

# A Law of Large Numbers for SIR on the Stochastic Block Model: A Proof via Herd Immunity

Christian Borgs<sup>1, 3</sup>   Karissa Huang<sup>2</sup>   Christian Ikeokwu<sup>1</sup>

October 9, 2024

## Abstract

In this paper, we study the dynamics of the susceptible-infected-recovered (SIR) model on a network with community structure, namely the stochastic block model (SBM). As usual, the SIR model is a stochastic model for an epidemic where infected vertices infect susceptible neighbors at some rate  $\eta$  and recover at rate  $\gamma$ , and the SBM is a random graph model where vertices are partitioned into a finite number of communities. The connection probability between two vertices depends on their community affiliation, here scaled in such a way that the average degrees have a finite limit as the network grows. We prove laws of large numbers (LLN) for the trajectory of the epidemic to a system of ordinary differential equations over any time horizon (finite or infinite), including in particular a LLN for the final size of the infection.

Our proofs rely on two main ingredients: (i) a new coupling of the SIR epidemic and the randomness of the SBM, revealing a vector valued random variable that drives the epidemic (related to what is usually called the “force of the infection” via a linear transformation), and (ii) a novel technique for analyzing the limiting behavior of the infinite time horizon for the infection, using the fact that once the infection passes the herd immunity threshold it dies out quickly and has a negligible impact on the overall size of the infection.

## 1 Introduction

The susceptible-infected-recovered (SIR) model was introduced by Kermack and McKendrick in 1927, and is a simple model used to describe the behavior of a disease in a closed, finite population [Kermack and McKendrick \[1927\]](#). In the simplest version of the model, the population is divided into three compartments – susceptible, infected, and recovered – whose sizes can be thought of as deterministic, time-varying, continuous variables  $s = s(t)$ ,  $i = i(t)$ ,  $r = r(t)$  adding up to one. The dynamics of the infection are governed by constants  $\eta$  and  $\gamma$ , and described by a well-known system of differential equations representing a “flow of individuals” from  $S$  to  $I$  and  $I$  to  $R$  at rates  $\eta si$  and  $\gamma i$ , respectively. In the slightly more realistic stochastic setting first considered by [Bartlett \[1949\]](#), the population consists of  $n$  individuals again divided into compartments of susceptible, infected, and recovered individuals. The time dynamics of the model are now stochastic, with infected individuals infecting susceptible individuals at rate  $\eta/n$ , and infected individuals recovering at rate  $\gamma$ . As  $n$  becomes large, the stochastic model is well approximated by the deterministic one,

---

<sup>1</sup>Department of Electrical Engineering and Computer Science, University of California, Berkeley.

<sup>2</sup>Department of Statistics, University of California, Berkeley.

<sup>3</sup>Bakar Institute of Digital Materials for the Planet (BIDMaP).

mathematically expressed as a law of large numbers (LLN). This implies that on any bounded time horizon, the fraction of susceptible, infected, and recovered individuals converge to deterministic functions  $s(t)$ ,  $i(t)$ , and  $r(t)$  given by the solutions of the usual differential equations, see, e.g., [Von Bahr and Martin-Löf \[1980\]](#) for the study of a somewhat related version in discrete time (the so-called Reed-Frost model) with a generalization to the Kermack-McKendrick SIR model, or the excellent review article [Darling and Norris \[2008\]](#) for a short, streamlined proof based on Gronwall’s inequality and Doob’s  $L^2$ -inequality.

It is well known that the homogeneity assumption inherent in the original Kermack and McKendrick model is not realistic in practice, as contact rates between individuals vary by age and other demographic factors [Jacquez et al. \[1988\]](#). While this can be, to some extent, addressed by creating more compartments stratified by these demographic factors, another property which a realistic model should take into account is the underlying graph structure of the contact network. Indeed, it is well known that the underlying graph structure plays a crucial role in the dynamics of the infection throughout the graph. As such, there has been interest in studying the SIR model on various random graphs including the Erdős-Rényi graph [[Ball and Britton, 2022](#); [Neal, 2003](#)], graphs with a given degree sequence (configuration models) [[Barbour and Reinert, 2013](#); [Bohman and Piccollelli, 2012](#); [Decreusefond et al., 2012](#); [Janson et al., 2014](#)], and graphs with local household structure [[Ball et al., 2010](#); [House, 2012](#)] or tunable local clustering, [Britton et al. \[2008\]](#).

In this paper, we consider the SIR epidemic on the stochastic block model (SBM), a network model with community structure. In the SBM, nodes are partitioned into  $K$  communities and the distribution of edges between vertices is determined by a  $K \times K$  matrix and is dependent on the group membership of the endpoints [Holland et al. \[1983\]](#). The SBM has appeared independently in many contexts in the statistics, computer science, and mathematics literature and there are many examples of networks with known community structure [[Bollobás et al., 2007](#); [Bui et al., 1987](#); [Girvan and Newman, 2002](#); [Holland et al., 1983](#)]. In addition there is a vast literature on the use of the stochastic block model in community detection problems [[Abbe, 2018](#); [Lee and Wilkinson, 2019](#)].

Despite its widespread study and use, somewhat surprisingly unlike other well-studied graph models, no rigorous work has established analogous law of large numbers results for a limiting system of ordinary differential equations for the stochastic SIR epidemic on the SBM. The rigorous derivation of these is the goal of the present paper.

Our work is motivated and inspired by the proofs of LLNs for the Erdős-Rényi graph  $G(n, p)$  and variations of  $G(n, p)$  involving the dropping and rewiring of edges as the infection spreads, as well as LLNs for graphs of a given degree sequence [[Ball and Britton, 2022](#); [Janson et al., 2014](#); [Neal, 2003](#)]. Our work differs from these in one important aspect, in all these works, the dynamics of the infection process are primarily driven by one random variable, often referred to as the force of the infection. For  $G(n, p)$  and its variants, this random variable is, roughly speaking, the number of edges between infected and susceptible vertices, and for the configuration model it is the number of not-yet explored or matched edges emanating from infected vertices.

In contrast to the configuration model where the resulting limiting equations were derived heuristically before being established mathematically rigorously, to our knowledge, there has been no conjecture for the correct differential equations describing a LLN for SIR on the SBM.

To address this, we build on ideas from [Ball and Britton \[2022\]](#) and [Janson, Luczak, and Windridge \[2014\]](#). First, we couple the randomness of the SBM and that of SIR by only exploring edges in the graph when they attempt to spread the SIR infection. More precisely, once a vertex

gets infected, we draw its degree (with a distribution that depends on the community type of the infected vertex), and then attach what we call active half-edges to it, each endowed with an exponential clock governing future infection events. Whenever these clocks click, we determine the label of the other endpoint, and “test” whether that endpoint is still susceptible before spreading the infection to a new vertex. This representation then naturally leads us to consider a LLN for three vector valued random variables: 1) the number of susceptible vertices in each community, 2) the number of infected vertices in each community and, 3) the number of not yet explored, active half-edges emanating from vertices of each community type.

Having derived this representation, both the heuristic derivation and the mathematical proof of a LLN for finite time horizons is straightforward, with proofs and quantitative error bounds following from standard techniques involving Gronwall’s inequality and Doob’s  $L^2$ -inequality as, e.g., laid out in [Darling and Norris \[2008\]](#). However, while these bounds are uniform for all times in a finite time interval  $[0, t_0]$ , the size of the error bounds turn out to be exponential in  $t_0$ , which prevents us from using the same techniques to prove a law of large numbers for an infinite time-horizon and, in particular, for what is usually called the final size of the infection.

The standard way of overcoming such a challenge is to apply a global time rescaling to the epidemic so that the entire epidemic happens on a finite time horizon [[Daley and Gani, 1999](#); [Ethier and Kurtz, 1986](#); [Kiss et al., 2017](#)]. However in the case of the SBM, because the infection rate differs depending on the community, it is unclear how to appropriately apply such a global time rescaling. To overcome this problem, we develop a new technique to prove the infinite time law of large numbers. This proof technique is based on the observation that the infection must eventually pass the herd immunity threshold and that after this point, the infection dies out quickly. Thus, for a large enough time  $t_1$  after the infection has passed herd immunity, any outbreak that occurs after time  $t_1$  has a negligible effect on the final state of the epidemic at time infinity (and, in fact, for all times between  $t_1$  and infinity). We believe that this technique should be applicable in much more generality than traditional time rescaling techniques allowing for the analysis of more models but will also greatly simplify proofs of the final size LLN for many models.

We close this introduction by briefly mentioning two alternative methods which could at least in principle be used to derive some of the results established in this paper, even though to our knowledge, they have not yet been applied to the stochastic block model. One of these methods involves the theory of local graph limits [[Alimohammadi et al., 2023](#); [Cocomello and Ramanan, 2023](#)], which can be used to study the final size of an epidemic starting from a constant fraction of initially infected vertices, provided these are chosen uniformly at random. Another method involves the method of forward and backward branching processes, see, e.g., [Barbour and Reinert \[2013\]](#). We will discuss these and other related works in [Section 2.4](#) below, where we put our work into the context of the existing literature.

This paper is organized as follows. In [Section 2](#) we define our model and state our main results, including the laws of large numbers for both the finite and infinite time horizon. In [Section 3](#), we define a multi-graph version of the SBM and then couple the SIR epidemic to the randomness of this model. As, e.g., in the case of the multi-graph vs. simple graph version of the configuration model, conditioning the multi-graph version of the SBM to be simple will give back the original simple graph version of the model, allowing us to transfer our results for the multi-graph model to the original one. We will also give the heuristic derivation of the LLN in this section. In [Section 4](#), we prove the law of large numbers in the finite time case where the number of initially infected vertices grows linearly with  $n$ , and in [Section 5](#) we extend the results from [Section 4](#) to the infinite time

horizon. Section 6 extends all results to the case where there is only  $o(n)$  many initially infected vertices, with some of the more technical proofs deferred to an appendix. In Section 7, we discuss the implications of our result, including a discussion of our notion of herd-immunity and its relation to the standard notion of the force of an infection, as well as an alternative heuristic derivation of our results using what is often called the pair-approximation. We kept Section 7 essentially self-contained, so after reading the summary of our results in Section 2 and possibly Section 3, the reader not interested in technical details might want to jump to the discussion section.

**Acknowledgements:** We are grateful to Tom Britton and Remco Van der Hofstad for their insightful discussions and for encouraging us to pursue this write-up, particularly in the early stages of developing the concept of using herd immunity to prove a law of large numbers for the epidemic’s final phase. Christian Borgs also extends thanks to the Simons Institute for the Theory of Computing for their support during the fall semester of 2022, and all three authors appreciate the Institute’s hospitality during that time. Lastly, Karissa Huang acknowledges support from the National Science Foundation under grant DGE 2146752.

## 2 Notation and Statements of Results

### 2.1 Notation

Throughout this paper, we use  $n$  to denote the number of individuals or number of vertices in the contact graph. We use  $\xrightarrow{P}$  for convergence in probability as  $n \rightarrow \infty$ , and we say that an event holds with high probability if it holds with probability tending to 1 as  $n \rightarrow \infty$ . For a positive integer  $K$ , we use  $[K]$  to denote the set  $[K] = \{1, \dots, K\}$ . We usually use bold letters for vectors, and if  $\mathbf{a} = (a_k)_{k \in K}$ , we use  $\text{diag}(\mathbf{a})$  for the diagonal matrix with entries  $(\text{diag}(\mathbf{a}))_{kk} = a_k$ . As usual, we will use  $\mathbb{N}$  to denote the set of natural numbers, and  $\mathbb{N}_0$  to denote the set of non-negative integers.

### 2.2 Model Definition

We consider an SIR epidemic spreading on a contact network of  $n$  individuals with a community structure modeled by a stochastic block model with  $K$  communities. To define the model, we first assign a label  $k(v) \in [K]$  to each individual in  $V = [n]$ , leading to a labeled set  $\mathbf{V} = (V, k(\cdot))$  of  $n$  vertices with labels in  $[K]$ . In a second step, we connect vertices at random, with the probability of an edge between two vertices  $u$  and  $v$  depending on the community labels  $k(u)$  and  $k(v)$ . This leads to the following formal definition.

**Definition 2.1** (Stochastic Block Model (SBM)). *Let  $K \in \mathbb{N}$ , let  $\mathbf{V} = (V, k(\cdot))$  be a set of vertices with labels in  $[K]$ , and let  $W$  be a symmetric, non-negative matrix such that each row contains at least one non-zero element. We say that a random graph  $G = (V, E)$  is drawn from a stochastic block model with affinity matrix  $W$ , and write  $G \sim \text{SBM}(\mathbf{V}, W)$ , if a pair of vertices  $\{u, v\}$  is joined by an edge with probability  $\frac{1}{n} W_{k(u)k(v)}$  independently for all  $\binom{n}{2}$  pairs  $\{u, v\}$ . We will use  $n_k$  to denote the number of vertices with label  $k$ , i.e.  $n_k = \sum_{v \in V} \mathbb{1}\{k(v) = k\}$ .*

**Remark 2.2.** *i) Note that the assumption that each row  $W_{k\cdot}$  of  $W$  contains at least one non-zero elements can be made without loss of generality, since otherwise vertices of color  $k$  would always be isolated, in which case we can just delete them from the graph  $G$ .*

ii) The SBM is often defined by choosing the labels of the vertices i.i.d. from some random distribution  $\boldsymbol{\rho} = (\rho_k)_{k \in [K]}$  over  $[K]$ . The results of this paper can easily be formulated for this version of the SBM as well; in fact, all our results hold for an arbitrary label distribution over  $[n]$ , as long as w.h.p., for all  $k \in [K]$  the ratio  $n_k/n$  stays bounded away from 0 uniformly in  $n$ .

In order to analyze SIR epidemic on the stochastic block model, it will be convenient to first study it on a slightly different model which allows for self-loops and multiple edges. We call it the Poisson Stochastic Block Model (PSBM).

**Definition 2.3** (Poisson Stochastic Block Model (PSBM)). *Let  $\mathbf{V}$  and  $W$  be as in Definition 2.1. We say that a random multi-graph  $G$  is drawn from a Poisson Stochastic Block Model,  $G \sim \text{PSBM}(\mathbf{V}, W)$ , if each pair of vertices  $\{u, v\}$  (for  $u, v \in [V]$ ) is joined by  $A_{uv}$  edges, and each vertex  $v$  has  $A_{vv}$  self-loops, where  $A_{uv} = A_{vu} \sim \text{Pois}\left(W_{k(v)k(v)}/n\right)$  independently for all  $v \leq u$ . We use  $D_k$*

$$D_k = \sum_{\ell \in [K]} W_{k\ell} \frac{n_\ell}{n} \quad (1)$$

to denote the average degree of a vertex of type  $k$ .

**Remark 2.4.** *The degree of a vertex  $v$  in  $G \sim \text{PSBM}(\mathbf{V}, W)$  is distributed according to  $d_v \sim \text{Pois}(D_{k(v)})$ , and its degree into the vertices with community label  $\ell$  is Poisson distributed with mean  $W_{k(v)\ell} \frac{n_\ell}{n}$ . Setting*

$$p_{k \rightarrow \ell} = \frac{1}{D_k} W_{k\ell} \frac{n_\ell}{n}, \quad (2)$$

we furthermore can couple these in such a way that conditioned on  $d_v$ , the degree into a community with label  $k$  can be obtained by choosing labels  $k_1, \dots, k_{d_v}$  i.i.d. with probability  $p_{k(v) \rightarrow k_i}$  and setting the degree into community  $\ell$  as the number of times  $\ell$  will appear among  $k_1, \dots, k_{d_v}$ .

Next, we formally define an SIR infection on a general multi-graph  $G$ , introducing at the same time some of the notation used later. Denote the number of edges between two vertices  $u$  and  $v$  by  $A_{uv}$ . The SIR model is then a continuous-time Markov process whose state at time  $t$  is given by the sets of susceptible, infected, and recovered vertices,  $V^S(t)$ ,  $V^I(t)$  and  $V^R(t)$ , respectively (with their union being equal to the set of vertices  $V$  in  $G$ ). Starting from an initial state  $V^S(0)$ ,  $V^I(0)$ , and  $V^R(0)$  at time  $t = 0$ , at time  $t > 0$  each  $u \in V^I(t)$  then infects a susceptible vertex  $v$  at rate  $\eta A_{uv}$ , and recovers at rate  $\gamma$ , independently for all vertices  $u \in V^I(t)$ . We adopt the standard assumption that our processes are càdlàg (or rcll), so that when vertex  $v$  gets infected at time  $t$ ,  $v \in V^S(t')$  for  $t' < t$  and  $v \in V^I(t') \cup V^R(t')$  for  $t' \geq t$  (and similarly for the recovery of a vertex  $u$  at time  $t$ ).

We will use the notation

$$V_k^S(t) = V^S(t) \cap V_k, \quad S_k(t) = |V_k^S(t)| \quad \text{and} \quad S(t) = |V^S(t)|.$$

We define  $V_k^I(t)$ ,  $I_k(t)$  and  $I(t)$ , and  $V_k^R(t)$ ,  $R_k(t)$  and  $R(t)$  analogously.

### 2.3 Statements of Results

In this paper, we separately discuss two types of initial conditions: one in which the number of initially infected vertices is proportional to  $n$ , and one where it is  $o(n)$ .

We start with the former, for which we will establish laws of large numbers for the time evolution as well as the final size of the infection. We will need the following assumptions on the initial state of the infection:

**Assumption 2.5.** *As  $n \rightarrow \infty$ ,*

$$i) \frac{1}{n}S_k(0) \xrightarrow{P} s_k(0) > 0 \text{ for all } k \in [K],$$

$$ii) \frac{1}{n}I_k(0) \xrightarrow{P} i_k(0) \geq 0 \text{ for all } k \in [K].$$

*We will assume that the initially infected and recovered vertices are chosen at random, possibly depending on their community labels, but not on the realization of  $\text{SBM}(\mathbf{V}, W)$  or  $\text{PSBM}(\mathbf{V}, W)$ .*

We are making the assumption on the random choice of initially infected and recovered vertices to avoid trivial examples, like an initial set of infected vertices that consists of a connected component in  $G$  (in which case all that happens is the independent recovery of all initially infected vertices). Note that equivalently, we could assume an arbitrary starting configuration  $V^S(0)$ ,  $V^I(0)$ , and  $V^R(0)$ , and draw  $G \sim \text{SBM}(\mathbf{V}, W)$  or  $G \sim \text{PSBM}(\mathbf{V}, W)$  independently of this starting configuration.

**Theorem 2.6** (Law of Large Numbers for Finite Time). *Consider the SIR epidemic on  $G \sim \text{PSBM}(\mathbf{V}, W)$  or  $G \sim \text{SBM}(\mathbf{V}, W)$ , with the initial state of the infection obeying the conditions from Assumption 2.5. Let  $s_k(t)$ ,  $i_k(t)$ , and  $x_k(t)$  be the unique solutions to*

$$\frac{ds_k(t)}{dt} = -\eta s_k(t) \sum_{\ell} x_{\ell}(t) W_{\ell k}, \quad (3)$$

$$\frac{dx_k(t)}{dt} = -(\eta + \gamma)x_k(t) + \eta s_k(t) \sum_{\ell} x_{\ell}(t) W_{\ell k}, \quad (4)$$

$$\frac{di_k(t)}{dt} = -\gamma i_k(t) + \eta s_k(t) \sum_{\ell} x_{\ell}(t) W_{\ell k}, \quad (5)$$

*with initial conditions  $s_k(0)$  and  $x_k(0) = i_k(0)$ , and let  $0 < t_0 < \infty$ , and let  $\varepsilon > 0$  be arbitrary. Then as  $n \rightarrow \infty$ ,*

$$\mathbb{P} \left( \sup_{t \in [0, t_0]} \left\| \left( \frac{1}{n} \mathbf{S}(t), \frac{1}{n} \mathbf{I}(t) \right) - (\mathbf{s}(t), \mathbf{i}(t)) \right\|_2 > \varepsilon \right) \rightarrow 0, \quad (6)$$

*where  $\mathbf{S}(t)$  is the vector  $(S_1(t), \dots, S_K(t))$ , and similarly for  $\mathbf{I}(t)$ ,  $\mathbf{s}(t)$  and  $\mathbf{i}(t)$ .*

**Remark 2.7.** *i) Assumption 2.5 does not have any condition on the set of initially recovered vertices, and our theorem does not state any LLN for  $\frac{1}{n}R_k(t)$ . However, if we add the assumption that for all  $k \in [K]$ ,  $n_k/n$  converges in probability to some  $\rho_k$ , we get the convergence of  $R_k(t)/n = n_k/n - S_k(t)/n - I_k(t)/n$  to  $r_k(t) = \rho_k - s_k(t) - i_k(t)$  for free.*

*ii) Note that the assumptions of the theorem include the case where  $\mathbf{i}(0) = 0$ , in which case the solution of the differential equations are constant,  $\mathbf{i}(t) \equiv 0$  and  $\mathbf{s}(t) \equiv \mathbf{s}(0)$  for all  $t$ . Note, however, that this does not imply that the actual infection dies out. Indeed, if  $\frac{\mathbf{I}(0)}{n} \rightarrow 0$  but  $I(0) \rightarrow \infty$  (and the basic reproduction number,  $\mathfrak{R}_0$ , is larger than one), the infection will eventually take off, see Theorem 2.13 below. It just takes an amount of time which diverges as  $n \rightarrow \infty$ ! By contrast,  $\frac{1}{n}(\mathbf{S}(t) - \mathbf{S}(0)) \rightarrow 0$  and  $\frac{1}{n}\mathbf{I}(t) \rightarrow 0$  uniformly for  $t$  on any bounded time interval, consistent with the statements of the theorem.*

The above remark raises the question of whether the bounds of Theorem 2.6 can be generalized to non-compact time intervals if we assume that  $\mathbf{i}(0) \neq 0$ . Theorem 2.8 below shows that this is indeed possible. It in particular gives a law of large numbers for the final state of the infection.

Before stating the theorem, we note that for fixed  $n$ , the infection will eventually die out, leaving just a set of vertices which never got infected (and hence are still susceptible) and a set of vertices that got infected at some point, but eventually recovered. The number of vertices in the latter set is usually referred to as the *final size of the infection*. Since the two sets are complements of each other, the final size of the infection (for the individuals in community  $k$ ) is therefore equal to  $I_k(0) + S_k(0) - S_k(\infty)$ . Under the convergence assumption for the initial state, Assumption 2.5, a law of large numbers for the final size of the infection is therefore equivalent to a law of large numbers for

$$\mathbf{S}(\infty) = \lim_{t \rightarrow \infty} \mathbf{S}(t).$$

The next theorem generalizes Theorem 2.6 to unbounded time intervals, and in particular relates the quantity  $\mathbf{S}(\infty)$  to the limit  $t \rightarrow \infty$  of  $\mathbf{s}(t)$ , where  $\mathbf{s}(t)$  and  $\mathbf{x}(t)$  are the solutions of the differential equations Eq. (3) and Eq. (4). Note that the existence of the limit  $\lim_{t \rightarrow \infty} \mathbf{s}(t)$  follows from the monotonicity of  $s_k(t)$  for all  $k$ , which in turn is a consequence of Eq. (3).

**Theorem 2.8.** *Under the additional assumptions that  $W$  is irreducible and  $\mathbf{i}(0) > 0$  as  $n \rightarrow \infty$ , the statement of Theorem 2.6 holds uniformly for all  $t \in \mathbb{R}_+$ ,*

$$\mathbb{P} \left( \sup_{t \in [0, \infty)} \left\| \left( \frac{1}{n} \mathbf{S}(t), \frac{1}{n} \mathbf{I}(t) \right) - (\mathbf{s}(t), \mathbf{i}(t)) \right\|_2 > \varepsilon \right) \rightarrow 0, \quad (7)$$

so in particular

$$\mathbb{P} \left( \left\| \frac{1}{n} \mathbf{S}(\infty) - \mathbf{s}(\infty) \right\|_2 > \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0.$$

where  $\mathbf{s}(\infty) = \lim_{t \rightarrow \infty} \mathbf{s}(t)$ .

The second statement of Theorem 2.8 gives a law of large numbers for the final size of the infection when the number of initially infected vertices is a constant fraction of the size of the graph. In Theorem 2.13, we consider the law of large numbers for the regime where the number of initially infected vertices is  $o(n)$ .

We start with the simple case of a single, initially infected vertex. We need some definitions.

**Definition 2.9** (Poisson Multi-type Branching Process). *Consider a  $K \times K$  matrix  $M$  with non-negative entries, and the multi-type, discrete time Poisson branching process where vertices of type  $k$  have  $\text{Pois}(M_{k\ell})$  children of type  $\ell$ . We will use  $\text{BP}_k(M)$  to denote the distribution of the labeled tree corresponding to this branching process starting from a root of type  $k$ .*

**Definition 2.10** (Infection Tree for arbitrary graphs). *Consider the SIR epidemic on an arbitrary graph  $G$  starting from a single vertex  $v$  in  $G$ . We use  $\mathcal{T}_{G,v}(t)$  to denote the infection tree on  $G$  at time  $t$ , defined as the tree on all vertices that were infected at or before time  $t$  (with labels indicating whether they are currently infected or recovered), and an edge between two vertices if it is an edge in  $G$  along which the infection has spread. If  $G$  carries community labels in some set  $[K]$ , we add these labels as a second label to the vertices in  $\mathcal{T}_{G,v}(t)$ .*

We consider two special cases: the case where  $G \sim \text{SBM}(\mathbf{V}, W)$  for an infection starting from single vertex  $v$  chosen uniformly at random from all vertices with label  $k$ , in which case we denote

the infection tree by  $\mathcal{T}_k^{\text{SBM}}(t)$ , and the case where the infection starts at the root of a Poisson branching process  $\text{BP}_k(M)$ , in which case we denote the infection tree by  $\mathcal{T}_k^{\text{BP}(M)}(t)$ , or simply by  $\mathcal{T}_k^{\text{BP}}(t)$  when  $M$  is clear from the context.

**Remark 2.11.** It is well known and easy to see that  $\mathcal{T}_k^{\text{BP}(M)}(\infty)$  can be represented as a standard, discrete time branching process as follows: starting from a vertex  $v$  of type  $k(v)$ , choose an exponentially distributed recovery time  $T_v \sim \text{exp}(\gamma)$  and then choose  $\text{Pois}(p_{T_v} M_{k(v)\ell})$  children of type  $\ell$ , where  $p_{T_v} = 1 - e^{-\eta T_v}$ . (To see this, just note that  $p_{T_v}$  is the probability that an  $\text{Exp}(\eta)$  distributed infection clock clicks before the recovery clock of the vertex  $v$  trying to infect its children runs out).

As a consequence, the expected number of children of  $v$  with label  $\ell$  in  $\mathcal{T}_k^{\text{BP}(M)}(\infty)$  is

$$C_{k(v)\ell} = \mathbb{E}_{T_v}[p_{T_v} M_{k(v)\ell}] = \frac{\eta}{\eta + \gamma} M_{k(v)\ell}.$$

Under the assumption that  $M$  is irreducible, standard branching process results then imply that the probability that the infection on  $\text{BP}_k(M)$  survives forever is non-zero if and only if the largest eigenvalue<sup>1</sup>,  $\lambda_{\max}(C)$ , is larger than 1. Defining

$$\pi_k = \mathbb{P}\left(\left|\mathcal{T}_k^{\text{BP}(M)}(\infty)\right| = \infty\right) \quad \text{and} \quad \mathfrak{R}_0 = \lambda_{\max}\left(\frac{\eta}{\eta + \gamma} M\right) \quad (8)$$

we therefore have that  $\pi_k > 0$  if and only if  $\mathfrak{R}_0 > 1$ .

**Lemma 2.12.** Consider the SIR epidemic on  $G \sim \text{PSBM}(\mathbf{V}, W)$  or  $G \sim \text{SBM}(\mathbf{V}, W)$ , with the initial state of the infection obeying the conditions from Assumption 2.5, and assume that  $V^I(0)$  consists of a single vertex  $v$  with label  $k$ . Let  $M$  be the matrix with matrix elements

$$M = W \text{diag}(\mathbf{s}(0)).$$

Then there exists a sequence  $N_n \rightarrow \infty$  as  $n \rightarrow \infty$  and a coupling of  $\mathcal{T}_{G,v}(t)$  and  $\mathcal{T}_k^{\text{BP}(M)}(t)$  such that

$$\mathbb{P}\left(\mathcal{T}_{G,v}(t) = \mathcal{T}_k^{\text{BP}(M)}(t) \text{ for all } t \text{ such that } |\mathcal{T}_k^{\text{SBM}}(t)| \leq N_n\right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

Lemma 2.12 allows us to relate the initial behavior of the epidemic to the behavior of the continuous time branching process  $\mathcal{T}_k^{\text{BP}(M)}(t)$ . In particular, if  $\mathfrak{R}_0 \leq 1$ , then the branching process  $\mathcal{T}_k^{\text{BP}(M)}(t)$  dies out eventually, and any outbreak on the original graph will be of constant size, and if  $\mathfrak{R}_0 > 1$  then  $\mathcal{T}_k^{\text{BP}(M)}(t)$  survives forever with probability  $\pi_k$ , and there will be an outbreak of size at least  $N_n$  on the original graph with asymptotic probability  $\pi_k$ .

In fact, it turns out that with high probability, the outbreak will be either of constant size, or of order  $\Theta(n)$ . The next theorem makes this precise, and also gives the relative size of an outbreak if it occurs. More precisely, it shows that conditioned on there being a large outbreak, the probability that a random vertex in  $V_k$  gets eventually infected is given by the survival probability of what is known as the backward branching process, see, e.g., [Barbour and Reinert, 2013; Barbour, 2014; Diekmann and Heesterbeek, 2000].

---

<sup>1</sup>Note that in general,  $M$  can have complex eigenvalues - but by the Perron-Frobenius Theorem, it has a single, non-degenerate eigenvalue which is positive and equal to the largest eigenvalue in absolute value., i.e., the spectral radius. We trust that the use of  $\lambda_{\max}(C)$  for this eigenvalue does not cause any confusion.



In our context, the backward branching process is the Poisson branching process  $\text{BP}_k(C)$  with transition matrix

$$C = \frac{\eta}{\eta + \gamma} W \text{diag}(\mathbf{s}(0)), \quad (9)$$

and its survival probability is

$$\theta_k = \mathbb{P}(|\text{BP}_k(C)| = \infty). \quad (10)$$

**Theorem 2.13.** *Assume that  $W$  is irreducible and consider the SIR epidemic on  $G \sim \text{SBM}(\mathbf{V}, W)$  or  $G \sim \text{PSBM}(\mathbf{V}, W)$  obeying Assumptions 2.5 with  $\mathbf{i}(0) = 0$  and assume further that*

$$1 - \prod_k (1 - \pi_k)^{I_k(0)} \xrightarrow{P} \pi$$

for some  $\pi \in [0, 1]$ . Then  $\frac{\mathbf{I}(0) + \mathbf{S}(0) - \mathbf{S}(\infty)}{n}$ , converges in probability to a random vector which is equal to  $\text{diag}(\boldsymbol{\theta})\mathbf{s}(0)$  with probability  $\pi$  and equal to 0 with probability  $1 - \pi$ .

Note that the theorem covers both the case where the vector  $(I_k(0))_{k \in [K]}$  has a finite limit, and the case where  $I(0) \rightarrow \infty$ . In the first case,  $\pi < 1$ , while in the second, we have that  $\pi = 1$  (provided  $\mathfrak{R}_0 > 1$ ), leading to an outbreak with probability tending to 1 as  $n \rightarrow \infty$ .

As a consequence of Lemma 2.12 and Theorem 2.13 we have the following corollary, which in particular implies that w.h.p., the epidemic starting from a single vertex either dies out after only infecting a finite number of vertices, or grows to size  $\Theta(n)$ .

**Corollary 2.14.** *Consider the setting from Theorem 2.13, with  $V^I(0)$  consisting of a single infected vertex of label  $k$ , and let  $N_n$  be an arbitrary sequence of integers with  $N_n \rightarrow \infty$  and  $N_n/n \rightarrow 0$  as  $n \rightarrow \infty$ . Then the following hold:*

- i) The probability that  $S(0) - S(\infty) > N_n$  converges to  $\pi_k$ .*
- ii) If  $\mathfrak{R}_0 \leq 1$ , the final size  $S(0) - S(\infty)$  converges in distribution to the size of  $\mathcal{F}_k^{\text{SBM}}(\infty)$ , showing in particular that the limiting distribution of  $S(0) - S(\infty)$  has exponentially decaying tails if  $\mathfrak{R}_0 < 1$ .*
- iii) Assume  $\mathfrak{R}_0 > 1$ . When conditioned on being at most  $N_n$ , the final size converges in distribution to  $|\mathcal{F}_k^{\text{SBM}}(\infty)|$  conditional on extinction, showing in particular that the conditional final size distribution has exponentially decaying tails.*
- iv) If  $\mathfrak{R}_0 > 1$  and we condition on the event that the final size is at least  $N_n$ , then  $\frac{S(0) - S(\infty)}{n}$ , converges in probability to  $\text{diag}(\boldsymbol{\theta})\mathbf{s}(0)$ , showing in particular that with high probability, the final size of the infection is of order  $\Theta(n)$  if we condition on it being at least  $N_n$ .*

**Remark 2.15.** *Since  $W$  is symmetric, even if it is not irreducible, it can be decomposed into isolated irreducible sub-components. As long as Assumption 2.5 is true in each sub-component, then Theorem 2.8 and Theorem 2.13 can be applied separately in each sub-component.*

## 2.4 Related Work

In this section we place our results and methods in the context of the existing literature. As discussed in the introduction, for both  $G(n, p)$  and the configuration model, the dynamics of the infection are primarily driven by a scalar-valued random variable determining the force of infection. For  $G(n, p)$  the force of the infection is the same for all susceptible vertices and proportional to the number of active half-edges, while for the configuration model, the force of the infection on vertices

of degree  $d$  depends on  $d$  and is proportional to  $d$  times the number of active half-edges. Due to this fact, for  $G(n, p)$  and the configuration model, a global time-rescaling can be applied to the stochastic epidemic such that the entire epidemic happens over the course of a finite time horizon. While our methods are inspired by [Ball and Britton, 2022; Janson et al., 2014; Neal, 2003], such a global time-rescaling cannot be applied in the case of the stochastic block model due to the fact that the force of infection is vector-valued rather than scalar-valued. Thus we develop new, fairly general, tools using the idea of herd immunity to derive the LLN for the final size in the stochastic block model.

Another method for deriving the trajectory of the epidemic curve is to analyze the forward and backward branching processes of the epidemic. In Barbour and Reinert [2013], the authors study the forward and backward branching processes in the case of graphs with bounded degree and apply their results to determine the epidemic curves for  $G(n, p)$  and the Volz configuration model Volz [2008]. Bhamidi et al. [2014] generalize the approach of Barbour and Reinert [2013] to the setting of sparse random graphs with degrees having finite second moments under a slightly less general epidemic setting. It is plausible that this approach could be generalized to the stochastic block model using multi-type Crump-Mode-Jagers continuous time branching processes<sup>2</sup> to prove a LLN for the epidemic curve. However, to our knowledge this has not yet been done nor has it been used even on a heuristic level to derive the correct differential equations describing the LLN, while these fall out naturally from our approach.

There also exist several methods of analyzing the SIR trajectories using local graph approximations, provided the initially infected set of vertices is a constant proportion of all vertices chosen uniformly at random. In Cocomello and Ramanan [2023], the authors study SIR and SEIR models with (possibly) time-varying rates on locally tree-like graphs, and derive a law of large numbers result. In Alimohammadi et al. [2023], the authors propose a local algorithm for approximation of the epidemic trajectory and leverage local graph limit theory to show that the SIR evolution on a sequence of graphs with a local limit converges to the same process in the graph limit. However, in these approaches, it is not easy to connect the differential equation phase of the epidemic to sub-linear initial conditions due to the fact that at the stopping time where a fixed fraction of the vertices has been infected, the state of the epidemic is not that of an epidemic started at uniformly chosen set of infected vertices.

By contrast, our work allows for general initial conditions of the epidemic (specifying the number of active half-edges emanating from each vertex separately), and in particular can be used to connect the initial phase starting from a sub-linear number of infected vertices to the phase where the differential equation approximation can be used to prove a LLN. But of course, the results from Cocomello and Ramanan [2023] and Alimohammadi et al. [2023], while not allowing to analyze the case with  $o(n)$  initially infected vertices, are much more general as far as the underlying graph model is concerned, and in fact do not even require an underlying stochastic graph model as long as the sequence has a local limit.

---

<sup>2</sup>We thank Júlia Komjáthy for discussing this approach after one of us presented the current work at a conference on random networks.

### 3 Preliminaries

#### 3.1 Poisson Stochastic Block Model Coupling

We will start with a simple lemma which will give us a coupling between the stochastic block model and the Poisson stochastic block model. Using this coupling, our proof of Theorem 2.6 for a graph drawn from  $\text{PSBM}(\mathbf{V}, W)$  will imply Theorem 2.6 for the standard stochastic block model  $\text{SBM}(\mathbf{V}, W)$ .

**Lemma 3.1.** *Let  $K$ ,  $W$ ,  $k(\cdot)$ , and  $n_k$  be as in Definition 2.1. Assume that  $n > \max_{k,\ell} W_{k,\ell}$ , and let  $W'_{k(u)k(v)} = \frac{W_{k(u)k(v)}}{1 - W_{k(u)k(v)}/n}$ , and consider a multi-graph  $G' \sim \text{PSBM}(\mathbf{V}, W')$ . If we condition on  $G'$  being simple, then the conditional distribution of  $G'$  is given by  $\text{SBM}(\mathbf{V}, W)$ . Furthermore, the probability that  $G'$  is simple is bounded away from 0, uniformly in  $n$ .*

As a consequence, all statements holding with high probability for  $G' \sim \text{PSBM}(\mathbf{V}, W')$  also hold with high probability for  $G \sim \text{SBM}(\mathbf{V}, W)$ , reducing the proof of any convergence in probability statements for  $G \sim \text{SBM}(\mathbf{n}, W)$  to those for  $G' \sim \text{PSBM}(\mathbf{n}, W')$ . Thus a LLN for the (easier to analyze) SIR epidemic on the Poisson stochastic block model will imply the same LLN on the standard stochastic block model.

*Proof.* If  $G'$  is a multi-graph with edge-multiplicities  $A_{uv}$ , then the probability of drawing  $G'$  from  $\text{PSBM}(\mathbf{V}, W')$  is equal to

$$\prod_{v \leq u} \frac{p_{k(u)k(v)}^{A_{uv}}}{A_{uv}!} e^{-p_{k(u)k(v)}}.$$

where  $p_{kk'} = n^{-1}W'_{kk'}$ . As a consequence,

$$\mathbb{P}(G' \text{ simple}) = \prod_{u < v} \left( e^{-p_{k(u)k(v)}} + e^{-p_{k(u)k(v)}} p_{k(u)k(v)} \right) \prod_{u=v} e^{-p_{k(u)k(v)}}.$$

Furthermore, for any simple graph  $G = (V, E)$

$$\mathbb{P}(G' = G \mid G' \text{ simple}) = \frac{\mathbb{P}(G' = G)}{\mathbb{P}(G' \text{ simple})} = \prod_{uv \in E} \frac{p_{k(u)k(v)}}{1 + p_{k(u)k(v)}} = \frac{W_{k(u)k(v)}}{n},$$

showing that the conditional distribution of  $G'$  is given by  $\text{SBM}(\mathbf{V}, W)$ . To lower bound  $\mathbb{P}(G' \text{ simple})$  we use that  $(1+p)e^{-p} \geq e^{-p^2/2}$ , a bound which follows, e.g., from the fact that the derivative of  $\log(1+p) - p + p^2/2$  is non-negative for all  $p \geq 0$ . Setting  $p = \max_{k,k'} p_{kk'}$ , we therefore get

$$\mathbb{P}(G' \text{ simple}) \geq e^{-\binom{n}{2} \frac{p^2}{2} - np} \geq e^{-\frac{(np)^2}{4} - np}.$$

Since  $pn$  is uniformly bounded away from 0 as  $n \rightarrow \infty$ , this completes the proof.  $\square$

**Remark 3.2.** *While the coupling of the full graph  $G \sim \text{SBM}(\mathbf{V}, W)$  to  $G' \sim \text{PSBM}(\mathbf{V}, W')$  requires moving from  $W$  to the modified matrix  $W'$  and conditioning on  $G'$  to be simple, it is not hard to show that if we are only interested in a small neighborhood of a given set of vertices, we can couple  $G \sim \text{SBM}(\mathbf{V}, W)$  to  $G' \sim \text{PSBM}(\mathbf{V}, W)$  such that within the small neighborhood,  $G$  and  $G'$  are identical with high probability, without having to condition  $G'$  or moving from  $W$  to  $W'$  (see Lemma 3.4 below, where we consider the set of vertices infected in the early stage of the infection).*

### 3.2 Coupling the SIR dynamic and the Poisson Stochastic Block Model

We study the SIR epidemic on a graph  $G$  drawn from the Poisson stochastic block model by dynamically revealing the edges of  $G$ . This turns out to be much easier for the Poisson stochastic block model than the original stochastic block model, due mainly to the independence inherent in the Poisson degrees as expressed in Remark 2.4.

We keep track of the exploration procedure by dynamically updating what one might call a decorated exploration forest on  $V$ , defined as a forest on  $V$ , augmented by a set of half-edges attached to the leaves of the trees. Vertices will be marked infected or recovered, and we refer to the number of half-edges attached to a leaf  $v$  as its active degree (denoted by  $d_v(t)$ ), and to the half-edges as active half-edges.

The exploration process starts with a set of roots, consisting of the initially infected set of vertices,  $V^I(0)$ , plus  $d_v(0) \sim \text{Pois}(D_{k(v)})$  half-edges attached to each  $v \in V^I(0)$ , chosen independently for all  $v$ . Given the exploration forest at some time  $t$ , for all active half-edges we independently explore a half-edge at rate  $\eta$  and independent from that, i.i.d. for all infected vertices, move an infected vertex to the recovered state at rate  $\gamma$ . When an active half-edge emanating from a vertex  $u$  is explored, we remove it, choose a label  $\ell$  according to  $p_{k(u) \rightarrow \ell}$ , and then choose a vertex  $v$  uniformly from all vertices with label  $k(v) = \ell$ . If this vertex is susceptible, we infect it, add the edge  $uv$  to the exploration tree, and add  $\text{Pois}(D_\ell)$  active half-edges to  $v$ . When a vertex  $v$  recovers, we remove all its active half-edges, i.e., we set  $d_v(t)$  to 0.

**Lemma 3.3.** *The above defined exploration process and the SIR-process on a random multi-graph  $G \sim \text{PSBM}(\mathbf{V}, W)$  can be coupled in such a way that for all times  $t$ , the sets of susceptible, infected, and recovered vertices are the same in both, with the forest traced out by the exploration process at time  $t$  being equal to the infection forest at time  $t$ . Furthermore, an edge  $uv$  in the infection forest (with  $uv$  oriented away from the roots) will appear in the exploration forest precisely at the time  $t' \leq t$  when  $v$  was infected by  $u$ .*

*Proof.* We start the proof with a cautionary remark: while the degree of any given vertex is given by  $d_v \sim \text{Pois}(D_{k(v)})$ , it is not true that this holds independently for any set of vertices containing more than one vertex, since the degrees of two vertices  $u$  and  $v$  share the randomness of  $A_{uv} = A_{vu}$ .

After this cautionary remark, we proceed with the actual proof. Consider a vertex  $u$  infected at time  $t$ , and let  $V_+$  consist of all vertices not yet discovered at this point. We can then couple  $d_u(t)$  to the Poisson random variables  $A_{uv} = A_{vu}$  defining the multi-graph  $G$  plus a second, independent copy  $\tilde{A}_{uv}$  as follows:

$$d_u(t) = \sum_{v \in V_+} A_{uv} + \sum_{v \in V \setminus V_+} \tilde{A}_{uv}.$$

We will refer to the half-edges corresponding to the second sum as compromised, and the vertices they point to as fake copies of the original vertex  $v$ . Furthermore, we will call a half-edge virulent at time  $t$  if its endpoint has not yet been explored and the hidden endpoint of the edge corresponding to it lies in  $S(t)$ . Finally, after the initial infection of  $u$ , whenever one of the half-edges corresponding to the first sum is explored, we will update the other half-edges pointing to the same end-point to be compromised as well.

After this setup, the proof of the lemma is straightforward: by construction, virulent half-edges will be not compromised, and for all vertices  $u$ , the total rate at which one of the Poisson clocks of virulent half-edges emanating from  $u$  clicks is just the rate at which  $u$  infects susceptible vertices. Finally, when a non-compromised half-edge is explored, its endpoint is a real, not fake

vertex, showing that the distribution of discovered, newly infected vertices is correct as well. This concludes the proof.  $\square$

**Lemma 3.4.** *Let  $N_n$  be a sequence such that  $N_n/n \rightarrow 0$  as  $n \rightarrow \infty$ , and fix a set of initially susceptible, infected, and recovered vertices. Then the SIR-process on the two graphs  $G \sim \text{SBM}(\mathbf{V}, W)$  and  $G' \sim \text{PSBM}(\mathbf{V}, W)$  can be coupled in such a way that with probability tending to 1 as  $n \rightarrow \infty$ , the sets of susceptible, infected, and recovered vertices are the same as long as  $|(V^I(t) \cup V^R(t)) \setminus V^R(0)| \leq N_n$ .*

*Proof.* On a high level, the lemma can be shown to hold by using the exploration process from the proof of Lemma 3.3 for both models, and using that we are only exploring  $O(nN_n)$  possible edges to determine the course of the infection up to the time where  $|V^I(t) \cup V^R(t) \setminus V^R(0)| = N_n$ . The fact that the total variation distance between a  $\text{Bern}(W_{k(u)k(v)}/n)$  and a  $\text{Pois}(W_{k(u)k(v)}/n)$  random variable is  $O(1/n^2)$  then implies the statement of the lemma.

Formally, we proceed as follows. Draw, independently for all oriented pairs  $(u, v) \in V \times V$ , the following random variables

1. a sample  $M_{uv} \sim \text{Bern}(W_{k(u)k(v)}/n)$
2. a sample  $M'_{uv} \sim \text{Pois}(W_{k(u)k(v)}/n)$
3. an infinite sequence of i.i.d. infection times  $T_{uv}^{(i)} \sim \text{Exp}(\eta)$ ,  $i = 1, 2, \dots$ ,

with  $M'_{uv}$  and  $M_{uv}$  coupled in such a way that  $M_{uv} = M'_{uv}$  with probability  $1 - O(1/n^2)$ , and  $T_{uv}^{(i)}$  drawn independently of the coupled random variables  $M'_{uv}$  and  $M_{uv}$ . Finally, draw independent samples of recovery times  $T_u \sim \text{Exp}(\gamma)$  for the vertices in  $V$ .

We then use the following coupling of the two processes: starting from the set of initially infected vertices, we attach two sets of (oriented) active edges to the vertices in  $V^I(0)$ : one red edge  $uv$  to each of the vertices  $v \in V$  with  $M_{uv} = 1$  and  $M'_{uv}$  orange edges between  $u$  and  $v$  for all vertices  $v$  with  $M'_{uv} > 0$ . We use the infection time  $T_{uv}^{(1)}$  on the red edges  $uv$  if  $M_{uv} = 1$ , and the times  $T_{uv}^{(1)}, \dots, T_{uv}^{(M'_{uv})}$  on the orange edges  $uv$  if  $M'_{uv} \geq 1$ . Using the random times  $T_u$  for recovery for both processes, we see that the infections process on  $G$  and  $G'$  is identical until the first time when there exists a newly infected vertex  $u$  and a vertex  $v \in V$  such that  $M_{uv} \neq M'_{uv}$ .

The rest is straightforward: order the vertices in  $V \setminus V^I(0)$  according to their infection times on  $G$ , and let  $A_i$  be the event that all edge choices up to the point where  $i$  got infected are such that the number of red and orange edges for each explored pair  $uv$  is the same (including the edges out of  $i$ ). We then get that

$$\mathbb{P}(A_i | A_{i-1}) = \prod_{v \in V} \mathbb{P}(M_{iv} = M'_{iv}) = (1 - O(1/n^2))^n = 1 - O(1/n)$$

with the initial probability of starting off with  $M_{uv} = M'_{uv}$  for all edges out of  $V^I(0)$  being

$$\mathbb{P}(A_0) = \prod_{u \in V^I(0)} \prod_{v \in V} \mathbb{P}(M_{uv} = M'_{uv}) = 1 - O(|V^I(0)|/n).$$

Since  $|V^I \cup V^R|$  is unaffected by recovery events, and grows by one precisely when a new vertex gets infected, we see that

$$\mathbb{P}(A_i) = 1 - O(\tilde{N}_i/n)$$

where  $\tilde{N}_i = |V^I(0)| + i$  is the size of  $|V^I(t) \cup V^R(t) \setminus V^R(0)|$  when  $i$  gets infected in  $G$ . This implies the statement of the lemma.  $\square$

When analyzing the SIR epidemic on a graph drawn from  $\text{PSBM}(\mathbf{V}, W)$ , we will henceforward work with the exploration process laid out at the beginning of this subsection, and will use

$$X_k(t) = \sum_{v:k(v)=k} d_v(t)$$

to denote the number of active half-edges attached to vertices of type  $k$  at time  $t$ . When proving Theorem 2.6, we will actually prove a LLN jointly for  $S_k(t)/n$ ,  $I_k(t)/n$  and  $X_k(t)/(D_k n)$ , with  $x_k(t)$  being the limit of  $X_k(t)/(nD_k)$ . Note that conditioned on  $I_k(0)$ , the initial number of active half-edges  $X_k(0)$  is equal in distribution to  $\text{Pois}(D_k I_k(0))$ , so in particular,  $X_k(0)/(D_k n)$  has the same expectation as  $I_k(0)/n$ , motivating the initial condition  $x_k(0) = i_k(0)$  in the theorem.

To motivate the differential equations Eq. (3), Eq. (4), and Eq. (5), we note that at any given time

- at rate  $\gamma X_k$ , we lose one active edge contributing to  $X_k$  to recovery,
- at rate  $\eta X_k$ , we explore one of the active half-edges contributing to  $X_k$ .

When we explore an active half-edge contributing to  $X_k$ ,  $X_k$  decreases by 1. To determine the rate of increase of  $X_k$ , we note that when a half-edge labeled  $\ell$  is explored, with probability  $p_{\ell k} = W_{\ell k} n_k / (n D_\ell)$ , we choose an endpoint with label  $k$ , and with probability  $S_k / n_k$ , this endpoint then hits a susceptible vertex, in which case  $S_k$  decreases by 1 and  $X_k$  increases by  $\text{Pois}(D_k)$ . Therefore, at an overall rate of  $\eta \frac{S_k}{n} \sum_{\ell} X_{\ell} \frac{W_{\ell k}}{D_{\ell}}$  we decrease  $S_k$  by 1 and increase  $X_k$  by  $\text{Pois}(D_k)$ . Defining

$$\hat{s}_k(t) = \frac{S_k(t)}{n}, \quad \hat{x}_k(t) = \frac{X_k(t)}{n D_k}, \quad \text{and} \quad \hat{i}_k(t) = \frac{I_k(t)}{n},$$

we see that the expected rate of change for  $\hat{s}_k$ ,  $\hat{x}_k$  and  $\hat{i}_k$ , conditioned on the current value of these quantities, is equal to

$$-\eta \hat{s}_k(t) \sum_{\ell} \hat{x}_{\ell}(t) W_{\ell k}, \quad -(\eta + \gamma) \hat{x}_k(t) + \eta \hat{s}_k(t) \sum_{\ell} \hat{x}_{\ell}(t) W_{\ell k} \quad \text{and} \quad -\gamma \hat{i}_k(t) + \eta \hat{s}_k(t) \sum_{\ell} \hat{x}_{\ell}(t) W_{\ell k},$$

respectively. Thus we expect that these quantities will obey a LLN given in terms of solutions to the differential equations Eq. (3), Eq. (4), and Eq. (5).

## 4 Law of Large Numbers on Bounded Time Intervals

In this section, we prove Theorem 2.6. To this end, we first state and prove a version of the LLN that allows for more general initial conditions than those implicit in Theorem 2.6. More specifically, we generalize the setting discussed in the last section, with  $X_k(0) \sim \text{Pois}(D_k I_k(0))$  to a setting where  $X_k(0)$  and  $I_k(0)$  can be specified separately. We will later use this more general theorem to discuss an epidemic starting from a single infected vertex at time 0.

#### 4.1 Formal statement for arbitrary initial conditions

We will write the differential equations Eq. (3), Eq. (4) and Eq. (5) in the compact form  $d\mathbf{y}_t/dt = \mathbf{b}(\mathbf{y}_t)$ , where  $\mathbf{y}_t \in \mathbb{R}_+^{3K}$  is the vector  $(s_1(t), \dots, s_K(t), i_1(t), \dots, i_K(t), x_1(t), \dots, x_K(t))$  and

$$b^j(\mathbf{y}) = \begin{cases} -\eta y^j \sum_{k=1}^K y^{k+2K} W_{kj} & j = 1, \dots, K \\ -\gamma y^j + \eta y^{j-K} \sum_{k=1}^K y^{k+2K} W_{k(j-K)} & j = K+1, \dots, 2K \\ -(\eta + \gamma) y^j + \eta y^{j-2K} \sum_{k=1}^K y^{k+2K} W_{k(j-2K)} & j = 2K+1, \dots, 3K. \end{cases} \quad (11)$$

We will consider solutions to these equations with initial conditions in the set

$$U_0 = \left\{ \mathbf{y} \in [0, \infty)^{3K} : \sum_{k=1}^K (y^k + y^{K+k}) \leq 1, \sum_{k=1}^K (y^k + y^{k+2K}) \leq 2 \right\},$$

which includes the initial conditions from Theorem 2.6, where we set  $x_k(0) = i_k(0)$  leaving just the condition  $\sum_{k=1}^K (y^k + y^{K+k}) \leq 1$ .

**Remark 4.1.** Note that any solution of the differential equation  $d\mathbf{y}_t/dt = \mathbf{b}(\mathbf{y}_t)$  will stay within  $U_0$  if  $\mathbf{y}_0 \in U_0$  by the fact that (i)  $d(y_t^k + y_t^{k+K})/dt \leq 0$  and  $d(y_t^k + y_t^{k+2K})/dt \leq 0$  for all  $k \in [K]$ , and (ii)  $dy_t^j/dt \geq 0$  if  $y_t^j = 0$ . Using furthermore that the vector field  $\mathbf{b}$  is Lipschitz in  $U_0$ , we see that for  $\mathbf{y}_0 \in U_0$ , the differential equation  $d\mathbf{y}_t/dt = \mathbf{b}(\mathbf{y}_t)$  has a unique solution  $\mathbf{y}_t \in U_0$  for all  $t \in [0, \infty)$ .

Our main theorem in this section gives a LLN for an arbitrary initial configuration of susceptible vertices, infected vertices, and active half-edges. We will formulate it for a deterministic initial configuration, and later use this theorem to discuss random initial conditions. We use  $\hat{\mathbf{y}}_t$  to denote the random vector

$$\hat{\mathbf{y}}_t = \left( \frac{S_1(t)}{n}, \dots, \frac{S_K(t)}{n}, \frac{I_1(t)}{n}, \dots, \frac{I_K(t)}{n}, \frac{X_1(t)}{nD_1}, \dots, \frac{X_K(t)}{nD_K} \right).$$

**Theorem 4.2** (Law of Large Numbers for Finite Time). *Let  $L$  be the Lipschitz constant of the vector field Eq. (11) on  $U = \{\mathbf{y} \in [0, \infty)^{3K} : \|\mathbf{y}\|_2 \leq 4\}$ , let  $t_0 < \infty$ , and choose  $\delta$  such that  $0 < \delta \leq \frac{1}{3}e^{-t_0L}$ . Choose  $\mathbf{y}_0 \in U_0$ , and let  $\mathbf{y}_t$  be the unique solution of*

$$\frac{d\mathbf{y}_t}{dt} = \mathbf{b}(\mathbf{y}_t)$$

*with initial condition  $\mathbf{y}_0$ . Consider the coupling of the SIR epidemic on  $\text{PSBM}(\mathbf{V}, W)$  as defined in Section 2, starting from an arbitrary initial configuration at time 0 obeying the conditions*

$$\max_{v \in [n]} d_v(0) \leq \log n \quad \text{and} \quad \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2 \leq \delta.$$

*If  $\min_k D_k \geq D_0 > 0$ , then*

$$\mathbb{P} \left( \sup_{t \in [0, t_0]} \|\hat{\mathbf{y}}_t - \mathbf{y}_t\|_2 > 3\delta e^{t_0L} \right) \leq \frac{\zeta}{n} \left( 1 + \frac{t_0(1 + \log n)^2}{\delta^2} \right), \quad (12)$$

*where  $\zeta$  is a constant depending on  $\eta, \gamma, \|W\|_\infty$  and  $D_0$ .*

## 4.2 Proof of Theorem 4.2

The proof of Theorem 4.2 is based on tools that are standard when proving laws of large numbers for continuous time Markov processes: Doob's  $L^2$ -inequality and Gronwall's lemma. We first introduce some notation, consider the continuous-time Markov process defined via the exploration process introduced in Section 3.2,

$$Y^n(t) := \left( (V_k^S(t))_{k \in [K]}, (V_k^I(t))_{k \in [K]}, (d_v^A(t))_{v \in [n]} \right), \quad (13)$$

with  $V_k^S(t), V_k^I(t), V_k^R(t)$  as in 2.2, and  $d_v^A(t)$  denoting the active degree of vertex  $v$  at time  $t$ . The state space of  $\{Y^n(t) : t \geq 0\}$  is given by

$$\mathcal{E}^n := \left\{ \left( (V_k^S)_{k \in [K]}, (V_k^I)_{k \in [K]}, (d_v^A)_{v \in V} \right) : \forall k, V_k^S, V_k^I \subseteq V_k \text{ and } V_k^S \cap V_k^I = \emptyset; \right. \\ \left. \forall v \in V, d_v^A \in \mathbb{N}_0 \text{ with } d_v^A = 0 \text{ if } v \notin \cup_{k \in [K]} V_k^I \right\}, \quad (14)$$

where  $V_k := \{v \in [n] : k(v) = k\}$ . We denote the elements of  $\mathcal{E}^n$  by  $\xi$ , and use  $q(\xi, \xi')$  to denote that rate at which  $\xi$  transitions to  $\xi'$ . We consider the following projection onto the coordinates for which we aim to establish the LLN,

$$\mathbf{y} = (y^1, \dots, y^{3K}) : \mathcal{E}^n \rightarrow \mathbb{R}^{3K} \\ \xi \rightarrow \frac{1}{n} \left( S_1, \dots, S_K, I_1, \dots, I_K, \frac{X_1}{D_1}, \dots, \frac{X_K}{D_K} \right)$$

where here we overload the notation using  $S_k = |V_k^S|$ ,  $I_k = |V_k^I|$  and  $X_k = \sum_{v \in V_k^I} d_v^A$ . In order to prove the LLN with the help of Doob's  $L^2$ -inequality, we need a bound on the variance of the jumps,

$$\alpha(\xi) = \sum_{\xi' \neq \xi} |\mathbf{y}(\xi') - \mathbf{y}(\xi)|^2 q(\xi, \xi').$$

**Lemma 4.3.** *Fix  $W$ ,  $\gamma$  and  $\eta$ , and assume that  $\min_k D_k \geq D_0$  for some  $D_0 > 0$ . Then there exists a constant  $\zeta_0 = \zeta_0(\|W\|_\infty, D_0, \gamma, \eta)$  such that*

$$\alpha(\xi) \leq \frac{\zeta_0}{n} \left( 1 + \max_v d_v^A \right)^2 \text{ for all } \zeta \in \mathcal{E}^n.$$

*Proof.* We write  $\alpha(\xi)$  as a sum of three contributions,

$$\alpha(\xi) = \alpha_R(\xi) + \alpha_A(\xi) + \alpha_I(\xi),$$

with the first term coming from transitions  $\xi \rightarrow \xi'$  corresponding to the recovery of a vertex, the second comes from the exploration of an active half-edge which does not lead to the infection of a new vertex, and the last one comes from infection events.

We start by analyzing the contribution to  $\alpha_R(\xi)$ . Vertices recover at rate  $\gamma$ . Each time a vertex recovers, the number of infected vertices decreases by 1 and the number of active infected edges



decreases by the active degree of the recovered vertex. Thus the contribution due to recoveries can be written as

$$\alpha_R(\xi) = \gamma \sum_k \sum_{v \in V_k^I} \left( n^{-2} + \frac{(d_v^A)^2}{n^2 D_k^2} \right) \leq \frac{\gamma}{n} \left( 1 + D_0^{-2} \max_v (d_v^A)^2 \right).$$

Next we analyze transitions where a vertex gets explored, but no vertex gets infected, in which case all that happens is that one of the active degrees decreases by 1. Since half-edges of type  $\ell$  get explored at rate  $\eta X_\ell$  and infect no vertex with probability

$$1 - \sum_{k=1}^K \frac{S_k W_{k\ell}}{n D_\ell} \leq 1,$$

we can bound

$$\alpha_A(\xi) \leq \eta \sum_{\ell=1}^K X_\ell \frac{1}{n^2 D_\ell^2} = \eta \sum_{\ell=1}^K \sum_{v \in V_\ell^I} d_v^A \frac{1}{n^2 D_\ell^2} \leq \frac{\eta}{n D_0^2} \max_v d_v^A.$$

Finally, consider the transitions where a vertex of type  $\ell$  infects a vertex of type  $k$ . In this case the number of infected vertices in community  $k$  increases by 1, the number of susceptible vertices in community  $k$  decreases by 1, and the number of free, infected, half-edges increases by  $\text{Pois}(D_k) - 1$  if the explored half-edge was of the same type. If it was of a different type,  $\ell$  then  $X_\ell$  decreases by 1 and  $X_k$  increases by  $\text{Pois}(D_k)$ . Thus the contributions to  $\alpha(\zeta)$  from infections can be written as

$$\alpha_I = \sum_{k=1}^K \alpha_{I,k}.$$

where

$$\alpha_{I,k} = \eta \frac{S_k}{n} \sum_{\ell} X_\ell \frac{W_{\ell k}}{D_\ell} \left( n^{-2} + n^{-2} + \frac{1}{n^2 D_k^2} \mathbb{E}[(Z_k - \delta_{k,\ell})^2 + (1 - \delta_{k,\ell})] \right),$$

with  $Z_k \sim \text{Pois}(D_k)$ . Thus

$$\begin{aligned} \alpha_{I,k} &\leq \eta n^{-2} \frac{S_k}{n} \sum_{\ell} X_\ell \frac{W_{\ell k}}{D_\ell} \left( 2 + \frac{1}{D_k^2} \mathbb{E}[Z_k^2 + 1] \right) \\ &= \eta n^{-2} \frac{S_k}{n} \sum_{\ell} X_\ell \frac{W_{\ell k}}{D_\ell} \left( 3 + D_k^{-1} + D_k^{-2} \right) \\ &\leq \frac{\eta}{n} \frac{S_k}{n} (3D_0^{-1} + D_0^{-2} + D_0^{-3}) \|W\|_\infty \max_v d_v^A \end{aligned}$$

As a consequence,

$$\alpha_I \leq \frac{\eta}{n} (3D_0^{-1} + D_0^{-2} + D_0^{-3}) \|W\|_\infty \max_v d_v^A$$

□

Next we prove the following lemma, which follows immediately from the construction in Section 3.

**Lemma 4.4.** Consider the coupling of the SIR epidemic on  $\text{PSBM}(\mathbf{V}, W)$  as defined in Section 2. Then the following statements hold with probability at least  $1 - \frac{1}{n}e^{7\|W\|_\infty}$ .

i) For an infection starting with a single, infected vertex  $v$  with  $d_v(0) \sim \text{Pois}(D_{k(v)})$ ,

$$\sup_{t \geq 0} \max_{v \in [n]} d_v(t) \leq \log n.$$

ii) For an arbitrary initial configuration

$$\sup_{t \geq 0} \max_{v \in [n]} d_v(t) \leq \max \left\{ \max_{v \in [n]} d_v(0), \log n \right\}.$$

*Proof.* i) Let  $Z_v$ ,  $v \in V$ , be independent  $\text{Pois}(D_{k(v)})$  random variables. Recall the construction of the coupled process and note that  $d_v(t)$  is decreasing in  $t$ , we see that  $\max_{v: k(v)=k} d_v(t)$  is stochastically dominated by  $\max_{v \in [n]} Z_v$ . We know that

$$\mathbb{P}(Z_v \geq \log n) = \mathbb{P}(e^{2Z_v} \geq n^2) \leq \frac{1}{n^2} \mathbb{E}[e^{2Z_v}] = \frac{1}{n^2} e^{(e^2-1)D_{k(v)}} \leq \frac{1}{n^2} e^{7\|W\|_\infty},$$

where we used that  $D_k \leq \|W\|_\infty$  for all  $k$ . The result follows by a union bound.

The proof of (ii) is a straightforward generalization of the proof of (i).  $\square$

*Proof of Theorem 4.2.* Given our preparations, this proof follows immediately from Theorem 4.1 in Darling and Norris [2008], which in turn is based on Gronwall's inequality and Doob's  $L^2$  inequality. Recall the definition of  $U$  from Theorem 4.2,

$$U = \left\{ \mathbf{y} \in [0, \infty)^{3K} : \|\mathbf{y}\|_2 \leq 4 \right\},$$

and set  $\varepsilon = 3\delta e^{t_0 L}$ . Note that  $U$  is a superset of  $U_0$  since  $\mathbf{y} \in U_0$  implies that  $\|\mathbf{y}\|_1 \leq 3$  and  $\|\mathbf{y}\|_2 \leq \|\mathbf{y}\|_1$ . Furthermore, we have that

$$\text{for all } t \in [0, \infty) \text{ and all } \xi \in \mathcal{E}^n, \quad \|\mathbf{y}(\xi) - \mathbf{y}_t\|_2 \leq \varepsilon \text{ implies that } \mathbf{y}(\xi) \in U. \quad (15)$$

To see this, note that  $\mathbf{y}_0 \in U_0$  implies that  $\mathbf{y}_t \in U_0$ , which in turn implies that  $\|\mathbf{y}_t\|_2 \leq \|\mathbf{y}_t\|_1 \leq 3$ . The claim then follows from the triangle inequality and the fact that  $\varepsilon \leq 1$  by our assumption that  $\delta \leq \frac{1}{3}e^{-t_0 L}$ . This puts us in the setting of Darling and Norris [2008], Section 4.

Theorem 4.1 of Darling and Norris [2008] then states that for all  $A \in \mathbb{R}_+$

$$\mathbb{P} \left( \sup_{t \leq t_0} \|\hat{\mathbf{y}}_t - \mathbf{y}_t\|_2 > \varepsilon \right) \leq \frac{4At_0}{\delta^2} + \mathbb{P} \left( \int_0^{T \wedge t_0} \alpha(Y^n(t)) dt > At_0 \right), \quad (16)$$

where  $T := \inf\{t \geq 0 : \hat{\mathbf{y}}_t \notin U\}$ . Set

$$A = \frac{\zeta_0}{n} (1 + \log n)^2,$$

where  $\zeta_0$  is the constant from Lemma 4.3. Combining the fact that  $d_v(0) \leq \log n$  by the assumption of the theorem with the first statement of Lemma 4.4 (ii) and Lemma 4.3, we then have that

$$\mathbb{P} \left( \int_0^{T \wedge t_0} \alpha(Y^n(t)) dt > At_0 \right) \leq \mathbb{P} \left( \sup_{t \in [0, t_0]} \max_{v \in [n]} d_v(t) \geq \log n \right) \leq \frac{1}{n} e^{7\|W\|_\infty} \quad (17)$$

This completes the proof.  $\square$

**Remark 4.5.** *Theorem 4.1 of Darling and Norris [2008] covers the more general situation where the drift vector of the continuous time Markov-chain is only approximately equal to  $\mathbf{b}(\hat{\mathbf{y}}_t)$ . To state this generalization, we assign a drift vector  $\beta(\xi)$  to each  $\xi \in \mathcal{E}^n$  via*

$$\beta(\xi) = \sum_{\xi' \neq \xi} (\mathbf{y}(\xi') - \mathbf{y}(\xi)) q(\xi, \xi').$$

*In the setting of Theorem 4.2, we had  $\beta(Y^n(t)) = \mathbf{b}(\hat{\mathbf{y}}_t)$  for all times  $t$ . If this is violated, Theorem 4.1 of Darling and Norris [2008] gives an additional error term of*

$$\mathbb{P} \left( \int_0^{T \wedge t_0} \|\mathbf{b}(\hat{\mathbf{y}}_t) - \beta(Y^n(t))\|_2 dt > \delta \right) \quad (18)$$

*on the right hand side of Eq. (16). We will need this generalization in the proof of Theorem 2.6.*

### 4.3 Proof of Theorem 2.6

Recall that Theorem 4.2 required that  $\|\hat{\mathbf{y}}_0 - \mathbf{y}_0\| \leq \delta$ , while Theorem 2.6 allows for arbitrary initial conditions as long as they obey Assumption 2.5. To prove Theorem 2.6, we therefore need to bound the probability

$$\mathbb{P}(\|\hat{\mathbf{y}}_0 - \mathbf{y}_0\|_2 > \delta),$$

where  $\mathbf{y}_0$  is the initial conditions as defined in the statement of Theorem 2.6. Using the fact that  $\mathbf{x}(0) = \mathbf{i}(0)$ , we have

$$\begin{aligned} \|\hat{\mathbf{y}}_0 - \mathbf{y}_0\|_2 &\leq \|\hat{\mathbf{s}}_0 - \mathbf{s}_0\|_2 + \|\hat{\mathbf{i}}_0 - \mathbf{i}_0\|_2 + \|\hat{\mathbf{x}}_0 - \mathbf{i}_0\|_2 \\ &\leq \|\hat{\mathbf{s}}_0 - \mathbf{s}_0\|_2 + 2\|\hat{\mathbf{i}}_0 - \mathbf{i}_0\|_2 + \|\hat{\mathbf{x}}_0 - \hat{\mathbf{i}}_0\|_2. \end{aligned}$$

By Assumption 2.5, the first two terms go to 0 in probability, furthermore  $\mathbb{E}[\hat{\mathbf{x}}_0] = \mathbb{E}[\hat{\mathbf{i}}_0]$ , thus by standard concentration bounds, the last term therefore also goes to 0 in probability. Showing that for any  $\delta > 0$ ,

$$\mathbb{P}(\|\hat{\mathbf{y}}_0 - \mathbf{y}_0\|_2 > \delta) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (19)$$

Next, as in the proof of Lemma 4.4,

$$\mathbb{P} \left( \max_{v \in [n]} d_v(0) > \log n \right) \leq \frac{1}{n} e^{7\|W\|_\infty},$$

which goes to 0 as  $n \rightarrow \infty$  as well. Finally, by Assumption 2.5 i) it follows that  $n_k/n$  is bounded away from 0 uniformly in  $n$ . Combining that fact with the assumption that  $\sum_\ell W_{k\ell} > 0$  for all  $k \in [K]$ , we know that  $\min_k D_k \geq D_0 > 0$  that does not depend on  $n$ . Therefore, under the conditions of Theorem 2.6, those of Theorem 4.2 hold with probability tending to 1 as  $n \rightarrow \infty$ , reducing Theorem 2.6 to Theorem 4.2 when  $G \sim \text{PSBM}(\mathbf{V}, W)$ .

To finish the proof of Theorem 2.6 when  $G \sim \text{SBM}(\mathbf{V}, W)$ , we use Lemma 3.1, which states that there exists a  $W'$  such that the distribution of  $G' \sim \text{PSBM}(\mathbf{V}, W')$  conditioned on being simple is that of  $G$ , and then apply Theorem 4.2 to SIR epidemics on  $G'$ . More precisely, we will use the

extension described in Remark 4.5 to take into account that the drift vector  $\boldsymbol{\beta}(\xi)$  of the continuous time Markov chain for the coupled SIR dynamics on  $G'$  is given in terms of  $W'$ ,

$$\boldsymbol{\beta}(\xi) = \left( \left( -\eta \frac{S_k}{n} \sum_{\ell} \frac{X_{\ell}}{n} \frac{W'_{\ell k}}{D_{\ell}} \right)_{k \in [K]}, \left( -\gamma \frac{I_k}{n} + \eta \frac{S_k}{n} \sum_{\ell} \frac{X_{\ell}}{n} \frac{W'_{\ell k}}{D_{\ell}} \right)_{k \in [K]}, \left( -(\eta + \gamma) \frac{X_k}{n D_k} + \eta \frac{S_k}{n} \sum_{\ell} \frac{X_{\ell}}{n} \frac{W'_{\ell k}}{D_{\ell}} \right)_{k \in [K]} \right), \quad (20)$$

while we want to establish the law of large numbers for the differential equations given in terms of the vector field Eq. (11) defined in terms of  $W$ .

We need an upper bound on the term in Eq. (18). Recall that  $W'_{\ell k} = W_{\ell k}/(1 - W_{\ell k}/n)$ , and assume that  $n \geq 2\|W\|_{\infty}$ . For any  $\xi \in \mathcal{E}^n$  we have

$$\|\mathbf{b}(\hat{\mathbf{y}}(\xi)) - \boldsymbol{\beta}(\xi)\|_2 \leq \|\mathbf{b}(\hat{\mathbf{y}}(\xi)) - \boldsymbol{\beta}(\xi)\|_1 \leq \sum_{k \in [K]} 3 \left\| \eta \hat{s}_k \sum_{\ell} \hat{x}_{\ell} (W'_{\ell k} - W_{\ell k}) \right\|_1 \leq \frac{\zeta}{n} \left\| \sum_{\ell} \hat{x}_{\ell} \right\|_1$$

for some constant  $\zeta$  depending on  $\|W\|_{\infty}$  and  $\eta$ . For  $n$  sufficiently large, we therefore have that

$$\begin{aligned} \mathbb{P} \left( \int_0^{T \wedge t_0} \|\mathbf{b}(\hat{\mathbf{y}}_t) - \boldsymbol{\beta}(Y^n(t))\|_2 dt > \delta \right) &\leq \mathbb{P} \left( \int_0^{T \wedge t_0} \frac{\zeta}{n} \|\hat{\mathbf{x}}(t)\|_1 dt > \delta \right) \\ &= \mathbb{P} \left( \frac{\zeta t_0}{n} \|\hat{\mathbf{x}}(t)\|_1 > \delta \right) \leq \mathbb{P} \left( D_0^{-1} \sup_{t \geq 0} \max_{v \in [n]} d_v(t) > \frac{\delta n}{\zeta t_0} \right) \\ &\leq n^{-1} e^{7\|W\|_{\infty}} \leq n^{-1} e^{14\|W\|_{\infty}} \end{aligned}$$

where the second to last inequality follows from a proof analogous to that of Lemma 4.4. This proves the desired law of large numbers for the epidemic on  $G' \sim \text{PSBM}(\mathbf{V}, W')$ . Conditioning on  $G'$  being simple, the statement of Theorem 2.6 for  $G \sim \text{SBM}(\mathbf{V}, W)$  follows with the help of Lemma 3.1.

## 5 Law of Large Numbers for the Final Size

In this section, we will prove Theorem 2.8, which gives a law of large numbers for the final state of the SIR epidemic on the stochastic block model.

### 5.1 Solution to Differential Equations

We start by establishing a few elementary facts about the solutions of the differential equations. Throughout this section,  $\mathbf{b}(\cdot)$  will be the vector field defined in Eq. (11).

**Lemma 5.1.** *For all finite  $t$ , the solution  $\mathbf{y}_t$  to the differential equation  $d\mathbf{y}_t/dt = \mathbf{b}(\mathbf{y}_t)$  is continuous in the initial conditions  $\mathbf{y}_0 \in U_0$ .*

*Proof.* The proof is standard and follows immediately from Gronwall's inequality and the fact that  $\mathbf{b}(\cdot)$  is Lipschitz continuous.  $\square$

**Lemma 5.2.** *Let  $\mathbf{y}_t = (s_1(t), \dots, s_K(t), i_1(t), \dots, i_K(t), x_1(t), \dots, x_K(t))$  be the solution to the differential equation  $d\mathbf{y}_t/dt = \mathbf{b}(\mathbf{y}_t)$  with initial condition  $\mathbf{y}_0 \in U_0$ . Then*

- (a) *As  $t \uparrow \infty$ ,  $s_k(t) \downarrow s_k(\infty) \geq 0$ .*

(b)  $\lim_{t \rightarrow \infty} x_k(t) = 0$

(c)  $\lim_{t \rightarrow \infty} i_k(t) = 0$

(d)  $s_k(\infty) > 0$  if and only if  $s_k(0) > 0$

(e) Assume that  $W$  is irreducible,  $s_k(0) > 0$  for all  $k$ , and that  $\mathbf{x}(0) \neq 0$ . Then

$$x_k(t) > 0, i_k(t) > 0 \text{ and } \frac{ds_k(t)}{dt} < 0$$

for all  $0 < t < \infty$  and all  $k \in [K]$ .

*Proof.* The proof follows from a straightforward analysis of the differential equations Eq. (3), Eq. (4), and Eq. (5) and is given in Appendix B.  $\square$

**Lemma 5.3.** For all initial conditions  $\mathbf{y}_0 \in U_0$ , the limit  $\mathbf{s}(\infty) = \lim_{t \rightarrow \infty} \mathbf{s}(t)$  obeys the implicit equation

$$s_k(\infty) = s_k(0) \exp \left\{ -\frac{\eta}{\eta + \gamma} \sum_{\ell} W_{\ell k} (s_{\ell}(0) - s_{\ell}(\infty) + x_{\ell}(0)) \right\}. \quad (21)$$

*Proof.* Defining  $\chi_{\ell}(t) = \int_0^t x_{\ell}(\tau) d\tau$ , we can integrate Eq. (3) to express  $s_k(\infty)$  as

$$s_k(\infty) = s_k(0) \exp \left\{ -\eta \sum_{\ell} \chi_{\ell}(\infty) W_{\ell k} \right\}. \quad (22)$$

Next we express the term  $(\eta + \gamma)x_k(t)$  in Eq. (4) as  $(\eta + \gamma)\frac{d\chi_k(t)}{dt}$  and then use Eq. (3) and Eq. (4) to conclude that the derivative of  $s_k(t) + x_k(t) + (\eta + \gamma)\chi_k(t)$  is 0, showing in particular that  $s_k(\infty) + (\eta + \gamma)\chi_k(\infty) = s_k(0) + x_k(0)$ . Inserted into the above expression for  $s_k(\infty)$  this proves the statement of the lemma.  $\square$

**Remark 5.4.** Assume that  $s_k(0) > 0$  for all  $k$ , and set  $q_k = s_k(\infty)/s_k(0)$ . We can then rewrite the implicit equation Eq. (21) as an implicit equation for  $q_k$ ,

$$q_k = e^{-a_k} \exp \left( -\sum_{\ell} C_{\ell k} (1 - q_{\ell}) \right) = e^{-a_k} G_k(\mathbf{q}) \quad (23)$$

where  $a_k = \frac{\eta}{\eta + \gamma} \sum_{\ell} W_{\ell k} x_{\ell}$ ,  $C = \frac{\eta}{\eta + \gamma} W \text{diag}(\mathbf{s}(0))$ , and  $G_k(\mathbf{q})$  is the generating function for the off-spring distribution of a Poisson multi-type branching process with mean matrix  $C$  starting from a root of type  $k$ . As we will see in the next section, for  $x(0) > 0$ , the implicit equation Eq. (23) has a unique solution which is the survival probability of a certain “backward” branching process expressing the probability that a random vertex of type  $k$  gets infected in the course of the infection.

## 5.2 Backward Branching Process

In the previous section we derived the implicit equation Eq. (23) by considering the infinite time limit of the solutions to the differential equations Eq. (3), Eq. (4), and Eq. (5). As we will see, they can also be understood as an implicit equation for the survival probability of a certain branching process. To motivate the definition of this branching process, we note that in order for an initially

susceptible vertex to be eventually infected, it either has to be infected directly by a vertex infected at time 0, or it has to be infected via a chain of infections going through initially susceptible vertices.

Consider an active edge contribution to  $X_\ell(0)$ . The probability that it directly infects a random vertex  $v$  of label  $k$  is  $\frac{\eta}{\eta+\gamma} \frac{W_{\ell k}}{nD_\ell}$ . Treating these infection events as approximately independent, we see that the probability that  $v$  is not directly infected is approximately equal to

$$\prod_{\ell} \left(1 - \frac{\eta}{\eta+\gamma} \frac{W_{\ell k}}{nD_\ell}\right)^{X_\ell} \approx \exp\left(-\frac{\eta}{\eta+\gamma} \sum_{\ell} W_{\ell k} x_\ell\right) = e^{-a_k}.$$

On the other hand, if a vertex  $w$  that is connected to  $v$  via an edge gets infected at some point during the course of the SIR epidemic, its probability to transmit the infection to  $v$  is equal to  $\frac{\eta}{\eta+\gamma}$ . Since  $v$  has  $\text{Pois}(W_{k\ell} \hat{s}_\ell(0)) \approx \text{Pois}(W_{\ell k} s_\ell(0))$  many neighbors of of label  $\ell$ , who themselves can again be either directly or indirectly infected, we are lead to consider the following ‘‘backward’’ branching process.

Starting from a susceptible vertex  $v$  of label  $k$ , we directly infect it with probability  $1 - e^{-a_k}$  (represented by a child of type  $K + 1$  in the backward branching process), while with probability  $e^{-a_k}$  we give it  $\text{Pois}\left(\frac{\eta}{\eta+\gamma} W_{\ell k} s_\ell(0)\right)$  children of color  $\ell$  which could transmit an infection to  $v$ . In later generations, we proceed iteratively with the same offspring distribution for parents of type  $k \in [K]$ , while parents which are infected stay infected (formally, with probability 1, each parent of type  $K + 1$  has exactly one child of type  $K + 1$ ). As argued above, the survival probability of the backward branching process starting from a root of type  $k$  should then be asymptotically equal to the probability that a randomly chosen vertex from the vertices with label  $k$  that are susceptible at time 0 gets infected at some time during the epidemic.

While we don’t derive this directly, we will prove it indirectly by showing that for  $\sum_{\ell} a_\ell > 0$  the implicit equation Eq. (23) has a unique solution, and that this solution is the survival probability of the backward branching process starting from a root of type  $k$ , see Appendix A for details.

### 5.3 $\mathfrak{R}_{\text{eff}}$ and herd immunity

To motivate the next definition and lemma, we note that the differential equation Eq. (4) can be written in the form

$$\frac{d\mathbf{x}(t)}{dt} = (\eta + \gamma)\mathbf{x}(t)(C(t) - \mathbf{1}), \quad (24)$$

where  $\mathbf{x}(t) = (x_1(t), \dots, x_K(t))$ ,  $\mathbf{1}$  is the  $K \times K$  identity matrix and  $C(t)$  is the matrix

$$C(t) = \frac{\eta}{\eta + \gamma} W \text{diag}(\mathbf{s}(t)).$$

Using  $\lambda_{\max}(C(t))$  to denote the largest eigenvalue of the matrix  $C(t)$ , we define

$$\mathfrak{R}_{\text{eff}}(t) := \lambda_{\max}(C(t)). \quad (25)$$

**Lemma 5.5.**  *$\mathfrak{R}_{\text{eff}}(t)$  is a continuous, non-increasing function of  $t$ . If we assume that  $W$  is irreducible,  $s_k(0) > 0$  for all  $k$ , and  $\mathbf{x}(0) \neq \mathbf{0}$ , then  $\mathfrak{R}_{\text{eff}}(t)$  is strictly monotone and  $\mathfrak{R}_{\text{eff}}(t) < 1$  for all large enough  $t$ .*

*Proof.* We start by proving monotonicity and continuity. Let  $\|\cdot\|_F$  be the frobenius norm. By Gelfand’s formula for the spectral radius,  $\mathfrak{R}_{\text{eff}}(t) = \lim_{m \rightarrow \infty} \|C^m(t)\|_F^{1/m}$ . Since  $s_k(t)$  is weakly

monotone decreasing in  $t$  for all  $k$ , it follows that  $(C(t))_{\ell k}$  is weakly monotone as well. Furthermore, the elements of  $C(t)$  are non-negative for all  $t$  so  $\|C^m(t)\|_F$  is weakly monotone, implying the desired weak monotonicity of  $\mathfrak{R}_{\text{eff}}(t)$ . To establish continuity, we use continuity of eigenvalues as a function of the matrix entries, [Bhatia et al. \[1990\]](#), combined with the continuity of  $s_k(\cdot)$ .

To prove strict monotonicity, we note that the eigenvalues of  $C(t)$  are equal to those of the symmetric matrix  $C'(t)$  with entries  $(C'(t))_{k\ell} = \frac{\eta}{\eta+\gamma} \sqrt{s_k(t)} W_{k\ell} \sqrt{s_\ell(t)}$ , giving the alternative representation

$$\mathfrak{R}_{\text{eff}}(t) = \max_{\mathbf{v}} \mathbf{v}^T C'(t) \mathbf{v}$$

where the maximum goes over all normalized vectors  $\mathbf{v}$ , which by Perron-Frobenius can be chosen to have all positive entries. Since  $s_k(t)$  is strictly monotone by [Lemma 5.2 \(e\)](#), the entries of  $C'(t)$  are strictly monotone, which implies that  $\mathfrak{R}_{\text{eff}}(t)$  is strictly monotone.

Finally, assume towards contradiction that  $\mathfrak{R}_{\text{eff}}(t) \geq 1$  for all  $t$ . By continuity, the same holds for the largest eigenvalue of  $C(\infty)$ ,  $\mathfrak{R}_{\text{eff}}(\infty)$ . Let  $\mathbf{v}$  be the eigenvector of  $C(\infty)$  corresponding to the eigenvalue  $\mathfrak{R}_{\text{eff}}(\infty) \geq 1$ . Again by Perron-Frobenius, we may assume that  $\mathbf{v}$  has all positive entries. Since the entries of  $C(t)$  are monotone decreasing in  $t$ , we have that  $C(t)\mathbf{v} \geq C(\infty)\mathbf{v} \geq \mathbf{v}$  component-wise. Inserted into [Eq. \(4\)](#), this implies  $\frac{d\mathbf{x}(t)}{dt} \cdot \mathbf{v} \geq 0$  for all  $0 \leq t < \infty$  which is a contradiction, as  $\mathbf{x}(0) \cdot \mathbf{v} > 0$  by the assumption that  $\mathbf{x}(0) \neq 0$  and  $\mathbf{x}(t) \cdot \mathbf{v} \rightarrow 0$  by [Lemma 5.2](#).  $\square$

**Remark 5.6.** *It is not hard to see that in the differential equation representation for the epidemic,  $\mathfrak{R}_{\text{eff}}(t) = 1$  is the herd immunity threshold of the epidemic. Indeed, if  $\mathfrak{R}_{\text{eff}}(t_0) < 1$  and  $\mathbf{v}$  is the corresponding right eigenvector with all positive entries, then*

$$\frac{d\mathbf{x}(t)}{dt} \cdot \mathbf{v} \leq (\eta + \gamma)(\mathfrak{R}_{\text{eff}}(t_0) - 1)\mathbf{x}(t) \cdot \mathbf{v} \quad \text{for all } t \geq t_0$$

*showing that  $\mathbf{x}(t) \cdot \mathbf{v}$  and hence  $\|\mathbf{x}(t)\|_1$  is decaying exponentially in  $t - t_0$  for any initial condition  $\mathbf{x}(t_0)$ . As a consequence, a small infection at time  $t_0$  will only have small overall impact on the final size of the infection. By contrast, if  $\mathfrak{R}_{\text{eff}}(t_0) > 1$ , then  $\mathbf{x}(t) \cdot \mathbf{v}$  will grow exponentially in  $t - t_0$  until  $\mathfrak{R}_{\text{eff}}(t) = 1$ . Thus no matter how small  $\mathbf{x}(t_0)$  is, the infection will always have a sizable effect on the final size, since the infection won't start to recede before it has infected enough susceptible vertices to drive  $\mathfrak{R}_{\text{eff}}(t)$  below one.*

In our subsequent proofs, it will be useful to have a version of [Lemma 5.5](#) which gives a lower bound on  $\mathfrak{R}_{\text{eff}}$  that holds uniformly over a suitable set of initial conditions.

**Lemma 5.7.** *Consider an arbitrary compact set  $\mathcal{C}$  of the initial conditions such that for all initial conditions in  $\mathcal{C}$ ,  $\mathbf{x}(0) \neq 0$  and  $s_k(0) > 0$  for all  $k$ , and assume that  $W$  is irreducible. Then there exists  $\delta > 0$  and  $t_0 < \infty$  such that  $\mathfrak{R}_{\text{eff}}(t_0) < 1 - \delta$  for all initial conditions in  $\mathcal{C}$ .*

*Proof.* Consider any starting conditions  $(\tilde{\mathbf{s}}(0), \tilde{\mathbf{x}}(0)) \in \mathcal{C}$ . By [Lemma 5.5](#), we can choose  $\tilde{t}_0 < \infty$  and  $\tilde{\delta} > 0$  such that  $\mathfrak{R}_{\text{eff}}(\tilde{t}_0) < 1 - 2\tilde{\delta}$ . By [Lemma 5.1](#) and the fact that eigenvalues are continuous in the elements of the matrix  $C(\tilde{t})$ , there exists an open ball  $B_{\tilde{r}}$  around  $(\tilde{\mathbf{s}}(0), \tilde{\mathbf{x}}(0))$  such that  $\mathfrak{R}_{\text{eff}}(\tilde{t}_0) < 1 - \tilde{\delta}$  for all initial conditions in  $B_{\tilde{r}} \cap \mathcal{C}$ . This uncountable collection of balls  $B_{\tilde{r}}$  indexed by starting conditions in  $\mathcal{C}$  gives an open cover of the set  $\mathcal{C}$  with the property that (i) each starting condition  $(\mathbf{s}(0), \mathbf{x}(0)) \in \mathcal{C}$  is in some ball  $B_{\tilde{r}}$  and (ii) for any starting condition in  $B_{\tilde{r}}$  we have that  $\mathfrak{R}_{\text{eff}}(\tilde{t}_0) < 1 - \tilde{\delta}$  for some  $\tilde{t}_0 < \infty$  and  $\tilde{\delta} > 0$ . Since  $\mathcal{C}$  is compact, this open cover has a finite subcover, giving the desired result.  $\square$

We will also use a similar lemma for the difference between  $\mathbf{y}(t)$  and its asymptotic value  $\mathbf{y}(\infty)$ .

**Lemma 5.8.** *Let  $W$  be irreducible and let  $\mathcal{C}$  be as in Lemma 5.7. Then there exists  $\delta_1 > 0$  such that  $\min_k s_k(\infty) \geq \delta_1$  for all initial conditions in  $\mathcal{C}$ . In addition, for all  $\varepsilon > 0$  there exists  $t_1 < \infty$  such that for all initial conditions in  $\mathcal{C}$ ,*

$$\|\mathbf{y}_t - \mathbf{y}_\infty\|_2 \leq \varepsilon \quad \text{for all } t \geq t_1.$$

*Proof.* Let  $t_0$  and  $\delta$  be as in Lemma 5.7, and let  $C'(t)$  be the symmetric matrix defined in the proof of Lemma 5.5. For  $t \geq t_0$ , we then have

$$\mathbf{x}(t) \leq \mathbf{x}(t_0)e^{(\eta+\gamma)(C(t_0)-1)(t-t_0)} = \mathbf{x}(t_0)A^{-1}e^{(\eta+\gamma)(C'(t_0)-1)(t-t_0)}A$$

where  $A$  is the diagonal matrix with entries  $\sqrt{s_k(t_0)}$ . Since  $1 \geq s_k(t_0) > 0$  for all initial conditions in  $\mathcal{C}$ , continuity and compactness imply the existence of some  $\delta' > 0$  such that  $1 \geq s_k(t_0) \geq \delta'$  for all initial conditions in  $\mathcal{C}$ . Combined with the fact that  $\mathbf{y}_{t_0} \in U_0$  implies  $\|\mathbf{x}(t_0)\|_2 \leq \|\mathbf{x}(t_0)\|_1 \leq 2$ , we conclude that for all  $t \geq t_0$ ,

$$\|\mathbf{x}(t)\|_2 \leq \frac{2}{\delta'} \|e^{(\eta+\gamma)(C'(t_0)-1)(t-t_0)}\|_{2 \rightarrow 2} = \frac{2}{\delta'} e^{(\eta+\gamma)(\mathfrak{R}_{\text{eff}}(t_0)-1)(t-t_0)} \leq \frac{2}{\delta'} e^{-(\eta+\gamma)\delta(t-t_0)}, \quad (26)$$

where  $\|\cdot\|_{2 \rightarrow 2}$  is the operator norm from  $\ell_2$  to  $\ell_2$ .

Observe that by the above bound,  $\int_{t_0}^\infty \|\mathbf{x}(t)\|_2 dt \leq \frac{2}{(\eta+\gamma)\delta\delta'}$ , which, by Eq. (21) and the lower bound on  $s_k(t_0)$  shown earlier, in turn implies that  $s_k(\infty)$  is bounded away from zero uniformly for all initial conditions in  $\mathcal{C}$ . Furthermore, again by Eq. (21) and the fact that  $1 - e^{-x} \leq x$ , we have that for all  $t \geq t_0$

$$\|\mathbf{s}(t) - \mathbf{s}(\infty)\|_2 \leq \eta \|W\|_\infty \int_t^\infty \|\mathbf{x}(t)\|_2 dt \leq \frac{2\eta \|W\|_\infty}{(\eta+\gamma)\delta\delta'} e^{-(\eta+\gamma)\delta(t-t_0)}. \quad (27)$$

To complete the proof, we choose  $\tilde{t}_1$  such that the right hand sides of Eq. (26) and Eq. (27) are smaller than  $\varepsilon/4$  for all  $t \geq \tilde{t}_1$  and then use the fact that  $\mathbf{i}(t) \rightarrow 0$  for all initial conditions in  $\mathcal{C}$  and another continuity & compactness argument to conclude that there exists a finite time  $t_1 \geq \tilde{t}_1$  such that  $\|\mathbf{i}(t_1)\|_2 \leq \varepsilon/4$  for all initial conditions in  $\mathcal{C}$ . Combined with the fact that  $i_k(t) + s_k(t)$  is increasing in  $t$ , we infer that for  $t \geq t_1$  we have

$$\|\mathbf{i}(t)\|_2 \leq \|\mathbf{i}(t_1)\|_2 + \|\mathbf{s}(t_1) - \mathbf{s}(t)\|_2 \leq \frac{\varepsilon}{4} + \|\mathbf{s}(t_1) - \mathbf{s}(\infty)\|_2 \leq \frac{\varepsilon}{2}$$

to complete the proof. □

Next we introduce the analog of  $\mathfrak{R}_{\text{eff}}$  for the actual SIR-infection on the Poisson stochastic block model. In particular, given the state of the infection at time  $t$  with  $S_k(t)$  susceptible vertices with label  $k$ , we define

$$\widehat{C}(t) = \frac{\eta}{\eta+\gamma} W \text{diag}(\hat{\mathbf{s}}(t)) \quad \text{where} \quad \hat{\mathbf{s}}(t) = \frac{1}{n} \mathbf{S}(t), \quad (28)$$

and set

$$\widehat{\mathfrak{R}}_{\text{eff}}(t) := \lambda_{\max}(\widehat{C}(t)). \quad (29)$$

The threshold  $\widehat{\mathfrak{R}}_{\text{eff}}(t) = 1$  can then be seen as the threshold for herd immunity for the (stochastic) SIR epidemic on the Poisson stochastic block model.



The following lemma shows that after passing this herd immunity threshold, a small initial outbreak (with a only a few active edges) will leave most of the vertices which are still susceptible untouched, while also not increasing the number of active half edges by more than a constant factor. It is a major ingredient in our proof of Theorem 2.8.

**Lemma 5.9.** *Consider an SIR-epidemic on  $\text{PSBM}(\mathbf{V}, W)$  and assume that at time  $t_0$  there are  $X_k(t_0)$  active edges and  $S_k(t_0)$  susceptible vertices of label  $k$ . If  $\widehat{\mathfrak{R}}_{\text{eff}}(t_0) < 1$ , then*

$$\mathbb{E}[\|\hat{\mathbf{s}}(t_0) - \hat{\mathbf{s}}(\infty)\|_2 \mid \mathcal{F}(t_0)] \leq \frac{1}{\min_k \sqrt{\hat{s}_k(t_0)}} \frac{\widehat{\mathfrak{R}}_{\text{eff}}(t_0)}{1 - \widehat{\mathfrak{R}}_{\text{eff}}(t_0)} \|\hat{\mathbf{x}}(t_0)\|_2$$

and

$$\mathbb{E}\left[\sup_{t \geq t_0} \|\hat{\mathbf{x}}(t)\|_2 \mid \mathcal{F}(t_0)\right] \leq \frac{1}{\min_k \sqrt{\hat{s}_k(t_0)}} \frac{1}{1 - \widehat{\mathfrak{R}}_{\text{eff}}(t_0)} \|\hat{\mathbf{x}}(t_0)\|_2,$$

where the expectation is conditioned on the state of the epidemic at time  $t_0$ .

*Proof.* Recall the coupling of the epidemic and the Poisson stochastic block model from Section 3.2. We will want to estimate the contribution from each active half-edge at time  $t_0$  to the the difference  $S_k(t_0) - S_k(\infty)$ , which is nothing but the total number of newly infected vertices emanating from this half-edge after time  $t_0$ . Assume that the half-edge in question has label  $\ell$ . With probability  $\frac{\eta}{\eta + \gamma}$  the infection clock on this half-edge will click at some time  $t$  before the recovery clock, at which point it will infect a vertex of label  $\ell'$  with probability  $\frac{W_{\ell\ell'}}{nD_\ell} S_{\ell'}(t)$ , for an overall probability of

$$\frac{\eta}{\eta + \gamma} \frac{W_{\ell\ell'}}{nD_\ell} S_{\ell'}(t) \leq \frac{1}{D_\ell} \widehat{C}_{\ell\ell'}(t_0)$$

of infecting a vertex of label  $\ell'$ . Each of these will give  $\text{Pois}(D'_\ell)$  new active half-edges of label  $\ell'$ , who in turn will each lead to a number of new infected vertices of label  $\ell''$  with probability at most  $\frac{1}{D'_\ell} \widehat{C}_{\ell'\ell''}(t_0)$ . Thus the expected number of infected vertices emanating from one of these is bounded by  $D_{\ell'} \frac{1}{D'_\ell} \widehat{C}_{\ell'\ell''}(t_0) = \widehat{C}_{\ell\ell''}(t_0)$ . Continuing by induction, we see that in expectation, the total number of infected vertices of label  $k$  in generation  $m$ ,  $m = 1, \dots$  is bounded above by

$$\sum_{\ell} \frac{X_\ell(t_0)}{D_\ell} (\widehat{C}(t_0)^m)_{\ell k}.$$

Summing up the contributions from all generations and observing that the total change in  $S_k$  from time  $t_0$  to the point where the infection dies out is just the total number of newly infected vertices after time  $t_0$ , we see that in expectation we have the bound

$$\mathbb{E}[\hat{s}_k(t_0) - \hat{s}_k(\infty) \mid \mathcal{F}(t_0)] \leq \sum_{m=1}^{\infty} (\hat{\mathbf{x}}(t_0) \widehat{C}(t_0)^m)_k = \left( \hat{\mathbf{x}}(t_0) \frac{\widehat{C}}{1 - \widehat{C}} \right)_k$$

Next we write  $\widehat{C}$  as  $\widehat{C} = \widehat{Q}^{-1} \widehat{M} \widehat{Q}$  where  $\widehat{Q}$  is the diagonal matrix with entries  $\widehat{Q}_{kk} = \sqrt{\hat{s}_k(t_0)}$  and  $\widehat{M} = \frac{\eta}{\eta + \gamma} \widehat{Q} W \widehat{Q}$  to write  $\frac{\widehat{C}}{1 - \widehat{C}}$  as  $\widehat{Q}^{-1} \frac{\widehat{M}}{1 - \widehat{M}} \widehat{Q}$ . This gives the bound

$$\begin{aligned}
\mathbb{E}[\|\hat{\mathbf{s}}(t_0) - \hat{\mathbf{s}}(\infty)\|_2 \mid \mathcal{F}(t_0)] &\leq \mathbb{E}[\|\hat{\mathbf{s}}(t_0) - \hat{\mathbf{s}}(\infty)\|_1 \mid \mathcal{F}(t_0)] \\
&\leq \sum_k \left( \hat{\mathbf{x}}(t_0) \frac{\widehat{C}}{1 - \widehat{C}} \right)_k = \sum_k \left( \hat{\mathbf{x}}(t_0) \widehat{Q}^{-1} \frac{\widehat{M}}{1 - \widehat{M}} \widehat{Q} \right)_k \\
&\leq \|\widehat{Q}^{-1}\|_\infty \hat{\mathbf{x}}(t_0) \frac{\widehat{M}}{1 - \widehat{M}} \mathbf{v}_M
\end{aligned}$$

where  $\mathbf{v}_M$  is the column vector with entries  $M_k$ . Using the fact that  $\widehat{M}$  is symmetric with the same eigenvalues as  $\widehat{C}(t_0)$ , we bound the operator norm of  $\frac{\widehat{M}}{1 - \widehat{M}}$  by  $\frac{\widehat{\mathfrak{R}}_{\text{eff}}(t_0)}{1 - \widehat{\mathfrak{R}}_{\text{eff}}(t_0)}$  to conclude that

$$\mathbb{E}[\|\hat{\mathbf{s}}(t_0) - \hat{\mathbf{s}}(\infty)\|_2 \mid \mathcal{F}(t_0)] \leq \|\widehat{Q}^{-1}\|_\infty \|\hat{\mathbf{x}}(t_0)\|_2 \frac{\widehat{\mathfrak{R}}_{\text{eff}}(t_0)}{1 - \widehat{\mathfrak{R}}_{\text{eff}}(t_0)} \|\mathbf{v}_M\|_2.$$

Bounding  $\|\mathbf{v}_M\|_2^2 = \sum_k \hat{s}_k(t_0) \leq 1$  and observing that  $\|\widehat{Q}^{-1}\|_\infty = \frac{1}{\min_k \sqrt{\hat{s}_k(t_0)}}$  this completes the proof of the first inequality.

To prove the second inequality, we note  $X_k(t)$  can be upper bounded by  $X_k(t_0)$  plus the total number of active half-edges created after time  $t_0$ . Proceed as in the proof of the bound on  $\hat{S}(\infty)$ , and denoting the number of new active half-edges of label  $k$  generated in generation  $m$  by  $Z_{m,k}$ , we get

$$\mathbb{E}[Z_{m,k} \mid (Z_{m-1,\ell})_{\ell \in [K]}] \leq D_k \sum_\ell \frac{\eta}{\eta + \gamma} Z_{m-1,\ell} \frac{W_{\ell k}}{nD_\ell} S_k(t_0) = D_k \sum_\ell \frac{Z_{m-1,\ell}}{D_\ell} \widehat{C}_{\ell k},$$

which by induction implies that

$$\mathbb{E} \left[ \frac{Z_{k,m}}{nD_k} \mid \mathcal{F}(t_0) \right] \leq \sum_\ell \frac{X_\ell(t_0)}{nD_\ell} (\widehat{C}^m)_{\ell k} = \sum_\ell \hat{x}_\ell(t_0) (\widehat{C}^m)_{\ell k}.$$

As a consequence,

$$\mathbb{E} \left[ \sup_{t \geq t_0} \hat{x}_k(t) \mid \mathcal{F}(t_0) \right] \leq \sum_{m \geq 0} \sum_\ell \hat{x}_\ell(t_0) (\widehat{C}^m)_{\ell k} = \left( \hat{\mathbf{x}}(t_0) \frac{1}{1 - \widehat{C}} \right)_k$$

Continuing as in the proof of the first bound, this implies the second bound of the lemma.  $\square$

**Remark 5.10.** Note that Lemma 5.9 implies a similar statement for  $I_k(t)$ . This follows immediately from the fact that both  $S_k(t)$  and  $S_k(t) + I_k(t)$  are non-increasing in  $t$ . As a consequence, we get that for all  $t \geq t_0$

$$I_k(t) \leq I_k(t_0) + S_k(t_0) - S_k(t) \leq I_k(t_0) + S_k(t_0) - S_k(\infty),$$

which together with the statements of Lemma 5.9 implies that past the herd immunity threshold, a small initial outbreak (with a small number of initially active edges and infected vertices) will stay small for all  $t \geq t_0$ .

## 5.4 Final Size of the Epidemic

In this section we prove Theorem 2.8, which in particular implies a law of large numbers for the final size of the infection. As in Section 4.1, we generalize the setting of Theorem 2.8, (with initial conditions which imply that  $X_k(0) \sim \text{Pois}(D_k I_k(0))$ ) to a setting where  $X_k(0)$  and  $I_k(0)$  can be specified separately, again using  $\hat{\mathbf{x}}(t)$  to denote the state of the epidemic of the SIR model coupled to  $G \sim \text{PSBM}(\mathbf{V}, W)$  as defined in Section 3.2. Theorem 2.8 will follow in the same way as Theorem 2.6 followed from Theorem 4.2.

Note that in contrast to Theorem 4.2, where all constants were independent of the initial conditions, the bounds in the next theorem are only uniform over the initial conditions in a compact subset  $\mathcal{C} \subset U_0$  obeying the condition

$$\mathbf{y}(0) \in \mathcal{C} \quad \Rightarrow \quad \mathbf{x}(0) \neq 0 \quad \text{and} \quad s_k(0) > 0 \quad \text{for all} \quad k \in [K]. \quad (30)$$

**Theorem 5.11.** *Assume that  $W$  is irreducible and  $\max_k D_k \geq D_0$  for some  $D_0 > 0$ . Let  $\mathcal{C} \subset U_0$  be a compact set obeying the conditions in Eq. (30), and let  $\varepsilon > 0$ . Then there exists constants  $\delta > 0$  and  $n_0 < \infty$  such that if  $\mathbf{y}_t$  is the solution of the differential equation  $d\mathbf{y}_t/dt = \mathbf{b}(\mathbf{y}_t)$  with initial conditions  $\mathbf{y}_0 \in \mathcal{C}$ , then*

$$\mathbb{P} \left( \sup_{t \geq 0} \|\hat{\mathbf{y}}_t - \mathbf{y}_t\|_2 \leq \varepsilon \right) \geq 1 - \varepsilon \quad (31)$$

provided  $n \geq n_0$  and the initial configuration at time 0 obeys the bound

$$\max_{v \in [n]} d_v(0) \leq \log n \quad \text{and} \quad \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2 \leq \delta. \quad (32)$$

*Proof.* The proof follows immediately from Theorem 4.2, the fact that  $\mathbf{y}_t \rightarrow \mathbf{y}_\infty$  as  $t \rightarrow \infty$ , and the fact that for  $t$  large enough, the infection has passed the herd immunity threshold and thus dies out quickly, implying that  $\hat{\mathbf{y}}_t - \hat{\mathbf{y}}_\infty$  is small when  $t$  is large. Lemmas 5.7, 5.8, 5.9 and Remark 5.10 provide the needed quantitative estimates.

Concretely, we first use Theorem 4.2 to see that for all  $t_0 < \infty$ , we can choose  $n_0 < \infty$  and  $\delta > 0$  such that if  $n \geq n_0$ ,  $\mathbf{y}_0 \in U_0$  and the conditions in Eq. (32) hold, then

$$\mathbb{P} \left( \sup_{0 \leq t \leq t_0} \|\hat{\mathbf{y}}_t - \mathbf{y}_t\|_2 \leq \varepsilon \right) \geq 1 - \varepsilon/4.$$

To prove Theorem 5.11, it is therefore enough to show that

$$\mathbb{P} \left( \sup_{t \geq t_0} \|\hat{\mathbf{y}}_t - \mathbf{y}_t\|_2 \leq \varepsilon \right) \geq 1 - 3\varepsilon/4, \quad (33)$$

uniformly for all  $\mathbf{y}_0 \in \mathcal{C}$ . To this end, we use the triangle inequality to conclude that for  $t \geq t_0$

$$\|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_2 \leq \|\mathbf{y}_t - \mathbf{y}_{t_0}\|_2 + \|\mathbf{y}_{t_0} - \hat{\mathbf{y}}_{t_0}\|_2 + \|\hat{\mathbf{y}}_{t_0} - \hat{\mathbf{y}}_t\|_2, \quad (34)$$

and then use the fact that  $\mathbf{y}_t$  converges as  $t \rightarrow \infty$  to bound the first term, Theorem 4.2 to bound the second, and Lemma 5.9 and Remark 5.10 to bound the last.

The details are tedious but straightforward. First, by Lemma 5.7, there exists  $\delta_0$  and  $t_1 < \infty$  such that for all initial conditions in  $\mathcal{C}$ ,  $\mathfrak{R}_{\text{eff}}(t_1) < 1 - \delta_0$ . By Lemma 5.8, there exists  $\delta_1 > 0$  such

that  $\min_k s_k(\infty) \geq \delta_1$ , and we can choose  $t_0 \geq t_1$  large enough such that for any starting condition in  $\mathcal{C}$ ,

$$\|\mathbf{x}(t_0)\|_2 \leq \frac{\varepsilon^2}{64} \frac{1 - \mathfrak{R}_{\text{eff}}(t_1)}{1 + \mathfrak{R}_{\text{eff}}(t_1)} \min_k \sqrt{s_k(t_0)}, \quad \|\mathbf{i}(t_0)\|_2 \leq \frac{\varepsilon}{8} \quad \text{and} \quad \|\mathbf{y}_\infty - \mathbf{y}_t\|_2 \leq \frac{\varepsilon}{8} \quad \text{for all } t \geq t_0.$$

In particular,  $\|\mathbf{y}_\infty - \mathbf{y}_t\|_2 \leq \varepsilon/8$  for all  $t \geq t_0$  implies that

$$\|\mathbf{y}_t - \mathbf{y}_{t_0}\|_2 \leq \|\mathbf{y}_t - \mathbf{y}_\infty\|_2 + \|\mathbf{y}_{t_0} - \mathbf{y}_\infty\|_2 \leq \frac{\varepsilon}{4} \quad \text{for all } t \geq t_0,$$

giving a uniform bound on the first term in Eq. (34).

Next we prove an upper bound on the right hand side of the bounds in Lemma 5.9. To this end, we use Theorem 4.2 to conclude that given any  $\varepsilon' > 0$  we can choose  $n_0$  large enough and  $\delta$  small enough, such that for all  $n \geq n_0$  and under the condition Eq. (32), we have that with probability at least  $1 - \varepsilon/4$

$$\|\hat{\mathbf{s}}(t_0) - \mathbf{s}(t_0)\|_2 \leq \varepsilon' \delta_1 \quad \text{and} \quad \|\hat{\mathbf{x}}(t_0)\|_2 \leq \|\mathbf{x}(t_0)\|_2 + \varepsilon'.$$

The first bound implies that  $\hat{s}_k(t_0) \leq (1 + \varepsilon')s_k(t_0)$  for all  $k$ , showing that  $\hat{C}(t_0) \leq (1 + \varepsilon')C(t_0)$  and thus  $\hat{\mathfrak{R}}_{\text{eff}}(t_0) \leq (1 + \varepsilon')\mathfrak{R}_{\text{eff}}(t_0) \leq (1 + \varepsilon')\mathfrak{R}_{\text{eff}}(t_1)$ , in addition to the lower bound  $\hat{s}_k(t_0) \geq (1 - \varepsilon')s_k(t_0)$ . By choosing  $\varepsilon'$  sufficiently small, we therefore have that with probability at least  $1 - \varepsilon/4$

$$\begin{aligned} \frac{1}{\min_k \sqrt{\hat{s}_k(t_0)}} \frac{1 + \hat{\mathfrak{R}}_{\text{eff}}(t_0)}{1 - \hat{\mathfrak{R}}_{\text{eff}}(t_0)} \|\hat{\mathbf{x}}(t_0)\|_2 &\leq \frac{3}{2} \frac{1}{\min_k \sqrt{s_k(t_0)}} \frac{1 + \mathfrak{R}_{\text{eff}}(t_1)}{1 - \mathfrak{R}_{\text{eff}}(t_1)} \|\hat{\mathbf{x}}(t_0)\|_2 \\ &\leq \frac{3}{2} \frac{1}{\min_k \sqrt{s_k(t_0)}} \frac{1 + \mathfrak{R}_{\text{eff}}(t_1)}{1 - \mathfrak{R}_{\text{eff}}(t_1)} \|\mathbf{x}(t_0)\|_2 + \frac{3}{2} \frac{1}{\sqrt{\delta_1}} \frac{2 + \delta_0}{\delta_0} \varepsilon' \leq \frac{3}{2} \frac{\varepsilon^2}{64} + \frac{1}{2} \frac{\varepsilon^2}{64} = \frac{\varepsilon^2}{32}. \end{aligned}$$

Combined with Lemma 5.9, this implies that with probability at least  $1 - \varepsilon/4$ .

$$\mathbb{E} \left[ \left( \|\hat{\mathbf{s}}(t_0) - \hat{\mathbf{s}}(\infty)\|_2 + \sup_{t \geq t_0} \|\hat{\mathbf{x}}(t)\|_2 \right) \middle| \mathcal{F}(t_0) \right] \leq \frac{\varepsilon^2}{32}. \quad (35)$$

Applying Markov's inequality to the conditional expectation, and then using that the above bound holds with probability at least  $1 - \varepsilon/4$ , we see that with probability at least  $1 - \varepsilon/2$ ,

$$\|\hat{\mathbf{s}}(t_0) - \hat{\mathbf{s}}(\infty)\|_2 + \sup_{t \geq t_0} \|\hat{\mathbf{x}}(t)\|_2 \leq \frac{\varepsilon}{8}.$$

Combined with the fact that  $\|\hat{\mathbf{s}}(t_0) - \hat{\mathbf{s}}(t)\|_2$  is monotone in  $t$  and the bound from Remark 5.10, we get that with probability at least  $1 - \varepsilon/2$ ,

$$\begin{aligned} \|\hat{\mathbf{y}}_{t_0} - \hat{\mathbf{y}}_t\|_2 &\leq \|\hat{\mathbf{s}}(t_0) - \hat{\mathbf{s}}(t)\|_2 + \|\hat{\mathbf{x}}(t_0)\|_2 + \|\hat{\mathbf{x}}(t)\|_2 + \|\hat{\mathbf{i}}(t_0)\|_2 + \|\hat{\mathbf{i}}(t)\|_2 \\ &\leq 2\|\hat{\mathbf{s}}(t_0) - \hat{\mathbf{s}}(\infty)\|_2 + \|\hat{\mathbf{x}}(t_0)\|_2 + \|\hat{\mathbf{x}}(t)\|_2 + 2\|\hat{\mathbf{i}}(t_0)\|_2 \\ &\leq \frac{\varepsilon}{4} + 2\|\hat{\mathbf{i}}(t_0)\|_2 + 2\|\hat{\mathbf{i}}(t_0) - \mathbf{i}(t_0)\|_2 \leq \frac{\varepsilon}{2} + 2\|\hat{\mathbf{i}}(t_0) - \mathbf{i}(t_0)\|_2. \end{aligned}$$

Inserting into Eq. (34) and using that  $\|\mathbf{y}_t - \mathbf{y}_{t_0}\|_2 \leq \frac{\varepsilon}{4}$  for all  $t \geq t_0$ , this gives that with probability at least  $1 - \varepsilon/2$

$$\sup_{t \geq t_0} \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_2 \leq \frac{3\varepsilon}{4} + \|\mathbf{y}_{t_0} - \hat{\mathbf{y}}_{t_0}\|_2 + 2\|\hat{\mathbf{i}}(t_0) - \mathbf{i}(t_0)\|_2$$

Using Theorem 4.2 a second time (and adjusting  $n_0$  and  $\delta$  if needed), the remaining terms on the right can be made smaller than  $\varepsilon/4$  with probability at least  $1 - \varepsilon/4$ , completing the proof.  $\square$

## 6 Law of Large Numbers for $o(n)$ Initially Infected Vertices

In this section we prove Theorem 2.13, the law of large numbers for the final size starting from one (or  $o(n)$ ) many initially infected vertices. In order to do so, we use a branching process approximation and martingale analysis for the initial phase of the epidemic to show that with high probability the epidemic either dies out or reaches size  $\Theta(n)$ . This puts us in the setting of Theorem 2.6, which we then use alongside results from Section 5 to prove the LLN result.

### 6.1 Branching Process Approximation for the Initial Phase

The infection tree of the initial phase of the epidemic on the Poisson SBM starting from one, randomly seeded, initially infected node is well-approximated by a continuous time branching process. Consider the following dynamics:

- The branching process starts with one infected node  $v_0$  chosen randomly from all nodes of label  $k$  and  $\text{Pois}(D_k)$  active half-edges attached to it.
- Infected nodes infect down active half-edges at rate  $\eta$ , independently for all active half-edges. When a half-edge emanating from a vertex  $u$  of label  $k'$  passes down the infection, a new infected vertex  $v$  with label  $\ell$  chosen with probability  $p_{k' \rightarrow \ell} = \frac{1}{D_k} W_{k'\ell} \frac{n_\ell}{n}$  is born, and the active half-edge gets replaced by an (inactive) full edge  $uv$ . Finally,  $v$  is given  $\text{Pois}(D_\ell)$  new active half-edges.
- Infected vertices recover at rate  $\gamma$ . Once a vertex recovers, all active half-edges emanating from this vertex get deleted, while the inactive (full) edges remain.

Observe that the tree generated from the dynamics above, represented by vertices labelled by group membership and state at time  $t$  (infected or recovered) and inactive edges between such vertices is equal to  $\mathcal{T}_k^{\text{BP}(\widehat{M})}(t)$ , with

$$\widehat{M} = W \text{diag}(\mathbf{n}/n).$$

Throughout this section, we will usually omit the reference to  $\widehat{M}$  and just use the notation  $\mathcal{T}_k^{\text{BP}}(t)$ . We will further slightly abuse notation by including the currently active half-edges as part of the tree  $\mathcal{T}_k^{\text{BP}}(t)$ , while in Section 2 these were not included.

Let  $\mathcal{T}_k^{\text{BP}}(t)$  be the tree generated by the dynamics described above at time  $t$ , represented as a tree with vertices labelled by group membership and the state at time  $t$  (infected or recovered), and edges labelled as active or inactive. Let  $\mathcal{T}_k^{\text{PSBM}}(t)$  be the SIR infection tree starting from a vertex in community  $k$  on  $G \sim \text{PSBM}(\mathbf{V}, W)$ , again including the active half-edges.

We can construct the process  $\mathcal{T}_k^{\text{PSBM}}(\cdot)$  from the process  $\mathcal{T}_k^{\text{BP}}(\cdot)$  as follows: generate a sample  $(\mathcal{T}_k^{\text{BP}}(t))_{t \geq 0}$  from the coupling described at the beginning of this section, but each time an infection passes along an edge  $uv$  in  $\mathcal{T}_k^{\text{PSBM}}(\cdot)$ , delete the edge with probability  $h_\ell(t^-)$ , where  $h_\ell(t^-) := 1 - S_\ell(t^-)/n_\ell$ ,  $t$  is the time the edge transitions from active to inactive in  $\mathcal{T}_k^{\text{BP}}(\cdot)$ ,  $t^-$  indicates a limit from the left, and  $\ell$  is the label of  $v$  in  $\mathcal{T}_k^{\text{BP}}(t)$ . This takes into account the fact that edges only appear in the SIR infection tree when the endpoint is connected to a susceptible individual, and couples the infection tree of PSBM,  $\mathcal{T}_k^{\text{PSBM}}(t)$ , to the branching process tree  $\mathcal{T}_k^{\text{BP}}(t)$  in such a way that  $\mathcal{T}_k^{\text{PSBM}}(t)$  is a sub tree of  $\mathcal{T}_k^{\text{BP}}(t)$  for all  $t \in [0, \infty)$ . Since  $S_\ell(t)$  decreases by 1 precisely when a new vertex with label  $\ell$  is added to  $\mathcal{T}_k^{\text{PSBM}}(t)$ , we have that  $h_\ell(t)$  is equal to  $\frac{1}{n_\ell}$  times

the number of vertices of label  $\ell$  in  $\mathcal{T}_k^{\text{PSBM}}(t)$ , showing in particular that  $\mathcal{T}_k^{\text{PSBM}}(t)$  is a Markov Process.

The above construction implies that  $|\mathcal{T}_k^{\text{PSBM}}(t)| \preceq |\mathcal{T}_k^{\text{BP}}(t)|$ , where  $\preceq$  means stochastic domination in the usual sense. It also implies that whenever  $|\mathcal{T}_k^{\text{PSBM}}(t)| = |\mathcal{T}_k^{\text{BP}}(t)|$ , then  $|\mathcal{T}_k^{\text{PSBM}}(t')| = |\mathcal{T}_k^{\text{BP}}(t')|$ , and in fact  $\mathcal{T}_k^{\text{PSBM}}(t') = \mathcal{T}_k^{\text{BP}}(t')$ , for all  $t' \leq t$ .

**Lemma 6.1.** *Let  $N_n$  be such that  $N_n/\sqrt{n} \rightarrow 0$  and assume that  $\liminf_{n \rightarrow \infty} \frac{n_k}{n} > 0$  for all  $k$ . Under the above coupling*

$$\mathbb{P} \left( \mathcal{T}_k^{\text{PSBM}}(t) = \mathcal{T}_k^{\text{BP}}(t) \text{ for all } t \text{ such that } |\mathcal{T}_k^{\text{BP}}(t)| \leq N_n \right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

*Proof.* Order the vertices in the branching process  $\mathcal{T}_k^{\text{BP}}(\cdot)$  by their time of arrival, with 1 being the label of the root, 2 the label of the next vertex to arrive, etc. Let  $t_i$  be the arrival time of vertex  $i$ ,  $k(i)$  its community label,  $v_i$  be the label of its parent in  $\mathcal{T}_k^{\text{BP}}(\cdot)$ , and  $f_\ell(i)$  be the number of vertices of group label  $\ell$  in  $\mathcal{T}_k^{\text{BP}}(\cdot)$  that have arrived **before** vertex  $i$  (not including  $i$ ).

We can imagine generating the SIR tree from  $\mathcal{T}_k^{\text{BP}}(\cdot)$  as follows: when vertex  $i$  gets born in  $\mathcal{T}_k^{\text{BP}}(\cdot)$ , leading to a new edge  $(v_i, i)$  in  $\mathcal{T}_k^{\text{BP}}(\cdot)$ , we assign a random variable  $Z_i \sim \text{Bern}(1 - h_{k(i)}(t_{i-1}))$  to the edge  $(v_i, i)$ . If  $Z_i = 1$ , then we keep the edge in the SIR tree, otherwise we remove the edge. More precisely, we obtain the tree  $\mathcal{T}_k^{\text{PSBM}}(t)$  by taking the connected component of the root obtained once we remove all edges with  $Z_i = 0$ . Our goal is to prove that

$$\mathbb{P}(Z_i = 1 \text{ for all } i = 1, \dots, N_n) \rightarrow 1 \text{ as } n \rightarrow \infty,$$

since that implies the statement of the lemma.

Note that in general, the random variables  $Z_i$  are not independent, even when conditioned on  $\mathcal{T}_k^{\text{BP}}(\cdot)$ , since  $h_{k(i)}(t_{i-1})$  depends on the tree  $\mathcal{T}_k^{\text{PSBM}}(t_{i-1})$ , which in turn depends on the previous variables  $Z_1, \dots, Z_{i-1}$ :

$$h_{k(i)}(t_{i-1}) = \frac{1}{n_{k(i)}} \left( f_{k(i)}(i) - \sum_{j=1}^{i-1} (1 - Z_j) \mathbf{1}_{k(j)=k(i)} \right).$$

However, if we condition on  $\mathcal{T}_k^{\text{BP}}(\cdot)$  and the event  $Z_1 = Z_2 = \dots = Z_{i-1} = 1$ , then  $h_{k(i)}(t_{i-1}) = f_{k(i)}(i)/n_{k(i)}$  and the conditional probability that  $Z_i = 1$  becomes

$$1 - f_{k(i)}(i)/n_{k(i)} \geq 1 - (i-1)/n_{k(i)}.$$

Rewriting the probability of the event that  $Z_1 = Z_2 = \dots = Z_{N_n} = 1$  as a product of conditional probabilities, we therefore get that

$$\begin{aligned} \mathbb{P} \left( \prod_{i=1}^{N_n} Z_i = 1 \mid \mathcal{T}_k^{\text{BP}}(\cdot) \right) &= \prod_{i=1}^{N_n} \left( 1 - \frac{f_{k(i)}(i)}{n_{k(i)}} \right) \geq \prod_{i=1}^{N_n} \left( 1 - \frac{i-1}{n_{k(i)}} \right) \geq 1 - \sum_{i=1}^{N_n} \frac{i-1}{n_{k(i)}} \\ &\geq 1 - \frac{N_n(N_n-1)}{\min_k n_k} = 1 - O(N_n^2/n). \end{aligned}$$

This implies

$$\mathbb{P} \left( \prod_{i=1}^{N_n} Z_i = 1 \right) = \mathbb{E}_{\mathcal{B}_k(\cdot)} \left[ \mathbb{P} \left( \prod_{i=1}^{N_n} Z_i = 1 \mid \mathcal{T}_k^{\text{BP}}(\cdot) \right) \right] \geq 1 - \frac{N_n(N_n-1)}{\min_k n_k} = 1 - O(N_n^2/n),$$

proving the lemma. □

Lemma 2.12 is an immediate corollary of Lemma 3.4 and Lemma 6.1.

*Proof of Lemma 2.12.* In view of Lemma 3.4 and Lemma 6.1, all we need to do is to couple the two process  $\mathcal{T}_k^{\text{BP}(M)}(t)$  and  $\mathcal{T}_k^{\text{BP}(\widehat{M})}(t)$  with  $M = W\text{diag}(\mathbf{s})$  and  $\widehat{M} = W\text{diag}(\mathbf{n}/n)$  in such a way that they are equal with high probability as long as, say, the first one has not too many vertices.

Formally, we proceed as follows: we label vertices in the two processes by their order of arrival in the infection tree, with both of them starting at vertex  $v = 1$  at time  $t = 0$ . We then iteratively couple the two processes as follows: if up to the time  $t_i$  when vertex  $i$  got infected in  $\mathcal{T}_k^{\text{BP}(M)}(\cdot)$  the two processes are identical, then we use the same random variable  $T_i \sim \text{Exp}(\gamma)$  for the recovery time of  $i$ . Furthermore, we couple the set of children of  $i$  in  $\text{BP}(M)$  and  $\text{BP}(\widehat{M})$  by optimally coupling the random variables  $\text{Pois}(M_{k(i)\ell})$  and  $\text{Pois}(\widehat{M}_{k(i)\ell})$  such that they are equal for all  $\ell$  with probability  $1 - O(\varepsilon_n)$ , where  $\varepsilon_n = \|\mathbf{s} - \frac{\mathbf{n}}{n}\|_1$ . If they are equal, we use the same random variables to determine the infection times along the edges emanating from  $i$  for both processes; if they are not, we run the two process independently from there on.

Let  $A_i$  be the event that up to the time when vertex  $i$  in  $\mathcal{T}_k^{\text{BP}(M)}(\cdot)$  gets infected, both process have identical histories. Then  $\mathbb{P}(A_{i+1} | A_i) = 1 - O(\varepsilon_n)$ , showing that

$$\mathbb{P}(A_i) = 1 - O(i\varepsilon_n).$$

To complete the proof, we use Assumption 2.5, to find a deterministic sequence  $N_n \rightarrow \infty$  with  $n \rightarrow \infty$  such that  $\varepsilon_n N_n \rightarrow 0$  in probability. This implies that

$$\mathbb{P}(A_{N_n}) \rightarrow 1 \quad \text{in probability as } n \rightarrow \infty.$$

(Note  $\mathbb{P}(A_{N_n})$  is random due to the possibly random starting conditions allowed by Assumption 2.5). Together with Lemma 3.4 and Lemma 6.1, this proves Lemma 2.12.  $\square$

**Remark 6.2.** *Observe that the coupling described at the beginning of this section and the results in Lemma 6.1 and Lemma 2.12 are robust to the starting conditions of the epidemic. Suppose that are  $\zeta = O(1)$  many initially infected vertices. Then couple the epidemic to a forest of branching processes in the following way: start a branching process  $\mathcal{T}_{k(v)}^{\text{BP}}(t)$  from each initially infected vertex  $v \in \zeta$ . Then proceed as described in the beginning of this section, removing edges with probability  $h_\ell(t^-)$  each time an infection event occurs into community  $\ell$ . It is clear that Lemma 6.1 immediately generalizes and the epidemic and forest are exactly equal with high probability until the size of the epidemic is of order  $N_n$ .*

Next, we will show that either the infection dies out, or there is a "large outbreak" of the SIR epidemic on the PSBM – in which case a constant fraction of the vertices are infected, and the differential equation approximation proved in Section 4 applies. To this end, it will be convenient to switch from the continuous time Markov chain defining the SIR epidemic to a discrete time Markov chain.

## 6.2 Martingale Bounds and Discrete Time Markov Chain

In this section we will prove the following theorem using a submartingale concentration bound for a suitably defined discrete time Markov chain. Throughout this subsection we will assume that  $W$  is irreducible, and that the initial conditions satisfy Assumption 2.5.

**Theorem 6.3.** Consider the SIR epidemic on  $G \sim \text{SBM}(\mathbf{V}, W)$  or  $G \sim \text{PSBM}(\mathbf{V}, W)$ , and let

$$\tau_\varepsilon = \min\{t : S(t) \leq S(0) - \lceil \varepsilon S(0) \rceil\}.$$

If  $\mathfrak{R}_0 > 1$ , then there exist constants  $\varepsilon_0 > 0$  and  $0 < \zeta_1 < \zeta_2 < \infty$  such that for  $0 < \varepsilon \leq \varepsilon_0$ , the following statements hold

(a) if  $I(0) \xrightarrow{P} \infty$  and  $I(0)/n \xrightarrow{P} 0$  as  $n \rightarrow \infty$ , then with probability tending to 1 as  $n \rightarrow \infty$

$$\tau_\varepsilon < \infty \quad \text{and} \quad \zeta_1 \varepsilon S(0) \leq \|\mathbf{X}(\tau_\varepsilon)\|_1 \leq \zeta_2 \varepsilon S(0).$$

(b) if  $I(0)$  is bounded in probability and  $N_n \xrightarrow{P} \infty$ , then with probability tending to 1 as  $n \rightarrow \infty$

$$\exists t \text{ s.th. } S(0) - S(t) \geq N_n \quad \Rightarrow \quad \tau_\varepsilon < \infty \quad \text{and} \quad \zeta_1 \varepsilon S(0) \leq \|\mathbf{X}(\tau_\varepsilon)\|_1 \leq \zeta_2 \varepsilon S(0).$$

To define the discrete time Markov chain used in the proof of the theorem, we observe that the continuous time SIR epidemic on the Poisson stochastic block model is a Markov Chain where elements,  $\xi$ , of the state space  $\Omega$  can be represented as follows:

$$\xi = \{S_k, I_k^d, R_k; k \in [K], d \in \mathbb{N}\}.$$

$S_k, R_k$  are the number of susceptible and recovered nodes in community  $k$ , respectively, and  $I_k^d$  is the number of infected nodes in community  $k$  with active degree  $d$ . First, we make a few observations. Notice that  $I_k = \sum_{d=0}^{\infty} I_k^d$  is the number of infected nodes in community  $k$ . Also,  $X_k = \sum_{d=0}^{\infty} d I_k^d$  is the number of active edges in community  $k$ . Since infected nodes with active degree zero have no bearing on the course of the epidemic, for our purposes, it will be useful to work with the following state space that omits such vertices:

$$\tilde{\xi} = \{S_k, A_k^d; k \in [K], d \in \mathbb{N}\}$$

where  $A_k^0 = R_k + I_k^0$  and  $A_k^d = I_k^d$  for all  $d \geq 1$ . We now construct a discrete time Markov Chain based on the jumps of the continuous time Markov Chain. Let  $A_k = \sum_{d \geq 1} A_k^d$ ,  $A = \sum_k A_k$ ,  $X_k = \sum_d d A_k^d$  and  $X = \sum_k X_k$ .

For this section, let  $T \in \mathbb{N}$  denote our discrete time steps, with  $T$  incrementing by 1 whenever the state of the continuous time chain changes. The rate at which this happens is then equal to  $\Gamma(t) = \gamma A(t) + \eta X(t)$ , where  $t$  denotes the time in the continuous-time epidemic. We use  $(S_k(T), A_k^d(T))_{k,d}$  to represent the state of the chain at time  $T$ . We start the chain at time  $T = 0$  with one initially infected vertex of label  $k_0$  and active degree  $d_0 \sim \text{Pois}(D_{k_0})$ , i.e.,  $S_k(0) = n_k - \delta_{kk_0}$  and  $A_k^d(0) = \delta_{kk_0} \delta_{dd_0}$  where  $\delta_{nm}$  is the Kronecker-Delta.

Whenever  $T$  increments in the discrete time chain, it corresponds to one of the following events:

- with probability  $\frac{\gamma A_k(T)}{\Gamma(T)}$ , a node chosen uniformly at random among all infected nodes with active degree  $d \geq 1$  in community  $k$  recovers. In particular, a node with active degree  $d \geq 1$  in community  $k$  is chosen with probability  $A_k^d(T)/A_k(T)$ . In this case,

$$A_k^d(T+1) = A_k^d(T) - 1, A_k^{d-1}(T+1) = A_k^{d-1}(T) + 1 \text{ and } X_k(T+1) = X_k(T) - d. \quad (36)$$



- with probability  $\frac{\eta X_k(T)}{\Gamma(T)}$ , choose an active edge uniformly at random among all active edges in community  $k$  to attempt infection across an edge. In particular, an edge is chosen with probability  $dA_k^d(T)/X_k(T)$  for  $d \geq 1$ . In this case,

$$A_k^d(T+1) = A_k^d(T) - 1, A_k^{d-1}(T+1) = A_k^{d-1}(T) + 1, \text{ and } X_k(T+1) = X_k(T) - 1. \quad (37)$$

- In this case, we also select the endpoint that the active edge infects and this infection event is successful with certain probability. With probability  $\frac{W_{k\ell}S_\ell}{nD_\ell}$ , we choose a susceptible vertex in community  $\ell$  to infect, and we give it  $d' \sim \text{Pois}(D_\ell)$  new active edges. In this case, the infection is successful and in addition to the changes to the system as in Eq. (37), we also have that

$$S_\ell(T+1) = S_\ell(T) - 1, A_\ell^{d'}(T+1) = A_\ell^{d'}(T) + 1, \text{ and } X_\ell(T+1) = X_\ell(T) + d'. \quad (38)$$

- If the infection is not successful, no further action is taken.

Let  $\mathcal{E}^n$  be a realization of this Markov chain, let  $\mathcal{F}_T$  be the natural filtration of this chain, and let  $\mathbf{X}(T)$  be the row vector  $\mathbf{X}(T) = (X_1(T), \dots, X_K(T))$ . Then the expectation of  $\mathbf{X}(T+1)$  conditioned on  $\mathcal{F}_T$  is easily calculated, giving

$$\mathbb{E}[\mathbf{X}(T+1) \mid \mathcal{F}_T] = \mathbf{X}(T) + \frac{\eta + \gamma}{\Gamma(T)} \left( \mathbf{X}(T)C_1(T) - \mathbf{X}(T) \right), \quad (39)$$

where  $C_1(T)$  is the matrix with matrix elements

$$(C_1)_{k\ell}(T) = \frac{\eta}{\eta + \gamma} D_k W_{k\ell} \frac{S_\ell(T)}{nD_\ell}.$$

Note that zero is an absorbing state for the chain  $\mathcal{E}^n$ ; if  $X(T)$  reaches 0 (and therefore  $A(T)$  also reaches 0), then the chain dies and remains at 0 forever. In order to aid in the analysis, we construct another Markov Chain on the same state space,  $\Omega$ , that restarts whenever the chain hits zero.

**Definition 6.4** (Restarted Discrete Time Chain). *Let  $\mathcal{E}_n$  be a draw from the Markov Chain described above. The restarted chain,  $\mathcal{E}_n^{(1)}$ , is a Markov Chain on  $\Omega$  that is identical to  $\mathcal{E}_n$  up to the random time  $\tau$  when  $\mathbf{X}(\tau) = 0$ . If  $S_{k_0}(\tau) = 0$ , the chain stops, otherwise restart  $\mathcal{E}_n^{(1)}$  with the current drift probabilities according to another copy of  $\mathcal{E}_n$  drawn independently, with  $S_k(\tau+1) = S_k(\tau) - \delta_{kk_0}$  and  $A_k^d(\tau+1) = \delta_{kk_0} \delta_{dd_0}$  where  $d_0$  is a fresh sample from  $\text{Pois}(D_{k_0})$ . If the restarted chain hits zero again, we again restart it in the same way, until  $S_{k_0} = 0$ , at which point the chain stops. We denote the active half-edges in the restarted chain by  $\mathbf{X}^{(1)}(T)$  and the natural filtration of the restarted chain by  $(\mathcal{F}_T^{(1)})_{T \geq 0}$ .*

Note Eq. (39) holds for the restarted chain as well, except when  $\mathbf{X}^{(1)}(T) = 0$ , in which case

$$\mathbb{E}[X_k^{(1)}(T+1) \mid \mathcal{F}_T^{(1)}] = \delta_{kk_0} D_{k_0}. \quad (40)$$

In a later analysis, we will construct a submartingale from  $\mathbf{X}^{(1)}(T)$  and prove a concentration result for that submartingale. Observe that at each discrete time step, the change in  $|\mathbf{X}^{(1)}(T)|$  is determined by one Poisson random variable, which has unbounded increments. In order to prove concentration for the submartingale, it will be convenient for us to introduce the following, slightly

modified, version of the discrete time chain, which has bounded increments. Here, we truncate the value of the Poisson increments by making the initially drawn degree of a newly infected vertex in community  $k$  the minimum of a  $\text{Pois}(D_k)$  random variable and some parameter  $\zeta_L$ , which we will later choose to be conveniently large enough such that the truncated process is equal to the original process with high probability.

**Definition 6.5** (Restarted, Truncated Discrete Time Chain). *Let  $\mathcal{E}_n^{(2)}$  be a Markov Chain on  $\Omega$ , constructed in the same way as  $\mathcal{E}_n^{(1)}$ , but when a susceptible vertex  $v$  is infected as described in the beginning of this section, its active degree is given by  $d' \sim \text{Pois}(D_{k(v)}) \wedge \zeta_L$  for some parameter  $\zeta_L$ . Denote the active edges of this process by  $\mathbf{X}^{(2)}(T)$  and the natural filtration by  $(\mathcal{F}_T^{(2)})_{T \geq 0}$ .*

Notice that Eq. (39) holds for  $\mathbf{X}^{(2)}(T)$  as well with  $C_1$  replaced by  $C_2$  where

$$(C_2)_{k\ell}(T) = \frac{\eta}{\gamma + \eta} \mathbb{E}[\text{Pois}(D_k) \wedge \zeta_L] W_{k\ell} \frac{S_\ell(T)}{nD_\ell}$$

except for when  $\mathbf{X}^{(2)}(T) = 0$ , in which case we have

$$\mathbb{E}[X_k^{(2)}(T+1) \mid \mathcal{F}_T^{(2)}] = \delta_{kk_0} \mathbb{E}[\text{Pois}(D_{k_0}) \wedge \zeta_L].$$

Recall that in the context of Section 6.1 and Lemma 2.12, we defined

$$\mathfrak{R}_0 = \lambda_{\max} \left( \frac{\eta}{\eta + \gamma} W \text{diag}(\mathbf{s}(0)) \right).$$

Since similar matrices have the same eigenvalues, we can also express  $\mathfrak{R}_0$  as

$$\mathfrak{R}_0 = \lambda_{\max}(C_0) \quad \text{where} \quad (C_0)_{k\ell} = \frac{\eta}{\gamma + \eta} D_k W_{k\ell} \frac{s_\ell(0)}{D_\ell}.$$

By Perron-Frobenius, we can find a normalized right eigenvector  $\mathbf{v}$  with strictly positive entries such that  $C_0 \mathbf{v} = \mathfrak{R}_0 \mathbf{v}$ . Furthermore, by Assumption 2.5, we have that  $C_1(0) \mathbf{v} \xrightarrow{p} \mathfrak{R}_0 \mathbf{v}$  as  $n \rightarrow \infty$ . Thus with probability tending to 1,

$$(C_1(0) \mathbf{v})_k \geq \frac{5\mathfrak{R}_0 + 1}{6} v_k \quad \text{for all } k \in [K].$$

**Lemma 6.6.** *Assume that  $\mathfrak{R}_0 > 1$  and  $C_1(0) \mathbf{v} \geq \frac{5\mathfrak{R}_0 + 1}{6} \mathbf{v}$  component wise. Choose  $\zeta_L$  sufficiently large so that*

$$\min_k \frac{\mathbb{E}[\text{Pois}(D_k) \wedge \zeta_L]}{D_k} \geq \frac{5}{6} + \frac{1}{6\mathfrak{R}_0}, \quad (41)$$

let

$$\varepsilon \leq \varepsilon_0 = \frac{\min_k S_k(0)}{6S(0)} \left( 1 - \frac{1}{\mathfrak{R}_0} \right),$$

let  $\tau_2$  be the stopping time

$$\tau_2 := \inf\{T \geq 0 : S^{(2)}(T) \leq \lfloor (1 - \varepsilon)S(0) \rfloor\},$$

and let  $H_2(T) = \mathbf{X}^{(2)}(T) \mathbf{v} - \zeta_0 T$  where

$$\zeta_0 = \frac{1}{2} \min\{\mathfrak{R}_0 - 1, D_{\min}\} v_{\min},$$

with  $D_{\min} = \min_k D_k$  and  $v_{\min} = \min_k v_k$ . Then  $H_2(T \wedge \tau_2)$  is a submartingale.

*Proof.* Let  $\mathcal{E}_n^{(2)}$  be a realization of the Markov Chain from Definition 6.5, and let

$$M_2(T) = H_2(T \wedge \tau_2).$$

We will prove that  $\mathbb{E}[M_2(T+1) \mid \mathcal{F}_T^{(2)}] \geq M_2(T)$ . Notice that  $\{\tau_2 \leq T\}$  is measurable with respect to  $\mathcal{F}_T^{(2)}$ . We consider three cases:

(i)  $T \geq \tau_2$

In this case,  $M_2(T+1) = M_2(T)$  and the inequality holds as an equality.

(ii)  $T < \tau_2$  and  $\mathbf{X}^{(2)}(T) = \mathbf{0}$

In this case, we have that  $M_2(T) = H_2(T)$  and  $M_2(T+1) = H_2(T+1)$ . If  $\mathbf{X}^{(2)} = \mathbf{0}$ , then we have

$$\mathbb{E}[H_2(T+1) \mid \mathcal{F}_T^{(2)}] = \mathbb{E}[\mathbf{X}^{(2)}(T+1)\mathbf{v} \mid \mathcal{F}_T^{(2)}] - (T+1)\zeta_0 \geq \zeta_0 T = H_2(T)$$

where we used that  $\mathbb{E}[\mathbf{X}^{(2)}(T+1)\mathbf{v} \mid \mathcal{F}_T^{(2)}] = \mathbb{E}[\text{Pois}(D_{k_0}) \wedge \zeta_L]v_{k_0} \geq \zeta_0$  by our assumption on  $\zeta_L$ .

(iii)  $T < \tau_2$  and  $\mathbf{X}^{(2)}(T) \neq \mathbf{0}$

In this case, we also have that  $M_2(T) = H_2(T)$  and  $M_2(T+1) = H_2(T+1)$ . If  $\mathbf{X}^{(2)} \neq \mathbf{0}$ , then

$$\mathbb{E}[H_2(T+1) \mid \mathcal{F}_T^{(2)}] = H_2(T) - \zeta_0 + \frac{\eta + \gamma}{\Gamma(T)} \left( \mathbf{X}^{(2)}(T)C_2(T) - \mathbf{X}^{(2)}(T) \right) \mathbf{v}.$$

Let  $\tilde{\varepsilon} = 1 - 1/\mathfrak{R}_0$ . Using the condition  $T < \tau_2$ , we know that  $S^{(2)}(T) \geq S(0)(1 - \varepsilon)$  which implies that for all  $\ell \in [K]$

$$S_\ell(0) - S_\ell(T) \leq S(0) - S(T) \leq \varepsilon_0 S(0) \leq \frac{\tilde{\varepsilon}}{6} S_\ell(0).$$

As a consequence,  $C_1(T)\mathbf{v} \geq \left(1 - \frac{\tilde{\varepsilon}}{6}\right) C_1(0)\mathbf{v} \geq \left(1 - \frac{\tilde{\varepsilon}}{6}\right) \frac{5\mathfrak{R}_0 + 1}{6} \mathbf{v}$  and

$$\begin{aligned} C_2(T)\mathbf{v} &\geq \left(\frac{5}{6} + \frac{1}{6\mathfrak{R}_0}\right) C_1(T)\mathbf{v} \geq \left(\frac{5}{6} + \frac{1}{6\mathfrak{R}_0}\right) \left(1 - \frac{\tilde{\varepsilon}}{6}\right) \frac{5\mathfrak{R}_0 + 1}{6} \mathbf{v} \\ &= \left(1 - \frac{\tilde{\varepsilon}}{6}\right)^3 \mathfrak{R}_0 \mathbf{v} \geq \left(1 - \frac{\tilde{\varepsilon}}{2}\right) \mathfrak{R}_0 \mathbf{v} = \frac{\mathfrak{R}_0 + 1}{2} \mathbf{v} \end{aligned}$$

component wise. This in turn implies that

$$\begin{aligned} \mathbb{E}[H_2(T+1) \mid \mathcal{F}_T^{(2)}] &\geq H_2(T) - \zeta_0 + \frac{\eta + \gamma}{\Gamma(T)} \frac{\mathfrak{R}_0 - 1}{2} \mathbf{X}^{(2)}(T)\mathbf{v} \\ &\geq H_2(T) - \zeta_0 + \frac{\mathfrak{R}_0 - 1}{2} v_{\min} \geq H_2(T) \end{aligned}$$

where in the last line we used that  $\Gamma(T) \leq (\eta + \gamma) \|\mathbf{X}^{(2)}(T)\|_1$ .

□

**Remark 6.7.** Note that the proof of the lemma never uses that we started with one initially infected vertex. In fact, we can start from an arbitrary starting configuration  $(S_k(0), A_k^d(0))_{k,d}$  and define the restarted chain as before, always restarting with one infected vertex with label  $k_0$ , even when the starting configuration at time  $T = 0$  is arbitrary. The previous lemma still holds. We will use this fact in the proof of the next lemma.

**Lemma 6.8** (Concentration for Sub-Martingale). *Let*

$$\tau_1 := \inf\{T \geq 0 : S^{(1)}(T) \leq \lfloor (1 - \varepsilon)S(0) \rfloor\}$$

Under the assumptions of Lemma 6.6,

$$\mathbb{P}(H_1(T \wedge \tau_1) - H_1(0) \leq -\lambda) \leq \exp\left(\frac{-\lambda^2}{2T(\zeta_L + \zeta_0)^2}\right) + (I(0) + (T \wedge \lceil \varepsilon S(0) \rceil)) e^{7\|W\|_\infty - 2\zeta_L}$$

where  $H_1(T) = \mathbf{X}^{(1)}(T)\mathbf{v} - \zeta_0 T$ .

*Proof.* Let  $\mathcal{E}_n^{(1)}$  be a draw from the restarted Markov Chain in Definition 6.4 and let  $\mathcal{E}_n^{(2)}$  be a draw from the truncated, restarted Markov Chain in Definition 6.5. Observe that

$$\begin{aligned} \mathbb{P}(H_1(T \wedge \tau_1) - H_1(0) \leq -\lambda) &\leq \mathbb{P}(H_2(T \wedge \tau_1) - H_2(0) \leq -\lambda) \\ &\quad + \mathbb{P}(H_1(T \wedge \tau_1) \neq H_2(T \wedge \tau_1) \text{ or } H_1(0) \neq H_2(0)). \end{aligned} \quad (42)$$

Since the increments of  $H_2(T)$  are bounded by  $\zeta_0 + \zeta_L$ , the first term in Eq. (42) can be bounded using Azuma-Hoeffding giving

$$\mathbb{P}(H_2(T \wedge \tau_1) - H_2(0) \leq -\lambda) \leq \exp\left(\frac{-\lambda^2}{2T(\zeta_L + \zeta_0)^2}\right).$$

Now consider the second term in Eq. (42). Observe that the only case in which  $H_1$  can differ from  $H_2$  is if one of the Poisson degrees are larger than  $\zeta_L$  up to time  $T \wedge \tau_1$ . Thus, letting  $Z_i \sim \text{Pois}(\max_k D_k)$  and observing that the number of Poisson random variables drawn in  $\mathcal{E}_n^{(1)}$  up to time  $T \wedge \tau_1$  is at most  $I(0) + T \wedge (S(\tau) - S(0)) \leq I(0) = T \wedge (\lceil \varepsilon S(0) \rceil) = N_T$ , we may proceed as in the proof of Lemma 4.4 to conclude that

$$\mathbb{P}(H_1(T \wedge \tau_2) \neq H_2(T \wedge \tau_2) \text{ or } H_1(0) \neq H_2(0)) \leq \mathbb{P}\left(\max_{i=1, \dots, N_T} Z_i > \zeta_L\right) \leq N_T e^{7\|W\|_\infty - 2\zeta_L}$$

□

We are now ready to prove Theorem 6.3.

*Proof of Theorem 6.3.* We start by proving the upper bounds on  $\|\mathbf{X}(T)\|_1$ . To this end, we observe that for every change in  $\|\mathbf{X}(T)\|_1$ , one of two events can happen: either a vertex recovers (removing however many active edges are left attached to it), or a half-edge successfully infects another vertex and gives rise to a new Poisson random variable. Since the total number of infection events between  $t = 0$  and  $t = \tau_\varepsilon$  is equal to  $S(\tau_\varepsilon) - S(0) = \lceil \varepsilon S(0) \rceil$  and the number of initially infected vertices is equal to  $I(0)$ , the random variable  $\|\mathbf{X}(T)\|_1$  is stochastically dominated by a sum of  $I(0) + \lceil \varepsilon S(0) \rceil$  independent  $\text{Pois}(D)$  random variables, where  $D := \max_k D_k$ . Since  $\varepsilon S(0) \xrightarrow{p} \infty$  and  $I(0)/\varepsilon S(0) \xrightarrow{p}$

0, Poisson concentration implies that with probability tending to 1,  $\|\mathbf{X}(T)\|_1 \leq \zeta_2 \varepsilon S(0)$  as long as  $\zeta_2 > D$ .

To prove the lower bounds and the finiteness of  $\tau_\varepsilon$ , we switch to the discrete time chains  $\mathcal{E}_n$  defined in Eq. (36), Eq. (37), and Eq. (38). As we will see, the submartingale concentration bounds for the restarted chain  $\mathcal{E}_n^{(1)}$  in Lemma 6.8 imply that

$$\mathbb{P}\left(\mathbf{X}^{(1)}(T)\mathbf{v} > \frac{\zeta_0 T}{2} \quad \text{if } T \leq \tau_1\right) \rightarrow 1 \quad (43)$$

in case (a), and

$$\mathbb{P}\left(\mathbf{X}^{(1)}(T)\mathbf{v} > \frac{\zeta_0 T}{2} \quad \text{if } N_n \leq T \leq \tau_1\right) \rightarrow 1 \quad (44)$$

in case (b). This shows that with high probability, the restarted chain stays bounded away from zero up to the stopping time  $\tau_1$  when  $S^{(1)}(\tau_1) = S(0) - \lceil \varepsilon S(0) \rceil$  in case (a), and for all times between the time  $T_0$  when  $S^{(1)}(T) = S(0) - N_n$  and the time  $\tau_1$  when  $S^{(1)}(\tau_1) = S(0) - \lceil \varepsilon S(0) \rceil$  in case (b) (note that  $T \geq T_0$  implies  $T \geq N_n$  since each step in the discrete chain decreases  $S$  by either 0 or 1). But if the restarted chain does not hit zero before  $S(T) \leq S(0) - \lceil \varepsilon S^{(1)}(0) \rceil$ , the original chain does not hit 0 before this time either, showing that up to this point, the two are the same, which in particular shows that  $\tau_0 = \min\{T \in \mathbb{N} : S(T) \leq S(0) - \lceil \varepsilon S(0) \rceil\} < \infty$  or equivalently that  $\tau_\varepsilon < \infty$  in the continuous time chain. The lower bounds in both cases follow by using the lower bounds in Eq. (43) and Eq. (44) at  $T = \tau_1$ , giving that with probability tending to 1,

$$\|\mathbf{X}(\tau_\varepsilon)\|_1 \geq \frac{\mathbf{X}(\tau_\varepsilon)\mathbf{v}}{v_{\max}} = \frac{\mathbf{X}^{(1)}(\tau_1)\mathbf{v}}{v_{\max}} \geq \frac{\mathbf{X}(0)\mathbf{v} + \zeta_0 \tau_1}{2v_{\max}} \geq \frac{\zeta_0}{2v_{\max}} \varepsilon S(0).$$

This gives the desired the lower bounds with  $\zeta_1 = \zeta_0/2v_{\max}$ .

We are left with the proof of the bounds Eq. (43) and Eq. (44). Let

$$M_1(T) = H_1(T \wedge \tau_1) \quad \text{with} \quad H_1(T) = \mathbf{X}^{(1)}(T)\mathbf{v} - \zeta_0 T.$$

Using Lemma 6.8 with  $\lambda = \frac{1}{2}(M_1(0) + \zeta_0 T) = \frac{1}{2}(\mathbf{X}^{(1)}(0)\mathbf{v} + \zeta_0 T)$  and an appropriately chosen  $\zeta_L$  (to be determined later), we get that

$$\mathbb{P}(M_1(T) - M_1(0) \leq -\lambda) \leq \exp\left(\frac{-(\mathbf{X}^{(1)}(0)\mathbf{v} + \zeta_0 T)^2}{8T(\zeta_L + \zeta_0)^2}\right) + (I(0) + T)e^{7\|W\|_\infty - 2\zeta_L} \quad (45)$$

provided  $\zeta_L$  is large enough to guarantee that Assumption 41 from Lemma 6.6 holds.

Starting with case (a), we set  $T_0 = \lceil I(0)^{1/3} \rceil$  and  $\zeta_L = T^{1/3} \vee T_0^{1/3}$ , observing that Assumption 41 from Lemma 6.6 holds for all large enough  $n$  by monotone convergence and the fact that  $\zeta_L \rightarrow \infty$ . Using the fact that  $\sum_{T \geq T_0} f(T) \leq \int_{T_0}^\infty f(T) dT + \sup_{T \geq T_0} f(T)$  whenever  $f$  is unimodal, the bound (45) implies that

$$\sum_{T \geq T_0} \mathbb{P}(M_1(T) - M_1(0) \leq -\lambda) \leq \int_{T_0}^\infty \left[ \exp\left(\frac{-\zeta_0^2 T}{8(\zeta_L + \zeta_0)^2}\right) + (I(0) + T)e^{7\|W\|_\infty - 2\zeta_L} \right] dT + o(1) \quad (46)$$

The integral of the first term is of order  $O\left(T_0^{2/3} e^{-\Theta(T_0^{1/3})}\right)$  which goes to 0 with  $n$ , while the integral of the second term in the bound is of order  $O\left(T_0^{2/3}(I(0) + T_0)e^{-T_0^{1/3}}\right)$ . Using the fact that  $I(0) \leq T_0^3$ , we see that this term goes to 0 as well.

Finally, we sum the bound Eq. (45) from 0 to  $T_0 - 1$  using our choice of  $\zeta_L$  and  $T_0$ :

$$\begin{aligned} \sum_{T < T_0} \mathbb{P}(M_1(T) - M_1(0) \leq -\lambda) &\leq \sum_{T < T_0} \left[ \exp\left(\frac{-(\mathbf{X}^{(1)}(0)\mathbf{v})^2}{8T(\zeta_L + \zeta_0)^2}\right) + (T_0^3 + T)e^{7\|W\|_\infty - 2\zeta_L} \right] \\ &\leq T_0 \exp\left(\frac{-(\mathbf{X}^{(1)}(0)\mathbf{v})^2}{8T_0(T_0^{1/3} + \zeta_0)^2}\right) + (T_0^4 + T_0^2)e^{7\|W\|_\infty - 2T_0^{1/3}}. \end{aligned}$$

Here, the second term clearly goes to 0 with  $n$ . To see that the first goes to zero, note that  $\mathbf{X}^{(1)}(0)\mathbf{v}$  is a weighted sum of  $I(0)$  independent Poisson random variables with expectation bounded from below by  $I(0)D_{\min}v_{\min}$ , showing that  $\mathbf{X}^{(1)}(0)\mathbf{v} \geq \Theta(I(0))$  with probability tending to 1 as  $n \rightarrow \infty$ . Together, these bounds imply that

$$\begin{aligned} \mathbb{P}\left(\mathbf{X}^{(1)}(T)\mathbf{v} > \frac{\mathbf{X}^{(1)}(0)\mathbf{v} + \zeta_0 T}{2} \quad \text{for } 0 \leq T \leq \tau_1\right) \\ \geq 1 - \mathbb{P}(\text{there exists } T \text{ s.t. } M_1(T) - M_1(0) \leq -\lambda) \xrightarrow{n \rightarrow \infty} 1, \end{aligned}$$

establishing the bound Eq. (43).

To prove the bound Eq. (44), we again use the bound Eq. (46), this time with  $T_0 = N_n$ . Indeed, bounding the integral of the two terms as before, and using that now  $I(0)$  is bounded, the right hand side is again of order  $o(1)$ , implying that

$$\mathbb{P}\left(\mathbf{X}^{(1)}(T)\mathbf{v} > \frac{\mathbf{X}^{(1)}(0)\mathbf{v} + \zeta_0 T}{2} \quad \text{for } N_n \leq T \leq \tau_1\right) \xrightarrow{n \rightarrow \infty} 1$$

as required.  $\square$

The arguments above prove Theorem 6.3 for  $G \sim \text{PSBM}(\mathbf{V}, W')$ . By conditioning on  $G$  being simple, we can immediately extend the results to any  $G \sim \text{SBM}(\mathbf{V}, W)$  with  $W$  depending on  $W'$  as in Lemma 3.1.

### 6.3 Final Size

As we can see from the results obtained in the last two subsections, if the epidemic starts with a constant, non-zero number of initially infected vertices, then a constant fraction of the population ends up being infected with probability converging to the survival probability of the branching process (forest)  $\bigcup_{v \in V^I(0)} \mathcal{F}_{k(v)}^{\text{BP}}$ , while for an initial infection of size which goes to infinity but is  $o(n)$ , this probability goes to 1 as long as  $\mathfrak{R}_0 > 1$ .

In this subsection, we will connect these results to those derived in Section 5 to prove Theorem 2.13 on the final size of an infection starting with  $o(n)$  initially infected vertices. We will need one additional lemma.

**Lemma 6.9.** *Assume that  $W$  is irreducible, and that  $s_k(0) > 0$  for all  $k$ . Let  $\theta_k$  be the survival probability of the backward process defined in Eq. (9) and Eq. (10), let  $\tilde{\mathbf{s}}(t)$  and  $\tilde{\mathbf{x}}(t)$  be the solution to the differential equations Eq. (3) and Eq. (4) with initial conditions  $\tilde{\mathbf{s}}(0)$  and  $\tilde{\mathbf{x}}(0) \neq 0$ , and let  $\tilde{\mathbf{s}}(\infty) = \lim_{t \rightarrow \infty} \tilde{\mathbf{s}}(t)$ . If  $\tilde{\mathbf{x}}(0) \rightarrow 0$  and  $\tilde{\mathbf{s}}(0) \rightarrow \mathbf{s}(0)$ , then  $\tilde{\mathbf{s}}(0) - \tilde{\mathbf{s}}(\infty) \rightarrow \text{diag}(\boldsymbol{\theta})\mathbf{s}(0)$ .*

We will prove the lemma in Appendix A by expressing  $\mathbf{s}(\infty)$  in terms of the backward process defined in Section 5.2 and then showing that the survival probability of that backwards process converges to that of the backward process defined in Eq. (9) and Eq. (10) as  $\tilde{\mathbf{x}}(0) \rightarrow 0$  and  $\tilde{\mathbf{s}}(0) \rightarrow \mathbf{s}(0)$ .

*Proof of Theorem 2.13.*

We first prove the theorem when  $\mathfrak{R}_0 \leq 1$ , in which case  $\boldsymbol{\theta} = 0$ , and our goal is to show that  $\frac{1}{n}(S(0) - S(\infty)) \rightarrow 0$  in probability as  $n \rightarrow \infty$ . As it turns out, this will be an easy consequence of Theorem 2.8 and the following claim, which states that the final size of the epidemic,  $|V^R(\infty)|$ , increases if the set of initially infected vertices increases.

**Claim 6.10.** *Consider the SIR epidemic on an arbitrary graph  $G = (V, E)$ , let  $V^I(0)$  and  $V^S(0)$  be the set of initially susceptible and infected vertices, respectively, and let  $V^R(\infty)$  be the final set of recovered vertices. If we add a set  $W \subset V^S(0)$  to the set of infected vertices, then the SIR dynamic with these new initial conditions can be coupled to the original one such that the new final set of recovered vertices,  $\tilde{V}^R(\infty)$ , is a superset of  $V^R(\infty)$ .*

*Proof.* Consider the following construction of the set  $V^R(\infty)$ , starting from  $V^R(\infty) = V^R(0)$ .

- (i) Add all vertices  $v \in V^I(0)$  to  $V^R(\infty)$  and color them red initially.
- (ii) For every red vertex, draw its recovery time  $T_v \sim \text{Exp}(\gamma)$  and for every edge incident to  $v$ , draw the infection time  $Z_{v,i} \sim \text{Exp}(\eta)$  where  $i \in \{1, \dots, |N(v)|\}$  and  $N(v)$  is the set of neighbors of  $v$ . Once a vertex has been explored, color it black.
- (iii) For all  $i \in N(v)$  such that  $Z_{v,i} < T_v$ , color them red and add  $i$  to  $V^R(\infty)$ . Repeat step (ii).
- (iv) Repeat steps (ii) and (iii) until the infection ends (i.e. there are no more new vertices reachable from step (ii)).

While the order in which vertices added to  $V^R(\infty)$  is not necessarily the same as the one in which they get added in the course of the continuous time epidemic, it is easy to see (and well known) that the final set  $V^R(\infty)$  is correctly obtained by the above construction. The construction also makes it obvious that  $|V^R(\infty)|$  can only become larger if we add vertices to the initial set of infected vertices.  $\square$

Thus, we may assume that we start with more infected vertices and the final size with these new starting conditions will be an upper bound on the final size with the original starting conditions. This in turn will put us into the setting of Theorem 2.8. Concretely, choose  $\hat{\varepsilon} > 0$  arbitrary. By Lemma 6.9 and the fact that  $\boldsymbol{\theta} = 0$ , we can choose  $\varepsilon > 0$  such that whenever  $\|\tilde{\mathbf{x}}(0)\|_2 \leq \varepsilon$  and  $\|\tilde{\mathbf{s}}(0) - \mathbf{s}(0)\|_2 \leq \varepsilon$ , we have that  $\tilde{s}(0) - \tilde{s}(\infty) \leq \hat{\varepsilon}/2$ . Next we change the initial conditions, raising the number of infected vertices from  $I_k(0) = o(n)$  to  $\tilde{I}_k(0) = \lfloor \frac{\varepsilon}{2K} n \rfloor$  and decreasing  $S_k(0)$  to  $\tilde{S}_k(0) = S_k(0) + I_k(0) - \tilde{I}_k(0)$ . With probability tending to 1 as  $n \rightarrow \infty$ , we then have that  $\tilde{I}_k(0) \geq I_k(0)$ , implying that

$$S_k(0) - S_k(\infty) \leq \tilde{S}_k(0) - \tilde{S}_k(\infty).$$

Furthermore, with the help of Assumption 2.5, we have that  $\frac{1}{n}\tilde{I}_k(0) \xrightarrow{P} \tilde{x}_k(0) = \frac{\varepsilon}{2K}$  and  $\frac{1}{n}\tilde{S}_k(0) \xrightarrow{P} \tilde{s}_k(0) = s_k(0) - \frac{\varepsilon}{2K}$ , which implies that with probability tending to 1, the solution of differential

equations with starting conditions  $\tilde{s}(0)$  and  $\tilde{\mathbf{x}}(0) = \tilde{\mathbf{i}}(0)$  obey the bound

$$\tilde{s}(0) - \tilde{s}(\infty) \leq \frac{\hat{\varepsilon}}{2}.$$

With the help of Theorem 2.8 we conclude that with probability tending to 1 as  $n \rightarrow \infty$

$$\frac{1}{n}(S_k(0) - S_k(\infty)) \leq \frac{1}{n}(\tilde{S}_k(0) - \tilde{S}_k(\infty)) \leq \tilde{s}(0) - \tilde{s}(\infty) + \frac{\hat{\varepsilon}}{2} \leq \hat{\varepsilon}.$$

Since  $\hat{\varepsilon} > 0$  was arbitrary, this implies that  $\frac{1}{n}(S_k(0) - S_k(\infty)) \xrightarrow{p} 0$ , as desired.

Next we consider the case when  $\mathfrak{R}_0 > 1$  and  $I(0)$  is bounded in probability. In this case, the existence of the limit  $\pi$  in Theorem 2.13 implies convergence of  $\mathbf{I}(0)$  in probability, with  $\pi$  being the survival probability of a branching forest consisting of the disjoint union of  $I(0)$  independent trees  $\mathcal{T}_{k(v)}^{\text{BP}}$ ,  $v \in V^I(0)$ . On the other hand, if  $N_n \rightarrow \infty$ , the survival probability of this branching forest is the limit of it growing to size at least  $N_n$ , which by Lemma 6.1 and Remark 6.2 is asymptotically equal to the probability that there exists some  $t < \infty$  such that  $S(0) - S(t) \geq N_n$  provided  $N_n$  is chosen appropriately. Consider the stopping time  $\tau_\varepsilon$  defined in Theorem 6.3. By the second statement of the theorem, we then get that for  $\varepsilon > 0$  small enough,

$$\pi = \lim_{n \rightarrow \infty} \mathbb{P}(\exists t < \infty \text{ s.th. } S(0) - S(t) \geq N_n) = \lim_{n \rightarrow \infty} \mathbb{P}(\tau_\varepsilon < \infty). \quad (47)$$

Furthermore, conditioned on  $\tau_\varepsilon < \infty$ , we have that with high probability

$$|s(0) - \hat{s}(\tau_\varepsilon)| \leq 2\varepsilon s(0) \quad \text{and} \quad \tilde{\zeta}_1 \varepsilon s(0) \leq \|\hat{\mathbf{x}}(\tau_\varepsilon)\|_1 \leq \tilde{\zeta}_2 \varepsilon s(0)$$

for some constants  $0 < \tilde{\zeta}_1 < \tilde{\zeta}_2 < \infty$ . Here we moved from the statements in Theorem 6.3 for  $S(t)$  and  $\mathbf{X}(t)$  to those for  $\hat{x}_k(t) = \frac{X_k(t)}{nD_k}$  and  $\hat{\mathbf{s}}(t) = \frac{\mathbf{S}(t)}{n}$  by absorbing factors of  $\frac{1}{D_{\max}}$  and  $\frac{1}{D_{\min}}$  into  $\tilde{\zeta}_1$  and  $\tilde{\zeta}_2$  and using Assumption 2.5 to control the difference between  $\hat{s}(0)$  and  $s(0)$ . Finally, note that with high probability

$$\hat{\mathbf{y}}_{\tau_\varepsilon} \in U_0 \quad \text{and} \quad \max_{u \in V} d_u(\tau_\varepsilon) \leq \log n$$

since by Lemma 4.4 these statements actually hold for all  $t$ . Defining the compact set

$$\mathcal{C}_\varepsilon = \{\hat{\mathbf{y}}(0) \in U_0 : |s(0) - \tilde{s}(0)| \leq 2\varepsilon s(0) \quad \text{and} \quad \tilde{\zeta}_1 \varepsilon s(0) \leq \|\hat{\mathbf{x}}(0)\|_1 \leq \tilde{\zeta}_2 \varepsilon s(0)\}$$

and the event

$$\mathcal{A}_\varepsilon = \{\tau_\varepsilon < \infty, \hat{\mathbf{y}}_{\tau_\varepsilon} \in \mathcal{C}_\varepsilon \text{ and } \max_{u \in V} d_u(\tau_\varepsilon) \leq \log n\}$$

we therefore have that

$$\pi = \lim_{n \rightarrow \infty} \mathbb{P}(\exists t < \infty \text{ s.th. } S(0) - S(t) \geq N_n) = \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{A}_\varepsilon). \quad (48)$$

Note that  $\mathcal{C}_\varepsilon$  obeys the conditions in Eq. (5.11) provided  $\varepsilon$  is small enough to guarantee that  $|s(0) - \tilde{s}(0)| \leq 2\varepsilon s(0)$  implies  $\tilde{s}_k(0) > 0$  for all  $k$  (if needed, we will decrease  $\varepsilon_0$  to make sure this holds).



Let  $\tilde{\mathbf{y}}_t$  denote the solution at time  $t$  to the differential equations starting at  $\tilde{\mathbf{y}}_0 = \hat{\mathbf{y}}_{\tau_\varepsilon}$ , and let  $\hat{\varepsilon} > 0$  be arbitrary. In view of Lemma 6.9, we can choose  $\varepsilon'_0 > 0$  such that  $|\tilde{\mathbf{s}}(0) - \tilde{\mathbf{s}}(\infty) - \text{diag}(\boldsymbol{\theta})\mathbf{s}(0)| \leq \varepsilon'$  whenever  $\varepsilon \leq \varepsilon'_0$  and  $\tilde{\mathbf{y}}(0) \in \mathcal{C}_\varepsilon$ . For  $\varepsilon \leq \min\{\varepsilon_0, \varepsilon'_0, \frac{1}{2}\hat{\varepsilon}\}$  we then have that

$$\begin{aligned} & \mathbb{P}(\|(\mathbf{s}(0) - \hat{\mathbf{s}}(\infty)) - \text{diag}(\boldsymbol{\theta})\mathbf{s}(0)\|_2 > 4\hat{\varepsilon} \mid \mathcal{A}_\varepsilon) \\ & \leq \mathbb{P}(\|\mathbf{s}(0) - \tilde{\mathbf{s}}(0)\|_2 > 2\hat{\varepsilon} \mid \mathcal{A}_\varepsilon) + \mathbb{P}(\|(\tilde{\mathbf{s}}(0) - \tilde{\mathbf{s}}(\infty)) - \text{diag}(\boldsymbol{\theta})\mathbf{s}(0)\|_2 > \hat{\varepsilon} \mid \mathcal{A}_\varepsilon) \\ & \quad + \mathbb{P}(\|\tilde{\mathbf{s}}(\infty) - \hat{\mathbf{s}}(\infty)\|_2 > \hat{\varepsilon} \mid \mathcal{A}_\varepsilon). \end{aligned}$$

The first term is 0 since  $\tilde{\mathbf{y}}_0 = \mathbf{y}_{\tau_\varepsilon} \in \mathcal{C}_\varepsilon$  implies  $\|\mathbf{s}(0) - \tilde{\mathbf{s}}(0)\|_2 \leq \|\mathbf{s}(0) - \tilde{\mathbf{s}}(0)\|_1 \leq 2\varepsilon s(0) \leq \hat{\varepsilon}$ , the second term is 0 by Lemma 6.9 and the fact that  $\varepsilon \leq \varepsilon'_0$ , and the last term goes to 0 by Theorem 5.11. Since  $\mathbb{P}(\mathcal{A}_\varepsilon) \rightarrow \pi$ , this proves Theorem 2.13 in the case when  $\mathfrak{R}_0 > 1$  and  $I(0)$  is bounded in probability.

When  $\mathfrak{R}_0 > 1$  and  $I(0)$  is not bounded in probability, we in fact have that  $I(0) \xrightarrow{P} \infty$  since otherwise the limit defining  $\pi$  would not exist. But then

$$\pi = 1 = \lim_{n \rightarrow \infty} (\mathcal{A}_\varepsilon),$$

where the second equality follows by the first statement of Theorem 6.3. The above argument then again implies the statement of Theorem 2.13.  $\square$

Finally, we prove Corollary 2.14.

*Proof.* For the proof of the corollary, we will choose a coupling of  $\mathcal{T}_{G,v}(t)$  and  $\mathcal{T}_k^{\text{BP}(M)}(t)$  and a sequence  $N'_n$  with  $N'_n \rightarrow \infty$  and  $N'_n \leq N_n$  such the statements of Lemma 2.12 hold for  $N'_n$ , i.e.,

$$\mathbb{P}\left(\mathcal{T}_{G,v}(t) = \mathcal{T}_k^{\text{BP}(M)}(t) \text{ for all } t \text{ such that } |\mathcal{T}_k^{\text{SBM}}(t)| \leq N'_n\right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

We start with the case  $\mathfrak{R}_0 \leq 1$ . Then  $\pi_k = 0$  and the branching process  $\mathcal{T}_k^{\text{BP}(M)}(\cdot)$  dies out in finite time, implying that with probability tending to 1,  $\mathcal{T}_k^{\text{BP}(M)}(\cdot) \leq N'_n$ . Combined with Lemma 2.12, we conclude that with probability tending to 1,  $\mathcal{T}_k^{\text{BP}(M)}(t) = \mathcal{T}_k^{\text{SBM}}(t)$  for all  $t$  and

$$\mathbb{P}(S(0) - S(\infty) > N_n) \leq \mathbb{P}(S(0) - S(\infty) > N'_n) \xrightarrow{n \rightarrow \infty} \pi_k.$$

This proves (ii) and (i) in the case where  $\mathfrak{R}_0 \leq 1$ , with the exception of the exponentially decaying tails statement for  $\mathfrak{R}_0 < 1$ . To prove that, we first note that the offspring distribution of a Poisson branching process has exponentially decaying tails. Since  $\mathcal{T}_k^{\text{BP}(M)}(\infty)$  was constructed as a sub-tree of  $\text{BP}_k(M)$ , its offspring distribution has exponentially decaying tails as well. For  $\mathfrak{R}_0 < 1$ ,  $\mathcal{T}_k^{\text{BP}(M)}(\infty)$  is therefore a sub-critical branching process whose offspring distribution has exponentially decaying tails, implying that  $|\mathcal{T}_k^{\text{BP}(M)}(\infty)|$  itself has exponentially decaying tails (see Lemma A.5 for a proof of this fact).

Turning to the case  $\mathfrak{R}_0 > 1$ , we first prove the following claim.

**Claim 6.11.** *Let  $N'_n = o(n)$  be a sequence for which the statements from Lemma 2.12 hold, and assume that  $\mathfrak{R}_0 > 1$ . If  $0 < \varepsilon < \boldsymbol{\theta} \cdot \mathbf{s}(0)$ , then*

$$\lim_{n \rightarrow \infty} \mathbb{P}(S(0) - S(\infty) \in (N'_n, \varepsilon n)) = 0.$$

*Proof.* Recall that  $\pi_k$  is the survival probability of the backward branching process. Then for an infection starting from one vertex with label  $k$ ,  $\pi = \pi_k$ . This implies that

$$\lim_{n \rightarrow \infty} \mathbb{P}(S(0) - S(\infty) \leq N'_n) = \lim_{n \rightarrow \infty} \mathbb{P}(|\mathcal{T}_k^{\text{SBM}}(\infty)| \leq N'_n) = 1 - \pi_k,$$

or equivalently that

$$\lim_{n \rightarrow \infty} \mathbb{P}(S(0) - S(\infty) > N'_n) = \pi_k.$$

Furthermore, Theorem 2.13 implies that

$$\lim_{n \rightarrow \infty} \mathbb{P}(S(0) - S(\infty) \geq \varepsilon n) = \pi_k.$$

Since  $N'_n = o(n)$ ,  $\{S(0) - S(\infty) \geq \varepsilon n\} \subseteq \{S(0) - S(\infty) > N'_n\}$  and we then see that

$$\lim_{n \rightarrow \infty} \mathbb{P}(S(0) - S(\infty) \in (N'_n, \varepsilon n)) = 0.$$

□

Using the just proven claim, we see that for  $\varepsilon > 0$  small enough, the probability that  $S(0) - S(\infty) > N_n$  is asymptotically equal to the probability that  $S(0) - S(\infty) > \varepsilon n$ , which converges to  $\pi_k$  by Theorem 2.13. This proves (i) in the case where  $\mathfrak{R}_0 > 1$ .

To prove (iii), we need to show that for all fixed  $N < \infty$ ,

$$\mathbb{P}(S(0) - S(\infty) = N \mid S(0) - S(\infty) \leq N_n) \rightarrow \mathbb{P}(|\mathcal{T}_k^{\text{BP}(M)}(\infty)| = N \mid |\mathcal{T}_k^{\text{BP}(M)}(\infty)| < \infty).$$

But this follows immediately from (i) and Lemma 2.12. Indeed, (i) implies that  $\mathbb{P}(S(0) - S(\infty) \leq N_n) \rightarrow 1 - \pi_k = \mathbb{P}(|\mathcal{T}_k^{\text{BP}(M)}(\infty)| < \infty)$  while Lemma 2.12 together with the fact that  $N_n \rightarrow \infty$  implies that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}(\{S(0) - S(\infty) = N\} \cap \{S(0) - S(\infty) \leq N_n\}) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(S(0) - S(\infty) = N) = \mathbb{P}(|\mathcal{T}_k^{\text{BP}(M)}(\infty)| = N). \end{aligned}$$

Finally, we prove (iv). It suffices to show that

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{\mathbf{S}(0) - \mathbf{S}(\infty)}{n} - \text{diag}(\boldsymbol{\theta})\mathbf{s}(0)\right| < \varepsilon\right) = \lim_{n \rightarrow \infty} \mathbb{P}(S(0) - S(\infty) > N_n).$$

Let  $\varepsilon \in (0, \boldsymbol{\theta} \cdot \mathbf{s}(0))$ . Observe that for  $N_n = o(n)$  and  $n$  large enough, the event

$$\left\{\left|\frac{\mathbf{S}(0) - \mathbf{S}(\infty)}{n} - \text{diag}(\boldsymbol{\theta})\mathbf{s}(0)\right| < \varepsilon\right\}$$

implies  $\{S(0) - S(\infty) > N_n\}$ . Thus

$$\mathbb{P}\left(\left\{\left|\frac{\mathbf{S}(0) - \mathbf{S}(\infty)}{n} - \text{diag}(\boldsymbol{\theta})\mathbf{s}(0)\right| < \varepsilon\right\} \cap \{S(0) - S(\infty) > N_n\}\right) = \mathbb{P}\left(\left|\frac{\mathbf{S}(0) - \mathbf{S}(\infty)}{n} - \text{diag}(\boldsymbol{\theta})\mathbf{s}(0)\right| < \varepsilon\right) \xrightarrow{n \rightarrow \infty} \pi_k$$

where the convergence is true by Theorem 2.13. Furthermore

$$\lim_{n \rightarrow \infty} \mathbb{P}(S(0) - S(\infty) > N_n) = \pi_k$$

by (i) and this gives the desired result. □

## 7 Discussion

In this section, we discuss our results and relate them to notions like herd immunity, the force of an infection, and the so-called pair-approximation, concluding with the discussion of possible extensions. Throughout, we consider the setting where  $s_k(0) > 0$  for all  $k$  and  $W$  is irreducible.

### 7.1 Herd Immunity

The herd immunity threshold for a disease is the threshold at which the infection starts slowing down due to the fact that a significant proportion of the (finite) population have been infected, and new infection attempts are unlikely to lead to a successful infection. Mathematically, this is the point at which the effective reproductive number falls below one, and new infections die out exponentially fast.

In the deterministic, homogeneous mixing setting for the SIR epidemic, described by the differential equations

$$\frac{ds}{dt} = -\eta si \tag{49}$$

$$\frac{di}{dt} = \gamma \left( \frac{\eta s}{\gamma} - 1 \right) i \tag{50}$$

$$\frac{dr}{dt} = \gamma i \tag{51}$$

the effective reproductive number is given by  $\mathfrak{R}_{\text{eff}}(t) = \eta s(t)/\gamma$ , and the herd immunity threshold is at  $\mathfrak{R}_{\text{eff}}(t) = 1$ , below which the number of infected vertices is exponentially decreasing.

To discuss the notions of herd immunity and effective reproductive number for SIR on the stochastic block model, we recall the vector form of the differential equations for the number of susceptible vertices and the number of active half-edges for SIR on the SBM:

$$\frac{d\mathbf{s}(t)}{dt} = -(\eta + \gamma)\mathbf{x}(t)C(t) \tag{52}$$

$$\frac{d\mathbf{x}(t)}{dt} = (\eta + \gamma)\mathbf{x}(t)(C(t) - 1) \tag{53}$$

where  $C(t) = \frac{\eta}{\eta + \gamma} W \text{diag}(\mathbf{s}(t))$ . As pointed out in Remark 5.6 in Section 5.3, which we rephrase here to keep our discussion self-contained, the largest eigenvalue of this matrix,  $\mathfrak{R}_{\text{eff}}(t)$ , is the appropriate generalization of the effective reproductive number to our setting, with  $\mathfrak{R}_{\text{eff}}(t) = 1$  determining the herd-immunity threshold.

Indeed, if  $\mathfrak{R}_{\text{eff}}(t_0) < 1$  and  $\mathbf{v}_0$  is the corresponding right eigenvector with all positive entries, then

$$\begin{aligned} \frac{d\mathbf{x}(t)}{dt} \cdot \mathbf{v}_0 &= (\eta + \gamma)\mathbf{x}(t)(C(t) - 1)\mathbf{v}_0 \\ &\leq (\eta + \gamma)\mathbf{x}(t)(C(t_0) - 1)\mathbf{v}_0 \\ &= (\eta + \gamma)(\mathfrak{R}_{\text{eff}}(t_0) - 1)\mathbf{x}(t) \cdot \mathbf{v}_0 \quad \text{for all } t \geq t_0 \end{aligned}$$

showing that  $\mathbf{x}(t) \cdot \mathbf{v}$  and hence  $\|\mathbf{x}(t)\|_1$  is decaying exponentially in  $t - t_0$  for any initial condition  $\mathbf{x}(t_0)$ . As a consequence, a small infection at time  $t_0$  will only have small overall impact on the final size of the infection. By contrast, if  $\mathfrak{R}_{\text{eff}}(t_0) > 1$ , then  $\mathbf{x}(t) \cdot \mathbf{v}$  will grow exponentially in  $t - t_0$

until  $\mathfrak{R}_{\text{eff}}(t) = 1$ . Thus no matter how small  $\mathbf{x}(t_0)$  is, the infection will always have a sizable effect on the final size, since the infection won't start to recede before it has infected enough susceptible vertices to drive  $\mathfrak{R}_{\text{eff}}(t)$  below one.

Notice that here, the parameters that drive the infection are not the number of infected vertices as in the case of the standard SIR, but rather the numbers of active half-edges in each community weighted by the connectivity between communities. This quantity differs by community based on the connectivity in between communities and gives rise to  $K$  parameters that govern the trajectory of the epidemic, one in each community. This implies that the trajectories of the epidemic curves in different communities may peak at different times, and furthermore that it is possible for the epidemic to be spreading in a community despite the global herd immunity threshold being reached in the population.

We make two observations about  $\mathfrak{R}_{\text{eff}}(t)$  and its relation to the number of active half-edges at time  $t$ . First, notice that if  $\mathfrak{R}_{\text{eff}}(t) > 1$  then there exists at least one community  $k$  such that  $dx_k(t)/dt > 0$ . In other words, before the herd immunity threshold, at least one community experiences exponential growth in the number of half-edges. Second, if  $\mathfrak{R}_{\text{eff}}(t) < 1$ , then in at least one community  $k$ ,  $dx_k(t)/dt < 0$  and there is exponential decay of half-edges in that community. We can see this as follows: let  $\mathbf{v}_0$  be the right Perron-Frobenius eigenvector of  $C(t)$ . If  $\mathfrak{R}_{\text{eff}}(t) > 1$  then

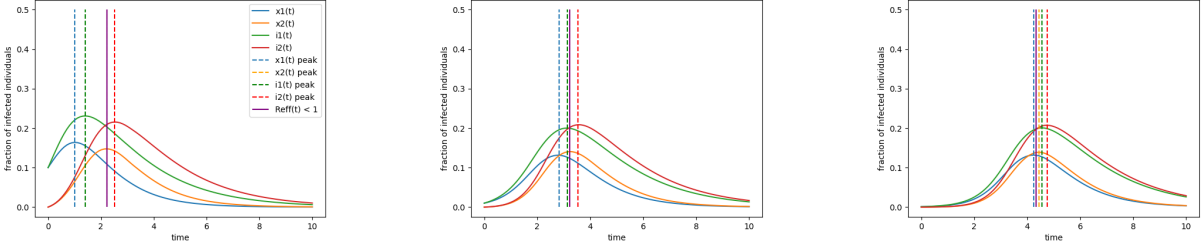
$$\begin{aligned} \frac{d\mathbf{x}(t)}{dt} \cdot \mathbf{v}_0 &= (\eta + \gamma)\mathbf{x}(t)(C(t) - 1)\mathbf{v}_0 \\ &= (\eta + \gamma)(\mathfrak{R}_{\text{eff}}(t) - 1)\mathbf{x}(t) \cdot \mathbf{v}_0 > 0. \end{aligned}$$

Since  $\mathbf{v}_0$  has all strictly positive entries, this implies that at least one component of  $d\mathbf{x}(t)/dt$  must be larger than 0. Similarly, in the case where  $\mathfrak{R}_{\text{eff}}(t) < 1$ , there must exist some  $k$  such that  $dx_k(t)/dt$  is less than 0.

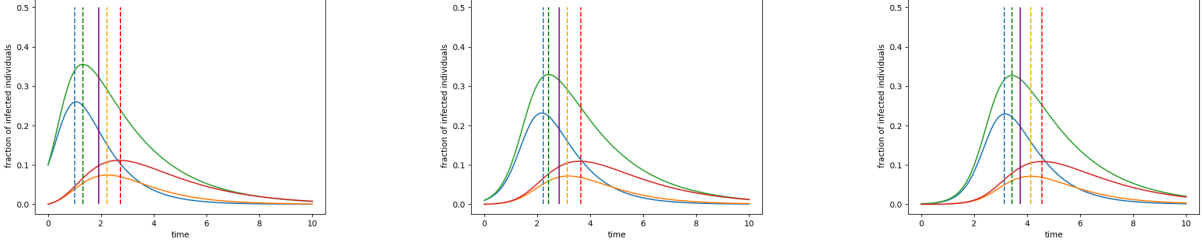
Thus, after the herd immunity threshold the number of half-edges in at least one community is decreasing. Note that eventually all components decay exponentially (we prove this in Lemma B.1 in Appendix B) furthermore, though a small infection at  $t_0$  will never grow large this does not preclude some components of  $\mathbf{x}(t)$  from growing even after the herd immunity threshold. All we know is that a small infection can only grow by at most a constant factor before dying out.

## 7.2 Numerical Simulations for $K = 2$

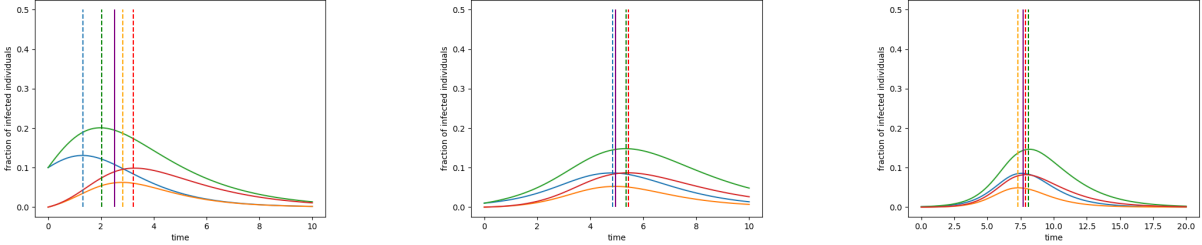
In this section, we demonstrate the relationship between the herd immunity threshold and the various maxima of the components of  $\mathbf{x}$  and  $\mathbf{i}$ . In Figure 1 we illustrate the relationship between  $\mathfrak{R}_{\text{eff}}(t)$  and  $\mathbf{x}(t)$  in the differential equations for a two-community stochastic block model. We start with a certain proportion,  $\varepsilon$ , of infected individuals in the first community and plot the curves for  $x_1(t)$ ,  $x_2(t)$ ,  $i_1(t)$ , and  $i_2(t)$ . The peaks of each curve are also plotted in relation to the first time  $t$  at which  $\mathfrak{R}_{\text{eff}}(t) < 1$ . Observe that the threshold  $\mathfrak{R}_{\text{eff}}(t) < 1$  always occurs between the first and last peak in the  $\mathbf{x}(t)$ . The first row simulates the ODE with the two communities being the same size, in which case as  $\varepsilon$  gets smaller, the trajectory of the curves for the two communities converge. The last row simulates the ODE for two communities of different sizes, with the contact matrix  $W$  adjusted so that the expected number of edges in each community is the same. For small  $\varepsilon$ , the peaks between the two communities are visibly separated, but as  $\varepsilon$  gets smaller, the peaks get closer.



(a) Simulations for two communities of equal size ( $s_1(0) = s_2(0) = 1/2$ ) and  $W = \begin{pmatrix} 10 & 1 \\ 1 & 10 \end{pmatrix}$ .



(b) Simulations for two communities of different sizes ( $s_1(0) = 2/3, s_2(0) = 1/3$ ) and  $W = \begin{pmatrix} 10 & 1 \\ 1 & 10 \end{pmatrix}$ .



(c) Simulations for two communities of different sizes ( $s_1(0) = 2/3, s_2(0) = 1/3$ ) and  $W = \begin{pmatrix} 5 & 1 \\ 1 & 10 \end{pmatrix}$  so the expected number of edges within each community is the same.

Figure 1: A simulation of our system of ordinary differential equations on a two-community stochastic block model. We vary parameter values of  $\varepsilon$ , the initial fraction of infected individuals going from 0.1, 0.01, 0.001 from the left column to the right column;  $W$ , the contact matrix; and  $s_1(0)$  and  $s_2(0)$ , the initial proportion of susceptible individuals in each community. For all simulations,  $\eta = \gamma = 1/2$ .

Notice that in all of the simulations, in each community the peak of the fraction of infected individuals occurs after the peak in the number of active half-edges for that community, and furthermore  $x_k$  stays below  $i_k$  for all times  $t > 0$ . As the following lemma shows, this is not a numerical coincidence, but a direct consequence of the differential equations for  $\mathbf{s}$ ,  $\mathbf{x}$  and  $\mathbf{i}$ .

**Lemma 7.1.** *Assume that  $W$  is irreducible, that  $s_k(0) > 0$  for all  $k \in [K]$ , that  $x(0) > 0$ , and that  $x_k(0) = i_k(0)$ . Then  $x_k(t) < i_k(t)$  for all  $t > 0$ . If we assume in addition that  $\left. \frac{dx_k}{dt} \right|_{t=0} > 0$ , then the first local maximum of  $x_k$ , occurs after the first local maximum of  $i_k$ .*

While the details of the proof are a little tedious, the basic idea is simple and based on the observation that

$$\frac{di_k}{dt} - \frac{dx_k}{dt} = (\eta + \gamma)x_k - \gamma i_k,$$

showing that at least near  $t = 0$ , where  $\frac{x_k}{i_k}$  starts at 1,  $x_k$  grows slower than  $i_k$ . To prove our statement on the local maxima, all one needs to show is that the first point where  $\frac{x_k}{i_k}$  falls below  $\frac{\gamma}{\eta + \gamma}$  happens after the first local maximum of  $x_k$ . We will prove this fact in Appendix B, where we give the complete proof of Lemma 7.1.

In Figure 2 we visualize the LLN behavior for the SIR epidemic on a stochastic block model with two communities. The figures contain 100 trajectories of the fraction of infected vertices in simulated epidemics with

$$W = \begin{pmatrix} 5 & 1 \\ 1 & 10 \end{pmatrix}, \quad \varepsilon = 0.01, \quad \eta = \gamma = 1/2.$$

We sampled networks with size  $n = 200, 2000$  and  $10,000$  nodes split equally amongst the two communities with  $n_1 = n_2 = n/2$ , where  $n_1$  and  $n_2$  are the number of nodes in community 1 and 2, respectively. A realization  $G$  of the stochastic block model with parameters  $W, (n_1, n_2)$  was drawn and then an epidemic, in which an  $\varepsilon$  fraction of the population in community 1 was initially infected and the remaining nodes were all susceptible, was run on  $G$ . The solid lines in the figure depict the corresponding ODE solutions for  $i_1(t)$  and  $i_2(t)$ . The stochastic epidemic was simulated using algorithms from Miller and Ting [2019].

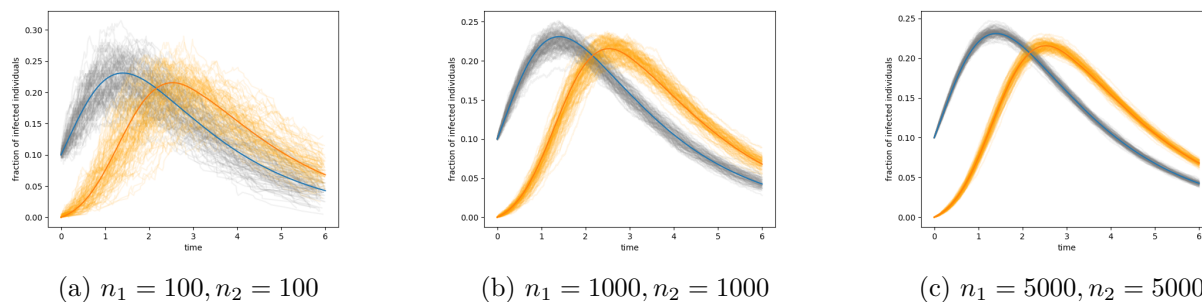


Figure 2: Simulations for one hundred stochastic SIR trajectories of the fraction of infected individuals on a stochastic block model with two communities of the same size. From left to right, we increase the number of nodes in the network keeping the ratio  $\frac{n_1}{n_2}$  the same. The solid blue and orange curves are the solution to the differential equations with the same parameters.

### 7.3 Force of the Infection

From a practical perspective, it is hard to estimate  $\mathbf{x}(t)$ , since this quantity is generated from a coupling of the SIR epidemic on the SBM and is difficult to interpret in the context of the epidemic. However, what can be observed is the force of infection. In the epidemiology literature, force of infection (also called hazard rate or incidence rate) refers to the rate at which susceptible individuals in a population are infected Vynnycky and White [2010]. In the standard SIR setting, the force of infection on the susceptible population is given by  $\eta \frac{I}{n}$ , while in our setting, the force of infection

on susceptible vertices in community  $k$  is given by  $\eta \sum_{\ell} \frac{X_{\ell}}{nD_k} W_{\ell k}$ . Again, observe that the force of infection is determined by the active half-edges and *not* by the number of infected individuals in each community. Denoting the limit of the force of the infection on individuals of with label  $k$  by  $F_k$ , we see that in terms of the vector  $\mathbf{F} = (F_k)_{k \in [K]}$  the limiting dynamics of the epidemic is described by

$$\frac{d\mathbf{s}(t)}{dt} = -\text{diag}(\mathbf{F}(t))\mathbf{s}(t) \quad (54)$$

$$\frac{d\mathbf{F}(t)}{dt} = (\eta + \gamma)\mathbf{F}(t)(C^T(t) - 1) \quad (55)$$

with initial conditions for  $\mathbf{F}$  given by  $F_k(0) = \eta \sum_{\ell} i_{\ell}(0) W_{\ell k}$ , and  $C^T(t)$  denoting the transpose of  $C(t)$ . Since  $C^T(t)$  has the same eigenvalues as  $C(t)$ , this shows that our discussion of the herd immunity threshold from the last subsection can equivalently be formulated in terms of the observable force of the infection, rather than the unobservable number of active half-edges.

## 7.4 Pair-Approximation

In this paper, we couple the stochastic epidemic on the stochastic block model to a dynamic process where the graph and the infection are revealed at the same time, allowing us to derive differential equations for the stochastic epidemic and to understand the parameters that drive the infection. An alternative, *a priori* not mathematically rigorous, approach to deriving a system of limiting ordinary differential equations for the stochastic epidemic is to draw a realization of the graph first and then use a mean-field approximation, relying on the number of edges connecting individuals in different compartments. Below, we derive a system of ODEs using the heuristics of one such mean-field approach: the pair-approximation (cf. Ch.7 of [Andersson and Britton \[2000\]](#) or, Ch. 4 of [Kiss et al. \[2017\]](#)). Note that this approach is conjectured to be exact for locally tree like graphs, but to our knowledge, this has not been established rigorously [[Kiss, Miller, and Simon, 2017](#); [Kiss, Morris, Sélley, Simon, and Wilkinson, 2015](#); [Kiss, Kenah, and Rempala, 2022](#); [Sharkey, Kiss, Wilkinson, and Simon, 2015](#)].

Our quantities of interest remain the expected number of susceptible, infected, and recovered vertices in community  $k$  at time  $t$ . Let  $G \sim \text{SBM}(\mathbf{V}, W)$  be a realization of the stochastic block model on  $n$  vertices, let  $u \in [n]$  be the labels of the nodes of  $G$ , and let  $g$  be the adjacency matrix of  $G$ . For  $A \in \{S, I, R\}$ , let  $I_u^A$  be the indicator function that  $u$  is susceptible, infected, or recovered, respectively, and let

$$[A_k]_n = \sum_{\substack{u \in [n] \\ k(u)=k}} I_u^A, \quad [A_k B_{\ell}]_n = \sum_{\substack{u, v \in [n] \\ k(u)=k \\ k(v)=\ell}} I_u^A g_{uv} I_v^B, \quad [A_k B_{\ell} C_m]_n = \sum_{\substack{u, v, w \in [n], u \neq w \\ k(u)=k \\ k(v)=\ell \\ k(w)=m}} I_u^A g_{uv} I_v^B g_{vw} I_w^C. \quad (56)$$

We say an edge  $uv$  is of type  $A_k B_{\ell}$  if  $u$  has the community label  $k$  and is in state  $A$  and  $v$  has label  $\ell$  and is in state  $B$ , i.e., if  $uv$  contributes to  $[A_k B_{\ell}]$ , and similarly for a type  $A_m B_k C_{\ell}$  path  $uvw$ .

Assume that all quantities in Eq. (56) when normalized by  $n$  converge to deterministic limits as  $n \rightarrow \infty$ . Observing that the rate at which a given susceptible vertex  $v$  becomes infected is equal to the number of edges joining  $v$  to an infected vertex, and that the rate at which infected vertices recover, is equal to  $\gamma$ , we heuristically obtain the following system of differential equations:

$$\frac{d}{dt}[S_k] = -\eta \sum_{\ell \in [K]} [I_{\ell} S_k] \quad (57)$$

$$\frac{d}{dt}[I_k] = \eta \sum_{\ell \in [K]} [I_\ell S_k] - \gamma[I_k] = -\frac{d}{dt}[S_k] - \gamma[I_k] \quad (58)$$

with errors which should be  $o(n)$  as  $n \rightarrow \infty$ . With a little bookkeeping, this heuristic also gives

$$\frac{d}{dt}[I_\ell S_k] = -\gamma[I_\ell S_k] + \eta \sum_{m \in [K]} [I_m S_\ell S_k] - \eta \left( [I_\ell S_k] + \sum_{m \in [K]} [I_m S_k I_\ell] \right) \quad (59)$$

$$\frac{d}{dt}[S_\ell S_k] = -2\eta \sum_{m \in [K]} [I_m S_\ell S_k]. \quad (60)$$

We give a heuristic explanation for Eq. (59), leaving the heuristic derivation of Eq. (60) to the reader. At rate  $\gamma[I_\ell S_k]$ , the infected endpoint in an  $I_\ell S_k$  edge recovers and the number of  $I_\ell S_k$  edges decreases by one. At rate  $\eta[I_m S_\ell S_k]$  for any  $m \in [K]$ , the infected vertex in an  $I_m S_\ell S_k$  infects its  $S_\ell$  neighbor and adds an  $I_\ell S_k$  edge. At rate  $\eta[I_m S_k I_\ell]$  for any  $m \in [K]$ , an  $I_m S_k I_\ell$  set of edges becomes an  $I_m I_k I_\ell$  set of edges and removes an  $S_k I_\ell$  edge. Finally, at rate  $\eta[I_\ell S_k]$  the  $I_\ell$  node infects the  $S_k$  node and removes an  $I_\ell S_k$  edge.

In the so-called pair-approximation, it is then assumed that we can make the approximation

$$[A_m B_k C_\ell] = \frac{[A_m B_k][B_k C_\ell]}{[B_k]},$$

giving us the following equations

$$\frac{d}{dt}[I_\ell S_k] = -(\eta + \gamma)[I_\ell S_k] + \eta \sum_{m \in [K]} \left( \frac{[I_m S_\ell][S_\ell S_k]}{[S_\ell]} - \frac{[I_m S_k][S_k I_\ell]}{[S_k]} \right) \quad (61)$$

$$\frac{d}{dt}[S_\ell S_k] = -2\eta \sum_{m \in [K]} \frac{[I_m S_\ell][S_\ell S_k]}{[S_\ell]}. \quad (62)$$

Finally, under the assumption that the initially susceptible vertices of type  $k$  are a random set of size  $[S_k]$ , we can assume that at time 0,

$$[S_k S_\ell]_{t=0} = \frac{W_{k\ell}}{n} [S_k]_{t=0} [S_\ell]_{t=0}.$$

On the other hand, combining Eq. (57) and Eq. (62) and using the symmetry  $k \leftrightarrow \ell$  of  $[S_k S_\ell]$ , we get

$$\frac{d}{dt} \left( [S_k S_\ell] - \frac{W_{k\ell}}{n} [S_k][S_\ell] \right) = -\eta \sum_{m \in [K]} \left( \frac{[I_m S_\ell]}{[S_k]} + \frac{[I_m S_k]}{[S_\ell]} \right) \left( [S_\ell S_k] - \frac{W_{k\ell}}{n} [S_\ell][S_k] \right) \quad (63)$$

which we can integrate to get that

$$[S_\ell S_k] = [S_\ell][S_k] \frac{W_{k\ell}}{n} \quad (64)$$

for all  $t$ .



Let  $[a_k] = n_{-1}[A_k]$  for  $A \in S, I, R$  and let  $[y_k] = n^{-1} \sum_{\ell \in [K]} [I_\ell S_k]$ . The pair approximation and Eq. (64) imply the following system of differential equations:

$$\frac{d}{dt}[s_k] = -\eta[y_k] \quad (65)$$

$$\frac{d}{dt}[i_k] = -\frac{d}{dt}[s_k] - \gamma[i_k] \quad (66)$$

$$\frac{d}{dt}[y_k] = -(\eta + \gamma)[y_k] + \eta \sum_{\ell} [y_\ell] W_{\ell k}[s_k] - \frac{\eta}{[s_k]} [y_k]^2. \quad (67)$$

As we will see, the time evolution for  $s_k$  and  $i_k$  following from these equations is identical to the time evolution derived rigorously in this paper. To show this, we consider the force of the infection,

$$F_k = -\frac{1}{[s_k]} \frac{d}{dt}[s_k] = \eta \frac{[y_k]}{[s_k]}$$

and use Eq. (64) and Eq. (67) to derive the differential equation

$$\begin{aligned} \frac{dF_k}{dt} &= -\eta \frac{[y_k]}{[s_k]^2} \frac{d[s_k]}{dt} + \eta \frac{1}{[s_k]} \frac{d[y_k]}{dt} = \left( \eta \frac{[y_k]}{[s_k]} \right)^2 - (\eta + \gamma) \left( \eta \frac{[y_k]}{[s_k]} \right) + \eta^2 \sum_{\ell} [y_\ell] W_{\ell k} - \left( \eta \frac{[y_k]}{[s_k]} \right)^2 \\ &= -(\eta + \gamma)F_k + \eta \sum_{\ell} F_\ell [s_\ell] W_{\ell k}, \end{aligned}$$

which is the time evolution for the force of infection derived rigorously in the previous section.

## 7.5 Extensions

In this work, we have given a rigorous derivation of the limiting system of ordinary differential equations for the stochastic SIR epidemic on the stochastic block model. While there are some general alternative approaches for approximating the epidemic trajectories using e.g. the forward and backward branching processes of the model (see [Barbour and Reinert \[2013\]](#); [Bhamidi et al. \[2014\]](#)) or local graph limits (see [Alimohammadi et al. \[2023\]](#); [Cocomello and Ramanan \[2023\]](#)), to our knowledge this paper is first to rigorously establish a system of differential equations for the limiting behavior allowing for general (including sublinear) starting conditions. Our law of large numbers results give convergence of the stochastic epidemic to the limiting ODE for all  $t$  finite and infinite, and demonstrate that the infection curves in different communities are driven by different parameters. In particular, we prove a law of large numbers for the  $t$  to infinity limit (final size) of the epidemic. To prove the final size result, we introduce a novel method using the idea of herd immunity and the  $\mathfrak{R}_{\text{eff}}(t)$  threshold. Our system of ODEs implies the system of ODEs derived using the pair-approximation, justifying the use of mean-field approximations in this setting.

In our model, we consider a constant rate of infection  $\eta$  across all communities, however a natural extension would be to vary the rate of infection by community and consider a matrix  $\boldsymbol{\eta} = (\eta_{k\ell})_{k,\ell \in [K]}$  and vector  $\boldsymbol{\gamma} = (\gamma_k)_{k \in [K]}$  instead, where  $\eta_{k\ell}$  is the rate at which an infected individual in community  $k$  infects its neighbors in community  $\ell$ , and  $\gamma_k$  is the rate at which vertices in community  $k$  recover. In such a model where the transmission rates vary by community, we can also derive a set of differential equations based on the notion of active half-edges. Let  $X_{k\ell}$  be the number of active half-edges going out of community  $k$  with an endpoint (not yet realized) in community  $\ell$ . Then at any given time

- at rate  $\gamma_k X_{k\ell}$ , we lose one active half-edge contributing to  $X_{k\ell}$
- at rate  $\eta_{k\ell} X_{k\ell}$ , we explore one of the active half-edges contributing to  $X_{k\ell}$ .

When we explore an active half-edge contributing to  $X_{k\ell}$ ,  $X_{k\ell}$  decreases by 1. Now, let us consider how  $X_{k\ell}$  increases. Note that at rate  $\eta_{jk} X_{jk}$  a half-edge in community  $j$  with (future) endpoint in community  $k$  is explored. With probability  $S_k/n_k$  the exploration leads to an infection, and in this case  $X_{k\ell}$  increases by  $\text{Pois}(D_{k\ell})$ , where we recall that  $D_{k\ell} = W_{k\ell} n_\ell/n$  is the expected number of half-edges from community  $k$  into community  $\ell$ . Therefore,

- at an overall rate  $\frac{S_k}{n_k} \sum_\ell \eta_{\ell k} X_{\ell k}$ ,  $S_k$  goes down by 1 and  $I_k$  goes up by 1
- at an overall rate  $(\eta_{k\ell} + \gamma_k) X_{k\ell}$ ,  $X_{k\ell}$  goes down by 1, and
- at an overall rate  $\sum_j \eta_{jk} X_{jk} \frac{S_k}{n_k}$ ,  $X_{k\ell}$  goes up by  $\text{Pois}(D_{k\ell})$ .

Furthermore, at time  $t = 0$ ,  $\mathbb{E}[X_{k\ell}(0)] = D_{k\ell} I_k(0)$  Normalizing

$$\hat{s}_k(t) = \frac{S_k(t)}{n}, \quad \hat{i}_k(t) = \frac{I_k(t)}{n}, \quad \text{and} \quad \hat{x}_{k\ell}(t) = \frac{X_{k\ell}(t)}{n D_{k\ell}},$$

the expected rates of change for  $\hat{s}_k$ ,  $\hat{x}_{k\ell}$ , and  $\hat{i}_k$  are given by

$$-\hat{s}_k \sum_j \eta_{jk} W_{jk} \hat{x}_{jk}, \quad -(\eta_{k\ell} + \gamma_k) \hat{x}_{k\ell} + \hat{s}_k \sum_j \eta_{jk} W_{jk} \hat{x}_{jk}, \quad -\gamma_k \hat{i}_k + \hat{s}_k \sum_j \eta_{jk} W_{jk} \hat{x}_{jk},$$

respectively. The methods developed in Section 4 immediately imply a LLN for finite time horizons, leading to the differential equations

$$\frac{ds_k}{dt} = -s_k \sum_j \eta_{jk} W_{jk} x_{jk} \tag{68}$$

$$\frac{dx_{k\ell}}{dt} = -(\eta_{k\ell} + \gamma_k) x_{k\ell} + s_k \sum_j \eta_{jk} W_{jk} x_{jk} \tag{69}$$

$$\frac{di_k}{dt} = -\gamma_k i_k - \frac{ds_k}{dt} \tag{70}$$

with initial conditions  $x_{k\ell}(0) = i_k(0)$ . The herd immunity threshold is now the threshold for which a new infection with  $x_{k\ell}(t_0) = i_k(t_0)$  does not grow by more than a constant. As the reader may easily verify, this threshold is now governed by the spectral radius  $\mathfrak{R}_{\text{eff}}(t)$  of the matrix

$$C_{k\ell}(t) = \frac{\eta_{k\ell}}{\eta_{k\ell} + \gamma_k} W_{k\ell} s_\ell(t).$$

While the technical details will be more complicated, we believe that our techniques will again allow one to prove an LLN which is uniform in the time horizon, but we have not worked out the details.

We further believe that our methods should allow one to analyze a dynamic version of the stochastic block model where susceptible-infected edges are rewired to susceptible-susceptible edges as in [Ball and Britton \[2022\]](#). Finally, while we believe that the study of SIR dynamics on many random graph models will be amenable to the technology developed here, others, as e.g., an LLN for the time evolution and final size of the SIR epidemic on dynamically grown models such as preferential attachment might be more challenging.

## References

- E. Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.
- Y. Alimohammadi, C. Borgs, R. van der Hofstad, and A. Saberi. Epidemic forecasting on networks: Bridging local samples with global outcomes. Technical report, Technical report, Working paper, 2023.
- H. Andersson. Limit theorems for a random graph epidemic model. *Annals of Applied Probability*, 8(4):1331–1349, 1998.
- H. Andersson and T. Britton. *Stochastic Epidemic Models and Their Statistical Analysis*, volume 151 of *Lecture Notes in Statistics*. Springer, New York, NY, 1 edition, 2000.
- F. Ball and T. Britton. Epidemics on networks with preventive rewiring. *Random Structures & Algorithms*, 61(2):250–297, 2022.
- F. Ball and P. Neal. Network epidemic models with two levels of mixing. *Mathematical Biosciences*, 212(1):69–87, 2008.
- F. Ball, D. Sirl, and P. Trapman. Analysis of a stochastic sir epidemic on a random network incorporating household structure. *Mathematical Biosciences*, 224(2):53–73, 2010. doi: 10.1016/j.mbs.2009.12.001.
- A. Barbour and G. Reinert. Approximating the epidemic curve. 2013.
- A. D. Barbour. Couplings for locally branching epidemic processes. *Journal of Applied Probability*, 51A:43–56, 2014. ISSN 00219002. URL <http://www.jstor.org/stable/43284107>.
- M. S. Bartlett. Some evolutionary stochastic processes. *Journal of the Royal Statistical Society: Series B*, 11:211, 1949.
- S. Bhamidi, R. van der Hofstad, and J. Komjáthy. The front of the epidemic spread and first passage percolation. *Journal of Applied Probability*, 51(A):101–121, 2014. doi: 10.1239/jap/1417528470.
- R. Bhatia, L. Elsner, and G. Krause. Bounds for the variation of the roots of a polynomial and the eigenvalues of a matrix. *Linear algebra and its applications*, 142:195–209, 1990.
- T. Bohman and M. Piccollelli. Sir epidemics on random graphs with a fixed degree sequence. *Random Structures & Algorithms*, 41(2):179–214, 2012. doi: 10.1002/rsa.20386.
- B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, June 2007. ISSN 1098-2418. doi: 10.1002/rsa.20168. URL <http://dx.doi.org/10.1002/rsa.20168>.
- T. Britton, M. Deijfen, A. N. Lageras, and M. Lindholm. Epidemics on random graphs with tunable clustering. *Journal of Applied Probability*, 45(3):743–756, 2008.
- T. A. Bui, S Chaudhuri, F. T. Leighton, and M. Sipser. Graph bisection algorithms with good average case behavior. *Combinatorica*, 7(2):171–191, 1987. ISSN 0209-9683. doi: 10.1007/BF02579448. URL <http://dx.doi.org/10.1007/BF02579448>.

- J. Cocomello and K. Ramanan. Exact description of limiting sir and seir dynamics on locally tree-like graphs. *arXiv preprint arXiv:2309.08829*, 2023.
- D. J. Daley and J. Gani. *Epidemic Modelling: An Introduction*. Cambridge University Press, Cambridge, 1999.
- R. W. R. Darling and J. R. Norris. Differential equation approximations for markov chains. *Probability Surveys*, 5:37–79, 2008.
- L. Decreusefond, J. S. Dhersin, P. Moyal, and V. C. Tran. Large graph limit for an sir process in random network with heterogeneous connectivity. *Annals of Applied Probability*, 22(2):541–575, 2012. doi: 10.1214/10-AAP745.
- O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Wiley Series in Mathematical & Computational Biology. Wiley, 2000. ISBN 9780471492412. URL <https://books.google.com/books?id=5VjSaAf35pMC>.
- S. N. Ethier and T. G. Kurtz. Markov processes. Wiley, 1986.
- M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002. doi: 10.1073/pnas.122653799. URL <https://doi.org/10.1073/pnas.122653799>.
- T. E. Harris et al. *The theory of branching processes*, volume 6. Springer Berlin, 1963.
- P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983. ISSN 0378-8733. doi: 10.1016/0378-8733(83)90021-7.
- T. House. Modelling epidemics on networks. *Contemp. Phys.*, 53(3):213–225, 2012.
- J. A. Jacquez, C. P. Simon, J. Koopman, L. Sattenspiel, and T. Perry. Modeling and analyzing hiv transmission: the effect of contact patterns. *Mathematical Biosciences*, 92:119–199, 1988. doi: 10.1016/0025-5564(88)90031-4. URL [https://doi.org/10.1016/0025-5564\(88\)90031-4](https://doi.org/10.1016/0025-5564(88)90031-4).
- S. Janson, M. Luczak, and P. Windridge. Law of large numbers for the sir epidemic on a random graph with given degrees. *Random Structures & Algorithms*, 45(4):726–763, 2014.
- W. O. Kermack and A. G. McKendrick. 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*, 115:700–721, 1927. doi: 10.1098/rspa.1927.0118. URL <http://doi.org/10.1098/rspa.1927.0118>.
- I. Z. Kiss, J. C. Miller, and P. L. Simon. *Mathematics of Epidemics on Networks*, volume 598 of *Springer Series in Statistics*. Springer, Cham, 2017.
- I.Z. Kiss, C.G. Morris, F. Sélley, P.L. Simon, and R.R. Wilkinson. Exact deterministic representation of markovian sir epidemics on networks with and without loops. *Journal of Mathematical Biology*, 70(3):437–464, 2015. doi: 10.1007/s00285-014-0772-0. URL <https://doi.org/10.1007/s00285-014-0772-0>. Epub 2014 Mar 4.

- I.Z. Kiss, E. Kenah, and G.A. Rempala. Necessary and sufficient conditions for exact closures of epidemic equations on configuration model networks, 2022. URL <https://arxiv.org/abs/2208.07983>.
- C. Lee and D. J. Wilkinson. A review of stochastic block models and extensions for graph clustering. *Appl Netw Sci*, 4:122, 2019. doi: 10.1007/s41109-019-0232-2. URL <https://doi.org/10.1007/s41109-019-0232-2>.
- J.C. Miller and T. Ting. Eon (epidemics on networks): a fast, flexible python package for simulation, analytic approximation, and analysis of epidemics on networks. *Journal of Open Source Software*, 4(44):1731, 2019. doi: 10.21105/joss.01731. URL <https://doi.org/10.21105/joss.01731>.
- P. Neal. Sir epidemics on a bernoulli random graph. *J. Appl. Probab.*, 40(3):779–782, 2003.
- M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):016128, 2002.
- K.J. Sharkey, I.Z. Kiss, R.R. Wilkinson, and P.L. Simon. Exact equations for sir epidemics on tree graphs. *Bulletin of Mathematical Biology*, 77(4):614–645, 2015. doi: 10.1007/s11538-013-9923-5. URL <https://doi.org/10.1007/s11538-013-9923-5>.
- E. Volz. Sir dynamics in random networks with heterogeneous connectivity. *Journal of mathematical biology*, 56:293–310, 2008.
- B. Von Bahr and A. Martin-Löf. Threshold limit theorems for some epidemic processes. *Advances in Applied Probability*, 12(2):319–349, 1980. URL <https://www.jstor.org/stable/1426600>.
- E. Vynnycky and R.G. White. *An Introduction to Infectious Disease Modelling*. Oxford University Press, 2010.

# Appendix

## A Branching Process Results

In this appendix, we derive several results about discrete time, multi-type branching processes used in the main body of this paper.

### A.1 Preliminaries

We start by establishing a few facts for multi-type branching processes which are usually formulated in the more restricted setting of “positively regular” branching processes, see, e.g., [Harris et al. \[1963\]](#), but hold more generally, often with proofs which are straightforward extensions of those in the literature.

We assume a finite type space  $[K]$  and the associated state-space  $\mathbb{N}_0^K$ , with the state corresponding to a single individual of type  $k$  represented by the unit vector  $\mathbf{e}_k \in \mathbb{N}_0^K$  in direction  $k$ . We use  $\mathbf{Z}_m$  to denote the state of the branching process in generation  $m$  and  $\mathbb{P}_k$  to denote the probability distribution of  $(\mathbf{Z}_m)_{m \geq 0}$  when starting from  $\mathbf{Z}_0 = \mathbf{e}_k$ ,

$$\mathbb{P}_k(\cdot) = \mathbb{P}(\cdot \mid \mathbf{Z}_0 = \mathbf{e}_k).$$

We define  $G_k(\cdot)$  to be the generating function of the offspring distribution  $p_k(\mathbf{z}) = \mathbb{P}_k(\mathbf{Z}_1 = \mathbf{z})$ , and more generally define the generating function of  $\mathbf{Z}_m$  by

$$G_k^{(m)}(\mathbf{s}) = \sum_{\mathbf{z}} \mathbb{P}_k(\mathbf{Z}_m = \mathbf{z}) \prod_{i=1}^K s_i^{z_i}.$$

Setting  $\mathbf{G}(\mathbf{s}) = (G_1(\mathbf{s}), \dots, G_K(\mathbf{s}))$  and  $\mathbf{G}^{(m)}(\mathbf{s}) = (G_1^{(m)}(\mathbf{s}), \dots, G_K^{(m)}(\mathbf{s}))$ , it is easy to see that  $G_k^{(m)}(\mathbf{s}) = G_k(\mathbf{G}^{(m-1)}(\mathbf{s}))$ , showing that  $\mathbf{G}^{(n)}(\cdot)$  is the  $n^{\text{th}}$  iterate of  $\mathbf{G}(\cdot) = \mathbf{G}^{(1)}(\cdot)$ , the generating function for the offspring distribution. Finally, we define the vector of extinction probabilities via

$$q_k = \mathbb{P}_k(\mathbf{Z}_m = \mathbf{0} \text{ for some } m < \infty).$$

**Lemma A.1.** *Let  $(\mathbf{Z}_m)_{m \geq 0}$  be a branching process such that all states in  $\mathbb{N}_0^K \setminus \{\mathbf{0}\}$  are transient. If  $0 \leq s_k < 1$  for all  $k$ , then  $\mathbf{G}^{(m)}(\mathbf{s}) \rightarrow \mathbf{q}$ , where  $\mathbf{q} = (q_1, \dots, q_k)$  is the vector of extinction probabilities.*

*Proof.* Expressing  $q_k$  as the limit of  $\mathbb{P}_k(\mathbf{Z}_m = \mathbf{0}) = G_k^{(m)}(\mathbf{0})$ , we need to prove that  $G_k^{(m)}(\mathbf{s}) - G_k^{(m)}(\mathbf{0}) \rightarrow 0$ . To this end, we decompose the sum representing this difference into a finite sum plus a tail,

$$G_k^{(m)}(\mathbf{s}) - G_k^{(m)}(\mathbf{0}) = \sum_{\mathbf{z} \neq \mathbf{0}} \mathbb{P}_k(\mathbf{Z}_m = \mathbf{z}) s_1^{z_1} \dots s_k^{z_k} \leq \sum_{\mathbf{z} \neq \mathbf{0}: \|\mathbf{z}\|_1 \leq R} \mathbb{P}_k(\mathbf{Z}_m = \mathbf{z}) + O(\max_k s_k^R). \quad (71)$$

Since all  $\mathbf{z} \neq \mathbf{0}$  are transient, the terms in the first sum go to 0 as  $m \rightarrow \infty$  for any fixed  $R$ . Choosing first  $R$  large enough, and then  $m$  large enough, the right hand side can be made as small as desired, proving the lemma.  $\square$

To state the next lemma, we define a branching process to be irreducible if for all  $i, j \in [K]$ , there exists an  $m < \infty$  such that  $\mathbb{P}(\mathbf{Z}_m \cdot \mathbf{e}_j > 0 \mid \mathbf{Z}_0 = \mathbf{e}_i) > 0$ .

**Lemma A.2.** *Let  $(\mathbf{Z}_m)_{m \geq 0}$  be an irreducible branching process such all states in  $\mathbb{N}_0^K \setminus \{\mathbf{0}\}$  are transient. If there exists a solution  $\mathbf{s} \neq \mathbf{1}$  to  $\mathbf{G}(\mathbf{s}) = \mathbf{s}$  in  $[0, 1]^K$ , then it is equal to  $\mathbf{q}$ , and  $q_k < 1$  for all  $k$ .*

*Proof.* In view of the previous lemma, all we need to show is that if  $\mathbf{s} = \mathbf{G}(\mathbf{s})$  and  $\mathbf{s} \neq \mathbf{1}$ , then  $s_k < 1$  for all  $k$ . Assume w.l.o.g. that  $s_1 < 1$ , and fix  $k$ . By irreducibility, there exists an  $m$  such that  $\mathbb{P}_k(\mathbf{Z}_m \cdot \mathbf{e}_1 > 0) > 0$ , implying that the generating function  $G_k^{(m)}(\mathbf{s})$  depends explicitly on  $s_1$  and

$$\begin{aligned} G_k^{(m)}(\mathbf{s}) &= \sum_{\mathbf{z}} \mathbb{P}_k(\mathbf{Z}_m = \mathbf{z}) s_1^{z_1} \dots s_K^{z_K} \leq \sum_{\mathbf{z}, z_1=0} \mathbb{P}_k(\mathbf{Z}_m = \mathbf{z}) + \sum_{\mathbf{z}, z_1>0} \mathbb{P}_k(\mathbf{Z}_m = \mathbf{z}) s_1^{z_1} \\ &< \sum_{\mathbf{z}, z_1=0} \mathbb{P}_k(\mathbf{Z}_m = \mathbf{z}) + \sum_{\mathbf{z}, z_1>0} \mathbb{P}_k(\mathbf{Z}_m = \mathbf{z}) = 1. \end{aligned}$$

But, since  $\mathbf{s}$  is a solution of  $\mathbf{s} = \mathbf{G}(\mathbf{s})$ , it is also a solution of  $\mathbf{s} = \mathbf{G}^{(m)}(\mathbf{s})$  implying  $s_k < 1$  and hence the claim.  $\square$

To use these lemmas, we will need to establish that all states  $\mathbf{z} \neq \mathbf{0}$  are transient. While not the most general lemma one can prove, the following will be sufficient for us.

**Lemma A.3.** *Let  $\mathbb{S} \subset \mathbb{N}_0^K$  be a set of states  $\mathbf{z}$  such that  $\mathbb{P}(\mathbf{Z}_1 = \mathbf{z} | \mathbf{Z}_0 = \mathbf{z}) = 1$ , and assume that for all  $\mathbf{z} \notin \mathbb{S}$  there exists a state  $\mathbf{z}' \in \mathbb{S}$  such that  $\mathbb{P}(\mathbf{Z}_1 = \mathbf{z}' | \mathbf{Z}_0 = \mathbf{z}) > 0$ . Then all states  $\mathbf{z} \notin \mathbb{S}$  are transient, so in particular  $\lim_{m \rightarrow \infty} \mathbb{P}_k(\mathbf{Z}_m = \mathbf{z}) = 0$  for all  $\mathbf{z} \notin \mathbb{S}$  and all  $k \in [K]$ .*

*Proof.* Fix  $\mathbf{z} \notin \mathbb{S}$  and  $\mathbf{z}' \in \mathbb{S}$  such that  $\mathbb{P}(\mathbf{Z}_1 = \mathbf{z}' | \mathbf{Z}_0 = \mathbf{z}) = \varepsilon > 0$ , and let  $T_{\mathbf{z}}$  be the first  $m$  such that  $\mathbf{Z}_m$  revisits  $\mathbf{z}$  when starting at  $\mathbf{z}$ , i.e., let  $T_{\mathbf{z}} = \min\{m > 0 : \mathbf{Z}_m = \mathbf{z}\}$  conditioned on  $\mathbf{Z}_0 = \mathbf{z}$ , where the minimum is defined to be  $\infty$  if there exists no such  $m$ . Then  $\mathbb{P}(T_{\mathbf{z}} < \infty | \mathbf{Z}_0 = \mathbf{z})$  is the probability that the chain revisits  $\mathbf{z}$  at least once, and

$$\mathbb{P}(T_{\mathbf{z}} = \infty | \mathbf{Z}_0 = \mathbf{z}) \geq \mathbb{P}(\mathbf{Z}_1 = \mathbf{z}' | \mathbf{Z}_0 = \mathbf{z}) = \varepsilon > 0.$$

On the other hand, by the strong Markov property of  $(\mathbf{Z}_m)_{m \geq 0}$ , the probability distribution of  $\mathbf{Z}_m$  after the stopping time  $T_{\mathbf{z}}$  is independent of the history up to time  $T_{\mathbf{z}}$  and identical to the original distribution (after the obvious time shift by  $T_{\mathbf{z}}$ ). As a consequence, the probability that the chain revisits  $\mathbf{z}$  at least  $k$  times is equal to

$$\mathbb{P}(T_{\mathbf{z}} < \infty | \mathbf{Z}_0 = \mathbf{z})^k \leq (1 - \varepsilon)^k \rightarrow 0 \quad \text{as } k \rightarrow \infty,$$

showing that  $\mathbf{z}$  is transient.  $\square$

**Remark A.4.** *The forward and backward branching processes of the SIR model introduced in the main part of this paper clearly obey the assumption of Lemma A.3, since any state has non-zero probability of going extinct in one step by the properties of the Poisson distribution. Furthermore, both are irreducible if  $W$  is irreducible, allowing us to use both Lemma A.1 and Lemma A.2 in this case.*

**Lemma A.5.** *Let  $\mathbf{Z} = \sum_m \mathbf{Z}_m$  be the total size of the branching process  $(\mathbf{Z}_m)_{m \geq 0}$  with offspring mean matrix  $C_{k\ell} = \mathbb{E}[\mathbf{Z}_1 \cdot \mathbf{e}_\ell | \mathbf{Z}_0 = \mathbf{e}_k]$ . If  $C$  has spectral radius  $\rho(C) < 1$  and all components of  $\mathbf{Z}_1$  have exponential tails, then all components of  $\mathbf{Z}$  have exponential tails as well.*

*Proof.* Let  $\tilde{\rho}$  be such that  $\rho(C) < \tilde{\rho} < 1$ , and let  $\mathbf{v}$  be a left eigenvector for the eigenvalue  $\rho(C)$ , chosen to have all positive entries (which is possible by Perron-Fröbenius). To prove the lemma, we will first prove the existence of a constant  $\alpha_0$  such that

$$\mathbb{E} \left[ e^{\alpha \mathbf{v} \cdot \mathbf{Z}_m} \mid \mathbf{Z}_{m-1} \right] \leq e^{\alpha \tilde{\rho} \mathbf{v} \cdot \mathbf{Z}_{m-1}} \quad \text{for all } \alpha \leq \alpha_0, \quad (72)$$

and then use this bound to show that

$$\mathbb{E} \left[ e^{\alpha \mathbf{v} \cdot \mathbf{Z}} \mid \mathbf{Z}_0 = \mathbf{e}_k \right] \leq e^{\frac{\alpha}{1-\tilde{\rho}} \mathbf{v} \cdot \mathbf{e}_k} \quad \text{for all } \alpha \leq \alpha_0(1-\tilde{\rho}), \quad (73)$$

establishing the existence of the moment-generating function of  $\mathbf{Z}$  in some neighborhood around 0, and hence proving the claim.

Let  $\mathbf{N}_k$  be a random variable such that  $\mathbb{P}(\mathbf{N}_k = \mathbf{z}) = p_k(\mathbf{z})$ . By the assumption that the offspring distribution has exponential tails, there exists some constant  $\zeta < \infty$  such that for all  $k$  and all sufficiently small  $\alpha$

$$\mathbb{E} \left[ e^{\alpha \mathbf{v} \cdot \mathbf{N}_k} (\mathbf{v} \cdot \mathbf{N}_k)^2 \right] \leq \zeta. \quad (74)$$

Combined with the bound  $e^x \leq 1 + x + \frac{x^2}{2} e^x$ , valid for all  $x \geq 0$ , and the fact that  $\mathbb{E}[\mathbf{v} \cdot \mathbf{N}_k] = \mathbf{v} C \mathbf{e}_k = \rho(C) \mathbf{v} \cdot \mathbf{e}_k$ , this implies that

$$\mathbb{E} \left[ e^{\alpha \mathbf{v} \cdot \mathbf{N}_k} \right] \leq 1 + \alpha \mathbb{E}[\mathbf{v} \cdot \mathbf{N}_k] + \frac{\alpha^2}{2} \mathbb{E} \left[ (\mathbf{v} \cdot \mathbf{N}_k)^2 e^{\alpha \mathbf{v} \cdot \mathbf{N}_k} \right] \leq 1 + \alpha \rho(C) \mathbf{v} \cdot \mathbf{e}_k + \frac{\alpha^2}{2} \zeta. \quad (75)$$

For  $\alpha$  sufficiently small, we thus get that

$$\mathbb{E} \left[ e^{\alpha \mathbf{v} \cdot \mathbf{N}_k} \right] \leq 1 + \alpha \left( \rho(C) \mathbf{v} \cdot \mathbf{e}_k + \frac{\alpha \zeta}{2} \right) \leq 1 + \alpha \tilde{\rho} \mathbf{v} \cdot \mathbf{e}_k \leq e^{\alpha \tilde{\rho} \mathbf{v} \cdot \mathbf{e}_k}. \quad (76)$$

Furthermore, letting  $N_{k,i} \sim N_k$  i.i.d., we have

$$\mathbb{E} \left[ e^{\alpha \mathbf{v} \cdot \mathbf{Z}_m} \mid \mathbf{Z}_{m-1} \right] = \prod_{\ell \in [K]} \prod_{i=1}^{(\mathbf{Z}_{m-1})_\ell} \mathbb{E} \left[ e^{\alpha \mathbf{v} \cdot N_{\ell,i}} \right] \leq \prod_{\ell \in [K]} \prod_{i=1}^{(\mathbf{Z}_{m-1})_\ell} e^{\alpha \tilde{\rho} \mathbf{v} \cdot \mathbf{e}_\ell} = e^{\alpha \tilde{\rho} \mathbf{v} \cdot \mathbf{Z}_{m-1}}.$$

This gives the bound Eq. (72).

As a consequence of Eq. (72) applied twice in a row, we we get that for  $\alpha(1+\tilde{\rho}) \leq \alpha_0$ ,

$$\mathbb{E} \left[ e^{\alpha \mathbf{v} \cdot (\mathbf{Z}_m + \mathbf{Z}_{m-1})} \mid \mathbf{Z}_{m-2} \right] = \mathbb{E} \left[ \mathbb{E} \left[ e^{\alpha \mathbf{v} \cdot \mathbf{Z}_m} \mid \mathbf{Z}_{m-1} \right] e^{\alpha \mathbf{v} \cdot \mathbf{Z}_{m-1}} \mid \mathbf{Z}_{m-2} \right] \leq e^{\alpha(\tilde{\rho} + \tilde{\rho}^2) \mathbf{v} \cdot \mathbf{Z}_{m-2}}.$$

Continuing by induction, this proves that for  $\alpha \sum_{i=0}^{m-1} \tilde{\rho}^i \leq \alpha_0$ ,

$$\mathbb{E} \left[ e^{\alpha \sum_{i=0}^m \mathbf{v} \cdot \mathbf{Z}_i} \mid \mathbf{Z}_0 = \mathbf{e}_k \right] \leq e^{\alpha \mathbf{v} \cdot \mathbf{e}_k \sum_{i=0}^m \tilde{\rho}^i}.$$

The bound Eq. (73) follows by by monotone convergence.  $\square$



## A.2 Backward Branching Process

In this section, we formally define the backwards branching process and use it to prove Lemma 6.9. The process has label space  $[K + 1]$ , with  $k = K + 1$  representing an infected state and labels  $k \in [K]$  representing susceptible vertices of type  $k$ , and state space  $\mathbb{N}_0^{K+1}$ . We will use the notation  $\bar{\mathbf{z}} = (\mathbf{z}, z_{K+1}) = (z_1, \dots, z_{K+1})$  for vectors in  $\mathbb{N}_0^{K+1}$ , and similarly for vectors  $\bar{\mathbf{s}} \in [0, 1]^{K+1}$ . The offspring distribution of the backward process is then defined as follows.

Let  $a_k \geq 0$  and let  $p_k(\cdot)$  be the Poisson offspring distribution

$$p_k(\mathbf{z}) = \prod_{\ell=1}^K \mathbb{P}(\text{Pois}(C_{k\ell}) = z_\ell).$$

For  $k \in [K]$ , we set

$$\bar{p}_k(\bar{\mathbf{z}}) = \begin{cases} e^{-a_k} p_k(\mathbf{z}) & \text{if } z_{K+1} = 0 \\ 1 - e^{-a_k} & \text{if } z_{K+1} = 1, \end{cases}$$

and for  $k = K + 1$  we set

$$\bar{p}_{K+1}(\bar{\mathbf{z}}) = 1 \text{ if } \bar{\mathbf{z}} = (0, \dots, 0, 1) \text{ and } 0 \text{ otherwise.}$$

We will denote the generating function of this offspring distribution by  $\bar{\mathbf{F}}$ , the state of the process in generation  $m$  by  $\bar{\mathbf{Z}}_m$ , and its generating function by  $\bar{\mathbf{F}}^{(m)}$ . It is easy to express  $\bar{\mathbf{F}}$  in terms of the generating function  $\mathbf{G}$  of the offspring distribution for the Poisson branching process, giving

$$\bar{F}_k(\bar{\mathbf{s}}) = \begin{cases} e^{-a_k} G_k(\mathbf{s}) + (1 - e^{-a_k}) s_{K+1} & \text{if } k \in [K] \\ s_{K+1} & \text{if } k = K + 1. \end{cases}$$

If  $s_{K+1} = 0$ , this expression, as well as the expression for the iterates takes a particular simple form, giving in particular

$$\bar{F}_k^{(m)}((\mathbf{s}, 0)) = \begin{cases} F_k^{(m)}(\mathbf{s}) & \text{if } k \in [K] \\ 0 & \text{if } k = K + 1, \end{cases}$$

where  $F^{(m)}$  is the  $m^{\text{th}}$  iterate of

$$F_k(\mathbf{s}) = e^{-a_k} G_k(\mathbf{s}).$$

Note that the function  $F_k$  already appeared in the implicit equation Eq. (23) for the final proportion of susceptible vertices, which we can now write in the form

$$s_k = F_k(\mathbf{s}). \tag{77}$$

It is clear that the survival probability of the backward branching process starting from the state  $\bar{\mathbf{Z}}_0 = \mathbf{e}_{K+1}$  is one. The following lemma gives an expression for the extinction (and equivalently survival) probability starting from  $e_k$  for  $k \in [K]$ .

**Lemma A.6.** *If  $0 \leq s_k < 1$  for all  $k \in [K]$ , then  $\mathbf{F}^{(m)}(\mathbf{s}) \rightarrow \mathbf{q}(\mathbf{a})$ , where  $\mathbf{q}(\mathbf{a}) = (q_1(\mathbf{a}), \dots, q_k(\mathbf{a}))$  is the vector of extinction probabilities, i.e.,  $q_k(\mathbf{a}) = \mathbb{P}(\bar{\mathbf{Z}}_m = 0 \text{ for some } m < \infty \mid \bar{\mathbf{Z}}_0 = \mathbf{e}_k)$ .*

*Proof.* First, we note that by the same arguments as those used in the proof of Lemma A.3, all states  $\bar{z} = (\mathbf{z}, z_{K+1})$  with  $\bar{z} \neq 0$  are transient. To see this, set  $T_{\bar{z}}$  to be the first time the chain revisits the state  $\bar{z}$ , and note that

$$\mathbb{P}(T_{\bar{z}} = \infty) \geq \mathbb{P}(\bar{\mathbf{Z}}_1 = (0, \dots, 0, z_{K+1}) \mid \mathbf{Z}_0 = \bar{z}) > 0$$

since the probability that Poisson offspring distribution has 0 children is bounded away from 0.

Using the transience of the states  $\bar{z} = (\mathbf{z}, z_{K+1})$  with  $\mathbf{z} \neq 0$ , we may then proceed as in the proof of Lemma A.1 to conclude that

$$\begin{aligned} F_k^{(m)}(\mathbf{s}) - \mathbb{P}_k(\bar{\mathbf{Z}}_m = 0) &= \bar{F}_k^{(m)}(\mathbf{s}, 0) - \bar{F}_k^{(m)}(\mathbf{0}, 0) \\ &= \sum_{\mathbf{z} \neq \mathbf{0}} \mathbb{P}_k(\bar{\mathbf{Z}}_m = (\mathbf{z}, 0)) s_1^{z_1} \dots s_k^{z_k} \leq \sum_{\mathbf{z} \neq \mathbf{0}: \|\mathbf{z}\|_1 \leq R} \mathbb{P}_k(\bar{\mathbf{Z}}_m = \mathbf{z}) + O(\max_k s_k^R). \end{aligned}$$

As before, the right hand side can be made arbitrary small by choosing first  $R$  and then  $m$  large enough, showing that the left hand side goes to 0 as  $m \rightarrow \infty$ .  $\square$

**Lemma A.7** (Uniqueness of Solution to Implicit Equation). *Assume that  $W$  is irreducible, and that  $\sum_\ell a_\ell > 0$ . Then the implicit equations Eq. (77) are uniquely solved by  $\mathbf{s} = \mathbf{q}(\mathbf{a})$ , where  $\mathbf{q}(\mathbf{a})$  is the vector of survival probabilities for the backward branching process, and  $q_k(\mathbf{a}) < 1$  for all  $k$ .*

*Proof.* In view of the last lemma, all we need to prove is that any solution  $\mathbf{s}$  of Eq. (77) is smaller than 1 component-wise. To this end, we first note that by the fact that  $G_k(\mathbf{s}) \leq 1$  and our assumption on  $\mathbf{a}$ , there exists at least one component  $s_\ell < 1$ . Next we use that  $\mathbf{s} = \mathbf{F}^{(m)}(\mathbf{s})$  and  $\mathbf{F} \leq \mathbf{G}$  component wise imply that

$$s_k \leq G_k^{(m)}(\mathbf{s}). \quad (78)$$

But  $s_\ell < 1$  and irreducibility of the Poisson branching process imply that there exists an  $m$  such that  $G_k^{(m)}(\mathbf{s}) < 1$ , see the proof of Lemma A.2. Since  $k \in [K]$  was arbitrary, this completes the proof.  $\square$

**Remark A.8.** *It follows immediately from the above proof that  $\mathbf{q}(\mathbf{a}) \leq \mathbf{q}$  component-wise, where  $\mathbf{q}$  is the smallest solution of the equation  $\mathbf{q} = \mathbf{G}(\mathbf{q})$  (which by Lemma A.2 is also the vector of survival probabilities of the Poisson branching process with mean matrix  $C$ ). Indeed, by Eq. (78),  $\mathbf{q}(\mathbf{a}) \leq \mathbf{G}^{(m)}(\mathbf{q}(\mathbf{a}))$ , and by Lemma A.1, the right hand side converges to  $\mathbf{q}$ .*

With these preparations, we are now ready to prove Lemma 6.9.

*Proof of Lemma 6.9.* Assume that  $(\tilde{\mathbf{s}}(0), \tilde{\mathbf{x}}(0)) \rightarrow (\mathbf{s}(0), \mathbf{x}(0))$  along some sequence. Let  $(\tilde{C}, \tilde{\mathbf{a}})$  be the corresponding values for the matrix  $C$  and the vector  $\mathbf{a}$ , and let  $\mathbf{q}(\tilde{C}, \tilde{\mathbf{a}})$  be the corresponding sequence of survival probabilities. We want to show that  $\mathbf{q}(\tilde{C}, \tilde{\mathbf{a}}) \rightarrow \mathbf{q}$ , where  $\mathbf{q}$  is the vector of survival probabilities of the Poisson branching process with mean matrix  $C$ . Consider some subsequence  $(C^j, \mathbf{a}^j)_{j \in \mathbb{N}}$  with  $(C^j, \mathbf{a}^j) \rightarrow (C, \mathbf{0})$  such that the limit of this subsequence  $\tilde{\mathbf{q}} := \lim_{j \rightarrow \infty} \mathbf{q}(C^j, \mathbf{a}^j)$  exists, and let  $F_k^j(\cdot)$  be the corresponding sequence of generating functions. Then

$$\tilde{q}_k = \lim_{j \rightarrow \infty} q_k(C^j, \mathbf{a}^j) = \lim_{j \rightarrow \infty} F_k^j(\mathbf{q}(C^j, \mathbf{a}^j)) = G_k(\tilde{\mathbf{q}}),$$

where the last equality is true by the fact that  $F_k^j(\mathbf{s})$  is jointly continuous in  $\mathbf{s}$  and the parameters  $C^j, \mathbf{a}^j$ . Thus  $\tilde{\mathbf{q}}$  satisfies  $\tilde{\mathbf{q}} = \mathbf{G}(\tilde{\mathbf{q}})$ . To complete the proof, we consider two cases. If the spectral

radius  $\rho(C)$  of  $C$  is at most one, we know that the survival probability of the Poisson branching process is 0 and the implicit equation has only one solution,  $\tilde{\mathbf{q}} = \mathbf{1}$ , proving the desired convergence. If  $\rho(C) > 1$ , we can find some  $\varepsilon > 0$  such that  $\rho(C(1 - \varepsilon)) > 1$ , and by the convergence of  $C^j$ , some  $j_0$  such that for  $j \geq j_0$ ,  $C^j \geq (1 - \varepsilon)C$ . Since the survival probabilities are monotone in the entries of the mean matrix, and  $q_k(C^j, \mathbf{a}^j) \leq q_k(C^j, 0)$  by Remark A.8, we conclude that  $\tilde{\mathbf{q}} \leq \mathbf{q}_\varepsilon$  component-wise, where  $\mathbf{q}_\varepsilon$  is survival probability of the Poisson branching process with mean matrix  $(1 - \varepsilon)C$ . Since the latter has a strictly positive survival probability, we conclude that  $\tilde{\mathbf{q}} < \mathbf{1}$  component wise. Therefore,  $\tilde{\mathbf{q}} = \mathbf{q}$  by uniqueness of the non-trivial solution to  $\mathbf{q} = \mathbf{G}(\mathbf{q})$  for probability generating functions. This gives the desired continuity result.  $\square$

## B Analysis of the Differential Equations for the SIR Epidemic on the SBM

We start this appendix with the proof of Lemma 5.2.

*Proof of Lemma 5.2.* We will write the differential equation  $d\mathbf{y}_t/dt = \mathbf{b}(\mathbf{y}_t)$  in the explicit form Eq. (3), Eq. (4), and Eq. (5) and note that  $\mathbf{y}_t \in U_0$  for all  $t < \infty$  if  $\mathbf{y}_0 \in U_0$ , see Remark 4.1.

By Eq. (3),  $ds_k(t)/dt \leq 0$  for all  $t$ . Since  $s_k(t)$  is non-negative, statement (a) follows by monotone convergence. Next we combined Eq. (3) and Eq. (4) to conclude that

$$\frac{dx_k(t)}{dt} + \frac{ds_k(t)}{dt} = -(\eta + \gamma)x_k(t) \leq 0 \quad (79)$$

and Eq. (3) and Eq. (5) to conclude that

$$\frac{di_k(t)}{dt} + \frac{ds_k(t)}{dt} = -\gamma i_k(t) \leq 0. \quad (80)$$

Monotone convergence then implies the existence of the limits  $\lim_{t \rightarrow \infty} (x_k + s_k)(t)$  and  $\lim_{t \rightarrow \infty} (i_k + s_k)(t)$ . Combined with the already established existence of the limit for  $s_k$ , this gives the existence of the limits in (b) and (c).

To prove that  $x_k(\infty) = 0$ , suppose towards contradiction that  $x_k(\infty) > 0$ . Then there exist some constants  $t_0 < \infty$  and  $\zeta > 0$  such that  $x_k(t) \geq \zeta$  for all  $t \geq t_0$ . Inserted into Eq. (79), this implies that

$$x_k(t) + s_k(t) \leq x_k(t_0) + s_k(t_0) - (\eta + \gamma)\zeta(t - t_0)$$

for all  $t \geq t_0$ , which is a contradiction since  $\mathbf{y}_t \in U_0$  so in particular  $x_k(t) + s_k(t) \geq 0$  for all  $t < \infty$ . This proves that  $x_k(\infty) = 0$ . The proof that  $i_k(\infty) = 0$  is identical.

Next we prove (d). We integrate Eq. (3) to express  $s_k(t)$  as

$$s_k(t) = s_k(0) \exp \left\{ -\eta \sum_{\ell} \chi_{\ell}(t) W_{\ell k} \right\}, \quad (81)$$

where

$$\chi_{\ell}(t) = \int_0^t x_{\ell}(\tau) d\tau = \frac{x_{\ell}(0) + s_{\ell}(0) - s_{\ell}(t) - x_{\ell}(t)}{\eta + \gamma}. \quad (82)$$

Here the identity in Eq. (82) defines  $\chi_{\ell}(t)$  and the second follows by integrating the differential equation Eq. (79) from 0 to  $t$ . This shows that  $\chi_{\ell}(t)$  is bounded uniformly in  $t$  implying claim (d).

To prove (e), we first note that  $\frac{d}{dt}x_k(t) \geq -(\eta + \gamma)x_k(t)$ , showing that  $x_k(t) > 0$  for all finite  $t \geq t_0$  if  $x_k(t_0) > 0$ , and similarly for  $i_k(t)$ . Consider the set  $\mathcal{K}(t)$  of indices  $k$  such that  $x_k(t) > 0$  at time  $t \geq 0$ . We want to prove that  $\mathcal{K}(t) = [K]$  for all  $t > 0$ . To this end, we will prove that if  $\mathcal{K}(t) \neq [K]$ , then we can find an  $\ell \notin \mathcal{K}(t)$  such that  $\ell \in \mathcal{K}(t')$  for all  $t' > t$ . To see this, we use that by irreducibility of  $W$ , we can find  $\ell \notin \mathcal{K}(t)$  and  $k \in \mathcal{K}(t)$  such that  $W_{k\ell} > 0$ , which shows that

$$\frac{dx_\ell}{dt} = -(\eta + \gamma)x_\ell(t) + \sum_{k'} x_{k'}(t)W_{k'\ell}s_\ell(t) > x_k(t)W_{k\ell}s_\ell(t) > 0.$$

This shows that  $\ell \in \mathcal{K}(t')$  for  $t' > t$  close enough to  $t$ , and by our previous observation that  $x_\ell(t'') > 0$  for all  $t'' \geq t'$  if  $x_\ell(t') > 0$ , we get  $\ell \in \mathcal{K}(t')$  for all  $t' > t$ . Starting from the observation that  $\mathcal{K}(0) \neq \emptyset$ , this implies by induction that  $\mathcal{K}(t) = [K]$  for all  $t > 0$ , i.e.,  $x_k(t) > 0$  for all  $k \in [K]$  and all  $t > 0$ . With the help of Eq. (5) (and the fact that  $i_k(t) > 0$  implies  $i_k(t') > 0$  for all  $t' > t$ ) this in turn implies that  $i_k(t) > 0$  for all  $k \in [K]$  and all  $t > 0$ . Finally,  $x_k(t) > 0$  for all  $0 < t < \infty$  implies that  $ds_k/dt < 0$  with the help of Eq. (3).  $\square$

Next we state and prove a lemma alluded to in the discussion section of this paper, showing that eventually, all components of  $\mathbf{x}(t)$  decay exponentially fast.

**Lemma B.1.** *Assume that  $W$  is irreducible and that  $s_k(0) > 0$  for all  $k \in [K]$ , let  $\mathfrak{R}_\infty$  be the largest eigenvalue of the matrix  $C(\infty) = \frac{\eta}{\eta + \gamma}W \text{diag}(\mathbf{s}(\infty))$ , and let  $\mathbf{v}_\infty$  be the corresponding left eigenvector, chosen to have all non-negative entries. If  $\mathbf{x}(0) > 0$ , then  $\mathfrak{R}_\infty < 1$  and*

$$\lim_{t \rightarrow \infty} \mathbf{x}(t)e^{(\eta + \gamma)(1 - \mathfrak{R}_\infty)t} = D\mathbf{v}_\infty,$$

and

$$\lim_{t \rightarrow \infty} \frac{d\mathbf{x}(t)}{dt}e^{(\eta + \gamma)(1 - \mathfrak{R}_\infty)t} = -D(\eta + \gamma)(1 - \mathfrak{R}_\infty)\mathbf{v}_\infty,$$

where  $D$  is a positive constant depending on the parameters of the model and the initial conditions (and the chosen normalization for  $\mathbf{v}_\infty$ ).

*Proof.* Recall the vector form of the differential equations for the number of active half-edges Eq. (53) in the SIR on SBM. By the results from Section 5.3, see in particular Lemma 5.5 and its proof,  $\mathfrak{R}_\infty < 1$ , and by the bound Eq. (27), the entries of  $C(t)$  differ from those of  $C(\infty)$  by an amount which is decaying exponentially fast in  $t$ . Combined with the Perron-Frobenius theorem, which implies that  $\mathbf{x}(t_0)e^{(\eta + \gamma)(C(\infty) - \mathfrak{R}_\infty)t}$  converges to  $\mathbf{v}_\infty$  times the scalar product of  $\mathbf{x}(t_0)$  with a suitably normalized right eigenvector and errors which are also exponentially small in  $t$ , this implies the first statement. Inserted back into the differential equation for  $\mathbf{x}(t)$ , this gives the second.  $\square$

We close this appendix with the proof of Lemma 7.1.

*Proof of Lemma 7.1.* Rescaling time so that  $\eta + \gamma = 1$  (and  $\eta = 1 - \gamma$ ), and using the differential equations for  $i_k$  and  $x_k$ , we write the difference of the derivative of  $i_k$  and  $x_k$  as

$$\frac{di_k}{dt} - \frac{dx_k}{dt} = x_k - \gamma i_k. \quad (83)$$

To prove the fact that  $x_k < i_k$  for all  $t > 0$ , we first analyze the ratio of  $x_k$  and  $i_k$  for small  $t$ , distinguishing the case  $x_k(0) > 0$  and  $x_k(0) = 0$ . If  $x_k(0) = i_k(0) > 0$ , we Taylor expand  $x_k(t)$

and  $i_k(t)$  as  $x_k(t) = x_k(0) + x'_k(0)t + O(t^2)$  and  $i_k(t) = x_k(0) + x'_k(0)t + (i'_k(0) - x'_k(0))t + O(t^2)$  where we used that  $i_k(0) = x_k(0)$ . Using the fact that  $i'_k(0) - x'_k(0) = x_k(0) - \gamma i_k(0) = \eta x_k(0)$  by Eq. (83), this implies that

$$\frac{x_k}{i_k} = 1 - \eta t + O(t^2).$$

when  $x_k(0) = i_k(0) > 0$ . If  $x_k(0) = i_k(0) = 0$  we have that  $x'_k(0) = i'_k(0) = (\mathbf{x}(0)C(0))_k > 0$  by the irreducibility of  $C(0)$  and the fact that  $x(0) > 0$ , and  $i''_k(0) - x''_k(0) = x'_k(0) - \gamma i'_k(0) = \eta x'_k(0)$ . As a consequence, we may expand  $\frac{x_k}{i_k}$  as

$$\frac{x_k}{i_k} = 1 - \frac{i_k - x_k}{i_k} = 1 - \frac{\eta x'_k(0) \frac{t^2}{2}}{x'_k(0)t} (1 + O(t)) = 1 - \frac{\eta}{2}t + O(t^2),$$

showing that in both cases  $\gamma < \frac{x_k}{i_k} < 1$  for all sufficiently small, positive  $t$ .

To prove that

$$\frac{x_k}{i_k} < 1 \quad \text{for all } t > 0,$$

we rewrite the differential equations for  $x_k$  and  $i_k$  as

$$\frac{d}{dt}x_k = (Z_k - 1)x_k \quad \text{and} \quad \frac{d}{dt}i_k = Z_k x_k - \gamma i_k$$

where

$$Z_k = \frac{(\mathbf{x}C(t))_k}{x_k}.$$

(Note that for all  $t > 0$ , both  $Z_k$  and the ratio  $\frac{x_k}{i_k}$  are well defined since  $i_k > 0$  and  $x_k > 0$  by Lemma 5.1). Next, we calculate the derivative of the ratio  $\frac{x_k}{i_k}$ , giving

$$\frac{d}{dt} \frac{x_k}{i_k} = \frac{x_k}{i_k} \left( \frac{x'_k}{x_k} - \frac{i'_k}{i_k} \right) = \frac{x_k}{i_k} \left( Z_k - 1 + \gamma - Z_k \frac{x_k}{i_k} \right) = \frac{x_k}{i_k} \left( Z_k \left( 1 - \frac{x_k}{i_k} \right) - \eta \right).$$

Assume now towards a contradiction that there exists a time  $0 < t < \infty$  such that  $\frac{x_k}{i_k} = 1$ , and let  $t^*$  be the first such time. For  $t = t^*$ , we then have that

$$\frac{d}{dt} \frac{x_k}{i_k} = 1 \cdot (Z_k \cdot 0 - \eta) = -\eta,$$

which implies that shortly before  $t^*$ ,  $\frac{x_k}{i_k} > 1$ . Since  $\frac{x_k}{i_k}$  is continuous on  $(0, \infty)$  and was strictly smaller than 1 for small positive  $t$ , there must be another time  $t'$  such that  $\frac{x_k}{i_k} = 1$ , a contradiction.

To formalize the statement about the local maxima, we let  $t_0$  be the location of the first local maximum of  $x_k$ , which means that  $\frac{dx_k}{dt} \geq 0$  for all  $t \leq t_0$ , with equality if and only if  $t = t_0$ . Our goal will be to prove that  $\frac{di_k}{dt} > 0$  for  $0 \leq t \leq t_0$ , including  $t = t_0$ .

We start by proving this fact for  $t = 0$ . To this end, we note that if  $i_k(0) = x_k(0) = 0$ , then  $i'_k(0) = (\mathbf{x}(0)C(0))_k > 0$  as we already observed earlier, while for  $i_k(0) = x_k(0) > 0$  the fact that  $x'_k(0) \geq 0$  and Eq. (83) imply that  $i'_k(0) \geq (1 - \eta)x_k(0) > 0$ .

Next consider the quantity  $Z_k$  for  $0 < t \leq t_0$ . Since the derivative of  $x_k$  is non-negative for these  $t$ , we conclude that  $Z_k \geq 1$ , and combined with the fact that  $\frac{x_k}{i_k} < 1$  for all  $0 < t < \infty$ , we conclude that

$$\frac{d}{dt} \frac{x_k}{i_k} \geq \frac{x_k}{i_k} \left( 1 - \frac{x_k}{i_k} - \eta \right) = \frac{x_k}{i_k} \left( \gamma - \frac{x_k}{i_k} \right) \quad \text{for all } 0 < t \leq t_0.$$

Recalling that  $\gamma < \frac{x_k}{i_k}$  for all sufficiently small  $t > 0$ , let  $t_\gamma$  be the first time that  $\gamma = \frac{x_k}{i_k}$  (if there is no such time, we set  $t_\gamma = \infty$ ). In view of Eq. (83), we want to show that  $t_\gamma > t_0$ , since this would imply that  $\frac{di_k}{dt} > \frac{dx_k}{dt} \geq 0$  for  $0 \leq t \leq t_0$ .

Assume towards a contradiction that  $t_\gamma \leq t_0$ , and define  $\varepsilon(t) = \log \frac{x_k(t)}{\gamma i_k(t)}$ . Then  $0 < \varepsilon(t) \leq \varepsilon(0) = -\log \gamma$  for  $0 < t < t_\gamma$  and  $\varepsilon(t_\gamma) = 0$ . Furthermore

$$\frac{d}{dt}\varepsilon(t) \geq \gamma(1 - e^{\varepsilon(t)}) \geq -\gamma\varepsilon(t)e^{\varepsilon(t)} \geq -\gamma^2\varepsilon(t) \quad \text{for all } 0 < t \leq t_\gamma,$$

implying that

$$0 = \varepsilon(t_\gamma) \geq e^{-\gamma^2 t_\gamma} \varepsilon(0) > 0,$$

a contradiction. This completes the proof of the lemma. □