DIPARTIMENTO DI MATEMATICA
"Francesco Brioschi"
POLITECNICO DI MILANO

Contact and voter processes on the infinite percolation cluster as models of host-symbiont interactions

Bertacchi, D.; Lanchier, N.; Zucca, F.

Collezione dei *Quaderni di Dipartimento*, numero **QDD 57**
Inserito negli *Archivi Digitali di Dipartimento* in data 18-11-2009

Piazza Leonardo da Vinci, 32 - 20133 Milano (Italy)

Contact and voter processes on the infinite percolation cluster as models of host-symbiont interactions

D. Bertacchi, N. Lanchier and F. Zucca

Abstract We introduce spatially explicit stochastic processes to model multispecies host-symbiont interactions. The host environment is static, modeled by the infinite percolation cluster of site percolation. Symbionts evolve on the infinite cluster through contact or voter type interactions, where each host may be infected by a colony of symbionts. In the presence of a single symbiont species, the condition for invasion as a function of the density of the habitat of hosts and the maximal size of the colonies is investigated in details. In the presence of multiple symbiont species, it is proved that the community of symbionts clusters in two dimensions whereas symbiont species may coexist in higher dimensions.

1. Introduction

The term symbiosis was coined by the mycologist Heinrich Anto de Bary to denote close and long-term physical and biochemical interactions between different species, in constrast with competition and predation that imply only brief interactions. Symbiotic relationships involve a symbiont species, smaller in size, that always benefits from the relationship, and a host species, larger in size, that may either suffer, be relatively unaffected, or also benefit from the relationship, which are referred to as parasistism, commensalism, and mutualism, respectively. Symbiotic relationships are ubiquitous in nature. For instance, more than 90% of terrestrial plants [26] live in association with mycorrhizal fungi, with the plant providing carbon to the fungus and the fungus providing nutrients to the plant, most herbivores have mutualistic gut fauna that help them digest plant matter, and almost all free-living animals are host to one or more parasite taxa [25].

To understand the role of spatial structure on the persistence of host-parasite and host-mutualist associations, Lanchier and Neuhauser [17, 18, 19] have initiated the study of multispecies host-symbiont systems including local interactions based on interacting particle systems. The stochastic process introduced in [18] describes the competition among specialist and generalist symbionts evolving in a deterministic static environment of hosts. The mathematical analysis of this model showed that fine-grained habitats promote generalist strategies, while coarse-grained habitats increase the competitiveness of specialists. The stochastic process introduced in [17, 19] includes in addition a feedback of the hosts, which is modeled by a dynamic-host system. This process has been further extended by Durrett and Lanchier [9]. The host population evolves, in the absence of symbionts, according to a biased voter model, while the symbiont population evolves in this dynamic environment of hosts according to a contact type process. The parameters of the process allow to model the effect of the symbionts on their host as well as the degree of specificity of the symbionts, thus resulting into a system of coupled interacting particle systems, each describing the evolution of a trophic level. The model is designed to understand the role of the symbionts in the spatial structure of plant communities. It is proved theoretically that generalist symbionts

AMS 2000 subject classifications: 60K35

Keywords and phrases: contact process, voter model, site percolation, logistic growth, branching random walks, random walks, host, symbiont, infrapopulation, metapopulation, infracommunity, component community.

have only a limited effect on the spatial structure of their habitat [19]. In contrast, the inclusion of specialist parasites promotes coexistence of the hosts, while the analysis of the corresponding mean-field model supported by numerical simulations suggests that in any dimension the inclusion of specialist mutualists translates into a clustering of the host environment [9].

Similarly to most spatial epidemic models such as the contact process, the state space of the stochastic processes introduced in [9, 17, 18, 19] indicates whether hosts are either healthy or infected, but does not distinguish between different levels of infection of the hosts. However, it is known from past research that the number of symbiont individuals, including ectosymbionts, i.e., symbionts living on their hosts or in their skin, associated to a single host individual may vary significantly. Mooring and Samuel [21] found for instance an average of 1,791 individuals of the species *Dermacentor albipictus*, commonly known as Winter Tick, on individual elk in Alberta, while some individual moose have been found with more than 50,000 ticks. In addition, symbionts are generally much smaller organisms than their hosts and reproduce much faster and in greater number. This motivates the development of spatially explicit multiscale models of host-symbiont interactions that describe the presence of symbionts through a level of infection of the hosts rather than binary random variables (infected versus healthy hosts) and include both inter-host symbiont dynamics and intra-host symbiont dynamics.

In diversity ecology, the infrapopulation refers to all the parasites of one species in a single individual host, while the metapopulation refers to all the parasites of one species in the host population. In systems involving multiple species of parasites, all the parasites of all species in a single individual host and in an entire host population are called infracommunity and component community, respectively. This terminology shall be employed in this article for symbionts in general, that is parasites, commensalists and mutualists, even though, strictly speaking, it only applies to parasites. Our main objective is to deduce from the microscopic evolution rules of the symbionts, described by transmission rates and reproduction rates, the long-term behavior of the metapopulation in a single-species invasion model, and the long-term behavior of the component community in a multispecies competition model. Since a host species and a symbiont species involved in a symbiotic relationship usually evolve at very different time scales (symbionts reproduce much faster than their hosts), we shall assume in both invasion and competition models that the discrete habitat of hosts is static. This habitat will be modeled by a realization of the infinite percolation cluster of supercritical site percolation. We shall also assume that symbionts can only survive when associated with a host (obligate relationship), which restricts their habitat to the infinite percolation cluster, and, to understand the role of space on the persistence of the symbiotic relationship, that symbionts can only transmit to nearby hosts, adding to the complexity of the interactions. In the single-species model, infrapopulations will evolve according to the logistic growth process, and the entire metapopulation according to a mixture of this model and its spatial analog, the contact process [13]. In the multispecies model, we will assume that infracommunities evolve according to the Moran model [22], and the entire component community according to a mixture of this model and its spatial analog, the voter model [5, 15]. Our analysis shows that the condition for survival of a metapopulation strongly depends on the carrying capacity of each infrapopulation. Exact calculations of the critical curve as a function of the reproduction and transmission rates are given when infrapopulations can be arbitrarily large which, as mentioned above, is a realistic biological assumption in many symbiotic relationships. In systems involving multiple symbiont species, the long-term behavior of the component community depends on the spatial dimension: the community clusters in two dimensions whereas coexistence is possible in higher dimensions.

2. Models and results

The models are constructed in two steps. First, the static random environment of hosts is fixed from a realization of the infinite percolation cluster of site percolation. This random environment naturally induces a random graph. The symbionts are then introduced into this universe where they evolve according to an interacting particle system on the random graph. The interactions are modeled based on two of the simplest interacting particle systems: the contact process and the voter model. The structure of the random graph implies that the infrapopulation dynamics are described by logistic growth processes, i.e., contact processes on a complete graph, while the infracommunity dynamics are described by Moran models, i.e., voter models on a complete graph.

Host environment. To define the habitat of hosts, we set $p \in (0, 1]$ and let ω be a realization of the site percolation process with parameter p on the d-dimensional regular lattice $\mathbb{Z}^d$, i.e., each site of the lattice is either permanently occupied by an individual host with probability p or permanently empty with probability $1 - p$. Let $\mathbb{H}(\omega)$ denote the set of open/occupied sites. By convention, elements of $\mathbb{Z}^d$ and processes with state space $S \subset \mathbb{Z}^d$ will be denoted in the following by capital Latin letters. We say that there is an open path between site X and site Y if there exists a sequence of sites $X = X_0, X_1, \ldots, X_n = Y$ such that the following two conditions hold:

1. For $i = 0, 1, \ldots, n$, we have $X_i \in \mathbb{H}(\omega)$, i.e., site X_i is open.
2. For $i = 0, 1, \ldots, n - 1$, we have $X_i \sim X_{i+1}$

where $X_i \sim X_{i+1}$ means that the Euclidean norm $\|X_i - X_{i+1}\| = 1$. Writing $X \rightleftharpoons Y$ the event that sites X and Y are connected by an open path, we observe that the binary relation $\rightleftharpoons$ is an equivalence relation on the random set $\mathbb{H}(\omega)$ thus inducing a partition of $\mathbb{H}(\omega)$. In dimensions $d \geq 2$ there exists a critical value $p_c \in (0, 1)$ that depends on d such that if $p > p_c$ then $\mathbb{H}(\omega)$ contains a unique infinite open cluster. The infinite open cluster is also called infinite percolation cluster and is denoted by $C_\infty(\omega)$ later. We assume that $p > p_c$ from now on. Sometimes, the infinite percolation cluster will be identified with the graph with vertex set $C_\infty(\omega)$ obtained by drawing an edge between sites of the cluster at Euclidean distance 1 from each other.

Random graph structure. In order to define the state space and dynamics of the stochastic processes, we first define a random graph $\mathcal{H}(\omega)$ as follows. Vertices of $\mathcal{H}(\omega)$ are to be interpreted as possible locations for the symbionts, while edges indicate how symbionts interact. Let N be an integer and $\mathbb{K}_N = \{1, 2, \ldots, N\}$. The vertex set of $\mathcal{H}(\omega)$ is $C_N(\omega) = C_\infty(\omega) \times \mathbb{K}_N$. By convention, elements of and processes with state space $C_N(\omega)$ will be denoted by small Latin letters. Let

$$\pi \ : \ C_N(\omega) \ \longrightarrow \ C_\infty(\omega) \quad \text{defined by} \quad \pi(x) = X \quad \text{for all } x = (X, i) \in C_N(\omega).$$

That is, $\pi(x)$ is the $C_\infty(\omega)$-coordinate of vertex x. Let $x, y \in C_N(\omega)$. Then, vertices x and y are connected by an edge if and only if one of the following two cases occurs.

1. If $\pi(x) = \pi(y)$ then x and y are connected by a vertical edge: we write $x \updownarrow y$. It is convenient to assume that each vertex is connected to itself by a vertical edge.
2. If $\pi(x) \sim \pi(y)$ then x and y are connected by a horizontal edge: we write $x \leftrightarrow y$.

In words, a complete graph with N vertices (which are connected to themselves) is attached to each site of the infinite percolation cluster. Edges of these complete graphs are said to be vertical while,

for any two sites of the infinite percolation cluster, vertices of the corresponding complete graphs are connected by edges which are said to be horizontal. Vertical and horizontal edges correspond respectively to potential reproduction events and transmission events of the symbionts.

Invasion of a single symbiont – contact process. To understand the conditions for survival of a single symbiont species, we introduce a generalization of the contact process [13] on the infinite random graph $\mathcal{H}(\omega)$. This defines a continuous-time Markov process whose state space consists of the set of the spatial configurations $\eta : C_N(\omega) \longrightarrow \{0,1\}$, and whose dynamics are described by the Markov generator L_1 defined on the set of the cylinder functions by

$$\begin{aligned} L_1 f(\eta) \;=\; & \sum_{x \in C_N(\omega)} [f(\eta_{x,0}) - f(\eta)] \\ & + \sum_{x \in C_N(\omega)} \left(\frac{\alpha}{N} \sum_{x \updownarrow y} \eta(y) \;+\; \frac{\beta}{N \deg \pi(x)} \sum_{x \leftrightarrow y} \eta(y) \right) [f(\eta_{x,1}) - f(\eta)] \end{aligned}$$

where $\deg \pi(x)$ is the degree of $\pi(x)$ as a site of the cluster $C_\infty(\omega)$, and where $\eta_{x,i}$ is the configuration obtained from η by assigning the value i to vertex x. Note that the degree of each site of the infinite percolation cluster is at least 1, therefore the dynamics are well defined. Thinking of vertices in state 0 as uninfected and vertices in state 1 as infected by a symbiont, the expression of the Markov generator above indicates that symbionts die independently of each other at rate 1, reproduce within their host at the reproduction rate α, and transmit their offspring to the nearby hosts at the transmission rate β. That is, each symbiont gives birth at rate α to an offspring which is then sent to a vertex chosen uniformly at random from the parent's host. If the vertex is uninfected then it becomes infected while if it is already infected then the birth is suppressed. Similarly, each symbiont gives birth at rate β to an offspring which is then sent to a vertex chosen uniformly at random from the hosts adjacent to the parent's host, which results as previously in an additional infection if and only if the vertex is not already infected. To study the single-species model, we will sometimes consider the stochastic process

$$\bar{\eta}_t(X) \;=\; \sum_{\pi(x) = X} \eta_t(x) \qquad \text{for all } X \in C_\infty(\omega)$$

where the sum is over the vertices $x \in C_N(\omega)$ such that $\pi(x) = X$. That is, $\bar{\eta}_t(X)$ keeps track of the level of infection of the host at X. This defines a Markov process whose state space consists of the functions that map $C_\infty(\omega)$ into $\{0, 1, \ldots, N\}$ and whose dynamics are described by

$$\begin{aligned} \bar{L}_1 f(\bar{\eta}) \;=\; & \sum_{X \in C_\infty(\omega)} \bar{\eta}(X) \, [f(\bar{\eta}_{X-}) - f(\bar{\eta})] \\ & + \sum_{X \in C_\infty(\omega)} \left(1 - \frac{\bar{\eta}(X)}{N} \right) \left(\alpha \; \bar{\eta}(X) \;+\; \frac{\beta}{\deg(X)} \sum_{X \sim Y} \bar{\eta}(Y) \right) [f(\bar{\eta}_{X+}) - f(\bar{\eta})] \end{aligned}$$

where the configurations $\bar{\eta}_{X-}$ and $\bar{\eta}_{X+}$ are obtained from the configuration $\bar{\eta}$ by respectively removing and adding a symbiont at site X. In view of the geometry of the graph $\mathcal{H}(\omega)$, the stochastic process $\{\bar{\eta}_t\}_t$ can be seen as a mixture of the contact process with infection parameter β on the infinite percolation cluster and logistic growth processes with parameter α.

To describe the predictions based on the invasion model, we let δ_i be the measure that concentrates on the "all i" configuration restricted to $C_N(\omega)$, i.e.,

$$\delta_i \{\eta(x) = i\} \;=\; 1 \quad \text{for all } x \in C_N(\omega) = C_\infty(\omega) \times \mathbb{K}_N.$$

We denote by $\bar{\mu}$ the upper invariant measure of the process $\{\eta_t\}_t$, which is also the limit starting from the measure δ_1 since the process is attractive. The process or metapopulation is said to survive whenever $\bar{\mu} \neq \delta_0$ and is said to die out otherwise.

First, we observe that, starting with a single infection at time 0, the number of symbionts in the system is dominated stochastically by the number of individuals in a birth and death process with birth parameter $b = \alpha + \beta$ and death parameter 1. Recurrence of one-dimensional symmetric random walks implies that such a process eventually dies out when $b \leq 1$. It follows that $\{\eta_t\}_t$ dies out for all values of N whenever $\alpha + \beta \leq 1$.

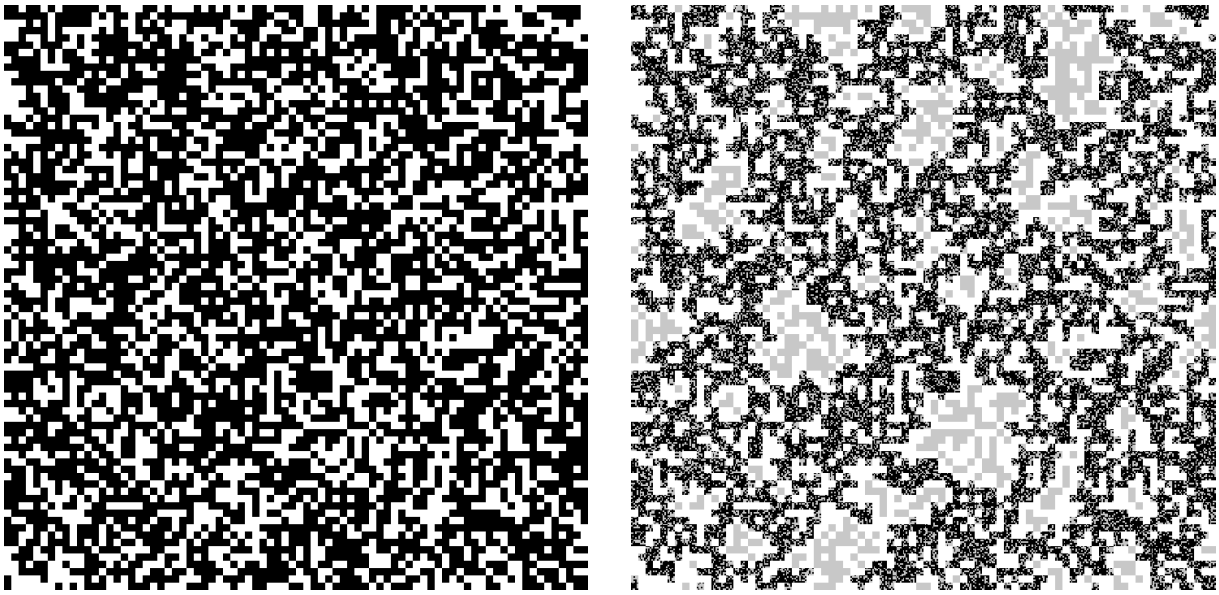
To find a general condition for survival of the infection, we now assume that $N = 1$ so that the value of the reproduction rate α becomes irrelevant, and compare the process with the one-dimensional contact process. Let Γ be an arbitrary infinite self-avoiding path in the infinite percolation cluster $C_\infty(\omega)$. Since for all sites $X \in \Gamma$ we have $2 \leq \deg(X) \leq 2d$, the process restricted to the infinite path Γ, i.e., symbionts sent outside Γ are instantaneously killed, dominates stochastically the contact process on Γ with infection parameter β/d. It follows that the process survives whenever $\beta > d\,\beta_c(1)$ where $\beta_c(1)$ is the critical value of the one-dimensional contact process, since the self-avoiding path is isomorphic to $\mathbb{Z}$. Monotonicity properties of the process finally implies that, for all values of N and α, survival occurs whenever $\beta > d\,\beta_c(1)$.

We now look at the long-term behavior of the metapopulation when N is large. As previously explained, this assumption is realistic in a number of symbiotic relationships, including the interactions between moose and Winter Ticks [21]. Under this assumption, at least when the number of symbionts is not too large, the stochastic process looks locally like a branching random walk on the random graph $\mathcal{H}(\omega)$, namely the process modified so that births onto infected vertices are allowed. In the context of large infrapopulations, global survival of the metapopulation occurs when the reproduction rate $\alpha > 1$ and the transmission rate $\beta > 0$. This and the comparison with a birth and death process imply that, when N is large and the transmission rate β is small, a situation which is common in parasitic relationships, the metapopulation undergoes a phase transition when the reproduction rate α approaches 1. Provided the density of the habitat is large enough, the phase transition occurs more generally when the sum of the reproduction and transmission rates approaches 1. These results are summarized in the following theorem.

Theorem 1 (contact interactions) *Assume that $p > p_c$ and $\beta > 0$.*

1. *For all $N > 0$, the metapopulation dies out if $\alpha + \beta \leq 1$ while it survives if $\beta > d\,\beta_c(1)$.*
2. *If $\alpha > 1$ or $\alpha + \beta/d > 1$ then the metapopulation survives for N large.*
3. *If $\alpha + \beta > 1$ and p is close to 1 then the metapopulation survives for N large.*

As previously explained, the first statement of part 1 follows from a comparison with a two-parameter branching random walk, and the second statement from a comparison with the contact process restricted to an infinite self-avoiding path embedded in the infinite percolation cluster. The second statement of part 2 can be proved similarly, by using part 3 and looking at the process restricted to an infinite self-avoiding path of hosts. The proofs of the first statement of part 2 and part 3 rely on the combination of estimates for the extinction time of the logistic growth process,



(a) $p = 0.6$ (b) $\alpha = 1$ and $\beta = 2$

FIGURE 1. (a) Realization of site percolation with parameter $p = 0.6$ on the 80×80 torus, with black squares referring to open sites, and white squares to closed sites. (b) Snapshot of the invasion model on the percolation structure starting with a single infected host at the center of the universe. Each site is represented by a 5×5 square, i.e., complete graph with $N = 25$ vertices. White squares refer to empty sites, i.e., sites which are not occupied by a host, black dots refer to symbionts, and grey dots to empty vertices.

random walks estimates, and block constructions to compare the process view under suitable space and time scales with oriented percolation. We refer to Section 3 for the details of the proof.

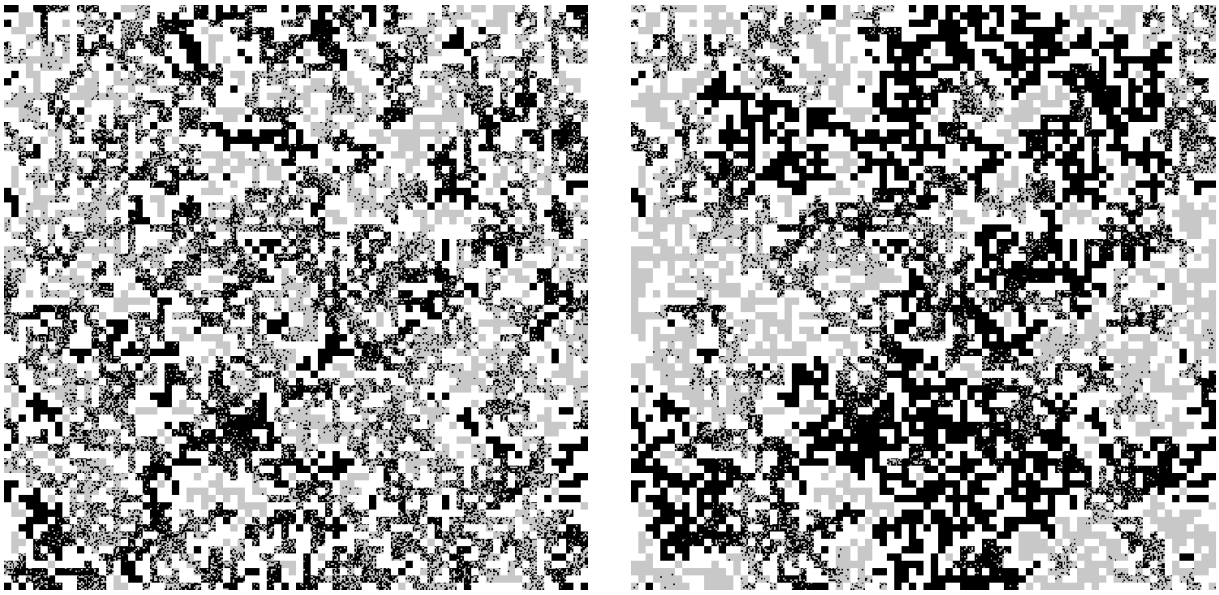
Competition among multiple symbionts – voter model. To study the interactions among multiple symbiont species, we introduce the analog of the previous spatial model replacing contact interactions with voter interactions [5, 15]. The state at time t is now $\xi_t : C_N(\omega) \longrightarrow \{1, 2\}$, i.e., each vertex is occupied by a symbiont of one of two types. Letting for $i = 1, 2$

$$f_i(x) \;=\; \frac{|\{y : \pi(y) = \pi(x) \text{ and } \xi_t(y) = i\}|}{N} \quad \text{and} \quad g_i(x) \;=\; \frac{|\{y : \pi(y) \sim \pi(x) \text{ and } \xi_t(y) = i\}|}{N \deg \pi(x)}$$

denote the fraction of type i symbionts at site $\pi(x)$ and its neighborhood, respectively, the evolution is described by the Markov generator L_2 defined on the set of the cylinder functions by

$$\begin{aligned} L_2 f(\xi) \;=\; & \sum_{x \in C_N(\omega)} \frac{\alpha_1 f_1(x) + \beta_1 g_1(x)}{\alpha_1 f_1(x) + \alpha_2 f_2(x) + \beta_1 g_1(x) + \beta_2 g_2(x)} \, [f(\xi_{x,1}) - f(\xi)] \\ & + \sum_{x \in C_N(\omega)} \frac{\alpha_2 f_2(x) + \beta_2 g_2(x)}{\alpha_1 f_1(x) + \alpha_2 f_2(x) + \beta_1 g_1(x) + \beta_2 g_2(x)} \, [f(\xi_{x,2}) - f(\xi)] \end{aligned}$$

where $\xi_{x,i}$ is the configuration obtained from ξ by assigning the value i to vertex x. The transition rates indicate that, regardless of its type, each symbiont dies at rate 1 and gets instantaneously



(a) $\alpha_1 = \alpha_2 = 1/2$ and $\beta_1 = \beta_2 = 1/2$ (b) $\alpha_1 = \alpha_2 = 1/2$ amd $\beta_1 = \beta_2 = 1/2$

FIGURE 2. Snapshots at time 100 and at time 1000, respectively, of the neutral competition model on site percolation with parameter $p = 0.6$ starting from a Bernoulli product measure with density $1/2$. Each site of the lattice is represented by a 5×5 square, i.e., complete graph with $N = 25$ vertices. White squares refer to empty sites, i.e., sites which are not occupied by a host, and black and grey dots to symbionts of type 1 and 2, respectively.

replaced by a symbiont whose type is chosen from the nearby symbionts according to the relative fecundities and transmissibilities of the two symbiont species. In the neutral case when the reproduction rates are both equal to say α and the transmission rates are both equal to say β the local evolution reduces to the following: the type of each symbiont is updated at rate 1 and the new type is chosen uniformly at random from the same host with probability $\alpha/(\alpha + \beta)$ or a nearby host with probability $\beta/(\alpha + \beta)$. The process $\{\xi_t\}_t$ can again be seen as a mixture of two well-known processes, namely, the Moran model, continuous-time version of the Wright-Fisher model, with selection, and its spatial analog, the biased voter model, on the infinite percolation cluster.

To state our results for the competition model, we set $\theta \in (0, 1)$ and denote by π_θ the product measure restricted to $C_N(\omega) = C_\infty(\omega) \times \mathbb{K}_N$ defined by

$$\pi_\theta \{\xi(x) = 1\} \;=\; \theta \quad \text{and} \quad \pi_\theta \{\xi(x) = 2\} \;=\; 1 - \theta \quad \text{for all } x \in C_N(\omega).$$

From now on, we assume that the process $\{\xi_t\}_t$ starts from the product measure π_θ. The process is said to cluster if there exists $a \in (0, 1)$ such that $\xi_t \Rightarrow a\delta_1 + (1 - a)\delta_2$ as $t \to \infty$ and

$$\lim_{t\to\infty} P\left(\xi_t(x) \neq \xi_t(y)\right) \;=\; 0 \quad \text{for all } \; x, y \in C_N(\omega),$$

and is said to coexist if $\xi_t \Rightarrow \nu_\theta$ as $t \to \infty$ for some ν_θ such that

$$\nu_\theta \{\xi(x) \neq \xi(y)\} \;\neq\; 0 \quad \text{for all } x, y \in C_N(\omega),\; x \neq y.$$

Here, $\Rightarrow$ denotes convergence in distribution. Finally, we say that type 1 outcompetes type 2 if the process converges in distribution to the measure δ_1.

Theorem 2 (voter interactions)

1. *Assume that $p > p_c$. In the neutral case $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, the component community clusters in two dimensions, whereas coexistence occurs in higher dimensions.*
2. *Assume that $p = 1$. If $\alpha_1 \geq \alpha_2$ and $\beta_1 > \beta_2$ then type 1 outcompetes type 2.*

The analysis of the neutral competition model relies on duality techniques. We will show, in the neutral case, that the process with voter type interactions is dual to a certain system of coalescing random walks evolving on the random graph $\mathcal{H}(\omega)$. The long-term behavior of the process is related to the so-called finite/infinite collision property of the graph, which will be studied in details in two dimensions and higher dimensions separately. In contrast, the process with selection will be studied without invoking the dual process but instead by comparing the evolution of the number of type 1 symbionts forward in time with the gambler's ruin model.

3. Proof of Theorem 1

This section is devoted to the invasion model and the proof of Theorem 1. The key to proving the first statement of part 2 and part 3 is to show that the logistic growth process with $\alpha > 1$ as well as the branching random walk with $\alpha + \beta > 1$ restricted to a large square persist an arbitrary long time provided N is large. In both cases, the combination of our estimates with a block construction will imply survival of the metapopulation restricted to an infinite self-avoiding path of hosts and an infinite self-avoiding path of large squares fully occupied by hosts, respectively. Such infinite self-avoiding paths exist whenever $p > p_c$ and p is close enough to 1, respectively.

The logistic growth process. Recall that the number of symbionts in each infrapopulation evolves according to a logistic growth process with parameter α, which is the continuous-time Markov process $\{X_t\}_t$ that makes the transitions

$$j \;\to\; \begin{cases} j+1 & \text{at rate} \quad \alpha\, j\,(1 - jN^{-1}) \\ j-1 & \text{at rate} \quad j \end{cases}$$

for all $j = 0, 1, \ldots, N$. Our main objective is to prove that the logistic growth process starting with a single particle persists an arbitrarily long time with positive probability when $\alpha > 1$ and N is sufficiently large. For our purpose, it is enough to prove that, for all N large,

$$T_{\text{ext}} \;=\; \inf\{t > 0 : X_t = 0\}$$

is larger than $2N^2$ with positive probability. Since $\alpha > 1$, there exists $\epsilon > 0$ small, fixed from now on, such that $4\epsilon < 1 - \alpha^{-1}$. For such $\epsilon > 0$, we have the following result.

Lemma 3 *There exist $C_1 < \infty$ and $\gamma_1 > 0$ such that*

$$P\left(T_{\text{ext}} < 2N^2 \mid X_0 > 4\epsilon N\right) \;\leq\; C_1 \exp(-\gamma_1 N).$$

Proof. We set $T_0 = 0$ and, for all $i \geq 0$, define recursively the stopping times

$$T_{2i+1} \;=\; \inf\{t > T_{2i} : X_t = \lfloor 2\epsilon N \rfloor\} \quad \text{and} \quad T_{2i+2} \;=\; \inf\{t > T_{2i+1} : X_t \notin [1, 4\epsilon N)\}$$

where $\lfloor \cdot \rfloor$ is for the integer part. Let $I = \min\{i \geq 0 : X_{T_{2i}} = 0\}$. Between T_{2i} and T_{2i+2} the process undergoes at least $2\epsilon N$ death events, each occurring at rate at most N, so large deviation estimates for the Poisson distribution imply that

$$P\left(T_{2i+2} - T_{2i} < \epsilon \mid i < I\right) \;\leq\; C_2 \exp(-\gamma_2 N) \tag{1}$$

for suitable $C_2 < \infty$ and $\gamma_2 > 0$. Observe that, for all states $j \in [1, 4\epsilon N)$, the ratio of the birth rate to the death rate in the logistic growth process can be bounded from below as follows:

$$\frac{\alpha\, j\,(1 - jN^{-1})}{j} \;=\; \alpha\left(1 - \frac{j}{N}\right) \;>\; \alpha\,(1 - 4\epsilon) \;=\; a \;>\; 1.$$

In particular, letting $\{Z_n\}_n$ be the asymmetric random walk with

$$P\left(Z_{n+1} = j \mid Z_n = i\right) \;=\; \begin{cases} a/(a+1) & \text{for } i = j+1 \\ 1/(a+1) & \text{for } i = j-1 \end{cases} \tag{2}$$

for all $i \in \mathbb{Z}$, we have

$$\begin{aligned} P\left(I = i \mid I > i-1\right) \;&=\; P\left(X_{T_{2i}} = 0 \mid X_{T_{2i-1}} = \lfloor 2\epsilon N \rfloor\right) \\ &\leq\; P\left(Z_n = 0 \text{ for some } n \geq 0 \mid Z_0 = \lfloor 2\epsilon N \rfloor\right) \;=\; a^{-2\epsilon N}. \end{aligned}$$

This implies that the random variable I is stochastically larger than a geometric random variable with success probability $a^{-2\epsilon N}$. Finally, using $T_{\text{ext}} = T_{2I}$ and the bound (1), we deduce

$$\begin{aligned} P\left(T_{\text{ext}} < 2N^2 \mid X_0 > 2\epsilon N\right) \;&\leq\; P\left(I < 2\epsilon^{-1}N^2\right) \\ &+ P\left(I \geq 2\epsilon^{-1}N^2 \text{ and } T_{2i+2} - T_{2i} < \epsilon \text{ for some } i = 0, 1, \ldots, 2\epsilon^{-1}N^2 - 1\right) \\ &\leq\; 1 \;-\; \left(1 - a^{-2\epsilon N}\right)^{2\epsilon^{-1}N^2} \;+\; 2\epsilon^{-1}N^2 \times C_2 \exp(-\gamma_2 N) \\ &\leq\; 2\epsilon^{-1}N^2\left(a^{-2\epsilon N} \;+\; C_2 \exp(-\gamma_2 N)\right) \;\leq\; C_1 \exp(-\gamma_1 N) \end{aligned}$$

for suitable constants $C_1 < \infty$ and $\gamma_1 > 0$. $\square$

The next step is to prove survival of the metapopulation restricted to an infinite self-avoiding path of hosts, say $\Gamma = \{\Gamma_i : i \in \mathbb{Z}\} \subset C_\infty(\omega)$. The idea is to compare the evolution of the invasion model viewed on suitable length and time scales along this infinite path with oriented percolation, which relies on a block construction. To couple the process with oriented percolation, we let

$$\mathcal{G} \;=\; \{(i, m) \in \mathbb{Z} \times \mathbb{Z}_+ : i + m \text{ is even}\},$$

and declare site $(i, m) \in \mathcal{G}$ to be good if at all times between times mN^2 and $(m+1)N^2$ the host at site Γ_i is infected by at least one symbiont. Let $\epsilon_0 > 0$ small and let

$$G_m \;=\; \{i \in \mathbb{Z} : (i, m) \text{ is a good site}\}$$

denote the set of good sites at level m. The proof of the first statement of part 2 is based on a rescaling technique introduced in [4]. It suffices to establish that the set of good sites at level m dominates the set W_m of wet sites at level m in a 1-dependent oriented site percolation process on the lattice $\mathcal{G}$ with parameter $1 - \epsilon_0$ and with initial condition $W_0 \subset G_0$ (see Durrett [8] for a complete description of oriented percolation). To study the invadibility of the metapopulation along the infinite self-avoiding path Γ, we let Y_t denote the number of symbionts associated to the host at site Γ_1 at time t. Then, we have the following result.

Lemma 4 *Assume that* $(0,0)$ *is good and let* $T^\star = \inf\{t > 0 : Y_t > 4\epsilon N\}$. *Then,*

$$P\left(T^\star > N^2 \mid Y_0 = 0\right) \;\le\; C_3 \exp(-\gamma_3 N)$$

for suitable $C_3 < \infty$ *and* $\gamma_3 > 0$.

Proof. First, we observe that, condition on the event that the host at site Γ_0 is infected by at least one symbiont, the process $\{Y_t\}_t$ makes the transitions

$$j \;\to\; \begin{cases} j+1 & \text{at rate} \quad \text{at least } (\alpha\, j + \beta/2d)\,(1 - jN^{-1}) \\ j-1 & \text{at rate} \quad j. \end{cases}$$

Let $0 < T_1 < T_2 < T^\star$ be two consecutive jumping times and $a = \alpha\,(1 - 4\epsilon) > 0$. Since the process jumps from state j to state $j+1$ at rate at least $\alpha\, j\,(1 - jN^{-1})$, we have

$$\mathbb{E}\left[Y_{T_2} - Y_{T_1} \mid Y_{T_1} = j\right] \;\ge\; \frac{\alpha\,(1 - jN^{-1}) - 1}{\alpha\,(1 - jN^{-1}) + 1} \;\ge\; \frac{a-1}{a+1} \;>\; 0$$

for all $j \in \{1, \ldots, \lfloor 4\epsilon N \rfloor\}$, while

$$\mathbb{E}\left[Y_{T_2} - Y_{T_1} \mid Y_{T_1} = 0\right] \;\ge\; \frac{\beta}{2d} \;>\; 0.$$

Let $b = (1/4)\min(1, \beta/2d) > 0$ and let J be the number of jumps performed by $\{Y_t\}_t$ by time N^2. Since jumps occur at rate at least $4b$, large deviation estimates for the Poisson distribution imply that there exist $C_4 < \infty$ and $\gamma_4 > 0$ such that

$$P\left(J \le 2bN^2\right) \;\le\; C_4 \exp(-\gamma_4 N). \tag{3}$$

In other respect, large deviation estimates for the Binomial distribution imply that

$$\begin{aligned} &P\left(Y_t \le 4\epsilon N \text{ for all } t \le N^2 \text{ and } J > 2bN^2 \mid Y_0 = 0\right) \\ &\qquad\le\; P\left(Z_n \le 4\epsilon N \text{ for all } n = 0, 1, \ldots, \lfloor 2bN^2 \rfloor \mid Z_0 = 0\right) \\ &\qquad\le\; P\left(Z_{\lfloor 2bN^2 \rfloor} \le 4\epsilon N \mid Z_0 = 0\right) \;\le\; P\left(Z_{\lfloor 2bN^2 \rfloor} \le \mathbb{E}\left[Z_1 - Z_0\right] bN^2 \mid Z_0 = 0\right) \\ &\qquad\le\; C_5 \exp(-\gamma_5 N) \end{aligned} \tag{4}$$

for suitable constants $C_5 < \infty$ and $\gamma_5 > 0$, and all N sufficiently large. Recall that $\{Z_n\}_n$ is the asymmetric random walk defined in (2) above. Finally, combining (3)-(4), we obtain

$$\begin{aligned} P\left(T^\star > N^2 \mid Y_0 = 0\right) \;&=\; P\left(Y_t \le 4\epsilon N \text{ for all } t \le N^2 \mid Y_0 = 0\right) \\ &\le\; P\left(J \le 2bN^2\right) \;+\; P\left(Y_t \le 4\epsilon N \text{ for all } t \le N^2 \text{ and } J > 2bN^2 \mid Y_0 = 0\right) \\ &\le\; C_4 \exp(-\gamma_4 N) \;+\; C_5 \exp(-\gamma_5 N) \;\le\; C_3 \exp(-\gamma_3 N). \end{aligned}$$

This completes the proof. □

Combining Lemmas 3 and 4, and using the fact that the evolution rules of the process $\{\eta_t\}_t$ are homogeneous in time imply that, for any site $(i,m) \in \mathcal{G}$,

$$\begin{aligned} P\left((i \pm 1, m+1) \text{ is not good} \mid (i,m) \text{ is good}\right) \\ &\leq P\left(Y_t = 0 \text{ for some } N^2 < t < 2N^2 \mid (0,0) \text{ is good}\right) \\ &\leq P\left(T^\star > N^2 \mid (0,0) \text{ is good}\right) \\ &\quad + P\left(Y_t = 0 \text{ for some } N^2 < t < 2N^2 \text{ and } T^\star \leq N^2\right) \\ &\leq P\left(T^\star > N^2 \mid (0,0) \text{ is good}\right) + P\left(T_{\text{ext}} < 2N^2 \mid X_0 > 4\epsilon N\right) \leq \epsilon_0 \end{aligned}$$

for sufficiently large N. This implies that the process and 1-dependent oriented percolation with parameter $1-\epsilon_0$ can be constructed on the same probability space in such a way that

$$P\left(W_m \subset G_m \text{ for all } m \geq 0 \mid W_0 \subset G_0\right) = 1.$$

Choosing $\epsilon_0 > 0$ such that $1-\epsilon_0$ is greater than the critical value of 1-dependent oriented percolation, and observing that, starting with all hosts associated with N symbionts, $\operatorname{card} G_0 = \infty$, we can conclude that $\bar{\mu}_1 \neq \delta_0$, which proves the first statement of part 2.

Branching random walk. We now prove part 3 of Theorem 1 relying on branching random walk estimates. We first establish the result when $p = 1$ and so $C_\infty(\omega) = \mathbb{Z}^d$. To begin with, we observe that, for all $M > 0$ and $\delta \in (0,1)$, the process $\{\bar{\eta}_t\}_t$ dominates, for N sufficiently large, the process $\{\zeta_t\}_t$ whose dynamics are described by the Markov generator

$$\begin{aligned} D_1 f(\zeta) = &\sum_{X \in C_\infty(\omega)} \zeta(X)\left[f(\zeta_{X-}) - f(\zeta)\right] + \sum_{X \in C_\infty(\omega)} (1-\delta)\, \mathbb{1}\{\zeta(X) \leq M\} \\ &\times \left(\alpha\, \zeta(X) + \frac{\beta}{2d} \sum_{X \sim Y} \zeta(Y)\right)\left[f(\zeta_{X+}) - f(\zeta)\right] \end{aligned}$$

where the configurations ζ_{X-} and ζ_{X+} are obtained from ζ by respectively removing and adding a symbiont at site X. Indeed, it suffices that $N \geq M/\delta$ since in that case

$$1 - jN^{-1} \geq 1 - \delta j M^{-1} \geq (1-\delta)\, \mathbb{1}\{j \in [0,M]\} \quad \text{for all } j = 0, 1, \ldots, N.$$

To see this, we observe that the process $\{\zeta_t\}_t$ is a truncated branching random walk that allows at most $M+1$ particles per site at the same time. See Figure 3 where we compare the reproduction rates to site Y of a particle living at site X for the processes $\{\bar{\eta}_t\}_t$ and $\{\zeta_t\}_t$: on the x-axis we have the number of particles at site Y, and the parameter χ is equal to α if $X = Y$ and β if $X \sim Y$.

Let $\delta > 0$ such that $(\alpha+\beta)(1-\delta) > 1$. We will prove, following [3], that for all M sufficiently large, the truncated branching random walk $\{\zeta_t\}_t$ survives with positive probability, by looking at the branching random walk $\{\bar{\zeta}_t\}_t$ whose dynamics are described by

$$\begin{aligned} \bar{D}_1 f(\bar{\zeta}) = &\sum_{X \in C_\infty(\omega)} \bar{\zeta}(X)\left[f(\bar{\zeta}_{X-}) - f(\bar{\zeta})\right] \\ &+ \sum_{X \in C_\infty(\omega)} \left(\bar{\alpha}\, \bar{\zeta}(X) + \bar{\beta} \sum_{X \sim Y} \bar{\zeta}(Y)\right)\left[f(\bar{\zeta}_{X+}) - f(\bar{\zeta})\right] \end{aligned}$$

starting with one particle at the origin, where $\bar{\alpha} = (1-\delta)\,\alpha$ and $\bar{\beta} = (1-\delta)\,\beta/2d$.

Lemma 5 *For $X \sim 0$, we have $\mathbb{E}\,[\bar\zeta_n(X) \mid \bar\zeta_0(0) = 1] > 1$ for n large enough.*

Proof. We observe that $\mathbb{E}\,(\bar\zeta_t(X))$ satisfies the differential equation (see [2, Section 4])

$$\frac{d}{dt}\,\mathbb{E}\,(\bar\zeta_t(X)) \;=\; -\,\mathbb{E}\,(\bar\zeta_t(X)) \;+\; \bar\alpha\,\mathbb{E}\,(\bar\zeta_t(X)) \;+\; \bar\beta\sum_{X\sim Y}\mathbb{E}\,(\bar\zeta_t(Y)),$$

whose solution is

$$\mathbb{E}\,(\bar\zeta_t(X)) \;=\; \sum_{n=0}^{\infty}\sum_{k=0}^{n-1}\mu^{(n,k)}(0,X)\;\frac{\bar\alpha^k\,\bar\beta^{n-k}\,t^n}{n!}\;e^{-t} \tag{5}$$

where $\mu^{(n,k)}(0,X)$ is the number of paths from site 0 to site X of length n with k loops. To estimate the right-hand side of equation (5), we let $\{U_k\}_k$ be the discrete-time random walk with

$$P\,(U_{k+1} = Z \mid U_k = Y) \;=\; \begin{cases} \bar\alpha\,(\bar\alpha + 2d\bar\beta)^{-1} & \text{for } Y = Z \\ \bar\beta\,(\bar\alpha + 2d\bar\beta)^{-1} & \text{for } Y \sim Z \end{cases}$$

and observe that, for any site $X \sim 0$,

$$\sum_{k=0}^{n-1}\mu^{(n,k)}(0,X)\;\frac{\bar\alpha^k\,\bar\beta^{n-k}}{(\bar\alpha+2d\bar\beta)^n} \;=\; P\,(U_n = X \mid U_0 = 0) \;\geq\; C_6\,n^{-d/2}$$

for a suitable $C_6 = C_6(\bar\alpha,\bar\beta) > 0$. We refer to [27, Corollary 13.11] for the asymptotic estimates of the n-step probabilities. In particular, for $X \sim 0$ and $t = n$, we obtain

$$\begin{aligned}
\mathbb{E}\,(\bar\zeta_n(X)) \;&\geq\; \sum_{k=0}^{n-1}\mu^{(n,k)}(0,X)\;\frac{\bar\alpha^k\,\bar\beta^{n-k}\,n^n}{n!}\;e^{-n} \\
&=\; \sum_{k=0}^{n-1}\mu^{(n,k)}(0,X)\;\frac{\bar\alpha^k\,\bar\beta^{n-k}}{(\bar\alpha+2d\bar\beta)^n}\;\frac{n^n(\bar\alpha+2d\bar\beta)^n}{n!}\;e^{-n} \\
&\overset{n\to\infty}{\sim}\; \frac{(\bar\alpha+2d\bar\beta)^n}{\sqrt{2\pi n}}\;\sum_{k=0}^{n-1}\mu^{(n,k)}(0,X)\;\frac{\bar\alpha^k\,\bar\beta^{n-k}}{(\bar\alpha+2d\bar\beta)^n} \;\geq\; \frac{(\bar\alpha+2d\bar\beta)^n}{\sqrt{2\pi n}}\;C_7\,n^{-d/2}.
\end{aligned}$$

for a suitable $C_7 > 0$. Finally, since $(\bar\alpha + 2d\bar\beta) = (\alpha+\beta)(1-\delta) > 1$, we deduce that

$$\mathbb{E}\,(\bar\zeta_n(X)) \;\geq\; \frac{(1-\delta)^n\,(\alpha+\beta)^n}{\sqrt{2\pi n}}\;C_7\,n^{-d/2} \;>\; 1$$

provided n is sufficiently large. □

Following the ideas of Lemma 5.3, Remark 5.2 and Theorem 5.1 in [3], and using Lemma 5 above in place of [3, Lemma 5.2], one proves that $\{\zeta_t\}_t$ survives when M is sufficiently large, and so does, by stochastic domination, the metapopulation when N is large and the density $p = 1$.

Lemma 6 *If $(\alpha+\beta)(1-\delta)>1$ then the process $\{\zeta_t\}_t$ survives when M is sufficiently large.*

PROOF. By additivity of $\{\bar\zeta_t\}_t$, if $X\sim 0$ then the Central Limit Theorem implies that

$$\lim_{K\to\infty}\left[P\left(\bar\zeta_n(X)\geq K \mid \bar\zeta_0(0)=K\right) -1+ \Phi\left(\frac{K-\mathbb{E}(\bar\zeta_n(X)\mid \bar\zeta_0(0)=1)\,K}{\sqrt{\mathrm{Var}(\bar\zeta_n(X)\mid \bar\zeta_0(0)=1)}\sqrt{K}}\right)\right] = 0,$$

where the function Φ is the cumulative distribution function of the standard normal. Since n is fixed, it follows that, for all $\epsilon>0$,

$$P\left(\bar\zeta_n(X)\geq K \text{ for all } X\sim 0 \mid \bar\zeta_0(0)=K\right) > 1-\epsilon$$

for K sufficiently large. Let $\{N_t\}_t$ be the branching process with birth rate $\bar\alpha+2d\bar\beta$ and death rate 0, which represents the total number of particles born up to time t. By the same argument as before, there exists $C_8>1$ such that

$$P\left(N_n\leq C_8\,K \mid N_0=K\right) \geq 1-\epsilon \quad \text{for all } K \text{ sufficiently large.}$$

Since, if $M\geq C_8\,K$ then $\{\bar\zeta_t\}_t$ and $\{\zeta_t\}_t$ coincide (up to time n) on $\{N_n\leq C_8\,K\}$, we have

$$P\left(\zeta_n(X)\geq K \mid \zeta_0(0)=K\right) > 1-2\epsilon.$$

In order to get

$$P\left(\zeta_n(X)\geq K \text{ for all } X\sim 0 \mid \zeta_0(0)=K\right) > 1-2\epsilon \tag{6}$$

we need to ensure that from time 0 to time n, in no site the process $\{\bar\zeta_t\}_t$ on $\{N_n\leq C_8\,K\}$ ever exceeds M particles. By geometric arguments (see [3, Step 3] for further details) one proves that it suffices to take $M\geq 2H_0C_8K=C_9K$ where H_0 is the number of paths of length n in $\mathbb{Z}^d$ crossing a fixed vertex. To complete the proof, we use equation (6) to couple the process $\{\zeta_t\}_t$ with a supercritical 1-dependent oriented site percolation process on $\mathbb{Z}^d\times\mathbb{Z}_+$ in a way such that the existence of an infinite cluster implies survival for $\{\zeta_t\}_t$. □

Since, in the proof of Lemma 5, we consider only the particles of generation n, equation (6) holds if, instead of the process $\{\zeta_t\}_t$, we deal with the process $\{\zeta_t^n\}_t$ obtained by deleting all the particles of generation $n'>n$. In addition, the process $\{\bar\eta_t^n\}_t$, obtained from the metapopulation model by assuming that symbionts sent outside $[n,n]^d$ are killed, clearly dominates $\{\zeta_t^n\}_t$.

Lemma 7 *Fix n so that Lemma 5 holds. Then, for all $\epsilon_0>0$,*

$$P\left(\bar\eta_n^n(X)\geq\sqrt{N} \text{ for all } X\sim 0 \mid \bar\eta_0^n(0)\geq\sqrt{N}\right) \geq 1-\epsilon_0$$

for all N sufficiently large.

PROOF. This follows from equation (6) (using $\{\bar\eta_t^n\}_t$ instead of $\{\zeta_t^n\}_t$) choosing $K=\sqrt{N}$, from stochastic domination when $C_9\sqrt{N}/\delta<N$, and from the monotonicity of $\{\bar\eta_t^n\}_t$. □

To deduce part 3 of Theorem 1 from the previous lemma, we follow the same strategy as before by comparing the evolution of the metapopulation along an infinite self-avoiding path with oriented percolation. To apply successfully Lemma 7, this infinite self-avoiding path must be contained in an infinite stripe with width at least $2n$. The existence of such a path follows by choosing the

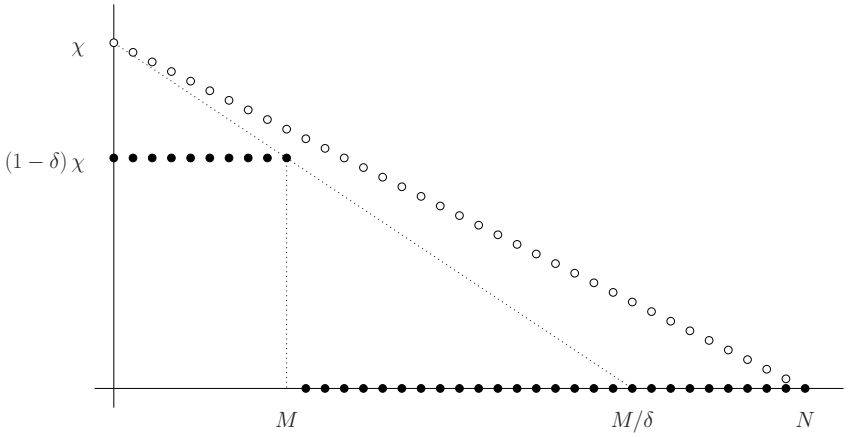


FIGURE 3. Reproduction rates for $\{\zeta_t\}_t$ and $\{\bar{\eta}_t\}_t$ (• and ∘, respectively).

parameter p close enough to 1. First, we fix n so that Lemma 5 holds (recall that n only depends on the reproduction rate α, the transmission rate β, and the spatial dimension d). Then, we fix the parameter $\epsilon_0 > 0$ such that $1 - \epsilon_0$ is greater than the critical value of n-dependent oriented percolation. We will prove part 3 when the density p of hosts is such that

$$p \;>\; \exp((2n+1)^{-d}\,\log p_c)$$

where p_c is the critical value of site percolation in d dimensions. We tile the d-dimensional regular lattice with cubes of edge length $2n+1$ by setting

$$B_0 \;=\; [-n,n]^d \qquad \text{and} \qquad B_Z \;=\; (2n+1)\,Z + B_0 \quad \text{for all } Z \in \mathbb{Z}^d.$$

Given a realization ω of the site percolation process with parameter p, we call a cube B_Z open if all the sites $X \in B_Z$ are occupied by a host, and closed otherwise. Our choice of p implies

$$P\,(B_Z \text{ is open}) \;=\; p^{(2n+1)^d} \;>\; p_c \qquad \text{for all } Z \in \mathbb{Z}^d.$$

In particular, there exists almost surely an infinite self-avoiding path of open cubes, i.e., there exists a self-avoiding path $\{Z_i : i \in \mathbb{Z}\} \subset \mathbb{Z}^d$ such that cube B_{Z_i} is open for all i. From this path of open cubes, we construct an infinite self-avoiding path of open sites $\Gamma = \{\Gamma_i : i \in \mathbb{Z}\}$ by including all the sites belonging to the straight lines connecting the centers of adjacent cubes, as shown in Figure 4 where grey squares refer to closed cubes, and white squares to open cubes. By construction,

1. For all $i \in \mathbb{Z}$ and all $X \in \Gamma_i + [-n,n]^d$, we have $X \in C_\infty(\omega)$.
2. For all $i \in \mathbb{Z}$, we have $\Gamma_i \sim \Gamma_{i+1}$.

We call site $(i,m) \in \mathcal{G}$ a good site whenever the host at site Γ_i is infected by at least $\sqrt{N}$ symbionts at time mn and as previously let G_m denote the set of good sites at level m. Then, Lemma 7 and the fact that the evolution rules of the process are homogeneous in time imply that

$$P\,((i-1,m+1) \text{ and } (i+1,m+1) \text{ are good} \mid (i,m) \text{ is good}) \;\geq\; 1-\epsilon_0 \tag{7}$$

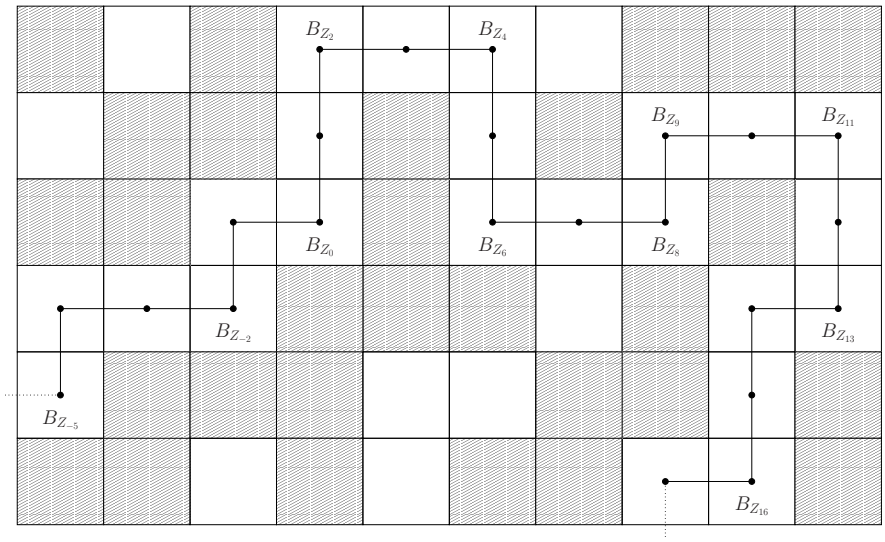


FIGURE 4. Picture of the self-avoiding path Γ.

for sufficiently large N. Denoting again by W_m the set of wet sites at level m in an n-dependent oriented site percolation process with parameter $1-\epsilon_0$ the inequality (7) implies that the processes can be constructed on the same probability space in such a way that

$$P\left(W_m \subset G_m \text{ for all } m \geq 0 \mid W_0 \subset G_0\right) \;=\; 1.$$

Since $1-\epsilon_0$ is greater than the critical value of oriented percolation, this implies as previously that the metapopulation survives, which completes the proof of Theorem 1.

4. Proof of Theorem 2

This section is devoted to the competition voter model and the proof of Theorem 2. In the neutral case, $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, the Harris' graphical representation induces a natural duality relationship between the spatial model and a system of coalescing random walks on the connected random graph $C_N(\omega)$. In particular, the first part of Theorem 2 will follow from certain collision properties of symmetric random walks on the infinite percolation cluster $C_\infty(\omega)$. In contrast, the analysis of the competition model in the non-neutral case will be performed without invoking duality but rather by comparing the forward evolution with a gambler's ruin model.

Duality with coalescing random walks. To define the dual process of the competition model under neutrality, we first construct the process graphically from collections of independent Poisson processes using an idea of Harris [12]. Each vertex $x \in C_N(\omega)$ is equipped with a Poisson process with parameter 1. Poisson processes attached to different vertices are independent. At the arrival times of the process at x, we toss a coin with success probability $\alpha/(\alpha+\beta)$ where α is the common reproduction parameter of both symbiont types and β the common transmission parameter. If there is a success, we choose a vertex uniformly at random from the host at site $\pi(x)$ and draw an arrow from this vertex to vertex x. If there is a failure, we choose a vertex uniformly at random from one

of the hosts adjacent to site $\pi(x)$ and draw an arrow from this vertex to vertex x. In view of the geometry of the graph and the number of vertices per host, this is equivalent to saying that

- For any pair of vertices $x, y \in C_N(\omega)$ with $x \updownarrow y$, we draw an arrow from y to x at the arrival times of an independent Poisson process with parameter $\alpha/(N(\alpha+\beta))$.
- For any pair of vertices $x, y \in C_N(\omega)$ with $x \leftrightarrow y$, we draw an arrow from y to x at the arrival times of an independent Poisson process with parameter $\beta/(N \deg \pi(x)(\alpha+\beta))$.

In any case, an arrow from vertex y to vertex x indicates that the symbiont at x dies and gets instantaneously replaced by a symbiont of the same species as the symbiont at y.

To define the dual process, we say that there is a path from $(y, T-s)$ to (x, T), which corresponds to a dual path from (x, T) to $(y, T-s)$, if there are sequences of times and vertices

$$s_0 \;=\; T-s \;<\; s_1 \;<\; \cdots \;<\; s_{n+1} \;=\; T \qquad \text{and} \qquad x_0 \;=\; y,\, x_1,\, \ldots,\, x_n \;=\; x$$

such that the following two conditions hold:

1. For $i = 1, 2, \ldots, n$, there is an arrow from x_{i-1} to x_i at time s_i and
2. For $i = 0, 1, \ldots, n$, there is no arrow that points at the segments $\{x_i\} \times (s_i, s_{i+1})$.

The dual process starting at (x, T) is the process defined by

$$\hat{\xi}_s(x, T) \;=\; \{y \in C_N(\omega) : \text{there is a dual path from } (x, T) \text{ to } (y, T-s)\}.$$

The dual process starting from a finite set of vertices $B \subset C_N(\omega)$ can be defined as well. In this case, the dual process starting at (B, T) is the set-valued process defined by

$$\begin{aligned} \hat{\xi}_s(B, T) \;&=\; \{y \in C_N(\omega) : \text{there is a dual path} \\ &\qquad\qquad \text{from } (x, T) \text{ to } (y, T-s) \text{ for some } x \in B\} \\ &=\; \{y \in C_N(\omega) : y \in \hat{\xi}_s(x, T) \text{ for some } x \in B\}. \end{aligned}$$

The dual process is naturally defined only for $0 \le s \le T$. However, it is convenient to assume that the Poisson processes in the graphical representation are defined for negative times so that the dual process can be defined for all $s \ge 0$. Note that, in view of the graphical representation of the competition model in the neutral case, the dual process starting at (x, T) performs a continuous-time random walk on the random graph $C_N(\omega)$ that makes transitions

$$y \;\to\; \begin{cases} z \;\text{ for } z \updownarrow y \quad \text{at rate} \quad \alpha/(N(\alpha+\beta)) \\ z \;\text{ for } z \leftrightarrow y \quad \text{at rate} \quad \beta/(N \deg \pi(y)(\alpha+\beta)). \end{cases}$$

The dual process starting from a finite set $B \subset C_N(\omega)$ consists of a system of $\mathrm{card}(B)$ such random walks, one random walk starting from each vertex in the set B. Any two of these random walks evolve independently of each other until they intersect when they coalesce. This induces a duality relationship between the model and coalescing random walks. We refer the reader to Figure 5 for an example of realization of the dual process.

The reason for introducing the dual process is that it allows us to deduce the configuration of the system at the current time based on the configuration at earlier times, but also how vertices at the current time are correlated, by keeping track of the ancestry of each symbiont. In particular, the

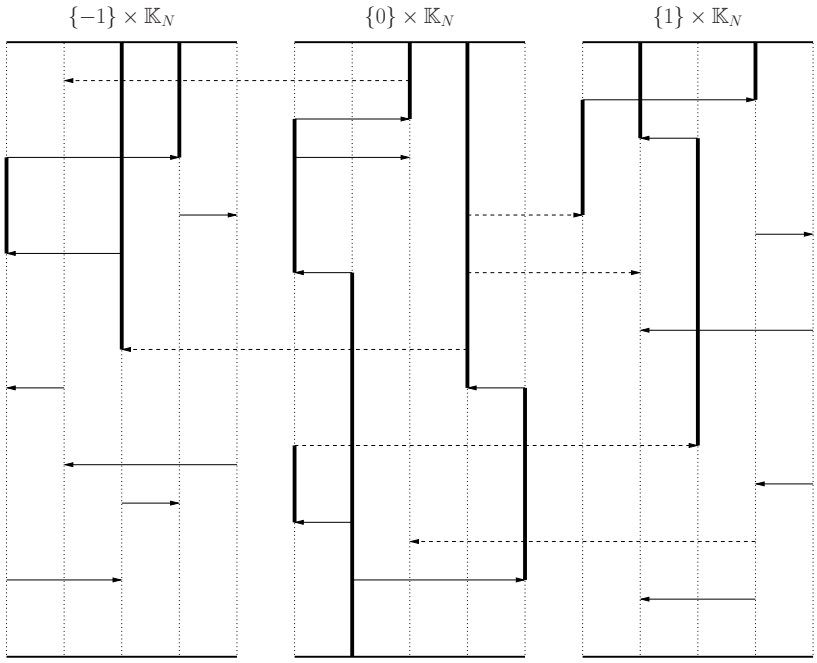


FIGURE 5. Example of realization of the dual process ($d = 1$ and $N = 5$) in bold lines, where we have drawn arrows within the same host in continuous line, and arrows connecting two adjacent hosts in dashed lines.

long-term behavior of the competition model (clustering versus coexistence) can be expressed in terms of collision properties of random walks on the infinite percolation cluster through the duality relationship between the model and coalescing random walks. We now explain this connection in details, starting with some key definitions.

Let $G = (V, E)$ be an infinite connected graph. We call simple symmetric random walk on this graph the continuous-time Markov process $\{X_t\}_t$ with state space V that jumps from u to v at rate one if and only if $(u, v) \in E$. Note that the embedded Markov chain associated to this Markov process is the discrete-time random walk $\{\mathcal{X}_n\}_n$ with transition probabilities

$$P\left(\mathcal{X}_{n+1} = v \mid \mathcal{X}_n = u\right) \;=\; \frac{1}{\deg(u)} \qquad \text{if and only if } (u, v) \in E.$$

Since the graph G is connected, the process $\{X_t\}_t$ is irreducible so either all the vertices of the graph are recurrent, in which case the graph is said to be recurrent, or all the vertices are transient, in which case the graph is said to be transient. Let X_t and Y_t be two independent random walks on the graph G. Using again the fact that the graph is connected and the Kolmogorov zero-one law, the probability that the two random walks intersect infinitely often, namely

$$P\left(\text{for all } t \text{ there exists } s \text{ such that } X_{t+s} = Y_{t+s}\right)$$

is either equal to 0 or 1 regardless of the initial positions of the random walks. The graph G is said to have the infinite collision property if the previous probability is equal to 1, and it is said to have the finite collision property if the previous probability is equal to 0. Such properties for the infinite percolation cluster $C_\infty(\omega)$ translate through the duality relationship with coalescing random walks into coexistence/clustering of the competition model, as shown in the next lemma.

Lemma 8 *We have the following alternative:*

1. *$C_\infty(\omega)$ has the infinite collision property and then the process clusters, or*
2. *$C_\infty(\omega)$ has the finite collision property and then coexistence occurs.*

Proof. Let $B \subset C_N(\omega)$ be finite and let $\Theta_t^i = \{x \in C_N(\omega) : \xi_t(x) = i\}$ denote the set of vertices occupied by a type i symbiont at time t. By duality,

$$P\left(\Theta_t^1 \cap B = \varnothing\right) \;=\; \mathbb{E}\,(1-\theta)^{|\hat{\xi}_t(B,t)|} \tag{8}$$

where θ is the initial density of type 1. Since the number of particles in $\hat{\xi}_t(B,t)$ is a nonincreasing function of t and has a limit, the bounded convergence theorem implies that the probability on the left-hand side of (8) also has a limit as $t \to \infty$. It follows that the process converges to a stationary distribution. To understand how different vertices are correlated under this stationary distribution, we take two vertices $x, y \in C_N(\omega)$, $x \neq y$, and consider the projections

$$X_s \;=\; \pi(\hat{\xi}_s(x,T)) \qquad \text{and} \qquad Y_s \;=\; \pi(\hat{\xi}_s(y,T)).$$

Let τ be the hitting time of the dual processes, i.e.,

$$\tau \;=\; \inf\{s > 0 : \hat{\xi}_s(x,T) = \hat{\xi}_s(y,T)\}.$$

Note that the processes X_s and Y_s evolve individually according to continuous-time random walks run at rate $q := \beta/(\alpha+\beta)$ on the infinite percolation cluster $C_\infty(\omega)$. They evolve independently of each other until time τ when they coalesce. We set $t_0 = 0$ and define inductively

$$\begin{aligned} s_i &= \inf\{s > t_{i-1} : X_s = Y_s\} \\ t_i &= \inf\{s > s_i : X_s \neq Y_s\} \;=\; \inf\{s > s_i : X_s \neq X_{s_i} \text{ or } Y_s \neq Y_{s_i}\} \end{aligned}$$

for $i \geq 1$. Note that, if the dual processes coalesce at time s then

$$t_i \;=\; s_{i+1} \;=\; \infty \quad \text{for all} \quad i \geq \max\{j : s_j < s\}.$$

Also, let M_i denote the total number of jumps during the interval of time (s_i, t_i) of either of the dual processes starting at vertex x or vertex y. Writing P_i for the conditional probability given the event that $s_i < \infty$ and using that each dual process jumps to one of the adjacent hosts at rate q and within each host at rate $1-q$, we obtain the following probability

$$\begin{aligned} P_i\,(\tau > t_i) \;&=\; \sum_{j=0}^{\infty} P_i\,(\tau > t_i \mid M_i = j)\; P_i\,(M_i = j) \\ &=\; \sum_{j=0}^{\infty} \left(1 - \frac{1}{N}\right)^j q\,(1-q)^j \;=\; \frac{q}{1-(1-q)(1-1/N)} \;=\; \frac{N\beta}{\alpha + N\beta}. \end{aligned} \tag{9}$$

Let $J = J(x,y,\omega) = \sup\{j : s_j < \infty\}$, and note that, on the event that the dual processes starting at x and y do not coalesce, J is equal in distribution to the number of intersections of two independent random walks on the infinite percolation cluster starting at $\pi(x)$ and $\pi(y)$. In

particular, if $C_\infty(\omega)$ has the infinite collision property and if I_t denotes the number of intersections up to time t of two independent random walks starting at $\pi(x)$ and $\pi(y)$ then (9) implies

$$\begin{aligned}\lim_{T\to\infty} P\left(\xi_T(x) \neq \xi_T(y)\right) &\leq \lim_{T\to\infty} P\left(\hat{\xi}_T(x,T) \neq \hat{\xi}_T(y,T)\right) \\ &= \lim_{T\to\infty} \sum_{j=0}^{\infty} \prod_{i=0}^{j} P\left(I_T = j\right) P_i\left(\tau > t_i\right) \\ &= \lim_{T\to\infty} \sum_{j=0}^{\infty} \left(\frac{N\beta}{\alpha+N\beta}\right)^j P\left(I_T = j\right) \;=\; 0\end{aligned}$$

by the bounded convergence theorem since $P\left(I_T = j\right) \to 0$ as $T \to \infty$ for all $j \in \mathbb{N}$. This shows that the process clusters. Alternatively, if the infinite percolation cluster has the finite collision property, then J is almost surely finite so equation (9) implies that

$$\begin{aligned}\lim_{T\to\infty} P\left(\xi_T(x) \neq \xi_T(y)\right) &= \theta(1-\theta) \lim_{T\to\infty} P\left(\hat{\xi}_T(x,T) \neq \hat{\xi}_T(y,T)\right) \\ &= \theta(1-\theta) \sum_{j=0}^{\infty} \lim_{T\to\infty} P\left(\hat{\xi}_T(x,T) \neq \hat{\xi}_T(y,T) \mid J = j\right) P\left(J = j\right) \\ &= \theta(1-\theta) \sum_{j=0}^{\infty} \left(\frac{N\beta}{\alpha+N\beta}\right)^j P\left(J = j\right) \;\geq\; c \;>\; 0,\end{aligned}$$

which shows that coexistence occurs. This completes the proof. □

Note that the previous lemma easily extends to any connected graph in which the degree of each vertex is uniformly bounded. That is, given such a graph $G = (V, E)$, the competition model can be naturally defined on the graph G_N with vertex set $V \times \mathbb{K}_N$ constructed from G in the same way as the graph $C_N(\omega)$ is constructed from the infinite percolation cluster. Then, the proof of the previous lemma implies that, for all N finite, the resulting process clusters when G has the infinite collision property, but coexists when G has the finite collision property.

Transience of the percolation cluster and coexistence. Motivated by Lemma 8, we now prove that the infinite percolation cluster $C_\infty(\omega)$ in dimensions $d \geq 3$ has the finite collision property. This follows from the fact that the infinite cluster is transient, a result due to Grimmett, Kesten and Zhang [11], and that the degree of each vertex is uniformly bounded. We also answer the same questions for the infinite percolation clusters in 2 dimensions since the proofs are similar, even through these results will not be used to establish clustering of the process.

Lemma 9 *The cluster $C_\infty(\omega)$ is recurrent in $d = 2$ and transient in $d \geq 3$.*

PROOF. The proof is based on the analogy between random walks and electrical networks. We refer the reader to [7] for more details about this analogy. The idea is to turn the infinite percolation cluster into an electrical network in which each edge has unit resistance. Then, simple random walks on the original graph are recurrent if and only if the effective resistance of the resulting electrical network between a given point and the points at infinity is infinite, as shown in [7]. To deal with the two-dimensional case, we first use Polya's theorem to deduce that the effective resistance of the

regular lattice $\mathbb{Z}^2$ between a nominated point and the points at infinity is infinite. By Rayleigh's monotonicity law, this implies that the effective resistance of any subgraph of $\mathbb{Z}^2$ between a given point and the points at infinity is again infinite, from which we deduce that the infinite percolation cluster in 2 dimensions is recurrent. The $d \geq 3$ case has been studied in details by Grimmett, Kesten and Zhang [11]. Their proof relies on the construction of a tree-like graph that can be embedded in the infinite percolation cluster and whose effective resistance between a given point and the points at infinity is finite. A new application of Rayleigh's monotonicity law implies that the infinite percolation cluster is transient in $d \geq 3$. The result in [11] applies to bond percolation but relies on geometric properties that are known for site percolation as well. □

Lemma 10 *Let $\{X_t\}_t$ and $\{Y_t\}_t$ be two independent random walks run at rate 1 on $C_\infty(\omega)$ both starting at vertex A, and denote by $I(X,Y)$ the number of their intersections. Then,*

$$\mathbb{E}\, I(X,Y) = \infty \text{ in } d = 2 \qquad \text{and} \qquad \mathbb{E}\, I(X,Y) < \infty \text{ in } d \geq 3.$$

Proof. Since the total rate of jump of both random walks equals 2,

$$\begin{aligned} \mathbb{E}\, I(X,Y) \;&=\; 2\,\mathbb{E}\left(\int_0^\infty \mathbb{1}\{X_t = Y_t\}\, dt\right) \;=\; 2\int_0^\infty P\left(X_t = Y_t\right) dt \\ &=\; 2\int_0^\infty \sum_{B \in C_\infty(\omega)} P\left(X_t = B\right) P\left(Y_t = B\right) dt \;=\; 2\int_0^\infty \sum_{B \in C_\infty(\omega)} \left(p_t(A,B)\right)^2 dt \end{aligned} \tag{10}$$

where $p_t(A,B) = P\left(X_t = B \mid X_0 = A\right)$. Now, we observe that the probability that a random walk follows a given directed path from vertex A to vertex B is equal to 1 divided by the product of the degrees of the vertices of this path excluding the final vertex B. Similarly, the probability that a random walk follows the reverse path from vertex B to vertex A is 1 divided by the product of the degrees of the vertices excluding the final vertex A, from which we deduce that

$$(2d)^{-1}\, p_t(B,A) \;\leq\; p_t(A,B) \;\leq\; 2d\, p_t(B,A) \quad \text{for all } B \in C_\infty(\omega) \tag{11}$$

since $1 \leq \deg(A), \deg(B) \leq 2d$. Therefore, when $C_\infty(\omega)$ is recurrent, (10)-(11) imply that

$$\mathbb{E}\, I(X,Y) \;\geq\; d^{-1} \int_0^\infty \sum_{B \in C_\infty(\omega)} p_t(A,B)\, p_t(B,A)\, dt \;=\; d^{-1} \int_0^\infty p_{2t}(A,A)\, dt \;=\; \infty$$

whereas when $C_\infty(\omega)$ is transient, (10)-(11) imply that

$$\mathbb{E}\, I(X,Y) \;\leq\; 4d \int_0^\infty \sum_{B \in C_\infty(\omega)} p_t(A,B)\, p_t(B,A)\, dt \;=\; 4d \int_0^\infty p_{2t}(A,A)\, dt \;<\; \infty.$$

The result then follows from Lemma 9. □

Lemma 10 indicates that $P\left(I(X,Y) < \infty\right) = 1$ in dimensions $d \geq 3$, i.e., $C_\infty(\omega)$ has the finite collision property, which, together with Lemma 8, implies that coexistence occurs. However, that the expected number of intersections is infinite does not imply that the number of intersections is infinite with positive probability (with probability 1 by the Kolmogorov zero-one law). In fact, it is known that recurrent graphs, even with bounded degree, do not necessarily have the infinite collision property. This has been proved by Krishnapur and Peres [16], looking at the comb lattice, that is the subgraph of $\mathbb{Z}^2$ obtained by deleting all the horizontal edges off the x-axis.

Infinite collision property of the percolation cluster. We now prove that the infinite percolation cluster $C_\infty(\omega)$ has the infinite collision property in $d = 2$, which, by Lemma 8, is equivalent to clustering of the neutral competition model in two dimensions. We use the same notations as before and let $\{X_t\}_t$ and $\{Y_t\}_t$ be two independent continuous-time random walks run at rate 1 on the infinite percolation cluster. Let $W_t = (X_t, Y_t)$ and $\mathcal{W}_n = (\mathcal{X}_n, \mathcal{Y}_n)$ denote the discrete-time Markov chain on $C_\infty(\omega) \times C_\infty(\omega)$ with transition probabilities

$$P\left(\mathcal{W}_{n+1} = (A', B') \mid \mathcal{W}_n = (A, B)\right) \;=\; \frac{1}{2}\left(q_1(A, A')\,\mathbb{1}\{B = B'\} + q_1(B, B')\,\mathbb{1}\{A = A'\}\right)$$

where $q_n(A, B)$ denotes the n-step transition probability of the lazy symmetric random walk on the infinite percolation cluster. That is, at each time step, one of the two coordinates of $\mathcal{W}_n$ is chosen at random with probability $1/2$. This coordinate then moves according to the uniform distribution on the neighbors or stands still, both with probability $1/2$, while the other coordinate does not change. Note that, at each step, with probability $1/2$, the process $\mathcal{W}_n$ does not move at all. Note also that the processes $\{W_t\}_t$ and $\{\mathcal{W}_n\}_n$ can be coupled in such a way that the sequences of states visited by both processes are equal. In particular, invoking in addition the Markov property, to prove the infinite collision property of the graph, it suffices to prove that

$$P\left(\mathcal{X}_n = \mathcal{Y}_n \text{ for some } n \geq 1 \mid \mathcal{W}_0 = (A, B)\right) \;=\; 1.$$

The first key to proving the infinite collision property of the cluster is the following theorem, which is the analog of Theorem 1 in [1]. We state the result in the general d-dimensional case, though we only deal with the 2-dimensional case in the rest of this section.

Theorem 11 *Let $p > p_c$. Then, there exist a subset Ω of the set of the realizations with probability one and a collection of random variables $\{S_A\}_{A \in \mathbb{Z}^d}$ such that the following holds.*

1. *We have $S_A(\omega) < \infty$ for each $\omega \in \Omega$ and $A \in C_\infty(\omega)$.*
2. *There are constants $c_1, c_2, c_3, c_4 > 0$ such that, for all $A, B \in C_\infty(\omega)$,*

$$\begin{array}{rcll} q_n(A, B) & \geq & c_1\, n^{-d/2}\, \exp(-c_2\, |A - B|^2/n) & \text{whenever} \quad |A - B| \vee S_A(\omega) \leq n \\ q_n(A, B) & \leq & c_3\, n^{-d/2}\, \exp(-c_4\, |A - B|^2/n) & \text{whenever} \quad S_A(\omega) \leq n. \end{array} \tag{12}$$

The proof of Theorem 11 follows the lines of the proof of its analog in [1] and only differs in two points: first, we consider a discrete-time lazy random walk instead of a continuous-time random walk, and second, processes under consideration evolve on the infinite percolation cluster of site percolation instead of bond percolation. To prove the subgaussian upper estimate, the idea is to use a discrete-time version of [20, Theorem 1.1] and the results of [6, Sections 5,6 and 8], while the proof of the subgaussian lower estimate follows closely the strategy of [1]. Note that the choice of a lazy random walk is motivated by the fact that one cannot expect the lower bound to hold for any time n for a standard simple random walk. This is due to the fact that it has period 2. In order to avoid unnecessary complications, we prefer to deal with an aperiodic random walk.

In the sequel, to simplify the notations, we write sums starting from (or ending at) possibly noninteger real numbers, but it is tacitly understood that one must consider their integer part. To

prove the infinite collision property, we define

$$F(n) \;=\; \sum_{j=0}^{n} 2^{-n} \binom{n}{j} q_j(A,X)\, q_{n-j}(B,X)$$

$$F_\rho(n) \;=\; \sum_{j=\rho n}^{(1-\rho)n} 2^{-n} \binom{n}{j} q_j(A,X)\, q_{n-j}(B,X)$$

where $\rho \in (0,1/2)$ and $A, B, X \in C_\infty(\omega)$.

Lemma 12 *Fix $\rho \in (0,1/2)$, $A, B, X \in C_\infty(\omega)$ and $\epsilon > 0$. Then $F(n) \le (1+\epsilon)\, F_\rho(n)$ when n is sufficiently large depending on ρ, A, B, X and ϵ.*

Proof. By the Hoeffding inequality [14, Theorem 1], we have

$$F(n) - F_\rho(n) \;\le\; \sum_{j=0}^{\rho n} 2^{-n} \binom{n}{j} \;+\; \sum_{j=(1-\rho)n}^{n} 2^{-n} \binom{n}{j} \;\le\; 2\, \exp(-2n(1/2-\rho)^2).$$

Taking n such that

$$\rho n \;\ge\; \sqrt{n} \vee S_A(\omega) \vee S_B(\omega) \;\ge\; |X-A| \vee |X-B| \vee S_A(\omega) \vee S_B(\omega),$$

we may use the first inequality in (12). Letting Φ denote the cumulative distribution function of the standard normal, and also applying the Central Limit Theorem, we obtain

$$\begin{aligned}
F_\rho(n) \;&\ge\; \sum_{j=\rho n}^{(1-\rho)n} 2^{-n} \binom{n}{j} \frac{c_1}{j} \exp\left(-\frac{c_2\,|X-A|^2}{j}\right) \frac{c_1}{n-j} \exp\left(-\frac{c_2\,|X-B|^2}{n-j}\right) \\
&\ge\; \sum_{k=\rho n}^{(1-\rho)n} 2^{-n} \binom{n}{j} \frac{c_1^2}{j(n-j)} \exp\left(-\frac{c_2\, n^2}{j(n-j)}\right) \qquad (13) \\
&\ge\; (2c_1/n)^2 \exp(-c_2/((1-\rho)\,\rho)) \sum_{j=\rho n}^{(1-\rho)n} 2^{-n} \binom{n}{j} \\
&\ge\; C_{10}\, n^{-2}\, (2\,\Phi((1-2\rho)\sqrt{n}) - 1)
\end{aligned}$$

for some $C_{10} < \infty$. To conclude, observe that

$$F(n) \;=\; \left(1 + \frac{F(n) - F_\rho(n)}{F_\rho(n)}\right) F_\rho(n)$$

while the previous estimates (13) imply

$$\lim_{n\to\infty} \frac{F(n) - F_\rho(n)}{F_\rho(n)} \;\le\; \lim_{n\to\infty} \frac{n^2}{C_{10}} \frac{\exp(-2n(1/2-\rho)^2)}{\Phi((1-2\rho)\sqrt{n}) - 1/2} \;=\; 0.$$

This completes the proof. □

With Theorem 11 and Lemma 12 in hands, we are now ready to prove that the infinite percolation cluster has the infinite collision property in the sense described above, i.e., considering continuous-time random walks run at a constant rate, say 1. Our proof relies in addition on an argument of Peres and Sousi [24] who studied the number of collisions of discrete-time random walks moving simultaneously at each time step. In order to understand the duality properties of the competition model, we need, in contrast, to consider a pair of random walks in which only one walk chosen uniformly at random can move while the other walk stands still, thus mimicking the evolution of a pair of independent continuous-time random walks.

Theorem 13 *Fix a realization ω. Then, for all $A, B \in C_\infty(\omega)$,*

$$P\left(\mathrm{card}\{n : \mathcal{X}_n = \mathcal{Y}_n\} = \infty \mid \mathcal{W}_0 = (A, B)\right) \;=\; 1.$$

PROOF. Let $\gamma > 0$ to be chosen later, and define

$$I_k \;=\; \sum_{n=k}^{k^2} \sum_{\substack{|X-A|\vee|X-B|<\sqrt{n} \\ S_X(\omega)\le\gamma}} I(X,n)$$

where $I(X,n) = 1$ if there is a collision at time n at site X, and $= 0$ otherwise. The first step is to find bounds for the first and second moments of I_k when k is large.

Lower bound: $\mathbb{E}(I_k) \ge C \log k$ for some constant $C > 0$ which does not depend on A, B and for all $k \ge k_1(A,B)$. First, we fix $\rho \in (0, 1/2)$ and observe that

$$\begin{aligned}\mathbb{E}^{A,B} I(X,n) \;&=\; P^{A,B}(\mathcal{X}_n = \mathcal{Y}_n = X) \;=\; \sum_{j=0}^{n} 2^{-n} \binom{n}{j} q_j(A,X)\, q_{n-j}(B,X) \\ &\ge \sum_{j=\rho n}^{(1-\rho)n} 2^{-n} \binom{n}{j} q_j(A,X)\, q_{n-j}(B,X).\end{aligned}$$

In the previous sum, j and $n-j$ are larger than ρn. Hence, for $n \ge k$, if

$$|X-A| \vee |X-B| < \sqrt{\rho n} \qquad \text{and} \qquad k \ge \rho^{-1}(S_A(\omega) \vee S_B(\omega))$$

then $j \wedge (n-j) \ge |X-A| \vee |X-B| \vee S_A(\omega) \vee S_B(\omega)$ so Theorem 11 implies

$$\begin{aligned}\mathbb{E}^{A,B}(I_k) \;&\ge\; \sum_{n=k}^{k^2} \sum_{\substack{|X-A|\vee|X-B|<\sqrt{\rho n} \\ S_X(\omega)\le\gamma}} \sum_{j=\rho n}^{(1-\rho)n} 2^{-n} \binom{n}{j} \frac{c_1^2}{j(n-j)} \exp\left[-c_2\left(\frac{\rho n}{j} + \frac{\rho n}{n-j}\right)\right] \\ &\ge\; \sum_{n=k}^{k^2} \sum_{\substack{|X-A|\vee|X-B|<\sqrt{\rho n} \\ S_X(\omega)\le\gamma}} \sum_{j=\rho n}^{(1-\rho)n} 2^{-n} \binom{n}{j} \left(\frac{2c_1}{n}\right)^2 \exp(-c_2/(1-\rho)).\end{aligned}$$

This and the Central Limit Theorem imply that, for k large depending on A, B,

$$\begin{aligned}\mathbb{E}^{A,B}(I_k) \;&\geq\; \exp(-c_2/(1-\rho)) \\ &\times \sum_{n=k}^{k^2} (2c_1/n)^2 \;\mathrm{card}\,\{X \in C_\infty(\omega) : |X-A| \vee |X-B| < \sqrt{\rho n},\; S_X(\omega) \leq \gamma\}.\end{aligned} \tag{14}$$

Now, by the ergodic theorem,

$$\begin{aligned}\lim_{n\to\infty} \frac{\mathrm{card}\,\{X \in C_\infty(\omega) : |X-A| \vee |X-B| < \sqrt{\rho n},\; S_X(\omega) \leq \gamma\}}{\mathrm{card}\,\{X \in \mathbb{Z}^2 : |X-A| \vee |X-B| < \sqrt{\rho n}\}} \\ = \; P\,(X \in C_\infty(\omega),\; S_X(\omega) \leq \gamma).\end{aligned}$$

In particular, there exists a constant $\delta > 0$ that only depends on the percolation parameter p such that for all γ and k sufficiently large, we have

$$\mathrm{card}\,\{X \in C_\infty(\omega) : |X-A| \vee |X-B| < \sqrt{\rho n},\, S_X(\omega) \leq \gamma\} \;\geq\; \delta\rho n \tag{15}$$

for $n \geq k$. By (14)-(15), there exists $k_1(A,B)$ large such that

$$\mathbb{E}^{A,B}(I_k) \;\geq\; c_1^2 \;\exp(-c_2/(1-\rho)) \sum_{n=k}^{k^2} \frac{\delta\rho}{n} \;\geq\; C_{11}\,(\log k^2 - \log k) \;=\; C_{11}\,\log k \tag{16}$$

for a suitable $C_{11} > 0$ not depending on A, B, and all $k \geq k_1(A,B)$.

Upper bound: $\mathbb{E}(I_k^2) \leq C\,(\log k)^2$ for some constant $C < \infty$ which does not depend on A, B and for all $k \geq k_2(A,B)$. First, we observe that, for $l \geq n$,

$$\begin{aligned}\mathbb{E}^{A,B}(I(X,n)\,I(Y,l)) \;&=\; P^{A,B}(\mathcal{X}_n = \mathcal{Y}_n = X,\; \mathcal{X}_l = \mathcal{Y}_l = Y) \\ &=\; \mathbb{E}^{A,B} I(X,n)\;\mathbb{E}^{X,X}(Y, l-n)\end{aligned}$$

from which it follows that

$$\mathbb{E}^{A,B}(I_k^2) \;\leq\; 2\sum_{n=k}^{k^2}\sum_{l=n}^{k^2} \sum_{\substack{|X-A|\vee|X-B|<\sqrt{n} \\ S_X(\omega)\leq\gamma}} \sum_{\substack{Y:|X-Y|<\sqrt{l-n} \\ S_Y(\omega)\leq\gamma}} \mathbb{E}^{A,B} I(X,n)\;\mathbb{E}^{X,X} I(Y,l-n).$$

Since $I(Y_1, l-n)\,I(Y_2, l-n) = 0$ whenever $Y_1 \neq Y_2$, we also have

$$\sum_{Y:|X-Y|<\sqrt{l-n}} \mathbb{E}^{X,X} I(Y,l-n)\;\mathbb{1}\{S_Y(\omega) \leq \gamma\} \;\leq\; 1.$$

Therefore, by applying Lemma 12 twice with $\epsilon = 1$, we deduce that there exists γ large such that for all k sufficiently large

$$\mathbb{E}^{A,B}(I_k^2) \;\le\; 2\sum_{n=k}^{k^2}\sum_{\substack{|X-A|\vee|X-B|<\sqrt{n}\\ S_X(\omega)\le\gamma}} 2\left(\sum_{j=\rho n}^{(1-\rho)n} 2^{-n}\binom{n}{j}\, q_j(A,X)\, q_{n-j}(B,X)\right)$$
$$\times\left(\frac{\gamma}{\rho} \;+\; 2\sum_{l=n+\gamma/\rho}^{k^2}\sum_{\substack{Y:|X-Y|<\sqrt{l-n}\\ S_Y(\omega)\le\gamma}}\sum_{i=\rho(l-n)}^{(1-\rho)(l-n)} 2^{-(l-n)}\binom{l-n}{i}\, q_i(X,Y)\, q_{l-n-i}(X,Y)\right).$$

Observing that in the sums over j and i above, we have

$$\begin{aligned} j\wedge(n-j) \;&\ge\; \rho n \;\ge\; \rho k \;\ge\; S_A(\omega)\vee S_B(\omega)\\ i\wedge(l-n-i) \;&\ge\; \rho(l-n) \;\ge\; \rho\gamma/\rho \;=\; \gamma \;\ge S_X(\omega), \end{aligned}$$

for all k large depending on A, B, X, Theorem 11 implies that

$$\mathbb{E}^{A,B}(I_k^2) \;\le\; 8\sum_{n=k}^{k^2} 5n\left(\sum_{j=\rho n}^{(1-\rho)n} 2^{-n}\binom{n}{j}\,\frac{c_3^2}{j(n-j)}\,\exp(-c_4\,n\,(j^{-1}+(n-j)^{-1})\right)$$
$$\times\left(\frac{\gamma}{\rho} \;+\; \sum_{l=n+\gamma/\rho}^{k^2}\sum_{m=0}^{\infty}\mathrm{card}\,\{Y : m\sqrt{l-n}\le|X-Y|<(m+1)\sqrt{l-n}\}\right.$$
$$\left.\times\sum_{i=\rho(l-n)}^{(1-\rho)(l-n)} 2^{-(l-n)}\binom{l-n}{i}\,\frac{c_3^2}{i(l-n-i)}\,\exp(-c_4 m^2\,(i^{-1}+(l-n-i)^{-1})\right).$$

In particular, there exists $k_2(A,B)$ large such that

$$\begin{aligned} \mathbb{E}^{A,B}(I_k^2) \;&\le\; 8\sum_{n=k}^{k^2}\frac{5\,c_3^2}{\rho(1-\rho)n}\\ &\times\left(\frac{\gamma}{\rho} \;+\; \sum_{l=n+\gamma/\rho}^{k^2}\sum_{m=0}^{\infty}\frac{5\,(m+2)^2\,c_3^2}{\rho(1-\rho)(l-n)}\,\exp(-4\,c_4\,(m+1)^2)\right) \qquad (17)\\ &\le\; C_{12}\sum_{n=k}^{k^2}\frac{1}{n}\left(\frac{\gamma}{\rho} \;+\; C_{13}\sum_{l=n+\gamma/\rho}^{k^2}\frac{1}{l-n}\right) \;\le\; C_{14}\,(\log k)^2 \end{aligned}$$

for suitable constants $C_{12}, C_{13}, C_{14} < \infty$ not depending on A, B and all $k \ge k_2(A,B)$.

Let $k(A,B) = k_1(A,B)\vee k_2(A,B)$. By (16)-(17) and the Paley-Zygmund inequality,

$$\begin{aligned} P\,(I_k > (C_{11}/2)\,\log k \mid \mathcal{W}_0 = (A,B)) \;&\ge\; P\,(I_k > \mathbb{E}(I_k)/2 \mid \mathcal{W}_0 = (A,B))\\ &\ge\; (\mathbb{E}^{A,B}(I_k))^2/4\,\mathbb{E}^{A,B}(I_k^2) \;\ge\; C_{11}^2/4\,C_{14} \;=\; c \;>\; 0 \end{aligned}$$

for $k = k(A, B)$ and where, as C_{11} and C_{14}, the constant $c > 0$ does not depend on the starting points of the random walks. Then, we define a sequence of stopping times and sites as follows: we start at $n_0 = 0$ and $(A_0, B_0) = (A, B)$, and for all $j \geq 1$ we define inductively

$$n_j \;=\; n_{j-1} + k(A_{j-1}, B_{j-1}) \qquad \text{and} \qquad (A_j, B_j) \;=\; (\mathcal{X}_{n_j}, \mathcal{Y}_{n_j}).$$

We say that there is a success at round $j \geq 1$ when

$$\text{card}\,\{n \in [n_{j-1}, n_j) : \mathcal{X}_n = \mathcal{Y}_n\} \;>\; (C_{11}/2)\,\log k(A_{j-1}, B_{j-1})$$

and observe that, at each round, the success probability is larger than $c > 0$. In particular, the probability mass function of the number of successes up to round $j \geq 1$ is stochastically larger than a Binomial random variable with parameters j and $c > 0$, from which it follows that the ultimate number of successes, thus the ultimate number of collisions, is almost surely infinite. □

As previously explained, clustering of the neutral competition model in two dimensions follows from the combination of Lemma 8 and Theorem 13.

Competition with selection. We now prove the second part of Theorem 2 which, in contrast with the first part, relies on an analysis of the evolution forward in time rather than backwards in time. For all $X \in \mathbb{Z}^d$, we let $\bar{\xi}_t(X)$ be the number of type 1 symbionts in the host at X and set

$$\begin{aligned} p_t(X) \;&=\; \frac{2d\,\alpha_1\,\bar{\xi}_t(X) + \beta_1 \sum_{Y\sim X}\,\bar{\xi}_t(Y)}{2d\,\alpha_1\,\bar{\xi}_t(X) + 2d\,\alpha_2\,(N - \bar{\xi}_t(X)) + \beta_1 \sum_{Y\sim X}\,\bar{\xi}_t(Y) + \beta_2 \sum_{Y\sim X}\,(N - \bar{\xi}_t(Y))} \\ q_t(X) \;&=\; \frac{2d\,\alpha_1\,\bar{\xi}_t(X) + \beta_1 \sum_{Y\sim X}\,\bar{\xi}_t(Y)}{2d\,N\,(\alpha_1 + \beta_1)}. \end{aligned}$$

Observe that $p_t(X) = q_t(X)$ in the neutral case $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, and that

$$p_t(X),\, q_t(X) \;=\; \begin{cases} 0 & \text{if and only if} \quad \bar{\xi}_t(X) = \bar{\xi}_t(Y) = 0 \quad \text{for all } Y \sim X \\ 1 & \text{if and only if} \quad \bar{\xi}_t(X) = \bar{\xi}_t(Y) = N \quad \text{for all } Y \sim X. \end{cases}$$

Note also that, since $q_t(X)$ can take at most $(N + 1)(2dN + 1)$ different values,

$$q^- \;:=\; \inf\{q_t(X) : q_t(X) \in (0, 1)\} \;>\; 0 \qquad \text{and} \qquad q^+ \;:=\; \sup\{q_t(X) : q_t(X) \in (0, 1)\} \;<\; 1.$$

Let N_t denote the number of type 1 symbionts present in the system at time t. If the number of symbionts of type 1 at time 0 is finite, then

$$N_t \;\to\; \begin{cases} N_t + 1 & \text{at rate} \quad \sum_X\,(N - \bar{\xi}_t(X))\,p_t(X) \\ N_t - 1 & \text{at rate} \quad \sum_X\,\bar{\xi}_t(X)\,(1 - p_t(X)) \end{cases}$$

where the sum is over all $X \in \mathbb{Z}^d$ such that $p_t(X) \in (0, 1)$. Now, we observe that in the neutral case when $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, the embedded Markov chain associated to the process $\{N_t\}_t$ is the simple symmetric random walk on $\mathbb{Z}_+$ absorbed at 0, therefore

$$\frac{1}{N} \sum_X\,\bar{\xi}_t(X) \;=\; \sum_X\,p_t(X).$$

This implies that for all α_1 and β_1 such that $\alpha_1 + \beta_1 \neq 0$, and for all configurations

$$\frac{1}{N} \sum_X \bar{\xi}_t(X) \;=\; \sum_X \frac{2d\,\alpha_1\,\bar{\xi}_t(X) + \beta_1 \sum_{Y\sim X} \bar{\xi}_t(Y)}{2d\,N\,(\alpha_1+\beta_1)} \;=\; \sum_X q_t(X). \tag{18}$$

Now, assume that $\alpha_1 \geq \alpha_2$ and $\beta_1 > \beta_2$. We want to show that in this case the embedded Markov chain associated to $\{N_t\}_t$ is a random walk with a positive drift. We say that

1. site $X \in \mathbb{Z}^d$ is bad at time t when $q_t(X) \in (0,1)$ and $\bar{\xi}_t(Y) = N$ for all $Y \sim X$ and
2. site $X \in \mathbb{Z}^d$ is good at time t when $q_t(X) \in (0,1)$ and $\bar{\xi}_t(Y) \neq N$ for some $Y \sim X$.

Note that if N_0 is finite then at any time t the sets of good and bad sites are both finite. The first ingredient to proving the result is to observe that for any site either good or bad

$$\begin{aligned} p_t(X) \;&=\; \frac{2d\,\alpha_1\,\bar{\xi}_t(X) + \beta_1 \sum_{Y\sim X} \bar{\xi}_t(Y)}{2dN\,\alpha_1 + 2dN\,\beta_1 - (\alpha_1-\alpha_2)\,(N - \bar{\xi}_t(X)) - (\beta_1-\beta_2)\sum_{Y\sim X}\,(N-\bar{\xi}_t(Y))} \\ &\geq\; \frac{2d\,\alpha_1\,\bar{\xi}_t(X) + \beta_1 \sum_{Y\sim X} \bar{\xi}_t(Y)}{2dN\,(\alpha_1+\beta_1)} \;=\; q_t(X) \end{aligned} \tag{19}$$

while if we assume in addition that X is a good site, then

$$p_t(X) \;\geq\; \frac{2d\,\alpha_1\,\bar{\xi}_t(X) + \beta_1 \sum_{Y\sim X} \bar{\xi}_t(Y)}{2dN\,(\alpha_1+\beta_1)} \left[1 - \frac{\beta_1-\beta_2}{2dN\,(\alpha_1+\beta_1)}\right]^{-1} \;=\; q_t(X)\,(1-c)^{-1} \tag{20}$$

where $c = (\beta_1-\beta_2)/(2dN\,(\alpha_1+\beta_1)) \in (0,1)$. The second ingredient is to observe that, by definition of the lower bound q^- and upper bound q^+, we have $q^-\,q_t(X_1) \leq q^+\,q_t(X_2)$ for all X_1, X_2 so

$$\begin{aligned} q^-\,(1-c)^{-1}\,q_t(X_1) \;+\; q^+\,q_t(X_2) \;&\leq\; q^-\,q_t(X_1) \;+\; q^+\,(1-c)^{-1}\,q_t(X_2) \\ (q^-\,(1-c)^{-1} + q^+)\,(q_t(X_1) + q_t(X_2)) \;&\leq\; (q^- + q^+)\,(q_t(X_1) + (1-c)^{-1}\,q_t(X_2)). \end{aligned} \tag{21}$$

The third ingredient is to observe also that if X is bad then all sites $Y \sim X$ are good, therefore the number of bad sites is at most equal to the number of good sites:

$$\mathrm{card}\,\{X : X \text{ is bad}\} \;\leq\; \mathrm{card}\,\{X : X \text{ is good}\}. \tag{22}$$

Combining (18)-(22), we obtain

$$\begin{aligned} \sum_X p_t(X) \;&=\; \sum_{X\,\mathrm{bad}} p_t(X) \;+\; \sum_{X\,\mathrm{good}} p_t(X) \\ &\geq\; \sum_{X\,\mathrm{bad}} q_t(X) \;+\; (1-c)^{-1} \sum_{X\,\mathrm{good}} q_t(X) \\ &\geq\; \frac{q^-\,(1-c)^{-1} + q^+}{q^- + q^+} \;\times\; \sum_X q_t(X) \\ &=\; \frac{q^-\,(1-c)^{-1} + q^+}{q^- + q^+} \;\times\; \frac{1}{N} \sum_X \bar{\xi}_t(X). \end{aligned}$$

In particular, we have

$$\begin{aligned}\sum_X (N - \bar{\xi}_t(X))\, p_t(X) &\geq \frac{q^-(1-c)^{-1}+q^+}{q^-+q^+} \times \sum_X \bar{\xi}_t(X) - \sum_X \bar{\xi}_t(X)\, p_t(X) \\ &\geq \frac{q^-(1-c)^{-1}+q^+}{q^-+q^+} \times \sum_X \bar{\xi}_t(X)\,(1-p_t(X)).\end{aligned}$$

By comparing the previous inequality with the transition rates of $\{N_t\}_t$ and applying the gambler's ruin formula, we can conclude that, starting with K symbionts of type 1,

$$\begin{aligned}&P\left(\lim_{t\to\infty} \xi_t(x) = 1 \text{ for all } x \in \mathbb{Z}^d \times \mathbb{K}_N\right) \\ &\qquad = P\left(\lim_{t\to\infty} N_t = \infty\right) \geq 1 - \left(\frac{q^-+q^+}{q^-(1-c)^{-1}+q^+}\right)^K.\end{aligned}$$

This completes the proof of Theorem 2.

Acknowledgment. The authors would like to thank Martin Barlow, Yuval Peres and Perla Sousi for fruitful discussions about random walks on random graphs.

References

[1] Barlow, M. (2004). Random walks on supercritical percolation clusters. *Ann. Probab.* **32** 3024–3084.
[2] Belhadji, L., Bertacchi, D. and Zucca, F. (2009). A self-regulating and patch subdivided population. arXiv:0811.1279.
[3] Bertacchi, D., Zucca, F. (2009). Approximating critical parameters of branching random walks. *J. Appl. Probab.* **46** 463–478.
[4] Bramson, M. and Durrett, R. (1988). A simple proof of the stability criterion of Gray and Griffeath. *Probab. Theory Related Fields* **80** 293–298.
[5] Clifford, P. and Sudbury, A. (1973). A model for spatial conflict. *Biometrika* **60** 581–588.
[6] Coulhon, T., Grigor'yan, A. and Zucca, F. (2005). The discrete integral maximum principle and its applications. *Tohoku Math. J.* **57** 559–587.
[7] Doyle, P. G. and Snell, E. J. (1984). *Random walks and electrical networks.* Carus Math. Monographs **22**, Math. Assoc. Amer,. Washington, D. C.
[8] Durrett, R. (1984). Oriented percolation in two dimensions. *Ann. Probab.* **12** 999–1040.
[9] Durrett, R. and Lanchier, N. (2008). Coexistence in host-pathogen systems. *Stochastic Process. Appl.* **118** 1004–1021.
[10] Grimmett, G. R. (1989). *Percolation.* Springer, Berlin.
[11] Grimmett, G. R., Kesten, H. and Zhang, Y. (1993). Random walk on the infinite cluster of the percolation model. *Probab. Theory Related Fields* **96** 33–44.
[12] Harris, T. E. (1972). Nearest neighbor Markov interaction processes on multidimensional lattices. *Adv. Math.* **9** 66–89.
[13] Harris, T. E. (1974). Contact interactions on a lattice. *Ann. Probability*, 2:969–988.
[14] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[15] Holley, R. A. and Liggett, T. M. (1975). Ergodic theorems for weakly interacting systems and the voter model. *Ann. Probab.* **3** 643–663.
[16] Krishnapur, M. and Peres, Y. (2008). Recurrent graphs where two independent random walks collide finitely often. *Elect. Comm. in Probab.* **9** 72–81.
[17] Lanchier, N. and Neuhauser, C. (2006a). Stochastic spatial models of host-pathogen and host-mutualist interactions I. *Ann. Appl. Probab.* **16** 448–474.
[18] Lanchier, N. and Neuhauser, C. (2006b). A spatially explicit model for competition among specialists and generalists in a heterogeneous environment. *Ann. Appl. Probab.* **16** 1385–1410.
[19] Lanchier, N. and Neuhauser, C. (2009). Stochastic spatial models of host-pathogen and host-mutualist interactions II. *Preprint.*
[20] Mathieu, P. and Remy, E. (2004). Isoperimetry and heat kernel decay on percolation clusters. *Ann. Prob.* **32** 100–128.
[21] Mooring, M. S. and Samuel, W, M. (1998). Tick-removal grooming by elk (*Cervus elaphus*): testing the principles of the programmed-grooming hypothesis. *Canadian Journal of Zoology* **76** 740–750.
[22] Moran, P. A. P. (1958). Random processes in genetics. *Proc. Cambridge Philos. Soc.* **54** 60–71.
[23] Neuhauser, C. (1992). Ergodic theorems for the multitype contact process. *Prob. Theory Relat. Fields* **91** 467–506.
[24] Peres, Y. and Sousi, P. (2009). *In progress.*
[25] Price, P. W. (1980). *Evolutionary Biology of Parasites.* Princeton University Press, Princeton.
[26] Trappe, J. M. (1987). Phylogenetic and ecologic aspects of mycotrophy in the angiosperms from an evolutionary standpoint. In: G.R. Safir (ed) *Ecophysiology of VA Mycorrhizal Plants.* CRC Press, Boca Ranton, pp. 5-25.
[27] Woess, W. (2000). *Random walks on infinite graphs and groups.* Cambridge Tracts in Mathematics, **138**, Cambridge Univ. Press, 2000.

UNIVERSITÀ DI MILANO-BICOCCA
DIPARTIMENTO DI MATEMATICA E APPLICAZIONI,
VIA COZZI 53, 20125 MILANO, ITALY.

SCHOOL OF MATHEMATICAL AND STATISTICAL SCIENCES,
ARIZONA STATE UNIVERSITY,
TEMPE, AZ 85287, USA.

DIPARTIMENTO DI MATEMATICA,
POLITECNICO DI MILANO,
PIAZZA LEONARDO DA VINCI 32, 20133 MILANO, ITALY.