Delft University of Technology
Faculty Eletrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Small-world and ultrasmall-world phenomena in kernel-based spatial random graphs

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

Stan Jonker

Delft, The Netherlands
March 2023

MSc THESIS APPLIED MATHEMATICS

Small-world and ultrasmall-world phenomena in kernel-based spatial random graphs

STAN JONKER

Delft University of Technology

Daily Supervisor

Dr. Júlia Komjáthy

Responsible Professor

Dr. Júlia Komjáthy

Other Thesis Committee Members

Dr. Richard Kraaij

Dr. Anurag Bishnoi

March, 2023 Delft

Small-world and ultrasmall-world phenomena in kernel-based spatial random graphs

A master thesis by Stan Jonker

Abstract

In this thesis, we examine the kernel-based spatial random graph (KSRG) model, which is a generalisation of many known models such as long-range percolation, scale-free percolation, the Poisson Boolean model and age-based spatial preferential attachment. We construct a KSRG from a vertex set $\mathcal{V} = \mathbb{Z}^d$, assigning each vertex $v \in \mathcal{V}$ a weight W_v according to a power-law with parameter $\tau - 1$ and connecting each pair of vertices conditionally independently according to

$$\mathbb{P}(u \leftrightarrow v | W_u, W_v) = \Theta\left(1 \wedge \left(\frac{\max\{W_u, W_v\}^{\sigma_1} \min\{W_u, W_v\}^{\sigma_2}}{|u - v|^d}\right)^{\alpha}\right).$$

Our first contribution is to show that, under certain choices of the parameters σ_1, σ_2, τ and α, the graph distance is at most poly-logarithmic or at most doubly logarithmic when compared to the spatial distance. Furthermore, the parameters σ_1 and σ_2 allow for extra degrees of freedom when compared to the models mentioned earlier, which yields a new exponent for poly-logarithmic distances and a generalised constant for doubly logarithmic distances. Our second contribution lies in the techniques used to prove these results. In particular, we show that with probability tending to 1 there is a subset of vertices that behaves pseudo-randomly with regards to the expected amount of vertices with a given weight in any radius. The presence of this subset, which we call a net, allows for the avoidance of the use of FKG-like inequalities in proofs. Primarily, it allows us to reveal all relevant weights in advance, which allows us to disregard many correlations that we would otherwise need to take into account during construction of paths.

Preface

This thesis was written as a final assignment of the degree Master of Science in Applied Mathematics. It has been the last task to crown five and a half years of hard work. While I am saddened that this time has come to an end, I am proud to have achieved this feat and look eagerly towards the future. During the year that I have written this thesis, I have had many ups and downs in my life, also outside of academia. The people that surround me have helped me tremendously during this time, so I would like to devote this page to thank these people.

Firstly, I would like to thank Júlia Komjáthy for supervising me during this year. She introduced me to an essentially new topic in a nice way, always gave good feedback and guidance and generally made for a really nice supervisor. If I ever decide to get back into academia, I hope we can work together again.

Then, I would like to thank my friends Ruby Brouwer, Joop Vermeulen, Thao Nguyen and Björn Titulaer for studying and writing together. They helped me stay motivated and gave me entertainment during the much-needed breaks in writing. Furthermore, I would like to thank Sanne Brouwer, who drew the front page. Of course, I also thank my other friends; while they did not directly help me with writing, they helped me in other parts of my life.

I would also like to thank my family for helping me get through the more unpleasant parts of the last year, but also just generally providing comfort. In particular, when I had to fight off a stomach infection for almost a third of the year, I could always come home when I was not feeling well.

Lastly, I want to thank you, the reader; I hope you enjoy your reading!

Stan Jonker
Delft, March 19, 2023

Contents

1. Notation

In this section, we denote some of the common notation we use throughout this thesis for later reference.

Graph Notation

Notation	Definition
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	We denote the entire graph by $\mathcal{G}$. A graph consists of a set of vertices $\mathcal{V}$ and a set of edges $\mathcal{E}$ between the vertices.
$\binom{\mathcal{V}}{2}$	We denote by $\binom{\mathcal{V}}{2}$ the set of all subsets of $\mathcal{V}$ that contain exactly 2 elements, i.e. $$\binom{\mathcal{V}}{2} := \{\{x, y\} \ : \ x, y \in \mathcal{V}, x \neq y\}.$$
$e = uv = \{u, v\}$	If the graph is *undirected*, i.e. if $\mathcal{E} \subseteq \binom{\mathcal{V}}{2}$, then any edge e between the vertices and u and v can be written as $e = \{u, v\}$. We may also use the shorthand notation $uv = \{u, v\}$.
$u \leftrightarrow v$	We denote $u \leftrightarrow v$ if u is connected to v, i.e. $uv \in \mathcal{E}$.
$u \leftrightarrow A$	If $A \subseteq \mathcal{V}$, then if there is at least one edge between u and a vertex of A, then we write $u \leftrightarrow A$.
$N_{\mathcal{G}}(u)$	For a vertex $u \in \mathcal{V}$, we use $N_{\mathcal{G}}(u)$ to denote all the vertices that are neighbours of u in $\mathcal{G}$, i.e. $$N_{\mathcal{G}}(u) = \{v \in \mathcal{V} : u \leftrightarrow v\}.$$
$d_{\mathcal{G}}(u, v)$	We denote with $d_{\mathcal{G}}(u, v)$ the *graph distance* between u and v, which is defined as the length of the shortest path between u and v in $\mathcal{G}$.

General probability notation

Notation	Definition
$X \perp Y$	If X is independent from Y, we denote this with $X \perp Y$. Here, X and Y may be either events or random variables.
With high probability	We say that a sequence of events $(E_i)_{i \in I}$ holds with high probability if $$\lim_{\substack{i \in I \\ i \to \infty}} \mathbb{P}(E_i) = 1.$$
$X \sim \mathrm{Bin}(n, p)$	If $X \sim \mathrm{Bin}(n, p)$, we say that X has a binomial distribution with parameters n and p. Then X has probability mass function $$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n$$
Stochastic domination, $X \overset{d}{\geq} Y$	We say that a random variable X stochastically dominates a random variable Y and write $X \overset{d}{\geq} Y$ iff $$\mathbb{P}(X > x) \geq \mathbb{P}(Y > x) \qquad \text{for all } x \in \mathbb{R}$$

Other notation

Notation	Definition
$\mathbb{N}, \mathbb{Z}, \mathbb{R}$	We denote the natural numbers (starting with 1) with $\mathbb{N}$, the integers with $\mathbb{Z}$ and the real numbers with $\mathbb{R}$. We also refer to the natural numbers as positive integers.
$x + A$	If $x \in \mathbb{R}$ is a real number and $A \subseteq \mathbb{R}$ is a set of real numbers, we denote $x + A$ as the set of all elements of A added to x, i.e. $$x + A = \{x + a : a \in A\}.$$
$a \wedge b, a \vee b$	For two real numbers a, b, we denote with $a \wedge b$ the minimum of a and b and with $a \vee b$ the maximum of a and b.
$[n]$	For a positive integer $n \in \mathbb{N}$, we denote with $[n]$ the set of all positive integers smaller than or equal to n, i.e. $$[n] = \{1, 2, \ldots, n-1, n\}.$$
$\{0,1\}^i$	For $i \in \mathbb{N}$, we denote $\{0,1\}^i$ as all sequences of length i with elements that are either 0 or 1. Furthermore, we denote these sequences without any punctuation, e.g. for the element $(0,1,0) \in \{0,1\}^3$ we simply write 010. If $i = 0$, then $\{0,1\}^i$ contains only an empty string.
$f(N) = o(g(N))$	We say that $f(N) = o(g(N))$ ("$f(N)$ is little o of $g(N)$") if $f(N)/g(N) \to 0$ if $N \to \infty$. In particular, we say that $f(N) = o(1)$ if $f(N) \to 0$ if $N \to \infty$.
$f(N) = \mathcal{O}(g(N))$	We say that $f(N) = \mathcal{O}(g(N))$ ("$f(N)$ is big O of $g(N)$") if there exists a constant $M > 0$ and an $\underline{N}$ such that for every $N \geq \underline{N}$ it holds that $f(N) \leq M g(N)$.
$f(N) = \Theta(g(N))$	We say that $f(N) = \Theta(g(N))$ if there exists an $\underline{N}$ and constants $C_1, C_2 > 0$ such that for all $N \geq \underline{N}$ it holds that $C_1 g(N) \leq f(N) \leq C_2 g(N)$.

2. Introduction

Over the previous years, misinformation has become a heavily discussed topic. Increasingly quickly and frequently, false information reaches a large proportion of people [45]. Once a person reads a piece of false information, they tend to spread it to their contacts. This happens either unintentionally, for example because the spreader does not know the information is false or does not understand the nuances of the subject, or intentionally, for example because it yields them social or economic benefit [44]. By repeating this process, the false information can be transmitted to a large group of people. The World Economic Forum has judged this as both a short-term and long-term global risk, regarding the severity of the problem to be similar to that of biodiversity loss and new potential pandemics [56].

Wikipedia is one of the most well-known websites in the world; according to the website ranking of similarweb it is the seventh most visited website globally [52]. The English Wikipedia consists of more than 6.6 million articles encompassing a broad selection of topics [54]. Each of these articles has its own web-page, often containing web-links to other articles. By traversing these links, it is very often possible to connect two arbitrarily chosen articles, even when the articles are about seemingly completely different topics [53]. Using this fact, the Data Science Lab at EPFL have set up 'Wikispeedia', a contest to use the fewest amount of web-links to find a path between two articles [55].

The above two examples might seem unrelated; both are, however, describable as a process on a network. In the first example, the underlying network is the social network of people and their connections, i.e., the people they know and/or can reach in some way. In the second example, we describe the entirety of Wikipedia as all its web-pages and the web-links from article to article as the connections. To understand a process on a network, it is helpful to first understand the network itself. As we will discuss in the next section, the study of networks can be done very well via mathematical modelling.

2.1. Graph theory & Network studies

Networks are a popular choice to describe and analyse many complex structures and data [18]. Any structure that can be described as a collection of elements and the connections between them, can be seen as a network. Networks are usually modelled as *graphs*. Informally, a graph is a group of vertices (also called nodes or simply 'points') and the edges between those vertices (also called links or bonds). These edges can be described in various ways. The simplest way is to just connect two vertices, in which case we call these edges *undirected*. For example, the network of people that are connected on Facebook is undirected, since if two people are connected they each appear the friend list of the other. As an antonym of undirected there is also the notion of directed edges, where an edge means that we can traverse from one vertex to another, but not necessarily backwards. An example of a network with directed edges is the network of academic papers and their citations. A paper can only cite another paper when it comes before it, so this network will have edges that can only be traversed backwards in time. There are also multigraphs — meaning that there may be multiple edges between two vertices — and hypergraphs — meaning that an edge may connect more than two vertices. Graphs that are not multi- or hypergraphs are often called simple graphs.

The start of the study of these graphs – called graph theory – is commonly attributed to Leonhard Euler and his 1736 solution of the 'Bridges of Köningsburg' problem [23, 26]. After this, graph theory has provided many real-world applications. A few examples:

- If given a distributive network where through each connection a certain amount of product can be moved per time unit, graph theory gives us a way to compute the maximal amount of product that may be moved from one point to another. A direct application of this is in cost minimisation of transport costs for large retail companies, for example.
- In any network, graph theory gives us a way to compute the shortest path between two points. This is for example used in the route planning algorithm of GPS systems.
- In a social network, graph theory gives a way to identify the most important individual(s) in the network. For example, this procedure has been used to identify the most important criminal(s) in a criminal network (See [46], pages 147-150).

The examples given above, however, are often done on relatively small networks. Due to the widespread availability of computers and communication networks, more and more data can be gathered. This allows for the analysis of networks on a scale that is much larger than previously possible. With the shift of analysing larger and larger networks also came a shift in the approach to the study of networks [23, 26]. Rather than analysing the properties of single vertices and edges, the focus was now on large-scale statistical properties of graphs. One of these reasons for this shift is that in large graphs, a single vertex or edge is usually relatively inconsequential. For example, the question "which vertex influences the network the most when removed?" is relevant in small networks (see the criminal network example above), but in large networks the removal of one vertex usually has very little effect on the network overall. However, the more statistical question "which percentage of vertices needs to be removed to significantly influence the network?" *does* have merit. A second reason is that large networks are often too hard to draw. Small networks (say tens or hundreds of vertices) may be drawn and visually inspected, but when there are millions or billions of vertices this becomes near impossible. Visual inspection is powerful tool for analysis, so its exclusion warrants other similarly powerful tools in the form of statistics.

With the introduction of statistical questions, the logical next step became to introduce randomness to model the studied networks. This is fully justified, because many real-world networks grow organically. Such procedures may well be modelled by introducing randomness [35]. This gave rise to the field of random graphs. The start of the field of random graphs is commonly attributed to the papers of Erdős and Rényi [1] and Gilbert [2], independently published in 1959. Erdős and Rényi considered a graph on n vertices and N edges chosen uniformly at random between those n vertices. Gilbert considered a graph on n vertices, and then let each of the $n(n-1)/2$ edges be present independently with probability p. While these models are interesting in their own right, they do not possess certain large-scale properties that were found in many real-world networks. These properties were found by the analysis of networks in other fields, such as physics, chemistry, biology and social psychology. We summarise some of these properties, which arguably are the most important. In particular, one property is the 'small-world phenomenon', which is very relevant to this thesis.

2.1.1. Small-world

The small-world phenomenon refers to the fact that in a large network, generally the distance between two elements in that network is relatively small. Furthermore, if the network grows, then the average distance between two elements grows several orders of magnitude slower than the growth of the network. The term 'small-world' is generally attributed to the social psychologist Stanley Milgram [8]. Milgram performed two experiments, now coined the 'small-world experiments'. In each experiment his team distributed envelopes in one city in the USA, asking people to deliver the envelope to a target person far away [4]. The catch: the person currently holding the envelope may give the envelope to someone they know. It was found that the resulting chains were relatively small, taking a median of 5 intermediate persons to reach the target person. Noting that many people likely did not choose the optimal route to reach the target person, Milgram conjectured that this chain often is not much longer than 5 intermediaries. From this, it would follow that it would take the average American only 6 links to reach (almost) every person in America. This idea is now called 'the six degrees of separation', which is now a widely used term also outside social psychology [16].

Many networks show this small-world phenomenon. For example, the network of film actors that have appeared in the same work [14], the network of authors of (mathematical) papers that have worked together [12] (see also the 'Erdős number'), the internet [20] and neural networks [10]. It has even been found that language may be constructed as a (random) network [28] which shows a small-world phenomenon [51].

In the study of the graphs that are used to model these networks, there is a precise definition of a graph being a small world. Suppose the network may be modelled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ that has $n := |\mathcal{V}|$ vertices. Then we say that $\mathcal{G}$ is a small world if the distance between two randomly chosen vertices grows proportional to $\ln(n)$ if n increases, i.e., if n increases and $u, v \in \mathcal{V}$ are uniformly randomly chosen, then

$$d_{\mathcal{G}}(u, v) \propto \ln(n). \tag{2.1}$$

Here $d_{\mathcal{G}}(u, v)$ is the graph distance between u and v, i.e., the length of the shortest path from u to v. There is also the related notion of a graph begin 'ultrasmall'. We say that $\mathcal{G}$ is an ultrasmall-world if two uniformly randomly chosen vertices $u, v \in \mathcal{V}$ satisfy

$$d_{\mathcal{G}}(u, v) \propto \ln\ln(n) \tag{2.2}$$

if n increases. We extend these definitions when the vertex set $\mathcal{V}$ is infinite and has an underlying metric space. In this case, we replace the n in (2.1) and (2.2) with the distance between u and v.

2.1.2. Scale-free property

In many networks, it was found that the vertices could have very high degree [10]. In particular, the empirical distribution of the degrees of these vertices are often thought to be modelled well by a power-law distribution with some parameter $\tau - 1$. That is, if D_v is the degree of a vertex randomly chosen vertex v, then this is thought to satisfy

$$\mathbb{P}(D_v \geq x) \propto x^{-(\tau-1)}.$$

If the vertex degrees of a network follow a power-law, we call this network *scale-free*. The name scale-free comes from the fact that the density function $f(x) \propto x^{-\tau}$ of a power-law is scale-free, i.e., for all $a > 0$ there exists a $b = b(a)$ such $f(ax) = bf(x)$.

Several real-world networks are found to be scale-free, and the literature has focused a lot on these networks [19, 22, 23]. For example, if we see the internet as a set of domains (vertices) and the links to other domains (edges), then it has been found that the degree distribution of the vertices approximately follows a power-law with parameter $\tau - 1 \approx 2.2$ (in November 1997) [13]. Other examples of scale-free networks include the network of scientific collaborations [17], the network of metabolical reactions [15], the network of protein coding genes [30] and the network of words in a language [28].

We do however give a disclaimer that the way many scale-free networks are modelled, may be too strict for many real-world networks [39]. However, other slightly less restrictive definitions have been proposed to remedy this, for example by introducing slowly varying prefactors to the density [41]. This debate about whether or not real-world networks are truly scale-free is ongoing.

2.1.3. Clustering

In a network it is often found that if vertex 1 is connected to vertex 2 and vertex 2 is connected 3, then it is more likely that vertex 1 is also connected to vertex 3 [23]. For example, in a social network, a friend of your friend is also more likely to be your friend. Similarly, if none of your friends are friends with a person, then it is less likely that this person is your friend. This behaviour is called clustering.

To analyse to which degree this clustering happens in a network, we analyse the clustering coefficient. Roughly speaking, the clustering coefficient CL is given by

$$\mathrm{CL} = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples of vertices}}.$$

Here a connected triple is a triple of vertices that is connected (see above, if 1 is connected to 2 and 2 to 3, then $\{1, 2, 3\}$ is a connected triple).

There is also a second related but distinct concept: the study of clusters. A *cluster* is simply a connected component in a graph. Examples of research on clusters is the size of the largest cluster or the second largest cluster [50]. In models with infinite vertex sets, it may also be investigated whether there exists a cluster of infinite size and if this cluster is unique.

2.2. Spatial random graphs

As already noted, the Erdős-Rényi model and Gilbert model do not satisfy all the properties given in Subsections 2.1.1-2.1.3. In particular, while both models may satisfy small-world phenomena, their degree distribution does not satisfy a power-law distribution (rather a binomial distribution with a possible Poisson limit) and clustering coefficients that tend to 0 in most cases [9]. Therefore, the search for other models that *do* satisfy these properties began. In one of these search directions, it was noted that many real-world networks have an underlying spatial structure; for example, the worldwide social network is determined heavily by worldwide topology. Because of this observation, *spatial* random graphs were introduced. In the following subsections, we discuss some examples of spatial random graphs. This list is not exhaustive. We show realisations of four of these models in Figure 1.

2.2.1. Random geometric graphs

One of the first spatial random graphs was introduced by Gilbert in 1961, which he called a random plane network [3]. In the literature, this model is now commonly known as a random geometric graph or the Gilbert disc model [24]. In this model, we let the vertices be generated by an infinite Poisson process with density λ per unit area. After this, we connect each pair of vertices if the pair is separated by a distance less than R, where R is some fixed number. A realisation of this model can be seen in Figure 1a.

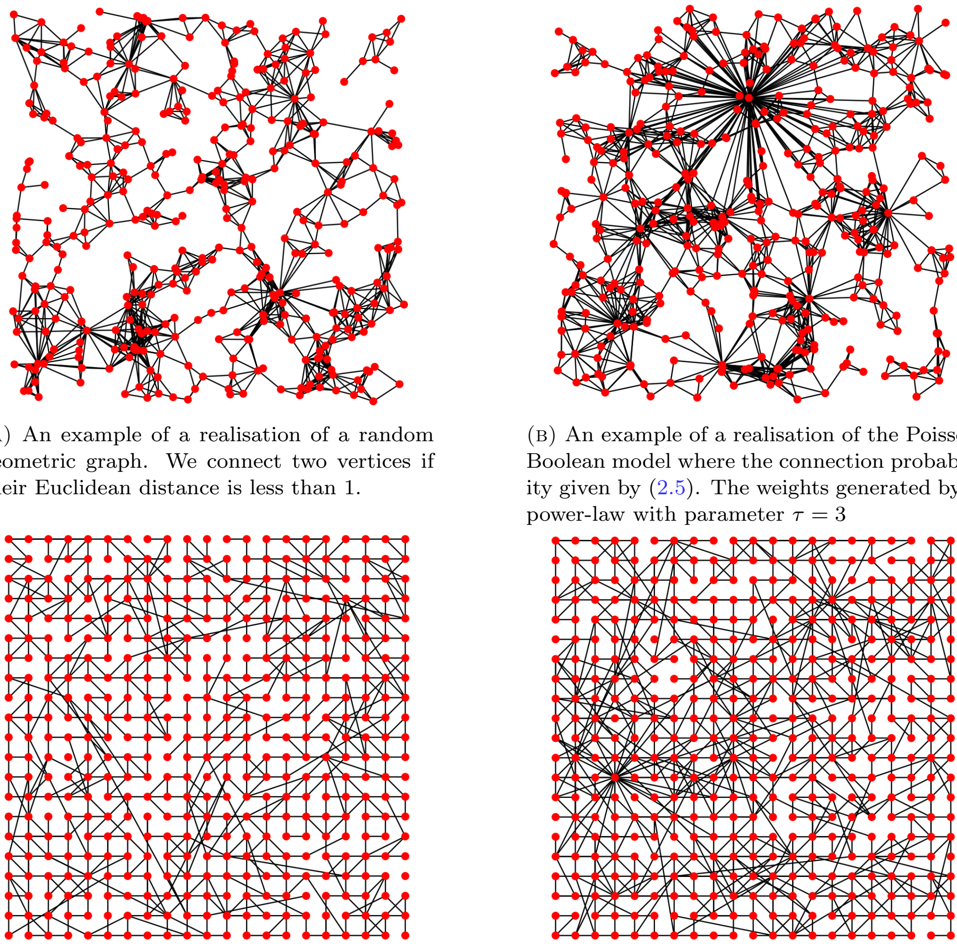
We note, however that this model does not show small-world phenomena or a power-law degree. This is easy to see:

- If u, v are such that $|u - v| =: N$, if there is a path from u to v it requires *at least* N/R edges to connect u and v, since each edge can at most of length R.
- The degree of a vertex u is given exactly by the amount of vertices in the circle of radius R around u. Since this amount of vertices is Poisson distributed, the vertex degree is also.

However, the random plane network *does* have a positive clustering coefficient [21].

2.2.2. Hyperbolic random graphs

In 2010, Krioukov et al. noted that the underlying geometry of many real-world complex networks may be described well with hyperbolic geometry [29]. In their work, Krioukov et al. considered the d-dimensional hyperbolic plane $\mathbb{H}^d$. To construct the vertex set, they place n points uniformly at random on a hyperbolic disc of radius R (where R may depend on n) and connect two vertices if their hyperbolic distance was less than R. While seemingly very similar to the random geometric model of Subsection 2.2.1, the change to hyperbolic geometry made a large difference. It was shown by Kiwi and Mitsche that if $R \propto \ln n$, the graph distance between two vertices in the same connected component is at most *poly-logarithmic*, meaning that the average distance is proportional to $\ln(n)^{\Delta}$ for some power Δ if n grows [34]. It was also shown by Gugelmann et al. that due to the hyperbolic geometry, the vertex degrees follow a power-law distribution [31]. It was shown by Krioukov et al. that this model shows a high clustering coefficient [29].



(A) An example of a realisation of a random geometric graph. We connect two vertices if their Euclidean distance is less than 1.

(B) An example of a realisation of the Poisson Boolean model where the connection probability given by (2.5). The weights generated by a power-law with parameter $\tau = 3$

(C) An example of a realisation of long-range percolation where the connection probability is given by (2.3) with $\alpha = 2$ and $\beta = 1$.

(D) An example for a realisation of scale free percolation where the connection probability is given by (2.4) with $\alpha = 2, \beta = 1$ and weights generated by a power-law with parameter $\tau = 3$.

Figure 1. Examples of a realisation of the random geometric graph, the Poisson Boolean model, long-range percolation and scale-free percolation. On each of these realisations, the vertex set is given by a Poisson Point Process with density parameter $\lambda = 1$ on $[-10, 10]^2$. Note in particular that the right-most two figures are essentially the same as the left-most two figures but with a connection probability that depends on weights. This generally yields more connections; this increase is often found in singular vertices with very high weight.

2.2.3. Long-range percolation

Initially, only one-dimensional long-range percolation was studied to solve problems in physics, such as understanding multi-spin interaction [6, 7]. However, Biskup extended this model to d dimensions in 2004 [25]. Biskup set the vertex set to be $\mathcal{V} = \mathbb{Z}^d$ and let two vertices $u, v \in \mathcal{V}$ be independently connected with probability

$$p_{u,v} = 1 - \exp\big[- \beta |u - v|^{-\alpha d} \big]. \tag{2.3}$$

Here $\beta > 0$ and $s > 0$ are parameters and $|u - v|$ is the Euclidean distance between u and v. Biskup showed that in this model the average graph distance is poly-logarithmic when $\alpha \in (1, 2)$. Furthermore, it has been established that the degrees are not scale-free [32].

A realisation of long-range percolation is given in Figure 1c.

2.2.4. Scale-free percolation

The scale-free percolation model was introduced by Maria Deijfen, Remco van der Hofstad and Gerard Hooghiemstra in 2013 as a model to study inhomogeneous long-range percolation [32]. We still consider the vertex set $\mathcal{V} = \mathbb{Z}^d$, but we also assign to each vertex $v \in \mathcal{V}$ a weight W_v according to a power-law. Then, two vertices $u, v \in \mathcal{V}$ are independently connected with probability

$$p_{u,v} = 1 - \exp\left[- \beta \frac{W_u W_v}{|u - v|^{\alpha d}} \right]. \tag{2.4}$$

It is shown by Deijfen et al. that this model shows ultrasmall-world behaviour depending on the parameters and that the vertex degrees are scale-free [32].

We note that the connection probability of (2.4) is essentially the same as the connection probability of long-range percolation (see (2.3) of Subsection 2.2.3), but with added weights. A realisation of scale-free percolation is given in Figure 1d.

2.2.5. Poisson Boolean model

The Poisson Boolean model, also called a scale-free Gilbert graph, was introduced by Christian Hirsch in 2017 [37]. To construct this graph, let $\mathcal{V}$ be a d-dimensional homogeneous Poisson Point Process and independently assign to each vertex $v \in \mathcal{V}$ a random radius R_v according to a power-law distribution with parameter $\tau - 1 > 0$. Then, for each pair of vertices $u, v \in \mathcal{V}$, connect them if $|u - v| \leq \max\{R_u, R_v\}$. Equivalently, the connection probability between u and v is given by

$$p_{u,v} = \mathbf{1}_{\{|u-v| \leq \max\{R_u, R_v\}\}}, \tag{2.5}$$

where $\mathbf{1}$ is the indicator function. It may be shown that the vertex degrees are again scale-free and show small-world phenomena with certain choices of parameters [37].

We note that the connection probability is very similar to that of a random geometric graph from Section 2.2.1, but with added weights. A realisation of this model is given in Figure 1b.

2.2.6. Geometric inhomogeneous random graphs

Introduced by Bringmann, Keusch and Lengler in 2018, geometric inhomogeneous random graphs (hereafter GIRG) is a model that shows the same qualitative behaviour as hyperbolic random graphs, but is simpler to work with [38]. GIRGs are a geometric generalisation of Chung-Lu random graphs. Briefly, a Chung-Lu random graph is a graph on n vertices (which we enumerate 1 to n), where each vertex i is assigned a weight $w_i > 0$. Then by setting $W := \sum_{i=1}^n w_i$ the total weight, a pair of vertices i, j, $i \neq j$ is independently connected with probability

$$p_{i,j} = \Theta\left(1 \wedge \frac{w_i w_j}{W}\right).$$

A GIRG generalises this notion by introducing an underlying d dimensional vector space (for example $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ [38]). We still consider the vertex set $[n] = \{1, \dots, n\}$ and the sequence of weights, but now the connection probability also depends on the spatial distance between vertices. More precisely, if we let x_i and x_j be the spatial positions of i and j respectively, then i and j are connected with probability

$$p_{i,j} = \Theta\left(1 \wedge \frac{w_i w_v / W}{|x_i - x_j|^d}\right)^{\alpha}. \tag{2.6}$$

Here, $\alpha > 0$ is some parameter that may be chosen. Bringman, Keusch and Lengler have shown that in a GIRG, if the weights $(w_i)_{i \leq n}$ are generated according to a power-law with parameter $\tau \in (2, 3)$, then the model is an ultrasmall-world [38]. In the same work, they note that while a Chung-Lu random graph does not show high clustering, a GIRG does. Furthermore, both models show power-law vertex degrees [38]. Lastly, we note that the connection probability given in (2.6) is very similar to that of scale-free percolation (see (2.4) in Subsection 2.2.4), with an extra normalising term given by the total weight W. To see this, note that $1 - \exp(x^{-1}) = \Theta(x^{-1})$.

2.2.7. Age-based spatial preferential attachment

Introduced by Gracar et al. in 2019, the age-based spatial preferential attachment model (also called the age-dependent random connection model) aims to model the way networks grow based on time and the spatial position and age of the vertices [40]. It is based on the spatial preferential attachment model introduced by Aiello et al. from 2008. In turn, the spatial preferential attachment model is based on the preferential attachment principle: vertices with high degrees are more likely to receive new edges [27]. The preferential attachment principle stems from Barabási and Albert [11].

The constructing principles of aged-based spatial preferential attachment model are to construct the network dynamically by adding nodes successively, and when a new node is introduced, it prefers to establish links to existing nods that are either old, powerful or in a similar spatial position to the new node [40]. To be precise, we construct this model in the following way:

- Start with a d-dimensional torus $\mathbb{T}_1^d = (-1/2, 1/2]^d$ endowed with metric $d(x, y) = \min\left\{|x - y + u| : u \in \{-1, 0, 1\}^d\right\}$.
- Vertices arrive according to a standard Poisson process in time and we place them independently uniformly on $\mathbb{T}_1^d$, which generates the time-dependent vertex set $\tilde{\mathcal{V}}_t$.
- Each time a vertex (x, t) arrives at time t with spatial position x, we connect it to each existing vertex $(y, s) \in \tilde{\mathcal{V}}_t$ independently with probability

$$p_{(x,t),(y,s)} = \phi\left(\frac{td(x, y)^d}{\beta(t/s)^{\gamma}}\right).$$

Here $\gamma \in (0,1), \beta \in (0,\infty)$ and $\phi : [0,\infty) \to [0,1]$ is a profile function that is non-increasing and integrates to 1. This process generates the time-dependent edge set $\tilde{\mathcal{E}}_t$.
- Then, the graph $\mathcal{G}_t$ results from rescaling all vertices from $\tilde{\mathcal{V}}_t$ by $(y,s) \mapsto (t^{1/d}y, s/t)$ and copying all edges from $\tilde{\mathcal{E}}_t$ onto the rescaled vertices.

It has been shown that the resulting graph $\mathcal{G}_t$ converges weakly locally to a graph $\mathcal{G}_\infty$ as $t \to \infty$ [40], where $\mathcal{G}_\infty = (\mathcal{V}_\infty, \mathcal{E}_\infty)$ is constructed according to

- Let $\mathcal{V}_\infty$ be given by a homogeneous Poisson Point Process with density 1 in $\mathbb{R}^d$.
- Assign to each $v \in \mathcal{V}_\infty$ a weight W_v according to a power-law with parameter $\tau - 1$, where $\tau > 2$.
- Given the weights, connect each pair of vertices $u, v \in \mathcal{V}_\infty$ independently with probability

$$p_{u,v} = \phi\left(\frac{\max\{W_u, W_v\}\min\{W_u, W_v\}^{\tau-2}}{\beta|u-v|^d}\right) \tag{2.7}$$

to construct $\mathcal{E}_\infty$.

It has also been shown that the resulting graphs $\mathcal{G}_\infty$ is scale-free and has a positive clustering coefficient [33, 40]. Furthermore, it has also been shown that the resulting graph is an ultra-small world depending on the parameters [42].

We may observe that the connection probability given in (2.7) is a modification of that of scale-free percolation (see (2.4) of Subsection 2.2.4). In particular, note that in the scale-free percolation connection probability, we may write $W_uW_v = \max\{W_u, W_v\}\min\{W_u, W_v\}$. Therefore, essentially the function $x \mapsto 1 - \exp[x]$ is replaced by ϕ and the exponent of the minimum is replaced by $\tau - 2$.

2.3. Setup of thesis

In this thesis, we investigate the small-world and ultrasmall-world properties of *kernel-based spatial random graphs* (hereafter abbreviated as KSRG(s)), which is a generalisation of many of the models described in Subsections 2.2.1-2.2.7. The most general definition of a KSRG appeared in the 2020 paper by Júlia Komjáthy and Bas Lodewijks [43] and was refined by Gracar et al. in 2022 [48]. The definition of a KSRG we use, is based on the work of Joost Jorritsma, Júlia Komjáthy and Dieter Mitsche [50], which is a reparametrisation of the definition given by Gracar et al. Another similar reparametrisation is the 'spatial inhomogeneous random graph' defined by Remco van der Hofstad, Pim van der Hoorn and Neeladri Maitra [49].

In brief, we investigate a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ such that

- $\mathcal{V} = \mathbb{Z}^d$,
- for each $v \in \mathcal{V}$, we assign a random weight W_v according to a power-law distribution with parameter $\tau - 1$, and
- each edge uv is conditionally independently present with

$$\mathbb{P}(u \leftrightarrow v | W_u, W_v) = \Theta\left(1 \wedge \left(\frac{\max\{W_u, W_v\}^{\sigma_1}\min\{W_u, W_v\}^{\sigma_2}}{|u-v|^d}\right)^\alpha\right). \tag{2.8}$$

We emphasise that the connection probability of this model generalises many of the models described in Subsections 2.2.1-2.2.7. In particular, the connection probability of long-range percolation ($\sigma_1 = \sigma_2 = 0$), scale-free percolation ($\sigma_1 = \sigma_2 = 1/\alpha$), Poisson Boolean model ($\sigma_1 = 1, \sigma_2 = 0, \alpha = \infty$) and $\mathcal{G}_\infty$ of age-based spatial preferential attachment ($\sigma_1 = 1, \sigma_2 = \tau - 2, \alpha = 1$) can be found using (2.8).

However, while the behaviour of the graph distance is known for these examples, it is not yet known for the KSRG with general parameters. My first contribution is to generalise the proofs that show these graph distances for a KSRG. This is non-trivial because of the different connection probability kernel and the fact that the presence of the additional parameters σ_1 and σ_2 allow for new phase transitions. Particularly, we show that under the assumption that $\sigma_1 = 1, \sigma_2 = \sigma \in (0,1)$, $\tau \in (2,3)$ and α such that $\tau - 1 < \alpha < (\tau-1)/(\tau-2)$ and $\alpha\sigma \leq \tau - 1$, then

$$\lim_{\substack{u,v \in \mathcal{V} \\ |u-v| \to \infty}} \mathbb{P}\big(d_{\mathcal{G}}(u,v) \leq (\ln(|u-v|))^{\Delta+\varepsilon}\big) = 1. \tag{2.9}$$

Here $\Delta = \ln(2)/\ln((\alpha + \tau - 1)/(\alpha(\tau-1)))$. We note that while this exponent Δ is different from other poly-logarithmic regimes (for example in scale free percolation the exponent is $\ln(2)/(\ln(2/\alpha))$).
Furthermore, we show that if $\sigma_1 = 1, \sigma_2 = \sigma > 0$, $\alpha > 1$ and $\tau \in (2, 2+\sigma)$, then for every δ it holds that

$$\lim_{|u-v| \to \infty} \mathbb{P}\left(d_{\mathcal{G}}(u,v) \leq \frac{2+\delta}{\ln\left(\frac{\sigma}{\tau-2}\right)} \ln\ln|u-v|\right) = 1. \tag{2.10}$$

The regimes are visualised in Figure 2.

Furthermore, we develop a new methodology that is separate of the FKG inequality or FKG-like inequalities (see [5]). These inequalities are for example used in the work of M. Biskup [25] and N. Hao and M. Heydenreich [47]. However, these inequalities are easily improperly applied, which may yield significant problems. We briefly describe such a problem with an example from the latter paper by Hao, which studies scale-free percolation. In this paper, the proof relies on the FKG-like inequality $\mathbb{P}(\pi \text{ exists}) \geq \mathbb{P}(\pi_1 \text{ exists})\mathbb{P}(\pi_2 \text{ exists})$ for any path π and any subpaths π_1, π_2 of π such that $\pi = \pi_1\pi_2$.

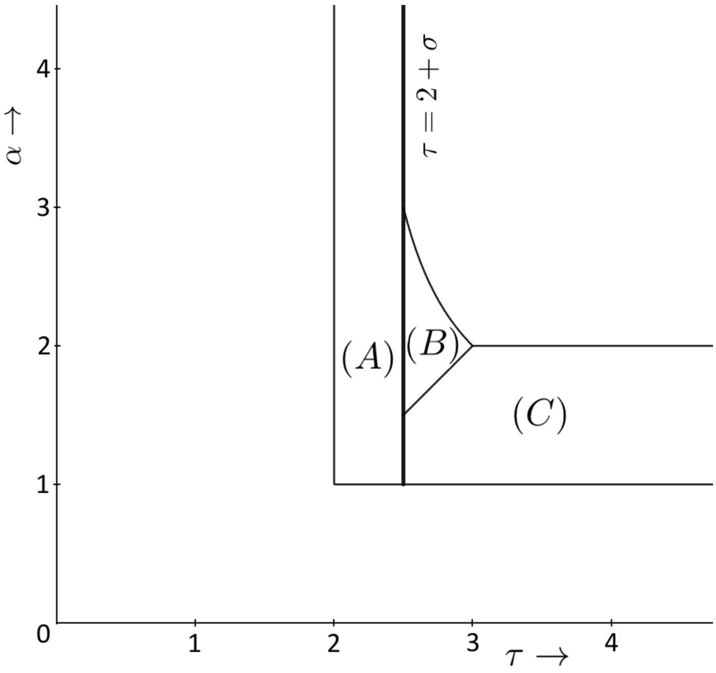


FIGURE 2. A figure showing the regimes in which different upper-bounds hold for the graph distance between two vertices. The line $\tau = 2 + \sigma$ is explicitly shown. In this image, $\sigma = 1/2$. If $\sigma > 1$, then regime (B) disappears. Region (A) is given by by $\alpha > 1$ and $\tau \in (2, 2+\sigma)$. In this region the distance between two vertices is doubly logarithmic as in (2.10). Region (B) is given by $\tau > 2 + \sigma$ and $\alpha \in (\tau - 1, (\tau - 2)/(\tau - 1)$. In this region, the distance between two vertices is poly-logarithmic as in (2.9). Note that the result in (2.9) also holds for $\tau \in (2, (2 + \sigma) \wedge 3)$, but since the doubly logarithmic upper-bound is much better than the poly-logarithmic bound, we truncate region (B). Region (C) is given by $\tau > 2+\sigma$ and $\alpha \in (1, (\tau-1)\wedge 2)$. While not proven in this thesis, the techniques used in this thesis may be used to show that in this region the distance is also poly-logarithmic with exponent $\tilde{\Delta} = \ln(2)/\ln(2/\alpha)$.

However, this holds only for paths where the spatial position of the vertices is fixed. In the proof, these paths are constructed randomly and throughout the construction, weights are revealed. These revealed weights influence the probability of the construction, which may be in a negative way. To see this, consider constructing a path in a vertex set $\tilde{\mathcal{V}}$ using a vertices with weight in some interval A. Then, the presence of the path implies the amount of vertices of $\tilde{\mathcal{V}}$ that have weight in A is likely higher, which in turn implies that the amount of vertices of $\tilde{\mathcal{V}}$ that do *not* have weight in A is likely smaller. For this reason, if we try to construct a second path using vertices that do not have weight in A, the presence of the first path will negatively affect the probability that this second construction succeeds. As such, the FKG-like inequality may not hold, or in the very least not be as obviously applied as done in [47].

My second contribution is a way to remedy the problem described above. To do this, we propose the notion of a *net*. Essentially, before we even start with constructing the paths, we reveal the weight of all relevant vertices. We then show that with high probability there is a subset of those vertices — a net — that behaves pseudo-randomly with regards to the expected amount of vertices with a given weight in any given radius. In other words, for every vertex v in a net, it holds that the amount of vertices in the net with any given weight that surround v is roughly the expected amount. Only when we have constructed such a net will we commence with the construction of the paths.

In Section 3, we give a proper definition of a KSRG and state all assumptions we use. Then, in Section 4 we give preliminary results and bounds that are useful in later proofs. After this, in Sections 5 and 6 we show that under certain choices of parameters, the graph distance is upper-bounded by a poly-logarithmic resp. doubly logarithmic bound. Furthermore, in Subsection 5.2 we also give a proper definition of the nets, as announced above. We finish with Section 7, where we discuss these results and give suggestions for further research.

3. Model definition

In this section, we define the general framework in which our results will be presented. First, we state the general definition of a kernel-based spatial random graph (KSRG) and the assumptions we place upon this graph. After this, we give a way to reparametrise the model and the assumptions we place on it.

3.1. Definition KSRG

We give the definition of a KSRG. The vertices of this graph are (possibly randomly) placed in a metric space and each given a random weight. The metric space is generally $\mathbb{R}^d$ with the standard Euclidean norm or some subset thereof, but other metric spaces such as a torus or hyperbolic space have also been considered [38, 29]. Each edge between the vertices is then randomly placed based on some function of the distance between the vertices ('spatial') and a function of the weights of the vertices ('kernel-based'). This model is based on [48] and [50].

Definition 3.1 (KSRG) *Let (M, d_M) be a metric space, P_W a probability distribution of a non-negative random variable and $p : M \times M \times \mathbb{R}_+ \times \mathbb{R}_+ \to [0, 1]$ a function that is symmetric when simultaneously switching its first two and last two components, i.e., for all $x_1, x_2 \in M, y_1, y_2 \in \mathbb{R}_+$*

$$p(x_1, x_2, y_1, y_2) = p(x_2, x_1, y_2, y_1). \tag{3.1}$$

*A **kernel-based spatial random graph** (hereafter abbreviated as **KSRG**) is an undirected simple random graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ generated by the following procedure:*

- *Let the vertex set $\mathcal{V} \subseteq M$ be some (possibly random) countable set.*
- *Assign to each $v \in \mathcal{V}$ a random weight W_v according to the distribution P_W, independently of all other vertices from $\mathcal{V}$.*
- *Let $\binom{\mathcal{V}}{2}$ be all unordered subsets of $\mathcal{V}$ containing exactly two distinct vertices. Then the **edge set** $\mathcal{E} \subseteq \binom{\mathcal{V}}{2}$ is created by inserting elements from $\binom{\mathcal{V}}{2}$ into $\mathcal{E}$ with a probability given by p. More specifically, let $u, v \in \mathcal{V}$ be any pair of vertices. We write $u \leftrightarrow v$ if $uv = \{u, v\} \in \mathcal{E}$. Given the weights $W_u = w_u, W_v = w_v$ of u and v respectively, the edge between the vertices is present with a probability given by*

$$\mathbb{P}\left(u \leftrightarrow v | \, W_u = w_u, W_v = w_v\right) = p(u, v, w_u, w_v). \tag{3.2}$$

  Furthermore, for any any subset $V \subseteq \mathcal{V}$ all edges between the vertices of V are present independently when conditioned on $(W_x)_{x \in V}$.

*We call the probability distribution P_W the **weight distribution** and p the **connectivity function**.*

We remark that the reason we require (3.1) is because the edges are undirected. In particular, this means that for all $u, v \in \mathcal{V}$ and $w_u, w_v \in \mathbb{R}_+$ we must have that

$$p(u, v, w_u, w_v) = \mathbb{P}(u \leftrightarrow v | W_u = w_u, W_v = w_v) = \mathbb{P}(v \leftrightarrow u | W_v = w_v, W_u = w_u) = p(v, u, w_v, w_u). \tag{3.3}$$

If p were not simultaneously symmetric in its first two and second two components, we would get that an undirected edge could possibly be present with two different probabilities. This obviously cannot happen.

Definition 3.1 is very broad, so to state any meaningful results we need to make assumptions. In particular, we set the metric space we work on to be $\mathbb{Z}^d$ for some $d \in \mathbb{N}$, restrict h and specify the weight distribution to be a power-law (see also Section 2.2 for justification).

Assumption 3.2 *Consider a KSRG from Definition 3.1. We fix $d \in \mathbb{N}$ and set the metric space to be $\mathbb{Z}^d$ equipped with the standard Euclidean norm $|\cdot|$. We refer to the constant d as the **dimension**. We assume that the weights follow a power-law distribution with parameter $\tau - 1 > 0$, i.e.,*

$$\mathbb{P}(W \geq w) = w^{-(\tau-1)}, \qquad w \geq 1. \tag{3.4}$$

*We refer to the parameter τ as the **power-law exponent** and $\tau - 1$ as the **tail exponent**. Next, let $\alpha, \sigma_1, \sigma_2 \geq 0$ be real-valued constants. We define $\kappa : [1, \infty) \times [1, \infty) \to [1, \infty)$,*

$$\kappa(w_u, w_v) := \max(w_u, w_v)^{\sigma_1} \min(w_u, w_v)^{\sigma_2} \tag{3.5}$$

and then use κ to define $\rho : \mathbb{R}_+ \times [1, \infty) \times [1, \infty) \to [0, 1]$,

$$\rho(|u - v|, w_u, w_v) = \min\left[1, \frac{\kappa(w_u, w_v)}{|u - v|^d}\right]^\alpha. \tag{3.6}$$

*We call κ the **weight kernel**, ρ the **regular connectivity function** and α the **long-range parameter**. Let $\underline{c}, \overline{C}$ be real-valued constants such that $0 < \underline{c} \leq \overline{C} \leq 1$. We assume that the connectivity function p from (3.2) is restricted by*

$$\underline{c}\rho(|u - v|, w_u, w_v) \leq p(u, v, w_u, w_v) \leq \overline{C}\rho(|u - v|, w_u, w_v). \tag{3.7}$$

For all $u, v \in \mathcal{V}$ and $w_u, w_v \geq 1$.

We remark that ρ is monotonically non-increasing in its first component and monotonically non-decreasing in its second and third component. Furthermore, by switching the roles of u and v,

$$\rho(|u-v|, w_u, w_v) = \rho(|v-u|, w_v, w_u).$$

This means that the lower- and upper-bound given in (3.7) does not change when switching the roles of u and v.

3.2. Reparametrising

In this subsection, we will show that the model defined by Definition 3.1 and Assumption 3.2 can be reparametrised to a form that has one fewer degree of freedom. This simplifies the proofs given in Sections 5 and 6.

Suppose that W has a power-law distribution with parameter $\tau-1$ as given in equation (3.4). Let $\eta > 0$ be some positive number. Then, by setting $\widetilde{W} = W^{\eta}$, we obtain that $\widetilde{W}$ also has a power-law distribution but with parameter $(\tau-1)/\eta$:

$$\mathbb{P}(\widetilde{W} \geq w) = \mathbb{P}(W \geq w^{1/\eta}) = w^{-(\tau-1)/\eta}, \qquad w \geq 1. \tag{3.8}$$

Now consider the regular connectivity function ρ as defined in (3.6). Provided that $\sigma_1 > 0$, we can rewrite ρ to

$$\rho(|u-v|, W_u, W_v) = \min\left[1, \frac{\max(W_u^{\sigma_1}, W_v^{\sigma_1})^1 \min(W_u^{\sigma_1}, W_v^{\sigma_1})^{\sigma_2/\sigma_1}}{|u-v|^d}\right]^{\alpha}. \tag{3.9}$$

By the observation made in (3.8), we observe that rather than parametrising the kernel with the parameters $(\sigma_1, \sigma_2, \tau-1)$, we can instead parametrise it with the parameters $(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\tau}-1) := (1, \sigma_2/\sigma_1, (\tau-1)/\sigma_1)$. Similarly, if $\sigma_1 = 0$ but $\sigma_2 > 0$, we can reparametrise the kernel with parameters $(\sigma_1, \sigma_2, \tau-1)$ to having parameters $(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\tau}-1) = (0, 1, (\tau-1)/\sigma_2)$. We call this last parametrisation the min-kernel, since the regular connectivitiy function of two vertices now only depends on the minimum of their weights. Lastly, if both $\sigma_1 = \sigma_2 = 0$, then notice that the kernel is not dependent on the weights, and in fact similar to the long-range percolation model. We summarise our findings in the following Claim 3.3.

Claim 3.3 *Consider a KSRG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from Definition 3.1 with parameters $(\sigma_1, \sigma_2, \tau, d, \alpha)$ satisfying Assumption 3.2. Then the following holds:*

(1) If $\sigma_1 > 0$, then $\mathcal{G}$ has the same distribution as $\mathcal{G}_1$, where $\mathcal{G}_1$ is a KSRG satisfying Assumption 3.2 with parameters $(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\tau}-1, d, \alpha) = (1, \sigma_2/\sigma_1, (\tau-1)/\sigma_1, d, \alpha)$.
(2) If $\sigma_1 = 0$ and $\sigma_2 > 0$, then $\mathcal{G}$ has the same distribution as $\mathcal{G}_2$, where $\mathcal{G}_2$ is a KSRG satisfying Assumption 3.2 with parameters $(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\tau}-1, d, \alpha) = (0, 1, (\tau-1)/\sigma_2, d, \alpha)$.
(3) If $\sigma_1 = \sigma_2 = 0$, then $\mathcal{G}$ has the same distribution as $\mathcal{G}_3$, where $\mathcal{G}_3$ is a KSRG satisfying Assumption 3.2 with parameters $(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\tau}-1, d, \alpha) = (0, 0, \tau', d, \alpha)$ for any $\tau' > 0$ fixed.

Considering the above, we may therefore without loss of generality assume that the parameters σ_1 and σ_2 defined in Assumption 3.2 either satisfy

- $(\sigma_1, \sigma_2) = (1, \sigma)$ for some $\sigma \geq 0$,
- $(\sigma_1, \sigma_2) = (0, 1)$, or
- $(\sigma_1, \sigma_2) = (0, 0)$.

In this thesis, we only consider the first case. We notice that in this first case, setting $\sigma = 1$ yields scale-free percolation from Subsection 2.2.4 and setting $\sigma = \tau - 2$ yields the limit of spatial preferential attachment from Subsection 2.2.7. Furthermore, we notice that the case $(\sigma_1, \sigma_2) = (0, 0)$ yields long-range percolation from Subsection 2.2.3.

4. Preliminary results

In this section we present useful bounds for the connection probability between two vertices that we use often in later sections. Furthermore, we compute bounds for the average degree of a single vertex.

4.1. Useful bounds and facts

Throughout this thesis, we often encounter connection probabilities where we do not know the weights of the vertices exactly, but we do know that they are bigger than a certain value. Similar to (3.7), we want to have a lower bound for these probabilities in terms of ρ. However, we cannot use the distribution of the weights as given in Assumption 3.2. The reason for this is that in those situations where we want to apply the bound, we also have more information that skews the distribution of the weights (such as when we know that that vertex already has an edge). The following lemma gives us this bound.

Lemma 4.1 *Let ρ be the regular connectivity function defined in* (3.6). *Consider a KSRG given by Definition 3.1 with arbitrary non-negative weight distribution function P_W and a connectivity function that satisfies* (3.7). *Let $s, t \geq 0$ and $u, v \in \mathcal{V}$ and let $\mathcal{F}$ be an event such that $\mathbb{P}(W_u \geq s, W_v \geq t|\mathcal{F}) > 0$ and for all $x \geq s, y \geq t$ it holds that*

$$\mathbb{P}(u \leftrightarrow v|W_u = x, W_y = y, \mathcal{F}) = \mathbb{P}(u \leftrightarrow v|W_u = x, W_y = y). \tag{4.1}$$

Then the following bound holds:

$$\mathbb{P}\left(u \leftrightarrow v\right| W_u \geq s, W_v \geq t, \mathcal{F}) \geq \underline{c}\rho(|u-v|, s, t). \tag{4.2}$$

Proof. We firstly bound

$$\mathbb{P}\left(u \leftrightarrow v\,|W_u \geq s, W_v \geq t, \mathcal{F}\right) \geq \inf_{\substack{x \geq s \\ y \geq t}} \mathbb{P}\left(u \leftrightarrow v\,|W_u = x, W_v = y, \mathcal{F}\right). \tag{4.3}$$

Now using (4.1) and (3.7)

$$\mathbb{P}\left(u \leftrightarrow v\,|W_u \geq s, W_v \geq t, \mathcal{F}\right) \geq \inf_{\substack{x \geq s \\ y \geq t}} \underline{c}\rho(|u-v|, x, y) = \underline{c}\rho(|u-v|, s, t), \tag{4.4}$$

where we have used that ρ is monotonically non-decreasing in both its second and third component. ■

We remark that the statement of Lemma 4.1 can easily be adapted for more general intervals than simply $[s, \infty)$ and $[t, \infty)$. Next, we elaborate the situation in which we apply Lemma 4.1. As explained, we often have more information about the distribution of the weight of u and v (as given in the lemma), which skews the distribution. The event $\mathcal{F}$ in Lemma 4.1 contains this information. However, we do require that $\mathcal{F}$ satisfies (4.1). Loosely, with this equation we demand that $\mathcal{F}$ may only affect the presence of the edge uv by affecting the distribution of W_u and W_v, but not directly. We give an example of such an $\mathcal{F}$. Suppose we know that one of the vertices is connected to other vertices, say v is connected to some vertex $a \neq u, v$. Then by Bayes' formula

$$\mathbb{P}(W_v \geq s \mid v \leftrightarrow a) = \frac{\mathbb{P}(v \leftrightarrow a \mid W_v \geq s)}{\mathbb{P}(v \leftrightarrow a)}\mathbb{P}(W_v \geq s) \tag{4.5}$$

Since in a KSRG we generally have that the weight W_v of v affects the probability that v is connected to a, we generally have that $\mathbb{P}(v \leftrightarrow a \mid W_v \geq s) \neq \mathbb{P}(v \leftrightarrow a)$. Therefore by (4.5) also $\mathbb{P}(W_v \geq s|v \leftrightarrow a) \neq \mathbb{P}(W_v \geq s)$. In fact, under Assumption 3.2 we expect to find that if s is large, $\mathbb{P}(v \leftrightarrow a \mid W_v \geq s) \gg \mathbb{P}(v \leftrightarrow a)$. Intuitively this is because by (3.7), if one of the weights of the two vertices increases, the connectivity function also increases. Therefore, if we have knowledge that suggests that one of the weights is high, we expect a higher connection probability than if we had no information about any of the weights. By (4.5), if we conversely know that an edge exists, we expect the weights of the vertices of that edge to be higher. Because of this, we therefore also expect that

$$\mathbb{P}(u \leftrightarrow v \mid v \leftrightarrow a) > \mathbb{P}(u \leftrightarrow v). \tag{4.6}$$

By similar reasoning, we may also conclude if we know that v is *not* connected to a, then the distribution of W_v skews to the lower values and hence

$$\mathbb{P}(u \leftrightarrow v \mid v \not\leftrightarrow a) < \mathbb{P}(u \leftrightarrow v). \tag{4.7}$$

We implicitly use the same reasoning as in the above example when taking certain steps in upcoming proofs. These steps may be identified by the application of Lemma 4.1. We therefore urge the reader to keep this example in mind.

Next, we expand the conditional independence that is given in Definition 3.1. Suppose we have three vertices x, y, z. Given $W_z = w_z$, the event $\{x \leftrightarrow z\}$ solely depends on W_x and the event $\{y \leftrightarrow z\}$ solely depends on W_y. Since W_x and W_y are independent, we therefore expect that given $W_z = w_z$, also $\{x \leftrightarrow z\} \perp \{y \leftrightarrow z\}$. We formalise this idea in Claim 4.2.

Claim 4.2 *Consider a KSRG given by Definition 3.1. Let $x, y, v \in \mathcal{V}$ be three distinct vertices. Then,*

$$\mathbb{P}(x \leftrightarrow v, y \leftrightarrow v \mid W_v = w_v) = \mathbb{P}(x \leftrightarrow v \mid W_v = w_v)\mathbb{P}(y \leftrightarrow v \mid W_v = w_v).$$

Proof. First, by the law of total probability we may write

$$\begin{aligned}\mathbb{P}(x \leftrightarrow v, y \leftrightarrow v | W_v = w_v) &= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{P}(x \leftrightarrow v, y \leftrightarrow v | W_x = w_x, W_y = w_y, W_v = w_v) dP_W(w_y) dP_W(w_x)\\ &=: \int_{\mathbb{R}}\int_{\mathbb{R}} q(w_x, w_y, w_x) dP_W(w_y) dP_W(w_x). \qquad (4.8)\end{aligned}$$

Next, notice that we may apply conditional independence on the integrant. Furthermore, we observe that the connection probability (3.2) only depends on the weights of the two vertices that the edge connects. As such, we may write

$$q(w_x, w_y, w_z) = \mathbb{P}(x \leftrightarrow v | W_x = w_x, W_v = w_v)\mathbb{P}(y \leftrightarrow v | W_y = w_y, W_v = w_v).$$

Substituting this into (4.8) and integrating gives the first result. ■

The proof of Lemma 4.2 suggests that this result can hold for more than just three vertices. Furthermore, the reader may verify that replacing one or both $\leftrightarrow$'s by $\not\leftrightarrow$'s, the statement is still true. We summarise these observations in the following Claim 4.3, which we give without proof.

Claim 4.3 *Consider a KSRG given by Definition 3.1. Let $v \in \mathcal{V}$ and $A, B \subseteq \mathcal{V} \setminus \{v\}$ be finite sets such that $A \cap B = \varnothing$. Then*

$$\mathbb{P}\left(\forall a \in A, b \in B : a \leftrightarrow v, b \not\leftrightarrow v \mid W_v = w_v\right) = \left[\prod_{a \in A} \mathbb{P}(a \leftrightarrow v \mid W_v = w_v)\right]\left[\prod_{b \in B} \mathbb{P}(b \not\leftrightarrow v | \ W_v = w_v)\right].$$

Proof. The proof is analogous to that of Claim 4.2. ■

We note that we cannot replace $W_v = w_v$ by $W_v \geq w_v$ in the statement of Lemma 4.3. The reason for this is that W_v is correlated with the events $x \leftrightarrow v$ and $y \leftrightarrow v$, as may may observe from the example earlier in this subsection.

4.2. Expected degrees of a single vertex

In this next section, we consider the expected degree of a single vertex. By observing that the connection probability is bounded by ρ as in (3.7) and the fact that ρ is increasing in the weight of the vertex, we would expect that the degree of a vertex is increasing in its weight. We show that if $\alpha > 1$ and $\tau > 2$, this expectation is correct. However, when $\alpha \leq 1$ or $\tau \leq 2$, we find that the expected degree is infinite. This is summarised in Claim 4.4 below.

We note that in the second item of Claim 4.4, we find that given the weight $W_v = w_v$ of a vertex v, the expected degree of v is $\Theta(w_v^{\sigma_1} + w_v^{\sigma_1+\sigma_2-(\tau-1)})$. By adapting the proof of Theorem 2.2 from [32], we may show that Claim 4.4 implies that the degree distribution (without conditioning on the weight) follows a power-law too. We do not provide a proof for this observation in this thesis.

Claim 4.4 *Consider a KSRG from Definition 3.1 satisfying Assumption 3.2 with parameters $\sigma_1, \sigma_2, \tau - 1, \alpha$ and d. For $v \in \mathcal{V} = \mathbb{Z}^d$, call $D_v := |\{uv \in \mathcal{E} | u \in \mathcal{V}\}|$ the degree of v. Then the following hold:*

(1) If $\alpha \leq 1$ or $\tau \leq 1 + \sigma_1$, then $\mathbb{E}[D_v] = \infty$.

(2) If $\alpha > 1$ and $\tau > 1 + \sigma_1$, then

$$\mathbb{E}[D_v \mid W_v = w_v] = \Theta\big(w_v^{\sigma_1} + w_v^{\sigma_1+\sigma_2-(\tau-1)}\big). \qquad (4.9)$$

Proof. We split this proof into two cases, one part where $\alpha \leq 1$ and one part where $\alpha > 1$.

(Case 1: $\alpha \in (0, 1]$) We show that in the case where $\alpha \leq 1$, every vertex almost surely has infinite degree. To do this, we apply the Borel-Cantelli lemma[1]. However, we cannot directly apply the Borel-Cantelli lemma to

$$\mathbb{P}(D_v = \infty) = \mathbb{P}\left(u \leftrightarrow v \text{ for infinitely many } u \neq v\right)$$

since the events of the form $\{u \leftrightarrow v\}$, $u \in \mathbb{Z}^d \setminus \{v\}$ are not independent[2]. However, conditioned on the weight of v, the events *are* independent (see Claim 4.3). We therefore first apply the Borel-Cantelli lemma

[1]This result can be found in many probability theory textbooks, such as 'Probability: A Graduate Course' by Allan Gut.

[2]Since $\mathbb{Z}^d$ is countable, we can enumerate all elements in $\mathbb{Z}^d \setminus \{v\}$, i.e., $\mathbb{Z}^d \setminus \{v\} = (u_n)_{n \in \mathbb{N}}$. Then

$$\begin{aligned}\{u \leftrightarrow v \text{ for infinitely many } u \neq v\} &= \{u_n \leftrightarrow v \text{ for infinitely many } n \in \mathbb{N}\}\\ &= \limsup_{n \to \infty} \{u_n \leftrightarrow v\}.\end{aligned}$$

Usually, the Borel-Cantelli lemma is formulated for the limit superior of a sequence of sets. The above shows that the formulation we use here is equivalent.

on

$$\mathbb{P}\big(D_v = \infty \mid W_v = w_v\big) = \mathbb{P}\left(u \leftrightarrow v \text{ for infinitely many } u \neq v \mid W_v = w_v\right), \tag{4.10}$$

where $w_v \geq 1$ is some fixed number. By Claim 4.3, the events $(\{u \leftrightarrow v\})_{u \in \mathbb{Z}^d \setminus \{v\}}$ are independent under $\mathbb{P}(\cdot | W_v = w_v)$. As such, by Borel-Cantelli the probability in (4.10) is 1 if and only if

$$\sum_{u \in \mathbb{Z}^d \setminus \{v\}} \mathbb{P}(u \leftrightarrow v | W_v = w_v) = \infty. \tag{4.11}$$

To show (4.11), we recall the proof of Lemma 4.1 (with $\mathcal{F}$ the entire sample space) and the fact that ρ is increasing in its second and third component to conclude

$$\mathbb{P}\left(u \leftrightarrow v | W_v = w_v\right) \geq \inf_{x \geq 1} \mathbb{P}\left(u \leftrightarrow v | W_u = x, W_v = w_v\right) \geq \underline{c}\rho(|u-v|, 1, w_v) \geq \underline{c}\rho(|u-v|, 1, 1). \tag{4.12}$$

By summing over $u \in \mathbb{Z}^d / \{v\}$ and noting that $\rho(|u-v|, 1, 1) = |u-v|^{-d\alpha}$ we obtain

$$\sum_{u \in \mathbb{Z}^d \setminus \{v\}} \mathbb{P}(u \leftrightarrow v | W_v = 1) \geq \underline{c} \sum_{u \in \mathbb{Z}^d \setminus \{v\}} \frac{1}{|u-v|^{d\alpha}} = \infty,$$

since $\sum_{u \in \mathbb{Z}^d \setminus \{v\}} |u-v|^{-p}$ diverges if $p \leq d$. We conclude that (4.11) holds and therefore that $\mathbb{P}(D_v = \infty | W_v = w_v) = 1$ for all $w_v \geq 1$ fixed. We therefore conclude that also

$$\mathbb{P}(D_v = \infty) = \int_1^\infty \mathbb{P}(D_v = \infty | W_v = w_v) dP_W(w_v) = 1.$$

Since v was arbitrarily chosen, we infer that all vertices have probability 1 to be connected to infinitely many other vertices. Lastly, we note that $\mathbb{P}(D_v = \infty) = 1$ directly implies that $\mathbb{E}[D_v] = \infty$.

(Case 2: $\alpha > 1$) We find two regimes: one where all vertices again have infinite degree and one where the vertices have finite expected degree.

Let $v \in \mathcal{V} = \mathbb{Z}^d$ be arbitrary and let $w_v \geq 1$. Below, we then apply (first line) the Monotone Convergence Theorem[3], (second line) the law of total probability, Definition 3.1 and (third line) the Fubini-Tonelli theorem

$$\begin{aligned}
\mathbb{E}[D_v | W_v = w_v] &= \sum_{\substack{u \in \mathcal{V} \\ u \neq v}} \mathbb{P}(u \leftrightarrow v | W_v = w_v) \\
&= \sum_{\substack{u \in \mathcal{V} \\ u \neq v}} \int_1^\infty p(u, v, w, w_v)(\tau - 1) w^{-\tau} dw \\
&= \int_1^\infty \sum_{\substack{u \in \mathcal{V} \\ u \neq v}} p(u, v, w, w_v)(\tau - 1) w^{-\tau} dw.
\end{aligned} \tag{4.13}$$

We want to bound the last quantity. To this end, note that by (3.7) we may bound p by $\underline{c}\rho$ and $\overline{C}\rho$. Hence, we first bound $\sum_{u \in \mathcal{V} \setminus \{v\}} \rho(|u-v|, w, w_v)$. Recall Assumption 3.2 and $\kappa(x, y) := \max(x, y)^{\sigma_1} \min(x, y)^{\sigma_2}$ from (3.5). We rewrite

$$\sum_{\substack{u \in \mathcal{V} \\ u \neq v}} \rho(|u-v|, w, w_v) = \sum_{\substack{u \in \mathcal{V} \\ 1 \leq |u-v|^d \leq \kappa(w, w_v)}} 1 + \sum_{\substack{u \in \mathcal{V} \\ |u-v|^d > \kappa(w, w_v)}} \frac{\kappa(w, w_v)^\alpha}{|u-v|^{d\alpha}}. \tag{4.14}$$

Then, we use that there are constants $\underline{\nu}_d, \overline{\nu}_d > 0$ such that

$$\underline{\nu}_d \kappa(w_0, w) \leq \left| \left\{ u \in \mathbb{Z}^d \ : \ 1 \leq |u-v|^d \leq \kappa(w_0, w) \right\} \right| \leq \overline{\nu}_d \kappa(w_0, w), \tag{4.15}$$

which can be seen by relating the amount of points between the two radii with the volume of the hollow sphere spanned by the two radii. Next, we use that we can approximate a sum by an integral, which gives us that there are constants $\underline{c}_d, \overline{c}_d > 0$ that do not depend on w or w_v such that

$$\underline{c}_d \int_{\left\{x \in \mathbb{R}^d : |x|^d > \kappa(w, w_v)\right\}} \frac{1}{|x|^{d\alpha}} dx \leq \sum_{\substack{u \in \mathcal{V} \\ |u-v|^d \leq \kappa(w, w_v)}} \frac{1}{|u-v|^{d\alpha}} \leq \overline{c}_d \int_{\left\{x \in \mathbb{R}^d : |x|^d > \kappa(w, w_v)\right\}} \frac{1}{|x|^{d\alpha}} dx \tag{4.16}$$

to state that

$$\frac{\underline{c}_d}{\kappa(w, w_v)^{\alpha - 1}} \leq \sum_{|u-v|^d > \kappa(w, w_v)} \frac{1}{|u-v|^{d\alpha}} \leq \frac{\overline{c}_d}{\kappa(w, w_v)^{\alpha - 1}}. \tag{4.17}$$

Here we have used that $\alpha > 1$ so the integral from (4.16) converges and can be computed.
We now combine the bounds from equation (3.7) with the above equations (4.13), (4.14), (4.15) and (4.17) to conclude that there are constants $\underline{C}_1, \underline{C}_2, \overline{C}_1, \overline{C}_2$ such that

$$\mathbb{E}[D_v | W_v = w_v] \leq \overline{C}(\overline{\nu}_d + \overline{c}_d)(\tau - 1) \int_1^\infty \kappa(w, w_v) w^{-\tau} dw \begin{cases} \leq \overline{C}_1 w_v^{\sigma_1} + \overline{C}_2 w_v^{\sigma_1 + \sigma_2 - \tau + 1} & \text{if } \tau - 1 > \sigma_1 \\ = \infty & \text{if } \tau - 1 \leq \sigma_1 \end{cases},$$

[3] This result can be found in measure-theoretic probability or real analysis textbooks, such as 'Real Analysis' by N. L. Carothers.

and in in exactly the same way

$$\mathbb{E}[D_v|W_v = w_v] \geq \underline{c}(\underline{\nu}_d + \underline{c}_d)(\tau - 1)\int_1^\infty \kappa(w, w_v)w^{-\tau}dw \begin{cases} \geq \underline{C}_1 w_v^{\sigma_1} + \underline{C}_2 w_v^{\sigma_1+\sigma_2-\tau+1} & \text{if } \tau - 1 > \sigma_1 \\ = \infty & \text{if } \tau - 1 \leq \sigma_1 \end{cases}.$$

Here we have used that

$$\int_1^\infty \kappa(w_v, w)w^{-\tau}dw = \int_1^{w_v} w_v^{\sigma_1} w^{\sigma_2-\tau}dw + \int_{w_v}^\infty w^{\sigma_1-\tau}w_v^{\sigma_2}dw$$
$$= \left(\frac{1}{\sigma_2 - \tau + 1} - \frac{1}{\sigma_1 - \tau + 1}\right) w_v^{\sigma_1+\sigma_2-\tau+1} + \frac{1}{\tau - 1 - \sigma_2} w_v^{\sigma_1} + \begin{cases} 0 & \text{if } \sigma_1 < \tau - 1 \\ \infty & \text{if } \sigma_1 \geq \tau - 1 \end{cases}.$$

One may verify that the constant before the dominant term (i.e., $w_v^{\sigma_1}$ if $\sigma_1 > \sigma_1 + \sigma_2 - (\tau - 1)$ and $w_v^{\sigma_1+\sigma_2-(\tau-1)}$ otherwise) is always positive. We conclude that (4.9) holds, which is what remained to show. ■

5. Poly-logarithmic upper-bound for distances

In this section, we show that with the right parameters, a KSRG satisfying Assumption 3.2 with high probability the graph distance between two vertices is poly-logarithmic in their spatial distance. More specifically, we show that if $u, v \in \mathcal{V} = \mathbb{Z}^d$, then there exists a Δ such that for every $\varepsilon > 0$, it holds that $d_{\mathcal{G}}(u,v) \le (\ln|u-v|)^{\Delta+\varepsilon}$ with high probability as $|u-v| \to \infty$. In particular, we show Theorem 5.1. When the parameters of the KSRG are such that the graph-distances are poly-logarithmic in the spatial distance, we refer to these parameters as being a *poly-logarithmic regime* (or 'polylog regime' for short).

Theorem 5.1 *Consider a KSRG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from Definition 3.1 satisfying Assumption 3.2. Assume that $\mathcal{V} = \mathbb{Z}^d$ and that the KSRG has parameters $d \in \mathbb{N}$, $\sigma_1 = 1$, $\sigma_2 = \sigma \in (0,1)$, $\tau \in (2,3)$ and α such that $\tau - 1 < \alpha < (\tau-1)/(\tau-2)$ and $\alpha\sigma \le \tau - 1$. Furthermore, assume that all nearest-neighbour edges are present in $\mathcal{E}$. Let $\varepsilon > 0$ and set*

$$\Delta = \frac{\ln 2}{\ln\left(\frac{\alpha+\tau-1}{\alpha(\tau-1)}\right)}. \tag{5.1}$$

Then it holds that

$$\lim_{\substack{u,v\in\mathcal{V}\\ |u-v|\to\infty}} \mathbb{P}\big(d_{\mathcal{G}}(u,v) \le (\ln(|u-v|))^{\Delta+\varepsilon}\big) = 1. \tag{5.2}$$

To prove this theorem, we use the following Proposition 5.2.

Proposition 5.2 *Consider a KSRG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from Definition 3.1 satisfying Assumption 3.2. Assume that $\mathcal{V} = \mathbb{Z}^d$ and that the KSRG has parameters $d \in \mathbb{N}$, $\sigma_1 = 1$, $\sigma_2 = \sigma \in (0,1)$, $\tau \in (2,3)$ and α such that $\tau - 1 < \alpha < (\tau-1)/(\tau-2)$ and $\alpha\sigma \le \tau - 1$. Furthermore, assume that all nearest-neighbour edges are present in $\mathcal{E}$. Let $\varepsilon > 0$ and set*

$$\Delta = \frac{\ln 2}{\ln\left(\frac{\alpha+\tau-1}{\alpha(\tau-1)}\right)}. \tag{5.3}$$

Let $u, v \in \mathcal{V}$. Then, there exists an $\underline{N}_{5.2}$ and a function $\mathrm{err}_{5.2}(|u-v|) = \mathrm{err}_{5.2}(|u-v|, \varepsilon, \alpha, \tau, d)$ that goes to 0 if $|u-v| \to \infty$, such that if $|u-v| \ge \underline{N}_{5.2}$, then

$$\mathbb{P}\big(d_{\mathcal{G}}(u,v) \le (\ln|u-v|)^{\Delta+\varepsilon}\big) \ge 1 - \mathrm{err}_{5.2}(|u-v|, \varepsilon, \alpha, \tau, d). \tag{5.4}$$

We note that given that Proposition 5.2 holds, Theorem 5.1 follows directly. Therefore, the remainder of this section is devoted to proving Proposition 5.2. To start, we firstly give the full setting of the proof to be used as reference. This contains the description of the KSRG and three additional parameters used throughout the proof.

Setting 5.3 *Consider a KSRG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from Definition 3.1 satisfying Assumption 3.2. Assume it has parameters $d \in \mathbb{N}, \sigma_1 = 1, \sigma_2 = \sigma \in [0,1), \tau \in (2,3)$ and α such that $\alpha > \tau - 1, \alpha\sigma \le \tau - 1$ and $\alpha < (\tau-1)/(\tau-2)$. Furthermore, fix γ, δ and η satisfying*

$$\gamma \in \left(\frac{\alpha(\tau-1)}{\alpha+\tau-1}, 1\right) =: I_\gamma, \qquad \delta \in \left(0, \frac{1-\gamma}{1+2\gamma}\right) =: I_\delta \qquad \textit{and} \qquad \eta \in \left(0, \frac{2}{3}\right). \tag{5.5}$$

Before we continue with the proof of Proposition 5.2, we discuss Setting 5.3. Firstly, we reason that we may indeed select α as described. To this end, notice that since $\tau \in (2,3)$, we also find that $(\tau-1)/(\tau-2) > \tau-1$. If $\sigma = 0$, then $0 = \alpha\sigma \le \tau - 1$ is always satisfied when $\tau \in (2,3)$. Hence, the restrictions on α may be rewritten as $\tau - 1 < \alpha < (\tau-1)/(\tau-2)$, which is clearly possible. Furthermore, if $\sigma \ne 0$, then we may rewrite

$$\alpha\sigma \le \tau - 1 \qquad \Leftrightarrow \qquad \alpha \le \frac{\tau-1}{\sigma}. \tag{5.6}$$

We find that the restrictions on α can be rewritten as

$$\tau - 1 < \alpha < \frac{\tau-1}{\sigma} \wedge \frac{\tau-1}{\tau-2}. \tag{5.7}$$

Since $\tau - 2 < 1$ and $\sigma < 1$, selecting an α that satisfies (5.7) is possible. From (5.7) we also notice that $\sigma = 1$ is not possible.

We continue with arguing that I_γ and I_δ are non-empty, which allows us to indeed fix γ and δ as described. To this end, notice that by rewriting

$$\alpha < \frac{\tau-1}{\tau-2} \qquad \Leftrightarrow \qquad \alpha(\tau-1) < \alpha + \tau - 1. \tag{5.8}$$

From the right-most inequality of (5.8) it immediately follows that I_γ is non-empty. Lastly, because γ is clearly positive and $\gamma < 1$, it also immediately follows that the upper end-point of I_δ is strictly positive. We therefore conclude that I_δ is indeed non-empty.

In the proof, we eventually choose γ as small as possible. For this reason, we examine the lower end-point of I_γ. To this end, set $\gamma_-(\alpha,\tau) = \alpha(\tau-1)/(\alpha+\tau-1)$. We examine γ_- on the domain

$$\mathscr{D} = \{\tau \in (2,3), \alpha > \tau-1, \alpha < (\tau-1)/(\tau-2), \alpha\sigma \le \tau-1\}. \tag{5.9}$$

We observe γ_- is continuously differentiable on $\mathscr{D}$ and therefore achieves its minimum either in a stationary point or on the boundary of $\mathscr{D}$. Next, we may observe that γ_- is strictly increasing in both of its variables, so it does not have any stationary points in $\mathscr{D}$. Furthermore, the fact that γ_- is strictly increasing in both variables suggests that

$$\inf_{\alpha,\tau\in\mathscr{D}} \gamma_-(\alpha,\tau) = \lim_{\substack{\alpha,\tau\in\mathscr{D}\\ \alpha\downarrow 1,\tau\downarrow 2}} \frac{\alpha(\tau-1)}{\alpha+\tau-1} = \frac{1}{2}. \tag{5.10}$$

From this, we may also conclude that

$$\gamma > \frac{1}{2}, \qquad \text{and hence} \qquad \delta < \frac{1}{4}. \tag{5.11}$$

5.1. Idea of the proof of Proposition 5.2

We continue with a sketch of the proof of Proposition 5.2 to motivate the steps we take in the actual proof. Visually, the sketch of the proof is given in Figure 3. Throughout this sketch, $C_1, C_2, \dots > 0$ denote strictly positive constants and let $\Theta(f(N))$ be as defined in Section 1. We want to connect two vertices $u, v \in \mathcal{V}$ that are distance $N := |u-v|$ apart. Before we start, we call $u = z_0$ and $v = z_1$ to give structure to the iteration scheme that follows (this may be viewed as the 'zeroth step'). In the first step, we then elongate the labels of $z_0 = z_{00}$ and $z_1 = z_{11}$. Then we consider a box $B^1(z_{00})$ around z_{00} and a box $B^1(z_{11})$ around z_{11} with sidelengths of size N^{γ^1}. Note that this for this reason we need $\gamma < 1$, since we want the boxes to shrink rather than grow. We then search for a vertex z_{01} in the box around z_{00} and a vertex z_{10} in the box around z_{11} such that $z_{01} \leftrightarrow z_{10}$, where z_{01} has weight $\Theta(1)$ and z_{10} has weight $\Theta\big(N^{d\gamma(1-\delta)/(\tau-1)}\big)$. We refer to the first vertex z_{01} as a 'constant-weight' vertex and denote all constant weight vertices with a green circle ●. The second vertex z_{10} is a 'high-weight' vertex, which we denote with a green square ■. We show that with high probability we may choose z_{01} and z_{10} as described above. To this end, notice that there are *no* vertices that satisfy the description of z_{01} and z_{10} exactly when all constant weight vertices in $B^1(z_{00})$ are not connected to any high-weight vertex of $B^1(z_{11})$. We bound the probability of the last event and show that it goes to 0 if $N \to \infty$. To this end, notice that there are roughly $N^{d\gamma}$ vertices in both $B^1(z_{00})$ and $B^1(z_{11})$. By the distribution of the weights from (3.4), we may obtain that

$$\mathbb{P}(W = \Theta(1)) = \Theta(1) \qquad \text{and} \tag{5.12}$$

$$\mathbb{P}\big(W = \Theta\big(N^{d\gamma(1-\delta)/(\tau-1)}\big)\big) = \Theta\big(N^{-d\gamma(1-\delta)}\big). \tag{5.13}$$

We conclude that there are roughly $\Theta(1)\cdot N^{d\gamma} = \Theta\big(N^{d\gamma}\big)$ constant weight vertices in the box around z_{00} and roughly $\Theta\big(N^{d\gamma(1-\delta)}\big)\cdot N^{d\gamma} = \Theta\big(N^{d\gamma\delta}\big)$ high-weight vertices in the box around z_{11}. Furthermore, notice that if $x \in B^1(z_{00})$ and $y \in B^1(z_{11})$, then $|x-y| = \Theta(N)$. Furthermore, if x is such that $W_x = \Theta(1)$ and $y = \Theta\big(N^{d\gamma(1-\delta)/(\tau-1)}\big)$, then by the assumptions of the connection probability given in (3.7), we have

$$\mathbb{P}(x \leftrightarrow y \mid W_x, W_y) \ge \underline{c}\rho(|x-y|, W_x, W_y) \ge \underline{c}\left(1 \wedge \frac{\Theta\big(N^{d\gamma\frac{1-\delta}{\tau-1}}\big)\Theta(1)^\sigma}{\Theta(N)^d}\right)^\alpha = \Theta\Big(N^{0\wedge d\gamma\alpha\left(\frac{1-\delta}{\tau-1}-\frac{1}{\gamma}\right)}\Big), \tag{5.14}$$

where ρ is from (3.6). Next, we examine the exponent of the last quantity of (5.14). Since $\gamma < 1, \tau-1 \in (1,2)$ by Setting 5.3 and $\delta \in (0,1/4)$ by (5.11), we find that $(1-\delta)/(\tau-1) - 1/\gamma < 0$. Hence from we find that $\mathbb{P}(x \leftrightarrow y) \ge \Theta\big(N^{d\gamma\alpha((1-\delta)/(\tau-1)-1/\gamma)}\big)$. Since there are $\Theta\big(N^{d\gamma}\big)$ constant-weight vertices in $B^1(z_{00})$ and roughly $\Theta\big(N^{d\gamma\delta}\big)$ high-weight vertices in $B^1(z_{11})$, the we find that (if we ignore dependence)

$$\begin{aligned}
&\mathbb{P}(x \not\leftrightarrow y \text{ for all constant weight vertices } x \in B^1(z_{00}) \text{ and high-weight vertices } y \in B^1(z_{11}))\\
&\approx \prod_{i=1}^{\Theta(N^{d\gamma})} \prod_{j=1}^{\Theta(N^{d\gamma\delta})} (1 - \mathbb{P}(x \leftrightarrow y)) \le \Big(1 - \Theta\Big(N^{d\gamma\alpha\left(\frac{1-\delta}{\tau-1}-\frac{1}{\gamma}\right)}\Big)\Big)^{\Theta(N^{d\gamma(1+\delta)})}\\
&\le \exp\Big[-\Theta\Big(N^{d\gamma\left(1+\delta+\alpha\frac{1-\delta}{\tau-1}-\frac{\alpha}{\gamma}\right)}\Big)\Big],
\end{aligned} \tag{5.15}$$

where we have used that $1-x \le e^{-x}$. To ensure that z_{01} and z_{10} exist, we want that (5.15) is small, in particular if N is large. Therefore, we examine the exponent of N in the right-most quantity of (5.15). Particularly, we want to show that

$$R := 1 + \delta + \alpha\frac{1-\delta}{\tau-1} - \frac{\alpha}{\gamma} > 0. \tag{5.16}$$

To this end, we rewrite to see that (5.16) holds if and only if

$$\gamma > \frac{\alpha}{(1+\delta) + \alpha\frac{1-\delta}{\tau-1}} = \frac{\alpha(\tau-1)}{\alpha+\tau-1+\delta(\tau-1-\alpha)} > \frac{\alpha(\tau-1)}{\alpha+\tau-1}. \tag{5.17}$$

Here we have used that $\alpha > \tau-1$, so that $\tau-1-\alpha < 0$. By definition of γ from Setting 5.3, (5.17) is true, which makes (5.16) true, which implies that the right-most quantity of (5.15) goes to 0 if $N \to \infty$. We conclude that with high probability, we may indeed find z_{01} and z_{10} as described. The edge between

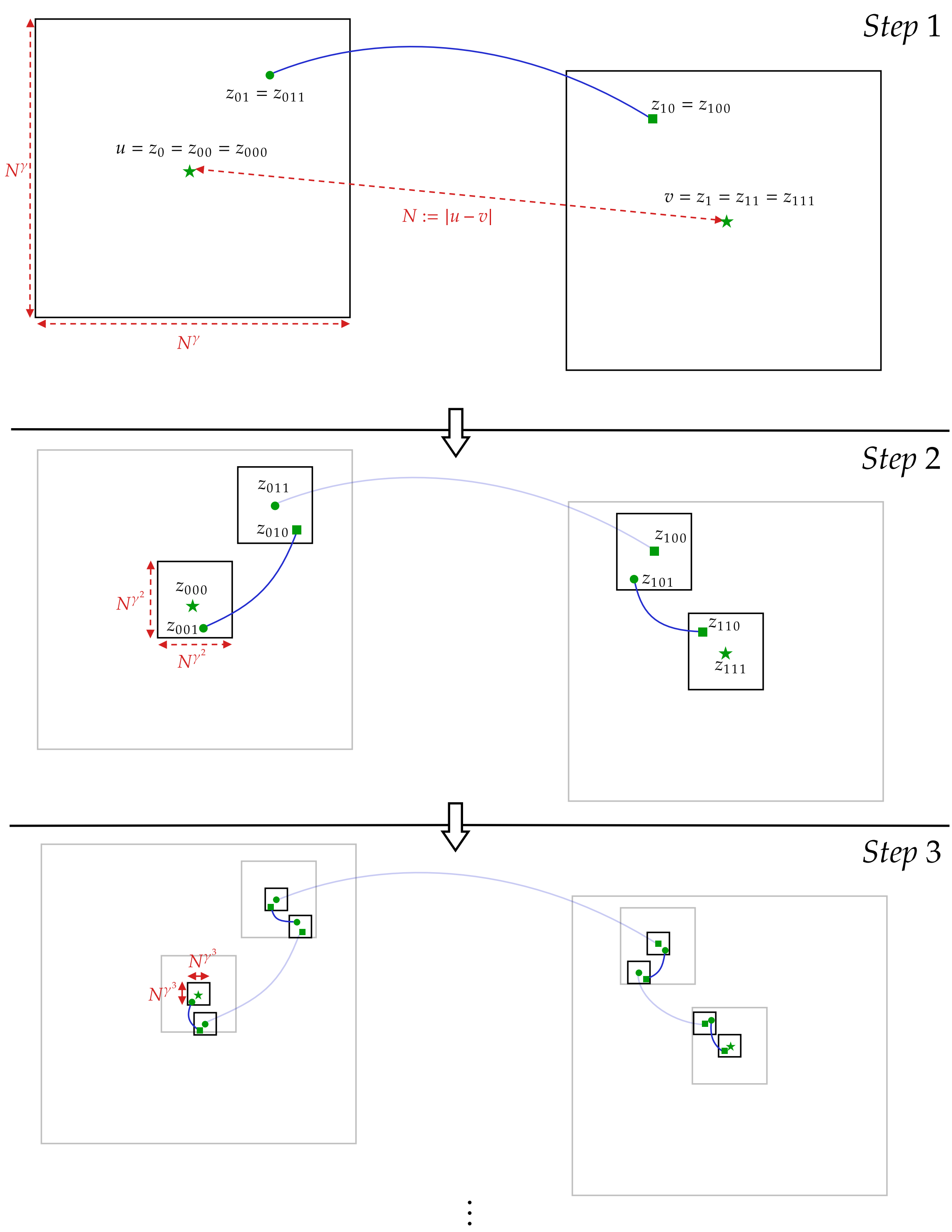


FIGURE 3. A figure showing the rough idea behind the proof of Proposition 5.2 in $d = 2$. This image is not to scale. Red arrows denote distances, green icons (i.e., $\star$, ■ and ●) denote vertices, blue curved lines denote edges and black-bordered squares denote box-shaped sets of vertices. Due to legibility, we omit labelling in the last step shown.

these two vertices is given in blue in Figure 3.

In step 2, we enhance the labelling: each label that ended in a 0 now ends in 00 and each vertex that ended in a 1 now ends in 11 (e.g. $z_0 = z_{00} = z_{000}$ and $z_{01} = z_{011}$) until their label is 3 characters long. Then around each vertex we constructed in step 1, we consider a box of size N^{γ^2} which we denote by B^2. In the same way as above, we now find an edge $z_{001}z_{010}$ between the box $B^2(z_{000})$ centred around $z_{00} = z_{000}$ and the box $B^2(z_{011})$ centred around $z_{01} = z_{011}$ such that z_{001} has constant weight (i.e., $\Theta(1)$) and z_{010} has weight $\Theta\big(N^{d\gamma^2(1-\delta)/(\tau-1)}\big)$. We again refer to the latter vertex as a 'high-weight vertex'. Similarly, we search for an edge $z_{101}z_{110}$ such that $z_{101} \in B^2(z_{100})$ has constant weight and $z_{110} \in B^2(z_{111})$ has high weight. We again analyse the probability that these edges are actually present. For each of the two edges separately, this is similar to the analysis done in step 1; however now the distance between the two vertices is $\Theta\big(N^{\gamma}\big)$ and the high-weight vertex is $\Theta\big(N^{d\gamma^2(1-\delta)/(\tau-1)}\big)$. Again by approximate computations, when ignoring dependence of the weights on step 1 and by setting R as in (5.16), we find that

$$\mathbb{P}(z_{001}, z_{010} \text{ do not exist}), \mathbb{P}(z_{101}, z_{110} \text{ do not exist}) \leq \exp\big[-\Theta\big(N^{d\gamma^2 R}\big)\big]. \tag{5.18}$$

By applying the union bound the probability that both $z_{001}z_{010}$ and $z_{101}z_{110}$ are present may be bounded by

$$\mathbb{P}(z_{001}z_{010} \text{ and } z_{101}z_{110} \text{ are present}) \geq 1 - 2\exp\big[-\Theta\big(N^{d\gamma^2 R}\big)\big]. \tag{5.19}$$

In particular, notice that there is an extra factor 2^1 in front of the exponential, and the exponent of N is multiplied with an extra factor γ.

We iterate this procedure. In the ith step, we elongate the labels of the vertices from the previous step by repeating the last digit of its label and put a box with sidelengths N^{γ^i} around it. Then, for every $\mathbf{t} \in \{0,1\}^{i-1}$, we search for an edge $z_{\mathbf{t}01}z_{\mathbf{t}10}$ between the boxes of $z_{\mathbf{t}00}$ and $z_{\mathbf{t}11}$ such that $z_{\mathbf{t}01}$ has constant weight and $z_{\mathbf{t}10}$ has weight between $\Theta\big(N^{d\gamma^i(1-\delta)/(\tau-1)}\big)$ (i.e., high weight). There are exactly 2^i such edges, and by the same analysis as in the previous two steps, for a fixed $\mathbf{t} \in \{0,1\}^{i-1}$ we have

$$\mathbb{P}(z_{\mathbf{t}01}z_{\mathbf{t}10} \text{ is present}) \geq 1 - \exp\big[-\Theta\big(N^{d\gamma^i R}\big)\big] \tag{5.20}$$

if we ignore dependence. By applying the union bound, we then obtain

$$\mathbb{P}(z_{\mathbf{t}01}z_{\mathbf{t}10} \text{ is present for all } \mathbf{t} \in \{0,1\}^{i-1}) \geq 1 - 2^{i-1}\exp\big[-\Theta\big(N^{d\gamma^i R}\big)\big]. \tag{5.21}$$

We iterate this until we have reached the kth step, where

$$k = \frac{\ln\ln N - \varepsilon \ln\ln\ln N}{\ln(1/\gamma)} \tag{5.22}$$

and $\varepsilon \in (0,1)$ some fixed number. Then, again by ignoring dependence, the probability that we have found each edge from each step can be found by just summing over (5.21) for $i = 1, \dots, k$ and using the union bound, i.e.

$$\begin{aligned} \mathbb{P}(\text{for all } i = 1, \dots, k, \text{ all edges in step } i \text{ are present}) &\geq 1 - \sum_{i=1}^{k} 2^{i-1}\exp\big[-\Theta\big(N^{d\gamma^i R}\big)\big] \\ &\geq 1 - 2^k \exp\Big[-\Theta\big(N^{d\gamma^k R}\big)\Big] \\ &=: 1 - g(N). \end{aligned} \tag{5.23}$$

Here we have used that $\gamma < 1$ and $\sum_{i=1}^{k} 2^{i-1} = 2^k - 1 \leq 2^k$. We want to show that g of (5.23) goes to 0 if $N \to \infty$. To this end, we notice that if k increases 2^k grows, while $\exp[-\Theta(N^{d\gamma^k R})]$ shrinks. In the following computation, we show that the choice of k from (5.22) is such that the latter shrinks 'faster' than the former, ensuring that $g(N)$ indeed goes to 0 if $N \to 0$, while still allowing k to be as large as possible. By substituting the definition of k from (5.22), rewriting and applying some elementary computations, we find that:

$$\begin{aligned} g(N) &= 2^{\frac{\ln\ln N - \varepsilon\ln\ln\ln N}{\ln(1/\gamma)}} \exp\Big[-\Theta\Big(N^{dR\gamma^{\frac{\ln\ln N - \varepsilon\ln\ln\ln N}{\ln(1/\gamma)}}}\Big)\Big] \\ &= \frac{(\ln N)^{\Delta(\gamma)}}{(\ln\ln N)^{\Delta(\gamma)\varepsilon}} \exp\big[-\Theta\big(\exp\big[dR(\ln\ln N)^{\varepsilon}\big]\big)\big] \\ &= \frac{1}{(\ln\ln N)^{\Delta(\gamma)\varepsilon}} \exp\big[\Delta(\gamma)\ln\ln N - \Theta\big(\exp\big[dR(\ln\ln N)^{\varepsilon}\big]\big)\big], \end{aligned} \tag{5.24}$$

where $\Delta(\gamma) = \ln(2)/\ln(1/\gamma) > 0$. Now notice that because $\Delta, R, d, \varepsilon > 0$ and $\gamma < 1$, it holds that $\Delta(\gamma)\ln\ln N - \Theta\big(\exp[dR(\ln\ln N)^{\varepsilon}]\big) \to -\infty$ if $N \to \infty$. Furthermore, $(\ln\ln N)^{-\Delta(\gamma)\varepsilon} \to 0$ if $N \to \infty$. We conclude that $g(N) \to 0$ if $N \to \infty$, which by (5.23) implies that with high probability, all edges in all steps are present.

It remains to count the amount of vertices it takes connect u and v. In the above, notice that in step i we add 2^{i-1} edges. In total, therefore, we have already used $\sum_{i=1}^{k} 2^{i-1} = 2^k - 1 \leq 2^k$ edges. Next, notice that each of the pairs $z_{\mathbf{t}_{k-1}01}$ and $z_{\mathbf{t}_{k-1}10}$, $\mathbf{t}_{k-1} \in \{0,1\}^{k-1}$ are not yet connected, but if they were, then we would have found a path between u and v. Then, notice that $|z_{\mathbf{t}_{k-1}01} - z_{\mathbf{t}_{k-1}10}| \leq \sqrt{d}N^{\gamma^k}$ by the iteration scheme. We are therefore able to connect each of these pairs with at most $C_1 N^{\gamma^k}$ nearest-neighbour edges. Here $C_1 = C_1(d) > 0$ is a positive constant. There are 2^k such pairs. In total, therefore, this path utilises at most

$$2^k + C_1 2^k N^{\gamma^k} = \frac{(\ln N)^{\Delta(\gamma)}}{(\ln\ln N)^{\Delta(\gamma)\varepsilon}}\big(1 + C_1\exp\big[(\ln\ln N)^{\varepsilon}\big]\big) = (\ln N)^{\Delta + o(1)} \tag{5.25}$$

edges by substituting the definition of k and applying Claim A.4. We therefore conclude that for every $\tilde{\varepsilon}_1$, there exists an $\underline{N}_1 = \underline{N}_1(\tilde{\varepsilon}_1)$ such that if $N \geq \underline{N}_1$ then $(\ln N)^{\Delta(\gamma)+o(1)} \leq (\ln N)^{\Delta(\gamma)+\tilde{\varepsilon}_1}$ Furthermore, since $\ln(2)/\ln(1/\gamma) \downarrow \ln(2)/\ln((\alpha+\tau-1)/(\alpha(\tau-1)))$ if $\gamma \downarrow (\alpha(\tau-1))/(\alpha+\tau-1)$, we may also conclude that for every $\tilde{\varepsilon}_2 > 0$ we may choose $\gamma = \gamma(\tilde{\varepsilon}_2)$ such that

$$\Delta(\gamma(\tilde{\varepsilon}_2)) \leq \frac{\ln 2}{\ln\left(\frac{\alpha+\tau-1}{\alpha(\tau-1)}\right)} + \tilde{\varepsilon}_2. \tag{5.26}$$

From this, we conclude that for every $\tilde{\varepsilon} = \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2 > 0$, there exists an $\underline{N}_2(\tilde{\varepsilon})$ such that

$$\mathbb{P}\big(d_{\mathcal{G}}(u,v) \leq (\ln N)^{\Delta+\tilde{\varepsilon}}\big) \geq 1 - g(N) \tag{5.27}$$

if $N \geq \underline{N}_2(\tilde{\varepsilon})$. This is the result we wanted to show.

In this sketch, we have ignored many important details. Particularly, we have often ignored dependence. The most egregious of these offences is that in each step after the first, we reveal weights that have already been revealed in previous steps. We have also ignored dependence when we reveal that all edges between two boxes are not present. The latter may be dealt with by applying Lemma 4.1. To resolve the former, in Subsection 5.2 we develop the notion of 'nets'. Roughly, a net is a subset of vertices that behaves pseudo-randomly with regards to the expected amount of vertices with a certain weight in any given radius. If such a net is present, then we are able to show that on a macro-level this dependence is largely irrelevant and the steps we have taken in the sketch are in fact largely justifiable.

5.2. Nets

In this section, to deal with the problem of dependence when revealing weights as explained above, we construct *nets*. Loosely, we say that $\mathcal{N}_v \subset \mathcal{V}$ is a net for v if $\mathcal{N}_v$ if for each vertex $x \in \mathcal{N}_v$ there are roughly the expected number of vertices with any weight around x. A bit more precise, a net should satisfy that in any set $\mathcal{X} \subset \mathcal{N}_v$ such that $|\mathcal{X}| =: n \gg 1$ and interval I, if each vertex has $\mathbb{P}(W \in I) =: p$ that its weight is in I, then there are roughly np vertices with weight in I in $\mathcal{X}$. The idea to do this comes from a renormalisation group technique — a method often used in physics. The essence of this idea is to look at the vertex set in boxes of multiple scales. We then call boxes good or bad based on the number of vertices of certain weight they contain, the number of good sub-boxes they contain and other properties.

To start the construction of the nets, we begin with making the boxes described in the sketch more precise. To this end, we define a sequence $(r_i)_i$ of sidelengths that slightly deviates from N^{γ^i}, but no more than a factor of 2.

Definition 5.4 (Sidelengths) *Consider γ from Setting 5.3, let $N > 1$ and let k be a (possibly N-dependent) positive integer. Recursively define*

$$r_k = r_k(N,\gamma) := \left\lceil N^{\gamma^k} \right\rceil, \text{ and } r_{i-1} = r_{i-1}(N,\gamma) := \left\lceil N^{\gamma^{i-1}}/r_i \right\rceil r_i \text{ for } i = 2,\dots,k. \tag{5.28}$$

Here $x \mapsto \lceil x \rceil$ denotes the ceiling function, i.e., $\lceil x \rceil = \min\{z \in \mathbb{Z} : x \leq z\}$. We refer to r_i as a ***sidelength****. We suppress the dependence on N and γ if this is clear from context.*

The reason we define the sequence of sidelengths is because these sidelengths have certain useful properties. These properties are shown in the coming Lemma 5.7. The properties, however, only hold when k is restricted. In the next definition, we give a function $k^\star_\varepsilon(N)$ that dictates this restriction.

Definition 5.5 *Consider γ from Setting 5.3. For any $\varepsilon > 0$ and $N > e$, we set*

$$k^\star_\varepsilon(N) = k^\star_\varepsilon(N,\gamma) = \frac{\ln\ln N - \varepsilon \ln\ln\ln N}{\ln(1/\gamma)}. \tag{5.29}$$

We suppress dependence on γ when this is clear from context.

In Definition 5.5, notice that we require $N > e$ since otherwise $k^\star_\varepsilon(N)$ does not exist.

We continue with a small claim that we use in multiple proofs in this section. It pertains to the fact that if $k < k^\star_\varepsilon(N)$ for any $\varepsilon > 0$, then for any constant $c > 0$ it holds that $N^{-c\gamma^k}$ goes to 0 if $N \to \infty$. This also implies that $N^{c\gamma^k}$ goes to infinity if $N \to \infty$. It is then also easily seen that this result then also holds for any positive integer $i \leq k$.

Claim 5.6 *Consider γ from Setting 5.3, let $N > e, \varepsilon > 0$, consider $k^\star_\varepsilon(N)$ from (5.29) of Definition 5.5 and let k be a (possibly N-dependent) positive integer satisfying $k < k^\star_\varepsilon(N)$. Then for any $c > 0$ and $i \leq k$ it holds that*

$$N^{-c\gamma^i} < \exp\left[-c(\ln\ln N)^\varepsilon\right] \qquad \text{and} \qquad N^{c\gamma^i} > \exp\left[c(\ln\ln N)^\varepsilon\right] \tag{5.30}$$

Proof. We show the first inequality of (5.30), the second inequality of (5.30) follows immediately from the first. Notice that because $N > 1, c > 0$ and $\gamma < 1$, the function $i \mapsto N^{-c\gamma^i}$ is increasing. Since $i \le k$, by assuming $k < k^\star_\varepsilon(N)$ holds, filling in the definition of $k^\star_\varepsilon$ from (5.29) and by applying some elementary computations we find that

$$N^{-c\gamma^i} \le N^{-c\gamma^k} < N^{-c\gamma^{\frac{\ln\ln N - \varepsilon \ln\ln\ln N}{\ln(1/\gamma)}}} = N^{-c\exp\left[-\ln(1/\gamma)\frac{\ln\ln N - \varepsilon\ln\ln\ln N}{\ln(1/\gamma)}\right]} = \exp\left[-c(\ln\ln N)^\varepsilon\right] \quad (5.31)$$

as required. ∎

We now continue with the properties of the sidelengths. In particular, we have chosen the sidelengths in such a way that each r_i is a (positive) integer, and each ratio r_{i-1}/r_i is too. This will be useful to *exactly* divide each box with side-lengths r_{i-1} up into $(r_{i-1}/r_i)^d$ sub-boxes with side-lengths r_i. Furthermore, as mentioned previously we show that each r_i differs no more than a factor of 2 from N^{γ^i}.

Lemma 5.7 (Properties of sidelengths) *Consider Setting 5.3, specifically γ. Let $N > 1$ and let k be a (possibly N-dependent) postive integer. Furthermore, let $\varepsilon > 0$ and let $k^\star_\varepsilon(N)$ be as in (5.29) of Definition 5.5.1 The sequence $(r_i)_{i\le k}$ from Definition 5.4 satisfies the following properties:*

(1) For every $i = 2, \dots, k$ it holds that r_{i-1}/r_i is a positive integer. Furthermore, every r_i is an integer for $i \le k$.

(2) For any $i \le k$ it holds that $N^{\gamma^i} \le r_i$. Additionally, for any $\varepsilon > 0$ there exists an $\underline{N}_{5.7} = \underline{N}_{5.7}(\varepsilon, \gamma)$ such that if $N \ge \underline{N}_{5.7}$ and $k < k^\star_\varepsilon(N)$ then for every $i = 1, \dots, k$ it holds that

$$N^{\gamma^i} \le r_i \le 2N^{\gamma^i}. \quad (5.32)$$

Proof. The first part of item (1) follows readily from

$$\frac{r_{i-1}}{r_i} = \frac{\left\lceil N^{\gamma^{i-1}}/r_i \right\rceil r_i}{r_i} = \left\lceil N^{\gamma^{i-1}}/r_i \right\rceil.$$

For the second part of (1), notice that r_k is an integer by definition. For any $i = 1, \dots, k-1$, we may then expand

$$r_i = r_k \frac{r_{k-1}}{r_k}\frac{r_{k-2}}{r_{k-1}} \cdots \frac{r_{i+1}}{r_{i+2}}\frac{r_i}{r_{i+1}}.$$

Since r_k and all fractions r_{j-1}/r_j are integers, r_i is an integer too.

Before we start with the proof of item (2), we firstly give the definition of $\underline{N}_{5.7}$:

$$\underline{N}_{5.7} := \exp\left[\exp\left[\left(\frac{\ln 2}{1-\gamma}\right)^{1/\varepsilon}\right]\right]. \quad (5.33)$$

In particular, notice that if $N \ge \underline{N}_{5.7}$, then it holds that

$$\exp\left[-(1-\gamma)(\ln\ln N)^\varepsilon\right] \le \frac{1}{2}. \quad (5.34)$$

Furthermore, $\underline{N}_{5.7} > e$, so that if $N \ge \underline{N}_{5.7}$ then $k^\star_\varepsilon(N)$ exists. We continue with the proof of item (2). Assume that $N \ge \underline{N}_{5.7}$. Starting with $i = k$, we see that by definition

$$N^{\gamma^k} \le r_k := \left\lceil N^{\gamma^k} \right\rceil \le N^{\gamma^k} + 1 \le 2N^{\gamma^k} \quad (5.35)$$

since $N > 1$. Next, we rewrite for $i = 1, \dots, k-1$:

$$N^{\gamma^i} \le r_i := \left\lceil N^{\gamma^i}/r_{i+1} \right\rceil r_{i+1} \le N^{\gamma^i} + r_{i+1}. \quad (5.36)$$

We notice that the first statement of item (2) holds regardless if $N \ge \underline{N}_{5.7}$ or not. To show the second statement of item (2), we show that the right-hand side of (5.36) is bounded by $2N^{\gamma^i}$. We work inductively backwards. The base case follows from (5.35). Now suppose that (5.32) holds for $i+1$. We use the induction hypothesis to rewrite (5.36) to

$$N^{\gamma^i} + r_{i+1} \le \left(1 + 2N^{-(1-\gamma)\gamma^i}\right) N^{\gamma^i}. \quad (5.37)$$

We show that the pre-factor of the right-hand side is smaller than 2. To this end, observe that $1 - \gamma > 0$ and we have assumed $k < k^\star_\varepsilon(N)$. Then, by Claim 5.6 we obtain that

$$N^{-(1-\gamma)\gamma^i} \le \exp\left[-(1-\gamma)(\ln\ln(N))^\varepsilon\right]. \quad (5.38)$$

Combining this with (5.34), we find that

$$1 + 2N^{-\gamma^i(1-\gamma)} \le 2$$

if $N \ge \underline{N}_{5.7}$, which shows that $r_i \le 2N^{\gamma^i}$ when combined with (5.36) and (5.37). By induction, this shows item (2). ∎

With the sidelengths now defined, we now create a boxing structure of nested partitions of boxes with exactly those sidelengths. To properly describe this boxing structure, we use the following definition.

Definition 5.8 *Let $a_1, b_1, \dots, a_d, b_d \in \mathbb{R}$ be $2d$ real-valued constants such that $a_i \le b_i$ for all $i \le d$. Sets of the form $(a_1, b_1] \times \cdots \times (a_d, b_d] \subseteq \mathbb{R}^d$ are called **half-open boxes**.*

In Definition 5.8, we remark here that there exists a more general notion of half-open boxes, where the constants may also be $\pm\infty$. We do not need this more general notion.

Using the above Definition 5.8, the boxing structure can now properly be described. It starts with a half-open box of with sidelength r_1 centred around some vertex x, which we call the half-open box of layer 1. We then define an iterating procedure. In each iteration, because of item (1) of Lemma 5.7 a half-open box of the previous layer $i-1$ can be partitioned into exactly $(r_{i-1}/r_i)^d$ boxes that do not overlap. This gives all boxes of the i'th layer. We continue partitioning this half-open box of level 1 into finer and finer layers, until we reach the kth layer. We refine this idea for a boxing structure as described above in Definition 5.9.

If it is not immediately clear that the boxes can really be partitioned so exactly, consider first $d = 1$ and any interval of the form $(a, a + r_{i-1}]$. By Lemma 5.7, we know that there is some integer ℓ such that $r_{i-1} = \ell r_i$. From this, it follows that we can rewrite

$$(a, a + r_{i-1}] = \bigcup_{j=1}^{\ell} (a + (j-1) r_i, a + j r_i].$$

It should be clear that all intervals of the form $(a + (j-1)r_i, a + jr_i]$ are pairwise disjoint. This idea can quite easily be generalised to d dimensions.

Definition 5.9 (Boxing structure) *Consider Setting 5.3, in particular γ. Let $N > 1$ and let k be a (possibly N-dependent) positive integer. Let $(r_i)_{i \le k}$ be the sequence of sidelengths given by Definition 5.4. Let half-open boxes be as defined in Definition 5.8. We define a boxing structure $\mathcal{S}_x = \mathcal{S}_x(N, k, \gamma, d)$ as a set of partitions of a box centered around $x \in \mathbb{R}^d$. More precisely, $\mathcal{S}_x$ is generated by the following procedure.*

- *Define B^1 to be the half-open box with sidelength r_1 centered around x, i.e.,*

$$B^1 = B^{1,x}(N, \gamma, d) := x + \left(-\frac{1}{2} r_1, \frac{1}{2} r_1\right]^d. \tag{5.39}$$

  This half-open box is called the layer 1 box.
- *Because of Lemma 5.7, we can divide $\mathcal{B}^1$ into $(r_1/r_2)^d$ half-open boxes with sidelength r_2. Index (for example, in lexicographic order) these half-open boxes and denote them*

$$B_j^2 = B_j^{2,x}(N, \gamma, d), \text{ for } j = 1, \dots, \frac{r_1^d}{r_2^d}. \tag{5.40}$$

  These half-open boxes are called boxes of layer 2.
- *Recursively, define the boxes of layer i as the subdivision of the boxes of layer $i-1$. More precisely, take any*

$$\mathbf{s}_{i-1} \in \left\{1, \dots, \left(\frac{r_1}{r_2}\right)^d\right\} \times \cdots \times \left\{1, \dots, \left(\frac{r_{i-2}}{r_{i-1}}\right)^d\right\}. \tag{5.41}$$

  Then any half-open box $B_{\mathbf{s}_{i-1}}^{i-1} = B_{\mathbf{s}_{i-1}}^{i-1,x}(N, \gamma, d)$ can be divided into exactly $(r_{i-1}/r_i)^d$ half-open boxes with sidelength r_i. We denote these boxes $B_{\mathbf{s}_{i-1},\ell}^i = B_{\mathbf{s}_{i-1},\ell}^{i,x}(N, \gamma, d)$ for ℓ ranging from 1 to $(r_{i-1}/r_i)^d$ and call them boxes of layer i. Repeat this procedure until layer k has been reached.

Then define the boxes $\mathcal{B}_{\mathbf{s}_i}^i = \mathcal{B}_{\mathbf{s}_i}^{i,x}(N, \gamma, d)$ to be exactly those vertices of the vertex set $\mathcal{V}$ that fall into $B_{\mathbf{s}_i}^i$, i.e., $\mathcal{B}_{\mathbf{s}_i}^i = B_{\mathbf{s}_i}^i \cap \mathcal{V}$. We also refer to $\mathcal{B}_{\mathbf{s}_i}^i$ as a box of layer i. The boxing structure $\mathcal{S}_x$ is defined as the set containing all these boxes:

$$\mathcal{S}_x = \left\{\mathcal{B}^{1,x}\right\} \cup \bigcup_{i=2}^{k} \left\{\mathcal{B}_{\mathbf{s}_i}^{i,x} : \mathbf{s}_i \in \left\{1, \dots, \left(\frac{r_1}{r_2}\right)^d\right\} \times \cdots \times \left\{1, \dots, \left(\frac{r_{i-1}}{r_i}\right)^d\right\}\right\}. \tag{5.42}$$

*We suppress the dependence on N, γ, k and x of $\mathcal{S}_x$ and the boxes if these parameters are clear from context. If a box $\mathcal{B}_{\mathbf{s}_j}^{j,x}$ of layer j is contained within a box $\mathcal{B}_{\mathbf{s}_i}^{i,x}$ of layer $i < j$, then we call $\mathcal{B}_{\mathbf{s}_j}^{j,x}$ a **sub-box** of $\mathcal{B}_{\mathbf{s}_i}^{i,x}$.*

Recall that because the boxes depend on the sidelengths $(r_i)_{i \le k}$. Since the sidelengths depend on N and γ, the boxes also depend on N and γ. A graphical representation of the boxing structure is given in Figure 4.

In later proofs, we need to know the amount of vertices in each box of the boxing structure. In the following Claim 5.10, we show that each box of layer i contains exactly r_i^d vertices.

Claim 5.10 *Consider Setting 5.3, let $N > 1$ and let $k \in \mathbb{N}$ be a (possibly N-dependent) positive integer. Let $(r_i)_{i \le k}$ be the sequence of sidelengths from Definition 5.4 and consider the boxing structure $\mathcal{S}_x$ centred around $x \in \mathcal{V}$. Fix $i \le k$ and let $\mathcal{B}_{\mathbf{s}_i}^{i,x} \in \mathcal{S}_x$ be any box of layer i from $\mathcal{S}_x$. Then*

$$|\mathcal{B}_{\mathbf{s}_i}^{i,x}| = r_i^d \tag{5.43}$$

Proof. First, we use Lemma 5.7 to obtain that every r_i is an integer. We note that $\mathcal{B}^i_{\mathbf{s}_i}$ can be written as a half-open box; for some $a_1, \dots, a_d \in \mathbb{R}$ we have

$$\mathcal{B}^i_{\mathbf{s}_i} = (a_1, a_1 + r_i] \times \cdots \times (a_d, a_d + r_i]. \tag{5.44}$$

We then rewrite

$$(a_1, a_1 + r_i] \times \cdots \times (a_d, a_d + r_i] = \bigcup_{j_1,\dots,j_d=1}^{r_i} (a_1 + (j_1 - 1), a_1 + j_1] \times \cdots \times (a_d + (j_d - 1), a_d + j_d]$$

Then, all half-open box of the form $(a_1+(j_1-1), a_1+j_1] \times \cdots \times (a_d+(j_d-1), a_d+j_d]$ are pairwise disjoint and must contain exactly one vertex from $\mathbb{Z}^d$. Since there are r_i^d such boxes, the result follows. ■

We move on by properly defining the constants that dictate which vertices are considered 'constant-weight' vertices and which are considered 'high-weight' vertices at layer i. Furthermore, we define constants that govern the amount of such vertices.

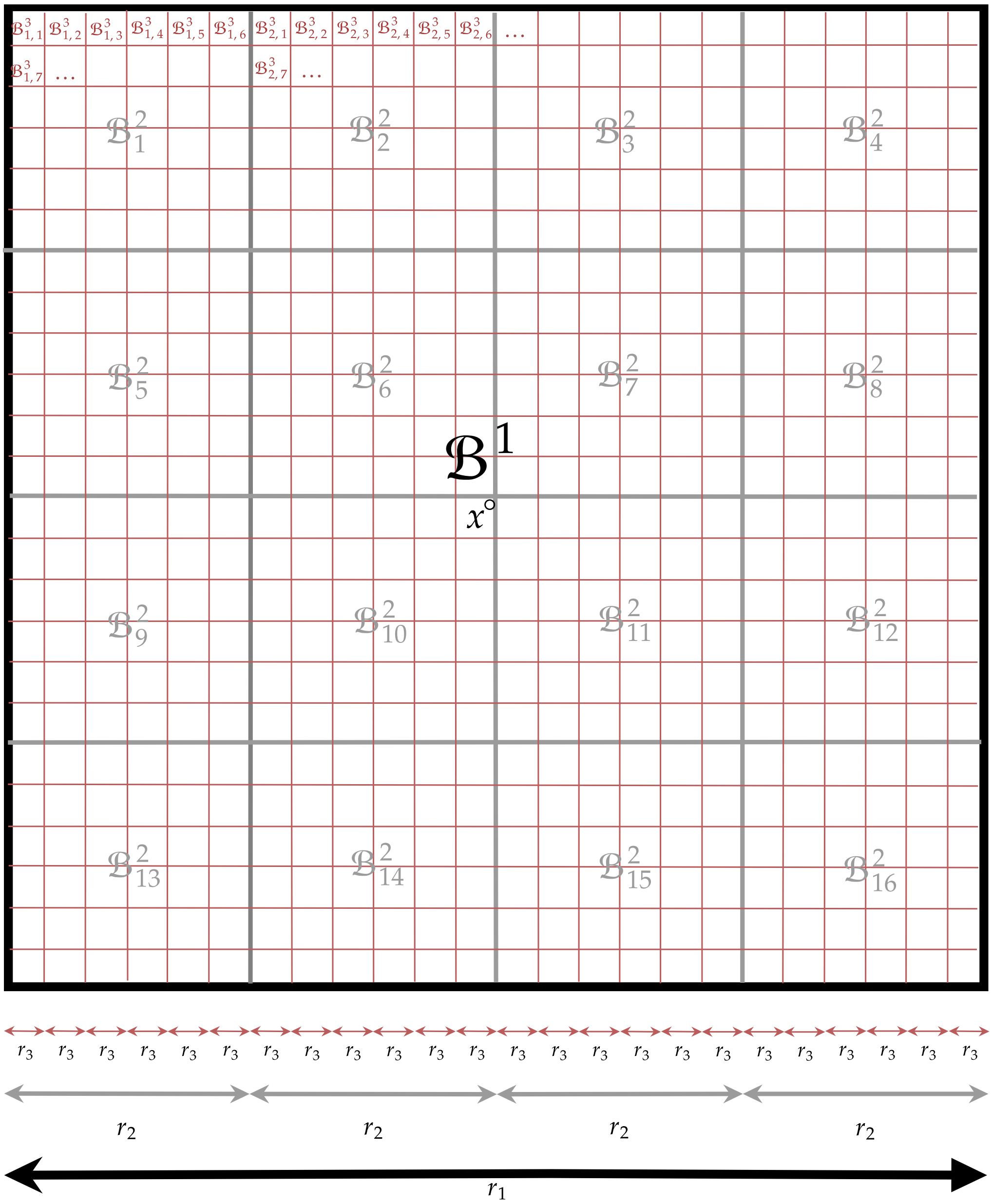


FIGURE 4. A graphical representation of the boxing scheme, where $d = 2$ and $k = 3$. The box of level 1 is given in black, the boxes of level 2 in gray and the boxes of level 3 in red. This figure is not to scale.

Definition 5.11 (Constants) *Consider Setting 5.3, especially δ and η. Let $N>1$ and let $k\in\mathbb{N}$ be a (possibly N-dependent) positive integer. Let $(r_i)_{i\le k}$ be the sequence of sidelengths from Definition 5.4. We define*

$$\underline{\eta}:=\frac{\eta}{2}, \qquad \overline{\eta}:=1-\frac{\eta}{2}. \tag{5.45}$$

Let $\phi=\left(1+\sqrt{5}\right)/2$ be the golden ratio. We define

$$\begin{aligned}\underline{M}_k&:=\frac{1}{\phi\eta}, & \underline{M}_i&:=\frac{1}{\phi\Big(\eta-\sum_{j=i+1}^k\left(\frac{1}{\underline{M}_j}-\frac{1}{\overline{M}_j}\right)r_j^{-d(1-\delta)}\Big)},\\ \overline{M}_k&:=\frac{\phi}{\eta}, & \overline{M}_i&:=\frac{\phi}{\eta-\sum_{j=i+1}^k\left(\frac{1}{\underline{M}_j}-\frac{1}{\overline{M}_j}\right)r_j^{-d(1-\delta)}},\end{aligned} \tag{5.46}$$

for $i\le k-1$. Next, we also recursively define

$$\underline{A}_k:=\underline{\eta}, \overline{A}_k:=\overline{\eta} \tag{5.47}$$

and

$$\begin{aligned}\underline{A}_i&:=1-\overline{\eta}\left[\prod_{j=i+1}^k\left(1-2r_j^{-d\delta}\right)\right]-\sum_{j=i+1}^k\left[\prod_{\ell=i+1}^{j-1}\left(1-2r_\ell^{-d\delta}\right)\right]\left[\overline{a}_j\left(1-2r_j^{-d\delta}\right)r_j^{-d(1-\delta)}+2r_j^{-d\delta}\right],\\ \overline{A}_i&:=1-\underline{\eta}\left[\prod_{j=i+1}^k\left(1-2r_j^{-d\delta}\right)\right]-\sum_{j=i+1}^k\left[\prod_{\ell=i+1}^{j}\left(1-2r_\ell^{-d\delta}\right)\right]\underline{a}_j r_j^{-d(1-\delta)}\end{aligned} \tag{5.48}$$

for $i\le k-1$, where

$$\underline{a}_i:=\frac{1}{2}\underline{A}_i, \qquad \overline{a}_i:=\frac{1}{2}\overline{A}_i+\frac{1}{2}. \tag{5.49}$$

We are now able to properly give the description of constant-weight vertices; they are exactly those vertices that have weight between 1 and $\eta^{-1/(\tau-1)}$. In particular, these vertices are called 'constant-weight' because their weight does not depend on the layer we are interested in. In contrast, we also consider 'high weight vertices', which are vertices whose weight depends layer we are interested in. More precisely, the high weight vertices of a layer i are those vertices that have weight between $\underline{M}_i^{1/(\tau-1)}r_i^{d(1-\delta)/(\tau-1)}$ and $\overline{M}_i^{1/(\tau-1)}r_i^{d(1-\delta)/(\tau-1)}$. Lastly, we use $\underline{\eta}$ and $\overline{\eta}$ to dictate the count of constant-weight vertices and $\underline{a}_i,\overline{a}_i,\underline{A}_i$ and $\overline{A}_i$ to both count and dictate the amount of high-weight vertices.

The reason behind the specific values of these sequences will become apparent in the proofs of Lemmas 5.15 and 5.16. Part of that reason is that they have certain properties, which we summarise and show in the following Claim 5.12.

Claim 5.12 (Properties constants) *Consider Setting 5.3, in particular γ,δ and η. Let $N>e,\varepsilon>0$ and consider $k_\varepsilon^\star$ from (5.29) of Definition 5.5. Let k be a (possibly N-dependent) positive integer satisfying $k<k_\varepsilon^\star(N)$. Let $(r_i)_{i\le k}$ be the sequence of sidelengths from Definition 5.4 and consider $\underline{\eta},\overline{\eta}$, $(\underline{M}_i)_{i\le k}$, $(\overline{M}_i)_{i\le k}$, $(\underline{a}_i)_{i\le k}$, $(\overline{a}_i)_{i\le k}$, $(\underline{A}_i)_{i\le k}$ and $(\overline{A}_i)_{i\le k}$ from Definition 5.11. Set $\phi=(1+\sqrt{5})/2$. There exists an $\underline{N}_{5.12}=\underline{N}_{5.12}(\varepsilon,\gamma,\delta,\eta,d)$ such that if $N\ge\underline{N}_{5.12}$, then the following all hold*

(1) For all $i=1,\dots,k$,

$$\frac{1}{\phi\eta}\le\underline{M}_i<\overline{M}_i\le\frac{2\phi}{\eta}. \tag{5.50}$$

(2) For all $i=1,\dots,k$,

$$\frac{1}{\underline{M}_i}-\frac{1}{\overline{M}_i}=\eta-\sum_{j=i+1}^k\left(\frac{1}{\underline{M}_j}-\frac{1}{\overline{M}_j}\right)r_j^{-d(1-\delta)}. \tag{5.51}$$

(3) For all $i=1,\dots,k$,

$$\frac{\eta}{4}\le\underline{A}_i<\overline{A}_i\le1-\frac{\eta}{4}. \tag{5.52}$$

Proof. We consider each item separately. For each item (i), $i=1,2,3$, we show that there exists an $\underline{N}_i$ such that the respective item holds if $N\ge\underline{N}_i$. Then taking $\underline{N}_{5.12}:=\underline{N}_1\vee\underline{N}_2\vee\underline{N}_3$ ensures that all three items holds if $N\ge\underline{N}_{5.12}$.

(Item (1)) We recall from (5.46) that

$$\underline{M}_i:=\frac{1}{\phi\left(\eta-\sum_{j=i+1}^k\left(\frac{1}{\underline{M}_j}-\frac{1}{\overline{M}_j}\right)r_j^{-d(1-\delta)}\right)} \text{ and } \overline{M}_i:=\frac{\phi}{\eta-\sum_{j=i+1}^k\left(\frac{1}{\underline{M}_j}-\frac{1}{\overline{M}_j}\right)r_j^{-d(1-\delta)}} \tag{5.53}$$

for $i\le k-1$. Observe that (5.50) follows immediately if we show that

$$\frac{1}{2}\eta\le\eta-\sum_{j=i+1}^k\left(\frac{1}{\underline{M}_j}-\frac{1}{\overline{M}_j}\right)r_j^{-d(1-\delta)}\le\eta. \tag{5.54}$$

We firstly define $\underline{N}_1 = \underline{N}_1(\varepsilon,\gamma,\delta,d)$, and afterwards show that (5.54) holds for all $i \le k$ if $N \ge \underline{N}_1$. To this end, we note that $d(1-\delta) > 0$ by (5.11) and hence by item (2) of Lemma 5.7 (properties of sidelengths), it holds that $kr_k^{-d(1-\delta)} \le kN^{-d(1-\delta)\gamma^k}$. Furthermore, by because $k < k_\varepsilon^\star(N)$ and Claim 5.6 we find that

$$kr_k^{-d(1-\delta)} \le kN^{-d(1-\delta)\gamma^k} \le \frac{\ln\ln(N) - \varepsilon \ln\ln\ln(N)}{\ln(1/\gamma)} \exp\big[-d(1-\delta)(\ln\ln N)^\varepsilon\big] =: g(N). \tag{5.55}$$

Observe that $g(N) \to 0$ if $N \to \infty$. For this reason, we may choose $\underline{N}_1$ such that if $N \ge \underline{N}_1$, it holds that $g(N) \le \phi^2/2$, where $\phi = (1+\sqrt{5})/2$.

Fix $\underline{N}_1$ as described and assume that $N \ge \underline{N}_1$. We continue by showing (5.54) using backwards induction. The base case is clear by the definition of $\underline{M}_k$ and $\overline{M}_k$ (see (5.46)). Suppose that (5.50) holds for all $j = i+1,\dots,k$ (induction hypothesis). We show that it also holds for i. We compute $\phi^2 - 2 = 1/\phi$, from which we can see that $\phi\eta - 2\eta/\phi = \eta/\phi^2$. Hence, under the induction hypothesis

$$0 < \frac{1}{\underline{M}_j} - \frac{1}{\overline{M}_j} \le \eta/\phi^2 \tag{5.56}$$

for all $j = i+1,\dots,k$. Since the sum in (5.54) consists only of positive terms, the right-most inequality of (5.54) follows. To show the left-hand side, we note that under the induction hypothesis

$$\eta - \sum_{j=i+1}^{k}\left(\frac{1}{\underline{M}_j} - \frac{1}{\overline{M}_j}\right) r_j^{-d(1-\delta)} \ge \eta\left(1 - \frac{1}{\phi^2}\sum_{j=i+1}^{k} r_j^{-d(1-\delta)}\right) \ge \eta\left(1 - \frac{kr_k^{-d(1-\delta)}}{\phi^2}\right) \ge \frac{1}{2}\eta \tag{5.57}$$

In the first inequality, we have used (5.56), in the second inequality we have used that $r_j \ge r_k$ for all $j \le k$, and in the last inequality we have used that $N \ge \underline{N}_1$ and hence $kr_k^{-d(1-\delta)} \le \phi^2/2$.

(Item (2)) This follows directly by substituting the definition of $\underline{M}_i$ and $\overline{M}_i$ (see (5.46) or (5.53)) into the left-hand side of (5.51), provided that the denominator of $\underline{M}_i$ and $\overline{M}_i$ is never 0. By the previous item, we ensure that this does not happen if $N \ge \underline{N}_1$. Therefore by setting $\underline{N}_2 := \underline{N}_1$ we also ensure that the second item holds if $N \ge \underline{N}_2$.

(Item (3)) Similarly to the proof of item (1), we first give the definition of $\underline{N}_3 = \underline{N}_3(\varepsilon,\gamma,\delta,\eta,d)$ and then show that (5.52) holds for all $i \le k$ if $N \ge \underline{N}_3$. By replacing $(1-\delta)$ by δ in (5.55), we may also show that $kr_k^{-d\delta} \to 0$ if $N \to \infty$. As such, it is possible to find an $\underline{N}_3$ such that if $N \ge \underline{N}_3$, then the following inequality holds:

$$kr_k^{-d\delta} < \frac{\eta}{12}. \tag{5.58}$$

Fix this $\underline{N}_3$ and let $N \ge \underline{N}_3$. Notice that it then also holds that

$$r_k^{-d\delta} \le kr_k^{-d\delta} < \frac{\eta}{12} < \frac{1}{4} \qquad \text{and thus} \qquad \frac{1}{2} < (1 - 2r_j^{-d\delta}) < 1 \quad \text{for all } j \le k, \tag{5.59}$$

since k is a positive integer, $\eta \in (0, 2/3)$, and for all $j \le k$ it holds that $r_j \ge r_k > 1$. We proceed with showing (5.52) for all $i \le k$ with backwards induction. The base case follows immediately from the definition of $\underline{A}_k := \underline{\eta} = \eta/2$ and $\overline{A}_k = \overline{\eta} = 1 - \eta/2$ and the fact that $\eta \in (0, 2/3)$. Suppose that (5.52) holds for all $j = i+1,\dots,k$ (induction hypothesis). We show that (5.52) then also holds for i. To this end, firstly recall the definition of $\underline{A}_i$ and $\overline{A}_i$:

$$\underline{A}_i := 1 - \overline{\eta}\left[\prod_{j=i+1}^{k}\left(1 - 2r_j^{-d\delta}\right)\right] - \sum_{j=i+1}^{k}\left[\prod_{\ell=i+1}^{j-1}\left(1 - 2r_\ell^{-d\delta}\right)\right]\left[\overline{a}_j\left(1 - 2r_j^{-d\delta}\right) r_j^{-d(1-\delta)} + 2r_j^{-d\delta}\right], \tag{5.60}$$

and

$$\overline{A}_i := 1 - \underline{\eta}\left[\prod_{j=i+1}^{k}\left(1 - 2r_j^{-d\delta}\right)\right] - \sum_{j=i+1}^{k}\left[\prod_{\ell=i+1}^{j}\left(1 - 2r_\ell^{-d\delta}\right)\right]\underline{a}_j r_j^{-d(1-\delta)}. \tag{5.61}$$

Furthermore, recall that $\underline{a}_i = \underline{A}_i/2$ and $\overline{A}_i = (\overline{A}_i + 1)/2$. We treat each of the three inequalities in (5.52) separately. To show the left-most inequality of (5.52), note that the induction hypothesis implies that $0 < \eta/8 \le \underline{a}_j < \overline{a}_j \le 1 - \eta/8 < 1$ for all $j = i+1,\dots,k$. Now because $N \ge \underline{N}_3$, by (5.59) it holds that $1/2 < (1 - 2r_j^{-d\delta}) \le 1$. Furthermore, by (5.11) we know that $\delta < 1/4$, which implies that $\delta < 1 - \delta$. Because $r_j > 1$ for all $j \le k$, this last inequality in turn results in $r_j^{-d(1-\delta)} \le r_j^{-d\delta}$. Combining these facts with the definition of $\underline{A}_i$ from (5.60), we find that

$$\underline{A}_i \ge 1 - \overline{\eta} - \sum_{j=i+1}^{k} 3r_j^{-d\delta} \ge 1 - \overline{\eta} - 3kr_k^{-d\delta} > 1 - \left(1 - \frac{\eta}{2}\right) - 3\frac{\eta}{12} = \frac{\eta}{4}. \tag{5.62}$$

In the second inequality we have used that $r_j \ge r_k$ for all $j \le k$. In the third inequality we have used that because $N \ge \underline{N}_3$, it holds that $kr_k^{-d\delta} < \eta/12$ by (5.58). This shows the left-most inequality of (5.52).

The middle inequality of (5.52) can be seen quickly by the fact that $\underline{A}_i$ has its negative terms multiplied with larger pre-factors (i.e., $\underline{\eta} < \overline{\eta}$ and $\underline{a}_j < \overline{a}_j$ by the induction hypothesis) and has an extra negative term when compared to $\overline{A}_i$.

We proceed with showing the right-most inequality of (5.52). Again we use that $1/2 < (1-2r_j^{-d\delta}) \leq 1$ when $N \geq \underline{N}_3$ by (5.59). Furthermore, we use that $1-x \geq e^{-2x}$, which is valid when $0 \leq x \leq 1/2$, and the fact that $r_j \geq r_k$ for $j \leq k$. Then we find that

$$\overline{A}_i \leq 1 - \underline{\eta}\left[\prod_{j=i+1}^{k}\left(1-2r_j^{-d\delta}\right)\right] \leq 1 - \underline{\eta}\left(1-2r_k^{-d\delta}\right)^k \leq 1 - \underline{\eta}\exp\left[-4kr_k^{-d\delta}\right]. \tag{5.63}$$

In the first inequality of (5.63), we ignore the right-most negative terms of (5.61). In the second inequality of (5.63), we use that $1-2r_j^{-d\delta} \geq 1-2r_k^{-d\delta}$ and $(1-2r_k^{-d\delta})^{k-i} \geq (1-2r_k^{-d\delta})^k$. In the third inequality of (5.63), we utilise that $2r_j^{-d\delta} \leq 2r_k^{-d\delta} \leq 1/2$ because $N \geq \underline{N}_3$. Next, notice that $\exp[-4x] \geq 1/2$ if $x < \ln(2)/4$. Since $N \geq \underline{N}_3$, by (5.58) we know that $kr_k^{-d\delta} < \eta/12 < 1/18 < \ln(2)/4$ (we may compute that $1/18 \approx 0.0555$ and $\ln(2)/4 \approx 0.1733$). By combining the previous two inequalities with (5.63) and the definition of $\underline{\eta} = \eta/2$ we find that $\overline{A}_i \leq 1 - \eta/4$. The right-most inequality of (5.52) follows.
This finishes the proof. ∎

Next, we introduce notation to distinguish vertices that have a given weight. This is comparable to $T_{\geq}$ from Definition 6.6. However, whereas in the construction of the path in Section 6 we were interested in vertices that just have weight larger than a given value, we now search for vertices that have weight in a given interval.

Definition 5.13 *Consider Setting 5.3. For each $E \subseteq \mathcal{V} = \mathbb{Z}^d$ and $I \subseteq \mathbb{R}_{\geq 1}$, set*

$$T(E,I) = \{v \in E : W_v \in I\}, \qquad \text{and} \qquad \#T(E,I) = |T(E,I)|. \tag{5.64}$$

Using Definitions 5.11 and 5.13, we may describe when a box behaves 'well', in the sense that it contains enough vertices with their weight in a certain interval. In particular, we make the idea described at the very start of this subsection precise.

Definition 5.14 (Lower-sufficient, upper-sufficient, good, bad) *Consider Setting 5.3, specifically $\gamma, \delta,$ and η. Fix $N > 1$ and let k be a (possibly N-dependent) positive integer. Consider the sequence of sidelengths $(r_i)_{i\leq k}$ from Definition 5.4, the boxing structure $\mathcal{S}_x(N,k,\gamma,d)$ around a vertex $x \in \mathcal{V}$ from Definition 5.9 and let T and $\#T$ be from (5.64) of Definition 5.13. Recall $\underline{\eta}$, $\overline{\eta}$, $(\underline{a}_i)_{i\leq k}$, $(\overline{a}_i)_{i\leq k}$, $(\underline{M}_i)_{i\leq k}$, $(\overline{M}_i)_{i\leq k}$ from Definition 5.11. For $i \leq k$, we call any box $\mathcal{B}^{i,x}_{\mathbf{s}_i} \in \mathcal{S}_x$ $(\gamma,\delta,\eta,\tau,d)$-**upper-sufficient** iff*

$$\underline{a}_i r_i^{d\delta} \leq \#T\Big(\mathcal{B}^{i,x}_{\mathbf{s}_i}, \Big[\underline{M}_i^{1/(\tau-1)} r_i^{d(1-\delta)/(\tau-1)}, \overline{M}_i^{1/(\tau-1)} r_i^{d(1-\delta)/(\tau-1)}\Big]\Big) \leq \overline{a}_i r_i^{d\delta}. \tag{5.65}$$

*We call any box $\mathcal{B}^{k,x}_{\mathbf{s}_k} \in \mathcal{S}_x$ of layer k $(\gamma,\delta,\eta,\tau,d)$-**lower-sufficient** iff*

$$\underline{\eta} r_k^d \leq \#T\big(\mathcal{B}^{k,x}_{\mathbf{s}_k}, \big[1, \eta^{-1/(\tau-1)}\big]\big) \leq \overline{\eta} r_k^d. \tag{5.66}$$

*For $i \leq k-1$, we call any box $\mathcal{B}^{i,x}_{\mathbf{s}_i} \in \mathcal{S}_x$ $(\gamma,\delta,\eta,\tau,\mathbf{d})$-**lower-sufficient** iff*

$$\#\left\{\ell = 1, \dots, \frac{r_i^d}{r_{i+1}^d} : \mathcal{B}^{i+1,x}_{\mathbf{s}_i\ell} \text{ is both upper-sufficient and lower-sufficient}\right\} \geq (1-2r_{i+1}^{-d\delta})\frac{r_i^d}{r_{i+1}^d}. \tag{5.67}$$

*We call boxes that are both lower-sufficient and upper-sufficient $(\gamma,\delta,\eta,\tau,d)$-**good**. Any box that is not $(\gamma,\delta,\eta,\tau,d)$-good is $(\gamma,\delta,\eta,\tau,d)$-**bad**. When γ,δ,η,τ and d are clear from context, we suppress their dependence and write upper-sufficient, lower-sufficient, good and bad.*

We remark that (5.67) should be interpreted as '$\mathcal{B}^i_{\mathbf{s}_i}$ is lower-sufficient if it contains at least $(1-2r_{i+1}^{-d\delta})(r_i/r_{i+1})^d$ good sub-boxes'.
In the following three lemmas, we work towards showing that any layer 1 box of the boxing structure is good with high probability as $N \to \infty$. To this end, we start from the bottom up with the smallest boxes (i.e., those of the form $\mathcal{B}^k_{\mathbf{s}_k}$). In the following Lemma 5.15, we show that these boxes are lower-sufficient.

Lemma 5.15 *Consider Setting 5.3, specifically γ, δ, η. Let $N > e$ and let $k \in \mathbb{N}$ be a (possibly N-dependent) positive integer. Let $(r_i)_{1\leq i\leq k}$ be as given in Definition 5.4 and let $\underline{\eta}$ and $\overline{\eta}$ be as given in Definition 5.11. Consider the boxing structure $\mathcal{S}_x(N,k,\gamma,d)$ around a vertex $x \in \mathcal{V}$ from Definition 5.9. Furthermore, let $\#T$ be from (5.64) of Definition 5.13 and consider $(\gamma,\delta,\eta,\tau,d)$-lower-sufficient from Definition 5.14. Fix $\mathcal{B}^{k,x}_{\mathbf{s}_k} \in \mathcal{S}_x$ any box of layer k from the boxing structure. There exists a $C_{5.15} = C_{5.15}(\eta)$ such that*

$$\mathbb{P}\big(\mathcal{B}^{k,x}_{\mathbf{s}_k} \text{ is } (\gamma,\delta,\eta,\tau,d)\text{-lower-sufficient}\big) \geq 1 - 2\exp\left[-C_{5.15}(\eta) r_k^d\right]. \tag{5.68}$$

Furthermore, the event in the left-hand side of (5.68) is independent of the weight of all vertices that are in $\mathcal{V}\backslash\mathcal{B}^{k,x}_{\mathbf{s}_k}$.

Proof. We write lower-sufficient rather than $(\gamma,\delta,\eta,\tau,d)$-lower-sufficient since γ,δ,η,τ and d are assumed to be fixed. Furthermore, we observe that this result holds regardless of the vertex x, so its dependence is suppressed. Throughout this proof, denote by X the amount of vertices in $\mathcal{B}^k_{\mathbf{s}_k}$ that have weight in $[1,\eta^{-1/(\tau-1)}]$, i.e.,

$$X := \#T\big(\mathcal{B}^k_{\mathbf{s}_k}, \big[1,\eta^{-1/(\tau-1)}\big]\big).$$

Recalling the definition of lower-sufficient for the boxes of layer k from (5.66) of Definition 5.14, we may then rewrite

$$Q := \mathbb{P}\Big(\mathcal{B}^k_{\mathbf{s}_k} \text{ is not lower-sufficient}\Big) = \mathbb{P}\left(X \notin \left[\underline{\eta} r_k^d, \overline{\eta} r_k^d\right]\right). \tag{5.69}$$

We bound Q. By Lemma 5.10 there are exactly r_k^d vertices in $\mathcal{B}^k_{\mathbf{s}_k}$. By the Definition 3.1 of a KSRG and the assumption on the weight distribution given by (3.4), each of those vertices independently has probability $1-\eta$ that its weight is between 1 and $\eta^{-1/(\tau-1)}$. We conclude that $X \sim \operatorname{Bin}\left(r_k^d, 1-\eta\right)$. Furthermore, notice that because $0<\eta<2/3$ we also find that $\underline{\eta} := \eta/2 < 1-\eta$ and $\overline{\eta} = 1-\eta/2 > 1-\eta$. We are therefore justified in applying the Chernoff bound for binomial random variables (Lemma B.4) to the right-most probability of (5.69). We therefore split (5.69) and rewrite it to match the format of the Chernoff bound to find that

$$\begin{aligned} Q &= \mathbb{P}\left(X < \left(1-\left(1-\frac{\underline{\eta}}{1-\eta}\right)\right)(1-\eta)r_k^d\right) + \mathbb{P}\left(X > \left(1+\left(\frac{\overline{\eta}}{1-\eta}-1\right)\right)(1-\eta)r_k^d\right) \\ &\le \exp\left[-\frac{1}{2}\left(1-\frac{\underline{\eta}}{1-\eta}\right)^2(1-\eta)r_k^d\right] + \exp\left[-\frac{1}{3}\left(\frac{\overline{\eta}}{1-\eta}-1\right)^2(1-\eta)r_k^d\right]. \end{aligned} \tag{5.70}$$

Now substituting the definition of $\underline{\eta}=\eta/2$ and $\overline{\eta} = 1-\eta/2$ from (5.45) into (5.70) and by applying some elementary bounds

$$Q \le 2\exp\left[-\left(\frac{(1-\frac{3}{2}\eta)^2}{2(1-\eta)} \wedge \frac{\eta^2}{12(1-\eta)}\right) r_k^d\right] =: 2\exp\left[-C_{5.15}(\eta) r_k^d\right].$$

Recalling (5.69) finishes the proof of (5.68). Furthermore, notice that throughout this proof, we have only considered the weight of vertices that are in $\mathcal{B}^k_{\mathbf{s}_k}$. Because the weights of all vertices are assumed to be independent by Definition 3.1 of a KSRG , the result in (5.68) is independent of all vertices in $\mathcal{V}\backslash\mathcal{B}^k_{\mathbf{s}_k}$. ■

Now that we have obtained a bound for the probability that a box of layer k is lower-sufficient, we continue by computing a bound for the probability that it is good. To compute this bound, we require one additional assumption; the boxing structure cannot have too many layers when compared to N. In particular, we need additionally require that $k < k^\star_\varepsilon(N)$ from (5.29) of Definition 5.5.

Lemma 5.16 *Consider the same setting as Lemma 5.15. Let $(\gamma,\delta,\eta,\tau,d)$-upper-sufficient and $(\gamma,\delta,\eta,\tau,d)$-good be as given in Definition 5.14. Furthermore, fix $\varepsilon>0$, consider $k^\star_\varepsilon(N)$ from (5.29) of Definition 5.5 and assume that k is such that $k<k^\varepsilon_\star(N)$. Then there exists a $\underline{N}_{5.16} = \underline{N}_{5.16}(\gamma,\delta,\eta,\tau,d)$ such that if $N \ge \underline{N}_{5.16}$, then for any box $\mathcal{B}^{k,x}_{\mathbf{s}_k} \in \mathcal{S}_x$ of layer k it holds that*

$$\mathbb{P}\left(\mathcal{B}^{k,x}_{\mathbf{s}_k} \textit{ is } (\gamma,\delta,\eta,\tau,d)\textit{-good}\right) \ge 1 - r_k^{-d\delta}. \tag{5.71}$$

Furthermore, the event in (5.71) is independent of the weight of all vertices in $\mathcal{V}\backslash\mathcal{B}^{k,x}_{\mathbf{s}_k}$.

Proof. Similar to the proof of Lemma 5.15, we suppress the dependence of x, N, γ and d of $\mathcal{S}_x$ and γ,δ,η,τ and d from upper-sufficient, lower-sufficient and good. We rewrite:

$$\begin{aligned} \mathbb{P}\left(\mathcal{B}^k_{\mathbf{s}_k} \text{ is good}\right) &= \mathbb{P}\left(\mathcal{B}^k_{\mathbf{s}_k} \text{ is upper-sufficient and } \mathcal{B}^k_{\mathbf{s}_k} \text{ is lower-sufficient}\right) \\ &= \underbrace{\mathbb{P}\left(\mathcal{B}^k_{\mathbf{s}_k} \text{ is upper-sufficient}\,\middle|\, \mathcal{B}^k_{\mathbf{s}_k} \text{ is lower-sufficient}\right)}_{=:1-\widetilde{Q}} \mathbb{P}\left(\mathcal{B}^k_{\mathbf{s}_k} \text{ is lower-sufficient}\right). \end{aligned} \tag{5.72}$$

Through Lemma 5.15 we already know a lower bound for the probability that $\mathcal{B}^k_{\mathbf{s}_k}$ is lower sufficient. We therefore bound $\widetilde{Q}$. Before we do this, we lay some groundwork.
Firstly, notice that because $k<k^\star_\varepsilon(N)$ and Claim 5.6, if follows that $r_k \to \infty$ if $N\to\infty$ (see also the proof of Claim 5.12). We recall the definition of $\underline{M}_k = 1/(\phi\eta)$ and $\overline{M}_k = \phi/\eta$ from (5.46) of Definition 5.11, where $\phi = (1+\sqrt{5})/2$. We apply item (2) of Lemma 5.7 to see that $r_k \ge N^{\gamma^k}$. From the assumption that $k<k^\star_\varepsilon(N)$, by Claim 5.6 we find $N^{\gamma^k} \to \infty$ if $N\to\infty$ and hence also $r_k\to\infty$ if $N\to\infty$. We conclude that there exists an $\underline{N}_1 = \underline{N}_1(\gamma,\delta,\eta,\tau,d)$ such that if $N\ge \underline{N}_1$, then

$$\eta^{-\frac{1}{\tau-1}} < \underline{M}_k^{\frac{1}{\tau-1}} r_k^{d\frac{1-\delta}{\tau-1}} < \overline{M}_k^{\frac{1}{\tau-1}} r_k^{d\frac{1-\delta}{\tau-1}}. \tag{5.73}$$

Fix this $\underline{N}_1$. Throughout the remainder of this proof, assume that $N\ge\underline{N}_1$.
Recall $\#T$ from (5.64) of Definition 5.13. By Lemma 5.10, there are exactly r_k^d vertices in $\mathcal{B}^k_{\mathbf{s}_k}$. Of those r_k^d vertices, there are $\#T\big(\mathcal{B}^k_{\mathbf{s}_k}, \big[1,\eta^{-1/(\tau-1)}\big]\big)$ vertices that have weight between 1 and $\eta^{-1/(\tau-1)}$, which

implies that there are $r_k^d - \#T\big(\mathcal{B}^k_{\mathbf{s}_k}, [1, \eta^{-1/(\tau-1)}]\big)$ remaining vertices that do *not* have weight between 1 and $\eta^{-1/(\tau-1)}$. Furthermore, the weight of each of those remaining vertices independently satisfy

$$\mathbb{P}\Big(W \in \Big[\underline{M}_k^{\frac{1}{\tau-1}} r_k^{d\frac{1-\delta}{\tau-1}}, \overline{M}_k^{\frac{1}{\tau-1}} r_k^{d\frac{1-\delta}{\tau-1}}\Big] \;\Big|\; W > \eta^{-\frac{1}{\tau-1}}\Big) = \eta^{-1}\left(\frac{1}{\underline{M}_k} - \frac{1}{\overline{M}_k}\right) r_k^{-d(1-\delta)} = r_k^{-d(1-\delta)}.$$

Here we have used that by (5.73) the left-most interval lies completely above $\eta^{-1/(\tau-1)}$, the conditional power-law distribution is given by Claim A.3 and the fact that $\underline{M}_k^{-1} - \overline{M}_k^{-1} = \eta(\phi - 1/\phi) = \eta$. By recalling the definition of upper-sufficient from (5.65) of Definition 5.14, we thus conclude that under the assumption that $N \geq \underline{N}_1$ it holds that

$$X := \#T\Big(\mathcal{B}^k_{\mathbf{s}_k}, \Big[\underline{M}_k^{\frac{1}{\tau-1}} r_k^{d\frac{1-\delta}{\tau-1}}, \overline{M}_k^{\frac{1}{\tau-1}} r_k^{d\frac{1-\delta}{\tau-1}}\Big]\Big) \sim \mathrm{Bin}\big(r_k^d - \#T\big(\mathcal{B}^k_{\mathbf{s}_k}[1, \eta^{-1/(\tau-1)}]\big), r_k^{-d(1-\delta)}\big). \tag{5.74}$$

Next, we define two random variables that stochastically bound X. The fact that $\mathcal{B}^k_{\mathbf{s}_k}$ is lower-sufficient by definition (see (5.66) of Definition 5.14) means that $\#T\big(\mathcal{B}^k_{\mathbf{s}_k}[1, \eta^{-1/(\tau-1)}]\big) \leq \overline{\eta} r_k^d$, which implies that given that $\mathcal{B}^k_{\mathbf{s}_k}$ is lower-sufficient, X stochastically dominates another binomial random variable:

$$(X|\mathcal{B}^k_{\mathbf{s}_k} \text{ is lower-sufficient}) \overset{d}{\geq} \underline{X}, \qquad \text{where} \qquad \underline{X} \sim \mathrm{Bin}\left((1-\overline{\eta})\, r_k^d, r_k^{-d(1-\delta)}\right). \tag{5.75}$$

Similarly, by using the lower bound for $\#T\big(\mathcal{B}^k_{\mathbf{s}_k}, [1, \eta^{-1/(\tau-1)}]\big)$ when $\mathcal{B}^k_{\mathbf{s}_k}$ is lower sufficient from (5.66), we find that

$$(X|\mathcal{B}^k_{\mathbf{s}_k} \text{ is lower-sufficient}) \overset{d}{\leq} \overline{X}, \qquad \text{where} \qquad \overline{X} \sim \mathrm{Bin}\left((1-\underline{\eta})\, r_k^d, r_k^{-d(1-\delta)}\right). \tag{5.76}$$

We return to $\widetilde{Q}$ by splitting it into two components, which we bound separately. By recalling the definition of upper-sufficient (Definition 5.14), we see that

$$\widetilde{Q} = \underbrace{\mathbb{P}\left(X < \underline{a}_k r_k^{d\delta} \,\middle|\, \mathcal{B}^k_{\mathbf{s}_k} \text{ is lower-sufficient}\right)}_{=:\widetilde{Q}_1} + \underbrace{\mathbb{P}\left(X > \overline{a}_k r_k^{d\delta} \,\middle|\, \mathcal{B}^k_{\mathbf{s}_k} \text{ is lower-sufficient}\right)}_{=:\widetilde{Q}_2}.$$

We then apply (5.75) and (5.76) to see that

$$\widetilde{Q}_1 \leq \mathbb{P}\left(\underline{X} < \underline{a}_k r_k^{d\delta}\right) \qquad \text{and} \qquad \widetilde{Q}_2 \leq \mathbb{P}\left(\overline{X} > \overline{a}_k r_k^{d\delta}\right). \tag{5.77}$$

Then, notice that by (5.49) of Definition 5.11,

$$\underline{a}_k = \frac{1}{2}(1-\overline{\eta}) < 1 - \overline{\eta}, \qquad \text{and} \qquad \overline{a}_k = 1 - \frac{1}{2}\underline{\eta} > 1 - \underline{\eta}. \tag{5.78}$$

We are therefore justified in applying the Chernoff bound for binomially distributed random variables (Lemma B.4) on both probabilities in (5.77). By doing so and rewriting we find that

$$\widetilde{Q}_1 \leq \exp\left[-\frac{1}{2}\left(1 - \frac{\underline{a}_k}{1-\overline{\eta}}\right)^2 (1-\overline{\eta})\, r_k^d r_k^{-d(1-\delta)}\right] = \exp\Bigg[-\underbrace{\frac{(1-\overline{\eta}-\underline{a}_k)^2}{2(1-\overline{\eta})}}_{=:C_1(\eta)} r_k^{d\delta}\Bigg] =: \widetilde{Q}_1^{\leq} \tag{5.79}$$

and

$$\widetilde{Q}_2 \leq \exp\left[-\frac{1}{3}\left(\frac{\overline{a}_k}{1-\underline{\eta}} - 1\right)^2 (1-\underline{\eta}) r_k^d r_k^{-d(1-\delta)}\right] = \exp\Bigg[-\underbrace{\frac{(\overline{a}_k + \underline{\eta} - 1)^2}{3(1-\underline{\eta})}}_{=:C_2(\eta)} r_k^{d\delta}\Bigg] =: \widetilde{Q}_2^{\leq}. \tag{5.80}$$

To see that $C_1(\eta)$ and $C_2(\eta)$ are actually constants that only depend on the constant η, recall from (5.49) of Definition 5.11 that $\underline{a}_k = \eta/4$ and $\overline{a}_k = 1-\eta/4$. We conclude that $1-\widetilde{Q} \geq 1 - \widetilde{Q}_1^{\leq} - \widetilde{Q}_2^{\leq}$. Finally, we return to equation (5.72). By combining the bounds for $\widetilde{Q}$, Lemma 5.15, the inequality $(1-x)(1-y) \geq 1-x-y$ which is valid if $x, y > 0$, and some elementary bounds, we obtain

$$\begin{aligned} \mathbb{P}\Big(\mathcal{B}^k_{\mathbf{s}_k} \text{ is good}\Big) &\geq (1 - \widetilde{Q}_1^{\leq} - \widetilde{Q}_2^{\leq})\left(1 - 2\exp\left[-C_{5.15}(\eta) r_k^d\right]\right) &(5.81)\\ &\geq 1 - 4\exp\left[-C_{5.16}(\eta) r_k^{d\delta}\right]. &(5.82) \end{aligned}$$

Here, $C_{5.16}(\eta) = C_1(\eta) \wedge C_2(\eta) \wedge C_{5.15}(\eta)$. For any constant $\hat{C}$, there is an $\underline{x}$ such that if $x > \underline{x}$, then $1 - 4\exp(-\hat{C}x) \geq 1 - x^{-1}$. It follows that there exists an $\underline{N}_2 = \underline{N}_2(\gamma, \delta, \eta, \tau, d)$ such that if $N > \underline{N}_2$, then

$$\mathbb{P}\big(\mathcal{B}^k_{\mathbf{s}_k} \text{ is good}\big) \geq 1 - r_k^{-d\delta}. \tag{5.83}$$

Setting $\underline{N}_{5.16} = \underline{N}_{5.16}(\gamma, \delta, \eta, \tau, d) := \underline{N}_1(\gamma, \delta, \eta, \tau, d) \vee \underline{N}_2(\gamma, \delta, \eta, d)$ finishes the proof of (5.71). Furthermore, similarly to the end of the proof of Lemma 5.15, we note that throughout this proof we have only used the weight of vertices within $\mathcal{B}^k_{\mathbf{s}_k}$ and as such, the event in (5.71) is independent of the weight of all vertices in $\mathcal{V} \backslash \mathcal{B}^k_{\mathbf{s}_k}$. ■

We have now found a bound for the probability that the boxes of the lowest level (i.e., the boxes of level k) are good. Next, our goal is to extend this result to boxes of any level $i \leq k$. We show this using induction; we start from the smallest boxes and work to the larger layers.

Lemma 5.17 *Consider Setting 5.3, in particular γ,δ and η. Fix $N>e$, $\varepsilon>0$, recall $k_\varepsilon^\star(N)$ from (5.29) of Definition 5.5 and let k be a positive integer satisfying $k<k_\varepsilon^\star(N)$. Let $(r_i)_{i\le k}$ be the sequence of sidelengths from Definition 5.4. Consider the boxing structure $\mathcal{S}_x(N,k,\gamma,d)$ around a vertex $x\in\mathcal{V}$ from Definition 5.9 and $(\gamma,\delta,\eta,\tau,d)$-good from Definition 5.14. Then there exists an $\underline{N}_{5.17}=\underline{N}_{5.17}(\varepsilon,\gamma,\delta,\eta,\tau,d)$ such that if $N>\underline{N}_{5.17}$, then for any $i\le k$ and any box $\mathcal{B}_{\mathbf{s}_i}^{i,x}\in\mathcal{S}_x$ of layer i it holds that*

$$\mathbb{P}\left(\mathcal{B}_{\mathbf{s}_i}^{i,x} \textit{ is } (\gamma,\delta,\eta,\tau,d)\textit{-good}\right)\ge 1-r_i^{-d\delta}. \tag{5.84}$$

Furthermore, the event in (5.84) is independent of the weight of all vertices within $\mathcal{V}\backslash\mathcal{B}_{\mathbf{s}_i}^{i,x}$.

Proof. Similar to the proof of Lemma 5.15, we suppress the dependence of x,N,γ and d of $\mathcal{S}_x$ and γ,δ,η,τ and d from upper-sufficient, lower-sufficient and good. Before we commence with the proof, we give the definition of $\underline{N}_{5.17}$. Firstly, by Lemma 5.16 we know that there exist an $\underline{N}_1=\underline{N}_1(\eta)$ such that (5.71) holds if $N\ge\underline{N}_1$ (which also implies that (5.84) holds for boxes of layer k). Furthermore, because $k<k_\varepsilon^\star(N)$ we may utilise item (2) of Lemma 5.7 (properties of sidelengths) to see that there exists an $\underline{N}_2=\underline{N}_2(\varepsilon,\gamma)$ such that

$$N^{\gamma^j}\le r_j\le 2N^{\gamma^j} \tag{5.85}$$

for all $j\le k$ if $N\ge\underline{N}_2$. Next, by Claim 5.12 there exists an $\underline{N}_3=\underline{N}_3(\varepsilon,\gamma,\delta,\eta,d)$ such that if $N\ge\underline{N}_3$ then all items of Claim 5.12 hold. Furthermore, because $k<k_\varepsilon^\star(N)$ and by recalling the definition of $(r_i)_{i\le k}$ from Definition 5.4, by applying Claim 5.6 to (5.85) we may observe that all the r_i grow rapidly with as i decreases. Furthermore, by (5.51) of item (2) of Claim 5.12, we note that all $\underline{M}_i$ and $\overline{M}_i$ are bounded if N is large. As such, we may observe that there exists an $\underline{N}_4=\underline{N}_4(\varepsilon,\gamma,\delta,\eta,\tau,d)$ such that if $N\ge\underline{N}_4$, then all of the intervals

$$\left[1,\eta^{-\frac{1}{\tau-1}}\right],\qquad\text{and}\qquad\left[\underline{M}_j^{\frac{1}{\tau-1}}r_j^{d\frac{1-\delta}{\tau-1}},\underline{M}_j^{\frac{1}{\tau-1}}r_j^{d\frac{1-\delta}{\tau-1}}\right],\ \text{where } j\le k, \tag{5.86}$$

are all disjoint. Lastly, we apply the same reasoning we have done to show (5.83) in the proof of Lemma 5.16. Notice that if we set

$$C_{5.17}=C_{5.17}(\eta):=\frac{1}{2^{1+d(1+3\delta)}}\wedge\frac{\eta}{32}\wedge\frac{\eta^2}{192-48\eta)}>0, \tag{5.87}$$

then there exists an $\underline{N}_5=\underline{N}_5(\varepsilon,\gamma,\delta,\eta,d)$ such that if $N\ge\underline{N}_5$, then

$$1-3\exp[-C_{5.17}r_j^{d\delta}]\ge 1-r_j^{-d\delta} \tag{5.88}$$

for all $j\le k$.

Next, we fix $\underline{N}_1,\underline{N}_2,\underline{N}_3,\underline{N}_4,\underline{N}_5$ and set $\underline{N}_{5.17}=\underline{N}_1\vee\underline{N}_2\vee\underline{N}_3\vee\underline{N}_4\vee\underline{N}_5$. In the remainder of this proof, we assume that $N\ge\underline{N}_{5.17}$, and hence all of the above holds.

To show that (5.84) holds for all $i\le k$, we apply (backwards) induction on layer i. Because $N\ge\underline{N}_{5.17}\ge\underline{N}_1$, we know that by Lemma 5.16 the base case is true: (5.84) holds for layer k. Next, we show that if the result holds for all layers $j=i+1,\dots,k$, then it also holds for any box of layer i. Fix any box $\mathcal{B}_{\mathbf{s}_i}^i$ of layer i. By writing out the definition of good (see Definition 5.14), we obtain that

$$\mathbb{P}(\mathcal{B}_{\mathbf{s}_i}^i\text{ is good})=\mathbb{P}\big(\mathcal{B}_{\mathbf{s}_i}^i\text{ is upper-sufficient}\;\big|\;\mathcal{B}_{\mathbf{s}_i}^i\text{ is lower-sufficient}\big)\mathbb{P}\big(\mathcal{B}_{\mathbf{s}_i}^i\text{ is lower-sufficient}\big) \tag{5.89}$$

The proof consists of four subsequent parts. In the first part, we bound $\mathbb{P}\big(\mathcal{B}_{\mathbf{s}_i}^i\text{ is lower-sufficient}\big)$. In the second part and third part, we compute $\mathbb{P}\big(\mathcal{B}_{\mathbf{s}_i}^i\text{ is upper-sufficient}\;\big|\;\mathcal{B}_{\mathbf{s}_i}^i\text{ is lower-sufficient}\big)$. In the fourth part, we bring everything together.

(1: Lower sufficiency) We compute a lower bound for the probability that $\mathcal{B}_{\mathbf{s}_i}^i$ is lower-sufficient. Call $\widetilde{\Gamma}$ the amount of good sub-boxes of layer $i+1$ in $\mathcal{B}_{\mathbf{s}_i}^i$. There are exactly $(r_i/r_{i+1})^d$ such sub-boxes by construction (see Definition 5.9). Furthermore, each of these sub-boxes are independently good with a probability lower-bounded by $1-r_{i+1}^{-d\delta}$ when we apply the induction hypothesis (see (5.84)). We conclude that $\widetilde{\Gamma}$ stochastically dominates Γ, where Γ is defined by

$$\Gamma\sim\mathrm{Bin}\left(\frac{r_i^d}{r_{i+1}^d},1-r_{i+1}^{-d\delta}\right). \tag{5.90}$$

By applying the Chernoff bound (Corollary B.4), we may compute

$$\begin{aligned}\mathbb{P}\left(\mathcal{B}_{\mathbf{s}_i}^i\text{ is lower-sufficient}\right)&=\mathbb{P}\left(\widetilde{\Gamma}\ge\left(1-2r_{i+1}^{-d\delta}\right)\frac{r_i^d}{r_{i+1}^d}\right)\\&\ge\mathbb{P}\left(\Gamma\ge\left(1-\left(1-\frac{1-2r_{i+1}^{-d\delta}}{1-r_{i+1}^{-d\delta}}\right)\right)(1-r_{i+1}^{-d\delta})\frac{r_i^d}{r_{i+1}^d}\right)\\&\ge 1-\exp\left[-\frac{1}{2}\left(1-\frac{1-2r_{i+1}^{-d\delta}}{1-r_{i+1}^{-d\delta}}\right)^2(1-r_{i+1}^{-d\delta})\frac{r_i^d}{r_{i+1}^d}\right].\end{aligned} \tag{5.91}$$

In the first line we have used the definition of lower-sufficient, in the second line we have used that $\widetilde{\Gamma}$ stochastically dominates Γ and rewritten to match the format of the Chernoff bound and in the third line we have applied the Chernoff bound. Our goal is to further bound the last quantity in (5.91). To this end, notice that $1-(1-2r_{i+1}^{-d\delta})/(1-r_{i+1}^{-d\delta}) = r_{i+1}^{-d\delta}/(1-r_{i+1}^{-d\delta})$ and $0 < 1-r_{i+1}^{-d\delta} < 1$. Substituting this into (5.91) and rewrite to see that

$$\mathbb{P}\left(\mathcal{B}^i_{\mathbf{s}_i} \text{ is lower-sufficient}\right) = 1-\exp\left[-\frac{1}{2}\frac{r_{i+1}^{-2d\delta}}{1-r_{i+1}^{-d\delta}}\frac{r_i^d}{r_{i+1}^d}\right] \geq 1-\exp\left[-\frac{1}{2}r_{i+1}^{-d(1+2\delta)}r_i^d\right]. \tag{5.92}$$

Furthermore, by the definition of δ (see Setting 5.3), we know that $\delta < (1-\gamma)/(1+2\gamma)$, which we may rewrite to $\delta < 1-\gamma(1+2\delta)$. Combining this with Lemma 5.7 (properties of sidelengths) and the fact that $N \geq \underline{N}_{5.17} \geq \underline{N}_2$, we find that

$$\frac{r_i^d}{r_{i+1}^{d(1+2\delta)}} \geq \frac{N^{d\gamma^i}}{2^{d(1+2\delta)}N^{d(1+2\delta)\gamma^{i+1}}} = \frac{1}{2^{d(1+2\delta)}}N^{d(1-\gamma(1+2\delta))\gamma^i} \geq \frac{1}{2^{d(1+2\delta)}}N^{d\delta\gamma^i} \geq \frac{1}{2^{d(1+3\delta)}}r_i^{d\delta}. \tag{5.93}$$

Notice that we have applied Lemma 5.7 twice in the left-most inequality and once in the right-most inequality. Applying all the above to (5.92) we obtain

$$\mathbb{P}\left(\mathcal{B}^i_{\mathbf{s}_i} \text{ is lower-sufficient}\right) \geq 1-\exp\left[-\frac{1}{2^{1+d(2+2\delta)}}r_i^{d\delta}\right] =: 1-\mathrm{err}_1 \tag{5.94}$$

which gives a lower bound for the probability that $\mathcal{B}^i_{\mathbf{s}_i}$ is lower-sufficient.

(2: Bounds for the amount of vertices that have already been revealed if $\mathcal{B}^i_{\mathbf{s}_i}$ is lower-sufficient) Now that we have found the probability that $\mathcal{B}^i_{\mathbf{s}_i}$ is lower-sufficient, we want to bound the probability that it is also upper-sufficient, given that it is lower-sufficient. This is analogous to the proof of Lemma 5.16. We have split this part of the proof into two parts. In this part of the proof, we bound the amount of vertices that can still contribute to $\mathcal{B}^i_{\mathbf{s}_i}$ being upper-sufficient, given that $\mathcal{B}^i_{\mathbf{s}_i}$ is lower-sufficient. In the next part, we use these bounds to bound the probability that $\mathcal{B}^i_{\mathbf{s}_i}$ is upper-sufficient, given that it is lower-sufficient.
Recall that for us to know $\mathcal{B}^i_{\mathbf{s}_i}$ to be lower-sufficient, we necessarily need to have revealed information about the weight of vertices of $\mathcal{B}^i_{\mathbf{s}_i}$. More precisely, we need to have revealed that there is a division of vertices: one portion of the vertices has weight either between 1 and $\eta^{-1/(\tau-1)}$ or between $\underline{M}_j^{1/(\tau-1)}r_j^{d(1-\delta)/(\tau-1)}$ and $\overline{M}_j^{1/(\tau-1)}r_j^{d(1-\delta)/(\tau-1)}$ for some $j = i+1,\dots,k$; the other portion consists of vertices that have a different weight. The first portion is exactly $\#T\big(\mathcal{B}^i_{\mathbf{s}_i},\mathscr{W}_{i+1}\big)$, where

$$\mathscr{W}_{i+1} := \left[1,\eta^{-\frac{1}{\tau-1}}\right] \cup \bigcup_{j=i+1}^{k}\left[\underline{M}_j^{\frac{1}{\tau-1}}r_j^{d\frac{1-\delta}{\tau-1}},\overline{M}_j^{\frac{1}{\tau-1}}r_j^{d\frac{1-\delta}{\tau-1}}\right] \tag{5.95}$$

and $\#T$ is as in (5.64) of Definition 5.13. The second portion is $\mathcal{B}^i_{\mathbf{s}_i}\backslash\#T\big(\mathcal{B}^i_{\mathbf{s}_i},\mathscr{W}_{i+1}\big)$. Because $N \geq \underline{N}_{5.17} \geq \underline{N}_4$ and (5.86), only the vertices of the second portion can have weight between $\underline{M}_i^{1/(\tau-1)}r_i^{d(1-\delta)/(\tau-1)}$ and $\overline{M}_i^{1/(\tau-1)}r_i^{d(1-\delta)/(\tau-1)}$ and hence can contribute to $\mathcal{B}^i_{\mathbf{s}_i}$ being upper-sufficient. As such, we bound the first portion of vertices and note that since both portions together make up $\mathcal{B}^i_{\mathbf{s}_i}$, this also bounds the second portion.

Throughout this entire part, assume that $\mathcal{B}^i_{\mathbf{s}_i}$ is lower-sufficient. We start with a lower-bound for $\#T\big(\mathcal{B}^i_{\mathbf{s}_i},\mathscr{W}_{i+1}\big)$. We firstly lower-bound the constant-weight vertices, i.e., those vertices that have weight between 1 and $\eta^{-1/(\tau-1)}$. There are exactly $\#T\big(\mathcal{B}^i_{\mathbf{s}_i},[1,\eta^{-1/(\tau-1)}]\big)$ such vertices. To find a lower-bound, we firstly count the fewest good layer j sub-boxes that are contained within $\mathcal{B}^i_{\mathbf{s}_i}$. By recalling the definition of lower-sufficient from (5.67), we see that $\mathcal{B}^i_{\mathbf{s}_i}$ must contain at least $(1-2r_{i+1}^{-d\delta})(r_i/r_{i+1})^d$ good sub-boxes of layer $i+1$. Since these sub-boxes are good, they are also lower-sufficient. It follows that each good sub-box of layer $i+1$ contains at least $(1-2r_{i+1}^{-d\delta})(r_{i+1}/r_{i+2})^d$ sub-boxes of layer $i+2$. We iterate this procedure to see that

$$\begin{array}{l}\text{the minimal amount of good sub-boxes of layer } j\\ \text{contained solely in good sub-boxes of } \mathcal{B}^i_{\mathbf{s}_i}\end{array} = \prod_{\ell=i+1}^{j}\left(1-2r_\ell^{-d\delta}\right)\frac{r_{\ell-1}^d}{r_\ell^d}, \tag{5.96}$$

for any $j = i+1,\dots,k$. Since each good sub-box of layer k contains at least $\underline{\eta}r_k^d$ vertices by definition (see (5.66) of Definition 5.14), we thus conclude that

$$\#T\big(\mathcal{B}^i_{\mathbf{s}_i},[1,\eta^{-1/(\tau-1)}]\big) \geq \underline{\eta}r_k^d\prod_{\ell=i+1}^{k}\left(\left(1-2r_\ell^{-d\delta}\right)\frac{r_{\ell-1}^d}{r_\ell^d}\right) = \underline{\eta}\left[\prod_{\ell=i+1}^{k}\left(1-2r_\ell^{-d\delta}\right)\right]r_i^d. \tag{5.97}$$

Here we have ignored that the bad sub-boxes of $\mathcal{B}^i_{\mathbf{s}_i}$ may also contribute with vertices that have weight between 1 and $\eta^{-1/(\tau-1)}$.

We reason similarly for vertices with weight between $\underline{M}_j^{1/(\tau-1)}r_j^{d(1-\delta)/(\tau-1)}$ and $\overline{M}_j^{1/(\tau-1)}r_j^{d(1-\delta)/(\tau-1)}$, where $j = i+1,\dots,k$. Combining (5.96) and the fact that each good box of layer j contains at least $\underline{a}_jr_j^{d\delta}$ vertices with the required weight yields

$$\#T\left(\mathcal{B}^i_{\mathbf{s}_i},\left[\underline{M}_j^{\frac{1}{\tau-1}}r_j^{d\frac{1-\delta}{\tau-1}},\overline{M}_j^{\frac{1}{\tau-1}}r_j^{d\frac{1-\delta}{\tau-1}}\right]\right) \geq \underline{a}_j\left[\prod_{\ell=i+1}^{j}\left(1-2r_\ell^{-d\delta}\right)\right]r_j^{-d(1-\delta)}r_i^d \tag{5.98}$$

for all $j \in \{i+1,\dots,k\}$. Again, we ignore the fact that the bad sub-boxes may also contain any vertices with weight between $\underline{M}_j^{1/(\tau-1)} r_j^{d(1-\delta)/(\tau-1)}$ and $\overline{M}_j^{1/(\tau-1)} r_j^{d(1-\delta)/(\tau-1)}$ for all $j \in \{i+1,\dots,k\}$. Combining (5.97), (5.98), the fact that $\mathscr{W}_{i+1}$ consists of disjoint intervals because $N \geq \underline{N}_{5.17} \geq \underline{N}_4$ and summing over $j = i+1,\dots,k$ we obtain

$$\begin{aligned}\#T\big(\mathcal{B}^i_{\mathbf{s}_i}, \mathscr{W}_{i+1}\big) &\geq \underline{\eta}\left[\prod_{j=i+1}^{k}\Big(1-2r_j^{-d\delta}\Big)\right] r_i^d + \sum_{j=i+1}^{k} \underline{a}_j \left[\prod_{\ell=i+1}^{j}\Big(1-2r_\ell^{-d\delta}\Big)\right] r_j^{-d(1-\delta)} r_i^d \\ &= (1-\overline{A}_i) r_i^d. \end{aligned} \tag{5.99}$$

Here $\overline{A}_i$ is as defined in (5.48) from Definition 5.11. We emphasise that (5.99) holds because we have conditioned on $\mathcal{B}^i_{\mathbf{s}_i}$ being lower-sufficient.

Next, we upper-bound the amount of vertices that have weight in $\mathscr{W}_{i+1}$. To this end, we approach layer by layer. Suppose that a box of layer j, $j = i,\dots,k-1$ is good. Then it must contain at least $(1 - 2r_j^{-d\delta})(r_j/r_{j+1})^d$ good sub-boxes by definition. The remaining $2r_j^{-d\delta}(r_j/r_{j+1})^2$ sub-boxes can either be

- bad, in which case it may consist fully of r_{j+1}^d vertices with weight in $\mathscr{W}_{i+1}$, or
- good, in which case it must consist of fewer than r_{j+1}^d vertices with weight in $\mathscr{W}_{i+1}$.

We therefore endeavour to find the largest amount of sub-boxes that are bad. We start with layer i. By the above reasoning, we know that we search for the maximum amount of bad sub-boxes of layer $i+1$, of which there are at most $2r_{i+1}^{-d\delta}(r_i/r_{i+1})^d$ by (5.96). We do not continue with the bad sub-boxes of layer $i+1$ to avoid double counting. We do continue with the fewest number of good sub-boxes of layer $i+1$, of which there are at least $(1-2r_{i+1}^{-d\delta})(r_i/r_{i+1})^d$. Then each of these sub-boxes of layer $i+1$ contains maximally $2r_{i+2}^{-d\delta}(r_{i+1}/r_{i+2})^d$ bad sub-boxes of layer $i+2$, so in total there are at most $(1-2r_{i+1}^{-d\delta})(r_i/r_{i+1})^d \cdot 2r_{i+2}^{-d\delta}(r_{i+1}/r_{i+2})^d$ sub-boxes of layer $i+2$ that we add to the amount of bad sub-boxes. More generally, we take a layer j, $j = i+1,\dots,k$ and assume that in all layers from i to $j-2$ we have already found the maximum amount of bad sub-boxes. Then by (5.96), there are $\prod_{\ell=i+1}^{j-1}((1-2r_\ell^{-d\delta})(r_{\ell-1}/r_\ell)^d)$ remaining good sub-boxes of layer $j-1$. Each of these good sub-boxes of layer $j-1$ may contain at most $2r_j^{-d\delta}(r_{j-1}/r_j)^d$ sub-boxes of layer j that are bad. We find that

$$\#\left\{\begin{matrix}\text{bad boxes of layer } j \text{ contained within}\\ \text{good sub-boxes of layer } j-1 \text{ in } \mathcal{B}^i_{\mathbf{s}_i}\end{matrix}\right\} \leq \left[\prod_{\ell=i+1}^{j-1}\Big(1-2r_\ell^{-d\delta}\Big)\frac{r_{\ell-1}^d}{r_\ell^d}\right] 2r_j^{-d\delta}\frac{r_{j-1}^d}{r_j^d}. \tag{5.100}$$

Each of these boxes consists of r_j^d vertices. By summing over $j = i+1,\dots,k$ and rewriting we find that

$$\text{the amount of vertices contained in bad sub-boxes of } \mathcal{B}^i_{\mathbf{s}_i} \leq \sum_{j=i+1}^{k}\left[\prod_{\ell=i+1}^{j-1}\Big(1-2r_\ell^{-d\delta}\Big)\right] 2r_j^{-d\delta} r_i^d \tag{5.101}$$

Next, we count the maximal amount of vertices that have weight in $\mathscr{W}_{i+1}$ that are in the remaining good sub-boxes. The computation is exactly the same for the lower bound from earlier in this part of the proof, but with $\underline{\eta}$ and $\underline{a}_j$ replaced by $\overline{\eta}$ and $\overline{a}_j$, respectively.

Combining everything and using the same reasoning done in finding the lower-bound, we obtain

$$\begin{aligned}\#T\big(\mathcal{B}^i_{\mathbf{s}_i}, \mathscr{W}_{i+1}\big) &\leq \underline{\eta}\left[\prod_{j=i+1}^{k}\Big(1-2r_j^{-d\delta}\Big)\right] r_i^d + \sum_{j=i+1}^{k} \underline{a}_j \left[\prod_{\ell=i+1}^{j}\Big(1-2r_\ell^{-d\delta}\Big)\right] r_j^{-d(1-\delta)} r_i^d \\ &\quad + \sum_{j=i+1}^{k}\left[\prod_{\ell=i+1}^{j-1}\Big(1-2r_\ell^{-d\delta}\Big)\right] 2r_j^{-d\delta} r_i^d \\ &= \left(1-\underline{A}_i\right) r_i^d \end{aligned} \tag{5.102}$$

Again, we emphasise that this result holds when conditioned on $\mathcal{B}^i_{\mathbf{s}_i}$ being lower-sufficient.

(3: upper-sufficiency) Now that we know bounds for the amount of vertices that potentially have a weight between $\underline{M}_i^{1/(\tau-1)} r_i^{d(1-\delta)/(\tau-1)}$ and $\overline{M}_i^{1/(\tau-1)} r_i^{d(1-\delta)/(\tau-1)}$ if $\mathcal{B}^i_{\mathbf{s}_i}$ is lower-sufficient, we may continue with computing a bound for the probability that $\mathcal{B}^i_{\mathbf{s}_i}$ is upper-sufficient, given that it is lower-sufficient. We do this in way that is similar to the proof of Lemma 5.16. To this end, note that $N \geq \underline{N}_{5.17} \geq \underline{N}_3 \vee \underline{N}_4$, which by item (2) of Claim 5.12 implies that

$$\frac{1}{\underline{M}_i} - \frac{1}{\overline{M}_i} = \eta - \sum_{j=i+1}^{k}\left(\frac{1}{\underline{M}_j} - \frac{1}{\overline{M}_j}\right) r_j^{-d(1-\delta)}. \tag{5.103}$$

Furthermore, by (5.86), we also obtain

$$\Big[\underline{M}_i^{\frac{1}{\tau-1}} r_i^{d\frac{1-\delta}{\tau-1}}, \overline{M}_i^{\frac{1}{\tau-1}} r_i^{d\frac{1-\delta}{\tau-1}}\Big] \cap \mathscr{W}_{i+1} = \varnothing. \tag{5.104}$$

Moreover, by (3.4) (see also the proof of Claim A.3) we find that $\mathbb{P}(W \notin \mathscr{W}_{i+1})$ is equal to the right-hand side of (5.103). Combining the above three observations, we find that

$$\mathbb{P}\Big(W \in \Big[\underline{M}_i^{\frac{1}{\tau-1}} r_i^{d\frac{1-\delta}{\tau-1}}, \overline{M}_i^{\frac{1}{\tau-1}} r_i^{d\frac{1-\delta}{\tau-1}}\Big] \;\Big|\; W \notin \mathscr{W}_{i+1}\Big) = \frac{\Big(\frac{1}{\underline{M}_i} - \frac{1}{\overline{M}_i}\Big) r_i^{-d(1-\delta)}}{\eta - \sum_{j=i+1}^{k} \Big(\frac{1}{\underline{M}_j} - \frac{1}{\overline{M}_j}\Big) r_j^{-d(1-\delta)}} = r_i^{-d(1-\delta)} \tag{5.105}$$

We note that each of the vertices that do *not* have weight in $\mathscr{W}_{i+1}$, each independently have probability given by (5.105) that their weight is between $\underline{M}_i^{1/(\tau-1)} r_i^{d(1-\delta)/(\tau-1)}$ and $\underline{M}_i^{1/(\tau-1)} r_i^{d(1-\delta)/(\tau-1)}$. Furthermore, by Lemma 5.10 the box $\mathcal{B}^i_{\mathbf{s}_i}$ contains r_i^d vertices, of which $\#T\big(\mathcal{B}^i_{\mathbf{s}_i}, \mathscr{W}_{i+1}\big)$ are vertices that have weight within $\mathscr{W}_{i+1}$. We may therefore define

$$X := \#T\Big(\mathcal{B}^i_{\mathbf{s}_i}, \Big[\underline{M}_i^{1/(\tau-1)} r_i^{d(1-\delta)/(\tau-1)}, \underline{M}_i^{1/(\tau-1)} r_i^{d(1-\delta)/(\tau-1)}\Big]\Big) \tag{5.106}$$

and observe that

$$X \sim \mathrm{Bin}\big(r_i^d - \#T\big(\mathcal{B}^i_{\mathbf{s}_i}, \mathscr{W}_{i+1}\big), r_i^{-d(1-\delta)}\big). \tag{5.107}$$

Then, by (5.99) and (5.102) we observe that

$$(X|\mathcal{B}^i_{\mathbf{s}_i} \text{ is lower-sufficient}) \overset{d}{\geq} \underline{X}, \qquad \text{where} \qquad \underline{X} \sim \mathrm{Bin}\big(\underline{A}_i r_i^d, r_i^{-d(1-\delta)}\big) \tag{5.108}$$

and

$$(X|\mathcal{B}^i_{\mathbf{s}_i} \text{ is lower-sufficient}) \overset{d}{\leq} \overline{X}, \qquad \text{where} \qquad \overline{X} \sim \mathrm{Bin}\big(\overline{A}_i r_i^d, r_i^{-d(1-\delta)}\big). \tag{5.109}$$

Next, we define

$$Q_i := \mathbb{P}\big(\mathcal{B}^i_{\mathbf{s}_i} \text{ is not upper-sufficient} \;\big|\; \mathcal{B}^i_{\mathbf{s}_i} \text{ is lower-sufficient}\big). \tag{5.110}$$

We then apply the definition of upper-sufficient (see 5.65 of Definition 5.14), equations (5.108) and (5.109) and rewrite

$$\begin{aligned} Q_i &= \mathbb{P}\big(X < \underline{a}_i r_i^{d\delta} \;\big|\; \mathcal{B}^i_{\mathbf{s}_i} \text{ is lower-sufficient}\big) + \mathbb{P}\big(X > \overline{a}_i r_i^{d\delta} \;\big|\; \mathcal{B}^i_{\mathbf{s}_i} \text{ is lower-sufficient}\big) \\ &\leq \mathbb{P}\big(\underline{X} < \underline{a}_i r_i^{d\delta}\big) + \mathbb{P}\big(\overline{X} > \overline{a}_i r_i^{d\delta}\big). \end{aligned} \tag{5.111}$$

We apply the Chernoff bound (see Lemma B.4) to both probabilities in (5.111) separately. Recall that $\underline{a}_i = \underline{A}_i/2 < \underline{A}_i$ and $\overline{a}_i = (\overline{A}_i + 1)/2 > \overline{A}_i$ by Definition 5.11 and item (3) of Lemma 5.12 (which we may apply since $N \geq N_{5.17} \geq \underline{N}_3$). Substituting these, recalling the definitions of $\underline{X}$ from (5.108) and $\overline{X}$ from (5.109) and rewriting in the format of the Chernoff bound, yields

$$\mathbb{P}\big(\underline{X} < \underline{a}_i r_i^{d\delta}\big) = \mathbb{P}\big(\underline{X} < (1 - 1/2)\underline{A}_i r_i^{d\delta}\big) \leq \exp\Big[-\frac{1}{8}\underline{A}_i r_i^{d\delta}\Big] \tag{5.112}$$

and

$$\mathbb{P}\big(\overline{X} > \overline{a}_i r_i^{d\delta}\big) = \mathbb{P}\Big(\overline{X} > \Big(1 + \frac{1 - \overline{A}_i}{2\overline{A}_i}\Big)\overline{A}_i r_i^{d\delta}\Big) \leq \exp\Big[-\frac{1}{12}\frac{(1-\overline{A}_i)^2}{\overline{A}_i} r_i^{d\delta}\Big]. \tag{5.113}$$

Then notice that, again by item (3) of Lemma 5.12, we have that $\underline{A}_i \geq \eta/4$ and $\overline{A}_i \leq 1 - \eta/4$. Furthermore, notice that $x \mapsto (1-x)^2/x$ is decreasing if $x \in (0,1)$. Combining this with (5.111), (5.112) and (5.113) and some elementary bounds then yields

$$Q_i \leq \exp\Big[-\frac{1}{8}\underline{A}_i r_i^{d\delta}\Big] + \exp\Big[-\frac{1}{12}\frac{(1-\overline{A}_i)^2}{\overline{A}_i} r_i^{d\delta}\Big] \leq 2\exp\Big[-\Big(\frac{\eta}{32} \wedge \frac{\eta^2}{192 - 48\eta}\Big) r_i^{d\delta}\Big] =: \mathrm{err}_2. \tag{5.114}$$

(4: Conclusion) We return to (5.89). By combining equations (5.94) and (5.114) we may now compute a lower bound for the probability that $\mathcal{B}^i_{\mathbf{s}_i}$ is good:

$$\begin{aligned} \mathbb{P}(\mathcal{B}^i_{\mathbf{s}_i} \text{ is good}) &= \mathbb{P}\left(\mathcal{B}^i_{\mathbf{s}_i} \text{ is upper-sufficient} \middle| \mathcal{B}^i_{\mathbf{s}_i} \text{ is lower-sufficient}\right) \mathbb{P}\left(\mathcal{B}^i_{\mathbf{s}_i} \text{ is lower-sufficient}\right) \\ &\geq (1 - \mathrm{err}_1)(1 - \mathrm{err}_2) \geq 1 - 3\exp(-C_{5.17} r_i^{d\delta}) \geq 1 - r_i^{-d\delta}. \end{aligned} \tag{5.115}$$

Here we have used that $(1-\mathrm{err}_1)(1-\mathrm{err}_2) \geq 1 - \mathrm{err}_1 - \mathrm{err}_2$ since $\mathrm{err}_1, \mathrm{err}_2 > 0$ and the fact that $N \geq \underline{N}_{5.17}$, so that (5.88) holds. By induction, this finishes the proof of (5.84). Furthermore, throughout this proof we have only used the weight of vertices that are within $\mathcal{B}^i_{\mathbf{s}_i}$ (recall that all sub-boxes of $\mathcal{B}^i_{\mathbf{s}_i}$ are fully contained within $\mathcal{B}^i_{\mathbf{s}_i}$). We conclude that the event in (5.84) is independent of all vertices within $\mathcal{V}\backslash\mathcal{B}^i_{\mathbf{s}_i}$. This finishes the proof. ■

Since r_i is increasing when i becomes smaller, the result in (5.84) suggests that the probability of a larger layer being good is higher than those of the lower layers. At first glance, this might seem counter-intuitive: the larger boxes depend on the smaller boxes. It is, however, to be expected when one considers that the larger boxes contain (much more) vertices than the smaller boxes. With the law of large numbers in mind, it is therefore to be expected that the large boxes deviate relatively little from their 'expected' behaviour (i.e., the distribution of the weight of the vertices) when compared to the smaller boxes.

Throughout the proof of Lemma 5.17, with (5.96) we have made one observation that will be useful in later proofs. We emphasise this result in Corollary 5.18

Corollary 5.18 *Consider Setting 5.3, in particular γ, δ and η. Let $N > 1$ and let k be a (possibly N dependent) positive integer. Let $(r_i)_{i\le k}$ be the sequence of sidelengths from Definition 5.4. Consider the boxing structure $\mathcal{S}_x(N,k,\gamma,d)$ around a vertex $x \in \mathcal{V}$ from Definition 5.9 and $(\gamma,\delta,\eta,\tau,d)$-good from Definition 5.14. Let $\mathcal{B}^{i,x}_{\mathbf{s}_i} \in \mathcal{S}_x(N,k,\gamma,d)$ be any box of layer $i \in \{1,\dots,k-1\}$ and assume that it is $(\gamma,\delta,\eta,\tau,d)$-good. Then*

$$\#\left\{\begin{matrix}(\gamma,\delta,\eta,\tau,d)\text{-good sub-boxes of } \mathcal{B}^{i,x}_{\mathbf{s}_i} \text{ of layer } j \text{ that are contained}\\ \text{in } (\gamma,\delta,\eta,\tau,d)\text{-good sub-boxes of } \mathcal{B}^{i,x}_{\mathbf{s}_i} \text{ of all layers } \ell = i,\dots,j\end{matrix}\right\} \ge \frac{r_i^d}{r_j^d}\prod_{\ell=i+1}^{j}\left(1-2r_\ell^{-d\delta}\right). \quad (5.116)$$

Proof. This follows immediately from (5.96) of part 2 of the proof of Lemma 5.17. ∎

In defining the nets, it will turn out to be very useful to be able to reference which box of the boxing structure a vertex is in. To this end, we define a labelling function, which for each vertex and each layer i returns the label of the box of layer i that the vertex is in.

Definition 5.19 *Consider Setting 5.3, particularly γ. Let $N > 1$ and k be a (possibly N-dependent) positive integer. Let $v \in \mathcal{V}$ and consider the boxing structure $\mathcal{S}_u = \mathcal{S}_u(N,k,\gamma,d)$ around u from Definition 5.9. For $x \in \mathcal{B}^{1,u}$ and $i \in \{1,\dots,k\}$, we let $\Lambda^i_u(u)$ be the label of the box of layer i in the boxing structure that contains u, i.e.,*

$$\text{if}\quad x \in \mathcal{B}^{i,u}_{\mathbf{s}_i} = \mathcal{B}^{i,u}_{\mathbf{s}_i}(N,\gamma,d) \quad\text{then}\quad \Lambda^i_u(x) = \mathbf{s}_i. \quad (5.117)$$

For $x \in \mathcal{B}^{1,u}$ and $i \in \{1,\dots,k\}$, we let $\mathcal{B}^i_\Lambda(x)$ denote the box of layer i that contains x, i.e.,

$$\mathcal{B}^i_\Lambda(x) = \mathcal{B}^{i,u}_{\Lambda^i_u(x)}. \quad (5.118)$$

Furthermore, suppose the setting is such that an additional boxing structure $\mathcal{S}_v = \mathcal{S}_v(N,k,\gamma,d)$ around v present, satisfying that for all $i,j = 1,\dots,k$, for all $\mathcal{B}^{i,u}_{\mathbf{s}_i} \in \mathcal{S}_u$ and $\mathcal{B}^{j,v}_{\mathbf{s}_j} \in \mathcal{S}_v$ it holds that $\mathcal{B}^{i,u}_{\mathbf{s}_i} \cap \mathcal{B}^{j,v}_{\mathbf{s}_j} = \varnothing$. For $x \in \mathcal{B}^{1,u} \cup \mathcal{B}^{1,v}$ and $i \in \{1,\dots,k\}$, we then define $\mathcal{B}^i_\Lambda(x)$ in the following way:

$$\mathcal{B}^i_\Lambda(x) = \begin{cases}\mathcal{B}^{i,u}_{\Lambda^i_u(x)} & \text{if } x \in \mathcal{B}^{1,u}\\ \mathcal{B}^{i,v}_{\Lambda^i_v(x)} & \text{if } x \in \mathcal{B}^{1,v}\end{cases}. \quad (5.119)$$

By convention $\mathcal{B}^i_\Lambda(\varnothing) = \varnothing$.

We discuss Definition 5.19, starting with $\Lambda^i_u(x)$. Notice that $\mathcal{B}^{1,v}$ is a superset of every other box in the boxing structure $\mathcal{S}_v$. Furthermore, by definition of the boxing structure $\mathcal{S}_u$ (see Definition 5.9), each box of the boxing structure is divided exactly into sub-boxes. Therefore, every vertex in $\mathcal{B}^{1,u}$ is contained in exactly one box of layer i, which means that $\Lambda^i_u(x)$ is well-defined. It follows that $\mathcal{B}^i_\Lambda(x)$ is also well-defined in the setting with only one boxing structure. Furthermore, it is obvious that $x \in \mathcal{B}^i_\Lambda(x)$ and $\mathcal{B}^i_\Lambda(x) \supset \mathcal{B}^{i+1}_\Lambda(x)$ when we consider the way the boxing structure is defined (see Definition 5.9).
Lastly, we briefly discuss $\mathcal{B}^i_\Lambda(x)$ in the setting where there are two boxing structures $\mathcal{S}_u, \mathcal{S}_v$ present. In this case, again $\mathcal{B}^{1,u} \cup \mathcal{B}^{1,v}$ is a superset of all boxes of $\mathcal{S}_u \cup \mathcal{S}_v$. Therefore, if the boxing structures are disjoint (i.e., each box of $\mathcal{S}_u$ is disjoint of each box of $\mathcal{S}_v$), there is exactly one level i box in either $\mathcal{S}_u$ or $\mathcal{S}_v$ that contains $x \in \mathcal{B}^{1,u} \cup \mathcal{B}^{1,v}$. Therefore, in this setting $\mathcal{B}^i_\Lambda(x)$ is again well-defined. Again, it is easily verifiable that $x \in \mathcal{B}^i_\Lambda(x)$ and $\mathcal{B}^i_\Lambda(x) \supset \mathcal{B}^{i+1}_\Lambda(x)$ also holds in this setting.

We continue with giving the proper definition of a net. Informally, a net is the set of all vertices u such that all boxes that contain u are good.

Definition 5.20 (Nets) *Consider Setting 5.3, particularly γ, δ, and η. Fix $N > 1$ and let k be a (possibly N-dependent) positive integer. Let $v \in \mathcal{V}$ and consider the boxing structure $\mathcal{S}_v = \mathcal{S}_v(N,k,\gamma,d)$ around v, let $\mathcal{B}^{1,v} \in \mathcal{S}_v$ be the layer 1 box of $\mathcal{S}_v$ and let $(\gamma,\delta,\eta,\tau,d)$-good be as in Definition 5.14. Furthermore, let $\mathcal{B}^i_\Lambda(x)$ be as defined in (5.118) of Definition 5.19. Then set*

$$\mathcal{N}_v = \mathcal{N}_v(N,k,\gamma,\delta,\eta,\tau,d) = \left\{u \in \mathcal{B}^{1,v} : \mathcal{B}^i_\Lambda(u) \text{ is } (\gamma,\delta,\eta,\tau,d)\text{-good for all } i = 1,\dots,k\right\}. \quad (5.120)$$

*We call $\mathcal{N}_v(N,k,\gamma,\delta,\eta,\tau,d)$ an $(N,k,\gamma,\delta,\eta,\tau,d)$-**net** for v if $v \in \mathcal{N}_v(N,k,\gamma,\delta,\eta,\tau,d)$. When $N,k,\gamma,\delta,\eta,\tau$ and d are clear from context, we write that $\mathcal{N}_v$ is a net for v.*

Notice that if $\mathcal{N}_v$ is a net for v, then every box of the boxing structure $\mathcal{S}_v$ in which v is contained, must be good. In particular, then $\mathcal{B}^{1,v}$ must be good. Therefore, if $v \in \mathcal{N}_v$ then by Corollary 5.18 there are at least $(r_1/r_k)^d\prod_{j=2}^k(1-2r_j^{-d\delta})$ boxes of layer k that are contained within good boxes of every layer. Since a box of layer k contains r_k^d vertices by Claim 5.10, there therefore must be at least $r_1^d\prod_{j=2}^k(1-2r_j^{-d\delta})$ vertices in the net.

We continue by showing that if $k < k^\star_\varepsilon$, then a net is present with high probability.

Claim 5.21 *Consider Setting 5.3, in particular γ, δ and η. Fix $N > e$, $\varepsilon > 0$, recall $k^\star_\varepsilon(N)$ from (5.29)* *of Definition 5.5 and let k be a (possibly N-dependent) positive integer satisfying $k < k_\varepsilon^\star(N)$. Let $(r_i)_{i\le k}$ be the sequence of sidelengths from Definition 5.4. Fix $v \in \mathcal{V} = \mathbb{Z}^d$ and let $\mathcal{S}_v = \mathcal{S}_v(N,k,\gamma,d)$ be the boxing structure from Definition 5.9 around v. Furthermore, let $\mathcal{N}_v = \mathcal{N}_v(N,k,\gamma,\delta,\eta,\tau,d)$ be as defined in (5.120) of Definition 5.20. There exists an $\underline{N}_{5.21} = \underline{N}_{5.21}(\varepsilon,\gamma,\delta,\eta,\tau,d)$ such that if $N > \underline{N}_{5.21}$, then*

$$\mathbb{P}(v \in \mathcal{N}_v(N,k,\gamma,\delta,\eta,\tau,d)) \ge 1 - kr_k^{-d\delta}. \tag{5.121}$$

Furthermore, the event in (5.121) is independent of all vertices of $\mathcal{V}\backslash\mathcal{B}^{1,v}$, where $\mathcal{B}^{1,v} \in \mathcal{S}_v$ is the layer 1 box of the boxing structure.

Proof. We suppress dependence of $v,\gamma,\delta,\eta,\tau$ and d when appropriate. Consider $\underline{N}_{5.17}$ from Lemma 5.17 and set $\underline{N}_{5.21} = \underline{N}_{5.17}$. For the remainder of this proof, suppose that $N \ge \underline{N}_{5.21}$. To show (5.121), from definition of $\mathcal{N}_v$ (see (5.120) of Definition 5.20), the definition of good (see Definition 5.14) and the union bound we find that

$$\mathbb{P}(v \in \mathcal{N}_v) = \mathbb{P}\big(\mathcal{B}_\Lambda^i(v) \text{ is good for all } i = 1,\dots,k\big) \ge 1 - \sum_{i=1}^k \big(1 - \mathbb{P}\big(\mathcal{B}_\Lambda^i(v) \text{ is good}\big)\big). \tag{5.122}$$

Then, because $N \ge \underline{N}_{5.21}$ we may apply (5.84) of Lemma 5.17 to (5.122) to see that

$$\mathbb{P}(v \in \mathcal{N}_v) \ge 1 - \sum_{i=1}^k r_i^{-d\delta} \ge 1 - kr_k^{-d\delta}, \tag{5.123}$$

which shows (5.121). Here we have used that $r_i \ge r_k$ for all $i \le k$. Furthermore, notice that the event in (5.121) only depends on vertices that are contained in the boxing structure. Since $\mathcal{B}^{1,v}$ encompasses all these vertices, the event in (5.121) is independent of all vertices from $\mathcal{V}\backslash\mathcal{B}^{1,v}$. This finishes the proof. ■

5.3. Proof of Proposition 5.2

In this section, we finalise the proof of Proposition 5.2. To this end, we apply the nets from the previous section to make the sketch of the proof from Subsection 5.1 precise. Throughout this subsection, we use the concept of *hierarchies*, which formalises the labelling of the vertices described in Subsection 5.1. This idea is inspired by M. Biskup, and we use a slightly modified definition (see Definition 2.1 in [25]).

Definition 5.22 (Hierarchy) *Consider an integer $k \in \mathbb{N}$ and let $u, v \in \mathcal{V}$ be distinct vertices. A collection*

$$\mathscr{H}_k(u,v) = \Big\{(z_{\mathbf{t}_i}) \in \mathcal{V} : i = 1,2,\dots,k,k+1, \mathbf{t}_i \in \{0,1\}^i\Big\} \tag{5.124}$$

*is called a **hierarchy of depth** k if it satisfies the following four properties:*

1. *$z_0 = u$ and $z_1 = v$.*
2. *For all $i = 0,1,\dots,k-1$, for all $\mathbf{t}_i \in \{0,1\}^i$ we have that $z_{\mathbf{t}_i0} = z_{\mathbf{t}_i00}$ and $z_{\mathbf{t}_i1} = z_{\mathbf{t}_i11}$.*
3. *For all $i = 0,1,\dots,k-1$, for all $\mathbf{t}_i \in \{0,1\}^i$, if $z_{\mathbf{t}_i01} \ne z_{\mathbf{t}_i10}$ we have that $z_{\mathbf{t}_i01} \leftrightarrow z_{\mathbf{t}_i10}$.*
4. *The edges specified in item (3) appear only once in $\mathscr{H}_k(u,v)$.*

When u, v and k are clear from context, we write $\mathscr{H}_k$ and call it a hierarchy.

We may now recognise that the construction that is described in Subsection 5.1 and visualised in Figure 3, exactly yields a hierarchy of depth k. Therefore, in the remainder of this section we show that a hierarchy is present with sufficient probability, similar to what has been done in Subsection 5.1. To this end, we introduce some notation. In the following Definition 5.23, we establish notation to obtain all edges between two sets of vertices.

Definition 5.23 *For $\mathcal{X},\mathcal{Y} \subset \mathcal{V}$ such that $\mathcal{X} \cap \mathcal{Y} = \varnothing$, we set*

$$\mathscr{E}(\mathcal{X},\mathcal{Y}) = \{(x,y) : x \in \mathcal{X}, y \in \mathcal{Y}, xy \in \mathcal{E}\} \tag{5.125}$$

Notice that $\mathscr{E}(\mathcal{X},\mathcal{Y})$ consists of ordered pair of vertices. The reason for this is that further in this section, we would like to extract a pair $(x,y) \in \mathscr{E}(\mathcal{X},\mathcal{Y})$ and ensure that $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. If $\mathscr{E}(\mathcal{X},\mathcal{Y})$ were to consist of edges or unordered pairs of vertices, this would not be as readily possible.

We continue by defining a construction that, if it succeeds, yields a hierarchy. This construction is given in Definition 5.24, and further explanation follows afterwards. In particular, we formalise the construction given in the sketch of Section 5.1. We show that this construction succeeds with high probability as $N \to \infty$ in Lemma 5.25.

Definition 5.24 *Consider Setting 7.3, in particular γ,δ and η. Let $u, v \in \mathcal{V} = \mathbb{Z}^d$ and suppose that $N := |u - v| > e$. Let $\varepsilon > 0$, recall $k_\varepsilon^\star(N)$ from (5.29) of Definition 5.5 and let k be a (possibly N-dependent) positive integer satisfying $k < k_\varepsilon^\star(N)$. Let $\mathcal{S}_u = \mathcal{S}_u(N,k,\gamma,d)$ and $\mathcal{S}_v = \mathcal{S}_v(N,k,\gamma,d)$ be two* *boxing structures centred around u resp. v as in Definition 5.9. Furthermore, let $\mathcal{N}_u = \mathcal{N}_u(N,k,\gamma,\delta,\eta,\tau,d)$ and $\mathcal{N}_v(N,k,\gamma,\delta,\eta,\tau,d)$ be from (5.120) of Definition 5.20. Then let $\mathcal{B}^i_\Lambda(x)$ be as defined in (5.119) of Definition 5.19, T and $\#T$ from Definition 5.13 and $\mathscr{E}(\mathcal{X},\mathcal{Y})$ as in (5.125) from Definition 5.23. We define a collection*

$$\mathcal{H}_k(u,v) = \Big\{(Z_{\mathbf{t}_i}) : i \in \{1,\dots,k+1\}, \mathbf{t}_i \in \{0,1\}^i\Big\} \tag{5.126}$$

of (random) vertices in the following iterative way: first set $Z_0 = u$ and $Z_1 = v$ and then for $i = 2,\dots,k+1$

(1) *For all $\mathbf{t}_{i-2} \in \{0,1\}^{i-2}$, set $Z_{\mathbf{t}_{i-2}00} = Z_{\mathbf{t}_{i-2}0}$ and $Z_{\mathbf{t}_{i-2}11} = Z_{\mathbf{t}_{i-2}1}$.*

(2) *For all $\mathbf{t}_{i-2} \in \{0,1\}^{i-2}$, define $Z_{\mathbf{t}_{i-2}01}$ and $Z_{\mathbf{t}_{i-2}10}$ in the following way:*
- *If $Z_{\mathbf{t}_{i-2}0} \neq \varnothing$ and $Z_{\mathbf{t}_{i-2}1} \neq \varnothing$, and if $\mathcal{B}^i_\Lambda(Z_{\mathbf{t}_{i-2}0}) = \mathcal{B}^i_\Lambda(Z_{\mathbf{t}_{i-2}1})$, then set $Z_{\mathbf{t}_{i-2}01} = Z_{\mathbf{t}_{i-2}10} = Z_{\mathbf{t}_{i-2}0}$.*
- *If $Z_{\mathbf{t}_{i-2}0} \neq \varnothing$ and $Z_{\mathbf{t}_{i-2}1} \neq \varnothing$, and if $\mathcal{B}^i_\Lambda(Z_{\mathbf{t}_{i-2}0}) \neq \mathcal{B}^i_\Lambda(Z_{\mathbf{t}_{i-2}1})$, then set*

$$\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0} = T\Big(\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}0}), \Big[1, \eta^{-\frac{1}{\tau-1}}\Big]\Big) \cap (\mathcal{N}_u \cup \mathcal{N}_v), \qquad \textit{and} \tag{5.127}$$

$$\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1} = T\Big(\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}1}), \Big[\underline{M}_{i-1}^{\frac{1}{\tau-1}} r_{i-1}^{d\frac{1-\delta}{\tau-1}}, \overline{M}_{i-1}^{\frac{1}{\tau-1}} r_{i-1}^{d\frac{1-\delta}{\tau-1}}\Big]\Big) \cap (\mathcal{N}_u \cup \mathcal{N}_v). \tag{5.128}$$

Then if $\mathscr{E}(\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}, \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}) \neq \varnothing$, choose any $(x,y) \in \mathscr{E}(\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}, \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1})$ randomly in a way that is independent of all other random terms in this procedure and set $Z_{\mathbf{t}_{i-2}01} = x$ and $Z_{\mathbf{t}_{i-2}10} = y$.
- *Otherwise, set $Z_{\mathbf{t}_{i-2}01} = Z_{\mathbf{t}_{i-2}10} = \varnothing$.*

If u and v are clear from context, we write $\mathcal{H}_k$.

We elaborate on Definition 5.24. We follow the same reasoning as given in Subsection 5.1. First, in the 'zeroth step', we set $Z_0 = u$ and $Z_1 = v$. Then, in the ith step, we firstly elongate every vertex-label as in item (1), which defines $Z_{\mathbf{t}_{i-2}00} = Z_{\mathbf{t}_{i-2}0}$ and $Z_{\mathbf{t}_{i-2}11} = Z_{\mathbf{t}_{i-2}1}$.

The complexity lies in defining $Z_{\mathbf{t}_{i-2}01}$ and $Z_{\mathbf{t}_{i-2}10}$. Recall from the sketch of the proof (see Subsection 5.1) that these vertices are supposed to bridge the box of layer $i-1$ containing $Z_{\mathbf{t}_{i-2}0}$ and the box of layer $i-1$ containing $Z_{\mathbf{t}_{i-2}1}$, in such a way that $Z_{\mathbf{t}_{i-2}01}$ is a constant-weight vertex and $Z_{\mathbf{t}_{i-2}10}$ is a high-weight vertex. In other words, we consider all vertices of $\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}0})$ that have weight between 1 and $\eta^{-1/(\tau-1)}$ and try to find edges to the vertices of $\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}1})$ that have weight between $\underline{M}_{i-1}^{1/(\tau-1)} r_{i-1}^{d(1-\delta)/(\tau-1)}$ and $\overline{M}_{i-1}^{1/(\tau-1)} r_{i-1}^{d(1-\delta)/(\tau-1)}$. However, to ensure that further steps also succeed, we only consider those vertices that are in good boxes all of all layers. This ensures that in further steps, we do not encounter situations where there are not 'enough' constant-weight or high-weight vertices to those steps also succeed. In particular, we search for vertices that are in the nets, i.e., either in $\mathcal{N}_u$ or $\mathcal{N}_v$. We conclude that we search for vertices exactly in $\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}$ and $\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}$ from equations (5.127) and (5.128). If there is an edge present between these sets (i.e., if $\mathscr{E}(\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}, \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}) \neq \varnothing$), then we may define $Z_{\mathbf{t}_{i-2}01}$ and $Z_{\mathbf{t}_{i-2}10}$ according to this edge. If there are multiple, we do however require that this happens in an independent way. Particularly, we require that the choice is independent of the everything else in the construction and the weight of all considered vertices.

There are two exceptions in defining $Z_{\mathbf{t}_{i-2}01}$ and $Z_{\mathbf{t}_{i-2}10}$. The first is if $Z_{\mathbf{t}_{i-2}0}$ and $Z_{\mathbf{t}_{i-2}1}$ are already in the same box of layer i, in which case there is little use in connecting their boxes of layer i. We then skip finding an edge by simply setting $Z_{\mathbf{t}_{i-2}01}$ and $Z_{\mathbf{t}_{i-2}10}$ to be $Z_{\mathbf{t}_{i-2}0}$. The second is if either $Z_{\mathbf{t}_{i-2}0}$ and $Z_{\mathbf{t}_{i-2}1}$ is equal to $\varnothing$, or if $\mathscr{E}(\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}, \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}) = \varnothing$. In this case, the construction has failed and we set $Z_{\mathbf{t}_{i-2}01} = Z_{\mathbf{t}_{i-2}10} = \varnothing$.

Before we continue, we make one further observation about Definition 5.24: if the construction succeeds, then $\mathcal{H}_k$ is a hierarchy of depth k. Furthermore, we have one simple test to see if the construction has failed; in this case $\mathcal{H}_k$ contains at least one element that is equal to $(\varnothing)$. That is,

$$\mathcal{H}_k \text{ is a hierarchy of depth } k \qquad \Leftrightarrow \qquad (\varnothing) \notin \mathcal{H}_k. \tag{5.129}$$

Furthermore, if the construction succeeds, then by definition all vertices $(Z_{\mathbf{t}_i})_{\mathbf{t}_i \in \{0,1\}^i, i \le k+1}$ are either in $\mathcal{N}_u$ or $\mathcal{N}_v$. This implies if the hierarchy succeeds, then all its vertices are contained in good boxes of every layer. Additionally, notice that by construction for all $i \in \{2,\dots,k+1\}$ it holds that $(Z_{\mathbf{t}_i})_{\mathbf{t}_i \in \{0,1\}^i}$ contains all information about $(Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1} \in \{0,1\}^{i-1}}$. Indeed, we may obtain every vertex from $(Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1} \in \{0,1\}^{i-1}}$ from $(Z_{\mathbf{t}_i})_{\mathbf{t}_i \in \{0,1\}^i}$ by applying step (1) from Definition 5.24 in reverse. We also observe that

$$\varnothing \notin (Z_{\mathbf{t}_i})_{\mathbf{t}_i \in \{0,1\}^i} \qquad \Rightarrow \qquad \varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1} \in \{0,1\}^{i-1}}. \tag{5.130}$$

Lastly, notice that by definition, if $Z_{\mathbf{t}_{i-2}01} \neq \varnothing$, then $Z_{\mathbf{t}_{i-2}01} \in \mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}0} = \mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-1}00})$. Similarly, if $Z_{\mathbf{t}_{i-2}10} \in \mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}11})$. We therefore observe that for all $\mathbf{t}_{i-1} \in \{0,1\}^{i-1}$ it holds that

$$\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-1}0}) = \mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-1}1}). \tag{5.131}$$

From item (1) of Lemma A.1 it then follows that

$$|Z_{\mathbf{t}_{i-1}0} - Z_{\mathbf{t}_{i-1}1}| \le \sqrt{d} r_{i-1}, \tag{5.132}$$

provided that $Z_{\mathbf{t}_{i-2}01}, Z_{\mathbf{t}_{i-2}10} \neq \varnothing$.
In the following lemma, we show that with high probability $\mathcal{H}_k$ in fact constitutes a hierarchy of depth k.

Lemma 5.25 *Consider Setting 5.3, in particular γ, δ and η. Let $u, v \in \mathcal{V} = \mathbb{Z}^d$. Let $N = |u - v|, \varepsilon > 0$, recall $k_\varepsilon^\star(N)$ from (5.29) of Definition 5.5 and let k be a (possibly N-dependent) positive integer satisfying $k < k_\varepsilon^\star(N)$. Let $(r_i)_{i \leq k}$ be the sequence of sidelengths from Definition 5.4. Fix $u, v \in \mathcal{V} = \mathbb{Z}^d$, let $\mathcal{H}_k(u, v)$ be as in Definition 5.24 and let a hierarchy of depth k be as given in Definition 5.22. Then there exists an $\underline{N}_{5.25} = \underline{N}_{5.25}(\varepsilon, \gamma, \delta, \eta, \tau, d)$ and a function $\mathrm{err}_{5.25}(N, \varepsilon, \gamma, \delta, \eta, \alpha, \tau, d)$ that goes to 0 if $N \to \infty$ such that if $N \geq \underline{N}_{5.25}$, then*

$$\mathbb{P}(\mathcal{H}_k(u, v) \text{ is a hierarchy of depth } k) \geq 1 - \mathrm{err}_{5.25}(N, \varepsilon, \gamma, \delta, \eta, \alpha, \tau, d). \tag{5.133}$$

Proof. As done before, we suppress dependence on γ, δ, η, N and k when appropriate. This proof consists of multiple parts.

(Part 1: defining $\underline{N}_{5.25}$) We firstly give the definition of $\underline{N}_{5.25}$. Firstly, note that $\gamma < 1$ (see Setting 5.3), from which it follows that there exists an $\underline{N}_1 = \underline{N}_1(\gamma)$ such that if $N \geq \underline{N}_1$, then it holds that $\lceil N^\gamma \rceil \leq N/\sqrt{d}$. Furthermore, by Claim 5.21 we find that if we set $\underline{N}_2 = \underline{N}_2(\varepsilon, \gamma, \delta, \eta, \tau, d) = \underline{N}_{5.21}(\varepsilon, \gamma, \delta, \eta, \tau, d)$, then if $N \geq \underline{N}_2$ it holds that $\mathcal{N}_u$ (resp. $\mathcal{N}_v$) is a net for v (resp. u) with probability greater than $1 - kr_k^{-d\delta}$. Furthermore, we set $\underline{N}_3 = \underline{N}_3(\varepsilon, \gamma, \delta, \eta, \tau, d) = \underline{N}_{5.17}(\varepsilon, \gamma, \delta, \eta, \tau, d)$ and note that if $N \geq \underline{N}_3$, then (5.112) holds. Then, by Claim 5.7 we may set $\underline{N}_4 = \underline{N}_4(\varepsilon, \gamma) = \underline{N}_{5.7}(\varepsilon, \gamma)$ such that if $N \geq \underline{N}_4$, then $N^{\gamma^i} \leq r_i \leq 2N^{\gamma^i}$ for all $i \leq k$. Next, by Claim 5.6 and item (2) of Lemma 5.7 it is possible to define $\underline{N}_5 = \underline{N}_5(\gamma, \delta, \varepsilon, d)$ such that if $N \geq \underline{N}_5$, then it holds that

$$\prod_{j=1}^{k}(1 - 2r_j^{-d\delta}) \geq 1 - \sum_{j=1}^{k} 2r_j^{-d\delta} \geq 1 - 2r_k^{-d\delta} \geq 1 - 4kN^{-d\delta\gamma^k} \geq 1 - 4\exp[-d\delta(\ln\ln N)^\varepsilon] \geq \frac{1}{2}. \tag{5.134}$$

Here we have repeatedly used that $(1-x)(1-y) \geq 1 - x - y$ if $x, y \in (0, 1)$ and the fact that $r_i \geq r_k$ for all $i \leq k$. Furthermore, by items (1) and (3) of Claim 5.12, we may define $\underline{N}_6 = \underline{N}_6(\varepsilon, \gamma, \delta, \eta, d) = \underline{N}_{5.12}(\varepsilon, \gamma, \delta, \eta, d)$ such that $\underline{M}_i \geq 1/(\phi\eta)$, $\underline{A}_i \geq \eta/4$ and $\underline{a}_i = \underline{A}_i/2 \geq \eta/8$ if $N \geq \underline{N}_6$.
We now set $\underline{N}_{5.25} = \max\{\underline{N}_1, \underline{N}_2, \underline{N}_3, \underline{N}_4, \underline{N}_5, \underline{N}_6\}$. For the remainder of this proof, assume that $N \geq \underline{N}_{5.25}$.

(Part 2: preliminary work) Before we show that $\mathcal{H}_k(u, v)$ is a hierarchy of depth k, we give some preliminary definitions and do some preliminary work to be used later in the proof. Firstly, let $\mathcal{S}_u = \mathcal{S}_u(N, k, \gamma, d)$ and $\mathcal{S}_v = \mathcal{S}_v(N, k, \gamma, d)$ two boxing structures centred around u and v respectively (see Definition 5.9). Furthermore, given these boxing structures let $\mathcal{N}_u = \mathcal{N}_u(N, k, \gamma, \delta, \eta, \tau, d)$ and $\mathcal{N}_v = \mathcal{N}_v(N, k, \gamma, \delta, \eta, \tau, d)$ be as defined in 5.120 in Definition 5.20. We suppress dependence on γ, δ and η and see the dependence on N and k as implicit, unless necessary for clarity. Recall that $\mathcal{N}_u$ (resp. $\mathcal{N}_v$) is called a net for u (resp. v) if all boxes of $\mathcal{S}_u$ (resp. $\mathcal{S}_v$) that contain u (resp. v) are good. Furthermore, recall that all boxes of $\mathcal{S}_u$ (resp. $\mathcal{S}_v$) are subsets of $\mathcal{B}^{1,u}$ (resp. $\mathcal{B}^{1,v}$). Furthermore, since $N \geq \underline{N}_{5.25}$, we have that $r_1 = \lceil N^\gamma \rceil \leq N/\sqrt{d}$, so that by item (2) of Lemma A.1 we my find that $\mathcal{B}^{1,u} \cap \mathcal{B}^{1,v} = \varnothing$. From this it follows that every box in $\mathcal{S}_u$ is disjoint from every box of $\mathcal{S}_v$, i.e.,

$$\text{for all } i, j = 1, \dots, k, \text{ for all } \mathcal{B}_{\mathbf{s}_i}^{i,u} \in \mathcal{S}_u \text{ and } \mathcal{B}_{\mathbf{s}_j}^{j,v} \in \mathcal{S}_v \text{ it holds that } \mathcal{B}_{\mathbf{s}_i}^{i,u} \cap \mathcal{B}_{\mathbf{s}_j}^{j,v} = \varnothing. \tag{5.135}$$

In turn, (5.135) implies that

$$\mathcal{N}_u \cap \mathcal{N}_v = \varnothing \tag{5.136}$$

by definition. Furthermore, (5.135) implies that $\mathcal{B}_\Lambda^i(x)$ from (5.119) of Definition 5.19 is well-defined.
(Part 3: rewriting and set-up) We return to (5.133) by denoting H as the event given in (5.133), i.e.,

$$H := \{\mathcal{H}_k \text{ is a hierarchy of depth } k\}. \tag{5.137}$$

Then, we rewrite (5.133) as

$$\begin{aligned}\mathbb{P}(H) &= \mathbb{P}(H | u \in \mathcal{N}_u, v \in \mathcal{N}_v)\mathbb{P}(u \in \mathcal{N}_u, v \in \mathcal{N}_v) + \mathbb{P}(H | (u \in \mathcal{N}_u, v \in \mathcal{N}_v)^c)\mathbb{P}((u \in \mathcal{N}_u, v \in \mathcal{N}_v)^c) \\ &\geq \mathbb{P}(H | u \in \mathcal{N}_u, v \in \mathcal{N}_v)\mathbb{P}(u \in \mathcal{N}_u, v \in \mathcal{N}_v),\end{aligned} \tag{5.138}$$

where we have applied the law of total probability. Now by Claim 5.21, since $\mathcal{N}_u$ and $\mathcal{N}_v$ are disjoint (see (5.136)), the events $\{u \in \mathcal{N}_u\}$ and $\{v \in \mathcal{N}_v\}$ are independent and furthermore since $N \geq \underline{N}_{5.25}$

$$\mathbb{P}(u \in \mathcal{N}_u, v \in \mathcal{N}_v) = \mathbb{P}(u \in \mathcal{N}_u)\mathbb{P}(v \in \mathcal{N}_v) \geq (1 - kr_k^{-d\delta})^2. \tag{5.139}$$

It therefore remains to analyse $\mathbb{P}(H | u \in \mathcal{N}_u, v \in \mathcal{N}_v)$ from (5.138). To this end, note that by (5.129) we may equivalently analyse $\mathbb{P}((\varnothing) \notin \mathcal{H}_k | u \in \mathcal{N}_u, v \in \mathcal{N}_v)$. In particular, we aim to show that

$$\mathcal{Q} := \mathbb{P}(\varnothing \notin (Z_{\mathbf{t}_i})_{i \in \{0,1\}^i} \text{ for all } i \in [k+1] \mid u \in \mathcal{N}_u, v \in \mathcal{N}_v) \geq 1 - \tilde{g}(N), \tag{5.140}$$

for some function $\tilde{g}(N)$ that satisfies $\tilde{g}(N) \to 0$ if $N \to \infty$. Here $[k+1] = \{1, 2, \dots, k, k+1\}$. In the remainder or this proof, we denote

$$\mathscr{N} = \{u \in \mathcal{N}_u, v \in \mathcal{N}_v\}. \tag{5.141}$$

We then telescopically apply conditional independence to (5.140) to see that

$$\mathscr{Q} = \mathbb{P}\big(\varnothing \notin (Z_{\mathbf{t}_1})_{\mathbf{t}_1\in\{0,1\}} \big| \mathscr{N}\big) \prod_{i=2}^{k+1} \mathbb{P}\big(\varnothing \notin (Z_{\mathbf{t}_i})_{\mathbf{t}_i\in\{0,1\}^i} \;\big|\; \varnothing \notin (Z_{\mathbf{t}_j})_{\mathbf{t}_j\in\{0,1\}^j} \text{ for all } j\in[i-1], \mathscr{N}\big). \tag{5.142}$$

Now notice that by (5.130), we obtain that

$$\Big\{\varnothing \notin (Z_{\mathbf{t}_j})_{j\in\{0,1\}^j} \text{ for all } j\in[i-1]\Big\} = \Big\{\varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^i}\Big\}, \tag{5.143}$$

Thus, if we define

$$\mathscr{Q}_1 := \mathbb{P}\big(\varnothing \notin (Z_{\mathbf{t}_1})_{\mathbf{t}_1\in\{0,1\}} \big| \mathscr{N}\big) \quad\text{and}\quad \mathscr{Q}_i := \mathbb{P}\big(\varnothing \notin (Z_{\mathbf{t}_i})_{\mathbf{t}_i\in\{0,1\}^i} \;\big|\; \varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}, \mathscr{N}\big), \tag{5.144}$$

then from (5.142) and (5.143) it follows that

$$\mathscr{Q} = \prod_{i=1}^{k+1} \mathscr{Q}_i. \tag{5.145}$$

(Part 4: examining $\mathscr{Q}_i$) To find the bound given in (5.140), we examine each $\mathscr{Q}_i$ separately. Firstly, notice that $Z_0 = u$ and $Z_1 = v$, so $\mathscr{Q}_1 = 1$. Therefore, consider any $\mathscr{Q}_i$ for $i \in \{2,\dots,k+1\}$. Rather than bound $\mathscr{Q}_i$, we bound $1-\mathscr{Q}_i$. To this end, we rewrite

$$\begin{aligned}\Big\{\varnothing \notin (Z_{\mathbf{t}_i})_{\mathbf{t}_i\in\{0,1\}^i}\Big\} &= \Bigg(\bigcap_{\mathbf{t}_{i-2}\in\{0,1\}^{i-2}} \big\{Z_{\mathbf{t}_{i-2}00} \neq \varnothing \text{ and } Z_{\mathbf{t}_{i-2}11} \neq \varnothing\big\}\Bigg) \\ &\qquad \cap \Bigg(\bigcap_{\mathbf{t}_{i-2}\in\{0,1\}^{i-2}} \big\{Z_{\mathbf{t}_{i-2}01} \neq \varnothing \text{ and } Z_{\mathbf{t}_{i-2}10} \neq \varnothing\big\}\Bigg).\end{aligned} \tag{5.146}$$

Then by applying (5.146) and the union bound, we find that

$$\begin{aligned}1-\mathscr{Q}_i \le &\sum_{\mathbf{t}_{i-2}\in\{0,1\}^{i-2}} \mathbb{P}\big(Z_{\mathbf{t}_{i-2}00} = \varnothing \text{ or } Z_{\mathbf{t}_{i-2}11} = \varnothing \;\big|\; \varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}, \mathscr{N}\big) \\ &+ \sum_{\mathbf{t}_{i-2}\in\{0,1\}^{i-2}} \mathbb{P}\big(Z_{\mathbf{t}_{i-2}01} = \varnothing \text{ or } Z_{\mathbf{t}_{i-2}10} = \varnothing \;\big|\; \varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}, \mathscr{N}\big).\end{aligned} \tag{5.147}$$

Then, observe that if $\varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}$, then also $Z_{\mathbf{t}_{i-2}00} \neq \varnothing$ and $Z_{\mathbf{t}_{i-2}11} \neq \varnothing$. Indeed, this follows immediately from the fact that $Z_{\mathbf{t}_{i-2}00} = Z_{\mathbf{t}_{i-2}0} \in (Z_{\mathbf{t}_{j-1}})_{\mathbf{t}_{j-1}\in\{0,1\}^{j-1}}$ and $Z_{\mathbf{t}_{i-2}11} = Z_{\mathbf{t}_{i-2}1} \in (Z_{\mathbf{t}_{j-1}})_{\mathbf{t}_{j-1}\in\{0,1\}^{j-1}}$. As such, we find that $\mathbb{P}\big(Z_{\mathbf{t}_{i-2}00} = \varnothing \text{ or } Z_{\mathbf{t}_{i-2}11} = \varnothing \;\big|\; \varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}, \mathscr{N}\big) = 0$ and the first sum in (5.147) vanishes. Thus, we examine

$$\Psi_{\mathbf{t}_{i-2}} = \mathbb{P}\big(Z_{\mathbf{t}_{i-2}01} = \varnothing \text{ or } Z_{\mathbf{t}_{i-2}10} = \varnothing \;\big|\; \varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}, \mathscr{N}\big) \tag{5.148}$$

for every $\mathbf{t}_{i-2} \in \{0,1\}^{i-2}$ and find an uniform bound for all these quantities. To this end, fix any $\mathbf{t}_{i-2} \in \{0,1\}^{i-2}$, assume that $Z_{\mathbf{t}_{i-2}0}, Z_{\mathbf{t}_{i-2}1} \neq \varnothing$ and $\mathscr{N}$ holds. Now by construction (see item (2) of Definition 5.24), if $\mathcal{B}^i_\Lambda(Z_{\mathbf{t}_{i-2}0}) = \mathcal{B}^i_\Lambda(Z_{\mathbf{t}_{i-2}1})$, we know that $Z_{\mathbf{t}_{i-2}01} = Z_{\mathbf{t}_{i-2}10} = Z_{\mathbf{t}_{i-2}0}$ and hence $\Psi_{\mathbf{t}_{i-2}} = 0$. So consider the case where this is not true. Then it is only possible that $Z_{\mathbf{t}_{i-2}} = \varnothing$ or $Z_{\mathbf{t}_{i-2}} = \varnothing$ if $\mathscr{E}(\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}, \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}) \neq \varnothing$. Recall here equations (5.127) and (5.128):

$$\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0} = T\Big(\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}0}), \Big[1, \eta^{-\frac{1}{\tau-1}}\Big]\Big) \cap (\mathcal{N}_u \cup \mathcal{N}_v), \qquad \text{and} \tag{5.149}$$

$$\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1} = T\Big(\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}1}), \Big[\underline{M}_{i-1}^{\frac{1}{\tau-1}} r_{i-1}^{d\frac{1-\delta}{\tau-1}}, \overline{M}_{i-1}^{\frac{1}{\tau-1}} r_{i-1}^{d\frac{1-\delta}{\tau-1}}\Big]\Big) \cap (\mathcal{N}_u \cup \mathcal{N}_v). \tag{5.150}$$

Since we condition on $\mathscr{N} = \{u \in \mathcal{N}_u, v \in \mathcal{N}_v\}$, we know that $\mathcal{N}_u$ and $\mathcal{N}_v$ are non-empty. We firstly examine $\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}$. We note that by construction, $Z_{\mathbf{t}_{i-2}0} \in \mathcal{N}_u \cup \mathcal{N}_v$ and hence all boxes that contain $Z_{\mathbf{t}_{i-2}0}$ are good. In particular, $\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}0})$ is good. By Corollary 5.18 we therefore know that $\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}0})$ contains at least $(r_{i-1}/r_k)^d \prod_{j=i}^k (1-2r_j^{-d\delta})$ sub-boxes of layer k that are contained within good boxes of layer $j = i-1,\dots,k$. Since also each box that contains $\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}0})$ is also good by construction, each of these sub-boxes of layer k are fully contained within $\mathcal{N}_u \cup \mathcal{N}_v$. Because each of these sub-boxes contains at least $\underline{\eta} r_k^d$ vertices that have weight between 1 and $\eta^{-1/(\tau-1)}$,

$$\big|\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}\big| \ge \underline{\eta} r_{i-1}^d \prod_{j=i}^k (1-2r_j^{-d\delta}) \tag{5.151}$$

holds when $Z_{\mathbf{t}_{i-2}0} \neq \varnothing$ and $\mathscr{N}$. Also notice that $Z_{\mathbf{t}_{i-2}0}$ is a high-weight vertex, so it does not subtract the number of constant-weight vertices, i.e., it does not subtract from the vertices of $\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}$ that may still be examined. Next, we examine $\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}$. To this end, note that by part 2 and 3 of the proof of Lemma 5.17, if $N \ge \underline{N}_{5.25}$ it follows that with probability greater than $1-\exp\big[-\eta r_{i-1}^{d\delta}/32\big]$ there are more than $\underline{a}_{i-1} r_{i-1}^{d\delta}$ vertices that are solely in good sub-boxes of layer $j = i,\dots,k$ of $\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}1})$ and have weight between $\underline{M}_{i-1}^{1/(\tau-1)} r_{i-1}^{d(1-\delta)/(\tau-1)}$ and $\overline{M}_{i-1}^{1/(\tau-1)} r_{i-1}^{d(1-\delta)/(\tau-1)}$. Since $Z_{\mathbf{t}_{i-2}1} \in \mathcal{N}_v \cup \mathcal{N}_u$, $\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}1})$ is good and solely contained in good boxes of layer 1 up to $i-1$. Therefore, all these vertices are solely in good sub-boxes and therefore in $\mathcal{N}_u \cup \mathcal{N}_v$. Furthermore, since $Z_{\mathbf{t}_{i-2}1}$ is a constant-weight vertex, it does not subtract from the number of high-weight vertices, i.e., it does not subtract from the number of vertices of $\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}$ that may be examined. We conclude that

$$\Psi^2_{\mathbf{t}_{i-2}} := \mathbb{P}\big(\big|\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}\big| \geq \underline{a}_{i-1} r^{d\delta}_{i-1} \;\big|\; \varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}, \mathscr{N}\big) \geq 1 - \exp\Big[-\frac{\eta}{32} r^{d\delta}_{i-1}\Big]. \tag{5.152}$$

Furthermore, by item (2) of Lemma 5.7 we know that $r_{i-1} \geq N^{\gamma^{i-1}}$ and hence

$$\Psi^2_{\mathbf{t}_{i-2}} \geq 1 - \exp\Big[-\frac{\eta}{32} N^{d\delta\gamma^{i-1}}\Big] =: 1 - \mathrm{err}_2(i-1). \tag{5.153}$$

We return to (5.148). We may now apply the law of total probability to find that

$$\begin{aligned}
1 - \Psi_{\mathbf{t}_{i-2}} &= \mathbb{P}\big(\mathscr{E}(\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}, \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}) \neq \varnothing \;\big|\; \varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}, \mathscr{N}\big)\\
&\geq \mathbb{P}\big(\mathscr{E}(\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}, \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}) \neq \varnothing \;\big|\; \varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}, \mathscr{N}, \big|\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}\big| \geq \underline{a}_{i-1} r^{d\delta}_{i-1}\big)\\
&\qquad \cdot \mathbb{P}\big(\big|\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}\big| \geq \underline{a}_{i-1} r^{d\delta}_{i-1} \big| \varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}, \mathscr{N}\big)\\
&=: \Psi^1_{\mathbf{t}_{i-2}} \Psi^2_{\mathbf{t}_{i-2}}.
\end{aligned} \tag{5.154}$$

Notice that we bound $1 - \Psi_{\mathbf{t}_{i-2}}$ and not $\Psi_{\mathbf{t}_{i-2}}$ in the last equation.

(Part 5: examining $\Psi^1_{\mathbf{t}_{i-2}}$) It remains to examine $\Psi^1_{\mathbf{t}_{i-2}}$, i.e., the probability that $\mathscr{E}(\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}, \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}) \neq \varnothing$ given that $\big|\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}\big| \geq \underline{a}_{i-1} r^{d\delta}_{i-1}$, $\varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}$ and $\mathscr{N}$ hold. We do this by examining the probability of the complement $1 - \Psi^1_{\mathbf{t}_{i-2}}$, i.e., where $\mathscr{E}(\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}, \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}) = \varnothing$. To ease notation in this part, we suppress the dependence on $\mathbf{t}_{i-2}$ when possible. Set $n := \big|\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}\big|$ and enumerate each vertex of $\mathcal{T}_{\mathbf{t}_{i-2}0}$ by $x_1, \ldots, x_n$ in a way that is independent from each other random element in the construction (for example by lexicographical ordering). Similarly, set $m := \big|\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}\big|$ and enumerate each vertex of $\mathcal{T}_{\mathbf{t}_{i-2}1}$ by $y_1, \ldots, y_m$. Then by (5.151), the fact that $\underline{\eta} = \eta/2$ (see (5.45) of Definition 5.11) and because $N \geq \underline{N}_{5.25}$, it holds that

$$n \geq \underline{\eta} r^d_{i-1} \prod_{j=i}^{k} (1 - 2r_j^{-d\delta}) \geq \underline{\eta} r^d_{i-1} \prod_{j=1}^{k} (1 - 2r_j^{-d\delta}) \geq \frac{\eta}{4} r^d_{i-1}. \tag{5.155}$$

Next, if we condition on $\big|\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}\big| \geq \underline{a}_{i-1} r^{d\delta}_{i-1}$, by applying item (3) of Claim 5.12 and the fact that $\underline{a}_{i-1} = \underline{A}_{i-1}/2$, we find that because $N \geq \underline{N}_{5.25}$

$$m \geq \underline{a}_{i-1} r^{d\delta}_{i-1} \geq \frac{\eta}{8} r^{d\delta}_{i-1}. \tag{5.156}$$

We return to $1 - \Psi^1_{\mathbf{t}_{i-2}}$. Notice that by the definition of $\mathscr{E}$ (see (5.125) of Definition 5.23) it is only possible that $\mathscr{E}(\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}, \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}) = \varnothing$ if there are *no* edges between $\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}$ and $\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}$, i.e., if for all $\tilde{n} \leq n$ and $\tilde{m} \leq m$ it holds that $x_{\tilde{n}} \not\leftrightarrow y_{\tilde{m}}$. Hence, by repeatedly applying the definition of conditional probability we may observe that

$$\begin{aligned}
1 - \Psi^1_{\mathbf{t}_{i-2}} =& \mathbb{P}\big(\text{for all } \tilde{n} \leq n, \tilde{m} \leq m \text{ we have } x_{\tilde{n}} \not\leftrightarrow y_{\tilde{m}} \big| \varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}, \mathscr{N}, \big|\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}\big| \geq \underline{a}_{i-1} r^{d\delta}_{i-1}\big)\\
=& \prod_{\substack{\tilde{n} \leq n\\ \tilde{m} \leq m}} \mathbb{P}(x_{\tilde{n}} \not\leftrightarrow y_{\tilde{m}} | \mathcal{F}_{\tilde{n},\tilde{m}}),
\end{aligned} \tag{5.157}$$

where we have defined as $\mathcal{F}_{\tilde{n},\tilde{m}}$ as follows:

$$\mathcal{F}_{\tilde{n},\tilde{m}} := \bigcap_{\substack{\mu < \tilde{m}\\ \nu \leq n}} \{x_\nu \not\leftrightarrow y_\mu\} \cap \bigcap_{\nu < \tilde{n}} \{x_\nu \not\leftrightarrow y_{\tilde{m}}\} \cap \mathscr{N} \cap \Big\{\varnothing \notin (Z_{\mathbf{t}_{i-1}})_{\mathbf{t}_{i-1}\in\{0,1\}^j}\Big\} \cap \Big\{\big|\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}\big| \geq \underline{a}_{i-1} r^{d\delta}_{i-1}\Big\}.$$

In particular, $\mathcal{F}_{\tilde{n},\tilde{m}}$ contains all the information we condition in in $\Psi^1_{\mathbf{t}_{i-2}}$ and the information that up to the combination $(\tilde{n}, \tilde{m})$, no edge has been found. We examine each $\mathbb{P}(x_{\tilde{n}} \not\leftrightarrow y_{\tilde{m}} | \mathcal{F}_{\tilde{n},\tilde{m}})$ in (5.157) separately. To this end, notice that for $x_{\tilde{n}} \in \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}$ and $y_{\tilde{m}} \in \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1}$ the following holds:

- Since nothing from $\mathcal{F}_{\tilde{n},\tilde{m}}$ prevents $x_{\tilde{n}}$ to have weight higher than 1 or $y_{\tilde{m}}$ to have weight higher than $\underline{M}^{1/(\tau-1)}_{i-1} r^{d(1-\delta)/(\tau-1)}_{i-1}$, it holds that $\mathbb{P}(W_{x_{\tilde{n}}} \geq 1, W_{y_{\tilde{m}}} \geq \underline{M}^{1/(\tau-1)}_{i-1} r^{d(1-\delta)/(\tau-1)}_{i-1} | \mathcal{F}_{\tilde{n},\tilde{m}}) > 0$.
- By the independence of edge presence if the weight is given, we may observe that for $w_{x_{\tilde{n}}} \in [1, \eta^{-1/(\tau-1)}]$ and $w_{y_{\tilde{m}}} \in [\underline{M}^{1/(\tau-1)}_{i-1} r^{d(1-\delta)/(\tau-1)}_{i-1}, \overline{M}^{1/(\tau-1)}_{i-1} r^{d(1-\delta)/(\tau-1)}_{i-1}]$ it holds that

$$\mathbb{P}(x_{\tilde{n}} \leftrightarrow y_{\tilde{m}} | W_{x_{\tilde{n}}} = w_{x_{\tilde{n}}}, W_{y_{\tilde{m}}} = w_{y_{\tilde{m}}}, \mathcal{F}_{\tilde{n},\tilde{m}}) = \mathbb{P}(x_{\tilde{n}} \leftrightarrow y_{\tilde{m}} | W_{x_{\tilde{n}}} = w_{x_{\tilde{n}}}, W_{y_{\tilde{m}}} = w_{y_{\tilde{m}}}). \tag{5.158}$$

- If $i = 2$, then $Z_{\mathbf{t}_{i-2}0} = Z_0 = u$ and $Z_{\mathbf{t}_{i-2}1} = Z_1 = v$. As such, we observe that $\mathcal{T}^1_0 \subseteq \mathcal{B}^{1,u}$ and $\mathcal{T}^1_1 \subseteq \mathcal{B}^{1,v}$ and hence $|x_{\tilde{n}} - y_{\tilde{m}}| \leq N + \sqrt{d} N^\gamma$. Since $N > 1$ and $\gamma < 1$, it holds that $N + \sqrt{d} N^\gamma \leq (1 + \sqrt{d}) N$. Hence if we set

$$\sqrt{d} r_0 := (1 + \sqrt{d}) N, \tag{5.159}$$

then $|x_{\tilde{n}} - y_{\tilde{m}}| \leq \sqrt{d} r_0$ holds.
If $i \in \{3, \ldots, k+1\}$, then by (5.131), $\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}0})$ and $\mathcal{B}^{i-1}_\Lambda(Z_{\mathbf{t}_{i-2}1})$ must be contained within the same layer $i-2$ box. It also follows that $\mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}, \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}1} \subset \mathcal{B}^{i-2}_\Lambda(Z_{\mathbf{t}_{i-2}0})$. By item (1) of Lemma A.1, $|x_{\tilde{n}} - y_{\tilde{m}}| \leq \sqrt{d} r_{i-2}$ holds.

The first two points justify applying Lemma 4.1. Furthermore, recall that $x_{\tilde n} \in \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}$ implies that $W_{x_{\tilde n}} \in [1, \eta^{-1/(\tau-1)}]$ and $y_{\tilde m} \in \mathcal{T}^{i-1}_{\mathbf{t}_{i-2}0}$ implies that $W_{y_{\tilde m}}$ is between $\underline{M}_{i-1}^{1/(\tau-1)} r_{i-1}^{d(1-\delta)/(\tau-1)}$ and $\overline{M}_{i-1}^{1/(\tau-1)} r_{i-1}^{d(1-\delta)/(\tau-1)}$. By applying Lemma 4.1 we find that

$$\mathbb{P}(x_{\tilde n} \not\leftrightarrow y_{\tilde n} | \mathcal{F}_{\tilde n, \tilde m}) \le 1 - \underline{c}\rho\Big(|x_{\tilde n} - y_{\tilde m}|, 1, \underline{M}_{i-1}^{\frac{1}{\tau-1}} r_{i-1}^{d\frac{1-\delta}{\tau-1}}\Big) \le 1 - \underline{c}\rho\Big(\sqrt{d} r_{i-2}, 1, \underline{M}_{i-1}^{\frac{1}{\tau-1}} r_{i-1}^{d\frac{1-\delta}{\tau-1}}\Big). \tag{5.160}$$

Here ρ is as in (3.6) of Assumption 3.2 and we have used that ρ is decreasing in its first component. Next, we know that because $N \ge \underline{N}_{5.25}$ for all $i \in \{3, \dots, k+1\}$ it holds that

$$r_{i-1} \ge N^{\gamma^{i-1}} \qquad \text{and} \qquad r_{i-2} \le 2N^{\gamma^{i-2}} \tag{5.161}$$

Furthermore, because $N \ge \underline{N}_{5.25}$, by item (1) of Claim 5.12 it holds that $\underline{M}_{i-1} \ge 1/(\phi\eta)$, where $\phi = (1+\sqrt{5})/2$. Then, filling in the definition of ρ and applying (5.161), we find that

$$\begin{aligned}\mathbb{P}(x_{\tilde n} \not\leftrightarrow y_{\tilde n} | \mathcal{F}_{\tilde n, \tilde m}) &= 1 - \underline{c}\bigg(1 \wedge \frac{\underline{M}_{i-1}^{1/(\tau-1)} r_{i-1}^{d(1-\delta)/(\tau-1)} \cdot 1^{\sigma}}{(\sqrt{d})^d r_{i-2}^d}\bigg)^{\alpha} \\ &\le 1 - C_1 N^{0 \wedge d\alpha\gamma^{i-1}\left(\frac{1-\delta}{\tau-1} - \frac{1}{\gamma}\right)} \\ &= 1 - C_1 N^{d\alpha\gamma^{i-1}\left(\frac{1-\delta}{\tau-1} - \frac{1}{\gamma}\right)}.\end{aligned} \tag{5.162}$$

Here $C_1 = C_1(\alpha, \tau, \eta, d) > 0$ is constant with respect to i and N. Furthermore, in the last equation of (5.162) we have used that $1 - \delta < 1$, $\tau - 1 > 1$ and $1/\gamma > 1$ by definition of δ, τ and γ (see Setting 5.3), so that $(1-\delta)/(\tau-1) - 1/\gamma < 0$. Then, we combine (5.161) with (5.155) and (5.156) to see that $n \ge \eta N^{d\gamma^{i-1}}/4$ and $m \ge \eta N^{d\delta\gamma^{i-1}}/8$. By substituting this into (5.157) and using $1 - x \le e^{-x}$, we obtain

$$\begin{aligned}1 - \Psi^1_{\mathbf{t}_{i-2}} = \prod_{\substack{\tilde n \le n \\ \tilde m \le m}} \mathbb{P}(x_{\tilde n} \not\leftrightarrow y_{\tilde m} | \mathcal{F}_{\tilde n, \tilde m}) &\le \prod_{\substack{\tilde n \le n \\ \tilde m \le m}} \Big(1 - C_1 N^{d\alpha\gamma^{i-1}\left(\frac{1-\delta}{\tau-1} - \frac{1}{\gamma}\right)}\Big) \\ &\le \exp\Big[- C_1 nm N^{d\alpha\gamma^{i-1}\left(\frac{1-\delta}{\tau-1} - \frac{1}{\gamma}\right)}\Big] \\ &\le \exp\Big[- C_2 N^{d\gamma^{i-1}\left(1+\delta+\alpha\frac{1-\delta}{\tau-1} - \frac{\alpha}{\gamma}\right)}\Big] =: \mathrm{err}_3(\mathrm{i}-1).\end{aligned} \tag{5.163}$$

Here $C_2 = C_2(\alpha, \tau, \eta, d) > 0$ is a constant. We set

$$R := 1 + \delta + \alpha\frac{1-\delta}{\tau-1} - \frac{\alpha}{\gamma} \tag{5.164}$$

and notice that $R > 0$ because of the same explanation that was given in (5.16).
(Part 6: conclusion) We now take everything together. By substituting (5.163) and (5.153) into (5.154), we obtain that $\Psi_{\mathbf{t}_{i-2}} \le 1 - (1 - \mathrm{err}_3(i-1))(1 - \mathrm{err}_2(i-1))$. We may further bound this by the inequality $(1-x)(1-y) \ge 1 - x - y$ to see that $\Psi_{\mathbf{t}_{i-2}} \le \mathrm{err}_2(i-1) + \mathrm{err}_3(i-1)$. This, in turn we substitute into (5.147) (with $\Psi_{\mathbf{t}_{i-2}}$ as defined in (5.148)) to observe that if $i \ge 2$

$$\mathcal{Q}_i \ge 1 - \sum_{\mathbf{t}_{i-2} \in \{0,1\}^{i-2}} (\mathrm{err}_2(i-1) + \mathrm{err}_3(i-1)) = 1 - 2^{i-2}(\mathrm{err}_2(i-1) + \mathrm{err}_3(i-1)). \tag{5.165}$$

Recall that $\mathcal{Q}_1 = 1$. Then, by substituting (5.165) into (5.145) and using repeatedly that $(1-x)(1-y) \ge 1 - x - y$ obtain that

$$\mathcal{Q} = \prod_{i=1}^{k+1} \mathcal{Q}_i \ge \prod_{i=2}^{k+1} (1 - 2^{i-2}(\mathrm{err}_2(i-1) + \mathrm{err}_3(i-1))) \ge 1 - \sum_{i=2}^{k+1} 2^{i-2}(\mathrm{err}_2(i-1) + \mathrm{err}_3(i-1)). \tag{5.166}$$

Now notice that $\mathrm{err}_2(i-1)$ and $\mathrm{err}_3(i-1)$ are increasing in i (see (5.153) resp. (5.163)). As such, for all $i \in \{2, \dots, k+1\}$ we have that $\mathrm{err}_2(i-1) + \mathrm{err}_3(i-1) \le \mathrm{err}_2(k) + \mathrm{err}_3(k)$. Furthermore, we may observe that $\sum_{i=2}^{k+1} 2^{i-2} = 2^k - 1 \le 2^k$. Then, recall $k^\star_\varepsilon(N) = (\ln\ln N - \varepsilon \ln\ln\ln N)/\ln(1/\gamma)$ from (5.29) of Definition 5.5 and set $\Delta(\gamma) := \ln(2)/\ln(1/\gamma)$. By elementary computation, we observe that

$$2^{k^\varepsilon_\star(N)} = 2^{\frac{\ln\ln N - \varepsilon \ln\ln\ln N}{\ln(1/\gamma)}} = \frac{(\ln N)^{\Delta(\gamma)}}{(\ln\ln N)^{\varepsilon\Delta(\gamma)}}. \tag{5.167}$$

Applying the last three observations to (5.166), applying that $k < k^\star_\varepsilon(N)$, using the definitions of err_2 and err_3 from (5.153) and (5.163) and Claim 5.6 to bound $N^{C\gamma^k}$, we see that

$$\begin{aligned}\mathcal{Q} &\ge 1 - 2^k(\mathrm{err}_2(k) + \mathrm{err}_3(k)) \\ &\ge 1 - \frac{(\ln N)^{\Delta(\gamma)}}{(\ln\ln N)^{\varepsilon\Delta(\gamma)}}\bigg(\exp\Big[-\frac{\eta}{32}\exp\big[d\delta(\ln\ln N)^{\varepsilon}\big]\Big] + \exp\Big[-C_2\exp\big[dR(\ln\ln N)^{\varepsilon}\big]\Big]\bigg) \\ &=: 1 - \tilde g(N, \varepsilon, \gamma, \delta, \eta, \alpha, \tau, d).\end{aligned} \tag{5.168}$$

One can show that $\tilde g(N,\gamma,\delta,\eta,\alpha,\tau,\varepsilon)\to 0$ if $N\to\infty$. Finally, we return to equations (5.137) and (5.138) to observe that

$$\begin{aligned}\mathbb{P}(\mathcal{H}_k \text{ is a hierarchy of depth } k) &\geq (1-\tilde g(N,\varepsilon,\gamma,\delta,\eta,\alpha,\tau,d)(1-kr_k^{-d\delta})^2\\ &\geq 1-\tilde g(N,\varepsilon,\gamma,\delta,\eta,\alpha,\tau,d)-2kr_k^{-d\delta}\\ &=: 1-\mathrm{err}_{5.25}(N,\varepsilon,\gamma,\delta,\eta,\alpha,\tau,d). \end{aligned}\tag{5.169}$$

To show that $\mathrm{err}_{5.25}\to 0$ if $N\to 0$, it remains to show that $2kr_k^{-d\delta}\to 0$ if $N\to 0$. To do this, we use that $r_k\geq N^{\gamma^k}$ by item (2) of Lemma 5.7, $k<k^\star_\varepsilon(N)$ and Claim 5.6 to observe that

$$kr_k^{-d\delta}\leq kN^{-d\delta\gamma^k}\leq \frac{\ln\ln N}{\ln(1/\gamma)}\exp\big[-d\delta(\ln\ln N)^\varepsilon\big].\tag{5.170}$$

Since the right-hand side of (5.170) goes to 0 if $N\to\infty$, we may conclude that $kr_k^{-d\delta}\to 0$ too if $N\to\infty$. We conclude that in (5.169) we have that $\mathrm{err}_{5.25}(N,\varepsilon,\gamma,\delta,\eta,\alpha,\tau,d)\to 0$ if $N\to\infty$. This finishes the proof. ∎

We are now ready to prove Proposition 5.2. We repeat it here for convenience.

Proposition 5.2 *Consider a KSRG $\mathcal{G}=(\mathcal{V},\mathcal{E})$ from Definition 3.1 satisfying Assumption 3.2. Assume that $\mathcal{V}=\mathbb{Z}^d$ and that the KSRG has parameters $d\in\mathbb{N}$, $\sigma_1=1$, $\sigma_2=\sigma\geq 0$, $\tau\in(2,3)$ and α such that $\tau-1<\alpha<(\tau-1)/(\tau-2)$. Furthermore, assume that all nearest-neighbour edges are present in $\mathcal{E}$. Let $\varepsilon>0$ and set*

$$\Delta=\frac{\ln 2}{\ln\big(\frac{\alpha+\tau-1}{\alpha(\tau-1)}\big)}.\tag{5.3}$$

Let $u,v\in\mathcal{V}$. Then, there exists an $\underline{N}_{5.2}$ and a function $\mathrm{err}_{5.2}(|u-v|)=\mathrm{err}_{5.2}(|u-v|,\varepsilon,\alpha,\tau,d)$ that goes to 0 if $|u-v|\to\infty$, such that if $|u-v|\geq\underline{N}_{5.2}$, then

$$\mathbb{P}\big(d_{\mathcal{G}}(u,v)\leq(\ln|u-v|)^{\Delta+\varepsilon}\big)\geq 1-\mathrm{err}_{5.2}(|u-v|,\varepsilon,\alpha,\tau,d).\tag{5.4}$$

Proof of Proposition 5.2. Observe that

$$\frac{\ln 2}{\ln(1/\gamma)}\downarrow\frac{\ln 2}{\ln\big(\frac{\alpha+\tau-1}{\alpha(\tau-1)}\big)}\qquad\text{if}\qquad\gamma\downarrow\frac{\alpha(\tau-1)}{\alpha+\tau-1}.\tag{5.171}$$

As such, it is possible to choose $\gamma_\star=\gamma_\star(\varepsilon,\alpha,\tau)>\alpha(\tau-1)/(\alpha+\tau-1)$ such that

$$\frac{\ln 2}{\ln(1/\gamma_\star)}\leq\frac{\ln 2}{\ln\big(\frac{\alpha+\tau-1}{\alpha(\tau-1)}\big)}+\frac{\varepsilon}{2}.\tag{5.172}$$

Next, choose any $\tilde\varepsilon\in(0,1)$. Set $N:=|u-v|$ and let $k<k^\star_{\tilde\varepsilon}(N)$. We use the law of total probability to see that

$$\begin{aligned}\mathbb{P}\big(d_{\mathcal{G}}(u,v)\leq(\ln|u-v|)^{\Delta+\varepsilon}\big)\geq\mathbb{P}\big(d_{\mathcal{G}}(u,v)\leq(\ln|u-v|)^{\Delta+\varepsilon}\;\big|\;\mathcal{H}_k(u,v)\text{ is a hierarchy of depth }k\big)\\ \cdot\,\mathbb{P}(\mathcal{H}_k(u,v)\text{ is a hierarchy of depth }k).\end{aligned}\tag{5.173}$$

We aim to reason that there exists an $\underline{\tilde N}_{5.2}$ such that if $N\geq\underline{\tilde N}_{5.2}$ and $\mathcal{H}_k$ from Definition 5.24 is a hierarchy of depth k, then $d_{\mathcal{G}}(u,v)\leq(\ln|u-v|)^{\Delta+\varepsilon}$. To this end, notice that by construction there are $\sum_{i=1}^{k-1}2^i=2^k-1\leq 2^k$ edges present between the vertices of $\mathcal{H}_k$. Furthermore, the vertices of $\mathcal{H}_k$ form a broken path between u and v, with gaps between the vertices of the form $Z_{\mathbf{t}_{k-1}00}$ and $Z_{\mathbf{t}_{k-1}01}$ or the vertices of the form $Z_{\mathbf{t}_{k-1}10}$ and $Z_{\mathbf{t}_{k-1}11}$. Shorthand, we may summarise that we only need to connect $Z_{\mathbf{t}_k0}$ with $Z_{\mathbf{t}_k1}$ for $\mathbf{t}_k\in\{0,1\}^k$. Recall $r_k=r_k(N,\gamma_\star)$ from Definition 5.4. Then, by (5.131) we see that for each $\mathbf{t}_k\in\{0,1\}^k$ it holds that $Z_{\mathbf{t}_k0}$ and $Z_{\mathbf{t}_k1}$ are contained within the same layer k box (i.e., $\mathcal{B}^k_\Lambda(Z_{\mathbf{t}_k0})=\mathcal{B}^k_\Lambda(Z_{\mathbf{t}_k1})$) and hence from (5.132) it follows that $|Z_{\mathbf{t}_k0}-Z_{\mathbf{t}_k1}|\leq\sqrt{d}r_k$. Set $\underline{N}_1=\underline{N}_1(\tilde\varepsilon,\gamma_\star)=\underline{N}_{5.7}(\tilde\varepsilon,\gamma_\star)$. Then by item (2) of Lemma 5.7, if $N\geq\underline{N}_1$ it holds that $r_k\leq 2N^{\gamma_\star^k}$. It follows that there exists a constant $C_1=C_1(d)>0$ such that for all $\mathbf{t}_k\in\{0,1\}^k$, we may connect $Z_{\mathbf{t}_k0}$ and $Z_{\mathbf{t}_k1}$ with at most $C_1N^{\gamma_\star^k}$ nearest-neighbour edges. Since there are 2^k such pairs, this requires at most $C2^kN^{\gamma_\star^k}$ edges. The path that we have constructed in this way connects u and v. It remains to rewrite the amount of edges required in the form given by (5.4). To this end, notice that the path from u to v we have constructed uses at most $2^k+C_12^kN^{\gamma_\star^k}$ vertices, which in turn implies that $d_{\mathcal{G}}(u,v)\leq 2^k+C_12^kN^{\gamma_\star^k}$. By using that $k<k^\star_{\tilde\varepsilon}(N)$ and applying Claim 5.6, with some rewriting and elementary computation we may rewrite

$$\begin{aligned}d_{\mathcal{G}}(u,v)&\leq\frac{(\ln N)^{\frac{\ln 2}{\ln(1/\gamma_\star)}}}{(\ln\ln N)^{\tilde\varepsilon\frac{\ln 2}{\ln(1/\gamma_\star)}}}+C_1\frac{(\ln N)^{\frac{\ln 2}{\ln(1/\gamma_\star)}}}{(\ln\ln N)^{\tilde\varepsilon\frac{\ln 2}{\ln(1/\gamma_\star)}}}\exp[(\ln\ln N)^{\tilde\varepsilon}]\\ &=(\ln N)^{\frac{\ln 2}{\ln(1/\gamma_\star)}}\Big(\underbrace{(\ln\ln N)^{-\tilde\varepsilon\frac{\ln 2}{\ln(1/\gamma_\star)}}+C_1(d)(\ln\ln N)^{-\tilde\varepsilon\frac{\ln 2}{\ln(1/\gamma_\star)}}\exp[(\ln\ln N)^{\tilde\varepsilon}]}_{=:f(N,\tilde\varepsilon,\gamma_\star,d)}\Big).\end{aligned}\tag{5.174}$$

One may verify that $f(N,\tilde\varepsilon,\gamma_\star,d)=o\big((\ln N)^{\varepsilon/2}\big)$ (see also Claim A.4). We may therefore choose $\underline{N}_2=\underline{N}_2(\varepsilon,\tilde\varepsilon,\gamma_\star,d)$ such that if $N\geq\underline{N}_2$, then $f(N,\tilde\varepsilon,\gamma_\star,d)\leq(\ln N)^{\varepsilon/2}$. We conclude that if $N\geq\underline{N}_1\vee\underline{N}_2$

and $\mathcal{H}_k$ is present, then $d_{\mathcal{G}}(u,v) \le (\ln N)^{\Delta+\varepsilon}$. Furthermore, by Lemma 5.25, there exists an $\underline{N}_3 = \underline{N}_3(\tilde{\varepsilon}, \gamma_\star, \delta, \eta, \tau, d) = \underline{N}_{5.25}(\tilde{\varepsilon}, \gamma_\star, \delta, \eta, \tau, d)$ such that

$$\mathbb{P}(\mathcal{H}_k(u,v) \text{ is a hierarchy of depth } k) \ge 1 - \mathrm{err}_{5.25}(N, \tilde{\varepsilon}, \gamma_\star, \delta, \eta, \alpha, \tau, d). \tag{5.175}$$

We return to (5.173). If we set $\underline{\tilde{N}}_{5.2} = \underline{\tilde{N}}_{5.2}(\varepsilon, \tilde{\varepsilon}, \gamma_\star, \delta, \eta, \tau, d) = \underline{N}_1(\tilde{\varepsilon}, \gamma_\star) \vee \underline{N}_2(\varepsilon, \tilde{\varepsilon}, \gamma_\star, d) \vee \underline{N}_3(\tilde{\varepsilon}, \gamma_\star, \delta, \eta, \tau, d)$, then if $N \ge \underline{\tilde{N}}_{5.2}$ by the above we find that

$$\mathbb{P}\big(d_{\mathcal{G}}(u,v) \le (\ln|u-v|)^{\Delta+\varepsilon} \bigm| \mathcal{H}_k(u,v) \text{ is a hierarchy of depth } k\big) = 1. \tag{5.176}$$

Combining this with (5.173) and (5.175) yields

$$\mathbb{P}\big(d_{\mathcal{G}}(u,v) \le (\ln|u-v|)^{\Delta+\varepsilon}\big) \ge 1 - \mathrm{err}_{5.25}(N, \tilde{\varepsilon}, \gamma_\star, \delta, \eta, \alpha, \tau, d). \tag{5.177}$$

Then, notice that $\gamma_\star$ depends on ε, α and τ. Fix $\delta_\star$ such that $\delta_\star \in I_\delta(\gamma_\star)$ (see Setting 5.3) and such that $R = 1 + \delta_\star + \alpha(1-\delta_\star)/(\tau-1) - \alpha/\gamma_\star > 0$ (note that this requirement is implicit used for $\mathrm{err}_{5.25}$). If we now fix $\tilde{\varepsilon} = 1/10$, $\eta = 1/2$, we may set

$$\underline{N}_{5.2} = \underline{N}_{5.2}(\varepsilon, \alpha, \tau, d) = \underline{\tilde{N}}_{5.2}(\varepsilon, 1/10, \gamma_\star(\varepsilon, \alpha, \tau), \delta_\star, 1/2, \alpha, \tau, d) \qquad \text{and} \tag{5.178}$$

$$\mathrm{err}_{5.2}(N, \varepsilon, \alpha, \tau, d) = \mathrm{err}_{5.25}(N, 1/10, \gamma_\star(\varepsilon, \alpha, \tau), \delta_\star, 1/2, \alpha, \tau, d) \tag{5.179}$$

to see that from (5.177) it follows that if $N \ge \underline{N}_{5.2}$, then

$$\mathbb{P}\big(d_{\mathcal{G}}(u,v) \le (\ln|u-v|)^{\Delta+\varepsilon}\big) \ge 1 - \mathrm{err}_{5.2}(N, \varepsilon, \alpha, \tau, d). \tag{5.180}$$

Furthermore, notice that by Lemma 5.25 the right-hand side of (5.179) goes to 0 if $N \to \infty$. From this it follows that also $\mathrm{err}_{5.2}(N, \varepsilon, \alpha, \tau, d) \to 0$ if $N \to \infty$. This finishes the proof. ■

We remark that in (5.178) and (5.179), the choice of $\tilde{\varepsilon} = 1/10$ and $\eta = 1/2$ is completely arbitrary. Similarly, if its constraints are satisfied the choice of $\delta_\star$ is also arbitrary. One may choose them differently and obtain a similar result. It should be noted, however, that varying one of the parameters $\tilde{\varepsilon}, \delta$ or η will likely have opposing effects on $\underline{N}_{5.2}$ and $\mathrm{err}_{5.2}$. In particular, if one wants the result of (5.4) to hold for more values of $N := |u-v|$ (i.e., decrease $\underline{N}_{5.2}$), then that necessarily requires increasing the error bound $\mathrm{err}_{5.2}$. The opposite also holds.

5.4. Other poly-logarithmic upper bounds

In the previous subsections, we have shown a poly-logarithmic upper bound for distances for a specific set of parameters, specifically $\sigma_1 = 1, \sigma_2 = \sigma \in [0,1)$, $\tau \in (2,3)$ and α such that $\tau - 1 < \alpha < (\tau-1)/(\tau-2)$ and $\alpha\sigma \le \tau - 1$. In this subsection, we reason that the same ideas from the previous subsections may be used to also find poly-logarithmic upper bounds for more values of σ, α and τ. We do not show these results as thoroughly as we have done earlier in this section, but we note that the same proof technique works with relatively few modifications.

Recall from that in the proof of Proposition 5.2 (in particular equations (5.164) and (5.171)), the optimal value of the exponent Δ depends on the lowest value that γ can attain. This lower bound is determined by the inequality $R > 0$, as for example given in (5.16) or (5.164). However, this value of R is specific to this set of parameters and the choice to search for one high-weight vertex and one constant-weight vertex. In this section, we generalise this idea.

We imitate the sketch given in Subsection 5.1, but immediately continue to the i'th step. Rather than searching for one high-weight vertex (i.e., with weight $\Theta\big(N^{d\gamma^i(1-\delta)/(\tau-1)}\big)$) and one constant-weight vertex (i.e., with weight $\Theta(1)$) in the i'th step as done in Subsection 5.1, we search for one vertex with weight $\Theta\big(N^{d\gamma^i\zeta_1/(\tau-1)}\big)$ and one vertex with weight $\Theta\big(N^{d\gamma^i\zeta_2/(\tau-1)}\big)$, where $\zeta_1, \zeta_2 \ge 0$. Note that without loss of generality, we may assume that $\zeta_1 \ge \zeta_2$, otherwise we interchange their role. In the i'th step, we put a box with sidelengths N^{γ^i} around $z_{\mathbf{t}_{i-1}00}$ and search for all vertices with weight $\Theta\big(N^{d\gamma^i\zeta_1/(\tau-1)}\big)$. We note that

$$\mathbb{P}\big(W = \Theta\big(N^{d\gamma^i\zeta_1/(\tau-1)}\big)\big) = \Theta\big(N^{-d\gamma^i\zeta_1}\big), \tag{5.181}$$

so that this box contains (roughly) $N^{d\gamma^i}\Theta\big(N^{-d\gamma^i\zeta_1}\big) = \Theta\big(N^{d\gamma^i(1-\zeta_1)}\big)$ vertices. Similarly, we put a box with sidelengths N^{γ^i} around $z_{\mathbf{t}_{i-1}11}$ and search for vertices with weight $\Theta\big(N^{d\gamma^i\zeta_2/(\tau-1)}\big)$; there are (roughly) $\Theta\big(N^{d\gamma^i(1-\zeta_2)}\big)$ such vertices. We note that for this to make sense, $\Theta\big(N^{d\gamma^i(1-\zeta_1)}\big)$ and $\Theta\big(N^{d\gamma^i(1-\zeta_2)}\big)$ needs to increase when N increases. Therefore, we impose that $\zeta_1, \zeta_2 < 1$. We continue by doing a similar computation to (5.14): if x is in the box around $z_{\mathbf{t}_{i-1}00}$ and such that $W_x = \Theta\big(N^{d\gamma^i\zeta_1/(\tau-1)}\big)$ and y is in the box around $z_{\mathbf{t}_{i-1}11}$ such that $W_y = \Theta\big(N^{d\gamma^i\zeta_2/(\tau-1)}\big)$, then $|x-y| = \Theta\big(N^{\gamma^{i-1}}\big)$ and

$$\mathbb{P}(x \leftrightarrow y | W_x, W_y) \ge \underline{c}\left(1 \wedge \frac{\Theta\big(N^{\gamma^i\frac{\zeta_1}{\tau-1}}\big)\Theta\big(N^{\gamma^i\frac{\zeta_2}{\tau-1}}\big)^\sigma}{\Theta\big(N^{\gamma^{i-1}}\big)^d}\right)^\alpha = \Theta\big(N^{0\wedge d\gamma^i\alpha\left(\frac{\zeta_1+\sigma\zeta_2}{\tau-1}-\frac{1}{\gamma}\right)}\big). \tag{5.182}$$

Note that here we have used that $\zeta_1 \geq \zeta_2$. Therefore, by mimicking (5.15) we find that the probability that *none* of the the vertices from the box around $z_{\mathbf{t}_{i-1}00}$ with weight $\Theta\big(N^{d\gamma^i\zeta_1/(\tau-1)}\big)$ is connected to any vertex of the box around $z_{\mathbf{t}_{i-1}11}$ with weight $\Theta\big(N^{d\gamma^i\zeta_2/(\tau-1)}\big)$ is (approximately) bounded by

$$\prod_{i=1}^{\Theta\left(N^{d\gamma^i(1-\zeta_1)}\right)} \prod_{j=1}^{\Theta\left(N^{d\gamma^i(1-\zeta_2)}\right)} \Big(1-\Theta\big(N^{0\wedge d\gamma^i\alpha\left(\frac{\zeta_1+\sigma\zeta_2}{\tau-1}-\frac{1}{\gamma}\right)}\big)\Big) \leq \exp\Big[-\Theta\Big(N^{dR\gamma^i}\Big)\Big], \tag{5.183}$$

where we have set

$$R = R(\zeta_1,\zeta_2,\gamma,\alpha,\tau,\sigma) = 2-\zeta_1-\zeta_2+0\wedge\left(\alpha\frac{\zeta_1+\sigma\zeta_2}{\tau-1}-\frac{\alpha}{\gamma}\right). \tag{5.184}$$

As suggested, if $R>0$ then we are able to show that the construction succeeds. We analyse for which values of ζ_1,ζ_2 and γ this is true. The goal is to do this in such a way that γ can achieve the lowest value possible, since ultimately if γ attains smaller values then the value of the exponent $\Delta(\gamma)=\ln(2)/\ln(1/\gamma)$ of the poly-logarithmic upper-bound becomes smaller. We denote this smallest possible value for γ by $\gamma_\star$. We claim that we may remove the minimisation with 0 in R and are only required to analyse

$$R = 2-\zeta_1-\zeta_2+\alpha\frac{\zeta_1+\sigma\zeta_2}{\tau-1}-\frac{\alpha}{\gamma}. \tag{5.185}$$

Indeed, if the minimisation *were* necessary, then this would be because ζ_1 and ζ_2 were large enough. However, then increasing ζ_1 and ζ_2 only decreases R because of the term $2-\zeta_1-\zeta_2$. In fact, we will later see that indeed all values we obtain justify removing the minimisation. We therefore continue with the expression in (5.185). With elementary rewriting, we may show that $R>0$ is equivalent to

$$\gamma > \frac{\alpha(\tau-1)}{(\alpha-(\tau-1))\zeta_1+(\alpha\sigma-(\tau-1))\zeta_2+2(\tau-1)} =: \tilde{\gamma}_\star(\zeta_1,\zeta_2,\alpha,\tau,\sigma). \tag{5.186}$$

Therefore, to obtain $\gamma_\star$ we need to find the minimal value of $\tilde{\gamma}_\star(\zeta_1,\zeta_2\alpha,\tau,\sigma)$. Recall however that we must do so under the requirements that $0\leq\zeta_2\leq\zeta_1<1$. We conclude that that

$$\gamma_\star = \inf_{0\leq\zeta_2\leq\zeta_1<1}\tilde{\gamma}_\star(\zeta_1,\zeta_2,\alpha,\tau,\sigma). \tag{5.187}$$

We compute this infimum by considering different cases for α,τ and σ. We do so with three examples, after which we note that all other cases can be reasoned in the same way.

- *(Ex. 1: $\alpha<\tau-1,\alpha\sigma<\tau-1$)* In this case, increasing ζ_1 and ζ_2 decreases the denominator, which increases $\tilde{\gamma}_\star$. Therefore, we find the infimum at $\zeta_1=\zeta_2=0$, and $\gamma_\star=\alpha/2$.
- *(Ex. 2: $\alpha>\tau-1,\alpha\sigma\leq\tau-1$)* In this case, increasing ζ_1 decreases $\tilde{\gamma}_\star$, while increasing ζ_2 increases $\tilde{\gamma}_\star$. Therefore, we find the infimum by letting $\zeta_1\uparrow 1$ and setting $\zeta_2=0$. This yields $\gamma_\star=\alpha(\tau-1)/(\alpha+\tau-1)$. We note that this is the setting of Proposition 5.2.
- *(Ex. 3: $\alpha<\tau-1,\alpha\sigma>\tau-1$)* In this case, we have to take care of the restraint $\zeta_1\geq\zeta_2$. We see that increasing ζ_1 increases $\tilde{\gamma}_\star$, while increasing ζ_2 decreases $\tilde{\gamma}_\star$. We conclude that to achieve the infimum, we need to choose ζ_1 as small as possible, i.e., we set $\zeta_1=\zeta_2$. We now arrive at a second disjunction:
  - If $|\alpha-(\tau-1)|>\alpha\sigma-(\tau-1)$, then $(\alpha-(\tau-1))\zeta_1+(\alpha\sigma-(\tau-1))\zeta_2=(\alpha-(\tau-1)+\alpha\sigma-(\tau-1))\zeta_2$ is decreasing in ζ_2 and hence setting $\zeta_2=0$ yields the infimum; $\gamma_\star=\alpha/2$.
  - However, if $|\alpha-(\tau-1)|<\alpha\sigma-(\tau-1)$, then $(\alpha-(\tau-1)+\alpha\sigma-(\tau-1))\zeta_2$ is increasing and therefore $\zeta_2\uparrow 1$ yields the infimum; $\gamma_\star=(\tau-1)/(1+\sigma)$.

In total, we obtain the following Table 1.

Restrictions on α,τ,σ	Optimal values ζ_1,ζ_2	$\gamma_\star$
$\alpha\leq\tau-1,\alpha\sigma\leq\tau-1$	$\zeta_1,\zeta_2=0$	$\frac{\alpha}{2}$
$\alpha>\tau-1,\alpha\sigma\leq\tau-1$	$\zeta_1\uparrow 1,\zeta_2=0$	$\frac{\alpha(\tau-1)}{\alpha+\tau-1}$
$\alpha\leq\tau-1,\alpha\sigma>\tau-1$, $\lvert\alpha-(\tau-1)\rvert<\alpha\sigma-(\tau-1)$	$\zeta_1,\zeta_2\uparrow 1$	$\frac{\tau-1}{1+\sigma}$
$\alpha\leq\tau-1,\alpha\sigma>\tau-1$, $\lvert\alpha-(\tau-1)\rvert=\alpha\sigma-(\tau-1)$	$\zeta_1=\zeta_2\in(0,1)$	$\frac{\alpha}{2}$
$\alpha\leq\tau-1,\alpha\sigma>\tau-1$, $\lvert\alpha-(\tau-1)\rvert>\alpha\sigma-(\tau-1)$	$\zeta_1,\zeta_2=0$	$\frac{\alpha}{2}$
$\alpha>\tau-1,\alpha\sigma>\tau-1$	$\zeta_1,\zeta_2\uparrow 1$	$\frac{\tau-1}{1+\sigma}$

TABLE 1. A table summarising the optimal lower-bound $\gamma_\star$ for γ such that $R>0$ in (5.185). In the left-most column we distinguish all cases based on the value of α,τ and σ. In the middle column we note which values for ζ_1 and ζ_2 should be taken in each case. The right-most column denotes the value of $\gamma_\star$ that holds in each case.

Our techniques may readily be generalised to show the same statement as Proposition 5.2, but with the restrictions replaced by the left-most column of Table 1 and Δ replaced by $\ln(2)/\ln(1/\gamma_\star)$ for the corresponding value of $\gamma_\star$ from the right-most column of Table 1, provided that $\gamma_\star < 1$. We note that $\gamma_\star < 1$ is necessary because $\gamma > \gamma_\star$ and we require that $\gamma < 1$.

Note that for the entries with $\gamma_\star = (\tau - 1)/(1 + \sigma)$, $\gamma_\star < 1$ implies that $\tau < 2 + \sigma$. As we will see in the following Section 6, there is much stricter upper bound in these cases. As such, these cases will not appear in the phase diagram given in Figure 2.

6. Doubly logarithmic upper-bound for distances

In this section, we prove that if we choose the parameters in a certain way, then a KSRG under Assumption 3.2 shows ultra-small behaviour. That is, the graph distance of two vertices $u, v \in \mathcal{V}$ grows doubly logarithmically with the spatial distance $|u-v|$ with high probability as $|u-v| \to \infty$. We show this under the condition that all nearest-neighbour edges are present, i.e., we assume that if $x, y \in \mathcal{V} = \mathbb{Z}^d$ are such that $|x-y| = 1$, then $xy \in \mathcal{E}$. In particular, we show Theorem 6.1. We do so for a KSRG with parameters $\sigma_1 = 1$ and $\sigma_2 = \sigma > 0$. We note that by Claim 3.3 the result then follows for all $\sigma_1, \sigma_2 > 0$.

Theorem 6.1 *Consider a KSRG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from Definition 3.1 satisfying Assumption 3.2 with parameters $d \in \mathbb{N}$, $\sigma_1 = 1$, $\sigma_2 = \sigma > 0$, $\alpha > 1$, $\tau \in (2, 2+\sigma)$ and $\mathcal{V} = \mathbb{Z}^d$ and suppose that all nearest-neighbour edges of $\mathbb{Z}^d$ are present in $\mathcal{E}$. Then for every $\delta > 0$ it holds that*

$$\lim_{|u-v|\to\infty} \mathbb{P}\bigg(d_{\mathcal{G}}(u,v) \le \frac{2+\delta}{\ln\left(\frac{\sigma}{\tau-2}\right)} \ln\ln|u-v| \bigg) = 1. \tag{6.1}$$

To prove Theorem 6.1, we use the following Proposition 6.2.

Proposition 6.2 *Consider a KSRG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from Definition 3.1 satisfying Assumption 3.2 with parameters $d \in \mathbb{N}$, $\sigma_1 = 1$, $\sigma_2 = \sigma > 0$, $\alpha > 1$, $\tau \in (2, 2+\sigma)$ and $\mathcal{V} = \mathbb{Z}^d$ and suppose that all nearest-neighbour edges of $\mathbb{Z}^d$ are present in $\mathcal{E}$. Let $u, v \in \mathcal{V}$. Take any $\varepsilon \in (0,1)$ and $\delta > 0$. Then there exists an $\underline{N}_{6.2} = \underline{N}_{6.2}(\delta, \varepsilon, \sigma, \alpha, \tau, d)$ such that if $|u-v| \ge \underline{N}_{6.2}$, then*

$$\mathbb{P}\bigg(d_{\mathcal{G}}(u,v) \le \frac{2+\delta}{\ln\left(\frac{\sigma}{\tau-2}\right)} \ln\ln|u-v| \bigg) \ge 1-\varepsilon. \tag{6.2}$$

We note that the proof of Theorem 6.1 follows immediately from Proposition 6.2. The following Subsection 6.1 is dedicated to proving Proposition 6.2.

6.1. Proof of Proposition 6.2

To show Proposition 6.2, we construct a path between u and v that has length at most $(2+\delta)\ln\ln(|u-v|)/\ln(\sigma/(\tau-2))$ and show that this construction succeeds with probability at least $1-\varepsilon$. In Figure 5, an illustration of the path we construct is given. The description of Figure 5 contains roughly the steps we take to construct this path. To start, we highlight the setting we are working in throughout the entire subsection. Additionally to all parameters of the KSRG $\mathcal{G}$ given in Theorem 6.1 and Proposition 6.2, we define two extra parameters A and B. Roughly, A determines the amount of vertices we may choose from for each step in the weight-increasing path (see step (1) from Figure 5). B determines the speed with which the size of the layers grows (see the black boxes in Figure 5).

Setting 6.3 *Consider a KSRG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from Definition 3.1 satisfying Assumption 3.2 with parameters $d \in \mathbb{N}$, $\sigma_1 = 1$, $\sigma_2 = \sigma > 0$, $\alpha > 1$ and $\tau \in (2, 2+\sigma)$. Fix $A, B > 1$ two real-valued constants such that*

$$A \in \left(1, \frac{1+\sigma}{\tau-1}\right) =: I_A, \quad B \in \left(\frac{\sigma}{(\tau-1)A-1}, \frac{\sigma}{(\tau-1)\left(A - \frac{A-1}{\alpha}\right)-1} \right) =: I_B(A) = I_B. \tag{6.3}$$

We always suppress the dependence on σ and τ from I_A and I_B and see it as implicit.

In Setting 6.3, it may not a priori be clear that I_A and I_B are non-empty or that $B > 1$. We therefore verify this. To show that I_A is non-empty, we show that $(1+\sigma)/(\tau-1) > 1$. To this end, note that $(1+\sigma)/(\tau-1) > 1$ is equivalent to $\tau < 2+\sigma$, which is true by assumption. We conclude that I_A is non-empty and we may therefore indeed choose A such as described above. Next, we show that I_B is non-empty and that if we choose $B \in I_B$, then $B > 1$. First, we analyse the lower-end point of I_B. We note that since $\tau > 2$ by assumption and $A > 1$, it is also clear that $(\tau-1)A - 1 > 0$. Therefore, by rewriting we observe that

$$\frac{\sigma}{(\tau-1)A-1} > 1 \quad \Leftrightarrow \quad A < \frac{1+\sigma}{\tau-1}, \tag{6.4}$$

which is true because of the choice of A. We conclude that the lower end-point of I_B is greater than 1. Next, we analyse the upper end-point of I_B, starting with showing that the denominator is positive. We again rewrite to see that

$$(\tau-1)\left(A - \frac{A-1}{\alpha}\right) - 1 > 0 \quad \Leftrightarrow \quad A < \frac{\alpha\frac{1+\sigma}{\tau-1} - 1}{\alpha - 1}. \tag{6.5}$$

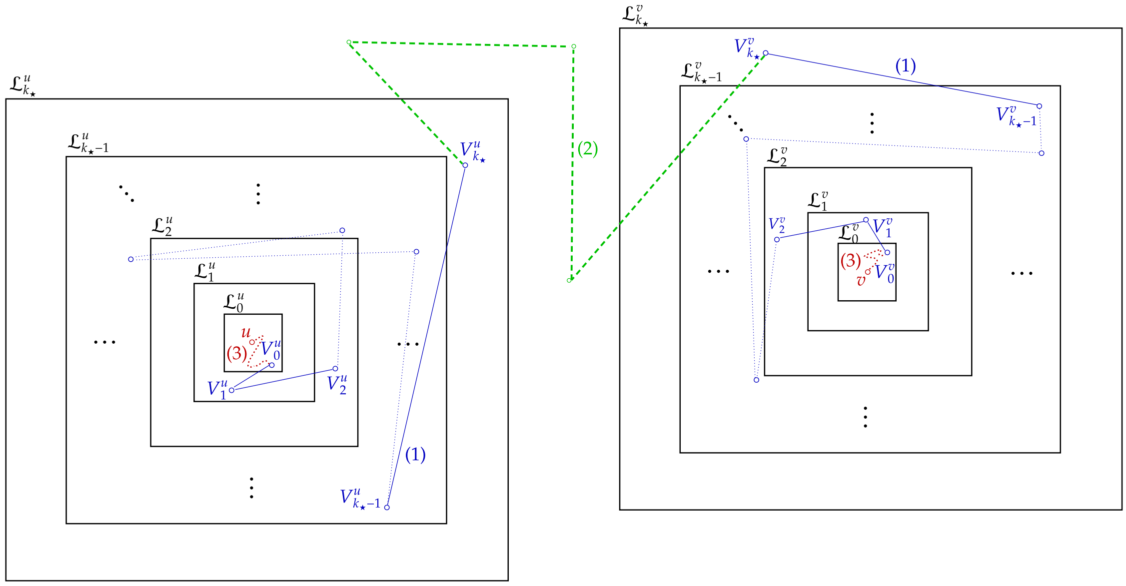


FIGURE 5. A figure illustrating the heuristic idea of the proof of Proposition 6.2. We want to construct a path between two distinct vertices $u, v \in \mathcal{V}$. To this end, we firstly define doubly exponentially increasing annuli $(\mathcal{L}^u_i)_i$ and $(\mathcal{L}^v_i)_i$ around u and v. These annuli are called layers and are given in black in the figure. In step (1), we construct a greedy weight-increasing path that has exactly one vertex in each layer, i.e., we sequentially build a path $V^u_0 V^u_1 \dots V^u_{k_\star-1} V^u_{k_\star}$ such that each V^u_i is in $\mathcal{L}^u_i$ and the weight of V^u_i is at least doubly exponential in i. Similarly, we construct such a path for v. Here $k_\star$ is a variable that signifies when we stop the process to ensure that step (2) can also be achieved. In particular, we set $k_\star$ to be $\mathcal{O}(\ln\ln|u-v|)$. This also implies that the length of the two paths we construct in step (1) is $\mathcal{O}(\ln\ln|u-v|)$. These paths are given in blue. Next, in step (2) we connect $V^u_{k_\star}$ and $V^v_{k_\star}$ in at most 4 edges. These are given in green. Lastly, in step (3) we connect u to V^u_0 and v to V^v_0 in a constant number of edges using nearest-neighbour edges. These paths are given in red. In total, therefore, this path from u to v utilises $\mathcal{O}(\ln\ln|u-v|)$ edges. Furthermore, each step may be achieved with high probability.

To show that the right-most inequality of (6.5) is true, we use that if $\alpha > 1$, then $(\alpha x - 1)/(\alpha - 1) > x$ for all $x > 1$. Since $(1+\sigma)/(\tau - 1) > A > 1$ and $(\alpha x - 1)/(\alpha - 1)$ is increasing in x, we therefore obtain

$$\frac{\alpha^{\frac{1+\sigma}{\tau-1}} - 1}{\alpha - 1} > \frac{\alpha A - 1}{\alpha - 1} > A. \tag{6.6}$$

We conclude that the right-most inequality of (6.5) is true, and hence that the denominator of the upper end-point of I_B is positive. From the fact that $(\tau-1)A - 1 > (\tau - 1)(A - (A-1)/\alpha) - 1$, it can quickly be observed that the upper end-point of I_B is indeed greater than the lower end-point. Taking everything together, we see that I_B is non-empty and lies completely above 1. We may therefore also choose B as described in (6.3).

Using the constants defined in Setting 6.3, for every vertex $v \in \mathcal{V}$ we define a sequence of doubly exponentially growing annuli centred around v. We call these annuli the layers around v. In Figure 5, these layers are illustrated in black.

Definition 6.4 (Layers) *Consider Setting 6.3, in particular A and B. Let $M_\varepsilon > 1$ and $v \in \mathcal{V} = \mathbb{Z}^d$. We define*

$$\mathcal{L}^v_0 = \mathcal{L}^v_0(M_\varepsilon, A, d) := \Big(v + \Big[-\frac{1}{2}M_\varepsilon^A, \frac{1}{2}M_\varepsilon^A\Big]^d\Big) \cap \mathcal{V} \qquad \text{and} \tag{6.7}$$

$$\mathcal{L}^v_i = \mathcal{L}^v_i(M_\varepsilon, A, B, d) := \Big(\Big(v + \Big[-\frac{1}{2}M_\varepsilon^{AB^i}, \frac{1}{2}M_\varepsilon^{AB^i}\Big]^d\Big) \cap \mathcal{V}\Big) \Big\backslash \bigcup_{j=0}^{i-1} \mathcal{L}^v_j \qquad \text{for } i \in \mathbb{N}. \tag{6.8}$$

*We refer to these sets as the **layers** around v. Furthermore, for $i \geq 0$, we refer to $\mathcal{L}^v_i$ as a **layer of order** i. When M_ε, A, B and d are clear from context we write $\mathcal{L}^v_0$ and $\mathcal{L}^v_i$.*

We continue by showing some useful properties that we use in later proofs of this section. In particular, we bound the distance between two vertices in the layers and bound the amount of vertices that each layer contains. These two properties are given in the following Claim 6.5.

Claim 6.5 (Properties of layers) *Consider Setting 6.3, in particular A and B. Let $M_\varepsilon > 1, v \in \mathcal{V} = \mathbb{Z}^d$ and let the layers $(\mathcal{L}_i^v)_{i\geq 0}$ be as given by Definition 6.4. The following statements hold:*

(1) Let $i, j \in \mathbb{N} \cup \{0\}$. Then, for all $x \in \mathcal{L}_i^v$ and $y \in \mathcal{L}_j^v$ it holds that $|x-y| \leq \sqrt{d} M_\varepsilon^{dAB^{i\vee j}}$.

(2) There exists a $\underline{M}_{6.5} = \underline{M}_{6.5}(A, B, d)$ such that for all $i \geq 0$ if $M_\varepsilon \geq M_{6.5}$, we have

$$|\mathcal{L}_i^v| \geq \frac{1}{2} M_\varepsilon^{dAB^i}. \tag{6.9}$$

Proof. Item (1) is a direct consequence of Lemma A.1. Next, we consider item (2). To this end, we note that for all $i \in \mathbb{N} \cup \{0\}$ it holds that

$$\left| \left(v + \left[-\frac{1}{2} M_\varepsilon^{AB^i}, \frac{1}{2} M_\varepsilon^{AB^i} \right]^d \right) \cap \mathcal{V} \right| = \left(2 \left\lfloor \frac{M_\varepsilon^{AB^i}}{2} \right\rfloor + 1 \right)^d. \tag{6.10}$$

We first consider a layer or order 0, i.e., $\mathcal{L}_0^v$. Then by (6.10) and the fact that $2\lfloor x/2 \rfloor \geq x - 2$, we obtain that

$$|\mathcal{L}_0^v| = \left(2 \left\lfloor \frac{M_\varepsilon^A}{2} \right\rfloor + 1 \right)^d \geq \left(M_\varepsilon^A - 1 \right)^d \geq \left(1 - M_\varepsilon^{-A} \right)^d M_\varepsilon^{dA} =: R_0(M_\varepsilon) M_\varepsilon^{dA}. \tag{6.11}$$

Similarly, if $i \geq 1$ we may again apply (6.10) and the fact that $x - 2 \leq 2\lfloor x/2 \rfloor \leq x$ to see that

$$\begin{aligned} |\mathcal{L}_i^v| &= \left(2 \left\lfloor \frac{M_\varepsilon^{AB^i}}{2} \right\rfloor + 1 \right)^d - \left(2 \left\lfloor \frac{M_\varepsilon^{AB^{i-1}}}{2} \right\rfloor + 1 \right)^d \\ &\geq \left(M_\varepsilon^{AB^i} - 1 \right)^d - \left(M_\varepsilon^{AB^{i-1}} + 1 \right) \\ &= \left[\left(1 - M_\varepsilon^{-AB^i} \right)^d - \left(M_\varepsilon^{-AB^{i-1}(B-1)} + M_\varepsilon^{-AB^i} \right)^d \right] M_\varepsilon^{dAB^i} =: R_i(M_\varepsilon) M_\varepsilon^{dB^i}. \end{aligned} \tag{6.12}$$

To show item 2, we want to show that there exists an $\underline{M}_{6.5}$ such that for all $j \geq 0$, $R_j(M_\varepsilon) \geq 1/2$ if $M_\varepsilon \geq M$. To this end, note that because $B > 1$, $R_j(M_\varepsilon)$ is increasing in j if $j \geq 1$. Hence

$$R_1(M_\varepsilon) \geq \frac{1}{2} \Rightarrow R_j(M_\varepsilon) \geq \frac{1}{2} \tag{6.13}$$

for all $j \geq 1$. Next, note that $R_0(M_\varepsilon) \to 1$ and $R_1(M_\varepsilon) \to 1$ if $M_\varepsilon \to \infty$. We conclude that there must be some $\underline{M}_1 = \underline{M}_1(A, d)$ such that if $M_\varepsilon \geq \underline{M}_1$, then $R_0(M_\varepsilon) \geq 1/2$. Similarly, there must be some $\underline{M}_2 = \underline{M}_2(A, B, d)$ such that if $M_\varepsilon \geq \underline{M}_2$, then $R_1(M_\varepsilon) \geq 1/2$. From (6.13) it then also follows that if $M_\varepsilon \geq \underline{M}_2$, then for all $j \geq 1$ it holds that $R_j(M_\varepsilon) \geq 1/2$. Thus, if we set $\underline{M}_{6.5} = \underline{M}_{6.5}(A, B, d) = \underline{M}_1(A, d) \vee \underline{M}_2(A, B, d)$, then if $M_\varepsilon \geq \underline{M}_{6.5}$ equation (6.9) holds, which is what we wanted to show. ■

Next, we continue with constructing paths through the layers. In particular, we build up a path such that its first vertex is in the layer of order 0 and each subsequent vertex is in a layer of one higher order. Furthermore, we do this in such a way that the weight along the path typically increases. For this reason, the path we construct is also called a *weight-increasing* path. This path is given in step (1) of Figure 5, where it is coloured blue. To develop this path, we firstly develop some notation. In particular, we define a short-hand notation to find all vertices in a set $E \subseteq \mathcal{V}$ that have a weight higher than a certain value K. Formally, we use the following Definition 6.6.

Definition 6.6 *Consider Setting 6.3. For any set $E \subseteq \mathcal{V} = \mathbb{Z}^d$ and $K \geq 1$ we set*

$$T_{\geq}(E, K) := \{ v \in E : W_v \geq K \}. \tag{6.14}$$

The function defined above essentially thins out a set of vertices based on their weight. For this reason, the function $T_{\geq}$ may be referred to as a *thinning function.*

Next, we develop notation for the random vertices that constitute the path described earlier. We denote these vertices $(V_i^v)_{i\geq 0}$ and we construct them in a specific way. We start by searching $\mathcal{L}_0^v$ for all vertices that have weight higher than $M_\varepsilon^{d/(\tau-1)}$. If such a vertex exists, we set V_0^v to be that vertex. If multiple such vertices exist, we set V_0^v to be the one with the highest weight. If no such vertex exists, we set V_0^v to be $\varnothing$. Next, we search for all vertices in $\mathcal{L}_1^v$ that have weight higher than $M_\varepsilon^{dB/(\tau-1)}$ *and* are connected to V_0^v. Again, we set V_1^v to be the vertex with the highest weight if such a vertex exist and $\varnothing$ otherwise. We then iterate this procedure for $i \geq 2$. In particular, we use the following Definition 6.7.

Definition 6.7 *Consider Setting 6.3, in particular A and B. Let $M_\varepsilon > 1$, $v \in \mathcal{V} = \mathbb{Z}^d$ and consider $(\mathcal{L}_i^v)_{i\geq 0}$ from Definition 6.4. Furthermore, let $T_\geq$ be from Definition 6.6. We define the random set $T_0^v = T_0^v(A,B,\tau,d) = T_\geq\left(\mathcal{L}_0^v, M_\varepsilon^{d/(\tau-1)}\right)$ and the random vertex*

$$V_0^v = V_0^v(T_0^v) = \begin{cases} \arg\max_{u\in T_0^v} W_u & \text{if } T_0^v \neq \varnothing \\ \varnothing & \text{if } T_0^v = \varnothing \end{cases}. \tag{6.15}$$

Then for $i \geq 1$ consecutively define $T_i^v = T_\geq\left(\mathcal{L}_i^v, M_\varepsilon^{dB^i/(\tau-1)}\right) \cap N_{\mathcal{G}}\left(V_{i-1}^v\right)$ and

$$V_i^v = V_i^v(V_{i-1}^v, T_i^v) = \begin{cases} \arg\max_{u\in T_i^v} W_u & \text{if } T_i^v \neq \varnothing \\ \varnothing & \text{if } T_i^v = \varnothing \end{cases}. \tag{6.16}$$

Suppose, in Definition 6.7 above, we have that $V_i^v \neq \varnothing$. Then it must hold that all V_j^v, $j \in \{0, i-1\}$ also satisfy $V_j^v \neq \varnothing$. Intuitively, this is because if at some point there is some V_j^v where the construction given in Definition 6.7 fails, then all the following vertices cannot connect to a previous vertex. More precisely, if there exists some j such that $V_j^v = \varnothing$, then $N_{\mathcal{G}}(V_j^0) = N_{\mathcal{G}}(\varnothing) = \varnothing$. It follows that $T_{j+1}^v = \varnothing$ and hence $V_{j+1}^v = \varnothing$. Repeating this observation shows that if $V_j^v = \varnothing$, then for all $\ell \geq j$ it holds that $V_\ell^v = \varnothing$. Since $V_i^v \neq \varnothing$, we therefore conclude that $V_j^v \neq \varnothing$ for all $j \in 0, i-1$. We continue by observing that if $V_i^v \neq \varnothing$, then it holds that $T_i^v \neq \varnothing$. From the fact that

$$V_i^v \in T_i^v \subseteq T_\geq\left(\mathcal{L}_i^v, M_\varepsilon^{dB^i/(\tau-1)}\right), \tag{6.17}$$

it immediately follows that $W_{V_i^v} \geq M_\varepsilon^{dB^i/(\tau-1)}$ by recalling Definition 6.6. These two properties described above are summarised in the following Claim 6.8.

Claim 6.8 (Properties of random vertices) *Consider Setting 6.3, in particular B. Let $M_\varepsilon > 1$, $v \in \mathcal{V}$ and consider $(V_i^v)_{i\geq 0}$ from Definition 6.7. If for $i \in \mathbb{N}\cup\{0\}$ it holds that $V_i^v \neq \varnothing$, then*

(1) for all $j < i$, $V_j^v \neq \varnothing$, and

(2) the weight of V_i^v satisfies $W_{V_i^v} \geq M_\varepsilon^{dB^i/(\tau-1)}$.

Proof. This follows immediately from Definition 6.7. ■

We note that if the construction of the random vertices given in Definition 6.7 succeeds, then this constitutes a weight-increasing path. Indeed, by item (2) of Claim 6.8, we see that the weight of the i'th random vertex typically grows with i. It remains to show that this construction succeeds with high probability. To this end, we firstly show that each step of the construction succeeds with high probability. More precisely, given that $V_{i-1}^v \neq \varnothing$, we bound the probability that also $V_i^v \neq \varnothing$. We also show that this bound goes to 1 both if $M_\varepsilon \to \infty$ and if $i \to \infty$. The latter suggests that the first few steps of the path are the 'hardest', in the sense that the success probability of these steps is much lower than those of the later steps.

Claim 6.9 *Consider Setting 6.3, in particular A and B. Let $M_\varepsilon > 1$, $v \in \mathcal{V}$, consider $(\mathcal{L}_i^v)_{i\geq 0}$ from Definition 6.4 and $(V_i^v)_{i\geq 0}$ and $(T_i^v)_{i\geq 0}$ from Definition 6.7. There exists a constant $C_{6.9} = C_{6.9}(\alpha, d) > 0$ and an $\underline{M}_{6.9} = \underline{M}_{6.9}(A,B,\alpha,d)$ such that if $M_\varepsilon \geq \underline{M}_{6.9}$, then for all $i \geq 1$*

$$\mathbb{P}\left(V_i^v \neq \varnothing \middle| V_{i-1}^v \neq \varnothing\right) \geq 1 - \exp\left[-C_{6.9} M_\varepsilon^{dB^i\alpha\left(\frac{B+\sigma}{B(\tau-1)} - A + \frac{A-1}{\alpha}\right)}\right]. \tag{6.18}$$

Furthermore, the event in (6.18) is independent of all vertices in $\mathcal{V}\setminus\bigcup_{j=0}^i \mathcal{L}_j^v$.

Proof. We aim to show that (6.18) holds when we condition on $V_{i-1}^v = v_{i-1}$, rather than $V_{i-1}^v \neq \varnothing$. That is, we show that for all $v_{i-1} \in \mathcal{L}_{i-1}^v$ then

$$\mathbb{P}\left(V_i^v \neq \varnothing \middle| V_{i-1}^v = v_{i-1}\right) \geq 1 - \exp\left[-C_{6.9} M_\varepsilon^{dB^i\alpha\left(\frac{B+\sigma}{B(\tau-1)} - A + \frac{A-1}{\alpha}\right)}\right]. \tag{6.19}$$

Since this bound is uniform over all v_{i-1}, it is valid for every realisation of V_i^v when conditioned on $V_i^v \neq \varnothing$. Equation (6.18) follows immediately.

Fix any $v_{i-1} \in \mathcal{L}_{i-1}^v$. We want to show (6.19). To do this, we find a bound for the complement of the event given in (6.19), i.e., we find a bound for $\mathbb{P}(V_i^v = \varnothing | V_{i-1}^v = v_{i-1})$. To this end, note that $V_i^v = \varnothing$ happens when $T_i^v = \varnothing$. By definition of T_i^v (see Definition 6.7) and the fact that $V_{i-1}^v = v_{i-1}$, this can only happen when each vertex of $\mathcal{L}_i^v$ is either not connected to v_{i-1} or does not have weight higher than

$M_\varepsilon^{dB^i/(\tau-1)}$. We conclude that

$$Q_{i-1}(v_{i-1}) := \mathbb{P}(V_i^v = \varnothing | V_{i-1}^v = v_{i-1}) = \mathbb{P}(\forall u \in \mathcal{L}_i^v: \ u \not\leftrightarrow v_{i-1} \text{ or } W_u < M_\varepsilon^{dB^i/(\tau-1)} | V_{i-1}^v = v_{i-1}). \tag{6.20}$$

We analyse $Q_{i-1}(v_{i-1})$. Firstly, note that $\mathcal{L}_i^v$ is finite, so we can enumerate all its vertices. Choose any labelling $\{u_j : j \in \{1, \ldots, |\mathcal{L}_i^v|\}\}$ of the vertices such that $\{u_j : j \in \{1, \ldots, |\mathcal{L}_i^v|\}\} = \mathcal{L}_i^v$ and the labelling is independent of the weight of the vertices[4]. Next, we define the following events for all $j \le |\mathcal{L}_i^v|$:

$$\mathcal{F}_{u_j, v_{i-1}} := \{V_{i-1}^v = v_{i-1}\} \cap \bigcap_{\ell=1}^{j-1} \left(\{u_\ell \not\leftrightarrow v_{i-1}\} \cap \left\{ W_{u_\ell} < M_\varepsilon^{dB^i/(\tau-1)} \right\} \right). \tag{6.21}$$

These sets give the information that $V_{i-1}^v = v_{i-1}$ and that all vertices before u_j are either not connected to v_{i-1} or do not have high enough weight. We consider for any $u_j \in \mathcal{L}_i^v$ the probability that $u_j \notin T_i^v$, i.e.,it is not either not connected to $V_{i-1}^v = v_{i-1}$ or its weight is not larger than $M_\varepsilon^{dB^i/(\tau-1)}$, given $\mathcal{F}_{u_j, v_{i-1}}$. Using this and telescopically applying the definition of conditional probability, we may write

$$\begin{aligned} Q_{i-1}(v_{i-1}) &= \mathbb{P}\bigg(\bigcap_{j=1}^{|\mathcal{L}_i^v|} \left(\{u_j \not\leftrightarrow v_{i-1}\} \cap \left\{ W_{u_j} < M_\varepsilon^{dB^i/(\tau-1)} \right\} \right) \ \Big| \ V_{i-1}^v = v_{i-1} \bigg) \\ &= \prod_{j=1}^{|\mathcal{L}_i^v|} \underbrace{\mathbb{P}(u_j \ne v_{i-1} \text{ or } W_{u_j} < M_\varepsilon^{dB^i/(\tau-1)} \mid \mathcal{F}_{u_j, v_{i-1}})}_{=: q_{i-1}(u_j, v_{i-1})}. \end{aligned} \tag{6.22}$$

Next, we analyse $q_{i-1}(u_j, v_{i-1})$. By rewriting and using conditional probability, we find that

$$q_{i-1}(u_j, v_{i-1}) = 1 - \mathbb{P}\left(u_j \leftrightarrow v_{i-1} \Big| W_{u_j} \ge M_\varepsilon^{dB^i/(\tau-1)}, \mathcal{F}_{u_j, v_{i-1}} \right) \mathbb{P}\left(W_{u_j} \ge M_\varepsilon^{dB^i/(\tau-1)} \right). \tag{6.23}$$

In the last probability, we have also used that $W_{u_j} \perp \mathcal{F}_{u_j, v_{i-1}}$, since W_{u_j} is independent of the weight of every other vertex and $\mathcal{F}_{u_j, v_{i-1}}$ only contains information about the weight of $(V_\ell^u)_{\ell \le i-1}$ and $(u_\ell)_{\ell \le j-1}$. Next, for any $s \ge 1, t \ge M_\varepsilon^{dB^{i-1}/(\tau-1)}$ we may verify that the assumptions of Lemma 4.1 hold:

(1) $\mathbb{P}(W_{u_j} \ge s, W_{v_{i-1}} \ge t | \mathcal{F}_{u_j, v_{i-1}}) > 0$

(2) $\mathbb{P}\left(u_j \leftrightarrow v_{i-1} | W_{u_j} = s, W_{v_{i-1}} = t, \mathcal{F}_{u_j, v_{i-1}} \right) = \mathbb{P}\left(u_j \leftrightarrow v_{i-1} | W_{u_j} = s, W_{v_{i-1}} = t \right).$

In the first inequality, we have used that the weight W_{u_j} of u_j is independent of $\mathcal{F}_{u_j, v_{i-1}}$ and $W_{v_{i-1}}$ and $W_{v_{i-1}} \ge t$ has positive probability under $\mathcal{F}_{u_j, v_{i-1}}$. In the second equality, we have used that edges are present independently of all other edges, given a realisation of the weights. We are therefore justified in applying Lemma 4.1. By doing so, from (6.23), the weight distribution of W_{u_j} given by (3.4) and the fact that $\mathcal{F}_{u_j, v_{i-1}}$ implies that $V_{i=1}^v = v_{i-1}$ which implies that $W_{v_{i-1}} \ge M_\varepsilon^{dB^{i-1}}$, we find that

$$q_{i-1}(u_j, v_{i-1}) \le 1 - \underline{c}\rho \left(|u_j - v_{i-1}|, M_\varepsilon^{dB^i/(\tau-1)}, M_\varepsilon^{dB^{i-1}/(\tau-1)} \right) M_\varepsilon^{-(\tau-1)dB^i/(\tau-1)}. \tag{6.24}$$

Then, notice that since $M_\varepsilon > 1, d \in \mathbb{N}, \tau - 1 > 1$ and $B > 1$ it holds that $M_\varepsilon^{dB^i/(\tau-1)} > M_\varepsilon^{dB^{i-1}/(\tau-1)}$. Furthermore, from item (1) of Claim 6.5, we find that $|u_j - v_{i-1}| \le \sqrt{d} M_\varepsilon^{dAB^i}$. Now by using the definition of ρ from (3.6) and using some elementary computations, from (6.24) we obtain that

$$\begin{aligned} q_{i-1}(u_j, v_{i-1}) &\le 1 - \underline{c}\bigg[1 \wedge \bigg(\frac{M_\varepsilon^{dB^i/(\tau-1)} M_\varepsilon^{\sigma dB^{i-1}/(\tau-1)}}{(\sqrt{d} M_\varepsilon^{AB^i})^d} \bigg) \bigg]^\alpha M_\varepsilon^{-dB^i} \\ &\le 1 - C_1 M_\varepsilon^{-dB^i - 0 \wedge d\alpha B^i \left(\frac{B+\sigma}{B(\tau-1)} - A \right)}. \end{aligned} \tag{6.25}$$

Here $C_1 = C_1(\alpha, d) = \underline{c}/d^{d\alpha/2} > 0$ is constant with respect to M_ε and i and we have used that $1 \wedge x^p = x^{0 \wedge p}$ if $x > 1$. Next, we analyse the exponent of M_ε in (6.25). By rewriting, we may show that

$$\frac{B+\sigma}{B(\tau-1)} - A < 0 \qquad \Leftrightarrow \qquad B > \frac{\sigma}{A(\tau-1)+1}. \tag{6.26}$$

The latter inequality is true by definition (see Setting 6.3), so the former is too. By going back to (6.25), we conclude that

$$q_{i-1}(u_j, v_{i-1}) \le 1 - C_1 M_\varepsilon^{dB^i (\alpha \frac{B+\sigma}{B(\tau-1)} - A\alpha - 1)}. \tag{6.27}$$

Continuing, by item (2) of Claim 6.5, there exists a $\underline{M}_{6.9} = \underline{M}_{6.9}(A, B, d) = \underline{M}_{6.5}(A, B, d)$ such that if $M_\varepsilon \ge \underline{M}_{6.5}$, then $|\mathcal{L}_i^v| \ge M_\varepsilon^{dB^i}/2$ for all $i \in \mathbb{N} \cup \{0\}$. If we then assume that $M_\varepsilon \ge \underline{M}_1$, by substituting

[4]For example, choose the lexicographical ordering of spatial position of the vertices.

(6.27) into (6.22) and using that $1-x\le e^{-x}$, we obtain

$$\begin{aligned}Q_{i-1}(v_{i-1}) &= \prod_{j=1}^{|\mathcal{L}_i^v|} q_{i-1}(u_j, v_{i-1})\\ &\le \Big(1 - C_1 M_\varepsilon^{-dB^i(\alpha\frac{B+\sigma}{B(\tau-1)}-A\alpha-1)}\Big)^{\frac{1}{2}M_\varepsilon^{dAB^i}}\\ &\le \exp\Big[-C_2 M_\varepsilon^{dB^i(\alpha\frac{B+\sigma}{B(\tau-1)}-\alpha A-1+A)}\Big]. \end{aligned}\tag{6.28}$$

Here $C_2 = C_2(\alpha,d) = C_1(\alpha,d)/2 > 0$ is constant with respect to i and M_ε. By recalling the definition of $Q_{i-1}(v_{i-1})$ from (6.20) and setting $C_{6.9} = C_{6.9}(\alpha,d) = C_2(\alpha,d)$, we thus conclude that (6.19) holds if $M_\varepsilon \ge \underline{M}_{6.9}$. As already reasoned at the start of the proof, this also shows that (6.18) holds. Lastly, note that in this proof we have only used weights of vertices that are in $\bigcup_{j=0}^{i}\mathcal{L}_j^v$. Since the weights of all vertices are independent, we conclude that the event in (6.18) is independent of all vertices in $\mathcal{V}\setminus\bigcup_{j=0}^{i}\mathcal{L}_j^v$. This finishes the proof.

■

We recall from Definition 3.1 that an edge between two vertices depends solely on the weight of those two vertices. The fact that the result in Claim 6.9 is independent of all vertices in $\mathcal{V}\setminus\bigcup_{j=0}^{i}\mathcal{L}_j^v$ therefore also means that the result is independent of all edges between vertices that are not in $\bigcup_{j=0}^{i}\mathcal{L}_j^v$.

Continuing, since we later on intend to take M_ε a large constant, the result (6.18) of Claim 6.9 is only useful when

$$R := \alpha\bigg(\frac{B+\sigma}{B(\tau-1)} - A + \frac{A-1}{\alpha}\bigg) > 0.\tag{6.29}$$

We therefore verify that (6.29) is correct. To do this, we note that by rewriting we can see that (6.29) is equivalent to

$$A < \frac{\alpha\frac{1+\sigma}{\tau-1}-1}{\alpha-1}.\tag{6.30}$$

We have already seen that this is true in equation (6.6). Therefore, the right-hand side of (6.18) indeed goes to 1 if $M_\varepsilon\to\infty$.

From the above Claim 6.9 we see that we can find edges from one layer to the next using vertices with typically increasing weight. Using this claim repeatedly, we can string all these edges together to construct a path from the layer of order 0 to a layer of any order. In the following Corollary 6.10, we compute a lower bound for the probability that this construction succeeds. Crucially, the bound we compute goes to 1 if $M_\varepsilon\to\infty$ and does *not* depend on the length of the path. The former is needed when we eventually need to construct a path with high probability; the latter is useful when we want to find increasingly longer paths.

Corollary 6.10 *Consider the same setting as Lemma 6.9 and fix any $k\in\mathbb{N}$. Then there exists a constant $C_{6.10} > 0$ and an $\underline{M}_{6.10} = \underline{M}_{6.10}(A,B,\alpha,d)$ such that if $M_\varepsilon \ge \underline{M}_{6.10}$, then*

$$\mathbb{P}(V_k^v \neq \varnothing) \ge 1 - 2\exp\Big[-\frac{1}{2}M_\varepsilon^{d(A-1)}\Big] - 4\exp\Big[-C_{6.10}M_\varepsilon^{d\alpha B\left(\frac{A-1}{\alpha}+\frac{B+\sigma}{B(\tau-1)}-1\right)}\Big].\tag{6.31}$$

Furthermore, the event $\{V_k^v\neq\varnothing\}$ is independent of the weight of all vertices in $\mathcal{V}\setminus\bigcup_{i=0}^{k}\mathcal{L}_i^v$.

Proof. Before we start with the proof, we firstly give the definition of $\underline{M}_{6.10}$. Firstly set $\underline{M}_1 = \underline{M}_1(A,B,d) = \underline{M}_{6.5}(A,B,d)$ and note that if $M_\varepsilon\ge\underline{M}_1$, then by item (2) of Claim 6.5, for all $i\in\mathbb{N}\cup\{0\}$ it holds that $|\mathcal{L}_i^v| \ge M_\varepsilon^{dAB^i}/2$. Furthermore, let $\underline{M}_2 = \underline{M}_2(A,B,\alpha,d) = \underline{M}_{6.9}(A,B,\alpha,d)$ and note that if $M_\varepsilon\ge\underline{M}_2$, then by Claim 6.9 for all $i\ge 1$ equation 6.18 holds. Next, define R as in (6.29), i.e.,

$$R = R(A,B,\alpha,\tau) = \alpha\bigg(\frac{B+\sigma}{B(\tau-1)} - A + \frac{A-1}{\alpha}\bigg).\tag{6.32}$$

Then by (6.29) we know that $R>0$. Furthermore, let $C_{6.9} = C_{6.9}(\alpha,d) > 0$ be the constant from Claim 6.9. Then, because $d\in\mathbb{N}$, $A>1$ and $B>1$, it is possible to find an $\underline{M}_3 = \underline{M}_3(A,B,\alpha,\tau,d)$ such that if $M_\varepsilon\ge\underline{M}_3$, then

$$\exp\big[-M_\varepsilon^{d(A-1)}\big] \le \frac{1}{2} \qquad\text{and}\qquad \exp\big[-C_{6.9}M_\varepsilon^{dRB}\big]\le\frac{1}{2}.\tag{6.33}$$

Specifically, we set

$$\underline{M}_3 = \ln(2)^{\frac{1}{d(A-1)}} \vee \Big(\frac{\ln 2}{C_{6.9}}\Big)^{\frac{1}{dRB}}.\tag{6.34}$$

Lastly, it is quickly verified that there exists an $\underline{M}_4 = \underline{M}_4(A,B,\alpha,\tau,d)$ such that if $M_\varepsilon\ge\underline{M}_4$, then for all $i\ge 0$ it holds that $iM_\varepsilon^{dRB}\le M_\varepsilon^{dRB^i}$. We now set $\underline{M}_{6.10} = \underline{M}_{6.10}(A,B,\alpha,\tau,d) = \underline{M}_1(A,B,d)\vee \underline{M}_2(A,B,\alpha,d)\vee\underline{M}_3(A,B,\alpha,\tau,d)\vee\underline{M}_4(A,B,\alpha,\tau,d)$. For the remainder of this proof, we assume that $M_\varepsilon\ge\underline{M}_{6.10}$.

We continue by showing (6.31). By the law of total probability, we observe that for any $i \geq 1$:

$$\begin{aligned}\mathbb{P}(V_i^v \neq \varnothing) &= \mathbb{P}\left(V_i^v \neq \varnothing \middle| V_{i-1}^v \neq \varnothing\right)\mathbb{P}(V_{i-1}^v \neq \varnothing) + \underbrace{\mathbb{P}\left(V_i^v \neq \varnothing \middle| V_{i-1}^v = \varnothing\right)}_{=0}\mathbb{P}(V_{i-1}^v = \varnothing) \\ &= \mathbb{P}\left(V_i^v \neq \varnothing \middle| V_{i-1}^v \neq \varnothing\right)\mathbb{P}(V_{i-1}^v \neq \varnothing). \end{aligned} \tag{6.35}$$

Here we have used that by the first item of Claim 6.8, it cannot happen that $V_{i-1}^v = \varnothing$ and $V_i^v \neq \varnothing$. Applying (6.35) k times, we obtain

$$\mathbb{P}(V_k^v \neq \varnothing) = \mathbb{P}(V_0^v \neq \varnothing)\prod_{i=1}^{k}\mathbb{P}\left(V_i^v \neq \varnothing \middle| V_{i-1}^v \neq \varnothing\right). \tag{6.36}$$

We first consider $\mathbb{P}(V_0^v \neq \varnothing)$. To this end, note that $\{V_0^v \neq \varnothing\}$ is only possible if $T_0^v \neq \varnothing$. By recalling the definition of T_0^v from Definition 6.7, we observe that this is possible if there is at least one vertex with weight higher than $M_\varepsilon^{d/(\tau-1)}$ in $\mathcal{L}_0^v$. Equivalently, $T_0^v \neq \varnothing$ if $\max_{x\in\mathcal{L}_0^v} W_x \geq M_\varepsilon^{d/(\tau-1)}$. Taking everything together, by additionally applying Claim A.2 and using that $M_\varepsilon \geq \underline{M}_{6.10}$ implies that $|\mathcal{L}_0^v| \geq M_\varepsilon^{dA}/2$, we find that

$$\mathbb{P}(V_0^v \neq \varnothing) = \mathbb{P}\Big(\max_{x\in\mathcal{L}_0^v} W_x \geq M_\varepsilon^{d/(\tau-1)}\Big) = 1 - \exp\big[-|\mathcal{L}_0^v| M_\varepsilon^{-d}\big] \leq 1 - \exp\Big[-\frac{1}{2}M_\varepsilon^{d(A-1)}\Big]. \tag{6.37}$$

Next, because $M_\varepsilon \geq \underline{M}_{6.10}$ and by Lemma 6.9, we know that (6.18) holds for every $\mathbb{P}(V_i^v \neq \varnothing | V_{i-1}^v \neq \varnothing)$. By returning to (6.36) we obtain

$$\mathbb{P}(V_k^v \neq \varnothing) \geq \left(1 - \exp\left[-\frac{1}{2}M_\varepsilon^{d(A-1)}\right]\right)\prod_{i=1}^{k}\left(1 - \exp\left[-C_{6.9}M_\varepsilon^{dRB^i}\right]\right), \tag{6.38}$$

where R is as in (6.32). Next, note that because $M_\varepsilon \geq \underline{M}_{6.10}$, we know that $\exp\big[-M_\varepsilon^{d(A-1)}/2\big] \leq 1/2$ and $\exp\big[-C_{6.9}M_\varepsilon^{dRB}\big] \leq 1/2$. Furthermore, since $B > 1$ from the last inequality we also obtain that for all $i \geq 2$ it holds that $\exp\big[-C_{6.9}M_\varepsilon^{dRB^i}\big] \leq 1/2$. Next, we apply the inequality $1 - x \geq e^{-2x}$, which is valid when $x \in [0, 1/2]$ [5]. By doing so to (6.38) we obtain

$$\begin{aligned}\mathbb{P}(V_k^v \neq \varnothing) &\geq \exp\left[-2\left(\exp\left[-\frac{1}{2}M_\varepsilon^{d(A-1)}\right] + \sum_{i=1}^{k}\exp\left[-C_{6.9}M_\varepsilon^{dRB^i}\right]\right)\right] \\ &\geq \exp\left[-2\left(\exp\left[-\frac{1}{2}M_\varepsilon^{d(A-1)}\right] + \sum_{i=1}^{\infty}\exp\left[-C_{6.9}M_\varepsilon^{dRB^i}\right]\right)\right]. \end{aligned} \tag{6.39}$$

In the second inequality we take the infinite sum rather than the partial sum up to k, as this yields a uniform bound over all k. Next, we examine the sum in (6.39). To this end, note that because $M_\varepsilon \geq \underline{M}_{6.10}$, for all $i \geq 0$ it holds that $iM_\varepsilon^{dRB} \leq M_\varepsilon^{dRB^i}$. By applying this and using that $\exp\big[-C_{6.9}M_\varepsilon^{dRB}\big] \leq 1/2$, we find that

$$\sum_{i=1}^{\infty}\exp\left[-C_{6.9}M_\varepsilon^{dRB^i}\right] \leq \sum_{i=1}^{\infty}\exp\left[-C_{6.9}M_\varepsilon^{dRB}\right]^i = \frac{\exp\left[-C_{6.9}M_\varepsilon^{dRB}\right]}{1 - \exp\left[-C_{6.9}M_\varepsilon^{dRB}\right]} \leq 2\exp\left[-C_{6.9}M_\varepsilon^{dRB}\right]. \tag{6.40}$$

By substituting (6.40) into (6.39) and using that $e^{-x} \geq 1 - x$, we obtain

$$\mathbb{P}(V_k^v \neq \varnothing) \geq 1 - 2\exp\left[-\frac{1}{2}M_\varepsilon^{d(A-1)}\right] - 4\exp\left[-C_{6.9}M_\varepsilon^{dRB}\right], \tag{6.41}$$

which is valid if $M_\varepsilon \geq \underline{M}_{6.10}$. By recalling the definition of R from (6.32) and setting $C_{6.10} = C_{6.10}(\alpha, d) = C_{6.9}(\alpha, d) > 0$ we see that this shows (6.31). Lastly, we note that the event $\{V_k^v \neq \varnothing\}$ only depends on the weight of vertices that are in $\bigcup_{i=0}^{k}\mathcal{L}_i^v$. Since the weights of all vertices are independent, we find that $\{V_k^v \neq \varnothing\}$ is independent of the weight of all vertices in $\mathcal{V}\setminus\bigcup_{i=0}^{k}\mathcal{L}_i^v$. This finishes the proof. ■

We note that with Corollary 6.10, we have shown that the construction of Definition 6.7 succeeds with high probability. Indeed, if $V_k^u \neq \varnothing$ with high probability, then by item (1) of Claim 6.8 we see that also the path $V_0^u V_1^u \dots V_k^u$ exists with high probability. As already announced, this result holds regardless of the value of k. We may therefore build the path from step (1) of Figure 5 up to any length. We choose a specific length $k_\star$ for this path. This $k_\star$ is defined in the following Definition 6.11.

Definition 6.11 *Consider Setting 6.3, in particular A and B. Let $u, v \in \mathcal{V}$ and set $N := |u - v|$. We define $k_\star = k_\star(N, M_\varepsilon, A, B, d)$ as the largest integer satisfying*

$$M_\varepsilon^{AB^{k_\star}} \leq \frac{1}{2\sqrt{d}}N. \tag{6.42}$$

When N, M_ε, A, B and d are clear from context, we suppress their dependence and write $k_\star$.

[5] This inequality is in fact valid for more values that just $x \in [0, 1/2]$. The lower-bound of 0 is valid, but the exact upper end-point for x is given by $1 + W(-2/e^2)/2 \approx 0.7968$. Here W is the principal branch of the Lambert W function, which is the inverse function of $y \mapsto ye^y$. Choosing $1/2$ as an upper bound is more convenient and does not significantly impact the proof.

Before we continue, we analyse Definition 6.11. Firstly, by taking logarithms twice in (6.42), we see that $k_\star$ needs to satisfy

$$\ln A + k_\star \ln B + \ln\ln M_\varepsilon \le \ln(\ln N - \ln(2\sqrt{d})). \tag{6.43}$$

Next, we notice that $B>1$ and that since $d\ge 1$, it holds that $\ln(\ln N - \ln(2\sqrt{d})) \le \ln\ln N$. Substituting this into (6.43) and rewriting yields

$$k_\star \le \frac{\ln\ln N - \ln A - \ln\ln M_\varepsilon}{\ln(B)}. \tag{6.44}$$

We indeed see that as announced, $k_\star = \mathcal{O}(\ln\ln(N))$. Next, notice that because $\mathcal{L}^u_{k_\star}$ and $\mathcal{L}^v_{k_\star}$ have side-lengths $M_\varepsilon^{AB^{k_\star}} \le N/(2\sqrt{d})$ and their centres u and v are N apart, by item (2) of Lemma A.1 we find that $\mathcal{L}^u_{k_\star} \cap \mathcal{L}^v_{k_\star} = \varnothing$. Since the layers are increasing, it quickly follows that

$$\left(\bigcup_{i=0}^{k_\star} \mathcal{L}^u_i\right) \cap \left(\bigcup_{i=0}^{k_\star} \mathcal{L}^v_i\right) = \varnothing. \tag{6.45}$$

From Corollary 6.10, we then also obtain that the events $\{V^u_{k_\star} \ne \varnothing\}$ and $\{V^v_{k_\star} \ne \varnothing\}$ are independent. This latter observation is exactly why we require that $k_\star$ satisfies (6.42); if $k_\star$ were (much) larger, then we cannot ensure that $\mathcal{L}^u_{k_\star} \cap \mathcal{L}^v_{k_\star} \ne \varnothing$ and hence that $\{V^u_{k_\star} \ne \varnothing\}$ and $\{V^v_{k_\star} \ne \varnothing\}$ are independent. There is, however, also reason to require $k_\star$ to be as large as possible. This becomes evident when we want to connect $V^u_{k_\star}$ to $V^v_{k_\star}$. If $k_\star$ were (much) smaller, then these vertices typically would have smaller weight. It would therefore be more difficult to connect them, i.e., the probability that these vertices would be connected in a constant number of edges would decrease. We avoid this decrease in probability by choosing $k_\star$ as large as possible, and noting that this yields no disadvantages elsewhere as the result of Corollary 6.10 is uniform over all k.

It remains to actually compute the probability that $V^u_{k_\star}$ and $V^v_{k_\star}$ are connected in a constant number of edges. More precisely, we show that with high probability, $d_{\mathcal{G}}(V^u_{k_\star}, V^v_{k_\star}) \le 4$. This result can be found in Corollary 6.17. However, to show this result we require an additional structure that we have not yet defined. In particular, we define a large set $\mathcal{D}_{u,v}$ of with sidelengths $\Theta\big(M_\varepsilon^{AB^{k_\star+1}}\big)$ centred around the midpoint of u and v. We do this in such a way that $\bigcup_{i=0}^{k_\star} \mathcal{L}^u_i$ and $\bigcup_{i=0}^{k_\star} \mathcal{L}^u_i$ are fully encompassed by $\mathcal{D}_{u,v}$. We then consider all vertices in $\mathcal{D}_{u,v}$ that have weight higher than $M_\varepsilon^{dB^{k_\star+1}/(\tau-1)}$, which we call $\mathcal{T}$. Then, we show that both $V^u_{k_\star}$ and $V_{k_\star}$ are connected to $\mathcal{T}$ and that the distance between two vertices in $\mathcal{T}$ is typically constant.

In the following Definition 6.12, we give the definition of $\mathcal{D}_{u,v}$ as described above. The reader may think of this set as the 'last layer' or an 'encompassing set'.

Definition 6.12 *Consider Setting 6.3, in particular A and B. Let $(\mathcal{L}^v_i)_i$ be as in Definition 6.4. Let $u,v \in \mathcal{V} = \mathbb{Z}^d$ be distinct vertices, set $N = |u-v|$ and let $k_\star = k_\star(N, M_\varepsilon, A, B, d)$ be as in Definition 6.11. We define*

$$\mathcal{D}_{u,v} = \mathcal{D}_{u,v}(N, M_\varepsilon, A, B, d) = \left(\frac{u+v}{2} + \left[-2\sqrt{d}M_\varepsilon^{AB^{k_\star+1}}, 2\sqrt{d}M_\varepsilon^{AB^{k_\star+1}}\right]^d\right) \cap \mathcal{V}\setminus \bigcup_{i=0}^{k_\star}(\mathcal{L}^u_i \cup \mathcal{L}^v_i). \tag{6.46}$$

When N, M_ε, A, B and d are clear from context, we write $\mathcal{D}_{u,v}$.

We immediately continue by giving the properties of $\mathcal{D}_{u,v}$ we use.

Claim 6.13 *Consider Setting 6.3, in particular A and B. Fix $u,v\in\mathcal{V}$, set $N=|u-v|$, let $M_\varepsilon>1$, consider $k_\star = k_\star(N, M_\varepsilon, A, B, d)$ from Definition 6.11 and consider $\mathcal{D}_{u,v} = \mathcal{D}_{u,v}(N, M_\varepsilon, A, B, d)$ from Definition 6.12. Then the following statements holds:*

(1) For every $M_\varepsilon > 1$ there exists an $\underline{N}_{6.13} = \underline{N}_{6.13}(M_\varepsilon, A, B, d)$ such that if $N \ge \underline{N}_{6.13}$, then

$$|\mathcal{D}_{u,v}| \ge (2\sqrt{d})^d M_\varepsilon^{dAB^{k_\star+1}}. \tag{6.47}$$

(2) For any $x,y \in \mathcal{D}_{u,v} \cup \bigcup_{i=0}^{k_\star}(\mathcal{L}^u_i \cup \mathcal{L}^v_i)$ it holds that

$$|x-y| \le 4dM_\varepsilon^{AB^{k_\star+1}}. \tag{6.48}$$

Proof. Item (2) follows immediately from first item of Lemma A.1. For item (1), we imitate the proof of Lemma 6.5. Because of the second item of Lemma A.1, we note that $\bigcup_{i=0}^{k_\star} \mathcal{L}^u_i$ is disjoint from $\bigcup_{i=0}^{k_\star} \mathcal{L}^v_i$.

Furthermore, we may compute (similarly to the proof of Lemma 6.5):

$$\begin{aligned}|\mathcal{D}_{u,v}| &= 2^d\Big\lfloor 2\sqrt{d}M_\varepsilon^{AB^{k_\star+1}}+1\Big\rfloor^d - 2\cdot 2^d\Big\lfloor \frac{1}{2}M_\varepsilon^{AB^{k_\star}}+1\Big\rfloor^d \\ &\geq \Big((4\sqrt{d})^d - 2^{d+1}\Big(M_\varepsilon^{-AB^{k^\star}(B-1)} + M_\varepsilon^{-AB^{k_\star+1}}\Big)\Big)M_\varepsilon^{dAB^{k_\star+1}} \\ &=: g(M_\varepsilon,k_\star)M_\varepsilon^{dAB^{k_\star+1}}. \end{aligned} \tag{6.49}$$

To show (6.47), it remains to show that $g(M_\varepsilon,k_\star)\geq(2\sqrt{d})^d$. Fix any $M_\varepsilon>1$. It is quickly verified that if N increases, then $k_\star$ is increasing and therefore $g(M_\varepsilon,k_\star)$ is also increasing. More precisely, if $N\uparrow\infty$ then $k_\star\uparrow\infty$ and if $k_\star\uparrow\infty$, then $g(M_\varepsilon,k_\star)\uparrow(4\sqrt{d})^d$. Therefore, for any fixed M_ε there indeed is such a $\underline{N}_{6.13}=\underline{N}_{6.13}(M_\varepsilon,A,B,d)$ such that if $N\geq\underline{N}_{6.13}$ then $g(M_\varepsilon,k_\star)\geq(2\sqrt{d})^d$. Furthermore, if $N\geq\underline{N}_{6.13}$ then also $|\mathcal{D}_{u,v}|\geq(2\sqrt{d})^dM_\varepsilon^{dAB^{k_\star+1}}$. This finishes the proof of item (1). ■

To increase legibility in later results, we introduce more notation. In particular, we give a name to the event that there are at least $\Theta\big(M_\varepsilon^{d(A-1)B^{k_\star+1}}\big)$ vertices that have weight $M_\varepsilon^{dB^{k_\star+1}/(\tau-1)}$ or greater within $\mathcal{D}_{u,v}$.

Definition 6.14 *Consider Setting 6.3 in particular A and B. Let $u,v\in\mathcal{V}=\mathbb{Z}^d$ be distinct, set $N=|u-v|$ and let $M_\varepsilon>1$. Consider $T_\geq$ from Definition 6.6, $k_\star=k_\star(N,M_\varepsilon,A,B,d)$ from Definition 6.11, $T_\geq$ from Definition 6.6 and $\mathcal{D}_{u,v}=\mathcal{D}_{u,v}(N,M_\varepsilon,A,B,d)$ from Definition 6.12. Denote by $\mathbf{W}_{u,v}=\mathbf{W}_{u,v}(N,M_\varepsilon,A,B,d)$ the (random) vector of weights of the vertices within $D_{u,v}$, i.e.,*

$$\mathbf{W}_{u,v}=(W_x)_{x\in\mathcal{D}_{u,v}}. \tag{6.50}$$

Then, set

$$\mathcal{T}=\mathcal{T}(\mathbf{W}_{u,v},N,M_\varepsilon,A,B,d)=T_\geq\Big(\mathcal{D}_{u,v},M_\varepsilon^{dB^{k_\star+1}/(\tau-1)}\Big). \tag{6.51}$$

If a realisation $\mathbf{w}=(w_x)_{x\in\mathcal{D}_{u,v}}$ of $\mathbf{W}_{u,v}$ is such that

$$|\mathcal{T}(\mathbf{w},N,M_\varepsilon,A,B,d)|\geq(\sqrt{d})^dM_\varepsilon^{d(A-1)B^{k_\star+1}}, \tag{6.52}$$

*we call this realisation (N,M_ε,A,B,d)-**good**. Furthermore, we define the following event*

$$\{\mathbf{W}_{u,v}\text{ is }(N,M_\varepsilon,A,B,d)\text{-good}\}:=\{\mathbf{W}_{u,v}\in\{\mathbf{w}:\mathbf{w}\text{ is }(N,M_\varepsilon,A,B,d)\text{-good}\}\} \tag{6.53}$$

When N,M_ε,A,B and d are clear from context, we write good rather than (N,M_ε,A,B,d)-good and we denote the event in (6.53) with $\{\mathbf{W}_{u,v}$ is good$\}$.

Next, using the notation from Definition 6.14, we compute the probability that a realisation of $\mathbf{W}_{u,v}$ is good. In particular, we show that this happens with high probability as $M_\varepsilon\to\infty$. This is be useful in later proofs, as this means that we may then condition on the fact that there are ‘enough’ (i.e., more than $(\sqrt{d})^dM_\varepsilon^{d(A-1)B^{k_\star+1}}$) vertices with weight higher that $M_\varepsilon^{dB^{k_\star+1}/(\tau-1)}$.

Claim 6.15 *Consider Setting 6.3, in particular A and B. Let $u,v\in\mathcal{V}=\mathbb{Z}^d$ be distinct, set $N=|u-v|$, fix any $M_\varepsilon>1$, consider $k_\star=k_\star(N,M_\varepsilon,A,B,d)$ from Definition 6.11 and let (N,M_ε,A,B,d)-good be as given in Definition 6.14. There is a $C_{6.15}=C_{6.15}(d)>0$ and an $\underline{N}_{6.15}=\underline{N}_{6.15}(M_\varepsilon,A,B,d)$ such that if $N\geq\underline{N}_{6.15}$, then*

$$\mathbb{P}\left(\mathbf{W}_{u,v}\text{ is }(N,M_\varepsilon,A,B,d)\text{-good}\right)\geq 1-\exp\Big(-C_{6.15}M_\varepsilon^{d(A-1)B^{k_\star+1}}\Big).$$

Furthermore, the event from (6.15) is independent of all vertices in $\bigcup_{i=0}^{k_\star}(\mathcal{L}_i^u\cup\mathcal{L}_i^v)$, where $(\mathcal{L}_i^u)_{i\geq0}$ is as given in Definition 6.4.

Proof. Throughout this proof, we suppress all dependencies unless necessary. Note that $\mathbf{W}_{u,v}$ is good if $|T|\geq(\sqrt{d})^dM_\varepsilon^{d(A-1)B^{k_\star+1}}$, where $\mathcal{T}$ is as defined in (6.51). We compute the probability that this inequality holds. To do this, note that there are $n:=|\mathcal{D}_{u,v}|$ vertices in $\mathcal{D}_{u,v}$. Furthermore, by (3.4) of Assumption 3.2 we find that each of these vertices independently have probability p that their weight is larger than $M_\varepsilon^{dB^{k_\star+1}}/(\tau-1)$, where p is given by:

$$p:=\mathbb{P}\Big(W\geq M_\varepsilon^{dB^{k_\star+1}/(\tau-1)}\Big)=M_\varepsilon^{-dB^{k_\star+1}}. \tag{6.54}$$

We conclude that

$$X:=|\mathcal{T}|=\Big|T_\geq\Big(\mathcal{D}_{u,v},M_\varepsilon^{dB^{k_\star+1}/(\tau-1)}\Big)\Big|\sim\mathrm{Bin}(n,p), \tag{6.55}$$

We remark that

$$\mathbf{W}_{u,v}\text{ is good}\qquad\Leftrightarrow\qquad X\geq(\sqrt{d})^dM_\varepsilon^{d(A-1)B^{k_\star+1}} \tag{6.56}$$

by Definition 6.14. Next, we set $\underline{N}_{6.15}=\underline{N}_{6.15}(M_\varepsilon,A,B,d)=\underline{N}_{6.13}(M_\varepsilon,A,B,d)$ and note that if $N\geq\underline{N}_{6.15}$, then by item (1) of Claim 6.13, it holds that $n=|D_{u,v}|\geq(2\sqrt{d})^dM_\varepsilon^{dAB^{k_\star+1}}$. From this, we may conclude that if $N \geq \underline{N}_{6.15}$, then X stochastically dominates another binomially distributed random variable $\tilde{X}$:

$$X \overset{d}{\geq} \widetilde{X} \sim \mathrm{Bin}\left((2\sqrt{d})^d M_\varepsilon^{dAB^{k_\star+1}}, M_\varepsilon^{-dB^{k_\star+1}}\right). \tag{6.57}$$

By using (6.57), the Chernoff bound for binomial random variables (see Lemma B.4) and rewriting to match the format of the Chernoff bound, we observe that

$$\begin{aligned}\mathbb{P}\left(X \geq \sqrt{d}^d M_\varepsilon^{dB^{k_\star+1}(A-1)}\right) &\geq \mathbb{P}\left(\tilde{X} \geq \sqrt{d}^d M_\varepsilon^{dB^{k_\star+1}(A-1)}\right)\\ &= 1 - \mathbb{P}\left(\widetilde{X} < (1-(1-2^{-d}))(2\sqrt{d})^d M_\varepsilon^{dAB^{k_\star+1}} M_\varepsilon^{dB^{k_\star+1}}\right)\\ &\geq 1 - \exp\Big[-\underbrace{\frac{(2\sqrt{d})^d(1-2^{-d})}{2}}_{=:C_{6.15}(d)=C_{6.15}} M_\varepsilon^{d(A-1)B^{k_\star+1}}\Big].\end{aligned} \tag{6.58}$$

Recalling (6.56) and defining $C_{6.15}$ as in (6.58) finishes the proof. ∎

In the following Claim 6.16, we show that both $V_{k_\star}^u$ and $V_{k_\star}^v$ are connected to $\mathcal{T}$ (see (6.51), i.e., the set of all vertices of $\mathcal{D}_{u,v}$ that have weight higher that $M_\varepsilon^{dB^{k_\star+1}/(\tau-1)}$.

Claim 6.16 *Consider Setting 6.3, in particular A and B. Let $u, v \in \mathcal{V} = \mathbb{Z}^d$, set $N = |u-v|$, let $M_\varepsilon > 1$ and let $k_\star = k_\star(N, M_\varepsilon, A, B, d)$ be as in Definition 6.11. Furthermore, let $T_\geq$ be as given in Definition 6.6, let $(V_i^u)_{i\leq k_\star}$ and $(V_i^v)_{i\leq k_\star}$ be as given in Definition 6.7 and $\mathcal{D}_{u,v} = \mathcal{D}_{u,v}(N, M_\varepsilon, A, B, d)$ as given in Definition 6.12. Let $\mathbf{W}_{u,v} = \mathbf{W}_{u,v}(N, M_\varepsilon, A, B, d)$, $\mathcal{T}$ and $(N, M_\varepsilon, A, B, d)$-good be as defined in 6.14. Set*

$$R := \alpha\left(\frac{B+\sigma}{B(\tau-1)} - A + \frac{A-1}{\alpha}\right) > 0. \tag{6.59}$$

There is a $C_{6.16} = C_{6.16}(\alpha, d)$ such that

$$\mathbb{P}\big(V_{k_\star}^u \leftrightarrow \mathcal{T}, V_{k_\star}^v \leftrightarrow \mathcal{T} \big|\ \mathbf{W}_{u,v} \text{ is good}, V_{k_\star}^u, V_{k_\star}^v \neq \varnothing\big) \geq 1 - 2\exp\Big[-C_{6.16} M_\varepsilon^{dRB^{k_\star+1}}\Big]. \tag{6.60}$$

Proof. Throughout this proof, we suppress any dependence on N, M_ε, A, B or d when these quantities are clear. We examine the complement of the event in (6.60), i.e.,

$$Q = \mathbb{P}\big(\{V_{k_\star}^u \not\leftrightarrow \mathcal{T}\} \cup \{V_{k_\star}^v \not\leftrightarrow \mathcal{T}\} \mid \mathbf{W}_{u,v} \text{ is good}, V_{k_\star}^u, V_{k_\star}^v \neq \varnothing\big) \tag{6.61}$$

Now by applying the union bound, we find that

$$Q \leq \underbrace{\mathbb{P}\big(V_{k_\star}^u \not\leftrightarrow \mathcal{T} \big|\ \mathbf{W}_{u,v} \text{ is good}, V_{k_\star}^u \neq \varnothing\big)}_{=:Q_u} + \underbrace{\mathbb{P}\big(V_{k_\star}^v \not\leftrightarrow \mathcal{T} \big|\ \mathbf{W}_{u,v} \text{ is good}, V_{k_\star}^v \neq \varnothing\big)}_{=:Q_v}. \tag{6.62}$$

Here have have dropped $V_{k_\star}^v \neq \varnothing$ from the conditioning in Q_u since $\{V_{k_\star}^u \leftrightarrow \mathcal{T}\}$ is independent of $V_{k_\star}^v$, also when conditioned on $V_{k_\star}^u \neq \varnothing$ and $\mathbf{W}_{u,v}$ is good. We do the same with Q_v but with the roles of u and v reversed. Notice that Q_u is not necessarily equal to Q_v, since the connection probability may depend on the locations of u and v. Regardless, we find the same lower-bound for Q_u and Q_v. We do not find this lower-bound when conditioned on $\mathbf{W}_{u,v}$ is good and $V_{k_\star}^x \neq \varnothing$, but on $\mathbf{W}_{u,v} = \mathbf{w}$ and $V_{k_\star}^x = v_{k_\star}$. In particular, for $x \in \{u, v\}$, $v_{k_\star} \in \mathcal{L}_{k_\star}^x$ and $\mathbf{w}$ a good weight vector we search for a lower bound for

$$Q_x(v_{k_\star}, \mathbf{w}) = \mathbb{P}\big(v_{k_\star} \not\leftrightarrow \mathcal{T} \big|\ \mathbf{W}_{u,v} = \mathbf{w}, V_{k_\star}^x = v_{k_\star}\big). \tag{6.63}$$

We note that if we find a uniform bound for $Q_x(v_{k_\star}, \mathbf{w})$ for all $x \in \{u, v\}$, $v_{k_\star} \in \mathcal{L}_{k_\star}^x$ and $\mathbf{w}$ that are good, then we have found a lower-bound for both Q_u and Q_v.

Fix any $x \in \{u, v\}$, any good $\mathbf{w}$ and $v_{k_\star} \in \mathcal{L}_{k_\star}^x$. Then, because $\mathbf{w}$ is good, we know that

$$|\mathcal{T}| = |\mathcal{T}(\mathbf{w})| = \left|T_\geq\left(\mathcal{D}_{u,v}, M_\varepsilon^{dB^{k_\star+1}/(\tau-1)}\right)\right| \geq (\sqrt{d})^d M_\varepsilon^{d(A-1)B^{k_\star+1}}. \tag{6.64}$$

Here we abuse notation and write $\mathcal{T}(\mathbf{w})$ rather than $\mathcal{T}(\mathbf{w}, N, M_\varepsilon, A, B, d)$. Enumerate all vertices in $\mathcal{T}$ by $\{t_\ell : \ell = 1, \ldots, |\mathcal{T}|\}$ in way that is independent from the weights, for example by lexicographical ordering of the spatial position. Then, by telescopically using conditional probability, we may write

$$Q_x(v_{k_\star}, \mathbf{w}) = \prod_{\ell=1}^{|\mathcal{T}(\mathbf{w})|} \underbrace{\mathbb{P}\big(v_{k_\star} \not\leftrightarrow t_\ell \mid \text{for all } m = 1, \ldots, \ell-1\ v_{k_\star} \not\sim t_m, \mathbf{W}_{u,v} = \mathbf{w}, V_{k_\star}^x = v_{k_\star}\big)}_{q_x(v_{k_\star}, \mathbf{w}, \ell)}. \tag{6.65}$$

Note that since we already condition on $\mathbf{w}$, we cannot apply Lemma 4.1 in its current form. However, we may imitate its proof to obtain a similar result. To this end, we write

$$\begin{aligned} q_x(v_{k_\star}, \mathbf{w}, \ell) &\leq \sup_{s \geq M_\varepsilon^{dB^{k_\star}/(\tau-1)}} \mathbb{P}\big(v_{k_\star} \not\leftrightarrow t_\ell \;\big|\; \forall m = 1, \ldots, \ell-1 : v_{k_\star} \not\leftrightarrow t_m, \mathbf{W}_{u,v} = \mathbf{w}, V_{k_\star}^x = v_{k_\star}, W_{v_{k_\star}} = s\big) \\ &= \sup_{s \geq M_\varepsilon^{dB^{k_\star}/(\tau-1)}} \mathbb{P}\big(v_{k_\star} \not\leftrightarrow t_\ell \;\big|\; W_{t_\ell} = w_{t_\ell}, W_{v_{k_\star}} = s\big) \\ &\leq \sup_{s \geq M_\varepsilon^{dB^{k_\star}/(\tau-1)}} 1 - \underline{c}\rho(|v_{k_\star} - t_\ell|, w_{t_\ell}, s). \end{aligned} \tag{6.66}$$

Here we have used in the second line that the event $\{v_{k_\star} \not\leftrightarrow t_\ell\}$ depends only on the weight W_{t_ℓ} and $W_{v_{k_\star}}$. Since $\mathbf{w}$ is a realisation of $\mathbf{W}_{u,v}$, it contains the realisation of the weight of t_ℓ, namely $W_{t_\ell} = w_{t_\ell}$. Furthermore, we condition on $W_{v_{k_\star}} = s$, and none of the other events we condition on have any influence over whether $\{v_{k_\star} \not\leftrightarrow t_\ell\}$ happens or not. Furthermore, in the third line we have used (3.7), where ρ is as defined in (3.6) in Assumption 3.2. Next, by item (2) of Claim 6.13, we find that $|v_{k_\star} - t_\ell| \leq 4dM_\varepsilon^{AB^{k_\star+1}}$. Furthermore, as $B > 1$ we find that $M_\varepsilon^{dB^{k_\star}/(\tau-1)} < M_\varepsilon^{dB^{k_\star+1}}$. Taking these two previous observations together, by substituting the definition of ρ and some elementary computations, we find that

$$\begin{aligned} q_x(v_{k_\star}, \mathbf{w}, \ell) &\leq 1 - \inf_{s \geq M_\varepsilon^{dB^{k_\star}/(\tau-1)}} \underline{c}\bigg(1 \wedge \frac{\max\{w_{t_\ell}, s\}^1 \min\{w_{t_\ell}, s\}^\sigma}{|v_{k_\star} - t_\ell|^d}\bigg)^\alpha \\ &\leq 1 - \underline{c}\bigg(1 \wedge \frac{M_\varepsilon^{dB^{k_\star+1}\frac{B+\sigma}{B(\tau-1)}}}{(4dM_\varepsilon^{dAB^{k_\star+1}})^d}\bigg)^\alpha \\ &\leq 1 - C_1 M_\varepsilon^{0 \wedge d\alpha B^{k_\star+1}\left(\frac{B+\sigma}{B(\tau-1)} - A\right)}. \end{aligned} \tag{6.67}$$

Here $C_1 = C_1(\alpha, d) > 0$ and we have used that $1 \wedge x^p = x^{0 \wedge p}$ for $x > 1$. Now, as we have already reasoned before (see for example (6.26)), it holds that $(B+\sigma)/(B(\tau-1)) - A < 0$, so

$$\begin{aligned} q_x(v_{k_\star}, \mathbf{w}, \ell) &\leq 1 - C_1 M_\varepsilon^{d\alpha B^{k_\star+1}\left(\frac{B+\sigma}{B(\tau-1)} - A\right)} \\ &\leq \exp\Big[-C_1 M_\varepsilon^{d\alpha B^{k_\star+1}\left(\frac{B+\sigma}{B(\tau-1)} - A\right)}\Big], \end{aligned} \tag{6.68}$$

where we have used that $1 - x \leq e^{-x}$. Now we substitute this into (6.65) and use that $\mathbf{w}$ is good to see that

$$\begin{aligned} Q_x(v_{k_\star}, \mathbf{w}) &\leq \exp\Big[-C_1 |\mathcal{T}(\mathbf{w})| M_\varepsilon^{d\alpha B^{k_\star+1}\left(\frac{B+\sigma}{B(\tau-1)} - A\right)}\Big] \\ &\leq \exp\Big[-C_2 M_\varepsilon^{dRB^{k_\star+1}}\Big]. \end{aligned} \tag{6.69}$$

Here $C_2 = C_2(\alpha, d) > 0$ and R is as given in (6.59). We now note that the bound for $Q_x(v_{k_\star}, \mathbf{w})$ is uniform over all $x \in \{u, v\}$, $v_{k_\star} \in \mathcal{L}_{k_\star}^x$ and good $\mathbf{w}$. From this and (6.62), we may conclude that

$$Q \leq Q_u + Q_v \leq 2\exp\Big[-C_2 M_\varepsilon^{dRB^{k_\star+1}}\Big]. \tag{6.70}$$

Setting $C_{6.16} = C_{6.16}(\alpha, d) = C_2(\alpha, d)$ and recalling the definition of Q from (6.61) finishes the proof. ■

As announced, we immediately continue with showing that if $V_{k_\star}^u \leftrightarrow \mathcal{T}$ and $V_{k_\star}^v \leftrightarrow \mathcal{T}$, then we may connect $V_{k_\star}^u$ and $V_{k_\star}^v$ through $\mathcal{T}$ using a constant number of vertices. More precisely, if $u_\star, v_\star \in \mathcal{T}$ are such that $V_{k_\star}^u \leftrightarrow u_\star$ and $V_{k_\star}^v \leftrightarrow v_\star$, then we show that with high probability there is a third vertex $t_\star \in \mathcal{T}$ such that $u_\star \leftrightarrow t_\star \leftrightarrow v_\star$. If this succeeds, then $V_{k_\star}^u u_\star t_\star v_\star V_{k_\star}^v$ is a path of length 4 that connects $V_{k_\star}^u$ and $V_{k_\star}^v$. Of course, if $u_\star \leftrightarrow v_\star$ or if $u_\star = v_\star$, we require even fewer edges. Furthermore, there may be other paths that do not solely utilise vertices from $\mathcal{T}$. In either case, however, it holds that $d_{\mathcal{G}}(V_{k_\star}^u, V_{k_\star}^v) \leq 4$ with high probability, as we show in the following Corollary 6.17.

Corollary 6.17 *Consider Setting 6.3, in particular A and B. Let $u, v \in \mathcal{V} = \mathbb{Z}^d$, set $N = |u - v|$, let $M_\varepsilon > 1$ and let $k_\star = k_\star(N, M_\varepsilon, A, B, d)$ be as in Definition 6.11. Furthermore, let $T_\geq$ be as given in Definition 6.6, let $(V_i^u)_{i \leq k_\star}$ and $(V_i^v)_{i \leq k_\star}$ be as given in Definition 6.7 and $\mathcal{D}_{u,v} = \mathcal{D}_{u,v}(N, M_\varepsilon, A, B, d)$ as given in Definition 6.12. Let $\mathbf{W}_{u,v} = \mathbf{W}_{u,v}(N, M_\varepsilon, A, B, d)$, $\mathcal{T}$ and $(N, M_\varepsilon, A, B, d)$-good be as defined in 6.14. Then there exists a function $\mathrm{err}_{6.16} = \mathrm{err}_{6.16}(N, M_\varepsilon, A, B, \sigma, \alpha, \tau, d)$ that satisfies $\mathrm{err}_{6.16} \to 0$ if $N \to \infty$ such that*

$$\mathbb{P}(d_{\mathcal{G}}(V_{k_\star}^u, V_{k_\star^v}) \leq 4 \mid \mathbf{W}_{u,v} \textit{ is good}, V_{k_\star}^u \neq \varnothing, V_{k_\star}^v \neq \varnothing) \geq 1 - \mathrm{err}_{6.16}(N, M_\varepsilon, A, B, \sigma, \alpha, \tau, d). \tag{6.71}$$

Proof. We set

$$Q_{\leq 4} := \mathbb{P}(d_{\mathcal{G}}(V_{k_\star}^u, V_{k_\star}^v) \leq 4 \mid \mathbf{W}_{u,v} \text{ is good}, V_{k_\star}^u \neq \varnothing, V_{k_\star}^v \neq \varnothing) \tag{6.72}$$

and then bound this using the law of total probability to see that

$$Q_{\le 4} \ge \underbrace{\mathbb{P}\big(d_{\mathcal{G}}(V^u_{k_\star}, V_{k^v_\star}) \le 4 \mid \mathbf{W}_{u,v} \text{ is good}, V^u_{k_\star} \neq \varnothing, V^u_{k_\star} \neq \varnothing, V^u_{k_\star} \leftrightarrow \mathcal{T}, V^v_{k_\star} \leftrightarrow \mathcal{T}\big)}_{=:Q_1}$$
$$\cdot \underbrace{\mathbb{P}\big(V^u_{k_\star} \leftrightarrow \mathcal{T}, V^v_{k_\star} \leftrightarrow \mathcal{T} \big| \mathbf{W}_{u,v} \text{ is good}, V^u_{k_\star}, V^v_{k_\star} \neq \varnothing\big)}_{=:Q_2}. \tag{6.73}$$

By Claim 6.16, we have already found a bound for the latter probability Ψ_2. We therefore analyse Ψ_1. We proceed similarly as the proof of Claim 6.16. First, for any good weight vector $\mathbf{w}$ and $u_\star, v_\star \in \mathcal{T}(\mathbf{w})$, set

$$\Psi_1(\mathbf{w}, u_\star, v_\star) := \mathbb{P}(d_{\mathcal{G}}(V^u_{k_\star}, V^v_{k_\star}) \le 4 \mid \mathbf{W}_{u,v} = \mathbf{w}, V^u_{k_\star} \neq \varnothing, V^v_{k_\star} \neq \varnothing, V^u_{k_\star} \leftrightarrow u_\star, V^v_{k_\star} \leftrightarrow v_\star). \tag{6.74}$$

Then note that if we find a uniform bound for Ψ_1, then we have also found a bound for Q_1. Fix any $\mathbf{w}$ and $u_\star, v_\star \in \mathcal{T}(\mathbf{w})$. Now if $u_\star = v_\star$, then we are done and $\Psi_1(\mathbf{w}, u_\star, v_\star) = 1$. Therefore, suppose that $u_\star$ and $v_\star$ are distinct. Then note that if we condition of $V^u_{k_\star} \leftrightarrow u_\star$ and $V^v_{k_\star} \leftrightarrow v_\star$, then $d_{\mathcal{G}}(V^u_{k_\star}, V^v_{k_\star}) \le 4$ happens if $u_\star$ and $v_\star$ are connected with at most one other vertex. Specifically, we search for any other vertex $t_\star \in \mathcal{T}(\mathbf{w}) \setminus \{u_\star, v_\star\}$ such that $u_\star \leftrightarrow t_\star \leftrightarrow v_\star$. To this end, we firstly bound

$$\begin{aligned}\Psi_1(\mathbf{w}, u_\star, v_\star) &\ge \mathbb{P}\big(\exists t_\star \in \mathcal{T}(\mathbf{w}) \setminus \{u_\star, v_\star\} : u_\star \leftrightarrow t_\star \leftrightarrow v_\star \;\big|\; \mathbf{W}_{u,v} = \mathbf{w}, V^u_{k_\star}, V^v_{k_\star} \neq \varnothing, V^u_{k_\star} \leftrightarrow u_\star, V^v_{k_\star} \leftrightarrow v_\star\big) \\ &= \mathbb{P}\big(\exists t_\star \in \mathcal{T}(\mathbf{w}) \setminus \{u_\star, v_\star\} : u_\star \leftrightarrow t_\star \leftrightarrow v_\star \;\big|\; \mathbf{W}_{u,v} = \mathbf{w}\big) =: \Psi_2(\mathbf{w}, u_\star, v_\star),\end{aligned} \tag{6.75}$$

where in the last equality we use that the event $\{\exists t_\star \in \mathcal{T}(\mathbf{w}) \setminus \{u_\star, v_\star\} : u_\star \leftrightarrow t_\star \leftrightarrow v_\star\}$ is independent of $\big\{V^u_{k_\star} \neq \varnothing, V^v_{k_\star} \neq \varnothing, V^u_{k_\star} \leftrightarrow u_\star, V^v_{k_\star} \leftrightarrow v_\star\big\}$ when conditioned on $\mathbf{W}_{u,v} = \mathbf{w}$. Next, we use that since we condition on $\mathbf{W}_{u,v} = \mathbf{w}$, for every $t_\star, \tilde{t}_\star \in \mathcal{T}(\mathbf{w}) \setminus \{u_\star, v_\star\}$ it holds that $\{u_\star \leftrightarrow t_\star \leftrightarrow v_\star\}$ is independent from $\big\{u_\star \leftrightarrow \tilde{t}_\star \leftrightarrow v_\star\big\}$. Furthermore, for any $t_\star \in \mathcal{T}(\mathbf{w}) \setminus \{u_\star, v_\star\}$, when conditioned on $\mathbf{W}_{u,v} = \mathbf{w}$ it holds that $\{u_\star \leftrightarrow t_\star\}$ is independent from $\{t_\star \leftrightarrow v_\star\}$. We may therefore rewrite $\Psi_2(\mathbf{w}, v_\star, u_\star)$ from (6.75) to

$$\Psi_2(\mathbf{w}, v_\star, u_\star) = 1 - \prod_{j=1}^{|\mathcal{T}(\mathbf{w})|-2} \big(1 - \mathbb{P}\big(u_\star \leftrightarrow t_j \mid \mathbf{W}_{u,v} = \mathbf{w}\big)\mathbb{P}\big(t_j \leftrightarrow v_\star \mid \mathbf{W}_{u,v} = \mathbf{w}\big)\big), \tag{6.76}$$

where $\{t_j : j = 1, \dots, |\mathcal{T}(\mathbf{w})| - 2\}$ is the lexicographical ordering in spatial position of $\mathcal{T}(\mathbf{w}) \setminus \{u_\star, v_\star\}$. Next, we apply (3.7) to see that

$$\Psi_2(\mathbf{w}, v_\star, u_\star) = 1 - \prod_{j=1}^{|\mathcal{T}(\mathbf{w})|-2} \big(1 - \underline{c}^2 \rho(|u_\star - t_j|, w_{u_\star}, w_{t_j}) \rho(|t_j - v_\star|, w_{t_j}, w_{v_\star})\big). \tag{6.77}$$

Next by item (2) of Claim 6.13 is holds that $|u_\star - t_j| \le 4dM - \varepsilon^{AB^{k_\star+1}}$ and $|t_j - v_\star| \le 4dM_\varepsilon^{AB^{k_\star+1}}$. Furthermore, because $u_\star, v_\star \in \mathcal{T}(\mathbf{w})$ and $t_j \in \mathcal{T}(\mathbf{w})$ for all $j = 1, \dots, |\mathcal{T}(\mathbf{w})| - 2$, we know that $w_{u_\star}, w_{v_\star}, w_{t_j} \ge M_\varepsilon^{dB^{k_\star+1}/(\tau-1)}$. Thus, we find that

$$\begin{aligned}\rho(|u_\star - t_j|, w_{u_\star}, w_{t_j}) &= \left(1 \wedge \frac{\max\{w_{u_\star}, w_{t_j}\}^1 \min\{w_{u_\star}, w_{t_j}\}^\sigma}{|u_\star - t_j|^d}\right)^\alpha \\ &\ge C_1 M_\varepsilon^{0 \wedge d\alpha B^{k_\star+1}\left(\frac{1+\sigma}{\tau-1} - A\right)},\end{aligned} \tag{6.78}$$

where $C_1 = C_1(\alpha, d) = (4d)^{-d\alpha} \in (0, 1/4)$. Now note that by Setting 6.3, it holds that $A < (1+\sigma)/(\tau-1)$. From (6.78) we therefore obtain that $\rho(|u_\star - t_j|, w_{u_\star}, w_{t_j}) \ge C_1$. In the same way, we also obtain $\rho\big(|t_j - v_\star|, w_{t_j}, w_{v_\star})\big) \ge C_1$. By setting $C_2 = C_2(\alpha, d) = \underline{c}^2 (4d)^{-2d\alpha} \in (0, 1/4)$, from (6.77) we obtain that

$$\Psi_2(\mathbf{w}, u_\star, v_\star) \ge 1 - \prod_{j=1}^{|\mathcal{T}(\mathbf{w})|-2} (1 - C_2) \ge 1 - \exp[-C_2(|\mathcal{T}(\mathbf{w})| - 2)] \ge 1 - \sqrt{e}\exp[-C_2|\mathcal{T}(\mathbf{w})|]. \tag{6.79}$$

Lastly, since $\mathbf{w}$ is good, we know that $|\mathcal{T}(\mathbf{w})| \ge (\sqrt{d})^d M_\varepsilon^{d(A-1)B^{k_\star+1}}$. Substituting this into (6.79) and recalling that Ψ_1 is bounded by Ψ_2 by (6.75), we obtain that

$$\Psi_1(\mathbf{w}, u_\star, v_\star) \ge \Psi_2(\mathbf{w}, u_\star, v_\star) \ge 1 - \sqrt{e}\exp\Big[-C_3 M_\varepsilon^{d(A-1)B^{k_\star+1}}\Big]. \tag{6.80}$$

Here $C_3 = C_3(\alpha, d) = (\sqrt{d})^d C_2(\alpha, d) = \underline{c}^2 4^{d\alpha} d^{d(1/2-\alpha)}$. Now note by (6.80) we have found a uniform bound for $\Psi_1(\mathbf{w}, u_\star, v_\star)$ for all good $\mathbf{w}$ and $u_\star, v_\star \in \mathcal{T}(\mathbf{w})$. As already reasoned, therefore the right-most quantity of (6.80) is also a lower-bound for Q_1 as in (6.73). Furthermore, by Claim 6.16 we also obtain a bound for Q_2 (see (6.73)). By applying both these bounds and setting $C_{6.17} = C_{6.17}(\alpha, d) = C_3(\alpha, d)$, we find that

$$\begin{aligned}Q_{\le 4} \ge Q_1 Q_2 &\ge \left(1 - \sqrt{e}\exp\Big[-C_3 M_\varepsilon^{d(A-1)B^{k_\star+1}}\Big]\right)\left(1 - 2\exp\Big[-C_{6.16} M_\varepsilon^{dB^{k_\star+1}\alpha\left(\frac{B+\sigma}{B(\tau-1)} - A + \frac{A-1}{\alpha}\right)}\Big]\right) \\ &=: 1 - \mathrm{err}_{6.16}(N, M_\varepsilon, A, B, \sigma, \alpha, \tau, d).\end{aligned} \tag{6.81}$$

Note that $\mathrm{err}_{6.16} = \mathrm{err}_{6.16}(N, M_\varepsilon, A, B, \sigma, \alpha, \tau, d)$ from (6.81) satisfies that $\mathrm{err}_{6.16} \downarrow 0$ if $k_\star \uparrow \infty$ (which is equivalent to $N \uparrow \infty$). Recalling the definition of $Q_{\le 4}$ from (6.72) finishes the proof. ■

We note that Corollary 6.17 essentially shows step (2) from Figure 5. Thus, by combining Corollary 6.10 (step (1)), Corollary 6.17 (step (2)) and the fact that all nearest-neighbour edges are present (step

(3)), we have completed the entire construction as described in Figure 5. We are therefore prepared to prove Proposition 6.2. We repeat it here for convenience.

Proposition 6.2 *Consider a KSRG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from Definition 3.1 satisfying Assumption 3.2 with parameters $d \in \mathbb{N}$, $\sigma_1 = 1$, $\sigma_2 = \sigma > 0$, $\alpha > 1$, $\tau \in (2, 2+\sigma)$ and $\mathcal{V} = \mathbb{Z}^d$ and suppose that all nearest-neighbour edges of $\mathbb{Z}^d$ are present in $\mathcal{E}$. Let $u, v \in \mathcal{V}$. Take any $\varepsilon \in (0,1)$ and $\delta > 0$. Then there exists an $\underline{N}_{6.2} = \underline{N}_{6.2}(\delta, \varepsilon, \sigma, \alpha, \tau, d)$ such that if $|u-v| \geq \underline{N}_{6.2}$, then*

$$\mathbb{P}\Big(d_{\mathcal{G}}(u,v) \leq \frac{2+\delta}{\ln\left(\frac{\sigma}{\tau-2}\right)} \ln\ln|u-v|\Big) \geq 1-\varepsilon. \tag{6.2}$$

Proof. Before we continue with the proof, we do some preliminary work and give preliminary definitions. Firstly, set

$$\hat{A} = \hat{A}(\delta, \sigma, \tau) = \frac{1 + \sigma\left(\frac{\tau-2}{\sigma}\right)^{\frac{2}{2+\delta/2}}}{\tau - 1}. \tag{6.82}$$

By recalling I_A from Setting 6.3, it is quickly verified that since $\tau - 2 < \sigma$ and $2/(2+\delta/2) \in (0,1)$ it holds that $\hat{A} \in I_A$. Furthermore, it may be verified that

$$\frac{\sigma}{(\tau-1)\hat{A} - 1} = \left(\frac{\sigma}{\tau-2}\right)^{\frac{2}{2+\delta/2}}. \tag{6.83}$$

Then, take any $\hat{B} = \hat{B}(\delta, \sigma, \tau) \in I_B(\hat{A})$, for example the mid-point of $I_B(\hat{A})$. Then by (6.83), it holds that $\hat{B} > (\sigma/(\tau-2))^{2/(2+\delta/2)}$ and hence

$$\frac{2}{\ln(\hat{B})} < \frac{2+\frac{\delta}{2}}{\ln\left(\frac{\sigma}{\tau-2}\right)}. \tag{6.84}$$

Next, set

$$1 - \mathrm{err}_1(M_\varepsilon, \hat{A}, \hat{B}, \sigma, \alpha, \tau, d) = 1 - 2\exp\left[-\frac{1}{2} M_\varepsilon^{d(\hat{A}-1)}\right] - 4\exp\left[-C_{6.10} M_\varepsilon^{d\hat{B}\left(\frac{\hat{A}-1}{\alpha} + \frac{\hat{B}+\sigma}{\hat{B}(\tau-1)} - 1\right)}\right], \tag{6.85}$$

where $C_{6.10}$ is from Corollary 6.10. We continue by noting that it is possible to choose $\hat{M}_\varepsilon = \hat{M}_\varepsilon(\delta, \sigma, \alpha, \tau, d)$ such that $\hat{M}_\varepsilon \geq \underline{M}_{6.10}(\hat{A}(\delta, \sigma, \tau), \hat{B}(\delta, \sigma, \tau), \alpha, d)$ and

$$1 - \mathrm{err}_1(\hat{M}_\varepsilon, \hat{A}(\delta, \sigma, \tau), \hat{B}(\delta, \sigma, \tau), \sigma, \alpha, \tau, d) \geq 1 - \frac{\varepsilon}{4}. \tag{6.86}$$

Fix such an $\hat{M}_\varepsilon$. Note that for this choice of $\hat{M}_\varepsilon$, by Corollary 6.10 it holds that for any $k \in \mathbb{N}$

$$\mathbb{P}(V_k^u \neq \varnothing) \geq 1 - \frac{\varepsilon}{4} \qquad \text{and} \qquad \mathbb{P}(V_k^v \neq \varnothing) \geq 1 - \frac{\varepsilon}{4}. \tag{6.87}$$

Next, set $N := |u-v|$ and let $k_\star = k_\star(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, d)$ be as given in Definition 6.11. Note that (6.87) also holds with k replaced by $k_\star$. Furthermore, let $(V_i^u)_{i \leq k_\star}$ and $(V_i^v)_{i \leq k_\star}$ be as given in Definition 6.7. Note also that $(V_i^u)_{i \leq k_\star}$ and $(V_i^v)_{i \leq k_\star}$ implicitly depend on $N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, \tau$ and d. Next, let $\mathbf{W}_{u,v} = \mathbf{W}_{u,v}(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, d)$ and $\Big\{\mathbf{W}_{u,v}$ is $(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, d)$-good$\Big\}$ be as in Definition 6.14. After this, set

$$1 - \mathrm{err}_2(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, d) := 1 - \exp\left(-C_{6.15} M_\varepsilon^{d(A-1)B^{k_\star+1}}\right), \tag{6.88}$$

where $C_{6.15} = C_{6.15}(d)$ if as given in Claim 6.15. Furthermore, notice that if $N \to \infty$, then $k_\star \to \infty$ and hence $\mathrm{err}_2(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, d) \to 0$. Also, by Claim 6.15 there exists a $\underline{N}_1 = \underline{N}_1(\delta, \varepsilon, \sigma, \alpha, \tau, d) = \underline{N}_{6.15}(\hat{M}_\varepsilon, \hat{A}, \hat{B}, d)$ such that

$$\mathbb{P}(\mathbf{W}_{u,v} \text{ is good}) \geq 1 - \mathrm{err}_2(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, d). \tag{6.89}$$

Lastly, by Corollary 6.16 there exists a function

$$1 - \mathrm{err}_3(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, \sigma, \alpha, \tau, d) = 1 - \mathrm{err}_{6.16}(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, \sigma, \alpha, \tau, d) \tag{6.90}$$

that satisfies $\mathrm{err}_3 \to 0$ if $N \to \infty$ such that

$$\mathbb{P}(d_{\mathcal{G}}(V_{k_\star}^u, V_{k_\star}^v) \leq 4 \mid \mathbf{W}_{u,v} \text{ is good}, V_{k_\star}^u \neq \varnothing, V_{k_\star}^v \neq \varnothing) \geq 1 - \mathrm{err}_3(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, \sigma, \alpha, \tau, d). \tag{6.91}$$

We continue with the proof of (6.2). To this end, first set

$$Q_{6.2} := \mathbb{P}\Big(d_{\mathcal{G}}(u,v) \leq \frac{2+\delta}{\ln\left(\frac{\sigma}{\tau-2}\right)} \ln\ln|u-v|\Big) \tag{6.92}$$

We firstly use the law of total probability and the definition of conditional probability multiple times to see that

$$Q_{6.2} \geq \underbrace{\mathbb{P}\Big(d_{\mathcal{G}}(u,v) \leq \frac{2+\delta}{\ln\left(\frac{\sigma}{\tau-2}\right)} \ln\ln|u-v| \;\Big|\; V^u_{k_\star} \neq \varnothing, V^v_{k_\star} \neq \varnothing, d_{\mathcal{G}}(V^u_{k_\star}, V^v_{k_\star}) \leq 4\Big)}_{=:\mathfrak{q}_4}$$
$$\cdot \underbrace{\mathbb{P}\Big(d_{\mathcal{G}}(V^u_{k_\star}, V^v_{k_\star}) \leq 4 \;\Big|\; V^u_{k_\star} \neq \varnothing, V^v_{k_\star} \neq \varnothing, \mathbf{W}_{u,v} \text{ is good}\Big)}_{=:\mathfrak{q}_3}$$
$$\cdot \underbrace{\mathbb{P}(\mathbf{W}_{u,v} \text{ is good})}_{=:\mathfrak{q}_2} \underbrace{\mathbb{P}(V^u_{k_\star} \neq \varnothing, V^v_{k_\star} \neq \varnothing)}_{=:\mathfrak{q}_1}. \tag{6.93}$$

We firstly analyse $\mathfrak{q}_1$. Recall $(\mathcal{L}^u_i)_{i\leq k_\star}$ and $(\mathcal{L}^v_i)_{i\leq k_\star}$ from Definition 6.4. We notice that by definition of $k_\star$ and item (2) of Lemma A.1, it holds that $\bigcup_{i=0}^{k_\star} \mathcal{L}^u_i$ is disjoint from $\bigcup_{i=0}^{k_\star} \mathcal{L}^v_i$. Corollary 6.10 then yields that $\{V^u_{k_\star} \neq \varnothing\}$ is independent from $\{V^v_{k_\star} \neq \varnothing\}$. Combining this with (6.87) and $(1-x)(1-y) \geq 1-x-y$ for $x, y \in (0,1)$ yields

$$\mathfrak{q}_1 = \mathbb{P}(V^u_{k_\star} \neq \varnothing)\mathbb{P}(V^v_{k_\star} \neq \varnothing) \geq \left(1 - \frac{\varepsilon}{4}\right)^2 \geq 1 - \frac{\varepsilon}{2}. \tag{6.94}$$

Now by (6.89), (6.91) and (6.94), we know that if $N \geq \underline{N}_1$ then

$$\mathfrak{q}_3\mathfrak{q}_2\mathfrak{q}_1 \geq (1 - \mathrm{err}_3(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, \sigma, \alpha, \tau, d))(1 - \mathrm{err}_2(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, d))\Big(1 - \frac{\varepsilon}{2}\Big) =: \mathfrak{q}^{\geq}_{321}. \tag{6.95}$$

Now since $\mathrm{err}_3, \mathrm{err}_2 \to 0$ if $N \to \infty$, there exists an $\underline{N}_2 = \underline{N}_2(\delta, \varepsilon, \sigma, \alpha, \tau, d)$ such that if $N \geq \underline{N}_2$, then

$$\mathfrak{q}^{\geq}_{321} \geq 1 - \varepsilon. \tag{6.96}$$

Lastly, we show that there exists an $\underline{N}_3 = \underline{N}_3(\delta, \varepsilon, \sigma, \alpha, \tau, d$ such that $\mathfrak{q}_4 = 1$ if $N \geq \underline{N}_3$. We note that if we have shown this, then by setting

$$\underline{N}_{6.2} = \underline{N}_{6.2}(\delta, \varepsilon, \sigma, \alpha, \tau, d) = \max\{\underline{N}_1(\delta, \varepsilon, \sigma, \alpha, \tau, d), \underline{N}_2(\delta, \varepsilon, \sigma, \alpha, \tau, d), \underline{N}_3(\delta, \varepsilon, \sigma, \alpha, \tau, d\} \tag{6.97}$$

we may conclude from (6.93), (6.95) and (6.96) that $Q_{6.2} \geq 1 - \varepsilon$. Recalling the definition of $Q_{6.2}$ from (6.92) then finishes the proof.

It remains to show that this $\underline{N}_3$ exists. To this end, we examine $\mathfrak{q}_4$. Notice that because $V^u_{k_\star} \neq \varnothing$, $V^u_0 \dots V^u_{k_\star}$ is a path of length $k_\star$ from V^u_0 to $V^u_{k_\star}$. Similarly, because $V^v_{k_\star} \neq \varnothing$, $V^v_0 \dots V^v_{k_\star}$ is a path of length $k_\star$ from V^v_0 to $V^v_{k_\star}$ by the construction of $(V^u_i)_{i\leq k_\star}$ and item (1) of Claim 6.8. Furthermore, because we condition on $d_{\mathcal{G}}(V^u_{k_\star}, V^v_{k_\star}) \leq 4$, there is a path utilising at most 4 edges from $V^u_{k_\star}$ to $V^v_{k_\star}$. Next, because $V^u_0 \in \mathcal{L}^u_0$, there is a path from u to V^u_0 utilising at most $d\hat{M}^{\hat{A}}_\varepsilon$ nearest neighbour edges. Similarly, there is a path from v to V^v_0 utilising at most $d\hat{M}^{\hat{A}}_\varepsilon$ edges. By combining all the paths above, we have found a path connecting u and v that utilises $2k_\star + 4 + 2d\hat{M}^{\hat{A}}_\varepsilon$ edges. Now by (6.44) and rewriting,

$$d_{\mathcal{G}}(u,v) \leq 2k_\star + 4 + 2d\hat{M}^{\hat{A}}_\varepsilon \leq \frac{2}{\ln(\hat{B})}\ln\ln(N) + \Bigg(\underbrace{\frac{-\ln(\hat{A}) - \ln\ln(\hat{M}_\varepsilon)}{\ln(\hat{B})\ln\ln(N)} + \frac{4 + 2d\hat{M}^{\hat{A}}_\varepsilon}{\ln\ln(N)}}_{=:g(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, d)}\Bigg)\ln\ln(N). \tag{6.98}$$

Now notice that by (6.84) it holds that $2/\ln(\hat{B}) < (2+\delta/2)/\ln(\sigma/(\tau-2))$. Furthermore, it is clear that $g(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, d) \to 0$ if $N \to \infty$. For this reason, there exists an $\underline{N}_3 = \underline{N}_3(\delta, \varepsilon, \sigma, \alpha, \tau, d)$ such that if $N \geq \underline{N}_3$, then $g(N, \hat{M}_\varepsilon, \hat{A}, \hat{B}, d) \leq \delta/(2\ln(\sigma/(\tau-2)))$. We conclude that if $N \geq \underline{N}_3$, then

$$d_{\mathcal{G}}(u,v) \leq \frac{2 + \frac{\delta}{2}}{\ln\left(\frac{\sigma}{\tau-2}\right)}\ln\ln(N) + \frac{\frac{\delta}{2}}{\ln\left(\frac{\sigma}{\tau-2}\right)} = \frac{2+\delta}{\ln\left(\frac{\sigma}{\tau-2}\right)}. \tag{6.99}$$

As such, if $N \geq \underline{N}_3$, $V^u_{k_\star} \neq \varnothing$, $V^v_{k_\star} \neq \varnothing$ and $d_{\mathcal{G}}(V^u_{k_\star}, V^v_{k_\star}) \leq 4$, it always holds that $d_{\mathcal{G}} \leq (2+\delta)/\ln(\sigma/(\tau-2))$. We conclude that $\mathfrak{q}_4 = 1$. As previously stated, defining $\underline{N}_{6.2}$ as in (6.97) finishes the proof. ∎

6.2. Presence of an infinite component

In this subsection, we reason that almost surely an infinite component is present in a KSRG satisfying Assumption 3.2 with $\mathcal{V} = \mathbb{Z}^d$ and parameters $d \in \mathbb{N}, \sigma_1 = 1, \sigma_2 = \sigma > 0$ $\alpha > 1$ and $\tau \in (2, 2+\sigma)$. In particular, we argue that this is a direct consequence of Corollary 6.10. Denote the event that an infinite component is present by $\{\mathcal{C}_\infty$ is present$\}$ and let $(V^u_i)_{i\in\mathbb{N}\cup\{0\}}$ be as defined in Definition 6.7. Then note that if *all* V^u_i are present (i.e., $V^u_i \neq \varnothing$ for all $i \in \mathbb{N}\cup\{0\}$), then the set $\{V^u_i\}$ forms an infinite component. Therefore, we reason that if with high probability all these vertices are present, then we have found that an infinite component is also almost surely present. To this end, by item (1) of Claim 6.8 we observe for all $k \in \mathbb{N}\cup\{0\}$ it holds $V^u_k \neq \varnothing$ implies that $V^u_j \neq \varnothing$ for $j < k$. By applying this and some elementary computation, we see that

$$\{V^u_i \neq \varnothing \text{ for all } i \in \mathbb{N}\cup\{0\}\} = \lim_{k\to\infty} \bigcap_{i\leq k} \{V^u_i \neq \varnothing\} = \lim_{k\to\infty} \{V^u_k \neq \varnothing\}. \tag{6.100}$$

We also note that $\bigcap_{i\le k}\{V_i^u \neq \varnothing\} = \{V_k^u \neq \varnothing\}$ is a decreasing set in k. Then, by combining all the above, continuity of probability measures and Corollary 6.10 we see that

$$\begin{aligned}\mathbb{P}(\mathcal{C}_\infty \text{ is present}) &\geq \lim_{k\to\infty} \mathbb{P}(V_k^u \neq \varnothing)\\ &\geq 1 - 2\exp\Big[-\frac{1}{2}M_\varepsilon^{d(A-1)}\Big] - 4\exp\Big[-C_{6.10}M_\varepsilon^{d\alpha B\left(\frac{A-1}{\alpha}+\frac{B+\sigma}{B(\tau-1)}-1\right)}\Big]. \end{aligned} \tag{6.101}$$

We note that $\{\mathcal{C}_\infty$ is present$\}$ is a tail-event, so by Kolmogorov's 0-1 law, its probability is either 0 or 1. We may now take M_ε in the right-hand side of (6.101) such that the right-hand side is positive. It follows that $\mathbb{P}(\mathcal{C}_\infty \text{ is present}) = 1$, which is what we wanted to show.

7. Conclusion

7.1. Summary

In this thesis, we have investigated the relation between the graph distance and the spatial distance between two vertices in the KSRG model. The KSRG model is a recent model that generalises known models as long-range percolation, scale-free percolation and age-based spatial preferential attachment. In this model, we choose the underlying vertex set to be $\mathbb{Z}^d$ and independently assign each vertex $v \in \mathcal{V}$ a weight W_v generated by a power-law with parameter $\tau - 1$. We then let each edge conditioned on the weight of the vertices independently present with probability

$$\mathbb{P}(u \leftrightarrow v | W_v, W_u) = \Theta\left(1 \wedge \frac{\max\{W_u, W_v\}^{\sigma_1} \min\{W_u, W_v\}^{\sigma_2}}{|u-v|^d}\right)^{\alpha},$$

where $\sigma_1, \sigma_2 \geq 0$, $\alpha \geq 0$ and $|u - v|$ is the Euclidean distance between u and v.

First, we analyse the expected vertex degree D_v given a realisation of the weight W_v. We show that if $\alpha \leq 1$ or $\tau \leq 1 + \sigma$, then the expected degrees are infinite. There is however a phase transition when $\alpha > 1$ and $\tau > 1 + \sigma$, where the expected degree of a vertex is polynomial in the weight of that vertex. This phase transition may be explained because

- for $\alpha \leq 1$, the spatial decaying factor is not summable, whereas for $\alpha > 1$ it is, and
- for $\tau \leq 1 + \sigma_1$, the weight kernel $\max\{W_u, W_v\}^{\sigma_1} \min\{W_u, W_v\}^{\sigma_2}$ has an infinite expected value, whereas for $\tau > 1 + \sigma_1$ it has a finite expected value.

These results may for example be compared with the work of Deijfen et al. [32].

Next, we show that under the assumption that $\sigma_1 = 1, \sigma_2 = \sigma \in (0, 1)$, $\tau \in (2, 3)$ and α such that $\tau - 1 < \alpha < (\tau - 1)/(\tau - 2)$ and $\alpha\sigma \leq \tau - 1$, then the graph distance is poly-logarithmic with exponent $\Delta = \ln(2)/\ln(\alpha(\tau - 1)/(\alpha + \tau - 1))$. We note that this exponent does not appear in other models such as long-range percolation, scale-free percolation and age-based spatial preferential attachment. Other poly-logarithmic upper-bounds were also found, in particular with exponent $\ln(2)/\ln(2/\alpha)$. This latter exponent *is* present in other models, such as in long-range percolation (see [25]). We note that in the poly-logarithmic regimes, the exponent never becomes 1. This suggest that growth of graph distances yields less than exponential growth in spatial distances. Furthermore, we show that if $\sigma_1 = 1, \sigma_2 = \sigma > 0$, $\alpha > 1$ and $\tau \in (2, 2 + \sigma)$, then the graph distances are most doubly logarithmic. This suggests that the spatial distances grow at least doubly exponentially with the graph distances.

The boundary between the poly-logarithmic phase and the doubly logarithmic phase is $\sigma = \tau - 2$, which yields the age-based spatial preferential attachment model.

To show the first regime, we develop the notion of *nets*. A net is a subset of vertices that behaves pseudo-randomly with regards to the expected amount of vertices with a certain weight in any given radius. In particular, if a vertex v is in a net, then in any radius the amount of other vertices in the net that have a certain weight in a ball of that radius surrounding v is roughly the expected amount. We develop nets to avoid having to rely on the FKG inequality or FKG-like inequalities.

7.2. Further research

We start by giving further research directions directly related to this thesis. To start, as already announced in Subsection 5.4, we note that the proof techniques used to show the poly-logarithmic upper-bound may be applied to find other (and potentially tighter) poly-logarithmic upper-bounds. Other upper-bounds for different choices of parameters may also be investigated; for example we conjecture that in the case that $\mathbb{E}(D_v) = \infty$ then almost surely the diameters of the graph is $\Theta(1)$ (generalising for example results found in the work by Heydenreich, Hulshof and Jorritsma [36]). Another direct open problem is to show matching lower-bounds with the upper-bounds. To do this, in the doubly-logarithmic case the proof of the paper by Deijfen et al. ([32]) may be adapted; in the poly-logarithmic case the paper by Biskup ([25]) may be adapted.

Next, by broadening our view, we note that the idea behind the nets may be generalised or adapted to other proofs to avoid having to use FKG-like inequalities. In particular, we note that if the underlying vertex space of the model is a homogeneous d-dimensional Poisson Point Process with density $\lambda \cdot \mathrm{Leb}_d$, then defining the nets likely requires less work. This is because we may then see combination of the vertex set and the weights of the vertices $(\mathcal{V}, (W_v)_{v \in \mathcal{V}})$ as a $d + 1$ dimensional inhomogeneous Poisson Point Process with density $\lambda \cdot \mathrm{Leb}_d \otimes P_W$. Here Leb_d is the d-dimensional Lebesgue measure and P_W is the distribution of the weights. When viewed this way, it is clear that the amount of vertices of a certain weight in a box is independent of the vertices with any other weight in the same box; each weight-range may therefore independently be examined.

Lastly, we suggest much broader further research. One of the first and most important open problems is to investigate whether or not a KSRG actually fits real-world networks well. Related to this problem is to devise ways to statistically estimate the parameters of the connection probability.

Further research may also be done on processes on a KSRG. For example, one may investigate a random walk on a KSRG, or the spread of an epidemic where each connection is modelled by a KSRG.

A. Appendix — Technical lemmas

Lemma A.1 *Fix $d \in \mathbb{N}$ and let $x, y \in \mathbb{R}^d$. Then fix $\xi > 0$. The following two points hold:*

(1) The hypercube

$$C_x(\xi) := x + \left[-\frac{\xi}{2}, \frac{\xi}{2}\right]^d \tag{A.1}$$

centered around x with sidelengths ξ satisfies

$$\operatorname{diam}(C_x(\xi)) = \sup_{a,b \in C_x(\xi)} |a - b| = \sqrt{d}\xi. \tag{A.2}$$

(2) If $\xi < |x - y|/\sqrt{d}$, then $C_x(\xi) \cap C_y(\xi) = \varnothing$.

Proof. *(1)* Consider two points $a, b \in C_x(\xi)$ and write $a = (a_1, \ldots, a_d), b = (b_1, \ldots, b_d)$. In the j-th coordinate, $j \in [d]$, the two points can differ at most ξ. That is, for all $j \in [d]$ it holds that $|a_j - b_j| \le \xi$. As such, we find that

$$|a - b| = \sqrt{\sum_{j=1}^{d} |a_j - b_j|^2} \le \sqrt{\sum_{j=1}^{d} \xi^2} = \sqrt{d}\xi. \tag{A.3}$$

Furthermore, we can find two points in which this inequality is equal; this happens when we pick two opposing corners of the hypercube.

(2) By definition, each point $a \in C_x(\xi)$ differs at most $\xi/2$ in each coordinate. By the same reasoning done in *(1)*, this means that $|a - x| \le \sqrt{d}\xi/2$. Similarly, $\sup_{b \in C_y(\xi)} |b - y| \le \sqrt{d}\xi/2$. Now suppose that there exists a $z \in C_x(\xi) \cap C_y(\xi)$. Then

$$|x - y| \le |x - z| + |z - y| \le 2\frac{\sqrt{d}\xi}{2} < |x - y|. \tag{A.4}$$

This is a contradiction, so $C_x(\xi) \cap C_y(\xi) = \varnothing$. ■

Lemma A.2 *Suppose that $(X_i)_{i=1}^n$ is a sequence of independent random variables with a power-law distribution with parameter $\tau - 1$. Then for every $c > 1$*

$$\mathbb{P}\left(\max_{i=1,\ldots,n} X_i \ge c\right) \ge 1 - e^{-n/c^{\tau-1}}. \tag{A.5}$$

As a consequence, with probability greater than $1 - e^{-n/c^{\tau-1}}$ there is at least one i such that $X_i > c$.

Proof. We make the following computation:

$$\begin{aligned}
\mathbb{P}\left(\max_{i=1,\ldots,n} X_i \ge c\right) &= 1 - \mathbb{P}\left(\max_{i=1,\ldots,n} X_i < c\right) \\
&= 1 - \left(1 - c^{-(\tau-1)}\right)^n \\
&\ge 1 - e^{-n/c^{\tau-1}}
\end{aligned}$$

where in the second line we use that $(X_i)_{i=1}^n$ is i.i.d., and in the third line we use that $1 - x \le e^{-x}$. ■

Claim A.3 *Let X be a random variable with a power-law distribution with parameter $\tau - 1 > 0$, i.e.*

$$\mathbb{P}(X \ge x) = x^{-(\tau-1)}, \qquad x \ge 1. \tag{A.6}$$

Let $I_1, \ldots, I_n, I_{n+1}$ be $n + 1$ disjoint intervals such that the i'th interval has lower end-point a_i and upper end-point b_i satisfying $1 \le a_i < b_i$, where only b_{n+1} is allowed to be infinite. Then

$$\mathbb{P}\Big(X \in I_{n+1} \Big| X \notin \bigcup_{i=1}^{n} I_i\Big) = \frac{a_{n+1}^{-(\tau-1)} - b_{n+1}^{-(\tau-1)}}{1 - \sum_{i=1}^{n} \left(a_i^{-(\tau-1)} - b_i^{-(\tau-1)}\right)}, \tag{A.7}$$

where by convention $\infty^{-(\tau-1)} = 0$.

Proof. We notice that for any interval I_i we obtain that $\mathbb{P}(X \in I_i) = a_i^{-(\tau-1)} - b_i^{-(\tau-1)}$. We then rewrite the left-hand side of A.7 and use that all I_i are disjoint to see that

$$\begin{aligned}\mathbb{P}\Big(X \in I_{n+1}\Big| X \notin \bigcup_{i=1}^{n} I_i\Big) &= \frac{\mathbb{P}\Big(\{X \in I_{n+1}\} \cap \{X \notin \bigcup_{i=1}^{n} I_i\}\Big)}{1 - \mathbb{P}\Big(X \in \bigcup_{i=1}^{n} I_i\Big)} \\ &= \frac{\mathbb{P}(X \in I_{n+1})}{1 - \sum_{i=1}^{n} \mathbb{P}(X \in I_i)} \\ &= \frac{a_{n+1}^{-(\tau-1)} - b_{n+1}^{-(\tau-1)}}{1 - \sum_{i=1}^{n} \left(a_i^{-(\tau-1)} - b_i^{-(\tau-1)}\right)}\end{aligned}$$

as required. ■

Claim A.4 *Let $\varepsilon \in (0,1)$. Then for all $C > 0$ it holds that*

$$\exp\left[C(\ln\ln N)^{\varepsilon}\right] = (\ln N)^{o(1)}. \tag{A.8}$$

Proof. We may rewrite

$$\exp\left[C(\ln\ln(N)^{\varepsilon})\right] = (\ln N)^{C\frac{(\ln\ln N)^{\varepsilon}}{\ln\ln N}} = (\ln N)^{C(\ln\ln N)^{-(1-\varepsilon)}} = (\ln N)^{o(1)}. \tag{A.9}$$

This finishes the proof. ■

B. Appendix — Chernoff bounds for binomial random variables

In this section we will set up the Chernoff bounds for random variables with a binomial distribution. These bounds will be used that the probability that a binomial random variable differs some fraction from its mean decays exponentially.
The proof of the Chernoff bound depends on Markov's inequality, which we will state first in the following lemma.

Lemma B.1 *(Markov's inequality) Let X be a non-negative random variable and $a > 0$. Then*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Proof. This statement follows directly from the following inequality:

$$\mathbb{E}(X) = \mathbb{E}(X|X \geq a)\mathbb{P}(X \geq a) + \mathbb{E}(X|X < a)\mathbb{P}(X < a) \geq a\mathbb{P}(X \geq a)$$

where we have used that $\mathbb{E}(X|X < a) \geq 0$ and $\mathbb{E}(X|X \geq a) \geq a$. ■

Markov's inequality usually is not very 'tight', meaning that the upper bound given by Markov's inequality can be much larger than the actual value of $\mathbb{P}(X \geq a)$. By appending Markov's inequality we can introduce an exponentially decaying term, which generally gives much tighter results. This is called the Chernoff bound.

Lemma B.2 *(General Chernoff bound) Let X be a real-valued random variable and $a \in \mathbb{R}$. Then the following two inequalities hold:*

$$\mathbb{P}(X \geq a) \leq \inf_{t\geq 0} \frac{\mathbb{E}\left(e^{tX}\right)}{e^{ta}}, \qquad \text{and} \tag{B.1}$$

$$\mathbb{P}(X \leq a) \leq \inf_{t\leq 0} \frac{\mathbb{E}\left(e^{tX}\right)}{e^{ta}}, \tag{B.2}$$

provided that $\mathbb{E}\left(e^{tX}\right)$ exists.

Proof. We consider the first statement. Let $t \geq 0$. By applying Markov's inequality (Lemma B.1) to e^{tX}, we find that

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}\left(e^{tX}\right)}{e^{ta}}$$

Since this holds for every $t \geq 0$, it also holds for the infimum. The second statement is shown in a similar way, but with $t \leq 0$. ■

Now that we have found the general Chernoff bound, we will apply this to binomial random variables.

Corollary B.3 *(Chernoff bound for binomial random variables) Let $n \in \mathbb{N}$ and p, α, β be such that $0 < \beta < p < \alpha < 1$. Consider $X \sim Bin(n,p)$ a random variable with a binomial distribution. Then*

$$\mathbb{P}(X \leq \beta n) \leq \left(\frac{1-p}{1-\beta}\right)^{n(1-\beta)} \left(\frac{p}{\beta}\right)^{n\beta}, \qquad \text{and} \tag{B.3}$$

$$\mathbb{P}(X \geq \alpha n) \leq \left(\frac{1-p}{1-\alpha}\right)^{n(1-\alpha)} \left(\frac{p}{\alpha}\right)^{n\alpha}. \tag{B.4}$$

Proof. We only show the first the inequality, the proof of the second is analogous. We want to apply the Chernoff bound. To this end, we first notice that X is a finite random variable, so that $\mathbb{E}\left(e^{tX}\right)$ exists for all $t \in \mathbb{R}$. In particular, we find that

$$\mathbb{E}\left(e^{tX}\right) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k q^{n-k} = \left(q + pe^t\right)^n,$$

where we have denoted $q = 1 - p$. Now applying the second statement of Lemma B.2, we obtain

$$\mathbb{P}(X \leq \beta n) \leq \inf_{t \leq 0} \left(qe^{-\beta t} + pe^{(1-\beta)t}\right)^n =: \inf_{t \leq 0} g(t)^n.$$

Notice that since $g(t) \geq 0$, to find the infimum of $g(t)^n$ we can just as well find the infimum of $g(t)$. Furthermore, notice that $t \mapsto g(t)$ is continuously differentiable and that $\lim_{t\to\pm\infty} g(t) = +\infty$, which means that $g(t)$ attains its minimum at a stationary point. A simple computation shows that $\underline{t} := \ln\left(\frac{\beta q}{(1-\beta)p}\right)$ is the unique stationary point of $g(t)$. That is,

$$\left.\frac{d}{dt} g(t)\right|_{t=\underline{t}} = 0$$

Furthermore, notice that $\beta < p$ and $q = 1 - p < 1 - \beta$, so that $\underline{t} < 0$ and the minimiser we have found is within the right region.

Filling $\underline{t}$ into g yields

$$\mathbb{P}(X \leq \beta n) \leq g(\underline{t}) = \left(\frac{1-p}{1-\beta}\right)^{n(1-\beta)} \left(\frac{p}{\beta}\right)^{n\beta},$$

which was the desired result. ∎

The Chernoff bounds given in Corollary B.4 can often not be applied directly and requires more treatment. For this reason, we state a slightly worse performing bound that is much more readily applicable. This result is used often in the literature.

Corollary B.4 *Suppose that $X \sim Bin(n,p)$ and let $0 \leq \delta < 1$. Then*

$$\mathbb{P}\left(X \leq (1-\delta)np\right) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{np} \leq e^{-\frac{\delta^2 np}{2}}. \tag{B.5}$$

If additionally δ is such that $(1+\delta)p < 1$,

$$\mathbb{P}\left(X \geq (1+\delta)np\right) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{np} \leq e^{-\frac{\delta^2 np}{2+\delta}}. \tag{B.6}$$

We further remark that for additional ease of use, because $2 + \delta < 3$ in the last inequality we may replace $2 + \delta$ by 3.

Proof of Corollary B.4. We firstly show (B.5). By putting $\beta = (1-\delta)p$ in Corollary B.3, we find that

$$\begin{aligned}
\mathbb{P}(X \leq (1-\delta)np) &\leq \left(\frac{1-p}{1-(1-\delta)p}\right)^{n(1-(1-\delta)p)} \left(\frac{p}{(1-\delta)p}\right)^{n(1-\delta)p} \\
&= \left(1 - \frac{\delta p}{1-(1-\delta)p}\right)^{n(1-(1-\delta)p)} \left(\frac{1}{(1-\delta)^{(1-\delta)}}\right)^{np} \\
&\leq e^{-\frac{\delta p}{1-(1-\delta)p} n(1-(1-\delta)p)} \left(\frac{1}{(1-\delta)^{(1-\delta)}}\right)^{np} = \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{np},
\end{aligned} \tag{B.7}$$

which shows the first inequality of (B.5). To show the second, we rewrite

$$\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} = \exp\left[-\delta - (1-\delta)\ln(1-\delta)\right] = \exp\left[-\delta + (1-\delta)\sum_{k=1}^{\infty} \frac{\delta^k}{k}\right]. \tag{B.8}$$

Here we have written out the Taylor expansion of $\ln(1-\delta)$ (which is valid when $|\delta| < 1$). By writing out the terms we notice that we can rewrite

$$-\delta + (1-\delta)\sum_{k=1}^{\infty} \frac{\delta^k}{k} = -\sum_{k=2}^{\infty} \frac{\delta^k}{k(k-1)} \leq -\frac{\delta^2}{2}. \tag{B.9}$$

Applying this to (B.7) and (B.8) yields the second inequality:

$$\left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{np} \leq e^{-\frac{\delta^2 np}{2}}.$$

Next, we consider the second sequence of inequalities given in (B.6). We again apply Corollary B.3 with $\alpha = (1+\delta)p$, which gives

$$\begin{aligned}
\mathbb{P}(X \geq (1+\delta)np) &\leq \left(\frac{1-p}{1-(1+\delta)p}\right)^{n(1-(1+\delta)p)} \left(\frac{p}{(1+\delta)p}\right)^{n(1+\delta)p)} \\
&= \left(1 + \frac{\delta p}{1-(1+\delta)p}\right)^{n(1-(1+\delta)p} \left(\frac{1}{(1+\delta)^{(1+\delta)}}\right)^{np} \\
&\leq e^{\frac{\delta p}{1-(1+\delta)p} n(1-(1+\delta)p)} \left(\frac{1}{(1+\delta)^{(1+\delta)}}\right)^{np} = \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{np}.
\end{aligned}$$

Differently than above, we now use the inequality $\frac{2\delta}{2+\delta} \leq \ln(1+\delta)$ (valid if $\delta \geq 0$). This inequality can be found by analysing the Taylor expansion of $e^{2x/(2+x)}$ around 0 and rewriting. Applying the inequality, we find that

$$\delta - (1+\delta)\ln(1+\delta) \leq \delta - (1+\delta)\frac{2\delta}{2+\delta} = -\frac{\delta^2}{2+\delta}.$$

This yields the last inequality:

$$\left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{np} \leq e^{-\frac{\delta^2 np}{2+\delta}}.$$

■

References

[1] P. Erdös and A. Rényi, "On random graphs," *Publicationes Mathematicae*, vol. 6, pp. 290–297, 1959.

[2] E. N. Gilbert, "Random graphs," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959, ISSN: 00034851. [Online]. Available: http://www.jstor.org/stable/2237458 (visited on 03/07/2023).

[3] ——, "Random plane networks," *Journal of the Society for Industrial and Applied Mathematics*, vol. 9, no. 4, pp. 533–543, 1961. DOI: 10.1137/0109045. eprint: https://doi.org/10.1137/0109045. [Online]. Available: https://doi.org/10.1137/0109045.

[4] S. Milgram, "The small world problem," *Psychology today*, vol. 2, no. 1, pp. 60–67, 1967.

[5] C. M. Fortuin, J. Ginibre, and P. W. Kasteleyn, "Correlation inequalities on some partially ordered sets," *Communications in Mathematical Physics*, vol. 22, no. 2, pp. 89–103, 1971.

[6] Z. Q. Zhang, F. C. Pu, and B. Z. Li, "Long-range percolation in one dimension," *Journal of Physics A: Mathematical and General*, vol. 16, no. 3, p. L85, 1983. DOI: 10.1088/0305-4470/16/3/002. [Online]. Available: https://dx.doi.org/10.1088/0305-4470/16/3/002.

[7] C. M. Newman and L. S. Schulman, "One dimensional $1/|j-i|^S$ percolation models: The existence of a transition for $S \leq 2$," *Communications in Mathematical Physics*, vol. 104, pp. 547–571, 1986. DOI: 10.1007/BF01211064. [Online]. Available: https://doi.org/10.1007/BF01211064.

[8] M. Kochen, *The Small World*. Ablex, 1989.

[9] B. Bollobás, *Random Graphs*. New York, NY: Springer New York, 1998. DOI: 10.1007/978-1-4612-0619-4_7. [Online]. Available: https://doi.org/10.1007/978-1-4612-0619-4_7.

[10] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998. DOI: 10.1038/30918.

[11] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[12] R. Castro and J. W. Grossman, "Famous trails to paul erdős," *The Mathematical Intelligencer*, vol. 3, no. 21, pp. 51–53, 1999.

[13] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," *SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 4, pp. 251–262, 1999, ISSN: 0146-4833. DOI: 10.1145/316194.316229. [Online]. Available: https://doi.org/10.1145/316194.316229.

[14] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, "Classes of small-world networks," *Proceedings of the national academy of sciences*, vol. 97, no. 21, pp. 11 149–11 152, 2000.

[15] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, "The large-scale organization of metabolic networks," *Nature*, vol. 407, pp. 651–654, 2000. DOI: 10.1038/35036627.

[16] M. E. Newman, "Models of the small world," *Journal of Statistical Physics*, vol. 101, pp. 819–841, 2000.

[17] H. Jeong, Z. Néda, and A. Barabási, "Measuring preferential attachment for evolving networks," *Submitted for publication*, 2001. DOI: cond-mat/0104131.

[18] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 2002. DOI: 10.1103/revmodphys.74.47. [Online]. Available: https://doi.org/10.1103%5C%2Frevmodphys.74.47.

[19] ——, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, pp. 47–97, 1 2002. DOI: 10.1103/RevModPhys.74.47. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.74.47.

[20] Q. Chen, H. Chang, R. Govindan, and S. Jamin, "The origin of power laws in internet topologies revisited," in *Proceedings. twenty-first annual joint conference of the ieee computer and communications societies*, IEEE, vol. 2, 2002, pp. 608–617.

[21] J. Dall and M. Christensen, "Random geometric graphs," *Physical review E*, vol. 66, no. 1, p. 016 121, 2002.

[22] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of networks," *Advances in Physics*, vol. 51, no. 4, pp. 1079–1187, 2002. DOI: 10.1080/00018730110112519. eprint: https://doi.org/10.1080/00018730110112519. [Online]. Available: https://doi.org/10.1080/00018730110112519.

[23] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003. DOI: 10.1137/S003614450342480. eprint: https://doi.org/10.1137/S003614450342480. [Online]. Available: https://doi.org/10.1137/S003614450342480.

[24] M. Penrose, *Random geometric graphs*. Oxford Studies in Probability, 2003. DOI: https://doi.org/10.1093/acprof:oso/9780198506263.001.0001.

[25] M. Biskup, "On the scaling of the chemical distance in long-range percolation models," *The Annals of Probability*, vol. 32, no. 4, pp. 2938–2977, 2004. DOI: 10.1214/009117904000000577. [Online]. Available: https://doi.org/10.1214/009117904000000577.

[26] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, no. 4, pp. 175–308, 2006, ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2005.10.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037015730500462X.

[27] W. Aiello, A. Bonato, C. Cooper, J. Janssen, and P. Prałat, "A spatial web graph model with local influence regions," *Internet Mathematics*, vol. 5, no. 1-2, pp. 175–196, 2008. DOI: 10.1080/15427951.

2008.10129305. eprint: https://doi.org/10.1080/15427951.2008.10129305. [Online]. Available: https://doi.org/10.1080/15427951.2008.10129305.

[28] M. Markošová, "Network model of human language," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 2, pp. 661–666, 2008, ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2007.09.027. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437107009855.

[29] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, "Hyperbolic geometry of complex networks," *Phys. Rev. E*, vol. 82, p. 036 106, 3 2010. DOI: 10.1103/PhysRevE.82.036106. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.82.036106.

[30] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: A network-based approach to human disease," *Nature Reviews Genetics*, vol. 12, pp. 56–68, 2011. DOI: 10.1038/nrg2918.

[31] L. Gugelmann, K. Panagiotou, and U. Peter, *Random Hyperbolic Graphs: Degree Sequence and Clustering*, A. Czumaj, K. Mehlhorn, A. Pitts, and R. Wattenhofer, Eds. Springer Berlin Heidelberg, 2012, pp. 573–585. [Online]. Available: https://arxiv.org/abs/1205.1470.

[32] M. Deijfen, R. Hofstad, van der, and G. Hooghiemstra, "Scale-free percolation," English, *Annales de l'institut Henri Poincare (B): Probability and Statistics*, vol. 49, no. 3, pp. 817–838, 2013, ISSN: 0246-0203. DOI: 10.1214/12-AIHP480.

[33] E. Jacob and P. Mörters, "Spatial preferential attachment networks: Power laws and clustering coefficients," *The Annals of Applied Probability*, vol. 25, no. 2, pp. 632–662, 2015. DOI: 10.1214/14-AAP1006. [Online]. Available: https://doi.org/10.1214/14-AAP1006.

[34] M. Kiwi and D. Mitsche, *A Bound for the Diameter of Random Hyperbolic Graphs*. 2015, pp. 26–39. DOI: 10.1137/1.9781611973761.3. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611973761.3.

[35] R. v. d. Hofstad, *Random Graphs and Complex Networks*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016. DOI: 10.1017/9781316779422.

[36] M. Heydenreich, T. Hulshof, and J. Jorritsma, "Structures in supercritical scale-free percolation," *The Annals of Applied Probability*, vol. 27, no. 4, pp. 2569–2604, 2017. DOI: 10.1214/16-AAP1270. [Online]. Available: https://doi.org/10.1214/16-AAP1270.

[37] C. Hirsch, "From heavy-tailed Boolean models to scale-free Gilbert graphs," *Brazilian Journal of Probability and Statistics*, vol. 31, no. 1, pp. 111–143, 2017. DOI: 10.1214/15-BJPS305. [Online]. Available: https://doi.org/10.1214/15-BJPS305.

[38] K. Bringmann, R. Keusch, and J. Lengler, "Geometric inhomogeneous random graphs," *Theoretical Computer Science*, vol. 760, pp. 35–54, 2019, ISSN: 0304-3975. DOI: https://doi.org/10.1016/j.tcs.2018.08.014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304397518305309.

[39] A. D. Broido and A. Clauset, "Scale-free networks are rare," *Nature Communications*, vol. 10, no. 1017, 2019. DOI: 10.1038/s41467-019-08746-5. [Online]. Available: https://doi.org/10.1038/s41467-019-08746-5.

[40] P. Gracar, A. Grauer, L. Lüchtrath, and P. Mörters, "The age-dependent random connection model," *Queueing Systems*, vol. 93, pp. 309–331, 2019. DOI: 10.1007/s11134-019-09625-y. [Online]. Available: https://doi.org/10.1007/s11134-019-09625-y.

[41] I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov, "Scale-free networks well done," *Physical Review Research*, vol. 1, no. 3, p. 033 034, 2019.

[42] C. Hirsch and C. Mönch, "Distances and large deviations in the spatial preferential attachment model," *Bernoulli*, vol. 26, no. 2, pp. 927–947, 2020. DOI: 10.3150/19-BEJ1121. [Online]. Available: https://doi.org/10.3150/19-BEJ1121.

[43] J. Komjáthy and B. Lodewijks, "Explosion in weighted hyperbolic random graphs and geometric inhomogeneous random graphs," *Stochastic Processes and their Applications*, vol. 130, no. 3, pp. 1309–1367, 2020, ISSN: 0304-4149. DOI: https://doi.org/10.1016/j.spa.2019.04.014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304414918301601.

[44] P. Meel and D. K. Vishwakarma, "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities," *Expert Systems with Applications*, vol. 153, p. 112 986, 2020, ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2019.112986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417419307043.

[45] K. Shu, *Disinformation, Misinformation, and Fake News in Social Media*, H. L. Suhang Wang Dongwon Lee, Ed. Springer Cham, 2020.

[46] L. Cavallaro, O. Bagdasar, P. De Meo, G. Fiumara, and A. Liotta, "Graph and network theory for the analysis of criminal networks," in *Data Science and Internet of Things: Research and Applications at the Intersection of DS and IoT*, G. Fortino, A. Liotta, R. Gravina, and A. Longheu, Eds. Cham: Springer International Publishing, 2021, pp. 139–156. DOI: 10.1007/978-3-030-67197-6_8. [Online]. Available: https://doi.org/10.1007/978-3-030-67197-6_8.

[47] N. Hao, M. Heydenreich, "Graph distances in scale-free percolation: the logarithmic case," 2021. [Online]. Available: https://arxiv.org/abs/2105.05709.

[48] P. Gracar, M. Heydenreich, C. Monch, and P. Morters, "Recurrence versus transience for weight-dependent random connection models," 2022, preprint on https://arxiv.org/abs/1911.04350. DOI: 10.48550/arXiv.1911.04350. [Online]. Available: https://doi.org/10.48550/arXiv.1911.04350.

[49] R. van der Hofstad, P. van der Hoorn, and N. Maitra, "Scaling of the clustering function in spatial inhomogeneous random graphs," *arXiv preprint arXiv:2212.12885*, 2023.

[50] J. Jorritsma, J. Komjáthy, and D. Mitsche, "Cluster-size decay in supercritical kernel-based spatial random graphs," 2023, preprint on https://arxiv.org/abs/2303.00724. DOI: 10.48550/arXiv.2303.00724. [Online]. Available: https://doi.org/10.48550/arXiv.2303.00724.

[51] S. R. Ferrer I Cancho R, "The small world of human language," *Proceedings of the Royal Society of London*, vol. 268, pp. 2261–2265, ISSN: 0962-8452. DOI: https://doi.org/10.1098/rspb.2001.1800.

[52] *Similarweb top website ranking*, https://www.similarweb.com/top-websites/, Accessed: 2023-01-11.

[53] *Six degrees of wikipedia*, https://www.sixdegreesofwikipedia.com/blog/search-results-analysis, Accessed: 2023-01-11.

[54] *Size of wikipedia*, https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia, Accessed: 2023-01-11.

[55] *Wikispeedia*, https://dlab.epfl.ch/wikispeedia/play/, Accessed: 2023-01-11.

[56] *World economic forum: Reports*, https://www.weforum.org/reports/, Accessed: 2023-01-11.