Exponential extinction time of the contact process on finite graphs

Thomas Mountford, Jean-Christophe Mourrat, Daniel Valesin, Qiang Yao

To cite this version:

EXPONENTIAL EXTINCTION TIME OF THE CONTACT PROCESS ON FINITE GRAPHS

THOMAS MOUNTFORD\textsuperscript{1}, JEAN-CHRISTOPHE MOURRAT\textsuperscript{1}, DANIEL VALESIN\textsuperscript{2,3} AND QIANG YAO\textsuperscript{4,5}

Abstract. We study the extinction time \( \tau \) of the contact process started with full occupancy on finite trees of bounded degree. We show that, if the infection rate is larger than the critical rate for the contact process on \( \mathbb{Z} \), then, uniformly over all trees of degree bounded by a given number, the expectation of \( \tau \) grows exponentially with the number of vertices. Additionally, for any increasing sequence of trees of bounded degree, \( \tau \) divided by its expectation converges in distribution to the unitary exponential distribution. These results also hold if one considers a sequence of graphs having spanning trees with uniformly bounded degree, and provide the basis for powerful coarse-graining arguments. To demonstrate this, we consider the contact process on a random graph with vertex degrees following a power law. Improving a result of Chatterjee and Durrett [CD09], we show that, for any non-zero infection rate, the extinction time for the contact process on this graph grows exponentially with the number of vertices.

MSC 2010: 82C22, 05C80.

Keywords: contact process, interacting particle systems, metastability.

1. Introduction

We study the contact process on finite graphs. Our main goal is to present robust results and techniques which justify that, for certain values of the infection parameter, the contact process survives for a time which is exponential in the number of vertices of the underlying graph. We start by introducing notations and recalling important facts.

Let \( G = (V, E) \) be a graph with undirected edges. The contact process on \( G \) with parameter \( \lambda > 0 \) is a continuous-time Markov process \((\xi_t)_{t \geq 0}\) on the space of subsets of \( V \) whose transitions are given by:

\[
\begin{align*}
\text{for every } x \in \xi_t, & \quad \xi_t \to \xi_t \setminus \{x\} \quad \text{with rate } 1, \\
\text{for every } x \notin \xi_t, & \quad \xi_t \to \xi_t \cup \{x\} \quad \text{with rate } \lambda |\{y \in \xi_t : \{x, y\} \in E\}|, 
\end{align*}
\]

where for a set \( A \), we write \(|A|\) to denote its cardinality.

Given \( A \subseteq V \), we write \((\xi^A_t)_{t \geq 0}\) to denote the contact process started from the initial configuration equal to \( A \). For \( x \in V \), we write \((\xi^x_t)\) rather than \((\xi^\{x\}_t)\); also, rather than \((\xi^V_t)\), we write \((\xi^1_t)\). When we write \((\xi_t)_{t \geq 0}\), with no superscript, the initial configuration will either be clear from the context or unimportant. We often

\textsuperscript{1}École Polytechnique Fédérale de Lausanne, Département de Mathématiques, 1015 Lausanne, Switzerland
\textsuperscript{2}University of British Columbia, Department of Mathematics, V6T1Z2 Vancouver, Canada
\textsuperscript{3}Research funded by the Post-Doctoral Research Fellowship of the Government of Canada
\textsuperscript{4}School of Finance and Statistics, East China Normal University, Shanghai 200241, China
\textsuperscript{5}Research partially supported by NSFC grants (No.11201150, No.11126236 and No.10901008), SSSTC grant (No. EG55-032010), the Fundamental Research Funds for the Central Universities (No.239201278210075), and the 111 Project (B14019)
abuse notation and associate configurations $\xi \subseteq V$ with the corresponding indicator functions (that is, elements of $\{0, 1\}^V$).

The contact process can be thought of as a model for the spread of an infection in a population. Vertices of the graph (sometimes referred to as sites) represent individuals. In a configuration $\xi \in \{0, 1\}^V$, individuals in state 1 are said to be infected, and individuals in state 0 are healthy. Pairs of individuals that are connected by edges in the graph are in proximity to each other in the population. With this terminology, the dynamics can be described as follows. First, infected individuals recover with rate 1. Second, an individual that is infected transmits its infection to a neighbouring site with rate $\lambda$ (assuming no multiple edges).

We begin by presenting some of the properties of the contact process for certain choices of the graph $G$, namely: the lattice $\mathbb{Z}^d$, regular infinite trees, and the finite counterparts of these graphs. For proofs of these properties and a detailed treatment of the topic, we refer the reader to [Li1, Li2].

On $\mathbb{Z}^d$ (equipped with its usual nearest-neighbour graph structure), there exists a number $\lambda_c = \lambda_c(\mathbb{Z}^d)$ such that the following holds. If $\lambda \leq \lambda_c$, then the contact process dies out, meaning that for any finite initial configuration, the empty configuration $\emptyset$ is almost surely eventually reached. On the other hand, if $\lambda > \lambda_c$, then the contact process survives strongly: for any non-empty initial configuration, the event $\{\xi_t \neq \emptyset \text{ for all } t\}$ has positive probability and, conditioned on this event, almost surely any vertex of $\mathbb{Z}^d$ becomes infected infinitely many times.

The interest in the contact process on trees was prompted when it was discovered that death and strong survival are not the only possibilities in this case [Pe92]. For $d \geq 2$, let $T_d$ denote the infinite $(d + 1)$-regular tree with a distinguished vertex $o$ called the root. The different phases of the process are captured by two constants $\lambda_1, \lambda_2$, which vary, as we now describe. To be precise, for $A \subseteq V$, let us write $\tau_A^d = \inf\{t : \xi_t^d = \emptyset\}$ for the extinction time of the process started from occupancy in $A$. We may omit the subscript $G$ when the context is clear enough, and simply write $\tau$ when the contact process is started from full occupancy, that is, $\tau = \tau^G$. Consider the graph $\{0, \ldots, n\}^d$ (viewed as a subgraph of $\mathbb{Z}^d$) and the distribution of $\tau$ for this graph, as $n$ goes to infinity. If $\lambda < \lambda_c(\mathbb{Z}^d)$, then $\tau/\log n$ converges in probability to a constant [DL88], [Ch94]; see also Theorem 3.3 in [Li2]. If $\lambda > \lambda_c$, then $\log E[\tau]/n^d$ converges to a positive constant, and $\tau/E[\tau]$ converges in distribution to the unit exponential distribution [DS88, Mo93, Mo99]. In particular, when $\lambda > \lambda_c$, the order of magnitude of the extinction time is exponential in the number of vertices of the graph; the process is said to exhibit metastability, meaning that it persists for a long time in a state that resembles an equilibrium and then quickly moves to its true equilibrium ($\emptyset$ in this case). Metastability for the contact process in this setting was also studied in [CGOV84] and [Sc85]. Finally, if $d = 1$, it is also known that, if $\lambda = \lambda_c$, then $\tau/n \to \infty$ and $\tau/n^4 \to 0$ in probability [DST89].

For the case of finite trees, the picture is less complete, and the available results concerning the extinction time are contained in [St01]. Fix $d \geq 2$, let $T_d^h$ be the finite subgraph of $T_d$ defined by considering up to $h$ generations from the root and again take the contact process started from full occupancy on this graph, with associated extinction time $\tau$. If $\lambda < \lambda_2$, then there exist constants $c, C > 0$
such that \( P(ch \leq \tau \leq Ch) \to 1 \) as \( h \to \infty \). If \( \lambda > \lambda_2 \), then for any \( \sigma < 1 \), \( P \left[ \tau > \exp((\sigma d)^h) \right] \to 1 \) as \( h \to \infty \). This implies that \( \tau \) is at least as large as a stretched exponential function of the number of vertices, \((d + 1)^h\). As far as we know, no rigorous results are available concerning more general classes of graphs.

For \( n \in \mathbb{N} \) and \( d > 0 \), let \( \Lambda(n, d) \) be the set of all (connected) trees with \( n \) vertices and degree bounded by \( d \), and let \( \mathcal{G}(n, d) \) be the set of graphs having a spanning tree in \( \Lambda(n, d) \). In this paper, we prove the following results.

**Theorem 1.1.** For any \( d \geq 2 \) and \( \lambda > \lambda_c(\mathbb{Z}) \), there exists \( c > 0 \) such that
\[
\lim_{n \to \infty} \inf_{T \in \Lambda(n, d)} P[\tau_T \geq c^n] = 1.
\]
In particular,
\[
\liminf_{n \to \infty} \inf_{T \in \Lambda(n, d)} \frac{\log E[\tau_T]}{n} \geq c.
\]

**Theorem 1.2.** Let \( d \geq 2 \), \( \lambda > \lambda_c(\mathbb{Z}) \), and \( (G_n)_{n \in \mathbb{N}} \) be a sequence of graphs with \( G_n \in \mathcal{G}(n, d) \). The distribution of \( \tau_{G_n}/E[\tau_{G_n}] \) converges to the unitary exponential distribution as \( n \) tends to infinity.

Theorem 1.2 is a (weak) way of exposing the metastability of the contact process (see part (3) of Proposition 3.2 for a finer statement; note also that from this distribution as \( \tau \) tends to infinity, one can replace \( \Lambda(n, d) \) by the set of all graphs having a subgraph in \( \Lambda(n, d) \), and in particular, one can replace \( \Lambda(n, d) \) by \( \mathcal{G}(n, d) \). For instance, the above results cover the case of any sequence of increasingly large connected subsets of \( \mathbb{Z}^d \). At the cost of requiring \( \lambda > \lambda_c(\mathbb{Z}) \), we thus recover and extend previously mentioned results, without any strong assumption on the regularity of the graph. For such values of \( \lambda \), this shows in particular that on regular trees with finite depth, the extinction time is actually exponentially large in the number of vertices.

This is however not quite the way in which we think our results are most useful. Rather, they are the basic ingredient of a general strategy to prove that the extinction time of the contact process is exponentially large in the number of vertices as soon as the infection parameter is above the natural critical value of the particular graphs we consider (instead of \( \lambda_c(\mathbb{Z}) \)). We now expose this strategy on certain random graphs whose degree distribution follows a power law. The case of finite homogeneous trees is discussed in [CMMV14].

We consider random graphs given by the configuration model with degree distribution equal to a power law. Let us explain what this means. For any \( n \in \mathbb{N} \), we construct a graph \( G_n \) on \( n \) vertices. The vertex set is simply \( \{1, \ldots, n\} \). The random set of edges will be constructed from a probability \( p \) on \( \{3, 4, \ldots\} \) with the property that, for some \( a > 2 \),
\[
0 < \lim \inf_{m \to \infty} m^a p(m) \leq \lim \sup_{m \to \infty} m^a p(m) < \infty.
\]
(We assume that \( p \) is supported on integers larger than 2 because in [vdH, Theorem 10.14], it is shown that under this assumption, it follows that \( G_n \) is connected and with probability tending to 1 as \( n \to \infty \).) We then let \( d_1, \ldots, d_n \) be independent random variables distributed according to \( p \). We want \( \sum_{i=1}^n d_i \) to be even, so if it is not, we simply add 1 to one of the \( d_i \) chosen uniformly at random. Next, from each vertex \( i \in \{1, \ldots, n\} \) we place \( d_i \) half-edges; when two half-edges are connected, an edge is formed. We pair up the \( d_1 + \cdots + d_n \) half-edges in a random way that is uniformly chosen among all possibilities.
Let us write $\mathbb{P}$ to denote a probability measure under which both the random graph and the contact process on this graph are defined. In [CD09], it is shown that, for any $\lambda > 0$ and any $\delta > 0$, we have $\mathbb{P}[\tau_{G_n} \geq e^{n^{1-\delta}}] \to 1$ as $n \to \infty$. In particular, in the limit when the number of vertices tends to infinity, the critical infection parameter for these graphs is 0. We strengthen this observation by showing:

**Theorem 1.3.** For any $\lambda > 0$, there exists $c > 0$ such that

$$\mathbb{P}[\tau_{G_n} \geq e^{cn}] \to 1 \text{ as } n \to \infty.$$  

Although it would be simple to deduce Theorem 1.3 from Theorem 1.1 assuming $\lambda > \lambda_c(Z)$, we stress again that Theorem 1.3 covers any non-zero infection parameter. We think that Theorem 1.3 is true for all $a > 1$, but we only give the proof for $a > 2$, which is the harder case (when we increase $a$, the degrees of the vertices become stochastically smaller, so the graph becomes less connected).

For finite boxes of $\mathbb{Z}^d$, the proof that the extinction time is exponential in the number of vertices relies on a coarse-graining argument. This coarse-graining enables to map the initial contact process into a coarse-grained (discrete-time) contact process with an increasingly large infection parameter. The remarkable feature of $\mathbb{Z}^d$ is its scale invariance, which ensures that the coarse-grained graph is still $\mathbb{Z}^d$ (or rather, a finite box of $\mathbb{Z}^d$). Now, simple percolation arguments show that on finite boxes of $\mathbb{Z}^d$, the time of survival is exponential in the number of vertices if the infection parameter is sufficiently large, say larger than $\lambda_{\text{target}}$. To sum up, the proof for finite boxes of $\mathbb{Z}^d$ consists in defining a coarse-graining scheme, and then in fixing a finite length of coarse-graining such that the coarse-grained system has infection parameter larger than $\lambda_{\text{target}}$.

For $G_n$ given by the configuration model, coarse-grained blocks will consist of certain stars, that is, vertices with a given large number of neighbours. The difficulty in trying to adapt the strategy to these graphs (or to finite homogeneous trees) is that there is no easy scale invariance as on $\mathbb{Z}^d$. It then becomes a very delicate matter to control the geometry of the coarse-grained graph, and thus to define a suitable equivalent of $\lambda_{\text{target}}$. However, Theorem 1.1 roughly tells us that we do not need to control this geometry, and that we can choose $\lambda_{\text{target}} = \lambda_c(Z)$.

The approach in [CD09] is also based on a coarse-graining procedure. There, the question of controlling the coarse-grained geometry was bypassed by choosing the coarse-grained scale so large as to ensure that the coarse-grained graph was a complete graph. Since the diameter of the graphs $G_n$ goes to infinity, this cannot be ensured unless the length scale of the coarse-graining diverges. In other words, in this approach, stars should have more and more vertices as $n$ increases, and the number of points in the coarse-grained scale must thus be $o(n)$. With our approach, we can choose instead a large but fixed size for the relevant stars in the graph, so that the coarse-grained graph still contains of order $n$ sites, and Theorem 1.3 follows from this construction.

We close this introduction by pointing to other related works and questions. To the best of our knowledge, the rigorous study of the behaviour of the contact process on random graphs with a power-law degree distribution began with [BBCS05]. The graph studied there is obtained by a generalization of the preferential attachment mechanism introduced in [BA99, BR04]. Vertices are added one by one. When a new vertex is added, it is connected to $m$ of the older vertices. These are chosen independently as follows: with probability $\alpha$, the vertex is chosen uniformly; with probability $1 - \alpha$, it is chosen with a probability proportional to its degree. That the degree distribution of these graphs follows a power law was proved in [CF03]. The structure of these graphs is more difficult to analyse than that of the configuration
model, and [BBCS05] could only show that the extinction time is at least $\exp(n^\beta)$ for some $\beta < 1$ depending on $\alpha$ (where $n$ is the number of vertices). It would be interesting to investigate whether the results presented here (possibly in combination with the recent [BBCS14]) enable to show that the extinction time is actually of the order of $\exp(cn)$ for some constant $c$. In another direction, [VM14] computes very precise asymptotics for the extinction time of the contact process on the complete graph.

The question of whether or not the critical infection parameter is zero for a given sequence of graphs still lacks a unified mathematical treatment. Results such as [DJ07, MV14] point towards the conjecture that the epidemic threshold of a sequence of random graphs is given by the lower critical infection parameter of the local graph limit of the sequence (see [MV14] for a more detailed discussion on this). The local limit of a sequence of random graphs given by the configuration model are Galton-Watson trees. However, we do not know of a good criterion for deciding whether the lower critical infection parameter of a given Galton-Watson tree is 0. It is suggested in [BCP13] that this threshold is zero as soon as the degree distribution decays slower than exponentially. Beyond this, physicists have investigated finite-size corrections to the asymptotic behaviour discussed here, see for instance [FCP12].

**Organization of the paper.** Section 2 is a brief reminder on some properties of the contact process that will be useful for our purposes. In Section 3, we show a weaker version of Theorem 1.1, which states that the expectation of the extinction time is larger than $e^{cn^\alpha}$ for some $\alpha > 0$. In order to do this, we consider two cases: either the tree contains a large segment, or it contains a large number of disjoint smaller segments. In the first case, the result follows from the known behavior of the extinction time on finite intervals of $\mathbb{Z}$. In the second case, we adapt an argument of [CD09] and show that, even if the segments are not too large, the time scale of extinction in individual segments is large enough for the infection to spread to other, possibly inactive, segments, so that the segments can jointly sustain activity for the desired amount of time. At this point, using a general metastability argument from [Mo93], we prove Theorem 1.2 (for completeness, we include a version of the argument of [Mo93] in an Appendix).

Given a tree $T \in \Lambda(n, d)$, we decompose it into two subtrees $T_1, T_2$ by removing an edge; we argue that this can be done so that $T_1$ and $T_2$ both contain a non-vanishing proportion of the vertices of $T$. In Section 4, we compare the contact process $(\xi_t)_{t \geq 0}$ on $T$ to a pair of processes $(\zeta_{T_1,t})_{t \geq 0}$ on $T_1$ and $(\zeta_{T_2,t})_{t \geq 0}$ on $T_2$. The process $\zeta_{T_1}$ evolves as a contact process on $T_1$ until extinction. Once extinct, the process stays extinct for some time, and then rises from the ashes (we call it a Phoenix contact process). This rebirth of the process reflects the fact that, as long as the true process $\xi$ has not died out, the tree $T_1$ constantly receives new infections that can restore its activity. The process $\zeta_{T_2}$ evolves independently, following the same rules. We show that the true process $\xi$ dominates $\zeta_{T_1} \cup \zeta_{T_2}$ up to the extinction of $\xi$, with probability close to 1. With this comparison at hand, we argue that, modulo a factor that is polynomial in the number of vertices, the expected extinction time for $T$ is larger than the product of the expected extinction times for $T_1$ and $T_2$. The polynomially growing error term can be washed away using the lower bound $e^{cn^\alpha}$ mentioned in the previous paragraph. We thus obtain part (1.3) of Theorem 1.1, from which we deduce part (1.2) using attractiveness.

In Section 5, we re-state some of the results explained above for a discrete-time version of the contact process. Finally, we prove Theorem 1.3 in Section 6.

**Notations.** For $x \in \mathbb{R}$, we write $\lfloor x \rfloor$ for the integer part of $x$. When talking about the size of a graph, we always mean its number of vertices.
2. A REMINDER ON THE CONTACT PROCESS

Graphical construction. We start this section by presenting the graphical construction of the contact process. Fix a graph $G = (V, E)$ and $\lambda > 0$. We take the following family of independent Poisson point processes on $[0, \infty)$:

\begin{align*}
(\mathcal{R}^x) &: x \in V \text{ with rate } 1; \\
(\mathcal{N}^e) &: e \in E \text{ with rate } \lambda.
\end{align*}

Let $H$ denote a realization of all these processes. Given $x, y \in V$, $s \leq t$, we say that $x$ and $y$ are connected by an infection path in $H$ (and write $(x, s) \leftrightarrow (y, t)$ in $H$) if there exist times $t_0 = s < t_1 < \cdots < t_k = t$ and vertices $x_0 = x, x_1, \ldots, x_{k-1} = y$ such that

- $R_x \cap (t_i, t_{i+1}) = \emptyset$ for $i = 0, \ldots, k - 1$;
- $\{x_i, x_{i+1}\} \in E$ for $i = 0, \ldots, k - 2$;
- $t_i \in N(x_i, x_{i+1})$ for $i = 1, \ldots, k - 1$.

Points of the processes $(\mathcal{R}^x)$ are called recovery marks and points of $(\mathcal{N}^e)$ are links; infection paths are thus paths that traverse links and do not touch recovery marks. $H$ is called a Harris system; we often omit dependence on $H$. For $A, B \subseteq V$, we write $A \times \{s\} \leftrightarrow B \times \{t\}$ if $(x, s) \leftrightarrow (y, t)$ for some $x \in A, y \in B$. We also write $A \times \{s\} \leftrightarrow (y, t)$ and $(x, s) \leftrightarrow B \times \{t\}$. Finally, given another set $C \subseteq V$, we write $A \times \{s\} \leftrightarrow B \times \{t\}$ inside $C$ if there is an infection path from a point in $A \times \{s\}$ to a point in $B \times \{t\}$ and the vertices of this path are entirely contained in $C$.

Given $A \subseteq V$, put

\begin{equation}
\xi_t^A(x) = \mathbf{1}_{\{A \times \{0\} \leftrightarrow (x, t)\}} \quad \text{for } x \in V, \ t \geq 0
\end{equation}

(here and in the rest of the paper, $\mathbf{1}$ denotes the indicator function). It is well-known that the process $(\xi_t^A)_{t \geq 0} = (\xi_t^A(H))_{t \geq 0}$ thus obtained has the same distribution as that defined by the infinitesimal generator (1.1). The advantage of (2.1) is that it allows us to construct in the same probability space versions of the contact processes with all possible initial distributions.

From now on, we always assume that the contact process is constructed from a Harris system, and will write $P_{G, \lambda}$ to refer to a probability measure under which such a system (on graph $G$ and with rate $\lambda$) is defined; we usually omit $G, \lambda$.

Time-shifted process. Given $s \geq 0$, $B \subseteq V$ and a Harris system $H$, define $(\xi_t^{B,s})_{t \geq s}$ by

\begin{equation}
\xi_t^{B,s}(x) = \mathbf{1}_{\{B \times \{s\} \leftrightarrow (x, t)\}} \quad \text{for } x \in V, \ t \geq s.
\end{equation}

It can then be readily checked that

\begin{equation}
\text{for all } A \subseteq V \text{ and } 0 \leq s \leq t, \ \xi_t^A = \xi_t^{A,s}.
\end{equation}

From this it follows that

\begin{equation}
\text{for all } A, B \subseteq V \text{ and } 0 \leq s \leq t, \text{ if } \xi_s^A = \xi_s^{B,s}, \text{ then } \xi_t^A = \xi_t^{B,s}.
\end{equation}

Attractiveness. An immediate consequence of (2.1) is that for all $A \subseteq V$ and $t \geq 0$, we have $\xi_t^A = \cup_{x \in A} \xi_t^x$. From this we obtain the attractiveness property of the contact process:

\begin{equation}
\text{for all } A \subseteq B \subseteq V \text{ and } t \geq 0, \ \xi_t^A \subseteq \xi_t^B.
\end{equation}

Using (2.4) and (2.5), we see that, if for some $x \in V$ and $s \geq 0$ we have $\xi_s^x = \xi_s^1$, then for any $A \subseteq V$ that contains $x$ and $t \geq s$, we must have $\xi_t^A = \xi_t^1$. Together
with a union bound, this yields
\begin{equation}
    P \left[ \xi_t^A \neq \emptyset, \xi_t^B \neq \emptyset \right] \leq \sum_{x \in A} P \left[ \xi_t^x \neq \emptyset, \xi_t^y \neq \emptyset \right].
\end{equation}

**Duality.** Fix $A \subseteq V$, $t > 0$ and a Harris system $H$. Let us define the *dual process* $(\hat{\xi}_s^{A,t})_{0 \leq s \leq t}$ by
\begin{equation}
    \hat{\xi}_s^{A,t}(y) = \mathbf{1}_{\{(y,t-s) \in A \times \mathbb{N} \text{ in } H\}}.
\end{equation}
If $A = \{x\}$, we write $(\hat{\xi}_s^{x,t})$. This process satisfies two important properties. First, its distribution (from time 0 to $t$) is the same as that of a contact process with same initial configuration. Second, it satisfies the *duality equation*
\begin{equation}
    \hat{\xi}_s^{A,t} \cap B \neq \emptyset \text{ if and only if } A \cap \hat{\xi}_s^{B,t} \neq \emptyset.
\end{equation}
In particular, $\hat{\xi}_t^x(x) = 1$ if and only if $\hat{\xi}_t^{x,t} \neq \emptyset$. We will also need the fact that
\begin{equation}
    \text{if } \xi_t^A = \xi_t^B \text{ and } \xi_t^{B,t} \neq \emptyset, \text{ then } \xi_t^A \cap B \neq \emptyset.
\end{equation}

**Contact process on an interval.** We now gather some classical results about the contact process on the graph with vertex set $\{0, \ldots, N\}$ (where $N \in \mathbb{Z}_+$) and nearest-neighbour edges. We start defining, for $x, y \in \{0, \ldots, N\}$ and the contact process on $\{0, \ldots, N\}$,
\begin{equation}
    \sigma_{x \rightarrow y}^N = \inf \{t : \xi_t^x(y) = 1\}.
\end{equation}

**Proposition 2.1.** For any $\lambda > \lambda_c(\mathbb{Z})$, there exist $\tau_1 > 0$ such that the following holds. For any $N$ and any $x \in \{0, \ldots, N\}$,
\begin{equation}
    P \left[ \max (\sigma_{x \rightarrow 0}^N, \sigma_{x \rightarrow N}^N) < \frac{N}{\tau_1} \right] > \tau_1
\end{equation}
and
\begin{equation}
    P \left[ \max (\sigma_{x \rightarrow 0}^N, \sigma_{x \rightarrow N}^N) \geq \frac{N}{\tau_1}, \xi_{N/\tau_1}^x \neq \emptyset \right] \leq e^{-N}.
\end{equation}

Additionally, for any $N$ and any $t \geq 0$,
\begin{equation}
    P \left[ \tau_{[0, \ldots, N]}^1 \leq t \right] \leq te^{-\pi_1 N}.
\end{equation}

This follows from the classical renormalization argument that compares the contact process with supercritical oriented percolation, see for instance the proof of [Li1, Corollary VI.3.22].

**Remark 2.2.** A crucial aspect of the contact process on an interval is the observation that for every starting point $x \in \{0, \ldots, N\}$, we have $\xi_t^x = \xi_t^y$ for every $t \geq \max (\sigma_{x \rightarrow 0}^N, \sigma_{x \rightarrow N}^N)$. In other words, when the contact process started from $x$ has reached the extremal points of the interval, it is coupled with the contact process that was started from the fully occupied configuration. To see this, assume that there exists an infection path $\gamma_1$ from $(x, 0)$ to $(0, \sigma_{x \rightarrow 0}^N)$ and an infection path $\gamma_2$ from $(x, 0)$ to $(N, \sigma_{x \rightarrow N}^N)$ (i.e. assume that the random times are finite). For any two sites $y$ and $z$ and $t \geq \max (\sigma_{x \rightarrow 0}^N, \sigma_{x \rightarrow N}^N)$, let $\gamma'$ be an infection path from $(y, 0)$ to $(z, t)$. Since the graph is an interval, the infection path $\gamma'$ is forced to intersect $\gamma_1$ or $\gamma_2$, and in both cases, this implies the existence of an infection path from $(x, 0)$ to $(z, t)$. 

Corollary 2.3. Assume $\lambda > \lambda_0(\mathbb{Z})$. Then, for any $N$, any $t \geq N/\tau_1$ and any non-empty $A \subseteq \{0, \ldots, N\}$,

\begin{align}
(2.12) \quad P_{\{0, \ldots, N\}, \lambda} \left[ \xi^A_t = \xi^F_t \right] & \geq \tau_1, \\
(2.13) \quad P_{\{0, \ldots, N\}, \lambda} \left[ \xi^A_t = \xi^F_t \neq 0 \right] & \geq \tau_1 - te^{-\tau_1 N}.
\end{align}

Additionally, for any $A \subseteq \{0, \ldots, N\}$ and any $t \geq N/\tau_1$,

\begin{align}
(2.12) \quad P_{\{0, \ldots, N\}, \lambda} \left[ \xi^A_t \neq 0, \xi^A_t \neq \xi^F_t \right] & \leq |A|e^{-N}.
\end{align}

Proof. (2.12) follows from fixing $x \in A$ and applying the comment that follows (2.5), Remark 2.2, (2.9) and (2.11). (2.13) follows from Remark 2.2, (2.6) and (2.10). \qed

In the course of the proof, we will need a slightly more technical result on contact processes on intervals. For the sake of clarity, we prefer to derive it now, although its motivation will become clear only later.

For $\lambda > 0$, $G = (V, E)$, $A \subseteq V$ and $t > 0$, define

\[ p(A, t) = p_G(A, t) = P_{G, \lambda} \left[ \xi^A_t = \xi^F_t \neq 0 \right]. \]

Obviously,

\[ p(V, t) = P \left[ \tau_G > t \right]. \]

Also, as a consequence of (2.8), for any $A, B \subseteq V$ and $t > 0$,

\[ P \left[ \xi^A \cap B \neq \emptyset \right] \geq P \left[ \xi^A = \xi^F \right] - P \left[ \xi^B = 0 \right] \geq p(A, t) + p(B, t) - 1. \]

Lemma 2.4. For any $\lambda > \lambda_0(\mathbb{Z})$, $N \in \mathbb{Z}_+$ and $t_0 \geq N/\tau_1$, the following holds.

(i.) For any $A \subseteq \{0, \ldots, N\}$ satisfying $p(A, t_0) \geq 1 - \epsilon$ and any $\kappa > 0$,

\[ P_{\{0, \ldots, N\}, \lambda} \left[ p(\xi^A_{t_0}, t_0) > 1 - \kappa \right] \geq 1 - 2e^{-\kappa N} + 2e^{-N}. \]

(ii.) For any non-empty $A \subseteq \{0, \ldots, N\}$ and any $\kappa > 0$,

\[ P_{\{0, \ldots, N\}, \lambda} \left[ p(\xi^A_{t_0}, t_0) > 1 - \kappa \right] \geq 1 - 2e^{-\kappa N} + 2e^{-N}. \]

Proof. Recalling the definition of the time-shifted process (2.2), we observe that

\[ P \left[ p(\xi^A_{t_0}, t_0) \leq 1 - \kappa \right] \leq \frac{1}{K} \sum_{B} P \left[ \xi^B_{t_0} = B \right] (1 - p(B, t_0)) \]

\[ = \frac{1}{K} \sum_{B} P \left[ \xi^B_{t_0} = B \right] \left( P \left[ \xi^B_{t_0} = 0 \right] + p(B, t_0) \right) \]

\[ = \frac{1}{K} \sum_{B} P \left[ \xi^B_{t_0} = B \right] \left( P \left[ \xi^B_{t_0} = 0 \right] + P \left[ \xi^B_{t_0} \neq \xi^F_{t_0} \right] \right) \]

\[ \leq \frac{1}{K} \left( P \left[ \xi^B_{t_0} = 0 \right] + \sum_{B} P \left[ \xi^B_{t_0} = B \right] P \left[ \xi^B_{t_0} \neq \xi^F_{t_0} \right] \right), \]

where we have used the Markov inequality in the first inequality and (2.3) in the last inequality. By (2.11) and (2.13) respectively,

\[ P \left[ \xi^B_{t_0} = 0 \right] \leq 2t_0e^{-\epsilon_1 N} \quad \text{and} \quad P \left[ \xi^B_{t_0} \neq \xi^F_{t_0} \right] \leq N e^{-N}, \]

so we get

\[ P \left[ p(\xi^A_{t_0}, t_0) \leq 1 - \kappa \right] \leq \frac{1}{K} \left( 2t_0e^{-\epsilon_1 N} + Ne^{-N} \right). \]

Parts (i.) and (ii.) now follow respectively from the definition of $p$ and (2.12). \qed
3. Metastability

In this section, we show that for \( \lambda > \lambda_c(\mathbb{Z}) \), the extinction time of the contact process is at least \( e^{an^\alpha} \) for some \( \alpha > 0 \), uniformly over \( T \in \Lambda(n, d) \). This is clear by Proposition 2.1 if the tree has diameter at least \( n^\alpha \). Else, we rely on a recursive decomposition of the tree into many subtrees each of which has size at least \( \sqrt{n} \).

We then pick long intervals (of logarithmic size) inside each of these trees, and study how these can sustain the infection. For a suitable choice of the parameters, the time it takes for such an interval, isolated from the rest of the graph, to turn extinct, is much larger than the time it takes for the infection to travel from one interval to another, since the diameter of the whole graph is assumed small. Over time, the number of infected intervals can be compared to a random walk on the integers with a drift to the right. Analyzing this walk gives the desired result. Using this construction, we also show metastability of the contact process, in the sense that after time \( n^2 \), either the contact process is extinct, or it is equal to the contact process that was started from full occupancy (and thus the initial configuration is forgotten). Since \( n^2 \) is much smaller than the extinction time, this is a quantitative way of presenting the metastability of the contact process, and indeed we will conclude the section by proving Theorem 1.2.

We begin with the following basic graph-theoretic observation, on which our recursive decomposition of the tree is based.

**Lemma 3.1.** For a tree \( T \in \Lambda(n, d) \), there exists an edge whose removal separates \( T \) into two subtrees \( T_1 \) and \( T_2 \) both of size at least \( \lfloor n/d \rfloor \).

**Proof.** Associate to each edge the value of the smallest cardinality of the two subtrees resulting from the edge’s removal. Let \( \{x, y\} \) be an edge having maximal value. We suppose that the subgraph \( T_y \) containing vertex \( y \) is the smaller and that the value of its subtree is no more than \( \lfloor n/d \rfloor - 1 \). Let the remaining edges of vertex \( x \) be \( \{x, x_1\}, \{x, x_2\}, \ldots, \{x, x_r\} \), where \( r \leq d - 1 \). Let \( T_j \) be the subtree containing \( x_j \) obtained by removing the edge \( \{x, x_j\} \), and let \( n_j \) be its cardinality. By maximality, all the \( n_j \) must be no more than \( \lfloor n/d \rfloor - 1 \), but we also have

\[
|T_y| = |T \setminus (\{x\} \cup T_1 \cup \cdots \cup T_r)| = n - (1 + n_1 + n_2 + \cdots + n_r) \leq \lfloor n/d \rfloor - 1.
\]

That is, \( n \leq (d - 1)(\lfloor n/d \rfloor - 1) + \lfloor n/d \rfloor \leq n - (d - 1) \), a contradiction (the case \( d = 1 \) being trivial). \( \square \)

**Proposition 3.2.** For any \( d > 0 \) and \( \lambda > \lambda_c(\mathbb{Z}) \), there exists \( \alpha > 0 \) and \( \tau_2 > 0 \) such that the following holds.

1. For any \( n \) large enough, any \( T \in \Lambda(n, d) \), any non-empty \( A \subseteq T \), one has

\[
P\left[ \tau^A \geq e^{an^\alpha} \right] \geq \tau_2.
\]

In particular, \( E[\tau^A] \geq \tau_2 e^{an^\alpha} \).

2. Moreover,

\[
P\left[ \tau \geq e^{an^\alpha} \right] \geq 1 - e^{-n^\alpha},
\]

where we recall that we write \( \tau \) as a shorthand for \( \tau_1^2 \).

3. For any \( n \) large enough, any \( G = (V, E) \in \mathcal{G}(n, d) \) and any \( A \subseteq V \),

\[
P_{G, \lambda} \left[ \xi^A_{n^2} \neq 0, \, \xi^A_{n^2} \neq \xi^A_{n^2} \right] \leq e^{-n^{\alpha/2}}.
\]

From now on, \( d \) is fixed and we consider a tree \( T \) of maximal degree \( d \) and size \( n \to \infty \). Let \( \beta > 0 \) be to be determined later, not depending on \( n \). Applying Lemma 3.1 repeatedly \( \beta \log n \) times, we obtain \( L_n = 2^\beta \log n \) disjoint subtrees each of size at least \( \frac{n}{(2d)^{\beta \log n}} \geq \sqrt{n} \), provided \( \beta \leq 1/(2 \log(2d)) \) (for clarity, we simply assume that
$L_n$ is an integer, without writing that the integer part should be taken). We write $T_1, \ldots, T_{L_n}$ for the trees thus obtained.

Since the tree $T$ has maximal degree bounded by $d$, so do the subtrees $(T_j)$. Now, the size of a tree with maximal degree $d$ is at most
\[
1 + d + \ldots + d^{\text{diam}} = \frac{d^{\text{diam}+1} - 1}{d - 1},
\]
where $\text{diam}$ denotes its diameter. As a consequence, for $n$ large enough, each $T_j$ must have a diameter at least $\frac{\log n}{4\log d}$, and thus contain a path of $\frac{\log n}{4\log d}$ distinct vertices. We write $I_j$ to denote such a path, which we identify with an interval of length $\frac{\log n}{4\log d}$.

In what follows, we will distinguish between the two possibilities:

(A) the diameter of $T$ is at least $n^\beta$,

(B) the diameter of $T$ is less than $n^\beta$.

Proof of parts (1-2) of Proposition 3.2. Assume that the tree $T$ satisfies (A). For part (1), by attractiveness, it suffices to consider initial configurations with a single occupied site $x$; we now fix $x$. Condition (A) ensures that one can find an interval of length $k \geq n^\beta$ on $T$; we write $[y, z]$ to denote such an interval, with $y$ and $z$ its endpoints. We also write $[x, y]$ to denote the shortest path (in graph distance) starting at $x$ and ending at $y$. Let $\sigma = \inf\{ t : \xi^x_t(y) = 1 \}$. Then,
\[
(3.1) \quad P \left[ \xi^x_{\exp(T_1, n^\beta/2)} \neq 0 \right] \geq P \left[ \sigma < \infty, \xi^y_{\sigma + \exp(T_1, n^\beta/2)} \neq 0 \right].
\]

Using (2.9) on the contact process on $[x, y]$, we have $P[\sigma < \infty] > \bar{\sigma}_1$. Using (2.12) on the contact process on $[y, z]$, we have
\[
P \left[ \xi^y_{\sigma + \exp(T_1, n^\beta/2)} \neq 0 \right] \geq \bar{\sigma}_1 - e^{-\tau_{2,1}n^\beta/2} - e^{-\tau_{1,k}}.
\]

Thus the right-hand side of (3.1) is larger than $\bar{\sigma}_1^2/2$ when $n$ is large enough, proving part (1). Part (2) follows from applying (2.11) to the restriction of the contact process on $[y, z]$.

We now assume that the graph satisfies (B), and adapt an approach due to [CD09]. For any $B \subseteq I_1$, we write $(\xi^B_{i,t})_{t \geq 0}$ for the contact process on $I_1$ with initial configuration $B$, that is,
\[
\xi^B_{i,t}(x) = I_{(B \times (0,+) + (x,t)) \cap t_i}, \quad x \in I_1, t \geq 0.
\]

When we write $(\xi^x_t)$ or $(\xi^{A^*}_{i,t})$ or $(\xi^{A\cap I_1}_{i,t})$ with $A \subseteq T$) with only the subscript corresponding to time, we mean the contact process on the whole graph $T$. Obviously, for any $A \subseteq T$,
\[
(3.2) \quad \xi^x_t \supset \cup_i \xi^{A\cap I_i}_{i,t} \quad \text{and} \quad (\xi^{A\cap I_i}_{i,t})_{t \geq 0} \text{ for } i = 1, \ldots, L_n \text{ are independent.}
\]

In accordance with the previous section, we write
\[
p_{i_1}(B, t) = P \left[ \xi^B_{i_1,t} = \xi^B_{i_1,t} \neq 0 \right], \quad B \subseteq I_1, t > 0.
\]

For $1 \leq i \leq L_n$, we say interval $I_i$ is good at time $t$ if $p_{I_i}(\xi_t \cap I_i, 2n^\beta/\tau_{1}) > 1 - n^{-2\beta}$. We interpret this as meaning that the set of infected sites $\xi_t \cap I_i$ is “large enough”. Here “large enough” means that, if this set is taken as the initial configuration for a contact process on $I_i$, then it gives probability higher than $1 - n^{-2\beta}$ to the event that the infection is sustained for time $2n^\beta/\tau_{1}$ and couples with the comparison process which evolves in $I_i$ starting from full occupancy.

We put $t_k = (2n^\beta/\tau_{1} + 1)$ and define $Y_{i,k}$ as the indicator function that interval $I_i$ is good at time $t_k$. We finally define $X_k = \sum_{i=1}^{L_n} Y_{i,k}$ as the number of good intervals at time $t_k$. 
Lemma 3.3. Assume $\beta < \frac{c_1}{20 \log d}$. Then, if $n$ is large enough, the following holds.

(i.) on $\{\xi_{t_k} \supset I_i\}$, $Y_{i,k} = 1$;

(ii.) for any $x \geq 0$, $P[X_{k+1} \leq x \mid \xi_{t_k}] \leq \mathbb{P} \left[ \text{Bin}(I_n, 2n^{-2\beta}) \geq x \right]$, where $\text{Bin}(m, q)$ denotes a Binomial random variable with parameters $m$ and $q$;

(iii.) on $\{\xi_{t_k} \neq \emptyset, X_k < L_n\}$, $P[X_{k+1} \geq X_k + 1 \mid \xi_{t_k}] \geq c_1^2/2$.

Proof. For (i.), from (2.14) we get that, on $\{\xi_{t_k} \supset I_i\}$,

$$p_{t_k}(I_i, 2n^{\beta}/\tau_1) > 1 - \frac{2n^{\beta}}{\tau_1} \cdot e^{-\tau_1 \log n / 4 \log d} > 1 - n^{-2\beta}$$

when $\beta < \frac{c_1}{20 \log d}$ and $n$ is large, so that $Y_{i,k} = 1$.

(ii.) will follow from (3.2) and the fact, which we now prove, that on $\{Y_{i,k} = 1\}$, $P[Y_{i,k+1} = 1 \mid \xi_{t_k}] > 1 - 2n^{-2\beta}$. This is a consequence of attractiveness and applying Lemma 2.4(i.) with the choice of variables:

$$A = \xi_{t_k} \cap I_i, \quad N = |I_i| = \frac{\log n}{4 \log d}, \quad t_0 = \frac{2n^{\beta}}{\tau_1}, \quad \epsilon = \kappa = n^{-2\beta}.$$

The first requirement that $t_0 > \frac{N}{\epsilon}$ is satisfied since $\frac{2n^{\beta}}{\tau_1} > \frac{\log n}{4 \log d}$ when $n$ is large. The second requirement that $p_{t_k}(A, t_0) \geq 1 - \epsilon$ holds since we restrict our attention to the event $\{Y_{i,k} = 1\}$. The conclusion of the lemma then reads

$$P[Y_{i,k+1} = 0 \mid \xi_{t_k}] \leq n^{-2\beta} + n^{2\beta} \left( \frac{4n^{\beta}}{\tau_1} \cdot e^{-\tau_1 \log n / 4 \log d} + \frac{\log n}{4 \log d} \cdot e^{-\tau_1 \log n / 4 \log d} \right)$$

and the right-hand side is smaller than $2n^{-2\beta}$ if $\beta < \frac{c_1}{20 \log d}$ and $n$ is large enough.

(iii.) is now proved.

Let us prove (iii.). On the event $\{\xi_{t_k} \neq \emptyset, X_k < L_n\}$, we can find $x$ and $i$ such that $\xi_{t_k}(x) = 1$ and $Y_{i,k} = 0$. Fix $y \in I_i$, and let $\sigma = \inf\{t > t_k : \xi_{t-k}(y) = 1\}$. Note that, since it is assumed that the diameter of the graph is less than $n^2$, we can use Proposition 2.1(i.) to get $P[\sigma < t_k + n^{\beta}/\tau_1] > \tau_1$. Then,

$$P[Y_{i,k+1} = 1 \mid \xi_{t_k}] \geq \tau_1 \cdot P \left[ p_{t_k} \left( \xi_{t-k+1}^{y, \sigma}, \frac{2n^{\beta}}{\tau_1} \right) > 1 - n^{-2\beta} \mid \xi_{t_k}, \sigma < t_k + n^{\beta}/\tau_1 \right].$$

Now, if $\sigma < t_k + n^{\beta}/\tau_1$, then $2n^{\beta}/\tau_1 > t_k + 1 - \sigma > n^{\beta}/\tau_1 > |I_i|/\tau_1$, so we can use Lemma 2.4(ii.) to conclude that the right-hand side is larger than

$$\tau_1 \left( \tau_1 - n^{-2\beta} \left( \frac{4n^{\beta}}{\tau_1} e^{-\tau_1 \log n / 4 \log d} + \frac{\log n}{4 \log d} \cdot e^{-\tau_1 \log n / 4 \log d} \right) \right) > \frac{3}{4} \tau_1$$

when $n$ is large enough. We can now end the proof as follows:

$$P[X_{k+1} > X_k \mid \xi_{t_k}] \geq P[Y_{i,k+1} = 1 \mid \xi_{t_k}] \geq P[X_{k+1} < X_k \mid \xi_{t_k}] \geq \tau_1^2/2$$

since, by part (ii.) and a union bound, $P[X_{k+1} < X_k \mid \xi_{t_k}] \leq L_n \cdot 2n^{-2\beta} = n^{\beta \log 2} \cdot 2n^{-2\beta} \to 0$ as $n \to \infty$. $\square$

The conclusion will now follow from the above lemma by a comparison with a random walk on $\mathbb{Z} \cap (-\infty, L_n]$ with a drift to the right. The necessary information on this drifted walk is contained in the following lemma.

Lemma 3.4. Let $(Z_t)_{t \in \mathbb{N}}$ be the random walk on $\mathbb{Z} \cap (-\infty, L_n]$ with transition probabilities

$$P[Z_{t+1} = x + k \mid Z_t = x < L_n] = \begin{cases} 0 & \text{if } k > 1, \\ \tau_1/2 & \text{if } k = 1, \\ e^{-n^{-\beta}} n^{-|k|/|k|} & \text{if } k \leq -1. \end{cases}$$
and
\[ P[Z_{i+1} = x + k \mid Z_i = L_n] = \begin{cases} 0 & \text{if } k \geq 1, \\ e^{-n^{-\beta}} \frac{n^{-|k|\beta}}{|k|!} & \text{if } k \leq -1. \end{cases} \]
Let also \( H_0 \) be the hitting time of \( Z_- = \mathbb{Z} \cap (-\infty, 0] \), and \( H_L \) be the hitting time of \( L_n \). For any \( n \) large enough and any \( x \leq L_n \), we have
\[ P[H_0 < H_L \mid Z_0 = x] \leq n^{-x\beta/2}. \]

Let us postpone the proof of this lemma, and see how it enables us to conclude. From Lemma 3.3(iii.), we learn that whatever the initial non-empty configuration, we have \( X_1 \geq 1 \) with probability bounded away from 0. On this event, we want to couple \((X_k)\) with the random walk of Lemma 3.4 started at \( X_1 \), so that \( X_{k+1} \geq Z_k \) for every \( 0 \leq k \leq H_0 \). In the r.h.s. of the inequality in Lemma 3.3(ii.), a Binomial random variable appears, while jumps to the left for \((Z_t)\) follow a Poisson random variable. Since a Bernoulli random variable of parameter \(-\log(1-p)\), it follows that Bin\((L_n, 2n^{-2\beta})\) is stochastically dominated by a Poisson random variable of parameter \(-n^{\beta} \log 2\log(1-1-p)\leq n^{-\beta}\).

This and Lemma 3.3(iii.) guarantee the existence of the coupling. With probability at least \( 1 - n^{-\beta/2} \geq 1/2 \), the random walk hits \( L_n \) before entering \( Z_- \).

Let \( \alpha = \frac{\log 2}{4} \beta \). The proof of part (1) will be complete if we can argue that starting from \( L_n \), with probability close to 1, the walk needs to hit and exit \( L_n \) at least \( e^n \) times before reaching \( Z_- \). Let us consider a sequence of \( e^n \) excursions from \( L_n \), and show that with high probability, none of them visits \( Z_- \). Each jump out of \( L_n \) is distributed according to a Poisson random variable of parameter \( n^{-\beta} \). By the exponential Chebyshev inequality, for \( n \) large enough we have \( \mathbb{P}[\text{Poi}(n^{-\beta}) > x] < 2e^{-x} \) for any \( x > 0 \). Thus, with probability tending to 1, the maximum over \( e^n \) such random variables does not exceed \( n^{2n} \leq L_n/4 \). In view of Lemma 3.4, given an excursion whose first step has size smaller than \( L_n/4 \), the excursion will visit \( Z_- \) with probability smaller than \( n^{-3L_n\beta/8} \leq e^{-2n^{\alpha}} \), and this finishes the proof of part (1).

As for part (2), the argument is similar, except that in this case \( X_0 = L_n \), and we couple with the random walk of Lemma 3.4 started from \( Z_0 = X_0 = L_n \). Consider \( e^n \) excursions from \( L_n \). None of these excursions has its first jump larger than \( n^{2\alpha} \leq L_n/4 \) with probability at least \( 1 - e^{n^{\alpha}} \cdot 2e^{-n^{\alpha}} > 1 - \frac{1}{4}e^{-n^{\alpha}} \). As noted above, given an excursion from \( L_n \) whose first step has size smaller than \( L_n/4 \), the excursion will visit \( Z_- \) with probability smaller than \( n^{-3L_n\beta/8} \leq \frac{1}{2} e^{-2n^{\alpha}} \), thus finishing the proof of part (2).

**Proof of Lemma 3.4.** Let \( h(x) = \mathbb{P}[H_0 < H_L \mid Z_0 = x] \), \( \tilde{h}(x) = n^{-x\beta/2} \), and let \( \mathcal{L} \) be the generator of the random walk:
\[ \mathcal{L}f(x) = \frac{\tau^2}{2} (f(x+1) - f(x)) + e^{-n^{-\beta}} \sum_{k=1}^{+\infty} \frac{n^{k\beta}}{k!} (f(x-k) - f(x)) \quad (x < L_n). \]
For \( x \in \mathbb{Z} \cap (0, L_n) \), we have \( \mathcal{L}h(x) = 0 \). On the other hand, for such \( x \), we have
\[ \mathcal{L}\tilde{h}(x) = \frac{\tau^2}{2} \left( n^{-\beta/2} - 1 \right) \tilde{h}(x) + e^{-n^{-\beta}} \sum_{k=1}^{+\infty} \frac{n^{k\beta}}{k!} (n^{k\beta/2} - 1) \tilde{h}(x) \]
\[ \leq \frac{\tau^2}{2} \left( n^{-\beta/2} - 1 \right) \tilde{h}(x) + \sum_{k=1}^{+\infty} \frac{n^{k\beta}}{k!} n^{k\beta/2} \tilde{h}(x) \]
\[ \leq \left[ \frac{\tau^2}{2} \left( n^{-\beta/2} - 1 \right) + e^{-n^{-\beta/2}} - 1 \right] \tilde{h}(x), \]
so $\mathcal{L}\hat{h}(x) \leq 0$ as soon as $n$ is large enough. As a consequence, $\mathcal{L}(h - \hat{h}) \geq 0$ on $\mathbb{Z} \cap (0, L_n)$. By the maximum principle,

$$\max_{\mathbb{Z} \cap (0, L_n)} (h - \hat{h}) \leq \max_{\mathbb{Z} \cup \{L_n\}} (h - \hat{h}) = 0,$$

and the lemma is proved. □

**Proof of part (3) of Proposition 3.2.** We continue with case (B), but considering that $T$ is the spanning tree of some graph $G = (V, E)$. For an arbitrary non-empty $A \subseteq V$, we wish to bound

$$P\left[\xi_{n^2}^A \neq \xi_{n^2}^\infty, \xi_{n^2}^A \neq 0\right].$$

The probability above is equal to $P[\exists \eta : \xi_{n^2}^{\eta}(y) \neq \xi_{n^2}^\infty(y), \xi_{n^2}^{\eta} \neq 0]$. For any fixed $y$, we will thus bound

$$P\left[\xi_{n^2}^{\eta}(y) \neq \xi_{n^2}^\infty(y), \xi_{n^2}^{\eta} \neq 0\right].$$

(3.3)

Letting $(\check{\xi}_{n^2})_{0 \leq \ell \leq n^2}$ be the dual contact process for time $n^2$ started with configuration $\{y\}$, we can rewrite this probability as

$$P\left[\xi_{n^2}^\infty(y) = 0, \check{\xi}_{n^2}^{y,n^2} \neq 0, \xi_{n^2}^\infty \neq 0\right].$$

As in the proof of part (1), we consider $X_k$ the number of good intervals at time $t_k = (2n^2/\tau_1)k$. By attractiveness, if an interval is good for the contact process on $T$, then it must be good for the contact process on $G$. Note that, for $H_L$ as in Lemma 3.4, a classical large deviation estimate on sums of i.i.d. random variables with an exponential moment gives us that

$$P[H_L > m] \leq e^{-\sqrt{m}};$$

and as a consequence,

$$P\left[\xi_{n^2}^\infty \neq 0, \xi_{n^2}^{\eta} \neq 0\right] \leq e^{-\sqrt{m}}.$$

(3.4)

Let $E_{3/4}$ be the event that starting from $A$ occupied, a proportion at least $3/4$ of all the intervals $(I_i)_{i \leq L_n}$ are good in $\xi_{n^2/2}^A$ (for simplicity, $n^2/2$ is assumed to be a multiple of $2n^2/\tau_1$). Arguing as in part (2), we see that once $X_k$ has reached $L_n$, the probability that it reaches a point below $3L_n/4$ before time $n^2/2$ is smaller than $e^{-n^2}$. Combining this with (3.4), we obtain

$$P\left[\xi_{n^2}^\infty \neq 0, E_{3/4}^c\right] \leq 2e^{-n^2},$$

where $E_{3/4}^c$ denotes the complement of $E_{3/4}$. Similarly, if we let $\hat{E}_{3/4}$ denote the event that for the dual process $\hat{\xi}_{n^2}^{y,n^2}$, a proportion at least $3/4$ of the intervals are good at time $n^2/2 - 2n^3/\tau_1$, then

$$P\left[\hat{\xi}_{n^2}^{y,n^2} \neq 0, \hat{E}_{3/4}\right] \leq 2e^{-n^2}.$$

Consider the event $\hat{E}_i$ defined by:

$$\hat{E}_i = \left\{\xi_{n^2/2}^A \times \{n^2/2\} \leftrightarrow \xi_{n^2/2}^{y,n^2} \times \{n^2/2 + 2n^3/\tau_1\} \text{ inside } I_i\right\}.$$

(see picture).

Let also $I$ be the set of indices $i$ such that $I_i$ is good both for the contact process at time $n^2/2$, and for its dual $\hat{\xi}_{n^2}^{y,n^2}$ at time $n^2/2 - 2n^3/\tau_1$. We have

$$P\left[\bigcap_{i \leq L_n} (\hat{E}_i)^c, E_{3/4}, \hat{E}_{3/4}\right] \leq P\left[\bigcap_{i \in I} (\hat{E}_i)^c, E_{3/4}, \hat{E}_{3/4}\right].$$
Given that $\mathcal{E}_{3/4}$ and $\hat{\mathcal{E}}_{3/4}$ both happen, at least $1/2$ of the intervals are good both for the contact process and its dual, or in other words, $|I| \geq L_n/2$. Moreover, the events $\mathcal{E}_{3/4}$ and $\hat{\mathcal{E}}_{3/4}$, and the set $I$, are independent of the state of the Harris system in the time layer $[n^2/2, n^2/2 + 2n^3/\tau_1]$. By the definition of being good and (2.15), we have $P[(\tilde{E})^c \mid i \in I] \leq 2n - 2\beta$. Note also that the events $(\tilde{E})^c$ are independent.

Hence

$$P\left[\bigcap_{i \leq L_n} (\tilde{E})^c, \mathcal{E}_{3/4}, \hat{\mathcal{E}}_{3/4}\right] \leq (2n^{-2\beta})L_n/2.$$  

Finally, note that when one of the $\tilde{E}_i$ happens, it must be that $A \times \{0\} \leftrightarrow (y, n^2)$, that is, $\xi_{n^2}(y) = 1$. We have thus proved that

$$P\left[\xi_{n^2}^A(y) = 0, \tilde{\xi}_{n^2} \neq 0, \xi_{n^2}^A \neq 0\right] \leq P\left[\xi_{n^2}^A(y) = 0, \mathcal{E}_{3/4}, \hat{\mathcal{E}}_{3/4}\right] + 4e^{-n^\alpha} \leq (2n^{-2\beta})L_n/2 + 4e^{-n^\alpha} \leq 5e^{-n^\alpha}.$$  

Recalling that the probability on the l.h.s. above is that appearing in (3.3), we have thus shown that

$$P\left[\xi_{n^2}^A \neq \xi_{n^2}, \xi_{n^2}^A \neq 0\right] \leq 5ne^{-n^\alpha}.$$  

For case (A), the reasoning is similar, only simpler, so we will just outline the steps involved in the proof. Let $I$ be an interval of length $n^3$ contained in $T$. We will say that $I$ is good at time $t$ if

$$P_I\left(\xi_t \cap I, \frac{2n}{\tau_1}\right) > 1 - e^{-n^\alpha/2}.$$  

Define $s_k = (2n/\tau_1)k$ and let $X_k$ be the indicator function of the event that $I$ is good at time $s_k$. Similarly to Lemma 3.3, we can then prove that

- on $\{\xi_{s_k} \neq 0\}$, $P[X_{k+1} = 1 \mid \xi_{s_k}] \geq \tau_1^2/2$;
- on $\{X_k = 1\}$, $P[X_{k+1} = 1 \mid \xi_{s_k}] \geq 1 - 2e^{-n^\alpha/2}$.

Using these facts, we can show that, for a small, non-empty $A$ and $y$,

- $P\left[\xi_{n^2}^A \neq 0, I \text{ is not good for } \xi_{n^2/2}^A\right] \leq e^{-n^\alpha}$,
- $P\left[\tilde{\xi}_{n^2} \neq 0, I \text{ is not good for } \tilde{\xi}_{n^2/2} - 2n^3/\tau_1\right] \leq e^{-n^\alpha}$.
and, conditioned on $I$ being good for both $\xi_{n^2/2}$ and $\check{\xi}_{n^2/2-2n^2}$, the probability of

$$\left\{ \xi_{n^2/2} \times \{ n^2/2 \} \leftrightarrow \check{\xi}_{n^2/2-2n^2} \times \{ n^2/2 + 2n/\tau_1 \} \text{ inside } I \right\}$$

is larger than $1 - 2e^{-n^{3/2}}$. The proof is then completed by summing over $y$. 

\[ \square \]

**Proof of Theorem 1.2.** The result follows from Lemma A.1 in the Appendix, using parts (2) and (3) of Proposition 3.2. 

\[ \square \]

4. Comparison with Phoenix contact processes

The aim of this section is to prove Theorem 1.1. To this end, we manufacture a “Phoenix contact process”. This process evolves as a contact process up to extinction, but has then the ability to recover activity. Separating a tree $T$ into $T_1$ and $T_2$ as in Lemma 3.1, we show that with high probability, the true contact process dominates the union of two Phoenix contact processes running independently on $T_1$ and $T_2$, as long as these two Phoenix contact processes are not simultaneously in the empty configuration. From this, we derive a recursive relation between $E[\tau_T]$ and the product $E[\tau_{T_1}]E[\tau_{T_2}]$, which enables us to conclude.

Let $T \in \Lambda(n, d)$. We say that the Harris system is trustworthy on the time interval $[0, n^4]$ if for any $(x, s) \in T \times [0, n^4/2]$, the following two conditions hold:

(C1) if $\xi^{x,s}_n$ survives up to time $n^4$, then $\xi^{x,s}_n = \xi^{x,s}_{n^4}$,

(C2) if $\xi^{x,s}_n$ survives up to time $s + 2n^2$, then it survives up to time $n^4$.

We say that the Harris system $H$ is trustworthy on the time interval $[t, t + n^4]$ if $\Theta_t H$ is trustworthy on the time interval $[0, n^4]$, where $\Theta_t H$ is the Harris system obtained by a time translation of $t$ [i.e. $(x, s) \leftrightarrow (y, v)$ in $\Theta_t H$ if and only if $(x, t + u) \leftrightarrow (y, t + v)$ in $H$].

For a given Harris system and for $(Y_t)_{t \in \mathbb{R}_+}$ a family of independent auxiliary random variables following a Bernoulli distribution of parameter $1/2$, independent of the Harris system, we define the Phoenix contact process $(\zeta_t)_{t \geq 0} = (\zeta_t)_{t \geq 0}$ on $\{0, 1\}^T$ as follows.

**Step 0.** Set $\zeta_0 = \mathbb{1}_0$ and go to Step 1.

**Step 1.** Evolve as a contact process according to the Harris system, up to reaching the state $\mathbb{1}_0$, and go to Step 2.

**Step 2.** Let $t$ be the time when Step 2 is reached. Stay at $\mathbb{1}_0$ up to time $t + n^4$ and

- if the Harris system is trustworthy on $[t, t + n^4]$ and $Y_t = 1$, then set $\zeta^{t,n^4}_t = \xi^{t,n^4}_t$ (where $\xi^{t,n^4}_t$ is the contact process started with full occupancy at time $t$ and governed by the Harris system), and go to Step 1;
- else, go to Step 2.

We say that the process is active when it is running Step 1; is quiescent when it is running Step 2. Note that after initialization, the process alternates between active and quiescent phases. If it happens that during Step 2, the Harris system is trustworthy on $[t, t + n^4]$ and $Y_t = 1$, but $\xi^{t,n^4}_t = \mathbb{1}_0$, we consider that the process is active at time $t + n^4$, and becomes inactive again immediately afterwards.

**Remark 4.1.** Note that since the time the process spends on state $\mathbb{1}_0$ is not exponentially distributed, $(\zeta_t)$ is not Markovian. It would however be easy to make the process Markovian, by enlarging its state space into $(\{0, 1\}^T \setminus \{\mathbb{1}_0\}) \cup (\{\mathbb{1}_0\} \times [0, n^4])$, so that when arriving in Step 2, the process is in the state $(\mathbb{1}_0, 0)$, and subsequently the second coordinate increases at unit speed.
Remark 4.2. The auxiliary randomization of $\zeta$ provided by the family $(Y_t)$ is a technical convenience, which guarantees that if $\zeta_t$ is quiescent at some time $t$, then with probability at least $1/2$ it remains so at least up to time $t + n^4$.

Remark 4.3. Each time the process becomes active again, its distribution at this time is that of $\xi_{n^4}$ conditioned on the event that the Harris system is trustworthy on the time interval $[0, n^4]$. We write $\nu$ to denote this distribution.

Our first goal is to ensure that the Phoenix contact process is quiescent at a given time $t$ with small probability, of order $1/E[\tau_T]$, with some polynomial multiplicative correction. This is achieved in Lemma 4.6. Two intermediate results are written as lemmas for later reference. Roughly speaking, Lemma 4.6 implies that the probability for two independent Phoenix contact processes running respectively on $T_1$ and $T_2$ to be both quiescent at a given time is bounded by $(E[\tau_{T_1}]E[\tau_{T_2}])^{-1}$.

Lemma 4.4. Let $T \in \Lambda(n, d)$. For any $n$ large enough and any $t$, the probability that the Harris system on $T$ is trustworthy on $[t, t + n^4]$ is larger than $1/2$.

Proof. It suffices to show the lemma for $t = 0$. We first consider condition (C1). By part (3) of Proposition 3.2, the probability that

\[ \forall z \in T, \; \xi_{n^4}^{z, n^4/2} \neq 0 \Rightarrow \xi_{n^4}^{z, n^4/2} = \xi_{n^4}^{1, n^4/2} \]

goes to 1 as $n$ tends to infinity. Let $(x, s) \in T \times [0, n^4/2]$, and assume that $\xi_{2, s}^{x, s}$ survives up to time $n^4$, that is,

\[ (x, s) \leftrightarrow T \times \{n^4\}. \]

Then there must exist $z \in T$ such that

\[ (x, s) \leftrightarrow (z, n^4/2) \leftrightarrow T \times \{n^4\}. \]

On the event (4.1), we thus have $\xi_{n^4}^{z, s} \geq \xi_{n^4}^{1, n^4/2}$. The converse comparison being clearly satisfied, we have in fact $\xi_{n^4}^{z, s} = \xi_{n^4}^{1, n^4/2}$. In order to show that condition (C1) is satisfied for any $(x, s) \in T \times [0, n^4/2]$ with probability tending to 1, it thus suffices to show that

\[ P\left(\xi_{n^4}^{1, n^4/2} \rightarrow 1 \text{ as } n \rightarrow \infty.\right) \]

In view of part (2) of Proposition 3.2, with probability tending to one, we have $\xi_{n^4}^{1, n^4/2} \neq 0$. On this event, by part (3) of Proposition 3.2, we also have $\xi_{n^4}^{1, n^4/2} = \xi_{n^4}^{2, s}$ with probability tending to 1, and thus (4.2) is proved.

We now turn to condition (C2). Note that the event $\xi_{s+2n^2}^{x, s} \neq 0$ can be rewritten as

\[ (x, s) \leftrightarrow T \times \{s + 2n^2\}, \]

and under such a circumstance, there must exist $z \in T$ such that

\[ (x, s) \leftrightarrow (z, s/n^2) \leftrightarrow T \times \{s + 2n^2\}. \]

It is thus sufficient to show that

\[ P\left(\exists z \in T, k \in \{0, \ldots, \lfloor n^2/2\rfloor\} : \xi_{(k+1)n^2}^{z, kn^2} \neq 0 \text{ but } \xi_{n^4}^{z, kn^2} = 0\right) \rightarrow 0 \text{ as } n \rightarrow \infty. \]

For a fixed $z \in T$ and integer $k$, we have by part (3) of Proposition 3.2 that

\[ P\left(\xi_{(k+1)n^2}^{z, kn^2} \neq 0 \text{ but } \xi_{(k+1)n^2}^{z, kn^2} \neq \xi_{(k+1)n^2}^{1, kn^2} \right) \leq e^{-n^{\alpha/2}}, \]

so the probability of the event

\[ \forall z \in T, k \in \{0, \ldots, \lfloor n^2/2\rfloor\} : \xi_{(k+1)n^2}^{z, kn^2} = 0 \text{ or } \xi_{(k+1)n^2}^{z, kn^2} = \xi_{(k+1)n^2}^{1, kn^2} \]

is bounded by $(E[\tau_{T_1}]E[\tau_{T_2}])^{-1}$.
tends to 1 as \( n \) tends to infinity. On the other hand, with probability tending to 1, \( \xi^k_1 \) survives up to time \( n^4 \), and \( \xi^k_1 \) is clearly dominated by \( \eta_{n^4}^{1,kn^2} \), for any \( k \leq \lfloor n^2/2 \rfloor \). On the conjunction of this event and the one described in (4.4), we thus have

\[
\forall z \in T, k \in \{0, \ldots, \lfloor n^2/2 \rfloor \}, \text{ either } \xi_{(k+1)n^2}^{z, kn^2} = 0 \text{ or } \xi_{n^4}^{z, kn^2} \geq \xi_{n^4}^k \\neq 0,
\]

and this proves (4.3). \( \square \)

For the next lemma, recall that \( \tau \) is the extinction time of the contact process started with full occupancy.

**Lemma 4.5.** For any \( s > 0 \), one has

\[
P[\tau \leq s] \leq \frac{s}{E[\tau]},
\]

Moreover, there exists a constant \( C \) such that for any \( T \in \Lambda(n, d), E[\tau] \leq e^{Cn} \).

**Proof.** Attractiveness of the contact process implies that for any \( r \in \mathbb{N} \),

\[
P[\tau \geq rs] \leq (P[\tau \geq s])^r.
\]

Since

\[
E[\tau] \leq s \sum_{r=0}^{+\infty} P[\tau \geq rs] \leq \frac{s}{1 - P[\tau \geq s]},
\]

it comes that

\[
P[\tau \geq s] \geq 1 - \frac{s}{E[\tau]},
\]

which proves the first part. For the second part, note that one can find \( C \) such that

\[
P[\tau \geq 1] \leq 1 - e^{-Cn}
\]

uniformly over \( T \in \Lambda(n, d) \). The conclusion thus follows from (4.6). \( \square \)

**Lemma 4.6.** For any \( n \) large enough, any \( T \in \Lambda(n, d) \) and any \( t \geq 0 \), one has

\[
P[\zeta = 0] \leq \frac{6n^6}{E[\tau]}
\]

**Proof.** Using Lemma 4.5 with \( s = n^6 \), it is clear that (4.8) holds for any \( n \) and any \( t \leq n^6 \). Note moreover that, writing \( \tau' \) for the extinction time of the contact process started from the distribution \( \nu \) defined in Remark 4.3, we have

\[
P[\tau' \leq n^6 - n^4] = P[\tau \leq n^6 \mid \text{Harris sys. trustworthy on } [0, n^4]] \leq \frac{2n^6}{E[\tau]}
\]

where we used Lemma 4.4 in the last step.

Suppose now that \( t > n^6 \), and consider the event \( \mathcal{E} \) defined by

\[
\exists s \in (t - n^6/2, t - n^6/4) \text{ such that } \zeta_s = 0.
\]

We write \( \tilde{\tau} \) for the first \( s \geq t - n^6/2 \) such that \( \zeta_s = 0 \). On the event \( \mathcal{E} \), we have \( \tilde{\tau} \leq t - n^6/4 \). The event \( \mathcal{E}' \) defined by

\[
\forall k \in \mathbb{N}, k < \lfloor n^2/4 \rfloor,
\]

Harris sys. not trustworthy on \([\tilde{\tau} + kn^4, \tilde{\tau} + (k + 1)n^4]\) or \( Y_{\tilde{\tau} + kn^4} \neq 1 \)

has probability smaller than \( (3/4)^{\lfloor n^2/4 \rfloor} \) by Lemma 4.4. When \( \mathcal{E} \) and \((\mathcal{E}')^c\) both hold, the process \( \zeta \) becomes active at some time \( t_{A} \in [t - n^6/2, t] \), and is distributed according to \( \nu \) at this time. Hence,

\[
P[\zeta_t = 0, \mathcal{E}] \leq P[\zeta_t = 0, \mathcal{E}, (\mathcal{E}')^c] + P[\mathcal{E}']
\]

\[
\leq P[\tau' \leq n^6/2] + P[\mathcal{E}'].
\]
Since \( P[\mathcal{E}'] \ll 1 / E[\tau] \) and in view of (4.9), we have indeed

\[
(4.10) \quad P[\zeta_t = 0, \mathcal{E}] \leq \frac{3n^6}{E[\tau]}
\]

for any large enough \( n \). It thus remains to bound

\[
(4.11) \quad P[\zeta_t = 0, \mathcal{E}^c].
\]

Let \( k \) be the first positive integer such that

\[
[t - n^6 / 2 + kn^4, t - n^6 / 2 + (k + 1)n^4].
\]

For the same reason as above, we may assume that \([a_k, b_k] \subseteq [t - n^6 / 2, t - n^6 / 4]\).

Since on the event \( \mathcal{E}^c \), the process \( \zeta \) remains active on the time interval \([a_k, b_k]\), and considering the definition of trustworthiness and of the Phoenix process, we know that \( \zeta_{b_k} = \xi_{a_k} \), and moreover, the latter random variable is distributed according to \( \nu \). Hence, up to a negligible event, the probability in (4.11) is bounded by

\[
P[\tau^c \leq n^6 / 2],
\]

and thus, using (4.9) again,

\[
(4.12) \quad P[\zeta_t = 0, \mathcal{E}^c] \leq \frac{3n^6}{E[\tau]}.
\]

The conclusion now follows, combining (4.10) and (4.12).

In order to justify that with high probability, the true contact process on \( T \) dominates the union of two independent Phoenix contact processes on the subtrees \( T_1 \) and \( T_2 \) until extinction, we need to make sure that if the contact process is alive in \( T_1 \) but not in \( T_2 \), it will try to infect \( T_2 \) many times in a short time interval (so that with high probability, a large-scale infection happens in \( T_2 \) before the Phoenix contact process in \( T_2 \) becomes active again). The next lemma ensures that any given vertex \( x \in T_1 \) is infected many times if the contact process remains alive for some short amount of time (we think of \( x \) as being the site bordering the cut of \( T \) into two pieces, so that having \( x \) infected gives the immediate opportunity to start an infection in \( T_2 \)).

**Lemma 4.7.** Let \( T \) be a tree with size at most \( n \) and maximal degree at most \( d \), and let \( x \in T \). Define recursively \( \gamma_0 = 0 \) and, for any \( i \in \mathbb{N} \),

\[
\gamma_{i+1} = \inf \{ t \geq \gamma_i + 2n^2 : \xi_t(x) = 1 \} \quad (+\infty \text{ if empty}).
\]

For \( n \) large enough, we have

\[
P[\gamma_{i+1} > n^4 / 2 \mid \xi_{i+1/2} \neq \emptyset] \leq e^{-n^2}.
\]

**Proof.** In view of part (1) of Proposition 2.1, for any non-empty \( A \subseteq T \), we have

\[
P[\exists s \leq \frac{n}{\tau_1} : \xi_s^A(x) = 1] \geq \tau_1.
\]

Let \( \mathcal{F}_i \) be the \( \sigma \)-field generated by \( \{\xi_t, t \leq \gamma_i\} \). By induction and the Markov property, we can thus show that for any \( k \in \mathbb{N} \),

\[
P[\gamma_{i+1} - (\gamma_i + 2n^2) \geq \frac{kn}{\tau_1}, \xi_{\gamma_i+2n^2+(k-1)n/\tau_1} \neq \emptyset \mid \mathcal{F}_i] \leq (1 - \tau_1)^k.
\]
Hence,
\[
P\left[\gamma_{n^2/8} > n^4/2, \xi_{n^4/2} \neq 0\right] = P\left[\sum_{i=0}^{n^2/8-1} (\gamma_{i+1} - (\gamma_i + 2n^2)) > n^4/4, \xi_{n^4/2} \neq 0\right]
\leq P\left[\sum_{i=0}^{n^2/8-1} B_i n/\tau_1 > n^4/4\right],
\]
where \((B_i)\) are independent geometric random variables of parameter \(1 - \tau_1\). For \(\lambda > 0\) small enough, we have
\[
e^{\phi(\lambda)} \equiv E[e^{\lambda B_1}] < +\infty,
\]
and we thus obtain
\[
P\left[\sum_{i=0}^{n^2/8-1} B_i > \tau_1 n^3/4\right] \leq \exp(\phi(\lambda)n^2/8 - \lambda\tau_1 n^3/4),
\]
which, together with part (1) of Proposition 3.2, proves the claim.

We are now ready to prove our coupling result.

**Proposition 4.8.** For \(n\) large enough, let \(T \in \Lambda(n, d)\) be split into two subtrees \(T_1, T_2\) as described by Lemma 3.1. Define the process \((\zeta_t)_{t \geq 0}\) by
\[
\tilde{\zeta}_t = \zeta_{T_1,t} \cup \zeta_{T_2,t} \quad (t \geq 0),
\]
where \(\zeta_{T_1}\) and \(\zeta_{T_2}\) are Phoenix processes defined on \(T_1\) and \(T_2\) respectively, using the Harris system on \(T\) together with two independent families of auxiliary random variables, independent of the Harris system. One has
\[
P \left[\forall t \leq \tau, \zeta_t \geq \tilde{\zeta}_t\right] \geq 1 - e^{-n^{3/2}}.
\]

**Proof.** Let \((\sigma_i)_{i \geq 1}\) be the sequence of (stopping) times when the process \(\zeta_{T_1}\) becomes quiescent. We start by showing that, for any \(i,\)
\[
P \left[\sigma_{i+n}^< < \zeta_{T_1,\sigma_i+n}^<, \xi_{\sigma_i+n}^> \neq 0\right] \leq e^{-n^{7/4}}.
\]
For some arbitrary \(x \in T_1\), consider the stopping times introduced in Lemma 4.7, but started with \(\gamma_0 = \sigma_i\), and let \(N\) be the largest index satisfying \(\gamma_N \leq \sigma_i + n^4/2\). By Lemma 4.7, we have
\[
P \left[N < n^2/8, \xi_{\sigma_i+n}^> \neq 0\right] \leq e^{-n^2}.
\]
Moreover, part (1) of Proposition 3.2 ensures that, for any \(j,\)
\[
P \left[\xi_{T_1}^{x,\gamma_j} \text{ survives up to time } \gamma_j + 2n^2 \mid \gamma_j < +\infty\right] \geq \tilde{\tau}_2,
\]
where \(\xi_{T_1}^{x,\gamma_j}\) denotes the contact process restricted to \(T_1\) started with \(x\) occupied at time \(\gamma_j\). We introduce the stopping times \(\tilde{\gamma}_j\) to deal with the fact that \(\gamma_j\) may be infinite. We let \(\tilde{\gamma}_j = \gamma_j\) if \(j \leq N, \tilde{\gamma}_{N+1} = \sigma_i + n^4/2 + 2n^2\), and then recursively, \(\tilde{\gamma}_{j+1} - \tilde{\gamma}_j = 2n^2\). We have
\[
P \left[\forall j \leq N, \xi_{T_1,\gamma_j+2n^2}^{x,\gamma_j} = 0, \xi_{\sigma_i+n}^> \neq 0\right]
\leq P \left[N < n^2/8, \xi_{\sigma_i+n}^> \neq 0\right] + P \left[\forall j \leq n^2/8, \xi_{T_1,\gamma_j+2n^2}^{x,\gamma_j} = 0\right]
\]
Since for any \(j\), we have \(\tilde{\gamma}_{j+1} \geq \tilde{\gamma}_j + 2n^2\), the events indexed by \(j\) appearing in the second probability on the r.h.s. of (4.16) are independent. Using also (4.14) and (4.15) (with \(\gamma_j\) replaced by \(\tilde{\gamma}_j\)), we thus arrive at
\[
P \left[\forall j \leq N, \xi_{T_1,\gamma_j+2n^2}^{x,\gamma_j} = 0, \xi_{\sigma_i+n}^> \neq 0\right] \leq e^{-n^2} + (1 - \tilde{\tau}_2)n^2/8.
\]
We now show that
\[(4.18) \exists j \leq N, \xi_{T_1,\sigma_i+2n^4} \neq 0 \Rightarrow \xi_{\sigma_i+n^4} \geq \xi_{T_1,\sigma_i+n^4}.\]
Indeed, in order for \(\xi_{T_1,\sigma_i+n^4}\) to be non-\(0\), it must be that the Harris system restricted to \(T_1\) is trustworthy on \([\sigma_i, \sigma_i+n^4]\). In this case, by the definition of trustworthiness, if there exists some \(j \leq N\) such that \(\xi_{T_1,\gamma_j+2n^4} \neq 0\), then it must be that
\[
\xi_{T_1,\sigma_i+n^4} = \xi_{T_1,\gamma_j+2n^4} \geq \xi_{T_1,\sigma_i+n^4}
\]
(the last two being equal when \(Y_{\sigma_i} = 1\), otherwise \(\xi_{T_1,\sigma_i+n^4} = 0\)). Since \(\xi_j(x) = 1\), it is clear that \(\xi_{\sigma_i+n^4} \geq \xi_{T_1,\sigma_i+n^4}\), thus justifying (4.18). This and (4.17) prove (4.13).

In order to conclude, we first show that \(\tau\) cannot be too large. It comes from (4.5) and (4.7) that
\[(4.19) P[\tau \geq n^4 e^{Cn}] \leq e^{-n^2},\]
where \(C\) can be chosen uniformly over \(T \in \Lambda(n, d)\). If \(\xi_{T_1}\) is active at time \(t\) and \(\xi\) dominates \(\xi_{T_1}\) at this time, then the domination is preserved during the whole phase of activity, since \(\xi_{T_1}\) is driven by a subset of the Harris system driving the evolution of \(\xi\). When \(\xi_{T_1}\) becomes quiescent, the domination is obviously preserved. As a consequence, if the domination of \(\xi_{T_1}\) by \(\xi\) is broken at some time, it must be when \(\xi_{T_1}\) turns from quiescent to active. We thus have
\[P[\exists t \leq \tau, \xi_t < \xi_{T_1,t}] = P[\exists i : \xi_{\sigma_i+n^4} < \xi_{T_1,\sigma_i+n^4} \text{ and } \xi_{\sigma_i+n^4} \neq 0].\]
Since \(\sigma_i+1 - \sigma_i \geq n^4\), on the event \(\tau \leq n^4 e^{Cn}\), there are at most \(e^{Cn}\) times when \(\xi_{T_1}\) turns from quiescent to active. Using (4.13), we thus obtain
\[P[\exists t \leq \tau, \xi_t < \xi_{T_1,t}] \leq P[\tau \geq n^4 e^{Cn}] + e^{Cn} e^{-n^{7/4}}.\]
The proposition is now proved, using (4.19) together with the fact that
\[P[\exists t \leq \tau, \xi_t < \xi_{\tilde T_2,t}] \leq P[\exists t \leq \tau, \xi_t < \xi_{T_1,t}] + P[\exists t \leq \tau, \xi_t < \xi_{T_2,t}].\]

**Corollary 4.9.** For \(n\) large enough, let \(T \in \Lambda(n, d)\) be split into two subtrees \(T_1, T_2\) as described by Lemma 3.1. We have
\[E[\tau_{T_2}] \geq n^{-9} E[\tau_{T_1}] E[\tau_{T_2}].\]

**Proof.** Let \(\tilde \sigma\) be the first time when \(\xi_{T_1}\) and \(\xi_{T_2}\) are simultaneously quiescent. By Proposition 4.8, for any \(t \geq 0\), we have
\[(4.20) P[\tau \leq t] \leq P[\tilde \sigma \leq t] + e^{-n^{3/2}}.\]
In view of Remark 4.2, at time \(\tilde \sigma\), both \(\xi_{T_1}\) and \(\xi_{T_2}\) remain quiescent for a time \(n^4\) with probability at least \(1/2\) (one of them just becomes quiescent at time \(\tilde \sigma\), while the other stays quiescent for a time \(n^4\) with probability at least \(1/2\)). As a consequence, for any \(t \geq 0\),
\[P[\tilde \sigma \leq t] \leq \frac{2}{n^4} \int_0^{t+n^4} P[\tilde \xi_s = 0] \, ds.\]
Since \(\xi_{T_1}\) and \(\xi_{T_2}\) are independent, and using Lemma 4.6, we thus obtain
\[(4.21) P[\tilde \sigma \leq t] \leq \frac{2}{n^4} (t + n^4) \frac{(6t^6)^2}{E[\tau_{T_1}] E[\tau_{T_2}]} = \frac{72n^8(t + n^4)}{E[\tau_{T_1}] E[\tau_{T_2}]}\]
Let us now fix
\[\hat t = \frac{2}{n^9} E[\tau_{T_1}] E[\tau_{T_2}].\]
Since we know from part (1) of Proposition 3.2 that \( \tilde{t} \) grows faster than any power of \( n \), (4.21) gives us that for \( n \) large enough,
\[
P [ \tilde{\sigma} \leq \tilde{t} ] \leq 1/4.
\]
In view of (4.20), we thus obtain
\[
P [ \tau \leq \tilde{t} ] \leq 1/4 + e^{-n^{3/2}} \leq 1/2,
\]
which implies that \( E[\tau] \geq \tilde{t}/2 \), and thus the corollary. \( \square \)

**Proof of Theorem 1.1.** Let \( \rho = 1 + 1/d \), and consider, for any \( r \in \mathbb{N} \), the quantity
\[
V_r = \inf_{n \in (\rho^{-1}/d, \rho)} \inf_{T \in \Lambda(n,d)} \log \frac{E[\tau(T)\mid T]}{|T|}
\]
Statement (1.3) of Theorem 1.1 will be proved if we can show that
\[
\liminf_{r \to \infty} V_r > 0
\]
Let \( r \) be a positive integer, and \( T \) be a tree of degree bounded by \( d \) and whose size lies in \( (\rho r, \rho r + 1) \). Since
\[
1 - \rho^{-1} = 1/(d + 1) < 1/d \text{ and in view of Lemma 3.1, for } r \text{ large enough,}
\]
we can split up \( T \) into two subtrees \( T_1, T_2 \) such that
\[
|T_1|, |T_2| \geq |T|(1 - \rho^{-1}).
\]
As a consequence,
\[
|T_1|, |T_2| \geq \rho^{-1}/d,
\]
and also
\[
|T_1| \leq |T| - |T_2| \leq |T| (1 - (1 - \rho^{-1})) \leq \rho^r,
\]
with the same inequality for \( T_2 \). Corollary 4.9 tells us that for \( r \) large enough,
\[
E[\tau(T)] \geq \frac{1}{|T|^9} E[\tau(T_1)] E[\tau(T_2)],
\]
that is to say,
\[
\log E[\tau(T)] \geq \log E[\tau(T_1)] + \log E[\tau(T_2)] - \log |T|^9.
\]
Observing that
\[
\log E[\tau(T_1)] + \log E[\tau(T_2)] \geq V_r (|T_1| + |T_2|) = V_r |T|,
\]
we arrive at
\[
(4.22) \quad \frac{\log E[\tau(T)]}{|T|} \geq V_r - \frac{\log |T|^9}{|T|}.
\]
Part (1) of Proposition 3.2 ensures that for \( r \) large enough, one has
\[
(4.23) \quad V_r \geq \frac{c}{\rho^r(1 - \alpha)}
\]
for some constant \( c > 0 \). Recalling that \( |T| \leq \rho^{r+1} \), we thus have
\[
\frac{\log |T|^9}{|T|} \leq V_r \rho^{r\alpha/2},
\]
and (4.22) turns into
\[
\frac{\log E[\tau(T)]}{|T|} \geq V_r \left(1 - \frac{1}{\rho^{r\alpha/2}}\right),
\]
for any large enough \( r \) and any tree whose size lies in \( (\rho^r, \rho^{r+1}) \). If the size of the tree lies in \( (\rho^r/d, \rho^r) \), then the inequality
\[
\frac{\log E[\tau(T)]}{|T|} \geq V_r
\]
is obvious, so we arrive at
\[ V_{r+1} \geq V_r \left( 1 - \frac{1}{\rho r^{\alpha/2}} \right). \]
Since \( V_r > 0 \) for any \( r \) large enough by (4.23), and
\[ \prod_r \left( 1 - \frac{1}{\rho r^{\alpha/2}} \right) > 0, \]
we have shown that \( \lim \inf_{r \to \infty} V_r = c > 0 \). This proves (1.3).

In order to prove (1.2), it suffices to note that, by Lemma 4.5,
\[ \sup_{T \in \Lambda(n,d)} P \left[ \tau \leq e^{cn/2} \right] \leq e^{-cn/4} \]
for \( n \) large enough. \( \square \)

5. Discrete time growth process

For comparison purposes, it is useful to consider a discrete-time analogue of the contact process; we will need to consider such a process in the next section. Though many different definitions may be proposed, we have decided on the following.

Fix \( p \in (0,1) \) and let \( \{ I_{r}(x,y) : r \in \{0,1,\ldots\}, x,y \in \mathbb{Z}, |x-y| \leq 1 \} \) be a family of independent Bernoulli(\( p \)) random variables. Fix \( \eta_0 \in \{0,1\}^\mathbb{Z} \) and, for \( r \geq 0 \), let
\[ \eta_{r+1}(x) = 1 \{ \exists y : |y-x| \leq 1, \eta_r(y)=1, I_{r}(y,x)=1 \}. \]

The following is standard.

**Proposition 5.1.** The above process is attractive and there exists \( p^{(1)}_c < 1 \) such that for \( p > p^{(1)}_c \) the process survives in the sense that, for any \( \eta_0 \neq 0 \),
\[ P[\eta_r \neq 0 \ \forall r] > 0 \]
and, if \( \eta_0 = 1 \), then \( \eta_r \) decreases stochastically to a non-zero limit.

This process generalizes to any locally finite graph \( G = (V,E) \) just as does the contact process: we take independent Bernoulli(\( p \)) random variables
\[ \{ I_{r}(x,y) : x,y \in V, \ \text{dist}(x,y) \leq 1 \} \]
(where \( \text{dist} \) denotes graph distance) and, given \( \eta_0 \in \{0,1\}^V \), let
\[ \eta_{r+1}(x) = 1 \{ \exists y : \text{dist}(x,y) \leq 1, \eta_r(y)=1, I_{r}(y,x)=1 \}, \quad r \geq 0, \ x \in V. \]

In particular, \( \{ \eta_r \}_{r \geq 0} \) has the self-duality property, and we can follow through the arguments of the preceding sections to arrive at:

**Proposition 5.2.** Let \( d \geq 2 \) and \( p > p^{(1)}_c \). There exists \( c > 0 \) such that
\[ \inf_{T \in \Lambda(n,d)} P[\tau_T \geq e^{cn}] \longrightarrow 1 \ \text{as} \ n \to \infty. \]
(Here, \( \tau_T \) is the extinction time for the discrete-time process on \( T \) started from full occupancy).
6. Extinction time for the configuration model

Let us briefly recall the definition of the random graph \( G_n = (V_n, E_n) \). We take \( V_n = \{1, 2, \ldots, n\} \) and suppose given a probability \( p(\cdot) \) on the positive integers greater than or equal to 3 with the property that there exist \( a > 2 \) and \( 0 < c_0 < C_0 \) such that, if \( m \) is large enough,

\[
(6.3) \quad c_0 < m^a \cdot p(m) < C_0.
\]

To generate \( G_n \), we first choose the degrees for the \( n \) vertices \( d_1, d_2, \ldots, d_n \), according to i.i.d. random variables of law \( p(\cdot) \) (and add 1 to one of the \( d_i \), if necessary, to make \( \sum d_i \) even). Given this realization, we choose the edges by first giving each vertex \( x \) \( d_x \) half-edges and then matching up the half-edges uniformly among all possible matchings, so that, say, a half-edge for vertex \( x \) matched with a half-edge of vertex \( y \) becomes an edge between \( x \) and \( y \). In [vdH, Theorem 10.14], it is shown that from the assumption that \( p \) is supported on integers larger than 2, it follows that \( G_n \) is a connected graph with probability tending to 1 as \( n \to \infty \).

We consider the contact process with parameter \( \lambda > 0 \) on \( G_n \). In order to do so, we need to slightly modify the generator given in (1.1) to accommodate the fact that the random graph obtained from the above distribution may have loops and parallel edges. We put

\[
(6.4) \quad \text{for every } x \in \xi_t, \quad \xi_t \to \xi_t \setminus \{x\} \quad \text{with rate } 1,
\]

\[
\text{for every } x \notin \xi_t, \quad \xi_t \to \xi_t \cup \{x\} \quad \text{with rate } \lambda \sum_{y \in \xi_t} |\{e \in E_n : x, y \in e\}|.
\]

By \( x, y \in e \) we mean simply that the extremities of \( e \) are \( x \) and \( y \) (this may now be true for more than one edge \( e \)). With this definition, loops have no effect on the dynamics and parallel edges are seen as independent channels for the transmission of the infection. The graphical construction defined in the beginning of Section 2 is compatible with (6.4) and requires no modification.

We will prove Theorem 1.3 under the assumption that \( a > 2 \), as mentioned in the Introduction. We will also assume that \( \lambda \) is small; this is not problematic to us because clearly it is sufficient to prove Theorem 1.3 for \( \lambda \) small enough.

We rely on the idea that the contact process is sustained for a long time in the vicinity of vertices of high degree ("stars"). Our approach is to exhibit a subgraph \( G'_n \) of \( G_n \) that has sufficiently many stars arranged so that none of them is very far from others, and to argue that \( G'_n \) provides a very fertile environment for the persistence of the infection.

Let us be a bit more specific in sketching the proof. Depending on \( \lambda \), we fix a degree threshold \( \kappa S \) – stars will be vertices with degree above \( S/2 \), a distance threshold \( D \) and a time scale \( \kappa \). Our subgraph \( G'_n \) is a (connected) tree composed essentially of \( O(n) \) stars and line segments connecting these stars (to be exact, \( G'_n \) will end up containing some other vertices which do not fall in either category, but let us ignore that for the moment). Each star is directly connected by segments to at most 3 other stars, and the lengths of the segments are all below \( D \). We then define a tree \( G''_n \) as a renormalized version of \( G'_n \) in a very natural way: stars of \( G'_n \) correspond to vertices of \( G''_n \) and segments of \( G'_n \) correspond to edges of \( G''_n \). In particular, the degrees in \( G''_n \) are bounded by 3. The contact process on \( G''_n \) then induces a discrete-time growth process \( (\eta_x)_x \) on \( G''_n \); roughly speaking, for a vertex \( x \) of \( G'_n \) we say that \( \eta_x(x) = 1 \) if the star of \( G'_n \) that corresponds to \( x \) has many infected vertices (or is infested, as we will write) at time \( \kappa \cdot r \). The parameters \( S, D \) and \( \kappa \) can be chosen so that, if a star is infested at time \( \kappa r \), then with high probability it keeps being infested until time \( \kappa(r + 1) \), and this in turn is long enough that, with high probability, the infection reaches nearby stars, which also become infested. For the growth process \( (\eta_x) \), this translates into saying that the
closure parameter \( p \) is very close to 1. We then apply Proposition 5.2 to \((\eta_r)\), thus concluding that it stays active for an amount of time that is exponential in the number of vertices of \( G''_n \), which as mentioned is \( O(n) \). The desired conclusion for the contact process on \( G_n \) is then immediate.

This section is organized as follows. We will first list the formal properties that we want for the subgraph \( G'_n \) endowed with a set of distinguished vertices \( J'_n \). We state in Proposition 6.1 that with high probability, large enough \( G'_n, J'_n \) satisfying these properties indeed exist. Next, we will show how this proposition implies Theorem 1.3. Finally, we will prove the proposition, showing how \( G'_n, J'_n \) can be constructed.

The first property on our list is

**Property 1:** \( G'_n \) is a (connected) tree (with no loops or multiple edges).

Before listing the other properties, we need some notation. Let \( \deg' \) denote the degree of a vertex in \( G'_n \), that is, \( \deg'(x) = |\{ y \in V'_n : \{x,y\} \in E'_n \}| \), and \( \text{dist}' \) denote the graph distance in \( G'_n \), that is, \( \text{dist}'(x,y) \) is the length of the minimal path from \( x \) to \( y \) contained in \( G'_n \). For \( x,y \in J'_n \), write \( x \sim y \) if the unique path contained in \( G'_n \) from \( x \) to \( y \) contains no vertices of \( J'_n \) other than \( x \) and \( y \). Let \( S \) be a large constant to be chosen later and \( D = \lambda^4 S \).

**Property 2:** \( \deg'(x) \geq \frac{S}{2} \) for all \( x \in J'_n \).

**Property 3:** \( \text{dist}'(x,y) \leq D \) for all \( x,y \in J'_n \) with \( x \sim y \).

**Property 4:** The graph \( G''_n = (V''_n, E''_n) \) given by

\[
V''_n = J'_n, \quad E''_n = \{ \{x, y\} : x, y \in J'_n, x \sim y \}
\]

is a (connected) tree with degree bounded by 3.

**Proposition 6.1.** For small enough \( \lambda > 0 \), if \( S \) is large enough (depending on \( \lambda \)), there exists \( \delta > 0 \) such that, with probability tending to 1 as \( n \to \infty \), \( G_n \) has a subgraph \( G'_n \) with a set of vertices \( J'_n \subset V'_n \) such that Properties 1-4 are satisfied and \( |J'_n| > \delta n \).

**Proof of Theorem 1.3.** Assume \( G'_n, J'_n \) are as in the above proposition and \( G''_n \) is as in Property 4. \( G'_n \) is thus a tree with more than \( \delta n \) vertices and degree bounded by 3.

We will couple the contact process \((\xi^t_r)_{t \geq 0} \) on \( G'_n \) (starting from full occupancy) and a discrete time growth process \((\eta_r)_{r \geq 0} \) on \( G''_n \) (again starting from full occupancy); this comes down to a coupling between the Harris system on \( G'_n \) and the Bernoulli random variables used to define the growth process. \((\eta_r)\) is to be thought of as a coarse-grained version of \((\xi^t_r)\). We define \( \kappa = e^{\lambda^4 S} \); one time unit for \((\eta_r)\) will correspond to a period of length \( \kappa \) for \((\xi^t_r)\). Our choice of parameters and Proposition 5.2 will guarantee that the extinction time for \((\eta_r)\) is exponential in \( n \), and the corresponding fact for \((\xi^t_r)\) will be immediate.

Given a vertex \( x \in V''_n \), we define the neighbourhoods

\[
N'(x) = \{ y \in V'_n : \text{dist}'(x,y) \leq 1 \},
\]

\[
N''(x) = \{ x \} \cup \{ y \in V''_n : x \sim y \}.
\]

For \( x,y \in V''_n \) with \( x \sim y \), let \( b(x,y) \) be the set of vertices of \( G'_n \) in the unique path from \( x \) to \( y \) (notice that this path has length less than \( D \)). Define

\[
\Lambda(x) = N'(x) \cup \left( \bigcup_{y \in N''(x), y \neq x} N'(y) \cup b(x,y) \right).
\]

Assume given a Harris system \( H \) for the contact process \((\xi^t_r)\). For fixed \( x \in V''_n \) and \( r \in \{0,1, \ldots \} \), we will now define an auxiliary process \((\Gamma[x,r] \colon r \kappa \leq t \leq (r + 1) \kappa)\)
on $\{0,1\}^{\Lambda(x)}$. For $rk \leq t \leq (r+1)k$ and $y \in \Lambda(x)$, we put
\[ \Gamma[x,r]_t(y) = 1 \{ 3z \in N''(x) : \xi'_{rk}(z) = 1, (z,kr) \leftrightarrow (y,t) \text{ inside } \Lambda(x) \text{ in } H \}. \]
In particular, note that $\Gamma[x,r]_{kr} = \xi''_{kr}$, $N''(x)$ and $\Gamma[x,r]_{kr} = 0$ on $\Lambda(x) \setminus N''(x)$
and $\Gamma[x,r]_t : rk \leq t \leq (r+1)k$ evolves as the contact process on $\Lambda(x)$.

Given a set of vertices $U$ and $\omega \in \{0,1\}^U$, we will say that $U$ is infested in $\omega$ if $|\{x \in U : \omega(x) = 1\}| \geq \frac{\lambda}{\beta k} |U|$.

**Lemma 6.2.** If $\lambda > 0$ is small enough, the following holds. For any $\sigma > 0$, there exists $S_0$ such that, if $S > S_0$,
\[ P \left[ \forall y \in N'''(x), N''(y) \text{ is infested in } \Gamma[x,r]_{(r+1)k} \mid N''(x) \text{ is infested in } \Gamma[x,r]_{rk} \right] > 1 - \sigma. \]

**Proof.** Lemma 3.1(ii.) and Lemma 3.2 in [MVY13] respectively imply
\[ P(\xi_{rk} \cap \Omega_{rk} \cap \xi'_{rk} \cap \xi''_{rk} = 1) = 1 \]
for each $rk \leq t \leq (r+1)k$. Arguing recursively, this implies that
\[ \Gamma[x,r]_{rk} \text{ and } N''(x) \text{ are disjoint, we have } \]
\[ P[\xi_{rk}] > 0 \text{ : if } S \text{ is large enough, then } \]
\[ N''(x) \text{ is infested in } \xi_{rk} \Rightarrow P[N''(y) \text{ infested in } \xi_{rk}] > 1 - e^{-c\lambda^2 S}; \]
(6.5) \[ \exists c > 0 : \text{ if } S \text{ is large enough, then } \]
\[ N''(x) \text{ is infested in } \xi_0 \Rightarrow P[N''(y) \text{ infested in } \xi_{rk}] > 1 - e^{-c\lambda^2 S} \]
(in fact, (6.5) and (6.6) are slightly different from the mentioned results in [MVY13],
but are readily seen to follow from their proof. We spare the reader the details.)
(6.5), (6.6) and a union bound then give the desired result. \[ \Box \]

We now define the Bernoulli random variables
\[ \{ I_{r:x,y} : x,y \in V_n'''(x), y \in N'''(x), r = 0,1,\ldots \} \]
from which the growth process will be defined. They will not be independent, but we
will be able to argue that the parameter $S$ may be chosen so that they stochastically
dominate a family of independent Bernoulli variables. Set $I_{r:x,y} = 1$ if one of the
following condition holds:
(C1) $N''(x)$ is infested in $\xi_{rk}$ and $N''(y)$ is in $\Gamma[x,r]_{(r+1)k}$;
(C2) $N''(x)$ is not infested in $\xi_{rk}$.
Set $I_{r:x,y} = 0$ otherwise. (Condition (C2) is only present to guarantee that the
parameters of the Bernoulli random variables are all close to 1).

Now define $(\eta_r)$ starting from $\eta_0 \equiv 1$ and as prescribed in (5.2). Assume that
$\eta_r(x) = 1$. Then, there exist a sequence $x_0, x_1, \ldots, x_{r-1}, x_r = x$ of vertices of $V_n'''$ such that $x_{i+1} \in N''(x_i)$ and $I_{r:x,x_{i+1}} = 1$ for each $i$. Since $\xi_0 \equiv 1$, $N''(x_0)$ is
infested in $\xi_{rk}$, so $I_{r:x_0,x_1} = 1$ can only hold if condition (C1) holds, and thus $N''(x_1)$ is
necessarily infested in $\xi_{rk}$. Arguing recursively, this implies that $N''(x_i)$ is infested
in $\xi_{rk}$ for each $i$ and, in particular, $N''(x)$ is infested in $\xi_{rk}$. This shows that, if
$\eta_r \neq 0$, then $\xi_{rk} \neq 0$.

Also note that, for fixed $\sigma > 0$, if we choose $S$ corresponding to $\lambda$ and $\sigma$ in
Lemma 6.2, and if $x_1, \ldots, x_m \in V_n'''$ are vertices such that the sets $\Lambda(x_1), \ldots, \Lambda(x_m)$
are disjoint, we have
\[ P[\eta_r(x_j) = 0, 1 \leq j \leq m \mid (\eta_s)_{0 \leq s < r}] \leq \sigma^m. \]
Now, using a result of Liggett, Schonmann and Stacey [LSS97] (see also Theorem B26 in [Li2]), given $p \in (0,1)$ we can choose $S$ large enough so that the measure of the field $\{ I_{r:x} : r > 0, x \in V_n''' \}$ stochastically dominates i.i.d. Bernoulli$(p)$ random variables. We then have

**Corollary 6.3.** For any $p > p_c(1)$, if $S$ is large enough, then $\{ \eta_r \}$ dominates a
growth process on $G_n'''$ defined from i.i.d. Bernoulli$(p)$ random variables.
This, the fact that $|V''| \geq \delta n$ and Proposition 5.2 give Theorem 1.3.

The rest of the section is devoted to the proof of Proposition 6.1. We start with some remarks concerning the random degree sequence $d_1, \ldots, d_n$. Recall that $S$ is the constant in the construction of $G'_n$; we will assume that $S$ is large enough in the sense of Corollary 6.3, and will often need to increase its value. Define

$$J_n = \{ x \in V_n : \deg(x) \geq S \}.$$

Let $\mu = \sum_{m=1}^{\infty} m \cdot p(m)$. Let us remark that, if the degrees are given by $d_1, \ldots, d_n$ and we choose a half-edge uniformly at random, then the probability that the corresponding vertex has degree $m$ is

$$\frac{m \cdot | \{ x : d_x = m \} |}{\sum_x d_x} \rightarrow \frac{m \cdot p(m)}{\mu} \text{ as } n \rightarrow \infty.$$

The probability $q(m) = m \cdot p(m)/\mu$ is called the size biased distribution. Recall that $c_0$ and $C_0$ are the constants of (6.3); by reducing $c_0$ and increasing $C_0$ if necessary, a comparison with an integral gives us, for $m$ large enough,

$$c_0 < m^{-a-1} \cdot p([m, \infty)) \leq C_0,$$

(6.7) we will also need the following facts.

**Lemma 6.4.** For large enough $S$, there exists $\epsilon > 0$ such that, with probability tending to 1 as $n \rightarrow \infty$,

(i) $| \{ x \in V_n : d_x > S \} | > \epsilon n$

and, for any $A \subseteq V_n$ with $|A| \leq \epsilon n$,

(ii) $\sum_{x \in A} d_x / \sum_{x \in V_n \cap A^c} d_x < S^{-a}$; (iii) $\sum_{x \in J_n \cap A^c} d_x / \sum_{x \in V_n} d_x > 1/2 \cdot c_0 S^{-(a-2)}$.

**Proof.** Assume that $S$ is large enough that

$$S^{-a} < \frac{c_0 \mu}{8} S^{-(a-2)}, \quad C_0 \mu S^{-2a} < \frac{S^{-a}}{2},$$

where $c_0, C_0$ are as in (6.3). Let $K = S^{2a}$ and $\epsilon < \frac{1}{4} p([K, \infty))$. Define the events

$$B_1 = \{ \{ x \in V_n : d_x \geq K \} > \epsilon n \}; \quad B_2 = \left\{ \frac{\mu}{2} n < \sum_{x \in V_n} d_x < \frac{3\mu}{2} n \right\};$$

$$B_3 = \left\{ \sum_{x \in J_n} d_x \geq \frac{7c_0 \mu}{8} n S^{-(a-2)} \right\}; \quad B_4 = \left\{ \sum_{x : d_x \geq K} d_x < n S^{-(a-2)} \right\}.$$

Since, for all $x \in [n]$, $\mathbb{E}(d_x) = \mu$, $\mathbb{P}(d_x \geq K) = p([K, \infty))$,

$$\mathbb{E}(d_x \cdot \mathbf{1}_{\{d_x \geq K\}}) = \mu q([K, \infty) < \mu c_0 K^{-(a-2)} = \mu c_0 S^{-2a} < \frac{S^{-a}}{2},$$

$$\mathbb{E}(d_x \cdot \mathbf{1}_{\{d_x \geq S\}}) = \mu q([S, \infty)) > \mu c_0 S^{-(a-2)},$$

we have $\mathbb{P}(B_1 \cap B_2 \cap B_3 \cap B_4) \rightarrow 1$ as $n \rightarrow \infty$. If $B_1$ occurs, since $S < K$, the event of (i) is satisfied. If $B_1$ and $B_4$ occur, we have

$$\sum_{x \in A} d_x < \sum_{x : d_x \geq K} d_x < n S^{-a}.$$
Now, if B₁, B₂ and B₄ occur, we have
\[
\frac{\sum_{x \in A} d_x}{\sum_{x \in V_n \cap A} d_x} = \frac{\sum_{x \in A} d_x}{\sum_{x \in V_n} d_x - \sum_{x \in A} d_x} \overset{(6.10), B_2}{{<}} \frac{S^{-a}n}{\frac{1}{2} \mu n - S^{-a}n} < S^{-a},
\]
since \( \mu \geq 3 \) and \( S^{-a} < \langle 1/2 \rangle \). If B₁, B₂, B₃ and B₄ occur, we also have
\[
\frac{\sum_{x \in J_n \cap A} d_x}{\sum_{x \in V_n} d_x} \overset{(6.10), B_2, B_3}{>} \frac{2c_\mu n S^{-(a-2)} - n S^{-a}}{\frac{3}{2} \mu n} \overset{(6.9)}{>} \frac{2c_\mu n S^{-(a-2)}}{\frac{3}{2} \mu n} = \frac{1}{2} c_\mu S^{-(a-2)}.
\]

□

In what follows, \( \epsilon \) is taken corresponding to \( S \) as in the lemma. Let us say that a degree sequence \( d = (d_1, \ldots, d_n) \) is robust if it satisfies (i.), (ii.) and (iii.). We will henceforth fix a robust sequence \( d \). We will write \( P_d \) to denote a probability measure under which the random graph \( G_n \) is constructed as follows: the degrees of the \( n \) vertices are deterministic, given by \( d \), and the half-edges are then matched in a manner that is chosen uniformly at random among all possibilities, as prescribed in the definition of the configuration model.

We now describe an alternative matching procedure that produces the same random graph. This procedure consists of matching the half-edges sequentially, pair by pair, so that, in each step, we are free to choose one of the half-edges involved in the matching, and the other is chosen at random. To be more precise, let us introduce some terminology. A semi-graph \( g = (V_n, H, E) \) is a triple consisting of the set of vertices \( V_n \), a set of half-edges \( H \) and a set of edges \( E \) (of course, if \( H = \emptyset \), then \( g \) is a graph). The degree of a vertex in a semi-graph is the number of its half-edges plus the number of edges that are incident to it. Given two half-edges \( h, h' \in H \), we will denote by \( h + h' \) a new edge produced by “attaching” \( h \) and \( h' \).

We will now show how to define a finite sequence of semi-graphs \( g_0, g_1, \ldots, g_k \) so that \( g_k \) is a graph with the desired distribution. \( g_0 = (V_n, H_0, E_0) \) is defined with \( E_0 = \emptyset \) and such that each vertex \( x \) has \( d_x \) half-edges. Assume \( g_i = (V_n, H_i, E_i) \) is defined and has half-edges. Fix an arbitrary half-edge \( h \in H_i \) (call this an elected half-edge) and randomly choose another half-edge \( h' \) uniformly in \( H_i \setminus \{h\} \). Then put \( g_{i+1} = (V_n, H_{i+1}, E_{i+1}) \), where \( H_{i+1} = H_i \setminus \{h, h'\} \) and \( E_{i+1} = E_i \cup \{h + h'\} \). When no half-edges are left, we are done, and the graph thus obtained is distributed as \( G_n \). Often, instead of updating the sets each time, say from \( H_i, E_i \) to \( H_{i+1}, E_{i+1} \), as above, we will hold the notation \( g = (V_n, H, E) \) and say (for example) that \( h, h' \) are deleted from \( H \) and \( h + h' \) is added to \( E \).

In each step of the above construction, we are free to indicate the elected half-edge. A full description of how to elect a half-edge given all previous steps in the construction (and thus the present state of the semi-graph) is an algorithm to construct the graph (or a subgraph of it, if we stop before exhausting all half-edges – this will be the case for us, since our objective is to construct the subgraph \( G'_n \)). We will present an algorithm that will help us find \( G'_n \) with high probability. The robustness property will come into play because we will have to deal with the set of half-edges after some matchings have been made.

Other than doing matchings, our algorithm writes labels on edges and vertices. Labels will serve both to guide the order of the matchings and to define \( G'_n \); once the algorithm stops running. At any given point while the algorithm is running, each edge may have no label at all or one of the labels: included, excluded or queued. The label queued will only be associated to vertices of \( J_n \). In the semi-graph \( g_0 \), from which the algorithm starts, there are no edges (as mentioned above) and vertices have no labels. Once the algorithm finishes running, included vertices and edges will be the ones which will constitute \( G'_n \).
A labelled semi-graph \( g = (V_n, \mathcal{H}, \mathcal{E}, \{\ell_x\}, \{\ell_e\}, \prec) \) is a semi-graph with labels \( \ell_x \) attached to some of the vertices \( x \), labels \( \ell_e \) associated to some of the edges \( e \) and a total order \( \prec \) on the set of queued vertices. It is worth remarking that since the algorithm only does matchings and labelling, it does not change the degree of any vertex. In particular, the definition of the set \( J_n \) does not change.

The algorithm repeatedly follows a subroutine called a pass, which is just a sequence of matchings of half-edges and labellings. When a new pass starts, it typically takes a queued vertex \( \bar{x} \in J_n \) and successively explores the graph around it by performing matchings. A successful pass is one in which, in fewer than \( S^{a-1} \) matchings, at least \( S/2 \) neighbours of \( \bar{x} \) and two new vertices of \( J_n \) are found. We interrupt the pass if any of the matchings it produces reveals a vertex that had been previously “seen” by the algorithm. This has to do with the requirement that \( G'_n \) be a tree.

We first define the pass, and later define the full algorithm. We will need the following values:

\[
N = S^{a-1}, \quad \epsilon' = \frac{\epsilon}{2N}, \quad \delta = \frac{\epsilon}{4N}.
\]
The pass
Input: \( g = (V_n, \mathcal{H}, \mathcal{E}, \{\ell_x\}, \{\ell_e\}, \prec) \) with at least one queued vertex.

(P1)
Let \( \bar{x} \) be the queued vertex of highest order. Change the label of \( \bar{x} \) to included.
If \( \bar{x} \) has less than \( \frac{S}{2} \) half-edges remaining:
The pass is stopped in status \( B_1 \).
ELSE:
Define the sets \( V^*, \mathcal{E}^* \) and \( \mathcal{H}^* \) as the sets of relevant vertices, edges and half-edges of the pass, respectively. Set \( V^* = \mathcal{E}^* = \emptyset \) and \( \mathcal{H}^* = \{ \text{half-edges attached to } \bar{x} \} \). Endow \( \mathcal{H}^* \) with a total order \( \prec^* \) chosen arbitrarily. Proceed to (P2).

(P2)
Let \( h \) be the half-edge of highest order in \( \mathcal{H}^* \). Choose another half-edge \( h' \) uniformly at random in \( \mathcal{H}\setminus\{h\} \) and let \( v' \) be the vertex of \( h' \). Delete \( h, h' \) from all sets that contain them (\( h \) from \( \mathcal{H} \) and \( \mathcal{H}^* \), \( h' \) from \( \mathcal{H} \) and possibly \( \mathcal{H}^* \)) and add \( h + h' \) to \( \mathcal{E}^* \) and \( \mathcal{E} \).
If \( v' \) has been placed in \( V^* \) earlier in this pass:
Give the label excluded to all vertices contained in \( V^* \) (this of course includes \( v' \)) and all edges contained in \( \mathcal{E}^* \). The pass is stopped in status \( B_2 \).
ELSEIF \( v' \) has been given a label by some earlier pass:
Give the label excluded to all vertices contained in \( V^* \) (this does not include \( v' \); the label of \( v' \) is left unchanged) and all edges contained in \( \mathcal{E}^* \). The pass is stopped in status \( B_2 \).
ELSE:
Add \( v' \) to \( V^* \).
If \( v' \) is not in \( J_n \):
Add the half-edges of \( v' \) to \( \mathcal{H}^* \) (note that at this point \( h' \) is no longer a half-edge of \( v' \)) so that, in the order \( \prec^* \), they have arbitrary order among themselves but lower order than all half-edges previously in \( \mathcal{H}^* \).
Proceed to (P3).

(P3)
IF \( |V^*| \geq N \):
Give the label excluded to all vertices of \( V^* \) and all edges of \( \mathcal{E}^* \). Stop the pass in status \( B_3 \).
ELSE IF \( V^* \) contains more than \( S/2 \) neighbours of \( \bar{x} \) and \( |V^* \cap J_n| \geq 2 \):
Give the label queued to the two vertices of \( J_n \) that were first included in \( V^* \); in the order \( \prec \), assign them lower order than all previously existing queued vertices and arbitrary order among themselves. Give the label included to all other vertices of \( V^* \) and all edges of \( \mathcal{E}^* \). Stop the pass in status \( G \).
ELSE:
Repeat (P2).

Output: updated labeled semi-graph.
Algorithm to find $G'_n$

**Input:** $g_0 = (V_n, \mathcal{H}_0, \mathcal{E}_0)$, the initial semi-graph with $n$ vertices, no edges, no labels and robust degree sequence $d = (d_1, \ldots, d_n)$.

**(A1)**

- **IF** there are no queued vertices:
  - Replace the label of all vertices that have been labelled so far and all edges that have been constructed so far by the algorithm with the label excluded (this re-labelling will be called a *clearing*).
  - Choose an arbitrary unlabelled vertex of $J_n$ and label it queued.

  Proceed to **(A2)**.

**(A2)**

- Run the pass. Proceed to **(A3)**.

**(A3)**

- **IF** there are more than $\delta n + 2$ queued vertices:
  - The algorithm succeeds and stops.

- ELSE **IF** the pass has been run more than $\epsilon' n$ times:
  - The algorithm fails and stops.

- ELSE:

  Proceed to **(A1)**.

We remark that, in step **(A1)**, if the algorithm requires the choice of an unlabelled vertex of $J_n$, it is always possible to find one such vertex. This follows from the fact that each pass labels at most $N$ vertices and the algorithm runs at most $\epsilon' n = \frac{\epsilon}{2} n$ passes; by the definition of robustness, $N \cdot \frac{\epsilon}{2} n = \frac{\epsilon}{2} n < |J_n|$.

Let us introduce some more notation. Let $i_\infty$ be the total number of passes the algorithm runs. For $1 \leq i \leq i_\infty$, let $\bar{x}(i)$ be the vertex of $J_n$ whose label is changed from queued to included in pass $i$. If pass $i$ ends in status $G$, then it gives the label queued to two new vertices $y_1, y_2$; we then write $D(\bar{x}(i)) = \{y_1, y_2\}$; if the pass ends in any other status, we write $D(\bar{x}(i)) = \emptyset$. Also define $Q_G = \{\bar{x}(i) : \text{Pass } i \text{ ends in status } G\}$.

Let $W_0 = 0$ and, for $i \geq 1$,

$$W_i = \text{Number of queued vertices at the time pass } i \text{ is complete},$$

$$X_i = \begin{cases} 1 & \text{if pass } i \text{ ends in status } G; \\ -1 & \text{otherwise.} \end{cases}$$

Assume that, at the time pass $i - 1$ ends, there are no queued vertices (so that $W_{i-1} = 0$). The algorithm then goes to step **(A3)**. Assume that the pass has been run for less than $\epsilon' n$ times; then the algorithm proceeds to **(A1)**. At this point, the algorithm performs a clearing, gives the label queued to an arbitrary vertex $\bar{x}$ of $J_n$ and starts performing pass $i$; this pass first re-labels $\bar{x}$ as included, then performs its matchings and either ends in status $G$ (in which case two new vertices are labelled queued, so that $W_i = 2$) or status $B_1, B_2$ or $B_3$ (in which case no new queued vertices are created and $W_i = 0$). Thus, when $W_{i-1} = 0$, $W_i = 1 + X_i$.

Now assume that, at the time pass $i - 1$ ends, $W_{i-1} > 0$ and assume once more that the algorithm does not terminate in **(A3)**, so that **(A1)** is reached. Then pass $i$ is immediately started, the queued vertex of highest order is re-labelled as included and again, either zero or two new queued vertices appear. Thus, when $W_{i-1} > 0$, $W_i = W_{i-1} + X_i$. 

These considerations show that, for $1 \leq i \leq i_\infty$, $W_i \geq \sum_{j=1}^i X_j$.

We define our subgraph $G'_n$ only on the event that the algorithm is successful. Define $V'_n$ and $E'_n$ as the set of vertices and edges that, at the moment the algorithm stops running, have the label included. Also define

$$J'_n = V'_n \cap \mathcal{Q}_c.$$  

We are now ready to prove

**Lemma 6.5.** $G'_n$ satisfies properties 1, 2, 3 and 4 and $|J'_n| > \delta n$.

**Proof.** Let $i_0 = 1 + \max\{i : W_i = 0\}$. Note that $V'_n$ and $E'_n$ are the sets of vertices and edges that are labelled as included in passes $i_0, i_0 + 1, \ldots, i_\infty$, since no clearing occurs between these passes.

- **Property 1:** Fix $x \in V'_n$ and let $i$ be the pass in which $x$ was labelled. We can then find vertices $y_1, \ldots, y_k \in J'_n$ such that $y_1 \in D(x(i_0)), y_2 \in D(y_1), \ldots, y_k \in D(y_k)$ and $x(i) \in D(y_k)$. Then, there is a path of included edges connecting $x(i_0)$ to $y_1$, $y_1$ to $y_2, \ldots, y_{k-1}$ to $y_k$, $y_k$ to $x(i)$ and finally, $x(i)$ to $x$. This shows that $G'_n$ is connected. The fact that it is a tree is guaranteed by the IF statement in (P2) in the definition of the pass: any edge that would produce a loop forces the pass to stop and is labelled excluded.

- **Property 2:** Fix $z \in J'_n$. There exists $i \geq i_0$ such that $z = x(i)$ and pass $i$ ends in status $\mathcal{G}$. This implies that at least $S/2$ neighbours of $z$ are in $V'_n$, that is, $\deg'(z) \geq S/2$.

- **Property 3:** If $y, z \in J'_n$ and $y \sim z$, we must either have $z \in D(y)$ or $y \in D(z)$; assume the former is the case and fix $i$ such that $y = x(i)$. Then, pass $i$ ends in status $\mathcal{G}$, so that at most $N$ edges are constructed in this pass and they form no cycles. Since each vertex has degree larger than 2, we have

$$\text{dist}'(y, z) \leq \lfloor \log_2 N \rfloor = \lfloor (a - 1) \log_2 S \rfloor < \lambda^3 S = D$$

when $S$ is large enough.

- **Property 4:** This follows immediately from the fact that $G'_n$ is a tree and each pass labels either zero or two vertices as queued.

- **$|J'_n| > \delta n.$** By the definition of $i_0$, pass $i_0$ ends in status $\mathcal{G}$, $W_{i_0} = 2$ and $W_{i} > 0$ for all $i \in \{i_0, \ldots, i_\infty\}$. Since the algorithm ends successfully, $W_{i_\infty} \geq \delta n + 2$, so there must be more than $\delta n + 2 - 2 = \delta n$ values of $i$ in $\{i_0 + 1, \ldots, i_\infty\}$ such that $X_i = 1$, and hence, $x(i) \in J'_n$.

In order to complete the proof of Proposition 6.1, we now need to prove that the probability that the algorithm ends successfully tends to 1 as $n \to \infty$. To this end, let us first treat individual passes.

**Lemma 6.6.** For $S$ large enough and $n$ large enough (depending on $S$), the following holds. Assume that, when the pass defines $\tilde{x}$, this vertex has more than $\frac{S}{2}$ half-edges. Then, the pass ends in status $\mathcal{G}$ with probability larger than $\frac{9}{10}$.

**Proof.** Let $\mathcal{L}$ be the set of labelled vertices at the moment the pass starts. Since the whole algorithm runs at most $e'n$ passes and each pass labels at most $N$ vertices, we have $|\mathcal{L}| < e'n \cdot N = en/2$. Assume the current pass has already made $k$ matchings, with $0 \leq k < N$, and has not yet terminated, and consider the set $V^*$ of vertices the pass has found in its exploration. We have $|\mathcal{L} \cup V^*| < en/2 + N$ and this is much smaller than $en$ when $n$ is large. We take the set $A$ in the definition of robustness as $\mathcal{L} \cup V^*$.

The half-edge chosen for the $(k+1)$-th matching then has probability:

1. smaller than $\frac{\sum_{x \in A} d_x}{\sum_{x \in V_n \cap A^c} d_x} < S^{-a}$ of belonging to a vertex of $A$;
(2) larger than \( \sum_{x \in J_n \cap A^c} d_x \) \( > \frac{1}{2} c_0 S^{-(a-2)} \) of belonging to a vertex of \( J_n \cap A^c \).

Using (1), the probability that the pass ends in status \( B_2 \) is less than \( N \cdot S^{-a} = S^{-1} \), which is less than \( 1/20 \) when \( S \) is large. Using (2), the probability that the pass ends in status \( B_3 \) is less than

\[
P \left( \text{Bin} \left( N, \frac{1}{2} c_0 S^{-(a-2)} \right) < 2 \right),
\]

which is also less than \( 1/20 \) when \( S \) is large, since the expectation of the Binomial is \( \frac{1}{2} c_0 S \).

\[\square\]

Now define \( Y_i = 1 \{ \text{Pass } i \text{ ends in status } B_1 \} \). If \( Y_i = 1 \), then \( X_i = -1 \). By the previous lemma, for any \( x_1, \ldots, x_{i-1}, y_1, \ldots, y_{i-1} \) we have

\[
Y_i = 1 \{ (x_j)_{j=1}^{i-1} = (x_j)_{j=1}^{i-1}, (Y_j)_{j=1}^{i-1} = (y_j)_{j=1}^{i-1}, Y_i = 0 \} > \frac{9}{10}.
\]

Let us now exclude the possibility that many passes end in status \( B_1 \).

**Lemma 6.7.** For \( S \) large enough, \( P _d \left[ \sum_{i=1}^{n} Y_i > \frac{1}{10} |\epsilon' n| \right] \to 0 \).

**Proof.** We start remarking that, for \( \{ Y_i = 1 \} \) to occur, there must exist a vertex \( x \in J_n \) such that

- the first time a half-edge of \( x \) is chosen for a matching occurs before pass \( i \);
- from this time to the beginning of pass \( i \), more than \( S/2 \) half-edges of \( x \) are chosen for matchings;
- \( x \) is the queued vertex of highest order when pass \( i \) starts.

Let \( h_1, \ldots, h_L \) be the sequence of half-edges chosen at random by the algorithm since the beginning of the first pass. We have \( L \leq c n \). By (ii.) of Lemma 6.4, at the moment \( h_j \) is chosen, the probability that it belongs to a vertex that has previously been “seen” by the algorithm (that is, a vertex that either has been labelled by an earlier pass or is in the set \( V^* \) of the current pass) is less than \( S^{-a} \). If this occurs, call it a wasted matching. For \( \{ \sum Y_i > (1/10)|\epsilon' n| \} \) to occur, more than \( \frac{1}{10} |\epsilon' n| \frac{S}{2} \) wasted matchings must occur. The probability of this is less than

\[
P \left( \text{Bin} \left( \lfloor \epsilon n \rfloor, S^{-a} \right) > \frac{1}{10} |\epsilon' n| \frac{S}{2} \right).
\]

If \( S \) is large, this probability vanishes as \( n \to \infty \) since \( \frac{S^{-a}}{S^{1/2}} = \frac{10}{7} \frac{1}{\sqrt{n+1}} = 20 \frac{1}{\sqrt{n}} \).

\[\square\]

**Proposition 6.8.** \( P _d \left[ W_{\lfloor \epsilon' n \rfloor} \geq \epsilon n \right] \to 1 \).

**Proof.** We start constructing auxiliary random variables \( X_1, \ldots, X_{\lfloor \epsilon' n \rfloor}, Y_1, \ldots, Y_{\lfloor \epsilon' n \rfloor} \) whose joint distribution is the same as that of \( X_1, \ldots, X_{\lfloor \epsilon' n \rfloor}, Y_1, \ldots, Y_{\lfloor \epsilon' n \rfloor} \). Given sequences \( (x_j)_{j=1}^{i-1}, (y_j)_{j=1}^{i-1} \) and \( s \in \{0, 1\} \), let

\[
\phi(s, (x_j)_{j=1}^{i-1}, (y_j)_{j=1}^{i-1}) = \begin{cases} -1 & \text{if } s \leq P _d \left[ X_i = -1 \mid (x_j)_{j=1}^{i-1} = (x_j)_{j=1}^{i-1}, (Y_j)_{j=1}^{i-1} = (y_j)_{j=1}^{i-1} \right] \\ 1 & \text{otherwise} \end{cases}
\]

Likewise, let

\[
\psi(s, (x_j)_{j=1}^{i-1}, (y_j)_{j=1}^{i-1}) = \begin{cases} 0 & \text{if } s \leq P _d \left[ Y_i = 0 \mid (x_j)_{j=1}^{i-1} = (x_j)_{j=1}^{i-1}, (Y_j)_{j=1}^{i-1} = (y_j)_{j=1}^{i-1} \right] \\ 1 & \text{otherwise} \end{cases}
\]

and set
(when we write only \(\phi(s), \psi(s)\), we mean the functions above for \(X_1\) and \(Y_1\), with no conditioning in the probabilities that define them). Let \(U_1, U_2, \ldots, V_1, V_2, \ldots\) be independent random variables with the uniform distribution on \((0, 1)\). Set \(X'_1 = \phi(U_1), Y'_1 = \psi(V_1)\) and recursively define, for \(1 < i < \epsilon'\),

\[
Y'_{i+1} = \psi \left( V_{i+1}, \{X'_{j}\}_{j=1}^i, \{Y'_{j}\}_{j=1}^i \right), \quad X'_{i+1} = \phi \left( U_{i+1}, \{X'_{j}\}_{j=1}^i, \{Y'_{j}\}_{j=1}^i \right).
\]

Now, clearly \(\{X'_i, Y'_i\}_{i=1}^{\epsilon' n}\) has the same distribution as \(\{X_i, Y_i\}_{i=1}^{\epsilon n}\). By (6.11), we have \(\{Y'_i = 0, X'_i \neq 1\} \subseteq \{U_i \leq \frac{1}{10}\}\). We can now estimate

\[
P_d \left[ W_{\epsilon n} < \frac{\epsilon'}{2} n \right] \leq P_d \left[ \sum_i Y_i > \frac{1}{10} \epsilon'n \right] + P \left[ \{i: Y_i = 0, X_i \neq 1\} > \frac{1}{5} \epsilon' n \right]
\]

\[
\leq P_d \left[ \sum_i Y_i > \frac{1}{10} \epsilon'n \right] + P \left[ \{i \leq \epsilon' n: U_i \leq \frac{1}{10}\} > \frac{1}{5} \epsilon' n \right].
\]

The first of these probabilities vanishes by Lemma 6.7, and the second by the Law of Large Numbers. \qed

**Appendix A. Metastability and Limit Exponential Distributions**

Here we state and prove a result that is required for the proof of Theorem 1.2. The contents of this appendix are very similar to Proposition 1.2 in [Mo93]; we have simply adapted that proposition to our setting and notation.

**Lemma A.1.** Let \((G_n) = (V_n, E_n))\) be a sequence of graphs and assume that, in each graph of the sequence, a graphical construction for the contact process with parameter \(\lambda\) is defined. Assume that there exist sequences of positive numbers \((a_n), (b_n)\) satisfying

1. \(\lim_{n \to \infty} a_n = \lim_{n \to \infty} b_n = \infty, \quad \lim_{n \to \infty} \frac{a_n}{b_n} = 0;\)

2. \(\lim_{n \to \infty} \sup_{A \subseteq G_n} P_{G_n, \lambda} \left[ \zeta_A \neq 0, \zeta_{A_n} \neq 1 \right] = 0;\)

3. \(\lim_{n \to \infty} P_{G_n, \lambda} [\tau_{G_n} < b_n] = 0.\)

Then, \(\tau_{G_n}/E[\tau_{G_n}]\) converges in distribution, as \(n \to \infty\), to the exponential distribution of parameter 1.

**Proof.** Fix \(\epsilon > 0\). For each \(n\), there exists a unique number \(w_n\) such that \(P[\tau_{G_n} \leq w_n] = \epsilon\). This follows from the fact that the distribution function of \(\tau_{G_n}\) is continuous, which in turn is a consequence of the fact that \(\tau_{G_n}\) is the hitting time of state 0 for the continuous-time Markov chain \((\zeta^{(n)})\). Note that, by (3), we have

\[
w_n \geq b_n \quad \text{for } n \text{ large enough.}
\]

We will now establish upper and lower bounds for \(E[\tau_{G_n}]\). Let us start with the upper bound, which is easier. Since for any \(m\),

\[
P \left[ \tau_{G_n} > m \cdot w_n \right] \leq (1 - \epsilon)^m,
\]

we have

\[
E[\tau_{G_n}] \leq w_n \cdot \sum_{m=0}^{\infty} P[\tau_{G_n} > m \cdot w_n] \leq w_n \cdot \sum_{m=0}^{\infty} (1 - \epsilon)^m \leq \frac{w_n}{\epsilon}.
\]

For the lower bound, for each \(n, m \geq 1\), define

\[
r_{n,m} = (m-1)(w_n-a_n), \quad s_{n,m} = (m-1)(w_n-a_n)+w_n, \quad J_{n,m} = [r_{n,m}, s_{n,m}],
\]

...
so that $|J_{n,m}| = w_n$ and $|J_{n,m} \cap J_{n,m+1}| = a_n$. For the contact process on $G_n$ and $A \subseteq G_n$, recall the notation

$$\xi_t^{A,s} = \{x : A \times \{s\} \leftrightarrow (x,t)\}, \quad \xi_t^{\mathbb{Z}^d,s} = \{x \in V_n \times \{s\} \leftrightarrow (x,t)\} \quad s \leq t.$$  

Define the events

$$E_{n,m} = \{\xi_t^{1_{\tau_n,m}} \neq 0\}, \quad F_{n,m} = \{\xi_t^{1_{\tau_n,m},m+1} = \xi_t^{1_{\tau_n,m}}\}.$$  

By the definition of $w_n$, we have $P[E_{n,m}] = 1 - \epsilon$. Also, (2) implies that the probability $P[E_{n,m} \cap (F_n,m)^c]$ (which does not depend on $m$) tends to $0$ as $n \to \infty$.

Now also fix $\delta < 1$. By (1) and (A.12), if $n$ is large enough we have $a_n/w_n < \delta$ and $P[E_{n,m} \cap (F_n,m)^c] < \delta$ for any $m$. Then,

$$P[\tau_G > m(1 - \delta) \cdot w_n] \geq P[\tau_G > m(w_n - a_n)]
\geq P[\tau_G > s_{n,m}]
\geq P[\cap_{i=1}^m (E_{n,i} \cap F_{n,i})]
\geq P[\cap_{i=1}^m E_{n,i}] - \sum_{i=1}^m P[E_{n,i} \cap (F_n,m)^c].$$

The first probability on the right-hand side is larger than $(1 - \epsilon)^m$ by the FKG inequality (the events $(E_{n,i})_{i \geq 1}$ are increasing). We thus get

$$P[\tau_G > m(1 - \delta) \cdot w_n] \geq (1 - \epsilon)^m - \delta m. \quad (A.15)$$

Then, for any integer $K > 0$,

$$\mathbb{E}[\tau_G] \geq (1 - \delta)w_n \cdot \sum_{m=1}^\infty P[\tau_G > m(1 - \delta) \cdot w_n]
\geq (1 - \delta)w_n \cdot \sum_{m=1}^K P[\tau_G > m(1 - \delta) \cdot w_n]
\geq (1 - \delta)w_n \cdot \sum_{m=1}^K [(1 - \epsilon)^m - \delta m].$$

Now, if $K$ is large enough and $\delta$ is small enough (depending on the earlier choice of $\epsilon$), the above finally gives our lower bound

$$\mathbb{E}[\tau_G] \geq (1 - 2\epsilon)\frac{w_n}{\epsilon}. \quad (A.16)$$

We are now ready to conclude. Fix $t > 0$. On the one hand, (A.16) and (A.13) give

$$P[\tau_G > t \cdot \mathbb{E}[\tau_G]] \leq P\left[\tau_G > \left[\frac{t(1 - 2\epsilon)}{\epsilon}\right] \cdot w_n\right] \leq (1 - \epsilon)^{\frac{t(1 - 2\epsilon)}{\epsilon}}. \quad (A.17)$$

On the other hand, (A.14) and (A.15) give

$$P[\tau_G > t \cdot \mathbb{E}[\tau_G]] \geq P\left[\tau_G > \left[\frac{t}{\epsilon(1 - \delta)}\right] (1 - \delta) \cdot w_n\right]
\geq (1 - \epsilon)\left[\frac{t}{\epsilon(1 - \delta)}\right] - \delta \cdot \left[\frac{t}{\epsilon(1 - \delta)}\right]. \quad (A.18)$$

With proper choices of $\epsilon$ and $\delta$, both the right-hand sides of (A.17) and (A.18) can be made arbitrarily close to $e^{-t}$. This completes the proof. \hfill \square
References


