Advances in Mathematics 354 (2019) 106739

Contents lists available at ScienceDirect

Advances in Mathematics

www.elsevier.com/locate/aim

Packing degenerate graphs ☆

Peter Allen [a,*,1], Julia Böttcher [a,2], Jan Hladký [b,3], Diana Piguet [c,3]

[a] Department of Mathematics, London School of Economics, Houghton Street, London, WC2A 2AE, UK
[b] Institute of Mathematics of the Czech Academy of Sciences. Žitná 25, 110 00, Praha, Czechia [4]
[c] Institute of Computer Science of the Czech Academy of Sciences, Pod Vodárenskou věží 2, 182 07 Prague, Czechia [5]

ARTICLE INFO

Article history:
Received 13 November 2017
Received in revised form 15 May 2019
Accepted 15 July 2019
Available online 2 August 2019
Communicated by Benjamin Sudakov

Keywords:
Graph packing
Randomised algorithm

ABSTRACT

Given D and $\gamma > 0$, whenever $c > 0$ is sufficiently small and n sufficiently large, if $\mathcal{G}$ is a family of D-degenerate graphs of individual orders at most n, maximum degrees at most $\frac{cn}{\log n}$, and total number of edges at most $(1-\gamma)\binom{n}{2}$, then $\mathcal{G}$ packs into the complete graph K_n. Our proof proceeds by analysing a natural random greedy packing algorithm.

☆ The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement number 628974.

* Corresponding author.
E-mail addresses: p.d.allen@lse.ac.uk (P. Allen), j.boettcher@lse.ac.uk (J. Böttcher), honzahladky@gmail.com (J. Hladký), piguet@cs.cas.cz (D. Piguet).

[1] PA was partially supported by the EPSRC, grant number EP/P032125/1.
[2] JB was partially supported by the EPSRC, grant number EP/R00532X/1.
[3] JH and DP were supported by the Czech Science Foundation, grant number GJ16-07822Y.
[4] The Institute of Mathematics of the Czech Academy of Sciences is supported by RVO:67985840.
[5] With institutional support RVO:67985807.

https://doi.org/10.1016/j.aim.2019.106739

1. Introduction

A *packing* of a family $\mathcal{G} = \{G_1, \dots, G_k\}$ of graphs into a graph H is a colouring of the edges of H with the colours $0, 1, \dots, k$ such that the edges of colour i form an isomorphic copy of G_i for each $1 \le i \le k$. The packing is *perfect* if no edges have colour 0. We will often say an edge is *covered* in a packing if it has colour at least 1, and *uncovered* if it has colour zero.

Packing problems have been studied in graph theory for several decades. Many classical theorems and conjectures of extremal graph theory can be written as packing problems. For example, Turán's theorem can be read as the statement that if the n-vertex G does not have too many edges (depending on r), then G and K_r pack into K_n. Putting extremal statements into this context often suggests interesting generalisations, such as asking for packings of more graphs. However packings in this context are usually very far from being perfect packings, with a large fraction of $E(H)$ uncovered. By contrast, in this paper we are interested in *near-perfect packings*, that is, packings in which $o\big(e(H)\big)$ edges are uncovered.

The first problems asking for perfect packings in graphs actually predate modern graph theory: Plücker [23] in 1835 found perfect packings of $\frac{1}{3}\binom{n}{2}$ copies of K_3 into K_n for various values of n, and more generally, Steiner [26] in 1853 asked the following question (phrased then in set-theoretic terms).

Question 1. *Given $2 \le k \le r$, for which values of n does the complete k-uniform hypergraph $K_n^{(k)}$ have a perfect packing with copies of $K_r^{(k)}$?*

A packing of this form is called a *combinatorial design*. There are some simple divisibility conditions on n which are necessary for an affirmative answer. Recently and spectacularly, Keevash [19] proved that for sufficiently large n these conditions are also sufficient. This result was reproved, using a more combinatorial method, by Glock, Kühn, Lo and Osthus [14], who were also able to extend the result to pack with arbitrary fixed hypergraphs in [15]. A related problem tracing back to Kirkman [21] in 1846 asks for packings with copies of the n-vertex $K_r^{(k)}$-factor (Kirkman posed specifically the case $k = 2$, $r = 3$, asking for K_n to be packed with $\frac{n-1}{6}$ copies of the graph consisting of $\frac{n}{3}$ disjoint triangles). Such packings are called *resolvable designs*, and although Ray-Chaudhuri and Wilson [24] solved Kirkman's problem (Kirkman's designs exist if and only if n is congruent to 3 modulo 6), in general the problem is wide open.

The focus of this paper is in packings of large connected graphs. In 1963 Ringel [25] conjectured that if T is any $(n+1)$-vertex tree, then $2n+1$ copies of T pack into K_{2n+1}, and in 1976 Gyárfás [16] made the Tree Packing Conjecture, that if T_i is an i-vertex tree for each $1 \le i \le n$ then $\{T_1, \dots, T_n\}$ packs into K_n. Note that both conjectures ask for perfect packings. These problems are both unsolved, although there are many partial results. It is easy (in both cases) to verify that the conjecture holds when the trees are all stars, or all paths. In both cases, the conjectures were also settled for some specific families of trees (see a rather outdated survey by Hobbs [17]), but until recently there existed no general results.

Intuitively, perfect packing results are hard precisely because every edge must be used. If the graphs $\mathcal{G}$ were embedded in order to H, on coming to the last graph of $\mathcal{G}$ we would need to find that a hole is left in H of precisely the right shape to accommodate it; this clearly requires some foresight in the packing. If some edges will remain uncovered at the end, this difficulty decreases. Bollobás [5] was the first to utilise this, making the observation that one can pack the $2^{-1/2}n$ smallest trees of the Tree Packing Conjecture, and assuming the Erdős–Sós Conjecture[6] even the $\sqrt{3}n/2$ smallest trees. More recently Balogh and Palmer [3] showed that for large n the $\frac{1}{4}n^{1/3}$ largest trees pack, provided their maximum degree is at most $2n^{2/3}$, and without degree restriction that the $\frac{1}{10}n^{1/4}$ largest trees pack in K_{n+1} (i.e. using an extra vertex). These results do not give near-perfect packings — a significant fraction of the complete graph is uncovered — but until recently they were the only general results on the Tree Packing Conjecture allowing high-degree trees.

The first approximate result on the tree packing conjectures is due to Böttcher, Hladký, Piguet and Taraz [6], who showed that one can pack into K_n any family of trees whose maximum degree is at most Δ, whose order is at most $(1-\delta)n$, and whose total number of edges is at most $(1-\delta)\binom{n}{2}$, provided that n is sufficiently large given Δ and $\delta > 0$. This provides approximate versions of both Ringel's Conjecture and the Tree Packing Conjecture for bounded degree graphs. A flurry of generalisations followed, beginning with Messuti, Rödl and Schacht [22], who showed that one can replace trees with graphs from any nontrivial minor-closed family (but still requiring the other conditions), and then by Ferber, Lee and Mousset [10] who showed that the restriction to at most $(1-\delta)n$ vertex graphs is unnecessary. Then, Kim, Kühn, Osthus and Tyomkyn [20] proved a near-perfect packing result for families of graphs with bounded maximum degree which are otherwise unrestricted. At last, Joos, Kim, Kühn and Osthus [18] obtained exact solutions of both Ringel's conjecture and the Tree Packing conjecture when all trees have degree bounded by a constant Δ and n is sufficiently large compared to Δ. This is an impressive and difficult result: what remains (which, unfortunately, is almost all cases) is to consider trees with some vertices of large degree.

Generalising in the direction of removing the restriction to bounded degree graphs, Ferber and Samotij [11] showed two near-perfect packing results for trees, one for spanning trees of maximum degree $O\big(n^{1/6}\log^{-6} n\big)$, and one for almost spanning trees of maximum degree $O\big(n/\log n\big)$. The latter result also follows in the particular case of Ringel's Conjecture from the work of Adamaszek, Allen, Grosu, Hladký [1]. The focus of [1] is the so-called Graceful Tree Conjecture but there is a well-known observation that this conjecture would imply Ringel's Conjecture, see [1, Section 1.1].

[6] The Erdős–Sós Conjecture states that if an n-vertex graph has more than $\frac{1}{2}(k-1)n$ edges then it contains each tree of order $k+1$. A proof (of a slightly weaker form of) the Erdős–Sós Conjecture was announced by Ajtai, Komlós, Simonovits and Szemerédi in the 1990s.

To state our main result, we need to define the *degeneracy* of a graph G. An ordering of $V(G)$ is D-degenerate if every vertex has at most D neighbours preceding it, and G is D-degenerate if $V(G)$ has a D-degenerate ordering. Every graph from a non-trivial minor-closed class has bounded degeneracy. In particular, trees are 1-degenerate, planar graphs are 5-degenerate. Of course, every bounded-degree graph has automatically bounded degeneracy.

Our main result then reads as follows.

Theorem 2. *For each $\gamma > 0$ and each $D \in \mathbb{N}$ there exists $c > 0$ and a number n_0 such that the following holds for each integer $n > n_0$. Suppose that $(G_t)_{t\in[t^*]}$ is a family of D-degenerate graphs, each of which has at most n vertices and maximum degree at most $\frac{cn}{\log n}$. Suppose further that the total number of edges of $(G_t)_{t\in[t^*]}$ is at most $(1-\gamma)\binom{n}{2}$. Then $(G_t)_{t\in[t^*]}$ packs into K_n.*

Theorem 2 thus strengthens the main results about packings into complete graphs from [6,22,10,20,11].[7] The main features of the result are that guest graphs may be spanning, expanding, and have very high maximum degree.

Moving away from packing into complete graphs, there are several classical conjectures which ask for packing results similar to the above when K_n is replaced by a graph of sufficiently high minimum degree, perhaps with additional constraints (such as regularity). Advances have recently been made on several of these, especially by the Birmingham Combinatorics group (see for example [8,4,13]). In particular, we should observe that the near-perfect packing for bounded degree graphs [20] mentioned above actually works in the setting of ε-regular partitions, which turned out to be necessary for the perfect packing results of [18].

Finally, in line with the current trend in extremal combinatorics of asking for random analogues of classical extremal theorems, one can ask for packing results when K_n is replaced by a typical binomial random graph $\mathbb{G}(n, p)$. This is actually the focus of the paper of Ferber and Samotij [11], and they are able to prove near-perfect packing results even in $\mathbb{G}(n, p)$ when p is not much above the threshold for connectivity. Our approach also proves near-perfect packing results (for the same family of graphs) in sufficiently quasirandom graphs of any positive constant edge density (see Theorem 11), and hence in Erdős–Rényi random graphs (see Theorem 12). It might be possible to modify our approach to work in somewhat sparse random graphs as well, but certainly not sparse enough to compete with [11].

Although our current progress with actually proving exact packing conjectures is limited, at least we have not found counterexamples. The existing conjectures point in the following direction.

[7] Some of these papers deal also with packings into non-complete graphs, and most of these results are summarised below.

Meta-Conjecture 3. *Let $\mathcal{G}$ be any family of sparse graphs, and H be an n-vertex dense graph. If there is no simple obstruction to packing $\mathcal{G}$ into H, then a packing exists.*

Some obvious examples of obstructions include the total number of edges in the family $\mathcal{G}$ being larger than $e(H)$, or any graph in $\mathcal{G}$ having more vertices than H. Certainly more subtle obstructions exist. For example it is possible that the total number of edges in graphs of $\mathcal{G}$ equals $e(H)$, but all graphs in $\mathcal{G}$ have only vertices of even degree, while some vertices of H have odd degree, so that there is a parity obstruction to packing $\mathcal{G}$ into H, or that $\mathcal{G}$ contains two graphs with vertices of degree $n-1$ (or more generally too many vertices of very high degree). More such examples exist, see for example the discussions in [6] (Section 9.1) and [18] (after Theorem 1.7). The meta-conjecture can be read as claiming that there is nevertheless a finite list.

Note that without restriction the problem of packing a given $\mathcal{G}$ into a given H is NP-complete (the survey [27] gives several NP-completeness results of which the one in [9] is arguably the most convincing), so in particular we do not expect to find any finite list of simple obstructions to the general packing problem. It follows that 'dense' in the meta-conjecture cannot simply mean large edge-density: one can artificially boost edge density without changing the outcome of this decision problem by taking the disjoint union with a very large clique and adding large connected graphs to $\mathcal{G}$ which perfectly pack the very large clique. However a typical random, or quasirandom, graph seems to be a reasonable candidate for 'dense', as does a graph with high minimum degree (in this case, the minimum degree bound must depend on parameters of the graphs $\mathcal{G}$ such as chromatic number, otherwise a reduction similar to the edge-density reduction exists).

Finally, on the topic of what constitutes a 'sparse graph', observe that bounded degeneracy is a fairly common and unrestrictive notion. One might ask whether degeneracy growing as a function of n is reasonable (of course, in Theorem 2 one can have a very slowly growing function). However, observe that we do not know the answer to Question 1 when r grows superlogarithmically, even for $k = 2$, and it seems reasonable to believe that the answer will often be 'no' even when the simple divisibility conditions are met. It is less clear that the maximum degree restriction of Theorem 2 is necessary, and we expect that it can at least be relaxed. However, with no degree restriction at all Theorem 2 becomes false, see Section 8.2.

Proof outline and organisation of the paper. Our proof of Theorem 2 amounts to the analysis of a quite natural randomised algorithm. We first describe a procedure which works if each graph in $\mathcal{G}$ has order at most $(1-\delta)n$. We take graphs in $\mathcal{G}$ in succession. For each G, we embed vertex by vertex into K_n in a degeneracy order, at each time embedding to a vertex of K_n chosen uniformly at random subject to the constraints that we do not re-use a vertex previously used in embedding G, or an edge used in embedding a previous graph. This procedure succeeds with high probability, and after each stage of embedding a graph, the unused edges in K_n are quasirandom (in a sense we will later make precise).

To allow for spanning graphs, we modify this slightly. We adjust the degeneracy order so that the last δn vertices are independent and all have the same degree; this can be done while at worst doubling the degeneracy of the order. Then for each graph we follow the above procedure to embed the first $(1-\delta)n$ vertices, and finally complete the embedding arbitrarily using a matching argument. We will see that this last step is with high probability always possible. The only slight subtlety is that we have to split $E(K_n)$ into a very dense main part, whose edges we use only for the embedding of the first $(1-\delta)n$ vertices, and a sparse reservoir which we use only for the completion; we do this randomly.

This paper is organised as follows. In Section 2 we introduce martingale concentration inequalities needed for the analysis of our algorithm. We also establish some basic properties of degenerate graphs. In Section 3 we state our main technical result (Theorem 11) and show how to deduce Theorem 2 from it. In Section 4 we describe in detail our packing algorithm, *PackingProcess*, and outline the main steps of its analysis. We also state our main lemmas and show how they imply Theorem 11. In Sections 5, 6 and 7 we prove these lemmas. Finally in Section 8 we give some concluding remarks.

2. Notation and preliminaries

2.1. Notation

When we write $x = y \pm \alpha$, we mean $x \in [y-\alpha, y+\alpha]$. When we write $y \pm \alpha = z \pm \beta$, we mean $[y-\alpha, y+\alpha] \subseteq [z-\beta, z+\beta]$. Note that the latter convention is not symmetric, that is, $y \pm \alpha = z \pm \beta$ is not the same as $z \pm \beta = y \pm \alpha$.

The neighbourhood of a vertex v in the graph G is denoted $\mathsf{N}_G(v)$. We write $\mathsf{N}_G(U) = \bigcap_{v \in U} \mathsf{N}_G(v)$ for the common neighbourhood of the set $U \subseteq V(G)$.

The definition of degenerate graphs naturally suggests to label the vertices of a graph by integers. Suppose that the vertices of a graph G are $V(G) = [\ell]$. Suppose that $i \in V(G)$. We write $\mathsf{N}^-(i) = \mathsf{N}(i) \cap [i-1]$ and $\deg^-(i) = |\mathsf{N}^-(i)|$ for the *left-neighbourhood* and the *left-degree* of i. We make use of the natural order on $[\ell]$ also in other ways, like referring to sets of the form $[\ell_1] \subseteq V(G)$ and $\{\ell_2, \ell_2+1, \dots, \ell\} \subseteq V(G)$ as *initial vertices* and *final vertices*, respectively. The *density* of a graph H is the quantity $e(H)/\binom{v(H)}{2}$.

The graphs to be packed in Theorem 2 are denoted G_t because they are **g**uest graphs. By contrast, during our packing procedure, we shall work with **h**ost graphs H_s which are obtained from the original K_n by removing what was used previously.

2.2. Probability

2.2.1. Probability basics

All probability spaces considered in this paper are finite. The implicit sigma-algebra underlying each such space is the sigma-algebra generated by all singletons; in particular, the notion of measurability is trivial in this setting. Recall that if Ω is finite probability space then a sequence of partitions $\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n$ of Ω is a *filtration* if each partition $\mathcal{F}_i$ refines its predecessor $\mathcal{F}_{i-1}$.[8] In this setting, a function $f : \Omega \to \mathbb{R}$ is called *$\mathcal{F}_i$-measurable* if f is constant on each part of $\mathcal{F}_i$.

Recall also that if Ω is a finite probability space and $f : \Omega \to \mathbb{R}$ is a function, then the *conditional expectation* $\mathbb{E}(f|\mathcal{F}) : \Omega \to \mathbb{R}$ and the *conditional variance* $\mathrm{Var}(f|\mathcal{F}) : \Omega \to \mathbb{R}$ of f with respect to a given partition $\mathcal{F}$ of Ω are defined by

$$\begin{aligned} \mathbb{E}(f|\mathcal{F})(x) &= \mathbb{E}(f|X), \\ \mathrm{Var}(f|\mathcal{F})(x) &= \mathrm{Var}(f|X), \end{aligned} \qquad \text{where } X \in \mathcal{F} \text{ is such that } X \ni x.$$

2.2.2. Sequential dependence and concentration

In this section we introduce some convenient consequences of standard martingale inequalities. These are generally useful in the analysis of randomised processes, so we try to provide some brief background and motivation.

Suppose that we have a randomised algorithm which proceeds in m rounds. We can then denote by $\Omega := \prod_{i=1}^{m} \Omega_i$ the probability space that underlies an execution of the algorithm. Here Ω_i is the set of all possible choices the algorithm may make in step i. It is important, however, that Ω as a probability space is not necessarily a product of probability spaces Ω_i; in other words, the algorithms can (and typically will) make choices for the step i depending on the choices it made in steps $1, \ldots, i-1$. By *history up to time t* we mean a set of the form $\{\omega_1\} \times \cdots \times \{\omega_t\} \times \Omega_{t+1} \times \cdots \Omega_m$, where $\omega_i \in \Omega_i$. We shall use the symbol $\mathscr{H}_t$ to denote any particular history of such a form. By a *history ensemble up to time t* we mean any union of histories up to time t; we shall use the symbol $\mathscr{L}$ to denote any one such. Observe that there are natural filtrations associated to such a probability space: given times $t_1 < t_2 < \ldots$ we let $\mathcal{F}_{t_i}$ denote the partition of Ω into the histories up to time t_i. We introduce formally a probability space of this type, which we use for the key part of our argument, in Section 4.1.

We recall that if $Y_1, \ldots, Y_n$ are a collection of independent random variables, whose ranges are not too large compared to n, we have Hoeffding's inequality for the tails of such sums:

$$\mathbb{P}\Big(\sum_{i=1}^{n} \big(Y_i - \mathbb{E}(Y_i)\big) \geq \varrho \Big) \leq \exp\Big(- \frac{2\varrho^2}{\sum_{i=1}^{n} (\max Y_i - \min Y_i)^2} \Big), \tag{2.1}$$

for each $\varrho > 0$. One should think of the squared range of Y_i as a crude upper bound for $\mathrm{Var}(Y_i)$. There are various improvements, such as the Bernstein inequalities, which take into account the actual values $\mathrm{Var}(Y_i)$ in order to obtain stronger concentration results such as

[8] Readers familiar with measure-theoretic probability will notice that the standard definition is a sequence of σ-algebras, namely those generated by our partitions; in the finite setting this is an unnecessary complication.

$$\mathbb{P}\Big(\sum_{i=1}^{n}\big(Y_i-\mathbb{E}(Y_i)\big)\geq\varrho\Big)\leq\exp\Big(-\frac{\varrho^2}{2R\varrho/3+2\sum_{i=1}^{n}\mathrm{Var}(Y_i)}\Big)\,, \tag{2.2}$$

valid when $0\leq Y_i\leq R$ for each i. When the sum of variances is much larger than $R\varrho$, this probability bound is optimal up to small order terms in the exponent; for most applications this means it cannot usefully be improved.

However when analysing randomised algorithms, usually one has to deal with a sum of random variables which are not independent, but rather are *sequentially dependent*, meaning that they come in an order in which earlier outcomes affect the later random variables. A good example is the following procedure (a variant of which we use in this paper) for embedding a graph G on vertex set $[n/2]$ into a graph H on n vertices. We simply embed vertices in order $1,\dots,n/2$, at each time t embedding vertex t uniformly at random to the set of all valid choices: that is, choices which give an embedding of $G[1,\dots,t]$. In order to show that this procedure is likely to succeed (which is true if G has small degeneracy and H is sufficiently quasirandom) we will want to know how vertices are embedded over time to some subsets $S\subseteq V(H)$. In other words, we define (in this case, Bernoulli) random variables Y_t to be 1 if t is embedded to S and 0 otherwise, and we want to know how the partial sums of these random variables, which are certainly not independent but are sequentially dependent, behave. The point of this section is to observe that in fact more or less the same concentration bounds hold as for independent random variables, except that one has to replace the sum of expectations with a sum of observed expectations, that is, $\sum_{i=1}^{n}\mathbb{E}\big(Y_i|\mathscr{H}_{i-1}\big)$, where $\mathscr{H}_{i-1}$ denotes the history up to time $i-1$, and the sum of variances with a sum of observed variances, similarly defined.

In combinatorial applications, one is usually interested in showing that a sum of random variables (which might in general not be Bernoulli) is close to its expectation μ. It is not *a priori* obvious that concentration bounds such as the above help: after all, the sum of observed expectations is itself a random variable and might not be concentrated near μ (it is easy to come up with examples in which it is not). We deal with this in what follows by defining a *good event* $\mathcal{E}$, within which the observed sum of expectations is $\mu\pm\nu$ for some (small) $\nu>0$. In applications $\mathcal{E}$ will often be a combinatorial statement about the process, and hence we refer to ν as the *combinatorial error*, to distinguish it from the *probabilistic error* $\varrho>0$, as in (2.1) and (2.2). It is important to note that $\mathcal{E}$ is usually not determined before the random variables Y_i (i.e. it may well not be $\mathcal{F}_i$-measurable for any member $\mathcal{F}_i$ of the filtration), so we do not condition on $\mathcal{E}$, rather we aim to estimate the probability that $\mathcal{E}$ holds and yet $\sum_{i=1}^{n}Y_i\neq\mu\pm(\nu+\varrho)$.

In order to avoid mentioning any particular process, it is convenient to state the following lemmas in terms of a finite probability space Ω with a filtration $(\mathcal{F}_0,\mathcal{F}_1,\dots,\mathcal{F}_n)$. We should stress that though in our applications we will always use the same probability space, which underlies our packing process, we will consider different filtrations, always given by the histories up to increasing times, depending on the random variables we wish to sum.

The following lemma, from [1], is a sequential dependence version of Hoeffding's inequality. Note that the lemma as stated in [1] includes the condition $\mathbb{P}(\mathcal{E}) > 0$. However if $\mathbb{P}(\mathcal{E}) = 0$ the lemma statement is trivially true, so we drop the condition below.

Lemma 4 *(Lemma 7, [1]). Let Ω be a finite probability space, and $(\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_n)$ be filtration. Suppose that for each $1 \le i \le n$ we have a nonnegative real number a_i, an $\mathcal{F}_i$-measurable random variable Y_i satisfying $0 \le Y_i \le a_i$, nonnegative real numbers μ and ν, and an event $\mathcal{E}$. Suppose that almost surely, either $\mathcal{E}$ does not occur or $\sum_{i=1}^n \mathbb{E}\left(Y_i \middle| \mathcal{F}_{i-1}\right) = \mu \pm \nu$. Then for each $\varrho > 0$ we have*

$$\mathbb{P}\left(\mathcal{E} \text{ and } \sum_{i=1}^n Y_i \neq \mu \pm (\nu + \varrho)\right) \le 2 \exp\left(-\frac{2\varrho^2}{\sum_{i=1}^n a_i^2}\right).$$

Furthermore, if we weaken the assumption, requiring only that either $\mathcal{E}$ does not occur or $\sum_{i=1}^n \mathbb{E}\left(Y_i \middle| \mathcal{F}_{i-1}\right) \le \mu + \nu$, then for each $\varrho > 0$ we have

$$\mathbb{P}\left(\mathcal{E} \text{ and } \sum_{i=1}^n Y_i > \mu + \nu + \varrho\right) \le \exp\left(-\frac{2\varrho^2}{\sum_{i=1}^n a_i^2}\right).$$

We should note that the probability bound in this lemma is what one would obtain from standard martingale inequalities for $\mathbb{P}(\sum_{i=1}^n Y_i \neq \mu \pm (\nu + \varrho))$ if the condition $\sum_{i=1}^n \mathbb{E}\left(Y_i \middle| \mathcal{F}_{i-1}\right) = \mu \pm \nu$ held almost surely. The rôle of $\mathcal{E}$ is that we can allow this condition to fail outside of $\mathcal{E}$ but still obtain the same concentration within $\mathcal{E}$; this is probabilistically fairly trivial but very useful. The same applies for the next lemma.

Lemma 4 gives close to optimal (up to a constant factor in the exponential) results when the random variables Y_i are relatively often close to 0 and a_i; in other words, when a_i^2 is not much larger than the variance $\mathrm{Var}(Y_i)$. This will turn out to be the case for most of the random sums we need to estimate in this paper. However, when it is not the case, at the cost of a second moment calculation the following version of Freedman's inequality [12] gives much stronger bounds, corresponding to a Bernstein inequality for independent random variables.

Lemma 5 *(Freedman's inequality on a good event). Let Ω be a finite probability space, and $(\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_n)$ be a filtration. Suppose that we have $R > 0$, and for each $1 \le i \le n$ we have an $\mathcal{F}_i$-measurable non-negative random variable Y_i, nonnegative real numbers μ, ν and σ, and an event $\mathcal{E}$. Suppose that almost surely, either $\mathcal{E}$ does not occur or we have $\sum_{i=1}^n \mathbb{E}\left(Y_i \middle| \mathcal{F}_{i-1}\right) = \mu \pm \nu$, and $\sum_{i=1}^n \mathrm{Var}\left(Y_i \middle| \mathcal{F}_{i-1}\right) \le \sigma^2$, and $0 \le Y_i \le R$ for each $1 \le i \le n$. Then for each $\varrho > 0$ we have*

$$\mathbb{P}\left(\mathcal{E} \text{ and } \sum_{i=1}^n Y_i \neq \mu \pm (\nu + \varrho)\right) \le 2 \exp\left(-\frac{\varrho^2}{2\sigma^2 + 2R\varrho}\right).$$

Furthermore, if we assume only that either $\mathcal{E}$ does not occur or we have $\sum_{i=1}^{n} \mathbb{E}\left(Y_i \middle| \mathcal{F}_{i-1}\right) \leq \mu + \nu$, and $\sum_{i=1}^{n} \operatorname{Var}\left(Y_i \middle| \mathcal{F}_{i-1}\right) \leq \sigma^2$, and $0 \leq Y_i \leq R$ for each $1 \leq i \leq n$, then for each $\varrho > 0$ we have

$$\mathbb{P}\left(\mathcal{E} \text{ and } \sum_{i=1}^{n} Y_i > \mu + \nu + \varrho\right) \leq \exp\Big(-\frac{\varrho^2}{2\sigma^2 + 2R\varrho}\Big).$$

As with the Bernstein inequality, this result is essentially optimal when the sum of observed variances is much larger than $R\varrho$. We would like to point out that since $\mathcal{E}$ is often a combinatorial statement which is not tailored to the specific random variables Y_i we are summing, when we use either lemma to estimate tail probabilities for several sums of random variables, we will often use the same event $\mathcal{E}$ repeatedly; since it will appear only once in union bounds, both lemmas are useful for showing that a.a.s. a collection of many (rapidly growing with n) sums are simultaneously close to their expectations, even when the probability of $\mathcal{E}$ only tends to one quite slowly with n.

We deduce Lemma 5 from Freedman's martingale inequality, which we now state.

Theorem 6 *(Proposition (2.1), [12]). Let Ω be a finite probability space, and $(\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_n)$ be a filtration. Suppose that for some $R > 0$, for each $1 \leq i \leq n$, we have an $\mathcal{F}_i$-measurable random variable Y_i that takes values in the range $-R \leq Y_i \leq R$, and we have $\mathbb{E}(Y_i|\mathcal{F}_{i-1}) = 0$ almost surely. Suppose that for some σ we have $\sigma^2 \geq \sum_{i=1}^{n} \operatorname{Var}(Y_i|\mathcal{F}_{i-1})$ almost surely. Then for each $\varrho > 0$, we have*

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq \varrho\right) \leq \exp\Big(-\frac{\varrho^2}{2\sigma^2 + 2R\varrho}\Big).$$

We now deduce Lemma 5, using a similar approach as was used in [1] to prove Lemma 4.

Proof of Lemma 5. We show the required upper bound

$$\mathbb{P}\left(\mathcal{E} \text{ and } \sum_{i=1}^{n} Y_i > \mu + \nu + \varrho\right) \leq \exp\Big(-\frac{\varrho^2}{2\sigma^2 + 2R\varrho}\Big), \tag{2.3}$$

and the corresponding lower bound follows by symmetry, replacing each Y_i with $R - Y_i$. This gives the desired two-sided result by the union bound.

Observe that if $\mathbb{P}(\mathcal{E}) = 0$, (2.3) holds trivially. We may thus assume $\mathbb{P}(\mathcal{E}) > 0$. Now, given $Y_1, \dots, Y_n$, we define random variables $U_1, \dots, U_n$ as follows. We set $U_i = \max(Y_i, R)$ if $\mathbb{P}(\mathcal{E}|\mathcal{F}_{i-1}) > 0$, and otherwise $U_i = 0$. Observe that U_i is constant on each part of $\mathcal{F}_i$ by definition. We claim that for each $1 \leq t \leq n$ we have almost surely

$$\sum_{i=1}^{t} \mathbb{E}(U_i|\mathcal{F}_{i-1}) \leq \mu + \nu \quad \text{and} \quad \sum_{i=1}^{t} \operatorname{Var}(U_i|\mathcal{F}_{i-1}) \leq \sigma^2. \tag{2.4}$$

Indeed, suppose that t is minimal such that this statement fails, and let F be a set in $\mathcal{F}_{t-1}$ with $\mathbb{P}(F) > 0$ witnessing its failure. By minimality of t, at least one of $\mathbb{E}(U_t|F)$ and $\mathrm{Var}(U_t|F)$ is strictly positive. By definition of U_t we have $\mathbb{P}(\mathcal{E}|F) > 0$. But since $\mathbb{E}(U_i|\mathcal{F}_{i-1})$ and $\mathrm{Var}(U_i|\mathcal{F}_{i-1})$ are nonnegative for each i, this shows that with probability at least $\mathbb{P}(F)\mathbb{P}(\mathcal{E}|F) > 0$, the event $\mathcal{E}$ occurs and one of the assumptions $\sum_{i=1}^n \mathbb{E}\left(Y_i\middle|\mathcal{F}_{i-1}\right) = \mu \pm \nu$ and $\sum_{i=1}^n \mathrm{Var}\left(Y_i\middle|\mathcal{F}_{i-1}\right) \le \sigma^2$ fails. This is a contradiction, so we conclude (2.4) holds almost surely for each t. Furthermore, we have $0 \le U_i \le R$ for each $1 \le i \le n$.

Next, define for each $1 \le i \le n$ the random variable $W_i = U_i - \mathbb{E}(U_i|\mathcal{F}_{i-1})$. We have $-R \le W_i \le R$ for each i, by definition W_i is $\mathcal{F}_i$-measurable, and by definition almost surely $\mathbb{E}(W_i|\mathcal{F}_{i-1}) = 0$ and $\mathrm{Var}(W_i|\mathcal{F}_{i-1}) = \mathrm{Var}(U_i|\mathcal{F}_{i-1})$. Thus by Theorem 6 we have

$$\mathbb{P}\left(\sum_{i=1}^n W_i \ge \varrho\right) \le \exp\left(-\frac{\varrho^2}{2\sigma^2 + 2R\varrho}\right).$$

Since almost surely we have $\sum_{i=1}^t \mathbb{E}(U_i|\mathcal{F}_{i-1}) \le \mu + \nu$, we obtain

$$\mathbb{P}\left(\sum_{i=1}^n U_i \ge \mu + \nu + \varrho\right) \le \exp\left(-\frac{\varrho^2}{2\sigma^2 + 2R\varrho}\right).$$

Finally, if $\mathcal{E}$ occurs then almost surely $Y_i = U_i$ for each $1 \le i \le n$, giving the desired upper bound (2.3). $\square$

Finally, let us note that we shall be using many statements of the form

$$\text{with probability at least } p\text{, provided event } \mathcal{A} \text{ we get event } \mathcal{B}. \tag{2.5}$$

We emphasize that such statements are not statements about conditional probabilities. That is, the meaning of (2.5) is $\mathbb{P}(\mathcal{A}\setminus\mathcal{B}) \le 1-p$. A prototypical example is *with probability at least* $1-o(1)$, *if a given randomised algorithm does not fail, then it produces an output with certain desired properties.*

2.3. Simple properties of degenerate graphs

We need to bound $\sum_{x\in V(G)} \deg(x)^2$ for degenerate graphs G. In several applications of Lemma 4 the numbers a_i will be upper bounded by the degrees of vertices in G, where G is one of the graphs to be packed, so that $\sum_{x\in V(G)} \deg(x)^2$ is an upper bound for the sum $\sum_i a_i^2$ appearing in Lemma 4.

Lemma 7. *Let G be an n-vertex graph with degeneracy D and maximum degree Δ. Then we have*

$$\sum_{x\in V(G)} \deg(x)^2 \le 2Dn\Delta\,.$$

Proof. We have

$$\sum_{x\in V(G)} \deg(x)^2 \le \sum_{x\in V(G)} \deg(x)\cdot\Delta = 2e(G)\cdot\Delta \le 2Dn\cdot\Delta\,. \quad \square$$

We also need to show that degenerate graphs contain large independent sets all of whose vertices have the same degree.

Lemma 8. *Let G be a D-degenerate n-vertex graph. Then there exists an integer $0 \le d \le 2D$ and a set $I \subseteq V(G)$ with $|I| \ge (2D+1)^{-3}n$ which is independent, and all of whose vertices have the same degree d in G.*

Proof. We first claim that at least $(2D+1)^{-1}n$ vertices of G have degree at most $2D$. Indeed, if this were false then there would be more than $2Dn/(2D+1)$ vertices of G all of whose degrees are at least $2D+1$, so that we obtain $e(G) > Dn$, which contradicts the D-degeneracy of G. Let $0 \le d \le 2D$ be chosen to maximise the number of vertices in G of degree d, and let S be the set of vertices in G with degree d. We thus have $|S| \ge (2D+1)^{-2}n$. Now let I be a maximal independent subset of S. Each vertex of I has at most $d \le 2D$ neighbours in S, so that $\big|I \cup \bigcup_{i\in I} \mathsf{N}(i)\big| \le (2D+1)|I|$. By maximality $I \cup \bigcup_{i\in I} \mathsf{N}(i)$ covers S, hence $|I| \ge (2D+1)^{-1}|S| \ge (2D+1)^{-3}n$, as desired. $\square$

3. Reducing the main theorem

We deduce Theorem 2 from the following technical result.

Theorem 9. *For each $\gamma > 0$ and each $D \in \mathbb{N}$ there exists $c > 0$ and a number n_0 such that the following holds for each integer $n > n_0$. Suppose that $s^* \le 2n$ and that for each $s \in [s^*]$ the graph G_s is a graph on vertex set $[n]$, with maximum degree at most $\frac{cn}{\log n}$, such that $\deg^-(x) \le D$ for each $x \in V(G_s)$ and such that the last $(D+1)^{-3}n$ vertices of $[n]$ form an independent set in G_s, and all have the same degree d_s in G_s. Suppose further that the total number of edges of $(G_s)_{s\in[s^*]}$ is at most $(1-3\gamma)\binom{n}{2}$. Then $(G_s)_{s\in[s^*]}$ packs into K_n.*

Actually, we prove Theorem 9 in a slightly more general form using the concept of quasirandomness which is crucial for our approach. This concept was introduced by several authors independently in the 1980s (of which the paper [7] is the most comprehensive) and captures a property that the edges of graph are distributed evenly among its vertices. We give a definition tailored for our needs which is somewhat stronger than the usual definition of quasirandom graphs.

Definition 10 *(quasirandom).* Suppose that H is a graph with n vertices and with density p. We say that such graph H is (α, L)-*quasirandom* if for every set $S \subseteq V(H)$ of at most L vertices we have $|\mathsf{N}_H(S)| = (1 \pm \alpha)p^{|S|}n$.

Theorem 11 *(Main technical result). For each $\gamma > 0$ and each $D \in \mathbb{N}$ there exist numbers $n_0 \in \mathbb{N}$ and $c, \xi > 0$ such that the following holds for each $n > n_0$. Suppose that $\widehat{H}$ is an $(\xi, 2D+3)$-quasirandom graph with n vertices and density $p > 0$. Suppose that $s^* \leq 2n$ and that for each $s \in [s^*]$ the graph G_s is a graph on vertex set $[n]$, with maximum degree at most $\frac{cn}{\log n}$, such that $\deg^-(x) \leq D$ for each $x \in V(G_s)$ and such that the last $(D+1)^{-3}n$ vertices of $[n]$ form an independent set in G_s, and all have the same degree d_s in G_s. Suppose further that the total number of edges of $(G_s)_{s\in[s^*]}$ is at most $(p-3\gamma)\binom{n}{2}$. Then $(G_s)_{s\in[s^*]}$ packs into $\widehat{H}$.*

Theorem 11 indeed generalises Theorem 9 because it can be easily checked that for any fixed $D \in \mathbb{N}$ and $\alpha > 0$, the graph K_n is $(\alpha, 2D+3)$-quasirandom for n sufficiently large. The reason why we give the proof in this greater generality is that it is clear that the only feature of K_n we actually use is its quasirandomness. We show that Theorem 9 implies Theorem 2. Note that starting with Theorem 11 the same deduction would yield a version of Theorem 2 for quasirandom host graphs. We state such a version for dense Erdős–Rényi random graphs $\mathbb{G}(n,p)$, an n-vertex graph, where each pair of vertices forms an edge independently with probability p. Those graphs are well-known to have asymptotically almost surely error in quasirandomness (even in our Definition 10) tending to zero.

Theorem 12. *For each $p, \gamma > 0$ and each $D \in \mathbb{N}$ there exists $c > 0$ such that the following holds asymptotically almost surely, as $n \to \infty$. Suppose that $(G_t)_{t\in[t^*]}$ is a family of D-degenerate graphs, each of which has at most n vertices and maximum degree at most $\frac{cn}{\log n}$. Suppose further that the total number of edges of $(G_t)_{t\in[t^*]}$ is at most $(p-\gamma)\binom{n}{2}$. Then $(G_t)_{t\in[t^*]}$ packs into $\mathbb{G}(n,p)$.*

Proof of Theorem 2. To deduce Theorem 2 from Theorem 9, observe that given an integer D and graphs $\mathcal{G} = (G_t)_{t\in[t^*]}$ to pack, we may assume without loss of generality that none of the graphs in $\mathcal{G}$ has isolated vertices, since such vertices can be erased and then easily packed in the last step.

We now successively modify the family $\mathcal{G}$ as follows. If there are two graphs $G, G' \in \mathcal{G}$ with $v(G), v(G') \leq n/2$, we replace G and G' with the disjoint union $G \cup G'$. We repeat this until no further such pairs exist, giving $\mathcal{G}'$.

Observe that the maximum degree and the degeneracy of the graphs in $\mathcal{G}$ is the same as in $\mathcal{G}'$. Furthermore a packing of $\mathcal{G}'$ is also a packing of $\mathcal{G}$. Finally, there is at most one graph in $\mathcal{G}'$ with less than $n/2$ vertices. Hence all but at most one graph has at least $n/4$ edges. We conclude that the total number s^* of graphs in $\mathcal{G}'$ satisfies $(s^*-1)n/4 \leq (1-\gamma)\binom{n}{2}$, and hence $s^* \leq 2n$. Finally, we let the graphs $(G'_s)_{s=1}^{s^*}$ be obtained from the graphs $\mathcal{G}'$ by adding if necessary isolated vertices to each in order to obtain n-vertex graphs.

Now, for each G'_s we choose an order on $V(G'_s)$ as follows. First, we pick an order witnessing D-degeneracy of G'_s. Next, we pick an integer $0 \leq d_s \leq 2D$ and an independent

I_s set of $(2D+1)^{-3}n$ vertices each of which has degree d_s in G'_s and change the order by moving these vertices to the end. Such an integer d_s and independent set exist by Lemma 8. The result is an ordering of $V(G'_s)$ with degeneracy at most $2D$, as required for Theorem 9 with input $2D$ and $\gamma/3$. Then Theorem 9 returns the desired packing. □

4. Proof of Theorem 11

For the proof of Theorem 11, we need some algorithms and definitions. We give these now along with a sketch of the proof.

We prove Theorem 11 by analysing a randomised algorithm, which we call *PackingProcess*, that packs the guest graphs G_s into $\widehat{H}$. We prove that this algorithm succeeds with high probability. In this algorithm we assume that the last δn vertices of each graph G_s form an independent set, where $\delta < (D+1)^{-3}$ is to be chosen later.

PackingProcess begins by splitting the edges of the input graph $\widehat{H}$ into a *bulk* H_0 and a *reservoir* H_0^* by independently selecting edges into the latter with probability chosen such that $e(H_0^*) \approx \gamma\binom{n}{2}$. As a result, the graphs H_0 and H_0^* are with high probability quasirandom.

Now *PackingProcess* proceeds in s^* stages. In each stage s, it runs a randomised embedding algorithm, called *RandomEmbedding* and explained below, to embed the first $n-\delta n$ vertices of G_s into the bulk H_{s-1}. Then in the *completion phase* the last δn vertices of G_s are embedded into the reservoir H_{s-1}^*. Since there are exactly δn vertices of G_s left to embed and exactly δn vertices of $V(\widehat{H})$ unused so far in this stage, we want to find a bijection between these. Since all neighbours of each yet unembedded vertex are already embedded, this completion amounts to choosing a system of distinct representatives. The completion phase does not use randomness: the system of disjoint representatives is obtained using Hall's theorem. Now H_s and H_s^* are defined simply by removing the edges used in this embedding.

Both *RandomEmbedding* and the completion phase may *fail* at any stage s; this means that it is not possible to embed a certain part of G_s. In that case *PackingProcess* fails, too. If *PackingProcess* does not fail then it always produces a valid packing of (G_s) into H. So, we need to show that *PackingProcess* (see Algorithm 1) succeeds with positive probability.

For describing our randomised embedding algorithm *RandomEmbedding* we need the following definitions. We shall use the symbol $\hookrightarrow$ to denote embeddings produced by *RandomEmbedding*. We write $G \hookrightarrow H$ to indicate that the graph G is to be embedded into H. Also, if $t \in V(G)$, $v \in V(H)$ and $A \subseteq V(H)$ then $t \hookrightarrow v$ means that t is embedded on v, and $t \hookrightarrow A$ means that t is embedded on a vertex of A.

Definition 13 *(partial embedding, candidate set).* Let G be a graph with vertex set $[v(G)]$, and H be a graph with $v(H) \geq v(G)$. Further, assume $\psi_j \colon [j] \to V(H)$ is a *partial embedding* of G into H for $j \in [v(G)]$, that is, ψ_j is a graph embedding of $G\big[[j]\big]$ into H.

Algorithm 1: *PackingProcess.*

Input: graphs $G_1, \dots, G_{s^*}$, with G_s on vertex set $[n]$ such that the last δn vertices of G_s form an independent set; a graph $\widehat{H}$ on n vertices
choose H_0^* by picking edges of $\widehat{H}$ independently with probability $\gamma\binom{n}{2}/e(\widehat{H})$;
let $H_0 = \widehat{H} - H_0^*$;
for $s = 1$ **to** s^* **do**
  run *RandomEmbedding* (G_s, H_{s-1}) to get an embedding ϕ_s of $G_s[[n-\delta n]]$ into H_{s-1};
  let H_s be the graph obtained from H_{s-1} by removing the edges of $\phi_s\big(G_s[[n-\delta n]]\big)$;
  choose an extension ϕ_s^* of ϕ_s embedding all of G_s and embedding the edges of $G_s - G_s[[n-\delta n]]$ into H_{s-1}^* ;
  let H_s^* be the graph obtained from H_{s-1}^* by removing the edges of $\phi_s^*\big(G_s - G_s[[n-\delta n]]\big)$;
end

Finally, let $t \in [v(G)]$ be such that $\mathsf{N}_G^-(t) \subseteq [j]$. Then the *candidate set of* t (with respect to ψ_j) is

$$C^j_{G\hookrightarrow H}(t) = \mathsf{N}_H\Big(\psi_j\big(\mathsf{N}_G^-(t)\big)\Big) .$$

When $j = t-1$, we call $C^j_{G\hookrightarrow H}(t)$ the *final candidate set of* t.

RandomEmbedding (see Algorithm 2) randomly embeds a guest graph G into a host graph H. The algorithm is simple: we iteratively embed the first $(1-\delta)n$ vertices of G randomly to one of the vertices of their candidate set which was not used for embedding another vertex already.

Algorithm 2: *RandomEmbedding.*

Input: graphs G and H, with $V(G) = [v(G)]$ and $v(H) = n$
$\psi_0 := \emptyset$;
$t^* := (1-\delta)n$;
for $t = 1$ **to** t^* **do**
  if $C^{t-1}_{G\hookrightarrow H}(t) \setminus \mathrm{im}(\psi_{t-1}) = \emptyset$ **then** halt with failure;
  choose $v \in C^{t-1}_{G\hookrightarrow H}(t) \setminus \mathrm{im}(\psi_{t-1})$ uniformly at random;
  $\psi_t := \psi_{t-1} \cup \{t \hookrightarrow v\}$;
end
return ψ_{t^*}

To show that *PackingProcess* does not fail at any stage, we shall show that the host graph H_s constructed in *PackingProcess* in embedding stage s is quasirandom in the sense of Definition 10. In fact, in order to analyse the completion phase of *PackingProcess* we need quasirandomness of the pair (H_s, H_0^*), where H_0^* is the initial reservoir. We now define this *coquasirandomness* of a pair of graphs. Recall that quasirandomness of one graph means that common neighbourhoods are always about the size one would expect in a random graph of a similar density. Coquasirandomness of two graphs means that the intersection of a common neighbourhood in the first graph and another in the second graph has about the size one would expect in two independent random graphs of the respective densities.

Definition 14 *(coquasirandom).* For $\alpha > 0$ and $L \in \mathbb{N}$, we say that a pair of graphs (F, F^*), both on the same vertex set V of order n and with densities p and p^*, respectively, is (α, L)-*coquasirandom* if for every set $S \subseteq V$ of at most L vertices and every subset $R \subseteq S$ we have

$$|\mathsf{N}_F(R) \cap \mathsf{N}_{F^*}(S \setminus R)| = (1 \pm \alpha)p^{|R|}(p^*)^{|S\setminus R|}n\,.$$

With this we can state the setting of our main lemmas and fix various constants which we will use in the remainder of the paper.

Setting 15. *Let $D, n \in \mathbb{N}$ and $\gamma > 0$ be given. We define*

$$\begin{aligned} \eta &= \frac{\gamma^D}{200D}\,, \quad \delta = \frac{\gamma^{10D}\eta}{10^6 D^4}\,, \quad C = 40D \exp\left(1000D\delta^{-2}\gamma^{-2D-10}\right), \\ \alpha_x &= \frac{\delta}{10^8 CD} \exp\left(\frac{10^8 CD^3\delta^{-1}(x-2n)}{n}\right) \qquad \text{for each } x \in \mathbb{R}, \\ \varepsilon &= \alpha_0\delta^2\gamma^{10D}/1000CD\,, \quad c = D^{-4}\varepsilon^4/100 \quad \text{and} \quad \xi = \alpha_0/100\,. \end{aligned} \tag{4.1}$$

Let $G_1, G_2, \dots, G_{s^*}$ (for some $s^* \le 2n$) be graphs on $[n]$, such that for each s and $x \in V(G_s)$ we have $\deg^-_{G_s}(x) \le D$, such that $\Delta(G_s) \le cn/\log n$, and such that the final δn vertices of G_s all have degree d_s and form an independent set.

Let H_0 and H_0^* be two edge-disjoint graphs on the same vertex set of order n such that (H_0, H_0^*) is $(\frac{1}{4}\alpha_0, 2D+3)$-coquasirandom, and $\sum_{s\in[s^*]} e(G_s) \le e(H_0) - \gamma n^2$.

Note that in (4.1) we give numbers α_x which we call ‘constant’ even though n appears in their definition. Observe that α_x is strictly increasing in x. We will be interested only in values $0 \le x \le 2n$ (though it is technically convenient to have the definition for all $x \in \mathbb{R}$), and it is easy to check that neither α_0 nor α_{2n} depends on n.

The main lemmas for the analysis of *PackingProcess* are now the following. Lemma 16 states that (H_0, H_0^*) is coquasirandom with high probability. Lemma 17 states that with high probability (H_s, H_0^*) continues to be coquasirandom for each stage s. To prove this lemma will be the main work of this paper. Lemma 18 states that, provided that H_s has the quasirandomness provided by Lemma 17, the *RandomEmbedding* of G_{s+1} into H_s is very likely to succeed. Lemma 19 states that with high probability very few edges of H_0^* are removed at each vertex to form H_s^*. This then implies that (H_s, H_s^*) is also likely to be coquasirandom (though with a much worse error parameter). Finally, in Lemma 20, using the coquasirandomness of (H_s, H_s^*), we argue that at each stage it is very likely that the completion phase is possible.

We start with the lemma concerning the coquasirandomness of the initial bulk and reservoir.

Lemma 16. *For each $D \in \mathbb{N}$ and each $\gamma > 0$, and for each n sufficiently large, let us suppose that the constants α_0 and ξ are as in Setting 15.*

Suppose that $\widehat{H}$ is a $(\xi, 2D+3)$-quasirandom graph of order n and density $p \geq 3\gamma$. Let H_0^* be a random subgraph of $\widehat{H}$ in which each edge of $\widehat{H}$ is kept with probability $q = \gamma/p$. Let H_0 be the complement of H_0^* in $\widehat{H}$. Then with probability at least $1 - n^{-6}$, we have that $e(H_0^*) = (1 \pm \alpha_0)\gamma\binom{n}{2}$ and the pair (H_0, H_0^*) is $\left(\frac{1}{4}\alpha_0, 2D+3\right)$-coquasirandom.

The next lemma states that coquasirandomness of (H_s, H_0^*) is preserved.

Lemma 17. *For each $D \in \mathbb{N}$ and each $\gamma > 0$, and for each n sufficiently large, the following holds with probability at least $1-n^{-5}$. Suppose that the constants and $G_1, G_2, \dots, G_{s^*}$ and the graph $H_0 \cup H_0^* = H$ are as in Setting 15. When* PackingProcess *is run, for each $s \in [s^*]$ either* PackingProcess *fails before completing stage s, or the pair (H_s, H_0^*) is $(\alpha_s, 2D+3)$-coquasirandom.*

The next lemma estimates the probability that a single execution of *RandomEmbedding* succeeds.

Lemma 18. *For each D, each $\gamma > 0$, and any sufficiently large n, let $\delta, \eta, \alpha_0, \alpha_{2n}, \varepsilon$ and c be as in Setting 15. Given any $\alpha_0 \leq \alpha \leq \alpha_{2n}$, let G be a graph on vertex set $[n]$ with maximum degree at most $cn/\log n$ such that $\deg^-(x) \leq D$ for each $x \in V(G)$, and let H be any $(\alpha, 2D+3)$-quasirandom n-vertex graph with at least $\gamma\binom{n}{2}$ edges. When* RandomEmbedding *is run then it fails with probability at most $2n^{-9}$.*

Our final two main lemmas concern the completion phase of *PackingProcess*. The first states that the completion phase is likely to delete very few edges at any vertex of H_0^*.

Lemma 19. *Given $D \in \mathbb{N}$ and $\gamma > 0$, let n be sufficiently large. Suppose that the constants and $G_1, G_2, \dots, G_{s^*}$ and H are as in Setting 15. When* PackingProcess *is run, with probability at least $1-n^{-50}$ one of the following three events occurs. First,* PackingProcess *fails. Second, there is some $s \in [s^*]$ such that (H_s, H_0^*) is not $(\alpha_s, 2D+3)$-coquasirandom. Third, for each $s \in [s^*]$ and $v \in V(H_s^*)$ we have $\deg_{H_0^*}(v) - \deg_{H_s^*}(v) \leq 50\gamma^{-D}D\delta n$, and (H_s, H_s^*) is $(\eta, 2D+3)$-coquasirandom.*

We will show in the proof of Theorem 11 that the first two events are unlikely, so that the likely event is the last.

Our last lemma states that with high probability, at any stage s, provided (H_{s-1}, H_{s-1}^*) is sufficiently coquasirandom, running *RandomEmbedding* to partially embed G_s into H_{s-1} is likely to give a partial embedding which can be completed to an embedding of G_s using H_s^*.

Lemma 20. *For each $D \in \mathbb{N}$ and each $\gamma > 0$, and for each n sufficiently large, let the constants be as in Setting 15. Suppose that G is a graph on $[n]$, such that we have $\deg^-(x) \leq D$ for each $x \in V(G)$, we have $\Delta(G) \leq cn/\log n$, and such that the final δn vertices of G form an independent set, and all have degree d. Suppose (H, H^*) are a pair* *of* $(\eta, 2D+3)$*-coquasirandom graphs on* n *vertices, and* H *is* $(\alpha_{s^*}, 2D+3)$*-quasirandom, with* $e(H) = p\binom{n}{2}$ *and* $e(H^*) = (1 \pm \eta)\gamma\binom{n}{2}$*, where* $p \geq \gamma$*. When* RandomEmbedding *is run to embed* $G[[n-\delta n]]$ *into* H*, with probability at least* $1 - 5n^{-9}$ *it returns a partial embedding* ϕ *which can be extended to an embedding* ϕ^* *of* G *into* $H \cup H^*$*, with all the edges using a vertex in* $\{n - \delta n + 1, \ldots, n\}$ *mapped to* H^**.*

Let us briefly explain why we cannot simply perform the whole embedding in the quasirandom $\widehat{H}$, but have to split it into a bulk and a reservoir. In order to analyse *RandomEmbedding*, we require that the bulk is very quasirandom, but *RandomEmbedding* is very well-behaved and preserves this good quasirandomness. In contrast, we are not able to show that the completion stage, where we choose a system of distinct representatives for the remaining vertices, is so well-behaved. If we used the bulk for this embedding the errors would rapidly become unacceptably large. However, to show that choosing such a system of distinct representatives is possible, we do not need much quasirandomness. Thus the reservoir H^*_s does rapidly lose its quasirandomness (compared to H_s), but it is sufficient for the completion.

We now argue that our main lemmas imply Theorem 11.

Proof of Theorem 11. We can assume that $p > 3\gamma$ as the statement is vacuous otherwise.

Suppose that we run *PackingProcess* on the input graphs $G_1, \ldots, G_{s^*}$. For the course of the analysis of this run, we shall first ignore possible failures during the completion phase. That is, if any failure during the completion phase occurs, we ignore it and continue embedding using *RandomEmbedding* into the bulk. Clearly, this does not change behaviour of future rounds of *RandomEmbedding* or the evolution of the bulk.

As we said earlier, we need to argue that with positive probability *PackingProcess* does not fail. Rather than proving this directly, we introduce additional quasirandomness conditions, and prove that with positive probability, all these conditions are satisfied up to any given stage, and that if we have the said quasirandomness conditions up to that stage, then *RandomEmbedding* will proceed successfully through the next stage. (Of course, it could happen that *PackingProcess* succeeds in the overall embedding even though some of our quasirandomness conditions failed during the course of the packing; we shall pessimistically view such an execution of *PackingProcess* as unsuccessful.) More precisely, it is clear that *PackingProcess* does not fail (in the *RandomEmbedding* stage) unless at least one of the following exceptional events occurs:

(i) (H_0, H^*_0) is not $(\frac{1}{4}\alpha_0, 2D+3)$-coquasirandom.
(ii) *RandomEmbedding* proceeded through stages $s = 1, \ldots, r$ (for some $r \in [s^* - 1]$) without failure, the pairs (H_s, H^*_0) are $(\alpha_s, 2D+3)$-coquasirandom for $s < r$, and (H_r, H^*_0) is not an $(\alpha_r, 2D+3)$-coquasirandom pair.
(iii) *RandomEmbedding* proceeded through stages $s = 1, \ldots, r$ (for some $r \in \{0, \ldots, s^* - 1\}$) without failure, the graphs H_s are $(\alpha_s, 2D+3)$-quasirandom for $s \leq r$. Then, in stage $r+1$, *RandomEmbedding* fails.

Lemma 16 gives an upper bound on the probability of the event in (i). Lemma 17 gives an upper bound on the probability of all the events in (ii). For each fixed $r \in \{0, \dots, s^* - 1\}$, the event in (iii) can be bounded using Lemma 18. Thus, the probability that *PackingProcess* fails in the *RandomEmbedding* part is at most $n^{-6} + n^{-5} + s^* \cdot 2n^{-9}$.

Let us now analyse the completion phases of *PackingProcess*. If *PackingProcess* fails in one of the completion phases then one of the following events occurs:

(iv) One of the events described under (i)-(iii).

(v) None of (i)-(iii) occurs. *RandomEmbedding* and the completion phase proceed successfully through the first r stages (for some $r \in \{1, \dots, s^* - 1\}$. For $s \in [r]$ all the pairs (H_s, H_0^*) are $(\alpha_s, 2D+3)$-coquasirandom. However, there is a stage $s \in [r]$ where (H_s, H_s^*) is not $(\eta, 2D+3)$-coquasirandom.

(vi) None of (i)-(iii) occurs. *RandomEmbedding* and the completion phase proceeds successfully through the first r stages (for some $r \in \{0, \dots, s^* - 1\}$, and throughout all the pairs (H_s, H_0^*) and (H_s, H_s^*) are $(\alpha_s, 2D+3)$-coquasirandom and $(\eta, 2D+3)$-coquasirandom, respectively. In stage $r+1$, *RandomEmbedding* successfully embeds but the completion phase fails.

Lemma 19 bounds the probability of the event in (v) by n^{-50}. Finally, Lemma 20 bounds the probability of events in (vi) for each given r by $5n^{-9}$. Thus, the total probability of failure due to (v) or (vi) is at most $n^{-50} + s^* \cdot 5n^{-9}$.

We conclude that *PackingProcess* packs the graphs $G_1, \dots, G_{s^*}$ into $\widehat{H}$ with positive probability. □

4.1. The probability space for RandomEmbedding

Algorithm 2 gives a sound definition of a randomised algorithm which either provides an embedding of $G_{[n-\delta n]}$ into H or fails, and the probability of any output can be in principle computed. To handle the analysis of *RandomEmbedding*, which is the most demanding part of this paper, it is useful to properly set up a probability space as indicated at the beginning of Section 2.2.2. Given G and H as in Algorithm 2 (recall that $V(G) = [n]$), let $\Omega^{G \hookrightarrow H} := (V(H) \cup \{☹\})^{n-\delta n}$. We now need to define the probability measure on $\Omega^{G \hookrightarrow H}$. Let $\boldsymbol{\omega} = (\omega_1, \dots, \omega_{n-\delta n}) \in \Omega^{G \hookrightarrow H}$ be given. Suppose first that $\boldsymbol{\omega}$ consists only of vertices of $V(H)$. Then we define $\mathbb{P}^{G \hookrightarrow H}(\boldsymbol{\omega})$ as the probability that *RandomEmbedding* succeeds embedding $G_{[n-\delta n]}$ into H, and maps each vertex $t \in [n-\delta n]$ of G on vertex ω_t. Suppose next that $\boldsymbol{\omega}$ contains some ☹'s, and that these form a terminal segment of $\boldsymbol{\omega}$, say starting from position t_0. Then we define $\mathbb{P}^{G \hookrightarrow H}(\boldsymbol{\omega})$ as the probability that *RandomEmbedding* succeeds in the first $t_0 - 1$ steps, and for each $t \in [t_0 - 1]$ it maps vertex t on ω_t, and then in step t it halts with failure. Last, suppose that $\boldsymbol{\omega}$ contains some ☹'s but these do not form a terminal segment of $\boldsymbol{\omega}$. We then define $\mathbb{P}^{G \hookrightarrow H}(\boldsymbol{\omega}) := 0$. It is clear that $\mathbb{P}^{G \hookrightarrow H}(\boldsymbol{\omega})$ is a probability measure on $\Omega^{G \hookrightarrow H}$ which corresponds to possible runs of *RandomEmbedding*.

We shall use the concept of histories and history ensembles, as introduced in Section 2.2.2, in connection with $\Omega^{G\hookrightarrow H}$.

4.2. Organisation of the technical part of the paper

It thus remains to prove all the main lemmas from this section. Lemmas 16 and 17 are proven in Section 6. Lemma 18 is stated here in a simplified form. In actuality, we prove a stronger statement (of which Lemma 18 is a straightforward consequence) in Lemma 24. This stronger form is also needed for proving Lemma 17, and its proof spans the entire Section 5. Lemmas 19 and 20 are proven in Section 7.

5. Staying on a diet

In this section we consider the running of *RandomEmbedding* to embed one degenerate graph G into a quasirandom graph H. The results of this section will always be used to analyse one stage s, when we take $G = G_s$ and $H = H_{s-1}$. We also analyse how *RandomEmbedding* behaves with respect to the graph $H^* = H^*_{s-1}$. We analyse carefully how fast common neighbourhoods of vertices in H are eaten up by *RandomEmbedding*, and how often individual vertices of H appear in candidate sets. To make this precise, we introduce the following two definitions.

The diet condition states that during the running of *RandomEmbedding*, for each $t \in [n-\delta n]$, the fraction of each set $\mathsf{N}_H(S)$ which is covered by $\mathrm{im}(\psi_t)$ is roughly as expected, that is, roughly proportional to $|\mathrm{im}(\psi_t)|/n$. As with (co)quasirandomness, we also require a codiet condition, considering the intersection of some vertex neighbourhoods in H and H^*.

Definition 21 *(diet condition, codiet condition).* Let H be a graph with n vertices and $p\binom{n}{2}$ edges, and let $X \subseteq V(H)$ be any vertex set. We say that the pair (H, X) satisfies the (β, L)-*diet condition* if for every set $S \subseteq V(H)$ of at most L vertices we have $|\mathsf{N}_H(S) \setminus X| = (1 \pm \beta)p^{|S|}(n - |X|)$.

Let H, H^* be two graphs with vertex set V of order n and $p\binom{n}{2}$ and $p^*\binom{n}{2}$ edges, respectively, and let $X \subseteq V$ be any vertex set. We say that the triple (H, H^*, X) satisfies the (β, L)-*codiet condition* if for every set $S \subseteq V$ of at most L vertices and for every subset $R \subseteq S$ we have

$$\Big|\big(\mathsf{N}_H(R) \cap \mathsf{N}_{H^*}(S \setminus R)\big) \setminus X\Big| = (1 \pm \beta)p^{|R|}(p^*)^{|S\setminus R|}(n - |X|) .$$

Observe that the (β, L)-diet condition holding for $(H, \emptyset)$ is simply the statement that H is (β, L)-quasirandom, and similarly for the codiet condition.

The cover condition, defined below, roughly states that for each v in the host graph H during the embedding of G into H by *RandomEmbedding*, the right fraction of vertices x of G have v in their final candidate set. For making precise what we mean by ‘the right fraction' some care is needed. Firstly, how likely it is that v is in the final candidate set of x depends on the number neighbours of x preceding x. Therefore we will partition $V(G)$ according to this number of previous neighbours. For technical reasons we actually further want to control this fraction in intervals of $V(G)$ of length εn, where n is the order of H. Hence we define for a given $\varepsilon > 0$ the set

$$X_{i,d} := \{x \in V(G) \colon i \le x < i + \varepsilon n, |\mathsf{N}^-(x)| = d\} .$$

When G is given with a D-degenerate ordering it is enough to consider $d \in \{0, 1, \dots, D\}$. So if H is quasirandom and has $p\binom{n}{2}$ edges, then for an arbitrary $v \in V(H)$, we would expect that about a p^d-fraction of vertices x in each $X_{i,d}$ have v in their final candidate sets (let us remind that the candidate set may include also vertices used by the embedding).

Definition 22 *(cover condition).* Suppose that G and H are two graphs such that H has order n, the vertex set of G is $[n]$, and H has density p. Suppose that numbers $\beta, \varepsilon > 0$ and $i \in [n - \varepsilon n]$ are given. We say that a partial embedding ψ of G into H, which embeds $\mathsf{N}^-(x)$ for each $i \le x < i + \varepsilon n$, satisfies the (ε, β, i)-*cover condition* if for each $v \in V(H)$, and for each $d \in \mathbb{N}$, if we have

$$\left|\left\{x \in X_{i,d} : v \in \mathsf{N}_H\big(\psi(\mathsf{N}^-(x))\big)\right\}\right| = (1 \pm \beta)p^d |X_{i,d}| \pm \varepsilon^2 n .$$

Note that a corresponding condition for $d = 0$ is trivial, even with zero error parameters.

We use Definitions 14, 21 and 22, to define key events $\mathsf{DietE}(\cdot;\cdot)$, $\mathsf{CoverE}(\cdot;\cdot)$, $\mathsf{CoDietE}(\cdot)$ on $\Omega^{G \hookrightarrow H}$.

Definition 23. Suppose that D, δ and ε are as in Setting 15. Suppose that $\lambda > 0$. Suppose that we have graphs G and H as in Algorithm 2. Suppose that we run *RandomEmbedding* to partially embed G into H. Let $(\psi_i)_{i \in [t_*]}$ be the partial embeddings of $G\big[[i]\big]$ into H, where $t_* = n - \delta n$ if *RandomEmbedding* succeeded, and otherwise $t_* + 1$ is the step in which *RandomEmbedding* halted with failure.

- For each $t \in [n - \delta n]$, let $\mathsf{DietE}(\lambda; t) \subseteq \Omega^{G \hookrightarrow H}$ correspond to executions of *RandomEmbedding* for which $t_* \ge t$ and the pair $(H, \operatorname{im} \psi_t)$ satisfies the $(\lambda, 2D + 3)$-diet condition.
- For each $t \in [n - \delta n]$, let $\mathsf{CoverE}(\lambda; t) \subseteq \Omega^{G \hookrightarrow H}$ correspond to executions of *RandomEmbedding* for which $t_* \ge t + \varepsilon n$ and the embedding ψ_{t^*} of G into H satisfies the $(\varepsilon, \lambda, t)$-cover condition.
- Suppose further that we have a graph H^* with $V(H) = V(H^*)$. For each $t \in [n - \delta n]$, let $\mathsf{CoDietE}(t) \subseteq \Omega^{G \hookrightarrow H}$ correspond to executions of *RandomEmbedding* for which $t_* \ge t$ and the triple $(H, H^*, \operatorname{im} \psi_t)$ satisfies the $(2\eta, 2D + 3)$-codiet condition.

Note that the events $\mathsf{DietE}(\cdot;t)$ and $\mathsf{CoDietE}(t)$ are determined by histories (as defined in Sections 2.2.2 and 4.1) up to time t. That is, for any $\lambda > 0$ and any history $\mathscr{H}_t$, we have that $\mathsf{DietE}(\lambda;t)$ either contains $\mathscr{H}_t$ or is disjoint from $\mathscr{H}_t$. We have similar the same property for $\mathsf{CoDietE}(t)$. The event $\mathsf{CoverE}(\cdot;t)$ is somewhat different since its definition involves the set $X_{t,d}$ which looks $\varepsilon n - 1$ many steps forward in time. So, for any history $\mathscr{H}_{t+\varepsilon n-1}$, we have that $\mathsf{CoverE}(\lambda;t)$ either contains $\mathscr{H}_{t+\varepsilon n-1}$ or is disjoint from $\mathscr{H}_{t+\varepsilon n-1}$.

The following lemma is the crucial accurate analysis of *RandomEmbedding* which we need in order to show that *RandomEmbedding* is likely to succeed and in order to derive further properties of the final embedding.

Lemma 24 *(Diet-and-cover lemma). For each $D \in \mathbb{N}$, each $\gamma > 0$, and any sufficiently large n, let $\delta, \eta, \alpha_0, \alpha_{2n}, \varepsilon$ and c, C be as in Setting 15. Let $\alpha \in [\alpha_0, \alpha_{2n}]$ be arbitrary. Let G be a graph on vertex set $[n]$ with maximum degree at most $cn/\log n$ such that $\deg^-(x) \le D$ for each $x \in V(G)$, and let H be any $(\alpha, 2D+3)$-quasirandom n-vertex graph with at least $\gamma\binom{n}{2}$ edges. Suppose in addition that H^* is a graph on $V(H)$ such that (H, H^*) is $(\eta, 2D+3)$-coquasirandom. Then we have*

$$\mathbb{P}^{G\hookrightarrow H}\left(\bigcap_{t\in[n-\delta n]} \mathsf{DietE}(C\alpha;t) \cap \bigcap_{t\in[n+1-\varepsilon n]} \mathsf{CoverE}(C\alpha;t) \cap \bigcap_{t\in[n-\delta n]} \mathsf{CoDietE}(t)\right)$$
$$\ge 1 - 2n^{-9}\,. \tag{5.1}$$

This lemma immediately implies Lemma 18.

Proof of Lemma 18. Recall that *RandomEmbedding* fails if and only if $C^{t-1}_{G\hookrightarrow H}(t) \setminus \mathrm{im}(\psi_{t-1}) = \emptyset$ for some t, and $\mathsf{DietE}(C\alpha;t-1)$ in particular gives a formula lower bounding the size of $C^{t-1}_{G\hookrightarrow H}(t) \setminus \mathrm{im}(\psi_{t-1})$ which is greater than 0. Since the likely event of Lemma 24 is contained in $\mathsf{DietE}(C\alpha;t-1)$ for each $t \ge 2$, and the same lower bound is trivially implied by $(\alpha, 2D+3)$-quasirandomness of H for $t = 1$ (since $\mathrm{im}\,\psi_0 = \emptyset$), we conclude that within the likely event of Lemma 24, *RandomEmbedding* does not fail. □

The main difficulty is to establish that the cover and diet conditions hold. We will see that the codiet condition is an easy byproduct. The reason for the difficulty is that the error terms in the cover and diet conditions for small times t feed back into the calculations which will establish the cover and diet conditions for larger times t, and we have to ensure that this feedback loop does not allow the errors to spiral out of control. To that end, we define a new sequence of error terms, which we need only in the proof of Lemma 24. The following constants $\{\beta_t : t \in \mathbb{R}\}$ are a carefully chosen increasing sequence (depending on α) such that $\beta_0 = \alpha$ and such that β_n/β_0 is bounded by a constant which does not depend on α (though it does depend on D, γ and δ). Given D and $\alpha, \delta, \gamma > 0$, we define

$$\beta_t := 2\alpha \exp\left(\tfrac{1000D\delta^{-2}\gamma^{-2D-10}t}{n}\right)\,. \tag{5.2}$$

We will mainly take t integer in the range $[0, n]$, but it is convenient to allow t to be any real number. In particular, for each $t \geq 0$, we have

$$\frac{1}{n}\int_{i=0}^{t} 1000D\delta^{-2}\gamma^{-2D-10}\beta_i \,\mathrm{d}i \leq 2\alpha \int_{i=-\infty}^{t} \frac{1000D\delta^{-2}\gamma^{-2D-10}}{n} \exp\left(\tfrac{1000D\delta^{-2}\gamma^{-2D-10}i}{n}\right) \mathrm{d}i = \beta_t\,. \tag{5.3}$$

Suppose that we have Setting 15, and suppose that $\alpha \geq \alpha_0$ is given. Then for each $t \geq 0$ we have

$$\beta_t\gamma^{2D+3}\delta \geq \beta_0\gamma^{2D+3} > \varepsilon\,. \tag{5.4}$$

We split the proof of Lemma 24 into two parts. The cover lemma (Lemma 25) states that if the $(\beta_t, 2D+3)$-diet condition holds for $(H, \operatorname{im}\psi_i)$ for each $i \in [t-1]$, then it is very unlikely that the $(\varepsilon, 20D\beta_t, t)$-cover condition fails for $\psi_{t+\varepsilon n-2}$. Note that the time $t+\varepsilon n-2$ is the first time at which the $(\varepsilon, 20D\beta_t, t)$-cover condition is guaranteed to be determined, since at this time all left-neighbours of all vertices $t, t+1, \dots, t+\varepsilon n-1$ have certainly been embedded.

Lemma 25 *(Cover lemma). For each D, each $\gamma > 0$ and sufficiently large n, let $\alpha_0, \alpha_{2n}, \varepsilon, \delta$ and c be as in Setting 15. Suppose that $\alpha_0 \leq \alpha \leq \alpha_{2n}$ and G is a graph on vertex set $[n]$, with $\deg^-(x) \leq D$ for each $x \in [n]$, with maximum degree at most $cn/\log n$, and suppose that H is an n-vertex graph of density at least γ. Let β_t for $0 \leq t \leq n$ be defined as in (5.2) and assume that $\beta_n \leq \frac{1}{10}$. Let t with $1 \leq t \leq n-\delta n-\varepsilon n+1$ be fixed.*

Then we have

$$\mathbb{P}^{G\hookrightarrow H}\left(\bigcap_{i=1}^{t-1} \mathsf{DietE}(\beta_t; i) \setminus \mathsf{CoverE}(20D\beta_t; t)\right) \leq n^{-10}\,.$$

Let us consider Setting 15. Suppose that for some $0 \leq t \leq n-\delta n-\varepsilon n$, *RandomEmbedding* runs up to time t and the $(\beta_t, 2D+3)$-diet condition holds for $(H, \operatorname{im}\psi_t)$. Let $p := e(H)/\binom{n}{2}$ and suppose that $p \geq \gamma$. Then for each $t+1 \leq j \leq t+\varepsilon n$, and each set $S \subseteq V(H)$ of at most $2D+3$ vertices, we have

$$\begin{aligned}
|\mathsf{N}_H(S) \setminus \operatorname{im}\psi_j| &\geq |\mathsf{N}_H(S) \setminus \operatorname{im}\psi_t| - \varepsilon n \\
{\scriptstyle (\text{diet for } (H, \operatorname{im}\psi_t))} \quad &\geq (1-\beta_t)p^{|S|}(n - |\operatorname{im}\psi_t|) - \varepsilon n \\
{\scriptstyle (\varepsilon < \beta_t\gamma^{2D+3}\delta \text{ by } (5.4))} \quad &\geq (1-2\beta_t)p^{|S|}(n - |\operatorname{im}\psi_t|)\,.
\end{aligned}$$

Hence, the $(2\beta_t, 2D+3)$-diet condition holds deterministically for $(H, \operatorname{im}\psi_j)$. In particular *RandomEmbedding* cannot fail before time $t+\varepsilon n$.

The diet lemma (Lemma 26) states that when the $(\beta_i, 2D+3)$-diet condition holds for $(H, \operatorname{im}\psi_i)$ for each $i \in [t-1]$, and the $(\varepsilon, 20D\beta_i, i)$-cover condition holds for $\psi_{i+\varepsilon n-2}$ for each $i \in [t+1-\varepsilon n]$, then it is unlikely that the $(\beta_t, 2D+3)$-diet condition fails for $(H, \operatorname{im}\psi_t)$. We also obtain the desired codiet condition.

Lemma 26 *(Diet lemma). For each D, each $\gamma > 0$, and any sufficiently large n, let $\alpha_0, \alpha_{2n}, \varepsilon, \delta$ and η be as in Setting 15. For any $t \le (1-\delta)n$, and $\alpha_0 \le \alpha \le \alpha_{2n}$ the following holds. Suppose that G is a graph on $[n]$ such that $\deg^-(x) \le D$ for each $x \in [n]$, and H is an $(\alpha, 2D+3)$-quasirandom graph with n vertices with $p\binom{n}{2}$ edges, with $p \ge \gamma$. Suppose furthermore that H^* is a graph on $V(H)$ and $\hat{p}\binom{n}{2}$ edges with $\hat{p} \ge (1-\eta)\gamma$, such that (H, H^*) satisfies the $(\eta, 2D+3)$-coquasirandomness condition. Let $\{\beta_\tau : \tau \in [0,n]\}$ be defined as in (5.2) and assume that $\beta_n \le \frac{1}{10}$. Let t with $1 \le t \le n - \delta n$ be fixed.*

Then we have

$$\mathbb{P}^{G\hookrightarrow H}\left(\bigcap_{j=1}^{t-1} \mathsf{DietE}(\beta_j; j) \cap \bigcap_{j=1}^{t+1-\varepsilon n} \mathsf{CoverE}(20D\beta_j; j) \setminus (\mathsf{DietE}(\beta_t; t) \cap \mathsf{CoDietE}(t))\right)$$
$$\le n^{-10}\,.$$

Since the graphs G and H are fixed in Lemmas 24, 25, and 26, in this section we drop the subscript in the notation $C^j_{G\hookrightarrow H}(x)$ and write simply $C^j(x)$. Likewise, we write $\mathbb{P}$ instead of $\mathbb{P}^{G\hookrightarrow H}$. Last, we write $(\psi_i)_{i\in t_*}$ for partial embeddings of G into H; here t_* is the time at which *RandomEmbedding* halts. Of course, t_* and $(\psi_i)_{i\in t_*}$ depend on a particular realization $\boldsymbol{\omega} \in \Omega^{G\hookrightarrow H}$ of the run of *RandomEmbedding*.

We now show that Lemmas 25 and 26, whose proofs are deferred to later in this section, imply Lemma 24.

Proof of Lemma 24. Suppose that we are given D and γ. Now, given $\alpha > 0$, we define β_t for each $0 \le t \le n$ as in (5.2). For $t = 0, \dots, n - \delta n$, define

$$\mathcal{A}_t := \bigcap_{j=1}^{t} \mathsf{DietE}(\beta_j; j) \cap \bigcap_{j=\varepsilon n}^{t} \mathsf{CoverE}(20D\beta_{t-\varepsilon n+1}; j-\varepsilon n+1) \cap \bigcap_{j=1}^{t} \mathsf{CoDietE}(j)\,. \quad (5.5)$$

Our strategy is first to show that $\mathbb{P}(\mathcal{A}_{t-1} \setminus \mathcal{A}_t)$ is tiny for each t. Since $\mathbb{P}(\mathcal{A}_0) = 1$, this will imply that $\mathbb{P}(\mathcal{A}_{n-\delta n})$ is very close to 1. Last, we shall show that $\mathcal{A}_{n-\delta n}$ is a subset of the event in (5.1).

Indeed, suppose that the event $\mathcal{A}_{t-1}$ holds. This in particular means that the $(\beta_j, 2D+3)$-diet condition holds for $(H, \operatorname{im}\psi_j)$ for each $1 \le j < t$, and the $(\varepsilon, 20D\beta_{j-\varepsilon n+1}, j-\varepsilon n+1)$-cover condition holds for ψ_j for each $\varepsilon n - 1 \le j < t$.

Because the $(\beta_{t-1}, 2D+3)$-diet condition holds for $(H,\ \psi_{t-1})$, picking $S = \psi_{t-1}(\mathsf{N}^-(t))$, we have $\left|C^{t-1}(t) \setminus \operatorname{im}\psi_{t-1}\right| = \left|\mathsf{N}_H(S) \setminus \operatorname{im}\psi_{t-1}\right| > 0$. It follows that *RandomEmbedding* cannot fail at time t.

Firstly, let us focus on the term $\mathsf{CoverE}(20D\beta_{t-\varepsilon n+1}; t-\varepsilon n+1)$ in (5.5). This term does not exist when $t < \varepsilon n$, so let us assume the contrary. Lemma 25 then tells us that

$$\mathbb{P}\left(\bigcap_{i=1}^{t-\varepsilon n} \mathsf{DietE}(\beta_{t-\varepsilon n+1}; i) \setminus \mathsf{CoverE}(20D\beta_{t-\varepsilon n+1}; t-\varepsilon n+1)\right) \leq n^{-10} .$$

In particular,

$$\mathbb{P}(\mathcal{A}_{t-1} \setminus \mathsf{CoverE}(20D\beta_{t-\varepsilon n+1}; t-\varepsilon n+1)) \leq n^{-10} . \tag{5.6}$$

Secondly, we use Lemma 26 to show that with high probability neither the diet condition nor the codiet condition fails at time t. Indeed, Lemma 26 tells us that

$$\mathbb{P}\left(\bigcap_{j=1}^{t-1} \mathsf{DietE}(\beta_j; j) \cap \bigcap_{j=\varepsilon n}^{t-1} \mathsf{CoverE}(20D\beta_{j+1-\varepsilon n}; j+1-\varepsilon n) \setminus (\mathsf{DietE}(\beta_t; t) \cap \mathsf{CoDietE}(t))\right)$$
$$\leq n^{-10}$$

In particular,

$$\mathbb{P}(\mathcal{A}_{t-1} \setminus (\mathsf{DietE}(\beta_t; t) \cap \mathsf{CoDietE}(t))) \leq n^{-10} . \tag{5.7}$$

Summing up (5.6) and (5.7), we conclude that $\mathbb{P}(\mathcal{A}_{t-1} \setminus \mathcal{A}_t) \leq 2n^{-10}$. Taking a union bound over the at most n choices of t, we see that with probability at least $1 - 2n^{-9}$ the good event from the statement of Lemma 24 holds, i.e., that *RandomEmbedding* does not fail, and by the choice of C and by (5.2), for each $1 \leq t \leq (1-\delta)n$ the pair $(H, \operatorname{im}\psi_t)$ satisfies the $(C\alpha, 2D+3)$-diet condition and the triple $(H, H^*, \operatorname{im}\psi_t)$ satisfies the $(2\eta, 2D+3)$-codiet condition, and for each $1 \leq t \leq n+1-\varepsilon n$ the embedding $\psi_{(1-\delta)n}$ satisfies the $(\varepsilon, C\alpha, t)$-cover condition, as desired. $\square$

We now prove the cover lemma.

Proof of Lemma 25. Let $e(G) = p\binom{n}{2} \geq \gamma\binom{n}{2}$. Let $\mathcal{D}$ be the event that the $(\beta_t, 2D+3)$-diet condition holds for each $(H, \operatorname{im}\psi_i)$ with $1 \leq i \leq t-1$, $\mathcal{D} := \bigcap_{i=1}^{t-1} \mathsf{DietE}(\beta_t; i)$. We fix a vertex $v \in V(H)$. We also fix $1 \leq d \leq D$. Define $\mathcal{B}_{v,d}$ as the event that $\mathcal{D}$ holds, and that v and d witness the failure of the $(\varepsilon, 20D\beta_t, t)$-cover condition for $\psi_{t+\varepsilon n-2}$. More formally,

$$\mathcal{B}_{v,d} := \mathcal{D} \cap \Big\{\boldsymbol{\omega} \in \Omega^{G\hookrightarrow H} :$$
$$\big|\big\{x \in X_{t,d} : v \in \mathsf{N}_H\big(\psi_{t+\varepsilon n-2}(\mathsf{N}^-(x))\big)\big\}\big| \neq (1 \pm 20D\beta_t)p^d|X_{t,d}| \pm \varepsilon^2 n\Big\} .$$

Our aim is to show that

$$\mathbb{P}(\mathcal{B}_{v,d}) \leq n^{-12}/D . \tag{5.8}$$

A union bound over the choices of v and d then gives the lemma.

Our strategy for proving (5.8) is as follows. Ideally, we would like to assert that for each $x \in X_{t,d}$ the probability of $v \in C^{x-1}(x)$ is roughly p^d and apply Lemma 4 to bound the probability of the bad event $\mathcal{B}_{v,d}$. To this end, we consider a dynamical version of candidate sets, where we track changes in the set potentially suitable to accommodate x as we gradually embed more and more left-neighbours of x. More precisely, for each $i \leq x-1$, let $C^{i,\mathsf{dyn}}(x) := \mathsf{N}_H\left(\psi_{x-1}\big([i] \cap \mathsf{N}^-_G(x)\big)\right)$. At time $i=0$, we have $v \in C^{i,\mathsf{dyn}}(x)$, and as i increases, the set $C^{i,\mathsf{dyn}}(x)$ shrinks exactly at times $y \in \mathsf{N}^-(x)$ when left-neighbours of x are embedded.

Unfortunately we are not able to carry out this ideal strategy, because when we apply Lemma 4 what we need to calculate is not the probability of $v \in C^{x-1}(x)$, but this probability in the conditioned space given by the history up to some earlier time. Because the sets $\mathsf{N}^-(x)$ interleave each other, this conditional probability will generally not be close to p^d and we were not able to find a good way to estimate it. Hence we refine this strategy by rewriting the event $\{v \in C^{x-1}(x)\}$ as

$$\bigcap_{k=1}^{d} \{y_1, y_2, \ldots, y_k \hookrightarrow \mathsf{N}_H(v)\}\,, \tag{5.9}$$

where $y_1, \ldots, y_d$ are the neighbours of x, ordered from left to right. The event $\{y_1, y_2, \ldots, y_d \hookrightarrow \mathsf{N}_H(v)\}$, of course, equals the entire intersection (5.9). However, this more complicated way of expressing (5.9) suggests to introduce, for each k, a sequence of random variables that count the events of the form $\{y_1, y_2, \ldots, y_k \hookrightarrow \mathsf{N}_H(v)\}$, ordered by y_k. Intuitively, conditioning on $\{y_1, y_2, \ldots, y_k \hookrightarrow \mathsf{N}_H(v)\}$ holding (which is determined by the history up to the time at which we embed y_k) we should expect that the probability that $\{y_1, y_2, \ldots, y_{k+1} \hookrightarrow \mathsf{N}_H(v)\}$ holds is about p. We will be able to demonstrate this is true, even if we condition on a typical history up to the time immediately before embedding y_{k+1}, and this allows us to use Lemma 4.

More formally, given $1 \leq k \leq d$ and $y \in V(G)$, we define random variables $Y_{k,1}, \ldots, Y_{k,t+\varepsilon n-2}$ as follows. Let $Y_{k,y}$ be the number of vertices $x \in X_{t,d}$ such that y is the k-th leftmost vertex of $\mathsf{N}^-(x)$ and the first k vertices of $\mathsf{N}^-(x)$ are all embedded to $\mathsf{N}_H(v)$. Further, for each $0 \leq k \leq d$, we let $\mathcal{Y}_k$ be the event that $(1 \pm 10\beta_t)^k p^k |X_{t,d}| \pm k\varepsilon^2 n/d$ vertices $x \in X_{t,d}$ have all of the first k vertices of $\mathsf{N}^-(x)$ embedded to $\mathsf{N}_H(v)$. Observe that the event $\mathcal{Y}_k$ is precisely the statement that

$$\sum_{y=1}^{t+\varepsilon n-2} Y_{k,y} = (1 \pm 10\beta_t)^k p^k |X_{t,d}| \pm k\varepsilon^2 n/d\,. \tag{5.10}$$

Our bad event then satisfies

$$\mathcal{B}_{v,d} \subseteq \mathcal{D} \setminus \mathcal{Y}_d\,,$$

because $(1 \pm 10\beta_t)^d = 1 \pm 20D\beta_t$. In order to bound the probability of $\mathcal{B}_{v,d}$ we cover $\mathcal{B}_{v,d}$ with d events, each of whose probabilities we can bound with Lemma 4. For this purpose we define the event

$$\mathcal{E}_k = \mathcal{Y}_{k-1} \cap \mathcal{D}$$

for each $1 \le k \le d$. Note that $\mathcal{E}_1 = \mathcal{D}$ since $\mathcal{Y}_0$ holds trivially with probability one. We thus have

$$\mathcal{B}_{v,d} \subseteq \mathcal{D} \setminus \mathcal{Y}_d \subseteq \bigcup_{1 \le k \le d} \left(\mathcal{E}_k \setminus \mathcal{Y}_k\right).$$

Our aim then is to show that for each $1 \le k \le d$ we have

$$\mathbb{P}(\mathcal{E}_k \setminus \mathcal{Y}_k) \le n^{-12}/(d \cdot D). \tag{5.11}$$

Note that this and a union bound over the d choices of k gives (5.8).

To establish (5.11) we would like to apply Lemma 4. Hence we need to argue that either $\mathcal{E}_k$ fails, or we can estimate $\sum_{y=1}^{t+\varepsilon n-2} \mathbb{E}\left(Y_{k,y} | \mathscr{H}_{y-1}\right)$, where $\mathscr{H}_{y-1}$ is the history of embedding decisions taken in *RandomEmbedding* up to and including the embedding of vertex $y-1$. To this end, for $y \in [t+\varepsilon n-2]$ let $Z_{k,y}$ be the number of vertices $x \in X_{t,d}$ such that y is the k-th leftmost vertex of $\mathsf{N}^-(x)$ and the first $k-1$ vertices of $\mathsf{N}^-(x)$ are embedded to $\mathsf{N}_H(v)$. Then the quantity $Z_{k,y}$ is determined by $\mathscr{H}_{y-1}$ and

$$\mathbb{E}\left(Y_{k,y} | \mathscr{H}_{y-1}\right) = Z_{k,y} \cdot \mathbb{P}\big(y \hookrightarrow \mathsf{N}_H(v) | \mathscr{H}_{y-1}\big). \tag{5.12}$$

Observe further that

$$\sum_{y=1}^{t+\varepsilon n-2} Z_{k,y} = \sum_{y=1}^{t+\varepsilon n-2} Y_{k-1,y}, \tag{5.13}$$

because both sums count the number of vertices $x \in X_{t,d}$ such that the first $k-1$ vertices of $\mathsf{N}^-(x)$ are embedded to $\mathsf{N}_H(v)$, in the first sum grouped by their k-th left neighbour, and in the second sum by their $(k-1)$-st left neighbour.

Assume now that $y \in V(G)$ is fixed and that $\mathscr{H}_{y-1}$ is such that $\mathscr{H}_{y-1} \cap \mathcal{E}_k \neq \emptyset$, and let us bound $\mathbb{P}\big(y \hookrightarrow \mathsf{N}_H(v) | \mathscr{H}_{y-1}\big)$. Since $\mathscr{H}_{y-1} \cap \mathcal{E}_k \neq \emptyset$ and $\mathcal{D} \supseteq \mathcal{E}_k$, by definition of $\mathcal{D}$ the $(\beta_t, 2D+3)$-diet condition holds for $(H, \operatorname{im}\psi_{y-\varepsilon n})$, where we have to subtract εn in the index of $\psi_{y-\varepsilon n}$ because y could be as large as $t+\varepsilon n-2$ (and we only know that the diet condition holds up to time $t-1$). This implies that for each set S of vertices in H with $|S| \le 2D+3$ we have

$$\begin{aligned}\big|\mathsf{N}_H(S) \setminus \operatorname{im}\psi_{y-1}\big| &= (1 \pm \beta_t)p^{|S|}(n-y+\varepsilon n) \pm \varepsilon n \\ &= (1 \pm \beta_t)p^{|S|}(n-y+1) \pm 2\varepsilon n = (1 \pm 2\beta_t)p^{|S|}(n-y+1),\end{aligned}$$

where the last inequality follows from $\gamma \le p$ and $\varepsilon \le \alpha\gamma^{2D+3} \le \frac{1}{2}\beta_t\gamma^{2D+3}$. We conclude that the $(2\beta_t, 2D+3)$-diet condition holds for $(H, \operatorname{im}\psi_{y-1})$. Since $\deg^-(y) \le D$ it follows that

$$\big|C^{y-1}(y) \setminus \operatorname{im}\psi_{y-1}\big| = (1 \pm 2\beta_t)p^{\deg^-(y)}(n-y+1) \quad \text{and}$$
$$\big|\mathsf{N}_H(v) \cap C^{y-1}(y) \setminus \operatorname{im}\psi_{y-1}\big| = (1 \pm 2\beta_t)p^{1+\deg^-(y)}(n-y+1)\,.$$

Therefore we have

$$\mathbb{P}\big(y \hookrightarrow \mathsf{N}_H(v)|\mathscr{H}_{y-1}\big) = \frac{\big|\mathsf{N}_H(v) \cap C^{y-1}(y) \setminus \operatorname{im}\psi_{y-1}\big|}{\big|C^{y-1}(y) \setminus \operatorname{im}\psi_{y-1}\big|} = (1 \pm 10\beta_t)p\,.$$

We conclude from (5.12) that

$$\sum_{y=1}^{t+\varepsilon n-2} \mathbb{E}(Y_{k,y}|\mathscr{H}_{y-1}) = (1 \pm 10\beta_t)p \sum_{y=1}^{t+\varepsilon n-2} Z_{k,y}\,, \tag{5.14}$$

unless $\mathcal{E}_k$ fails. Further, unless $\mathcal{E}_k$ fails, we have

$$\sum_{y=1}^{t+\varepsilon n-2} Z_{k,y} \overset{(5.13)}{=} \sum_{y=1}^{t+\varepsilon n-2} Y_{k-1,y} \overset{(5.10)}{=} (1 \pm 10\beta_t)^{k-1}p^{k-1}|X_{t,d}| \pm (k-1)\varepsilon^2 n/d\,.$$

Plugging this in (5.14), we get that $\mathcal{E}_k$ fails or we have

$$\sum_{y=1}^{t+\varepsilon n-2} \mathbb{E}\left(Y_{k,y}|\mathscr{H}_{y-1}\right) = (1 \pm 10\beta_t)^k p^k |X_{t,d}| \pm (k-1)\varepsilon^2 n/d\,.$$

Since $0 \le Y_{k,y} \le \deg(y)$ for each y, we can thus apply Lemma 4 with the event $\mathcal{E} = \mathcal{E}_k$, with $\mu \pm \nu = (1 \pm 10\beta_t)^k p^k |X_{t,d}| \pm (k-1)\varepsilon^2 n/d$, and with $\varrho = \varepsilon^2 n/d$ to conclude that

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_k \text{ and not } \mathcal{Y}_k\right) &= \mathbb{P}\left(\mathcal{E}_k \text{ and } \sum_{y=1}^{t+\varepsilon n-2} Y_{k,y} \neq \mu \pm (\nu + \varrho)\right) \\ &\le 2\exp\left(-\frac{2\varrho^2}{\sum_{y=1}^{t+\varepsilon n-2}\deg(y)^2}\right)\,.\end{aligned}$$

By Lemma 7 applied to G, and because $\Delta(G) \le cn/\log n$, we have

$$\frac{2\varrho^2}{\sum_{y=1}^{t+\varepsilon n-2}\deg(y)^2} = \frac{2\varepsilon^4 n^2}{d^2\sum_{y=1}^{t+\varepsilon n-2}\deg(y)^2} \ge \frac{\varepsilon^4 \log n}{d^2 Dc}\,,$$

and hence, because $c \le D^{-4}\varepsilon^4/100$ and $d \le D$, we obtain (5.11) as desired. □

Finally, we prove the diet lemma.

Proof of Lemma 26. First observe that if ψ_{t-1} satisfies the $(\beta_{t-1}, 2D+3)$-diet condition, *RandomEmbedding* cannot fail at time t, so ψ_t exists. We first state a claim that if the diet condition holds up to time $t-\varepsilon n$, then for any given large set $T \subseteq V(H)$, with high probability either the cover condition fails at some time before $t-\varepsilon n$, or ψ_{t-1} embeds about the expected fraction of each interval of εn vertices to T.

Claim 26.1. *For every $1 \le j \le t-\varepsilon n+1$, and for every $T \subseteq V(H) \setminus \operatorname{im}\psi_j$ with $|T| \ge \frac{1}{2}\gamma^{2D+3}\delta n$, if the $(\beta_j, 2D+3)$-diet condition holds for $(H, \operatorname{im}\psi_j)$, then with probability at least $1-n^{-2D-19}$, one of the following occurs.*

(a) *ψ_t does not have the $(\varepsilon, 20D\beta_j, j)$-cover condition, or*
(b) $\big|\{x : j \le x < j+\varepsilon n, \psi_{t-1}(x) \in T\}\big| = (1 \pm 40D\beta_j)\frac{|T|\varepsilon n}{n-j}$.

We defer the proof of this claim until later, and move on to state a second claim, which we will deduce from Claim 26.1. Let $\ell = \lfloor \frac{t}{\varepsilon n} \rfloor$. We claim that either we witness a failure of the diet or cover conditions before time t, or the set $\mathsf{N}_H(R) \cap \mathsf{N}_{H^*}(S \setminus R) \setminus \operatorname{im}\psi_{\ell\varepsilon n}$ has about the expected size for each $R \subseteq S \subseteq V(H)$ with $|S| \le 2D+3$.

Claim 26.2. *With probability at least $1-n^{-10}$, one of the following holds.*

(a) *The $(\beta_j, 2D+3)$-diet condition fails for $(H, \operatorname{im}\psi_j)$ for some $1 \le j \le t-1$, or*
(b) *the $(\varepsilon, 20D\beta_j, j)$-cover condition fails for ψ_{t-1} for some $1 \le j \le t+1-\varepsilon n$, or*
(c) *for every $R \subseteq S \subseteq V(H)$ with $|S| \le 2D+3$, we have*

$$\begin{aligned} &\big|\mathsf{N}_H(R) \cap \mathsf{N}_{H^*}(S \setminus R) \setminus \operatorname{im}\psi_{\ell\varepsilon n}\big| = \\ &\big|\mathsf{N}_H(R) \cap \mathsf{N}_{H^*}(S \setminus R)\big| \prod_{k=0}^{\ell-1}\Big(1 - \big(1 \pm 40D\beta_{k\varepsilon n}\big)\tfrac{\varepsilon n}{n-k\varepsilon n}\Big)\,. \end{aligned} \tag{5.15}$$

Before proving these claims, we show that Claim 26.2 implies the lemma. We want to show that (5.15) holding implies that we do not have witnesses for a failure of the diet condition nor the codiet condition at time t. Indeed, taking logs, we have

$$\begin{aligned} &\log\big|\mathsf{N}_H(R) \cap \mathsf{N}_{H^*}(S \setminus R) \setminus \operatorname{im}\psi_{\ell\varepsilon n}\big| \\ &= \log\big|\mathsf{N}_H(R) \cap \mathsf{N}_{H^*}(S \setminus R)\big| + \sum_{k=0}^{\ell-1} \log\Big(1 - (1 \pm 40D\beta_{k\varepsilon n})\tfrac{\varepsilon n}{n-k\varepsilon n}\Big) \\ &= \log\big|\mathsf{N}_H(R) \cap \mathsf{N}_{H^*}(S \setminus R)\big| + \sum_{k=0}^{\ell-1}\Big(\log\tfrac{n-(k+1)\varepsilon n}{n-k\varepsilon n} + \log\big(1 \pm \tfrac{40D\beta_{k\varepsilon n}\varepsilon n}{n-(k+1)\varepsilon n}\big)\Big) \end{aligned}$$

$$= \log \left|\mathsf{N}_H(R) \cap \mathsf{N}_{H^*}(S \setminus R)\right| + \log\left(1-\ell\varepsilon\right) \pm 2\sum_{k=0}^{\ell-1} \tfrac{40D\beta_{k\varepsilon n}\varepsilon}{1-(k+1)\varepsilon}\,,$$

where the final equality holds since $1-(k+1)\varepsilon \geq \delta$, and hence by choice of ε the quantity $\frac{40D\beta_{k\varepsilon n}\varepsilon}{1-(k+1)\varepsilon}$ is close to 0. Since at most εn vertices are removed from $\mathsf{N}_H(R)\cap\mathsf{N}_{H^*}(S\setminus R)\setminus \operatorname{im}\psi_{\ell\varepsilon n}$ to obtain $\mathsf{N}_H(R)\cap\mathsf{N}_{H^*}(S\setminus R)\setminus\operatorname{im}\psi_t$, we conclude

$$\begin{aligned}&\left|\mathsf{N}_H(R)\cap\mathsf{N}_{H^*}(S\setminus R)\setminus\operatorname{im}\psi_t\right|\\&= |\mathsf{N}_H(R)\cap\mathsf{N}_{H^*}(S\setminus R)|\cdot\frac{n-t\pm\varepsilon n}{n}\cdot\exp\Big(\pm 80D\delta^{-1}\varepsilon\sum_{k=0}^{\ell-1}\beta_{k\varepsilon n}\Big)\pm\varepsilon n\,.\end{aligned}\tag{5.16}$$

We first consider the case $R = S$, when $\mathsf{N}_H(R)\cap\mathsf{N}_{H^*}(S\setminus R) = \mathsf{N}_H(S)$, and deduce that S does not witness a failure of the $(\beta_t, 2D+3)$-diet condition for $(H, \operatorname{im}\psi_t)$. Indeed, from (5.16) we have

$$\begin{aligned}\left|\mathsf{N}_H(S)\setminus\operatorname{im}\psi_t\right| &= |\mathsf{N}_H(S)|\cdot\frac{n-t\pm\varepsilon n}{n}\cdot\exp\Big(\pm 80D\delta^{-1}\varepsilon\sum_{k=0}^{\ell-1}\beta_{k\varepsilon n}\Big)\pm\varepsilon n\\&= (1\pm\alpha)p^{|S|}(n-t\pm\varepsilon n)\Big(1\pm 200D\delta^{-1}\varepsilon\sum_{k=0}^{\ell-1}\beta_{k\varepsilon n}\Big)\Big(1\pm\frac{2\varepsilon n}{p^{|S|}(n-t)}\Big)\end{aligned}$$

where the second equality uses the fact that H is $(\alpha, 2D+3)$-quasirandom. We thus have

$$\begin{aligned}\left|\mathsf{N}_H(S)\setminus\operatorname{im}\psi_t\right| &= (1\pm\alpha)p^{|S|}(n-t)\Big(1\pm 200D\delta^{-1}\varepsilon\sum_{k=0}^{\ell-1}\beta_{k\varepsilon n}\Big)(1\pm 4\varepsilon\delta^{-1}\gamma^{-|S|})\\&\overset{(5.3)}{=} (1\pm\alpha)p^{|S|}(n-t)(1\pm\beta_t/4)(1\pm 4\varepsilon\delta^{-1}\gamma^{-|S|})\\&= (1\pm\beta_t)p^{|S|}(n-t)\,.\end{aligned}$$

Now, we let R be any subset of S and aim to establish the codiet condition. Again from (5.16), we have

$$\begin{aligned}&\left|\mathsf{N}_H(R)\cap\mathsf{N}_{H^*}(S\setminus R)\setminus\operatorname{im}\psi_t\right|\\&\quad= |\mathsf{N}_H(R)\cap\mathsf{N}_{H^*}(S\setminus R)|\cdot\frac{n-t\pm\varepsilon n}{n}\cdot\exp\Big(\pm 80D\delta^{-1}\varepsilon\sum_{k=0}^{\ell-1}\beta_{k\varepsilon n}\Big)\pm\varepsilon n\\&\quad= (1\pm\eta)p^{|R|}\hat{p}^{|S\setminus R|}(n-t\pm\varepsilon n)\Big(1\pm 200D\delta^{-1}\varepsilon\sum_{k=0}^{\ell-1}\beta_{k\varepsilon n}\Big)\Big(1\pm\frac{2\varepsilon n}{p^{|S|}(n-t)}\Big)\end{aligned}$$

since (H, H^*) is $(\eta, 2D+3)$-coquasirandom. Therefore

$$\begin{aligned}
&\big|\mathsf{N}_H(R)\cap\mathsf{N}_{H^*}(S\setminus R)\setminus \operatorname{im}\psi_t\big| \\
&\qquad = (1\pm\eta)p^{|R|}\hat{p}^{|S\setminus R|}(n-t)\Big(1\pm 200D\delta^{-1}\varepsilon\sum_{k=0}^{\ell-1}\beta_{k\varepsilon n}\Big)(1\pm 4\varepsilon\delta^{-1}\gamma^{-|S|}) \\
&\qquad \overset{(5.3)}{=} (1\pm\eta)(1\pm\beta_k)(1\pm 4\varepsilon\delta^{-1}\gamma^{-|S|})p^{|R|}\hat{p}^{|S\setminus R|}(n-t) \\
&\qquad = (1\pm 2\eta)p^{|R|}\hat{p}^{|S\setminus R|}(n-t)\,.
\end{aligned}$$

This concludes the proof of the lemma, modulo the proofs of Claim 26.1 and Claim 26.2, which we now provide.

Proof of Claim 26.1. Let j and T be as in the statement. Fix $0\le d\le D$. We want to show how to make use of the $(\varepsilon,20D\beta_j,j)$-cover condition for ψ_j (which we have when Part (a) fails) to deduce that the assertion of Part (b) holds with high probability. That is, we consider the number of vertices in $X_{j,d}$ embedded to T. In order to apply Lemma 4, we want to estimate the sum over $x\in X_{j,d}$ of the probability that x is embedded to T, conditioning on ψ_{x-1}, that is, we need to estimate the number

$$\frac{\big|T\cap C^{x-1}(x)\setminus \operatorname{im}\psi_{x-1}\big|}{\big|C^{x-1}(x)\setminus \operatorname{im}\psi_{x-1}\big|}\,. \tag{5.17}$$

By the diet condition, we have $\big|C^{x-1}(x)\setminus\operatorname{im}\psi_j\big|=(1\pm\beta_j)p^d(n-j)$. Since $j<t\le(1-\delta)n$, since $x\le j+\varepsilon n$, since $p\ge\gamma$, and by choice of ε, we have

$$\big|C^{x-1}(x)\setminus \operatorname{im}\psi_{x-1}\big| = (1\pm 2\beta_j)p^d(n-j)\;, \tag{5.18}$$

thus providing a bound on the denumerator in (5.17). (Note that this bound on the denumerator does not depend on the choice of $x\in X_{j,d}$.) Now x is embedded uniformly at random into $C^{x-1}(x)\setminus\operatorname{im}\psi_{x-1}$, so it remains to determine the sum of the numerators in (5.17),

$$\begin{aligned}
\sum_{x\in X_{j,d}}\big|T\cap C^{x-1}(x)\setminus\operatorname{im}\psi_{x-1}\big| &= \sum_{x\in X_{j,d}}\big|T\cap C^{x-1}(x)\setminus\operatorname{im}\psi_{j}\big| \pm\varepsilon|X_{j,d}|n \\
&= \sum_{x\in X_{j,d}}\big|T\cap C^{x-1}(x)\big| \pm\varepsilon^2 n^2\,,
\end{aligned} \tag{5.19}$$

where the first equality uses $j\le x<j+\varepsilon n$, and the second the fact that $T\subseteq V(H)\setminus\operatorname{im}\psi_j$ and that $|X_{j,d}|\le\varepsilon n$. But now if the $(\varepsilon,20D\beta_j,j)$-cover condition holds for ψ_j, then summing over $v\in T$ we obtain

$$\sum_{x\in X_{j,d}}\big|T\cap C^{x-1}(x)\big| = |T|(1\pm 20D\beta_j)p^d|X_{j,d}|\pm\varepsilon^2|T|n\,,$$

which, combined with (5.19), gives

$$\sum_{x\in X_{j,d}} \left|T\cap C^{x-1}(x)\setminus \operatorname{im}\psi_{x-1}\right| = (1\pm 20D\beta_j)p^d|T||X_{j,d}| \pm 2\varepsilon^2n^2 \,. \tag{5.20}$$

We can thus apply Lemma 4, setting $\mathcal{E}$ to be the event that the $(\varepsilon, 20D\beta_j, j)$-cover condition holds for ψ_j. The random variables whose sum we are estimating are the Bernoulli random variables indicating whether each $x\in X_{j,d}$ is embedded to T, so the sum of squares of their ranges is at most εn. Combining (5.18) and (5.20), the expected number of vertices of $X_{j,d}$ embedded to T is

$$\frac{(1\pm 20D\beta_j)p^d|T||X_{j,d}|\pm 2\varepsilon^2n^2}{(1\pm 2\beta_j)p^d(n-j)} = (1\pm 30D\beta_j)\frac{|T||X_{j,d}|}{n-j} \pm 4\varepsilon^2\gamma^{-d}\delta^{-1}n\,,$$

where we use $n-j\geq \delta n$ and $p\geq\gamma$. The probability that the $(\varepsilon, 20D\beta_j, j)$-cover condition holds for ψ_j and the outcome differs from this by more than $\varepsilon^2 n$ is at most $2\exp(-2\varepsilon^3 n)\leq n^{-2D-20}$, so taking the union bound over the $D+1$ choices of d and summing, we conclude that with probability at most n^{-2D-19} the $(\varepsilon, 20D\beta_j, j)$-cover condition holds for ψ_j and the number of vertices x with $j\leq x<j+\varepsilon n$ embedded to T is not equal to

$$(1\pm 30D\beta_j)\frac{|T|\varepsilon n}{n-j} \pm 4(D+1)\varepsilon^2\gamma^{-D}\delta^{-1}n \pm (D+1)\varepsilon^2 n = (1\pm 40D\beta_j)\frac{|T|\varepsilon n}{n-j}\,,$$

where the final equality uses our lower bound on $|T|$ and the choice of ε. This is what we wanted to show. $\square$

Proof of Claim 26.2. Given a set $S\subseteq V(H)$ with $|S|\leq 2D+3$ and a subset $R\subseteq S$, for each integer $0\leq k<\ell$, we set $T_k = \mathsf{N}_H(R)\cap \mathsf{N}_{H^*}(S\setminus R)\setminus \operatorname{im}\psi_{k\varepsilon n}$. Observe that as (H,H^*) is $(\eta, 2D+3)$-coquasirandom, we have

$$|T_0|\geq (1-\eta)p^{|R|}\hat{p}^{|S\setminus R|}n \geq (1-\eta)^{2D+4}\gamma^{2D+3}n\,.$$

For each $0\leq k<\ell$, suppose that

$$\begin{aligned}|T_k| &\geq (1-\eta)^{2D+4}(1-80D\beta_n\delta^{-1}\varepsilon)^k\gamma^{2D+3}(n-k\varepsilon n)\\ &\geq \tfrac{9}{10}(1-80D\beta_n\delta^{-1}\varepsilon)^{1/\varepsilon}\gamma^{2D+3}\delta n \geq \tfrac{9}{10}\exp\big(-200D\beta_n\delta^{-1}\big)\gamma^{2D+3}\delta n\\ &> \frac{1}{2}\gamma^{2D+3}\delta n\,,\end{aligned}$$

where the final line follows since $200D\beta_n\delta^{-1}\leq 400CD\alpha\delta^{-1}<1/100$ by choice of α. We can thus apply Claim 26.1 with $T=T_k$ and obtain that with probability at least $1-n^{-2D-19}$ either we have a failure of the diet or the cover condition is witnessed before time k, or we have

$$|T_{k+1}| = |T_k|\Big(1 - \big(1 \pm 40D\beta_{k\varepsilon n}\big)\frac{\varepsilon n}{n - k\varepsilon n}\Big) .$$

Observe that then

$$\begin{aligned}|T_{k+1}| &\geq |T_k|\left(1 - \frac{\varepsilon n}{n - k\varepsilon n} - 40D\beta_n\delta^{-1}\varepsilon\right)\\ &> (1-\eta)^{2D+4}(1 - 80D\beta_n\delta^{-1}\varepsilon)^{k+1}\gamma^{2D+3}\big(n - (k+1)\varepsilon n\big) ,\end{aligned}$$

providing the assumption for using of Claim 26.1 in step $k+1$.

Repeating this process for each $0 \leq k \leq \ell - 1$ we get that with probability at least $1 - \varepsilon^{-1}n^{-2D-19}$ either a failure of the diet or cover condition is witnessed before time $\ell\varepsilon n$, or we have

$$\big|T_\ell\big| = \Big|\mathsf{N}_H(R) \cap \mathsf{N}_{H^*}(S \setminus R)\Big| \prod_{k=0}^{\ell-1}\Big(1 - (1 \pm 40D\beta_{k\varepsilon n})\tfrac{\varepsilon n}{n - k\varepsilon n}\Big) .$$

Taking a union bound over the at most $(2D+3)n^{2D+3}$ choices of S and the at most 2^{2D+3} choices of $R \subseteq S$, we see that with probability at least $1 - n^{-10}$ either a failure of the diet or cover condition is witnessed before time t, or the above equation holds for all $|S| \leq 2D + 3$ and $R \subseteq S$. □

6. Maintaining quasirandomness

In this section we provide the proofs of Lemma 16 and Lemma 17.

6.1. Initial coquasirandomness

We begin with the easy proof of Lemma 16, which states that splitting the edges of a quasirandom graph randomly gives a coquasirandom pair with high probability.

Proof of Lemma 16. Using (2.1) we see that the densities p_0 and p_0^* of H_0 and H_0^* satisfy

$$p_0 = (1 \pm \tfrac{\alpha_0}{1000D})(p - \gamma) \quad \text{and} \quad p_0^* = (1 \pm \tfrac{\alpha_0}{1000D})\gamma \tag{6.1}$$

with probability at least $1 - n^{-10}$, giving the first part of Lemma 16.

Now, let $R \subseteq S \subseteq V(\widehat{H})$ be two sets of size at most $2D + 3$. By quasirandomness of $\widehat{H}$ we have $|\mathsf{N}_{\widehat{H}}(S)| = (1 \pm \xi)p^{|S|}n$. Observe that each vertex of $\mathsf{N}_{\widehat{H}}(S)$ appears with probability $q^{|R|}(1-q)^{|S\setminus R|}$ in $\mathsf{N}_{H_0^*}(R) \cap \mathsf{N}_{H_0}(S \setminus R)$. Hence,

$$\mathbb{E}\left(\left|\mathsf{N}_{H_0^*}(R) \cap \mathsf{N}_{H_0}(S \setminus R)\right|\right) = q^{|R|}(1-q)^{|S\setminus R|}(1 \pm \xi)p^{|S|}n .$$

Observe also that for distinct vertices in $\mathsf{N}_{\widehat{H}}(S)$ the events whether these appear in $\mathsf{N}_{H_0^*}(R) \cap \mathsf{N}_{H_0}(S \setminus R)$ are independent. Using again (2.1), with probability at least $1 - n^{-2D-10}$ we have that

$$\left|\mathsf{N}_{H_0^*}(R)\cap\mathsf{N}_{H_0}(S\setminus R)\right| = q^{|R|}(1-q)^{|S\setminus R|}(1\pm 2\xi)p^{|S|}n\,. \tag{6.2}$$

Taking the union bound we conclude that (6.2) holds for all $S\subseteq V(\widehat{H})$ with $|S|\le 2D+3$ and $R\subseteq S$ with probability at least $1-n^{-6}$.

Now, assume that (6.1) holds. Then the right-hand side of (6.2) can be rewritten as

$$\begin{aligned}(1\pm 2\xi)\gamma^{|R|}(p-\gamma)^{|S\setminus R|}n &= (1\pm 2\xi)\left(\frac{p_0^*}{1\pm\frac{\xi_0}{1000D}}\right)^{|R|}\left(\frac{p_0}{1\pm\frac{\alpha_0}{1000D}}\right)^{|S\setminus R|}n\\ &= (1\pm 2\xi)(1\pm\tfrac{\alpha_0}{100})(p_0^*)^{|R|}p_0^{|S\setminus R|} = \left(1\pm\tfrac{1}{10}\alpha_0\right)(p_0^*)^{|R|}p_0^{|S\setminus R|}\,.\end{aligned}$$

We conclude that (H_0^*,H_0) is $\left(\frac{1}{10}\alpha_0, 2D+3\right)$-coquasirandom with probability at least $1-n^{-5}$. □

6.2. *Maintaining coquasirandomness*

In this subsection we prove Lemma 17. We need to show that, provided coquasirandomness is maintained up to stage $s-1$ and *RandomEmbedding* does not fail, it is likely that coquasirandomness holds after stage s, when G_s is embedded into H_{s-1} and we obtain H_s. Let us briefly sketch the idea (for convenience focusing only on quasirandomness of H_s). We fix a set $R\subseteq V(\widehat{H})$ with $|R|\le 2D+3$, and consider the running of *PackingProcess* up to stage s. We want to show that it is very unlikely that R witnesses the failure of H_s to be quasirandom, since then the union bound over choices of R tells us that it is likely that H_s is quasirandom. In other words, we want to know that $\left|N_{H_s}(R)\right|$ is very likely close to the expected size. We write

$$\left|N_{H_s}(R)\right| = \left|N_{H_0}(R)\right| - Y_1 - \cdots - Y_s\,,$$

where $Y_i = \left|N_{H_i-1}(R)\right| - \left|N_{H_i}(R)\right|$ is the change at step i, and apply Lemma 5 to show that the sum $Y_1+\cdots+Y_s$ is very likely to be close to its expectation. So proving Lemma 17 boils down to estimating accurately $\mathbb{E}(Y_i|H_{i-1})$ and finding a reasonable upper bound for $\mathbb{E}(Y_i^2|H_{i-1})$. The latter turns out to be relatively straightforward and is done in Lemma 31. We now outline the route to the former estimation.

Observe that Y_i is equal to the number of stars in H_{i-1} whose leaves are the vertices in R and at least one of whose edges is used in embedding G_i to H_{i-1}. By linearity of expectation, $\mathbb{E}(Y_i|H_{i-1})$ is equal to the sum, over stars in H_{i-1} whose leaves are R, of the probability that at least one edge in the star is used in embedding G_i. We will see that this probability is about the same for any given star S, and the problem is to calculate it. To do this we need to consider the running of *RandomEmbedding*.

We begin in Lemma 27 by estimating the chance that a given vertex, or one of a given pair of vertices, is used in a short time interval in *RandomEmbedding*. From this we deduce in Lemma 28 the probability that a given vertex, or one of a pair, is used in any given time interval. This helps us to establish, in Lemma 29, that any given edge of H_{i-1} is about equally likely to be used in the embedding of G_i. Finally, in Lemma 30 we show that the chance of two or more edges in S being used in the embedding of G_i is tiny, from which it follows that the chance of one or more is about $|R|$ times the probability of any given edge being used.

All of these estimations depend upon H_{i-1} being sufficiently quasirandom, and the errors depend upon the quasirandomness α_{i-1}. Because the errors add up over time, it is important that the α_s increase quite fast with s. Here it is very important that the dependence of the error term in Lemma 24 is linear in the input α and not much worse: otherwise it would not be possible to choose any sequence α_s such that the error remains bounded by α_s at each stage s.

As the main work is to estimate the probability that, for a given H_{s-1} and G_s, and R and v, *RandomEmbedding* uses an edge of the star with centre v and leaves R when embedding G_s into H_{s-1}, for most of this section we will consider fixed graphs G and H. We now embark upon this probability estimation.

First, for given $u, v \in V(H)$, we estimate the probability that *RandomEmbedding* embeds a vertex to $\{u, v\}$ in the short interval of time $[t, t + \varepsilon n)$, conditioning on not having done so before time t, and the probability that *RandomEmbedding* embeds a vertex to v in the interval of time $[t, t + \varepsilon n)$, conditioning on not having done so before time t. In both cases, we need to assume that the history $\mathscr{H}_{t-1}$ of embedding up to time $t - 1$ is typical (in a sense which we now make precise).

Lemma 27. *Given $D \in \mathbb{N}$ and $\gamma > 0$, let $\delta, \alpha_0, \alpha_{2n}, C, \varepsilon$ be as in Setting 15. The following holds for any $\alpha_0 \le \alpha \le \alpha_{2n}$ and all sufficiently large n. Suppose that G is a graph on $[n]$ such that $\deg^-(x) \le D$ for each $x \in V(G)$, and H is an $(\alpha, 2D + 3)$-quasirandom graph with n vertices and $p\binom{n}{2}$ edges, with $p \ge \gamma$. Suppose that u and v are two distinct vertices of H. When* RandomEmbedding *is run to embed $G[[n-\delta n]]$ into H, for any $1 \le t \le n + 1 - (\delta + \varepsilon)n$ we have the following two statements.*

(a) *Suppose the history $\mathscr{H}_{t-1}$ up to and including embedding $t - 1$ is such that $v \notin \operatorname{im}\psi_{t-1}$, the $(C\alpha, 2D + 3)$-diet condition holds for $(H, \operatorname{im}\psi_{t-1})$, and*

$$\mathbb{P}^{G\hookrightarrow H}\left(\mathsf{CoverE}(C\alpha; t)|\mathscr{H}_{t-1}\right) \le n^{-3}\,.$$

Then we have

$$\mathbb{P}^{G\hookrightarrow H}\left(v \in \operatorname{im}\psi_{t+\varepsilon n-1}\middle|\mathscr{H}_{t-1}\right) = (1 \pm 10C\alpha)\tfrac{\varepsilon n}{n-t}\,.$$

(b) *Suppose the history $\mathscr{H}_{t-1}$ up to and including embedding $t - 1$ is such that $u, v \notin \operatorname{im}\psi_{t-1}$, the $(C\alpha, 2D + 3)$-diet condition holds for $(H, \operatorname{im}\psi_{t-1})$, and*

$$\mathbb{P}^{G\hookrightarrow H}\left(\mathsf{CoverE}(C\alpha; t)|\mathscr{H}_{t-1}\right) \le n^{-3}\,.$$

Then we have

$$\mathbb{P}^{G\hookrightarrow H}\Big(\big|\{u,v\}\cap \operatorname{im}\psi_{t+\varepsilon n-1}\big|\geq 1\Big|\mathscr{H}_{t-1}\Big)=(1\pm 10C\alpha)\tfrac{2\varepsilon n}{n-t}\,.$$

Before proving Lemma 27, we first sketch its proof. For Lemma 27(a), the idea is that either the cover condition fails, or v is in candidate sets of roughly $p^d|X_{t,d}|$ vertices x of $X_{t,d}$ (for each d). Because the diet condition holds at time $t-1$, each of these vertices x is embedded uniformly at random to a set of roughly $p^d(n-t)$ vertices. One would like to say that it follows that the probability that x is embedded to v is thus about $1/(p^d(n-t))$ and the desired result follows by summing these probabilities. Unfortunately this is not true: the probability that x is embedded to v also depends on the probability that no previous vertex was embedded to v. In order to get around this, we define the following *ModifiedRandomEmbedding*, which generates a sequence of embeddings with an identical distribution to *RandomEmbedding*, but which in addition generates a sequence of *reported* vertices. The modification we make is simple: at each time $1\leq t'\leq n-\delta n$, *RandomEmbedding* chooses a vertex of $C^{t'-1}_{G\hookrightarrow H}(t')\setminus \operatorname{im}\psi_{t'-1}$. In *ModifiedRandomEmbedding*, we instead choose a vertex w of $C^{t'-1}_{G\hookrightarrow H}(t')\setminus(\operatorname{im}\psi_{t'-1}\setminus\{v\})$, and report this vertex. If the reported vertex w is not in $\operatorname{im}\psi_{t'-1}$, we set $\psi_{t'}=\psi_{t'-1}\cup\{t'\hookrightarrow w\}$, as in *RandomEmbedding*. If the reported vertex is in $\operatorname{im}\psi_{t'-1}$ (which happens only if $w=v$) we choose w' uniformly at random in $C^{t'-1}_{G\hookrightarrow H}(t')\setminus \operatorname{im}\psi_{t'-1}$, and set $\psi_{t'}=\psi_{t'-1}\cup\{t'\hookrightarrow w'\}$. We will see that it is easy to calculate the expected number of times v is reported, and also easy to show that the contribution due to v being reported multiple times is tiny. The point is that the probability of *RandomEmbedding* using v is the same as the probability that *ModifiedRandomEmbedding* reports v at least once, which we can thus calculate.

Lemma 27(b) is established similarly, using a slightly different version of *ModifiedRandomEmbedding*.

Proof of Lemma 27(a). Instead of *RandomEmbedding*, we consider *ModifiedRandomEmbedding* as defined above, which creates the same embedding distribution. For each i, let $r(i)$ be the vertex reported by *ModifiedRandomEmbedding* at time i. We shall use the following two auxiliary claims.

Define E as the random variable counting the times when v is reported by *ModifiedRandomEmbedding* in the interval $t\leq x<t+\varepsilon n$,

$$E=\Big|\,\big\{x\in[t,t+\varepsilon n)\,:\,r(x-1)=v\big\}\,\Big|\,.$$

The probability that *RandomEmbedding* uses v in the interval $t\leq x<t+\varepsilon n$, conditioning on $\mathscr{H}_{t-1}$, is equal to the probability that *ModifiedRandomEmbedding* reports v at least once in that interval, which probability is by definition at least

$$\mathbb{E}\left(E\mid\mathscr{H}_{t-1}\right)-\sum_{k=2}^{\varepsilon n}\mathbb{P}\big(v\text{ is reported at least }k\text{ times in the interval }[t,t+\varepsilon n)\big|\mathscr{H}_{t-1}\big)\,.$$

Our first claim estimates $\mathbb{E}(E \mid \mathscr{H}_{t-1})$.

Claim 27.1. *We have that*

$$\mathbb{E}(E \mid \mathscr{H}_{t-1}) = (1 \pm 4C\alpha)\frac{\varepsilon n}{n-t} \pm 4(D+1)\varepsilon^2\gamma^{-D}\delta^{-1} .$$

Our second claim is that the sum in the expression above is small.

Claim 27.2. *We have that*

$$\sum_{k=2}^{\varepsilon n} \mathbb{P}\Big(\big|\{x \in [t, t+\varepsilon n) \,:\, r(x-1) = v\}\big| \geq k\Big) \leq 8\varepsilon^2\gamma^{-2D}\delta^{-2} .$$

By choice of ε, we have $16(D+1)\varepsilon^2\gamma^{-2D}\delta^{-2} < C\alpha\varepsilon$. Thus the two claims give Lemma 27(a). We now prove the auxiliary Claims 27.1 and 27.2.

Proof of Claim 27.1. Note that since the $(C\alpha, 2D+3)$-diet condition holds for $(H, \operatorname{im}\psi_{t-1})$, for each $t \leq x < t+\varepsilon n$, setting $S = \psi_{x-1}(\mathsf{N}^-(x))$, we have[9]

$$\begin{aligned} \big|C^{x-1}(x) \setminus \operatorname{im}\psi_{x-1}\big| \pm 2 &= \big|\mathsf{N}_H(S) \setminus \operatorname{im}\psi_{t-1}\big| \pm \varepsilon n \pm 2 \\ &= (1 \pm C\alpha)p^{|\mathsf{N}^-(x)|}(n-t) \pm \varepsilon n \pm 2 \\ &= (1 \pm 2C\alpha)p^{|\mathsf{N}^-(x)|}(n-t) . \end{aligned} \tag{6.3}$$

By linearity of expectation, we have

$$\begin{aligned} \mathbb{E}\Big[E \mid \mathscr{H}_{t-1}\Big] &= \sum_{x=t}^{t+\varepsilon n-1} \mathbb{P}\big(v \text{ is reported at time } x \big| \mathscr{H}_{t-1}\big) \\ &= \sum_{x=t}^{t+\varepsilon n-1} \mathbb{E}\left(\frac{\mathbb{1}\{v \in C^{x-1}(x)\}}{|C^{x-1}(x) \setminus (\operatorname{im}\psi_{x-1} \setminus \{v\})|} \,\Big|\mathscr{H}_{t-1}\right) \\ &= \sum_{x=t}^{t+\varepsilon n-1} \mathbb{E}\left(\frac{\mathbb{1}\{v \in C^{x-1}(x)\}}{|C^{x-1}(x) \setminus \operatorname{im}\psi_{x-1}| \pm 1} \,\Big|\mathscr{H}_{t-1}\right) . \end{aligned} \tag{6.4}$$

Using (6.3), we get

$$\mathbb{E}(E \mid \mathscr{H}_{t-1}) = \sum_{x=t}^{t+\varepsilon n-1} \frac{\mathbb{P}\big(v \in C^{x-1}(x) \big| \mathscr{H}_{t-1}\big)}{(1 \pm 2C\alpha)p^{|\mathsf{N}^-(x)|}(n-t)} .$$

Splitting this sum up according to $|\mathsf{N}^-(x)|$, and again using linearity of expectation, we have

[9] We remark that in (6.3), the calculations are included with an error "± 2" and for this proof "± 1" would have sufficed. We reuse (6.3) in the proof of Lemma 27(b) where the bigger error is needed.

$$\mathbb{E}\left(E \mid \mathscr{H}_{t-1}\right) = \sum_{d=0}^{D} \frac{\mathbb{E}\left(|\{x \in X_{t,d} : v \in C^{x-1}(x)\}|\big|\mathscr{H}_{t-1}\right)}{(1 \pm 2C\alpha)p^d(n-t)}\,.$$

Now for each $0 \le d \le D$, since the $(\varepsilon, C\alpha, t)$-cover condition holds with probability at least $1 - n^{-3}$ conditioning on $\mathscr{H}_{t-1}$, we have

$$\begin{aligned}\mathbb{E}\left(|\{x \in X_{t,d} : v \in C^{x-1}(x)\}|\big|\mathscr{H}_{t-1}\right) &= (1-n^{-3})\big((1 \pm C\alpha)p^d|X_{t,d}| \pm \varepsilon^2 n\big) \pm n^{-3}\cdot \varepsilon n \\ &= (1 \pm C\alpha)p^d|X_{t,d}| \pm 2\varepsilon^2 n\,.\end{aligned}$$

Substituting this in, we have

$$\mathbb{E}\left(E \mid \mathscr{H}_{t-1}\right) = \sum_{d=0}^{D} \frac{(1 \pm C\alpha)p^d|X_{t,d}| \pm 2\varepsilon^2 n}{(1 \pm 2C\alpha)p^d(n-t)} = (1 \pm 4C\alpha)\tfrac{\varepsilon n}{n-t} \pm 4(D+1)\varepsilon^2\gamma^{-D}\delta^{-1}\,,$$

where the last equality uses $p \ge \gamma$ and $n - t \ge \delta n$. □

Proof of Claim 27.2. Since the $(C\alpha, 2D+3)$-diet condition holds for $(H, \operatorname{im}\psi_{t-1})$, since $p \ge \gamma$, and since $n-t \ge \delta n$, for each $x \in [t, t+\varepsilon n)$, when we embed x we report a uniform random vertex from a set of size at least $\frac{1}{2}\gamma^D\delta n$. The probability of reporting v when we embed x is thus at most $2\gamma^{-D}\delta^{-1}n^{-1}$, conditioning on $\mathscr{H}_{t-1}$ and any embedding of the vertices $[t, x)$. Since the conditional probabilities multiply, the probability that at each of a given k-set of vertices in $[t, t+\varepsilon n)$ we report v is at most $2^k\gamma^{-kD}\delta^{-k}n^{-k}$. Taking the union bound over choices of k-sets, we have

$$\begin{aligned}&\sum_{k=2}^{\varepsilon n}\mathbb{P}\big(v \text{ is reported at least } k \text{ times in the interval } [t, t+\varepsilon n)\big|\mathscr{H}_{t-1}\big) \\ &\le \sum_{k=2}^{\varepsilon n}\binom{\varepsilon n}{k}2^k\gamma^{-kD}\delta^{-k}n^{-k} \le \sum_{k=2}^{\varepsilon n}\big(2\varepsilon\gamma^{-D}\delta^{-1}\big)^k \le \tfrac{4\varepsilon^2\gamma^{-2D}\delta^{-2}}{1-2\varepsilon\gamma^{-D}\delta^{-1}} \le 8\varepsilon^2\gamma^{-2D}\delta^{-2}\,,\end{aligned}$$

where we use the bound $\binom{\varepsilon n}{k} \le (\varepsilon n)^k$ and sum the resulting geometric series. □

The proof of Lemma 27(b) is similar, and we only focus on the differences.

Proof of Lemma 27(b). We define *MoreModifiedRandomEmbedding* this time reporting a uniform random vertex of $C^{t-1}_{G\hookrightarrow H}(t) \setminus (\operatorname{im}\psi_{t-1} \setminus \{u, v\})$ at each time step t, and either embedding t to it (if it is not in $\operatorname{im}\psi_{t-1}$) or otherwise picking as before a uniform random vertex of $C^{t-1}_{G\hookrightarrow H}(t) \setminus \operatorname{im}\psi_{t-1}$ to embed t to. As before, the embedding distribution generated by this procedure is the same as for *RandomEmbedding*. We let E' be the number of times u or v are reported in the interval $t \le x < t+\varepsilon n$. Again, the probability that *RandomEmbedding* uses either u or v is equal to the probability that *MoreModifiedRandomEmbedding* reports u or v at least once, which by definition is

$$\mathbb{E}\left(E' \mid \mathscr{H}_{t-1}\right) - \sum_{k=2}^{\varepsilon n} \mathbb{P}\big(u \text{ or } v \text{ is reported at least } k \text{ times in the interval } [t, t+\varepsilon n)\big|\mathscr{H}_{t-1}\big).$$

By linearity of expectation, $\mathbb{E}\left(E'|\mathscr{H}_{t-1}\right)$ is equal to the expected number of times u is reported plus the expected number of times v is reported. We now argue that these latter quantities are $(1 \pm 4C\alpha)\frac{\varepsilon n}{n-t} \pm 4(D+1)\varepsilon^2\gamma^{-D}\delta^{-1}$. This follows from calculations in Claim 27.1, with a small change which we now describe. Note that Claim 27.1 deals with *ModifiedRandomEmbedding*, where reported vertices are taken from $C^{x-1}(x) \setminus (\operatorname{im}\psi_{x-1} \setminus \{v\})$ and not from $C^{x-1}(x) \setminus (\operatorname{im}\psi_{x-1} \setminus \{u, v\})$. This is corrected if we rewrite (6.4) as

$$\mathbb{E}\left(E' \mid \mathscr{H}_{t-1}\right) = \sum_{x=t}^{t+\varepsilon n-1} \mathbb{E}\left(\frac{\mathbb{1}\{u \in C^{x-1}(x)\} + \mathbb{1}\{v \in C^{x-1}(x)\}}{|C^{x-1}(x) \setminus \operatorname{im}\psi_{x-1}| \pm 2}\,\Big|\mathscr{H}_{t-1}\right).$$

Then the rest of the calculations in Claim 27.1 applies (see Footnote 9) We thus have

$$\mathbb{E}\left(E' \mid \mathscr{H}_{t-1}\right) = (1 \pm 4C\alpha)\tfrac{2\varepsilon n}{n-t} \pm 8(D+1)\varepsilon^2\gamma^{-D}\delta^{-1}.$$

Again, it remains to show that the effect of reporting u or v multiple times is small. This time the probability at any step x that one of u and v is reported, conditioning on the history up to time $x-1$, is at most $4\gamma^{-2D}\delta^{-2}n^{-1}$, and by the same calculation as above we conclude that the summation is bounded above by $16\varepsilon^2\gamma^{-2D}\delta^{-2}$, which as before gives Lemma 27(b). □

We now use Lemma 27 to estimate the probability of embedding a vertex to v, or to $\{u, v\}$, in the interval $(t_0, t_1]$ (which may be of any length). This time, we do not condition on one typical embedding history up to time t_0, but rather on a history ensemble up to time t_0 which is not very unlikely. This allows us to drop the typicality restriction, simply because only very few histories can be atypical.

Lemma 28. *Given $D \in \mathbb{N}$ and $\gamma > 0$, let $\delta, \alpha_0, \alpha_{2n}, C, \varepsilon$ be as in Setting 15. Then the following holds for any $\alpha_0 \le \alpha \le \alpha_{2n}$ and all sufficiently large n. Suppose that G is a graph on $[n]$ such that $\deg^-(x) \le D$ for each $x \in V(G)$, and H is an $(\alpha, 2D+3)$-quasirandom graph with n vertices and $p\binom{n}{2}$ edges, with $p \ge \gamma$. Let $0 \le t_0 < t_1 \le n - \delta n$. Let $\mathscr{L}$ be a history ensemble of* RandomEmbedding *up to time t_0, and suppose that $\mathbb{P}(\mathscr{L}) \ge n^{-4}$. Then the following hold for any distinct vertices $u, v \in V(H)$.*

(a) *If $v \notin \operatorname{im}\psi_{t_0}$ then we have*

$$\mathbb{P}^{G \hookrightarrow H}(v \notin \operatorname{im}\psi_{t_1}|\mathscr{L}) = (1 \pm 100C\alpha\delta^{-1})\tfrac{n-1-t_1}{n-t_0}.$$

(b) *If $u, v \notin \operatorname{im}\psi_{t_0}$ then we have*

$$\mathbb{P}^{G \hookrightarrow H}(u, v \notin \operatorname{im}\psi_{t_1}|\mathscr{L}) = (1 \pm 100C\alpha\delta^{-1})\big(\tfrac{n-1-t_1}{n-t_0}\big)^2.$$

Proof. We write $\mathbb{P}$ for $\mathbb{P}^{G\hookrightarrow H}$. We shall first address part (a). We divide the interval $(t_0, t_1]$ into $k := \lceil (t_1 - t_0)/\varepsilon n \rceil$ intervals, all but the last of length εn. Let $\mathscr{L}_0 := \mathscr{L}$. Let, for each $1 \le i < k$, the set $\mathscr{L}_i$ be the embedding histories up to time $t_0 + i\varepsilon n$ of *RandomEmbedding* which extend histories in $\mathscr{L}_{i-1}$ and are such that $v \notin \psi_{t_0+i\varepsilon n}$. Let $\mathscr{L}_k$ be the embedding histories up to time t_1 extending those in $\mathscr{L}_{k-1}$ such that $v \notin \psi_{t_1}$. Thus we have

$$\mathbb{P}(v \notin \operatorname{im} \psi_{t_1} | \mathscr{L}) = \mathbb{P}(\mathscr{L}_k)/\mathbb{P}(\mathscr{L}_0)\,.$$

Finally, for each $1 \le i \le k$, let the set $\mathscr{L}'_{i-1}$ consist of all histories in $\mathscr{L}_{i-1}$ such that the $(C\alpha, 2D+3)$-diet condition holds for $(H, \operatorname{im} \psi_{t_0+(i-1)\varepsilon n})$ and the probability that the $(\varepsilon, C\alpha, t_0 + 1 + (i-1)\varepsilon n)$-cover condition fails, conditioned on $\psi_{t_0+(i-1)\varepsilon n}$, is at most n^{-3}. In other words, $\mathscr{L}'_i$ is the subset of $\mathscr{L}_i$ consisting of typical histories, satisfying the conditions of Lemma 27.

We now determine $\mathbb{P}(\mathscr{L}_k)$ in terms of $\mathbb{P}(\mathscr{L}_0)$, and in particular we show inductively that $\mathbb{P}(\mathscr{L}_i) > n^{-5}$ for each i. Observe that for any time t, the probability (not conditioned on any embedding) that either the $(C\alpha, 2D+3)$-diet condition fails for $(H, \operatorname{im} \psi_i)$ for some $i \le t$ or that the $(\varepsilon, C\alpha, t+1)$-cover condition has probability greater than n^{-3} of failing, is at most $2n^{-6}$ by Lemma 24. In other words, for each i we have $\mathbb{P}(\mathscr{L}_i \setminus \mathscr{L}'_i) \le 2n^{-6}$. Thus by Lemma 27(a) we have

$$\begin{aligned}\mathbb{P}(\mathscr{L}_i) &= \big(1 - (1 \pm 10C\alpha)\tfrac{\varepsilon n}{n - t_0 - (i-1)\varepsilon n}\big)\mathbb{P}(\mathscr{L}'_{i-1}) \pm 2n^{-6}\\ &= \big(1 - (1 \pm 10C\alpha)\tfrac{\varepsilon n}{n - t_0 - (i-1)\varepsilon n}\big)\big(\mathbb{P}(\mathscr{L}_{i-1}) \pm 2n^{-6}\big) \pm 2n^{-6}\\ &= \big(1 - (1 \pm 20C\alpha)\tfrac{\varepsilon n}{n - t_0 - (i-1)\varepsilon n}\big)\mathbb{P}(\mathscr{L}_{i-1})\,,\end{aligned}$$

where the final equality uses the lower bound $\mathbb{P}(\mathscr{L}_{i-1}) \ge n^{-5}$. Similarly, we have $\mathbb{P}(\mathscr{L}_k) = \big(1 \pm (1 + 20C\alpha)\frac{\varepsilon n}{n - t_1}\big)\mathbb{P}(\mathscr{L}_{k-1})$.

Putting these observations together, we can compute $\mathbb{P}(\mathscr{L}_k)$:

$$\mathbb{P}(\mathscr{L}_k) = \big(1 \pm (1 + 20C\alpha)\tfrac{\varepsilon n}{n - t_1}\big)\mathbb{P}(\mathscr{L}_0) \prod_{i=1}^{k-1} \Big(1 - (1 \pm 20C\alpha)\tfrac{\varepsilon n}{n - t_0 - (i-1)\varepsilon n}\Big)\,.$$

Observe that the approximation $\log(1 + x) = x \pm x^2$ is valid for all sufficiently small x. In particular, since $n - t_0 - (i-1)\varepsilon n \ge n - t_1 \ge \delta n$ and by choice of ε, for each i we have

$$\log\Big(1 - (1 \pm 20C\alpha)\tfrac{\varepsilon n}{n - t_0 - (i-1)\varepsilon n}\Big) = -(1 \pm 30C\alpha)\tfrac{\varepsilon n}{n - t_0 - (i-1)\varepsilon n}\,.$$

Thus we obtain

$$\log \mathbb{P}(\mathscr{L}_k) = \log \mathbb{P}(\mathscr{L}_0) \pm (1 + 30C\alpha)\tfrac{\varepsilon n}{n - t_1} - \sum_{i=1}^{k-1}(1 \pm 30C\alpha)\tfrac{\varepsilon n}{n - t_0 - (i-1)\varepsilon n}$$

$$= \log \mathbb{P}(\mathscr{L}_0) \pm 2\delta^{-1}\varepsilon - (1 \pm 40C\alpha) \int_{x=0}^{(k-1)\varepsilon n} \tfrac{1}{n-t_0-x}\,\mathrm{d}x$$
$$= \log \mathbb{P}(\mathscr{L}_0) \pm 2\delta^{-1}\varepsilon - (1 \pm 50C\alpha)\big(\log(n-t_0) - \log(n-1-t_1)\big)$$
$$= \log \mathbb{P}(\mathscr{L}_0) + \log \tfrac{n-1-t_1}{n-t_0} \pm 2\delta^{-1}\varepsilon \pm 50C\alpha \log \delta^{-1}\,, \tag{6.5}$$

where we use $t_1 \le n - \delta n$, and we justify that the integral and sum are close by observing that for each i in the summation, if $(i-1)\varepsilon n \le x \le i\varepsilon n$ then we have

$$\tfrac{1}{n-t_0-i\varepsilon n} \le \tfrac{1}{n-t_0-x} \le \tfrac{1}{n-t_0-(i-1)\varepsilon n} \le (1+\alpha)\tfrac{1}{n-t_0-i\varepsilon n}\,,$$

where the final inequality uses $n - t_0 - i\varepsilon n \le n - t_1 \le \delta n$ and the choice of ε. By choice of ε, this gives part (a). Furthermore, (6.5), and the fact $t_1 \le n - \delta n$, imply that $\mathbb{P}(\mathscr{L}_k) \ge n^{-5}$. Since the $\mathscr{L}_i$ form a decreasing sequence of events the same bound holds for each $\mathscr{L}_i$.

For part (b), we use the identical approach, replacing Lemma 27(a) with Lemma 27(b). Since the difference between these equations is a factor of 2, we obtain twice all the terms other than the term $\log \mathbb{P}(\mathscr{L}_0)$ in the above equation, and hence the second statement of the claim. □

Next, we estimate the probability that the edge $uv \in E(H)$ is used by *RandomEmbedding* when embedding G to H. The idea is the following. In order for uv to be used, there must be some $xy \in G$ such that x is embedded to u and y to v, or vice versa. These events are disjoint, and so it suffices to estimate the probability of each separately and sum them. Without loss of generality, we can assume x is embedded before y. We need to calculate the probability that x is embedded to u and y to v. In other words, we need that all left-neighbours of x are embedded to neighbours of u, all left-neighbours of y are embedded to vertices of v, other vertices are not embedded to $\{u, v\}$, and when we come to embed x and y we actually do embed them to u and v. The point of phrasing it like this is that, provided the diet condition holds, we can estimate accurately all the (conditional) probabilities of embedding individual vertices in $N(x) \cup N(y) \cup \{x, y\}$ to neighbourhoods or to u or v, while Lemma 28 gives accurate estimates for the probability of any other vertex being embedded to u or v. Putting this together yields the desired accurate estimate for the probability that we have $x \hookrightarrow u$ and $y \hookrightarrow v$.

Lemma 29. *Given $D \in \mathbb{N}$, and $\gamma > 0$, let constants $\delta, \varepsilon, C, \alpha_0, \alpha_{2n}$ be as in Setting 15. Then the following holds for any $\alpha_0 \le \alpha \le \alpha_{2n}$ and all sufficiently large n. Suppose that G is a graph on $[n]$ such that $\deg^-(x) \le D$ for each $x \in V(G)$, and H is an $(\alpha, 2D+3)$-quasirandom graph with n vertices and $p\binom{n}{2}$ edges, with $p \ge \gamma$. Let uv be an edge of H. When* RandomEmbedding *is run to embed $G[[n-\delta n]]$ into H, the probability that an edge of G is embedded to uv is*

$$\big(1 \pm 500C\alpha\delta^{-1}\big)^{4D+2} p^{-1} n^{-2} \cdot 2e(G)\,.$$

Proof. We first calculate the probability that a given pair (x, y), such that xy is an edge of G, is embedded to (u, v), in that order. Without loss of generality, suppose that $x < y$. Let $z_1, \ldots, z_k$ be the vertices $\mathsf{N}^-(x) \cup \mathsf{N}^-(y) \setminus \{x, y\}$ in increasing order. Let $j \in \{0, \ldots, k\}$ be such that $z_j < x < z_{j+1}$ (where the case $j = 0$ and $j = k$ corresponds to the situations when all z_i's are to the right or to the left of x, respectively; in these cases some notation below has to be modified in a straightforward way). Define time intervals using $z_1, \ldots, z_j, x, z_{j+1}, \ldots, z_k, y$ as separators: $I_0 = [1, z_1 - 1]$, $I_1 = [z_1 + 1, z_2 - 1]$, ..., $I_j = [z_j + 1, x - 1]$, $I_{j+1} = [x + 1, z_{j+1} - 1]$, ..., $I_{k+1} = [z_k + 1, y - 1]$.

We now define a nested collection of events, the first being the trivial (always satisfied) event and the last being the event $\{x \hookrightarrow u, y \hookrightarrow v\}$, whose probability we wish to estimate. These events are simply that we have not yet (by given increasing times in *RandomEmbedding*) made it impossible to have $\{x \hookrightarrow u, y \hookrightarrow v\}$. We will see that we can estimate accurately the probability of each successive event, conditioned on its predecessor.

Let $\mathscr{L}'_{-1}$ be the trivial (always satisfied) event. If $\mathscr{L}'_{i-1}$ is defined, we let $\mathscr{L}_i$ be the event that $\mathscr{L}'_{i-1}$ holds intersected with the event that

(A1) (if $i \leq j$:) no vertex of G in the interval I_i is mapped to u or v, or
(A2) (if $i > j$:) no vertex of G in the interval I_i is mapped to v.

In other words, $\mathscr{L}_i$ is the event that we have not covered u or v in the interval I_i. It turns out that we do not need to know anything else about the embeddings in the interval I_i.

If $\mathscr{L}_i$ is defined, we let $\mathscr{L}'_i$ be that event that $\mathscr{L}_i$ holds and that

(B1) (if $i < j$:)
  (i) (subcase $z_{i+1} \in \mathsf{N}^-(x) \setminus \mathsf{N}^-(y)$:) we have the event $z_{i+1} \hookrightarrow \mathsf{N}_H(u) \setminus \{v\}$,
  (ii) (subcase $z_{i+1} \in \mathsf{N}^-(y) \setminus \mathsf{N}^-(x)$:) we have the event $z_{i+1} \hookrightarrow \mathsf{N}_H(v) \setminus \{u\}$,
  (iii) (subcase $z_{i+1} \in \mathsf{N}^-(x) \cap \mathsf{N}^-(y)$:) we have the event $z_{i+1} \hookrightarrow \mathsf{N}_H(u) \cap \mathsf{N}_H(v)$,
(B2) (if $i = j$:) we have the event $x \hookrightarrow u$,
(B3) (if $j < i \leq k$:) we have the event $z_i \hookrightarrow \mathsf{N}_H(v) \setminus \{u\}$ (unlike the range $i < j$, there are no subcases here, as necessarily $z_i \in \mathsf{N}^-(y) \setminus \mathsf{N}^-(x)$),
(B4) (if $i = k + 1$:) we have the event $y \hookrightarrow v$.

Again, in order for $\{x \hookrightarrow u, y \hookrightarrow v\}$ to occur we obviously need that a neighbour of x is embedded to a neighbour of u and so on, hence the above conditions.

By definition, we have $\mathscr{L}'_{k+1} = \{x \hookrightarrow u, y \hookrightarrow v\}$. Since we have $\mathscr{L}'_i \subseteq \mathscr{L}_i \subseteq \mathscr{L}'_{i-1}$ for each i and $\mathscr{L}'_{-1}$ is the sure event, we see

$$\mathbb{P}\left(x \hookrightarrow u, y \hookrightarrow v\right) = \prod_{i=0}^{k+1} \frac{\mathbb{P}(\mathscr{L}_i)}{\mathbb{P}(\mathscr{L}'_{i-1})} \cdot \frac{\mathbb{P}(\mathscr{L}'_i)}{\mathbb{P}(\mathscr{L}_i)} = \prod_{i=0}^{k+1} \mathbb{P}\left(\mathscr{L}_i \mid \mathscr{L}'_{i-1}\right) \mathbb{P}\left(\mathscr{L}'_i \mid \mathscr{L}_i\right) . \quad (6.6)$$

Thus, we need to estimate the factors in (6.6). This is done in the two claims below. In each claim we assume $\mathbb{P}(\mathscr{L}'_i), \mathbb{P}(\mathscr{L}_i) > n^{-4}$. This assumption is justified, using an implicit induction, since the smallest of all the events we consider is $\mathscr{L}'_{k+1}$, whose probability according to the following (6.10) is bigger than n^{-4}.

Claim 29.1. *We have*

$$\prod_{i=0}^{k+1} \mathbb{P}\left(\mathscr{L}_i \mid \mathscr{L}'_{i-1}\right) = (1 \pm 200C\alpha\delta^{-1})^{2k+2} \cdot \frac{(n-x)(n-y)}{n^2} .$$

Proof. By definition of (A1), for each $i = 0, \dots, j$, we have

$$\mathbb{P}\left(\mathscr{L}_i \mid \mathscr{L}'_{i-1}\right) = (1 \pm 200C\alpha\delta^{-1}) \cdot \frac{(n-1-\max(I_i))^2}{(n-\min(I_i)+1)^2} \tag{6.7}$$

by Lemma 28(b), with $\mathscr{L} = \mathscr{L}'_{i-1}$. Note that looking at two consecutive indices i and $i+1$ in (6.7) we have cancellation of the former nominator and the latter denominator, $n - 1 - \max(I_i) = n - \min(I_{i+1}) + 1$. Thus,

$$\prod_{i=0}^{j} \mathbb{P}\left(\mathscr{L}_i \mid \mathscr{L}'_{i-1}\right) = (1 \pm 200C\alpha\delta^{-1})^{2j+2} \cdot \frac{(n-x)^2}{n^2} . \tag{6.8}$$

To express $\prod_{i=j+1}^{k+1} \mathbb{P}\left(\mathscr{L}_i \mid \mathscr{L}'_{i-1}\right)$, by definition of (A2) we have to repeat the above replacing Lemma 28(b) by Lemma 28(a). We get that

$$\prod_{i=j+1}^{k+1} \mathbb{P}\left(\mathscr{L}_i \mid \mathscr{L}'_{i-1}\right) = (1 \pm 200C\alpha\delta^{-1})^{2(k-j)+2} \cdot \frac{n-y}{n-x} . \tag{6.9}$$

Putting (6.8) and (6.9) together, we get the statement of the claim. □

Claim 29.2. *We have*

$$\prod_{i=0}^{k+1} \mathbb{P}\left(\mathscr{L}'_i \mid \mathscr{L}_i\right) = (1 \pm 100C\alpha)^{2D} \cdot \frac{1}{p(n+1-x)(n+1-y)} .$$

Proof. Suppose that we have embedded up to vertex $\max(I_i)$, and that $\mathscr{L}_i$ holds. The probability of the event $\mathscr{L}'_i$ depends on which of the cases in (B1)-(B3) applies. When $\mathscr{L}'_i$ is defined using (B1)(i) then the probability $\mathbb{P}(\mathscr{L}'_i|\mathscr{L}_i)$ is equal to $\mathbb{P}(\{z_{i+1} \hookrightarrow \mathsf{N}_H(u) \setminus \{v\}\}|\mathscr{L}_i)$. Let $X := N_H\big(\psi(\mathsf{N}^-_G(z_{i+1}))\big) \setminus \operatorname{im}\psi_{z_{i+1}-1}$ be the set of vertices in H to which we could embed z_{i+1}, given the embedding of all vertices before z_{i+1}. Suppose that the $(C\alpha, 2D+3)$-diet condition holds for $(H, \operatorname{im}\psi_{z_{i+1}-1})$. Then we have

$$\begin{aligned}\mathbb{P}\left(z_{i+1} \hookrightarrow \mathsf{N}_H(u) \setminus \{v\} | \mathscr{L}_i\right) &= \frac{|(\mathsf{N}_H(u) \setminus \{v\}) \cap X|}{|X|} = \frac{|\mathsf{N}_H(u) \cap X| \pm 1}{|X|} \\ &= \frac{(1 \pm C\alpha)p^{1+\deg^-(z_{i+1})}(n-(z_{i+1}-1)) \pm 1}{(1 \pm C\alpha)p^{\deg^-(z_{i+1})}(n-(z_{i+1}-1))} = (1 \pm 4C\alpha)p\,,\end{aligned}$$

where the last line uses the $(C\alpha, 2D+3)$-diet condition for $(H, \operatorname{im}\psi_{z_{i+1}-1})$ twice, in the denominator with the set $\psi(\mathsf{N}^-(z_{i+1}))$ and in the numerator with the set $\{u\} \cup \psi(\mathsf{N}^-(z_{i+1}))$. Recall that we assume the event $\mathscr{L}_i$, and so we have $u \notin \operatorname{im}\psi_{z_{i+1}-1}$. Therefore, the set $\{u\} \cup \psi(\mathsf{N}^-_G(z_{i+1}))$ has indeed size $1 + \deg^-(z_{i+1})$.

Likewise, when $\mathscr{L}'_i$ is defined using (B1)(ii), using (B1)(iii), or using (B3) then $\mathbb{P}(\mathscr{L}'_i|\mathscr{L}_i)$ is the probability of $\{z_{i+1} \hookrightarrow \mathsf{N}_H(v) \setminus \{u\}\}$, of $\{z_{i+1} \hookrightarrow \mathsf{N}_H(u,v)\}$, or of $\{z_i \hookrightarrow \mathsf{N}_H(v) \setminus \{u\}$, respectively. If the $(C\alpha, 2D+3)$-diet condition holds for $(H, \operatorname{im}\psi_{z_{i+1}-1})$, this probability is equal to $(1 \pm 4C\alpha)p$, $(1 \pm 4C\alpha)p^2$, or $(1 \pm 4C\alpha)p$, respectively.

Let us now deal with the terms $\mathbb{P}\left(\mathscr{L}'_j \mid \mathscr{L}_j\right)$ and $\mathbb{P}\left(\mathscr{L}'_{k+1} \mid \mathscr{L}_{k+1}\right)$ which correspond to (B2) and (B4), respectively. Suppose first that $\mathscr{L}_j$ holds. In particular, $\mathsf{N}^-(x)$ is embedded to $\mathsf{N}_H(u)$. Suppose first that the $(C\alpha, 2D+3)$-diet condition for $(H, \operatorname{im}\psi_{x-1})$ holds. With this, conditioning on the embedding up to time $x-1$, the probability of embedding x to u is $(1 \pm 2C\alpha)p^{-\deg^-(x)}\frac{1}{n+1-x}$. Similarly, if the $(C\alpha, 2D+3)$-diet condition for $(H, \operatorname{im}\psi_{y-1})$ holds, the probability of embedding y to v, provided $\mathsf{N}^-(y)$ is embedded to $\mathsf{N}_H(v)$, and conditioning on the embedding up to time $y-1$, is $(1 \pm 2C\alpha)p^{-\deg^-(y)}\frac{1}{n+1-y}$.

Thus, letting $\mathcal{F}$ be the event that the $(C\alpha, 2D+3)$-diet condition fails at least once for $(H, \operatorname{im}\psi_t)$, where t runs between 1 and y, we have

$$\begin{aligned}\prod_{i=0}^{k+1} \mathbb{P}\left(\mathscr{L}'_i \mid \mathscr{L}_i\right) = \Big(&\left((1 \pm 4C\alpha)p\right)^{\ell_1} \cdot \left((1 \pm 4C\alpha)p^2\right)^{\ell_2} \\ &\cdot (1 \pm 2C\alpha)p^{-\deg^-(x)}\tfrac{1}{n+1-x} \cdot (1 \pm 2C\alpha)p^{-\deg^-(y)}\tfrac{1}{n+1-y}\Big) \pm \mathbb{P}(\mathcal{F})\,,\end{aligned}$$

where we write ℓ_1 for the number of times (B1)(i), (B1)(ii), or (B3) applies, and ℓ_2 for the number of times (B1)(iii) applies. We have $\ell_1 + 2\ell_2 = \deg^-(x) + \deg^-(y) - 1$. Indeed, ℓ_1 and ℓ_2 count the left neighbours of x and y, but x, which is a left neighbour of y, is omitted. Finally, $\mathbb{P}(\mathcal{F}) \leq 2n^{-9}$ by Lemma 24. Thus we obtain

$$\prod_{i=0}^{k+1} \mathbb{P}\left(\mathscr{L}'_i \mid \mathscr{L}_i\right) = (1 \pm 4C\alpha)^{\ell_1+\ell_2+2}p^{-1} \cdot \tfrac{1}{n+1-x} \cdot \tfrac{1}{n+1-y} \pm 2n^{-9}\,,$$

which gives the claim since $\ell_1 + \ell_2 + 2 \leq 2D + 1$. □

Plugging Claims 29.1 and 29.2 into (6.6), we get

$$\mathbb{P}\left(x \hookrightarrow u, y \hookrightarrow v\right) = (1 \pm 500C\alpha\delta^{-1})^{4D+2} \cdot p^{-1}n^{-2}\,. \tag{6.10}$$

We now sum over the choices of (x,y) such that $xy \in E(G)$. There are $2e(G)$ such choices, so we conclude that the probability that some edge of G is embedded by *RandomEmbedding* to uv is

$$\left(1 \pm 500C\alpha\delta^{-1}\right)^{4D+2} p^{-1}n^{-2} \cdot 2e(G)$$

as desired. □

We can now estimate the probability that, again for fixed G and H, at least one edge in a given star in H is used by *RandomEmbedding*.

Lemma 30. *Given $D \in \mathbb{N}$ and $\gamma > 0$, let the constants $\delta, \varepsilon, \alpha_0, \alpha_{2n}, C$ be as in Setting 15. Then the following holds for any $\alpha_0 \le \alpha \le \alpha_{2n}$ and all sufficiently large n. Suppose that G is a graph on $[n]$ such that $\deg^-(x) \le D$ for each $x \in V(G)$, with at least $n/4$ edges and maximum degree $\Delta(G) \le n/\log n$, and H is an $(\alpha, 2D+3)$-quasirandom graph with n vertices and $p\binom{n}{2}$ edges, where $p \ge \gamma$. Let $u_1, \dots, u_k, v$ be vertices of H for some $k \le 2D+3$, and suppose u_iv is an edge of H for each i. When* RandomEmbedding *is run to embed $G[[n-\delta n]]$ into H, the probability that there is at least one u_iv to which some edge of G is embedded is*

$$\left(1 \pm 1000C\alpha\delta^{-1}\right)^{4D+2} p^{-1}n^{-2} \cdot 2ke(G)\,.$$

Proof. Given $u_1, \dots, u_k, v$ and G and H, let S be the event that there is at least one u_iv to which some edge of G is embedded.

The expected number of edges u_iv embedded to by *RandomEmbedding* is, by Lemma 29 and linearity of expectation,

$$E := \left(1 \pm 500C\alpha\delta^{-1}\right)^{4D+2} p^{-1}n^{-2} \cdot 2kE(G)\,,$$

and by inclusion-exclusion, we have

$$E - \sum_{1 \le i < i' \le k} \mathbb{P}\left(u_iv \text{ and } u_{i'}v \text{ are embedded to by } \textit{RandomEmbedding}\right) \le \mathbb{P}(S) \le E\,.$$

We thus simply have to show that the above sum, which has $\binom{k}{2} \le \binom{2D+3}{2}$ terms, is small. We will show that the probability of *RandomEmbedding* embedding to any two fixed edges $uv, u'v$ is small. This probability is equal to the sum over triples $x, x', y \in V(G)$ such that $xy, x'y \in E(G)$ of the probability that $x \hookrightarrow u$, $x' \hookrightarrow u'$ and $y \hookrightarrow v$. For any given $y \in V(G)$ there are at most $\deg_G(y)^2$ choices of (x, x'), so by Lemma 7, there are at most $2Dn\Delta(G)$ such triples. It is now enough to make the estimate for one such triple. Assuming the $(C\alpha, 2D+3)$-diet condition holds throughout *RandomEmbedding*, we embed each of x, x' and y uniformly at random into a set of size at least $\frac{1}{2}p^D\delta n \ge \frac{1}{2}\gamma^D\delta n$, so the probability of the event $x \hookrightarrow u, x' \hookrightarrow u', y \hookrightarrow v$ is at most $8\gamma^{-3D}\delta^{-3}n^{-3}$.

Finally, the probability of the $(C\alpha, 2D+3)$-diet condition failing for some $(H, \operatorname{im}\psi_i)$ is by Lemma 26 at most $2n^{-9}$. Putting this together, we have

$$\mathbb{P}(S) = \left(1 \pm 500C\alpha\delta^{-1}\right)^{4D+2} p^{-1}n^{-2} \cdot 2ke(G) \pm \binom{2D+3}{2} \cdot 2Dn\Delta(G) \cdot 8\gamma^{-3D}\delta^{-3}n^{-3} \pm 2n^{-9}\,.$$

Because $e(G) \geq n/4$ the first term in the above is $\Theta(n^{-1})$, while since $\Delta(G) \leq n/\log n$ the other two terms are of asymptotically smaller order. Since n is sufficiently large, this gives the desired result. □

In Lemma 30 we estimated the probability of using an edge in a star with a given centre and a given set R of ends. In particular, looking at all stars in H whose ends are R, we get an estimate of the expected number of them from which an edge is used in the embedding. In the following lemma we prove an upper bound on the second moment of this random variable.

Lemma 31. *Let $D \in \mathbb{N}$ and let $\gamma > 0$. Let $\delta, \varepsilon, c, C, \alpha_0, \alpha_{2n}$ be as in Setting 15. Then the following holds for any $\alpha_0 \leq \alpha \leq \alpha_{2n}$ and all sufficiently large n. Suppose that G is a graph on $[n]$ such that $\deg^-(x) \leq D$ for each $x \in V(G)$, with at least $n/4$ edges and maximum degree $\Delta(G) \leq cn/\log n$, and H is an $(\alpha, 2D+3)$-quasirandom graph with n vertices and $p\binom{n}{2}$ edges, where $p \geq \gamma$. Given $R \subseteq V(H)$ with $|R| \leq 2D+3$ and any subset T of $\mathsf{N}_H(R)$, let X count the number of vertices $v \in T$ such that an edge from v to R is used by* RandomEmbedding *when embedding G to H. Then we have*

$$\mathbb{E}(X^2) \leq 2^{30}D^4\Delta(G)\gamma^{-4D}\delta^{-4}\,.$$

Proof. We can write $X = \sum_{v\in T} W_v$, where W_v is the indicator random variable of the event that some edge from R to v is used in embedding G. We have

$$\mathbb{E}(X^2) = \sum_{(v,v')\in T^2} \mathbb{E}(W_v W_{v'}) = \mathbb{E}(X) + 2\sum_{\{v,v'\}\subseteq T} \mathbb{E}(W_v W_{v'})\,.$$

Since $e(G) \leq Dn$, by Lemma 30, applied with $\{u_1, \ldots, u_k\} = R$ and for each $v \in T$, we have

$$\mathbb{E}(X) \leq \left(1 + 1000C\alpha\delta^{-1}\right)^{4D+2} p^{-1}n^{-2} \cdot 2|R| \cdot Dn \cdot |T| \leq 4\gamma^{-1}D(2D+3)\,,$$

where we use $|R| \leq 2D+3$ and $|T| \leq n$. Thus the main task is thus to estimate $\mathbb{E}(W_v W_{v'})$ for $v \neq v'$. Now $W_v W_{v'}$ is equal to 1 if and only if there is an edge of G embedded to some edge between R and v, and another to an edge between R and v'. So, in order to refine our strategy, for $v \in T$ and $u \in R$, let $Y_{v,u}$ be the indicator random variable of the event that the edge uv is used in embedding G. For each $\{v, v'\} \subseteq T$ we have

$$\mathbb{E}(W_v W_{v'}) = \sum_{u,u'\in R, u\neq u'} \mathbb{E}(Y_{v,u}Y_{v',u'}) + \sum_{u\in R} \mathbb{E}(Y_{v,u}Y_{v',u})\,. \tag{6.11}$$

First, we focus on the first term of the right-hand side of (6.11). That is, we need to find an upper bound for the probability that two given disjoint edges xy and $x'y'$ of G are embedded to respectively uv and $u'v'$ for some fixed $u, u' \in R$ and fixed v, v'. As *RandomEmbedding* runs, either for some t we observe that the $(C\alpha, 2D+3)$-diet condition fails for $(H, \operatorname{im}\psi_t)$, or it is successful and at each time t, the vertex t is embedded uniformly at random into a set of size at least $\frac{1}{2}\gamma^D\delta n$. The probability of the former occurring is at most $2n^{-9}$ by Lemma 24, while in the latter case the probability of embedding x, y, x', y' to u, v, u', v' in that order is at most $16\gamma^{-4D}\delta^{-4}n^{-4}$. Putting these together the probability of $xy, x'y'$ being embedded to $uv, u'v'$ in that order is at most $32\gamma^{-4D}\delta^{-4}n^{-4}$. Summing over the at most $8\binom{e(G)}{2} \leq 8\binom{Dn}{2}$ choices of edges $xy, x'y'$ and their orderings, we get

$$\mathbb{E}(Y_{v,u}Y_{v',u'}) \leq 8\binom{Dn}{2} \cdot 32\gamma^{-4D}\delta^{-4}n^{-4} .$$

There are exactly $|R|^2 - |R| \leq (2D+3)^2$ choices of distinct vertices $u, u' \in R$. Hence

$$\sum_{u,u' \in R, u \neq u'} \mathbb{E}(Y_{v,u}Y_{v',u'}) \leq (2D+3)^2 \cdot 8\binom{Dn}{2} \cdot 32\gamma^{-4D}\delta^{-4}n^{-4} \leq 2^{15}D^4\gamma^{-4D}\delta^{-4}n^{-2} . \tag{6.12}$$

Next, we focus on the second term of the right-hand side of (6.11). That is, we now find an upper bound for the probability that *RandomEmbedding* uses both uv and uv' for some $u \in R$. The only way this can happen is that for some $x, y, y' \in V(G)$ with $xy, xy' \in E(G)$, the vertex x is embedded to u and y, y' to v, v'. Again, by Lemma 24, the probability that a fixed such triple x, y, y' are embedded to u, v, v' is at most $2n^{-9} + 8\gamma^{-3D}\delta^{-3}n^{-3}$. By Lemma 7 there are at most $2Dn\Delta(G)$ such triples. Hence, we get

$$\mathbb{E}(Y_{v,u}Y_{v',u}) \leq 2Dn\Delta(G) \cdot (2n^{-9} + 8\gamma^{-3D}\delta^{-3}n^{-3}) \leq 2Dn\Delta(G) \cdot 16\gamma^{-3D}\delta^{-3}n^{-3} .$$

There are exactly $|R| \leq 2D+3$ choices of u, so the probability that *RandomEmbedding* uses both uv and uv' for some $u \in R$ is at most

$$\sum_{u \in R} \mathbb{E}(Y_{v,u}Y_{v',u}) \leq (2D+3) \cdot 2Dn\Delta(G) \cdot 16\gamma^{-3D}\delta^{-3}n^{-3} \leq 2^{10}D^2\Delta(G)\gamma^{-3D}\delta^{-3}n^{-2} . \tag{6.13}$$

We can now plug in (6.12) and (6.13) into (6.11),

$$\mathbb{E}(W_vW_{v'}) \leq 2^{20}D^4\Delta(G)\gamma^{-4D}\delta^{-4}n^{-2} .$$

Summing over the at most n^2 choices of $v, v' \in T$, we obtain the desired bound. □

We are now in a position to prove Lemma 17.

Proof of Lemma 17. We define $\hat{p}$ by $e(H_0^*) = \hat{p}\binom{n}{2}$. By assumption we have $\hat{p} = (1 \pm \eta)\gamma$.

Our aim is to show that with high probability, for any given s, either *PackingProcess* fails before completing stage s or the pair (H_s, H_0^*) is $(\alpha_s, 2D+3)$-coquasirandom. Let S be a set of at most $2D+3$ vertices in $V(H_0^*)$, and let $R \subseteq S$. Recall that for (H_s, H_0^*) to be $(\alpha_s, 2D+3)$-coquasirandom means that $\mathsf{N}_{H_s}(R) \cap \mathsf{N}_{H_0^*}(S \setminus R)$ has about the size one would expect if both graphs were random. For each $1 \le i \le s$, let

$$Y_i = \left|\mathsf{N}_{H_{i-1}}(R) \cap \mathsf{N}_{H_0^*}(S \setminus R) \setminus \mathsf{N}_{H_i}(R)\right| .$$

In other words, Y_i is the number of vertices which are removed to form $\mathsf{N}_{H_i}(R) \cap \mathsf{N}_{H_0^*}(S \setminus R)$ when we embed $G_i[[n-\delta n]]$ to H_{i-1}. To prove coquasirandomness of (H_s, H_0^*), what we want is for $\sum_{i=1}^{s} Y_i$ to be sufficiently concentrated to take a union bound over choices of R and S. For this purpose we aim to apply Lemma 5 with $\mathcal{E}$ being the event that after each stage $i = 0, \ldots, s-1$ the pair (H_i, H_0^*) is $(\alpha_i, 2D+3)$-coquasirandom. The probability space in which we work is the set of all possible histories of *RandomEmbedding*, and the sequence of partitions required by Lemma 5 is given by the histories up to increasing times $1 \le i \le s$ of *RandomEmbedding*. We thus have to estimate $\mathbb{E}(Y_s|H_{s-1})$ and $\mathrm{Var}(Y_s|H_{s-1})$ only in the case (H_{s-1}, H_0^*) is $(\alpha_{s-1}, 2D+3)$-coquasirandom.

So suppose that (H_{s-1}, H_0^*) is $(\alpha_{s-1}, 2D+3)$-coquasirandom. Let p_s be such that $p_s\binom{n}{2} = e(H_s) = e(H_0) - \sum_{i=1}^{s} e(G_i[[n-\delta n]])$. Then by Lemma 30 and linearity of expectation, we have

$$\begin{aligned}
&\mathbb{E}(Y_s|H_{s-1}) \\
&= (1 \pm \alpha_{s-1}) p_{s-1}^{|R|} \hat{p}^{|S\setminus R|} n \cdot \left(1 \pm 1000C\alpha_{s-1}\delta^{-1}\right)^{4D+2} p_{s-1}^{-1} n^{-2} \cdot 2|R| e(G_s[[n-\delta n]]) \\
&= \left(2|R| \pm 10^6 C D^2 \delta^{-1} \alpha_{s-1}\right) p_{s-1}^{|R|-1} \hat{p}^{|S\setminus R|} e(G_s[[n-\delta n]])/n . \qquad (6.14)
\end{aligned}$$

We now need to estimate the sum $\sum_{i=1}^{s} \mathbb{E}(Y_i|H_{i-1})$, on the assumption that each (H_{i-1}, H_0^*) is $(\alpha_{i-1}, 2D+3)$-coquasirandom. We first estimate the sum of the main terms of (6.14). Using the facts that, and that $p_{i-1} - p_i \le 4D/n$:

$$\begin{aligned}
&\sum_{i=1}^{s} 2|R| p_{i-1}^{|R|-1} \hat{p}^{|S\setminus R|} e(G_i[[n-\delta n]])/n \\
\text{(we have } e(G_i[[n-\delta n]]) = (p_{i-1}-p_i)\tbinom{n}{2}\text{)} \quad &= \sum_{i=1}^{s} |R| p_{i-1}^{|R|-1} (p_{i-1} - p_i) \hat{p}^{|S\setminus R|} (n-1) . \qquad (6.15)
\end{aligned}$$

Note that for every $x, h \in [0,1]$ and $a \in \mathbb{N}$, we have $(x+h)^a - x^a = ah(x+h)^{a-1} \pm 2^a h^2$. We use this with $x := p_i$, $h := p_{i-1} - p_i$, and $a := |R|$, and continue (6.15) as follows:

$$\begin{aligned}
\sum_{i=1}^{s} 2|R| p_{i-1}^{|R|-1} \hat{p}^{|S\setminus R|} e(G_i[[n-\delta n]])/n &= (n-1)\hat{p}^{|S\setminus R|} \sum_{i=1}^{s} \left(\left(p_{i-1}^{|R|} - p_i^{|R|}\right) \pm 16D^2 2^{|R|}/n^2\right) \\
&= (n-1)\hat{p}^{|S\setminus R|}\left(p_0^{|R|} - p_s^{|R|}\right) \pm 64D^2 2^{|R|}
\end{aligned}$$

$$= \big(p_0^{|R|} - p_s^{|R|}\big)\hat{p}^{|S\setminus R|}n \pm 100D^2 2^{2D+3}\,. \tag{6.16}$$

Next, we bound the sum of the error terms of (6.14):

$$\begin{aligned} &\sum_{i=1}^{s} 10^6 CD^2\delta^{-1}\alpha_{i-1}p_{s-1}^{|R|-1}\hat{p}^{|S\setminus R|}e(G_i[[n-\delta n]])/n \\ \text{(we have } e(G_s)\le Dn\text{)} \quad &\le \int_{-\infty}^{s} 10^7 CD^3\delta^{-1}\alpha_x\,\mathrm{d}x \\ \text{(by (4.1))} \quad &\le \alpha_s n/4\,. \end{aligned} \tag{6.17}$$

Plugging (6.16) and (6.17) into (6.14), we get

$$\sum_{i=1}^{s}\mathbb{E}(Y_i|H_{i-1}) = \big(p_0^{|R|} - p_s^{|R|}\big)\hat{p}^{|S\setminus R|}n \pm \alpha_s n/2\,,$$

provided that H_{i-1} is $(\alpha_{i-1}, 2D+3)$-quasirandom for each $1 \le i \le s$.

Let us write $\Delta := cn/\log n$.

We wish to estimate $\mathrm{Var}(Y_i|H_{i-1})$. Trivially, we have $\mathrm{Var}(Y_s|H_{s-1}) \le \mathbb{E}(Y_i^2|H_{i-1})$. By Lemma 31,

$$\mathbb{E}(Y_s^2|H_{s-1}) \le 2^{30}D^4\Delta(G_s)\gamma^{-4D}\delta^{-4} \le 2^{30}D^4\Delta\gamma^{-4D}\delta^{-4}\,.$$

Summing this up, we obtain

$$\sum_{i=1}^{s}\mathbb{E}(Y_s^2|H_{s-1}) \le 2^{31}D^4\Delta\gamma^{-4D}\delta^{-4}n =: \sigma^2\,.$$

Furthermore, the range of each Y_i is at most $|S|\Delta(G_i) \le |S|\Delta$. We apply Lemma 5 with σ^2 as above, $\varrho = \varepsilon n$ and $\mathcal{E}$ the event that the pair (H_i, H_0^*) is $(\alpha_i, 2D+3)$-coquasirandom for each $0 \le i \le s-1$. We obtain that the probability that

$$\sum_{i=1}^{s} Y_i \ne \big(p_0^{|R|} - p_s^{|R|}\big)\hat{p}^{|S\setminus R|}n \pm (\alpha_s n/2 + \varepsilon n) = \big(p_0^{|R|} - p_s^{|R|}\big)\hat{p}^{|S\setminus R|}n \pm \tfrac{3}{4}\alpha_s n$$

is at most

$$2\exp\left(\frac{-\varepsilon^2 n^2}{2^{31}D^4\Delta\gamma^{-4D}\delta^{-4}n + 2(2D+3)\Delta\varepsilon n}\right) < n^{-2D-30}\,,$$

where the last inequality is by choice of c.

Taking the union bound over all choices of $R \subsetneq S$ and S of size at most $2D+3$, and applying Lemma 26, we see that the following event has probability at most $3n^{-9}$.

The pair (H_i, H_0^*) is $(\alpha_i, 2D+3)$-coquasirandom for each $0 \le i \le s-1$, but either *RandomEmbedding* fails to embed G_s or (H_s, H_0^*) is not $(\alpha_s, 2D+3)$-coquasirandom. Taking now the union bound over all choices of $1 \le s \le s^*$, and recalling that (H_0, H_0^*) is by assumption $\left(\frac{1}{4}\alpha_0, 2D+3\right)$-coquasirandom, we conclude that the probability that for any $1 \le s \le s^*$, *RandomEmbedding* fails to embed G_s or the pair (H_s, H_0^*) fails to be $(\alpha_s, 2D+3)$-coquasirandom is at most $1.5n^{-8}$. This completes the proof. □

7. Completing the embedding

Recall that we complete the embedding of each graph G_s by embedding the final δn vertices using only edges of H_{s-1}^*. From Setting 15, these unembedded of G_s vertices form an independent set and each of them has degree d_s. Lemma 19 states that it is very likely, provided *PackingProcess* does not fail and provided (H_s, H_0^*) is coquasirandom for each s, that only a few edges of H_0^* are used at any given vertex to form H_s^*, and hence (H_s, H_s^*) is also coquasirandom. Complementing this, Lemma 20 states that this coquasirandomness guarantees that completing the embedding is possible. We prove these two lemmas in this section.

To prove Lemma 19, we give an upper bound for the expected number of edges used at v in each stage, and apply Lemma 5 to show that the actual outcome is with high probability not much larger than this upper bound. For each $x \in V(G_s)$, we define the *completion degree* of x, written $\deg^*(x)$, to be the degree of x in the bipartite graph $G_s\big[[n-\delta n], [n]\setminus[n-\delta n]\big]$. Then the number of edges of H_0^* at v used in stage s is $\deg^*(x)$ where x is the vertex of G_s embedded to v. Note that since $\sum_{x=n-\delta n+1}^{n} \deg^*(x) = \delta n d_s$, the hand-shaking lemma tells us that

$$\sum_{x=1}^{n-\delta n} \deg^*(x) = \delta n d_s \; . \tag{7.1}$$

We note that the number of edges of H_{s-1}^* used in stage s at any given vertex v does not depend upon how the embedding of G_s is completed, but only on how *RandomEmbedding* embeds the first $n-\delta n$ vertices, so the proof of Lemma 19 will only need to analyse *RandomEmbedding*. Indeed, if some vertex $x \in V(G_s)$, $x \le n-\delta n$ is mapped onto v, then this number is $\deg^*(x)$. If on the other hand, v is not in the image of $G_s\big[[n-\delta n]\big]$ then v will be used is the completion phase. In this case, the number of edges used at v will be d_s irrespective of which particular vertex v will host.

Proof of Lemma 19. Fix $v \in V(H_0^*)$. For each $s \in [s^*]$, let Y_s be the number of edges of H_0^* at v used in stage s. We have

$$Y_s = \sum_{x \in V(G_s)} \deg^*(x)\mathbb{1}_{x \hookrightarrow v} = \sum_{x=1}^{n-\delta n} \deg^*(x)\mathbb{1}_{x \hookrightarrow v} + \sum_{x=n-\delta n+1}^{n} \deg^*(x)\mathbb{1}_{x \hookrightarrow v} \; . \tag{7.2}$$

We define $\mathcal{E}$ to be the event that *PackingProcess* succeeds and (H_{s-1}, H_0^*) is $(\alpha_{s-1}, 2D+3)$-coquasirandom for each $1 \le s \le s^*$. In other words, $\mathcal{E}$ is the complement of the first two events in the statement of Lemma 19, so to prove Lemma 19 we want to show that the probability of $\mathcal{E}$ occurring and the third event not occurring is very small.

Suppose that $\mathscr{H}_{s-1}$ is an arbitrary history of *PackingProcess* up to and including stage $s-1$ for which (H_{s-1}, H_0^*) is $(\alpha_{s-1}, 2D+3)$-coquasirandom. We begin by estimating $\mathbb{E}(Y_s|\mathscr{H}_{s-1})$.

To estimate the desired expectation, we first aim to show

$$\mathbb{P}(x \hookrightarrow v|\mathscr{H}_{s-1}) \le 5\gamma^{-D}n^{-1} \qquad \text{if } 1 \le x \le n-\delta n, \text{ and} \tag{7.3}$$

$$\mathbb{P}(\nexists x \in [1, n-\delta n] : x \hookrightarrow v|\mathscr{H}_{s-1}) \le 2\delta. \tag{7.4}$$

In order to establish (7.3) and (7.4), we need the following consequence of Lemma 28. Conditioning on $\mathscr{H}_{s-1}$, for each $1 \le t \le n-\delta n$, the probability that *RandomEmbedding* does not embed any of the first t vertices of G_s to v is at most $2\frac{n-1-t}{n} < 2\frac{n-t}{n}$. This readily establishes (7.4).

Furthermore, under the same conditioning, by Lemma 24, for each $1 \le t \le n-\delta n$, with probability at least $1-2n^{-9}$, we have $\left|C^{t-1}_{G_s \hookrightarrow H_{s-1}}(t)\right| \ge \frac{1}{2}\gamma^D(n+1-t)$. Now, for each $1 \le t \le n-\delta n$, the probability that *RandomEmbedding*, conditioning on $\mathscr{H}_{s-1}$, embeds t to v is the probability that no vertex is embedded to v at time $t-1$ times the probability of picking v when choosing uniformly from the candidate set of t. This is at most

$$2n^{-9} + 2\frac{n+1-t}{n} \cdot \frac{1}{\left|C^{t-1}_{G_s \hookrightarrow H_{s-1}}(t)\right|} \le 2n^{-9} + \frac{2}{\frac{1}{2}\gamma^D n}\,.$$

This establishes (7.3).

Now, we are going to substitute (7.3) and (7.4) into (7.2). To this end, recall that for each $x \in [n-\delta n+1, n]$ we have $\deg^*(x) = d_s$. It follows that

$$\begin{aligned}\mathbb{E}(Y_s|\mathscr{H}_{s-1}) &\le 5\gamma^{-D}n^{-1} \sum_{1 \le x \le n-\delta n} \deg^*(x) + d_s \sum_{x=n-\delta n+1}^{n} \mathbb{P}(x \hookrightarrow v|\mathscr{H}_{s-1}) \\ \text{(by (7.1), (7.4))} \quad &\le 5\gamma^{-D}n^{-1} \cdot \delta n d_s + d_s \cdot 2\delta \le 7\gamma^{-D}D\delta\,. \end{aligned} \tag{7.5}$$

Next, we obtain a similar upper bound for the second moment. Since only one vertex gets embedded to v, we have

$$\begin{aligned}\mathbb{E}(Y_s^2|\mathscr{H}_{s-1}) &= \sum_{x \in V(G_s)} \deg^*(x)^2 \mathbb{P}(x \hookrightarrow v|\mathscr{H}_{s-1}) \\ &\le \Delta(G_s) \cdot \sum_{x \in V(G_s)} \deg^*(x) \mathbb{P}(x \hookrightarrow v|\mathscr{H}_{s-1}) = \Delta(G_s) \cdot \mathbb{E}(Y_s|\mathscr{H}_{s-1})\end{aligned}$$

$$\overset{(7.5)}{\leq} 7\gamma^{-D} D\delta \cdot \Delta(G_s)\,.$$

Since $0 \leq Y_s \leq \Delta(G_s) \leq \Delta$ holds for each s, and since $s^* \leq 2n$, we can apply Lemma 5, with $\varrho = \delta n$ and with $\mathcal{E}$ as defined above, to give

$$\mathbb{P}\left(\mathcal{E} \text{ and } \sum_{i=1}^{s^*} Y_s > 50\gamma^{-D} D\delta n\right) \leq \exp\left(-\tfrac{\delta^2 n^2}{28\gamma^{-D} D\delta\cdot\Delta n + 2\Delta\delta n}\right) < n^{-100}\,,$$

where the final inequality is since $\Delta = cn/\log n$ and by choice of c. Taking the union bound over all choices of v, we see that the probability that $\mathcal{E}$ occurs and yet more than $50\gamma^{-D} D\delta n$ edges of H_0^* are deleted at any vertex in the running of *PackingProcess* is at most n^{-99}. Because the degree of each vertex in H_s^* is monotone decreasing as s increases, in particular this implies that the probability that there exists $1 \leq s \leq s^*$ such that *PackingProcess* completes stage s, and (H_i, H_0^*) is $(\alpha_i, 2D+3)$-coquasirandom for each $i < s$, yet more than $50\gamma^{-D} D\delta n$ edges of H_0^* are deleted at any vertex of H_s^*, is at most n^{-99}.

It remains to argue that since few edges are deleted at each vertex of H_0^* to form H_s^*, the pair (H_s, H_s^*) is coquasirandom. Suppose now that $\Delta\big(H_0^* - H_s^*\big) \leq 50\gamma^{-D} D\delta n$ for some s, and that (H_s, H_0^*) is $(\alpha_s, 2D+3)$-coquasirandom. Then for any $R \subseteq S \subseteq V(H_s)$ with $|S| \leq 2D+3$, we have

$$\big|\mathsf{N}_{H_s}(R) \cap \mathsf{N}_{H_0^*}(S \setminus R)\big| = (1 \pm \alpha_s) p^{|R|} \gamma^{|S\setminus R|} n$$

and hence

$$\begin{aligned}\big|\mathsf{N}_{H_s}(R) \cap \mathsf{N}_{H_s^*}(S \setminus R)\big| &= (1 \pm \alpha_s) p^{|R|} \gamma^{|S\setminus R|} n \pm (2D+3) \cdot 50\gamma^{-D} D\delta n \\ &= \big(1 \pm \eta\big) p^{|R|} \gamma^{|S\setminus R|} n\end{aligned}$$

where the final line is by choice of δ in (4.1) and since $p \geq \gamma$, so that (H_s, H_s^*) is $(\eta, 2D+3)$-coquasirandom, as desired. $\square$

Recall that Lemma 20 states that it is likely that the partial embedding ϕ_s of each G_s provided by *RandomEmbedding* can be extended to an embedding ϕ_s^* of G_s, with the completion edges used for the extension lying in H^*. Since the neighbours of each of the last δn vertices of G_s are embedded by ϕ_s, the set of candidate vertices

$$C_s^*(x) := \big\{v \in V(H_{s-1}^*) \setminus \operatorname{im}\phi_s : \phi_s(y) \in \mathsf{N}_{H_{s-1}^*}(v) \text{ for each } y \in \mathsf{N}_{G_s}(x)\big\}$$

for each x of these last δn vertices in $V(H_{s-1}^*) \setminus \operatorname{im}\phi_s$ are already fixed, and the desired ϕ_s^* exists if and only if there is a system of distinct representatives for the $C_s^*(x)$ as x ranges over the last δn vertices of G_s. Recall that Lemma 24 states in particular that $(H^*, \operatorname{im}\phi_s)$ is likely to satisfy the $(2\eta, 2D+3)$-diet condition, which implies both that

$C_s^*(x)$ is of size roughly $p^{d_s}\delta n$ for each of these last x, and also that the collection of sets is well-distributed (in a sense we will make precise later). We will see that this is almost enough to verify Hall's condition for the existence of a system of distinct representatives, but we need in addition to know that every vertex of $H_{s-1}^* - \operatorname{im}\phi_s$ is in sufficiently many of these candidate sets. The following lemma states that this typically is the case.

Lemma 32. *Let $D \in \mathbb{N}$ and let $\gamma > 0$. Let $\eta, \delta, \varepsilon, c$ and α_x be as in Setting 15. Suppose that G is a graph on vertex set $[n]$, with $\deg^-(x) \le D$ for each $x \in V(G)$, with maximum degree at most $cn/\log n$ and whose last δn vertices all have degree d, where $0 \le d \le D$, and form an independent set. Suppose that H is an $(\alpha_{s^*}, 2D+3)$-quasirandom n-vertex graph and that H^* is a graph on $V(H)$ with $(1 \pm \eta)\gamma\binom{n}{2}$ edges such that (H, H^*) forms an $(\eta, 2D+3)$-coquasirandom pair. When* RandomEmbedding *is run to embed $G[[n-\delta n]]$ into H, with probability at least $1 - 3n^{-9}$ we have that for all $v \in V(H^*)$*

$$\Big|\big\{x \in V(G) \,:\, n - \delta n < x \le n,\ \psi_{n-\delta n}(\mathsf{N}^-(x)) \subseteq \mathsf{N}_{H^*}(v)\big\}\Big| = (1 \pm 10D\eta)\gamma^d \delta n \,.$$

The proof of this lemma is similar to the proof of Lemma 25.

Proof. Fix $v \in V(H^*)$ and let I be the last δn vertices of G, which by assumption form an independent set. Denote by $\mathsf{N}_k^-(x)$ the first k neighbours of $\mathsf{N}^-(x)$. Let $\mathcal{Y}_k$ be the event that the vertices $\mathsf{N}_k^-(x)$ are all embedded to $\mathsf{N}_{H^*}(v)$ for about as many $x \in I$ as one would expect, more formally that

$$\Big|\big\{x \in I \,:\, \psi_{n-\delta n}(\mathsf{N}_k^-(x)) \subseteq \mathsf{N}_{H^*}(v)\big\}\Big| = (1 \pm 10k\eta)\gamma^k \delta n \,. \tag{7.6}$$

Let $\mathcal{B}$ be the event that the $(2\eta, 2D+3)$-codiet condition fails at some time $t \le n - \delta n$. Let

$$Z_{k,t} := \Big|\big\{x \in I \,:\, \psi_{n-\delta n}(\mathsf{N}_{k-1}^-(x)) \subseteq \mathsf{N}_{H^*}(v) \text{ and } t \text{ is the } k\text{th vertex of } \mathsf{N}^-(x)\big\}\Big| \,.$$

In other words, when we embed the vertex t, if it is embedded to $\mathsf{N}_{H^*}(v)$ it will add $Z_{k,t}$ more vertices to the set in (7.6). Let $Y_{k,t} := Z_{k,t} \cdot \mathbb{1}_{\psi_{n-\delta n}(t) \in \mathsf{N}_{H^*}(v)}$.

We want to show that if $\mathcal{Y}_{k-1}$ occurs, then $\mathcal{Y}_k$ is very likely to occur. We will then show this implies the lemma. Observe that $\mathcal{Y}_k$ is the event that $\sum_{t=1}^{n-\delta n} Y_{k,t} = (1 \pm 10k\eta)\gamma^k \delta n$. Furthermore, $\mathcal{Y}_{k-1}$ implies that $\sum_{t=1}^{n-\delta n} Z_{k,t} = (1 \pm 10(k-1)\eta)\gamma^{k-1}\delta n$. We would like to calculate $\sum_{t=1}^{n-\delta n} \mathbb{E}(Y_{k,t} | \mathscr{H}_{t-1})$, where $\mathscr{H}_{t-1}$ denotes the embedding history of *RandomEmbedding* up to and including embedding $t-1$. Given a time t, if t is the kth vertex of $\mathsf{N}^-(x)$, then at time $t-1$ the first $k-1$ vertices of $\mathsf{N}^-(x)$ have already been embedded, so $Z_{k,t}$ is determined. Thus we have

$$\mathbb{E}(Y_{k,t} | \mathscr{H}_{t-1}) = \mathbb{P}\Big(\psi_t(t) \in \mathsf{N}_{H^*}(v) \big| \mathscr{H}_{t-1}\Big) \cdot Z_{k,t} \,.$$

Suppose that at time $t-1$ we have not seen a witness that $\mathcal{B}$ fails. Then, using the $(2\eta, 2D+3)$-codiet condition once with $S = \mathsf{N}^-(t) \cup \{v\}$ and $R = \mathsf{N}^-(t) \subseteq S$ and once with $S = R = \mathsf{N}^-(t)$, we obtain

$$\mathbb{P}\Big(\psi_t(t) \in \mathsf{N}_{H^*}(v) \big| \mathscr{H}_{t-1}\Big) = \frac{(1 \pm 2\eta)(1 \pm \eta)\gamma p^{|\mathsf{N}^-(t)|}(n-t+1)}{(1 \pm 2\eta)p^{|\mathsf{N}^-(t)|}(n-t+1)} = (1 \pm 6\eta)\gamma\ .$$

Therefore, if $\overline{\mathcal{B}}$ and $\mathcal{Y}_{k-1}$ hold, we have

$$\sum_{t=1}^{n-\delta n} \mathbb{E}(Y_{k,t} \big| \mathscr{H}_{t-1}) = (1 \pm 10(k-1)\eta)(1 \pm 6\eta)\gamma^k \delta n\,.$$

Applying Lemma 4 with $\varrho = \eta\gamma^k\delta n$, we deduce that the probability that $\mathcal{Y}_k$ fails is very small. Indeed, the probability that $\overline{\mathcal{B}}$ holds but $\sum_{t=1}^{n-\delta n} Y_{k,t} \neq (1 \pm 10k\eta)\gamma^k\delta n$ is at most $2\exp\big(-\frac{\eta^2\gamma^{2k}\delta^2 n^2 \log n}{2Dcn^2}\big) \le n^{-20}$, where we use that $Y_{k,t} \le \deg(t)$ and observe that Lemma 7 gives $\sum_{t=1}^{n-\delta n} \deg(t)^2 \le 2D\Delta(G)n \le 2Dcn^2/\log n$.

As $\mathcal{Y}_0$ holds trivially with probability one, by a union bound over the choices of k and v we obtain that the probability that $\overline{\mathcal{B}}$ holds but there is some $1 \le k \le d$ for which $\mathcal{Y}_k$ fails is at most $2dn^{-19}$. Finally, Lemma 24 states that $\mathcal{B}$ holds with probability at most $2n^{-9}$, giving the lemma statement by the union bound. □

We are now in a position to prove the completion lemma, Lemma 20.

Proof of Lemma 20. Suppose H is an n-vertex $(\alpha_{s^*}, 2D+3)$-quasirandom graph, and (H, H^*) is $(\eta, 2D+3)$-coquasirandom, with $e(H) = p\binom{n}{2}$ and $e(H^*) = (1 \pm \eta)\gamma\binom{n}{2}$. Let G be a graph on $[n]$ with $\deg^-(x) \le D$ for each $x \in [n]$ and such that the last δn vertices of G form an independent set all of whose vertices have degree d. When *RandomEmbedding* is run to produce a partial embedding ϕ of G into H, by Lemma 24 with probability at least $1 - 2n^{-9}$ the algorithm succeeds and the triple $(H, H^*, \operatorname{im}\phi)$ satisfies the $(2\eta, 2D+3)$-diet condition. By Lemma 32, with probability at least $1 - 3n^{-9}$ in addition we have, for every vertex v of $V(H^*) \setminus \operatorname{im}\phi$,

$$\Big|\big\{x \in V(G) : n - \delta n < x \le n,\ \phi(\mathsf{N}^-(x)) \subseteq \mathsf{N}_{H^*}(v)\big\}\Big| = (1 \pm 10D\eta)\gamma^d\delta n\ . \tag{7.7}$$

Suppose that both good events occur, which happens with probability at least $1-5n^{-9}$. We will now show that (deterministically) this implies the existence of a system of distinct representatives for the candidate sets $\big\{C^*(x) : n-\delta n+1 \le x \le n\big\}$, which trivially gives an embedding ϕ^* of G into $H \cup H^*$ such that all edges in $[n-\delta n]$ are embedded to H and the rest to H^*, as desired.

We prove the existence of a system of distinct representatives by verifying Hall's condition. To that end, let X be a subset of $\{n - \delta n + 1, \dots, n\}$. We need to show

$$\left|\bigcup_{x\in X} C^*(x)\right| \geq |X|\,. \tag{7.8}$$

We separate three cases. The two easy cases are $|X| \leq \frac{1}{2}\gamma^D\delta n$ and $|X| \geq \delta n - \frac{1}{2}\gamma^D\delta n$. For the former, if $X = \emptyset$ the statement is trivial. If not, pick any $x \in X$. We have

$$\big|C^*(x)\big| \geq (1-2\eta)(1-\eta)^d\gamma^d\delta n \geq \tfrac{1}{2}\gamma^D\delta n \tag{7.9}$$

since $\mathsf{N}_G(x)$ is a set of $d \leq D$ vertices and $(H^*, \operatorname{im}\phi)$ satisfies the $(\eta, 2D+3)$-diet condition, which in particular verifies (7.8). For the latter, by (7.7) and choice of η, every vertex of $V(H^*)\setminus\operatorname{im}\phi$ is in more than $\frac{3}{4}\gamma^d\delta n$ of the sets $C^*(x)$ for $x \in \{n-\delta n+1, \dots, n\}$. In particular, every vertex $v \in V(H^*)\setminus\operatorname{im}\phi$ is in $C^*(x)$ for some $x \in X$, giving (7.8).

The final, harder, case is $\frac{1}{2}\gamma^D\delta n < |X| < \delta n - \frac{1}{2}\gamma^D\delta n$. Given X in this size range, let X' be a maximal subset of X with the property $\mathsf{N}_G(x)\cap\mathsf{N}_G(x') = \emptyset$ for each $x, x' \in X'$. Since each vertex of X' has $d \leq D$ neighbours, the set $Y = \bigcup_{x\in X'}\mathsf{N}_G(x)$ has size at most $D|X'|$. By maximality of X', every vertex in X is adjacent to some vertex of Y. Since no vertex of Y has degree more than $\Delta(G) \leq cn/\log n$, we conclude

$$\tfrac{1}{2}\gamma^D\delta n < |X| \leq \Delta(G)|Y| \leq \Delta(G)D|X'| \leq cnD|X'|/\log n\,,$$

and hence $|X'| \geq \log n$ by choice of c in (4.1). We will now argue that $Z := \bigcup_{x\in X'} C^*(x)$ satisfies $|Z| \geq \big(1-\frac{1}{2}\gamma^D\big)\delta n$, which implies (7.8).

Suppose for a contradiction that $|Z| < \big(1-\frac{1}{2}\gamma^D\big)\delta n$. By definition, we have $C^*(x) \subseteq Z$ for each $x \in X'$. We now aim to estimate the number N of triples (x, x', z) with $x, x' \in X$ distinct and $z \in Z$ satisfying $z \in C^*(x)\cap C^*(x')$. For each z, let $d_z = \big|\{x \in X' : z \in C^*(x)\}\big|$. Using Jensen's inequality (since $\binom{\cdot}{2}$ is convex), we have

$$\begin{aligned}
N = \sum_{z\in Z}\binom{d_z}{2} &\geq |Z|\cdot\binom{|Z|^{-1}\sum_{z\in Z} d_z}{2} \\
\text{(by (7.9))}\qquad &\geq |Z|\cdot\binom{|Z|^{-1}|X'|(1-2\eta)(1-\eta)^d\gamma^d\delta n}{2} \\
&= \tfrac{1}{2}|X'|(1-2D\eta)\gamma^d\delta n\big(|Z|^{-1}|X'|(1-2D\eta)\gamma^d\delta n - 1\big) \\
&\geq \tfrac{1}{2}(1-2D\eta)^3|X'|^2|Z|^{-1}\gamma^{2d}\delta n \\
&\geq \tfrac{1}{2}(1-2D\eta)^3|X'|^2\big(1-\tfrac{1}{2}\gamma^D\big)^{-1}\gamma^{2d}n\,,
\end{aligned}$$

where the penultimate inequality holds since $|Z| < \delta n$ and $|X'| \geq \log n$ is sufficiently large, and the final inequality uses our assumed upper bound on $|Z|$. On the other hand, since $\mathsf{N}_G(x)$ and $\mathsf{N}_G(x')$ are disjoint, we have

$$N = \sum_{x,x'\in X'}\big|C^*(x)\cap C^*(x')\big| \leq \tbinom{|X'|}{2}(1+2\eta)(1+\eta)^{2d}\gamma^{2d}\delta n \leq \tfrac{1}{2}|X'|^2(1+4D\eta)\gamma^{2d}\delta n$$

using the $(2\eta, 2D+3)$-diet condition which $(H^*, \operatorname{im}\phi)$ satisfies. We conclude

$$\tfrac{1}{2}(1-2D\eta)^3|X'|^2\big(1-\tfrac{1}{2}\gamma^D\big)^{-1}\gamma^{2d}n \le \tfrac{1}{2}|X'|^2(1+4D\eta)\gamma^{2d}\delta n$$

which is false since by choice of η in (4.1) we have $(1-2D\eta)^3(1+4D\eta)^{-1} > 1-\frac{1}{2}\gamma^D$. Thus (7.8) holds for all X, so the desired ϕ^* exists. □

8. Concluding remarks

8.1. Constants in Theorem 2

Given γ and D in Theorem 2, the constant c is set in Setting 15. All the dependencies in (4.1) are polynomial, except for the exponentials used to define C and α_x. As a result, c depends roughly doubly-exponentially on D and γ, more precisely $c \approx \exp(-\exp(D^{5+o(1)} \cdot \gamma^{-24D-10+o(1)}))$ (where $o(1) \to 0$ as $D, 1/\gamma \to \infty$). This of course puts an implicit requirement on n_0, as instances of the result for which the maximum degree bound $\frac{cn}{\log n}$ are less than 1 are vacuous.

By way of brief comparison with other recent packing results, we believe most of the results we cited earlier obtain broadly similar or better constant dependencies to our results (though these bounds are generally not given explicitly and we did not check carefully), unless the Regularity Lemma is used.

8.2. Limits of the method

As Ferber and Samotij [11] point out, a randomised strategy such as the one we use here will not succeed in packing graphs with many vertices of degree $\omega\big(\frac{n}{\log n}\big)$, because it is likely to put these vertices unevenly into the host graph and after packing only half the guest graphs one vertex will probably have degree substantially less than the average. If the remaining graphs are for example Hamilton cycles, this vertex will become a bottleneck which causes the strategy to fail. One might try to pick vertices non-uniformly in order to correct such imbalances as they form, but analysing such a strategy would be challenging and it is not clear that it would work: common neighbourhoods of several vertices will also occasionally be far from the expected size.

Although it might well be that we can obtain near-perfect packings of graphs with degeneracy much bigger than $\log n$ into K_n, any strategy like the one we use here will certainly not succeed in doing so. The reason is simply that strategies like ours work by maintaining quasirandomness, and hence work equally well starting with a dense random graph rather than the complete graph. Take H to be a clique of order $3\log_2 n$. Then a well-known calculation shows that $\mathbb{G}\big(n, \frac{1}{2}\big)$ typically does not even contain one copy of H.

We have not tried to analyse our approach more carefully in order to work with sparse random or quasirandom graphs. We are confident that (with substantially more work, and using ideas from [2]) one could prove a near-perfect packing result for typical $\mathbb{G}(n,p)$, where $p > n^{-\varepsilon}$ for some $\varepsilon > 0$ depending on the degeneracy bound D. But we suspect that our approach would not then allow for maximum degrees of the guest graphs as large as $\Omega(pn/\log n)$, even if we asked only to pack almost-spanning graphs, and certainly we cannot take ε as big as $\frac{1}{2D+3}$, since at this point $\mathbb{G}(n,p)$ itself is typically not $(\frac{1}{2}, 2D+3)$-quasirandom. In particular, our approach cannot challenge the tree packing results of [11] in sparse random graphs.

8.3. *Perfect packings*

It is easy to check that the graph of uncovered edges in the packing of Theorem 11 is $(2\eta, 2D+3)$-quasirandom, and η can be chosen arbitrarily small by increasing D if necessary. In particular, this means that the result of Joos, Kim, Kühn and Osthus [18] applies to this leftover. Thus we can extend the result of [18] on the Tree Packing Conjecture to allow many trees where the maximum degree is bounded only by $\frac{cn}{\log n}$, provided that it is bounded by D in the remainder. This is however a rather peculiar condition.

Acknowledgments

We thank Pavel Hladký for looking after our little children while we were proving these results. We thank the anonymous referees for their very helpful comments.

The contents of this publication reflects only the authors' views and not necessarily the views of the European Commission of the European Union.

References

[1] A. Adamaszek, P. Allen, C. Grosu, J. Hladký, Almost all trees are almost graceful, arXiv:1608.01577.
[2] P. Allen, J. Böttcher, H. Hàn, Y. Kohayakawa, Y. Person, Blow-up lemmas for sparse graphs, arXiv:1612.00622.
[3] J. Balogh, C. Palmer, On the tree packing conjecture, SIAM J. Discrete Math. 27 (4) (2013) 1995–2006.
[4] B. Barber, D. Kühn, A. Lo, R. Montgomery, D. Osthus, Fractional clique decompositions of dense graphs and hypergraphs, J. Combin. Theory Ser. B 127 (2017) 148–186.
[5] B. Bollobás, Some remarks on packing trees, Discrete Math. 46 (2) (1983) 203–204.
[6] J. Böttcher, J. Hladký, D. Piguet, A. Taraz, An approximate version of the tree packing conjecture, Israel J. Math. 211 (1) (2016) 391–446.
[7] F.R.K. Chung, R.L. Graham, R.M. Wilson, Quasi-random graphs, Combinatorica 9 (4) (1989) 345–362.
[8] B. Csaba, D. Kühn, A. Lo, D. Osthus, A. Treglown, Proof of the 1-factorization and Hamilton decomposition conjectures, Mem. Amer. Math. Soc. 244 (1154) (2016) 170.
[9] D. Dor, M. Tarsi, Graph decomposition is NP-complete: a complete proof of Holyer's conjecture, SIAM J. Comput. 26 (4) (1997) 1166–1187.
[10] A. Ferber, C. Lee, F. Mousset, Packing spanning graphs from separable families, Israel J. Math. 219 (2) (2017) 959–982.
[11] A. Ferber, W. Samotij, Packing trees of unbounded degrees in random graphs, J. Lond. Math. Soc. 99 (3) (2019) 653–677, arXiv:1607.07342.
[12] D.A. Freedman, On tail probabilities for martingales, Ann. Probab. 3 (1975) 100–118.

[13] S. Glock, D. Kühn, A. Lo, R. Montgomery, D. Osthus, On the decomposition threshold of a given graph, arXiv:1603.04724, J. Combin. Theory Ser. B (2019), in press.
[14] S. Glock, D. Kühn, A. Lo, D. Osthus, The existence of designs via iterative absorption, arXiv:1611.06827.
[15] S. Glock, D. Kühn, A. Lo, D. Osthus, Hypergraph F-designs for arbitrary F, arXiv:1706.01800.
[16] A. Gyárfás, J. Lehel, Packing trees of different order into K_n, in: Combinatorics, Proc. Fifth Hungarian Colloq., Keszthely, 1976, in: Colloq. Math. Soc. János Bolyai, vol. 18, North-Holland, Amsterdam, 1978, pp. 463–469.
[17] A.M. Hobbs, Packing trees, in: Proceedings of the Twelfth Southeastern Conference on Combinatorics, Graph Theory and Computing, vol. II, Baton Rouge, La., 1981, in: Congressus Numerantium, vol. 33, 1981, pp. 63–73.
[18] F. Joos, J. Kim, D. Kühn, D. Osthus, Optimal packings of bounded degree trees, arXiv:1606.03953, J. Eur. Math. Soc. (2019), in press.
[19] P. Keevash, The existence of designs, arXiv:1401.3665.
[20] J. Kim, D. Kühn, D. Osthus, M. Tyomkyn, A blow-up lemma for approximate decompositions, Trans. Amer. Math. Soc. 371 (7) (2019) 4655–4742.
[21] T.P. Kirkman, On a problem in combinations, Cambridge and Dublin Math. J. 2 (1847) 191–204.
[22] S. Messuti, V. Rödl, M. Schacht, Packing minor-closed families of graphs into complete graphs, J. Combin. Theory Ser. B 119 (2016) 245–265.
[23] J. Plücker, System der analytischen Geometrie, auf neue Betrachtungsweisen gegründet, und insbesondere eine ausführliche Theorie der Curven dritter Ordnung enthaltend, Duncker und Humboldt, Berlin, 1835.
[24] D.K. Ray-Chaudhuri, R.M. Wilson, Solution of Kirkman's schoolgirl problem, in: Combinatorics, Proc. Sympos. Pure Math., vol. XIX, Univ. California, Los Angeles, Calif., 1968, Amer. Math. Soc., Providence, R.I., 1971, pp. 187–203.
[25] G. Ringel, Problem 25, in: Theory of Graphs and Its Applications, Proc. Int. Symp., Smolenice, 1963, Czech. Acad. Sci., Prague, 1963.
[26] J. Steiner, Combinatorische aufgabe, J. Reine Angew. Math. 45 (1853) 181–182.
[27] R. Yuster, Combinatorial and computational aspects of graph packing and graph decomposition, Comput. Sci. Rev. 1 (1) (2007) 12–24.