the real world, thanks to its accommodating of arbitrary degree distributions. It turns out that the link probability between node u and v in this variant will be the same as Eq. 2.20 with the expected values replaced. From now on, whenever we say a configuration model, we mean a soft version, specifically the Chung and Lu variant [44].

If we sample the expected degrees from a Poisson distribution, we will recover an ER network [61]. For more realistic modeling, specifically in modeling the contact network for disease spreading, we can use a negative binomial distribution as we can control its variance and produce a heavier tail [154]. This flexibility allows for generating random networks with more degree heterogeneity.

2.3.1 Existence of the Giant Component

The configuration model has a giant component if and only if $\langle q \rangle > 1$ [151] or in terms of moments of the degree distribution, if and only if

$$\left\langle k^2 \right\rangle - \langle k \rangle > 0. \tag{2.25}$$

This is known as the Molloy and Reed condition [143, 103]. When there is a giant component, starting from a typical node in a configuration model network, we can expect to have N_d d-order neighbors [151], given as

$$N_d = \langle q \rangle^{d-1} \langle k \rangle . \tag{2.26}$$

2.3.2 Size of Giant Component and Generating Functions

We can find the size of the giant component, if any, similar to how we argued for the ER model in Sec. 2.2.1 [151]. We begin from some node and follow one of its neighbors. Assume the probability that the neighbor is not in the giant component is ϕ. A node does not belong to the giant component when its k neighbors are not, which happens with probability ϕ^k. So, the size of the giant component would be given by

$$\Sigma = 1 - \left\langle \phi^k \right\rangle, \tag{2.27}$$

where we have averaged ϕ^k over the entire network with degree distribution p_k. It is common to name $\left\langle \phi^k \right\rangle$ as $g_0(\phi)$ where the function g_0 is called the *probability generating function* for the probability distribution

p_k [151];

$$g_0(\phi) = \sum_k p_k \phi^k = \langle \phi^k \rangle . \tag{2.28}$$

The probability ϕ was introduced for a neighbor node rather than a randomly chosen node. Therefore, its degree follows the excess degree distribution. We can write a self-consistent equation for ϕ as

$$\phi = \langle \phi^k \rangle_Q , \tag{2.29}$$

such that

$$g_1(\phi) = \sum_k Q_k \phi^k = \langle \phi^k \rangle_Q , \tag{2.30}$$

where we have averaged ϕ^k over the excess degree distribution Q_k and g_1 is the probability generating function of Q. Given the degree distribution p_k of a network, we can calculate g_0 and consequently, using Eq. 2.23, g_1 can be written as

$$g_1(\phi) = \frac{1}{\langle k \rangle} g_0'(\phi), \tag{2.31}$$

where g_0' is the first derivative of g_0 with respect to its argument [151]. When evaluated at 1, g_0' would give the average degree

$$g_0'(1) = \langle k \rangle , \tag{2.32}$$

leading to

$$g_1(\phi) = \frac{g_0'(\phi)}{g_0'(1)} . \tag{2.33}$$

When there is no giant component, the average size of the component a node belongs to would be specified entirely by the first and second moments of the degree distribution. With some algebra, we can show that

$$\sigma = 1 + \frac{g_0'(1)}{1 - g_1'(1)} = 1 + \frac{\langle k \rangle^2}{2\langle k \rangle + \langle k^2 \rangle} \tag{2.34}$$

where we have used Eq. 2.18 for the evaluating $g_1'(1)$. The good news is that the expected small component size, unlike the giant component size, can be evaluated without calculating any generating functions.

The diameter of a configuration model network would also scale logarithmically with the size of the network [44] such that

$$D = \frac{\ln N}{\ln \langle q \rangle} + \text{constant}. \tag{2.35}$$

2.4 Random Graph Models For Networks with Group Structure

Random graph models for networks with groups play a pivotal role in modeling epidemic diseases. These models effectively capture the complexity of

social interactions and the dynamics of disease spread within and between different groups in a population. By incorporating group structures, these models offer a more nuanced understanding of how diseases propagate through networks, accounting for variables like contact patterns, group sizes, and connectivity. This enhanced realism is crucial for predicting outbreaks, assessing the impact of interventions, and informing public health policies. In the following sections, we will delve into specific models that provide us with detailed group structures, further elucidating their significance in epidemiological modeling and disease control strategies.

2.4.1 Bipartite Network

Most social networks contain groups of densely connected subgraphs. One way to model them is to use a bipartite network model, also known as a two-mode or bi-modal network, which represents a type of graph where nodes are divided into two distinct sets, and links only exist between nodes of different sets. In other words, there are no links between nodes within the same set. This structure is useful in modeling systems where relationships occur only between distinct categories. Common examples of bipartite networks include author-paper relationships in academic databases (where authors are in one set and papers in another, and links indicate authorship) or customer-product purchase histories in market research (where customers are in one set, products in another, and links indicate a purchase). The bipartite nature of these networks offers unique analytical challenges and opportunities, distinct from traditional one-mode networks. We will utilize a refined version of the bipartite network model as outlined in Publication III. This model distinctively categorizes one set of nodes as individuals and another set as the groups to which these individuals are affiliated. By projecting this network model, we aim to intricately map the complex interactions between individuals and their respective groups.

2.4.2 Random Clique Networks

In studying epidemic processes, cliques are commonly used to represent social groups within contact networks. Within social networks, c-cliques are complete subgraphs indicating a group of c people who are all interconnected and can potentially infect one another. These structures have been identified in real-world social networks, challenging the standard tree-like assumption often applied in epidemic studies [192, 166, 209]. These networks can be constructed by dividing a graph into two sections: node groups and clique groups. Since it's a bipartite structure, connections only form between these groups. In the thermodynamic limit, the unipartite projection of this graph forms a contact network with a negligible number of self-loops or multi-links. See [205, 29] for further details on these net-

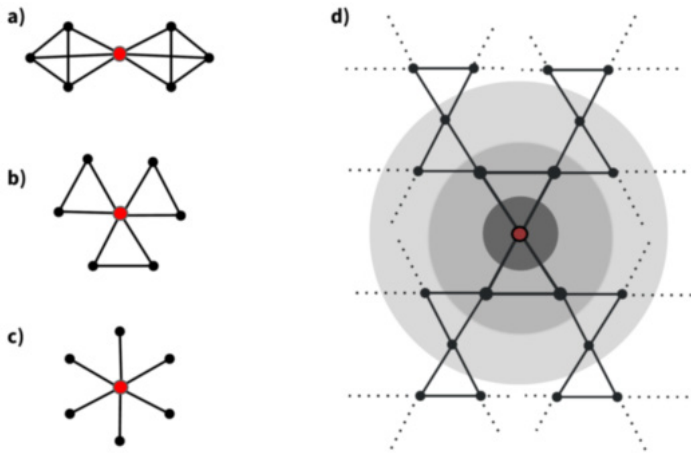

Figure 2.4. Schematic of 6-regular c-clique networks from Publication III. Panels (a-c) illustrate the neighborhoods of a (highlighted) node within networks of 4, 3, and 2-cliques, respectively. These configurations recur around each node in the random network with cliques. Panel (d) depicts a segment of the broader neighborhood in a 3-clique network, highlighting an expanded area around the panel. It's notable that each node possesses a degree of 4, with only a portion of the node connections being displayed. Such a pattern is typical in a large clique network.

work structures. Consequently, every c-clique adds $c-1$ links to a node's degree. As Fig. 2.4 shows, in cases where $c=2$, the model reduces to a *random regular graph*. A *regular graph* is a graph where each node has the same number of neighbors. We will use these types of structures in Publication III.

2.4.3 Stochastic Block Models

As we transition from examining small, closely-knit groups in network models, our focus shifts to encompassing substantially larger groups that cover a significant portion of the entire network [108]. This shift in scale necessitates a more versatile modeling approach, which is where the Stochastic Block Model (SBM) [95, 125] becomes particularly valuable. Unlike smaller groups that are typically densely connected, larger blocks in a network can exhibit a broader range of connectivity patterns. The SBM's flexibility in modeling these large, complex structures makes it an ideal choice, not only for our study but also for a wide array of applications. The SBM has demonstrated its versatility in various fields, including federated learning [104, 180], graph clustering [125], social network analysis [94], and community detection [1, 160].

Now, consider Eq. 2.4, which presents an adjacency matrix for a network comprising two distinct communities, as shown in Fig. 2.1. A *community*

in this context refers to a group of nodes that share similar connection patterns with other groups. The structure of matrix Eq. 2.4 is meticulously arranged to accentuate the block configuration of the graph, thereby illuminating regions of high density (indicative of strong intra-community ties) and areas of sparsity (reflecting weaker or non-existent inter-community connections).

In general, when constructing a network with B communities using the SBM, nodes are allocated into blocks such that $b_u \in \{1,\dots,B\}$ signifies the community or block to which node u is assigned. This model allows for the application of probabilistic rules governing the likelihood of connections both within and across these communities. For instance, in the stark scenario depicted by Fig. 2.1, nodes within a single community are completely interconnected with a probability of $p = 1$, while there are no inter-community links. Conversely, in scenarios like the one illustrated in Fig. 1.9, there is a non-zero probability for the formation of links between different communities, demonstrating the SBM's adaptability to various network configurations.

Quantitatively, the number of links between two blocks r and s (or within a single block) can be expressed as:

$$e_{rs}(G) = \sum_{u<v} A_{uv}\delta_{b_u,r}\delta_{b_v,s}, \tag{2.36}$$

where A_{uv} represents the adjacency matrix, and $\delta_{\bullet,\bullet}$ is the Kronecker delta function ensuring that nodes u and v belong to blocks r and s respectively. The ensemble of graphs generated by the SBM is, in essence, the maximum entropy ensemble [157], where the expected values of these link numbers conform to specified values, allowing for a broad range of network structures.

2.4.4 Networks with Homophily

Figure 1.9 illustrates the phenomenon of homophily regarding vaccination status within a community. Let us now discuss a particular version of block models that emphasizes the higher probability of connections within two groups than between them. This idea is, of course, consistent with the Stochastic Block Model (SBM) framework.

Consider a population where a fraction π_{v} is vaccinated, and the remainder, $\pi_{\mathrm{u}} = 1-\pi_{\mathrm{v}}$, is not. This scenario examines the propensity of individuals with the same vaccination status to connect. Specifically, π_{vv} denotes the probability of a vaccinated individual connecting with another vaccinated person, while π_{uu} represents the same for unvaccinated individuals. Crucially, these probabilities are not independent but are influenced by the overall vaccination rate, π_{v}. To model this network, we need only establish π_{vv}, as the remaining probabilities can be inferred: $\pi_{\mathrm{vu}} = 1-\pi_{\mathrm{vv}}$, $\pi_{\mathrm{uv}} = \frac{\pi_{\mathrm{v}}}{1-\pi_{\mathrm{v}}}(1-\pi_{\mathrm{vv}})$, and $\pi_{\mathrm{uu}} = \frac{1-\pi_{\mathrm{v}}-\pi_{\mathrm{v}}(1-\pi_{\mathrm{vv}})}{1-\pi_{\mathrm{v}}}$. The second equation, in par-

ticular, balances the links from vaccinated to unvaccinated individuals, conforming to $\pi_v N \pi_{vu} \langle k \rangle = (1-\pi_v) N \pi_{uv} \langle k \rangle$. For more details, see Publication I and Publication IV.

Using these connection probabilities, we can use the Coleman homophily index [49], originally proposed for social network analysis and defined by

$$h = \frac{\pi_{vv} - \pi_v}{1-\pi_v} = \frac{\pi_{uu} - \pi_u}{1-\pi_u}, \tag{2.37}$$

to determine the level of homophily in the network. This index provides a consistent measure of homophily across varying π_v values. The index has three key properties: it increases with both π_{vv} and π_{uu}, is symmetrical for vaccinated and unvaccinated groups, and ranges from 0 (no homophily) to 1 (complete homophily). Negative values suggest heterophilic networks based on vaccination status. Note that the connection probabilities $\pi_{vv} = \pi_v + \pi_u h$ and $\pi_{uu} = \pi_u + \pi_v h$ must be positive, and therefore, the Coleman homophily index is bounded from below as $h \geq \max(-\pi_v/\pi_u, -\pi_u/\pi_v)$. We will use networks with homophily in Publication I and Publication IV.

2.5 Spatial Random Graphs

Spatial networks are crucial for modeling disease spread because they incorporate the essential element of space into network structures [21, 22]. In real-world scenarios, like transportation, social contact, and mobility networks, space plays a significant role in influencing how nodes (individuals or locations) interact. The cost associated with the length of connections (links) in these networks directly impacts their topological structure. This spatial dimension is critical in understanding how diseases propagate, as it affects various dynamics like contact frequency, mobility patterns, and connectivity, all of which are key factors in the spread of diseases. Understanding the spatial constraints of networks, therefore, provides valuable insights into disease transmission and informs effective strategies for managing epidemics [87, 74]. The most basic forms of spatial networks can be exemplified by lattices or random geometric graphs. In these structures, nodes are uniformly and randomly placed across a two-dimensional area, and connections are formed between nodes if their Euclidean distance is less than a specified radius.

2.5.1 Random Geometric Graphs

Random Geometric Graphs (RGGs) represent a specific spatial graph type wherein nodes are randomly distributed within a defined space [162]. Connections between nodes are established based on spatial proximity, specifically if they fall within a predetermined distance or radius from each other. This model is particularly adept at representing networks

where spatial closeness dictates connectivity [6]. Unlike certain networks from the configuration model family, such as random regular graphs and Erdős–Rényi (ER) networks, which lack a defined spatial organization, RGGs, similar to lattices, exhibit a more systematic arrangement. These graphs are characterized by a higher degree of spatial embeddedness, reflecting the inherent spatiality in their structure. This contrast highlights the diverse ways networks can be structured, ranging from completely random to highly spatially oriented configurations.

2.6 Lattices

Lattices are a type of spatial network where nodes are arranged in a regular, repeating pattern, often resembling a grid. Unlike random geometric graphs where nodes are placed randomly, lattices have a structured and predictable layout. Each node in a lattice is typically connected to its nearest neighbors based on the lattice structure, which can vary (e.g., square, hexagonal, triangular) [5]. Lattices are used in various scientific fields, including physics, for modeling phenomena in a structured yet simplified environment [23]. They are particularly useful in studying spatial relationships and processes in a controlled, orderly framework.

2.7 Temporal Networks

So far, our discussion has centered on complex networks that do not change or evolve over time. We use temporal frameworks in Publication V, VI and Ref. [177]. In other publications, we use static complex networks. Static network models struggle to encapsulate time-dependent properties observed in many systems [97, 144, 110, 71, 33, 131, 170, 11, 178]. In models like the ER network, a node's number of neighbors might differ from other nodes, but it stays constant over time. To accurately represent the dynamics of/on social networks, we should take into account not just topological heterogeneities but also those that occur over time. Empirical observations challenge the common assumption that entities in static network models interact consistently. Many real-world systems, ranging from human behavior to natural phenomena, display bursts of high activity followed by inactivity periods [109, 18, 43, 82, 52, 25, 16, 51]. To address these limitations, time-dependent systems have been depicted as layers of static networks, sometimes termed snapshot graph sequences, each capturing behavior within a specific time window [117, 100, 77, 32, 186, 15, 210]. The core concept of temporal networks enhances static models by retaining temporal information about network interactions. In these models, an event signifies a single interaction instance between two nodes. These

interactions can be instantaneous or can span a duration. Figure 2.5(a) illustrates the temporal network using a time-line visualization. In this representation, individual nodes are depicted as horizontal dashed lines and events as vertical line segments joining two interacting nodes.

A temporal network is mathematically represented as $G = (\mathcal{V}, \mathcal{E}, \mathcal{T})$, where $\mathcal{V}$ denotes the set of nodes, $\mathcal{E}$ the set of events, and $\mathcal{T}$ the time window. Each event e in $\mathcal{E}$ is defined as $e = (\mathbf{u}, \mathbf{v}, t_{\text{start}}, t_{\text{end}})$, with t_{start} and t_{end} indicating the start and end times of an interaction. The network encapsulates dynamic interactions between entities over time. Temporal networks can be further conceptualized as an event graph $D = (\mathcal{E}, E_D, \Delta t(e, e'))$, where nodes are events and edges represent time differences between events. Adjacent events share a common node and are sequenced in time. Therefore, the event graph [137, 138, 118, 179, 101] encapsulates the full set of paths in the network, making it easier to study reachability and other properties.

In temporal networks, unlike static ones, connections are not inherently transitive. For instance, if node i connects to j, and j to k, this doesn't automatically ensure an effect from i can reach k unless the timings of these connections align sequentially. *Temporal adjacency* is defined where events e and e' are adjacent if $\mathbf{v} \cap \mathbf{u}' \neq \emptyset$ and $t'_{\text{start}} > t_{\text{end}}$, allowing the formation of *time-respecting paths*. This dependency on time and node specifics leads to the concept of temporal clusters or components distinct from those in static networks. Temporal network *reachability* requires nuanced analytical approaches. The concept of limited waiting-time reachability, or restless reachability, adds temporal constraints to paths, defining δt-adjacency based on a maximum time δt between events. This approach models various processes, including the SIS process, in a temporal network framework. For more details, see Publication V, VI, and Ref. [177].

The event graph representation encapsulates all time-respecting paths [121]. This approach facilitates the application of static network analysis methods to the structural study of temporal networks. Every path in the event graph, aligning with a time-respecting path in the temporal network, reveals the range of vertices and times they are accessible from a starting event, considering temporal adjacency. Therefore, the out-components in the event graph represent the spread or reachability from an initial event. Conversely, the in-components illustrate which events can reach a particular vertex, essentially mapping the influence or reachability to that event. This approach allows for a detailed understanding of how effects propagate and interact within a temporal network.

It is worth noting that to determine the out-component of a single event node, a straightforward breadth-first search (BFS) suffices. However, the task becomes more computationally expensive if we aim to grasp broader reachability patterns in temporal networks, such as how the reachability cluster evolves on average across multiple start points. One could sample

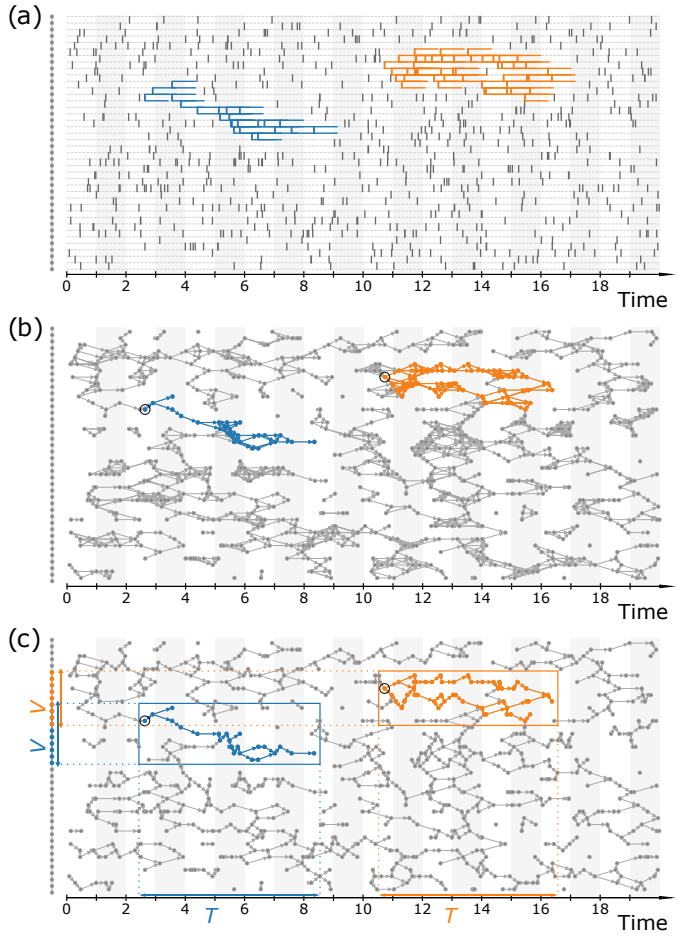

Figure 2.5. Representation of a temporal network and two limited-waiting-time spreading processes on that network, from Publication VI. The network is built using a static path graph, which is shown vertically to the left, and each link has activation times determined by the Poisson process. In part (a), the temporal network is depicted through a time-line chart. Here, every horizontal dashed line stands for an individual node, while vertical solid line segments indicate immediate interactions or events between two nodes at a given moment. Two specific spreading processes with limited waiting times, initiated from two separate events, are highlighted using different colors. These colored horizontal segments indicate the duration of infectivity for each node, and the highlighted events mark potential paths of transmission.
(b) The event graph of the temporal network is illustrated as a directed acyclic graph. In this representation, every event from the original network becomes a node in the event graph. Two nodes are linked if their corresponding events are next to each other. (c) This event graph can be simplified by removing unnecessary feed-forward loops, all the while preserving the characteristics of the spreading process. The **Cluster volume**, which refers to the distinct count of infected nodes in the temporal network, and the **Cluster lifetime**, representing the duration from the beginning to the end of the process, are determinable from the properties of the event graph. Meanwhile, the **Cluster mass** — the cumulative node-hours of infectivity, equivalent to the total of colored horizontal lines seen in panel (a) — can be gauged in the event graph by counting the distinct events in the infected cluster.

a selection of starting events and execute a BFS for each, but this method grows costly in terms of computational resources. This is mainly because each BFS demands operations on the order of $O(|\mathscr{E}_D|)$, where $\mathscr{E}_D$ represents the set of links or adjacency relationships within the event graph. Ref. [14] introduces a method that efficiently computes estimates of sizes of in- or out-components in temporal networks. This method has far-reaching implications, as we show in Publication V and VI.

> A stone is a prototypical "thing": we can ask ourselves where it will be tomorrow. Conversely, a kiss is an "event." It makes no sense to ask where the kiss will be tomorrow. **The world is made up of networks of kisses, not of stones.**
>
> Carlo Rovelli, *The Order of Time*

2.8 Percolation Theory

Percolation theory is a mathematical framework used in statistical physics, materials science, and network theory to study the behavior of connected components in a random graph [173]. It provides a way to analyze how various properties of a system change as the number of connections or occupied sites increases. The theory is particularly famous for describing phase transitions, which are sudden changes in the system's properties.

Percolation theory is also useful in understanding disease propagation in social networks, offering analytical tools that will be specifically utilized in Publication III and IV. This theory emphasizes the need to consider both the structural aspects of networks and the nuances of individual disease transmission behaviors. Our development of this framework, as detailed in Publication V, VI, and Ref. [177], extends its application to temporal networks. We map spreading phenomena onto directed percolation problems, thereby enhancing our understanding of how diseases proliferate over time in complex network structures.

2.8.1 Bond Percolation

In Sec. 2.2, we introduced the ER network $G(N,p)$ where each pair of nodes where connected with probability p. That model results in a network that undergoes a phase transition, as illustrated by Fig. 2.2, from a disconnected to a connected system such that if $p > p_c$, the giant component of the network spans a non-zero fraction of all nodes. The wiring probability p stochastically connects any pair of nodes and acts as the *control parameter* here and adjusts the network's giant component size, the *order parameter*

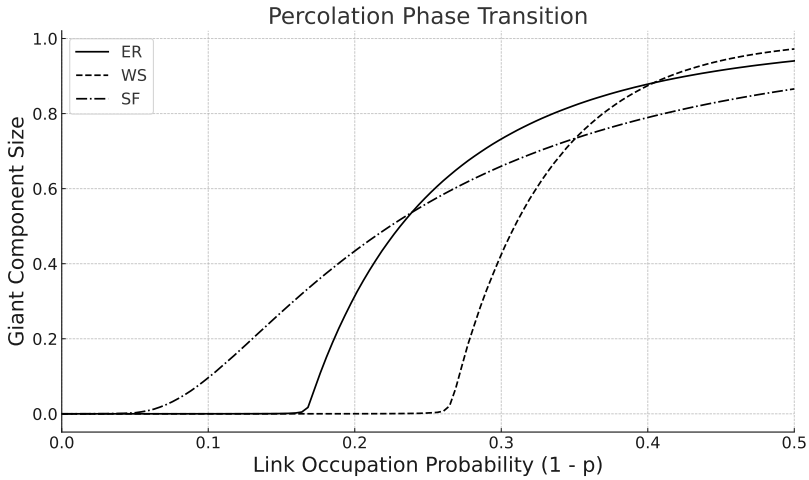

Figure 2.6. Simulation results for the percolation phase transition in three networks (each of size 6×10^5 and average degree 6). The solid line represents the Erdős-Rényi (ER) network, with a critical probability ($p_c = 1/6$), as discussed in Sec. 2.2.1. The dash-dotted line illustrates a configuration model network with a heavy-tail degree distribution ($p_k \sim k^{-3.01}$, see Sec. 2.3), which is more robust against link removal. The critical probability for this network can be derived using Eq. 2.38. This network exhibits more degree heterogeneity compared to an ER network and demonstrates lower critical p. The dashed line shows a Watts–Strogatz small-world network with a rewiring probability of 0.1 (see Sec. 2.2.2), which has a higher percolation threshold than the ER network. Due to its high clustering coefficient, the standard approach for calculating the epidemic threshold (Eq. 2.38) is not applicable for this network.

of the problem. We can do the opposite: Begin with any graph G, remove any link uniformly at random with probability p (or leave the link occupied with probability $1-p$), and check how the giant component size varies as a consequence of link removal. Fig. 2.6 shows what happens to the size of the giant component of three different networks as we remove their links uniform at random with p. As we can see, different networks undergo different changes depending on their topology. When the giant component emerges by including a non-zero fraction of the nodes in the network, we say the network has percolated.

In the realm of physics, this approach is known as bond-percolation, where only a subset of the system's original bonds (or links) remain active [47, 128]. Most connected graphs experience a phase transition at a particular probability, p_c, which is referred to as the bond percolation threshold. In the case of tree-like networks, this threshold is determined by observing how the percolation probability p affects the average excess degree, given that the transition occurs when $\langle q \rangle = 1$. It's worth noting that various properties, such as the size of the giant component, small component size distribution, network percolation characteristics, and more, are all intrinsically linked to this tree-like property.

For configuration model networks, we can find the size of the giant component or the percolation threshold using the same arguments we followed in Sec. 2.3.2 such that

$$p_c = \frac{1}{g_1'(1)} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}, \tag{2.38}$$

where g_1 is the generating function of the excess degree distribution of the configuration model network.

2.8.2 Site Percolation

If instead of removing or keeping the links, we remove each node and its connecting links from the network, a process known as *site-percolation* [168], we will again go through a phase transition but this time through a different mechanism such that the threshold happens at different values. Vaccination can be modeled as a site-percolation process since by vaccinating people, we remove them from the transmission network [202, 129].

2.8.3 Percolation as a Critical Phenomenon

Percolation phase transitions are a fascinating area of study in statistical physics, characterized by the emergence of a giant component that connects a significant portion of a system [128, 148]. This transition is governed by *scaling laws* and *critical exponents* [81], which offer a quantitative framework for understanding these phenomena. These laws describe the behavior of various physical quantities as the system approaches the critical threshold of percolation, denoted as p_c. Near the critical point, properties of the system exhibit power-law scaling, indicating that they can be expressed as a power of the distance from p_c. Critical exponents are key to understanding these scaling laws [81]. They describe how certain properties of the system change as the percolation threshold is approached [62]. Accordingly, the size of the giant component and the average size of small components scale with their specific critical exponents at the critical point [63]. This can be expressed as:

$$\sigma \sim |p - p_c|^{-\gamma}$$

The symbol $\sim$ denotes that the left-hand side scales as a power of the right-hand side, and γ is the critical exponent that describes how the average size of components diverges as the percolation threshold p_c is approached. Another singular property at the critical point is the *correlation length*. The correlation length ξ indicates the scale of spatial correlations in a system, say, the typical size of components, and its divergence signifies the onset of long-range order. The scaling of the correlation length can also be expressed as follows:

$$\xi \sim |p - p_c|^{-\nu}$$

where ν is the critical exponent associated with the correlation length.

These scaling laws and critical exponents are not just mathematical curiosities; they reveal the underlying *universality* of phase transitions [81]. A *universality class* in the context of statistical physics and critical phenomena is a classification that groups various physical systems and models based on their similar critical behavior, particularly as they approach a phase transition. A fridge magnet and boiling water exhibit similar behaviors near their critical temperature, despite their differences [81]. This concept suggests that different systems, even with diverse microscopic structures or dynamics, can exhibit identical large-scale statistical behavior near their critical points. Besides, it highlights the profound interconnectedness of physical and mathematical worlds, where the same fundamental rules can govern both material states and abstract structures.

Taking percolation as an example, systems within the same universality class will share critical exponents that dictate the scaling behavior of key properties like component sizes or correlation lengths near the percolation threshold. This remarkable similarity means that by studying the critical behavior of one system in a given universality class, we can infer the behavior of other systems in the same class despite their microscopic differences. This principle of universality is crucial in simplifying and unifying our understanding of complex phenomena across a wide range of disciplines, from condensed matter physics to network theory. Ref. [39, 86, 187, 63] provide comprehensive insights into the mathematical foundations and physical interpretations of scaling laws and critical exponents in percolation theory.

2.8.4 Directed Percolation

Following the exploration of percolation in various network topologies, as discussed in the context of the ER network and configuration model networks, we now turn our attention to a specific type of percolation process known as *directed* percolation. This process is a cornerstone example of continuous, non-equilibrium phase transitions, distinct from *isotropic* phase transitions due to its imposition of a preferred direction in one of the system's dimensions [90, 91].

Consider the process of brewing espresso as an illustrative example of a percolation process. In this scenario, the network is the finely ground coffee held within a portafilter, a porous medium. The percolation occurs when hot or steamed water is forced through the coffee grounds, extracting the coffee from the bottom. Unlike isotropic percolation processes, the liquid in espresso-making travels predominantly in a downward direction, guided by the combined forces of the espresso machine's pressure and gravity. This directional flow of water through the coffee grounds exemplifies a directed percolation process, where the flow is not random or isotropic but instead

follows a specific, predetermined path. This directional bias in the percolation process significantly alters the dynamics of the system, disrupting the usual symmetry between the horizontal and other spatial dimensions. This phenomenon, where the percolation process is guided in a specific direction, can be likened to the concept of directed percolation in physics [90], as opposed to isotropic percolation, where the fluid or the agents would move uniformly in all directions. This directed percolation process, as seen in espresso brewing, is a metaphoric analogy to the theoretical models of directed percolation discussed in non-equilibrium physics. It provides a tangible example of how directional constraints in a percolation process can fundamentally change the behavior and outcome of the system.

Directed percolation can also be conceptualized as a $d+1$-dimensional system with one dimension having a preferred direction, time evolution, given the unidirectional nature of time [91, 90]. In this framework, the evolution of the system in discrete-time directed bond percolation is governed by two simple rules: 1) each occupied node leads to the occupation of neighboring nodes, and 2) any occupation not reinforced in a turn results in the node becoming unoccupied. This dynamic leads to three distinct phenomena: death (or extinction at dead-ends), multiplication (where an occupied node infects a neighboring one), and coalescence (where two occupied nodes interact, leaving only one occupied).

Similar to the isotropic percolation discussed earlier, the evolution of directed percolation systems is characterized by changes in order parameters and characteristic quantities in response to variations in the control parameter p, the probability of link existence. At a critical threshold $p = p_c$, the system undergoes a phase transition. Below this threshold, percolating clusters starting from a single node are finite in both lifetime and size. As the threshold is approached, these clusters grow, and beyond it, they remain indefinitely active in an infinite system [90, 91]. This behavior is quantified through various parameters such as the probability of survival $P(t)$, mean cluster mass M, volume V, and lifetime T, each scaling with different critical exponents.

In Publication V and VI, we will discuss in more detail that in scenarios where the spreading process begins from all nodes at $t = 0$, the system remains active for a finite duration at $\tau < 0$ but indefinitely for $\tau > 0$. This leads to the definition of occupation density $\rho(t)$ and stationary density ρ_{stat}, which scale differently based on the system's state relative to the critical threshold. The influence of an external field on ρ_{stat} is particularly noteworthy near the critical threshold, where even minor changes can significantly affect the system, as indicated by the susceptibility χ.

While these quantities have distinct critical exponents, not all are independent. For example, the static occupation density ρ_{stat} and the ultimate probability of survival P_∞ are closely related under a time-reversal transformation, leading to the equality of certain critical exponents [90].

Despite its theoretical elegance and extensive study, practical observations of directed percolation systems in real-world scenarios were elusive until recently. This gap is partly due to the difficulty in replicating certain features of classical directed percolation systems, such as a non-fluctuating absorbing phase, in natural settings [90]. However, recent experimental breakthroughs have begun to bridge this gap, demonstrating the applicability of directed percolation concepts in various fields ranging from star formation to biological evolution [190, 175, 126]. Ref. [177] revolves around analyzing the temporal evolution of spreading dynamics on networks. We contribute to forming a concrete connection between temporal network reachability and percolation theory in Publication V and VI. We show that limited-waiting-time reachability, a feature of constrained connectivity in temporal networks, undergoes a directed percolation phase transition. This suggests that spreading behaviors on these networks can be analyzed using directed percolation universality class, a finding consistent across various models and actual temporal networks. Building on this, we aim to connect the reachability in temporal networks to epidemic spreading similarly as connectivity in static networks is related to late-stage disease-spreading results [150, 114].

3. Epidemics on Networks

The COVID-19 pandemic has made us come to grips with the epidemic problems, as they offer not only a set of practical problems to solve but also interesting theoretical puzzles. The interconnectedness of individuals in society plays a pivotal role in determining the scale and likelihood of an outbreak [158, 115]. Integrating more realistic assumptions into our models while employing advanced mathematical and computational tools to study these complex systems is imperative. In this chapter, we discuss how different diseases unfold in social networks and how, and in what ways, various interventions can be effective considering the networked structure of human populations.

3.1 Identifying Key Nodes in Networks

As discussed in Section 2.2.2, the friendship paradox is a key concept in understanding the dynamics on social networks, particularly in the context of disease spread [115]. In social networks, certain individuals, known as *hubs*, are characterized by their high number of connections [158]. These hubs play a crucial role in the propagation of diseases. Due to their extensive network of connections, hubs are more susceptible to contracting illnesses. The significance of hubs in accelerating the spread of diseases within networks is twofold. Firstly, due to their higher-than-average number of connections, these hubs have a greater probability of encountering and contracting infections. Secondly, once infected, their extensive connections facilitate them to transmit the disease to a larger number of individuals compared to a typical node.

In the case of Severe Acute Respiratory Syndrome (SARS), a viral respiratory disease caused by a SARS-associated coronavirus, data from Singapore revealed that the index patient, who had a large number of contacts, transmitted the disease to many people [102]. Not all of these contacts further spread the disease, but a few acted as hubs and were responsible for its widespread transmission. This highlights the critical role of these hubs

in propagating the disease. Therefore, an interesting strategy would be to focus vaccination efforts specifically on these *super-spreaders*. The idea of *targeted vaccination*[159, 76], suggests that immunizing these key individuals could disrupt the transmission chain, effectively curbing the disease spread with fewer vaccine doses. The complexities and strategies of vaccination campaigns, including this targeted method, are explored in detail in Publication I and Chapter II.

However, identifying these hubs poses a challenge. Traditional vaccination methods often rely on random selection, which may not be optimal. This is because the majority of the population tends to have average or below-average connections, whereas the few true hubs have a disproportionate impact on spreading the disease [158]. To address this, an innovative approach called *acquaintance immunization* was proposed [48]. Rather than vaccinating individuals chosen at random, this method involves asking the selected person to nominate a friend or acquaintance for vaccination. By doing so, the likelihood of vaccinating a hub increases, making the process more efficient and targeted in curtailing the spread of diseases like SARS.

The logic behind the effectiveness of acquaintance immunization is similar to the way that the friendship paradox works. At first, this might seem counter-intuitive, but there's a logical reasoning behind it. If a person has a high number of connections (a hub), they have a higher chance of being nominated by any one of their numerous acquaintances. On the other hand, a person with fewer connections has fewer chances of being nominated. Therefore, this method naturally leads to the identification and vaccination of hubs more frequently than a random approach would. This method has been tested through simulations and works well there. However, implementing the acquaintance immunization strategy in real-world scenarios presents distinct challenges. First, there is a technical hurdle: acquiring accurate network data can be complex and demanding. This data is crucial for identifying key individuals within the network and ensuring the strategy's effectiveness. Second, a moral or legal challenge arises when asking individuals to nominate friends. This request may encounter resistance due to privacy concerns, as people might be hesitant to share personal connections, or due to logistical issues in gathering and managing such sensitive information. Despite these challenges, the potential effectiveness of this strategy, as demonstrated in simulations, highlights its promise as a valuable tool in combating infectious diseases.

3.2 Modeling Epidemics on Networks

To delve into epidemics within structured populations, it's essential to weave together the compartmental models presented in Chapter 1 with

the network system methodologies discussed in Chapter 2. Achieving analytical arguments necessitates analyzing the interplay between the structure of contact patterns and the dynamics of disease propagation. The primary hurdle is that compartment models are typically characterized by coupled non-linear differential equations. Directly solving these complicated equations on a graph is a daunting task, both numerically and analytically.

Considering the inherently high-dimensional nature of dynamical systems on networks, we need to find a more manageable, lower-dimensional representation that retains the essence of the original system. Such representations should ideally be more intuitive and potentially amenable to analytical approaches, like deriving approximate closed-form expressions for quantities of interest [167]. *Mean-field* theories stand out as a promising strategy to achieve this [115]. When it comes to dynamical systems on networks, these theories and similar methods of approximation are of special importance [19].

3.3 Mean-field Approximations

In Physics, a mean-field approximation is a technique where the state of a specific particle (for us, a node or a link, depending on the context, in a network) is viewed in interaction with the collective average state of the rest or, extending the idea, the distribution of states among all particles. This approach is a valuable starting point for deciphering complex systems. The approximation becomes even more accurate when interactions are broadly of the same magnitude, reminiscent of those in a densely populated random graph [167].

Mean-field approximation and heterogeneous mean-field theory are commonly used in the study of epidemic spread on networks [158]. Generally, the mean-field approximation assumes that each node in the network interacts with every other node in an identical manner, which simplifies the model but may not capture network heterogeneity. On the other hand, heterogeneous mean-field theory takes into account the different characteristics and behaviors of nodes within a network, providing a more nuanced and accurate representation of epidemic spread [201].

3.4 From the Contact Networks to Transmission Networks

If we model the contact network of a human population with network G, the actual disease transmission might occur on a different network G', depending on the dynamics of the disease. The initial network, G, is termed the *contact network*, while G' is referred to as the *transmission network*

[41]. In practice, accurately describing contact networks is challenging due to difficulties in measuring them [20]. Furthermore, there's still much to learn about contact networks' topological characteristics and temporal evolution. Gathering sufficient observations and data for a comprehensive analysis remains a hurdle [155]. It is worth noting that transitioning to the transmission network that dictates disease dynamics is not straightforward, even with a clear understanding of a contact network.

Consider a rather simple scenario where a disease spreads across a contact network G. The disease's dynamics manifest on this network such that every link in the contact network has a probability p of actively transmitting the disease [83]. To transition from the contact network to the transmission network, one would replicate the original nodes, retaining each link based on probability p. The resulting graph represents the transmission network. Each node in this network represents an individual infected at some point, and the number of nodes in the giant component of this new network can serve as an indication of the epidemic size [150, 114, 113]. This mapping process is called the *bond-percolation* method. We discussed percolation in depth in Sec. 2.8.

To execute a bond-percolation mapping, it's essential to understand the structure of the original contact network. In many real-world scenarios, our knowledge about the contact network's structure is limited. Yet, we might possess a reliable estimate of the transmission network. Knowing the transmission network simplifies many analytical calculations. For instance, given a transmission network from the configuration model family, with degree distribution p_k, we can identify the basic reproduction number as the mean excess degree of the transmission network [143, 149, 191] as

$$R_0 = \langle q \rangle . \tag{3.1}$$

This is because the average excess degree of a transmission network yields the expected number of secondary cases produced by a typical infectious individual over the course of their infectious period in a fully susceptible population. For locally tree-like networks with no correlations in their connections, the basic reproduction number can be written as a function of the first and the second moments of the degree distribution of the transmission network, given by Eq. 2.19.

This understanding significantly aids in the application of our mean-field approximations. Depending on the quantity of interest, we can write different equations regarding the network structure. In the subsequent sections, we will delve into the practical applications of transmission networks.

3.5 Mapping Epidemics to Percolation

Disease spread in network models, similar to fully mixed ones in Chapter 1, operates on the contact network and is defined by a transmission rate, β, indicating the infection probability per unit time between two connected individuals. However, β in network models differs from fully mixed models as it refers to contacts between an individual and their direct connections, not the entire population. The transmission rate is influenced both by the disease's nature and social behaviors, affecting how often and closely people interact within their communities.

Network models can simulate how diseases like the SI model spread over time, with β determining the infection spread to connected susceptible individuals. While calculating the disease's spread over a general network is complex, the eventual size of an outbreak is predictable: it will include all individuals reachable from the initial infection through network paths. This behavior, where an outbreak's size depends on the network's structure and the initial infection's position, introduces stochastic elements into the model, leading to different outcomes even with identical parameters.

These network and percolation models capture more realistic dynamics of disease spread than fully mixed models, acknowledging that not all exposures lead to an epidemic. Realistically, diseases can fizzle out if initial cases do not transmit to others. The models recognize the randomness in transmission and the variability in outbreak patterns, yielding probabilities or averages rather than precise predictions.

3.5.1 SIR Model and the Configuration Model

The SIR model of Sec. 1.2.2 adds the parameter of recovery time to the network model, where the disease may not always spread to a susceptible individual before the infected one recovers. The transmission probability p is based on the transmission rate β and recovery time τ, simplifying to [150]

$$p = 1 - e^{-\beta\tau}. \tag{3.2}$$

Therefore, using the percolation threshold, the critical probability p_c, we can rearrange Eq. 2.38 to give [150]

$$\beta\tau = -\ln(1-p_c) = \ln\frac{\langle k^2\rangle - \langle k\rangle}{\langle k^2\rangle - 2\langle k\rangle}. \tag{3.3}$$

Using a technique akin to bond percolation, where links are active or occupied with a probability p, we can predict that an outbreak will spread across connected occupied links, corresponding to potential disease transmission [150, 114]. The bond percolation model also indicates that the giant component size corresponds to the final epidemic size, resembling how diseases spread through populations. Similar to Sec. 2.3.2, given the

transmission probability p, the total probability ϕ that a node does not belong to the giant component can be written as [151]

$$\phi = \left\langle 1 - p + p\phi^k \right\rangle_Q = 1 - p + pg_1(\phi), \tag{3.4}$$

and the size of the giant component can be calculated as

$$\Sigma = 1 - \left\langle \phi \right\rangle = 1 - g_0(\phi), \tag{3.5}$$

where g_0 and g_1 are the probability generating functions for the degree and excess degree distributions of the contact network, respectively.

When the product of the transmission rate and the infectious period surpasses a certain threshold, an epidemic becomes a feasible event. However, this does not guarantee an outbreak will occur, as the disease's initial host might not be part of the giant component of the transmission graph where the epidemic could take hold. Conversely, if the product of β and τ does not reach the critical threshold, an epidemic will not unfold, irrespective of the seed of infection within the population. The likelihood and potential extent of an epidemic provided that it is feasible, are quantified by Eq. 3.5.

The progression towards an epidemic within this model is influenced by the value of $\beta\tau$. An increase in either the time an individual remains infectious τ or the rate at which the disease is transmitted β can propel the system toward an epidemic state. The specific value at which this transition occurs, along with the chances and magnitude of an outbreak, is intricately linked to the network's configuration, particularly the average number of connections $\langle k \rangle$ and connections squared $\left\langle k^2 \right\rangle$. This consideration of network structure marks a stark contrast from the fully mixed model, which does not take into account the effects of a network's configuration [140, 113].

In Publication I, we utilize this technique to calculate the number of individuals infected after a vaccination campaign. Building on this approach, in Publication IV, we further apply the same percolation trick to accurately determine the final outbreak size, incorporating the impact of contact tracing. This demonstrates the adaptability and robustness of our methodology in various public health scenarios.

3.6 Spreading on Temporal Networks

So far, our investigations have focused exclusively on static networks, which remain unchanged over time. While this approach offers a reasonable approximation for modeling social networks, it is both theoretically and practically beneficial to integrate temporal heterogeneities. as we discussed in Sec.2.7, into these networks [100]. By doing so, we can more accurately assess the impact of the temporal characteristics of contact networks on disease transmission.

Temporal networks differ from static ones in that they inherently include interaction timings, capturing real-world correlations and inhomogeneities like diurnal patterns and burstiness of activity, which significantly influence spreading scenarios [134, 99, 97]. These networks also reflect structural changes over time, crucial in studies like sexually transmitted diseases, where relationship dynamics or geographical moves can alter contact networks [134, 24]. The results from simulating spreading processes on temporal networks often differ significantly from those on static networks due to factors like linger time distribution and inter-event time distribution [97, 144, 110, 71, 33, 131, 170].

3.6.1 Directed Percolation and Spreading Phenomena

Expanding upon our previous discussions, there is a well-established connection between reachability and disease spread in static networks, underpinned by percolation theory as highlighted in Sec. 2.8. However, this interplay is less examined in temporal networks. Our research, detailed in Publication V and VI, is dedicated to analyzing the temporal evolution of spreading dynamics on networks and forging a solid link between reachability in temporal networks and percolation theory. We have uncovered that in these networks, limited-waiting-time reachability, a kind of constrained connectivity discussed in Sec. 2.7, exhibits a directed percolation phase transition as described in Sec. 2.8.4. This discovery implies that spreading behaviors in temporal networks can be insightfully studied using the directed percolation universality class, a finding consistent across a variety of models and actual temporal networks.

Transferring analytical tools from static to temporal networks is a complex task. We aim to bridge this gap by correlating reachability in temporal networks with epidemic spreading, akin to how connectivity in static networks is associated with late-stage disease transmission, as extensively researched in [150, 114]. Adapting some methodologies, we can interpret reachability in temporal networks through event graph representation, as cited in [118, 179, 101]. In these networks, 'reachability'—similar to connectivity in static contexts—encompasses the capability to trace a chronological path from one node to another amidst the fluctuating nature of network connections over time, as described in [96]. The event graph method outlined in Sec. 2.7 effectively translates time-sensitive reachability data into a higher-order static directed acyclic graph (DAG), elaborated further in Ref. [177]. This graph's out-component provides insights on nodes and periods that are reachable, whereas its in-component sheds light on feasible origins and timings to reach certain destinations.

The directed percolation phase transition and its universality class, introduced in Sec. 2.8.4, provide a valuable framework for converting these mathematical concepts into a tangible physical model. This framework

greatly improves our comprehension of the dynamics within temporal networks, especially concerning the spread of diseases and the flow of information. In Publication V, VI, and Ref. [177], we have developed an extensive framework that maps specific disease dynamics to a directed percolation problem. This significant development enriches our understanding of how temporal changes in network structures can influence both reachability and, ultimately, the spread of diseases in temporal environments.

3.7 Homophily and Herd Immunity Threshold

A good example to see the power of a mean-field approach is to consider a locally tree-like homophilic transmission network similar to the one in Sec. 2.4.4. Let's say we are interested in the epidemic threshold as a function of network structure and spreading parameters. Assume a naive population that we can vaccinate π_{v} fraction of the individuals given this condition that vaccinated and unvaccinated nodes are connected to each other with different probabilities as depicted in Fig. 1.9 and the network degree distribution follows p_k. We will go through more details and calculations in Publication II.

Our focus is on a group of epidemic models in which infection induces complete and permanent immunity, whereas the immunity induced by vaccines is generally incomplete. We represent the number of infections among vaccinated and unvaccinated groups at generation m (stemming from the initial infected case) as $I_{\mathrm{v}}^{(m)}$ and $I_{\mathrm{u}}^{(m)}$, respectively. Assuming ideal vaccine efficacy, the vaccinated group would experience no infections. Hence, under a mean-field approximation, we can express the situation using the following recurrence formulas:

$$I_{\mathrm{v}}^{(m+1)} = 0, \tag{3.6}$$

$$I_{\mathrm{u}}^{(m+1)} = R_0 \pi_{\mathrm{uu}} I_{\mathrm{u}}^{(m)}, \tag{3.7}$$

which is a typical branching process, with branching factor $R_0\pi_{\mathrm{uu}}$. Remember that π_{uu} is the probability that an unvaccinated individual is connected to another unvaccinated in our homophilic network (Fig. 1.9) model of Sec. 2.4.4.

It's important to recognize that vaccines might not be flawless. Two particular effects of vaccine protection are crucial when thinking about herd immunity [88, 70]. One, the vaccine might lower the chance that someone gets infected when exposed. This decrease is known as the efficacy against susceptibility, labeled as f_{S} [130, 106]. In our discussion, we're suggesting that the vaccine provides complete immunity to a portion f_{S} of those vaccinated, while the rest remain entirely vulnerable. Our model views the vaccine as being *all-or-nothing*, unlike *leaky* vaccines that slightly reduce everyone's susceptibility. These models are equivalent under the

assumption that the network is locally tree-like, on which our study is based [88, 70]. Secondly, those who get infected even after vaccination might be less likely to pass on the infection. We capture this using the efficacy against infectiousness, f_I, which indicates how much the rate of secondary infection is reduced. Given these two new parameters, the previous mean-field equations can be written as:

$$I_\mathrm{v}^{(m+1)} = (1-f_\mathrm{S})R_0[(1-f_\mathrm{I})\pi_{\mathrm{vv}}I_\mathrm{v}^{(m)} + \pi_{\mathrm{uv}}I_\mathrm{u}^{(m)}], \tag{3.8}$$

$$I_\mathrm{u}^{(m+1)} = R_0[(1-f_\mathrm{I})\pi_{\mathrm{vu}}I_\mathrm{v}^{(m)} + \pi_{\mathrm{uu}}I_\mathrm{u}^{(m)}], \tag{3.9}$$

where $\pi_{\mathrm{uv}} = 1-\pi_{\mathrm{uu}}$ and $\pi_{\mathrm{vu}} = 1-\pi_{\mathrm{vv}}$ are the conditional probabilities that a link from one group points to the other. We can rewrite these equations in a matrix form such that

$$\mathbf{I}^{(m+1)} = \mathbf{M}\mathbf{I}^{(m)}, \tag{3.10}$$

where $\mathbf{I}^{(m)} = (I_\mathrm{v}^{(m)}, I_\mathrm{u}^{(m)})^\intercal$ and

$$\mathbf{M} = R_0 \begin{pmatrix} (1-f_\mathrm{S})(1-f_\mathrm{I})\pi_{\mathrm{vv}} & (1-f_\mathrm{S})\pi_{\mathrm{uv}} \\ (1-f_\mathrm{I})\pi_{\mathrm{vu}} & \pi_{\mathrm{uu}} \end{pmatrix},$$

we see that the infection eventually dies out after a finite number of generations if all the eigenvalues of the next-generation matrix $\mathbf{M}$ have an absolute value of less than one. That is, at the critical point, the spectral radius $\rho(\mathbf{M}) = 1$.

In Publication I, we show that by adjusting the connection probabilities using π_v and Coleman homophily index h from Eq. 2.37, the critical vaccine coverage for achieving herd immunity so that $R_0 = 1$ can be expressed as:

$$\pi_\mathrm{v}^\mathrm{c} = \frac{1-\epsilon R_0 h}{(1-\epsilon)(1-h)}\left(1-\frac{1}{R_0}\right), \tag{3.11}$$

Here, ϵ is defined as $(1-f_\mathrm{S})(1-f_\mathrm{I})$ and it's necessary that $\epsilon \leq 1/R_0$. If $\epsilon > 1/R_0$, the threshold for vaccination becomes obsolete, making herd immunity unreachable.

In cases of a perfect vaccine, where either $f_\mathrm{S} = 1$ or $f_\mathrm{I} = 1$ or both:

$$\pi_\mathrm{v}^\mathrm{c} = \frac{1}{1-h}\left(1-\frac{1}{R_0}\right), \tag{3.12}$$

This equation narrows down to the commonly referenced threshold in Eq. (1.22) when there's homogeneous mixing and $h = 0$. The equation suggests that as the intensity of homophily h grows, the critical vaccine coverage $\pi_\mathrm{v}^\mathrm{c}$ for achieving herd immunity also increases (as illustrated in Fig. 3.1). Simply put, having more homophily makes it more challenging to achieve herd immunity. It's key to highlight that when

$$h \geq \frac{1}{R_0}, \tag{3.13}$$

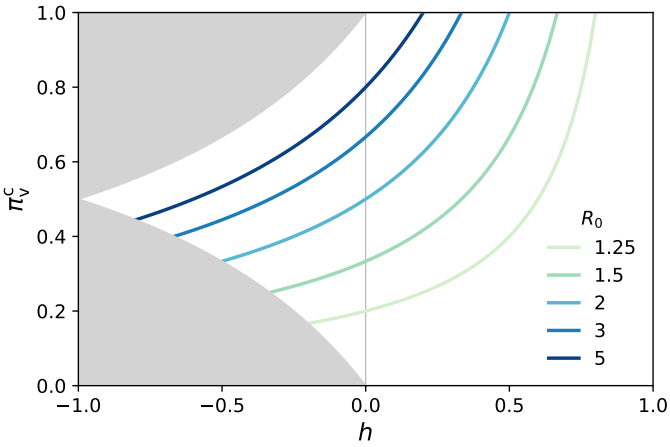

Figure 3.1. Critical coverage π_v^c of a perfect vaccine required for herd immunity as a function of homophily strength h for different values of basic reproduction number R_0. Positive values of h indicate homophily, while negative values point to heterophily. The gray-shaded area in the figure denotes the parameter space where the network cannot be realized.This figure is from Publication I

herd immunity can't be achieved unless everyone is vaccinated. In other words, regardless of how tiny the group of unvaccinated individuals might be, there's always a risk of a significant outbreak within that group.

We built a network-based framework to accommodate homophily related to interventions. Our results show that a small level of homophily in vaccination status can considerably increase the threshold required for herd immunity and even make herd immunity impossible to reach. To truly understand and predict the dynamics of spreading processes in the presence of vaccination homophily, it is essential to incorporate these factors into our models and analyses. For more details, see Publication I.

3.8 Disease-induced Herd Immunity

When immunity arises from a natural infection, the inherent heterogeneity in human interactions can significantly alter the dynamics of disease spread. In contrast to vaccine-induced immunity, natural immunity exhibits variations in its strength and nature. In particular, diseases often target highly connected individuals in the early stages, leading to the efficient immunization of these influential nodes [35]. However, while epidemics tend to target and remove high-degree nodes, enhancing herd immunity, the localized nature of these infections within the network can simultaneously weaken it.

We will investigate the strengths and weaknesses of disease-induced

herd immunity in Publication II. Our analysis is anchored in understanding various network structures, with particular emphasis on two primary aspects: degree heterogeneity (which looks at the variations in how nodes connect) and spatial embeddedness (which focuses on the placement of these networks within a defined space). For a clearer picture, see Figure 3.2.

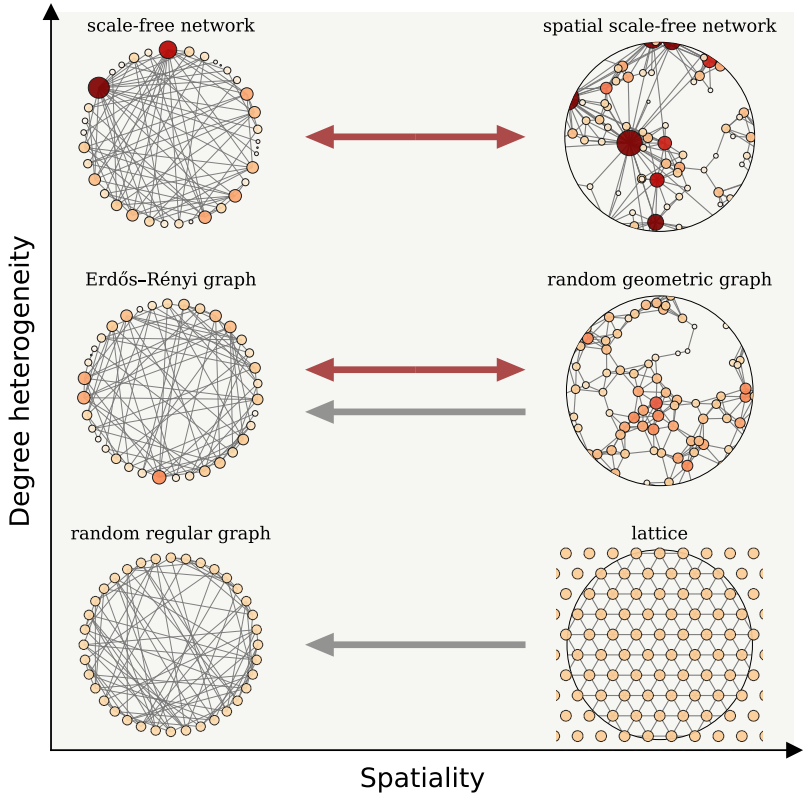

Figure 3.2. Network models according to their level of degree heterogeneity and spatiality from Publication II. The three arrows indicate directions of sweeps conducted through rewiring randomization. Spatial embeddedness describes how a network is geometrically arranged in a set space. At one end, networks like RRGs and Erdős-Rényi networks show minimal spatiality, representing the most randomness. On the opposite end, lattices and RGGs have the most spatial structure

We use link randomization techniques [204] to bridge the gap between networks with high and low spatiality. Starting with a highly spatial network, like a lattice or a random geometric graph, we modify a fraction of its links. This modification can be done in two ways: double link swap (which keeps the node's degree constant) and random rewiring (which maintains the total link count but not individual node degrees). Fully randomized links transform a lattice into a random regular graph and

a random geometric graph into an ER network. However, any network becomes an ER network with complete random rewiring, regardless of its starting point.

During these modifications, the network starts to show small-world traits [204], thanks to the formation of long-range connections. This characteristic persists until the network's local structures dissolve, but the average node degree remains unchanged. Refer to Publication II for more details. Our results reveal that when there's no degree heterogeneity, disease-induced immunity has a milder impact compared to random immunization. However, for Erdős-Rényi networks, both methods offer similar herd immunity levels. This outcome stems from two opposing forces at play in disease-induced herd immunity. On one hand, epidemics tend to target and remove nodes with many connections, bolstering the herd immunity effect. On the other hand, the spread is concentrated and localized within the contact network, which weakens herd immunity. See Fig. 1.10 for a visualization. These competing forces are reflected in the number of connections between susceptible and removed nodes and the average connections of removed nodes. In networks with slight heterogeneity, the localization's influence is more pronounced than the targeted removal of well-connected nodes.

3.9 Epidemic Spreading and Contact Tracing

Ibn Sina (Avicenna, 980–1037 CE), a Persian polymath, suspected that some diseases were spread by microorganisms. He highlighted the concept of quarantine in his influential work, *The Canon of Medicine* [17]. Contrary to Galen but in line with Aristotle, he identified tuberculosis as contagious [185]. It is clear that since the old times, quarantine as a restriction on the movement and interactions between people has been around as an effective non-pharmaceutical intervention for infectious diseases.

Contact tracing involves identifying and isolating individuals who have been exposed to infected persons [34]. This method is not only effective in containing the spread of diseases but also practical [208]. In this section, we explore models that integrate epidemic spread with the implementation of contact tracing. Contact tracing as a targeted approach proves cost-effective and allows for the possibility of relaxing certain social distancing measures, striking a balance between public health and economic considerations [3, 199, 116]. Additionally, contact tracing is instrumental in predicting future outbreaks, identifying new disease clusters, and tracing the origin of infections [26, 79, 119, 120]. For a detailed analysis of contact tracing, refer to Publication III and IV.

Note that in our study, we treat the concepts of isolation and quarantine as being equivalent and thus use these terms interchangeably to reflect

their similar roles in our model. However, it's important to note that in the context of public health, isolation and quarantine are considered distinct approaches to controlling the spread of infectious diseases [153]. Isolation involves separating individuals diagnosed with a contagious disease from those who are healthy to prevent the transmission of the infection. On the other hand, quarantine refers to the separation and limitation of movement of people exposed to a contagious disease to monitor if they develop symptoms. It focuses on individuals who might be infected but are not yet confirmed to be sick. Therefore, in public health, isolation is applied to ill and contagious people. In contrast, quarantine is used for individuals at risk of falling ill due to their exposure to the disease [153].

3.9.1 Digital Contact Tracing

With the rise of affordable wearable health devices and mobile apps, digital contact tracing has become more precise and efficient, addressing the challenges of traditional manual tracing, such as slowness and reluctance to share contacts due to various concerns [174, 147, 182, 156, 146, 169, 141, 189]. These modern tools not only facilitate contact tracing but also provide real-time health data that can be used for other health strategies [8]. In this section, following Publication IV, we focus on digital contact tracing. It is the use of digital tools, often smartphone apps, to identify and notify individuals who have been in close proximity to someone diagnosed with a contagious disease, such as COVID-19. This technology aims to quickly and efficiently track potential disease exposure and prevent further spread.

We aim to understand how the size and threshold of the epidemic vary when apps are installed based on specific criteria like homophily (Sec. 2.4.4). In a manner akin to the homophily observed in people's vaccination status, as mentioned in Sec. 3.7, individuals using the app tend to have connections with other app users. In our representation, if one app user infects another, the infected person is likely to self-isolate, thereby halting further transmission. By adopting this method, we offer a cautious assessment of the influence of app-based contact tracing, especially when a significant portion of the population is infected at once.

Consider a network analogous to the one in the vaccination scenario. However, instead of individuals being vaccinated or not, they either use contact tracing apps or don't. These apps function effectively with a probability of f_{app}. Assuming that π_{a} proportion of individuals utilize the apps, we can derive a mean-field approximation for the epidemic size. This approximation is based on the branching process, similar to how it's done in Eq. 3.8 and 3.9;

$$I_{\text{n}}^{(t+1)} = R_0[\pi_{\text{nn}} I_{\text{n}}^{(t)} + \pi_{\text{an}} I_{\text{a}}^{(t)}], \tag{3.14}$$

$$I_{\mathrm{a}}^{(t+1)} = R_0[\pi_{\mathrm{na}} I_{\mathrm{n}}^{(t)} + \pi_{\mathrm{aa}}(1 - f_{\mathrm{app}}) I_{\mathrm{a}}^{(t)}]. \tag{3.15}$$

The probabilities $\pi_{..}$ indicate the likelihood of interactions within and across groups of app-users and non-app-users. Specifically, π_{aa} represents the probability of a connection between two app-users, π_{an} denotes the probability of a connection from an app-user to a non-app-user, π_{na} signifies the probability of a connection from a non-app-user to an app-user, and π_{nn} corresponds to the probability of a connection between two non-app-users.

We can again write these equations in a matrix form and find the spectral radius of the next-generation matrix to find the critical app coverage to curb the epidemic, similar to what we did for the herd immunity threshold. We skip this as the calculations would be very similar to what we did in the previous section. However, it is worth noting that when apps are working perfectly, $f_{\mathrm{app}} = 1$, for each value of $0 < \pi_{\mathrm{a}} < 1$, there is a non-trivial optimum value for homophily h^* that leads to the largest epidemic threshold such that:

$$h^* = \frac{2 - 2\pi_{\mathrm{a}}}{3\pi_{\mathrm{a}}^2 - 7\pi_{\mathrm{a}} + 4}. \tag{3.16}$$

This formula suggests that as the number of app users in a population increases, the optimal level of homophily also increases steadily. In practical terms, this means that to maximize the benefit of a digital contact tracing initiative, it's essential to distribute the apps so that the user group's homophily matches $h^*(\pi_{\mathrm{a}})$. So, a population with fewer app users will require a lower level of homophily to achieve the best results in terms of increasing the epidemic threshold. Moreover, for any π_{a}, the optimal homophily is larger than the homogenous mixing condition where $h = 0$.

In our Publication IV, we explore how effective contact tracing is by modifying the conventional approach to analyzing percolation and connectivity in contact networks. This involves applying our modified methodology to networks with different characteristics like degree distributions, user numbers of the application, and chances of quarantine breaches. Additionally, we consider populations with distinct social structures characterized by homophily and heterophily and the potential for targeting specific degrees in application distribution. Our findings are derived from a mix of direct simulations and mean-field analysis. They reveal significant variations in the size and probability of epidemics compared to standard SIR processes. Notably, the variation in connections within the network (degree heterogeneity) plays a crucial role in determining the epidemic threshold, though it has less impact on the size of the epidemic. The likelihood of tracing leading to quarantine is not as critical as the rate of application adoption. Moreover, we found that a strong preference for or against (homophily and heterophily, respectively) adopting the application can be harmful. In conclusion, the dynamics of epidemics are highly sensitive to all the tested parameters, highlighting the complex, multidimensional nature of

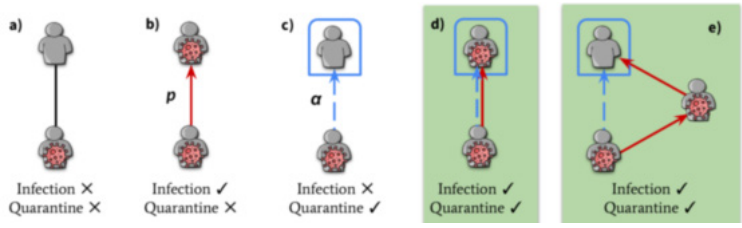

Figure 3.3. Illustration of contact tracing and transmission dynamics. The diagram shows scenarios without loops (a-d) and with local loops (e). Successful infections are indicated by solid red connections, while effective contact tracing is represented by dashed blue links. Following each exposure event, a susceptible node (S) either isolates itself with a probability of α or becomes infected with a probability of p, with these outcomes being independent. In scenarios without loops, the interplay of infections and contact tracing can be simplified to a single link, leading to four distinct outcomes: (a) no event occurs, (b) the infection is transmitted to the adjacent node, but contact tracing is unsuccessful, (c) the infection does not transmit, yet contact tracing is effective, or (d) both the transmission of the infection and the success of contact tracing occur. The final scenario, marked with a green background, demonstrates the advantage of contact tracing in interrupting indirect transmission routes, especially in clustered networks. In scenario (e), which involves local loops, the situation depicted in panel (c) becomes advantageous. This is because the quarantine, implemented near the site of infection, can hinder the spread of the infection to the neighboring node via a local loop. Essentially, timely quarantine in the vicinity of an infection source can prevent the disease from reaching the neighbor through these looped pathways.

estimating the impact of digital contact tracing.

3.9.2 Contact Tracing and Social Groups

In our previous examples, we assumed that the transmission networks locally look like trees. That is the main reason why we could use simple branching process arguments when we were interested in the phenomena around the epidemic threshold. In Publication IV Real-world social networks deviate from simple tree-like structures [133, 107], especially regarding disease spread, as we discussed in Sec. 2.2.2. This can be seen in densely connected groups like families or workplaces. Clique structures play a crucial role in the spread of behaviors in complex contagion processes where prior exposures increase adoption chances. This is different from disease spread, where infections are independent of past exposures.

As we see in Publication III, disease spread combined with contact tracing acts similarly to behavior spread, a.k.a complex contagion in networks, with prior exposures reducing subsequent infections. This highlights the significance of group structures in contact tracing.

To explore the dynamics of contact tracing on outbreaks based on a model that emphasizes group interactions, we use Sec. 2.4.2 networks with equal connections per node but varying group sizes. When someone in a social group gets infected, they infect some members, and contact tracing isolates others. Effective contact tracing targets those both infected and

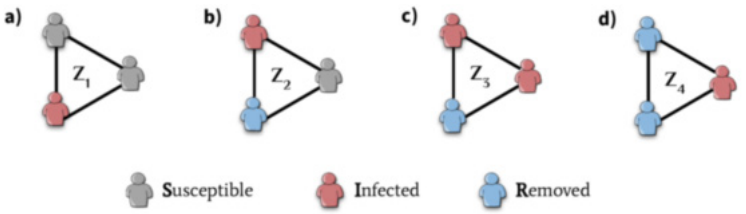

Figure 3.4. Every 3-clique is capable of exhibiting one of four distinct life stages or diffusion patterns, each including at least a single infected node. In our model, nodes that are either recovered or quarantined are grouped into the R compartment. Within a 6-regular 3-clique network, a node designated as Z_1 has the potential to form a Z_2 motif alongside two other Z_1 nodes, as well as a Z_3 motif when combined with four Z_1 nodes. Furthermore, a node identified as Z_2 is capable of creating a Z_4 motif in conjunction with two Z_1 nodes. It's important to note that nodes within the Z_1, Z_2, and Z_4 motifs have the possibility of transitioning to infection-eliminated states, such as R,S,S, which are not depicted here. This figure is from Publication III.

isolated, breaking transmission chains. However, even isolating uninfected individuals can help control the spread. If contact tracing isn't perfect, isolating members can stop further infections. We will see that group structures boost contact tracing effectiveness in Publication III. Specifically, contact tracing in networks with cliques impacts the spread more than in tree-like structures.

Let us continue with a basic mathematical model that uses a discrete-time SIR approach to represent disease dynamics. In this model, infected individuals spread the disease to nearby susceptible ones based on a **transmission probability** p. These infected individuals then move to a recovered state.

Contact tracing methods vary, such as using phone apps or manual tracing, and their effectiveness depends on factors like recall of contacts, delays in tracing, app adoption, and adherence to isolation guidelines [208]. In Publication III, these complexities are represented by an **isolation probability** α, which indicates the chance a nearby node successfully isolates to prevent further infections. Infected nodes can move susceptible neighboring nodes to a quarantine compartment based on this probability. This compartment can contain both infected and susceptible individuals, as depicted in Fig. 3.3 when $\alpha = 0$ the model reduces to a simple SIR process.

The infection and contact tracing processes are considered independent in our model. The order of these processes in the model slightly affects the epidemic size but not the epidemic threshold. We compute the epidemic size by evaluating each infected-susceptible link, considering both the disease spread and contact tracing. The total size of an epidemic is determined by adding the number of infected individuals, both in and out of quarantine.

Using a multi-type branching process [111] to represent our SIR+Q model, we can derive the relationship between p, α, and clique size, c, concerning the epidemic threshold. In this model, we identify different

clique motifs that represent potential states of susceptibles, infected, or recovered nodes in any clique. These motifs are denoted by Z_i, as seen in Fig. 3.4 for a 3-node clique example.

i,j	m_{ij}
1,1	$4p(1-\alpha)$
1,2	$2p(1-\alpha)$
2,1	$2p(1-\alpha)^2(1-p)$
3,1	$p^2(1-\alpha)^2$
4,1	$2\alpha p(1-\alpha)$
4,2	$p(1-\alpha)$

Table 3.1. Non-zero elements of the next-generation matrix $\mathbf{M}_{4\times4}$ for a 3-clique network. m_{ij} gives the expected number of Z_i cliques from a Z_j clique, as shown in Fig. 3.4. This table is from Publication III.

Fig. 3.4 illustrates the four stages of a 3-clique's life, where isolated and recovered nodes are combined into one compartment (R) and Table 3.1 provides non-zero elements of $\mathbf{M}$, detailing how motifs transition. The matrix $\mathbf{M}$ represents the transitions between these motifs. For example, the infected node in Z_1 can infect one or two neighbors, corresponding to motifs Z_2 and Z_3. Further, n_c-1 new Z_1 motifs are produced every time such an infection takes place. That is, when Z_1 turns into Z_2 there are also n_c-1 new Z_1 motifs, and when it turns into Z_2 there are $2(n_c-1)$ new Z_1 motifs created.

It is always feasible to compute the average expected number of new infections across all potential types of infection, utilizing our multi-type branching process. This calculation is based on the next-generation matrix, denoted as $\mathbf{M}$, which is also known as the mean or population projection matrix. The epidemic threshold in any clique network can be identified by finding values for p and α where $\rho(\mathbf{M})=1$.

Fig. 3.5a presents an epidemic phase diagram for networks with cliques. The curves within this diagram separate it into sub- and super-critical regions, determining the potential of an outbreak. Networks with larger cliques have a more significant sub-critical region and a smaller super-critical one. As the efficacy of contact tracing grows, the distinction between critical p values widens.

This model demonstrates the enhanced efficiency of contact tracing in clustered networks compared to tree-like structures or fully mixed populations. Specifically, when contact tracing is less reliable (for lower α values), increasing clique sizes notably reduces the outbreak size. This suggests that standard epidemic models might undervalue the efficacy of contact tracing in networks with more clustering.

Our study, along with Publication III, simplifies disease transmission

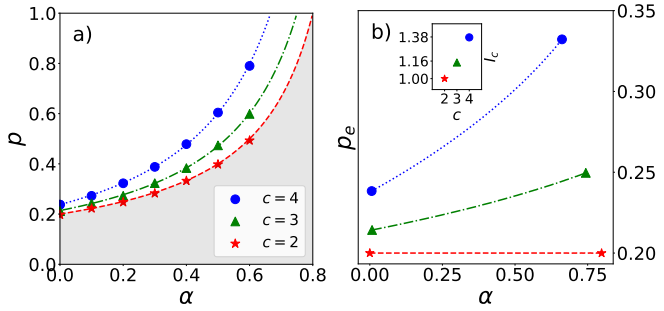

Figure 3.5. Phase diagram showing that increasing the clique size increases the epidemic threshold and effectiveness of contact tracing. (a) The critical curves where $R_e = 1$ in the αp-plane for $c \in \{2,3,4\}$. The shaded area is the sub-critical region for $c = 2$ where the infection eventually dies out after a finite number of generations for any clique size. (b) The same phase diagram in αp_e-plane, where $p_e = p(1-\alpha)$ is the effective transmission probability. The inset in panel b shows the relative maximum increase in the effective epidemic threshold for different networks with cliques. Each point in the inset is the ratio of the p_e values at the endpoints of each curve outside the inset, such that $I_c = p_e(\alpha_{max})/p_e(\alpha_{min})$. This figure is from Publication III.

and contact tracing modeling but highlights the importance of realistic social structures. Key findings include the efficiency of contact tracing in clustered networks and the impact of group sizes and contact tracing efficiency on outbreak control. Our model, which contrasts with traditional fully mixed or tree-like network assumptions, shows greater effectiveness in networks with clustering. It also reveals the importance of considering group structures in contact tracing, akin to complex contagion models where previous exposures affect adoption/infection chances. This is particularly beneficial in early disease spread stages, helping to limit infection paths. Our model also indicates that ignoring group structures in contact tracing can lead to an oversimplified understanding of infection risk and thresholds. We used a simplified dynamical model, integrating various factors into a single parameter, which may not fully capture the nuances of real-world contact tracing and disease spread.

Future research could incorporate more sophisticated SEIR models, but our focus was on the critical transition from disease-free to endemic states. The SIR model was chosen for its broad insights into epidemic outcomes, allowing us to examine the impact of clique structures and contact tracing on simple disease spread. In essence, our findings emphasize the need to consider realistic social network structures in epidemic modeling, showing that contact tracing is more efficient in clustered networks and that group size and tracing efficacy significantly influence epidemic control.

4. Conclusion

The COVID-19 pandemic has underscored the critical importance of comprehending epidemic dynamics, particularly in anticipation of future pandemics. As we discussed earlier and will explore in upcoming publications, the structure of social networks plays a pivotal role in the spread and intensity of epidemics, necessitating a deeper understanding of how various diseases propagate in society and the effectiveness of interventions within these networked structures. My research has focused on developing theoretical and computational approaches to gain insights into epidemic spreading in the presence of realistic social network structures, particularly in the context of vaccination and contact tracing, and how these interventions can be optimized given the networked nature of human populations.

The topology of contact networks significantly influences the dynamics of spreading phenomena. Group structures within these networks can either contain or facilitate the spread of information or diseases. A key element in these structures is homophily, the tendency of similar individuals to connect, which can create areas of increased vulnerability or resistance within the network. This phenomenon is particularly relevant in health behaviors like adopting digital contact tracing apps or taking vaccines. Traditional infectious disease models fall short of capturing the complexities of human interactions. Per contra, network-based frameworks give us a more detailed understanding.

The heterogeneity in human interactions also plays a crucial role when immunity arises from natural infection. Diseases initially target highly connected individuals, which can lead to efficient immunization of these nodes but also create localized weaknesses within the network. The interaction between network structure and epidemic dynamics, particularly the effects of network communities and higher-order interactions, remains a vital area of research. The impact of contact network structural and spatial properties on the threshold and robustness of herd immunity remains an area of exploration. In this thesis, I began to address these challenges by examining the effects of spatial structure on herd immunity

in various network models. Future research should extend these findings to networks with greater heterogeneity, such as scale-free networks. Non-pharmaceutical interventions like contact tracing are essential in controlling outbreaks, especially before vaccines become widely available. My research has delved into the dynamics of contact tracing within social groups and its influence on outbreak size and epidemic thresholds. Integrating real-world complexities, such as tracing delays and adherence to isolation protocols into models, remains a significant challenge.

This thesis delved into the temporal dynamics of spreading processes in networks, establishing a crucial link between temporal network reachability and percolation theory. It reveals that when constrained by limited waiting times, reachability exhibits a phase transition characteristic of directed percolation, suggesting a new approach to analyzing spreading behaviors. The research aims to uncover universal dynamics in spreading processes and reachability in complex systems, enhancing our understanding and aiding decision-making in policy and engineering. By integrating temporal network analysis with non-equilibrium statistical mechanics, the work identifies key measurable aspects of reachability, establishes their connection to directed percolation parameters, and demonstrates their alignment with directed percolation scaling behaviors. However, this is just the beginning, as further research is needed to refine methods for studying temporal network spreading and to observe these phenomena in empirical networks, thereby deepening our understanding of complex systems. A deeper understanding of reachability and spreading processes will be instrumental in navigating and comprehending complex systems and challenges.

As I draw to a close, I want to impart two profound insights that have guided me.

Modeling an epidemic is not rocket science, it's harder!

Nigel Goldenfeld, *Pritzker's COVID-19 press conference*

> The most terrifying fact about the universe is not that it is hostile but that it is indifferent; but if we can come to terms with this indifference and accept the challenges of life within the boundaries of death—however mutable man may be able to make them—our existence as a species can have genuine meaning and fulfillment. However vast the darkness, we must supply our own light.

Stanley Kubrick, *Interviews (2001)*

References

[1] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

[2] Mohammed N Alenezi, Fawaz S Al-Anzi, Haneen Alabdulrazzaq, Ammar Alhusaini, and Abdullah F Al-Anzi. A study on the efficiency of the estimation models of covid-19. *Results in Physics*, 26:104370, 2021.

[3] Alberto Aleta, David Martin-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E Dean, M Elizabeth Halloran, Ira M Longini Jr, Stefano Merler, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of covid-19. *Nature Human Behaviour*, 4(9):964–971, 2020.

[4] Amir A Aliabadi, Steven N Rogak, Karen H Bartlett, Sheldon I Green, et al. Preventing airborne disease transmission: review of methods for ventilation design in health care facilities. *Advances in preventive medicine*, 2011, 2011.

[5] Philip W Anderson. *Basic notions of condensed matter physics*. CRC Press, 2018.

[6] Alberto Antonioni and Marco Tomassini. Degree correlations in random geometric graphs. *Physical Review E*, 86(3):037101, 2012.

[7] Elisha B Are, Kiffer G Card, and Caroline Colijn. The role of vaccine status homophily in the covid-19 pandemic: A cross-sectional survey with modeling. *medRxiv*, pages 2023–06, 2023.

[8] H Ceren Ates, Ali K Yetisen, Firat Güder, and Can Dincer. Wearable devices for the detection of covid-19. *Nature Electronics*, 4(1):13–14, 2021.

[9] Binish Ather, Taaha M Mirza, and Peter F Edemekong. Airborne precautions. 2018.

[10] Krishna B. Athreya and Peter E. Ney. *Branching Processes*. Springer-Verlag, New York, 1972.

[11] Arash Badie-Modiri. Error and attack tolerance of public transportation networks: A temporal networks approach. Master's thesis, Aalto University School of Science, 2018. URL: http://urn.fi/URN:NBN:fi:aalto-201806293772.

[12] Arash Badie-Modiri, Abbas K. Rizi, Márton Karsai, and Mikko Kivelä. Directed percolation in random temporal network models with heterogeneities. *Physical Review E*, 105(5):054313, 2022.

[13] Arash Badie-Modiri, Abbas K. Rizi, Márton Karsai, and Mikko Kivelä. Directed percolation in temporal networks. *Physical Review Research*, 4(2):L022047, 2022.

[14] Arash Badie-Modiri, Márton Karsai, and Mikko Kivelä. Efficient limited-time reachability estimation in temporal networks. *Physical Review E*, 101(5):052303, 2020.

[15] Paolo Bajardi, Alain Barrat, Fabrizio Natale, Lara Savini, and Vittoria Colizza. Dynamical patterns of cattle trade movements. *PloS one*, 6(5):e19869, 2011. doi:10.1371/journal.pone.0019869.

[16] Per Bak, Kim Christensen, Leon Danon, and Tim Scanlon. Unified scaling law for earthquakes. *Physical Review Letters*, 88(17):178501, 2002. doi:10.1103/PhysRevLett.88.178501.

[17] Laleh Bakhtiar et al. *The canon of medicine*, volume 1. Kazi Publications, 1999.

[18] Albert-Laszlo Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005. doi:10.1038/nature03459.

[19] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge university press, 2008.

[20] Alain Barrat, Ciro Cattuto, Mikko Kivelä, Sune Lehmann, and Jari Saramäki. Effect of manual and digital contact tracing on covid-19 outbreaks: A study on empirical contact data. *Journal of the Royal Society Interface*, 18(178):20201000, 2021.

[21] Marc Barthélemy. Spatial networks. *Physics reports*, 499(1-3):1–101, 2011.

[22] Marc Barthelemy. *Spatial Networks: A Complete Introduction: From Graph Theory and Statistical Physics to Real-World Applications*. Springer Nature, 2022.

[23] Rodney J Baxter. *Exactly solved models in statistical mechanics*. Elsevier, 2016.

[24] Peter S Bearman, James Moody, and Katherine Stovel. Chains of affection: The structure of adolescent romantic and sexual networks. *American journal of sociology*, 110(1):44–91, 2004. doi:10.1086/386272.

[25] John M Beggs and Dietmar Plenz. Neuronal avalanches in neocortical circuits. *Journal of neuroscience*, 23(35):11167–11177, 2003. doi:10.1523/JNEUROSCI.23-35-11167.2003.

[26] James Bell, Ginestra Bianconi, David Butler, Jon Crowcroft, Paul CW Davies, Chris Hicks, Hyunju Kim, Istvan Z Kiss, Francesco Di Lauro, Carsten Maple, et al. Beyond covid-19: network science and sustainable exit strategies. *Journal of Physics: Complexity*, 2(2):021001, 2021.

[27] Md Arif Billah, Md Mamun Miah, and Md Nuruzzaman Khan. Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence. *PloS one*, 15(11):e0242128, 2020.

[28] Béla Bollobás. The diameter of random graphs. *Transactions of the American Mathematical Society*, 267(1):41–52, 1981.

[29] Béla Bollobás. Random graphs. In *Modern graph theory*, pages 215–252. Springer, 1998. doi:10.1007/978-1-4612-0619-4_7.

[30] Lera Boroditsky and Michael Ramscar. First, we assume a spherical cow... *Behavioral and Brain Sciences*, 24(4):656–657, 2001.

[31] Pierre-Yves Boëlle and Eugenio Valdano. The importance of increasing primary vaccinations against covid-19 in europe. *Infectious Disease Modelling*, 9(1):1–9, 2024. URL: https://www.sciencedirect.com/science/article/pii/S2468042723000982, doi:https://doi.org/10.1016/j.idm.2023.11.008.

[32] Dan Braha and Yaneer Bar-Yam. Time-dependent complex networks: Dynamic centrality, dynamic motifs, and cycles of social interactions. In *Adaptive Networks*, pages 39–50. Springer, 2009. doi:10.1007/978-3-642-01284-6_3.

[33] Aaron Bramson and Benjamin Vandermarliere. Dynamical properties of interaction data. *Journal of Complex Networks*, 4(1):87–114, 2016. doi:10.1093/comnet/cnv009.

[34] Allan M Brandt. The history of contact tracing and the future of public health, 2022.

[35] Tom Britton, Frank Ball, and Pieter Trapman. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science*, 369(6505):846–849, August 2020.

[36] Anna D Broido and Aaron Clauset. Scale-free networks are rare. *Nature communications*, 10(1):1017, 2019.

[37] Andrew F Brouwer. Why the spectral radius? an intuition-building introduction to the basic reproduction number. *Bulletin of Mathematical Biology*, 84(9):96, 2022.

[38] Giulio Burgio, Benjamin Steinegger, and Alex Arenas. Homophily impacts the success of vaccine roll-outs. *Communications Physics*, 5(1):70, 2022.

[39] John Cardy. *Scaling and renormalization in statistical physics*, volume 5. Cambridge university press, 1996.

[40] Carlos W Castillo-Garsow and Carlos Castillo-Chavez. A tour of the basic reproductive number and the next generation of researchers. *An Introduction to Undergraduate Research in Computational and Mathematical Biology: From Birdsongs to Viscosities*, pages 87–124, 2020.

[41] Shi Chen and Cristina Lanzas. Distinction and connection between contact network, social network, and disease transmission network. *Preventive veterinary medicine*, 131:8–11, 2016.

[42] Zesheng Chen. Discrete-time vs. continuous-time epidemic models in networks. *IEEE Access*, 7:127669–127677, 2019.

[43] Jonathan R Chubb, Tatjana Trcek, Shailesh M Shenoy, and Robert H Singer. Transcriptional pulsing of a developmental gene. *Current biology*, 16(10):1018–1025, 2006. doi:10.1016/j.cub.2006.03.092.

[44] Fan Chung and Linyuan Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, 2002.

[45] Fan Chung and Linyuan Lu. Connected components in random graphs with given expected degree sequences. *Annals of combinatorics*, 6(2):125–145, 2002. doi:10.1007/PL00012580.

[46] Giulio Cimini, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and Guido Caldarelli. The statistical physics of real-world networks. *Nature Reviews Physics*, 1(1):58–71, 2019.

[47] Reuven Cohen and Shlomo Havlin. Percolation in complex networks. *Complex Media and Percolation Theory*, pages 419–431, 2021.

[48] Reuven Cohen, Shlomo Havlin, and Daniel Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Physical review letters*, 91(24):247901, 2003.

[49] James S Coleman. Relational Analysis: The Study of Social Organizations with Survey Methods. *Human Organization*, 17:28–36, 1958. doi:10.17730/humo.17.4.q5604m676260q8n7.

[50] Ian Cooper, Argha Mondal, and Chris G Antonopoulos. A sir model assumption for the spread of covid-19 in different communities. *Chaos, Solitons & Fractals*, 139:110057, 2020.

[51] Alvaro Corral. Local distributions and rate fluctuations in a unified scaling law for earthquakes. *Physical Review E*, 68(3):035102, 2003. doi:10.1103/PhysRevE.68.035102.

[52] Leonardo Dalla Porta and Mauro Copelli. Modeling neuronal avalanches and long-range temporal correlations at the emergence of collective oscillations: Continuously varying exponents mimic m/eeg results. *PLoS computational biology*, 15(4):e1006924, 2019. doi:10.1371/journal.pcbi.1006924.

[53] Leon Danon, Ashley P Ford, Thomas House, Chris P Jewell, Matt J Keeling, Gareth O Roberts, Joshua V Ross, Matthew C Vernon, et al. Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011, 2011.

[54] Paul L Delamater, Erica J Street, Timothy F Leslie, Y Tony Yang, and Kathryn H Jacobsen. Complexity of the basic reproduction number (r0). *Emerging infectious diseases*, 25(1):1, 2019.

[55] Odo Diekmann, Hans Heesterbeek, and Tom Britton. *Mathematical tools for understanding infectious disease dynamics*, volume 7. Princeton University Press, 2013.

[56] Odo Diekmann, JAP Heesterbeek, and Michael G Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the royal society interface*, 7(47):873–885, 2010.

[57] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. On the definition and the computation of the basic reproduction ratio r 0 in models for infectious diseases in heterogeneous populations. *Journal of mathematical biology*, 28:365–382, 1990.

[58] Reinhard Diestel. Graduate texts in mathematics: Graph theory. 2018.

[59] Klaus Dietz. The estimation of the basic reproduction number for infectious diseases. *Statistical methods in medical research*, 2(1):23–41, 1993.

[60] Klaus Dietz and JAP Heesterbeek. Daniel bernoulli's epidemiological model revisited. *Mathematical biosciences*, 180(1-2):1–21, 2002.

[61] Aolan Ding and Rick Durrett. The phase transition in chung-lu graphs. *arXiv:2305.08203*, 2023.

[62] Cyril Domb. *Phase transitions and critical phenomena*. Elsevier, 2000.

[63] Sergey N Dorogovtsev, Alexander V Goltsev, and José FF Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics*, 80(4):1275, 2008.

[64] Sergey N Dorogovtsev and José FF Mendes. *The nature of complex networks*. Oxford University Press, 2022.

[65] Freeman Dyson et al. A meeting with enrico fermi. *Nature*, 427(6972):297–297, 2004.

[66] Young-Ho Eom and Hang-Hyun Jo. Generalized friendship paradox in complex networks: The case of scientific collaboration. *Scientific reports*, 4(1):1–6, 2014.

[67] Byron D Erath and Andrea R Ferro. Infectious disease transmission from bioaerosols. *Journal of Exposure Science & Environmental Epidemiology*, 32(5):645–646, 2022.

[68] P ERDdS and A R&wi. On random graphs i. *Publ. math. debrecen*, 6(290-297):18, 1959.

[69] Péter Érdi and János Tóth. *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*. Manchester University Press, 1989.

[70] C. P. Farrington. On vaccine efficacy and reproduction numbers. *Mathematical Biosciences*, 185:89–109, 2003. URL: https://www.sciencedirect.com/science/article/pii/S0025556403000610, doi:https://doi.org/10.1016/S0025-5564(03)00061-0.

[71] NH Fefferman and KL2365573 Ng. How disease models in static networks can fail to approximate disease in dynamic networks. *Physical Review E*, 76(3):031919, 2007. doi:10.1103/PhysRevE.76.031919.

[72] Scott L Feld. Why your friends have more friends than you do. *American journal of sociology*, 96(6):1464–1477, 1991. doi:10.1086/229693.

[73] Finnish Institute for Health and Welfare. Influenza vaccine, 2023. Accessed on [Insert Access Date Here]. URL: https://thl.fi/en/web/infectious-diseases-and-vaccinations/vaccines-a-to-z/influenza-vaccine.

[74] Matthias Flückiger and Markus Ludwig. Spatial networks and the spread of covid-19: results and policy implications from germany. *Review of Regional Research*, pages 1–27, 2023.

[75] B Fosdick, D Larremore, J Nishumura, and J Ugander. Configuring random graph models with fixed degree sequences. preprint, 2016.

[76] Sylvain Gandon and Sébastien Lion. Targeted vaccination and the speed of sars-cov-2 adaptation. *Proceedings of the National Academy of Sciences*, 119(3):e2110666119, 2022.

[77] Laetitia Gauvin, Mathieu Génois, Márton Karsai, Mikko Kivelä, Taro Takaguchi, Eugenio Valdano, and Christian L. Vestergaard. Randomized reference models for temporal networks. *SIAM Review*, 64(4):763–830, 2022. doi:10.1137/19M1242252.

[78] Mohammad Ghaderi. Public health interventions in the face of pandemics: Network structure, social distancing, and heterogeneity. *European Journal of Operational Research*, 298(3):1016–1031, 2022.

[79] Marius Gilbert, Mathias Dewatripont, Eric Muraille, Jean-Philippe Platteau, and Michel Goldman. Preparing for a responsible lockdown exit strategy. *Nature medicine*, 26(5):643–644, 2020.

[80] Julia R Gog and T Déirdre Hollingsworth. Epidemic interventions: insights from classic results. *Philosophical Transactions of the Royal Society B*, 376(1829):20200263, 2021.

[81] Nigel Goldenfeld. *Lectures on phase transitions and the renormalization group*. CRC Press, 2018.

[82] Ido Golding, Johan Paulsson, Scott M Zawilski, and Edward C Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6):1025–1036, 2005. doi:10.1016/j.cell.2005.09.031.

[83] Peter Grassberger. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, 63(2):157–172, 1983.

[84] John Graunt and Walter Francis Willcox. Natural and political observations made upon the bills of mortality. *(No Title)*, 1939.

[85] Trisha Greenhalgh, Jose L Jimenez, Kimberly A Prather, Zeynep Tufekci, David Fisman, and Robert Schooley. Ten scientific reasons in support of airborne transmission of sars-cov-2. *The lancet*, 397(10285):1603–1605, 2021.

[86] Geoffrey Grimmett and Geoffrey Grimmett. Near the critical point: Scaling theory. *Percolation*, pages 232–253, 1999.

[87] Bnaya Gross and Shlomo Havlin. Epidemic spreading and control strategies in spatial modular network. *Applied network science*, 5(1):1–14, 2020.

[88] M. Elizabeth Halloran, Ira M. Longini, Jr., and Claudio J. Struchiner. Design and Interpretation of Vaccine Field Studies. *Epidemiologic Reviews*, 21:73–88, 03 1999. doi:10.1093/oxfordjournals.epirev.a017990.

[89] Tiberiu Harko, Francisco SN Lobo, and MK3197716 Mak. Exact analytical solutions of the susceptible-infected-recovered (sir) epidemic model and of the sir model with equal death and birth rates. *Applied Mathematics and Computation*, 236:184–194, 2014.

[90] Malte Henkel, Haye Hinrichsen, Sven Lübeck, and Michel Pleimling. *Non-equilibrium phase transitions*, volume 1. Springer, 2008. doi:10.1007/978-1-4020-8765-3.

[91] Haye Hinrichsen. Non-equilibrium critical phenomena and phase transitions into absorbing states. *Advances in physics*, 49(7):815–958, 2000. doi:10.1080/00018730050198152.

[92] Takayuki Hiraoka, Abbas K. Rizi, Zahra Ghadiri, Mikko Kivelä, and Jari Saramäki. The strength and weakness of disease-induced herd immunity. *arXiv:2307.04700*.

[93] Takayuki Hiraoka, Abbas K. Rizi, Mikko Kivelä, and Jari Saramäki. Herd immunity and epidemic size in networks with vaccination homophily. *Physical Review E*, 105(5):L052301, 2022. doi:10.1103/PhysRevE.105.L052301.

[94] Elizabeth A Hobson, Matthew J Silk, Nina H Fefferman, Daniel B Larremore, Puck Rombach, Saray Shai, and Noa Pinter-Wollman. A guide to choosing and implementing reference models for social network analysis. *Biological Reviews*, 96(6):2716–2734, 2021.

[95] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[96] Petter Holme. Network reachability of real-world contact sequences. *Physical Review E*, 71(4):046119, 2005. doi:10.1103/PhysRevE.71.046119.

[97] Petter Holme. Modern temporal network theory: a colloquium. *The European Physical Journal B*, 88(9):1–30, 2015. doi:10.1140/epjb/e2015-60657-4.

[98] Petter Holme. Rare and everywhere: Perspectives on scale-free networks. *Nature communications*, 10(1):1016, 2019.

[99] Petter Holme. Fast and principled simulations of the sir model on temporal networks. *Plos one*, 16(2):e0246961, 2021. doi:10.1371/journal.pone.0246961.

[100] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012. doi:10.1016/j.physrep.2012.03.001.

[101] Petter Holme and Jari Saramäki. A map of approaches to temporal networks. In *Temporal Network Theory*, pages 1–24. Springer, 2019. doi:10.1007/978-3-030-23495-9_1.

[102] Li-Yang Hsu, Cheng-Chuan Lee, Justin A Green, Brenda Ang, Nicholas I Paton, Lawrence Lee, Jorge S Villacian, Poh-Lian Lim, Arul Earnest, and Yee-Sin Leo. Severe acute respiratory syndrome (sars) in singapore: clinical features of index patient and initial contacts. *Emerging infectious diseases*, 9(6):713, 2003.

[103] Felix Joos, Guillem Perarnau, Dieter Rautenbach, and Bruce Reed. How to determine if a random graph with a fixed degree sequence has a giant component. *Probability Theory and Related Fields*, 170(1-2):263–310, 2018.

[104] Alexander Jung and Yasmin SarcheshmehPour. Local graph clustering with network lasso. *IEEE Signal Processing Letters*, 28:106–110, 2020.

[105] Abbas K. Rizi, Ali Faqeeh, Arash Badie-Modiri, and Mikko Kivelä. Epidemic spreading and digital contact tracing: Effects of heterogeneous mixing and quarantine failures. *Physical Review E*, 105(4):044313, 2022.

[106] Rebecca Kahn, Stephanie J Schrag, Jennifer R Verani, and Marc Lipsitch. Identifying and alleviating bias due to differential depletion of susceptible people in postmarketing evaluations of covid-19 vaccines. *American journal of epidemiology*, 191(5):800–811, 2022.

[107] Brian Karrer and Mark EJ Newman. Random graphs containing arbitrary distributions of subgraphs. *Physical Review E*, 82(6):066118, 2010.

[108] Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical review E*, 83(1):016107, 2011.

[109] Márton Karsai, Hang-Hyun Jo, Kimmo Kaski, et al. *Bursty human dynamics*. Springer, 2018.

[110] Márton Karsai, Mikko Kivelä, Raj Kumar Pan, Kimmo Kaski, János Kertész, Albert-László Barabási, and Jari Saramäki. Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E*, 83(2):025102, 2011. doi:10.1103/PhysRevE.83.025102.

[111] Leah A Keating, James P Gleeson, and David JP O'Sullivan. Multitype branching process method for modeling complex contagion on clustered networks. *Physical Review E*, 105(3):034306, 2022.

[112] Matt J Keeling and Ken TD Eames. Networks and epidemic models. *Journal of the royal society interface*, 2(4):295–307, 2005.

[113] Eben Kenah, Joel C Miller, et al. Epidemic percolation networks, epidemic outcomes, and interventions. *Interdisciplinary perspectives on infectious diseases*, 2011, 2011.

[114] Eben Kenah and James M Robins. Second look at the spread of epidemics on networks. *Physical Review E*, 76(3):036113, 2007. doi:10.1103/PhysRevE.76.036113.

[115] István Z Kiss, Joel C Miller, Péter L Simon, et al. Mathematics of epidemics on networks. *Cham: Springer*, 598:31, 2017.

[116] Stephen M Kissler, Christine Tedijanto, Edward Goldstein, Yonatan H Grad, and Marc Lipsitch. Projecting the transmission dynamics of sars-cov-2 through the postpandemic period. *Science*, 368(6493):860–868, 2020.

[117] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014. doi:10.1093/comnet/cnu016.

[118] Mikko Kivelä, Jordan Cambe, Jari Saramäki, and Márton Karsai. Mapping temporal-network percolation to weighted, static event graphs. *Scientific reports*, 8(1):1–9, 2018. doi:10.1038/s41598-018-29577-2.

[119] Nicole E. Kogan, Leonardo Clemente, Parker Liautaud, Justin Kaashoek, Nicholas B. Link, Andre T. Nguyen, Fred S. Lu, Peter Huybers, Bernd Resch, Clemens Havas, Andreas Petutschnig, Jessica Davis, Matteo Chinazzi, Backtosch Mustafa, William P. Hanage, Alessandro Vespignani, and Mauricio Santillana. An early warning approach to monitor covid-19 activity with multiple digital traces in near real time. *Science Advances*, 7(10), 2021. URL: https://advances.sciencemag.org/content/7/10/eabd6989, doi:10.1126/sciadv.abd6989.

[120] Sadamori Kojaku, Laurent Hébert-Dufresne, Enys Mones, Sune Lehmann, and Yong-Yeol Ahn. The effectiveness of backward contact tracing in networks. *Nature Physics*, pages 1–7, 2021.

[121] Gautier Krings, Márton Karsai, Sebastian Bernhardsson, Vincent D Blondel, and Jari Saramäki. Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science*, 1(1):1–16, 2012. doi:10.1140/epjds4.

[122] Ninaad Lasrado and Dan H Barouch. Sars-cov-2 hybrid immunity: The best of both worlds, 2023.

[123] Ignazio Lazzizzera. An analytic approximate solution of the sir model. *arXiv preprint arXiv:2011.07494*, 2020.

[124] Melissa Leach, Hayley MacGregor, Ian Scoones, and Annie Wilkinson. Post-pandemic transformations: How and why covid-19 requires us to rethink development. *World development*, 138:105233, 2021.

[125] Clement Lee and Darren J Wilkinson. A review of stochastic block models and extensions for graph clustering. *Applied Network Science*, 4(1):1–50, 2019.

[126] Grégoire Lemoult, Liang Shi, Kerstin Avila, Shreyas V Jalikop, Marc Avila, and Björn Hof. Directed percolation phase transition to sustained turbulence in couette flow. *Nature Physics*, 12(3):254–258, 2016. doi:10.1038/nphys3675.

[127] Nancy HL Leung. Transmissibility and transmission of respiratory viruses. *Nature Reviews Microbiology*, 19(8):528–545, 2021.

[128] Ming Li, Run-Ran Liu, Linyuan Lü, Mao-Bin Hu, Shuqi Xu, and Yi-Cheng Zhang. Percolation on complex networks: Theory and application. *Physics Reports*, 907:1–68, 2021.

[129] Shuping Li, Xiaorong Zhao, and Ruixia Zhang. Site-bond percolation model of epidemic spreading with vaccination in complex networks. *Journal of Mathematical Biology*, 85(5):49, 2022.

[130] Marc Lipsitch and Rebecca Kahn. Interpreting vaccine efficacy trial results for infection and transmission. *Vaccine*, 39(30):4082–4088, 2021.

[131] Su-Yu Liu, Andrea Baronchelli, and Nicola Perra. Contagion dynamics in time-varying metapopulation networks. *Physical Review E*, 87(3):032805, 2013. doi:10.1103/PhysRevE.87.032805.

[132] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

[133] R Duncan Luce and Albert D Perry. A method of matrix analysis of group structure. *Psychometrika*, 14(2):95–116, 1949.

[134] Naoki Masuda and Petter Holme. Predicting and controlling infectious disease epidemics using temporal networks. *F1000prime reports*, 5, 2013. doi:10.12703/P5-6.

[135] Auni Aslah Mat Daud. Five common misconceptions regarding flattening-the-curve of covid-19. *History and Philosophy of the Life Sciences*, 44(3):41, 2022.

[136] Laura Matrajt and Tiffany Leung. Evaluating the effectiveness of social distancing interventions to delay or flatten the epidemic curve of coronavirus disease. *Emerging infectious diseases*, 26(8):1740, 2020.

[137] Andrew Mellor. The temporal event graph. *Journal of Complex Networks*, 6(4):639–659, 2018. doi:10.1093/comnet/cnx048.

[138] Andrew Mellor. Event graphs: Advances and applications of second-order time-unfolded temporal network models. *Advances in Complex Systems*, 22(03):1950006, 2019. doi:10.1142/S0219525919500061.

[139] Sergey Melnik, Adam Hackett, Mason A Porter, Peter J Mucha, and James P Gleeson. The unreasonable effectiveness of tree-based theory for networks with clustering. *Physical Review E*, 83(3):036112, 2011.

[140] Joel C Miller. Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. *Physical Review E*, 76(1):010101, 2007. doi:10.1103/PhysRevE.76.010101.

[141] Tejaswini Mishra, Meng Wang, Ahmed A Metwally, Gireesh K Bogu, Andrew W Brooks, Amir Bahmani, Arash Alavi, Alessandra Celli, Emily Higgs, Orit Dagan-Rosenfeld, et al. Pre-symptomatic detection of covid-19 from smartwatch data. *Nature Biomedical Engineering*, 4(12):1208–1220, 2020.

[142] Jeta Molla, Alejandro Ponce de León Chávez, Takayuki Hiraoka, Tapio Ala-Nissila, Mikko Kivelä, and Lasse Leskelä. Adaptive and optimized covid-19 vaccination strategies across geographical regions and age groups. *PLoS computational biology*, 18(4):e1009974, 2022.

[143] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180, 1995.

[144] James Moody. The importance of relationship timing for diffusion. *Social forces*, 81(1):25–56, 2002. doi:10.1353/sof.2002.0056.

[145] Alfredo Morabia. Epidemiology's 350th anniversary: 1662–2012. *Epidemiology (Cambridge, Mass.)*, 24(2):179, 2013.

[146] Eden Morales-Narváez and Can Dincer. The impact of biosensing in a pandemic outbreak: Covid-19. *Biosensors and Bioelectronics*, 163:112274, 2020.

[147] Aravind Natarajan, Hao-Wei Su, and Conor Heneghan. Assessment of physiological signs associated with covid-19 measured using wearable devices. *NPJ digital medicine*, 3(1):1–8, 2020.

[148] Alexander Nesterov and Pablo Héctor Mata Villafuerte. Critical phenomena in complex networks: from scale-free to random networks. *The European Physical Journal B*, 96(11):1–14, 2023.

[149] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118, Jul 2001. URL: https://link.aps.org/doi/10.1103/PhysRevE.64.026118, doi:10.1103/PhysRevE.64.026118.

[150] Mark EJ Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002. doi:10.1103/PhysRevE.66.016128.

[151] Mark EJ Newman. *Networks*. Oxford university press, 2018.

[152] Michael Nikolaou. Revisiting the standard for modeling the spread of infectious diseases. *Scientific Reports*, 12(1):7077, 2022.

[153] US Department of Health, Human Services, et al. What is the difference between isolation and quarantine?, 2020.

[154] Augustine Okolie, Johannes Müller, and Mirjam Kretzschmar. Parameter estimation for contact tracing in graph-based models. *Journal of the Royal Society Interface*, 20(208):20230409, 2023.

[155] Nuria Oliver, Bruno Lepri, Harald Sterly, Renaud Lambiotte, Sébastien Deletaille, Marco De Nadai, Emmanuel Letouzé, Albert Ali Salah, Richard Benjamins, Ciro Cattuto, et al. Mobile phone data for informing public health actions across the covid-19 pandemic life cycle, 2020.

[156] Daniel P Oran and Eric J Topol. Prevalence of asymptomatic sars-cov-2 infection: a narrative review. *Annals of internal medicine*, 173(5):362–367, 2020.

[157] Juyong Park and Mark EJ Newman. Statistical mechanics of networks. *Physical Review E*, 70(6):066117, 2004.

[158] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015. doi:10.1016/j.physrep.2020.12.003.

[159] Romualdo Pastor-Satorras and Alessandro Vespignani. Immunization of complex networks. *Physical review E*, 65(3):036104, 2002.

[160] Leto Peel, Daniel B Larremore, and Aaron Clauset. The ground truth about metadata and community detection in networks. *Science advances*, 3(5):e1602548, 2017.

[161] Lorenzo Pellis, Neil M Ferguson, and Christophe Fraser. The relationship between real-time and discrete-generation models of epidemic spread. *Mathematical biosciences*, 216(1):63–70, 2008.

[162] Mathew Penrose. *Random geometric graphs*, volume 5. OUP Oxford, 2003.

[163] Nicola Perra. Non-pharmaceutical interventions during the covid-19 pandemic: A review. *Physics Reports*, 913:1–52, 2021.

[164] Nicola Perra, Duygu Balcan, Bruno Gonçalves, and Alessandro Vespignani. Towards a characterization of behavior-disease models. *PloS one*, 6(8):e23084, 2011.

[165] Nicola Perra, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Activity driven modeling of time varying networks. *Scientific reports*, 2(1):469, 2012.

[166] Carlo Piccardi. Social networks and the spread of epidemics. *Lettera Matematica*, 1:119–126, 2013.

[167] Mason A Porter and James P Gleeson. Dynamical systems on networks.

[168] MJ Powell. Site percolation in random networks. *Physical Review B*, 21(8):3725, 1980.

[169] Giorgio Quer, Jennifer M Radin, Matteo Gadaleta, Katie Baca-Motes, Lauren Ariniello, Edward Ramos, Vik Kheterpal, Eric J Topol, and Steven R Steinhubl. Wearable sensor data and self-reported symptoms for covid-19 detection. *Nature Medicine*, 27(1):73–77, 2021.

[170] Luis EC Rocha, Fredrik Liljeros, and Petter Holme. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS computational biology*, 7(3):e1001109, 2011. doi:10.1371/journal.pcbi.1001109.

[171] Kenneth J Rothman. The rise and fall of epidemiology, 1950–2000 ad. *International journal of epidemiology*, 36(4):708–710, 2007.

[172] Carlo Rovelli. *There Are Places in the World Where Rules Are Less Important Than Kindness: And Other Thoughts on Physics, Philosophy and the World*. Penguin, 2023.

[173] Abbas Ali Saberi. Recent advances in percolation theory and its applications. *Physics Reports*, 578:1–32, 2015.

[174] Marcel Salathe, Linus Bengtsson, Todd J Bodnar, Devon D Brewer, John S Brownstein, Caroline Buckee, Ellsworth M Campbell, Ciro Cattuto, Shashank Khandelwal, Patricia L Mabry, et al. Digital epidemiology. *PLoS Comput Biol*, 8(7):e1002616, 2012.

[175] Masaki Sano and Keiichi Tamai. A universal transition to turbulence in channel flow. *Nature Physics*, 12(3):249–253, 2016. doi:10.1038/nphys3659.

[176] Clodomir Santana, Federico Botta, Hugo Barbosa, Filippo Privitera, Ronaldo Menezes, and Riccardo Di Clemente. Covid-19 is linked to changes in the time–space dimension of human mobility. *Nature Human Behaviour*, pages 1–11, 2023.

[177] Jari Saramäki, Arash Badie-Modiri, Abbas K. Rizi, Mikko Kivelä, and Márton Karsai. Weighted temporal event graphs and temporal-network connectivity. In *Temporal Network Theory*, pages 107–130. Springer, 2023.

[178] Jari Saramäki and Petter Holme. *Temporal Network Theory*. Springer International Publishing, Cham, 2023. doi:10.1007/978-3-031-30399-9.

[179] Jari Saramäki, Mikko Kivelä, and Márton Karsai. Weighted temporal event graphs. In *Temporal Network Theory*, pages 107–128. Springer, 2019. doi:10.1007/978-3-030-23495-9_6.

[180] Yasmin SarcheshmehPour, Yu Tian, Linli Zhang, and Alexander Jung. Clustered federated learning via generalized total variation minimization. *IEEE Transactions on Signal Processing*, 2023.

[181] Jonathan Schaffer. What not to multiply without necessity. *Australasian Journal of Philosophy*, 93(4):644–664, 2015.

[182] Dhruv R Seshadri, Evan V Davies, Ethan R Harlow, Jeffrey J Hsu, Shanina C Knighton, Timothy A Walker, James E Voos, and Colin K Drummond. Wearable sensors for covid-19: a call to action to harness our digital infrastructure for remote patient monitoring and virtual assessments. *Frontiers in Digital Health*, 2:8, 2020.

[183] James Sethna. *Statistical mechanics: entropy, order parameters, and complexity*, volume 14. Oxford University Press, USA, 2021.

[184] Emily Seto, Priyanka Challa, and Patrick Ware. Adoption of covid-19 contact tracing apps: A balance between privacy and effectiveness. *Journal of medical Internet research*, 23(3):e25726, 2021.

[185] Roy J Shephard. *An illustrated history of health and fitness, from pre-history to our post-modern world*. Springer, 2015.

[186] Amir H Shirazi, Arash Badie-Modiri, Sara Heydari, Jennifer L Rohn, Gholam R Jafari, and Ali R Mani. Evolution of communities in the medical sciences: evidence from the medical words network. *Plos one*, 11(12):e0167546, 2016. doi:10.1371/journal.pone.0167546.

[187] Didier Sornette. *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools*. Springer Science & Business Media, 2006.

[188] Tiziano Squartini, Joey de Mol, Frank den Hollander, and Diego Garlaschelli. Breaking of ensemble equivalence in networks. *Physical review letters*, 115(26):268701, 2015.

[189] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, et al. High-resolution measurements of face-to-face contact patterns in a primary school. *PloS one*, 6(8):e23176, 2011.

[190] Kazumasa A Takeuchi, Masafumi Kuroda, Hugues Chaté, and Masaki Sano. Experimental realization of directed percolation criticality in turbulent liquid crystals. *Physical Review E*, 80(5):051116, 2009. doi:10.1103/PhysRevE.80.051116.

[191] Pieter Trapman, Frank Ball, Jean-Stéphane Dhersin, Viet Chi Tran, Jacco Wallinga, and Tom Britton. Inferring R_0 in emerging epidemics—the effect of common population structure is small. *Journal of The Royal Society Interface*, 13:20160288, 2016. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2016.0288, doi:10.1098/rsif.2016.0288.

[192] Lucas Daniel Valdez, Lautaro Vassallo, and Lidia Adriana Braunstein. Epidemic control in networks with cliques. *Physical Review E*, 107(5):054304, 2023.

[193] Pauline Van den Driessche. Reproduction numbers of infectious disease models. *Infectious disease modelling*, 2(3):288–303, 2017.

[194] Remco Van Der Hofstad. *Random graphs and complex networks*, volume 43. Cambridge university press, 2016.

[195] Pim van der Hoorn, Gabor Lippner, and Dmitri Krioukov. Sparse maximum-entropy random graphs with a given power-law degree distribution. *Journal of Statistical Physics*, 173:806–844, 2018.

[196] Richard S Varga. Iterative analysis. *New Jersey*, 322, 1962.

[197] Emilia Vynnycky and Richard White. *An introduction to infectious disease modelling*. OUP oxford, 2010.

[198] Benjamin Wacker and Jan Schlüter. Time-continuous and time-discrete sir models revisited: theory and applications. *Advances in Difference Equations*, 2020(1):556, 2020.

[199] Patrick GT Walker, Charles Whittaker, Oliver J Watson, Marc Baguelin, Peter Winskill, Arran Hamlet, Bimandra A Djafaara, Zulma Cucunubá, Daniela Olivera Mesa, Will Green, et al. The impact of covid-19 and strategies for mitigation and suppression in low-and middle-income countries. *Science*, 369(6502):413–422, 2020.

[200] Chia C Wang, Kimberly A Prather, Josué Sznitman, Jose L Jimenez, Seema S Lakdawala, Zeynep Tufekci, and Linsey C Marr. Airborne transmission of respiratory viruses. *Science*, 373(6558):eabd9149, 2021.

[201] Wei Wang, Ming Tang, H Eugene Stanley, and Lidia A Braunstein. Unification of theoretical approaches for epidemic spreading on complex networks. *Reports on progress in physics*, 80(3):036603, 2017.

[202] Zhen Wang, Chris T Bauch, Samit Bhattacharyya, Alberto d'Onofrio, Piero Manfredi, Matjaž Perc, Nicola Perra, Marcel Salathé, and Dawei Zhao. Statistical physics of vaccination. *Physics Reports*, 664:1–113, 2016.

[203] Stanley Wasserman, Katherine Faust, et al. *Social network analysis: Methods and applications*. Cambridge university press, 1994.

[204] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *nature*, 393(6684):440–442, 1998. doi:10.1038/30918.

[205] Nicholas C Wormald et al. Models of random regular graphs. *London Mathematical Society Lecture Note Series*, pages 239–298, 1999.

[206] Ana Pastore y Piontti, Nicola Perra, Luca Rossi, Nicole Samay, and Alessandro Vespignani. *Charting the next pandemic: modeling infectious disease spreading in the data science age*. Springer, 2019.

[207] W Yang, D Zhang, L Peng, C Zhuge, and L Hong. Rational evaluation of various epidemic models based on the covid-19 data of china [internet]. *Nueva York: ArXiv*, 2020.

[208] Dongni Zhang and Tom Britton. Epidemic models with digital and manual contact tracing. *arXiv preprint arXiv:2211.12869*, 2022.

[209] Zhaoyang Zhang, Honggang Wang, Chonggang Wang, and Hua Fang. Modeling epidemics spreading on social contact networks. *IEEE transactions on emerging topics in computing*, 3(3):410–419, 2015.

[210] Qiankun Zhao, Yuan Tian, Qi He, Nuria Oliver, Ruoming Jin, and Wang-Chien Lee. Communication motifs: a tool to characterize social communications. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1645–1648, 2010. doi:10.1145/1871437.1871694.

Aalto-DT 48/2024

The thesis focuses on epidemic spreading and the relation between the underlying network structure and the spread of the epidemic but has a broad scope. ***The introductory chapters, in particular, give a very nice summary of the state of the art in network epidemiology, making the thesis accessible for researchers or students who have never worked with epidemiology before.***

Clara Stegehuis, Mathematician & Associate Professor at Twente University

The dissertation begins with a short introduction motivating the thesis (which includes a couple of great quotes that I will steal), followed by ***three pedagogical chapters preparing any reader with a solid scientific background to understand the papers presented in the final part of the dissertation.*** *These three "micro-courses" cover i) mathematical epidemiology, ii) network theory, and iii) the integration of these two subjects in the third chapter.* ***The writing is engaging, and I sincerely appreciate the effort to render the reading of this extensive body of text a more pleasant journey.***

Riccardo Gallotti, Head of the Complex Human Behaviour Lab, FBK

ISBN 978-952-64-1704-2 (printed)
ISBN 978-952-64-1705-9 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Computer Science
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL THESES