A proof of Mader's conjecture on large clique subdivisions in C_4-free graphs

Hong Liu [*]   Richard Montgomery [†]

February 1, 2018

Abstract

Given any integers $s, t \geq 2$, we show there exists some $c = c(s, t) > 0$ such that any $K_{s,t}$-free graph with average degree d contains a subdivision of a clique with at least $cd^{\frac{1}{2}\frac{s}{s-1}}$ vertices. In particular, when $s = 2$ this resolves in a strong sense the conjecture of Mader in 1999 that every C_4-free graph has a subdivision of a clique with order linear in the average degree of the original graph. In general, the widely conjectured asymptotic behaviour of the extremal density of $K_{s,t}$-free graphs suggests our result is tight up to the constant $c(s, t)$.

1   Introduction

Given a graph H, a *subdivision* of H is a graph obtained from H by subdividing each of its edges into internally vertex-disjoint paths. A graph G contains an H-subdivision if G contains a subdivision of H as a subgraph. Subdivisions have proved a key notion in the connections between graph theory and topology. Indeed, perhaps the most important historical result in topological graph theory is Kuratowski's theorem from 1930 that planar graphs are exactly those graphs which do not contain a subdivision of the complete graph with five vertices or a subdivision of the complete bipartite graph with three vertices in each class [15].

In 1967, Mader [16] proved the fundamental extremal result that, for each integer $d \geq 1$, there is some $c > 0$ such that any graph G with average degree $d(G)$ at least c contains a subdivision of the complete graph with d vertices, K_d. That is, we may define for each $d \geq 1$

$$s(d) = \inf\{c : d(G) \geq c \implies G \text{ contains a } K_d\text{-subdivision}\}.$$

Mader [16], and independently Erdős and Hajnal [5], conjectured that $s(d) = O(d^2)$. This conjecture matched the known lower bound for $s(d)$, as, for example, the disjoint union of complete regular bipartite subgraphs demonstrates that $s(d) \geq d^2/8$ for $d \geq 3$ (as first observed by Jung [7]).

In 1972, Mader [17] showed that $s(d) = O(2^d)$, but it was not until 1994 that further progress was made by Komlós and Szemerédi [13], who showed that $s(d) = O(d^2 \log^\eta d)$, for any fixed $\eta > 14$. Shortly afterwards, Bollobás and Thomason [4] finally confirmed that $s(d) = O(d^2)$, before Komlós and Szemerédi [14] were able to improve their own methods to give an independent proof. These methods involved graph expansion and have formed the basis for many constructions introduced both here and elsewhere (e.g. [2, 8, 9, 10, 19]). Currently, it is known that

$$(1 + o(1))9d^2/64 \leq s(d) \leq (1 + o(1))10d^2/23, \tag{1}$$

where the upper bound is due to Kühn and Osthus [11] and the lower bound is due to an example by Łuczak.

The extremal examples used to prove the lower bounds mentioned above consist of the disjoint union of dense bipartite graphs. Mader [18] conjectured that any C_4-free graph contains a subdivision of a clique with order linear in its average degree. Towards this, Kühn and Osthus [8, 10] proved that if a graph has sufficiently large girth, then it contains a subdivision of a clique with order linear in its average degree. In fact, they deduced this from the stronger result that such a graph, G say, contains a subdivision of a clique with $\delta(G) + 1$ vertices. Recently, Balogh, Sharifzadeh and the first author [2] showed that for any fixed $k \geq 3$, each C_{2k}-free graph contains a subdivision of a clique with order linear in its average degree.

Approaching Mader's conjecture from a different direction, Kühn and Osthus [9] showed that each C_4-free graph G contains a subdivision of a clique with at least $d(G)/\log^{12}(d(G))$ vertices, when $d(G)$ is sufficiently large. In fact, they were able to show that for all integers $s, t \geq 2$ there exists a $d_0 = d_0(s, t)$ such that every $K_{s,t}$-free graph G of average degree $d \geq d_0$ contains a subdivision of a clique with at least $d^{\frac{1}{2}\frac{s}{s-1}}/\log^{12} d$ vertices [9]. For each s and t, this is likely to be tight up to the logarithmic term (as discussed below, and in [9, Section 4]). Inspired by [2] and [19], we will give new constructions of clique subdivisions to show the following improvement.

Theorem 1.1. *For all integers $t \geq s \geq 2$ there exists some constant $c = c(s, t)$ so that the following holds for every $d > 0$. Every $K_{s,t}$-free graph G with average degree d contains a subdivision of a clique with at least $cd^{\frac{1}{2}\frac{s}{s-1}}$ vertices.*

Note that Theorem 1.1 demonstrates not only that Mader's conjecture is true, but in fact for any fixed $t \geq 2$ the exclusion of $K_{2,t}$-subgraphs in a graph is sufficient to guarantee a subdivision of a clique with order linear in the average degree of the graph.

In [9], Kühn and Osthus proved the following proposition.

Proposition 1.2. *[9, Proposition 13] For every $c > 0$ and all integers $t \geq s \geq 2$ there is a constant $C = C(c, s, t)$ such that no $K_{s,t}$-free graph G with $e(G) \geq c|G|^{2-1/s}$ contains a subdivision of a complete graph of order at least $Cd(G)^{\frac{1}{2}\frac{s}{s-1}}$.*

It is widely expected that for each $t \geq s \geq 2$ there exists some $c' = c'(s, t) > 0$ for which there are arbitrarily large $K_{s,t}$-free graphs G with at least $c'|G|^{2-1/s}$ edges. If this is true, then Proposition 1.2 implies that Theorem 1.1 is tight up to the constant $c(s, t)$. In particular, such graphs are known to exist when $s = 2, 3$ (see e.g. [6]) and when $t \geq (s-1)! + 1$ (see [1]). Thus, in these cases Theorem 1.1 is tight up to the constant $c(s, t)$. Due to Kővári, Sós and Turán [12], it has long been known that there exists a constant $C'' = C''(s, t)$ for which every graph G with more than $C''|G|^{2-1/s}$ edges contains a copy of $K_{s,t}$, and we use a key lemma leading to this result in our proof (Lemma 2.5).

Lastly, we would like to discuss whether the conditions we consider (that is, the exclusion of some fixed bipartite graph) are the most natural in forcing a large clique subdivision. For example, considering the extremal example for the upper bound in (1), one might ask whether a graph with no small subgraphs which are almost as dense as the parent graph must contain a large clique subdivision. More precisely, for any graph G, let $f(G)$ be the number of vertices in the smallest subgraph H of G with $d(H) \geq d(G)/100$. It would be possible to use similar constructions to those given here to show that there exists some constant $c > 0$ such that every graph G contains a subdivision of a clique with at least $c \cdot \min\{\sqrt{f(G)/\log(f(G))}, d(G)\}$ vertices. We had hoped to improve this by removing the logarithmic term, thus generalising both Theorem 1.1 and the result of Bollobás and Thomason [4], and Komlós and Szemerédi [14], that $s(d) = O(d^2)$. This, however, is not possible, as shown by the following counterexample constructed for some large d with $1000|d$.

Counterexample: Take a random 1000-regular graph H with d vertices and blow up each vertex by an empty clique with $d/1000$ vertices. Let the resulting graph be G.

If G contains a subdivision of a clique with t core vertices in one of the empty cliques, then, considering the paths between these core vertices and the size of their neighbourhood, we must have that $\binom{t}{2} \leq d$. Therefore, any clique subdivision in G has at most $\sqrt{2d}$ core vertices in the blowup of each vertex in H. There are at most (say) $d^{1/3}$ vertices within a distance $(\log d)/100$ of any single vertex in H. Therefore, if a clique subdivision in G has more than $4d^{5/6}$ core vertices then at least half of the subdivided paths must contain at least $(\log d)/100$ vertices. If G contains a K_t-subdivision with $t \geq 4d^{5/6}$, then, as the paths in a K_t-subdivision are internally vertex-disjoint, $\binom{t}{2}/2 \leq 200|G|/\log d$. As G has $d^2/1000$ vertices, it therefore has no clique subdivision with more than $d/\sqrt{\log d}$ core vertices. Finally, note that with high probability we expect $f(G) = \Omega(|G|) = \Omega(d^2)$, demonstrating that for any fixed $c > 0$ a graph G may contain no subdivision of a clique with at least $c \cdot \min\{\sqrt{f(G)}, d(G)\}$ vertices.

In Section 2 we introduce the results from the literature we need for our proof, before dividing the proof of Theorem 1.1 into three cases and giving an overview of the rest of the paper.

Notation: Given a graph G, denote its average degree by $d(G)$. For a set of vertices $X \subseteq V(G)$, denote its *external neighbourhood* by $N(X) := \{u \notin X : uv \in E(G) \text{ for some } v \in X\}$. Furthermore, set $N_G^1(X) := N(X)$ and for each $i \geq 1$, define $N_G^{i+1}(X) := N(N_G^i(X))$ iteratively. Denote by $B_G^r(X)$ the ball of radius r around X, i.e. $B_G^r(X) = \cup_{i \leq r} N_G^i(X)$. Let $B_G(X) = B_G^1(X)$, and for each r and vertex v, let $N_G^r(\{v\}) = N_G^r(v)$ and $B_G^r(\{v\}) = B_G^r(v)$. We omit the index G if the underlying graph is clear from context. For $k \in \mathbb{N}$, denote by $X^{(k)}$ the family of all k-sets in X. A vertex in an H-subdivision is a *core* vertex if it corresponds to a vertex in H, i.e. it is not an internal vertex in any of the vertex-disjoint paths corresponding to the edges in H.

We will omit floor and ceiling signs when they are not crucial.

2 Preliminaries

2.1 Graph expansion

We need the following notion of graph expansion, which was introduced by Komlós and Szemerédi [13]. For $\varepsilon_1 > 0$ and $k > 0$, we first let $\varepsilon(x)$ be the function

$$\varepsilon(x) = \varepsilon(x, \varepsilon_1, k) := \begin{cases} 0 & \text{if } x < k/5 \\ \varepsilon_1/\log^2(15x/k) & \text{if } x \geq k/5, \end{cases} \tag{2}$$

where, when it is clear from context we will not write the dependency on ε_1 and k of $\varepsilon(x)$. Note that $\varepsilon(x) \cdot x$ is increasing for $x \geq k/2$.

Definition 2.1. (ε_1, k)**-expander:** A graph G is an (ε_1, k)-*expander* if $|N(X)| \geq \varepsilon(|X|) \cdot |X|$ for all subsets $X \subseteq V(G)$ of size $k/2 \leq |X| \leq |V(G)|/2$.

Komlós and Szemerédi [13, 14] showed that every graph G contains an (ε_1, k)-expander subgraph H that is almost as dense as G.

Theorem 2.2. *Let $k > 0$ and choose $\varepsilon_1 > 0$ sufficiently small (independently of k) so that $\varepsilon(x) = \varepsilon(x, \varepsilon_1, k)$ defined in (2) satisfies $\int_1^\infty \frac{\varepsilon(x)}{x} dx < \frac{1}{8}$. Then every graph G has an (ε_1, k)-expander subgraph H with $d(H) \geq d(G)/2$ and $\delta(H) \geq d(H)/2$.*

Note that the subgraph H in Theorem 2.2 might be much smaller than G. For example if G is a vertex-disjoint collection of many small cliques, then H could be one of those cliques.

We can find a relatively short path between two sufficiently large sets in an (ε_1, k)-expander, even after the deletion of an arbitrary, but smaller, set of vertices. This is formally captured in the following result (see Corollary 2.3 in [14]).

Lemma 2.3. *If G is an n-vertex (ε_1, k)-expander, then any two vertex sets, each of size at least $x \geq k$, are of distance at most*

$$diam := diam(n, \varepsilon_1, k) = \frac{2}{\varepsilon_1} \log^3(15n/k)$$

apart. This remains true even after deleting $x\varepsilon(x)/4$ arbitrary vertices from G.

4

2.2 Bipartite $K_{s,t}$-free graphs

It will simplify our constructions to work within a bipartite graph. This will be possible due to the following well-known result.

Lemma 2.4. *Within any graph G there is a bipartite subgraph H with $d(H) \geq d(G)/2$.*

We will use the $K_{s,t}$-free property of our graphs primarily through the following result of Kővári, Sós and Turán [12] (see also [3, IV, Lemma 9]).

Lemma 2.5. *Let $G = (A, B)$ be a bipartite graph that does not contain a copy of $K_{s,t}$ with t vertices in A and s vertices in B. Then*

$$|A|\binom{d(A)}{s} \leq t\binom{|B|}{s},$$

where $d(A) = \sum_{v \in A} \frac{d(v)}{|A|}$ is the average degree in G of the vertices in A.

We will use this lemma mainly through the following corollary.

Corollary 2.6. *Let $G = (A, B)$ be a bipartite graph that does not contain a copy of $K_{s,t}$ with t vertices in A and s vertices in B, and in which every vertex in A has at least δ neighbours in B. Then, $|B| \geq \delta|A|^{1/s}/et$.*

Proof. As $d(A) \geq \delta$, using Lemma 2.5 we have

$$|A|\left(\frac{\delta}{s}\right)^s \leq |A|\binom{d(A)}{s} \leq t\binom{|B|}{s} \leq t\left(\frac{e|B|}{s}\right)^s \leq \left(\frac{et|B|}{s}\right)^s.$$

Taking an appropriate root and rearranging gives the required inequality. $\square$

We also use more directly the following version of Kővári, Sós and Turán's theorem [12].

Theorem 2.7. *For each $s, t \geq 2$, and every $K_{s,t}$-free graph G, we have $2t|G| \geq (d(G))^{s/s-1}$.*

2.3 Division of the proof of Theorem 1.1 into cases

We will divide the proof of Theorem 1.1 into three main lemmas. Using Theorem 2.2, we will find within our graph G a subgraph which is almost as dense as G but also has some expansion properties. The first main lemma, Lemma 2.8, will either find the required subdivision or a large dense subgraph which retains some useful expansion properties while additionally having a small maximum degree.

Thus, this reduces the problem to finding a subdivision in a graph with a certain expansion property and a small maximum degree. The construction we use differs according to the density of the subgraph. The dense case is covered by Lemma 2.9, while the sparse case is covered by Lemma 2.10.

The first of these lemmas is adapted and generalised from a lemma in [2].

5

Lemma 2.8. *For any $0 < \varepsilon_1 < 1$, and integers $t \geq s \geq 2$, there exists $\varepsilon_2' := \varepsilon_2'(\varepsilon_1, t) > 0$ such that for any $0 < \varepsilon_2 \leq \varepsilon_2'$ and $K \geq 100/\varepsilon_2$, there exists some $c_0 := c_0(\varepsilon_1, \varepsilon_2, s, K) > 0$ for which the following holds for any d. Let G be a bipartite, $K_{s,t}$-free, $(\varepsilon_1, \varepsilon_2 d^{s/(s-1)})$-expander with $\delta(G) \geq d/8$. Then either G contains a subdivision of a clique of order $c_0 d^{\frac{1}{2}\frac{s}{s-1}}$ or a subgraph H with $\delta(H) \geq d/16$, $|H| \geq Kd^{s/(s-1)}$ and $\Delta(H) \leq d\log^{10s}(|H|/d^{s/(s-1)})$ which is an $(\varepsilon_1/2, \varepsilon_2 d^{s/(s-1)})$-expander.*

When the subgraph found using Lemma 2.8 is dense, we will use the following lemma.

Lemma 2.9. *Let $0 < \varepsilon_1, \varepsilon_2 < 1$ and let $t \geq s \geq 2$ be integers. Then, for sufficiently large $K := K(\varepsilon_1, \varepsilon_2, s, t) > 0$, the following holds for any integers n and d with $n \geq Kd^{s/(s-1)}$ and $d \geq \log^{20s} n$. If G is an n-vertex bipartite, $K_{s,t}$-free, $(\varepsilon_1, \varepsilon_2 d^{s/(s-1)})$-expander with $\delta(G) \geq d/16$ and $\Delta(G) \leq d\log^{10s}(n/d^{s/(s-1)})$, then G contains a $K_{\ell/10^4 t}$-subdivision for $\ell = d^{\frac{1}{2}\frac{s}{s-1}}$.*

When the subgraph found using Lemma 2.8 is sparse, we will in fact find a subdivision of a clique of order linear in the average degree of the subgraph, as shown by the following lemma.

Lemma 2.10. *Let $0 < \varepsilon_1 < 1$, $0 < \varepsilon_2 \leq 1/10^5 t$, and let $t \geq s \geq 2$ be integers. Then there is some $c_1 := c_1(\varepsilon_1, \varepsilon_2, s, t) > 0$ for which the following holds for any integers n and d with $d \leq \log^{20s} n$. If G is an n-vertex bipartite, $K_{s,t}$-free, $(\varepsilon_1, \varepsilon_2 d^{s/(s-1)})$-expander with $\delta(G) \geq d/16$ and $\Delta(G) \leq d\log^{10s} n$, then G contains a $K_{c_1 d}$-subdivision.*

Due to the number of different constants involved, we will carefully show, as follows, that Theorem 1.1 follows from Lemmas 2.8, 2.9 and 2.10.

Proof of Theorem 1.1. We first fix all the parameters we need, as follows.

- Let ε_0 be a constant with the property in Theorem 2.2, and $\varepsilon_1 = \min\{\varepsilon_0/2, 1/8\}$.

- Let $t \geq s \geq 2$ be integers.

- Let $\varepsilon_2 = \min\{\varepsilon_2'(2\varepsilon_1, t), 1/10^5 t\}$, where ε_2' is a constant with the property in Lemma 2.8.

- Let $K' = \max\{K(\varepsilon_1, \varepsilon_2, s, t), 100/\varepsilon_2\}$, where K is a constant with the property in Lemma 2.9.

- Finally, let $c = \min\{1/10^5 t, c_0(2\varepsilon_1, \varepsilon_2, s, K'), c_1(\varepsilon_1, \varepsilon_2, s, t)\}$, where c_0 and c_1 are constants with the properties in Lemmas 2.8 and 2.10 respectively.

We will prove Theorem 1.1 with the constant c. Let G then be a $K_{s,t}$-free graph with average degree $d > 0$, and let $\ell = d^{\frac{1}{2}\frac{s}{s-1}}$. We seek a $K_{c\ell}$-subdivision in G.

By Lemma 2.4, G contains a bipartite subgraph G_1 with $d(G_1) \geq d/2$. By the choice of $2\varepsilon_1 \leq \varepsilon_0$ with Theorem 2.2, there is a subgraph G_2 of G_1 with $\delta(G_2) \geq d(G_2)/2 \geq d(G_1)/4 \geq d/8$ which is a $(2\varepsilon_1, \varepsilon_2 d^{s/(s-1)})$-expander.

By Lemma 2.8, if G_2 does not contain a $K_{c\ell}$-subdivision, then we can find a subgraph G_3 of G_2 with $\delta(G_3) \geq d/16$, $|G_3| \geq K'd^{s/(s-1)}$ and $\Delta(G_3) \leq d\log^{10s}(|G_3|/d^{s/(s-1)})$ which

is an $(\varepsilon_1, \varepsilon_2 d^{s/(s-1)})$-expander. If $d \geq \log^{20s} |G_3|$, then by the choice of ε_2 and K' with Lemma 2.9, G_3 contains a $K_{c\ell}$-subdivision. If $d \leq \log^{20s} |G_3|$, then by the choice of ε_2 and c with Lemma 2.10, G_3 contains a $K_{c\ell}$-subdivision. Therefore, in all cases, G contains a $K_{c\ell}$-subdivision. $\qquad\square$

Our remaining task is therefore to prove Lemmas 2.8, 2.9 and 2.10, which we do in Sections 3, 4 and 5 respectively. In the rest of this section we will sketch the constructions we will use. In Section 6 we will make some concluding remarks.

2.4   Sketch of constructions

Here we sketch an overview of the three constructions we will use to prove Lemmas 2.8, 2.9 and 2.10. For simplicity, we will assume that $s = t = 2$, and say we wish to construct a K_{cd}-subdivision in an n-vertex C_4-free $(\varepsilon_1, \varepsilon_2 d^2)$-expander G with minimum degree d, for some small constants c, ε_1 and ε_2. The main variables we have are n and d, but we also consider the following important variable depending on them:

$$diam := (2/\varepsilon_1) \log^3(15n/\varepsilon_2 d^2).$$

By Lemma 2.3, in our graph G we will be able to find a path with length at most $diam$ between any two vertex sets A and B if $|A|, |B| \geq \varepsilon_2 d^2$. Furthermore, this will remain true if we delete any vertex set W from the graph, as long as $|A|, |B| \geq |W| \cdot diam$. In particular, we will use that if we have found paths using at most $8c^2 d^2 diam$ vertices, then we can avoid these vertices while finding a new path with length at most $diam$ between any two vertex sets with size at least $d^2(diam)^2$.

We will take in turn the additional assumptions that i) G has many vertices with degree at least $d(diam)^3$, ii) G has maximum degree at most $d(diam)^3$, $d \geq \log^{40} n$ and $n \geq d^2(diam)^8$, and iii) G has maximum degree at most $d(diam)^3$ and $d \leq \log^{40} n$. These assumptions take us close to the conditions of Lemmas 2.8, 2.9 and 2.10 respectively, and permit us to sketch the constructions in a slightly simplified situation.

i) Lemma 2.8 construction. We assume that G has many vertices with degree at least $d(diam)^3$. This will allow us to find $2cd$ disjoint $4d(diam)^2$-stars in G. We choose the root vertices of these stars as our candidate core vertices. We then greedily connect as many different pairs of potential core vertices as possible by paths with length at most $4diam$ whose internal vertices do not contain any potential core vertices and are disjoint between paths. When this process is finished there will trivially be at most $8c^2 d^2 diam$ vertices in the paths found. Therefore, there will be at least cd *nice* core vertices v_i among the candidate core vertices which have a set U_i of at least $2d(diam)^2$ neighbours not in any of the paths we found (as the original stars were disjoint). The graph G has minimum degree at least d, and no two vertices in U_i share a neighbour other than v_i (as G is C_4-free). Therefore, $U_i \subset N(v_i)$ will have a neighbourhood of size at least $d^2(diam)^2$, large enough to find a path connecting it to any other such set U_j while avoiding all the vertices in previously found paths (by Lemma 2.3). Therefore, in the process we must have found a path between any two nice core vertices, which gives a K_{cd}-subdivision, as required.

ii) Lemma 2.9 construction. We assume that G has maximum degree at most $d(diam)^3$, $d \geq \log^{40} n$ and $n \geq d^2(diam)^8$. The main consequence of the low maximum degree is that it allows us to iteratively find structures with up to $d^2(diam)^4$ vertices in total. Indeed, G has at least $nd/2$ edges, and $n/(diam)^4$ vertices are incident to at most $nd/diam$ edges. Thus, deleting at most $n/(diam)^4 \geq d^2(diam)^4$ vertices gives a subgraph almost as dense as G in which we can find further structure. In the following sketch we describe the construction of $2cd$ *units* which each have at most $d(diam)^4$ vertices.

Broadly speaking, our construction for Lemma 2.9 is similar to that used for Lemma 2.8, but with each potential core vertex replaced disjointly by a structure known as a 'unit'. Instead of a single vertex which has $4d(diam)^2$ neighbours, we use $2cd$ *connecting* vertices each with their own $4(diam)^3$ *assigned* neighbours, along with disjoint paths with length at most $2diam$ linking the connecting vertices back to a potential core vertex v_i. This structure will further have the property that any set of at least cd connecting vertices and at least $2(diam)^3$ of each of their assigned neighbours have together at least $d^2(diam)^2$ neighbours in the graph.

We will then expand and connect up different pairs of these units using disjoint paths which go to a connecting vertex and then down to the corresponding core vertex. By averaging, when this is done there will be at least cd *nice* units where at least cd connecting vertices, their paths, and at least $2(diam)^3$ of their assigned neighbours, are untouched by the paths we found. By expanding the untouched connecting vertices of two nice units we would be able to find disjointly another path between them, so we must have found a path between any two nice units. This will give a K_{cd}-subdivision, as required.

iii) Lemma 2.10 construction. We assume that G has maximum degree at most $d(diam)^3$ and $d \leq \log^{40} n$. We start with core vertices that are chosen to be a large graph distance apart in G, and greedily find internally vertex-disjoint paths between the core vertices under certain conditions. When we go to find a new path between two core vertices, v and w say, we begin by expanding around v (and similarly around w) in three stages - out to distance $(\log \log n)^5$, $\log n/100 \log \log n$ and $diam$ respectively. At first we consider successive neighbourhoods expanding out from v while avoiding the vertices in the paths we have found which connect other core vertices to v. In this expansion we pick up enough vertices that in the second stage we can expand while avoiding all the vertices in the paths we have found so far (such vertices will not previously be encountered as each path will not come near core vertices which are not one of its endpoints). After this expansion we will have picked up enough vertices that we can avoid all these vertices as well as any other vertices that are close to the other core vertices (such vertices will not previously be encountered as the other core vertices are far from v). After performing a similar expansion around w this allows us to find a short path connecting v to w that does not come close to the other core vertices. In this manner, we will be able to connect up all the different pairs of core vertices and find a K_{cd}-subdivision.

In all this, the maximum degree condition will be critical - in particular it both guarantees that there are not too many vertices close to each core vertex, so that such vertices can be avoided in the third expansion, and guarantees that there exist enough vertices which are

pairwise far enough apart to use as core vertices.

3 Constructing subdivisions when many vertices have high degree

In this section, we prove Lemma 2.8. That is, in a $K_{s,t}$-free graph G with a certain expansion condition and a minimum degree condition, we either find the subdivision we seek or we may assume an additional maximum degree condition. Essentially, if there are few vertices of high degree, then we can delete these vertices and obtain a subgraph of G with almost the same expansion condition and minimum degree condition, but which additionally has a small maximum degree (see Claim 3.1). If to the contrary there are many vertices of high degree, we can then construct the desired subdivision using some of these high-degree vertices as core vertices. In this construction, we first choose appropriate sets $S_1(v)$ of neighbours of some *selected* high-degree vertices v (see Claim 3.2), before using that G is $K_{s,t}$-free to conclude that $N(S_1(v))$ is large, even when some of the vertices in $S_1(v)$ are deleted. We then connect one-by-one the pairs of the high-degree vertices v through the matching sets $N(S_1(v))$ using a short path (which will exist by Lemma 2.3), deleting any vertices we use from other sets $S_1(w)$ to ensure the paths we find are internally disjoint (see Algorithm P). We do not expect to be able to connect all the pairs of selected high-degree vertices in this manner, but we will show that at the end of the process most of the sets $S_1(v)$ will still be large (see Claim 3.3) and that we will have found a path between any pair of vertices v with a large matching set $S_1(v)$ (see Claim 3.4). These paths will form the desired subdivision.

Proof of Lemma 2.8. Given $0 < \varepsilon_1 < 1$ and integers $t \geq s \geq 2$, take

$$\varepsilon_2' = \varepsilon_2'(\varepsilon_1, t) := \frac{1}{10t} \min \left\{ e^{-100et}, e^{-100/\varepsilon_1} \right\}. \tag{3}$$

Given further constants $0 < \varepsilon_2 \leq \varepsilon_2'$ and $K \geq 100/\varepsilon_2$, let

$$d_0 = d_0(\varepsilon_1, \varepsilon_2, s) := \left(\frac{100}{\varepsilon_1 \varepsilon_2} \right)^s \quad \text{and} \quad c_0 = c_0(\varepsilon_1, \varepsilon_2, s, K) := \min \left\{ \frac{1}{10 \log^{10s}(30K/\varepsilon_2)}, \frac{1}{d_0} \right\}. \tag{4}$$

We will show that the lemma holds with c_0. Let $d > 0$ and $n \in \mathbb{N}$. Let G be an n-vertex bipartite $K_{s,t}$-free $(\varepsilon_1, \varepsilon_2 d^{s/s-1})$-expander graph with $\delta(G) \geq d/8$. Note that if $d \leq d_0$, then, as G contains a K_1-subdivision as $\delta(G) \geq d/8 > 0$, and $c_0 d^{\frac{1}{2}\frac{s}{s-1}} \leq 1$, we have the required subdivision. We can thus assume that $d \geq d_0$.

Let

$$\Delta := \max \left\{ d/8, \quad c_0 d \left(\log \frac{15n}{\varepsilon_2 d^{s/(s-1)}} \right)^{10s} \right\}, \tag{5}$$

and let $L \subseteq V(G)$ be the set of vertices in G with degree at least Δ.

Claim 3.1. *If $|L| \le d/16$, then $H := G - L$ is an $(\varepsilon_1/2, \varepsilon_2 d^{s/(s-1)})$-expander satisfying $\delta(H) \ge d/16$, $|H| \ge K d^{s/(s-1)}$ and $\Delta(H) \le d \log^{10s}(|H|/d^{s/(s-1)})$.*

Proof of Claim 3.1. Suppose that $|L| \le d/16$. As $|L| < \delta(G) < |G|$, we know $L \ne V(G)$, and therefore, by the definition of L, $\Delta > \delta(G) \ge d/8$. Thus, from (5), we have

$$\Delta = c_0 d \left(\log \frac{15n}{\varepsilon_2 d^{s/(s-1)}} \right)^{10s} \ge \frac{d}{8} \quad \Rightarrow \quad \left(\log \frac{15n}{\varepsilon_2 d^{s/(s-1)}} \right)^{10s} \ge \frac{1}{8c_0} \overset{(4)}{\ge} \left(\log \frac{30K}{\varepsilon_2} \right)^{10s}$$

$$\Rightarrow \quad n \ge 2K d^{s/(s-1)}. \tag{6}$$

As $n > \delta(G) \ge d/8$, we have $|H| \ge n - |L| \ge n - d/16 \ge n/2$, and thus $|H| \ge K d^{s/(s-1)}$ by (6), as required. Furthermore, $\delta(H) \ge \delta(G) - |L| \ge d/16$.

Using that $|H|/d^{s/(s-1)} \ge K \ge 100/\varepsilon_2$ and $|H| \ge n/2$, the maximum degree of H is at most

$$\Delta \overset{(5)}{\le} c_0 d \cdot \log^{10s} \left(\frac{30|H|}{\varepsilon_2 d^{s/(s-1)}} \right) \overset{(4)}{\le} d \cdot \left(\frac{\log \left(\frac{30}{\varepsilon_2} \cdot \frac{|H|}{d^{s/(s-1)}} \right)}{\log(30K/\varepsilon_2)} \right)^{10s}$$

$$\le \quad d \cdot \left(\frac{2 \cdot \log \left(\frac{|H|}{d^{s/(s-1)}} \right)}{\log(30K/\varepsilon_2)} \right)^{10s} \overset{(3)}{\le} d \log^{10s} \left(\frac{|H|}{d^{s/(s-1)}} \right).$$

To finish the proof of the claim it is left to show that H is an $(\varepsilon_1/2, \varepsilon_2 d^{s/(s-1)})$-expander. Set $k := \varepsilon_2 d^{s/(s-1)}$. Since G is an (ε_1, k)-expander and $\varepsilon(x) \cdot x$ is increasing when $x \ge k/2$, for any set X in H of size $x \ge k/2$ with $x \le |H|/2 \le |G|/2$, we have

$$|N_G(X)| \ge x \cdot \varepsilon(x, \varepsilon_1, k) \ge \frac{k}{2} \cdot \varepsilon \left(\frac{k}{2}, \varepsilon_1, k \right) = \frac{\varepsilon_2 d^{s/(s-1)}}{2} \cdot \frac{\varepsilon_1}{\log^2(15/2)}$$

$$\ge \frac{\varepsilon_2 \varepsilon_1}{100} \cdot d^{s/(s-1)} \ge d \ge 2|L|,$$

where ε is defined in (2) and the second last inequality follows as $d \ge d_0 = \left(\frac{100}{\varepsilon_1 \varepsilon_2} \right)^s$. Hence, $|N_H(X)| \ge \frac{1}{2}|N_G(X)| \ge \frac{1}{2} x \cdot \varepsilon(x, \varepsilon_1, \varepsilon_2 d^{s/(s-1)}) = x \cdot \varepsilon(x, \varepsilon_1/2, \varepsilon_2 d^{s/(s-1)})$, as required. $\square$

Therefore, if $|L| \le d/16$ then we have can find a subgraph H that would satisfy the lemma. Thus, we may assume that $|L| \ge d/16$. Let $\ell = c_0 d^{\frac{1}{2} \frac{s}{s-1}}$. We will now find in G a K_ℓ-subdivision, completing the proof of the lemma.

By Theorem 2.7, and since $\delta(G) \ge d/8$, we have $n/d^{s/(s-1)} \ge 1/128t$. Using this, let

$$m := \log \frac{15n}{\varepsilon_2 d^{s/(s-1)}} \ge \log \frac{15}{128t \cdot \varepsilon_2} \overset{(3)}{\ge} \max \left\{ 100 et, \frac{100}{\varepsilon_1} \right\}. \tag{7}$$

As $K \ge 100/\varepsilon_2$, (3) implies that $c_0 \le 1/64$, so that $|L| \ge d/16 \ge 4\ell$. Therefore, as G is bipartite, we can find a subset $L' \subseteq L$ of 2ℓ vertices in the same partite set. We will use some vertices from L' as the core vertices of our K_ℓ-subdivision.

10

Claim 3.2. *For each $v \in L'$, we can pick a subset $S_1(v) \subseteq N(v)$ such that*
(i) $|S_1(v)| = \Delta/2$, and
(ii) each vertex $u \in S_1(v)$ is adjacent to at most $c_0 d/\ell$ vertices in L'.

Proof. Fixing $v \in L'$, let $A = \{w \in N(v) : |N(w) \cap L'| \geq d^{\frac{1}{2}\frac{s-2}{s-1}} + 1\}$. Note that $d^{\frac{1}{2}\frac{s-2}{s-1}} = c_0 d/\ell$, so that if $w \in N(v) \setminus A$ then, by the definition of A, w has at most $c_0 d/\ell$ neighbours in L'. Now, there is no copy of $K_{s-1,t}$ with t vertices in A and $s-1$ vertices in $L' \setminus \{v\}$ since such a copy of $K_{s-1,t}$ together with v would form a copy of $K_{s,t}$. Therefore, using Corollary 2.6 with $\delta = d^{\frac{1}{2}\frac{s-2}{s-1}}$ and $B = L' \setminus \{v\}$, we have

$$d^{\frac{1}{2}\frac{s-2}{s-1}}|A|^{1/(s-1)}/et \leq |L' \setminus \{v\}| \leq 2\ell = 2c_0 d^{\frac{1}{2}\frac{s}{s-1}}.$$

Hence, using (7) and (5), we have $|A| \leq d(2c_0 et)^{(s-1)} \leq c_0^{s-1} dm^s/2 \leq \Delta/2$. Thus, $|N(v) \setminus A| \geq \Delta/2$, and we may pick a set $S_1(v) \subset N(v) \setminus A$ with size $\Delta/2$. Picking such a set for each $v \in L'$ satisfies the claim. $\square$

Algorithm P: Take sets $S_1(v)$, $v \in L'$, with the properties in Claim 3.2. Connect different pairs of core vertices $v \in L'$ greedily through these sets under the following rules:

P1 At each step, build a path of length at most $2m^4$ connecting a new pair of core vertices, avoiding vertices used in previous connections and in L'.

P2 During the whole process, discard a core vertex $v \in L'$ if more than $\Delta/4$ vertices in $S_1(v)$ are used in previous connections.

This process will result in a K_ℓ-subdivision, as shown by the following two claims.

Claim 3.3. *At the end of the process at least ℓ core vertices remain undiscarded.*

Claim 3.4. *At the end of the process all possible paths have been created between the remaining core vertices.*

Proof of Claim 3.3. By **P1**, at most $\binom{2\ell}{2} \cdot 2m^4 \leq 4\ell^2 m^4$ vertices have been used to create connections in the entire process. By **P2**, if a core vertex $v \in L'$ has been discarded then at least $\Delta/4$ vertices in $S_1(v)$ have been used in creating connections. Note that by Claim 3.2 (ii), each vertex is in at most $c_0 d/\ell$ of the sets $S_1(v)$, $v \in L'$. Hence, at the end of process the number of discarded core vertices is at most

$$\frac{4\ell^2 m^4 \cdot c_0 d/\ell}{\Delta/4} \overset{(5),(7)}{\leq} \frac{16\ell m^4 c_0 d}{c_0 dm^{10s}} \overset{(7)}{\leq} \ell.$$

Thus, at least $|L'| - \ell = \ell$ core vertices remain at the end of the process. $\square$

Proof of Claim 3.4. Suppose that, having perhaps connected some pairs of core vertices subject to **P1** and **P2**, the current pair of undiscarded core vertices to be connected is $\{v, v'\}$. Let $A \subseteq S_1(v)$ be the set of vertices in $S_1(v)$ not used in previous connections. Since v has not been discarded, by **P2**, $|A| \geq |S_1(v)| - \Delta/4 = \Delta/4$. Let $B := N(A) \setminus \{v\} \subseteq N(S_1(v)) \setminus \{v\}$ and note that $G[A, B]$ does not contain a copy of $K_{s-1,t}$ with t vertices in A and $s-1$ vertices in B. Then, by Corollary 2.6, we have

$$|B| \geq \frac{1}{et}\left(\frac{d}{8} - 1\right)\left(\frac{\Delta}{4}\right)^{1/(s-1)} \geq \frac{d}{64et} \cdot \Delta^{1/(s-1)},$$

and hence, by (5), (7) and (3), we have

$$|B| \geq \frac{d}{64et} \cdot c_0 d^{\frac{1}{s-1}} m^{10} \geq \ell^2 m^9 \quad \text{and} \quad |B| \geq \frac{d}{64et} \cdot \left(\frac{d}{8}\right)^{\frac{1}{s-1}} \geq 2\varepsilon_2 d^{s/(s-1)}. \tag{8}$$

To make connections according to **P1**, we have to exclude vertices from L' as well as used vertices from B. Therefore, the number of available vertices in B is, using (8), at least

$$y := |B| - |L'| - 4\ell^2 m^4 \geq |B|/2 \geq \max\{\ell^2 m^9/2, \varepsilon_2 d^{s/(s-1)}\}.$$

Letting $k = \varepsilon_2 d^{s/(s-1)}$, by (7), we have that $\varepsilon(n, \varepsilon_1, k) = \varepsilon_1/m^2 \geq 1/m^3$, where ε is defined in (2). Since $\varepsilon(x, \varepsilon_1, k)$ is decreasing when $x \geq k$, we have

$$\frac{1}{4} \cdot y \cdot \varepsilon(y, \varepsilon_1, k) \geq \frac{y}{4} \cdot \varepsilon(n, \varepsilon_1, k) \geq \frac{\ell^2 m^9/2}{4} \cdot \frac{1}{m^3} > 5\ell^2 m^4.$$

Similarly, if A' is the set of available vertices in $S_1(v')$ and y' is the number of available vertices in $N(A') \setminus \{v'\}$, then $\frac{1}{4} \cdot y' \cdot \varepsilon(y', \varepsilon_1, k) > 5\ell^2 m^4$. Recall that the number of vertices to avoid in **P1** is at most $|L'| + 4\ell^2 m^4 \leq 5\ell^2 m^4$. Thus, by Lemma 2.3, there is a path of length at most $2\log^3(15n/\varepsilon_2 d^{s/(s-1)})/\varepsilon_1 \leq m^4$ between the set of available vertices in $N(A)$ and $N(A')$, which yields a v, v'-path of length at most $m^4 + 4 \leq 2m^4$, as desired. $\square$

Thus, the process creates a K_ℓ-subdivision. $\square$

4 Constructing subdivisions in a dense expander

In this section, we prove Lemma 2.9, which covers the case for Theorem 1.1 where the subgraph that may be found using Lemma 2.8 is dense. Here this means that $d \geq \log^{20s} n$. We need the following two definitions, which are depicted in Figure 1.

(h_1, h_2)-**Hub.** Given integers $h_1, h_2 > 0$, an (h_1, h_2)-*hub* is a graph consisting of a center vertex u, a set $S_1(u) \subseteq N(u)$ of size h_1, and pairwise disjoint sets $S_1(z) \subset N(z)$ of size h_2 for each $z \in S_1(u)$. Denote by $H(u)$ a hub with center vertex u and write $B_1(u) = \{u\} \cup S_1(u)$ and $S_2(u) = \bigcup_{z \in S_1(u)} S_1(z)$. For any $z \in S_1(u)$, write $B_1(z) = \{z\} \cup S_1(z)$.
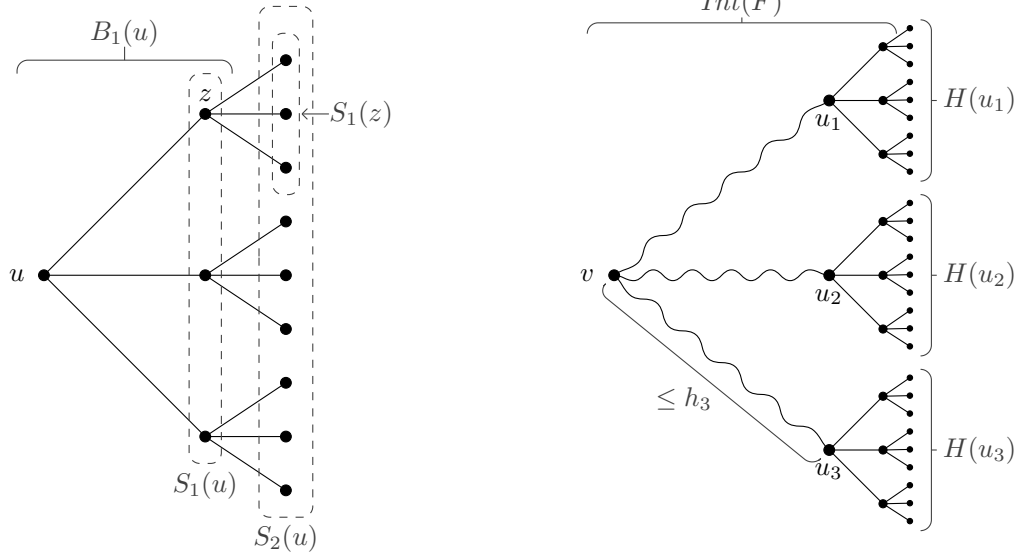
Figure 1: An (h_1, h_2)-hub $H(u)$ on the left with $h_1 = h_2 = 3$ and an (h_0, h_1, h_2, h_3)-unit F on the right with $h_0 = h_1 = h_2 = 3$. The straight lines represent edges, while the wavy lines represent paths (here with length at most h_3).

(h_0, h_1, h_2, h_3)-**Unit.** Given integers $h_0, h_1, h_2, h_3 > 0$, an (h_0, h_1, h_2, h_3)-*unit* F is a graph consisting of a core vertex v, h_0 vertex-disjoint (h_1, h_2)-hubs $H(u_1), \ldots, H(u_{h_0})$ and pairwise disjoint v, u_j-paths of length at most h_3. By the *exterior* of the unit, denoted by $\mathrm{Ext}(F)$, we mean $\bigcup_{j=1}^{h_0} S_2(u_j)$. Denote by $\mathrm{Int}(F) := V(F) \setminus \mathrm{Ext}(F)$ the *interior* of the unit.

An outline of the proof of Lemma 2.9 is as follows. We need to construct a collection of units whose interiors are pairwise disjoint (see Section 4.2). We do so iteratively by starting with many vertex-disjoint hubs (found in Section 4.1), and connecting them (using Algorithm Q in Section 4.2) in such a way that one of the hubs will be linked to many others (see Claim 4.3), forming the desired unit. Once this is done, we then connect pairs of units through their exteriors while avoiding all the previously used vertices and the vertices in all these v, u_j-paths in the units (see Section 4.3, and in particular Algorithm R), so that their centre vertices become the core vertices of the required subdivision.

Our use of units is inspired by a similar structure introduced in [19]. Roughly speaking, we face two challenges when attempting to connect some potential core vertices to get a subdivision. Firstly, the paths from each core vertex v must be disjoint, a challenge near v, and secondly the paths must connect v to all the other core vertices. Essentially, through the use of units we compartmentalise these two problems and deal with them separately. In finding the units, we find many paths emerging from a potential core vertex, where the paths end in hubs to help us extend them further. Once we have found the units, we then concern ourselves with extending some of these paths to connect different units until we have formed a subdivision.

4.1 Constructing hubs

In order to construct units, we will first find many vertex-disjoint hubs. These are shown to exist by the following lemma.

Lemma 4.1. *For each $2 \le s \le t$ there is some d_0 such that the following holds for each $d \ge d_0$. Let $h_1, h_2 \le d^{\frac{1}{2}\frac{s}{s-1}}/400t$ be integers and let G be an n-vertex $K_{s,t}$-free bipartite graph with $\delta(G) \ge d/16$. Given any set $W \subseteq V(G)$ with size at most $nd/128\Delta(G)$, there are in $G - W$ pairwise disjoint (h_1, h_2)-hubs of total size at most $nd/128\Delta(G)$.*

Proof. Let $c = 1/400t$ and $\ell = d^{\frac{1}{2}\frac{s}{s-1}}$. It suffices to show that given any set $W' \subseteq V(G)$ with size $nd/64\Delta(G)$, there is a $(c\ell, c\ell)$-hub in $G - W'$. Note that

$$d(G - W') \ge \frac{n \cdot \delta(G) - |W'| \cdot 2\Delta(G)}{n - |W'|} \ge \frac{nd/16 - nd/32}{n} \ge \frac{d}{32}.$$

There is then a subgraph $G' \subseteq G - W'$ with $\delta(G') \ge d(G - W')/2 \ge d/64$. We will find a $(c\ell, c\ell)$-hub in G'.

Let v be an arbitrary vertex in G' and let $A = N(v)$, so that $|A| \ge d/64$. It suffices to find $c\ell$ disjoint stars in $G' - \{v\}$, each with $c\ell$ leaves and its center in A, since such stars together with v form a $(c\ell, c\ell)$-hub. Let $A' \subset A$ be a maximal subset such that we can find $|A'|$ disjoint stars in $G' - \{v\}$, each with $c\ell$ leaves and its center in A'. Let B be the union of the leaves of such a collection of stars.

If $|A'| \ge c\ell$, then we are done, so suppose $|A'| < c\ell$, so that $|A \setminus A'| \ge d/64 - c\ell \ge d/128$. Each vertex in $A \setminus A'$ has fewer than $c\ell$ neighbours in $V(G') \setminus (B \cup \{v\})$, otherwise that vertex could be added to A' to reach a contradiction. Therefore, as $\delta(G') \ge d/64$, each vertex in $A \setminus A'$ has at least $d/64 - c\ell - 1 \ge d/128$ neighbours in B.

Note that $G' - \{v\}$ does not contain a copy of $K_{s-1,t}$ with t vertices in $A \setminus A'$ and $s - 1$ vertices in B, since otherwise such a copy together with v forms a copy of $K_{s,t}$ in G'. Therefore, by Lemma 2.6, we have

$$|B| \ge (d/128)|A \setminus A'|^{1/(s-1)}/et \ge (d/128)^{s/(s-1)}/et \ge \ell^2/10^5 t \ge c^2\ell^2.$$

As $|B| = |A'|c\ell$, we thus have $|A'| \ge c\ell$, a contradiction. $\square$

4.2 Constructing units from hubs

We will now expand hubs to connect them into a unit.

Lemma 4.2. *For each $0 < \varepsilon_1, \varepsilon_2 < 1$ and integers $2 \le s \le t$, there exists $K := K(\varepsilon_1, \varepsilon_2, s, t)$ such that the following holds for all n and d with $d \ge \log^{20s} n$ and $n \ge Kd^{s/(s-1)}$. Let $c = 1/800t$, $\ell = d^{\frac{1}{2}\frac{s}{s-1}}$ and $m = \log^{2s}(n/\ell^2)$, and suppose G is a bipartite n-vertex $K_{s,t}$-free $(\varepsilon_1, \varepsilon_2\ell^2)$-expander with $\delta(G) \ge d/16$ and $\Delta(G) \le dm^5$. Then G contains ℓ $(c\ell, m^2, c\ell, 2m)$-units $F_1, \ldots, F_\ell$ with core vertices $v_1, \ldots, v_\ell$ so that the interiors of all the units F_i, that is, the sets $\text{Int}(F_i)$, are pairwise disjoint.*

14

Proof. We will construct the units iteratively. Let W be the set of vertices in the interiors of the $(c\ell, m^2, c\ell, 2m)$-units constructed so far. The interior of each such unit has size at most $c\ell \cdot (2m + m^2) \le 2c\ell m^2$. Thus, $|W| \le 2c\ell^2 m^2$.

Since $n/\ell^2 \ge K$, for sufficiently large K, we have that

$$\frac{n}{\ell^2} \ge 128 \log^{20s}\left(\frac{n}{\ell^2}\right) = 128 m^{10}. \tag{9}$$

Note that $d \ge \log^{20s} n \ge m^{10}$ and, consequently, since $m \ge \log^{2s} K$, for sufficiently large K,

$$c\ell = cd^{\frac{1}{2}\frac{s}{s-1}} > cd^{1/2} \ge 8m^4. \tag{10}$$

Therefore, by Lemma 4.1, we can find in $G - W$ vertex-disjoint hubs $H(w_1), \ldots, H(w_{m^3})$ and $H(u_1), \ldots, H(u_{\ell m^3})$ such that each $H(w_i)$, $1 \le i \le m^3$, is a $(2c\ell, 2c\ell)$-hub and each $H(u_j)$, $1 \le j \le \ell m^3$, is a $(2m^2, 2c\ell)$-hub. Indeed, this is possible since $2c \le 1/400t$, by (10) we have that $2m^2 \le c\ell$,

$$|W| \le 2c\ell^2 m^2 \overset{(9)}{\le} \frac{n}{128 m^5} \le \frac{nd}{128\Delta(G)},$$

and the total number of vertices in all these hubs is at most

$$2(2c\ell)^2 \cdot m^3 + 2(2c\ell)(2m^2) \cdot \ell m^3 \le \ell^2 m^5 \overset{(9)}{\le} \frac{nd}{128\Delta(G)}.$$

We will construct a unit using some vertex w_i as the core vertex and some subgraphs of the hubs $H(u_j)$ as the hubs in the unit.

Algorithm Q: Greedily, connect as many different pairs of vertices $\{w_i, u_j\}$ as possible under the following rules.

Q1 Each connection uses a path of length at most $2m$, avoiding vertices used in previous connections.

Q2 Each connection avoids using vertices in any set $B_1(w_{i'})$ or $B_1(u_{j'})$, except for at most two vertices each in $B_1(w_i)$ and $B_1(u_j)$ when $\{w_i, u_j\}$ is the pair of vertices being connected.

Claim 4.3. *There exists a vertex w_i connected to at least $c\ell$ vertices u_j.*

Proof of Claim 4.3. Suppose to the contrary that each vertex w_i is connected to fewer than $c\ell$ vertices u_j at the end of the process. Let P be the set of all interior vertices in all the connections. Then, by **Q1**,

$$|P| \le 2m \cdot m^3 \cdot c\ell = 2m^4 c\ell. \tag{11}$$

Let W' contain the vertices in P as well as all the vertices in each set $B_1(u_j)$ if u_j has been connected to at least one of the vertices w_i. As there are at most $m^3 \cdot c\ell$ such vertices u_j, using (10) we have

$$|W'| \le |P| + m^3 c\ell \cdot (2m^2 + 1) \overset{(11)}{\le} 2m^4 c\ell + 4m^5 c\ell \overset{(10)}{\le} c^2 \ell^2 m. \tag{12}$$

15

For each $1 \leq i \leq m^3$, let $T_i = B_1(w_i) \setminus W'$, so that, by **Q2**, $|T_i| \geq c\ell$. As the graphs $H(w_i)$ are vertex-disjoint $(2c\ell, 2c\ell)$-hubs, we have $|N(\cup_i T_i)| \geq 2c\ell \cdot c\ell \cdot m^3$, and hence, we have

$$|N(\cup_i T_i) \setminus W'| \geq 2c\ell \cdot c\ell \cdot m^3 - |W'| \overset{(12)}{\geq} c^2\ell^2 m^3.$$

At least $\ell m^3 - m^3 \cdot c\ell \geq \ell m^3/2$ vertices u_i have not been involved in the connections made. Call these vertices $u'_1, \ldots, u'_p$, where $p \geq \ell m^3/2$. By **Q2**, the set W' is disjoint from $\cup_i B_1(u'_i)$, and note that, as $H(u'_i)$ are disjoint $(2m^2, 2c\ell)$-hubs, we have

$$| \cup_i H(u'_i) \setminus W'| \geq 2m^2 \cdot 2c\ell \cdot \frac{\ell m^3}{2} - c^2\ell^2 m^3 \geq c^2\ell^2 m^3.$$

We will apply Lemma 2.3 to connect $N(\cup_i T_i) \setminus W'$ and $\cup_i H(u'_i) \setminus W'$, while avoiding the vertices in W'. Recall that $\varepsilon(x)$ is decreasing, and, since $n/\ell^2 \geq K$ is sufficiently large and $s \geq 2$,

$$\varepsilon(n) = \frac{\varepsilon_1}{\log^2(15n/\varepsilon_2\ell^2)} \geq \frac{4}{\log^3(n/\ell^2)} \geq \frac{4}{m^{3/4}}. \tag{13}$$

Hence, setting $y := c^2\ell^2 m^3$, we have

$$\frac{1}{4} \cdot \varepsilon(y) \cdot y \geq \frac{1}{4} \cdot \varepsilon(n) \cdot y \overset{(13)}{\geq} \frac{1}{4} \cdot \frac{4}{m} \cdot y = c^2\ell^2 m^2.$$

Thus, by (12) and Lemma 2.3, there is a path of length at most

$$\frac{2}{\varepsilon_1} \log^3\left(\frac{15n}{\varepsilon_2\ell^2}\right) + 1 \leq \log^4\left(\frac{n}{\ell^2}\right) \leq m$$

from some T_i to some $H(u'_j)$ which avoids all the vertices in W'. Taking some shortest such path, we may take a w_i, u'_j-path of length at most $1+m+2 \leq 2m$ in $G-W'$ which connects two pairs of vertices unconnected by the process and satisfies **Q1** and **Q2**, a contradiction. $\square$

By Claim 4.3 we can take a vertex w_i with $c\ell$ vertices u_j, $1 \leq j \leq c\ell$, connected to w_i under the rules **Q1** and **Q2**. It suffices to find, in each $(2m^2, 2c\ell)$-hub $H(u_j)$, an $(m^2, c\ell)$-hub which is disjoint from any vertices used in all the connections with w_i, since such a hub together with all the w_i, u_j-paths forms a $(c\ell, m^2, c\ell, 2m)$-unit. By **Q2**, at most 1 vertex in $S_1(u_j)$ is used in the connections with w_i. At most $c\ell \cdot 2m = 2c\ell m$ vertices are used in the connections with w_i. Therefore, at most $2m$ vertices v in $S_1(u_j)$ can have more than $c\ell$ vertices from $S_1(v)$ in the connections with w_i. We can then take a set of m^2 vertices v in $S_1(u_j)$ along with $c\ell$ vertices in each $S_1(v)$ which avoid the connections with w_i, allowing us to form the desired $(m^2, c\ell)$-hub. $\square$

4.3 Constructing subdivisions from units

Finally in this section, we will expand and connect units to form a subdivision.

Proof of Lemma 2.9. Let G be an n-vertex bipartite, $K_{s,t}$-free $(\varepsilon_1, \varepsilon_2 d^{s/(s-1)})$-expander with $\delta(G) \geq d/16$ and $\Delta(G) \leq d \log^{10s}(n/d^{s/(s-1)})$, where $d \geq \log^{20s} n$. Let $c = 1/800t$, $\ell = d^{\frac{1}{2}\frac{s}{s-1}}$, $m = \log^{2s}(n/\ell^2)$ and let $K := K(\varepsilon_1, \varepsilon_2, s, t)$ be sufficiently large that it has the property in Lemma 4.2 and that, if $n \geq K^{s/(s-1)}$, then $\varepsilon(n) \geq 4/m^{3/4}$ (that is, (13) holds) and $m \geq 8/c$. Supposing $n \geq K^{s/(s-1)}$ then, we may find in G $c\ell/2$ $(c\ell, m^2, c\ell, 2m)$-units with disjoint interiors. Let these units be $F_1, \ldots, F_{c\ell/2}$ with core vertices $v_1, \ldots, v_{c\ell/2}$ and denote by $u_{i,j}$ the center of the j-th hub in F_i, $1 \leq j \leq c\ell$. We will construct a $K_{c\ell/4}$-subdivision, which is sufficient to prove the lemma with the constant $c/4$.

Let W be the union of the vertices in all the $v_i, u_{i,j}$-paths in all the units, including their endvertices, so that $|W| \leq c\ell \cdot (2m + 1) \cdot \frac{c\ell}{2} \leq 2c^2\ell^2 m$.

Algorithm R: Greedily connect pairs of hubs from different units using paths between the sets $S_1(u_{i,j})$ and $S_1(u_{i',j'})$ under the following rules.

R1 Each connection uses a path of length at most $6m$, avoiding vertices used in previous connections and vertices in W.

R2 Each hub is in at most one connection.

R3 For each pair of units, there is at most one connection between their respective hubs.

R4 If more than ℓm vertices in the interior of any unit have been used in connections, then discard that unit (along with the hubs it contains).

Note first that by **R1** and **R3**, the number of vertices used in all the connections is at most $(6m + 1) \cdot \binom{c\ell/2}{2} \leq c^2\ell^2 m$. Thus, as the interior of the units are disjoint, we discard at most $\frac{c^2\ell^2 m}{\ell m} \leq c\ell/4$ units by **R4**.

Claim 4.4. *For every pair of remaining units there is a connection between two of their respective hubs.*

Proof of Claim 4.4. Contrary to the claim, suppose there is a pair of units F_i and F_j with no connection between two of their respective hubs. By **R3**, there are at least $c\ell/2$ hubs in F_i not involved in connections. Say these hubs are $H(u_{i,i'})$, $i' \in I$. Let A_i be the set of vertices in $\cup_{i' \in I} S_1(u_{i,i'})$ not involved in connections. As F_i has not been discarded, by **R4** we have

$$|A_i| \geq |\cup_{i' \in I} S_1(u_{i,i'})| - \ell m \geq \frac{c\ell}{2} \cdot m^2 - \ell m \geq \frac{c\ell m^2}{4}.$$

As the hubs in F_i are disjoint, we have

$$|N(A_i) \setminus W| \geq \frac{c\ell m^2}{4} \cdot c\ell - 2c^2\ell^2 m \geq \frac{c^2\ell^2 m^2}{8} := y.$$

17

Similarly, if A_j is defined comparably to A_i, then $|N(A_j) \setminus W| \geq y$. The number of vertices to avoid in **R1** is at most
$$c^2 \ell^2 m + |W| \leq 3c^2 \ell^2 m.$$
Using that $\varepsilon(n) \geq 4/m^{3/4}$ and that $\varepsilon(x)$ is decreasing, we have

$$\frac{1}{4} \cdot \varepsilon(y) \cdot y \geq \frac{1}{4} \cdot \varepsilon(n) \cdot y \geq \frac{1}{4} \cdot \frac{4}{m^{3/4}} \cdot y = \frac{c^2 \ell^2 m^{5/4}}{8} > 3c^2 \ell^2 m.$$

Therefore, by Lemma 2.3, there is a path of length at most m between $N(A_i) \setminus W$ and $N(A_j) \setminus W$ that avoids the vertices specified by **R1**, a contradiction. $\qquad\square$

Finally, we note that we can put the paths we have found together to create a $K_{c\ell/4}$-subdivision. For each pair of remaining core vertices v_i and v_j, take the connection between two of their respective hubs along with the paths from the units between the centre vertex of those hubs and the vertices v_i and v_j, to create a v_i, v_j-path. For each pair of core vertices v_i and v_j we then have a v_i, v_j-path and by **R1** and **R2** these paths are disjoint outside of their endvertices. As there are at least $c\ell/4$ units remaining, we have a $K_{c\ell/4}$-subdivision, as required. $\qquad\square$

5   Constructing subdivisions in a sparse expander

In this section, we prove Lemma 2.10, which covers the case when the subgraph found using Lemma 2.8 is sparse. Here, this means that $d \leq \log^{20s} n$. We break its proof into the following two results. First, in Lemma 5.1, we show that if the graph is sparse, then under some mild expansion property (i.e. if it is an $(\varepsilon_1, \varepsilon_2 d)$-expander) the graph contains a subdivision of order linear in the minimum degree, even without the $K_{s,t}$-free condition. Then, in Proposition 5.2, we show that any graph satisfying the hypothesis of Lemma 2.10 has the expansion property required in Lemma 5.1 (i.e. every $K_{s,t}$-free $(\varepsilon_1, \varepsilon_2 d^{s/(s-1)})$-expander is an $(\varepsilon_1, \varepsilon_2 d)$-expander).

Lemma 5.1. *For each $0 < \varepsilon_1 < 1$, $0 < \varepsilon_2 \leq 1/20$, and $s \geq 2$, there exists $c_1 = c_1(\varepsilon_1, \varepsilon_2, s) > 0$ for which the following holds for each $d > 0$ and $n \in \mathbb{N}$ with $d \leq \log^{20s} n$. If G is an n-vertex, bipartite, $(\varepsilon_1, \varepsilon_2 d)$-expander, with $\delta(G) \geq d/16$ and $\Delta(G) \leq d \log^{10s} n$, then G contains a $K_{c_1 d}$-subdivision.*

Proposition 5.2. *Let $0 < \varepsilon_1 < 1$, $0 < \varepsilon_2 < 1/10^5 t$, and let $t \geq s \geq 2$ be integers. If G is a $K_{s,t}$-free, $(\varepsilon_1, \varepsilon_2 d^{s/(s-1)})$-expander with $\delta(G) \geq d/16$, then G is also an $(\varepsilon_1, \varepsilon_2 d)$-expander.*

Note that together these two results imply Lemma 2.10.

5.1   Proof of Lemma 5.1

We first sketch here the idea of the proof. Using that the graph is sparse and has bounded maximum degree, we first find a collection of vertices v_i that are pairwise far apart and will

serve as the core vertices of our clique subdivision (see Proposition 5.3). We then robustly grow two balls around each vertex v_i, one inner ball of medium size called $B^r(v_i)$ (as in Lemma 5.5) and one outer ball of large size called $B^k(v_i)$ (as in Lemma 5.6). Due to the fact that all the vertices v_i are pairwise far apart, all the sets $B^k(v_i)$ are pairwise disjoint. To construct the desired clique subdivision, we connect pairs v_i, v_j using a shortest path between the outer balls around them while avoiding all vertices in the inner balls of other core vertices, i.e. $\cup_{p \neq i,j} B^r(v_p)$. Using the robust expansion guaranteed by Lemmas 5.5 and 5.6, we will be able to grow new inner and outer balls around v_i to enable us to connect v_i to more core vertices.

We start therefore with Proposition 5.3, which shows there are many vertices which are pairwise far apart in the graph.

Proposition 5.3. *Let $s \geq 1$. Suppose a graph G has n vertices and maximum degree at most $\log^{30s} n$. For sufficiently large n, G contains at least $n^{1/5}$ vertices which are pairwise a distance at least $\log n/(50s \log \log n)$ apart.*

Proof. Let $k = \log n/(50s \log \log n)$. Let Y be a maximal set of vertices of G which are pairwise a distance at least k apart and suppose, for contradiction, that $|Y| \leq n^{1/5}$. Then

$$|B^k(Y)| \leq 2|Y|(\Delta(G))^k \leq 2n^{1/5}(\log^{30s} n)^k \leq 2n^{1/5} \exp\left(\frac{30 \log n}{50}\right) < n.$$

Thus there exists some vertex $v \notin B^k(Y)$, which must be a distance at least k away from each of the vertices in Y, a contradiction. $\qquad\square$

In the next lemma we show that if a vertex v has paths leading into it which do not have many vertices near v (see Definition 5.4) then we can expand out from v while avoiding the interior vertices of these paths. This lemma is a development of techniques used in [19].

Definition 5.4. We say that paths $P_1, \ldots, P_q$, each starting with the vertex v and contained in the vertex set W, are *consecutive shortest paths from v in W* if, for each i, $1 \leq i \leq q$, the path P_i is a shortest path between its endpoints in the set $W - \cup_{j<i} P_j + v$.

Lemma 5.5. *Let $0 < \varepsilon_1 < 1$, $0 < \varepsilon_2 < 1/20$ and $s \geq 1$. Then there is some $c > 0$ and $d_0 \in \mathbb{N}$ for which the following holds for any n and d with $d_0 \leq d \leq \log^{20s} n$. Suppose H is an n-vertex $(\varepsilon_1, \varepsilon_2 d)$-expander with $\delta(H) \geq d/16$. Letting $r = (\log \log n)^5$ and $P = \cup_i V(P_i)$, if $q \leq cd$ and $P_1, \ldots, P_q$ are consecutive shortest paths from v in $B^r(v)$ then*

$$|B^r_{H-P+v}(v)| \geq d^2 \log^7 n.$$

Proof. We will choose $c := c(\varepsilon_1, \varepsilon_2)$ sufficiently small and $d_0 := d_0(\varepsilon_1, \varepsilon_2, s)$ sufficiently large later. Note that as $\log^{20s} n \geq d \geq d_0$ we can also make n sufficiently large with respect to ε_1, ε_2 and s.

Let $F = H - P + v$. We will show by induction on $p \geq 1$ that, if $|B^p_F(v)| \leq d^2 \log^7 n$, then we have $p < r$ and

$$|N_F(B^p_F(v))| \geq \frac{1}{2}|B^p_F(v)| \cdot \varepsilon(|B^p_F(v)|). \tag{14}$$

19

We will also show that $|B_F(v)| \geq d/20$, which together with this inductive statement will prove the lemma. Indeed, if this holds, then for sufficiently large $d \geq d_0$ and hence sufficiently large n, we have for each $1 \leq p < r$ that

$$
\begin{aligned}
|N_F(B_F^p(v))| &\geq \frac{1}{2}|B_F^p(v)| \cdot \varepsilon(|B_F^p(v)|) = \frac{\varepsilon_1 |B_F^p(v)|}{2\log^2\left(\frac{15|B_F^p(v)|}{\varepsilon_2 d}\right)} \geq |B_F^p(v)| \cdot \frac{\varepsilon_1}{2\log^2\left(\frac{15 d^2 \log^7 n}{\varepsilon_2 d}\right)} \\
&\geq |B_F^p(v)| \cdot \frac{\varepsilon_1}{2\log^2\left(\log^{30s} n\right)} \geq \frac{|B_F^p(v)|}{(\log\log n)^3},
\end{aligned}
$$

where we have used that $|B_F^p(v)| \geq |B_F(v)| \geq d/20 \geq \varepsilon_2 d/2$ to apply the expansion property. Thus, we have

$$
|B_F^r(v)| > \left(1 + \frac{1}{(\log\log n)^3}\right)^{r-1} \geq \exp\left(\frac{r-1}{2(\log\log n)^3}\right) \gg \log^{50s} n \geq d^2 \log^7 n.
$$

Thus we need only prove the inductive statement holds and $|B_F(v)| \geq d/20$. Observe that, if $0 \leq p < r$, then, as the paths P_i are consecutive shortest paths from v in $B^r(v)$, only the first $p + 2$ vertices of each path P_i, including v, can belong in $N_H(B_{H-\cup_{j<i}(V(P_j)\backslash\{v\})}^p(v))$. Therefore, if $p < r$, as $F = H - P + v$, only the first $p + 2$ vertices of each of the paths P_i, including the vertex v, can belong in $N_H(B_F^p(v))$. Hence, as we have at most cd paths P_i, if $p < r$, then $|N_H(B_F^p(v)) \cap (P \backslash \{v\})| \leq (p+1)cd$, so that

$$
|N_H(B_F^p(v)) \backslash N_F(B_F^p(v))| \leq (p+1)cd. \tag{15}
$$

Thus, if $|B_F^p(v)| \leq d^2 \log^7 n$, then, once we have shown that $p < r$, we can use the inequality in (15).

In particular, when $p = 0$, combining (15) and the minimum degree condition for H implies that $|B_F(v)| \geq |N_F(v)| \geq d/16 - cd \geq d/20$.

We first verify the base case of (14) when $p = 1$. Since $|B_F(v)| \geq d/20 \geq \varepsilon_2 d$ and that $\varepsilon(x) \cdot x$ is increasing when $x \geq \varepsilon_2 d$, we have that, for sufficiently small c,

$$
2cd \leq \frac{1}{2} \cdot \frac{\varepsilon_1}{\log^2(3/4\varepsilon_2)} \cdot \frac{d}{20} = \frac{1}{2} \cdot \varepsilon\left(\frac{d}{20}\right) \cdot \frac{d}{20} \leq \frac{1}{2}\varepsilon(|B_F(v)|)|B_F(v)|.
$$

The base case then follows from the expansion property, as

$$
|N_F(B_F(v))| \overset{(15)}{\geq} |N_H(B_F(v))| - 2cd \geq \varepsilon(|B_F(v)|)|B_F(v)| - 2cd \geq \frac{1}{2}\varepsilon(|B_F(v)|)|B_F(v)|.
$$

Now, suppose that $p \geq 2$, $|B_F^p(v)| \leq d^2 \log^7 n$, and that the induction hypothesis holds for all $1 \leq p' < p$. Let α be defined by $|B_F^p(v)| = \alpha \varepsilon_2 d/15$, and note that $\alpha \geq 3$. Then

$$
\varepsilon(|B_F^p(v)|) = \frac{\varepsilon_1}{\log^2\left(\frac{15}{\varepsilon_2 d} \cdot \frac{\alpha\varepsilon_2 d}{15}\right)} = \frac{\varepsilon_1}{\log^2 \alpha}. \tag{16}
$$

The induction hypothesis for each p', $1 \leq p' < p$, limits how large p can be. The size of the ball $B_F^p(v)$ has increased by at most a factor of $\frac{|B_F^p(v)|}{|B_F(v)|} \leq \frac{\alpha \varepsilon_2 d/15}{d/20} \leq \alpha$ from $|B_F(v)|$. On the other hand, by the induction hypothesis and (16), and as $\varepsilon(x)$ is decreasing in x when $x \geq \varepsilon_2 d$, at each increase in radius the size of $B_F^p(v)$ increases by at least a factor of $\left(1 + \varepsilon_1/2\log^2 \alpha\right)$. Thus, $\left(1 + \varepsilon_1/2\log^2 \alpha\right)^{p-1} \leq \alpha$, so that

$$p - 1 \leq \frac{\log \alpha}{\log\left(1 + \varepsilon_1/2\log^2 \alpha\right)} \leq \frac{4\log^3 \alpha}{\varepsilon_1}, \tag{17}$$

where we have used that $\log(1+x) \geq x/2$ when x is small. As for sufficiently large $d \geq d_0$, and hence sufficiently large n, we have

$$\alpha = \frac{15|B_F^p(v)|}{\varepsilon_2 d} \leq \frac{15d^2 \log^7 n}{\varepsilon_2 d} \leq \log^{30s} n,$$

by (17) we have that $p - 1 \leq 4(\log(\log^{30s} n))^3/\varepsilon_1 \ll r/2$, and hence $p < r$. Note that when $\alpha \geq 3$ the function $\alpha \mapsto \log^5 \alpha/\alpha$ is bounded above by some universal constant, K say. Therefore, for sufficiently large d, we have

$$(p+1)cd \overset{(17)}{\leq} \frac{8\log^3 \alpha}{\varepsilon_1} \cdot cd \leq \frac{8cd}{\varepsilon_1} \cdot \frac{K \cdot \alpha}{\log^2 \alpha} = \frac{120cK}{\varepsilon_1 \varepsilon_2} \cdot \frac{\alpha \varepsilon_2 d}{15\log^2 \alpha}$$

$$\overset{(16)}{=} \frac{120cK}{\varepsilon_1^2 \varepsilon_2} \cdot \varepsilon(|B_F^p(v)|) \cdot |B_F^p(v)| \leq \frac{1}{2}\varepsilon(|B_F^p(v)|) \cdot |B_F^p(v)|, \tag{18}$$

for c sufficiently small depending on ε_1, ε_2 and the universal constant K. By (15), the expansion property, and (18), we have

$$\begin{aligned}
|N_F(B_F^p(v))| &\geq |N_H(B_F^p(v))| - (p+1)cd \geq \varepsilon(|B_F^p(v)|)|B_F^p(v)| - (p+1)cd \\
&\geq \frac{1}{2}\varepsilon(|B_F^p(v)|)|B_F^p(v)|.
\end{aligned}$$

Therefore, (14) holds for p, and thus the inductive hypothesis for p holds. $\qquad\square$

Having expanded around a vertex while avoiding the paths leading into it, we will expand again while avoiding more vertices. Proposition 5.6 confirms this will result in a large set.

Proposition 5.6. *For each $s \geq 1$, $0 < \varepsilon_1 < 1$ and $\varepsilon_2 > 0$, there exists $d_0 := d_0(\varepsilon_1, \varepsilon_2, s)$ so that the following is true for each $d \geq d_0$. Suppose that H is an $(\varepsilon_1, \varepsilon_2 d)$-expander with n vertices and let $k = \log n/100s \log\log n$. If $Y, W \subset V(G)$ are disjoint sets with $|Y| \geq d^2 \log^7 n$ and $|W| \leq d^2 \log^4 n$, then $|B_{G-W}^k(Y)| \geq \exp(\sqrt[4]{\log n})$.*

Proof. Note that $n \geq |Y| \geq d$, and thus by requiring $d \geq d_0$ to be large, we can make n large. For any $p \geq 0$, if $\varepsilon_2 d/5 \leq |B_{G-W}^p(Y)| \leq \exp(\sqrt[4]{\log n})$ then

$$\varepsilon(|B_{G-W}^p(Y)|) \geq \frac{\varepsilon_1}{\sqrt{\log n}}.$$

21

Thus, as $|B^p_{G-W}(Y)| \geq |Y| \geq d^2 \log^7 n$, we have $\varepsilon(|B^p_{G-W}(Y)|)|B^p_{G-W}(Y)| \geq 2|W|$. Therefore, if $|B^p_{G-W}(Y)| \leq \exp(\sqrt[4]{\log n})$, then, as H is an $(\varepsilon_1, \varepsilon_2 d)$-expander,

$$|N_{G-W}(B^p_{G-W}(Y))| \geq \frac{\varepsilon_1}{2\sqrt{\log n}}|B^p_{G-W}(Y)|.$$

We must have that $|B^k_{G-W}(Y)| < \exp(\sqrt[4]{\log n})$, for otherwise we are done, whereupon

$$|B^k_{G-W}(Y)| \geq \left(1 + \frac{\varepsilon_1}{2\sqrt{\log n}}\right)^k \geq \exp\left(\frac{\varepsilon_1 k}{4\sqrt{\log n}}\right) \geq \exp\left(\sqrt[4]{\log n}\right),$$

a contradiction. $\qquad\square$

Finally, as sketched at the start of this section, we can prove Lemma 5.1.

Proof of Lemma 5.1. Define parameters as follows:

$$r = (\log\log n)^5, \qquad k = \frac{\log n}{100s \log\log n}, \qquad c_1 = \min\{c(\varepsilon_1, \varepsilon_2), 1/d_0(\varepsilon_1, \varepsilon_2)\},$$

where c and d_0 come from Lemma 5.5 and Proposition 5.6 respectively. Note that as $\delta(G) \geq d/16 > 0$, G contains a K_1-subdivision, and hence we may assume that $d \geq d_0$, for otherwise we are done. Since $\Delta(G) \leq d\log^{10s} n \leq \log^{30s} n$, by Proposition 5.3, we may find in G vertices $v_1, \ldots, v_{c_1 d}$ which are pairwise a distance at least $2k$ apart.

Let $I \subset [c_1 d]^{(2)}$ be a maximal subset for which we can find paths Q_e, $e \in I$, so that the following hold.

(i) For each $ij \in I$, Q_{ij} is a v_i, v_j-path with length at most $2\log^4 n$.

(ii) For each $e, e' \in I$, the paths Q_e and $Q_{e'}$ are disjoint except for, potentially, their end vertices.

(iii) For each i, there is some ordering of $\{Q_e[B^r(v_i)] : e \in I, i \in e\}$ such that they are consecutive shortest paths from v_i in $B^r(v_i)$.

(iv) For each $e \in I$ and $i \notin e$, $B^r(v_i)$ and Q_e are disjoint.

If $I = [c_1 d]^{(2)}$, then by (ii) the paths Q_e, $e \in I$, form a $K_{c_1 d}$-subdivision as required. Thus, suppose there is some $ij \in [c_1 d]^{(2)} \setminus I$. Let $W = (\cup_{e \in I} V(Q_e)) \setminus \{v_i, v_j\}$. By (iii), (iv) and Lemma 5.5, we have $|B^k_{G-W}(v_i)| \geq d^2 \log^7 n$. By Proposition 5.6, and as $|W| \leq d^2 \log^4 n$ (due to (i)), we have $|B^{k+r}_{G-W}(v_i)| \geq x := \exp(\sqrt[4]{\log n})$.

Let $W' = W \cup (\cup_{i' \notin \{i,j\}} B^r(v_{i'}))$, i.e. W' is the set of all vertices that are either used in some connection or are in some inner ball of other core vertices. As we chose the vertices $v_{i'}$ to be pairwise at least a distance $2k > k + 2r$ apart, both $B^{k+r}_{G-W}(v_i)$ and $B^{k+r}_{G-W}(v_j)$ are disjoint from W'. Therefore, we have that $|B^{k+r}_{G-W'}(v_i)|, |B^{k+r}_{G-W'}(v_j)| \geq x$. As $\Delta(G) \leq \log^{30s} n$, we then have

$$|W'| \leq |W| + c_1 d \cdot 2(\log^{30s} n)^r \leq \log^{40sr} n = \exp(40s(\log\log n)^6) \ll \varepsilon(x)x/4.$$

22

Therefore, by Lemma 2.3, there is a $B_{G-W'}^{k+r}(v_i), B_{G-W'}^{k+r}(v_j)$-path in $G - W'$ with length at most $\log^4 n$. Thus, if we let Q_{ij} be a shortest v_i, v_j-path in $G - W'$, then Q_{ij} has length at most $\log^4 n + 2k + 2r \leq 2\log^4 n$ in $G - W'$. The paths Q_e, $e \in I \cup \{ij\}$ satisfy the conditions (i)–(iv) above, contradicting the choice of I. $\square$

5.2 Proof of Proposition 5.2

It is left then only to prove Proposition 5.2.

Proof of Proposition 5.2. Let $0 < \varepsilon_1 < 1$ and $0 < \varepsilon_2 \leq 1/10^5 t$. We need to show that for every set $X \subset V(G)$ with $\varepsilon_2 d/2 \leq |X| \leq n/2$, we have $|N(X)| \geq \varepsilon(|X|, \varepsilon_1, \varepsilon_2 d) \cdot |X|$. Recall that $\varepsilon(x, \varepsilon_1, k)$ is an increasing function in k and G is an $(\varepsilon_1, \varepsilon_2 d^{s/(s-1)})$-expander. Thus, for every set $X \subset V(G)$ of size $x \geq \frac{1}{2}\varepsilon_2 d^{s/(s-1)}$ with $x \leq n/2$,

$$|N(X)| \geq \varepsilon(x, \varepsilon_1, \varepsilon_2 d^{s/(s-1)}) \cdot x \geq \varepsilon(x, \varepsilon_1, \varepsilon_2 d) \cdot x.$$

Fix now a set $X \subset V(G)$ of size x with $\frac{1}{2}\varepsilon_2 d \leq x \leq \frac{1}{2}\varepsilon_2 d^{s/(s-1)}$.

Claim 5.7. $|N(X)| \geq |X|$.

Proof of Claim 5.7. As G is bipartite, we can pick a subset $X' \subset X$ in one class of the partition with $|X'| \geq |X|/2$. If there is no vertex v with $d/64$ neighbours in X', then, as $\delta(G) \geq d/16$, $|N(X)| \geq (d/16)|X'|/(d/64) = 4|X'|$, and thus $|N(X)| \geq |N(X')| - |X| \geq |X|$. Therefore, we may assume there is a vertex v with at least $d/64$ neighbours in X'. Letting $A = N(v) \cap X'$ and $B = N(X') \setminus \{v\}$, there is no copy of $K_{s-1,t}$ with $s - 1$ vertices in B and t vertices in A. Therefore, by Corollary 2.6 we have

$$|N(X')| \geq |B| \geq (d/16 - 1)(d/64)^{1/(s-1)}/et \geq \varepsilon_2 d^{s/(s-1)} \geq 2|X|,$$

and thus $|N(X)| \geq |N(X')| - |X| \geq |X|$. $\square$

Therefore, combining this claim with the fact that $\varepsilon(x, \varepsilon_1, \varepsilon_2 d)$ is a decreasing function in x,

$$|N(X)| \geq x > \frac{\varepsilon_1}{\log^2(15/2)} \cdot x = \varepsilon\left(\frac{\varepsilon_2 d}{2}, \varepsilon_1, \varepsilon_2 d\right) \cdot x \geq \varepsilon(x, \varepsilon_1, \varepsilon_2 d) \cdot x,$$

as desired. $\square$

6 Concluding remarks

We have shown that every $K_{s,t}$-free graph with average degree d has a subdivision of a clique of order $\Omega(d^{\frac{1}{2}\frac{s}{s-1}})$. When $s = 2$, this settles in a strong sense a conjecture of Mader that every C_4-free graph contains a clique subdivision with order linear in average degree of the parent graph. As discussed in the introduction, in general the $K_{s,t}$-free condition seems like the most natural to force a subdivision larger than that guaranteed in a general graph

with average degree d. It would be interesting to generalise Theorem 1.1 to non-complete bipartite forbidden subgraphs. Here our methods are limited as we do not have a comparable version of Lemma 2.5. Note that forbidding 3-colourable graphs does not force a larger clique subdivision as the extremal examples for $s(d)$ mentioned in the introduction are bipartite.

Acknowledgements

References

[1] N. Alon, L. Rónyai and T. Szabó, Norm-Graphs: Variations and Applications. *J. Combin. Theory Ser. B*, 76, (1999), 280–290.

[2] J. Balogh, H. Liu and M. Sharifzadeh, Subdivisions of a large clique in C_6-free graphs. *J. Combin. Theory Ser. B*, 112, (2015), 18–35.

[3] B., Bollobás, Modern Graph Theory. Springer, 1998

[4] B. Bollobás, A. Thomason, Proof of a conjecture of Mader, Erdős and Hajnal on topological complete subgraphs. *European J. Combin.*, 19, (1998), 883–887.

[5] P. Erdős, A. Hajnal, On topological complete subgraphs of certain graphs. *Annales Univ. Sci. Budapest*, 7 (1969), 193–199.

[6] Z. Füredi, M. Simonovits, The history of degenerate (bipartite) extremal graph problems. *Bolyai Soc. Math. Stud. Erdős Centennial*, 25 (2013), 169–264.

[7] H.A. Jung, Eine Verallgemeinerung des n-fachen Zusammenhangs für Graphen. *Math. Ann.*, 187, (1970), 95–103.

[8] D. Kühn, D. Osthus, Topological minors in graphs of large girth. *J. Combin. Theory Ser. B*, 86, (2002), 364–380.

[9] D. Kühn, D. Osthus, Large topological cliques in graphs without a 4-cycle. *Combin. Probab. Comput.*, 13, (2004), 93–102.

[10] D. Kühn, D. Osthus, Improved bounds for topological cliques in graphs of large girth. *SIAM J. Discrete Math.*, 20, (2006), 62–78.

[11] D. Kühn, D. Osthus, Extremal connectivity for topological cliques in bipartite graphs. *J. Combin. Theory Ser. B*, 96, (2006), 73–99.

[12] T. Kővári, V.T. Sós and P. Turán, On a problem of K. Zarankiewicz. *Colloquium Math.*, 3, (1954), 50–57.

[13] J. Komlós, E. Szemerédi, Topological cliques in graphs. *Combin. Probab. Comput.*, 3, (1994), 247–256.

[14] J. Komlós, E. Szemerédi, Topological cliques in graphs II. *Combin. Probab. Comput.*, 5, (1996), 79–90.

[15] K. Kuratowski, Sur le probleme des courbes gauches en topologie. *Fund. Math.*, 16, (1930), 271–283.

[16] W. Mader, Homomorphieeigenschaften und mittlere Kantendichte von Graphen. *Math. Annalen*, 174, (1967), 265–268.

[17] W. Mader, Hinreichende Bedingungen für die Existenz von Teilgraphen, die zu einem vollständigen Graphen homöomorph sind. *Math. Nachr.*, 53, (1972), 145–150.

[18] W. Mader, An extremal problem for subdivisions of K_5^-. *J. Graph Theory*, 30, (1999), 261–276.

[19] R.H. Montgomery, Logarithmically small minors and topological minors. *J. Lond. Math. Soc. (2)*, 91, (2015), 71–88.