

# CONTACT AND VOTER PROCESSES ON THE INFINITE PERCOLATION CLUSTER AS MODELS OF HOST-SYMBIONT INTERACTIONS

BY D. BERTACCHI, N. LANCHIER<sup>1</sup> AND F. ZUCCA

*Università di Milano-Bicocca, Arizona State University and Politecnico di Milano*

We introduce spatially explicit stochastic processes to model multi-species 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 contrast 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 parasitism, commensalism, and mutualism, respectively. The degree of specificity of the symbiont is another important factor: while some symbionts may live in association with a wide range of host species, in which case the symbiont is called a generalist, others are highly host-specific indicating that they can only benefit from few host species. Symbiotic relationships, either pathogenic or mutualistic, are ubiquitous in nature. For instance, more than 90% of terrestrial plants [28] 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 [27].

To understand the role of spatial structure on the persistence of host-parasite and host-mutualist associations, Lanchier and Neuhauser [21–23] have initiated

---

Received November 2009; revised September 2010.

<sup>1</sup>Supported in part by NSF Grant DMS-10-05282.

*AMS 2000 subject classification.* 60K35.

*Key words and phrases.* Contact process, voter model, site percolation, logistic growth, branching random walks, random walks, host, symbiont, infrapopulation, metapopulation, infracommunity, component community.

the study of multispecies host-symbiont systems including local interactions based on interacting particle systems. The stochastic process introduced in [22] 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 [21, 23] 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 [13]. 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 in a system of coupled interacting particle systems, each describing the evolution of a trophic level. The model is designed for the understanding of the role of the symbionts in the spatial structure of plant communities. It is proved theoretically that generalist symbionts have only a limited effect on the spatial structure of their habitat [23]. In contrast, the inclusion of specialist parasites promotes coexistence of the hosts in terms of the existence of a stationary distribution under which the density of each host type is positive, 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 [13].

Similarly to most spatial epidemic models such as the contact process, the state space of the stochastic processes introduced in [13, 21–23] 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, that is, symbionts living on their hosts or in their skin, associated to a single host individual may vary significantly. Mooring and Samuel [25] found for instance an average of 1791 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 [14]. 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 [17]. In the multispecies model, we will assume that infracommunities evolve according to the Moran model [26], and the entire component community according to a mixture of this model and its spatial analog, the voter model [8, 19]. 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 [14]. 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 particle systems: the contact process [17] and the voter model [8, 19]. The structure of the random graph implies that the infrapopulation dynamics are described by logistic growth processes, that is, contact processes on a complete graph, and the infracommunity dynamics by Moran models, that is, voter models on a complete graph.

*Host environment.* To define the habitat of hosts, we set  $p \in (0, 1]$  and let  $\omega$  be a realization of the site percolation process with parameter  $p$  on the  $d$ -dimensional regular lattice  $\mathbb{Z}^d$ , that is, 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  $X = X_0, X_1, \dots, X_n = Y$  such that the following two conditions hold:

1. For  $i = 0, 1, \dots, n$ , we have  $X_i \in \mathbb{H}(\omega)$ , that is, site  $X_i$  is open.
2. For  $i = 0, 1, \dots, n - 1$ , 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. For more details about site percolation, we refer the reader to Grimmett [14].

*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  $\mathbb{K}_N = \{1, 2, \dots, N\}$ . The vertex set of  $\mathcal{H}(\omega)$  is

$$C_N(\omega) = \{(X, i) : X \in C_\infty(\omega) \text{ and } i \in \mathbb{K}_N\}.$$

By convention, elements of and processes with state space  $C_N(\omega)$  will be denoted by small Latin letters. To define the edge set, we also introduce

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

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 [17] 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  $\eta : C_N(\omega) \rightarrow \{0, 1\}$ , and whose dynamics are described by the Markov generator  $L_1$  defined on the set of the cylinder functions by

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

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  $\eta$  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  $\alpha$ , and transmit their offspring to the nearby hosts at the transmission rate  $\beta$ . That is, each symbiont gives birth at rate  $\alpha$  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  $\beta$  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. See Figure 1 for simulation pictures of this contact process. To study the single-species model, we will sometimes consider the stochastic process

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

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

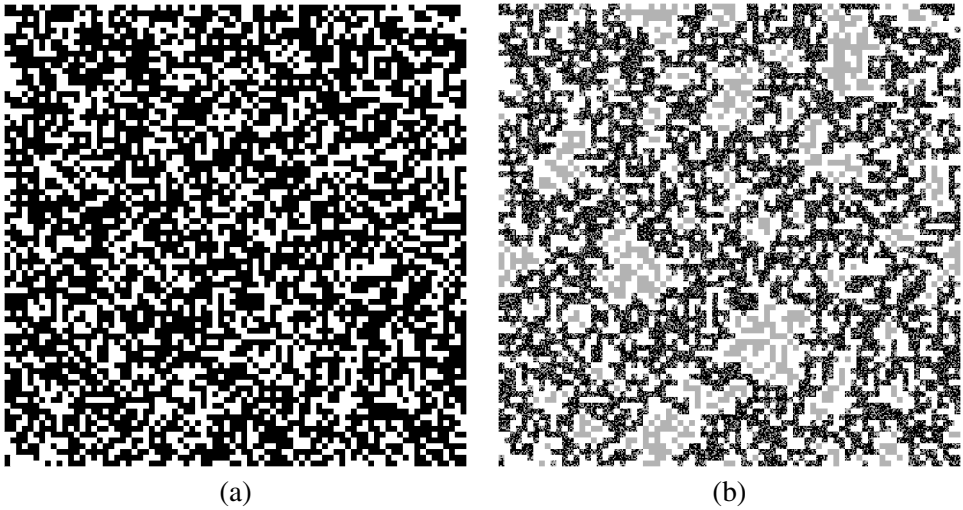


FIG. 1. (a) Realization of site percolation with parameter  $p = 0.6$  on the  $80 \times 80$  torus, with black squares referring to open sites, and white squares to closed sites. (b) Snapshot of the invasion model on the percolation structure starting with a single infected host at the center of the universe ( $\alpha = 1$  and  $\beta = 2$ ). Each site is represented by a  $5 \times 5$  square, that is, complete graph with  $N = 25$  vertices. White squares refer to empty sites, that is, sites which are not occupied by a host, black dots refer to symbionts, and gray dots to empty vertices.

whose state space consists of the functions that map  $C_\infty(\omega)$  into  $\{0, 1, \dots, N\}$  and whose dynamics are described by

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

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  $\beta$  on the infinite percolation cluster and logistic growth processes with parameter  $\alpha$ .

To describe the predictions based on the invasion model, we let  $\delta_i$  be the measure that concentrates on the “all  $i$ ” configuration restricted to  $C_N(\omega)$ , that is,

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

We denote by  $\bar{\mu}$  the upper invariant measure of the process  $\{\eta_t\}_t$ , which is also the limit starting from the measure  $\delta_1$  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  $\alpha$  becomes irrelevant, and compare the process with the one-dimensional contact process. Let  $\Gamma$  be an arbitrary infinite self-avoiding path in the infinite percolation cluster  $C_\infty(\omega)$ . Since for all sites  $X \in \Gamma$  the degree of  $X$  ranges from 2 to  $2d$ , the process restricted to the infinite path  $\Gamma$ , that is, symbionts sent outside  $\Gamma$  are instantaneously killed, dominates stochastically the contact process on  $\Gamma$  with infection parameter  $\beta/d$ . 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  $\mathbb{Z}$ . Standard coupling arguments also imply that the survival probability of the infection is nondecreasing with respect to both the reproduction rate  $\alpha$  and the maximum number of symbionts per host  $N$ . It follows directly from these monotonicity properties that, for all values of  $N$  and  $\alpha$ , 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 [25]. 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  $\beta$  is small, a situation which is common in parasitic relationships, the metapopulation undergoes a phase transition when the reproduction rate  $\alpha$  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 where “survival” means strong survival of the stochastic process, that is the existence of a stationary distribution under which the density of symbionts is positive.

**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  $\alpha + \beta > 1$  and  $p$  is close to 1, then the metapopulation survives for  $N$  large.*
3. *If  $\alpha + \beta/d > 1$  and  $p > p_c$ , 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 a self-avoiding path embedded in the infinite percolation cluster. The proof of the second part relies on the combination of random walk estimates and block constructions to compare the process view under suitable space and time scales with oriented percolation, and we refer to Section 3 for more details. Survival when  $\alpha > 1$  and  $p > p_c$  in the presence of large infrapopulations can be proved based on estimates for the extinction time of the logistic growth process and a new block construction. However, the third part, which indicates survival under the weaker assumption  $\alpha + \beta/d > 1$ , can be directly deduced from the proof of the second part by again looking at the process restricted to an infinite self-avoiding path of hosts. Let  $\Gamma$  be an infinite self-avoiding path, which exists almost surely under the assumption  $p > p_c$ , and observe that, since the degree of each site along this path ranges from 2 to  $2d$ , the process restricted to  $\Gamma$  dominates stochastically the one-dimensional process with parameters  $\alpha$  and  $\beta/d$ . The latter survives if  $\alpha + \beta/d > 1$  since, under this assumption, the proof of the second part indicates that survival occurs when  $p = 1$  in any dimension, including  $d = 1$ . The third part of the theorem clearly follows.

*Competition among multiple symbionts—voter model.* To study the interactions among multiple symbiont species, we introduce the analog of the previous model replacing contact interactions with voter interactions [8, 19]. The state at time  $t$  is now  $\xi_t : C_N(\omega) \rightarrow \{1, 2\}$ , that is, each vertex is occupied by a symbiont of one of two types. Letting for  $i = 1, 2$

$$f_i(x) = \text{card}\{y : \pi(y) = \pi(x) \text{ and } \xi_t(y) = i\} / N,$$

$$g_i(x) = \text{card}\{y : \pi(y) \sim \pi(x) \text{ and } \xi_t(y) = i\} / (N \deg \pi(x))$$

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

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

$$+ \sum_{x \in C_N(\omega)} \frac{\alpha_2 f_2(x) + \beta_2 g_2(x)}{\alpha_1 f_1(x) + \alpha_2 f_2(x) + \beta_1 g_1(x) + \beta_2 g_2(x)} [f(\xi_{x,2}) - f(\xi)],$$

where  $\xi_{x,i}$  is the configuration obtained from  $\xi$  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 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  $\alpha$  and the transmission rates are both equal to say  $\beta$ , the local evolution



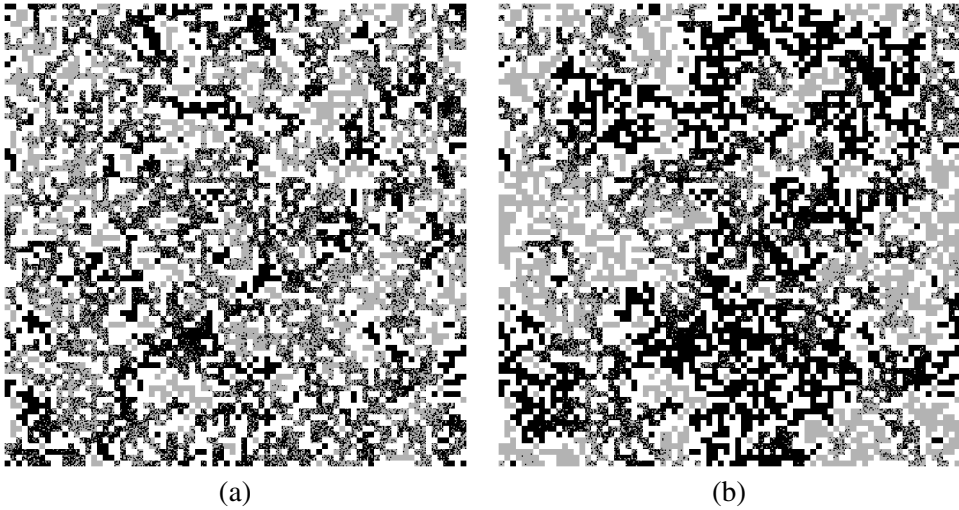


FIG. 2. Snapshots at time 100 (a) and at time 1000 (b), respectively, of the neutral competition model on site percolation with parameter  $p = 0.6$  starting from a Bernoulli product measure with density  $1/2$  ( $\alpha_1 = \alpha_2 = 1/2$  and  $\beta_1 = \beta_2 = 1/2$ ). Each site of the lattice is represented by a  $5 \times 5$  square, that is, complete graph with  $N = 25$  vertices. White squares refer to empty sites, that is, sites which are not occupied by a host, and black and gray dots to symbionts of type 1 and 2, respectively.

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  $\beta/(\alpha + \beta)$ . Note that the process  $\{\xi_t\}_t$  can again be seen as a mixture of two well-known processes, namely, the Moran model with selection, and its spatial analog, the biased voter model [6, 7] on the infinite percolation cluster. See Figure 2 for simulation pictures of this voter model.

To state our results for the competition model, we set  $\theta \in (0, 1)$  and denote by  $\pi_\theta$  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  $\{\xi_t\}_t$  starts from the product measure  $\pi_\theta$  and let  $\Rightarrow$  stand for convergence in distribution. The process is said to cluster if

$$\text{there exists } a \in (0, 1) \quad \text{such that } \xi_t \Rightarrow a\delta_1 + (1 - a)\delta_2 \quad \text{as } t \rightarrow \infty.$$

In particular, we have

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

The process is said to coexist if in contrast  $\xi_t \Rightarrow \nu_\theta$  as  $t \rightarrow \infty$  for some  $\nu_\theta$  such that

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

Type 1 is said to invade type 2 if

$$P\left(\lim_{t \rightarrow \infty} N_t = \infty \mid N_0 = 1\right) > 0 \quad \text{where } N_t = \text{card}\{x \in C_N(\omega) : \xi_t(x) = 1\},$$

indicating that, starting with a single symbiont of type 1 in the infinite percolation cluster, there is a positive probability that the number of type 1 keeps growing indefinitely. Finally, type 1 is said to outcompete type 2 whenever we have the stronger condition  $\xi_t \Rightarrow \delta_1$ .

**THEOREM 2** (Voter interactions). *Assume that  $p > p_c$ . If  $\alpha_1 = \alpha_2$  and  $\beta_1 = \beta_2$ , the component community clusters in two dimensions, whereas coexistence occurs in higher dimensions.*

The analysis of the neutral competition model relies on duality techniques. We show, in the neutral case, that the process is dual to a certain system of coalescing random walks evolving on the random graph induced by the infinite percolation cluster. The long-term behavior of the process is related to the so-called finite/infinite collision property of the graph, which is studied in details in two dimensions and higher dimensions separately in Section 4.

**THEOREM 3** (Voter interactions with selection). *Assume that  $p = 1$ .*

1. *If  $\alpha_1 \geq \alpha_2$  and  $\beta_1 > \beta_2$ , then type 1 invades type 2.*
2. *If  $\alpha_2 = 0$  and  $\beta_1 > \beta_2$ , then type 1 outcompetes type 2.*

Note that, in contrast with the neutral case, when one symbiont type has a selective advantage, the transition rates of the competition model are no longer linear functions of the local frequencies of symbiont types. This leads a priori to a failure of duality techniques. The first part of the theorem is established by comparing the number of type 1 symbionts with a gambler's ruin model whereas the second part relies on the analysis of a certain system of branching coalescing random walks which is dual to a particle system related to the model with selection when  $\alpha_2 = 0$ . Before going into the proofs, we note that, though the condition  $\alpha_2 = 0$  may appear biologically unrealistic, the second part obviously holds when  $N = 1$  since in this case the value of the reproduction parameters becomes unimportant. In particular, the result in the second part is relevant for species in which only one symbiont individual can associate with a host individual. Finally, we point out that the techniques to prove Theorem 3 also apply to the case when  $p > p_c$  but lead to conditions on the parameters which are far from being optimal and to very tedious calculations that only make the key ideas unclear. Therefore, for simplicity, we focus on the case  $p = 1$  only.

**3. Proof of Theorem 1.** This section is devoted to the analysis of the invasion model, and more precisely to the proof of the second part of Theorem 1. The key

idea is to show that the branching random walk restricted to a large square persist an arbitrary long time provided  $\alpha + \beta > 1$  and  $N$  is large. The combination of our estimates with a block construction implies survival of the metapopulation restricted to an infinite self-avoiding path of large squares fully occupied by hosts. It is also proved that such a path exists whenever the parameter  $p$  is close enough to 1.

The first step is to prove branching random walk estimates in order to establish the result in any dimension when  $p = 1$ , that is, the infinite percolation cluster consists of the entire lattice. 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

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

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

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

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  $Y$ , and the parameter  $\chi$  is equal to  $\alpha$  if  $X = Y$  and  $\beta$  if  $X \sim Y$ .

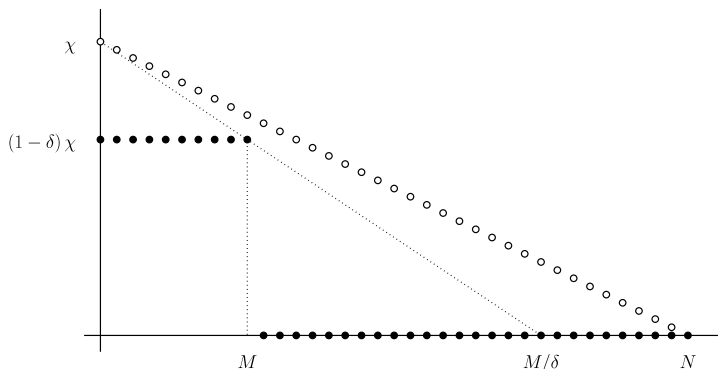


FIG. 3. Reproduction rates for  $\{\zeta_t\}_t$  and  $\{\bar{\eta}_t\}_t$  (• and o, resp.).

Let  $\delta > 0$  such that  $(\alpha + \beta)(1 - \delta) > 1$ . We will prove, following [4], 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

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

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

LEMMA 4. *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 [3], 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

$$(1) \quad \mathbb{E}(\bar{\zeta}_t(X)) = \sum_{n=0}^{\infty} \sum_{k=0}^{n-1} \mu^{(n,k)}(0, X) \frac{\bar{\alpha}^k \bar{\beta}^{n-k} t^n}{n!} e^{-t},$$

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 (1), 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  $X \sim 0$ ,

$$\begin{aligned} \sum_{k=0}^{n-1} \mu^{(n,k)}(0, X) \frac{\bar{\alpha}^k \bar{\beta}^{n-k}}{(\bar{\alpha} + 2d\bar{\beta})^n} &= P(U_n = X \mid U_0 = 0) \\ &\geq C_1 n^{-d/2} \end{aligned}$$

for a suitable  $C_1 = C_1(\bar{\alpha}, \bar{\beta}) > 0$ . We refer to [29], Corollary 13.11, for the asymptotic estimates of the  $n$ -step probabilities. In particular, for  $X \sim 0$  and  $t = n$ , we

obtain

$$\begin{aligned} \mathbb{E}(\bar{\zeta}_n(X)) &\geq \sum_{k=0}^{n-1} \mu^{(n,k)}(0, X) \frac{\bar{\alpha}^k \bar{\beta}^{n-k} n^n}{n!} e^{-n} \\ &= \sum_{k=0}^{n-1} \mu^{(n,k)}(0, X) \frac{\bar{\alpha}^k \bar{\beta}^{n-k}}{(\bar{\alpha} + 2d\bar{\beta})^n} \frac{n^n (\bar{\alpha} + 2d\bar{\beta})^n}{n!} e^{-n} \\ &\stackrel{n \rightarrow \infty}{\sim} \frac{(\bar{\alpha} + 2d\bar{\beta})^n}{\sqrt{2\pi n}} \sum_{k=0}^{n-1} \mu^{(n,k)}(0, X) \frac{\bar{\alpha}^k \bar{\beta}^{n-k}}{(\bar{\alpha} + 2d\bar{\beta})^n} \\ &\geq \frac{(\bar{\alpha} + 2d\bar{\beta})^n}{\sqrt{2\pi n}} C_2 n^{-d/2} \end{aligned}$$

for a suitable  $C_2 > 0$ . Finally, since  $(\bar{\alpha} + 2d\bar{\beta}) = (\alpha + \beta)(1 - \delta) > 1$ , we deduce that

$$\mathbb{E}(\bar{\zeta}_n(X)) \geq \frac{(1 - \delta)^n (\alpha + \beta)^n}{\sqrt{2\pi n}} C_2 n^{-d/2} > 1$$

provided  $n$  is sufficiently large.  $\square$

Following the ideas of Lemma 5.3, Remark 5.2 and Theorem 5.1 in [4], and using Lemma 4 above in place of [4], 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  $p = 1$ . These ideas are developed in more details in the following lemma.

LEMMA 5. *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

$$\lim_{K \rightarrow \infty} \left[ P(\bar{\zeta}_n(X) \geq K \mid \bar{\zeta}_0(0) = K) - 1 + \Phi\left(\frac{K - \mathbb{E}(\bar{\zeta}_n(X) \mid \bar{\zeta}_0(0) = 1)K}{\sqrt{\text{Var}(\bar{\zeta}_n(X) \mid \bar{\zeta}_0(0) = 1)}\sqrt{K}}\right) \right] = 0,$$

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

$$P(\bar{\zeta}_n(X) \geq K \text{ for all } X \sim 0 \mid \bar{\zeta}_0(0) = K) > 1 - \varepsilon$$

for  $K$  sufficiently large. Let  $\{N_t\}_t$  be the branching process with birth rate  $\bar{\alpha} + 2d\bar{\beta}$  and death rate zero, which represents the total number of particles born up to time  $t$ . By the same argument as before, there exists  $C_3 > 1$  such that

$$P(N_n \leq C_3 K \mid N_0 = K) \geq 1 - \varepsilon \quad \text{for all } K \text{ sufficiently large.}$$

Since, if  $M \geq C_3K$  then  $\{\bar{\zeta}_t\}_t$  and  $\{\zeta_t\}_t$  coincide (up to time  $n$ ) on  $\{N_n \leq C_3K\}$ , we have

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

In order to get

$$(2) \quad P(\zeta_n(X) \geq K \text{ for all } X \sim 0 \mid \zeta_0(0) = K) > 1 - 2\varepsilon$$

we need to ensure that from time 0 to time  $n$ , in no site the process  $\{\bar{\zeta}_t\}_t$  on  $\{N_n \leq C_3K\}$  ever exceeds  $M$  particles. By geometric arguments (see [4], Step 3, for further details), one proves that it suffices to take  $M \geq 2H_0C_3K = C_4K$  where  $H_0$  is the number of paths of length  $n$  in  $\mathbb{Z}^d$  crossing a fixed vertex. To complete the proof, we couple the process  $\{\zeta_t\}_t$  with a supercritical 1-dependent oriented site percolation process on  $\mathbb{Z} \times \mathbb{Z}_+$  in a way such that the existence of an infinite cluster implies survival for  $\{\zeta_t\}_t$  relying on the standard rescaling technique introduced in [5]. Let

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

and declare site  $(z, m) \in \mathcal{G}$  to be good if the host at site  $ze_1$  is infected by at least  $K$  symbionts at time  $m \times n$ , where  $e_1$  denotes the first unit vector. Also, let

$$\mathcal{G}_m = \{z \in \mathbb{Z} : (z, m) \text{ is a good site}\}$$

denote the set of good sites at level  $m$ . Then, inequality (2) above implies that  $\mathcal{G}_m$  dominates stochastically 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  $1 - \varepsilon$  and with initial condition  $W_0 \subset \mathcal{G}_0$  (see Durrett [11] for a complete description of oriented percolation). The result then follows by choosing  $\varepsilon > 0$  sufficiently small to make the oriented percolation process supercritical.  $\square$

Since, in the proof of Lemma 4, we consider only the particles of generation  $n$ , equation (2) 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 6. *Fix  $n$  so that Lemma 4 holds. Then, for all  $\varepsilon > 0$ ,*

$$P(\bar{\eta}_n^n(X) \geq \sqrt{N} \text{ for all } X \sim 0 \mid \bar{\eta}_0^n(0) \geq \sqrt{N}) \geq 1 - \varepsilon$$

*for all  $N$  sufficiently large.*

PROOF. This follows from (2) (using  $\{\bar{\eta}_t^n\}_t$  instead of  $\{\zeta_t^n\}_t$ ) choosing  $K = \sqrt{N}$ , from stochastic domination when  $C_4\sqrt{N}/\delta < N$ , and from the monotonicity of  $\{\bar{\eta}_t^n\}_t$ .  $\square$

To deduce the second part of Theorem 1 from the previous lemma, we use another block construction in order to compare the evolution of the metapopulation along an infinite self-avoiding path with oriented percolation. To apply successfully Lemma 6, any site within distance  $n$  of this infinite self-avoiding path must be open. The existence of such a path follows by choosing the parameter  $p$  close enough to 1. First, we fix  $n$  so that Lemma 4 holds (recall that  $n$  only depends on the reproduction rate  $\alpha$ , the transmission rate  $\beta$ , and the spatial dimension  $d$ ). Then, we fix the parameter  $\varepsilon > 0$  such that  $1 - \varepsilon$  is greater than the critical value of  $n$ -dependent oriented percolation. We prove the result when the density  $p$  of hosts satisfies

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

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

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

Given a realization  $\omega$  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

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

In particular, there exists almost surely an infinite self-avoiding path of open cubes, that is, 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  $\Gamma = \{\Gamma_z : z \in \mathbb{Z}\}$  by including all the sites belonging to the straight lines connecting the centers of adjacent cubes, as shown in Figure 4 where gray squares refer to closed cubes, and white squares to open cubes. By construction:

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

Site  $(z, m) \in \mathcal{G}$  is now said to be good whenever the host at  $\Gamma_z$  is infected by at least  $\sqrt{N}$  symbionts at time  $m \times n$ . As previously, we let  $\mathcal{G}_m$  denote the set of good sites at level  $m$ . Then Lemma 6 and the fact that the evolution rules of the process are homogeneous in time imply that

$$(3) \quad P((z - 1, m + 1) \text{ and } (z + 1, m + 1) \text{ are good} \mid (z, m) \text{ is good}) \geq 1 - \varepsilon$$

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  $1 - \varepsilon$  the inequality (3) implies that the processes can be constructed on the same probability space in such a way that

$$P(W_m \subset \mathcal{G}_m \text{ for all } m \geq 0 \mid W_0 \subset \mathcal{G}_0) = 1.$$

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

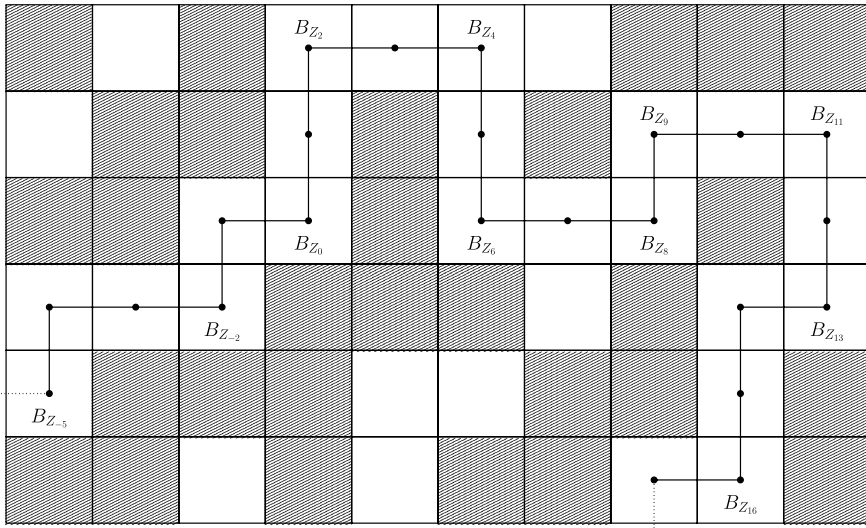


FIG. 4. Picture of the self-avoiding path  $\Gamma$ .

**4. Proof of Theorem 2.** This section is devoted to the analysis of the competition model under neutrality. The process can be constructed graphically relying on an idea of Harris [16] from a collection of independent Poisson processes. In the neutral case, because the transition rates are linear functions of the local frequencies, the graphical representation induces a natural duality relationship between the spatial model and a system of coalescing random walks on  $C_N(\omega)$ , and Theorem 2 follows from certain collision properties of symmetric random walks on the infinite percolation cluster.

*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 [16]. 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  $\alpha$  is the common reproduction parameter of both symbiont types and  $\beta$  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  $x \uparrow y$ , 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

$$s_0 = T - s < s_1 < \dots < s_{n+1} = T \quad \text{and} \quad x_0 = y, x_1, \dots, x_n = x$$

such that the following two conditions hold:

1. for  $i = 1, 2, \dots, n$ , there is an arrow from  $x_{i-1}$  to  $x_i$  at time  $s_i$  and
2. for  $i = 0, 1, \dots, n$ , 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

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

The dual process is naturally defined only for  $0 \leq s \leq T$ . 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  $s \geq 0$ . 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

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

The dual process starting from a finite set  $B \subset C_N(\omega)$  consists of a system of  $\text{card}(B)$  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 the left-hand side of Figure 5 for an example of realization of the dual process in the neutral case.

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

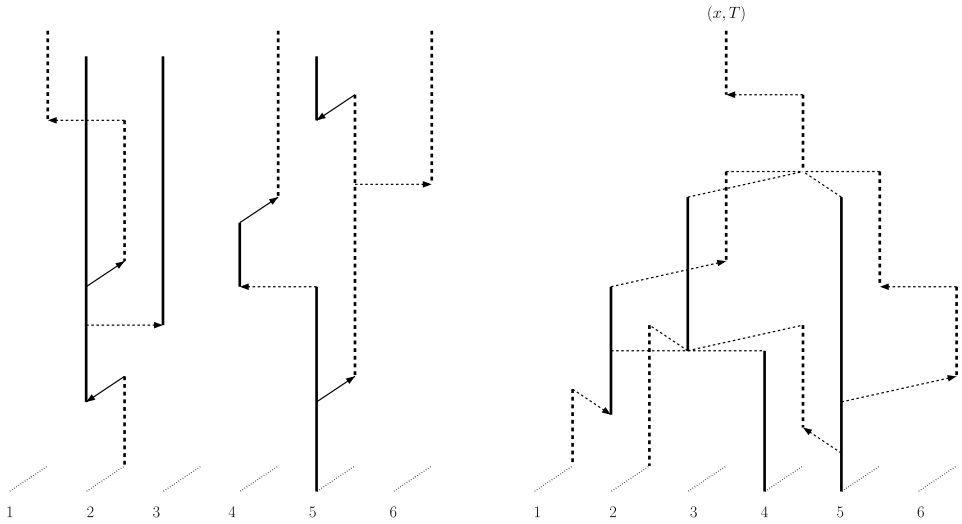


FIG. 5. Dual process in the neutral case and branching random walk  $\hat{\zeta}_s(x, T)$ . In both pictures,  $N = 2$  and for simplicity we have set  $p = 1$  and  $d = 1$ . Time goes up, and arrows within the same host are drawn in continuous line, while arrows connecting two adjacent hosts are drawn in dashed lines.

times, but also how vertices at the current time are correlated, by keeping track of the ancestry of each symbiont. In particular, the 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

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

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

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

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 7. *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,

$$(4) \quad P(\Theta_t^1 \cap B = \emptyset) = \mathbb{E}(1 - \theta)^{|\hat{\xi}_t(B,t)|},$$

where  $\theta$  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 (4) also has a limit as  $t \rightarrow \infty$ . 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

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

Let  $\tau$  be the hitting time of the dual processes, that is,

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

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  $\tau$  when they coalesce. We set  $t_0 = 0$  and define inductively

$$s_i = \inf\{s > t_{i-1} : X_s = Y_s\},$$

$$t_i = \inf\{s > s_i : X_s \neq Y_s\} = \inf\{s > s_i : X_s \neq X_{s_i} \text{ or } Y_s \neq Y_{s_i}\}$$

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

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

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:

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

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 (5) implies

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

by the bounded convergence theorem since  $P(I_T = j) \rightarrow 0$  as  $T \rightarrow \infty$  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 (5) implies that

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

which shows that coexistence occurs. This completes the proof.  $\square$

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 7, 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 [15], 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 though this result will not be used to establish clustering of the process.

LEMMA 8. *The cluster  $C_\infty(\omega)$  is recurrent in  $d = 2$  and transient in  $d \geq 3$ .*

PROOF. Both statements follow from the fact that any subgraph of a recurrent graph is recurrent, and equivalently any supergraph of a transient graph is transient. This is proved in details in Doyle and Snell [10] based on the analogy between random walks and electrical networks so we only give an outline of their proof. The idea is to turn the graph under consideration into an electrical network in which each edge has unit resistance. Then, simple random walks on this 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 [10]. In other respects, Rayleigh's monotonicity law states that if the resistances of a circuit are increased, respectively, decreased, then the effective resistance between any two points can only increase, respectively, decrease. In particular, removing edges induces an increase of the effective resistance between any two points, therefore any subgraph of a recurrent graph is recurrent. Again, we refer to [10] for the details.

With the previous result in hands, the first statement follows directly from the fact that the infinite percolation cluster in two dimensions is a subgraph of the two-dimensional lattice, which is recurrent by Polya's theorem. Transience of the infinite percolation cluster in higher dimensions has been established by Grimmett, Kesten and Zhang [15]. Their proof relies on the construction of a transient tree-like graph that can be embedded in the infinite percolation cluster. Note that their result applies to bond percolation but relies on geometric properties that are known for site percolation as well, so their proof easily extends to our context.  $\square$

LEMMA 9. *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,

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

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

$$\begin{aligned} \mathbb{E}I(X, Y) &= 2\mathbb{E}\left(\int_0^\infty \mathbb{1}\{X_t = Y_t\} dt\right) = 2 \int_0^\infty P(X_t = Y_t) dt \\ (6) \qquad &= 2 \int_0^\infty \sum_{B \in C_\infty(\omega)} P(X_t = B)P(Y_t = B) dt \\ &= 2 \int_0^\infty \sum_{B \in C_\infty(\omega)} (p_t(A, B))^2 dt, \end{aligned}$$

where  $p_t(A, B) = P(X_t = B \mid X_0 = A)$ . 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

$$(7) \quad (2d)^{-1} p_t(B, A) \leq p_t(A, B) \leq 2d p_t(B, A) \quad \text{for all } B \in C_\infty(\omega)$$

since  $1 \leq \text{deg}(A), \text{deg}(B) \leq 2d$ . Therefore, when  $C_\infty(\omega)$  is recurrent, (6) and (7) imply that

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

whereas when  $C_\infty(\omega)$  is transient, (6) and (7) imply that

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

The result then follows from Lemma 8.  $\square$

Lemma 9 indicates that  $P(I(X, Y) < \infty) = 1$  in dimensions  $d \geq 3$ , that is,  $C_\infty(\omega)$  has the finite collision property, which, together with Lemma 7, 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 [20], 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 7, is equivalent to clustering of the neutral competition model in two dimensions. We use the same notation 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

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

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 and the Borel–Cantelli lemma, to prove the infinite collision property, it suffices to prove that

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

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 two-dimensional case in the rest of this section.

**THEOREM 10.** *Let  $p > p_c$ . Then, there exist a subset  $\Omega$  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  $A, B \in C_\infty(\omega)$ ,

$$(8) \quad \begin{aligned} q_n(A, B) &\geq c_1 n^{-d/2} \exp(-c_2|A - B|^2/n) && \text{whenever } |A - B| \vee S_A(\omega) \leq n, \\ q_n(A, B) &\leq c_3 n^{-d/2} \exp(-c_4|A - B|^2/n) && \text{whenever } S_A(\omega) \leq n. \end{aligned}$$

The proof of Theorem 10 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 sub-Gaussian upper estimate, the idea is to use a discrete-time version of [24], Theorem 1.1, and the results of [9], Sections 5, 6 and 8, while the proof of the sub-Gaussian 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 notation, 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 11. *Fix  $\rho \in (0, 1/2)$ ,  $A, B, X \in C_\infty(\omega)$  and  $\varepsilon > 0$ . Then  $F(n) \leq (1 + \varepsilon)F_\rho(n)$  when  $n$  is sufficiently large depending on  $\rho, A, B, X$  and  $\varepsilon$ .*

PROOF. By the Hoeffding inequality ([18], Theorem 1), we have

$$F(n) - F_\rho(n) \leq \sum_{j=0}^{\rho n} 2^{-n} \binom{n}{j} + \sum_{j=(1-\rho)n}^n 2^{-n} \binom{n}{j}$$

$$\leq 2 \exp(-2n(1/2 - \rho)^2).$$

Taking  $n$  such that

$$\rho n \geq \sqrt{n} \vee S_A(\omega) \vee S_B(\omega)$$

$$\geq |X - A| \vee |X - B| \vee S_A(\omega) \vee S_B(\omega),$$

we may use the first inequality in (8). Letting  $\Phi$  denote the cumulative distribution function of the standard normal, and also applying the central limit theorem, we



obtain

$$\begin{aligned}
 F_\rho(n) &\geq \sum_{j=\rho n}^{(1-\rho)n} 2^{-n} \binom{n}{j} \frac{c_1}{j} \exp\left(-\frac{c_2|X-A|^2}{j}\right) \frac{c_1}{n-j} \exp\left(-\frac{c_2|X-B|^2}{n-j}\right) \\
 &\geq \sum_{k=\rho n}^{(1-\rho)n} 2^{-n} \binom{n}{j} \frac{c_1^2}{j(n-j)} \exp\left(-\frac{c_2 n^2}{j(n-j)}\right) \\
 (9) \quad &\geq (2c_1/n)^2 \exp(-c_2/((1-\rho)\rho)) \sum_{j=\rho n}^{(1-\rho)n} 2^{-n} \binom{n}{j} \\
 &\geq C_5 n^{-2} (2\Phi((1-2\rho)\sqrt{n}) - 1)
 \end{aligned}$$

for some  $C_5 < \infty$ . 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 (9) imply

$$\lim_{n \rightarrow \infty} \frac{F(n) - F_\rho(n)}{F_\rho(n)} \leq \lim_{n \rightarrow \infty} \frac{n^2 \exp(-2n(1/2 - \rho)^2)}{C_5 \Phi((1-2\rho)\sqrt{n}) - 1/2} = 0.$$

This completes the proof.  $\square$

With Theorem 10 and Lemma 11 in hand, we are now ready to prove that the infinite percolation cluster has the infinite collision property in the sense described above, that is, considering continuous-time random walks run at a constant rate, say 1. Our proof relies in addition on an argument of Barlow, Peres and Sousi [2] 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 12.** *Fix a realization  $\omega$ . Then, for all  $A, B \in C_\infty(\omega)$ ,*

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

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

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

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

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

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

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

then  $j \wedge (n - j) \geq |X - A| \vee |X - B| \vee S_A(\omega) \vee S_B(\omega)$  so Theorem 10 implies

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

This and the central limit theorem imply that, for  $k$  large depending on  $A, B$ ,

$$\begin{aligned} \mathbb{E}^{A,B}(I_k) &\geq \exp(-c_2/(1-\rho)) \\ &\quad \times \sum_{n=k}^{k^2} (2c_1/n)^2 \text{card}\{X \in C_\infty(\omega) : \\ (10) \quad &|X - A| \vee |X - B| < \sqrt{\rho n}, \\ &S_X(\omega) \leq \gamma\}. \end{aligned}$$

Now, by the ergodic theorem,

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

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

$$(11) \quad \text{card}\{X \in C_\infty(\omega) : |X - A| \vee |X - B| < \sqrt{\rho n}, S_X(\omega) \leq \gamma\} \geq \delta \rho n$$

for  $n \geq k$ . By (10) and (11), there exists  $k_1(A, B)$  large such that

$$(12) \quad \mathbb{E}^{A,B}(I_k) \geq c_1^2 \exp(-c_2/(1-\rho)) \sum_{n=k}^{k^2} \frac{\delta \rho}{n} \geq C_6(\log k^2 - \log k) = C_6 \log k$$

for a suitable  $C_6 > 0$  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

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

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 11 twice with  $\varepsilon = 1$ , we deduce that there exists  $\gamma$  large such that for all  $k$  sufficiently large

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

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

$$j \wedge (n - j) \geq \rho n \geq \rho k \geq S_A(\omega) \vee S_B(\omega),$$

$$i \wedge (l - n - i) \geq \rho(l - n) \geq \rho\gamma/\rho = \gamma \geq S_X(\omega)$$

for all  $k$  large depending on  $A, B, X$ , Theorem 10 implies that

$$\begin{aligned} \mathbb{E}^{A,B}(I_k^2) &\leq 8 \sum_{n=k}^{k^2} 5n \left( \sum_{j=\rho n}^{(1-\rho)n} 2^{-n} \binom{n}{j} \frac{c_3^2}{j(n-j)} \exp(-c_4n(j^{-1} + (n-j)^{-1})) \right) \\ &\quad \times \left( \frac{\gamma}{\rho} + \sum_{l=n+\gamma/\rho}^{k^2} \sum_{m=0}^{\infty} \text{card}\{Y : m\sqrt{l-n} \leq |X - Y| \right. \\ &\quad \left. < (m+1)\sqrt{l-n} \right) \\ &\quad \times \sum_{i=\rho(l-n)}^{(1-\rho)(l-n)} 2^{-(l-n)} \binom{l-n}{i} \frac{c_3^2}{i(l-n-i)} \\ &\quad \times \exp(-c_4m^2(i^{-1} + (l-n-i)^{-1})). \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}^{A,B}(I_k^2) &\leq 8 \sum_{n=k}^{k^2} \frac{5c_3^2}{\rho(1-\rho)n} \\ &\quad \times \left( \frac{\gamma}{\rho} + \sum_{l=n+\gamma/\rho}^{k^2} \sum_{m=0}^{\infty} \frac{5(m+2)^2 c_3^2}{\rho(1-\rho)(l-n)} \right. \\ (13) \quad &\quad \left. \times \exp(-4c_4(m+1)^2) \right) \\ &\leq C_7 \sum_{n=k}^{k^2} \frac{1}{n} \left( \frac{\gamma}{\rho} + C_8 \sum_{l=n+\gamma/\rho}^{k^2} \frac{1}{l-n} \right) \leq C_9(\log k)^2 \end{aligned}$$

for suitable constants  $C_7, C_8, C_9 < \infty$  not depending on  $A, B$  and all  $k \geq k_2(A, B)$ .

Let  $k(A, B) = k_1(A, B) \vee k_2(A, B)$ . By (12) and (13) and the Paley–Zygmund inequality,

$$\begin{aligned} P(I_k > (C_6/2) \log k \mid \mathcal{W}_0 = (A, B)) &\geq P(I_k > \mathbb{E}(I_k)/2 \mid \mathcal{W}_0 = (A, B)) \\ &\geq (\mathbb{E}^{A,B}(I_k))^2 / 4\mathbb{E}^{A,B}(I_k^2) \geq C_6^2/4C_9 = c > 0 \end{aligned}$$

for  $k = k(A, B)$  and where, as  $C_6$  and  $C_9$ , 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

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

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

$$\text{card}\{n \in [n_{j-1}, n_j) : \mathcal{X}_n = \mathcal{Y}_n\} > (C_6/2) \log k(A_{j-1}, B_{j-1})$$

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.  $\square$

As previously explained, clustering of the neutral competition model in two dimensions follows from the combination of Lemma 7 and Theorem 12.

**5. Proof of Theorem 3.** This section is devoted to the proof of Theorem 3 and the analysis of the competition model in the presence of selection. In the asymmetric case, the main difficulty arises from the fact that the transition rates are no longer linear with respect to the local frequencies of each symbiont type, which leads a priori to a failure of duality techniques. In particular, the invadability of type 1 is proved in the next subsection by comparing directly the forward evolution of the competition model with a gambler’s ruin model. In contrast, under the additional assumption  $\alpha_2 = 0$ , extinction of the symbionts of type 2 is established by invoking duality techniques which are available for what we shall call a threshold version of the competition model. With this duality relationship in hands, the result follows as in the previous section from random walk estimates.

*Invasion of type 1.* In order to prove the first part of Theorem 3, we first let  $\bar{\xi}_t(X)$  be the number of type 1 symbionts in the host at  $X \in \mathbb{Z}^d$  and set

$$\begin{aligned} p_t(X) &= \left( 2d\alpha_1 \bar{\xi}_t(X) + \beta_1 \sum_{Y \sim X} \bar{\xi}_t(Y) \right) \\ &\quad \times \left( 2d\alpha_1 \bar{\xi}_t(X) + 2d\alpha_2 (N - \bar{\xi}_t(X)) \right. \\ &\quad \left. + \beta_1 \sum_{Y \sim X} \bar{\xi}_t(Y) + \beta_2 \sum_{Y \sim X} (N - \bar{\xi}_t(Y)) \right)^{-1}, \\ q_t(X) &= \frac{2d\alpha_1 \bar{\xi}_t(X) + \beta_1 \sum_{Y \sim X} \bar{\xi}_t(Y)}{2dN(\alpha_1 + \beta_1)}. \end{aligned}$$

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

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

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

$$q^- := \inf\{q_t(X) : q_t(X) \in (0, 1)\} > 0$$

and

$$q^+ := \sup\{q_t(X) : q_t(X) \in (0, 1)\} < 1.$$

Denote by  $N_t$  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

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

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  $\{N_t\}_t$  is the simple symmetric random walk on  $\mathbb{Z}_+$  absorbed at 0 [note that on each edge  $(x, y)$  the rates of invasion between  $x$  and  $y$  are symmetric], therefore the two rates above are equal and

$$\begin{aligned} \frac{1}{N} \sum_X \bar{\xi}_t(X) &= \frac{1}{N} \sum_X \bar{\xi}_t(X)(1 - p_t(X)) + \frac{1}{N} \sum_X \bar{\xi}_t(X)p_t(X) \\ &= \frac{1}{N} \sum_X (N - \bar{\xi}_t(X))p_t(X) + \frac{1}{N} \sum_X \bar{\xi}_t(X)p_t(X) \\ &= \sum_X p_t(X). \end{aligned}$$

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

$$(14) \quad \frac{1}{N} \sum_X \bar{\xi}_t(X) = \sum_X \frac{2d\alpha_1 \bar{\xi}_t(X) + \beta_1 \sum_{Y \sim X} \bar{\xi}_t(Y)}{2dN(\alpha_1 + \beta_1)} = \sum_X q_t(X).$$

Note that the expression of  $q_t(X)$  depends neither on  $\alpha_2$  nor on  $\beta_2$  therefore the equation above does not hold in the neutral case only. It is always true regardless of the choice of the reproduction and transmission rates. Now, assume that  $\alpha_1 \geq \alpha_2$  and  $\beta_1 > \beta_2$ . We want to show that in this case the process  $\{N_t\}_t$  has 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

$$\begin{aligned}
 p_t(X) &= \left( 2d\alpha_1\bar{\xi}_t(X) + \beta_1 \sum_{Y \sim X} \bar{\xi}_t(Y) \right) \\
 &\quad \times \left( 2dN\alpha_1 + 2dN\beta_1 - (\alpha_1 - \alpha_2)(N - \bar{\xi}_t(X)) \right. \\
 &\quad \left. - (\beta_1 - \beta_2) \sum_{Y \sim X} (N - \bar{\xi}_t(Y)) \right)^{-1} \\
 (15) \qquad &\geq \frac{2d\alpha_1\bar{\xi}_t(X) + \beta_1 \sum_{Y \sim X} \bar{\xi}_t(Y)}{2dN(\alpha_1 + \beta_1)} = q_t(X),
 \end{aligned}$$

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

$$\begin{aligned}
 p_t(X) &\geq \frac{2d\alpha_1\bar{\xi}_t(X) + \beta_1 \sum_{Y \sim X} \bar{\xi}_t(Y)}{2dN(\alpha_1 + \beta_1)} \left[ 1 - \frac{\beta_1 - \beta_2}{2dN(\alpha_1 + \beta_1)} \right]^{-1} \\
 (16) \qquad &\geq q_t(X)(1 - c)^{-1} \qquad \text{where } c = \frac{\beta_1 - \beta_2}{2dN(\alpha_1 + \beta_1)}.
 \end{aligned}$$

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

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

The third ingredient is to observe that if  $X$  is bad then all sites  $Y \sim X$  are good, so the number of bad sites is at most equal to the number of good sites. In particular, letting

$$B = \{X : X \text{ is bad}\}$$

and

$$G = \{X : X \text{ is good}\},$$

there exists a subset  $G^* \subset G$  with  $\text{card}(G^*) = \text{card}(B)$ . Then, combining (14)–(17) gives

$$\begin{aligned} \sum_X p_t(X) &= \sum_{X \text{ bad}} p_t(X) + \sum_{X \text{ good}} p_t(X) \\ &\geq \sum_{X \in B} q_t(X) + (1-c)^{-1} \sum_{X \in G^*} q_t(X) + (1-c)^{-1} \sum_{X \in G \setminus G^*} q_t(X) \\ &\geq \frac{q^-(1-c)^{-1} + q^+}{q^- + q^+} \sum_{X \in B \cup G^*} q_t(X) + (1-c)^{-1} \sum_{X \in G \setminus G^*} q_t(X) \\ &\geq \frac{q^-(1-c)^{-1} + q^+}{q^- + q^+} \sum_X q_t(X) = \frac{q^-(1-c)^{-1} + q^+}{q^- + q^+} \frac{1}{N} \sum_X \bar{\xi}_t(X). \end{aligned}$$

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, we have

$$P\left(\lim_{t \rightarrow \infty} N_t = \infty\right) \geq 1 - \left(\frac{q^- + q^+}{q^-(1-c)^{-1} + q^+}\right)^K > 0.$$

This establishes the first part of Theorem 3.

*Extinction of type 2.* We now prove that, under the extra assumption  $\alpha_2 = 0$ , type 1 outcompetes type 2, which is the second part of Theorem 3. Letting  $C_N$  denote the Cartesian product of the regular lattice  $\mathbb{Z}^d$  and  $\mathbb{K}_N$  the expression of the Markov generator reduces to

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

Introduce  $\beta = \beta_2$  and  $\kappa = \beta_1 - \beta_2 > 0$ . Let  $\gamma = \kappa \cdot (2dN)^{-1}$  and observe that

$$g_1(x) \neq 0 \text{ implies that } \kappa g_1(x) \geq \kappa \cdot (N \deg \pi(x))^{-1} = \kappa \cdot (2dN)^{-1} = \gamma$$



from which it follows that

$$\frac{\alpha_1 f_1(x) + \beta_1 g_1(x)}{\alpha_1 f_1(x) + \beta_1 g_1(x) + \beta_2 g_2(x)} = \frac{\alpha_1 f_1(x) + \beta g_1(x) + \kappa g_1(x)}{\alpha_1 f_1(x) + \beta + \kappa g_1(x)} \geq \frac{\beta g_1(x) + \gamma}{\beta + \gamma},$$

$$\frac{\beta_2 g_2(x)}{\alpha_1 f_1(x) + \beta_1 g_1(x) + \beta_2 g_2(x)} = \frac{\beta g_2(x)}{\alpha_1 f_1(x) + \beta + \kappa g_1(x)} \leq \frac{\beta g_2(x)}{\beta + \gamma}.$$

In particular, it suffices to prove that type 1 symbionts outcompete type 2 symbionts when the dynamics are described by the new process  $\{\zeta_t\}_t$  with Markov generator

$$L_3 f(\zeta) = \sum_{x \in C_N} \frac{\beta g_1(x) + \gamma}{\beta + \gamma} \mathbb{1}_{\{g_1(x) \neq 0\}} [f(\zeta_{x,1}) - f(\zeta)]$$

$$+ \sum_{x \in C_N} \left( \frac{\beta g_2(x)}{\beta + \gamma} \mathbb{1}_{\{g_1(x) \neq 0\}} + \mathbb{1}_{\{g_1(x) = 0\}} \right) [f(\zeta_{x,2}) - f(\zeta)],$$

where as previously  $\zeta_{x,i}$  is obtained from  $\zeta$  by assigning the value  $i$  to vertex  $x$  and leaving the state of all the other vertices unchanged. The process can be constructed graphically from the three collections of independent processes introduced in Table 1 in the following manner. First, we equip the set  $C_N$  with a total order relation. Then, at time  $T_n(z)$ , we have the following alternative:

1. In the case when  $U_n(z) < p := \beta(\beta + \gamma)^{-1}$ , we select a neighboring vertex uniformly at random, say  $y$ , using in an obvious manner the uniform random variable  $V_n(z)$  and the total order on the set  $C_N$ , and then draw an arrow from vertex  $y$  to vertex  $z$ .
2. In the case when  $U_n(z) > p := \beta(\beta + \gamma)^{-1}$ , we draw a set of  $2dN$  arrows starting from each of the vertices adjacent to the host at site  $\pi(z)$  and pointing at vertex  $z$ .

We call the events in rules 1 and 2 above jumping event and branching event, respectively. The type of vertex  $z$  is updated at times  $T_n(z)$ ,  $n \geq 1$ , with  $z$  becoming of type 1 if at least one of the arrows that point at vertex  $z$  originates from a type 1, and of type 2 otherwise. In particular, if all the neighbors are of type 2 then the new type is 2 while if at least one neighbor is of type 1 then the new type is

TABLE 1  
Graphical representation

Notation	Description	Interpretation
$T_n(x)$	Poisson process with parameter 1	Times of an update at vertex $x$
$U_n(x)$	Uniform random variable on $(0, 1)$	Determining the new symbiont type
$V_n(x)$	Uniform random variable on $\{1, 2, \dots, 2dN\}$	Determining the new symbiont type

chosen uniformly at random from the neighbors with probability  $p$  or type 1 with probability  $1 - p$ , which produces the suitable transition rates. We say that there exists a path from space–time point  $(y, T - s)$  to point  $(x, T)$  if there are sequences of times and vertices

$$s_0 = T - s < s_1 < \dots < s_{n+1} = T \quad \text{and} \quad x_0 = y, \quad x_1, \dots, x_n = x,$$

such that the following two conditions hold:

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

We define a set-valued process  $\{\hat{\zeta}_s(x, T)\}_s$  by setting for all  $0 \leq s \leq T$

$$(18) \quad \hat{\zeta}_s(x, T) = \{y \in C_N : \text{there is a path from } (y, T - s) \text{ to } (x, T)\}.$$

Note that the process  $\{\hat{\zeta}_s(x, T)\}_s$  consists of a system of branching coalescing random walks in which particles independently jump at rate  $p$  and branch at rate  $1 - p$ . We refer to the right-hand side of Figure 5 on page 1232 for a picture. The introduction of the process (18) is motivated by the following lemma, which is somewhat reminiscent of the duality relationship between the biased voter model and branching coalescing random walks.

LEMMA 13. *Assume that  $\zeta_0(z) = 1$  for some  $z \in \hat{\zeta}_T(x, T)$ . Then,  $\zeta_T(x) = 1$ .*

PROOF. Let  $z \in \hat{\zeta}_T(x, T)$  with  $\zeta_0(z) = 1$ . Then there is a unique path from  $(z, 0)$  to  $(x, T)$ . Using the same notation as in the definition of a path, we introduce the jump process

$$X_t = x_i \quad \text{for all } s_i \leq t < s_{i+1} \quad \text{and} \quad X_T = x.$$

From the construction of the process  $\{\zeta_t\}_t$ , we have  $\zeta_t(X_t) = 1$  for all  $0 \leq t \leq T$ . The lemma follows immediately by applying the equation at time  $t = T$  since  $X_T = x$ .  $\square$

The rest of the proof relies on standard random walk estimates supplemented with a rescaling argument similar to the one described in Section 3. In short, introducing the spatial regions

$$B(X, K) = \{z \in C_N : \pi(z) \in X + (-K, K)^d\} \quad \text{for } X \in \mathbb{Z}^d \text{ and } K \in \mathbb{Z}_+,$$

the next objective is to show that, for all  $\varepsilon > 0$  and  $T = K^2$ , there exists  $K$  large such that

$$(19) \quad \begin{aligned} &P(\zeta_T(x) = 1 \text{ for all } x \in B(X, 3K) \mid \zeta_0(x) = 1 \text{ for all } x \in B(X, K)) \\ &\geq 1 - \varepsilon. \end{aligned}$$

In view of Lemma 13, inequality (19) follows directly from the following result.

LEMMA 14. *Let  $T = K^2$ . Then, there exists  $C_{10} < \infty$  and  $\gamma_{10} > 0$  such that  $P(\hat{\xi}_T(x, T) \cap B(X, K) = \emptyset) \leq C_{10} \exp(-\gamma_{10}K)$  for all  $x \in B(X, 3K)$ .*

PROOF. Let  $x \in B(X, 3K)$ . The idea is to define a random walk  $\{W_s\}_s$  embedded in the system of branching coalescing random walks and connecting, with probability close to one, the space-time point  $(x, T)$  to a vertex in the ball  $B(X, K)$  at time 0. The process starts at  $W_0 = x$ . To define the dynamics, we also introduce the projection  $\mathcal{W}_s = \pi(W_s)$ . Then, the random walk jumps at each time  $s$  such that  $T - s = T_n(W_{s-})$  when we have the following alternative:

1. Jumping event: when  $U_n(W_s) < p$ , there is an arrow from a vertex, say  $z$ , to  $W_{s-}$ . Then, the random walk jumps to vertex  $z$ , that is, we set  $W_s = z$ .
2. Branching event: when  $U_n(W_s) > p$ , there are  $2dN$  arrows that point at  $W_{s-}$ . Then, the random walk jumps to one of the tails, chosen randomly and uniformly, that make the random walk's projection  $\mathcal{W}_s$  closer to the center  $X$  of the ball.

Note that  $W_s \in \hat{\xi}_s(x, T)$  for all  $s \in (0, T)$ , as desired. Introduce

$$Y_s^i = |\pi_i(\mathcal{W}_s) - \pi_i(X)| \quad \text{for } i = 1, 2, \dots, d,$$

where  $\pi_i$  is the projection on the  $i$ th axis in  $\mathbb{Z}^d$ . Since each vertex has at most  $dN$  neighbors which are closer to the center of the target ball for a total of  $2dN$  neighbors, we have

$$\begin{aligned} \lim_{h \rightarrow 0} h^{-1} \cdot P(Y_{s+h}^i = Y_s^i + 1 \mid Y_s^i > 0) &\leq p/(2d) =: r, \\ \lim_{h \rightarrow 0} h^{-1} \cdot P(Y_{s+h}^i = Y_s^i - 1 \mid Y_s^i > 0) &\geq p/(2d) + (1 - p)/d =: l. \end{aligned}$$

Therefore,  $\{Y_s^i\}_s$  is stochastically smaller than the random walk  $\{Z_s\}_s$  with a reflecting boundary at zero and that otherwise jumps to the right at rate  $r$  and to the left at rate  $l$ . Let  $\tau$  denote the first time the random walk  $Z_s$  hits the boundary 0. Since  $r < l$ , standard large deviation estimates for the Poisson distribution imply that

$$(20) \quad P(\tau > C_{11}K \mid Z_0 \leq 3K) \leq C_{12} \exp(-\gamma_{12}K)$$

for suitable constants. Finally, we introduce the reverse asymmetric random walk  $\{\bar{Z}_s\}_s$  with state space  $\mathbb{Z}$  that jumps to the right at rate  $l$  and to the left at rate  $r$ . Letting

$$u_k = P(\bar{Z}_s = 0 \text{ for some } s > 0 \mid \bar{Z}_0 = k)$$

a first-step analysis gives  $l(u_{k+1} - u_k) = r(u_k - u_{k-1})$  and then

$$(1 - a)(u_k - u_0) = (1 - a) \sum_{j=0}^{k-1} a^j (u_1 - u_0) = (1 - a^k)(u_1 - u_0),$$

where  $a = r \cdot l^{-1}$ . It is straightforward to deduce that  $u_k = a^k$ . In particular,

$$(21) \quad \begin{aligned} P(Z_T > K \mid \tau < T) &\leq P(Z_T > K \mid Z_0 = 0) \\ &\leq P(\bar{Z}_s = 0 \text{ for some } s < T \mid \bar{Z}_0 > K) \leq u_K = a^K. \end{aligned}$$

In conclusion, recalling the definition of the processes  $\{Y_s^i\}_s$ , using the stochastic domination mentioned above, and applying (20) and (21), we obtain

$$\begin{aligned} P(\hat{\zeta}_T(x, T) \cap B(X, K) = \emptyset) &\leq P(W_T \notin B(X, K)) \leq d \times P(Y_T^i > K) \\ &\leq d \times P(\tau > T \mid Z_0 \leq 3K) + d \times P(Z_T > K \mid \tau < T) \\ &\leq d \times C_{12} \exp(-\gamma_{12}K) + d \times a^K \end{aligned}$$

for all  $K$  large. Since  $a = r \cdot l^{-1} < 1$ , the lemma follows.  $\square$

Lemmas 13 and 14 imply that, when viewed under suitable scales, the set of space–time boxes which are void of type 2 dominates oriented site percolation with parameter  $1 - \varepsilon$ . This almost produces the second part of Theorem 3. The last problem is that oriented site percolation has a positive density of unoccupied sites. To prove that there is an in-all-directions expanding region which is indeed void of type 2 symbionts, we apply a result from Durrett [12] which shows that unoccupied sites do not percolate when  $\varepsilon$  is close enough to 0. Since symbionts of either type cannot appear spontaneously, once a region is void of one type, this type can only reappear in the region through invasion from the outside. This then implies that our process has the desired property and completes the proof of the second part of Theorem 3.

**Acknowledgments.** The authors would like to thank Martin Barlow, Yuval Peres and Perla Sousi for fruitful discussions about random walks on random graphs, as well as two anonymous referees for comments that helped to improve the article.

## REFERENCES

- [1] BARLOW, M. T. (2004). Random walks on supercritical percolation clusters. *Ann. Probab.* **32** 3024–3084. [MR2094438](#)
- [2] BARLOW, M., PERES, Y. and SOUSI, P. (2009). Collisions of random walks. Available at [arXiv:1003.3255](#).
- [3] BELHADJI, L., BERTACCHI, D. and ZUCCA, F. (2009). A self-regulating and patch subdivided population. *Adv. in Appl. Probab.* **42** 899–912.
- [4] BERTACCHI, D. and ZUCCA, F. (2009). Approximating critical parameters of branching random walks. *J. Appl. Probab.* **46** 463–478. [MR2535826](#)

- [5] BRAMSON, M. and DURRETT, R. (1988). A simple proof of the stability criterion of Gray and Griffeath. *Probab. Theory Related Fields* **80** 293–298. [MR0968822](#)
- [6] BRAMSON, M. and GRIFFEATH, D. (1980). On the Williams–Bjerknes tumour growth model. II. *Math. Proc. Cambridge Philos. Soc.* **88** 339–357. [MR0578279](#)
- [7] BRAMSON, M. and GRIFFEATH, D. (1981). On the Williams–Bjerknes tumour growth model. I. *Ann. Probab.* **9** 173–185. [MR0606980](#)
- [8] CLIFFORD, P. and SUDBURY, A. (1973). A model for spatial conflict. *Biometrika* **60** 581–588. [MR0343950](#)
- [9] COULHON, T., GRIGOR'YAN, A. and ZUCCA, F. (2005). The discrete integral maximum principle and its applications. *Tohoku Math. J. (2)* **57** 559–587. [MR2203547](#)
- [10] DOYLE, P. G. and SNELL, J. L. (1984). *Random Walks and Electric Networks*. *Carus Mathematical Monographs* **22**. Math. Assoc. America, Washington, DC. [MR0920811](#)
- [11] DURRETT, R. (1984). Oriented percolation in two dimensions. *Ann. Probab.* **12** 999–1040. [MR0757768](#)
- [12] DURRETT, R. (1992). Multicolor particle systems with large threshold and range. *J. Theoret. Probab.* **5** 127–152. [MR1144730](#)
- [13] DURRETT, R. and LANCHIER, N. (2008). Coexistence in host-pathogen systems. *Stochastic Process. Appl.* **118** 1004–1021. [MR2418255](#)
- [14] GRIMMETT, G. (1989). *Percolation*. Springer, New York. [MR0995460](#)
- [15] 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. [MR1222363](#)
- [16] HARRIS, T. E. (1972). Nearest-neighbor Markov interaction processes on multidimensional lattices. *Adv. Math.* **9** 66–89. [MR0307392](#)
- [17] HARRIS, T. E. (1974). Contact interactions on a lattice. *Ann. Probab.* **2** 969–988. [MR0356292](#)
- [18] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. [MR0144363](#)
- [19] HOLLEY, R. A. and LIGGETT, T. M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probab.* **3** 643–663. [MR0402985](#)
- [20] KRISHNAPUR, M. and PERES, Y. (2004). Recurrent graphs where two independent random walks collide finitely often. *Electron. Comm. Probab.* **9** 72–81 (electronic). [MR2081461](#)
- [21] LANCHIER, N. and NEUHAUSER, C. (2006). Stochastic spatial models of host-pathogen and host-mutualist interactions. I. *Ann. Appl. Probab.* **16** 448–474. [MR2209349](#)
- [22] LANCHIER, N. and NEUHAUSER, C. (2006). A spatially explicit model for competition among specialists and generalists in a heterogeneous environment. *Ann. Appl. Probab.* **16** 1385–1410. [MR2260067](#)
- [23] LANCHIER, N. and NEUHAUSER, C. (2010). Stochastic spatial models of host-pathogen and host-mutualist interactions II. *Stoch. Models* **26** 399–430. [MR2739344](#)
- [24] MATHIEU, P. and REMY, E. (2004). Isoperimetry and heat kernel decay on percolation clusters. *Ann. Probab.* **32** 100–128. [MR2040777](#)
- [25] 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.
- [26] MORAN, P. A. P. (1958). Random processes in genetics. *Proc. Cambridge Philos. Soc.* **54** 60–71. [MR0127989](#)
- [27] PRICE, P. W. (1980). *Evolutionary Biology of Parasites*. Princeton Univ. Press, Princeton.
- [28] TRAPPE, J. M. (1987). Phylogenetic and ecologic aspects of mycotrophy in the angiosperms from an evolutionary standpoint. In *Ecophysiology of VA Mycorrhizal Plants*. (G. R. Safir, ed.) 5–25. CRC Press, Boca Raton, FL.

- [29] WOESS, W. (2000). *Random Walks on Infinite Graphs and Groups*. *Cambridge Tracts in Mathematics* **138**. Cambridge Univ. Press, Cambridge. MR1743100

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

N. LANCHIER  
SCHOOL OF MATHEMATICAL  
AND STATISTICAL SCIENCES  
ARIZONA STATE UNIVERSITY  
TEMPE, ARIZONA 85287  
USA  
E-MAIL: [lanchier@math.asu.edu](mailto:lanchier@math.asu.edu)

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