The diameter of KPKVB random graphs

Tobias Müller* Merlijn Staps†

August 1, 2017

Abstract

We consider a random graph model that was recently proposed as a model for complex networks by Krioukov et al. [15]. In this model, nodes are chosen randomly inside a disk in the hyperbolic plane and two nodes are connected if they are at most a certain hyperbolic distance from each other. It has been previously shown that this model has various properties associated with complex networks, including a power-law degree distribution and a strictly positive clustering coefficient. The model is specified using three parameters: the number of nodes N, which we think of as going to infinity, and $\alpha, \nu > 0$, which we think of as constant. Roughly speaking α controls the power law exponent of the degree sequence and ν the average degree.

Earlier work of Kiwi and Mitsche [14] has shown that when $\alpha < 1$ (which corresponds to the exponent of the power law degree sequence being < 3) then the diameter of the largest component is a.a.s. at most polylogarithmic in N. Friedrich and Krohmer [9] have shown it is a.a.s. $\Omega(\log N)$ and they improved the exponent of the polynomial in $\log N$ in the upper bound. Here we show the maximum diameter over all components is a.a.s. $O(\log N)$ thus giving a bound that is tight up to a multiplicative constant.

1 Introduction

The term *complex networks* usually refers to various large real-world networks, occurring diverse fields of science, that appear to exhibit very similar graph theoretical properties. These include having a constant average degree, a so-called power-law degree sequence, clustering and "small distances". In this paper we will study a random graph model that was recently proposed as a model for complex networks and has the above properties. We refer to it as the Krioukov-Papadopoulos-Kitsak-Vahdat-Boguñá model, or KPKVB model,

*Mathematical Institute, Utrecht University, t.muller@uu.nl. Supported in part by the Netherlands Organisation for Scientific Research (NWO) under project nos 612.001.409 and 639.032.529

†Mathematical Institute, Utrecht University, m.staps@uu.nl. This paper is the result of this author's MSc thesis research project, carried out under the supervision of the first author. The thesis is available from [16].

after its inventors [15]. We should however maybe point out that many authors simply refer to the model as "hyperbolic random geometric graphs" or even "hyperbolic random graphs". In the KPKVB model a random geometric graph is constructed in the hyperbolic plane. We use the Poincaré disk representation of the hyperbolic plane, which is obtained when the unit disk $\mathbb{D} = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 < 1\}$ is equipped with the metric given by the differential form $\mathrm{d}s^2 = 4\frac{\mathrm{d}x^2+\mathrm{d}y^2}{(1-x^2-y^2)^2}$. (This means that the length of a curve $\gamma : [0,1] \to \mathbb{D}$ under the metric is given by $2\int_0^1 \frac{\sqrt{(\gamma_1'(t))^2+(\gamma_2'(t))^2}}{1-\gamma_1^2(t)-\gamma_2^2(t)}\mathrm{d}t$.) For a more elaborate, readable introduction to hyperbolic geometry and the various models and properties of the hyperbolic plane, the reader could consult the book of Stillwell [17]. Throughout this paper we will represent points in the hyperbolic plane by polar coordinates (r,ϑ), where r denotes the hyperbolic distance of a point to the origin, and ϑ denotes its angle with the positive x-axis.

We now discuss the construction of the KPKVB random graph. The model has three parameters: the number of vertices N and two additional parameters $\alpha, \nu > 0$. Usually the behavior of the random graph is studied for $N \to \infty$ for a fixed choice of α and ν. We start by setting $R = 2\log(N/\nu)$. Inside the disk $\mathcal{D}_R$ of radius R centered at the origin in the hyperbolic plane we select N points, independent from each other, according to the probability density f on $[0,R] \times (-\pi,\pi]$ given by

$$f(r,\vartheta) = \frac{1}{2\pi}\frac{\alpha\sinh(\alpha r)}{\cosh(\alpha R)-1} \qquad (r \in [0,R], \vartheta \in (-\pi,\pi]).$$

We call this distribution the (α,R)-quasi uniform distribution. For $\alpha = 1$ this is corresponds to the uniform distribution on $\mathcal{D}_R$. We connect points if and only if their hyperbolic distance is at most R. In other words, two points are connected if their hyperbolic distance is at most the (hyperbolic) radius of the disk that the graph lives on. We denote the random graph we have thus obtained by $G(N;\alpha,\nu)$.

As observed by Krioukov et al. [15] and rigorously shown by Gugelmann et al. [10], the degree distribution follows a power law with exponent $2\alpha+1$, the average degree tends to $2\alpha^2\nu/\pi(\alpha-1/2)^2$ when $\alpha > 1/2$, and the (local) clustering coefficient is bounded away from zero a.a.s. (Here and in the rest of the paper a.a.s. stands for *asymptotically almost surely*, meaning with probability tending to one as $N \to \infty$.) Earlier work of the first author with Bode and Fountoulakis [4] and with Fountoulakis [7] has established the "threshold for a giant component": when $\alpha < 1$ then there always is a unique component of size linear in N no matter how small ν (and hence the average degree) is; when $\alpha > 1$ all components are sublinear no matter the value of ν; and when $\alpha = 1$ then there is a critical value ν_c such that for $\nu < \nu_c$ all components are sublinear and for $\nu > \nu_c$ there is a unique linear sized component (all of these statements holding a.a.s.). Whether or not there is a giant component when $\alpha = 1$ and $\nu = \nu_c$ remains an open problem.

In another paper of the first author with Bode and Fountoulakis [5] it was shown that $\alpha = 1/2$ is the threshold for connectivity: for $\alpha < 1/2$ the graph is a.a.s. connected, for $\alpha > 1/2$ the graph is a.a.s. disconnected and when $\alpha = 1/2$ the probability of being connected tends to a continuous, nondecreasing function of ν which is identically one for $\nu \geq \pi$ and strictly less than one for $\nu < \pi$.

Friedrich and Krohmer [8] studied the size of the largest clique as well as the number of cliques of a given size. Boguña et al. [6] and Bläsius et al. [3] considered fitting the KPKVB model to data using maximum likelihood estimation. Kiwi and Mitsche [13] studied the spectral gap and related properties, and Bläsius et al. [2] considered the treewidth and related parameters of the KPKVB model.

Abdullah et al. [1] considered typical distances in the graph. That is, they sampled two vertices of the graph uniformly at random from the set of all vertices and consider the (graph-theoretic) distance between them. They showed that this distance between two random vertices, conditional on the two points falling in the same component, is precisely $(c + o(1)) \cdot \log\log N$ a.a.s. for $1/2 < \alpha < 1$, where $c := -2\log(2\alpha - 1)$.

Here we will study another natural notion related to the distances in the graph, the graph diameter. Recall that the diameter of a graph G is the supremum of the graph distance $d_G(u, v)$ over all pairs u, v of vertices (so it equals infinity if the graph is disconnected). It has been shown previously by Kiwi and Mitsche [14] that for $\alpha \in (\frac{1}{2}, 1)$ the largest component of $G(N; \alpha, \nu)$ has a diameter that is $O\left((\log N)^{8/(1-\alpha)(2-\alpha)}\right)$ a.a.s. This was subsequently improved by Friedrich and Krohmer [9] to $O\left((\log N)^{1/(1-\alpha)}\right)$. Friedrich and Krohmer [9] also gave an a.a.s. lower bound of $\Omega(\log N)$. We point out that in these upper bounds the exponent of $\log N$ tends to infinity as α approaches one.

Here we are able to improve the upper bound to $O(\log N)$, which is sharp up to a multiplicative constant. We are able to prove this upper bound not only in the case when $\alpha < 1$ but also in the case when $\alpha = 1$ and ν is sufficiently large.

Theorem 1. *Let $\alpha, \nu > 0$ be fixed. If either*

(i) $\frac{1}{2} < \alpha < 1$ and $\nu > 0$ is arbitrary, or;

(ii) $\alpha = 1$ and ν is sufficiently large,

then, a.a.s. as $N \to \infty$, every component of $G(N; \alpha, \nu)$ has diameter $O(\log(N))$.

We remark that our result still leaves open what happens for other choices of α, ν as well as several related questions. See Section 5 for a more elaborate discussion of these.

1.1 Organization of the paper

In our proofs we will also consider a Poissonized version of the KPKVB model, where the number of points is not fixed but is sampled from a Poisson distribution with mean N. This model is denoted $G_{\mathrm{Po}}(N; \alpha, \nu)$. It is convenient to work with this Poissonized version of the model as it has the advantage that the numbers of points in disjoint regions are independent (see for instance [12]).

The paper is organized as follows. In Section 2 we discuss a somewhat simpler random geometric graph Γ, introduced in [7], that behaves approximately the same as the (Poissonized) KPKVB model. The graph Γ is embedded into a rectangular domain $\mathcal{E}_R$ in the Euclidean plane $\mathbb{R}^2$. In Section 3.1 we discretize this simplified model by dissecting $\mathcal{E}_R$ into small rectangles. In Section 3.2 we show how to construct relatively short paths in Γ. The constructed paths have length $O(\log(N))$ unless there exist large regions that do not contain any vertex of Γ. In Section 3.3 we use the observations of Section 3.2 to formulate sufficient conditions for the components of the graph Γ to have diameter $O(\log N)$. In Section 4 we show that the probability that Γ fails to satisfy these conditions tends to 0 as $N \to \infty$. We also translate these results to the KPKVB model, and combine everything into a proof of Theorem 1.

2 The idealized model

We start by introducing a somewhat simpler random geometric graph, introduced in [7], that will be used as an approximation of the KPKVB model. Let $X_1, X_2, \ldots \in \mathcal{D}_R$ be an infinite supply of points chosen according the (α, R)-quasi uniform distribution on $\mathcal{D}_R$ described above. Let $G = G(N; \alpha, \nu)$ and $G_{\text{Po}} = G_{\text{Po}}(N; \alpha, \nu)$. Let $Z \sim \text{Po}(N)$ be the number of vertices of G_{Po}. By taking $\{X_1, \ldots, X_N\}$ as the vertex set of G and $\{X_1, \ldots, X_Z\}$ as the vertex set of G_{Po}, we obtain a coupling between G and G_{Po}.

We will compare our hyperbolic random graph to a random geometric graph that lives on the Euclidean plane. To this end, we introduce the map $\Psi : \mathcal{D}_R \to \mathbb{R}^2$ given by $\Psi : (r, \vartheta) \mapsto \left(\vartheta \cdot \frac{1}{2} e^{R/2}, R - r\right)$. The map Ψ works by taking the distance of a point to the boundary of the disk as y-coordinate and the angle of the point as x-coordinate (after scaling by $\frac{1}{2}e^{R/2}$) The image of $\mathcal{D}_R$ under Ψ is the rectangle $\mathcal{E}_R = (-\frac{\pi}{2}e^{R/2}, \frac{\pi}{2}e^{R/2}] \times [0, R] \subset \mathbb{R}^2$ (Figure 1).

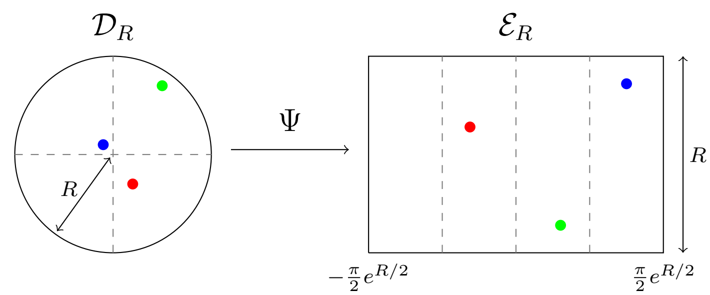

Figure 1: Ψ maps $\mathcal{D}_R$ to a rectangle $\mathcal{E}_R \subset \mathbb{R}^2$.

On $\mathcal{E}_R$ we consider the Poisson point process $\mathcal{P}_{\alpha,\lambda}$ with intensity function $f_{\alpha,\lambda}$ defined by $f_{\alpha,\lambda}(x, y) = \lambda e^{-\alpha y}$. We will denote by $V_{\alpha,\lambda}$ denote the point set of this Poisson process. We also introduce the graph $\Gamma_{\alpha,\lambda}$, with vertex set $V_{\alpha,\lambda}$, where points $(x, y), (x', y') \in V_{\alpha,\lambda}$ are connected if and only if $|x - x'|_{\pi e^{R/2}} \leq e^{\frac{1}{2}(y+y')}$. Here $|a - b|_d = \inf_{k \in \mathbb{Z}} |a - b + kd|$ denotes the distance between a and b modulo d.

If we choose $\lambda = \frac{\nu\alpha}{\pi}$ it turns out that V_λ can be coupled to the image of the vertex set of G_{Po} under Ψ and that the connection rule of Γ_λ approximates the connection rule of G_{Po}. In particular, we have the following:

Lemma 2 ([7], Lemma 27)**.** *Let $\alpha > \frac{1}{2}$. There exists a coupling such that a.a.s. $V_{\alpha,\nu\alpha/\pi}$ is the image of the vertex set of G_{Po} under Ψ.*

Let $\tilde{X}_1, \tilde{X}_2, \ldots \in \mathcal{E}_R$ be the images of $X_1, X_2, \ldots$ under Ψ. On the coupling space of Lemma 2, a.a.s. we have $V_\lambda = \{\tilde{X}_1, \ldots, \tilde{X}_Z\}$.

Lemma 3 ([7], Lemma 30)**.** *Let $\alpha > \frac{1}{2}$. On the coupling space of Lemma 2, a.a.s. it holds for $1 \le i, j \le Z$ that*

(i) if $r_i, r_j \ge \frac{1}{2}R$ and $\tilde{X}_i\tilde{X}_j \in E(\Gamma_{\alpha,\nu\alpha/\pi})$, then $X_iX_j \in E(G_{\mathrm{Po}})$.

(ii) if $r_i, r_j \ge \frac{3}{4}R$, then $\tilde{X}_i\tilde{X}_j \in E(\Gamma_{\alpha,\nu\alpha/\pi}) \iff X_iX_j \in E(G_{\mathrm{Po}})$.

Here r_i and r_j denote the radial coordinates of $X_i, X_j \in \mathcal{D}_R$.

For $(A_i)_i, (B_i)_i$ two sequences of events with A_i and B_i defined on the same probability space $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$, we say that A_i happens a.a.s. conditional on B_i if $\mathbb{P}(A_i \mid B_i) \to 1$ as $i \to \infty$. By a straightforward adaptation of the proofs given in [7], it can be shown that also:

Corollary 4. *The conclusions of Lemmas 2 and 3 also hold conditional on the event $Z = N$.*

In other words, the corollary states that the probability that the conclusions of Lemmas 2 and 3 fail, given that $Z = N$, is also $o(1)$. For completeness, we prove this as Lemmas 19 and 20 in the appendix. An example of G_{Po} and $\Gamma_{\nu\alpha/\pi}$ is shown in Figure 2.

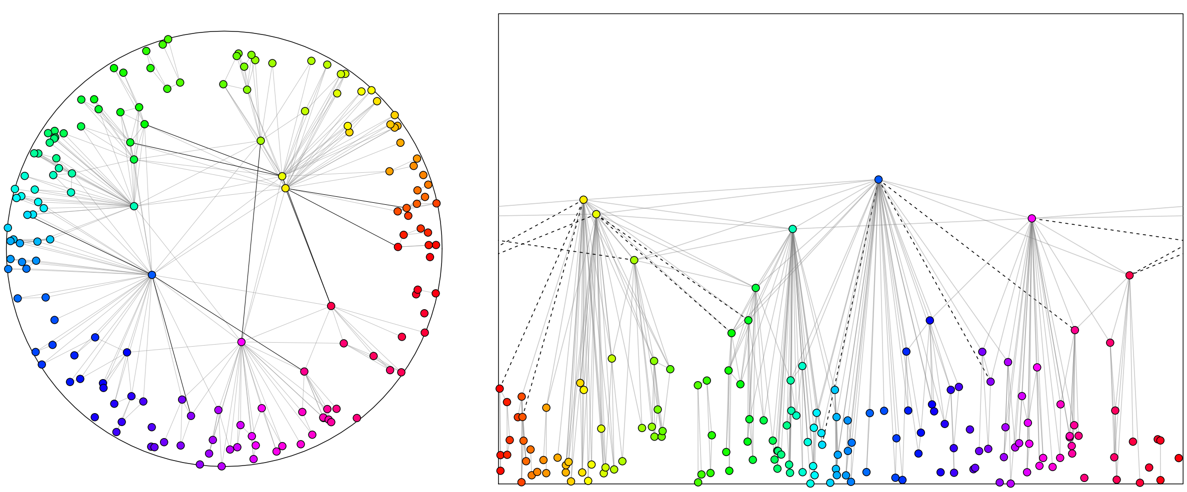

Figure 2: An example of the Poissonized KPKVB random graph G_{Po} (left) and the graph $\Gamma_{\alpha,\nu\alpha/\pi}$ (right), under the coupling of Lemma 2. The graph G_{Po} is drawn in the native model of the hyperbolic plane, where a point with hyperbolic polar coordinates (r, ϑ) is plotted with Euclidean polar coordinates (r, ϑ). Points are colored based on their angular coordinate. The edges for which the coupling fails are drawn in black in the picture of G_{Po} and as dotted lines in the picture of $\Gamma_{\nu\alpha/\pi}$. The parameters used are $N = 200$, $\alpha = 0.8$ and $\nu = 1.3$.

3 Deterministic bounds

For the moment, we continue in a somewhat more general setting, where $V \subset \mathcal{E}_R$ is any finite set of points and Γ is the graph with vertex set V and connection rule $(x, y) \sim (x', y') \iff |x - x'|_{\pi e^{R/2}} \leq e^{\frac{1}{2}(y+y')}$.

3.1 A discretization of the model

We dissect $\mathcal{E}_R$ into a number of rectangles, which have the property that vertices of Γ in rectangles with nonempty intersection are necessarily connected by an edge. This is done as follows. First, divide $\mathcal{E}_R$ into $\ell + 1$ layers L_0, L_1, ..., L_ℓ, where

$$L_i = \{(x, y) \in \mathcal{E}_R : i \log(2) \leq y < (i + 1) \log(2)\}$$

for $i < \ell$ and $L_\ell = \{(x, y) \in \mathcal{E}_R : y \geq \ell \log(2)\}$. Here ℓ is defined by

$$\ell := \left\lfloor \frac{\log(6\pi) + R/2}{\log(2)} \right\rfloor . \tag{1}$$

Note that this gives $6\pi e^{R/2} \geq 2^\ell > 3\pi e^{R/2}$. We divide L_i into $2^{\ell - i}$ (closed) rectangles of equal width $2^{i-\ell} \cdot \pi e^{R/2} = 2^i \cdot b$, where $b = 2^{-\ell} \cdot \pi e^{R/2} \in [\frac{1}{6}, \frac{1}{3})$ is the width of a rectangle in the lowest layer L_0 (Figure 3). In each layer, one of the rectangles has its left edge on the line $x = 0$. We have now partitioned $\mathcal{E}_R$ into $2^{\ell+1} - 1 = O(N)$ boxes.

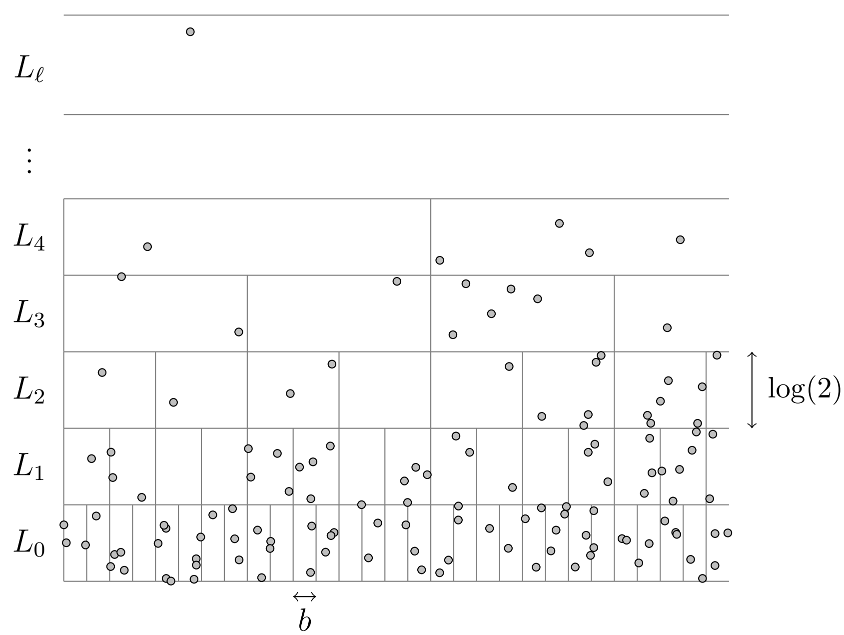

Figure 3: Partitioning $\mathcal{E}_R$ with boxes. All layers except L_ℓ have height log(2). The boxes in layer L_i have width $2^i b$, where $b \in [\frac{1}{6}, \frac{1}{3})$ is the width of a box in L_0. The small circles serve as an example of V.

The boxes are the vertices of a graph $\mathcal{B}$ in which two boxes are connected if they share at least a corner (Figure 4, left). Here we identify the left and right edge of $\mathcal{E}_R$ with each other, so that (for example) also the leftmost and rightmost box in each layer become neighbors. We will call a box *active* if it contains a vertex of Γ and *inactive* otherwise. The dissection has the following properties:

Lemma 5. *The following hold for $\mathcal{B}$ and Γ:*

(i) If vertices of Γ lie in boxes that are neighbors in $\mathcal{B}$, then they are connected by an edge in Γ.

(ii) The number of boxes that lie (partly) above the line $y = R/2$ is at most 63.

Proof: We start with (i). Consider two points (x, y) and (x', y') that lie in boxes that are neighbors in $\mathcal{B}$. Suppose that the lowest of these two points lies in L_i. Then $y, y' \geq i \log(2)$. Furthermore, the horizontal distance between (x, y) and (x', y') is at most 3 times the width of a box in L_i. It follows that

$$|x - x'|_{\pi e^{R/2}} \leq 3 \cdot 2^i \cdot b \leq 2^i \leq e^{\frac{1}{2}(y+y')},$$

so (x, y) and (x', y') are indeed connected in Γ.

To show (ii), we note that the first layer L_i that extends above the line $y = R/2$ has index $i = \lfloor \frac{R/2}{\log 2} \rfloor$. Therefore, we must count the boxes in the layers L_i, L_{i+1}, ..., L_ℓ, of which there are $2^{\ell-i+1} - 1$. We have

$$\ell - i + 1 = \left\lfloor \frac{\log(6\pi) + R/2}{\log 2} \right\rfloor - \left\lfloor \frac{R/2}{\log 2} \right\rfloor + 1 \leq \left\lceil \frac{\log(6\pi)}{\log 2} \right\rceil + 1 = 6,$$

so there are indeed at most $2^6 - 1 = 63$ boxes that extend above the line $y = R/2$. ■

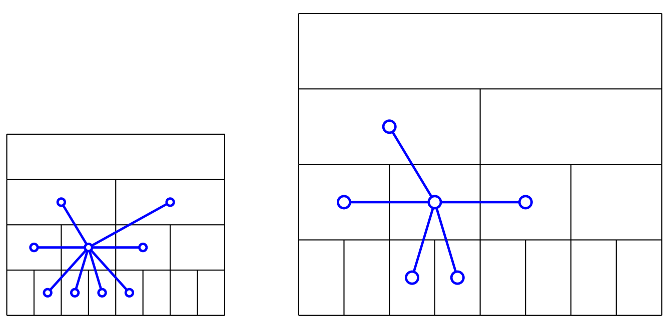

Figure 4: The connection rules of $\mathcal{B}$ and $\mathcal{B}^*$. Left: a box with its 8 neighbors in $\mathcal{B}$. Right: a box with its 5 neighbors in $\mathcal{B}^*$.

Let $\mathcal{B}^*$ be the subgraph of $\mathcal{B}$ where we remove the edges between boxes that have only a single point in common (Figure 4, right). Note that $\mathcal{B}^*$ is a planar graph and that $\mathcal{B}$ is obtained from $\mathcal{B}^*$ by adding the diagonals of each face ([11] deals with a more general notation of *matching pairs* of graphs). We make the following observation (Figure 5; compare Proposition 2.1 in [11]) for later reference.

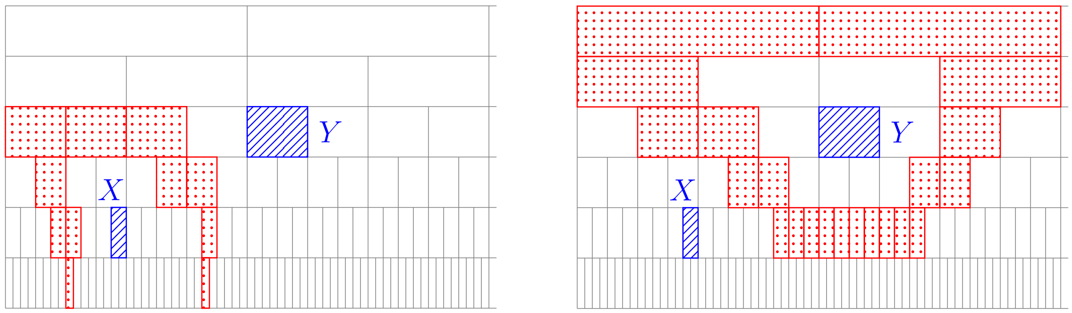

Figure 5: If two blue boxes X and Y are not connected by a path of blue (striped) boxes, then a red (dotted) walk exists that intersects every path in $\mathcal{B}$ from X from Y (Lemma 6). This walk can be chosen such that it either connects two boxes in L_0 (left) or is cyclic (right).

Lemma 6. *Suppose each box in $\mathcal{B}$ is colored red or blue. If there is no path of blue boxes in $\mathcal{B}$ between two blue boxes X and Y, then $\mathcal{B}^*$ (and hence also $\mathcal{B}$) contains a walk of red boxes Q that intersects every walk (and hence also every path) in $\mathcal{B}$ from X to Y.* ■

We leave the straightforward proof of this last lemma to the reader. It can for instance be derived quite succinctly from the Jordan curve theorem. A proof can be found in the MSc thesis of the second author [16].

3.2 Constructing short paths

We will use the boxes defined in the previous subsection to construct short paths between vertices of Γ. Recall that $V \subset \mathcal{E}_R$ is an arbitrary finite set of points and Γ is the graph with vertex set V and connection rule $(x, y) \sim (x', y') \iff |x - x'|_{\pi e^{R/2}} \leq e^{\frac{1}{2}(y+y')}$.

One can easily deduce that the graph Γ has the following geometric properties:

Lemma 7 ([7], Lemma 3)**.** *Let x, y, z, $w \in V$.*

(i) If $xy \in E(\Gamma)$ and z lies above the line segment $[x, y]$ (i.e. $[x, y]$ intersects the segment joining z and the projection of z on the horizontal axis), then at least one of xz and yz is also present in Γ.

(ii) If xy, $wz \in E(\Gamma)$ and the segments $[x, y]$ and $[z, w]$ intersect, then at least one of the edges xw, xz, yw and yz is also present in Γ. In particular, $\{x, y, z, w\}$ is a connected subset of Γ.

We now prove a lemma that allows us to compare paths in Γ with walks in $\mathcal{B}$. This will enable us to translate information about Γ (such as that two boxes contain vertices in the same component of Γ) to information about the states of the boxes.

Lemma 8. *Suppose boxes $X, Y \in \mathcal{B}$ contain vertices $x, y \in V$ respectively that lie in the same component of Γ. Then $\mathcal{B}$ contains a walk $X = B_0, B_1, \ldots, B_n = Y$ with the following property:*

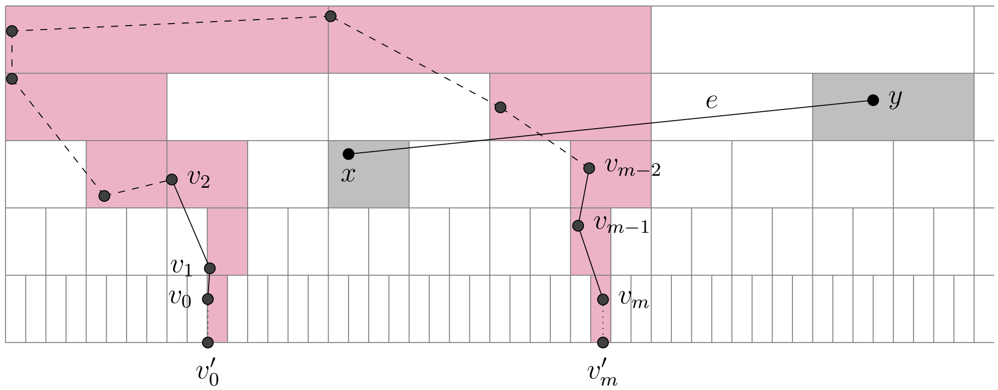

Figure 6: Proof of Lemma 8. The edge e of Γ connects vertices x and y. If a red walk of boxes exists that separates the boxes containing x and y, then e intersects one of the segments $[v_i, v_{i+1}]$, $[v_0, v_0']$ or $[v_m, v_m']$. This contradicts the assumption that no red box contains a neighbor of x or y.

($*$) *if B_i and B_j are active but B_{i+1}, B_{i+2}, …, B_{j-1} are not, then Γ has vertices $a \in B_i$, $b \in B_j$ that are connected in Γ by a path of length at most* 3.

Proof: We prove the statement by induction on the length of the shortest path from x to y in Γ.

First suppose that this length is 1, so that there is an edge e connecting x and y. We claim that a walk $X = B_0, B_1, \ldots, B_n = Y$ in $\mathcal{B}$ exists with the property that if B_i is active, then B_i contains a neighbor of x or y. For this we use Lemma 6. We color a box blue if it is either **a)** inactive or **b)** active and it contains a neighbor of x or y. All other boxes are colored red. Note that X and Y are blue, because X contains a neighbor of y (namely x) and Y contains a neighbor of x (namely y). We intend to show that $\mathcal{B}$ contains a blue path from X to Y. Aiming for a contradiction, we suppose that this is not the case. By Lemma 6, there must then exist a red walk $S = S_0, S_1, \ldots, S_m$ that intersects each path in $\mathcal{B}$ from X to Y. If we remove S from $\mathcal{E}_R$ then $\mathcal{E}_R \backslash S$ falls apart in a number of components. Because there is no path in $\mathcal{B}$ from X to Y that does not intersect S, X and Y lie in different components. (We say S *separates* X and Y.) We choose vertices $v_i \in V \cap S_i$ for all i (these vertices exist because all red boxes are active; see Figure 6). By Lemma 5(i), v_i and v_{i+1} are neighbors in Γ for each i.

We may assume that either S_0 and S_m are both boxes in the lowest layer L_0, or S_0 and S_m are adjacent in $\mathcal{B}$ (Figure 5). In the latter case, we consider the polygonal curve γ consisting of the line segments $[v_0, v_1]$, $[v_1, v_2]$, …, $[v_m, v_0]$. This polygonal curve consists of edges of Γ. Let us observe that each of these edges passes through boxes in S and maybe also boxes adjacent to boxes in S, but the edges cannot intersect any box that is neither on S nor adjacent to a box of S. So in particular, none of these edges can pass through the box X, because X is not adjacent to a box in S (this box should then have been blue by Lemma 5(i)). From this it follows that γ also separates x and y. Therefore, the edge e crosses an edge $[v_i, v_{i+1}]$ of Γ (Figure 6). By Lemma 7(ii) this means that v_i or v_{i+1} neighbors x or y, which is a contradiction because v_i and v_{i+1} do not lie in a blue box.

We are left with the case that S_0 and S_m lie in the lowest layer L_0. Let v'_0 and v'_m denote the projections of v_0 and v_m, respectively, on the horizontal axis. By an analogous argument, we find that the polygonal line through v'_0, v_0, v_1, …, v_m, v'_m separates x and y. We now see that e either crosses an edge $[v_i, v_{i+1}]$ (we then find a contradiction with Lemma 7(ii)) or one of the segments $[v_0, v'_0]$ and $[v_m, v'_m]$ (we then find a contradiction with Lemma 7(i)). From the contradiction we conclude that a blue path must exist connecting X and Y.

We have now shown that if x and y are neighbors in Γ, there exists a walk $X = B_0$, B_1, …, $B_n = Y$ such that the B_i that are active contain a neighbor of x or y. This means that if B_i, B_{i+1}, …, B_j are such that B_i and B_j are active but B_{i+1}, …, B_{j-1} are not, then B_i contains a vertex a that neighbors x or y and B_j contains a vertex b that neighbors x or y. Now $d_\Gamma(a, b) \leq 3$ follows from the fact that both a and b neighbor an endpoint of the same edge e. We conclude that if x and y are neighbors in Γ, then a walk satisfying $(*)$ exists.

Now suppose that the statement holds whenever x and y satisfy $d_\Gamma(x, y) \leq k$ and consider two vertices x and y with $d_\Gamma(x, y) = k+1$. Choose a neighbor y' of y such that $d_\Gamma(x, y') = k$. Let Y' be the active box containing y'. By the induction hypothesis, there exists walks from X to Y' and from Y' to Y satisfying $(*)$. By concatenating these two walks we obtain a walk from X to Y satisfying $(*)$, as desired. ∎

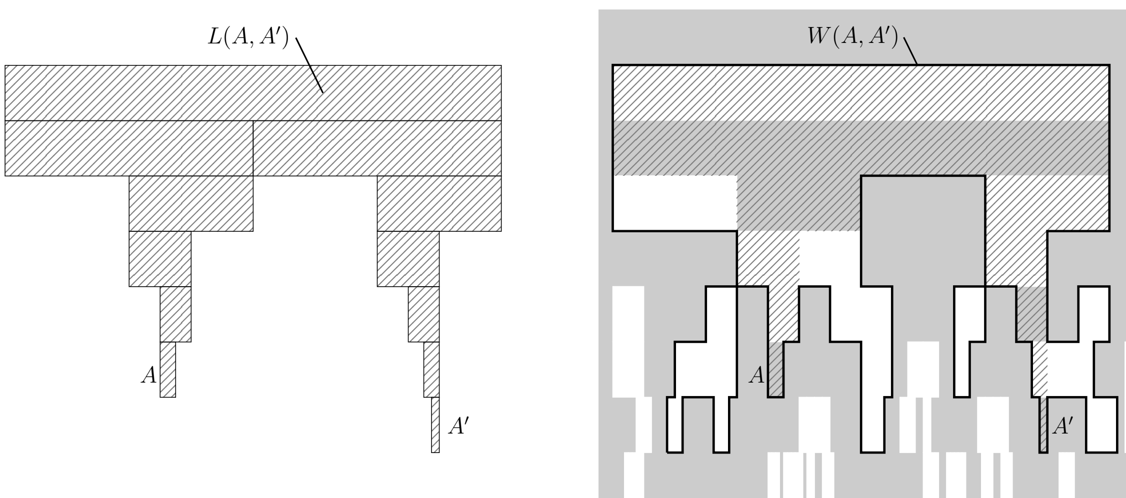

Figure 7: Two boxes A, A' in $\mathcal{B}$ and the path $L(A, A')$ that connects them (left). We can form $L(A, A')$ by concatenating the shortest paths from A and A' to the lowest box lying above both A and A'. In the right image active boxes are colored gray and inactive boxes are colored white. The union of $L(A, A')$ and the inactive components intersecting $L(A, A')$ is called $W(A, A')$ and outlined in black.

From every box in $\mathcal{B}$ there is a path of length $\ell \leq R$ to the unique box in the highest layer L_ℓ. Therefore, for every two boxes $A, A' \in \mathcal{B}$ there exists a path $L(A, A')$ from A to A' of length at most $2R$, consisting of boxes lying above A or A' (Figure 7, left). We define $W(A, A')$ as the set of boxes that either lie in $L(A, A')$ or from which an inactive path exists to a box in $L(A, A')$ (Figure 7, right). Note that $W(A, A')$ is a connected subset of $\mathcal{B}$, consisting of all boxes in $L(A, A')$ and all inactive components intersecting $W(A, A')$ (by an inactive component we mean a component of the induced subgraph of $\mathcal{B}$ on the inactive boxes). We will show that if vertices $x \in A$, $x' \in A'$ of Γ lie in the same component of Γ, then their graph-theoretic distance can be bounded in terms of the size of $W(A, A')$. This gives an upper bound on the diameter of each component of Γ.

Lemma 9. *There exists a constant c such that the following holds (for all finite $V \subseteq \mathcal{E}_R$ with Γ constructed as above). If the vertices $x, x' \in V$ and the boxes $A, A' \in \mathcal{B}$ are such that*

- **(i)** *$x \in A$, $x' \in A'$, and;*
- **(ii)** *x, x' lie in the same component of Γ,*

then $d_\Gamma(x, x') \leq c|W(A, A')|$.

Proof: We claim that there is a walk $S = S_0, S_1, \ldots, S_n$ in $\mathcal{B}$ from A to A' satisfying

- (i) if S_i and S_j are active but $S_{i+1}, \ldots, S_{j-1}$ are not, then Γ has vertices $a \in S_i$, $b \in S_j$ that are connected in Γ by a path of length at most 3;
- (ii) if S_i is active, then either S_i itself or an inactive box adjacent to S_i belongs to $W(A, A')$.

We define $\mathcal{B}_x$ to be the set of active boxes that contain vertices of the component of Γ that contains x and x'. By assumption we have $A, A' \in \mathcal{B}_x$. If A and A' are the only boxes in $L(A, A')$, then we can take $S = (A, A')$ and we're done.

Next, let us suppose A and A' are the only boxes in $L(A, A')$ that belong to $\mathcal{B}_x$. Then the boxes in between A and A' on $L(A, A')$ are either inactive, or they are active but contain vertices of a different component of Γ. The box B in $L(A, A')$ directly following A must be inactive and belongs to some inactive component F (recall that an inactive component is a component of the induced subgraph of $\mathcal{B}$ on the inactive boxes). We will prove the stronger statement that a walk $S = S_0, S_1, \ldots, S_n$ from A to A' exists satisfying (i) and

- (ii') if S_i is active, then S_i is adjacent to a box in F.

By Lemma 8 there exists a walk S from A to A' satisfying (i). We will modify S such that also (ii') holds. We proceed in two steps. In Step 1 we modify S such that all inactive boxes in S that are not in F are removed. In Step 2 we remove all active boxes from S that are not adjacent to a box in F.

Step 1. There is a walk S satisfying (i) that contains no inactive boxes outside F.

We start with the walk S that Lemma 8 provides. This walk satisfies (i). Suppose S contains some inactive box X not in F (Figure 8, left). Because $B \in F$, there can then be no inactive path in $\mathcal{B}$ from X to B. It follows from Lemma 6 that there is an active walk Q that intersects all walks in $\mathcal{B}$ from X to B (we apply Lemma 6 with the inactive boxes colored blue and all other boxes colored red). One such walk from X to B is obtained by following S towards A (which is a neighbor of B). Another possible walk is obtained by first following S towards A' and then following $L(A, A')$ towards B. We define boxes E and E' such that Q intersects the walk in $\mathcal{B}$ from X to B via S and A in E and the walk in $\mathcal{B}$ from X to B via S, A' and $L(A, A')$ in E' (Figure 8, left). Because E belongs to S, E also belongs to $\mathcal{B}_x$. It follows that E' also belongs to $\mathcal{B}_x$, which implies that E' lies in S (the boxes in $L(A, A')$ between A and A' do not lie in $\mathcal{B}_x$ by assumption). We see that Q contains two active boxes E and E' that lie on either side of X. Because Q contains only active boxes, we can replace the part of S from E to E' by a walk of active boxes from E to E'. Doing so we find a walk that still satisfies (i) but from which the box X is removed. By repeatedly applying this procedure, we remove all such boxes X from S. The resulting walk satisfies (i) and contains no inactive boxes outside F.

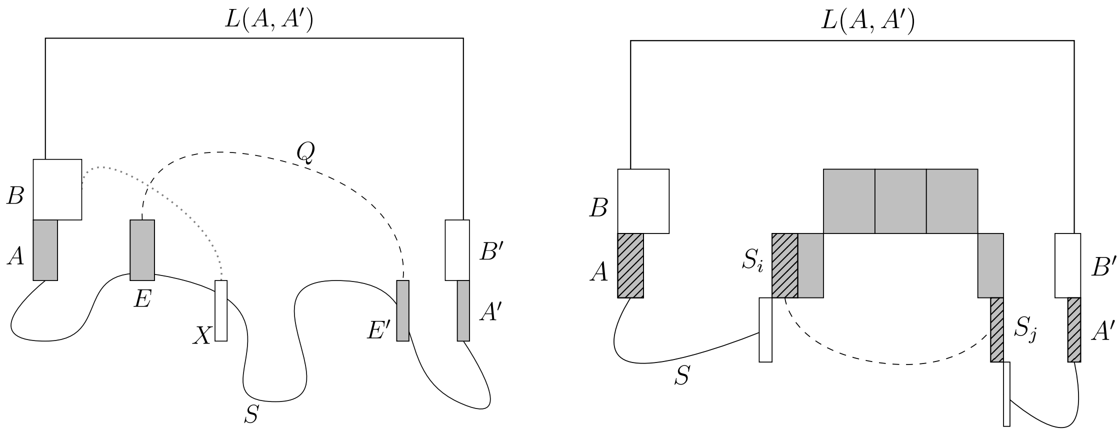

Figure 8: Proof of Lemma 8. Left: Step 1. The walk S satisfies (i) and connects A with A'. If from an inactive box X there is no inactive path to B (dotted line), then there is an active walk Q (dashed line) that connects active boxes E and E' in S on either side of X. Right: Step 2. The walk S satisfies (i) and contains no inactive boxes outside F. The boxes A, S_i, S_j and A' (striped) all belong to F'. The proof works by finding a path in F' from S_i to S_j (dashed line). In both figures active boxes are colored gray and inactive boxes are colored white.

Step 2. There is a walk S satisfying (i) that contains no active boxes outside F', where F' is the set of active boxes adjacent to a box in F.

We start with the walk constructed in Step 1. Since A is adjacent to B it belongs to F'. Let B' be the box in $L(A, A')$ directly preceding A'. We claim that B' belongs to F. Note that B' is inactive. We use Lemma 6 to show that an inactive path from B to B' exists. If such a path would not exist, then an active walk Q would exist that intersects all walks from B to B'. In particular, Q would contain an active box in $L(A, A') \setminus \{A, A'\}$ (which does not lie in $\mathcal{B}_x$, because by assumption A and A' are the only boxes in $L(A, A')$ that belong to $\mathcal{B}_x$) and an active box in S (which lies in $\mathcal{B}_x$, because we know there is a path in Γ from a vertex in this box to a vertex in A). This is a contradiction, because by Lemma 5(i) there cannot be an active walk between a box in $\mathcal{B}_x$ and an active box not in $\mathcal{B}_x$. It follows that an inactive path from B to B' exists, so B' belongs to F. Furthermore, every box in S that has an inactive neighbor in S also lies in F', because this inactive neighbor lies in F by Step 1.

Now consider active boxes S_i, S_{i+1}, ..., S_j in S such that S_i and S_j lie in F' but S_{i+1}, ..., S_{j-1} do not (Figure 8, right). We claim that there is a path in F' from S_i to S_j. Color all boxes in F' blue and all other boxes red. Then our claim is that $\mathcal{B}$ contains a blue path from S_i to S_j. We use Lemma 6 and argue by contradiction. If this blue path would not exist, then there would exist a red walk Q that intersects every walk from S_i to S_j. Because S_i and S_j lie in F', there exists such a walk that apart from S_i and S_j contains only boxes in F. Because Q does not contain S_i and S_j (which are blue) it must contain a box in F. Furthermore, Q also contains one of the active boxes S_{i+1}, ..., S_j. Therefore, Q is a connected set of boxes that contains a box in F and an active box. This implies that Q must also contain a box in F', which contradicts the fact that Q consists of red boxes. This contradiction shows that there must be a blue path in $\mathcal{B}$ from S_i to S_j, i.e. a path in F' from S_i to S_j. We replace the boxes S_{i+1}, ..., S_{j-1} of S by this path, thereby removing the boxes S_{i+1}, ..., S_{j-1} from S. Repeatedly applying this operation, we remove all active boxes that do not lie in F' from S. This completes Step 2.

The walk constructed in Step 2 satisfies (i) and (ii'), so we are now done with the case that $L(A, A')$ contains no boxes in $\mathcal{B}_x$ other than A and A'.

Now suppose A and A' are not the only boxes in $L(A, A')$ that belong to $\mathcal{B}_x$; let $A = A_0$, A_1, ..., $A_n = A'$ be all the boxes in $L(A, A')$ that belong to $\mathcal{B}_x$ (ordered by their position in $L(A, A')$). All these boxes contain vertices in the same component of Γ. For all i we have $L(A_i, A_{i+1}) \subset L(A, A')$ and furthermore A_i and A_{i+1} are the only boxes in $L(A_i, A_{i+1})$ that belong to $\mathcal{B}_x$. Therefore, a walk from A_i to A_{i+1} satisfying (i) and (ii) exists. By concatenating these walks for all i we find a walk S from A to A' satisfying (i) and (ii).

We now construct a path in Γ from x to x' of length at most $37|W(A, A')|$. We may assume that the active boxes in S are all distinct, because if S contains an active box twice we can remove the intermediate part of S. The number of active boxes in S is at most $9|W(A, A')|$ because each active box in S lies in $W(A, A')$ or is one of the at most 8 neighbors of an inactive box in $W(A, A')$. Suppose S_i and S_j are active boxes in S such that S_{i+1}, ..., S_{j-1} are all inactive. Then for every vertex $v \in S_i$ there is a path in Γ of length at most 4 to a vertex in S_j: by (i) there are vertices $a \in S_i$, $b \in S_j$ such that $d_\Gamma(a, b) \leq 3$ and furthermore v and a are neighbors because they lie in the same box. It follows that there is a path of length at most $36|W(A, A')|$ from x to a vertex in A', hence a path of length at most $36|W(A, A')| + 1 \leq 37|W(A, A')|$ from x to x'. This shows that we may take $c = 37$. ■

3.3 Bounding the diameter

In this subsection we continue with the general setting where $V \subseteq \mathcal{E}_R$ is an arbitrary finite set, and Γ is the graph with vertex set V and connection rule $(x, y) \sim (x', y') \iff |x - x'|_{\pi e^{R/2}} \leq$

$e^{\frac{1}{2}(y+y')}$. Here we will translate the bounds from the previous section into results on the maximum diameter of a component of Γ. We start with a general observation on graph diameters.

Lemma 10. *Suppose H_1, H_2 are induced subgraphs of G such that $V(G) = V(H_1) \cup V(H_2)$ (but H_1, H_2 need not be vertex disjoint). If every component of H_1 has diameter at most d_1 and H_2 (is connected and) has diameter at most d_2, then every component of G has diameter at most $2d_1 + d_2 + 2$.*

In particular, if H_2 is a clique, then every component of G has diameter at most $2d_1 + 3$.

Proof: Let C be a component of G. If C contains no vertices of H_2, then C is a component of H_1 as well. So in this case C has diameter at most d_1. If C is not a component of H_1, then for any vertex $v \in C$ there is a path of length at most $d_1 + 1$ from v to a vertex in H_2. Thus, since there is a path of length at most d_2 between any two vertices in H_2, any two vertices $u, v \in C$ have distance at most $(d_1 + 1) + d_2 + (d_1 + 1) = 2d_1 + d_2 + 2$ in G, as required. ■

We let $\tilde{\ell} := \lfloor R/2\log 2 \rfloor - 1$ be the largest i such that layer i is completely below the horizontal line $y = R/2$; we set $\tilde{V} := V \cap \{y \leq \tilde{\ell} \cdot \log 2\}$, and we let $\tilde{\Gamma} := \Gamma[\tilde{V}]$ be the subgraph of Γ induced by $\tilde{V}$. For $A, A' \in \mathcal{B}$ we let $\tilde{W} = \tilde{W}(A, A')$ denote the set $W(A, A')$ but corresponding to $\tilde{V}$ instead of V. (I.e. boxes in layers $\tilde{\ell} + 1, \ldots, \ell$ are automatically inactive. Note that this could potentially increase the size of W substantially.)

The following lemma gives sufficient conditions for an upper bound on the diameter of each component of $\tilde{\Gamma}$. The lemma also deals with graphs that can be obtained by $\tilde{\Gamma}$ by adding a specific type of edges.

Lemma 11. *There exists a constant c such that the following holds. Let $\tilde{\Gamma}, \tilde{W}$ be as above, and let $K = \{(x, y) \in \mathcal{E}_R : y > R/4\}$. Consider the following two conditions:*

(i) *For any two boxes A and A' we have $|\tilde{W}(A, A')| \leq D$;*

(ii) *There is no inactive path (wrt. $\tilde{V}$) in $\mathcal{B}$ connecting a box in L_0 with a box in K.*

If **(i)** *holds, then each component of $\tilde{\Gamma}$ has diameter at most cD. If furthermore* **(ii)** *holds then, for any any graph Γ' that is obtained from $\tilde{\Gamma}$ by adding an arbitrary set of edges E' each of which has an endpoint in K, every component of Γ' has diameter cD.*

Proof: The first statement directly follows from Lemma 9.

If furthermore **(ii)** holds, there exists a cycle of active boxes in $\mathcal{E}_R \backslash K$ that separates K from L_0. Since vertices in neighboring boxes are connected in $\tilde{\Gamma}$, this means that there is a cycle in $\tilde{\Gamma}$ that separates K from L_0. Every vertex in K lies above some edge in this cycle and thereby lies in the component C of this cycle by Lemma 7(i). Thus, every edge of Γ' that is not present in $\tilde{\Gamma}$ has an endpoint in the component C of $\tilde{\Gamma}$.

Let d be the maximum diameter over all components of $\tilde{\Gamma}$. An application of Lemma 10 (with C as one of the two subgraphs; note that we may assume that no added edge connects vertices in the same component, because this can only lower the diameter), we see that the diameter of Γ' is at most $3d + 2$. This proves the second statement (with a larger value of c). ■

4 Probabilistic bounds

We are now ready to use the results from the previous sections to obtain (probabilistic) bounds on the diameters of components in the KPKVB model. Recall from Section 2 that $\Gamma_{\alpha,\lambda}$ is a graph with vertex set $V_{\alpha,\lambda}$, where two vertices (x, y) and (x', y') are connected by an edge if and only if $|x - x'|_{\pi e^{R/2}} \leq e^{\frac{1}{2}(y+y')}$. Here $V_{\alpha,\lambda}$ is the point set of the Poisson process with intensity $f_{\alpha,\lambda} = \mathbf{1}_{\mathcal{E}_R} \lambda e^{-\alpha y}$ on $\mathcal{E}_R = (-\frac{\pi}{2}e^{R/2}, \frac{\pi}{2}e^{R/2}] \times [0, R] \subset \mathbb{R}^2$.

Consistently with the previous sections, we define the subgraph $\tilde{\Gamma}_{\alpha,\lambda}$ of $\Gamma_{\alpha,\lambda}$, induced by the vertices in

$$\tilde{V}_{\alpha,\lambda} := \{(x, y) \in V_{\alpha,\lambda} : y \leq (\tilde{\ell} + 1) \log 2\}.$$

In the remainder of this section all mention of active and inactive (boxes) will be wrt. $\tilde{V}_{\alpha,\lambda}$.

Our plan for the proof of Theorem 1 is to first show that for $\lambda = \nu\alpha/\pi$ the graph $\tilde{\Gamma}_{\alpha,\lambda}$ satisfies the conditions in Lemma 11 for some $D = O(R)$. In the final part of this section we spell out how this result implies that a.a.s. all components of the KPKVB random graph $G(N, \alpha, \nu)$ have diameter $O(R)$.

We start by showing that property (i) of Lemma 11 is a.a.s. satisfied by $\tilde{\Gamma}_{\alpha,\lambda}$. To do so, we need to estimate the probability that a box is active if V is the point set $V_{\alpha,\lambda}$ of the Poisson process above. For $0 \leq i \leq \tilde{\ell}$, let us write

$$p_i = p_{i,\alpha,\lambda} := \mathbb{P}_{\alpha,\lambda}(B \text{ is active}), \tag{2}$$

where $B \in L_i$ is an arbitrary box in layer L_i.

Lemma 12. *For each $0 \leq i < \tilde{\ell}$ we have:*

$$\begin{aligned} p_i &= 1 - \exp\left[-b \cdot \tfrac{2^{1-\alpha}}{\alpha} \cdot \lambda \cdot 2^{(1-\alpha)i}\right] \\ &\geq 1 - \exp\left[-\tfrac{1}{12}\lambda 2^{(1-\alpha)i}\right]. \end{aligned}$$

Proof: The expected number of points of $\tilde{V}_{\alpha,\lambda}$ that fall inside a box B in layer L_i satisfies

$$\mathbb{E}(|B \cap V_{\alpha,\lambda}|) = \int_{i \log(2)}^{(i+1)\log(2)} \int_0^{2^i b} \lambda e^{-\alpha y} \, \mathrm{d}x \, \mathrm{d}y = \lambda \cdot b \cdot \frac{1 - 2^{-\alpha}}{\alpha} \cdot 2^{(1-\alpha)i}$$

Since the number of points that fall in B follows a Poisson distribution and because $b \cdot \frac{1-2^{-\alpha}}{\alpha} \geq \frac{1}{12}$, the result follows. ■

Lemma 13. *There exists a λ_0 such that if $\alpha = 1$ and $\lambda > \lambda_0$ then the following holds. Let E denote the event that there exists a connected subgraph $C \subseteq \mathcal{B}$ with $|C| > R$ such that least half of the boxes of C are inactive. Then $\mathbb{P}_{1,\lambda}(E) = O(N^{-1000})$.*

Proof: The proof is a straightforward counting argument. If C is a connected subset of the boxes graph $\mathcal{B}$ and $A \in C$ is a box of C, then there exists a walk P, starting at A, through all boxes in C, that uses no edge in $\mathcal{B}$ more than twice (this is a general property of a connected graph). Since the maximum degree of $\mathcal{B}$ is 8, the walk P visits no box more than 8 times. Thus, the number of connected subgraphs of $\mathcal{B}$ of cardinality i is no more than $|\mathcal{B}| \cdot 8^{8i} = (2^{\ell+1} - 1) \cdot 2^{24i} = e^{O(R)} \cdot 2^{24i}$ (using the definition (1) of ℓ). Given such a connected subgraph C of cardinality $i > R$ there are $\binom{i}{i/2}$ ways to choose a subset of cardinality $i/2$. Out of any such subset at most 63 boxes lie above level $\tilde{\ell}$ by Lemma 5 and by Lemma 12 each of the remaining $i/2 - 63 > i/4$ is inactive with probability at most $e^{-\lambda/12}$. This gives

$$\begin{aligned}
\mathbb{P}_{1,\lambda}(E) &\leq \textstyle\sum_{i>R} |\mathcal{B}| \cdot 8^{8i} \cdot \binom{i}{i/2} \cdot e^{-(i/2-63)\cdot\lambda/12} \\
&\leq e^{O(R)} \cdot \textstyle\sum_{i>R} 8^{8i} \cdot 2^i \cdot e^{-i/4\cdot\lambda/12} \\
&= e^{O(R)} \cdot O\left(\left(2^{25} \cdot e^{-\lambda/48}\right)^R\right) \\
&= \exp[O(R) - \lambda \cdot \Omega(R)] \\
&= O(N^{-1000}),
\end{aligned}$$

where the third and fifth line follow provided λ is chosen sufficiently large. ■

Corollary 14. *There exist constants c, λ_0 such that if $\alpha = 1$ and $\lambda > \lambda_0$ then*

$$\mathbb{P}_{1,\lambda}\left(\textit{there exist boxes } A, A' \textit{ with } |\tilde{W}(A, A')| > cR\right) = O(N^{-1000}).$$

Proof: We let λ_0 be as provided by Lemma 13 and we take $c := 5$. We note that for every two boxes A, A' the set $\tilde{W}(A, A')$ is a connected set, and all boxes except for some of the at most $2R$ boxes on $L(A, A')$ must be inactive by definition of $\tilde{W}$. Hence if it happens that $|\tilde{W}(A, A')| > 5R$ for some pair of boxes A, A' then event E defined in Lemma 13 holds. The corollary thus follows directly from Lemma 13. ■

We now want to show that in the case when $\frac{1}{2} < \alpha < 1$ and $\lambda > 0$ we also have that, with probability very close to one, $|\tilde{W}(A, A')| = O(R)$ holds for all A, A'. Recall that the probability that a box in layer i is inactive is upper bounded by $\exp(-\frac{\lambda}{12} 2^{(1-\alpha)i})$ (this bound now depends on i), which decreases rapidly if i increases. However, for small values of i this expression could be very close to 1, depending on the value of λ. In particular we cannot hope for something like Lemma 13 to hold for all $\frac{1}{2} < \alpha < 1$ and $\lambda > 0$.

To gain control over the boxes in the lowest layers, we merge boxes in the lowest layers into larger blocks. An *h-block* is defined as the union of a box in L_{h-1} and all $2^h - 2$ boxes lying below this box (Figure 9, left). In other words, an h-block consists of $2^h - 1$ boxes in the lowest h layers that together form a rectangle. The following Lemma shows that the

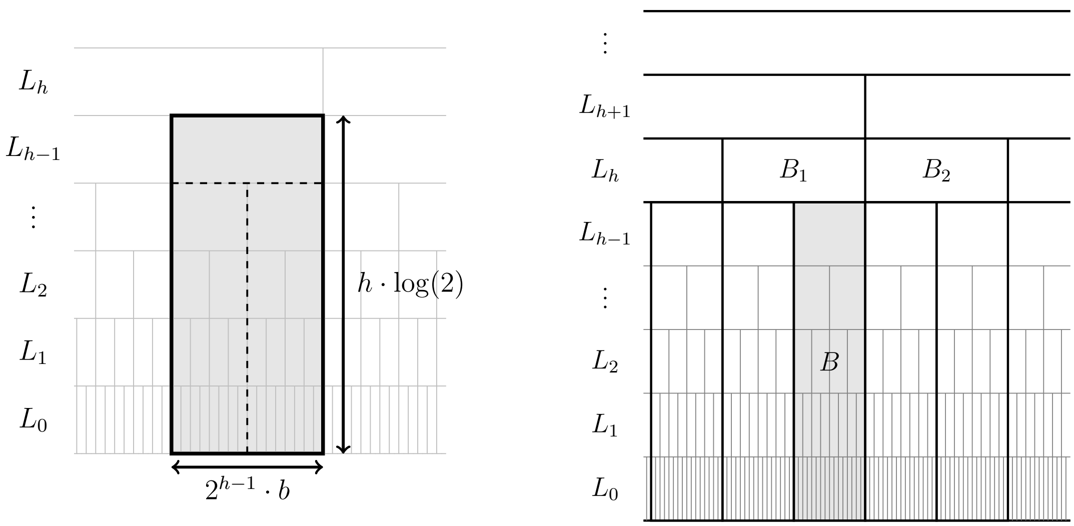

Figure 9: Left: an h-block (in the figure $h = 5$). An h-block is the union of $2^h - 1$ boxes in the lowest h layers. Right: definition of a lonely block (used in the proof of Lemma 16). The lowest h layers are partitioned into h-blocks. An h-block B in W' is called *lonely* if both boxes B_1 and B_2 lying above it are not in W'. If $|W'| > 3$ and B is lonely, one of the blocks adjacent to B contains a horizontal path in W.

probability that an h-block contains a horizontal inactive path can be made arbitrarily small by taking h large.

Let us denote by q_h the probability:

$$q_h = q_{h,\alpha,\lambda} := \mathbb{P}_{\alpha,\lambda}\left(H \text{ has a vertical, active crossing}\right), \tag{3}$$

where H is an arbitrary h-block; and a "vertical, active crossing" means a path of active boxes (in $\mathcal{B}^*$) inside the block connecting the unique box in the highest layer to a box in the bottom layer.

Lemma 15. *If $\alpha < 1$ and $\lambda > 0$ then, for every $\varepsilon > 0$, there exists an $h_0 = h_0(\varepsilon, \alpha, \lambda)$ such that $q_h > 1 - \varepsilon$ for all $h_0 \le h \le \tilde{\ell}$.*

Proof: In the proof that follows, we shall always consider blocks that do not extend above the horizontal line $y = R/2$, (i.e. boxes in layers $h \le R/2 \log 2 - 1$) so that we can use Lemma 12 to estimate the probability that a box is active.

An $(h+1)$-block H consists of one box B in layer L_h and two h-blocks H_1, H_2. There is certainly a vertical, active crossing in H if B is active and either H_1 or H_2 has a vertical, active crossing. In other words,

$$q_{h+1} \ge p_h \cdot (2q_h - q_h^2), \tag{4}$$

where $p_h \ge 1 - \exp[-\frac{1}{12}\lambda 2^{(1-\alpha)h}]$ is the probability that B is active. We choose $\delta = \delta(\varepsilon)$ small, to be made precise shortly. Clearly there is an h_0 such that $p_h > 1 - \delta$ for all

$h_0 \leq h \leq \tilde{\ell}$. Thus (4) gives that $q_{h+1} \geq f(q_h)$ for all such h, where $f(x) = (1-\delta)(2x - x^2)$. It is easily seen that f has fixed points $x = 0, \frac{1-2\delta}{1-\delta}$, that $x < f(x) < \frac{1-2\delta}{1-\delta}$ for $0 < x < \frac{1-2\delta}{1-\delta}$ and $\frac{1-2\delta}{1-\delta} < f(x) < x$ for $\frac{1-2\delta}{1-\delta} < x \leq 1$. Therefore, using that clearly $0 < q_{h_0} < 1$ (there is for instance a strictly positive probability all boxes of the block H are active, resp. inactive), we must have $f^{(k)}(q_{h_0}) \to \frac{1-2\delta}{1-\delta}$ as $k \to \infty$, where $f^{(k)}$ denotes the k-fold composition of f with itself. Hence, provided we chose $\delta = \delta(\varepsilon)$ sufficiently small, there is a $k_0 = k_0(\varepsilon)$ such that $q_{h_0+k} \geq f^{(k)}(q_{h_0}) > 1 - \varepsilon$ for all $k_0 \leq k \leq \tilde{\ell} - h_0$. ■

Lemma 16. *For every $\alpha < 1$ and $\lambda > 0$ there exists a $c = c(\alpha, \lambda)$ such that*

$$\mathbb{P}_{\alpha,\lambda}\Big(\text{there exist boxes } A \text{ and } A' \text{ with } |\tilde{W}(A, A')| > cR\Big) = O(N^{-1000}).$$

Proof: Let p_i be as defined in (2) and q_i as defined in (3). Let $\varepsilon > 0$ be arbitrary, but fixed, to be determined later on in the proof. By Lemmas 12 and 15, there exists an h such that

$$p := \min\{q_h, p_i : h \leq i \leq \tilde{\ell}\} > 1 - \varepsilon.$$

We now create a graph $\mathcal{B}'$, modified from the boxes graph $\mathcal{B}$, as follows. The vertices of $\mathcal{B}'$ are the boxes above layer h, together with the h-blocks. Boxes or blocks are neighbors in $\mathcal{B}'$ if they share at least a corner. Note that the maximum degree of $\mathcal{B}'$ is at most 8.

Given two boxes $A, A' \in \mathcal{B}$ we define $W'(A, A') \subseteq \mathcal{B}'$ as the natural analogue of $\tilde{W}(A, A')$, i.e. the set of all boxes of $\tilde{W}(A, A')$ above layer h together with all h-blocks that contain at least one element of $\tilde{W}(A, A')$. Note that $W'(A, A')$ is a connected set in $\mathcal{B}'$ and that $|W'(A, A')| \geq |\tilde{W}(A, A)|/(2^h - 1)$.

We will say that an h-block B is *lonely* if the two boxes in L_h adjacent to B both do not lie in $\tilde{W}(A, A')$ (Figure 9, right). Observe that if B is lonely, then at least one of the two neighbouring blocks must have a horizontal, inactive crossing. This shows that:

$$|\text{blocks without an active, vertical crossing}| \geq |\text{lonely blocks}|/2. \tag{5}$$

Consider two boxes $A, A' \in \mathcal{B}$ and assume that $|\tilde{W}(A, A')| > cR$, where c is a large constant to be made precise later. By a previous observation $|W'(A, A')| \geq (\frac{c}{2^h - 1})R =: dR$. We distinguish two cases.

Case a): at least $|W'(A, A')|/100$ of the elements of $W'(A, A')$ are boxes (necessarily above layer h). Subtracting the at most 63 boxes of levels $\tilde{\ell} + 1, \ldots, \ell$ and the at most $2R$ boxes of $L(A, A')$, we see that at least $|W'(A, A)|/100 - (2R + 63) \geq |W'(A, A')|/1000$ boxes of $W'(A, A')$ must be inactive and lie in levels $h, \ldots, \tilde{\ell}$. (Here the inequality holds assuming $|W'(A, A')| \geq dR$ with d sufficiently large.)

Case b): at most $|W'(A, A')|/100$ of the elements of $W'(A, A')$ are boxes. Hence, at least $\frac{99}{100}|W'(A, A')|$ of the elements of $W'(A, A')$ are h-blocks. Of these, at least $\frac{97}{100}|W'(A, A')|$ blocks must be lonely, since each box of W' is adjacent to no more than two h-blocks of W'. Thus, by the previous observation (5) at least $\frac{97}{200}|W'(A, A')| \geq |W'(A, A')|/1000$ elements of W' are blocks without a vertical, active crossing.

Combining the two cases, we see that either W' contains $|W'(A,A')|/1000$ inactive boxes in the levels $h, \ldots, \tilde{\ell}$, or W' contains $|W(A,A')|/1000$ blocks without a vertical, active crossing. Summing over all possible choices of A, A' and all possible sizes of $W'(A,A')$, we see that

$$\begin{aligned}
\mathbb{P}_{\alpha,\lambda}\Big(\text{there exist } A, A' \text{ with } |\tilde{W}(A,A')| > cR\Big) &\leq |\mathcal{B}|^2 \cdot \textstyle\sum_{i\geq dR} 8^{8i} \cdot (1-p)^{i/1000} \\
&\leq |\mathcal{B}|^2 \cdot \textstyle\sum_{i\geq dR} \left(8^8 \cdot \varepsilon^{1/1000}\right)^i \\
&= |\mathcal{B}|^2 \cdot O\left(\left(8^8 \cdot \varepsilon^{1/1000}\right)^{dR}\right) \\
&= \exp\left[O(R) - d \cdot \Omega(R)\right] \\
&= O(N^{-1000}),
\end{aligned}$$

where the factor 8^{8i} in the first line is a bound on the number of connected subsets of $\mathcal{B}'$ of cardinality i that contain A, A'; the third line holds provided ε is sufficiently small ($\varepsilon < 8^{-8000}$ will do); and the last line holds provided c (and thus also $d = c/(2^h - 1)$) was chosen sufficiently large. ■

We now turn to the proof of (ii) of Lemma 11.

Lemma 17. *If either*

(i) $\frac{1}{2} < \alpha < 1$ *and* $\lambda > 0$ *is arbitrary, or;*

(ii) $\alpha = 1$ *and* λ *is sufficiently large,*

then, it holds with probability $1 - O(N^{-1000})$ *that there are no inactive paths in* $\mathcal{B}$ *from* L_0 *to* $K := \{(x,y) \in \mathcal{E}_R : y > R/4\}$.

Proof: Since only the boxes below the line $y = R/2$ are relevant, we can freely use Lemma 12. Note that an inactive path in $\mathcal{B}$ from L_0 to K would have length at least $R/4$ (the height of each layer equals $\log 2 < 1$) and that it would have a subpath of length at least $R/8$ that lies completely in $\{(x,y) : y > R/8\}$. Let q be the maximum probability that a box between the lines $y = R/8$ and $y = R/2$ is inactive. Since there are $\exp(O(R))$ boxes and at most 9^k paths of length k starting at any given box, the probability that such a subpath exists is at most $\exp(O(R))9^{R/8}q^{R/8} = \exp(O(R) + \log(q)R/8)$. If $\alpha = 1$ then $q \leq \exp(-\lambda/12)$, which can be chosen arbitrarily small by choosing λ sufficiently large. For sufficiently small q we then have $\exp(O(R)+\log(q)R/8) \leq \exp(-R/2) = O(N^{-1000})$ and therefore such a path does not exist with probability $1 - O(N^{-1000})$. If $\alpha < 1$ we have $q \leq \exp(-\lambda/12 \cdot 2^{(1-\alpha)R/8})$ and it follows that $\exp(O(R) + \log(q)R/8) = \exp(O(R) - \lambda/12 \cdot 2^{(1-\alpha)R/8} \cdot R/8) = \exp(-\omega(R))$, so we can draw the same conclusion. ■

We are almost ready to finally prove Theorem 1, but it seems helpful to first prove a version of the theorem for G_{Po}, the Poissonized version of the model.

Proposition 18 (Theorem 1 for G_{Po})**.** *If either* *(i)* $\frac{1}{2} < \alpha < 1$ *and* $\nu > 0$ *is arbitrary, or;*

(ii) $\alpha = 1$ *and* ν *is sufficiently large,*

then, a.a.s., every component of $G_{\mathrm{Po}}(N; \alpha, \nu)$ *has diameter* $O(\log(N))$*.*

Proof: Let $\tilde{G}_{\mathrm{Po}}$ be the subgraph of G_{Po} induced by the vertices of radius larger than $R - (\tilde{\ell} + 1)\log 2$. (Here $\tilde{\ell} := \lfloor R/2\log 2 \rfloor - 1$ is as before.) By the triangle inequality, all vertices of G_{Po} with distance at most $R/2$ from the origin form a clique. Moveover, by Lemmas 2, 3 and 5, a.a.s. the vertices of G_{Po} with radii between $R/2$ and $R-(\tilde{\ell}+1)\log 2$ can be partitioned into up to 63 cliques corresponding to the boxes above level $\tilde{\ell}$. In other words, a.a.s., $\tilde{G}_{\mathrm{Po}}$ can be obtained from G_{Po} by successively removing up to 64 cliques. Therefore, by up to 64 applications of Lemma 10, it suffices to show that a.a.s. every component of $\tilde{G}_{\mathrm{Po}}$ has diameter $O(\log N)$. Again invoking Lemmas 2 and 3 as well as Lemma 11, it thus suffices to show that a.a.s. $\tilde{\Gamma}_{\alpha,\nu\alpha/\pi}$ satisfies the conditions (i) and (ii) of Lemma 11. This is taken care of by Corollary 14 and Lemma 17 in the case when $\alpha = 1$ and ν is sufficiently large and Lemmas 16 and 17 in the case when $\alpha < 1$ and $\nu > 0$ is arbitrary. ■

Finally, we are ready to give a proof of Theorem 1.

Proof of Theorem 1: Let us point out that G_{Po} conditioned on $Z = N$ has exactly the same distribution as $G = G(N; \alpha, \nu)$. We can therefore repeat the previous proof, where we substitute the use of Lemmas 2 and 3 by Corollary 4, but with one important additional difference. Namely, now we have to show that $\tilde{\Gamma}_{\alpha,\nu\alpha/\pi}$ satisfies the conditions (i) and (ii) of Lemma 11 a.a.s. **conditional on** $Z = N$.

To this end, let E denote the event that $\tilde{\Gamma}_{\alpha,\nu\alpha/\pi}$ fails to have one or both of properties (i) or (ii). By Corollary 14, resp. Lemma 16, and Lemma 17 we have $\mathbb{P}(E) = O(N^{-1000})$. Using the standard fact that $\mathbb{P}(\mathrm{Po}(N) = N) = \Omega(N^{-1/2})$, it follows that

$$\mathbb{P}(E \mid Z = N) \leq \mathbb{P}(E)/\mathbb{P}(Z = N) = O(N^{-1000})/\Omega(N^{-1/2}) = o(1),$$

as required. ■

5 Discussion and further work

In this paper we have given an upper bound of $O(\log N)$ on the diameter of the components of the KPKVB random graph, which holds when $1/2 < \alpha < 1$ and $\nu > 0$ is arbitrary and when $\alpha = 1$ and ν is sufficiently large. Our upper bound is sharp up to the leading constant hidden inside the $O(.)$-notation.

The proof proceeds by considering the convenient idealized model introduced by Fountoulakis and the first author [7], and a relatively crude discretization of this idealized model. The discretization is obtained by dissecting the upper half-plane into rectangles ("boxes") and declaring a box active if it contains at least one point of the idealized model. With a mix of combinatorial and geometric arguments, we are then able to give a deterministic upper bound on the component sizes of the idealized model in terms of the combinatorial structure of the active set of boxes, and finally we apply Peierls-type arguments to give a.a.s. upper bounds for the diameters of all components of the idealized model and the KPKVB model.

We should also remark that the proof in [5] that $G(N;\alpha,\nu)$ is a.a.s. connected when $\alpha < 1/2$ in fact shows that the diameter is a.a.s. $O(1)$ in this case, and that also when $\alpha = 1/2$ the largest component has diameter $O(1)$ a.a.s. Minor variations on the proofs in [5] should give that when $\alpha = 1$, although the graph has a positive probability of being disconnected for some values of ν, for every ν all components will have diameter $O(1)$ a.a.s.

What happens when $\alpha > 1$ or $\alpha = 1$ and ν is arbitrary on the other hand is an open question. In the latter case our methods seem to break down at least partially. We expect that the relatively crude discretization we used in the current paper will not be helpful here and a more refined proof technique will be needed.

Another natural question is to see whether one can say something about the difference between the diameter of the largest component and the other components. We remark that by the work of Friedrich and Krohmer [9] the largest component has diameter $\Omega(\log N)$ a.a.s., but their argument can easily be adapted to show that there will be components other than the largest component that have diameter $\Omega(\log N)$ as well.

Another natural, possibly quite ambitious, goal for further work would be to determine the leading constant (i.e. a constant $c = c(\alpha,\nu)$ such that the diameter of the largest component is $(c + o(1)) \log N$ a.a.s.) if it exists, or indeed even just to establish the existence of a leading constant without actually determining it. We would be especially curious to know if anything special happpens as ν approaches ν_c.

We have only mentioned questions directly related to the graph diameter here. To the best of our knowledge the study of (most) other properties of the KPKVB model is largely virgin territory.

Acknowledgements

We warmly thank Nikolaos Fountoulakis for helpful discussions during the early stages of this project.

References

[1] M. A. Abdullah, M. Bode, and N. Fountoulakis. Typical distances in a geometric model for complex networks, 2015. Preprint. ArXiv version: `https://arxiv.org/abs/1506.07811`.

[2] T. Bläsius, T. Friedrich, and A. Krohmer. Hyperbolic random graphs: Separators and treewidth. In *European Symposium on Algorithms (ESA)*, pages 15:1–15:16, 2016.

[3] T. Bläsius, T. Friedrich, A. Krohmer, and S. Laue. Efficient embedding of scale-free graphs in the hyperbolic plane. In *European Symposium on Algorithms (ESA)*, pages 16:1–16:18, 2016.

[4] M. Bode, N. Fountoulakis, and T. Müller. On the largest component of a hyperbolic model of complex networks. *The Electronic Journal of Combinatorics*, 22(3):P3.24, 2015.

[5] M. Bode, N. Fountoulakis, and T. Müller. The probability of connectivity in a hyperbolic model of complex networks. *Random Structures & Algorithms*, 49(1):65–94, 2016.

[6] M. Boguñá, F. Papadopoulos, and D. Krioukov. Sustaining the internet with hyperbolic mapping. *Nature Communications*, 1(62), 2010.

[7] N. Fountoulakis and T. Müller. Law of large numbers for the largest component in a hyperbolic model of complex networks. Annals of Applied Probability, to appear. ArXiv version: `https://arxiv.org/pdf/1604.02118.pdf`.

[8] T. Friedrich and A. Krohmer. Cliques in hyperbolic random graphs. In *International Conference on Computer Communications (INFOCOM)*, pages 1544–1552. IEEE, 2015.

[9] T. Friedrich and A. Krohmer. On the diameter of hyperbolic random graphs. In M. M. Halldórsson, K. Iwama, N. Kobayashi, and B. Speckmann, editors, *Automata, Languages, and Programming: 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part II*, pages 614–625, Springer Berlin Heidelberg, 2015. ArXiv version: `https://arxiv.org/abs/1512.00184`.

[10] L. Gugelmann, K. Panagiotou, and U. Peter. Random hyperbolic graphs: Degree sequence and clustering. In A. Czumaj, K. Mehlhorn, A. Pitts, and R. Wattenhofer, editors, *Automata, Languages, and Programming: 39th International Colloquium, ICALP 2012, Warwick, UK, July 9-13, 2012, Proceedings, Part II*, pages 573–585, Springer Berlin Heidelberg, 2012. ArXiv version: `https://arxiv.org/abs/1205.1470`.

[11] H. Kesten. *Percolation theory for mathematicians*. Birkhäuser, Boston, 1982.

[12] J. F. C. Kingman. *Poisson Processes*. Oxford Studies in Probability. Clarendon Press, 1993.

[13] M. Kiwi and D. Mitsche. Spectral gap of random hyperbolic graphs and related parameters. Annals of Applied Probabilty, to appear.

[14] M. Kiwi and D. Mitsche. A bound for the diameter of random hyperbolic graphs. In *Proceedings of the Meeting on Analytic Algorithmics and Combinatorics*, pages 26–39, SIAM, 2015. ArXiv version: `https://arxiv.org/abs/1408.2947`.

[15] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá. Hyperbolic geometry of complex networks. *Physical Review E*, 82:036106, 2010.

[16] M. Staps. The diameter of hyperbolic random graphs. MSc thesis, Utrecht University, 2017. Available from `http://studenttheses.library.uu.nl/search.php?language=en`.

[17] J. Stillwell. *Geometry of Surfaces*. Universitext. Springer New York, 1992.

A The proof of Corollary 4

We show that Lemma 2 and 3 are also true when conditioned on $Z = N$. Recall that we say that an event A happens a.a.s. conditional on B if $\mathbb{P}(A \mid B) \to 1$ as $N \to \infty$.

Lemma 19 (Lemma 2 conditional on $Z = N$)**.** *Let $\alpha > \frac{1}{2}$. On the coupling space of Lemma 2, conditional on $Z = N$, a.a.s. $V_{\nu\alpha/\pi}$ is the image of the vertex set of G_{Po} under Ψ.*

Proof: Write $V = \{X_1, \ldots, X_Z\}$ and $\tilde{V} = V_{\nu\alpha/\pi}$. As in [7], there are independent Poisson processes $\mathcal{P}_0$, $\mathcal{P}_1$, $\mathcal{P}_2$ on $\mathcal{E}_R$ such that $\Psi(V) = \mathcal{P}_0 \cup \mathcal{P}_1$, $\tilde{V} = \mathcal{P}_0 \cup \mathcal{P}_2$ and $\mathbb{E}|\mathcal{P}_1|$, $\mathbb{E}|\mathcal{P}_2| = o(1)$. We now find

$$\begin{aligned}\mathbb{P}(\tilde{V} = \Psi(V) \mid Z = N) &= \mathbb{P}(|\mathcal{P}_1| = |\mathcal{P}_2| = 0 \mid |\mathcal{P}_0| + |\mathcal{P}_1| = N) \\ &= \mathbb{P}(|\mathcal{P}_1| = 0 \mid |\mathcal{P}_0| + |\mathcal{P}_1| = N)\mathbb{P}(|\mathcal{P}_2| = 0)\end{aligned}$$

because $\mathcal{P}_0$, $\mathcal{P}_1$ and $\mathcal{P}_2$ are independent. From $\mathbb{E}|\mathcal{P}_2| = o(1)$ it follows that $\mathbb{P}(|\mathcal{P}_2| = 0) = 1 - o(1)$. Furthermore, since the conditional distribution of a Poisson distributed variable given its sum with an independent Poisson distributed variable is binomial, we have $\mathbb{P}(|\mathcal{P}_1| = 0 \mid |\mathcal{P}_0| + |\mathcal{P}_1| = N) = \binom{N}{N}(1 - \mathbb{E}|\mathcal{P}_1|/N)^N = (1 - o(1)/N)^N = 1 - o(1)$, from which it follows that $\mathbb{P}(\tilde{V} = \Psi(V) \mid Z = N) = 1 - o(1)$. ■

Lemma 20 (Lemma 3 conditional on $Z = N$)**.** *Let $\alpha > \frac{1}{2}$. On the coupling space of Lemma 2, conditional on $Z = N$, a.a.s. it holds for $1 \leq i, j \leq Z$ that*

(i) if $r_i, r_j \geq \frac{1}{2}R$ and $\tilde{X}_i\tilde{X}_j \in E(\Gamma_{\alpha,\nu\alpha/\pi})$, then $X_iX_j \in E(G_{\mathrm{Po}})$.

(ii) if $r_i, r_j \geq \frac{3}{4}R$, then $\tilde{X}_i\tilde{X}_j \in E(\Gamma_{\alpha,\nu\alpha/\pi}) \iff X_iX_j \in E(G_{\mathrm{Po}})$.

Here r_i and r_j denote the radial coordinates of $X_i, X_j \in \mathcal{D}_R$.

Proof: Let A denote the event that (i) or (ii) fails for some $i, j \leq Z$ and let B denote the event that (i) or (ii) fails for some $i, j \leq \min(N, Z)$. It follows that

$$\mathbb{P}(B \mid Z \geq N) \leq \frac{\mathbb{P}(B)}{\mathbb{P}(Z \geq N)} \leq \frac{\mathbb{P}(A)}{\mathbb{P}(Z \geq N)} \overset{N\to\infty}{\longrightarrow} \frac{0}{1/2} = 0, \tag{6}$$

because $\mathbb{P}(A) \to 0$ by Lemma 3 and $\mathbb{P}(Z \geq N) \to \frac{1}{2}$ by the central limit theorem. Next, let us observe that $\mathbb{P}(B \mid Z = N) = \mathbb{P}(B \mid Z = N + 1) = \ldots$ since the points with index greater than $\min(N, Z)$ are irrelevant for the event B. From this it follows that

$$\mathbb{P}(B \mid Z \geq N) = \frac{\sum_{i \geq N} \mathbb{P}(B \mid Z = i)\mathbb{P}(Z = i)}{\sum_{i \geq N} \mathbb{P}(Z = i)} = \mathbb{P}(B \mid Z = N). \tag{7}$$

Combining (6) and (7) we see that $\mathbb{P}(B \mid Z = N) = o(1)$, as desired. ■