

Stochastic epidemics on random networks and competition in growth

Carolina Fransson



Stochastic epidemics on random networks and competition in growth

Carolina Fransson

Academic dissertation for the Degree of Doctor of Philosophy in Mathematical Statistics at Stockholm University to be publicly defended on Friday 10 February 2023 at 13.00 in Lecture room 4, 2nd floor, house 1, Albano, Albanovägen 28.

Abstract

The COVID-19 pandemic has dramatically demonstrated the importance of epidemic models in understanding and predicting disease spread and in assessing the effectiveness of interventions. The overarching topic of this thesis is stochastic epidemic modelling, with the main focus on the role of the underlying social structure in infectious disease spread.

In Paper I we study the spread of stochastic SIR-epidemics on an extended version of the configuration model with group structure. We present expressions for the basic reproduction number R_0 , the probability of a major outbreak and the expected final size, and investigate random vaccination with a perfect vaccine. We weaken the assumptions of earlier results for epidemics on this type of graph by allowing for heterogeneous infectivity both in individual infectivity and between different kinds of edges. An important special case of this model is the spread of a disease with arbitrary infectious period distribution in continuous time.

Paper II concerns multi-type competition in a variant of Pólya's urn model with interaction, where balls of different colours/types annihilate upon contact. The model dynamics are governed by the structure of an underlying graph. In the special case of a cycle graph, this urn model is equivalent to a planar growth model with competing pathogens. It has earlier been shown that in the two-type case, indefinite coexistence has probability 0 for any (finite and connected) underlying graph, while for $K \geq 3$ types the possibility of coexistence depends on the structure of this graph. We show that for $K \geq 3$ types competing on a cycle graph, there is with probability 1 eventually only one remaining type.

In Paper III we study the real-time growth rate of SIR epidemics on random intersection graphs with mixed Poisson degree distribution. We show that during the early stage of the epidemic, the number of infected individuals grows exponentially and the Malthusian parameter is shown to satisfy a version of the Euler-Lotka equation. These results are obtained via an approximating embedded single-type Crump-Mode-Jagers branching process. In addition, we provide a lower bound on the cumulative number of individuals that get infected before the branching process approximation breaks down.

In Paper IV we consider stochastic SIR epidemics on inhomogeneous random graphs with degree-dependent contact rates. In this model, the per-neighbour contact rate of an individual decrease but its overall expected contact rate increases with its expected number of neighbours. We provide the basic reproduction number R_0 , the probability of a large outbreak and the final size of an epidemic. We show that reducing heterogeneity in contact rates results in a higher value of the basic reproduction number R_0 , and demonstrate that this result does not generally extend to the probability of a major outbreak and the final size.

Keywords: *random graphs, branching processes, SIR epidemics, malthusian parameter, urn model.*

Stockholm 2023

<http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-212974>

ISBN 978-91-8014-144-4

ISBN 978-91-8014-145-1



Stockholm
University

Department of Mathematics

Stockholm University, 106 91 Stockholm

STOCHASTIC EPIDEMICS ON RANDOM NETWORKS AND
COMPETITION IN GROWTH

Carolina Fransson



Stochastic epidemics on random networks and competition in growth

Carolina Fransson

©Carolina Fransson, Stockholm University 2023

ISBN print 978-91-8014-144-4

ISBN PDF 978-91-8014-145-1

Printed in Sweden by Universitetservice US-AB, Stockholm 2022

List of Papers

This thesis is based on the following papers.

I: SIR epidemics and vaccination on random graphs with clustering

Fransson, C. & Trapman, P. (2019).
In: *Journal of Mathematical Biology*
DOI:10.1007/s00285-019-01347-2

II: Multicolour competition with reinforcement

Ahlberg, D. & Fransson, C. (2022).
The manuscript is submitted for publication.
DOI:10.48550/arXiv.2206.00400

III: The real-time growth rate of stochastic epidemics on random intersection graphs

Fransson, C. (2022).
The manuscript is submitted for publication.
DOI:10.48550/arXiv.2211.11034

IV: Stochastic epidemics on inhomogenous random graphs with degree-dependent contact rates

Fransson, C. & Trapman, P. (2022).

Authors' contributions: In Paper I, Fransson contributed with the mathematical analysis and wrote the paper with feedback from Trapman. Trapman suggested the model. A simpler version of Paper I is part of Fransson's master's thesis. Fransson initiated Paper II and formulated the early proof strategy with continuous discussions with Ahlberg. Overall the authors contributed equally to the proofs, and also to the writing of the paper. Fransson carried out every part in the making of Paper III, with feedback from Trapman. In Paper IV, Trapman formulated an early draft of the model. Fransson formulated the exact model and suggested and proved Theorem 6. The authors contributed equally both to the proofs and to the writing of the paper.

Acknowledgments

First of all, I would like to express my deep gratitude to my main supervisor Pieter Trapman for the continuous support, for always having an open door and for encouraging me to develop as a mathematician. I could not have asked for a better mentor. Many thanks to my supervisor Daniel Ahlberg for his positive support, encouragement and generous advice.

Thanks to my colleagues - past and present - for making the department such a nice workplace.

Thank you David for always listening, I cannot imagine doing this without you. Thanks also to my family - mum, dad, Elin and Klara - I appreciate the support I received from you.

Finally, I would like to thank Vetenskapsrådet for the financial support that made this research possible.

Contents

| | |
|--|-----------|
| List of Papers | 1 |
| Acknowledgments | 3 |
| I Introduction | 1 |
| 1 Epidemic modelling | 3 |
| 1.1 The SIR epidemic model | 4 |
| 1.2 Graphs as social networks | 5 |
| 1.3 Branching processes approximations | 6 |
| 1.3.1 Forward and backward processes | 7 |
| 1.3.2 Reproduction numbers and herd immunity | 8 |
| 1.3.3 Malthusian law of exponential growth | 9 |
| 2 Random graphs | 11 |
| 2.1 The Erdős-Rényi-Gilbert model | 11 |
| 2.1.1 Phase transition and asymptotics | 12 |
| 2.2 The configuration model | 12 |
| 2.3 Random intersection graphs | 14 |
| 2.4 Inhomogeneous random graphs | 15 |
| 3 Growth and competition models | 17 |
| 3.1 Spatial growth models | 17 |
| 3.1.1 Competition in spatial growth | 20 |
| 3.2 Pólya urn models | 21 |

| | | |
|-----------|------------------------------|-----------|
| 4 | Overview of papers | 23 |
| 4.1 | Paper I | 23 |
| 4.2 | Paper II | 24 |
| 4.3 | Paper III | 25 |
| 4.4 | Paper IV | 26 |
| 5 | Svensk sammanfattning | 29 |
| II | Papers | 37 |

Part I

Introduction

1. Epidemic modelling

The rapid spread of COVID-19 has been met with an unprecedented range of government action around the world. The pandemic has put the spotlight on the value of mathematical epidemic models, not least in guiding control measures and facilitating in planning medical resources. Around the globe, epidemic models have been used to assess the effectiveness of interventions such as school closures and a shift to online education, contact tracing, and restrictions on travelling and public gatherings.

Using mathematical models to understand infectious disease spread has several benefits. As mentioned above, epidemic models help in assessing the efficacy of control measures for disease mitigation and prevention. During an ongoing outbreak, time is of essence in evaluating control policies. Even if we ignore the time aspect, direct observations of the effects of control measures are often impossible, as counterfactual outcomes are inherently unobservable. This means that it is not possible to directly evaluate the outcome in the counterfactual event that different (or no) control measures were taken. Unlike in many other scientific fields, controlled trials with repeated observations of disease outbreaks and the outcome of different control measures are generally unfeasible.

More broadly, mathematical epidemic models provide a tool for systematic understanding of sociological and biological mechanisms of infectious disease spread. A mathematical model is, by definition, a simplification of reality. The development of an infectious disease outbreak is the result of many, often highly complex, factors. Those factors include pathogen-specific properties, the health and immune status of individuals, and human behaviour both on the population and individual levels. A mathematical model helps to translate knowledge of those factors to a description of disease spread in a population and can provide insights about things that are not directly observable. For a general introduction to mathematical epidemic modelling, see e.g. Diekmann et al. (2013).

In this chapter we present the basic building blocks of many stochastic epidemic models. The focus is mainly on the methods and results used in Papers I, III and IV. In Section 1.1 we present the SIR epidemic model, which is one of the most popular models for characterizing the disease status of individuals. Section 1.2 contains a description of how the underlying social structure of a population can be incorporated into an epidemic model in the form of a

network. In Section 1.3 we give an overview of the role of branching process approximations in studying the behaviour of infectious diseases.

1.1 The SIR epidemic model

One of the most widely used models for infectious disease spread is the SIR model. It is based on the paper by Kermack and McKendrick (1927) where a more general model was introduced. In the SIR model, each individual is classified as either susceptible (S), infectious (I) or recovered (R) according to its current health status. At the start of the epidemic, one (or more) infectious individual is introduced into an otherwise susceptible population. As time progresses, individuals may move to another class if their health status changes. The disease spreads from the infectious to the susceptible individuals by contact transmission; if a susceptible individual comes into contact with an infectious individual it contracts the disease and immediately becomes infectious. An infectious individual may eventually cease to be infectious and become recovered. Once recovered, an individual has full and lasting immunity and plays no further role in the spread of the disease. The population is assumed to be closed, which means that there is no immigration, no births, and no deaths.

In its simplest form, the SIR model is deterministic with fixed transmission and recovery rates and homogeneous mixing, which means that any two individuals of the population interact according to the same pattern. It can then be described via a system of differential equations that governs the evolution of the disease spread over time. In this thesis, we focus on stochastic SIR models. Stochastic models are particularly useful when the number of infected individuals is small since randomness then plays an important role in the behavior of the epidemic. Stochastic models are, for instance, suitable to analyze the probability of a large outbreak in the early phase of an epidemic. They are also useful for models where small communities play an important role.

An important component of a stochastic epidemic model is the contact patterns of the population; instead of the fixed transmission rate of the deterministic SIR model described above, contacts between individuals occur at random points in time. The contact pattern among individuals typically depends on a number of factors, such as age, sex, geographical location etc. Heterogeneity in contacts may manifest both in terms of who has contact with whom and contact frequencies. Another component of a stochastic epidemic model is the infectious periods. Instead of the fixed recovery rate of the deterministic model, each infected individual remains infectious for some random period of time and then recovers.

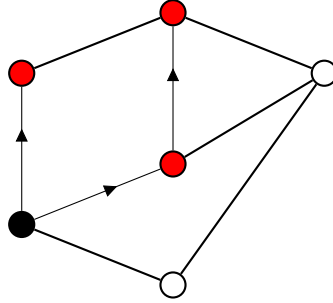


Figure 1.1: The spread of a disease on a graph. An edge symbolizes a relationship between the individuals that correspond to the two terminal nodes. The initial case is represented by the black node and an arrowhead signifies transmission. A red node represents a non-initial case and a white node represent an individual that ultimately escapes infection.

1.2 Graphs as social networks

A key factor in the spread of an infectious disease is the contact pattern between individuals of the population. The simple models with homogeneous mixing described above can be extended to better capture social structures. One such extension is to incorporate the underlying social network. For a wide range of diseases, the patterns of social interactions (e.g. social proximity and the frequency and duration of contacts) in the population play a major role in disease transmission (see e.g. Read et al. (2012) and the references therein). To incorporate this into an epidemic model we may use a graph G to model the social network; each vertex of G represents an individual and an edge of G represents a close social relationship and a potential channel of disease transmission. If an infectious individual is in close contact with a susceptible neighbour transmission occurs. Starting with one (or many) infected node(s) at time 0, the disease spreads on the network along edges. The epidemic spread on a graph is illustrated in Figure 1.1. In this thesis, we study epidemics on static networks (i.e. networks that do not change as time progresses). This means that the social structure remains intact during an outbreak. This may, for instance, reflect that the outbreak takes place over a relatively short time period.

To facilitate analysis of an epidemic on a graph, it is helpful to construct a so-called epidemic generated graph G^* that fully captures the disease spread. The epidemic generated graph G^* is a directed graph, which reflects the directed nature of disease transmission. An edge (u, v) of G^* represents a potential channel of transmission; if the initial vertex u of (u, v) gets infected (which might or might not occur) it transmits the disease to the end vertex v unless v has already been infected via some other route. The individuals ultimately infected are the individuals that can be reached from an initial case by following a path of transmission on G^* (a path on G^* is an ordered sequence e_1, e_2, \dots, e_k of edges such that the end vertex of e_i is the initial vertex of e_{i+1} , $i = 1, \dots, k-1$).

An example of an epidemic generated graph is depicted in Figure 1.2 in Section 1.3.1. To construct G^* , start with the underlying undirected social network G . Replace each undirected edge of G with two parallel directed edges, pointing in the opposite direction. Next, equip each individual v_i with an infectious period I_i , $i = 1, \dots, n$. For any two neighbours v_i and v_j , let $T_{i,j}$ be the time from the event (which might or might not occur) that v_i becomes infected to a contact between v_i and v_j . The standard choice is to let the infectious periods be i.i.d. copies of some non-negative random variable I and to let the $T_{i,j}$ be i.i.d. and exponentially distributed. In the next step, we delete edges where the initial vertex is not in contact with the end vertex during its infectious period. That is, an edge (v_i, v_j) is only kept if $T_{i,j} \leq I_i$.

1.3 Branching processes approximations

For many stochastic epidemic models, branching processes can be used to approximate the early spread of the disease. Questions that can be addressed via branching process approximations include “How does an epidemic behave in its early phase?”, “What is the probability of a large outbreak” and “If a large outbreak occurs, how many will get infected?”. In the case of epidemics on networks, the key network property that underpins this approximation is that, for a wide range of random graph models, the network structure is locally treelike in the limit as the number of vertices tends to infinity.

A branching process is a stochastic process that describes the demographic evolution of a population in time as individuals are born, reproduce and die. A salient feature of branching processes is that once born, the life of an individual is independent of the lives of other individuals. One of the simplest branching processes is the single-type Galton-Watson process, which describes the evolution of a population with non-overlapping generations. Let $Y = \{Y_n\}_{n \in \mathbb{N}}$ be a single-type Galton-Watson process, where Y_n is the population size at time n , $n \geq 1$. Individuals of Y reproduce at age 1 and then immediately dies, which means that Y_n is the size of generation n . Let the random variable X be distributed as the number of offspring of an individual of Y , and let $\{X_{n,i}\}_{n,i \in \mathbb{N}}$ be a collection of independent copies of X . Here $X_{n,i}$ is the number of children born to individual i (according to some ordering) of generation n for $i = 1, \dots, Y_n$. The number of individuals of generation n , Y_{n+1} , is then given by

$$Y_{n+1} = \sum_{i=1}^{Y_n} X_{n,i}.$$

Ultimately, a Galton-Watson process either goes extinct at some point or survives forever. The expected number of offspring $E(X)$ of an individual of Y is a threshold parameter for the fate of the branching process population; the population of Y has a strictly positive probability of survival if and only if $E(X) > 1$. If $E(X) > 1$ we say that Y is supercritical, if $E(X) = 1$ we say that

Y is critical and if $E(X) < 1$ we say that Y is subcritical. Assume that Y has a single ancestor (that is $Y_1 = 1$) and let q be the probability that Y eventually goes extinct. Since Y goes extinct if and only if none of the children of the ancestor gives rise to a branching process population that survives indefinitely, the extinction probability q satisfies

$$q = \sum_{k=0}^{\infty} q^k P(X = k) = E(q^X)$$

where X follows the offspring distribution. It is easily verified that $1 = E(1^X)$. It can be shown that the equation $q = E(q^X)$ has at most two solutions in $(0, 1]$ and the extinction probability is given by the smallest solution in $(0, 1]$ (see e.g. (Haccou et al., 2005)).

For many epidemic models, a multitype branching process approximation is required. For an introduction to multitype branching processes, we refer the reader to Haccou et al. (2005). Galton-Watson processes can be used to approximate the disease spread on a generation-basis by ignoring the real-time dynamics of transmission. To investigate the real-time spread a continuous branching process approximation can be used. Continuous branching processes are presented in Section 1.3.3.

1.3.1 Forward and backward processes

In the early phase of an outbreak, when the number of individuals reached by the disease is still small, the epidemic process can be approximated by a branching process. A birth in the approximating branching process represents a transmission in the epidemic process and a death represents a recovery. This approximation provides a tool to describe the behaviour of the disease in the early phase. In particular, it yields an approximation of the probability that a large outbreak occurs via the corresponding branching process survival probability. The stochastic process that describes the course of an outbreak as time progresses is called the forward epidemic process, and the approximating branching process is referred to as the forward branching process.

Branching processes can also be used to approximate the final size of a large outbreak. Central to this approach are the concepts of so-called susceptibility sets (see e.g. Ball and Lyne (2001)) and backward processes. Recall the epidemic generated graph G^* described in Section 1.2. The susceptibility set S_v of a vertex v consists of the set of vertices that has a directed path of transmission to v . In other words, a vertex u belongs to S_v if and only if v would contract the disease in the (possibly counterfactual) event that u is an initial case. An example of a susceptibility set can be found in Figure 1.2. The final size of an outbreak is the number of vertices whose susceptibility set contains an initial case. The susceptibility set S_v of v can be explored by following edges of G^* backwards, starting at v . The process that is obtained by tracing (potential)

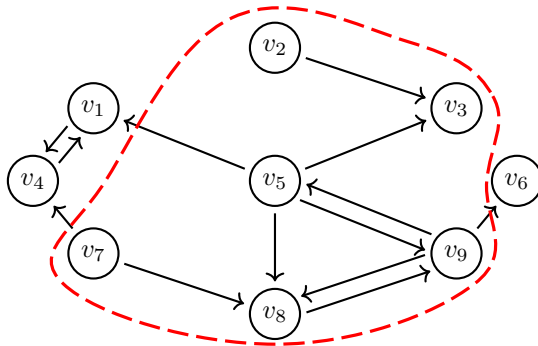


Figure 1.2: An epidemic generated graph. The red dashed line encircles the nodes of the susceptibility set of v_3 .

transmissions backwards is called the backward epidemic process. For many models, the backward epidemic process can be approximated by a branching process. In the backward branching process, the offspring of a particle x represents the individuals that would have transmitted the disease to the individual that corresponds to x .

Let p_f and p_b be the survival probabilities of the forward and backward branching processes, respectively, and let G_n be a graph on n vertices. Under some technical conditions (that are mainly designed to ensure local convergence of the graph structure), the approach outlined above typically gives a variant of the following theorem (e.g. Ball et al. (2014)).

Theorem 1. *For an epidemic on G_n , let S_n be the fraction of the population that ultimately gets infected. As $n \rightarrow \infty$, S_n converges in distribution to a random variable S that satisfies*

$$P(S = 0) = 1 - p_f \text{ and } P(S = p_b) = p_f.$$

The key step which gives some variant of Theorem 1 is to construct a coupling such that, in the limit as the population size $n \rightarrow \infty$, the forward and backward epidemic processes connect precisely when the corresponding branching processes survive.

1.3.2 Reproduction numbers and herd immunity

In this section the basic reproduction number R_0 and the effective reproduction number R_t are presented. These numbers serve as threshold parameters for herd immunity. As such, they play a crucial role in epidemic modelling and helps guide intervention policy decisions. The basic reproduction number R_0 represents the average number of cases caused by a typical infectious individual

in the early phase of an outbreak when the number of individuals reached by the disease is still small. It is a threshold parameter for herd immunity in the sense that a large outbreak occurs with positive probability if and only if $R_0 > 1$. For models where the epidemic process can be approximated by a single-type Galton-Watson process, the basic reproduction number is typically taken to be

$$R_0 = E(X),$$

where X follows the offspring distribution.

The circumstances that are relevant for disease spread are generally not static. For instance, as control measures are implemented and individuals gain immunity (acquired from pathogen exposure or vaccination) prospects for disease containment and mitigation improve. A quantity that is closely related to R_0 is the effective reproduction number R_t , $t \geq 0$, which is the expected number of cases caused by a typical infectious individual at time $t \geq 0$. The effective reproduction number has similar threshold properties as R_0 . By reducing the effective reproduction number below unity the disease can be eliminated.

1.3.3 Malthusian law of exponential growth

In his late 18th century book "An Essay on the Principle of Population", the economist Thomas Malthus warned of famine, disease and war that would ensue if birth rates did not decrease. This prediction was based on the observation that human population size tended to follow a pattern of exponential increase in times of abundance, followed by a sharp decline when the per capita resources became scarce. Fortunately, Malthus' prediction did not materialize (Trewavas, 2002). Since his time, humanity has achieved unprecedented levels of prosperity and large-scale Malthusian disasters have been avoided.

Like the population in Malthus' predictions, the size of a branching process population grows exponentially in time. In the beginning of Section 1.3, we described how generation-based branching processes can be used to analyze the final outcome of an epidemic. To analyze the real-time dynamics of the early phase of an outbreak, a continuous branching process can be employed. Let $Z = \{Z_t\}_{t \geq 0}$ be a single-type branching process in continuous time, where Z_t is the number of individuals of Z that are alive at time t . An individual x of Z reproduces according to a point process on the time axis $[0, \infty)$, where the points represent x 's ages at reproduction. The life (i.e. the reproduction process and life length) of an individual is assumed to be independent of the lives of all other individuals of Z , and the lives of the individuals of Z all follow the same law. Let $\{\xi(t)\}_t$ and T be the reproduction process and life length of a generic individual x of Z . Denote the total number of children of x by N , and let τ_1, \dots, τ_N be the ages of x at reproduction (some of the τ_i may coincide, since we allow for litters of size larger than one). For simplicity, here we assume that $T < \infty$ and $E(N) < \infty$. Then

$$\xi(a) = \sum_{i=1}^{\infty} \mathbb{1}(\tau_i \leq a).$$

Denote the intensity measure of the offspring process by μ . That is, for $t \geq 0$, $\mu(t) = E(\xi(t))$ is the expected number of offspring born to an individual at age t . We say that Z is supercritical if the embedded Galton-Watson process is supercritical.

Definition 1 (Malthusian parameter). Let Z be supercritical. The Malthusian parameter is the unique solution $\alpha > 0$ to

$$\int_0^{\infty} e^{-\alpha t} \mu(dt) = 1. \tag{1.1}$$

The equation in (1.1) is often referred to the Euler-Lotka equation.

Under some technical conditions, the branching process population either dies out or grows infinitely large at an exponential rate α (Iksanov and Meiners, 2015, Thrm 2.4). To be more precise, it holds almost surely that

$$e^{-\alpha t} Z_t \rightarrow W$$

as $t \rightarrow \infty$, where W is an integrable stochastic variable. Moreover, it holds almost surely that $W = 0$ precisely when Z goes extinct (and $W > 0$ precisely when Z survives forever).

2. Random graphs

Real world social networks are typically large with a complicated structure. This makes mathematical modelling challenging since many social networks do not lend themselves to tractable analysis. In addition to this, the exact structure of a network is often not available (van der Hofstad, 2016). One way to circumvent these issues is to use a random graph in place of the social network. In the random graph literature there is a plethora of models that capture various properties exhibited in real world networks. Those properties are typically local (which means that they pertain to a specific vertex and its near neighbours), or global. Some examples of local properties include the degree distribution of the vertices and community structure.

In this chapter we give an overview of the field of random graphs and present the models used in this thesis. Section 2.1 is devoted to one of the earliest random graph models, the Erdős-Rényi model, and some key aspects of random graphs. In Section 2.2 we describe the configuration model. This type of model is central to Paper I, where a generalisation of the configuration model is used. Section 2.3 contains an overview of a random intersection graph model. This graph model is a simpler version of the model employed in Paper III. In Section 2.4 we describe a special case of the inhomogeneous graph model of Bollobás et al. (2007), which is used in Paper IV.

2.1 The Erdős-Rényi-Gilbert model

One of the earliest random graph models is the Erdős-Rényi-Gilbert model, often referred to as the Erdős-Rényi model. It was introduced independently (in slightly different versions) by Erdős and Rényi (1959) and Gilbert (1959). Denote an Erdős-Rényi graph by $G(n, p)$, where n is the number of vertices of the graph and the parameter $p \in [0, 1]$ is the (occupational) edge probability. To construct $G(n, p)$, start with the vertex set $\{v_1, \dots, v_n\}$. For each (unordered) pair of vertices v_i and v_j , toss a p -biased coin. We assume independence between coin tosses. The edge (v_i, v_j) is present if the coin comes up heads and absent if it comes up tails. In Figure 2.1, the construction of an Erdős-Rényi graph on five vertices is illustrated.

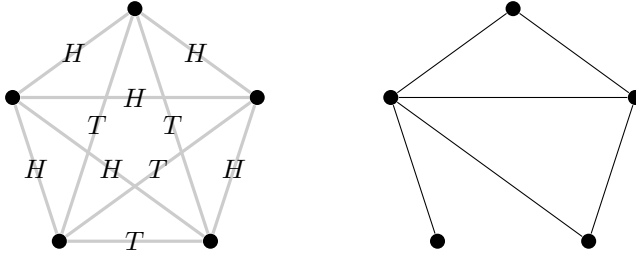


Figure 2.1: Construction of an Erdős-Rényi graph on five vertices. Left: A coin is tossed for each pair of vertices. H indicates heads, T indicates tails. Right: The resulting graph. Only edges whose coin toss come up heads are present.

2.1.1 Phase transition and asymptotics

Many random graph models (including the Erdős-Rényi model and the models considered in this thesis) exhibit so-called phase transitions. In a broader context, a phase transition is an abrupt transition between different states brought on by a slight change in external conditions. Some examples include transitions between solid, liquid and gas form caused by a change in temperature or pressure. In the context of random graphs, phase transitions usually pertains to a sharp transition in the size of the largest connected component (a connected component is a connected subgraph in which no vertex has a neighbour that is not a member of this subgraph). In the limit as the graph size tends to infinity, a wide range of random graph models have a critical threshold; above this threshold the size of the largest component is of the same order as the number of vertices, and at or below this threshold only relatively small components are present. The degree distribution of an Erdős-Rényi graph $G(n, \frac{c}{n})$ converges to a Poisson distribution with mean c as $n \rightarrow \infty$ (see e.g. van der Hofstad (2016, Theorem 5.12)) and the asymptotic phase transition occurs at the critical threshold where the mean degree is $c = 1$. In the limit as $n \rightarrow \infty$ the graph $G(n, \frac{c}{n})$ is locally treelike (see e.g. van der Hofstad (2016, Section 4)). This property plays a vital role in branching process approximations of disease spread.

2.2 The configuration model

The configuration model is a random graph model for generating graphs with a specified degree sequence (or distribution). It was introduced by Bollobás (1980). A graph generated according to the configuration model on n vertices is generated as follows. Let the degree sequence $\{d_1, \dots, d_n\}$ be a given sequence of non-negative integers. Assign d_i half-edges to vertex v_i , $i = 1, \dots, n$, as illustrated in Figure 2.2. Here d_i represents the (future) degree of vertex i . Next, we connect the half-edges by pairing them uniformly at random. When

half-edges are paired, self-loops and multiple edges may be created. As the size n of the graph tends to infinity, the number of self-loops and multiple edges is (for many purposes) negligible (van der Hofstad, 2016, Proposition 7.13). One way to deal with these imperfections is to remove self-loops and merge multiple edges. Similarly, if the number of half-edges is odd we remove the remaining unpaired half-edge. The procedure described above can also be used to generate a graph with a specified degree distribution. This can be achieved by letting d_1, \dots, d_n be i.i.d. realizations from the desired distribution.

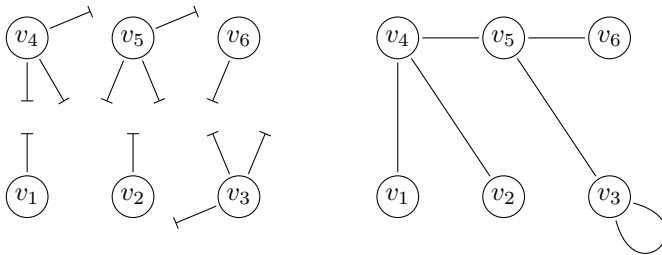


Figure 2.2: Construction of a configuration random graph. Left: Each vertex is assigned a number of half-edges according to its degree. Right: The resulting graph after the half-edges have been paired uniformly at random.

As is the case for the Erdős-Rényi model, configuration model graphs exhibit a phase transition. If the empirical distribution of the degree sequence converges to the distribution of a random variable D as the number of vertices $n \rightarrow \infty$ then under mild regularity conditions the phase transition occurs at the critical threshold $E(D^2)/E(D) - 1 = 1$ (see e.g. van der Hofstad (2022, Theorem 4.9)). The heuristics behind this threshold is that $E(D^2)/E(D)$ is the (asymptotic) expected degree of the vertex of a uniformly selected half-edge. Thus, if we explore the structure of the graph by following edges from vertex to vertex, then the phase transition threshold $E(D^2)/E(D) - 1$ is the expected number of additional neighbours of a vertex reached in this exploration process.

The Erdős-Rényi model and the configuration model share the property that, asymptotically, they are locally treelike. In the social network representation, this means that the people in the social vicinity of an individual tend to not know each other. Real-world networks, however, typically have a non-negligible amount of short circuits (Newman, 2003). In Paper I, we consider an extended version of the classical configuration model where the neighbours of an individual may be neighbours as well.

2.3 Random intersection graphs

As mentioned in the previous section, real world networks are often not locally treelike. Clustering in social networks may, for instance, arise from groups such as schools, households and workplaces. To generate random graphs with groups in the form of highly connected subgraphs (which we refer to as cliques here) the random intersection graph model can be used. It first appeared in the literature in the thesis by Singer (1995) and later in Karoński et al. (1999). Other random graph models that allow for connected groups are Bollobás et al. (2011); Karoński et al. (1999); Newman (2002).

Let G be a finite and simple (i.e. no multiple edges or self-loops) graph on n vertices, which we denote by v_1, \dots, v_n . We say that G is an intersection graph if there exists sets S_1, \dots, S_n such that $S_i \cap S_j \neq \emptyset$ precisely when G has an edge between v_i and v_j , $i, j = 1, \dots, n$. The characterization of an intersection graph G in terms of S_1, \dots, S_n is illustrated in Figure 2.3.

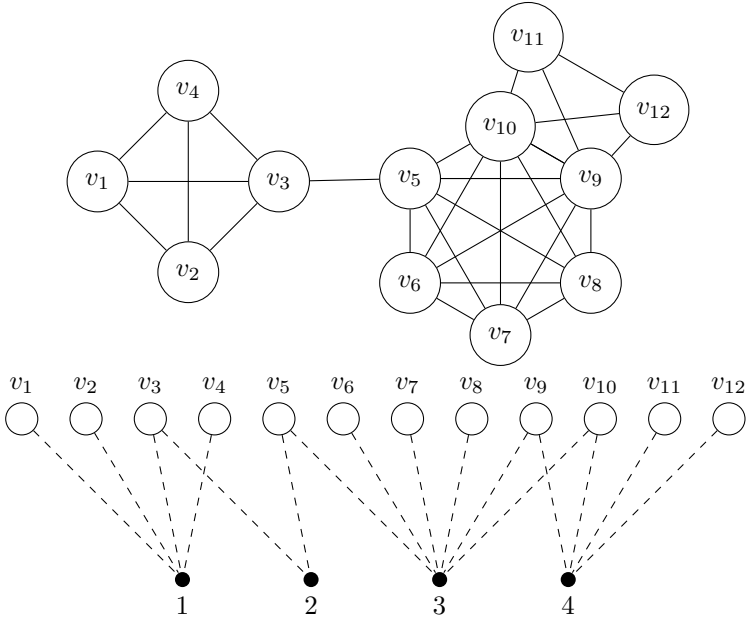


Figure 2.3: Top: An intersection graph on 12 vertices. Bottom: Illustration of the corresponding sets S_1, \dots, S_{12} . A dashed line between v_i and $k \in \mathbb{N}$ means that $k \in S_i$, $i = 1, \dots, 12$. The intersection graph has four cliques, which correspond to the four intersections $\{1\} = S_1 \cap S_2 \cap S_3 \cap S_4$, $\{2\} = S_3 \cap S_5$, $\{3\} = S_5 \cap \dots \cap S_{10}$ and $\{4\} = S_1 \cap S_2 \cap S_3 \cap S_4$.

A random intersection graph on n vertices can be created by generating the sets S_1, \dots, S_n randomly. Let $G(n, m, p)$ be a random intersection graph with parameters n, m and p , where n and m are positive integers and $p \in [0, 1]$. Here

n and m are the numbers of vertices and cliques of $G(n, m, p)$, respectively. To construct $G(n, m, p)$, we start by selecting the vertex v_1 and repeatedly flip an independent p -biased coin m times. If the coin comes up heads in the k th flip we add the element k to S_1 , $k = 1, \dots, m$. In the resulting graph, this means that v_1 is a member of the clique that corresponds to k . If the coin comes up tails, we do nothing. This procedure is then repeated for node v_i for $i = 2, \dots, n$; in the i th step we add elements in $\{1, \dots, m\}$ to S_i based on a series of independent p -biased coin tosses.

2.4 Inhomogeneous random graphs

Inhomogeneous random graphs are constructed in a similar vein as Erdős-Rényi graphs and random intersection graphs. Unlike these two graph models, however, an inhomogeneous random graph does not have heterogeneous occupational edge probabilities. In this model, each vertex is assigned a weight. Given the weights, edges are present or absent independently. The model is inhomogeneous in the sense that the occupational probability of an edge depends on the weights of its terminal vertices. Here, we give an informal introduction to the special case of the inhomogeneous graph model considered in Paper IV. For the full model we refer the reader to Bollobás et al. (2007).

To construct an inhomogeneous random graph on n vertices, start with the vertex set $\{v_1, \dots, v_n\}$. Let the weights w_1, \dots, w_n be a sequence in \mathbb{R} . Assign the weight $w_i \geq 0$ to each vertex v_i , $i = 1, \dots, n$. Let the function $p : \mathbb{R}^2 \rightarrow [0, 1]$ be continuous and symmetric. Given the weights, the probability that the vertices v_i and v_j share an edge is $p(w_i, w_j)$, independently between pairs of vertices. Here we assume that p takes the form

$$p(x, y) = \min\left(\frac{\varphi(x, y)}{n}, 1\right) \quad (2.1)$$

where the function $\varphi : \mathbb{R}^2 \rightarrow [0, \infty)$ is non-decreasing in both variables (i.e. $\varphi(x_1, y_1) \leq \varphi(x_2, y_2)$ whenever $x_1 \leq x_2$ and $y_1 \leq y_2$), continuous and symmetric. Similarly to the random graph models presented above, homogeneous random graphs experience a phase transition (Bollobás et al., 2007, Thrm 3.1).

In the context of a social network, the monotonicity of φ means that the weights can be interpreted as a measure of sociability; the higher the weight of an individual, the more prone it is to form social connections with others.

3. Growth and competition models

Relationships between species play a vital role in any ecological ecosystem. There is a plethora of various kinds of short- and long-term interspecific relationships, where the effect on the species involved ranges from unequivocally beneficial to unequivocally harmful. Some relationships are indirect (e.g. competition for resources) or characterized by short-term interactions between organisms (e.g. prey-predator interactions). Other interspecific relationships take the form of long-term interactions, such as parasitism (a symbiotic relationship between two species that is beneficial to one of the species and harmful to the other) and symbiotic mutualism (i.e. long-term mutually beneficial interactions).

The main focus of this chapter is on models for spatial growth and competition. In particular, we give an overview of models and theory related to Paper II. In short, in Paper II we show that for a special case of the so-called Pólya urn model with competition long-term coexistence has probability zero. This particular urn model can be reformulated as a spatial growth process with competition. Section 3.1 contains an overview of related spatial growth models. In Section 3.2, we present Pólya urn models.

3.1 Spatial growth models

One of the earliest models for spatial growth is the Eden model. It was introduced by Eden (1961), who proposed it as a model of the growth of cell colonies on the plane. In the Eden model, a colony consisting of homogeneous cells is considered. As cells divide, the colony grows spatially. In his 1961 paper, Eden was primarily interested in the structural properties of large cell colonies. To be precise about the model, it describes a discrete-time Markov stochastic process in $\{0, 1\}^{\mathbb{Z}^2}$ where each time step corresponds to the division of a cell. Each site of \mathbb{Z}^2 is either occupied (1) or vacant (0). In Eden (1961), there is initially a single occupied site. In each time step, precisely one vacant site becomes occupied. The site that becomes occupied is selected uniformly at random among

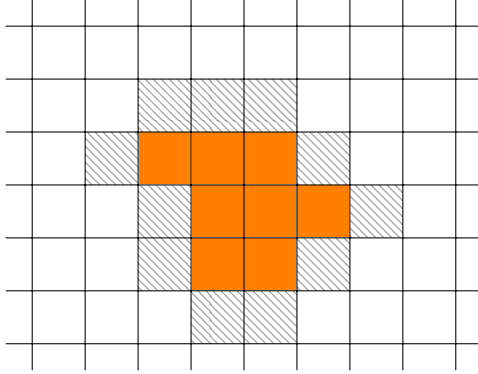


Figure 3.1: The Eden model. Orange sites are occupied, white sites are vacant. The vacant sites that might become occupied in the next step are marked with gray stripes. Each of these sites has an equal probability of becoming occupied in the next step.

the sites that are adjacent to at least one occupied site (we say that two sites $x, y \in \mathbb{Z}^d$, $d \geq 1$, are adjacent if $|x - y| = 1$). Once occupied, a site cannot become vacant again. That is, in this model cells are immortal. An example of this process is displayed in Figure 3.1.

The Eden model is closely related to first passage percolation (FPP). A FPP model describes the spread of an entity through an environment. This entity could, for instance, be an infectious pathogen, a fluid percolating through a medium or a species extending its habitat. The FPP model was introduced by Hammersley and Welsh (1965) and is defined as follows. Let G be a connected graph (with a finite or countable infinite vertex set). Classify each vertex of G as either initially occupied or initially vacant. The entity spreads from occupied to vacant sites along the edges of G . Once a site has been occupied it remains so forever. To construct the time dynamics of this process, assign a random passage time τ_e to each edge e of G . Here τ_e represents the time it takes for the entity to traverse e . The passage times are often assumed to be i.i.d realizations of some non-negative random variable, and here we mainly restrict our attention to this case. A path Γ on G is an ordered sequence of edges e_1, e_2, \dots, e_k of G such that e_i and e_{i+1} share precisely one end vertex, $i = 1, \dots, k - 1$, and the passage time of Γ is

$$T(\Gamma) = \sum_{i=1}^k \tau_{e_i}.$$

Similarly, the passage time between two vertices $x, y \in \mathbb{Z}^d$ is

$$T(x, y) = \inf\{T(\Gamma) : \Gamma \text{ is a path from } x \text{ to } y\}$$

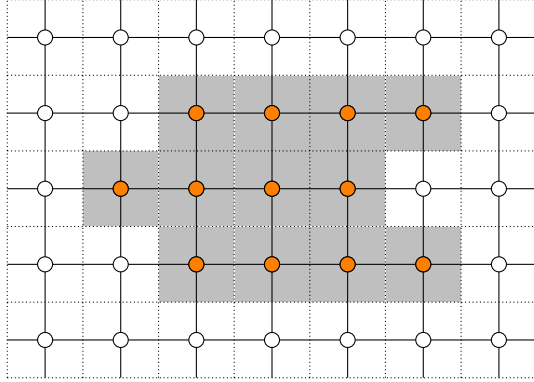


Figure 3.2: First passage percolation on the square lattice \mathbb{Z}^2 . Orange vertices are occupied, white vertices are vacant. In the corresponding process on \mathbb{R}^2 , the points of the gray area are occupied and the points of the white area are vacant.

(with $T(x, x) = 0$). A vertex y is occupied at time $t \geq 0$ if $T(x, y) \leq t$ for some initially occupied vertex x . In Figure 3.2 a FPP process on the square lattice \mathbb{Z}^2 is shown.

Considerable interest has been directed towards FPP on \mathbb{Z}^d , $d \geq 2$. This includes the so-called shape results, which describe the asymptotic shape of the growing entity. To formalise this, we first point out that a FPP process on \mathbb{Z}^d can be viewed as the spread of some entity in \mathbb{R}^d ; for $x \in \mathbb{Z}^d$, let the points of the unit square $x + [-\frac{1}{2}, \frac{1}{2}]^d$ be occupied if x is occupied and vacant if x is vacant. This is illustrated in Figure 3.2. For $t \geq 0$, let $B(t)$ be the set of points of \mathbb{R}^d that are occupied at time t . The asymptotic shape of $B(t)$ as $t \rightarrow \infty$ has been studied under different assumptions on the passage time distribution (see e.g. Cox and Durrett (1981); Boivin (1990)). For ease of exposition, we present a shape result only in the special case where the passage times are realisations of exponential i.i.d. random variables, which is known as the Richardson model. This model was introduced (among other, more general models) by Richardson (1973). For this model there is a deterministic set $\mathcal{B} \subset \mathbb{R}^d$ such that for any $\varepsilon > 0$

$$P \left((1 - \varepsilon)\mathcal{B} \subset \frac{1}{t}B(t) \subset (1 + \varepsilon)\mathcal{B} \right) \rightarrow 1$$

as $t \rightarrow \infty$ (Richardson, 1973, Thrm 2). The limiting shape \mathcal{B} is bounded, closed, convex, symmetric about the axes of \mathbb{R}^d and a strict superset of $\{0\}$. Other than this, little is known about the asymptotic shape \mathcal{B} . It is believed to be strictly convex, but so far it remains a conjecture. Alm and Deijfen (2015) investigated the asymptotic shape via simulations for a number of passage time distributions, including exponential passage times.

3.1.1 Competition in spatial growth

The FFP model described above can be extended to a multi-type setting. In this section we consider the two-type Richardson model on \mathbb{Z}^d , $d \geq 2$. This model was first studied by Häggström and Pemantle (1998), who used it to establish existence results for so-called geodesics (a geodesic is a time-minimising path that, heuristically, can be thought of as a fast “highway”). In a two-type FFP model there are two types of entities. At time zero, each site is either vacant or occupied by precisely one of the two types. As time progresses, the two entities spread along the edges of the nearest-neighbour lattice \mathbb{Z}^d at exponential rates. The passage times are independent, and for type i the passage times are exponentially distributed with intensity $\lambda_i > 0$, $i = 1, 2$. If a vacant site is reached by type i it immediately becomes occupied by i , $i = 1, 2$. That is, a vacant site becomes occupied by type i at rate λ_i times the number of adjacent sites that are occupied by type i . A site can only be occupied by at most one type. As in the single type case, once a site is occupied by type i it remains occupied by i forever, $i = 1, 2$.

The two-type Richardson model can be viewed as a model for competition between the types. We say that coexistence occurs if both type 1 and 2 ultimately occupy an infinite number of sites. Heuristically, coexistence occurs if no type gets completely surrounded by sites occupied by the other type. Häggström and Pemantle (1998) showed that, in the two-type Richardson model (i.e. FFP with exponential passage times) on \mathbb{Z}^2 coexistence occurs with positive probability if the passage times of the two types have the same intensity. This holds for any finite initial configuration of occupied sites, except for the trivial case where one type “surrounds” the other at time 0. This result was extended to coexistence of multiple types by Hoffman (2008).

For the two-type Richardson model on \mathbb{Z}^d with different passage time intensities, coexistence is believed to occur with probability 0. This conjecture was partially solved by Häggström and Pemantle (2000), who showed that if the passage time intensities λ_1 and λ_2 satisfy $\lambda_1 \neq \lambda_2$ coexistence occurs with probability zero, except for at most countably many values of $\frac{\lambda_1}{\lambda_2}$. Ahlberg et al. (2020) showed that, for the two-type Richardson model on the half-plane square lattice $\mathbb{Z} \times \mathbb{N}$, coexistence occur with probability 0 if $\frac{\lambda_1}{\lambda_2} \neq 1$ if each type initially occupies one site on the horizontal axis.

In Paper II, we study a multi-type competition model that is closely related to the Richardson model on \mathbb{Z}^2 . In this model, which was introduced in a more general form by Ahlberg et al. (2019), $k \in \mathbb{N}$ types of entities spread on the nearest-neighbour lattice \mathbb{Z}^2 according to the following rules. If a vacant site has precisely one adjacent site that is occupied by type i it becomes occupied by type i at Poisson rate 1, $i = 1, \dots, k$. If a vacant site has at least two adjacent sites that are occupied by type i then it immediately becomes occupied by type i . Ahlberg et al. (2019) showed that, for the two-type model, coexistence occurs with probability 0. The planar growth model studied in Paper II is equivalent to a Pólya urn model with competition (for a description of the connection

between these models, see Ahlberg et al. (2019) or Paper II).

3.2 Pólya urn models

A Pólya urn model describes a type of self-reinforcing stochastic process. The original Pólya urn model was proposed by Eggenberger and Pólya (1923). Their model, which describes a stochastic process in discrete-time, is defined as follows. Consider an urn that contains some number of black balls and some number of red balls. Pick a ball selected uniformly at random from the urn and put it back together with an extra ball of the same colour. In each time step, this procedure is repeated. This gives rise to a reinforcement mechanism; in each time step, the colour of the selected ball is reinforced by one additional ball. Let X_n be the proportion of red balls in the urn right after the n th draw, $n \in \mathbb{N}$. It is known that X_n has an almost sure limit X as $n \rightarrow \infty$. The limiting random variable X has distribution $\beta(a, b)^1$, where a and b are the initial numbers of red and black balls (see for instance Pemantle (2007), Thrm 2.1).

There is a vast literature on generalizations of the basic urn model described above, which we only touch on briefly here. Instead of a single urn with balls of two colours, we may consider a model with several urns and colours. Another generalization concerns the replacement rule of the original Pólya urn model; instead of putting the selected ball back together with an extra ball of the same colour, we add some other constellation of balls to the urns. Particularly useful techniques for analysing urn model are branching process embeddings (Athreya and Karlin, 1968) and martingale methods via stochastic approximation. For an overview of stochastic approximation and generalisations of the original basic urn model we refer the reader to Pemantle (2007).

Pólya urn models have many applications. One example relates to dynamical networks via the preferential attachment model. A preferential attachment network grows dynamically in time. When a new vertex joins the network, it connects to some of the existing vertices. There are various attachment schemes in the literature. Often, the larger the degree of an existing vertex, the more likely it is to connect to the newly arrived vertex, which is known as the rich get richer effect. Barabási and Albert (1999) proposed a preferential attachment model where the probability that an existing vertex is selected is proportional to its degree. The evolution of the empirical degree distribution of a preferential attachment network can be fully captured by a dynamic urn model. To construct the urn process each vertex is equipped with an urn. Initially, the number of balls in an urn is given by the degree of the corresponding vertex. Selecting an existing vertex u and connecting it to a newly arrived vertex v corresponds to selecting a ball from the urn of u . The selected ball is returned to its urn together with one extra ball in the urn of u and one in the urn of v .

¹Here $\beta(a, b)$ denotes the beta distribution with density $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$, where Γ is the gamma function.

For this coupling the empirical degree distribution of the preferential attachment network coincides with the empirical distribution of the number of balls contained in the urns. The preferential attachment model gives rise to graphs with asymptotically scale-free degree distribution (Barabási and Albert, 1999; Bollobás et al., 2001). A network is scale-free if its degree distribution follows a power law. Roughly speaking, this means that there is some $\alpha > 1$ such that the fraction of vertices with degree k is proportional to $k^{-\alpha}$ for large $k \in \mathbb{N}$.

Stochastic processes with reinforcement can also be used to model market adoption of competing technologies (Arthur, 1989). Market adoption is often subject to positive feedback, which means that the larger the market share of a technology, the more likely it is to be adopted by new users. As a technology is adopted, it attracts investments that leads to improvements, which in turn results in further adoption. Via this feedback loop, even a small advantage for an emerging technology can have a significant effect on its eventual market share. The reinforcement mechanism may give rise to a so-called lock-in effect, which is a phenomenon where one technology comes to dominate the market and thereby locks other technologies out. Arthur (1989) used an urn processes to investigate, among other things, under what circumstances market shares are locked-in to the monopoly of a technology.

4. Overview of papers

4.1 Paper I

In Paper I we study SIR-epidemics on an extended version of the standard configuration model (described in Section 2.2) with group structure, where the neighbours of an individual may in turn be neighbours. Graphs generated according to the standard configuration model are asymptotically unclustered, unlike many real world social networks (Newman, 2003). This means that in a graph generated from the configuration model, short loops are rare and the graph is locally treelike. Here we study an SIR epidemic on a variant of the configuration model where group structure is incorporated in the form of triangles (i.e. circuits with three edges) with non-overlapping edges. We present (implicit) expressions for the basic reproduction number R_0 , the probability of a major outbreak and the expected final size. Additionally, we analyse how vaccination affects the probability of a major outbreak and the final size. To this end, we investigate the spread of a contagious disease in a population where a fraction of the population is vaccinated against the disease. The vaccinated individuals are selected uniformly at random and the vaccine is assumed to be perfect, which means that a vaccinated individual is fully immune and cannot contract the disease. We show that the critical vaccination coverage, which is the proportion of the population that needs to be vaccinated to reach herd immunity, equals $1 - 1/R_0$. This is typically the case in the unclustered case, but does not hold in general for graphs with clustering (see e.g. Ball et al. (2016)).

In the configuration model with clustering considered here, there are two types of edges: single edges and triangle edges. A graph of this model is constructed by assigning a single edge degree and a triangle edge degree to each vertex. The triangle degree of a vertex specifies the number of triangles that the vertex in question is a member of. SIR epidemics on configuration graphs with clustering have earlier been studied under the assumption of homogeneous infectivity (that is, an infected individual infects each of its neighbours independently with some fixed probability) by Miller (2009), and fixed transmission and recovery rates (Volz et al., 2011). In Paper I we weaken these assumptions in two regards. First, we allow for differences in infectivity between triangle and single edge neighbours. Second, we obtain the above-mentioned results for a model that

allows for heterogeneity in individual infectivity. This means that each vertex is equipped with a (random) weight that determines its probability of disease transmission. Given its weight, (if infected) an individual transmits the disease to its neighbours independently with a probability that is determined by the weight. An important potential source of such heterogeneity is differences between individuals in the length of the infectious period. To obtain the results described above we use branching process approximations. Since branching process approximations generally rely on the locally treelike structure of the graph, obtaining results on epidemics on networks with clustering requires a more involved approach than for epidemics on unclustered networks.

Open questions. We restrict the analysis to groups of three, but the configuration model with clustering could readily be extended to a model with larger group sizes. We expect a similar approach as the one used in Paper I to work for models with larger groups. It is, however, likely to get increasingly complicated with larger group sizes which would make tractable analysis challenging.

4.2 Paper II

Random processes with reinforcement is an important class of stochastic processes, where previous events may reinforce the future evolution of the process. Pólya's urn model is a classical example of a reinforcement process. In Paper II we study a multi-type variant of the Pólya urn model with interaction and competition, introduced by Ahlberg et al. (2019). The time evolution of the number of balls in this model corresponds to the spread in a spatial growth model on the square lattice, which we describe in some more detail below. In the urn model, the urns are located on the vertices of an underlying graph that determines the replacement rules. Balls of different types interact by annihilation (on a one-to-one basis) upon contact. In this paper, we show that for $k \geq 2$ types on a cycle graph coexistence occurs with probability zero, thus settling a conjecture posed by Ahlberg et al. (2019). Ahlberg et al. (2019) studied two-type competition and coexistence and showed that in the two-type case, coexistence occurs with probability 0 for any (finite and connected) graph. They also showed by explicit example that this result does not extend to general (finite and connected) graphs for $k \geq 3$ types. That is, in general the anatomy of the underlying graph does affect the possibility of coexistence.

In this urn model, the balls interact according to the following dynamics. Let G be a finite and connected graph and suppose that precisely one urn is located on each vertex of G . We assume that the initial configuration is such that no urn contains balls of more than one type. In each step, a ball is selected uniformly at random. The selected ball is then returned to its urn and, in addition, a copy of the selected ball is placed in each of the neighbouring urns (according to the structure of the underlying graph). If a ball is added to an urn that contains balls of a different type then the newly added ball and one of the existing balls immediately annihilate (destroy each other on a one-to-one basis). Coexistence

occurs if no type that is initially present is fully eradicated.

As mentioned above, the urn process can be viewed as a spatial growth model with reinforcement. The connection between these two processes was first described by Ahlberg et al. (2019). The dynamics of the spatial model are similar to the Richardson model in Section 3.1. The difference between these two models lies in the reinforcement mechanism: if a vacant site has at least two adjacent sites that are occupied by the same type then it immediately becomes occupied by that type. To obtain these results, we analyse an auxiliary auto-annihilative process on the cycle graph.

Open questions. For $k \geq 3$ types, the conditions under which coexistence has positive probability remain an open question for most underlying graphs. Additionally, in the urn model considered in Paper II a ball is selected uniformly at random. It would be interesting to investigate a model where the probability that a ball is selected depends on its type and/or its urn.

4.3 Paper III

One of the most accessible characteristics of a contagious disease outbreak is the growth rate of the number of infected individuals. Arguably, it is also one of the quantities that best conveys the severity of the epidemic. In the early stages of an outbreak this number often grows exponentially. This has been seen both empirically (Nishiura and Chowell, 2014; Dye et al., 2015) as well as in many theoretical models. This paper is concerned with the growth rate of SIR epidemics on graphs with group structure. Unlike Paper I, where groups of size three are present in the underlying social network, here we allow for groups of unbounded size. We show that, in the supercritical case, the Malthusian parameter solves a version of the Euler-Lotka equation in Section 1.3.3. In addition, we establish a lower bound on the number of infected cases before the coupling of the epidemic process and approximating branching process breaks down via an upper bound on the coupling error between the processes.

In Paper III, the underlying social network is represented by an extended version of the random intersection graph of Karoński et al. (1999) that is presented in Section 2.3. This graph model deviates from the original random intersection graph model in that the probability that a vertex is a member of a clique is not homogeneous between vertices and cliques. To construct a graph according to this model, each vertex and clique is assigned a (random) non-negative weight in $[0, \infty)$. The probability that a vertex with weight a belongs to a clique with weight b is given by $\frac{ab}{n} \wedge 1$. This results in graphs with (asymptotic) mixed Poisson degree distribution.

SIR epidemics on graphs of this model have earlier been studied by Ball et al. (2014), who established results on the probability of a major outbreak and the final size. We point out that the real-time dynamics of epidemics on graphs with clustering are generally more difficult to analyse than the ultimate outcome (e.g.

the occurrence of a major outbreak and the final size) since it requires a more involved approach where the exact chain of transmission is considered. Related results have also been presented by Pellis et al. (2011) who considered epidemics in a population with small highly connected communities of bounded size. They mainly focus on methods to estimate the so-called household reproduction number from the observed real-time growth rate and also obtain heuristic results for the Malthusian parameter as the solution to the Euler-Lotka equation. Previous results only deal with the case where the group sizes are bounded and extensions to models with unbounded group sizes are not straightforward. Paper III considers the unbounded case and provides rigorous proofs of these results.

Open questions. The methods used in Paper III rely on the tree-like structure (at the level of cliques) of the underlying graph. The dynamics of epidemic spread on graphs that are not asymptotically tree-like are in many regards not well understood and generally challenging to analyse. Additionally, while we provide an implicit expression for the Malthusian parameter α we have not looked into methods for explicitly calculating α . This could involve approximate methods for the transmission time between two nodes in small to medium size cliques (c.f. Pellis et al. (2011)) as well as large cliques (see Janson (1999)). We believe that the methods employed here can be applied to a more general class of models, for example to epidemics on graphs with other community structures than fully connected cliques (e.g. van der Hofstad et al. (2022)) and/or models where transmission rates within a community depend on community-specific attributes such as community size (c.f. Paper IV).

4.4 Paper IV

Many models for epidemics on networks assume homogeneity in contacts between relationships. In other words, potentially infectious interactions are assumed to follow the same pattern across all edges. Much interest has been directed towards incorporating a tendency that the number of per-relationship interactions decreases from additional connections. In the case of sexual contacts, this phenomenon is known as coital dilution and has been found to hold empirically (see e.g. Althaus et al. (2015); Blower and Boe (1993); Nordvik and Liljeros (2006); Britton et al. (2007) and the references in Sawers et al. (2011)). This tendency has, for instance, been modelled by considering epidemics on weighted networks (Kamp et al., 2013; Moslonka-Lefebvre et al., 2012; Britton et al., 2011), via models for sexual interactions with casual and steady relationships where the contact pattern depends on the type of the relationship (Britton et al., 2007), and by assuming explicit dependence between contact rates and degree (Baxter and Timár, 2021). Relatedly, Komjáthy et al. (2020) studied transmission on inhomogeneous spatial random graphs with power law degree distributions where each edge has a transmission cost/time that depends on the expected degrees of its terminal vertices.

Here, we study a model for SIR epidemics on networks with degree-dependent contact rates where, roughly speaking, vertices with high degrees tends to have lower per-neighbour contact frequency but higher overall contact frequency than individuals with small degrees. The underlying social network is generated via a special case of the inhomogeneous random graph model of Bollobás et al. (2007). In this model, graphs are constructed by assigning a weight to each vertex and letting each potential edge be present with an occupational probability that is calculated via a symmetric function φ : an edge with whose terminal vertices have weights w_1 and w_2 is present with probability $\varphi(w_1, w_2)/n \wedge 1$ where n is the number of vertices of the graph. We assume that φ is increasing. Thus, the weight of an individual has the interpretation of a measure of its proclivity to form social relationships. The dependence between the degrees and contact rates enters via the vertex weights: two neighbours with weight w_1 and w_2 interact at a Poisson rate $\kappa(w_1, w_2)$ where κ is a symmetric function. We make assumptions that, heuristically, entail that individuals with many acquaintances tend to have fewer contacts per acquaintance but more contacts overall than individuals with few acquaintances. Formally, this behaviour is achieved by assuming that the functions φ and κ satisfy suitable monotonicity requirements.

In this paper, we use branching process approximations to show that the probability of a major outbreak and the final size can be found via the survival probability of the approximating forward and backward branching process, respectively. Moreover, we prove that reducing heterogeneity in transmission rates (while keeping the expected number of overall contacts of an individual constant) yields a higher value of the basic reproduction number R_0 . As an important special case, our result implies that ignoring heterogeneity by assuming constant transmission rates results in an overstated R_0 . Our result substantiates earlier observational findings (see e.g. Nordvik and Liljeros (2006); Britton et al. (2007, 2011)) concerning the relationship between R_0 and heterogeneity in transmission. In addition, we demonstrate by explicit example that corresponding inequalities for the probability of a major outbreak and the final size do not hold in general.

Open questions. As we show in Paper IV, reducing heterogeneity in contact rates does not always result in a larger probability of a major outbreak and final size. It would be interesting to further investigate the relationship between contact heterogeneity and these two quantities. In paper IV, we restrict our analysis to the case where the per-neighbour contact rates of individuals are negatively correlated with their degree, while the overall contact rates are positively correlated with their degree. One might, however, allow for other relations between degree and contact frequency.

5. Sammanfattning

COVID-19-pandemin har satt strålkastarljuset på vikten av att kunna förstå och förutsäga spridningen av smittsamma sjukdomar. Matematiska modeller har spelat en central roll i de beslut om kontrollåtgärder som vidtagits för att hindra virusets spridning. I Sverige har Folkhälsomyndigheten bidragit till beslutsunderlaget vid beslut om och utvärdering av smittohämmande åtgärder.

Den här avhandlingen handlar om stokastiska modeller för spridningen av smittsamma sjukdomar. Den består av fyra artiklar numrerade I till IV. Tyngdpunkten ligger huvudsakligen på hur sociala strukturer inverkar på sjukdomsspridning. För att införliva denna struktur i en smittspridningsmodell används ofta en graf som representerar befolkningens underliggande sociala nätverk, där varje nod motsvarar en individ och en kant motsvarar en social relation som kan resultera i smittoöverföring. Sociala nätverk är vanligtvis stora och har en invecklad struktur som ofta är okänd och svår att kartlägga. Detta medför att även i de fall där strukturen hos ett socialt nätverk är fullständigt känd lämpar det sig i praktiken ofta inte för analys. Ett sätt att kringgå dessa problem är att låta en slumpgraf (en slumpmässigt genererad graf) representera det sociala nätverket. I den akademiska litteraturen förekommer en uppsjö av slumpgrafmodeller som fångar olika egenskaper hos nätverk. Några exempel på sådana egenskaper är gradfördelningen och gruppstrukturer.

I Artikel I, III och IV använder vi den så kallade SIR-modellen för att beskriva personers smittorelaterade faser, där SIR står för "Susceptible Infectious Recovered". I SIR-modellen klassificeras varje person som antingen motaglig (susceptible), smittsam (infectious), eller återhämtad (recovered), beroende på dess hälsostatus. Vanligtvis antas det att initialt är ett fåtal personer smittsamma och övriga mottagliga. En mottaglig person kan smittas om hen har kontakt med en smittsam person och blir då i sin tur genast smittsam. En smittsam person kan så småningom återhämta sig och får då permanent immunitet. Detta innebär att när en person väl klassificerats som återhämtad kan hen aldrig bli mottaglig eller smittsam.

Ett centralt tal inom epidemimodellering är det basala reproduktionstalet R_0 . Det anger antalet personer som en typisk smittsam person förväntas smitta om resten av populationen är mottaglig. Talet R_0 är en så kallad tröskelparameter som avslöjar om en smittsam sjukdom har potential att få allmän spridning; ett

stort utbrott är möjligt om och endast om R_0 ligger över den kritiska tröskeln 1.

I den första artikeln undersöker vi SIR-epidemier i en population med gruppstruktur, där en persons vänner i sin tur kan känna varandra. Vi presenterar ett (implicit) uttryck för reproduktionstalet R_0 , sannolikheten för en stort utbrott och, givet att ett stort utbrott inträffar, den förväntade andelen av befolkningen som insjuknar. Vi analyserar även hur vaccination påverkar sannolikheten för stort utbrott och hur stort det i så fall blir. Det sociala nätverket representeras av en generaliserad version av konfigurationsmodellen som, till skillnad från den klassiska konfigurationsmodellen med dess lokala trädstruktur, har grupper av storlek tre (d.v.s. kanter som formar trianglar) med icke-överlappande kanter. Vi analyserar spridningen av en smittsam sjukdom i en befolkning där en given andel av befolkningen vaccineras mot sjukdomen med ett vaccin som ger fullständig och permanent immunitet. De vaccinerade individerna väljs uniformt slumpmässigt. Vi visar att den kritiska vaccinationsgraden, som är den andel av befolkningen som behöver vaccineras för att uppnå flockimmunitet, är $1 - 1/R_0$.

Artikel II handlar om en tävlingsmodell som kan ses som en spridningsmodell för konkurrerande smittor. Modellen har två ekvivalenta formuleringar: den kan antingen ses som en spridningsmodell i planet med en förstärkningsmekanism, eller som en version av Pólyas urnmodell (se avsnitt 3.2) med interaktion. I modellen finns k olika typer av smittor. Ahlberg et al. (2019) har tidigare (bland annat) visat att för $k = 2$ typer inträffar samexistens med sannolikhet 0, vilket innebär att sannolikheten att en typ så småningom helt utplånas är 1. I Artikel II visar vi att sammexistens har sannolikheter 0 för $k \geq 2$ typer.

I urnversionen av denna process är urnorna utplacerade på noderna hos en cyklisk graf och återläggningsreglerna bestäms av den underliggande grafen. I varje steg väljs en befintlig boll slumpmässigt. Den valda bollen läggs sedan tillbaka i sin urna och en kopia av den valda bollen adderas till varje grannurna. Bollar av olika typ interagerar genom ömsesidig utplåning enligt en "en-mot-en"-princip; om en boll adderas till en urna som redan innehåller en eller flera bollar av en annan typ utplånas den genast tillsammans med precis en befintlig boll. I den spatiala versionen av modellen sprider sig k typer av smittor mellan punkterna i det tvådimensionella heltalsgittret \mathbb{Z}^2 . Varje punkt i \mathbb{Z}^2 har antingen status vakant eller är ockuperad av typ i för någon typ $i = 1, \dots, k$. En punkt kan endast ockuperas av en typ, och när en punkt väl har blivit ockuperad kan den inte bli vakant. Varje typ sprider sig till angänsande vakanta punkter (d.v.s. mellan punkter med euklidiskt avstånd 1) med Poisson-intensitet 1. Modellen har dock en förstärkningsmekanism; en vakant punkt som har minst två angänsande punkter som är ockuperade av samma typ blir omedelbart ockuperad av typen i fråga. Samexistens i urnversionen av processen motsvarar att ingen typ helt omringas av andra typer i den spatiala versionen.

I den tredje artikeln undersöker vi tillväxthastigheten hos en SIR-epidemi i en population med gruppstruktur. I den inledande fasen av ett utbrott växer ant-

alet sjukdomsfall ofta exponentiellt. Detta har observerats både empiriskt och i många teoretiska modeller. Till skillnad från Artikel I, där grupper av storlek tre förekommer, tillåter vi här grupper av obegränsad storlek. Vi visar att i det superkritiska fallet är den Malthusiska parametern lösningen till en version av Euler-Lotka-ekvationen i avsnitt 1.3.3. Dessutom etablerar vi en undre gräns för antalet personer som smittas innan kopplingen av epidemiprocessen och den approximerande förgreningsprocessen bryter samman via en övre gräns för kopplingsfelet mellan processerna. Vi inkorporerar gruppstrukturen i modellen genom att använda en slumpmässig skärningsgraf för att modellera befolkningens sociala nätverk. Den asymptotiska gradfördelning för grafer genererade enligt denna modell ges av en blandad Poissonfördelning.

I Artikel IV analyserar vi en modell med heterogena kontaktfrekvenser. I många modeller för epidemier görs antagandet att potentiellt smittsamma interaktioner är homogena mellan relationer, vilket innebär att de följer samma mönster i alla sociala förbindelser. Detta är i många fall ett förenklande antagande. Exempelvis tenderar ofta personer med många sexuella partners att genomföra färre akter per relation än individer med få partners (Althaus et al., 2015; Blower and Boe, 1993; Nordvik and Liljeros, 2006; Britton et al., 2007), vilket har betydelse för spridningen av sexuellt överförbara infektioner. Här lättar vi på antagandet om homogena interaktioner genom att låta kontaktintensiteten mellan två bekanta personer vara beroende av hur många övriga relationer personerna har. Vi använder en inhomogen slumpgraf för att modellera befolkningens sociala struktur. Vi gör antaganden som (något förenklat) innebär att ju fler relationer en person har, desto högre total kontaktintensitet men lägre kontaktintensitet per relation tenderar hen att ha. Vi presenterar sannolikheten för ett stort utbrott och utbrottets totala storlek. Vi visar att en minskning i kontakterintensiteternas heterogenitet (samtidigt som det förväntade antalet kontakter hålls konstant) resulterar i en minskning av det basala reproduktionstalet R_0 . Speciellt innebär detta att förenklande antaganden som ignorerar heterogenitet i kontaktfrekvens här resulterar i en modell där R_0 överskattas. Vidare demonstrerar vi genom explicita exempel att motsvarande olikheter för sannolikheten att ett större utbrott inträffar och den slutliga storleken inte håller i allmänhet.

Bibliography

- Ahlberg, D., M. Deijfen, and C. Hoffman (2020). The two-type Richardson model in the half-plane. *The Annals of Applied Probability* 30(5), 2261–2273.
- Ahlberg, D., S. Griffiths, S. Janson, and R. Morris (2019). Competition in growth and urns. *Random Structures & Algorithms* 54(2), 211–227.
- Alm, S. E. and M. Deijfen (2015). First passage percolation on \mathbb{Z}^2 : A simulation study. *Journal of statistical physics* 161(3), 657–678.
- Althaus, C. L., M. Choisy, S. Alizon, and C. group (2015). How sex acts scale with the number of sex partners: evidence from chlamydia trachomatis data and implications for control. *PeerJ PrePrints* 3, e940v2.
- Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The economic journal* 99(394), 116–131.
- Athreya, K. B. and S. Karlin (1968). Embedding of urn schemes into continuous time markov branching processes and related limit theorems. *The Annals of Mathematical Statistics* 39(6), 1801–1817.
- Ball, F. and O. D. Lyne (2001). Stochastic multi-type sir epidemics among a population partitioned into households. *Advances in Applied Probability* 33(1), 99–123.
- Ball, F., L. Pellis, and P. Trapman (2016). Reproduction numbers for epidemic models with households and other social structures II: Comparisons and implications for vaccination. *Mathematical Biosciences* 274, 108 – 139.
- Ball, F., D. Sirl, and P. Trapman (2014). Epidemics on random intersection graphs. *The Annals of Applied Probability* 24(3), 1081–1128.
- Barabási, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *science* 286(5439), 509–512.
- Baxter, G. and G. Timár (2021). Degree dependent transmission rates in epidemic processes. *Journal of Statistical Mechanics: Theory and Experiment* 2021(10), 103501.

- Blower, S. M. and C. Boe (1993). Sex acts, sex partners, and sex budgets: implications for risk factor analysis and estimation of hiv transmission probabilities. *Journal of acquired immune deficiency syndromes* 6(12), 1347–1352.
- Boivin, D. (1990). First passage percolation: the stationary case. *Probability theory and related fields* 86(4), 491–499.
- Bollobás, B., S. Janson, and O. Riordan (2007, August). The phase transition in inhomogeneous random graphs. *Random Struct. Algorithms* 31(1), 3–122.
- Bollobás, B. e., O. Riordan, J. Spencer, and G. Tusnády (2001). The degree sequence of a scale-free random graph process. *Random Structures & Algorithms* 18(3), 279–290.
- Bollobás, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics* 1(4), 311 – 316.
- Bollobás, B., S. Janson, and O. Riordan (2011). Sparse random graphs with clustering. *Random Structures & Algorithms* 38(3), 269–323.
- Britton, T., M. Deijfen, and F. Liljeros (2011). A weighted configuration model and inhomogeneous epidemics. *Journal of statistical physics* 145(5), 1368–1384.
- Britton, T., M. K. Nordvik, and F. Liljeros (2007). Modelling sexually transmitted infections: the effect of partnership activity and number of partners on R_0 . *Theoretical population biology* 72(3), 389–399.
- Cox, J. T. and R. Durrett (1981). Some limit theorems for percolation processes with necessary and sufficient conditions. *The Annals of Probability*, 583–603.
- Diekmann, O., H. Heesterbeek, and T. Britton (2013). *Mathematical tools for understanding infectious disease dynamics*, Volume 7. Princeton University Press.
- Dye, C., B. Aylward, P. Barboza, L. Bawo, E. Bertherat, P. Bilivogui, I. Blake, R. Brennan, S. Briand, J. Chakauya, K. Chitala, R. M. Conteh, A. Cori, A. Croisier, D. Jean-Marie, B. Diallo, C. A. Donnelly, T. Eckmanns, N. M. Ferguson, and Z. Yoti (2015, 01). Ebola virus disease in west africa—the first 9 months. *The New England journal of medicine* 372, 189.
- Eden, M. (1961). A two-dimensional growth process. In *Proc. 4th Berkeley Symp. Math. Statist. Probability*, Volume 4, pp. 223. Univ. California Press, Berkeley.
- Eggenberger, F. and G. Pólya (1923). Über die statistik verketteter vorgänge. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* 3(4), 279–289.
- Erdős, P. and A. Rényi (1959, 1959). On random graphs i. *Publicationes Mathematicae Debrecen* 6, 290–297.

- Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics* 30(4), 1141–1144.
- Haccou, P., P. Jagers, and V. Vatutin (2005). *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press.
- Hägglström, O. and R. Pemantle (1998). First passage percolation and a model for competing spatial growth. *Journal of Applied Probability* 35(3), 683–692.
- Hägglström, O. and R. Pemantle (2000). Absence of mutual unbounded growth for almost all parameter values in the two-type richardson model. *Stochastic Processes and their Applications* 90(2), 207–222.
- Hammersley, J. M. and D. J. Welsh (1965). First-passage percolation, sub-additive processes, stochastic networks, and generalized renewal theory. In *Bernoulli 1713, Bayes 1763, Laplace 1813*, pp. 61–110. Springer.
- Hoffman, C. (2008). Geodesics in first passage percolation. *The Annals of Applied Probability* 18(5), 1944–1969.
- van der Hofstad, R. (2016). *Random Graphs and Complex Networks Volume I*. <https://www.win.tue.nl/~rhofstad/NotesRGCN.pdf>.
- van der Hofstad, R. (2022). Random graphs and complex networks volume 2. Retrieved November 28 from <https://www.win.tue.nl/~rhofstad/NotesRGCN.html>.
- van der Hofstad, R., J. Komjáthy, and V. Vadon (2022). Phase transition in random intersection graphs with communities. *Random Structures & Algorithms* 60(3), 406–461.
- Iksanov, A. and M. Meiners (2015). Rate of convergence in the law of large numbers for supercritical general multi-type branching processes. *Stochastic Processes and their Applications* 125(2), 708 – 738.
- Janson, S. (1999). One, two and three times $\log n/n$ for paths in a complete graph with random weights. *Combinatorics, Probability and Computing* 8(4), 347–361.
- Kamp, C., M. Moslonka-Lefebvre, and S. Alizon (2013, 12). Epidemic spread on weighted networks. *PLoS computational biology* 9, e1003352.
- Karoński, M., E. Scheinerman, and K. Singer-cohen (1999). On random intersection graphs: The subgraph problem. *Combinatorics, Probability and Computing* 8(1-2), 131–159.
- Kermack, W. O. and A. G. McKendrick (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115(772), 700–721.

- Komjáthy, J., J. Lapinskas, and J. Lengler (2020). Stopping explosion by penalising transmission to hubs in scale-free spatial random graphs. *arXiv preprint arXiv:2004.01149*.
- Miller, J. (2009, Aug). Percolation and epidemics in random clustered networks. *Phys. Rev. E* 80, 020901.
- Moslonka-Lefebvre, M., S. Bonhoeffer, and S. Alizon (2012). Weighting for sex acts to understand the spread of STI on networks. *Journal of theoretical biology* 311, 46–53.
- Newman, M. E. (2002). Spread of epidemic disease on networks. *Physical review E* 66(1), 016128.
- Newman, M. E. J. (2003, Aug). Properties of highly clustered networks. *Phys. Rev. E* 68, 026121.
- Nishiura, H. and G. Chowell (2014). Early transmission dynamics of ebola virus disease (evd), west africa, march to august 2014. *Eurosurveillance* 19(36).
- Nordvik, M. K. and F. Liljeros (2006). Number of sexual encounters involving intercourse and the transmission of sexually transmitted infections. *Sexually transmitted diseases*, 342–349.
- Pellis, L., N. M. Ferguson, and C. Fraser (2011). Epidemic growth rate and household reproduction number in communities of households, schools and workplaces. *Journal of mathematical biology* 63(4), 691–734.
- Pemantle, R. (2007). A survey of random processes with reinforcement. *Probability surveys* 4, 1–79.
- Read, J., W. Edmunds, S. Riley, J. Lessler, and D. Cummings (2012). Close encounters of the infectious kind: methods to measure social mixing behaviour. *Epidemiology & infection* 140(12), 2117–2130.
- Richardson, D. (1973). Random growth in a tessellation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 74, pp. 515–528. Cambridge University Press.
- Sawers, L., A. G. Isaac, and E. Stillwaggon (2011). Hiv and concurrent sexual partnerships: modelling the role of coital dilution. *Journal of the International AIDS Society* 14(1), 1–9.
- Singer, K. B. (1995). *Random intersection graphs*. The Johns Hopkins University.
- Trevas, A. (2002). Malthus foiled again and again. *Nature* 418(6898), 668–670.
- Volz, E., J. Miller, A. Galvani, and L. Meyers (2011, 06). Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. 7, e1002042.

Part II

Papers