ABSTRACT

Title of Dissertation: ERASURE CORRECTION FOR GRAPH CODES AS A PROBLEM OF BOOTSTRAP PERCOLATION

Fengxing Zhu
Doctor of Philosophy, 2025

Dissertation directed by: Professor Alexander Barg
Department of Electrical and Computer Engineering

Distributed storage systems are designed to have the capability to recover the contents of failed nodes by accessing the contents of other functional nodes in the storage cluster. To achieve this property in storage systems, engineers have focused considerable effort on developing codes that enable local recovery. An encoding method is said to have locality r if the contents of a storage node can be reconstructed by accessing r other nodes in the storage cluster. An additional constraint faced by the designer is accounting for the underlying topology of the system that may affect the functioning of the recovery method. In this research, we study certain aspects of storage systems with graphical structure, modeling them as storage codes on graphs. A storage code is an assignment of bits to the vertices of a graph with the property that every vertex can recover its value from its neighbors.

Specializing the problem further, we assume that several vertices in the graph represent failed storage nodes, and the task of code construction is to support their recovery. Suppose that every vertex is erased with some probability p independently from the other vertices, and that the recovery proceeds in rounds, whereby the vertices with more than r functional neighbors are assumed to be able to recover themselves. As the number of erased vertices decreases, additional vertices may be able to restore their values until all the erasures are corrected. In graph-theoretic terms this procedure is known as *bootstrap percolation* with infection threshold r (it is often called infection spread in graphs).

The research presented in this thesis extends earlier works on bootstrap percolation. We determine asymptotics of the critical probability p_c of bootstrap percolation on the binary n-dimensional hypercube, solving for several regimes of dependence of r on the dimension n. We also determine the maximum time to erasure recovery (the maximum percolation time) on the q-ary hypercube graph. Extending the classical setting, we modify the hypercube graph to expand the neighborhood to allow vertices at distance 2 or more, and compute the asymptotics of critical probability and obtain other similar asymptotic results.

ERASURE CORRECTION FOR GRAPH CODES AS A PROBLEM OF BOOTSTRAP PERCOLATION

by

FENGXING ZHU

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2025

Committee Members:
PROFESSOR ALEXANDER BARG (CHAIR)
PROFESSOR BEHTASH BABADI
PROFESSOR ARMAND MAKOWSKI
PROFESSOR PRAKASH NARAYAN
PROFESSOR ARAVIND SRINIVASAN (DEAN'S REPRESENTATIVE)

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my research advisor, Prof. Alexander Barg, for his unwavering support and encouragement. He introduced me to the topic of bootstrap percolation and consistently guided me throughout this research. Without his insight and mentorship, this thesis would not have been possible. His constant reassurance helped me stay motivated, and I am deeply thankful for his availability and willingness to meet with me whenever I needed assistance.

I am also extremely grateful to my family for their love and support, without which I could not have come this far. In particular, my weekly conversations with my parents have been a source of strength and comfort, consistently nourishing and recharging me.

I would like to thank Yihan Zhang for his ongoing engagement with my research. His willingness to discuss ideas and meet with me almost every week has made this journey more enjoyable and intellectually stimulating. I also appreciate Ke Wu for taking the time to meet with me and explain deletion and insertion codes in great detail.

I have had many enriching conversations of academic and general nature with Prof. Eyad Abed and Prof. Richard La, and I am thankful for their time, insights, and encouragement.

My interactions with fellow students in my research group, as well as with others, have significantly shaped my academic experience. From them, I have learned not only important research topics but also valuable life lessons.

Finally, I would like to thank the members of my committee for their willingness to serve.

Contents

Chapter 1

INTRODUCTION

Distributed storage systems have been extensively utilized in various fields such as cloud computing, big data processing, and scientific computing, among others. Given their ability to handle large volumes of data across multiple nodes in a distributed environment, it is imperative to design robust distributed storage systems capable of efficient information storage and recovery. One commonly employed model for distributed storage systems has the ability to locally recover from a single node failure, or even withstand multiple node failures simultaneously. In practical terms, the most frequent error event in a distributed storage system is the failure of a single node. When such a failure occurs, the objective is to reconstruct the content of the failed node from the remaining properly functioning nodes, minimizing the number of nodes involved in the recovery process. In order to handle this challenge, the study of codes with locality for was initiated by Gopalan et al. [32]. We say that the ith coordinate of a codeword c in a code $\mathcal{C}$ has locality r if c_i can be recovered from $\{c_j\}_{j\in S}$ where $|S| = r$ and $S \subset [n\backslash i]$. However, this model does not account for the underlying topology of a distributed system and in practice that aspect has to be dealt with. Motivated by this consideration and also by earlier works by Alon et al. [3], Mazumdar [40] [41] and Shanmugam and Dimakis [45] introduced the concept of storage codes for graphs around 2014.

Let us formally define storage codes for a graph G. Given a connected graph $G(V, E)$ with $|V| = N$, denote by $\mathcal{N}(v) = \{u : (u, v) \in E\}$ the set of neighbors of the vertex v on G. A storage code for G is a set $\mathcal{C}$ of vectors $\mathbf{c} = (\mathbf{c}_v)_{v\in V} \in \mathbb{F}_q^N$ such that for every $\mathbf{c} \in \mathcal{C}$ and $v \in V$ the value of $\mathbf{c}_v$ is uniquely

determined by the values of $\mathbf{c}_w$ with $w \in \mathcal{N}(v)$. More precisely, for every vertex $v \in V$ there is a function $f_v : \mathbb{F}_q^{\mathcal{N}(v)} \to \mathbb{F}_q$ such that for every $\mathbf{c} \in \mathcal{C}$ and every $v \in V$ we have $\mathbf{c}_v = f_v((\mathbf{c}_w)_{w \in \mathcal{N}(v)})$. The definition of storage codes implies the ability to recover an erased coordinate of a codeword from its neighborhood. This highlights the primary distinction between codes with locality and storage codes for graphs. While codes with locality can correct an erased coordinate of a codeword using the values of other coordinates (repair group), aiming to minimize the size of the repair group, storage codes for a graph G impose an additional constraint on the repair group for each coordinate of a codeword. In other words, the repair group for each coordinate of a codeword consists of its neighbors as defined by the underlying graph.

While the most probable error event is the failure of a single node, it is also possible to encounter multiple node failures. From an erasure correction perspective, a single node failure corresponds to correcting a single erasure, while multiple node failures correspond to correcting multiple erasures. Moreover, the problem of correcting multiple erasures naturally extends the problem of single-erasure correction. Building upon this concept, Mazumdar [41] introduced the problem of correcting multiple erasures. In this discussion, our focus will be on codes constructed from the edge-vertex construction, defined below, using a d-regular underlying graph G. Various known constructions of storage codes for a given graph G can be found in [20], [44], [4], and some of these constructions, including the matching construction, edge-vertex construction, clique-vertex construction, and fractional clique cover, were mentioned in [15].

Now we will describe the edge-vertex construction for a d-regular graph $G(V, E)$ with N vertices, as introduced in [20] and [24]. Consider the space $(\mathbb{F}_q)^{|E|}$ of vectors indexed by the edges of G. Next, we map this space onto $(\mathbb{F}_q^d)^N \cong \mathbb{F}_{q^d}^N$ by assigning to each vertex v a d-vector formed by the symbols written on the edges incident to it. Equivalently, the code $\mathcal{C}(G) = c_v, v \in V$ is obtained as the image of the following mapping:

$$\mathbb{F}_q^{|E|} \to \mathbb{F}_{q^d}^N$$

$$(x_e)_{e \in E} \to (c_v)_{v \in V}, \text{ where } c_v = (x_e)_{e \in \partial(v)}.$$

Here, $\partial(v)$ represents the edge neighborhood of v in G. According to this definition, for a given $v \in V$, c_v from $\mathcal{C}(G)$ is an element in $\mathbb{F}_q^d$, or it can be viewed as a d-vector over $\mathbb{F}_q$ obtained from the values placed on the edges incident to $v \in V$. Now, it is required for $\mathcal{C}(G)$ to be able to correct multiple erasures. In other words, for every $v \in V$ and every $c \in \mathcal{C}(G)$, c_v can be recovered from $c_{v'}$, where $v' \in \mathcal{N}(v)$, as long as at most t of them are erased. One way to achieve this is by enforcing the symbols on the edges incident to each vertex $v \in V$ to belong to a linear code D of length d over $\mathbb{F}_q$. Additionally, we require the linear code D to be capable of correcting t erasures. This defines a linear code $\mathcal{C}(D, G)$ that enables each vertex to recover its value from $d - t$ or more non-erased neighbors.

The problem of correcting multiple erasures in a deterministic manner using storage codes for a given graph G appears interesting. However, it is also worthwhile to explore the same problem in a random setting, as formulated by [15]. Let's consider the edge-vertex construction with a local code capable of correcting t erasures. Under this setting, each vertex is independently and randomly erased with a probability $\bar{p}$, resulting in a set of erased vertices. It is important to note that erased vertices can be recovered iteratively until all of them are corrected. In other words, a vertex is not erased with a probability $p = 1 - \bar{p}$, which is independent of any other vertices, and an erased vertex can recover its value if it has at least $d - t$ non-erased neighbors. Barg and Zémor [15] established a connection between correcting multiple erasures using storage codes for a graph G and bootstrap percolation on the same graph G.

The process of r-neighbor bootstrap percolation on an undirected graph $G(V, E)$, with an integer $r \geqslant 1$, was introduced by Chalupa, Leith, and Reich [23]. In this process, each vertex is either infected or healthy, and once a vertex becomes infected, it remains infected forever. Initially, a set of vertices A_0 is infected, and let A_i denote the set of infected vertices up to step i. The bootstrap percolation process evolves in discrete steps as follows: for $i > 0$,

$$A_i = A_{i-1} \cup \{v \in V : |\mathcal{N}(v) \cap A_{i-1}| \geqslant r\},$$

where $\mathcal{N}(v)$ represents the neighborhood of vertex v. In simple terms, a vertex that is not initially infected becomes infected at step i if it has at least r infected neighbors at step $i-1$. Here, we define r as the infection threshold for all $v \in V$ and a contagious set as a set of initially infected vertices that leads to the complete infection of the entire graph. Percolation occurs if all vertices have been infected by the end of the process. In random bootstrap percolation, each vertex is independently and randomly infected at the beginning with a probability p. The central question in the problem of random bootstrap percolation is to determine the critical probability, denoted as $p_c(G, r)$:

$$p_c(G, r) = \sup\left\{p \in (0, 1) : \mathbb{P}_p(A \text{ percolates on } G) \leqslant \frac{1}{2}\right\},$$

where A represents the set of initially infected vertices.

Indeed, the erasure correction process in the random setting can be seen as equivalent to random bootstrap percolation on a graph G, just described in different terms. In our context, percolation occurs when all erased vertices have been corrected through the iterative process. The critical probability can be defined as follows:

$$p_c(G, d-t) = \sup\left\{p \in (0, 1) : \mathbb{P}_p(\text{all erased vertices are corrected by the iterative process}) \leqslant \frac{1}{2}\right\}.$$

Bootstrap percolation in the deterministic setting has attracted considerable attention, as evident in works such as [25] and [30]. In the deterministic setting, two main questions have been investigated. The first question focuses on finding the minimum number of initially infected vertices required for the entire graph to eventually become infected. The second question involves determining the maximum number of steps needed for percolation to occur, considering all possible contagious sets.

Rephrasing bootstrap percolation in the deterministic setting for the erasure correction process, *one question* is to determine the minimum size of the non-erased vertex set required to ensure the correction of all

erased vertices in the end. This corresponds to finding the smallest set of non-erased vertices that guarantees complete correction. The *second question* relates to determining the maximum correction time, which refers to the maximum number of steps required to correct all erased vertices. This involves considering all choices of initially non-erased vertices, ensuring that the chosen set guarantees the eventual full correction of all erased vertices. By addressing these questions in this specific context, we can gain insights into the minimum requirements for complete erasure correction and the maximum time needed for the correction process across different choices of non-erased vertices.

Mathematically, the grid and the hypercube are fundamental structures that arise in a wide range of areas in mathematics, such as combinatorics, graph theory, and theoretical computer science, due to their rich structural properties, inherent symmetries, and recursive definitions. In terms of practical applications, grid and hypercube topologies serve as main network architectures in computer networking and data storage systems, where they are employed to facilitate efficient data routing and communication among multiple processing units or computers [1].

Motivated by the mathematical importance and applications for the storage systems, we study the process of bootstrap percolation on the hypercube, the hypercube with the extended neighborhood, and the discrete torus. More precisely, under the probabilistic setting, we study the critical probability for the n-dimensional hypercube when the infection threshold is of the form n^a where $0 < a < 1$ and for the n-dimensional hypercube with extended neighbors when the infection threshold is half of the degree of a vertex. Additionally, we study the percolation time on the d-dimensional torus under the probabilistic setting. Under the deterministic setting, we derive the maximum percolation time on the n-dimensional q-ary hypercube. These results can help us estimate approximately how long it will take for the system to recover from multiple node failures, and how many node failures the system can tolerate while still being able to support data recovery. The results of this thesis appear in preprints [46–49], submitted for publication.

Chapter 2

EXISTING RESULTS AND OUR CONTRIBUTIONS

2.1 EXISTING RESULTS

For bootstrap percolation under the random setting, Aizenman and Lebowitz [2] demonstrated that when the process occurs within a finite box $\{1, 2, \ldots, n\}^d$, the critical probability scales as $\Theta((\log(n))^{1-d})$ with $d \geqslant r = 2$. In their work [7], Balogh and Bollobás established the sharp phase transition for bootstrap percolation on a d-dimensional box, building upon the results from [31]. Notably, Holroyd [27] achieved a significant breakthrough by identifying the precise threshold for $d = r = 2$. Subsequent advancements were made by Gravner and Holroyd [33] as well as Hartarsky and Morris [36], who determined the scaling of the second-order term,

$$p_c([n]^2, 2) = \frac{\pi}{18 \log n} - \frac{\Theta(1)}{(\log n)^{\frac{3}{2}}}.$$

For the case where $d = r = 3$, Cerf and Cirillo [22] established that the critical probability scales as $\Theta\left(\frac{1}{\log \log(n)}\right)$. In the scenario where $d \geqslant r \geqslant 3$, Cerf and Manzo [21] demonstrated that the critical probability follows a behavior of $\Theta\left((\log^{(r-1)}(n))^{r-d-1}\right)$, where $\log^{(r-1)}$ represents the logarithm iterated $r-1$ times. The sharp threshold for the r-neighbor process on $[n]^d$ was ultimately determined by Balogh, Bollobás, Duminil-Copin, and Morris [9]. For every $d \geqslant r \geqslant 2$, they identified an explicit constant $\lambda(d, r) > 0$

such that the critical probability can be expressed as follows:

$$p_c([n]^d, r) = \left[\frac{\lambda(d,r) + o(1)}{\log^{(r-1)} n}\right]^{d-r+1}.$$

While a considerable amount of attention has been devoted to the study of random bootstrap percolation in the d-dimensional box, it has also been extensively explored on the hypercube. Let $Q(n)$ be a hypercube graph whose vertices are the binary n-dimensional vectors. Balogh and Bollobás [8] established the following result: there exist constants $c_1 > 0$ and $c_2 > 0$ such that

$$\frac{c_1}{n^2} 2^{-2\sqrt{n}} \leqslant p_c(Q(n), r = 2) \leqslant \frac{c_2}{n^2} 2^{-2\sqrt{n}}.$$

In a subsequent study, Balogh, Bollobás, and Morris further improved upon the bound presented in [12] and obtained the following result: there exists an explicit constant $\lambda > 0$ such that

$$\frac{16\lambda}{n^2}\left(1 + \frac{\log n}{\sqrt{n}}\right) 2^{-2\sqrt{n}} \leqslant p_c(Q(n), 2) \leqslant \frac{16\lambda}{n^2}\left(1 + \frac{5(\log n)^2}{\sqrt{n}}\right) 2^{-2\sqrt{n}}.$$

For majority bootstrap percolation on the hypercube Q_n, where the infection threshold r is equal to $\frac{1}{2}n$, Balogh, Bollobás and Morris in [11] and [10] proved that:

$$\frac{1}{2} - \frac{1}{2}\sqrt{\frac{\log n}{n}} - 2\frac{\log\log n}{\sqrt{n\log n}} \leqslant p_c\Big(Q_n, \frac{n}{2}\Big) \leqslant \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\log n}{n}} + \frac{\log\log n}{2\sqrt{n\log n}} + o\left(\frac{\log\log n}{2\sqrt{n\log n}}\right).$$

In addition to the d-dimensional box and the hypercube, some efforts have been devoted to investigating bootstrap percolation on random graphs. Feige, Krivelevich, and Reichman [28] examined the r-neighbor bootstrap percolation on the binomial random graph $G(n,p)$, where $p = \frac{d}{n}$ and $1 \ll d \ll \left(\frac{n\log\log n}{\log^2 n}\right)^{\frac{r-1}{r}}$. Assuming $r > 1$ to be a constant independent of n, they proved that the minimum size of a contagious set,

denoted by $m(G(n,p),r)$, satisfies

$$m(G,r) = \Theta\left(\frac{n}{d^{\frac{r}{r-1}}\log d}\right),$$

with high probability. Furthermore, they determined that the threshold probability for $m(G,r) = r$ to hold is

$$p^* = \Theta\left(\frac{1}{\left(n\log^{r-1} n\right)^{\frac{1}{r}}}\right).$$

Subsequently, Holmgren, Juskevicius and Kettle [37] studied the majority bootstrap percolation on the binomial random graph $G(n,p)$. Their approach involved selecting an initially infected set A_0 of size m uniformly from $[n]^m$ (all subsets of $[n]$ of size m), followed by the selection of edges for $G(n,p)$. This process is referred to as $MB(n,p;m)$. We say that the $MB(n,p;m)$ percolates if all vertices eventually become infected. They proved the following theorem:

Theorem 2.1.1. *Fix some number $\epsilon > 0$. Assume that for n large enough, $(1+\epsilon)\log n \leqslant p(1-p)n$ and $p \leqslant 1 - \frac{\log^4 n}{n}$. Let $d = \frac{np}{1-p}$. If the initially infected set A_0 has size*

$$m = \frac{n}{2} - \frac{n}{2}\sqrt{\frac{\log d}{d}} + \lambda n\frac{\log\log\log d}{\sqrt{d\log d}} + o\left(\frac{\log\log\log d}{\sqrt{d\log d}}\right),$$

then as $n \to \infty$,

$$\mathbb{P}(MB(n,p;m) \textit{ percolates}) = \begin{cases} 1, & \textit{if } \lambda > \frac{1}{2}, \\ 0, & \textit{if } \lambda < 0. \end{cases}$$

Let's now consider the deterministic setting for bootstrap percolation. By employing the techniques developed in [6], Morrison and Noel [42] determined the leading term of the minimum size of the contagious set for the r-neighbor process on the hypercube $Q(n)$ with n vertices. This quantity is denoted as $m(Q(n),r)$. In the case where n tends to infinity while keeping $r > 2$ fixed, they established the following

expression:

$$m(Q(n), r) = \frac{1 + o(1)}{r} \binom{n}{r-1}.$$

Another significant class of graphs comprises graphs with good expansion. It is well known that the expansion property of a graph is controlled by the spectral gap, i.e., the difference between the largest eigenvalue of its adjacency matrix λ_1 and the absolute value of the second eigenvalue. Heuristically, in a graph with a strong expansion property, every set of vertices is expected to have a large number of external neighbors, leading to the possibility of small-sized minimum contagious sets. Coja-Oghlan et al. [25] formally proved this statement. To describe their result, consider d-regular graphs on n vertices with eigenvalues $\lambda_1 > \lambda_2 \geqslant \ldots \geqslant \lambda_n$, and let $\lambda := \max\{|\lambda_2|, |\lambda_n|\}$. In order to get a sense of their results, let us mention one of the theorems that establishes an upper bound for the minimum size of contagious sets.

Theorem 2.1.2. *Let G be an (n, d, λ)-graph. If $\lambda = O(\sqrt{d})$, then $m(G, r = 2) = O(\frac{n}{d^2})$.*

While the expansion property of a graph roughly ensures that every set of vertices has a significant number of external neighbors, a graph with high density, where the minimum degree of vertices is very large, can achieve the same effect. Formally, Gunderson [34] characterized the minimum size of contagious sets in terms of the minimum degree of a graph. The following theorem was proved in [34].

Theorem 2.1.3. *In the r-neighbor bootstrap percolation on any graph G with n vertices (n large enough) and with minimum degree $\lfloor \frac{n}{2} \rfloor + 1$, we have $m(G, r = 3) = 3$. Moreover, for $r \geqslant 4$, in the r-neighbor bootstrap percolation on any graph G with n vertices (n large enough) and with minimum degree $\lfloor \frac{n}{2} \rfloor + r - 3$, we have $m(G, r) = r$. The minimum-degree requirements are sharp.*

One important class of dense graphs is Ore graphs. An Ore graph $G(V, E)$ is defined such that for any $u, v \in V$ with $uv \notin E$, it holds that $\deg(u) + \deg(v) \geqslant n$. Efforts have been made to determine the minimum size of contagious sets on Ore graphs. Freud, Poloczek, and Reichman [30] proved the following theorem:

Theorem 2.1.4. *In the 2-neighbor bootstrap percolation on an Ore graph G with n vertices (where n is sufficiently large), we have $m(G, r = 2) = 2$.*

Another direction of research in bootstrap percolation involves determining the maximum percolation time for the process across all choices of initially infected sets that result in the infection of the entire graph. Let Q_n denote the hypercube on n vertices. Przykucki proved the following theorem regarding the maximum percolation time for the 2-neighbor process in [43].

Theorem 2.1.5. *Assume the 2-neighbor bootstrap percolation on Q_n. If $A \in Q_n$ percolates, then it percolates in at most $\left\lfloor \frac{n^2}{3} \right\rfloor$ steps.*

Following Przykucki's work, Hartarsky [35] determined the maximum percolation time on Q_n up to a $\log n$ factor for the r-neighbor process, where $r \geqslant 3$. The following result was shown by Hartarsky in [35].

Theorem 2.1.6. *Assume the r-neighbor bootstrap percolation on Q_n with $r \geqslant 3$. If $A \in Q_n$ percolates, then it percolates in at most $\frac{2^n}{n}(\log n)^{-O(1)}$ steps. Moreover, this bound is tight for all n.*

By comparing Theorem 2.1.5 and Theorem 2.1.6, we observe a significant increase in the maximum percolation time on the hypercube when transitioning from $r = 2$ to $r \geqslant 3$.

2.2 OUR CONTRIBUTIONS

We consider bootstrap percolation on the Hamming cube, including a generalized cube, and on the n-dimensional torus. The n-dimensional Hamming cube, denoted by $Q_n(V, E)$, is a graph with vertices given by the n-vectors over a q-ary alphabet, $V = \{0, 1, \ldots, q-1\}^n$ and edges $E = \{xy : d_H(x, y) = 1\}$, where d_H is the Hamming distance. The generalized n-dimensional Hamming cube, denoted by $Q_{n,k}(V, E)$, has the same set of vertices and edges $E = \{xy : 1 \leqslant d_H(x, y) \leqslant k\}$, where $k \geqslant 2$. The d-dimensional torus is the graph, denoted by $\mathbb{T}_n^d$, with vertex set $V = \{0, \ldots, n-1\}^d$ and $x, y \in V$ are connected if and only if for some $i \in \{1, \ldots, d\}$, $x_i = y_i \pm 1 \pmod{n}$ and $x_j = y_j$ for all $j \neq i$.

In the thesis, we consider the cases of the generalized binary cube, the q-ary n-cube, $q \geqslant 2$ and the torus.

2.2.1 BOOTSTRAP PERCOLATION ON A GENERALIZED HAMMING CUBE I

We study the bootstrap percolation process on a generalized Hamming cube $Q_{n,k}$. As our main contributions, we characterize percolating sets and determine the critical probability for the process when the infection threshold is 2 and 3. Additionally, we determine the size of minimum percolating sets in certain cases and establish bounds in others.

Our arguments start with a simple characterization of initially infected sets that fail to percolate. An initially infected set fails to percolate if and only if there exists a set of uninfected vertices B such that every vertex in B has fewer than r neighbors outside of B. To derive a lower bound on the minimum percolating sets, we use this description of initially infected sets to derive a lower bound for the case of large r. The bound turns out to be exponentially tight when $r = \Theta(n^k)$.

These results are presented in Ch. 3 and appear in [46].

2.2.2 BOOTSTRAP PERCOLATION ON A GENERALIZED HAMMING CUBE II

While percolation on the n-dimensional hypercube with infection threshold $r = \frac{1}{2}n$ (*majority bootstrap percolation*) has been studied in [11] and [26], the case where the threshold is $r = n^a$ for some $a < 1$ had not been investigated prior to our work. In Section 9 of [26], two open questions were posed: characterizing d-regular graphs for which the critical probability is of the form d^{-a} with $0 < a < 1$, and investigating the behavior of the critical probability for d-regular graphs with infection threshold $r = \sqrt{d}$. We resolve the first question for the n-dimensional Hamming cube with infection threshold $r = n^a$ for $\frac{2}{3} < a < 1$.

We also examine bootstrap percolation on the n-dimensional hypercube with extended neighborhoods, another problem highlighted as an open question in Section 9 of [26]. We determine the leading term and establish bounds on the second-order term of the critical probability on $Q_{n,k}$. In the case $k = 2$, the second-order term is tight up to a factor of $\sqrt{\log n}$.

We combine the tools used in [11] and [26] with the additional tools such as the FKG inequality, the De Moivre–Laplace theorem and the factorization theorem of the complete k-uniform hypergraph to derive the desired results.

These results are presented in Ch. 4 and appear in [47].

2.2.3 MAXIMUM PERCOLATION TIME ON THE Q-ARY HYPERCUBE

We study the maximum percolation time on the q-ary Hamming cube $Q_n(V, E), q \geqslant 3$ with infection threshold 2. To derive a recursive formula for this quantity, we introduce a set of new operations designed to count the number of nonzero elements in a q-ary string of length n under certain constraints. These operations enable the formulation of several technical lemmas that characterize the spread of infection under various configurations of initially infected vertices, ultimately leading to the desired recursive formula for the maximum percolation time.

The derivation of the recursive formula is significantly more intricate than in the case of the binary

hypercube. In several parts of the proof, we introduce some more involved inductive arguments to establish important claims.

By using the recursive formula with the initial condition for $n = 0, 1, 2$, we successfully derive the maximum percolation time for every $n \geqslant 0$.

These results are presented in Ch.5 and appear in [48].

2.2.4 THE TIME OF BOOTSTRAP PERCOLATION IN HIGH DIMENSIONS

We study the percolation process on the d-dimensional torus, defined above, with infection threshold d. In two dimensions [5], the key intuition is that the event in which a vertex x remains uninfected at time t is "equivalent to" the existence of a line segment of uninfected vertices of length roughly t "near" x. In higher dimensions, it is not immediately clear whether this key intuition still holds. However, we show that it remains valid in higher dimensions, provided that the infection threshold equals the dimension.

In the two-dimensional case, the idea from [5] of analyzing the intersection between a path of uninfected vertices and squares of a certain size was effective in reducing the combinatorial factor from counting the number of such paths. In higher dimensions, however, it is not obvious which geometric objects should be used to achieve a similar reduction. We find that replacing two-dimensional squares with their higher-dimensional analogs (cubes) is a viable approach.

Another challenge lies in generalizing the definition of the interior of a square from two dimensions to higher dimensions in a way that serves our purposes. In the two-dimensional setting, the notion of a square's interior was introduced to properly define certain events, and a similar concept is required in higher dimensions. We address these challenges and provide the necessary definitions to extend the analysis.

These results are presented in Ch. 6 and appear in [49].

2.2.5 IMPLICATIONS FOR DISTRIBUTED STORAGE SYSTEMS

The q-ary hypercube is a natural generalization of the binary hypercube that preserves its fundamental structure. The torus can be viewed as a grid with periodic boundary conditions, where opposite edges are identified, resulting in a structure without boundary. Since the underlying topologies of many distributed storage systems are hypercubes or grids [1], our results about critical probability help to evaluate the robustness and resilience of such systems for a wider range of the parameter recovery threshold, i.e, the minimum number of functional neighboring nodes required to recover a failed node. More specifically, they enable us to estimate the number of failed nodes that a system can tolerate and return to full functionality without data rewrite. Our results about percolation time give insights into the expected time to complete recovery of the system.

Chapter 3

BOOTSTRAP PERCOLATION ON A GENERALIZED HAMMING CUBE, I

3.1 INTRODUCTION

In this chapter we study bootstrap percolation on the hypercube $\{0,1\}^n$ with edges added between the pairs of vertices at Hamming distance $\leqslant k$ for a fixed $k \geqslant 2$. We denote this graph by $Q_{n,k}(V,E)$, where $V = \{0,1\}^n$ and $E = \{xy : 1 \leqslant d_{yx} \leqslant k\}$, where $d_{xy} := d_H(x,y)$ is the Hamming distance and k is a fixed integer. Given a subgraph $G \subset Q_{n,k}$ and $v \in V$, we denote by $N_v(G)$ the set of neighbors of v in G, omitting the reference to G if it is the entire cube.

Our main results concern the critical probability $p_c(Q_{n,k}, r)$ for $r = 2, 3$, for which we obtain tight estimates of the form $2^{-n/r}n^{-k}$ as $n \to \infty$. These results are formally stated and proved in Sections 3.2 and 3.3. In Section 3.4 we prove some results about the size of the minimum contagious set on $Q_{n,k}$, with the most detailed answers for $k = 2$.

3.2 CRITICAL PROBABILITY FOR THE 2-NEIGHBOR PROCESS

We begin with two simple lemmas used repeatedly in this chapter.

Lemma 3.2.1. *Let $r, k \geqslant 2, n \geqslant r-1$ and let $A_0 \subset Q_{n,k}$ be a set of already infected vertices. If A_0 contains an $(r-1)$-dimensional subcube of $Q_{n,k}$, then the r-process on $Q_{n,k}$ will percolate after several*

additional steps.

Proof. W.l.o.g., consider the subcube[1] $Q^0_{r-1,k} = \{x = (*^{r-1}0^{n-r+1})\}$ formed of all the vertices that end with $n-r+1$ 0's and assume that all of its vertices are infected. We will show by induction that all the subcubes $Q^0_{i,k}, i = r, \dots, n$ become infected, implying the percolation claim. For the base, consider the vertex $v = (0^{r-1}10^{n-r})$ and observe that $|N_v(Q^0_{r-1,k})| \geqslant r$ for all $k \geqslant 2$: the origin 0^n and all the vertices of the form $x = (z|0^{n-r+1})$ with $|z| = 1$ satisfy $d(v,x) \leqslant 2$, so the infection spreads to v. Similarly, all the vertices of the form $x = (*^{r-1}|1|0^{n-r})$ have $\geqslant r$ neighbors in the set $Q^0_{r-1,k}$, so the entire set $Q^0_{r,k}$ gets infected, completing the base step. The transition from $Q^0_{i-1,k}$ to $Q^0_{i,k}, r+1 \leqslant i \leqslant n$ is argued in the same way (in fact, it becomes easier to propagate the infection as the dimension grows because the set of close neighbors in $Q^0_{i-1,k}$ grows in size as i increases). □

Because of this observation, to prove that percolation takes place, it suffices to show that with the appropriate choice of the initially infected set $A_0 \subset V$ several steps of the process will infect a subcube $Q_{r-1,k}$. We will use variations of this approach in the proofs of the main results.

Lemma 3.2.2. *Let $X_n, n \geqslant 1$ be a collection of random variables such that $X_n \in \mathbb{N}_0$, $EX_n \to \infty$ and $\mathrm{Var}(X_n) = o(\mathbb{E}X_n^2)$ as $n \to \infty$. Then $\mathbb{P}(X_n = 0) = o(1)$ and*

$$\mathbb{P}\Big(\frac{\mathbb{E}X_n}{2} \leqslant X_n \leqslant \frac{3}{2}\mathbb{E}X_n\Big) = 1 - o(1).$$

Proof. The first statement follows by an application of the second moment method and the second one by Chebyshev's inequality. □

Below by $\omega(n)$ and $C(k)$ we denote an arbitrarily slowly growing function of n and a constant which may depend on k respectively. We will use $\omega(n)$ and $C(k)$ as generic notations, where the specific form of

[1] by abuse of notation, we may write $Q_{r-1,k}$ to refer to an $(r-1)$-dimensional subcube of $Q_{n,k}$, where the locations of $*$ are not necessarily consecutive.

these functions may be different in different expressions. In this section our goal is to prove the following theorem.

Theorem 3.2.3. *(Critical probability for the 2-neighbor process). We have*

$$\frac{1}{\omega(n)}2^{-\frac{n}{2}}n^{-k} \leqslant p_c(Q_{n,k}, 2) \leqslant \omega(n)2^{-\frac{n}{2}}n^{-k}. \tag{3.1}$$

Proof. We assume that at the start of the process, every vertex is infected independently of the others with probability p and is not infected with probability $1-p$. Let A_0 be the set of initially infected vertices. We claim that percolation occurs if and only if there is a pair $x, y \in A_0$ with $d_{xy} \leqslant 2k$. If not, then no vertex will have 2 neighbors in $Q_{n,k}$, so the process cannot evolve. Now suppose the assumption holds true. We will prove that the process percolates, differentiating between the following two cases.

Suppose that $d_{xy} = t$, where $1 \leqslant t \leqslant k$. Let us write $x = (x_1, \ldots, x_t, x_{t+1}, \ldots x_n)$ and $y = (\bar{x}_1\bar{x}_2 \ldots \bar{x}_t x_{t+1} \ldots x_n)$ where $\bar{x}$ refers to the negation of the bit x. Making one step of the process ensures that all the vertices in the subcube $(* * \cdots * \,| x_{t+1}x_{t+2} \ldots x_n)$ are infected, and the claim follows by Lemma 3.2.1.

Now suppose that $k+1 \leqslant t \leqslant 2k$. Writing x and y as above, consider the vertex

$$u = (x_1x_2 \ldots x_k\bar{x}_{k+1}\bar{x}_{k+2} \ldots \bar{x}_t x_{t+1} \ldots x_n).$$

We have $d_{ux} = t - k \leqslant k$ and $d_{uy} = k$, so after one step, u will be infected. Considering the pair (u, x), we have reduced the percolation task to the previous case, so our claim is proved.

Now let $X_i, i = 1, 2, \ldots, 2k$ denote the number of initially infected pairs (x, y) with $d_{xy} = i$. Clearly,

$$\mathbb{E}\sum_{i=1}^{2k} X_i = \sum_{i=1}^{2k} 2^{n-1}\binom{n}{i}p^2.$$

For a pair $T = (x, y) \in V \times V$, let $X_T = \mathbb{1}_{\{T \text{ initially infected}\}}$. Let $\mathscr{T}_i := \{(x, y) \in V \times V : d_{xy} = i\}$.

$$\operatorname{Var}\Big(\sum_{i=1}^{2k} X_i\Big) = \sum_{i<j} \sum_{T\in\mathscr{T}_i, T'\in\mathscr{T}_j} \operatorname{Cov}(X_T, X_{T'}) + \sum_{i=1}^{2k} \sum_{T,T'\in\mathscr{T}_i} \operatorname{Cov}(X_T, X_{T'}). \tag{3.2}$$

Let $T \in \mathscr{T}_i$ and $T' \in \mathscr{T}_j$ with $i < j$. Then we have

$$\begin{aligned}
\operatorname{Cov}(X_T, X_{T'}) &= \mathbb{E}[X_T X_{T'}] - \mathbb{E}[X_T]\mathbb{E}[X_{T'}] \\
&= p^{|T\cup T'|} - p^{|T|+|T'|} \\
&= \begin{cases} 0 & \text{if } |T \cap T'| = 0 \\ p^3 - p^4 & \text{if } |T \cap T'| = 1. \end{cases}
\end{aligned}.$$

Summing over all choices of the pairs T_i, T_i', we obtain

$$\sum_{T\in T_i . T'\in T_j} \operatorname{Cov}(T, T') = \binom{n}{i}\binom{n}{j}(p^3 - p^4)2^n. \tag{3.3}$$

Now let $T, T' \in \mathscr{T}_i$. We have

$$\begin{aligned}
\operatorname{Cov}(X_T, X_{T'}) &= \mathbb{E}[X_T X_{T'}] - \mathbb{E}[X_T]\mathbb{E}[X_{T'}] \\
&= p^{|T\cup T'|} - p^{|T|+|T'|} \\
&= \begin{cases} p^2 - p^4 & \text{if } T = T' \\ p^3 - p^4 & \text{if } |T \cap T'| = 1 \\ 0 & \text{if } |T \cap T'| = 0. \end{cases}
\end{aligned}$$

Summing over all choices of the pairs T_i, T_i', we obtain

$$\sum_{T,T' \in \mathscr{T}_i} \mathrm{Cov}(T,T') = (p^2-p^4)2^n \binom{n}{i} + (p^3-p^4)2^n \binom{n}{i} \Big(\sum_{l=0}^{i-1} \binom{i}{l} \binom{n-i}{i-l} \Big). \tag{3.4}$$

Let $X = \sum_{i=1}^{2k} X_i$ (the dependence on n is suppressed). Therefore,

$$\mathbb{E}X = C(k) 2^n n^{2k} p^2.$$

To fit the assumptions of Lemma 3.2.2, we will choose p so that $\mathrm{Var}(X) = o(\mathbb{E}X^2)$. Write the variance using (3.3) and (3.4) in (3.2) to observe that

$$\mathrm{Var}(\sum_{i=1}^{2k} X_i) \leqslant C(k)(p^2 2^n n^{2k} + p^3 2^n n^{4k}).$$

It suffices to take

$$(p^3 2^n n^{4k})^{\frac{1}{2}} \ll 2^n n^{2k} p^2,$$

or equivalently

$$p \gg 2^{-n},$$

and

$$(p^2 2^n n^{2k})^{\frac{1}{2}} \ll 2^n n^{2k} p^2,$$

or equivalently

$$p \gg 2^{-\frac{n}{2}} n^{-k}.$$

The second of these conditions is more restrictive, giving the lower bound in the claim (3.1) of the theorem.

At the same time, by Markov's inequality, if $p \ll 2^{-\frac{n}{2}} n^{-k}$, then $\mathbb{E}X \to 0$ if $n \to \infty$, and thus,

percolation does not occur with probability approaching 1. This proves the upper bound in (3.1). □

3.3 CRITICAL PROBABILITY FOR THE 3-NEIGHBOR PROCESS

3.3.1 THE CASE OF $k = 2$

We start with $k = 2$ and treat general k in the second part of this section. Thus, in this part we will study $p_c(Q_{n,2}, 3)$. The following lemma gives a necessary and sufficient condition on the initially infected set $A_0 \subset V$ for the r-process to percolate.

Lemma 3.3.1. *3-neighbor percolation on $Q_{n,2}$ takes place if and only if there exist vertices $x_1, x_2, x_3 \in A_0$ with pairwise distances at most 4.*

Proof. To prove the only if part, note that if no three vertices in A_0 form a triangle with sides at most 4, then the process cannot evolve beyond the initially infected set.

To prove sufficiency, suppose that $x_1, x_2, x_3 \in A_0$ are three initially infected vertices. W.l.o.g, we can place x_1 at the origin. The proof proceeds by a case study depending on the values of the distances d_{12}, d_{13} and d_{23}. By inspection, if we disregard the order, there are 9 types of triangles that three vertices can form[2]. We will consider each of them, showing that after several steps of the process there emerges an infected subcube $Q_{2,2}$, and then percolation follows from Lemma 3.2.1.

Case 1: $(d_{12}, d_{13}, d_{23}) = (1, 1, 2)$. We already assumed that $x_1 = 0^n$. W.l.o.g. assume that $x_2 = 10^{n-1}$, and $x_3 = 010^{n-2}$. The subcube $Q_{2,2}$ will be formed by vertices x_1, x_2, x_3, x_4, where $x_4 = 110^{n-2}$. Observe that x_4 is added in the next step of the process because $x_1, x_2, x_3 \in N_{x_4}$ (recall that N_v is the neighborhood of v in $Q_{n,k}$). We write this concisely as

$$(x_1, x_2, x_3) \to x_4.$$

[2]These are all the unordered triples of integers i, j, k with $1 \leqslant i, j, k \leqslant 4$ such that $|i - j| \leqslant k \leqslant i + j$ and $i - j + k$ even.

Case 2: $(d_{12}, d_{13}, d_{23}) = (2, 2, 2)$. Assume that $x_2 = 110^{n-2}$ and $x_3 = 0110^{n-3}$. We have $Q_{2,2} = (x_2, x_3, x_4, x_5)$, where

$$x_4 = 010^{n-2}, \quad x_5 = 1^3 0^{n-3}.$$

These vertices are infected following the sequence

$$(x_1, x_2, x_3) \to x_4, \quad (x_2, x_3, x_4) \to x_5.$$

Case 3: $(d_{12}, d_{13}, d_{23}) = (2, 2, 4)$. Assume that $x_2 = 110^{n-2}$ and $x_3 = 00110^{n-4}$. We have $Q_{2,2} = (x_1, x_3, x_5, x_6)$ where

$$x_5 = 0010^{n-3}, \quad x_6 = 00010^{n-4}.$$

These vertices are infected following the sequence

$$(x_1, x_2, x_3) \to x_4, \quad (x_1, x_3, x_4) \to x_5, \quad (x_1, x_3, x_5) \to x_6,$$

where $x_4 = 0110^{n-3}$.

Case 4: $(d_{12}, d_{13}, d_{23}) = (1, 2, 3)$. Assume that $x_2 = 0010^{n-2}$ and $x_3 = 110^{n-2}$. We have $Q_{2,2} = (x_2, x_4, x_5, x_6)$, where

$$x_4 = 0110^{n-3}, \quad x_5 = 1110^{n-3},$$

$$x_6 = 1010^{n-3}.$$

These vertices are infected following the sequence

$$(x_1, x_2, x_3) \to x_4, \quad (x_2, x_3, x_4) \to x_5, \quad (x_2, x_4, x_5) \to x_6.$$

Case 5: $(d_{12}, d_{13}, d_{23}) = (2, 3, 3)$. Assume that $x_2 = 110^{n-2}$ and $x_3 = 01110^{n-4}$. We have $Q_{2,2} = (x_3, x_4, x_5, x_6)$, where

$$x_4 = 0110^{n-3}, \quad x_5 = 1110^{n-3}, \quad x_6 = 11110^{n-4}.$$

These vertices are infected following the sequence

$$(x_1, x_2, x_3) \to x_4, \quad (x_2, x_3, x_4) \to x_5, \quad (x_3, x_4, x_5) \to x_6.$$

Case 6: $(d_{12}, d_{13}, d_{23}) = (4, 4, 4)$. Assume that $x_2 = 11110^{n-4}$ and $x_3 = 0011110^{n-6}$. We have $Q_{2,2} = (x_4, x_5, x_6, x_7)$, where

$$x_4 = 00110^{n-4}, \quad x_5 = 011110^{n-5},$$
$$x_6 = 001110^{n-5}, \quad x_7 = 01110^{n-4}.$$

These vertices are infected following the sequence

$$(x_1, x_2, x_3) \to x_4, \quad (x_2, x_3, x_4) \to x_5,$$
$$(x_3, x_4, x_5) \to x_6, \quad (x_4, x_5, x_6) \to x_7.$$

Case 7: $(d_{12}, d_{13}, d_{23}) = (4, 2, 4)$. Assume that $x_2 = 11110^{n-4}$ and $x_3 = 000110^{n-5}$. We have

$Q_{2,2} = (x_3, x_4, x_5, x_6)$, where

$$x_4 = 00110^{n-4}, \quad x_5 = 00010^{n-4}, \quad x_6 = 001110^{n-5}.$$

These vertices are infected following the sequence

$$(x_1, x_2, x_3) \to x_4, \quad (x_1, x_3, x_4) \to x_5, \quad (x_3, x_4, x_5) \to x_6. \tag{3.5}$$

Case 8: $(d_{12}, d_{13}, d_{23}) = (4, 1, 3)$. Assume that $x_2 = 11110^{n-4}$ and $x_3 = 00010^{n-4}$. We have $Q_{2,2} = (x_3, x_4, x_5, x_6)$, where

$$x_4 = 00110^{n-4}, \quad x_5 = 01010^{n-4}, \quad x_6 = 01110^{n-4}.$$

The infection sequence is the same as in (3.5).

Case 9: $(d_{12}, d_{13}, d_{23}) = (4, 3, 3)$. Assume that $x_2 = 11110^{n-4}$ and $x_3 = 001110^{n-5}$. We have $Q_{2,2} = (x_3, x_4, x_5, x_6)$, where

$$x_4 = 01110^{n-4}, \quad x_5 = 00110^{n-4}, \quad x_6 = 011110^{n-5}.$$

The infection sequence is again the same as in (3.5). □

Now we are in a position to find the critical probability for the graph G with the infection threshold $r = 3$.

Theorem 3.3.2 (Critical probability for the 3-neighbor process with $k = 2$)**.** *We have*

$$\frac{1}{\omega(n)} 2^{-\frac{n}{3}} n^{-2} \leqslant p_c(Q_{n,k}, 3) \leqslant \omega(n) 2^{-\frac{n}{3}} n^{-2}. \tag{3.6}$$

Proof. Call a triple of vertices $x_1, x_2, x_3 \in Q_{n,k}$ *good* if each of $d_{12}, d_{13}, d_{23} \leqslant 4$. By Lemma 3.3.1, every vertex will be infected by the r-neighbor bootstrap percolation on $Q_{n,k}$ if and only if the starting set A_0 contains a good triple. Let X be the number of good triples in A_0. To compute $\mathbb{E}X$, we count the number of triangles whose sides (d_{12}, d_{13}, d_{23}) fall under one of the nine cases enumerated in Lemma 3.3.1. For instance, the number of triples with $(d_{12}, d_{13}, d_{23}) = (1, 1, 2)$ is at least $\frac{2^n n(n-1)}{6}$ and at most $2^n n(n-1)$ since the number of such ordered triple is $2^n n(n-1)$. Indeed, once we fix x_1, say to 0, there are n options for x_2 and $n-1$ options for x_3 to form the ordered triple. For other values of (d_{12}, d_{13}, d_{23}) the count of triples is found in a similar way, and we obtain

$$\begin{aligned}\mathbb{E}[X] \geqslant \frac{1}{6}p^3 2^n \Big[& n(n-1) + 2\binom{n}{2}(n-2) + \binom{n}{2}\binom{n-2}{2} + n\binom{n-1}{2} \\ & + 2\binom{n}{2}\binom{n-2}{2} + 4\binom{n}{4}\binom{n-4}{2} + 4\binom{n}{4}(n-4) \\ & + 4\binom{n}{4} + \binom{n}{4}\binom{4}{2}(n-4)\Big]\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[X] \leqslant p^3 2^n \Big[& n(n-1) + 2\binom{n}{2}(n-2) + \binom{n}{2}\binom{n-2}{2} + n\binom{n-1}{2} \\ & + 2\binom{n}{2}\binom{n-2}{2} + 4\binom{n}{4}\binom{n-4}{2} + 4\binom{n}{4}(n-4) \\ & + 4\binom{n}{4} + \binom{n}{4}\binom{4}{2}(n-4)\Big].\end{aligned}$$

Let $T_1 = \{x_1, x_2, x_3\}$ and $T_1' = \{x_1', x_2', x_3'\}$ be a pair of good triples. For a triple T let $X_T = \mathbb{I}_{\{T \in A_0\}}$.

We have

$$\begin{aligned}\mathrm{Cov}(X_{T_1}, X_{T_1'}) &= \mathbb{E}[X_{T_1}X_{T_1'}] - \mathbb{E}[X_{T_1}]\mathbb{E}[X_{T_1'}] \\ &= p^{|T_1 \cup T_1'|} - p^{|T_1|+|T_1'|} \\ &= \begin{cases} 0 & \text{if } |T_1 \cap T_1'| = 0 \\ p^5 - p^6 & \text{if } |T_1 \cap T_1'| = 1 \\ p^4 - p^6 & \text{if } |T_1 \cap T_1'| = 2 \\ p^3 - p^6 & \text{if } |T_1 \cap T_1'| = 3. \end{cases}\end{aligned}$$

Estimating the variance, we obtain

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{T_1, T_1'} \mathrm{Cov}(X_{T_1}, X_{T_1'}) \\ &= \sum_{|T_1 \cap T'|=1} \mathrm{Cov}(X_{T_1}, X_{T'}) + \sum_{|T_1 \cap T'|=2} \mathrm{Cov}(X_{T_1}, X_{T'}) + \sum_{|T_1 \cap T'|=3} \mathrm{Cov}(X_{T_1}, X_{T'}) \\ &\leqslant C\Big\{(p^5 - p^6)2^n \binom{n}{4}^2 \binom{n}{2}^2 + (p^4 - p^6)2^n \binom{n}{4}^2 \binom{n}{2} \\ &\qquad + (p^3 - p^6)2^n \binom{n}{4}\binom{n}{2}\Big\}.\end{aligned}$$

where C is a constant. For the first term above, the number of good triples T_1 is at most $C(k)2^n\binom{n}{4}\binom{n}{2}$ and T_1' is at most $C(k)\binom{n}{4}\binom{n}{2}$ since $|T_1 \cap T_1'| = 1$. Other terms can be derived in a similar way.

In order to satisfy the assumption that $\mathrm{Var}(X) = o((\mathbb{E}X)^2)$, it suffices to take

$$p^3 2^n \binom{n}{4}\binom{n}{2} \gg \left(p^5 2^n \binom{n}{4}^2 \binom{n}{2}^2\right)^{1/2}$$

$$p^3 2^n \binom{n}{4}\binom{n}{2} \gg \left(p^4 2^n \binom{n}{4}^2 \binom{n}{2}\right)^{1/2}$$

and

$$p^3 2^n \binom{n}{4}\binom{n}{2} \gg \left(p^3 2^n \binom{n}{4}\binom{n}{2}\right)^{1/2}.$$

A direct check shows that if $p \gg 2^{-\frac{n}{3}} n^{-2}$, then $\mathrm{Var}(X) = o((\mathbb{E}X)^2)$ and $\mathbb{E}X \to \infty$. By Lemma 3.2.2, $\mathbb{P}(X > 0) \to 1$, and then Lemma 3.3.1 implies that percolation occurs, proving the lower bound in (3.6). At the same time, if $p \ll 2^{-\frac{n}{3}} n^{-2}$, then by Markov's inequality, $\mathbb{P}(X \geqslant 1) \leqslant \mathbb{E}[X] = o(1)$. Thus with probability $\to 1$ there are no good triples in A_0, so by Lemma 3.3.1 the process does not percolate. This proves the upper bound. □

Remark 1. This theorem forms a particular case of Theorem 3.3.4. At the same time, that theorem relies on Lemma 3.3.3, which does not apply to $k = 2$, so we kept the presentation of the case $k = 2$ self-contained.

3.3.2 GENERAL $k \geqslant 3$

We begin with an extension of Lemma 3.3.1 beyond $k = 2$.

Lemma 3.3.3. *3-neighbor percolation on $Q_{n,k}$ takes place if and only if there exist vertices $x_1, x_2, x_3 \in A_0$ with pairwise distances at most $2k$.*

Proof. Without loss of generality, assume that $x_1 = 0^n$, $x_2 = 1^{a_1+a_2}0^{n-a_1-a_2}$ and $x_3 = 0^{a_1}1^{a_2+a_3}0^{n-a_1-a_2-a_3}$. The distance assumptions translate into $a_i + a_j \leqslant 2k$ with $(i, j) = (1, 2), (1, 3)$ or $(2, 3)$. The proof proceeds by case analysis in accordance with $a_2 \lessgtr k$.

1. $a_2 > k$ In this case $a_1 < k$ and $a_3 < k$.

1.1 $a_1 > 0$ and $a_3 > 0$. Our plan is to assemble an infected 2-dimensional subcube and use Lemma 3.2.1.

The subcube will be formed by the vertices, which are added to the infected subset one by one.

$$x_4 = 0^{a_1}1^k0^{n-a_1-k}, \quad x_5 = 0^{a_1-1}1^{k+1}0^{a_2-k}10^{n-a_1-a_2-1},$$
$$x_6 = 0^{a_1-1}1^{k+1}0^{n-a_1-k}, \quad x_7 = 0^{a_1}1^k0^{a_2-k}10^{n-a_1-a_2-1}.$$

We follow the writing pattern used in the proof of Lemma 3.3.1. The infection sequence is as follows:

$$(x_1, x_2, x_3) \to x_4, \quad (x_2, x_3, x_4) \to x_5,$$
$$(x_2, x_4, x_5) \to x_6, \quad (x_4, x_5, x_6) \to x_7.$$

The remaining cases follow the same pattern.

1.2 $a_1 = 0$ and $a_3 > 0$. The subcube that proves overall percolation is $Q_{k,2} = (x_4, x_5, x_6, x_7)$, where

$$x_4 = 1^k0^{n-k}, \quad x_5 = 1^{k+1}0^{n-k-1},$$
$$x_6 = 1^{k+2}0^{n-k-2}, \quad x_7 = 1^k010^{n-k-2}.$$

We have

$$\begin{aligned}(x_1, x_2, x_3) \to x_4, \quad (x_2, x_3, x_4) \to x_5,\\ (x_3, x_4, x_5) \to x_6, \quad (x_4, x_5, x_6) \to x_7.\end{aligned} \tag{3.7}$$

1.3 $a_1 > 0$ and $a_3 = 0$. The proof is the same as in the previous case by symmetry.

2. $a_2 = k$ In this case $a_1 \leqslant k$ and $a_3 \leqslant k$.

2.1 $0 < a_1 \leqslant k$ and $0 < a_3 < k$. We have $Q_{k,2} = (x_4, x_5, x_6, x_7)$, where

$$x_4 = 0^{a_1}1^{a_2}0^{n-a_1-a_2}, \quad x_5 = 0^{a_1-1}1^{a_2+1}0^{n-a_1-a_2},$$
$$x_6 = 0^{a_1}1^{a_2+1}0^{n-a_1-a_2-1}, \quad x_7 = 0^{a_1-1}1^{a_2+2}0^{n-a_1-a_2-1}$$

and these vertices are infected in 4 steps starting with x_1, x_2, x_3 and following the sequence in (3.7).

2.2 $a_1 = 0$ and $0 < a_3 < k$. We have $Q_{k,2} = (x_2, x_4, x_5, x_6)$, where

$$x_4 = 1^{k-1}0^{n-k+1}, \quad x_5 = 1^{k+1}0^{n-k-1}, \quad x_6 = 1^{k-1}010^{n-k-1}$$

and these vertices are infected in 3 steps starting with x_1, x_2, x_3 and following the sequence

$$(x_1, x_2, x_3) \to x_4, \quad (x_2, x_3, x_4) \to x_5, \quad (x_2, x_4, x_5) \to x_6. \tag{3.8}$$

2.3 $0 < a_1 \leqslant k$ and $a_3 = 0$. If $0 < a_1 < k$ the proof is the same as in 2.2 due to symmetry. Therefore we can assume that $a_1 = k$. We have $Q_{k,2} = (x_3, x_4, x_5, x_6)$, where

$$x_4 = 0^{k-1}1^k0^{n-2k+1}, \quad x_5 = 0^{k-1}1^{k+1}0^{n-k},$$
$$x_6 = 0^k1^{k-1}0^{n-2k+1},$$

and these vertices are infected in 3 steps starting with x_1, x_2, x_3 and following the sequence (3.8).

2.4 $a_1 = a_3 = k$. We have $Q_{k,2} = (x_4, x_5, x_6, x_7)$, where

$$x_4 = 0^k 1^k 0^{n-2k}, \quad x_5 = 0^{k-1} 1^{k+2} 0^{n-2k-1},$$
$$x_6 = 0^k 1^{k+1} 0^{n-2k-1}, \quad x_7 = 0^{k-1} 1^{k+1} 0^{n-2k}.$$

These vertices are infected in 4 steps starting with x_1, x_2, x_3 and following the sequence (3.7).

3. $a_2 < k$

3.1 $a_1 + a_3 \leqslant k$.

3.1.1 $a_1 > 0, a_2 > 0$, and $a_3 > 0$. We have $Q_{k,2} = (x_4, x_5, x_6, x_7)$, where

$$x_4 = 0^{a_1} 1^{a_2} 0^{n-a_1-a_2}, \quad x_5 = 0^{a_1-1} 1^{a_2+2} 0^{n-a_1-a_2-1},$$
$$x_6 = 0^{a_1} 1^{a_2+1} 0^{n-a_1-a_2-1}, \quad x_7 = 0^{a_1-1} 1^{a_2+1} 0^{n-a_1-a_2}.$$

These vertices are infected in 4 steps starting with x_1, x_2, x_3 and following the sequence in (3.7).

3.1.2 $a_1 > 0, a_2 = 0$, and $a_3 > 0$. We have $Q_{k,2} = (x_1, x_4, x_5, x_6)$, where

$$x_4 = 10^{n-1}, \quad x_5 = 010^{n-2}, \quad x_6 = 110^{n-2}.$$

These vertices are infected in a single step starting with x_1, x_2, x_3:

$$(x_1, x_2, x_3) \to x_4, \quad (x_1, x_2, x_3) \to x_5, \quad (x_1, x_2, x_3) \to x_6.$$

3.1.3 $a_1 > 0, a_2 = 0$, and $a_3 > 0$. We have $Q_{k,2} = (x_2, x_4, x_5, x_6)$, where

$$x_4 = 1^{a_2+1}0^{n-a_2-1}, \quad x_5 = 1^{a_2+2}0^{n-a_2-2}, \quad x_6 = 1^{a_2}010^{n-a_2-2}$$

and these vertices are infected in 2 steps starting with x_1, x_2, x_3 and following the sequence

$$(x_1, x_2, x_3) \to x_4, \quad (x_2, x_3, x_4) \to x_5,$$

$$(x_2, x_3, x_4) \to x_6.$$

3.1.4 $a_1 > 0, a_2 > 0$, and $a_3 = 0$. Due to symmetry the proof is the same as in Case 3.1.3.

3.2 $a_1 + a_3 > k$, and $k \geqslant 3$.

3.2.1 $a_3 > k$, $a_1 < k$, $a_2 < k - 1$ and $a_1 \geqslant a_2 + 2$. Let $x_4 = 0^{a_1-l_1}1^{a_2+l_2+l_1}0^{n-a_1-a_2-l_2}$.

The vertices x_1, x_2 and x_3 will fall in $N(x_4)$ if

$$\left.\begin{aligned} l_1 + l_2 + a_2 &\leqslant k, \\ l_1 + a_3 - l_2 &\leqslant k, \\ l_2 + a_1 - l_1 &\leqslant k. \end{aligned}\right\} \tag{3.9}$$

Let

$$l_1 = \left\lfloor \frac{a_1 - a_2}{2} \right\rfloor, \quad l_2 = k - \left\lceil \frac{a_2 + a_1}{2} \right\rceil.$$

It is easy to check that l_1 and l_2 satisfy (3.9). It is also easy to see that $a_1 \geqslant l_1$ and $a_3 \geqslant l_2$. Finally, $a_1 \geqslant a_2 + 2$ implies that $l_1 \geqslant 1$ and $a_2 + 2 \leqslant a_1 < k$ implies $l_2 \geqslant 1$.

3.2.2 $a_3 > k$, $a_1 < k$, $a_2 < k - 1$ and $a_1 \leqslant a_2$. Let $l_1 = 0$. and $l_2 = k - a_2$. It is easy to check that if $a_1 \leqslant a_2$ then l_1 and l_2 satisfy (3.9).

3.2.3 $a_3 > k$, $a_1 < k$, $a_2 < k - 1$, $a_1 = a_2 + 1$ and $a_2 \geqslant \frac{k}{3}$.

Let

$$l_1 = k - \left\lceil \frac{a_2 + a_3}{2} \right\rceil$$

and

$$l_2 = \left\lfloor \frac{a_3 - a_2}{2} \right\rfloor.$$

It is easy to check that l_1 and l_2 satisfy (3.9). Note that $l_1 \geqslant 1$ since

$$a_2 + a_3 \leqslant 2k - 2$$

and that $a_1 \geqslant l_1$ which requires

$$2a_1 + a_2 + a_3 \geqslant 2k. \tag{3.10}$$

In order for (3.10) to be satisfied it is sufficient that $a_2 \geqslant \frac{k}{3}$.

3.2.4 $a_3 > k$, $a_1 < k$, $a_2 < k - 1$, $a_1 = a_2 + 1$ and $a_2 < \frac{k}{3}$. Let $l_1 = 0$ and $l_2 = k - a_1$. It is easy to check that l_1 and l_2 satisfy (3.9). It is also easy to see that $l_2 \geqslant 2$ as long as $k \geqslant 3$, and this implies that $a_1 \leqslant k - 2$.

Now for Cases 3.2.1–3.2.4, we need to construct vertices x_5, x_6 and x_7 such that each of them will be infected by the percolation process, and together with x_4 they form a subcube $Q_{2,k}$. Given $l_1 \geqslant 1$ and $l_2 \geqslant 1$, we define these vertices in the form

$$x_5 = 0^{a_1 - l_1 + 1} 1^{a_2 + l_2 + l_1 - 1} 0^{n - a_1 - a_2 - l_2},$$

$$x_6 = 0^{a_1 - l_1} 1^{a_1 + l_2 + l_1 - 1} 0^{n - a_1 - a_2 - l_2 + 1},$$

$$x_7 = 0^{a_1 - l_1 + 1} 1^{a_2 + l_2 + l_1 - 2} 0^{n - a_1 - a_2 + 1}$$

and note that they are infected starting with x_1, x_2, x_3 and following the sequence

$$(x_1, x_3, x_4) \to x_5, \quad (x_1, x_2, x_4) \to x_6, \quad (x_4, x_5, x_6) \to x_7.$$

Now given $l_1 = 0$ and $l_2 \geqslant 2$, we have

$$x_5 = 0^{a_1}1^{l_2+a_2-1}0^{n-a_1-a_2-l_2+1}, \quad x_6 = 0^{a_1}1^{l_2+a_2-2}10^{n-a_1-a_2-l_2+1},$$
$$x_7 = 0^{a_1}1^{l_2+a_2-2}0^{n-a_1-a_2+2},$$

where these vertices are infected starting with x_1, x_2, x_3 and following the sequence

$$(x_1, x_2, x_4) \to x_5, \quad (x_1, x_2, x_4) \to x_6, \quad (x_4, x_5, x_6) \to x_7.$$

3.2.5 $a_3 > k$, $a_1 < k$ and $a_2 = k - 1$

Since $a_2 = k - 1$, then $a_3 = k + 1$. We have $Q_{k,2} = (x_4, x_5, x_6, x_7)$, where

$$x_4 = 0^{a_1}1^k0^{n-a_1-k}, \quad x_5 = 0^{a_1+1}1^k0^{n-a_1-k-1},$$
$$x_6 = 0^{a_1-1}1^{k+1}0^{n-a_1-k}, \quad x_7 = 0^{a_1+1}1^{k-1}0^{n-a_1-k},$$

and these vertices are infected starting with x_1, x_2, x_3 and following the sequence

$$(x_1, x_2, x_3) \to x_4, \quad (x_1, x_3, x_4) \to x_5,$$
$$(x_3, x_4, x_5) \to x_6, \quad (x_4, x_5, x_6) \to x_7.$$

3.2.6 $a_3 \leqslant k$, $a_1 \leqslant k$ and $a_2 > 0$. We have $Q_{k,2} = (x_4, x_5, x_6, x_7)$, where

$$x_4 = 0^{a_1}1^{a_2}0^{n-a_1-a_2}, \quad x_5 = 0^{a_1}1^{a_2+1}0^{n-a_1-a_2-1},$$
$$x_6 = 0^{a_1-1}1^{a_2+1}0^{n-a_1-a_2}, \quad x_7 = 0^{a_1-1}1^{a_2+2}0^{n-a_1-a_2-1}.$$

These vertices will be infected starting with x_1, x_2, x_3 and following the sequence

$$(x_1, x_2, x_3) \to x_4, \quad (x_1, x_3, x_4) \to x_5,$$
$$(x_2, x_4, x_5) \to x_6, \quad (x_4, x_5, x_6) \to x_7.$$

3.2.7 $a_3 \leqslant k$, $a_1 \leqslant k$ and $a_2 = 0$. We have $Q_{k,2} = (x_1, x_4, x_5, x_6)$, where the vertices

$$x_4 = 0^{a_1-1}110^{n-a_1-1}, \quad x_5 = 0^{a_1}10^{n-a_1-1},$$
$$x_6 = 0^{a_1+1}10^{n-a_1-2}$$

will be infected starting with x_1, x_2, x_3 and following the sequence

$$(x_1, x_2, x_3) \to x_4, \quad (x_1, x_3, x_4) \to x_5,$$
$$(x_1, x_4, x_5) \to x_6.$$

This concludes the proof. □

This lemma enables us to find the main term of the critical probability of percolation.

Theorem 3.3.4 (Critical probability for the 3-neighbor process)**.** *We have*

$$\frac{1}{\omega(n)}2^{-\frac{n}{3}}n^{-k} \leqslant p_c(Q_{n,k}, 3) \leqslant \omega(n)2^{-\frac{n}{3}}n^{-k}.$$

Proof. By Lemma 3.3.3, percolation occurs if and only if the initially infected set A_0 contains a triangle with all sides at most $2k$. Let X be the number of such triangles in A_0. Let $x_1 = 0^n$, $x_2 = 1^{a_1+a_2}0^{n-a_1-a_2}$, and $x_3 = 0^{a_1}1^{a_2+a_3}0^{n-a_1-a_2-a_3}$, where all the pairs (a_i, a_j) satisfy $a_i + a_j \leqslant 2k$. Then

$$\mathbb{E}[X] \leqslant p^3 2^n \sum_{\substack{a_1,a_2,a_3 \\ a_i+a_j \leqslant 2k}} \binom{n}{a_1+a_2}\binom{a_1+a_2}{a_2}\binom{n-a_1-a_2}{a_3}$$

and

$$\mathbb{E}[X] \geqslant \frac{1}{6} p^3 2^n \sum_{\substack{a_1,a_2,a_3 \\ a_i+a_j \leqslant 2k}} \binom{n}{a_1+a_2}\binom{a_1+a_2}{a_2}\binom{n-a_1-a_2}{a_3}.$$

Since k is a constant, for large n the largest term in the sum is given by $\binom{n}{2k}\binom{2k}{k}\binom{n-2k}{k}$.

Next let us estimate the variance of X. Let $T_1 = \{x_1, x_2, x_3\}$ and $T_2 = \{x'_1, x'_2, x'_3\}$ be two triangles with all the sides at most $2k$. For a triple T define the indicator random variable $X_T = \mathbb{1}_{\{T \in A_0\}}$. We have

$$\begin{aligned}
\operatorname{Cov}(X_{T_1}, X_{T_2}) &= \mathbb{E}[X_{T_1} X_{T_2}] - \mathbb{E}[X_{T_1}]\mathbb{E}[X_{T_2}] \\
&= p^{|T_1 \cup T_2|} - p^{|T_1|+|T_2|} \\
&= \begin{cases} 0 & \text{if } |T_1 \cap T_2| = 0 \\ p^5 - p^6 & \text{if } |T_1 \cap T_2| = 1 \\ p^4 - p^6 & \text{if } |T_1 \cap T_2| = 2 \\ p^3 - p^6 & \text{if } |T_1 \cap T_2| = 3. \end{cases}
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{T_1,T_2} \mathrm{Cov}(X_{T_1}, X_{T_2}) \\ &= \sum_{|T_1\cap T_2|=1} \mathrm{Cov}(X_{T_1}, X_{T_2}) + \sum_{|T_1\cap T_2|=2} \mathrm{Cov}(X_{T_1}, X_{T_2}) + \sum_{|T_1\cap T_2|=3} \mathrm{Cov}(X_{T_1}, X_{T_2}) \\ &\leqslant C(k)2^n \binom{n}{2k}\binom{n}{k}\Big[(p^5-p^6)\binom{n}{2k}\binom{n}{k} + (p^4-p^6)\binom{n}{2k} \\ &\quad + (p^3-p^6)\Big].\end{aligned}$$

where $C(k)$ is a constant that depends only on k. For the first term above, the number of good triples T_1 is at most $C(k)2^n\binom{n}{2k}\binom{n}{k}$ and T_2 is at most $C(k)\binom{n}{2k}\binom{n}{k}$ since $|T_1 \cap T_2| = 1$. Other terms can be derived in a similar way.

In order to fit the assumption that $\mathrm{Var}(X) = o((\mathbb{E}X)^2)$, it suffices to take

$$p^3 2^n \binom{n}{2k}\binom{n}{k} \gg \left(p^5 2^n \binom{n}{2k}^2 \binom{n}{k}^2\right)^{1/2}$$

$$p^3 2^n \binom{n}{2k}\binom{n}{k} \gg \left(p^4 2^n \binom{n}{2k}^2 \binom{n}{k}\right)^{1/2}$$

and

$$p^3 2^n \binom{n}{2k}\binom{n}{k} \gg \left(p^3 2^n \binom{n}{2k}\binom{n}{k}\right)^{1/2}.$$

It is easy to check that if $p \gg 2^{\frac{-n}{3}} n^{-k}$, then $\mathrm{Var}(X) = o((\mathbb{E}X)^2)$ and $\mathbb{E}X \to \infty$, and the lower bound on p_c follows by Lemma 3.2.2. On the other hand, if $p \ll 2^{\frac{-n}{3}} n^{-k}$, then by Markov's inequality, $\mathbb{P}(X \geqslant 1) \leqslant \mathbb{E}[X] = o(1)$. □

Lemma 3.3.3 gives a sufficient condition for bounding the critical probability, and one would want to

extend it to all r. Unfortunately, this extension fails, as will be shown next.

Lemma 3.3.5. *Let $r \geqslant 4$ and $rk \leqslant n$. There exists a subset $A_0 = \{x_1, x_2, \ldots, x_r\}$ of r vertices with all pairwise distances $\leqslant 2k$ such that the r-neighbor percolation process that starts with A_0 fails to infect the entire graph $Q_{n,k}$.*

Proof. We will show that the set $A_0 = \{x_1, x_2, \ldots, x_{r-1}, x_r = 0^n\}$ with $x_i = 0^{(i-1)k}1^k0^{n-ik}, i = 1, \ldots, r-1$ and $d(x_i, x_j) \leqslant 2k$ for all $1 \leqslant i, j \leqslant r$ satisfies the claim of the lemma. We claim that

$$\bigcap_{i=1}^{r} N_{x_i}(Q_{n,k}) = \emptyset.$$

Suppose the contrary and let $y \in \cap_{i=1}^{r} N(x_i)$ be a vertex common to all the neighborhoods. Suppose that the Hamming weight of the subvector $(y_{(j-1)k+1}, \ldots, y_{jk})$ is a_j, where $0 \leqslant a_j \leqslant k$ and $j = 1, \ldots, r-1$. We have

$$0 < \sum_{i=1}^{r-1} a_i \leqslant k$$

and for $j \in \{1, \ldots, r-1\}$,

$$\sum_{\substack{i=1 \\ i \neq j}}^{r-1} a_i + k - a_j \leqslant k.$$

Then we have

$$(r-3)\left(\sum_{i=1}^{r-1} a_i\right) \leqslant 0,$$

which is a contradiction since it forces $\sum_{i=1}^{r-1} a_i \leqslant 0$. □

3.4 MINIMUM PERCOLATING SETS

In this section, we study the size of the smallest percolating sets $m(r)$ for the graph $Q_{n,k}$ where r is the infection threshold and $k \geqslant 2$. Our first main result in this section is as follows.

Theorem 3.4.1. *The size of the smallest percolating set $m(r)$ satisfies*

$$m(r) = r \qquad \text{if } k = 2 \text{ and } 2 \leqslant r \leqslant 6 \tag{3.11}$$

$$m(r) \leqslant \binom{r}{m_r} \quad \text{if } k = 2 \text{ and } 10 \leqslant r \leqslant n \tag{3.12}$$

$$m(r) \leqslant \binom{n+1}{m_r} \quad \text{if } k = 2 \text{ and } n \leqslant r \leqslant \frac{n^2}{8} \tag{3.13}$$

$$m(r) \leqslant \sum_{j=0}^{k-1} \binom{r-1}{m_{k,r}+j} \qquad \text{if } k \geqslant 3,\ r = cn \text{ with } 0 < c < 1, \text{ and } r \geqslant 2^{\frac{k^2}{k-1}} \tag{3.14}$$

$$m(r) \leqslant \sum_{j=0}^{k-1} \binom{n}{m_{k,r}+j} \qquad \text{if } k \geqslant 3 \text{ and } n \leqslant r \leqslant \frac{n^k}{(2k)^k} \tag{3.15}$$

$$m(r) > r \qquad \text{if } k = 2 \text{ and } r > 15 \tag{3.16}$$

$$m(r) = r \qquad \text{if } k \geqslant 3 \text{ and } r = o(n), \tag{3.17}$$

where $m_r := \lfloor \sqrt{2r} \rfloor$ *and* $m_{k,r} = \lfloor kr^{1/k} \rfloor$.

Proof. Let us prove (3.12). Clearly, $m(r) \geqslant r$ because otherwise the initial set $\mathscr{A}_0$ will never change. Repeating the proof of Lemma 3.2.1, we claim that if the initially infected set includes the subcube $Q^0_{r-1,k}$, then the r-neighbor process will infect the entire graph. Again, in the first step, infection spreads to every vertex v in the subcube $\{*^{r-1}10^{n-r}\}$ since $|N_v \cap Q^0_{r-1,k}| = r$, and induction completes the argument.

Thus, if $Q^0_{r-1,k} \subset \mathscr{A}_0$, then percolation occurs. Denote by V the set of vertices of $Q^0_{r-1,k}$ and let us write $V = \cup_{j=0}^{r-1} V_j$, where V_j is the set of vertices of Hamming weight $j = 0, \ldots, r-1$ in $Q^0_{r-1,k}$. We claim that if $(V_{m_r-1} \cup V_{m_r}) \subset \mathscr{A}_0$, then the process propagates to the entire subcube $Q^0_{r-1,k}$. Indeed, every vertex of weight $m_r - i$ is adjacent to $\binom{r-1-m_r+i}{2}$ vertices of weight $m_r - i + 2$. By a direct check, if $r \geqslant 10$ then

$$\binom{r-1-m_r+i}{2} \geqslant r. \tag{3.18}$$

Moreover, every vertex of weight $m_r + i$ is adjacent to $\binom{m_r+i}{2}$ vertices of weight $m_r + i - 2$. We have

$$\binom{m_r + i}{2} \geqslant r. \tag{3.19}$$

Once every vertex in $V_{m_r-1} \cup V_{m_r}$ is initially infected, then in the next step every vertex on V_{m_r-2} and V_{m_r+1} will be infected. After that the infection will go on due to the monotonicity in i on the left side of (3.18) and (3.19). Therefore, we have $m(r) \leqslant \binom{r-1}{m_r} + \binom{r-1}{m_r-1}$, which shows (3.12).

Now let us prove (3.13), which is very similar to that for (3.12). First let us write $V(Q_{n,2}) = \cup_{j=0}^{n} V_j(Q_{n,2})$, where $V_j(Q_{n,2})$ is the set of vertices of Hamming weight $j = 0, \ldots, n$ in $V(Q_{n,2})$. We claim that if $(V_{m_r} \cup V_{m_r+1}) \subset \mathscr{A}_0$, the process infects the entire graph $Q_{n,2}$. Indeed every vertex of weight $m_r - i$ is adjacent to $\binom{n-(m_r-i)}{2}$ vertices of weight $m_r - i + 2$. As long as $r \leqslant \frac{n^2}{8}$,

$$\binom{n - m_r + i}{2} \geqslant r, \tag{3.20}$$

Moreover, every vertex of weight $m_r + i$ is adjacent to $\binom{m_r+i}{2}$ vertices of weight $m_r + i - 2$. We have

$$\binom{m_r + i}{2} \geqslant r. \tag{3.21}$$

Once every vertex in $V_{m_r} \cup V_{m_r+1}$ is initially infected, then in the next step every vertex on V_{m_r-1} and V_{m_r+2} will be infected. After that the infection will go on due to the monotonicity in i on the left side of (3.20) and (3.21). Therefore, we have $r \leqslant m(r) \leqslant \binom{n}{m_r} + \binom{n}{m_r+1}$, which shows (3.12).

Now let us prove (3.14). Again, if $Q^0_{r-1,k} \subset \mathscr{A}_0$, then percolation occurs. Denote by V the set of vertices of $Q^0_{r-1,k}$ and let us write $V = \cup_{j=0}^{r-1} V_j$, where V_j is the set of vertices of Hamming weight $j = 0, \ldots, r-1$ in $Q^0_{r-1,k}$. We claim that if $\cup_{j=0}^{k-1} V_{m_{k,r}+j} \subset \mathscr{A}_0$, then every vertex in $Q^0_{r_1,k}$ will be infected by the process. Indeed, every vertex of weight $m_{k,r} - i$ is adjacent to $\binom{r-1-(m_{k,r}-i)}{k}$ vertices of weight $m_{k,r} - i + k$. If

$r \geqslant 2^{\frac{k^2}{k-1}}$, by using the $\binom{n}{m} \geqslant \left(\frac{n}{m}\right)^m$, we have

$$\binom{r-1-(m_{r,k}-i)}{k} \geqslant r. \tag{3.22}$$

Moreover, every vertex of weight $m_{k,r}+i$ is adjacent to $\binom{m_{k,r}+i}{k}$. We have for $i \geqslant k$

$$\binom{m_{k,r}+i}{k} \geqslant r. \tag{3.23}$$

Once every vertex in $\cup_{j=0}^{k-1} V_{m_{k,r}+j}$ is initially infected, then in the next step every vertex on $V_{m_{k,r}-1}$ and $V_{m_{k,r}+k}$ will be infected. After that the infection will go on due to the monotonicity in i on the left side of (3.22) and (3.23). Therefore, we have $m(r) \leqslant \sum_{j=0}^{k-1} \binom{r-1}{m_{k,r}+j}$, which shows (3.14). Similarly, (3.15) can be proved.

Before proceeding, we will need the following lemma.

Lemma 3.4.2. *Suppose that $k \geqslant 2, n \geqslant r \geqslant 2$, and $2 \leqslant r \leqslant 2k$. Then*

$$m(r) = r.$$

Proof. By Lemma 3.2.1, it suffices to exhibit a set $\mathscr{A}_0$ of size r that implies percolation. First suppose that $r-1 \leqslant k$. Let $\mathscr{A}_0 = \{x_1, \ldots, x_r\}$, where $x_1 = 0^n$, $x_2 = 1^{r-1}0^{n-r+1}$, and $x_3 = 1^{r-2}0^{n-r+2}$, $x_4 = 1^{r-3}010^{n-r+1}$, etc., $x_r = 101^{r-3}0^{n-r+1}$.

Since $r-1 \leqslant k$ every binary sequences in $\{0,1\}^{r-1}$ are within Hamming distance k from $x_1, \ldots, x_r$. Therefore, every vertex on this $r-1$ dimensional cube will be infected.

Now let us consider the case where $r-1 \geqslant k$. Again let us initially infect $x_1, \ldots, x_r$. It is easy to see that every binary sequence of length $r-1$ and of weight k will be infected by the process since those sequences have $x_1, \ldots, x_r$ as their neighbors. Then every binary sequence of length $r-1$ and of weight

$k+1$ will be infected by the process and every binary sequence of length $r-1$ and of weight up to $r-2$ will be infected by the process. Therefore, every vertex in this $r-1$ dimensional cube is infected by this process. □

Returning to the proof of Theorem 3.4.1, we will now address (3.11), which has been mostly taken care of by Lemma 3.4.2. What remains is to show that $m(5) = 5$, and $m(6) = 6$. Let $x_i = 1^{i-1}0^{n-i+1}, i = 1, \ldots, 5$. We claim that if $\mathscr{A}_0$ contains these 5 vectors, the process will percolate to the entire graph. Indeed, observe that the first steps of the process propagate to the subcube $Q^0_{4,k}$, and after that Lemma 3.2.1 implies the full claim $m(5) = 5$.

Moving to the case $r = 6$, consider the set

$$\mathscr{A}_0 = \{0^n, 01^20^{n-3}, 0^21^30^{n-5}, 1^40^{n-4}, 01^40^{n-5}, 0^21^40^{n-6}\}.$$

The proof is completed by observing that the first steps infect the subcube $Q^0_{6,k}$ and using Lemma 3.2.1 to prove the full claim.

Next we address (3.16). Let $V(Q_{n,k}) = V_0 \cup V_1 \cup \cdots \cup V_n$ be the set of vertices of the cube with V_i denoting the subset of all vertices of Hamming weight i, for all i. Without loss of generality assume that $0^n \in \mathscr{A}_0$, then the remaining $r-1$ vertices in $\mathscr{A}_0$ should be of weights 1 to 4. Let $v \in V_5$, then $|N_v \cap V_4| = 5$ and $|N_v \cap V_4| = 10$. Similarly, if $w \in V_6$, then $|N_w \cap V_4| = 15$. Thus even if $\cup_{i=0}^4 V_i \subset \mathscr{A}_0$, percolation does not occur.

It remains to prove (3.17). As before, let $V_i, i = 0, \ldots, n$ be the subset of vertices of weight i. We claim that for the r-process to percolate it suffices that $|V_1 \cap \mathscr{A}_0| = r$. Indeed, since the vertices in V_0 and V_1 form a clique, the process clearly spreads to every vertex in $V_0 \cup V_1$. Next, since a vertex in V_2 is connected to all of $V_0 \cup V_1$, so the process further spreads to V_2. Continuing in this way, every vertex in V_i has at least $\binom{i}{i-1} + \binom{i}{i-2}(n-i)$ neighbors in V_{i-1} since $k \geqslant 3$. This suffices to claim percolation. □

While our first main result focuses on the upper bound of $m(r)$, our second main result addresses its lower bound.

Theorem 3.4.3. *Suppose that $k \geqslant 2$ and n is sufficiently large. If $\sum_{i=1}^{k} \binom{n'}{i} \leqslant cr^{\ell}$ and $\sum_{i=1}^{k-1} \sum_{j=1}^{k-i} \binom{n'}{j}\binom{n-n'}{i} \leqslant r - cr^{\ell}$ with $0 < \ell \leqslant 1$ and $0 < c < 1$, then*

$$m(r) \geqslant \delta 2^{n'},$$

where $\delta > 0$.

Proof. It is not difficult to see that percolation fails to occur from some initially infected set if and only if there exists a set B of healthy vertices such that every vertex in B has fewer than r neighbors outside of B. Therefore our task is to show the existence of a such set B with $|\mathscr{A}_0| = \delta 2^{n'}$ regardless of the locations of the vertices in $\mathscr{A}_0$.

Consider a partition of $V(Q_{n,k})$ into $2^{n-n'}$ parts, i.e.,

$$V(Q_{k,n}) = \bigcup_{a \in \{0,1\}^{n-n'}} [*]^{n'} a,$$

where $[*]^{n'}a$ denotes $\underbrace{* * \cdots *}_{n'} a$. Let us denote the induced subgraphs $\{G_a\}_{a \in \{0,1\}^{n-n'}}$ with $V(G_a) = [*]^{n'}a$ and $E(G_a) = \{xy : 1 \leqslant d_{xy} \leqslant k \text{ and } x, y \in V(G_a)\}$.

Note that each induced subgraph G_a is an n'-dimensional subcube and each vertex of G_a has $cr^{\ell} = \sum_{i=1}^{k} \binom{n'}{i}$ neighbors in G_a, where $0 < c < 1$. We will argue that if $|\mathscr{A}_0| = \delta 2^{n'}$ with $\delta > 0$, then there exists a set B of healthy vertices such that every vertex in that set has less than r neighbors outside that set and thus percolation does not happen.

Since $|\mathscr{A}_0| = \delta 2^{n'}$, there is a set B_a of healthy vertices of size $(1-\delta)2^{n'}$ in G_a for each $a \in \{0,1\}^{n-n'}$. Additionally, the vertices in the sets $(B_a)_{a \in \{0,1\}^{n-n'}}$ satisfy the following condition: if $x_1 x_2 \cdots x_{n'} a$ is in

B_a then $x_1x_2\cdots x_{n'}a'$ is in $B_{a'}$ for all $a,a'\in\{0,1\}^{n-n'}$. Therefore, $B=\cup_{a\in\{0,1\}^{n-n'}}B_a$.

Now let us count the number of possibly infected neighbors not in $V(G_a)$ for a vertex in B_a. W.l.o.g, we can assume that $a=[0]^{n-n'}$. Let us denote the number of possibly infected neighbors not in $V(G_a)$ for a vertex v in $B_{[0]^{n-n'}}$ by $N(v)$. It is not difficult to see that

$$N(v)\leqslant\sum_{i=1}^{k-1}\sum_{j=1}^{k-i}\binom{n'}{j}\binom{n-n'}{i}. \tag{3.24}$$

Therefore, as long as $cr^\ell+N(v)\leqslant r$, we can conclude $m(r)\geqslant\delta 2^{n'}$ where $\delta>0$. □

The following corollary highlights some implications of Theorem 3.4.3 and its proof is a direction calculation.

Corollary 3.4.4. *Let n be sufficiently large, and let $0<\delta<1$, $0<\ell<1$, and $0<c<1$. We have*

$$m(r)\geqslant\delta 2^{\sqrt{2cr}}\qquad \text{if } k=2 \text{ and } r\geqslant n^{\frac{2}{2-\ell}}\left(\frac{\sqrt{2c}}{2(1-c)}\right)^{\frac{2}{2-\ell}} \tag{3.25}$$

$$m(r)\geqslant\delta 2^{\sqrt{2cr}}\qquad \text{if } k=2 \text{ and } r\geqslant \frac{n^2}{\left(\sqrt{2c}+\frac{(1-c)}{\sqrt{2c}}\right)^2} \tag{3.26}$$

$$m(r)\geqslant\delta\exp\left(r^{\ell/k}\frac{k}{e}(\frac{c}{k})^{1/k}\right)\qquad \text{if } k\geqslant 3 \text{ and } r\geqslant n^{\frac{k(k-1)}{k-\ell}}\left(\frac{e}{k-1}\right)^{\frac{(k-1)k}{k-\ell}}\left(\frac{c}{2}\right)^{\frac{k}{k(k-\ell)}\frac{1}{1-c}} \tag{3.27}$$

$$m(r)\geqslant\delta\exp\left(r^{1/k}\frac{k}{e}\left(\frac{c}{2}\right)^{1/k}\right)\qquad \text{if } k\geqslant 3 \text{ and } r\geqslant n^k\left(\frac{2}{1-c}\left(\frac{e}{k-1}\right)^{k-1}\left(\frac{c}{2}\right)^{1/k}\frac{k}{e}\right)^{\frac{k}{k-1}}. \tag{3.28}$$

Proof. Let us show (3.25). For $k=2$, we have $\binom{n'}{2}+n'\leqslant cr^l$ and thus $n'\leqslant\sqrt{2cr}$. From (3.24), we have $(n-n')n'\leqslant(1-c)r$ and thus $n\leqslant\sqrt{2cr^\ell}+\frac{(1-c)r^{1-\ell/2}}{\sqrt{2c}}\leqslant\frac{2(1-c)r^{1-\ell/2}}{\sqrt{2c}}$ since $1-\frac{\ell}{2}\geqslant\frac{\ell}{2}$ and n is sufficiently large. We have the desired result.

(3.26) can be proved very similarly.

Now let us prove (3.27). For $k\geqslant 3$, we have $\sum_{i=1}^{k}\binom{n'}{i}\leqslant cr^\ell$ and thus $n'\leqslant r^{\ell/k}\frac{k}{e}\left(\frac{c}{2}\right)^{1/k}$. From (3.24), we have $\sum_{i=1}^{k-1}\sum_{j=1}^{k-i}\binom{n'}{j}\binom{n-n'}{i}\leqslant(1-c)r$. It suffices to have $2\binom{n-n'}{k-1}n'\leqslant(1-c)r$ since n is sufficiently

large.

(3.28) can be proved similarly. □

3.5 DISCUSSION AND OPEN PROBLEMS

We conclude with several problems left open in our research.

1. What is $m(Q_{n,k}, r)$ when $k \geqslant 2$ and $r = r(n)$ is large ?

   We conjecture that if $r = r(n)$ is a relatively fast growing function of n, then $m(Q_{n,k}, r) \geqslant \delta 2^{n'}$ where n' satisfies that $\sum_{i=1}^{k} \binom{n'}{i} \leqslant cr$ with $0 < c < 1$ and $\delta > 0$. Note that if this holds, then we obtain an exponentially tight bound for $m(Q_{n,k}, r)$ over a wide range values of r, whereas Theorem 3.4.3 gives such a bound only when $r = \theta(n^k)$.

   In order to explain our conjecture it is necessary to introduce the notion of the weak saturation introduced by Bollobás [17]. Given fixed graphs G and H, a spanning subgraph F of G is weakly (G, H)-saturated if the edges of $E(G)\backslash E(F)$ can be added to F, one edge at a time, in such a way that each edge creates a new copy of H when it is added. The weak saturation number of H in G is defined by

   $$\text{wsat}(G, H) := \min\{|E(F)| : F \text{ is weakly } (G, H)\text{-saturated}\}.$$

   Our conjecture is based on the following bound:

   $$m(G, r) \geqslant \frac{\text{wsat}(G, S_{r+1})}{r}, \tag{3.29}$$

   where S_{r+1} denotes a star with $r + 1$ rays.

   In [42], Morrison and Noel used bound (3.29) to determine tight asymptotics for $m(G, r)$, where G is the n-dimensional Boolean cube and r is a fixed constant. In their proof they relied on a linear-

algebraic method developed in [13] to obtain an asymptotically tight lower bound for wsat(G,S_{r+1}). We conjecture that wsat$(Q_{n,k}, S_{r+1}) \geqslant \delta r 2^{n'}$. To support this conjecture, consider a partitioning of $Q_{n,k}$ into $2^{n-n'}$ isomorphic parts so that each induced subgraph is an n'-dimensional subcube. Each vertex of an n'-dimensional subcube has $cr = \sum_{i=1}^{k} \binom{n'}{i}$ neighbors, and the total number of edges in the subcube is $2^{n'-1}cr$. We believe that the lower bound wsat$(Q_{n,k}, S_{r+1}) \geqslant \delta 2^{n'}$ can be proved relying on a version of the linear algebraic method.

2. What is $p_c(G, r)$ for a fixed k and r?

   Although Lemma 3.3.3 does not hold when $r \geqslant 4$, it might still be possible to derive asymptotics of the critical probability using alternative characterizations of percolating sets.

Chapter 4

BOOTSTRAP PERCOLATION ON A GENERALIZED HAMMING CUBE, II

4.1 INTRODUCTION

While majority bootstrap percolation on the hypercube has been studied in [11] and [26], where the infection threshold $r = \frac{1}{2}n$, the case of infection threshold $r = n^a$ with $a < 1$ has not been investigated prior to our work. In fact, this has been mentioned as an open question in Section 9 of [26]. By utilizing key techniques from [11] with some new ideas, we extend the results established in [11] to the case where the infection threshold $r = n^a$ with $a < 1$. More precisely, we prove the following two theorems.

Theorem 4.1.1. *As* $n \to \infty$,

$$p_c(Q_n, n^a) = n^{a-1} + o(n^{a-1}),$$

where $\frac{2}{3} < a < 1$.

In fact if $\frac{3}{4} < a < 1$, more precise information about the second-order term of the critical probability can be derived.

Theorem 4.1.2. *As* $n \to \infty$,

$$n^{a-1} - n^{\frac{a}{2}-1+\delta} \leqslant p_c(Q_n, n^a) \leqslant n^{a-1} - n^{\frac{a}{2}-1},$$

where $\frac{3}{4} < a < 1$ *and* $\delta > \frac{1}{2}(1-a)$.

Holroyd, Liggett, and Romik [38] considered percolation on a generalized grid graph with each vertex having more than four neighbors. Inspired by their work we consider an n-dimensional hypercube graph with extended neighborhoods and prove estimates for the critical probability of majority bootstrap percolation. Again this has been mentioned as an open question in Section 9 of [26].

Let $Q_{k,n}$ be a graph with vertices $V(Q_{k,n}) = \{0,1\}^n$ and edges $E(Q_{k,n}) = \{xy : d_H(x,y) = 1, 2, \cdots, k\}$, where $k \geqslant 2$. We prove the following theorem.

Theorem 4.1.3. *Consider a positive integer* $k \geqslant 2$ *such that* k *divides* n*, and let* $N = \sum_{i=1}^{k} \binom{n}{i}$*.Then*

$$\frac{1}{2} - \frac{1}{n^a} \leqslant p_c\left(Q_{k,n}, \frac{N}{2}\right) \leqslant \frac{1}{2} - \left(\frac{1}{n}\right)^{\frac{k}{2}},$$

where $a < 1$.

Moreover, for $k = 2$*, we have*

$$\frac{1}{2} - \frac{C\sqrt{\log n}}{n} \leqslant p_c\left(Q_{2,n}, \frac{N}{2}\right) \leqslant \frac{1}{2} - \frac{1}{n},$$

where C *is a sufficiently large constant independent of* n.

4.2 PROOF OF THEOREM 1

In order to prove Theorem 4.1.1, we will prove several lemmas.

We will use the standard Chernoff bound (See Remark 2.5 in [39]) which was also used in [11].

Lemma 4.2.1. *Let* $n \in \mathbb{N}$, $0 < p < 1$, $t > 0$ *and* $S(n) \sim$ *Bin*(n,p)*. Then,*

$$\mathbb{P}(S(n) \geqslant np + t) \leqslant \exp\left(-\frac{2t^2}{n}\right)$$

and

$$\mathbb{P}(S(n) \leqslant np - t) \leqslant \exp\left(-\frac{2t^2}{n}\right).$$

Fact 4.2.2. *If $S \subset V(Q_n)$ satisfies $d_H(y,z) \geqslant 2k+1$ for every $y, z \in S$, then the events $\{y \in A_k\}_{y \in S}$ are independent.*

We will also use the following three lemmas which were proved in [10].

Lemma 4.2.3. *Let $p \in (0,1)$ and $n \in \mathbb{N}$ satisfy $pn^2 < 1$, and let $S(n) \sim Bin(n,p)$. Then*

$$\mathbb{P}(S(n) \geqslant m) \leqslant 2p^{\frac{m}{2}},$$

for every $m \in [n]$. In particular, if $c, \epsilon > 0$ and $p \leqslant e^{-cn}$, then for $b = b(x,\epsilon)$ not depending on n,

$$\mathbb{P}(S(n) \geqslant \epsilon n) \leqslant e^{-bn^2}.$$

Lemma 4.2.4. *Let G be a graph , let $k, m \in \mathbb{N}$, and suppose that for each $x \in V(G)$,*

$$B(x,k) := |y \in V(G) : d(x,y) \leqslant k| \leqslant m.$$

Then there exists a partition

$$V(G) = B_1 \cup \cdots \cup B_m$$

of $V(G)$, such that if $y, z \in B_i$, then $d(y,z) \geqslant k+1$. Here, $d(x,y)$ denotes the graph-theoretic distance between the vertices x and y.

Lemma 4.2.5. *Let $n, k \in \mathbb{N}$, and $x \in V(Q_n)$. Then there exists a partition*

$$S(x,k) := \{y \in V : d_H(y,x) = k\} = B_1 \cup \cdots \cup B_m$$

of $S(x,k)$ *into* $m \leqslant k\binom{n}{k-1} \leqslant 2n^{k-1}$ *sets, such that if* $x, y \in B_j$ *for some* j*, then* $d_H(x,y) \geqslant 2k$*.*

The key observation is that after running the process for only a few steps, the probability of any uninfected vertices remaining uninfected becomes exponentially small in n. Hence, it suffices to analyze the initial few steps of the process. Following Lemma 4.3 in [11], we prove the following lemma.

Lemma 4.2.6. *Let* $c > 0$*,* $\delta > 0$*, and* k *be absolute constants. Suppose that initially, each vertex is infected with probability* p*, independent of the other vertices. Suppose that for every* $x \in V(Q_n)$

$$\mathbb{P}(x \notin A_k) \leqslant e^{-cn^{\delta}}.$$

Then there exists a constant $d > 0$ *such that*

$$\mathbb{P}(x \notin A_{2k+1}) \leqslant e^{-dn^{1+\delta}}.$$

Proof. Suppose that $x \notin A_{2k+1}$. Let $S(x,l) = \{y \in V : d_H(y,x) = l\}$. We claim that for each $t \in [0, k+1]$, there exist a set $T(t) \subset S(x,t)$ and a function $0 < \alpha(t) < 1$ such that

$$T(t) \cap A_{2k+1-t} = \emptyset, \tag{4.1}$$

and

$$|T(t)| \geqslant \alpha(t)n^t - O(n^{t-1}). \tag{4.2}$$

We will prove the claim by induction on t. In the base step, we have $t = 0$ and so the claim is trivially true. Suppose that $T(t)$ satisfies (4.1) and (4.2) for some $t \geqslant 1$. Then every $y \in T(t)$ has at most n^a neighbors in $S(x,t+1) \cap A_{2k-t}$ and so at least $n - t - n^a$ neighbors in $S(x,t+1) \backslash A_{2k-t}$. Each element

of $S(x, t+1)$ has $t+1$ neighbors in $S(x,t)$. Therefore, there exists a set $T(t+1) \subset S(x,t+1)$ such that $T(t+1) \cap A_{2k-t} = \emptyset$ and

$$|T(t+1)| \geqslant \frac{(n-t-n^a)|T(t)|}{t+1}.$$

Therefore, we have

$$|T(t+1)| \geqslant \alpha(t+1)n^{t+1} - O(n^t),$$

which concludes the proof of the claim.

From Lemmas 4.2.4 and 4.2.5, we can observe that there exists a partition of $V(Q_n)$ as follows:

$$V = \cup_{i=1}^{m} B_i,$$

where, for each $i \in [m]$, and for any x and $y \in B_i$, we have $d_H(x,y) \geqslant 2k+1$. Furthermore, the number of subsets m in this partition satisfies $m \leqslant 2n^k$.

Now we claim that there exists $j \in [m]$ and a sufficiently small $\epsilon > 0$ such that

$$|B_j| \geqslant \epsilon n$$

and

$$|T(k+1)| \geqslant \epsilon |B_j|.$$

Otherwise

$$|T(k+1)| \leqslant \epsilon mn + \epsilon \binom{n}{k+1},$$

which is a contradiction since ϵ can be taken to be arbitrarily small.

Recall that $T(k+1) \cap A_k = \emptyset$ and $\mathbb{P}(x \notin A_k) \leqslant e^{-cn^\delta}$. We have, by Lemma 4.2.3,

$$\begin{aligned}\mathbb{P}\left(|T(k+1) \cap B_j| \geqslant \epsilon |B_j|\right) &\leqslant 2(e^{-cn^\delta})^{\epsilon |B_j|/2} \\ &\leqslant 2(e^{-cn^{\delta+1}\epsilon^2/2}).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P}(x \notin A_{2k+1}) &\leqslant \mathbb{P}[\exists j \text{ with } |B_j| \geqslant \epsilon n \text{ and } |T(k+1) \cap B_j| \geqslant \epsilon |B_j|] \\ &\leqslant m2(e^{-cn^{1+\delta}\epsilon^2/2}) \\ &\leqslant e^{-dn^{1+\delta}}\end{aligned}$$

for some $d > 0$. □

In order to proceed we will introduce a process closely related to the bootstrap percolation process (which was originally introduced in [11]). The only difference is the less strict requirement for a vertex to be infected at step 1. Let A_i denote the set of infected vertices up to step i. We define the process Boot1(t) as follows:

* Initially, each vertex is infected with probability p, independent of other vertices.

* If a vertex becomes infected then it remains infected forever.

*A healthy vertex will be infected at step 1 if it has at least $n^a - t$ infected neighbors at the initial step i.e., $x \in A_1$ if $x \in A_0$ or $|\mathcal{N}(x) \cap A_0| \geqslant n^a - t$.

* Let $i \geqslant 1$. A healthy vertex will be infected at step $i+1$ if it has at least n^a infected neighbors at step i, i.e., $x \in A_{i+1}$ if $x \in A_i$ or $|\mathcal{N}(x) \cap A_i| \geqslant n^a$.

Let us refer to the original n^a-neighbor process as Boot. Note that the process Boot1(t) dominates the process Boot, in the sense that given the same initial set A_0, then for each $k \in \mathbb{N}$, the set A_k given by

Boot1(t) contains that given by Boot. Therefore, if percolation does not happen (with high probability) in the process Boot1(t) then percolation will not happen in the process Boot as long as the initial probability for each vertex to be infected is the same for both processes.

Lemma 4.2.7. *Let $\epsilon_1, \epsilon_2 > 0$, $\epsilon_1 > 2\epsilon_2$ and $\frac{2}{3} < a < 1$. Consider the Boot1(t) process, with $p = n^{a-1} - \epsilon_1 n^{a-1}$ and $t = \epsilon_2 n^a$. Then $A_2 = A_1$ with high probability in the process Boot1(t).*

Proof. Let $x \in V(Q_n)$ and suppose $x \in A_2 \backslash A_1$. Then by definition we have

$$|\mathcal{N}(x) \cap A_1| \geqslant n^a$$

and

$$|\mathcal{N}(x) \cap A_0| \leqslant n^a - t.$$

Thus there exists a set $T \subset \mathcal{N}(x)$ with $|T| = t$, and $T \subset A_1 \backslash A_0$.

Let $S(x, l) = \{y \in V : d_H(y, x) = l\}$ and consider the set $\mathcal{N}(T) \cap S(x, 2)$. It has a set (denoted as B) of $\binom{t}{2}$ vertices with two neighbors in T, and a set (denotes as C) of $t(n - t)$ elements with one neighbor in T.

Since $T \subset A_1 \backslash A_0$, we have

$$|C \cap A_0| \geqslant t(n^a - t) - 2\binom{t}{2} \tag{4.3}$$

$$= tn^a - 2t^2 + t \tag{4.4}$$

$$= \epsilon_2 n^{2a} - 2\epsilon_2^2 n^{2a} + \epsilon_2 n^a. \tag{4.5}$$

Indeed, this follows by counting the number of edges between T and $\mathcal{N}(T) \cap S(x, 2)$. Note that every element in T has at least $n^a - t$ neighbors in A_0 and even if every element in B is in A_0 (4.4) is still valid.

However, since $|C \cap A_0| \sim \text{Bin}(t(n-t), p)$, we have

$$|C \cap A_0| - \mathbb{E}[|C \cap A_0|] \geqslant -2\epsilon_2^2 n^{2a} + \epsilon_2 n^a + n^{3a-1}\epsilon_2^2 + \epsilon_1\epsilon_2 n^{2a} - \epsilon_1\epsilon_2^2 n^{3a-1}.$$

By Lemma 4.2.1 and the fact that $\frac{2}{3} < a < 1$, we have

$$\begin{aligned}\mathbb{P}\left(|C \cap A_0| \geqslant t(n^a - t) - 2\binom{t}{2}\right) &\leqslant \exp\left(-\frac{2(\epsilon_1\epsilon_2 - 2\epsilon_2^2)^2 n^{4a}}{t(n-t)}\right) \\ &\leqslant \exp\left(-\frac{2(\epsilon_1\epsilon_2 - 2\epsilon_2^2)^2 n^{3a-1}}{\epsilon_2}\right).\end{aligned}$$

Since we have at most $\binom{n}{n^a}$ choices for the set T, thus for some $c > 0$,

$$\mathbb{P}\left(x \notin A_2 \backslash A_1\right) \leqslant \binom{n}{n^a}\mathbb{P}\left(|C \cap A_0| \geqslant t(n^a - t) - 2\binom{t}{2}\right) \leqslant \exp(-cn^{3a-1}).$$

Moreover, since $|V(Q_n)| = 2^n$, we have

$$\mathbb{P}(A_2 \backslash A_1 \neq \emptyset) \leqslant 2^n \mathbb{P}(x \in A_2 \backslash A_1) \leqslant 2^n \exp\left(-cn^{3a-1}\right) = o(1). \quad \square$$

Now we are ready to prove Theorem 4.1.1.

Proof. Let us first show the upper bound. Let the probability p for every vertex to be initially infected be $n^{a-1} + \epsilon n^{a-1}$ assuming $\epsilon > 0$, and $x \in V(Q_n)$, by Lemma 4.2.1 we have,

$$\begin{aligned}\mathbb{P}(x \notin A_1) &= \mathbb{P}(\{x \notin A_0\} \cap \{|N(x) \cap A_0| \leqslant n^a\}) \\ &\leqslant \mathbb{P}(|\mathcal{N}(x) \cap A_0| \leqslant n^a) \\ &= \mathbb{P}(\text{Bin}(n, p) \leqslant n^a) \\ &\leqslant \exp(-2\epsilon^2 n^{2a-1}).\end{aligned}$$

By Lemma 4.2.6, with $\delta = 2a - 1 > 0$ we have

$$\mathbb{P}(x \notin A_3) < \exp(-dn^{2a}).$$

Therefore, since $\frac{2}{3} < a < 1$

$$\mathbb{P}\left(\exists x \in Q_n \text{ such that } \ x \notin A_3\right) \leqslant 2^n \exp(-dn^{2a}) \leqslant o(1).$$

Now we will move on to the lower bound. Consider the process Boot1(t) and let p be the probability that each vertex is initially infected, where $p = n^{a-1} - \epsilon n^{a-1}$. If we can prove $\mathbb{P}(A_1 = V(Q_n)) = o(1)$ for Boot1(t), then together with Lemma 4.2.7, this would imply that the process Boot1(t) does not percolate with high probability, and therefore the process Boot does not percolate with high probability. Indeed, We have

$$\begin{aligned}
\mathbb{P}(x \notin A_1) &\geqslant \mathbb{P}(x \notin A_0)\mathbb{P}(|N(x) \cap A_0| < n^a) \\
&\geqslant 0.9(1 - o(1)) \\
&\geqslant 0.8.
\end{aligned}$$

From Fact 4.2.2 it is clear that if $S \subset V(Q_n)$ satisfies $d_H(y, z) \geqslant 3$ for every $y, z \in S$, then the events $\{y \in A_1\}_{y \in S}$ are independent.

From Lemmas 4.2.4 and 4.2.5, we can observe that there exists a partition of V as follows:

$$V = \cup_{i=1}^{m} B_i,$$

where, for each $i \in [m]$ and for any x and $y \in B_i$, we have $d_H(x, y) \geqslant 3$. Furthermore, the number of

subsets m in this partition satisfies $m = \sum_{i=1}^{3} \binom{n}{i}$.

Thus there exists a set B_i with $|B_i| \geqslant \frac{2^n}{n^4}$ such that for any x and $y \in B_i$, we have $d_H(x, y) \geqslant 3$. Therefore, with the independence condition in place, we have

$$\mathbb{P}(x \in A_1 \text{ for all } x \in B_i) \leqslant (0.2)^{|B_i|}.$$

Thus,

$$\mathbb{P}(A_1 = V(Q_n)) = o(1).$$

The theorem is proved. □

4.3 PROOF OF THEOREM 2

Let's begin by proving the upper bound. To do so, we require a slightly more general concept. Let $A_{i,r}$ represent the set of infected vertices up to and including step i when the infection threshold is r. In order to establish the upper bound, we will rely on a few lemmas.

The key observation is that after running the process for only a few steps, the probability of any non-infected vertex remaining noninfected becomes exponentially small in n. Hence, it suffices to analyze the initial few steps of the process.

We need to introduce more tools before proceeding.

Lemma 4.3.1. *(De Moivre-Laplace Theorem) Let $S(n) \sim Bin(n, p)$. Then*

$$\lim_{n\to\infty} \mathbb{P}\{\frac{S(n) - np}{\sqrt{np(1-p)}} \leqslant z\} = \Phi(z),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Before stating the next lemma we need to introduce some concepts and notation.

Suppose that $\Omega = \{0,1\}^{2^n}$, and each coordinate is assigned independently. In other words, we have a collection of independent Bernoulli random variables $X_1, ..., X_{2^n}$.

Define a partial order on the elements of Ω as follows:

$$(x_1, ..., x_{2^n}) \geqslant (y_1, ..., y_{2^n})$$

if and only if $x_i \geqslant y_i$ for all $1 \leqslant i \leqslant 2^n$.

We say that an event $A \subset \Omega$ is increasing(decreasing) if $x \in A$ and $y \geqslant x (y \leqslant x)$ implies that $y \in A$.

There is a one-to-one correspondence between the initial infection configuration and Ω. Let us enumerate all the vertices in $V(Q_n)$ as $v_1, v_2, \cdots, v_{2^n}$ and denote $x_i = 1$ if v_i is initially infected and otherwise $x_i = 0$.

Lemma 4.3.2. *(FKG Inequality [29]) In the above setting, if both events A and B are increasing (or decreasing), then*

$$\mathbb{P}(A \cap B) \geqslant \mathbb{P}(A)\mathbb{P}(B).$$

Lemma 4.3.3. *Let $x \in V(Q_n)$ and let the initial infection probability $p = n^{a-1} - n^{\frac{a}{2}-1}$ with $\frac{3}{4} < a < 1$. We set the infection threshold to $r = n^a + 3$. Then there exists an absolute constant $\delta > 0$ such that as $n \to \infty$,*

$$\mathbb{P}(x \in A_{1,r}) \geqslant \delta.$$

Proof. We have

$$\begin{aligned}\mathbb{P}(x \in A_{1,r}) &= \mathbb{P}(x \in A_0) + \mathbb{P}(|N(x) \cap A_0| \geqslant r | x \notin A_0)\mathbb{P}(x \notin A_0) \\ &= \mathbb{P}(x \in A_0) + \mathbb{P}(|N(x) \cap A_0| \geqslant r)\mathbb{P}(x \notin A_0).\end{aligned}$$

Let $S(n) = |N(x) \cap A_0|$ and note that $S(n) \sim \text{Bin}(n, p)$. Then, we have

$$\mathbb{P}(S(n) \geqslant r) = \mathbb{P}\left(\frac{S(n) - np}{\sqrt{np(1-p)}} \geqslant \frac{r - np}{\sqrt{np(1-p)}}\right).$$

Let $z(n) = \frac{r-np}{\sqrt{np(1-p)}}$. Then it is easy to see that

$$\lim_{n\to\infty} z(n) = 1.$$

By Lemma 4.3.1, we have

$$\lim_{n\to\infty} \mathbb{P}(S(n) \geqslant r) = 1 - \Phi(1).$$

Therefore there exists an absolute constant $\delta > 0$

$$\mathbb{P}(x \in A_{1,r}) \geqslant \delta. \quad \square$$

Lemma 4.3.4. *Let $\delta > 0$. Let p be the same as that in Lemma 4.3.3 and then $\mathbb{P}(x \in A_{1,n^a+3}) \geqslant \delta$. Then there exist a constant $c > 0$ and $\delta' > 0$ such that*

$$\mathbb{P}(x \in A_{4,n^a}) \geqslant 1 - e^{-cn^{\delta'}}$$

Proof. Let $x \in V(Q_n)$. There exists a set S,$|S| = n^a$ such that $S \subset \mathcal{N}(x)\backslash A_0$. Otherwise, it follows that x belongs to A_1. Let $T = \mathcal{N}(S)\backslash\{x\}$. Denote $T_i = \{y \in T : |\mathcal{N}(y) \cap S| = i\}$ for $i \in \{1, 2\}$. It is easy to see that

$$|T_1| = (n - n^a)n^a$$

and

$$|T_2| = \binom{n^a}{2}.$$

Let $|T_1 \cap A_2| = b$ and $|T_2 \cap A_2| = c$. Note that if $S \cap A_3 = \emptyset$, then $b + 2c < n^{2a}$.

Note that $|T_1 \cap A_0| \sim \text{Bin}(n^{1+a} - n^{2a}, n^{a-1} - n^{\frac{a}{2}-1})$. By Lemma 4.2.1, there exits $c' > 0$ such that

$$\mathbb{P}(||T_1 \cap A_0| - n^{2a}| \geqslant 2n^{3a-1}) \leqslant \exp(-c'n^{5a-3}).$$

Let $F(S)$ denote this event that $||T_1 \cap A_0| - n^{2a}| \leqslant 2n^{3a-1}$ and thus we have $\mathbb{P}(F(S)) \geqslant 1 - \exp(-c'n^{5a-3})$. From now on we assume the event $F(S)$ occurs. Note that if $|T_1 \cap A_2 \backslash A_0| \geqslant 8n^{3a-1}$, then $b \geqslant n^{2a}$ and thus $S \cap A_3 \neq \emptyset$.

Claim 4.3.5. *If $F(S)$ holds, then*

$$\mathbb{P}(|T_1 \cap A_2 \backslash A_0| = O(n^{3a-1})) \leqslant 9n \exp(-c''\delta^2 \epsilon n^{1+a}),$$

where $c'' > 0$.

Proof. Consider a bipartite graph H with $V(H) = W \cup Y$ where $W = T_1 \backslash A_0$, $Y = S(x, 3)$ and $E(H) = \{wy : wy \in E(Q_n)\}$. Color an edge of H red if its end in the set Y is in A_1 and color an edge of H blue if its end in the set Y is not in A_1.

It is easy to see that $|E(H)| = (n-2)|W|$. Since $F(S)$ holds, we have $|W| = (n - n^a)n^a - n^{2a} + O(n^{3a-1})$.

Now suppose $|T_1 \cap A_2 \backslash A_0| = O(n^{3a-1})$. Since only $O(n^{3a-1})$ vertices in $T_1 \backslash A_0$ can have at least n^a neighbors in Y, we have

$$e_R(H) \leqslant n^a |W| + O(n^{3a})$$

and

$$e_B(H) \geqslant (n-2)|W| - n^a|W| - O(n^{3a}), \tag{4.6}$$

where $e_R(H)$ and $e_B(H)$ denotes the number of red edges and blue edges in H respectively.

From Lemmas 4.2.4 and 4.2.5 , we can observe that there exists a partition of the set $S(x,3)$ as follows:

$$S(x,3) = \cup_{i=1}^{m} B_i$$

where, for each $i \in [m]$, and for any x and $y \in B_i$, we have $d_H(x,y) \geqslant 3$. Furthermore, the number of subsets m in this partition satisfies $m \leqslant 3n$.

Let us define for $i \in [3]$ and $j \in [m]$

$$B_j(i) = \{x \in B_j : |N(x) \cap T_1| = i\}.$$

Claim 4.3.6. *For a sufficiently small $\epsilon > 0$, then there exists $i \in [3]$ and $j \in [m]$ such that*

(a) l edges of H are incident with $B_j(i)$ where $l \geqslant \epsilon n^{1+a}$ and

(b) at most $\frac{\delta}{2}l$ of those edges are red.

Proof. Suppose the claim is false. We will find a contradiction by counting the number of blue edges. Indeed, we have at most $3m\epsilon n^{1+a} + e(H)(1-\frac{\delta}{2}) \leqslant 3m\epsilon n^{1+a} + (1-\frac{\delta}{2})n|W|$ blue edges. Since ϵ is small enough, this is a contradiction to the bound in (4.6). □

Note that $\{y \in A_1\}_{y \in B_j(i)}$ are independent but this event is not independent of which members of $B(x,2)$ are in A_0. Since $y \in B_j(i)$ has at most 3 neighbors in $B(x,2)$ and $\mathbb{P}(y \in A_{1,n^a+3}) \geqslant \delta$, then $\mathbb{P}(y \in A_{1,n^a}) \geqslant \delta$ for any set $B(x,2) \cap A_0$.

Let $E_j(i)$ denote the event that $B_j(i)$ satisfies (a) and (b). Then

$$\begin{aligned}\mathbb{P}(\cup_{i,j} E_j(i)) &\leqslant \sum_{i,j} \mathbb{P}\left(\text{Bin}\left(\frac{\epsilon n^{1+a}}{i}, \delta\right) \leqslant \frac{\delta}{2}\frac{\epsilon n^{1+a}}{i}\right) \\ &\leqslant 9n\exp(-c''\delta^2\epsilon n^{1+a}).\end{aligned}$$

for $c'' > 0$. □

The number of choices for S is at most

$$\begin{aligned}\binom{n}{n^a} &\leqslant \left(\frac{en}{n^a}\right)^{n^a} \\ &= e^{n^a + n^a\log n - an^a\log n}.\end{aligned}$$

Let $M(S)$ denote the event $|T_1 \cap (A_2 \backslash A_0)| \geqslant 8n^{3a-1}$. Note that if for every $S \subset \mathcal{N}(x)\backslash A_0$ with $|S| = n^a$ both $F(S)$ and $M(S)$ hold, then $x \in A_4$.

Therefore,

$$\mathbb{P}(x \notin A_4) \leqslant \sum_S \mathbb{P}[F(S)^c] + \sum_S \mathbb{P}[M(S)^c] \tag{4.7}$$

$$\leqslant \exp(-c'''n^{\delta'}), \tag{4.8}$$

where $c''' > 0$ and $\delta' > 0$.

It is easy to check that (4.8) is valid since $a > \frac{3}{4}$. □

Now we are in a position to prove the upper bound in Theorem 4.1.2

Proof. By Lemmas 4.3.3, 4.3.4, and 4.2.6 there exists $d > 0$ and $\delta'' > 0$ such that

$$\begin{aligned}\mathbb{P}(A_0 \text{ does not percolate}) &\leqslant \mathbb{P}(\cup_{x\in V(G)}(x \notin A_9)) \\ &\leqslant \sum_{x\in V(G)} \mathbb{P}(x \notin A_9) \\ &\leqslant 2^n e^{-dn^{1+\delta''}} \\ &= o(1).\end{aligned}$$

□

In order to prove the lower bound on the probability that percolation does not happen we will introduce a new process closely related to the bootstrap percolation process (which was originally introduced in [11]). The only difference is the less strict requirement for a vertex to be infected at steps 1 and 2. Let A_i denote the set of vertices infected up to step i. We define the process Boot2(t) as follows:

* Initially, each vertex is infected with probability p, independent of the other vertices.

* If a vertex becomes infected, then it remains infected forever.

*A healthy vertex will be infected at step 1 if it has at least $n^a - 2t$ infected neighbors at the initial step i.e., $x \in A_1$ if $x \in A_0$ or $|\mathcal{N}(x) \cap A_0| \geqslant n^a - 2t$.

*A healthy vertex will be infected at step 2 if it has at least $n^a - t$ infected neighbors at the first step i.e., $x \in A_2$ if $x \in A_1$ or $|\mathcal{N}(x) \cap A_1| \geqslant n^a - t$.

* Let $i \geqslant 2$. A healthy vertex will be infected at step $i + 1$ if it has at least n^a infected neighbors at step i, i.e., $x \in A_{i+1}$ if $x \in A_i$ or $|\mathcal{N}(x) \cap A_i| \geqslant n^a$.

Note again that the process Boot2(t) dominates the process Boot, in the sense that given the same initial set A_0, then for each $k \in \mathbb{N}$, the set A_k given by Boot2(t) contains that given by Boot. Therefore, if percolation does not happen with high probability in the process Boot2(t) then percolation will not happen with high probability in the process Boot as long as the initial probability for each vertex to be infected is the same for both processes.

In order to prove the lower bound we need the following two lemmas. The first lemma is taken from [11], which generalizes Lemma 4.2.1 to a weighted binomial distribution. In the second lemma we will show that $\mathbb{P}(A_3 = A_2) = 1 - o(1)$ by drawing on ideas from [11].

Lemma 4.3.7. *Let $t, k, d_1, ..., d_k \in \mathbb{N}$ and $p \in (0,1)$. Let $X_i \sim Bin(d_i, p)$ for each $i \in [k]$, let $Y_k = \sum_{i=1}^{k} iX_i$, and let $D(k) = \sum_{i=1}^{k} i^2 d_i$. Then*

$$\mathbb{P}(Y_k \geqslant \mathbb{E}(Y_k) + t) \leqslant (2t)^{k-1} \exp\left(\frac{-2t^2}{D(k)}\right).$$

Lemma 4.3.8. *Consider the Boot2(t) process, where the value of the initial infection probability $p = n^{a-1} - n^{\frac{a}{2}-1+\delta}$ with $\frac{3}{4} < a < 1$, and $\delta > \frac{1}{2}(1-a)$. Then with high probability $A_3 = A_2$ in the process Boot2(t) with $t = \frac{1}{10} n^{\frac{a}{2}+\delta}$.*

Proof. Suppose $x \in A_3 \backslash A_2$. Then we have

$$|\mathcal{N}(x) \cap A_2| \geqslant n^a$$

and

$$|\mathcal{N}(x) \cap A_1| \leqslant n^a - t.$$

Thus there exists a set $T \subset \mathcal{N}(x)$ such that $|T| = t$ and $T \subset A_2 \backslash A_1$. Let $y \in T$, and then we have

$$|\mathcal{N}(y) \cap A_1| \geqslant n^a - t$$

and

$$|\mathcal{N}(y) \cap A_0| \leqslant n^a - 2t.$$

Thus every $y \in T$ must have at least t neighbors in $A_1 \backslash A_0$. Therefore there exists a set S such that

$S \subset \mathcal{N}(T) \cap A_1 \backslash A_0$ and

$$\frac{t^2}{2} \leqslant |S| \leqslant t^2.$$

Let $|S| = s$, $D = \mathcal{N}(S) \cap S(x, 3)$, and let d be the number of edges between S and D. Next partition D as $D = \cup_{i=1}^{3} D_i$ where each element in D_i has i neighbors in S and let $|D_i| = d_i$. Then we have

$$d = \sum_{i=1}^{3} i d_i = s(n-2).$$

Let $R_i = A_0 \cap D_i$ and $r_i = |R_i|$. Since every vertex in S has at least $n^a - 2t$ neighbors in A_0, we have

$$r := \sum_{i=1}^{3} i r_i \geqslant s(n^a - 2t - 2).$$

Now consider $\mathbb{E}(r)$. We have

$$\begin{aligned} \mathbb{E}(r) &= pd \\ &= ps(n-2) \\ &= (n^a - n^{\frac{a}{2}+\delta} - 2n^{a-1} + 2n^{\frac{a}{2}-1+\delta})s. \end{aligned}$$

Since $\{x \in A_3 \backslash A_2\}$ implies that $\{r \geqslant \mathbb{E}(r) + \frac{3}{4} n^{\frac{a}{2}+\delta} s\}$, we have

$$\mathbb{P}\left(\cup_{x \in V(Q_n)} x \in A_3 \backslash A_2\right) \leqslant 2^n \binom{n}{t} \binom{nt}{s} t^2 \sum_{d_1, d_2, d_3} \mathbb{P}\left(r \geqslant \mathbb{E}[r] + \frac{3}{4} n^{\frac{a}{2}+\delta} s\right).$$

Then by Lemma 4.3.7 we have

$$\mathbb{P}\left(r \geqslant \mathbb{E}[r] + \frac{3}{4} n^{\frac{a}{2}+\delta} s\right) \leqslant \frac{9}{4} n^{a+2\delta} s^2 \exp\left(-\frac{n^{a+2\delta} s^2}{D(3)}\right),$$

where $D(3) = \sum_{i=1}^{3} i^2 d_i$.

Moreover, there exists a $C > 0$ such that

$$\begin{aligned} D(3) &= \sum_{i=1}^{3} i d_i + \sum_{i=1}^{3} (i^2 - i) d_i \\ &= s(n-2) + \sum_{i=1}^{3} (i^2 - i) d_i \\ &\leqslant Csn. \end{aligned}$$

Therefore, there exists a $C' > 0$ such that

$$\mathbb{P}\left(r \geqslant \mathbb{E}[r] + \frac{3}{4} n^{\frac{a}{2}+\delta} s \right) \leqslant \frac{9}{4} n^{a+2\delta} s^2 \exp(-C' n^{a+2\delta-1} s).$$

Putting everything together we have

$$\begin{aligned} \mathbb{P}(\cup_{x \in V(Q_n)} x \in A_3 \backslash A_2) &\leqslant 2^n \binom{n}{t} \binom{nt}{s} t^2 \sum_{d_1,d_2,d_3} \frac{9}{4} n^{a+2\delta} s^2 \exp(-C' n^{a+2\delta-1} s) \\ &\leqslant 2^n \binom{n}{t} \binom{nt}{s} t^2 (ns)^3 \frac{9}{4} n^{a+2\delta} s^2 \exp(-C' n^{a+2\delta-1} s) \\ &= o(1), \end{aligned}$$

as long as $\delta > \frac{1}{2}(1-a)$. □

Now we are ready to prove the lower bound in Theorem 4.1.2

Proof. Consider the process Boot2(t) with $t = \frac{1}{10} n^{\frac{a}{2}+\delta}$, where each vertex is initially infected with probability p. Let $p = n^{a-1} - n^{\frac{a}{2}-1+\delta}$, where $\frac{3}{4} < a < 1$, and $\delta > \frac{1}{2}(1-a)$. If we prove that $\mathbb{P}(A_2 = V(Q_n)) = o(1)$ for Boot2(t), together with Lemma 4.3.8, then this would imply that the process Boot2(t) does not percolate with high probability, therefore also that the process Boot does not percolate with high

probability. Indeed, we have

$$\mathbb{P}(x \notin A_2) \geqslant \mathbb{P}(x \notin A_0)\mathbb{P}(|\mathcal{N}(x) \cap A_0| < n^a - 2t)\mathbb{P}(|\mathcal{N}(x) \cap (A_1 \backslash A_0)| = \emptyset).$$

It is obvious that $\mathbb{P}(x \notin A_0) > 1 - o(1)$. By Lemma 4.2.1, we have

$$\mathbb{P}(|\mathcal{N}(x) \cap A_0| < n^a - 2t) \geqslant 1 - \exp(-n^{\delta'}),$$

where $\delta' > 0$.

Let $y \in \mathcal{N}(x) \backslash A_0$. Then $\mathbb{P}(y \in A_1) = \mathbb{P}(|\mathcal{N}(y) \cap A_0| \geqslant n^a - 2t) = \epsilon(n)$, where $\epsilon(n) = \exp(-n^{\delta'})$. Since the events $\{y \notin A_1\}_{y \in \mathcal{N}(x) \backslash A_0}$ are decreasing events and thus by Lemma 4.3.2,

$$\begin{aligned}
\mathbb{P}(|\mathcal{N}(x) \cap (A_1 \backslash A_0)| = \emptyset) &\geqslant \prod_{y \in \mathcal{N}(x)} \mathbb{P}(y \notin A_1) \\
&\geqslant (1 - \epsilon(n))^n \\
&\geqslant e^{-2\epsilon(n)n}.
\end{aligned}$$

From Lemmas 4.2.4 and 4.2.5, we can observe that there exists a partition of the set $V(Q_n)$ as follows:

$$V = \cup_{i=1}^{m} B_i,$$

where, for each $i \in [m]$, and for any x and $y \in B_i$, we have $d_H(x, y) \geqslant 6$. Furthermore, the number of subsets m in this partition satisfies $m = \sum_{i=1}^{5} \binom{n}{i}$.

Let $B \subset V$. Assume for any $x \in B$ and $y \in B$, we have $d_H(x, y) \geqslant 6$. Then from Fact 4.2.2, the events $\{x \in A_2\}_{x \in B}$ are independent.

Given the partition of V, there exists a set B_i such that $|B_i| \geqslant \frac{2^n}{n^6}$. Furthermore, for any x and y

belonging to B_i, it holds that $d_H(x, y) \geqslant 6$. With the independence condition in place, we have

$$\begin{aligned}\mathbb{P}(x \in A_2 \text{ for all } x \in B_i) &= (1 - e^{-2\epsilon(n)n})^{|B_i|} \\ &\leqslant (2\epsilon(n)n)^{|B_i|} \\ &= o(1),\end{aligned}$$

since $\epsilon(n)$ is an exponentially decaying function of n. □

4.4 PROOF OF THEOREM 3

Let us start by proving the upper bound. To do so, we require a slightly more general concept. Let $A_{i,r}$ represent the set of infected vertices up to and including step i when the infection threshold is r. Hence, $A_i = A_{i,\frac{N}{2}}$. In order to establish the upper bound, we will rely on a few lemmas.

Lemma 4.4.1. *Let $x \in V(Q_{k,n})$, let $p = \frac{1}{2} - (\frac{1}{n})^{\frac{k}{2}}$, and let the infection threshold be $r = \frac{N}{2} + \binom{3k}{2k}$, where $k \geqslant 2$ is an integer. Then there exists an absolute constant $\delta > 0$ such that*

$$\mathbb{P}(x \in A_{1,r}) \geqslant \frac{1}{2} + \delta.$$

Proof. We have

$$\begin{aligned}\mathbb{P}(x \in A_{1,r}) &= \mathbb{P}(x \in A_0) + \mathbb{P}(|\mathcal{N}(x) \cap A_0| \geqslant r | x \notin A_0)\mathbb{P}(x \notin A_0) \\ &= \mathbb{P}(x \in A_0) + \mathbb{P}(|\mathcal{N}(x) \cap A_0| \geqslant r)\mathbb{P}(x \notin A_0).\end{aligned}$$

Let $S(N) = |\mathcal{N}(x) \cap A_0|$ and so $S(N) \sim \text{Bin}(N, p)$. Then, we have

$$\mathbb{P}(S(N) \geqslant r) = \mathbb{P}\left(\frac{S(N) - Np}{\sqrt{Np(1-p)}} \geqslant \frac{r - Np}{\sqrt{Np(1-p)}}\right).$$

Let $z(N) = \frac{r-Np}{\sqrt{Np(1-p)}}$. Then it is easy to see that

$$\lim_{N\to\infty} z(N) = \frac{2\sqrt{k!}}{k!}.$$

By Lemma 4.3.1, we have

$$\lim_{N\to\infty} \mathbb{P}(S(N) \geqslant r) = 1 - \Phi\Big(\frac{2\sqrt{k!}}{k!}\Big).$$

Therefore there exists an absolute constant $\delta > 0$

$$\mathbb{P}(A_{1,r}) \geqslant \frac{1}{2} + \delta.$$

□

Before stating the next lemma, we will introduce two definitions and one theorem.

Definition 4.4.1. *A 1-factor of a hypergraph is a spanning 1-regular sub-hypergraph.*

Definition 4.4.2. *A decomposition of a hypergraph into edge-disjoint 1-factors is a 1-factorization.*

The following Theorem was originally proved in [14] by Baranyai.

Theorem 4.4.2. *Let $k \geqslant 2$. A 1-factorization of the complete k-uniform hypergraph on n vertices exists if and only if $k|n$.*

Now we are ready to state the next lemma.

Lemma 4.4.3. *Let $\delta > 0$ be a real number and $k \geqslant 2$ be an integer. Additionally, we assume that k divides n. Additionally assume that with the initial infection probability $p = \frac{1}{2} - (\frac{1}{n})^{k/2}$ we can establish $\mathbb{P}\left(x \in A_{1,\frac{N}{2}+\binom{3k}{2k}}\right) \geqslant \delta$. Then there exists a constant $c > 0$ such that*

$$\mathbb{P}(x \in A_{4,\frac{N}{2}}) \geqslant 1 - e^{-cn^{k-1}}.$$

Proof. Let $x \in V(Q_{k,n})$.There exists a partition $S(x,k) = \cup_{i=1}^{m} B_i$ such that supp$(x) \cap$ supp$(y) = \emptyset$ for all $x, y \in B_i$ and all $i \in [m]$. Moreover, $|B_i| = \frac{n}{k}$ for all $i \in [m]$ and $m = \binom{n-1}{k-1}$.

Indeed, we can consider $S(x,k)$ as the edges of a k-uniform hypergraph on n vertices. According to Theorem 4.4.2, such a partition of $S(x,k)$ exists given that $k|n$.

There exists an $i \in [m]$ such that there is a set $S \subset B_i \backslash A_0$ of size $|S| = \frac{n}{2k} + 1$. If such a set S cannot be found, this implies that for every $i \in [m]$, $|B_i \cap A_0| \geqslant \frac{n}{2k}$. If for every $i \in [m]$, $|B_i \cap A_0| \geqslant \frac{n}{2k} + 1$, then $x \in A_1$, and we are done. Since the "+1" term will not affect the calculations, we can safely assume that there exists a set $S \subset B_i \backslash A_0$ with $|S| = \frac{n}{2k}$.

Let $T = \mathcal{N}(S) \cap S(x, 2k)$. Denote $T_i = \{y \in T : |\mathcal{N}(y) \cap S| = i\}$ for $i \in \{1, 2\}$. It is easy to see that

$$|T_1| = \binom{n-k}{k}\frac{n}{2k} - 2\binom{\frac{n}{2k}}{2} = \frac{n^{k+1}}{2kk!} - \frac{(3k-1)(k^2-k)n^k}{4kk!} + O(n^{k-1})$$

and

$$|T_2| = \binom{\frac{n}{2k}}{2} = \frac{n^2}{8k^2} - \frac{n}{4k}.$$

Let $|T_1 \cap A_2| = a$ and $|T_2 \cap A_2| = b$. Note that if $S \cap A_3 = \emptyset$, then $a + 2b < \frac{n}{2k}(\frac{1}{2}N) = \frac{n^{k+1}}{4kk!} + \frac{k(1-0.5(k-1)(k-2))n^k}{4kk!} + O(n^{k-1})$.

Note that

$$\mathbb{E}|T_1 \cap A_0| = \frac{n^{k+1}}{4kk!} - \frac{(3k-1)(k^2-k)n^k}{8kk!} + O(n^{k-1}).$$

By Lemma 4.2.1, that there exits $c' > 0$ such that

$$\mathbb{P}(||T_1 \cap A_0| - \frac{n^{k+1}}{4kk!}| \geqslant kn^k) \leqslant \exp(-c'n^{k-1}).$$

Let $G(S)$ denote this event doesn't happen, i.e., $||T_1 \cap A_0| - \frac{n^{k+1}}{4kk!}| \leqslant kn^k$ and we have $\mathbb{P}(G(S)) \leqslant 1 - \exp(-c'n^{k-1})$.

From now on, we assume that the event $G(S)$ holds. If $|T_1 \cap A_2 \backslash A_0| \geqslant 2kn^k$, then $a \geqslant \frac{n^{k+1}}{4kk!} + kn^k$ and thus $S \cap A_3 \neq \emptyset$.

Claim 4.4.4. *If $G(S)$ holds, then*

$$\mathbb{P}(|T_1 \cap A_2 \backslash A_0| = O(n^k)) \leqslant C\binom{3k}{2k} n^k \exp(-c''\delta^2 \epsilon n^{1+k}).$$

Proof. Consider a bipartite graph H with $V(H) = H_1 \cup H_2$ where $H_1 = T_1 \backslash A_0$, $H_2 = S(x, 3k)$ and $E(H) = \{uv : uv \in E(Q_{k,n})\}$.We will color edges of H red and blue. If the end point in H_2 is in A_1, then color edges xy red for every $x \in H_1$ and if the end point in H_2 is not in A_1, then color edges xy blue for every $x \in H_1$.

It is easy to see that $|E(H)| = \binom{n-2k}{k}|H_1|$. Since $G(s)$ holds, we have $|H_1| = \frac{n^{k+1}}{4kk!} + O(n^k)$.

Now suppose that $|T_1 \cap A_2 \backslash A_0| = O(n^k)$. Since at most $O(n^k)$ vertices in H_1 can have at least $\frac{N}{2}$ neighbors in A_1, the number of red edges in H, denoted $E_R(H)$, satisfies

$$|E_R(H)| \leqslant \frac{1}{2}\binom{n}{k}|H_1| + O(n^{2k}),$$

and thus the number of blue edges in H, denoted by $E_B(H)$, has to satisfy

$$|E_B(H)| \geqslant |H_1|\binom{n-2k}{k} - \frac{1}{2}\binom{n}{k}|H_1| - O(n^2k). \tag{4.9}$$

From Lemmas 4.2.4 and 4.2.5, we can observe that there exists a partition of the set $S(x,3k)$ as follows:

$$S(x,3k) = \cup_{i=1}^{m} D_i,$$

where, for each $i \in [m]$, and for any x and $y \in B_i$, we have $d_H(x,y) \geqslant 2k+1$. Furthermore, the number of subsets m in this partition satisfies $m \leqslant 1 + \sum_{i=1}^{k} \binom{3k}{3k-i}\binom{n-3k}{i} \leqslant Cn^k$.

Let us define, for $i \in [\binom{3k}{2k}]$ and $j \in [m]$,

$$D_j(i) = \{x \in D_j : |N(x) \cap T_1| = i\}$$

Claim 4.4.5. *If $\epsilon > 0$ is arbitrarily small, then there exists an $i \in [\binom{3k}{2k}]$ and a $j \in [m]$ such that*

(a) l edges of H are incident with $D_j(i)$ where $l \geqslant \epsilon n^{1+k}$ and

(b) at most $(\frac{1}{2} + \frac{\delta}{2})l$ of those edges are red.

Proof: Suppose the claim is false. Let us count the number of blue edges. We have

$$|E_B(H)| \leqslant \binom{n-2k}{k}|H_1|(\frac{1}{2} - \frac{\delta}{2}) + C\binom{3k}{2k}n^k \epsilon n^{1+k},$$

which is contradiction to (4.9) since ϵ is chosen arbitrarily small.

Note that events $\{y \in A_1\}_{y \in D_j(i)}$ are independent, but they are not independent of which members of $B(x,2k)$ are in A_0. Since $y \in B_j(i)$ has at most $\binom{3k}{2k}$ neighbors in $B(x,2)$ and $\mathbb{P}(y \in A_{1,\frac{N}{2}+\binom{3k}{2k}}) \geqslant \frac{1}{2} + \delta$, then $\mathbb{P}(y \in A_{1,\frac{N}{2}}) \geqslant \frac{1}{2} + \delta$ for any set $B(x,2) \cap A_0$.

Let $E_j(i)$ denote the event that $D_j(i)$ satisfies (a) and (b). Thus

$$\begin{aligned}\mathbb{P}(\cup_{i,j}E_j(i)) &\leqslant \sum_{i,j}\mathbb{P}\left(\text{Bin}\left(\frac{\epsilon n^{1+k}}{i},\frac{1}{2}+\delta\right)\leqslant\left(\frac{1}{2}+\frac{\delta}{2}\right)\frac{\epsilon n^{1+k}}{i}\right)\\ &\leqslant C\binom{3k}{2k}n^k\exp(-c''\delta^2\epsilon n^{1+k}),\end{aligned}$$

for $c''>0$. □

The number of choices for S is at most

$$\binom{n/k}{n/2k}\leqslant 2^{\frac{n}{k}}.$$

Let $M(S)$ denote the event $|T_1\cap(A_2\backslash A_0)|\geqslant 2kn^k$. Note that if for every $i\in[\frac{n-1}{t-1}]$ and every subset $S\subset B_i\backslash A_0$ of size $\frac{n}{2t}$ both $G(S)$ and $M(S)$ hold, then $x\in A_4$.

Therefore,

$$\begin{aligned}\mathbb{P}(x\notin A_4)&\leqslant\sum_i\sum_S\mathbb{P}[G(S)^c]+\sum_i\sum_S\mathbb{P}[M(S)^c]\\ &\leqslant\exp(-c'''n^{k-1}),\end{aligned}$$

where $c'''>0$. □

Now we are ready to prove the upper bound for $k>2$ in Theorem 4.1.3.

Proof. By Lemmas 4.4.1 and 4.4.3 we have, for $k > 2$,

$$\begin{aligned}\mathbb{P}(A_0 \text{ does not percolate}) &\leqslant \mathbb{P}(\cup_{x\in V(G)}(x \notin A_4)) \\ &\leqslant \sum_{x\in V(G)} \mathbb{P}(x \notin A_4) \\ &\leqslant 2^n e^{-c'''n^{k-1}} \\ &= o(1).\end{aligned}$$

□

For $k = 2$, we can observe that Lemma 4.4.3 is insufficient to prove the statement in Theorem 4.1.3. Therefore, it is necessary to prove an additional lemma in order to prove the statement for $k = 2$.

Lemma 4.4.6. *Let $\delta > 0$ be a real number. Additionally, we assume n to be even. Let the initial infection probability be $p = \frac{1}{2} - \frac{1}{n}$ for each vertex. With this value of p we can establish $\mathbb{P}(x \in A_{1,r=\frac{N}{2}+21}) \geqslant \frac{1}{2} + \delta$. Then*

$$\mathbb{P}(x \in A_{4,r=\frac{N}{2}}) \geqslant 1 - e^{-cn^{0.1}}.$$

Proof. Assume n is even. Recall $S(x, l) = \{y \in V(Q_{2,n}) : d_H(x, y) = l\}$.

From Theorem 4.4.2, there exists a partition of $S(x, 2)$ of the form $S(x, 2) = \cup_{i=1}^{m} B_i$ such that Support$(x) \cap$ Support$(y) = \emptyset$ for all $x, y \in B_i$ and all $i \in [m]$. Moreover, $|B_i| = \frac{n}{2}$ for all $i \in [m]$ and $m = n - 1$.

From Lemma 4.2.1 we have

$$\mathbb{P}(|B_i \cap A_0| \leqslant \frac{n}{4} - n^{\frac{1.1}{2}}) \leqslant \exp\Big(-n^{0.1}\frac{1}{50}\Big).$$

Let $J(i, x)$ denote this event does not happen, i.e $|B_i \cap A_0| \geqslant \frac{n}{4} - n^{\frac{1.1}{2}}$. From now on, we assume that $J(i, x)$ holds and then choose $S \subset B_i \backslash A_0$ with $|S| = \frac{n}{4} + 1$. We will show that $S \cap A_3 = \emptyset$ is highly

unlikely. Since the $+1$ term will not affect the calculation, we will drop it.

Let $T = \mathcal{N}(S) \cap S(x,4)$ and $T_i = \{y \in T : |\mathcal{N}(y) \cap S| = i\}$ for $i \in \{1,2\}$. It is easy to see that

$$|T_2| = \binom{\frac{n}{4}}{2} = \frac{1}{32}n^2 - \frac{1}{8}n$$

and

$$|T_1| = \binom{n-2}{2}\frac{n}{4} - 2\binom{\frac{n}{4}}{2} = \frac{n^3}{8} - \frac{11}{16}n^2 + n.$$

Let $a = |T_1 \cap A_2|$ and $b = |T_2 \cap A_2|$. Observe that if $S \cap A_3 = \emptyset$, then $a + 2b \leqslant \frac{n}{4}\frac{1}{2}(\binom{n}{2} + n) = \frac{n^3}{16} + \frac{n^2}{16}$.

From Lemma 4.2.1, we have

$$\mathbb{P}\Big(||T_1 \cap A_0| - \frac{n^3}{16}| \geqslant n^2\Big) \leqslant \exp\Big(-\frac{1}{3000}n\Big).$$

Let $G(S)$ denote the event that $||T_1 \cap A_0| - \frac{n^3}{16}| \leqslant n^2$ and so $\mathbb{P}(G(S)) \geqslant 1 - \exp(-\frac{1}{3000}n)$. Assume that $G(S)$ holds. Therefore, if $|T_1 \cap A_2 \backslash A_0| \geqslant 2n^2$, then with high probability, $a \geqslant \frac{n^3}{16} + n^2$ and thus $S \cap A_3 \neq \emptyset$.

Claim 4.4.7. *If $G(S)$ holds, then*

$$\mathbb{P}(|T_1 \cap A_2 \backslash A_0| = O(n^2)) \leqslant n^3 \exp(-c'\delta^2\epsilon n).$$

Proof. Consider a bipartite graph H with the vertex set $H_1 \cup H_2$, where $H_1 = T_1 \backslash A_0 \subset S(x,4)$ and $H_2 = S(x,6)$, and the edge set $\{uv : uv \in E(G)\}$. We have

$$|E(H)| = |H_1|\binom{n-4}{2}$$

and

$$|H_1| = \frac{n^3}{16} + O(n^2),$$

since $G(S)$ holds.

We will color the edges of H red and blue. If $y \in H_2$ is in A_1, then color xy red for every $x \in H_1$ and if $y \in H_2$ is not in A_1, then color xy blue for every $x \in H_1$. Now suppose $|T_1 \cap A_2 \backslash A_0| = O(n^2)$. Thus, only $O(n^2)$ vertices of H_1 can have at least $\frac{n^2}{4} + n$ neighbors in A_1. Thus

$$|E_R(H)| \leqslant \frac{n^2}{4}|H_1| + O(n^4) = \frac{n^5}{64} + O(n^4)$$

and

$$|E_B(H)| \geqslant \binom{n-4}{2}|H_1| - \frac{n^2}{4}|H_1| - O(n^4) = \frac{n^5}{64} - O(n^4), \tag{4.10}$$

where $E_R(H)$ denotes the red edges of H and $E_B(H)$ denotes the blue edges of H.

From Lemmas 4.2.4 and 4.2.5, we can observe that there exists a partition of the set $S(x, 6)$ as follows:

$$S(x, 6) = \cup_{i=1}^{m} D_i,$$

where, for each $i \in [m]$, and for any x and $y \in B_i$, we have $d_H(x, y) \geqslant 9$. Furthermore, the number of subsets m in this partition satisfies $m = \binom{6}{2}\binom{n-6}{4} + \binom{6}{4}\binom{n-6}{2} + \binom{6}{3}\binom{n-6}{3} + \binom{6}{5}\binom{n-6}{1} + 1$.

Define a refinement of D_i as $D_j(i) = \{z \in D_j : |N(z) \cap T_1| = i\}$ for $i \in [15]$.

Claim 4.4.8. *For a sufficiently small $\epsilon > 0$, there exist an $i \in [15]$ and $j \in [m]$ such that:*

(a) l edges of H are incident to $D_j(i)$ and $l \geqslant \epsilon n$. And

(b) at most $(\frac{1}{2} + \frac{\delta}{2})l$ of those edges are red.

Proof. Suppose the claim is false. We can have at most

$$|E_B(H)| \leqslant |E(H)|(\frac{1}{2} - \frac{\delta}{2}) + 15\epsilon nm$$

blue edges, which is a contradiction to the bound in (4.10). □

Note that $\{y \in A_1\}_{y \in D_j(i)}$ are independent, but this event is not independent of which members of $B(x,5)$ are in A_0. Since $y \in D_j(i)$ has 21 neighbors in $B(x,5)$ and $\mathbb{P}(x \in A_{1,r=\frac{N}{2}+21}) \geqslant \frac{1}{2} + \delta$, then $\mathbb{P}(y \in A_{1,\frac{N}{2}}) \geqslant \frac{1}{2} + \delta$ for all set $B(x,5) \cap A_0$.

Let $E_j(i)$ denote the event that $D_j(i)$ satisfies (a) and (b). Thus, there exists a constant $c' > 0$ such that

$$\begin{aligned}\mathbb{P}(\cup_{i,j} E_j(i)) &\leqslant \sum_{i,j} \mathbb{P}\left(\text{Bin}\left(\frac{\epsilon n}{i}, \frac{1}{2} + \delta\right) \leqslant \left(\frac{1}{2} + \frac{\delta}{2}\right)\frac{\epsilon n}{i}\right) \\ &\leqslant 15n^4 \exp(-c'\delta^2 \epsilon n).\end{aligned}$$ □

The number of choices for S is at most

$$\begin{aligned}\binom{\frac{n}{4} + n^{0.55}}{\frac{n}{4}} &\leqslant \left(e\left(1 + \frac{n^{1.1/2}}{n/4}\right)\right)^{n^{0.55}} \\ &\leqslant \exp(n^{0.55}) \exp(\frac{4}{n^{0.9/2}} n^{0.55}) \\ &\leqslant \exp(n^{0.7}).\end{aligned}$$

Let $M(S)$ denote the event $|T_1 \cap (A_2 \backslash A_0)| \geqslant 2n^2$. Note that if for every $i \in [n-1]$ and for every subset $S \subset B_i \backslash A_i$ of size $\frac{n}{4}$, $J(i,x)$,$G(S)$, and $M(S)$ hold, then $x \in A_4$.

Therefore,

$$\begin{aligned}\mathbb{P}(x \notin A_4) &\leqslant \sum_{i\in[n-1]} \mathbb{P}(J(x)^c) + \sum_{i\in[n-1]} \sum_{S} \mathbb{P}[G(S)^c] + \sum_{i\in[n-1]} \sum_{S} \mathbb{P}[M(S)^c] \\ &\leqslant \exp(-cn^{0.1}).\end{aligned}$$

□

Lemma 4.4.9. *Suppose that for every $x \in V(Q_{2,n})$ there exists an absolute constant $c > 0$ such that*

$$\mathbb{P}(x \notin A_k) \leqslant e^{-cn^{0.1}}.$$

Then for some $d > 0$, we have

$$\mathbb{P}(x \notin A_{3k+2}) \leqslant e^{-dn^{1.1}}.$$

Proof. Suppose that $x \notin A_{3k+2}$. Assume that k is an absolute constant. We claim that for each even $t \in [0, 2k+2]$, there exists a set $T(t) \subset S(x,t)$ and a function $0 < \alpha(t) < 1$ such that

$$T(t) \cap A_{3k+2-t} = \emptyset \tag{4.11}$$

and

$$|T(t)| \geqslant \alpha(t) n^t - O(n^{t-1}). \tag{4.12}$$

We will prove the claim by induction on t. When $t = 0$, $|T(t)| = 1$ so the statement is trivially true. Suppose that $T(t)$ satisfies 4.11 and 4.12. Then for every $y \in T(t)$ it has at most $\frac{1}{2}(n + \binom{n}{2})$ neighbors in $S(x, t+2) \cap A_{3k-t}$ and so at least $\binom{n-t}{2} - \frac{1}{2}(n + \binom{n}{2})$ neighbors in $S(x,t+2) \backslash A_{3k-t}$. Note that each element of $S(x, t+2)$ has $\binom{t+2}{2}$ neighbors in $S(x,t)$.

Therefore, there exists a set $T(t+2) \subset S(x, t+2)$ such that $T(t+2) \cap A_{3k-t} = \emptyset$ and

$$|T(t+2)| \geqslant \frac{[\binom{n-t}{2} - \frac{1}{2}(n + \binom{n}{2})]|T(t)|}{\binom{t+2}{2}}.$$

Moving forward, We need the following fact.

From Fact 4.2.2, it is easy to see that if $S \subset V(Q_{2,n})$ satisfies $d_H(y, z) \geqslant 4k+1$ for every $y, z \in S$, then the events $\{y \in A_k\}_{y \in S}$ are independent.

From Lemmas 4.2.4 and 4.2.5, we can observe that there exists a partition of the set $S(x, 2k+2)$ as follows:

$$S(x, 2k+2) = \cup_{i=1}^{m} B_i,$$

where, for each $i \in [m]$, and for any x and $y \in B_i$, we have $d_H(x, y) \geqslant 4k+1$. Furthermore, the number of subsets m in this partition satisfies $m = 2(k+1)\binom{n}{2k+1}$.

Now we claim that there exist a $j \in [m]$ and a sufficiently small $\epsilon > 0$ such that

$$|B_j| \geqslant \epsilon n$$

and

$$|T(2k+2)| \geqslant \epsilon |B_j|.$$

Indeed, if the claim is not true then

$$|T(2k+2)| \leqslant \epsilon mn + \epsilon \binom{n}{2k+2},$$

which is a contraction since ϵ is sufficiently small.

Recall that $T(2k+2) \cap A_k = \emptyset$ and $\mathbb{P}(x \notin A_k) \leqslant e^{-cn^{0.1}}$. From Lemma 7 in [10], we have

$$\begin{aligned}\mathbb{P}(|T(2k+2) \cap B_j| \geqslant \epsilon|B_j|) &\leqslant 2(e^{-cn^{0.1}})^{\epsilon|B_j|/2} \\ &\leqslant 2(e^{-cn^{1.1}\epsilon^2/2}).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P}(x \notin A_{3k+2}) &\leqslant \mathbb{P}\left(\exists j \text{ with } |B_j| \geqslant \epsilon n \text{ and } |T(2k+2) \cap B_j| \geqslant \epsilon|B_j|\right) \\ &\leqslant m2(e^{-cn^{1.1}\epsilon^2/2}) \\ &\leqslant e^{-dn^{1.1}}.\end{aligned}$$

□

Now we are ready to prove the upper bound for $k = 2$ in Theorem 4.1.3.

Proof. By Lemmas 4.4.1, 4.4.6, and 4.4.9 we have

$$\begin{aligned}\mathbb{P}(A_0 \text{ does not percolate}) &\leqslant \mathbb{P}(\cup_{x \in V(G)}(x \notin A_{14})) \\ &\leqslant \sum_{x \in V(G)} \mathbb{P}(x \notin A_{14}) \\ &\leqslant 2^n e^{-dn^{1.1}} \\ &= o(1).\end{aligned}$$

□

Next we will prove the lower bound of Theorem 4.1.3. Let us introduce a new process closely related to the bootstrap percolation process. The only difference is a less strict requirement for a vertex to be infected at step 1 and step 2. Let A_i denote the set of infected vertices up to step i. We define the process Boot3(t) as follows:

* Initially, each vertex is infected with probability p, independent of other vertices.

* If a vertex becomes infected, then it remains infected forever.

*A healthy vertex will be infected at step 1 if it has at least $\frac{N}{2} - 2t$ infected neighbors at the initial step, i.e., $x \in A_1$ if $x \in A_0$ or $|\mathcal{N}(x) \cap A_0| \geqslant \frac{N}{2} - 2t$.

*A healthy vertex will be infected at step 2 if it has at least $\frac{N}{2} - t$ infected neighbors at the first step, i.e., $x \in A_2$ if $x \in A_1$ or $|\mathcal{N}(x) \cap A_1| \geqslant \frac{N}{2} - t$.

* Let $i \geqslant 2$. A healthy vertex will be infected at step $i+1$ if it has at least $\frac{N}{2}$ infected neighbors at step i, i.e., $x \in A_{i+1}$ if $x \in A_i$ or $|\mathcal{N}(x) \cap A_i| \geqslant \frac{N}{2}$.

Let us refer to the original $\frac{N}{2}$-neighbor process as Boot($\frac{N}{2}$).

To establish the lower bound, we will use Lemma 4.3.7 to show that $\mathbb{P}(A_3 = A_2) = 1 - o(1)$. Subsequently, we will establish that $\mathbb{P}(A_2 = V) = o(1)$.

Lemma 4.4.10. *Consider the Boot3(t) process, where the initial infection probability $p = \frac{1}{2} - \frac{1}{n^a}$ with $a < 1$ and $t = n^b$ with $k - 1 < b < k$. Then $A_3 = A_2$ with high probability.*

Proof. Suppose $x \in A_3 \backslash A_2$. Then we have

$$|\mathcal{N}(x) \cap A_2| \geqslant \frac{N}{2}$$

and

$$|\mathcal{N}(x) \cap A_1| \leqslant \frac{N}{2} - t.$$

Thus there exists a set $T \subset \mathcal{N}(x)$ with $|T| = t$ such that $T \subset A_2 \backslash A_1$ and

$$|T \cap S(x,k)| \geqslant t - \sum_{i=1}^{k-1} \binom{n}{i}.$$

Let $y \in T$, and then we have

$$|\mathcal{N}(y) \cap A_1| \geqslant \frac{N}{2} - t$$

and

$$|\mathcal{N}(y) \cap A_0| \leqslant \frac{N}{2} - 2t.$$

Thus every $y \in T$ must have at least t neighbors in $A_1 \backslash A_0$.

For every $y \in (T \cap S(x,k))$, it must have at least t neighbors in $A_1 \backslash A_0$ among which at least $t - O(n^{k-1})$ are in $S(x,2k)$. This is because each vertex in $S(x,k)$ has at most $O(n^{k-1})$ neighbors outside $S(x,2k)$.Therefore, there exists a set S such that

$$S \subset (A_1 \backslash A_0) \cap S(x,2k) \cap \mathcal{N}(T \cap S(x,k))$$

and

$$\frac{(t - \sum_{i=1}^{k-1} \binom{n}{i})(t - O(n^{k-1}))}{\binom{2k}{k}} \leqslant |S| \leqslant t^2,$$

since each vertex in $S(x,2k)$ has at most $\binom{2k}{k}$ neighbors in $S(x,k)$.

Now, let us consider the set $D := \mathcal{N}(S) \cap S(x,3k)$. Let d be the number of edges between S and D. We can express D as a union of disjoint sets D_i, where $D = \cup_{i=1}^{\binom{3k}{k}} D_i$. Each element in D_i has i neighbors in S. Let $|D_i| = d_i$ and $|S| = s$. Then we have

$$d = \sum_{i=1}^{\binom{3k}{k}} i d_i = s\binom{n-2k}{k}.$$

Let $R_i = A_0 \cap D_i$ and $|R_i| = r_i$. We have

$$r := \sum_{i=1}^{\binom{3k}{k}} i r_i \geqslant s\left(\frac{N}{2} - 2t - O\left(n^{k-1}\right)\right)$$

and also

$$\begin{aligned}\mathbb{E}(r) &= pd \\ &= p\binom{n-2k}{k}s \\ &= (\frac{1}{2}-\delta)\binom{n-2k}{k}s \\ &= \left(\frac{n^k}{2k!}-\frac{\delta n^k}{k!}+O\left(n^{k-1}\right)\right)s,\end{aligned}$$

where $\delta = \frac{1}{n^a}$.

Since the event $\{x \in A_3 \backslash A_2\}$ implies the event $\{r \geqslant \mathbb{E}(r) + (\frac{\delta n^k}{k!} - 2t)s\}$, we have

$$\mathbb{P}(\cup_{x\in V(Q_{k,n})} x \in A_3\backslash A_2) \leqslant 2^n\binom{n^k}{t}\binom{n^k t}{s}t^2 \sum_{d_1,d_2,\ldots d_{\binom{3k}{k}}} \mathbb{P}\left(r \geqslant \mathbb{E}[r] + \left(\frac{\delta n^k}{k!}-2t\right)s\right).$$

Indeed, we have at most 2^n vertices in $Q_{k,n}$, $\binom{n^k}{t}$ choices for the set T, $\binom{n^k t}{s}$ choices for the set S and at most t^2 values that s can take.

Assume $n^k\delta \gg t$, i.e., $k-a > b$. Thus, from Lemma 4.3.7, there exists a constant $c > 0$ such that

$$\mathbb{P}\left(r \geqslant \mathbb{E}[r] + (\frac{\delta n^k}{k!}-2t)s\right) \leqslant \left(2\frac{\delta n^k}{k!}\right)^{\binom{3k}{k}} \exp\left(-c\frac{(\delta n^k s)^2}{D\left(\binom{3k}{k}\right)}\right).$$

Since $D\left(\binom{3k}{k}\right) = \sum_{i=1}^{\binom{3k}{k}} i^2 d_i$ and $\sum_{i=1}^{\binom{3k}{k}} i d_i = s\binom{n-2k}{k}$, there exists a big absolute constant C such that

$$D\left(\binom{3k}{k}\right) = s\binom{n-2k}{k} + \sum_{i=1}^{\binom{3k}{k}}(i^2-i)d_i \leqslant Cs\binom{n-2k}{k}.$$

Thus, we have

$$\mathbb{P}[\exists x \in A_3 \backslash A_2] = o(1). \quad \square$$

Finally, we are ready to prove the lower bound of Theorem 4.1.3

Proof. Consider the process Boot3(t), where the initial infection probability $p = \frac{1}{2} - \frac{1}{n^a}$, where $0 < a < 1$. Let $k \geqslant 2$ be an integer independent of n and $t = n^b$ where $k - 1 < b < k$. Furthermore, we assume $k - a > b$. If we can prove that $\mathbb{P}(A_2 = V(Q_{k,n})) = o(1)$ for Boot3(t), in conjunction with Lemma 4.4.10, it implies that the process Boot3(t) does not percolate with high probability. Consequently, this suggests that the process Boot($\frac{N}{2}$) does not percolate with high probability.

Indeed, we have

$$\mathbb{P}\left(x \notin A_2\right) \geqslant \mathbb{P}(x \notin A_0)\mathbb{P}\left(|\mathcal{N}(x) \cap A_0| < \frac{N}{2} - 2t\right) \mathbb{P}\left(|\mathcal{N}(x) \cap (A_1 \backslash A_0)| = \emptyset\right).$$

It is obvious to see that $\mathbb{P}(x \notin A_0) > \frac{1}{2}$ and by Lemma 4.2.1, that

$$\mathbb{P}\left(|\mathcal{N}(x) \cap A_0| < \frac{N}{2} - 2t\right) \geqslant 1 - \exp(-n^{\delta}),$$

where $\delta > 0$.

Consider $y \in \mathcal{N}(x) \backslash A_0$. Then we have $\mathbb{P}(y \in A_1) = \mathbb{P}\left(|\mathcal{N}(y) \cap A_0| \geqslant \frac{N}{2} - 2t\right) = \epsilon(n)$, where $\epsilon(n) = \exp(-n^{\delta})$. Since the events $\{y \notin A_1\}_{y \in \mathcal{N}(x) \backslash A_0}$ are decreasing and thus by Lemma 4.3.2,

$$\begin{aligned}
\mathbb{P}(|\mathcal{N}(x) \cap (A_1 \backslash A_0)| = \emptyset) &\geqslant \prod_{y \in \mathcal{N}(x)} \mathbb{P}(y \notin A_1) \\
&\geqslant (1 - \epsilon(n))^{n^k} \\
&\geqslant e^{-2\epsilon(n) n^k}.
\end{aligned}$$

From Lemmas 4.2.4 and 4.2.5, we can observe that there exists a partition of the set $V(Q_{k,n})$ as follows:

$$V(Q_{k,n}) = \cup_{i=1}^{m} B_i,$$

where, for each $i \in [m]$, and for any x and $y \in B_i$, we have $d_H(x, y) \geqslant 4k + 1$. Furthermore, the number of subsets m in this partition satisfies $m = \sum_{i=1}^{4k+1} \binom{n}{i}$.

Let $B \subset V(Q_{k,n})$. Assume for every $x \in B$ and $y \in B$, $d_H(x, y) \geqslant 4k + 1$. Then the events $\{x \in A_2\}_{x \in B}$ are independent.

Given the partition of $V(Q_{k,n})$, there exists a set B_i such that $|B_i| \geqslant \frac{2^n}{n^{4k+2}}$. Furthermore, for any x and $y \in B_i$, it holds that $d_H(x, y) \geqslant 4k + 1$. With the independence condition in place, we have

$$\begin{aligned}
\mathbb{P}(x \in A_2 \text{ for all } x \in B_i) &= (1 - e^{2\epsilon(n)n^k})^{|B_i|} \\
&\leqslant (2\epsilon(n)n^k)^{|B_i|} \\
&= o(1).
\end{aligned}$$

since $\epsilon(n)$ is an exponentially decaying function of n. □

Now let us prove the lower bound for Theorem 4.1.3 when $k = 2$.

Lemma 4.4.11. *Consider the Boot3(t) process, where the initial infection probability $p = \frac{1}{2} - \frac{C\sqrt{\log n}}{n}$ and $t = 10n$. Then $A_3 = A_2$ with high probability.*

Proof. Suppose $x \in A_3 \backslash A_2$. Then we have

$$|\mathcal{N}(x) \cap A_2| \geqslant \frac{N}{2}$$

and

$$|\mathcal{N}(x) \cap A_1| \leqslant \frac{N}{2} - t.$$

Thus there exists a set $T \subset \mathcal{N}(x)$ with $|T| = t$ such that $T \subset A_2 \backslash A_1$ and

$$|T \cap S(x, 2)| \geqslant t - n,$$

since $|S(x, 1)| = n$.

Let $y \in T$, and then we have

$$|\mathcal{N}(y) \cap A_1| \geqslant \frac{N}{2} - t$$

and

$$|\mathcal{N}(y) \cap A_0| \leqslant \frac{N}{2} - 2t.$$

Thus every $y \in T$ must have at least t neighbors in $A_1 \backslash A_0$.

For every $y \in (T \cap S(x, 2))$ must have at least t neighbors in $A_1 \backslash A_0$ among which at least $t - 2(n - 2) - (n - 2) - 2$ are in $S(x, 4)$ since every vertex in $S(x, 2)$ have $2(n - 2)$ neighbors in $S(x, 2)$, $n - 2$ neighbors in $S(x, 3)$ and 2 neighbors in $S(x, 1)$.Therefore, there exists a set S such that

$$S \subset (A_1 \backslash A_0) \cap S(x, 4) \cap \mathcal{N}(T \cap S(x, 2))$$

and

$$\frac{(t - n)(t - 3(n - 2) - 2)}{6} \leqslant |S| \leqslant t^2,$$

since every vertex in $S(x, 4)$ has 6 neighbors in $S(x, 2)$.

Now consider $D = \mathcal{N}(S) \cap S(x, 6)$. Let d be the number of edges between S and D. We can express D as a union of disjoint sets D_i, where $D = \cup_{i=1}^{15} D_i$. Each element in D_i has i neighbors in S. Let $|D_i| = d_i$

and $|S| = s$. Then we have

$$d = \sum_{i=1}^{15} id_i = s\binom{n-4}{2}.$$

Let $R_i = A_0 \cap D_i$ and $|R_i| = r_i$. We have

$$r := \sum_{i=1}^{15} ir_i \geqslant s\left(\frac{N}{2} - 2t - 4 - \binom{4}{2} - (n-4) - \binom{4}{3}(n-4)\right),$$

since every vertex in S has at least $\frac{N}{2} - 2t$ neighbors in A_0 and at most $4 + \binom{4}{2} + (n-4) + \binom{4}{3}(n-4)$ neighbors in $A_0 \backslash S(x, 6)$.

We also have

$$\begin{aligned}
\mathbb{E}(r) &= pd \\
&= p\binom{n-4}{2}s \\
&= (\frac{1}{2} - \delta)\binom{n-4}{2}s \\
&= (\frac{n^2}{4} - \frac{9n}{4} + 5 - \frac{n^2\delta}{2} + \frac{9n\delta}{2} + 10\delta)s,
\end{aligned}$$

where $\delta = \frac{C\sqrt{\log n}}{n}$. Note that the event $\{x \in A_3 \backslash A_2\}$ implies the event $\{r \geqslant \mathbb{E}(r) + \frac{\delta n^2}{4}s\}$. Therefore, we have

$$\mathbb{P}(\cup_{x \in V(Q_{2,n})} x \in A_3 \backslash A_2) \leqslant 2^n \binom{n^2}{t}\binom{n^2 t}{s} t^2 \sum_{d_1, d_2, \ldots d_{15}} \mathbb{P}\left(r \geqslant \mathbb{E}[r] + \frac{\delta n^2}{4}s\right).$$

Indeed, we have at most 2^n vertices in the graph G, $\binom{n^2}{t}$ choices for the set T, $\binom{n^2 t}{s}$ choices for the set S and t^2 values that s can take.

Thus, from Lemma 4.3.7, we have

$$\mathbb{P}\left(r \geqslant \mathbb{E}[r] + \frac{\delta n^2}{4}s\right) \leqslant \left(\frac{\delta n^2}{2}\right)^{14} \exp\left(-\frac{(\delta n^2 s)^2}{16D(15)}\right).$$

Since $D(15) = \sum_{i=1}^{15} i^2 d_i$ and $\sum_{i=1}^{15} i d_i = s\binom{n-4}{2}$, we have

$$D(15) = s\binom{n-4}{2} + \sum_{i=1}^{15}(i^2 - i)d_i \leqslant C's\binom{n-4}{2},$$

where $C' > 0$. Thus we have

$$\mathbb{P}(\exists x \in A_3 \backslash A_2) = o(1). \quad \square$$

Finally, we are ready to prove the lower bound of Theorem 4.1.3 when $k = 2$.

Proof. Consider the process Boot3(t), where the initial infection probability $p = \frac{1}{2} - \frac{C\sqrt{\log n}}{n}$. If we can prove that $\mathbb{P}(A_2 = V(Q_{2,n})) = o(1)$ for Boot3(t), in conjunction with Lemma 4.4.11, it implies that the process Boot3(t) does not percolate with high probability. Consequently, this suggests that the process Boot($\frac{N}{2}$) does not percolate with high probability.

Indeed, we have

$$\mathbb{P}\left(x \notin A_2\right) \geqslant \mathbb{P}(x \notin A_0)\mathbb{P}\left(|\mathcal{N}(x) \cap A_0| < \frac{N}{2} - 2t\right)\mathbb{P}\left(|\mathcal{N}(x) \cap (A_1 \backslash A_0)| = \emptyset\right).$$

It is obvious to see that $\mathbb{P}(x \notin A_0) > \frac{1}{2}$ and by Lemma 4.2.1, that

$$\mathbb{P}\left(|\mathcal{N}(x) \cap A_0| < \frac{N}{2} - 2t\right) \geqslant 1 - \frac{1}{n^4},$$

since C is sufficiently large.

Consider $y \in \mathcal{N}(x) \backslash A_0$. Then we have $\mathbb{P}(y \in A_1) = \mathbb{P}\left(|\mathcal{N}(y) \cap A_0| \geqslant \frac{N}{2} - 2t\right) \leqslant \epsilon(n)$, where

$\epsilon(n) = \frac{1}{n^4}$. Since the events $\{y \notin A_1\}_{y \in \mathcal{N}(x) \setminus A_0}$ are decreasing and thus by Lemma 4.3.2,

$$\begin{aligned}
\mathbb{P}(|\mathcal{N}(x) \cap (A_1 \setminus A_0)| = \emptyset) &\geqslant \prod_{y \in \mathcal{N}(x)} \mathbb{P}(y \notin A_1) \\
&\geqslant (1 - \epsilon(n))^{n^2} \\
&\geqslant e^{-2\epsilon(n)n^2}.
\end{aligned}$$

From Lemmas 4.2.4 and 4.2.5, we can observe that there exists a partition of the set $V(Q_{2,n})$ as follows:

$$V(Q_{2,n}) = \cup_{i=1}^{m} B_i,$$

where for each $i \in [m]$, and for any x and $y \in B_i$, we have $d_H(x, y) \geqslant 9$. Furthermore, the number of subsets m in this partition satisfies $m = \sum_{i=1}^{9} \binom{n}{i}$.

Let $B \subset V(Q_{2,n})$. Assume for every $x \in B$ and $y \in B$, $d_H(x, y) \geqslant 9$. Then the events $\{x \in A_2\}_{x \in B}$ are independent.

Given the partition of $V(Q_{k,n})$, there exists a set B_i such that $|B_i| \geqslant \frac{2^n}{n^{10}}$. Furthermore, for any x and $y \in B_i$, it holds that $d_H(x, y) \geqslant 9$. With the independence condition in place, we have

$$\begin{aligned}
\mathbb{P}(x \in A_2 \text{ for all } x \in B_i) &= (1 - e^{2\epsilon(n)n^2})^{|B_i|} \\
&\leqslant (2\epsilon(n)n^2)^{|B_i|} \\
&= o(1).
\end{aligned}$$

since $\epsilon(n) = \frac{1}{n^4}$. □

4.5 DISCUSSION AND OPEN PROBLEMS

1. Question 1. What is the exact second-order term of $p_c(Q_n, n^a)$ where $\frac{2}{3} < a < 1$?

   For $\frac{2}{3} < a < 1$ we derived the first-order term of $p_c(Q_n, n^a)$ and obtained some insights into the second-order term. It would be interesting to extend this analysis and derive an exact expression for the second-order term.

2. Question 2. What is the exact first-order term of $p_c(Q_n, n^a)$ where $0 < a \leqslant \frac{2}{3}$?

   We conjecture that $p_c(Q_n, n^a) = n^{a-1}(1 + o_n(1))$ for $0 < a \leqslant \frac{2}{3}$. For the infection threshold in this regime, determining the first-order term of the critical probability seems to be beyond the reach of existing techniques, suggesting that entirely new ideas are needed to make progress.

3. Question 3. What is the exact second-order term of $p_c(Q_{n,k}, \frac{N}{2})$ where $N = \sum_{i=1}^{k} \binom{n}{i}$?

   We conjecture that it is on the order $-\frac{\sqrt{\log n}}{n^{k/2}}$, up to a positive constant.

   Let us explain the limitations of our approach. In our proof of the lower bound the first key step is to show that for some $\delta > 0$

   $$\mathbb{P}(x \in A_{1,r}) \geqslant \frac{1}{2} + \delta,$$

   where $x \in V(Q_{k,n})$ and $A_{i,r}$ is the set of infected vertices up to and including step i when the infection threshold is r.

   It is clear that the De Moivre-Laplace Theorem can not yield the desired result because $\Phi(z(N))$ approaches 1 as $n \to \infty$ if $p = \frac{1}{2} - C\frac{\sqrt{\log n}}{n^{k/2}}$ with some constant C.

In [11], the moment method was used to show that

$$\mathbb{P}(x \in A_{2,r}) \geqslant \frac{1}{2} + \delta.$$

The proof was based on the relations $\mathcal{N}(y) \cap \mathcal{N}(z) = \{x, w\}$ and $\mathcal{N}(x) \cap \mathcal{N}(y) = \emptyset$, where $y, z \in \mathcal{N}(x)$ and $x \in V(Q_n)$ which made the application of the second-moment analysis manageable. In our case, these relations are absent, so application of the second-moment analysis lacks a good starting point.

Chapter 5

MAXIMUM PERCOLATION TIME ON THE Q-ARY HYPERCUBE

5.1 INTRODUCTION

In this paper we address a combinatorial question in bootstrap percolation, namely, the maximum percolation time over all contagious sets on q-ary hypercubes with the infection threshold $r = 2$. More specifically we extend the main theorem in [43] about the maximum percolation time with threshold $r = 2$ on the binary hypercubes to the q-ary case.

Let us define the n-dimensional q-ary hypercube, denoted by $Q_{n,q}$, as $V(Q_{n,q}) = \{0, 1, \ldots, q-1\}^n$ and $E(Q_{n,q}) = \{xy : x, y \in \{0, 1, \ldots, q-1\}^n, |\{i : x_i \neq y_i\}| = 1\}$.

In order to simplify the notation, instead of writing $M(Q_{n,q}, 2)$, we write $M_q(n)$ to denote the maximum percolation time on the n-dimensional q-ary hypercube under 2-neighbor bootstrap percolation.

Theorem 5.1.1. *For $q \geqslant 3$, $M_q(0) = 0$, $M_q(1) = 1$ and $M_q(2) = 3$. For larger n,*

$$M_3(n) = \begin{cases} \frac{n^2}{3} + \frac{2n}{3} & \text{if } n = 3, 6, 9, \cdots, \\ \frac{n^2}{3} + \frac{2n}{3} & \text{if } n = 4, 7, 10, \cdots, \\ \frac{n^2}{3} + \frac{2n}{3} + \frac{1}{3} & \text{if } n = 5, 8, 11, \cdots. \end{cases}$$

For $q \geqslant 4$,

$$M_q(n) = \begin{cases} \frac{n^2}{3} + n & \text{if } n = 3, 6, 9, \cdots, \\ \frac{n^2}{3} + n - \frac{1}{3} & \text{if } n = 4, 7, 10, \cdots, \\ \frac{n^2}{3} + n - \frac{1}{3} & \text{if } n = 5, 8, 11, \cdots. \end{cases}$$

5.2 PRELIMINARY RESULTS

We adopt the notation introduced in [8] and [43]. Specifically, we denote by Q_l any of the $\binom{n}{l}q^{n-l}$ subcubes of dimension l in $Q_{n,q}$. For $x = (x_i)_{i=1}^n \in \{0, 1, \ldots, q-1, *\}^n$, let Q^x be the subcube $\{z = (z_i)_{i=1}^n \in \{0, 1, \ldots, q-1\}^n : z_i = x_i \text{ if } x_i \neq *\}$.

We will write $x = x_1 x_2 \cdots x_n$ and $x = (x_1, x_2 \ldots, x_n)$ interchangeably. For convenience and clarity we will write $x = \underbrace{a \cdots a}_{k} \underbrace{b \cdots b}_{l}$ as $[a]^k[b]^l$.

Define the distance between two coordinates

$$d(x_i, y_i) = \begin{cases} 1 & \text{if } x_i \neq y_i, x_i \neq * \text{ and } y_i \neq *, \\ 0 & \text{if either } x_i = * \text{ or } y_i = *, \\ 0 & \text{if } x_i = y_i. \end{cases}$$

Using the above definition we can define the distance between two subcubes Q^x and Q^y in Q_n as

$$d(Q^x, Q^y) = \sum_{i=1}^{n} d(x_i, y_i),$$

where x and y represent subcubes Q^x and Q^y.

Let $\langle A_0 \rangle = \cup_{i=0}^{\infty} A_t$ be the set of all eventually infected vertices under the 2-neighbor process when A_0

is the set of initially infected vertices.

A set $S \subset V(Q_{q,n})$ is closed if every vertex $x \in V(Q_{n,q}) \backslash S$ has at most one neighbor in S. Note that every subcube of a hypercube is closed.

Let us prove some preliminary results by following [8]. Before stating the first lemma we need to give a definition.

For x and $y \in \{0, 1, \ldots, q-1, *\}^n$, define

$$x \vee y = z = (z_i)_{i=1}^n,$$

where

$$(z_i) = \begin{cases} x_i & \text{if } x_i = y_i, \\ * & \text{otherwise.} \end{cases}$$

For $x \in \{0, 1, \ldots, q-1, *\}^n$ we define the dimension of x as $\dim(x) = |\{i : x_i = *\}|$.

If the dimension of a subcube Q^x is l we will sometimes write it as Q_l^x. Similarly if we only know that the cube has dimension at most l we may write it as $Q_{\leqslant l}^x$.

The following three lemmas were proven for the binary hypercube in [8]. For the non-binary case the proofs follow similarly and we give the proofs for completeness.

Lemma 5.2.1. *If the hypercube $Q_{n,q}$ contains a subcube Q_l with $S \subset Q_l$, then $\langle S \rangle \subset Q_l$.*

Proof. Let $S \subset Q_l \subset Q_n$, where Q_l is a subcube of $Q_{n,q}$. Note that $\langle S \rangle$ is the intersection of all closed sets containing S.

Since the subcube Q_l is a closed set containing S, it contains the intersection of all closed sets containing S and thus it contains $\langle S \rangle$. □

Lemma 5.2.2. *Let x and y be vectors $\in \{0, 1, \cdots, q-1, *\}^n$ with $z = x \vee y$ and $d(x, y) \leqslant 2$. If $S \subset Q_l^x \cup Q_k^y$, then $\langle S \rangle \subset Q^z_{\leqslant l+k+2}$.*

Proof. Since $S \subset Q_l^x \cup Q_k^y \subset Q^z$ and the subcube Q^z is closed, we have $\langle S \rangle \subset Q^z$. Since the number of $*$ in the coordinates of z is at most $l + k + d(x, y)$, the dimension of the subcube Q^z is $\leqslant l + k + 2$. □

Lemma 5.2.3. *For x and $y \in \{0, 1, \ldots, q-1, *\}^n$ with $d(x, y) \leqslant 2$, we have*

$$\langle Q^x \cup Q^y \rangle = Q^{x \vee y}.$$

Proof. Since the subcube $Q^{x \vee y}$ is closed and $Q^x \cup Q^y \subset Q^{x \vee y}$, we have $\langle Q^x \cup Q^y \rangle \subset Q^{x \vee y}$. by Lemma 5.2.2. What remains is to show $Q^{x \vee y} \subset \langle Q^x \cup Q^y \rangle$.

W.l.o.g we can assume that

$$x = [*]^k [0]^{n-k}$$

$$y = [*]^j [0]^{k-j} [*]^{l-j} [0]^{n-k-l+j-2} ab,$$

where $a, b \in \{1, 2, ..., q-1\}$ and $0 \leqslant j \leqslant k \leqslant n - l + j - 2$.

We will use the following claim several times in the proof.

Claim 5.2.4. *Let A be a subset of $V(Q_{n,q})$. If a vertex u has at least 2 neighbors in A, then $u \in \langle A \rangle$.*

Proof. Let $u = u_1 u_2 \cdots u_n$. Since u has at least 2 neighbors in A, w.l.o.g the two neighbors y and z of u are $y = a u_2 u_3 \cdots u_n$ and $z = b u_2 \cdots u_n$ where $a \notin \{b, u_1\}$ and $b \notin \{a, u_1\}$ or $y = a u_2 u_3 \cdots u_n$ and $z = u_1 b u_3 \cdots u_n$ where $a \neq u_1$ and $b \neq u_2$.

If $y = a u_2 u_3 \cdots u_n$ and $z = b u_1 u_2 \cdots u_n$, then $\langle y \cup z \rangle = * u_2 u_3 \cdots u_n$ and thus $u \in \langle A \rangle$.

If $y = a u_2 u_3 \cdots u_n$ and $z = u_1 b u_3 \cdots u_n$, then $\langle y \cup z \rangle = * * u_3 u_4 \cdots u_n$ and thus $u \in \langle A \rangle$. □

For $0 \leqslant i \leqslant k+l-2j$, let z_i be

$$z_i = [*]^{i+j}0^{n-i-j-2} * *.$$

We will prove by induction on i that $Q^{z_i} \subset \langle Q^x \cup Q^y \rangle$ for $0 \leqslant i \leqslant k+l-2j$, which is sufficient to prove this lemma since $z_{k+l-2j} = x \vee y$.

Consider a vertex u in $Q^{z_0} \backslash (Q^x \cup Q^y)$. If $u = u_1u_2 \cdots u_{n-2}a0$ or $u = u_1u_2 \cdots u_{n-2}0b$ then it has one neighbor in Q^x and one neighbor in Q^y. Then by Claim 5.2.4, $u \in \langle Q^x \cup Q^y \rangle$.

For a vertex $u = u_1u_2 \cdots u_n$ we will write $u^{ab} = u_1u_2 \cdots u_{n-2}ab$. Now if $u = u_1u_2 \cdots u_{n-2}aj$ with $j \notin \{0, b\}$ then it has two neighbors $u^{ab} = u_1u_2 \cdots u_{n-2}ab$ and $u^{a0} = u_1u_2 \cdots u_{n-2}a0$ and thus $u \in \langle Q^x \cup Q^y \rangle$. Similarly, a vertex $u = u_1u_2 \cdots u_{n-2}jb$ for $j \notin \{0, a\}$ is also in $\langle Q^x \cup Q^y \rangle$. Finally if $u = u_1u_2 \cdots u_{n-2}ij$ where $i \notin \{0, a\}$ and $j \notin \{0, b\}$, it has two neighbors in Q^x and Q^y and thus $u \in \langle Q^x \cup Q^y \rangle$. Indeed, one of them is $u_1u_2 \cdots u_{n-2}u_{n-1}b$ and another one is $u = u_1u_2 \cdots u_{n-2}au_n$.

Now assume that for $i \geqslant 0$, $Q^{z_i} \subset \langle Q^x \cup Q^y \rangle$ and our goal is to show that $Q^{z_{i+1}} \subset \langle Q^x \cup Q^y \rangle$.

Let a vertex $u \in Q^{z_{i+1}} \backslash (Q^{z_i} \cup Q^x \cup Q^y)$. Thus u has a neighbor in Q_{z_i} and thus we would like to find another neighbor of u in $\langle Q^x \cup Q^y \rangle$.

First let us consider the case when $i \leqslant k-j-1$. Since $u \notin Q^x$, the last two digits of of u cannot be both 0. If $u = u_1u_2 \cdots u_{n-2}a0$, then it has a neighbor is Q^x and by Lemma 5.2.4 $u \in \langle Q^x \cup Q^y \rangle$. Similarly, $u^{0b} \in \langle Q^x \cup Q^y \rangle$ and it is easy to see that $u^{**} \in \langle Q^x \cup Q^y \rangle$.

Now assume that $i \geqslant k-j$. Let $u = u_1u_2 \cdots u_{n-2}ab$. Let r be the smallest value in the range $0 \leqslant r \leqslant k-j$ such that $u_{j+r+1} = \cdots = u_k = 0$. We apply induction on r to prove that $u \in \langle Q^x \cup Q^y \rangle$. Let $r = 0$, then $u \in Q^y \subset \langle Q^x \cup Q^y \rangle$. Let us turn to the induction step. Thus $1 \leqslant r \leqslant k-j$,$u_{j+r} = 1$ and $u_{j+r+1} = \cdots = u_k = 0$. Since $u \notin Q^{z_i}$, we have $u_{i+j+1} \neq 0$, and so switching u_{i+j+1} to 0 we obtain a neighbor of u in Q^{z_i}. Since $r \leqslant k-j \leqslant i$, then $j+r \neq i+j+1$. Therefore if we change the

$(j+r)$th digit of u to 0 then by the induction hypothesis we get a vertex in $\langle Q^x \cup Q^y \rangle$. Thus by Claim 5.2.4 $u \in \langle Q^x \cup Q^y \rangle$.

Then it is easy to see that $u^{ij} \in \langle Q^x \cup Q^y \rangle$ for all $i, j \in \{0, 1, \cdots, q-1\}$. □

The following three lemmas were proven for the binary hypercube in [8] and for the non-binary case the proofs are identical, by applying Lemmas (5.2.1),(5.2.2), and (5.2.3). Therefore we ommit the proofs.

Lemma 5.2.5. *(i)The only subsets of the hypercube that are closed under 2-neighbor percolation are those which are a union of disjoint subcubes that are at distance at least 3 from each other.*

(ii) If a subcube $Q_l = \langle S \rangle$, then $|S| \geqslant \frac{l}{2} + 1$.

Before stating the next lemma, we need to give a definition. For a given bootstrap percolation process on $Q_{n,q}$ with an initially infected set of vertices $A \subset Q_{n,q}$, a subcube $Q_l \subset Q_{n,q}$ is said to be internally spanned if $\langle A \cap Q_l \rangle = Q_l$.

Lemma 5.2.6. *If the initially infected set percolates on the hypercube $Q_{n,q}$, then for each $k \leqslant n$ there is an integer l such that $k \leqslant l \leqslant 2k$ and the hypercube $Q_{n,q}$ contains an internally spanned cube of dimension l.*

Lemma 5.2.7. *Let $A \subset Q_{n,q}$ be such that $\langle A \rangle = Q_{n,q}$. Then there is a nested sequence $Q_0 = Q_{i_1}^{x_{i_1}} \subset Q_{i_2}^{x_{i_2}} \subset \cdots \subset Q_{i_t}^{x_{i_t}} = Q_{n,q}$, of internally spanned subcubes with respect to A, where $2i_j + 2 \geqslant i_{j+1}$ for all j, $0 \leqslant j \leqslant t-1$. Furthermore, for $j \geqslant 2$ each subcube $Q_{i_j}^{x_{i_j}}$ is spanned by two internally spanned cubes, namely by $Q_{i_{j-1}}^{x_{i_{j-1}}}$ and a subcube $Q_{m_{j-1}}$ with $m_{j-1} \leqslant i_{j-1}$ which is not a member of the sequence.*

Let A_0 be the set of initially infected vertices in a graph G and let

$$T_G(A_0) = \min\{t : A_t = \langle A_0 \rangle\}.$$

For the n-dimensional q-ary hypercube its maximum percolation time is defined to be

$$M_q(n) := \max_{A:\langle A\rangle = Q_{n,q}} T_{Q_{n,q}}(A).$$

5.3 MAXIMUM PERCOLATION TIME

In this section, we prove Theorem 5.1.1 by drawing on ideas from [43], beginning with a lemma on the monotonicity of the maximum percolation time. We then define several new operations, which are slight modifications of those introduced in [43]. With those defined operations, we then establish six technical lemmas (which are similar to those established in [43] but more technical) that characterize the spread of infection on $Q_{n,q}$ based on the set of initially infected vertices. Using these lemmas along with some preliminary results, we derive a recursive formula for the maximum percolation time, ultimately obtaining the desired closed-form expression for $M_{n,q}$.

Lemma 5.3.1. *For any $n \in \mathbb{N}$, $M_q(n) \leqslant M_q(n+1)$.*

Proof. Let A_0 be the set of initially infected vertices such that $\langle A_0\rangle = Q_{n,q}$ and $T_{Q_{n,q}}(A_0) = M_q(n)$. Let

$$A_0^{'} = \{(a_1, \ldots, a_n, i) : (a_1, \ldots, a_n) \in A_0 \text{ and } i \in \{0, \ldots, q-1\}\}.$$

Then $\langle A_0^{'}\rangle = Q_{n+1,q}$ and $M_q(n) = T_{Q_{n,q}}(A_0) = T_{Q_{n+1}}(A_0^{'}) \leqslant M_q(n+1)$. □

We need to define the following operation. Let $n, d, n_1, n_2 \in \{1, 2, \ldots\}$ with $n \geqslant d + n_1 + n_2$ and $a_i \in \{1, 2, \ldots, q-1\}$ for all $i \in [d]$. For each $x \in \{0, 1, \cdots, q-1\}^n$ with $x = x_1 x_2 \ldots x_n$, set

$$\|x\|_{n_1,n_2}^{a_1,\cdots,a_d} = \left(\sum_{i=1}^{n-d} \mathbb{I}_{\{x_i \neq 0\}}\right) \mathbb{I}_{\{\sum_{i=1}^{n_1} x_i > 0\}} \mathbb{I}_{\{\sum_{i=n_1+1}^{n_1+n_2} x_i > 0\}} \prod_{i=1}^{d} \mathbb{I}_{\{x_{n-d+i} = a_i\}}.$$

In words, $\|x\|_{n_1,n_2}^{a_1,\cdots,a_d}$ is the number of the none-zero elements in $\{x_1, \ldots, x_{n-d}\}$ if the following three

conditions are satisfied:

(1) There is at least one nonzero element in $\{x_1, x_2, \ldots, x_{n_1}\}$,

(2) There is at least one nonzero element in $\{x_{n_1+1}, \ldots, x_{n_1+n_2}\}$,

and

(3) The last d digits of x are exactly equal to $a_1, a_2, \ldots, a_d$, i.e., $x_{n-d+i} = a_i$ for all $i \in [d]$.

By the same spirit we will define

$$||x|| := \sum_{i=1}^{n} \mathbb{I}_{\{x_i \neq 0\}},$$

$$||x||^{a_1,\ldots,a_d} := \left(\sum_{i=1}^{n-d} \mathbb{I}_{\{x_i \neq 0\}}\right) \prod_{i=1}^{d} \mathbb{I}_{\{x_{n-d+i}=a_i\}},$$

$$\|x\|_{n_1}^{a_1,\cdots,a_d} := \left(\sum_{i=1}^{n-d} \mathbb{I}_{\{x_i \neq 0\}}\right) \mathbb{I}_{\{\sum_{i=1}^{n_1} x_i > 0\}} \prod_{i=1}^{d} \mathbb{I}_{\{x_{n-d+i}=a_i\}},$$

$$\|x\|_{n_1} := \left(\sum_{i=1}^{n} \mathbb{I}_{\{x_i \neq 0\}}\right) \mathbb{I}_{\{\sum_{i=1}^{n_1} x_i > 0\}},$$

and

$$\|x\|_{n_1,n_2} := \left(\sum_{i=1}^{n} \mathbb{I}_{\{x_i \neq 0\}}\right) \mathbb{I}_{\{\sum_{i=1}^{n_1} x_i > 0\}} \mathbb{I}_{\{\sum_{i=n_1+1}^{n_1+n_2} x_i > 0\}}.$$

We now proceed to prove six technical lemmas that provide useful characterizations of the spread of infection for various configurations of initially infected vertices.

Recall that A_0 denotes the set of initially infected vertices on $Q_{n,q}$.

Lemma 5.3.2. *Let $k, l \in \mathbb{N}_0$, $n = k + l$. Let $S = [*]^k[0]^l$ and $T = [0]^k[*]^l$. With $A_0 = S \cup T$, then we have*

$$A_t \supset \{x \in \{0, 1, \ldots, q-1\}^n : \|x\| \leqslant t + 1\},$$

for all $t \in \mathbb{N}_0$.

Proof. Let $t = 0$. For every vertex x with $\|x\| \leqslant 1$, we have $x \in A_0$.

Now assume that for $t \geqslant 0$

$$A_t \supset \{x \in \{0, 1, \ldots, q-1\}^n : \|x\| \leqslant t+1\},$$

Consider a vertex x with $\|x\| = t + 2 \geqslant 2$. This vertex x has at least 2 neighbors y with $\|y\| = t + 1$. Therefore, $x \in A_{t+1}$. Thus

$$A_{t+1} \supset \{x \in \{0, 1, \ldots, q-1\}^n : \|x\| \leqslant t+2\}.$$

□

Lemma 5.3.3. *Let $k, l \in \mathbb{N}_0$, $n = k + l + 1$. Let $S = [*]^k[0]^{l+1}$ and $T = [0]^k[*]^l i$, where $i \neq 0$.With $A_0 = S \cup T$, then we have*

$$A_t \supset \left([0]^{k+l}*\right) \bigcup_{j=0,i} \{x \in \{0, 1, \ldots, q-1\}^n : 1 \leqslant \|x\|^j \leqslant t\} \bigcup_{j\neq 0,i} \{x \in \{0, 1, \ldots, q-1\}^n : 1 \leqslant \|x\|^j \leqslant t-1\},$$

for all $t \in \mathbb{N}$.

Proof. Let $t = 1$. It is easy to see every vertex in $[0]^{k+l}*$ has two neighbors in $S \cup T$. Therefore, $[0]^{k+l}* \subset A_1$. Now consider x such that $\|x\|^0 = 1$. Either $x \in S$ or there exists exactly j with $k < j \leqslant n-1$ such that $x_j \neq 0$. For the latter case, x has one neighbor in S and one neighbor in T. For the case where $\|x\|^i$, the analysis is very similar. Therefore

$$A_1 \supset \left([0]^{k+l}*\right) \bigcup_{j=0,i} \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|^j = 1\}.$$

Assume for $t \geqslant 1$,

$$A_t \supset \bigcup_{j=0,i} \{x \in \{0,1,\ldots,q-1\}^n : 1 \leqslant \|x\|^j \leqslant t\}.$$

Consider $\|x\|^0 = t+1$. It is easy to see that x has two neighbors y with $\|y\|^0 = t$. Therefore, $x \in A_{t+1}$. Similarly, $x \in A_{t+1}$ with $\|x\|^i = t+1$. Thus

$$A_{t+1} \supset \bigcup_{j=0,i} \{x \in \{0,1,\ldots,q-1\}^n : 1 \leqslant \|x\|^j \leqslant t+1\}.$$

Now let $t \geqslant 2$ and consider x with $\|x\|^j = t-1$ where $j \notin \{0,i\}$. Note that x has one neighbor y_1 in $\{y : \|y\|^0 = t-1\}$ and one neighbor y_2 in $\{y : \|y\|^i = t-1\}$. Moreover, $y_1, y_2 \in A_{t-1}$. Therefore $x \in A_t$. □

Lemma 5.3.4. *Let $k, l \in \mathbb{N}_0$, $n = k+l+2$. Let $S = [*]^k[0]^{l+2}$ and $T = [0]^k[*]^l ab$. With $A_0 = S \cup T$, where $a, b \neq 0$, then we have*

$$A_1 \supset \{[0]^{k+l}0b, [0]^{k+l}a0, [0]^{k+l}ab, [0]^{k+l}00\},$$

$$A_2 \supset \{[0]^{k+l}0*, [0]^{k+l}a*, [0]^{k+l}*0, [0]^{k+l}*b\},$$

$$A_3 \supset \{[0]^{k+l}**\},$$

and for all $t \geqslant 2$,

$$A_t \supset \{[0]^{k+l} * *\} \bigcup \{x \in \{0,1,\ldots,q-1\}^n : 1 \leqslant \|x\|^{0b} \leqslant t-1\} \tag{5.1}$$

$$\bigcup \{x \in \{0,1,\ldots,q-1\}^n : 1 \leqslant \|x\|^{a0} \leqslant t-1\} \tag{5.2}$$

$$\bigcup_{i \neq b} \{x \in \{0,1,\ldots,q-1\}^n : 1 \leqslant \|x\|^{0i} \leqslant t-2\} \tag{5.3}$$

$$\bigcup_{j \neq a} \{x \in \{0,1,\ldots,q-1\}^n : 1 \leqslant \|x\|^{j0} \leqslant t-2\} \tag{5.4}$$

$$\bigcup_{i \neq 0} \{x \in \{0,1,\ldots,q-1\}^n : 1 \leqslant \|x\|^{ai} \leqslant t-2\} \tag{5.5}$$

$$\bigcup_{j \neq 0} \{x \in \{0,1,\ldots,q-1\}^n : 1 \leqslant \|x\|^{jb} \leqslant t-2\} \tag{5.6}$$

$$\bigcup_{\substack{c \neq 0,a \\ d \neq 0,b}} \{x \in \{0,1,\ldots,q-1\}^n : 1 \leqslant \|x\|^{cd} \leqslant t-3\}. \tag{5.7}$$

Proof. Consider the vertex $[0]^{k+l}0b$ and it has one neighbor in S and one neighbor in T. Therefore this vertex is in A_1. The same holds for the vertex $[0]^{k+l}a0$. Note that both $[0]^{k+l}ab$ and $[0]^{k+l}00$ are in A_0.

First consider the vertex $[0]^{k+l}0j$ where $j \notin \{0,b\}$. It has two neighbors in A_1, namely, $[0]^{k+l}0b$ and $[0]^{k+1}00$. The same holds for the vertices in $\{[0]^{k+l}0*, [0]^{k+l}a*, [0]^{k+l} * 0, [0]^{k+l} * b\}$. Now Consider the vertex $[0]^{k+l}ij$ where $i \notin \{0,a\}$ and $j \notin \{0,b\}$.This vertex has at least 2 neighbors in A_2. Therefore, this vertex is in A_3.

Now let $t = 2$ and consider a vertex x with $\|x\|^{0b} = 1$. It has at least 2 neighbors in A_1. Indeed, one of the neighbors of vertex x is $[0]^{k+l}0b$ and another one is in S if there exists a j with $1 \leqslant j \leqslant k$ such that $x_j \neq 0$ or in T if there exists a j with $k+1 \leqslant j \leqslant k+l$ such that $x_j \neq 0$. The analysis is the same for x with $\|x\|^{a0} = 1$.

Now assume that for $t \geqslant 2$,

$$A_t \supset \{x \in \{0,1,\cdots,q-1\}^n : 1 \leqslant \|x\|^{0b} \leqslant t-1\} \bigcup \{x \in \{0,1,\cdots,q-1\}^n : 1 \leqslant \|x\|^{a0} \leqslant t-1\}.$$

Consider a vertex x with $\|x\|^{0b} = t$ and it has 2 neighbors y satisfying $\|y\|^{0b} = t-1$. By the hypothesis both neighbors are in A_t and thus $x \in A_{t+1}$. The analysis for the case where the vertex x satisfies $\|x\|^{a0} = t$ is the same. This finishes the proof of (5.1) and (5.2).

Now let us prove (5.3)–(5.6). Let $t = 3$ and consider a vertex x with $\|x\|^{0i} = 1$. The vertex x has 2 neighbors in A_2. Indeed one is $[0]^{k+l}0i$ and another neighbor is y satisfying $\|y\|^{0b} = 1$. Therefore, the vertex x is in A_3. The analysis is the same for the vertex x satisfying (5.4),(5.5), or (5.6). Thus, for $t = 3$, (5.3)–(5.6) are satisfied.

Let us assume for $t \geqslant 3$,

$$A_t \supset \bigcup_{i \neq b} \{x \in \{0,1,\cdots,q-1\}^n : 1 \leqslant \|x\|^{0i} \leqslant t-2\}.$$

Now consider a vertex x satisfying $\|x\|^{0i} = t-1$. It is easy to see that x has at least 2 neighbors y satisfying $\|y\|^{0i} = t-2$. Therefore, x is in A_{t+1}. Similarly, (5.4),(5.5), and (5.6) can be proved.

Now we will prove (5.7). Let $t = 4$ and consider a vertex x satisfying $\|x\|^{cd} = 1$ with $c \notin \{0,a\}$ and $d \notin \{0,b\}$. Note that x has at least two neighbors in A_3, of one which is $[0]^{k+l}cd$ and another is y satisfying $\|y\|^{c0} = 1$.

We assume that for $t \geqslant 4$,

$$A_t \supset \bigcup_{\substack{c \neq 0,a \\ d \neq 0,b}} \{x \in \{0,1,\ldots,q-1\}^n : 1 \leqslant \|x\|^{cd} \leqslant t-3\}.$$

Consider a vertex x satisfying $\|x\|^{cd} = t-2$ and it has at least two neighbors y satisfying $\|y\|^{cd} = t-3$.

Thus the vertex x is in A_{t+1}. □

Lemma 5.3.5. *Let $k, l \in \mathbb{N}_0$, $n = k + l$. Let $S = [*]^k[0]^l$ and $T = [0]^k[*]^l$. With $A_0 = S \cup T$, then we have*

$$A_t \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_{k,l} \geqslant t + 2\} = \emptyset,$$

for all $0 \leqslant t \leqslant k + l - 2$.

Proof. Let $t = 0$ and consider a vertex x with $\|x\|_{k,l} = 2$. Then it is clear that $x \notin A_0$. Now let $t \geqslant 0$ and assume that

$$A_t \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_{k,l} \geqslant t + 2\} = \emptyset.$$

Consider a vertex x with $\|x\|_{k,l} \geqslant t + 3$ and it has at most one neighbor y satisfying $\|y\|_{k,l} < t + 2$. Indeed if $\sum_{j=1}^{k} \mathbb{I}_{\{x_j \neq 0\}} = 1$, we have $y = [0]^k x_{k+1} \cdots x_n$ and if $\sum_{j=k+1}^{n} \mathbb{I}_{\{x_j \neq 0\}} = 1$, we have $y = x_1 \cdots x_k [0]^l$. Therefore, $x \notin A_{t+1}$. □

Lemma 5.3.6. *Let $k, l \in \mathbb{N}_0$, $n = k + l + 1$. Let $S = [*]^k[0]^{l+1}$ and $T = [0]^k[*]^l i$ where $i \neq 0$. With $A_0 = S \cup T$, then we have the following.*

For all $0 \leqslant t \leqslant k + l - 1$,

$$A_t \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^i \geqslant t + 1\} = \emptyset.$$

For $1 \leqslant t \leqslant k + l$ and $j \notin \{0, i\}$,

$$A_t \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^j \geqslant t\} = \emptyset.$$

Proof. We will first prove that for all $0 \leqslant t \leqslant k+l-1$,

$$A_t \cap \{x \in \{0,1,\ldots,q-1\}^n : \|x\|_k^j \geqslant t+1\} = \emptyset,$$

where $j \neq 0$.

Let $t = 0$ and consider a vertex x with $\|x\|_k^j \geqslant 1$ where $j \neq 0$. It is easy to see that $x \notin A_0$. Now let $t \geqslant 0$ and assume that for $j \neq 0$,

$$A_t \cap \{x \in \{0,1,\ldots,q-1\}^n : \|x\|_k^j \geqslant t+1\} = \emptyset.$$

Then consider a vertex x with $\|x\|_k^j \geqslant t+2$ with $j \neq 0$. It has $q-1$ neighbors y such that $\|y\|_k^j \leqslant t$, obtained by changing x_n to one of the other $q-1$ symbols in $\{0,1,\ldots,q-1\}$. By the induction hypothesis, among these $q-1$ neighbors, at most one $y = x_1x_2\cdots x_{n-1}0$ may be in A_t. If $\sum_{i=1}^{k} \mathbb{I}_{\{x_i \neq 0\}} = 1$, another neighbor z of x satisfying $\|z\|_k^j = 0$ has not been considered. Indeed, $z = [0]^k x_{k+1}\cdots x_{n-1}j$.

We claim that if z exists then $z \notin A_t$. If such a z exits, we have $\sum_{i=1}^{k} \mathbb{I}_{\{x_i \neq 0\}} = 1$ and $\sum_{i=k+1}^{k+l} \mathbb{I}_{\{x_i \neq 0\}} \geqslant t+1$. Now let $z' = (z_{k+1},\ldots,z_{k+l},z_1,\cdots,z_k,z_{k+l+1})$. Let $S' = [*]^l[0]^{k+1}$ and $T' = [0]^l[*]^k i$. Then we have $\|z'\|_l^j \geqslant t+1$. By the induction hypothesis, it follows that $z' \notin A_t^{'}$, where $A_t^{'}$ is the set of infected vertices after t steps by the 2 neighbor process with the initially infected set $S' \cup T'$. By symmetry, $z \notin A_t$. Therefore, $x \notin A_{t+1}$.

Now we will try to prove for $1 \leqslant t \leqslant k+l$ and $j \notin \{0,i\}$,

$$A_t \cap \{x \in \{0,1,\ldots,q-1\}^n : \|x\|_k^j \geqslant t\} = \emptyset. \tag{5.8}$$

Let $t = 1$ and consider a vertex x with $\{0,1,\ldots,q-1\}^n : \|x\|_k^j \geqslant t\}$ where $j \notin \{0,i\}$. It is clear that this vertex x has at most one neighbor in A_0 specifically in S. Therefore, $x \notin A_1$.

Now assume that for $t \geqslant 1$, (5.8) is satisfied. Now consider a vertex x with $\|x\|_k^j \geqslant t+1$. Consider the neighbors y of x satisfying $\|x\|_k^j \leqslant t-1$. Specifically, we have y satisfying $y = x_1 \cdots x_{n-1} m$ for $m \neq j$. If $m = i$, we have $\|y\|_k^i \geqslant t+1$ implying that $y \notin A_t$. If $m = 0$, y might be in A_t. If $m \notin \{0, i, j\}$, we have $\|y\|_k^i \geqslant t+1$, by the induction hypothesis, implying that $y \notin A_t$.

If $\sum_{i=1}^k \mathbb{I}_{\{x_i \neq 0\}} = 1$, another neighbor z of x satisfying $\|z\|_k^j = 0$ has not been considered. Indeed, $z = [0]^k x_{k+1} \cdots x_{n-1} j$. By the same argument, $z \notin A_t$. Therefore, $x \notin A_{t+1}$. □

Lemma 5.3.7. *Let $k, l \in \mathbb{N}_0$, $n = k + l + 2$. Let $S = [*]^k[0]^{l+2}$ and $T = [0]^k[*]^l ab$ where $a, b \neq 0$. With $A_0 = S \cup T$, then we have the following.*

For $t \geqslant 1$,

$$A_t \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{0b} \geqslant t\} = \emptyset,$$

and

$$A_t \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{a0} \geqslant t\} = \emptyset.$$

For $j \notin \{b, 0\}$ and $t \geqslant 2$,

$$A_t \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{0j} \geqslant t-1\} = \emptyset.$$

For $i \notin \{a, 0\}$ and $t \geqslant 2$,

$$A_t \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{i0} \geqslant t-1\} = \emptyset.$$

For $j \neq 0$ and $t \geqslant 2$,

$$A_t \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{aj} \geqslant t-1\} = \emptyset.$$

For $i \neq 0$ and $t \geqslant 2$,

$$A_t \cap \{x \in \{0,1,\ldots,q-1\}^n : \|x\|_k^{ib} \geqslant t-1\} = \emptyset.$$

For $i \notin \{a,0\}$, $j \notin \{b,0\}$, and $t \geqslant 3$,

$$A_t \cap \{x \in \{0,1,\ldots,q-1\}^n : \|x\|_k^{ij} \geqslant t-2\} = \emptyset.$$

Proof. Let $t = 1$ and consider a vertex x with $\|x\|_k^{0b} \geqslant 1$ where $b \neq 0$. It is easy to see that this vertex x has at most one neighbor in A_0 specifically in S. Thus $x \notin A_1$. Similarly, the vertex y with $\|y\|_k^{a0} \geqslant 1$ is not in A_1.

Now let $t = 2$ and consider a vertex x with $\|x\|_k^{0j} \geqslant 1$ where $j \notin \{0,b\}$. Note that the vertex x has at most one neighbor in A_0 specifically in S. Thus $x \notin A_1$. Similarly, the vertex y with $\|y\|_k^{i0} \geqslant 1$, where $i \notin \{a,0\}$, is not in A_1.

Again let $t = 2$ and consider a vertex x with $\|x\|_k^{aj} \geqslant 1$ where $j \neq 0$. Note that the vertex x has no neighbor in A_0 unless $\sum_{i=1}^k \mathbb{I}_{\{x_i>0\}} = 1$ and $x_{n-1} = a$ and $x_n = b$. It is easy to see that even for a such x it has just one neighbor in A_0. Thus $x \notin A_1$. Similarly, the vertex y with $\|y\|_k^{ib} \geqslant 1$ where $i \neq 0$, is not in A_1.

Now let $t = 3$ and consider a vertex x with $\|x\|_k^{ij} \geqslant 1$ where $i \neq a, 0$ and $j \notin \{0,b\}$. Note that the vertex x has no neighbor in A_0. Thus $x \notin A_1$.

The neighbors y of a vertex x with $\|x\|_k^{0j} \geqslant 1$ satisfying $\|y\|_k^{0j} = 0$ will be considered since the other neighbors of x will not be in A_1. Specifically, we only need to analyze $y = x_1 \cdots x_{n-2}00$, $y = x_1 \cdots x_{n-2}0i$ with $i \notin \{0,j\}$, $y = x_1 \cdots x_{n-2}ij$ with $i \neq 0$ and $y = [0]^k x_{k+1} \cdots x_{n-2}0j$ if $\sum_{i=1}^k \mathbb{I}_{\{x_i \neq 0\}} = 1$. The only neighbors y of x which are possibly in A_1 are the ones satisfying $y = x_1 \cdots x_{n-2}00$ and $y = [0]^k x_{k+1} \cdots x_{n-2}0j$. However, it is clear that the vertex y with $y = [0]^k x_{k+1} \cdots x_{n-2}0j$ is not in

A_1 since it has at most one neighbor in A_0 specifically in S. Therefore, the vertex $x \notin A_2$. Similarly, the vertex x with $\|x\|_k^{i0} \geqslant 1$ where $i \notin \{a, 0\}$, is not in A_2.

Now we will show that the vertex x with $\|x\|_k^{aj} \geqslant 1$ where $j \neq 0$, is not in A_2. Again we only need to consider the neighbors y of a vertex x with $\|x\|_k^{aj} \geqslant 1$ satisfying $\|y\|_k^{aj} = 0$. Specifically, $y = x_1 \cdots x_{n-2}ai$ with $i \neq j$, $y = x_1 \cdots x_{n-2}ij$ with $i \neq a$ and $y = [0]^k x_{k+1} \cdots x_{n-2}aj$ if $\sum_{i=1}^{k} \mathbb{I}_{\{x_i \neq 0\}} = 1$. The only neighbor y of x possibly in A_1 is the one satisfying $y = [0]^k x_{k+1} \cdots x_{n-2}aj$.Therefore, the vertex $x \notin A_2$. Similarly, the vertex x with $\|x\|_k^{ib} \geqslant 1$ where $i \neq 0$, is not in A_2.

Now let us turn to proving that the vertex x with $\|x\|_k^{ij} \geqslant 1$, where $i \notin \{0, a\}$ and $j \notin \{0, b\}$, is not in A_3. Note that x does not have any neighbor in A_0 so that $x \notin A_1$. Let us first show that $x \notin A_2$. The only neighbors y of x possibly in A_1 satisfy $\|y\|_k^{ij} = 0$. Specifically, $y = x_1 \cdots x_{n-2}aj$ with $a \neq i$, $y = x_1 \cdots x_{n-2}ia$ with $a \neq j$ and $y = [0]^k x_{k+1} \cdots x_{n-2}ij$ if $\sum_{i=1}^{k} \mathbb{I}_{\{x_i \neq 0\}} = 1$. The only neighbor y of x possibly in A_1 is the one satisfying $y = [0]^k x_{k+1} \cdots x_{n-2}ij$. Therefore, $x \notin A_2$. Note that the vertex y with $y = [0]^k x_{k+1} \cdots x_{n-2}ij$ is not in A_1 since it does not have any neighbor in A_0. Very similarly, the only neighbors y of x possibly in A_2 satisfy $\|y\|_k^{ij} = 0$. Specifically, $y = x_1 \cdots x_{n-2}aj$ with $a \neq i$, $y = x_1 \cdots x_{n-2}ia$ with $a \neq j$ and $y = [0]^k x_{k+1} \cdots x_{n-2}ij$ if $\sum_{i=1}^{k} \mathbb{I}_{\{x_i \neq 0\}} = 1$. The only neighbor y of x possibly in A_2 is the one satisfying $y = [0]^k x_{k+1} \cdots x_{n-2}ij$. Therefore, $x \notin A_3$.

Now we will move on to the inductive step. Now assume we have the following.

For $t \geqslant 1$,

$$A_t \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{0b} \geqslant t\} = \emptyset,$$

and

$$A_t \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{a0} \geqslant t\} = \emptyset.$$

For $j \notin \{b, 0\}$ and $t \geqslant 1$,

$$A_{t+1} \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{0j} \geqslant t\} = \emptyset.$$

For $i \notin \{a, 0\}$ and $t \geqslant 1$,

$$A_{t+1} \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{i0} \geqslant t\} = \emptyset.$$

For $j \neq 0$ and $t \geqslant 1$,

$$A_{t+1} \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{aj} \geqslant t\} = \emptyset.$$

For $i \neq 0$ and $t \geqslant 1$,

$$A_{t+1} \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{ib} \geqslant t\} = \emptyset.$$

For $i \notin \{a, 0\}$, $j \notin \{b, 0\}$, and $t \geqslant 1$,

$$A_{t+2} \cap \{x \in \{0, 1, \ldots, q-1\}^n : \|x\|_k^{ij} \geqslant t\} = \emptyset.$$

Let us consider a vertex x satisfying $\|x\|_k^{0b} \geqslant t+1$. The neighbors y of the vertex x must satisfy $\|y\|_k^{0b} \geqslant t$, $\|y\|_k^{0j} \geqslant t+1$ for $j \notin \{b, 0\}$, $\|y\|_k^{00} \geqslant t+1$, $\|y\|_k^{ib} \geqslant t+1$ where $i \neq 0$ or $y = [0]^k x_{k+1} \cdots x_n$. From the induction hypothesis only two neighbors y and z of x satisfying $\|y\|_k^{00} \geqslant t+1$ and $z = [0]^k x_{k+1} \cdots x_n$ can possibly be in A_t. However we claim that $z \notin A_t$.

Now let us prove the claim. In order for a such z to exist, we must have $\sum_{i=1}^{k} \mathbb{I}_{\{x_i \neq 0\}} = 1$ and $\sum_{i=k+1}^{k+l} \mathbb{I}_{\{x_i \neq 0\}} \geqslant t$. Let $z' = z_{k+1} \cdots z_{k+l} z_1 \cdots z_k z_{k+l+1} z_{k+l+2}$ and then $\|z'\|_l^{0b} \geqslant t+1$. Now let

$S' = [*]^l[0]^{k+2}$ and $T' = [0]^l[*]^k ab$. Then by the induction hypothesis, we have $z' \notin A'_t$, where A'_t is the set of infected vertices with the initial infected set $S' \cup T'$. Thus $z \notin A_t$. Now we can conclude $x \notin A_{t+1}$. Similarly, the vertex x satisfying $\|x\|_k^{a0} \geqslant t+1$ is not in A_{t+1}.

Now consider a vertex x satisfying $\|x\|_k^{0j} \geqslant t+1$ where $j \notin \{b, 0\}$. The neighbors y of the vertex x must satisfy $\|y\|_k^{0i} \geqslant t+1$ for all $i \neq j$, $\|y\|_k^{0j} \geqslant t$, $\|y\|_k^{ij} \geqslant t+1$ for all $i \neq 0$, or $y = [0]^k x_{k+1} \cdots x_n$. By the induction hypothesis and what we have already proved, only two neighbors y and z of x satisfying $\|y\|_k^{00} \geqslant t+1$ and $z = [0]^k x_{k+1} \cdots x_n$ can possibly be in A_{t+1}. However by using a very similar argument, we can show that that $z \notin A_{t+1}$. Therefore, $x \notin A_{t+2}$. In a very similar way, we can prove that $x \notin A_{t+2}$ if x satisfies $\|x\|_k^{i0} \geqslant t+1$ for $i \notin \{a, 0\}$, $\|x\|_k^{aj} \geqslant t+1$ for $j \neq 0$, or $\|x\|_k^{ib} \geqslant t+1$ for $i \neq 0$.

Let us consider a vertex x satisfying $\|x\|_k^{ij} \geqslant t+1$ where $i \notin \{a, 0\}$ and $j \notin \{b, 0\}$. We would like to show that $x \notin A_{t+3}$. The neighbors y of the vertex x must satisfy $\|y\|_k^{ij} \geqslant t$, $\|y\|_k^{mj} \geqslant t+1$ where $m \neq i$, $\|y\|_k^{im} \geqslant t+1$ where $m \neq j$ or $y = [0]^k x_{k+1} \cdots x_n$. From the induction hypothesis and what we have already proved, the neighbor y of x satisfying $y = [0]^k x_{k+1} \cdots x_n$ is the only possible vertex in A_{t+2}. Therefore, $x \notin A_{t+3}$. □

Now let us summarize Lemmas (5.3.2)–(5.3.7) in the following lemma.

Lemma 5.3.8. *For $x, y \in \{0, 1, \cdots, q-1\}^n$ such that dim(x)=k, dim(y)=l, dim($x \vee y$)=m, where $k, l < m \leqslant n$, $d(x, y) = d \leqslant 2$, and $|\{i : x_i = y_i = *\}| = p$, then the infection time of $Q^x \cup Q^y$ in $Q_{n,q}$ is:*

For $d = 0$,

$$T_{Q_{n,q}}(Q^x \cup Q^y) = m - p - 1.$$

For $d = 1$,

$$T_{Q_{n,q}}(Q^x \cup Q^y) = m - p.$$

For $d = 2$,

$$T_{Q_{n,q}}(Q^x \cup Q^y) = m - p + 1.$$

Proof. Note that we have $m = p + (k - p) + (l - p) + d = k + l - p + d$.

Let us consider the case where $d = 0$. By Lemma 5.3.2, we have $T_{Q_{m,q}}(Q^x \cup Q^y) \leqslant k-p+l-p-1 = m - p - 1$. By Lemma 5.3.5, we have $k - p + l - p - 1 = m - p - 1$.

Now let us consider the case where $d = 1$. By Lemma 5.3.3, we have $T_{Q_{m,q}}(Q^x \cup Q^y) \leqslant k - p + l - p + 1 = m - p$. By Lemma 5.3.6, we have $k - p + l - p + 1 = m - p$.

Finally let us consider the case where $d = 2$. By Lemma 5.3.4, we have $T_{Q_{m,q}}(Q^x \cup Q^y) \leqslant k-p+l-p+3 = m - p + 1$. By Lemma 5.3.7, we have $k - p + l - p + 3 = m - p + 1$. □

We are now in a position to prove the main theorem of this section, where we establish a recursive formula for the maximum percolation time, leading to its closed-form expression.

Theorem 5.3.9. *For $q \geqslant 3$, $M_q(0) = 0$ $M_q(1) = 1$, and $M_q(2) = 3$. Moreover, for all $n \geqslant 3$*

$$M_3(n) - M_3(n-3) = 2n - 1,$$

and for $q \geqslant 4$

$$M_q(n) - M_q(n-3) = 2n.$$

Proof. It is obvious that $M_q(0) = 0$ and $M_q(1) = 1$. Now consider the case where $n = 2$. W.l.o.g we can assume $00 \in A_0$ and first consider $A_0 = \{00, ab\}$ where $a, b \neq 0$. It takes 3 steps to infect every vertex in $Q_q(2)$. Another initially infected set that we need to consider is $A_0 = \{00, a0, 0b\}$ where $a, b \neq 0$ and again it takes 3 steps to percolate. These are the only sets of initially infected vertices that we need to consider since adding any other vertex to these sets will not increase the percolation time and the process can not be initiated if any vertex is removed from these sets.

We will first prove the lower bound of $M_q(n)$.

Case 1. Let A^{n-2} be a set that internally spans Q^x_{n-2} for $x = [*]^{n-2}00$ in time $M_q(n-2)$ such that

$[0]^n$ is infected at time $M_q(n-2)$. Now let $A' = A^{n-2} \cup [0]^{n-2}11$ and assume that we initially infect every vertex in A'. Note that $\langle A' \rangle = Q_{n,q}$ and before the time $M_q(n-2)$, none of the vertex in $[0]^{n-2} * *$ has interaction with the infection process in $[*]^{n-2}00$.

Now assume that every vertex in $[*]^{n-2}00$ has been infected. Then it takes exactly $n+1$ additional steps to infect the rest of the vertices in $Q_{n,q}$. Indeed, by Lemma 5.3.8, we have $d([*]^{n-2}00, [0]^{n-2}11) = 2$, $|\{i : x_i = y_i = *\}| = 0$ where $x = [*]^{n-2}00$ and $y = [0]^{n-2}11$, and $\langle [*]^{n-2}00, [0]^{n-2}11) = 2 \rangle = Q_{n,q}$. Therefore,

$$T_{Q_{n,q}}(A') = M_q(n-2) + n + 1.$$

Case 2. Let A^{n-3} be a set that internally spans Q^x_{n-3} for $x = [*]^{n-3}000$ in time $M_q(n-3)$ such that $[0]^n$ is infected at time $M_q(n-3)$. Let $A' = [*]^{n-3}000 \cup [0]^{n-3}110 \cup [2]^n$ and assume we initially infect every vertex in A'. Note that $\langle A' \rangle = Q_{n,q}$ and the set of sites infected after $M_q(n-3)$ steps is $[*]^{n-3}000 \cup [0]^{n-3}110 \cup [2]^n$.

After $M_q(n-3) + n - 1$ steps, we claim that $[2]^n$ is the only vertex infected in $\cup_{j \neq 0}([*]^{n-1}j)$. Indeed, the vertices y in $[*]^{n-1}0$ that are at distance 2 from $[2]^n$ satisfy for $i \neq 2$

$$\|y\|^{i20} = n - 3, \tag{5.9}$$

for $j \neq 2$

$$\|y\|^{2j0} = n - 3, \tag{5.10}$$

$$\|y\|^{220} = n - 4, \tag{5.11}$$

or

$$\|y\|^{220} = n - 3. \tag{5.12}$$

By Lemma 5.3.4 and Lemma 5.3.7, none of those vertices y satisfying (5.9)–(5.12) is infected at time $\leqslant M_q(n-3)+n-2$. In fact some of those vertices y are infected exactly at time $M_q(n-3)+n-1$ and the rest of those vertices will be infected at time $M_q(n-3)+n$. Therefore, at time $M_q(n-3)+n-1$, $[2]^n$ is the only vertex infected in $\cup_{j\neq 0}([*]^{n-1}j)$. Now we assume that every vertex in $[*]^{n-1}0 \cup [2]^n$ has been infected, then by Lemma 5.3.8, it takes exactly n steps to infected the entire $Q_{n,q}$ since $d([*]^{n-1}0, [2]^n) = 1$, $|\{i : x_i = y_i = *\}| = 0$ where $x = [*]^{n-1}0$ and $y = [2]^n$, and $\langle [*]^{n-1}0 \cup [2]^n \rangle = Q_{n,q}$.

Therefore, we have

$$T_{Q_{n,q}}(A') \geqslant M_q(n-3) + 2n - 1.$$

Additionally we would like to show that for $q \geqslant 4$,

$$T_{Q_{n,q}}(A') \geqslant M_q(n-3) + 2n. \tag{5.13}$$

In order to prove this tighter bound we need to further analyze the process right after $M_q(n-3)$ steps.

Now we need to show the following claim for $q \geqslant 4$.

Claim 5.3.10. *For $1 \leqslant k \leqslant n-1$, every k dimensional cube in $[*]^{n-1}2$ will be fully infected at time $M_q(n-3)+n+k$ and at time $\leqslant M_q(n-3)+n+k-1$ none of the k dimensional cubes in $[*]^{n-1}2$ has been fully infected. In particular, the vertices x with $x_n = 2$ satisfying the following condition, denoted as Condition(k),are not infected at time $\leqslant M_q(n-3)+n+k-1$:*

there exists a subset $I \subset [n-1]$ *of size at least* k *such that*

$$x_i = \begin{cases} \notin \{0,2\} & \text{if } i \in I \text{ and } i \notin \{n-2, n-1\}, \\ \notin \{0,1,2\} & \text{if } i \in I \text{ and } i \in \{n-2, n-1\}, \\ \in \{0,2\} & \text{if } i \notin I \text{ and } i \notin \{n-2, n-1\}, \\ \in \{0,1,2\} & \text{if } i \notin I \text{ and } i \in \{n-2, n-1\}. \end{cases}$$

For $l \neq 0, 2$ *and for* $1 \leqslant k \leqslant n-1$*, none of the* k *dimensional cubes in* $[*]^{n-1}l$ *has been fully infected at time* $\leqslant M_q(n-3) + n + k$*. In particular, the vertices* x *with* $x_n = l$ *satisfying Condition*(k) *are not infected at time* $\leqslant M_q(n-3) + n + k$*.*

Proof. Let $k = 1$. Note that by Lemma 5.3.4 and Lemma 5.3.7, after $M_q(n-3) + n - 1$ steps every vertex x in $[*]^{n-1}0$ has been infected except for the vertices x satisfying

$$\|x\|^{ij0} = n - 3,$$

where $i, j \notin \{0, 1\}$ and after one additional step, the cube $[*]^{n-1}0$ has been fully infected.

Again $[2]^n$ is the only vertex infected in $\cup_{j \neq 0}([*]^{n-1}j)$ at time $M_q(n-3) + n - 1$ and after one more step there are no infected vertices in $\cup_{j \notin \{0,2\}}([*]^{n-1}j)$ besides $[2]^n$. The following vertices are infected in $[*]^{n-1}2$ at time $M_q(n-3) + n$:

$$[2]^{n-3}122 \quad [2]^{n-3}022$$

$$[2]^{n-3}202 \quad [2]^{n-3}212,$$

and

$$0[2]^{n-1}$$

$$202^{n-2}$$

$$2202^{n-3}$$

...

$$[2]^{n-4}0[2]^3.$$

Indeed the vertex $[2]^{n-3}122$ has two neighbors infected at time $\leqslant M_q(n-3)+n-1$, namely $[2]^n$ and $[2]^{n-3}120$ and it is similar for other vertices listed above.

At time $M_q(n-3)+n+1$, every 1 dimensional cube in $[*]^{n-1}2$ is fully infected since the vertices in a 1 dimensional cube are either infected at time $M_q(n-3)+n$ or having at least 2 neighbors infected at time $M_q(n-3)+n+1$.

Now consider the vertices in $\cup_{j\notin\{0,2\}}([*]^{n-1}j)$. At time $M_q(n-3)+n+1$, the following vertices are infected:

$$\cup_{j\notin\{0,2\}}[2]^{n-1}j$$

since the vertex $[2]^{n-1}j$ has two neighbors infected at time $\leqslant M(n-3)+n$ namely $[2]^n$ and $[2]^{n-1}0$. Additionally, at time $M_q(n-3)+n+1$, for all $l\notin\{0,2\}$

$$[2]^{n-3}12l \quad [2]^{n-3}02l$$

$$[2]^{n-3}20l \quad [2]^{n-3}21l,$$

and

$$0[2]^{n-2}l$$

$$202^{n-3}l$$

$$2202^{n-4}l$$

...

$$[2]^{n-4}0[2]^2l$$

are infected. It is easy to see that the rest of the vertices in $\cup_{j\notin\{0,2\}}([*]^{n-1}j)$ are not infected at time $M_q(n-3)+n+1$.

Therefore, the statement is true for $k=1$.

Now assume the statement is true for $k \geqslant 1$. Consider a vertex x with $x_n = 2$ satisfying Condition(k+1). By the induction hypothesis this vertex x has only one neighbors at time $M_q(n-3)+n+k-1$, namely, $x_1x_2\cdots x_{n-1}0$.

Indeed, let us choose a particular $i \in [n-1]$ where $i \notin \{n-1, n-2\}$ and $i \notin I$ and w.l.o.g we can assume that $x_i = 0$. Consider a neighbor y of x where $y_j = x_j$ for $j \neq i$ and $y_i = 2$. By the induction hypothesis, the vertex y is not infected at time $M_q(n-3)+n+k-1$. Consider another neighbor z of x where $z_j = x_j$ for $j \neq i$ and $z_i = a$ where $a \notin \{0,2\}$. By the induction hypothesis, z is not infected at time $M_q(n-3)+n-k-1$ since the vertex z satisfies Condition($k+2$). The analysis is very similar for $i \in \{n-1, n-2\}$ and $i \notin I$.

Now we choose a particular $i \in [n-1]$ where $i \notin \{n-1, n-2\}$ and $i \in I$. Consider a neighbor y of x where $y_j = x_j$ for $j \neq i$ and $y_i \notin \{x_i, 0, 2\}$. By the induction hypothesis y is not infected at time $M_q(n-3)+n+k-1$ since the vertex y satisfies Condition($k+1$). Consider another neighbor z of x with $z_j = x_j$ for $j \neq i$ and $z_i = a$ where $a \in \{0,2\}$. By the induction hypothesis the vertex z is not

infected at time $\leqslant M_q(n-3)+n+k-1$ since z satisfies Condition(k). The analysis is very similar for $i \in \{n-1, n-2\}$ and $i \notin I$.

Now consider a neighbor y of x where $y_j = x_j$ for $j \neq n$ and $y_n \notin \{0, 2\}$.By the induction hypothesis, the vertex y is not infected at time $M_q(n-3)+n-k$ since y satisfies Condition(k).

Consider a vertex x with $x_n = l$ satisfying Condition($k+1$). By the induction hypothesis this vertex x has only one infected neighbor at time $M_q(n-3)+n+k$, namely, $x_1x_2\cdots x_{n-1}0$. The analysis is very similar to the case where $x_n = 2$. □

It is easy to see that claim 5.3.10 suffices to show (5.13).

Now let us try to prove the upper bound of $M_q(n)$. Let A be a set of initially infected vertices spanning the hypercube $Q_{n,q}$ by the process.

Let $Q_0 = Q_{i_1}^{x_{i_1}} \subset Q_{i_2}^{x_{i_2}} \subset \cdots \subset Q_{i_t}^{x_{i_t}} = Q_{n,q}$ be a nested sequence of internally spanned subcubes with respect to A and $Q_{m_1}^{z_{m_1}}$,$Q_{m_2}^{z_{m_2}}$,$\cdots$,$Q_{m_{t-2}}^{z_{m_{t-2}}}$,$Q_{m_{t-1}}^{z_{m_{t-1}}}$ be the cubes that marge with cubes $Q_{i_j}^{x_{i_j}}$ as described in Lemma 5.2.7. Assume w.l.o.g both sequences are of maximal lengths.

W.l.o.g, assume that A is minimal (under containment) set spanning $Q_{n,q}$. Note that $i_{t-1} \leqslant n$. We will proceed by analyzing different cases based on the the values of i_{t-1} and i_{t-2}. By Lemma 5.2.7, for all $1 \leqslant j \leqslant t-1$ we have $i_j \geqslant m_j$

Case 1. If $i_{t-1} \leqslant n-2$, then after at most $M_q(i_{t-1}) \leqslant M_q(n-2)$ steps, both $Q_{i_{t-1}}^{x_{i_{t-1}}}$ and $Q_{m_{t-1}}^{z_{m_{t-1}}}$ are fully infected. Since $\langle Q_{i_{t-1}}^{x_{i_{t-1}}} \cup Q_{m_{t-1}}^{z_{m_{t-1}}} \rangle = Q_{n,q}$, after at most additional $n+1$ steps every vertex in $Q_{n,q}$ will be infected by Lemma 5.3.8. Indeed, we have dim($Q_{i_{t-1}}^{x_{i_{t-1}}} \vee Q_{m_{t-1}}^{z_{m_{t-1}}}$)=n and consider the worst case where $d(x,y) = 2$, and $|\{i : x_i = y_i = *\}| = 0$ with $x = Q_{i_{t-1}}^{x_{i_{t-1}}}$ and $y = Q_{m_{t-1}}^{z_{m_{t-1}}}$. Therefore, we have

$$T_{Q_{n,q}}(A) \leqslant M(n-2)+n+1.$$

Case 2. If $i_{t-1} = n-1$ and $i_{t-2} = n-2$, then there is a vertex $v \in A \cap Q_{m_{t-2}}^{z_{m_{t-2}}}$ such that $d(x_{t-2}, v) = 1$

and $\langle Q_{i_{t-2}}^{x_{i_{t-2}}} \cup v\rangle = Q_{i_{t-1}}^{x_{i_{t-1}}}$. Moreover, there exists a vertex $w \in A \cap Q_{m_{t-1}}^{z_{m_{t-1}}}$ such that $\langle Q_{i_{t-1}}^{x_{i_{t-1}}} \cup w\rangle = Q_{n,q}$.

Note that $d(x_{i_{t-2}}, w) = 1$ or $d(x_{i_{t-2}}, w) = 2$. If $(x_{i_{t-2}}, w) = 2$, then $\langle Q_{i_{t-2}}^{x_{i_{t-2}}} \cup w\rangle = Q_{n,q}$, which contradicts the minimality of A since $\langle A\backslash\{v\}\rangle = Q_{n,q}$. Thus $d(x_{i_{t-2}}, w) = 1$.

W.l.o.g we can assume that

$$x_{i_{t-2}} = [*]^{n-2}00,$$

$$x_{i_{t-2}} \vee v = [*]^{n-1}0,$$

and

$$x_{i_{t-2}} \vee w = [*]^{n-2}0*.$$

It is easy to see that after at most $M_q(n-2)$ steps the cube $Q_{i_{t-2}}^{x_{i_{t-2}}}$ is fully infected. After at most $(n-1)$ additional steps both $Q_{n-1}^{x_{i_{t-2}} \vee v}$ and $Q_{n-1}^{x_{i_{t-2}} \vee w}$ are fully infected by Lemma 5.3.8. After 2 more steps every vertex in $Q_{n,q}$ will be infected again by Lemma 5.3.8. Therefore, we have

$$T_{Q_{n,q}}(A) \leqslant M_q(n-2) + n - 1 + 2 = M_q(n-2) + n + 1.$$

Case 3. If $i_{t-1} = n-1$, $i_{t-2} \leqslant n-3$, and $d(x_{i_{t-2}}, z_{m_{t-2}}) \leqslant 1$, then after $M(n-3)$ steps both $Q_{i_{t-2}}^{x_{i_{t-2}}}$ and $Q_{m_{t-2}}^{z_{m_{t-2}}}$ are fully infected. Since $\langle Q_{i_{t-2}}^{x_{i_{t-2}}} \cup Q_{m_{t-2}}^{z_{m_{t-2}}}\rangle = Q_{i_{t-1}}^{x_{i_{t-1}}}$, after at most i_{t-1} more steps the cube $Q_{i_{t-1}}^{x_{i_{t-1}}}$ is fully infected by Lemma 5.3.8.

Since $i_{t-1} = n-1$, $d(x_{i_{t-1}}, Z_{m_{t-1}}) \leqslant 1$. Again by Lemma 5.3.8 after at most n more steps every vertex in $Q_{n,q}$ will be infected. Therefore, we have

$$T_{Q_{n,q}}(A) \leqslant M_q(n-3) + n - 1 + n = M_q(n-3) + 2n - 1.$$

Case 4. If $i_{t-1} = n-1$,$i_{t-2} \leqslant n-3$, $d(x_{i_{t-2}}, z_{m_{t-2}}) = 2$, and $m_{t-2} = n-3$, then after at most

$M(n-3)$ steps both $Q_{i_{t-2}}^{x_{i_{t-2}}}$ and $Q_{m_{t-2}}^{z_{m_{t-2}}}$ are fully infected. Note that $\langle Q_{i_{t-2}}^{x_{i_{t-2}}} \cup Q_{m_{t-2}}^{z_{m_{t-2}}} \rangle = Q_{i_{t-1}}^{x_{i_{t-1}}}$ and since $i_{t-2} \geqslant m_{t-2}$ then $i_{t-2} = n-3$. After at most 3 additional steps the cube $Q_{i_{t-1}}^{x_{i_{t-1}}}$ is infected. Indeed observe that $|\{i : x_i = y_i = *\}| = n-3$ where $x = Q_{i_{t-2}}^{x_{i_{t-2}}}$ and $y = Q_{i_{t-2}}^{x_{i_{t-2}}}$ since one of the coordinates must be the same for x and y. since $d(x,y) = 2$ and $\dim(x \vee y) = n-1$, $T_{Q_{n,q}}(Q_{i_{t-2}}^{x_{i_{t-2}}} \cup Q_{m_{t-2}}^{z_{m_{t-2}}}) = 3$ by Lemma 5.3.8.

Since $i_{t-1} = n-1, d(x_{i_{t-1}}, z_{m_{t-1}}) \leqslant 1$. By lemma 5.3.8 after at most n more steps every vertex on $Q_{n,q}$ will be infected. Therefore, we have

$$T_{Q_{n,q}}(A) \leqslant M_q(n-3) + 3 + n.$$

Additionally we would like to show

$$T_{Q_{n,3}}(A) \leqslant M_3(n-3) + 2 + n.$$

W.l.o.g, we can assume that

$$x_{i_{t-1}} = [*]^{n-1}0,$$

$$x_{i_{t-2}} = [*]^{n-3}000,$$

and

$$z_{m_{t-2}} = [*]^{n-3}ij0,$$

where $i \neq 0$ and $j \neq 0$.

Again after $M_3(n-3)$ steps, both $Q_{i_{t-2}}^{x_{i_{t-2}}}$ and $Q_{m_{t-2}}^{z_{m_{t-2}}}$ are fully infected. After at most 2 additional steps every vertex in $Q_{i_{t-1}}^{x_{i_{t-1}}}$ satisfying $[*]^{n-3}kl0$ for all $k, l \in \{0,1,2\}$ but $k \notin \{0,i\}$ and $l \notin \{0,j\}$.

There is a vertex $y \in A$ such that $A \in [*]^{n-1}l$ where $l \neq 0$. Otherwise the vertices not in the cube

$Q_{i_{t-1}}^{x_{i_{t-1}}}$ will not be infected. Let $y = y_1 y_2 \cdots y_{n-1} l$.

Therefore, after $M_3(n-3)+2$ steps, the vertices v and w in $Q_{i_{t-1}}^{x_{i_{t-1}}}$ are infected, where

$$v = y_1 y_2 \cdots y_{n-3}\ i\ y_{n-1} 0$$

and

$$w = y_1 y_2 \cdots y_{n-3}\ 0\ y_{n-1} 0.$$

Assume $y_{n-2} \notin \{0, i\}$. Then after $M_3(n-3)+n$, vertices v^* and w^* in $[*]^{n-1} l$ are infected, where

$$v^* = y_i y_2 \cdots y_{n-3}\ i\ y_{n-1} l$$

and

$$w^* = y_i y_2 \cdots y_{n-3}\ 0\ y_{n-1} l.$$

Therefore, after $M_3(n-3)+3$, $[*]^{n-1}0$ and $y_1 y_2 \cdots y_{n-3} * y_{n-1} l$ have been infected. By Lemma 5.3.8 after at most $n-1$ steps every vertex in $Q_{n,3}$ will be infected.

Now assume $y_{n-2} = 0$. If $y_{n-1} \notin \{0, j\}$, similarly it can be shown that after $M_3(n-3)+n+2$ every vertex in $Q_{n,3}$ will be infected. Thus we can assume $y_{n-1} \in \{0, j\}$. First let $y_{n-1} = 0$. Note that after $M_3(n-3)+n-1$ steps, the vertices u and z in $Q_{i_{t-1}}^{x_{i_{t-1}}}$ are infected where

$$u = y_1 y_2 \cdots y_{n-3} k 00$$

with $k \notin \{i, 0\}$ and

$$z = y_1 y_2 \cdots y_{n-3} i 00.$$

Then after one more step, the vertices u^* and z^* in $[*]^{n-1}l$ are infected, where

$$u^* = y_1 y_2 \cdots y_{n-3} k 0 l$$

and

$$z^* = y_1 y_2 \cdots y_{n-3} i 0 l.$$

Therefore, after $M_3(n-3)+3$ steps, $[*]^{n-1}0$ and $y_1 y_2 \cdots y_{n-3} * 0l$ have been infected. By Lemma 5.3.8 after at most $n-1$ steps every vertex in $Q_{n,3}$ will be infected.

Due to symmetry for other values of y_{n-2} and y_{n-1} the argument is the same.

Therefore, we have

$$T_{Q_{n,3}} \leqslant M_3(n-3) + n + 2.$$

Case 5. If $i_{t-1} = n-1$,$i_{t-2} \leqslant n-3$, $d(x_{i_{t-2}}, z_{m_{t-2}}) = 2$, and $m_{t-2} < n-3$, after at most $M_q(n-3)$ steps, both $Q_{i_{t-2}}^{x_{i_{t-2}}}$ and $Q_{m_{t-2}}^{z_{m_{t-2}}}$ are fully infected. After at most n more steps the cube $Q_{i_{t-1}}^{x_{i_{t-1}}}$ is fully infected by Lemma 5.3.8. Then after at most n more steps every vertex on $Q_{n,q}$ is infected again by Lemma 5.3.8 since $d(x_{i_{t-2}}, z_{m_{t-2}}) \leqslant 1$ and $\langle Q_{i_{t-1}}^{x_{i_{t-1}}} \cup Q_{m_{t-1}}^{z_{m_{t-1}}} \rangle = Q_{n,q}$. Therefore, we have for $q \geqslant 3$

$$T_{Q_{n,q}}(A) \leqslant M_q(n-3) + n + n = M_q(n-3) + 2n.$$

Additionally we would like to show that

$$T_{Q_{n,3}}(A) \leqslant M_3(n-3) + 2n - 1.$$

In order to show this tighter bound we need to analyze the process after $M_3(n-3)$ steps in a more careful way. Assume that $q = 3$.

W.l.o.g we can assume that

$$x_{i_{t-1}} = [*]^{n-1}0,$$

$$x_{i_{t-2}} = [*]^{i_{t-2}}[0]^{n-i_{t-2}},$$

and

$$z_{m_{t-2}} = [*]^{p}[0]^{i_{t-2}-p}[*]^{m_{t-2}-p}ij0,$$

where $i \neq 0$ and $j \neq 0$.

Again after at most $M_3(n-3)$ steps both $Q_{i_{t-2}}^{x_{i_{t-2}}}$ and $Q_{m_{t-2}}^{z_{m_{t-2}}}$ are fully infected. After at most $n-1$ additional steps, by Lemma 5.3.4 and Lemma 5.3.7, vertices x in $Q_{i_{t-1}}^{x_{i_{t-1}}}$ satisfying $1 \leqslant \|x\|^{kl0} \leqslant n-3$ for all $k, l \in \{0,1,2\}$ but $k \notin \{0,i\}$ and $l \notin \{0,j\}$ are infected.

There is a vertex $y \in A$ such that $A \in [*]^{n-1}l$ where $l \neq 0$. Otherwise the vertices not in the cube $Q_{i_{t-1}}^{x_{i_{t-1}}}$ will not be infected. Let $y = y_1y_2\cdots y_{n-1}l$.

Assume that $\sum_{i=1}^{n-3}\mathbb{I}_{\{y_i\neq 0\}} \geqslant 1$. Therefore, after $M_3(n-3)+n-1$ steps, the vertices v and w in $Q_{i_{t-1}}^{x_{i_{t-1}}}$ are infected, where

$$v = y_1y_2\cdots y_{n-3}\ i\ y_{n-1}0$$

and

$$w = y_1y_2\cdots y_{n-3}\ 0\ y_{n-1}0.$$

Assume $y_{n-2} \notin \{0,i\}$. Then after $M_3(n-3)+n$, vertices v^* and w^* in $[*]^{n-1}l$ are infected, where

$$v^* = y_iy_2\cdots y_{n-3}\ i\ y_{n-1}l$$

and

$$w^* = y_iy_2\cdots y_{n-3}\ 0\ y_{n-1}l.$$

Therefore, after $M_3(n-3)+n$ steps, $[*]^{n-1}0$ and $y_1y_2\cdots y_{n-3}*y_{n-1}l$ have been infected. By Lemma 5.3.8 after at most $n-1$ steps every vertex in $Q_{n,3}$ will be infected.

Now assume $y_{n-2}=0$. If $y_{n-1}\notin\{0,j\}$, similarly it can be shown that after $M_3(n-3)+2n-1$ every vertex in $Q_{n,3}$ will be infected. Thus we can assume $y_{n-1}\in\{0,j\}$. First let $y_{n-1}=0$. Note that after $M_3(n-3)+n-1$ steps, the vertices u and z in $Q_{i_{t-1}}^{x_{i_{t-1}}}$ are infected where

$$u=y_1y_2\cdots y_{n-3}k00$$

with $k\notin\{i,0\}$ and

$$z=y_1y_2\cdots y_{n-3}i00.$$

Then after one more step, the vertices u^* and z^* in $[*]^{n-1}l$ are infected, where

$$u^*=y_1y_2\cdots y_{n-3}k0l$$

and

$$z^*=y_1y_2\cdots y_{n-3}i0l.$$

Therefore, after $M_3(n-3)+n$ steps, $[*]^{n-1}0$ and $y_1y_2\cdots y_{n-3}*0l$ have been infected. By Lemma 5.3.8 after at most $n-1$ steps every vertex in $Q_{n,3}$ will be infected.

Due to symmetry for other values of y_{n-2} and y_{n-1} the argument is the same.

Therefore, we have

$$T_{Q_{n,3}}\leqslant M_3(n-3)+2n-1.$$

Assume $y=[0]^{n-3}y_{n-2}y_{n-1}l$. Therefore, after $M_3(n-3)+n-1$ steps, as long as $n-1\geqslant 2$, the

vertices v and w in $Q_{i_{t-1}}^{x_{i_{t-1}}}$ are infected, where

$$v = [0]^{n-3}\ i\ y_{n-1}0$$

and

$$w = [0]^{n-3}\ 0\ y_{n-1}0.$$

Indeed, the vertex $[0]^n$ and $[0]^{n-3}ij0$ are infected at time $\leqslant M_3(n-3)$. From now on the proof is identical to the case where $\sum_{i=1}^{n-3} \mathbb{I}_{\{y_i \neq 0\}} \geqslant 1$.

Finally we have

$$T_{Q_{n,3}} \leqslant M_3(n-3) + 2n - 1.$$

We need one more claim to conclude the proof.

Claim 5.3.11. *If* $M_q(0) = 0, M_q(1) = 1$ *and* $M_q(2) = 3$ *and for* $n \geqslant 3$

$$M_3(n) = \max\{M_3(n-2) + n + 1, M_3(n-3) + 2n - 1\}$$

and for $q \geqslant 4$

$$M_q = \max\{M_q(n-2) + n + 1, M_q(n-3) + 2n\},$$

then

$$M_3(n) = M_3(n-3) + 2n - 1$$

and for $q \geqslant 4$

$$M_q(n) = M_q(n-3) + 2n.$$

Proof. Assume that $q = 3$. We have $M_3(3) = M_3(0) + 5 = 5$ or $M_3(3) = M_3(1) + 4 = 5$

$M_3(4) = M_3(1) + 7 = 8$ or $M_3(4) = M_3(2) + 5 = 8$.

$M_3(5) = M_3(2) + 9 = 12$ or $M_3(5) = M_3(3) + 6 = 11$.

Thus the lemma holds for $1 \leqslant n \leqslant 3$.

For $n \geqslant 3$, we assume that the lemma holds for $n, n+1, n+2$. We have

$$M_3(n-3) = \max\{M_3(n) + 2(n+3) - 1, M_3(n+1) + n + 4\} \tag{5.14}$$

$$= \max\{M_3(n-3) + 2n - 1 + 2(n+3) - 1, M_3(n-2) + 2(n+1) - 3 + n + 4\} \tag{5.15}$$

$$= \max M_3(n-3) + 4n + 4, M_3(n-2) + 3n + 3 \tag{5.16}$$

$$= M_3(n-3) + 4n + 4 \tag{5.17}$$

$$= M_3(n) - 2n + 1 + 4n + 4 \tag{5.18}$$

$$= M_3(n) + 2n + 5, \tag{5.19}$$

where (5.17) follows from

$$\begin{aligned} M_3(n-3) + 4n + 4 &= M_3(n-3) - 2n + 1 + 2n + 5 \\ &\geqslant M_3(n-2) + n + 1 + 2n + 5 \\ &> M_3(n-2) + 3n + 3. \end{aligned}$$

For $q \geqslant 4$, the calculation is almost the same. □

Therefore, we have our desired results. □

Now we are ready to prove Theorem 5.1.1.

Proof. Since $M_3(n) = M_3(n-3) + 2n - 1$, we have $M_3(n) = \frac{n^2}{3} + \frac{2n}{3} + C$ where C is an undetermined real constant. Since $M_3(0) = 0$, we have $C = 0$ and thus $M_3(n) = \frac{n^2}{3} + \frac{2n}{3}$ for $n = 3, 6, 9, \cdots$.

Other expressions in Theorem 5.1.1 can be derived in the same way. □

5.4 OPEN PROBLEMS

We conclude with several open problems.

1. What is the maximum percolation time on the n-dimensional q-ary hypercube when the infection threshold $r \geqslant 3$?

   The value $M(Q_{n,2}, r \geqslant 3) = \frac{2^n}{n}(\log n)^{-O(1)}$ was obtained in [35] by discovering a link between this problem and a generalization of the Snake-in-the-Box problem. Improving upon this result to achieve an asymptotically tight, up to a constant, is an interesting question. Finding an exact expression for $M(Q_{n,q}, r \geqslant 3)$ remains as a challenging open problem.

2. What is the maximum percolation time, denoted by $M([n]^d, r)$, on the $[n]^d$ grid when the infection threshold $2 \leqslant r \leqslant 2d$?

   The value $M([n]^2, 2) = \frac{3n^2}{8}$ was established in [16]. However, for the other values of d and r, the behavior of $M([n]^d, r)$ has not been explored.

Chapter 6

THE TIME OF BOOTSTRAP PERCOLATION IN HIGH DIMENSIONS

6.1 INTRODUCTION

While significant efforts have been dedicated to studying the critical probability of the process, estimating percolation time, denoted by $T = \min\{t : A_t = V(G)\}$, has also garnered attention. In the deterministic setting, papers [43], [35], and [16] estimated the maximum percolation time for the process on the n-dimensional hypercube and the $[n]^2$ grid. In the probabilistic setting, papers [19], [5], and [18] estimated the distribution of percolation time on a $[n]^d$ grid.

In particular, from [5] we have the following theorem.

Theorem 6.1.1. *Let $0 < p = p(n) < 1$ be such that $\liminf p \log\log n > 2\lambda$ and $1 - p = n^{-o(1)}$ (that is $\log 1/(1-p) \ll \log n$). Let T denote the percolation time of a p-random subset of $[n]^2$ under the 2-neighbor bootstrap percolation process and $\lambda = \frac{\pi^2}{18}$. Then we have*

$$T = \frac{(1 + o(1)) \log n}{2 \log(1/(1-p))},$$

with high probability as $n \to \infty$.

In this chapter we investigate the distribution of the time of bootstrap percolation on the d-dimensional

discrete torus[1] $\mathbb{T}_n^d$. As our main result we extend Theorem 6.1.1 in [5] from the two-dimensional case to d-dimensions.

We highlight the primary challenges in extending Theorem 6.1.1 to higher dimensions, as well as our main contributions. In two dimensions [5], the key intuition is that the event in which a vertex x remains uninfected at time t is "equivalent to" the existence of an empty line segment of length roughly t "near" x. In the higher dimensions, it is not immediately clear whether this key intuition still holds. However, we show that it remains valid in higher dimensions, provided that the infection threshold equals the dimension.

In the two-dimensional case, the idea of analyzing the intersection between a path of uninfected vertices and squares of a certain size was effective in reducing the combinatorial factor from counting the number of such paths. In higher dimensions, however, it is not obvious what geometric objects should be used to achieve a similar reduction. We find that replacing two-dimensional squares with their higher-dimensional analogs (cubes) is a viable approach.

Another challenge lies in generalizing the definition of the interior of a square from two dimensions to higher dimensions in a way that serves our purposes. In the two-dimensional setting, the notion of a square's interior was introduced to properly define certain events, and a similar concept is required in higher dimensions. We address these challenges and provide the necessary definitions to extend the analysis.

Our main result is given by the following theorem.

Theorem 6.1.2. *Consider a graph G as $G = \mathbb{T}_n^d$ and let T be the percolation time. Assume that every vertex on the graph G is initially infected with probability $p(n)$ independent of any other vertex and the infection threshold $r = d$. Assume that $p(n) \geqslant \frac{C}{\log^{(d)}(n)}$ and $1 - p = n^{-o(1)}$, where $C > 0$ is sufficiently large and $\log^{(r)}(.)$ denotes iterating the logarithm r times. Then we have*

$$T = \Theta\left(\log_{\frac{1}{1-p}} n\right),$$

[1] A graph with vertex set $V = \{0, \ldots, n-1\}^d$ and $x, y \in V$ are connected if and only if for some $i \in \{1, \ldots, d\}$, $x_i = y_i \pm 1 \pmod{n}$ and $x_j = y_j$ for all $j \neq i$.

with high probability as $n \to \infty$.

Let us explain our strategy in the proof of the lower bound of the percolation time. For the two-dimensional case, a natural example of an event that would prevent percolation from happening by time t is the existence of an initially uninfected $[2t+1] \times [2]$ rectangle. For the d-dimensional case, an event that there exists an initially uninfected $[2t+1] \times [2]^{d-1}$ rectangle initially would prevent percolation from happening by time t by [19]. One can easily show that the largest t for which such a rectangle is likely to exist is about $\log_{\frac{1}{1-p}} n$. This observation essentially proves the lower bound of Theorem 6.1.2.

We briefly outline the key ideas behind the proof of the upper bound in Theorem 6.1.2. The strategy is to show that if a vertex x remains uninfected at time t, then this deterministically implies the existence of a "path" of nearby $[L]^d$ cubes, each of which is not internally spanned. By choosing L appropriately (and hence controlling the length of the path), we can show that the probability of this event is small.

However, this summary is so brief and somewhat misleading. In fact, requiring each $[L]^d$ cube to be internally spanned is so strong that it is not necessary for bounding the probability of interest and, moreover, the probability of this stronger event differs significantly from the probability of interest. As a result, we must consider a weakened version of the condition, which introduces additional technical complexity. A more detailed sketch of the proof will be provided in Section 6.2.

Below by $C(d)$ we denote a constant which may depend on d. We will use $C(d)$ as generic notation, where the specific form of this function may be different in different expressions.

We write $[A_0] = \cup_{t=0}^{\infty} A_t$ and call $[A]$ the span of A. We say A percolates on G if $[A] = V(G)$ and a set $X \subset [n]^d$ is internally spanned if $X \subset [A \cup X]$.

A set of vertices is defined as empty if all its vertices are initially uninfected and call it non-empty otherwise.

6.2 SKETCH OF THE PROOF FOR THE UPPER BOUND

The upper bound in Theorem 6.1.2 is roughly saying that the initially uninfected $[2t+1] \times [2]^{d-1}$ rectangles are the only obstacles to percolation by time t.

Suppose a site x in $\mathbb{T}_n^d$ is uninfected at time t. It is easy to see that at least $d+1$ neighbors of x must be uninfected at time $t-1$. Thus at least one of its neighbors among $(x_1+1, x_2, \cdots, x_d)$, $(x_1, x_2+1, \cdots, x_d)$, $\cdots$, $(x_1, x_2, \cdots, x_d+1)$ is uninfected at time $t-1$ and say it is y. Then at least one of y's neighbors among $(y_1+1, y_2, \cdots, y_d)$, $(y_1, y_2+1, \cdots, y_d)$, $\cdots$, $(y_1, y_2, \cdots, y_d+1)$ is uninfected at time $t-2$. In fact, it is easy to see that there must exist a sequence of t initially uninfected vertices, starting with x, and continuing on the direction parallel to the standard vector $e_1, e_2, \cdots$, or e_d each time.

We would like to show that by far the most probable way for this to occur is for this path to be aligned to form a line segment starting at x, or more specifically, we would like to show that the probability that the uninfected paths exist is not much more than $(1-p)^t$, which is the probability that a given line segment of length t is initially uninfected.

One possible attempt of a proof might go as follows. Assume x is uninfected at time t and thus there is a path of uninfected vertices starting at x and continuing on the direction parallel to $e_1, e_2, \cdots, e_d$ each time. This path, denoted by P, of uninfected vertices may not be a line segment and then this would cause a big combinatorial factors about the choice of such path.

However, it is possible to modify this idea. Following the ideas from [5], rather than counting paths along directions parallel to the standard vectors $e_1, e_2, \cdots, e_d$ of vertices individually, we look at the intersection of these paths with cubes $[L]^d$ and count these. First we allow an initial time $t' = o(t)$. By this time we expect nearly all internally spanned cubes $[L]^d$ to have been infected. Now consider just the first $t - t'$ vertices in the path P: at time t' they are still uninfected, and they intersect a path of cubes $[L]^d$ all of which are not internally spanned; we call such squares bad. Now we have an optimization question: how large

should L be to minimize the probability of this event that there is a path of bad cubes $[L]^d$? In order to have any hope, the probability that a cube $[L]^d$ is bad should be at most $(1-p)^{(1+c)L}$, for some $c > 0$. This is because we would like to show that the probability there exists a path of bad cubes $[L]^d$ is about $(1-p)^t$, so we need the additional c to overcome the combinatorial factor coming from taking a union bound over all paths. Thus, L must be large enough for the probability that a cube $[L]^d$ is bad to be small. Another reason L should be large is to minimize the combinatorial factor. As L increases, there are fewer paths of cubes $[L]^d$ inside a cube $[t-t']^d$, so the combinatorial factor decreases. On the other hand, L cannot be too large, because the error time $t' = L^d$ must be $o(t)$.

This second attempt of a proof is also not quite right: the probability that a cube $[L]^d$ is bad, as we have defined it, is at least $(1-p)^L$ because if any of the $2^{d-1}d$ "sides" of the cube is empty then the cube cannot be internally spanned. On the other hand we have said that the probability needs to be at most $(1-p)^{(1+c)L}$, so our definition of bad cannot be the right one. The way around this is as follows. Let us define a cube $[L]^d$ is bad if it is not internally spanned except for its "sides", which will be defined precisely in Section 6.4 . It turns out this definition is the right one to use for our purpose.

6.3 LOWER BOUND

To prove the lower bound for the percolation time T we will use a theorem by Bollobás, Smith and Uzzell [19].

First let us observe that the states of vertices at l_1 distance greater than t from the vertex x cannot affect whether or not the vertex x is infected at time t. Therefore we can restrict ourselves to the l_1-ball

$$B_t(x) = \{y \in [n]^d : |x-y|_{l_1} \leqslant t\}.$$

The following theorem from [19] will be used in the proof.

Theorem 6.3.1. *Let*

$$ex_{d,r}(t) := \min\{|B_t(0)\backslash A_0| : 0 \notin A_t\}.$$

and

$$P_{d,r}(t) := \{x \in B_t(0) : x_{d-r+2}, \ldots, x_d \in \{0,1\}\}.$$

where $x = (x_1, x_2, \ldots, x_d)$.

Then for every $2 \leqslant r \leqslant d$

$$ex_{d,r}(t) = |P_{d,r}(t)|.$$

Using this result, let us prove the lower bound for the percolation time. Assume that $d = r$.

Proof. Let E_1 be the event that every vertex in $\{x \in B_t(0) : x_2, ..., x_d \in \{0,1\}\}$ is initially uninfected and E_2 be the event that the vertex $(0, 0, ..., 0)$ is uninfected at time t.

From Theorem 6.3.1, the event E_1 implies the event E_2. Therefore, the event E_1 implies the event $\{T > t\}$.

Let E_3 be the event that every rectangle $[2t+1] \times [2]^{d-1}$ is nonempty. It is easy to see that the event $\{T \leqslant t\}$ implies the event E_3. Let us divide $[n]^d$ into $\frac{n^d}{2^d t}$ disjoint rectangles $[2]^{d-1} \times [2t+1]$. Let E_4 be an event that every such disjoint rectangle $[2]^{d-1} \times [2t+1]$ is nonempty. It is easy to see that the event E_3 implies the event E_4.

Therefore, we have

$$\begin{aligned}
\mathbb{P}(T \leqslant t) &\leqslant \left(1 - (1-p)^{2^d t}\right)^{\frac{n^d}{2^d t}} \\
&\leqslant \exp\left(-(1-p)^{2^d t} \frac{n^d}{2^d t}\right) \\
&= \exp\left(-\exp\left(d\log(n) - \log(2^d t) - 2^d t \log\left(\frac{1}{1-p}\right)\right)\right) \\
&= o(1),
\end{aligned}$$

if

$$\frac{2^d t \log(\frac{1}{1-p})}{d \log(n)} \leqslant 1.$$

Thus, with high probability,

$$T \geqslant C(d) \frac{\log(n)}{\log(\frac{1}{1-p})}. \quad \square$$

6.4 UPPER BOUND

Before proceeding, we introduce some elements of notation and definitions.

Define a subcube

$$[(a_1, b_1), (a_2, b_2), ..., (a_d, b_d)] := \{(x_1, x_2, ..., x_d) \in [n]^d : a_i \leqslant x_i \leqslant b_i \ \forall i \in [d]\}.$$

Define a sub-grid

$$[(a_1, b_1), (a_2, b_2), ..., (a_{j-1}, b_{j-1}), (a_j), (a_{j+1}, b_{j+1}), ..., (a_d, b_d)] :=$$

$$\{(x_1, x_2, ..., x_d) \in [n]^d : x_j = a_j, a_i \leqslant x_i \leqslant b_i \ \forall i \in [d] \backslash \{j\}\}.$$

Define a line segment

$$[(a_1), (a_2), \cdots, (a_{j-1}), (a_j, b_j), (a_{j+1})..., (a_d)] := \{(x_1, x_2, ..., x_d) \in [n]^d :$$

$$x_i = a_i, a_j \leqslant x_j \leqslant b_j \ \forall i \in [d] \backslash \{j\}\}.$$

Define a side of the cube $[m]^d$ as of the form $[(a_1), (a_2), ..., (a_{j-1}), (1, m), (a_{j+1}), ..., (a_d)]$ with $a_i \in \{1, m\}$ for all $i \in \{[d] \backslash j\}$ for some $j \in [d]$. Note that there are in total $2^{d-1} d$ sides of $[m]^d$.

Let $\mathfrak{S}_d$ be the symmetric group on d coordinates. Then define:

$$\text{Perm}_d(x) := \{\sigma x : \sigma \in \mathfrak{S}_d\},$$

where for $x = [(x_1), \ldots, (x_d)]$, we define:

$$\sigma x := [\big(x_{\sigma(1)}\big), \cdots, \big(x_{\sigma(d)}\big)].$$

Note that by abuse of the notation if x_i is an interval for some $i \in [d]$, regarding the permutation it is treated as a single value.

We will give the definition of interior($[m]^d$) such that it serves our purpose.

Define the interior of $[m]^d$ as

$$\text{int}([m]^d) := [m]^d \backslash \{\text{Perm}_d[((a_1), (a_2), ..., (a_{j-1}), (1, m), (a_{j+1}), ..., (a_d))] : a_i \in \{1, m\} \text{ for } i \in [d]\backslash\{j\}\},$$

i.e., int($[m]^d$) includes every vertex in $[m]^d$ except for its sides.

Let $A \sim \text{Bin}([m]^d, p)$ be the initial set. Define the cube $D = [m]^d$ to be good if its span contains its interior. Formally, the cube D is good if $\text{int}(D) \subset [D \cap A]$. The cube D is strongly good if it is internally spanned, i.e., $D \subset [D \cap A]$. At last, D is semi-good if it is good but not strongly good, and bad if it is not good.

We denote by $\eta_{m,r}$ the probability that $[m]^d$ is bad with the infection threshold r and in order to simplify the notation, sometimes we may drop the dependence on r if it is clear on the context.

We need the following theorem by Bollobás, Balogh, Duminil-Copin, and Morris [9].

Theorem 6.4.1. *For every $d \geqslant r \geqslant 2$, there exists an explicit constant $\lambda(d, r) > 0$ such that the critical*

probability for $[n]^d$ with infection threshold r,denoted by $p_c([n]^d, r)$, can be expressed as follows:

$$p_c([n]^d, r) = \left[\frac{\lambda(d,r) + o(1)}{\log^{(r-1)} n}\right]^{d-r+1}.$$

6.4.1 $d = r = 3$

Before stating the next lemma let us state one fact which will be used very often in this section.

Fact 6.4.2. *Assume that $\mathbb{P}(B) \leqslant a$, $\mathbb{P}(C) \leqslant b$, $\mathbb{P}(A \cap B) = 0$, and $\mathbb{P}(A \cap C) = 0$. Moreover, assume that the event B and the event C are independent. Then*

$$\mathbb{P}(A) \leqslant 2ab.$$

Proof.

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A \cap B \cap C) + \mathbb{P}(A \cap (B \cap C)^c) \\
&= \mathbb{P}(A \cap (B \cap C)^c) \\
&\leqslant \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap C^c) \\
&\leqslant \mathbb{P}(A \cap B^c \cap C) + \mathbb{P}(A \cap B^c \cap C^c) + \mathbb{P}(A \cap C^c \cap B) + \mathbb{P}(A \cap C^c \cap B^c) \\
&\leqslant 2\mathbb{P}(B^c \cap C^c) \\
&= 2ab.
\end{aligned}$$

□

When $d = r = 3$, we use Lemma 6.4.3 to derive a recursive relation on the probability that $[m]^3$ is bad.

Lemma 6.4.3. *Let $\eta_{m,3}$ be the probability that $[m]^3$ is bad with the infection threshold 3. We have*

$$\eta_{2m,3} \leqslant C(d)\eta_{m,3}^3 + C(d)m^2(1-p)^{4m-8}.$$

The intuition behind the proof of Lemma 6.4.3 is as follows. We begin by partitioning the cube $[2m]^3$ into eight disjoint subcubes and break the proof into several subcases according to the number of bad subcubes in this 8-tuple. Specifically, we analyze the cases where exactly one or two of the eight subcubes are bad as well as the case where none are bad. The main technical part of the proof is devoted to the analysis to these three subcases, where we study the interaction arising among the 8 subcubes. Further the probability of having at least three bad subcubes can be estimated by a very rough upper bound of $C\eta_m^3$, as established in Lemma 6.4.3.

Before proving this lemma we need to introduce notation and prove several auxiliary lemmas.

Let us partition the cube $[2m]^3$ into 8 disjoint subcubes as shown in Figure 6.1.

$$C_1 = [(1,m),(1,m),(1,m)], C_2 = [(m+1,2m),(1,m),(1,m)],$$

$$C_3 = [(1,m),(m+1,2m),(1,m)], C_4 = [(m+1,2m),(m+1,2m),(1,m)],$$

$$C_5 = [(1,m),(1,m),(m+1,2m)], C_6 = [(m+1,2m),(1,m),(m+1,2m)],$$

$$C_7 = [(1,m),(m+1,2m),(m+1,2m)], C_8 = [(m+1,2m),(m+1,2m),(m+1,2m)].$$

We will first define a few events.

- $E := \{[2m]^3$ is good $\}$.
- A:=$\{$all cubes $C_1, C_2, \cdots, C_8$ are good$\}$.
- A_i:=$\{$exactly i subcubes among C_1,C_2,...,C_8 are bad$\}$ for $i \in \{1, \ldots, 8\}$.

Note that

$$\mathbb{P}(E^c) = \mathbb{P}(E^c \cap A) + \mathbb{P}(E^c \cap A^c). \tag{6.1}$$

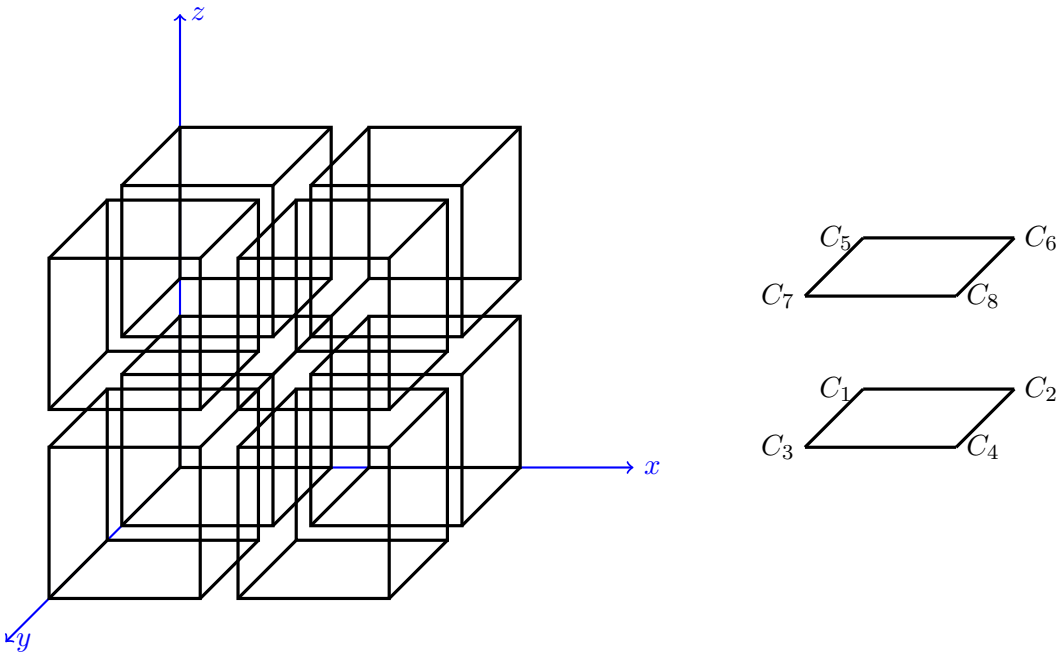

Figure 6.1: Subcubes $C_1, C_2, \cdots, C_8$

Let us consider the term $\mathbb{P}(E^c \cap A)$ in (6.1) and this is the case where all 8 subcubes $C_1, \ldots, C_8$ are good. Since the sides of these 8 subcubes may still remain uninfected the event A does not necessarily imply that the cube $[2m]^d$ is good. Therefore, we have to analyze the interaction between the events that some vertices in the sides of $C_1, \ldots, C_8$ are initially infected and the event A. Moreover, in order for our method to work the probability that the complement of the events that some vertices in the edges of $C_1, \ldots, C_8$ are initially infected needs to be small and in fact it needs to be $\leqslant (1-p)^{4m}$. Let us define these events precisely.

Let $B(m, (1, 2m), a_3)$ for $a_3 \in \{1, m+1, 2m\}$ denote the event that at least one of the following 2 line segments is non-empty:

$$[(m), (1, 2m), (a_3)], \text{or} \quad [(m+1), (1, 2m), (a_3)].$$

Note that the above 2 line segments belong to the sides of $C_5, \ldots, C_8$ if $a_3 = 2m$, and belong to the sides of $C_1, \ldots, C_4$ if $a_3 = 1$.

$B(\pi(m), \pi((1, 2m)), \pi(a_3))$ are defined similarly for all π where π is an ordering of the set $\{m, (1, 2m), a_3\}$.

Let $B(m, m, (c_1, c_2))$ denote the event that at least one of the following 4 line segments is nonempty:

$$\{[(m), (m), (c_1, c_2)], [(m), (m+1), (c_1, c_2)],$$

$$[(m+1), (m), (c_1, c_2)], [(m+1), (m+1), (c_1, c_2)]\}.$$

$B(\pi(m), \pi(m), \pi((c_1, c_2)))$ are defined similarly for all π where π is an ordering of the set $\{m, (m), (c_1, c_2)\}$.

Let us define an event B such that if the events A and B occur then the event E occurs and $\mathbb{P}(B^c) \leqslant C(1-p)^{4m}$ with $C > 0$.

Let

$$B_1 = \cap_{a_3 \in \{2m, m+1, m, 1\}} B(m, (1, 2m), a_3),$$

$$B_2 = \cap_{a_3 \in \{2m, m+1, m, 1\}} B((1, 2m), m, a_3),$$

$$B_3 = B(m, m, (m+1, 2m) \cap B(m, m, (1, m)),$$

$$B_4 = B(m, 1, (1, 2m)) \cap B((1, 2m), 1, m),$$

$$B_5 = B(2m, m, (1, 2m)) \cap B(2m, (1, 2m), m),$$

$$B_6 = B((1, 2m), 2m, m) \cap B(m, 2m, (1, 2m)),$$

$$B_7 = B(1, m, (1, 2m)) \cap B(1, (1, 2m), m).$$

Let $B = \cap_{i \in [7]} B_i$. Note that

$$\mathbb{P}(E^c) = \mathbb{P}(E^c \cap A \cap B) + \mathbb{P}(E^c \cap A^c \cap B) + \mathbb{P}(E^c \cap A \cap B^c) + \mathbb{P}(E^c \cap A^c \cap B^c). \tag{6.2}$$

Therefore we will try to bound the 4 terms on the right side of (6.2). If we can show that $A \cap B = E$,

then the first term on the right side of (6.2) vanishes. The next lemma shows this.

Lemma 6.4.4. *If the events A and B occur, then the event E occurs, i.e., $[2m]^3$ is good.*

Proof. Note that since all 8 subcubes $C_1, C_2, ..., C_8$ are good, the interior of C_i has been infected after m^3 steps for all $i \in [8]$. Thus after m^3 steps the only possibly uninfected vertices are the vertices on the sides for each cube C_i. Without loss of generality, since the event $B(m, (1, 2m), 2m)$ occurs, we can assume that there is a vertex $a = (a_1, a_2, a_3)$ in $[(m), (1, m), (2m)]$ is initially infected.

Let us consider the vertices in $[(1, 2m), (1, 2m), (2m)]$ after m^3 steps. Since all 8 subcubes $C_1, C_2, \cdots, C_8$ are good, every vertex in $[(m), (2, m-1), 2m]$ has at least 2 infected neighbors after m^3 steps. Since the vertex a is initially infected, it takes at most $m^3 + m$ steps to infect every vertex in $[(m), (2, m-1), 2m]$.

Since the vertex a is initially infected and all 8 subcubes are good, the vertex $y = (a_1 + 1, a_2, a_3)$ has 3 infected neighbors after m^3 so it will be infected in the next step and then it takes at most $m^3 + m$ steps to infect every vertex in $[(m+1), (2, m-1), 2m]$.

Note that every vertex in

$$[(m), (m), (m+2, 2m-1)],$$

$$[(m+1), (m), (m+2, 2m-1)],$$

$$[(m), (m+1), (m+2, 2m-1)],$$

and

$$[(m+1), (m+1), (m+2, 2m-1)]$$

already has 2 infected neighbors since all 8 small cubes are good. Therefore since the event $B(m, m, (m+1, 2m))$ occurs, it takes at most $m^3 + m$ steps to infect every vertex in

$$[(m), (m), (m+2, 2m-1)],$$

$$[(m+1),(m),(m+2,2m-1)],$$

$$[(m),(m+1),(m+2,2m-1)],$$

and

$$[(m+1),(m+1),(m+2,2m-1)].$$

Then note that after m^3+m steps, the vertices $(m,m,2m)$ and $(m+1,m,2m)$ have at least two infected neighbors and the vertices $(m,m+1,2m)$ and $(m+1,m+1,2m)$ have at least one infected neighbor.

Since the event $B((1,2m),m,2m)$ occurs, a vertex b in

$$[(1,m),(m),(2m)],$$

$$[(1,m),(m+1),(2m)],$$

$$[(m+1,2m),(m),(2m)],$$

or

$$[(m+1,2m),(m+1),(2m)]$$

is initially infected. Then following the same logic, every vertex in $[(2,m-1),(m),2m]$ and $[(2,m-1),(m+1),(2m)]$ will be infected after at most m^3+m steps. Thus the vertex $(m,m,2m)$ has 3 infected neighbors after at most m^3+m steps and will be infected in the next step and after that the vertices $(m,m+1,2m)$ and $(m+1,m,2m)$ will be infected. Since the vertex $(m+1,m,2m)$ has been infected after at most m^3+m steps it is easy to see that every vertex in $[(m+1,2m-1),(m),(2m)]$ will be infected after at most m^3+2m steps. Then every vertex in $[(m+1,2m-1),(m+1),(2m)]$ has 3 infected neighbors

after at most $m^3 + 2m$ steps and will be infected in the next step. Similarly, every vertex in

$$[(m), (m+1, 2m-1), (2m),] \quad [(m+1), (m+1, 2m-1), (2m)]$$

will be infected after at most $m^3 + 3m$ steps.

The rest of the uninfected vertices on the edges of $C_1, \ldots, C_8$ can be handled in a very similar way and so we have the desired result. □

Using Lemma 6.4.4, we have

$$\begin{aligned}
\mathbb{P}(E^c) &= \mathbb{P}(E^c \cap A \cap B) + \mathbb{P}(E^c \cap A^c \cap B) + \mathbb{P}(E^c \cap A \cap B^c) + \mathbb{P}(E^c \cap A^c \cap B^c) \\
&\leqslant \mathbb{P}(E^c \cap A^c \cap B) + 2\mathbb{P}(B^c) \\
&= 2\mathbb{P}(B^c) + \mathbb{P}(\cup_{i=1}^{8}(E^c \cap A_i \cap B)) \\
&\leqslant C'(1-p)^{4m} + \mathbb{P}(E^c \cap A_1 \cap B) + \mathbb{P}(E^c \cap A_2 \cap B) + \sum_{i=3}^{8} \mathbb{P}(A_i) \\
&\leqslant C'(1-p)^{4m} + \mathbb{P}(E^c \cap A_1 \cap B) + \mathbb{P}(E^c \cap A_2 \cap B) + \sum_{i=3}^{8} \binom{8}{i} \eta_{m,3}^{i} \\
&\leqslant C'(1-p)^{4m} + \mathbb{P}(E^c \cap A_1 \cap B) + \mathbb{P}(E^c \cap A_2 \cap B) + C\eta_{m,3}^{3},
\end{aligned}$$

where $C > 0$ and $C' > 0$ are absolute constants. Now let us estimate the term $\mathbb{P}(E^c \cap A_1 \cap B)$.

Before proceeding, let us define a few events.

$$B_8 = \cap_{a_2 \in \{2m, m+1\}} B(m, a_2, (1, 2m)) \cap B((1, 2m), a_2, m) \cap B(m, (m+1, 2m), m),$$

$$B_9 = \cap_{a_1 \in \{2m, m+1\}} B(a_1, m, (1, 2m)) \cap B(a_1, (1, 2m), m) \cap B((m+1, 2m), m, m),$$

and

$$B_{10} = \cap_{a_3 \in \{2m, m+1\}} B(m, (1, 2m), a_3) \cap B((1, 2m), m, a_3) \cap B(m, m, (m+1, 2m)).$$

Lemma 6.4.5. *The probability that $[2m]^3$ is bad along with the occurrence of the event that there is exactly one bad subcube is correlated with double empty line segments of length $2m$. More precisely, we have*

$$\mathbb{P}(E^c \cap A_1) \leqslant C(1-p)^{4m}.$$

Proof. Due to symmetry, without loss of generality, we may assume that the subcube C_1 is bad. The main idea of the proof is as follows. We will construct 4 set of events $S_{start}, S_{x=1}, S_{y=1}$ and $S_{z=1}$ such that, if the events in these 4 sets occur along with events A_1 and B, then the event E also occurs; that is, the cube $[2m]^d$ is good. Moreover, the probability that some of these events in these 4 sets fails to occur is approximately $(1-p)^{4m}$. If the probability that any of the events in these four sets fails to occur is significantly larger than $(1-p)^{4m}$, then a more detailed analysis is required. The rough idea is that other events, not contained in S_{start}, $S_{x=1}$, $S_{y=1}$, or $S_{z=1}$, can also lead to percolation. The probability of the complement of these additional events, together with the complement of the events in S_{start}, $S_{x=1}$, $S_{y=1}$, and $S_{z=1}$, is $\leqslant (1-p)^{4m}$. The exact details are presented below.

The infection spreads as follows. First, after at most m^3 steps, the interior of $C_2, \ldots, C_8$ have been infected since $C_2, \ldots, C_8$ are good. Then the sides of the subcubes $C_2, \ldots, C_8$ become infected (except for those on $[(1, 2m), (1), (1, 2m)]$, $[(1), (1, 2m), (1, 2m)]$ and $[(1, 2m), (1, 2m), (1)]$ and those belonging to the sides of $[2m]^3$)after at most $m^3 + Cm$ steps, due to the occurrence of the events in S_{start}. Then, after at most $2m^3 + Cm$ steps the interior of C_1 have been infected. Finally, the uninfected vertices on $[(1, 2m), (1), (1, 2m)]$, $[(1), (1, 2m), (1, 2m)]$ and $[(1, 2m), (1, 2m), (1)]$ will be infected after at ost $2m^3 + Cm'$ steps, due to the occurrence of the event in $S_{x=1}, S_{y=1}$ and $S_{z=1}$.

Now we will start defining the events in the set S_{start}. It turns out that the following B_8, B_9, and B_{10}

serve our purpose.

Now we can focus on analyzing how to infect the vertices on $[(1, 2m), (1), (1, 2m)]$, $[1, (1, 2m), (1, 2m)]$, and $[(1, 2m), (1, 2m), (1)]$. Due to symmetry we only need to consider infecting vertices on $[(1, 2m), (1), (1, 2m)]$.

Consider the vertices in the region $[(1, 2m), (1), (1, 2m)]$ as shown in Figure 6.2, where the vertices in the shaded area are infected, and those in the white area remain uninfected. Note that after $2m^3 + Cm$ steps among those vertices the only possible uninfected vertices are on $[(1, m), (1), (1, m)]$ and the sides of C_2, C_5, C_6.

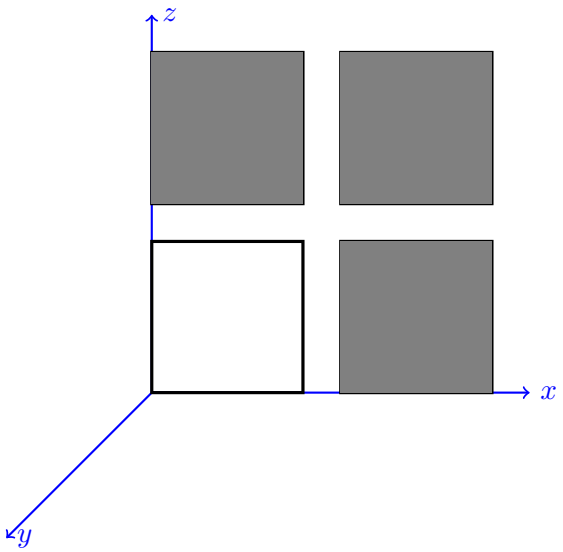

Figure 6.2: Configuration on $[(1, 2m), (1), (1, 2m)]$ after $2m^3 + Cm$ steps.

Let us define the event $D(y = 1)$ in $S_{y=1}$ as

$$D(y = 1) := \{[(1, m), (1), (m)] \text{ is non-empty}\} \cap \{[(m), (1), (1, m)] \text{ is non-empty}\}.$$

It is easy to see that if the events A_1, B, B_i for $i \in \{8, 9, 10\}$ and $D(y = 1)$ occurs, then every vertex on $[(1, 2m), (1), (1, 2m)]$ will be infected except for those on the sides of $[2m]^3$.

We will define the event $D(x = 1)$ in $S_{x=1}$ and the event $D(z = 1)$ in $S_{z=1}$ accordingly.

Therefore, we have

$$\mathbb{P}(E^c \cap A_1 \cap B \cap (\cap_{i=8}^{10} B_i) \cap D(x=1) \cap D(y=1) \cap D(z=1)) = 0.$$

and thus we have

$$\begin{aligned}\mathbb{P}(E^c \cap A_1 \cap B) &= \mathbb{P}(E^c \cap A_1 \cap B \cap (\cap_{i=8}^{10} B_i) \cap D(x=1) \cap D(y=1) \cap D(z=1)) \\ &\quad + \mathbb{P}(E^c \cap A_1 \cap B \cap ((\cap_{i=8}^{10} B_i) \cap D(x=1) \cap D(y=1) \cap D(z=1))^c) \\ &= \mathbb{P}(E^c \cap A_1 \cap B \cap ((\cap_{i=8}^{10} B_i) \cap D(x=1) \cap D(y=1) \cap D(z=1))^c).\end{aligned}$$

We need to further analyze the event $E^c \cap A_1 \cap B \cap ((\cap_{i=8}^{10} B_i) \cap D(x=1) \cap D(y=1) \cap D(z=1))^c$. Note that $\mathbb{P}(B_i^c) \leqslant C(1-p)^{4m}$ for $i \in \{8, 9, 10\}$.

Since $\mathbb{P}(D(y=1)^c) = 2(1-p)^m$, we need to construct some other events. Let us consider the event $D(y=1)^c$ and we have

$$D(y=1)^c = D_1 \cup D_2 \cup D_3,$$

where

$$D_1 = \{[(1,m),(1),(m)] \text{ is non-empty}\} \cap \{[(m),(1),(1,m)] \text{ is empty}\},$$

$$D_2 = \{[(1,m),(1),(m)] \text{ is empty}\} \cap \{[(m),(1),(1,m)] \text{ is non-empty}\},$$

and

$$D_3 = \{[(1,m),(1),(m)] \text{ is empty}\} \cap \{[(m),(1),(1,m)] \text{ is empty}\}.$$

We will need to do further analysis based on the event $D(y=1)^c$.

Case 1: Assume that the event D_1 occurs.

In order to proceed we need to define another event. Let $F_1(y=1)$ be the event that one of $[(m),(1),(m+1,2m)]$ or $[(m+1),(1),(m+1,2m)]$ is non-empty. It is easy to see that if the events $A_1, B, D_1, F_1(y=1)$ and B_i for $i \in \{8,9,10\}$ occur, then every vertex on $[(1,2m),(1),(1,2m)]$ will be infected after at most $2m^2 + Cm$ steps, except for those on the sides of $[2m]^3$.

Now we need to consider $A_1 \cap B \cap_{i=8}^{10} B_i \cap D_1 \cap F_1(y=1)^c$. Let $F_2(y=1)$ be the event that $[(m+1),(1),(1,m)]$ is non-empty. It is easy to see that if the events $A_1, B, D_1, F_1^c(y=1), F_2(y=1), B_i$ for $i \in \{8,9,10\}$ occur, then every vertex on $[(1,2m),(1),(1,2m)]$ will be infected except for the sides of $[2m]^3$.

Since $\mathbb{P}(D_1 \cap F_1(y=1)^c \cap F_2(y=1)^c) \leqslant (1-p)^{4m}$ and Fact 6.4.2, we have the desired result.

Case 2: Assume that the event D_2 happens.

The analysis is the same as Case 1 due to symmetry.

Case 3: Assume that the event D_3 happens.

Before proceeding we need to define another event. Define $F_3(y=1)$ to be the event that both $[(m+1),(1),(m+1,2m)]$ and $[(m+1,2m),(1),(m+1)]$ are non-empty. It is easy to see that if the events A_1, B, D_3, $F_3(y=1)$ and B_i for $i \in \{8,9,10\}$ occur, then every vertex on $[(1,2m),(1),(1,2m)]$ have been infected after at most $2m^3 + Cm$ steps, except for those on the sides of $[2m]^3$.

Now we need to analyze $F_3(y=1)^c$. Define the event H be the event that $[(1,m),(1),(m+1)]$ is non-empty.

We have

$$F_3(y=1)^c = F_1 \cup F_2 \cap F_3,$$

where

$$F_1 = \{[(m+1),(1),(m+1,2m)] \text{ is non-empty}\} \cap \{[(m+1,2m),(1),(m+1)] \text{ is empty}\},$$

$$F_2 = \{[(m+1),(1),(m+1,2m)] \text{ is empty}\} \cap \{[(m+1,2m),(1),(m+1)] \text{ is non-empty}\},$$

and

$$F_3 = \{[(m+1),(1),(m+1,2m)] \text{ is empty}\} \cap \{[(m+1,2m),(1),(m+1)] \text{ is empty}\}.$$

Assume that F_1 occurs. Then again it is easy to see that if the events A_1, B, D_3, F_1, H and B_i for $i \in \{8,9,10\}$ occur, then every vertex on $[(1,2m),(1),(1,2m)]$ have been infected after at most $2m^3 + Cm$ steps, except for those on the sides of $[2m]^3$. Note that $\mathbb{P}(D_3 \cap F_1 \cap H^c) = (1-p)^{4m}$.

Assume that F_2 occurs. The analysis is the same as the case where F_1 occurs.

Assume that F_3 occurs. Note that $\mathbb{P}(D_3 \cap F_3) = (1-p)^{4m}$. □

Now let us move on to estimating the term $\mathbb{P}(E^c \cap A_2 \cap B)$.

Lemma 6.4.6. *The probability that $[2m]^3$ is bad along with the occurrence of the event that there are exactly two subcubes are bad is correlated with double empty line segments of length $2m$. More precisely, we have*

$$\mathbb{P}(E^c \cap A_2) \leqslant C(d)m^2(1-p)^{4m-8}.$$

Proof. In order to estimate $\mathbb{P}(E^c \cap A_2 \cap B)$ we will separate cases based on the locations of the two bad subcubes.

Case 1: Two bad cubes are adjacent.

Without loss of generality assume the subcubes C_1 and C_2 are bad, as in Figure 6.3.

The main idea of the proof is somewhat similar to that of Lemma 6.4.5. We will construct 5 sets of

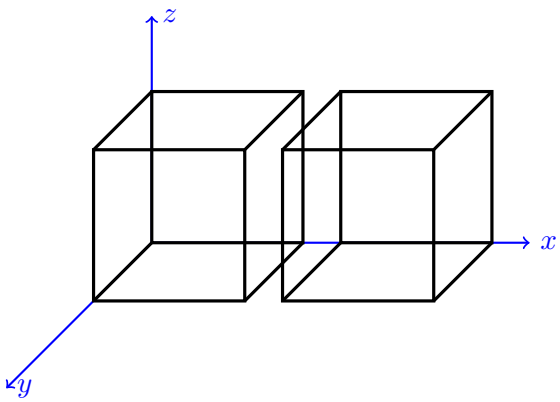

Figure 6.3: Bad subcubes C_1 and C_2.

events $T_{start}, T_{x=1}, T_{y=1}$, $T_{z=1}$ and $T_{x=2m}$ such that, if the events in these 5 sets occur along with events A_2 and B, then the event E also occurs; that is, the cube $[2m]^d$ is good. Moreover, the probability that some of these events in these 5 sets fails to occur is approximately $(1-p)^{4m}$. If the probability that some of the events in these 5 sets fails to occur is significantly larger than $(1-p)^{4m}$, then a more detailed analysis is required. The rough idea is that other events, not contained in T_{start}, $T_{x=1}$, $T_{y=1}$, $T_{z=1}$,or $T_{x=2m}$ can also lead to percolation. The probability of the complement of these additional events, together with the complement of the events in T_{start}, $T_{x=1}$, $T_{y=1}$,$T_{z=1}$,and $T_{x=2m}$ is $\leqslant (1-p)^{4m}$. The exact details are presented below.

The infection spreads as follows. First, after at most m^3 steps, the interior of $C_3, \ldots, C_8$ have been infected since $C_3, \ldots, C_8$ are good. Then the sides of the subcubes $C_3, \ldots, C_8$ become infected (except for those on $[(1,2m),(1),(1,2m)]$, $[(1),(1,2m),(1,2m)]$, $[(1,2m),(1,2m),(1)]$ and $[(2m),(1,2m),(1,2m)]$ and those belonging to the sides of $[2m]^d$)after at most $m^3 + Cm$ steps, due to the occurrence of the events in S_{start}. Here it comes to the difference between this case and the case where there is only one subcube is bad. We will need to define a new event H' to be {every two adjacent line segments of the form $[(2,2m-1),(y),(z)]$ of length $2m-2$ in $[(1,2m),(1,m),(1,m)]$ parallel to the x-axis is non-empty}, where $y, z \in \{1,2,\ldots,m\}$.

Then, after at most $3m^3 + Cm$ steps the interior of C_1 and C_2 have been infected due to the occurrence

of the events H'. Finally, the uninfected vertices on $[(1,2m),(1),(1,2m)]$, $[(1),(1,2m),(1,2m)]$ and $[(1,2m),(1,2m),(1)]$ will be infected after at most $3m^3 + Cm'$ steps, due to the occurrence of the event in $T_{x=1}, T_{y=1}, T_{z=1}$ and $T_{x=2m}$.

It is easy to see that $\mathbb{P}(H'^c) \leqslant Cm^2(1-p)^{4m-4}$.

Now we will start defining the events in the set T_{start}. It turns out that B_8 and B_{10} in the set T_{start} serve our purpose. It is easy to see that after at most $2m^3 + Cm$ steps, every vertex in $[(1,2m),(1,2m),(1,2m)]$ will be infected except for the vertices in $[(1),(1,2m),(1,2m)]$,$[(1,2m),(1),(1,2m)]$, $[(1,2m),(1,2m),(1)]$, and $[(2m),(1,2m),(1,2m)]\}$ and the sides of $[2m]^3$ due to the occurrence of A_2, B_8, B_{10} and H'.

We can apply the same method from Lemma 6.4 to infect uninfected vertices in $[(1),(1,2m),(1,2m)]$, $[(1,2m),(1,2m),(1)]$, and $[(2m),(1,2m),(1,2m)]\}$. Therefore, we just need to focus on infecting the vertices on $[(1,2m),(1),(1,2m)]$.

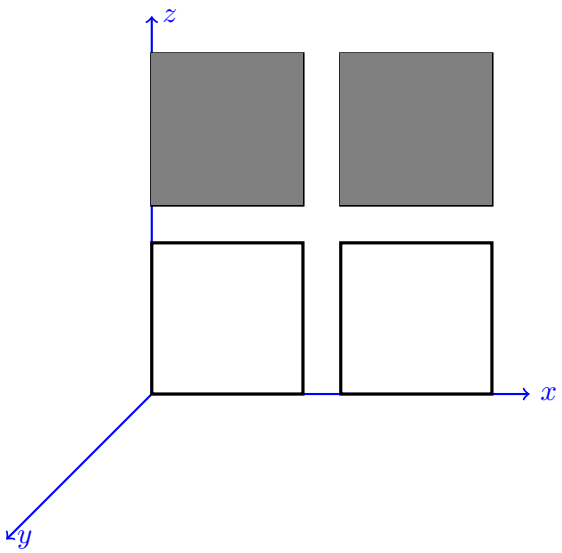

Figure 6.4: Configuration on $[(1,2m),(1),(1,2m)]$ after $3m^3 + Cm^2$ steps.

Consider the vertices in the region $[(1,2m),(1),(1,2m)]$ as shown in Figure 6.4, where the vertices in the shaded area are infected, and those in the white area remain uninfected. Note that since the subcubes C_5 and C_6 are good, the only uninfected vertices are in the sides of C_5 and C_6 and $[(1,2m)(1)(1,m)]$ after

$3m^3 + Cm^2$ steps.

Let us define the events in $T_{y=1}$. Let

$E_1(y = 1)$:={every two adjacent line segments of length $m - 2$ in $[(2, m - 1), (1), (1, 2m)]$ parallel to the x-axis is non-empty},

$E_2(y = 1)$:={one of $[(m), (1), (2, 2m - 1)]$ or $[(m + 1), (1), (2, 2m - 1)]$ is non-empty},

and $E_3(y = 1)$:={every two adjacent line segments of length $m - 2$ in $[(m + 2, 2m - 1), (1), (1, 2m)]$ parallel to the x-axis is non-empty}.

It is clear that if the event $E_2(y = 1)$ occurs along with either the event $E_1(y = 1)$ or the event $E_3(y = 1)$ then every vertex in $[(1, 2m), (1), (1, 2m)]$ will be infected except for the sides of $[2m]^3$.

Note that $\mathbb{P}(E_1(y = 1)^c \cap E_3(y = 1)^c) = (1 - p)^{4m-8}$ and $\mathbb{P}(E_2(y = 1)^c) = (1 - p)^{4m-4}$. By Fact 6.4.2, we have the desired result.

Case 2: Two bad subcubes are non-adjacent and not located diagonally.

Without loss of generality, assume that the subcubes C_1 and C_4 are bad as in Figure 6.5.

The main idea is as follows. We will define two new events E_4 and E_5 such that along with the occurrence of the events A_2 and B, every vertex in $[(1, 2m), (1, 2m), (1, 2m)]$ has been infected except for those belonging to $[(1), (1, 2m), (1, 2m)]$, $[(1, 2m), (1), (1, 2m)]$, $[(1, 2m), (1, 2m), (1)]$, $[(2m), (1, 2m), (1, 2m)]$, $[(1, 2m), (2m), (1, 2m)]$ and the edges of $[2m]^3$ after at most $2m^3 + Cm$ steps. Another requirement for the events E_4 and E_5 is that $\mathbb{P}(E_4^c \cap E_5^c)$ is roughly $(1 - p)^{4m}$. Then consider the uninfected vertices on $[(1), (1, 2m), (1, 2m)]$, $[(1, 2m), (1), (1, 2m)]$, $[(1, 2m), (1, 2m), (1)]$, $[(2m), (1, 2m), (1, 2m)]$, $[(1, 2m), (2m), (1, 2m)]$ and the sides of $[2m]^3$ after $2m^3 + Cm$ steps. The analysis of infection for these uninfected vertices is the same as in the cases where the two bad subcubes are adjacent or where there is only one bad subcube.

Now let us define events E_4 and E_5.

$E_4 :=$ {every two adjacent line segments of length $m-2$ on $[(m+2, 2m-1), (2, m+1), (m)]$ parallel to the x-axis is non-empty}, and

$E_5 :=$ { at least one of the two line segments $[(m), (2, 2m-1), (m)]$ or $[(m+1), (2, 2m-1), (m)]$ is non-empty}.

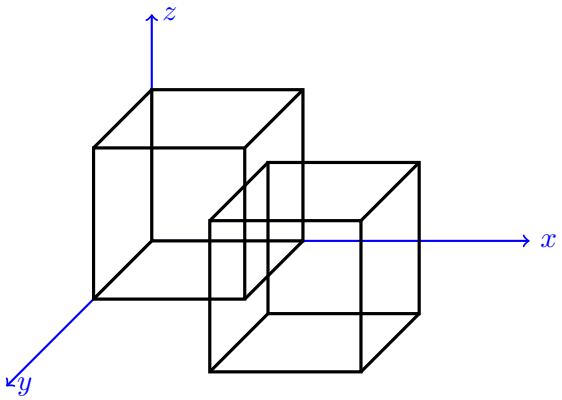

Figure 6.5: Bad subcubes C_1 and C_4.

Case 3: Two bad subcubes are located diagonally.

Without loss of generality we can assume the subcubes C_3 and C_6 are bad as in Figure 6.6.

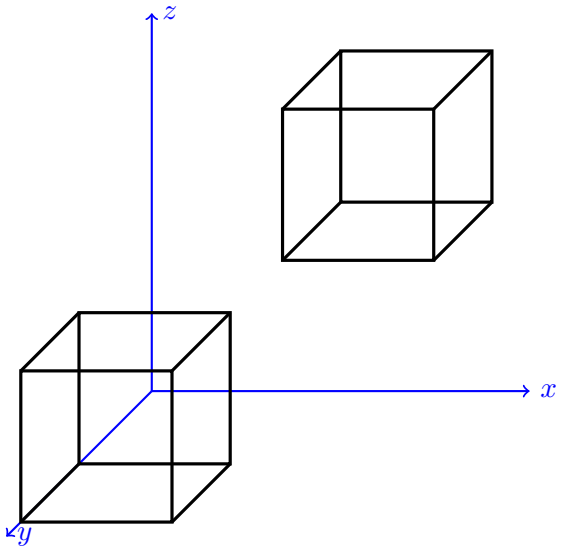

Figure 6.6: Bad subcubes C_3 and C_6.

The main idea is as follows. We will define three new events E_6 , E_7 and E_8 such that along with the

occurrence of the events A_2 and B, every vertex in $[(1,2m),(1,2m),(1,2m)]$ has been infected except for those belonging to $[(1),(1,2m),(1,2m)]$, $[(1,2m),(1),(1,2m)]$, $[(1,2m),(1,2m),(1)]$, $[(2m),(1,2m),(1,2m)]$, $[(1,2m),(2m),(1,2m)]$, $[(1,2m),(1,2m),(2m)]$ and the sides of $[2m]^3$ after at most $2m^3+Cm$ steps. Another requirement for the events E_6, E_7 E_8 is that $\mathbb{P}(E_i^c)$ is roughly $(1-p)^{4m}$ for $i \in \{6,7,8\}$. Then consider the uninfected vertices on $[(1),(1,2m),(1,2m)]$, $[(1,2m),(1),(1,2m)]$, $[(1,2m),(1,2m),(1)]$, $[(2m),(1,2m),(1,2m)]$, $[(1,2m),(2m),(1,2m)]$ except for these on the sides of $[2m]^3$ after $2m^3+Cm$ steps. The analysis of infection for these uninfected vertices is the same as in the case where there is only one bad subcube.

Let us define the events E_6,E_7, and E_8.

E_6:= {at least one of the 4 line segments of $[(m),(1,m),(m)]$, $[(m),(1,m),(m{+}1)]$, $[(m{+}1),(1,m),(m)]$, or $[(m+1),(1,m),(m+1)]$ is non-empty},

E_7:= {at least one of the 4 line segments $[(1,m),(m),(m)][(1,m),(m),(m{+}1)]$, $[(1,m),(m{+}1),(m)]$, or $[(1,m),(m+1),(m+1)]$ is non-empty},

E_8={at least one of the 4 line segments $[(m+1,2m),(m),(m)]$, $[(m+1,2m),(m),(m+1)]$, $[(m+1,2m),(m+1),(m)]$, or $[(m+1,2m),(m+1),(m+1)]$ is non-empty}. □

Now we are in a position to prove Lemma 6.4.3, which is the main lemma in this subsection.

Proof. In order to estimate the probability of the event E^c we will write the event E^c in a different way and

so we have, by using Lemma 6.4.4, Lemma 6.4.5 and Lemma 6.4.6,

$$\begin{aligned}
\mathbb{P}(E^c) &= \mathbb{P}(E^c \cap A \cap B) + \mathbb{P}(E^c \cap A^c \cap B) + \mathbb{P}(E^c \cap A \cap B^c) + \mathbb{P}(E^c \cap A^c \cap B^c) \\
&\leqslant 2\mathbb{P}(B^c) + \mathbb{P}(E^c \cap A^c \cap B) \\
&= 2\mathbb{P}(B^c) + \mathbb{P}(\cup_{i=1}^{8}(E^c \cap A_i \cap B)) \\
&\leqslant C'(1-p)^{4m} + \mathbb{P}(E^c \cap A_1 \cap B) + \mathbb{P}(E^c \cap A_2 \cap B) + \sum_{i=3}^{8} \mathbb{P}(A_i) \\
&\leqslant C'(1-p)^{4m} + \mathbb{P}(E^c \cap A_1 \cap B) + \mathbb{P}(E^c \cap A_2 \cap B) + \sum_{i=3}^{8} \binom{8}{i} \eta_{m,3}^{i} \\
&\leqslant C'(1-p)^{4m} + \mathbb{P}(E^c \cap A_1 \cap B) + \mathbb{P}(E^c \cap A_2 \cap B) + C\eta_{m,3}^{3} \\
&\leqslant Bm^2(1-p)^{4m-8} + C\eta_{m,3}^{3},
\end{aligned}$$

where C and C' are absolute constants. □

6.4.2 $d = r \geqslant 3$

This lemma gives a recursive relation on the probability that $[m]^d$ is bad. As mentioned in the introduction, we prove Lemma 6.4.7 using induction, where Lemma 6.4.3 is the base case.

Lemma 6.4.7. *Let $\eta_{m,r}$ be the probability that $[m]^d$ is bad with the infection threshold $r = d$. We have*

$$\eta_{2m,d} \leqslant C(d)\eta_{m,d}^{3} + C(d)m^{d-1}(1-p)^{4m-8}.$$

The intuition behind the proof of Lemma 6.4.7 is as follows. We begin by partitioning the cube $[2m]^{d+1}$ into 2^{d+1} disjoint subcubes and break the proof into several subcases according to the number of bad subcubes. Specifically, we analyze the cases where exactly one or two of the 2^{d+1} subcubes are bad as well as the case where none are bad. The main technical part of the proof is devoted to the analysis to these three subcases, where we study the interaction arising among the 2^{d+1} subcubes by using the induction on the

dimension d. Further the probability of having at least three bad subcubes can be estimated by a very rough upper bound of $C\eta^3_{m,d+1}$, as established in Lemma 6.4.7.

Before proving this lemma we need to introduce some notation and prove several auxiliary lemmas.

Consider partitioning $[2m]^d$ into 2^d subcubes, i.e

$$[(a_1, a_1 + m - 1), (a_2, a_2 + m - 1), ..., (a_d, a_d + m - 1)],$$

where $a_i = \{1, m + 1\}$ for $i \in [d]$.

First we need to define a few events.

- $E^d := [2m]^d$ is good.
- A^d :=$\{$all 2^d subcubes are good $\}$.
- A^d_i :=$\{$exactly i subcubes among 2^d subcubes are bad $\}$ for $i \in [2^d]$.

We would like to prove that for $d \geqslant 3$

$$\mathbb{P}((E^d)^c \cap A^d) \leqslant C(d)(1 - p)^{4m}, \tag{6.3}$$

$$\mathbb{P}((E^d)^c \cap A^d_1) \leqslant C(d)m^{d-3}(1 - p)^{4m}, \tag{6.4}$$

and

$$\mathbb{P}((E^d)^c \cap A^d_2) \leqslant C(d)m^{d-1}(1 - p)^{4m-8}. \tag{6.5}$$

Then we have

$$\begin{aligned}\mathbb{P}((E^d)^c) &= \mathbb{P}((E^d)^c \cap A^d) + \mathbb{P}((E^d)^c \cap (A^d)^c) \\ &= \mathbb{P}((E^d)^c \cap A^d) + \mathbb{P}(\cup_{i=1}^{2^d}((E^d)^c \cap A_i^d)) \\ &= \mathbb{P}((E^d)^c \cap A^d) + \mathbb{P}((E^d)^c \cap A_1^d) + \mathbb{P}((E^d)^c \cap A_2^d) + \sum_{i=3}^{2^d} \mathbb{P}(A_i^d) \\ &\leqslant C(d)m^{d-1}(1-p)^{4m-8} + C(d)\eta_{m,3}^3.\end{aligned}$$

When $d = 3$ it is easy to see that (6.3)(6.4)(6.5) are satisfied by Lemmas (6.4.3)(6.4.5)(6.4.6). Now assume that (6.3)(6.4)(6.5) are satisfied for $d \geqslant 3$ we would like to show they are satisfied for $d+1$.

Before proceeding, we need some notation. Let a d-dimensional subgrid $[\underbrace{(1,2m),(1,2m),\ldots,(1,2m)}_{d},(i)]$ be denoted by $B((1,2m),d,i)$ and a d-dimensional subgrid $[\underbrace{(1,m),(1,m),\ldots,(1,m)}_{d},(i)]$ by $B((1,m),d,i)$.

Let us define E_i^d to be the event that $(B(1,2m),d,i)$ is good.

Let $B\left(a_1,a_2,\cdots,a_{i-1},a_{i+1},\cdots,a_{j-1},a_{j+1},\cdots,a_d,(1,m)\right)$ be the event that at least one of the 4 line segments

$$[(a_1),(a_2)\cdots,(a_{i-1}),(m),(a_{i+1}),\cdots,(a_{j-1}),(m),(a_{j+1}),\cdots,(a_d),(1,m)],$$

$$[(a_1),(a_2)\cdots,(a_{i-1}),(m),(a_{i+1}),\cdots,(a_{j-1}),(m+1),(a_{j+1}),\cdots,(a_d),(1,m)],$$

$$[(a_1),(a_2)\cdots,(a_{i-1}),(m+1),(a_{i+1}),\cdots,(a_{j-1}),(m),(a_{j+1}),\cdots,(a_d),(1,m)],$$

or

$$[(a_1),(a_2)\cdots,(a_{i-1}),(m+1),(a_{i+1}),\cdots,(a_{j-1}),(m+1),(a_{j+1}),\cdots,(a_d),(1,m)]$$

is non-empty, where $a_l \in \{1,m,m+1,2m\}$ for $l \in [d]\backslash\{i,j\}$.

The event $B\left(a_1,a_2,\cdots,a_{i-1},a_{i+1},\cdots,a_{j-1},a_{j+1},\cdots,a_d,(m+1,2m)\right)$ is defined similarly.

Let $B(\sigma\,(a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_{j-1}, a_{j+1}, \cdots, a_d, (1, m)))$ denote the event that at least one of the 4 line segments

$$\sigma[(a_1), (a_2) \cdots, (a_{i-1}), (m), (a_{i+1}), \cdots, (a_{j-1}), (m), (a_{j+1}), \cdots, (a_d), (1, m)],$$

$$\sigma[(a_1), (a_2) \cdots, (a_{i-1}), (m), (a_{i+1}), \cdots, (a_{j-1}), (m+1), (a_{j+1}), \cdots, (a_d), (1, m)],$$

$$\sigma[(a_1), (a_2) \cdots, (a_{i-1}), (m+1), (a_{i+1}), \cdots, (a_{j-1}), (m), (a_{j+1}), \cdots, (a_d), (1, m)],$$

or

$$\sigma[(a_1), (a_2) \cdots, (a_{i-1}), (m+1), (a_{i+1}), \cdots, (a_{j-1}), (m+1), (a_{j+1}), \cdots, (a_d), (1, m)]$$

is nonempty, where $a_l \in \{1, m, m+1, 2m\}$ for $l \in [d] \backslash \{i, j\}$ and $\sigma \in \mathfrak{S}_{d+1}$.

The event $B(\sigma\,(a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_{j-1}, a_{j+1}, \cdots, a_d, (m+1, 2m)))$ is defined similarly, where $\sigma \in \mathfrak{S}_{d+1}$.

Let us define $D_\sigma(1, m)$ and $D_\sigma(1, 2m)$ to be

$$D_\sigma(1, m) := B(\sigma\,(a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_{j-1}, a_{j+1}, \cdots, a_d, (1, m)))$$

and

$$D_\sigma(m+1, 2m) := B(\sigma\,(a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_{j-1}, a_{j+1}, \cdots, a_d, (m+1, 2m))).$$

Let us define D to be

$$D := \cap_{l \in [d+1] \backslash \{i,j\}} \cap_{a_l \in \{1, m, m+1, 2m\}} \cap_{\sigma \in \mathfrak{S}_{d+1}} \big(D_\sigma(1, m) \cap D_\sigma(m+1, 2m)\big).$$

Let B' be defined as

$$B' := \cap_{l \in [d+1] \setminus \{i,j\}} \cap_{a_l \in \{1,m,m+1,2m\}} \cap_{\sigma \in \mathfrak{S}_{d+1}} D_\sigma(m+1, 2m).$$

Let $D\left(a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_d, m\right)$ be the event that at least one of the 4 line segments

$$[(a_1), (a_2) \cdots, (a_{i-1}), (1, m), (a_{i+1}), \cdots, (a_d), (m)],$$

$$[(a_1), (a_2) \cdots, (a_{i-1}), (m+1, 2m), (a_{i+1}), \cdots, (a_d), (m)],$$

$$[(a_1), (a_2) \cdots, (a_{i-1}), (1, m), (a_{i+1}), \cdots, (a_d), (m+1)],$$

or

$$[(a_1), (a_2) \cdots, (a_{i-1}), (m+1, 2m), (a_{i+1}), \cdots, (a_d), (m+1)]$$

is non-empty, where $a_i \in \{1, 2m\}$ for $i \in [d]$.

The event $D(\sigma(a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_d, m))$ is defined similarly.

Let us define D' to be

$$D' := \cap_{j \in [d+1] \setminus i]} \cap_{a_j \in \{1,2m\}} \cap_{\sigma \in \mathfrak{S}_{d+1}} D\left(a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_d, m\right)$$

Let $E\left(a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_d, m\right)$ be an event that one of the 4 segments

$$[(a_1), (a_2) \cdots, (a_{i-1}), (m), (a_{i+1}), \cdots, (a_d), (1, m)],$$

$$[(a_1), (a_2) \cdots, (a_{i-1}), (m+1), (a_{i+1}), \cdots, (a_d), (1, m)],$$

$$[(a_1), (a_2) \cdots, (a_{i-1}), (m), (a_{i+1}), \cdots, (a_d), (m+1, 2m)],$$

or

$$[(a_1), (a_2) \cdots, (a_{i-1}), (m+1), (a_{i+1}), \cdots, (a_d), (m+1, 2m)]$$

is non-empty, where $a_i \in \{1, 2m\}$ for $i \in [d]$.

The event $E\left(\sigma(a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_d, m)\right)$ is defined similarly.

Let $E(1, 2m)$ to be defined as

$$E(1,2m) := \cap_{l \in [d+1] \setminus \{i\}} \cap_{a_l \in \{1,2m\}} \cap_{\sigma \in \mathfrak{S}_{d+1}} E\left(\sigma(a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_d, m)\right).$$

Let us first deal with the case where all 2^{d+1} subcubes are good, where the subcubes are $[(a_1, a_1 + m - 1), (a_2, a_2 + m - 1), ..., (a_{d+1}, a_{d+1} + m - 1)]$ where $a_i = \{1, m+1\}$ for $i \in [d+1]$.

Lemma 6.4.8. *The probability that $[2m]^d$ is bad along with the occurrence of the event that all subcubes are good is correlated with double empty line segments of length $2m$. More precisely, we have*

$$\mathbb{P}((E^d)^c \cap A^d) \leqslant C(d)(1-p)^{4m}.$$

Proof. The idea for the proof is as follows. We will construct some events in the set B_{start} such that with the occurrence of these events and the event A^{d+1} then the event E^{d+1} occur. Additionally, the probability of the complement of these events is roughly $(1-p)^{4m}$.

Assume that all 2^{d+1} subcubes are good, i.e

$$[(a_1, a_1 + m - 1), (a_2, a_2 + m - 1), ..., (a_{d+1}, a_{d+1} + m - 1)],$$

where $a_i = \{1, m+1\}$ for $i \in [d+1]$, are good.

After m^{d+1} steps, the only possibly uninfected vertices on $B((1, 2m), d, i)$ for $i \notin \{1, m, m+1, 2m\}$

are $[(a_1), (a_2), \cdots, (a_d), (i)]$, where $a_j \in \{1, m, m+1, 2m\}$ for $j \in [d]$.

Now consider the possibly uninfected vertices on $B((1, 2m), d, i)$ for $i \in \{1, m, m+1, 2m\}$ after m^{d+1} steps. Note that the vertices on the sides of subgrids

$$[(a_1, a_1 + m - 1), (a_2, a_2 + m - 1), ..., (a_d, a_d + m - 1), (j)],$$

where $a_i \in \{1, m+1\}$ for $i \in [d]$ and $j \in \{1, m, m+1, 2m\}$, may have not been infected.

Observe that the vertices in the sides of the subcubes $B((1, 2m), d, i)$ where $i \in \{m, m+1\}$ do not belong to the sides of $[2m]^{d+1}$ except for the vertices of the form

$$[(a_1), (a_2), \cdots, (a_d), (i)],$$

where $a_j \in \{1, m, m+1, 2m\}$ for $j \in [d]$ and $i \in \{m, m+1\}$.

It turns out that the events D, D' and $E(1, 2m)$ in B_{start} serve our purpose.

With the occurrence of D and A^{d+1}, after at most $m^{d+1} + C(d)m$ steps, every vertex on the subcube $B((1, 2m), d, i)$ with $i \notin \{1, m, m+1, 2m\}$ has been infected except for those belonging to the sides of $[2m]^{d+1}$ and the vertices in $[(a_1), (a_2), \cdots, (a_{i-1}), (m), (a_{i+1}), \cdots, (a_d), (1, 2m)]$ and $[(a_1), (a_2), \cdots, (a_{i-1}), (m+1), (a_{i+1}), \cdots, (a_d), (1, 2m)]$ where $a_j \in \{1, 2m\}$ for $j \in [d]$.

Now every vertex, except possibly for the vertices on the sides of $B((1, 2m), d, i)$, on subcubes $B((1, 2m), d, i)$ for $i \in \{1, m, m+1, 2m\}$ has exactly one infected neighbor which does not belong to $B((1, 2m), d, i)$ for

$i \in \{1, m, m+1, 2m\}$. Hence we can use the induction hypothesis and thus we have

$$\begin{aligned}
\mathbb{P}((E^{d+1})^c \cap A^{d+1}) &\leqslant \mathbb{P}((E^{d+1})^c \cap A^{d+1} \cap D \cap E_1^d \cap E_m^d \cap E_{m+1}^d \cap E_{2m}^d) \\
&\quad + \mathbb{P}((E^{d+1})^c \cap A^{d+1}(\cap D \cap E_1^d \cap E_m^d \cap E_{m+1}^d \cap E_{2m}^d)^c) \\
&\leqslant \mathbb{P}((E^{d+1})^c \cap A^{d+1} \cap D \cap E_1^d \cap E_m^d \cap E_{m+1}^d \cap E_{2m}^d) + 4\mathbb{P}((E_1^d)^c) + \mathbb{P}(D^c) \\
&\leqslant \mathbb{P}((E^{d+1})^c \cap A^{d+1} \cap D \cap E_1^d \cap E_m^d \cap E_{m+1}^d \cap E_{2m}^d) + C(1-p)^{4m}.
\end{aligned}$$

Now we need to handle the term $\mathbb{P}((E^{d+1})^c \cap D \cap A^{d+1} \cap E_1^d \cap E_m^d \cap E_{m+1}^d \cap E_{2m}^d)$.

It clear that with the occurrence of the events $D, A^{d+1}, E_1^d, E_m^d, E_{m+1}^d, E_{2m}^d$ and D', every vertex on $[2m]^{d+1}$ has been infected after at most $m^{d+1} + C(d)m$ steps, except for those on the sides of $[2m]^{d+1}$.

Thus we have

$$\mathbb{P}((E^{d+1})^c \cap D \cap A^{d+1} \cap E_1^d \cap E_m^d \cap E_{m+1}^d \cap E_{2m}^d \cap D' \cap E(1, 2m)) = 0$$

Note that

$$\mathbb{P}((D')^c) \leqslant C(d)(1-p)^{4m},$$

and

$$\mathbb{P}(E(1, 2m)^c) \leqslant C(d)(1-p)^{4m}.$$

By Fact 6.4.2, we have the desired result. □

Now we will move onto the case where there is exactly one bad subcube and we will show the following lemma.

Lemma 6.4.9. *The probability that $[2m]^d$ is bad along with the occurrence of the event that there is exactly*

one bad subcube is correlated with double empty line segments of length $2m$*. More precisely, we have*

$$\mathbb{P}((E^d)^c \cap A_1^d) \leqslant C(d)m^{d-3}(1-p)^{4m}.$$

Proof. There is exactly one bad subcube, i.e

$$[(a_1, a_1+m-1), (a_2, a_2+m-1), ..., (a_{d+1}, a_{d+1}+m-1)],$$

where $a_i = \{1, m+1\}$ for $i \in [d+1]$ are good except for one.

W.l.o.g, we can assume that the subcube $[m]^{d+1}$ is bad.

The main idea of the proof is somewhat similar to that in the proof of Lemma 6.4.5. We will construct an event such that every vertex in $B((1,2m),d,i)$ has been infected for all $i \in \{m+2, \cdots, 2m-1\}$ after at most $m^{d+1} + C(d)m$ steps except for those vertices on the sides of $B((1,2m),d,i)$ for $i \in \{m+2, \cdots, 2m-1\}$. For the uninfected vertices on $B((1,2m),d,m+1)$ and $B((1,2m),d,2m)$ we can use induction hypothesis because every vertex in $B((1,2m),d,m+1)$ and $B((1,2m),d,2m)$ has at least one infected neighbor which does not belong to $B((1,2m),d,m+1)$ and $B((1,2m),d,2m)$. The same can be applied to infecting vertices on $B((1,2m),d,i)$ for $i \in [m]$. Then we will deal with the rest of the uninfected vertices that are not in the sides of $[2m]^{d+1}$

The exact details are presented below.

After m^{d+1} steps let us describe the uninfected vertices on $[2m]^{d+1}$. Every vertex on $B((1,2m),d,i)$ for $i \in \{m+2, m+3, \cdots, 2m-1\}$ has been infected except for those of the form $[(a_1),(a_2),\cdots,(a_d),(i)]$, where $a_j \in \{1, m, m+1, 2m\}$ for $j \in [d]$.

For the vertices on $B((1,2m),d,j)$ where $j \in \{m+1, 2m\}$, all of them have been infected except for

those on the sides of

$$[(a_1, a_1 + m - 1), (a_2, a_2 + m - 1), ..., (a_d, a_d + m - 1), (j)],$$

where $a_i = \{1, m + 1\}$ for $i \in [d]$.

For the vertices on $B((1, 2m), d, j)$ where $j \in \{1, m\}$, the vertices on $B((1, m), d, 1)$and $B((1, m), d, m)$ have not been infected. Additionally, the vertices on the edges of

$$[(a_1, a_1 + m - 1), (a_2, a_2 + m - 1), ..., (a_d, a_d + m - 1), (j)],$$

where $a_i = \{1, m + 1\}$ for $i \in [d]$ have not been infected either.

For the vertices on $B((1, 2m), d, j)$ for $j \in \{2, 3, \cdots, m - 1\}$, the vertices on $B((1, m), d, j)$ have not been infected. Additionally, the vertices on the edges of

$$[(a_1, a_1 + m - 1), (a_2, a_2 + m - 1), ..., (a_d, a_d + m - 1), (j)],$$

where $a_i = \{1, m + 1\}$ for $i \in [d]$ have not been infected either.

With the occurrence of the events A_1^{d+1} and B' then every vertex on $B((1, 2m), d, i)$ for $i \in \{m + 2, m + 3, \cdots, 2m - 1\}$ has been infected after $m^{d+1} + C(d)m$ steps. except for those belonging to the sides of $[2m]^{d+1}$ and the vertices in $[(a_1), (a_2), \cdots, (a_{i-1}), (m), (a_{i+1}), \cdots, (a_d), (1, 2m)]$ and $[(a_1), (a_2), \cdots, (a_{i-1}), (m + 1), (a_{i+1}), \cdots, (a_d), (1, 2m)]$ where $a_j \in \{1, 2m\}$ for $j \in [d]$.

Now note that every vertex in the subcube $B((1, 2m), d, m + 1)$ has at least one infected neighbor that is not on $B((1, 2m), d, m + 1)$ except for those that are in the sides of $B((1, 2m(, d, m + 1)$. So does every

vertex in the subcube $B((1,2m),d,2m)$. Therefore we can use the induction to have

$$\mathbb{P}((E^d_{m+1})^c \cap A^d_1) \leqslant C(d)m^{d-3}(1-p)^{4m},$$

and

$$\mathbb{P}((E^d_{2m})^c \cap A^d_1) \leqslant C(d)m^{d-3}(1-p)^{4m}.$$

Since every vertex in the subcube $B((1,2m),d,m)$ has at least one infected neighbor that is not on $B((1,2m),d,m)$ except for those that are in the sides of $B((1,2m(,d,m+1)$, then the induction hypothesis can be applied and we have

$$\mathbb{P}((E^d_{m})^c \cap A^d_1) \leqslant C(d)m^{d-3}(1-p)^{4m}.$$

Inductively, every vertex on $[(1,2m),(1,2m),\cdots,(1,2m),(i)]$ for $i \in \{1,2,\cdots,m-1\}$ has been infected infected except for those on the edges of $B((1,2m),d,i)$ for $i \in \{1,2,\cdots,m-1\}$ after at most $2m^{d+1} + C(d)m^2$ steps.

Now we need to deal with the uninfected vertices that are not in the sides of $[2m]^{d+1}$ after at most $2m^{d+1} + C(d)m^2$ steps. Let us describe them. For the vertices on $B((1,2m),d,i)$ for $i \in \{m+2, m+3,\cdots,2m-1\}$, the remaining uninfected vertices are on the sides of $[2m]^{d+1}$ and of the form $[(a_1),(a_2),\cdots,(a_{i-1}),(m),(a_{i+1}),\cdots,(a_d),(i)]$ and $[(a_1),(a_2),\cdots,(a_{i-1}),(m+1),(a_{i+1}),\cdots,(a_d),(i)]$ where $a_j \in \{1,2m\}$ for $j \in [d]$.

For the vertices on $B((1,2m),d,j)$ where $j \in \{m+1,2m\} \cup [m]$, those remaining uninfected vertices are on the sides of $B((1,2m),d,j)$. More precisely, for the vertices on $B((1,2m),d,j)$, where $j \in \{2,3,\cdots,m-1\}$, the remaining uninfected vertices are $((1,m),(a_2),\cdots,(a_d),i)$ where $a_i = 1$ for $i \in [d]\backslash\{1\}$ and of the form $[(a_1),(a_2),\cdots,(a_{i-1}),(m),(a_{i+1}),\cdots,(a_d),(i)]$ and $[(a_1),(a_2),\cdots,(a_{i-1}),(m+1),(a_{i+1}),\cdots,(a_d),(i)]$ where $a_j \in \{1,2m\}$ for $j \in [d]\backslash\{i\}$ and $[(a_1),(a_2),\cdots,(a_d),(i)]$ where $a_j \in$

$\{1, 2m\}$ for $j \in [d]$.

Let us consider the uninfected vertices on $\text{Perm}_{d+1}[(1, 2m), \underbrace{(1), \cdots, (1)}_{d-1}, (1, 2m)]$. W.l.o.g, it is sufficient to consider the uninfected vertices on $[(1, 2m), \underbrace{(1), \cdots, (1)}_{d-1}, (1, 2m)]$.

Let $\mathscr{A}$ be event that both line segments $[(m+1, 2m), (1), \cdots, (1), (m)]$ and $[(m), (1), (1), (m+1, 2m)]$ are non-empty. This event is the same as the event $D(y = 1)$ in the proof of Lemma 6.4.5. Therefore, the further analysis is the same as that in the proof of Lemma 6.4.

The other uninfected vertices can be handled in the same way in Lemma 6.4.8. □

Lemma 6.4.10. *The probability that $[2m]^d$ is bad along with the occurrence of the event that there are exactly two bad subcubes is correlated with double empty line segments of length $2m$. More precisely, we have*

$$\mathbb{P}((E^d)^c \cap A_2^d) \leqslant C(d)m^{d-1}(1-p)^{4m-8}.$$

Proof. There are exactly two bad subcubes, i.e

$$[(a_1, a_1 + m - 1), (a_2, a_2 + m - 1), ..., (a_{d+1}, a_{d+1} + m - 1)],$$

where $a_i = \{1, m + 1\}$ for $i \in [d + 1]$ are good except for exactly two of them.

Case 1. Assume that two bad subcubes are $[m]^{d+1}$ and

$$[(m + 1, 2m), \underbrace{(1, m), (1, m), \cdots, (1, m)}_{d}].$$

After m^{d+1} steps let us describe the uninfected vertices on $[2m]^{d+1}$.

Every vertex on $B((1, 2m), d, i)$ for $i \in \{m+2, m+3, \cdots, 2m-1\}$ has been infected except for those of the form $[(a_1), (a_2), \cdots, (a_d), (i)]$, where $a_j \in \{1, m, m + 1, 2m\}$ for $j \in [d]$.

For the vertices on $B((1,2m),d,j)$, where $j \in \{m+1,2m\}$, all of them have been infected except for those on the sides of

$$[(a_1,a_1+m-1),(a_2,a_2+m-1),...,(a_d,a_d+m-1),(j)],$$

where $a_i = \{1,m+1\}$ for $i \in [d]$.

For the vertices on $B((1,2m),d,j)$ where $j \in \{1,m\}$, the vertices on $B((1,m),d,j)$ and $[(m+1,2m),\underbrace{(1,m),\cdots,(1,m)}_{d-1},(j)]$ have not been infected. Additionally, the vertices on the sides of

$$[(a_1,a_1+m-1),(a_2,a_2+m-1),...,(a_d,a_d+m-1),(j)]$$

have not been infected, where $a_i = \{1,m+1\}$ for $i \in [d]$.

For the vertices on $B((1,2m),d,j)$ for $j \in \{2,3,\cdots,m-1\}$, the vertices on $B((1,m),d,j)$ and $[(m+1,2m),\underbrace{(1,m),\cdots,(1,m)}_{d-1},(j)]$ have not been infected. Additionally, the vertices of the form $[(a_1),(a_2),\cdots,(a_d),(j)]$ have not been infected, where $a_j \in \{1,m,m+1,2m\}$ for $j \in [d]$.

Now assume that B' occurs. Then after at most $m^{d+1}+C(d)m$ steps, every vertex on $[\underbrace{(1,2m),(1,2m),\cdots,(1,2m)}_{d},(j)]$ for $j \in \{m+2,\cdots,2m-1\}$ has been infected except possibly for those on the sides of $[2m]^{d+1}$ and are of the form $[(a_1,a_2,\cdots,a_{i-1},m,a_{i+1},\cdots,a_d,j)]$ and $[(a_1,a_2,\cdots,a_{i-1},m+1,a_{i+1},\cdots,a_d,j)]$ where $a_i \in \{1,2m\}$ for $i \in [d]$.

Therefore, every vertex on $B((1,2m),d,2m)$ and $B((1,2m),d,m+1)$ except for those on the sides of $B((1,2m),d,2m)$ and $B((1,2m),d,m+1)$ has one infected vertex that do not belong to $B((1,2m),d,2m)$ and $B((1,2m),d,m+1)$. We can use the induction to have that

$$\mathbb{P}((E^d_{m+1})^c \cap A^{d+1}_2) \leqslant C(d)(1-p)^{4m},$$

and

$$\mathbb{P}((E_{2m}^d)^c \cap A_2^{d+1}) \leqslant C(d)(1-p)^{4m}.$$

Similarly, every vertex on the subcube $B((1,2m),d,m)$ has an infected neighbor in $B((1,2m),d,m+1)$ except for those that are in the sides of $B((1,2m),d,m)$. The induction hypothesis can be applied and we have

$$\mathbb{P}(E_m^d \cap A_2^{d+1}) \leqslant C(d)m^{d-1}(1-p)^{4m-8}.$$

Similarly, then every vertex on the subcube $B((1,2m),d,m-1)$ has an infected neighbor in $B((1,2m),d,m)$ except for those that are in the sides of $B((1,2m),d,m-1)$. The induction hypothesis can be applied and we have

$$\mathbb{P}(E_{m-1}^d \cap A_2^{d+1}) \leqslant C(d)m^{d-1}(1-p)^{4m-8}.$$

Inductively, we have for $i \in [m]$

$$\mathbb{P}(E_i^d \cap A_2^{d+1}) \leqslant C(d)m^{d-1}(1-p)^{4m-8}.$$

Let us describe the possibly uninfected vertices after $2m^{d+1} + C(d)m^2$ steps.

The vertices on $B((1,2m),d,1)$ and $B((1,2m),d,2m)$ have been infected except for those on the edges of $[2m]^{d+1}$.

The vertices on $B((1,2m),d,j)$ where $j \in \{2,3,\cdots,m,m+1\}$ have been infected except for those that are of the form

$$[(a_1),(a_2),\cdots,(a_{i-1}),(1,2m),(a_{i+1}),\cdots,(a_d),(j)],$$

where $a_i = 1$ for $i \in [d]$, and of the form

$$[(a_1), (a_2), \cdots, (a_{i-1}), (m), (a_{i+1}), \cdots, (a_d), (j)],$$

$$[(a_1), (a_2), \cdots, (a_{i-1}), (m+1), (a_{i+1}), \cdots, (a_d), (j)],$$

where $a_i \in \{1, 2m\}$ for $i \in [d]$.

The vertices on $B((1, 2m), d, j)$ where $j \in \{m+2, m+3, \cdots, 2m-1\}$ have been infected except for those of the form

$$[(a_1), (a_2), \cdots, (a_{i-1}), (m), (a_{i+1}), \cdots, (a_d), (j)]$$

and

$$[(a_1), (a_2), \cdots, (a_{i-1}), (m+1), (a_{i+1}), \cdots, (a_d), (j)],$$

where $a_i \in \{1, 2m\}$ for $i \in [d]$ and $j \in \{m+2, m+3, \cdots, 2m-1\}$.

Let us consider the uninfected vertices on $\text{Perm}_{d+1}[(1, 2m), \underbrace{(1), \cdots, (1)}_{d-1}, (1, 2m)]$. Due to symmetry we only need to analyze the uninfected vertices on $[(1, 2m), \underbrace{(1), \cdots, (1)}_{d-1}, (1, 2m)]$. The analysis is in the same fashion of that in the proof of Lemma 6.4.6, specifically in the case where two bad cubes are adjacent.

The other uninfected vertices can be handled in the same way as Lemma 6.4.8.

Case 2. Assume that two bad subcubes are $[m]^{d+1}$ and $[\underbrace{(m+1, 2m), (m+1, 2m), \cdots, (m+1, 2m)}_{d+1}]$.

Let us describe the uninfected vertices after m^{d+1} steps.

Every vertex on $B((1, 2m), d, i)$ for $i \in \{m+2, m+3, \cdots, 2m-1\}$ has been infected except for those on $[(m+1, 2m), (m+1, 2m), \cdots, (m+1, 2m), (i)]$ and of the form $[(a_1), (a_2), \cdots, (a_d), (i)]$, where $a_j \in \{1, m, m+1, 2m\}$ for $j \in [d]$.

For the vertices on $[(1, 2m), (1, 2m), \cdots, (1, 2m), (j)]$, where $j \in \{m+1, 2m\}$, all of them have been

infected except for those on $[(m+1,2m),(m+1,2m),\cdots,(m+1,2m),(j)]$ and the sides of

$$[(a_1,a_1+m-1),(a_2,a_2+m-1),...,(a_d,a_d+m-1),(j)],$$

where $a_i=\{1,m+1\}$ for $i\in[d]$.

For the vertices on $B((1,2m),d,j^{'})$ where $j^{'}\in\{1,m\}$, the vertices on $B((1,m),d,j^{'})$ have not been infected. Additionally, the vertices on the sides of

$$[(a_1,a_1+m-1),(a_2,a_2+m-1),...,(a_d,a_d+m-1),(j^{'})],$$

where $a_i=\{1,m+1\}$ for $i\in[d]$ have not been infected.

For the vertices on $B((1,2m),d,j^{''})$ for $j^{''}\in\{2,3,\cdots,m-1\}$, the vertices on $B((1,m),d,j^{''})$ have not been infected. Additionally, the vertices of the form $[(a_1),(a_2),\cdots,(a_d),(j^{''})]$, where $a_j\in\{1,m,m+1,2m\}$ for $j\in[d]$, have not been infected.

Assume that the event D happens and then let us describe the uninfected vertices after $m^{d+1}+C(d)m$ steps.

Every vertex on $B((1,2m),d,i)$ for $i\in\{m+2,m+3,\cdots,2m-1\}$ has been infected except for those of the form $[(a_1),(a_2),\cdots,(a_d),(i)]$, where $a_j\in\{1,2m\}$ for $j\in[d]$. Additionally, the vertices on $[(a_1),(a_2),\cdots,a_{j-1},(1,m),a_{j+1},\cdots,(a_d),(i)]$ for $i\in\{m+2,m+3,\cdots,2m-1\}$ where $a_j=2m$ for $j\in[d]$, have not been infected.

For the vertices on $B((1,2m),d,j)$, where $j\in\{m+1,2m\}$, all of them have been infected except for those on the sides of

$$[(a_1,a_1+m-1),(a_2,a_2+m-1),...,(a_d,a_d+m-1),(j)],$$

where $a_i = \{1, m+1\}$ for $i \in [d]$.

For the vertices on $B((1,2m), d, j^{'})$where $j^{'} \in \{1, m\}$, the vertices on $[(1,m),(1,m),\cdots,(1,m),(j^{'})]$ have not been infected. Additionally, the vertices on the sides of

$$[(a_1, a_1+m-1),(a_2, a_2+m-1),...,(a_d, a_d+m-1),(j^{'})],$$

where $a_i = \{1, m+1\}$ for $i \in [d]$ have not been infected.

For the vertices on $B((1,2m), d, j^{''})$ for $j^{''} \in \{2, 3, \cdots, m-1\}$, the vertices of the form $[(a_1),(a_2),\cdots,(a_d),(j^{''})]$, where $a_j \in \{1, 2m\}$ for $j^{''} \in [d]$. Additionally, the vertices on $[(a_1),(a_2),\cdots,a_{j-1},(1,m),a_{j+1}\cdots,(a_d),(i)]$ for $i \in \{m+2, m+3, \cdots, 2m-1\}$ where $a_j = 1$, have not been infected.

Then we can apply exactly the same method from the case where there is only one bad subcube to handle the rest of the uninfected vertices.

Case 3. Assume that the two bad subcubes are $[m]^{d+1}$ and

$$[\underbrace{(m+1,2m),(m+1,2m),\cdots,(m+1,2m)}_{l},\underbrace{(1,m),(1,m),\cdots,(1,m)}_{d-l},(1,m)],$$

where $2 \leqslant l \leqslant d$.

Then let us describe the uninfected vertices after m^{d+1} steps.

Every vertex on $B((1,2m), d, i)$ for $i \in \{m+2, m+3, \cdots, 2m-1\}$ has been infected except for those of the form $[(a_1),(a_2),\cdots,(a_d),(i)]$, where $a_j \in \{1, m, m+1, 2m\}$ for $j \in [d]$.

For the vertices on $B((1,2m), d, j)$, where $j \in \{m+1, 2m\}$, all of them have been infected except for those on the sides of

$$[(a_1, a_1+m-1),(a_2, a_2+m-1),...,(a_d, a_d+m-1),(j)]$$

where $a_i = \{1, m+1\}$ for $i \in [d]$.

For the vertices on $B((1,2m), d, j^{'})$ where $j^{'} \in \{1, m\}$, the vertices on $B((1,m), d, j^{'})$ and

$$[\underbrace{(m+1,2m),(m+1,2m),\cdots,(m+1,2m)}_{l},\underbrace{(1,m),(1,m),\cdots,(1,m)}_{d-l},(j^{'})]$$

have not been infected. Additionally, the vertices on the sides of

$$[(a_1, a_1+m-1),(a_2, a_2+m-1),...,(a_d, a_d+m-1),(j^{'})]$$

where $a_i = \{1, m+1\}$ for $i \in [d]$ have not been infected.

For the vertices on $B((1,2m), d, j^{''})$ for $j^{''} \in \{2, 3, \cdots, m-1\}$, the vertices on $B((1,m), d, j^{''})$ and

$$[\underbrace{(m+1,2m),(m+1,2m),\cdots,(m+1,2m)}_{l},\underbrace{(1,m),(1,m),\cdots,(1,m)}_{d-l},(j^{''})]$$

have not been infected. Additionally, the vertices of the form $[(a_1),(a_2),\cdots,(a_d),(j^{''})]$, where $a_j \in \{1, m, m+1, 2m\}$ for $j \in [d]$, have not been infected.

Now assume that the event D' occurs. Then let us describe the uninfected vertices after $m^{d+1} + C(d)m$ steps.

Every vertex on $B((1,2m), d, i)$ for $i \in \{m+2, m+3, \cdots, 2m-1\}$ has been infected except for those of the form $[(a_1),(a_2),\cdots,(a_d),(i)]$, where $a_j \in \{1, 2m\}$ for $j \in [d]$.

For the vertices on $B((1,2m), d, j)$, where $j \in \{m+1, 2m\}$, all of them have been infected except for those on the sides of

$$[(a_1, a_1+m-1),(a_2, a_2+m-1),...,(a_d, a_d+m-1),(j)],$$

where $a_i = \{1, m+1\}$ for $i \in [d]$.

For the vertices on $B((1,2m), d, j^{'})$ where $j^{'} \in \{1, m\}$, the vertices on $[(1,m),(1,m),\cdots,(1,m),(j^{'})]$ and

$$[\underbrace{(m+1,2m),(m+1,2m),\cdots,(m+1,2m)}_{l},\underbrace{(1,m),(1,m),\cdots,(1,m)}_{d-l},(j^{'})]$$

have not been infected. Additionally, the vertices on the edges of

$$[(a_1, a_1+m-1),(a_2, a_2+m-1),...,(a_d, a_d+m-1),(j^{'})],$$

where $a_i = \{1, m+1\}$ for $i \in [d]$ have not been infected.

For the vertices on $B((1,2m), d, j^{''})$ for $j^{''} \in \{2, 3, \cdots, m-1\}$, the vertices on $B((1,m), d, j^{''})$ and

$$[\underbrace{(m+1,2m),(m+1,2m),\cdots,(m+1,2m)}_{l},\underbrace{(1,m),(1,m),\cdots,(1,m)}_{d-l},(j^{''})]$$

have not been infected. Additionally, the vertices of the form $[(a_1),(a_2),\cdots,(a_d),(j^{''})]$, where $a_j \in \{1, m, m+1, 2m\}$ for $j \in [d]$, have not been infected.

Note that after $m^{d+1} + C(d)m$ steps, every vertex on $B((1,2m), d, j)$ where $j \in \{m+1, 2m\}$ except for those on the sides of $B((1,2m), d, j)$ where $j \in \{m+1, 2m\}$ has one infected neighbor which does not belong to $B((1,2m), d, j)$ where $j \in \{m+1, 2m\}$. Thus we can use the induction hypothesis to deal with the uninfected vertices that do not belong to the sides of $B((1,2m), d, j)$ where $j \in \{m+1, 2m\}$.

Similarly for the uninfected vertices on $B((1,2m), d, j)$ where $j \in [m]$ we can use the induction hypothesis to deal with the uninfected vertices that do not belong to the sides of $B((1,2m), d, j)$ where $j \in [m]$.

For the rest of the uninfected vertices, we can use the same approach from the case where there is only one bad subcube to handle. □

Now we will use the inequality derived in Lemma 6.4.7 to derive an important property of the grid size

L, which states that the probability percolation fails to happen in a grid $[L]^d$ correlates to the probability of the existence of an empty double line segments of length L.

Before stating the following lemma, we need to define a quantity $K = K(p)$ which will be used multiple times from now on in this section. Define

$$K(p) := \begin{cases} \exp^{(d-1)}(\frac{2\lambda}{p}) & \text{if } p \leqslant p_0, \\ \exp^{(d-1)}(\frac{2\lambda}{p_0}) & \text{if } p \geqslant p_0, \end{cases}$$

where $\exp^l(\cdot)$ denotes iterating the exponential function l times and p_0 will be defined later.

Lemma 6.4.11. *If $L \geqslant 16K^3 \left(\log \frac{1}{1-p}\right)^2 \left(\frac{1}{\log \frac{1}{\delta}}\right)^2$, then the probability η_L that a grid $[L]^d$ is bad satisfies*

$$\eta_L \leqslant B(d)L^{d-1}(1-p)^{2L-8},$$

where $B(d) > 0$, and $\delta \leqslant \frac{1}{C(d)}$.

Proof. We will make use of Theorem 6.4.1. Since $p_c([n]^d, r = d) = \frac{\lambda(1+o(1))}{\log^{(d-1)} n}$, by taking p_0 small enough, we have

$$R([K]^d, r = d, p) = 1 - o(1),$$

where $R([K]^d, d, p)$ is the probability that every vertex on $[K]^d$ will be infected by the percolation process with the infection threshold $r = d$ if each vertex is initially infected with probability p.

Therefore, let $\delta > 0$ and we have

$$\eta_K \leqslant \delta.$$

From Lemma 6.4.7, we have

$$\eta_{2m} \leqslant C(d)\eta_m^3 + B(d)m^{d-1}(1-p)^{4m-8},$$

and thus

$$\eta_{2m} \leqslant 2\max\{C(d)\eta_m^3, B(d)m^{d-1}(1-p)^{4m-8}\}$$

We will use C and B to denote $C(d)$ and $B(d)$ respectively in the proof.

Note that if

$$C\eta_m^3 \leqslant Bm^{d-1}(1-p)^{4m-8}, \tag{6.6}$$

then we have

$$\eta_{2m} \leqslant Bm^{d-1}(1-p)^{4m-8},$$

and thus the desired result.

From the recursive relation $\eta_{2m} \leqslant C\eta_m^3$ with the initial condition $\eta_K \leqslant \delta$, we have

$$\eta_{2^r K} \leqslant C^{\frac{3^r-1}{2}}\delta^{3^r}.$$

In order to satisfy (6.6) it suffices to have

$$C^{\frac{3^r-1}{2}}\delta^{3^r} \leqslant B(2^{r-1}K)^{d-1}(1-p)^{4\times 2^{r-1}K-8},$$

which is equivalent to

$$\frac{3^r-1}{2}\log C - 3^r\log\frac{1}{\delta} \leqslant \log B + (d-1)\log(2^{r-1}K) - 2^{r+1}K\log\frac{1}{1-p}. \tag{6.7}$$

Assume $\delta \leqslant \frac{1}{C}$ by taking p_0 small enough. Then it is easy to see that as long as

$$\frac{3^r}{2}\log\frac{1}{\delta} \geqslant 2^{r+1}K\log\frac{1}{1-p},$$

is satisfied, (6.7) is satisfied.

Therefore, we have

$$\frac{3^r}{2^r} \geqslant 4K \left(\log \frac{1}{1-p}\right) \left(\frac{1}{\log \frac{1}{\delta}}\right),$$

which is equivalent to

$$\left(\frac{3}{2}\right)^{r \log_{\frac{3}{2}} 2} K \geqslant K \left(4K \left(\log \frac{1}{1-p}\right) \left(\frac{1}{\log \frac{1}{\delta}}\right)\right)^{\log_{\frac{3}{2}} 2}.$$

Thus if $L \geqslant K \left(4K \left(\log \frac{1}{1-p}\right) \left(\frac{1}{\log \frac{1}{\delta}}\right)\right)^{\log_{\frac{3}{2}} 2}$, then

$$\eta_L \leqslant BL^{d-1}(1-p)^{2L-8}. \quad \square$$

Now fix $L = 16K^3 \left(\log \frac{1}{1-p}\right)^2 \left(\frac{1}{\log \frac{1}{\delta}}\right)^2$.

Let us provide some explanations before stating and proving the following lemma. We would like to estimate the probability that a particular vertex, say the origin, is uninfected at time t. We show that this particular probability is roughly the same as the probability of having an empty line segment of length t starting on the origin. As mentioned in the introduction, the existence of an initially uninfected $[2t+1] \times [2]^{d-1}$ rectangle "near" the origin implies that the origin will not be infected at time t. Thus the following lemma is essentially saying that the probability from other configurations that also prevent the origin from being infected at time t is negligibly small.

Lemma 6.4.12. *Let t be an integer and $t' = L^d$. Every vertex of $[n]^d$ is initially infected with probability p independent of any other vertex. Then*

$$\mathbb{P}(\textit{the origin is uninfected at time } t) \leqslant C(d) \frac{(1-p)^{t-t'}}{p}.$$

Proof. Let us introduce some notation before proceeding. Let $[(a)^k, (b)^l] := [\underbrace{(a), \cdots, (a)}_{k}, \underbrace{(b), \cdots, (b)}_{l}]$.

Suppose that the origin is uninfected at time t. Then at least one of the vertices among $[(1), (0)^{d-1}]$, $[(0), (1), (0)^{d-2}]$,$\cdots$, and $[(0)^{d-1}, (1)]$ has to be uninfected at time $t-1$ since the origin has $2d$ neighbors and the infection threshold d. Without loss of generality if $[(1), (0)^{d-1}]$ is uninfected at time $t-1$ then one of the vertices among $[(2), (0)^{d-1}]$, $[(1), (1), (0)^{d-2}]$,$\cdots$, and $[(1), (0)^{d-2}, (1)]$ } has to be uninfected at time $t-2$ by the same reason. It is easy to see that there has to exist a path starting at the origin and moving along only in the directions of the standard basis vectors $e_1, e_2, \cdots, e_d$ of initially uninfected vertices of length t.

Therefore, there exists a path $y_1, y_2, \cdots, y_{t-t'}$, starting at the origin, of initially uninfected vertices of length $t-t'$ at time t'. Let $M > 0$ be the maximal such that at least two coordinates of $y_{t-t'-M}$ are non-zero. By definition if $M = 0$, then the path $y_1, y_2, \cdots, y_{t-t'}$ starts at the origin and keeps going straight along the standard basis vector e_1, e_2,..., or e_d. We will show that the most likely way to guarantee the origin to be uninfected at time t is to have the path $y_1, y_2, \cdots, y_{t-t'}$ starting at the origin and parallel to the standard basis vector e_1, e_2, $\cdots$, or e_d which corresponds to $M = 0$. This corresponds to the heuristic that the minimal configuration of uninfected vertices for the origin being uninfected at time t is the most likely way to guarantee the origin to be uninfected at time t.

The path $y_{t-t'-M}, \cdots, y_{t-t'}$ intersects an L-path which consists of disjoint cubes $D_1, D_2, ..., D_l$ of size $[L]^d$ with $D_1 = [(y^1_{t-t'-M}, y^1_{t-t'-M} + L - 1), (y^2_{t-t'-M}, y^2_{t-t'-M} + L - 1), ..., (y^d_{t-t'-M,}, y^d_{t-t'-M} + L - 1)]$ where $y_{t-t'-M} = (y^1_{t-t'-M}, y^2_{t-t'-M}, \cdots, y^d_{t-t'-M})$. It is easy to see that $\frac{M}{L} - 2 \leqslant l \leqslant \frac{M}{L}$.

The cubes $D_1, D_2, ..., D_l$ are either semi-good or bad since at time $t' = L^d$, $y_1, ...y_{t-t'}$ are uninfected.

Let $F_2(i, j)$ denote the event that the cubes $D_i, \cdots, D_j$ are semi-good, and that there exists a path $y_k, \cdots, y_h$ of uninfected vertices at time t', which is entirely contained in the sides of these cubes. The

initial vertex y_k lies in one of the following sets:

$$[(a_1, a_1 + L - 1), (a_2), \ldots, (a_d)], \quad [(a_1), (a_2, a_2 + L - 1), \ldots, (a_d)], \ \ldots,$$

$$\text{or} \quad [(a_1), (a_2), \ldots, (a_{d-1}), (a_d, a_d + L - 1)],$$

where

$$D_i = [(a_1, a_1 + L - 1), (a_2, a_2 + L - 1), \ldots, (a_d, a_d + L - 1)].$$

Similarly, the final vertex y_h lies in one of the following sets:

$$[(b_1 - (L - 1), b_1), (b_2), \ldots, (b_d)], \quad [(b_1), (b_2 - (L - 1), b_2), \ldots, (b_d)], \ \ldots$$

$$\text{or} \quad [(b_1), (b_2), \ldots, (b_{d-1}), (b_d - (L - 1), b_d)],$$

where

$$D_j = [(b_1 - (L - 1), b_1), (b_2 - (L - 1), b_2), \ldots, (b_d - (L - 1), b_d)].$$

If a path of uninfected vertices goes through one side $[(a_1, a_1 + l - 1), (a_2), ..., (a_d)]$ of a cube of size $[L]^d$ at time t' and the interior of this cube has been infected at time t', then the sides $[(a_1, a_1 + l - 1), (a_2), (a_3), ..., (a_d)]$, $[(a_1, a_1 + l - 1), (a_2 + 1), (a_3), ..., (a_d)]$,...,$[(a_1, a_1 + l - 1), (a_2), (a_3), ..., (a_d + 1)]$ will have to be uninfected. Since the interior of this cube has been infected by time t' every vertex on the sides has $d - 1$ infected neighbors by the time t'.

We need to give a definition before proceeding. Given a cube $D = [(a_1, b_1), (a_2, b_2), \cdots, (a_d, b_d)]$ where $b_i - a_i = m - 1$ for $i \in [d]$, let a buffer of D for the side $[(a_1), (a_2, b_2), (a_3), \cdots, (a_d)]$ be the $[2] \times [m-2] \times \underbrace{[1] \cdots \times [1]}_{d-2}$ rectangle $[(a_1 - 1, a_1), (a_2 + 1, b_2 - 1), (a_3), \cdots, (a_d)]$. Define the set of buffers

of D for the side $[(a_1), (a_2, b_2), (a_3), \cdots, (a_d)]$ to be the set

$$\{[(a_1 - 1, a_1), (a_2 + 1, b_2 - 1), (a_3), \cdots, (a_d)],$$

$$[(a_1), (a_2 + 1, b_2 - 1), (a_3 - 1, a_3), \cdots, (a_d)],$$

$$\cdots$$

$$[(a_1), (a_2 + 1, b_2 - 1), (a_3), \cdots, (a_d - 1, a_d)]\}.$$

The set of buffer for the other sides of D is defined similarly. Let $\mathbf{B}$ be the set of buffers of $D_i, \cdots, D_{i+r-1}$.

Now suppose that there is a consecutive r semi-good cubes $D_i, ..., D_{i+r-1}$ so that the event $F_2(i, i + r - 1)$ happens. Since the interior of the cubes $D_i, ..., D_{i+r-1}$ have been infected at time t', the existence of a path of uninfected vertices along the standard basis vectors $e_1, e_2, \cdots, e_d$ going along the sides of $D_i, \cdots, D_{i+r-1}$ implies that at least $r - 1$ of the buffers in $\mathbf{B}$ are uninfected. Since each buffer is a set of $d(L - 2)$ vertices, we have

$$\mathbb{P}(F_2(i, i + r - 1)) \leqslant d^{r-1}(1 - p)^{d(r-1)(L-2)},$$

If $r = 1$, we have

$$\mathbb{P}(F_2(i, i)) \leqslant \frac{2^d d}{2}(1 - p)^L.$$

Indeed since D_i is semi-good at least one of its sides is initially uninfected.

Let $F_1(i, j)$ denote the event that the cubes $D_i, ...D_j$ are bad. From Lemma 6.4.11, we have

$$\mathbb{P}(F_1(i, i + r - 1)) \leqslant (BL^d(1 - p)^{-8})^r(1 - p)^{2Lr}.$$

There exists a finite sequence $0 = b_1 < s_1 < b_2 < s_2 < b_3 < ...$, where the last term is l, such that the

event

$$F = F_1(b_1, s_1 - 1) \cap F_2(s_1, b_2 - 1) \cap F_1(b_2, s_2 - 1) \cap F_2(s_2, b_3 - 1) \cap ...$$

happens. Assume that the last term of the sequence is $l = b_{u+1} - 1$. Let v be the number of i such that $b_{i+1} - 1 = s_i$, i.e, v is the number of times that there are three consecutive cubes in the sequence $D_1, ..., D_l$ that are of the form bad, semi-good, bad. We have

$$\begin{aligned}
\mathbb{P}(F) &= \prod_{i=1}^{u} \mathbb{P}\left(F_1(b_i, s_i - 1) \cap F_2(s_i, b_{i+1} - 1)\right) \\
&\leqslant (2^{d-1}d)^v (1-p)^{Lv} \prod_{i=1}^{u} g(p)^{s_i - b_i} (1-p)^{2L(s_i - b_i)} d^{b_{i+1} - s_i - 1} (1-p)^{d(L-2)(b_{i+1} - s_i - 1)} \\
&\leqslant (2^{d-1}d)^v (1-p)^{Lv} \prod_{i=1}^{u} g(p)^{s_i - b_i} (1-p)^{2L(s_i - b_i)} d^{b_{i+1} - s_i - 1} (1-p)^{2(L-2)(b_{i+1} - s_i - 1)} \\
&\leqslant (2^{d-1}d^2 g(p))^l (1-p)^{(L-2)(2l-2u+v)},
\end{aligned}$$

where $g(p) = BL^d(1-p)^{-8}$.

Moreover, we have $2v + 3(u - v) \leqslant l$ by partitioning sequences of consecutive semi-good cubes of size $[L]^d$ into those of length 1 and those of length greater than 1. Thus we have $2u - v \leqslant \frac{2l}{3}$. Therefore,

$$\mathbb{P}(F) \leqslant (2^{d-1}d^2 g(p))^l (1-p)^{\frac{4(L-2)l}{3}}.$$

For a given l, there are d^l choices of the path along along the standard basis vector e_1, e_2,..., or e_d of cubes of size $[L]^d$ and 2^l ways of choosing whether each cube is good or bad. Therefore, if we let H be an event that there exists a path along e_1, e_2,...,or e_d direction of length M starting from a given vertex, then we have

$$\mathbb{P}(H) \leqslant (2^d d^3 g(p))^{\frac{M}{L}} (1-p)^{\frac{4M(1-\frac{2}{L})}{3}} \leqslant (2^d d^3 g(p))^{\frac{M}{L}} (1-p)^{1.3M}.$$

Therefore, with G being an event that the origin is uninfected at time t, we have

$$\mathbb{P}(G) \leqslant d \sum_{M=0}^{t-t'} (1-p)^{t-t'-M} (2^d d^3 g(p))^{\frac{M}{L}} (1-p)^{1.3M} \tag{6.8}$$

$$\leqslant \frac{C(1-p)^{t-t'}}{1-(1-p)^{0.2}} \tag{6.9}$$

$$\leqslant \frac{C(1-p)^{t-t'}}{p}. \tag{6.10}$$

The inequality 6.10 is true since $(2^d d^3 g(p))^{\frac{1}{L}} (1-p)^{0.3} \leqslant (1-p)^{0.2}$ and p_0 can be made sufficiently small. □

Now we are ready to prove the upper bound for the percolation time.

Proof. Let $t' = L^d$ and H' be an event that there exists a vertex in $[n]^d$ which is uninfected at time t. We have

$$\begin{aligned}
\mathbb{P}(T \geqslant t) &= \mathbb{P}(H') \\
&\leqslant n^d \frac{(1-p)^{t-t'}}{p} \\
&\leqslant \exp\left(d \log n + \log \frac{1}{p} - (t-t') \log \frac{1}{1-p} \right) \\
&= o(1),
\end{aligned}$$

if $\frac{d \log n}{\log \frac{1}{1-p}} \leqslant t - t'$.

Therefore, with high probability

$$T \leqslant C \frac{\log n}{\log \frac{1}{1-p}},$$

since $t' = o\left(\frac{\log n}{\log \frac{1}{1-p}} \right)$.

Thus the upper bound is proved. □

6.5 OPEN PROBLEMS

In this paper we have extended one of the two main theorems in [5] to the higher dimensional case when the initial infection probability $p(n)$ is large and the infection threshold $r = d$. There are a few questions that can be further investigated.

1. Question 1: What is the distribution of the percolation time T on $[n]^d$ if $\frac{\lambda}{\log^{d-1} n} \leqslant p(n) \leqslant \frac{C}{\log^d(n)}$ and $r = d$?

   In [5], this question was addressed for the case $d = 2$ and $r = 2$. The main difficulty in adapting the ideas from [5] to higher dimensions lies in the fact that the percolation process behaves quite differently when $d = r \geqslant 3$ compared to the two-dimensional case. For $d = r = 2$, the results in [5] rely heavily on the so-called rectangular process. In higher dimensions, however, it is unclear how to generalize this notion, and indeed, all existing results on the critical probability rely on induction on the dimension d. Nevertheless, to analyze the percolation time in higher dimensions, this inductive approach appears to be insufficient.

2. Question 2: What is the distribution of the percolation time T on $[n]^d$ if $r < d$ and $p(n) \geqslant \frac{\lambda}{\log^d(n)}$?

   This question remains largely open, and very little is known except in the case where the initially infected set is very dense, i.e., when $p(n)$ is close to 1, as studied in [18] and [19]. The first nontrivial case arises when $d = 3$ and $r = 2$.

Chapter 7

CONCLUSION AND FUTURE RESEARCH

In this thesis, we studied bootstrap percolation, motivated by its application to erasure correction in graph-based codes. In particular, we extended existing results about the critical probability of bootstrap percolation on the binary hypercube to the generalized hypercube, and investigated the behavior of the critical probability under varying infection thresholds. We also characterized the percolation time on the torus with high probability. Additionally, we determined the maximum percolation time on the q-ary hypercube. Some open questions related to our results have been listed in the end of Ch. 3 (Sec. 3.5), Ch. 4 (Sec. 4.5), Ch. 5 (Sec. 5.4), and Ch. 6 (Sec. 6.5), which can be explored in future research.

In addition to the open questions discussed in the previous chapters, another promising direction for further research is the study of graphs with strong expansion properties. The work of Balogh, Bollobas, and Morris [11] provided valuable insights by deriving the first-order term of the critical probability for the majority bootstrap percolation process on a class of regular graphs. While the conditions for that specific class of graphs were strong, their findings suggest the possibility of deriving the critical probability for the process on other classes of regular graphs as well. One such conjecture, proposed by Holmgren, Juskevicius, and Kettle [37], pertains to d-regular graphs on n vertices where $d = \log n$ and possessing strong expansion properties. The conjecture states that the critical probability for these graphs satisfies the following

expression:

$$p_c = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\log d}{d}} + o\left(\sqrt{\frac{\log d}{d}}\right).$$

Formulating this conjecture more precisely and subsequently proving it would be of significant interest.

BIBLIOGRAPHY

[1] D. Abts and J. Kim. *High Performance Datacenter Networks: Architectures, Algorithms, and Opportunities*. Morgan and Claypool, 2011.

[2] M. Aizenman and J. L. Lebowitz. Metastability effects in bootstrap percolation. *Journal of Physics A: Mathematical and General*, 21(19):3801, Oct 1988.

[3] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, and A. Hassidim. Broadcasting with side information. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 823–832, 2008.

[4] F. Arbabjolfaei and Y.-H. Kim. Fundamentals of index coding. *Foundations and Trends in Communications and Information Theory*, 14(3-4):163–346, 2018.

[5] P. Balister, B. Bollobás, and P. Smith. The time of bootstrap percolation in two dimensions. *Probab. Theory Relat. Fields*, 166:321–364, 2016.

[6] B. Balogh, Józsefand Bollobás, R. Morris, and O. Riordan. Linear algebra and bootstrap percolation. *Journal of Combinatorial Theory, Series A*, 119(6):1328–1335, 2012.

[7] J. Balogh and B. Bollobás. Sharp thresholds in bootstrap percolation. *Physica A: Statistical Mechanics and its Applications*, 326(3):305–312, 2003.

[8] J. Balogh and B. Bollobás. Bootstrap percolation on the hypercube. *Probability Theory and Related Fields*, 134:624–648, 2006.

[9] J. Balogh, B. Bollobás, H. Duminil-Copin, and R. Morris. The sharp threshold for bootstrap percolation in all dimensions. *Transactions of the American Mathematical Society*, 364, 10 2010.

[10] J. Balogh, B. Bollobás, and R. Morris. Majority bootstrap percolation on the hypercube, 2007. arXiv.math/0702373 [math.CO].

[11] J. Balogh, B. Bollobás, and R. Morris. Majority bootstrap percolation on the hypercube. *Combinatorics, Probability and Computing*, 18(1-2):17–51, 2009.

[12] J. Balogh, B. Bollobás, and R. Morris. Bootstrap percolation in high dimensions. *Combinatorics, Probability and Computing*, 19(5-6):643–692, 2010.

[13] J. Balogh, B. Bollobás, R. Morris, and O. Riordan. Linear algebra and bootstrap percolation. *Journal of Combinatorial Theory, Series A*, 119:1328–1335, 2012.

[14] Z. Baranyai. On the factorization of the complete uniform hypergraph. In *Infinite and finite sets(Colloq. Keszthely,1973: dedicated to Paul Erdös on his 60th birthday)*, volume I, pages 91–108, 1973.

[15] A. Barg and G. Zémor. High-rate storage codes on triangle-free graphs. *IEEE Transactions on Information Theory*, 68(12):7787–7797, 2022.

[16] F. Benevides and M. Przykucki. Maximum percolation time in two-dimensional bootstrap percolation. *SIAM Journal on Discrete Mathematics*, 29(1):224–251, 2015.

[17] B. Bollobás. Weakly k-saturated graphs. In *Beiträge zur Graphentheorie (Kolloquium, Manebach,1967)*, pages 25–31, Leipzig, 1968. Teubner.

[18] B. Bollobás, C. Holmgren, P. Smith, and A. J. Uzzell. The time of bootstrap percolation with dense initial sets. *The Annals of Probability*, 42(4):1337–1373, 2014.

[19] B. Bollobás, P. Smith, and A. J. Uzzell. The time of bootstrap percolation with dense initial sets for all thresholds. *Random Structures & Algorithms*, 47:1–29, 2014.

[20] P. Cameron, A. Dang, and S. Riis. Guessing games on triangle-free graphs. *The Electronic Journal of Combinatorics*, 23, 10 2014.

[21] R. Cerf and F. Manzo. The threshold regime of finite volume bootstrap percolation. *Stochastic Processes and their Applications*, 101(1):69–82, 2002.

[22] R. Cerf and E. N. M. Cirillo. Finite size scaling in three-dimensional bootstrap percolation. *The Annals of Probability*, 27(4):1837–1850, 1999.

[23] J. Chalupa, P. L.Leath, and G. R.Reich. Bootstrap percolation on a Bethe lattice. *Journal of Physics C: Solid State Physics*, 12(1):L31, Jan 1979.

[24] D. Christofides and K. Markström. The guessing number of undirected graphs. *The Electronic Jounral of Combinatorics*, 18, 09 2011.

[25] A. Coja-Oghlan, U. Feige, M. Krivelevich, and D. Reichman. Contagious sets in expanders. In *Proceedings of the 2015 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1953–1987.

[26] M. Collares, J. Erde, A. Geisler, and M. Kang. Universal behaviour of majority bootstrap percolation on high-dimensional geometric graphs. 2024. arXiv:2406.17486 [math.CO].

[27] A. E. Holroyd. Sharp metastability threshold for two-dimensional bootstrap percolation. *Probability Theory and Related Fields*, 125:195–224, 2002.

[28] U. Feige, M. Krivelevich, and D. Reichman. Contagious sets in random graphs. *The Annals of Applied Probability*, 27(5):2675–2697, 2017.

[29] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2):89–103, 1971.

[30] D. Freund, M. Poloczek, and D. Reichman. Contagious sets in dense graphs. *European Journal of Combinatorics*, 68:66–78, 2018.

[31] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc Amer Math Soc*, 124, 06 1999.

[32] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin. On the locality of codeword symbols. *IEEE Transactions on Information Theory*, 58(11):6925–6934, 2012.

[33] J. Gravner and A. E. Holroyd. Slow convergence in bootstrap percolation. *The Annals of Applied Probability*, 18(3):909–928, 2008.

[34] K. Gunderson. Minimum degree conditions for small percolating sets in bootstrap percolation. *The Electronic Jounral of Combinatorics*, 27:2, 2017.

[35] I. Hartarsky. Maximal bootstrap percolation time on hypercube via generalized snake-in-the-box. *The Electronic Journal of Combinatorics*, 25(3), 07 2018.

[36] I. Hartarsky and R. Morris. The second term for two-neighbour bootstrap percolation in two dimensions. *Transactions of the American Mathematical Society*, 372:1, 02 2019.

[37] C. Holmgren, T. Juskevicius, and N. Kettle. Majority bootstrap percolation on $G(n, p)$. *The Electronic Journal of Combinatorics*, 24(1), 2017.

[38] A. E. Holroyd, T. M. Liggett, and D. Romik. Integrals, partitions, and cellular automata. *Transactions of the American Mathematical Society*, 356(8):3349–3368, 2004.

[39] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley and Sons, Inc., New York, 2000.

[40] A. Mazumdar. On a duality between recoverable distributed storage and index coding. In *2014 IEEE International Symposium on Information Theory*, pages 1977–1981, 2014.

[41] A. Mazumdar. Storage capacity of repairable networks. *IEEE Transactions on Information Theory*, 61(11):5810–5821, 2015.

[42] N. Morrison and J. A. Noel. Extremal bounds for bootstrap percolation in the hypercube. *Journal of Combinatorial Theory, Series A*, 156:61–84, 2018.

[43] M. Przykucki. Maximal percolation time in hypercubes under 2-bootstrap percolation. *The Electronic Journal of Combinatorics*, 19(2), 06 2012.

[44] S. Riis. Information flows, graphs and their guessing numbers. In *2006 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, WiOpt 2006*, pages 1–9, 05 2006.

[45] K. Shanmugam and A. G. Dimakis. Bounding multiple unicasts through index coding and locally repairable codes. *2014 IEEE International Symposium on Information Theory*, pages 296–300, 2014.

[46] F. Zhu. Bootstrap percolation on a generalized Hamming cube, 2024. arXiv:2412.20982, submitted to *Random Structures and Algorithms*.

[47] F. Zhu. Bootstrap percolation on a generalized Hamming cube II, 2025. arXiv:2503.12607, submitted to *Random Structures and Algorithms*.

[48] F. Zhu. Maximum percolation time on the q-ary hypercube, 2025. arXiv:2503.00990, submitted to *Combinatorics, Probability, and Computing*.

[49] F. Zhu. The time of bootstrap percolation in high dimensions, 2025. arXiv:2505.11410, submitted to *Combinatorics, Probability, and Computing*.