RESEARCH THESIS

Author's Declaration

Full name (block capitals, surname first): ESPUNY DÍAZ, ALBERTO

Full title of thesis/dissertation (block capitals): HAMILTONICITY PROBLEMS IN RANDOM GRAPHS

College/School/Department (block capitals): SCHOOL OF MATHEMATICS

Date of award of degree (leave blank):

1. I understand that one printed and one electronic copy of my thesis/dissertation (the Work) will be deposited in the University Library (the Library) and in a suitable electronic repository, for permanent retention.
2. Without changing the content, the Library or any third party with whom it has an agreement to do so, may convert either copy into a format or medium for the purpose of long-term retention, accessibility and preservation.
3. The Library will publish, and/or arrange with appropriate third parties for the non-exclusive publication of, a bibliographic description of the thesis/dissertation, and the author's abstract.
4. Unless arrangements are made to the contrary, (see paragraph 6. below), the Library is authorised to make the Work available for consultation in the Library, and via a recognised inter library loans system. The Library is authorised to make the electronic copy of the Work freely accessible to individuals and institutions - including automated agents - via the Internet.
5. Rights granted to the University of Birmingham through this agreement are entirely non-exclusive. I retain all my rights in the Work in its present version or future derivative works.
6. I understand that I may apply to the University to retain the right to withhold access to the content of my thesis/dissertation. Access to the paper version may be withheld for a period which shall not normally exceed four calendar years from the congregation at which the degree is conferred. The actual length of the period will be specified in the application, together with the precise reasons for making that application. The electronic copy may be withheld from dissemination via the web or other networks for any period.
7. I have obtained permission for any use made of substantial amounts of published or unpublished copyright material (text, illustrations, etc) where the rights are owned by a third party, other than as permitted under either The Copyright Designs and Patents Act 1988 (as modified by any related or successor legislation) or the Terms and Conditions of any Licence governing its use.
8. The content of the copies I shall deposit with the Library will be the final version of my thesis, as approved by the Examiners.
9. I understand that the Library and administrators of any electronic theses repository do not hold any obligation to take legal action on behalf of myself, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.
10. I understand that, in the event of my thesis/dissertation being not approved by the Examiners, this declaration will become null and void.

Signature: A. Espuny **Date:** 22/06/2020

For Library use (please leave blank):

Classmark: **Accession number:** **Control number:**

eTheses Repository url:

HAMILTONICITY PROBLEMS IN RANDOM GRAPHS

by

ALBERTO ESPUNY DÍAZ

A thesis submitted to the University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Mathematics
College of Engineering and
Physical Sciences
University of Birmingham
June 2020

ABSTRACT

In this thesis, we present some of the main results proved by the author while fulfilling his PhD. While we present all the relevant results in the introduction of the thesis, we have chosen to focus on two of the main ones.

First, we show a very recent development about Hamiltonicity in random subgraphs of the hypercube, where we have resolved a long standing conjecture dating back to the 1980s.

Second, we present some original results about correlations between the appearance of edges in random regular hypergraphs, which have many applications in the study of subgraphs of random regular hypergraphs. In particular, these applications include subgraph counts and property testing.

Para Laura.

ACKNOWLEDGEMENTS

First and foremost, I need to express my deep gratitude to my supervisors, Daniela and Deryk. They have guided my research during the past four years and have provided invaluable insights and advice. They have chosen the most interesting problems to work on, and have always pushed me to pursue a bit more. Without them, nothing in this thesis would have been possible.

I also need to thank all my coauthors: Felix, Jaehoon, António and, of course, Padraig. Our discussions have provided many of my favourite mathematical moments of the past years, and they have always done their best to solve each of the problems we have faced.

I am also grateful to the European Research Council (ERC), from which I have received partial funding under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 786198).

I doubt the past few years would have been so enjoyable without all of my colleagues from Birmingham. While I risk forgetting many names, I want to thank my fellow PhD students, office mates and colleagues (Jon, Andy, 'Good' Matt, Joel, Tejas, Ryan, Joseph, Gianmarco, Raúl, Launghoi, 'Bad' Matt, Tássio, Stefan, Jack, Rob, Frederik, Andrea, Amarja, Nicolás, Arthur, Bertille, Yahya, Vincent, Stephen, Euan, John, Alex...), as well as everyone in the combinatorics group (Andrew, Richard, Allan, Richard, Guillem, Nikolaos, Henning, Johannes, Eoin, Michal, Richard, Michelle, Andrey and Aria, and Tom) for the very good times I had in Birmingham and the fond memories I will keep forever. They have always been available to discuss maths, chat, or generally have a good time.

I also want to thank all my friends from Spain for dealing with me whenever I needed help, with a special shout-out to Isaac, for choosing to accompany me in so many projects and helping me out so very often. I also need to mention Fran, Roberto, Chema and Alba, as well as Guillermo, Marta, Ander, Francesc and Álex. I also need to thank my lecturers and the people who have helped reach this point, with a special mention to Oriol and Juanjo. I really cannot stress their influence enough.

Finally, I need to thank my family and, of course, my dear Laura, for their continued support and affection. I would not have made it so far without them.

CONTENTS

CHAPTER 1

A QUEST FOR HAMILTONICITY

One of the most well-known and well-studied properties in graph theory is *Hamiltonicity*. A graph G is said to be *Hamiltonian* whenever it contains a cycle which covers all of the vertices of G. We refer to such a cycle as a *Hamilton cycle*.

As happens with many graph properties, one would like to have a general characterization of all graphs which contain a Hamilton cycle. However, the problem of determining whether or not a graph is Hamiltonian is NP-complete, as shown by Karp [74]. Thus, unless P = NP, there is little hope of obtaining such a characterization, and so the study of Hamiltonicity focuses on finding sufficient conditions. Many of these sufficient conditions have taken the form of degree conditions. One of the best-known examples of such results is Dirac's theorem [38], which states that every graph G on $n \geq 3$ vertices with minimum degree at least $n/2$ is Hamiltonian.

On a different direction, a lot of research has been devoted to understanding the *typical* properties of graphs, via the study of random graphs. Hamiltonicity has been the focus of much of this research, and this will be the main topic of this thesis. We defer an overview of some of the main results in this area, as well as the main contributions of this thesis, to the upcoming sections.

While Hamiltonicity concerns itself with the *existence* of a Hamilton cycle, many other questions can be asked about Hamilton cycles in graphs. For instance, the containment of a certain number of edge-disjoint Hamilton cycles and the problem of counting the number of Hamilton cycles in a given graph will also be relevant for this thesis. Other properties that have been considered are packings of Hamilton cycles or decompositions of graphs into Hamilton cycles. Furthermore, all of these problems can be generalised to other settings, such as directed graphs, tournaments, or hypergraphs. For a general overview of Hamiltonicity results and open problems, as well as related problems in the area, we recommend some of the excellent surveys that Gould [60, 61, 62] and Kühn and Osthus [90, 91] have written about the topic, as well as the references therein.

Hamiltonicity is often tied with other graph properties. In the exposition of this thesis, the containment of perfect matchings and connectivity of graphs will be especially relevant. A *matching* in a graph G is a set of edges of G no pair of which share an endpoint. A matching is said to be *perfect* if its edges cover all of the vertices of G. Connectivity (that is, the property that any vertex of G can be reached from any other vertex of G by moving along edges), on the other hand, is clearly a necessary condition for Hamiltonicity.

1.1. Hamiltonicity in random graphs

As has already been mentioned, Hamiltonicity has been a central topic in the study of random graphs. We now provide a short survey of some of the main results in this area, as well as the main contributions of this thesis. In particular, results can be very varied depending on the model of random graph that is being considered. In section 1.1.1 we will discuss the binomial random graph $G_{n,p}$, as well as some generalisations and related problems where it plays a central role, and in section 1.1.2 we will mention some results about Hamiltonicity in random regular graphs. We will then mention hitting time results in section 1.1.3, before talking about random binomial subgraphs of the hypercube and hitting time results in subgraphs of the hypercube in section 1.1.4. Finally, we will devote section 1.1.5 to briefly talking about randomly perturbed graphs.

Note that this is not a comprehensive survey of all results about Hamiltonicity in random graphs, and we do not attempt to cover all models of random graphs where results about Hamiltonicity are known (for instance, random geometric graphs or preferential attachment graphs do not appear here). For a general overview of the state of the art of Hamiltonicity in random graphs, we recommend the survey of Frieze [55].

1.1.1. The binomial random graph

One of the most studied random graph models is the *binomial random graph* model $\mathcal{G}_{n,p}$, where we have a (labelled) set of n vertices and $G_{n,p}$ is obtained by including each edge with probability p, independently of all other edges. This is deeply tied with the $\mathcal{G}_{n,m}$ model introduced by Erdős and Rényi [47, 48], where a (labelled) graph $G_{n,m}$ is chosen uniformly at random from the set of all graphs on n vertices with exactly m edges. In the second of their seminal papers, Erdős and Rényi already asked for the necessary number of edges m (as a function of n) for which $G_{n,m}$ asymptotically almost surely (a.a.s.) contains a Hamiltonian line (that is, a path which contains all vertices of the graph). This is widely regarded as the beginning of the study of Hamiltonicity in random graphs.

The question about the necessary number of edges can be formalised via the concept of *thresholds*. Let us consider these in the $\mathcal{G}_{n,p}$ model. Given some monotone increasing graph property $\mathcal{P}$, a function $p^* = p^*(n)$ is said to be a (coarse) *threshold* for $\mathcal{P}$ if $\mathbb{P}[G_{n,p} \in \mathcal{P}] \to 1$ whenever $p/p^* \to \infty$ and $\mathbb{P}[G_{n,p} \in \mathcal{P}] \to 0$ whenever $p/p^* \to 0$. One can define the stronger notion of a sharp threshold similarly: $p^* = p^*(n)$ is said to be a *sharp threshold* for $\mathcal{P}$ if, for all $\varepsilon > 0$, we have that $\mathbb{P}[G_{n,p} \in \mathcal{P}] \to 1$ whenever $p \geq (1+\varepsilon)p^*$ and $\mathbb{P}[G_{n,p} \in \mathcal{P}] \to 0$ whenever $p \leq (1-\varepsilon)p^*$. Bollobás and Thomason [24] proved that every (non-trivial) monotone graph property has a threshold in $\mathcal{G}_{n,p}$. Most of the natural graph properties that we work with satisfy these assumptions, so we know that their thresholds exist; it is now a matter of determining what those thresholds are.

Erdős and Rényi [49] proved that the threshold for the containment of a perfect matching in $\mathcal{G}_{n,p}$ is $\log n/n$ (if n is even). The threshold for the containment of a Hamilton cycle was found independently by Pósa [100] and Koršunov [82, 83], who actually determined the sharp threshold to be $p^* = \log n/n$. This is also the sharp threshold for the property of having minimum degree at least 2. In this sense, the results about Hamilton cycles in $\mathcal{G}_{n,p}$ can be interpreted as saying that the natural obstruction of having sufficiently high minimum degree is also an "almost sufficient" condition. As a generalisation, the sharp threshold for the containment of k edge-disjoint Hamilton cycles in $\mathcal{G}_{n,p}$, for some $k \in \mathbb{N}$ independent of n, is $p^* = \log n/n$, i.e., the same as the threshold for Hamiltonicity.

One more recent approach to extend the classical extremal results to random graphs is based on the following concept of *resilience*. The *local resilience* of a graph G with respect to some property $\mathcal{P}$ is the maximum number r such that, for any subgraph $H \subseteq G$ with $\Delta(H) < r$, we have $G \setminus H \in \mathcal{P}$. One may view this concept as a measure of the damage an adversary can commit at each vertex of G, without destroying the property $\mathcal{P}$. The systematic study of local resilience was initiated by Sudakov and Vu [110]. Restated in this terminology, Dirac's theorem says that the local resilience of the complete graph K_n with respect to Hamiltonicity is $\lfloor n/2 \rfloor$.

This concept of resilience naturally suggests a generalisation of Dirac's theorem in the setting of random graphs. Lee and Sudakov [93] proved that, when $p = C \log n/n$ and C is sufficiently large, the local resilience of the random graph $G_{n,p}$ with respect to Hamiltonicity is a.a.s. at least $(1/2 - \varepsilon)np$, extending Dirac's theorem to random graphs. This improved on earlier bounds [11, 56, 110], and is equivalent to the following statement.

Theorem 1.1.1 (Lee and Sudakov [93]). *For every positive ε, there exists a constant $C = C(\varepsilon)$ such that, for $p \geq C \log n/n$, a.a.s. every subgraph of $G_{n,p}$ with minimum degree at least $(1/2 + \varepsilon)np$ is Hamiltonian.*

In this context, together with Condon, Kim, Kühn and Osthus [35], we provide a generalization of theorem 1.1.1 where we allow the degrees of roughly half of the vertices to be lower, effectively extending Pósa's theorem [101] to random graphs.

Theorem 1.1.2 ([35]). *For every $\varepsilon > 0$, there exists a constant $C = C(\varepsilon)$ such that, for $p \geq C \log n/n$, a.a.s. every subgraph of $G_{n,p}$ with degree sequence $d_1 \leq \ldots \leq d_n$ which for all $i < n/2$ satisfies that $d_i \geq (i + \varepsilon n)p$ is Hamiltonian.*

We also obtained a more general degree sequence result for perfect matchings.

Theorem 1.1.3 ([35]). *For every $\varepsilon > 0$, there exists a constant $C = C(\varepsilon)$ such that, for $p \geq C\log n/n$, a.a.s. every subgraph of $G_{n,p}$ with degree sequence $d_1 \leq \ldots \leq d_n$ which for all $i < n/2$ satisfies that $d_i \geq (i + \varepsilon n)p$ or $d_{n-i-\varepsilon n} \geq (n - i + \varepsilon n)p$ contains a matching of size $\lfloor n/2 \rfloor$.*

We further conjecture that the conditions stated in theorem 1.1.3 should suffice to guarantee the a.a.s. existence of a Hamilton cycle. If true, this would be an extension of Chvátal's theorem [32] to random graphs. Both results above, while stated in terms of degree sequences, can be rephrased in terms of local resilience (see the original paper [35] for details).

On a different but related direction, rather than considering subgraphs of a random graph, we can consider random subgraphs of a given graph G. This leads to the study of *robustness* of graph properties, which has also attracted much attention recently. For instance, given a graph G which is known to satisfy some property $\mathcal{P}$, consider a random subgraph G_p obtained by deleting each edge of G with probability $1 - p$, independently of all other edges. The problem then is to determine the range of p for which G_p satisfies $\mathcal{P}$ with high probability. In this setting, a result of Krivelevich, Lee, and Sudakov [87] asserts that, for any n-vertex graph G with minimum degree at least $n/2$, the graph G_p is asymptotically almost surely Hamiltonian whenever $p \gg \log n/n$. This can be viewed as a robust version of Dirac's theorem on Hamilton cycles.

1.1.2. Random regular graphs

The search for Hamilton cycles in other random graph models has proven more difficult. This is the case, for instance, with random *regular* graphs model $\mathcal{G}_{n,d}$: given $n, d \in \mathbb{N}$ such that $d < n$ and nd is even, $G_{n,d}$ is chosen uniformly at random from the set of all d-regular graphs on n vertices (a graph is said to be d-regular if every one of its vertices has degree d). The study of this model is often more challenging than that of $\mathcal{G}_{n,p}$ due to the fact that the presence and absence of edges in $G_{n,d}$ are correlated. Several different techniques have been developed to deal with this model, such as the configuration model or edge-switching techniques. Robinson and Wormald [103] proved that $G_{n,3}$ is a.a.s. Hamiltonian, and later extended this result to $G_{n,d}$ for any fixed $d \geq 3$ [104]. This is in contrast to $G_{n,p}$, where the average degree must be logarithmic in n to ensure Hamiltonicity. These results were later generalised by Cooper, Frieze, and Reed [37] and Krivelevich, Sudakov, Vu, and Wormald [89] for the case when d is allowed to grow with n, up to $d \leq n-1$.

The resilience of random regular graphs with respect to Hamiltonicity has also been considered. Ben-Shimon, Krivelevich, and Sudakov [10] proved that, for large (but constant) d, a.a.s. $G_{n,d}$ is $(1-\varepsilon)d/6$-resilient with respect to Hamiltonicity. However, they conjectured that the true value of the (likely) local resilience of $G_{n,d}$ with respect to Hamiltonicity should be closer to $d/2$, and also suggested to study the same problem when d is allowed to grow with n. In this direction, it is worth pointing out that, for any fixed $\varepsilon > 0$ and $d \gg \log n$, the fact that a.a.s. $G_{n,d}$ is $(1/2-\varepsilon)d$-resilient with respect to Hamiltonicity follows by combining results of Sudakov and Vu [110], Krivelevich, Sudakov, Vu, and Wormald [89], Cook, Goldstein, and Johnson [36] and Tikhomirov and Youssef [111], and results of Kim and Vu [76] and Lee and Sudakov [93] (see [33] for details).

In joint work with Condon, Girão, Kühn, and Osthus [33], we completely resolved the question about the local resilience of random regular graphs. This can be seen as a version of Dirac's theorem for random regular graphs. Our main result gives the following lower bound (the upper bound follows from edge distribution properties of random regular graphs).

Theorem 1.1.4. *For every $\varepsilon > 0$ there exists D such that, for every $D < d \leq \log^2 n$, the random graph $G_{n,d}$ is a.a.s. $(1/2-\varepsilon)d$-resilient with respect to Hamiltonicity.*

While we do not try to optimise the dependency of D on ε, we remark that D in theorem 1.1.4 can be taken to be polynomial in ε^{-1}. This is essentially best possible in the sense that theorem 1.1.4 fails if $d \leq (2\varepsilon)^{-1}$.

Theorem 1.1.5. *For any odd $d > 2$, the random graph $G_{n,d}$ is not a.a.s. $(d-1)/2$-resilient with respect to Hamiltonicity.*

We point out that the methods we used for theorem 1.1.2 and theorem 1.1.4 follow the rotation-extension technique, which exploits expansion properties of graphs. For theorem 1.1.2, we use a "thinning" idea introduced by Ben-Shimon, Krivelevich, and Sudakov [10] and improve on an analysis of Lee and Sudakov [93]. For theorem 1.1.4, we combine the configuration model with the "thinning" approach and an idea of Montgomery [97]. For the detailed proofs, see the original papers [35, 33].

1.1.3. Hitting time results

Remarkably, the intuition that having the necessary minimum degree is an "almost sufficient" condition for Hamiltonicity, as we established in section 1.1.1, can be strengthened greatly via so-called hitting time results. These are expressed in terms of random graph processes (which can be seen as a generalization of the $\mathcal{G}_{n,m}$ model). The general setting is as follows. For an integer n, we denote $[n] \coloneqq \{i \in \mathbb{Z} : 1 \leq i \leq n\}$ and $[n]_0 \coloneqq \{i \in \mathbb{Z} : 0 \leq i \leq n\}$. Let G be an n-vertex graph with $m = m(n)$ edges, and consider an arbitrary labelling $E(G) = \{e_1, \dots, e_m\}$. The *G-process* is defined as a random sequence of nested graphs $\tilde{G}(\sigma) = (G_t(\sigma))_{t=0}^m$, where σ is a permutation of $[m]$ chosen uniformly at random and, for each $i \in [m]_0$, we set $G_i(\sigma) = (V(G), E_i)$, where $E_i \coloneqq \{e_{\sigma(j)} : j \in [i]\}$. Given any monotone increasing graph property $\mathcal{P}$ such that $G \in \mathcal{P}$, the *hitting time* for $\mathcal{P}$ in the above G-process is the random variable $\tau_{\mathcal{P}}(\tilde{G}(\sigma)) \coloneqq \min\{t \in [m]_0 : G_t(\sigma) \in \mathcal{P}\}$.

Let us denote the properties of containing a perfect matching by $\mathcal{PM}$, and Hamiltonicity by $\mathcal{HAM}$. For any $k \in \mathbb{N}$, let δk denote the property of having minimum degree at least k, and let $\mathcal{HM}k$ denote the property of containing $\lfloor k/2 \rfloor$ edge-disjoint Hamilton cycles and, if k is odd, one matching of size $\lfloor n/2 \rfloor$ which is edge-disjoint from these Hamilton cycles. With this notion of hitting times, many of the results about thresholds presented in section 1.1.1 can be strengthened significantly. For instance, Bollobás and Thomason [23] showed that, if n is even, then a.a.s. $\tau_{\mathcal{PM}}(\tilde{K}_n(\sigma)) = \tau_{\delta 1}(\tilde{K}_n(\sigma))$. Ajtai, Komlós, and Szemerédi [1] and Bollobás [18] independently proved that a.a.s. $\tau_{\mathcal{HAM}}(\tilde{K}_n(\sigma)) = \tau_{\delta 2}(\tilde{K}_n(\sigma))$. Later, Bollobás and Frieze [21] proved that, given any $k \in \mathbb{N}$, for n even a.a.s. $\tau_{\mathcal{HM}k}(\tilde{K}_n(\sigma)) = \tau_{\delta k}(\tilde{K}_n(\sigma))$.

In the setting of local resilience, very recently, Montgomery [97] as well as Nenadov, Steger, and Trujić [98] independently obtained a hitting time version of theorem 1.1.1, that is, they proved that, in the K_n-process, a.a.s. the first graph which has minimum degree 2 is resiliently Hamiltonian (Nenadov, Steger and Trujić also obtained such a hitting time version for perfect matchings [98]).

For graphs other than the complete graph, Johansson [72] recently obtained a robustness version of the hitting time results for Hamiltonicity. In particular, for any n-vertex graph G with $\delta(G) \geq (1/2+\varepsilon)n$, he proved that a.a.s. $\tau_{\mathcal{HAM}}(\tilde{G}(\sigma)) = \tau_{\mathcal{S}2}(\tilde{G}(\sigma))$. This was later extended to a larger class of graphs G and to hitting times for $\mathcal{HM}2k$, for all $k \in \mathbb{N}$ independent of n, by Alon and Krivelevich [6].

1.1.4. Random subgraphs of the hypercube

When considering the setting of robustness of graph properties, one of the graphs G which has been studied more thoroughly is the hypercube $\mathcal{Q}^n$. The n-dimensional *hypercube* $\mathcal{Q}^n$ is the graph whose vertex set consists of all n-bit 01-strings, where two vertices are joined by an edge whenever their corresponding strings differ by a single bit. One of the main contributions in this thesis is precisely in this setting: we show that the hypercube is robustly Hamiltonian, with results in both the random binomial and the hitting time settings (see chapter 2).

We will denote by $\mathcal{Q}^n_p$ the random subgraph of the hypercube obtained by removing each edge of $\mathcal{Q}^n$ with probability $1-p$ independently of every other edge. In this context, Bollobás [19] determined the threshold and hitting time for perfect matchings by showing that a.a.s. $\tau_{\mathcal{PM}}(\tilde{\mathcal{Q}}^n(\sigma)) = \tau_{\mathcal{S}1}(\tilde{\mathcal{Q}}^n(\sigma))$.

The main goal of chapter 2 is to study the analogous problem for Hamiltonicity in random subgraphs of the hypercube. This is joint work with Condon, Girão, Kühn, and Osthus [34]. There is a folklore conjecture that the sharp threshold for Hamiltonicity in $\mathcal{Q}^n_p$ should be 1/2, the same as the threshold for having minimum degree at least 2. Even more, there is also a folklore conjecture that the hitting time for Hamiltonicity should be given by the necessary degree condition. These questions were explicitly asked by Bollobás [20] at several conferences in the 1980's. One of our main results is a hitting time result for Hamiltonicity (and, more generally, property $\mathcal{HM}k$) in $\mathcal{Q}^n$-processes, which resolves both conjectures.

Theorem 1.1.6 ([34]). *For all $k \in \mathbb{N}$, a.a.s. $\tau_{\mathcal{HM}k}(\tilde{\mathcal{Q}}^n(\sigma)) = \tau_{\delta k}(\tilde{\mathcal{Q}}^n(\sigma))$.*

1.1.5. Randomly perturbed graphs

A relatively recent area at the interface of extremal combinatorics and random graph theory is the study of *randomly perturbed graphs*. Generally speaking, we consider a deterministic dense n-vertex graph H and a binomial random graph $G_{n,p}$ on the same vertex set and ask whether H is close to satisfying some given property $\mathcal{P}$ in the sense that a.a.s. $H \cup G_{n,p} \in \mathcal{P}$ for some small p. This line of research was initiated by Bohman, Frieze, and Martin [15], who showed that, if H is an n-vertex graph with $\delta(H) \geq \alpha n$, for any constant $\alpha > 0$, then a.a.s. $H \cup G_{n,p}$ is Hamiltonian for all $p \geq C(\alpha)/n$. Very recently, Hahn-Klimroth, Maesaka, Mogge, Mohr, and Parczyk [65] extended this result to allow α to tend to 0 with n (that is, to allow graphs H which are not dense). A common phenomenon in this model is that, by considering the union with a graph H with linear degrees, the probability threshold of different properties is significantly lower than that in the classical $\mathcal{G}_{n,p}$ model.

In joint work with Condon, Girão, Kühn, and Osthus [34], we consider randomly perturbed graphs in the setting of subgraphs of the hypercube. To be precise, we take an arbitrary spanning subgraph H of the hypercube, with linear minimum degree, and a random subgraph $\mathcal{Q}^n_\varepsilon$, and consider $H \cup \mathcal{Q}^n_\varepsilon$. In this setting, we show the following result.

Theorem 1.1.7. *For all $\varepsilon, \alpha \in (0,1]$ and $k \in \mathbb{N}$, the following holds. Let H be a spanning subgraph of $\mathcal{Q}^n$ such that $\delta(H) \geq \alpha n$. Then, a.a.s. $H \cup \mathcal{Q}^n_\varepsilon$ contains k edge-disjoint Hamilton cycles.*

We can also allow H to have much smaller degrees, at the cost of requiring a larger probability to find the Hamilton cycles.

Theorem 1.1.8. *For every integer $k \geq 2$, there exists $\varepsilon > 0$ such that a.a.s., for every spanning subgraph H of $\mathcal{Q}^n$ with $\delta(H) \geq k$, the graph $H \cup \mathcal{Q}^n_{1/2-\varepsilon}$ contains a collection of $\lfloor k/2 \rfloor$ Hamilton cycles and $k - 2\lfloor k/2 \rfloor$ perfect matchings, all pairwise edge-disjoint.*

In fact, it is from (a stronger version of) theorem 1.1.8 that we will easily derive theorem 1.1.6 (see chapter 2 for full details about this).

1.2. Hamiltonicity in random hypergraphs

A hypergraph is a pair $H = (V, E)$ where V is a set of vertices and E is a set of subsets of V, called the set of edges. Whenever for all $e \in E$ we have that $|e| = r$ for some $r \in \mathbb{N}$, we say that H is an *r-uniform hypergraph* (or *r-graph*, for short). Hypergraphs are a generalisation of graphs in the sense that 2-graphs are simply graphs.

Hamiltonicity in (uniform) hypergraphs is also a very important topic, but its definition is not so clear-cut as it is for standard graphs. In particular, the notion of a cycle can be generalised in several different ways, and there is a notion of Hamiltonicity attached to each of these. For this thesis, we will concentrate on the notion of ℓ-overlapping cycles (or ℓ-cycles, for short). Given $\ell, r \in \mathbb{N}$ with $1 \leq \ell < r$, an ℓ-cycle in an r-graph H is a set of edges of H which can be labelled (cyclically) in such a way that any two consecutive edges share exactly ℓ vertices. Of particular interest are the cases where $\ell = 1$ or $\ell = r - 1$, when ℓ-cycles are referred to as *loose cycles* and *tight cycles*, respectively. An ℓ-cycle of H is Hamiltonian if it spans all vertices of H. Of course, there are some trivial divisibility conditions for an r-graph to contain such an ℓ-cycle (namely, we must have that $r - \ell \mid n$); we will assume this throughout without further mention. We will now give some of the main results about Hamiltonicity in random hypergraphs. For a perspective about some of the main extremal results in this setting, we recommend the survey of Kühn and Osthus [91].

1.2.1. Binomial random r-graphs

Given any $r \in \mathbb{N}$, a binomial random r-graph $G_{n,p}^{(r)}$ is obtained analogously as a binomial random graph: we consider a labelled set of n vertices and add each of the $\binom{n}{r}$ edges to the graph with probability p independently of all other edges. All the results we discuss here assume that r (and thus ℓ) is fixed, while n tends to infinity. The main results have to do with thresholds for Hamiltonicity in $\mathcal{G}_{n,p}^{(r)}$, for any $r \in \mathbb{N}$.

For loose cycles, Frieze [53], Dudek and Frieze [39], and Dudek, Frieze, Loh, and Speiss [41] proved that the threshold for Hamiltonicity in $\mathcal{G}_{n,p}^{(r)}$ is $\log n/n^{r-1}$. Dudek and Frieze [40] later proved that, for all $2 \leq \ell < r$, the threshold for the containment of a Hamilton ℓ-cycle in $\mathcal{G}_{n,p}^{(r)}$ is $1/n^{r-\ell}$.

More recently, the Hamiltonicity problem has been considered in the randomly perturbed hypergraph setting. In particular, Han and Zhao [66] showed that, given any r-graph H of minimum degree at least αn^{r-1}, the threshold for the containment of a Hamilton ℓ-cycle in $H \cup G_{n,p}^{(r)}$ is lower than $1/n^{r-\ell+\varepsilon}$, for some $\varepsilon = \varepsilon(\alpha)$.

1.2.2. Random regular r-graphs

Random regular r-graphs have also been considered. We denote by $G_{n,d}^{(r)}$ a d-regular r-graph chosen uniformly at random from the set of all d-regular r-graphs. Dudek, Frieze, Ruciński, and Šileikis [42, 43] proved that there is a joint distribution of $\mathcal{G}_{n,p}^{(r)}$ and $\mathcal{G}_{n,d}^{(r)}$ (for suitable values and ranges of d and p) such that $G_{n,p}^{(r)} \subseteq G_{n,d}^{(r)}$. This allowed them to prove results about Hamiltonicity in random regular r-graphs, for all $r \geq 3$. In particular, they proved that, if $d \gg \log n$, then a.a.s. $G_{n,d}^{(r)}$ contains a loose Hamilton cycle, and for all $2 \leq \ell < r$ and $d \gg n^{\ell-1}$, a.a.s. $G_{n,d}^{(r)}$ contains a Hamilton ℓ-cycle. They further conjectured that this should be best possible in the sense that, for all $2 \leq \ell < r$, if $d \ll n^{\ell-1}$, then a.a.s. $G_{n,d}^{(r)}$ does not contain a Hamilton ℓ-cycle. However, for loose Hamilton cycles they conjectured that the threshold should be constant. Altman, Greenhill, Isaev, and Ramadurai [7] investigated this problem and provided an exact threshold for the containment of a loose Hamilton cycle.

They further provided a partial result towards Hamiltonicity for ℓ-cycles when $2 \leq \ell < r$ by showing that, if $d \ll n^{1/2}$, then a.a.s. $G_{n,d}^{(r)}$ does not contain a Hamilton ℓ-cycle.

In joint work with Joos, Kühn, and Osthus [51], we developed an edge-switching technique which allowed us to estimate the probabilities that different edges appear in a random d-regular r-graph (thus extending results of Kim, Sudakov, and Vu [77] for graphs). Among several other applications (see chapter 3), this allowed us to verify the above conjecture. Combining our results with those of Dudek, Frieze, Ruciński, and Šileikis [43] and Altman, Greenhill, Isaev, and Ramadurai [7] effectively gives a threshold for Hamilton ℓ-cycles in random regular r-graphs.

1.3. A detour: testing subgraph freeness

Property testing is an area at the interface of combinatorics and computer science. Roughly speaking, one could say that property testing is the study of superfast (randomised) algorithms for approximate decision making. Indeed, the goal of a tester (that is, a property testing algorithm) for some property $\mathcal{P}$ is to establish whether an input graph satisfies $\mathcal{P}$ or is "far" from this property, and such algorithms are allowed to give a wrong answer with some probability, but must be right with probability more than 1/2. Depending on the notion of "distance" that we use, we can define several different models for property testing, with the dense graphs model, the bounded degree graphs model and the general graphs model being three of the main ones.

While the dense graphs model is fairly well understood and there are also many results in the bounded degree graphs model, our understanding of the general graphs model is far from satisfactory. Using the results about correlations in random regular r-graphs that we developed with Joos, Kühn, and Osthus [51], we managed to provide lower bounds for the query complexity of testing subgraph freeness in the general hypergraphs model, extending some results of Alon, Kaufman, Krivelevich, and Ron [3]. For full details about this, see chapter 3.

1.4. Organisation of the thesis

Throughout the past few years, we have produced four different papers [51, 35, 33, 34]. Due to space constraints for the thesis, we have been forced to reduce the scope to only two of these papers. We have chosen to present in detail the results about random subgraphs of the hypercube [34] and about subgraphs of random regular uniform hypergraphs [51].

In chapter 2, we reproduce most of the results we present in [34]. Again, due to space limitations, we have been forced to remove some parts of the paper. Thus, while this chapter is deeply rooted in the contents of [34], we have made some necessary changes for the sake of readability, which include the deletion and reordering of some sections of the original manuscript.

In chapter 3, we reproduce the results shown in [51]. This includes both the results about correlations in random regular r-graphs and their applications to subgraph counts as well as the application to subgraph testing which we discussed in section 1.3.

The author of the thesis has substantially contributed to all of the papers above. In particular, the development of mathematical ideas has been equally shared by all the authors, and the writing has been done exclusively by the author in [51], and in a shared effort with Condon in the remaining three papers [35, 33, 34]. In particular, for instance, in [34] the author was responsible for the writing of Section 6, while Condon was responsible for Section 7, and all other sections were written jointly by both of us.

CHAPTER 2

HAMILTONICITY IN RANDOM SUBGRAPHS OF THE HYPERCUBE

We study Hamiltonicity in random subgraphs of the hypercube $\mathcal{Q}^n$. Our first main theorem is an optimal hitting time result. Consider the random process which includes the edges of $\mathcal{Q}^n$ according to a uniformly chosen random ordering. Then, with high probability, as soon as the graph produced by this process has minimum degree $2k$, it contains k edge-disjoint Hamilton cycles, for any fixed $k \in \mathbb{N}$. Secondly, we obtain a perturbation result: if $H \subseteq \mathcal{Q}^n$ satisfies $\delta(H) \geq \alpha n$ with $\alpha > 0$ fixed and we consider a random binomial subgraph $\mathcal{Q}^n_p$ of $\mathcal{Q}^n$ with $p \in (0,1]$ fixed, then with high probability $H \cup \mathcal{Q}^n_p$ contains k edge-disjoint Hamilton cycles, for any fixed $k \in \mathbb{N}$. In particular, both results resolve a long standing conjecture, posed e.g. by Bollobás, that the threshold probability for Hamiltonicity in the random binomial subgraph of the hypercube equals $1/2$. Our techniques also show that, with high probability, for all fixed $p \in (0,1]$ the graph $\mathcal{Q}^n_p$ contains an almost spanning cycle. Our methods involve branching processes, the Rödl nibble, and absorption.

2.1. Introduction

The n-dimensional *hypercube* $\mathcal{Q}^n$ is the graph whose vertex set consists of all n-bit 01-strings, where two vertices are joined by an edge whenever their corresponding strings differ by a single bit. The hypercube and its subgraphs have attracted much attention in graph theory and computer science, e.g. as a sparse network model with strong connectivity properties. It is well known that hypercubes contain spanning paths (also called *Gray codes* or *Hamilton paths*) and, for all $n \geq 2$, they contain spanning cycles (also referred to as *cyclic Gray codes* or *Hamilton cycles*). Classical applications of Gray codes in computer science are described in the surveys of Savage [109] and Knuth [79]. Applications of hypercubes to parallel computing are discussed in the monograph of Leighton [94].

2.1.1. Spanning subgraphs in hypercubes

The systematic study of spanning paths, trees and cycles in hypercubes was initiated in the 1970's. There is by now an extensive literature about subtrees of the hypercube; see, for instance, results of Bhatt, Chung, Leighton, and Rosenberg [14] about embedding subdivided trees (instigated by processor allocation in distributed computing systems).

As a generalization of Hamilton paths, Caha and Koubek [29] considered the problem of finding a collection of spanning vertex-disjoint paths, given a prescribed set of endpoints. After several improvements [31, 64], this problem was recently resolved by Dvořák, Gregor, and Koubek [45].

The applications of hypercubes as networks in computer science inspired questions about the reliability of its properties. This led to considering "faulty" hypercubes in which some edges or vertices are missing. For instance, Chan and Lee [30] showed that, if $\mathcal{Q}^n$ has at most $2n-5$ faulty edges and every vertex has (non-faulty) degree at least 2, then there is a Hamilton cycle in $\mathcal{Q}^n$ which avoids all faulty edges (and this condition is best possible). They also showed that the general problem of determining the Hamiltonicity of $\mathcal{Q}^n$ with a larger number of faulty edges is NP-complete. More generally, Dvořák and Gregor [44] studied the existence of spanning collections of vertex-disjoint paths with prescribed endpoints in faulty hypercubes. (We will apply these results in our proofs, see section 2.7.4 for details.) These can be seen as extremal results about the *robustness* of the hypercube with respect to containing spanning collections of paths and cycles.

2.1.2. Hamilton cycles in binomial random graphs

One of the most studied random graph models is the binomial random graph $\mathcal{G}_{n,p}$. Here we have a (labelled) set of n vertices and we include each edge with probability p independently of all other edges.

Given some monotone increasing graph property $\mathcal{P}$, a function $p^* = p^*(n)$ is said to be a (coarse) *threshold* for $\mathcal{P}$ if $\mathbb{P}[G_{n,p} \in \mathcal{P}] \to 1$ whenever $p/p^* \to \infty$ and $\mathbb{P}[G_{n,p} \in \mathcal{P}] \to 0$ whenever $p/p^* \to 0$. One can define the stronger notion of a sharp threshold similarly: $p^* = p^*(n)$ is said to be a *sharp threshold* for $\mathcal{P}$ if, for all $\varepsilon > 0$, we have that $\mathbb{P}[G_{n,p} \in \mathcal{P}] \to 1$ whenever $p \geq (1+\varepsilon)p^*$ and $\mathbb{P}[G_{n,p} \in \mathcal{P}] \to 0$ whenever $p \leq (1-\varepsilon)p^*$. The problem of finding the threshold for the containment of a Hamilton cycle was solved independently by Pósa [100] and Koršunov [82]. Furthermore, Koršunov [82] determined the sharp threshold for Hamiltonicity to be $p^* = \log n/n$. These results were later made even more precise by Komlós and Szemerédi [81]. It is worth noting that $p^* = \log n/n$ is also the sharp threshold for the property of having minimum degree at least 2. In this sense, the results about Hamilton cycles in $G_{n,p}$ can be interpreted as saying that the natural obstruction of having sufficiently high minimum degree is also an "almost sufficient" condition.

A property that generalises Hamiltonicity is that of containing k edge-disjoint Hamilton cycles, for some $k \in \mathbb{N}$. We will present more results in this direction in section 2.1.4; for now, let us simply note that the sharp threshold for the containment of k edge-disjoint Hamilton cycles in $G_{n,p}$, for some $k \in \mathbb{N}$ independent of n, is $p^* = \log n/n$, i.e. the same as the threshold for Hamiltonicity.

The study of *robustness* of graph properties has also attracted much attention recently. For instance, given a graph G which is known to satisfy some property $\mathcal{P}$, consider a random subgraph G_p obtained by deleting each edge of G with probability $1-p$, independently of all other edges. The problem then is to determine the range of p for which G_p satisfies $\mathcal{P}$ with high probability. In this setting, a result of Krivelevich, Lee, and Sudakov [87] asserts that, for any n-vertex graph G with minimum degree at least $n/2$, the graph G_p is asymptotically almost surely Hamiltonian whenever $p \gg \log n/n$. This can be viewed as a robust version of Dirac's theorem on Hamilton cycles.

2.1.3. Hamilton cycles in binomial random subgraphs of the hypercube

Throughout this chapter, we will consider random subgraphs of the hypercube and show that the hypercube is robustly Hamiltonian in the above sense. We will denote by $\mathcal{Q}^n_p$ the random subgraph of the hypercube obtained by removing each edge of $\mathcal{Q}^n$ with probability $1-p$ independently of every other edge.

The random graph $\mathcal{Q}^n_p$ was first studied by Burtin [28], who proved that the sharp threshold for connectivity is 1/2. This result was later made more precise by Erdős and Spencer [50] and Bollobás [17]. As a related problem, Dyer, Frieze, and Foulds [46] determined the sharp threshold for connectivity in subgraphs of $\mathcal{Q}^n$ obtained by removing both vertices and edges uniformly at random. Later, Bollobás [19] proved that 1/2 is also the sharp threshold for the containment of a perfect matching in $\mathcal{Q}^n_p$. As with the $\mathcal{G}_{n,p}$ model, this also coincides with the threshold for having minimum degree at least 1.

The main goal of this chapter is to study the analogous problem for Hamiltonicity in random subgraphs of the hypercube. There is a folklore conjecture that the sharp threshold for Hamiltonicity in $\mathcal{Q}^n_p$ should be 1/2, i.e. the same as the threshold for having minimum degree at least 2. This question was explicitly asked by Bollobás [20] at several conferences in the 1980's, in the ICM surveys of Frieze [54] and Kühn and Osthus [91], as well as the recent survey of Frieze [55]. A special case of our first result resolves this problem.

Theorem 2.1.1. *For any $k \in \mathbb{N}$, the sharp threshold for the property of containing k edge-disjoint Hamilton cycles in $\mathcal{Q}^n_p$ is $p^* = 1/2$.*

For $k = 1$, this can be seen as a probabilistic version of the result on faulty hypercubes [30], and also as a statement about the robustness of Hamiltonicity in the hypercube.

While, for $p < 1/2$, with high probability $\mathcal{Q}^n_p$ will not contain a Hamilton cycle, it turns out that the reason for this is mostly due to local obstructions (e.g., vertices with degree zero or one). More precisely, we prove that, for any constant $p \in (0, 1/2)$, a.a.s. the random graph $\mathcal{Q}^n_p$ contains an almost spanning cycle.

Theorem 2.1.2. *For any $\delta, p \in (0, 1]$, a.a.s. the graph $\mathcal{Q}^n_p$ contains a cycle of length at least $(1-\delta)2^n$.*

We believe that the probability bound is far from optimal, in the sense that random subgraphs of the hypercube where edges are picked with vanishing probability should also satisfy this property.

Conjecture 2.1.3. *Suppose that $p = p(n)$ satisfies that $pn \to \infty$. Then, a.a.s. $\mathcal{Q}^n_p$ contains a cycle of length $(1 - o(1))2^n$.*

Similarly, it would be interesting to determine which (long) paths and (almost spanning) trees can be found in $\mathcal{Q}^n_p$. Moreover, our methods might also be useful to embed other large subgraphs, such as F-factors.

Conjecture 2.1.4. *Suppose $\varepsilon > 0$ and an integer $\ell \geq 2$ are fixed and $p \geq 1/2 + \varepsilon$. Then, a.a.s. $\mathcal{Q}^n_p$ contains a C_{2^ℓ}-factor, that is, a set of vertex-disjoint cycles of length 2^ℓ which together contain all vertices of $\mathcal{Q}^n$.*

2.1.4. Hitting time results

Remarkably, the above intuition that having the necessary minimum degree is an "almost sufficient" condition for the containment of edge-disjoint perfect matchings and Hamilton cycles can be strengthened greatly via so-called hitting time results. These are expressed in terms of random graph processes. The general setting is as follows. Let G be an n-vertex graph with $m = m(n)$ edges, and consider an arbitrary labelling $E(G) = \{e_1, \dots, e_m\}$. The *$G$-process* is defined as a random sequence of nested graphs $\tilde{G}(\sigma) = (G_t(\sigma))_{t=0}^m$, where σ is a permutation of $[m]$ chosen uniformly at random and, for each $i \in [m]_0$, we set $G_i(\sigma) = (V(G), E_i)$, where $E_i := \{e_{\sigma(j)} : j \in [i]\}$. Given any monotone increasing graph property $\mathcal{P}$ such that $G \in \mathcal{P}$, the *hitting time* for $\mathcal{P}$ in the above G-process is the random variable $\tau_{\mathcal{P}}(\tilde{G}(\sigma)) := \min\{t \in [m]_0 : G_t(\sigma) \in \mathcal{P}\}$.

Let us denote the properties of containing a perfect matching by $\mathcal{PM}$, Hamiltonicity by $\mathcal{HAM}$, and connectivity by $\mathcal{CON}$, respectively. For any $k \in \mathbb{N}$, let δk denote the property of having minimum degree at least k, and let $\mathcal{HM}k$ denote the property of containing $\lfloor k/2 \rfloor$ edge-disjoint Hamilton cycles and, if k is odd, one matching of size $\lfloor n/2 \rfloor$ which is edge-disjoint from these Hamilton cycles. With this notion of hitting times, many of the results about thresholds presented in sections 2.1.2 and 2.1.3 can be strengthened significantly. For instance, Bollobás and Thomason [23] showed that a.a.s. $\tau_{\mathcal{CON}}(\tilde{K_n}(\sigma)) = \tau_{\delta 1}(\tilde{K_n}(\sigma))$ and, if n is even, then a.a.s. $\tau_{\mathcal{PM}}(\tilde{K_n}(\sigma)) = \tau_{\delta 1}(\tilde{K_n}(\sigma))$. Ajtai, Komlós, and Szemerédi [1] and Bollobás [18] independently proved that a.a.s. $\tau_{\mathcal{HAM}}(\tilde{K_n}(\sigma)) = \tau_{\delta 2}(\tilde{K_n}(\sigma))$. This was later generalised by Bollobás and Frieze [21], who proved that, given any $k \in \mathbb{N}$, for n even a.a.s. $\tau_{\mathcal{HM}k}(\tilde{K_n}(\sigma)) = \tau_{\delta k}(\tilde{K_n}(\sigma))$.

A hitting time result for the property of having k edge-disjoint Hamilton cycles when k is allowed to grow with n is still not known, even in K_n-processes. As a slightly weaker notion, consider property $\mathcal{H}$, where we say that a graph G satisfies property $\mathcal{H}$ if it contains $\lfloor \delta(G)/2 \rfloor$ edge-disjoint Hamilton cycles, together with an additional edge-disjoint matching of size $\lfloor n/2 \rfloor$ if $\delta(G)$ is odd. Knox, Kühn, and Osthus [78], Krivelevich and Samotij [88] as well as Kühn and Osthus [92] proved results for different ranges of p which, together, show that $G_{n,p}$ a.a.s. satisfies property $\mathcal{H}$.

For graphs other than the complete graph, Johansson [72] recently obtained a robustness version of the hitting time results for Hamiltonicity. In particular, for any n-vertex graph G with $\delta(G) \geq (1/2+\varepsilon)n$, he proved that a.a.s. $\tau_{\mathcal{HAM}}(\tilde{G}(\sigma)) = \tau_{\delta 2}(\tilde{G}(\sigma))$. This was later extended to a larger class of graphs G and to hitting times for $\mathcal{HM}2k$, for all $k \in \mathbb{N}$ independent of n, by Alon and Krivelevich [6].

In the setting of random subgraphs of the hypercube, Bollobás [19] determined the hitting time for perfect matchings by showing that a.a.s. $\tau_{\mathcal{PM}}(\tilde{\mathcal{Q}^n}(\sigma)) = \tau_{\mathcal{CON}}(\tilde{\mathcal{Q}^n}(\sigma)) = \tau_{\delta 1}(\tilde{\mathcal{Q}^n}(\sigma))$. One of our main results (which implies theorem 2.1.1) is a hitting time result for Hamiltonicity (and, more generally, property $\mathcal{HM}k$) in $\mathcal{Q}^n$-processes. Again, this question was raised by Bollobás [20] at several conferences.

Theorem 2.1.5. *For all $k \in \mathbb{N}$, a.a.s. $\tau_{\mathcal{HM}k}(\tilde{\mathcal{Q}}^n(\sigma)) = \tau_{\delta k}(\tilde{\mathcal{Q}}^n(\sigma))$, that is, the hitting time for the containment of a collection of $\lfloor k/2 \rfloor$ Hamilton cycles and $k - 2\lfloor k/2 \rfloor$ perfect matchings, all pairwise edge-disjoint, in $\mathcal{Q}^n$-processes is a.a.s. equal to the hitting time for the property of having minimum degree at least k.*

We also wonder whether this is true if k is allowed to grow with n, and propose the following conjecture which, if true, would be an approximate version of the results of [88, 92, 78] in the hypercube.

Conjecture 2.1.6. *For all $p \in (1/2, 1]$ and $\eta > 0$, a.a.s. $\mathcal{Q}^n_p$ contains $(1/2 - \eta)\delta(\mathcal{Q}^n_p)$ edge-disjoint Hamilton cycles.*

2.1.5. Randomly perturbed graphs

A relatively recent area at the interface of extremal combinatorics and random graph theory is the study of *randomly perturbed graphs*. Generally speaking, the idea is to consider a deterministic dense n-vertex graph H (usually satisfying some minimum degree condition) and a random graph $G_{n,p}$ on the same vertex set as H. The question is whether H is close to satisfying some given property $\mathcal{P}$ in the sense that a.a.s. $H \cup G_{n,p} \in \mathcal{P}$ for some small p. This line of research was sparked off by Bohman, Frieze, and Martin [15], who showed that, if H is an n-vertex graph with $\delta(H) \geq \alpha n$, for any constant $\alpha > 0$, then a.a.s. $H \cup G_{n,p}$ is Hamiltonian for all $p \geq C(\alpha)/n$. Other properties that have been studied in this context are e.g. the existence of powers of Hamilton cycles and general bounded degree spanning graphs [27], F-factors [9] or spanning bounded degree trees [86, 26]. One common phenomenon in this model is that, by considering the union with a dense graph H (i.e. a graph H with linear degrees), the probability threshold of different properties is significantly lower than that in the classical $\mathcal{G}_{n,p}$ model. The results for Hamiltonicity [15] were very recently generalised by Hahn-Klimroth, Maesaka, Mogge, Mohr, and Parczyk [65] to allow α to tend to 0 with n (that is, to allow graphs H which are not dense).

We consider randomly perturbed graphs in the setting of subgraphs of the hypercube. To be precise, we take an arbitrary spanning subgraph H of the hypercube, with linear minimum degree, and a random subgraph $\mathcal{Q}^n_\varepsilon$, and consider $H \cup \mathcal{Q}^n_\varepsilon$. (Note here that $\mathcal{Q}^n_\varepsilon$ is a "dense" subgraph of $\mathcal{Q}^n$, but for $\varepsilon < 1/2$ it will contain both isolated vertices and vertices of very low degrees.) In this setting, we show the following result.

Theorem 2.1.7. *For all $\varepsilon, \alpha \in (0, 1]$ and $k \in \mathbb{N}$, the following holds. Let H be a spanning subgraph of $\mathcal{Q}^n$ such that $\delta(H) \geq \alpha n$. Then, a.a.s. $H \cup \mathcal{Q}^n_\varepsilon$ contains k edge-disjoint Hamilton cycles.*

We can also allow H to have much smaller degrees, at the cost of requiring a larger probability to find the Hamilton cycles.

Theorem 2.1.8. *For every integer $k \geq 2$, there exists $\varepsilon > 0$ such that a.a.s., for every spanning subgraph H of $\mathcal{Q}^n$ with $\delta(H) \geq k$, the graph $H \cup \mathcal{Q}^n_{1/2-\varepsilon}$ contains a collection of $\lfloor k/2 \rfloor$ Hamilton cycles and $k - 2\lfloor k/2 \rfloor$ perfect matchings, all pairwise edge-disjoint.*

Note that theorem 2.1.8 can be viewed as a "universality" result for H, meaning that it holds for all choices of H simultaneously. It would be interesting to know whether such a result can also be obtained for the lower edge probability assumed in theorem 2.1.7, i.e., is it the case that, for all $\varepsilon, \alpha \in (0, 1]$, a.a.s. $G \sim \mathcal{Q}^n_\varepsilon$ has the property that, for every spanning $H \subseteq \mathcal{Q}^n$ with $\delta(H) \geq \alpha n$, $G \cup H$ is Hamiltonian?

While we do not show this here, theorem 2.1.1 follows straightforwardly from theorem 2.1.7, and it follows trivially from theorem 2.1.5. In turn, theorem 2.1.5 follows from theorem 2.1.8. On the other hand, theorems 2.1.2, 2.1.7 and 2.1.8, while being proved with similar ideas, are incomparable. It is worth noting that we do not explicitly prove theorems 2.1.2 and 2.1.7 in this exposition. These can be proved with similar methods as used for the proof of theorem 2.1.8, albeit with fewer technicalities; for full proofs, please see the original manuscript [34].

2.1.6. Percolation on the hypercube

To build Hamilton cycles in random subgraphs of the hypercube, we will consider a random process which can be viewed as a branching process or percolation process on the hypercube.

With high probability, for constant $p > 0$, this process results in a bounded degree tree in $\mathcal{Q}^n_p$ which covers most of the neighbourhood of every vertex in $\mathcal{Q}^n$, and thus spans almost all vertices of $\mathcal{Q}^n$. The version stated below is a special case of theorem 2.5.11.

Theorem 2.1.9. *For any fixed* $\varepsilon, p \in (0,1]$, *there exists* $D = D(\varepsilon)$ *such that a.a.s.* $\mathcal{Q}^n_p$ *contains a tree* T *with* $\Delta(T) \leq D$ *and such that* $|V(T) \cap N_{\mathcal{Q}^n}(x)| \geq (1-\varepsilon)n$ *for every* $x \in V(\mathcal{Q}^n)$.

Further results concerning the local geometry of the giant component in $\mathcal{Q}^n_p$ for constant $p \in (0, 1/2)$ were proved recently by McDiarmid, Scott, and Withers [95].

The random process we consider in the proof of theorem 2.1.9 can be viewed as a branching random walk (with a bounded number of branchings at each step). Simpler versions of such processes (with infinite branchings allowed) have been studied by Fill and Pemantle [52] and Kohayakawa, Kreuter, and Osthus [80], and we will base our analysis on these. Motivated by our approach, we raise the following question, which seems interesting in its own right.

Question 2.1.10. *Does a non-returning random walk on* $\mathcal{Q}^n$ *a.a.s. visit almost all vertices of* $\mathcal{Q}^n$*?*

More generally, there are many results and applications concerning random walks on the hypercube (but allowing for returns). For example, motivated by a processor allocation problem, Bhatt and Cai [13] studied a walk algorithm to embed large (subdivided) trees into the hypercube. Moreover, the analysis of (branching) random walks is a critical ingredient in the study of percolation thresholds for the existence of a giant component in $\mathcal{Q}^n_p$. These have been investigated, e.g., by Bollobás, Kohayakawa, and Łuczak [22], Borgs, Chayes, van der Hofstad, Slade, and Spencer [25], and van der Hofstad and Nachmias [68].

2.1.7. Organisation of the chapter

In section 2.2 we provide an overview of our ideas and proof methods. In section 2.3 we introduce the notation we will use throughout this chapter. In section 2.4 we state the different probabilistic tools, as well as some other well-known results, that we will call on, and in section 2.5 we collect various results on matchings and random subgraphs of the hypercube, including theorem 2.5.11, our main near-spanning tree result. In section 2.6 we prove theorem 2.6.6, our main cube tiling result (see section 2.2 for more details on tilings and near-spanning trees). Then, in section 2.7 we present some structures that will be crucial for our proof, including the different absorbing structures that we will consider, and prove our connecting lemmas. Finally, in section 2.8 we prove both theorem 2.1.8 and our hitting time result (theorem 2.1.5).

2.2. Outline of the main proofs

2.2.1. Overall outline

We now sketch the key ideas for the proof of theorem 2.1.8. We will first prove the case $k = 1$, and later use this to deduce the case when $k > 1$. Most of these ideas are common with those required for the proof of theorem 2.1.7 (omitted in this thesis), so for simplicity we briefly adopt the language necessary for this result. Recall we are given $H \subseteq \mathcal{Q}^n$ with $\delta(H) \geq \alpha n$, and $G \sim \mathcal{Q}^n_\varepsilon$, with $\alpha, \varepsilon \in (0, 1]$. Our aim is to show that a.a.s. $H \cup G$ is Hamiltonian.

Our approach for finding a Hamilton cycle is to first obtain a spanning tree. By passing along all the edges of a spanning tree T (with a vertex ordering prescribed by a depth first search), one can create a closed spanning walk W which visits every edge of T twice. The idea is then to modify such a walk into a Hamilton cycle. (This approach is inspired by the approximation algorithm for the travelling salesman problem which returns a tour of at most twice the optimal length.) More precisely, our approach will be to obtain a near-spanning tree of $\mathcal{Q}^{n-s}$, for some suitable constant s, and to blow up vertices of this tree into s-dimensional cubes (see figure 2.1). These cubes can then be used to move along the tree without revisiting vertices, which will result in a near-Hamilton cycle $\mathfrak{H}$. All remaining vertices which are not included in $\mathfrak{H}$ will be absorbed into $\mathfrak{H}$ via absorbing structures that we carefully put in place beforehand.

In sections 2.2.2 to 2.2.4 we outline in more detail how we find a long cycle in G (theorem 2.1.2). Note that in theorem 2.1.2 we have $G \sim \mathcal{Q}^n_\varepsilon$, so a.a.s. G will have isolated vertices which prevent any Hamilton cycle occurring as a subgraph. In section 2.2.5 we outline how we build on this approach to obtain the case $k = 1$ of theorem 2.1.7. Finally, in section 2.2.6 we sketch how we obtain theorem 2.1.8 (and thus theorem 2.1.5).

2.2.2. Building block I: trees via branching processes.

We view each vertex in $\mathcal{Q}^n$ as an n-dimensional 01-coordinate vector. By fixing the first s coordinates, we fix one of 2^s *layers* $L_1, \dots, L_{2^s}$ of the hypercube, where $s \in \mathbb{N}$ will be constant (note that this is different from the usual notion of *layer* in hypercubes, which corresponds to those vertices having the same number of coordinates equal to 1). Thus, $L \cong \mathcal{Q}^{n-s}$ for each layer L. By considering a Hamilton cycle in $\mathcal{Q}^s$, we may assume that consecutive layers differ only by a single coordinate on the unique elements of $\mathcal{Q}^s$ which define them. Let $G \sim \mathcal{Q}^n_\varepsilon$. For each layer L, we let $L(G) \coloneqq G[V(L)]$ and (by momentarily viewing these layers as different subgraphs on the vertex set of $\mathcal{Q}^{n-s}$) define the *intersection graph* $I(G) \coloneqq \bigcap_{i=1}^{2^s} L_i(G)$. Hence, $I(G) \sim \mathcal{Q}^{n-s}_{\varepsilon^{2^s}}$. We view $I(G)$ as a subgraph of $\mathcal{Q}^{n-s}$. We first show that $I(G)$ contains a near-spanning tree T (theorem 2.5.11). Thus, a copy of T is present in each of $L_1(G), \dots, L_{2^s}(G)$ simultaneously.

Since the walk W mentioned in section 2.2.1 passes through each vertex x of T a total of $d_T(x)$ times, it will be important later for T to have bounded degree. In order to guarantee this, we run bounded degree branching processes from several far apart "corners" of the hypercube. Roughly speaking, T will be formed by taking a union of these processes and removing cycles. Crucially, the model we introduce for these processes has a joint distribution with $\mathcal{Q}^{n-s}_{\varepsilon^{2^s}}$, so that T will in fact appear as a subgraph of $I(G)$. In applying theorem 2.5.11, we obtain a bounded degree tree $T \subseteq I(G)$ which contains almost all of the neighbours of every vertex of $I(G)$. We also obtain a "small" *reservoir* set $R \subseteq V(I(G))$, which T avoids and which will play a key role later in the absorption of vertices which do not belong to our initial long cycle. At this point, both T and R are now present in every layer of the hypercube simultaneously.

2.2.3. Building block II: cube tilings via the nibble.

Let $\ell < s$ and $0 < \delta \ll 1$ be fixed. In order to gain more local flexibility when traversing the near-spanning tree T, we augment T by locally adding a near-spanning ℓ-cube factor of $I(G)$. One can use classical results on matchings in almost regular uniform hypergraphs of small codegree to show that $I(G)$ contains such a collection of $\mathcal{Q}^\ell$ spanning almost all vertices of $I(G)$. However, we require the following stronger properties, namely that there exists a collection $\mathcal{C}$ of vertex disjoint copies of $\mathcal{Q}^\ell$ in $I(G)$ so that, for each $x \in V(I(G))$,

(i) $\mathcal{C}$ covers almost all vertices in $N_{\mathcal{Q}^n}(x)$;

(ii) the directions spanned by the cubes intersecting $N_{\mathcal{Q}^n}(x)$ do not correlate too strongly with any given set of directions.

The precise statement is given in theorem 2.6.6. Neither (i) nor (ii) follow from existing results on hypergraph matchings and the proofs strongly rely on geometric properties intrinsic to the hypercube.

To prove theorem 2.6.6, we build on the so-called Rödl nibble. More precisely, we consider the hypergraph $\mathcal{H}$, with $V(\mathcal{H}) = V(\mathcal{Q}^{n-s})$, where the edge set is given by the copies of $\mathcal{Q}^\ell$ in $I(G)$. We run a random iterative process where at each stage we add a "small" number of edges from $\mathcal{H}$ to $\mathcal{C}$, before removing all those remaining edges of $\mathcal{H}$ which "clash" with our selection. A careful analysis and an application of the Lovász local lemma yield the existence of an instance of this process which terminates in the near-spanning ℓ-cube factor with the properties required for theorem 2.6.6.

2.2.4. Constructing a long cycle.

Roughly speaking, we will use T as a backbone to provide "global" connectivity, and will use the near-spanning ℓ-cube factor $\mathcal{C}$ and the layer structure to gain high "local" connectivity and flexibility. We show a representation of this structure in figure 2.1. Let $T \cup \bigcup_{C\in\mathcal{C}} C =: \Gamma' \subseteq I(G)$ and let $\Gamma \subseteq \Gamma'$ be formed by removing all leaves and isolated cubes in Γ'. It follows by our tree and nibble results that almost all vertices of $I(G)$ are contained in Γ. Note that,

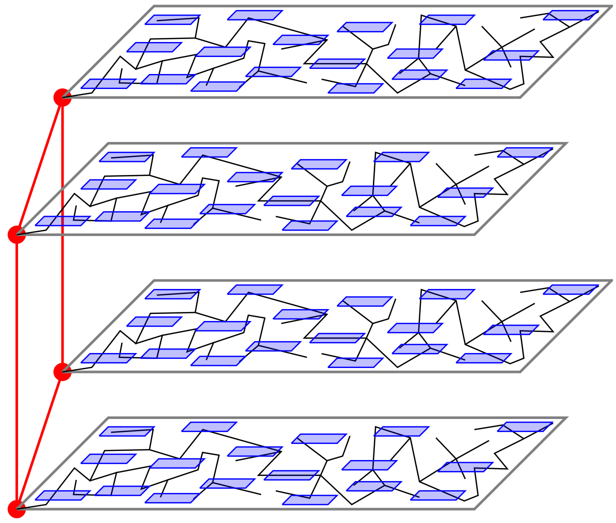

Figure 2.1. A representation of the main structure used for the proofs. We think of $\mathcal{Q}^n$ as a 'product' of two smaller cubes. Each 'horizontal' cube represents a layer (that is, a copy of $\mathcal{Q}^{n-s}$), and the red 'vertical' cube represents $\mathcal{Q}^s$. All 'horizontal' cubes contain a copy of the same tree T and the same cube tiling $\mathcal{C}$ (which are consistently distributed with respect to the 'vertical' cube; this gives rise to 'cube molecules'). When finding a long cycle, cube molecules are highly connected and can be covered by few paths, and the tree is used to join cube molecules to one another.

for each $v \in V(\mathcal{Q}^{n-s}) = V(I(G))$, there is a unique vertex in each of the 2^s layers which corresponds to v. We refer to these 2^s vertices as *clones* of v and to the collection of these 2^s clones as a *vertex molecule*. Similarly, each ℓ-cube $C \in \mathcal{C}$ contained in Γ gives rise to a *cube molecule* (i.e., a cube molecule is a collection of 2^s cubes, each of them contained in a layer and comprised of clones of a given cube in $I(G)$). We construct a cycle in G which covers all of the cube molecules (and, therefore, almost all vertices in $\mathcal{Q}^n$).

Let Γ^* be the graph obtained from Γ by contracting each ℓ-cube $C \subseteq \Gamma$ into a single vertex. We refer to such vertices in Γ^* as *atomic vertices*, and to all other vertices as *inner tree vertices*. We run a depth-first search on Γ^* to give an order to the vertices. Next, we construct a *skeleton* which will be the backbone for our long cycle. The skeleton is an ordered sequence of vertices in $\mathcal{Q}^n$ which contains the vertices via which our cycle will enter and exit each molecule. That is, given an *exit vertex* v for some molecule in the skeleton, the vertex u which succeeds v in the skeleton will be an *entry vertex* for another molecule, and such that $uv \in E(G)$. Here, a vertex in the skeleton belonging to an inner tree vertex molecule is referred to as both an entry and exit vertex. (Actually, we will first construct an "external skeleton", which encodes this information. The skeleton then also prescribes some edges within molecules which go between different layers.) We use the ordering of the vertices of Γ^* to construct the skeleton in a recursive way starting from the lowest ordered vertex. It is crucial that our tree T has bounded degree (much smaller than 2^s), so that no molecule is overused in the skeleton.

Once the skeleton is constructed, we apply our "connecting lemmas" (lemmas 2.7.11 and 2.7.12). These connecting lemmas, applied to a cube molecule with a bounded number of pairs of entry and exit vertices as input (given by the skeleton), provide us with a sequence of vertex-disjoint paths which cover this molecule, where each path has start and end vertices consisting of an input pair. The union of all of these paths combined with all edges in G between the successive exit and entry vertices of the skeleton will then form a cycle $\mathfrak{H} \subseteq G$ which covers all vertices lying in the cube molecules (thus proving theorem 2.1.2).

2.2.5. Constructing a Hamilton cycle.

In order to construct a Hamilton cycle in $H \cup G$, we will absorb the vertices of $V(\mathcal{Q}^n) \setminus V(\mathfrak{H})$ into $\mathfrak{H}$. We achieve this via absorbing structures that we identify for each vertex (see definition 2.7.1). To construct these absorbing structures, we will need to use some edges of H. Roughly speaking, to each vertex v we associate a left ℓ-cube $C_v^l \subseteq \mathcal{Q}^n$ and a right ℓ-cube $C_v^r \subseteq \mathcal{Q}^n$, where C_v^l, C_v^r are both clones of some ℓ-cubes $C^l, C^r \in \mathcal{C}$ contained in Γ. We choose these cubes so that v will have a neighbour $u \in V(C_v^l)$ and a neighbour $u' \in V(C_v^r)$, to which we refer as *tips* of the absorbing structure. Furthermore, u will have a neighbour $w \in V(C_v^r)$, which is also a neighbour of u'. Our near-Hamilton cycle $\mathfrak{H}$ will satisfy the following properties:

(a) $\mathfrak{H}$ covers all vertices in $C_v^l \cup C_v^r$ except for u, and

(b) $wu' \in E(\mathfrak{H})$.

These additional properties will be guaranteed by our connecting lemmas discussed in section 2.2.4. We can then alter $\mathfrak{H}$ to include the segment $wuvu'$ instead of the edge wu', thus absorbing the vertices u and v into $\mathfrak{H}$. The following types of vertices will require absorption.

(i) Every vertex that is not covered by a clone of either some inner tree vertex or of some cube $C \in \mathcal{C}$ which is contained in Γ.

(ii) The cycle $\mathfrak{H}$ does not cover all the clones of inner tree vertices and, thus, the uncovered vertices of this type will also have to be absorbed.

However, we will not know precisely which of the vertices described in (ii) will be covered by $\mathfrak{H}$ and which of these vertices will need to be absorbed until after we have constructed the (external) skeleton. Moreover, many potential absorbing structures are later ruled out as candidates (for example, if they themselves contain vertices that will need to be absorbed). Therefore, it is important that we identify a "robust" collection of many potential absorbing structures for every vertex in $\mathcal{Q}^n$ at a preliminary stage of the proof. The precise absorbing structure eventually assigned to each vertex will be chosen via an application of our rainbow matching lemma (lemma 2.5.5) at a late stage in the proof.

We will now highlight the purpose of the reservoir R. Suppose $v \in V(\mathcal{Q}^n)$ is a vertex which needs to be absorbed via an absorbing structure with left ℓ-cube C_v^l and left tip $u \in V(C_v^l)$. Recall that both u and C_v^l are clones of some $u^* \in V(\Gamma)$ and $C^l \in \mathcal{C}$, where $u^* \in V(C^l)$. If u^* has a neighbour w^* in $T - V(C^l)$, then it is possible that the skeleton will assign an edge from u to w for the cycle $\mathfrak{H}$ (where w is the clone of w^* in the same layer as u). Given that u is now incident to a vertex outside of C_v^l, we can no longer use the absorbing structure with u as a (left) tip (otherwise, we might disconnect T). To avoid this problem, we show that most vertices have many potential absorbing structures whose tips lie in the reservoir R (which T avoids). Here we make use of vertex degrees of H. A small number of *scant vertices* will not have high enough degree into R. For these vertices we fix an absorbing structure whose tips do not lie in R, and then alter T slightly so that these tips are deleted from T and reassigned to R. The fact that scant vertices are few and well spread out from each other will be crucial in being able to achieve this (see lemma 2.5.13).

Let us now discuss two problems arising in the construction of the skeleton. Firstly, let $\mathcal{M}_C \subseteq \mathcal{Q}^n$ with $C \in \mathcal{C}$ be a cube molecule which is to be covered by $\mathfrak{H}$. Furthermore, suppose one of the clones C_v^l of C belongs to an absorbing structure for some vertex v. Let u be the tip of C_v^l and suppose that u has even parity. We would like to apply the connecting lemmas to cover $\mathcal{M}_C - \{u\}$ by paths which avoid u. But this would now involve covering one fewer vertex of even parity than of odd parity. This, in turn, has the effect of making the construction of the skeleton considerably more complicated (this construction is simplest when successive entry and exit vertices have opposite parities). To avoid this, we assign absorbing structures in pairs, so that, for each $C \in \mathcal{C}$, either two or no clones of C will be used in absorbing structures. In the case where two clones are used, we enforce that the tips of these clones will have opposite parities, and therefore each molecule $\mathcal{M}_C$ will have the same number of even and odd parity vertices to be covered by $\mathfrak{H}$. We use our robust matching lemma (see lemma 2.5.2) to pair up the clones of absorbing structures in this way. To connect up different layers of a cube molecule, we will of course need to have suitable edges between these. Molecules which do not satisfy this requirement are called "bondless" and are removed from Γ before the absorption process (so that their vertices are absorbed).

Secondly, another issue related to vertex parities arises from inner tree vertex molecules. Depending on the degree of an inner tree vertex $v \in V(T)$, the skeleton could contain an odd number of vertices from the molecule $\mathcal{M}_v$ consisting of all clones of v. All vertices in $\mathcal{M}_v$ outside the skeleton will need to be absorbed. But since the number of these vertices is odd, it would be impossible to pair up (in the way described above) the absorbing structures assigned to these vertices. To fix this issue, we effectively impose that $\mathfrak{H}$ will "go around T twice". That is, the skeleton will trace through every molecule beginning and finishing at the lowest ordered vertex in Γ^*. It will then retrace its steps through these molecules in an almost identical way, effectively doubling the size of the skeleton. This ensures that the skeleton contains an even number of vertices from each molecule, half of them of each parity.

Finally, once we have obtained an appropriate skeleton, we can construct a long cycle $\mathfrak{H}$ as described in section 2.2.4. For every vertex in $\mathcal{Q}^n$ which is not covered by $\mathfrak{H}$ we have put in place an absorbing structure, which is covered by $\mathfrak{H}$ as described in (a) and (b). Thus, as discussed before, we can now use these structures to absorb all remaining vertices into $\mathfrak{H}$ to obtain a Hamilton cycle $\mathfrak{H}' \subseteq H \cup G$, thus proving the case $k = 1$ of theorem 2.1.7.

2.2.6. Hitting time for the appearance of a Hamilton cycle.

In order to prove theorem 2.1.5, we consider $G \sim \mathcal{Q}^n_{1/2-\varepsilon}$. We show that a.a.s., for any graph H with $\delta(H) \geq 2$, the graph $G \cup H$ is Hamiltonian (i.e., theorem 2.1.8 holds). The main additional difficulty faced here is that $G \cup H$ may contain vertices having degree as low as 2. For the set $\mathcal{U}$ of these vertices we cannot hope to use the previous absorption strategy: the neighbours of $v \in \mathcal{U}$ may not lie in cubes from $\mathcal{C}$. (In fact, v may not even have a neighbour within its own layer in $G \cup H$.) To handle such small degree vertices, we first prove that they will be few and well spread out (see lemma 2.7.7). In section 2.7.2 we define three types of new "special absorbing structures". The type of the special absorbing structure $SA(v)$ for v will depend on whether the neighbours a, b of v in H lie in the same layer as v. In each case, $SA(v)$ will consist of a short path P_1 containing the edges av and bv, and several other short paths designed to "balance out" P_1 in a suitable way. (This is further discussed in section 2.7.2, see figure 2.2.) These paths will be incorporated into the long cycle $\mathfrak{H}$ described in section 2.2.4. In particular, this allows us to "absorb" the vertices of $\mathcal{U}$ into $\mathfrak{H}$. To incorporate the paths P_i forming $SA(v)$, we will proceed as follows.

Firstly, we make use of the fact that theorem 2.5.11 allows us to choose our near-spanning tree T in such a way that it avoids a small ball around each $v \in \mathcal{U}$. Thus, (all clones of) T will avoid $SA(v)$, which has the advantage there will be no interference between T and the special absorbing structures. To link up each $SA(v)$ with the long cycle $\mathfrak{H}$, for each endpoint w of a path in $SA(v)$, we will choose an ℓ-cube in $I(G)$ which suitably intersects T and which contains w (or more precisely, the vertex in $I(G)$ corresponding to w). Altogether, these ℓ-cubes allow us to find paths between $SA(v)$ and vertices of $\mathfrak{H}$ which are clones of vertices in

T. The remaining vertices in molecules consisting of clones of these ℓ-cubes will be covered in a similar way as in section 2.2.4. All vertices in these balls around $\mathcal{U}$ which are not part of the special absorbing structures will be absorbed into $\mathfrak{H}$ via the same absorbing structures used in the proof of theorem 2.1.7 to once again obtain a Hamilton cycle $\mathfrak{H}'$.

2.2.7. Edge-disjoint Hamilton cycles.

The results on k edge-disjoint Hamilton cycles can be deduced from suitable versions (theorem 2.8.1) of the case $k = 1$. Those versions are carefully formulated to allow us to repeatedly remove a Hamilton cycle from the original graph. We deduce theorem 2.1.5 from theorem 2.8.1 in section 2.8.2.

2.3. Notation

For $n \in \mathbb{Z}$, we denote $[n] \coloneqq \{k \in \mathbb{Z} : 1 \leq k \leq n\}$ and $[n]_0 \coloneqq \{k \in \mathbb{Z} : 0 \leq k \leq n\}$. Whenever we write a hierarchy of parameters, these are chosen from right to left. That is, whenever we claim that a result holds for $0 < a \ll b \leq 1$, we mean that there exists a non-decreasing function $f\colon [0,1) \to [0,1)$ such that the result holds for all $a > 0$ and all $b \leq 1$ with $a \leq f(b)$. We will not compute these functions explicitly. Hierarchies with more constants are defined in a similar way.

A *hypergraph* H is an ordered pair $H = (V(H), E(H))$ where $V(H)$ is called the vertex set and $E(H) \subseteq 2^{V(H)}$, the edge set, is a set of subsets of $V(H)$. If $E(H)$ is a multiset, we refer to H as a *multihypergraph*. We say that a (multi)hypergraph H is *r-uniform* if for every $e \in E(H)$ we have $|e| = r$. In particular, 2-uniform hypergraphs are simply called *graphs*. Given any set of vertices $V' \subseteq V(H)$, we denote the subhypergraph of H *induced* by V' as $H[V'] \coloneqq (V', E')$, where $E' \coloneqq \{e \in E(H) : e \subseteq V'\}$. We write $H - V' \coloneqq H[V \setminus V']$. Given any set $\hat{E} \subseteq E(H)$, we will sometimes write $V(\hat{E}) \coloneqq \{v \in V : \text{there exists } e \in \hat{E} \text{ such that } v \in e\}$.

Given any (multi)hypergraph H and any vertex $v \in V(H)$, let $E(H, v) \coloneqq \{e \in E(H) : v \in e\}$. We define the *neighbourhood* of v as $N_H(v) \coloneqq \bigcup_{e \in E(H,v)} e \setminus \{v\}$, and we define the *degree* of v by $d_H(v) \coloneqq |E(H, v)|$. We denote the minimum and maximum degrees of (the vertices in) H by $\delta(H)$ and $\Delta(H)$, respectively. Given any pair of vertices $u, v \in V(H)$, we define $E(H, u, v) \coloneqq \{e \in E(H) : \{u, v\} \subseteq e\}$. The *codegree* of u and v in H is given by $d_H(u, v) \coloneqq |E(H, u, v)|$. Given any set of vertices $W \subseteq V(H)$, we define $N_H(W) \coloneqq \bigcup_{w \in W} N_H(w)$. We denote $E(H, v, W) \coloneqq \{e \in E(H) : v \in e, e \setminus \{v\} \subseteq W\}$, $N_H(v, W) \coloneqq \bigcup_{e \in E(H,v,W)} e \setminus \{v\}$ and $d_H(v, W) \coloneqq |E(H, v, W)|$; we refer to the latter two as the neighbourhood and degree of v into W, respectively. Given $A, B \subseteq V(H)$ we denote $E_H(A, B) \coloneqq \{e \in E(H) : e \subseteq A \cup B, e \cap A \neq \varnothing, e \cap B \neq \varnothing\}$ and $e_H(A, B) \coloneqq |E_H(A, B)|$. Whenever $A = \{v\}$ is a singleton, we abuse notation and write $E_H(v, B)$ and $e_H(v, B)$. Thus, $e_H(v, B)$ and $d_H(v, B)$ may be used interchangeably.

Given any graph G and two vertices $u, v \in V(G)$, the *distance* $\mathrm{dist}_G(u, v)$ between u and v in G is defined as the length of the shortest path connecting u and v (and it is said to be infinite if there is no such path). Similarly, given any sets $A, B \subseteq V(G)$, the *distance* between A and B is given by $\mathrm{dist}_G(A, B) \coloneqq \min_{u \in A, v \in B} \mathrm{dist}_G(u, v)$. For any $r \in \mathbb{N}$, we denote $B^r_G(u) \coloneqq \{v \in V(G) : \mathrm{dist}_G(u, v) \leq r\}$ and $B^r_G(A) \coloneqq \{v \in V(G) : \mathrm{dist}_G(A, v) \leq r\}$; we refer to these sets as the *balls* of radius r around u and A, respectively.

A *directed graph* (or *digraph*) is a pair $D = (V(D), E(D))$, where $E(D)$ is a set of ordered pairs of elements of $V(D)$. If no pair of the form (v, v) with $v \in V(D)$ belongs to $E(D)$, we say that D is *loopless*. Given any $v \in V(D)$, we define its *inneighbourhood* as $N^-_D(v) \coloneqq \{u \in V(D) : (u, v) \in E(D)\}$, and its *outneighbourhood* as $N^+_D(v) \coloneqq \{u \in V(D) : (v, u) \in E(D)\}$. The *indegree* and *outdegree* of v are defined as $d^-_D(v) \coloneqq |N^-_D(v)|$ and $d^+_D(v) \coloneqq |N^+_D(v)|$, respectively. The minimum in- and outdegrees of (the vertices in) D are denoted by $\delta^-(D)$ and $\delta^+(D)$, respectively.

Given any multihypergraph or directed graph (V, E), a set $M \subseteq E$ is called a *matching* if its elements are pairwise disjoint. If the edges of M cover all of V, then it is said to be a *perfect matching*. Given an edge-colouring c of H, we say that a matching of H is *rainbow* if each of its edges has a different colour in c.

We often refer to the n-dimensional hypercube $\mathcal{Q}^n$ as an *n-cube* (the n is dropped whenever clear from the context). Given two vertices $v_1, v_2 \in V(\mathcal{Q}^n) = \{0, 1\}^n$, we write $\mathrm{dist}(v_1, v_2)$ for the Hamming distance between v_1 and v_2. Thus, $\{v_1, v_2\} \in E(\mathcal{Q}^n)$ if and only if $\mathrm{dist}(v_1, v_2) = 1$. Whenever the dimension n is clear from the context, we will use $\mathbf{0}$ to denote the vertex $\{0\}^n$. Given any $v \in \{0,1\}^n$, we will say that its *parity* is *even* if $\mathrm{dist}(v, \mathbf{0}) \equiv 0 \pmod 2$, and we will say that it is *odd* otherwise. This gives a natural partition of $V(\mathcal{Q}^n)$ into the sets of vertices with even and odd parities. Given any two vertices $v_1, v_2 \in \{0,1\}^n$, we will write $v_1 =_{\mathrm{p}} v_2$ if they have the same parity, and $v_1 \neq_{\mathrm{p}} v_2$ otherwise.

We will often consider the natural embedding of $V(\mathcal{Q}^n)$ into $\mathbb{F}_2^n$, which will allow us to use operations on the vertex set: whenever we write $v + u$, for some $u, v \in \{0,1\}^n$, we refer to their sum in $\mathbb{F}_2^n$. Given a vertex $v \in \{0,1\}^n$ and an edge $e = \{x, y\} \in E(\mathcal{Q}^n)$, we define $v + e$ to be the edge with endvertices $v + x$ and $v + y$. Given any two sets $A, B \subseteq \{0,1\}^n$, we will use the sumset notation $A + B \coloneqq \{a + b : a \in A, b \in B\}$, and we will abbreviate the k-fold sumset $A + \ldots + A$ by kA. Similarly, given any sets $A \subseteq \{0,1\}^n$ and $E \subseteq E(\mathcal{Q}^n)$, we write $A + E \coloneqq \{a + e : a \in A, e \in E\}$. Given a graph $G \subseteq \mathcal{Q}^n$ and a set of vertices $A \subseteq \{0,1\}^n$, $A + G$ will denote the graph with vertex set $A + V(G)$ and edge set $A + E(G)$. Note that this should never be confused with the notation $G - A$, which will be used exclusively to consider induced subgraphs of G. We will call the unitary vectors in $\mathbb{F}_2^n$ the *directions* of the hypercube. The set of directions will be denoted by $\mathcal{D}(\mathcal{Q}^n)$. Thus, $\mathcal{D}(\mathcal{Q}^n) = \{\hat{e} \in \{0,1\}^n : \mathrm{dist}(\hat{e}, \mathbf{0}) = 1\}$. Note that two vertices $v_1, v_2 \in \{0,1\}^n$ are adjacent in $\mathcal{Q}^n$ if and only if there exists $\hat{e} \in \mathcal{D}(\mathcal{Q}^n)$ such that $v_1 = v_2 + \hat{e}$. Given any vertex $v \in \{0,1\}^n$ and any set $\mathcal{D} \subseteq \mathcal{D}(\mathcal{Q}^n)$, we will denote by $\mathcal{Q}^n(v, \mathcal{D}) \coloneqq \mathcal{Q}^n[v + n(\mathcal{D} \cup \{\mathbf{0}\})]$ the subcube of $\mathcal{Q}^n$ which contains v and all vertices in $\{0,1\}^n$ which can be reached from v by only adding directions in $\mathcal{D}$. Given any subcube $Q \subseteq \mathcal{Q}^n$, we will write $\mathcal{D}(Q)$ to denote the subset of $\mathcal{D}(\mathcal{Q}^n)$ such that, for any $v \in V(Q)$, we have $Q = \mathcal{Q}^n(v, \mathcal{D}(Q))$. Given any direction $\hat{e} \in \mathcal{D}(Q)$, we will sometimes informally say that Q *uses* $\hat{e}$. Given two vertices $v_1, v_2 \in \{0,1\}^n$, their *differing directions* are all directions in $\mathcal{D}(v_1, v_2) \coloneqq \{\hat{e} \in \mathcal{D}(\mathcal{Q}^n) : \mathrm{dist}(v_1 + \hat{e}, v_2) < \mathrm{dist}(v_1, v_2)\}$. Observe that, if $\mathrm{dist}(v_1, v_2) = d$, then $|\mathcal{D}(v_1, v_2)| = d$ and $\mathcal{Q}^n(v_1, \mathcal{D}(v_1, v_2))$ is the smallest subcube of $\mathcal{Q}^n$ which contains both v_1 and v_2.

When considering random experiments for a sequence of graphs $(G_n)_{n\in\mathbb{N}}$ with $|V(G_n)|$ tending to infinity with n, we say that an event $\mathcal{E}$ holds *asymptotically almost surely* (*a.a.s.*) for G_n if $\mathbb{P}[\mathcal{E}] = 1-o(1)$. When considering asymptotic statements, we will ignore rounding whenever this does not affect the argument.

2.4. Probabilistic tools

Here we list some probabilistic tools that we will use throughout the chapter. Throughout, we will be interested in proving concentration results for different random variables. We will often need the following Chernoff bound (see e.g. [71, Corollary 2.3]).

Lemma 2.4.1. *Let X be the sum of n mutually independent Bernoulli random variables and let $\mu := \mathbb{E}[X]$. Then, for all $0 < \delta < 1$ we have that $\mathbb{P}[X \geq (1+\delta)\mu] \leq e^{-\delta^2\mu/3}$ and $\mathbb{P}[X \leq (1-\delta)\mu] \leq e^{-\delta^2\mu/2}$. In particular, $\mathbb{P}[|X-\mu| \geq \delta\mu] \leq 2e^{-\delta^2\mu/3}$.*

The following bound will also be used repeatedly (see e.g. [5, Theorem A.1.12]).

Lemma 2.4.2. *Let X be the sum of n mutually independent Bernoulli random variables. Let $\mu := \mathbb{E}[X]$, and let $\beta > 1$. Then, $\mathbb{P}[X \geq \beta\mu] \leq (e/\beta)^{\beta\mu}$. In particular, we have $\mathbb{P}[X \geq 7\mu] \leq e^{-\mu}$.*

Given any sequence of random variables $X = (X_1, \ldots, X_n)$ taking values in a set Ω and a function $f\colon \Omega^n \to \mathbb{R}$, for each $i \in [n]_0$ define $Y_i := \mathbb{E}[f(X) \mid X_1, \ldots, X_i]$. The sequence $Y_0, \ldots, Y_n$ is called the *Doob martingale* for f and X. All the martingales that appear in this chapter will be of this form. To deal with them, we will need the following version of the well-known Azuma-Hoeffding inequality.

Lemma 2.4.3 (Azuma's inequality [8, 67])**.** *Let $X_0, X_1, \ldots$ be a martingale and suppose that $|X_i - X_{i-1}| \leq c_i$ for all $i \in \mathbb{N}$. Then, for any $n, t \in \mathbb{N}$,*

$$\mathbb{P}[|X_n - X_0| \geq t] \leq 2\exp\left(\frac{-t^2}{2\sum_{i=1}^{n} c_i^2}\right).$$

The following lemma, which concerns further bounds for martingales, is due to Alon, Kim, and Spencer [4] (see also [5, Theorem 7.4.3]). Here, we describe a version which is tailored to our purposes. Let $r \in \mathbb{N}$ and let $\mathcal{H}$ be an r-uniform hypergraph. Let $\mathcal{H}' \subseteq \mathcal{H}$ be a random subgraph chosen according to any distribution for which the inclusion of edges are mutually independent. Let X be a random variable whose value is determined by the presence or absence of the edges of some collection $E' = \{e_1, \dots, e_k\} \subseteq E(\mathcal{H})$ in $\mathcal{H}'$. Let p_i be the probability that e_i is present in $\mathcal{H}'$. Let c_i be the maximum value X could change, for some given choice of $\mathcal{H}'$, by changing the presence or absence of e_i. Let $C \coloneqq \max_{i\in[k]} c_i$ and $\sigma^2 \coloneqq \sum_{i\in[k]} p_i(1-p_i)c_i^2$.

Lemma 2.4.4 (Alon, Kim, and Spencer [4]). *For all $\alpha > 0$ with $\alpha C < 2\sigma$ we have that*

$$\mathbb{P}[|X - \mathbb{E}[X]| > \alpha\sigma] \leq 2e^{-\alpha^2/4}.$$

We will also need the following special case of Talagrand's inequality (see e.g. [5, Theorem 7.7.1]). Let $\Omega \coloneqq \prod_{i=1}^n \Omega_i$, where each Ω_i is a probability space. We say that $f\colon \Omega \to \mathbb{R}$ is *K-Lipschitz*, for some $K \in \mathbb{R}$, if for every $x, y \in \Omega$ which differ only on one coordinate we have $|f(x) - f(y)| \leq K$. We say that f is *h-certifiable*, for some $h\colon \mathbb{N} \to \mathbb{N}$, if, for every $x \in \Omega$ and $s \in \mathbb{R}$, whenever $f(x) \geq s$, there exists $I \subseteq [n]$ with $|I| \leq h(s)$ such that every $y \in \Omega$ that agrees with x on the coordinates in I satisfies $f(y) \geq s$.

Lemma 2.4.5 (Talagrand's inequality). *Let $\Omega \coloneqq \prod_{i=1}^n \Omega_i$, where each Ω_i is a probability space. Let $X\colon \Omega \to \mathbb{N}$ be K-Lipschitz and h-certifiable, for some $K \in \mathbb{N}$ and $h\colon \mathbb{N} \to \mathbb{N}$. Then, for all $b, t \in \mathbb{R}$,*

$$\mathbb{P}\left[X \leq b - tK\sqrt{h(b)}\right]\mathbb{P}[X \geq b] \leq \exp\left(\frac{-t^2}{4}\right).$$

Finally, the Lovász local lemma will come in useful. Let $\mathfrak{E} \coloneqq \{\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_m\}$ be a collection of events. A *dependency graph* for $\mathfrak{E}$ is a graph H on vertex set $[m]$ such that, for all $i \in [m]$, $\mathcal{E}_i$ is mutually independent of $\{\mathcal{E}_j : j \neq i, j \notin N_H(i)\}$, that is, if $\mathbb{P}[\mathcal{E}_i] = \mathbb{P}[\mathcal{E}_i \mid \bigwedge_{j\in J} \mathcal{E}_j]$ for all $J \subseteq [m] \setminus (N_H(i) \cup \{i\})$. We will use the following version of the local lemma (it follows e.g. from [5, Lemma 5.1.1]).

Lemma 2.4.6 (Lovász local lemma)**.** *Let* $\mathfrak{E} := \{\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_m\}$ *be a collection of events and let* H *be a dependency graph for* $\mathfrak{E}$*. Suppose that* $\Delta(H) \leq d$ *and* $\mathbb{P}[\mathcal{E}_i] \leq p$ *for all* $i \in [m]$*. If* $ep(d+1) \leq 1$*, then*

$$\mathbb{P}\left[\bigwedge_{i=1}^{m} \overline{\mathcal{E}_i}\right] \geq (1-ep)^m.$$

2.5. Auxiliary results

2.5.1. Results about matchings

We will need three auxiliary results to help us find suitable absorbing cube pairs for different vertices. We will need to preserve the alternating parities of vertices that are absorbed by each molecule. The first lemma (lemma 2.5.2) presented in this section, as well as its corollary, will help us to show that all vertices can be paired up in such a way that these parities can be preserved. The second lemma (lemma 2.5.4) will be used to show that, for each such pair of vertices, there are many possible pairs of absorption cubes. Finally, the third lemma (lemma 2.5.5) will allow us to assign one of those pairs of absorption cubes to each pair of vertices we need to absorb in such a way that these cube pairs are pairwise vertex disjoint.

To prove lemma 2.5.2, as well as theorem 2.5.12, the following consequence of Hall's theorem will be useful.

Lemma 2.5.1. *Let* G *be a bipartite graph with vertex partition* $A \mathbin{\dot\cup} B$*. Assume that there is some integer* $\ell \geq 0$ *such that, for all* $S \subseteq A$*, we have* $|N(S)| \geq |S| - \ell$*. Then,* G *contains a matching which covers all but at most* ℓ *vertices in* A*.*

Given any graph G and a bipartition $(\mathfrak{A}, \mathfrak{B})$ of $V(G)$, we say that $(\mathfrak{A}, \mathfrak{B})$ is an *r-balanced bipartition* if $||\mathfrak{A}| - |\mathfrak{B}|| \leq r$. Let G be a graph on n vertices, and let $r, d \in \mathbb{N}$ with $r \leq d$. We say that G is *d-robust-parity-matchable* with respect to an r-balanced bipartition $(\mathfrak{A}, \mathfrak{B})$ if, for every $S \subseteq V(G)$ such that $|S| \leq d$ and $|\mathfrak{A} \setminus S| = |\mathfrak{B} \setminus S|$, the graph $G - S$ contains a perfect matching M with the property that every edge $e \in M$ has one endpoint in $\mathfrak{A} \setminus S$ and one endpoint in $\mathfrak{B} \setminus S$.

Given two disjoint sets of vertices A and B, the binomial random bipartite graph $G(A,B,p)$ is obtained by adding each possible edge with one endpoint in A and the other in B with probability p independently of every other edge. Given any two bipartite graphs on the same vertex set, $G_1 = (A,B,E_1)$ and $G_2 = (A,B,E_2)$, and any $\alpha \in \mathbb{R}$, we define $\Gamma^{\alpha}_{G_1,G_2}(A)$ as the graph with vertex set A where any two vertices $x,y \in A$ are joined by an edge whenever $|N_{G_1}(x) \cap N_{G_2}(y)| \geq \alpha|B|$ or $|N_{G_1}(y) \cap N_{G_2}(x)| \geq \alpha|B|$.

Lemma 2.5.2. *Let $d,k,r \in \mathbb{N}$ and $\alpha,\varepsilon,\beta > 0$ be such that $r \leq d$, $1/k \ll 1/d, \varepsilon, \alpha$ and $\beta \ll \varepsilon, \alpha$. Then, any bipartite graph $G = G(A,B,E)$ with $|B| = n \geq |A| \geq k$ such that $d_G(x) \geq \alpha n$ for every $x \in A$ satisfies the following with probability at least $1 - 2^{-10n}$: for any r-balanced bipartition of A into $(\mathfrak{A},\mathfrak{B})$, the graph $\Gamma^{\beta}_{G,G(A,B,\varepsilon)}(A)$ is d-robust-parity-matchable with respect to $(\mathfrak{A},\mathfrak{B})$.*

Proof. Let $\Gamma := \Gamma^{\beta}_{G,G(A,B,\varepsilon)}(A)$. Let Γ' be the auxiliary digraph with vertex set A where, for any pair of vertices $x,y \in A$, there is a directed edge from x to y if $|N_G(x) \cap N_{G(A,B,\varepsilon)}(y)| \geq \beta n$. Observe that the graph obtained from Γ' by ignoring the directions of its edges and identifying the possible multiple edges is exactly Γ, which means that $\delta(\Gamma) \geq \delta^+(\Gamma')$.

Given any two vertices $x,y \in A$, by lemma 2.4.1 we have that

$$\mathbb{P}[(x,y) \notin E(\Gamma')] = \mathbb{P}[|N_G(x) \cap N_{G(A,B,\varepsilon)}(y)| < \beta n] \leq e^{-\varepsilon\alpha n/3}.$$

Furthermore, for a fixed $x \in A$, observe that the events that $(x,y) \notin E(\Gamma')$, for all $y \in A \setminus \{x\}$, are mutually independent. Therefore, $d^+_{\Gamma'}(x)$ is a sum of independent Bernoulli random variables. Let $m := |A|$. If $d^+_{\Gamma'}(x) < 4m/5$, that means that there is a set of $m/5$ vertices $Y \subseteq A \setminus \{x\}$ such that $(x,y) \notin E(\Gamma')$ for all $y \in Y$. We then conclude that

$$\mathbb{P}[d^+_{\Gamma'}(x) < 4m/5] \leq \sum_{Y \in \binom{A\setminus\{x\}}{m/5}} \mathbb{P}[(x,y) \notin E(\Gamma') \text{ for all } y \in Y] \leq \binom{m}{m/5} e^{-\varepsilon\alpha n m/15} \leq 2^{-20n}.$$

By a union bound over the choice of x, we conclude that

$$\mathbb{P}[\delta(\Gamma) < 4m/5] \leq \mathbb{P}[\delta^+(\Gamma') < 4m/5] \leq m2^{-20n} \leq 2^{-10n}.$$

Now, condition on the event that the previous holds. Fix any r-balanced bipartition $(\mathfrak{A}, \mathfrak{B})$ of A and let $\Gamma_{(\mathfrak{A},\mathfrak{B})}$ be the bipartite subgraph of Γ induced by this bipartition. Fix any set $S \subseteq A$ with $|S| \leq d$ and $|\mathfrak{A} \setminus S| = |\mathfrak{B} \setminus S|$. We have that $\delta(\Gamma_{(\mathfrak{A},\mathfrak{B})} - S) \geq 4m/5 - m/2 - d - r > m/4$. Therefore, by lemma 2.5.1, $\Gamma_{(\mathfrak{A},\mathfrak{B})} - S$ contains a perfect matching. □

For the proof of theorem 2.8.1 in section 2.8 we will need to use the following corollary 2.5.3. Let G be a graph on $2n$ vertices. Let $(\mathfrak{A}, \mathfrak{B})$ be a balanced bipartition of $V(G)$ and let $(V_1, \dots, V_k)$, for some $k \in \mathbb{N}$, be a partition of $V(G)$. Given any $d \in \mathbb{N}$, we say that G is d-robust-parity-matchable with respect to $(\mathfrak{A}, \mathfrak{B})$ *clustered in* $(V_1, \dots, V_k)$ if, for every $S \subseteq V(G)$ with $|S| \leq d$ and $|S \cap \mathfrak{A}| = |S \cap \mathfrak{B}|$, the graph $G - S$ contains a perfect matching M such that every edge $e \in M$ has one endpoint in $\mathfrak{A} \setminus S$ and one endpoint in $\mathfrak{B} \setminus S$ and, for every $e = \{x, y\} \in M$, if $x \in V_j$ then $y \in V_{j-1} \cup V_j \cup V_{j+1}$ (where we take indices cyclically).

Corollary 2.5.3. *Let $d, k, t \in \mathbb{N}$ and $\alpha, \varepsilon, \beta > 0$ be such that $1/k \ll 1/d, \varepsilon, \alpha$ and $\beta \ll \varepsilon, \alpha$. Let $G = G(A, B, E)$ be a bipartite graph and $(A_1, \dots, A_t)$ be a partition of A such that*

- $|B| = n \geq |A|$,
- *for every $i \in [t]$, we have that $|A_i| \geq k$ is even,*
- *for every $x \in A$, we have $d_G(x) \geq \alpha n$.*

Then, the following holds with probability at least $1 - 2^{-9n}$: for each $i \in [t]$ and for any balanced bipartition of A_i into $(\mathfrak{A}_i, \mathfrak{B}_i)$, the graph $\Gamma^{\beta}_{G,G(A,B,\varepsilon)}(A)$ is d-robust-parity-matchable with respect to $(\bigcup_{i=1}^t \mathfrak{A}_i, \bigcup_{i=1}^t \mathfrak{B}_i)$ clustered in $(A_1, \dots, A_t)$.

Proof. Given any set $C \subseteq A$, for each $i \in [t]$, let $C_i \coloneqq C \cap A_i$. Given any bipartition $(\mathfrak{A}, \mathfrak{B})$ of A, we write $C^{\mathfrak{A}} \coloneqq C \cap \mathfrak{A}$ and $C^{\mathfrak{B}} \coloneqq C \cap \mathfrak{B}$. Throughout this proof, we consider the indices in $[t]$ to be taken cyclically.

For each set $D \subseteq A$ with $|D| \leq d$, and for each $i \in [t]$, we apply lemma 2.5.2 to the graph $G[D_i \cup A_{i-1}, B]$, with $2d$, k, d, α, ε and β playing the roles of d, k, r, α, ε and β, respectively. Then, by a union bound over all choices of D and all choices of $i \in [t]$, the following holds with probability at least $1 - 2^{-9n}$. For each $i \in [t]$, consider any balanced bipartition $(\mathfrak{A}_i, \mathfrak{B}_i)$ of A_i. Consider any $D \subseteq A$ with $|D_i| \leq d$ for each $i \in [t]$. Then, for each $i \in [t]$, the graph $\Gamma^{\beta}_{G[D_i \cup A_{i-1}, B], G(D_i \cup A_{i-1}, B, \varepsilon)}(D_i \cup A_{i-1})$ is $2d$-robust-parity-matchable with respect to $((\mathfrak{A}_i \cap D) \cup \mathfrak{A}_{i-1}, (\mathfrak{B}_i \cap D) \cup \mathfrak{B}_{i-1})$. Condition on the event that the above holds.

Now, for each $i \in [t]$, fix a balanced bipartition $(\mathfrak{A}_i, \mathfrak{B}_i)$ of A_i. Let $\mathfrak{A} := \bigcup_{i=1}^{t} \mathfrak{A}_i$ and $\mathfrak{B} := \bigcup_{i=1}^{t} \mathfrak{B}_i$. Let $S \subseteq A$ be a subset of size $|S| \leq d$ such that $|S^{\mathfrak{A}}| = |S^{\mathfrak{B}}|$. We want to show that $\Gamma^{\beta}_{G, G(A,B,\varepsilon)}(A) - S$ contains a perfect matching M such that every edge $e \in M$ has one endpoint in $\mathfrak{A} \setminus S$ and one endpoint in $\mathfrak{B} \setminus S$ and, for every $e = \{x, y\} \in M$, if $x \in A_j$ then $y \in A_{j-1} \cup A_j \cup A_{j+1}$. We begin by proving the following claim.

Claim 2.1. *There exists a set $D \subseteq A \setminus S$ satisfying the following properties:*

(RM1) *for every $i \in [t]$ we have $|D_i| \leq d$, and*

(RM2) *for every $i \in [t]$ we have $|D^{\mathfrak{A}}_{i+1} \cup D^{\mathfrak{B}}_i \cup S^{\mathfrak{B}}_i| = |D^{\mathfrak{B}}_{i+1} \cup D^{\mathfrak{A}}_i \cup S^{\mathfrak{A}}_i|$.*

Proof. We will construct one such set D by constructing the sets $D_i \subseteq A_i$ inductively. We will argue by induction on $i \in [t]$ in decreasing order. Let $D_t := \varnothing$. Now, suppose that, for some $i \in [t-1]$, we have already constructed the sets $D_j \subseteq A_j$ for all $j \in [t] \setminus [i]$. Then, let $D_i \subseteq A_i \setminus S_i$ be a smallest set such that

$$|D^{\mathfrak{A}}_{i+1} \cup D^{\mathfrak{B}}_i \cup S^{\mathfrak{B}}_i| = |D^{\mathfrak{B}}_{i+1} \cup D^{\mathfrak{A}}_i \cup S^{\mathfrak{A}}_i|. \tag{2.1}$$

Observe that either $D_i = D^{\mathfrak{A}}_i$ or $D_i = D^{\mathfrak{B}}_i$. Furthermore, observe that $||D^{\mathfrak{A}}_{i+1} \cup S^{\mathfrak{B}}_i| - |D^{\mathfrak{B}}_{i+1} \cup S^{\mathfrak{A}}_i|| \leq |D_{i+1} \cup S_i|$. Therefore, there exists a set $D_i \subseteq A_i$ as required with $|D_i| \leq |D_{i+1} \cup S_i|$.

In order to prove that this results in a set D which satisfies the required properties, consider the following. First, by following the induction above, we have that $|D_t| = 0$ and $|D_i| \leq |D_{i+1}| + |S_i|$, hence $|D_i| \leq \sum_{j=1}^{t} |S_j| = |S| \leq d$ for all $i \in [t]$, thus (RM1) holds. On the other hand, (RM2) holds by (2.1) for all $i \in [t-1]$, so we must prove that it also holds for $i = t$. But this follows by summing (2.1) over all $i \in [t-1]$, and using the fact that $|S^{\mathfrak{A}}| = |S^{\mathfrak{B}}|$. ◀

Let D be the set given by Claim 2.1. Now, for each $i \in [t]$, let $J_i := D_i \cup S_i$. By Claim 2.1 (RM1) we have that $|J_i| \leq 2d$. Furthermore, by Claim 2.1 (RM2) it follows that $|(\mathfrak{A}_i \cup D^{\mathfrak{A}}_{i+1}) \setminus J_i| = |(\mathfrak{B}_i \cup D^{\mathfrak{B}}_{i+1}) \setminus J_i|$. By the conditioning above, this means that $\Gamma^{\beta}_{G[D_{i+1} \cup A_i, B], G(D_{i+1} \cup A_i, B, \varepsilon)}(D_{i+1} \cup A_i) - J_i$ contains a perfect matching M_i such that every edge of M_i has one endpoint in $(\mathfrak{A}_i \cup D^{\mathfrak{A}}_{i+1}) \setminus J_i$ and one endpoint in $(\mathfrak{B}_i \cup D^{\mathfrak{B}}_{i+1}) \setminus J_i$. Finally, let $M := \bigcup_{i=1}^{t} M_i$. It is clear that M satisfies the required conditions. The statement follows. □

The second lemma will be stated in terms of directed graphs.

Lemma 2.5.4. *Let $c, C > 0$ and let $\alpha \in (0, 1/(1 + c/C))$. Let D be a loopless n-vertex digraph such that*

(i) *for every $A \subseteq V(D)$ with $|A| \geq \alpha n$ we have $\sum_{v \in A} d^{-}(v) \geq c\alpha n$, and*

(ii) *for every $B \subseteq V(D)$ with $|B| \leq c\alpha n/C$ we have $\sum_{v \in B} d^{+}(v) \leq c\alpha n$.*

Then, D contains a matching M with $|M| > c\alpha n/(2C)$.

Proof. Assume for a contradiction that the largest matching M in D has size $|M| \leq c\alpha n/(2C)$. Since $\alpha < 1/(1 + c/C)$, there exists a set $A \subseteq V(D) \setminus V(M)$ with $|A| \geq \alpha n$, and thus, by (i), $\sum_{v \in A} d^{-}(v) \geq c\alpha n$. Since M is the largest matching, all edges that enter A must come from vertices of M (otherwise, we could add one such edge to M, finding a larger matching). However, by (ii), the number of edges going out of $V(M)$ is less than $c\alpha n$, a contradiction. □

For convenience, we state the third lemma in terms of rainbow matchings in hypergraphs.

Lemma 2.5.5. *Let $n, r \in \mathbb{N}$ and let $\mathcal{H}$ be an n-edge-coloured r-uniform multihypergraph. Then, for any $m \geq 10$, the following holds. Suppose $\mathcal{H}$ satisfies the following two properties:*

(i) *For every $i \in [n]$, there are at least m edges of colour i.*

(ii) $\Delta(\mathcal{H}) \leq m/(6r)$.

Then, there exists a rainbow matching of size n.

Proof. The idea is to pick a random edge from each colour class and prove that with non-zero probability this results in a rainbow matching. First, for each $i \in [n]$, let M_i be a set of m edges of colour i. We choose an edge from each M_i uniformly at random, independently of the other choices. For any $i, j \in [n]$ with $i \neq j$ and for any two edges $e \in M_i$ and $e' \in M_j$ for which $e \cap e' \neq \varnothing$, we denote by $A_{e,e'}$ the event that both e and e' are picked. We observe that

$$\mathbb{P}[A_{e,e'}] = \left(\frac{1}{m}\right)^2.$$

Moreover, note that every event $A_{e,e'}$ is independent of all other events $A_{f,f'}$ but at most $2m \cdot r \cdot \Delta(\mathcal{H}) \leq m^2/3$. Indeed, this holds because $A_{e,e'}$ can only depend on those events which involve at least one edge from either colour i or colour j. Applying now lemma 2.4.6, we deduce that with non-zero probability no event $A_{e,e'}$ occurs, as required. □

2.5.2. Properties of random subgraphs of the hypercube

In this section we state some basic properties of random subgraphs of the hypercube. Many of the proofs follow standard probabilistic arguments; due to space limitations, we have removed them from this exposition. For the sake of completeness, we include the proofs in appendix A.1.

The first property guarantees that the degrees of all vertices are linear in the dimension.

Lemma 2.5.6. *Let* $0 < \delta \ll \varepsilon \leq 1/2$*. Then, we a.a.s. have that* $\delta(\mathcal{Q}^n_{1/2+\varepsilon}) \geq \delta n$*.*

It will be important to have that a.a.s. there are not too many vertices whose degree deviates from the expected "close" to any given vertex.

Lemma 2.5.7. *Let* $\varepsilon \in (0,1)$*,* $a \in (2/3, 1)$ *and* $\ell \in \mathbb{N}$*. Then, for any* $b > 2 - 2a$*, a.a.s. there are no vertices* $v \in \{0,1\}^n$ *for which* $|\{u \in B^\ell(v) : d_{\mathcal{Q}^n_\varepsilon}(u) \neq \varepsilon n \pm n^a\}| \geq n^b$*.*

We are also interested in the number of subcubes in which each vertex lies. Given a graph G, a vertex v and any $\ell \in \mathbb{N}$, we denote the number of copies of $\mathcal{Q}^\ell$ in G which contain v by

$d_G^\ell(v)$. It is easy to give trivial upper bounds for this number by considering its value in $\mathcal{Q}^n$. Indeed, for all $v \in \{0,1\}^n$ we have that

$$d^\ell_{\mathcal{Q}^n_\varepsilon}(v) \leq \binom{n}{\ell}. \tag{2.2}$$

Lemma 2.5.8. *Let $\varepsilon \in (0,1)$, $a \in (1/2,1)$ and $\ell \in \mathbb{N}$. Then, a.a.s. all but at most $2^n e^{-n^{2a-1}/(6\varepsilon)}$ vertices $v \in V(\mathcal{Q}^n_\varepsilon)$ satisfy*

$$d^\ell_{\mathcal{Q}^n_\varepsilon}(v) = (1 \pm \mathcal{O}(n^{a-1}))\frac{\varepsilon^{2^{\ell-1}\ell}}{\ell!}n^\ell. \tag{2.3}$$

Remark 2.5.9. *In particular, the proof of lemma 2.5.8 shows that a.a.s. all vertices $v \in \{0,1\}^n$ which satisfy $d_{\mathcal{Q}^n_\varepsilon}(v) = \varepsilon n \pm n^a$ also satisfy (2.3). Therefore, by lemma 2.5.7, for any $r \in \mathbb{N}$ and $a \in (2/3,1)$, a.a.s. in any ball of radius r, all but at most $n^{2-2a+\eta}$ vertices satisfy (2.3), where $\eta > 0$ is an arbitrarily small constant.*

In more generality, we will need to bound the number of subcubes which contain a given pair of vertices. Given a graph G, two vertices u and v, and any $\ell \in \mathbb{N}$, we denote the number of copies of $\mathcal{Q}^\ell$ in G which contain both u and v by $d^\ell_G(u,v)$. Again, we can easily give upper bounds for this number in $\mathcal{Q}^n_\varepsilon$ by considering its value in $\mathcal{Q}^n$. Indeed, for all $u,v \in \{0,1\}^n$ we have that

$$d^\ell_{\mathcal{Q}^n}(u,v) \leq \binom{n}{\ell - \mathrm{dist}(u,v)} \leq n^{\ell-\mathrm{dist}(u,v)} \tag{2.4}$$

(here, we understand that $\binom{n}{a} = 0$ for all $a < 0$).

We will also need the property that the cubes containing a given vertex use different directions quite evenly. More precisely, given any graph $G \subseteq \mathcal{Q}^n$, any set $S \subseteq \mathcal{D}(\mathcal{Q}^n)$, two vertices $x,y \in \{0,1\}^n$, an integer $\ell \in \mathbb{N}$ and a real $t \in \mathbb{R}$, we denote by $d^\ell_{G,S,t,x}(y)$ the number of copies C of $\mathcal{Q}^\ell$ which contain y, do not contain x, and satisfy $|\mathcal{D}(C) \cap S| \geq t$.

Lemma 2.5.10. *Let $0 < 1/\ell \ll \delta < 1$, with $\ell \in \mathbb{N}$. Let $\varepsilon, \eta \in (0,1)$ and $a \in (2/3,1)$. Then, a.a.s. the following holds for every $x \in \{0,1\}^n$: for any set $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $|S| \geq \delta n$, all but at* *most* $n^{2(1+\eta-a)}$ *vertices* $y \in N_{\mathcal{Q}^n}(x)$ *satisfy*

$$d^{\ell}_{\mathcal{Q}^n_\varepsilon,S,\ell^{1/2},x}(y) \geq \frac{\varepsilon^{2^{\ell-1}\ell}}{2\ell!}n^{\ell}.$$

2.5.3. Near-spanning trees in random subgraphs of the hypercube

In this section we present our results on bounded degree near-spanning trees in $\mathcal{Q}^n_\varepsilon$. Due to space constraints, we will simply state the results which we need for the proof of theorem 2.1.8. All the details about the proofs of these results can be found in [34, Section 7]. Theorem 2.5.11 implies that with high probability there exists a near-spanning bounded degree tree in $\mathcal{Q}^n_\varepsilon$, which covers most of the neighbourhood of every vertex whilst avoiding a small random set of vertices, to which we refer as a reservoir. Theorem 2.5.12 allows us to extend the tree using vertices of the reservoir so that (amongst others) the proportion of uncovered vertices is even smaller. Finally, lemma 2.5.13 (to which we refer as the *repatching lemma*) states that, if some number of small local obstructions is prescribed, the tree given by theorem 2.5.12 can be slightly modified to avoid these obstructions.

Given a graph G and $\delta \in [0,1]$, let $Res(G,\delta)$ be a probability distribution on subsets of $V(G)$, where $R \sim Res(G,\delta)$ is obtained by adding each vertex $v \in V(G)$ to R with probability δ, independently of every other vertex. We will refer to this set R as a *reservoir*.

Theorem 2.5.11. *Let* $0 < 1/D, \delta \ll \varepsilon' \leq 1/2$*, and let* $\varepsilon, \gamma \in (0,1]$ *and* $k \in \mathbb{N}$*. Then, the following holds a.a.s. Let* $\mathcal{A} \subseteq V(\mathcal{Q}^n)$ *with the following two properties:*

(P1) *for any distinct* $x, y \in \mathcal{A}$ *we have* $\mathrm{dist}(x,y) \geq \gamma n$*, and*

(P2) $B^{k+2}_{\mathcal{Q}^n}(\mathcal{A}) \cap \{\{0\}^n, \{1\}^n, \{1\}^{\lceil n/2\rceil}\{0\}^{n-\lceil n/2\rceil}, \{0\}^{\lceil n/2\rceil}\{1\}^{n-\lceil n/2\rceil}\} = \varnothing$.

Let $R \sim Res(\mathcal{Q}^n,\delta)$*. Then, there exists a tree* $T \subseteq \mathcal{Q}^n_\varepsilon - (R \cup B^k_{\mathcal{Q}^n}(\mathcal{A}))$ *such that*

(T1) $\Delta(T) < D$,

(T2) *for all* $x \in V(\mathcal{Q}^n) \setminus B^k_{\mathcal{Q}^n}(\mathcal{A})$*, we have that* $|N_{\mathcal{Q}^n}(x) \cap V(T)| \geq (1-\varepsilon')n$*.*

Theorem 2.5.12. *For all* $0 < 1/n \ll 1/\ell, \varepsilon \leq 1$, *where* $n, \ell \in \mathbb{N}$, *the following holds. Let* $R, W \subseteq V(\mathcal{Q}^n)$ *and let* $T' \subseteq \mathcal{Q}^n - (R \cup W)$ *be a tree. For each* $x \in V(\mathcal{Q}^n) \setminus W$, *let* $Z(x) \subseteq N_{\mathcal{Q}^n}(x) \cap V(T')$ *be such that* $|Z(x)| \geq 3n/4$. *Then, a.a.s. there exists a tree* T *with* $T' \subseteq T \subseteq (\mathcal{Q}^n_\varepsilon \cup T') - W$ *such that*

(TC1) $\Delta(T) \leq \Delta(T') + 1$;

(TC2) *for all* $x \in V(\mathcal{Q}^n)$, *we have that* $|B^\ell_{\mathcal{Q}^n}(x) \setminus (V(T) \cup W)| \leq n^{3/4}$, *and*

(TC3) *for each* $x \in V(T) \cap R$, *we have that* $d_T(x) = 1$ *and the unique neighbour* x' *of* x *in* T *is such that* $x' \in Z(x)$.

Lemma 2.5.13. *Let* $0 < 1/n \ll c, \varepsilon, 1/f, 1/D$, *where* $f, D \in \mathbb{N}$. *Given a fixed* $x \in V(\mathcal{Q}^n)$, *let* $C(x) \subseteq N_{\mathcal{Q}^n}(x) \times N_{\mathcal{Q}^n}(x)$ *be such that* $|C(x)| \geq cn$ *and, for all distinct* $(y_1, z_1), (y_2, z_2) \in C(x)$, *we have* $\{y_1, z_1\} \cap \{y_2, z_2\} = \varnothing$. *Furthermore, for each* $(y, z) \in C(x)$, *let* $B(y, z) \subseteq (N_{\mathcal{Q}^n}(y) \cup N_{\mathcal{Q}^n}(z)) \setminus \{x\}$ *with* $|B(y, z)| < D$. *Then, with probability at least* $1 - e^{-5n}$, *for every* $F \subseteq V(\mathcal{Q}^n)$ *with* $|F| \leq f$, *there exist a pair* $(y, z) \in C(x)$ *with* $y, z \notin F$ *and a graph* $P \subseteq \mathcal{Q}^n_\varepsilon - \{y, z\}$ *with* $|V(P)| < 5D$ *such that*

(R1) $B(y, z) \cap N_{\mathcal{Q}^n}(y)$ *is connected in* P, *and so is* $B(y, z) \cap N_{\mathcal{Q}^n}(z)$.

(R2) $V(P) \cap F = \varnothing$.

2.6. Tiling random subgraphs of the hypercube with small cubes

Throughout this section, we will consider auxiliary hypergraphs to obtain information about subgraphs of the n-dimensional hypercube. The general idea will be to apply the so-called "Rödl nibble" to achieve this. Roughly speaking, the Rödl nibble is a randomised iterative process which, given an almost regular uniform hypergraph with small codegrees, finds a matching covering all but a small proportion of the vertices. The basic idea is the following. Let H be an almost regular uniform hypergraph with small codegrees. Consider a random subset of the edges $E \subseteq E(H)$, where each edge is taken independently with the same probability. If this probability is chosen carefully, then one can show that, with high probability, E is "almost" a matching and that the hypergraph resulting after the deletion of all vertices covered by E is still almost regular and has small codegrees. This allows one to iterate the process until all but a small fraction of the vertices have been covered. This approach is the basis for the proof of our main result in this section, theorem 2.6.6. The main auxiliary result is lemma 2.6.5, which shows that in each iteration of the process we have the properties we require. In particular, we require our matching to satisfy several additional "local" properties. This means our application of the nibble will require strong concentration results, as well as the use of the Lovász local lemma. It is also worth noting that our result relies strongly on the geometry of the hypercube, and cannot be stated for general hypergraphs.

2.6.1. The Rödl nibble

Given $\ell \in \mathbb{N}$ and any graph $G \subseteq \mathcal{Q}^n$, we will denote by $H_\ell(G)$ the 2^ℓ-uniform hypergraph with vertex set $V(G)$ where a set of vertices $W \subseteq \{0,1\}^n$ with $|W| = 2^\ell$ forms a hyperedge if and only if $G[W] \cong \mathcal{Q}^\ell$. Observe that the vertex set of $H_\ell(G)$ is (a subset of) $\{0,1\}^n$. Hence, we can use the underlying notation of directions we have considered for hypercubes so far. In particular, given any pair of vertices $x, y \in V(H_\ell(G))$, any set $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ and a real $t \in \mathbb{R}$, we denote by $d_{H_\ell(G),S,t,x}(y)$ the number of hyperedges $e \in E(H_\ell(G))$ which contain y, do not contain x, and satisfy $|\mathcal{D}(G[e]) \cap S| \geq t$. Note that, with the notation from lemma 2.5.10, $d_{H_\ell(G),S,t,x}(y) = d^\ell_{G,S,t,x}(y)$. In order to simplify notation, for any vertex $x \in \{0,1\}^n$ and any sets $Y \subseteq N_{\mathcal{Q}^n}(x)$ and $E \subseteq E(H_\ell(G))$, we let $E_x(Y) := \{e \in E : \operatorname{dist}(x,e) = 1, e \cap Y \neq \varnothing\}$. If E is the set of all edges of a given hypergraph $H \subseteq H_\ell(G)$, we may sometimes denote this by $E_x(H,Y)$. Furthermore, it is worth noting that, for hypergraphs $H_\ell(G)$ defined as above, the inequality $d_H(x) \leq \sum_{y \in V(H)\setminus\{x\}} d_H(x,y)$, which holds for all hypergraphs H and all $x \in V(H)$, can be improved to the following: for every $\ell \geq 2$ and every $x \in V(H_\ell(G))$,

$$d_{H_\ell(G)}(x) \leq \sum_{y \in V(H_\ell(G)) \cap N_{\mathcal{Q}^n}(x)} d_{H_\ell(G)}(x, y). \tag{2.5}$$

The following observations will also come in useful.

Remark 2.6.1. *Let $\ell, t \in \mathbb{N}$ and $G \subseteq \mathcal{Q}^n$, and let $H \coloneqq H_\ell(G)$. Let $x \in V(H)$ and $e \in E(H)$ be such that* $\operatorname{dist}(x, e) = t$*. Then, there is a unique vertex $y \in e$ such that* $\operatorname{dist}(x, y) = t$*. Furthermore, for every $e' \in E(H)$ such that $x \in e'$ we have that $e \cap e' \neq \varnothing$ if and only if $y \in e'$. In particular, $|\{e' \in E(H) : x \in e', e \cap e' \neq \varnothing\}| = d_H(x, y)$.*

Remark 2.6.2. *Let $\ell, t \in \mathbb{N}$ and $G \subseteq \mathcal{Q}^n$, and let $H \coloneqq H_\ell(G)$. Let $x \in V(H)$ and $Y \subseteq N_{\mathcal{Q}^n}(x)$. Let $e \in E(H)$ be such that* $\operatorname{dist}(x, e) = \operatorname{dist}(Y, e) = t$*. Let $Y' \coloneqq \{y \in Y : \operatorname{dist}(y, e) = t\}$. Then, $|Y'| \leq \ell$ and none of the edges in $E_x(H, Y \setminus Y')$ intersects e.*

Remark 2.6.3. *Let $\ell \in \mathbb{N}$ and $G \subseteq \mathcal{Q}^n$, and let $H \coloneqq H_\ell(G)$. Let $x \in V(H)$ and $Y \subseteq N_{\mathcal{Q}^n}(x)$. Then, for any $e \in E_x(H, Y)$, we have $|\{e' \in E_x(H, Y \setminus e) : e \cap e' \neq \varnothing\}| = \mathcal{O}(n^{\ell-1})$.*

Remark 2.6.4. *Let $k, n \in \mathbb{N}$ and $A \subseteq \{0,1\}^n$. Then,*

$$|\{v \in \{0,1\}^n : \operatorname{dist}(v, A) = k\}| \leq |A|\binom{n}{k}.$$

Consider $\ell \in \mathbb{N}$, $G \subseteq \mathcal{Q}^n$ and $H \coloneqq H_\ell(G)$. Recall that each edge of H corresponds to an ℓ-dimensional subcube of G. Let $e \in E(H)$, $E \subseteq E(H)$ and $S \subseteq \mathcal{D}(\mathcal{Q}^n)$. We define the *significance of e in S* as $\sigma(e, S) \coloneqq |\mathcal{D}(e) \cap S|$. Given any $t \in \mathbb{R}$, we say that e is *t-significant in S* if $\sigma(e, S) \geq t$. We define the *significance of E in S* as $\sigma(E, S) \coloneqq \sum_{e \in E} \sigma(e, S)$. We denote $\Sigma(E, S, t) \coloneqq \{e \in E : \sigma(e, S) \geq t\}$. In particular, $\sigma(E, S) \geq t|\Sigma(E, S, t)|$.

With this, we are now ready to state the main auxiliary result in this section. This shows that, given $H = H_\ell(G)$, under suitable conditions about the degrees, the codegrees and the local distribution of the edges of H along the directions of the cube (namely, that the edges are significant in every large set of directions), one iteration of our nibble process will yield a subset of edges which is locally close to a matching, satisfies several local properties that we require of our matching (namely, the edges given by the nibble are sufficiently significant in large sets of directions, and not too significant in any given direction), and its deletion yields a hypergraph which still satisfies almost the same suitable conditions for further iterations.

Lemma 2.6.5. *Let $\ell, k, K \in \mathbb{N}$ with $k > \ell \geq 2$ and let $\beta \in (0,1]$. Let $G \subseteq \mathcal{Q}^n$ and let $H := H_\ell(G)$. Fix $x \in \{0,1\}^n$. Let $A_0 := N_{\mathcal{Q}^n}(x)$ and, for each $i \in [K]$, let $A_i \subseteq A_0$ be a set of size $|A_i| \geq \beta n$. Assume that there exist two constants $a \in (3/4, 1)$ and $\gamma \in (0,1]$, and $D = \Theta(n^\ell)$, such that*

(P1) *for every $y \in V(H) \cap B^k_{\mathcal{Q}^n}(x)$ we have $d_H(y) = (1 \pm \mathcal{O}(n^{a-1}))D$;*

(P2) *for every $i \in [K]_0$ we have $|V(H) \cap A_i| = (1 \pm \mathcal{O}(n^{a-1}))\gamma|A_i|$.*

Then, for all $\varepsilon \ll 1/\ell$, the following holds.

Let $E' \subseteq E(H)$ be a random subset of $E(H)$ obtained by adding each edge with probability ε/D, independently of every other edge. Let $E'' \subseteq E'$ be the set of all edges not intersecting any other edge of E'. Then, E', E'', $V' := V(H) \setminus V(E')$ and $H' := H[V']$ satisfy the following:

(N1) *with probability at least $1 - e^{-\Theta(n^{1/2})}$, for every $i \in [K]_0$ we have*

$$|V' \cap A_i| = (1 \pm \mathcal{O}(n^{a-1}))e^{-\varepsilon}\gamma|A_i|;$$

(N2) *with probability at least $1 - e^{-\Theta(n^{1/2})}$, for every $i \in [K]_0$ we have*

$$|V(E'') \cap A_i| \geq \varepsilon(1 - 2^{\ell+1}\varepsilon)\gamma|A_i|;$$

(N3) *with probability at least $1 - e^{-\Theta(n^{1/2})}$, for every $\hat{e} \in \mathcal{D}(\mathcal{Q}^n)$ we have*

$$|\Sigma(E'_x(A_0), \{\hat{e}\}, 1)| = o(n^{1/2});$$

(N4) *with probability at least $1 - e^{-\Theta(n^{2a-1})}$, for every $y \in V(H') \cap B^{k-\ell}_{\mathcal{Q}^n}(x)$ we have*

$$d_{H'}(y) = (1 \pm \mathcal{O}(n^{a-1}))e^{-(2^\ell-1)\varepsilon}D.$$

If, in addition to (P1) and (P2), there exist $c, \delta \in (0,1]$ such that

(P3) *for every $i \in [K]_0$ and every $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $|S| \geq \delta n$ we have*

$$|\Sigma(E_x(H, A_i), S, \ell^{1/2})| \geq (1 - \mathcal{O}(n^{a-1}))c\gamma|A_i|D,$$

then E' and H' also satisfy the following:

(N5) *with probability at least $1 - e^{-\Theta(n^{1/2})}$, for every $i \in [K]_0$ and every $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $|S| \geq \delta n$ we have*

$$|\Sigma(E_x(H', A_i), S, \ell^{1/2})| \geq (c - \varepsilon)e^{-(2^\ell - 1)\varepsilon}\gamma|A_i|D;$$

(N6) *for any fixed $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $|S| \geq \delta n$, with probability at least $1 - e^{-\varepsilon c\gamma\beta n/100}$, for every $i \in [K]_0$ and n sufficiently large we have*

$$|V(\Sigma(E'_x(A_i), S, \ell^{1/2})) \cap A_i| \geq \varepsilon c^2\gamma|A_i|/8.$$

Proof. We begin by noting that, since $H = H_\ell(G)$, the following two properties hold:

(P4) for all $y, z \in V(H)$ such that $\mathrm{dist}(y, z) > \ell$, we have that $d_H(y, z) = 0$;

(P5) for each $i \in [\ell]$, for all $y, z \in V(H)$ with $\mathrm{dist}(y, z) = i$, we have that $d_H(y, z) = \mathcal{O}(D/n^i)$.

Indeed, by the definition of H, both follow from (2.4). These will be used repeatedly throughout the proof.

We next observe another simple property which will be useful later in the proof. Fix any $i \in [K]_0$. Note that, by (P1) and (P2), $|E_x(H, A_i)| = \sum_{y \in A_i \cap V(H)} d_H(y) \pm \ell n^\ell = (1 \pm \mathcal{O}(n^{a-1}))\gamma|A_i|D$. Therefore, $\mathbb{E}[|E'_x(A_i)|] = (1 \pm \mathcal{O}(n^{a-1}))\varepsilon\gamma|A_i|$ and, by lemma 2.4.1,

$$\mathbb{P}[|E'_x(A_i)| \neq (1 \pm \mathcal{O}(n^{a-1}))\varepsilon\gamma|A_i|] = e^{-\Theta(n^{2a-1})}. \tag{2.6}$$

By a union bound over all $i \in [K]_0$, we conclude that, with probability $1 - e^{-\Theta(n^{2a-1})}$, we have $|E'_x(A_i)| = (1 \pm \mathcal{O}(n^{a-1}))\varepsilon\gamma|A_i|$ for all $i \in [K]_0$.

(N1): On the number of vertices in $\mathbf{A_i}$ remaining in $\mathbf{H'}$.

In order to prove that (N1) holds, fix $i \in [K]_0$. Let $Y_i := V(H) \cap A_i$ and fix any vertex $y \in Y_i$. By (P1) we have that $d_H(y) = (1 \pm \mathcal{O}(n^{a-1}))D$. Therefore,

$$\mathbb{P}[y \in V'] = (1 - \varepsilon/D)^{(1 \pm \mathcal{O}(n^{a-1}))D} = (1 \pm \mathcal{O}(n^{a-1}))e^{-\varepsilon}.$$

Thus, by (P2), $\mathbb{E}[|V' \cap A_i|] = (1 \pm \mathcal{O}(n^{a-1}))e^{-\varepsilon}\gamma|A_i|$. We must now prove that $|V' \cap A_i|$ concentrates with high probability. However, the events $\{y \notin V(E')\}_{y \in Y_i}$ are not necessarily independent.

In order to consider independent events, let $E^* := \{e \in E(H) : x \notin e, e \cap A_i \neq \varnothing\}$ and, for each $y \in Y_i$, let $d^*_H(y) := |\{e \in E^* : y \in e\}|$. By (P5), we have $d^*_H(y) = d_H(y) \pm \mathcal{O}(D/n) = (1 \pm \mathcal{O}(n^{a-1}))D$. Let $V^* := Y_i \setminus V(E' \cap E^*)$. For every $y \in Y_i$ we have that

$$\mathbb{P}[y \notin V(E' \cap E^*)] = (1 - \varepsilon/D)^{(1 \pm \mathcal{O}(n^{a-1}))D} = (1 \pm \mathcal{O}(n^{a-1}))e^{-\varepsilon},$$

hence $\mathbb{E}[|V^*|] = (1 \pm \mathcal{O}(n^{a-1}))e^{-\varepsilon}\gamma|A_i|$. Furthermore, the events $\{y \notin V(E' \cap E^*)\}_{y \in Y_i}$ are mutually independent, so by lemma 2.4.1 we have that

$$\mathbb{P}[|V^*| \neq (1 \pm \mathcal{O}(n^{a-1}))e^{-\varepsilon}\gamma|A_i|] \leq e^{-\Theta(n^{2a-1})}. \tag{2.7}$$

As $V' \cap A_i \subseteq V^*$, we conclude that

$$\mathbb{P}[|V' \cap A_i| \geq (1 + \mathcal{O}(n^{a-1}))e^{-\varepsilon}\gamma|A_i|] \leq e^{-\Theta(n^{2a-1})}. \tag{2.8}$$

In order to obtain the lower tail concentration, observe that

$$|V' \cap A_i| = |V^*| - |V^* \cap V(E' \setminus E^*)|,$$

so it will suffice to show that the last term in the previous expression is small with high probability. Let $\hat{E} := \{e \in E(H) : |e \cap A_i| > 1\}$. Note that $|V^* \cap V(E' \setminus E^*)| \leq 2^\ell |\hat{E} \cap E'|$, so it suffices to bound this quantity. By (P5), $|\hat{E}| = \mathcal{O}(D)$. Since edges are picked independently, we have that $\mathbb{E}[|\hat{E} \cap E'|] = \mathcal{O}(1)$ and, by lemma 2.4.2, $\mathbb{P}[|\hat{E} \cap E'| > n^{1/2}] \leq e^{-\Theta(n^{1/2})}$. Combining this with (2.7), we conclude that $\mathbb{P}[|V' \cap A_i| \leq (1-\mathcal{O}(n^{a-1}))e^{-\varepsilon}\gamma|A_i|] \leq e^{-\Theta(n^{1/2})}$. Together with (2.8), the previous yields

$$\mathbb{P}[|V' \cap A_i| \neq (1 \pm \mathcal{O}(n^{a-1}))e^{-\varepsilon}\gamma|A_i|] \leq e^{-\Theta(n^{1/2})}. \tag{2.9}$$

The statement of (N1) follows by a union bound over all $i \in [K]_0$.

(N2): On the number of vertices in $\mathbf{A_i}$ covered by the matching.

We now prove (N2). Fix $i \in [K]_0$ and let $Y_i := V(H) \cap A_i$. Observe that

$$|V(E'') \cap A_i| = |(V(H) \setminus V') \cap A_i| - |V(E' \setminus E'') \cap A_i|. \tag{2.10}$$

By (2.9) and (P2) we have that $|(V(H)\setminus V')\cap A_i| = (1\pm\mathcal{O}(n^{a-1}))(1-e^{-\varepsilon})\gamma|A_i|$ with probability at least $1 - e^{-\Theta(n^{1/2})}$, so let us consider the last term in (2.10).

Given any vertex $y \in Y_i$, by abusing notation, let $d_{E'}(y) := |\{e \in E' : y \in e\}|$. Observe that $y \in V(E' \setminus E'')$ if and only if $d_{E'}(y) \geq 2$ or $d_{E'}(y) = 1$ and, for the edge $e \in E'$ such that $y \in e$, there exists $z \in e \setminus \{y\}$ such that $d_{E'\setminus\{e\}}(z) \geq 1$. Let $\mathcal{B}(y)$ be the event that, conditioned on $d_{E'}(y) = 1$, there exists such a vertex z. Then, for any $y \in Y_i$ we have

$$\mathbb{P}[y \in V(E' \setminus E'')] \leq \mathbb{P}[d_{E'}(y) \geq 2] + \mathbb{P}[d_{E'}(y) = 1]\mathbb{P}[\mathcal{B}(y)]. \tag{2.11}$$

Observe that, by (P1), $d_{E'}(y) \sim \mathrm{Bin}((1 \pm \mathcal{O}(n^{a-1}))D, \varepsilon/D)$. Then, it is easy to check that

(2.12)

$$\mathbb{P}[d_{E'}(y) = 1] = (1 \pm \mathcal{O}(n^{a-1}))\varepsilon e^{-\varepsilon} \quad \text{and} \quad \mathbb{P}[d_{E'}(y) \geq 2] = 1 - (1 \pm \mathcal{O}(n^{a-1}))(1 + \varepsilon)e^{-\varepsilon}.$$

By a union bound and the fact that $\mathbb{P}[d_{E'\setminus e}(z) \geq 1] \leq \mathbb{P}[d_{E'}(z) \geq 1]$ for every $e \in E(H)$ and $z \in e \setminus \{y\}$, we also have that

$$\mathbb{P}[\mathcal{B}(y)] \leq (1 \pm \mathcal{O}(n^{a-1}))(2^{\ell} - 1)(1 - e^{-\varepsilon}). \tag{2.13}$$

Combining (2.11)–(2.13), for n sufficiently large we have that $\mathbb{P}[y \in V(E' \setminus E'')] \leq 2^{\ell}\varepsilon^2$. Hence, by considering all $y \in Y_i$ and (P2), we conclude that

$$\mathbb{E}[|V(E' \setminus E'') \cap A_i|] \leq (1 + \mathcal{O}(n^{a-1}))2^{\ell}\varepsilon^2\gamma|A_i|. \tag{2.14}$$

In order to prove concentration, we will resort to Talagrand's inequality. Consider $X :=$ $|V(E' \setminus E'') \cap A_i|$. This is a random variable on the probability space given by the product of the probability spaces associated with each edge of H being present in E'. In this setting, it is easy to see that X is a $\ell 2^{\ell}$-Lipschitz function. Furthermore, X is h-certifiable for $h\colon \mathbb{N} \to \mathbb{N}$ given by $h(s) = 2s$. Thus, by lemma 2.4.5, for any real values b and t we have that

$$\mathbb{P}\left[X \leq b - t\ell 2^{\ell}\sqrt{2b}\right]\mathbb{P}[X \geq b] \leq e^{-t^2/4}.$$

By considering the change of variables $c = b - t\ell 2^{\ell}\sqrt{2b}$, we conclude that, for any reals c and t,

$$\mathbb{P}\left[X \leq c\right]\mathbb{P}\left[X \geq \left(t\ell 2^{\ell+1/2} + \left(t^2\ell^2 2^{2\ell+1} + 4c\right)^{1/2}\right)^2/4\right] \leq e^{-t^2/4}. \tag{2.15}$$

Let $c := 3 \cdot 2^{\ell-1}\varepsilon^2\gamma|A_i|$ and $t := \Theta(n^{a-1/2})$. By Markov's inequality, we have that $\mathbb{P}[X \leq c] \geq$ $1/4$ for n sufficiently large. By substituting these into (2.15), we conclude that

$$\mathbb{P}\left[X \geq (1 + \mathcal{O}(n^{a-1}))c\right] \leq e^{-\Theta(n^{2a-1})}. \tag{2.16}$$

From this and (2.9), it follows that

$$\mathbb{P}[|V(E'')\cap A_i|\leq \varepsilon(1-2^{\ell+1}\varepsilon)\gamma|A_i|]\leq e^{-\Theta(n^{1/2})}.$$

The statement of (N2) follows by a union bound over all $i\in[K]_0$.

(N3): On the significance of E′ in any direction.

In order to prove (N3), we first observe that there are "few" edges in $E_x(H,A_0)$ which use any given direction. Indeed, given any vertex $y\in V(H)\cap A_0$ and any direction $\hat{e}\in\mathcal{D}(\mathcal{Q}^n)$, the number of edges $e\in E(H)$ containing y and such that $\hat{e}\in\mathcal{D}(e)$ equals the codegree of y and $y+\hat{e}$. Therefore, by (P5), there are $\mathcal{O}(D/n)$ such edges and, adding over all vertices $y\in V(H)\cap A_0$, we conclude that $|\Sigma(E_x(H,A_0),\{\hat{e}\},1)|=\mathcal{O}(D)$. Since $|\Sigma(E'_x(A_0),\{\hat{e}\},1)|\sim\mathrm{Bin}(|\Sigma(E_x(H,A_0),\{\hat{e}\},1)|,\varepsilon/D)$, it immediately follows that $\mathbb{E}[|\Sigma(E'_x(A_0),\{\hat{e}\},1)|]=\mathcal{O}(1)$ and, by lemma 2.4.2,

$$\mathbb{P}[|\Sigma(E'_x(A_0),\{\hat{e}\},1)|=\Omega(n^{1/2})]\leq e^{-\Theta(n^{1/2})}.$$

The statement of (N3) follows by a union bound over all directions $\hat{e}\in\mathcal{D}(\mathcal{Q}^n)$.

(N4): On the degrees in H′.

We now want to bound the degrees of vertices in H' in order to prove (N4). Consider any vertex $y\in V(H)$ such that $\mathrm{dist}(x,y)\leq k-\ell$. Condition on the event that $y\in V'$. First, observe that, by (P1) and (P5),

$$\mathbb{E}[d_{H'}(y)]=(1\pm\mathcal{O}(n^{a-1}))D(1-\varepsilon/D)^{(2^\ell-1)(1\pm\mathcal{O}(n^{a-1}))D}=(1\pm\mathcal{O}(n^{a-1}))e^{-(2^\ell-1)\varepsilon}D. \tag{2.17}$$

In order to bound the probability that $d_{H'}(y)$ deviates from its expectation, we will apply lemma 2.4.4. Observe that the value of $d_{H'}(y)$ is determined by the presence or absence of the edges of $E^\bullet:=\{e\in E(H): \text{there exists } e'\in E(H) \text{ such that } y\in e'\setminus e, e\cap e'\neq\varnothing\}$ in E'. Note that, for each $e\in E^\bullet$, the maximum possible change in the value of $d_{H'}(y)$ due to the presence or absence of e is $c_e:=|\{e'\in E(H): y\in e', e\cap e'\neq\varnothing\}|$. Let $C:=\max_{e\in E^\bullet}c_e$ and $\sigma^2:=\sum_{e\in E^\bullet}(\varepsilon/D)(1-\varepsilon/D)c_e^2$. We must now estimate the value of σ.

Partition $E^\bullet$ into sets E_i, $i \in [\ell]$, given by $E_i \coloneqq \{e \in E^\bullet : \operatorname{dist}(y,e) = i\}$. Observe that, by (P1) and remark 2.6.4, for all $i \in [\ell]$ we have

$$|E_i| = \mathcal{O}(n^i D). \tag{2.18}$$

Furthermore, for each $i \in [\ell]$ and each $e \in E_i$, it follows from remark 2.6.1 and (P5) that

$$c_e = \mathcal{O}(D/n^i). \tag{2.19}$$

In order to apply lemma 2.4.4, we will need to show that σ is not too small. For this, we claim that

$$\text{there exist } \Theta(nD) \text{ edges } e \in E_1 \text{ such that } c_e = \Theta(D/n). \tag{2.20}$$

Indeed, an averaging argument using (2.5) together with (P1) shows that there are $\Theta(n)$ vertices $z \in V(H) \cap N_{\mathcal{Q}^n}(y)$ such that $d_H(y,z) = \Theta(D/n)$. Let Z be the set of all those vertices z. For each $z \in V(H) \cap N_{\mathcal{Q}^n}(y)$, let $E_1(z) \coloneqq \{e \in E_1 : z \in e\}$. By remark 2.6.1, for every $e \in E_1(z)$, z is the unique vertex in e such that $\operatorname{dist}(y,z) = 1$, so this gives a partition of E_1. By (P1) and (P5), we have that $|E_1(z)| = \Theta(D)$ for every $z \in Z$. Again by remark 2.6.1, for every $z \in X$ and every $e \in E_1(z)$ we have that $c_e = d_H(y,z)$. (2.20) now follows.

In particular, (2.20) combined with (2.19) shows that $C = \Theta(D/n)$. Combining (2.18)–(2.20), it follows that

$$\sigma^2 = \Theta(D^2/n).$$

Now, by setting $\alpha \coloneqq \Theta(n^{a-1/2})$, we observe that $\alpha = o(\sigma/C)$ and, thus, by lemma 2.4.4 and (2.17),

$$\mathbb{P}[d_{H'}(y) \neq (1 \pm \mathcal{O}(n^{a-1}))e^{-(2^\ell-1)\varepsilon}D] \leq e^{-\Theta(n^{2a-1})}.$$

The statement of (N4) follows by a union bound over all vertices $y \in V(H) \cap B^{k-\ell}_{\mathcal{Q}^n}(x)$.

(N5): **On the significance of H′ in any large set of directions.**

We now turn our attention to (N5). Fix $i \in [K]_0$ and $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $|S| \geq \delta n$. By (2.6) we have that

$$|E'_x(A_i)| = (1 \pm \mathcal{O}(n^{a-1}))\varepsilon\gamma|A_i| \tag{2.21}$$

with probability at least $1 - e^{-\Theta(n^{2a-1})}$. Furthermore, by (2.9), we have that

$$|V' \cap A_i| = (1 \pm \mathcal{O}(n^{a-1}))e^{-\varepsilon}\gamma|A_i| \tag{2.22}$$

with probability at least $1 - e^{-\Theta(n^{1/2})}$. Reveal $E'_x(A_i)$, as well as all edges in E' which contain x and intersect A_i, and condition on the event that (2.21) and (2.22) hold (note that this event is determined by the edges we have revealed). For the remainder of the proof of (N5), all probabilistic statements refer to probabilities when revealing all other edges in E'.

Let $X' := |\Sigma(E_x(H', A_i), S, \ell^{1/2})|$. Note that X' is a sum of indicator random variables, one for each edge in $\Sigma(E_x(H, A_i), S, \ell^{1/2})$; we will refer to those edges $e \in \Sigma(E_x(H, A_i), S, \ell^{1/2})$ for which we have $\mathbb{P}[e \in \Sigma(E_x(H', A_i), S, \ell^{1/2})] \neq 0$ as *potential edges*. The set of potential edges is denoted by E_P. We now want to prove a lower bound on $|E_P|$. We know that $\Sigma(E_x(H', A_i), S, \ell^{1/2}) \subseteq \Sigma(E_x(H, A_i), S, \ell^{1/2})$. By (P3) we have that $|\Sigma(E_x(H, A_i), S, \ell^{1/2})| \geq (1 - \mathcal{O}(n^{a-1}))c\gamma|A_i|D$. Any edge of $\Sigma(E_x(H, A_i), S, \ell^{1/2})$ whose endpoint in A_i does not lie in V' is not a potential edge. By (2.22), (P1) and (P2), the number of such edges is at most $(1 + \mathcal{O}(n^{a-1}))(1 - e^{-\varepsilon})\gamma|A_i|D$. Furthermore, some of the edges in $E'_x(A_i)$ may intersect other edges in $\Sigma(E_x(H, A_i), S, \ell^{1/2})$ (and, if this happens, the latter are not potential edges). By remark 2.6.3 and (2.21), the number of such non-potential edges is $\mathcal{O}(D)$. Combining these bounds, we conclude that $|E_P| \geq (1 - \mathcal{O}(n^{a-1}))(c - (1 - e^{-\varepsilon}))\gamma|A_i|D$. Now, each of these potential edges contributes to X' if and only if none of its vertices lie in any edge in E'. By (P1) and (P5), it follows that, for each $e \in E_P$, $\mathbb{P}[e \in \Sigma(E_x(H', A_i), S, \ell^{1/2})] = (1 \pm \mathcal{O}(n^{a-1}))e^{-(2^\ell - 1)\varepsilon}$ and, therefore,

$$\mathbb{E}[X'] \geq (1 - \mathcal{O}(n^{a-1}))(c - (1 - e^{-\varepsilon}))e^{-(2^\ell - 1)\varepsilon}\gamma|A_i|D. \tag{2.23}$$

In order to prove concentration we will resort once more to lemma 2.4.4. Let $E_i^\bullet := \{e \in E(H) : e \cap A_i = \varnothing$ and there exists $e' \in \Sigma(E_x(H,A_i), S, \ell^{1/2})$ such that $e \cap e' \neq \varnothing\}$. The value of X' is determined uniquely by the presence or absence of the edges of $E_i^\bullet$ in E'. For each $e \in E_i^\bullet$, the maximum change in the value of X' due to the presence or absence of e can be bounded by $c_e := |\{e' \in \Sigma(E_x(H,A_i), S, \ell^{1/2}) : e' \cap e \neq \varnothing\}|$. Let $C := \max_{e \in E_i^\bullet} c_e$ and $\sigma^2 := \sum_{e \in E_i^\bullet} (\varepsilon/D)(1-\varepsilon/D)c_e^2$. We must now estimate the value of σ.

Partition $E_i^\bullet$ into sets E_i^j, $j \in [\ell]$, given by $E_i^j := \{e \in E_i^\bullet : \mathrm{dist}(e, A_i) = j\}$. Observe that, by (P1) and remark 2.6.4, for all $j \in [\ell]$ we have

$$|E_i^j| = \mathcal{O}(n^{j+1}D). \tag{2.24}$$

Furthermore, for each $j \in [\ell]$ and each $e \in E_i^j$, it follows from remarks 2.6.1 and 2.6.2 and (P5) that

$$c_e = \mathcal{O}(D/n^j). \tag{2.25}$$

In particular, we claim that

$$\text{there exist } \Theta(n^2D) \text{ edges } e \in E_i^1 \text{ such that } c_e = \Theta(D/n). \tag{2.26}$$

Indeed, by (P1), (P2) and (P3), there are at least $c\gamma|A_i|/2$ vertices $y \in A_i \cap V(H)$ such that $d_{H,S,\ell^{1/2},x}(y) \geq cD/2$. Let U_i denote the set of these vertices. Then, an averaging argument using (2.5) together with (P1) shows that, for each $y \in U_i$, there are $\Theta(n)$ vertices $z \in V(H) \cap (N_{Q^n}(y) \setminus \{x\})$ such that $d_{H,S,\ell^{1/2},x}(y,z) = \Theta(D/n)$. For each $y \in U_i$, let $Z_i(y)$ be the set of such vertices. Now, fix any $y \in U_i$ and, for each $z \in Z_i(y)$, let $E_i^1(z) := \{e \in E_i^1 : z \in e\}$. By (P1) and (P5), we have that $|E_i^1(z)| = \Theta(D)$ for every $z \in Z_i(y)$; furthermore, by remark 2.6.1, for every $e \in E_i^1(z)$, z is the unique vertex in e such that $\mathrm{dist}(y,z) = \mathrm{dist}(y,e) = 1$. Then, for every $z \in Z_i(y)$ and every $e \in E_1(z)$ we have that

$$c_e \geq d_{H,S,\ell^{1/2},x}(y,z) = \Theta(D/n).$$

(2.26) now follows by considering all vertices $y \in U_i$.

In particular, (2.26) combined with (2.25) shows that $C = \Theta(D/n)$. Combining (2.24)–(2.26), it follows that

$$\sigma^2 = \Theta(D^2).$$

Now, by setting $\alpha := n^{a-1}\mathbb{E}[X']/\sigma = \Theta(n^a)$, we observe that $\alpha = o(\sigma/C)$ and, thus, by lemma 2.4.4 and (2.23),

$$\mathbb{P}[X' < (c-\varepsilon)e^{-(2^\ell-1)\varepsilon}\gamma|A_i|D] \leq e^{-\Theta(n^{2a})}.$$

Since this holds for every $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $|S| \geq \delta n$, by a union bound we conclude that the same holds simultaneously for every such set S. Recall, however, that this holds after conditioning on the event that (2.21) and (2.22) hold, which happens with probability $1 - e^{-\Theta(n^{1/2})}$. Taking this into account and using a union bound over all choices of $i \in [K]_0$, the statement of (N5) follows.

(N6): On the significance of E′ in a large fixed set of directions.

We finally turn our attention to (N6). Fix $i \in [K]_0$. Let U_i denote the set of vertices $y \in A_i \cap V(H)$ such that $d_{H,S,\ell^{1/2},x}(y) \geq cD/2$. By (P1), (P2) and (P3) we have $|U_i| \geq c\gamma|A_i|/2$. Let $V_i := V(\Sigma(E'_x(A_i), S, \ell^{1/2})) \cap U_i$. For each $y \in U_i$ we have that

$$\mathbb{P}[y \notin V_i] \leq (1-\varepsilon/D)^{cD/2} = (1 \pm \mathcal{O}(1/D))e^{-\varepsilon c/2}.$$

Thus, we conclude that

$$\mathbb{E}[|V_i|] \geq (1-\mathcal{O}(1/D))(1-e^{-\varepsilon c/2})c\gamma|A_i|/2.$$

Note that the events $\{y \notin V_i\}_{y\in U_i}$ are mutually independent. Hence, by lemma 2.4.1,

$$\mathbb{P}[|V_i| \leq (1-e^{-\varepsilon c/2})c\gamma|A_i|/3] \leq e^{-\varepsilon c\gamma\beta n/99}.$$

Finally, note that $(1-e^{-\varepsilon c/2})c\gamma|A_i|/3 \geq (\varepsilon c/2-\varepsilon^2c^2/8)c\gamma|A_i|/3 \geq \varepsilon c^2\gamma|A_i|/8$. The statement of (N6) follows by a union bound over all $i \in [K]_0$. □

2.6.2. Iterating the nibble

By making use of lemma 2.6.5, we can now prove the main result of this section. Roughly speaking, theorem 2.6.6 states that, for any constant $\varepsilon > 0$ and $\ell \in \mathbb{N}$, with high probability the random graph $\mathcal{Q}^n_\varepsilon$ contains a set of ℓ-dimensional cubes which are vertex-disjoint, cover all but a small proportion of the vertices of $\mathcal{Q}^n_\varepsilon$, and are "sufficiently significant" with respect to every large set of directions, while not being "too significant" with respect to any given direction.

By analogy with the notation introduced before lemma 2.6.5, given any $\ell \in \mathbb{N}$, any $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ and any copy C of $\mathcal{Q}^\ell$ with $C \subseteq \mathcal{Q}^n$, we define the *significance of C in S* as $\sigma(C,S) := |\mathcal{D}(C)\cap S|$. Similarly, given any set $\mathcal{C}$ of ℓ-dimensional cubes in $\mathcal{Q}^n$, we define the *significance of $\mathcal{C}$ in S* as $\sigma(\mathcal{C},S) := \sum_{C\in\mathcal{C}}\sigma(C,S)$. We also denote $\Sigma(\mathcal{C},S,t) := \{C \in \mathcal{C} : \sigma(C,S) \geq t\}$. Given any $x \in \{0,1\}^n$ and any $Y \subseteq N_{\mathcal{Q}^n}(x)$, we denote $\mathcal{C}_x(Y) := \{C \in \mathcal{C} : \mathrm{dist}(x,C) = 1, V(C)\cap Y \neq \varnothing\}$. In particular, we will write $\mathcal{C}_x := \mathcal{C}_x(N_{\mathcal{Q}^n}(x))$.

Theorem 2.6.6. *Let $\varepsilon,\delta,\alpha,\beta \in (0,1)$ and $K,\ell \in \mathbb{N}$ be such that $1/\ell \ll \alpha \ll \beta$. For each $x \in \{0,1\}^n$, let $A_0(x) := N_{\mathcal{Q}^n}(x)$ and, for each $i \in [K]$, let $A_i(x) \subseteq A_0(x)$ be a set of size $|A_i(x)| \geq \beta n$. Then, the graph $\mathcal{Q}^n_\varepsilon$ a.a.s. contains a collection $\mathcal{C}$ of vertex-disjoint copies of $\mathcal{Q}^\ell$ such that the following properties are satisfied for every $x \in \{0,1\}^n$:*

(M1) $|A_0(x)\cap V(\mathcal{C})| \geq (1-\delta)n$;

(M2) *for every $\hat{e} \in \mathcal{D}(\mathcal{Q}^n)$ we have $|\Sigma(\mathcal{C}_x,\{\hat{e}\},1)| = o(n^{1/2})$;*

(M3) *for every $i \in [K]_0$ and every $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $\alpha n/2 \leq |S| \leq \alpha n$ we have*

$$|\Sigma(\mathcal{C}_x(A_i(x)), S, \ell^{1/2})| \geq |A_i(x)|/3000.$$

Proof. Let $n_0, k \in \mathbb{N}$ and $\varepsilon' > 0$ be such that $1/n_0 \ll 1/k \ll \varepsilon' \ll 1/\ell, \delta$, and let $n \geq n_0$. Let $H := H_\ell(\mathcal{Q}^n_\varepsilon)$. Observe that, with the notation from lemmas 2.5.8 and 2.5.10, for any $x \in \{0,1\}^n$, $y \in A_0(x)$, $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ and $t \in \mathbb{R}$ we have that $d_H(x) = d^\ell_{\mathcal{Q}^n_\varepsilon}(x)$ and $d_{H,S,t,x}(y) = d^\ell_{\mathcal{Q}^n_\varepsilon,S,t,x}(y)$. Let $D_1 := \varepsilon^{2^{\ell-1}\ell} n^\ell/\ell!$.

Claim 2.2. *We a.a.s. have that, for every $x \in \{0,1\}^n$ and every $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $|S| \geq \alpha n/2$,*

(C1) $|\{y \in B^{2\ell}_{\mathcal{Q}^n}(x) : d_H(y) \neq (1 \pm \mathcal{O}(n^{-1/8}))D_1\}| \leq n^{7/8}$, *and*

(C2) $|\{y \in A_0(x) : d_{H,S,\ell^{1/2},x}(y) < D_1/2\}| \leq n^{3/4}$.

Proof. (C1) holds a.a.s. by remark 2.5.9 applied with $a = 7/8$ and 2ℓ playing the role of r. (C2) holds a.a.s. by applying lemma 2.5.10 with $a = 3/4$ and $\alpha/2$ playing the role of δ. ◀

Now, we condition on (C1) and (C2) and will show that there exists a collection $\mathcal{C}$ of vertex-disjoint copies of $\mathcal{Q}^\ell$ in $\mathcal{Q}^n_\varepsilon$ satisfying (M1)–(M3), as desired. In order to do this, we would like to apply lemma 2.6.5 to H with $a = 7/8$ and $D = D_1$. However, H does not satisfy all the required properties. It is worth noting that it does satisfy (P4) and (P5), which follow immediately from (2.4). The argument now will be to modify H slightly so that lemma 2.6.5 can be applied, independently of the choice of $x \in \{0,1\}^n$, and then iterate.

Claim 2.3. *There exists $H_1 \subseteq H$ which satisfies (P1)–(P3) with $7/8$, D_1, 1, $1/2$ and $\alpha/2$ playing the roles of a, D, γ, c and δ, respectively, for every $x \in \{0,1\}^n$ and every value of $k > \ell$.*

Proof of Claim 2.3. We construct H_1 by removing from H all vertices $y \in \{0,1\}^n$ such that $d_H(y) \neq (1 \pm \mathcal{O}(n^{-1/8}))D_1$. We first need to show that this deletion does not substantially decrease the degrees of other vertices. In fact, we claim that, for any $y \in \{0,1\}^n$,

(2.27) $$\text{if } d_H(y) = (1 \pm \mathcal{O}(n^{-1/8}))D_1, \text{ then } d_{H_1}(y) = (1 \pm \mathcal{O}(n^{-1/8}))D_1.$$

Indeed, consider any vertex $y \in \{0,1\}^n$ which satisfies $d_H(y) = (1 \pm \mathcal{O}(n^{-1/8}))D_1$. By (P4), the removal of any vertex z such that $\mathrm{dist}(y,z) > \ell$ does not affect the degree of y. Furthermore, by (C1), the number of vertices $z \in B^{\ell}_{\mathcal{Q}^n}(y)$ such that $d_H(z) \neq (1 \pm \mathcal{O}(n^{-1/8}))D_1$ is at most $n^{7/8}$. Let Z be the set of all such vertices. By (P5), the number of edges incident to y which are removed because of some fixed $z \in Z$ is $\mathcal{O}(D_1/n)$. By adding over all $z \in Z$, we have that the number of edges incident to y which have been removed is $\mathcal{O}(D_1 n^{-1/8})$, and the claim follows. One can similarly prove that, for any $y \in V(H_1)$ and any $S \subseteq \mathcal{D}(\mathcal{Q}^n)$,

$$\text{if } d_{H,S,\ell^{1/2},x}(y) \geq D_1/2, \text{ then } d_{H_1,S,\ell^{1/2},x}(y) \geq (1 - \mathcal{O}(n^{-1/8}))D_1/2. \tag{2.28}$$

By (2.27), we now have that H_1 satisfies (P1) with the parameters stated in Claim 2.3 for every $x \in \{0,1\}^n$. It also follows that (P2) holds for every $x \in \{0,1\}^n$ since, by (C1), at most $n^{7/8}$ vertices are removed from the neighbourhood of any vertex in H. Finally, by (2.28) and (C2), we also have that H_1 satisfies (P3) for every $x \in \{0,1\}^n$. ◀

Therefore, we are in a position to apply lemma 2.6.5. The argument now will be as follows. In order to prove that a collection of cubes as described in the statement exists, we will take a random collection of cubes in an iterative manner. We will prove that such a collection satisfies the desired properties locally with high probability. Then, we will apply the local lemma to extend the properties to the whole hypergraph. For this, it is important to define the probability space we are working with.

Fix a vertex $x \in \{0,1\}^n$. We now proceed iteratively. Let G_1 be the graph obtained by deleting from $\mathcal{Q}^n_\varepsilon$ all vertices $y \in \{0,1\}^n$ which do not satisfy $d^{\ell}_{\mathcal{Q}^n_\varepsilon}(y) = (1 \pm \mathcal{O}(n^{-1/8}))D_1$. Note that $H_1 = H_\ell(G_1)$. Let $i \in [k]$ and suppose that we have already defined G_i and H_i, where $H_i = H_\ell(G_i)$. Choose a random set of edges $E_i \subseteq E(H_i)$ by adding each edge in $E(H_i)$ to E_i independently with probability ε'/D_i. Then, define $H_{i+1} \coloneqq H_i - V(E_i)$. (Observe that $H_{i+1} = H_\ell(G_{i+1})$, where $G_{i+1} \coloneqq G_i - V(E_i)$.) Finally, let $D_{i+1} \coloneqq e^{-(2^\ell-1)\varepsilon'}D_i$, and iterate for k steps.

The randomized process above defines a probability space on the sequences of outcomes of each iteration of the process. Formally, the process, when iterated, results in a random sequence $E^k := (E_1, \dots, E_k)$ of sets of edges of H_1. Note that, for each $i \in [k]$, the hypergraph H_{i+1} is uniquely determined by $(E_1, \dots, E_i)$, and H_1 does not depend on any of these sets; thus, the sequence E^k encodes all the information about the outcome of the iterative process. For any $i \in [k]$, we will write $E^i := (E_1, \dots, E_i)$. We will write $\mathbb{P}_{E^0}[E_1] := \mathbb{P}[\text{process outputs } E_1 \text{ on input } H_1]$ and, for each $i \in [k] \setminus \{1\}$, we will write $\mathbb{P}_{E^{i-1}}[E_i] := \mathbb{P}[\text{process outputs } E_i \text{ on input } H_i]$ (where H_i is determined by E^{i-1} for all $i \geq 2$). Whenever needed, we will treat E^0 as an empty sequence. Note that the choice of the process in any iteration affects the probability distribution on all subsequent iterations. For each $i \in [k]_0$, let Ω^i be the set of all sequences $E^i = (E_1, \dots, E_i)$ such that, for all $j \in [i]$, $\mathbb{P}_{E^{j-1}}[E_j] > 0$, and let $\Omega := \Omega^k$. Given any $\omega = E^k = (E_1, \dots, E_k) \in \Omega$, we write $\omega^i := E^i$. Consider any $\omega = E^k = (E_1, \dots, E_k) \in \Omega$. The probability distribution on the outputs of the iterative process is given by $\mathbb{P}_\Omega[\omega] := \prod_{j=1}^k \mathbb{P}_{E^{j-1}}[E_j]$. Similarly, the distribution on the outputs after i iterations of the process is given by $\mathbb{P}_{\Omega^i}[\omega^i] := \prod_{j=1}^i \mathbb{P}_{E^{j-1}}[E_j]$. Observe that, given $i \in [k]$ and $\omega' \in \Omega^i$, we have that

$$\mathbb{P}_\Omega[\omega^i = \omega'] = \mathbb{P}_{\Omega^i}[\omega']. \tag{2.29}$$

In particular, we wish to apply lemma 2.6.5 in each iteration of the process. In order to do so, we will restrict ourselves to a suitable subspace of Ω by conditioning (again, in an iterative way). Let $k_1 := \lfloor 1/(3\varepsilon') \rfloor$, $\gamma_1 := 1$ and $c_1 := 1/2$. For each $i \in [k_1]$, we proceed as follows. Given $E^{i-1} \in \Omega^{i-1}$ (and thus the hypergraph H_i), let $\mathcal{A}_i(E^{i-1})$ be the event that E_i, $E_i' := \{e \in E_i : e \cap V(E_i \setminus \{e\}) = \varnothing\}$, $V_i' := V(H_i) \setminus V(E_i)$ and $H_{i+1} = H_i[V_i']$ satisfy (N1)–(N5) with D_i, $(k - i + 1)\ell + 1$, γ_i, c_i, ε' and $\alpha/2$ playing the roles of D, k, γ, c, ε and δ, respectively (in all iterations we will use $a = 7/8$). Then, let $\gamma_{i+1} := e^{-\varepsilon'}\gamma_i$ and $c_{i+1} := c_i - \varepsilon'$, and iterate.

Claim 2.4. *For any $i \in [k_1]$, let $E^{i-1} \in \Omega^{i-1}$ be such that H_i satisfies (P1)–(P3) with D_i, $(k-i+1)\ell+1$, γ_i, c_i and $\alpha/2$ playing the roles of D, k, γ, c and δ, respectively. Then,*

$$\mathbb{P}_{E^{i-1}}[\mathcal{A}_i(E^{i-1})] \geq 1 - e^{-\Theta(n^{1/2})}.$$

Proof. This follows immediately from lemma 2.6.5. ◀

This will naturally lead us into applying lemma 2.6.5 iteratively. Indeed, in any given iteration, assume that H_i satisfies (P1)–(P3) with D_i, $(k-i+1)\ell+1$, γ_i, c_i and $\alpha/2$ playing the roles of D, k, γ, c and δ, respectively. Note that, for $i=1$, by Claim 2.3, these properties hold for every choice of $x \in \{0,1\}^n$ (but recall that we have now fixed x). If $\mathcal{A}_i(E^{i-1})$ holds, then, because of (N1), (N4) and (N5), the next hypergraph H_{i+1} satisfies (P1)–(P3) with D_{i+1}, $(k-i)\ell+1$, γ_{i+1}, c_{i+1} and $\alpha/2$ playing the roles of D, k, γ, c and δ, respectively, so lemma 2.6.5 can be applied again.

As discussed above, in order to apply lemma 2.6.5 fully in each iteration, we must condition on the event that certain properties are satisfied after the previous iteration (namely, the corresponding event $\mathcal{A}_i$ holds). For each $j \in [k_1]_0$, let $\Omega_*^j \coloneqq \{(E_1,\dots,E_j) \in \Omega^j : E_i \in \mathcal{A}_i(E^{i-1}) \text{ for all } i \in [j]\}$. We denote $\Omega_* \coloneqq \Omega_*^{k_1}$. Using Claim 2.4, it now easily follows by induction that, for any $i \in [k_1]$,

$$\mathbb{P}_{\Omega}[\omega^i \in \Omega_*^i] \geq 1 - e^{-\Theta(n^{1/2})}. \tag{2.30}$$

Now fix a set of directions $S \subseteq \mathcal{D}(Q^n)$ with $\alpha n/2 \leq |S| \leq \alpha n$. For each $i \in [k_1]$, let $\mathcal{B}_i(S) \subseteq \Omega^i$ be the event that $|V(\Sigma(E_{i_x}(A_j(x)), S, \ell^{1/2})) \cap A_j(x)| < \varepsilon' c_i^2 \gamma_i |A_j(x)|/8$ for some $j \in [K]_0$. In other words, $\mathcal{B}_i(S)$ is the event that, in the i-th iteration, (N6) fails for S. Let

$$\begin{aligned}
a_i &\coloneqq \max_{E^{i-1}\in\Omega_*^{i-1}} \mathbb{P}_{\Omega^i}[\omega \in \mathcal{B}_i(S) \mid \omega^{i-1} = E^{i-1}] = \max_{E^{i-1}\in\Omega_*^{i-1}} \mathbb{P}_{E^{i-1}}[\mathcal{B}_i(S)],\\
b_i &\coloneqq \max_{E^{i-1}\in\Omega_*^{i-1}} \mathbb{P}_{\Omega^i}[\omega \in \mathcal{B}_i(S) \cap \mathcal{A}_i(E^{i-1}) \mid \omega^{i-1} = E^{i-1}] = \max_{E^{i-1}\in\Omega_*^{i-1}} \mathbb{P}_{E^{i-1}}[\mathcal{B}_i(S) \cap \mathcal{A}_i(E^{i-1})],\\
c_i &\coloneqq \max_{E^{i-1}\in\Omega_*^{i-1}} \mathbb{P}_{\Omega^i}[\mathcal{A}_i(E^{i-1}) \mid \omega^{i-1} = E^{i-1}] = \max_{E^{i-1}\in\Omega_*^{i-1}} \mathbb{P}_{E^{i-1}}[\mathcal{A}_i(E^{i-1})].
\end{aligned}$$

By lemma 2.6.5(N6), for each $i \in [k_1]$ we have that

$$b_i \leq a_i \leq e^{-\varepsilon'\beta n/900} =: d.$$

Let $\mathcal{I}(S)$ be the set of indices $i \in [k_1]$ in which $\mathcal{B}_i(S)$ holds. Note that, for any set $\mathcal{I} \subseteq [k_1]$, by (2.29) we have that $\mathbb{P}_\Omega[\mathcal{I}(S) = \mathcal{I}] = \mathbb{P}_{\Omega^{k_1}}[\mathcal{I}(S) = \mathcal{I}]$. Using the definitions above and induction on k_1, it follows that

$$f := \mathbb{P}_{\Omega^{k_1}}[(\mathcal{I}(S) = \mathcal{I}) \wedge \Omega_*] \leq \prod_{i\in\mathcal{I}} b_i \prod_{i\in[k_1]\setminus\mathcal{I}} c_i \leq d^{|\mathcal{I}|}.$$

Let $X = X(S) := |\mathcal{I}(S)|$. By adding over all sets $\mathcal{I} \subseteq [k_1]$ with $|\mathcal{I}| \geq k_1/2$, we conclude that

$$\mathbb{P}_{\Omega^{k_1}}[(X \geq k_1/2) \wedge \Omega_*] \leq 2^{k_1} d^{k_1/2} \leq e^{-\beta n/7000}. \tag{2.31}$$

Let $\mathcal{B} \subseteq \Omega^{k_1}$ be the event that there exists a set $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $\alpha n/2 \leq |S| \leq \alpha n$ such that the event $\mathcal{B}_i(S)$ holds in at least $k_1/2$ iterations. A union bound on (2.31) over all choices of $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $\alpha n/2 \leq |S| \leq \alpha n$ allows us to conclude that $\mathbb{P}_{\Omega^{k_1}}[\mathcal{B} \wedge \Omega_*] \leq e^{-\Theta(n)}$. Finally, combining this with (2.29) and (2.30), we have that

$$\mathbb{P}_\Omega[\omega^{k_1} \in \mathcal{B}] \leq \mathbb{P}_\Omega[\omega^{k_1} \in \mathcal{B} \mid \omega^{k_1} \in \Omega_*] + \mathbb{P}_\Omega[\omega^{k_1} \in \overline{\Omega_*}] \leq e^{-\Theta(n^{1/2})}. \tag{2.32}$$

Now, for all $i \in [k] \setminus [k_1]$, we iterate as above with the difference that we no longer require (P3). Thus, we can no longer guarantee that (N5) or (N6) hold with high probability, but we still have (N1)–(N4) as above. To be more precise, given any $i \in [k] \setminus [k_1]$ and $E^{i-1} \in \Omega^{i-1}$ (and thus the hypergraph H_i), let $\mathcal{A}_i(E^{i-1})$ be the event that E_i, $E_i' := \{e \in E_i : e \cap V(E_i \setminus \{e\}) = \varnothing\}$, $V_i := V(H_i) \setminus V(E_i)$ and $H_{i+1} = H_i[V_i]$ satisfy (N1)–(N4) with D_i, $(k-i+1)\ell + 1$, ε' and γ_i playing the roles of D, k, ε and γ, respectively. Then, let $\gamma_{i+1} := e^{-\varepsilon'}\gamma_i$ and iterate. Similarly to Claim 2.4, we can now show the following.

Claim 2.5. *For any $i \in [k] \setminus [k_1]$, let E^{i-1} be such that H_i satisfies (P1) and (P2) with D_i, $(k-i+1)\ell+1$ and γ_i playing the roles of D, k and γ, respectively. Then,*

$$\mathbb{P}_{E^{i-1}}[\mathcal{A}_i(E^{i-1})] \geq 1 - e^{-\Theta(n^{1/2})}.$$

Proof. This follows immediately from lemma 2.6.5. ◀

Now, assume that H_i satisfies (P1) and (P2) with D_i, $(k-i+1)\ell+1$ and γ_i playing the roles of D, k and γ, respectively. Note that, for $i = k_1 + 1$, we have these properties by conditioning on $\mathcal{A}_{k_1}(E^{k_1-1})$. Then, if $\mathcal{A}_i(E^{i-1})$ holds, because of (N1) and (N4), the hypergraph H_{i+1} satisfies (P1) and (P2) with D_{i+1}, $(k-i)\ell+1$ and γ_{i+1} playing the roles of D, k and γ, respectively, so we may apply lemma 2.6.5 again.

Let $\mathcal{A} := \{(E_1, \dots, E_k) \in \Omega : E^{k_1} \in \Omega_* \cap \overline{\mathcal{B}}, E_i \in \mathcal{A}_i(E^{i-1})$ for all $i \in [k] \setminus [k_1]\}$. By combining (2.30), (2.32) and Claim 2.5, observe that $\mathbb{P}_\Omega[\mathcal{A}] \geq 1 - e^{-\Theta(n^{1/2})}$. For any $(E_1, \dots, E_k) \in \mathcal{A}$, let $E := \bigcup_{i=1}^{k} E_i$ and $E' := \bigcup_{i=1}^{k} E'_i$. Note that E' is a matching by construction, that is, it corresponds to a collection $\mathcal{C}'$ of vertex-disjoint copies of $\mathcal{Q}^\ell$ in $\mathcal{Q}^n_\varepsilon$. We will now show that $\mathcal{C}'$ satisfies (M1)–(M3) for our fixed vertex x. Indeed, (M1) and (M2) hold for x since $(E_1, \dots, E_k) \in \mathcal{A}$ implies that (N2) and (N3) hold in each iteration (note that (M1) follows from the case $i = 0$ of (N2)). In order to prove (M3) for x, consider the following. For each $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $\alpha n/2 \leq |S| \leq \alpha n$, there are at least $k_1/2$ iterations $i \in [k_1]$ in which (N6) holds (for all $j \in [K]_0$). For each such set S, let $\mathcal{I}^*(S) \subseteq [k_1]$ be the set of indices of all such iterations. In particular, for each $i \in \mathcal{I}^*(S)$ we have that $|V(\Sigma(E_{i_x}(A_j(x)), S, \ell^{1/2})) \cap A_j(x)| \geq \varepsilon' c_i^2 \gamma_i |A_j(x)|/8 \geq \varepsilon' |A_j(x)|/432$ for all $j \in [K]_0$. Furthermore, by (N1) and (N2), we know that the number of vertices of $A_j(x)$ covered by $E_i \backslash E'_i$ satisfies $|V(E_i \backslash E'_i) \cap A_j(x)| \leq 2^{\ell+2}(\varepsilon')^2 |A_j(x)|$, so $|V(\Sigma(E'_{i_x}(A_j(x)), S, \ell^{1/2})) \cap A_j(x)| \geq (1/432 - 2^{\ell+2}\varepsilon')\varepsilon' |A_j(x)|$ for all $j \in [K]_0$. By adding this over all $i \in \mathcal{I}^*(S)$, we conclude that $|V(\Sigma(E'_x(A_j(x)), S, \ell^{1/2})) \cap A_j(x)| \geq |A_j(x)|/3000$ for all $j \in [K]_0$, as we wanted to see.

In order to show that there exists a matching which satisfies (M1)–(M3) simultaneously for all $x \in \{0,1\}^n$, let $\mathcal{E}(x)$ be the event that they hold for x. By the above discussion, we have that $\mathbb{P}_\Omega[\overline{\mathcal{E}(x)}] \leq \mathbb{P}_\Omega[\overline{\mathcal{A}}] \leq e^{-\Theta(n^{1/2})}$ for each $x \in \{0,1\}^n$. Furthermore, throughout the iterative process, the presence or absence in E of any edges e such that $\mathrm{dist}(x,e) > k\ell + 1$ does not have any effect on $\mathcal{E}(x)$, so $\mathcal{E}(x)$ is mutually independent of all events $\mathcal{E}(y)$ with $\mathrm{dist}(x,y) \geq 3k\ell$. Thus, by lemma 2.4.6, we conclude that there is a choice of E which satisfies (M1)–(M3) for every $x \in \{0,1\}^n$. □

2.7. Connecting lemmas and other auxiliary results

The first step of the proofs of theorems 2.1.7 and 2.1.8 will be to consider a particular partition of the hypercube into subcubes. The structure of this partition will be used extensively throughout the rest of the chapter, so we first introduce the necessary notation in the next subsection. Then, in section 2.7.2 we present the different absorbing structures that we will need. In section 2.7.3 we prove several results regarding this structure, concerning its properties in $\mathcal{Q}^n_\varepsilon$ and with respect to a reservoir $R \sim \mathit{Res}(\mathcal{Q}^n,\delta)$. In section 2.7.4 we prove our *connecting lemmas*, which provide sets of paths in (sub)cubes which (roughly speaking) link up pairs of vertices and, together, span all vertices of these (sub)cubes.

2.7.1. Layers, molecules and atoms

Throughout this section, given any two vectors u and v, we will write uv for their concatenation. Consider $\mathcal{Q}^n$ and some $s \in \mathbb{N}$, with $s < n$. We divide $\mathcal{Q}^n$ into 2^s vertex-disjoint copies of $\mathcal{Q}^{n-s}$ as follows: for each $a \in \{0,1\}^s$, we consider the set of vertices $V_a := \{av : v \in \{0,1\}^{n-s}\}$, and consider the graph $\mathcal{Q}(a) := \mathcal{Q}^n[V_a]$. We will refer to each $\mathcal{Q}(a)$ as an *s-layer* of $\mathcal{Q}^n$ (s will be dropped whenever clear from the context). Given $\ell \leq n-s$, we will refer to any copy of a cube $\mathcal{Q}^\ell$ in one of the s-layers as an *ℓ-atom* (again, ℓ will be dropped whenever clear from the context).

Fix a Hamilton cycle $\mathcal{C}$ of $\mathcal{Q}^s$. By abusing notation, whenever necessary, we assume that the coordinate vector of each vertex of $\mathcal{C}$ is concatenated with $n-s$ 0's. $\mathcal{C}$ induces a cyclical ordering on $\{0,1\}^s$, which we will label as $a_1,\ldots,a_{2^s}$. In turn, this gives a cyclical ordering on the set of layers. For each $i \in [2^s]$, we denote $L_i := \mathcal{Q}(a_i)$. Given an ℓ-atom $\mathcal{A}$ in an s-layer

$\mathcal{Q}(a)$, we refer to $\mathcal{M}(\mathcal{A}) \coloneqq \mathcal{A} + V(\mathcal{C})$ as an (s,ℓ)-*molecule* (again, the parameters will be dropped when clear from the context). Thus $\mathcal{M}(\mathcal{A})$ is the vertex-disjoint union of 2^s copies of $\mathcal{Q}^\ell$. We refer to an $(s,1)$-molecule as a *vertex molecule* and an (s,ℓ)-molecule for $\ell > 1$ as a *cube molecule*. Observe that, if we label the atoms in a molecule cyclically following the labelling of the layers, then $\mathcal{Q}^n$ contains a perfect matching between any two consecutive atoms where all edges are in the same direction as the corresponding edge in $\mathcal{C}$. Whenever we work with molecules, we consider this cyclical order implicitly. In particular, whenever we refer to a molecule $\mathcal{M} = \mathcal{M}(\mathcal{A}) = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_{2^s}$, the cyclical order $\mathcal{A}_1 \cup \cdots \cup \mathcal{A}_{2^s}$ of the $\mathcal{A}_i$ is that induced by $\mathcal{C}$. Given a molecule $\mathcal{M}(\mathcal{A}) = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_{2^s}$, a *slice* $\mathcal{M}^* \subseteq \mathcal{M}(\mathcal{A})$ will consist of the subgraph of $\mathcal{M}(\mathcal{A})$ induced by its intersection with some number of consecutive layers, i.e. $\mathcal{M}^* = \mathcal{A}_{a+1} \cup \cdots \cup \mathcal{A}_{a+t}$ for some $a, t \in [2^s]$. Alternatively, given any $a \in V(\mathcal{C})$, any path $P \subseteq \mathcal{C}$ and any atom $\mathcal{A} \subseteq \mathcal{Q}(a)$, P determines a slice of $\mathcal{M}(\mathcal{A})$ by setting $\mathcal{M}^* \coloneqq \mathcal{A} + V(P)$.

Consider $i \in [2^s]$ and the cyclical ordering of the layers given by $\mathcal{C}$ (throughout this section, we will assume this ordering is given and fixed). Given any subgraph $G \subseteq \mathcal{Q}^n$, we will often denote the restriction of G to the i-th layer by $L_i(G)$, that is, $L_i(G) \coloneqq G[V(L_i)]$. Given any $v \in \{0,1\}^{n-s}$, we will refer to the vertex $a_i v$ as the i-th *clone* of v. In general, when it is clear from the context, we will also refer to the i-th clone of a cube $C \subseteq \mathcal{Q}^{n-s}$ (as well as other subgraphs), which, analogously, will be the corresponding copy in L_i of C. In particular, the i-th layer L_i is the i-th clone of $\mathcal{Q}^{n-s}$.

Consider the fixed ordering of the layers $L_1, \ldots, L_{2^s}$ of $\mathcal{Q}^n$ induced by $\mathcal{C}$. If we view these layers as different subgraphs on the vertex set of $\mathcal{Q}^{n-s}$, we can define the *intersection graph* of the layers $I \coloneqq \bigcap_{i=1}^{2^s} L_i$ (note that $I \cong \mathcal{Q}^{n-s}$) and, for any $G \subseteq \mathcal{Q}^n$, we denote $I(G) \coloneqq \bigcap_{i=1}^{2^s} L_i(G)$. Note that, if $\mathcal{G} \subseteq I(G)$, then there is a clone of $\mathcal{G}$ in $L_i(G)$, for each $i \in [2^s]$.

2.7.2. Absorbing structures

As we already discussed in section 2.2, in order to prove our results we will first construct a near-spanning cycle and then absorb the remaining vertices into this cycle. We will primarily achieve this by using the following absorbing structure.

Definition 2.7.1 (Absorbing ℓ-cube pair)**.** Let $\ell, n \in \mathbb{N}$, and let $G \subseteq \mathcal{Q}^n$. Given a vertex $x \in V(\mathcal{Q}^n)$, an *absorbing ℓ-cube pair for x* in G, which we denote by (C^l, C^r), is a subgraph of G which consists of two vertex-disjoint ℓ-dimensional cubes $C^l, C^r \subseteq G$ and three edges $e, e^l, e^r \in E(G)$ satisfying the following properties:

(AP1) $|V(C^l) \cap N_{\mathcal{Q}^n}(x)| = |V(C^r) \cap N_{\mathcal{Q}^n}(x)| = 1$;

(AP2) e^l and e^r are the unique edges from x to C^l and C^r, respectively;

(AP3) the unique vertex $y \in V(C^l) \cap N_{\mathcal{Q}^n}(x)$ satisfies $\mathrm{dist}(y, C^r) = 1$, and

(AP4) e is the unique edge from y to C^r.

We will refer to C^l as the *left absorption cube* and to C^r as the *right absorption cube*. Given an absorbing ℓ-cube pair (C^l, C^r) we refer to y as the *left absorber tip*, and to the unique vertex $z \in V(C^r) \cap N_{\mathcal{Q}^n}(x)$ as the *right absorber tip*. We refer to the unique vertex $z' \in e \setminus \{y\}$ as the *third absorber vertex*.

However, again as discussed in section 2.2, this will not suffice to absorb all vertices in the proof of theorem 2.1.8. In particular, there will be vertices of very low degree (as low as degree 2). We cannot hope to absorb these via absorbing ℓ-cube pairs, as they may not have any neighbours which lie in the cube factor we construct. Thus, we first construct alternative absorbing structures for these vertices.

To be more precise, in order to absorb a vertex of low degree we will define several paths. One path will contain the vertex of low degree, while the others are used to compensate the parities of vertices in this first path. Moreover, the paths will be constructed in such a way that they end in vertices which can be paired up so that each pair consists of clones of a given vertex of the intersection graph I and these two clones lie in a (bonded) cube molecule. These cube molecules can then be used to connect these paths. It is worth mentioning, however, that we cannot guarantee that the pairing of the vertices can be done within a single slice, so we need to alter our approach to deal with this.

Suppose $x \in V(\mathcal{Q}^n)$ is incident to one edge with direction a and one edge with direction b. We will construct three different types of special absorbing structures, to handle the cases where both, only one, or none of these two edges lie in the same layer as x. Representations of these three types of special absorbing structures can be found in figure 2.2. Let L be the layer such that $x \in V(L)$. Given a path $P = x_1 x_2 \dots x_k$ in $\mathcal{Q}^n$, we define $\mathrm{end}(P) := \{x_1, x_k\}$.

Type I. Assume that $x + a, x + b \in V(L)$. Let $f\colon V(\mathcal{Q}^n) \to V(\mathcal{Q}^n)$ be defined as follows: for each $i \in [2^s]$ and each $y \in V(L_i)$, if i is even, we set $f(y) := y + \hat{e}_{i-1}$; otherwise, we set $f(y) := y + \hat{e}_i$. By abusing notation, for any $F \subseteq L_i$, we also consider the graph $f(F)$, where for each edge $e = \{y, z\} \in E(F)$ we define $f(e) := \{f(y), f(z)\}$. Let $(c, d, d_1, d_2, d_3, d_4) \in (\mathcal{D}(L) \setminus \{a, b\})^6$ be a tuple of distinct directions and define the following paths in $\mathcal{Q}^n$:

- $P_1 := (x + a + d_1, x + a, x, x + b, x + b + d_2)$;
- $P_2 := (f(x + b + d_2), f(x + b), f(x + b + c))$;
- $P_3 := (x + c + b, x + c, x + c + d_3)$;
- $P_4 := (f(x + c + d_3), f(x + c), f(x), f(x + d), f(x + d + d_4))$;
- $P_5 := (x + d + d_4, x + d, x + d + a)$;
- $P_6 := (f(x + a + d), f(x + a), f(x + a + d_1))$.

Observe that, for every $y \in V(P_1 \cup \dots \cup P_6)$, we have that $f(y) = f^{-1}(y) \in V(P_1 \cup \dots \cup P_6)$. We say that $CS(x, a, b) := (P_1, \dots, P_6)$ is an (x, a, b)-*consistent system of paths*, and let

$$\mathrm{end}(CS(x, a, b)) := \bigcup_{i=1}^{6} \mathrm{end}(P_i).$$

Let $(P_1, \dots, P_6)$ be an (x, a, b)-consistent system of paths. Let $\mathcal{D} := \{c, d, d_1, d_2, d_3, d_4\} \subseteq \mathcal{D}(L)$ be the set of directions such that $P_1, \dots, P_6$ are as defined above. Let $\mathbf{C} := \{C_1, \dots, C_{12}\}$ be a collection of vertex-disjoint ℓ-cubes which satisfy the following:

(PI.1) for all $i \in [12]$, we have that either $C_i \subseteq L$ or $C_i \subseteq f(L)$;

(PI.2) for all $i \in [6]$, we have $f(C_{2i}) = C_{2i+1}$, where indices are taken modulo 12;

(PI.3) for all $i \in [6]$, C_{2i-1} contains the first vertex of P_i, and C_{2i} contains the last vertex of P_i;

(PI.4) for all $i \in [12]$, we have $\mathcal{D}(C_i) \cap (\mathcal{D} \cup \{a, b\}) = \varnothing$.

We say that $SA(x, a, b) := (P_1, \dots, P_6, C_1, \dots, C_{12})$ is an (x, a, b)-*special absorbing structure*.

Finally, given $G \subseteq \mathcal{Q}^n$ and an (x, a, b)-consistent system of paths $CS(x, a, b) = (P_1, \dots, P_6)$, we say that $CS(x, a, b)$ *extends* to an (x, a, b)-special absorbing structure in G if there is a collection $\mathbf{C} = \{C_1, \dots, C_{12}\}$ with $C_i \subseteq G$ such that $SA(x, a, b) = (P_1, \dots, P_6, C_1, \dots, C_{12})$ is an (x, a, b)-special absorbing structure.

Type II. Assume now that $x + a, x + b \notin V(L)$. Let $(d_1, d_2) \in (\mathcal{D}(L))^2$ be a pair of distinct directions and define the following paths in $\mathcal{Q}^n$:

- $P_1 := (x + a + d_1, x + a, x, x + b, x + b + d_2)$;
- $P_2 := (x + a + b + d_2, x + a + b, x + a + b + d_1)$.

We say that $CS(x, a, b) := (P_1, P_2)$ is an (x, a, b)-*consistent system of paths*, and let

$$\mathrm{end}(CS(x, a, b)) := \mathrm{end}(P_1) \cup \mathrm{end}(P_2).$$

Let (P_1, P_2) be an (x, a, b)-consistent system of paths. Let $\mathcal{D} := \{d_1, d_2\} \subseteq \mathcal{D}(L)$ be the set of directions such that P_1 and P_2 are as defined above. Let L_{ab}, L_a and L_b be the layers such that $x + a + b \in V(L_{ab})$, $x + a \in V(L_a)$ and $x + b \in V(L_b)$, respectively. Let $\mathbf{C} := \{C_1, C_2, C_3, C_4\}$ be a collection of vertex-disjoint ℓ-cubes which satisfy the following:

(PII.1) $C_1 \subseteq L_a$, $C_2 \subseteq L_b$ and $C_3, C_4 \subseteq L_{ab}$;

(PII.2) $C_1 + b = C_4$ and $C_2 + a = C_3$;

(PII.3) for all $i \in [2]$, C_{2i-1} contains the first vertex of P_i, and C_{2i} contains the last vertex of P_i;

(PII.4) for all $i \in [4]$, we have $\mathcal{D}(C_i) \cap \mathcal{D} = \varnothing$.

We say that $SA(x, a, b) := (P_1, P_2, C_1, \dots, C_4)$ is an (x, a, b)-*special absorbing structure*.

Finally, given any graph $G \subseteq \mathcal{Q}^n$ and an (x, a, b)-consistent system of paths $CS(x, a, b) = (P_1, P_2)$, we say that $CS(x, a, b)$ *extends* to an (x, a, b)-special absorbing structure in G if there is a collection $\mathbf{C} = \{C_1, \dots, C_4\}$ with $C_i \subseteq G$ such that $SA(x, a, b) = (P_1, P_2, C_1, \dots, C_4)$ is an (x, a, b)-special absorbing structure.

Type III. Finally, assume that $x + a \notin V(L)$ and $x + b \in V(L)$. For each vertex $y \in V(L)$, let $f(y) := y + a$. By abusing notation, for any $F \subseteq L$, we also consider the graph $f(F)$, where for each edge $e = \{y, z\} \in E(F)$ we define $f(e) := \{f(y), f(z)\}$. Let $(d_1, d_2, d_3) \in (\mathcal{D}(L) \setminus \{b\})^3$ be a tuple of distinct directions and define the following paths in $\mathcal{Q}^n$:

- $P_1 := (f(x + d_1 + d_2), f(x + d_1), f(x), x, x + b, x + b + d_3)$;
- $P_2 := (f(x + b + d_3), f(x + b), f(x + b + d_1))$;
- $P_3 := (x + d_1 + b, x + d_1, x + d_1 + d_2)$.

We say that $CS(x, a, b) := (P_1, P_2, P_3)$ is an (x, a, b)-*consistent system of paths*, and let

$$\mathrm{end}(CS(x, a, b)) := \mathrm{end}(P_1) \cup \mathrm{end}(P_2) \cup \mathrm{end}(P_3).$$

Let (P_1, P_2, P_3) be an (x, a, b)-consistent system of paths. Let $\mathcal{D} := \{d_1, d_2, d_3\} \subseteq \mathcal{D}(L)$ be the set of directions such that P_1, P_2 and P_3 are as defined above. Let $\mathbf{C} := \{C_1, \dots, C_6\}$ be a set of vertex-disjoint ℓ-cubes which satisfy the following:

(PIII.1) for all $i \in [6]$, we have that $C_i \subseteq L$ or $C_i \subseteq f(L)$;

(PIII.2) for all $i \in [3]$, we have $C_{2i} = f(C_{2i+1})$, where indices are taken modulo 6;

(PIII.3) for all $i \in [3]$, C_{2i-1} contains the first vertex of P_i, and C_{2i} contains the last vertex of P_i;

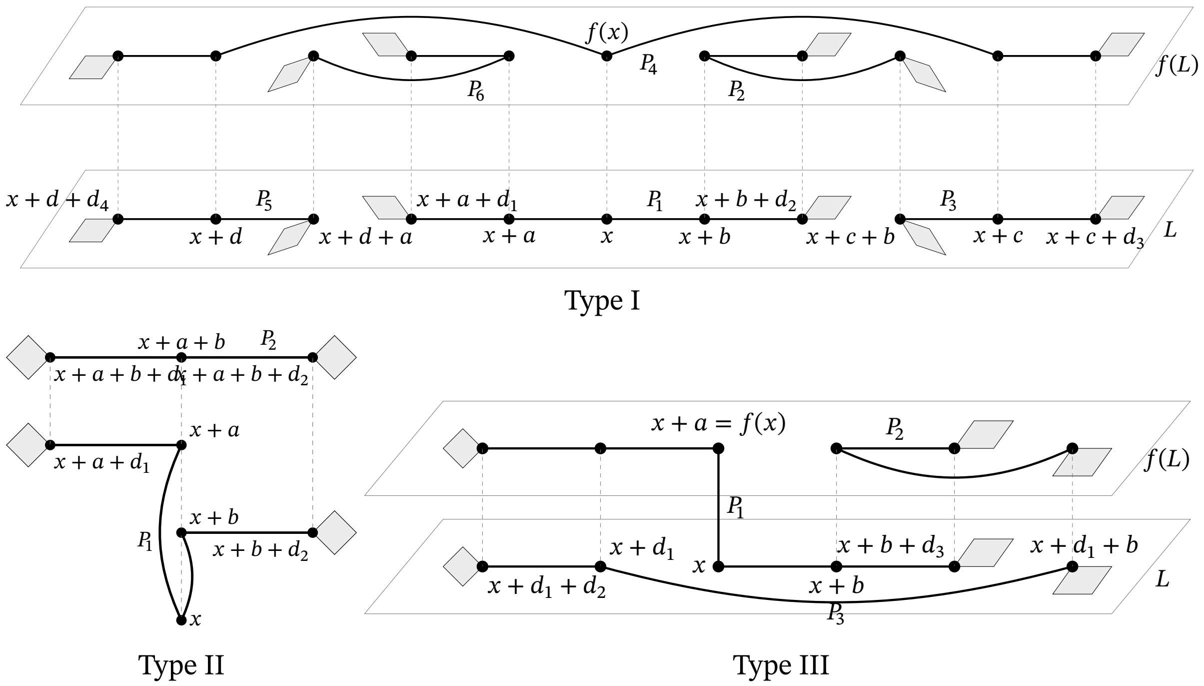

Figure 2.2. A representation of the special absorbing structures. In each case, vertices (or cubes) represented in vertical lines are clones of the same vertex (or cube) of the intersection graph I, and any vertices in the same horizontal line lie in the same layer of $\mathcal{Q}^n$.

(PIII.4) for all $i \in [6]$, we have $\mathcal{D}(C_i) \cap (\mathcal{D} \cup \{b\}) = \varnothing$.

We say that $SA(x,a,b) \coloneqq (P_1,P_2,P_3,C_1,\dots,C_6)$ is an (x,a,b)-*special absorbing structure*.

Finally, given any graph $G \subseteq \mathcal{Q}^n$ and an (x,a,b)-consistent system of paths $CS(x,a,b) = (P_1,P_2,P_3)$, we say that $CS(x,a,b)$ *extends* to an (x,a,b)-special absorbing structure in G if there is a collection $\mathbf{C} = \{C_1,\dots,C_6\}$ with $C_i \subseteq G$ such that $SA(x,a,b) = (P_1,P_2,P_3,C_1,\dots,C_6)$ is an (x,a,b)-special absorbing structure.

Whenever x, a and b are clear from the context, we will simply write CS and SA instead of $CS(x,a,b)$ and $SA(x,a,b)$. Given any consistent system of paths CS, we let $\mathrm{endmol}(CS)$ be the set of vertices $v \in V(I)$ such that some clone of v lies in $\mathrm{end}(CS)$. We write $\mathcal{D}(CS)$ to denote the set of directions $\mathcal{D} \cup \{a,b\}$ used to define the paths which comprise CS as above. If CS extends to a special absorbing structure SA, we denote $\mathrm{end}(SA) \coloneqq \mathrm{end}(CS)$. Moreover, we denote by $\mathbf{C}(SA)$ the collection of cubes associated with SA. Observe that (PI.4), (PII.4) and (PIII.4) imply that

(AS) each cube $C \in \mathbf{C}(SA)$ is vertex-disjoint from the paths in SA except for the unique vertex in $\mathrm{end}(SA)$ contained in C.

We will sometimes abuse notation and treat CS and SA as graphs; in particular, we will write $V(CS)$ to denote the vertices of the union of the paths which comprise CS, and $V(SA)$ to denote the vertices of the union of the paths and cubes which comprise SA, and similarly for $E(CS)$ and $E(SA)$.

2.7.3. Bondless and bondlessly surrounded molecules and robust subgraphs

Before moving on, we need to prove several properties about the structures we have defined so far. Again, due to space limitations, we cannot present the proofs here; we include them in appendix A.2.

Given any graph $G \subseteq \mathcal{Q}^n$, we will say that an (s,ℓ)-molecule $\mathcal{M} = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_{2^s} \subseteq \mathcal{Q}^n$, where $\mathcal{A}_i$ is the i-th clone of some ℓ-cube $\mathcal{A} \subseteq \mathcal{Q}^{n-s}$, is *bonded* in G if, for all $i \in [2^s]$, G contains at least 100 edges between $\mathcal{A}_i$ and $\mathcal{A}_{i+1}$ whose endpoint in $\mathcal{A}_i$ has even parity and at least 100 such edges whose endpoint in $\mathcal{A}_i$ has odd parity. Otherwise, we call it *bondless* in G. Furthermore, given a collection $\mathcal{U}$ of (s,ℓ)-molecules in G, we say that $\mathcal{M} \in \mathcal{U}$ is *bondlessly surrounded* in G (with respect to $\mathcal{U}$) if there exists some vertex $v \in V(\mathcal{M})$ which has at least $n/2^{\ell+5s}$ neighbours in $\mathcal{Q}^n$ which are part of (s,ℓ)-molecules of $\mathcal{U}$ which are bondless in G. Both bondless and bondlessly surrounded molecules create difficulties in applying the rainbow matching lemma (they make it harder to bound the maximum degree of a certain auxiliary graph, which is one of the required conditions to apply lemma 2.5.5), which in turn is used to assign absorption structures to vertices. Therefore, it will become important that we bound the number of each, and show that they are well spread out.

Lemma 2.7.2. *Let $\varepsilon > 0$ and $\ell, s, n \in \mathbb{N}$ be such that $s < n$, $\ell \leq n - s$ and $1/\ell \ll \varepsilon$. Then, for any (s,ℓ)-molecule $\mathcal{M} \subseteq \mathcal{Q}^n$, the probability that it is bondless in $\mathcal{Q}^n_\varepsilon$ is at most $2^{s+1-\varepsilon 2^\ell/4}$.*

When taking random subgraphs of the hypercube, we will need to guarantee that, given a vertex in I and a large collection of cubes in I incident to this vertex, some of the cube molecules given by these cubes are bonded.

Lemma 2.7.3. *Let $\varepsilon, \gamma \in (0,1)$ and $\ell, n \in \mathbb{N}$ with $0 < 1/n \ll 1/\ell \ll \varepsilon, \gamma$, and let $s \coloneqq 10\ell$. Let $x \in V(I)$ and let $\mathcal{C}$ be a collection of ℓ-cubes $C \subseteq I$ such that $|\mathcal{C}| \geq \gamma n^\ell$ and, for all $C \in \mathcal{C}$, we have $x \in V(C)$. For each $C \in \mathcal{C}$, let $\mathcal{M}_C$ denote the cube molecule of C in $\mathcal{Q}^n$. For any graph $G \subseteq \mathcal{Q}^n$, let*

$$B(G) \coloneqq \{C \in \mathcal{C} : \mathcal{M}_C \text{ is bonded in } G\}.$$

Then, with probability at least $1 - 2^{-10n}$, we have $|B(\mathcal{Q}^n_\varepsilon)| \geq \gamma n^\ell/4$.

Lemma 2.7.4. *Let $\varepsilon \in (0,1)$ and $\ell, n \in \mathbb{N}$ with $0 < 1/n \ll 1/\ell \ll \varepsilon$, and let $s \coloneqq 10\ell$. Let $\mathfrak{M}$ be a collection of vertex-disjoint (s,ℓ)-molecules $\mathcal{M} \subseteq \mathcal{Q}^n$. For each $x \in V(\mathcal{Q}^n)$, let $N^{\mathfrak{M}}(x) \coloneqq \{\mathcal{M} \in \mathfrak{M} : \operatorname{dist}(x, \mathcal{M}) = 1\}$. Assume that the following holds for every $x \in V(\mathcal{Q}^n)$:*

(BS) *for any direction $\hat{e} \in \mathcal{D}(\mathcal{Q}^n)$, there are at most $\sqrt{n}$ molecules $\mathcal{M} \in N^{\mathfrak{M}}(x)$ such that $\hat{e} \in \mathcal{D}(\mathcal{A})$ for all atoms $\mathcal{A} \in \mathcal{M}$.*

Then, with probability at least $1 - 2^{-n^{9/8}}$, for every $x \in V(\mathcal{Q}^n)$ we have that $B^{\ell^2}_{\mathcal{Q}^n}(x)$ intersects at most $n^{1/3}$ molecules from $\mathfrak{M}$ which are bondlessly surrounded in $\mathcal{Q}^n_\varepsilon$.

The following observation will also be used repeatedly.

Remark 2.7.5. *Let $n, \ell \in \mathbb{N}$ and $\eta, \eta' \in (0,1)$ with $1/n \ll \eta' < \eta$. Let $x \in V(\mathcal{Q}^n)$. Let $\mathcal{C}$ be a collection of ℓ-cubes $C \subseteq \mathcal{Q}^n$ such that $x \in V(C)$ for all $C \in \mathcal{C}$ and $|\mathcal{C}| \geq \eta n^\ell$. Let $\mathcal{D}' \subseteq \mathcal{D}(\mathcal{Q}^n)$ be a set of directions with $|\mathcal{D}'| \leq \eta' n$. Then, there exists a cube $C \in \mathcal{C}$ with $\mathcal{D}(C) \cap \mathcal{D}' = \varnothing$.*

As discussed in section 2.2.6, a crucial requirement for the proof of theorem 2.1.8 will be that vertices of very low degree in $\mathcal{Q}^n_{1/2-\varepsilon}$ are few and far apart. Moreover, we will also require some more properties about the distribution of these vertices, and that, for all of them, we can find many candidates for special absorbing structures. We express all this information in the following definition.

Definition 2.7.6. Let $n, s, \ell \in \mathbb{N}$ with $1/n \ll 1/s \leq 1/\ell$, and let $\varepsilon_1, \varepsilon_2, \gamma \in [0,1]$. Fix an ordering of the layers $L_1, \ldots, L_{2^s}$ of $\mathcal{Q}^n$ induced by any Hamilton cycle in $\mathcal{Q}^s$ (as defined in section 2.7.1). Let $G \subseteq \mathcal{Q}^n$ be a spanning subgraph. For any $\varepsilon > 0$, let $\mathcal{U}(G, \varepsilon) := \{x \in V(\mathcal{Q}^n) : d_G(x) < \varepsilon n\}$. Let $\mathcal{U} \subseteq V(\mathcal{Q}^n)$ be a set of size $|\mathcal{U}| \leq 2^{\varepsilon_2 n}$. We say that G is $(s, \ell, \varepsilon_1, \varepsilon_2, \gamma, \mathcal{U})$-*robust* if the following properties are satisfied:

(R1) $\mathcal{U}(G, \varepsilon_1) \subseteq \mathcal{U}$.

(R2) For all $x \in \mathcal{U}$ and every $y \in B^{s+5\ell}_{\mathcal{Q}^n}(x) \setminus \{x\}$, we have $d_G(y) \geq \gamma n$.

(R3) For all $x \in V(\mathcal{Q}^n)$, we have $|\mathcal{U} \cap B^{\gamma n}_{\mathcal{Q}^n}(x)| \leq 1$.

(R4) For all $x \in \mathcal{U}$ and any distinct directions $a, b \in \mathcal{D}(\mathcal{Q}^n)$, there exists a collection $\mathfrak{C}(x, a, b)$ of (x, a, b)-consistent systems of paths in $G \cup \{\{x, x+a\}, \{x, x+b\}\}$ which satisfies the following. Let L be the layer containing x.

(R4.I) Suppose $x + a, x + b \in V(L)$. Then, there exists a collection $\mathcal{D}^{(2)}(x, a, b)$ of disjoint pairs of distinct directions $c, d \in \mathcal{D}(L) \setminus \{a, b\}$ such that $|\mathcal{D}^{(2)}(x, a, b)| \geq \gamma n$ and, for every $(c, d) \in \mathcal{D}^{(2)}(x, a, b)$, there is a collection $\mathcal{D}^{(4)}(x, a, b, c, d)$ of disjoint 4-tuples of distinct directions in $\mathcal{D}(L) \setminus \{a, b, c, d\}$ with $|\mathcal{D}^{(4)}(x, a, b, c, d)| \geq \gamma n$ satisfying the following property: for each $(c, d) \in \mathcal{D}^{(2)}(x, a, b)$ and each $(d_1, d_2, d_3, d_4) \in \mathcal{D}^{(4)}(x, a, b, c, d)$, the (x, a, b)-consistent system of paths $CS(c, d, d_1, d_2, d_3, d_4) = (P_1, \ldots, P_6)$ defined as in section 2.7.2 belongs to $\mathfrak{C}(x, a, b)$.

(R4.II) Suppose $x + a, x + b \notin V(L)$. Then, there exists a collection $\mathcal{D}^{(2)}(x, a, b)$ of disjoint pairs of distinct directions $d_1, d_2 \in \mathcal{D}(L)$ such that $|\mathcal{D}^{(2)}(x, a, b)| \geq \gamma n$ and, for every $(d_1, d_2) \in \mathcal{D}^{(2)}(x, a, b)$, the (x, a, b)-consistent system of paths $CS(d_1, d_2) = (P_1, P_2)$ defined as in section 2.7.2 belongs to $\mathfrak{C}(x, a, b)$.

(R4.III) Suppose $x + a \notin V(L)$ and $x + b \in V(L)$. Then, there exists a set $\mathcal{D}(x, a, b)$ of directions $d_1 \in \mathcal{D}(L)$ such that $|\mathcal{D}(x, a, b)| \geq \gamma n$ and, for every $d_1 \in \mathcal{D}(x, a, b)$, there exists a collection $\mathcal{D}^{(2)}(x, a, b, d_1)$ of disjoint pairs of distinct directions in $\mathcal{D}(L) \setminus \{b, d_1\}$ with $|\mathcal{D}^{(2)}(x, a, b, d_1)| \geq \gamma n$ satisfying the following property:

for each $d_1 \in \mathcal{D}^{(2)}(x,a,b)$ and each $(d_2,d_3) \in \mathcal{D}^{(2)}(x,a,b,d_1)$, the (x,a,b)-consistent system of paths $CS(d_1,d_2,d_3) = (P_1,P_2,P_3)$ defined as in section 2.7.2 belongs to $\mathfrak{C}(x,a,b)$.

(R5) Let $x_1 \coloneqq \{0\}^n$, $x_2 \coloneqq \{1\}^n$, $x_3 \coloneqq \{1\}^{\lceil n/2\rceil}\{0\}^{n-\lceil n/2\rceil}$ and $x_4 \coloneqq \{0\}^{\lceil n/2\rceil}\{1\}^{n-\lceil n/2\rceil}$. Then, for each $i \in [4]$ we have $\mathcal{U} \cap B^{s+\ell}_{\mathcal{Q}^n}(x_i) = \varnothing$.

Lemma 2.7.7. *Let $1/n \ll 1/s \leq 1/\ell \ll \varepsilon_1 \ll \varepsilon \ll \varepsilon_2 \ll \gamma \ll 1/r$ with $n,s,\ell,r \in \mathbb{N}$. Then,*

(i) a.a.s. $G \sim \mathcal{Q}^n_{1/2-\varepsilon}$ is $(s,\ell,\varepsilon_1,\varepsilon_2,\gamma,\mathcal{U}(G,\varepsilon_1))$-robust, and

(ii) given any $\mathcal{U} \subseteq V(\mathcal{Q}^n)$ with $|\mathcal{U}| \leq 2^{\varepsilon_2 n}$ and any $H \subseteq \mathcal{Q}^n$ which is $(s,\ell,\varepsilon_1,\varepsilon_2,\gamma,\mathcal{U})$-robust, there exists an edge-decomposition $H = H_1 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} H_r$ such that for each $i \in [r]$ we have that $H_i \subseteq H$ is spanning and $(s,\ell,\varepsilon_1/(2r),\varepsilon_2,\gamma/r^{10},\mathcal{U})$-robust.

Finally, we need that "scant" molecules are not too clustered. (We will later define a vertex molecule as "scant" –with respect to a graph H and a reservoir R– if one of its vertices v_i has the property that few of its neighbours lie in the i-th clone of R.) Recall that $Res(\mathcal{Q}^n,\delta)$ was defined in section 2.5.3.

Lemma 2.7.8. *Let $0 < 1/n \ll 1/C \ll \varepsilon_1,\varepsilon_2 \ll \gamma,\delta \leq 1$ and $1/n \ll 1/s \leq 1/\ell$, where $n,C,s,\ell \in \mathbb{N}$. Let $H \subseteq \mathcal{Q}^n$ and $\mathcal{U} \subseteq V(\mathcal{Q}^n)$ be such that H is $(s,\ell,\varepsilon_1,\varepsilon_2,\gamma,\mathcal{U})$-robust. Let $\mathcal{U}_I \subseteq V(I)$ be the set of vertices $u \in V(I)$ such that $\mathcal{U}$ contains some clone of u. For each $v \in V(I)$ and each $i \in [2^s]$, let v_i be the i-th clone of v, and let $\mathcal{M}_v \coloneqq \{v_i : i \in [2^s]\}$. Let $R \sim Res(I,\delta)$ and, for each $i \in [2^s]$, let R_i be the i-th clone of R. Let*

$$B \coloneqq \{v \in V(I) : \text{there exists } i \in [2^s] \text{ with } v_i \notin \mathcal{U} \text{ and } e_H(v_i,R_i) < \varepsilon_1\delta n/4\}.$$

Let $\mathcal{E}_1$ be the event that there exists some $u \in V(I)$ such that $|B^{10\ell}_I(u) \cap B| \geq C$. Let $\mathcal{E}_2$ be the event that there exists some $u \in \mathcal{U}_I$ such that $|B^{5\ell}_I(u) \cap B| \geq 1$. Then, $\mathbb{P}[\mathcal{E}_1 \vee \mathcal{E}_2] \leq 1/n$.

2.7.4. Connecting cubes

The hypercube satisfies some robust connectivity properties. The problem of (almost) covering $\mathcal{Q}^n$ with disjoint paths has been extensively studied.

In order to create a long cycle, which can be used to absorb all remaining vertices, while preserving the absorbing structure, we will make use of the robust connectivity properties of the hypercube. In particular, we will need several results which guarantee that, given any prescribed pairs of vertices in a slice, there is a spanning collection of vertex-disjoint paths, each of which uses the vertices of one of the given pairs as endpoints. We will also need similar results for almost spanning collections of paths, where these paths avoid a given prescribed vertex. Throughout this subsection we denote by uv the edge between two given adjacent vertices u and v (instead of $\{u, v\}$).

The following lemma will be essential for us. It follows from some results of Dvořák and Gregor [44, Corollary 5.2].

Lemma 2.7.9. *For all $n \geq 100$, the graph $\mathcal{Q}^n$ satisfies the following.*

(i) *Let $m \in [25]$ and let $\{u_i, v_i\}_{i\in[m]}$ be disjoint pairs of vertices with $u_i \neq_{\mathrm{p}} v_i$ for all $i \in [m]$. Then, there exist m vertex-disjoint paths $\mathcal{P}_1, \dots, \mathcal{P}_m \subseteq \mathcal{Q}^n$ such that, for each $i \in [m]$, $\mathcal{P}_i$ is a (u_i, v_i)-path, and $\bigcup_{i\in[m]} V(\mathcal{P}_i) = V(\mathcal{Q}^n)$.*

(ii) *Let $x \in V(\mathcal{Q}^n)$. Let $m \in [25]$ and let $\{u_i, v_i\}_{i\in[m]}$ be disjoint pairs of vertices of $\mathcal{Q}^n - \{x\}$ such that $u_1, v_1 \neq_{\mathrm{p}} x$ and $u_i \neq_{\mathrm{p}} v_i$ for all $i \in [m] \setminus \{1\}$. Then, there exist m vertex-disjoint paths $\mathcal{P}_1, \dots, \mathcal{P}_m \subseteq \mathcal{Q}^n$ such that, for each $i \in [m]$, $\mathcal{P}_i$ is a (u_i, v_i)-path, and $\bigcup_{i\in[m]} V(\mathcal{P}_i) = V(\mathcal{Q}^n) \setminus \{x\}$.*

(iii) *Let $\{u_i, v_i\}_{i\in[2]}$ be disjoint pairs of vertices with $u_i =_{\mathrm{p}} v_i$ for all $i \in [2]$ and $u_1 \neq_{\mathrm{p}} u_2$. Then, there exist two vertex-disjoint paths $\mathcal{P}_1, \mathcal{P}_2 \subseteq \mathcal{Q}^n$ such that, for each $i \in [2]$, $\mathcal{P}_i$ is a (u_i, v_i)-path, and $V(\mathcal{P}_1) \cup V(\mathcal{P}_2) = V(\mathcal{Q}^n)$.*

We now motivate the statement (as well as the proof) of lemma 2.7.11, which is the main result of this subsection. We are given a slice $\mathcal{M}^*$ of a molecule $\mathcal{M} \subseteq \mathcal{Q}^n$ which is bonded in a graph $G \subseteq \mathcal{Q}^n$. Furthermore, we are given collections of vertices L, R (which are part of absorbing cube structures), and S (which, when constructing a long cycle, will be used to enter and leave $\mathcal{M}^*$). More specifically, we have that

- L will have size 0 or 2, and will consist of left absorber tips. If it has size 2, the vertices will have opposite parities. These must be avoided by our connecting paths, so that we can make use of the absorbing structures we have put in place (see the discussion in section 2.2).
- R will consist of the pairs of right absorber tip and third absorber vertex. These must be connected via an edge with the paths we find.
- S will consist of a set of pairs of vertices $\{u, v\}$ with $u \neq_{\mathrm{p}} v$. Later, when creating a long cycle, u will be a vertex through which we enter $\mathcal{M}^*$ from a different molecule, and v will be the next vertex from which we leave $\mathcal{M}^*$ (with respect to some ordering). Each of our paths will be a (u, v)-path, for some such pair $\{u, v\}$.

In order to find our paths, we will call on lemma 2.7.9. To illustrate this, suppose $\mathcal{M}^*$ consists of the atoms $\mathcal{A}_1, \dots, \mathcal{A}_t$, for some $t \in \mathbb{N}$. Suppose that $S = \{u, v\}$ with $u \in V(\mathcal{A}_1)$ and $v \in V(\mathcal{A}_t)$. Furthermore, suppose that $L, R = \varnothing$. To construct a path from u to v, we will first specify the edges used to pass between different atoms. For all $k \in [t-1]$, we choose an edge $v_k^{\uparrow} u_{k+1}^{\uparrow}$ from $\mathcal{A}_k$ to $\mathcal{A}_{k+1}$, thus $v_k^{\uparrow} \neq_{\mathrm{p}} u_{k+1}^{\uparrow}$. For technical reasons, we aim to have all the vertices $u_{k+1}^{\uparrow}$ of the same parity as u. We can then apply lemma 2.7.9 to find a path from $u_{k+1}^{\uparrow}$ to $v_{k+1}^{\uparrow}$ which covers all of $V(\mathcal{A}_{k+1})$. Together with the edges $v_k^{\uparrow} u_{k+1}^{\uparrow}$, all these paths will form a single path from u to v which spans $V(\mathcal{M}^*)$. In the more general setting where $u \in V(\mathcal{A}_i)$ and $v \in V(\mathcal{A}_j)$ with $1 < i < j < t$, the (u, v)-path we construct would first pass down to $\mathcal{A}_1$, then up to $\mathcal{A}_t$ and, finally, back down to $\mathcal{A}_j$.

When $L \neq \varnothing$, due to vertex parities, the following issue can arise. Suppose $L = \{x, y\}$ with $x \in V(\mathcal{A}_1)$, $u \in V(\mathcal{A}_2)$, $y \in V(\mathcal{A}_3)$ and $v \in V(\mathcal{A}_j)$ for some $j > 3$ (and $R = \varnothing$). Furthermore, suppose that both u and x have odd parity. In line with the above description, the vertex $u_1^{\downarrow}$, through which we enter $\mathcal{A}_1$, would have odd parity. It follows that, since x also has odd parity, we cannot hope to construct a path which starts at $u_1^{\downarrow}$ and covers all of $V(\mathcal{A}_1) \setminus \{x\}$. The solution will be instead to pass up to $\mathcal{A}_3$ first (and, in general, to whichever atom contains y). Recall that, since x has odd parity, y must have even parity. We specify a vertex $u_3^{\uparrow}$ of odd parity, through which we enter $\mathcal{A}_3$, but then also specify a vertex $v_3^{\downarrow}$ of odd parity from which we will leave $\mathcal{A}_3$ to reenter $\mathcal{A}_2$. We now arrive back in $\mathcal{A}_2$ with a vertex $u_2^{\downarrow}$ of even parity. We will specify another vertex $v_2^{\downarrow}$ of odd parity from which we leave $\mathcal{A}_2$ and a vertex $u_1^{\downarrow}$ of even parity through which we enter $\mathcal{A}_1$. In this way, we can now apply lemma 2.7.9 to find a path which starts at $u_1^{\downarrow}$ and covers all of $V(\mathcal{A}_1) \setminus \{x\}$, and which can be extended into a path from u to v covering all of $V(\mathcal{M}^*) \setminus L$.

There are several other instances which must be dealt with in a similar way. This is formalised by lemma 2.7.11. Before proving this lemma, however, we need the following definition.

Definition 2.7.10 ((u, j, F, R)-alternating parity sequence)**.** Let $\ell, s, t, n \in \mathbb{N}$ with $t \leq 2^s$ and $2 \leq \ell \leq n - s$. Let $G \subseteq \mathcal{Q}^n$. Let $\mathcal{M} = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_{2^s} \subseteq \mathcal{Q}^n$ be an (s, ℓ)-molecule and let $\mathcal{M}^* = \mathcal{A}_{a+1} \cup \cdots \cup \mathcal{A}_{a+t}$, for some $a \in [2^s]$, be a slice of $\mathcal{M}$. Let $u \in V(\mathcal{A}_i)$, for some $i \in [a + t] \setminus [a]$. Let $j \in [a + t] \setminus [a]$, and let $F, R \subseteq V(\mathcal{M}^*)$. Suppose $i \leq j$. Let $I_R := \{k \in [j - i]_0 : |R \cap V(\mathcal{A}_{i+k})| \geq 1\}$. Assume that the following properties hold:

- For all $k \in [j - i]_0$ we have that $|R \cap V(\mathcal{A}_{i+k})| \in \{0, 2\}$.
- For each $k \in I_R$, the vertices in $R \cap V(\mathcal{A}_{i+k})$ are adjacent in $\mathcal{Q}^n$, and we write $R \cap V(\mathcal{A}_{i+k}) = \{w_k, z_k\}$ so that $w_k \neq_{\mathrm{p}} u$.

Let $\mathcal{S}' = (u_0, v_1, u_1, \dots, v_{j-i}, u_{j-i})$ be a sequence of vertices satisfying the following properties:

(P0) If $u \in R$, then $u_0 := w_0$; otherwise, $u_0 := u$.

(P1) For each $k \in [j - i]$ we have that $u_k =_{\mathrm{p}} u$.

(P2) For each $k \in [j - i]$ we have that $v_k \in V(\mathcal{A}_{i+k-1})$, $u_k \in V(\mathcal{A}_{i+k})$ and $v_k u_k \in E(G)$.

(P3) The vertices of $\mathcal{S}'$ other than u_0 avoid $F \cup R$.

A (u, j, F, R)*-alternating parity sequence* $\mathcal{S}$ *in* G is a sequence obtained from any sequence $\mathcal{S}'$ which satisfies (P0)–(P3) as follows. For each $k \in I_R \cap [j-i]$, replace each segment (v_k, u_k) of $\mathcal{S}'$ by (v_k, u_k, w_k, z_k).

The case $i > j$ is defined similarly by replacing each occurrence of $[j-i]$ and $[j-i]_0$ in the above by $[i-j]$ and $[i-j]_0$, and each occurrence of $\mathcal{A}_{i+k}$ and $\mathcal{A}_{i+k-1}$ by $\mathcal{A}_{i-k}$ and $\mathcal{A}_{i-k+1}$.

Given an alternating parity sequence $\mathcal{S}$, we will denote by $\mathcal{S}^-$ the sequence obtained from $\mathcal{S}$ by deleting its initial element.

Lemma 2.7.11. *Let $n, s, \ell \in \mathbb{N}$ be such that $s \geq 4$ and $100 \leq \ell \leq n-s$. Let $G \subseteq \mathcal{Q}^n$ and consider any (s, ℓ)-molecule $\mathcal{M} = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_{2^s} \subseteq \mathcal{Q}^n$ which is bonded in G. Let $\mathcal{M}^* = \mathcal{A}_{a+1} \cup \cdots \cup \mathcal{A}_{a+t}$, for some $a \in [2^s]$ and $t \geq 10$, be a slice of $\mathcal{M}$. Moreover, consider the following sets.*

(C1) *Let $L \subseteq V(\mathcal{M}^*)$ be a set of size $|L| \in \{0, 2\}$ such that, if $L = \{x, y\}$, then $x \in V(\mathcal{A}_i)$ and $y \in V(\mathcal{A}_j)$ with $i \neq j$ and $x \neq_{\mathrm{p}} y$.*

(C2) *Let $R \subseteq V(\mathcal{M}^*) \setminus L$ be a (possibly empty) set of vertices with $|R| \leq 10$ such that, for all $k \in [a+t] \setminus [a]$, we have $|R \cap V(\mathcal{A}_k)| \in \{0, 2\}$ and, if $|R \cap V(\mathcal{A}_k)| = 2$, then $R \cap V(\mathcal{A}_k) = \{w_k, z_k\}$ satisfies that $w_k z_k \in E(\mathcal{M}^*)$ and, if $|L| = 2$, then $k \notin \{i, j\}$.*

(C3) *Let $m \in [14]$ and consider m vertex-disjoint pairs $\{u_r, v_r\}_{r \in [m]}$, where $u_r, v_r \in V(\mathcal{M}^*) \setminus L$ and $u_r \neq_{\mathrm{p}} v_r$ for all $r \in [m]$, such that, for each $r \in [m]$, we have $u_r \in V(\mathcal{A}_{i_r})$ and $v_r \in V(\mathcal{A}_{j_r})$. Assume, furthermore, that for each $t' \in [t]$ we have that $|\bigcup_{r \in [m]} \{u_r, v_r\} \cap V(\mathcal{A}_{a+t'}) \cap R| \leq 1$.*

Then, there exist vertex-disjoint paths $\mathcal{P}_1, \ldots, \mathcal{P}_m \subseteq \mathcal{M}^* \cup G$ such that, for each $r \in [m]$, $\mathcal{P}_r$ is a (u_r, v_r)-path, $\bigcup_{r \in [m]} V(\mathcal{P}_r) = V(\mathcal{M}^*) \setminus L$, and every pair $\{w_k, z_k\}$ with $k \in [a+t] \setminus [a]$ is an edge of some $\mathcal{P}_r$.

Proof. By relabelling the atoms, we may assume that $\mathcal{M}^* = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_t$. Let $S := \{u_r, v_r : r \in [m]\}$. By relabelling the vertices, we may assume that $i_r \leq j_r$ for all $r \in [m]$ and (if

$L \neq \varnothing$) $i < j$. Let $I_L \coloneqq \{k \in [t] : L \cap V(\mathcal{A}_k) \neq \varnothing\}$, $I_R \coloneqq \{k \in [t] : R \cap V(\mathcal{A}_k) \cap S \neq \varnothing\}$ and $R^* \coloneqq R \setminus \bigcup_{k \in I_R} V(\mathcal{A}_k)$. Note that $I_L = \varnothing$ or $I_L = \{i, j\}$ and $I_L \cap I_R = \varnothing$. For each $r \in [m]$, let $I_R^r \coloneqq \{k \in \{i_r, j_r\} : R \cap V(\mathcal{A}_k) \cap \{u_r, v_r\} \neq \varnothing\}$, so that $I_R = \bigcup_{r=1}^m I_R^r$. Without loss of generality, we may also assume that, for each $r \in [m]$, if $u_r \in R$, then $u_r = z_{i_r}$, and if $v_r \in R$, then $v_r = w_{j_r}$. Similarly, for each $k \in [t] \setminus I_R$, if $R \cap V(\mathcal{A}_k) = \{w_k, z_k\}$, we may assume that $w_k \neq_{\mathrm{p}} u_1$.

For each $r \in [m]$, we will create a list $\mathcal{L}_r$ of vertices. We will refer to $\mathcal{L}_r$ as the *skeleton* for $\mathcal{P}_r$. We will later use these skeletons to construct the vertex-disjoint paths via lemma 2.7.9. For each $r \in [m]$, we will write L_r^* for the (unordered) set of vertices in $\mathcal{L}_r$. In order to construct each $\mathcal{L}_r$, we will start with an empty list and update it in (possibly) several steps, by concatenating alternating parity sequences. Whenever $\mathcal{L}_r$ is updated, we implicitly update L_r^*. In the end, for each $r \in [m]$ we will have a list of vertices $\mathcal{L}_r = (x_1^r, \dots, x_{\ell_r}^r)$. For each $r \in [m]$ and $k \in [t]$, let $I_r(k) \coloneqq \{h \in [\ell_r - 1] : 2 \nmid h \text{ and } x_h^r, x_{h+1}^r \in V(\mathcal{A}_k)\}$. We will require the $\mathcal{L}_r$ to be pairwise vertex-disjoint. Furthermore, we will require that they satisfy the following properties:

($\mathcal{L}$1) For all $r \in [m]$ we have that ℓ_r is even.

($\mathcal{L}$2) For all $r \in [m]$ and $h \in [\ell_r - 1]$, if h is odd, then $x_h^r, x_{h+1}^r \in V(\mathcal{A}_k)$, for some $k \in [t]$; if h is even, then $x_h^r x_{h+1}^r \in E(G \cup \mathcal{M}^*)$.

($\mathcal{L}$3) For all $k \in [t]$ we have that $1 \leq |I_1(k)| \leq 6$ and $|I_r(k)| \leq 1$ for all $r \in [m] \setminus \{1\}$.

($\mathcal{L}$4)$_1$ For each $k \in [t] \setminus (I_L \cup I_R^1)$ and each $h \in I_1(k)$, we have $x_h^1 \neq_{\mathrm{p}} x_{h+1}^1$. For each $k \in I_L \cup I_R^1$, for all but one $h \in I_1(k)$ we have $x_h^1 \neq_{\mathrm{p}} x_{h+1}^1$, while for the remaining index $h \in I_1(k)$ we have that $x_h^1 =_{\mathrm{p}} x_{h+1}^1$ and their parity is opposite to that of the unique vertex in $L \cap V(\mathcal{A}_k)$ if $k \in I_L$ and to that of the unique vertex in $\{w_k, z_k\} \cap \{u_1, v_1\}$ if $k \in I_R^1$.

($\mathcal{L}$4)$_r$ For each $r \in [m] \setminus \{1\}$, the following holds. For each $k \in [t] \setminus I_R^r$ and each $h \in I_r(k)$, we have $x_h^r \neq_{\mathrm{p}} x_{h+1}^r$. For each $k \in I_R^r$, for all but one $h \in I_r(k)$ we have $x_h^r \neq_{\mathrm{p}} x_{h+1}^r$, while for the remaining index $h \in I_r(k)$ we have that $x_h^r =_{\mathrm{p}} x_{h+1}^r$ and their parity is opposite to that of the unique vertex in $\{w_k, z_k\} \cap \{u_r, v_r\}$.

($\mathcal{L}5$) For each $r \in [m]$, we have the following. If $u_r \notin R$, then $u_r = x_1^r$. If $v_r \notin R$, then $v_r = x_{\ell_r}^r$. If $u_r \in R$ (and thus $u_r = z_{i_r}$), then $w_{i_r} = x_1^r$ and $u_r \notin L_1^* \cup \dots \cup L_m^*$. If $v_r \in R$ (and thus $v_r = w_{j_r}$), then $z_{j_r} = x_{\ell_r}^r$ and $v_r \notin L_1^* \cup \dots \cup L_m^*$.

($\mathcal{L}6$) Every pair (w_k, z_k) with $\{w_k, z_k\} \subseteq R^*$ is contained in $\mathcal{L}_1$ and z_k directly succeeds w_k.

We begin by constructing $\mathcal{L}_1$. Let $\mathcal{L}_1 := \varnothing$ and let $F := L \cup R \cup S$. If $i_1 = 1$ and $R^* \cap V(\mathcal{A}_1) = \{w_1, z_1\}$, then let $\mathcal{S}_1 := (u_1, w_1, z_1)$. If $i_1 = 1$ and $u_1 \in R$, then let $\mathcal{S}_1 := (u_1)$. Otherwise, let $\mathcal{S}_1$ be a $(u_1, 1, F, (R \cap V(\mathcal{A}_{i_1})) \cup (R^* \cap V(\mathcal{A}_1)))$-alternating parity sequence. Let $\mathcal{L}_1 := \mathcal{S}_1$. Note that the existence of such a sequence $\mathcal{S}_1$ is guaranteed by our assumption that $\mathcal{M}$ is bonded in G. To see this, note that all edges of G required by $\mathcal{S}_1$ (that is, the pairs $\{v_k, u_k\}$ in definition 2.7.10) need to be chosen so that they do not have an endpoint in F; given any particular pair of consecutive atoms, this forbids at most 30 edges between these two atoms (26 because of S and 4 because of $L \cup R$).

We will now update $\mathcal{L}_1$. While doing so, we will update F and consider several alternating parity sequences. The existence of each of these follows a similar argument to the above. For any given pair of consecutive atoms, every time we update F, the set of forbidden edges will increase its size by at most 3. We will update F at most four times, so F will forbid at most 42 edges between any pair of consecutive atoms. Thus, by the definition of bondedness, each of the alternating parity sequences required below actually exists.

Let $u_1^\downarrow$ be the last vertex in $\mathcal{L}_1$. Note that $u_1^\downarrow =_{\mathrm{p}} u_1$ by definition 2.7.10(P1). We update F as $F := F \cup L_1^*$. For the next step in the construction of $\mathcal{L}_1$, there are three cases to consider, depending on the size of L and, if $|L| = 2$, the relative parities of x and u_1. If $i_1 = 1$ and $u_1 \in R$, let $R^\diamond := R^* \cup \{w_1, z_1\}$; otherwise, let $R^\diamond := R^*$.

Case 1: $L = \varnothing$.

Let $\mathcal{S}_2$ be a $(u_1^\downarrow, t, F, R^\diamond)$-alternating parity sequence. If $i_1 = 1$ and $u_1 \in R$, update $\mathcal{L}_1$ as $\mathcal{L}_1 := \mathcal{S}_2$. Otherwise, update $\mathcal{L}_1$ as $\mathcal{L}_1 := \mathcal{L}_1 \mathcal{S}_2^-$. Update $F := F \cup L_1^*$.

Case 2: $|L| = 2$ and $x \neq_{\mathrm{p}} u_1$.

Let $\mathcal{S}_2$ be a $(u_1^\downarrow, i, F, R^\diamond)$-alternating parity sequence. If $i_1 = 1$ and $u_1 \in R$, update $\mathcal{L}_1$ as $\mathcal{L}_1 := \mathcal{S}_2$. Otherwise, update $\mathcal{L}_1 := \mathcal{L}_1 \mathcal{S}_2^-$. Update $F := F \cup L_1^*$. Choose any vertex $u_i^* \in V(\mathcal{A}_i)$

with $u_i^* \neq_{\mathrm{p}} u_1$, and let $\mathcal{S}_3$ be a $(u_i^*, j, F, R^\diamond)$-alternating parity sequence. Update $\mathcal{L}_1 \coloneqq \mathcal{L}_1\mathcal{S}_3^-$ and $F \coloneqq F \cup L_1^*$. Let v^- be the final vertex of $\mathcal{S}_2$, and let v^+ be the second vertex of $\mathcal{S}_3$. Note that v^- and v^+ appear consecutively in $\mathcal{L}_1$ and that $v^- =_{\mathrm{p}} v^+ =_{\mathrm{p}} u_1 \neq_{\mathrm{p}} x$. Finally, choose any vertex $u_j^* \in V(\mathcal{A}_j)$ with $u_j^* =_{\mathrm{p}} u_1$, let $\mathcal{S}_4$ be a $(u_j^*, t, F, R^\diamond)$-alternating parity sequence, and update $\mathcal{L}_1 \coloneqq \mathcal{L}_1\mathcal{S}_4^-$ and $F \coloneqq F \cup L_1^*$. Let w^- be the final vertex of $\mathcal{S}_3$, and let w^+ be the second vertex of $\mathcal{S}_4$. We then have that w^- and w^+ appear consecutively in $\mathcal{L}_1$, and $w^- =_{\mathrm{p}} w^+ \neq_{\mathrm{p}} y, u_1$. Moreover, the final vertex $u_t^\uparrow$ of $\mathcal{L}_1$ satisfies $u_t^\uparrow =_{\mathrm{p}} u_1$.

Case 3: $|L| = 2$ and $x =_{\mathrm{p}} u_1$.

Let $\mathcal{S}_2$ be a $(u_1^\downarrow, j, F, R^\diamond)$-alternating parity sequence. If $i_1 = 1$ and $u_1 \in R$, update $\mathcal{L}_1$ as $\mathcal{L}_1 \coloneqq \mathcal{S}_2$; otherwise, update $\mathcal{L}_1 \coloneqq \mathcal{L}_1\mathcal{S}_2^-$. Update $F \coloneqq F \cup L_1^*$. Next, let $u_j^* \in V(\mathcal{A}_j)$ be a vertex with $u_j^* \neq_{\mathrm{p}} u_1$ and let $\mathcal{S}_3$ be a $(u_j^*, i, F, \varnothing)$-alternating parity sequence. Update $\mathcal{L}_1 \coloneqq \mathcal{L}_1\mathcal{S}_3^-$ and $F \coloneqq F \cup L_1^*$. Finally, let $u_i^* \in V(\mathcal{A}_i)$ be a vertex with $u_i^* =_{\mathrm{p}} u_1$ and let $\mathcal{S}_4$ be a $(u_i^*, t, F, R^* \cap \bigcup_{k=j+1}^{t} V(\mathcal{A}_k))$-alternating parity sequence. Update $\mathcal{L}_1 \coloneqq \mathcal{L}_1\mathcal{S}_4^-$ and $F \coloneqq F \cup L_1^*$.

In each of the three cases, let $u_t^\uparrow$ denote the last vertex in $\mathcal{L}_1$. Note that, by definition 2.7.10(P1), we have $u_t^\uparrow =_{\mathrm{p}} u_1$, and recall that $v_1 \neq_{\mathrm{p}} u_1$. Let $\mathcal{S}_5$ be a $(u_t^\uparrow, j_1, F, \varnothing)$-alternating parity sequence. Update $\mathcal{L}_1 \coloneqq \mathcal{L}_1\mathcal{S}_5^-$. Again by definition 2.7.10(P1), we have that the final vertex u^* of $\mathcal{L}_1$ is such that $u^* =_{\mathrm{p}} u_t^\uparrow =_{\mathrm{p}} u_1 \neq_{\mathrm{p}} v_1$. Finally, if $v_1 \in R$, update $\mathcal{L}_1 \coloneqq \mathcal{L}_1(z_{j_1})$; otherwise, update it as $\mathcal{L}_1 \coloneqq \mathcal{L}_1(v_1)$. Observe that $\mathcal{L}_1$ satisfies $(\mathcal{L}1)$–$(\mathcal{L}3)$, $(\mathcal{L}4)_1$, $(\mathcal{L}5)$ and $(\mathcal{L}6)$ for the case $r = 1$ by construction.

We now construct $\mathcal{L}_r$ for all $r \in [m] \setminus \{1\}$. For each $r \in [m] \setminus \{1\}$, we proceed iteratively as follows. Let $\mathcal{L}_r \coloneqq \varnothing$ and $F_r \coloneqq L \cup R \cup S \cup \bigcup_{r' \in [r-1]} L_{r'}^*$. Let $\mathcal{S}^r$ be a $(u_r, j_r, F_r, R \cap V(\mathcal{A}_{i_r}))$-alternating parity sequence and update $\mathcal{L}_r$ as $\mathcal{L}_r \coloneqq \mathcal{S}^r$. If $v_r \in R$, update $\mathcal{L}_r \coloneqq \mathcal{L}_r(z_{j_r})$; otherwise, update $\mathcal{L}_r \coloneqq \mathcal{L}_r(v_r)$. Note that each sequence $\mathcal{S}^r$ requires the existence of at most one edge of G, which has to avoid F_r, between any pair of consecutive atoms of $\mathcal{M}^*$. In a similar way to what was discussed above, at most three choices of such edges can be forbidden every time we add a new alternating parity sequence to F. Since for each $r \in [m] \setminus \{1\}$ we consider one new sequence, by the time we consider F_m we have increased the number of forbidden edges by at most $3(m-1) \leq 39$. This gives a total of at most 81 forbidden edges and, thus, the existence of the sequences $\mathcal{S}^r$ is guaranteed by the assumption that $\mathcal{M}$ is bonded in G. Moreover, the lists $\mathcal{L}_1, \dots, \mathcal{L}_r$ now satisfy ($\mathcal{L}$1)–($\mathcal{L}$6).

We are now in a position to apply lemma 2.7.9. For each $k \in [t]$, let $t_k := \sum_{r\in[m]} |I_r(k)|$. Furthermore, for any $r \in [m]$ and $k \in [t]$, for each $h \in I_r(k)$, we refer to the pair x^r_h, x^r_{h+1} as a *matchable pair*. By ($\mathcal{L}$3), $(\mathcal{L}4)_1$, $(\mathcal{L}4)_r$ and lemma 2.7.9(i), each atom $\mathcal{A}_k$ with $k \in [t]\setminus(I_L \cup I_R)$ can be covered by t_k vertex-disjoint paths, each of whose endpoints are a matchable pair contained in $\mathcal{A}_k$. Similarly, by ($\mathcal{L}$3), $(\mathcal{L}4)_1$, $(\mathcal{L}4)_r$ and lemma 2.7.9(ii), each atom $\mathcal{A}_k$ with $k \in I_L \cup I_R$ contains t_k vertex-disjoint paths, each of whose endpoints are a matchable pair in $\mathcal{A}_k$ such that the union of these t_k paths covers precisely $V(\mathcal{A}_k) \setminus (L \cup (S \cap R))$. (Recall that by (C2) and (C3) the set $V(\mathcal{A}_k) \cap (L \cup (S \cap R))$ consists of a single vertex if $k \in I_L \cup I_R$.) For each matchable pair x^r_h, x^r_{h+1} in $\mathcal{A}_k$, let us denote the corresponding path by $\mathcal{P}_{x^r_h, x^r_{h+1}}$.

The paths $\mathcal{P}_1, \dots, \mathcal{P}_m$ required for lemma 2.7.11 can now be constructed as follows. For each $r \in [m]$, let $\mathcal{P}_r$ be the path obtained from the concatenation of the paths $\mathcal{P}_{x^r_h, x^r_{h+1}}$, for each odd $h \in [\ell_r]$, via the edges $x^r_h x^r_{h+1}$ for $h \in [\ell_r - 1]$ even. By ($\mathcal{L}$5), if $\mathcal{P}_r$ does not contain u_r, then $\mathcal{P}_r$ starts in w_{i_r}, and u_r does not lie in any other path; therefore, we can update $\mathcal{P}_r$ as $\mathcal{P}_r := u_r\mathcal{P}_r$. Similarly, if $\mathcal{P}_r$ does not contain v_r, then $\mathcal{P}_r$ ends in z_{j_r} and v_r does not lie in any other path, and thus we can update $\mathcal{P}_r$ as $\mathcal{P}_r := \mathcal{P}_r v_r$. It follows that $\bigcup_{r\in[m]} V(\mathcal{P}_r) = V(\mathcal{M}^*) \setminus L$, and thus the paths $\mathcal{P}_r$ are as required in lemma 2.7.11. □

We also need the following simpler result. Its proof follows similar ideas as those present in the proof of lemma 2.7.11. For the sake of completeness, we include the proof of lemma 2.7.12 in appendix A.3. We point out here that lemma 2.7.9(iii) is only needed for this proof.

Lemma 2.7.12. *Let $n, s, \ell \in \mathbb{N}$ be such that $4 \leq s$ and $100 \leq \ell \leq n - s$. Let $G \subseteq \mathcal{Q}^n$ and consider any (s, ℓ)-molecule $\mathcal{M} = \mathcal{A}_1 \cup \dots \cup \mathcal{A}_{2^s} \subseteq \mathcal{Q}^n$ which is bonded in G. Let $\mathcal{M}^* = \mathcal{A}_{a+1} \cup \dots \cup \mathcal{A}_{a+t}$, for some $a \in [2^s]$ and $t \geq 10$, be a slice of $\mathcal{M}$. Moreover, consider the following sets.*

(C′1) *Let* $L \subseteq V(\mathcal{M}^*)$ *be a set of size* $|L| \in \{0,2\}$ *such that, if* $L = \{x, y\}$*, then* $x \in V(\mathcal{A}_i)$ *and* $y \in V(\mathcal{A}_j)$*, with* $i \neq j$ *and* $x \neq_{\mathrm{p}} y$*.*

(C′2) *Let* $R \subseteq V(\mathcal{M}^*) \setminus L$ *be a (possibly empty) set of vertices with* $|R| \leq 10$ *such that, for all* $k \in [a + t] \setminus [a]$*, we have* $|R \cap V(\mathcal{A}_k)| \in \{0, 2\}$ *and, if* $|R \cap V(\mathcal{A}_k)| = 2$*, then* $R \cap V(\mathcal{A}_k) = \{w_k, z_k\}$ *satisfies that* $w_k z_k \in E(\mathcal{M}^*)$ *and, if* $|L| = 2$*, then* $k \notin \{i, j\}$*.*

(C′3) *Consider two vertex-disjoint pairs* $\{u_r, v_r\}_{r\in[2]}$ *with* $u_1, u_2 \in V(\mathcal{A}_{a+1}) \setminus L$ *and* $v_1, v_2 \in V(\mathcal{A}_{a+t}) \setminus L$ *such that* $u_1 \neq_{\mathrm{p}} u_2$*,* $v_1 \neq_{\mathrm{p}} v_2$*,* $u_1 =_{\mathrm{p}} v_1$*, and* $|\{u_1, u_2\} \cap R|, |\{v_1, v_2\} \cap R| \leq 1$*.*

Then, there exist two vertex-disjoint paths $\mathcal{P}_1, \mathcal{P}_2 \subseteq \mathcal{M}^* \cup G$ *such that, for each* $r \in [2]$*,* $\mathcal{P}_r$ *is a* (u_r, v_r)*-path,* $V(\mathcal{P}_1) \cup V(\mathcal{P}_2) = V(\mathcal{M}^*) \setminus L$*, and every pair of the form* $\{w_k, z_k\} \subseteq R$ *with* $k \in [a + t] \setminus [a]$ *is an edge of either* $\mathcal{P}_1$ *or* $\mathcal{P}_2$*.*

2.8. Proofs of the main results

In this section we combine all the tools we have developed so far to prove theorems 2.1.5 and 2.1.8. For the proofs of theorems 2.1.1, 2.1.2 and 2.1.7, please see [34]. Theorem 2.1.8 actually follows from a slightly more general result (see theorem 2.8.1 below). We first prove the result for the property of being Hamiltonian, and then use this to prove the general result. To state the result, we need the following notation.

Given any integers $s \leq n$, we say that $d \in \mathcal{D}(\mathcal{Q}^n)$ is an *s-direction* if its only non-zero coordinate is one of the first s coordinates. Given a graph $F \subseteq \mathcal{Q}^n$, a set $\mathcal{U} \subseteq V(\mathcal{Q}^n)$ and $\ell, s \in \mathbb{N}$, we say that F is $(\mathcal{U}, \ell, s)$*-good* if, for each $x \in \mathcal{U}$, the set $E_F(x) := \{e \in E(F) : e \cap N_{\mathcal{Q}^n}(x) \neq \varnothing\}$ satisfies that, for each $d \in \mathcal{D}(\mathcal{Q}^n)$ which is not an s-direction, we have $|\{e \in E_F(x) : \mathcal{D}(e) = d\}| \leq n/\ell$. Thus, a graph is good if locally the directions of its edges are not too correlated (ignoring s-directions). The goodness of the “forbidden” graph F below will be needed when finding the special absorbing structures (see Step 11).

Theorem 2.8.1. *Let* $0 < 1/\ell \ll \varepsilon_1 \ll \varepsilon_2 \ll \gamma \leq 1$ *and* $1/\ell \ll \eta, 1/c \leq 1$*, with* $\ell \in \mathbb{N}$*. Let* $s := 10\ell$ *and* $n \in \mathbb{N}$*. Then, there exists* $\Phi \in \mathbb{N}$ *such that the following holds.*

Let $H \subseteq \mathcal{Q}^n$ and $\mathcal{U} \subseteq V(\mathcal{Q}^n)$ be such that H is an $(s, \ell, \varepsilon_1, \varepsilon_2, \gamma, \mathcal{U})$-robust subgraph, and let $Q \sim \mathcal{Q}^n_\eta$. Then, a.a.s. there is a $(\mathcal{U}, \ell^2, s)$-good subgraph $Q' \subseteq Q$ with $\Delta(Q') \leq \Phi$ such that

- *for every $H' \subseteq \mathcal{Q}^n$, where $d_{H'}(x) \geq 2$ for every $x \in \mathcal{U}$, and*
- *for every $F \subseteq \mathcal{Q}^n$ with $\Delta(F) \leq c\Phi$ which is $(\mathcal{U}, \ell, s)$-good,*

we have that $((H \cup Q) \setminus F) \cup H' \cup Q'$ contains a $(\mathcal{U}, \ell^2, s)$-good Hamilton cycle C such that, for all $x \in \mathcal{U}$, both edges of C incident to x belong to H'.

Note that theorem 2.8.1, together with lemma 2.7.7, directly implies the case $k = 2$ of theorem 2.1.8 (by choosing H in theorem 2.1.8 to play the role of H' in theorem 2.8.1 and F to be empty). In fact, in section 2.8.2 we will use theorem 2.8.1 to prove theorem 2.1.8 in full generality. For this derivation, we will need the stronger conditions imposed in the statement of theorem 2.8.1. More precisely, the formulation of theorem 2.8.1 involving a "forbidden" graph F and a "protected" graph G' is designed to make repeated applications of theorem 2.8.1 possible in order to take out k edge-disjoint Hamilton cycles. When finding the i-th Hamilton cycle, the protected graph will contain all the essential ingredients for this, while the forbidden graph will contain all previously chosen Hamilton cycles as well as the protected graphs for the entire set of Hamilton cycles. We prove theorem 2.8.1 in section 2.8.1, and then prove theorems 2.1.5 and 2.1.8 in section 2.8.2.

2.8.1. Hamilton cycles in robust subgraphs of the cube

Proof of theorem 2.8.1. Let $1/D, \delta' \ll 1$, let $\varepsilon_1 \ll \varepsilon_2 \ll \gamma, \delta' \leq 1$, and let

$$0 < 1/n_0 \ll \delta, \lambda \ll 1/\ell \ll 1/k^*, \alpha' \ll \beta, 1/S' \ll 1/c, 1/D, \eta, \varepsilon_1, \delta',$$

where $n_0, \ell, k^*, S', D \in \mathbb{N}$. Our proof assumes that n tends to infinity; in particular, $n \geq n_0$. Let $\Phi := 60\ell^4$ and $\Psi := c\Phi$.

Observe that $\mathcal{Q}^n[\{0,1\}^s \times \{0\}^{n-s}] \cong \mathcal{Q}^s$ contains a Hamilton cycle. We fix an ordering of layers $L_1, \dots, L_{2^s}$ of $\mathcal{Q}^n$ induced by this Hamilton cycle. Recall that we consider the intersection graph I and, for any $G \subseteq \mathcal{Q}^n$, the graph $I(G)$ can be seen as the "intersection" of the subgraphs that G induces on each layer. For each layer L and each $\mathcal{G} \subseteq I$, we denote by $\mathcal{G}_L$ the clone of $\mathcal{G}$ in L. Recall that, for any $\varepsilon \in [0,1]$, we have $I(\mathcal{Q}^n_\varepsilon) \sim \mathcal{Q}^{n-s}_{\varepsilon^{2^s}}$. We will sometimes write G_I for the subgraph of I where, for each $e \in E(I)$, we have $e \in E(G_I)$ whenever G contains some clone of e.

Let $\mathcal{U}$ be as in the statement of theorem 2.8.1. In particular, $\mathcal{U}(H,\varepsilon_1) \subseteq \mathcal{U}$ by (R1). Let $\mathcal{U}_I \subseteq V(I)$ be the set of vertices $x \in V(I)$ such that there is some clone x' of x with $x' \in \mathcal{U}$. Note that, by property (R2), for each $x \in \mathcal{U}_I$, there is exactly one clone x' of x with $x' \in \mathcal{U}$.

For each $i \in [8]$, let $\eta_i := \eta/8$ and $G_i \sim \mathcal{Q}^n_{\eta_i}$, where these graphs are chosen independently. It is easy to see that $\bigcup_{i=1}^8 G_i \sim \mathcal{Q}^n_{\eta'}$ for some $\eta' < \eta$, so it suffices to show that a.a.s. there is a $(\mathcal{U},\ell^2,s)$-good subgraph $Q' \subseteq \bigcup_{i=1}^8 G_i$ with $\Delta(Q') \leq \Phi$ and such that, for every $H' \subseteq \mathcal{Q}^n$, where $d_{H'}(x) \geq 2$ for every $x \in \mathcal{U}$, and every $F \subseteq \mathcal{Q}^n$ with $\Delta(F) \leq \Psi$ which is $(\mathcal{U},\ell,s)$-good, the graph $((H \cup \bigcup_{i=1}^8 G_i) \setminus F) \cup H' \cup Q'$ contains a Hamilton cycle of the form in the statement of the theorem. We now split our proof into several steps.

Step 1. Finding a tree and a reservoir. Consider the probability space $\Omega := \mathcal{Q}^{n-s}_{\eta_1^{2^s}} \times \mathit{Res}(\mathcal{Q}^{n-s},\delta')$ (with the latter defined as in section 2.5.3), so that, given $R \sim \mathit{Res}(I,\delta')$, we have that $(I(G_1),R) \sim \Omega$. Let $\mathcal{E}_1$ be the event that there exists a tree $T \subseteq I(G_1) - (R \cup B^5_I(\mathcal{U}_I))$ such that the following hold:

(TR1) $\Delta(T) < D$,

(TR2) for all $x \in V(I) \setminus B^5_I(\mathcal{U}_I)$, we have that $|N_I(x) \cap V(T)| \geq 4(n-s)/5$.

Note that, by (R3), for all $x,y \in \mathcal{U}_I$ we have that $\mathrm{dist}(x,y) \geq \gamma n/2$. Furthermore, by (R5), we have that, if we see $x_1 := \{0\}^{n-s}$, $x_2 := \{1\}^{n-s}$, $x_3 := \{1\}^{\lceil (n-s)/2\rceil}\{0\}^{n-s-\lceil (n-s)/2\rceil}$ and $x_4 := \{0\}^{\lceil (n-s)/2\rceil}\{1\}^{n-s-\lceil (n-s)/2\rceil}$ as vertices of I, then $\mathcal{U}_I \cap B^\ell_I(x_i) = \varnothing$ for all $i \in [4]$. Thus, it follows from theorem 2.5.11, with $n-s$, D, δ', $1/5$, $\eta_1^{2^s}$, $\gamma/2$, 5 and $\mathcal{U}_I$ playing the roles of n, D, δ, ε', ε, γ, k and $\mathcal{A}$, respectively, that $\mathbb{P}_\Omega[\mathcal{E}_1] = 1 - o(1)$.

Step 2. Identifying scant molecules. For each $v \in V(I)$, let $\mathcal{M}_v$ denote the vertex molecule of all clones of v in $\mathcal{Q}^n$. We say $\mathcal{M}_v$ is *scant* if there exist some layer L and some vertex $x \in V(\mathcal{M}_v \cap L) \setminus \mathcal{U}$ such that $e_H(x, R_L) < \varepsilon_1\delta' n/10$, where R_L is the clone of R in L. Let $\mathcal{E}_2$ be the event that there exists some $x \in V(I)$ such that there are at least S' vertices $v \in B_I^{10\ell}(x)$ with the property that $\mathcal{M}_v$ is scant. Let $\mathcal{E}_3$ be the event that there exist $x \in \mathcal{U}_I$ and $v \in B_I^{5\ell}(x)$ such that $\mathcal{M}_v$ is scant. It follows from lemma 2.7.8 with S' and δ' playing the roles of C and δ that $\mathbb{P}_\Omega[\mathcal{E}_2 \vee \mathcal{E}_3] = o(1)$. Let $\mathcal{E}_1^* := \mathcal{E}_1 \wedge \overline{\mathcal{E}_2} \wedge \overline{\mathcal{E}_3}$. Then, $\mathbb{P}_\Omega[\mathcal{E}_1^*] = 1 - o(1)$.

Condition on $\mathcal{E}_1^*$ holding. Then, there exist a set $R \subseteq V(I)$ and a tree $T \subseteq I(G_1) - (R \cup B_I^5(\mathcal{U}_I))$ such that the following hold:

(T1) $\Delta(T) < D$;

(T2) for all $x \in V(I) \setminus B_I^5(\mathcal{U}_I)$, we have that $|N_I(x) \cap V(T)| \geq 4(n-s)/5$;

(T3) for every $x \in V(I)$, we have $|\{v \in B_I^{10\ell}(x) : \mathcal{M}_v \text{ is scant}\}| \leq S'$, and

(T4) for every $x \in \mathcal{U}_I$ and every $v \in B_I^{5\ell}(x)$, we have that $\mathcal{M}_v$ is not scant.

Recall this implies clones of T and R satisfying (T1)–(T4) exist simultaneously in each layer of G_1.

Step 3: Finding clustered robust matchings for each molecule. Recall from section 2.2.5 that we will absorb vertices in pairs, where each pair consists of two clones x', x'' of the same vertex $x \in V(I)$. In this step, for each $x \in V(I)$ and for each set of clones of x that may need to be absorbed, we find a pairing of these clones so that we can later build suitable absorbing ℓ-cube pairs for each such pair of clones.

We now partition the set of layers into sets of consecutive layers as follows. Let

$$q := 2^{10Dk^*} \qquad \text{and} \qquad t := 2^s/q. \tag{2.33}$$

For each $j \in [t]$, let $S_j := \bigcup_{i=(j-1)q+1}^{jq} L_i$. Given any molecule $\mathcal{M}$, we consider the slices $\mathcal{S}_j(\mathcal{M}) := S_j \cap \mathcal{M}$. We denote by $\mathcal{S}(\mathcal{M})$ the collection of all these slices of $\mathcal{M}$.

Let $V_{\mathrm{sc}} \subseteq V(I)$ be the set of all vertices $x \in V(I)$ such that $\mathcal{M}_x$ is scant. In particular, by (T4) we have that $V_{\mathrm{sc}} \cap \mathcal{U}_I = \varnothing$. Recall $G_2 \sim \mathcal{Q}^n_{\eta_2}$. For each $v \in V(I) \setminus (V_{\mathrm{sc}} \cup \mathcal{U}_I)$, we define the following auxiliary bipartite graphs. Let $H(v) := (V(\mathcal{M}_v), N_I(v), E_H)$, where E_H is defined as follows. Consider $v' \in V(\mathcal{M}_v)$ and let $L^{v'}$ be the layer which contains v'. Let $w \in N_I(v)$, and let $w_{L^{v'}}$ be the clone of w in $L^{v'}$. Then, $\{v', w\} \in E_H$ if and only if $w \in R$ and $\{v', w_{L^{v'}}\} \in E(H)$. Note that $d_{H(v)}(v') \geq \varepsilon_1\delta' n/10$ for all $v' \in V(\mathcal{M}_v)$ since $\mathcal{M}_v$ is a not a scant molecule and does not contain any $x \in \mathcal{U}$. Similarly, we define $G_2(v) := (V(\mathcal{M}_v), N_I(v), E_{G_2})$, where $\{v', w\} \in E_{G_2}$ if and only if $\{v', w_{L^{v'}}\} \in E(G_2)$.

For each $v \in \mathcal{U}_I$, we also define two such auxiliary graphs. Let $H(v) := (V(\mathcal{M}_v), N_I(v), E^*_H)$, where E^*_H is defined as follows. As above, consider $v' \in V(\mathcal{M}_v)$ and let $L^{v'}$ be the layer which contains v'. Let $w \in N_I(v)$, and let $w_{L^{v'}}$ be the clone of w in $L^{v'}$. Then, if $v' \in \mathcal{U}$, we add $\{v', w\}$ to E^*_H (these can be seen as purely auxiliary edges, and we will ignore their effect later). Otherwise, $\{v', w\} \in E^*_H$ if and only if $w \in R$ and $\{v', w_{L^{v'}}\} \in E(H)$. In particular, $d_{H(v)}(v') \geq \varepsilon_1\delta' n/10$ for all $v' \in V(\mathcal{M}_v)$ since $\mathcal{M}_v$ is a not a scant molecule. As above, we define $G_2(v) := (V(\mathcal{M}_v), N_I(v), E_{G_2})$, where $\{v', w\} \in E_{G_2}$ if and only if $\{v', w_{L^{v'}}\} \in E(G_2)$.

For every $v \in V(I) \setminus V_{\mathrm{sc}}$ and every slice $\mathcal{S} \in \mathcal{S}(\mathcal{M}_v)$, note that the partition of $V(\mathcal{S})$ into vertices of even and odd parity is a balanced bipartition. Define the graph $\Gamma^\beta_{H(v),G_2(v)}(V(\mathcal{M}_v))$ as in section 2.5.1. Note that, by definition, we have that $V(\Gamma^\beta_{H(v),G_2(v)}(V(\mathcal{M}_v))) = V(\mathcal{M}_v)$. Furthermore, by definition,

(RM) given any $w_1, w_2 \in V(\mathcal{M}_v)$, we have that $\{w_1, w_2\} \in E(\Gamma^\beta_{H(v),G_2(v)}(V(\mathcal{M}_v)))$ if and only if $|N_{H(v)}(w_1) \cap N_{G_2(v)}(w_2)| \geq \beta(n-s)$ or $|N_{G_2(v)}(w_1) \cap N_{H(v)}(w_2)| \geq \beta(n-s)$.

For each $i \in [t]$, let $\mathfrak{A}_i(v)$ consist of all vertices of $V(\mathcal{S}_i(\mathcal{M}_v))$ of even parity, and let $\mathfrak{B}_i(v)$ consist of those of odd parity. By applying corollary 2.5.3 with $d = 100D$, $\alpha = \varepsilon_1\delta'/10$, $\varepsilon = \eta_2$, $n = n - s$, $k = q = 2^{10Dk^*}$, $\beta = \beta$, $t = t$, $G = H(\mathcal{M}_v)$ and $V(\mathcal{S}_1(\mathcal{M}_v)) \cup \ldots \cup V(\mathcal{S}_t(\mathcal{M}_v))$ as a partition of $V(\mathcal{M}_v)$, we obtain that, with probability at least $1 - 2^{-9(n-s)} \geq 1 - 2^{-8n}$, the graph $\Gamma^\beta_{H(v),G_2(v)}(V(\mathcal{M}_v))$ is $100D$-robust-parity-matchable with respect to $(\bigcup_{i=1}^t \mathfrak{A}_i(v), \bigcup_{i=1}^t \mathfrak{B}_i(v))$ clustered in $(V(\mathcal{S}_1(\mathcal{M}_v)), \ldots, V(\mathcal{S}_t(\mathcal{M}_v)))$.

We would like to proceed as above for scant molecules; however, recall that scant molecules contain vertices with few or no neighbours in the reservoir, and therefore we must adapt our approach. For each $v \in V_{\mathrm{sc}}$, we define auxiliary bipartite graphs $H(v)$ and $G_2(v)$ as we did for vertices in $V(I) \setminus (V_{\mathrm{sc}} \cup \mathcal{U}_I)$, except that we omit the condition that $w \in R$ for the existence of an edge in $H(v)$. Again, for each $i \in [t]$, let $\mathfrak{A}_i(v)$ consist of the vertices of $V(\mathcal{S}_i(\mathcal{M}_v))$ of even parity, and let $\mathfrak{B}_i(v)$ consist of those of odd parity. By applying corollary 2.5.3 again, we obtain that, with probability at least $1 - 2^{-8n}$, the graph $\Gamma^{\beta}_{H(v),G_2(v)}(V(\mathcal{M}_v))$ is $100D$-robust-parity-matchable with respect to $(\bigcup_{i=1}^{t} \mathfrak{A}_i(v), \bigcup_{i=1}^{t} \mathfrak{B}_i(v))$ clustered in $(V(\mathcal{S}_1(\mathcal{M}_v)), \dots, V(\mathcal{S}_t(\mathcal{M}_v)))$.

By a union bound over all $v \in V(I)$, a.a.s. $\Gamma^{\beta}_{H(v),G_2(v)}(V(\mathcal{M}_v))$ is $100D$-robust-parity-matchable with respect to $(\bigcup_{i=1}^{t} \mathfrak{A}_i(v), \bigcup_{i=1}^{t} \mathfrak{B}_i(v))$ clustered in $(V(\mathcal{S}_1(\mathcal{M}_v)), \dots, V(\mathcal{S}_t(\mathcal{M}_v)))$ for every $v \in V(I)$. We condition on this event holding and call it $\mathcal{E}_2^*$. Thus, for each $v \in V(I)$ and each set $S \subseteq V(\mathcal{M}_v)$ with $|S| \leq 100D$ which contains as many odd vertices as even vertices, there exists a perfect matching $\mathfrak{M}(\mathcal{M}_v, S)$ in the bipartite graph with parts consisting of the even and odd vertices of $V(\mathcal{M}_v) \setminus S$, respectively, and edges given by $\Gamma^{\beta}_{H(v),G_2(v)}(V(\mathcal{M}_v))$, with the property that, for each $e = \{w_{\mathrm{e}}, w_{\mathrm{o}}\} \in \mathfrak{M}(\mathcal{M}_v, S)$, if $w_{\mathrm{e}} \in V(\mathcal{S}_i(\mathcal{M}_v))$ for some $i \in [t]$, then $w_{\mathrm{o}} \in V(\mathcal{S}_{i-1}(\mathcal{M}_v)) \cup V(\mathcal{S}_i(\mathcal{M}_v)) \cup V(\mathcal{S}_{i+1}(\mathcal{M}_v))$ (where indices are taken cyclically). For each $v \in V(I) \setminus \mathcal{U}_I$, we denote by $\mathfrak{M}(v)$ the set of edges contained in the union (over all S) of the matchings $\mathfrak{M}(\mathcal{M}_v, S)$ (without multiplicity). For each $v \in \mathcal{U}_I$, we let $u(v)$ be the unique vertex in $\mathcal{M}_v$ such that $u(v) \in \mathcal{U}$, and we denote by $\mathfrak{M}(v)$ the set of edges contained in the union of the matchings $\mathfrak{M}(\mathcal{M}_v, S)$ over all S such that $u(v) \in S$ (again, without multiplicity). Furthermore, for each $v \in V(I)$ and each $e = \{w_{\mathrm{e}}, w_{\mathrm{o}}\} \in \mathfrak{M}(v)$, we let $N(e) := (N_{H(v)}(w_{\mathrm{e}}) \cap N_{G_2(v)}(w_{\mathrm{o}})) \cup (N_{G_2(v)}(w_{\mathrm{e}}) \cap N_{H(v)}(w_{\mathrm{o}}))$. By (RM), we have $|N(e)| \geq \beta(n-s) \geq \beta n/2$. Let $K := \max_{v \in V(I)} |\mathfrak{M}(v)|$. In particular, we have that $K \leq \binom{2^s}{2}$.

Step 4: Obtaining an appropriate cube factor via the nibble. For each $x \in V(I)$, consider the multiset $\mathfrak{A}(x) := \{N(e) : e \in \mathfrak{M}(x)\}$. If $|\mathfrak{A}(x)| < K$, we artificially increase its size to K by repeating any of its elements. Label the sets in $\mathfrak{A}(x)$ arbitrarily as $\mathfrak{A}(x) = \{A_1(x), \dots, A_K(x)\}$. Thus, if $x \in V(I) \setminus V_{\mathrm{sc}}$, then $A_i(x) \subseteq R$ for all $i \in [K]$.

Let $\mathcal{C}$ be any collection of subgraphs $C \subseteq I$ such that $C \cong \mathcal{Q}^\ell$ for all $C \in \mathcal{C}$. For any vertex $x \in V(I)$ and any set $Y \subseteq N_I(x)$, let $\mathcal{C}_x(Y) \subseteq \mathcal{C}$ be the set of all $C \in \mathcal{C}$ such that $x \notin V(C)$ and $Y \cap V(C) \neq \varnothing$, and let $\mathcal{C}_x := \mathcal{C}_x(N_I(x))$.

Recall $G_3 \sim \mathcal{Q}^n_{\eta_3}$ and $I(G_3) \sim \mathcal{Q}^{n-s}_{\eta_3^{2s}}$. We now apply theorem 2.6.6 to the graph $I(G_3)$, with $\eta_3^{2s}, \alpha', \delta/2, \beta/2, K$ and ℓ playing the roles of $\varepsilon, \alpha, \delta, \beta, K$ and ℓ, respectively, and using the sets $A_i(x)$ given above, for each $x \in V(I)$ and $i \in [K]$. Thus, a.a.s. we obtain a collection $\mathcal{C}$ of vertex-disjoint copies of $\mathcal{Q}^\ell$ in $I(G_3)$, such that the following properties hold for every $x \in V(I)$:

(N1) $|\mathcal{C}_x| \geq (1-\delta)n$.

(N2) For every direction $\hat{e} \in \mathcal{D}(I)$ we have that $|\Sigma(\mathcal{C}_x, \{\hat{e}\}, 1)| = o(n^{1/2})$.

(N3) For every $i \in [K]$ and every $S \subseteq \mathcal{D}(I)$ with $\alpha'(n-s)/2 \leq |S| \leq \alpha'(n-s)$ we have

$$|\Sigma(\mathcal{C}_x(A_i(x)), S, \ell^{1/2})| \geq |A_i(x)|/3000 \geq \beta n/6000.$$

Condition on the above event holding and call it $\mathcal{E}_3^*$.

Step 5: Absorption cubes. For each $x \in V(I)$ and $i \in [K]$, we define an auxiliary digraph $\mathfrak{D} = \mathfrak{D}(A_i(x))$ on vertex set $A_i(x) - \{x\}$ (seen as a set of directions of $\mathcal{D}(I)$) by adding a directed edge from $\hat{e}$ to $\hat{e}'$ if there is a cube $C^r \in \mathcal{C}_x(A_i(x))$ such that $x + \hat{e} \in V(C^r)$ and $\hat{e}' \in \mathcal{D}(C^r)$. In this way, an edge from $\hat{e}$ to $\hat{e}'$ in $\mathcal{D}$ indicates that the cube C^r could be used as a right absorber cube for x, if combined with a vertex-disjoint left absorber cube with tip $x + \hat{e}'$. Observe that, for all $\hat{e} \in A_i(x) - \{x\}$,

$$d^+_{\mathfrak{D}}(\hat{e}) \in [\ell]_0. \tag{2.34}$$

Furthermore, it follows by (N3) that any set $S \subseteq V(\mathfrak{D})$ with $|S| = \alpha' n/2$ satisfies

$$e_{\mathfrak{D}}(V(\mathfrak{D}), S) \geq \ell^{1/2}\beta n/6000 > \ell^{1/2}\beta^2 n. \tag{2.35}$$

Recall that $A_i(x) = N(\{x_1, x_2\})$ for some $\{x_1, x_2\} \in \mathfrak{M}(x)$. Note that $x_1, x_2 \in \mathcal{M}_x$, and let L^j be the layer containing x_j for each $j \in [2]$. We say that x_1 and x_2 are the vertices (or clones of x) which *correspond* to the pair (x, i). Let $(\hat{e}, \hat{e}') \in E(\mathfrak{D})$ and, for each $j \in [2]$, let e_j be the clone of $\{x + \hat{e}', x + \hat{e}' + \hat{e}\}$ in L^j. It follows that there is a cube $C^r \in \mathcal{C}_x(A_i(x))$ such that e_j connects the clone C_j of C^r to the clone of $x + \hat{e}'$ in L^j.

Recall $G_4 \sim \mathcal{Q}^n_{\eta_4}$. Let $\mathfrak{D}' \subseteq \mathfrak{D}$ be the subdigraph which retains each edge $(\hat{e}, \hat{e}') \in E(\mathfrak{D})$ if and only if the edges e_1, e_2 described above are both present in G_4. Note that each edge of $\mathfrak{D}$ is therefore retained independently of every other edge with probability η_4^2. By lemma 2.4.1, (2.34) and (2.35), it follows that $\mathfrak{D}'$ satisfies the following with probability at least $1 - e^{-10n}$:

(DG1) for every $A \subseteq V(\mathfrak{D})$ with $|A| = \alpha' n/2$ we have $\sum_{v \in A} d^-_{\mathfrak{D}'}(v) \geq \eta_4^3 \beta^2 \ell^{1/2} n$, and

(DG2) for every $B \subseteq V(\mathfrak{D})$ we have that $\sum_{v \in B} d^+_{\mathfrak{D}'}(v) \leq \ell |B|$.

Recall that $\mathfrak{D} = \mathfrak{D}(A_i(x))$. By a union bound, (DG1) and (DG2) hold a.a.s. for all $x \in V(I)$ and $i \in [K]$. We condition on this event and call it $\mathcal{E}_4^*$.

For each $x \in V(I)$ and $i \in [K]$, recall that (RM) and the definition of $A_i(x)$ imply that $|A_i(x)| \geq \beta(n - s)$. Thus, it follows by lemma 2.5.4 with $|A_i(x)|$, $2\alpha'/\beta$, $\eta_4^3 \beta^3 \ell^{1/2}/(2\alpha')$ and ℓ playing the roles of n, α, c and C, respectively, that there exists a matching $M''(A_i(x))$ of size at least $\frac{\eta_4^3 \beta^2}{2\ell^{1/2}} |A_i(x)| \geq \eta_4^3 \beta^3 n/(3\ell^{1/2})$ in $\mathfrak{D}'(A_i(x))$.

Next, for each $x \in V(I)$ and $i \in [K]$, we remove from $M''(A_i(x))$ all edges $(\hat{e}, \hat{e}') \in M''(A_i(x))$ such that $x + \hat{e}'$ does not lie in any cube of $\mathcal{C}_x(A_i(x))$. We denote the resulting matching by $M'(A_i(x))$. Note that, by (N1), we have

$$|M'(A_i(x))| \geq \eta_4^3 \beta^3 n/(3\ell^{1/2}) - \delta n \geq n/\ell. \qquad (2.36)$$

Consider $A_i(x)$, for some $x \in V(I)$ and $i \in [K]$, and let x_1, x_2 be the clones of x which correspond to (x, i). As before, for each $j \in [2]$, let L^j be the layer containing x_j. Recall definition 2.7.1 and note that by construction we have the following.

(AB1) For each edge $(\hat{e}, \hat{e}') \in M'(A_i(x))$, there is an absorbing ℓ-cube pair (C^l, C^r) for x in I such that, for each $j \in [2]$, the clone (C_j^l, C_j^r) of (C^l, C^r) in L^j is an absorbing ℓ-cube pair for x_j in $H \cup G_2 \cup G_3 \cup G_4$. In particular, the edge joining the left absorber tip to the third absorber vertex lies in G_4. Moreover, $C^l, C^r \in \mathcal{C}_x(A_i(x)) \subseteq \mathcal{C}$ and (C^l, C^r) has left and right absorber tips $x + \hat{e}'$ and $x + \hat{e}$, respectively. Furthermore, for each $x \in V(I) \setminus V_{\mathrm{sc}}$, these tips lie in R. We refer to (C_1^l, C_1^r) and (C_2^l, C_2^r) as the absorbing ℓ-cube pairs for x_1 and x_2 *associated* with $(\hat{e}, \hat{e}')$.

Thus, the graph $H \cup G_2 \cup G_3 \cup G_4$, contains at least n/ℓ absorbing ℓ-cube pairs for each of the clones x_1 and x_2 of x associated with edges in $M'(A_i(x)) \subseteq \mathfrak{D}(A_i(x))$. Moreover, since $M'(A_i(x))$ is a matching, for each $j \in [2]$ these absorbing ℓ-cube pairs for x_j are pairwise vertex-disjoint apart from x_j. Furthermore, recall by the construction in Step 3 that, for each $x \in V(I)$ and $i \in [K]$, we have $x_1, x_2 \notin \mathcal{U}$. In particular, this means that we do not choose absorbing ℓ-cube pairs for the vertices in $\mathcal{U}$ and, thus, the auxiliary edges at the vertices in $\mathcal{U}$ introduced in the definition of the sets E_H^* in Step 3 will never be used. As discussed before, the vertices in $\mathcal{U}$ will instead be incorporated into the Hamilton cycle using the special absorbing structures introduced in section 2.7.2.

For ease of notation, we will often consider the absorbing ℓ-cube pair (C^l, C^r) for x in I which (C_1^l, C_1^r) and (C_2^l, C_2^r) are clones of, and use it as a placeholder for both of its clones. By slightly abusing notation, we will refer to (C^l, C^r) as the *absorbing ℓ-cube pair associated* with $(\hat{e}, \hat{e}')$. Note, however, that (C^l, C^r) is not necessarily an absorbing ℓ-cube pair for x in $I(H \cup G_2 \cup G_3 \cup G_4)$.

Step 6: Removing bondless molecules. Recall $G_5 \sim \mathcal{Q}^n_{\eta_5}$. In this step, we consider the edges between the different layers.

For each $C \in \mathcal{C}$, let $\mathcal{M}_C$ denote the cube molecule consisting of the clones of C. Let $\mathcal{C}' \subseteq \mathcal{C}$ be the set of cubes $C \in \mathcal{C}$ for which $\mathcal{M}_C$ is bonded in G_5. By an application of lemma 2.7.2 with η_5 playing the role of ε, for each $C \in \mathcal{C}$ we have that

$$\mathbb{P}[C \notin \mathcal{C}'] = \mathbb{P}[\mathcal{M}_C \text{ is bondless in } G_5] \leq 2^{s+1-\eta_5 2^{\ell}/4} \leq 2^{-\eta_5 2^{\ell}/5}.$$

For each $x \in V(I)$, let $A_0(x) := N_I(x)$. For each $i \in [K]_0$, let $\mathcal{E}(x,i)$ be the event that $|\mathcal{C}_x(A_i(x)) \setminus \mathcal{C}'| > n/\ell^4$. Since the cubes $C \in \mathcal{C}$ are vertex-disjoint, the events that the molecules $\mathcal{M}_C$ are bondless in G_5 are independent. Therefore, we have that

$$\mathbb{P}[\mathcal{E}(x,i)] \leq \binom{n}{n/\ell^4}(2^{-\eta_5 2^{\ell}/5})^{n/\ell^4} \leq 2^{-10n}.$$

Let $\mathcal{E}_4 := \bigvee_{x\in V(I)} \bigvee_{i\in[K]_0} \mathcal{E}(x,i)$. By a union bound over all $x \in V(I)$ and $i \in [K]_0$, it follows that

$$\mathbb{P}[\mathcal{E}_4] \leq 2^{-8n}. \tag{2.37}$$

Let $\mathcal{C}_{\mathrm{bs}} \subseteq \mathcal{C}$ be the set of all $C \in \mathcal{C}$ such that $\mathcal{M}_C$ is bondlessly surrounded in G_5 (with respect to $\{\mathcal{M}_{C'} : C' \in \mathcal{C}\}$). For each $x \in V(I)$, let $\mathcal{E}(x)$ be the event that there are more than $n^{1/3}$ cubes $C \in \mathcal{C}_{\mathrm{bs}}$ which intersect $B_I^{\ell^2}(x)$. Let $\mathcal{E}_5 := \bigvee_{x\in V(I)} \mathcal{E}(x)$. By (N2), we may apply lemma 2.7.4 with η_5 playing the role of ε to conclude that

$$\mathbb{P}[\mathcal{E}_5] \leq 2^{-n^{9/8}}. \tag{2.38}$$

Now let $\mathcal{E}_5^* := \overline{\mathcal{E}_4} \wedge \overline{\mathcal{E}_5}$. It follows from (2.37) and (2.38) that $\mathcal{E}_5^*$ occurs a.a.s. Condition on this event.

Let $\mathcal{C}'' := \mathcal{C}' \setminus \mathcal{C}_{\mathrm{bs}}$. For each $x \in V(I)$ and each $i \in [K]$, let

(AB2) $M(A_i(x)) \subseteq M'(A_i(x))$ consist of all edges $(\hat{e},\hat{e}') \in M'(A_i(x))$ whose associated absorbing ℓ-cube pair (C^l, C^r) satisfies that $C^r, C^l \in \mathcal{C}''$.

By combining (2.36) with the further conditioning, it follows that, for each $x \in V(I)$ and each $i \in [K]$,

$$|M(A_i(x))| \geq n/\ell - n/\ell^4 - n^{1/3} \geq n/\ell^2. \tag{2.39}$$

Consider any $x \in V(I)$ and $i \in [K]$, and let x_1, x_2 be the two clones of x corresponding to (x, i). Then, for each $j \in [2]$, $H \cup G_2 \cup G_3 \cup G_4$ contains at least n/ℓ^2 vertex-disjoint (apart from x_j) absorbing ℓ-cube pairs for x_j such that each of these absorbing ℓ-cube pairs (C^l, C^r) is associated with an edge of $M(A_i(x))$, and for each $C \in \{C^l, C^r\}$ the corresponding cube molecule $\mathcal{M}_C$ is bonded in G_5 and (within the collection $\{\mathcal{M}_{C'} : C' \in \mathcal{C}\}$ of all cube molecules) $\mathcal{M}_C$ is not bondlessly surrounded in G_5.

Step 7: Extending the tree T. For each $x \in V(I) \setminus B_I^5(\mathcal{U}_I)$, let $Z(x) \coloneqq N_I(x) \cap V(T) \cap (\bigcup_{C\in\mathcal{C}''} V(C))$. It follows by (T2), (N1) and our conditioning on the event $\mathcal{E}_5^*$ that, for each $x \in V(I) \setminus B_I^5(\mathcal{U}_I)$, we have that

$$|Z(x)| \geq 4(n-s)/5 - \delta n - n/\ell^4 - n^{1/3} \geq 3n/4. \tag{2.40}$$

Recall $G_6 \sim \mathcal{Q}^n_{\eta_6}$. We apply theorem 2.5.12 with $\eta_6^{2^s}$, 2ℓ, T, R, $B_I^5(\mathcal{U}_I)$ and the sets $Z(x)$ playing the roles of ε, ℓ, T', R, W and $Z(x)$, respectively. Combining this with (T1), we conclude that a.a.s. there exists a tree T' such that $T \subseteq T' \subseteq (I(G_6) \cup T) - B_I^5(\mathcal{U}_I)$ and the following hold:

(ET1) $\Delta(T') < D + 1$;

(ET2) for all $x \in V(I)$, we have that $|B_I^{2\ell}(x) \setminus (V(T') \cup B_I^5(\mathcal{U}_I))| \leq n^{3/4}$;

(ET3) for each $x \in V(T') \cap R$, we have that $d_{T'}(x) = 1$ and the unique neighbour x' of x in T' is such that $x' \in Z(x)$.

We condition on the above event holding and call it $\mathcal{E}_6^*$.

At this point, for each $x \in V(I)$ and each $i \in [K]$, we redefine the set $M(A_i(x))$.

(AB3) Let $M(A_i(x))$ retain only those edges whose associated absorbing ℓ-cube pair (C^l, C^r) satisfies that both C^l and C^r intersect T' in at least two vertices.

It follows from (2.39), (ET2) and (R3) that

$$|M(A_i(x))| \geq n/\ell^2 - n^{3/4} \geq 4n/\ell^3. \tag{2.41}$$

Step 8. Consistent systems of paths and cubes. Recall $G_7 \sim \mathcal{Q}^n_{\eta_7}$. For each $v \in V(I)$, let $\mathcal{C}(v) := \{C \subseteq I(G_7) : C \cong \mathcal{Q}^\ell, v \in V(C)\}$. Let $\mathcal{P} := \{v \in V(I) : |\mathcal{C}(v)| \geq \lambda n^\ell\}$. By remark 2.5.9 (applied with $r = 10$ and $\eta_7^{2^s}$ playing the role of ε), the following property holds a.a.s.

(D1) For every $v \in V(I)$ we have $|B_I^{10}(v) \setminus \mathcal{P}| \leq n^{7/8}$.

For each $v \in \mathcal{P}$, a straightforward application of lemma 2.7.3 with λ and $\mathcal{C}(v)$ playing the roles of γ and $\mathcal{C}$, respectively, shows that the following holds with probability at least $1 - 2^{-10n}$: there exists a subcollection $\mathcal{C}'(v) \subseteq \mathcal{C}(v)$, with $|\mathcal{C}'(v)| \geq \lambda n^\ell/4$, with the property that, for every $C \in \mathcal{C}'(v)$, the molecule $\mathcal{M}_C$ is bonded in G_7. By a simple union bound over all vertices in $\mathcal{P}$, we obtain that the following holds a.a.s.

(D2) For every $v \in \mathcal{P}$, there exists a collection $\mathcal{C}'(v) \subseteq \mathcal{C}(v)$ with $|\mathcal{C}'(v)| \geq \lambda n^\ell/4$ such that, for every $C \in \mathcal{C}'(v)$, we have that $\mathcal{M}_C$ is bonded in G_7.

Condition on the event that (D1) and (D2) hold and call it $\mathcal{E}_7^*$.

We will show that we may extend many of the consistent systems of paths given by (R4) into special absorbing structures. Recall that, since H is $(s, \ell, \varepsilon_1, \varepsilon_2, \gamma, \mathcal{U})$-robust, for every $x \in \mathcal{U}$ and each pair of directions $a, b \in \mathcal{D}(\mathcal{Q}^n)$, there exists a collection $\mathfrak{C}(x, a, b)$ of (x, a, b)-consistent systems of paths in $H \cup \{\{x, x+a\}\{x, x+b\}\}$ satisfying (R4). By (D1), for every $x \in \mathcal{U}$ and $a, b \in \mathcal{D}(\mathcal{Q}^n)$

(CS) there exists a subcollection $\mathfrak{C}'(x, a, b) \subseteq \mathfrak{C}(x, a, b)$ which satisfies (R4) with $\gamma/2$ playing the role of γ and such that, for every $CS \in \mathfrak{C}'(x, a, b)$, we have $\mathrm{endmol}(CS) \subseteq \mathcal{P}$.

Let L be any layer of $\mathcal{Q}^n$. For each $x \in \mathcal{U}$, let $\mathrm{End}(x) := \bigcup_{a,b\in\mathcal{D}(\mathcal{Q}^n)} \bigcup_{CS\in\mathfrak{C}'(x,a,b)} \mathrm{endmol}(CS)$, let $\mathrm{End\,II}(x) := \bigcup_{a,b\in\mathcal{D}(\mathcal{Q}^n)\setminus\mathcal{D}(L)} \bigcup_{CS\in\mathfrak{C}'(x,a,b)} \mathrm{endmol}(CS)$, and let $x_I \in V(I)$ be such that

x is a clone of x_I. Given any tree $T^\bullet \subseteq I$ and any cube $C \subseteq V(I)$, we say that C *meets* $T^\bullet$ if $V(C) \cap V(T^\bullet) \neq \varnothing$. Recall that, for any $y, z \in V(I)$, we denote the set of directions in a shortest path in I between y and z by $\mathcal{D}(y, z)$.

Claim 2.6. *For each $x \in \mathcal{U}$ and each $z \in \mathrm{End}(x)$, there exists a collection of cubes $\mathcal{C}''(z) \subseteq \mathcal{C}'(z)$ with $|\mathcal{C}''(z)| = 20\ell$ which satisfies the following properties.*

(i) *Every $C \in \mathcal{C}''(z)$ meets T'.*

(ii) *For every $C \in \mathcal{C}''(z)$, we have $\mathcal{D}(C) \cap \mathcal{D}(x_I, z) = \varnothing$.*

(iii) *For every $C_1, C_2 \in \mathcal{C}''(z)$, we have $V(C_1) \cap V(C_2) = \{z\}$.*

(iv) *For each $x \in \mathcal{U}$, let $\mathcal{C}''(x) := \bigcup_{z \in \mathrm{End\,II}(x)} \mathcal{C}''(z)$. For each $d \in \mathcal{D}(L)$ and each $x \in \mathcal{U}$, let $\alpha(d, x) := |\{C \in \mathcal{C}''(x) : d \in \mathcal{D}(C)\}|$. For all $d \in \mathcal{D}(L)$ and $x \in \mathcal{U}$, we have that $\alpha(d, x) \leq n/\ell^3$.*

Proof. Fix any vertex $x \in \mathcal{U}$. Let $\mathfrak{m} := |\mathrm{End}(x)|$ and $m := |\mathrm{End\,II}(x)|$, and consider an arbitrary labelling $z_1, \dots, z_m, z_{m+1}, \dots, z_{\mathfrak{m}}$ of the vertices in $\mathrm{End}(x)$ such that all vertices in $\mathrm{End\,II}(x)$ come first. Observe that, by the definition of Type II consistent systems of paths (see section 2.7.2), we have that $m \leq n$. We will now iteratively define the sets $\mathcal{C}''(z_i)$ for each $i \in [\mathfrak{m}]$.

Fix first any $i \in [m]$, and suppose that a set $\mathcal{C}''(z_j)$ satisfying the claim is already defined for all $j \in [i-1]$. Let $\mathcal{C}''(x, i) := \bigcup_{j=1}^{i-1} \mathcal{C}''(z_j)$. For each $d \in \mathcal{D}(L)$, let $\alpha(d, x, i) := |\{C \in \mathcal{C}''(x, i) : d \in \mathcal{D}(C)\}|$. Let $\mathcal{D}^* := \{d \in \mathcal{D}(L) : \alpha(d, x, i) \geq n/\ell^3\}$. Observe that $|\mathcal{D}^*| \leq 20\ell^5$. Let $\mathcal{C}'''(z_i)$ be the set of all cubes $C \in \mathcal{C}'(z_i)$ such that C meets T' and $\mathcal{D}(C) \cap (\mathcal{D}(x_I, z_i) \cup \mathcal{D}^*) = \varnothing$ (i.e., they satisfy (i) and (ii) and, if added to the collection, would not violate (iv)). We claim that $|\mathcal{C}'''(z_i)| \geq \lambda n^\ell/5$. Indeed, observe that $\mathrm{dist}(x_I, z_i) \leq 2$ and, thus, the number of cubes $C \in \mathcal{C}'(z_i)$ such that $\mathcal{D}(C) \cap (\mathcal{D}(x_I, z_i) \cup \mathcal{D}^*) \neq \varnothing$ is at most $(20\ell^5 + 2)n^{\ell-1}$, and, by (ET2) and (R3), the number of such cubes which do not meet T' is at most $n^{\ell-2}$. The bound then follows by (D2).

We can now construct $\mathcal{C}''(z_i)$ by obtaining cubes $C_1(z_i), \dots, C_{20\ell}(z_i)$ iteratively. Note that, for any pair of cubes $C_1, C_2 \in \mathcal{C}'(z_i)$, we have that $z \in V(C_1) \cap V(C_2)$. Then, (iii) is equivalent to having that $\mathcal{D}(C_1) \cap \mathcal{D}(C_2) = \varnothing$. For each $k \in [20\ell]$, we proceed as follows. Let $\mathcal{D}'_k := \bigcup_{j=1}^{k-1} \mathcal{D}(C_j(z_i))$. Note that $|\mathcal{D}'_k| \leq 20\ell^2 \leq \lambda n/8$. Now, applying remark 2.7.5 with $n-s, z, \lambda/5, \lambda/8$ and $\mathcal{D}'_k$ playing the roles of n, x, η, η' and $\mathcal{D}'$, respectively, we deduce that there is a cube $C_k(z_i) \in \mathcal{C}'''(z_i)$ with $\mathcal{D}(C_k(z_i)) \cap \mathcal{D}'_k = \varnothing$. By enforcing that (iii) holds, it follows that each direction is used at most once in the cubes that were added in this step. It then follows that (iv) holds as well.

Consider now any $i \in [\mathfrak{m}] \setminus [m]$, and suppose that a set $\mathcal{C}''(z_j)$ satisfying the claim is already defined for all $j \in [i-1]$. Let $\mathcal{C}'''(z_i)$ be the set of all cubes $C \in \mathcal{C}'(z_i)$ such that C meets T' and $\mathcal{D}(C) \cap \mathcal{D}(x_I, z_i) = \varnothing$ (i.e., they satisfy (i) and (ii)). As above, we claim that $|\mathcal{C}'''(z_i)| \geq \lambda n^\ell/5$. Indeed, the number of cubes $C \in \mathcal{C}'(z_i)$ such that $\mathcal{D}(C) \cap \mathcal{D}(x_I, z_i) \neq \varnothing$ is at most $2n^{\ell-1}$, and, again, the number of such cubes which do not meet T' is at most $n^{\ell-2}$. The bound then follows by (D2).

We can now construct $\mathcal{C}''(z_i)$ as above. For each $k \in [20\ell]$, we proceed as follows. Let $\mathcal{D}'_k := \bigcup_{j=1}^{k-1} \mathcal{D}(C_j(z_i))$. Note that $|\mathcal{D}'_k| \leq 20\ell^2 \leq \lambda n/8$. Now, applying remark 2.7.5 with $n-s, z, \lambda/5, \lambda/8$ and $\mathcal{D}'_k$ playing the roles of n, x, η, η' and $\mathcal{D}'$, respectively, we deduce that there is a cube $C_k(z_i) \in \mathcal{C}'''(z_i)$ with $\mathcal{D}(C_k(z_i)) \cap \mathcal{D}'_k = \varnothing$. ◀

Let $J^1 := \bigcup_{x \in \mathcal{U}} \bigcup_{z \in \mathrm{End}(x)} \bigcup_{C \in \mathcal{C}''(z)} \mathcal{M}_C$, where $\mathcal{C}''(z)$ are the sets given by Claim 2.6, and let $G_7^* \subseteq G_7$ consist of all edges of G_7 which have endpoints in different layers.

Claim 2.7. *$J^1 \cup G_7^*$ is $(\mathcal{U}, \ell^3, s)$-good and* $\Delta(J^1 \cup G_7^*) \leq 50\ell^4$.

Proof. In order to see that $J^1 \cup G_7^*$ is $(\mathcal{U}, \ell^3, s)$-good, observe first that the edges of G_7^* do not affect this definition, so it suffices to see that J^1 is $(\mathcal{U}, \ell^3, s)$-good. By Claim 2.6(ii), for all $x \in \mathcal{U}, z \in \mathrm{End}(x)$ and $C \in \mathcal{C}''(z)$ we have that $\mathrm{dist}(x_I, C) = \mathrm{dist}(x_I, z)$. In particular, by the definition of the different consistent systems of paths (see section 2.7.2), it follows that the only cubes which affect whether J^1 is $(\mathcal{U}, \ell^3, s)$-good or not are those of the collection $\mathcal{C}''(x)$

described in Claim 2.6(iv). But then, by Claim 2.6(iv), we have that no direction $d \in \mathcal{D}(L)$ is used more than n/ℓ^3 times, as required.

Note that $\Delta(G_7^*) \leq s = 10\ell$, by construction. We will now show that $\Delta(J^1) \leq 45\ell^4$. Observe that J^1 does not contain any edges with endpoints in different layers. In particular, J^1 consists of clones of the same subgraph of I, that is $I(J^1) = \bigcup_{x\in\mathcal{U}} \bigcup_{z\in\mathrm{End}(x)} \bigcup_{C\in\mathcal{C}''(z)} C$. By this observation, it is enough to show that $\Delta(I(J^1)) \leq 45\ell^4$.

Recall that, by (R3), given any distinct $x, y \in \mathcal{U}_I$, we have that $\mathrm{dist}(x, y) \geq \gamma n/2$. In particular, by this observation and Claim 2.6(ii), it follows that, for all $x \in \mathcal{U}_I$, we have $d_{I(J^1)}(x) = 0$. We also note that, for every $z \in V(I)$ for which $\mathrm{dist}(z, \mathcal{U}_I) \geq \ell + 3$, we have $d_{I(J^1)}(z) = 0$. Now, fix any $x \in \mathcal{U}_I$ and $z \in V(I)$ such that $\mathrm{dist}(z, x) = t$, for some $t \in [\ell + 2]$. We claim that $d_{I(J^1)}(z) \leq 2t^2 \cdot 20\ell^2 \leq 45\ell^4$.

Suppose first that $t = 1$. Then, by Claim 2.6(ii), for every edge $e \in E(I(J^1))$ incident with z, we have $e \in E(C)$ for some $C \in \mathcal{C}''(z)$, and hence $d_{I(J^1)}(z) \leq \ell|\mathcal{C}''(z)| = 20\ell^2$, as we wanted to show. Suppose now that $t \geq 2$ and let $\mathcal{D}(z, x) = \{d_1, \dots, d_t\}$. Every edge e incident with z must come from the edges of a cube $C \in \mathcal{C}''(w)$, for some $w \in \mathrm{End}(x)$. Moreover, by Claim 2.6(ii), we must have $\mathcal{D}(x, w) \subsetneq \mathcal{D}(x, z)$. As there are at most $t + t^2 \leq 2t^2$ vertices $w \in V(I)$ such that $\mathcal{D}(x, w) \subsetneq \mathcal{D}(x, z)$ and $\mathrm{dist}(x, w) \in [2]$, we have have that $d_{I(J^1)}(z) \leq 2t^2 20\ell^2$, which concludes the proof of the claim. ◀

Step 9: Fixing a collection of absorbing ℓ-cube pairs for the vertices in scant molecules. Recall $G_8 \sim \mathcal{Q}^n_{\eta_8}$. Consider any $x \in V_{\mathrm{sc}}$ and $j \in [K]$. Recall from Step 3 that the tips of the cubes of the absorbing ℓ-cube pair associated with a given edge in $M(A_j(x))$ may not lie in the reservoir R. Roughly speaking, we will alter T' so that the tips are relocated from the tree T' to the reservoir.

We start by redefining the matchings $M(A_j(x))$ as follows: for each $x \in V_{\mathrm{sc}}$ and each $j \in [K]$, remove from $M(A_j(x))$ all edges $(\hat{e}, \hat{e}')$ such that $N_{T'}(x) \cap \{x + \hat{e}, x + \hat{e}'\} \neq \varnothing$. It follows from (2.41) and (ET1) that, for all $x \in V_{\mathrm{sc}}$ and $j \in [K]$,

$$|M(A_j(x))| \geq 4n/\ell^3 - D > 2n/\ell^3. \tag{2.42}$$

For each $x \in V_{\mathrm{sc}}$, each $j \in [K]$ and each matching $M' \subseteq M(A_j(x))$ with $|M'| \geq n/\ell^3$, let $\mathcal{E}'(x, j, M')$ be the following event:

> for every set $B \subseteq V(I)$ with $|B| < 2^{\ell+s+3}\Psi KS'$, there exists an edge $\vec{e} \in M'$, whose associated absorbing ℓ-cube pair (C^l, C^r) has tips x^l and x^r, for which there exists a subgraph $P(\vec{e}, B) \subseteq I(G_8) - \{x^l, x^r\}$ such that $|V(P(\vec{e}, B))| < 21D/2$, $V(P(\vec{e}, B)) \cap B = \varnothing$, and both $N_{T'}(x^l)$ and $N_{T'}(x^r)$ are connected in $P(\vec{e}, B)$.

For a graph $P(\vec{e}, B)$ as above, we will refer to x^l and x^r as the tips *associated* with $P(\vec{e}, B)$, and refer to (C^l, C^r) as the absorbing ℓ-cube pair *associated* with $P(\vec{e}, B)$. (Recall that, if $\vec{e} = (\hat{e}, \hat{e}')$, then $x^l = x + \hat{e}$ and $x^r = x + \hat{e}'$.)

By invoking lemma 2.5.13 with $n - s$, $\eta_8^{2^s}$, $1/\ell^3$, $2^{\ell+s+3}\Psi KS'$, $2D + 2$ and the sets $\{(x + \hat{e}, x + \hat{e}') : (\hat{e}, \hat{e}') \in M'\}$ and $(N_{T'}(x + \hat{e}) \cup N_{T'}(x + \hat{e}'))_{(\hat{e},\hat{e}') \in M'}$ playing the roles of n, ε, c, f, D, $C(x)$ and $(B(y, z))_{(y,z) \in C(x)}$, respectively, we have that $\mathcal{E}'(x, j, M')$ holds with probability at least $1 - 2^{-5(n-s)}$. Let $\mathcal{E}_8^* := \bigwedge_{x \in V_{\mathrm{sc}}} \bigwedge_{j \in [K]} \bigwedge_{M' \subseteq M(A_j(x)) : |M'| \geq n/\ell^3} \mathcal{E}'(x, j, M')$. By a union bound over all $x \in V_{\mathrm{sc}}$, $j \in [K]$ and $M' \subseteq M(A_j(x))$ such that $|M'| \geq n/\ell^3$, it follows that $\mathbb{P}[\mathcal{E}_8^*] \geq 1 - 2^{-2n}$.

Condition on the event that $\mathcal{E}_8^*$ holds. It follows that, for each $x \in V_{\mathrm{sc}}$, $j \in [K]$, $M' \subseteq M(A_j(x))$ with $|M'| \geq n/\ell^3$ and any $B \subseteq V(I)$ with $|B| < 2^{\ell+s+3}\Psi KS'$, there exists a subgraph $P(x, j, M', B) \subseteq I(G_8)$ with $|V(P(x, j, M', B))| < 21D/2$ which avoids $B \cup \{x^l, x^r\}$, where x^l and x^r are the tips associated with $P(x, j, M', B)$, and such that both $N_{T'}(x^l)$ and $N_{T'}(x^r)$ are connected in $P(x, j, M', B)$. Moreover, by choosing $P(x, j, M', B)$ minimal, we may assume that it consists of at most two components, and each such component contains either $N_{T'}(x^l)$ or $N_{T'}(x^r)$.

Let $\iota := |V_{\mathrm{sc}}|$ and let $x_1, \dots, x_\iota$ be an ordering of V_{sc}. For each $i \in [\iota]$, $j \in [K]$ and $k \in [2^{s+1}\Psi]$, by ranging over i first, then j, and then k, we will iteratively fix a graph $P(x_i, j, k, M'_{i,j,k}, B_{i,j,k})$ as above. In particular, this graph will have an absorbing ℓ-cube pair with tips $x^l_{i,j,k}$ and $x^r_{i,j,k}$ associated with it. After the graph $P(x_i, j, k, M'_{i,j,k}, B_{i,j,k})$ is fixed, so are these tips. Let $\mathcal{J}_{i,j,k} := ([i-1] \times [K] \times [2^{s+1}\Psi]) \cup \{(i, j', k') : (j', k') \in [j-1] \times [2^{s+1}\Psi]\} \cup \{(i, j, k'') : k'' \in [k-1]\}$ and suppose that we have already fixed $P(x_{i'}, j', k', M'_{i',j',k'}, B_{i',j',k'})$

for all $(i',j',k')\in\mathcal{J}_{i,j,k}$ such that these $P(x_{i'},j',k',M'_{i',j',k'},B_{i',j',k'})$ are vertex-disjoint from each other and from the set $\{x^l_{i',j',k'},x^r_{i',j',k'}:(i',j',k')\in\mathcal{J}_{i,j,k}\}$ of tips associated with all these $P(x_{i'},j',k',M'_{i',j',k'},B_{i',j',k'})$. In order to fix $P(x_i,j,k,M'_{i,j,k},B_{i,j,k})$, we first define the sets $B_{i,j,k}$ and $M'_{i,j,k}$. Let $M'_{i,j,k}$ be obtained from $M(A_j(x_i))$ as follows. Remove all edges whose associated absorbing ℓ-cube pair (C^l,C^r) satisfies $(V(C^l)\cup V(C^r))\cap\{x^l_{i',j',k'},x^r_{i',j',k'}:(i',j',k')\in\mathcal{J}_{i,j,k}\}\neq\varnothing$. Remove all edges $(\hat{e},\hat{e}')\in M(A_j(x_i))$ such that $\{x_i+\hat{e},x_i+\hat{e}'\}\cap\bigcup_{(i',j',k')\in\mathcal{J}_{i,j,k}}V(P(x_{i'},j',k',M'_{i',j',k'},B_{i',j',k'}))\neq\varnothing$ too. Note that, by (2.42) and (T3), it follows that $|M'_{i,j,k}|\geq n/\ell^3$. Let $B_{i,j,k}$ be the set of vertices $y\in B^{\ell/2}_I(x_i)$ such that at least one of the following holds:

(P1) there exists $(i',j',k')\in\mathcal{J}_{i,j,k}$ such that $y\in V(P(x_{i'},j',k',M'_{i',j',k'},B_{i',j',k'}))$;

(P2) there exists $(i',j',k')\in\mathcal{J}_{i,j,k}$ such that y lies in the absorbing ℓ-cube pair associated with $P(x_{i'},j',k',M'_{i',j',k'},B_{i',j',k'})$.

Note that $|B_{i,j,k}|<2^{s+\ell+3}\Psi KS'$ by (T3). We then fix $P(x_i,j,k,M'_{i,j,k},B_{i,j,k})$ to be the graph guaranteed by our conditioning on $\mathcal{E}^*_8$. Observe that, by the choice of $B_{i,j,k}$, we have that $P(x_i,j,k,M'_{i,j,k},B_{i,j,k})$ is vertex-disjoint from $\bigcup_{(i',j',k')\in\mathcal{J}_{i,j,k}}P(x_{i'},j',k',M'_{i',j',k'},B_{i',j',k'})$. We denote by $(C^l(x_i,j,k),C^r(x_i,j,k))$ the absorbing ℓ-cube pair for x_i associated with $P(x_i,j,k,M'_{i,j,k},B_{i,j,k})$. By the choice of $M'_{i,j,k}$ we have that

(CD) for all $(i',j',k')\in\mathcal{J}_{i,j,k}$, $C^l(x_i,j,k)$ and $C^r(x_i,j,k)$ are both vertex-disjoint from $C^l(x_{i'},j',k')$ and $C^r(x_{i'},j',k')$.

For each $i\in[\iota]$ and $j\in[K]$, let $\mathcal{C}^{\mathrm{sc}}_1(x_i,j):=\{(C^l(x_i,j,k),C^r(x_i,j,k)):k\in[2^{s+1}\Psi]\}$, and let $\mathcal{C}^{\mathrm{sc}}_1:=\bigcup_{(i,j)\in[\iota]\times[K]}\mathcal{C}^{\mathrm{sc}}_1(x_i,j)$. Let $P':=\{x^l_{i,j,k},x^r_{i,j,k}:(i,j,k)\in[\iota]\times[K]\times[2^{s+1}\Psi]\}$ and $P:=\bigcup_{(i,j,k)\in[\iota]\times[K]\times[2^{s+1}\Psi]}P(x_i,j,k,M'_{i,j,k},B_{i,j,k})$. Recall that $P(x_i,j,k,M'_{i,j,k},B_{i,j,k})$ avoids the tips $x^l_{i,j,k}$ and $x^r_{i,j,k}$ associated with it. It follows from this, (P2), and the definition of $M'_{i,j,k}$ that $P'\cap V(P)=\varnothing$. Finally, observe that (T4) and the assumption on the size of the graphs $P(x_i,j,k,M'_{i,j,k},B_{i,j,k})$ imply that $(P'\cup V(P))\cap B^{2\ell}(\mathcal{U}_I)=\varnothing$. Let $T^{\mathrm{IV}}:=T'[V(T')\setminus P']\cup P$. Note that T^{IV} is connected by the definition of $\mathcal{E}'(x,j,M')$. Let T'' be a spanning tree of T^{IV}.

In particular, it follows from the above and the definitions of T and T' in Steps 1 and 7 that

$$T'' \subseteq I(G_1 \cup G_6 \cup G_8) - B_I^5(\mathcal{U}_I) \subseteq I - B_I^5(\mathcal{U}_I). \tag{2.43}$$

By (ET1) and the fact that the graphs $P(x_i, j, k, M'_{i,j,k}, B_{i,j,k})$ are vertex-disjoint and satisfy $|V(P(x_i, j, k, M'_{i,j,k}, B_{i,j,k}))| < 21D/2$, it follows that

$$\Delta(T'') \leq 12D. \tag{2.44}$$

Define the (new) reservoir $R' \coloneqq (R \cup P') \setminus V(P)$.

For each $x \in V(I) \setminus B_I^5(\mathcal{U}_I)$, let $Z'(x) \coloneqq Z(x) \cap V(T'')$ (where $Z(x)$ is as defined in Step 7). It follows by (2.40) and (T3) that

$$|Z'(x)| \geq 3n/4 - 4 \cdot 2^s \Psi K S' \geq n/2.$$

Choose any vertex $x \in V(I)$ with $\mathrm{dist}(x, \mathcal{U}_I) \geq 3\ell$. Again by (T3), there are at most $4 \cdot 2^{s+\ell} \Psi K S'$ vertices in $Z'(x)$ which lie in cubes of absorbing ℓ-cube pairs of $\mathcal{C}_1^{\mathrm{sc}}$. Choose any vertex $y \in Z'(x)$ which does not lie in any of those cubes. Denote the cube $C \in \mathcal{C}''$ which contains y by $C^\bullet$.

At this point, for each $x \in V(I) \setminus V_{\mathrm{sc}}$ and each $i \in [K]$, we redefine the set $M(A_i(x))$ as follows.

(AB4) Let $M(A_i(x))$ retain only those edges whose associated absorbing ℓ-cube pair (C^l, C^r) satisfies that both C^l and C^r are different from $C^\bullet$ and vertex-disjoint from both cubes of all absorbing ℓ-cube pairs of $\mathcal{C}_1^{\mathrm{sc}}$, and both tips x^l and x^r satisfy that $x^l, x^r \in R \setminus V(P) \subseteq R'$.

Note that, by (T3), we have $|B_I^{\ell+1}(x) \cap V(P)| \leq 21 \cdot 2^s \Psi D K S'$ and $|B_I^{\ell+1}(x) \cap V(\bigcup_{(C^l, C^r) \in \mathcal{C}_1^{\mathrm{sc}}} (C^l \cup C^r))| \leq 4 \cdot 2^{\ell+s} \Psi K S'$. Combining this with (2.41) and (AB1), it follows that

$$|M(A_i(x))| \geq 4n/\ell^3 - (21D + 4 \cdot 2^\ell) 2^s \Psi K S' - 1 \geq 2n/\ell^3. \tag{2.45}$$

Step 10: Fixing a collection of absorbing ℓ-cube pairs for vertices in non-scant molecules and vertices near $\mathcal{U}$. At this point, it is not yet clear which vertices will need to eventually be absorbed into the long cycle we construct. For vertices in I which are "far" from $\mathcal{U}_I$, we can already determine those which will have clones that will need to be absorbed (though we cannot yet determine the precise clones). However, for vertices which are "near" $\mathcal{U}_I$, we still cannot say which of them will have clones that need to be absorbed (this depends on the special absorbing structure which is fixed once the edges of H' are revealed). As a result, we proceed as if all of the clones of vertices in I near $\mathcal{U}_I$ will need to be absorbed. Recall that $\mathcal{C}'$ and $\mathcal{C}''$ were defined in Step 6. Recall also the notation $\mathcal{C}_x$ introduced in Step 4. Let

$$\mathcal{C}''' \coloneqq \{C \in \mathcal{C}' : V(C) \cap V(T'') \neq \varnothing\} \quad \text{and} \quad V'_{\mathrm{abs}} \coloneqq (V(I) \setminus \bigcup_{C \in \mathcal{C}'''} V(C)) \cup B_I^{3\ell}(\mathcal{U}_I).$$

We will now fix a collection of absorbing ℓ-cube pairs for all vertices in each vertex molecule $\mathcal{M}_x$ with $x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$, except for the vertices of $\mathcal{U}$.

First, recall from (T3) that, for all $x \in V(I)$, we have that $|B_I^{10\ell}(x) \cap V_{\mathrm{sc}}| \leq S'$. Thus, in constructing T'', we removed at most $2^{s+2}\Psi K S'$ vertices in $B_I^{2\ell}(x)$ from T'. Therefore, it follows from (ET2) that, for all $x \in V(I)$, we have

$$|B_I^{2\ell}(x) \setminus (V(T'') \cup B_I^{5}(\mathcal{U}_I))| \leq 2n^{3/4}. \tag{2.46}$$

For all $x \in \bigcup_{C \in \mathcal{C}''} V(C)$, by combining (N1), (2.46) and (R3) with the definition of bondlessly surrounded molecules, we have that

$$|\mathcal{C}_x \cap \mathcal{C}'''| \geq (1 - 2^{-\ell - 5s + 1})n. \tag{2.47}$$

Recall that, for any $x \in V(I)$, each index $i \in [K]$ is given by a unique edge $e \in \mathfrak{M}(x)$ via the relation $N(e) = A_i(x)$. Recall also the definition of $\mathfrak{M}(x)$ from Step 3.

Claim 2.8. *For each $x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$ and each $e \in \mathfrak{M}(x)$, there exists a set $\mathcal{C}_1^{\mathrm{abs}}(e)$ of $2^{s+1}\Psi$ absorbing ℓ-cube pairs $(C_k^l(e), C_k^r(e)) \subseteq I$, one for each $k \in [2^{s+1}\Psi]$, which satisfies the following:*

(i) *for all $x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$, $e \in \mathfrak{M}(x)$ and $k \in [2^{s+1}\Psi]$, the absorbing ℓ-cube pair $(C_k^l(e), C_k^r(e))$ is associated with some edge in $M(A_j(x))$, for some $j \in [K]$;*

(ii) *for all $x \in B_I^{3\ell}(\mathcal{U}_I)$, $e \in \mathfrak{M}(x)$ and $k \in [2^{s+1}\Psi]$, the absorbing ℓ-cube pair $(C_k^l(e), C_k^r(e))$ has tips $x_k^l(e)$ and $x_k^r(e)$ which satisfy that $\mathrm{dist}(x, \mathcal{U}_I) < \mathrm{dist}(x_k^l(e), \mathcal{U}_I)$ and $\mathrm{dist}(x, \mathcal{U}_I) < \mathrm{dist}(x_k^r(e), \mathcal{U}_I)$, and*

(iii) *for all $x, x' \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$, all $e \in \mathfrak{M}(x)$ and $e' \in \mathfrak{M}(x')$, and all $k, k' \in [2^{s+1}\Psi]$ with $(x, e, k) \neq (x', e', k')$, the absorbing ℓ-cube pairs $(C_k^l(e), C_k^r(e))$ and $(C_{k'}^l(e'), C_{k'}^r(e'))$ satisfy that $(V(C_k^l(e)) \cup V(C_k^r(e))) \cap (V(C_{k'}^l(e')) \cup V(C_{k'}^r(e'))) = \varnothing$.*

Proof. Let $\mathcal{V} := \bigcup_{x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}} \mathfrak{M}(x)$. Let $K' := |\mathcal{V}|$, and let $f_1, \dots, f_{K'}$ be an ordering of the edges in $\mathcal{V}$. Given any $i \in [K']$, the edge f_i corresponds to a pair $(x, j(i))$, where $x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$ and $j(i) \in [K]$. If $x \notin B_I^{3\ell}(\mathcal{U}_I)$, let $\mathfrak{C}_i$ be the collection of absorbing ℓ-cube pairs for x in I associated with some edge of $M(A_{j(i)}(x))$. Otherwise, let $\mathfrak{C}_i$ be the same collection, after removing all those absorbing ℓ-cube pairs for which (ii) does not hold. Since each $x \in B_I^{3\ell}(\mathcal{U}_I)$ has at most 3ℓ neighbours $y \in N_I(x)$ such that $\mathrm{dist}(x, \mathcal{U}_I) \geq \mathrm{dist}(y, \mathcal{U}_I)$, it follows by (2.45) that $|\mathfrak{C}_i| \geq n/\ell^3$ for all $i \in [K']$. In particular, by (AB2), each of the absorbing ℓ-cube pairs (C^l, C^r) in any of the collections $\mathfrak{C}_i$ satisfies that $C^l, C^r \in \mathcal{C}''$.

Let $\mathcal{H}$ be the $2^{s+1}\Psi K'$-edge-coloured auxiliary multigraph with $V(\mathcal{H}) := \mathcal{C}''$, which contains one edge of colour $(i, k) \in [K'] \times [2^{s+1}\Psi]$ between C and C' whenever $(C, C') \in \mathfrak{C}_i$ or $(C', C) \in \mathfrak{C}_i$. In particular, $\mathcal{H}$ contains at least n/ℓ^3 edges of each colour. We now bound $\Delta(\mathcal{H})$. Consider any $C \in V(\mathcal{H})$. Note that, for each edge e of $\mathcal{H}$ incident to C, there exists some $x = x(e) \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$ such that C together with some other cube $C' \in V(\mathcal{H})$ forms an absorbing ℓ-cube pair for x. In particular, x must be adjacent to C in I. Let $\overline{\partial}(C) := \{x(e) : e \in E(\mathcal{H}) \text{ is incident to } C\}$. Moreover, if e has colour (i, z), then $f_i \in \mathfrak{M}(x)$ (and f_i has corresponding pair $(x, j(i))$ for some $j(i) \in [K]$). Since $f_i \in \mathfrak{M}(x)$ and $|\mathfrak{M}(x)| \leq \binom{2^s}{2}$, it follows that each vertex y which is adjacent to C in I can play the role of x for at most $2^{s+1}\Psi \cdot 2^{2s}$ edges of $\mathcal{H}$ incident to C. Thus, $d_{\mathcal{H}}(C) \leq 2^{3s+1}\Psi|\bar{\partial}(C)|$.

Fix a cube $C \in V(\mathcal{H})$. In order to bound $|\bar{\partial}(C)|$, consider first $|\bar{\partial}(C) \cap B_I^{3\ell}(\mathcal{U}_I)|$. Recall that, by (R3), there is at most one vertex $z \in \mathcal{U}_I \cap B_I^{3\ell}(V(C))$. Furthermore, since the property described in (ii) holds for all absorbing ℓ-cube pairs for vertices in $B_I^{3\ell}(\mathcal{U}_I)$ represented in $\mathcal{H}$, it follows that each vertex $x \in V(C)$ has at most $3\ell+1$ neighbours in $B_I^{3\ell}(\mathcal{U}_I) \cap \bar{\partial}(C)$. Thus, in total, $|\bar{\partial}(C) \cap B_I^{3\ell}(\mathcal{U}_I)| \leq (3\ell+1)2^{\ell}$. Consider now $|\bar{\partial}(C) \setminus B_I^{3\ell}(\mathcal{U}_I)|$. Note that $\bar{\partial}(C) \setminus B_I^{3\ell}(\mathcal{U}_I) \subseteq (V'_{\mathrm{abs}} \cap N_I(V(C))) \setminus B_I^{3\ell}(\mathcal{U}_I) \subseteq N_I(V(C)) \setminus \bigcup_{C' \in \mathcal{C}'''} V(C')$. By (2.47), the number of vertices in $V'_{\mathrm{abs}} \setminus B_I^{3\ell}(\mathcal{U}_I)$ which are adjacent to C is at most $2|C|n/2^{\ell+5s}$, that is, $|\bar{\partial}(C) \setminus B_I^{3\ell}(\mathcal{U}_I)| \leq 2n/2^{5s}$. We conclude that $|\bar{\partial}(C)| \leq 3n/2^{5s}$ and, thus, $d_{\mathcal{H}}(C) \leq 2^{3s+1}\Psi 3n/2^{5s} \leq n/\ell^4$.

Since each colour class has size at least n/ℓ^3 and $\Delta(\mathcal{H}) \leq n/\ell^4$, by lemma 2.5.5, $\mathcal{H}$ contains a rainbow matching of size $2^{s+1}\Psi K'$. For each $(i,z) \in [K'] \times [2^{s+1}\Psi]$, let $(C_z^l(f_i), C_z^r(f_i)) \in \mathfrak{C}_i$ be the absorbing ℓ-cube pair of colour (i,z) in this rainbow matching. This ensures that (iii) holds, while (i) and (ii) follow by construction. ◀

For each $x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$ and each $i \in [K]$, let $\mathcal{C}_1^{\mathrm{abs}}(x,i) := \mathcal{C}_1^{\mathrm{abs}}(e)$ be the set of absorbing ℓ-cube pairs guaranteed by Claim 2.8, where $e \in \mathfrak{M}(x)$ is the unique edge such that $A_i(x) = N(e)$. Similarly, for each $k \in [2^{s+1}\Psi]$, let $(C^l(x,i,k), C^r(x,i,k)) := (C_k^l(e), C_k^r(e))$. Let $\mathcal{C}_1^{\mathrm{abs}} := \bigcup_{x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}} \bigcup_{i \in [K]} \mathcal{C}_1^{\mathrm{abs}}(x,i)$.

Let $G := \bigcup_{i=1}^{8} G_i$. For each $x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$ and each $i \in [K]$, let $G^*(x,i) \subseteq I$ be the graph consisting of all edges between the left absorber tip and third absorber vertex of every absorbing ℓ-cube pair in $\mathcal{C}_1^{\mathrm{abs}}(x,i)$. Let $G^\bullet \subseteq I$ be the graph consisting of all edges between the left absorber tip and third absorber vertex of every absorbing ℓ-cube pair in $\mathcal{C}_1^{\mathrm{sc}}$. Let $G^* := G^\bullet \cup \bigcup_{x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}} \bigcup_{i \in [K]} G^*(x,i) \subseteq I$. Recall that, given any graph $\mathcal{G} \subseteq I$, for each layer L, we denote by $\mathcal{G}_L$ the clone of $\mathcal{G}$ in L. Let $G_4^* := G_4 \cap \bigcup_{i=1}^{2^s} G_{L_i}^*$. Furthermore, let $G_5^* \subseteq G_5$ consist of all edges of G_5 which have endpoints in different layers. Recall that G_7^* and J^1

were defined in Step 8. We let $Q' \subseteq G$ be the spanning subgraph with edge set

$$E(Q') := E(J^1) \cup E(G_4^*) \cup E(G_5^*) \cup E(G_7^*) \cup \bigcup_{C \in \mathcal{C}'} E(\mathcal{M}_C) \cup \bigcup_{i=1}^{2^s} E(T''_{L_i}).$$

Note that, using Claim 2.7 and (2.44), we have that $\Delta(Q') \leq \Phi$.

Claim 2.9. *Q' is $(\mathcal{U}, 2\ell^2, s)$-good.*

Proof. Indeed, observe that this fact only depends on those edges contained within a layer which are incident to a neighbour of x in $\mathcal{Q}^n$, for some $x \in \mathcal{U}$. Therefore, the graph G_5^* has no effect here. By property (C3) below, the graph $\bigcup_{i=1}^{2^s} T''_{L_i}$ also has no effect. Now, by Claim 2.7 we have that $J^1 \cup G_7^*$ is $(\mathcal{U}, \ell^3, s)$-good, and $\bigcup_{C \in \mathcal{C}'} \mathcal{M}_C$ is $(\mathcal{U}, \ell^3, s)$-good by (N2) combined with (R3). Finally, consider G_4^*. For each $x \in V'_{\mathrm{abs}} \cup V_{\mathrm{sc}}$, $i \in [K]$ and $k \in [2^{s+1}\Psi]$, let $e(x, i, k)$ be the edge between the left absorber tip and the third absorber vertex of $(C^l(x, i, k), C^r(x, i, k)) \in \mathcal{C}_1^{\mathrm{sc}} \cup \mathcal{C}_1^{\mathrm{abs}}$. Observe that, for all $x \in V'_{\mathrm{abs}} \cup V_{\mathrm{sc}}$ such that $\mathrm{dist}(x, \mathcal{U}_I) \geq 5$, all $i \in [K]$ and all $k \in [2^{s+1}\Psi]$, we have that $e(x, i, k)$ does not affect whether Q' is $(\mathcal{U}, 2\ell^2, s)$-good or not. In particular, by (T4), this is true for all $x \in V_{\mathrm{sc}}$. Now consider each $x \in V'_{\mathrm{abs}}$ such that $\mathrm{dist}(x, \mathcal{U}_I) < 5$. By Claim 2.8(ii), it follows that $e(x, i, k)$ only affects our claim when $x \in \mathcal{U}_I$. Observe that, for each $i \in [K]$ and $k \in [2^{s+1}\Psi]$, the direction of $e(x, i, k)$ is the same as that of the edge $e'(x, i, k)$ joining x to the right absorber tip of $(C^l(x, i, k), C^r(x, i, k))$. By Claim 2.8(iii), all cubes of absorbing ℓ-cube pairs in $\bigcup_{i \in [K]} \mathcal{C}_1^{\mathrm{abs}}(x, i)$ are vertex disjoint, which implies that each edge $e'(x, i, k)$ with $i \in [K]$ and $k \in [2^{s+1}\Psi]$ uses a different direction. Hence, G_4^* is $(\mathcal{U}, n, s)$-good, and the claim follows. ◀

Note that $T'' \subseteq I(Q')$, $R' \subseteq V(I)$, and $C \subseteq I(Q')$ for all $C \in \mathcal{C}'$. Recall the definitions of C'' from Step 6 and C''' from Step 10. For any $u \in \mathcal{U}$, recall the definitions of $\mathrm{End}(x)$ given in Step 8. Recall also the definitions of P, P' and $C^\bullet$ from Step 9. Combining all the previous steps, we claim that the following hold (conditioned on the events $\mathcal{E}_1^*, \dots, \mathcal{E}_8^*$, which occur a.a.s.).

(C1) $\Delta(T'') \leq 12D$.

(C2) Any vertex $x \in R' \cap V(T'')$ is a leaf of T''. Furthermore, if $x \in R' \cap V(T'')$, then $x \notin V(T)$ and its unique neighbour x' in T'' satisfies that $x' \in Z(x)$ (where $Z(x)$ is as defined in Step 7).

(C3) $B_I^5(\mathcal{U}_I) \cap V(T'') = \varnothing$.

(C4) For all $x \in V(I)$ we have that $|\mathcal{C}_x \cap \mathcal{C}'''| \geq (1 - 3/2\ell^4)n$.

(C5) For each $x \in V_{\mathrm{sc}}$ and $i \in [K]$, there is a collection $\mathcal{C}_1^{\mathrm{sc}}(x,i)$ of $2^{s+1}\Psi$ absorbing ℓ-cube pairs $(C^l(x,i,k), C^r(x,i,k))$ for x in I (defined in Step 9), each of which is associated with some edge $e \in M(A_i(x))$. In particular, $(C^l(x,i,k), C^r(x,i,k))$ is as described in (AB1) (recall also (AB2)), that is, there are two absorbing ℓ-cube pairs $(C_1^l(x,i,k), C_1^r(x,i,k))$ and $(C_2^l(x,i,k), C_2^r(x,i,k))$ in $H \cup G$, associated with $e \in M(A_i(x))$, for the clones x_1 and x_2 of x which correspond to (x,i). Moreover, each of these absorbing ℓ-cube pairs $(C^l(x,i,k), C^r(x,i,k))$ satisfies the following:

(C5.1) $(C_1^l(x,i,k), C_1^r(x,i,k)) \cup (C_2^l(x,i,k), C_2^r(x,i,k)) - V(\mathcal{M}_x) \subseteq Q'$;

(C5.2) the tips of $C^l(x,i,k)$ and $C^r(x,i,k)$ lie in $R' \setminus V(T'')$;

(C5.3) $C^l(x,i,k), C^r(x,i,k) \in \mathcal{C}'' \cap \mathcal{C}'''$, and

(C5.4) for any $x' \in V_{\mathrm{sc}}$, $i' \in [K]$ and $k' \in [2^{s+1}\Psi]$ with $(x',i',k') \neq (x,i,k)$, we have that $C^l(x,i,k)$, $C^r(x,i,k)$, $C^l(x',i',k')$ and $C^r(x',i',k')$ are vertex-disjoint.

(C6) For each $x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$ and $i \in [K]$, there is a collection $\mathcal{C}_1^{\mathrm{abs}}(x,i)$ of $2^{s+1}\Psi$ absorbing ℓ-cube pairs $(C^l(x,i,k), C^r(x,i,k))$ for x in I (defined in Step 10), each of which is associated with an edge $e \in M(A_i(x))$. In particular, $(C^l(x,i,k), C^r(x,i,k))$ is as described in (AB1) (recall also (AB2)), that is, there are two absorbing ℓ-cube pairs $(C_1^l(x,i,k), C_1^r(x,i,k))$ and $(C_2^l(x,i,k), C_2^r(x,i,k))$ in $H \cup G$, associated with $e \in M(A_i(x))$, for the clones x_1 and x_2 of x which correspond to (x,i). Moreover, each of these absorbing ℓ-cube pairs $(C^l(x,i,k), C^r(x,i,k))$ satisfies the following:

(C6.1) $(C_1^l(x,i,k), C_1^r(x,i,k)) \cup (C_2^l(x,i,k), C_2^r(x,i,k)) - V(\mathcal{M}_x) \subseteq Q'$;

(C6.2) the tips of $C^l(x,i,k)$ and $C^r(x,i,k)$ lie in R';

(C6.3) $C^l(x,i,k), C^r(x,i,k) \in \mathcal{C}'' \cap \mathcal{C}'''$;

(C6.4) for any $x' \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$, $i' \in [K]$ and $k' \in [2^{s+1}\Psi]$ with $(x',i',k') \neq (x,i,k)$, we have that $C^l(x,i,k)$, $C^r(x,i,k)$, $C^l(x',i',k')$ and $C^r(x',i',k')$ are vertex-disjoint, and

(C6.5) both $C^l(x,i,k)$ and $C^r(x,i,k)$ are vertex-disjoint from all cubes of absorbing ℓ-cube pairs in $\mathcal{C}_1^{\mathrm{sc}}$.

(C7) For every $x \in \mathcal{U}$ and every $z \in \mathrm{End}(x)$, there exists a collection of cubes $\mathcal{C}''(z)$ in $I(Q')$ with $|\mathcal{C}''(z)| = 20\ell$ which satisfies the following properties:

(C7.1) for every $C \in \mathcal{C}''(z)$, we have that $z \in V(C)$;

(C7.2) for every $C \in \mathcal{C}''(z)$, the molecule $\mathcal{M}_C$ is bonded in Q';

(C7.3) every $C \in \mathcal{C}''(z)$ meets T'';

(C7.4) for every $C_1, C_2 \in \mathcal{C}''(z)$, we have $V(C_1) \cap V(C_2) = \{z\}$, and

(C7.5) for every $C^* \in \mathcal{C}_1^{\mathrm{sc}}$ and $C \in \mathcal{C}''(z)$, we have $V(C) \cap V(C^*) = \varnothing$.

(C8) $C^\bullet$ intersects $V(T) \cap V(T'')$ and is different from all cubes described in (C5), (C6) and (C7).

Indeed, (C1) is given by (2.44). (C2) holds by (ET3) and the fact that $P' \cap V(T'') = \varnothing$. (C3) follows directly by (2.43). (C4) follows by combining (N1), the conditioning on $\mathcal{E}_5^*$, (2.46) and (R3). (C5) follows from the construction of P and T'' in Step 9. Indeed, (C5.1) follows from the definition of Q' combined with (AB1), and (C5.2) holds by the definition of R' and T'' combined with (AB1), while (C5.3) follows because of the definition of the set $M(A_i(x))$ in (AB2) and (AB3), and (C5.4) holds by (CD). Consider now (C6). For each $x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$ and $i \in [K]$, consider $\mathcal{C}_1^{\mathrm{abs}}(x,i)$. All absorbing ℓ-cube pairs of $\mathcal{C}_1^{\mathrm{abs}}(x,i)$ satisfy (C6.1) and (C6.2) by the definition of Q', (AB1) and (AB4). Similarly, they satisfy (C6.3) by (AB2), (AB3) and the fact that, by (AB4), their intersection with T'' contains their intersection with T'. Moreover, (C6.4) holds by Claim 2.8, and (C6.5) holds because of (AB4). Now, (C7) holds by Claim 2.6 and (T4). Indeed, let $\mathcal{C}''(z)$ be the collection of cubes given by Claim 2.6, so (C7.1), (C7.2) and (C7.4) follow directly. (C7.3) follows by using again the observation that, by (T4) and the construction of P, for any $x \in \mathcal{U}_I$, we have that T' and T'' coincide in $B_I^{2\ell}(x)$. Now recall that, by (T4), all vertices $x \in \mathcal{U}_I$ are at distance at least 5ℓ from V_{sc}, so (C7.5) follows by construction. Finally, consider (C8). The fact that $C^\bullet$ intersects $V(T) \cap V(T'')$ follows by its definition in Step 9, as does the fact that it is different from all cubes described in (C5). The fact that it is different from all cubes in (C6) follows by (AB4). Finally, the fact that it is different from the cubes in (C7) follows since $\mathrm{dist}(V(C^\bullet), \mathcal{U}_I) \geq \ell$ by the definition of $C^\bullet$ in Step 9.

Step 11: Fixing special absorbing structures. From this point onward, every step will be deterministic. Let $F \subseteq \mathcal{Q}^n$ be any graph with $\Delta(F) \leq \Psi$ which is $(\mathcal{U}, \ell, s)$-good, that is, for each $x \in \mathcal{U}$, the set $E_F(x) \coloneqq \{e \in E(F) : e \cap N_{\mathcal{Q}^n}(x) \neq \varnothing\}$ satisfies the following:

(F^*) for each layer L of $\mathcal{Q}^n$ and all $d \in \mathcal{D}(L)$, we have $|\{e \in E_F(x) : \mathcal{D}(e) = d\}| \leq n/\ell$.

Let $H' \subseteq \mathcal{Q}^n$ be any graph such that, for every $x \in \mathcal{U}$, we have $d_{H'}(x) \geq 2$. For each $x \in \mathcal{U}$, let $\{x, x + a(x)\}, \{x, x + b(x)\} \in E(H')$, where $a(x), b(x) \in \mathcal{D}(\mathcal{Q}^n)$. Our goal is to find a $(\mathcal{U}, \ell^2, s)$-good Hamilton cycle in $((H \cup G) \setminus F) \cup H' \cup Q'$ which, for each $x \in \mathcal{U}$, contains the edges $\{x, x + a(x)\}$ and $\{x, x + b(x)\}$. Recall that $\mathfrak{C}'(x, a(x), b(x))$ was defined in Step 8.

Claim 2.10. *For every $x \in \mathcal{U}$, there exists an $(x, a(x), b(x))$-consistent system of paths $CS(x) \in \mathfrak{C}'(x, a(x), b(x))$ such that $(E(CS(x)) \setminus \{\{x, x + a(x)\}, \{x, x + b(x)\}\}) \cap E(F) = \varnothing$.*

Proof. Suppose $x \in V(L)$, for some layer L. Suppose first that $x + a(x), x + b(x) \in V(L)$. Thus, we must show the existence of an $(x, a(x), b(x))$-consistent system of paths of Type I in $\mathfrak{C}'(x, a(x), b(x))$ with the desired property. Recall all the notation for consistent systems of paths introduced in section 2.7.2, as well as definition 2.7.6. By (CS), there is a collection $\mathcal{D}^{(2)}(x, a(x), b(x))$ of at least $\gamma n/2$ disjoint pairs of distinct directions

$c,d \in \mathcal{D}(L) \setminus \{a(x), b(x)\}$ such that, for each $(c,d) \in \mathcal{D}^{(2)}(x, a(x), b(x))$, there is a collection $\mathcal{D}^{(4)}(x, a(x), b(x), c, d)$ of at least $\gamma n/2$ disjoint 4-tuples of distinct directions in $\mathcal{D}(L) \setminus \{a(x), b(x), c, d\}$ satisfying the following: for each $(c,d) \in \mathcal{D}^{(2)}(x, a(x), b(x))$ and each $(d_1, d_2, d_3, d_4) \in \mathcal{D}^{(4)}(x, a(x), b(x), c, d)$, $\mathfrak{C}'(x, a(x), b(x))$ contains the $(x, a(x), b(x))$-consistent system of paths $CS(c, d, d_1, d_2, d_3, d_4)$ defined as in section 2.7.2. We will now show that there are many such consistent systems of paths which avoid F.

The choice of $(c,d) \in \mathcal{D}^{(2)}(x, a(x), b(x))$ determines six edges of the consistent system of paths: $e_1 := \{f(x + a(x)), f(x + a(x) + d)\}$, $e_2 := \{f(x + b(x)), f(x + b(x) + c)\}$, $e_3 := \{f(x+c), f(x)\}$, $e_4 := \{f(x), f(x+d)\}$, $e_5 := \{x+c, x+c+b(x)\}$ and $e_6 := \{x+d, x+d+a(x)\}$. Since $f(x)$, $f(x + a(x))$ and $f(x + b(x))$ are fixed and $\Delta(F) \leq \Psi$, for each $i \in [4]$ there are at most Ψ choices of $(c,d) \in \mathcal{D}^{(2)}(x, a(x), b(x))$ such that $e_i \in E(F)$. Furthermore, by (F^*), for each $i \in \{5, 6\}$ there are at most n/ℓ choices $(c,d) \in \mathcal{D}^{(2)}(x, a(x), b(x))$ such that $e_i \in E(F)$. Thus, there exist at least $\gamma n/2 - 4\Psi - 2n/\ell \geq \gamma n/4$ choices $(c,d) \in \mathcal{D}^{(2)}(x, a(x), b(x))$ such that $e_i \in E(H) \setminus E(F)$ for all $i \in [6]$. For any such choice of (c,d), the choice of $(d_1, d_2, d_3, d_4) \in \mathcal{D}^{(4)}(x, a(x), b(x), c, d)$ now determines the remaining eight edges of an $(x, a(x), b(x))$-consistent system of paths, each with a unique endpoint in $\{x + a(x), x + b(x), f(x + b(x)), x + c, f(x + c), f(x + d), x + d, f(x + a(x))\}$. It now follows by the fact that $\Delta(F) \leq \Psi$ that there are at most 8Ψ choices of $(d_1, d_2, d_3, d_4) \in \mathcal{D}^{(4)}(x, a(x), b(x), c, d)$ such that some of these eight edges lies in $E(F)$. In particular, we may fix a consistent system of paths $CS(x) \in \mathfrak{C}'(x, a(x), b(x))$ which satisfies the statement of the claim.

The cases where $x + a(x) \notin L$, $x + b(x) \in L$ and where $x + a(x), x + b(x) \notin L$ can be shown similarly. ◀

Note that $CS(x) \subseteq (H \setminus F) \cup H'$ for each $x \in \mathcal{U}$.

Claim 2.11. *For every $x \in \mathcal{U}$, we can extend $CS(x)$ into an $(x, a(x), b(x))$-special absorbing structure $SA(x)$ such that the following hold:*

(SA$_{\mathrm{i}}$) *for every $C \in \mathbf{C}(SA(x))$, we have that $\mathcal{M}_C \subseteq Q'$ and $\mathcal{M}_C$ is bonded in Q', and*

(SA$_{\mathrm{ii}}$) *every $C \in \mathbf{C}(SA(x))$ meets T''.*

Proof. For each $x \in \mathcal{U}$, we iterate through each $z \in \mathrm{endmol}(CS(x))$ fixing a cube $C(z) \in \mathcal{C}''(z)$. This will then determine $\mathbf{C}(SA(x))$, by taking the appropriate clones of $C(z)$. To see that this can be done, note that $|\mathrm{endmol}(CS(x))| \leq 6$. For each $z \in \mathrm{endmol}(CS(x))$, by (C7), there exist at least 20ℓ choices of $C(z) \in \mathcal{C}''(z)$ for which (SA$_\mathrm{i}$) and (SA$_\mathrm{ii}$) hold. Finally, by (C7.4) it follows that we can fix $C(z) \in \mathcal{C}''(z)$ such that $\mathcal{D}(C(z)) \cap \mathcal{D}(CS(x)) = \varnothing$. In particular, this implies that $C(z)$ is vertex-disjoint from all $C(z')$ already fixed with $z \neq z' \in \mathrm{endmol}(CS(x))$ and, therefore, this process forms a valid extension of $CS(x)$ into an $(x, a(x), b(x))$-special absorbing structure. ◀

For each $x \in \mathcal{U}$, let $SA(x)$ be an $(x, a(x), b(x))$-special absorbing structure which extends $CS(x)$, as determined by Claim 2.11. Note that, by (R3) and the fact that $V(SA(x)) \subseteq B^{2\ell}_{\mathcal{Q}^n}(x)$, the special absorbing structures in the collection $\{SA(x) : x \in \mathcal{U}\}$ are pairwise vertex-disjoint. Denote by $SA^v := \bigcup_{x\in\mathcal{U}} V(SA(x))$. Recall that, for any $C \in \mathbf{C}(SA(x))$, $C_I \subseteq I$ denotes the cube of which C is a clone. Given any tree $T^\bullet \subseteq I$ and any $x \in \mathcal{U}$, we say that $SA(x)$ *meets* $T^\bullet$ if, for all $C \in \mathbf{C}(SA(x))$, we have $V(C_I) \cap V(T^\bullet) \neq \varnothing$.

Recall that $\mathcal{C}'$ and $\mathcal{C}''$ were defined in Step 6. Let $\mathcal{C}^*_1 := \{C \in \mathcal{C}' : V(\mathcal{M}_C) \cap SA^v \neq \varnothing\}$. Note that, by (R3),

(CB) for each $x \in \mathcal{U}$, there are at most $2^{2\ell}$ ℓ-cubes $C \in \mathcal{C}^*_1$ such that $V(C) \cap B^{10\ell}(x) \neq \varnothing$.

Let $\mathcal{C}'_1 := \mathcal{C}' \setminus \mathcal{C}^*_1$ and $\mathcal{C}''_1 = \mathcal{C}'' \setminus \mathcal{C}^*_1$. We now define a tree $T''' \subseteq T''$ in the following way. Consider each $x \in R' \cap V(T'')$ such that $x \in V(C)$ for some $C \in \mathcal{C}'_1$. By (C2), we have that x has a unique neighbour x' in T'', and $x' \in Z(x)$. By the definition of $Z(x)$ (see Step 7), it follows that $x' \in V(C')$ for some $C' \in \mathcal{C}''$. If $C' \notin \mathcal{C}''_1$, then we remove x from T''. We denote the resulting tree by T'''. Let $\mathcal{C}'''_1 := \{C \in \mathcal{C}'_1 : V(C) \cap V(T''') \neq \varnothing\}$. By using (C1)–(C6), the definition of $\mathcal{C}'_1$, $\mathcal{C}''_1$ and $\mathcal{C}'''_1$, the construction of T''', and the maximum degree of F, we claim that the following now hold.

(C′1) $\Delta(T''') \leq 12D$.

(C′2) Any vertex $x \in R' \cap V(T''')$ is a leaf of T'''. Furthermore, if $x \in R' \cap V(T''')$, then $x \notin V(T)$ and its unique neighbour x' in T''' satisfies that $x' \in N_I(x) \cap V(T) \cap (\bigcup_{C \in \mathcal{C}_1''} V(C)) \subseteq Z(x)$.

(C′3) $B_I^5(\mathcal{U}_I) \cap V(T''') = \varnothing$.

(C′4) For all $x \in V(I)$ we have that $|\mathcal{C}_x \cap \mathcal{C}_1'''| \geq (1 - 2/\ell^4)n$.

(C′5) For each $x \in V_{\mathrm{sc}}$ and $i \in [K]$, there is an absorbing ℓ-cube pair $(C^l(x,i), C^r(x,i))$ for x in I, which is associated with some edge $e \in M(A_i(x))$. In particular, $(C^l(x,i), C^r(x,i))$ is such that there are two absorbing ℓ-cube pairs $(C_1^l(x,i), C_1^r(x,i))$ and $(C_2^l(x,i), C_2^r(x,i))$ in $H \cup G$, associated with $e \in M(A_i(x))$, for the clones x_1 and x_2 of x which correspond to (x,i). Additionally, each of these absorbing ℓ-cube pairs $(C^l(x,i), C^r(x,i))$ satisfies the following:

(C′5.1) $(C_1^l(x,i), C_1^r(x,i)) \cup (C_2^l(x,i), C_2^r(x,i)) - V(\mathcal{M}_x) \subseteq Q'$;

(C′5.2) the tips x^l of $C^l(x,i)$ and x^r of $C^r(x,i)$ lie in $R' \setminus V(T''')$, and $\{x, x^l\}, \{x, x^r\} \notin E(F_I)$; in particular, the tips x_1^l, x_1^r of $(C_1^l(x,i), C_1^r(x,i))$ and x_2^l, x_2^r of $(C_2^l(x,i), C_2^r(x,i))$ satisfy that $\{x_1, x_1^l\}, \{x_1, x_1^r\}, \{x_2, x_2^l\}, \{x_2, x_2^r\} \in E((H \cup G) \setminus F)$;

(C′5.3) $C^l(x,i), C^r(x,i) \in \mathcal{C}_1'' \cap \mathcal{C}_1'''$, and

(C′5.4) for any $x' \in V_{\mathrm{sc}}$ and $i' \in [K]$ with $(x', i') \neq (x, i)$, we have that $C^l(x,i)$, $C^r(x,i)$, $C^l(x',i')$ and $C^r(x',i')$ are vertex-disjoint.

Let $\mathcal{C}^{\mathrm{sc}}$ denote the collection of these absorbing ℓ-cube pairs.

(C′6) For each $x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$ and $i \in [K]$, there is an absorbing ℓ-cube pair $(C^l(x,i), C^r(x,i))$ for x in I, which is associated with an edge $e \in M(A_i(x))$. In particular, $(C^l(x,i), C^r(x,i))$ is such that there are two absorbing ℓ-cube pairs $(C_1^l(x,i), C_1^r(x,i))$ and $(C_2^l(x,i), C_2^r(x,i))$ in $H \cup G$, associated with $e \in M(A_i(x))$, for the clones x_1 and x_2 of x which correspond to (x,i). Moreover, each of these absorbing ℓ-cube pairs $(C^l(x,i), C^r(x,i))$ satisfies the following:

(C′6.1) $(C_1^l(x,i), C_1^r(x,i)) \cup (C_2^l(x,i), C_2^r(x,i)) - V(\mathcal{M}_x) \subseteq Q'$;

(C′6.2) the tips x^l of $C^l(x,i)$ and x^r of $C^r(x,i)$ lie in R', and $\{x,x^l\},\{x,x^r\} \notin E(F_I)$; in particular, the tips x_1^l, x_1^r of $(C_1^l(x,i), C_1^r(x,i))$ and x_2^l, x_2^r of $(C_2^l(x,i), C_2^r(x,i))$ satisfy that $\{x_1,x_1^l\},\{x_1,x_1^r\},\{x_2,x_2^l\},\{x_2,x_2^r\} \in E((H\cup G)\setminus F)$;

(C′6.3) $C^l(x,i), C^r(x,i) \in \mathcal{C}_1'' \cap \mathcal{C}_1'''$;

(C′6.4) for any $x' \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$ and $i' \in [K]$ with $(x',i') \neq (x,i)$, we have that $C^l(x,i)$, $C^r(x,i)$, $C^l(x',i')$ and $C^r(x',i')$ are vertex-disjoint, and

(C′6.5) both $C^l(x,i)$ and $C^r(x,i)$ are vertex-disjoint from all cubes of absorbing ℓ-cube pairs in $\mathcal{C}^{\mathrm{sc}}$.

Let $\mathcal{C}^{\neg\mathrm{sc}}$ denote the set of these absorbing ℓ-cube pairs.

(C′7) For every $x \in \mathcal{U}$, there is an $(x,a(x),b(x))$-consistent system of paths $CS(x)$ in $(H\setminus F)\cup H'$ which extends into an $(x,a(x),b(x))$-special absorbing structure $SA(x)$ which meets T''' and with the property that, for every $C \in \mathbf{C}(SA(x,a,b))$, we have that $\mathcal{M}_{C_I} \subseteq Q'$ and $\mathcal{M}_{C_I}$ is bonded in Q'. Moreover, $\{x,x+a(x)\},\{x,x+b(x)\} \in E(H')$.

(C′8) $C^\bullet$ intersects $V(T)\cap V(T''')$ (so, in particular, $C^\bullet \in \mathcal{C}_1'''$) and is different from all cubes described in (C′5), (C′6) and (C′7).

(C′9) $\bigcup_{C\in\mathcal{C}'''\setminus\mathcal{C}_1'''} V(C) \subseteq B_I^{3\ell}(\mathcal{U}_I)$.

Indeed, since $T''' \subseteq T''$, (C′1)–(C′3) follow immediately by (C1)–(C3), respectively. (C′4) follows from (C4), (R3) and (CB). Now, for each $x \in V_{\mathrm{sc}}$ and $i \in [K]$, consider the set $\mathcal{C}_1^{\mathrm{sc}}(x,i)$ described in (C5). We first remove from this set all absorbing ℓ-cube pairs any of whose cubes do not belong to $\mathcal{C}_1'$. Then, we remove all absorbing ℓ-cube pairs any of whose cubes do not intersect T'''. Finally, we remove all absorbing ℓ-cube pairs such that any of the edges joining its tips to x belong to F_I. Observe that by (CB) and (C′1) it follows that, for any $y \in V(I)$, we have $|B_I^{10\ell}(y)\cap(V(T'')\setminus V(T'''))| \leq 12D\cdot 2^{3\ell}$. Using this fact, (CB), and the fact that $\Delta(F) \leq \Psi$ (and, thus, $\Delta(F_I) \leq 2^s\Psi$), it follows that there is at least one absorbing ℓ-cube pair remaining in the collection. Let $(C^l(x,i), C^r(x,i))$ be such an absorbing ℓ-cube pair. Then, (C′5.2) and (C′5.3) hold by the choice above, and (C′5.1) and (C′5.4) hold by (C5.1) and (C5.4), respectively. For each $x \in V'_{\mathrm{abs}} \setminus V_{\mathrm{sc}}$ and $i \in [K]$, we proceed similarly from the set $\mathcal{C}_1^{\mathrm{abs}}(x,i)$ to fix an absorbing ℓ-cube pair $(C^l(x,i), C^r(x,i)) \in \mathcal{C}_1^{\mathrm{abs}}(x,i)$ which satisfies (C′6.2) and (C′6.3). Then, (C′6.1), (C′6.4) and (C′6.5) hold by (C6.1), (C6.4) and (C6.5), respectively. Furthermore, (C′7) holds by Claim 2.10, Claim 2.11 and the construction of T''' above. (Indeed, to see that each $SA(x)$ still meets T''', note that $(V(T'')\setminus V(T'''))\cap SA^v = \varnothing$.) For (C′8), by construction $C^\bullet \in \mathcal{C}'$ and $V(C^\bullet) \cap B_I^{3\ell/2}(\mathcal{U}_I) = \varnothing$. Therefore, $C^\bullet \in \mathcal{C}'_1$. The fact that $C^\bullet$ intersects $V(T) \cap V(T''')$ follows by (C8) and the fact that, in constructing T''', none of the leaves which are removed are vertices of T. Thus, in particular, $C^\bullet \in \mathcal{C}'''_1$. The rest of (C′8) follows immediately from (C8). Finally, (C′9) follows by the definition of $\mathcal{C}_1^*$. Indeed, consider the set $SA_I^v \subseteq V(I)$ of vertices such that each vertex in SA^v is a clone of some vertex in SA_I^v. It follows by construction (see section 2.7.2) that for any $x \in SA_I^v$ we have $\mathrm{dist}(x, \mathcal{U}_I) \leq \ell + 2$. The claim follows since any ℓ-cube $C \in \mathcal{C}_1^*$ must intersect SA_I^v and any two vertices in C are at distance at most ℓ.

Let $\mathcal{C}'_2 \coloneqq \bigcup_{x\in\mathcal{U}}\{C_I : C \in \mathbf{C}(SA(x))\}$ and $R'' \coloneqq R' \setminus \bigcup_{C\in\mathcal{C}'_2} V(C)$. Finally, let $\mathcal{C}'''_3 \coloneqq \mathcal{C}'''_1 \cup \mathcal{C}'_2$. Note that, by construction, any two cubes in $\mathcal{C}'''_3$ are vertex-disjoint.

Step 12: Constructing auxiliary trees $\mathbf{T}^*$ and $\boldsymbol{\tau_0}$. Let T^* be obtained from T''' by removing all leaves of T''' which lie in R''. In particular, by (C′2) and (C′8), we have that $C^\bullet$ intersects T^*.

We will now construct an auxiliary tree τ_0, which will be used in the construction of an almost spanning cycle. We start by defining an auxiliary multigraph Γ' as follows. First, let $\Gamma_1 \coloneqq T^* \cup \bigcup_{C\in\mathcal{C}'''_3} C$. Let Γ_2 be the graph obtained by iteratively removing all leaves from Γ_1 until all vertices have degree at least 2. Observe that, after this is achieved, the resulting graph still contains all cubes $C \in \mathcal{C}'''_3$. Let Γ_3 be obtained from Γ_2 by removing all connected components which consist of a single cube $C \in \mathcal{C}'''_3$. Now, let Γ' be the multigraph obtained by contracting each cube $C \in \mathcal{C}'''_3$ such that $C \subseteq \Gamma_3$ into a single vertex. We refer to the vertices resulting from contracting such cubes as *atomic vertices*, and to the remaining vertices in Γ' as *inner tree vertices*. Given $C \in \mathcal{C} \cup \mathcal{C}'_2$ and $j \in [2^s]$, we call $\mathcal{A} = \mathcal{M}_C \cap L_j$ an *atom*. We continue to identify each inner tree vertex v with the vertex $v \in V(I)$ from which it originated in Γ_1. Observe that Γ' is connected, and (C′1) implies that

(2.48) $$d_{\Gamma'}(v) \leq 12D \text{ for all inner tree vertices, and } \Delta(\Gamma') \leq 12 \cdot 2^{\ell} D.$$

Given an atomic vertex $v \in V(\Gamma')$, let $C(v) \in \mathcal{C}_3'''$ be the cube which was contracted to v in the construction of Γ', and let $\mathcal{M}(v) := \mathcal{M}_{C(v)}$. Furthermore, for each $j \in [2^s]$, let $\mathcal{A}_j(v) := \mathcal{M}(v) \cap L_j$. Similarly, for any $v \in V(\Gamma')$ which is an inner tree vertex, we define $\mathcal{M}(v) := \mathcal{M}_v$. Observe that every edge $e \in E(\Gamma')$ corresponds to a unique edge $e' \in I(Q')$. We say that *e originates* from e'. We denote by $D(e) \in \mathcal{D}(I)$ the direction of e' in I. By abusing notation, we will sometimes also view $D(e)$ as a direction in $\mathcal{Q}^n$.

Let $v_0 \in V(\Gamma')$ be the vertex which resulted from contracting $C^{\bullet}$. We define an auxiliary labelled rooted tree $\tau_0 = \tau_0(v_0)$ by performing a depth-first search on Γ' rooted at v_0 and then iteratively removing all leaves which are inner tree vertices. This results in a tree τ_0 rooted at an atomic vertex v_0 and all whose leaves are atomic vertices. Let $m := |V(\tau_0)| - 1$, and let the vertices of τ_0 be labelled as $v_0, v_1, \dots, v_m$, with the labelling given by the order in which each vertex is explored by the depth-first search performed on Γ'. For each $i \in [m]$, we define τ_i as the maximal subtree of τ_0 which contains v_i and all whose vertices have labels which are at least as large as i. Given any vertex $x \in V(I)$, we say that x is *represented* in τ_0 if $x \in V(\tau_0)$ or there exists some atomic vertex $v \in V(\tau_0)$ such that $x \in V(C(v))$. Similarly, we say that a cube $C \in \mathcal{C} \cup \mathcal{C}_2'$ is *represented* in τ_0 if there exists an atomic vertex $v \in V(\tau_0)$ such that $C = C(v)$. We will sometimes also say that $\mathcal{M}_x$ or $\mathcal{M}_C$ are represented in τ_0, respectively.

The tree τ_0 will be the backbone upon which we construct our near-spanning cycle. First, we need to set up some more notation. For each $i \in [m]_0$, let $p_i := d_{\tau_i}(v_i)$ and let $N_{\tau_i}(v_i) = \{u_1^i, \dots, u_{p_i}^i\}$. It follows from (2.48) that

(2.49) $$p_i \leq 12D - 1 \text{ if } v_i \text{ is an inner tree vertex, and } \Delta(\tau_0) \leq 12 \cdot 2^{\ell} D.$$

For each $i \in [m]_0$ and $k \in [p_i]$, let $e^i_k := \{v_i, u^i_k\}$, let $f^i_k := D(e^i_k)$, and let j^i_k be the label of u^i_k in τ_0, that is, $u^i_k = v_{j^i_k}$. For any $k \in [p_i]$, we will sometimes refer to i as the *parent index* of j^i_k. Furthermore, for each $i \in [m]_0$ such that v_i is an atomic vertex, and for each $k \in [p_i]$, consider the edge in $I(G)$ from which e^i_k originates and let ν^i_k be its endpoint in $C(v_i)$. Finally, for each $i \in [m]_0$, we define a parameter $\Delta(v_i)$ recursively by setting

$$\Delta(v_i) := \begin{cases} 0 & \text{if } v_i \text{ is an atomic vertex which is a leaf of } \tau_0, \\ \sum_{k=1}^{p_i} \Delta(u^i_k) & \text{if } v_i \text{ is an atomic vertex which is not a leaf of } \tau_0, \\ p_i + 1 + \sum_{k=1}^{p_i} \Delta(u^i_k) & \text{if } v_i \text{ is an inner tree vertex.} \end{cases} \tag{2.50}$$

This parameter $\Delta(v_i)$ will be used to keep track of parities throughout the following steps. Note that $\Delta(v_i)$ counts the number of times a depth first search of τ_i (starting and ending at v_i) traverses an inner tree vertex.

Consider the partition of all molecules into slices of size q introduced at the beginning of Step 3, where q is as defined in (2.33). Given any $v \in V(\tau_0)$, we denote the slices of its molecule by $\mathcal{M}_1(v), \dots, \mathcal{M}_t(v)$, where t is as defined in (2.33). Thus, for each $i \in [t]$ we have that $\mathcal{M}_i(v) = \bigcup_{j=(i-1)q+1}^{iq} \mathcal{A}_j(v)$. For each $i \in [m]_0$, we are going to assign an *input slice* $\mathcal{M}_{b(i)}(v_i)$ to each vertex v_i. We do so by recursively assigning an *input index* $b(i) \in [t]$ to each $i \in [m]_0$. We begin by letting $b(0) := 1$. Then, for each $i \in [m]_0$ and each $k \in [p_i]$, we set

$$b(j^i_k) := \begin{cases} b(i) & \text{if } v_i \text{ is an inner tree vertex,} \\ b(i) + k - 1 \pmod{t} & \text{if } v_i \text{ is an atomic vertex.} \end{cases}$$

Note that the bound on $\Delta(\tau_0)$ in (2.49) and the definition of t in (2.33) imply that $b(j^i_k) \neq b(j^i_{k'})$ whenever v_i is an atomic vertex and $k \neq k'$.

Step 13: Finding an external skeleton for $\mathbf{T}^*$. Our next goal is to find an almost spanning cycle in $(H \setminus F) \cup H' \cup Q'$ by using τ_0 to explore different molecules in a given order. We will construct this cycle in such a way that it already contains all vertices of $\mathcal{U}$.

For this, we are going to generate a *skeleton*; this will be an ordered list of vertices which we will denote by $\mathcal{L}$. In order to construct $\mathcal{L}$, we will construct disjoint *partial skeletons* $\mathcal{L}_i$ and $\hat{\mathcal{L}}_i$ for all $i \in [m]$ in an inductive way. Each of these skeletons will start and end in the input slice for the vertex v_i which is being considered. These partial skeletons will depend on the starting and ending vertices of $\mathcal{M}_{b(i)}(v_i)$ which are provided for each of them. Therefore, given two distinct starting vertices $x, \hat{x} \in V(\mathcal{M}_{b(i)}(v_i))$ and two distinct ending vertices $y, \hat{y} \in V(\mathcal{M}_{b(i)}(v_i))$, we will denote the partial skeletons by $\mathcal{L}_i(x, y)$ and $\hat{\mathcal{L}}_i(\hat{x}, \hat{y})$, respectively. Using all these partial skeletons, we will construct a first skeleton and then extend it by adding some more segments (see Step 17).

The first step in the construction of $\mathcal{L}$ is to construct a set of vertices $L^{\bullet}$, to which we will refer as an *external skeleton*, and for which we will in turn construct *partial external skeletons* in an inductive way. The external skeleton will be essential in determining which vertices will not be covered by the almost spanning cycle, and hence need to be absorbed. Roughly speaking, the external skeleton will contain

(i) all vertices where the almost spanning cycle enters and leaves each cube molecule represented in τ_0, and

(ii) all vertices which are not in cube molecules and are needed to connect cube molecules to each other (that is, some clones of inner tree vertices).

On the other hand, all vertices in a vertex molecule represented in τ_0 by an inner tree vertex which do not belong to the external skeleton will have to be absorbed.

For each $i \in [m]$, given the starting and ending vertices $x, y, \hat{x}, \hat{y} \in V(\mathcal{M}_{b(i)}(v_i))$ for $\mathcal{L}_i(x, y)$ and $\hat{\mathcal{L}}_i(\hat{x}, \hat{y})$, we will denote the corresponding partial external skeleton by $L_i^{\bullet}(x, y, \hat{x}, \hat{y})$.

The external skeleton is constructed recursively. The partial external skeletons are the result of each recursive step, assuming that the starting and ending points have been defined. Roughly speaking, for each $i \in [m]$, we will define partial external skeletons for any possible starting and ending vertices. The starting and ending vertices which we actually use are then fixed by the partial external skeleton whose index is the parent of i. Ultimately, all of them will be fixed when defining the external skeleton $L^\bullet$.

Let $\mathcal{M}_{\mathrm{Res}} \subseteq V(\mathcal{Q}^n)$ be the union of all the clones of R''. For each $x \in \mathcal{U}$, consider the graph $CS(x)_I \subseteq I$, and let $\mathcal{M}_{CS} \subseteq \mathcal{Q}^n$ be the union of all the clones of $\bigcup_{x\in\mathcal{U}} CS(x)_I$. We will construct an external skeleton $L^\bullet$ which satisfies the following properties:

(ES1) For each $i \in [m]$ such that v_i is an inner tree vertex, $L^\bullet \cap V(\mathcal{M}_{b(i)}(v_i))$ contains exactly $2p_i + 2$ vertices, half of them of each parity, and $L^\bullet \cap (V(\mathcal{M}(v_i)) \setminus V(\mathcal{M}_{b(i)}(v_i))) = \varnothing$.

(ES2) For each $i \in [m]$ such that v_i is an atomic vertex, $L^\bullet \cap V(\mathcal{M}(v_i))$ contains exactly $4p_i+4$ vertices. If v_i is not a leaf of τ_0, eight of these vertices (four of each parity) lie in $V(\mathcal{M}_{b(i)}(v_i))$, and four (two of each parity) lie in each $V(\mathcal{M}_{b(i)+k}(v_i))$ with $k \in [p_i-1]$. If v_i is a leaf, then all four of these vertices lie in $V(\mathcal{M}_{b(i)}(v_i))$.

(ES3) $L^\bullet \cap V(\mathcal{M}(v_0))$ contains exactly $4p_0$ vertices, four of them (two of each parity) lying in each $V(\mathcal{M}_k(v_0))$ with $k \in [p_0]$.

(ES4) The sets described in (ES1)–(ES3) partition $L^\bullet$.

(ES5) $L^\bullet \cap (\mathcal{M}_{\mathrm{Res}} \cup V(\mathcal{M}_{CS})) = \varnothing$.

We now proceed to define the partial external skeletons formally. The construction proceeds by induction on $i \in [m]$ in decreasing order, starting with $i = m$. We define a *valid connection sequence* $(x^i, y^i, \hat{x}^i, \hat{y}^i)$ *for* v_i as any set of distinct vertices $x^i, y^i, \hat{x}^i, \hat{y}^i \in V(\mathcal{M}_{b(i)}(v_i))$ which satisfy the following:

(V1) $x^i \neq_{\mathrm{p}} y^i$ if $\Delta(v_i)$ is even, and $x^i =_{\mathrm{p}} y^i$ otherwise;

(V2) $\hat{x}^i \neq_{\mathrm{p}} x^i$, and

(V3) $\hat{y}^i \neq_{\mathrm{p}} y^i$.

Given any valid connection sequence $(x^i, y^i, \hat{x}^i, \hat{y}^i)$, we will refer to x^i and $\hat{x}^i$ as *starting* vertices, and to y^i and $\hat{y}^i$ as *ending* vertices. Throughout the construction ahead, observe that, every time we use a partial external skeleton to build a larger one, its starting and ending vertices form a valid connection sequence by construction. The vertices x^i, y^i, etc. will be part of $\mathcal{L}_i(x^i, y^i)$, and the vertices $\hat{x}^i$, $\hat{y}^i$, etc. will be part of $\hat{\mathcal{L}}_i(\hat{x}^i, \hat{y}^i)$. The vertices x^i, y^i, $\hat{x}^i$, $\hat{y}^i$ will be used by the skeleton to move from the molecule represented by v_i in τ_0 to the molecule represented by its parent. Given these vertices, the following construction provides the vertices w^i_k and $\hat{w}^i_k$ (as well as z^i_k and $\hat{z}^i_k$, if applicable) which are used to move to molecules represented by the children of v_i. Given any vertices $(x, y, \hat{x}, \hat{y})$ in $\mathcal{Q}^n$ and any direction $f \in \mathcal{D}(\mathcal{Q}^n)$, we write $f + (x, y, \hat{x}, \hat{y}) = (f + x, f + y, f + \hat{x}, f + \hat{y})$.

Now suppose that $i \in [m]$ and that, for each $i' \in [m] \setminus [i]$, we have already constructed a partial external skeleton $L^{\bullet}_{i'}(x^{i'}, y^{i'}, \hat{x}^{i'}, \hat{y}^{i'})$ for $v_{i'}$ and every valid connection sequence $(x^{i'}, y^{i'}, \hat{x}^{i'}, \hat{y}^{i'})$ for $v_{i'}$. We will now construct a partial external skeleton for v_i and every valid connection sequence for v_i. We consider several cases.

Case 1: $v_i \in V(\tau_0)$ is a leaf of τ_0. Assume that $(x^i, y^i, \hat{x}^i, \hat{y}^i)$ is a valid connection sequence for v_i. Then, the partial external skeleton for this connection sequence is given by $L^{\bullet}_i(x^i, y^i, \hat{x}^i, \hat{y}^i) \coloneqq \{x^i, y^i, \hat{x}^i, \hat{y}^i\}$.

Case 2: $v_i \in V(\tau_0)$ is an inner tree vertex. We construct a set of partial external skeletons for v_i as follows.

1. Suppose $(x^i, y^i, \hat{x}^i, \hat{y}^i)$ is a valid connection sequence for v_i. Let $w^i_0 \coloneqq x^i$, $w^i_{p_i} \coloneqq y^i$, $\hat{w}^i_0 \coloneqq \hat{x}^i$ and $\hat{w}^i_{p_i} \coloneqq \hat{y}^i$. Let $W^i_0 \coloneqq \{w^i_0, w^i_{p_i}, \hat{w}^i_0, \hat{w}^i_{p_i}\}$.

2. For each $k \in [p_i - 1]$, iteratively choose two vertices $w^i_k, \hat{w}^i_k \in V(\mathcal{M}_{b(i)}(v_i)) \setminus W^i_{k-1}$ such that $f^i_k + (w^i_{k-1}, w^i_k, \hat{w}^i_{k-1}, \hat{w}^i_k)$ is a valid connection sequence for u^i_k, and let $W^i_k \coloneqq W^i_{k-1} \cup \{w^i_k, \hat{w}^i_k\}$.

Note that the definition of q in (2.33) and the bound on p_i in (2.49) ensure that we have sufficiently many vertices to choose from (similar comments apply in the other cases). Moreover, (2.50) implies that $f^i_{p_i} + (w^i_{p_i-1}, w^i_{p_i}, \hat{w}^i_{p_i-1}, \hat{w}^i_{p_i})$ is a valid connection sequence for

$u^i_{p_i}$. The partial external skeleton for v_i and connection sequence $(x^i, y^i, \hat{x}^i, \hat{y}^i)$ is defined as

$$L^\bullet_i(x^i, y^i, \hat{x}^i, \hat{y}^i) := \{x^i, \hat{x}^i\} \cup \bigcup_{k=1}^{p_i} \left(\{w^i_k, \hat{w}^i_k\} \cup L^\bullet_{j^i_k}(f^i_k + (w^i_{k-1}, w^i_k, \hat{w}^i_{k-1}, \hat{w}^i_k))\right).$$

Case 3: $v_i \in V(\tau_0)$ is an atomic vertex which is not a leaf. We construct a set of partial external skeletons for v_i as follows.

1. Assume $(x^i, y^i, \hat{x}^i, \hat{y}^i)$ is a valid connection sequence for v_i. Let $w^i_0 := x^i$.
2. For each $k \in [p_i]$, iteratively choose distinct vertices $z^i_k, w^i_k, \hat{z}^i_k, \hat{w}^i_k \in (V(\mathcal{M}_{b(i)+k-1}(v_i)) \cap V(\mathcal{M}_{\nu^i_k})) \setminus \{x^i, y^i, \hat{x}^i, \hat{y}^i\}$ satisfying that $z^i_k \neq_{\mathrm{p}} w^i_{k-1}$ and $f^i_k + (z^i_k, w^i_k, \hat{z}^i_k, \hat{w}^i_k)$ is a valid connection sequence for u^i_k.

Then, the partial external skeleton for v_i and connection sequence $(x^i, y^i, \hat{x}^i, \hat{y}^i)$ is defined as

$$L^\bullet_i(x^i, y^i, \hat{x}^i, \hat{y}^i) := \{x^i, y^i, \hat{x}^i, \hat{y}^i\} \cup \bigcup_{k=1}^{p_i} \left(\{z^i_k, w^i_k, \hat{z}^i_k, \hat{w}^i_k\} \cup L^\bullet_{j^i_k}(f^i_k + (z^i_k, w^i_k, \hat{z}^i_k, \hat{w}^i_k))\right).$$

After having constructed all these partial external skeletons for all v_i with $i \in [m]$, we are now ready to construct $L^\bullet$.

1. Choose any vertex $w^0_0 \in V(\mathcal{A}_1(v_0))$.
2. For each $k \in [p_0]$, iteratively choose four distinct vertices $z^0_k, \hat{z}^0_k, w^0_k, \hat{w}^0_k \in (V(\mathcal{M}_k(v_0)) \cap V(\mathcal{M}_{\nu^0_k}))$ satisfying that $z^0_k \neq_{\mathrm{p}} w^0_{k-1}$ and $f^0_k + (z^0_k, w^0_k, \hat{z}^0_k, \hat{w}^0_k)$ is a valid connection sequence for u^0_k.

Then, we define

$$L^\bullet := \bigcup_{k=1}^{p_0} \left(\{z^0_k, w^0_k, \hat{z}^0_k, \hat{w}^0_k\} \cup L^\bullet_{j^0_k}(f^0_k + (z^0_k, w^0_k, \hat{z}^0_k, \hat{w}^0_k))\right).$$

Observe that (ES1)–(ES4) hold by construction. In turn, (ES5) holds because of the definition of τ_0. Indeed, by (C′2) and (C′3) together with the definition of R'' and T^*, observe that $V(T^*) \cap (R'' \cup \bigcup_{x \in \mathcal{U}} V(CS(x)_I)) = \varnothing$. Moreover, by the construction above, all vertices in $L^\bullet$

are incident to some edge in a clone of the tree T^*, and thus, they cannot lie in $\mathcal{M}_{\mathrm{Res}} \cup V(\mathcal{M}_{CS})$.

Step 14: Constructing an auxiliary tree τ_0'. In order to extend the external skeleton into the skeleton and construct an almost spanning cycle, we first need to extend τ_0 to a new auxiliary tree τ_0' which encodes information about some additional molecules.

We construct τ_0' by appending some new leaves to τ_0. Note that τ_0 was built by encoding all the information about T^*, and τ_0' will encode the information about T'''. Recall the definition of $\mathcal{C}_3'''$ from the end of Step 11. In particular, by (C′2), each cube $C \in \mathcal{C}_3'''$ which does not intersect T^* contains at least one vertex u which is joined to T^* by an edge $e' = \{u, v\} \in E(T''')$ such that $v \in V(C')$, where $C \neq C' \in \mathcal{C}_1''$. Note that the construction of τ_0 implies that C' is represented in τ_0. For each such cube C, choose one such vertex u and append a new vertex to the atomic vertex representing C' in τ_0 via an edge e which originates as $e' \in E(T''')$. We say that this newly added vertex is atomic and *represents* C. The resulting tree after all these leaves are appended is τ_0'. In particular, $\tau_0 \subseteq \tau_0'$, and it follows that precisely the $C \in \mathcal{C}_3'''$ are represented in τ_0'. Furthermore, it follows from (C′1) that

$$
\begin{aligned}
&d_{\tau_0'}(v) \leq 12D \text{ for all } v \in V(\tau_0') \text{ which are inner tree vertices, and}\\
&\Delta(\tau_0') \leq 12 \cdot 2^{\ell} D.
\end{aligned}
\tag{2.51}
$$

For all vertices of τ_0', we will use the same notation for the vertices, cubes and molecules that they represent as we did for the vertices of τ_0. Note that, by (C′5.3) and (C′6.3),

(CP) every cube C belonging to some absorbing ℓ-cube pair in $\mathcal{C}^{\mathrm{sc}} \cup \mathcal{C}^{\neg\mathrm{sc}}$ is represented in τ_0'.

It will be important for us that τ_0' represents "most" vertices of the hypercube. In particular, for each $x \in V(I)$, let $\zeta(x)$ denote the number of vertices $y \in N_I(x)$ which are represented in τ_0' by atomic vertices. By (C′4), we have that

$$\zeta(x) \geq (1 - 2/\ell^4)n. \tag{2.52}$$

By an averaging argument, it follows that at least $(1 - 2/\ell^4)2^{n-s}$ vertices $x \in V(I)$ are represented in τ_0' by an atomic vertex. We will construct an almost spanning cycle in G which contains all the clones of these vertices.

Let $m' \coloneqq |V(\tau_0')| - 1$. Label $V(\tau_0') \setminus V(\tau_0) = \{v_{m+1}, \dots, v_{m'}\}$ arbitrarily. For each $i \in [m]$, we define τ_i' as the maximal subtree of τ_0' which contains v_i and all of whose vertices have labels at least as large as i. For each $i \in [m]_0$, let $p_i' \coloneqq d_{\tau_i'}(v_i)$ and let $N_{\tau_i'}(v_i) = \{u_1^i, \dots, u_{p_i'}^i\}$ (where the labelling is consistent with that of $N_{\tau_i}(v_i)$). For each $i \in [m]_0$ and $k \in [p_i'] \setminus [p_i]$, let $e_k^i \coloneqq \{v_i, u_k^i\}$, let $f_k^i \coloneqq D(e_k^i)$, and let j_k^i be the label of u_k^i in τ_0'. Furthermore, for each $i \in [m]_0$ such that v_i is an atomic vertex, and for each $k \in [p_i'] \setminus [p_i]$, consider the unique edge which e_k^i originates from in $I(G)$ and let ν_k^i be its endpoint in $C(v_i)$. Finally, for each $i \in [m'] \setminus [m]$ we set $\Delta(v_i) \coloneqq 0$.

As in Step 12, we consider the partition into slices for the new molecules arising from the newly added cubes represented by τ_0'. For each $i \in [m'] \setminus [m]$, we assign an input index $b(i) \in [t]$. To do so, for each $i \in [m]_0$ such that v_i is an atomic vertex and each $k \in [p_i'] \setminus [p_i]$, we set $b(j_k^i) \coloneqq b(i) + k - 1 \pmod t$. Similarly to Step 12, (2.33) and (2.51) imply that in this case $b(j_k^i) \neq b(j_{k'}^i)$ for all $k \neq k'$. For each $i \in [m'] \setminus [m]$, let ℓ_i be the label in τ_0' of the unique vertex adjacent to v_i (i.e., the parent label of i), and let m_i be the label of v_i in $N_{\tau_{\ell_i}'}(v_{\ell_i})$. Note that $b(i) = b(\ell_i) + m_i - 1$.

Step 15: Fixing absorbing ℓ-cube pairs for the vertices that need to be absorbed. At this point, we can determine every vertex in $V(\mathcal{Q}^n)$ that will have to be absorbed via absorbing ℓ-cube pairs into the almost spanning cycle we are going to construct. Recall from Step 11 that $SA^v = \bigcup_{y \in \mathcal{U}} V(SA(y))$. For every vertex $x \in V(I)$ not represented in τ_0', we will have to absorb all vertices in $\mathcal{M}_x \setminus SA^v$. Furthermore, for each $v \in V(\tau_0)$ which is an inner tree vertex, we will also need to absorb all vertices in $\mathcal{M}_v \setminus L^\bullet = \mathcal{M}_v \setminus (L^\bullet \cup SA^v)$. (The fact that $\mathcal{M}_v \cap SA^v = \varnothing$ follows by (C′3).) Recall the definition of V_{abs}' from Step 10. Let $V_{\mathrm{abs}} \subseteq V(I)$ be the set of all vertices which are not represented in τ_0' by an atomic vertex. Therefore, V_{abs} is the set of all vertices $x \in V(I)$ such that some clone of x needs to be absorbed. Moreover,

$V_{\text{abs}} = V(I) \setminus \bigcup_{C \in \mathcal{C}_3'''} V(C)$ and, thus, (C′9) and the definition of $\mathcal{C}_3'''$ at the end of Step 11 imply that $V_{\text{abs}} \subseteq V'_{\text{abs}}$. It follows from (2.52) that

$$|V_{\text{abs}}| \leq 2^{n-s+1}/\ell^4. \tag{2.53}$$

Now, for each $x \in V_{\text{abs}}$, we will pair all those vertices in $\mathcal{M}_x$ which need to be absorbed (each pair consisting of one vertex of each parity) and fix an absorbing ℓ-cube pair for each such pair of vertices. The absorbing ℓ-cube pair that we fix will be the one given by (C′5) or (C′6) for this pair of vertices, depending on whether $x \in V_{\text{sc}}$ or not.

For each $x \in V_{\text{abs}}$, let $S(x) := V(\mathcal{M}_x) \cap (L^{\bullet} \cup SA^{v}) = V(\mathcal{M}_x) \cap (L^{\bullet} \cup V(\mathcal{M}_{CS}))$. It follows by (ES1)–(ES5), (R3) and the definition of our special absorbing structures that $|S(x)| \leq 25D$ and $S(x)$ contains the same number of vertices of each parity. (Here we also use that $p_i \leq 12D - 1$ for every inner tree vertex v_i by (2.49) and (2.51).) Therefore, the matching $\mathfrak{M}(\mathcal{M}_x, S(x))$ defined in Step 3 is well defined, and we can use this matching to define our pairing of the vertices in $V(\mathcal{M}_x) \setminus S(x)$. Recall that each edge $e \in \mathfrak{M}(\mathcal{M}_x, S(x))$ gives rise to a unique index $i \in [K]$ via the relation $N(e) = A_i(x)$. (Here we ignore all those indices $i' \in [K]$ arising by artificially increasing the size of $\mathfrak{A}(x)$, see the beginning of Step 4.) For each $x \in V_{\text{abs}}$, let $\mathfrak{I}_x \subseteq [K]$ be the set of indices $i \in [K]$ which correspond to edges in $\mathfrak{M}(\mathcal{M}_x, S(x))$.

For each $x \in V_{\text{abs}}$ and $i \in \mathfrak{I}_x$, as stated in (C′5) and (C′6), we have already fixed an absorbing ℓ-cube pair for the clones of x corresponding to (x, i). Let

$$V^{\text{abs}} := \bigcup_{x \in V_{\text{abs}}} V(\mathcal{M}_x) \setminus (L^{\bullet} \cup SA^{v}).$$

(Thus, in particular, $V^{\text{abs}} \cap \mathcal{U} = \varnothing$.) As discussed above, this is the set of all vertices that need to be absorbed via absorbing ℓ-cube pairs. Recall that Q' was defined before (C1)–(C8). It follows from (C′5) and (C′6) that $((H \cup G) \setminus F) \cup Q'$ contains a set $\mathcal{C}^{\text{abs}} = \{(C^l(u), C^r(u)) : u \in V^{\text{abs}}\}$ of absorbing ℓ-cube pairs such that

(C_1) for all distinct $u, v \in V^{\mathrm{abs}}$, the absorbing ℓ-cube pairs $(C^l(u), C^r(u))$ and $(C^l(v), C^r(v))$ for u and v are vertex-disjoint and $(C^l(u), C^r(u)) \cup (C^l(v), C^r(v)) - \{u, v\} \subseteq Q'$;

(C_2) there exists a pairing $\mathcal{V} = \{f_1, \dots, f_{K'}\}$ of V^{abs} such that

($C_{2.1}$) for all $i \in [K']$, if $f_i = \{u_i, u_i'\}$, then $u_i \neq_{\mathrm{p}} u_i'$;

($C_{2.2}$) if $f_i = \{u_i, u_i'\}$, then there is a vertex $v \in V_{\mathrm{abs}}$ such that u_i and u_i' are clones of v which lie in either the same or consecutive slices of $\mathcal{M}_v$, and $(C^l(u_i), C^r(u_i))$ and $(C^l(u_i'), C^r(u_i'))$ are clones of the same absorbing ℓ-cube pair for v in I such that $(C^l(u_i), C^r(u_i))$ lies in the same layer as u_i and $(C^l(u_i'), C^r(u_i'))$ lies in the same layer as u_i';

($C_{2.3}$) if $u, u' \in V^{\mathrm{abs}}$ do not form a pair $f \in \mathcal{V}$, then $(C^l(u), C^r(u))$ and $(C^l(u'), C^r(u'))$ are clones of vertex-disjoint absorbing ℓ-cube pairs in I (except in the case when u, u' are clones of the same vertex $v \in V_{\mathrm{abs}}$, in which case $(C^l(u), C^r(u))$ and $(C^l(u'), C^r(u'))$ are clones of absorbing ℓ-cube pairs in I which intersect only in v);

(C_3) if we let $\mathcal{C}^* := \bigcup_{(C^l(u), C^r(u)) \in \mathcal{C}^{\mathrm{abs}}} \{C^l(u), C^r(u)\}$, then $\mathcal{C}^*$ contains either two or no clones of each cube $C \in \mathcal{C}_1'' \cap \mathcal{C}_1'''$, and every cube in $\mathcal{C}^*$ is a clone of some cube $C \in \mathcal{C}_1'' \cap \mathcal{C}_1'''$.

The pairing described in (C_2) is given by the matchings $\mathfrak{M}(\mathcal{M}_x, S(x))$. Furthermore, it follows from (C′5.2), (C′6.2) and (ES5) that

(C_4) the set of all tips of the absorbing ℓ-cube pairs in $\mathcal{C}^{\mathrm{abs}}$ is disjoint from $L^{\bullet}$.

We denote by $\mathfrak{L}$, $\mathfrak{R}_1$ and $\mathfrak{R}_2$ the collections of all left absorber tips, right absorber tips, and third absorber vertices, respectively, of the absorbing ℓ-cube pairs in $\mathcal{C}^{\mathrm{abs}}$. Observe that by ($C_1$)–($C_3$) the following properties are satisfied:

(C^*1) For all $i \in [m']_0$ such that v_i is an atomic vertex, we have that $|\mathfrak{L} \cap V(\mathcal{M}(v_i))| \in \{0, 2\}$. If $|\mathfrak{L} \cap V(\mathcal{M}(v_i))| = 2$, then these two vertices u, u' lie in different atoms of either the same or consecutive slices of $\mathcal{M}(v_i)$, and satisfy that $u \neq_{\mathrm{p}} u'$.

(C*2) For all $i \in [m']_0$ such that v_i is an atomic vertex, we have that $|(\mathfrak{R}_1 \cup \mathfrak{R}_2) \cap V(\mathcal{M}(v_i))| \in \{0, 4\}$. If $|(\mathfrak{R}_1 \cup \mathfrak{R}_2) \cap V(\mathcal{M}(v_i))| = 4$, then these four vertices form two pairs such that one vertex of each pair belongs to $\mathfrak{R}_1$ and the other to $\mathfrak{R}_2$. Each of these pairs lies in a different atom of the same or consecutive slices of $\mathcal{M}(v_i)$ and satisfies that its two vertices are adjacent in Q'.

(C*3) For all $i \in [m']_0$ such that v_i is an atomic vertex, if $\mathfrak{L} \cap V(\mathcal{M}(v_i)) \neq \emptyset$, then $(\mathfrak{R}_1 \cup \mathfrak{R}_2) \cap V(\mathcal{M}(v_i)) = \emptyset$.

(C*4) The sets described in (C*1) and (C*2) partition $\mathfrak{L}$ and $\mathfrak{R}_1 \cup \mathfrak{R}_2$, respectively.

Indeed, (C*1)–(C*3) follow from (C_1)–(C_3), and (C*4) follows by (CP).

For each $u \in V^{\mathrm{abs}}$, we denote the edge consisting of the right absorber tip and the third absorber vertex of $(C^l(u), C^r(u))$ by $e_{\mathrm{abs}}(u)$, and we denote by $\mathcal{P}^{\mathrm{abs}}(u)$ the path of length three formed by the third absorber vertex, the left absorber tip, u, and the right absorber tip, visited in this order. Note that $e_{\mathrm{abs}}(u) \in E(Q')$ by (C_1). Moreover, recall that $\mathcal{C}^{\mathrm{abs}}$ consists of absorbing ℓ-cube pairs in $((H \cup G) \setminus F) \cup Q')$. Thus, $\mathcal{P}^{\mathrm{abs}}(u) \subseteq ((H \cup G) \setminus F) \cup Q'$.

Step 16: Constructing the skeleton. We can now define the skeleton for the almost spanning cycle. Intuitively, this skeleton builds on the external skeleton by adding more structure that the cycle will have to follow. In particular, the skeleton adds the edges used to traverse from each slice in a cube molecule to its neighbouring slices, and it also incorporates the cube molecules represented in τ_0' which were not represented in τ_0. (The reason why these were not incorporated earlier is the following: if we already choose the valid connection sequences for these cube molecules in Step 13, then the tips of the absorbing cubes chosen in Step 15 might have non-empty intersection with the external skeleton, which we want to avoid, see (S7) below.) Furthermore, the skeleton gives an ordering to its vertices, and the cycle will visit the vertices of the skeleton in this order.

We will build a skeleton $\mathcal{L} = (x_1, \dots, x_r)$, for some $r \in \mathbb{N}$, and write $\mathcal{L}^\bullet := \{x_1, \dots, x_r\}$. We will construct $\mathcal{L}$ in such a way that the following properties hold:

(S1) For all distinct $k, k' \in [r]$, we have that $x_k \neq x_{k'}$.

(S2) $\{x_1, x_r\} \in E(Q')$.

(S3) For every $k \in [r-1]$, if x_k and x_{k+1} do not both lie in the same slice of a cube molecule represented in τ_0', then $\{x_k, x_{k+1}\} \in E(Q')$. Moreover, in this case, if x_{k+1} lies in a cube molecule represented in τ_0', then x_{k+2} lies in the same slice of this cube molecule as x_{k+1}.

(S4) For every $i \in [m']_0$ and every $j \in [t]$, no three consecutive vertices of $\mathcal{L}$ lie in $\mathcal{M}_j(v_i)$ (here $\mathcal{L}$ is viewed as a cyclic sequence of vertices).

(S5) For every $i \in [m']$ such that v_i is an atomic vertex and every $j \in [t]$, we have that $|V(\mathcal{M}_j(v_i)) \cap \mathcal{L}^\bullet|$ is even and $4 \leq |V(\mathcal{M}_j(v_i)) \cap \mathcal{L}^\bullet| \leq 12$. In particular, $|V(\mathcal{M}_t(v_0)) \cap \mathcal{L}^\bullet| = 4$.

(S6) For all $k \in [r]$ except two values, we have that $x_k \neq_{\mathrm{p}} x_{k+1}$. The remaining two values $k_1, k_2 \in [r]$ correspond to two pairs of vertices $x_{k_1}, x_{k_1+1}, x_{k_2}, x_{k_2+1} \in V(\mathcal{M}_t(v_0))$. For these two values, we have that $x_{k_1} \neq_{\mathrm{p}} x_{k_2}$ and either

  (i) $x_{k_1} =_{\mathrm{p}} x_{k_1+1}$ and $x_{k_2} =_{\mathrm{p}} x_{k_2+1}$, or

  (ii) $x_{k_1} \neq_{\mathrm{p}} x_{k_1+1}$ and $x_{k_2} \neq_{\mathrm{p}} x_{k_2+1}$,

where $x_{k_1}, x_{k_2} \in V(\mathcal{A}_{(t-1)q+1}(v_0))$ and $x_{k_1+1}, x_{k_2+1} \in V(\mathcal{A}_{tq}(v_0))$.

(S7) $\mathcal{L}^\bullet \cap (\mathfrak{L} \cup \mathfrak{R}_1 \cup V^{\mathrm{abs}} \cup V(\mathcal{M}_{CS})) = \varnothing$ and $L^\bullet \subseteq \mathcal{L}^\bullet$.

As happened with the external skeleton, the skeleton is built recursively from partial skeletons, which are defined first for the leaves. This recursive construction means that the overall order in which the molecules are visited will be determined by a depth first search of the tree τ_0'. Moreover, as discussed in section 2.2.5, for parity reasons the skeleton will actually traverse τ_0' twice. These two traversals will be "tied together" in the final step of the construction of the skeleton.

Note that, for each $i \in [m]$, the starting and ending vertices x^i, $\hat{x}^i$, y^i, $\hat{y}^i$ for the partial skeletons for v_i are determined by the external skeleton. For each $i \in [m'] \setminus [m]$, the starting and ending vertices for the partial skeletons of v_i will be determined when constructing the partial skeleton for the parent vertex v_{ℓ_i} of v_i. In particular, when constructing the partial skeleton for v_{ℓ_i}, we will define vertices $z^{\ell_i}_{m_i}, \hat{z}^{\ell_i}_{m_i}, w^{\ell_i}_{m_i}, \hat{w}^{\ell_i}_{m_i} \in \mathcal{M}_{b(i)}(v_{\ell_i})$. Then, the starting and ending vertices for the partial skeleton of v_i will be

$$(x^i, y^i, \hat{x}^i, \hat{y}^i) \coloneqq f^{\ell_i}_{m_i} + (z^{\ell_i}_{m_i}, w^{\ell_i}_{m_i}, \hat{z}^{\ell_i}_{m_i}, \hat{w}^{\ell_i}_{m_i}). \tag{2.54}$$

(Recall that ℓ_i, m_i, $b(i)$ and $f^{\ell_i}_{m_i}$ were defined at the end of Step 14.)

We are now in a position to define the partial skeletons formally. The construction proceeds by induction on $i \in [m']$ in decreasing order, starting with $i = m'$. Recall from the beginning of Step 13 that, for all $i \in [m]$, $x^i, y^i \in V(\mathcal{M}_{b(i)}(v_i))$ are the starting and ending vertices for the first partial skeleton $\mathcal{L}(x^i, y^i)$ for v_i, respectively, and $\hat{x}^i, \hat{y}^i \in V(\mathcal{M}_{b(i)}(v_i))$ are the starting and ending vertices for the second partial skeleton $\hat{\mathcal{L}}(\hat{x}^i, \hat{y}^i)$ for v_i, respectively. The vertices $x^i, y^i, \hat{x}^i, \hat{y}^i$ were fixed in the construction of the external skeleton, and they form a valid connection sequence. For each $i \in [m'] \setminus [m]$, the vertices $x^i, y^i, \hat{x}^i, \hat{y}^i \in V(\mathcal{M}_{b(i)}(v_i))$ defined in (2.54) will also form a valid connection sequence.

Let $\mathcal{F} \coloneqq \mathfrak{L} \cup \mathfrak{R}_1 \cup L^\bullet \cup V(\mathcal{M}_{CS})$. For each $k \in [2^s]$, let $\hat{e}_k$ be the direction of the edges in $\mathcal{Q}^n$ between L_k and L_{k+1}. Throughout the following construction, we will often choose vertices which are used to transition between neighbouring slices, all while avoiding the set $\mathcal{F}$. Similarly to the proof of lemma 2.7.11, all of these choices can be made by (ES2), (ES3), (C*1), (C*2), (R3), and because all cube molecules considered here are bonded in G_5 and, therefore, also in Q'. (The latter holds since for each atomic vertex $v \in V(\tau_0')$ the corresponding cube $C(v)$ satisfies $C(v) \in \mathcal{C}_3'''$.) Whenever we mention a vertex that we do not define here, we refer to the vertex with the same notation defined when constructing the external skeleton in Step 13.

Suppose that $i \in [m']$ and that for every $i' \in [m'] \setminus ([i] \cup [m])$ and every valid connection sequence $(x^{i'}, y^{i'}, \hat{x}^{i'}, \hat{y}^{i'})$ for $v_{i'}$ we have already defined two partial skeletons $\mathcal{L}(x^{i'}, y^{i'})$, $\hat{\mathcal{L}}(\hat{x}^{i'}, \hat{y}^{i'})$ for $v_{i'}$ with this connection sequence. (As discussed above, eventually we will only use the two partial skeletons for $v_{i'}$ with connection sequence as defined in (2.54).) Moreover, suppose that for every $i' \in [m] \setminus [i]$ we have already defined two partial skeletons $\mathcal{L}(x^{i'}, y^{i'})$, $\hat{\mathcal{L}}(\hat{x}^{i'}, \hat{y}^{i'})$ for $v_{i'}$ with connection sequence $(x^{i'}, y^{i'}, \hat{x}^{i'}, \hat{y}^{i'})$ (fixed by the external skeleton). If $i \in [m]$, let $(x^i, y^i, \hat{x}^i, \hat{y}^i)$ be the connection sequence for v_i fixed by the external skeleton. If $i \in [m'] \setminus [m]$, let $(x^i, y^i, \hat{x}^i, \hat{y}^i)$ be any connection sequence for v_i. We will now define the two partial skeletons for v_i with connection sequence $(x^i, y^i, \hat{x}^i, \hat{y}^i)$. We consider several cases.

Case 1: v_i is a leaf of τ_0'. We construct the partial skeletons as follows. Let $x_0^i \coloneqq x^i$ and $\hat{x}_0^i \coloneqq \hat{x}^i$. For each $k \in [t-1]_0$, iteratively choose any two vertices $y_k^i, \hat{y}_k^i \in V(\mathcal{A}_{(b(i)+k)q}(v_i)) \setminus (\mathcal{F} \cup \{x^i, y^i, \hat{x}^i, \hat{y}^i\})$ satisfying that

1. $y_k^i \neq_{\mathrm{p}} x_k^i$ and $\hat{y}_k^i \neq_{\mathrm{p}} \hat{x}_k^i$;
2. $x_{k+1}^i \coloneqq y_k^i + \hat{e}_{(b(i)+k)q} \notin \mathcal{F} \cup \{x^i, y^i, \hat{x}^i, \hat{y}^i\}$ and $\hat{x}_{k+1}^i \coloneqq \hat{y}_k^i + \hat{e}_{(b(i)+k)q} \notin \mathcal{F} \cup \{x^i, y^i, \hat{x}^i, \hat{y}^i\}$, and
3. $\{y_k^i, x_{k+1}^i\}, \{\hat{y}_k^i, \hat{x}_{k+1}^i\} \in E(Q')$.

Recall that we use $\bigtimes$ to denote the concatenation of sequences. The first and second partial skeletons for v_i with connection sequence $(x^i, y^i, \hat{x}^i, \hat{y}^i)$ are given by

$$\mathcal{L}_i(x^i, y^i) \coloneqq (x^i)\left(\bigtimes_{k=0}^{t-1}(y_k^i, x_{k+1}^i)\right)(y^i) \quad \text{and} \quad \hat{\mathcal{L}}_i(\hat{x}^i, \hat{y}^i) \coloneqq (\hat{x}^i)\left(\bigtimes_{k=0}^{t-1}(\hat{y}_k^i, \hat{x}_{k+1}^i)\right)(\hat{y}^i).$$

Case 2: $v_i \in V(\tau_0)$ is an inner tree vertex. Then, the first and second partial skeletons for v_i with connection sequence $(x^i, y^i, \hat{x}^i, \hat{y}^i)$ are defined as

$$\mathcal{L}_i(x^i, y^i) \coloneqq (x^i)\bigtimes_{k=1}^{p_i}(\mathcal{L}_{j_k^i}(x^{j_k^i}, y^{j_k^i}), w_k^i) \quad \text{and} \quad \hat{\mathcal{L}}_i(\hat{x}^i, \hat{y}^i) \coloneqq (\hat{x}^i)\bigtimes_{k=1}^{p_i}(\hat{\mathcal{L}}_{j_k^i}(\hat{x}^{j_k^i}, \hat{y}^{j_k^i}), \hat{w}_k^i),$$

where j_k^i was defined in Step 12.

Case 3: $v_i \in V(\tau_0)$ is an atomic vertex which is not a leaf of τ_0'. We construct the partial skeletons for v_i as follows. (Recall that, for each $k \in [p_i'] \setminus [p_i]$, the vertex ν_k^i was defined in Step 14.)

1. For each $k \in [p_i]$, iteratively choose distinct vertices $y_k^i, \hat{y}_k^i \in V(\mathcal{A}_{(b(i)+k-1)q}(v_i)) \setminus \mathcal{F}$ such that

   1.1. $y_k^i \neq_{\mathrm{p}} w_k^i$ and $\hat{y}_k^i \neq_{\mathrm{p}} \hat{w}_k^i$;

   1.2. $x_{k+1}^i := y_k^i + \hat{e}_{(b(i)+k-1)q} \notin \mathcal{F}$ and $\hat{x}_{k+1}^i := \hat{y}_k^i + \hat{e}_{(b(i)+k-1)q} \notin \mathcal{F}$, and

   1.3. $\{y_k^i, x_{k+1}^i\}, \{\hat{y}_k^i, \hat{x}_{k+1}^i\} \in E(Q')$.

2. If $p_i = 0$, let $x_1^i := x^i$ and $\hat{x}_1^i := \hat{x}^i$. For each $k \in [p_i'] \setminus [p_i]$, iteratively choose distinct vertices $z_k^i, w_k^i, \hat{z}_k^i, \hat{w}_k^i \in (V(\mathcal{M}_{b(i)+k-1}(v_i)) \cap V(\mathcal{M}_{\nu_k^i})) \setminus (\mathcal{F} \cup \{x_k^i, \hat{x}_k^i\})$ and distinct vertices $y_k^i, \hat{y}_k^i \in V(\mathcal{A}_{(b(i)+k-1)q}(v_i)) \setminus (\mathcal{F} \cup \{z_k^i, w_k^i, \hat{z}_k^i, \hat{w}_k^i\})$ satisfying that

   2.1. $z_k^i, \hat{w}_k^i \neq_{\mathrm{p}} x_k^i$ and $\hat{z}_k^i, w_k^i =_{\mathrm{p}} x_k^i$;

   2.2. $x^{j_k^i}, y^{j_k^i}, \hat{x}^{j_k^i}, \hat{y}^{j_k^i} \notin \mathcal{F}$, where $x^{j_k^i}$, $y^{j_k^i}$, $\hat{x}^{j_k^i}$ and $\hat{y}^{j_k^i}$ are defined as in (2.54);

   2.3. $y_k^i \neq_{\mathrm{p}} w_k^i$ and $\hat{y}_k^i \neq_{\mathrm{p}} \hat{w}_k^i$;

   2.4. $x_{k+1}^i := y_k^i + \hat{e}_{(b(i)+k-1)q} \notin \mathcal{F}$ and $\hat{x}_{k+1}^i := \hat{y}_k^i + \hat{e}_{(b(i)+k-1)q} \notin \mathcal{F}$, and

   2.5. $\{y_k^i, x_{k+1}^i\}, \{\hat{y}_k^i, \hat{x}_{k+1}^i\} \in E(Q')$.

   As discussed earlier, observe that a choice satisfying 2.2. exists by (C*1), (C*2), (ES2) and (R3).

3. For each $k \in [t] \setminus [p_i']$, iteratively choose distinct vertices $y_k^i, \hat{y}_k^i \in V(\mathcal{A}_{(b(i)+k-1)q}(v_i)) \setminus \mathcal{F}$ satisfying that

   3.1. $y_k^i \neq_{\mathrm{p}} x_k^i$ and $\hat{y}_k^i \neq_{\mathrm{p}} \hat{x}_k^i$;

   3.2. $x_{k+1}^i := y_k^i + \hat{e}_{(b(i)+k-1)q} \notin \mathcal{F}$ and $\hat{x}_{k+1}^i := \hat{y}_k^i + \hat{e}_{(b(i)+k-1)q} \notin \mathcal{F}$, and

   3.3. $\{y_k^i, x_{k+1}^i\}, \{\hat{y}_k^i, \hat{x}_{k+1}^i\} \in E(Q')$.

Then, we may define the first and second partial skeletons for v_i with connection sequence $(x^i, y^i, \hat{x}^i, \hat{y}^i)$ as

$$\mathcal{L}_i(x^i, y^i) := (x^i)\left(\bigtimes_{k=1}^{p_i'}(z_k^i, \mathcal{L}_{j_k^i}(x^{j_k^i}, y^{j_k^i}), w_k^i, y_k^i, x_{k+1}^i)\right)\left(\bigtimes_{k=p_i'+1}^{t}(y_k^i, x_{k+1}^i)\right)(y^i),$$

$$\hat{\mathcal{L}}_i(\hat{x}^i, \hat{y}^i) := (\hat{x}^i)\left(\bigtimes_{k=1}^{p_i'}(\hat{z}_k^i, \hat{\mathcal{L}}_{j_k^i}(\hat{x}^{j_k^i}, \hat{y}^{j_k^i}), \hat{w}_k^i, \hat{y}_k^i, \hat{x}_{k+1}^i)\right)\left(\bigtimes_{k=p_i'+1}^{t}(\hat{y}_k^i, \hat{x}_{k+1}^i)\right)(\hat{y}^i).$$

We are now ready to construct $\mathcal{L}$. The idea is similar to that of Case 3, except that we now tie together the first and second partial skeletons in Step 1.2 below.

1. Choose any two vertices $x_1^0, \hat{x}_1^0 \in V(\mathcal{A}_1(v_0)) \setminus \mathcal{F}$ such that

   1.1. $x_1^0 =_\mathrm{p} w_0^0$ and $\hat{x}_1^0 \neq_\mathrm{p} w_0^0$;

   1.2. $y_t^0 := \hat{x}_1^0 + \hat{e}_{2^s} \notin \mathcal{F}$ and $\hat{y}_t^0 := x_1^0 + \hat{e}_{2^s} \notin \mathcal{F}$, and

   1.3. $\{x_1^0, \hat{y}_t^0\}, \{\hat{x}_1^0, y_t^0\} \in E(Q')$.

2. For each $k \in [p_0]$, iteratively choose two distinct vertices $y_k^0, \hat{y}_k^0 \in V(\mathcal{A}_{kq}(v_0)) \setminus \mathcal{F}$ such that

   2.1. $y_k^0 \neq_\mathrm{p} w_k^0$ and $\hat{y}_k^0 \neq_\mathrm{p} \hat{w}_k^0$;

   2.2. $x_{k+1}^0 := y_k^0 + \hat{e}_{kq} \notin \mathcal{F}$ and $\hat{x}_{k+1}^0 := \hat{y}_k^0 + \hat{e}_{kq} \notin \mathcal{F}$, and

   2.3. $\{y_k^0, x_{k+1}^0\}, \{\hat{y}_k^0, \hat{x}_{k+1}^0\} \in E(Q')$.

3. For each $k \in [p_0']\setminus[p_0]$, iteratively choose distinct vertices $z_k^0, w_k^0, \hat{z}_k^0, \hat{w}_k^0 \in (V(\mathcal{M}_k(v_0)) \cap V(\mathcal{M}_{v_{k+1}^0}))\setminus(\mathcal{F}\cup\{x_k^0, \hat{x}_k^0\})$ and distinct vertices $y_k^0, \hat{y}_k^0 \in V(\mathcal{A}_{kq}(v_0))\setminus(\mathcal{F}\cup\{z_k^0, w_k^0, \hat{z}_k^0, \hat{w}_k^0\})$ satisfying that

   3.1. $z_k^0, \hat{w}_k^0 \neq_\mathrm{p} x_k^0$ and $\hat{z}_k^0, w_k^0 =_\mathrm{p} x_k^0$;

   3.2. $x^{j_k^0}, y^{j_k^0}, \hat{x}^{j_k^0}, \hat{y}^{j_k^0} \notin \mathcal{F}$, where $x^{j_k^0}$, $y^{j_k^0}$, $\hat{x}^{j_k^0}$ and $\hat{y}^{j_k^0}$ are defined as in (2.54);

   3.3. $y_k^0 \neq_\mathrm{p} w_k^0$ and $\hat{y}_k^0 \neq_\mathrm{p} \hat{w}_k^0$;

   3.4. $x_{k+1}^0 := y_k^0 + \hat{e}_{kq} \notin \mathcal{F}$ and $\hat{x}_{k+1}^0 := \hat{y}_k^0 + \hat{e}_{kq} \notin \mathcal{F}$, and

3.5. $\{y_k^0, x_{k+1}^0\}, \{\hat{y}_k^0, \hat{x}_{k+1}^0\} \in E(Q')$.

4. For each $k \in [t-1] \setminus [p_0']$, iteratively choose any two vertices $y_k^0, \hat{y}_k^0 \in V(\mathcal{A}_{kq}(v_0)) \setminus \mathcal{F}$ satisfying that

4.1. $y_k^0 \neq_{\mathrm{p}} x_k^0$ and $\hat{y}_k^0 \neq_{\mathrm{p}} \hat{x}_k^0$;

4.2. $x_{k+1}^0 \coloneqq y_k^0 + \hat{e}_{kq} \notin \mathcal{F}$ and $\hat{x}_{k+1}^0 \coloneqq \hat{y}_k^0 + \hat{e}_{kq} \notin \mathcal{F}$, and

4.3. $\{y_k^0, x_{k+1}^0\}, \{\hat{y}_k^0, \hat{x}_{k+1}^0\} \in E(Q')$.

The final definition of $\mathcal{L}$ is given by

$$\mathcal{L} \coloneqq (x_1^0)\left(\bigtimes_{k=1}^{p_0'}(z_k^0, \mathcal{L}_{j_k^0}(x^{j_k^0}, y^{j_k^0}), w_k^0, y_k^0, x_{k+1}^0)\right)\left(\bigtimes_{k=p_0'+1}^{t-1}(y_k^0, x_{k+1}^0)\right)(y_t^0, \hat{x}_1^0)$$
$$\left(\bigtimes_{k=1}^{p_0'}(\hat{z}_k^0, \hat{\mathcal{L}}_{j_k^0}(\hat{x}^{j_k^0}, \hat{y}^{j_k^0}), \hat{w}_k^0, \hat{y}_k^0, \hat{x}_{k+1}^0)\right)\left(\bigtimes_{k=p_0'+1}^{t-1}(\hat{y}_k^0, \hat{x}_{k+1}^0)\right)(\hat{y}_t^0).$$

Observe that (S1)–(S6) hold by construction. In particular, (2.50) together with (V1) ensure that in Case 3 the final two vertices of the two partial skeletons satisfy $x_{t+1}^i \neq_{\mathrm{p}} y^i$ and $\hat{x}_{t+1}^i \neq_{\mathrm{p}} \hat{y}^i$. Moreover, the pairs x_t^0, y_t^0 and $\hat{x}_t^0, \hat{y}_t^0$ will play the roles of the pairs x_{k_1}, x_{k_1+1} and x_{k_2}, x_{k_2+1} in the second part of (S6). Similarly, (S7) holds by combining the construction of $\mathcal{L}$ (more specifically the definition of the set $\mathcal{F}$), (C_4), (ES5) and the definition of V^{abs}.

Recall that we write $\mathcal{L} = (x_1, \dots, x_r)$. For each $i \in [m']_0$ such that v_i is an atomic vertex and each $j \in [t]$, let $\mathfrak{I}_{i,j} \coloneqq \{k \in [r] : x_k, x_{k+1} \in V(\mathcal{M}_j(v_i))\}$ and $S_{i,j} \coloneqq \{\{x_k, x_{k+1}\} : k \in \mathfrak{I}_{i,j}\}$.

Step 17. Incorporating special absorbing structures into the skeleton. In this step, we are going to incorporate all special absorbing structures fixed in Step 11 into the skeleton we just constructed. Note that all cube molecules referred to are represented by atomic vertices in τ_0', and all slices referred to are one of the t slices of each of these molecules defined in Step 3. For each $x \in \mathcal{U}$, consider the consistent system of paths $CS(x)$ and the special absorbing structure $SA(x)$ given by (C′7). By (S7), we have that $\mathcal{L}$ avoids $CS(x)$. Moreover, by the definition of $\mathcal{C}_3'''$ at the end of Step 11, each $C' \in \mathbf{C}(SA(x))$ is a clone of some $C \in \mathcal{C}_3'''$. Thus, by (S5) we have that $\mathcal{L}$ has positive intersection with each slice which contains a vertex of $\mathrm{end}(CS(x))$.

Recall that a special absorbing structure is a tuple of paths and cubes (see section 2.7.2). For each $z \in \mathcal{U}$, let P_1^z denote the first path of $SA(z)$. Let $x(z) \in \mathcal{L}^\bullet$ be the first vertex in $\mathcal{L}$ that is contained in the slice which contains the first vertex of P_1^z. Let $x'(z)$ be the successor of $x(z)$ in $\mathcal{L}$ (in particular, by (S3), both $x(z)$ and $x'(z)$ lie in the same slice). Now, for each $z \in \mathcal{U}$, depending on the type of the special absorbing structure $SA(z)$, we will update $\mathcal{L}$ in different ways.

(I) If $SA(z)$ is a special absorbing structure of Type I, proceed as follows. Let $P_1^z, \dots, P_6^z$ be the six paths of $SA(z)$. Let $S := \bigtimes_{i=1}^{6} P_i^z$ and let S^{-1} denote the same sequence of vertices in reverse order. If $x(z)$ has opposite parity to the initial vertex of P_1^z, then we replace the segment $(x(z), x'(z))$ of $\mathcal{L}$ by $(x(z), S, x'(z))$; otherwise, we we replace the segment $(x(z), x'(z))$ by $(x(z), S^{-1}, x'(z))$.

(II) If $SA(z)$ is a special absorbing structure of Type II, we proceed as follows. Let P_1^z and P_2^z be the two paths of $SA(z)$. Let y^1 and x_0^1 be the first and last vertices of P_1^z, and let y^2 and x_0^2 be the first and last vertices of P_2^z, respectively. Let $v, v' \in V(\tau_0')$ and $t_1, t_2, t_3 \in [t]$ be such that $y^1 \in V(\mathcal{M}_{t_1}(v))$, $x_0^1 \in V(\mathcal{M}_{t_2}(v'))$, and $y^2 \in V(\mathcal{M}_{t_3}(v'))$ (this implies $x_0^2 \in V(\mathcal{M}_{t_3}(v))$).

We now define two sequences of vertices S_1^z and S_2^z following similar ideas to Step 16. Recall that, for each $i \in [2^s]$, we use $\hat{e}_i$ to denote the direction of the edges between L_i and L_{i+1}. If $t_3 \geq t_2$, let $m_1 := t_3 - t_2$; otherwise, let $m_1 := t - (t_2 - t_3)$. For each $k \in [m_1 - 1]_0$, iteratively choose a vertex $y_k^1 \in V(\mathcal{A}_{(t_2+k)q}(v')) \setminus \mathcal{L}^\bullet$ satisfying that

1. $y_k^1 \neq_{\mathrm{p}} x_k^1$;
2. $x_{k+1}^1 := y_k^1 + \hat{e}_{(t_2+k)q} \notin \mathcal{L}^\bullet$, and
3. $\{y_k^1, x_{k+1}^1\} \in E(Q')$.

We set $S_1^z := \bigtimes_{k=0}^{m_1-1} (y_k^1, x_{k+1}^1)$.

In order to construct S_2^z, we proceed similarly. If $t_3 \geq t_1$, let $m_2 \coloneqq t_3 - t_1$; otherwise, let $m_2 \coloneqq t - (t_1 - t_3)$. For each $k \in [m_2 - 1]_0$, iteratively choose a vertex $y_k^2 \in V(\mathcal{A}_{(t_3-k-1)q+1}(v)) \setminus \mathcal{L}^\bullet$ satisfying that

1. $y_k^2 \neq_{\mathrm{p}} x_k^2$;

2. $x_{k+1}^2 \coloneqq y_k^2 + \hat{e}_{(t_3-k-1)q} \notin \mathcal{L}^\bullet$, and

3. $\{y_k^2, x_{k+1}^2\} \in E(Q')$.

Now, let $S_2^z \coloneqq \bigtimes_{k=0}^{m_2-1}(y_k^2, x_{k+1}^2)$.

Let $S \coloneqq P_1^z \times S_1^z \times P_2^z \times S_2^z$, and let S^{-1} denote the same sequence of vertices in reverse order. Finally, we replace the segment $(x(z), x'(z))$ of $\mathcal{L}$ by $(x(z), S, x'(z))$ if $x(z)$ has parity opposite to the initial vertex of P_1^z; otherwise, we replace $(x(z), x'(z))$ by $(x(z), S^{-1}, x'(z))$.

(III) If $SA(z)$ is a special absorbing structure of Type III, we proceed as follows. Let P_1^z, P_2^z and P_3^z be the three paths of $SA(z)$. For each $i \in [3]$, let y^i and x_0^i be the first and last vertices of P_i^z, respectively. Let $v_1, v_2, v_3 \in V(\tau_0')$ and $t_1, t_2 \in [t]$ be such that $y^1 \in V(\mathcal{M}_{t_1}(v_1))$, $y^2 \in V(\mathcal{M}_{t_1}(v_2))$ and $y^3 \in V(\mathcal{M}_{t_2}(v_3))$ (note this implies that $x_0^1 \in V(\mathcal{M}_{t_2}(v_2))$, $x_0^2 \in V(\mathcal{M}_{t_1}(v_3))$ and $x_0^3 \in V(\mathcal{M}_{t_2}(v_1))$).

For each $i \in [3]$, we define a sequence S_i^z as follows. If $t_2 \geq t_1$, let $m^* \coloneqq t_2 - t_1$; otherwise, let $m^* \coloneqq t - (t_1 - t_2)$. For each $k \in [m^* - 1]_0$, iteratively choose three vertices $y_k^1 \in V(\mathcal{A}_{(t_2-k-1)q+1}(v_2)) \setminus \mathcal{L}^\bullet$, $y_k^2 \in V(\mathcal{A}_{(t_1+k)q}(v_3)) \setminus \mathcal{L}^\bullet$ and $y_k^3 \in V(\mathcal{A}_{(t_2-k-1)q+1}(v_1)) \setminus \mathcal{L}^\bullet$ satisfying that

1. $y_k^i \neq_{\mathrm{p}} x_k^i$ for all $i \in [3]$;

2. $x_{k+1}^1 \coloneqq y_k^1 + \hat{e}_{(t_2-k-1)q} \notin \mathcal{L}^\bullet$, $x_{k+1}^2 \coloneqq y_k^2 + \hat{e}_{(t_1+k)q} \notin \mathcal{L}^\bullet$ and $x_{k+1}^3 \coloneqq y_k^3 + \hat{e}_{(t_2-k-1)q} \notin \mathcal{L}^\bullet$, and

3. $\{y_k^i, x_{k+1}^i\} \in E(Q')$ for all $i \in [3]$.

Then, for each $i \in [3]$, we define $S_i^z \coloneqq \bigtimes_{k=0}^{m^*-1}(y_k^i, x_{k+1}^i)$.

Let $S \coloneqq \bigtimes_{i=1}^{3}(P_i^z \times S_i^z)$, and let S^{-1} denote the same sequence in reverse order. Finally, we replace the segment $(x(z), x'(z))$ of $\mathcal{L}$ by $(x(z), S, x'(z))$ if $x(z)$ has parity opposite to the initial vertex of P_1^z; otherwise, we replace $(x(z), x'(z))$ by $(x(z), S^{-1}, x'(z))$.

Write $\mathcal{L} = (x_1, \dots, x_r)$, for some $r \in \mathbb{N}$, for the extended skeleton into which all the special absorbing structures $SA(z)$ for $z \in \mathcal{U}$ have been incorporated, and let $\mathcal{L}^\bullet \coloneqq \{x_1, \dots, x_r\}$. It follows from (S1)–(S7) and the construction above (together with the choice of v_0 in Step 12) that the following properties hold:

(S′1) For all distinct $k, k' \in [r]$, we have that $x_k \neq x_{k'}$.

(S′2) $\{x_1, x_r\} \in E(Q')$.

(S′3) For every $k \in [r-1]$, if x_k and x_{k+1} do not both lie in the same slice of a cube molecule represented in τ_0', then $\{x_k, x_{k+1}\} \in E((H \setminus F) \cup H' \cup Q')$. Moreover, in this case $\{x_k, x_{k+1}\} \in E(Q')$ unless both x_k and x_{k+1} lie in SA^v.

(S′4) For every $i \in [m']_0$ and every $j \in [t]$, no three consecutive vertices of $\mathcal{L}$ lie in $\mathcal{M}_j(v_i)$ (here $\mathcal{L}$ is viewed as a cyclic sequence of vertices).

(S′5) For every $i \in [m']$ such that v_i is an atomic vertex and every $j \in [t]$, we have that $|V(\mathcal{M}_j(v_i)) \cap \mathcal{L}^\bullet|$ is even and $4 \leq |V(\mathcal{M}_j(v_i)) \cap \mathcal{L}^\bullet| \leq 14$. Moreover, $|V(\mathcal{M}_t(v_0)) \cap \mathcal{L}^\bullet| = 4$.

(S′6) For all $k \in [r]$ except two values, we have that $x_k \neq_{\mathrm{p}} x_{k+1}$. The remaining two values $k_1, k_2 \in [r]$ correspond to two pairs of vertices $x_{k_1}, x_{k_1+1}, x_{k_2}, x_{k_2+1} \in V(\mathcal{M}_t(v_0))$. For these two values, we have that $x_{k_1} \neq_{\mathrm{p}} x_{k_2}$ and either

  (i) $x_{k_1} =_{\mathrm{p}} x_{k_1+1}$ and $x_{k_2} =_{\mathrm{p}} x_{k_2+1}$, or

  (ii) $x_{k_1} \neq_{\mathrm{p}} x_{k_1+1}$ and $x_{k_2} \neq_{\mathrm{p}} x_{k_2+1}$,

where $x_{k_1}, x_{k_2} \in V(\mathcal{A}_{(t-1)q+1}(v_0))$ and $x_{k_1+1}, x_{k_2+1} \in V(\mathcal{A}_{tq}(v_0))$.

(S′7) $\mathcal{L}^\bullet \cap (\mathfrak{L} \cup \mathfrak{R}_1 \cup V^{\mathrm{abs}}) = \varnothing$ and $L^\bullet \subseteq \mathcal{L}^\bullet$.

Indeed, for (S′3)–(S′6) we make use of the properties of the paths P_j^z defined in section 2.7.2 as well as (AS) (see section 2.7.2) and (C′7). We also use that the set of all cube molecules represented in τ_0' is precisely $\mathcal{C}_3''' = C_1''' \cup \mathcal{C}_2'$. Recall that, by the definition of $\mathcal{C}_1'''$, for each $C \in \mathcal{C}_1'''$ we have $V(\mathcal{M}_C) \cap SA^v = \varnothing$. Moreover, by (R3) and (AS), for each $C \in \mathcal{C}_2'$ we have $|V(\mathcal{M}_C) \cap SA^v| = 2$, and these two vertices $x_C^1, x_C^2 \in V(\mathcal{M}_C) \cap SA^v$ satisfy the following properties:

(i) there is some $z \in \mathcal{U}$ and two consecutive paths P_i^z, P_{i+1}^z in $SA(z)$ (with indices taken cyclically) such that x_C^1 is the final vertex of P_i^z and x_C^2 is the first vertex of P_{i+1}^z, and

(ii) $\mathbf{C}(SA(z))$ contains two clones C_1, C_2 of C, where $x_C^1 \in V(C_1)$ and $x_C^2 \in V(C_2)$.

Moreover, to check (S′5) for case (I), note that the definition of f in a consistent system of paths of Type I in section 2.7.2 implies that, for each $i \in [6]$, the final vertex of P_i^z and the first vertex of P_{i+1}^z lie in the same slice.

Step 18: Constructing an almost spanning cycle. We will now apply the connecting lemmas to obtain an almost spanning cycle in $(H \setminus F) \cup H' \cup Q'$ from $\mathcal{L} = (x_1, \dots, x_r)$. For each $i \in [m']$ such that v_i is an atomic vertex, by (C*1) there is at most one value $k(i) \in [t]$ such that $|\mathfrak{L} \cap V(\mathcal{M}_{k(i)}(v_i))| = 1$ and $|\mathfrak{L} \cap V(\mathcal{M}_{k(i)+1}(v_i))| = 1$. If such $k(i)$ exists, then we denote by $k^*(i)$ an additional index not in $[t]$ and let $\mathfrak{I}_{i,k^*(i)} \coloneqq \mathfrak{I}_{i,k(i)} \cup \mathfrak{I}_{i,k(i)+1}$, $S_{i,k^*(i)} \coloneqq S_{i,k(i)} \cup S_{i,k(i)+1}$ and $\mathcal{M}_{k^*(i)}(v_i) \coloneqq \mathcal{M}_{k(i)}(v_i) \cup \mathcal{M}_{k(i)+1}(v_i)$. Let

$$\mathfrak{T}(i) \coloneqq \begin{cases} [t] & \text{if there is no } k(i) \text{ as above,} \\ ([t] \cup \{k^*(i)\}) \setminus \{k(i), k(i)+1\} & \text{otherwise.} \end{cases}$$

Observe that the definition of v_0 in Step 12 together with (C′8) ensures that $\mathfrak{T}(0) = [t]$.

For each $i \in [m']_0$ such that v_i is an atomic vertex and for each $j \in \mathfrak{T}(i)$, except the pair $(0, t)$, we apply lemma 2.7.11 to the slice $\mathcal{M}_j(v_i)$ and the graph Q', with $\mathfrak{L} \cap V(\mathcal{M}_j(v_i))$, $(\mathfrak{R}_1 \cup \mathfrak{R}_2) \cap V(\mathcal{M}_j(v_i))$ and $S_{i,j}$ playing the roles of L, R and the pairs of vertices described in lemma 2.7.11(C3), respectively. Note that the conditions of lemma 2.7.11 can be verified as follows. (C1) and (C2) hold by (C*1) and (C*2) combined with (C*3). (C3) holds by (S′1) and (S′3)–(S′7). For $\mathcal{M}_t(v_0)$, we apply lemma 2.7.11 or lemma 2.7.12 depending on whether (ii) or (i) holds in (S′6) (the conditions for lemma 2.7.12 can be checked analogously). For each $i \in [m']_0$ such that v_i is an atomic vertex and each $j \in \mathfrak{T}(i)$, this yields $|\mathfrak{I}_{i,j}|$ vertex-disjoint paths $(\mathcal{P}_k^{i,j})_{k\in\mathfrak{I}_{i,j}}$ in $\mathcal{M}_j(v_i) \cup Q' = Q'$ such that, for each $k \in \mathfrak{I}_{i,j}$,

(i) $\mathcal{P}_k^{i,j}$ is an (x_k, x_{k+1})-path,

(ii) $\bigcup_{k\in\mathfrak{I}_{i,j}} V(\mathcal{P}_k^{i,j}) = V(\mathcal{M}_j(v_i)) \setminus \mathfrak{L}$, and

(iii) any pair of right absorber tip and third absorber vertex in $\mathfrak{R}_1 \cup \mathfrak{R}_2$ contained in the same atom of $\mathcal{M}_j(v_i)$ form an edge in one of the paths.

Now consider the path obtained as follows by going through $\mathcal{L}$. Start with x_1. For each $k \in [r]$, if there exist $i \in [m']_0$ and $j \in \mathfrak{T}(i)$ such that $\{x_k, x_{k+1}\} \in S_{i,j}$, add $\mathcal{P}_k^{i,j}$ to the path; otherwise, add the edge $\{x_k, x_{k+1}\}$ (this must be an edge of $(H \setminus F) \cup H' \cup Q'$ by (S′3)). Finally, add the edge $\{x_r, x_1\}$ of Q' (this is given by (S′2)) to the path to close it into a cycle $\mathfrak{H}$ in $(H \setminus F) \cup H' \cup Q'$. This cycle $\mathfrak{H}$ satisfies the following properties (recall that $e_{\mathrm{abs}}(u)$ was defined at the end of Step 15):

(HC1) $|V(\mathfrak{H})| \geq (1 - 4/\ell^4)2^n$.

(HC2) $V(\mathfrak{H}) \mathbin{\dot\cup} \mathfrak{L} \mathbin{\dot\cup} V^{\mathrm{abs}}$ partitions $V(\mathcal{Q}^n)$.

(HC3) For all $u \in V^{\mathrm{abs}}$, we have that $e_{\mathrm{abs}}(u) \in E(\mathfrak{H})$.

(HC4) For all $x \in \mathcal{U}$, we have that $\{x, x + a(x)\}, \{x, x + b(x)\} \in E(\mathfrak{H})$.

(HC5) For all $x \in V(\mathfrak{H}) \setminus SA^{\upsilon}$, each of the two edges of $\mathfrak{H}$ incident to x lies in Q'.

Indeed, note that $\mathfrak{H}$ covers all vertices in $L^{\bullet}$ (since $L^{\bullet} \subseteq \mathcal{L}^{\bullet}$ by (S′7)) as well as all vertices lying in cube molecules represented in τ_0' except for those in $\mathfrak{L}$ (by (ii)). Together with the definition of V^{abs}, this implies (HC2). Moreover, since $|\mathfrak{L}| = |V^{\mathrm{abs}}|$, (HC1) follows from (2.53). (HC3) follows by (iii), and (HC4) follows immediately by the definition of P_1 in each of the three types of special absorbing structures defined in section 2.7.2. Finally, (HC5) follows from (S′3).

Step 19: Absorbing vertices to form a Hamilton cycle. For each $u \in V^{\mathrm{abs}}$, replace the edge $e_{\mathrm{abs}}(u)$ by the path $\mathcal{P}_{\mathrm{abs}}(u)$ (recall from the end of Step 15 that $\mathcal{P}_{\mathrm{abs}}(u)$ lies in $((H \cup G) \setminus F) \cup Q')$. Clearly, this incorporates all vertices of $\mathfrak{L} \cup V^{\mathrm{abs}}$ into the cycle and, by (HC2) and (HC3), the resulting cycle $\mathfrak{H}'$ is Hamiltonian. Moreover, since by (C_3) the endvertices of each edge $e_{\mathrm{abs}}(u)$ lie in cubes belonging to $\mathcal{C}_1'''$, all these endvertices avoid $\mathcal{U}$. Thus, by (HC4), for each $x \in \mathcal{U}$ the edges at x in $\mathfrak{H}'$ are still $\{x, x + a(x)\}$ and $\{x, x + b(x)\}$, and so, in particular, by (C′7) these edges belong to H'.

It now remains to show that $\mathfrak{H}'$ is $(\mathcal{U}, \ell^2, s)$-good. Fix any vertex $x \in \mathcal{U}$. Let $Y_x := N_{\mathcal{Q}^n}(x) \setminus (V(SA(x)) \cup V^{\mathrm{abs}})$ (that is, by (C′3), Y_x is the set of all vertices in $N_{\mathcal{Q}^n}(x) \setminus SA^v = N_{\mathcal{Q}^n}(x) \setminus V(SA(x))$ which lie in clones of cubes which are represented in τ_0' by atomic vertices). By (2.52), we have that $|Y_x| \geq (1 - 2/\ell^4)n - |V(SA(x))| \geq (1 - 1/\ell^3)n$. Claim 2.8(ii) implies that $Y_x \cap (\mathfrak{L} \cup \mathfrak{R}_1 \cup \mathfrak{R}_2) = \varnothing$, so by definition we have that $Y_x \cap \bigcup_{u \in V^{\mathrm{abs}}} V(\mathcal{P}_{\mathrm{abs}}(u)) = \varnothing$. It then follows by (HC5) that, for each $y \in Y_x$, each of the two edges of $\mathfrak{H}'$ incident to y lies in Q'. But Q' is $(\mathcal{U}, 2\ell^2, s)$-good by Claim 2.9. Now, even if all the edges incident to the remaining vertices $y \in N_{\mathcal{Q}^n}(x) \setminus Y_x$ used the same pair of directions, it follows that the edges of $\mathfrak{H}'$ incident to the vertices in $N_{\mathcal{Q}^n}(x)$ use each direction of $\mathcal{Q}^n$ which is not an s-direction at most $n/\ell^3 + n/(2\ell^2) \leq n/\ell^2$ times. □

2.8.2. Proofs of theorems 2.1.5 and 2.1.8

We now deduce theorems 2.1.5 and 2.1.8 from theorem 2.8.1.

Proof of theorem 2.1.8. Let $0 < 1/n \ll 1/\ell \ll \varepsilon_1 \ll \varepsilon \ll \varepsilon_2 \ll \gamma \ll 1/k \leq 1$. Let $s := 10\ell$. Let $H^* \sim \mathcal{Q}^n_{1/2-2\varepsilon}$ and $Q \sim \mathcal{Q}^n_{\varepsilon}$. Observe that $H^* \cup Q \sim \mathcal{Q}^n_{1/2-\varepsilon'}$ for some $\varepsilon' \geq \varepsilon$, so it suffices to prove that $H \cup H^* \cup Q$ contains the desired Hamilton cycles and perfect matchings.

By lemma 2.7.7 with 2ε playing the role of ε, we have that a.a.s. H^* is $(s, \ell, \varepsilon_1, \varepsilon_2, \gamma, \mathcal{U}(H^*, \varepsilon_1))$-robust. Condition on this event and let $\mathcal{U} := \mathcal{U}(H^*, \varepsilon_1)$. By an application of lemma 2.7.7(ii), it follows that there exists a decomposition of H^* into $r := \lceil k/2 \rceil$ edge-disjoint spanning subgraphs $H_1^*, \dots, H_r^*$ such that, for every $i \in [r]$, we have that H_i^* is $(s, \ell, \varepsilon_1/(2r), \varepsilon_2, \gamma/r^{10}, \mathcal{U})$-robust.

Consider a random decomposition of Q into r edge-disjoint spanning subgraphs $Q_1, \dots, Q_r$ in such a way that, if $e \in Q$, then e is assigned to one of the Q_i chosen uniformly at random and independently of all other edges. It follows that, for all $i \in [r]$, we have $Q_i \sim \mathcal{Q}^n_{\varepsilon/r}$.

Let Φ be a constant such that theorem 2.8.1 holds with $\varepsilon_1/(2r)$, γ/r^{10}, ε/r and $r+2$ playing the roles of ε_1, γ, η and c, respectively. (In particular, $\Phi \geq r$.) For each $i \in [r]$, apply theorem 2.8.1 with H_i^*, Q_i, $\varepsilon_1/(2r)$, γ/r^{10}, ε/r and $r+2$ playing the roles of H, Q, ε_1, γ, η and c, respectively, to conclude that a.a.s. there is a $(\mathcal{U}, \ell^2, s)$-good subgraph $Q_i' \subseteq Q_i$ with $\Delta(Q_i') \leq \Phi$ such that, for every $H' \subseteq \mathcal{Q}^n$ such that $d_{H'}(x) \geq 2$ for every $x \in \mathcal{U}$, and every $F \subseteq \mathcal{Q}^n$ with $\Delta(F) \leq (r+2)\Phi$ which is $(\mathcal{U}, \ell, s)$-good, we have that $((H_i^* \cup Q_i) \setminus F) \cup H' \cup Q_i'$ contains a $(\mathcal{U}, \ell^2, s)$-good Hamilton cycle C such that, for all $x \in \mathcal{U}$, both edges of C incident to x belong to H'. Condition on the event that this holds for all $i \in [r]$ (which holds a.a.s. by a union bound).

Now consider the graph H from the statement of theorem 2.1.8. By (R3) in definition 2.7.6, we can greedily find r edge-disjoint subgraphs $H_1, \dots, H_r \subseteq H$ such that

(i) for each $i \in [\lfloor k/2 \rfloor]$, we have that $|E(H_i)| = 2|\mathcal{U}|$ and $d_{H_i}(x) = 2$ for every $x \in \mathcal{U}$, and

(ii) if $2r = k+1$, then H_r is a matching of size $|\mathcal{U}|$ such that $d_{H_r}(x) = 1$ for all $x \in \mathcal{U}$.

Suppose first that $2r = k$. We are going to find r edge-disjoint $(\mathcal{U}, \ell^2, s)$-good Hamilton cycles $C_1, \dots, C_r$ with $H_i \subseteq C_i$ iteratively. Suppose that for some $i \in [r]$ we have already found $C_1, \dots, C_{i-1}$. Let $F_i := \bigcup_{j=1}^{r} Q_j' \cup \bigcup_{j=1}^{i-1} C_j$. It follows by construction that F_i is $(\mathcal{U}, \ell, s)$-good and $\Delta(F_i) \leq r(\Phi+2) \leq (r+2)\Phi$. Then, by the conditioning above, the graph $((H_i^* \cup Q_i) \setminus F_i) \cup H_i \cup Q_i'$ must contain a $(\mathcal{U}, \ell^2, s)$-good Hamilton cycle C_i such that, for each $u \in \mathcal{U}$, both edges of C_i incident to x belong to H_i. In particular, $H_i \subseteq C_i$. Take one such cycle and proceed.

In order to see that these r cycles are pairwise edge-disjoint, suppose that there exist $i, j \in [r]$ with $i < j$ such that $E(C_i) \cap E(C_j) \neq \varnothing$, and let $e \in E(C_i) \cap E(C_j)$. Observe that

$e \notin E(H_i) \cup E(H_j)$ because, otherwise, we would have e incident to some vertex $x \in \mathcal{U}$, and we know that both edges incident to x in C_i and C_j belong to H_i and H_j, respectively, which are edge-disjoint. Therefore, since $e \in E(C_i)$ and $Q_j' \subseteq F_i \setminus Q_i'$, we must have that $e \notin E(Q_j')$. However, since $e \in E(C_j)$ and $e \in E(F_j)$ by definition, we must have $e \in E(Q_j')$, a contradiction.

Suppose now that $2r = k+1$. Let $F_1 := \bigcup_{j=1}^{r} Q_j'$, so $\Delta(F_1) \leq r\Phi$ and it is $(\mathcal{U}, \ell, s)$-good. By the conditioning above, $((H_1^* \cup Q_1) \setminus F_1) \cup H_1 \cup Q_1'$ contains a $(\mathcal{U}, \ell^2, s)$-good Hamilton cycle C with $H_1 \subseteq C$. We split C into two perfect matchings $M_1 \cup M_2$ (observe that both of them are $(\mathcal{U}, \ell^2, s)$-good) and redefine $H_r := H_r \cup \{e \in M_2 : \mathcal{U} \cap e \neq \varnothing\}$, so that H_r now satisfies (i). Now, for each $i \in \{2, \dots, r\}$, we proceed as follows. Let $F_i := M_1 \cup \bigcup_{j=1}^{r} Q_j' \cup \bigcup_{j=2}^{i-1} C_j$. It follows by construction that F_i is $(\mathcal{U}, \ell, s)$-good and $\Delta(F_i) \leq r(\Phi + 2) \leq (r+2)\Phi$. Then, by the conditioning above, the graph $((H_i^* \cup Q_i) \setminus F_i) \cup H_i \cup Q_i'$ must contain a $(\mathcal{U}, \ell^2, s)$-good Hamilton cycle C_i with $H_i \subseteq C_i$. Take one such cycle and proceed. The fact that the graphs $M_1, C_2, \dots, C_r$ are pairwise edge-disjoint can be proved as in the previous case. □

We now prove theorem 2.1.5. Recall from section 2.1.4 that, for any $k \in \mathbb{N}$ and any graph $G \subseteq \mathcal{Q}^n$, we say that $G \in \delta k$ if $\delta(G) \geq k$, and $G \in \mathcal{HM}k$ if it contains $\lfloor k/2 \rfloor$ edge-disjoint Hamilton cycles and $k - 2\lfloor k/2 \rfloor$ perfect matchings which are edge-disjoint from these cycles. We say that $G \in \mathcal{P}k$ if, for every spanning subgraph $H \subseteq \mathcal{Q}^n$ with $H \in \delta k$, we have $G \cup H \in \mathcal{HM}k$.

Proof of theorem 2.1.5. The case $k = 1$ of the statement was proved by Bollobás [19], so we may assume $k \geq 2$. Let $0 < \varepsilon \ll 1/k$ and $G \sim \mathcal{Q}^n_{1/2-\varepsilon}$. By theorem 2.1.8, we have $\mathbb{P}[G \in \mathcal{P}k] = 1 - o(1)$. Also note that, by lemma 2.4.1, we have that $\mathbb{P}[e(G) \geq (1/2 - \varepsilon/2)n2^{n-1}] = o(1)$. Hence,

$$\mathbb{P}[\{G \in \mathcal{P}k\} \wedge \{e(G) < (1/2 - \varepsilon/2)n2^{n-1}\}] = 1 - o(1).$$

Thus, by a simple conditioning argument, there exists a positive integer $m < (1/2-\varepsilon/2)n2^{n-1}$ such that

$$\mathbb{P}[G \in \mathcal{P}k \mid e(G) = m] = 1 - o(1). \tag{2.55}$$

Let $G_m \subseteq \mathcal{Q}^n$ be a uniformly random subgraph of $\mathcal{Q}^n$ with exactly m edges. Since $\mathbb{P}[G \in \mathcal{P}k \mid e(G) = m] = \mathbb{P}[G_m \in \mathcal{P}k]$, by (2.55) we have $\mathbb{P}[G_m \in \mathcal{P}k] = 1 - o(1)$. Now, because a.a.s. $\tau_{\mathcal{P}k}(\tilde{\mathcal{Q}}^n(\sigma)) \geq (1/2 - \varepsilon/4)n2^{n-1}$, the result follows. □

CHAPTER 3

EDGE CORRELATIONS IN RANDOM REGULAR HYPERGRAPHS AND APPLICATIONS TO SUBGRAPH TESTING

Compared to the classical binomial random (hyper)graph model, the study of random regular hypergraphs is made more challenging due to correlations between the occurrence of different edges. We develop an edge-switching technique for hypergraphs which allows us to show that these correlations are limited for a large range of densities. This extends some previous results of Kim, Sudakov, and Vu for graphs. From our results we deduce several corollaries on subgraph counts in random d-regular hypergraphs. We also prove a conjecture of Dudek, Frieze, Ruciński, and Šileikis on the threshold for the existence of an ℓ-overlapping Hamilton cycle in a random d-regular r-graph.
Moreover, we apply our results to prove bounds on the query complexity of testing subgraph-freeness. The problem of testing subgraph-freeness in the general graphs model was first studied by Alon, Kaufman, Krivelevich, and Ron, who obtained several bounds on the query complexity of testing triangle-freeness. We extend some of these previous results beyond the triangle setting and to the hypergraph setting.

3.1. Introduction

3.1.1. Random regular graphs

While the consideration of random d-regular graphs is very natural and has a long history, this model is much more difficult to analyze than the seemingly similar $\mathcal{G}(n,p)$ and $\mathcal{G}(n,m)$ models due to the dependencies between edges (here $\mathcal{G}(n,p)$ refers to the binomial n-vertex random graph model with edge probability p and $\mathcal{G}(n,m)$ refers to the uniform distribution on all n-vertex graphs with m edges). For small d, the configuration model (due to Bollobás [16]) has led to numerous results on random d-regular graphs. Moreover, the switching method introduced by McKay and Wormald [96] has led to results for a much larger range of d than can be handled by the configuration model. For example, Kim, Sudakov, and Vu [77] used such ideas to show that the classical results on distributions of small subgraphs in $\mathcal{G}(n,p)$ carry over to random regular graphs.

In this chapter we develop an edge-switching technique for random regular r-uniform hypergraphs (also called r-graphs). More precisely, we show that correlations between the existence of edges in a random regular r-graph are small even if we condition on the (non-)existence of some further edges (see corollary 3.2.3). This allows us to generalise results of Kim, Sudakov, and Vu [77] on the appearance of fixed subgraphs in a random regular graph to the hypergraph setting (see corollary 3.3.3). Moreover, even in the graph case, we can condition on the (non-)existence of a significantly larger edge set than in [77].

A general result of Dudek, Frieze, Ruciński, and Šileikis [43] implies that one can transfer many statements from the binomial model to the random regular hypergraph model (see theorem 3.3.5). This allows them to deduce (from the main result of Dudek and Frieze [40]) the following: if $2 \leq \ell < r$ and $n^{\ell-1} \ll d \ll n^{r-1}$, then a random d-regular r-graph asymptotically almost surely (a.a.s.) contains an ℓ-overlapping Hamilton cycle, that is, a Hamilton cycle in which consecutive edges overlap in precisely ℓ vertices (these cycles are defined formally in section 3.1.4). They conjectured that the lower bound provides the correct threshold in the following sense:

(3.1) *if* $2 \leq \ell < r$ *and* $d \ll n^{\ell-1}$*, then a.a.s. a random* d*-regular* r*-graph contains no* ℓ*-overlapping Hamilton cycle.*

Our correlation results from section 3.2 allow us to confirm this conjecture (see corollary 3.3.13). The threshold for a loose Hamilton cycle (i.e. a 1-overlapping Hamilton cycle) in a random d-regular r-graph was recently determined (via the configuration model) by Altman, Greenhill, Isaev, and Ramadurai [7]. This improved earlier bounds by Dudek, Frieze, Ruciński, and Šileikis [42]. Altman, Greenhill, Isaev, and Ramadurai [7] also investigated the above conjecture and proved that (3.1) holds under the much stronger condition that $d \ll n$ if $r \geq 4$ and $d \ll n^{1/2}$ if $r = 3$ (we do rely on their result when d is constant to establish (3.1)). The graph case $r = 2$ where d is fixed is a classical result by Robinson and Wormald [103, 104]: if $d \geq 3$ is fixed, then a.a.s. a random d-regular graph has a Hamilton cycle. This was extended to larger d by Cooper, Frieze, and Reed [37].

In a similar way, we can transfer several classical counting results for random graphs to the regular setting. We illustrate this for Hamilton cycles, where we extend the density range of a counting result of Krivelevich [85]: for $\log n \ll d \ll n$, a.a.s. the number of Hamilton cycles in a random d-regular n-vertex graph is fairly close to $n!(d/n)^n$ (see corollary 3.3.8). The results by Krivelevich [85] imply the same behaviour for $d \gg e^{(\log n)^{1/2}}$. For constant d, this problem was studied by Janson [70]. Similarly, we transfer a general counting result for spanning subgraphs in $\mathcal{G}(n, m)$ due to Riordan [102] to the setting of random regular graphs.

3.1.2. Property testing

The running time of any "exact" algorithm that checks whether a given combinatorial object has a given property must be at least linear in the size of the input. Property testing algorithms have the potential to give much quicker answers, although at the cost of not knowing for certain if the desired property is satisfied by the object. A property testing algorithm is usually given oracle access to the combinatorial object and answers whether the object satisfies the property or is "far" from satisfying it.

To be precise, following, e.g., Goldreich, Goldwasser, and Ron [59], we define testers as follows. Given a property $\mathcal{P}$, a *tester* for $\mathcal{P}$ is a (possibly randomized) algorithm that is given a distance parameter ε and oracle access to a structure S. If $S \in \mathcal{P}$, then the algorithm must accept with probability at least 2/3. If S is ε-far from $\mathcal{P}$, then the algorithm should reject with probability at least 2/3. If the algorithm is allowed to make an error in both cases, we say it is a *two-sided error tester*; if, on the contrary, the algorithm always gives the correct answer when S has the property, we say it is a *one-sided error tester*.

For graphs (and, more generally, r-graphs) there have been two classical models for testers: one of them is the dense model, and the other is the bounded-degree model. In the dense model, the density of the r-graph is assumed to be bounded away from 0, and we say that an r-graph G is ε-far from having property $\mathcal{P}$ if at least εn^r edges have to be modified (added or deleted) to turn G into a graph that satisfies $\mathcal{P}$. Many results have been proved for the dense model. In particular, there exists a characterization of all properties which are testable with constant query complexity (by Alon, Fischer, Newman, and Shapira [2] in the graph case and Joos, Kim, Kühn, and Osthus [73] in the r-graph case). For the bounded-degree graphs model (which assumes that the maximum degree of the input graphs is bounded by a fixed constant), several general results have also been obtained (see, for example, the results of Benjamini, Schramm, and Shapira [12] as well as Newman and Sohler [99]).

Here, we consider the general graphs model and its generalization to r-graphs. In the general graphs model (introduced by Kaufman, Krivelevich, and Ron [75]), a graph G with m edges is ε-far from having property $\mathcal{P}$ if at least εm edges have to be modified for the graph to satisfy $\mathcal{P}$. Furthermore, we also assume that the edges are labelled in the sense that for each vertex there is an ordering of its incident edges. It is natural to consider the following two types of queries. First, we allow vertex-pair queries, where any algorithm may take two vertices and ask whether they are joined by an edge in the graph or not. Second, we allow neighbour queries, where any algorithm may take a vertex and ask which vertex is its i-th neighbour.

These notions generalise to hypergraphs in a straightforward way. More precisely, we will consider the following general hypergraphs model, where a hypergraph with m edges is ε-far from having property $\mathcal{P}$ if at least εm edges must be added or deleted to ensure the resulting hypergraph satisfies $\mathcal{P}$. As in the graph case, we will consider two types of queries:

- Vertex-set queries: Any algorithm may take a set of r vertices and ask whether they constitute an edge in the r-graph. The answer must be either yes or no.

- Neighbour queries: Any algorithm may take a vertex and ask for its i-th incident edge (according to the labelling of the edges). The answer is either a set of $r-1$ vertices or an error message if the degree of the queried vertex is smaller than i.

In this chapter we consider the property $\mathcal{P}$ of being F-free for fixed r-graphs F. In the dense setting, the theory of hypergraph regularity (as developed by Rödl and Skokan [108], Rödl and Schacht [105, 106, 107], as well as Gowers [63]) implies the existence of testers with constant query complexity for this problem.

However, the problem is still wide open for general graphs and hypergraphs. Alon, Kaufman, Krivelevich, and Ron [3] studied the problem of testing triangle-freeness. In section 3.4, we provide lower and upper bounds for testing F-freeness which apply to large classes of hypergraphs F. In particular, we observe that testing F-freeness cannot be achieved in a constant number of queries whenever F is not a weak forest and the density of the graphs G to be tested is somewhat below the Turán threshold for F (see proposition 3.4.1). Based on the results of sections 3.2 and 3.3.1, we also provide a lower bound (see theorem 3.4.5) which improves on proposition 3.4.1 for a large range of parameters and r-graphs. Roughly speaking, theorem 3.4.5 provides better bounds than proposition 3.4.1 if the average degree d of the input r-graph G is not too small. On the other hand, the class of admissible F is more restricted. We also provide three upper bounds on the query complexity (see section 3.4.3).

Kaufman, Krivelevich, and Ron [75] also studied the problem of testing bipartiteness in general graphs. It would be interesting to obtain results for the general (hyper)graphs model covering further properties and to improve the lower and upper bounds we present for testing F-freeness.

3.1.3. Outline of the chapter

The remainder of the chapter is organised as follows. In section 3.2 we develop a hypergraph generalisation of the edge-switching technique to prove a correlation result (corollary 3.2.3) for the event that a given edge is present in a random d-regular r-graph even if we condition on the (non-)existence of some further edges.

Section 3.3 builds on this to obtain subgraph count results in random d-regular r-graphs. In particular, in section 3.3.1 we consider the counting problem for small fixed graphs F, for which we prove a concentration result, thus also obtaining the threshold for their appearance, which generalises a result of Kim, Sudakov, and Vu [77] for graphs. We also derive bounds on the number of edge-disjoint copies of fixed subgraphs F in a random d-regular r-graph, which we use in section 3.4.2. In section 3.3.2, we combine the results from section 3.2 with known results for $\mathcal{G}^{(r)}(n, p)$ and $\mathcal{G}^{(r)}(n, m)$ to count the number of suitable spanning subgraphs (such as Hamilton cycles) in random d-regular r-graphs.

Finally, section 3.4 provides lower and upper bounds on the query complexity for testing subgraph freeness for small, fixed r-graphs F. The proof of the main lower bound relies on corollary 3.2.3 and the counting results derived in section 3.3.1.

3.1.4. Definitions and notation

Given any $n \in \mathbb{N}$, we will write $[n] := \{1, \dots, n\}$. Throughout the chapter, we will use the standard $\mathcal{O}$ notation to compare asymptotic behaviours of functions. Whenever this is used, we implicitly assume that the functions are non-negative. Given $a, b, c \in \mathbb{R}$, we will write $c = a \pm b$ if $c \in [a - b, a + b]$.

An *r-graph* (or r-uniform hypergraph) $H = (V, E)$ is an ordered pair where V is a set of vertices, and $E \subseteq \binom{V}{r}$ is a set of r-subsets of V, called edges. We always assume that r is a fixed integer greater than 1. When $r = 2$, we will simply refer to these as graphs and omit the presence of r in any notation. To indicate the vertex set and the edge set of a certain r-graph H we will use the notation $V(H)$ and $E(H)$, respectively. We will often abuse notation and write $e \in H$ to mean $e \in E(H)$, or use $E(H)$ instead of H to denote the r-graph. In particular, we write $|H|$ for $|E(H)|$. The *order* of an r-graph H is $|V(H)|$ and the *size* of H is $|E(H)|$. For a fixed r-graph H, we sometimes denote its number of vertices by v_H, while e_H will denote the number of edges.

Given a vertex $v \in V(H)$, the *degree* of v in H is $\deg_H(v) := |\{e \in H : v \in e\}|$. When H is clear from the context, it may be dropped from the notation. We will use $\Delta(H)$ to denote the maximum (vertex) degree of H, $\delta(H)$ to denote the minimum (vertex) degree of H and $d(H)$ to denote its average (vertex) degree. We say that H is *d-regular* if $\deg_H(v) = d$ for all $v \in V(H)$. The set of vertices lying in a common edge with v is called its *neighbourhood* and denoted by $N_H(v)$.

The *complete* r-graph of order n is denoted by $K_n^{(r)}$. If its vertex set V is given, we denote this by $K_V^{(r)}$. We say that an r-graph H is *k-partite* if there exists a partition of $V(H)$ into k sets such that every edge $e \in H$ contains at most one vertex in each of the sets. A *path* P between vertices u and v, also called a *(u, v)-path*, is an r-graph whose vertices admit a labelling $u, v_1, \dots, v_k, v$ such that any two consecutive vertices lie in an edge of P and each edge consists of consecutive vertices. An r-graph H is *connected* if there exists a path joining any two vertices in $V(H)$. The *distance* between vertices u and v in H is defined by $\mathrm{dist}_H(u, u) := 0$ and $\mathrm{dist}_H(u, v) := \min\{|P| : P \text{ is an } (u, v)\text{-path}\}$ whenever $u \neq v$. If there is no such path, the distance is said to be infinite. The distance between sets of vertices S and T is $\mathrm{dist}_H(S, T) := \min\{\mathrm{dist}_H(s, t) : s \in S, t \in T\}$. The *diameter* of an r-graph H is $D(H) := \max_{(u,v)\in V(H)^2} \mathrm{dist}_H(u, v)$. An r-graph C is a *k-overlapping cycle* of length ℓ if $|C| = \ell$ and the vertices of C admit a cyclic labelling such that each edge in C consists of r consecutive vertices and any two consecutive edges have exactly k vertices in common (in the natural cyclic order induced on the edges of C). When $k = 1$, we refer to C as a *loose cycle*. When $k = r-1$, C is called a *tight cycle*. A k-overlapping cycle C is said to be *Hamiltonian* for an r-graph H if $E(C) \subseteq E(H)$ and $V(C) = V(H)$. We will write C_n^k for a k-overlapping cycle of order n. We say that a connected r-graph H is a *weak tree* if $|e \cap f| \leq 1$ for all $e, f \in H$ with $e \neq f$, and H contains no loose cycles. We say that an r-graph is a *weak forest* if it is the union of vertex-disjoint weak trees. Note that, for graphs, this is the usual definition of a forest. Given any r-graph H, its complement is denoted as $\overline{H}$.

The Erdős-Rényi random r-graph, also called the binomial model, is denoted by $\mathcal{G}^{(r)}(n,p)$, for $n \in \mathbb{N}$ and $p \in [0,1]$. An r-graph $G^{(r)}(n,p)$ on vertex set V with $|V| = n$ chosen according to this model is obtained by including each $e \in \binom{V}{r}$ with probability p independently from the other edges. For $n \in \mathbb{N}$ and $m \in [\binom{n}{r}] \cup \{0\}$, we denote by $\mathcal{G}^{(r)}(n,m)$ the set of all r-graphs on n vertices that have exactly m edges, and denote by $G^{(r)}(n,m)$ an r-graph chosen uniformly at random from this set. We denote the set of all d-regular r-graphs on vertex set V with $|V| = n$ by $\mathcal{G}_{n,d}^{(r)}$, for $n \in \mathbb{N}$ and $d \in [\binom{n-1}{r-1}] \cup \{0\}$, and denote by $G_{n,d}^{(r)}$ an r-graph chosen uniformly at random from $\mathcal{G}_{n,d}^{(r)}$. If H and H' are two r-graphs on vertex set V, we define $\mathcal{G}_{n,d,H,H'}^{(r)}$ as the set of all r-graphs $G \in \mathcal{G}_{n,d}^{(r)}$ such that $H \subseteq G$ and $H' \subseteq \overline{G}$. With a slight abuse of notation, we sometimes also treat $\mathcal{G}_{n,d,H,H'}^{(r)}$ as the event that $G_{n,d}^{(r)} \in \mathcal{G}_{n,d,H,H'}^{(r)}$. Given a sequence of events $\{\mathcal{A}_n\}_{n\geq 1}$, we will say that $\mathcal{A}_n$ holds *asymptotically almost surely*, and write a.a.s., if $\lim_{n\to\infty} \mathbb{P}[\mathcal{A}_n] = 1$.

Throughout the chapter, we will often use the following observation.

Remark 3.1.1. *Let $r \geq 2$ be an integer, and let $d = o(n^{r-1})$ be such that $r \mid nd$. Then, there exist d-regular r-graphs on n vertices.*

Indeed, since $r \mid nd$, we can write $r = r_1 r_2$ such that $r_1 \mid n$ and $r_2 \mid d$. Then an $(r-r_1)$-overlapping cycle is r_2-regular, and thus an edge-disjoint union of d/r_2 $(r-r_1)$-overlapping cycles on the same vertex set is d-regular. Since $d = o(n^{r-1})$, such a set of d/r_2 edge-disjoint cycles can be found iteratively (see, e.g., [58, Theorem 2]).

The condition that $r \mid nd$ is necessary, and throughout the chapter we will always implicitly assume it to hold.

3.2. Edge correlation in random regular r-graphs

This section is devoted to estimating the probability that any fixed r-set of vertices forms an edge in a random d-regular r-graph, even if we require certain edges to (not) be present. More precisely, we obtain accurate bounds on $\mathbb{P}[e \in G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}]$ for a large range of d as long as H, H' are sparse (see corollary 3.2.3). This result is the core ingredient for all the results in section 3.3 and it will be used in the proof of our lower bound on the query complexity for testing F-freeness, for a fixed r-graph F, in section 3.4.2.

Corollary 3.2.3 follows immediately from lemma 3.2.1 (which provides the upper bound) and lemma 3.2.2 (which provides the lower bound). To prove lemmas 3.2.1 and 3.2.2 we develop a hypergraph generalization of the method of edge-switchings, which was introduced for graphs by McKay and Wormald [96]. The switchings we consider in the proof of lemma 3.2.1 are similar to those used by Dudek, Frieze, Ruciński, and Šileikis [43]. The switchings we use in lemma 3.2.2 are more complex, however. Moreover, to bound the number of certain 'bad' configurations, the proof of lemma 3.2.2 relies on lemma 3.2.1. The special case of lemmas 3.2.1 and 3.2.2 when $r = 2$ and H, H' have bounded size (which is much simpler to prove) was obtained by Kim, Sudakov, and Vu [77].

Lemma 3.2.1. *Let $r \geq 2$ be a fixed integer. Assume that $d = o(n^{r-1})$. Suppose $H, H' \subseteq \binom{V}{r}$ are two edge-disjoint r-graphs such that $|H| = o(nd)$ and $\Delta(H') = o(n^{r-1})$. Then, for all $e \in \binom{V}{r} \setminus (H \cup H')$, we have*

$$\mathbb{P}\left[e \in G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}\right] \leq (r-1)!\frac{d}{n^{r-1}}\left(1 + \mathcal{O}\left(\frac{1}{n} + \frac{d}{n^{r-1}} + \frac{|H|}{nd} + \frac{\Delta(H')}{n^{r-1}}\right)\right).$$

Proof. Write $e = \{v_1, \dots, v_r\}$ and fix this labelling of the vertices in e. Let $e_1 := e$ and let $e_2, \dots, e_r \in \binom{V}{r}$ be pairwise disjoint and also disjoint from e_1. Let $f_1, \dots, f_r \in \binom{V}{r}$ be pairwise disjoint and such that $f_i \cap e_1 = \{v_i\}$ for all $i \in [r]$. We say that $\Lambda_e := (e_1, \dots, e_r)$ is an *out-switching configuration* and that $\Lambda_{\overline{e}} := (f_1, \dots, f_r)$ is an *in-switching configuration*. If, furthermore, $|e_i \cap f_j| = 1$ for all $i, j \in [r]$, we say that Λ_e and $\Lambda_{\overline{e}}$ are *related*.

Given $\Lambda_e = (e_1, \dots, e_r)$, we denote the number of in-switching configurations related to Λ_e by $\lambda_{\text{in}} = \lambda_{\text{in}}(\Lambda_e)$; we claim that

$$\lambda_{\text{in}} = (r!)^{r-1}. \tag{3.2}$$

Indeed, for each $i \in [r] \setminus \{1\}$, write $e_i = \{v_1^i, \dots, v_r^i\}$ and let $\pi_i \colon [r] \to [r]$ be a permutation. For each $i \in [r]$, let $f_i := \{v_i, v^2_{\pi_2(i)}, \dots, v^r_{\pi_r(i)}\}$. Then, $\Lambda_{\overline{e}} := (f_1, \dots, f_r)$ is related to Λ_e. In this way, each (ordered) $(r-1)$-tuple of permutations $(\pi_2, \dots, \pi_r)$ defines a unique in-switching configuration. On the other hand, each $\Lambda_{\overline{e}} = (f_1, \dots, f_r)$ related to Λ_e gives rise to a different $(r-1)$-tuple of permutations $(\pi_2, \dots, \pi_r)$ by setting, for each $i \in [r] \setminus \{1\}$ and $j \in [r]$, $\pi_i(j)$ to be the subscript of the vertex in $e_i \cap f_j$. There are $(r!)^{r-1}$ such tuples of permutations, so (3.2) follows.

Similarly, given $\Lambda_{\overline{e}} = (f_1, \dots, f_r)$, we denote the number of out-switching configurations related to $\Lambda_{\overline{e}}$ by $\lambda_{\text{out}} = \lambda_{\text{out}}(\Lambda_{\overline{e}})$. We claim that

$$\lambda_{\text{out}} = ((r-1)!)^r. \tag{3.3}$$

Indeed, for each $i \in [r]$, write $f_i = \{v_i, v_2^i, \dots, v_r^i\}$ and let $\sigma_i \colon [r] \setminus \{1\} \to [r] \setminus \{1\}$ be a permutation. For each $i \in [r] \setminus \{1\}$, let $e_i := \{v^1_{\sigma_1(i)}, \dots, v^r_{\sigma_r(i)}\}$. Then, $\Lambda_e := (e_1, \dots, e_r)$ is related to $\Lambda_{\overline{e}}$. Each r-tuple of permutations $(\sigma_1, \dots, \sigma_r)$ defines a unique Λ_e. On the other hand, each $\Lambda_e = (e_1, \dots, e_r)$ related to $\Lambda_{\overline{e}}$ gives rise to a unique r-tuple of permutations $(\sigma_1, \dots, \sigma_r)$. Thus (3.3) holds.

Let $\Omega_1, \Omega_2 \subseteq \binom{V}{r}$. We define a function ψ on the set of all r-graphs G on V by $\psi(G, \Omega_1, \Omega_2) := (G \setminus \Omega_1) \cup \Omega_2$. Now let G be an r-graph on V. Let Λ_e and $\Lambda_{\overline{e}}$ be related out- and in-switching configurations, respectively, such that $\Lambda_e \subseteq G$ and $\Lambda_{\overline{e}} \subseteq \overline{G}$. An *out-switching* on G from Λ_e to $\Lambda_{\overline{e}}$ is obtained by applying the operation $\psi(G, \Lambda_e, \Lambda_{\overline{e}})$ (here Λ_e and $\Lambda_{\overline{e}}$ are viewed as (unordered) sets of edges). We denote this out-switching by the triple $(G, \Lambda_e, \Lambda_{\overline{e}})$. Similarly, if Λ_e and $\Lambda_{\overline{e}}$ are related out- and in-switching configurations, respectively, such that $\Lambda_e \subseteq \overline{G}$ and $\Lambda_{\overline{e}} \subseteq G$, an *in-switching* on G from $\Lambda_{\overline{e}}$ to Λ_e is the operation $\psi(G, \Lambda_{\overline{e}}, \Lambda_e)$

and is denoted by $(G, \Lambda_{\overline{e}}, \Lambda_e)$. Note that $\psi(\psi(G, \Lambda_{\overline{e}}, \Lambda_e), \Lambda_e, \Lambda_{\overline{e}}) = G$, that is, switchings are involutions. Furthermore, both types of switchings preserve the vertex degrees of the r-graph G on which they act.

Let $\mathcal{F}_e \subseteq \mathcal{G}^{(r)}_{n,d,H,H'}$ be the set of all r-graphs $G \in \mathcal{G}^{(r)}_{n,d,H,H'}$ such that $e \in G$, and let $\mathcal{F}_{\overline{e}} := \mathcal{G}^{(r)}_{n,d,H,H'} \setminus \mathcal{F}_e$. We define an auxiliary bipartite multigraph Γ with bipartition $(\mathcal{F}_e, \mathcal{F}_{\overline{e}})$ as follows. For each $G \in \mathcal{F}_e$, consider all possible out-switchings on G whose image is in $\mathcal{G}^{(r)}_{n,d,H,H'}$ (that is, all triples $(G, \Lambda_e, \Lambda_{\overline{e}})$ such that $\Lambda_e \subseteq G \setminus H$ and $\Lambda_{\overline{e}} \subseteq \overline{G} \setminus H'$ are related) and add an edge between G and $\psi(G, \Lambda_e, \Lambda_{\overline{e}})$ for each such triple $(G, \Lambda_e, \Lambda_{\overline{e}})$. Similarly, one could consider each $G \in \mathcal{F}_{\overline{e}}$ and every possible in-switching $(G, \Lambda_{\overline{e}}, \Lambda_e)$ on G with $\psi(G, \Lambda_{\overline{e}}, \Lambda_e) \in \mathcal{G}^{(r)}_{n,d,H,H'}$, and add an edge between G and $\psi(G, \Lambda_{\overline{e}}, \Lambda_e)$. Both constructions result in the same multigraph Γ.

We will use switchings to bound $\mathbb{P}[e \in G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}] = |\mathcal{F}_e|/|\mathcal{G}^{(r)}_{n,d,H,H'}|$ from above in terms of $\mathbb{P}[e \notin G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}]$. In order to obtain this bound, we will use a double-counting argument involving the edges of Γ.

Assume first that $G \in \mathcal{F}_{\overline{e}}$. Let $S_{\mathrm{in}}(G)$ be the number of in-switchings $(G, \Lambda_{\overline{e}}, \Lambda_e)$ on G, thus $\deg_\Gamma(G) \leq S_{\mathrm{in}}(G)$. We claim that

$$S_{\mathrm{in}}(G) \leq ((r-1)!)^r d^r. \tag{3.4}$$

Clearly, $S_{\mathrm{in}}(G)$ is at most the number of in-switching configurations $\Lambda_{\overline{e}} \subseteq G$ multiplied by λ_{out}. As G is d-regular and $\Lambda_{\overline{e}}$ must contain an edge incident to each $v_i \in e$, there are at most d^r such in-switching configurations. This, together with (3.3), yields (3.4).

Assume now that $G \in \mathcal{F}_e$. Let $\ell := |H|$ and $k' := \Delta(H')$, and let $\eta := \eta(n, d, \ell, k') = \frac{1}{n} + \frac{d}{n^{r-1}} + \frac{\ell}{nd} + \frac{k'}{n^{r-1}}$. Let $S_{\mathrm{out}}(G)$ be the number of possible out-switchings $(G, \Lambda_e, \Lambda_{\overline{e}})$ on G with $\psi(G, \Lambda_e, \Lambda_{\overline{e}}) \in \mathcal{G}^{(r)}_{n,d,H,H'}$; thus, $\deg_\Gamma(G) = S_{\mathrm{out}}(G)$. We claim that

$$S_{\mathrm{out}}(G) \geq ((r-1)!)^{r-1} (nd)^{r-1} \left(1 - \mathcal{O}(\eta)\right). \tag{3.5}$$

In order to have $\psi(G, \Lambda_e, \Lambda_{\overline{e}}) \in \mathcal{G}^{(r)}_{n,d,H,H'}$ we must have $\Lambda_e \subseteq G \setminus H$ and $\Lambda_{\overline{e}} \subseteq \overline{G} \setminus H'$. Let $\lambda_e(G)$ be the number of out-switching configurations Λ_e with $\Lambda_e \subseteq G \setminus H$. We first give a lower bound on $\lambda_e(G)$.

Choose $\Lambda_e = (e_1, \ldots, e_r)$ by sequentially choosing $e_2, \ldots, e_r \in G \setminus H$ in such a way that e_i is disjoint from $e_1, \ldots, e_{i-1}$, for $i \in [r] \setminus \{1\}$. As each vertex is incident to exactly d edges, the number of choices for e_i is at least $(nd/r - \ell - (r-1)rd)$. Thus,

$$\lambda_e(G) \geq \left(\frac{nd}{r} - \ell - (r-1)rd\right)^{r-1}. \tag{3.6}$$

We say that an out-switching configuration $\Lambda_e \subseteq G \setminus H$ is *good* (for G) if there are λ_{in} in-switching configurations $\Lambda_{\overline{e}} \subseteq \overline{G} \setminus H'$ related to Λ_e, and *bad* (for G) otherwise. Let $\lambda_{e,\text{bad}}(G)$ denote the number of bad out-switching configurations $\Lambda_e \subseteq G \setminus H$. We now provide an upper bound on this quantity. An out-switching configuration $\Lambda_e \subseteq G \setminus H$ can be bad only if

(a) one of the edges in some $\Lambda_{\overline{e}}$ related to Λ_e, say, g, lies in G, or

(b) one of the edges in some $\Lambda_{\overline{e}}$ related to Λ_e, say, h, lies in H'.

In case (a), the edge g has to intersect e, so there are at most rd possible such edges g. Furthermore, $g \setminus e$ must intersect every edge in $\Lambda_e \setminus \{e\}$, so each edge g can make at most $(r-1)!d^{r-1}$ out-switching configurations bad. Thus, there are at most $r!d^r$ out-switching configurations which are bad because of (a). In case (b), the edge h has to intersect e, so there are at most rk' such edges. As above, it follows that there are at most $r!k'd^{r-1}$ out-switching configurations which are bad because of (b). Overall,

$$\lambda_{e,\text{bad}}(G) \leq r!d^r + r!k'd^{r-1}. \tag{3.7}$$

By combining (3.2), (3.6), and (3.7), we have that

$$\begin{aligned} S_{\text{out}}(G) &\geq (r!)^{r-1}\left(\left(\frac{nd}{r} - \ell - (r-1)rd\right)^{r-1} - r!d^r - r!k'd^{r-1}\right) \\ &= ((r-1)!)^{r-1}(nd)^{r-1}\left(1 - \mathcal{O}(\eta)\right). \end{aligned}$$

As (3.4) and (3.5) hold for every $G \in \mathcal{F}_{\overline{e}}$ and $G \in \mathcal{F}_e$, respectively, we can use these expressions to estimate the number $|\Gamma|$ of edges in Γ. We conclude that

$$((r-1)!)^{r-1}(nd)^{r-1}(1-\mathcal{O}(\eta))|\mathcal{F}_e| \leq |\Gamma| \leq ((r-1)!)^r d^r |\mathcal{F}_{\overline{e}}|.$$

Noting that $|\mathcal{F}_{\overline{e}}| \leq |\mathcal{G}^{(r)}_{n,d,H,H'}|$ and dividing this by $|\mathcal{G}^{(r)}_{n,d,H,H'}|$ implies that

$$((r-1)!)^{r-1}(nd)^{r-1}(1-\mathcal{O}(\eta)) \cdot \mathbb{P}\left[e \in G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}\right] \leq ((r-1)!)^r d^r.$$

Thus, we conclude that

$$\mathbb{P}\left[e \in G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}\right] \leq (r-1)!\frac{d}{n^{r-1}}(1+\mathcal{O}(\eta)). \qquad \square$$

Lemma 3.2.2. *Let $r \geq 2$ be a fixed integer. Suppose that $d = \omega(1)$ and $d = o(n^{r-1})$. Let $H, H' \subseteq \binom{V}{r}$ be two edge-disjoint r-graphs such that $\Delta(H), \Delta(H') = o(d)$. Then, for all $e \in \binom{V}{r} \setminus (H \cup H')$,*

$$\mathbb{P}\left[e \in G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}\right] \geq (r-1)!\frac{d}{n^{r-1}}\left(1-\mathcal{O}\left(\frac{1}{n}+\frac{1}{d}+\frac{d}{n^{r-1}}+\frac{\Delta(H)}{d}+\frac{\Delta(H')}{d}\right)\right).$$

Proof. Our strategy is similar to that in lemma 3.2.1, but we change the definition of a switching configuration. Write $e = \{v_1, \dots, v_r\}$. Let $e_1, \dots, e_r \in \binom{V}{r}$ be such that, for each $i \in [r]$, $v_i \notin e_i$ and there is a vertex $u_i \in e_i \setminus e$ such that $u_i \notin e_j$ for all $j \in [r] \setminus \{i\}$. Let $f_1, \dots, f_r \in \binom{V}{r} \setminus \{e\}$ be distinct such that $v_i \in f_i$, and let $f \in \binom{V}{r}$ be disjoint from $f_1, \dots, f_r$. We say that $\Lambda_e := (e, e_1, \dots, e_r)$ is an *out-switching configuration* and that $\Lambda_{\overline{e}} := (f_1, \dots, f_r, f)$ is an *in-switching configuration*. We say that Λ_e and $\Lambda_{\overline{e}}$ are *related* if, for each $i \in [r]$, one can find a set $A_i \in \binom{V}{r-1}$ such that $e_i \cap f_i = A_i$, and $f = (e_1 \setminus A_1) \cup \dots \cup (e_r \setminus A_r)$ (note that in this case we must have $A_i = f_i \setminus \{v_i\}$). See figure 3.1 for an illustration. Given related out- and in-switching configurations $\Lambda_e = (e, e_1, \dots, e_r)$ and $\Lambda_{\overline{e}} = (f_1, \dots, f_r, f)$, we will always write $A_i := e_i \cap f_i$ and $\{u_i\} := e_i \setminus f_i$ for $i \in [r]$. It is easy to check that this definition of u_i implies

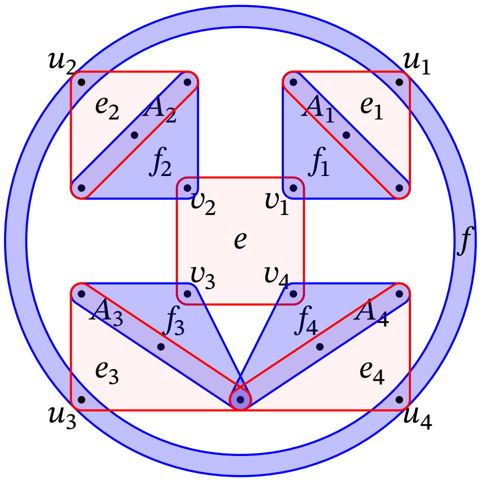

Figure 3.1. Representation of a switching for lemma 3.2.2 in the case $r = 4$. Shaded (blue) edges represent an in-switching configuration, while clear (red) ones represent an out-switching configuration.

that $\{u_i\} = e_i \cap f$ and $u_i \notin e_j$ for all $j \in [r] \setminus \{i\}$. So u_i is indeed as required in the definition of an out-switching configuration.

Given $\Lambda_e = (e, e_1, \dots, e_r)$, we denote the number of in-switching configurations related to Λ_e by $\lambda_{\text{in}}(\Lambda_e)$. We claim that

$$\lambda_{\text{in}}(\Lambda_e) \leq r^r. \tag{3.8}$$

Indeed, in order to obtain an in-switching configuration $\Lambda_{\overline{e}} = (f_1, \dots, f_r, f)$ related to Λ_e one has to choose $u_i \in e_i$ for each $i \in [r]$. There are at most r choices for each u_i. Each (admissible) choice of u_i uniquely determines f_i, and thus they determine f.

Similarly, given $\Lambda_{\overline{e}} = (f_1, \dots, f_r, f)$, we denote the number of out-switching configurations related to $\Lambda_{\overline{e}}$ by $\lambda_{\text{out}} = \lambda_{\text{out}}(\Lambda_{\overline{e}})$. We claim that

$$\lambda_{\text{out}} = r!. \tag{3.9}$$

This holds because, for each $i \in [r]$, the edge e_i must contain $f_i \setminus \{v_i\} = A_i$ and one vertex $u_i \in f$, hence each permutation of the labels of the vertices in f results in a different Λ_e.

We define $\psi(G, \Lambda_e, \Lambda_{\overline{e}})$, $\mathcal{F}_e$, $\mathcal{F}_{\overline{e}}$ and Γ as in the proof of lemma 3.2.1. As before, neither out- nor in-switchings on an r-graph G change the vertex degrees.

Assume first that $G \in \mathcal{F}_e$. Let $S_{\text{out}}(G)$ be the number of possible out-switchings $(G, \Lambda_e, \Lambda_{\overline{e}})$ on G satisfying that $\psi(G, \Lambda_e, \Lambda_{\overline{e}}) \in \mathcal{G}^{(r)}_{n,d,H,H'}$. Thus $\deg_\Gamma(G) = S_{\text{out}}(G)$. Let $S_{\text{out}} := \sum_{G\in\mathcal{F}_e} S_{\text{out}}(G)$ be the number of edges incident to $\mathcal{F}_e$ in Γ. We claim that

$$S_{\text{out}}(G) \leq (nd)^r. \tag{3.10}$$

Indeed, (3.8) implies that $S_{\text{out}}(G)$ is at most the number of out-switching configurations $\Lambda_e \subseteq G$ multiplied by r^r. The number of such out-switching configurations is given by the choice of $(e_1, \dots, e_r)$, so there are at most $(nd/r)^r$ such configurations. This yields (3.10). As this is true for every G,

$$S_{\text{out}} \leq |\mathcal{F}_e|(nd)^r. \tag{3.11}$$

Consider now any r-graph $G \in \mathcal{F}_{\overline{e}}$. Let $S_{\text{in}}(G)$ be the number of possible in-switchings $(G, \Lambda_{\overline{e}}, \Lambda_e)$ on G satisfying that $\psi(G, \Lambda_{\overline{e}}, \Lambda_e) \in \mathcal{G}^{(r)}_{n,d,H,H'}$. Thus $\deg_\Gamma(G) = S_{\text{in}}(G)$. Let $S_{\text{in}} := \sum_{G\in\mathcal{F}_{\overline{e}}} S_{\text{in}}(G)$ be the number of edges incident to $\mathcal{F}_{\overline{e}}$ in Γ. Let $T_{\text{in}}(G)$ denote the number of in-switching configurations $\Lambda_{\overline{e}} \subseteq G$. As an in-switching configuration is given by r edges, one incident to each of the vertices of e, and one more edge which is disjoint from the previous ones, by choosing each edge in turn and taking into consideration that G is d-regular, we conclude that

$$T_{\text{in}}(G) \leq \frac{nd^{r+1}}{r}. \tag{3.12}$$

For a lower bound on $T_{\text{in}}(G)$, observe that there are exactly d choices for f_1. Then, f_2 can be chosen in at least $d-1$ ways. More generally, there are at least $(d-r)^r$ choices for $(f_1, \dots, f_r)$. Finally, f must be chosen disjoint from $f_1, \dots, f_r$, so there are at least $nd/r - r^2 d$ choices.

Overall,

$$T_{\mathrm{in}}(G) \geq (d-r)^r\left(\frac{nd}{r} - r^2 d\right) = \frac{nd^{r+1}}{r}\left(1 - \mathcal{O}\left(\frac{1}{d} + \frac{1}{n}\right)\right). \tag{3.13}$$

We say that an in-switching configuration $\Lambda_{\overline{e}} \subseteq G$ is *good* (for G) if there are λ_{out} out-switching configurations $\Lambda_e \subseteq \overline{G}$ related to $\Lambda_{\overline{e}}$ which satisfy $\psi(G, \Lambda_{\overline{e}}, \Lambda_e) \in \mathcal{G}^{(r)}_{n,d,H,H'}$. We say that $\Lambda_{\overline{e}}$ is *bad* (for G) otherwise. An in-switching configuration $\Lambda_{\overline{e}} = (f_1, \ldots, f_r, f)$ is bad for G if and only if any of the following occur:

(a) $(f_i \setminus \{v_i\}) \cup \{v\} \in H$ for some $i \in [r]$ and $v \in f$.

(b) $(f_i \setminus \{v_i\}) \cup \{v\} \in H'$ for some $i \in [r]$ and $v \in f$.

(c) $f_i \in H$ for some $i \in [r]$ or $f \in H$.

(d) Neither (a) nor (b) hold, but $(f_i \setminus \{v_i\}) \cup \{v\} \in G$ for some $i \in [r]$ and $v \in f$.

For each $G \in \mathcal{F}_{\overline{e}}$, let $\mathcal{L}(G)$ denote the set of in-switching configurations $\Lambda_{\overline{e}}$ with $\Lambda_{\overline{e}} \subseteq G$. Consider the set $\Omega := \{(G, \Lambda_{\overline{e}}) \mid G \in \mathcal{F}_{\overline{e}}, \Lambda_{\overline{e}} \in \mathcal{L}(G)\}$. We say that a pair $(G, \Lambda_{\overline{e}})$ is *bad* if $\Lambda_{\overline{e}}$ is bad for G.

Let $k := \Delta(H)$, $k' := \Delta(H')$. We first count the number of in-switching configurations in $\mathcal{L}(G)$ which are bad because of (a)–(c). For this, fix an r-graph $G \in \mathcal{F}_{\overline{e}}$. Let $T_{\mathrm{a}}(G)$ be the number of in-switching configurations which are bad because of (a). Fix $e^* \in H$ and $i \in [r]$. To count the number of in-switching configurations $\Lambda_{\overline{e}} = (f_1, \ldots, f_r, f) \in \mathcal{L}(G)$ with $(f_i \setminus \{v_i\}) \cup \{v\} = e^*$ for some $v \in f$, note that there are at most r choices for v, and then at most d choices for f (since $v \in f$). Then we must have $f_i = (e^* \setminus \{v\}) \cup \{v_i\}$. Finally, there are at most d choices for each f_j with $j \in [r] \setminus \{i\}$ (since $v_j \in f_j$). Therefore, $T_{\mathrm{a}}(G) \leq |H| \cdot r \cdot r \cdot d \cdot d^{r-1} \leq rnkd^r$. Let $T_{\mathrm{a}} := \sum_{G \in \mathcal{F}_{\overline{e}}} T_{\mathrm{a}}(G)$ be the number of pairs $(G, \Lambda_{\overline{e}})$ which are bad because of (a). Then,

$$T_{\mathrm{a}} \leq |\mathcal{F}_{\overline{e}}| rnkd^r. \tag{3.14}$$

Similarly, for $G \in \mathcal{F}_{\overline{e}}$, let $T_{\mathrm{b}}(G)$ be the number of in-switching configurations which are bad because of (b). As above, one can show that $T_{\mathrm{b}}(G) \leq |H'| \cdot r \cdot r \cdot d \cdot d^{r-1} \leq rnk'd^r$. Let $T_{\mathrm{b}} \coloneqq \sum_{G \in \mathcal{F}_{\overline{e}}} T_{\mathrm{b}}(G)$ be the number of pairs $(G, \Lambda_{\overline{e}})$ which are bad because of (b). Then,

$$T_{\mathrm{b}} \leq |\mathcal{F}_{\overline{e}}| rnk'd^r. \tag{3.15}$$

Next, for $G \in \mathcal{F}_{\overline{e}}$, let $T_{\mathrm{c}}(G)$ be the number of in-switching configurations which are bad because of (c). Given $i \in [r]$, there are at most k choices for $f_i \in H$ (as $v_i \in f_i$), and the remaining edges in the in-switching configuration can be chosen in at most $d^{r-1}nd/r$ ways. Similarly, if $f \in H$, then the remaining edges in the in-switching configuration can be chosen in at most d^r ways. Therefore, $T_{\mathrm{c}}(G) \leq r \cdot k \cdot d^{r-1}nd/r + |H| \cdot d^r \leq (r+1)nkd^r/r$. Let $T_{\mathrm{c}} \coloneqq \sum_{G \in \mathcal{F}_{\overline{e}}} T_{\mathrm{c}}(G)$ be the number of pairs $(G, \Lambda_{\overline{e}})$ which are bad because of (c). Then,

$$T_{\mathrm{c}} \leq |\mathcal{F}_{\overline{e}}| \frac{(r+1)nkd^r}{r}. \tag{3.16}$$

Finally, we count the number of in-switching configurations which are bad because of (d). For this, fix $\Lambda_{\overline{e}} = (f_1, \dots, f_r, f) \in \bigcup_{G \in \mathcal{F}_{\overline{e}}} \mathcal{L}(G)$. Note that this implies that $\Lambda_{\overline{e}} \cap H' = \varnothing$. We now apply lemma 3.2.1 with $H \cup \Lambda_{\overline{e}}$ playing the role of H and $H' \cup \{e\}$ playing the role of H' to bound the number of pairs $(G, \Lambda_{\overline{e}})$ that are bad because of (d). We denote this number by T_{d}. Lemma 3.2.1 implies that, for any $\hat{e} \in \binom{V}{r} \setminus (H \cup H' \cup \Lambda_{\overline{e}} \cup \{e\})$,

$$\mathbb{P}\left[\hat{e} \in G_{n,d}^{(r)} \mid \mathcal{G}_{n,d,H \cup \Lambda_{\overline{e}}, H' \cup \{e\}}^{(r)}\right] \leq 2(r-1)! \frac{d}{n^{r-1}}.$$

In particular, this holds for all r-sets of the form $(f_i \setminus \{v_i\}) \cup \{v\}$ for some $i \in [r]$ and $v \in f$ (as long as they are not in H or H', which is guaranteed for condition (d)). Therefore, a union bound yields an upper bound on the probability that $\Lambda_{\overline{e}}$ is bad for G because of (d). Indeed, let $\mathcal{B}(G, \Lambda_{\overline{e}})$ denote the event that the pair $(G, \Lambda_{\overline{e}})$ is bad because of (d). Then,

$$\mathbb{P}\left[\mathcal{B}(G_{n,d}^{(r)}, \Lambda_{\overline{e}}) \mid \mathcal{G}_{n,d,H \cup \Lambda_{\overline{e}}, H' \cup \{e\}}^{(r)}\right] \leq 2r^2(r-1)! \frac{d}{n^{r-1}}. \tag{3.17}$$

The same approach works for all $\Lambda_{\overline{e}}$. By (3.12) we have that $|\Omega| \leq |\mathcal{F}_{\overline{e}}| n d^{r+1}/r$. Moreover, note that

$$|\Omega| = \sum_{\Lambda_{\overline{e}} \in \bigcup_{G \in \mathcal{F}_{\overline{e}}} \mathcal{L}(G)} |\mathcal{G}^{(r)}_{n,d,H \cup \Lambda_{\overline{e}}, H' \cup \{e\}}|. \tag{3.18}$$

Hence, for the number T_{d} of pairs that are bad because of (d), by (3.17) and (3.18) it follows that

$$T_{\mathrm{d}} = \sum_{\Lambda_{\overline{e}} \in \bigcup_{G \in \mathcal{F}_{\overline{e}}} \mathcal{L}(G)} |\mathcal{G}^{(r)}_{n,d,H \cup \Lambda_{\overline{e}}, H' \cup \{e\}}| \cdot \mathbb{P}\left[\mathcal{B}(G^{(r)}_{n,d}, \Lambda_{\overline{e}}) \mid \mathcal{G}^{(r)}_{n,d,H \cup \Lambda_{\overline{e}}, H' \cup \{e\}}\right] \leq |\mathcal{F}_{\overline{e}}| 2r! \frac{d^{r+2}}{n^{r-2}}. \tag{3.19}$$

By (3.13) we have that $|\Omega| \geq |\mathcal{F}_{\overline{e}}| \frac{nd^{r+1}}{r} \left(1 - \mathcal{O}\left(\frac{1}{d} + \frac{1}{n}\right)\right)$. Let $\varepsilon := \varepsilon(n,d,k,k') = \frac{1}{n} + \frac{1}{d} + \frac{d}{n^{r-1}} + \frac{k}{d} + \frac{k'}{d}$. By (3.9) and (3.14)–(3.19), we conclude that

$$S_{\mathrm{in}} \geq \lambda_{\mathrm{out}}(|\Omega| - T_{\mathrm{a}} - T_{\mathrm{b}} - T_{\mathrm{c}} - T_{\mathrm{d}}) = |\mathcal{F}_{\overline{e}}|(r-1)! n d^{r+1} (1 - \mathcal{O}(\varepsilon)). \tag{3.20}$$

Combining (3.11) and (3.20), we conclude that

$$|\mathcal{F}_{\overline{e}}|(r-1)! n d^{r+1} (1 - \mathcal{O}(\varepsilon)) \leq S_{\mathrm{in}} = S_{\mathrm{out}} \leq |\mathcal{F}_e| (nd)^r.$$

Dividing this by $|\mathcal{G}^{(r)}_{n,d,H,H'}|$ implies that

$$(r-1)! n d^{r+1} (1 - \mathcal{O}(\varepsilon)) \mathbb{P}\left[e \notin G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}\right] \leq (nd)^r \mathbb{P}\left[e \in G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}\right].$$

Taking into account that $\mathbb{P}[e \notin G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}] = 1 - \mathbb{P}[e \in G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}]$, we conclude that

$$\mathbb{P}\left[e \in G^{(r)}_{n,d} \mid \mathcal{G}^{(r)}_{n,d,H,H'}\right] \geq (r-1)! \frac{d}{n^{r-1}} (1 - \mathcal{O}(\varepsilon)). \qquad \square$$

Together, lemmas 3.2.1 and 3.2.2 imply the following result.

Corollary 3.2.3. *Let $r \geq 2$ be a fixed integer. Suppose that $d = \omega(1)$ and $d = o(n^{r-1})$. Let $H, H' \subseteq \binom{V}{r}$ be two edge-disjoint r-graphs such that $\Delta(H), \Delta(H') = o(d)$. Then, for all $e \in \binom{V}{r} \setminus (H \cup H')$ we have*

$$\mathbb{P}\left[e \in G_{n,d}^{(r)} \mid \mathcal{G}_{n,d,H,H'}^{(r)}\right] = (r-1)!\frac{d}{n^{r-1}}\left(1 \pm \mathcal{O}\left(\frac{1}{n} + \frac{1}{d} + \frac{d}{n^{r-1}} + \frac{\Delta(H)}{d} + \frac{\Delta(H')}{d}\right)\right).$$

3.3. Counting subgraphs of random regular r-graphs

In this section we use the results of section 3.2 to count the number of copies of certain r-graphs F inside a random d-regular r-graph. In section 3.3.1 we consider the case when F is fixed. In particular, we will derive results on the number of edge-disjoint copies of F, which will be used in section 3.4.2. In section 3.3.2 we apply our results to count the number of copies of sparse but possibly spanning r-graphs such as Hamilton cycles.

3.3.1. Counting small subgraphs

For an r-graph F, let $\mathrm{aut}(F)$ denote the number of automorphisms of F. Let $X_F(G)$ denote the number of (unlabelled) copies of F in an r-graph G. We will often just write X_F whenever G is clear from the context. Observe that X_F is a random variable whenever G is randomly chosen from some set $\mathcal{G}$. We will consider the uniform distribution on the set $\mathcal{G}_{n,d}^{(r)}$. Furthermore, we define

$$p := (r-1)!\frac{d}{n^{r-1}} \qquad \text{and} \qquad \varepsilon_{n,d} := \frac{1}{n} + \frac{1}{d} + \frac{d}{n^{r-1}}.$$

Corollary 3.3.1. *Let $r \geq 2$ and $t \geq 1$ be fixed integers, and let F be a fixed r-graph. Suppose that $d = \omega(1)$ and $d = o(n^{r-1})$. Then,*

(i) *for any set $\mathcal{E} \subseteq \binom{V}{r}$ of size t, $\mathbb{P}[\mathcal{E} \subseteq G_{n,d}^{(r)}] = p^t\left(1 \pm \mathcal{O}\left(\varepsilon_{n,d}\right)\right)$,*

(ii) $\mathbb{E}[X_F] = \binom{n}{v_F}\frac{v_F!}{\mathrm{aut}(F)}p^{e_F}\left(1 \pm \mathcal{O}\left(\varepsilon_{n,d}\right)\right)$.

Proof. Enumerate the edges in $\mathcal{E}$ as $e_1, \dots, e_t$. Item (i) follows by applying corollary 3.2.3 repeatedly. This in turn implies (ii). □

The next lemma implies that X_F is concentrated around $\mathbb{E}[X_F]$ whenever $\Phi_F = \omega(1)$, where

$$\Phi_F := \min\{\mathbb{E}[X_K] : K \subseteq F, e_K > 0\}.$$

Lemma 3.3.2. *Let $r \geq 2$ be a fixed integer. Suppose that $d = \omega(1)$ and $d = o(n^{r-1})$. Then, for any fixed r-graph F with $e_F \geq 1$, we have that* $\mathrm{Var}[X_F] = \mathcal{O}(\varepsilon_{n,d} + \Phi_F^{-1})\mathbb{E}[X_F]^2$.

The proof follows a straightforward second moment approach (based on corollary 3.3.1), so we omit the details (for a proof of the same statement in $\mathcal{G}_{n,p}$, see for instance [71, Lemma 3.5]). Corollary 3.3.1, lemma 3.3.2 and Chebyshev's inequality imply the following result. In particular, this determines the threshold for the appearance of a copy of a fixed F in $\mathcal{G}_{n,d}^{(r)}$.

Corollary 3.3.3. *Let $r \geq 2$ be a fixed integer. Suppose that $d = \omega(1)$ and $d = o(n^{r-1})$. Then, for any fixed r-graph F with $\Phi_F = \omega(1)$, we a.a.s. have*

$$X_F = (1 \pm o(1))\binom{n}{v_F}\frac{v_F!}{\mathrm{aut}(F)}p^{e_F}.$$

The next result addresses the problem of counting edge-disjoint copies of an r-graph F in $G_{n,d}^{(r)}$. Its proof builds on an idea of Kreuter [84] for counting vertex-disjoint copies in the binomial random graph model (see also [71, Theorem 3.29]). The approach is to consider an auxiliary graph whose vertex set consists of the copies of F in $G_{n,d}^{(r)}$ and where an independent set corresponds to a set of edge-disjoint copies of F. To estimate the number of vertices and edges of this graph (with a view to apply Turán's theorem), one makes use of corollary 3.3.1, lemma 3.3.2 and corollary 3.3.3. For the sake of completeness, we include the details in appendix A.4.

Lemma 3.3.4. *Let F be a fixed r-graph. Assume that $d = \omega(1)$ and $d = o(n^{r-1})$. Let D_F be the maximum number of edge-disjoint copies of F in an r-graph chosen uniformly from $\mathcal{G}_{n,d}^{(r)}$. If $\Phi_F = \omega(1)$, then $D_F = \Theta(\Phi_F)$ a.a.s.*

3.3.2. Counting spanning graphs

Let $H = \{H_i\}_{i \geq 1}$ be a sequence of r-graphs with $|V(H_i)|$ strictly increasing. When we say that H is a subgraph of G, for some G of order n, we mean that the corresponding H_i of order n is a subgraph of G. This makes sense only when $n = |V(H_i)|$ for some i; we will implicitly assume this is the case, and study the asymptotic behaviour as i tends to infinity.

Our main tool for this section is the following result of Dudek, Frieze, Ruciński, and Šileikis [43], which allows to translate results on the $\mathcal{G}^{(r)}(n,p)$ and $\mathcal{G}^{(r)}(n,m)$ random graph models to $\mathcal{G}^{(r)}_{n,d}$. Roughly speaking, their result asserts that $G^{(r)}(n,p) \subseteq G^{(r)}_{n,d}$ a.a.s. provided that p is at least a little smaller than $d/\binom{n-1}{r-1}$. For the graph case, a similar result was proved by Kim and Vu [76] (for a more restricted range of d).

Theorem 3.3.5 ([43]). *For every $r \geq 2$ there exists a constant $C > 0$ such that if for some positive integer $d = d(n)$,*

$$\delta_{n,d} \coloneqq C\left(\left(\frac{d}{n^{r-1}} + \frac{\log n}{d}\right)^{1/3} + \frac{1}{n}\right) < 1, \tag{3.21}$$

then there is a joint distribution of $G^{(r)}(n,p_d)$ and $G^{(r)}_{n,d}$ such that

$$\lim_{n\to\infty} \mathbb{P}\left[G^{(r)}(n,p_d) \subseteq G^{(r)}_{n,d}\right] = 1,$$

where $p_d \coloneqq (1-\delta_{n,d})d/\binom{n-1}{r-1}$. The analogous statement also holds with $G^{(r)}(n,p_d)$ replaced by $G^{(r)}(n,m_d)$ for $m_d \coloneqq (1-\delta_{n,d})nd/r$.

In order to be able to apply theorem 3.3.5, from now on we always assume that $d = o(n^{r-1})$ and $d = \omega(\log n)$. We now combine theorem 3.3.5 with our results from section 3.2 to obtain a general result relating subgraph counts in $\mathcal{G}^{(r)}_{n,d}$ to those in $G^{(r)}(n,p_d)$ and $G^{(r)}(n,m_d)$.

Theorem 3.3.6. *Let $r \geq 2$ be a fixed integer and V be a set of n vertices. Assume that $d = \omega(\log n)$ and $d = o(n^{r-1})$. Let H be an r-graph on V with $\Delta(H) = \mathcal{O}(1)$. Suppose that*

$\eta = \eta(n) = o(1)$ is such that

$$\varepsilon_{n,d} = o(\eta), \qquad \delta_{n,d} = o(\eta), \qquad \eta = \omega(1/n), \tag{3.22}$$

and $X_H(G^{(r)}(n, p_d)) = (1 \pm \eta)^{|H|}\mathbb{E}[X_H(G^{(r)}(n, p_d))]$ a.a.s. Then, a.a.s.

$$X_H(G^{(r)}_{n,d}) = (1 \pm 3\eta)^{|H|}\mathbb{E}[X_H(G^{(r)}(n, p_d))]. \tag{3.23}$$

Similarly, if (3.22) holds and $X_H(G^{(r)}(n, m_d)) = (1 \pm \eta)^{|H|}\mathbb{E}[X_H(G^{(r)}(n, m_d))]$ a.a.s., then a.a.s.

$$X_H(G^{(r)}_{n,d}) = (1 \pm 3\eta)^{|H|}\mathbb{E}[X_H(G^{(r)}(n, m_d))]. \tag{3.24}$$

Proof. Observe first that, by corollary 3.2.3, for any fixed copy H' of H we have

$$\mathbb{P}\left[H' \subseteq G^{(r)}_{n,d}\right] = ((1 \pm \mathcal{O}(\varepsilon_{n,d}))(r-1)!d/n^{r-1})^{|H|}. \tag{3.25}$$

Therefore,

$$\frac{\mathbb{E}[X_H(G^{(r)}_{n,d})]}{\mathbb{E}[X_H(G^{(r)}(n, p_d))]} = (1 \pm \mathcal{O}(\varepsilon_{n,d} + \delta_{n,d}))^{|H|} \leq (1+\eta)^{|H|}. \tag{3.26}$$

By using Markov's inequality and (3.26) we conclude that

$$\begin{aligned} &\mathbb{P}\left[X_H(G^{(r)}_{n,d}) \geq (1+3\eta)^{|H|}\mathbb{E}\left[X_H(G^{(r)}(n, p_d))\right]\right] \\ \leq &\mathbb{P}\left[X_H(G^{(r)}_{n,d}) \geq (1+\eta)^{|H|}\mathbb{E}\left[X_H(G^{(r)}_{n,d})\right]\right] \leq 1/(1+\eta)^{|H|} = o(1). \end{aligned} \tag{3.27}$$

Note that, as $G^{(r)}(n, p_d) \subseteq G^{(r)}_{n,d}$ a.a.s. by theorem 3.3.5, then $X_H(G^{(r)}_{n,d}) \geq X_H(G^{(r)}(n, p_d))$ a.a.s. Thus, by assumption,

$$\begin{aligned} &\mathbb{P}\left[X_H(G^{(r)}_{n,d}) \leq (1-\eta)^{|H|}\mathbb{E}\left[X_H(G^{(r)}(n,p_d))\right]\right] \\ &\leq \mathbb{P}\left[X_H(G^{(r)}(n,p_d)) \leq (1-\eta)^{|H|}\mathbb{E}\left[X_H(G^{(r)}(n,p_d))\right]\right] + o(1) = o(1). \end{aligned} \tag{3.28}$$

Combining equations (3.27) and (3.28) yields (3.23).

Finally, one can prove (3.24) in a very similar way. □

We may apply theorem 3.3.6 to obtain estimates on the number of copies of certain spanning subgraphs. This requires concentration results in the $\mathcal{G}^{(r)}(n, p)$ model or the $\mathcal{G}^{(r)}(n, m)$ model in order to obtain results for $\mathcal{G}^{(r)}_{n,d}$.

We start with the following result of Glebov and Krivelevich [57] on counting Hamilton cycles in $\mathcal{G}(n, p)$. For a more restricted range of densities, Janson [69] proved more precise results in $\mathcal{G}(n, m)$.

Theorem 3.3.7 ([57]). *Let V be a set of n vertices. Let H be a Hamilton cycle on V. If $p \geq \frac{\ln n + \ln\ln n + \omega(1)}{n}$, then a.a.s.*

$$X_H(G(n, p)) = (1 \pm o(1))^n n! p^n.$$

Together with theorem 3.3.6 this implies the following result.

Corollary 3.3.8. *Let V be a set of n vertices. Let H be a Hamilton cycle on V. Assume $d = \omega(\log n)$ and $d = o(n)$; then, a.a.s.*

$$X_H(G_{n,d}) = (1 \pm o(1))^n n! \left(\frac{d}{n-1}\right)^n.$$

Corollary 3.3.8 improves a previous result of Krivelevich [85] by increasing the range of d in which the number of Hamilton cycles is estimated from $d = \omega(e^{(\log n)^{1/2}})$ to $d = \omega(\log n)$.

Note that, on the other hand, the results of Krivelevich [85] also cover pseudo-random d-regular graphs.

A very general result due to Riordan [102] allows us to count the number of copies of H as a spanning subgraph of $G(n,m)$ for a large class of graphs H. We state only a special case of this result here. Let $\alpha_1(H) := |H|/\binom{n}{2}$, $\alpha_2(H) = X_{P_2}(H)/(3\binom{n}{3})$ (where P_2 stands for a path of length 2), $e_H(k) := \max\{|F| : F \subseteq H, |V(F)| = k\}$, $\gamma_1(H) := \max_{3\leq k\leq n}\{e_H(k)/(k-2)\}$ and $\gamma_2(H) := \max_{5\leq k\leq n}\{(e_H(k)-4)/(k-4)\}$.

Theorem 3.3.9 ([102]). *Let V be a set of n vertices. Let $p = \omega(\max\{1/n^{1/2}, 1/n^{1/\gamma_1}, 1/n^{1/\gamma_2}\})$, $p = o(1/\log n)$, $m := p\binom{n}{2}$, and let H be a triangle-free spanning graph on V with $|H| \geq n$, $\Delta(H) = \mathcal{O}(1)$ and $|\alpha_2(H) - \alpha_1(H)^2| = \Omega(1/n^2)$. Then, $X_H(G(n,m))$ follows a normal distribution such that* $\mathrm{Var}[X_H(G(n,m))]/\mathbb{E}[X_H(G(n,m))]^2 = o(1)$.

Together with theorem 3.3.6, we can deduce the following.

Corollary 3.3.10. *Let V be a set of n vertices. Assume that $d = \omega(\max\{n^{1/2}, n^{1-1/\gamma_1}, n^{1-1/\gamma_2}\})$, $d = o(n/\log n)$, and let H be a triangle-free spanning graph on V with $|H| \geq n$, $\Delta(H) = \mathcal{O}(1)$ and $|\alpha_2(H) - \alpha_1(H)^2| = \Omega(1/n^2)$. Then, $X_H(G_{n,d}) = (1 \pm o(1))^n \mathbb{E}[X_H(G(n, m_d))]$ a.a.s., where $m_d = (1-o(1))dn/2$ is defined as in theorem 3.3.6.*

As a particular case of this, we can estimate the number of spanning square lattices in a random d-regular graph. A square lattice L_k is defined by setting $V(L_k) = [k] \times [k]$ and $L_k = \{\{(x,y),(u,v)\} : u,v,x,y \in [k], \|(x,y)-(u,v)\| = 1\}$.

Corollary 3.3.11. *Let $n = k^2$. Let $d = \omega(1)$, $d = o(n/\log n)$ and $p := d/(n-1)$.*

(i) *If $d = o(n^{1/2})$, then $\mathbb{P}[X_{L_k}(G_{n,d}) > 0] = o(1)$.*

(ii) *If $d = \omega(n^{1/2})$, then, $X_{L_k}(G_{n,d}) = (1 \pm o(1))^n n! p^{|L_k|}$ a.a.s.*

In particular, as $|L_k| = 2n \pm \mathcal{O}(n^{1/2})$, this determines the threshold for the existence of a spanning square lattice L_k in $G_{n,d}$. Corollary 3.3.11(i) follows from corollary 3.2.3 and Markov's inequality, while corollary 3.3.11(ii) follows from corollary 3.3.10.

Much less is known for r-graphs when $r \geq 3$. For Hamilton cycles, we can apply the following result of Dudek and Frieze [40] on ℓ-overlapping Hamilton cycles.

Theorem 3.3.12 ([40], Section 2)**.** *Let $r > \ell \geq 2$ and assume that $(r-\ell) \mid n$. Assume $p = \omega(1/n^{r-\ell})$. Then, a.a.s.*

$$X_{C_n^\ell}(G^{(r)}(n,p)) = (1 \pm o(1))^n n! p^{n/(r-\ell)}.$$

Together with theorem 3.3.6, corollary 3.2.3 and Markov's inequality, this implies the following result.

Corollary 3.3.13. *Let $r > \ell \geq 2$ and assume that $(r-\ell) \mid n$. Let $p := d/\binom{n-1}{r-1}$.*

(i) *If $d = o(n^{\ell-1})$ then $\mathbb{P}[X_{C_n^\ell}(G_{n,d}^{(r)}) > 0] = o(1)$.*

(ii) *If $d = \omega(n^{\ell-1})$ and $d = o(n^{r-1})$, then a.a.s. $X_{C_n^\ell}(G_{n,d}^{(r)}) = (1 \pm o(1))^n n! p^{n/(r-\ell)}$.*

In particular, this determines the threshold for the existence of C_n^ℓ in $\mathcal{G}_{n,d}^{(r)}$ for $\ell \in [r-1] \setminus \{1\}$, solving a conjecture of Dudek, Frieze, Ruciński, and Šileikis [43]. We note that Altman, Greenhill, Isaev, and Ramadurai [7] recently determined the threshold for the appearance of loose Hamilton cycles in random regular r-graphs. Their results imply that for every $r \geq 3$ there exists a value d_0 (which is calculated explicitly in [7]) such that if $d \geq d_0$, then $G_{n,d}^{(r)}$ a.a.s. has a loose Hamilton cycle. For $\ell \in [r-1] \setminus \{1\}$, they also proved that $\mathbb{P}[X_{C_n^\ell}(G_{n,d}^{(r)}) > 0] = o(1)$ holds under the much stronger condition that $d = o(n)$ if $r \geq 4$ and $d = o(n^{1/2})$ if $r = 3$ (but to deduce corollary 3.3.13(i) we do rely on their result when d is constant; we rely on corollary 3.2.3 when $d = \omega(1)$).

3.4. Testing F-freeness in general r-graphs

We now give lower and upper bounds on the query complexity of testing F-freeness in the general r-graphs model, where F is a fixed r-graph. In the special case when F is a triangle, these (and other) bounds were already obtained by Alon, Kaufman, Krivelevich, and Ron [3]. Our proofs build on ideas from their paper.

In section 3.4.1, we observe a simple lower bound for the query complexity of any F-freeness tester. In section 3.4.2, we use our results from sections 3.2 and 3.3 to improve this bound for input r-graphs whose density is larger than a certain threshold. The bound that we obtain, however, holds only for one-sided error testers; extending it to two-sided error testers, as Alon, Kaufman, Krivelevich, and Ron [3] do with their triangle-freeness tester, would be an interesting problem. Finally, section 3.4.3 is devoted to upper bounds on the query complexity.

3.4.1. A lower bound for sparser r-graphs

In this section we provide a lower bound on the query complexity of testing F-freeness which is stronger than that in section 3.4.2 when the r-graphs that are being tested are sparser (the range of the average degree d for which this holds depends on the particular r-graph F). Recall that our algorithms are allowed to perform two types of queries: vertex-set queries and neighbour queries. For a fixed r-graph F, let $ex(n, F)$ denote the maximum number of edges of an F-free r-graph G on n vertices.

Proposition 3.4.1. *Let $r \geq 2$ and F be an r-graph. Let $c, a > 0$ be fixed constants such that $c \cdot n^a \leq ex(n, F)$ and suppose that $d = \Omega(1)$ and $d = o(n^{a-1})$. Then, any F-freeness tester in r-graphs must perform $\Omega\left(n^{1-1/a} d^{-1/a}\right)$ queries, when restricted to input r-graphs on n vertices of average degree $d \pm o(d)$.*

Observe that the assumptions in the statement imply that $1 < a \leq r$. In particular, the result only applies for r-graphs F such that $ex(n, F)$ is superlinear.

Proof. It suffices to construct two families of r-graphs on n vertices $\mathcal{F}_1$ and $\mathcal{F}_2$ such that the following hold:

(i) All r-graphs in $\mathcal{F}_1$ are F-free.

(ii) All r-graphs in $\mathcal{F}_2$ are $\Theta(1)$-far from F-free.

(iii) All r-graphs in both families have average degree $d \pm o(d)$.

(iv) Consider an r-graph G chosen from $\mathcal{F}_1 \cup \mathcal{F}_2$ according to the following rule. First choose $i \in [2]$ uniformly at random. Then choose $G \in \mathcal{F}_i$ uniformly at random. Then any algorithm that determines with probability at least $2/3$ whether $G \in \mathcal{F}_1$ or $G \in \mathcal{F}_2$ must perform at least $\Omega(n^{1-1/a}d^{-1/a})$ queries.

Let H be an F-free r-graph on $(nd/(cr))^{1/a}$ vertices with nd/r edges. Let $\mathcal{F}_1$ be the family of all labelled r-graphs consisting of the disjoint union of H on $(nd/(cr))^{1/a}$ vertices and $n-(nd/(cr))^{1/a}$ isolated vertices. Let $\mathcal{F}_2$ be the family of all labelled r-graphs consisting of the disjoint union of a complete r-graph on a set of $(nd(r-1)!)^{1/r}$ vertices and $n-(nd(r-1)!)^{1/r}$ isolated vertices.

A simple computation shows that all r-graphs in both families have average degree $d \pm o(d)$. All r-graphs in $\mathcal{F}_1$ are F-free by definition. Since the number of distinct $K_{v_F}^{(r)}$ in $K_k^{(r)}$ is $\Theta(k^{v_F})$, it is easy to check that all r-graphs in $\mathcal{F}_2$ are $\Theta(1)$-far from being $K_{v_F}^{(r)}$-free, and hence $\Theta(1)$-far from being F-free. Thus, conditions (i), (ii) and (iii) hold.

Now consider any algorithm ALG that, given an r-graph G chosen at random from either $\mathcal{F}_1$ or $\mathcal{F}_2$ as in (iv), tries to determine with probability at least $2/3$ whether $G \in \mathcal{F}_1$ or $G \in \mathcal{F}_2$. If $G \in \mathcal{F}_1$, then the probability of finding a vertex with positive degree with any given query is $\mathcal{O}(n^{1/a-1}d^{1/a})$. Similarly, if $G \in \mathcal{F}_2$, the probability of finding a vertex with positive degree with any given query is $\mathcal{O}(n^{1/a-1}d^{1/a})$. Hence, if the number of queries is $Q = o(n^{1-1/a}d^{-1/a})$, by the union bound, one has that the probability of finding any such vertex is $o(1)$. So a.a.s. ALG only finds a set of isolated vertices, of size $\mathcal{O}(Q)$, after the first Q queries. Thus we conclude that, for $i \in [2]$, $\mathbb{P}[G \in \mathcal{F}_i \mid \text{ALG finds only isolated vertices}] = 1/2 \pm o(1)$. Therefore, the algorithm cannot distinguish between r-graphs in $\mathcal{F}_1$ and $\mathcal{F}_2$ with sufficiently high probability with only Q queries. □

If F is a non-r-partite r-graph, then $ex(n, F) = \Theta(n^r)$. Using this, proposition 3.4.1 asserts that, for any non-r-partite r-graph F, testing F-freeness needs $\Omega((n^{r-1}/d)^{1/r})$ queries. This implies that for all non-r-partite r-graphs F there is no constant time F-freeness tester for input r-graphs G on n vertices with $d = o(n^{r-1})$ and $d = \Omega(1)$, as opposed to the constant time algorithms existing for dense r-graphs.

In more generality, proposition 3.4.1 shows that there can be no F-freeness tester that requires a constant number of queries whenever the input r-graph G has average degree $d = o(ex(n, F)/n)$ and $d = \Omega(1)$. On the other hand, if the number of edges of the input r-graph is larger than the Turán number of F, then there is a trivial F-freeness tester: an algorithm that rejects every input, which has constant query complexity. As another example, it is well-known that $ex(n, C_4) = \Theta(n^{3/2})$. With this, we conclude that any algorithm testing C_4-freeness in graphs with average degree d, when $d = o(n^{1/2})$ and $d = \Omega(1)$, must perform at least $\Omega((n/d^2)^{1/3})$ queries.

The asymptotic growth of $ex(n, F)$ is not known for every F. Let $\beta(F) := \frac{v_F - r}{e_F - 1}$. An easy probabilistic argument shows that $ex(n, F) = \Omega\left(n^{r-\beta(F)}\right)$. This bound is superlinear in n as long as $\beta(F) < r - 1$, which holds for every connected F that is not a weak tree. Using this bound on $ex(n, F)$, proposition 3.4.1 asserts that for any connected r-graph F other than a weak tree the number of queries performed by any F-freeness tester on input r-graphs on at least $\Omega(n)$ and at most $o\left(n^{r-\beta(F)}\right)$ edges is $\Omega((n^{r-1-\beta(F)}/d)^{1/(r-\beta(F))})$.

3.4.2. A lower bound for denser r-graphs

The lower bound on the query complexity of F-freeness testers we present here improves the bound in section 3.4.1 when d is large enough and either $r = 2$ or $r \geq 3$ and F is non-r-partite. However, this approach works only for one-sided error algorithms. The answer given by one-sided error algorithms must always be correct when the input r-graph is F-free, so any algorithm we consider must accept if it cannot rule out the possibility of G being F-free. Thus, in order to prove that the query complexity is at least Q, say, (roughly speaking) the idea is to find a family $\mathcal{F}$ of r-graphs which are far from being F-free and such that any algorithm, given an r-graph chosen uniformly at random from $\mathcal{F}$ as an input, must perform at least Q queries in order to find a copy of F (with high probability). As we will prove, the family $\mathcal{F}^{(r)}_{n,d(n)}$ described below has the required properties.

Let F be an r-graph other than a weak forest. Recall that $X_F(G)$ denotes the number of copies of F in G. Let $\Phi_{F,n,d} := \min\{\mathbb{E}[X_K(G^{(r)}_{n,d})] : K \subseteq F, e_K > 0\}$. Taking K to be an edge shows that $\Phi_{F,n,d'} \leq nd'/r$ for any d'.

Assume now that $d(n) = \omega(1)$ and $d(n) = o(n^{r-1})$. Choose $\eta(n)$ such that $\eta(n) = o(1)$. Let

$$n_* := \max\{n_0 \leq n : \Phi_{F,n_0,d(n)} \geq (1-\eta(n))n_0 d(n)/r\}.$$

We claim that n_* always exists. Indeed, let $n_1 \leq n$ be such that there exists an r-graph G^* on n_1 vertices with average degree $d(n)$ and at least $(1-\eta(n)^2)\binom{n_1}{r}$ edges. Thus, $n_1 = (1 \pm o(1))((r-1)!d(n))^{1/(r-1)}$ and, since $d = \omega(1)$, we have $n_1 = \omega(1)$. Consider any G^* as above. Given any $K \subseteq F$, note that the number of copies of K in G^* is given by $(1 \pm \eta(n))\binom{n_1}{v_K}\frac{v_K!}{\mathrm{aut}(K)}$. (This can be seen by observing that G^* is "almost complete" and that every edge that is removed from a complete r-graph on n_1 vertices affects at most $n_1^{v_K-r}$ copies of K; since only $\eta(n)^2\binom{n_1}{r}$ edges are removed, this gives a total of at most $\eta(n)^2\binom{n_1}{r}n_1^{v_K-r} = o(\eta(n)\binom{n_1}{v_K})$ copies of K affected by the missing edges.) Among all $K \subseteq F$ with $e_K \geq 1$, this expression achieves its minimum (if n is sufficiently large) for a single edge. Hence $\Phi_{F,n_1,d(n)} \geq (1-\eta(n))n_1 d(n)/r$ and $n_* \geq n_1$ must exist[1].

Lemma 3.4.2. *Let F be a fixed r-graph other than a weak forest and let $d(n)$ be such that $d(n) = \omega(1)$ and $d(n) = o(n^{r-1})$. Then $d(n) = o(n_*^{r-1})$.*

Proof. For any fixed r-graph K with $e_K > 1$, let $d^*(n,K)$ be the smallest integer such that $\mathbb{E}[X_K(G^{(r)}_{n,d^*(n,K)})] \geq nd^*(n,K)/r$. Let $d^*_F(n) := \max_{K\subseteq F : e_K>1}\{d^*(n,K)\}$. We claim that $d^*_F(n) = o(n^{r-1})$. To prove the claim, note that, by corollary 3.3.1(ii), for any $K \subseteq F$ with $e_K > 1$ we have that

$$d^*(n,K) = \Theta\left(\left\lceil n^{\frac{(r-1)e_K - v_K + 1}{e_K - 1}}\right\rceil\right).$$

In particular, $d^*(n,K) = o(n^{r-1})$ as $v_K > r$. The claim follows by taking the maximum over all K.

[1]Note that here we are using the fact that there exist very dense $d(n)$-regular r-graphs. This follows from remark 3.1.1 by considering the complement.

Returning to the main proof, we now consider two cases. If $n_* = n$, then $d(n) = o(n_*^{r-1})$ by assumption. So suppose $n_* < n$. Let $n_+ > n_*$ be the smallest integer such that there exists a $d(n)$-regular r-graph on n_+ vertices. So $n_+ \leq 2n_*$ (since a $d(n)$-regular r-graph on $2n_*$ vertices can be constructed by duplicating one on n_* vertices) and $n_+ \leq n$ (because $d(n) = o(n^{r-1})$; see remark 3.1.1). By the definition of n_*, $\Phi_{F,n_+,d(n)} < (1-\eta(n))n_+d(n)/r$. In particular, there exists $K \subseteq F$ with $e_K \geq 2$ such that $\mathbb{E}[X_K(G^{(r)}_{n_+,d(n)})] < (1-\eta(n))n_+d(n)/r$. By the definition of $d^*(n,K)$ and corollary 3.3.1(ii), we then have that $d(n) < 2d^*(n_+,K)$. This in turn implies that $d(n) < 2d^*_F(n_+)$. But $d^*_F(n_+) = o(n_+^{r-1})$ by the above claim, and thus $d(n) = o(n_*^{r-1})$. □

Let $t := \lfloor n/n_* \rfloor$. Define $\mathcal{F}^{(r)}_{n,d(n)}$ by considering all possible partitions of V into sets $V_1, \dots, V_t$ of size

$$\tilde{n} := n/t \tag{3.29}$$

and, for each of them, all possible labelled $d(n)$-regular r-graphs G_i on each of the sets V_i. By lemma 3.4.2, $d(n) = o(\tilde{n}^{r-1})$ and so the G_i are well-defined (see remark 3.1.1). With these definitions, all the results in sections 3.2 and 3.3.1 can be applied to each family $\mathcal{F}^{(r)}_{n,d(n)}[V_i]$ consisting of the subgraphs of each $G \in \mathcal{F}^{(r)}_{n,d(n)}$ restricted to vertex set V_i, and hence to $\mathcal{F}^{(r)}_{n,d(n)}$ by summing over all $i \in [t]$.

Lemma 3.4.3. *Let F be a fixed, connected r-graph other than a weak tree and let $d(n)$ be such that $d(n) = \omega(1)$ and $d(n) = o(n^{r-1})$. Let $\tilde{n}$ and $\mathcal{F}^{(r)}_{n,d(n)}$ be as defined above. Then, an r-graph $G \in \mathcal{F}^{(r)}_{n,d(n)}$ chosen uniformly at random contains $\Theta(nd(n))$ edge-disjoint copies of F a.a.s.*

Note that this immediately implies that a.a.s. a graph $G \in \mathcal{F}^{(r)}_{n,d(n)}$ chosen uniformly at random is ε-far from being F-free for some fixed $\varepsilon > 0$.

Proof. Let $D_F(G)$ denote the maximum number of edge-disjoint copies of F in an r-graph G. Recall that $\mathcal{F}^{(r)}_{n,d(n)}$ is obtained by partitioning the set of vertices into sets $V_1, \dots, V_t$ of size $\tilde{n}$, where $t = n/\tilde{n}$, and considering $d(n)$-regular r-graphs G_i on each of the V_i, where each

G_i is chosen uniformly at random from $\mathcal{G}^{(r)}_{\tilde{n},d(n)}$, independently of each other. Note that $n_* \leq \tilde{n} \leq 2n_*$. Together with the definition of n_* and corollary 3.3.1(ii), this implies that the value of $\Phi_{F,\tilde{n},d(n)}$ in each G_i satisfies $\Phi_{F,\tilde{n},d(n)} = \Theta(\tilde{n}d(n))$. Then, by lemma 3.3.4, for any fixed $i \in [t]$, the maximum number of edge-disjoint copies of F in G_i is $D_F(G_i) = \Theta(\tilde{n}d(n))$ a.a.s.

We now claim that a graph $G \in \mathcal{F}^{(r)}_{n,d(n)}$ chosen uniformly at random a.a.s. satisfies that $D_F(G) = \Theta(nd(n))$. Observe that the bound $D_F(G) = \mathcal{O}(nd(n))$ is trivial, as G has exactly $nd(n)/r$ edges. For the lower bound, since $D_F(G_i) = \Theta(\tilde{n}d(n))$ a.a.s. for each $i \in [t]$, by the independence of the choice of G_i we have that a.a.s. at least half of the graphs G_i satisfy this equality. Therefore, $D_F(G) = \Omega(t\tilde{n}d(n)) = \Omega(nd(n))$. □

We now provide a proof for the lower bound on the complexity of any algorithm that tests F-freeness in r-graphs (for graphs and non-r-partite r-graphs F with $r \geq 3$). In order to do so, consider any algorithm ALG that performs Q queries given an input r-graph G on n vertices with average degree $d(n) \pm o(d(n))$. ALG will retrieve some information about G from the queries it performs, namely a set of r-sets $E_1 \subseteq E(G)$, a set of r-sets $E_2 \subseteq E(\overline{G})$ and (potentially) some vertex degrees of G, i.e., a set $\mathcal{D} \subseteq \{(v, d_v) : v \in V(G), d_v = \deg_G(v)\}$. We call the information retrieved by ALG after Q queries the *history* of G seen by ALG, and denote it as $(E_1, E_2, \mathcal{D})$. We say that the history of G seen by ALG is *simple* if E_1 forms a weak forest and for all $(v, d_v) \in \mathcal{D}$ we have that $d_v = \mathcal{O}(d(n))$.

We will allow our algorithm to find weak forests in the input graphs. Thus we assume that F is not a weak forest, that is, F contains at least two edges whose intersection has size at least 2 or a loose cycle. In order to prove our bound we first show the following result.

Lemma 3.4.4. *Let F be an r-graph which is not a weak forest and define $\tilde{n}$ as in (3.29). Assume that $d(n) = \omega(1)$ and $d(n) = o(n^{r-1})$. Suppose ALG is an algorithm whose input is an r-graph $G \in \mathcal{F}^{(r)}_{n,d(n)}$ and which for at least $1/3$ of the r-graphs $G \in \mathcal{F}^{(r)}_{n,d(n)}$ sees with probability at least $1/3$ a history which is not simple. Then, ALG must perform $\Omega(\min\{d(n), \tilde{n}^{r-1}/d(n), \tilde{n}^{1/2}\})$ queries.*

To prove lemma 3.4.4, we will show that an algorithm that performs only Q queries, $Q = o(\min\{d(n), \tilde{n}^{r-1}/d(n), \tilde{n}^{1/2}\})$, will usually not succeed with the desired probability. For this, we consider a suitable randomised process P that answers the queries of the algorithm.

Proof. Suppose $Q = o(\min\{d(n), \tilde{n}^{r-1}/d(n), \tilde{n}^{1/2}\})$. Let ALG be a (possibly adaptive and randomised) algorithm that performs Q queries and searches for some history of the input $G \in \mathcal{F}^{(r)}_{n,d(n)}$ which is not simple. Since we have, for any history $(E_1, E_2, \mathcal{D})$ seen by any algorithm on any $G \in \mathcal{F}^{(r)}_{n,d(n)}$, that any pair $(v, d_v) \in \mathcal{D}$ satisfies $d_v = d(n)$, the only condition for $(E_1, E_2, \mathcal{D})$ being simple is that E_1 forms a weak forest. Therefore, ALG tries to find a set $E_1 \subseteq E(G)$ which forms an r-graph which is not a weak forest.

The queries performed by ALG are answered by a randomised process P. We denote the queries asked by ALG as $q_1, q_2, \ldots$, and the answers given by P as $a_1, a_2, \ldots$. After t queries, we refer to all the previous queries from ALG and all the answers provided by P as the *query-answer history*. The process P uses the query-answer history to build what we call the *history book*, defined for each $t \geq 0$ and denoted by $H^t = (V^t, E^t_*, \bar{E}^t)$, where $V^t \subseteq V$, $\bar{E}^t \subseteq \binom{V}{r}$ and E^t_* is a set of labelled r-sets in $\binom{V}{r}$ such that each r-set $e \in E^t_*$ has r labels $i_1, \ldots, i_r$, one for each vertex in e. We denote by E^t the set of edges consisting of the r-sets in E^t_*. Given an edge $e = \{v_1, \ldots, v_r\} \in E^t$, its labels in E^t_* indicate, for each vertex $v_j \in e$, that e is the i_j-th edge in the incidence list of v_j.

Initially, V^0, E^0_* and $\bar{E}^0$ are set to be empty. Note that we may always assume that in the t-th step ALG never asks a query whose answer can be deduced from the history book H^{t-1}. Given two r-graphs H and H', define $\mathcal{F}^{(r)}_{n,d(n),H,H'} := \{G \in \mathcal{F}^{(r)}_{n,d(n)} : H \subseteq G, H' \subseteq \overline{G}\}$. We abuse notation to write $\mathcal{F}^{(r)}_{n,d(n),H,H'}$ as the event that $G \in \mathcal{F}^{(r)}_{n,d(n),H,H'}$. The process P answers ALG's queries and builds the history book as follows.

If $q_t = \{v_1, \ldots, v_r\}$ is a vertex-set query, then P answers "yes" with probability $\mathbb{P}[q_t \in G \mid \mathcal{F}^{(r)}_{n,d(n),E^{t-1},\bar{E}^{t-1}}]$, and "no" otherwise. If the answer is "yes", then the history book is updated by setting $V^t := V^{t-1} \cup q_t$, $\bar{E}^t := \bar{E}^{t-1}$ and adding q_t together with its labels $j_1, \ldots, j_r$ to E^{t-1}_* to obtain E^t_*, where the labels $j_1, \ldots, j_r$ are chosen uniformly at random among all possible labellings which are consistent with the labels in E^{t-1}_*. In this case, the labels are also given to ALG as part of the answer. Otherwise, the history book is updated by setting $V^t := V^{t-1} \cup q_t$, $E^t_* := E^{t-1}_*$ and $\bar{E}^t := \bar{E}^{t-1} \cup \{q_t\}$.

If $q_t = (u, i)$ is a neighbour query, P replies with $a_t := (v_1, \dots, v_{r-1}, j_1, \dots, j_{r-1})$, where a_t is chosen such that $e := \{u, v_1, \dots, v_{r-1}\}$ is an edge and for each $k \in [r-1]$, the number j_k is the position of e in the incidence list of v_k (we may assume that, as the r-graphs are $d(n)$-regular, the algorithm never queries $i > d(n)$). To determine its answer a_t, the process P will first choose an r-graph $G_t \in \mathcal{F}^{(r)}_{n,d(n),E^{t-1},\bar{E}^{t-1}}$ uniformly at random, and then choose a labelling of the edges of G_t which is consistent with H^{t-1} uniformly at random. The edge $e = \{u, v_1, \dots, v_{r-1}\}$ will be the i-th edge at u in G_t (in the chosen labelling) and j_s will be the label of e in the incidence list of v_s (for each $s \in [r-1]$). Note that the random labelling ensures that, given G_t, e is chosen uniformly at random from a set of edges of size at least $d(n) - t$ (namely from the set of those edges of G_t incident to u which have no label at u in H^{t-1}). This in turn means that for all $f \in G_t$ with $u \in f$, the probability that the label of u in f is i is at most $1/(d(n) - t)$. The history book is updated by setting $V^t := V^{t-1} \cup e$, $\bar{E}^t := \bar{E}^{t-1}$ and adding e together with the labels $i, j_1, \dots, j_{r-1}$ to E^{t-1}_* to obtain E^t_*.

Once P has answered all Q queries, it chooses an r-graph $G^* \in \mathcal{F}^{(r)}_{n,d(n),E^Q,\bar{E}^Q}$ uniformly at random. Note that P gives extra information to the algorithm in the form of labels that have not been queried. This extra information can only benefit the algorithm, so any lower bound on the query complexity in this setting will also be a lower bound in the general setting.

We claim that G^* is chosen uniformly at random in $\mathcal{F}^{(r)}_{n,d(n)}$. Indeed, let $s_0 := |\mathcal{F}^{(r)}_{n,d(n)}|$. Given a query-answer history $\mathcal{H} = (q_1, a_1, \dots, q_Q, a_Q)$, for each $t \in [Q] \cup \{0\}$, write $\mathcal{F}^t(\mathcal{H})$ for the set of all those graphs $G \in \mathcal{F}^{(r)}_{n,d(n)}$ which are "consistent" with $\mathcal{H}$ for at least the first t steps, i.e., all $G \in \mathcal{F}^{(r)}_{n,d(n),E^t,\bar{E}^t}$, where $(V^t, E^t_*, \bar{E}^t)$ is the history book associated with the first t steps of $\mathcal{H}$. Thus $\mathcal{F}^t$ is a random variable and $\mathcal{F}^0(\mathcal{H}) = \mathcal{F}^{(r)}_{n,d(n)}$ for each $\mathcal{H}$. Now consider any sequence $\mathcal{S} = (s_1, \dots, s_Q)$ such that $s_t \in \mathbb{N}$ and $\mathbb{P}[|\mathcal{F}^t| = s_t] > 0$ for all $t \in [Q]$. Write $\mathbb{P}_{\mathcal{S}}$ for the probability space consisting of all those query-answer histories $\mathcal{H} = (q_1, a_1, \dots, q_Q, a_Q)$ which satisfy $|\mathcal{F}^t| = s_t$ for all $t \in [Q]$. Take any fixed r-graph $G \in \mathcal{F}^{(r)}_{n,d(n)}$. Note that our choice of the t-th answer a_t given by P implies that $\mathbb{P}_{\mathcal{S}}\left[G \in \mathcal{F}^t \mid G \in \mathcal{F}^{t-1}\right] = s_t/s_{t-1}$ for all

$t \in [Q]$. Thus,

$$\mathbb{P}_{\mathcal{S}}[G^* = G] = \mathbb{P}_{\mathcal{S}}\left[G \in \mathcal{F}^Q\right]/s_Q = \frac{1}{s_Q}\left(\prod_{t=1}^{Q} \mathbb{P}_{\mathcal{S}}\left[G \in \mathcal{F}^t \mid G \in \mathcal{F}^{t-1}\right]\right)\mathbb{P}_{\mathcal{S}}\left[G \in \mathcal{F}^0\right]$$
$$= \frac{1}{s_Q}\prod_{t=1}^{Q}\frac{s_t}{s_{t-1}} = \frac{1}{s_0}.$$

Thus $\mathbb{P}[G^* = G] = 1/|\mathcal{F}^{(r)}_{n,d(n)}|$ by the law of total probability.

Now let us prove that ALG will a.a.s. only see a simple history $(E_1, E_2, \mathcal{D})$. Note that $E_1 = E^Q$ and $E_2 = \bar{E}^Q$. Hence it suffices to show that E^Q is a weak forest a.a.s. Recall that we can write each $G \in \mathcal{F}^{(r)}_{n,d(n)}$ as the disjoint union of $G_1, \dots, G_s$, where $s = n/\tilde{n}$, each G_j is uniformly distributed in $\mathcal{G}^{(r)}_{\tilde{n},d(n)}$ and G_j has vertex set V_j.

Assume q_t is a vertex-set query. The probability that P answers "yes" is given by corollary 3.2.3 as $\mathcal{O}(d(n)/\tilde{n}^{r-1})$, as long as $t = o(d(n))$. Thus, because the number of queries is $Q = o(\tilde{n}^{r-1}/d(n))$, then, by a union bound, the probability that any edge is found with vertex-set queries is $o(1)$.

Assume now that $q_t = (u, i)$ is a neighbour query, where $u \in V_j$, $j \in [s]$. We will bound the probability that some vertex returned by P in the t-th answer a_t lies in V^{t-1}. Note that such a vertex will always lie in V_j. To bound this probability, for any given vertex $v \in V^{t-1} \cap V_j$, let $S_v := \{f \in \binom{V_j}{r} : u, v \in f, f \notin E^{t-1} \cup \bar{E}^{t-1}\}$. Note that $|S_v| = \mathcal{O}(\tilde{n}^{r-2})$. By corollary 3.2.3, the probability that a given r-set in S_v is an edge of G_t is $\Theta(d(n)/\tilde{n}^{r-1})$. Furthermore, if we condition on $e \in S_v$ being an edge of G_t, recall that the probability that its label belonging to u equals i is at most $1/(d(n) - t)$. Thus, by a union bound over all elements of S_v, the probability that some r-set in S_v is the i-th edge in the incidence list of u is $\mathcal{O}(d(n)/(\tilde{n}(d(n) - t))) = \mathcal{O}(1/\tilde{n})$. Note that $|V^{t-1}| \leq rt = \mathcal{O}(Q)$. Thus, by a union bound over all $v \in V^{t-1}$, the probability that the answer to the t-th query results in E^t being not a weak forest is $\mathcal{O}(Q/\tilde{n})$. By a union bound over all queries, the probability that any of the at most Q neighbour queries finds any vertex in the current history book is $\mathcal{O}\left(Q^2/\tilde{n}\right) = o(1)$. This in turn implies that the probability that a neighbour query detects anything other than a weak forest is $o(1)$.

Combining the conditions and lower bounds for both types of queries, we have that the probability (taken over all queries of ALG and choices of P in the above process) that E^Q is not a weak forest is $o(1)$. The statement follows since we have shown that the r-graph G^* returned by P is chosen uniformly at random from $\mathcal{F}^{(r)}_{n,d(n)}$. □

Theorem 3.4.5. *The following statements hold:*

(i) *Let F be a connected graph which is not a tree. Assume that $d(n) = \omega(1)$ and $d(n) = o(n)$. Assume, furthermore, that $nd(n)/2 \leq ex(n, F)$. Then, any one-sided error F-freeness tester must perform $\Omega(\min\{d(n), \tilde{n}/d(n), \tilde{n}^{1/2}\})$ queries when restricted to n-vertex inputs of average degree $d(n) - o(d(n))$, where $\tilde{n}$ is as defined in (3.29).*

(ii) *Let $r \geq 3$. Let F be a connected non-r-partite r-graph. Assume that $d(n) = \omega(1)$ and $d(n) = o(n^{r-1})$. Then, any one-sided error F-freeness tester in r-graphs must perform $\Omega(\min\{d(n), \tilde{n}^{r-1}/d(n), \tilde{n}^{1/2}\})$ queries when restricted to n-vertex inputs of average degree $d(n) \pm o(d(n))$, where $\tilde{n}$ is as defined in (3.29).*

Proof. We first prove (ii) and later discuss which modifications are needed to prove (i). Let $Q = o(\min\{d(n), \tilde{n}^{r-1}/d(n), \tilde{n}^{1/2}\})$. Consider any algorithm ALG that performs Q queries given an input r-graph G on n vertices with average degree $d(n) \pm o(d(n))$. Assume that ALG is given an r-graph $G \in \mathcal{F}^{(r)}_{n,d(n)}$ as an input. By lemma 3.4.4 we know that any algorithm that performs at most Q queries will see a simple history $(E_1, E_2, \mathcal{D})$ of G with probability at least $2/3$ for at least $2/3$ of the graphs $G \in \mathcal{F}^{(r)}_{n,d(n)}$. Note that any such simple history $(E_1, E_2, \mathcal{D})$ is such that $|E_1 \cup E_2| \leq Q$, $|\mathcal{D}| \leq Q$ and for every $(v, d_v) \in \mathcal{D}$, $d_v = d(n)$. We now show that there is a family $\mathcal{F}_2$ of F-free r-graphs that, for every such simple history, contains at least one r-graph for which ALG will see the same history with positive probability.

- For each simple history $(E_1, E_2, \mathcal{D})$, let H be the r-graph that has vertex set $\bigcup_{e \in E_1 \cup E_2} e$ and edge set E_1. Note that H is a weak forest with (possibly) some isolated vertices and $|V(H)| \leq rQ$. Consider a partition of $V(H)$ into $V_1, \dots, V_r$ such that for every $e \in E(H)$ and $i \in [r]$, we have $|e \cap V_i| = 1$, which can be constructed inductively by adding the edges of H one by one and distributing their vertices into different parts. Consider pairwise disjoint sets of vertices $W_1, \dots, W_r$ of size $d(n)^{1/(r-1)}$ which are disjoint from $V(H)$.

- Define an r-graph K with vertex set $V(H) \cup W_1 \cup \dots \cup W_r$. Note that for each $v \in V_i$ there are $d(n)$ r-sets f such that $v \in f$ and $|f \cap W_j| = 1$ for all $j \in [r] \setminus \{i\}$. Define $E(K)$ by including $E(H)$ and adding $d(n) - \deg_H(v)$ of these r-sets incident to each vertex $v \in V(H)$. Note that K is r-partite and, thus, F-free.

- Finally, for any such K, consider the r-graph G obtained as the vertex-disjoint union of K and any F-free r-graph on $n - |V(K)|$ vertices with average degree $d(n) - o(d(n))$ (to see that this is possible, note that $|V(K)| = o(n)$ and $ex(n, F) = \Theta(n^r)$ since F is non-r-partite).

We define $\mathcal{F}_2$ as the family that consists of all r-graphs G that can be constructed as above and all possible relabellings of their vertices. Note that each $G \in \mathcal{F}_2$ has n vertices and average degree $d(n) \pm o(d(n))$ and is F-free. Moreover, for every $G \in \mathcal{F}^{(r)}_{n,d(n)}$ and any simple history $(E_1, E_2, \mathcal{D})$ seen by ALG on G, there is some r-graph $G \in \mathcal{F}_2$ such that ALG would have seen $(E_1, E_2, \mathcal{D})$ on G.

Now suppose ALG is a one-sided error F-freeness tester for r-graphs of average degree $d(n) \pm o(d(n))$ that performs Q queries. Assume that ALG is given inputs as follows. With probability 99/100, the input is an r-graph $G \in \mathcal{F}^{(r)}_{n,d}$ chosen uniformly at random. With probability 1/100, the input is an r-graph $G \in \mathcal{F}_2$ chosen uniformly at random. By lemma 3.4.4, the proportion of r-graphs $G \in \mathcal{F}^{(r)}_{n,d(n)}$ for which with probability at least 2/3 ALG only sees a simple history is at least 2/3. Moreover, since ALG is a one-sided error tester, it can only reject an input G if ALG can guarantee the existence of a copy of F in G. Thus, if after Q queries ALG has seen a simple history $(E_1, E_2, \mathcal{D})$, then it cannot reject the input, as there are r-graphs $G \in \mathcal{F}_2$ which are F-free and for which ALG may see the same history with positive probability. So given a random input as described above, the probability that ALG accepts is at least $(99/100)(2/3)^2 > 2/5$.

On the other hand, by lemma 3.4.3, the proportion of r-graphs in $\mathcal{F}_{n,d(n)}^{(r)}$ that are ε-far from being F-free is at least 99/100. Since ALG is a one-sided error F-freeness tester, it must reject these inputs with probability at least 2/3. Therefore, given a random input G, the probability that ALG rejects G must be at least $(99/100)^2(2/3) > 3/5$. This is a contradiction to the previous statement, so ALG cannot be a one-sided error F-freeness tester.

In order to prove (i), let $Q = o(\min\{d(n), \tilde{n}/d(n), \tilde{n}^{1/2}\})$. If F is not bipartite, then (ii) already shows the desired statement. In order to deal with bipartite graphs F, define a new family $\mathcal{F}_1$ (which also works for non-bipartite F) as follows. Given a simple history $(E_1, E_2, \mathcal{D})$, define H as above. For each $v \in V(H)$, consider $d(n) - \deg_H(v)$ new vertices and add an edge between v and each of them. Denote the resulting graph by K. Finally, consider the graph G obtained as the disjoint union of K and any F-free graph on $n - |V(K)|$ vertices with average degree $d(n) - o(d(n))$. We define $\mathcal{F}_1$ as the family that consists of all graphs G that can be constructed as above and all possible relabellings of their vertices. The remainder of the proof works in the same way. □

Note that if, for instance, $d(n) = 2ex(n, F)/n$ and $F = C_4$, then theorem 3.4.5(i) (together with corollary 3.3.1(ii)) implies a lower bound of $\Omega(n^{1/2})$. The bound on the number of queries in theorem 3.4.5 is stronger than in proposition 3.4.1 as long as d is not too small.

3.4.3. Upper bounds

Here, we present several upper bounds on the query complexity for testing F-freeness. All the testers we present here are one-sided error testers. Note that there is always the trivial bound of $\mathcal{O}(nd)$ queries; the forthcoming results are relevant only whenever the presented bound is smaller than this. Proposition 3.4.6 provides a bound on the query complexity which applies to input r-graphs G in which the maximum degree does not differ too much from the average degree. Proposition 3.4.7 improves proposition 3.4.6 for special r-graphs F. Finally, theorem 3.4.8 provides a bound which works for arbitrary F and G. Propositions 3.4.6 and 3.4.7 give stronger bounds for very sparse r-graphs G, whereas theorem 3.4.8 gives stronger bounds for denser r-graphs.

We will say that a tester for a property $\mathcal{P}$ is an ε'-tester if it is a valid tester for $\mathcal{P}$ for all distance parameters $\varepsilon \geq \varepsilon'$ (recall that ε stands for the proportion of edges of a graph G that needs to be modified to satisfy a given property $\mathcal{P}$ in order for G to be considered far from $\mathcal{P}$). The techniques of our algorithms are based on two strategies: random sampling and local exploration. We will always write V for the vertex set of the input r-graph G and d for its average degree. Given any $S \subseteq V$, we denote by $G[S] := \{e \in G : e \subseteq S\}$ the subgraph of G spanned by S. Thus $V(G[S]) = S$. We denote by $G\{S, \rho\} := \{e \in G : \exists\, v \in e : \mathrm{dist}(S, v) < \rho\}$ the graph obtained from G by performing a breadth-first search of depth ρ from S. Throughout this section, the hidden constants in the $\mathcal{O}$ notation will be independent of both ε and n. When the constants depend on ε, we will denote this by writing $\mathcal{O}_\varepsilon$.

Proposition 3.4.6. *For every $\varepsilon > 0$, the following holds. Let F be a fixed, connected r-graph and let D be its diameter. For the class consisting of all input r-graphs G on n vertices with average degree d and maximum degree $\Delta(G) = \mathcal{O}(d)$, there exists an ε-tester for F-freeness with $\mathcal{O}_\varepsilon(d^{D+1})$ queries.*

Proof. We consider a one-sided error F-freeness ε-tester. The procedure is as follows. First choose a set $S \subseteq V(G)$ of size $\Theta(1/\varepsilon)$ uniformly at random. For each $v \in S$, find $G\{v, D+1\}$ by performing neighbour queries. If any of the graphs $G\{v, D+1\}$ contains a copy of F, the algorithm rejects G. Otherwise, it accepts it. Clearly, the complexity is $\mathcal{O}(d^{D+1}/\varepsilon)$ and the procedure will always accept G if it is F-free.

Assume now that the input is ε-far from being F-free. Then, it contains at least $\varepsilon nd/r$ edges that belong to copies of F. It follows that the number of vertices that belong to some copy of F is $\Omega(\varepsilon nd/\Delta(G)) = \Omega(\varepsilon n)$. Therefore, if the implicit constant in the bound on $|S|$ is large enough, the algorithm will choose one of the vertices that belong to a copy of F with probability at least $2/3$. If it chooses such a vertex, then, as F has diameter D, it rejects the input. □

We can improve the bound in proposition 3.4.6 for a certain class of r-graphs F. Given any r-graph F, let D_F be its diameter. Consider the partition of its vertices given by choosing an edge $e \in F$, taking $V_0(e) \coloneqq e$ and $V_i(e) \coloneqq \{u \in V(F) : \mathrm{dist}(e,u) = i\}$ for $i \in [D_F]$. We let $\mathcal{F}_E \coloneqq \{F : |F[V_{D_F}(e)]| = 0$ for all $e \in F\}$. The class $\mathcal{F}_E$ contains, for instance, complete r-partite r-graphs, loose cycles and tight cycles. If $r = 2$, then $\mathcal{F}_E$ also contains hypercubes, for example.

Proposition 3.4.7. *For every $\varepsilon > 0$, the following holds. Let $F \in \mathcal{F}_E$ be an r-graph and let D be its diameter. For the class consisting of all input r-graphs G with average degree d and maximum degree $\Delta(G) = \mathcal{O}(d)$, there exists an ε-tester for F-freeness with $\mathcal{O}_\varepsilon(d^D)$ queries.*

Proof. We consider a one-sided error ε-tester, which works in a very similar way as in the proof of proposition 3.4.6. The F-freeness tester chooses a set $S \subseteq V$ of size $\Theta(1/\varepsilon)$ uniformly at random. It then chooses an edge e incident to each $v \in S$ uniformly at random and finds $G\{e, D\}$ by performing neighbour queries; then, it searches for a copy of F. If any copy of F is found, the algorithm rejects the input; otherwise, it accepts. The query complexity is clearly $\mathcal{O}(d^D/\varepsilon)$. The analysis of the algorithm is similar to that of proposition 3.4.6, so we omit the details. □

We conclude with the following bound, which works for arbitrary G and any F without isolated vertices. Given an r-graph F, we define its *vertex-overlap index* $\ell(F)$ as the minimum integer ℓ such that two graphs isomorphic to F sharing ℓ vertices must share at least one edge; if this does not hold for any $\ell \in [v_F]$, we then set $\ell = v_F + 1$. For instance, $\ell(K_k^{(r)}) = r$, and for a matching M we have $\ell(M) = |V(M)| + 1$ if $|V(M)| \geq 2r$.

Theorem 3.4.8. *For every $\varepsilon > 0$, the following holds. Let $r \geq 2$ and let F be an r-graph without isolated vertices. Let $\ell \coloneqq \ell(F)$. For the class consisting of all input r-graphs G on n vertices with average degree d and maximum degree Δ, there exists an ε-tester for F-freeness with $\mathcal{O}_\varepsilon(\max\{(n/(nd)^{1/v_F})^r, (n^{\ell-2}\Delta/d)^{r/(\ell-1)}\})$ queries.*

In the case when $F = K_k^{(r)}$ and the input r-graph G satisfies $\Delta(G) = \mathcal{O}(d)$, the bound in theorem 3.4.8 becomes $\mathcal{O}_\varepsilon((n/(nd)^{1/k})^r)$ whenever $d = o(n^{k/(r-1)-1})$, and $\mathcal{O}_\varepsilon(n^{r(r-2)/(r-1)})$ otherwise.

Proof. Choose a constant c which is large enough compared to v_F and e_F. We present a one-sided error ε-tester in algorithm 1. In this proof, the constants in the $\mathcal{O}$ notation are independent of c.

Algorithm 1. An F-freeness ε-tester for r-graphs.

```
1: procedure CANONICAL F TESTER
2:     Let s = c max{n/(εnd)^{1/v_F}, (n^{ℓ−2}Δ/εd)^{1/(ℓ−1)}}.
3:     Choose a set S ⊆ V of size s uniformly at random.
4:     Find G[S] by performing all vertex-set queries.
5:     if G[S] contains a copy of F, then reject.
6:     otherwise, accept.
7: end procedure
```

It is easy to see that we may assume s is large compared to v_F. If G is F-free, the algorithm will never find a copy of F and will always accept the input. Assume now that G is ε-far from being F-free. Then, G must contain a set $\mathcal{F}$ of $\varepsilon nd/e_F$ edge-disjoint copies of F. For each $W \subseteq V$, we define $\deg_{\mathcal{F}}(W) := |\{F' \in \mathcal{F} : W \subseteq V(F')\}|$. It is clear that

$$\deg_{\mathcal{F}}(W) \leq \min_{v \in W} \deg(v) \leq \Delta. \tag{3.30}$$

For any fixed $F' \in \mathcal{F}$, we have $\mathbb{P}[F' \in G[S]] = (1 \pm 1/2)(s/n)^{v_F}$. We denote by X the number of $F' \in \mathcal{F}$ such that $F' \in G[S]$. We conclude that

$$\mathbb{E}[X] = (1 \pm 1/2)|\mathcal{F}|\left(\frac{s}{n}\right)^{v_F} = \Theta\left(\frac{\varepsilon d s^{v_F}}{n^{v_F - 1}}\right). \tag{3.31}$$

The variance of X can be estimated by observing that we only need to consider r-graphs $F', F'' \in \mathcal{F}$ whose vertex sets intersect, as otherwise the events are negatively correlated. Hence,

$$\mathrm{Var}[X] \leq \sum_{\substack{(F',F'') \in \mathcal{F}\times\mathcal{F} \\ V(F') \cap V(F'') \neq \emptyset}} \mathbb{P}[F' \cup F'' \subseteq G[S]] = \sum_{i=1}^{v_F} \sum_{\substack{(F',F'') \in \mathcal{F}\times\mathcal{F} \\ |V(F') \cap V(F'')| = i}} \mathbb{P}[F' \cup F'' \subseteq G[S]]. \tag{3.32}$$

Let us estimate this quantity for each $i \in [v_F]$. For $i \in [v_F - 1]$ we can apply a double counting argument to see that

$$|\{(F', F'') \in \mathcal{F} \times \mathcal{F} : |V(F') \cap V(F'')| = i\}| \leq 2 \sum_{W \in \binom{V}{i}} \binom{\deg_{\mathcal{F}}(W)}{2}, \tag{3.33}$$

while for $i = v_F$ we have that

$$|\{(F', F'') \in \mathcal{F} \times \mathcal{F} : |V(F') \cap V(F'')| = v_F\}| \leq |\mathcal{F}| + 2 \sum_{W \in \binom{V}{v_F}} \binom{\deg_{\mathcal{F}}(W)}{2}. \tag{3.34}$$

Note that

$$\sum_{W \in \binom{V}{i}} \deg_{\mathcal{F}}(W) = \mathcal{O}(|\mathcal{F}|) = \mathcal{O}(\varepsilon n d). \tag{3.35}$$

By assumption on F, we have that $\deg_{\mathcal{F}}(W) \leq 1$ for all W such that $|W| \geq \ell$, which implies that $\binom{\deg_{\mathcal{F}}(W)}{2} = 0$. Moreover, by (3.30) and (3.35), for each $i \in [\ell - 1]$ we obtain

$$\sum_{W \in \binom{V}{i}} \binom{\deg_{\mathcal{F}}(W)}{2} \leq \Delta \sum_{W \in \binom{V}{i}} \deg_{\mathcal{F}}(W) = \mathcal{O}(\varepsilon n d \Delta). \tag{3.36}$$

Combining (3.33)–(3.36), the estimation in (3.32) yields

(3.37)

$$\mathrm{Var}[X] = \mathcal{O}\left(\varepsilon n d \left(\frac{s}{n}\right)^{v_F}\right) + \sum_{i=1}^{\ell-1} \mathcal{O}\left(\varepsilon n d \Delta \left(\frac{s}{n}\right)^{2v_F - i}\right) = \varepsilon n d \cdot \mathcal{O}\left(\left(\frac{s}{n}\right)^{v_F} + \Delta \left(\frac{s}{n}\right)^{2v_F - \ell + 1}\right).$$

By Chebyshev's inequality, $\mathbb{P}[X = 0] \leq \mathrm{Var}[X]/\mathbb{E}[X]^2$. Using (3.31), (3.37) and the fact that c is large compared to v_F and e_F, one can check that $\mathrm{Var}[X]/\mathbb{E}[X]^2 < 1/3$. Thus $G[S]$ contains a copy of F with probability at least $2/3$. Therefore, G will be rejected with probability at least $2/3$, which shows that algorithm 1 is an F-freeness ε-tester.

The query complexity of the algorithm is given by performing all $\binom{s}{r}$ vertex-set queries. This yields the stated complexity. □

APPENDIX A

APPENDICES

A.1. Proofs of properties of random subgraphs of the hypercube

Proof. Proof of lemma 2.5.6 Let $p := 1/2 + \varepsilon$. Fix any $v \in \{0,1\}^n$. Throughout this proof, we write $d(v)$ to refer to the degree of v in $\mathcal{Q}_p^n$.

Note that $d(v)$ follows a binomial distribution with parameters n and p. Since $\delta < 1/2$, it follows that

$$\mathbb{P}[d(v) \leq \delta n] \leq \delta n \binom{n}{\delta n} p^{\delta n}(1-p)^{(1-\delta)n}.$$

Using the Stirling formula, we conclude that

$$\mathbb{P}[d(v) \leq \delta n] \leq (1 + \mathcal{O}(n^{-1}))\sqrt{\frac{\delta n}{2\pi(1-\delta)}}\left(\left(\frac{p}{\delta}\right)^{\delta}\left(\frac{1-p}{1-\delta}\right)^{1-\delta}\right)^n.$$

By the union bound, it now suffices to show that

$$\left(\frac{p}{\delta}\right)^{\delta}\left(\frac{1-p}{1-\delta}\right)^{1-\delta} = \left(\frac{1+2\varepsilon}{2\delta}\right)^{\delta}\left(\frac{1-2\varepsilon}{2(1-\delta)}\right)^{1-\delta} < \frac{1}{2},$$

but this follows since $\delta \ll \varepsilon$. □

Proof of lemma 2.5.7. First, note that we may assume that $b < a$ (otherwise, choose a value $b' \in (2-2a, a)$ and prove the statement for this value, which in turn implies the result for b). Throughout this proof, we write $d(v)$ for $d_{\mathcal{Q}^n_\varepsilon}(v)$. Fix any vertex $v \in \{0,1\}^n$. Let $X(v) := |\{u \in B^\ell(v) : d(u) \neq \varepsilon n \pm n^a\}|$.

If $X(v) \geq n^b$, there exists a set $A \subseteq B^\ell(v)$ of size $|A| = n^b$ such that $d(u) \neq \varepsilon n \pm n^a$ for all $u \in A$. We call such a set A *bad*. We now bound the probability that such a bad set exists. Given any set $A \in \binom{B^\ell(v)}{n^b}$, for each $u \in A$ let $d^A(u) := |N_{\mathcal{Q}^n_\varepsilon}(u) \setminus A|$. Observe that $d^A(u) = d(u) \pm n^b$ and, since $b < a$, for any $u \in A$ we have that, if $d(u) \neq \varepsilon n \pm n^a$, then $d^A(u) \neq \varepsilon n \pm n^a/2$.

Fix a set $A \in \binom{B^\ell(v)}{n^b}$. Observe that $\mathbb{E}[d^A(u)] \in [\varepsilon n(1-n^{b-1}), \varepsilon n]$ for all $u \in A$. Furthermore, the variables $\{d^A(u) : u \in A\}$ are mutually independent, and each of them follows a binomial distribution. By lemma 2.4.1, for each $u \in A$ we have that

$$\mathbb{P}[d(u) \neq \varepsilon n \pm n^a] \leq \mathbb{P}[d^A(u) \neq \varepsilon n \pm n^a/2] \leq 2e^{-n^{2a-1}/(40\varepsilon)}.$$

Since the variables $d^A(u)$ are mutually independent, it follows that

$$\mathbb{P}[A \text{ is bad}] \leq \left(2e^{-n^{2a-1}/(40\varepsilon)}\right)^{n^b}.$$

Now consider a union bound over all possible choices of A and all choices of v. It suffices to prove that

$$\binom{\ell n^\ell}{n^b}\left(2e^{-n^{2a-1}/(40\varepsilon)}\right)^{n^b} < 2.1^{-n}.$$

Since $\binom{\ell n^\ell}{n^b} \leq (e\ell n^{\ell-b})^{n^b}$, it suffices to show that

$$n^b(1 + \ln 2 + \ln \ell + (\ell - b)\ln n - n^{2a-1}/(40\varepsilon)) < -n \ln 2.1.$$

But this follows for n sufficiently large, from the fact that $b > 2 - 2a$. □

For the proof of lemma 2.5.8, it will be important to show that the number of vertices whose degree deviates from the expected degree is small.

Lemma A.1.1. *Let $\varepsilon \in (0,1)$ and $a \in (1/2,1)$. Let X be the number of vertices $v \in \{0,1\}^n$ for which $d_{\mathcal{Q}^n_\varepsilon}(v) \neq \varepsilon n \pm n^a$. Then,*

$$\mathbb{P}[X \geq e^{-n^{2a-1}/(6\varepsilon)}2^n] \leq 2e^{-n^{2a-1}/(6\varepsilon)}.$$

Proof. Throughout this proof, we use $d(v)$ for $d_{\mathcal{Q}^n_\varepsilon}(v)$. Note that $d(v)$ follows a binomial distribution with parameters n and ε, so $\mathbb{E}[d(v)] = \varepsilon n$ and, by lemma 2.4.1,

$$\mathbb{P}[d(v) \neq \varepsilon n \pm n^a] \leq 2e^{-n^{2a-1}/(3\varepsilon)}.$$

We then have that $\mathbb{E}[X] \leq 2^{n+1}e^{-n^{2a-1}/(3\varepsilon)}$, and the statement follows by Markov's inequality. □

Remark A.1.2. *In particular, for any $a \in (1/2,1)$ we have that a.a.s. the number of vertices whose degree deviates from the expectation by more than n^a is at most $e^{-n^{2a-1}/(6\varepsilon)}2^n$.*

Proof of lemma 2.5.8. Throughout this proof, we will write $d(v)$ for $d_{\mathcal{Q}^n_\varepsilon}(v)$, $N(v)$ for $N_{\mathcal{Q}^n_\varepsilon}(v)$ and $d^\ell(v)$ for $d^\ell_{\mathcal{Q}^n_\varepsilon}(v)$.

Fix a vertex $v \in \{0,1\}^n$, reveal all the edges incident to v, and condition on the event that $d(v) = \varepsilon n \pm n^a$. We then have that

$$\mathbb{E}[d^\ell(v)] = \binom{\varepsilon n \pm n^a}{\ell}\varepsilon^{(2^{\ell-1}-1)\ell} = (1 \pm \mathcal{O}(n^{a-1}))\frac{n^\ell}{\ell!}\varepsilon^{2^{\ell-1}\ell}. \tag{A.1}$$

Let $\mathcal{D}(v) \subseteq \mathcal{D}(\mathcal{Q}^n)$ be the set of directions such that $N(v) = v + \mathcal{D}(v)$. Consider the graph $\Gamma_\ell(v) := \mathcal{Q}^n[v + \ell(\mathcal{D}(v) \cup \{\mathbf{0}\})]$. For each $i \in [\ell]$, let $L_i \subseteq E(\Gamma_\ell(v))$ be the set of edges which are at distance $i-1$ from v in $\mathcal{Q}^n$. Note that these sets partition $E(\Gamma_\ell(v))$ and that $|L_i| = i\binom{d(v)}{i}$. Let $m := |E(\Gamma_\ell(v))|$ and $m_j := \sum_{i=1}^{j} |L_i|$ for all $j \in [\ell]$. Label the edges of $\Gamma_\ell(v)$ as $e_1, \ldots, e_m$ in such a way that all the edges in L_1 come first, then the edges in L_2, and so on, until covering all the edges in L_ℓ. For each $j \in [m]$, let X_j be the indicator random variable that $e_j \in E(\mathcal{Q}^n_\varepsilon)$ (recall that we condition on the neighbourhood of v being revealed and v being good). We now consider an edge-exposure martingale given by the variables $Y_j \coloneqq \mathbb{E}[d^\ell(v) \mid X_1, \dots, X_j]$, for $j \in [m]_0$. This is a Doob martingale with $Y_{d(v)} = \mathbb{E}[d^\ell(v)]$ and $Y_m = d^\ell(v)$.

We must now bound the differences $|Y_j - Y_{j-1}|$, for all $j \in [m]$. Observe that the maximum change in the expected number of ℓ-dimensional cubes in $\mathcal{Q}^n_\varepsilon$ containing v when a new edge e_i is revealed is bounded from above by the number of such cubes in $\Gamma_\ell(v)$ containing e_i. Given any $k \in [\ell] \setminus \{1\}$ and any $i \in [m_k] \setminus [m_{k-1}]$, we claim that the number of copies of $\mathcal{Q}^\ell$ in $\Gamma_\ell(v)$ containing e_i is bounded by $\binom{d(v)-k}{\ell-k}$ (recall that for all $i \in [m_1]$ we have that $Y_i = Y_{i-1}$). Indeed, let $e_i = \{x, y\}$ with $\mathrm{dist}(x, v) = k - 1$, and let $\mathcal{D}_k(x) \subseteq \mathcal{D}(v)$ be the set of $k - 1$ directions such that $\mathcal{Q}^n(v, \mathcal{D}_k(x))$ contains x. Then, any copy $\mathcal{Q}$ of $\mathcal{Q}^\ell$ in $\Gamma_\ell(v)$ containing e_i must satisfy that $\mathcal{D}(\mathcal{Q}) \subseteq \mathcal{D}(v)$ contains $\mathcal{D}_k(x)$, the direction given by $y - x$, and any other $\ell - k$ of the directions in $\mathcal{D}(v)$, for which there is the claimed number of choices. Therefore, we conclude that

$$\text{(A.2)} \qquad \sum_{i=1}^{m} |Y_i - Y_{i-1}|^2 \leq \sum_{k=2}^{\ell} k \binom{d(v)}{k} \binom{d(v) - k}{\ell - k}^2 = \frac{1}{((\ell-2)!)^2} d(v)^{2\ell-2} \left(1 \pm \mathcal{O}\left(\frac{1}{d(v)}\right)\right).$$

Hence, by lemma 2.4.3, for n sufficiently large we have that

$$\mathbb{P}\left[|d^\ell(v) - \mathbb{E}[d^\ell(v)]| \geq \sqrt{\frac{2}{\varepsilon}} \frac{1}{(\ell-2)!} d(v)^{\ell-1/2} \right] \leq 2.1^{-n}.$$

Finally, by remark A.1.2 combined with a union bound on all vertices v such that $d(v) = \varepsilon n \pm n^a$, we conclude that a.a.s. all such vertices satisfy that $d^\ell(v) = (1 \pm \mathcal{O}(n^{a-1})) n^\ell \varepsilon^{2^{\ell-1}\ell}/\ell!$.

□

In more generality than lemma A.1.1, for the proof of lemma 2.5.10 we will need to use the fact that in $\mathcal{Q}^n_\varepsilon$ the directions in which the neighbours of a vertex lie are not correlated too much between vertices. Given a graph $G \subseteq \mathcal{Q}^n$, for any set $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ and any vertex $x \in \{0,1\}^n$, we denote $N_{G,S}(x) \coloneqq \{x + \hat{e} : \hat{e} \in S, \{x, x + \hat{e}\} \in E(G)\}$ and $d_{G,S}(x) \coloneqq |N_{G,S}(x)|$. Similarly, for any $y \in \{0,1\}^n$ such that $\mathrm{dist}(x, y) = 1$, we denote $N_{G,S,x}(y) \coloneqq N_{G,S}(y) \setminus \{x\}$ and $d_{G,S,x}(y) \coloneqq |N_{G,S,x}(y)|$.

Lemma A.1.3. *For every $\varepsilon,\delta\in(0,1)$, a.a.s. the following holds for every $x\in\{0,1\}^n$: for any set $S\subseteq\mathcal{D}(\mathcal{Q}^n)$ with $|S|\geq\delta n$, all but at most $100/(\varepsilon\delta)$ vertices $y\in N_{\mathcal{Q}^n}(x)$ satisfy $d_{\mathcal{Q}^n_\varepsilon,S,x}(y)\geq 2\varepsilon|S|/3$.*

Proof. Fix a vertex $x\in\{0,1\}^n$ and a set $S\subseteq\mathcal{D}(\mathcal{Q}^n)$ with $|S|\geq\delta n$. Choose any vertex $y\in N_{\mathcal{Q}^n}(x)$, and consider the variable $X(y)\coloneqq d_{\mathcal{Q}^n_\varepsilon,S}(y)$. It suffices to prove that a.a.s. $X(y)>2\varepsilon|S|/3$ for all but $100/(\varepsilon\delta)$ vertices $y\in N_{\mathcal{Q}^n}(x)$. Observe that $X(y)\sim\mathrm{Bin}(|S|,\varepsilon)$, so $\mathbb{E}[X(y)]=\varepsilon|S|$ and, by lemma 2.4.1,

$$\mathbb{P}[X(y)\leq 2\varepsilon|S|/3]\leq e^{-\varepsilon|S|/18}\leq e^{-\varepsilon\delta n/18}.$$

Observe, furthermore, that $N_{\mathcal{Q}^n}(x)$ is an independent set in $\mathcal{Q}^n$, hence the variables $\{X(y):y\in N_{\mathcal{Q}^n}(x)\}$ are mutually independent. It follows that the probability that at least $100/(\varepsilon\delta)$ vertices $y\in N_{\mathcal{Q}^n}(x)$ do not satisfy the bound is at most $\binom{n}{100/(\varepsilon\delta)}e^{-5n}$. Finally, by a union bound over all choices of S and x, we conclude that the statement fails with probability at most $2^{3n}e^{-5n}=o(1)$. □

Proof of lemma 2.5.10. Throughout this proof, for any $x\in\{0,1\}^n$ and any $y\in N_{\mathcal{Q}^n}(x)$, we will write $N(y)$ for $N_{\mathcal{Q}^n_\varepsilon}(y)$, $d(y)$ for $d_{\mathcal{Q}^n_\varepsilon}(y)$, $d_{S,x}(y)$ for $d_{\mathcal{Q}^n_\varepsilon,S,x}(y)$, $d^\ell(y)$ for $d^\ell_{\mathcal{Q}^n_\varepsilon}(y)$ and $d^\ell_{S,\ell^{1/2},x}(y)$ for $d^\ell_{\mathcal{Q}^n_\varepsilon,S,\ell^{1/2},x}(y)$.

Fix a set $S\subseteq\mathcal{D}(\mathcal{Q}^n)$ with $|S|\geq\delta n$. Let $D\coloneqq\varepsilon^{2^{\ell-1}\ell}n^\ell/\ell!$ and fix a vertex $x\in\{0,1\}^n$. Consider any $y\in N_{\mathcal{Q}^n}(x)$; reveal all edges incident to y and condition on the event that $d(y)=\varepsilon n\pm n^a$ and $d_{S,x}(y)\geq\varepsilon|S|/2$. By lemmas 2.5.7 and A.1.3, we have that a.a.s. all but at most $n^{2(1+\eta-a)}$ vertices $y\in N_{\mathcal{Q}^n}(x)$ satisfy this event.

Let $\mathcal{D}(y)$ be the set of directions such that $N(y)\setminus\{x\}=y+\mathcal{D}(y)$. Thus, $|\mathcal{D}(y)|=d(y)\pm 1$. Let $\alpha\coloneqq|S\cap\mathcal{D}(y)|/n$, and note that $\varepsilon\delta/2\leq\alpha\leq\varepsilon+n^{a-1}$. Similar to the proof of (A.1), we have that

$$\mathbb{E}[d^\ell_{S,\ell^{1/2},x}(y)]=\varepsilon^{(2^{\ell-1}-1)\ell}\sum_{i=\lceil\ell^{1/2}\rceil}^{\ell}\binom{\alpha n}{i}\binom{\varepsilon n-\alpha n\pm(n^a+1)}{\ell-i}\geq 3D/4.$$

Consider the graph $\Gamma_\ell(y) := \mathcal{Q}^n[y + \ell(\mathcal{D}(y) \cup \{\mathbf{0}\})]$. For each $i \in [\ell]$, let $L_i \subseteq E(\Gamma_\ell(y))$ be the set of edges which are at distance $i-1$ from y. Note that these sets partition $E(\Gamma_\ell(y))$ and that $|L_i| = i\binom{d(y)\pm 1}{i}$. Let $m := |E(\Gamma_\ell(y))|$ and $m_j := \sum_{i=1}^{j} |L_i|$ for all $j \in [\ell]$. Label the edges of $\Gamma_\ell(y)$ as $e_1, \dots, e_m$ in such a way that all the edges in L_1 come first, then the edges in L_2, and so on, until covering all the edges in L_ℓ. For each $j \in [m]$, let X_j be the indicator random variable that $e_j \in E(\mathcal{Q}^n_\varepsilon)$. We now consider an edge-exposure martingale given by the variables $Y_j := \mathbb{E}[d^\ell_{S,\ell^{1/2},x}(y) \mid X_1, \dots, X_j]$, for $j \in [m]_0$. This is a Doob martingale with $Y_{d(y)} = \mathbb{E}[d^\ell_{S,\ell^{1/2},x}(y)]$ and $Y_m = d^\ell_{S,\ell^{1/2},x}(y)$.

In order to bound the differences $|Y_j - Y_{j-1}|$, for all $j \in [m]$, observe that the maximum change in the expected number of ℓ-dimensional cubes in $\mathcal{Q}^n_\varepsilon$ containing y when a new edge e_i is revealed is bounded from above by the number of such cubes in $\Gamma_\ell(y)$ containing e_i. In particular, this is an upper bound for the maximum change in the expected number of ℓ-dimensional cubes in $\mathcal{Q}^n_\varepsilon$ containing y, not containing x, and whose directions intersect S in a set of size at least $\ell^{1/2}$, when a new edge e_i is revealed. Thus, similarly as in (A.2), it follows that

$$\sum_{i=1}^{m} |Y_i - Y_{i-1}|^2 \le \frac{1}{((\ell-2)!)^2} d(y)^{2\ell-2}\left(1 \pm \mathcal{O}\left(\frac{1}{d(y)}\right)\right).$$

Hence, by lemma 2.4.3, we have that

$$\mathbb{P}\left[d^\ell_{S,\ell^{1/2},x}(y) \le D/2\right] \le e^{-\Theta(n^2)}.$$

Finally, the statement follows by a union bound on all sets $S \subseteq \mathcal{D}(\mathcal{Q}^n)$ with $|S| \ge \delta n$, on all vertices $x \in \{0,1\}^n$, and on all vertices $y \in N_{\mathcal{Q}^n}(x)$ such that $d(y) = \varepsilon n \pm n^a$ and $d_{S,x}(y) \ge \varepsilon|S|/2$. $\square$

A.2. Bondless and bondlessly surrounded molecules and robust subgraphs

Proof of lemma 2.7.2. Fix an (s,ℓ)-molecule $\mathcal{M} = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_{2^s} \subseteq \mathcal{Q}^n$. Consider a pair of consecutive atoms $\mathcal{A}_i, \mathcal{A}_{i+1} \subseteq \mathcal{M}$, for some $i \in [2^s]$. Let X_i be the number of edges between $\mathcal{A}_i$ and $\mathcal{A}_{i+1}$ in $\mathcal{Q}^n_\varepsilon$ whose endpoint in $\mathcal{A}_i$ is odd, and let Y_i be the number of such edges whose endpoint in $\mathcal{A}_i$ is even. We have that $X_i, Y_i \sim \mathrm{Bin}(2^{\ell-1}, \varepsilon)$. By lemma 2.4.1, it follows that

$$\mathbb{P}[X_i < 100] \leq 2^{-\varepsilon 2^\ell/4},$$

and the same bound holds for $\mathbb{P}[Y_i < 100]$. By a union bound over all $i \in [2^s]$, we conclude that

$$\mathbb{P}[\mathcal{M} \text{ is bondless in } \mathcal{Q}^n_\varepsilon] \leq 2^{s+1-\varepsilon 2^\ell/4}. \qquad \square$$

Proof of lemma 2.7.3. Let $G \sim \mathcal{Q}^n_\varepsilon$ and let $\mathcal{C}' := \{C - (N_I(x) \cup \{x\}) : C \in \mathcal{C}\}$. Given $C' \in \mathcal{C}'$, for each $i \in [2^s]$, let C'_i be the i-th clone of C'. We denote $\mathcal{M}_{C'} := C'_1 \cup \cdots \cup C'_{2^s}$, and refer to it as the molecule of C' in $\mathcal{Q}^n$. We say that $\mathcal{M}_{C'}$ is *bonded in* G if, for each $i \in [2^s]$, the graph G contains at least 100 edges between C'_i and C'_{i+1} whose endpoints in C'_i have odd parity and 100 edges whose endpoints in C'_i have even parity, where indices are taken cyclically. Note that, if $C' = C - (N_I(x) \cup \{x\})$ for some $C \in \mathcal{C}$ and C' is bonded in G, then C must be bonded in G. Moreover, $|V(C')| = |V(C)| - \ell - 1 > 9|V(C)|/10$. Therefore, similarly to the proof of lemma 2.7.2, by lemma 2.4.1 and a union bound we have that

$$\mathbb{P}[\mathcal{M}_{C'} \text{ is bondless in } G] \leq 2^{s-\varepsilon 2^\ell/100} \leq 1/2.$$

Let $X := |\{C' \in \mathcal{C}' : \mathcal{M}_{C'} \text{ is bonded in } G\}|$. It follows that

$$\mathbb{E}[|B(G)|] \geq \mathbb{E}[X] \geq \gamma n^\ell/2. \tag{A.3}$$

Let $V := \bigcup_{C' \in \mathcal{C}'} V(\mathcal{M}_{C'})$. Let $e_1, \ldots, e_m$ be an arbitrary ordering of the edges of $E := \bigcup_{i \in [2^s]} E_{Q^n}(L_i \cap V, L_{i+1} \cap V)$. For each $j \in [m]$, let X_j be the indicator variable which takes value 1 if $e_j \in E(G)$ and 0 otherwise. Consider the edge-exposure martingale $Y_j := \mathbb{E}[X \mid X_1, \ldots, X_j]$ for $j \in [m]_0$. This is a Doob martingale with $Y_0 = \mathbb{E}[X]$ and $Y_m = X$.

We will now bound the differences $|Y_j - Y_{j-1}|$, for all $j \in [m]$. For each $i \in [\ell] \setminus \{1\}$, let $N^i(x) := \{y \in \bigcup_{C' \in \mathcal{C}'} V(C') : \operatorname{dist}(x, y) = i\}$. Let $E^i \subseteq E$ be the collection of edges $e = (u, v)$ where both u and v are clones of a vertex $z \in N^i(x)$. Note that the sets $E^2, \ldots, E^\ell$ partition E. Moreover, for each $j \in [m]$, if $e_j \in E^i$, then $|Y_j - Y_{j-1}| \leq n^{\ell - i}$. Furthermore, $|E^i| = 2^s |N^i(x)| \leq 2^s n^i$ and, thus,

$$\sum_{j \in [m]} |Y_j - Y_{j-1}|^2 \leq \sum_{i=2}^{\ell} |E^i(x)| n^{2\ell - 2i} \leq \sum_{i=2}^{\ell} 2^s n^i n^{2\ell - 2i} = \mathcal{O}(n^{2\ell - 2}).$$

Therefore, we can apply lemma 2.4.3 and combine it with (A.3) to obtain

$$\mathbb{P}[|B(G)| < \gamma n^\ell / 4] \leq \mathbb{P}[X < \gamma n^\ell / 4] \leq \mathbb{P}[|X - \mathbb{E}[X]| \geq \mathbb{E}[X]/2] \leq e^{-10n}. \qquad \square$$

Proof of lemma 2.7.4. We begin by fixing an arbitrary vertex $x \in V(Q^n)$ and an arbitrary set $B \subseteq \mathfrak{M}$ of $n^{1/3}$ molecules which intersect $B^{\ell^2}_{Q^n}(x)$. We will estimate the probability that all of the molecules in B are bondlessly surrounded in Q^n_ε, by considering the neighbourhoods of the different vertices which make up these molecules. If the probability of being bondlessly surrounded was independent over different molecules and vertices, then this would be a straightforward calculation. However, there are dependencies which we must consider: namely, when two different molecules have edges to the same third molecule. We will first bound the number of such configurations in Q^n. Since the molecules in $\mathfrak{M} \supseteq B$ are vertex-disjoint, it follows that, if two of these molecules are adjacent in Q^n, then all of their atoms are pairwise adjacent in each of the layers, via clones of the same edges. Thus, we can restrict the analysis to a single layer.

Fix a layer L and let $\mathfrak{A}$ be the collection of atoms obtained by intersecting each molecule $\mathcal{M} \in \mathfrak{M}$ with L. Let $\mathfrak{A}_B \subseteq \mathfrak{A}$ be the set of such atoms whose molecules lie in B. Fix an atom

$\mathcal{A} \in \mathfrak{A}_B$, and let $y \in V(\mathcal{A})$ be a fixed vertex. We say an atom $\mathcal{A}' \in \mathfrak{A}$ is *y-dependent* if there exists $\mathcal{A}'' \in \mathfrak{A}_B$, $\mathcal{A}'' \neq \mathcal{A}$, such that $\mathrm{dist}(y, \mathcal{A}') = \mathrm{dist}(\mathcal{A}', \mathcal{A}'') = 1$. The following claim will allow us to bound the number of y-dependent atoms.

Claim A.1. *Fix $\mathcal{A}'' \in \mathfrak{A}_B$ with $\mathcal{A}'' \neq \mathcal{A}$. Then, the number of $\mathcal{A}' \in \mathfrak{A}$ for which $\mathrm{dist}(y, \mathcal{A}') = \mathrm{dist}(\mathcal{A}', \mathcal{A}'') = 1$ is at most $2^\ell(2 + \sqrt{n})$.*

Proof. Let $z \in V(\mathcal{A}'')$ and let $\hat{e} \in \mathcal{D}(y, z)$. Let $\mathcal{A}' \in \mathfrak{A}$ be such that $\mathrm{dist}(y, \mathcal{A}') = \mathrm{dist}(z, \mathcal{A}') = 1$. Suppose first that $\hat{e} \notin \mathcal{D}(\mathcal{A}')$. Then we must have either $y + \hat{e} \in V(\mathcal{A}')$ or $z + \hat{e} \in V(\mathcal{A}')$. Since all the atoms in $\mathfrak{A}$ are vertex-disjoint, this leaves only two possibilities for $\mathcal{A}'$. Alternatively, suppose $\hat{e} \in \mathcal{D}(\mathcal{A}')$. Then, by (BS) applied with y playing the role of x, we have at most $\sqrt{n}$ possibilities for $\mathcal{A}'$. Finally, by considering all $z \in V(\mathcal{A}'')$ we prove the claim. ◀

By considering all possibilities for $\mathcal{A}'' \in \mathfrak{A}_B$, since $|\mathfrak{A}_B| = n^{1/3}$, it follows by Claim A.1 that the number of y-dependent atoms is at most $n^{6/7}$. For each $y \in V(\mathcal{A})$, let $N'(y) \subseteq N^{\mathfrak{M}}(y)$ be given by removing from $N^{\mathfrak{M}}(y)$ all molecules which contain a y-dependent atom. It follows that $|N'(y)| = |N^{\mathfrak{M}}(y)| - o(n)$ for every $y \in V(\mathcal{A})$.

Let $\mathcal{M}_\mathcal{A} \in \mathfrak{M}$ be the molecule containing $\mathcal{A}$. For each vertex $y \in V(\mathcal{A})$, let $\mathcal{E}_y$ be the event that $N'(y)$ contains at least $n/2^{\ell+5s+1}$ molecules $\mathcal{M} \in \mathfrak{M}$ which are bondless in $\mathcal{Q}^n_\varepsilon$. Then, $|N'(y)| \geq n/2^{\ell+5s+2}$. Moreover, we only consider here those vertices $y \in V(\mathcal{A})$ for which $|N^{\mathfrak{M}}(y)| \geq n/2^{\ell+5s+1}$, since otherwise y cannot contribute towards $\mathcal{M}_\mathcal{A}$ being bondlessly surrounded. Fix such a vertex y. Let Y be the number of atoms $\mathcal{A} \in N'(y)$ which correspond to molecules which are bondless in $\mathcal{Q}^n_\varepsilon$. Note that Y is a sum of independent indicator variables. By lemma 2.7.2, we have that $\mathbb{E}[Y] \leq 2^{s+1-\varepsilon 2^\ell/4}n$. In order to derive a lower bound for $\mathbb{E}[Y]$, note that the probability that an (s, ℓ)-molecule $\mathcal{M}$ is bondless can be bounded from below by the probability that there are no edges between two fixed consecutive atoms $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{M}$, whose endpoints in $\mathcal{A}_1$ are even. This occurs with probability $(1-\varepsilon)^{2^{\ell-1}}$. Thus,

$$\mathbb{E}[Y] \geq (1-\varepsilon)^{2^{\ell-1}}|N'(y)| \geq (1-\varepsilon)^{2^{\ell-1}}(n/2^{\ell+5s+2}).$$

By lemma 2.4.2, we have that $\mathbb{P}[\mathcal{E}_y] \leq 2^{-cn}$, for some constant $c > 0$ which depends on ℓ and ε. For each atom $\mathcal{A} \in \mathfrak{A}_B$, let $\mathcal{B}_{\mathcal{A}}$ be the event that there exists a vertex $y \in V(\mathcal{A})$ such that $\mathcal{E}_y$ holds. Let $\mathcal{B} := \bigwedge_{\mathcal{A} \in \mathfrak{A}_B} \mathcal{B}_{\mathcal{A}}$. Note that the definition of $N'(y)$ ensures that the events $\mathcal{B}_{\mathcal{A}}$ with $\mathcal{A} \in \mathfrak{A}_B$ are pairwise independent. Thus,

$$\mathbb{P}[\mathcal{E}] \leq (2^{\ell - cn})^{n^{1/3}} < 2^{-n^{5/4}}.$$

In turn, this means that the probability that all molecules $\mathcal{M} \in B$ are bondlessly surrounded is bounded from above by $2^{-n^{5/4}}$. lemma 2.7.4 now follows by a union bound over the 2^n choices for x and the at most $\binom{n^{\ell^2}}{n^{1/3}}$ choices for B. □

Proof of lemma 2.7.7. We begin with a proof of (i). Let $\mathcal{U} := \mathcal{U}(G, \varepsilon_1)$. For any $x \in V(\mathcal{Q}^n)$, we have that

$$\mathbb{P}[x \in \mathcal{U}] \leq \binom{n}{\varepsilon_1 n}(1/2 + \varepsilon)^{n - \varepsilon_1 n} < 2^{-n + 20\varepsilon n}. \tag{A.4}$$

It follows that $\mathbb{E}[|\mathcal{U}|] < 2^n 2^{-n+20\varepsilon n} = 2^{20\varepsilon n}$. Therefore, by Markov's inequality we have that

$$\mathbb{P}[|\mathcal{U}| \geq 2^{\varepsilon_2 n}] < 2^{20\varepsilon n}/2^{\varepsilon_2 n} < 2^{-\varepsilon_2 n/2},$$

so a.a.s. $|\mathcal{U}| \leq 2^{\varepsilon_2 n}$. (R1) holds trivially by the choice of $\mathcal{U}$. Furthermore, we have that

$$\left| \bigcup_{i \in [4]} B^{s+\ell}_{\mathcal{Q}^n}(x_i) \right| \leq 5n^{s+\ell},$$

so, by (A.4) and a union bound, (R5) also holds a.a.s.

To see that (R2) holds, fix $x \in V(\mathcal{Q}^n)$ and $y \in B^{s+5\ell}_{\mathcal{Q}^n}(x)$. Then, $\mathbb{E}[d_{G-\{x\}}(y)] = (1/2 - \varepsilon)n \pm 1$. Thus, by lemma 2.4.1,

$$\mathbb{P}[d_{G-\{x\}}(y) \leq \gamma n - 1] \leq 2^{-n/20}.$$

Therefore, by (A.4), we have that $\mathbb{P}[x \in \mathcal{U} \wedge d_{G-\{x\}}(y) \leq \gamma n - 1] \leq 2^{-n+20\varepsilon n - n/20} \leq 2^{-31n/30}$. A union bound over all $x \in V(\mathcal{Q}^n)$ and over all $y \in B^{s+5\ell}_{\mathcal{Q}^n}(x)$ shows that (R2) holds a.a.s.

The fact that (R3) holds a.a.s. can be shown similarly.

Finally, consider (R4). Let $x \in V(\mathcal{Q}^n)$, and suppose $x \in V(L)$, for some layer L. First, let $a, b \in \mathcal{D}(L)$ be distinct. We are going to show that a.a.s. we can find the desired collection of (x, a, b)-consistent systems of paths.

Recall that an (x, a, b)-consistent system of paths $(P_1, \dots, P_6)$, as defined in section 2.7.2, is determined uniquely by a 6-tuple of directions $(c, d, d_1, d_2, d_3, d_4)$. In order to show that (R4.I) is satisfied, we will first consider the directions c and d, and then the rest of the tuple. Recall that all (x, a, b)-consistent systems of paths contain the two edges $\{x, x + a\}$ and $\{x, x + b\}$. Then, once c and d are fixed, this determines a total of 6 more edges. The remaining 8 edges will be determined by the choice of (d_1, d_2, d_3, d_4).

Consider a collection $\mathcal{W}$ of disjoint pairs of distinct directions (c, d) with $c, d \in \mathcal{D}(L) \setminus \{a, b\}$ such that $|\mathcal{W}| \geq n/4$. For each $(c, d) \in \mathcal{W}$, let $E^*(c, d) \subseteq E(\mathcal{Q}^n)$ be the set of six edges of an (x, a, b)-consistent system of paths determined by these two directions. Observe that, since the pairs in $\mathcal{W}$ are disjoint, it follows that, for any distinct $(c, d), (c', d') \in \mathcal{W}$, we have $E^*(c, d) \cap E^*(c', d') = \varnothing$. Now let $\mathcal{W}_G \coloneqq \{(c, d) \in \mathcal{W} : E^*(c, d) \subseteq E(G)\}$ and $X \coloneqq |\mathcal{W}_G|$. We have that $\mathbb{E}[X] \geq (1/2 - \varepsilon)^6 n/4$ and, by lemma 2.4.1, it follows that $\mathbb{P}[X \leq \gamma n] \leq 2^{-\gamma n}$.

For each $(c, d) \in \mathcal{W}$, let $\mathcal{V}(c, d)$ be a collection of disjoint 4-tuples of distinct directions (d_1, d_2, d_3, d_4) with $d_1, d_2, d_3, d_4 \in \mathcal{D}(L) \setminus \{a, b, c, d\}$ such that $|\mathcal{V}(c, d)| \geq n/5$. For each $(d_1, d_2, d_3, d_4) \in \mathcal{V}(c, d)$, let $E^*(c, d, d_1, d_2, d_3, d_4) \subseteq E(\mathcal{Q}^n)$ be the set of eight edges of an (x, a, b)-consistent system of paths determined by $(c, d, d_1, d_2, d_3, d_4)$ which are not in $E^*(c, d) \cup \{\{x, x + a\}, \{x, x + b\}\}$. In particular, since the tuples in $\mathcal{V}(c, d)$ are disjoint, it follows that, for any distinct $(d_1, d_2, d_3, d_4), (d_1', d_2', d_3', d_4') \in \mathcal{V}(c, d)$, we have $E^*(c, d, d_1, d_2, d_3, d_4) \cap E^*(c, d, d_1', d_2', d_3', d_4') = \varnothing$. Now let

$$\mathcal{V}_G(c, d) \coloneqq \{(d_1, d_2, d_3, d_4) \in \mathcal{V}(c, d) : E^*(c, d, d_1, d_2, d_3, d_4) \subseteq E(G)\},$$

and let $Y(c,d) := |\mathcal{V}_G(c,d)|$. We then have that $\mathbb{E}[Y(c,d)] \geq (1/2-\varepsilon)^8 n/5$ and, again by lemma 2.4.1, it follows that $\mathbb{P}[Y(c,d) \leq \gamma n] \leq 2^{-\gamma n}$. Thus, by a union bound, with probability at least $1-2^{-\gamma n/2}$, for every $(c,d) \in \mathcal{W}$ we have $Y(c,d) \geq \gamma n$.

Let $\mathcal{E}(x,a,b)$ be the event that $G \cup \{\{x,x+a\},\{x,x+b\}\}$ contains a collection $\mathfrak{C}(x,a,b)$ of (x,a,b)-consistent systems of paths satisfying (R4.I). By combining all the above, it follows that

$$\mathbb{P}[\mathcal{E}(x,a,b)] \geq 1-2^{-\gamma n/4}. \tag{A.5}$$

The same bound can be proved for the cases where $a \notin \mathcal{D}(L), b \in \mathcal{D}(L)$ and $a,b \notin \mathcal{D}(L)$. Observe that, for any $x \in V(\mathcal{Q}^n)$ and $a,b \in \mathcal{D}(\mathcal{Q}^n)$, the event $\mathcal{E}(x,a,b)$ is independent of the event that $x \in \mathcal{U}$. Now, by combining (A.5) with (A.4) and a union bound over all choices of $a,b \in \mathcal{D}(\mathcal{Q}^n)$, and then a union bound over all $x \in V(\mathcal{Q}^n)$, we conclude that (R4) holds a.a.s.

The proof of (ii) is similar. Let $H \subseteq \mathcal{Q}^n$ be given and consider a random partition of $E(H)$ into r parts $H_1,\dots,H_r$, in such a way that each edge is assigned to one of the parts uniformly and independently of all other edges. For each $x \in V(\mathcal{Q}^n) \setminus \mathcal{U}$, let $\mathcal{B}(x)$ be the event that there exists some $i \in [r]$ such that $d_{H_i}(x) < d_H(x)/(2r)$. Observe that, if $\overline{\mathcal{B}(x)}$ holds for all $x \in V(\mathcal{Q}^n) \setminus \mathcal{U}$, then no vertex outside $\mathcal{U}$ will be contained in $\mathcal{U}(H_i,\varepsilon_1/(2r))$ for any $i \in [r]$. It would then follow that (R1), (R2), (R3) and (R5) all hold with the desired constants for each H_i. Fix $x \in V(\mathcal{Q}^n) \setminus \mathcal{U}$ and $i \in [r]$. Let $X := d_{H_i}(x)$. Then, $\mathbb{E}[X] = d_H(x)/r \geq \varepsilon_1 n/r$. Thus, by lemma 2.4.1, $\mathbb{P}[X \leq \mathbb{E}[X]/2] \leq e^{-\varepsilon_1^2 n}$. A union bound over all $i \in [r]$ shows that $\mathbb{P}[\mathcal{B}(x)] \leq re^{-\varepsilon_1^2 n} \leq e^{-\varepsilon_1^3 n}$.

We now consider the property (R4). For each $x \in \mathcal{U}$, let $\mathcal{B}(x)$ be the event that there exist $i \in [r]$ and distinct directions $a,b \in \mathcal{D}(\mathcal{Q}^n)$ such that (R4) does not hold for H_i with γ/r^{10} playing the role of γ.

Fix $x \in \mathcal{U}$, $i \in [r]$ and distinct directions $a,b \in \mathcal{D}(\mathcal{Q}^n)$. Similarly as in (i), using lemma 2.4.1 one can show that the probability that H_i does not satisfy (R4) for x with γ/r^{10}

playing the role of γ is at most $2^{-\gamma^2 n} + 2^{-\gamma^3 n}$. Therefore, by a union bound over all choices of $a, b \in \mathcal{D}(\mathcal{Q}^n)$ and over each $i \in [r]$, we have that $\mathbb{P}[\mathcal{B}(x)] \leq rn^2(2^{-\gamma^2 n} + 2^{-\gamma^3 n}) \leq e^{-\varepsilon_1^3 n}$.

Finally, we are interested in the event where $\mathcal{B}(x)$ does not occur for any $x \in V(\mathcal{Q}^n)$. We will invoke lemma 2.4.6. Note that each event $\mathcal{B}(x)$ is mutually independent of all but at most n^{10} other events. We have that $\mathbb{P}[\mathcal{B}(x)] \leq e^{-\varepsilon_1^3 n}$ for every $x \in V(\mathcal{Q}^n)$ and $e \cdot e^{-\varepsilon_1^3 n}(n^{10}+1) < 1$, so by lemma 2.4.6 there exists an edge-decomposition of H with the desired properties. □

Proof of lemma 2.7.8. Let $u \in V(I)$ and let $D \subseteq B_I^{10\ell}(u) \setminus \mathcal{U}_I$ be a set of C vertices. Let $D' := \bigcup_{x,y \in D : x \neq y} N_I(x) \cap N_I(y)$. Since any pair of distinct vertices in I share at most two neighbours, we have that $|D'| \leq 2\binom{C}{2}$. For each $i \in [2^s]$, we denote the i-th clone of D' by D'_i, and let $R'_i := R_i \setminus D'_i$.

For each $x \in V(\mathcal{Q}^n)$, let $i(x)$ be the unique index $i \in [2^s]$ such that $x \in V(L_i)$. Observe that, by (R1), we have $e_H(x, V(L_{i(x)})) > 2\varepsilon_1 n/3$ for every $x \in V(\mathcal{Q}^n) \setminus \mathcal{U}$. For each $x \in V(\mathcal{Q}^n)$, let $\mathcal{E}_x$ be the event that $e_H(x, R_{i(x)}) \leq \varepsilon_1 \delta n/4$, and let $\mathcal{E}'_x$ be the event that $e_H(x, R'_{i(x)}) \leq \varepsilon_1 \delta n/4$. It follows by lemma 2.4.1 that $\mathbb{P}[\mathcal{E}'_x] \leq e^{-\varepsilon_1 \delta n/16}$ for all $x \in V(\mathcal{Q}^n) \setminus \mathcal{U}$. For each $v \in V(I)$, let $\mathcal{E}_v$ and $\mathcal{E}'_v$ be the events that there exists $i \in [2^s]$ with $v_i \notin \mathcal{U}$ such that $\mathcal{E}_{v_i}$ and $\mathcal{E}'_{v_i}$ hold, respectively. By a union bound, it follows that $\mathbb{P}[\mathcal{E}'_v] \leq 2^s e^{-\varepsilon_1 \delta n/16}$ for all $v \in V(I)$. Finally, let $\mathcal{E}_D$ and $\mathcal{E}'_D$ be the events that $\mathcal{E}_v$ and $\mathcal{E}'_v$, respectively, hold for every $v \in D$. Note that the events in the collection $\{\mathcal{E}'_v : v \in V(I)\}$ are mutually independent. Furthermore, since the event $\mathcal{E}_x$ implies $\mathcal{E}'_x$ for all $x \in V(\mathcal{Q}^n)$, we have that

$$\mathbb{P}[\mathcal{E}_D] \leq \mathbb{P}[\mathcal{E}'_D] \leq (2^s e^{-\varepsilon_1 \delta n/16})^C < e^{-5n}.$$

By a union bound over all $u \in V(I)$ and over all choices of D, we have $\mathbb{P}[\mathcal{E}_1] \leq e^{-n}$.

Consider now any $u \in \mathcal{U}_I$. Observe that, if $v \in B_I^{5\ell}(u)$, then for every $i, j \in [2^s]$ we have that $\mathrm{dist}(u_i, v_j) \leq 5\ell + s$. Therefore, by (R2), for all $v \in B_I^{5\ell}(u)$ and $i \in [2^s]$ such that $v_i \notin \mathcal{U}$, we have $d_H(v_i) \geq \gamma n$. For each $v \in B_I^{5\ell}(u)$ and each $i \in [2^s]$ with $v_i \notin \mathcal{U}$, let $\mathcal{F}_{v_i}$ be the event that $e_H(v_i, R_i) \leq \varepsilon_1 \delta n/4$, and let $\mathcal{F}_v$ be the event that there exists some $i \in [2^s]$ with $v_i \notin \mathcal{U}$ such that $\mathcal{F}_{v_i}$ holds. By lemma 2.4.1 and a union bound, it follows that $\mathbb{P}[\mathcal{F}_v] \leq 2^{-\gamma \delta n/16}$.

Then, by a union bound over all $u \in \mathcal{U}_I$ and $v \in B_I^{5\ell}(u)$,

$$\mathbb{P}[\mathcal{E}_2] \leq |\mathcal{U}_I| \cdot |B_I^{5\ell}(\mathcal{U}_I)| \cdot \mathbb{P}[\mathcal{F}_v] \leq 2n^{5\ell} 2^{\varepsilon_2 n} 2^{-\gamma\delta n/16}. \qquad \square$$

A.3. Connecting cubes: proof of lemma 2.7.12

Proof of lemma 2.7.12. The proof is similar (but easier) to that of lemma 2.7.11. As before, by relabelling the atoms, we may assume that $\mathcal{M}^* = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_t$. Without loss of generality, we may assume that, for each $r \in [2]$, if $u_r \in R$, then $u_r = z_1$, and if $v_r \in R$, then $v_r = w_t$. Moreover, we may assume that $x \neq_{\mathrm{p}} u_1$. Let $S := \{u_1, v_1, u_2, v_2\}$. Let $I_R := \{k \in [t] : R \cap V(\mathcal{A}_k) \cap S \neq \varnothing\}$, $R^* := R \setminus \bigcup_{k \in I_R} V(\mathcal{A}_k)$ and $I_{R^*} := \{k \in [t] : R^* \cap V(\mathcal{A}_k) \neq \varnothing\}$. For each $r \in [2]$, let $I_R^r \subseteq \{1, t\}$ be such that $1 \in I_R^r$ if and only if $u_r \in R$ and $t \in I_R^r$ if and only if $v_r \in R$. Note that $I_R = I_R^1 \cup I_R^2$. Fix an index $t^* \in [t-1] \setminus (I_{R^*} \cup \{1\})$. If $|L| = 2$, let $I_L^1 := \{i\}$, $I_L^2 := \{j\}$ and $I_L := \{i, j\}$; otherwise, let $I_L^1 := I_L^2 := I_L := \{t^*\}$.

Once more, for each $r \in [2]$, we create an ordered list $\mathcal{L}_r$ of vertices, which will be used to construct the vertex-disjoint paths $\mathcal{P}_r$. Again, given any list of vertices $\mathcal{L}_r$, we write L_r^* to denote the (unordered) set of vertices in $\mathcal{L}_r$, and whenever $\mathcal{L}_r$ is updated, we implicitly update L_r^*. In the end, for each $r \in [2]$ we will have a list of vertices $\mathcal{L}_r = (x_1^r, \dots, x_{\ell_r}^r)$. For each $r \in [2]$ and $k \in [t]$, let $I_r(k) := \{h \in [\ell_r - 1] : 2 \nmid h \text{ and } x_h^r, x_{h+1}^r \in V(\mathcal{A}_k)\}$. We will require $\mathcal{L}_1$ and $\mathcal{L}_2$ to be vertex-disjoint and to satisfy the following properties:

($\mathcal{L}'$1) ℓ_1 and ℓ_2 are even.

($\mathcal{L}'$2) For each $r \in [2]$, for all $h \in [\ell_r - 1]$, if h is odd, then $x_h^r, x_{h+1}^r \in V(\mathcal{A}_k)$, for some $k \in [t]$; if h is even, then $x_h^r x_{h+1}^r \in E(G \cup \mathcal{M}^*)$.

($\mathcal{L}'$3) For all $k \in [t]$ we have that $|I_1(k)|, |I_2(k)| \geq 1$ and $2 \leq |I_1(k)| + |I_2(k)| \leq 3$.

($\mathcal{L}'$4) For each $r \in [2]$, the following holds. For each $k \in [t] \setminus (I_L^r \cup I_R^r)$ and each $h \in I_r(k)$, we have $x_h^r \neq_{\mathrm{p}} x_{h+1}^r$. For each $k \in I_L^r \cup I_R^r$, we have that $|I_r(k)| = 1$ and for the unique index $h \in I_r(k)$ we have $x_h^r =_{\mathrm{p}} x_{h+1}^r$, with the same parity as u_r in the case when

$k \in I_L^r$, and with parity opposite to that of the unique vertex in $\{w_k, z_k\} \cap \{u_r, v_r\}$ in the case when $k \in I_R^r$.

($\mathcal{L}'5$) For each $r \in [2]$, we have the following. If $u_r \notin R$, then $u_r = x_1^r$. If $v_r \notin R$, then $v_r = x_{\ell_r}^r$. If $u_r \in R$ (and thus $u_r = z_1$), then $w_1 = x_1^r$ and $u_r \notin L_1^* \cup L_2^*$. If $v_r \in R$ (and thus $v_r = w_t$), then $z_t = x_{\ell_r}^r$ and $v_r \notin L_1^* \cup L_2^*$.

($\mathcal{L}'6$) Every pair (w_k, z_k) with $\{w_k, z_k\} \subseteq R^*$ is contained in $\mathcal{L}_1$ and z_k directly succeeds w_k or vice versa.

If $R^* \cap V(\mathcal{A}_1) = \{w_1, z_1\}$, then let $\mathcal{L}_1 \coloneqq (u_1, w_1, z_1)$, where we assume that $w_1 \neq_{\mathrm{p}} u_1$; otherwise, let $\mathcal{L}_1 \coloneqq (u_1)$. Observe once more that, in what follows, the existence of each alternating parity sequence follows from the bondedness of $\mathcal{M}$.

Let $F_1 \coloneqq L \cup R \cup S$ and let $t_1^\bullet \in I_L^1$. Let $\mathcal{S}_1$ be a $(u_1, t_1^\bullet, F_1, R)$-alternating parity sequence. If $u_1 \in R$, update $\mathcal{L}_1 \coloneqq \mathcal{S}_1$; otherwise, update $\mathcal{L}_1 \coloneqq \mathcal{L}_1\mathcal{S}_1^-$. Choose any vertex $u_{t_1^\bullet} \in V(\mathcal{A}_{t_1^\bullet})$ with $u_{t_1^\bullet} \neq_{\mathrm{p}} u_1$, and let $\mathcal{S}_2$ be a $(u_{t_1^\bullet}, t, F_1, R^*)$-alternating parity sequence. Update $\mathcal{L}_1 \coloneqq \mathcal{L}_1\mathcal{S}_2^-$. If $v_1 \in R$, update $\mathcal{L}_1 \coloneqq \mathcal{L}_1(z_t)$. Otherwise, update $\mathcal{L}_1 \coloneqq \mathcal{L}_1(v_1)$.

Next, let $F_2 \coloneqq F_1 \cup L_1^*$ and let $t_2^\bullet \in I_L^2$. Let $\mathcal{S}_3$ be a $(u_2, t_2^\bullet, F_2, R \cap V(\mathcal{A}_1))$-alternating parity sequence, and let $\mathcal{L}_2 \coloneqq \mathcal{S}_3$. Choose any vertex $u'_{t_2^\bullet} \in V(\mathcal{A}_{t_2^\bullet})$ with $u'_{t_2^\bullet} \neq_{\mathrm{p}} u_2$, and let $\mathcal{S}_4$ be a $(u'_{t_2^\bullet}, t, F_2, \varnothing)$-alternating parity sequence. Update $\mathcal{L}_2 \coloneqq \mathcal{L}_2\mathcal{S}_4^-$. Finally, if $v_2 \in R$, update $\mathcal{L}_2 \coloneqq \mathcal{L}_2(z_t)$. Otherwise, update $\mathcal{L}_2 \coloneqq \mathcal{L}_2(v_2)$.

Observe that $\mathcal{L}_1$ and $\mathcal{L}_2$ satisfy ($\mathcal{L}'1$)–($\mathcal{L}'6$). We are now in a position to apply lemma 2.7.9. For each $k \in [t]$, let $t_k \coloneqq |I_1(k)|+|I_2(k)|$. Again, for any $r \in [2]$ and $k \in [t]$, for each $h \in I_r(k)$, we refer to the pair x_h^r, x_{h+1}^r as a *matchable pair*. By ($\mathcal{L}'3$), ($\mathcal{L}'4$) and lemma 2.7.9(i), each $\mathcal{A}_k$ with $k \in [t] \setminus (I_L \cup I_R)$ can be covered by t_k vertex-disjoint paths, each of whose endpoints are a matchable pair contained in $\mathcal{A}_k$. Similarly, by ($\mathcal{L}'3$), ($\mathcal{L}'4$) and lemma 2.7.9(ii), each $\mathcal{A}_k$ with $k \in I_R$ contains t_k vertex-disjoint paths, each of whose endpoints are a matchable pair in $\mathcal{A}_k$, such that the union of these t_k paths covers precisely $V(\mathcal{A}_k) \setminus (S \cap R)$. Similarly, if $L \neq \varnothing$ and $k \in I_L$, then $\mathcal{A}_k$ contains t_k paths, each of whose endpoints are a matchable pair in $\mathcal{A}_k$, such that the union of these t_k paths covers precisely $V(\mathcal{A}_k) \setminus L$. Finally, by ($\mathcal{L}'3$),

$(\mathcal{L}'4)$ and lemma 2.7.9(iii), if $L = \varnothing$ and $k \in I_L$ (that is, $k = t^*$), then $\mathcal{A}_k$ can be covered by t_k paths, each of whose endpoints are a matchable pair in $\mathcal{A}_k$. For each matchable pair x_h^r, x_{h+1}^r in $\mathcal{A}_k$, let us denote the corresponding path by $\mathcal{P}_{x_h^r, x_{h+1}^r}$.

The paths $\mathcal{P}_r$ required for lemma 2.7.12 can now be constructed as follows. For each $r \in [2]$, let $\mathcal{P}_r$ be the path obtained from the concatenation of the paths $\mathcal{P}_{x_h^r, x_{h+1}^r}$, for each odd $h \in [\ell_r]$, via the edges $x_h^r x_{h+1}^r$ for $h \in [\ell_r - 1]$ even. By $(\mathcal{L}'5)$, if $\mathcal{P}_r$ does not contain u_r, then $\mathcal{P}_r$ starts in w_1, and u_r does not lie in any other path; therefore, we can update $\mathcal{P}_r$ as $\mathcal{P}_r := u_r\mathcal{P}_r$. Similarly, if $\mathcal{P}_r$ does not contain v_r, then $\mathcal{P}_r$ ends in z_t and v_r does not lie in any other path, hence we can update $\mathcal{P}_r$ as $\mathcal{P}_r := \mathcal{P}_r v_r$. It follows that $V(\mathcal{P}_1 \cup \mathcal{P}_2) = V(\mathcal{M}^*) \setminus L$, and thus the paths $\mathcal{P}_r$ are as required for lemma 2.7.12. □

A.4. Proof of lemma 3.3.4

Proof of lemma 3.3.4. Clearly, $D_K \geq D_F$ for each $K \subseteq F$, and $D_K \leq X_K$. Thus, $D_F \leq \min\{X_K : K \subseteq F, e_K > 0\}$. Corollary 3.3.3 applied to each K implies that $\min\{X_K : K \subseteq F, e_K > 0\} = \Theta(\Phi_F)$ a.a.s., so $D_F = O(\Phi_F)$ a.a.s.

It now suffices to show that $D_F = \Omega(\Phi_F)$ a.a.s. To do so, for each $G \in \mathcal{G}_{n,d}^{(r)}$ we define an auxiliary graph $\Gamma = \Gamma(G)$ whose vertices are all the copies of F in G, and in which $F_1, F_2 \in V(\Gamma)$ are adjacent if and only if F_1 and F_2 share at least one edge. Let us denote by $\sum^*$ the sum over all graphs $\tilde{F}$ which can be written as $\tilde{F} = F' \cup F''$, where $F', F'' \subseteq K_V^{(r)}$, $F', F'' \cong F$ and $E(F') \cap E(F'') \neq \varnothing$. This means that $v_\Gamma = X_F$ and

$$e_\Gamma = \mathcal{O}\left(\sum\nolimits^* X_{\tilde{F}}\right).$$

Note that the size of the largest independent set in Γ equals D_F. By Turán's theorem we have that

$$D_F \geq \frac{X_F^2}{X_F + \mathcal{O}\left(\sum^* X_{\tilde{F}}\right)}.$$

Using corollary 3.3.3, one can check that, a.a.s.,

$$\frac{X_{\tilde{F}}^2}{X_F + \mathcal{O}\left(\sum^* X_{\tilde{F}}\right)} = \Omega(\Phi_F) \iff X_{\tilde{F}} = \mathcal{O}\left(n^{2v_F} p^{2e_F} \Phi_F^{-1}\right) \tag{A.6}$$

for all $\tilde{F} = F' \cup F''$ such that $e_{F' \cap F''} > 0$. So it suffices to prove the final bound in (A.6).

For any fixed r-graph K, let $\Psi_K := n^{v_K} p^{e_K}$. Note that if $K \subseteq F$, then $\Psi_K = \Theta(\mathbb{E}[X_K])$ (by corollary 3.3.1(ii)). Furthermore, for any two r-graphs K and L on a vertex set V,

$$\Psi_L \Psi_K = \Psi_{L\cup K} \Psi_{L \cap K}. \tag{A.7}$$

Consider two copies F' and F'' of F whose intersection has at least one edge. Let $\tilde{F} := F' \cup F''$ and $K := F' \cap F''$, so $e_K > 0$. Thus, by corollary 3.3.1(ii),

$$\mathbb{E}[X_{\tilde{F}}] = \Theta\left(n^{2v_F - v_K} p^{2e_F - e_K}\right) = \Theta\left(\frac{\Psi_F^2}{\Psi_K}\right) = O\left(\frac{\Psi_F^2}{\Phi_F}\right). \tag{A.8}$$

We now claim that

$$\Phi_{\tilde{F}} = \min\{\mathbb{E}[X_L] : L \subseteq \tilde{F}, e_L > 0\} = \Omega\left(\min\left\{\frac{\Phi_F^3}{\Psi_K^2}, \frac{n\Phi_F^2}{\Psi_K^2}\right\}\right). \tag{A.9}$$

Indeed, consider any r-graph $L \subseteq \tilde{F}$ with $e_L > 0$ and let $L' := L \cap F'$ and $L'' := L \cap F''$. Note that $L \cup K = (L' \cup K) \cup (L'' \cup K)$ and $K = (L' \cup K) \cap (L'' \cup K)$, so two applications of (A.7) yield

$$\Psi_L = \frac{\Psi_{L\cup K}\Psi_{L\cap K}}{\Psi_K} = \frac{\Psi_{L'\cup K}\Psi_{L''\cup K}\Psi_{L\cap K}}{\Psi_K^2}. \tag{A.10}$$

If $e_{L\cap K} > 0$, then the values of $\Psi_{L'\cup K}$, $\Psi_{L''\cup K}$ and $\Psi_{L\cap K}$ can be lower bounded by $\Omega(\Phi_F)$ (as $L' \cup K$, $L'' \cup K$ and $L \cap K$ are all subgraphs of F). Thus $\Psi_L = \Omega(\Phi_F^3/\Psi_K^2)$. So suppose that $e_{L\cap K} = 0$. Then L' and L'' are edge-disjoint, and at least one of them has at least one edge. We may assume that $e_{L'} > 0$ without loss of generality. Consider three cases. If $e_{L''} = 0$, then $\Psi_{L''} = n^{v_{L''}} \geq \Psi_{L'\cap L''}$ and, by (A.7), $\Psi_L = \Psi_{L'}\Psi_{L''}/\Psi_{L'\cap L''} = \Omega(\Psi_{L'}) = \Omega(\Phi_F) = \Omega(\Phi_F^3/\Psi_K^2)$, where the final equality holds since $\Psi_K = \Omega(\Phi_F)$. If $e_{L''} > 0$ but $L' \cap L'' = \varnothing$ then $\Psi_L =$

$\Psi_{L'}\Psi_{L''} = \Omega(\Psi_{L'}) = \Omega(\Phi_{\tilde{F}}^3/\Psi_K^2)$. Otherwise, we have that $e_{L''} > 0$ and $L' \cap L'' \neq \varnothing$. We use (A.10) taking into account that $\Psi_{L \cap K} = n^{v_{L \cap K}} = \Omega(n)$ to conclude that $\Psi_L = \Omega(n\Phi_{\tilde{F}}^2/\Psi_K^2)$. This proves the claim.

By lemma 3.3.2 we have that $\mathrm{Var}(X_{\tilde{F}}) = \mathbb{E}[X_{\tilde{F}}]^2\mathcal{O}(\varepsilon_{n,d} + \Phi_{\tilde{F}}^{-1})$. As $\varepsilon_{n,d} = o(1)$ by assumption, by (A.8) Chebyshev's inequality implies that the final bound in (A.6) holds a.a.s. if $\Phi_{\tilde{F}}^{-1} = \mathcal{O}(\varepsilon_{n,d})$. Therefore, we may assume that $\mathrm{Var}(X_{\tilde{F}}) = \mathcal{O}(\mathbb{E}[X_{\tilde{F}}]^2/\Phi_{\tilde{F}}) = \mathcal{O}(\Psi_{\tilde{F}}^2/\Phi_{\tilde{F}})$. Consequently, by (A.9) and (A.7) we have

$$\mathrm{Var}(X_{\tilde{F}}) = \mathcal{O}\left(\frac{\Psi_{\tilde{F}}^2\Psi_K^2}{\Phi_F^3} + \frac{\Psi_{\tilde{F}}^2\Psi_K^2}{n\Phi_F^2}\right) = \mathcal{O}\left(\frac{\Psi_F^4}{\Phi_F^3} + \frac{\Psi_F^4}{n\Phi_F^2}\right).$$

Thus, Chebyshev's inequality gives

$$\mathbb{P}\left[X_{\tilde{F}} \geq \mathbb{E}[X_{\tilde{F}}] + \frac{\Psi_F^2}{\Phi_F}\right] = \mathcal{O}\left(\Phi_F^{-1} + 1/n\right) = o(1)$$

by assumption. Hence $X_{\tilde{F}} = O\left(\Psi_F^2/\Phi_F\right)$ a.a.s., as required. □

BIBLIOGRAPHY

[1] M. Ajtai, J. Komlós, and E. Szemerédi. “First occurrence of Hamilton cycles in random graphs”. In: *Cycles in graphs (Burnaby, B.C., 1982)*. Vol. 115. North-Holland Math. Stud. North-Holland, Amsterdam, 1985, pp. 173–178. DOI: 10.1016/S0304-0208(08)73007-X.

[2] N. Alon, E. Fischer, I. Newman, and A. Shapira. “A combinatorial characterization of the testable graph properties: it’s all about regularity”. *SIAM J. Comput.* 39.1 (2009), pp. 143–167. ISSN: 0097-5397. DOI: 10.1137/060667177.

[3] N. Alon, T. Kaufman, M. Krivelevich, and D. Ron. “Testing triangle-freeness in general graphs”. *SIAM J. Discrete Math.* 22 (2008), pp. 786–819. ISSN: 0895-4801. DOI: 10.1137/07067917X.

[4] N. Alon, J.-H. Kim, and J. Spencer. “Nearly perfect matchings in regular simple hypergraphs”. *Israel J. Math.* 100 (1997), pp. 171–187. ISSN: 0021-2172. DOI: 10.1007/BF02773639.

[5] N. Alon and J. H. Spencer. *The probabilistic method*. Fourth. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, 2016, pp. xiv+375. ISBN: 978-1-119-06195-3.

[6] Y. Alon and M. Krivelevich. “Hitting time of edge disjoint Hamilton cycles in random subgraph processes on dense base graphs”. *arXiv e-prints* (2019). arXiv: 1912.01251.

[7] D. Altman, C. Greenhill, M. Isaev, and R. Ramadurai. “A threshold result for loose Hamiltonicity in random regular uniform hypergraphs”. *J. Combin. Theory Ser. B* 142 (2020), pp. 307–373. ISSN: 0095-8956. DOI: 10.1016/j.jctb.2019.11.001.

[8] K. Azuma. “Weighted sums of certain dependent random variables”. *Tôhoku Math. J. (2)* 19 (1967), pp. 357–367. ISSN: 0040-8735. DOI: 10.2748/tmj/1178243286.

[9] J. Balogh, A. Treglown, and A. Z. Wagner. “Tilings in randomly perturbed dense graphs”. *Combin. Probab. Comput.* 28.2 (2019), pp. 159–176. ISSN: 0963-5483. DOI: 10.1017/S0963548318000366.

[10] S. Ben-Shimon, M. Krivelevich, and B. Sudakov. "Local resilience and Hamiltonicity maker-breaker games in random regular graphs". *Combin. Probab. Comput.* 20.2 (2011), pp. 173–211. ISSN: 0963-5483. DOI: 10.1017/S0963548310000453.

[11] S. Ben-Shimon, M. Krivelevich, and B. Sudakov. "On the resilience of Hamiltonicity and optimal packing of Hamilton cycles in random graphs". *SIAM J. Discrete Math.* 25.3 (2011), pp. 1176–1193. ISSN: 0895-4801. DOI: 10.1137/110821299.

[12] I. Benjamini, O. Schramm, and A. Shapira. "Every minor-closed property of sparse graphs is testable". *Adv. Math.* 223.6 (2010), pp. 2200–2218. ISSN: 0001-8708. DOI: 10.1016/j.aim.2009.10.018.

[13] S. Bhatt and J.-Y. Cai. "Take a walk, grow a tree". In: *29th Annual Symposium on Foundations of Computer Science (STOC).* 1988, pp. 469–478.

[14] S. N. Bhatt, F. R. K. Chung, F. T. Leighton, and A. L. Rosenberg. "Efficient embeddings of trees in hypercubes". *SIAM J. Comput.* 21 (1992), pp. 151–162. ISSN: 0097-5397. DOI: 10.1137/0221012.

[15] T. Bohman, A. Frieze, and R. Martin. "How many random edges make a dense graph Hamiltonian?" *Random Structures Algorithms* 22.1 (2003), pp. 33–42. ISSN: 1042-9832. DOI: 10.1002/rsa.10070.

[16] B. Bollobás. "A probabilistic proof of an asymptotic formula for the number of labelled regular graphs". *European J. Combin.* 1.4 (1980), pp. 311–316. ISSN: 0195-6698. DOI: 10.1016/S0195-6698(80)80030-8.

[17] B. Bollobás. "The evolution of the cube". In: *Combinatorial mathematics (Marseille-Luminy, 1981).* Vol. 75. North-Holland Math. Stud. North-Holland, Amsterdam, 1983, pp. 91–97.

[18] B. Bollobás. "The evolution of sparse graphs". In: *Graph theory and combinatorics (Cambridge, 1983).* Academic Press, London, 1984, pp. 35–57.

[19] B. Bollobás. "Complete matchings in random subgraphs of the cube". *Random Structures Algorithms* 1.1 (1990), pp. 95–104. ISSN: 1042-9832. DOI: 10.1002/rsa.3240010107.

[20] B. Bollobás. Personal communication. 2020.

[21] B. Bollobás and A. M. Frieze. "On matchings and Hamiltonian cycles in random graphs". In: *Random graphs '83 (Poznań, 1983).* Vol. 118. North-Holland Math. Stud. North-Holland, Amsterdam, 1985, pp. 23–46.

[22] B. Bollobás, Y. Kohayakawa, and T. Łuczak. "The evolution of random subgraphs of the cube". *Random Structures Algorithms* 3.1 (1992), pp. 55–90. ISSN: 1042-9832. DOI: 10.1002/rsa.3240030106.

[23] B. Bollobás and A. Thomason. "Random graphs of small order". In: *Random graphs '83 (Poznań, 1983)*. Vol. 118. North-Holland Math. Stud. North-Holland, Amsterdam, 1985, pp. 47–97.

[24] B. Bollobás and A. Thomason. "Threshold functions". *Combinatorica* 7.1 (1987), pp. 35–38. ISSN: 0209-9683. DOI: 10.1007/BF02579198.

[25] C. Borgs, J. T. Chayes, R. van der Hofstad, G. Slade, and J. Spencer. "Random subgraphs of finite graphs. III. The phase transition for the n-cube". *Combinatorica* 26.4 (2006), pp. 395–410. ISSN: 0209-9683. DOI: 10.1007/s00493-006-0022-1.

[26] J. Böttcher, J. Han, Y. Kohayakawa, R. Montgomery, O. Parczyk, and Y. Person. "Universality for bounded degree spanning trees in randomly perturbed graphs". *Random Structures Algorithms* 55.4 (2019), pp. 854–864. ISSN: 1042-9832. DOI: 10.1002/rsa.20850.

[27] J. Böttcher, R. Montgomery, O. Parczyk, and Y. Person. "Embedding spanning bounded degree graphs in randomly perturbed graphs". *Mathematika* 66 (2020), pp. 422–447. ISSN: 2041-7942. DOI: 10.1112/mtk.12005.

[28] J. D. Burtin. "The probability of connectedness of a random subgraph of an n-dimensional cube". *Problemy Peredači Informacii* 13.2 (1977), pp. 90–95. ISSN: 0555-2923.

[29] R. Caha and V. Koubek. "Spanning multi-paths in hypercubes". *Discrete Math.* 307 (2007), pp. 2053–2066. ISSN: 0012-365X. DOI: 10.1016/j.disc.2005.12.050.

[30] M. Y. Chan and S.-J. Lee. "On the existence of Hamiltonian circuits in faulty hypercubes". *SIAM J. Discrete Math.* 4 (1991), pp. 511–527. ISSN: 0895-4801. DOI: 10.1137/0404045.

[31] X.-B. Chen. "Paired many-to-many disjoint path covers of the hypercubes". *Inform. Sci.* 236 (2013), pp. 218–223. ISSN: 0020-0255. DOI: 10.1016/j.ins.2013.02.028.

[32] V. Chvátal. "On Hamilton's ideals". *J. Combinatorial Theory Ser. B* 12 (1972), pp. 163–168. ISSN: 0095-8956. DOI: 10.1016/0095-8956(72)90020-2.

[33] P. Condon, A. Espuny Díaz, A. Girão, D. Kühn, and D. Osthus. "Dirac's theorem for random regular graphs". *Combin. Probab. Comput.* (2020). DOI: 10.1017/S0963548320000346.

[34] P. Condon, A. Espuny Díaz, A. Girão, D. Kühn, and D. Osthus. "Hamiltonicity of random subgraphs of the hypercube". *arXiv e-prints* (2020). arXiv: 2007.02891.

[35] P. Condon, A. Espuny Díaz, J. Kim, D. Kühn, and D. Osthus. "Resilient degree sequences with respect to Hamilton cycles and matchings in random graphs". *Electron. J. Combin.* 26.4 (2019), Paper No. 4.54, 22. DOI: 10.37236/8279.

[36] N. Cook, L. Goldstein, and T. Johnson. "Size biased couplings and the spectral gap for random regular graphs". *Ann. Probab.* 46.1 (2018), pp. 72–125. ISSN: 0091-1798. DOI: 10.1214/17-AOP1180.

[37] C. Cooper, A. Frieze, and B. Reed. "Random regular graphs of non-constant degree: connectivity and Hamiltonicity". *Combin. Probab. Comput.* 11.3 (2002), pp. 249–261. ISSN: 0963-5483. DOI: 10.1017/S0963548301005090.

[38] G. A. Dirac. "Some theorems on abstract graphs". *Proc. London Math. Soc. (3)* 2 (1952), pp. 69–81. ISSN: 0024-6115. DOI: 10.1112/plms/s3-2.1.69.

[39] A. Dudek and A. Frieze. "Loose Hamilton cycles in random uniform hypergraphs". *Electron. J. Combin.* 18.1 (2011), Paper 48, 14. DOI: 10.37236/535.

[40] A. Dudek and A. Frieze. "Tight Hamilton cycles in random uniform hypergraphs". *Random Structures Algorithms* 42.3 (2013), pp. 374–385. ISSN: 1042-9832. DOI: 10.1002/rsa.20404.

[41] A. Dudek, A. Frieze, P.-S. Loh, and S. Speiss. "Optimal divisibility conditions for loose Hamilton cycles in random hypergraphs". *Electron. J. Combin.* 19.4 (2012), Paper 44, 17. DOI: 10.37236/2523.

[42] A. Dudek, A. Frieze, A. Ruciński, and M. Šileikis. "Loose Hamilton cycles in regular hypergraphs". *Combin. Probab. Comput.* 24.1 (2015), pp. 179–194. ISSN: 0963-5483. DOI: 10.1017/S0963548314000406.

[43] A. Dudek, A. Frieze, A. Ruciński, and M. Šileikis. "Embedding the Erdős-Rényi hypergraph into the random regular hypergraph and Hamiltonicity". *J. Combin. Theory Ser. B* 122 (2017), pp. 719–740. ISSN: 0095-8956. DOI: 10.1016/j.jctb.2016.09.003.

[44] T. Dvořák and P. Gregor. "Partitions of faulty hypercubes into paths with prescribed endvertices". *SIAM J. Discrete Math.* 22 (2008), pp. 1448–1461. ISSN: 0895-4801. DOI: 10.1137/060678476.

[45] T. Dvořák, P. Gregor, and V. Koubek. "Generalized Gray codes with prescribed ends". *Theoret. Comput. Sci.* 668 (2017), pp. 70–94. ISSN: 0304-3975. DOI: 10.1016/j.tcs.2017.01.010.

[46] M. E. Dyer, A. M. Frieze, and L. R. Foulds. "On the strength of connectivity of random subgraphs of the n-cube". In: *Random graphs '85 (Poznań, 1985)*. Vol. 144. North-Holland Math. Stud. North-Holland, Amsterdam, 1987, pp. 17–40.

[47] P. Erdős and A. Rényi. "On random graphs. I". *Publ. Math. Debrecen* 6 (1959), pp. 290–297. ISSN: 0033-3883.

[48] P. Erdős and A. Rényi. "On the evolution of random graphs". *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 5 (1960), pp. 17–61. ISSN: 0541-9514.

[49] P. Erdős and A. Rényi. "On the existence of a factor of degree one of a connected random graph". *Acta Math. Acad. Sci. Hungar.* 17 (1966), pp. 359–368. ISSN: 0001-5954. DOI: 10.1007/BF01894879.

[50] P. Erdős and J. Spencer. "Evolution of the n-cube". *Comput. Math. Appl.* 5.1 (1979), pp. 33–39. ISSN: 0898-1221. DOI: 10.1016/0898-1221(81)90137-1.

[51] A. Espuny Díaz, F. Joos, D. Kühn, and D. Osthus. "Edge correlations in random regular hypergraphs and applications to subgraph testing". *SIAM J. Discrete Math.* 33.4 (2019), pp. 1837–1863. ISSN: 0895-4801. DOI: 10.1137/18M1177159.

[52] J. A. Fill and R. Pemantle. "Percolation, first-passage percolation and covering times for Richardson's model on the n-cube". *Ann. Appl. Probab.* 3 (1993), pp. 593–629. ISSN: 1050-5164.

[53] A. Frieze. "Loose Hamilton cycles in random 3-uniform hypergraphs". *Electron. J. Combin.* 17.1 (2010), Note 28, 4. DOI: 10.37236/477.

[54] A. Frieze. "Random structures and algorithms". In: *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. 1*. Kyung Moon Sa, Seoul, 2014, pp. 311–340.

[55] A. Frieze. "Hamilton cycles in random graphs: a bibliography". *arXiv e-prints* (2019). arXiv: 1901.07139.

[56] A. Frieze and M. Krivelevich. "On two Hamilton cycle problems in random graphs". *Israel J. Math.* 166 (2008), pp. 221–234. ISSN: 0021-2172. DOI: 10.1007/s11856-008-1028-8.

[57] R. Glebov and M. Krivelevich. "On the number of Hamilton cycles in sparse random graphs". *SIAM J. Discrete Math.* 27.1 (2013), pp. 27–42. ISSN: 0895-4801. DOI: 10.1137/120884316.

[58] R. Glebov, Y. Person, and W. Weps. "On extremal hypergraphs for Hamiltonian cycles". *European J. Combin.* 33.4 (2012), pp. 544–555. ISSN: 0195-6698. DOI: 10.1016/j.ejc.2011.10.003.

[59] O. Goldreich, S. Goldwasser, and D. Ron. "Property testing and its connection to learning and approximation". *J. ACM* 45.4 (1998), pp. 653–750. ISSN: 0004-5411. DOI: 10.1145/285055.285060.

[60] R. J. Gould. "Updating the Hamiltonian problem—a survey". *J. Graph Theory* 15.2 (1991), pp. 121–157. ISSN: 0364-9024. DOI: 10.1002/jgt.3190150204.

[61] R. J. Gould. “Advances on the Hamiltonian problem—a survey”. *Graphs Combin.* 19.1 (2003), pp. 7–52. ISSN: 0911-0119. DOI: 10.1007/s00373-002-0492-x.

[62] R. J. Gould. “Recent advances on the Hamiltonian problem: Survey III”. *Graphs Combin.* 30.1 (2014), pp. 1–46. ISSN: 0911-0119. DOI: 10.1007/s00373-013-1377-x.

[63] W. T. Gowers. “Hypergraph regularity and the multidimensional Szemerédi theorem”. *Ann. of Math. (2)* 166.3 (2007), pp. 897–946. ISSN: 0003-486X. DOI: 10.4007/annals.2007.166.897.

[64] P. Gregor and T. Dvořák. “Path partitions of hypercubes”. *Inform. Process. Lett.* 108 (2008), pp. 402–406. ISSN: 0020-0190. DOI: 10.1016/j.ipl.2008.07.015.

[65] M. Hahn-Klimroth, G. S. Maesaka, Y. Mogge, S. Mohr, and O. Parczyk. “Random perturbation of sparse graphs”. *arXiv e-prints* (2020). arXiv: 2004.04672.

[66] J. Han and Y. Zhao. “Hamiltonicity in randomly perturbed hypergraphs”. *J. Combin. Theory Ser. B* (2020). ISSN: 0095-8956. DOI: 10.1016/j.jctb.2019.12.005.

[67] W. Hoeffding. “Probability inequalities for sums of bounded random variables”. *J. Amer. Statist. Assoc.* 58 (1963), pp. 13–30. ISSN: 0162-1459.

[68] R. van der Hofstad and A. Nachmias. “Hypercube percolation”. *J. Eur. Math. Soc. (JEMS)* 19 (2017), pp. 725–814. ISSN: 1435-9855. DOI: 10.4171/JEMS/679.

[69] S. Janson. “The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph”. *Combin. Probab. Comput.* 3.1 (1994), pp. 97–126. ISSN: 0963-5483. DOI: 10.1017/S0963548300001012.

[70] S. Janson. “Random regular graphs: asymptotic distributions and contiguity”. *Combin. Probab. Comput.* 4.4 (1995), pp. 369–405. DOI: 10.1017/S0963548300001735.

[71] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000, pp. xii+333. ISBN: 0-471-17541-2. DOI: 10.1002/9781118032718.

[72] T. Johansson. “On Hamilton cycles in Erdős-Rényi subgraphs of large graphs”. *arXiv e-prints* (2018). arXiv: 1811.03501.

[73] F. Joos, J. Kim, D. Kühn, and D. Osthus. “A characterization of testable hypergraph properties”. *arXiv e-prints* (2017). arXiv: 1707.03303.

[74] R. M. Karp. “Reducibility among combinatorial problems”. In: *Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972)*. Springer US, 1972, pp. 85–103. ISBN: 978-1-4684-2001-2. DOI: 10.1007/978-1-4684-2001-2_9.

[75] T. Kaufman, M. Krivelevich, and D. Ron. "Tight bounds for testing bipartiteness in general graphs". *SIAM J. Comput.* 33.6 (2004), pp. 1441–1483. ISSN: 0097-5397. DOI: 10.1137/S0097539703436424.

[76] J. H. Kim and V. H. Vu. "Sandwiching random graphs: universality between random graph models". *Adv. Math.* 188.2 (2004), pp. 444–469. ISSN: 0001-8708. DOI: 10.1016/j.aim.2003.10.007.

[77] J. H. Kim, B. Sudakov, and V. Vu. "Small subgraphs of random regular graphs". *Discrete Math.* 307 (2007), pp. 1961–1967. ISSN: 0012-365X. DOI: 10.1016/j.disc.2006.09.032.

[78] F. Knox, D. Kühn, and D. Osthus. "Edge-disjoint Hamilton cycles in random graphs". *Random Structures Algorithms* 46.3 (2015), pp. 397–445. ISSN: 1042-9832. DOI: 10.1002/rsa.20510.

[79] D. E. Knuth. *The Art of Computer Programming, Volume 4, Fascicle 2: Generating All Tuples and Permutations (Art of Computer Programming)*. Addison-Wesley Professional, 2005. ISBN: 0201853930.

[80] Y. Kohayakawa, B. Kreuter, and D. Osthus. "The length of random subsets of Boolean lattices". *Random Structures Algorithms* 16 (2000), pp. 177–194. ISSN: 1042-9832. DOI: 10.1002/(SICI)1098-2418(200003)16:2<177::AID-RSA4>3.0.CO;2-9.

[81] J. Komlós and E. Szemerédi. "Limit distribution for the existence of Hamiltonian cycles in a random graph". *Discrete Math.* 43.1 (1983), pp. 55–63. ISSN: 0012-365X. DOI: 10.1016/0012-365X(83)90021-3.

[82] A. D. Koršunov. "Solution of a problem of P. Erdős and A. Rényi on Hamiltonian cycles in undirected graphs". *Dokl. Akad. Nauk SSSR* 228.3 (1976), pp. 529–532. ISSN: 0002-3264.

[83] A. D. Koršunov. "Solution of a problem of P. Erdős and A. Rényi on Hamiltonian cycles in undirected graphs". *Metody Diskretn. Anal.* 31 (1977), pp. 17–56. ISSN: 0136-1228.

[84] B. Kreuter. "Threshold functions for asymmetric Ramsey properties with respect to vertex colorings". *Random Structures Algorithms* 9.3 (1996), pp. 335–348. ISSN: 1042-9832. DOI: 10.1002/(SICI)1098-2418(199610)9:3<335::AID-RSA5>3.3.CO;2-0.

[85] M. Krivelevich. "On the number of Hamilton cycles in pseudo-random graphs". *Electron. J. Combin.* 19.1 (2012), Paper 25, 14. DOI: 10.37236/1177.

[86] M. Krivelevich, M. Kwan, and B. Sudakov. "Bounded-degree spanning trees in randomly perturbed graphs". *SIAM J. Discrete Math.* 31.1 (2017), pp. 155–171. ISSN: 0895-4801. DOI: 10.1137/15M1032910.

[87] M. Krivelevich, C. Lee, and B. Sudakov. “Robust Hamiltonicity of Dirac graphs”. *Trans. Amer. Math. Soc.* 366.6 (2014), pp. 3095–3130. ISSN: 0002-9947. DOI: 10.1090/S0002-9947-2014-05963-1.

[88] M. Krivelevich and W. Samotij. “Optimal packings of Hamilton cycles in sparse random graphs”. *SIAM J. Discrete Math.* 26.3 (2012), pp. 964–982. ISSN: 0895-4801. DOI: 10.1137/110849171.

[89] M. Krivelevich, B. Sudakov, V. H. Vu, and N. C. Wormald. “Random regular graphs of high degree”. *Random Structures Algorithms* 18.4 (2001), pp. 346–363. ISSN: 1042-9832. DOI: 10.1002/rsa.1013.

[90] D. Kühn and D. Osthus. “A survey on Hamilton cycles in directed graphs”. *European J. Combin.* 33.5 (2012), pp. 750–766. ISSN: 0195-6698. DOI: 10.1016/j.ejc.2011.09.030.

[91] D. Kühn and D. Osthus. “Hamilton cycles in graphs and hypergraphs: an extremal perspective”. In: *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV*. Kyung Moon Sa, Seoul, 2014, pp. 381–406.

[92] D. Kühn and D. Osthus. “Hamilton decompositions of regular expanders: applications”. *J. Combin. Theory Ser. B* 104 (2014), pp. 1–27. ISSN: 0095-8956. DOI: 10.1016/j.jctb.2013.10.006.

[93] C. Lee and B. Sudakov. “Dirac’s theorem for random graphs”. *Random Structures Algorithms* 41.3 (2012), pp. 293–305. ISSN: 1042-9832. DOI: 10.1002/rsa.20419.

[94] F. T. Leighton. *Introduction to parallel algorithms and architectures*. Arrays, trees, hypercubes. Morgan Kaufmann, San Mateo, CA, 1992, pp. xx+831. ISBN: 1-55860-117-1.

[95] C. McDiarmid, A. Scott, and P. Withers. “The component structure of dense random subgraphs of the hypercube”. *arXiv e-prints* (2018). arXiv: 1806.06433.

[96] B. D. McKay and N. C. Wormald. “Uniform generation of random regular graphs of moderate degree”. *J. Algorithms* 11.1 (1990), pp. 52–67. ISSN: 0196-6774. DOI: 10.1016/0196-6774(90)90029-E.

[97] R. Montgomery. “Hamiltonicity in random graphs is born resilient”. *J. Combin. Theory Ser. B* 139 (2019), pp. 316–341. ISSN: 0095-8956. DOI: 10.1016/j.jctb.2019.04.002.

[98] R. Nenadov, A. Steger, and M. Trujić. “Resilience of perfect matchings and Hamiltonicity in random graph processes”. *Random Structures Algorithms* 54.4 (2019), pp. 797–819. ISSN: 1042-9832. DOI: 10.1002/rsa.20827.

[99] I. Newman and C. Sohler. “Every property of hyperfinite graphs is testable”. *SIAM J. Comput.* 42.3 (2013), pp. 1095–1112. ISSN: 0097-5397. DOI: 10.1137/120890946.

[100] L. Pósa. "Hamiltonian circuits in random graphs". *Discrete Math.* 14.4 (1976), pp. 359–364. ISSN: 0012-365X. DOI: 10.1016/0012-365X(76)90068-6.

[101] L. Pósa. "A theorem concerning Hamilton lines". *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 7 (1962), pp. 225–226. ISSN: 0541-9514.

[102] O. Riordan. "Spanning subgraphs of random graphs". *Combin. Probab. Comput.* 9.2 (2000), pp. 125–148. ISSN: 0963-5483. DOI: 10.1017/S0963548399004150.

[103] R. W. Robinson and N. C. Wormald. "Almost all cubic graphs are Hamiltonian". *Random Structures Algorithms* 3.2 (1992), pp. 117–125. ISSN: 1042-9832. DOI: 10.1002/rsa.3240030202.

[104] R. W. Robinson and N. C. Wormald. "Almost all regular graphs are Hamiltonian". *Random Structures Algorithms* 5.2 (1994), pp. 363–374. ISSN: 1042-9832. DOI: 10.1002/rsa.3240050209.

[105] V. Rödl and M. Schacht. "Property testing in hypergraphs and the removal lemma [extended abstract]". In: *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*. ACM, New York, 2007, pp. 488–495. DOI: 10.1145/1250790.1250862.

[106] V. Rödl and M. Schacht. "Regular partitions of hypergraphs: counting lemmas". *Combin. Probab. Comput.* 16.6 (2007), pp. 887–901. ISSN: 0963-5483. DOI: 10.1017/s0963548307008565.

[107] V. Rödl and M. Schacht. "Regular partitions of hypergraphs: regularity lemmas". *Combin. Probab. Comput.* 16.6 (2007), pp. 833–885. ISSN: 0963-5483. DOI: 10.1017/s0963548307008553.

[108] V. Rödl and J. Skokan. "Regularity lemma for k-uniform hypergraphs". *Random Structures Algorithms* 25.1 (2004), pp. 1–42. ISSN: 1042-9832. DOI: 10.1002/rsa.20017.

[109] C. Savage. "A survey of combinatorial Gray codes". *SIAM Rev.* 39 (1997), pp. 605–629. ISSN: 0036-1445. DOI: 10.1137/S0036144595295272.

[110] B. Sudakov and V. H. Vu. "Local resilience of graphs". *Random Structures Algorithms* 33.4 (2008), pp. 409–433. ISSN: 1042-9832. DOI: 10.1002/rsa.20235.

[111] K. Tikhomirov and P. Youssef. "The spectral gap of dense random regular graphs". *Ann. Probab.* 47.1 (2019), pp. 362–419. ISSN: 0091-1798. DOI: 10.1214/18-AOP1263.