Forum of Mathematics, Sigma (2026), **Vol. 14**:e6 1–85
doi:10.1017/fms.2025.10144

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Degree-penalized contact processes

Zsolt Bartha[1], Júlia Komjáthy[2] and Daniel Valesin[3]

[1]Alfréd Rényi Institute of Mathematics, Hungary; E-mail: bartha@renyi.hu
[2]Delft University of Technology, Netherlands; E-mail: j.komjathy@tudelft.nl
[3]University of Warwick, UK; E-mail: daniel.valesin@warwick.ac.uk (Corresponding author)

Received: 20 February 2024; **Revised:** 1 September 2025; **Accepted:** 1 September 2025

2020 Mathematics Subject Classification: *Primary* – 82C22; *Secondary* – 60K35, 05C80, 60J85

Abstract

In this paper we study degree-penalized contact processes on Galton-Watson (GW) trees and the configuration model. The model we consider is a modification of the usual contact process on a graph. In particular, each vertex can be either infected or healthy. When infected, each vertex heals at rate one. Also, when infected, a vertex u with degree d_u infects its neighboring vertex v with degree d_v with rate $\lambda/f(d_u, d_v)$ for some positive function f. In the case $f(d_u, d_v) = \max(d_u, d_v)^\mu$ for some $\mu \geq 0$, the infection is slowed down to and from high-degree vertices. This is in line with arguments used in social network science: people with many contacts do not have the time to infect their neighbors at the same rate as people with fewer contacts.

We show that new phase transitions occur in terms of the parameter μ (at $1/2$) and the degree distribution D of the GW tree.

- When $\mu \geq 1$, the process goes extinct for all distributions D for all sufficiently small $\lambda > 0$;
- When $\mu \in [1/2, 1)$, and the tail of D weakly follows a power law with tail-exponent less than $1 - \mu$, the process survives globally but not locally for all λ small enough;
- When $\mu \in [1/2, 1)$, and $\mathbb{E}[D^{1-\mu}] < \infty$, the process goes extinct almost surely, for all λ small enough;
- When $\mu < 1/2$, and D is heavier than stretched exponential with stretch-exponent $1 - 2\mu$, the process survives (locally) with positive probability for all $\lambda > 0$.

We also study the product case, where $f(d_u, d_v) = (d_u d_v)^\mu$. In that case, the situation for $\mu < 1/2$ is the same as the one described above, but $\mu \geq 1/2$ always leads to a subcritical contact process for small enough $\lambda > 0$ on all graphs. Furthermore, for finite random graphs with prescribed degree sequences, we establish the corresponding phase transitions in terms of the length of survival.

Contents

1 Introduction **2**
1.1 Degree-penalized infection processes: main definitions 3
1.2 Definition of the underlying graphs 6
2 Results **9**
2.1 Background, discussion and overview of proof techniques 13
3 Preliminaries **16**
3.1 Graphical construction of the contact process 16
3.2 Genealogic branching random walks 17

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

4 Extinction proofs via particle counting and martingales **21**
4.1 Product penalty: global extinction for all graphs when $\mu \geq 1/2$ via martingales 21
4.2 Product-penalty: fast extinction on the configuration model when $\mu \geq 1/2$ 23
4.3 Max-penalty: global extinction for all graphs when $\mu \geq 1$ 23
4.4 Max-penalty: fast extinction on the configuration model when $\mu \geq 1$ 23
4.5 Max-penalty: Loop erasure in particle counting when $\mu \in [1/2, 1)$ 24
4.6 Max-penalty: global extinction on trees when growth is limited 29
4.7 Max-penalty: fast extinction when $\mu \in [1/2, 1)$ 31
5 The configuration model: fast extinction via loop erasure **34**
6 Proofs of survival on Galton-Watson trees **45**
6.1 Max-penalty: global survival via infinite infection rays on heavy tailed GW trees 46
6.2 Product penalty: local survival using a row of star-graphs when $\mu < 1/2$ 48
6.2.1 Embedding stars in the Galton-Watson tree 49
6.2.2 Contact process on an infinite line of stars 52
6.2.3 Local survival through renormalization 55
7 The configuration model: k-cores sustain the infection when stars do not **59**
7.1 The subgraph spanned on low-degree vertices contains a k-core 60
8 The configuration model: survival through a network of stars **69**
A Appendix: Proofs of technical lemmas **73**
A.1 Proof of the statement in Example 1.13 73
A.2 Proof of long survival on stars 74
A.3 Proofs about the degrees in the configuration model 77
References **82**

1. Introduction

The contact process (CP) is a model for epidemics on graphs, described by a continuous-time Markovian dynamics, in which each vertex is in one of two states: infected or healthy. Infected vertices infect each of their healthy neighbors with a constant rate λ, while also healing at a constant rate 1. The model was first introduced by Harris in 1974 [32], who studied it on the integer lattice. Since then, much work has been done to characterize the behavior of the process also on infinite trees and locally tree-like finite graphs. The focus of this line of research has been to establish phase transitions in the long-term behavior of the process, as the spreading rate λ varies. A series of works [46, 57, 65] showed that the process on the infinite d-ary tree ($d \geq 2$), with an initial infection at the root, has three possible phases separated by two critical values $0 < \lambda_{c,1} < \lambda_{c,2}$: when $\lambda < \lambda_{c,1}$ the process undergoes eventual extinction, when $\lambda \in (\lambda_{c,1}, \lambda_{c,2})$ there is "global but not local" survival, and when $\lambda > \lambda_{c,2}$ there is "local" survival of the infection (see Definition 1.3). More recently, studying the process on Galton-Watson trees, the combination of the results in [34] and [6] showed that models with exponentially decaying offspring distributions always have an extinction phase ($\lambda_{c,1} > 0$), whereas subexponentially decaying offspring distributions lead to local survival for any positive value of λ due to the persistence of the infection around high-degree vertices, that is, $\lambda_{c,1} = \lambda_{c,2} = 0$ in this case.

Motivated by the latter results, we introduce a variant of the original contact process, where we slow down the spread of the infection around high-degree vertices in a degree-dependent way, in order not to let "superspreaders" scale up the infection rate linearly in their degree. Our results show that this change in the dynamics can reveal *topological features* of the graphs hidden from the classical versions, whose behaviours tend to depend strongly on the highest-degree vertices. Further, it allows us to observe different phases of the process *on the same underlying graph* caused by only a slight change in the process dynamics.

Our results, informally. In the degree-dependent contact process, the total infection rate from a high-degree infected vertex shall only grow polynomially with its degree, with an exponent less than one. Gradually increasing the penalty on the infection rate, we prove that the new process qualitatively differs

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

from the classical version. In particular, we obtain new phase diagrams for Galton-Watson trees: as soon as the total infection rate from a high-degree vertex scales less than the square root of its degree, high-degree vertices no longer maintain the infection, but their local surroundings heal quickly, and the process shows local extinction for small λ, yielding $\lambda_{c,2} > 0$, on *any* tree in fact (not just Galton-Watson trees). On Galton-Watson trees, if the offspring distribution is sufficiently heavy tailed (i.e., heavier than $x^{-\alpha_c}$ for some critical α_c depending on the degree-dependent penalty on the infection rate), then the degree-penalized CP survives globally but not locally (i.e., $\lambda_{c,1} = 0$ but $\lambda_{c,2} > 0$). However, if the tail is lighter, i.e., the offspring distribution has finite α_c-th moment (with $\alpha_c < 1$), then CP has an extinction phase (i.e., $\lambda_{c,1} > 0$). Here we find it surprising that subexponential distributions as heavy as infinite mean power laws can also show extinction. We also establish the corresponding phase diagrams for large finite random graphs with prescribed degree distributions (the configuration model), in terms of the length of time the infection survives on them. Here, tree-based recursion techniques break down, and we develop new methods to treat the extinction phase when $\lambda_{c,1} > 0$, which work as soon as the offspring distribution has finite variance. In the phase when high-degree vertices no longer maintain the infection for a long time, but the Galton-Watson tree show global survival for small $\lambda > 0$, we find new structures – k-cores existing on constant degree vertices only – that maintain the infection globally on the graph for a long time. All our results are also valid for the corresponding branching random walks as well. See a summary of our main results in Table 1 where we briefly explain the main parameters. We defer mentioning more related work to Section 2.1. From the point of view of epidemic modeling, an important message of our results is that the change in the phases can be obtained by *only changing the dynamics of the process around high-degree vertices* (i.e., increasing the degree-penalization), while keeping the underlying graph/contact network intact.

Applied and theoretical motivation for our model. While this paper is theoretical in nature, the choice of degree-dependent transmission rates comes directly from applications. Actual contacts do not scale linearly with network connectivity due to *limited time or awareness* [27, 44, 71]. Even individuals who spread an atypically large number of pathogens cause only sublinearly many new cases even via indirect spreading [64]. Degree-dependent transmission rates have been used to model the *sublinear impact of superspreaders* as a function of contacts in applications ranging from infection spreading to information spread in communication networks [29, 39, 49]. Two versions of the degree-dependent contact process were proposed and studied empirically in [71, 72]. Also related are the degree-dependent bond percolation and Ising model [2, 33] and topology-biased random walks in the applied literature [13, 23, 45, 59, 73], in which the transition probabilities from a vertex depend on the degrees of its neighbors. All these works assume a polynomial dependence on the degrees. On the theoretical side, the recent degree-dependent first passage percolation (dd-FPP) [40, 41, 42] uses the same "degree-penalization" that we shall assume, combined with the first passage percolation dynamics where reinfections to a vertex are not possible. Our results show that the phase-transition points of degree-penalized CP differ from those of dd-FPP. Reinfection in CP makes both the results and the proof techniques different. See more in Section 2.1 below.

1.1. Degree-penalized infection processes: main definitions

We now define the processes considered in this paper. These processes take place on an underlying graph, which is undirected, but not necessarily simple, that is, we allow multiple edges and loops, see Section 1.2 for the underlying graphs we use. We use the convention that the *degree* of a vertex is the number of nonloop edges incident to it (counted with multiplicity) plus twice the number of loops incident to it. More formally, for a graph $G = (V, E)$ we denote by $e(u, v)$ the number of edges between vertices $u, v \in V$, and by $N(v)$ the *neighborhood* of $v \in V$, the set of vertices u for which $e(u, v) \geq 1$. For a vector $\underline{x} \in \mathbb{N}^V$ (where $\mathbb{N} = \{0, 1, 2, \ldots\}$), we let $|\underline{x}| := \sum_{v \in V} x(v)$ be its 1-norm.

Definition 1.1 (Degree-penalized contact process). Consider a graph $G = (V, E)$, with d_v denoting the degree of vertex $v \in V$. Let $f(x, y) > 1$ be a function of two variables, $\lambda > 0$, and $\underline{\xi}_0 \in \{0, 1\}^V$. For

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Table 1. *Summary of our main results: phases of degree-dependent contact process. Let u, v be two vertices with degrees d_u, d_v, respectively, connected by and edge. Then the infection rate across the edge (u, v) is $\lambda/f(d_u, d_v) = \lambda/(d_u d_v)^\mu$ in the case of the product penalty, and $\lambda/f(d_u, d_v) = \lambda/\max\{d_u, d_v\}^\mu$ in the case of the max penalty. The second column shows the phases when the underlying graph is a Galton-Watson tree with offspring distribution D, and initially only the root is infected. Here, α denotes the power-law tail-exponent, that is, $\mathbb{P}(D \geq z) \asymp z^{-\alpha}$. The third column shows the phases when the underlying graph is a configuration model with degree sequence $\underline{d}_n$, and initially all the vertices are infected. Here, τ denotes the exponent of the limiting mass function, that is, $\mathbb{P}(D \geq z) \asymp z^{-(\tau-1)}$. We allow not just pure power laws, see Definitions 1.7–1.8 and Assumptions 1.10–1.12 for weaker assumptions. Some technical conditions are omitted in the table. For $\mu \in [1/2, 1)$ on the configuration model, fast extinction occurs when $\tau > 3$, including any other lighter tails, not just power laws.*

Product penalty	Galton-Watson tree $\mathcal{T}_D$	Configuration model $\mathrm{CM}(\underline{d}_n)$
$\mu < 1/2$	**Local survival** for any $\lambda > 0$ for tail heavier than stretched-exponential with $\zeta = 1 - 2\mu$	**Survival** until $\Theta_{\mathbb{P}}(\exp(Cn))$ time for any $\lambda > 0$ for tail heavier than stretched-exponential with $\zeta = 1 - 2\mu$
$\mu \geq 1/2$	**Extinction** for $\lambda < 1$ for any graph	**Extinction** in $O_{\mathbb{P}}(\mathrm{poly}(n))$ time for $\lambda < 1$ whenever $\sum_{i=1}^n d_i^{1-\mu} = O_{\mathbb{P}}(\mathrm{poly}(n))$
Max penalty	**Galton-Watson tree $\mathcal{T}_D$**	**Configuration model $\mathrm{CM}(\underline{d}_n)$**
$\mu < 1/2$	**Local survival** for any $\lambda > 0$ for tail heavier than stretched-exponential with $\zeta = 1 - 2\mu$	**Survival** until $\Theta_{\mathbb{P}}(\exp(Cn))$ time for any $\lambda > 0$ for tail heavier than stretched-exponential with $\zeta = 1 - 2\mu$
$\mu \in [1/2, 1)$	**Only global survival** for $\lambda < 1/2$ for weak power law with tail-exponent $\alpha < 1 - \mu$ **Extinction** for small λ when $\mathbb{E}[D^{1-\mu}] < \infty$	**Survival** until $\Theta_{\mathbb{P}}(\exp(Cn))$ time for any $\lambda > 0$ for power-law empirical degrees with $\mu < 3 - \tau$ **Extinction** in $\Theta_{\mathbb{P}}(\log(n))$ time for small λ for power-law empirical degrees with $\tau > 3$ (or lighter)
$\mu \geq 1$	**Extinction** for $\lambda < 1$ for any graph	**Extinction** in $O_{\mathbb{P}}(\mathrm{poly}(n))$ time for $\lambda < 1$ whenever $\sum_{i=1}^n d_i^{1-\mu} = O_{\mathbb{P}}(\mathrm{poly}(n))$

$u, v \in V$ let $r(u, v) = \lambda \cdot e(u, v)/f(d_u, d_v)$. We define $\mathrm{CP}_{f,\lambda}(G, \underline{\xi}_0) = (\underline{\xi}_t)_{t \geq 0} = (\xi_t(v))_{v \in V, t \geq 0}$ to be the following continuous-time Markov process on the state space $\{0, 1\}^V$. The process starts from the state $\underline{\xi}_0$ at time $t = 0$, and evolves according to the following transition rates:

$$\underline{\xi} \longrightarrow \underline{\xi} - \mathbb{1}_v \quad \text{with rate 1; for all } v \text{ with } \xi(v) = 1, \tag{1}$$

$$\underline{\xi} \longrightarrow \underline{\xi} + \mathbb{1}_v \quad \text{with rate } \sum_{u \in N(v)} \xi(u) r(u, v); \text{ for all } v \text{ with } \xi(v) = 0, \tag{2}$$

where $\mathbb{1}_v \in \{0, 1\}^V$ denotes the vector with entry 1 at position v, and zero entries at all other positions.

Strictly speaking, the above is not an actual mathematical definition. In case the graph is finite, the description using jump rates is entirely satisfactory (one can think of exponential waiting times governing the dynamics). However, as is well-known in the particle systems literature, the treatment of infinite graphs is more subtle [47]. We include the above only as a first indication of how the process behaves, but we define the contact process to be the process obtained from the Poisson graphical construction (see Section 3.1 below).

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

We refer to vertices v with $\xi_t(v) = 1$ as *infected* at time t, and to all other vertices as *healthy* at time t, and consequently $|\xi_t|$ is the number of infected vertices at time t. Describing the process less formally, each infected vertex u heals at rate 1, and during the time it is infected, it infects each of its healthy neighbors v at rate $r(u,v) = \lambda \cdot e(u,v)/f(d_u, d_v)$, where $e(u,v)$ is the number of edges between u and v. A common choice for $\underline{\xi}_0$ we take is $\underline{1}_G$, the all-1 vector on the vertex set V of G. This choice is a theoretical tool in our analysis, as the process starting from this initial state stochastically dominates the process starting from any other initial state.

A process related to the contact process is the branching random walk on the same graph. Branching random walks are known to stochastically dominate the contact process, since they consider the vertices of the graph as locations that infected particles can occupy, and they allow more than one infected particles per vertex. In comparison, in the contact process only one particle per vertex is allowed. In our setting, the *degree-penalized branching random walk* turns out to be useful for upper bounds when proving extinction.

Definition 1.2 (Degree-penalized branching random walk). Consider a graph $G = (V, E)$, with d_v denoting the degree of vertex $v \in V$ and $e(u,v)$ the number of edges between u and v. Let $f(x,y) > 1$ be a function of two variables, $\lambda > 0$, and $\underline{x}_0 \in \mathbb{N}^V$. For $u, v \in V$ let $r(u,v) = \lambda \cdot e(u,v)/f(d_u, d_v)$. We define $\mathrm{BRW}_{f,\lambda}(G, \underline{x}_0) = (\underline{x}_t)_{t\geq 0} = (x_t(v))_{v\in V, t\geq 0}$ to be the following continuous-time Markov process on the state space $\mathbb{N}^V$. The process starts from the state $\underline{x}_0$ at time $t = 0$, and evolves according to the following transition rates:

$$\underline{x} \longrightarrow \underline{x} - \mathbb{1}_v \quad \text{with rate } x(v) \text{ for all } v \in V, \tag{3}$$

$$\underline{x} \longrightarrow \underline{x} + \mathbb{1}_v \quad \text{with rate } \sum_{u\in N(v)} x(u) r(u,v) \text{ for all } v \in V. \tag{4}$$

Similarly to Definition 1.1, this definition works for finite graphs; we give a more general mathematical definition using particle genealogies in Definitions 3.4–3.5 below. Informally, we think of $x_t(v)$ as the number of particles at location v at time t. Then each particle dies at rate 1, independently of everything else, and each particle located at u reproduces to every neighboring vertex v at rate $r(u,v) = \lambda \cdot e(u,v)/f(d_u, d_v)$.

In what follows we study the qualitative long-term behavior of the above processes, for small infection parameters $\lambda > 0$. The following definition summarizes the possible phases that can occur on graphs, first with (countably) infinitely many vertices, and then on graphs with finitely many vertices. Here, and in the following, $\underline{0}$ denotes the all-zero vector (on the relevant index set).

Definition 1.3 (Modes of survival). Given a graph $G = (V, E)$, a penalty function $f(x,y) > 0$ and some $\lambda > 0$, consider either the process $(\underline{\xi}_t)_{t\geq 0} = \mathrm{CP}_{f,\lambda}(G, \underline{\xi}_0)$ or the process $(\underline{x}_t)_{t\geq 0} = \mathrm{BRW}_{f,\lambda}(G, \underline{x}_0)$ with respective fixed starting states $\underline{\xi}_0 \in \{0,1\}^V$ and $\underline{x}_0 \in \mathbb{N}^V$. If $|V| = \infty$, we say that the process exhibits

(i) *almost sure extinction* if, with probability 1, there exists some $T < \infty$ such that $\underline{\xi}_t = \underline{0}$ (respectively, $\underline{x}_t = \underline{0}$) for all $t \geq T$,
(ii) *global survival* if, with positive probability, $\underline{\xi}_t \neq \underline{0}$ (respectively $\underline{x}_t \neq \underline{0}$), for all $t \geq 0$.
(iii) *local survival* if, with positive probability, there exists $v \in V$ such that for any $t \geq 0$ there exists some $s > t$ such that $\xi_s(v) = 1$ (respectively, $x_s(v) \geq 1$).

For any underlying graph G and respective initial states $\underline{\xi}_0 \in \{0,1\}^V$ and $\underline{x}_0 \in \mathbb{N}^V$ of $\mathrm{CP}_{f,\lambda}$ and $\mathrm{BRW}_{f,\lambda}$, let us define the (possibly infinite) *extinction time*, and for a vertex $v \in G$ the *local extinction time at v*

$$T^{\mathrm{cp}}_{\mathrm{ext}}(G, \underline{\xi}_0) = \inf\{t \geq 0 : \underline{\xi}_t = \underline{0}\}, \quad T^{\mathrm{cp}}_{\mathrm{ext}}(G, \underline{\xi}_0, v) = \inf\{t \geq 0 : \xi_{t'}(v) = 0 \ \forall t' \geq t\},$$
$$T^{\mathrm{brw}}_{\mathrm{ext}}(G, \underline{x}_0) = \inf\{t \geq 0 : \underline{x}_t = \underline{0}\}, \quad T^{\mathrm{brw}}_{\mathrm{ext}}(G, \underline{x}_0, v) = \inf\{t \geq 0 : x_{t'}(v) = 0 \ \forall t' \geq t\}.$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

We note some remarks: First, local survival in (iii) implies global survival in (ii). Second, *only global* (but not local) survival means that (ii) holds, whereas for any choice $v \in V$ almost surely there exists some $T_v > 0$ such that $\xi_t(v) = 0$ (resp., $x_t(v) = 0$) for all $t > T_v$. Finally, provided that $0 < |\underline{\xi}_0| < \infty$ (resp., $0 < |\underline{x}_0| < \infty$), and that the graph G is connected, the phase that occurs among (i)–(iii) does not depend on the initial state $\underline{\xi}_0$ (resp., $\underline{x}_0$).

1.2. Definition of the underlying graphs

Next, we define the graph models that we focus on.

Definition 1.4 (Galton-Watson tree). Given a non-negative integer-valued random variable D, we define the *Galton-Watson (GW) tree with offspring distribution D* as follows. Let $\varnothing$ be a distinguished vertex, called the *root* of the tree. $\{\varnothing\}$ is generation 0 of the tree, and its cardinality is $Z_0 = 1$. Let $(D_{i,j})_{i=0,j=1}^{\infty}$ be an array of iid copies of D. Then we recursively define generation $i+1$ of the tree for $i = 0, 1 \dots$ in the following way. For each vertex j ($j = 1, \dots, Z_i$) of generation i we assign $D_{i,j}$ many offspring, connect them to vertex j, forming together generation $i+1$, that is, generation $i+1$ has cardinality $Z_{i+1} = \sum_{j=1}^{Z_i} D_{i,j}$. We call the resulting finite or infinite tree a realization of the Galton-Watson tree.

Our results, in an important regime, extend to any random or deterministic tree as well, as long as it grows at most exponentially almost surely, a concept which we define now.

Definition 1.5 (Branching number of a tree). Let $\mathcal{T}$ be an infinite tree, and let $Z_N(\mathcal{T}) := |\mathrm{Gen}_N(\mathcal{T})|$ be the size of generation N. Then we define the (possibly infinite) "upper" branching number of T as

$$\overline{\mathrm{br}}(T) := \limsup_{N\to\infty} Z_N(\mathcal{T})^{1/N}. \tag{5}$$

Definition 1.6 (Spherically symmetric tree). Given a positive integer-valued sequence $\underline{d} := (d_0, d_1, d_2, \dots)$, we define the *Spherically Symmetric Tree (SST) with degree sequence* $\underline{d}$, $\mathrm{SST}(\underline{d})$ as follows. Let $\varnothing$ be the *root* of the tree having $d_\varnothing := d_0$ many offspring. Then $\mathrm{SST}(\underline{d})$ is the tree where each vertex in generation i has d_i many offspring.

The following two definitions describe two important classes of degree distributions that we use for Galton-Watson trees.

Definition 1.7 (Weak power-law tails). Consider a distribution D on $\{0, 1, \dots\}$. We say that the tail of D weakly follows a power law with tail-exponent $\alpha > 0$ if for all fixed $\varepsilon > 0$ there exists a constant $z_0(\varepsilon) > 1$, such that whenever $z > z_0(\varepsilon)$,

$$\frac{1}{z^{\alpha+\varepsilon}} \le \mathbb{P}(D \ge z) \le \frac{1}{z^{\alpha-\varepsilon}}. \tag{6}$$

In the numerators in (6) we could have allowed a slowly varying function as well, but those can be ignored by adjusting $z_0(\varepsilon)$, due to Potter's theorem [8], since any slowly varying function $\ell(x)$ satisfies $x^{-\varepsilon} \ll \ell(x) \ll x^{\varepsilon}$ for all $\varepsilon > 0$ as $x \to \infty$. Pure power-law distributions satisfy (6) with $\varepsilon = 0$, in this case the constant 1 in the numerators of the upper and lower bounds may change. The next definition considers a similar domination, but now with stretched exponential tails:

Definition 1.8 (Heavier than stretched exponential tails). Consider a distribution D on $\{0, 1, \dots\}$. We say that D is heavier than stretched exponential with stretch-exponent $\zeta > 0$ if there exists a function $g : \mathbb{N} \to [0, \infty)$ such that $g(x) \to 0$ as $x \to \infty$, and an infinite sequence of nonnegative numbers $z_1 < z_2 < \dots$ such that for $i \ge 1$,

$$\mathbb{P}\big(D = z_i\big) \ge \exp(-g(z_i) z_i^{\zeta}). \tag{7}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

An equivalent statement to (7) is

$$\liminf_{z\to\infty} \frac{-\log(\mathbb{P}(D=z))}{z^{\zeta}} = 0.$$

We comment that in case of stretched exponential distributions, the tail $\mathbb{P}(D \geq K)$ and the mass function $\mathbb{P}(D = K)$ are a polynomial prefactor away, which can be incorporated in the function g.

The next definition gives the *finite* random graph model that we consider in this paper: the configuration model with a given degree sequence [12, 50].

Definition 1.9 (Configuration model). Given a positive integer n, and a sequence $\underline{d}_n := (d_1, \dots, d_n)$ of nonnegative integers with $h_n := \sum_{i=1}^n d_n$ even, we define the *configuration model* $\mathrm{CM}(\underline{d}_n)$ as a distribution on (multi)graphs constructed as follows. We take n vertices, and assign $d_1, d_2, \dots, d_n$ "half-edges" to them, respectively. Then we take a uniformly random pairing of the set of half-edges, and to each such pair we associate an edge in $\mathrm{CM}(\underline{d}_n)$ between the respective vertices.

In Definition 1.9, in the degree sequence $\underline{d}_n = (d_1^{(n)}, d_2^{(n)}, \dots, d_n^{(n)})$ we allow that the degrees depend on n. If it is not confusing we drop the superscript (n) from the degree sequence. When the degree sequence is random, (e.g., coming from an iid sequence $D_1, D_2, \dots$), then one may add an extra half-edge to D_n when $\sum_{i=1}^n D_i$ is odd. This will not affect the "regularity" assumptions on the degree sequence below. The configuration model is a locally tree-like graph: its *local weak limit* is a Galton-Watson tree [1, 4]. We expect that our results extend to other nongeometric graph models with branching processes as their local weak limit, for example, the Erdős-Rényi random graph, the Chung-Lu or Norros-Reitu model, rank-1 inhomogeneous random graphs [26, 18, 60, 11], and so on.

We define the *empirical mass function* ν_n of the degrees and the corresponding cumulative distribution function (cdf) for all $z \geq 0$ as

$$\nu_n(z) := \frac{n_z}{n} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{d_i=z\}} \quad \text{and} \quad F_n(z) = \nu_n([0,z]) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{d_i\leq z\}}. \tag{8}$$

Let D_n be a random variable with distribution ν_n. To be able to relate different elements of the sequence $\mathrm{CM}(\underline{d}_n)$ to each other, we pose the following regularity assumption, common in the literature [50, 51, 37].

Assumption 1.10 (Regularity assumptions on the degrees). Consider the configuration model in Definition 1.9. We assume that the sequence $(\underline{d}_n)_{n\geq 1} = ((d_1, d_2, \dots, d_n))_{n\geq 1}$ satisfies the following:

a) D_n with cdf $F_n(z)$ in (8) converges in distribution to some a.s. finite random variable D with $\mathbb{E}[D] \in (0, \infty)$. We denote the cdf of D by F_D.
b) $\lim_{n\to\infty} \mathbb{E}[D_n] = \mathbb{E}[D]$. In particular, for any constant $M \geq 0$,

$$\lim_{n\to\infty} \mathbb{E}[D_n \mathbb{1}_{\{D_n\geq M\}}] = \mathbb{E}[D\mathbb{1}_{\{D\geq M\}}].$$

Formulating power-law assumptions about a sequence of empirical distributions is slightly different than about a single distribution, since the minimal mass in the model with n vertices is $1/n$ and the maximal degree is n-dependent and finite. Hence, we formulate the next assumption, which ensures that the empirical distribution F_n follows a (possibly truncated) weak power law.

Assumption 1.11 (Power-law empirical degrees). We say that the empirical distribution of $(\underline{d}_n)_{n\geq 1}$ follows a weak (possibly truncated) power law with exponent $\tau > 1$ with exponent-error $\varepsilon \geq 0$, if there exist constants $c_\ell, c_u, z_0 = z_0(\varepsilon), n_0(\varepsilon) > 0$ and a function $z_{\max}^{(\ell)}(\varepsilon, n) \to \infty$ as $n \to \infty$ such that for all $n \geq n_0(\varepsilon)$, $F_n(z)$ in (8) satisfies

$$\frac{c_\ell}{z^{(\tau-1)(1+\varepsilon)}} \leq 1 - F_n(z) \leq \frac{c_u}{z^{(\tau-1)(1-\varepsilon)}}, \tag{9}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

for all $z \in [z_0, z_{\max}^{(\ell)}(\varepsilon, n)]$, while the upper bound holds for all $z \geq z_0$. In this case we call $\tau - 1$ the tail-exponent, consistent with Definition 1.7.

When the degrees are coming from an iid sample of a distribution D that satisfies (6) with some τ, ε, then one can use Chernoff bounds to show that Assumption 1.11 is also satisfied with a slightly larger ε and $z_{\max}(\varepsilon, n)$ can be chosen slightly below the typical maximum degree among iid degrees, which is $n^{(1-\varepsilon)/(\tau-1)}$ with high probability. However, in Assumption 1.11 we also allow for much lower $z_{\max}^{(\ell)}(\varepsilon, n)$. In such cases we talk about truncated power-law degrees. Since the truncation value $z_{\max}^{(\ell)}(\varepsilon, n) \to \infty$ as $n \to \infty$, the limiting distribution D satisfies (9) for all (fixed) $z \geq z_0$. We also comment that if $\varepsilon > 0$, by slightly increasing ε and z_0 if necessary, one may choose $c_\ell = c_u = 1$. Further, if instead of (9), one has the bounds

$$\ell_1(z) z^{-(\tau-1)} \leq 1 - F_n(z) \leq \ell_2(z) z^{-(\tau-1)} \tag{10}$$

for some slowly varying functions ℓ_1, ℓ_2, then (9) holds for any $\varepsilon > 0$, since $z^{-\varepsilon} \ll \ell_1(z) \leq \ell_2(z) \ll z^{\varepsilon}$ by Potter's theorem [8]. Then z_0 may depend on ε. In one of our results below, we additionally require the following assumption on the maximum degree and the empirical mass function.

Assumption 1.12. We assume that there is an $\varepsilon > 0$ such that there exists constants $n_0(\varepsilon), z_0(\varepsilon), C_u > 0$, such the empirical measure ν_n in (8) satisfies, for all $n > n_0(\varepsilon)$,

$$\nu_n(z) \leq \frac{C_u}{z^{\tau(1-\varepsilon)}} \quad \text{for all } z \geq z_0(\varepsilon), \tag{11}$$

$$\max_{i \leq n} d_i \leq C_u n^{1/(\tau(1-\varepsilon)-1)}. \tag{12}$$

The first condition implies the upper bound in Assumption 1.11, since (11) implies that $\nu_n((z, \infty)) \leq \sum_{i \geq z} c_u i^{-\tau(1-\varepsilon)} = c_u' z^{-(\tau-1)+\tau\varepsilon} = c_u' z^{-(\tau-1)(1-\varepsilon')}$ with $\varepsilon' := \varepsilon\tau/(\tau-1)$. The second condition is also quite natural, and both conditions hold for the empirical measure of iid degrees whp, as the following example shows. The proof can be found on page 74 in the Appendix.

Example 1.13 (Iid degrees). Suppose $\underline{d}_n = (D_{n,1}, \ldots, D_{n,n} + \mathbb{1}\{\sum_{i \leq n} D_{n,i} \text{ odd}\})$ where $(D_{n,i})_{i \leq n}$ are iid from a distribution D satisfying Definition 1.7 with some α. Then $(\underline{d}_n)_{n \geq 1}$ with high probability satisfies Assumptions 1.10, 1.11 with $\tau = \alpha + 1$ and any $\varepsilon > 0$, and $z_{\max}^{(\ell)}(\varepsilon, n) = n^{1/(\alpha(1+\varepsilon))}$ in Assumption 1.11, that is, with $z_0(\varepsilon/2)$ from Definition 1.7,

$$\mathbb{P}\left(\begin{array}{c} \forall z \geq z_0(\varepsilon/2) : 1 - F_n(z) \leq z^{-\alpha(1-\varepsilon)} \text{ and} \\ \forall z \in [z_0(\varepsilon/2), n^{1/(\alpha(1+\varepsilon))}] : 1 - F_n(z) \geq z^{-\alpha(1+\varepsilon)} \end{array} \right) \to 1. \tag{13}$$

Further, D satisfying Definition 1.7 for some α implies that (12) holds whp with $\tau = \alpha + 1$ and any $\varepsilon > 0$, that is, $\mathbb{P}(\max D_{n,i} \leq n^{1/(\alpha(1-\varepsilon))}) \to 1$. If D satisfies also that for all $\varepsilon > 0$ there exists $z_0(\varepsilon)$, such that for all $z \geq z_0(\varepsilon)$,

$$\mathbb{P}(D = z) \leq z^{-\tau(1-\varepsilon)}, \tag{14}$$

then the empirical measure $\nu_n(z)$ of $\underline{d}_n$ also satisfies (11) with any $\varepsilon > 1/\tau$. That is, for all $\varepsilon' > 0$,

$$\mathbb{P}\Big(\forall z \geq z_0(\varepsilon) : \nu_n(z) \leq z^{-\tau(1-1/\tau+\varepsilon)} = z^{-(\tau-1+\varepsilon')}\Big) \to 1. \tag{15}$$

Finally, if one considers *truncated* power-law distributions with $\max_{n,i} D_{n,i} = o(n^{1/\tau})$, then for all $\varepsilon > 0$

$$\mathbb{P}\Big(\forall z \geq z_0(\varepsilon) : \nu_n(z) \leq z^{-\tau(1-\varepsilon)}\Big) \to 1. \tag{16}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

While (15) seems rather weak, it is essentially best possible. Namely, using the lower bound one can show that the vertices with maximal degree are of order $n^{(1+o(1))/(\tau-1)}$, and when there is a single vertex with degree in this range, then the upper bound in (15) can be sharp. Examples on *truncated* power-law degree distributions can be found in [68, Example 1.20, 1.21] where graph distances are discussed under truncation. Here, as soon as the maximal degree is $o(n^{1/\tau})$, the true τ can be recovered also for point-masses with any $\varepsilon > 0$ in (16).

2. Results

We focus on the behavior of degree-penalized CP and BRW for small values of $\lambda > 0$. Table 1 contains a simplified summary of our results. We first state our results on the *product penalty*, that is, when $f(x, y) = (xy)^\mu$ for some $\mu \geq 0$ in Definitions 1.1 and 1.2. We based this choice on a slightly related model, degree-dependent first passage percolation [40], where this penalty function is proven to show rich phenomena for first passage percolation. Some of our results extend to *polynomial* penalty functions as well; see Remark 2.4 below. We start with results on Galton-Watson trees. On a Galton-Watson tree, the degree of a nonroot vertex v equals its number of offspring plus 1. Survival proofs for the contact process are often based on the "star"-graph strategy. This means that an infected high-degree vertex of degree K survives $\exp(\Theta(\lambda^2 K))$ long time with high probability where the infection is sustained by repeated reinfections from the surrounding K neighbors. If the rate is changed to $\lambda' = \lambda K^{-\mu}$ around this vertex, then a vertex of degree K survives $\exp(\Theta(\lambda^2 K^{1-2\mu}))$ long, which grows with the degree K only if $\mu < 1/2$. This intuition suggest a phase transition at $\mu = 1/2$ that we confirm in the following theorems:

Theorem 2.1 (Product penalty with $\mu < 1/2$ on Galton-Watson trees). *Let $\mathcal{T}$ be an infinite Galton-Watson tree with offspring distribution D, so that $p_0 = \mathbb{P}(D = 0) = 0$. Consider the degree-penalized contact process $\mathrm{CP}_{f,\lambda}$ and branching random walk $\mathrm{BRW}_{f,\lambda}$ with penalty function $f(x, y) = (xy)^\mu$ in Definitions 1.1 and 1.2 for some $\mu \in [0, 1/2)$.*

When the tail of D is heavier than stretched-exponential with stretch-exponent $1 - 2\mu$ (as in Definition 1.8), then for all $\lambda > 0$, $\mathrm{CP}_{f,\lambda}(\mathcal{T}, \mathbb{1}_\varnothing)$ and $\mathrm{BRW}_{f,\lambda}(\mathcal{T}, \mathbb{1}_\varnothing)$ both show local survival, for almost all realizations $\mathcal{T}$ of the Galton-Watson tree, that is, $\lambda_{c,1} = \lambda_{c,2} = 0$.

By setting $\mu = 0$, we recover the result for classical CP: if the tail of D is heavier than exponential then there is local survival [34]. Theorem 2.1 generalizes this result for any $\mu < 1/2$, and we see a phase transition point at $\mu = 1/2$. The counterpart of this theorem for $\mu \geq 1/2$ holds generally on *any* graph.

Theorem 2.2 (Product penalty with $\mu \geq 1/2$). *Consider the degree-penalized contact process $\mathrm{CP}_{f,\lambda}$ and branching random walk $\mathrm{BRW}_{f,\lambda}$ with penalty function $f(x, y) = (xy)^\mu$ in Definitions 1.1 and 1.2 for some $\mu \geq 1/2$. Then $\lambda_{c,1} > 1$, equivalently, for all $\lambda < 1$, $\mathrm{CP}_{f,\lambda}(G, \underline{\xi}_0)$ and $\mathrm{BRW}_{f,\lambda}(G, \underline{\xi}_0)$ both go extinct almost surely on any (finite or infinite) graph G whenever $|\underline{\xi}_0| < \infty$ (respectively, $|\underline{x}_0| < \infty$) almost surely. Further,*

$$\mathbb{E}[T^{\mathrm{cp}}_{\mathrm{ext}}(G, \underline{\xi}_0) \mid G, \underline{\xi}_0] \leq \mathbb{E}[T^{\mathrm{brw}}_{\mathrm{ext}}(G, \underline{\xi}_0) \mid G, \underline{\xi}_0] \leq \sum_{v \in V} \xi_0(v) d_v^{1-\mu}/(1 - \lambda) \tag{17}$$

and $\mathbb{P}(T^{\mathrm{cp}}_{\mathrm{ext}}(G, \underline{\xi}_0) > t)$ and $\mathbb{P}(T^{\mathrm{brw}}_{\mathrm{ext}}(G, \underline{\xi}_0) > t)$ both decay (at least) exponentially in t at a rate of at least $1 - \lambda$.

This result is novel. Intuitively, it shows that when the average number of infections to neighbors is at most λ times the square root of the degree, then $\lambda_{c,1} = 0$ on *any* graph. This is especially counterintuitive on graphs/trees with power-law degree distribution, since without penalization those have $\lambda_{c,1} = \lambda_{c,2} = 0$ by [34], and the penalization for $\mu \leq 1$ is not yet strong enough to suppress the power laws: since the average number of infections out of a vertex is polynomial of the degrees $(\Theta(\lambda \deg(v)^{1-\mu})$, which still follows a power law when $\deg(v)$ does so. The bound (17) bounds the mean extinction time as a function of the initially infected set. If G is finite, $\xi_0(v) = 1$ for all v, then the bound is linear in the number of edges of G.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Our next theorem is about the same processes on the configuration model. For the sake of simplicity, we assume that $\min_{i\le n} d_i \ge 3$, ensuring that for all sufficiently large n, $\mathrm{CM}(\underline{d}_n)$ on n vertices has a giant component $\mathcal{C}_n^{(1)}$ containing $n(1-o(1))$ many vertices with probability that tends to 1 as $n\to\infty$, see [50, 51]. We use the $O_{\mathbb{P}}, \Theta_{\mathbb{P}}$-notation in the standard way, see notation at the end of Section 2.1. By $\mathrm{poly}(n)$ we denote polynomial functions of n (with an arbitrary but finite exponent).

Theorem 2.3 (Product penalty on CM). *Let $G_n := \mathrm{CM}(\underline{d}_n)$ be the configuration model in Definition 1.9 on the degree sequence $\underline{d}_n = (d_1,\dots,d_n)$. Consider the degree-penalized contact process $\mathrm{CP}_{f,\lambda}$ and branching random walk $\mathrm{BRW}_{f,\lambda}$ with penalty function $f(x,y) = (xy)^\mu$ for some $\mu \ge 0$ from Definition 1.1 and 1.2.*

(a) *Let $\mu < 1/2$, and $\underline{d}_n$ satisfy the regularity assumptions in Assumption 1.10 with $\min_{i\le n} d_i \ge 3$, so that D has heavier tails than stretched-exponential with stretch-exponent $1-2\mu$ (as in Definition 1.8). Then for all $\lambda > 0$, both $\mathrm{CP}_{f,\lambda}(G_n, \underline{1}_{G_n})$ and $\mathrm{BRW}_{f,\lambda}(G_n, \underline{1}_{G_n})$ survive at least until $\Theta_{\mathbb{P}}(\exp(Cn))$ long time.*

(b) *Let $\mu \ge 1/2$. Then for all fixed $\lambda < 1$, both $\mathrm{CP}_{f,\lambda}(G_n, \underline{1}_{G_n})$ and $\mathrm{BRW}_{f,\lambda}(G_n, \underline{1}_{G_n})$ go extinct in $O_{\mathbb{P}}(\sum d_i^{1-\mu}) = O_{\mathbb{P}}(|E(G_n)|)$.*

There results are stated in the annealed setting, as the $O_{\mathbb{P}}, \Theta_{\mathbb{P}}$ notation can accommodate the errors coming from bad realizations of G_n. However, part (b) is a direct application of Theorem 2.2, and as such it can be strengthened to the quenched setting, and noting that $1-\mu \le 1$, the bound on the extinction time is linear in the number of edges of G_n.

Part (a) here recovers the result of [6] for classical CP by setting $\mu = 0$, and generalizes it for $\mu \in (0, 1/2)$. The phase transition at $\mu = 1/2$ occurs again: Part (b) is again novel and it is the finite graph analogue of Theorem 2.2. It shows that on finite graphs extinction happens quickly when $\mu \ge 1/2$. Starting from the all-infected state on G_n is not a serious restriction. In part (a), when started from a single vertex, that is, $\underline{\xi}_0 = \mathbb{1}_v$, the process has a positive probability of reaching a large pandemic, and the same result – long survival – is valid with positive probability. See [6] on how to move between a single vertex and all vertices as starting states.

Remark 2.4 (Polynomial penalties). The proof of Theorems 2.2 and 2.3 (b) are based on supermartingale arguments. They also work more generally for any penalty function $f_1(x,y) = x^\mu y^\nu$ with $\mu+\nu \ge 1$ under the same conditions, that is, for all graphs G, whenever $\lambda < 1$ and initial infected set ξ_0 is finite. In particular, with $x_t(v)$ the number of particles on vertex v in the BRW, the supermartingale is of the form $M_t = \sum_v x_t(v) d_v^\beta$ for some $\beta \in [1-\mu, \nu]$. Using the same supermartingale, it is thus straightforward to extend the result from monomials to polynomials of the form

$$f_2(x,y) = \sum_{i\in\mathbb{N}} a_i x^{\mu_i} y^{\nu_i}$$

with at least one term, say the first one, satisfying $\mu_1+\nu_1 \ge 1$, and all $a_i \ge 0$. In this case we can guarantee extinction whenever $\lambda < a_1$, using the stochastic domination of $\mathrm{CP}_{f_2,\lambda}$ by $\mathrm{CP}_{a_1 f_1,\lambda} = \mathrm{CP}_{f_1,\lambda/a_1}$, since the penalty is higher in process with f_2, leading to smaller infection rates, see (20) below. By the same reasoning, the proof of Theorem 2.2 also extends to processes with penalty function

$$f_3(x,y) := 1 \Big/ \sum_{i\in\mathbb{N}} a_i x^{-\mu_i} y^{-\nu_i}, \quad \text{with} \quad \sum_{i\in\mathbb{N}} a_i < \infty$$

whenever $(\mu_i,\nu_i)_{i\in\mathbb{N}}$ are such that and there is a *unique dominant term* (say the first one) in the following sense: $\mu_1 \le \mu_i$ and $\nu_1 \le \nu_i$ for every $i \in \mathbb{N}$ and $\mu_1+\nu_1 \ge 1$. We then bound the infection rates from above as follows:

$$\lambda / f_3(d_u,d_v) = \lambda \sum_i a_i d_u^{-\mu_i} d_v^{-\nu_i} \le \lambda \Big(\sum_i a_i\Big) d_u^{-\mu_1} d_v^{-\nu_1} = \lambda \Big(\sum_i a_i\Big) \Big/ f_1(d_u,d_v),$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

with $f_1(x,y) = x^{\mu_1} y^{\mu_1}$. So, using stochastic domination, whenever $\lambda < (\sum_i a_i)^{-1}$, Theorem 2.2 is still valid by the first part of the remark.

It turns out that – instead of the product penalty – switching to a class of penalty functions f that are monomials of $\max(x,y)$ shows a richer behavior, and we see an extra phase when μ crosses 1.

Theorem 2.5 (Max penalty on GW trees). *Let $\mathcal{T}$ be an infinite Galton-Watson tree with offspring distribution D, so that $\mathbb{P}(D = 0) = 0$. Consider the degree-penalized contact process $\mathrm{CP}_{f,\lambda}$ and branching random walk $\mathrm{BRW}_{f,\lambda}$ with penalty function $f(x,y) = \max(x,y)^\mu$ for some $\mu \geq 0$ in Definitions 1.1 and 1.2.*

(a) *Let $\mu < 1/2$, and the tail of D be heavier than stretched-exponential with stretch-exponent $1-2\mu$, (as in Definition 1.8). Then $\lambda_{c,2} = \lambda_{c,2} = 0$, that is, for all $\lambda > 0$, the contact process $\mathrm{CP}_{f,\lambda}(\mathcal{T}, \mathbb{1}_\varnothing)$ and $\mathrm{BRW}_{f,\lambda}(\mathcal{T}, \mathbb{1}_\varnothing)$ both show local survival, for almost all realizations $\mathcal{T}$ of the Galton-Watson tree.*

(b) *Let $\mu \in [1/2, 1)$, $\alpha \in (0, 1-\mu)$, and the tail of D weakly follow a power law with tail-exponent α (as in Definition 1.7). Then $\lambda_{c,1} = 0$ and $\lambda_{c,2} > 0$. In particular, for $\lambda \in (0, 1/2)$, $\mathrm{CP}_{f,\lambda}(\mathcal{T}, \mathbb{1}_\varnothing)$ and $\mathrm{BRW}_{f,\lambda}(\mathcal{T}, \mathbb{1}_\varnothing)$ both show local extinction and global survival, for almost all realizations $\mathcal{T}$ of the Galton-Watson tree.*

(c) *Let $\mu \in [1/2, 1)$, and $\mathbb{E}[D^{1-\mu}] < \infty$. Then $\lambda_{c,1} > 0$. In particular, for $\lambda < 1/(2\mathbb{E}[D^{1-\mu}])$, the processes $\mathrm{CP}_{f,\lambda}(\mathcal{T}, \mathbb{1}_\varnothing)$ and $\mathrm{BRW}_{f,\lambda}(\mathcal{T}, \mathbb{1}_\varnothing)$ both go extinct almost surely, for almost all realizations $\mathcal{T}$ of the Galton-Watson tree.*

Part (a) here again recovers classical results [34] when $\mu = 0$. Whenever $\mu \geq 1/2$, we see two new phases: if the offspring distribution has very heavy tails (part (b)), then local extinction still occurs (see Theorem 2.6 below) and the process survives by escaping to infinity for any $\lambda > 0$, that is, $\lambda_{c,1} = 0$ for almost all realizations of the Galton-Watson tree. If the offspring distribution has slightly lighter tails (but could still be a power law with infinite mean) in part (c), then global extinction occurs for small λ. The fact that the boundary between these two phases depends on the *exact* power-law tail has not been observed before in the contact process literature on static graphs [17]. Note that $\alpha < 1-\mu$ in part (b) means that $\mathbb{E}[D^{1-\mu}] = \infty$, and for power-law degrees with $\alpha > 1-\mu$, we have $\mathbb{E}[D^{1-\mu}] < \infty$. In this sense part (b) and (c) are almost matching and we leave out only the case $\alpha = 1-\mu$, where the (potentially present) slowly varying function multiplying the power-law decay shall play a decisive role in survival vs extinction (see below (6)). To avoid technical difficulties of tail-estimates, we decided to leave out this boundary case. Part (c) above is also valid more generally, see Corollary 2.7 below. To prove both local extinction (in part (b)) and global extinction (part (c)), we develop a new technique that we call *loop erasure of infection paths*, see Section 2.1 and Figure 1. This technique is robust, and can be used to obtain stronger results on extinction more generally, hence we state them separately. Note that there is some overlap between Theorem 2.5 and the theorem below.

Theorem 2.6 (Max penalty on trees and graphs). *Let $\mathcal{T}$ be any (possibly infinite) rooted tree with root $\varnothing$. Consider the degree-penalized contact process $\mathrm{CP}_{f,\lambda}$ and branching random walk $\mathrm{BRW}_{f,\lambda}$ with penalty function $f(x,y) = \max(x,y)^\mu$ for some $\mu \geq 0$.*

(a) *Let $\mu \geq 1/2$. Then $\lambda_{c,2} > 0$, that is, for all $\lambda < 1/2$, the processes $\mathrm{CP}_{f,\lambda}(\mathcal{T}, \underline{\xi}_0)$ and $\mathrm{BRW}_{f,\lambda}(\mathcal{T}, \underline{x}_0)$ both show local extinction almost surely, whenever $|\underline{\xi}_0| < \infty$ (resp., $|\underline{x}_0| < \infty$) almost surely. In this case we further have that for any $v \in \mathcal{T}$, the tail-distributions of the local extinction times $T^{\mathrm{cp}}_{\mathrm{ext}}(\mathcal{T}, \underline{\xi}_0, v)$, $T^{\mathrm{brw}}_{\mathrm{ext}}(\mathcal{T}, \underline{x}_0, v)$ decay exponentially in t.*

(b) *Let $\mu \geq 1$. Then $\lambda_{c,1} > 0$, that is, for all $\lambda < 1$, the processes $\mathrm{CP}_{f,\lambda}(G, \underline{\xi}_0)$ and $\mathrm{BRW}_{f,\lambda}(G, \underline{x}_0)$ both go extinct almost surely on any (finite or infinite) graph G whenever $|\underline{\xi}_0| < \infty$ (resp., $|\underline{x}_0| < \infty$) almost surely, hence also on any tree $\mathcal{T}$. Further, the bound (17) is also valid here on the extinction times, which decay at least exponentially in t with rate at least $1-\lambda$.*

Part (b) here is the max-penalty analogue of Theorem 2.2 which considers the product penalty with $\mu \geq 1/2$. In the regime $\mu \in [1/2, 1)$ we see a surprising difference: for the max-penalty we can only

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

guarantee local extinction on trees, whereas for the product penalty we see global extinction on all graphs. For the max-penalty CP on Galton-Watson trees, global survival but local extinction occurs for all small λ when $\mathbb{E}[D^{1-\mu}] = \infty$, see Theorem 2.5(b), which is a new phase in the CP literature. The reason for this difference is that under the product penalty it is much harder for the infection to spread between two superspreaders than under the max-penalty, and in the max-penalty case with $\mathbb{E}[D^{1-\mu}] = \infty$ the infection can escape to infinity via a ray of superspreaders of growing degree.

Here, we prove Theorem 2.6(a) using again the *loop erasure of infection paths* technique of Theorem 2.5(b-c). It follows from the *proof* of Theorems 2.5(c) and Theorem 2.6(a) that (local-global) extinction for small $\lambda > 0$ happens on any tree with at most exponential growth. Recall the upper branching number $\overline{\mathrm{br}}(\mathcal{T})$ from Definition 1.5.

Corollary 2.7 (Trees with finite branching number). *Let $\mathcal{T}$ be a rooted tree with $\overline{\mathrm{br}}(\mathcal{T}) := b < \infty$, and consider $\mathrm{CP}_{f,\lambda}$ and $\mathrm{BRW}_{f,\lambda}$ on $\mathcal{T}$ with penalty function $f(x,y) = \max(x,y)^{\mu}$ with $\mu \geq 1/2$. Then for all $\lambda < b^{-1}/2$, the processes $\mathrm{CP}_{f,\lambda}(\mathcal{T}, \mathbb{1}_{\varnothing})$ and $\mathrm{BRW}_{f,\lambda}(\mathcal{T}, \mathbb{1}_{\varnothing})$ both go extinct almost surely.*

Let $\mathcal{T}$ be a spherically symmetric tree with degree sequence $\underline{d} = (d_0, d_1, d_2, \dots)$ satisfying $\overline{\mathrm{br}}(\mathcal{T}) := b < \infty$. Then for all $\lambda < b^{-(1-\mu)}/2$, the processes $\mathrm{CP}_{f,\lambda}(\mathcal{T}, \mathbb{1}_{\varnothing})$ and $\mathrm{BRW}_{f,\lambda}(\mathcal{T}, \mathbb{1}_{\varnothing})$ both go extinct almost surely.

For spherically symmetric trees, finiteness of the upper branching number $\overline{\mathrm{br}}(\mathcal{T})$ is equivalent to requiring that $\log \overline{\mathrm{br}}(\mathcal{T}) = \limsup_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \log(d_i) < \infty$. The requirement on λ in Corollary 2.7 for SSTs is slightly milder than for arbitrary trees with finite upper branching number. Our last theorems describes the behavior of degree-penalized processes with maximum penalty on the configuration model.

Theorem 2.8 (Max penalty on CM, long-survival regimes). *Let $G_n := \mathrm{CM}(\underline{d}_n)$ be the configuration model in Definition 1.9 on the degree sequence $\underline{d}_n = (d_1, \dots, d_n)$ that satisfies the regularity assumptions in Assumption 1.10. Consider the degree-penalized contact process $\mathrm{CP}_{f,\lambda}$ and branching random walk $\mathrm{BRW}_{f,\lambda}$ with penalty function $f(x,y) = \max(x,y)^{\mu}$.*

(a) *Let $\mu < 1/2$, and the tail of D be heavier than stretched-exponential with stretch-exponent $1 - 2\mu$ (in the sense of Definition 1.8), and $\min_{i\leq n} d_i \geq 3$. Then for all $\lambda > 0$ the process $\mathrm{CP}_{f,\lambda}(G_n, \underline{1}_{G_n})$ survives at least until $\Theta_{\mathbb{P}}(\exp(Cn))$ long time.*

(b) *Let $\mu \in [1/2, 1)$, and $(\underline{d}_n)_{n\geq 1}$ satisfy the power-law empirical degree Assumption 1.11 with exponent τ and exponent-error $\varepsilon \geq 0$, with*

$$\mu < \big(3 - \tau - \varepsilon(\tau - 1)\big) \cdot \frac{1-\varepsilon}{1+\varepsilon}. \tag{18}$$

Then for all $\lambda > 0$ the process $\mathrm{CP}_{f,\lambda}(G_n, \underline{1}_{G_n})$ survives until $\Theta_{\mathbb{P}}(\exp(Cn))$ long time.

Part (a) and (b) here both show long survival; the difference is that when $\mu < 1/2$, the requirement on the degree distribution is very mild, while one needs sufficiently heavy power-law degrees for long survival when $\mu \leq 1/2$ (essentially, $\tau < 3 - \mu$). We emphasize that the $\Theta_{\mathbb{P}}$ notation implies that the results are annealed over the graphs, bad realizations are swallowed by the error there, and the proofs indeed find structures that sustain the infection for a long time, that are only "whp" present in G_n but are not present in "almost all realizations" of G_n for fixed n. Part (a) here again recovers the result for the classical CP [6] by setting $\mu = 0$. Part (b) is a novel phase; it is the finite-graph analogue of Theorem 2.5(b), as we explain now. As the error in the power-law exponent $\varepsilon \downarrow 0$, the condition in (18) simplifies to $\mu < 3 - \tau$, which is equivalent to the condition that $\alpha := \tau - 2 < 1 - \mu$ in Theorem 2.5. Here $\alpha = \tau - 2$ is the tail-exponent of the size-biased version of D, say $\widetilde{D}$, which can be shown to weakly follow a power law with $\alpha = \tau - 2 > 0$ in the sense of Definition 1.7. The local weak limit of the configuration model is a Galton-Watson tree with a version of the size-biased degree distribution $\widetilde{D}$. Theorem 2.5(b) describes that when $\mu \in [1/2, 1)$, on a weak power-law GW tree the processes both survive globally exactly when $\alpha < 1 - \mu$. Hence, this theorem reflects the analogous Theorem 2.5(b) on Galton-Watson trees, showing that global survival (but local extinction) there implies long survival

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

for the corresponding configuration model. Our last theorem states fast extinction on the configuration model, and admittedly it has the most involved proof.

Theorem 2.9 (Max penalty on CM, fast extinction regimes). *Consider the configuration model* $G_n := \mathrm{CM}(\underline{d}_n)$ *in Definition 1.9 on the degree sequence* $\underline{d}_n = (d_1, \dots, d_n)$. *Consider the degree-penalized contact process* $\mathrm{CP}_{f,\lambda}$ *and branching random walk* $\mathrm{BRW}_{f,\lambda}$ *with penalty function* $f(x, y) = \max(x, y)^{\mu}$.

(a) *Let* $\mu \in [1/2, 1)$, *and* $(\underline{d}_n)_{n\geq 1}$ *satisfy the regularity assumptions in Assumption 1.10, and the power-law empirical degrees of Assumption 1.11–1.12 with exponent* τ *and exponent-error* $\varepsilon \geq 0$ *with* $\tau(1-\varepsilon) > 3$. *Then for all* λ *small enough the processes* $\mathrm{CP}_{f,\lambda}(G_n, \underline{1}_{G_n})$ *and* $\mathrm{BRW}_{f,\lambda}(G_n, \underline{1}_{G_n})$ *both go extinct in* $\Theta_{\mathbb{P}}(\log n)$ *time.*

(b) *Let* $\mu \geq 1$. *Then for all* $\lambda < 1$, *the processes* $\mathrm{CP}_{f,\lambda}(G_n, \underline{1}_{G_n})$ *and* $\mathrm{BRW}_{f,\lambda}(G_n, \underline{1}_{G_n})$ *both go extinct in* $O_{\mathbb{P}}(\mathrm{poly}(n))$ *time, whenever it holds for* $(\underline{d}_n)$ *that* $\sum_{i=1}^{n} d_i^{1-\mu} = O_{\mathbb{P}}(\mathrm{poly}(n))$.

The results of this theorem, especially part (a) are novel in the contact process literature. First, CP dies out on power-law configuration models when, on average, a vertex transmits the infection to fewer neighbors than the square root of its degree. Second, we could prove that the extinction happens extremely fast, in $\Theta(\log n)$ time, using our new technique of *loop erasure of infection paths* combined with new structural results on the configuration model with $\tau > 3$ itself. Namely, we developed the loop erasure technique for trees where loops of infection paths are back-and-forth, and we erase these back-and-forth steps gradually. However, we cannot erase nontrivial loops. To be able to push the technique through for the configuration model with $\tau > 3$ power law degrees, we develop strong bounds on the surplus edges of neighborhoods, which also controls the number of nontrivial loops, see below. The best currently known bounds for the extinction time on graphs whose degree distribution has infinite support are polynomial [6], using recursive techniques on subtrees. On d-regular graphs, extinction of CP in its subcritical regime also happens in $\Theta(\log n)$ time [56]. Theorem 2.9(a) is the counterpart of Theorem 2.8(b), that is, it shows fast extinction on the configuration model with power-law degrees with sufficiently light tail. For long survival, Theorem 2.8(b) essentially requires $\mu < 3 - \tau$, equivalently, $\tau > 3 - \mu$. Here in Theorem 2.9(a) to prove extinction we need essentially $\tau > 3$, that is, we leave the cases when $\tau \in (3 - \mu, 3)$ open. We show that when $\tau > 3$, the local weak limit GW tree can be embedded until $\Theta(\log n)$ generations and with only a *bounded number of surplus edges* for all n vertices all-at-once, see Proposition 5.1, which might be interesting in its own right. We can then relate extinction of the CP/BRW on this new structure using a modified version of our methodology of *loop erasure of infection paths* also accommodating the presence of a few cycles in $\Theta(\log n)$ neighborhoods (see below in Section 2.1) so that CP/BRW never reaches the last generation of the "local weak limit + few surplus edges" approximation.

When $\tau < 3$, the configuration model looks structurally very different: the Galton-Watson tree forming the local weak limit of the configuration model has infinite mean, so it grows doubly-exponentially, and can be embedded into the configuration model only until $\Theta(\log\log n)$ generations, and with many surplus edges (i.e., edges beyond the number of vertices−1 that form the tree). On the one hand the $\Theta(\log\log n)$ generations of the embedding are too short and leave a good probability for CP/BRW to escape the embedded tree, and on the other hand there are too many additional cycles on the embedded tree that might boost the performance of CP/BRW. This causes the gap in the theorem, and so to prove extinction when $\tau \in (3 - \mu, 3)$ on the configuration model remains open.

2.1. Background, discussion and overview of proof techniques

In the following we highlight our novel proof techniques and their relation to the literature. The overview follows the structure of the rest of the paper.

Novel methodology: loop erasure in the space of infection paths (Sections 4 and 5). In Sections 4.5, 4.6 for the proof of Theorems 2.5(c) and 2.6(a) we develop a new recursive path counting argument on the space of infection paths, where we essentially carry out a (probability-weighted) loop erasure on the set

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

of possible infection paths. Then we relate the probability that $\mathrm{BRW}_{f,\lambda}$ survives on $\mathcal{T}$ to the product of degrees $\prod_{i=1}^{t} d_{\pi_i}^{1-\mu}$ summed over nonbacktracking paths, called rays $\pi = (\pi_0 = \varnothing, \pi_1, \pi_2, \dots,)$ on the tree, that is, paths that always go downwards. See Figure 1 for an illustration.

To extend the same result to the configuration model, that is, to prove Theorem 2.9(a) (in Section 5), we need to handle loops in the underlying graph. First in Lemma 5.2 we develop a new moment bound for the total size of GW trees with power-law offspring distribution with n-dependent maximum degree, (i.e., coming from the empirical degrees of the configuration model) valid for all $\tau > 3$. We use this new bound to show that whp the following holds for configuration models with $\tau > 3$ on n vertices: for some small $\delta > 0$, the $\delta \log n$ graph-neighborhood of every vertex only has at most a constant ℓ many *surplus edges*, that is, upon removing at most ℓ edges the $\delta \log n$ neighborhood becomes a tree. This result, Proposition 5.1, may be of independent interest. Returning to the degree-penalized contact process on the configuration model, we extend the (probability weighted) loop-erasure method that we developed for trees, to graphs with a bounded number of surplus edges, which is a nontrivial adaptation itself.

Survival on GW-trees with stretched-exponential-tailed offspring (Section 6). Theorem 2.1 is the generalization of the result by Huang and Durrett [34], where the authors prove that the classical contact process shows local survival on Galton-Watson trees whenever the offspring distribution has no exponential moments, that is, for all $c > 0$, it holds that $\mathbb{E}[\mathrm{e}^{cD}] = \infty$. When we set $\mu = 0$ in our degree penalised CP, we get back this result. For the degree-penalized versions, (i.e., $\mu \geq 0$) due to the penalties, the same condition is not sufficient for the proofs to carry through. For our proofs to hold, we need that D has heavier tails than stretched exponential with stretch exponent that is strictly less than $1 - 2\mu$, as in Definition 1.8. We leave it an open question whether this condition in Theorem 2.1 is sharp. For the classical contact process on Galton-Watson trees, the no-exponential-moments condition is sharp, as shown by Bhamidi, Nam, Nguyen and Sly [6].

The combination of Theorems 2.1 and 2.2 shows that the product penalty has a phase transition at $\mu = 1/2$. The usual argument that star-graph maintain the infection, as introduced by Chatterjee and Durrett [17], gives a back-of-the-envelope calculation that suggests this phase transition. Namely, a star-graph has a central vertex of degree say K, connected to K leaves or very low-degree vertices. The degree-penalized contact process on this structure survives typically for a time that is $\Omega_{\mathbb{P}}(\exp(\lambda^2 K^{1-2\mu}))$. Hence, whenever $1 - 2\mu > 0$, star-graphs survive long enough to infect other star-graphs embedded in the graph, provided these stars are not too far away from each other, that is, at most the logarithm of the survival time, giving at most $o(K^{1-2\mu})$ away. The stretched-exponential condition on the tail of D ensures that we can find stars within this distance of each other. For the infection to be able to pass between the stars, we also need to ensure that the path connecting the stars only contain low-degree vertices, so that the penalty does not hinder the infection from passing. This is new compared to the classical contact process, see Section 6.2.

Local extinction and global survival for small λ on power-law GW-trees. The combination of Theorems 2.5 and 2.6 shows that for the *max-penalty* when $\mu \in (1/2, 1)$, on a Galton-Watson tree, local extinction but global survival happens for any small $\lambda > 0$ and D has a power-law tail with tail-exponent $\alpha < 1 - \mu$. The behavior for large rates ($\lambda > 1$) may depend on the exact offspring distribution, and the contact process and the branching random walk may differ in behavior, see the work of Pemantle and Stacey [58]. Comparing Theorems 2.5 and 2.6 for the max-penalty with the corresponding Theorems 2.1 and 2.2 for the *product* penalty, we see that the phase of $\mu \geq 1/2$ for the max-penalty is subdivided into three different sub-phases, and the almost-sure extinction on arbitrary graphs requires $\mu \geq 1$ for the max-penalty, c.f. $\mu \geq 1/2$ for the product penalty. The subphases of max penalty with $\mu \in [1/2, 1)$ (Theorem 2.5(b)–(c)) are novel, since they provide the first natural static graph model where the contact process on power-law degree graphs *can* be subcritical (c) and show *only global survival* (b); and the exact condition also depends on the exact power-law exponent. For dynamical graphs a similar phenomenon occurs, see the recent work of Jacob, Linker and Mörters [36].

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Survival proofs: k-cores sustain the infection when stars heal quickly (Section 7). When $\mu \geq 1/2$, in the degree-penalized contact process, star-graphs heal essentially immediately and hence the usual arguments that they maintain the infection for a long time break down. In this regime on the GW tree, when the offspring distribution is sufficiently heavy-tailed (so that the $(1-\mu-\varepsilon)$th moment is infinite for some $\varepsilon > 0$), we prove that contact process shows local extinction but global survival by escaping to infinity, by Theorem 2.6(a) and Theorem 2.5(b).

In the configuration model with the same local weak limit, we find a new sub-graph that maintains the infection exponentially long in n. Extending the results of Janson and Luczak [37] we show that a k-core $H_n \subseteq G_n$ is present whp whenever $\tau \in (2,3)$, with size linear in n, on vertices with degree $k^{(1+\eta)/(3-\tau)}$, for some small $\eta = \eta(\varepsilon)$. The heuristic idea is that within H_n, the expected number of vertices that an infected vertex infects before healing is (ignoring the η error in the exponent, and denoting by $\deg_G(v)$ the degree of a vertex in the graph G):

$$\deg_{H_n}(u) r(u,v) = \deg_{H_n}(u) \lambda (\deg_{G_n}(u) \vee \deg_{G_n}(v))^{-\mu} \approx k\lambda k^{-\mu/(3-\tau)} \approx \lambda k^{1-\mu/(3-\tau)},$$

which grows with k whenever $\mu < 3-\tau$. We then show that when we choose k a large λ-dependent constant, the graph H_n sustains the contact process exponentially long. As far as we know this is the first model where k-cores are directly used to maintain the infection process.

Long survival on the configuration model with stretched exponential degree distribution (Section 8). In the regime where $\mu < 1/2$, a star-graph of degree j maintains the infection long enough to pass it to a neighboring star-graph if the graph-distance between them is $o(j^{1-2\mu})$. This idea will lead to Theorem 2.3 (a) and, as a consequence, Theorem 2.8 (a). Our proof here is an almost direct adaptation of the argument in [6] where we embed an expander-graph of stars with degree approximately j into the original graph so that each edge of the expander corresponds to a path of length $o(j^{1-2\mu})$. This leads to the condition of heavier than stretched exponential degree distributions with the exponent at most $1-2\mu$.

Another CP-model with degree-dependent transmission rates. Wei Su in [66] studies a degree-penalized contact process and branching random walk with the *asymmetric* penalty function $f(x,y) = x$. This penalty function implies that the total rate of infection from every vertex v is a constant $\lambda > 0$, irrespective of the degree of v. In this case, CP can be coupled to a "usual" un-penalized BRW on the GW tree with Poisson(λ) total offspring, and finer results can be obtained on Galton-Watson trees, not just the small $\lambda > 0$ behavior. For BRW, extinction occurs when $\lambda < 1$, and local vs. only global survival depends on whether $\lambda > 1/r(\mathcal{T})$ or not, where $r(\mathcal{T})$ is the spectral radius of the underlying tree with respect to symmetric random walk. For the contact process, the minimal degree in the Galton-Watson tree is decisive, see [66, Theorems 3.1, 4.2].

Comparison to degree-dependent FPP. In a sequence of papers, Komjáthy *et al.* [40, 41, 42] study non-Markovian degree-dependent first passage percolation (dd-FPP) on spatial graphs with power-law degrees with exponent τ. In dd-FPP, there is no healing, and hence, reinfections are not allowed. When one considers exponentially distributed transmission times, the degree-dependence there is similar to the product penalty here. Despite the similar transition rates, the results are entirely different for the two processes. The main phase transition point in our results, $\mu = 1/2$ is completely *absent* in dd-FPP: this transition point emanates from reinfecting the same high-degree vertex over and over. Explosion (reaching infinitely many individuals in finite time) stops happening in dd-FPP when $\mu < (3-\tau)/2$, see [40]. The other two papers [41, 42] study the rate of growth in time of the infection cluster on geometric inhomogeneous random graphs. The underlying geometry there is crucial, and the three phases are: stretched exponential, polynomial faster than the dimension, and polynomial growth proportional to t^d. The proof techniques mostly consist of renormalization techniques. We leave the question of spatial underlying graphs for degree-penalized CP for future projects.

Further directions. We believe that most of our results can be relatively easily adapted to graphs with GW trees as local weak limits, for example, the Chung-Lu or Norros-Reitu models or even to general inhomogeneous random graphs [11, 18, 60]. Our current proof techniques pose the restriction that they

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

all rely on tree-based arguments or "almost" tree-based arguments. It would be interesting to see how far this can be relaxed. Sparse random intersection graphs [9, 10, 20, 38, 63] or random intersection graphs with communities (where not every community is a complete graph [69, 70]) provide a natural candidate for this. These graphs are no longer locally tree-like, yet there is an embedded tree-like structure formed by the communities [69]. Another interesting direction is to develop robust techniques that can extend our results (beyond the $\mu \geq 1$ case) to spatial graphs with inhomogeneous degree distributions, for instance to geometric inhomogeneous random graphs [14], scale-free percolation [21], or the hyperbolic random graph [43]. A coupling argument to the related degree-dependent first passage percolation [40], which explodes also exactly when $\alpha := \tau - 2 < 1 - \mu$, indicates that at least Theorem 2.5(b) on global survival must carry through for these graphs. Considering the recent growth phases of degree-dependent first passage percolation (1-FPP) in [41, 42], it is an intriguing question to ask whether the front of the degree-dependent contact process started from the origin and conditioned to survive, follows the same universality classes of growth as the 1-FPP spreading process.

Metastable behavior of the original contact process on finite graphs is a lively field of research starting with [16]; see also [25, 52, 54, 55, 61, 62]. See [5, 15, 17, 53] for results on power-law preferential attachment models and configuration models, [48] on hyperbolic random graphs, [19, 35, 36] on dynamically evolving graphs, and [30] on spatial random graphs. Further studying metastability of the degree-penalized processes here (for instance, investigating metastable densities) is an interesting future direction.

Organization of the rest of the paper: Before the proofs we introduce some necessary terminology and preliminary facts about the contact process and branching random walks in Section 3. Then, in Section 4 we give the proofs of Theorems 2.2, 2.3(b), 2.5(c), 2.6(a), (b) and 2.9(c). In Section 5 we prove Theorem 2.9(a). Section 6 contains the proofs of Theorems 2.1 and 2.5(a), (b). In Section 7 we provide the proof of Theorem 2.8(b). Finally, in Section 8 we give a sketch of the proofs of Theorems 2.3(a) and 2.8(a).

Notation: When we compare degrees of vertices in graphs on the same vertex set, we use the notation $\deg_G(v)$ for the degree of vertex v within graph G. Unless specified, we always think of graphs as undirected. With a slight abuse of notation, we use $|G|$ as a shorthand for $|V(G)|$, the number of vertices in G.

We use the abbreviations "rhs" and "lhs" for "right-hand side" and "left-hand side" (of an equation), "iid" for "independent and identically distributed" and "whp" for "with high probability," that is, with probability converging to 1 as the size of the underlying graph (the number of its vertices) tends to infinity. For a deterministic function $g(n)$, we say that a sequence of random variables $X_n = o_{\mathbb{P}}(g(n))$, if the sequence $(X_n/g(n))_{n\geq 1}$ tends to 0 in probability, and we say that $X_n = O_{\mathbb{P}}(g(n))$ if $(X_n/g(n))_{n\geq 1}$ is a tight sequence of random variables. Similarly, $X_n = \Omega_{\mathbb{P}}(g(n))$ if $(g(n)/X_n)_{n\geq 1}$ is a tight sequence, and finally, we say that $X_n = \Theta_{\mathbb{P}}(g(n))$ if $X_n = O_{\mathbb{P}}(g(n))$ and $X_n = \Omega_{\mathbb{P}}(g(n))$ both hold.

3. Preliminaries

In this section we describe some basic properties of the contact process and the underlying random graphs that will be used throughout the paper.

3.1. Graphical construction of the contact process

We briefly discuss the *graphical construction* of the contact process, based on Section 6.2 of [31]. The graphical construction provides the mathematical definition of the contact process, and is useful for various coupling arguments. The idea is to record the infection and healing events of the contact process $\mathrm{CP}_{f,\lambda}(G, \underline{\xi}_0)$ on the space-time domain $V \times [0, \infty)$. For a Poisson point process PPP on $[0, \infty)$, we say that $t \in \mathrm{PPP}$ if t is an arrival time (a point) in the given PPP. Further, $\mathrm{PPP}(I)$ denotes the set of points that fall in the set $I \subseteq R$.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Definition 3.1 (Graphical construction of CP). For a graph $G = (V, E)$, consider for each $v \in V$ an independent Poisson process PPP_v with rate 1, and, independently of these, further independent Poisson processes PPP_{uv} for each $u, v \in V$ with corresponding rate $r(u, v) = \lambda \cdot e(u, v)/f(d_u, d_v)$. The healing events in (1) form a subset of the arrival times of $(\mathrm{PPP}_v)_{v\in V}$, and the infection events in (2) form a subset of the arrival times of $(\mathrm{PPP}_{uv})_{\{u,v\}\in E}$ that we describe now.

We define an *infection path* as a sequence $\{(v_0, t_0), (v_0, t_1), (v_1, t_1), (v_1, t_2), \ldots, (v_k, t_{k+1})\}$ with vertices $v_0, v_1, \ldots, v_k \in V$ and times $t_0 \le t_1 \le \ldots \le t_{k+1}$ such that

(i) $\mathrm{PPP}_{v_i}([t_i, t_{i+1}]) = \emptyset$ for each $i \in \{0, \ldots, k\}$, and
(ii) $t_i \in \mathrm{PPP}_{v_{i-1}v_i}$ for each $i \in \{1, \ldots, k\}$.

Then, we set

$$\xi_t(u) = \mathbb{1}\{\text{there is an infection path from } (v, 0) \text{ to } (u, t) \text{ for some } v \in \underline{\xi}_0\}, \quad u \in V,\ t \ge 0, \tag{19}$$

that is, we say u is infected at time t if the space-time point (u, t) can be reached by an infection path started at some infected v at time 0. Here, and in the future, with a slight abuse of notation we use the convention that $v \in \underline{\xi}_t$ means that $\underline{\xi}_t(v) = 1$ for the 0-1 vector $\underline{\xi}_t$ (i.e., we also treat $\underline{\xi}_t$ as a set).

We define the contact process to be the process obtained from (19). Note that we do not exclude the possibility of finite-time explosion, meaning that a process started from finitely many infections reaches infinitely many infections in finite time.

This definition is useful for coupling contact processes with different initial conditions and different spreading rates. The following is an easy consequence of the graphical construction.

Corollary 3.2. *For two penalty functions f_1, f_2 for which $f_1(x, y) \ge f_2(x, y)$ holds for all $x, y \ge 1$, it holds on any graph G and arbitrary initial starting state $\underline{\xi}_0$ and any $\lambda > 0$ that*

$$\mathrm{CP}_{f_1,\lambda}(G, \underline{\xi}_0) \overset{d}{\le} \mathrm{CP}_{f_2,\lambda}(G, \underline{\xi}_0). \tag{20}$$

Proof. The stochastic domination in (20) is the consequence of a standard coupling argument: construct the graphical construction of $\mathrm{CP}_{f_2,\lambda}(G, \underline{\xi}_0)$, that is, of the process with higher infection rates $(\lambda/f_2(u, v))_{u,v}$. Then, independently for different pairs uv, on PPP_{uv}, keep every infection event (point) with probability $(\lambda/f_1(u, v))/(\lambda/f_2(u, v)) = f_2(u, v)/f_1(u, v)$, independently across points. The thinned PPP has rate $\lambda f_1(u, v)$, hence we obtain a graphical construction of $\mathrm{CP}_{f_1,\lambda}(G, \underline{\xi}_0)$. This joint realization of the two processes gives a coupling of $\mathrm{CP}_{f_1,\lambda}(G, \underline{\xi}_0)$ and $\mathrm{CP}_{f_2,\lambda}(G, \underline{\xi}_0)$, so that every infection event in the former process is also an infection event in the latter process. This finishes the proof of (20). □

For the branching random walk, we adopt a different definition: we construct the process via particle genealogies. Heuristically speaking, every infected particle can trace back its infection via a finite-length chain of particles to a particle infected initially. In the next section we make this notion precise.

3.2. Genealogic branching random walks

We now describe a construction of branching random walks that keeps track of not only the number of particles per site, but also of the genealogy of particles. The advantage of defining the process via genealogies is that it allows us to treat the process both "locally" as well as "globally": it is possible that the process locally is well-behaving (even dies out) while at the same time it escapes to infinity in finite time, called explosion. Even if this latter event happens, the genealogical definition allows for (local) particle-chains also beyond the explosion time. This will be useful for proofs to show both local and global extinction, which are based on counting particles with given genealogies. Recall that for two vertices u and v in a graph G, we write $\mathrm{e}(u, v)$ to denote the number of edges between u and v.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Definition 3.3 (Set of genealogical labels). Given a graph $G = (V, E)$, we let $\mathscr{T} = \mathscr{T}(G)$ be the set

$$\mathscr{T} := \{(u_0, \dots, u_m) : m \in \mathbb{N},\ u_0, \dots, u_m \in V,\ e(u_i, u_{i+1}) > 0 \text{ for all } i\}.$$

An element $\pi = (u_0, \dots, u_m) \in \mathscr{T}$ will be a genealogical label attributed to certain particles that occupy u_m, the final vertex in the sequence. More specifically, a particle occupying u_m receives label π if it has the following genealogical history: its oldest ancestor particle (present at time 0) was at u_0 and gave birth to its next ancestor particle at u_1, which then gave birth to its next ancestor particle at $u_2, \dots,$ which then gave birth to the particle in question, at u_m. Hence, the label π lists the vertices occupied by the ancestors of the particle (and the particle itself), in chronological order. In particular, a particle present at vertex v at time 0 receives the label (v).

For $\pi = (u_0, \dots, u_m) \in \mathscr{T}$, we define

$$\begin{aligned} \mathfrak{l}(\pi) &:= m \quad \text{(length of } \pi\text{)}, \\ \mathfrak{s}(\pi) &:= u_m \ \text{(end-location of } \pi\text{)}. \end{aligned} \tag{21}$$

In case $m \geq 1$, we also let

$$\mathfrak{p}(\pi) := (u_0, \dots, u_{m-1}) \ \text{(parent path of } \pi\text{)}.$$

Definition 3.4 (Degree-penalized genealogic branching random walk). Consider a graph $G = (V, E)$, with d_v denoting the degree of vertex $v \in V$. Let $f(x, y) \geq 1$ be a function of two variables and $\lambda > 0$; for $u, v \in V$ let $r(u, v) = \lambda \cdot \mathrm{e}(u, v)/f(d_u, d_v)$. Also let $\underline{x}_0 \in \mathbb{N}^V$. We define $\mathrm{GBRW}_{f,\lambda}(G, \underline{x}_0) = (\underline{y}_t)_{t\geq 0} = (y_t(\pi))_{\pi\in\mathscr{T}, t\geq 0}$ to be the following continuous-time Markov process on the state space $\mathbb{N}^{\mathscr{T}}$. The process starts at time $t = 0$ from the state $\underline{y}_0$ defined by

$$y_0(\pi) = \begin{cases} x_0(\mathfrak{s}(\pi)) & \text{if } \mathfrak{l}(\pi) = 0; \\ 0 & \text{otherwise,} \end{cases}$$

and evolves according to the following transition rates:

$$\underline{y} \longrightarrow \underline{y} - \mathbb{1}_\pi \quad \text{with rate } y(\pi) \text{ for all } \pi \in \mathscr{T}; \tag{22}$$

$$\underline{y} \longrightarrow \underline{y} + \mathbb{1}_\pi \quad \text{with rate } y(\mathfrak{p}(\pi)) \cdot r(\mathfrak{s}(\mathfrak{p}(\pi)), \mathfrak{s}(\pi)) \text{ for all } \pi \in \mathscr{T} \text{ with } \mathfrak{l}(\pi) \geq 1. \tag{23}$$

We interpret $y_t(\pi)$ as the number of particles with label π at time t. Guided by this interpretation, we obtain the degree-penalized branching random walk from $\mathrm{GBRW}_{f,\lambda}$ as a projection, defined next.

Definition 3.5. Let $(\underline{y}_t)_{t\geq 0} = \mathrm{GBRW}_{f,\lambda}(G, \underline{x}_0)$, and define

$$x_t(v) = \sum_{\pi\in\mathscr{T}:\ \mathfrak{s}(\pi)=v} y_t(\pi), \quad t > 0,\ v \in V. \tag{24}$$

Then, we call $(\underline{x}_t)_{t\geq 0} = (x_t(v))_{v\in V, t\geq 0}$ the degree-penalized branching random walk on G with rate λ, penalization function f, and initial configuration $\underline{x}_0$.

We show that this definition is consistent with Defintion 1.2 by computing the transition rates. Let $(\underline{x}_t)_{t\geq 0}$ be the process obtained from $(\underline{y}_t)_{t\geq 0}$ as in (24). For each $v \in V$, the transition $\underline{x} \longrightarrow \underline{x} - \mathbb{1}_v$ occurs with rate

$$\sum_{\pi\in\mathscr{T}:\ \mathfrak{s}(\pi)=v} y(\pi) = x(v),$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

and the transition $\underline{x} \longrightarrow \underline{x} + \mathbb{1}_v$ occurs with rate

$$\sum_{\substack{\pi \in \mathscr{T}:\, \mathfrak{s}(\pi)=v,\\ \mathfrak{l}(\pi)\geq 1}} y(\mathfrak{p}(\pi)) \cdot r(\mathfrak{s}(\mathfrak{p}(\pi)), \mathfrak{s}(\pi)) = \sum_{w \in V} \sum_{\substack{\pi' \in \mathscr{T}:\\ \mathfrak{s}(\pi')=w}} y(\pi') \cdot r(w, v) = \sum_{w \in V} x(w) \cdot r(w, v),$$

where we used (24) to obtain the last equality.

In the statement of the following lemma, we interpret products of the form $\prod_{i=0}^{m-1}$ as 1 when $m = 0$.

Lemma 3.6 (Expectation formulas for genealogic branching random walks). *Let* $(\underline{y}_t)_{t\geq 0} = \mathrm{GBRW}_{f,\lambda}(G, \underline{x}_0)$*, and let* $\pi = (u_0, \dots, u_m) \in \mathscr{T}$*.*

(a) *For any* $t \geq 0$*, we have*

$$\mathbb{E}[y_t(\pi)] = \frac{t^m}{m!} \mathrm{e}^{-t} \cdot \Big(x_0(u_0) \prod_{i=0}^{m-1} r(u_i, u_{i+1}) \Big). \tag{25}$$

(b) *Define*

$$Z(\pi) := \begin{cases} x_0(u_0) & \text{if } \pi = (u_0); \\ \#\{t > 0 : y_t(\pi) = y_{t-}(\pi) + 1\} & \text{otherwise,} \end{cases} \tag{26}$$

that is, in case π *has length zero (so that* $\pi = (u_0)$*),* $Z(\pi)$ *is the number of initial particles* $x_0(u_0)$*, and in case* $m = \mathfrak{l}(\pi) \geq 1$*,* $Z(\pi)$ *is the number of particles with label* π *ever born. Then,*

$$\mathbb{E}[Z(\pi)] = z(\pi) = x_0(u_0) \prod_{i=0}^{m-1} r(u_i, u_{i+1}). \tag{27}$$

Before the proof we mention that the factor $\mathrm{e}^{-t} t^m / m!$ is the density of a Gamma random variable with parameters 1 and $m + 1$, that is, the convolution of $m + 1$ iid Exp(1) random variables. Intuitively this factor comes from the convolution of the healing times of the $m+1$ vertices on the path $u_0, \dots, u_m$.

Proof. Proof of part (a). We argue by induction in $m = \mathfrak{l}(\pi)$. In case $m = 0$, we have $\pi = (u_0)$, and the process $(y_s(\pi))_{s\geq 0}$ is a continuous-time Markov chain that starts at $y_0(\pi) = x_0(u_0)$ at time 0 and can only decrease, doing so with rate $y_s(\pi)$ at any time $s \geq 0$. If we interpret the state of this chain as a number of particles, where each particle dies with rate 1 (and no particles are born), then the probability that a particle is still alive at time t is e^{-t}, so the expected number of living particles at time t is $x_0(u_0)\mathrm{e}^{-t}$, as desired.

Now assume that $m \geq 1$ and the statement in (25) holds for all $\pi' \in \mathscr{T}$ with $\mathfrak{l}(\pi') \leq m - 1$. Let

$$\pi_0 = (u_0), \quad \pi_1 = (u_0, u_1), \quad \dots, \quad \pi_m = \pi = (u_0, \dots, u_m),$$

and let $\mathcal{F}$ be the σ-algebra generated by

$$\{y_s(\pi_i) : 1 \leq i \leq m - 1,\ s \geq 0\}.$$

Conditioned on $\mathcal{F}$, the process $(y_s(\pi))_{s\geq 0}$ is an $\mathbb{N}$-valued (time-inhomogeneous) Markov process that starts at 0 at time 0 and, at any time $s \geq 0$, increases by 1 with rate

$$y_s(\pi_{m-1}) r(\mathfrak{s}(\mathfrak{p}(\pi)), \mathfrak{s}(\pi)) = y_s(\pi_{m-1}) r(u_{m-1}, u_m),$$

and decreases by 1 with rate $y_s(\pi)$, by (23) and (22). Again seeing this process as counting particles (which as before die with rate 1, but now can also be born with a time-dependent rate), the conditional expectation of the number of particles at time t is

$$\mathbb{E}[y_t(\pi) \mid \mathcal{F}] = r(u_{m-1}, u_m) \cdot \int_0^t y_s(\pi_{m-1}) \cdot \mathrm{e}^{-(t-s)}\, \mathrm{d}s.$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Taking expectation and using Tonelli's theorem, this gives

$$\mathbb{E}[y_t(\pi)] = r(u_{m-1}, u_m) \cdot \int_0^t \mathbb{E}[y_s(\pi_{m-1})] \cdot \mathrm{e}^{-(t-s)} \, \mathrm{d}s.$$

Using this recursively m times, and then using the base induction case $\mathbb{E}[y_s(\pi_0)] = x_0(u_0)\mathrm{e}^{-s}$, we obtain

$$\begin{aligned}
&\mathbb{E}[y_t(\pi)] \\
&\quad = x_0(u_0) \prod_{i=0}^{m-1} r(u_{i-1}, u_i) \int_0^t \int_{s_1}^t \cdots \int_{s_{m-1}}^t \mathrm{e}^{-s_1} \mathrm{e}^{-(s_2-s_1)} \cdots \mathrm{e}^{-(s_m - s_{m-1})} \mathrm{e}^{-(t-s_m)} \, \mathrm{d}s_m \cdots \mathrm{d}s_1 \\
&\quad = x_0(u_0) \prod_{i=0}^{m-1} r(u_{i-1}, u_i) \cdot \mathrm{e}^{-t} \cdot \frac{t^m}{m!}.
\end{aligned}$$

Proof of part (b). In case $\mathfrak{I}(\pi) = 0$, the statement is obvious, using the fact that $y_0((v)) = x_0(v)$ for all v. Assume that $\mathfrak{I}(\pi) = m \geq 1$, and write $\pi = (u_0, \ldots, u_m)$. Since the transition $\underline{y} \to \underline{y} + \mathbb{1}_\pi$ occurs with rate $\underline{y}(\mathfrak{p}(\pi)) \cdot r(u_{m-1}, u_m)$ by (23), we have

$$\begin{aligned}
\mathbb{E}[Z(\pi)] &= r(u_{m-1}, u_m) \cdot \mathbb{E}\left[\int_0^\infty y_t(\mathfrak{p}(\pi)) \, \mathrm{d}t\right] \\
&= r(u_{m-1}, u_m) \cdot \mathbb{E}\left[\int_0^\infty y_t((u_0, \ldots, u_{m-1})) \, \mathrm{d}t\right].
\end{aligned}$$

Using Tonelli's theorem and (25) on the right-hand side, we obtain

$$\mathbb{E}[Z(\pi)] = r(u_{m-1}, u_m) \cdot \left(x_0(u_0) \prod_{i=0}^{m-2} r(u_i, u_{i+1})\right) \cdot \underbrace{\int_0^\infty \frac{t^{m-1}}{(m-1)!} \mathrm{e}^{-t} \, \mathrm{d}t}_{=1},$$

as desired. □

Corollary 3.7. *Let* $(\underline{y}_t)_{t \geq 0} = \mathrm{GBRW}_{f,\lambda}(G, \underline{x}_0)$*, and let* $\pi \in \mathcal{T}$ *with* $m = \mathfrak{I}(\pi)$*. Let* X_{m+1} *be a* $\mathrm{Gamma}(1, m+1)$ *random variable, that is, with density* $f_m(s) = e^{-s}s^m/m!$*. For any* $t \geq 0$*, we have*

$$\mathbb{P}(y_s(\pi) > 0 \text{ for some } s \geq t) \leq e \cdot \mathbb{E}[Z(\pi)] \cdot \mathbb{P}(X_{m+1} \geq t). \tag{28}$$

Proof. Let

$$\tau := \inf\{s \geq t : \ y_s(\pi) > 0\},$$

so that the left-hand side of (28) equals $\mathbb{P}(\tau < \infty)$. Next, define the event

$$\mathcal{A} := \{\tau < \infty, \ y_s(\pi) > 0 \text{ for all } s \in [\tau, \tau + 1]\}.$$

It is easy to check that

$$\mathbb{P}(\mathcal{A}) \geq \mathbb{P}(\tau < \infty) \cdot \mathrm{e}^{-1}. \tag{29}$$

Using first Tonelli's theorem and then the fact that $y_s(\pi)$ is integer-valued, we have

$$\int_t^\infty \mathbb{E}[y_s(\pi)] \, \mathrm{d}s = \mathbb{E}\left[\int_t^\infty y_s(\pi) \, \mathrm{d}s\right] \geq \mathbb{E}\left[\int_t^\infty \mathbb{1}_{\{y_s(\pi) > 0\}} \, \mathrm{d}s\right].$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Now, the right-hand side is bounded from below by

$$\mathbb{E}\left[\mathbb{1}_{\mathcal{A}} \cdot \int_{\tau}^{\tau+1} \mathbb{1}_{\{y_s(\pi)>0\}}\, \mathrm{d}s\right] = \mathbb{P}(\mathcal{A}),$$

where the equality follows from the definition of $\mathcal{A}$. Also using (29), we have thus obtained

$$\mathbb{P}(y_s(\pi) > 0 \text{ for some } s \geq t) = \mathbb{P}(\tau < \infty) \leq \mathrm{e} \int_t^{\infty} \mathbb{E}[y_s(\pi)]\, \mathrm{d}s.$$

The desired bound in (28) now follows from (25) in Lemma 3.6 (a). □

The last statement in this section states the stochastic domination between the contact process and branching random walk.

Lemma 3.8 (Domination of contact process by branching random walk). *Given any graph $G = (V, E)$, parameters $f : \mathbb{R}^2 \to [0, \infty)$, $\lambda > 0$, and starting state $\underline{\xi}_0 \in \{0, 1\}^V$, it holds that*

$$(\underline{\xi}_t)_{t\geq 0} = \mathrm{CP}_{f,\lambda}(G, \underline{\xi}_0) \overset{d}{\leq} \mathrm{BRW}_{f,\lambda}(G, \underline{\xi}_0) = (\underline{x}_t)_{t\geq 0}. \tag{30}$$

This is a well-known result which can be proved either by comparison of transition rates or a coupling using a graphical construction. See [47, p.34] for details of the latter approach; here we omit further details.

4. Extinction proofs via particle counting and martingales

In this section we prove several results relating to global, local or fast extinction. We start by showing Theorem 2.2 on global extinction for the product penalty with $\mu \geq 1/2$ in Section 4.1. Theorem 2.3 (in Section 4.2), Theorem 2.6(b) (in Section 4.3) and Theorem 2.9(b) (in Section 4.4) will all be straightforward consequences. Then we establish the other extinction phases for the max-penalty, showing local extinction on all trees – Theorem 2.6(a) – for $\mu \geq 1/2$ in Section 4.5. We then prove global extinction on GW trees with finite $(1-\mu)$th moment – Theorem 2.5(c) – in Section 4.6.

4.1. Product penalty: global extinction for all graphs when $\mu \geq 1/2$ via martingales

We start by establishing the subcritical phase for the product penalty (Theorem 2.2). Here, the result holds generally for any underlying graph, not just a Galton-Watson tree, and any monomial penalty function with polynomial-degree at least 1:

Claim 4.1 (Supermartingale for global extinction). Let $f(x, y) = ax^{\mu}y^{\nu}$ for some $a > 0$ and $\mu, \nu \geq 0$ such that $\mu+\nu \geq 1$, and let $G = (V, E)$ be an arbitrary finite graph. Consider the process $(x_t(v))_{t\geq 0, v\in V} = \mathrm{BRW}_{f,\lambda}(G, \underline{x}_0)$ for $\lambda > 0$ on a finite graph G, starting from a given state $\underline{x}_0 \in \mathbb{N}^V$. Define, for any $\alpha \in [1-\mu, \nu]$,

$$M_t := \sum_{v\in V} x_t(v) d_v^{\alpha}.$$

Then, whenever $\sum_{v\in V} x_0(v) d_v^{\alpha} < \infty$, the process $(M_t)_{t\geq 0}$ is a supermartingale with respect to the filtration $\mathcal{F}_t = \sigma\big((x_s(v))_{v\in V(G), s\leq t}\big)$ for all $\lambda \in (0, a]$ and a strict supermartingale when $\lambda \in (0, a)$.

Proof. We start by observing that the interval $[1-\mu, \nu]$ is nonempty since $\mu + \nu \geq 1$. To prove the supermartingale property we analyze the expected increments of $(M_t)_{t\geq 0}$, using the definition of $\mathrm{BRW}_{f,\lambda}$ in Def. 1.2. Here we use the transition rates for computations instead of the construction in Definition 3.5. This is justified since we assume that the graph is finite. The change in M_t may come from either a particle disappearing at v due to a death event, or from a new particle appearing at v due to reproduction events from neighboring particles. We obtain, using the rates in (3) and (4) with $r(u, v) = \lambda e(u, v)/f(d_u, d_v)$, that

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

$$\begin{aligned}\mathbb{E}[M_{t+\mathrm{d}t} - M_t \mid \mathcal{F}_t] &= \mathbb{E}\left[\sum_{v\in V(G)} (x_{t+\mathrm{d}t}(v) - x_t(v)) d_v^\alpha \,\middle|\, \mathcal{F}_t\right] \\ &= \sum_{v\in V(G)} \left(-x_t(v) + \sum_{u\in N(v)} x_t(u) r(u,v)\right) \mathrm{d}t \cdot d_v^\alpha \\ &= -M_t \mathrm{d}t + \sum_{v\in V(G)} \sum_{u\in N(v)} x_t(u) \cdot [\lambda e(u,v)/f(d_u,d_v)] \cdot d_v^\alpha \mathrm{d}t.\end{aligned}$$

We substitute $f(d_u, d_v) = a d_u^\mu d_v^\nu$ in the last line above, and use that $d_v^{\alpha-\nu} \le 1$ by the assumption that $\alpha \le \nu$ to obtain that:

$$\begin{aligned}\mathbb{E}[M_{t+\mathrm{d}t} - M_t \mid \mathcal{F}_t] &= -M_t \mathrm{d}t + (\lambda/a) \cdot \sum_{v\in V(G)} \sum_{u\in N(v)} x_t(u) e(u,v) d_u^{-\mu} d_v^{\alpha-\nu} \mathrm{d}t \\ &\le -M_t \mathrm{d}t + (\lambda/a) \cdot \sum_{v\in V(G)} \sum_{u\in N(v)} x_t(u) e(u,v) d_u^{-\mu} \mathrm{d}t.\end{aligned}$$

Exchanging the sums and using that $\sum_{v\in V} e(u,v) = d_u$ (see Notation in Section 1), we obtain

$$\mathbb{E}[M_{t+\mathrm{d}t} - M_t \mid \mathcal{F}_t] \le -M_t \mathrm{d}t + (\lambda/a) \cdot \sum_{u\in V(G)} x_t(u) d_u^{1-\mu} \mathrm{d}t.$$

Finally, since $d_u \ge 0$ is an integer, $d_u^{1-\mu} \le d_u^\alpha$ holds by the assumption $1-\mu \le \alpha$. Hence,

$$\mathbb{E}[M_{t+\mathrm{d}t} - M_t \mid \mathcal{F}_t] \le -M_t \mathrm{d}t + (\lambda/a) \cdot \sum_{u\in V(G)} x_t(u) d_u^\alpha \mathrm{d}t = [(\lambda/a) - 1] \cdot M_t \mathrm{d}t. \tag{31}$$

Since $M_t \ge 0$, for $\lambda \le a$ we obtain the supermartingale property, as $[(\lambda/a) - 1] \cdot M_t \mathrm{d}t \le 0$, with strict inequality when $\lambda < a$. The finiteness of the initial state M_0 is ensured by the assumption that $M_0 = \sum_{v\in V} x_0(v) d_v^\alpha < \infty$. This finishes the proof. □

Proof of Theorem 2.2. Let G be a finite graph. Without loss of generality we may assume that all vertices in G have degree at least 1. Indeed, if G contained a (finite) number of vertices with degree 0, the contact process on those, starting from any $\underline{\xi}_0$ with finitely many infected vertices, reduces to a pure death process where each particle dies at rate 1. This is because infection cannot happen to and from these vertices. This process goes almost surely extinct. Hence we assume wlog that $d_v \ge 1$ for all $v \in V$.

By Lemma 3.8, it is sufficient to prove the almost sure extinction of $\mathrm{BRW}_{f,\lambda}(G, \underline{\xi}_0)$ for any ξ_0 that is almost surely finite, that is, $\sum_{v\in V} \xi_0(v) < \infty$ almost surely. Fix now any such realization of the initial state. Then, since only finitely many coordinates are nonzero, $\sum_{v\in V} \xi_0(v) d_v^\alpha = \sum_{v\in V} x_0(v) d_v^\alpha < \infty$ also holds for any $\alpha > 0$. We assumed $\lambda \in (0,1)$ also in Theorem 2.2. Hence, the conditions of Claim 4.1 are satisfied with $\nu = \mu \ge 1/2$, and we can set $\alpha = 1 - \mu$ there to obtain the non-negative (strict) supermartingale $(M_t)_{t\ge 0}$, (i.e., not a martingale).

Apply Doob's martingale convergence theorem for the non-negative supermartingale $(M_t)_{t\ge 0}$. Since $(M_t)_{t\ge 0} \ge 0$ cannot take values in $(0,1)$, (as $d_v \ge 1$ and $x_t(v) \in \mathbb{N}$ for all v) its almost sure limit can only be 0. Therefore, almost surely $M_t = 0$ for large enough t. By the coupling between $\mathrm{CP}_{f,\lambda}$ and $\mathrm{BRW}_{f,\lambda}$ and that $d_v^{1-\mu} \ge 1$ whenever $d_v \ge 1$, we obtain that

$$\sum_{v\in V} \xi_t(v) \le \sum_{v\in V} x_t(v) d_v^{1-\mu} = M_t \xrightarrow{a.s.} 0$$

implying global extinction. To compute the extinction time, by definition, $T_{\mathrm{ext}}(G, \underline{\xi}_0) \ge t$ implies the existence of at least one infected particle at time t. Since $1 - \mu > 0$ and $d_v \ge 1$ for all v, the existence

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

of at least one infected particle at time t in turn implies $M_t \geq 1$. By Markov's inequality, and since $M_0 = |\underline{\xi}_0|$, taking expectation of (31) and solving the resulting differential equation for $\mathbb{E}[M_t \mid G, \underline{\xi}_0]$ yields for all $\lambda \leq 1$:

$$\begin{aligned}\mathbb{P}(T_{\text{ext}}(G, \underline{\xi}_0) \geq t \mid G, \underline{\xi}_0) &\leq \mathbb{P}(M_t \geq 1 \mid G, \underline{\xi}_0) \leq \mathbb{E}[M_t \mid G, \underline{\xi}_0] \\ &\leq \Big(\sum_{v \in V} \xi_0(v) d_v^{1-\mu} \Big) \exp(-(1-\lambda)t).\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{E}[T_{\text{ext}}(G, \underline{\xi}_0) \mid G, \underline{\xi}_0] &\leq \int_0^\infty \Big(\sum_{v \in V} \xi_0(v) d_v^{1-\mu} \Big) \exp(-(1-\lambda)t) \mathrm{d}t \\ &= \Big(\sum_{v \in V} \xi_0(v) d_v^{1-\mu} \Big) / (1-\lambda).\end{aligned}$$

This finishes the proof for finite graphs. To extend the result to infinite graphs, we can take an exhausting sequence of sets V_n (increasing finite subgraphs whose union is the whole graph) and work with the transition rates inherited from the original infinite graph and take a monotone limit; we omit the details.

The extensions in Remark 2.4 follow immediately from the stochastic domination in (20) and then the martingale argument applied to the monomial obtained. □

4.2. Product-penalty: fast extinction on the configuration model when $\mu \geq 1/2$

We obtain Theorem 2.3(b) as an immediate consequence of Theorem 2.2, since it applies for arbitrary finite graphs as well.

Proof of Theorem 2.3(b). The bound in (17) in Theorem 2.2 applied to the configuration model G_n yields that $\mathbb{E}[T_{\text{ext}}^{\text{cp}}(G_n, \underline{1}_{G_n}) | (d_i)_{i \leq n}] \leq \sum_{i=1}^n d_i^{1-\mu}/(1-\lambda)$. Fast extinction now follows by using the assumption that $\sum_{i=1}^n d_i^{1-\mu} = O_{\mathbb{P}}(\text{poly}(n))$. Assumption 1.10 implies that $\max_{i \leq n} d_i = o(n)$, so then this condition is automatically satisfied, but it holds even in a much larger class of degree sequences $(\underline{d}_n)$ that do not grow superpolynomially. □

4.3. Max-penalty: global extinction for all graphs when $\mu \geq 1$

Global extinction in Theorem 2.6(b) is a straightforward consequence of that in Theorem 2.2.

Proof of Theorem 2.6(b). For all $\mu \geq 0$,

$$f_1(x, y) := \max(x, y)^\mu \geq x^{\mu/2} y^{\mu/2} =: f_2(x, y)$$

holds for all $x, y \geq 1$. Hence, the stochastic domination in (20) applies and $\text{CP}_{f_1,\lambda} \overset{d}{\leq} \text{CP}_{f_2,\lambda}$. Since the exponent in f_2 is $\mu/2 \geq 1/2$ by the assumption that $\mu \geq 1$, Theorem 2.2 applies for $\text{CP}_{f_2,\lambda}$, and the process goes extinct for all $\lambda < 1$. Hence, so does $\text{CP}_{f_1,\lambda}$. □

4.4. Max-penalty: fast extinction on the configuration model when $\mu \geq 1$

Fast extinction in Theorem 2.9(b) follows from Theorem 2.3(b) in a similarly straightforward way.

Proof of Theorem 2.9,(b). The stochastic domination between the product and max-penalties discussed in the proof of Theorem 2.6(b) above implies the result from Theorem 2.3(b). □

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

4.5. *Max-penalty: Loop erasure in particle counting when* $\mu \in [1/2, 1)$

To prove local extinction, and also global extinction later under the max-penalty, we go back to the construction of genealogic branching random walks from Section 3.2. We use Lemma 3.6 and bound the number of total particles ever born, decomposed along genealogical paths. We first give some definitions.

For a graph $G = (V, E)$, we will take throughout this section the infection-rate function to be

$$r(u, v) = \frac{\lambda \cdot \mathrm{e}(u, v)}{\max(d_u, d_v)^\mu}, \quad u, v \in V. \tag{32}$$

Recall that $\mathscr{T} = \mathscr{T}(G)$ denotes the set of genealogical labels in G, as in Definition 3.3, and $Z(\pi)$ from (26). We define, for $\pi = (\pi_0, \dots, \pi_m) \in \mathscr{T}$,

$$z(\pi) := \prod_{i=0}^{m-1} r(\pi_i, \pi_{i+1}) = \lambda^m \cdot \prod_{i=0}^{m-1} \frac{\mathrm{e}(\pi_i, \pi_{i+1})}{\max(d_{\pi_i}, d_{\pi_{i+1}})^\mu}, \tag{33}$$

with $z(\pi) = 1$ if the length of the path $\mathrm{l}(\pi) = 0$. Note that, by Lemma 3.6(b), $z(\pi) = \mathbb{E}[Z(\pi)]$, the expected number of particles with label π ever born, in a genealogical branching process with birth rate λ, maximum-penalty function with exponent μ, and started with a single particle with label (π_0).

Definition 4.2 (Backtracking steps). Let $G = (V, E)$ be a graph. Given a path $\pi = (\pi_1, \dots, \pi_m) \in \mathscr{T}$ with length $\mathrm{l}(\pi) = m \geq 2$, we define

$$\tau(\pi) := \min\{i \geq 2 : \ \pi_i = \pi_{i-2} \neq \pi_{i-1}\} \tag{34}$$

(with the convention $\min \varnothing = \infty$). That is, $\tau(\pi)$ is the first index on the path when π returns to a vertex u right after having jumped away from it to a different vertex v. We informally refer to this kind of motion $u \to v \to u$ (with $u \neq v$) as a *backtracking step*. For π with $\tau(\pi) < \infty$, we define

$$g(\pi) := (\pi_0, \dots, \pi_{\tau-2}, \pi_{\tau+1}, \dots, \pi_{\mathrm{l}(\pi)}), \tag{35}$$

that is, $g(\pi)$ is the path obtained by removing the first backtracking step of π. We define $g^{-1}(\pi) = \{\pi' : g(\pi') = \pi\}$ as the set of paths that map to π under g.

We clarify that traversal of self-loops, even multiple times, is not considered a backtracking step for the above definition.

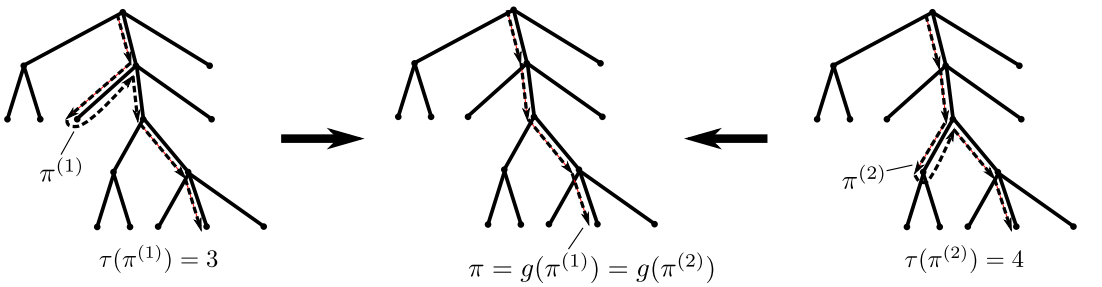


Figure 1. *Illustration of the loop erasure technique: two potential infection paths $\pi^{(1)}$ (left) and $\pi^{(2)}$ (right) both lead to the same rectified path π (middle). The figure also shows the definition of τ and g in Definition 4.2.*

Claim 4.3 (Removal of one backtracking step). Let $G = (V, E)$ be a graph, $\lambda > 0$ and $\mu \geq 1/2$. Let $z(\cdot)$ be as in (33). For any $\pi \in \mathscr{T}$ and any index $a \in \mathbb{N}$, we have

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

$$\sum_{\pi' \in g^{-1}(\pi):\ \tau(\pi')=a} z(\pi') \le \lambda^2 \cdot z(\pi) \max_{v: v \neq \pi_{a-2}} e(\pi_{a-2}, v). \tag{36}$$

Proof. Fix π and a as in the statement of the lemma. Write $\pi = (\pi_0, \dots, \pi_m)$, where $m = \mathrm{I}(\pi)$. We assume that the set $\{\pi' \in g^{-1}(\pi) :\ \tau(\pi') = a\}$ is nonempty, as the desired inequality is trivial otherwise. By (34) we then have $a \in \{2, \dots, m+2\}$, and any $\pi' \in g^{-1}(\pi)$ with $\tau(\pi') = a$ and π are of the form

$$\pi' = (\pi_0, \dots, \pi_{a-3}, u, v, u, \pi_{a-1}, \dots, \pi_m), \qquad \pi = (\pi_0, \dots, \pi_{a-3}, u, \pi_{a-1}, \dots, \pi_m), \tag{37}$$

where $u = \pi_{a-2} = \pi a - 2 = \pi'_a$ and v is a neighbor of u (with $v \neq u$). (We obtain that the next vertex on the path π has index $a-1$ by the erasure of the $(a-2)$nd (u) and $(a-1)$th vertex (v) on π'). Then, by (33),

$$z(\pi') = r(u,v)^2 \cdot z(\pi) = \left(\frac{\lambda \cdot \mathrm{e}(u,v)}{\max(d_u, d_v)^\mu} \right)^2 \cdot z(\pi) \le \frac{\lambda^2 \cdot \mathrm{e}(u,v)^2}{(d_u)^{2\mu}} \cdot z(\pi),$$

and

$$\begin{aligned} \sum_{\pi' \in g^{-1}(\pi):\ \tau(\pi')=a} z(\pi') &\le \frac{\lambda^2 \cdot z(\pi)}{(d_u)^{2\mu}} \sum_{v: v \neq u} \mathrm{e}(u,v)^2 \le \frac{\lambda^2 \cdot z(\pi)}{(d_u)^{2\mu}} \cdot d_u \cdot \max_{v: v \neq u} e(u,v) \\ &= \lambda^2 \cdot z(\pi) \cdot (d_u)^{1-2\mu} \cdot \max_{v: v \neq u} e(u,v) \le \lambda^2 \cdot z(\pi) \cdot \max_{v: v \neq \pi_{a-2}} e(\pi_{a-2}, v), \end{aligned}$$

where the last inequality follows from $d_u \ge 1$ and $\mu \ge 1/2$, and that $u = \pi_{a-2}$. □

With $g(\pi)$ the path obtained from π by the erasure of its first backtracking step in (35), let us write

$$g^{(1)} := g, \quad g^{(k+1)} := g \circ g^{(k)},\ k \ge 0.$$

In the statement and proof of the following lemma, to avoid summations with long subscripts, for any set A and any function $h : A \to \mathbb{R}$, we write $\sum\{h(x) : x \in A\} = \sum_{x \in A} h(x)$ (with the convention that this is zero when A is empty).

Lemma 4.4 (Removal of multiple backtracking steps). *Let G, λ, μ, f and $z(\cdot)$ be as in Claim 4.3. Fix $\pi \in \mathcal{T}$. Then, for any $k \ge 1$ and any sequence of positive integers $(a_1, \dots, a_k)$, we have*

$$\begin{aligned} &\sum \left\{ z(\pi') :\ \begin{array}{l} \pi' \in (g^{(k)})^{-1}(\pi), \\ \tau(\pi') = a_1,\ \tau(g(\pi')) = a_2,\ \dots,\ \tau(g^{(k-1)}(\pi')) = a_k \end{array} \right\} \\ &\quad \le \left(\max_{\substack{u,v \in G \\ v \neq u}} e(u,v) \right)^k \lambda^{2k} \cdot z(\pi). \end{aligned}$$

Proof. The proof is by induction on k, the case $k = 1$ being Claim 4.3. Assume the statement has been proved for k, and fix $\pi \in \mathcal{T}$ and a sequence $(a_1, \dots, a_{k+1})$. Then, since τ gives the location of the first backtracking step,

$$\begin{aligned} &\sum \left\{ z(\pi') :\ \pi' \in (g^{(k+1)})^{-1}(\pi),\ \tau(\pi') = a_1,\ \dots,\ \tau(g^{(k)}(\pi')) = a_{k+1} \right\} \\ &\quad = \sum \left\{ \sum \left\{ z(\pi') :\ \begin{array}{l} \pi' \in g^{-1}(\pi''), \\ \tau(\pi') = a_1 \end{array} \right\} :\ \begin{array}{l} \pi'' \in (g^{(k)})^{-1}(\pi), \\ \tau(\pi'') = a_2,\ \dots,\ \tau(g^{(k-1)}(\pi'')) = a_{k+1} \end{array} \right\}. \end{aligned} \tag{38}$$

By Claim 4.3, for each π'', the inner sum above is smaller than

$$\lambda^2 z(\pi'') \max_{v: v \neq \pi''_{a_1-2}} e(\pi''_{a_1-2}, v) \le \lambda^2 z(\pi'') \max_{u,v \in G: v \neq u} e(u,v),$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

so the double sum in (38) is smaller than

$$\max_{u,v\in G:v\neq u} e(u,v)\lambda^2 \cdot \sum\Big\{z(\pi'') : \pi'' \in (g^{(k)})^{-1}(\pi),\ \tau(\pi'') = a_2,\ \ldots,\ \tau(g^{k-1}(\pi'')) = a_{k+1}\Big\}.$$

Using the induction hypothesis, this is smaller than $\big(\max_{u,v\in G:v\neq u}\big)^{k+1}\lambda^{2(k+1)}z(\pi)$, as required. □

We would now like to use the above lemma to obtain a bound involving all possible sequences $(a_1,\ldots,a_k)$. Before doing so, we prove the following simple fact.

Claim 4.5. Let G be a graph and $\pi \in \mathscr{T}$ be such that $\tau(\pi) < \infty$ and $\tau(g(\pi)) < \infty$. Then,

$$\tau(g(\pi)) \geq \tau(\pi) - 1.$$

Proof. This follows from the observation that the sub-path $(\pi_0,\ldots,\pi_{\tau-2})$ remains intact after applying g to π, and this sub-path contains no backtracking steps by the minimality of $\tau(\pi)$. □

Corollary 4.6. *Let G, λ, μ and f be as in Claim 4.3. Fix $\pi \in \mathscr{T}$ and $k \geq 1$. Then,*

$$\sum_{\pi'\in(g^{(k)})^{-1}(\pi)} z(\pi') \leq 2^{\mathrm{I}(\pi)} \cdot \Big(4\lambda^2\big(\max_{\substack{u,v\in G\\ v\neq u}} e(u,v)\big)\Big)^k \cdot z(\pi). \tag{39}$$

Proof. Fix π and k as in the statement. Define

$$\mathcal{A} := \{(\tau(\pi'), \tau(g(\pi')), \ldots, \tau(g^{(k-1)}(\pi'))) :\ \pi' \in (g^{(k)})^{-1}(\pi)\}.$$

That is, for a single $\pi' \in (g^{(k)})^{-1}(\pi)$, the sequence $(\tau(\pi'), \tau(g(\pi')), \ldots, \tau(g^{(k-1)}(\pi')))$ gives the locations – that is, not the vertex but its index on the "current" path – of loop erasure when we sequentially apply g, k times, on the path π'. $\mathcal{A}$ is then the set of all sequences of length k that can be obtained by taking $\pi' \in (g^{(k)})^{-1}(\pi)$ and applying τ, $\tau\circ g$, …, $\tau\circ g^{(k-1)}$ to π'. By Lemma 4.4, the left-hand side of (39) is smaller than

$$\sum_{\pi'\in(g^{(k)})^{-1}(\pi)} z(\pi') \leq \Big(\max_{u,v\in G:v\neq u} e(u,v)\Big)^k \lambda^{2k} z(\pi) \cdot |\mathcal{A}|.$$

The desired bound will then follow from the inequality $|\mathcal{A}| \leq 2^{\mathrm{I}(\pi)+2k}$, which we now prove.

For each $\pi' \in (g^{(k)})^{-1}(\pi)$, we add $2(i-1)$ to the location of the ith erasure in the sequential application of loop erasure g on π', which, by Claim 4.5 leads to a a strictly increasing sequence of numbers, that is, we define

$$c_i(\pi') := \tau(g^{(i-1)}(\pi')) + 2(i-1), \quad i \in \{1,\ldots,k\}$$

(with $g^{(0)}(\pi') = \pi'$). Note that

$$\begin{aligned} c_k(\pi') &= \tau(g^{(k-1)}(\pi')) + 2(k-1) \leq \mathrm{I}(g^{(k-1)}(\pi')) + 2(k-1) \\ &= \mathrm{I}(\pi) + 2 + 2(k-1) = \mathrm{I}(\pi) + 2k. \end{aligned}$$

Moreover, for $i \in \{1,\ldots,k-1\}$,

$$c_{i+1}(\pi') - c_i(\pi') = \tau(g^{(i)}(\pi')) - \tau(g^{(i-1)}(\pi')) + 2,$$

which is positive by Claim 4.5. These considerations show that $(c_1(\pi'),\ldots,c_k(\pi'))$ is an increasing sequence in $\{1,\ldots,\mathrm{I}(\pi)+2k\}$. Therefore, $\mathcal{A}$ can be mapped injectively into the set of increasing sequences with k elements in $\{1,\ldots,\mathrm{I}(\pi)+2k\}$. It is a combinatorial exercise to show that the number of such sequences is $\binom{\mathrm{I}(\pi)+2k}{k} \leq 2^{\mathrm{I}(\pi)+2k}$. □

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Proof of Theorem 2.6(a). By Lemma 3.8 it is enough to prove the result for the branching random walk. Assume that $\mu \geq 1/2$ and $\lambda < 1/2$. Let $\mathcal{T}$ be a tree with a root $\varnothing$. For each vertex u of $\mathcal{T}$, let $\pi_{\downarrow u}$ denote the geodesic path from $\varnothing$ to u. Consider the branching random walk on $\mathcal{T}$ with penalty function $f(x, y) = \max(x, y)^{\mu}$, birth rate μ and initial configuration consisting of a single particle, located at the root. For this process, let $Z(\cdot)$ be as in (26) and $z(\cdot)$ be as in (33); note that by (27), we have $\mathbb{E}[Z(\pi)] = z(\pi)$ for any $\pi \in \mathcal{T}$. Further, let $\mathcal{T}_0 = \{\pi \in \mathcal{T} : \pi_0 = \varnothing\}$ denote the set of paths in $\mathcal{T}$ that start at the root. Then, since $\mathcal{T}$ is a tree and $e(u, v) \in \{0, 1\}$ for all pairs $u, v \in \mathcal{T}$,

$$\begin{aligned}\sum_{\pi \in \mathcal{T}_0:\, \mathfrak{s}(\pi) = u} \mathbb{E}[Z(\pi)] &= \sum_{\pi \in \mathcal{T}_0:\, \mathfrak{s}(\pi) = u} z(\pi) = \sum_{k=0}^{\infty} \sum_{\pi \in (g^{(k)})^{-1}(\pi_{\downarrow u})} z(\pi) \\ &\leq z(\pi_{\downarrow u}) \cdot 2^{\mathfrak{l}(\pi_{\downarrow u})} \cdot \sum_{k=0}^{\infty} (4\lambda^2)^k = \frac{2^{\mathfrak{l}(\pi_{\downarrow u})}}{1 - 4\lambda^2} \cdot z(\pi_{\downarrow u}), \end{aligned} \tag{40}$$

where the inequality follows from Corollary 4.6. Since the right-hand side above is finite, we see that the expectation of the number of particles ever born at u is finite, so this number is almost surely finite. This proves local extinction for the initial configuration in which there is a single particle at the root. As already observed, this implies local extinction for the branching random walk, and also the contact process, started from any finite initial configuration.

To prove the exponential decay of the local extinction time, we will use Corollary 3.7 to write, for any $t > 0$,

$$\begin{aligned}\mathbb{P}\Big(T^{\mathrm{brw}}_{\mathrm{ext}}(\mathcal{T}, \mathbb{1}_{\varnothing}, u) > t\Big) &= \sum_{\pi \in \mathcal{T}_0:\, \mathfrak{s}(\pi) = u} \mathbb{P}(y_s(\pi) > 0 \text{ for some } s \geq t) \\ &\leq \sum_{\pi \in \mathcal{T}_0:\, \mathfrak{s}(\pi) = u} e \cdot \mathbb{E}[Z(\pi)] \cdot \mathbb{P}(X_{\mathfrak{l}(\pi)+1} \geq t), \end{aligned} \tag{41}$$

where X_m is a Gamma$(1, m)$ variable for any $m > 0$. Let $\alpha \in (0, 1)$ be a constant specified later. Further bounding the right-hand side of (41), we write

$$\mathbb{P}\Big(T^{\mathrm{brw}}_{\mathrm{ext}}(\mathcal{T}, \mathbb{1}_{\varnothing}, u) > t\Big) \leq \sum_{\substack{\pi \in \mathcal{T}_0:\, \mathfrak{s}(\pi) = u, \\ \mathfrak{l}(\pi) < \lfloor \alpha t \rfloor}} e \cdot z(\pi) \cdot \mathbb{P}(X_{\mathfrak{l}(\pi)+1} \geq t) + \sum_{\substack{\pi \in \mathcal{T}_0:\, \mathfrak{s}(\pi) = u, \\ \mathfrak{l}(\pi) \geq \lfloor \alpha t \rfloor}} e \cdot z(\pi). \tag{42}$$

First, we bound the first sum on the right-hand side of (42). Noting that $X_{\lfloor \alpha t \rfloor}$ stochastically dominates $X_{\mathfrak{l}(\pi)+1}$ when $\mathfrak{l}(\pi) < \lfloor \alpha t \rfloor$, and using Corollary 4.6 we get

$$\begin{aligned}\sum_{\substack{\pi \in \mathcal{T}_0:\, \mathfrak{s}(\pi) = u, \\ \mathfrak{l}(\pi) < \lfloor \alpha t \rfloor}} e \cdot z(\pi) \cdot \mathbb{P}(X_{\mathfrak{l}(\pi)+1} \geq t) &\leq e \cdot \mathbb{P}(X_{\lfloor \alpha t \rfloor} \geq t) \cdot \sum_{r = \mathfrak{l}(\pi_{\downarrow}(u))}^{\lfloor \alpha t \rfloor - 1} \sum_{\substack{\pi \in \mathcal{T}_0:\, \mathfrak{s}(\pi) = u, \\ \mathfrak{l}(\pi) = r}} z(\pi) \\ &\leq e \cdot \mathbb{P}(X_{\lfloor \alpha t \rfloor} \geq t) \cdot z(\pi_{\downarrow}(u)) \cdot 2^{\mathfrak{l}(\pi_{\downarrow}(u))} \sum_{k=0}^{(\lfloor \alpha t \rfloor - 1 - \mathfrak{l}(\pi_{\downarrow}(u)))/2} (4\lambda^2)^k. \end{aligned} \tag{43}$$

Since $\lambda < 1/2$, the sum on the right-hand side of (43) is bounded by $1/(1 - 4\lambda^2)$. By (33), we have

$$z(\pi_{\downarrow}(u)) = \lambda^{\mathfrak{l}(\pi_{\downarrow}(u))} \prod_{i=0}^{\mathfrak{l}(\pi_{\downarrow}(u)) - 1} (\max(d_{\pi_i}, d_{\pi_{i+1}}))^{-\mu} \leq (2^{-\mu}\lambda)^{\mathfrak{l}(\pi_{\downarrow}(u))}. \tag{44}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Combining (43) and (44) to further upper bound the right-hand side of (43) yields

$$\sum_{\substack{\pi\in\mathcal{T}_0:\ \mathfrak{s}(\pi)=u,\\ \mathrm{I}(\pi)<\lfloor\alpha t\rfloor}} e\cdot z(\pi)\cdot\mathbb{P}(X_{\mathrm{I}(\pi)+1}\geq t)\leq\frac{e}{1-4\lambda^2}\cdot\mathbb{P}(X_{\lfloor\alpha t\rfloor}\geq t)\cdot(2^{1-\mu}\lambda)^{\mathrm{I}(\pi_\downarrow(u))}. \tag{45}$$

To bound the probabilistic term on the right-hand side of (45), we use the large deviation principle for Gamma variables to write

$$\mathbb{P}(X_{\lfloor\alpha t\rfloor}\geq t)\leq e^{-\lfloor\alpha t\rfloor I_{\exp}(1/\alpha)},$$

where $I_{\exp}$ is the large deviation rate function of the exponential distribution with parameter 1, defined as

$$I_{\exp}(a)=a-1+\log(1/a) \tag{46}$$

for $a>1$. As a result, we get

$$\sum_{\substack{\pi\in\mathcal{T}_0:\ \mathfrak{s}(\pi)=u,\\ \mathrm{I}(\pi)<\lfloor\alpha t\rfloor}} e\cdot z(\pi)\cdot\mathbb{P}(X_{\mathrm{I}(\pi)+1}\geq t)\leq\frac{e\cdot(2^{1-\mu}\lambda)^{\mathrm{I}(\pi_\downarrow(u))}}{1-4\lambda^2}\cdot e^{-\lfloor\alpha t\rfloor I_{\exp}(1/\alpha)}. \tag{47}$$

Next, we bound the second sum on the right-hand side of (42). Similarly to (43), again using Corollary 4.6, we get

$$\sum_{\substack{\pi\in\mathcal{T}_0:\ \mathfrak{s}(\pi)=u,\\ \mathrm{I}(\pi)\geq\lfloor\alpha t\rfloor}} e\cdot z(\pi)\leq e\cdot z(\pi_\downarrow(u))\cdot 2^{\mathrm{I}(\pi_\downarrow(u))}\sum_{k=(\lfloor\alpha t\rfloor-\mathrm{I}(\pi_\downarrow(u)))/2}^{\infty}(4\lambda^2)^k. \tag{48}$$

Bounding $z(\pi_\downarrow(u))$ as in (44), and evaluating the geometric sum in (48) yields

$$\begin{aligned}\sum_{\substack{\pi\in\mathcal{T}_0:\ \mathfrak{s}(\pi)=u,\\ \mathrm{I}(\pi)\geq\lfloor\alpha t\rfloor}} e\cdot z(\pi)&\leq e\cdot(2^{1-\mu}\lambda)^{\mathrm{I}(\pi_\downarrow(u))}\cdot\frac{(4\lambda^2)^{(\lfloor\alpha t\rfloor-\mathrm{I}(\pi_\downarrow(u)))/2}}{1-4\lambda^2}\\ &=\frac{e\cdot 2^{-\mu\mathrm{I}(\pi_\downarrow(u))}}{1-4\lambda^2}\cdot(2\lambda)^{\lfloor\alpha t\rfloor}.\end{aligned} \tag{49}$$

Substituting the bounds (47) and (49) into (42) yields

$$\mathbb{P}\Big(T^{\mathrm{brw}}_{\mathrm{ext}}(\mathcal{T},\mathbb{1}_\varnothing,u)>t\Big)\leq\frac{e\cdot(2^{1-\mu}\lambda)^{\mathrm{I}(\pi_\downarrow(u))}}{1-4\lambda^2}\cdot e^{-\lfloor\alpha t\rfloor I_{\exp}(1/\alpha)}+\frac{e\cdot 2^{-\mu\mathrm{I}(\pi_\downarrow(u))}}{1-4\lambda^2}\cdot(2\lambda)^{\lfloor\alpha t\rfloor}. \tag{50}$$

For $\lambda<1/2$, (50) shows the exponential decay of the local extinction time at u. Since the first term on the right-hand side is increasing in α, whereas the second term is decreasing, the optimized bound is given by $\alpha=\alpha^\star$, where $\alpha^\star$ is the solution of

$$e^{-\lfloor\alpha^\star t\rfloor I_{\exp}(1/\alpha^\star)}=(2\lambda)^{\lfloor\alpha^\star t\rfloor}. \tag{51}$$

Using (46), (51) simplifies to

$$1/\alpha^\star-1+\log(\alpha^\star)=-\log(2\lambda). \tag{52}$$

Since the left-hand side of (52) is strictly decreasing from ∞ to 0 as $\alpha^\star$ increases from 0 to 1, there is exactly one solution $\alpha^\star\in(0,1)$ for any given $\lambda<1/2$. This finishes the proof for $\underline{x}_0=\mathbb{1}_\varnothing$.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

To extend the argument to any starting state $\underline{x}_0$ with $|\underline{x}_0| < \infty$, we make two observations. First, since the above argument is valid for *any* tree $\mathcal{T}$ with *any* fixed root $\varnothing$, (by rerooting the tree) this implies that

$$\mathbb{P}\Big(T^{\mathrm{brw}}_{\mathrm{ext}}(\mathcal{T}, \mathbb{1}_v, u) > t\Big) \le c_1(v)e^{-c_2 t} \tag{53}$$

for any $u, v \in \mathcal{T}$ and $t > 0$ (for $\lambda < 1/2$). Here, the constant $c_1(v)$ further depends on u, λ, μ, while c_2 depends on λ, but, importantly, not on v. Second, when $(\underline{x}_t)_{t\ge 0} = \mathrm{BRW}_{f,\lambda}(\mathcal{T}, \underline{x}_0)$, then by the independent behavior of the particles in BRW, we have that

$$(\underline{x}_t)_{t\ge 0} \overset{d}{=} \left(\sum_{v: x_0(v)>0} \sum_{i=1}^{x_0(v)} \underline{x}_t^{(v,i)} \right)_{t\ge 0}, \tag{54}$$

where $(\underline{x}_t^{(v,i)})_{v,i}$ are independent realizations of the processes $\mathrm{BRW}_{f,\lambda}(\mathcal{T}, \mathbb{1}_v)$. Hence, if $T^{(v,i,u)}_{\mathrm{ext}}$ denotes the local extinction time of $(\underline{x}_t^{(v,i)})_{t\ge 0}$ at u, then a union bound combined with (53) gives

$$\begin{aligned}\mathbb{P}\Big(T^{\mathrm{brw}}_{\mathrm{ext}}(\mathcal{T}, \underline{x}_0, u) > t\Big) &= \mathbb{P}\Big(\max_{v,i} T^{(v,i,u)}_{\mathrm{ext}} > t\Big) \\ &\le \sum_{v: x_0(v)>0} \sum_{i=1}^{x_0(v)} \mathbb{P}\Big(T^{(v,i,u)}_{\mathrm{ext}} > t\Big) \le \sum_{v: x_0(v)>0} \sum_{i=1}^{x_0(v)} c_1(v)e^{-c_2 t},\end{aligned}$$

that is, exponential decay of the distribution of the local extinction time (with the same constant in the exponent for any $|\underline{x}_0|$). This finishes the proof. □

4.6. Max-penalty: global extinction on trees when growth is limited

In this section, we consider rooted trees. The root will always be denoted by $\varnothing$. We always assume that trees have no loops or parallel edges. For any vertex u of $\mathcal{T}$, we keep using the notation $\pi_{\downarrow u}$ for the geodesic from $\varnothing$ to u. Given $\mu > 0$, for each vertex u in $\mathcal{T}$ we let

$$\zeta(u) := \prod_{i=0}^{\mathfrak{l}(\pi)-1} \max(d_{\pi_i}, d_{\pi_{i+1}})^{-\mu}, \quad \text{where } \pi = \pi_{\downarrow u}, \tag{55}$$

so that (recalling (33), and recalling that we exclude parallel edges, so that $\mathrm{e}(\pi_i, \pi_{i+1}) = 1$) we have

$$z(\pi_{\downarrow u}) = \lambda^{\mathfrak{l}(\pi_{\downarrow u})} \cdot \zeta(u). \tag{56}$$

We will write $\mathrm{Gen}_N(\mathcal{T})$ for the set of vertices at graph distance N from $\varnothing$, for $N \in \mathbb{N}$.

Lemma 4.7. *Let $\mathcal{T}$ be a tree with root $\varnothing$. Fix $\mu \in [1/2, 1)$, $\lambda > 0$ and assume that*

$$\sum_{N=0}^{\infty} (2\lambda)^N \sum_{u \in \mathrm{Gen}_N(\mathcal{T})} \zeta(u) < \infty. \tag{57}$$

Then, $\mathrm{BRW}_{f,\lambda}(\mathcal{T}, \mathbb{1}_\varnothing)$ *with penalty function $f(x, y) = \max(x, y)^\mu$ goes extinct globally.*

Proof. We continue using the notation $\mathcal{T}_0 = \{\pi \in \mathcal{T} : \pi_0 = \varnothing\}$ for the set of paths in $\mathcal{T}$ that start at the root. Repeating the estimate in (40) and using (56), for any $N \in \mathbb{N}$ and any vertex $u \in \mathrm{Gen}_N(\mathcal{T})$ we have $\mathfrak{l}(\pi_{\downarrow u}) = N$, so summing over all infection paths ending at u gives

$$\sum_{\pi \in \mathcal{T}_0:\, \mathfrak{s}(\pi) = u} \mathbb{E}[Z(\pi)] = \sum_{\pi \in \mathcal{T}_0:\, \mathfrak{s}(\pi) = u} z(\pi) \le \frac{(2\lambda)^N}{1 - 4\lambda^2} \cdot \zeta(u).$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Then, when summing over all infection paths in the tree, we have

$$\sum_{\pi\in\mathcal{T}_0}\mathbb{E}[Z(\pi)] = \sum_{N=0}^{\infty}\sum_{u\in\mathrm{Gen}_N(\mathcal{T})}\sum_{\pi\in\mathcal{T}_0:\,\mathfrak{s}(\pi)=u}\mathbb{E}[Z(\pi)]$$
$$\leq \frac{1}{1-4\lambda^2}\sum_{N=0}^{\infty}(2\lambda)^N\sum_{u\in\mathrm{Gen}_N(\mathcal{T})}\zeta(u) < \infty$$

by the assumption. This shows that, starting from a single particle at the root, the expected number of particles ever born (overall in $\mathcal{T}$) is finite, so this number is finite almost surely. This implies global extinction. □

In the applications we have in mind, rather than verifying (57) directly, we will verify that

$$\sum_{N=1}^{\infty}(2\lambda)^N\sum_{u\in\mathrm{Gen}_N(\mathcal{T})}\tilde{\zeta}(u) < \infty, \tag{58}$$

where $\tilde{\zeta}(u)$ is defined for all $u \neq \varnothing$ by

$$\tilde{\zeta}(u) := (d_\varnothing)^{-\mu}\cdot\prod_{i=1}^{\mathrm{l}(\pi)-1}(d_{\pi_i}-1)^{-\mu}, \quad \text{where } \pi = \pi_{\downarrow u} \tag{59}$$

where $d_{\pi_i} \geq 2$ as we assumed no vertices are leaves in the tree. We leave $\tilde{\zeta}$ undefined at the root. Clearly, by (55), $\zeta(u) \leq \tilde{\zeta}(u)$ for all $u \neq \varnothing$, so (58) implies (57).

Proof of Theorem 2.5(c). We assume that the offspring distribution of the Galton-Watson tree satisfies $\mathbb{E}[D^{1-\mu}] < \infty$. We claim that, for any $N \geq 1$,

$$\mathbb{E}\left[\sum_{u\in\mathrm{Gen}_N(\mathcal{T})}\tilde{\zeta}(u)\right] = (\mathbb{E}[D^{1-\mu}])^N. \tag{60}$$

This is obvious in case $N = 1$. Assume that it has been proved for N. Recalling that $d_v \geq 2$ for all v except possibly the root, for the induction step, by (59), we note that

$$\sum_{u\in\mathrm{Gen}_{N+1}(\mathcal{T})}\tilde{\zeta}(u) = \sum_{v\in\mathrm{Gen}_N(\mathcal{T})}\tilde{\zeta}(v)\sum_{\substack{u\in\mathrm{Gen}_{N+1}(\mathcal{T}):\\ u\sim v}}(d_v-1)^{-\mu}$$
$$= \sum_{v\in\mathrm{Gen}_N(\mathcal{T})}\tilde{\zeta}(v)\cdot(d_v-1)\cdot(d_v-1)^{-\mu} = \sum_{v\in\mathrm{Gen}_N(\mathcal{T})}\tilde{\zeta}(v)\cdot(d_v-1)^{1-\mu}. \tag{61}$$

Let $\mathcal{T}_N$ denote the truncation of $\mathcal{T}$ at generation N, that is, $\mathcal{T}_N$ is the subgraph of $\mathcal{T}$ induced by the set of vertices at graph distance at most N from $\varnothing$. Note that $\mathcal{T}_N$ does not include information about the offsprings of vertices in generation N, and conditioned on $\mathcal{T}_N$, the sizes of these offsprings are iid, with same law as D. Taking expectations in (61), we have

$$\mathbb{E}\left[\sum_{u\in\mathrm{Gen}_{N+1}(\mathcal{T})}\tilde{\zeta}(u)\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{v\in\mathrm{Gen}_N(\mathcal{T})}\tilde{\zeta}(v)\cdot(d_v-1)^{1-\mu}\middle|\mathcal{T}_N\right]\right]$$
$$= \mathbb{E}\left[\sum_{v\in\mathrm{Gen}_N(\mathcal{T})}\tilde{\zeta}(v)\cdot\mathbb{E}\left[(d_v-1)^{1-\mu}\middle|\mathcal{T}_N\right]\right]$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

$$= \mathbb{E}\left[\sum_{v \in \mathrm{Gen}_N(\mathcal{T})} \tilde{\zeta}(v)\right] \cdot \mathbb{E}[D^{1-\mu}] = (\mathbb{E}[D^{1-\mu}])^{N+1},$$

where the last equality follows from the induction hypothesis. This completes the proof of (60).

Now, if $\lambda < (2\mathbb{E}[D^{1-\mu}])^{-1}$, then

$$\mathbb{E}\left[\sum_{N=1}^{\infty} (2\lambda)^N \cdot \sum_{u \in \mathrm{Gen}_N(\mathcal{T})} \tilde{\zeta}(u)\right] = \sum_{N=1}^{\infty} (2\lambda \cdot \mathbb{E}[D^{1-\mu}])^N < \infty.$$

Hence, $\sum_{N=1}^{\infty} (2\lambda)^N \cdot \sum_{u \in \mathrm{Gen}_N(\mathcal{T})} \tilde{\zeta}(u)$ is finite for almost all realizations of $\mathcal{T}$. It then follows from Lemma 4.7 (and the observation following its proof) that there is global extinction of the penalized branching random walk for almost every realization of $\mathcal{T}$. □

We now see further applications of Lemma 4.7, the proof of Corollary 2.7.

Proof of Corollary 2.7. The case of trees with finite upper branching number b follows from verifying condition (57) with the simple bound $\zeta(u) \le 1$ for all u. For the case of spherically symmetric trees, we can verify condition (58) directly instead of working with the branching number. Note that, for any $N \ge 1$, we have

$$\tilde{\zeta}(u) = (d_0)^{-\mu} \prod_{i=1}^{N-1} (d_i - 1)^{-\mu} \quad \text{for any } u \in \mathrm{Gen}_N(\mathcal{T}),$$

so

$$\sum_{u \in \mathrm{Gen}_N(\mathcal{T})} \tilde{\zeta}(u) = (d_0)^{-\mu} \prod_{i=1}^{N-1} (d_i - 1)^{-\mu} \cdot |\mathrm{Gen}_N(\mathcal{T})| = (d_0)^{1-\mu} \prod_{i=1}^{N-1} (d_i - 1)^{1-\mu},$$

and then

$$(2\lambda)^N \sum_{u \in \mathrm{Gen}_N(\mathcal{T})} \tilde{\zeta}(u) = \exp\left\{N\left(\log(2) + \log(\lambda) + \frac{(1-\mu)(\log d_0)}{N} + \frac{1-\mu}{N} \sum_{i=1}^{N-1} \log(d_i - 1)\right)\right\}.$$

Now, it is easy to check that $\limsup 1/N \cdot \sum_{i=1}^{N-1} \log(d_i - 1) \le \log \overline{\mathrm{br}}(\mathcal{T})$, so if $\lambda < \mathrm{e}^{-(1-\mu)\log \overline{\mathrm{br}}(\mathcal{T})}/2$, then there exists $c < 0$ such that the expression inside parentheses above is smaller than c for N large enough. It readily follows that (58) is satisfied, so global extinction follows from Lemma 4.7. □

4.7. Max-penalty: fast extinction when $\mu \in [1/2, 1)$

We close this section by proving a result that bounds the survival of $\mathrm{BRW}_{f,\lambda}$ for $f(x, y) = \max(x, y)^\mu$, $\mu \in [1/2, 1)$ on any graph, both in space and in time. We will use this result in Section 5 to prove Theorem 2.9(a), stating that the max-penalty contact process goes quickly extinct on the configuration model whenever $\tau > 3$.

We again go back to the genealogic branching random walk construction of Section 3.2. For a graph $G = (V, E)$, recall the definition of the set of genealogical labels $\mathcal{T}$ from Definition 3.3, the notations $\mathfrak{l}(\pi)$ and $\mathfrak{s}(\pi)$, the construction of $(\underline{y}_t)_{t \ge 0}$ in Definition 3.4 and its relation to the branching random walk $(\underline{x}_t)_{t \ge 0}$ given in Lemma 3.5. Here we will take these processes with birth rate λ and max-penalty function with exponent μ, $f(x, y) = \max(x, y)^\mu$, so that $r(\cdot, \cdot)$ is as in (32). As before, for π with

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

$\mathrm{I}(\pi) \geq 1$, $Z(\pi)$ denotes the number of particles with label π born in the whole history of the process. We let $z(\cdot)$ be as in (33). Finally, recall the first backtracking index $\tau(\cdot)$ and the backtracking erasure function $g(\cdot)$ from Definition 4.2.

Lemma 4.8. *Let $\mu \in [1/2, 1)$ and $G = (V, E)$ be a graph with a distinguished vertex $\bar{v}$. Assume that for some constant $\ell > 0$, $\mathrm{e}(u, v) \leq \ell$ for any $u, v \in V$. Fix $N \geq 2$ and let b_N denote the number of non-backtracking paths of length at most N started at $\bar{v}$,*

$$b_N := |\{\pi \in \mathscr{T} : \pi_0 = \bar{v},\ \mathrm{I}(\pi) \leq N,\ \tau(\pi) = \infty\}|. \tag{62}$$

Consider the penalized branching random walk $(\underline{x}_t^{(\bar{v})})_{t \geq 0}$ on G with penalization function $f(x, y) = \max(x, y)^{\mu}$, birth rate $\lambda < 1/(4\ell)$ and started from a single particle, located at $\bar{v}$. Then, for any fixed constant $C > 1$,

$$\begin{aligned} &\mathbb{P}\left(\begin{array}{l} (\underline{x}_t^{(\bar{v})}) \text{ dies before time } CN, \text{ and never reaches} \\ \text{any vertex at graph distance } N \text{ from } \bar{v} \end{array} \right) \\ &\quad > 1 - 2b_N \left(\mathrm{e}\ell \cdot (4\ell\lambda)^N + \mathrm{e}^{-N(C-1)^2/(2C)} \right). \end{aligned} \tag{63}$$

Proof. Let $(\underline{y}_t)_{t \geq 0}$ be the genealogic branching random walk corresponding to $(\underline{x}_t)_{t \geq 0}$ as in Lemma 3.5; in particular, $y_0((\bar{v})) = 1$ and $y_0(\pi) = 0$ for any $\pi \neq \bar{v}$. We note that

$$\begin{aligned} &\left\{ (\underline{x}_t) \text{ is alive at time } CN, \text{ or reaches some vertex at distance } N \text{ from } \bar{v} \right\} \\ &\subset \{ y_{CN}(\pi) > 0 \text{ for some } \pi \in \mathscr{T} \text{ with } \pi_0 = \bar{v},\ \mathrm{I}(\pi) < N \} \\ &\quad \cup \{ y_t(\pi) > 0 \text{ for some } \pi \in \mathscr{T} \text{ with } \pi_0 = \bar{v},\ \mathrm{I}(\pi) = N \text{ and some } t > 0 \}. \end{aligned}$$

Using a union bound and the inequalities $\mathbb{P}(y_{CN}(\pi) > 0) \leq \mathbb{E}[y_{CN}(\pi)]$ and $\mathbb{P}(y_t(\pi) > 0 \text{ for some } t) \leq \mathrm{e} \cdot \mathbb{E}[Z(\pi)] = \mathrm{e} \cdot z(\pi)$ from Corollary 3.7, we have

$$\mathbb{P}\left(\begin{array}{l} (\underline{x}_t) \text{ is alive at time } CN, \text{ or reaches} \\ \text{some vertex at distance } N \text{ from } \bar{v} \end{array} \right) \leq \sum_{\substack{\pi \in \mathscr{T}: \\ \pi_0 = \bar{v}, \\ \mathrm{I}(\pi) < N}} \mathbb{E}[y_{CN}(\pi)] + \mathrm{e} \cdot \sum_{\substack{\pi \in \mathscr{T}: \\ \pi_0 = \bar{v}, \\ \mathrm{I}(\pi) = N}} z(\pi). \tag{64}$$

We bound the two sums in the rhs separately. Using (33) the following bound holds for any path:

$$z(\pi) \leq (\ell\lambda)^{\mathrm{I}(\pi)}, \tag{65}$$

which follows from $\max(d_u, d_v)^{\mu} \geq 1$ and the assumption that $\mathrm{e}(u, v) \leq \ell$.

We first deal with the second sum in (64). Recall that if $\pi' \in (g^{(k)})^{-1}(\pi)$, then $\mathrm{I}(\pi') = \mathrm{I}(\pi) + 2k$. Then, we break the sum as follows:

$$\sum_{\substack{\pi: \pi_0 = \bar{v}, \\ \mathrm{I}(\pi) = N}} z(\pi) = \sum_{\substack{(m,k): \\ m+2k=N}} \sum_{\substack{\pi: \pi_0 = \bar{v}, \\ \mathrm{I}(\pi) = m, \\ \tau(\pi) = \infty}} \sum_{\pi' \in (g^{(k)})^{-1}(\pi)} z(\pi').$$

Using (39) in Corollary 4.6, the right-hand side is at most

$$\sum_{\substack{\pi: \pi_0 = \bar{v}, \\ \mathrm{I}(\pi) = N}} z(\pi) \leq \sum_{\substack{(m,k): \\ m+2k=N}} \sum_{\substack{\pi: \pi_0 = \bar{v}, \\ \mathrm{I}(\pi) = m, \\ \tau(\pi) = \infty}} 2^m \cdot (4\lambda^2\ell)^k \cdot z(\pi).$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Using (65) and b_N from (62), this is at most

$$\sum_{\substack{(m,k):\\ m+2k=N}} \sum_{\substack{\pi:\pi_0=\bar{v},\\ \mathfrak{l}(\pi)=m,\\ \tau(\pi)=\infty}} 2^m \cdot (4\ell\lambda^2)^k \cdot (\ell\lambda)^m$$

$$= \sum_{\substack{(m,k):\\ m+2k=N}} (4\ell)^{m+k} \cdot \lambda^{m+2k} \cdot |\{\pi : \pi_0 = \bar{v},\ \mathfrak{l}(\pi) = m,\ \tau(\pi) = \infty\}|$$

$$\le b_N \cdot \sum_{\substack{(m,k):\\ m+2k=N}} (4\ell)^{m+k} \cdot \lambda^{m+2k}.$$

Using that $m+2k = N$ implies that $m+k = (N+m)/2$ for each $m \in 0, \dots, N$, the above sum is at most

$$\sum_{\substack{\pi:\pi_0=\bar{v},\\ \mathfrak{l}(\pi)=N}} z(\pi) \le b_N \cdot \sum_{m=0}^{N} (4\ell)^{(N+m)/2} \cdot \lambda^N = (2\ell^{1/2}\lambda)^N \cdot b_N \cdot \sum_{m=0}^{N} (2\ell^{1/2})^m$$

$$\le 2\ell^{1/2}(4\ell\lambda)^N \cdot b_N. \tag{66}$$

We now turn to the first term in (64). Using (25), we have

$$\sum_{\substack{\pi:\pi_0=\bar{v},\\ \mathfrak{l}(\pi)<N}} \mathbb{E}[y_{CN}(\pi)] \le \left(\max_{0\le m<N} \frac{(CN)^m}{m!} \mathrm{e}^{-CN} \right) \cdot \sum_{\substack{\pi:\pi_0=\bar{v},\\ \mathfrak{l}(\pi)<N}} z(\pi). \tag{67}$$

Let us bound the sum in the right-hand side using (39) with $\max e(u,v) \le \ell$ and then (65) as

$$\sum_{\substack{\pi:\pi_0=\bar{v},\\ \mathfrak{l}(\pi)<N}} z(\pi) \le \sum_{\substack{\pi:\pi_0=\bar{v},\\ \mathfrak{l}(\pi)<N,\\ \tau(\pi)=\infty}} \sum_{k=0}^{\infty} \sum_{\pi' \in (g^{(k)})^{-1}(\pi)} z(\pi')$$

$$\overset{(39)}{\le} \frac{1}{1-4\ell\lambda^2} \sum_{\substack{\pi:\pi_0=\bar{v},\\ \mathfrak{l}(\pi)<N,\\ \tau(\pi)=\infty}} 2^{\mathfrak{l}(\pi)} z(\pi) \overset{(65)}{\le} \frac{1}{1-4\ell\lambda^2} \sum_{\substack{\pi:\pi_0=\bar{v},\\ \mathfrak{l}(\pi)<N,\\ \tau(\pi)=\infty}} (2\ell\lambda)^{\mathfrak{l}(\pi)}.$$

Since $\lambda < 1/(4\ell)$ with $\ell \ge 1$, we have $\frac{1}{1-4\ell\lambda^2} < 2$ and $2\ell\lambda < 1/2 < 1$, so the last factor in (67) is smaller than

$$2|\{\pi : \pi_0 = \bar{v},\ \mathfrak{l}(\pi) < N,\ \tau(\pi) = \infty\}| \le 2b_N. \tag{68}$$

Next, the expression inside the maximum in (67) equals $\mathbb{P}(W = m)$ for W having the Poisson(CN) distribution. We bound

$$\max_{0\le m<N} \mathbb{P}(W = m) \le \mathbb{P}(W \le N).$$

We use a Chernoff bound for Poisson random variables: for $X \sim \text{Poisson}(\nu)$ we have $\mathbb{P}(X \le \nu - t) \le \mathrm{e}^{-t^2/(2\nu)}$, see [67, Exercise 2.21]. This gives

$$\mathbb{P}(W \le N) \le \exp\left\{ -\frac{(CN-N)^2}{2CN} \right\} = \exp\left\{ -\frac{(C-1)^2}{2C} \cdot N \right\}.$$

Combining this with (68) in (67) and (66) completes the proof of (63). □

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

5. The configuration model: fast extinction via loop erasure

In this section we prove Theorem 2.9(a). This theorem says that the contact process $\mathrm{CP}_{f,\lambda}$ and the branching random walk $\mathrm{BRW}_{f,\lambda}$ go extinct quickly for small $\lambda > 0$ on the configuration model when $f(x,y) = \max(x,y)^{\mu}$ with $\mu \in [1/2, 1)$ and the degree distribution is lighter than a power-law with exponent $\tau > 3$. The proof idea is the following. Fixing a large constant ℓ, first, we show that with probability $1 - o(1/n)$, there are at most ℓ surplus edges in the r-neighborhood $B_r(u_n)$ of a uniformly chosen vertex u_n with $r = \delta \log n$ for some small $\delta > 0$. That is, one can remove at most ℓ edges from $B_{\delta \log n}(u_n)$ to obtain a tree. Then, we apply Lemma 4.8 to show that the expected number of particles of $\mathrm{BRW}_{f,\lambda}$ on infection paths in $B_{\delta \log n}(u_n)$ that reach the boundary $\partial B_{\delta \log n}(u_n)$ decays exponentially for small λ. This implies that $\mathrm{BRW}_{f,\lambda}$ dies out inside $B_{\delta \log n}(u_n)$ before reaching $\partial B_{\delta \log n}(u_n)$ with probability at least $1 - o(1/n)$. A union bound over the n vertices then finishes the proof.

Our first goal is to prove a statement about the surplus edges of $B_{\delta \log n}(u_n)$, and then we move on to the analysis of infection paths of $\mathrm{BRW}_{f,\lambda}$. The number of surplus edges of a (sub)graph $H = (V_H, E_H)$ is given by $|E_H| - (|V_H| - 1)$. Recall the configuration model from Definition 1.9 and that $e(u,v)$ denotes the number of edges between vertices u, v.

Proposition 5.1. *Consider the configuration model with degree sequence $\underline{d}_n$ satisfying Assumption 1.10, and Assumptions 1.11 and 1.12 with some $\tau, \varepsilon, c_u, z_0$ (for all sufficiently large n) with $\tau(1-\varepsilon) > 3$. Fix some $\delta > 0$. Let u_n be a uniformly chosen vertex in $[n]$ and let $\mathrm{Surp}_{\delta \log n}(u_n)$ denote the number of surplus edges in $B_{\delta \log n}(u_n)$. Then, for all $\varepsilon' \in (0, (\tau(1-\varepsilon)-3)/2$ there exists $\delta > 0$ and $\delta' > 0$ so that for any $\ell > (\tau(1-\varepsilon)-1)/(\tau(1-\varepsilon)-3-2\varepsilon')$*

$$\mathbb{P}(|B_{\delta \log n}(u_n)| \geq n^{(1+\varepsilon')/(\tau(1-\varepsilon)-1)} \text{ or } \mathrm{Surp}_{\delta \log n}(u_n) \geq \ell) \leq n^{-1-\delta'} \tag{69}$$

Finally, for any $\ell > 3 \vee (\tau(1-\varepsilon)-1)/(\tau(1-\varepsilon)-3)$, there exists some $\delta' > 0$ that

$$\mathbb{P}(\max_{u,v \in [n]} e(u,v) \geq \ell) \leq n^{-1-\delta'}. \tag{70}$$

Observe that with probability $1/n$ the root's degree is the maximal degree in the graph, which can be as high as $O(n^{1/(\tau(1-\varepsilon)-1)})$, so $\varepsilon' > 0$ in (69) is necessary for the bound to be true. The condition $\varepsilon' \in (0, (\tau(1-\varepsilon)-3)/2$ ensures on the one hand that $\zeta := (1+\varepsilon')/(\tau(1-\varepsilon)-1) < 1/2$ and on the other hand that the required lower bound $(\tau(1-\varepsilon)-1)/(\tau(1-\varepsilon)-3-2\varepsilon') = 1/(1-2\zeta)$ on ℓ is positive. If one aims to bound the maximal multiplicity of edges inside $B_{\delta \log n}$, the inequality (69) also includes that, since multiple edges also count as surplus edges. For generality we include the stronger result in (70) here.

The proof is based on a breadth-first-search exploration process of $B_{\delta \log n}(u_n)$, and a coupling to a (power-law) branching process tree $\mathcal{T}^{\#}_{\delta \log n}$ so that the tree contains $B_{\delta \log n}(u_n)$. First we give a good bound on the size of the tree that holds with probability $1 - o(1/n)$. When the offspring distribution decays exponentially, this is fairly easy, but when it follows for instance a power law, we need to develop some new bounds.

Hence, the next lemma bounds the kth moment of the size of (truncated) power-law BP trees, but before that, we give some definitions. Let $(\zeta_n)_{n \geq 1}$ be a sequence of discrete measures on $\mathbb{N}$ that satisfies

$$\tau' \notin \mathbb{N}: \quad \zeta_n(z) \leq c_u z^{-(\tau'-1)}, \qquad M_n := \max \mathrm{support}(\zeta_n) \leq C_u n^{1/(\tau'-1)}, \tag{71}$$

Usually ζ_n is the size-biased measure of an empirical degree sequence $\underline{d}_n$ satisfying Assumptions 1.10 and (1.12). For each integer $k \geq 1$, there exists $C_k > 0$ such that, if n is large enough, the k-th moment can be bounded by an integral of the rhs of (71) yields that the kth moment

$$\sum_{z=1}^{M_n} z^k \cdot \zeta_n(z) \leq c_u + \sum_{z=1}^{M_n} c_u z^{k-(\tau'-1)} \leq C_k \cdot n^{h_k} \tag{72}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

where

$$h_k := \max\{(k+1)/(\tau'-1) - 1,\ 0\}. \tag{73}$$

Whenever $\tau' > 2$, the coefficient of $k+1$ in h_k is positive but less than 1. Thus, $k \mapsto h_k$ is non-decreasing and, due to the additive term -1, for any k, ℓ, the super-additivity property holds:

$$h_k + h_\ell \le h_{k+\ell}. \tag{74}$$

Lemma 5.2. *Let $\mathcal{T}^{\#}$ be a Galton-Watson tree with offspring distribution ζ_n satisfying (71) with $\tau' > 3$, and for each r, let Z_r be the size of its generation r. For any integer $k \ge 1$, there exists $\mathfrak{C}_k > 0$ such that the following holds for all sufficiently large n:*

$$\mathbb{E}[(Z_r)^k] \le \mathfrak{C}_k \cdot n^{h_k} \cdot \mathrm{e}^{\mathfrak{C}_k r} \qquad \textit{for all } r \ge 0. \tag{75}$$

The criterion $\tau' > 3$ is important: this guarantees that the mean offspring $\mathbb{E}[\mathcal{X}] = \mathbb{E}[\mathcal{X}_n]$ does not grow with n. BPs with $\tau' \in (2,3)$ grow doubly-exponentially, and (75) does not hold for them. The importance here is that the rhs of (75) only depends on the generation number r exponentially, that is, the constant $\mathfrak{C}_k$ in the exponential growth does not depend on n. This is non-trivial, since the k-th moment of the offspring distribution itself does, but it only enters the bound *once*, as the prefactor n^{h_k}.

Proof of Lemma 5.2. We will argue by induction over k. Let $\mathcal{X}$ be a random variable distributed as ζ_n (we will generally omit the dependence on n).

For the base case $k = 1$, recalling (75), note that $h_1 = 0$ since $\tau' > 2$; hence, $\mathbb{E}[\mathcal{X}]$ is bounded by the constant C_1 which does not depend on n (equivalently, in h_k the maximum is at 0 in (73)). The right-hand side of (75) is satisfied in this case since

$$\mathbb{E}[Z_r] = \mathbb{E}[\mathcal{X}]^r \le (C_1)^r = n^{h_1} \cdot \mathrm{e}^{\log(C_1) r}.$$

Now assume that we have proved (75) for $j = 1, \dots, k-1$, that is, assume that we have already found constants $\mathfrak{C}_1, \dots, \mathfrak{C}_{k-1}$ such that

$$\mathbb{E}[(Z_r)^j] \le \mathfrak{C}_j \cdot n^{h_j} \cdot \mathrm{e}^{\mathfrak{C}_j r} \qquad \text{for all } j \in \{1, \dots, k-1\} \text{ and all } r \ge 0, \tag{76}$$

and we want to find $\mathfrak{C}_k$ so that (75) holds. Let $\mathfrak{f}(s)$ denote the probability-generating function of $\mathcal{X}$,

$$\mathfrak{f}(s) := \sum_{z \ge 1} s^z \cdot \mathbb{P}(\mathcal{X} = z) =: \sum_{z \ge 1} s^z \cdot \widehat{\nu}_n(z), \qquad s \in \mathbb{R}.$$

Since ζ_n has finite support, $\mathfrak{f}$ is well defined for any s; it is also infinitely differentiable, with derivative of order m at $s = 1$ satisfying

$$\mathfrak{f}^{(m)}(1) = \mathbb{E}[\mathcal{X}(\mathcal{X}-1)\cdots(\mathcal{X}-m+1)].$$

For any $r \in \mathbb{N}$, let $\mathfrak{f}_r$ denote the r-fold composition of $\mathfrak{f}$ with itself (i.e., $\mathfrak{f}_0$ is the identity function, $\mathfrak{f}_1 = \mathfrak{f}$ and $\mathfrak{f}_r = \mathfrak{f} \circ \mathfrak{f}_{r-1}$ for $r > 1$). It is well-known that $\mathfrak{f}_r$ is the probability-generating function of Z_r [3], which is again well defined and infinitely differentiable for all s,

$$\mathfrak{f}_r(s) = \sum_{z=1}^{\infty} s^z \cdot \mathbb{P}(Z_r = z), \quad s \in \mathbb{R}, \quad \text{and} \quad \mathfrak{f}_r^{(m)}(1) = \mathbb{E}[Z_r(Z_r-1)\cdots(Z_r-m+1)].$$

We claim that there exists $\mathfrak{C}'_k > 0$ such that

$$\mathfrak{f}_r^{(k)}(1) \le \mathfrak{C}'_k \cdot n^{h_k} \cdot \mathrm{e}^{\mathfrak{C}'_k r} \qquad \text{for all } r \ge 0. \tag{77}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Before proving this, let us show how to use it together with the induction hypothesis to obtain (75) (with a constant $\mathfrak{C}_k$ that is possibly different from $\mathfrak{C}'_k$). We bound

$$\begin{aligned}\mathbb{E}[(Z_r)^k] &\le \mathbb{E}[Z_r(Z_r-1)\cdots(Z_r-(k-1))] \\ &\quad + |\mathbb{E}[(Z_r)^k] - \mathbb{E}[Z_r(Z_r-1)\cdots(Z_r-(k-1))]| \\ &\le \mathfrak{f}_r^{(k)}(1) + \sum_{j=1}^{k-1} |a_{k-1,j}| \cdot \mathbb{E}[Z_r^j],\end{aligned}$$

where $a_{k-1,j}$ is the coefficient of x^j in the polynomial $x(x-1)\cdots(x-(k-1))$. By (77) and the induction hypothesis, the right-hand side above is smaller than

$$\mathbb{E}[(Z_r)^k] \le \mathfrak{C}'_k \cdot n^{h_k} \cdot \mathrm{e}^{\mathfrak{C}'_k r} + \sum_{j=1}^{k-1} |a_{k-1,j}| \cdot \mathfrak{C}_j \cdot n^{h_j} \cdot \mathrm{e}^{\mathfrak{C}_j r}.$$

Since $j \mapsto h_j$ is increasing, we can choose $\mathfrak{C}_k$ (not depending on n or r) such that the above expression is smaller than $\mathfrak{C}_k \cdot n^{h_k} \cdot \mathrm{e}^{\mathfrak{C}_k r}$ for all r. This proves (75) once (77) is proved. To prove (77), fix $r \ge 1$. We start by writing

$$\mathfrak{f}_r^{(k)}(1) = (\mathfrak{f} \circ \mathfrak{f}_{r-1})^{(k)}(1). \tag{78}$$

We will use the chain rule for higher-order derivatives (also known as Faà di Bruno's formula); let us briefly state it. Let $f, g : I \to \mathbb{R}$ be functions defined in an open interval I containing $s \in \mathbb{R}$. Fix $k \in \mathbb{N}$ and assume that f and g are k times differentiable in s. Let $\mathcal{P}_k$ denote the set of partitions of $\{1, \dots, k\}$. For some $\mathcal{P} = \{B_1, \dots, B_\ell\} \in \mathcal{P}_k$, we let $|\mathcal{P}| = \ell$ be the number of blocks in $\mathcal{P}$, and for $B \in \mathcal{P}$ similarly we write $|B|$ for the number of elements in B. Let then $\mathcal{P}_{k,\ell} \subset \mathcal{P}_k$ be the set of partitions containing ℓ blocks. Then,

$$\begin{aligned}(f \circ g)^{(k)}(s) &= \sum_{\mathcal{P} \in \mathcal{P}_k} f^{(|\mathcal{P}|)}(g(s)) \cdot \prod_{B \in \mathcal{P}} g^{(|B|)}(s) \\ &= \sum_{\ell=1}^{k} \sum_{\{B_1,\dots,B_\ell\} \in \mathcal{P}_{m,\ell}} f^{(\ell)}(g(s)) \cdot \prod_{j=1}^{\ell} g^{(|B_j|)}(s),\end{aligned}$$

Using this formula with $f = \mathfrak{f}$ and $g = \mathfrak{f}_{r-1}$ (together with $\mathfrak{f}_{r-1}(1) = 1$) in (78), we have

$$\mathfrak{f}_r^{(k)}(1) = (\mathfrak{f} \circ \mathfrak{f}_{r-1})^{(k)}(1) = \sum_{\ell=1}^{k} \sum_{\{B_1,\dots,B_\ell\} \in \mathcal{P}_{k,\ell}} \mathfrak{f}^{(\ell)}(1) \cdot \prod_{j=1}^{\ell} \mathfrak{f}_{r-1}^{(|B_j|)}(1). \tag{79}$$

We now inspect each term in (79). The value $\ell = 1$ gives the trivial partition which consists of a single block $\{1, \dots, k\}$. The corresponding term is

$$\mathfrak{f}'(1) \cdot \mathfrak{f}_{r-1}^{(k)}(1) = \mathbb{E}[\mathcal{X}] \cdot \mathfrak{f}_{r-1}^{(k)}(1) \qquad \text{when } \ell = 1. \tag{80}$$

Now fix a partition $\mathcal{P} = \{B_1, \dots, B_\ell\}$ with $\ell \ge 2$. The corresponding term in (79) equals

$$\begin{aligned}\mathfrak{f}^{(\ell)}(1) \cdot \prod_{j=1}^{\ell} \mathfrak{f}_{r-1}^{(|B_j|)}(1) &= \mathbb{E}[\mathcal{X}(\mathcal{X}-1)\cdots(\mathcal{X}-\ell+1)] \\ &\quad \cdot \prod_{j \le \ell} \mathbb{E}[Z_{r-1}(Z_{r-1}-1)\cdots(Z_{r-1}-|B_j|+1)] \\ &\le \mathbb{E}[\mathcal{X}^\ell] \cdot \prod_{j \le \ell} \mathbb{E}[(Z_{r-1})^{|B_j|}] \overset{(72)}{\le} C_\ell \cdot n^{h_\ell} \cdot \prod_{j \le \ell} \mathbb{E}[(Z_{r-1})^{|B_j|}].\end{aligned}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Since $\ell \geq 2$, each block has size $|B_j| < k$. We thus use the induction hypothesis (76) to bound the rhs as

$$\begin{aligned}\mathfrak{f}^{(\ell)}(1) \cdot \prod_{j=1}^{\ell} \mathfrak{f}_{r-1}^{(|B_j|)}(1) &\leq C_\ell \cdot n^{h_\ell} \cdot \prod_{j=1}^{\ell} \mathfrak{C}_{|B_j|} \cdot n^{h_{|B_j|}} \cdot \mathrm{e}^{\mathfrak{C}_{|B_j|}(r-1)} \\ &= C_\ell \Big(\prod_{j=1}^{\ell} \mathfrak{C}_{|B_j|} \cdot \mathrm{e}^{\sum_{j=1}^{\ell} \mathfrak{C}_{|B_j|} \cdot (r-1)} \Big) \cdot n^{h_\ell + \sum_{j=1}^{\ell} h_{|B_j|}} \\ &\leq c' \mathrm{e}^{C'(r-1)} \cdot n^{h_\ell + \sum_{j=1}^{\ell} h_{|B_j|}}, \end{aligned} \tag{81}$$

where c', C' are constants that neither depend on r nor on the partition $\mathcal{P}$, and are given by

$$c' := (\max_{i \leq k} C_i) \cdot (\max_{i \leq k-1} \mathfrak{C}_i)^k, \qquad C' := k \cdot \max_{i \leq k-1} \mathfrak{C}_i.$$

We inspect the exponent of n that appears in (81), and set out to prove the inequality

$$h_\ell + \sum_{j \leq \ell} h_{|B_j|} \leq h_k. \tag{82}$$

We consider two cases. The first case is when $h_\ell = 0$. The superadditivity (74) yields that

$$h_\ell + \sum_{j \leq \ell} h_{|B_j|} \leq h_{\sum |B_j|} = h_k.$$

The second case is $h_\ell > 0$, with a more involved proof. Recall h_k from (73). We write $\alpha := \frac{1}{\tau'-1}$ and $\beta := 1 - \frac{1}{\tau'-1}$, so that $h_i = \max(\alpha i - \beta, 0)$ for any i, and carry out some formal rearrangements:

$$\begin{aligned} h_\ell + \sum_{j \leq \ell} h_{|B_j|} &= \alpha\ell - \beta + \sum_{j : h_{|B_j|} > 0} (\alpha |B_j| - \beta) = -\beta + \sum_{j=1}^{\ell} \alpha + \sum_{j : h_{|B_j|} > 0} (\alpha |B_j| - \beta) \\ &= -\beta + \sum_{j : h_{|B_j|} = 0} \alpha + \sum_{j : h_{|B_j|} > 0} (\alpha |B_j| + \alpha - \beta) \\ &\leq -\beta + \sum_{j : h_{|B_j|} = 0} \alpha |B_j| + \sum_{j : h_{|B_j|} > 0} (\alpha |B_j| + \alpha - \beta). \end{aligned}$$

By the assumption in the lemma that $\tau' > 3$, $\alpha - \beta < 0$. Since $\{B_1, \dots, B_\ell\} \in \mathcal{P}_{k,\ell}$, that is, the blocks partition $\{1, \dots, k\}$, $\sum_{j \leq \ell} |B_j| = k$ holds which gives that

$$h_\ell + \sum_{j \leq \ell} h_{|B_j|} \leq -\beta + \sum_{j : h_{|B_j|} = 0} \alpha |B_j| + \sum_{j : h_{|B_j|} > 0} \alpha |B_j| = \alpha k - \beta = h_k.$$

This completes the proof of (82). We substitute it as an upper bound in (81) to obtain that for any $\mathcal{P} \in \mathcal{P}_{k,\ell}$ for any $\ell \geq 2$,

$$\mathfrak{f}^{(\ell)}(1) \cdot \prod_{B \in \mathcal{P}} \mathfrak{f}_{r-1}^{(|B|)}(1) \leq c' \mathrm{e}^{C'(r-1)} \cdot n^{h_k}.$$

Substituting this bound into (79) and using (80) for $\ell = 1$, we arrive at

$$\mathfrak{f}_r^{(k)}(1) \leq c'' \mathrm{e}^{C'(r-1)} \cdot n^{h_k} + \mathbb{E}[\mathcal{X}] \cdot \mathfrak{f}_{r-1}^{(k)}(1),$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

where $c'' := c' \cdot |\mathcal{P}_k| = c' 2^k$. This bound can now be used recursively: the same inequality (with $(r, r-1)$ replaced by $(r-1, r-2)$) can be used to bound $\mathfrak{f}_{r-1}^{(k)}(1)$ on the right-hand side, and then further. This gives

$$\begin{aligned}\mathfrak{f}_r^{(k)}(1) &\le c'' \mathrm{e}^{C'(r-1)} \cdot n^{h_k} \cdot (1 + \mathbb{E}[\mathcal{X}] + \mathbb{E}[\mathcal{X}]^2 + \cdots + \mathbb{E}[\mathcal{X}]^r) \\ &\le c'' \mathrm{e}^{C'(r-1)} \cdot \mathbb{E}[\mathcal{X}]^{r+1} \cdot n^{h_k}.\end{aligned}$$

Now, we can choose $\mathfrak{C}_k' > 0$ such that the right-hand side above is smaller than $\mathfrak{C}_k' \mathrm{e}^{\mathfrak{C}_k' r} \cdot n^{h_k}$ for all r. This completes the proof of (77). □

We now proceed to embed $B_r(u_n)$ in Proposition 5.1 to a branching process that satisfies the conditions of Lemma 5.2. Recall $\nu_n(z) = n_z/n$ from (8). Define the *size-biased version* and the *down-shifted size-biased version* of ν_n as

$$\nu_n^\star(z) := \frac{z\nu_n(z)}{\mathbb{E}[D_n]}, \qquad \text{and} \qquad \widetilde{\nu}_n(z) := \frac{(z+1)\nu_n(z+1)}{\mathbb{E}[D_n]} = \frac{(z+1)n_{z+1}}{\sum_{i \le n} d_i}. \tag{83}$$

If $D_n \sim \nu_n$, $D_n^\star \sim \nu_n^\star$, then $D_n^\star - 1 \sim \widetilde{\nu}_n$. It is well-known that $\nu_n \overset{d}{\le} \nu_n^\star$, that is, the size-biased version of a random variable on $\mathbb{N}$ stochastically dominates the original measure. This follows from Harris' inequality: $\mathbb{P}(D_n^\star > z)\mathbb{E}[D_n] = \mathbb{E}[D_n \mathbb{1}_{\{D_n > z\}}] \ge \mathbb{E}[D_n]\mathbb{P}(D_n > z)$ for any $z > 0$. The next definition makes the tail of any starting distribution ν having a $q > 1$ moment slightly heavier so that it also stochastically dominates ν.

Definition 5.3 (η-heavier-transformation of a probability measure). Let ν be a probability measure so that $\nu(z) \le z^{-\tau'}$ holds for all sufficiently large $z > 0$. Let η satisfy that $\tau'(1-\eta) > 1$, and given a distribution ν, let $z_0^\# \ge 1$ be the smallest integer that satisfy the following:

$$\min_{z \ge z_0^\# : \nu(i) \ne 0} \nu(i)^{-\eta} \ge 8/7, \qquad \sum_{z \ge z_0^\#} \nu(i)^{1-\eta} < 7/8. \tag{84}$$

Choose a normalizing factor $Z := Z(\eta, \nu)$ so that the following measure is a probability measure:

$$\nu^\#(z) := \begin{cases} 0 & \text{if } z \le z_0^\#, \\ \nu(z)^{1-\eta}/Z & \text{if } z > z_0^\# \end{cases} \tag{85}$$

The choice $7/8$ is quite arbitrary in (84), any number strictly less than 1 would serve our purposes.

Claim 5.4 (Stochastic domination between ν and $\nu^\#$). Let ν be a probability measure so that for some $\tau' > 1$, $\nu(z) \le z^{-\tau'}$ holds for all sufficiently large $z > 0$. Then the measure $\nu^\#$ exists and stochastically dominates ν for all η satisfying $\tau'(1-\eta) > 1$, and has finite q-th moment for all $q < \tau'(1-\eta) - 1$. Finally, $Z < 7/8$.

Proof. Suppose the measure exists. Then $Z < 7/8$ follows from the second criterion in (84) since $Z = \sum_{i > z_0^\#} \nu(i)^{1-\eta} \le 7/8$. For $z \le z_0^\#$, $\nu^\#([0,z]) \le \nu([0,z])$ is immediate from the first row in (85). For $z > z_0^\#$, we aim to show $\nu((z,\infty)) \le \nu^\#((z,\infty))$, which is equivalent to

$$Z \sum_{i > z} \nu(i) \le \sum_{i > z} \nu(i)^{1-\eta},$$

which holds since $Z < 7/8$ and $\nu(i) < 1$ implies that $\nu(i) \le \nu(i)^{1-\eta}$ for each $i > z$. To see the moment conditions, for all $z \ge 1$ it holds that $Z\nu^\#(z) \le \nu(z)^{1-\eta}$, and so the qth moment is finite whenever $\sum_{z \ge z_0^\#} z^q \nu(z)^{1-\eta} < \infty$, which in turn is at most $\sum_{z \ge z_0^\#} z^q z^{-\tau'(1-\eta)}$. This sum is convergent if $q - \tau'(1-\eta) < -1$, equivalently if $q < \tau'(1-\eta) - 1$. This also gives with $q = 0$ that $\tau'(1-\eta) > 1$ is indeed sufficient for $z_0^\#$ in (84) to exist and the normalizing factor Z to be finite. □

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

The following exploration process gradually constructs the configuration model by matching half-edges sequentially in a way that reveals the graph neighborhood of a vertex v_0, $B_r(v_0)$, in a breadth-first search manner. The exploration also immediately couples the r-neighborhood $B_r(v_0)$ to the first r generations of a random rooted tree $\mathcal{T}_r^{\mathrm{expl}}$ so that $|B_r(v_0)| \le |\mathcal{T}_r^{\mathrm{expl}}|$ holds a.s. under the coupling.

Construction 5.5 (Exploration of the neighborhood of a vertex). We take as input a degree sequence $\underline{d}_n$, a starting vertex v_0, a target radius r, and an additional offspring distribution ζ. The coupled exploration of $B_r(v_0)$ in the configuration model $\mathrm{CM}(\underline{d}_n)$ is then as follows:

Step 0. Initialization. To initialize, we set v_0 active and reveal its half-edges (say $h_1, \dots h_{d_{v_0}}$) and set also all of its half-edges active. We introduce the list of the active vertices $A_v(0) := \{v_0\}$ and of the active half-edges $A_h(0) := \{h_1, \dots, h_{d_{v_0}}\}$, and we set $\mathrm{Ex}_v(0) := \emptyset, \mathrm{Ex}_h(0) := \emptyset$ for the list of explored vertices and half-edges, respectively.

Step s. Exploring a half-edge. In each discrete step $s \ge 1$ we take the first half-edge h_s from $A_h(s-1)$, in a first-in-first-out (breadth-first search) order, and reveal the half-edge $m(h_s)$ it is matched to. We then append h_s and $m(h_s)$ to the end of the list of explored half-edges $\mathrm{Ex}_h(s-1)$, obtaining $\mathrm{Ex}_h(s)$, and we remove h_s from the active half-edges $A_h(s-1)$, and also remove $m(h_s)$ from it if it happened to belong to $A_h(s-1)$. Then we carry out three more substeps:

Substep s.(i): Adding newly discovered vertices. If the vertex $v(m(h_s))$ that $m(h_s)$ is attached to is a new vertex, that is, not in $A_v(s-1)$, then we append $v(m(h_s))$ to the end of the list $A_v(s-1)$, obtaining $A_v(s)$, and we append the remaining $X_s^{(n)}$ many half-edges of $v(m(h_s))$ to the end of the active half-edge list, obtaining $A_h(s)$. We call $X_s^{(n)}$ the forward degree of the vertex discovered in step s.

Substep s.(ii) Handling loops and creating ghost subtrees. If, however, the half-edge $m(h_s)$ is already active and it is attached to an active vertex $v(m(h_s))$, then we call this a *collision* at step s. This creates a loop and hence a surplus edge in $B_r(v_0)$. We then do the following: in $B_r(v_0)$ we create the loop formed by $(h_s, m(h_s))$, and in $\mathcal{T}_r^{\mathrm{expl}}$ we create two "ghost" subtrees as follows. Let $r_1 := d_G(v_0, v(h_s)), r_2 := d_G(v_0, v(m(h_s)))$, respectively. We then sample two independent branching processes, $\mathcal{T}_{r-r_1}^{\#,s1}$ and $\mathcal{T}_{r-r_2}^{\#,s2}$ with offspring distribution ζ, (the first one has depth $r - r_1$ while the second one has depth $r - r_2$) and attach their root to the half-edges h_s and $m(h_s)$ respectively, and add these ghost-subtrees to $\mathcal{T}_r^{\mathrm{expl}}$.

Substep s.(iii): Checking for vertices being fully explored. If the half-edges of the vertices $v(h_s)$ and/or $v(m(h_s))$ are *all* explored after substep s.(ii), then we append $v(h_s)$ and/or $v(m(h_s))$ also to the set of explored vertices $\mathrm{Ex}_v(s)$, otherwise we keep them active.

Stopping condition. The exploration stops when we have matched all half-edges belonging to vertices at graph distance $r-1$ from v_0. We denote the number of needed steps by $t(r)$.

Output. The output is the graph $B_r(v_0)$ and the tree $\mathcal{T}_r^{\mathrm{expl}}$. We denote the number of half-edges added in step s to the active half-edges by $X_s^{(n)}$, giving the random sequence $X_1^{(n)}, X_2^{(n)}, \dots, X_{t(r)}^{(n)}$, with the convention that we set $X_s^{(n)} := 0$ if a collision have occurred at step s and no new vertex was added. We denote by $\mathrm{Coll}_r(v_0)$ the number of collisions that occurred during the process.

Observation 5.6. The exploration reveals the whole graph (including all loops) within $B_{r-1}(v_0)$, and also the size of $B_r(v_0)$. To see the latter, by the stopping condition, we have explored all vertices in generation $r-1$, and their forward degrees, say $X_{s_{r-1}}^{(n)}, \dots, X_{t_{r-1}}^{(n)}$ are thus known. Matching then all these half-edges reveals edges between at least one vertex in generation $r-1$, and the other vertex can be either in generation $r-1$ or r. For each edge where the other vertex is also in generation $r-1$, a loop between two vertices in generation $r-1$ arises, and the size of $B_r(v_0)$ is reduced by 2 compared to $\sum_{i\in[s_{r-1},t_{r-1}]} X_i^{(n)}$. Each collision where two edges lead to the same vertex in generation r, reduces the size of $B_r(v_0)$ compared to $\sum_{i\in[s_{r-1},t_{r-1}]} X_i^{(n)}$ by 1. Note that $|B_r(v_0)| \le |\mathcal{T}_r^{\mathrm{expl}}|$ for any offspring distribution ζ.

Observation 5.7. All surplus edges are either self-loops, multiple edges, or between two vertices, say v, v' so that the distance between $|d_G(u_n, v) - d_G(u_n, v')| \le 1$. Indeed, when a surplus edge is created,

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

the half-edge h_s is matched to an active half-edge in $A_h(s-1)$. All half-edges in $A_h(s-1)$ either belong to the same generation as $v(h_s)$ or they belong to the next generation.

Recall the size biasing from (83) and the hash-transformation of a measure from (85) in Definition 5.3.

Lemma 5.8. *Consider Construction 5.5 started from a uniformly chosen vertex $v_0 := u_n$ on the configuration model $\mathrm{CM}(\underline{d}_n)$ so that $(\underline{d}_n)_{n\geq 1}$ satisfies Assumptions 1.10 and 1.12 with some $\tau, \varepsilon, c_u, z_0$ for all sufficiently large n so that $\tau(1-\varepsilon) > 2$ in (11). Let $\eta > 0$ be so that $(\tau(1-\varepsilon)-1)(1-\eta) > 1$. Assume that the number of exploration steps $t(r) \leq \sum_{i\in[n]} d_i/17$. Then, for all sufficiently large n, the forward-degree sequence $(X_s^{(n)})_{s\leq t(r)}$ is stochastically dominated by an iid sequence $(Y_s)_{s\leq t(r)}$ from $(\nu_n^\star)^\#$ defined from (83) and (85). Under Assumption 1.12 this measure satisfies for some constant c_u':*

$$(\nu_n^\star)^\#(z) \leq c_u' z^{-(\tau(1-\varepsilon)-1)(1-\eta)}. \tag{86}$$

As a result, there exists a coupling $B_r(u_n) \subseteq \mathcal{T}_r^{\mathrm{expl}} \subseteq \mathcal{T}_r^\#$ where $\mathcal{T}_r^\#$ is the first r generations of a branching process having iid offspring from $(\nu_n^\star)^\#(z)$.

Remark 5.9. With the same method it could also be proved that $(X_s^{(n)})_{s\leq t(r)}$ is stochastically dominated by an iid sequence $(Z_s)_{s\leq t(r)}$ from $(\widetilde{\nu}_n)^\#$ defined from (83) and (85), the η-heavier transformation of the down-shifted size-biased version of ν_n. In that case, however, the root's degree d_{u_n} cannot necessarily be dominated by $(\widetilde{\nu}_n)^\#$. Further, $(\widetilde{\nu}_n)^\#$ and $(\nu_n^\star)^\#$ both satisfy the same inequality (86), so for simplicity we dominate by a "usual" GW tree $\mathcal{T}_r^\#$ where all vertices have the same offspring distribution.

The proof will follow from the following statement and Construction 5.5.

Claim 5.10 (Domination and size-biasing during the exploration). Let ν_n be the empirical measure of $\underline{d}_n = (d_1, \dots d_n)$ in (8) satisfying that $\nu_n^\star(z) \leq c_u z^{-\tau'}$ for all $z \geq z_0$ for some $\tau' > 1$ in (83). For a subset $\Delta \subset [\sum_{i\leq n} d_i]$, remove the half-edges with label in Δ to obtain a new degree sequence $\underline{d}_n^\Delta := (d_1' \mathbb{1}_{d_1'\neq 0}, \dots d_n' \mathbb{1}_{d_n'\neq 0})$, and let $\nu_{n,\Delta}^\star$ denote the size-biased version of the empirical distribution of $\underline{d}_n^\Delta$. Then, for any choice of Δ with $|\Delta| \leq (\sum_{i\in[n]} d_i)/8$, $\nu_{n,\Delta}^\star$ is stochastically dominated by $(\nu_n^\star)^\#$ for any η so that $\tau'(1-\eta) > 1$.

Proof. We assume here that $\nu_n^\star(z) \leq c_u z^{-\tau'}$ for all $z \geq z_0$. Then, Claim 5.4 gives that $\nu_n^\star$ is stochastically dominated by $(\nu_n^\star)^\#$ whenever $\tau'(1-\eta) > 1$. So when $\Delta = \emptyset$ then the statement holds. Recall that $\nu_n(z) = n_z/n$, and let $h_n := \sum_{i\in[n]} d_i$. Then $\sum_{i\in[n]} d_i' \mathbb{1}_{d_i'\neq 0} = h_n - |\Delta|$ since we removed $|\Delta|$ many half-edges. Recall $z_0^\#$ from (84) and (85). Let us first consider any $z < z_0^\#$. Clearly $\nu_{n,\Delta}^\star([0,z]) \geq 0$ while $(\nu_n^\star)^\#([0,z]) = 0$ so the criterion for stochastic domination $\nu_{n,\Delta}^\star([0,z]) \geq (\nu_n^\star)^\#([0,z])$ holds in this case. Let now $z \geq z_0^\#$. Observe that all degrees can only decrease by removing half-edges, hence writing $n_i' := \sum_{j\in[n]} \mathbb{1}_{d_j'=i}$ for the number of vertices of degree i after removing the half-edges with label in Δ, it holds that $\sum_{i>z} i n_i' \leq \sum_{i>z} i n_i$ relating to (83). Now we look at the upper tail using that $|\Delta| \leq n\mathbb{E}[D_n]/8$

$$\nu_{n,\Delta}^\star((z,\infty)) = \frac{\sum_{i>z} i n_i'}{h_n - |\Delta|} \leq \frac{\sum_{i>z} i n_i}{n\mathbb{E}[D_n](1-1/8)} = \nu_n^\star((z,\infty)) \cdot 8/7.$$

At the same time, using that $Z < 7/8$ in Claim 5.4, the tail of $(\nu_n^\star)^\#$ satisfies:

$$(\nu_n^\star)^\#((z,\infty)) = \frac{1}{Z}\Big(\sum_{i>z} \nu_n^\star(i)^{1-\eta}\Big) \leq \nu_n^\star((z,\infty))/Z \geq \nu_n^\star((z,\infty)) \cdot 8/7.$$

Hence the stochastic domination criterion $\nu_{n,\Delta}^\star((z,\infty)) \leq (\nu_n^\star)^\#((z,\infty))$ is satisfied. □

Proof of Lemma 5.8. Let us write $h_n := \sum_{i\leq n} d_i$. Consider step (s) of Construction 5.5, when we match half-edge h_s. Its pair $m(h_s)$ is chosen uniformly among the available $h_n - 2s - 1$ many half-edges at step s. At this point the half-edges *not* available for matching to h_s form the set $\Delta_s := \mathrm{Ex}_h(s-1) \cup \{h_s\}$. Consider the "available" degrees at this moment, say $\underline{d}_n^{(s)} := (d_1^{(s)} \mathbb{1}_{d_1^{(s)}\neq 0}, \dots, d_n^{(s)} \mathbb{1}_{d_n^{(s)}\neq 0})$, where $d_j^{(s)}$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

is the number of not-matched half-edges of vertex j before step s if h_s is not attached to vertex j and 1 less if h_s is attached to vertex j. Since we choose the half-edge $m(h_s)$ uniformly at random from the currently available half-edges, the vertex $v(m(h_s))$ that $m(h_s)$ is attached to is chosen *size-biasedly* from $\underline{d}_n^{(s)}$, conditionally independently of previous matchings, that is, its forward degree then

$$\mathbb{P}(X_s^{(n)} = i-1) = \mathbb{P}(d^{(s)}_{v(m(h_s))} = i) = \frac{i\sum_{j\in[n]} \mathbb{1}_{\{d_j^{(s)}=i\}}}{h_n - 2s - 1} = \nu^\star_{n,\Delta_s}(i),$$

with $\Delta_s := \mathrm{Ex}_h(s-1) \cup \{h_s\}$. In particular $X_s^{(n)} + 1$ follows the measure $\nu^\star_{n,\Delta_s}(i)$ in Claim 5.10. Thus, let us apply Claim 5.4 with $\Delta_s := \mathrm{Ex}_h(s-1) \cup \{h_s\}$, that is, removing the set of unavailable half-edges. Since $t_r \le h_n/17$, we have $|\Delta_s| \le 2h_n/17 + 1 \le h_n/8$ so Claim 5.10 applies. By Claim 5.10, the measure $\nu^\star_{n,\Delta_s}(i)$ is stochastically dominated by $(\nu_n^\star)^\#$ for each s, so let Y_s be such a random variable. Using the conditional independence of the consecutive matchings, one can thus construct a coupling where $X_s^{(n)} \le X_s^{(n)} + 1 \le Y_s$ and Y_s are iid from $(\nu_n^\star)^\#$. Further, since u_n is a vertex chosen uniformly at random, the root's degree d_{u_n} has distribution ν_n. By below (83), the measure $\nu_n^\star$ stochastically dominates ν_n. So it holds that

$$\nu_n \overset{d}{\le} \nu_n^\star \overset{d}{\le} (\nu_n^\star)^\#,$$

and thus one can construct a coupling where $d_{u_n} \le Y_0$ with Y_0 from $(\nu_n^\star)^\#$. To finish, recall that whenever the exploration discovers a loop at some step s, it appends two ghost subtrees to the half-edges h_s and $m(h_s)$ exactly so that their last generation ends at distance r from u_n. Setting the offspring distribution of these branching processes to be also $(\nu_n^\star)^\#$ gives then a coupling where $B_r(0)$ is embedded in $\mathcal{T}_r^{\mathrm{expl}}$ which are both embedded in $\mathcal{T}_r^\#$, a branching process where all vertices have iid degree from $\widehat{\nu}_n$.

Using Assumption 1.12 we now bound $(\nu_n^\star)^\#(z)$ for all $z \ge z_0^\# \vee z_0$. Since we assumed $\nu_n(z) \le c_u z^{-\tau(1-\varepsilon)}$ with $\tau(1-\varepsilon) > 2$, it holds for some finite constant $\overline{m}$ that $\mathbb{E}[D_n] < \overline{m} < \infty$ uniformly for all n, and Assumption 1.10 also ensures that $\mathbb{E}[D_n] \ge \underline{m}$ for some $\underline{m}$, uniformly for all n. Hence $\nu_n^\star(z) \le c_u(z+1)z^{-\tau(1-\varepsilon)}/\underline{m}$ for all n and all $z \ge z_0$. Finally, for all $z \ge z_0^\# \vee z_0$

$$(\nu_n^\star)^\#(z) = \frac{1}{Z}\nu_n^\star(z)^{1-\eta} = \frac{c_u^{1-\eta}}{Z\underline{m}^{1-\eta}} z^{1-\eta} z^{-\tau(1-\varepsilon)(1-\eta)} \le c_u' z^{-(\tau(1-\varepsilon)-1)(1-\eta)},$$

which proves (86). The condition $(\tau(1-\varepsilon)-1)(1-\eta) > 1$ is necessary for the hash-measure to exist in Claim 5.4. □

We are ready to prove Proposition 5.1.

Proof of Proposition 5.1. We start by applying Lemma 5.8. This gives that $B_r(u_n)$ is contained in a BP tree $\mathcal{T}_r^\#$ as long as the number of half-edges explored is $t(r) < n\mathbb{E}[D_n]/17$, with offspring distribution $\widehat{\nu}_n$ defined in (86). Next, we ensure that this measure satisfies the conditions (71) so that we can use the moment bounds of Lemma 5.2. To see (71) is satisfied, we observe that $\widehat{\nu}_n$ has power-law exponent $\tau' - 1 := (\tau(1-\varepsilon)-1)(1-\eta) > 2$, that is, $\tau' > 3$, and we can easily ensure that $\tau' \notin \mathbb{N}$ by changing η if necessary. The condition on the maximum of the support in (71) follows from Assumption 1.12 since the exponent $1/(\tau(1-\varepsilon)-1)$ there is less than $1/(\tau'-1)$ which is allowed in (71). Hence Lemma 5.2 is applicable for the BP tree $\mathcal{T}_r^\#$ in Lemma 5.8.

By Observation 5.6, in order to also bound the surplus edges in $B_{\delta\log n}$ we need to reveal the size of one more generation, and so we set out to bound $|\mathcal{T}^\#_{\delta\log n+1}|$ for some $\delta > 0$. Set $r_n := \delta\log n + 1$. Let $k \in \mathbb{N}$, and $\zeta > 0$ to be determined later. We use first the increasing function x^k, then Markov's inequality, and then Minkowski's inequality in the second inequality:

$$\mathbb{P}\big(|\mathcal{T}^\#_{r_n}| \ge n^\zeta\big) = \mathbb{P}\big((|\mathcal{T}^\#_{r_n}|)^k \ge n^{k\zeta}\big) \le n^{-k\zeta}\mathbb{E}\left[\left(\sum_{i\le k_n} Z_i\right)^k\right] \le n^{-k\zeta}\left(\sum_{i\le r_n}\mathbb{E}\big[Z_i^k\big]^{1/k}\right)^k.$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

We now apply Lemma 5.2 on $\mathbb{E}[Z_i^k]$ for each $i \le r_n$:

$$\mathbb{P}\big((|\mathcal{T}^{\#}_{r_n}|)^k \ge n^{k\zeta}\big) \le n^{-k\zeta}\left(\sum_{i\le r_n}(\mathfrak{C}_k \cdot n^{h_k}\cdot \mathrm{e}^{\mathfrak{C}_k i})^{1/k}\right)^k = n^{-k\zeta}\cdot \mathfrak{C}_k \cdot n^{h_k}\left(\sum_{i\le r_n}\mathrm{e}^{\mathfrak{C}_k i/k}\right)^k.$$

The sum on the rhs is geometric and since $\mathfrak{C}_k > 0$, it is at most $C'\mathrm{e}^{(\mathfrak{C}_k/k)r_n}$ for some constant C', with $r_n = \delta \log n + 1$, which gives

$$\mathbb{P}\big((|\mathcal{T}^{\#}_{r_n}|)^k \ge n^{k\zeta}\big) \le n^{-k\zeta}\cdot \mathfrak{C}_k \cdot n^{h_k} C'^k n^{\delta \mathfrak{C}_k} = C n^{-k\zeta + h_k + \delta\mathfrak{C}_k}.$$

We inspect the exponent of n. Recall that $h_k = ((k+1)/(\tau'-1) - 1) \vee 0$ from (73). Since $\tau' - 1 > 2$, we may write

$$-k\zeta + h_k + \delta\mathfrak{C}_k = -k(\zeta - \tfrac{1}{\tau'-1}) - (1 - \tfrac{1}{\tau'-1}) + \delta\mathfrak{C}_k. \tag{87}$$

The exponent of n can be made strictly less than -1 for sufficiently large k if $\zeta > 1/(\tau'-1)$. Since $\tau' - 1 = (\tau(1-\varepsilon) - 1)(1-\eta)$ with η arbitrarily small, this yields the formulation $|B_{\delta\log n}(u_n)| > n^{(1+\varepsilon')/(\tau(1-\varepsilon)-1)}$ in (69) of the proposition. For any such ζ one can now choose $k \in \mathbb{N}$ so large that the exponent goes below -1, in particular any k satisfying $k > \zeta/(\tau'-1) - 1$ is a good choice. Given k, one now chooses δ small enough so that the whole exponent in (87) still stays below -1, giving also $\delta' > 0$.

By the coupling $B_{\delta\log n}(u_n) \subseteq B_{\delta\log n+1}(u_n) \subseteq \mathcal{T}^{\#}_{\delta\log n+1}$, we have just proved

$$\mathbb{P}(\mathcal{A}_{\mathrm{size}}) := \mathbb{P}\big(|B_{\delta\log n+1}(u_n)| \le n^{(1+\varepsilon')/(\tau(1-\varepsilon)-1)}\big) \ge 1 - n^{-1-\delta'}, \tag{88}$$

and then by monotonicity $\{|B_{\delta\log n}(u_n)| \le n^{(1+\varepsilon')/(\tau(1-\varepsilon)-1)}\}$ also holds with the same error probability. Now we start bounding the surplus edges inside $B_{\delta\log n}(u_n)$. On the event $\mathcal{A}_{\mathrm{size}}$, the exploration in Construction 5.5 finishes in $t(\delta\log n) \le n^{\zeta}$ with $\zeta := (1+\varepsilon')/(\tau(1-\varepsilon)-1)$ steps, and by Observation 5.6, the exploration reveals $B_{\delta\log n}(u_n)$ and all surplus edges inside. We estimate the probability of a collision from above at each step of the exploration. When the exploration is at step s, a collision happens if the half-edge h_s is matched to one of the active half-edges in $A_h(s-1)$, see substep s.(ii) in Construction 5.5. The size of $A_h(s-1)$ is at any time no more than the total size of $B_{\delta\log n+1}(u_n)$, that is, at most n^{ζ}. Hence, since $s \le n^{\zeta}$ on $\mathcal{A}_{\mathrm{size}}$ also, and so

$$\mathbb{P}(\text{a surplus edge is created at step } s) \le \frac{n^{\zeta}}{h_n - 2s - 1} \le \frac{2}{\mathbb{E}[D_n]} n^{\zeta-1} := c n^{\zeta-1},$$

uniformly for all $s \le n^{\zeta}$, and conditionally independent of other steps. One can thus dominate the sequence of indicators of whether a surplus edge is created at step s by an iid sequence of n^{ζ} many Bernoulli random variables with mean $2n^{\zeta-1}\mathbb{E}[D_n]$. Thus, the number of collisions is at most $\mathrm{Bin}(n^{\zeta}, cn^{\zeta-1})$. Since $\zeta = (1+\varepsilon')/(\tau(1-\varepsilon)-1)$, and we assumed $\tau(1-\varepsilon) - 1 > 2$ and $\varepsilon' \in (0, (\tau(1-\varepsilon)-3)/2$, we have $\zeta < 1/2$, and so the mean, $\Theta(n^{2\zeta-1})$ tends to zero for ε' in this interval. For some ℓ to be chosen later, we bound

$$\begin{aligned}\mathbb{P}(\mathrm{Surp}_{\delta\log n}(u_n) \ge \ell \mid \mathcal{A}_{\mathrm{size}}) &\le \mathbb{P}(\mathrm{Bin}(n^{\zeta}, cn^{\zeta-1}) \ge \ell)\\ &\le \sum_{i\ge\ell}\binom{n^{\zeta}}{i}(cn^{\zeta-1})^i \le \sum_{i=\ell}^{\infty}(cn^{2\zeta-1})^i \le c' n^{(2\zeta-1)\ell},\end{aligned}$$

where we used that $\binom{n^{\zeta}}{i} \le n^{\zeta i}$, and that the geometric sum in the middle has base less than 1 for all sufficiently large n since $2\zeta - 1 < 0$. Choose now ℓ so large that the exponent of n on the rhs, $(2\zeta-1)\ell < -1$, that is, $\ell > 1/(1-2\zeta)$. Then one has for some $\delta' > 0$ that

$$\mathbb{P}(\mathrm{Surp}_{\delta\log n}(u_n) \ge \ell \mid \mathcal{A}_{\mathrm{size}}) \le n^{-1-\delta'}. \tag{89}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

One can compute using ζ that $\ell \geq \frac{\tau(1-\varepsilon)-1}{\tau(1-\varepsilon)-3-2\varepsilon'}$ which also shows that $\tau(1-\varepsilon) > 3$ is necessary for the argument to work. Combining now (88) with (89) with a union bound finishes the proof of (69). Finally we estimate the maximal multiplicity of the edges in the whole graph. We introduce $\mathbb{1}_{u,v}^{(\ell)} := 1$ if there are at least ℓ edges between vertex u and v. Then by Markov's inequality, and pairing ℓ chosen half-edges from u and from v together yields that

$$\begin{aligned}\mathbb{P}\Big(\max_{u,v\in[n]} e(u,v) \geq \ell\Big) &= \mathbb{P}\Big(\sum_{u,v\in[n]} \mathbb{1}_{u,v}^{(\ell)} \geq 1\Big) \leq \mathbb{E}\Big[\sum_{u,v\in[n]} \mathbb{1}_{u,v}^{(\ell)}\Big] \\ &\leq \sum_{u,v} \frac{d_u^\ell d_v^\ell}{(h_n - 2\ell - 1)^\ell} \leq cn^2\mathbb{E}[D_n^\ell]^2/n^\ell\end{aligned} \tag{90}$$

for some constant $c > 0$. Using (11) and (12) in Assumption 1.12, with $M_n := C_u n^{1/(\tau(1-\varepsilon)-1)}$ and so one bounds the moment as

$$\begin{aligned}\mathbb{E}[D_n^\ell] &\leq \sum_{z\leq M_n} c_u z^{\ell-\tau(1-\varepsilon)} \leq c\int_1^{M_n} z^{\ell-\tau(1-\varepsilon)}\mathrm{d}z \\ &\leq CM_n^{(\ell+1-\tau(1-\varepsilon))\vee 0} = n^{(\ell/(\tau(1-\varepsilon)-1)-1)\vee 0},\end{aligned}$$

similarly to h_ℓ in (73). If now the maximum is at 0 in the exponent, one obtains $\ell > 3$ is necessary for the exponent to be below -1, and if the maximum is at the other term $\ell/(\tau(1-\varepsilon)-1) - 1$ then one obtains $\ell > (\tau(1-\varepsilon)-1)/(\tau(1-\varepsilon)-3)$ then the exponent in (90) is less than -1. Hence $\ell > 3 \vee (\tau(1-\varepsilon)-1)/(\tau(1-\varepsilon)-3)$ is a sufficient choice, finishing the proof of (70) and thus the proposition. □

With Proposition 5.1 at hand, we now move on to analyze the contact process on $B_{\delta\log n}(u_n)$. On the event in (69), $B_{\delta\log n}(u_n)$ has at most ℓ surplus edges. By Observation 5.7, all the surplus edges created during the exploration are either self-loops, multiple edges, or the distance between the root u_n and the two end-vertices of the surplus edge differ by at most 1. We will apply the next lemma to bound the number of nonbacktracking infection paths of the contact process on $B_{\delta\log n}(u_n)$.

Recall from Definition 3.3 that $\mathcal{T}(G)$ denotes the genealogical label of particles in the contact process, equivalently, the set of possible infection paths π on G. Recall also that $\mathfrak{l}(\pi)$ is the length of the path (number of edges) from (21), while $\tau(\pi)$ in (34) denotes the location of the first backtracking step on the path, with the convention that $\tau(\pi) = \infty$ if the path is nonbacktracking.

Lemma 5.11. *Let $\mathcal{T} = (V, E)$ be a tree with root $\varnothing$; assume that $\mathcal{T}$ has no self-loops or parallel edges. Let $N, k \in \mathbb{N}$. Let $u_1, v_1, u_2, v_2, \ldots, u_k, v_k \in V$ be (not necessarily distinct) vertices such that for all $i \in \{1, \ldots, k\}$*

$$0 \leq \mathrm{dist}_{\mathcal{T}}(\varnothing, u_i) \leq \mathrm{dist}_{\mathcal{T}}(\varnothing, v_i) \leq \mathrm{dist}_{\mathcal{T}}(\varnothing, u_i) + 1. \tag{91}$$

Consider another graph $\mathcal{T}^{(k)}$ on the same vertex set V, with edge set $E' := E \cup \{u_1, v_1\} \cup \ldots \cup \{u_k, v_k\}$. Let $\mathcal{T}_N := \{v \in V : \mathrm{dist}_{\mathcal{T}}(\varnothing, v) \leq N\}$ as before, and define

$$\mathcal{B}_N := \{\pi \in \mathcal{T}(\mathcal{T}^{(k)}) : \pi_0 = \varnothing,\ \mathfrak{l}(\pi) \leq N,\ \tau(\pi) = \infty\}. \tag{92}$$

Then $|\mathcal{B}_N| \leq (2k+1)^N |\mathcal{T}_N|$.

The lemma allows for self-loops and multiple edges, these also satisfy (91).

Proof. We start by introducing a labeling of the directed edges of any path $\pi \in \mathcal{B}_N$, describing whether the edge uses a surplus edge in one of the two possible directions, or the edge is not a surplus edge. So

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

introducing the symbol o for the latter, we define the set of possible labels $\mathcal{L}$, and we then introduce Seq_N as the set of length-N sequences with elements from $\mathcal{L}$ with a vertex in $\mathcal{T}$ appended at the end:

$$\mathcal{L} := \bigcup_{1\le i\le k} \{(u_i, v_i), (v_i, u_i)\} \cup \{o\}, \tag{93}$$

$$\mathrm{Seq}_N := \{(s_1, s_2, \dots, s_N, v) : \ s_j \in \mathcal{L}, v \in V, \mathrm{dist}_{\mathcal{T}}(\varnothing, v) \le N\}. \tag{94}$$

Observe that $|\mathcal{L}| \le 2k + 1$ (self-loops and multiple edges can make this inequality strict) and thus $|\mathrm{Seq}_N| \le (2k+1)^N |\mathcal{T}_N|$. Therefore, if we show that there is an injection from $\mathcal{B}_N$ to Seq_N, it will yield

$$|\mathcal{B}_N| \le |\mathrm{Seq}_N| \le (2k+1)^N |\mathcal{T}_N|,$$

proving the lemma. We now construct this injection.

Fix any $\pi = (\pi_0, \pi_1, \dots, \pi_m) \in \mathcal{B}_N$, where $m = \mathfrak{l}(\pi)$. We think of this path as the sequence $(e_1, e_2, \dots, e_m)$ with $e_j = (\pi_{j-1}, \pi_j)$ a directed edge. By the definition of $\mathcal{B}_N$ in (92), $m \le N$. Recalling the labels from (93), for each $1 \le j \le m$ define

$$s_j := \begin{cases} (u_i, v_i) & \text{if } e_j = (u_i, v_i) \text{ for some } i \le k, \\ (v_i, u_i) & \text{if } e_j = (v_i, u_i) \text{ for some } i \le k, \\ o & \text{otherwise.} \end{cases}$$

Furthermore, define $s_j = o$ for each $m + 1 \le j \le N$, and finally, let $v = \pi_m$. By the condition (91), each edge in π can only change the distance from $\varnothing$ by at most 1, thus $\mathrm{dist}_{\mathcal{T}}(\varnothing, v) \le N$. Hence, we associate a vector $L(\pi) := (s_1, \dots, s_N, v) \in \mathrm{Seq}_N$ to each $\pi \in \mathcal{B}_N$. We will show that this mapping is injective, that is, $(s_1, \dots, s_N, v)$ uniquely encodes the path π.

For each $1 \le j \le N$ the label s_j reveals whether the edge e_j crosses one of the surplus edges $\{u_i, v_i\}$, and if so, in which direction. Between two consecutive crossings, π is a nonbacktracking path on the edges of the tree $\mathcal{T}$, hence it is uniquely determined, since in a tree there is a single nonbacktracking path between any two vertices: for example, if $s_j = (u_i, v_i)$ and $s_{j'} = (u_{i'}, v_{i'})$ for $j < j'$ and $s_{j+1} = \dots = s_{j'-1} = o$, then $(\pi_j, \dots, \pi_{j'-1})$ is the unique geodesic (i.e., nonbacktracking shortest path) in $\mathcal{T}$ from v_i to $u_{i'}$. A similar argument shows that if $j_{\max} = \max\{j : s_j \ne o\}$, then $(\pi_{j_{\max}}, \dots, \pi_{\mathfrak{l}(\pi)})$ is the unique geodesic in $\mathcal{T}$ from the endpoint of $s_{j_{\max}}$ to v, the endpoint of π. This shows that the defined map is indeed injective, finishing the proof. □

Proof of Theorem 2.9(a). Let G_n be a realization of $\mathrm{CM}(\underline{d}_n)$. Recalling from Lemma 3.8 the stochastic domination between CP and BRW, and that a branching random walk with initial configuration $\underline{\xi}_0$ can be realized as the sum of independent BRWs, each started from a single particle present in $\underline{\xi}_0$, we obtain that

$$\mathrm{CP}_{f,\lambda}(G_n, \underline{1}_{G_n}) \overset{d}{\le} \mathrm{BRW}_{f,\lambda}(G_n, \underline{1}_{G_n}) = \sum_{v\in[n]} \mathrm{BRW}_{f,\lambda}(G_n, \mathbb{1}_v) =: \sum_{v\in[n]} \underline{x}_t^{(v)},$$

where the branching random walks $\underline{x}_t^{(v)}$ are independent given G_n. Let now T_{ext} denote the extinction time of $\mathrm{BRW}_{f,\lambda}(G_n, \underline{1}_{G_n})$, and let $T_{\mathrm{ext}}^{(v)}$ denote the extinction time of $\underline{x}_t^{(v)}$. Then $T_{\mathrm{ext}} = \max_{v\in[n]} T_{\mathrm{ext}}^{(v)}$. Hence for any $t > 0$,

$$\mathbb{P}\big(T_{\mathrm{ext}} > t\big) = \mathbb{P}\big(\exists v \in [n] : T_{\mathrm{ext}}^{(v)} > t\big) \le n \cdot \Big(\frac{1}{n} \sum_{v\in[n]} \mathbb{P}\big(T_{\mathrm{ext}}^{(v)} > t\big)\Big) = n \cdot \mathbb{P}\big(T_{\mathrm{ext}}^{(u_n)} > t\big), \tag{95}$$

where u_n is a uniformly chosen vertex. We will show that for some $C > 0$, $\mathbb{P}\big(T_{\mathrm{ext}}^{(u_n)} > C \log n\big) = o(1/n)$. which then shows that the extinction time is $O_{\mathbb{P}}(\log n)$ by (95).

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

We first apply Proposition 5.1, which is applicable since its conditions coincide with that of Theorem 2.9(a). Proposition 5.1 then gives constants $\delta, \delta', \varepsilon', \ell > 0$ and $\zeta := (1+\varepsilon')/(\tau(1-\varepsilon)-1) < 1/2$ so that the event

$$\mathcal{A}_{\mathrm{good}}(u_n) := \{\max_{u,v\in[n]} e(u,v) \le \ell\} \cap \left\{|B_{\delta\log n}(u_n)| \le n^{\zeta}\right\} \cap \left\{\mathrm{Surp}_{\delta\log n}(u_n) \le \ell\right\}$$

holds with probability $1 - 2n^{-1-\delta'}$. On the event $\mathcal{A}_{\mathrm{good}}(u_n)$, there are at most ℓ surplus edges in $B_{\delta\log n}(u_n)$, so we may apply Lemma 5.11 to see that the set of nonbacktracking infections paths in $B_{\delta\log n}(u_n) = \mathcal{T}^{(\ell)}$ starting at u_n of length $N = \delta\log n$, defined in (92) satisfies on the event $\mathcal{A}_{\mathrm{good}}(u_n)$ that

$$|\mathcal{B}_{\delta\log n}| \le (2\ell+1)^{\delta\log n}|\mathcal{T}_{\delta\log n}| \le n^{\delta\log(2\ell+1)}|B_{\delta\log n}(u_n)| \le n^{\zeta+\delta\log(2\ell+1)}.$$

Now we apply Lemma 4.8, with ℓ as the maximal number of multiple edges and $\bar{v} := u_n$. The main result there, (63) turns into, with $N = \delta\log n$ and $\lambda < 1/(4\ell)$,

$$\begin{aligned}\mathbb{P}\Bigg(&\begin{array}{l}(\underline{x}_t^{(u_n)}) \text{ dies before time } C\delta\log n, \text{ and never reaches}\\ \text{any vertex at graph distance } \delta\log n \text{ from } u_n\end{array}\Bigg)\\ &> 1 - 2|\mathcal{B}_{\delta\log n}|\Big(\mathrm{e}\ell\cdot(4\ell\lambda)^{\delta\log n} + \mathrm{e}^{-\delta\log n(C-1)^2/(2C)}\Big)\\ &\ge 1 - 2n^{\zeta+\delta\log(2\ell+1)}\Big(\mathrm{e}\ell n^{-\delta|\log(4\ell\lambda)|} + n^{-\delta(C-1)^2/2C}\Big).\end{aligned} \tag{96}$$

Distributing the brackets, there are two error terms, the first one is

$$2e\ell\cdot n^{\zeta+\delta\log(2\ell+1)-\delta|\log(4\ell\lambda)|} \le n^{-1-\delta'}$$

whenever $4\ell\lambda$ is small enough so that the exponent of n goes below $-1-\delta'$, in particular when

$$\lambda < \frac{1}{4\ell}\exp\Big(-\tfrac{1}{\delta}(1+\delta'+\zeta+\delta\log(2\ell+1))\Big). \tag{97}$$

The second error term is

$$2n^{\zeta+\delta\log(2\ell+1)-\delta(C-1)^2/(2C)} \le n^{-1-\delta'}$$

whenever C is so large that the exponent of n goes below $-1-\delta'$, in particular using that $(C-1)^2/(2C) > (C-1)/4$ the exponent is below -1 whenever

$$C > 1 + 4\tfrac{1}{\delta}(1+\delta'+\zeta+\delta\log(2\ell+1)).$$

This shows that for all λ sufficiently small (satisfying (97)), the event in (96) holds with probability at least $1-n^{-1-\delta'}$. On this event, the process $\underline{x}_t^{(u_n)}$ never leaves the ball $B_{\delta\log n}(u_n)$, in particular the process never sees other parts of the graph. In other words, extinction of $\underline{x}_t^{(u_n)}$ on $B_{\delta\log n}(u_n)$ without reaching the boundary of $B_{\delta\log n}(u_n)$ implies extinction of $\underline{x}_t^{(u_n)}$ on G_n. Hence, the event $\{T_{\mathrm{ext}}^{(u_n)} > C\delta\log n\}$ is covered by the complement of the event in (96), $\mathbb{P}(T_{\mathrm{ext}}^{(u_n)} > C\delta\log n) \le 2n^{-1-\delta'}$. Substituting this back to (95) finishes the proof. □

6. Proofs of survival on Galton-Watson trees

In this section we present proofs of survival regimes. We start with (only) global survival – Theorem 2.5(b), then we prove Theorems 2.1 and 2.5 (a) in Section 6.2.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

6.1. Max-penalty: global survival via infinite infection rays on heavy tailed GW trees

To prove global survival for the max-penalty with $\mu \in [1/2, 1)$ on GW trees with sufficiently fat-tailed offspring distributions, that is, Theorem 2.5(b), we will show the existence of a (random) infinite *ray* in the Galton-Watson tree on which the infection survives forever.

Definition 6.1 (Down-directed contact process). Let $\mathcal{T}$ be any (given) tree with root $\varnothing$. Consider the directed graph $\mathcal{T}^{\downarrow}$ where each edge $\{u, v\}$ of $\mathcal{T}$ is directed away from the root, that is, from parent to child. Then we denote by $\mathrm{CP}^{\downarrow}_{f,\lambda}(\mathcal{T}, \underline{\xi}_0) = (\underline{\xi}^{\downarrow}_t)_{t\geq 1}$ the degree-penalized contact process in Definition 1.1 on the *directed* graph $\mathcal{T}^{\downarrow}$ with initial state $\underline{\xi}_0$.

One can obtain the down-directed contact process $\mathcal{T}^{\downarrow}(\mathcal{T}, \underline{\xi}_0)$ from the graphical construction of the original $\mathrm{CP}_{f,\lambda}(\mathcal{T}, \underline{\xi}_0)$ by deleting the Poisson point processes that represent infections from child to parent (i.e., upward in the tree), and leaving only those infection paths intact which only contain parent-to-child infection events. Hence, for every given tree $\mathcal{T}$ and starting state $\underline{\xi}_0 \in \{0,1\}^{V(\mathcal{T})}$ it holds that

$$\mathrm{CP}^{\downarrow}_{f,\lambda}(\mathcal{T}, \underline{\xi}_0) \overset{d}{\leq} \mathrm{CP}_{f,\lambda}(\mathcal{T}, \underline{\xi}_0). \tag{98}$$

The next proposition shows that $\mathrm{CP}^{\downarrow}_{f,\lambda}$ survives globally with positive probability on a Galton-Watson tree:

Proposition 6.2. *Let $\mathcal{T}$ be a Galton-Watson tree with offspring distribution D satisfying Definition 1.7 for some $\alpha > 0$ and $\mathbb{P}(D \geq 1) = 1$. Suppose $f(x, y) = \max(x, y)^{\mu}$, and moreover $\mu + \alpha < 1$. Then the down-directed contact process $\mathrm{CP}^{\downarrow}_{f,\lambda}(\mathcal{T}, \mathbb{1}_{\varnothing})$ exhibits global survival with positive probability on $\mathcal{T}$ for any $\lambda > 0$, for almost all realizations $\mathcal{T}$ of the Galton-Watson tree.*

Proof of Theorem 2.5(b). The result follows from Proposition 6.2 by using the stochastic domination in (98). □

Proof of Proposition 6.2. In this proof we denote by $D_v = d_v - 1$ the out-degree (number of children) of the vertex v in $\mathcal{T}^{\downarrow}$. Let $\mathcal{A}_{\mathrm{glob}}$ be the event that $\mathrm{CP}^{\downarrow}_{f,\lambda}(\mathcal{T}, \mathbb{1}_{\varnothing})$ survives globally. Let $\mathcal{B}_K = \{\exists t_0 \geq 0, \exists v \in V(\mathcal{T}), \deg(v) \geq K : \xi^{\downarrow}_{t_0}(v) = 1\}$ be the event that $\mathrm{CP}^{\downarrow}_{f,\lambda}$ ever reaches a vertex with degree at least K for a large enough K decided later. This event has strictly positive probability p_K with lower bound depending only on K, since $p_K \geq \mathbb{P}(D_{\varnothing} \geq K) > 0$.

$$\mathbb{P}(\mathcal{A}_{\mathrm{glob}}) \geq \mathbb{P}(\mathcal{B}_K)\mathbb{P}(\mathcal{A}_{\mathrm{glob}} \mid \mathcal{B}_K), \tag{99}$$

so it is enough to show that $\mathbb{P}(\mathcal{A}_{\mathrm{glob}} \mid \mathcal{B}_K) > 0$ for some large enough K. Fix some constants $1 < s_1 < s_2$ to be chosen later.

Consider a vertex v with degree $D_v = L \geq K$ in the Galton-Watson tree and let $\mathcal{N}(v, [L^{s_1}, L^{s_2}])$ and $N(v, [L^{s_1}, L^{s_2}])$ be the set and number of children of v in $\mathcal{T}$ with degrees in $[L^{s_1}, L^{s_2}]$, respectively. Since the children have iid degrees, $N(v, [L^{s_1}, L^{s_2}]))$ is binomially distributed with parameters L and $\mathbb{P}(D \in [L^{s_1}, L^{s_2}])$. We bound its mean from below using (6). Given some $\varepsilon \in [0, \alpha(s_2 - s_1)/(s_2 + s_1))$, assuming $L > K_0(\varepsilon)$ so that (6) holds,

$$\begin{aligned}\mathbb{E}[N(v, [L^{s_1}, L^{s_2}]) \mid D_v = L] &= L\big(\mathbb{P}(D \geq L^{s_1}) - \mathbb{P}(D \geq L^{s_2})\big) \\ &\geq L\Big(\frac{1}{L^{s_1(\alpha+\varepsilon)}} - \frac{1}{L^{s_2(\alpha-\varepsilon)}}\Big) = L^{1-\alpha s_1 - \varepsilon s_1}\Big(1 - L^{-\alpha(s_2-s_1)+\varepsilon(s_2+s_1)}\Big).\end{aligned}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

By the assumption that $s_2 > s_1$ and $\varepsilon < \alpha(s_2 - s_1)/(s_2 + s_1)$, we obtain the existence of $K_1(\varepsilon, s_2, s_1, \alpha)$ such that the second factor on the rhs above is at least $1/2$ for all $L > K_1(\varepsilon, s_2, s_1, \alpha) \vee K_0(\varepsilon)$. Hence for all such L,

$$\mathbb{E}\big[N(v, [L^{s_1}, L^{s_2}]) \mid D_v = L\big) \geq L^{1-\alpha s_1 - \varepsilon s_1}/2. \tag{100}$$

We now require that s_1, ε is such that $1 - \alpha s_1 - \varepsilon s_1 > 0$, then the mean tends to infinity with L. Using now Chernoff's bound on this Binomial random variable we obtain that

$$\begin{aligned}\mathbb{P}(\mathcal{A}_1(v, L) | D_v = L) &:= \mathbb{P}\big(N(v, [L^{s_1}, L^{s_2}]) > L^{1-\alpha s_1 - \varepsilon s_1}/4 \mid D_v = L\big) \\ &\geq 1 - \exp\big(-L^{1-\alpha s_1 - \varepsilon s_1}/48\big) =: 1 - \mathrm{err}_1(L).\end{aligned} \tag{101}$$

Assume now that $\mathrm{CP}^{\downarrow}_{f,\lambda}$ has reached vertex v at some time, and that $\mathcal{A}_1(v, L)$ holds for v. Let now $\mathcal{A}_2(v, L)$ be the event that v infects at least one of the first $L^{1-\alpha s_1 - \varepsilon s_1}/4$ many children within the set $\mathcal{N}(v, [L^{s_1}, L^{s_2}])$ before healing. We bound the complement of this event using that the degree of such a child is in the interval $[L^{s_1}, L^{s_2}]$, which gives that the infection rate from v to any child $u \in \mathcal{N}(v, [L^{s_1}, L^{s_2}])$ is at least $r(v, u) = \lambda \max(L, D_u)^{-\mu} \geq \lambda L^{-\mu s_2}$ (since we assumed that $s_2 > s_1 > 1$). We obtain that

$$\begin{aligned}&\mathbb{P}(\neg\mathcal{A}_2(v, L) \mid v \text{ ever infected}, D_v = L, \mathcal{A}_1(v, L)) \\ &= \frac{1}{1 + \sum\{r(v, u_i) : u_i \in \mathcal{N}(v, [L^{s_1}, L^{s_2}]), i \leq L^{1-\alpha s_1 - \varepsilon s_1}/4\}} \leq \frac{1}{1 + \lambda L^{1-\alpha s_1 - \mu s_2 - \varepsilon s_1}/4} \\ &\leq 8\lambda^{-1} L^{-(1-\alpha s_1 - \mu s_2 - \varepsilon s_1)} =: \mathrm{err}_2(L),\end{aligned} \tag{102}$$

where we used that L is sufficiently large, and the assumption that $1 - \alpha s_1 - \mu s_2 - \varepsilon s_1 > 0$ to obtain the last line. This assumption can be satisfied with $s_2 > s_1 > 1$ and $\varepsilon > 0$ small enough whenever $1 - \alpha - \mu > 0$, which is true since we assumed $\alpha + \mu < 1$. Also note that it cannot be satisfied when $\alpha + \mu \geq 1$.

We use the error bound in (102) repeatedly. Let now v_0 be the first vertex reached by $\mathrm{CP}^{\downarrow}_{f,\lambda}$ with degree at least K in the event $\mathcal{B}_K$ in (99), and let D_{v_0} denote its random degree. We now define a random infection ray $(v_0, v_1, \ldots, v_m, v_{m+1} \ldots)$ recursively. Suppose we already defined $(v_0, \ldots, v_m)$ for some $m \geq 0$, and their degrees $(D_{v_0}, \ldots, D_{v_m})$. We now check whether the event $\mathcal{A}_1(v_m, D_{v_m}) \cap \mathcal{A}_2(v_m, D_{v_m})$ holds, and if so, then we choose any vertex $v_{m+1} \in \mathcal{N}(v_m, [D_{v_m}^{s_1}, D_{v_m}^{s_2}])$ that is infected by v_m before v_m heals. We now obtain the existence of an infinite ray by taking the limit of the nested sequence of events:

$$\begin{aligned}\mathbb{P}\big((v_0, \ldots, v_m, \ldots) \text{ exists}\big) &= \lim_{m_0 \to \infty} \mathbb{P}\Big(\cap_{m \leq m_0} \{v_{m+1} \text{ exists}\}\Big) \\ &= \lim_{m_0 \to \infty} \prod_{m=0}^{m_0} \mathbb{P}\Big(v_{m+1} \text{ exists} \mid (v_0, \ldots, v_m) \text{ exists}\Big),\end{aligned}$$

We denote by $\mathcal{F}_m$ the sigma-algebra generated by

$$\cup_{i \leq m-1} \{\mathcal{A}_1(v_i, D_{v_i}), \mathcal{A}_2(v_i, D_{v_i}), v_i, D_{v_i}\} \cup \{v_m, D_{v_m}\}.$$

That is, we reveal the degree and existence of v_m, but not whether $\mathcal{A}_1(v_m, D_{v_m}) \cap \mathcal{A}_2(v_m, D_{v_m})$ holds since those events already give v_{m+1}. Using this sigma-algebra, we can use the Markov property of $\mathrm{CP}^{\downarrow}_{f,\lambda}$, lower bound the probability of existence of v_0 by $\mathbb{P}(\mathcal{B}_K)$, and that of v_{m+1} by the conditional probability of $\mathcal{A}_1(v_m, D_{v_m}) \cap \mathcal{A}_2(v_m, D_{v_m})$ to obtain

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

$$\begin{aligned}
&\mathbb{P}\big((v_0,\dots,v_m,\dots)\text{ exists}\big)\\
&\quad\geq \lim_{m_0\to\infty}\mathbb{P}(\mathcal{B}_K)\mathbb{E}\left[\prod_{m=0}^{m_0}\mathbb{P}\Big(\mathcal{A}_1(v_m,D_{v_m})\cap\mathcal{A}_2(v_m,D_{v_m})\mid\mathcal{F}_m\Big)\right]\\
&\quad\geq \mathbb{P}(\mathcal{B}_K)\lim_{m_0\to\infty}\left[\prod_{m=1}^{m_0}\mathbb{P}\Big(\mathcal{A}_1(v_m,D_{v_m})\cap\mathcal{A}_2(v_m,D_{v_m})\mid v_m\text{ ever infected},D_{v_m}\Big)\right].
\end{aligned}\tag{103}$$

Observe that now the calculations in (101) and (102) apply, and the mth factor is, conditionally on D_{v_m}, at least $1-\mathrm{err}_1(D_{v_m})-\mathrm{err}_2(D_{v_m})$. We inductively show that the mth factor in the product above is at least

$$1-\mathrm{err}_1(K^{s_1^m})-\mathrm{err}_2(K^{s_1^m}),\tag{104}$$

by showing that $D_{v_m}\geq K^{s_1^m}$ whenever v_m exists. Monotonicity of $\mathrm{err}_1(L)+\mathrm{err}_2(L)$ in L then immediately yields the lower bound (104), as follows. Since we assumed $D_{v_0}\geq K=K^{s_1^0}$, the induction starts. Assume now that $D_{v_{m-1}}\geq K^{s_1^{m-1}}$. Then per definition, (see below (102)), $D_{v_m}\in[D^{s_1}_{v_{m-1}},D^{s_2}_{v_{m-1}}]$. Using now the induction hypothesis immediately gives (104). Hence, we return to (103), for a.e. realization in the conditional expectation the lower bound in (104) holds, hence,

$$\begin{aligned}
\mathbb{P}\big((v_0,\dots,v_m,\dots)\text{ exists}\big)&\geq\mathbb{P}(\mathcal{B}_K)\prod_{i=1}^{\infty}(1-\mathrm{err}_1(K^{s_1^m})-\mathrm{err}_2(K^{s_1^m}))\\
&\geq\mathbb{P}(\mathcal{B}_K)\Big(1-\sum_{m=0}^{\infty}\mathrm{err}_1(K^{s_1^m})+\mathrm{err}_2(K^{s_1^m})\Big).
\end{aligned}\tag{105}$$

Using the values of $\mathrm{err}_1(K^{s_1^m})+\mathrm{err}_2(K^{s_1^m})$ from (101), (102), given that

$$1<s_1<s_2,\qquad 1-\alpha s_1-\mu s_2-\varepsilon s_1>0,\qquad 1-\alpha s_1-\varepsilon s_1>0,\tag{106}$$

the sum on the right hand side is summable in m, and both terms decrease faster then geometrically in m, hence they are dominated by a constant times their first term:

$$\begin{aligned}
&\sum_{m=0}^{\infty}\exp(-K^{s_1^m(1-\alpha s_1-\varepsilon s_1)}/48)+\sum_{m=0}^{\infty}8\lambda^{-1}K^{-s_1^m(1-\alpha s_1-\mu s_2-\varepsilon s_1)}\\
&\qquad\leq C\exp(-K^{1-\alpha s_1-\varepsilon s_1}/48)+C\lambda^{-1}K^{-(1-\alpha s_1-\mu s_2-\varepsilon s_1)}.
\end{aligned}\tag{107}$$

One can check that the system of inequalities in (106) is solvable whenever $1-\alpha-\mu>0$. Namely, choose first $1<s_1<s_2$ close enough to 1 so that $1-\alpha s_1-\mu s_2>0$ holds. Choose then $\varepsilon>0$ small enough so that (100) and (106) hold as well, and finally one can set K sufficiently large so that all inequalities above are valid. In particular, given now any $\lambda>0$ (i.e., small), one can choose K sufficiently large so that the sum in (107) is at most $1/2$, and then we obtain in (105) that an infinite infection ray exists with probability at least $\mathbb{P}(\mathcal{A}_K)/2$, which is strictly positive. Hence, global survival occurs with strictly positive probability, whenever $\alpha+\mu<1$, finishing the proof. □

6.2. *Product penalty: local survival using a row of star-graphs when* $\mu<1/2$

We will prove local survival of $\mathrm{CP}_{f,\lambda}$ (for both product and maximum penalty) when $\mu\in[0,1/2)$ on the Galton-Watson tree, with at last stretched exponential offspring distributions, that is, Theorem 2.1 in multiple steps.

The idea is the following: As a direct consequence of known results about star graphs that goes back to Berger, Borgs, Chayes and Saberi [5], in Claim 6.6 we prove that when $\mu<1/2$, the infection survives

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

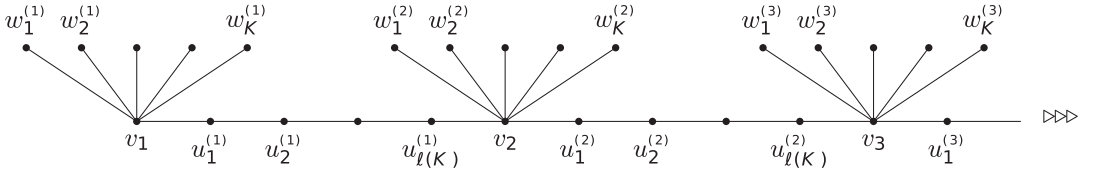


Figure 2. *The graph $H_{K,\ell(K)}$.*

on a star-graph of degree K, which consists of a degree-K vertex and its degree-1 neighbors, for a time $T_K = \exp(\Theta(\lambda^2 K^{1-2\mu}))$ with probability very close to 1. Moreover, throughout this time the star will be *infested*, by which we mean that a sufficiently high fraction of its vertices are infected.

We then show that a star-graph that is infested for time T_K, sends the infection through a path of length ℓ to another such star-graph with probability close to 1 if and only if $\ell = o(\log T_K)$. Hence we need that $\ell = o(K^{1-2\mu})$ so that the infection successfully infests another star-graph.

Let $H_{K,\ell(K)}$ be a graph that consists of a one-ended infinite row of star-graphs of degree K, $(v_1, v_2, \dots)$, with paths of length $\ell(K) = o(K^{1-2\mu})$ between two consecutive stars. We show that the degree-penalized contact process survives forever on $H_{K,\ell(K)}$ with positive probability, as long as K is sufficiently large compared to λ. We do this by mapping the process on $H_{K,\ell(K)}$ to a discrete time analog of the contact process on $\mathbb{N}_+ = \{1, 2, \dots\}$ corresponding to the infinite row of star-graphs $(v_1, v_2, \dots)$.

Finally, we prove that $H_{K,\ell(K)}$ can be embedded almost surely in a Galton-Watson tree $\mathcal{T}_D$ in a way that in the embedding, every vertex in $H_{K,\ell(K)}$ has degree at most M times its degree in $H_{K,\ell(K)}$. This only changes λ in the arguments above by a constant factor, that is, to $\tilde{\lambda} := \lambda/M^{2\mu}$, so if $\mathrm{CP}_{f,\lambda}$ survives on $H_{K,\ell(K)}$ whenever K is sufficiently large, then the same is true for $\mathrm{CP}_{f,\tilde{\lambda}}$ by increasing K if necessary. For the embedding to be possible, the tail of D must be heavier than stretched exponential with stretch-exponent $1 - 2\mu$, in the sense of Definition 1.8, which is the mildest condition possible for this proof to work.

6.2.1. Embedding stars in the Galton-Watson tree

We now make the former outline precise, starting with the definition of the infinite row of star-graphs and the embedding that does not increase degrees too much.

Definition 6.3 (Infinite path of stars and M-embedding). Given two integers $K, \ell \geq 1$, let $H = H_{K,\ell}$ be an infinite graph defined as follows: we start by taking an infinite path $(v_1, \mathcal{P}_1, v_2, \mathcal{P}_2, \dots, v_i, \mathcal{P}_i, v_{i+1}, \dots)$, where for all $i \geq 1$ the paths $\mathcal{P}_i = (u_1^{(i)}, \dots, u_\ell^{(i)})$ have length ℓ, and then to each $v_i, i \in \mathbb{N}$ we attach K additional neighbors $w_1^{(i)}, \dots, w_K^{(i)}$, each with $\deg_H(w_j^{(i)}) = 1$, which we call leaves. We call K the star-degree of $H_{K,\ell}$ and ℓ the connecting-path length, which might depend on K. See Figure 2.

We say that $H = H_{K,\ell}$ is (degree-factor) M-embedded in a graph G if G contains $H_{K,\ell}$ as subgraph, and for all vertices $v \in H_{K,\ell} \subseteq G$ it holds that

$$\frac{\deg_G(v)}{\deg_H(v)} \leq M. \tag{108}$$

The next lemma shows that for large K, $H_{K,\ell}$ can be M-embedded almost surely into a Galton-Watson tree $\mathcal{T}$ with offspring distribution D. The proof reveals that the tail of D determines the minimal $\ell = \ell(K)$ that is possible for the embedding to hold almost surely.

Lemma 6.4. *Let $\mathcal{T}$ be a Galton-Watson tree with degree distribution D so that the tail of D is heavier than stretched exponential with stretch-exponent $1 - 2\mu$, in the sense of Definition 1.8, along the infinite sequence $(z_i)_{i\geq 1}$, and prefactor $g(z) \to 0$ as $z \to 0$. Then there exists a constant $M \geq 1$, such that $H_{K,\ell(K)}$ can be M-embedded in $\mathcal{T}$ for all sufficiently large K such that $2K \in \{z_i, i \geq 1\}$, for almost all realizations of $\mathcal{T}$, whenever*

$$\ell(K) \geq 2^{1-2\mu}\sqrt{g(2K)}K^{1-2\mu} = o(K^{1-2\mu}). \tag{109}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Proof. First, fix some small constant $\varepsilon > 0$ decided later. Let $\mathbb{E}[D] := q > 1$, and define $D_M := D\mathbb{1}_{D<M}$, that is, the distribution where $\mathbb{P}(D_M = 0) = \mathbb{P}(D = 0) + \mathbb{P}(D \geq M)$ and $\mathbb{P}(D_M = k) = \mathbb{P}(D = k)$ for all $k \in [1, M)$. Given $\varepsilon > 0$, we choose $M > 2$ such that both of the following inequalities hold:

$$\begin{aligned} q_M := \mathbb{E}[D_M] = \mathbb{E}[D\mathbb{1}\{D < M\}] \geq \mathbb{E}[D] - \varepsilon > 1 + 2\varepsilon, \\ \mathbb{P}(D < M) \geq 1 - \varepsilon. \end{aligned} \tag{110}$$

It is clear that D_M can be coupled to D such that $\mathbb{P}(D_M \leq D) = 1$, and this embedding can be done for each vertex of the original Galton Watson tree $\mathcal{T}$, obtaining a sub-forest $\mathcal{F}_M$ of $\mathcal{T}$. The embedding can be done by first sampling $D_v \sim D$ many children for each vertex v, and then accepting the number of offspring as it is when D_v is between 0 and $M - 1$, but setting the degree of v in $\mathcal{F}_M$ to be 0 when $D_v \geq M$. We will denote the distribution of a single tree in $\mathcal{F}_M$ by $\mathcal{T}_M$, which is a branching process with offspring distribution D_M.

Define the event, for $2K \in \{z_i\}_{i\geq 1}$,

$$\mathcal{A}_1 := \{\exists v \in \mathcal{T} : D_v = 2K\}.$$

Since we assumed $\mathbb{P}(D = 0) = 0$, $\mathcal{T}$ survives almost surely and so $\mathbb{P}(\mathcal{A}_1) = 1$. Take then the vertex $v \in \mathcal{T}$ that is closest to the root $\varnothing$ and has $D_v = 2K$, and set it to v_1 in $H_{K,\ell}$ of the embedding. Clearly v_1 then satisfies (108) since its degree in $\mathcal{T}$ is $2K \leq MK$ by our assumption that $M \geq 2$.

Similarly as in the proof of Proposition 6.2 below (99), let $\mathcal{N}(v, [a, b]), N(v, [a, b])$ denote the set and number of children of a vertex $v \in \mathcal{T}$ with offspring in the interval $[a, b]$. Consider now the event $\mathcal{A}_{\text{child}}(v_1) := \{N(v_1, [0, M)) \geq K + 1\}$. Since $D_{v_1} = 2K$ per assumption, and the children of v_1 have iid degrees, using (110), each of these children has offspring less than M with probability at least $1 - \varepsilon$. Hence, using the concentration of Binomial random variables (e.g., a Chernoff's bound), whenever $\varepsilon < 1/8$ (which we safely assume), for all K sufficiently large,

$$\begin{aligned} \mathbb{P}\big(\mathcal{A}_{\text{child}}(v_1)\big) &= \mathbb{P}\big(N(v_1, [0, M)) \geq K + 1\big) \\ &\geq \mathbb{P}(\,\text{Bin}(2K, 1 - \varepsilon) > K) \geq 1 - \mathrm{e}^{-K/12}. \end{aligned} \tag{111}$$

On the event $\mathcal{A}_{\text{child}}(v_1)$, we label by $w_1, w_2, \ldots, w_{K+1}$ the first $K + 1$ children in $\mathcal{N}(v_1, [0, M))$. Including the edge towards v_1, the total degree of any of these vertices in $\mathcal{T}$ is at most M, satisfying thus the degree factor M in (108). So, v_1 and any K out of the children $w_1, \ldots, w_{K+1}$ may serve as the embedding of $w_1^{(1)}, \ldots, w_K^{(1)}$ of $H_{K,\ell}$, and any one of these children may take the role of $u_1^{(1)}$ of the path $\mathcal{P}_1$ in $H_{K,\ell}$.

From each of these vertices w_i we start the (embedded) branching process $\mathcal{T}_M(w_i) \subseteq \mathcal{T}(w_i)$ with offspring distribution D_M. Let the number of descendants of w_i in $\mathcal{T}_M(w_i)$ in generation ℓ (that is, of distance ℓ from w_i) be $Z_\ell^{(i)}$ for each $\ell \geq 1$. It is well-known that $W_\ell^{(i)} := Z_\ell^{(i)}/q_M^\ell$ is a martingale for each i [3], and that $\lim_{\ell\to\infty} W_\ell^{(i)} = W_\infty^{(i)}$ exists a.s. Since $\mathbb{E}[D_M] = q_M > 1 + 2\varepsilon$, this branching process is supercritical, and because D_M is bounded by M, the Kesten-Stigum Theorem gives that $\eta := \mathbb{P}(W_\infty^{(i)} \neq 0) > 0$ is the probability that the corresponding branching process $\mathcal{T}_M$ survives indefinitely. It follows then that, for any i,

$$\lim_{\ell\to\infty} \mathbb{P}(Z_\ell^{(i)} \geq (q_M - \varepsilon)^\ell) = \lim_{\ell\to\infty} \mathbb{P}\left(\frac{Z_\ell^{(i)}}{q_M^\ell} \geq \left(\frac{q_M - \varepsilon}{q_M}\right)^\ell\right) = \mathbb{P}(W_\infty^{(i)} > 0) = \eta.$$

By (110), $q_M - \varepsilon > 1 + \varepsilon$ and consequently, there exists a (deterministic) ℓ_0 only depending on D_M (but not on K) such that for all $\ell > \ell_0$ we have

$$\mathbb{P}(\mathcal{B}_i) := \mathbb{P}(Z_\ell^{(i)} \geq (q_M - \varepsilon)^\ell) \geq \eta/2. \tag{112}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Denote the set of individuals in the ℓ-th generation of $\mathcal{T}_M(w_i)$ by $\mathcal{G}_\ell^{(i)}$ for each $i = 1, 2, \dots, K$, and let $\mathcal{G}_\ell = \cup_{i=1}^K \mathcal{G}_\ell^{(i)}$. Since $(w_i)_{i\le K}$ are siblings, $\mathcal{G}_\ell$ is embedded in $\mathcal{T}$ also in the same (possibly other than ℓ) generation. We now return to the original branching process $\mathcal{T}$ for a single generation. For each $v \in \mathcal{G}_\ell$ consider i.i.d. copies D_v of D (that is, without the truncation at M used so far), and define the events for $i \le K$:

$$\widetilde{\mathcal{B}}_i := \{\exists u_{(i)} \in \mathcal{G}_\ell^{(i)} : \ D_{u_{(i)}} = 2K\} \tag{113}$$

for each $i = 1, \dots, K+1$. By (112) we have $\mathbb{P}(\mathcal{B}_i) \ge \eta/2$. Furthermore, since on the event $\mathcal{B}_i$

$$\begin{aligned}\mathbb{P}(\neg\widetilde{\mathcal{B}}_i \mid \mathcal{B}_i) &\le \big(1 - \mathbb{P}(D = 2K)\big)^{(q_M-\varepsilon)^\ell} \\ &\le \exp\Big(- \mathbb{P}(D = 2K)(q_M - \varepsilon)^\ell\Big).\end{aligned}$$

Since we have assumed $2K \in \{z_i\}_{i\ge1}$ in Definition 1.8, we can use the bound

$$\mathbb{P}(D = 2K) \ge \exp(-g(2K)(2K)^{1-2\mu})$$

for the function $g(2K) \to 0$ as $K \to 0$ in Definition 1.8. Hence, $g(2K) = o(\sqrt{g(2K)})$ but at the same time $\sqrt{g(2K)} \to 0$ as $K \to \infty$. We then also use that $q_M - \varepsilon > 1 + \varepsilon$ by assumption, and so by choosing $\ell = \ell(K) \ge \sqrt{g(2K)}(2K)^{1-2\mu}$, one can compute that $(2K)^{1-2\mu}(\sqrt{g(2K)}\log(q_M - \varepsilon) - g(2K)) \to \infty$ and so for all sufficiently large K it holds that

$$\begin{aligned}\mathbb{P}(\neg\widetilde{\mathcal{B}}_i \mid \mathcal{B}_i) &\le \exp\Big(- \mathrm{e}^{-g(2K)(2K)^{1-2\mu}}(q_M - \varepsilon)^{\ell(K)}\Big) \\ &\le \exp\big(- \mathrm{e}^{(2K)^{1-2\mu}(\sqrt{g(2K)}\log(q_M-\varepsilon)-g(2K))}\big) \le 1/2.\end{aligned} \tag{114}$$

Combining (112) and (114) yields

$$\mathbb{P}(\widetilde{\mathcal{B}}_i) \ge \mathbb{P}(\mathcal{B}_i) \cdot \mathbb{P}(\widetilde{\mathcal{B}}_i \mid \mathcal{B}_i) \ge (\eta/2) \cdot (1/2) \ge \eta/4.$$

Now we define the event that at least two events $\widetilde{\mathcal{B}}_i, \widetilde{\mathcal{B}}_j$ happen for v_1:

$$\widetilde{\mathcal{A}}(v_1) := \{\exists i, j : \ i \ne j : \widetilde{\mathcal{B}}_i \cap \widetilde{\mathcal{B}}_j \text{ holds}\}. \tag{115}$$

Now consider the number of indices $i \le K+1$ for which $\widetilde{\mathcal{B}}_i$ holds. By (113), on the event $\mathcal{A}_{\text{child}}(v_1)$ in (111), this number stochastically dominates a binomial random variable with parameters $K+1$ and $\eta/4$. Hence, by the definition of $\widetilde{\mathcal{A}}(v_1)$ in (115), it holds for some constant $c(\eta) > 0$ that

$$\begin{aligned}\mathbb{P}(\widetilde{\mathcal{A}}(v_1) \mid \mathcal{A}_{\text{child}}(v_1)) &\ge \mathbb{P}(\mathrm{Bin}(K+1, \eta/4) \ge 2) \\ &= 1 - (1-\eta/4)^{K+1} - K(\eta/4)(1-\eta/4)^K \ge 1 - \mathrm{e}^{-c(\eta)K}.\end{aligned}$$

Combining this with (111), we obtain that for all sufficiently large K,

$$\mathbb{P}(\mathcal{A}_{\text{child}}(v_1) \cap \widetilde{\mathcal{A}}(v_1)) \ge 1 - \mathrm{e}^{-c(\eta)K} - \mathrm{e}^{-K/12} \ge 1 - \varepsilon. \tag{116}$$

On the event $\tilde{\mathcal{A}}(v_1) \cap \mathcal{A}_{\text{child}}(v_1)$, there are two vertices $v_{2,1}, v_{2,2}$ such that their most recent common ancestor is the starting vertex v_1, and $\deg(v_{2,1}), \deg(v_{2,2}) = 2K$, and $d_G(v_1, v_{2,1}) = d_G(v_1, v_{2,2}) = \ell(K) \ge \sqrt{g(2K)}(2K)^{1-2\mu}$ with $\ell(K) = o(K^{1-2\mu})$, and the paths $\mathcal{P}_{1,1}, \mathcal{P}_{1,2}$ joining v with $v_{2,1}$ and $v_{2,2}$ respectively are edge-disjoint with all internal vertices having degree at most M. Observe that $(v_1, \mathcal{P}_{1,1}, v_{2,1})$ and $(v_1, \mathcal{P}_{1,2}, v_{2,2})$ both serve as a factor M-embedding of the vertices in $(v_1, \mathcal{P}_1, v_2)$ in $H_{K,\ell(K)}$, hence we may choose any of them for the embedding. Further, the vertices $v_{2,1}$ and $v_{2,2}$ have degree $2K$ in $\mathcal{T}$, hence, using the argument between (111) and (115), one can repetitively apply the

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

procedure of checking whether the events $\mathcal{A}_{\text{child}}(\cdot) \cap \tilde{\mathcal{A}}(\cdot)$ hold for these vertices, and the vertices then found by either $\mathcal{A}_{\text{child}}(v_{2,1}) \cap \tilde{\mathcal{A}}(v_{2,1})$ or $\mathcal{A}_{\text{child}}(v_{2,2}) \cap \tilde{\mathcal{A}}(v_{2,2})$ may all serve as the embedding of the path $\mathcal{P}_2$ and v_3, and so on.

We thus consider an auxiliary "renormalised" branching process. We say that v_1 has 2 children (in this case $v_{2,1}, v_{2,2}$) with probability (at least) $1-\varepsilon$ in (116) and 0 otherwise. Observe that the path leading to any vertex in generation j of this branching process serves as an M-embedding of $(v_1, \mathcal{P}_1, v_2, \dots, \mathcal{P}_{j-1}, v_j)$. This renormalised branching process is supercritical. Hence, it survives with positive probability, giving that the M-embedding of the infinite graph $H_{K,\ell(K)}$ exists in $\mathcal{T}$, starting from v_1, with positive probability. Kolmogorov's 0-1 law finishes the proof that $\mathcal{T}$ then has a proper M-embedding of $H_{K,\ell(K)}$ somewhere in $\mathcal{T}$ with probability 1. □

We now define star-graphs (subgraphs of $H_{K,\ell}$) and the notion of infested stars.

Definition 6.5. A *star-graph* S of degree K is a graph which consists of one vertex v of degree $\deg_S(v) = K$ (its center) and its K neighbors $(w_i)_{i\le K}$, each of degree $\deg_S(w_i) = 1$ that we call leaves. Consider the classical contact process with infection rate r on S. We will call such a star *r-infested* at some time t by the contact process if at least $rK/(16e^2)$ of its leaves are infected.

The next claim adapts [53, Lemma 3.1] to the degree-penalized contact process on S. The claim shows that starting with only the center infected, a star-graph is $\lambda K^{-\mu}$*-infested* for a time interval of length $T_K \ge \exp(cr^2K) = \exp(c\lambda^2K^{1-2\mu})$ with high probability, and during this time-interval the center vertex v is infected more than half of the time. Writing r for the rate of infection of the classical contact process on a star-graph, [53, Lemma 3.1] holds under the condition that r^2K is uniformly bounded away from 0. Since in the degree-penalized CP, the rate across the edges of the star-graph is $r = \lambda K^{-\mu}$, we shall require that $\lambda^2K^{1-2\mu}$ is uniformly bounded away from 0.

Claim 6.6 (Lemma 3.1 of [53] adapted). Assume $\mu < 1/2$, $\lambda < 1$. Consider a star-graph S of degree K with center v. Let ξ_t denote the contact process CP on S where $r(v,u) = r(u,v) = \lambda/K^{\mu}$. Then there exists a constant $c_1 > 0$ such that

$$\mathbb{P}\Big(|\underline{\xi}_1| \ge \lambda K^{1-\mu}/(4e) \mid \xi_0(v) = 1\Big) \ge (1 - e^{-c_1\lambda K^{1-\mu}})/e. \tag{117}$$

Further, let $T_K := \exp(c_1\lambda^2K^{1-2\mu})$. If $\lambda^2K^{1-2\mu} > 32e^2$, then

$$\mathbb{P}\Big(\underline{\xi}_{T_K} \ne \underline{0} \,\Big|\, |\underline{\xi}_0| \ge \lambda K^{1-\mu}/(8e)\Big) \ge 1 - e^{-c_1\lambda^2K^{1-2\mu}} =: 1 - \mathrm{err}_{\lambda,K}. \tag{118}$$

Moreover,

$$\mathbb{P}\left(S \text{ is } \lambda K^{-\mu}\text{-infested for all } t \in [0,T_K] \text{ and } \int_0^{T_K} \xi_t(v) \ge T_K/2 \,\middle|\, |\underline{\xi}_0| \ge \lambda K^{1-\mu}/(8e)\right)$$
$$\ge 1 - \mathrm{err}_{\lambda,K}. \tag{119}$$

The proof of Claim 6.6 is very similar to [53, Lemma 3.1], therefore we include it in the Appendix.

6.2.2. Contact process on an infinite line of stars

We continue by studying the spread of the infection on $H_{K,\ell(K)}$. In particular, we prove that the probability that an infested star passes on the infestation to a neighboring star in $H_{K,\ell(K)}$ can be made arbitrarily close to 1 with the right choice of the parameters.

Claim 6.7. For each fixed small $\lambda > 0$ and $\delta > 0$ there is a $K_{\lambda,\delta}$ such that the following holds for all $K \ge K_{\lambda,\delta}$. Consider the degree-penalized contact process $\mathrm{CP}_{f,\lambda}$ on $H_{K,\ell(K)}$ with $f = (xy)^{\mu}$ for some $\mu < 1/2$. Consider two consecutive stars v_i, v_{i+1} in $H_{K,\ell(K)}$ in Definition 6.3, with $\ell(K) = o(K^{1-2\mu})$, and let $T_K := \exp(c_1\lambda^2K^{1-2\mu})$ from Claim 6.6. Suppose that v_i is $\lambda K^{-\mu}$-infested at some time t_0. Then at time $t_0 + T_K$, v_{i+1} is $\lambda K^{-\mu}$-infested with probability at least $1-\delta$.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Proof. In Definition 6.3, we denoted the vertices on the path $\mathcal{P}_i$ connecting v_i to v_{i+1} by $u_1^{(i)}, u_2^{(i)}, \dots u_\ell^{(i)}$. In this proof we will omit the superscript and denote absolute constant factors by c that we specify on the go. Further, $|\cdot|$ means the Lebesgue measure of a set in $\mathbb{R}$. We define the event and bound its probability from below using (119):

$$\mathbb{P}(\mathcal{A}_1(v_i)) := \{v_i \text{ is } \lambda K^{-\mu}\text{-infested for all } t \in [t_0, t_0 + T_K]\} \geq 1 - \mathrm{e}^{c_1 \lambda^2 K^{1-2\mu}} \geq 1 - \delta/8, \tag{120}$$

whenever $K \geq \log(8/\delta)\lambda^{-2/(1-2\mu)}/c_1 =: K_0(\delta)$. For some t_K to be determined later, partition the time interval $[t_0, t_0 + T_K]$ into $m_K = \lfloor T_K/t_K \rfloor$ disjoint consecutive intervals of length t_K followed by one potentially shorter time interval, denoted by $J_1, \dots J_{m_K}$ and J_{m_K+1} (for the remaining time of length $T_K - m_K t_K \leq t_K$). We would like to use the infested status of the star around v_i to help transmit the infection along the path $v_i, u_1^{(i)}, \dots, u_\ell^{(i)}$. For this we start with establishing that v_i itself is in an infected state shortly after the beginning of the time interval J_j. While the probability of such an event is not explicitly mentioned in Claim 6.6, it can be obtained from its proof. Namely, since v_i is infested at time t_0 by assumption, it is also infested for all times $t \in [t_0, t_0 + T_K]$, and the proof of Claim 6.6 reveals that $\xi_t(v_i)$ stochastically dominates a two-state Markov chain (say η_t) on $\{0, 1\}$ with transition rate $q_{0,1} = \lambda^2 K^{1-2\mu}/(16\mathrm{e}^2)$ and $q_{1,0} = 1$. So, regardless of the value of $\xi_{J_j^-}(v_i)$, the value of $\xi_{J_j^-+1}(v_i)$ stochastically dominates the value of $\eta_1 | \{\eta_0 = 0\}$, which equals 1 with probability more than $1/2$. Formally, for each interval $J_j = [J_j^-, J_j^+)$ with $j \leq m_K$, let τ_j denote the first time in J_j when $\xi_t(v_i) = 1$. Define then the event that

$$\mathcal{A}_2(J_j) := \{\tau_j \leq J_j^- + 1\}. \tag{121}$$

Then by the above argument, $\mathbb{P}(\mathcal{A}_2(J_j)) \geq 1/2$ for all J_j, and the Markov property of the process ensures that $\mathcal{A}_2(J_j)_{j \leq m_k}$ are independent. Then Chernoff's bound yields that

$$\begin{aligned} \mathbb{P}(\mathcal{A}_3(v_i)) &:= \mathbb{P}\big(\#\{j \leq m_K : \mathcal{A}_2(J_j) \text{ holds}\} \geq m_K/4\big) \\ &\geq \mathbb{P}\big(\mathrm{Bin}(m_K, 1/2) \geq m_K/4\big) \geq 1 - \mathrm{e}^{-c m_k}, \end{aligned} \tag{122}$$

for $c = 1/48$. Consider now $\{j : \mathcal{A}_2(J_j) \text{ holds}\}$, and for each such j, call such J_j *successful* if there is some time $t \in J_j$ when at least $\lambda K^{1-\mu}/(4e)$ many leaves in the star-graph of v_{i+1} are infected. We now lower bound the probability of the event that J_j is successful conditioned on $\mathcal{A}_2(J_j)$, as follows. Define a sequence of time-moments $s_h := \tau_j + h4^\mu$ for $h \in \{0, \dots, \ell + 1\}$, and for $h = 1, \dots, \ell$ we recursively check whether u_h is infected at time s_h, given that u_{h-1} is infected at s_{h-1} (setting $u_0 := v_i$), and that whether $v_{i+1} =: u_{\ell+1}$ is infected at time $s_{\ell+1}$ given that u_ℓ is infected at time s_ℓ. We also set $s_{\ell+2} := s_{\ell+1} + 1$ and check whether at least $\lambda K^{1-\mu}/(4e)$ many leaves in the star of v_{i+1} are infected at time $s_{\ell+2}$, given that v_{i+1} is infected at time $s_{\ell+1}$. We shall thus bound, for some constant c, the time-interval lengths and their number as

$$t_K := 4^\mu(\ell + 2) + 2 \leq (4^\mu \vee 2)(\ell + 3), \qquad m_K = \lfloor T_K/t_K \rfloor \geq c\, T_K/\ell. \tag{123}$$

Returning to an interval J_j being successful, denote the infection status of the set of leaves in the star around v_{i+1} by $\underline{\xi}_t^{(i+1)}$. Then, using the strong Markov property, we can lower bound

$$\mathbb{P}(J_j \text{ successful} \mid \mathcal{A}_2(J_j)) \geq \mathbb{P}\big(|\underline{\xi}_{s_{\ell+2}}^{(i+1)}| \geq \lambda K^{1-\mu}/(4e) \mid \xi_{\tau_j}(v_i) = 1\big) \tag{124}$$

$$\geq \mathbb{P}(\xi_{s_1}(u_1) = 1 \mid \xi_{\tau_j}(v_i) = 1) \prod_{h=2}^{\ell+1} \mathbb{P}\Big(\xi_{s_h}(u_h) = 1 \mid \xi_{s_{h-1}}(u_{h-1}) = 1\Big) \tag{125}$$

$$\cdot\, \mathbb{P}(|\underline{\xi}_{s_{\ell+2}}^{(i+1)}| \geq \lambda K^{1-\mu}/(4e) \mid \xi_{s_{\ell+1}}(v_{i+1}) = 1). \tag{126}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

On the last factor we shall use Claim 6.6 shortly, but first we bound the probability of each other factor in (125) from below by requiring that the sender vertex u_{h-1} infects u_h during a time interval of length 4^μ from below, and then u_h stays infected for the rest of the time-interval. More generally, along an edge (u, v), for any two time-moments $t < t'$, with infection rate r along the edge,

$$\mathbb{P}(\xi_{t'}(v) = 1 \mid \xi_t(u) = 1) \geq \int_{\tau=0}^{t'-t} (\mathrm{e}^{-\tau})(r\mathrm{e}^{-r\tau})\mathrm{e}^{-((t'-t)-\tau)}\mathrm{d}\tau = \mathrm{e}^{-(t'-t)}\Big(1 - \mathrm{e}^{-r(t'-t)}\Big).$$

On the path $(u_0, u_1, u_2, \ldots, u_\ell, u_{\ell+1})$ (with $u_0 := v_i$, $u_{\ell+1} := v_{i+1}$), we apply this lower bound with $t' - t = 4^\mu$ along each edge, with rates $r(u_{h-1}, u_h) = \lambda/4^\mu$ for all $h \in \{2, \ldots, \ell\}$, and $r(u_0, u_1) = r(u_\ell, u_{\ell+1}) = \lambda/(2K)^\mu$. For (126), we recall that $s_{\ell+2} - s_{\ell+1} = 1$, so here (117) directly applies, hence

$$\begin{aligned}\mathbb{P}(J_j \text{ is successful} \mid \mathcal{A}_2(J_j)) \geq \mathrm{e}^{-1}(1 - \mathrm{e}^{-c_1\lambda K^{1-\mu}}) \\ \cdot \Big(\mathrm{e}^{-4^\mu}\big(1 - \mathrm{e}^{-4^\mu\lambda/(2K)^\mu}\big)\Big)^2 \prod_{h=1}^{\ell} \mathrm{e}^{-4^\mu}\Big(1 - \mathrm{e}^{-4^\mu\cdot\lambda/4^\mu}\Big).\end{aligned}$$

Then we may apply that $1 - \mathrm{e}^{-x} \geq x/2$ for all $x < 1/2$ to arrive at

$$\begin{aligned}\mathbb{P}(J_j \text{ is successful} \mid \mathcal{A}_2(J_j)) &\geq (1 - \mathrm{e}^{-c_1\lambda K^{1-\mu}})e^{-4^\mu(\ell+2)-1}(\lambda/2)^\ell(2K)^{-2\mu} \\ &\geq c(c_2\lambda)^\ell K^{-2\mu} =: q_K,\end{aligned} \tag{127}$$

for some constant $c > 0$ and $c_2 := \mathrm{e}^{-4^\mu}/2$, as long as $(1 - \mathrm{e}^{-c_1\lambda K^{1-\mu}}) \geq 1/2$ which is ensured since we already assumed $K \geq K_0(\delta)$ at (120). Since the time-intervals are disjoint, on $\mathcal{A}_3(v_i)$ from (122), by the strong Markov property, the indicators of the events $\{J_j \text{ successful}\}$ stochastically dominate $m_K/4$ independent trials (with m_K from (123)), each with success probability q_K from (127). Let $\mathcal{A}_4(v_i)$ be the event that at least one of the intervals is successful. Then

$$\mathbb{P}(\mathcal{A}_4(v_i) \mid \mathcal{A}_1(v_i) \cap \mathcal{A}_3(v_i)) \geq 1 - (1 - q_K)^{m_K/4} \geq 1 - \mathrm{e}^{-m_K q_K/4}, \tag{128}$$

where we used that $1 - x \geq \mathrm{e}^{-x/2}$ for all $x < 1/4$, which is applicable since q_K in (127) tends to 0 with K. We now analyze the exponent $m_K q_K$ as a function of K on the rhs of (128).The assumption in this claim is that $\ell(K) = o(K^{1-2\mu})$ (in contrast to (109) which is more specific). So, we may assume wlog that $\ell(K)$ can be written in the form

$$\ell(K) := \widetilde{g}(K)K^{1-2\mu} \qquad \text{for} \qquad \widetilde{g}(K) \to 0 \text{ as } K \to \infty. \tag{129}$$

Recalling from (119) that $T_K = \exp(c_1\lambda^2 K^{1-2\mu})$, and $m_K \geq cT_K/\ell(K)$ from (123), as well as (127), we obtain using the fact that $1/\ell(K) \geq K^{-(1-2\mu)}$:

$$\begin{aligned}m_K q_K \geq c(T_K/\ell)\cdot(c_2\lambda)^\ell K^{-2\mu} &= c\exp\Big(c_1\lambda^2 K^{1-2\mu} + \widetilde{g}(K)K^{1-2\mu}\log(c_2\lambda)\Big)K^{-1} \\ &= c\exp\Big(\lambda^2 K^{1-2\mu}(c_1 - \widetilde{g}(K)|\log(c_2\lambda)|/\lambda^2) - \log(K)\Big).\end{aligned}$$

We now argue that for any small fixed $\lambda > 0$ we can choose K sufficiently large so that the rhs tends to infinity. First choose $K(g, \lambda)$ so large that for all $K \geq K(g, \lambda)$ the inequality

$$\widetilde{g}(K)|\log(c_2\lambda)|/\lambda^2 < c_1/2$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

holds. This is doable since $\widetilde{g}(K) \to 0$. For all $K > K(g, \lambda)$ we thus have

$$m_K q_K \geq C \exp\Big(\lambda^2 K^{1-2\mu} c_1/2 - \log(K)\Big).$$

We now further increase $K(g, \lambda)$ if necessary so that $m_K q_K \geq \log(8/\delta)/c$. (We comment that by wlog assuming a monotonically decreasing $\widetilde{g}$, the minimal $K(g, \lambda)$ can be chosen as a constant multiple of

$$\widetilde{g}^{(-1)}\big(c_1\lambda^2/(2|\log(c_2\lambda)|)\big) \vee \lambda^{-(2+\varepsilon)/(1-2\mu)}$$

for some $\varepsilon > 0$, for all λ sufficiently small.) Returning to (128), and using (120), (122) we see that

$$\mathbb{P}(\mathcal{A}_4(v_i) \mid \mathcal{A}_3(v_i) \cap \mathcal{A}_1(v_i)) \geq 1 - \delta/8 \quad \text{and} \quad \mathbb{P}(\mathcal{A}_4(v_i) \cap \mathcal{A}_3(v_i) \cap \mathcal{A}_1(v_i)) \geq 1 - \delta/2.$$

On the event $\mathcal{A}_4(v_i) \cap \mathcal{A}_3(v_i) \cap \mathcal{A}_1(v_i)$, at least one J_j is successful, and by (124), that means that at least $\lambda K^{1-2\mu}/(4e)$ leaves in the star of v_{i+1} are infected at some time in the interval $[t_0, t_0 + T_K]$. Using the strong Markov property, and applying now (119), v_{i+1} stays $\lambda K^{-\mu}$ infested during the rest of the time interval $[t_0, t_0 + T_K]$ with probability $1 - \exp(-c_1\lambda^2 K^{1-2\mu}) \geq 1 - \delta/4$ by our initial assumption that $K \geq K_0(\delta)$. This finishes the proof. □

6.2.3. Local survival through renormalization

Having established Lemma 6.4 and Claim 6.7 we are in a position to prove Theorems 2.1 and 2.5(a) by showing that the embedded structure $H_{K,\ell(K)}$ sustains the infection (locally) indefinitely with positive probability. This is formalized below in Lemma 6.8. The proof of this has two steps. The first step is a time-renormalization. Based on the results of Claim 6.7, we prove that on $H_{K,\ell(K)}$ the infection moves between neighboring centers with large enough probability on a specified discrete time-scale, leading to a renormalized version of the contact process on $\mathbb{N}$. The second step is to establish a relationship between this renormalized contact process and a certain oriented percolation model, which then can be analyzed by techniques from percolation theory, involving a Peierls-type argument. This connection was already used in [24] to derive various results for the contact process on $\mathbb{Z}$.

Lemma 6.8. *For any fixed $\mu < 1/2$ and $\lambda > 0$, there is a $K_0(\lambda)$ such that the following holds for all $K > K_0(\lambda)$. Let $H = H_{K,\ell(K)}$ be the graph defined in Definition 6.3 with $\ell(K) = o(K^{1-2\mu})$ and with v_1 being the center of its first star. Consider the penalty function $f(x, y) = (xy)^\mu$. Then both the contact process $\mathrm{CP}_{f,\lambda}(H, \mathbb{1}_{v_1})$ and $\mathrm{BRW}_{f,\lambda}(H, \mathbb{1}_{v_1})$ exhibit local survival with positive probability.*

Proof. By the stochastic domination between $\mathrm{CP}_{f,\lambda}(H, \mathbb{1}_{v_1})$ and $\mathrm{BRW}_{f,\lambda}(H, \mathbb{1}_{v_1})$ in Lemma 3.8, it is enough to prove the statement for $\mathrm{CP}_{f,\lambda}(H, \mathbb{1}_{v_1})$. For fixed $\lambda > 0$, we choose a small $\delta > 0$ specified later. Then we choose K large enough such that $K \geq K_{\lambda,\delta}$ as in Claim 6.7. Finally, let $T_K = \exp(c_1\lambda^2 K^{1-2\mu})$ as in Claim 6.6. Then, Claim 6.7 yields the following: for any v_i in $H_{K,\ell(K)}$, if v_i is $\lambda K^{-\mu}$-infested at some time t_0, then v_{i+1} is $\lambda K^{-\mu}$-infested by v_i at time $t_0 + T_K$ with probability at least $1 - \delta$, and the same holds for v_{i-1} when $i \geq 2$. (However, these two events are not necessarily independent.) Throughout this proof, the term "infested" will refer to "$\lambda K^{-\mu}$-infested."

Now we construct an oriented percolation model, which we couple with $\mathrm{CP}_{f,\lambda}(H, \mathbb{1}_{v_1})$ so that it dominates from below $\mathrm{CP}_{f,\lambda}(H, \mathbb{1}_{v_1})$ restricted to the vertices $\{v_1, v_2, \ldots\}$ at times $\{T_K, 2T_K, \ldots\}$. Let $\mathcal{H}$ be an oriented graph on the vertex set

$$V_{\mathcal{H}} = \{(x, y) \in \mathbb{Z}^+ \times \mathbb{Z}^+ : x + y \text{ even}\}$$

with the oriented (equivalently, directed) edge set

$$E_{\mathcal{H}} = \{((x_1, y_1), (x_2, y_1 + 1)) \in V_{\mathcal{H}} \times V_{\mathcal{H}} : |x_2 - x_1| = 1\}. \tag{130}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Observe that $\mathcal{H}$ is isomorphic to a subgraph (a cone) of $\mathbb{Z}^+ \times \mathbb{Z}^+$ as a graph but the edges are "diagonal" and have Euclidean length $\sqrt{2}$. In $\mathcal{H}$, we will refer to the vertex sets $\{(x,1)\}_{x\in\mathbb{Z}^+}$ and $\{(1,y)\}_{y\in\mathbb{Z}^+}$ as the x- and y-axis, respectively. For every oriented edge $e = ((x_1,y_1),(x_2,y_2))$ – where $y_2 = y_1 + 1$ and $x_2 = x_1 \pm 1$ by (130) – define the event $\mathcal{A}_e = \mathcal{A}_{(x_1,y_1),(x_2,y_2)}$ that either v_{x_1} is not infested at time $y_1 T_K$, or v_{x_1} is infested at time $y_1 T_K$ and it infests v_{x_2} by time $y_2 T_K$ in the sense of Claim 6.7. The same claim shows that

$$\mathbb{P}(\mathcal{A}_e) \geq 1 - \delta \quad \text{for every } e \in E_{\mathcal{H}}. \tag{131}$$

Now let $\eta : V_{\mathcal{H}} \to \{0,1\}$ be a function on the vertices of $\mathcal{H}$ defined recursively as

$$\begin{aligned} \eta((x,1)) &= \mathbb{1}\{x = 1\}, \\ \eta((x,y+1)) &= \begin{cases} 1 & \text{if } \eta(x-1,y) = 1 \text{ and } \mathcal{A}_{(x-1,y),(x,y+1)} \text{ holds, or} \\ & \quad \eta(x+1,y) = 1 \text{ and } \mathcal{A}_{(x+1,y),(x,y+1)} \text{ holds,} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{132}$$

Define the event

$$\mathcal{I}_1 = \{v_1 \text{ is infested at time } T_K \text{ in } \mathrm{CP}_{f,\lambda}(H, \mathbb{1}_{v_1})\}, \tag{133}$$

which exactly corresponds to $\eta((1,1)) = 1$. Then, conditioned on $\mathcal{I}_1$, $\eta(x,y) = 1$ exactly when there is an "infestation" path π of length $\mathrm{I}(\pi) = y$ through stars $(v_{\pi_1} = v_1, v_{\pi_2}, \ldots, v_{\pi_y} = v_x)$ so that v_{π_j} is infested by $v_{\pi_{j-1}}$ at time jT_K. So, on $\mathcal{I}_1$,

$$\big(\eta(x,y)\big)_{(x,y)\in V_{\mathcal{H}}} \overset{d}{\leq} \big(\xi_{yT_K}(v_x)\big)_{(x,y)\in V_{\mathcal{H}}}. \tag{134}$$

We now define a subgraph of $\mathcal{H}$. Let us declare each edge $e \in E_{\mathcal{H}}$ open if and only if $\mathbb{1}\{\mathcal{A}_e\} = 1$, closed otherwise, and denote the graph of open edges by $G(\mathcal{H})$. This is a percolation model, where the outgoing edges from a vertex (x,y) are dependent, however, the outgoing edges from distinct vertices are independent due to the strong Markov property and Claim 6.7. The open connected component of $(1,1)$ is

$$\mathcal{C}_{(1,1)} = \{(x,y) \in V_{\mathcal{H}} : \text{ there is an oriented path of open edges from } (1,1) \text{ to } (x,y)\}. \tag{135}$$

Then, comparing $\mathcal{C}_{(1,1)}$ to $\{(x,y) : \eta(x,y)\} = 1\}$ in (132), which is defined recursively as precisely those vertices that are accessible from $(1,1)$ via an oriented path of open edges in $\mathcal{H}$, we obtain that $\{(x,y) : \eta(x,y)\} = 1\} = \mathcal{C}_{(1,1)}$.

Now we carry out a Peierls-type argument to prove local survival of $\mathrm{CP}_{f,\lambda}$. Due to the coupling and stochastic domination in (134), and (135), it is enough to show that with positive probability $\mathcal{C}_{(1,1)}$ contains infinitely many vertices of the form $(1,y)$. This implies for $\mathrm{CP}_{f,\lambda}$ that v_1 is infested at times yT_K, for infinitely many y, which guarantees local survival. Let

$$Y_{\max} = \sup\{y \in \mathbb{Z}^+ : (1,y) \in \mathcal{C}_{(1,1)}\}. \tag{136}$$

We will prove that for small enough $\delta > 0$ in (131) it holds that $\mathbb{P}(Y_{\max} = \infty) > 3/4$.

Assume to the contrary that $\{Y_{\max} = k\}$ for some $k < \infty$. We now construct a path of length k, which starts from the y-axis next to $(1,k)$, and forms a part of the boundary of $\mathcal{C}(1,1)$ containing enough closed edges in $\mathcal{H}$. Define for each edge $e = ((x_1,y_1),(x_2,y_2)) \in E_{\mathcal{H}}$ its (unoriented) dual $e' = \{(x_1,y_2),(x_2,y_1)\}$. The dual edges connect vertices on the dual lattice $\mathcal{H}' := \{(x,y) \in \mathbb{Z}^+ \times \mathbb{Z}^+ : x + y \text{ odd}\}$. We declare the dual edge e' closed if e is closed, and open if e is open. We then define the

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

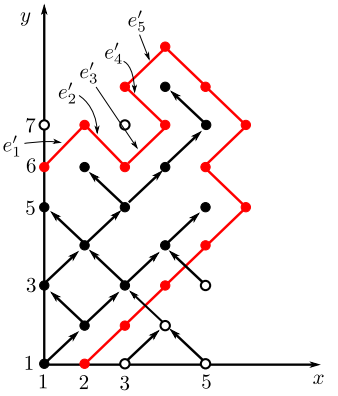


Figure 3. *This example shows a finite oriented cluster of the origin $\mathcal{C}_{(1,1)}$: filled black circles are vertices in $\mathcal{C}_{(1,1)}$ while empty black circles are vertices that do not belong to $\mathcal{C}_{(1,1)}$. The oriented, black edges are open in $\mathcal{H}$, while the closed edges of $\mathcal{H}$ are not drawn. The red contour and red vertices belong to the dual lattice $\mathcal{H}'$. Since $Y_{max} = 5$, the dual contour π_∂ starts from $(1, 6)$, and follows the closed dual edges colored red, ending at $(2, 1)$. Edges of $\mathcal{H}$ pointing out of $\mathcal{C}_{(1,1)}$ are all closed (not drawn), whereas edges pointing into $\mathcal{C}_{(1,1)}$ may be open – such as the edge $((5, 3), (4, 4))$ – or closed.*

(outer edge-) boundary of $\mathcal{C}_{(1,1)}$ as the set of dual edges

$$\partial\mathcal{C}_{(1,1)} = \{e' : \text{ exactly one of the two endpoints of } e \text{ is in } \mathcal{C}_{(1,1)}\}. \tag{137}$$

Since $\mathcal{H}$ is a cone in $\mathbb{Z}^+ \times \mathbb{Z}^+$, and $\mathcal{C}_{(1,1)}$ is connected per definition, $\partial\mathcal{C}_{(1,1)}$ is a union of connected contours in $\mathcal{H}'$, which along with (parts of) the x- and y-axes encircle $\mathcal{C}_{(1,1)}$. Assume now that the event $\{Y_{\max} = k\}$ occurs. This implies that $(1, k) \in \mathcal{C}_{(1,1)}$ and $(1, k + 2) \notin \mathcal{C}_{(1,1)}$. So, define the edges and their duals

$$\hat{e}_{k,1} = ((1, k), (2, k + 1)), \qquad \hat{e}'_{k,1} = \{(1, k + 1), (2, k)\},$$
$$\hat{e}_{k,2} = ((2, k + 1), (1, k + 2)), \qquad \hat{e}'_{k,2} = \{(1, k + 1), (2, k + 2)\}.$$

Now, if $(2, k + 1) \notin \mathcal{C}_{(1,1)}$, then since $(1, k) \in \mathcal{C}_{(1,1)}$, the dual edge $\hat{e}'_{k,1} \in \partial\mathcal{C}_{(1,1)}$ (and $\hat{e}'_{k,2} \notin \partial\mathcal{C}_{(1,1)}$). In this case, define $\hat{e}_k = \hat{e}_{k,1}$. On the other hand, if $(2, k + 1) \in \mathcal{C}_{(1,1)}$, then since $(1, k + 2) \notin \mathcal{C}_{(1,1)}$, the dual edge $\hat{e}'_{k,2} \in \partial\mathcal{C}_{(1,1)}$ (and $\hat{e}'_{k,1} \in \partial\mathcal{C}_{(1,1)}$). In this case, define $\hat{e}_k = \hat{e}_{k,2}$. In both of these cases, the vertex $(1, k + 1)$ is the starting point of the dual $\hat{e}'_k$, which is in $\partial\mathcal{C}_{(1,1)}$, and the other dual edge with endpoint $(1, k+1)$ is not in $\partial\mathcal{C}_{(1,1)}$. Then we start exploring $\partial\mathcal{C}_{(1,1)}$, starting from $e'_1 := \hat{e}'_k$ by following the dual edges in this connected component of $\partial\mathcal{C}_{(1,1)}$. That is, the next dual edge in the path, denoted by e'_2, is incident to $(2, k)$ if $e'_1 = \{(1, k + 1), (2, k)\}$ and to $(2, k + 2)$ if $e'_1 = \{(1, k + 1), (2, k + 2)\}$. Then we continue from the other endpoint of e'_2, and so on. We continue this exploration process either indefinitely (if $\mathcal{C}_{(1,1)}$ is infinite), or until we reach the x-axis (if $\mathcal{C}_{(1,1)}$ is finite). As we explain next, these are the only two possible outcomes. For an example of the second outcome, see Figure 3.

Denote by $\pi_\partial = (e'_1, e'_2, \ldots)$ the path (as a sequence of dual edges) obtained this way. It is possible that π_∂ visits the y-axis above $(1, k + 1)$ (say at $(1, y')$ with $y' > k$), but since $Y_{\max} = k$, this can only happen when $(2, y') \in \mathcal{C}_{(1,1)}$ and $(1, y' + 1) \notin \mathcal{C}_{(1,1)}$, and then we can always continue the path π_∂ by traversing the dual edge $\{(1, y'), (2, y' + 1)\}$. However, π_∂ cannot visit the y-axis below $(1, k)$, since then we would have encircled the entire $\mathcal{C}_{(1,1)}$, starting from $(1, k + 1)$, without containing $(1, 1)$, a contradiction. Hence, one of the two remaining cases happens. We either find an infinite path π_∂ in $\partial\mathcal{C}_{(1,1)}$, and then we set $\pi_\partial(k)$ to be the sequence of its first k edges. Or, we find a finite path π_∂ that reaches the x-axis, in particular, the dual vertex $(2, 1)$. This path has length at least k, since the path starts at $(1, k + 1)$, and the y coordinate only changes by ± 1 between consecutive vertices on the path. In this case we again set $\pi_\partial(k)$ to be the sequence of the first k edges of π_∂.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

We now categorize edges of $\pi_\partial(k)$ (all are in $\partial\mathcal{C}_{(1,1)}$) as follows. Recall that edges of $\mathcal{H}$ in (130) are oriented (directed), and recall (137). Given $\mathcal{C}_{(1,1)}$ let us call the dual edge $e' \in \partial\mathcal{C}_{(1,1)}$ an *outward* dual edge, if for the edge $e = ((x_1, y_1), (x_2, y_2))$ it holds that $(x_1, y_1) \in \mathcal{C}_{(1,1)}$ and $(x_2, y_2) \notin \mathcal{C}_{(1,1)}$ and let us call e' an *inward* dual edge if $(x_1, y_1) \notin \mathcal{C}_{(1,1)}$ and $(x_2, y_2) \in \mathcal{C}_{(1,1)}$. Per definition of $\mathcal{C}_{(1,1)}$ in (135), the outward edges and their duals are all closed. However, $\mathcal{C}_{(1,1)}$ does not determine the status of inward dual edges.

We now prove that for any realization of $\mathcal{C}_{(1,1)}$, at least half of the edges of $\pi_\partial(k)$ are outward dual edges, and hence closed. Let us introduce the notation $\pi_0 = (1, k+1), \pi_1, \pi_2, \ldots, \pi_k, \ldots$ for the vertices of the path π_∂ in order, and define the directed edge $\underline{e}'_i = (\pi_{i-1}, \pi_i)$ for all $i \geq 1$ (the directed version of e'_i). Then for all outward dual edges $e'_i \in \pi_\partial$, $\underline{e}'_i$ is pointing to the right (in the direction of increasing x coordinate), and for all inward dual edges $e'_i \in \pi_\partial$, $\underline{e}'_i$ is pointing to the left (in the direction of decreasing x coordinate). Since $\pi_\partial(k)$ starts from $(1, k+1)$, which is part of the y-axis, and remains in the positive quadrant, at least half of its dual edges have to be directed to the right, thus, duals of outward edges. Hence, at least $k/2$ dual edges in $\pi_\partial(k)$ are closed. Further, since every vertex in $V_\mathcal{H}$ has at most two outgoing (nondual) edges, in every possible realization $(e'_1, e'_2, \ldots, e'_k)$ of $\pi_\partial(k)$ we can find $k/4$ edges that are all closed and that their oriented nondual edges in $\mathcal{H}$ all start from different vertices.

By (131), the probability that a given edge (and its dual) is closed is at most δ. As mentioned before (135), the status of different edges are not independent, however, $\mathcal{A}_{(x_1,y_1),(x_2,y_2)}$ is independent of $\mathcal{A}_{(x'_1,y'_1),(x'_2,y'_2)}$ if $(x_1, y_1) \neq (x'_1, y'_1)$. That is, two edges $e_1, e_2 \in E_\mathcal{H}$ are open or closed independently if their starting points are distinct.

We call a given connected path $(e'_1, \ldots, e'_k)$ of dual edges eligible if it is a possible realization of $\pi_\partial(k)$ (of which one requirement is that one of the endpoints of e'_1 is $(1, (k+1))$). Then, for all such $(e'_1, e'_2, \ldots, e'_k)$,

$$\mathbb{P}\big(\pi_\partial(k) = (e'_1, \ldots, e'_k)\big) \leq \delta^{k/4}. \tag{138}$$

Next, we upper bound the number of eligible paths $(e'_1, \ldots, e'_k)$. Since $(e'_1, \ldots, e'_k)$ is a path starting from $(1, k+1)$ on the dual lattice $\mathcal{H}'$ isomorphic to a quadrant of $\mathbb{Z}^2$, each of the k steps in the exploration of $\pi_\partial(k)$ can be taken in one of at most three directions. This yields that the number of possible trajectories is at most 3^k. Therefore, by a union bound,

$$\mathbb{P}(Y_{\max} = k) \leq \mathbb{P}\left(\bigcup_{(e'_1,\ldots,e'_k) \text{ eligible}} \{\pi_\partial(k) = (e'_1, \ldots, e'_k)\} \right) \leq 3^k \delta^{k/4}. \tag{139}$$

Then (139) implies that

$$\mathbb{P}(Y_{\max} < \infty) = \sum_{k=1}^{\infty} \mathbb{P}(Y_{\max} = k) \leq \sum_{k=1}^{\infty} 3^k \delta^{k/4} < 1/4, \tag{140}$$

whenever $\delta \in (0, (1/15)^4)$. Consequently, $\mathbb{P}(Y_{\max} = \infty) > 3/4$. Finally, recalling $\mathcal{I}_1$ from (133), $\mathbb{P}(\mathcal{I}_1) > 1/3$ for large enough K by Claim 6.6. By the stochastic dominance in (134), it follows from a union bound that

$$\mathbb{P}\big(\mathrm{CP}_{f,\lambda}(H, \mathbb{1}_{v_1}) \text{ survives locally at } v_1\big) \geq \mathbb{P}\big(\mathcal{I}_1 \cap \{Y_{\max} = \infty\}\big) \geq 1 - 2/3 - 1/4 > 0.$$

This proves local survival of $\mathrm{CP}_{f,\lambda}(H, \mathbb{1}_{v_1})$ with positive probability. □

Proof of Theorem 2.1. Lemma 6.4 states that for some $M \geq 1$ there exists K_1 such that for $K > K_1$ and $\ell(K)$ as in (109) $H_{K,\ell(K)}$ can be M-embedded in $\mathcal{T}$ almost surely. Set $\bar{\lambda} = \lambda/M^{2\mu}$ and let $K_0(\bar{\lambda})$ be given by Lemma 6.8. Now let $K > \max(K_0(\bar{\lambda}), K_1)$. Then Lemma 6.4 yields that $H_{K,\ell(K)}$ can be M-embedded in $\mathcal{T}$ almost surely. Let v_1 be the center of the first star in the embedded $H_{K,\ell(K)}$. Recalling (108) in

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Definition 6.3, we observe that the process $\mathrm{CP}_{f,\lambda}(\mathcal{T}, \mathbb{1}_{v_1})$ restricted to the vertices of the embedded $H_{K,\ell(K)}$ is stochastically dominated from below by the process $\mathrm{CP}_{f,\bar{\lambda}}(H_{K,\ell(K)}, \mathbb{1}_{v_1})$ on a standalone copy of $H_{K,\ell(K)}$. Combining this with Lemma 6.8 and $K \geq K_0(\bar{\lambda})$ implies that $\mathrm{CP}_{f,\bar{\lambda}}(H_{K,\ell(K)}, \mathbb{1}_{v_1})$ survives locally with positive probability. This, along with the fact that with positive probability, $\mathrm{CP}_{f,\lambda}(\mathcal{T}, \mathbb{1}_{\varnothing})$ infects v_1 at some point in time finishes the proof. □

Proof of Theorem 2.5(a). This is an easy consequence of Theorem 2.1 by stochastic domination, noting that $\max(d_u, d_v)^{\mu} \leq (d_u d_v)^{\mu}$. □

7. The configuration model: k-cores sustain the infection when stars do not

In this section we will prove part (b) of Theorem 2.8. A crucial difference between the classical contact process and the degree-dependent version in this regime is that star-graphs do not sustain the infection, in fact they heal quickly when $\mu > 1/2$, by Claim 6.6. However, we know from Section 6.1 that the approximating Galton-Watson tree shows global survival (only), which suggests long survival on the configuration model. So we set out to find a new structure – a subgraph – embedded in the configuration model that sustains the degree-dependent contact process for a long time. Generally speaking, any (sparse) graph can sustain the infection linearly long in its number of edges (or vertices), hence to prove exponentially long survival in n we aim for this subgraph to have *linearly* many vertices in n.

To find such a subgraph, we need to take into account that vertices that have either too high or too low degree cannot sustain the infection, either because the penalty f on them is too high or because λ is assumed to be close to 0. The subgraph we found is the k-core – a maximal subgraph of the configuration model where each vertex has degree at least k inside the same subgraph – but with a twist: in the original configuration model with fat-tailed degrees, the k-core contains vertices of very high degree (e.g., polynomials of n). However, the degree-dependent CP near these vertices would have too high penalty f, so we need to exclude them from the k-core.

As a result we look at the k-core of *not* the original configuration model, but the subgraph obtained after *removing all vertices of degree above* a threshold value M, where now M is a constant depending only on k but not on the total number of vertices n. It is a priori unclear whether such a low truncation value even produces a connected graph, let alone contains a linear-sized k-core (i.e., containing at least some constant times n vertices). So our first step is to study the dependence between $M = M(k)$ and k so that a linear sized k-core still exists in the configuration model where all vertices of degree above M are removed. As we will see, the exact relation between k and $M(k)$ will be crucial on whether the infection manages to spread: indeed, any vertex in the k-core can spread to (typically) k vertices while it (typically) experiences rate $\lambda M(k)^{-\mu}$ coming from the original degrees. Intuitively, CP will survive on the k-core if $kM(k)^{-\mu}$ is growing with k, which limits the value of $M(k)$ to a polynomial of k. We will show that "essentially the lowest" power of k we can achieve so that a linear k-core exists is $M(k) = k^{(1+o(1))/(3-\tau)}$, which then readily yields the $\mu < 3 - \tau - o(1)$ criterion for survival in Theorem 2.8.

After finding the linear-sized k-core on vertices of degree at most $M(k)$, we show that $\mathrm{CP}_{f,\lambda}$ survives on this k-core. For this step, our proof is a nontrivial adaptation of the proof of [56, Theorem 1.2(b)], which shows long survival in the original contact process model on $(d+1)$-regular random graphs, when λ is above the lower critical $\lambda_1(\mathbb{T}^d)$ on d-regular trees needed for global survival [57]. However, in our case we have λ arbitrarily close to 0. Fortunately, we *can* choose k as a function of λ that makes the process locally supercritical. This also makes the proof different from that in [56] even beyond finding the k-cores. First, we define the k-core of a graph.

Definition 7.1. Let G be any simple, finite graph. For a fixed positive integer k, the k-core of G is the largest induced subgraph $\mathrm{Core}_k(G)$ of G such that every vertex in $\mathrm{Core}_k(G)$ has degree at least k within $\mathrm{Core}_k(G)$.

It is not hard to see that the k-core $\mathrm{Core}_k(G)$ in Definition 7.1 is well-defined – but may be empty – by the following algorithm producing it. First, delete all vertices of G that have degree less than k along

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

with their incident edges. Then do the same with the resulting graph, repeatedly, until no new vertex is deleted. The output of this algorithm is the unique largest induced subgraph of G with all degrees at least k. Note that the k-core of a graph might be empty or may contain more than one component.

7.1. The subgraph spanned on low-degree vertices contains a k-core

Our first goal is to prove the existence of a k-core in the configuration model after we remove all vertices with too high degrees. Throughout this section, we work with the configuration model $\mathrm{CM}(\underline{d}_n) =: G$ in Definition 1.9 on the degree sequence $\underline{d}_n = (d_1, \dots, d_n)$ that satisfies the regularity assumptions in Assumption 1.10, and the weak power-law empirical degrees of Assumption 1.11 with exponent $\tau > 2$ and error $\varepsilon > 0$.

We now set up the procedure of removing all vertices above some degree M and the edges attached to those in $\mathrm{CM}(\underline{d}_n)$. This is often called a *targeted attack* on the graph. Because the graph is formed by a random matching, and the half-edges that have one endpoint at a vertex with degree larger than M and another endpoint at a vertex with degree at most M are also removed, the degrees in the remaining graph are random.

Definition 7.2 (Configuration model under targeted attack). Consider the configuration model $\mathrm{CM}(\underline{d}_n)$ in Definition 1.9 on the degree sequence $\underline{d}_n = (d_1, \dots, d_n)$. Fix some value $M \geq 0$. Denote

$$\begin{aligned} \mathcal{V}_{\leq M} &:= \{i \leq n : d_i \leq M\}, \quad V_{\leq M} := |\mathcal{V}_{\leq M}|, \qquad H_{\leq M} := \sum_{i=1}^{n} d_i \mathbb{1}_{\{d_i \leq M\}}, \\ \mathcal{V}_{>M} &:= \{i \leq n : d_i > M\}, \quad V_{>M} := |\mathcal{V}_{\geq M}|, \qquad H_{>M} := \sum_{i=1}^{n} d_i \mathbb{1}_{\{d_i > M\}}, \end{aligned} \tag{141}$$

Let $G_n[\mathcal{V}_{\leq M}]$ denote the (random) subgraph of $\mathrm{CM}(\underline{d}_n)$ that is spanned on the vertex set $\mathcal{V}_{\leq M}$. For any $v \in \mathcal{V}_{\leq M}$, we denote the *random* degree of v in $G_n[\mathcal{V}_{\leq M}]$ by $\widetilde{d}_v$, and we write $\widetilde{n}_i$ for the number of vertices with degree i in $G_n[\mathcal{V}_{\leq M}]$. For any $z \geq 0$ define

$$\widetilde{F}_{n,M}(z) := \frac{1}{V_{\leq M}} \sum_{v \in \mathcal{V}_{\leq M}} \mathbb{1}_{\{\widetilde{d}_i \leq z\}} = \frac{1}{V_{\leq M}} \sum_{i \leq z} \widetilde{n}_i, \tag{142}$$

and let $\widetilde{D}_{n,M}$ denote a random variable with cdf $\widetilde{F}_{n,M}(z)$.

Observe that $\widetilde{F}_{n,M}(z)$ is the new empirical distribution of the degrees, after the targeted attack. This distribution is random, caused by the random matching that generated the graph before the attack. The quantities in (141) all depend on n, which we suppress in notation. We are ready to state the existence of the k-core in the configuration model under attack.

Theorem 7.3. *Consider the configuration model* $\mathrm{CM}(\underline{d}_n) =: G_n$ *in Definition 1.9 on the degree sequence* $\underline{d}_n = (d_1, \dots, d_n)$ *that satisfies the regularity assumptions in Assumption 1.10, and the weak power-law empirical degrees of Assumption 1.11 with exponent* $\tau \in (2, 3)$ *and error* $\varepsilon > 0$. *Let*

$$\eta_{\min} := \frac{(3-\tau)}{(3-\tau) - \varepsilon(\tau-1)} \cdot \frac{1+\varepsilon}{1-\varepsilon} - 1, \tag{143}$$

and assume τ, ε *are such that* $\eta_{\min} \in [0, \infty)$. *Fix a large enough positive integer* k, *and for any* $\eta > \eta_{\min}$ *let* $M := M_{k,\eta} = k^{(1+\eta)/(3-\tau)}$. *Let* $G_n[\mathcal{V}_{\leq M}]$ *be the configuration model under attack in Definition 7.2, and denote by* $\mathrm{Core}_k(G_n[\mathcal{V}_{\leq M}])$ *its* k*-core. Then there exists some* $\rho = \rho(k) > 0$ *such that*

$$\lim_{n \to \infty} \mathbb{P}\Big(|\mathrm{Core}_k(G_n[\mathcal{V}_{\leq M}])| \geq \rho n \Big) = 1, \tag{144}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Further, conditioned on its vertex set and degree sequence, $\mathrm{Core}_k(G_n[\mathcal{V}_{\le M}])$ *is itself a configuration model.*

Remark 7.4 (Asymptotics of ρ). The proof of Theorem 7.3 shows that there exists a constant $c' > 0$ such that

$$\rho(k) > c' k^{-\frac{(\tau-1)(1+\varepsilon)}{(2-(\tau-1)(1+\varepsilon))}}. \tag{145}$$

Note that in this lower estimate only the lower bound exponent in Assumption 1.11 appears. We comment that $\eta_{\min} \in [0, \infty)$ implies that $\varepsilon < (3-\tau)/(\tau-1)$, which is exactly the condition that the lower bound on the tail-exponent, $(\tau-1)(1+\varepsilon)$ in (9), stays strictly below 2. Hence a k-core exists for all k when the estimates on the empirical power law are so that the tail is always heavier than a power law with infinite variance. Without the truncation at M, that is, for pure power laws, such a result is already known, see [37] and [28]. Here we specify the truncation value M for which the result stays valid. We comment that when $\varepsilon = 0$ in Assumption 1.11, then our proof can be strengthened so that $M = \Theta(k^{1/(3-\tau)})$ guarantees the existence of a large k-core after the targeted attack.

We will prove Theorem 7.3 below using the following two lemmas and the results of Janson and Luczak [37] that we will state soon. The first lemma says that the random empirical distribution of $G_n[\mathcal{V}_{\le M}]$ converges in probability, assuming the regularity assumptions on the original degrees. We use notation from Definition 7.2. Given the degree sequence $\underline{d}_n$, D_n stands for the random variable that follows the empirical distribution F_n of $\underline{d}_n$ in (8), and D is the random variable following the limiting distribution in Assumption 1.10. Define then

$$q_{n,M} := \mathbb{E}[D_n \mathbb{1}_{\{D_n \le M\}}]/\mathbb{E}[D_n], \qquad q_M := \mathbb{E}[D \mathbb{1}_{\{D \le M\}}]/\mathbb{E}[D], \tag{146}$$

and we collect the errors below M between the n-dependent degree distribution D_n and the limit D as follows:

$$\begin{aligned} \delta_n := \max \Big\{ & |q_{n,M}/q_M - 1|,\ |(1-q_{n,M})/(1-q_M) - 1|, \\ & \max_{i \le M, \mathbb{P}(D=i)=0} \mathbb{P}(D_n = i), \max_{i \le M, \mathbb{P}(D=i)\neq 0} |\mathbb{P}(D_n = i)/\mathbb{P}(D=i) - 1| \Big\}, \end{aligned} \tag{147}$$

with $\delta_n \to 0$ when Assumption 1.10 holds. Typically, for M large q_M is close to 1 so the relative error of $1 - q_{M,n}$ to $1 - q_M$ is driving the maximum in the first row, while the second row is only over values $i \le M$.

Lemma 7.5 (Degree distribution of CM under attack). *Consider the configuration model* $\mathrm{CM}(\underline{d}_n)$ *in Definition 1.9 on the degree sequence* $\underline{d}_n = (d_1, \dots, d_n)$ *that satisfies the regularity assumptions in Assumption 1.10. Fix any* $M > 0$ *constant, and let* $q_{n,M}, q_M, \delta_n$ *as in (146) and (147). Define the following random variable* $\widetilde{D}_M$*: for all* $i \le M$*, let*

$$\begin{aligned} p_M(i) := \mathbb{P}(\widetilde{D}_M = i) &= \sum_{j=i}^{M} \frac{\mathbb{P}(D=j)}{\mathbb{P}(D \le M)} \binom{j}{i} q_M^i (1-q_M)^{j-i} \\ &= \mathbb{P}(\mathrm{Bin}(D, q_M) = i \mid D \le M). \end{aligned} \tag{148}$$

Let $X_{n,M}(i) := \widetilde{n}_i / V_{\le M} = \mathbb{P}(\widetilde{D}_{n,M} = i \mid G_n[\mathcal{V}_{\le M}])$ *be the random empirical degree distribution of* $G_n[\mathcal{V}_{\le M}]$*. Then for all* $\varepsilon_n > 0$ *that satisfies* $\varepsilon_n \gg \max\{\delta_n, 1/\sqrt{n}\}$*,*

$$\mathbb{P}\Big(\sup_{i \le M} \big| X_{n,M}(i) - p_M(i) \big| \ge \varepsilon_n \Big) = O\Big(\frac{M^3}{n \varepsilon_n^2} \Big) \to 0. \tag{149}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Further, $\lim_{n\to\infty} \mathbb{E}[\widetilde{D}_{n,M} \mid G_n[\mathcal{V}_{\le M}]] = \mathbb{E}[\widetilde{D}_M]$ *in probability. So, the empirical degree distribution* $\widetilde{D}_{n,M}$ *of* $G_n[\mathcal{V}_{\le M}]$ *satisfies Assumption 1.10 with probability tending to* 1. *Furthermore,* $G_n[\mathcal{V}_{\le M}]$ *is itself a configuration model on its vertex set, conditioned on the degrees of its vertices.*

The second lemma proves that the limiting degree distribution of $G_n[\mathcal{V}_{\le M}]$ is a truncated weak power law with truncation close to M if the original degree distribution satisfied the weak power law assumption.

Lemma 7.6 (Truncated power laws after targeted attack). *Consider the configuration model* $\mathrm{CM}(\underline{d}_n)$ *in Definition 1.9 on the degree sequence* $\underline{d}_n = (d_1, \dots, d_n)$ *that satisfies the regularity assumptions in Assumption 1.10, and the power-law empirical degrees of Assumption 1.11 with exponent* τ *and exponent-error* $\varepsilon \ge 0$. *Let* $M > 0$ *be a constant (i.e., not depending on n, but it may depend on* ε*), and let*

$$\widetilde{z}_{\max}(M) := 2^{-1}(c_\ell/(2c_u))^{\frac{1}{(\tau-1)(1+\varepsilon)}} M^{(1-\varepsilon)/(1+\varepsilon)}. \tag{150}$$

Consider the limiting degree distribution $\widetilde{F}_M(z) =: \sum_{i\le z} p_M(i)$ *in (148) of* $G_n[\mathcal{V}_{\le M}]$ *in Lemma 7.5. Then there exist constants* $\widetilde{c}_\ell, \widetilde{c}_u, M_0$, *such that whenever* $M \ge M_0$, *for all* $z \in [z_0, \widetilde{z}_{\max}(M)]$, *it holds that*

$$\frac{\widetilde{c}_\ell}{z^{(\tau-1)(1+\varepsilon)}} \le 1 - \widetilde{F}_M(z) \le \frac{\widetilde{c}_u}{z^{(\tau-1)(1-\varepsilon)}}. \tag{151}$$

The proof shows that $\widetilde{c}_\ell = c_\ell 2^{-(\tau-1)(1+\varepsilon)-2}$ and $\widetilde{c}_u = 2c_u$ are valid choices (although they may not be optimal). Since the proofs of Lemmas 7.5 and 7.6 are fairly standard, we provide them in the Appendix on pages 78 and 82.

With these lemmas at hand, the proof of Theorem 7.3 relies on the result of Janson and Luczak [37], describing the k-core of the configuration model. To state this result, we introduce some notation.

For a random variable D and $p \in [0,1]$, we let $X_{D,p}$ denote a random variable with Binomial(D,p) distribution. That is,

$$\mathbb{P}(X_{D,p} = r) = \sum_{l=r}^{\infty} \mathbb{P}(D = l)\binom{l}{r} p^r (1-p)^{l-r}.$$

We then define the following functions:

$$h(D,p) := \mathbb{E}[X_{D,p}\mathbb{1}\{X_{D,p} \ge k\}], \qquad h_1(D,p) := \mathbb{P}(X_{D,p} \ge k). \tag{152}$$

Note that both h and h_1 are increasing in p, and $h(D,0) = h_1(D,0) = 0$. Moreover, $h(D,1) = \mathbb{E}[D\mathbb{1}\{D \ge k\}] \le \mathbb{E}[D]$, and $h_1(D,1) = \mathbb{P}(D \ge k) \le 1$.

Then the theorem of Janson and Luczak is as follows. They use the same regularity Assumption 1.10 as we do.

Theorem 7.7 (Theorem 2.3 in [37]). *Consider the configuration model* $G_n := \mathrm{CM}(\underline{d}_n)$ *in Definition 1.9 on the degree sequence* $\underline{d}_n = (d_1, \dots, d_n)$ *that satisfies the regularity assumptions in Assumption 1.10. For* $k \ge 2$ *be fixed, let* $\mathrm{Core}_k := \mathrm{Core}_k(G_n)$ *be the k-core of* G_n. *Let*

$$\hat{p} := \max\{p \le 1 : \mathbb{E}[D]p^2 = h(D,p)\}. \tag{153}$$

Then, if $\hat{p} > 0$ *and* $\mathbb{E}[D]p^2 < h(D,p)$ *for p in some interval* $(\hat{p} - \varepsilon, \hat{p})$, *then* $\mathrm{Core}_k(G_n)$ *is nonempty whp, and*

$$|\mathcal{V}(\mathrm{Core}_k)|/n \xrightarrow{\mathbb{P}} h_1(D,\hat{p}), \qquad |\mathcal{E}(\mathrm{Core}_k)|/n \xrightarrow{\mathbb{P}} h(D,\hat{p})/2 = \mathbb{E}[D]\hat{p}^2/2. \tag{154}$$

We first need a small extension of this theorem.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Claim 7.8. Suppose there is a value p_- where $\mathbb{E}[D]p_-^2 < h(D, p_-)$ holds (see below (153)). Then, there is a nonzero fixed point $p_\star > p_-$ of (153) so that in the interval $(p_\star, p_\star - \varepsilon)$ the inequality $\mathbb{E}[D]p^2 < h(D, p)$ holds. Then, $\mathrm{Core}_k(G_n)$ is nonempty and

$$\begin{aligned}&\mathbb{P}\Big(|\mathcal{V}(\mathrm{Core}_k)|/n \le h_1(D, p_-)(1-\varepsilon)\Big)\\ &\quad\le \mathbb{P}\Big(|\mathcal{V}(\mathrm{Core}_k)|/n \le h_1(D, p_\star)(1-\varepsilon)\Big) \to 0\end{aligned} \tag{155}$$

as $n \to \infty$.

Sketch of proof. The first statement, namely that $p^\star$ exists, follows from the continuity of the function $\mathbb{E}[D]p^2 - h(D, p)$. The second statement, that the size of the k-core is at least $h_1(D, p_\star)(1-\varepsilon)$, follows from the proof of [37, Theorem 2.3]. Namely, the only case where the proof of [37, Theorem 2.3] does not apply directly is where the function $f(p) = E[D]p^2 - h(D, p)$ does not cross the 0-line at its maximal 0 but rather, it touches it. Nevertheless, if one finds a smaller value p_- where the function $f(p)$ is in the negative, it implies that there must be zero-point $p_\star$ of f where the function crosses the 0 level line. In this case the proof there yields that the density of the k-core is at least $h_1(D, p_\star)$, hence (155) holds. This can be found on [37, page 56-57], where a value p_- for which $f(p_-) < 0$ implies the upper bound on the stopping time of a pruning algorithm generating the k-core for (154). That is, the continuous time pruning algorithm of sequentially removing vertices of degrees at least k and their outgoing edges is guaranteed to stop by time $t = -\log(p_-)$, that is, one can set $t_2 = -\log(p_-)$ at the bottom [37, page 9]. An upper bound on the stopping of the pruning algorithm gives a lower bound on the number of remaining vertices forming the k-core. In our case, by not knowing whether p_- is adjacent to the maximal fixed point and whether f touches or crosses 0 there, we lose the upper bound on the k-core size.

With Lemmas 7.5 and 7.6 at hand, we are ready to prove Theorem 7.3 by checking the conditions of Theorem 7.7 and Claim 7.8.

Proof of Theorem 7.3. First, we prove (144) holds: we will check that the conditions of Theorem 7.7 hold for $G_n[\mathcal{V}_{\le M}]$ with probability tending to 1. First, Lemma 7.5 implies that $G_n[\mathcal{V}_{\le M}]$ is a configuration model (conditioned on its vertices and their degrees), and its (random) degree sequence satisfies Assumption 1.10 with probability tending to 1. We will use the notations of Lemmas 7.5 – 7.6, so, $\widetilde{D}_M$ denotes the limiting degree distribution of $G_n[\mathcal{V}_{\le M}]$. Since $h(\widetilde{D}_M, p)$ in (152) is a continuous function of p, it is enough for us to find a particular choice of p with $\mathbb{E}[\widetilde{D}_M]p^2 < h(\widetilde{D}_M, p)$. Based on the tail probabilities of $\widetilde{D}_M$ from (151), in particular the exponent $\tau \in (2, 3)$ and the constant $\widetilde{c}_\ell$ in the lower bound, which holds for $z \in [z_0, \widetilde{z}_{\max}(M)]$ with $\widetilde{z}_{\max}(M)$ defined in (150), our goal is to find two positive constants $a_- < a_+$ and $\xi > 3 - \tau$ and an interval

$$I_p := [p_-, p_+] := \Big[a_- k^{-(\xi/(3-\tau)-1)}, a_+ k^{-(\xi/(3-\tau)-1)}\Big]. \tag{156}$$

We will show that when $p \in I_p$, then $\mathbb{E}[\widetilde{D}_M]p^2 < h(\widetilde{D}_M, p)$. Using (152), and that $X_{l_1,p}$ stochastically dominates $X_{l_2,p}$ when $l_2 > l_1$, we estimate, for some constant β and exponent $\xi > 3 - \tau$ to be chosen later,

$$\begin{aligned}h(\widetilde{D}_M, p) &= \sum_{l=k}^{M} \mathbb{P}(\widetilde{D}_M = l) \sum_{r=k}^{l} r \cdot \mathbb{P}(X_{l,p} = r)\\ &\ge \sum_{l=\beta k^{\xi/(3-\tau)}}^{M} \mathbb{P}(\widetilde{D}_M = l)\mathbb{E}\big[X_{l,p}\mathbb{1}\{X_{l,p} \ge k\}\big]\\ &\ge \mathbb{P}\big(\widetilde{D}_M \ge \beta k^{\xi/(3-\tau)}\big)\mathbb{E}\big[X_{\beta k^{\xi/(3-\tau)},p}\mathbb{1}\{X_{\beta k^{\xi/(3-\tau)},p} \ge k\}\big].\end{aligned} \tag{157}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

We bound the first factor on the rhs of (157) first. Recalling that the tail-bound on $\mathbb{P}(\widetilde{D}_M > z)$ in (151), we get

$$\mathbb{P}\big(\widetilde{D}_M \geq \beta k^{\xi/(3-\tau)}\big) \geq \widetilde{c}_\ell \Big(\beta k^{\xi/(3-\tau)}\Big)^{-(\tau-1)(1+\varepsilon)}, \tag{158}$$

on the condition that $\beta k^{\xi/(3-\tau)} \leq \widetilde{z}_{\max}(M)$ which we now check. (This is the place where we need that the truncation point M is high enough.) We expand $\widetilde{z}_{\max}(M)$ in (150) as a function of k using that $M = k^{(1+\eta)/(3-\tau)}$ below (143). We write $\widetilde{C}$ for the prefactor in (150) that only depends on $c_\ell, c_u, \tau, \varepsilon$:

$$\widetilde{z}_{\max}(M) = \widetilde{C} M^{(1-\varepsilon)/(1+\varepsilon)} = \widetilde{C} k^{\left(\frac{1+\eta}{3-\tau}\right)(1-\varepsilon)/(1+\varepsilon)}. \tag{159}$$

Treating $\beta, \widetilde{C}$ as constants while k can be chosen arbitrarily large, the rhs of (159) is larger than $\beta k^{\xi/(3-\tau)}$ for all sufficiently large k when

$$\xi < (1+\eta)(1-\varepsilon)/(1+\varepsilon), \tag{160}$$

which shall lead to the assumption that $\eta > \eta_{\min}$ in (143) shortly. Next, we bound the second factor on the rhs of (157). For any variable X, it holds that $\mathbb{E}[X\mathbb{1}_{\{X\geq k\}}] = \mathbb{E}[X] - \mathbb{E}[X\mathbb{1}_{\{X<k\}}] \geq \mathbb{E}[X] - k\mathbb{P}(X < k)$. In (157) $X \sim \mathrm{Bin}(\beta k^{\xi/(3-\tau)}, p)$, and with the choice $a_- := 2/\beta$, we can lower bound its mean using that $p > p_-$ in (156) as $\beta k^{\xi/(3-\tau)} p > 2k^{\xi/(3-\tau)} k^{-(\xi/(3-\tau)-1)} = 2k$. Hence, a Chernoff bound applies and we obtain that

$$k\mathbb{P}\big(X_{\beta k^{\xi/(3-\tau)},p} < k\big) \leq k \exp\Big(-\beta k^{\xi/(3-\tau)} p/8\Big) \leq k \exp(-k/4), \tag{161}$$

for all $p > p_-$ in (156). Using again that $p > p_-$ implies $\beta k^{\xi/(3-\tau)} p \geq 2k$, the second factor in (157) can be bounded from below for all sufficiently large k as

$$\mathbb{E}[X_{\beta k^{\xi/(3-\tau)},p}\mathbb{1}\{X_{\beta k^{\xi/(3-\tau)},p} \geq k\}] \geq \beta k^{\xi/(3-\tau)} p - k\exp(-k/4) \geq \beta k^{\xi/(3-\tau)} p/2. \tag{162}$$

Substituting (158) and (162) into (157) gives, for all β, $p > p_-$ in (156) and all $\xi > 3-\tau > 0$ that

$$h(\widetilde{D}_M, p) \geq (\widetilde{c}_\ell/2) \cdot \beta^{1-(\tau-1)(1+\varepsilon)} k^{(1-(\tau-1)(1+\varepsilon))\xi/(3-\tau)} p. \tag{163}$$

Thus, $h(\widetilde{D}_M, p) > \mathbb{E}[\widetilde{D}_M]p^2$ holds when

$$(\widetilde{c}_\ell/2)\beta^{1-(\tau-1)(1+\varepsilon)} k^{(1-(\tau-1)(1+\varepsilon))\xi/(3-\tau)} > \mathbb{E}[\widetilde{D}_M]p. \tag{164}$$

At this point we still have the freedom of choosing β and $\xi > 3-\tau$ provided that the relation between η, ξ in (160) holds. Since $p < a_+ k^{-(\xi/(3-\tau)-1)}$ in (156), first we compare the powers of k on both sides. The inequality (164) holds for all sufficiently large k if

$$(1-(\tau-1)(1+\varepsilon))\xi/(3-\tau) \geq -(\xi/(3-\tau)-1).$$

After elementary computations, the smallest ξ that satisfies this inequality, and hence the threshold η for (160) is

$$\xi \geq \xi_{\min} := \frac{3-\tau}{3-\tau-\varepsilon(\tau-1)}, \qquad \eta > \eta_{\min} = \frac{\xi_{\min}(1+\varepsilon)}{1-\varepsilon} - 1, \tag{165}$$

which equals $\eta_{\min}$ in (143). Comparing now constants on the two sides of (164) yields that

$$a_+ := (\widetilde{c}_\ell/2\mathbb{E}[\widetilde{D}_M])\beta^{1-(\tau-1)(1+\varepsilon)}.$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Solving the inequality $a_- = 2/\beta < a_+$ gives that the interval I_p is nonempty whenever

$$\beta > \left(4\mathbb{E}[\widetilde{D}_M]/\widetilde{c}_\ell\right)^{1/[2-(\tau-1)(1+\varepsilon)]}.$$

Summarizing, we have found that whenever β satisfies this inequality, and p is in the interval

$$I_p = \left[(2/\beta)\cdot k^{-(1/(3-\tau-\varepsilon(\tau-1))-1)}, (\widetilde{c}_\ell/2\mathbb{E}[\widetilde{D}_M])\beta^{1-(\tau-1)(1+\varepsilon)}k^{-(1/(3-\tau-\varepsilon(\tau-1))-1)}\right],$$

then the required inequality for the existence of the k-core holds. This implies that $\hat{p} > p_+$, and we can estimate the asymptotic proportion of the k-core (154), $h_1(\widetilde{D}_M, \hat{p}) \geq h_1(\widetilde{D}_M, p_+)$ following similar steps as in (157):

$$\begin{aligned} h_1(\widetilde{D}_M, \hat{p}) &\geq \mathbb{P}(\widetilde{D}_M \geq k^{\xi/(3-\tau)})\mathbb{P}\big(X_{k^{\xi/(3-\tau)},p_+} \geq k\big) \\ &\geq \widetilde{c}_\ell k^{-\xi(\tau-1)(1+\varepsilon)/(3-\tau)}(1-\exp(-k/4)), \end{aligned}$$

using the same $\xi = \xi_{\min}$ and Chernoff bound as in (165) and in (161), yielding (145) in Remark 7.4.

Finally, we need to check that conditioned on its vertex set and degree sequence, $\mathrm{Core}_k(G_n[\mathcal{V}_{\leq M}])$ is itself a configuration model. This follows from the fact that every matching of half-edges within $\mathrm{Core}_k(G_n[\mathcal{V}_{\leq M}])$, given its degree sequence, has equal probability by the construction of the configuration model. □

This finishes the first combinatorial part, that is, the existence of a large k-core. We now (slowly) transition to studying the contact process on the k-core. The proof of Theorem 2.8(b) is based on a structural property of $\mathrm{Core}_k(G_n[\mathcal{V}_{\leq M}])$ that we define next. This structural property guarantees that an infected set of vertices can pass the infection to many other vertices in a unit time step.

Definition 7.9 ((δ,k)-expansion). Fix any $\delta \in (0,1)$ and an even positive integer k. We say that a (multi)graph G on n vertices is (δ,k)-good if for every set $\{v_1,\dots,v_{\lfloor \delta n\rfloor}\}$ of $\lfloor \delta n\rfloor$ vertices in G, we can choose a subset $\mathcal{I}_g$ of the indices of size $|\mathcal{I}_g| \geq \lfloor \delta n\rfloor/8$ such that each $v_i : i \in \mathcal{I}_g$ has $k/2$ neighbors $w_{i,1},\dots,w_{i,k/2}$ in G such that the vertices $v_i, i \in \mathcal{I}_g$ and $w_{i,j}, i \in \mathcal{I}_g, j \leq k/2$ are all distinct.

A graph being (δ,k)-good is somewhat stronger than requiring that the 1-neighborhood of any $\lfloor \delta n\rfloor$ many vertices expands by a factor $k/16$, since we need enough individual vertices that expand to $k/2$ different vertices. The following lemma proves that $\mathrm{Core}_k(G_n[\mathcal{V}_{\leq M}])$ has the (δ,k)-good property for small enough $\delta > 0$.

Lemma 7.10. *Consider the configuration model* $\mathrm{CM}(\underline{d}_n) =: G_n$ *in Definition 1.9 on the degree sequence* $\underline{d}_n = (d_1,\dots,d_n)$ *so that for an even integer* $k \geq 128$ *and constant* $\zeta > 1$, $d_i \in [k, k^\zeta]$ *holds for all* $i \in [n]$. *Then there exists some* $\delta_0 = \delta_0(k,\zeta) > 0$ *independent of n, such that for all* $\delta < \delta_0$,

$$\mathbb{P}(G_n \text{ is } (\delta,k)\text{-good}) > 1 - e^{-n\delta\log(1/\delta)/8}. \tag{166}$$

Proof. Let $v_1,\dots,v_{\lfloor \delta n\rfloor}$ be distinct fixed vertices in G_n. We will explore, that is, gradually reveal the neighbors of these vertices, as follows. In the first exploration step, we reveal the first k edges adjacent to v_1 (according to an arbitrary ordering), one by one. When revealing an edge, we say that a *collision* happens at v_1 if the revealed edge either leads to one of $v_1, v_2, \dots, v_{\lfloor \delta n\rfloor}$, or is parallel to an edge revealed earlier (note that we allow self-loops and multiple edges in G_n). During this first step, as soon as the number of collisions at v_1 reaches two, we stop revealing the connections of v_1 and color v_1 red. If the number of collisions does not reach two by the end of step 1, we color v_1 green, and we assign the revealed distinct neighbors of v_1, outside the set $\{v_1,\dots,v_{\lfloor \delta n\rfloor}\}$, the labels $w_{1,1}, w_{1,2}, \dots, w_{1,n_1}$. Here, $n_1 \in [k-1,k]$, since there was at most one collision.

In the second step we reveal the first k edges adjacent to v_2, one by one, *including* the potential edges (at most two) that lead to v_1 and have already been revealed. Now we say that a collision happens at v_2

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

if a revealed edge either leads to one of $v_1, v_2, \dots, v_{\lfloor \delta n \rfloor}$, (except when it was already revealed starting from v_1, and thus the collision happened at v_1 in which case we do not count it as a new collision), or it leads to one of $w_{1,1}, w_{1,2}, \dots, w_{1,n_1}$ (in case v_1 was colored green), or is parallel to an edge already revealed at v_2. Again, as soon as the number of collisions at v_2 reaches two during this step, we stop revealing the edges of v_2 and color v_2 red. If the number of collisions at v_2 does not reach two by the end of the step, we color v_2 green, and assign the revealed distinct neighbors of v_2, outside the set $\{v_1, \dots, v_{\lfloor \delta n \rfloor}, w_{1,1}, w_{1,2}, \dots, w_{1,n_1}\}$ the labels $w_{2,1}, w_{2,2}, \dots, w_{2,n_2}$. Here, $n_2 \in [k-3, k]$, since at most three edges caused collisions at either v_1 (these can connect to v_2) or v_2.

We then continue this procedure, in each step revealing the first k connections of $v_3, \dots, v_{\lfloor \delta n \rfloor}$ analogously to the above, with one modification: if at the beginning of step i, when starting to reveal the neighbors of vertex v_i ($i \geq 2$), v_i already has at least $k/4$ adjacent revealed edges coming from the already processed vertex set $\{v_1, v_2, \dots, v_{i-1}\}$, then we do not reveal any new connections at v_i, but color it blue, and continue to the next step $i+1$, with v_{i+1}.

After all the $\lfloor \delta n \rfloor$ steps are done, let $\mathcal{I}_g := \{i_1, \dots, i_g\}$ denote the indices and $\{v_{i_1}, \dots, v_{i_g}\}$ be the set of green vertices (subset of $\{v_1, \dots, v_{\lfloor \delta n \rfloor}\}$). We will prove that with probability at least $1 - \exp(-Cn)$ for some constant $C > 0$, $|\mathcal{I}_g| \geq \lfloor \delta n \rfloor / 8$ and $n_i \geq k/2$ for all $i \in \mathcal{I}_g$. So, the green vertices along with their revealed neighbors $\{w_{i,j} : i \in \mathcal{I}_g, j \leq n_i\}$ demonstrate the (δ, k)-good property of G_n in Definition 7.9.

Later, we take a union bound over all subsets of size $\lfloor \delta n \rfloor$, but for now we fix a choice of $\{v_1, \dots, v_{\lfloor \delta n \rfloor}\}$. First, we bound the number of blue vertices. When at step j, we reveal at most two edges that connect v_j to some $v_{j'}$, for $j' > j$. Hence, we reveal at most $2\lfloor \delta n \rfloor$ edges with both endpoints in the set $\{v_1, \dots, v_{\lfloor \delta n \rfloor}\}$, which we call *internal* edges. These involve at most $4\lfloor \delta n \rfloor$ half-edges at $\{v_1, \dots, v_{\lfloor \delta n \rfloor}\}$. Since more than $16\lfloor \delta n \rfloor / k$ vertices adjacent to at least $k/4$ internal edges would involve more than $4\lfloor \delta n \rfloor$ half-edges, by the pigeonhole principle, for all $k \geq 128$:

$$\begin{aligned} |\text{Blue vertices}| &= |\{i \in [\lfloor \delta n \rfloor] : v_i \text{ is adjacent to at least } k/4 \text{ internal edges}\}| \\ &\leq 4\lfloor \delta n \rfloor / (k/4) = 16\lfloor \delta n \rfloor / k \leq \lfloor \delta n \rfloor / 8. \end{aligned} \tag{167}$$

Hence, the exploration reveals the neighborhood of at least $7\lfloor \delta n \rfloor / 8$ and at most $\lfloor \delta n \rfloor$ vertices that can be either red or green. Next, we bound the number of red vertices. Here we use that G_n is a configuration model, with all degrees in the interval $[k, k^\zeta]$. Thus we can carry out the exploration process above by matching the first (at most) k half-edges of each vertex under consideration. After revealing the jth edge, for $j \leq k\lfloor \delta n \rfloor - 1$, we have discovered at most j new vertices and so half-edges attached to at most $\lfloor \delta n \rfloor + j$ vertices can cause a collision when matching the $j+1$th half-edge. And, there are at least $nk - 2j - 1$ remaining unmatched half-edges to choose from. Let us denote by $\mathcal{F}_j$ the σ-algebra generated by the outcome of the matching of the first j half-edges. Then, for all $k > 2$ and sufficiently small $\delta = \delta(k) > 0$, and for any realization in $\mathcal{F}_j$

$$\mathbb{P}(\text{collision at } j+1^{\text{st}} \text{ edge} \mid \mathcal{F}_j) \leq \frac{(\lfloor \delta n \rfloor + j)k^\zeta}{nk - 2j - 1} \leq \frac{(\delta n + \delta n k)k^\zeta}{(1-2\delta)nk} \leq 2\delta k^\zeta.$$

Let $Y_j = 1$ if revealing the j^{th} edge causes a collision and $Y_j = 0$ otherwise. Then $(Y_1, Y_2, \dots)$ is dominated by a sequence of i.i.d. Bernoulli variables with parameter $2\delta k^\zeta$. We color v_i red if at least two collisions happen at step i, that is, if at least 2 of the Y_j variables corresponding to the at most k revealed edges at v_i are 1. So, with $X_{n,p}$ a binomial variable as before, independently across different v_i,

$$\mathbb{P}(v_i \text{ is red}) \leq \mathbb{P}(X_{k, 2\delta k^\zeta} \geq 2) \leq k^2 4\delta^2 k^{2\zeta} = 4\delta^2 k^{2+2\zeta}. \tag{168}$$

Combining (167) and (168) yields that the number of red vertices is stochastically dominated by a Binomial random variable with parameters $\lfloor \delta n \rfloor$ and $4\delta^2 k^{2+2\zeta} =: q$. Hence, by a crude upper bound on the binomial coefficients,

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

$$\mathbb{P}(|i : v_i \text{ is red}| \geq 3\lfloor \delta n \rfloor/4) \leq \mathbb{P}(X_{\lfloor \delta n \rfloor, q} > 3\lfloor \delta n \rfloor/4) = \sum_{r > 3\lfloor \delta n \rfloor/4} \binom{\lfloor \delta n \rfloor}{r} q^r (1-q)^{\lfloor \delta n \rfloor - r}$$
$$\leq \lfloor \delta n \rfloor 2^{\lfloor \delta n \rfloor} q^{3\lfloor \delta n \rfloor/4} = \lfloor \delta n \rfloor 2^{\lfloor \delta n \rfloor} \left(4\delta^2 k^{2+2\zeta}\right)^{3\lfloor \delta n \rfloor/4}. \tag{169}$$

after substituting the value of q. Rewriting (169) we obtain for small enough $\delta = \delta(k) > 0$,

$$\begin{aligned} &\mathbb{P}(|i : v_i \text{ is red}| \geq 3\lfloor \delta n \rfloor/4) \\ &\leq \lfloor \delta n \rfloor \exp\left((3/2)\log(\delta)\lfloor \delta n \rfloor + (5/2)\log(2)\lfloor \delta n \rfloor + (3/4)\log(k^{2+2\zeta})\lfloor \delta n \rfloor\right) \\ &\leq C \exp\left(-(5/4)\log(1/\delta)\delta n\right). \end{aligned} \tag{170}$$

We bound the number of ways to choose the $\lfloor \delta n \rfloor$ vertices $S = \{v_1, \ldots, v_{\lfloor \delta n \rfloor}\}$:

$$\binom{n}{\lfloor \delta n \rfloor} \leq \frac{n^{\lfloor \delta n \rfloor}}{(\lfloor \delta n \rfloor)!} \leq \frac{n^{\delta n}}{\exp\left(\delta n \log(\delta n) - \delta n\right)} = \exp\left(\delta n(1 + \log(1/\delta))\right). \tag{171}$$

Combining (170) and (171), we obtain for some positive constant C that for all small enough $\delta = \delta(k) > 0$,

$$\begin{aligned} &\mathbb{P}(\exists S \subset G : |S| = \lfloor \delta n \rfloor, \text{ at least } 3\lfloor \delta n \rfloor/4 \text{ red vertices in } S) \\ &\leq \exp\left(\delta n(1 + \log(1/\delta)) - (5/4)\log(1/\delta)\right) < \exp\left(-n\delta \log(1/\delta)/8\right). \end{aligned}$$

Combining this with (167), we obtain that with probability at least $1 - \exp(-Cn)$, for any choice of $v_1, \ldots, v_{\lfloor \delta n \rfloor}$, there are at least $\lfloor \delta n \rfloor/8$ green vertices among $v_1, \ldots, v_{\lfloor \delta n \rfloor}$. The green vertices, per design, have at most one collision among their at least $3k/4$ revealed edges. Hence, each green vertex has at least $k/2$ neighbors in G_n, that are all distinct from each other and from $v_1, \ldots, v_{\lfloor \delta n \rfloor}$, demonstrating the (δ, k)-good property. This finishes the proof. □

The next lemma studies a contact process with lower infection rate than $\mathrm{CP}_{f,\lambda}$ with $f = \max(x, y)^\mu$ on a (δ, k)-good graph and shows that when $\lfloor \delta n \rfloor$ vertices are infected, their neighborhood sustains the infection for a unit of time:

Lemma 7.11. *Fix some $\lambda > 0$, $\mu \in [1/2, 1)$ and $\zeta > 1$ satisfying $\mu\zeta < 1$. Then there exist constants $C' > 0$ and $k_0 = k_0(\lambda, \mu, \zeta)$ so that for all $k > k_0$ even, the following holds. Let G_n be any multi-graph with degree sequence $\underline{d}_n = (d_1, \ldots, d_n)$ satisfying $d_i \in [k, k^\zeta]$ for all $i \in [n]$, so that G_n is (δ, k)-good for some fixed $\delta > 0$. Let $(\underline{\tilde{\xi}}_t)_{t \geq 0}$ be a contact process $\mathrm{CP}_{f_+,\lambda}$ with $f_+(x, y) \equiv k^{\zeta\mu}$ on G_n. Then, for all sufficiently large n, and any $t \geq 0$,*

$$\mathbb{P}\left(|\underline{\tilde{\xi}}_{t+1}| \geq \lfloor \delta n \rfloor \;\middle|\; |\underline{\tilde{\xi}}_t| \geq \lfloor \delta n \rfloor\right) \geq 1 - \exp(-n\delta/(193e)). \tag{172}$$

By (20) in Corollary 3.2, the process $\mathrm{CP}_{f_+,\lambda}$ on G_n dominates from below the contact process $\mathrm{CP}_{f,\lambda}$ with $f(x, y) = \max(x, y)^\mu$, since $f(d_u, d_v) = \max(d_u, d_v)^\mu \leq k^{\zeta\mu} = (\max_{i \leq n} d_i)^\mu$.

Proof. We shall fix $k > 400$. Since $|\underline{\tilde{\xi}}_t| \geq \lfloor \delta n \rfloor$ in the conditioning in (172), denote the first $\lfloor \delta n \rfloor$ infected vertices by $S_t := \{v_1, \ldots, v_{\lfloor \delta n \rfloor}\}$. Since G_n is (δ, k)-good, choose the index set $\mathcal{I}_g$ with size $|\mathcal{I}_g| \geq \lfloor \delta n \rfloor/8$ guaranteed by the (δ, k)-good property in Definition 7.9 and write $w_{i,1}, \ldots, w_{i,k/2}$ for the distinct neighbors of each $v_i, i \in \mathcal{I}_g$. For each $i \in \mathcal{I}_g$ define the event $\mathcal{A}(v_i)$ as

$$\begin{aligned} \mathcal{A}(v_i) := \{&v_i \text{ infects at least 87 vertices among } w_{i,1}, \ldots, w_{i,k/2} \\ &\text{that stay infected by time } t + 1\}. \end{aligned} \tag{173}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

We will prove that

$$\mathbb{P}(\mathcal{B}) := \mathbb{P}\Big(\sum_{i \in \mathcal{I}_g} \mathbb{1}_{\mathcal{A}(v_i)} \geq \lfloor \delta n \rfloor / (32e) \Big) \geq 1 - \exp(-\delta n/(193e)). \tag{174}$$

Then, on the event $\mathcal{B}$, at least $87\lfloor \delta n \rfloor/(32e)$ vertices among $\{w_{i,j}\}_{1 \leq i \leq \lfloor \delta n \rfloor, 1 \leq j \leq k/2}$ are infected at time $t+1$, and since $32e \approx 86.98$, this implies that $|\tilde{\xi}_{t+1}| \geq \lfloor \delta n \rfloor$ holds in (172), proving the lemma.

For (174), we first give a lower bound on $\mathbb{P}(\mathcal{A}(v_i))$ in (173). The probability that v_i does not heal in the time interval $[t, t+1]$ is $1/e$. Given that v_i does not heal, it infects each of $w_{i,1}, \ldots, w_{i,k/2}$, in the time interval $[t, t+1]$, with probability at least $1 - \exp(-\lambda k^{-\mu\zeta})$, as the infection rate $r(v_i, w_{i,j})$ from v_i to $w_{i,j}$ is $\lambda k^{-\mu\zeta}$. A given $w_{i,j}$ infected in the time interval $[t, t+1]$ stays infected until $t+1$ with conditional probability at least $1/e$. So, given that v_i does not heal until time $t+1$, the number of infected vertices among $w_{i,1}, \ldots, w_{i,k/2}$ at time $t+1$ is stochastically dominated from below by a Binomial random variable with parameters $k/2$ and $(1/e)(1 - \exp(-\lambda k^{-\mu\zeta})) \geq (1/e)(\lambda k^{-\mu\zeta}/2) := p$. This lower bound holds whenever $k \geq \lambda^{-1/\mu\zeta}$, which holds for all $k \geq 2$ when $\lambda < 1$ and for all sufficiently large k when $\lambda > 1$. Hence,

$$\mathbb{P}(\mathcal{A}(v_i)) \geq \mathbb{P}(\tilde{\xi}_s(v_i) = 1 \ \forall s \in [t, t+1]) \cdot \mathbb{P}(X_{k/2,p} \geq 87) \geq \mathrm{e}^{-1} \cdot \mathbb{P}(X_{k/2,p}) \geq 87).$$

The mean $\mathbb{E}[X_{k,p}] = \lambda k^{1-\mu\zeta}/(4e)$ and since $\mu\zeta < 1$, this quantity grows with k, and we can choose k large enough so that $\mathbb{E}[X_{k,p}] \geq 2 \cdot 87$. Then, by a Chernoff bound,

$$\mathbb{P}(\mathcal{A}(v_i)) \geq e^{-1} \cdot \mathbb{P}(X_{k/2,p} \geq 87) \leq \mathrm{e}^{-1}(1 - \mathrm{e}^{-2 \cdot 87/12}) \geq 1/(2e).$$

Now we use Corollary 3.2 to obtain $\mathbb{1}_{\mathcal{A}_i}, i \in \mathcal{I}_g$ is stochastically dominated from below by independent events with success probability $1/(2e)$. Thus, another Chernoff bound finishes the proof of (174):

$$\mathbb{P}(\mathcal{B}) \geq \mathbb{P}(X_{\lceil \lfloor \delta n \rfloor / 8 \rceil, 1/(2e)} \geq \lfloor \delta n \rfloor / (32e)) \geq 1 - \exp\big(- \lfloor \delta n \rfloor / (16 \cdot 12e) \big),$$

completing the proof of the lemma with $C' := 1/(193e)$ where we increased $16 \cdot 12 = 192$ by one to compensate for dropping the integer part. □

With Theorem 7.3, and Lemmas 7.10 and 7.11 at hand, we are ready to prove Theorem 2.8(b).

Proof of Theorem 2.8(b). Observe that in (143) in Theorem 7.3,

$$\zeta_{\min} := \frac{\eta_{\min} + 1}{3 - \tau} = \frac{1}{3 - \tau - \varepsilon(\tau - 1)} \cdot \frac{1 + \varepsilon}{1 - \varepsilon}. \tag{175}$$

The inequality (18), that is, that $\mu < (3 - \tau - \varepsilon(\tau - 1))(1 + \varepsilon)/(1 - \varepsilon)$ and (175) together imply that for all μ satisfying (18) one can choose $\zeta > \zeta_{\min}$ so that $\zeta\mu < 1$ also holds. Fix such a ζ. Then, Theorem 7.3 states that for all sufficiently large but fixed k even, a linear sized k-core of $\mathrm{CM}(\underline{d}_n)$ exists after removing all vertices of degree larger than $M = k^{(1+\eta)/(3-\tau)} =: k^{\zeta}$, that is, for all $\varepsilon' > 0$, for all sufficiently large n,

$$\mathbb{P}(\mathcal{A}_n) := \lim_{n \to \infty} \mathbb{P}\Big(|\mathrm{Core}_k(G_n[\mathcal{V}_{\leq k^{\zeta}}])| \geq \rho n \Big) = 1 - \varepsilon'/3, \tag{176}$$

and conditioned on its vertex set and degree sequence, $\mathrm{Core}_k(G_n[\mathcal{V}_{\leq k^{\zeta}}])$ is itself a configuration model. Applying Lemma 7.10 on $\mathrm{Core}_k(G_n[\mathcal{V}_{\leq k^{\zeta}}])$ then yields that for all small enough $\delta > 0$

$$\begin{aligned} \mathbb{P}(\mathcal{B}_n \mid \mathcal{A}_n) &:= \mathbb{P}(\mathrm{Core}_k(G_n[\mathcal{V}_{\leq k^{\zeta}}]) \text{ is } (\delta, k)\text{-good} \mid \mathcal{A}_n) \\ &> 1 - e^{-n\rho(k)\delta/8} > 1 - \varepsilon'/4. \end{aligned} \tag{177}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Consider the process $(\underline{\xi}_t)_{t\geq 0} \sim \mathrm{CP}_{f,\lambda}$ with $f(x,y) = \max(x,y)^\mu$ on G_n. For any $t \geq 0$ define the event

$$\mathcal{I}_t := \{\text{at least } \delta\rho n \text{ vertices of } \mathrm{Core}_k(G_n[\mathcal{V}_{\leq k^\zeta}]) \text{ are infected at time } t\}.$$

On the event $\mathcal{A}_n \cap \mathcal{B}_n$, all vertices in $H_n := \mathrm{Core}_k(G_n[\mathcal{V}_{\leq k^\zeta}])$ have original degrees in the interval $[k, k^\zeta]$ within G_n, hence $\mathrm{CP}_{f,\lambda}$ restricted to H_n is dominated from below by a contact process on H_n with $f(x,y) = k^{\zeta\mu}$, exactly as in Lemma 7.11. Hence, Lemma 7.11 applies for $H_n := \mathrm{Core}_k(G_n[\mathcal{V}_{\leq k^\zeta}])$,

$$\mathbb{P}(\mathcal{I}_{t+1} |\, \mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{I}_t) > 1 - e^{-n\rho(k)\delta/(193e)}. \tag{178}$$

whenever k is larger than $k_0 = k_0(\lambda, \mu, \eta)$. This latter condition dictates our choice of k. Starting from the all-infected initial condition, (178) implies that on the event $\mathcal{A}_n \cap \mathcal{B}_n$ the extinction time of $\mathrm{CP}_{f,\lambda}$ is dominated from below by a geometric random variable with success probability $\exp(-C'n)$. Hence, the process survives until time $\exp(nC'/2)$ with probability at least $1 - \varepsilon'/3$. Combining this with (176) and (177) yields that the process $\mathrm{CP}_{f,\lambda}(G_n, \underline{1}_{G_n})$ exhibits long survival, finishing the proof. □

8. The configuration model: survival through a network of stars

The proof of Theorem 2.3 (a) (which then implies Theorem 2.8(a)) follows the proof of Theorem 4 in [6], that is, the proof of the exponentially long survival of the classical contact process on the configuration model with subexponentially tailed degree distributions. We need some modifications to adapt the proof there to the degree-penalized model. Since the proof in [6] is rather lengthy, we only provide an outline of the main steps, and we focus on explaining the necessary modifications for the degree-penalized version. We direct the interested reader to [6, Section 6, 7] for a full proof.

A common way to show exponentially long survival for the classical contact process is to find $\Theta(n)$ many embedded star-graphs in the configuration model with paths of bounded degree vertices connecting them (similarly as we proved local survival on Galton Watson trees in Section 6 here for the penalized version). The exact structure in this case, corresponding to [6, Definition 5.1], is an embedded expander-graph.

For a graph H, and a subset of vertices $A \subset V(H)$ we denote by $\mathcal{N}_H(A, r)$ the set of vertices at most distance r from A. For some $m \geq 1$, let also $\deg_{G,\leq m}(u)$ denote the number of neighbors of vertex u in G that have degree at most m.

Definition 8.1. Let $H = (V(H), E(H))$ be a graph with $|V(H)| \leq |V(G)|$. We say that H is an α-expander if for every subset $A \subset V(H)$ with $|A| \leq \alpha|V(H)|$,

$$|\mathcal{N}_H(A, 1)| \geq 2|A|. \tag{179}$$

Let G be a connected graph with $|V(G)| \geq |V(H)|$. We say that H is an (R, m, j)-embedded α-expander in G if there is a choice of vertices $W_0 \subseteq V(G)$ with $|W_0| = |V(H)|$ with a one-to-one map between W_0 and $V(H)$, so that there exist, for each edge $(u,v) \in E(H)$ an associated path $\pi^G_{u,v}$ in G between $u, v \in W_0$ that satisfies

$$|\pi^G_{u,v}| \leq R \text{ for all } u, v \in E(H_{W_0}), \tag{180}$$

$$\deg_G(w) \in [2, m] \quad \text{for all } w \in \pi^G_{u,v} \setminus \{u, v\}, \quad u, v \in W_0, \tag{181}$$

$$\deg_G(u) \in [j, 2j], \text{ and } \deg_{G,\leq m}(u) \geq j/2 \quad \text{for all } u \in W_0. \tag{182}$$

Observe that (179) is the expansion property of H, while (182) ensures that the embedded vertices of W_0 serve as star-graphs in G, that is, they have sufficiently high degree. Meanwhile, (180) and (181) ensure that the paths corresponding to each edge of H are fairly short and occur on low-degree vertices, so that even the degree-penalized contact process can pass through them with good probability. Next, we prove the following structural lemma, corresponding to [6, Lemma 6.1].

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Lemma 8.2. *Consider the configuration model $G_n := \mathrm{CM}(\underline{d}_n)$ in Definition 1.9 on the degree sequence $\underline{d}_n = (d_1, \dots, d_n)$ that satisfies the regularity assumptions in Assumption 1.10. Suppose that its limiting degree distribution D has heavier tails than stretched-exponential with stretch-exponent ζ, for some $\zeta > 0$, as in Definition 1.8. Then, for any sufficiently large $m > 0$ there exists a $j_0 > m$ such that whenever $j > j_0$ then there exists $\alpha, \beta, R > 0$ with $R \le o(j^{\zeta})$ such that the following holds whp. The graph G_n contains an (R, m, j)-embedded α-expander $H := H_{W_0}$ on the vertex set W_0 with $|W_0| \ge \beta n$.*

Proof. We choose m so high that

$$\bar{b} := \frac{\mathbb{E}[D(D-1)\mathbb{1}_{\{D \le m\}}]}{\mathbb{E}[D\mathbb{1}_{\{D \le m\}}]} \ge \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}(1-\varepsilon) \quad \text{and} \quad \mathbb{E}[D\mathbb{1}_{\{D \le m\}}] \ge (1-\varepsilon)\mathbb{E}[D]. \tag{183}$$

The proof is similar to the proof of [6, Lemma 6.1], and consists of the following steps.

Step 1. Targeted attack. Recall the configuration model under targeted attack from Definition 7.2. Here we carry out the attack above degree $2j$ (considering (182) and (181)), and we denote the remaining graph by $G_n[\mathcal{V}_{\le 2j}]$, and the degree of a vertex v in $G_n[\mathcal{V}_{\le 2j}]$ by $\tilde{d}_v$. This ensures that all remaining degrees are at most $2j$.

The second criterion in (183) ensures that each half-edge of a vertex with degree in the interval $[j, 2j]$ is matched to a vertex with degree below m with probability at least $1 - \varepsilon$. Hence, denoting by $u_j := \mathbb{P}(D \in [j, 2j])$, a Chernoff bound similar as in [6, Lemma 7.1, part (4) and Claim 7.2] ensures that there are at least $\varepsilon n u_j$ many vertices with $\deg_G(u) \in [j, 2j]$ and $\deg_{G,\le m}(u) \ge j/2$, as required in (182).

Step 2. Exploration. Let $W := \{v \in G_n : \deg_{G_n}(v) \in [j, 2j], \deg_{G_n,\le m}(v) \in [j/2, 2j]\}$. We find the vertex set W_0 of H_{W_0} as a subset of W. We explore the R-neighborhood in $G_n[\mathcal{V}_{\le 2j}]$ of each vertex $w \in W$ simultaneously, but we refrain from exploring vertices that have degree (within G_n) higher than m, as we explain now in more detail. As is usual for the configuration model, we construct the graph along with the exploration, see, for example, [67]. That is, we put all half-edges attached to vertices in W –initially in an arbitrary order – in an active list and call these unmatched, also all other half-edges in the graph are initially unmatched (but not active). Due to the uniform matching property, we then sequentially may choose the currently first half-edge h_* in the active list and match it to a uniformly chosen yet unmatched half-edge, its pair $p(h_*)$, forming a new edge. We remove these two half-edges from the set of unmatched/active list. We call $v(p(h_*))$ the vertex that $p(h_*)$ belongs to. If the degree of $v(p(h_*))$ is at most m, if any, we append the not- active unmatched half-edges adjacent to $v(p(h_*))$ to the end of the active list. If however the vertex $v(p(h_*))$ has degree above m, we keep the rest of its unmatched half-edges in the unmatched not-active list. This way we explore the neighborhood of W in $\mathcal{V}_{\le 2j}$ generation by generation, but we do not follow the neighborhood of vertices with degree $> m$, which become leaves in the exploration tree (unless there is an overlap and such a vertex is matched to more than once). As we explore in a breadth-first-search manner, there is an associated exploration tree and thus it is possible to keep track of parents/generation numbers during this exploration. As the exploration is indexed by discrete steps, there is a (random) step number when we finish exploring the Rth generation, that is, revealing all vertices that are at distance at most R from W_0 reachable from paths with vertices of degree $\le m$ except the first and possibly the last vertex. Each at-this-moment active half-edge can be associated to the edge boundary of the R-neighborhood of some vertex $w \in W_0$. We denote these exploration-R-neighborhoods by $\mathcal{N}_{\le 2j,m}(w, R)$ for each $w \in W_0$. We need to bound the overlap of these neighborhoods. For this, we follow [6] and notice that an overlap between $\mathcal{N}_{\le 2j,m}(w, R)$ and $\mathcal{N}_{\le 2j,m}(w', R)$ means that w' is part of $\mathcal{N}_{\le 2j,m}(w, 2R)$.

The criteria in (183) implies, by standard coupling arguments (see, e.g., [6] or [7]) that the exploration inside $\mathcal{V}_{\le 2j}$ with refraining to explore vertices with degree $> m$ (except those in W_0 that we claimed initially active), can be approximated by a supercritical branching process with mean offspring $\bar{b}$ from the first generation onward. Also, the total number of half-edges in $G_n[\mathcal{V}_{\le 2j}] \ge (1-\varepsilon)n\mathbb{E}[D]$ since

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

$j > m$), which can be used to derive an upper bound on the probabilities that we match to a particular vertex at any time during the exploration. Following notation in [6], here we introduce a new parameter r that (contrary to usual notation for radius) is controlling the number of allowed overlaps between neighborhoods of vertices in W_0. Given j, R, we choose the value of the integer r so that for a $w \in W$ the expected number of vertices of W_0 that lie in the neighborhood $\mathcal{N}_{\leq 2j,m}(w, 2R)$ in the exploration is small compared to r. Then it will be unlikely that different neighborhoods $\mathcal{N}_{\leq 2j,m}(w, R)$ intersect in more than r vertices. To determine r, we note that degrees of explored vertices in the exploration are $\leq m$ while we also may match to vertices of degree $\leq 2j$ and some of these (with degree in $[j, 2j]$) may be the ones being part of W_0. So we estimate the size of the $(2R-1)$-st generation of the branching process and then we sample the degrees in generation $2R$ according to size-biased distribution $D_j^\star$ of $D\mathbb{1}_{D\leq 2j}$:

$$\mathbb{E}[|\mathcal{N}_{\leq 2j,m}(v, 2R) \cap W_0|] \approx \mathbb{E}[\partial\mathcal{N}(v, 2R-1)]\mathbb{P}(D_j^\star \in [j, 2j]) \approx j\bar{b}^{2R-1} \cdot \frac{u_j j}{d} \tag{184}$$

where $u_j = \mathbb{P}(j \leq D \leq 2j)$ and $d = \mathbb{E}[D]$. Since the above contains ε errors both in the numerator and denominator, we set the requirement that

$$\frac{\bar{b}^{2R-1} j^2 u_j}{d} \leq \frac{r}{10}. \tag{185}$$

Step 3. Graph contraction. We carry out a graph contraction on $G_n[\mathcal{V}_{\leq 2j}]$ as follows: We associate a vertex v_w to each of the neighborhoods $\mathcal{N}_{\leq 2j,m}(w, R)$, $w \in W$, forming the (contracted) vertex set V'. We associate to $v_w \in V'$ as many half-edges as there are unmatched half-edges adjacent to any vertex in $\mathcal{N}_{\leq 2j,m}(w, R)$ after the exploration process in Step 2 finishes exploring generation R of W_0. Furthermore, let V'' be the set of vertices of $G_n[\mathcal{V}_{\leq 2j}]$ that have not been touched in the exploration process, i.e., vertices that belong to none of the neighborhoods $\cup_{w\in W}\mathcal{N}_{\leq 2j,m}(w, R)$. Then the graph G_n' is obtained by matching the half-edges of the vertex set $V' \cup V''$ uniformly at random. About the degrees of vertices in V' in G_n', that is, the number of unmatched half-edges in each $\mathcal{N}_{\leq 2j,m}(w, R)$, [6, Lemma 7.5] proves the following:

There exists positive constant $\varepsilon', \varepsilon'', R_0$, depending only on the degree sequences $(\underline{d}_n)_{n\geq 1}$, such that for all bounded positive numbers R_1, R, r satisfying

$$R_0 \leq \min\{R_1, R - R_1\}, \qquad 800r \leq \varepsilon'^2(\bar{b}(1-\varepsilon''))^{R_1-1} j, \tag{186}$$

$$\frac{\bar{b}^{2R_1-1} j^2 u_j}{d} \leq \frac{1}{10^4}, \qquad \frac{\bar{b}^{2R-1} j^2 u_j}{d} \leq \frac{r}{10}, \tag{187}$$

the number of vertices in V' with degree at least M is at least $(\varepsilon'/2)|V'|$ whp, where

$$M = \frac{\varepsilon'^3(\bar{b}(1-\varepsilon''))^{R-1} j}{8}.$$

Note that this is reasonable, as a typical degree of a $w_v, v \in V'$ has cca $\approx j\bar{b}^{R-1}$ many unmatched half-edges after finishing generation R, via the coupling to a branching process, and M is on the same scale but much smaller. Note also that M grows with j, in fact all R, R_1, r are dependent on and growing with j, while $R_0, \varepsilon, \varepsilon''$ are not.

Step 4. Given that the conditions (186)-(187) are satisfied, [6] proves the existence of a high-degree core in G_n', which is an (R, m, j)-embedded α-expander in G_n (the core-number is chosen so that vertices that are not in V' have too low degree to be part of the core of G_n', so the core will be a subset of W). Here we mean core in the sense of Definition 7.1. [6] chooses r, R, R_1 as the solution to the following equations:

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

$$\frac{\bar{b}^{2R-1} j^2 u_j}{d} = \frac{r}{10}, \tag{188}$$

$$\varepsilon'^2 (\bar{b}(1-\varepsilon''))^{R_1-1} j = 800r, \tag{189}$$

$$\bar{b}^{2R_1-1} = \frac{d}{10^4 j^2 u_j}. \tag{190}$$

It is relatively easy to check that for large j this set of choices satisfies then (186)-(187). We also set r, R, R_1 given by (188)–(190), and now compute the value of R: Combining (188) and (189) gives the relation between R_1 and R:

$$2R - 1 = (R_1 - 1) \cdot \frac{\log(\bar{b}(1-\varepsilon''))}{\log(\bar{b})} + \frac{\log\left(\frac{d\varepsilon'^2}{8000 j u_j}\right)}{\log(\bar{b})}. \tag{191}$$

Next, we note that (190) yields

$$2R_1 - 1 = \frac{\log\left(\frac{d}{10^4 j^2 u_j}\right)}{\log(\bar{b})}. \tag{192}$$

Since we assume that $\mathbb{E}[D^2] < \infty$, it holds that $\lim_{j\to\infty} j^2 u_j = \lim_{j\to\infty} j^2 \mathbb{P}(D \in [j, 2j]) = 0$. So R_1 can be chosen arbitrarily large by increasing j. Using this in (191) yields

$$\begin{aligned} 2R - 1 &\approx \frac{\log\left(\frac{d}{10^4 j^2 u_j}\right)}{2\log(\bar{b})} \cdot \frac{\log(\bar{b}(1-\varepsilon''))}{\log(\bar{b})} + \frac{\log\left(\frac{d\varepsilon'^2}{8000 j u_j}\right)}{\log(\bar{b})}, \\ R &\approx \frac{\log\left(\frac{d}{10^4 j^2 u_j}\right)}{4\log(\bar{b})} + \frac{\log\left(\frac{d\varepsilon'^2}{8000 j u_j}\right)}{2\log(\bar{b})}. \end{aligned} \tag{193}$$

In [6, Theorem 4], the degree distribution of G_n is subexponential, that is, $u_j = e^{-o(j)}$. Then, the rhs of (193) is $o(j)$. In our case, the degree distribution has heavier tails than stretched-exponential with stretch-exponent ζ, that is, $u_j = e^{-o(j^\zeta)}$. Therefore, the rhs of (192) is $o(j^\zeta)$, finishing the proof. □

Proof of Theorem 2.3 (a), outline. With Lemma 8.2 at hand, the proof can be word-by-word adapted from the proof of [6, Theorem 4] with the difference that we use Claim 6.7 for the degree-penalized process, in place of [6, Lemma 6.2]. Both [6, Lemma 6.2] and our Claim 6.7 ensure that given that a star is infested, the infection reaches the next star at most $2R$ away with probability close to 1. For us, $2R = o(j^{1-2\mu})$ is necessary for Claim 6.7, hence the assumption of heavier than stretched exponential decay with exponent $1 - 2\mu$ for the degree-penalized process. In comparison, in [6], $R = o(j)$ is necessary for [6, Lemma 6.2], which leads to the assumption of subexponential tails there. We note that j depends on the infection rate λ. □

Proof of Theorem 2.8(a). This is an easy consequence of Theorem 2.3(a) by stochastic domination, noting that $\max(d_u, d_v)^\mu \le (d_u d_v)^\mu$. □

Remark 8.3. Here we highlight the difference between the expander that [6, Theorem 4] uses vs. what we describe in Lemma 8.2 and the reason for the choice of difference. In Section 6.2, we have seen the following: a star-graph of degree $j = j(\lambda)$ that survives until $\exp(cj^{1-2\mu})$ long time can transfer the infection along a path of length $o(j^{1-2\mu})$ *if* the path contains only constant degree vertices (say, at most degree m, neither depending on j nor on λ). If we would allow the path to contain vertices of any degree up to j, the penalty along the path increases and along such a path whp transmission within $\exp(cj^{1-2\mu})$ long time only happens up to distance $o(j^{1-2\mu}/\log j)$, which can be seen by adapting the

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

proof of Claim 6.7. Thus, to obtain a sharp result, in Definition 8.1, in addition to the constraints (179), (180), (182) that are all already present in [6], we have added (181), that restricts the embedded paths connecting the stars of degree j to contain only low-degree vertices of degree at most m. Without the restriction in (181), the proof in [6] word-by-word carries through for the degree-penalized CP as well, but gives a weaker result: R can only be set in the proof to $R = o(j^{1-2\mu}/\log j)$, which then, by (193), would result in the slightly stronger assumption on the degree distribution

$$\mathbb{P}(D = K) \geq \exp\{-g(K)K^{1-2\mu}/\log(K)\} \tag{194}$$

along an infinite subsequence $(K_i)_{i\geq 1}$ and with some function g such that $g(x) \to 0$ as $x \to \infty$. For limiting degree distributions satisfying (194), the proof of [6, Theorem 4] goes through for the degree-penalized version without any modifications. The modification (181) thus eliminates the extra $1/\log(K)$ factor in the tail-requirement on D in (194) so that the same assumption as for GW trees, Definition 1.8 with $\zeta = 1 - 2\mu$ is enough.

A. Appendix: Proofs of technical lemmas

A.1. Proof of the statement in Example 1.13

Assumption 1.10 is a consequence of the law of large numbers. To prove Assumptions 1.11, and 1.12, we also need to consider n-dependent values for $\nu_n(z)$ and $1 - F_n(z)$ which makes the statement nontrivial. We bound the maximum degree first, this immediately gives (12) in Assumption 1.12. Here we use that $\mathbb{P}(D \geq z) \leq 1/z^{\alpha-\varepsilon'} \leq 1/z^{\alpha(1-\varepsilon')}$ holds for all ε' to estimate that the probability that (12) fails to hold for given $n, C_u, \varepsilon_1 > 0$ is

$$\begin{aligned}\mathbb{P}\Big(\max_{i\leq n} D_{n,i} > n^{1/(\alpha(1-\varepsilon_1))}\Big) &\leq n\mathbb{P}\Big(D > n^{1/(\alpha(1-\varepsilon_1))}\Big)\\ &\leq nn^{-(\alpha(1-\varepsilon')/((\alpha(1-\varepsilon_1))} = n^{1-(1-\varepsilon')/(1-\varepsilon_1)}.\end{aligned} \tag{A.1}$$

For any fixed $\varepsilon_1 > 0$, choose $\varepsilon' := \varepsilon_1/2$ and then the exponent of n is negative. Hence, with probability tending to 1, we have $\max d_i \leq n^{1/(\alpha(1-\varepsilon_1))} =: z_{\max}(\varepsilon_1)$ for *any* fixed ε_1. We can rewrite the exponent to obtain that $\alpha = \tau - 1$. This means that $\nu_n(z)$ has discrete support on $[0, n^{1/\alpha(1-\varepsilon_1)}]$ with probability tending to 1, hence it is enough to consider $z \in \mathbb{N}$ in this range. We now recall that for any Binomial random variable with parameters n and p, and any $c > 1$,

$$\mathbb{P}(\mathrm{Bin}(n,p) \geq cnp) \leq \exp(-np(c\log c + 1 - c)) = \exp(-npc(\log c + 1/c - 1)). \tag{A.2}$$

Clearly the right-hand side is tending to 0 as long as $npc \to \infty$ and $\log c \to \infty$ both hold. We start by estimating the upper tail for Assumption 1.11, so that we prove (13). Our goal is to show that for all $z \leq n^{1/(\alpha(1-\varepsilon_1))} = z_{\max}$, for some $\varepsilon_2 > 0$ that is still arbitrarily small, whp

$$\mathbb{P}\big(\forall z \in [z_0(\varepsilon_2/2), z_{\max}(\varepsilon_1)] : 1 - F_n(z) \leq z^{-\alpha(1-\varepsilon_2)}\big) \to 1. \tag{A.3}$$

Note that $n(1 - F_n(z))$ is the number of vertices with degree above z, which has $\mathrm{Bin}(n, \mathbb{P}(D > z))$ distribution. For all $z \geq z_0(\varepsilon')$ this is stochastically dominated from above by a $\mathrm{Bin}(n, z^{-\alpha(1-\varepsilon')})$ distribution. Hence,

$$\mathbb{P}\Big(1 - F_n(z) \geq z^{-\alpha(1-\varepsilon_2)}\Big) \leq \mathbb{P}\Big(\mathrm{Bin}(n, z^{-\alpha(1-\varepsilon')}) \geq nz^{-\alpha(1-\varepsilon_2)}\Big).$$

Now we apply (A.2) with $p = z^{-\alpha(1-\varepsilon')}$ and $c = z^{-\alpha(1-\varepsilon_2)+\alpha(1-\varepsilon')} = z^{\alpha(\varepsilon_2-\varepsilon')}$. The exponent of z is positive whenever $\varepsilon_2 > \varepsilon'$ which we already we may safely assume since ε' can be chosen arbitrarily, hence $\log c \to \infty$. Further, $npc = nz^{-\alpha(1-\varepsilon_2)}$ tends to ∞ exactly when $z = o(n^{1/(\alpha(1-\varepsilon_2))})$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

which can be made always true in the range $[1, n^{1/(\alpha(1-\varepsilon_1))}]$ of the empirical distribution by choosing $\varepsilon_2 \geq \varepsilon_1 \geq \varepsilon' := \varepsilon_2/2$, but all of them arbitrarily small. At $z_{\max}(\varepsilon_1)$ the exponent in (A.2) becomes minimal and is at least constant times $npc = nz_{\max}^{-\alpha(1-\varepsilon_2)} = nn^{-(1-\varepsilon_2)/(1-\varepsilon_1)} = n^{+\delta}$. Taking a union bound over all $z \in [1, n^{1/(\alpha(1-\varepsilon_1))}]$ and the bound in (A.1) we obtain that

$$\begin{aligned} 1 - \mathbb{P}(\forall z \geq 1 : 1 - F_n(z) \leq z^{-\alpha(1-\varepsilon_2)}) &= \mathbb{P}(\exists i \leq n : D_i > z_{\max}(\varepsilon_1)) \\ &\quad + \mathbb{P}\big(\exists z \leq z_{\max}(\varepsilon_1) : 1 - F_n(z) \leq z^{-\alpha(1-\varepsilon_2)}\big) \\ &\leq n^{1-(1-\varepsilon')/(1-\varepsilon_1)} + n^{1/(\alpha(1-\varepsilon_1))} \exp(-n^{\delta}) \to 0. \end{aligned}$$

This finishes the proof of (A.3) and the upper bound in Assumption 1.11. To prove the lower bound we need the opposite direction, that is, for all $c \leq 1/2$,

$$\mathbb{P}(\mathrm{Bin}(n,p) \leq cnp) \leq \exp(-np/8), \tag{A.4}$$

as long as $np \to \infty$. We now estimate $n(1 - F_n(z)) = \mathrm{Bin}(n, \mathbb{P}(D > z))$ stochastically from below by $\mathrm{Bin}(n, z^{-(\alpha+\varepsilon')}) \geq \mathrm{Bin}(n, z^{-\alpha(1+\varepsilon')})$ which is true for all fixed ε' and all $z > z_0(\varepsilon')$, since the lower bound here is coming from (6). So let us set $z_{\max}^{(\ell)}(\varepsilon, n)$ in Assumption 1.11 to be $n^{1/(\alpha(1+\varepsilon))}$, and then the mean $n(z_{\max}^{(\ell)}(\varepsilon, n))^{-(\alpha(1+\varepsilon'))} = n^{1-(1+\varepsilon')/(1+\varepsilon)}$ tends to infinity whenever $\varepsilon' < \varepsilon$. Further, if $\varepsilon' < \varepsilon$ then also $nz^{-\alpha(1+\varepsilon)} \leq nz^{-\alpha(1+\varepsilon')}/2$ for all $z \leq z_{\max}^{(\ell)}(\varepsilon, n)$, and so (A.4) applies with $p = z^{-\alpha(1+\varepsilon')}$. By a union bound then

$$\mathbb{P}\Big(\exists z \in [z_0, z_{\max}^{(\ell)}(\varepsilon, n)] : 1 - F_n(z) \leq z^{-\alpha(1+\varepsilon)}\Big) \leq n^{1/(\alpha(1+\varepsilon))} \exp(-n^{1-(1+\varepsilon')/(1+\varepsilon)}/8),$$

which tends to 0. It remains to prove (11) in Assumption 1.12, and here we can use the extra assumption (14). Namely, following the bound on the maximum in (A.1). We want to prove that

$$\mathbb{P}\Big(\exists z \in [z_0, z_{\max}(\varepsilon_1)] : \nu_n(z) \geq z^{-\tau(1-\varepsilon)}\Big) \to 0.$$

In this case $n\nu_n(z) = n_z = \mathrm{Bin}(n, \mathbb{P}(D = z))$ which is stochastically dominated by $\mathrm{Bin}(n, z^{-\tau(1-\varepsilon')})$, and returning to (A.2), now $c = z^{\tau(\varepsilon-\varepsilon')}$ tends to infinity whenever $\varepsilon > \varepsilon'$, and now $nz^{-\tau(1-\varepsilon)}$ takes the role of npc. This tends to infinity whenever $z = o(n^{1/(\tau(1-\varepsilon))})$, which is much less than the maximum degree $z_{\max}(\varepsilon_1) = n^{1/(\tau-1)(1-\varepsilon_1)}$ for *small* $\varepsilon > 0$. Nevertheless, we can set a reasonable ε, namely, whenever we set $\varepsilon > 1/\tau$, for example, set $\varepsilon := 1/\tau + \delta$, then $nz^{-\tau(1-\varepsilon)} = nz^{-\tau(1-1/\tau-\delta)} = nz^{-(\tau-1-\delta)}$, and so for $z_{\max} = n^{1/(\tau-1)(1-\varepsilon_1)}$ this is $nn^{-(\tau-1-\delta)/(\tau-1)(1-\varepsilon_1)}$, which has a positive exponent whenever $\delta > \varepsilon_1(\tau - 1)$. Since ε_1 was arbitrarily small, δ is thus also arbitrarily small. This, together with a union bound with (A.1) finishes the proof of (15):

$$\begin{aligned} 1 - \mathbb{P}(\forall z \geq 1 : \nu_n(z) \leq z^{-\tau(1-1/\tau+\delta)}) &= \mathbb{P}(\exists i \leq n : D_i > z_{\max}(\varepsilon_1)) \\ &\quad + \mathbb{P}\big(\exists z \leq z_{\max}(\varepsilon_1) : \nu_n(z) \leq z^{-\tau(1-1/\tau+\delta)}\big) \\ &\leq n^{1-(1-\varepsilon')/(1-\varepsilon_1)} + n^{1/(\tau-1)(1-\varepsilon_1))} \exp(-n^{\delta}) \to 0. \end{aligned}$$

If one considers truncated power-law distributions with maximal degree $z_{\max,\mathrm{tr}} = o(n^{1/\tau})$, then $nz_{\max,\mathrm{tr}}^{-\tau(1-\varepsilon)} \to \infty$ for all possible values z, hence the proof above works with $\varepsilon > 0$ arbitrary.

A.2. Proof of long survival on stars

Proof of Claim 6.6. Denote the neighbors of v by $w_1, \ldots, w_K$. Define

$$\begin{aligned} \mathcal{A}_1 &= \{\xi_t^v(v) = 1 \text{ for all } t \in [0,1]\}, \\ \mathcal{A}_2 &= \big\{\big|\{i : \xi_1^v(w_i) = 1\}\big| \geq \lambda K^{1-\mu}/(4e)\big\}. \end{aligned}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Since v recovers at rate 1, $\mathbb{P}(\mathcal{A}_1 \mid \xi_0(v) = 1) = 1/e$. Conditioning on $\mathcal{A}_1$, v infects each of w_i during $[0,1]$ with rate $\lambda K^{-\mu}$, independently of each other. Hence, for $i = 1, \dots, K$,

$$\mathbb{P}(v \text{ infects } w_i \text{ at some } t \in [0,1]) = 1 - e^{-\lambda K^{-\mu}} \geq \lambda K^{-\mu}/2,$$

using that $\lambda K^{-\mu} < 1$. Each u_i that becomes infected during $[0,1]$ is still infected at time 1 with conditional probability at least $1/e$. Hence,

$$\left|\left\{i : \xi_1^v(w_i) = 1\right\}\right| \mid \mathcal{A}_1 \succcurlyeq X \sim \mathrm{Bin}(K, \lambda K^{-\mu}/(2e)),$$

where $\succcurlyeq$ stands for stochastic domination. By a standard Chernoff bound, this yields

$$\mathbb{P}(\mathcal{A}_2 \mid \mathcal{A}_1) \geq \mathbb{P}\Big(X \geq \lambda K^{1-\mu}/(4e)\Big) \geq 1 - e^{-\lambda K^{1-\mu}/(16e)}. \tag{A.5}$$

Therefore,

$$\mathbb{P}(\mathcal{A}_2) \geq \mathbb{P}(\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_2 \mid \mathcal{A}_1) \geq \Big(1 - e^{-\lambda K^{1-\mu}/(16e)}\Big)/e,$$

finishing the proof of (117) in Claim 6.6.

We now turn to proving (118) and (119). Starting from time 0, we declare each unit time-interval $[s, s+1]$ for $s \in \mathbb{N}$ *successful* if the following events jointly occur:

$$\begin{aligned}
\mathcal{B}_s^1 &= \left\{|\{i : \xi_s(w_i) = 1\}| \geq \lambda K^{1-\mu}/(8e)\right\},\\
\mathcal{B}_s^2 &= \left\{|\{i : \xi_t(w_i) = 1 \text{ for all } t \in [s, s+1]\}| \geq \lambda K^{1-\mu}/(16e^2)\right\},\\
\mathcal{B}_s^3 &= \left\{\int_s^{s+1} \xi_t(v)\,\mathrm{d}t \geq 0.55\right\},\\
\mathcal{B}_s^4 &= \left\{|\{i : \xi_{s+1}(w_i) = 1\}| \geq \lambda K^{1-\mu}/(8e)\right\}.
\end{aligned} \tag{A.6}$$

Here $\mathcal{B}_s^1$ is the event that a large enough number of leaves of the star are infected at the beginning of the time interval $[s, s+1]$, which will be enough to sustain the infestation during the whole period, while $\mathcal{B}_s^4$ is the corresponding event for the end of the time interval. $\mathcal{B}_s^2$ is the event that the star is infested during $[s, s+1]$, while $\mathcal{B}_s^3$ is the event that the center is infected a bit more than half the time during the time interval $[s, s+1]$. Note that $\mathcal{B}_s^4 = \mathcal{B}_{s+1}^1$ for all s and that $\mathcal{B}_0^1$ holds by the condition of the Lemma.

One can see that if $\mathcal{B}_s^1 \cap \mathcal{B}_s^2 \cap \mathcal{B}_s^3 \cap \mathcal{B}_s^4$ holds for all $s \in \{0, 1, \dots, \lfloor T_K \rfloor\}$, given that $|\xi_0| \geq \lambda K^{1-\mu}/8\mathrm{e}$, then the event on the left-hand side of (119) holds (we demand in $\mathcal{B}_s^3$ a little bit more than $1/2$ of the time being infected, so that even if v is healthy during $[\lfloor T_K \rfloor, T_K$, this does not cause a problem on the total infected time being above $T_K/2$. Further, (118) is a direct consequence of (119), so it is enough to bound the probability of the intersection of these events.

We now fix some $s \in \mathbb{N}$ and bound the conditional probabilities of each of these events given the previous ones. First, any leaf of the star that is infected at time s will stay infected during the whole interval $[s, s+1]$ with conditional probability at least $1/e$, conditioned on any trajectory of the process on the other vertices. Formally, for any $i = 1, \dots, K$,

$$\inf_{\eta} \mathbb{P}\left(\begin{array}{l}\xi_t(w_i) = 1 \text{ for all } t \in [s, s+1] \mid \xi_s(w_i) = 1,\\ \xi \equiv \eta \text{ on } [s, s+1] \text{ on all vertices apart from } u_i\end{array}\right) \geq 1/e.$$

Hence, by a Chernoff bound similar to (A.5),

$$\mathbb{P}(\mathcal{B}_s^2 \mid \mathcal{B}_s^1) \geq 1 - e^{-\lambda K^{1-\mu}/(64e^2)}. \tag{A.7}$$

We will now give a bound on $\mathbb{P}(\mathcal{B}_s^3 \mid \mathcal{B}_s^1 \cap \mathcal{B}_s^2)$, using that an infested star has enough leaves infected at every time to send back the infection to the center frequently enough to keep it infected for at least half

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

of the time. Formally, given $\mathcal{B}_s^1 \cap \mathcal{B}_s^2$, $(\xi_t(v))_{t\in[s,s+1]}$ is a Markov process on the state space $\{0,1\}$ with transition rates

$$Q_{01} \geq \lambda K^{1-\mu} \cdot \lambda K^{-\mu}/(16e^2) = \lambda^2 K^{1-2\mu}/(16e^2), \qquad Q_{10} = 1,$$

and some starting state $\xi_s(v)$. Let us introduce auxiliary Markov processes $(Y_t)_{t\geq 0}$, $(Y'_t)_{t\geq 0}$ and $(Y''_t)_{t\geq 0}$ on $\{0,1\}$, all starting from the same initial state $\xi_s(v)$, with transition rates

$$\begin{array}{lll} q_{01} = \lambda^2 K^{1-2\mu}/(16e^2), & q_{10} = 1 & \text{of } Y, \\ q'_{01} = 1, & q'_{10} = 16e^2/(\lambda^2 K^{1-2\mu}) & \text{of } Y', \\ q''_{01} = 1, & q''_{10} = 1/2 & \text{of } Y'', \end{array}$$

respectively. Note that Y' is a time-changed (slowed-down) version of Y, and Y'' is stochastically dominated by Y' when $(16e^2)/(\lambda^2 K^{1-2\mu}) < 1/2$. Then, recalling (A.6), we have

$$\begin{aligned} \mathbb{P}(\mathcal{B}_s^3 \mid \mathcal{B}_s^1 \cap \mathcal{B}_s^2) &\geq \mathbb{P}\left(\int_0^1 Y_t \,\mathrm{d}t \geq 0.55\right) = \mathbb{P}\left(\frac{16e^2}{\lambda^2 K^{1-2\mu}} \int_0^{\frac{\lambda^2 K^{1-2\mu}}{16e^2}} Y'_t \,\mathrm{d}t \geq 0.55\right) \\ &\geq \mathbb{P}\left(\frac{16e^2}{\lambda^2 K^{1-2\mu}} \int_0^{\frac{\lambda^2 K^{1-2\mu}}{16e^2}} Y''_t \,\mathrm{d}t \geq 0.55\right). \end{aligned} \tag{A.8}$$

Note that the stationary distribution of Y'' is $(\pi_0, \pi_1) = (1/3, 2/3)$. The large deviation principle for Markov chains (see for example [22]) yields that the time average of Y''_t on the right-hand side of (A.8) is close to π_1 with large probability, as $K \to \infty$:

$$\mathbb{P}\left(\frac{16e^2}{\lambda^2 K^{1-2\mu}} \int_0^{\frac{\lambda^2 K^{1-2\mu}}{16e^2}} Y''_t \,\mathrm{d}t \geq 0.55\right) \geq 1 - \exp\{-c\lambda^2 K^{1-2\mu}\}. \tag{A.9}$$

Combining (A.8) and (A.9) gives

$$\mathbb{P}(\mathcal{B}_s^3 \mid \mathcal{B}_s^1 \cap \mathcal{B}_s^2) \geq 1 - \exp\{-c\lambda^2 K^{1-2\mu}\} \tag{A.10}$$

for some $c > 0$.

Given $\mathcal{B}_s^3$, during $[s, s+1]$, v spends at least $1/2$ time in total in state 1, during which it infects all the leaves with rate $\lambda K^{1-\mu}$. Each leaf infected this way is still infected at $s+1$ with conditional probability at least $1/e$. Hence, for $\mathcal{B}_s^4$ given by (A.6), another Chernoff bound, similar to (A.5), yields

$$\mathbb{P}(\mathcal{B}_s^4 \mid \mathcal{B}_s^1 \cap \mathcal{B}_s^2 \cap \mathcal{B}_s^3) \geq 1 - e^{-\lambda K^{1-\mu}/(32e)}. \tag{A.11}$$

Combining (A.7), (A.10) and (A.11) yields

$$\mathbb{P}(\mathcal{B}_s^1 \cap \mathcal{B}_s^2 \cap \mathcal{B}_s^3 \cap \mathcal{B}_s^4 \mid \mathcal{B}_s^1) \geq 1 - \exp\{-c'\lambda^2 K^{1-2\mu}\} \tag{A.12}$$

for some $c' > 0$. In words, (A.12) states the following: given that a large enough number of leaves (at least $\lambda K^{1-\mu}/(8e)$) are infected at time s, the conditional probability that the time-interval $[s, s+1]$ will be successful (in the sense discussed around (A.6)) is at least $1 - \exp\{-c'\lambda^2 K^{1-2\mu}\}$. The fact that the time-interval $[s, s+1]$ is successful includes the event $\mathcal{B}_s^4 = \mathcal{B}_{s+1}^1$, that is, that a large enough number of leaves are infected at time $s+1$ as well. Hence, using (A.12) iteratively (for $s = 0, 1, \ldots$) shows that, given $\mathcal{B}_0^1 = \{|\underline{\xi}_0| \geq \lambda K^{1-\mu}/(8e)\}$, the number of consecutive successful time-intervals $[0,1], [1,2], \ldots$ stochastically dominates a Geometric random variable with parameter $\exp\{-c'\lambda^2 K^{1-2\mu}\}$. Consequently (using a union bound with (A.12)), there exists a constant $c_1 > 0$ such that, with $T_K := \exp(c_1 \lambda^2 K^{1-2\mu})$,

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

$$\mathbb{P}\left([0,1],[1,2],\dots,[\lfloor T_K\rfloor,\lceil T_K\rceil] \text{ are all successful time-intervals} \,\middle|\, |\underline{\xi}_0| \geq \lambda K^{1-\mu}/(8e)\right)$$
$$\geq 1-\exp(-c_1\lambda^2 K^{1-2\mu}). \tag{A.13}$$

Recalling how successful time intervals are defined in terms of the events in (A.6), (A.13) implies both (118) and (119) in the claim. □

A.3. Proofs about the degrees in the configuration model

Proof of Lemma 7.5. We will consider $\widetilde{n}_i := \sum_{j=1}^n \mathbb{1}_{\{\widetilde{d}_j=i, d_j\leq M\}}$ which gives the number of vertices with degree i in $G_n[\mathcal{V}_{\leq M}]$. Then the (random) empirical mass function of $\widetilde{F}_{n,M}$ can be written as

$$X_{n,M}(i) = \mathbb{P}(\widetilde{D}_{n,M}=i \mid G_n[\mathcal{V}_{\leq M}]) = \frac{\widetilde{n}_i}{V_{\leq M}} = \left(\frac{V_{\leq M}}{n}\right)^{-1}\cdot\frac{\widetilde{n}_i}{n}. \tag{A.14}$$

We can now analyze both factors on the rhs separately. The first factor is already given by (141), and can be exactly described using D_n with cdf in (8)

$$\frac{V_{\leq M}}{n} = \frac{nF_n(M)}{n} = F_n(M) = \mathbb{P}(D_n\leq M). \tag{A.15}$$

By the definition of δ_n in (147), this falls in the range $\mathbb{P}(D\leq M)\pm M\delta_n$. Turning to the second factor $\widetilde{n}_i/n$ in (A.14), we introduce $n_\ell := |\mathcal{V}_\ell| = \sum_{j=1}^n \mathbb{1}_{\{d_j=\ell\}}$ the number of degree-ℓ vertices in the original graph. Then we can carry out a first and second moment method, that is, we take expectation over the realization of the matching and hence the graph $G_n[\mathcal{V}_{\leq M}]$. We start with the first moment:

$$\frac{1}{n}\mathbb{E}[\widetilde{n}_i] = \frac{1}{n}\sum_{\ell=i}^M\sum_{v\in\mathcal{V}_\ell}\mathbb{E}[\mathbb{1}_{\{\widetilde{d}_v=i\}}\mid d_v=\ell] = \frac{1}{n}\sum_{\ell=i}^M\sum_{v\in\mathcal{V}_\ell}\mathbb{P}(\widetilde{d}_v=i\mid d_v=\ell). \tag{A.16}$$

To analyze $\mathbb{P}(\widetilde{d}_v=i\mid d_v=\ell)$, we first deal with self-loops at $v\in\mathcal{V}_\ell$. Labeling the half-edges of v as $h_1,h_2,\dots,h_\ell$, the number of self-loops at v is $S_v := \sum_{1\leq s,t\leq\ell}\mathbb{1}_{\{h_s\leftrightarrow h_t\}}$, with $\leftrightarrow$ standing for the event that the two half-edges are matched to each other. We denote the total number of half-edges in the graph by $H_n=\mathbb{E}[D_n]n$, and then a first moment method yields, as $\ell\leq M$,

$$\mathbb{P}(S_v\geq 1)\leq\mathbb{E}[S_v] = \binom{\ell}{2}\frac{1}{H_n-1}\leq\frac{M^2}{\mathbb{E}[D_n]n}. \tag{A.17}$$

Recall from (141) in Definition 7.2 that we denote by $H_{\leq M}$ and $H_{>M}$ the number of half-edges attached to vertices of degree at most M and larger than M, respectively. Partition now the ℓ half-edges of v into (arbitrary) two groups of size i and $\ell-i$, respectively: $h_{s_1},\dots,h_{s_i}$ and $h_{s_{i+1}},\dots,h_{s_\ell}$, and let us write informally

$$\mathcal{A}_{\{s_1,\dots,s_i\}} := \left\{\{h_{s_1},\dots,h_{s_i}\}\leftrightarrow\mathcal{V}_{\leq M}, \{h_{s_{i+1}},\dots,h_{s_\ell}\}\leftrightarrow\mathcal{V}_{>M}, S_v=0\right\}$$

for the event that the half-edges $h_{s_1},\dots,h_{s_i}$ are all matched to half-edges belonging to vertices in $\mathcal{V}_{\leq M}$, the half-edges $h_{s_{i+1}},\dots,h_{s_\ell}$ are all matched to half-edges belonging to vertices in $\mathcal{V}_{>M}$, and there is no self-loop created among $h_{s_1},\dots,h_{s_i}$. Then, matching half-edges one by one, we come to

$$\mathbb{P}(\mathcal{A}_{\{s_1,\dots,s_i\}}) = \prod_{a=0}^{i-1}\frac{H_{\leq M}-\ell-a}{H_n-2a-1}\cdot\prod_{b=0}^{\ell-i-1}\frac{H_{>M}-b}{H_n-2(i+b)-1}.$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Observe that per definition $H_{\le M} = n\mathbb{E}[D_n \mathbb{1}_{\{D_n \le M\}}]$ and $H_n = n\mathbb{E}[D_n]$ so one can compute, using also that $\ell \le M$, that each factor in the first product is $q_{n,M}(1 + O(M/n))$ and each factor in the second product is $(1 - q_{n,M})(1 + O(M/n))$. Considering all the possible partitions of the half-edges into two groups of i and $\ell - i$ half-edges, and using that there are $\ell \le M$ factors in the two products together, we arrive at

$$\begin{aligned}\mathbb{P}(\widetilde{d}_v = i \mid d_v = \ell) &\ge \sum_{\{s_1,\dots,s_i\}\subset[\ell]} \mathbb{P}\big(\mathcal{A}_{\{s_1,\dots,s_i\}}\big) \\ &= \binom{\ell}{i} q_{n,M}^i (1 - q_{n,M})^{\ell - i}\big(1 + O\big(\tfrac{M^2}{n}\big)\big).\end{aligned} \tag{A.18}$$

A similar upper bound holds: we account for the error caused by the event that there might be self-loops at v in (A.17),

$$\begin{aligned}\mathbb{P}(\widetilde{d}_v = i \mid d_v = \ell) &\le \mathbb{P}(S_v \ge 1) + \sum_{\{s_1,\dots,s_i\}\subset[\ell]} \mathbb{P}\big(\mathcal{A}_{\{s_1,\dots,s_i\}}\big) \\ &= O\big(\tfrac{M^2}{n}\big) + \binom{\ell}{i} q_{n,M}^i (1 - q_{n,M})^{\ell - i}\big(1 + O\big(\tfrac{M^2}{n}\big)\big).\end{aligned} \tag{A.19}$$

Using these bounds in (A.16), and that $|\mathcal{V}_\ell|/n = \mathbb{P}(D_n = \ell)$, we arrive at

$$\begin{aligned}\frac{1}{n}\mathbb{E}[\widetilde{n}_i] &= \sum_{\ell=i}^{M} \mathbb{P}(D_n = \ell)\left(O\big(\tfrac{M^2}{n}\big) + \binom{\ell}{i} q_{n,M}^i (1 - q_{n,M})^{\ell - i}\big(1 + O\big(\tfrac{M^2}{n}\big)\big)\right) \\ &= O\big(\tfrac{M^2}{n}\big)\mathbb{P}(D_n \le M) + \big(1 + O\big(\tfrac{M^2}{n}\big)\big) \sum_{\ell=i}^{M} \mathbb{P}(D_n = \ell)\binom{\ell}{i} q_{n,M}^i (1 - q_{n,M})^{\ell - i}.\end{aligned}$$

Combining this with (A.14) and (A.15), (recalling also that given $(d_1, \dots d_n)$, $V_{<M}$ is deterministic), we obtain that

$$\begin{aligned}\mathbb{E}[X_{n,M}(i)] &= \frac{1}{V_{\le M}}\mathbb{E}[\widetilde{n}_i] \\ &= O\big(\tfrac{M^2}{n}\big) + \big(1 + O\big(\tfrac{M^2}{n}\big)\big) \sum_{\ell=i}^{M} \frac{\mathbb{P}(D_n = \ell)}{\mathbb{P}(D_n \le M)}\binom{\ell}{i} q_{n,M}^i (1 - q_{n,M})^{\ell - i}.\end{aligned} \tag{A.20}$$

We can here observe that the rhs gives the probability $\mathbb{P}(\mathrm{Bin}(D_n, q_{n,M}) = i \mid D_n \le M)$. Since $\mathbb{P}(D_n \le M) \to \mathbb{P}(D \le M)$ and $q_{n,M} \to q_M$ by Assumption 1.10, the rhs of (A.20) tends to

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}[X_{n,M}(i)] &= \sum_{\ell=i}^{M} \frac{\mathbb{P}(D = \ell)}{\mathbb{P}(D \le M)}\binom{\ell}{i} q_M^i (1 - q_M)^{\ell - i} \\ &= \mathbb{P}(\mathrm{Bin}(D, q_M) = i \mid D \le M) =: \mathbb{P}(\widetilde{D}_M = i) =: p_M(i),\end{aligned} \tag{A.21}$$

where we recognised that the formula on the right hand-side of the first row equals the second row, which ensures that this is a proper random variable. As $p_M(i)$ is $\lim_{n\to\infty} \mathbb{E}[X_{n,M}(i)]$, we now set out to prove that the *random* empirical distribution $\widetilde{F}_{n,M}$ – with pointmasses $X_{n,M}(i)$ at i – converges pointwise for each $i \le M$ to $p_M(i)$, in probability. Then (148) will be the limit random variable in the lemma statement.

To achieve this, first we bound the difference between $\mathbb{E}[X_{n,m}(i)]$ and its limit $p_M(i)$. Recalling the definition of δ_n from (147) we may write $q_{n,M} = q_M(1 \pm \delta_n)$, $1 - q_{n,M} = (1 - q_M)(1 \pm \delta_n)$ and similarly we can use that $\mathbb{P}(D_n = \ell) = \mathbb{P}(D = \ell)(1 \pm \delta_n)$ when the limit $\mathbb{P}(D = \ell)$ is nonzero and

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

otherwise $\mathbb{P}(D_n = \ell) \le \delta_n$. So, we subtract (A.20) from the right-hand side of the first row in (A.21) to obtain after elementary error-estimates that

$$\begin{aligned}\big|\mathbb{E}[X_{n,M}(i)] - p_M(i)\big| &\le O\big(\tfrac{M^2}{n}\big) + O\big(\tfrac{M^2}{n} + M^2\delta_n\big)\mathbb{P}(\widetilde{D}_M = i)\\ &= O\big(\tfrac{M^2}{n} + \delta_n M^2\big).\end{aligned} \tag{A.22}$$

This finishes comparing the first moments. We now turn to the variance in (A.14). Clearly

$$\mathrm{Var}\Big(X_{n,M}(i)\Big) = \mathrm{Var}\Big(\frac{\widetilde{n}_i}{V_{\le M}}\Big) = \Big(\frac{V_{\le M}}{n}\Big)^{-2} \cdot \frac{\mathrm{Var}(\widetilde{n}_i)}{n^2}. \tag{A.23}$$

The first factor on the rhs is $\mathbb{P}(D_n \le M)^{-2}$. Using the indicator representation of $\widetilde{n}_i$, we compute using the covariance formula that

$$\begin{aligned}\frac{\mathrm{Var}(\widetilde{n}_i)}{n^2} = \frac{1}{n^2}\sum_{\ell,\ell'=i}^{M}\sum_{\substack{v\in\mathcal{V}_\ell,\\ u\in\mathcal{V}_{\ell'}}}\Big(&\mathbb{P}\big(\widetilde{d}_v = i, \widetilde{d}_u = i \mid d_v = \ell, d_u = \ell'\big)\\ &- \mathbb{P}\big(\widetilde{d}_v = i, \mid d_v = \ell\big)\mathbb{P}\big(\widetilde{d}_u = i \mid d_u = \ell'\big)\Big).\end{aligned} \tag{A.24}$$

When $u = v$, the two vertices are the same, the (co)variance is at most 1, and the summation contains only at most n terms, hence the error coming from coinciding u, v is $O(1/n)\mathbb{P}(D_n \le M)$ when summed also over $\ell = \ell'$. Now we treat the case when $u \ne v$. For $\mathbb{P}\big(\widetilde{d}_v = i \mid d_v = \ell\big)$ and $\mathbb{P}\big(\widetilde{d}_u = i \mid d_u = \ell'\big)$ we can use the bounds in (A.18) and (A.19). Similarly to there, we compute the first term $\mathbb{P}\big(\widetilde{d}_v = i, \widetilde{d}_u = i \mid d_v = \ell, d_u = \ell'\big)$ as well. Let $S_{u,v}$ denote the number of self-loops at the two vertices u, v together plus the number of edges between u and v. Then a first moment method yields

$$\mathbb{P}(S_{u,v} \ge 1) \le \mathbb{E}[S_{u,v}] = \frac{1}{H_n - 1}\left(\binom{\ell}{2} + \binom{\ell'}{2} + \ell\ell'\right) \le \frac{2M^2}{\mathbb{E}[D_n]n}. \tag{A.25}$$

Now we label the half-edges $h_1^{(v)}, \dots, h_\ell^{(v)}$ and $h_1^{(u)}, \dots, h_{\ell'}^{(u)}$ of v and u, respectively, and partition them into two subsets each, defined by the index sets $\{s_1, \dots, s_i\}, \{s_{i+1}, \dots, s_\ell\} \subset [\ell]$ and $\{t_1, \dots, t_i\}, \{t_{i+1}, \dots, t_{\ell'}\} \subset [\ell']$. We introduce the event

$$\begin{aligned}\mathcal{A}_{\{s_1,\dots,s_i,t_1,\dots,t_i\}} := \Big\{&\{h_{s_1}^{(v)}, \dots, h_{s_i}^{(v)}, h_{t_1}^{(u)}, \dots, h_{t_i}^{(u)}\} \leftrightarrow \mathcal{V}_{\le M},\\ &\{h_{s_{i+1}}^{(v)}, \dots, h_{s_\ell}^{(v)}, h_{t_{i+1}}^{(u)}, \dots, h_{t_{\ell'}}^{(u)}\} \leftrightarrow \mathcal{V}_{>M}, S_{u,v} = 0\Big\},\end{aligned}$$

the event that the half-edges $h_{s_1}^{(v)}, \dots, h_{s_i}^{(v)}, h_{t_1}^{(u)}, \dots, h_{t_i}^{(u)}$ are all matched to half-edges belonging to vertices in $\mathcal{V}_{\le M}$, the half-edges $h_{s_{i+1}}^{(v)}, \dots, h_{s_\ell}^{(v)}, h_{t_{i+1}}^{(u)}, \dots, h_{t_{\ell'}}^{(u)}$ are all matched to half-edges belonging to vertices in $\mathcal{V}_{>M}$, and there is no self-loop and edge created at and between u and v. Then

$$\mathbb{P}(\mathcal{A}_{\{s_1,\dots,s_i,t_1,\dots,t_i\}}) = \prod_{a=0}^{2i-1}\frac{H_{\le M} - \ell - \ell' - a}{H_n - 2a - 1} \cdot \prod_{b=0}^{\ell+\ell'-2i-1}\frac{H_{>M} - b}{H_n - 2(2i+b) - 1}.$$

Using that $\ell, \ell' \le M$, one can compute that each factor in the first product is $q_{n,M}(1 + O(M/n))$ and each factor in the second product is $(1 - q_{n,M})(1 + O(M/n))$. Hence, similarly to (A.18) and (A.19), summing over all possible partitions, we obtain the lower bound

$$\begin{aligned}&\mathbb{P}\big(\widetilde{d}_v = i, \widetilde{d}_u = i \mid d_v = \ell, d_u = \ell'\big)\\ &\quad\ge \binom{\ell}{i}q_{n,M}^i(1 - q_{n,M})^{\ell-i}\binom{\ell'}{i}q_{n,M}^i(1 - q_{n,M})^{\ell'-i}\big(1 + O(\tfrac{M^2}{n})\big).\end{aligned}$$

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

and the upper bound is the same as the rhs with an additive $O\big(\frac{M^2}{n}\big)$ coming from (A.25). We see that this is the same bound as the one in (A.19), multiplied together for u and v. Hence, returning to (A.24), when we take the difference of the two terms, the summand 1 in the $(1+O\big(\frac{M^2}{n}\big))$ factor cancels, and each summand can be upper bounded as

$$\begin{aligned}\Big|\mathbb{P}\big(\widetilde{d}_v = i, \widetilde{d}_u = i \mid d_v = \ell, d_u = \ell'\big) - \mathbb{P}\big(\widetilde{d}_v = i, \mid d_v = \ell\big)\mathbb{P}\big(\widetilde{d}_u = i \mid d_u = \ell'\big)\Big| \\ \leq O\big(\tfrac{M^2}{n}\big) + O\big(\tfrac{M^2}{n}\big)\binom{\ell}{i} q_{n,M}^i (1-q_{n,M})^{\ell-i}\binom{\ell'}{i} q_{n,M}^i (1-q_{n,M})^{\ell'-i}.\end{aligned}$$

We account for the $O(1/n)\mathbb{P}(D_n \leq M)$ error coming from $u = v$, and use $|\mathcal{V}_\ell|/n = \mathbb{P}(D_n = \ell)$ and $|\mathcal{V}_{\ell'}|/n = \mathbb{P}(D_n = \ell')$, then we obtain in (A.24) that

$$\begin{aligned}\frac{\mathrm{Var}(\widetilde{n}_i)}{n^2} &\leq O(\tfrac{1}{n})\mathbb{P}(D_n \leq M) + O\big(\tfrac{M^2}{n}\big) \sum_{\ell,\ell'=i}^{M} \mathbb{P}(D_n = \ell)\mathbb{P}(D_n = \ell') \\ &\quad \cdot \left(1 + \binom{\ell}{i} q_{n,M}^i (1-q_{n,M})^{\ell-i}\binom{\ell'}{i} q_{n,M}^i (1-q_{n,M})^{\ell'-i}\right) \\ &\leq O\big(\tfrac{M^2}{n}\big)\Big(\mathbb{P}(D_n \leq M) + 2\mathbb{P}(D_n \leq M)^2\Big),\end{aligned}$$

where the last row is a far-from-sharp upper bound.

Wlog we may assume M is large enough for $\mathbb{P}(D_n \leq M) \geq 1/2$ to hold. Using the previous inequality in (A.23), and that the first factor there is $1/\mathbb{P}(D_n \leq M)^2$, we come to

$$\mathrm{Var}\Big(X_{n,M}(i)\Big) \leq O\big(\tfrac{M^2}{n}\big)\Big(2 + 1/\mathbb{P}(D_n \leq M)\Big) = O\big(\tfrac{M^2}{n}\big), \tag{A.26}$$

which is true *uniformly* in i, that is, the factor $O(M^2/n)$ is not depending on i.

We now are ready to prove the bound in (149) in Lemma 7.5. Recall that $X_{n,M}(i) := \widetilde{n}_i/V_{\leq m} = \mathbb{P}(\widetilde{D}_{n,M} = i \mid G_n[\mathcal{V}_{\leq M}])$. We can replace the supremum on the left-hand side of (149) by a "there exists," followed by a union bound and a triangle inequality:

$$\begin{aligned}&\mathbb{P}\Big(\sup_{i\leq M}\big|X_{n,M}(i) - p_M(i)\big| \geq \varepsilon_n\Big) \leq \sum_{i\leq M}\mathbb{P}\Big(\big|X_{n,M}(i) - p_M(i)\big| \geq \varepsilon_n\Big) \\ &\sum_{i\leq M}\left(\mathbb{P}\Big(\big|X_{n,M}(i) - \mathbb{E}[X_{n,M}(i)]\big| \geq \varepsilon_n/2\Big) + \mathbb{P}\Big(\big|\mathbb{E}[X_{n,M}(i)] - p_M(i)\big| \geq \varepsilon_n/2\Big)\right).\end{aligned} \tag{A.27}$$

The second probability in the second row is either 0 or 1 as it involves only deterministic quantities. Recall that we computed $\mathbb{E}[X_{n,M}(i)]$ in (A.20), and $\lim_{n\to\infty}\mathbb{E}[X_{n,M}(i)] = \mathbb{P}(\widetilde{D}_M = i) := p_M(i)$ in (A.21) per our definition of $p_M(i)$ in (A.21) and (148). By (A.22), $|\mathbb{E}[X_{n,M}(i)] - p_M(i)| \leq O(M^2/n + M^2\delta_n)$. Thus, whenever $\varepsilon_n \gg O(M^2/n + M^2\delta_n)$ (that we assumed in Lemma 7.5), the second probability on the right-hand side is 0 simultaneously for all $i \leq M$, for all n is sufficiently large. On each term in the first sum we can apply Chebyshev's inequality, and use the bound on the variance of $X_{n,M}(i)$ in (A.26) as follows:

$$\begin{aligned}\mathbb{P}\Big(\sup_{i\leq M}\big|X_{n,M}(i) - p_M(i)\big| \geq \varepsilon_n\Big) &\leq \sum_{i\leq M}\mathbb{P}\Big(\big|X_{n,M}(i) - \mathbb{E}[X_{n,M}(i)]\big| \geq \varepsilon_n/2\Big) \\ &\leq \sum_{i\leq M} 4\varepsilon_n^{-2}\mathrm{Var}(X_{n,M}(i)) = O\Big(\frac{M^3}{n\varepsilon_n^2}\Big),\end{aligned} \tag{A.28}$$

where we summed over $i \leq M$ to obtain the last bound. The rhs tends to zero as $n \to \infty$ by the assumption that $\varepsilon_n \gg 1/\sqrt{n}$ implying $n\varepsilon_n^2 \to \infty$. This finishes the proof of (149).

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

Recall that $\widetilde{D}_{n,M}$ denotes a random variable with random mass function $X_{n,M}(i)$. We compute the limit of the (random) mean of the empirical distribution $\widetilde{F}_{n,M}$ of $G_n[\mathcal{V}_{\le M}]$ using (A.28)

$$\mathbb{E}\big[\widetilde{D}_{n,M} \mid G_n[\mathcal{V}_{\le M}]\big] = \sum_{i=1}^{M} iX_{n,M}(i) \xrightarrow{\mathbb{P}} \sum_{i=1}^{M} ip_M(i) = \mathbb{E}\big[\widetilde{D}_M\big].$$

Finally, the fact that $G_n[\mathcal{V}_{\le M}]$ is a configuration model, conditioned on its vertices and their degrees, follows from the fact that every matching of its half-edges have equal probability under the law of the configuration model G_n. This finishes the proof of Lemma 7.5. □

Proof of Lemma 7.6. We analyze now the limiting distribution in (148) under the assumption that the original empirical distribution sequence $(F_n)_{n\ge1}$ satisfies both Assumptions 1.10 and 1.11. In particular, Assumption 1.11 implies that the cdf of the limiting distribution F_D of D_n satisfies (9) for all $\varepsilon > 0$ such that for all $n \ge n_0(\varepsilon)$, and for all $z \ge z_0$ that

$$\frac{c_\ell}{z^{(\tau-1)(1+\varepsilon)}} \le 1 - F_D(z) \le \frac{c_u}{z^{(\tau-1)(1-\varepsilon)}}. \tag{A.29}$$

and $\mathbb{E}[D] < \infty$ by assumption. We observe first that $\widetilde{D}_M$ in (148) is a binomial thinning of $(D|D \le M)$, hence $\widetilde{D}_M$ is stochastically dominated from above by $(D|D \le M)$. So, by the definition of stochastic domination,

$$1 - \widetilde{F}_M(z) = \mathbb{P}(\widetilde{D}_M > z) \le \mathbb{P}(D > z \mid D \le M) = \frac{(1 - F_D(z)) - (1 - F_D(M))}{F_D(M)}. \tag{A.30}$$

Using now (A.29), estimating the numerator from above and the denominator from below, assuming that M is such that $c_u M^{-(\tau-1)/2} \le 1/2$, for all $\varepsilon \in (0, (\tau-1)/2]$ it holds for all $z \in [z_0, M]$ that

$$1 - \widetilde{F}_M(z) \le \frac{c_u z^{-(\tau-1)(1-\varepsilon)}}{1 - c_u M^{-(\tau-1)(1-\varepsilon)}} \le \frac{c_u z^{-(\tau-1)(1-\varepsilon)}}{1 - c_u M^{-(\tau-1)/2}} \le 2c_u z^{-(\tau-1)(1-\varepsilon)} \tag{A.31}$$

which finishes the proof of the upper bound in (151) with $\widetilde{c}_u = 2c_u$. For the lower bound in (151) we will also need a lower bound on $\mathbb{P}(D > z \mid D \le M)$. Using the rhs of (A.30), estimating the denominator by at most 1, and the numerator using (A.29), we obtain

$$\begin{aligned}\mathbb{P}(D > z \mid D \le M) &\ge c_\ell z^{(\tau-1)(1+\varepsilon)} - c_u M^{-(\tau-1)(1-\varepsilon)} \\ &= c_\ell z^{-(\tau-1)(1+\varepsilon)}\big(1 - (z^{(\tau-1)(1+\varepsilon)}/M^{(\tau-1)(1-\varepsilon)}) \cdot (c_u/c_\ell)\big).\end{aligned} \tag{A.32}$$

Here we require that the second factor is at least, say, $1/2$, which leads to

$$\begin{aligned}&\mathbb{P}(D > z \mid D \le M) \ge (c_\ell/2) z^{-(\tau-1)(1+\varepsilon)} \\ &\qquad \text{for all} \quad z \le (c_\ell/(2c_u))^{\frac{1}{(\tau-1)(1+\varepsilon)}} M^{1-\frac{2\varepsilon}{1+\varepsilon}} =: \widetilde{z}'_{\max}(M).\end{aligned} \tag{A.33}$$

Observe that even without considering the binomial thinning in (148), one cannot hope to prove a lower bound for z too close to M. Nevertheless, $\widetilde{z}'_{\max}(M)$ is growing with M for all $\varepsilon < 1$, and it gets closer to $\Theta(M)$ as ε is smaller, which intuitively means that the sharper bound one has on the tail of D, the

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

sharper bound we can also get on probabilities of D falling in given intervals. Nevertheless, even for $\varepsilon = 0$, we must require $z \le c_2 M$ for some constant $c_2 \le 1$.

Now we compute the thinning probability in (148):

$$\begin{aligned} 1 - q_M &= \mathbb{E}[D]^{-1}\mathbb{E}[D\mathbb{1}_{\{D>M\}}] = \mathbb{E}[D]^{-1}\Big(M\mathbb{P}(D > M) + \sum_{j=M}^{\infty} \mathbb{P}(D > j)\Big) \\ &\le \mathbb{E}[D]^{-1}(c_u M^{1-(\tau-1)(1-\varepsilon)} + \sum_{j=M}^{\infty} c_u j^{-(\tau-1)(1-\varepsilon)}) \le c_{u,1} M^{1-(\tau-1)(1-\varepsilon)}, \end{aligned} \tag{A.34}$$

for some constant $c_{u,1} > 0$ (that does not depend on M) and a similar lower bound holds $1 - q_M \ge c_{\ell,1} M^{1-(\tau-1)(1+\varepsilon)}$. Then, using the Binomial representation in (148), and then stochastic domination of $\mathrm{Bin}(j, q)$ by $\mathrm{Bin}(j^\star, q)$ when $j \le j^\star$, we obtain that for all $j^\star \ge z$,

$$\begin{aligned} \mathbb{P}(\widetilde{D}_M \ge z) &= \sum_{j=z}^{M} \frac{\mathbb{P}(D = j)}{\mathbb{P}(D \le M)} \mathbb{P}\big(\mathrm{Bin}(j, q_M) \ge z\big) \\ &\ge \mathbb{P}\big(\mathrm{Bin}(j^\star, q_M) \ge z\big)\mathbb{P}\big(D \ge j^\star \mid D \le M\big), \end{aligned} \tag{A.35}$$

and we can optimize the value $j^\star = j^\star(z) \ge z$ to obtain a sharp enough bound. For the second factor on the rhs we may use (A.33). Moving to the "complement" binomial, we estimate the first factor in (A.35) as

$$\begin{aligned} \mathbb{P}(\mathrm{Bin}(j^\star, q_M) \ge z) &= \mathbb{P}(\mathrm{Bin}(j^\star, 1 - q_M) \le j^\star - z) \\ &= 1 - \mathbb{P}(\mathrm{Bin}(j^\star, 1 - q_M) > j^\star - z). \end{aligned} \tag{A.36}$$

We observe that the thinning probability $1 - q_M$ in (A.34) tends to zero with M. So, when $z \le \widetilde{z}'_{\max}/2$, we may take $j^\star(z) := 2z$ and use Markov's inequality on the rhs in (A.36) to obtain

$$1 - \mathbb{P}\big(\mathrm{Bin}(2z, 1 - q_M) > z\big) \ge 1 - \frac{2z(1 - q_M)}{z} = 1 - 2(1 - q_M) \ge 1/2$$

for all M large enough so that $c_{u,1} M^{1-(\tau-1)(1-\varepsilon)} < 1/4$. Using this bound in (A.35), along with (A.33), we obtain for all $z < \widetilde{z}_{\max}/2$ that

$$\mathbb{P}(\widetilde{D}_M \ge z) \ge \mathbb{P}\big(D \ge 2z \mid D \le M\big)/2 \ge (c_\ell/4)(2z)^{-(\tau-1)(1+\varepsilon)},$$

which finishes the proof by choosing

$$\widetilde{c}_\ell := 2^{-(\tau-1)(1+\varepsilon)-2} c_\ell \quad \text{and} \quad \widetilde{z}_{\max}(M) := 2^{-1}(c_\ell/(2c_u))^{\frac{1}{(\tau-1)(1+\varepsilon)}} M^{1-\frac{2\varepsilon}{1+\varepsilon}}. \qquad \square$$

Competing interests. The authors have no competing interest to declare.

Financial support. ZB was supported by the ERC Consolidator Grant 772466 "NOISE."

Ethical standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

[1] D. Aldous and J. M. Steele, 'The objective method: probabilistic combinatorial optimization and local weak convergence', in *Probability on Discrete Structures* (Springer, 2004), pp. 1–72.

[2] R. F. S. Andrade, J. S. AndradeJr and H. J. Herrmann, 'Ising model on the apollonian network with node-dependent interactions', *Phys. Rev. E* **79**(3) (2009), 036105.

[3] K. B. Athreya, P. E. Ney and P. Ney, *Branching Processes* (Courier Corporation, New York Heidelberg Berlin, 2004).

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

[4] I. Benjamini and O. Schramm, 'Recurrence of distributional limits of finite planar graphs', in *Selected Works of Oded Schramm* (Springer, 2011), 533–545.

[5] N. Berger, C. Borgs, J. T. Chayes and A. Saberi, 'On the spread of viruses on the internet', in *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (Society for Industrial and Applied Mathematics, 2005), 301–10.

[6] S. Bhamidi, D. Nam, O. Nguyen and A. Sly, 'Survival and extinction of epidemics on random graphs with general degree', *Ann. Probab.* **49**(1) (2021), 244–286.

[7] S. Bhamidi, R. van der Hofstad and G. Hooghiemstra, 'Universality for first passage percolation on sparse random graphs', *Ann. Probab.* (2017), 2568–2630.

[8] N. H. Bingham, C. M. Goldie, J. L. Teugels and J. Teugels, *Regular Variation* (Cambridge University Press, Cambridge, 1989), no. 27.

[9] M. Bloznelis, 'Component evolution in general random intersection graphs', *SIAM J. Discrete Math.* **24** (10)(2) (2010), 639–654. [Online]. https://doi.org/10.1137/080713756

[10] M. Bloznelis, 'Degree and clustering coefficient in sparse random intersection graphs', *Ann. Appl. Probab.* **23** (10)(3) (2013), 1254–1289. [Online]. Available: https://doi.org/10.1214/12-AAP874

[11] B. Bollobás, S. Janson and O. Riordan, 'The phase transition in inhomogeneous random graphs', *Random Struct. Algor.* **31** (2007), 3–122.

[12] B. Bollobás, 'A probabilistic proof of an asymptotic formula for the number of labelled regular graphs', *Eur. J. Comb.* **1**(4) (1980), 311–316. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0195669880800308

[13] M. Bonaventura, V. Nicosia and V. Latora, 'Characteristic times of biased random walks on complex networks', *Phys. Rev. E* **89**(1) (2014), 012803.

[14] K. Bringmann, R. Keusch and J. Lengler, 'Geometric inhomogeneous random graphs', *Theor. Comput. Sci.* **760** (2019), 35–54.

[15] V. H. Can, 'Metastability for the contact process on the preferential attachment graph', *Internet Mathematics*, 2017.

[16] M. Cassandro, A. Galves, E. Olivieri and M. E. Vares, 'Metastable behavior of stochastic dynamics: a pathwise approach', *J. Stat. Phys.* **35** (1984), pp. 603–634.

[17] S. Chatterjee and R. Durrett, 'Contact processes on random graphs with power law degree distributions have critical value 0', *Ann. Probab.* **37**(6) (2009), pp. 2332–2356.

[18] F. Chung and L. Lu, 'The average distance in a random graph with given expected degrees', *Internet Math.* **1** (2003), pp. 91–113.

[19] G. L. B. da Silva, R. I. Oliveira and D. Valesin, 'The contact process over a dynamical d-regular graph', arXiv Preprint, 2021, arXiv:2111.11757.

[20] M. Deijfen and W. Kets, 'Random intersection graphs with tunable degree distribution and clustering', *Probab. Eng. Inf. Sci.* **23**(4) (2009), 661–674. [Online]. https://doi.org/10.1017/S0269964809990064

[21] M. Deijfen, R. Van der Hofstad and G. Hooghiemstra, 'Scale-free percolation', *Annales de l'IHP Probabilités et statistiques*, **49**(3) (2013), 817–838.

[22] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* vol. 38 (Springer Science & Business Media, New York Heidelberg Berlin, 2009).

[23] C. Ding and K. Li, 'Centrality ranking in multiplex networks using topologically biased random walks', *Neurocomputing* **312** (2018), 263–275.

[24] R. Durrett and D. Griffeath, 'Supercritical contact processes on z', *Ann. Probab.* (1983), **1–15**.

[25] R. Durrett and R. H. Schonmann, 'The contact process on a finite set. ii', *Ann. Probab.* (1988), 1570–1583.

[26] P. Erdős and A. Rényi, 'On the evolution of random graphs', *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, pp. 17–61, 1960.

[27] J. Feldman and J. Janssen, 'High degree vertices and spread of infections in spatially modelled social networks', in *International Workshop on Algorithms and Models for the Web-Graph* (Springer, 2017), 60–74.

[28] D. Fernholz and V. Ramachandran, 'The giant k-core of a random graph with a specified degree sequence', manuscript, UT-Austin (2003).

[29] C. Giuraniuc, J. Hatchett, J. Indekeu, M. Leone, I. P. Castillo, B. Van Schaeybroeck and C. Vanderzande, 'Criticality on networks with topology-dependent interactions', *Phys. Rev. E* **74**(3) (2006), 036108.

[30] P. Gracar and A. Grauer, 'The contact process on scale-free geometric random graphs', arXiv Preprint, 2022, arXiv:2208.08346.

[31] G. Grimmett, *Probability on Graphs: Random Processes on Graphs and Lattices* vol. 8 (Cambridge University Press, Cambridge, 2018).

[32] T. E. Harris, 'Contact interactions on a lattice', *Ann. Probab.* **2**(6) (1974), 969–988.

[33] H. Hooyberghs, B. Van Schaeybroeck, A. A. Moreira, J. S. Andrade Jr, H. J. Herrmann and J. O. Indekeu, 'Biased percolation on scale-free networks', *Phys. Rev. E* **81**(1) (2010), 011102.

[34] X. Huang and R. Durrett, 'The contact process on random graphs and Galton Watson trees', *ALEA* **17** (2020), 159–182.

[35] E. Jacob, A. Linker and P. Moerters, 'Metastability of the contact process on fast evolving scale-free networks', *Ann. Appl. Probab.* **29**(5) (2019), 2654–2700.

[36] E. Jacob, A. Linker and P. Mörters, 'The contact process on dynamical scale-free networks', arXiv Preprint, 2022, arXiv:2206.01073.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

[37] S. Janson and M. J. Luczak, 'A simple solution to the k-core problem', *Random Struct. Algorithms* **30**(1–2) (2007), 50–62.
[38] M. Karonski, E. R. Scheinerman and K. B. Singer-Cohen, 'On random intersection graphs: The subgraph problem', *Comb. Probab. Comput.* **8**(1&2) (1999), 131–159.
[39] M. Karsai, R. Juhász and F. Iglói, 'Nonequilibrium phase transitions and finite-size scaling in weighted scale-free networks', *Phys. Rev. E* **73**(3) (2006), 036116.
[40] J. Komjáthy, J. Lapinskas and J. Lengler, 'Penalising transmission to hubs in scale-free spatial random graphs', in *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, **57**(4) (2021), 1968–2016.
[41] J. Komjáthy, J. Lapinskas, J. Lengler and U. Schaller, 'Four universal growth regimes in degree-dependent first passage percolation on spatial random graphs i', arXiv Preprint, 2023, arXiv:2309.11840.
[42] J. Komjáthy, J. Lapinskas, J. Lengler and U. Schaller, 'Four universal growth regimes in degree-dependent first passage percolation on spatial random graphs i', arXiv Preprint, 2023, arXiv:2309.11880.
[43] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat and M. Boguná, 'Hyperbolic geometry of complex networks', *Phys. Rev. E.* **82**(3) (2010), 036106.
[44] K. Kroy, 'Superspreading and heterogeneity in epidemics', in *Diffusive Spreading in Nature, Technology and Society* (Springer, 2023), 473–507.
[45] S. Lee, S.-H. Yook and Y. Kim, 'Centrality measure of complex networks using biased random walks', *Eur. Phys. J. B* **68**(2) (2009), 277–281.
[46] T. M. Liggett, 'Multiple transition points for the contact process on the binary tree', *Ann. Probab.* **24**(4) (1996), 1675–1710.
[47] T. M. Liggett et al., *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes* vol. 324 (Springer Science & Business Media, New York Heidelberg Berlin, 1999).
[48] A. Linker, D. Mitsche, B. Schapira and D. Valesin, 'The contact process on random hyperbolic graphs: metastability and critical exponents', *Ann. Probab.* **49**(3) (2021), 1480–1514.
[49] G. Miritello, E. Moro, R. Lara, R. Martínez-López, J. Belchamber, S. G. Roberts and R. I. Dunbar, 'Time as a limited resource: Communication strategy in mobile phone networks', *Social Networks* **35**(1) (2013), 89–95.
[50] M. Molloy and B. Reed, 'A critical point for random graphs with a given degree sequence', *Random Struct. Algorithms*, **6**(2–3) (1995), 161–180. [Online]. https://doi.org/10.1002/rsa.3240060204
[51] M. Molloy and B. Reed, 'The size of the giant component of a random graph with a given degree sequence', *Comb. Probab. Comput.* **7**(3) (Sep. 1998), 295–305. [Online]. https://doi.org/10.1017/S0963548398003526
[52] T. Mountford, J.-C. Mourrat, D. Valesin and Q. Yao, 'Exponential extinction time of the contact process on finite graphs', *Stoch. Process. Their Appl.* **126**(7) (2016), 1974–2013.
[53] T. Mountford, D. Valesin and Q. Yao, 'Metastable densities for the contact process on power law random graphs', *Electron. J. Probab.* **18** (2013), 1–36.
[54] T. Mountford, 'Existence of a constant for finite system extinction', *J. Stat. Phys.* **96** (1999), 1331–1341.
[55] T. Mountford, 'A metastable result for the finite multidimensional contact process', *Can. Math. Bull.* **36**(2) (1993), 216–226.
[56] J.-C. Mourrat and D. Valesin, 'Phase transition of the contact process on random regular graphs', *Electron. J. Probab.* **21** (2016), 1–17.
[57] R. Pemantle, 'The contact process on trees', *Ann. Probab.* (1992), 2089–2116.
[58] R. Pemantle and A. M. Stacey, 'The branching random walk and contact process on Galton-Watson and nonhomogeneous trees', *Ann. Probab.* **29**(4) (2001), 1563–1590.
[59] C. Pu, S. Li, and J. Yang, 'Epidemic spreading driven by biased random walks', *Physica A Stat. Mech. Appl.* **432** (2015), 230–239.
[60] H. Reittu and I. Norros, 'On the power-law random graph model of massive data networks', *Perform. Eval.* **55**(1–2) (Jan 2004), 3–23. [Online]. http://doi.org/10.1016/S0166-5316(03)00097-X
[61] B. Schapira and D. Valesin, 'Extinction time for the contact process on general graphs', *Probab. Theory Relat. Fields* **169** (2017), 871–899.
[62] R. H. Schonmann, 'Metastability for the contact process', *J. Stat. Phys.* **41** (1985), 445–464.
[63] K. B. Singer, *Random intersection graphs*. ProQuest LLC, Ann Arbor, MI, 1996, thesis (Ph.D.) – The Johns Hopkins University. [Online]. URL: http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqdiss&rft_dat=xri:pqdiss:9617602
[64] N. Slater, R. M. Mitchell, R. H. Whitlock, T. Fyock, A. K. Pradhan, E. Knupfer, Y. H. Schukken and Y. Louzoun, 'Impact of the shedding level on transmission of persistent infections in mycobacterium avium subspecies paratuberculosis (map)', *Vet. Res.* **47**(1) (2016), 1–12.
[65] A. M. Stacey, 'The existence of an intermediate phase for the contact process on trees', *Ann. Probab.* **24**(4) (1996), 1711–1726.
[66] W. Su, 'Branching random walks and contact processes on Galton-Watson trees', *Electron. J. Probab.* **19** (2014), 1–12.
[67] R. van der Hofstad, *Random Graphs and Complex Networks* vol. 43 (Cambridge University Press, Cambridge, 2016).
[68] R. van der Hofstad and J. Komjáthy, 'When is a scale-free graph ultra-small?', *J. Stat. Phys.* **169** (2017), 223–264.
[69] R. Van Der Hofstad, J. Komjáthy and V. Vadon, 'Random intersection graphs with communities', *Adv. Appl. Prob.* **53**(4) (2021), 1061–1089.
[70] R. van der Hofstad, J. Komjáthy and V. Vadon, 'Phase transition in random intersection graphs with communities', *Random Struct. Algorithms* **60**(3) (2022), 406–461.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press

[71] J. Wang, W. Xiong, R. Wang, S. Cai, D. Wu, W. Wang and X. Chen, 'Effects of the information-driven awareness on epidemic spreading on multiplex networks', *Chaos*, **32**(7) (2022).
[72] R. Yang, T. Zhou, Y.-B. Xie, Y.-C. Lai and B.-H. Wang, 'Optimal contact process on complex networks', *Phys. Rev. E* **78**(6) (2008), 066109.
[73] V. Zlatić, A. Gabrielli and G. Caldarelli, 'Topologically biased random walk and community finding in networks', *Phys. Rev. E* **82**(6) (2010), 066109.

https://doi.org/10.1017/fms.2025.10144 Published online by Cambridge University Press