Hamilton cycles and algorithms

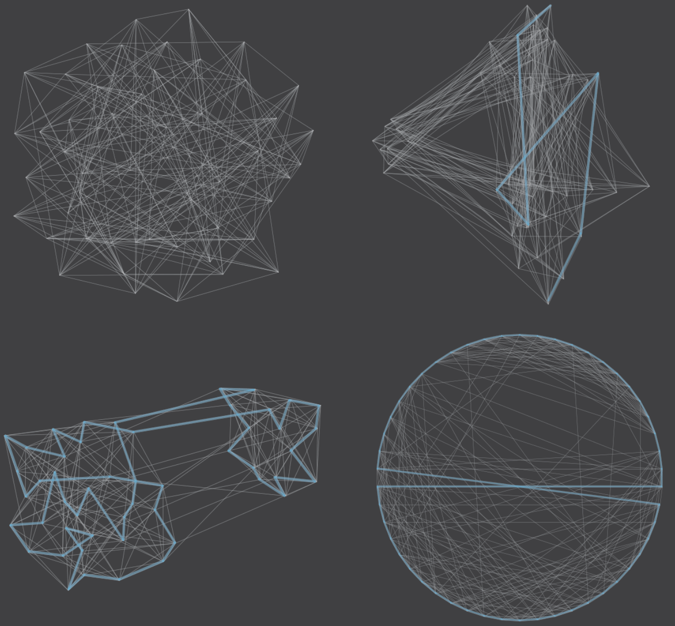
Stroh, F.J.M.

Publication date
2022

Document Version
Final published version

Hamilton Cycles and Algorithms

Fabian Stroh



Hamilton Cycles and Algorithms

Fabian Stroh

Hamilton cycles and algorithms

Fabian Stroh

Cover design: Anna Bleeker, persoonlijkproefschrift.nl
Layout: Fabian Stroh
Typesetting: LaTeX using the `classicthesis` template.
Printed by: Proefschriften.nl, Deventer

This thesis is produced on FSC®-certified materials.

The research for this doctoral thesis received financial assistance from the Netherlands Organisation for Scientific Research (NWO) with TOP grant 613.001.601. The work was carried out at the Korteweg-de Vries Institute for Mathematics (KdVI) at the University of Amsterdam.

Hamilton cycles and algorithms

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gesag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 6 april 2022, te 13:00 uur

door Fabian Jonas Michael Stroh
geboren te Speyer

Promotiecommissie

Promotor:	prof. dr. J.A. Ellis-Monaghan	Universiteit van Amsterdam
Copromotor:	dr. V.S. Patel	Universiteit van Amsterdam
Overige leden:	prof. dr. ir. H.J. Broersma	University of Twente
	prof. dr. M.R.H. Mandjes	Universiteit van Amsterdam
	prof. dr. J.V. Stokman	Universiteit van Amsterdam
	prof. dr. A. Schrijver	Universiteit van Amsterdam
	dr. G. Regts	Universiteit van Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

CONTENTS

1 INTRODUCTION

1.1 A SHORT INTRODUCTION FOR THE LAYPERSON

This thesis is concerned with the mathematical field of graph theory. Graphs can be used to model many different situations or represent different types of networks (e.g. transport networks, social networks, communications networks). They consist of a collection of vertices and edges, where an edge is a connection between two vertices. Sometimes the edges are directed and sometimes they are weighted.

The main theme in this thesis is Hamilton cycles. A *cycle* is a cyclic sequence of non-repeating vertices where any two successive vertices are connected by an edge. See Figure 1 for an example. A *Hamilton cycle* in a graph is a cycle that contains all vertices of the graph. Hamilton cycles are one of the simplest, most natural spanning structures, that is, structures that contain every vertex. Therefore, understanding Hamilton cycles can help in understanding more complicated spanning structures. Another reason to study Hamilton cycles is their connection to the famous traveling salesman problem, which we now describe.

Imagine you are a delivery driver, and you have a certain number of deliveries to make in your area and must then return to your starting point. You know all the places you need to visit, and your cell phone can tell you how long it takes to drive between any two delivery addresses, but it is not clear in which order you should visit your destinations in order to

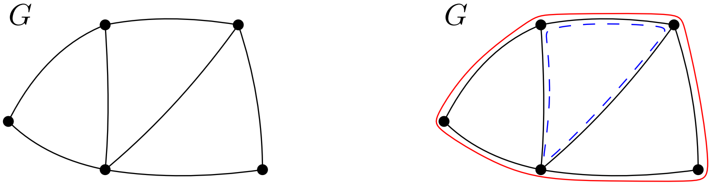


Figure 1: Left: An example of a graph G. Right: two cycles in G: one cycle is indicated by dashed blue lines, the other by a solid red line. The red cycle is a Hamilton cycle, the blue cycle is not.

complete your deliveries as quickly as possible. This problem is known as the traveling salesman problem or TSP[1], and has been widely studied [14].

In order to view this through the lens of graph theory, we take the destinations and our starting point as vertices, and connect every pair of vertices with an edge that is weighted according to the travel time between those two points. In order to find an optimal route, we can now look for a minimum weight Hamilton cycle in our graph. In other words, we want to find a way of 'traveling' along the edges of the graph such that we visit every vertex exactly once, we finish our route in the vertex we start at and we choose the edges we traverse so as to minimize their total weight.

Trying to work out an example by hand will quickly convince you that this is work best left to computers. So one is interested in algorithms for solving the traveling salesman problem as quickly as possible. An algorithm is a list of precise instructions that can be followed by computers. One important property of an algorithm is how quickly (i.e. with how many elementary steps) it completes its calculation. This is usually measured in terms of the size of the input data.

TSP belongs to the class of $\mathcal{NP}$-hard problems [49], which are problems believed to be computationally difficult. In fact, even the apparently easier problem of deciding whether an (unweighted) graph has a Hamilton cycle is $\mathcal{NP}$-complete [49]. In practice, this means that it is highly unlikely that there is an algorithm that, given any graph as input, is able to decide whether the graph has a Hamilton cycle efficiently.[2] This computational intractability is part of what gives the study of Hamilton cycles its richness.

In this thesis we consider three problems. They are quite different, but they are all unified by the theme of Hamilton cycles. One is motivated by algorithmically finding Hamilton cycles in graphs, one is motivated by counting Hamilton cycles in graphs, and one is related to Hamilton decompositions of graphs, i.e. partitioning the edges of a graph into Hamilton cycles.

1 Strictly speaking, what we describe here is known as *metric TSP*, a closely related variant.

2 'Efficiently' here means that the number of elementary steps needed by the algorithm (for any input graph on n vertices) can be bounded by a polynomial in n. Such an algorithm for the Hamilton cycle problem does not exist if, as is widely believed, $\mathcal{P} \neq \mathcal{NP}$.

1.2 BASIC NOTATION

In this section we fix some standard graph theory notation that will be used throughout.

A *graph* G is a tuple $G = (V, E)$ consisting of a set V of vertices and a set $E \subseteq \{\{x, y\} \mid x, y \in V, x \neq y\}$ of edges, where each edge is a pair of distinct vertices. We sometimes write $V(G)$ for the vertex set of G and $E(G)$ for its edge set. We denote an edge $e \in E$ that contains two vertices $v, w \in V$ as vw (rather than $\{v, w\}$); in this case we say v and w are *adjacent*. We say two edges are *incident* if they share a vertex. We call the number of vertices in a graph G the *order* of G and denote it by $|G|$.

A graph $H = (V', E')$ such that $V' \subseteq V$ and $E' \subseteq E$ is a *subgraph* of G, denoted $H \subseteq G$. We also say G contains H. A *cycle* is a graph with vertex set $\{v_1, \ldots, v_n\}$ and edge set $\{v_1v_2, v_2v_3, \ldots, v_{n-1}v_n, v_nv_1\}$; see Figure 1. We say that a graph G contains a *Hamilton cycle*, or is Hamiltonian, if it contains a cycle that contains all vertices in G. A graph G is *connected*, if for any two vertices $v, v' \in V(G)$ there is a sequence of vertices $v = v_0, v_1, \ldots, v_k = v'$ such that v_i and v_{i+1} are adjacent for $i = 0, \ldots, k-1$. The *degree* $d_G(v)$ of a vertex $v \in V(G)$ is the number of vertices adjacent to v, i.e. $d_G(v) := |\{u \in V \mid uv \in E\}|$. For a graph G we set $\delta(G) = \min_{v \in V} d_G(v)$ and $\Delta(G) = \max_{v \in V} d_G(v)$, called respectively the minimum and maximum degree of G. If every vertex in G has the same degree r, we say that G is regular, or r-regular.

A *directed graph*, or *digraph* is a tuple $D = (V, E)$ consisting again of a set V of vertices and a set E of *directed edges*. A directed edge is an ordered pair (x, y) of two different vertices $x, y \in V$ and we understand the edge to be directed from x to y. We set $d^+_G(v) = |\{w \mid (v, w) \in E(G)\}|$ as the *outdegree* of v and $d^-_G(v) = |\{w \mid (w, v) \in E(G)\}|$ as the *indegree* of v. For any graph theory definitions not mentioned here we refer the reader to e.g. Diestel [20].

We give Dirac's seminal theorem on Hamilton cycles, which will be referred to several times.

Theorem 1.2.1 ([21])**.** Every graph with $n \geq 3$ vertices and minimum degree at least $n/2$ has a Hamilton cycle.

1.3 OUTLINE OF CONTENTS

This thesis is based on the following works:

(1) Alberto Espuny Díaz, Viresh Patel, Fabian Stroh. Path decompositions of random directed graphs (2021). arXiv: `2109.13565`. In: *Extended Abstracts EuroComb 2021*, (2021), pp. 702–706. Submitted to *Random Structures & Algorithms.*

(2) V. Patel, F. Stroh. A polynomial-time algorithm to determine (almost) Hamiltonicity of dense regular graphs (2020). arXiv: `2007.14502`. To appear in *SIAM Journal on Discrete Mathematics.*

(3) P. Kleer, V. Patel, F. Stroh. Switch-based Markov chains for sampling Hamiltonian cycles in dense graphs. *Electronic Journal of Combinatorics* 27.4 (2020), Paper No. 4.29, 25.

Each of the authors contributed equally to each of the publications. Chapter 2 is based on (1), Chapter 3 is based on (2) and Chapter 4 is based on (3). The chapters are self-contained and can be read in any order. They each begin with an introduction to the problem, followed by the statement of the main results and some background and context. Then, we give preliminaries, followed by the proofs and a short concluding section. We conclude this chapter by giving a short overview of each of the main chapters.

Chapter 2: Path decompositions of random directed graphs

An area of extremal combinatorics that has seen a lot of activity both historically and recently is the study of decompositions of combinatorial structures. The prototypical question in this area asks whether, for some given class $\mathcal{C}$ of graphs, directed graphs, or hypergraphs, the edge set of each $H \in \mathcal{C}$ can be decomposed into parts satisfying some given property. The goal is usually to minimize the number of parts.

One classical decomposition problem concerns edge colorings of graphs. A proper edge coloring of a graph G is an assignment of colors to its edges such that incident edges receive different colors. Notice that the color classes form a partition of the edges into matchings (a matching being a set of edges in which no two edges are incident). The chromatic index of a graph G, denoted $\chi'(G)$, is the smallest number of colors needed in a proper edge coloring of G. Notice that $\chi'(G) \geq \Delta(G)$. The classical theorem of Vizing [75] asserts that $\chi'(G) \in \{\Delta(G), \Delta(G)+1\}$. This gives us a lot of

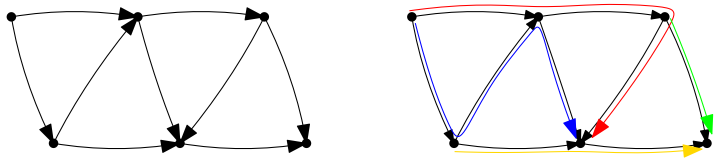


Figure 2: Left: An example of a digraph with excess 4. Right: One way to decompose the edges into four edge-disjoint paths.

information about optimal decompositions of graphs into matchings, but it is generally $\mathcal{NP}$-complete[3] to determine whether $\chi'(G) = \Delta(G)$ or $\chi'(G) = \Delta(G) + 1$ [39], so we should not expect a simple characterization.

However, *almost all* graphs achieve $\chi'(G) = \Delta(G)$. To explain what we mean by almost all we introduce random graphs. For $n \in \mathbb{N}$ and $p \in [0, 1]$, let $G_{n,p}$ be the random graph on n (labeled) vertices constructed as follows: we start with n isolated vertices, and for each of the $\binom{n}{2}$ possible edges, we include each edge with probability p and make these $\binom{n}{2}$ choices independently. So $G_{n,p}$ is a probability distribution on n-vertex graphs and is called the Erdös-Rényi random graph model. Note that $G_{n,1/2}$ is the uniformly random distribution on n-vertex graphs. There is a large literature on the properties of $G_{n,p}$, see [9, 44], even if we restrict ourselves to properties relating to graph decompositions, e.g. [13, 25, 29, 33, 37].

Returning to the chromatic index, Erdös and Wilson [25] showed that for $G = G_{n,1/2}$, it holds that $\mathbb{P}[\chi'(G) = \Delta(G)] \to 1$ as $n \to \infty$, i.e. almost all n-vertex graphs achieve the natural lower bound for the chromatic index as n increases. So, while the chromatic index is not generally easy to understand, we see that for most graphs, we know what value it takes. Our goal in Chapter 2 is to obtain a similar result for path decompositions of directed graphs.

In Chapter 2, we consider the problem of partitioning the edges of directed graphs D into as few directed paths as possible. The number of paths in such a partition is called the path number of D and is denoted by $pn(D)$. See Figure 2 for an example. Again, there is a natural lower bound for $pn(D)$ called the excess of D and denoted $\text{ex}(D)$. The excess, $\text{ex}(D)$, is

3 We will not define $\mathcal{NP}$-complete formally, but instead mention that we do not expect to find an efficient algorithm to solve these problems unless $\mathcal{P} = \mathcal{NP}$. Again, efficient here means the algorithm has a running time polynomial in the size of the input.

easy to compute: it is simply half of the sum over all vertices of the absolute difference of the in- and outdegrees, i.e.

$$\mathrm{ex}(D) = \frac{1}{2} \sum_{v \in V} |d^+(v) - d^-(v)|.$$

See the introduction of Chapter 2 to see why this is a lower bound. Similarly to $G_{n,p}$, we define $D_{n,p}$ as the random directed graph constructed by starting with n isolated vertices and randomly and independently adding directed edges with probability p. However, for $D_{n,p}$ there are $n(n-1)$ possible *directed* edges, up to two between each pair of vertices. The main result of Chapter 2 is as follows:

Theorem 2.1.2. Let $\log^4 n/n^{1/3} \leq p \leq 1 - \log^{5/2} n/n^{1/5}$. Then, $\mathbb{P}[\mathrm{ex}(D_{n,p}) = pn(D_{n,p})] \to 1$ as $n \to \infty$.

The bounds on p obtained are unlikely to be optimal (but include $p = 1/2$). This is discussed further in Chapter 2. So far, it is not obvious what the connection of this chapter is with Hamilton cycles. We go into more detail about this in the introduction of Chapter 2.

Chapter 3: Almost-Hamiltonicity in dense regular graphs

A basic problem in algorithmic graph theory is to decide whether a given input graph contains some desired subgraph. For example, there is a polynomial-time algorithm (with running time $O(n^3)$) to decide whether a graph has a triangle: simply check whether any triple of vertices forms a triangle or not. The problem becomes more difficult if we wish to detect a specific spanning subgraph. A perfect matching of an n-vertex graph (with n even) is a spanning matching, i.e. a collection of $n/2$ edges, no two of which are incident. Edmonds' perfect matching algorithm [24] is a polynomial-time algorithm to decide whether a given graph contains a perfect matching. Not all subgraphs are easy to detect, however. As mentioned earlier it is $\mathcal{NP}$-complete to decide whether a given graph contains a Hamilton cycle, and so we do not expect to find a polynomial-time algorithm to detect Hamilton cycles. Hamilton cycles are one of the simplest spanning structures that are $\mathcal{NP}$-complete to detect.

Yet, this problem is still an active and important area of research. One goal is to find graph classes and situations in which Hamilton cycles are guaranteed or easier to find. Dirac's theorem 1.2.1 gives us one such graph class. It is trivial to decide Hamiltonicity in graphs of minimum degree at least $n/2$ (such graphs are guaranteed to contain a Hamilton cycle), and the proof of Dirac's theorem supplies a straightforward polynomial-time algorithm for finding a Hamilton cycle in such graphs. There are many results beyond Dirac's Theorem that give conditions under which graphs have Hamilton cycles; see e.g. the surveys [34, 59]. Usually, such results translate into efficient algorithms to find Hamilton cycles.

Chapter 3 concerns the Hamiltonicity of regular graphs with linear degree. Given $\alpha \in (0, 1]$ let $\mathcal{G}_\alpha$ be the set of graphs G such that every vertex of G has degree exactly D and $D \geq \alpha n$, where $n = |G|$. The question is whether for each α there is a polynomial-time algorithm to decide whether graphs in $\mathcal{G}_\alpha$ have a Hamilton cycle. This is motivated by a question in extremal combinatorics, which we discuss in the introduction of Chapter 3. We cannot solve this question, but we can answer a closely related question affirmatively. Specifically, we replace Hamilton cycles with *almost Hamilton cycles*. Almost Hamilton cycles are cycles that contain all but a very small number of vertices of a graph. Given $\alpha \in (0, 1]$, we give a number $c(\alpha)$ and a polynomial-time algorithm that determines whether a graph in $\mathcal{G}_\alpha$ contains a cycle on all but a constant number $c = c(\alpha)$ of vertices. Further, we give a randomized polynomial-time algorithm to find such a cycle if it exists.

Note that the result cannot be improved in the sense that, if we allow irregular graphs (of linear minimum degree) it becomes $\mathcal{NP}$-complete to detect (almost) Hamilton cycles, and similarly if we allow regular graphs of arbitrary degree.

Chapter 4: Reconfiguration of Hamilton cycles under k-switches In reconfiguration problems, we study a collection of objects and their relationship under a reconfiguration operation transforming one object into another. Typically, the objects in question will be solutions to some combinatorial problem and the operation will usually correspond to some minor change. The most fundamental question is then, can any such object be transformed into any other, and if so, how many steps are needed? We may understand the objects as the vertices of the *reconfiguration graph* $\mathbb{G}$, and connect them by an edge if one arises from the other by our chosen operation. Then the questions of reconfiguration can be phrased as questions of the properties of $\mathbb{G}$. Is $\mathbb{G}$ connected? What is the diameter of $\mathbb{G}$ (i.e., what is the furthest any two objects are apart)? Given two objects, can we efficiently find a path from one to the other in $\mathbb{G}$?

An example where it is easy to show the reconfiguration graph is connected is the case of proper k-colorings of a graph G. In a proper k-coloring, each vertex is assigned one of k colors, such that no two adjacent vertices share a color. Note that if $k \geq \Delta(G) + 1$, we will always be able to find a proper k-coloring, e.g. by successively coloring each vertex with a color not yet used among its neighbors. Our reconfiguration operation in this case consists of changing the color of a single vertex such that the resulting coloring is also proper. In this example, we can see that the reconfiguration graph is connected if $k \geq \Delta(G) + 2$. To see this, we construct a path between two arbitrary colorings, i.e. we transform one coloring into another by successively recoloring single vertices. We can always change the color of any vertex, as there are always at least two colors not among its neighbors. We transform one coloring into another by handling the vertices in an arbitrary order. Each vertex v, in order, is recolored to its target color i by first recoloring all of v's neighbors that are currently colored i and then coloring v with color i. Note that we do not recolor v once it has received its target color.

Other examples of objects that have been studied in the context of reconfiguration include triangulations of planar graphs, independent sets and vertex covers. More on reconfiguration problems can be found in [65]. Mostly we are interested in graphical objects, and often the reconfiguration graph is very large. More specifically, if our underlying graph G has n vertices, the number of vertices in the reconfiguration graph is usually exponential in n.

In the first part of Chapter 4 we study the reconfiguration of Hamilton cycles of a graph. Our reconfiguration operation is the k-switch. Given a graph G and a Hamilton cycle H of G, we perform a k-switch by removing up to k edges from H and adding the same number of edges from G such that the resulting subgraph H' is a Hamilton cycle of G again. See Figure 3 for an example. The switch operation is one of the simplest reconfiguration operations for Hamilton cycles and is used e.g. in the k-opt heuristic for TSP [61]. One of our main results is as follows:

Theorem 4.1.1 Let G be a graph on n vertices with $\delta(G) \geq n/2 + 7$. Then the k-switch reconfiguration graph on Hamilton cycles of G is connected for $k \geq 10$.

We give examples to show that the minimum degree cannot be lowered much. We expect that the bound of 10 can be reduced, but we show that

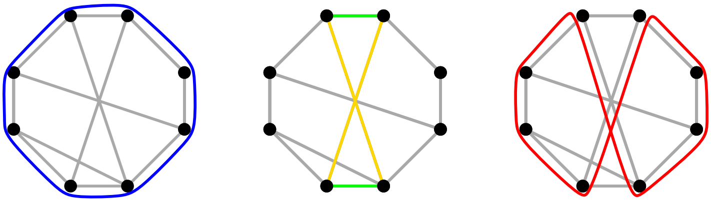


Figure 3: Left and right: A graph and two of its Hamilton cycles, indicated in blue and red. Center: four edges that can be used to transform the cycles into one another. We obtain the red Hamilton cycle from the blue Hamilton cycle by a 2-switch by removing the green edges and adding the yellow edges.

it is not possible to replace 10 with 2 without significantly increasing the minimum degree bound. These examples are found in Subsection 4.1.4.

One of the motivations to study reconfiguration of Hamilton cycles is in an application to computational counting and sampling. In computational counting, one is interested in algorithms for (approximately) computing the number of solutions of a combinatorial problem. A closely related problem is that of sampling a uniformly random solution. A powerful method to achieve this is to set up a suitable Markov chain on the reconfiguration graph so that its stationary distribution is the uniform distribution on the vertices of the reconfiguration graph. If such a Markov chain converges quickly to its stationary distribution (this is known as rapidly mixing, defined in Section 4.2.1), then we have a means to quickly sample an (approximate) uniformly random solution. This can often be used to approximate the number of solutions that we wish to count; we give the informal argument on how to do this in Section 4.2.2.

One of the applications of our reconfiguration result is to show that the natural Markov chain that arises from the Hamilton cycle reconfiguration under k-switches is rapidly mixing for the class of dense monotone graphs. We postpone the statement of this result to Chapter 4. This rapid mixing result can be used to give an efficient approximate algorithm that samples and counts Hamilton cycles in such graphs.

2 PATH DECOMPOSITIONS OF RANDOM DIRECTED GRAPHS

2.1 INTRODUCTION

Let D be a directed graph (or digraph for short) with vertex set $V(D)$ and edge set $E(D)$. A path decomposition of D is a collection of directed paths $P_1, \ldots, P_k$ of D whose edge sets $E(P_1), \ldots, E(P_k)$ partition $E(D)$. Given any directed graph D, the minimum number of paths in a path decomposition of D is called the *path number* of D and is denoted $\mathrm{pn}(D)$. A natural lower bound on $\mathrm{pn}(D)$ is obtained by examining the degree sequence of D. For each vertex $v \in V(D)$, write $d_D^+(v)$ (resp. $d_D^-(v)$) for the number of edges exiting (resp. entering) v. The *excess* at vertex v is defined to be $\mathrm{ex}_D(v) := d_D^+(v) - d_D^-(v)$. We note that, in any path decomposition of D, at least $|\mathrm{ex}_D(v)|$ paths must start (resp. end) at v if $\mathrm{ex}_D(v) \geq 0$ (resp. $\mathrm{ex}_D(v) \leq 0$). Therefore, we have

$$\mathrm{pn}(D) \geq \mathrm{ex}(D) := \frac{1}{2} \sum_{v \in V(D)} |\mathrm{ex}_D(v)|,$$

where $\mathrm{ex}(D)$ is called the *excess* of D. Any digraph for which equality holds above is called *consistent*. Clearly, not every digraph is consistent; in particular, any Eulerian digraph D has excess 0 and so cannot be consistent.

For the class of tournaments (that is, orientations of the complete graph), Alspach, Mason, and Pullman [3] conjectured that every tournament with an even number of vertices is consistent. Tournaments with an odd number of vertices may be regular and so have excess 0.

Conjecture 2.1.1. Every tournament T with an even number of vertices is consistent.

Many cases of this conjecture were resolved by Lo, Patel, Skokan, and Talbot [62], and the conjecture has very recently been completely resolved (for sufficiently large tournaments) by Girão, Granet, Kühn, Lo, and Osthus [32]. Both results relied on the robust expanders technique, developed by Kühn and Osthus with several coauthors, which has been instrumental in resolving several conjectures about edge decompositions of graphs and directed graphs; see, e.g., [17, 57, 58].

The conjecture seems likely to hold for many digraphs other than tournaments: indeed, the conjecture was stated only for even tournaments probably because it considerably generalized the following conjecture of Kelly, which was wide open at the time. Kelly's conjecture states that every regular tournament has a decomposition into Hamilton cycles (see [64]). We briefly describe how Kelly's conjecture follows from Conjecture 2.1.1. Given a regular tournament T, delete an arbitrary vertex v and its incident edges from T to obtain the subtournament $T - v$. As regular tournaments have an odd number of vertices, this yields an even tournament, which is consistent if Conjecture 2.1.1 holds. The paths in the path decomposition of $T - v$ can then be completed to Hamilton cycles in T by including v. The solution of Kelly's conjecture for sufficiently large tournaments was one of the first applications of the robust expanders technique [57].

A natural question then arises from Conjecture 2.1.1: which directed graphs are consistent? It is $\mathcal{NP}$-complete to determine whether a digraph is consistent [76], and so we should not expect to have a simple characterization of consistent digraphs. Nonetheless, here we begin to address this question by showing that the large majority of digraphs are consistent.

We consider the random digraph $D_{n,p}$. This is constructed by taking n isolated vertices and inserting each of the $n(n-1)$ possible directed edges independently with probability p. Typically statements about $D_{n,p}$ claim that, perhaps for some bounds on p, some property $\mathcal{P}$ holds for $D_{n,p}$ *asymptotically almost surely* (a.a.s.), which means that $\mathbb{P}[\mathcal{P} \text{ holds for } D_{n,p}] \to 1$ for $n \to \infty$. Our main result is the following theorem.

Theorem 2.1.2. Let $\log^4 n/n^{1/3} \leq p \leq 1 - \log^{5/2} n/n^{1/5}$. Then, a.a.s. $D_{n,p}$ is consistent.

Notice that some upper bound on p, as in the above theorem, is necessary because, when $p = 1$, we have that $\text{ex}(D_{n,p}) = 0$ (with probability 1) and so $D_{n,p}$ cannot be consistent. Moreover the property of being consistent is not a monotone property, that is, adding edges to a consistent digraph does not imply the resulting digraph is consistent. Therefore, unlike many other properties (see [10]), we should not necessarily expect a threshold for the consistency of random digraphs. We believe that the theorem holds for much smaller (and larger) values of p. For this reason, we have not tried to optimize the polylogarithmic terms in our bounds on p.

Recall from the example for $\chi'(G)$ in the introduction that Erdős and Wilson [25] showed that a.a.s. the random graph $G = G_{n,p}$ satisfies $\chi'(G) = \Delta(G)$ for $p = 1/2$. Frieze, Jackson, McDiarmid, and Reed [29] extended this to all constant values of $p \in (0,1)$. Recently, this was extended to all $p = o(1)$ by Haxell, Krivelevich, and Kronenberg [37]. This is an example of a graph decomposition result of random graphs that holds for all p, and suggests the possibility that perhaps no lower bound on p is necessary in Theorem 2.1.2.

The proof of Theorem 2.1.2 does not use randomness in a very significant way. In fact, we give a set of sufficient conditions for a digraph to be consistent and show that the random digraph (for suitable p) satisfies these conditions asymptotically almost surely. Here we give a simplified version of our main deterministic result (see Theorem 2.4.3 for the full statement).

For a digraph D, a subset of vertices $S \subseteq V(D)$, and a vertex $v \in V(D)$, we write $e_D(v, S)$ (resp. $e_D(S, v)$) for the number of outneighbors (resp. inneighbors) of v in S.

Theorem 2.1.3. There exist constants n_0 and c such that the following holds. Let $D = (V, E)$ be a digraph on $n \geq n_0$ vertices. Set $t := c(n \log n)^{2/5}$ and let

$$\begin{aligned} A^+ &:= \{v \in V \mid \mathrm{ex}_D(v) \geq t\}, \\ A^- &:= \{v \in V \mid \mathrm{ex}_D(v) \leq -t\}, \text{ and} \\ A^0 &:= V \setminus (A^+ \cup A^-). \end{aligned}$$

Assume there is some $d \geq t$ such that

(i) for every $v \in A^+$ we have $d/4 \leq e_D(v, A^-) \leq d$,

(ii) for every $v \in A^-$ we have $d/4 \leq e_D(A^+, v) \leq d$,

(iii) for every $v \in A^+ \cup A^-$ we have $e_D(v, A^0), e_D(A^0, v) \leq \min\{d/3, t^2/10^6\}$, and

(iv) for every $v \in A^0$ we have $e_D(A^+, v), e_D(v, A^-) \geq d/3$.

Then, D is consistent.

Here is a concrete class of examples to which Theorem 2.1.3 applies. Take the edge-disjoint union of $D = (V, E)$ and $D' = (V, E')$, where D is any digraph obtained by taking a regular bipartite graph of degree $t \geq c(n \log n)^{2/5}$ and orienting all edges from one part to the other, and D' is any Eulerian digraph of maximum degree at most $3t$. One can easily check that Theorem 2.1.3 applies to such digraphs (here A^0 is empty), and so such digraphs are consistent.

Informally, when working with random (di)graphs, a usual strategy is to make use of expansion or pseudorandom properties, see e.g. [52, 57] (meaning the graph is well connected). However, we do not make use of such techniques. Therefore Theorem 2.1.3 can be applied to many digraphs that are far from having any expansion or pseudorandom properties; e.g., digraphs satisfying the conditions of Theorem 2.1.3 could easily be disconnected or weakly connected.

Broadly speaking, our proof relies on the use of the so-called absorption technique, an idea due to Rödl, Ruciński, and Szemerédi [67] (with special forms appearing in earlier work, e.g., [53]). We adapt and refine some of the absorption ideas used in [62], but we also require several new ingredients. We explain the main ideas of our proof in Section 2.2 below. In contrast to the previous work on this question [32, 62], our proof does not make use of robust expanders. Some preliminary ideas for this work came from de Vos [76].

The rest of this chapter is organized as follows. We give a sketch of the proof of Theorem 2.1.2 in Section 2.2. Section 2.3 is dedicated to giving common definitions and citing results we use. In Section 2.4 we describe the *absorbing structure* and we show how to use it to decompose directed graphs D satisfying certain properties into $\mathrm{ex}(D)$ paths. Finally, in Section 2.5 we show that a.a.s. the random digraph contains the absorbing structure and satisfies the properties required to use the absorbing structure for decomposition. The proof of Theorem 2.1.2 appears in Section 2.5 and the proof of Theorem 2.1.3 appears in Section 2.4.

Beginning in Section 2.4 we will sometimes defer details of calculations to endnotes at the end of this chapter in order to improve readability. Endnote markers are superscript numbers in square brackets, like this:[1].

2.2 PROOF SKETCH

Let $D = D_{n,p}$ with p as in Theorem 2.1.2. We divide the vertices of D into sets A^+, A^- and A^0 depending on whether $\mathrm{ex}_D(v) \geq t$, $\mathrm{ex}_D(v) \leq -t$, or $-t < \mathrm{ex}_D(v) < t$, respectively, for a suitable choice of t (as a function of n and p). One can show that, a.a.s., A^+ and A^- have roughly the same size and A^0 is small.

We start by setting aside an absorbing structure $\mathcal{A}$ which consists of a set of edge-disjoint (short) paths of D. Each vertex $v \in V(D)$ will have a set of paths $f(v)$ from $\mathcal{A}$ assigned to it, where the sets $f(v)$ partition $\mathcal{A}$. In particular, for each $v \in A^+$ (resp. $v \in A^-$), the set $f(v)$ consists of single-edge paths from v to A^- (resp. A^+ to v) and, for each $v \in A^0$, the set $f(v)$ consists of a path with two edges which goes from A^+ to A^- through v. We think of $\mathcal{A}$ interchangeably as a set of paths and as a digraph that is the union of those paths. We will require that $|f(v)|$ is sufficiently large for every vertex v but at the same time that $\mathrm{ex}_{\mathcal{A}}(v) \leq \mathrm{ex}_D(v)$ for every vertex v. We give a set of conditions that ensure the existence of one such absorbing structure in Definition 2.4.1 (see Lemmas 2.4.5 and 2.4.6), and Section 2.5 is devoted to showing, by using concentration inequalities for martingales, that $D_{n,p}$ fulfills these conditions (a.a.s.) for all values of p in the desired range (and, in fact, for a slightly larger range than stated in Theorem 2.1.2).

Next it is straightforward to obtain a set of edge-disjoint paths $\mathcal{P}$ in $D \setminus E(\mathcal{A})$ such that $|\mathcal{P}| + |\mathcal{A}| = \mathrm{ex}(D)$, and such that, writing $D' := D \setminus (E(\mathcal{A}) \cup E(\mathcal{P}))$, we have $\mathrm{ex}(D') = 0$. So $\mathcal{P} \cup \mathcal{A}$ gives the correct number (i.e., $\mathrm{ex}(D)$) of edge-disjoint paths but the edges in D' are not covered, and moreover D' is Eulerian. Our goal now is to slowly combine edges of $\mathcal{A}$ with edges of D' to create longer paths in such a way that we maintain exactly $\mathrm{ex}(D)$ paths at every stage (*absorbing* the edges of D'). If we manage to combine all the edges of D' in this way, then we have decomposed D into $\mathrm{ex}(D)$ paths, thus proving that D is consistent.

To begin the process of absorption, we apply a recent result of Knierim, Larcher, Martinsson and Noever [51] (improving on an earlier result of Huang, Ma, Shapira, Sudakov and Yuster [41]) which allows us to decompose the edges of D' into $O(n \log n)$ cycles. The core idea then is to combine certain paths from $\mathcal{A}$ with each cycle C given by the decomposition, and to decompose their union into paths; we refer to this as *absorbing* the cycle. Crucially, in order to keep the number of paths invariant, we will combine

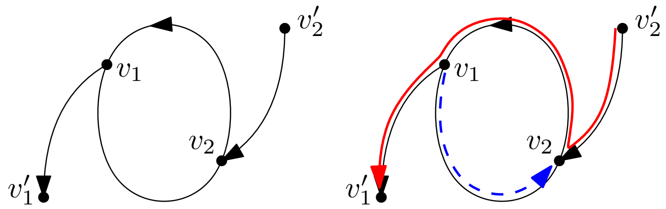


Figure 4: Left: One example of absorbing a cycle using two absorbing edges. We have v_1, v_2 on our cycle C with $v_1 \in A^+$, $v_2 \in A^-$. We find paths $(v_1, v_1') \in f(v_1)$ and $(v_2', v_2) \in f(v_2)$ with $v_1' \in A^- \setminus V(C)$ and $v_2' \in A^+ \setminus V(C)$.
Right: The solid red and dashed blue lines show the two paths $P_1 := v_2' v_2 C v_1 v_1'$ and $P_2 := v_1 C v_2$, which use all involved edges.
Note that under certain circumstances, if v_1', v_2' lie on C, we can still decompose all involved edges into two paths.

each cycle C with a set $\mathcal{A}_C$ of two paths from $\mathcal{A}$ and decompose $C \cup \mathcal{A}_C$ into two paths, as illustrated in Figure 4. Thereafter, the edges $\mathcal{A}_C$ are no longer available for use in absorbing other cycles.

Therefore, we must allocate suitable absorbing paths to the cycles. The two main challenges here are the following.

(i) The absorbing paths need to *fit* the specific cycle, meaning they and the cycle can be decomposed into two paths. Generally, given a cycle C, if we can find vertices $v_1, v_2 \in V(C) \setminus A^0$ and paths $P_1 \in f(v_1)$ and $P_2 \in f(v_2)$ where P_1 and P_2 have distinct endpoints not on C, then P_1 and P_2 will fit C (see Figure 4 for an example). If both endpoints are on C, it is still sometimes possible (but not always) that P_1 and P_2 fit C. If v_1 or v_2 lie in $V(C) \cap A^0$, a similar idea can be used to find fitting paths.

(ii) We only have a limited number of absorbing paths available at each vertex.

In order to address (i), we prepare more absorbing paths than we plan to use, as having the option to select from a sufficiently large number ensures that at least two fit a given cycle. Any paths from $\mathcal{A}$ that we do not eventually use to absorb a cycle remain as paths in the final decomposition. In order to address point (ii), we employ different strategies to assign absorbing edges to cycles, depending on the number of vertices that the cycle has in $A^+ \cup A^-$.

For cycles C that are long (meaning they have many vertices in $A^+ \cup A^-$), we greedily choose two paths that fit the cycle. This is possible as each cycle contains a large number of vertices, so there are many choices for the possible absorbing paths, and we can always find two that fit the cycle. Here, we allow both endpoints of the paths to be on C.

For cycles of medium length, we use a flow problem to assign vertices to cycles in such a way that each cycle is assigned a suitably large number of vertices dependent on its length, but such that no vertex is assigned to too many cycles. This choice of assignment allows us to find two assigned vertices v_1 and v_2 per cycle and pick paths $P_i \in f(v_i)$ for $i = 1, 2$ that fit the cycle. This strategy is wasteful in the sense that we sometimes assign more than two vertices to a cycle and thereby reserve more absorbing paths than we use.

For cycles that are short, it is easier to find fitting paths, as we are guaranteed to find absorbing paths that have their other endpoint off the cycle, as in the example in Figure 4. However, it is harder to ensure that we do not use too many paths per vertex. In this case, we also use a flow problem to assign vertices to cycles, but we take multiple rounds and only decompose certain 'safe' cycles in each round. In addition, we absorb certain closed walks in each round, so we need to apply the result by Knierim et al. between rounds in order to re-decompose the remaining edges into cycles, and this may generate new cycles which are long or of medium length. Absorbing the short cycles is the most complicated process of the three, but it is the process we apply first so that the long and medium cycles that are produced as a byproduct can be absorbed by the appropriate processes described above. It is also the only process in which we use the absorbing paths attached to vertices in A^0.

2.3 PRELIMINARIES

2.3.1 *Basic definitions and notation*

For any $n \in \mathbb{Z}$, we will write $[n] := \{i \in \mathbb{Z} \mid 1 \leq i \leq n\}$ and $[n]_0 := \{i \in \mathbb{Z} \mid 0 \leq i \leq n\}$. Whenever we write $a = b \pm c$ for any $a, b, c \in \mathbb{R}$, we mean that $a \in [b - c, b + c]$. Given any set X, we let 2^X denote the set of all subsets of X. Our logarithms are always natural logarithms. We use the standard $\mathcal{O}$-notation for asymptotic statements, where the asymptotics will always be with respect to a parameter n. Throughout, we ignore rounding whenever it does not affect our arguments.

In this chapter, a *digraph* $D = (V(D), E(D))$ is a loopless directed graph where, for each pair of distinct vertices $x, y \in V(D)$, we allow up to two edges between them, at most one in each direction. We usually denote edges $(x, y) \in E(D)$ simply as xy. The *complement* of D is a digraph on the same vertex set as D which contains exactly all the edges which are not contained in D. Given any digraph D, we write $H \subseteq D$ to mean that H is a *subdigraph* of D, that is, $V(H) \subseteq V(D)$ and $E(H) \subseteq E(D)$. If $\mathcal{H}$ is a set of subdigraphs of D, we will sometimes abuse notation and treat $\mathcal{H}$ as the digraph obtained as the union of the digraphs which comprise $\mathcal{H}$. In particular, we will write $V(\mathcal{H}) := \bigcup_{H\in\mathcal{H}} V(H)$ and $E(\mathcal{H}) := \bigcup_{H\in\mathcal{H}} E(H)$. Given any disjoint sets $A, B \subseteq V(D)$, we denote $E_D(A) := \{ab \in E(D) \mid a, b \in A\}$ and $E_D(A, B) := \{ab \in E(D) \mid a \in A, b \in B\}$. If one of the sets consists of a single element (say, $A = \{a\}$), we will simplify the notation by setting $E(a, B) := E(\{a\}, B)$, and similarly for the rest of the notation. We will write $e_D(A) := |E_D(A)|$ and $e_D(A, B) := |E_D(A, B)|$. We denote $D[A] := (A, E_D(A))$ for the subdigraph *induced* by A and, similarly, $D[A, B] := (A \cup B, E_D(A, B))$ for the bipartite subdigraph *induced* by (A, B). Given any $E \subseteq E(D)$, we write $D \setminus E := (V(D), E(D) \setminus E)$. Given any vertex $x \in V(D)$, we define its *outneighborhood* and *inneighborhood* as $N_D^+(x) := \{y \in V(D) \mid xy \in E(D)\}$ and $N_D^-(x) := \{y \in V(D) \mid yx \in E(D)\}$, respectively. The *outdegree* and *indegree* of x are given by $d_D^+(x) := |N_D^+(x)|$ and $d_D^-(x) := |N_D^-(x)|$, respectively. Throughout, we may sometimes abuse notation by referring to a digraph by its edge set, especially in subscripts; the vertex set of such digraphs will always be clear from context.

As in the introduction, we define the *excess* at x to be $\mathrm{ex}_D(x) := d_D^+(x) - d_D^-(x)$, and similarly define the *positive excess* and *negative excess* at x as $\mathrm{ex}_D^+(x) := \max\{\mathrm{ex}_D(x), 0\}$ and $\mathrm{ex}_D^-(x) := \max\{-\mathrm{ex}_D(x), 0\}$, respectively. Observe that $\sum_{x\in V(D)} \mathrm{ex}_D(x) = 0$. We define the *excess* of D as

$$\mathrm{ex}(D) := \sum_{x\in V(D)} \mathrm{ex}_D^+(x) = \sum_{x\in V(D)} \mathrm{ex}_D^-(x) = \frac{1}{2} \sum_{x\in V(D)} |\mathrm{ex}_D(x)|.$$

When we refer to paths, cycles, and walks in digraphs, we mean *directed* paths, cycles, and walks, i.e., the edges are oriented consistently. Given a digraph D, a *walk* W in D is given by a sequence of (not necessarily distinct) vertices $W = v_1v_2 \cdots v_k$ where $v_1v_2, v_2v_3, \ldots, v_{k-1}v_k$ are distinct edges of D. We also think of W as being a subdigraph of D with vertex set $\{v_1, \dots, v_k\}$ and edge set $\{v_1v_2, v_2v_3, \dots, v_{k-1}v_k\}$. We also call W a (v_1, v_k)-walk and sometimes denote it by v_1Wv_k to emphasise that it starts at v_1 and ends at v_k, and we say W is *closed* if $v_1 = v_k$. For two edge-disjoint walks $W_1 = aW_1b = av_1 \cdots v_kb$ and $W_2 = bW_2c = bv'_1 \cdots v'_\ell c$, we write $W_1W_2 = av_1 \cdots v_kbv'_1 \cdots v'_\ell c$ for the concatenation of W_1 and W_2. This notation extends in the natural way for concatenating more than two walks. For a walk $W = v_1 \cdots v_k$, and $1 \leq i < j \leq k$, we write v_iWv_j for the (v_i, v_j)-walk $v_iv_{i+1} \cdots v_j$ between v_i and v_j.

In fact, we will mostly be concerned with paths and cycles rather than walks. A walk $W = v_1 \cdots v_k$ is a *path* if $v_1, \dots, v_k$ are distinct vertices, and it is a *cycle* if $v_1, \dots, v_k$ are distinct except that $v_1 = v_k$. The length of a walk, path, or cycle is the number of edges it contains. We sometimes also consider degenerate single-vertex paths. Note that, if P_1 is an (a, b)-path and P_2 is a (b, c)-path, where P_1 and P_2 are vertex-disjoint except at b, then P_1P_2 is an (a, c)-path. For sets of vertices X and Y, we say that a path P is an (X, Y)-path if it starts in X and ends in Y.

In this chapter, we say a digraph D is *Eulerian* if $d^+_D(v) = d^-_D(v)$ for every $v \in V(D)$ or, equivalently, if $\mathrm{ex}(D) = 0$.[1] A well-known consequence of this definition is the fact that the edge set of any Eulerian digraph can be decomposed into cycles.

We will sometimes need to consider a *multidigraph* D, which is allowed to have multiple edges between any two vertices, in both directions (but it is still loopless). Whenever D is a multidigraph, all edge sets should be seen as multisets, while all vertex sets will remain simple sets. The notation and terminology above extend in the natural way to multidigraphs.

2.3.2 *Path and cycle decompositions*

The following definitions are convenient.

Definition 2.3.1. A *perfect decomposition* of a digraph D is a set $\mathcal{P} = \{P_1, \dots, P_r\}$ of edge-disjoint paths of D that together cover $E(D)$ with $r = \mathrm{ex}(D)$. (Thus, a digraph D is consistent if and only if it has a perfect decomposition.)

We will need the following basic facts.

1 This is different from the standard definition, which also asks that D is strongly connected.

Proposition 2.3.2. Let D be a digraph with $\mathrm{ex}(D) > 0$. Then, there exists a path in D from a vertex of positive excess to a vertex of negative excess.

Proof. First, repeatedly remove cycles from D until this is no longer possible and call the resulting digraph D'; note that this does not affect the excess of any vertex. Now any maximal path P in D' starts at a vertex that has no inneighbors (so it has positive excess) and ends at a vertex that has no outneighbors (so it has negative excess). □

Proposition 2.3.3. Suppose D is a digraph, and let $X, Y \subseteq V(D)$ be disjoint. If $P_1, \ldots, P_k$ are edge-disjoint (X,Y)-paths and $E(P_1) \cup \ldots \cup E(P_k) = E(D)$, then $\{P_1, \ldots, P_k\}$ is a perfect decomposition of D.

Proof. If we construct D by adding the k paths one at a time, we notice that the excess increases by one each time a path is added, so that $\mathrm{ex}(D) = k$. □

As mentioned in Section 2.2, we will use *absorbing structures* (see Definition 2.4.4) to absorb Eulerian digraphs. For this, we will first decompose the Eulerian digraphs into cycles. We will use Theorem 2.3.4 of Knierim, Larcher, Martinsson and Noever [51] to achieve this.

Theorem 2.3.4. There exists a constant c' such that every Eulerian digraph D on n vertices can be decomposed into at most $c'n \log n$ edge-disjoint cycles.[2]

2.3.3 *Flows*

We recall some common definitions and facts about flow networks. We note that flows are only used in the proofs of Lemmas 2.4.11 and 2.4.14.

A *flow network* is a tuple (F, w, s, t), where $F = (V, E)$ is a digraph, $w \colon E \to \mathbb{R}$ is the *capacity* function, and $s \in V$ is a *source* (i.e., it only has outedges incident to it) and $t \in V$ is a *sink* (i.e., it only has inedges incident to it). A *flow* for the flow network (F, w, s, t) is a function $\phi \colon E \to \mathbb{R}_{\geq 0}$ such that, for all $e \in E$, we have $\phi(e) \leq w(e)$ and, for all $v \in V \setminus \{s, t\}$, we have $\sum_{u \in N_F^-(v)} \phi(uv) = \sum_{u \in N_F^+(v)} \phi(vu)$. We define the *value* of ϕ as

2 In fact, the result of Knierim, Larcher, Martinsson and Noever [51] is slightly stronger, in the sense that $\log n$ can be replaced by $\log \Delta$, where Δ is the maximum (out- or in-)degree of D.

$val(\phi) := \sum_{v \in N_F^+(s)} \phi(sv)$. A *maximum flow* on a given flow network is a flow ϕ that maximises $val(\phi)$.

A partition (U, W) of V with $s \in U$, $t \in W$ is called a *cut*, and we call the edge set $E_F(U, W)$ its corresponding *cut-set*. The *capacity* $w((U, W))$ of a cut (U, W) is the sum of the capacities of the edges of its cut-set, i.e., $w((U, W)) := w(E_F(U, W)) := \sum_{e \in E_F(U,W)} w(e)$. A *minimum cut* of the given flow network is a cut of minimum capacity. We make use of the following well-known theorem.

Theorem 2.3.5 (Max-flow min-cut [28])**.** For every flow network with maximum flow ϕ and minimum cut (U, W) we have that $val(\phi) = w((U, W))$.

An easy consequence is that, if all edge capacities are integers, then there exists a maximum flow such that all flow values are integers.

Given a flow ϕ on a flow network (F, w, s, t), we define the *residual digraph* G_ϕ of G under ϕ as a directed graph with vertex set V and edge set $\{uv \in E \mid \phi(uv) < w(uv)\} \cup \{vu \mid uv \in E,\ \phi(uv) > 0\}$. An (s, t)-path in a residual graph G_ϕ is called an *augmenting path*, and it is easy to see that an augmenting path exists in G_ϕ if and only if ϕ is not a maximum flow.

2.3.4 *Random digraphs and probabilistic estimates*

In Section 2.5, we begin working with random digraphs in the binomial model (although we also introduce slight variants of this model in the proofs of Lemmas 2.4.5 and 2.4.6). We denote by $D_{n,p}$ a random digraph on vertex set $[n]$ obtained by adding each of the possible $n(n-1)$ edges with probability p, independently of all other edges. Most of our results will be asymptotic in nature. In particular, given a (di)graph property $\mathcal{P}$ and a sequence of random (di)graphs $\{G_i\}_{i>0}$ with $|V(G_i)| \to \infty$ as $i \to \infty$, we say that G_i satisfies $\mathcal{P}$ *asymptotically almost surely* (a.a.s.) if $\mathbb{P}[G_i \in \mathcal{P}] \to 1$ as $i \to \infty$.

We will need to prove concentration results for different random variables. For this, we will often use Chernoff bounds (see, e.g., the book of Janson, Łuczak and Ruciński [44, Corollary 2.3]).

Lemma 2.3.6. Let X be the sum of n mutually independent Bernoulli random variables, and let $\mu := \mathbb{E}[X]$. Then, for all $\delta \in (0, 1)$ 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 Chernoff-type bound extends Lemma 2.3.6 to allow us to bound probabilities of large deviations (see, e.g., the book of Alon and Spencer [2, Theorem A.1.12]).

Lemma 2.3.7. Let X be the sum of n mutually independent Bernoulli random variables. Let $\mu := \mathbb{E}[X]$, and let $\beta > 1$. Then, $\mathbb{P}[X \geq \beta\mu] \leq (e/\beta)^{\beta\mu}$.

We will sometimes consider random variables which are not independent, in which case we cannot obtain concentration results as above. For such random variables we will need the following version of the well-known Azuma-Hoeffding inequality (see, e.g., [44, Theorem 2.25]). 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*[3] for f and X. All the martingales that appear in this chapter will be of this form.

Lemma 2.3.8 (Azuma's inequality)**.** Let $Y_0, \ldots, Y_n$ be a martingale and suppose $|Y_i - Y_{i-1}| \leq c_i$ for all $i \in [n]$. Then, for any $t > 0$,

$$\mathbb{P}[|Y_n - Y_0| \geq t] \leq 2\exp\left(\frac{-t^2}{2\sum_{i=1}^n c_i^2}\right).$$

We will also make use of the following well-known inequality; see, e.g., [36, Theorem 368].

Lemma 2.3.9 (rearrangement inequality)**.** Let $n \in \mathbb{N}$, and let $x_1 \leq \ldots \leq x_n$ and $y_1 \leq \ldots \leq y_n$ be real numbers. Let $\sigma \in \mathfrak{S}_n$ be an arbitrary permutation. Then,

$$\sum_{i=1}^n x_i y_{n+1-i} \leq \sum_{i=1}^n x_i y_{\sigma(i)} \leq \sum_{i=1}^n x_i y_i.$$

2.4 OPTIMAL PATH DECOMPOSITIONS OF DIGRAPHS

In this section we give sufficient conditions for a digraph to be consistent. These conditions will ensure that our digraph has a certain absorbing structure, and the absorbing structure will help us to decompose D into $\mathrm{ex}(D)$ paths.

3 The definition here is all we require and so we will not define martingales.

We begin by defining the classes of digraphs we will be working with throughout the rest of the chapter.

Definition 2.4.1. Fix $p \in [0,1]$ and $0 \leq \lambda, \kappa \leq n$. We say that $D = (V, E)$ is an (n, p, κ, λ)*-digraph* if $|V| = n$ and the vertex set V can be partitioned into three parts, A^+, A^- and A^0 (where A^0 may be empty), in such a way that the following properties are satisfied:

(P1) For every $v \in A^+$ we have $\mathrm{ex}_D(v) \geq 155\kappa$ and $np/4 \leq e_D(v, A^-) \leq np$.

(P2) For every $v \in A^-$ we have $\mathrm{ex}_D(v) \leq -155\kappa$ and $np/4 \leq e_D(A^+, v) \leq np$.

(P3) For every $v \in A^+ \cup A^-$ we have $e_D(v, A^0), e_D(A^0, v) \leq \lambda$.

(P4) For every $v \in A^0$ we have $e_D(A^+, v) \geq np/3$ and $e_D(v, A^-) \geq np/3$.

We say that D is an (n, p, κ, λ)*-pseudorandom digraph* if it is an (n, p, κ, λ)-digraph and, additionally, the following property holds:

(P5) For every set $U \subseteq V$ with $|U| \geq \log n/(50p)$ we have $e_D(U) \leq 100|U|^2 p$.

Whenever we are given an (n, p, κ, λ)-digraph, we implicitly consider a partition of its vertex set into sets A^+, A^- and A^0 which satisfy the properties described in Definition 2.4.1. This partition is not necessarily unique; throughout this section, we simply assume that one such partition is given. We will write $\dot{A} := A^+ \cup A^-$.

Remark 2.4.2. If D is an (n, p, κ, λ)-(pseudorandom) digraph and $\kappa' \leq \kappa$ and $\lambda' \geq \lambda$, then D is an $(n, p, \kappa', \lambda')$-(pseudorandom) digraph.

We will see in Section 2.5 that a.a.s. $D_{n,p}$ is an (n, p, κ, λ)-pseudorandom digraph, for a suitable choice of parameters. Our goal in this section is to prove the following theorem.

Theorem 2.4.3. There exists $n_0 \in \mathbb{N}$ with the following property. Suppose $n \in \mathbb{N}$, $p \in (0,1)$ and $\kappa, \lambda \in \mathbb{R}$ are parameters satisfying $n \geq n_0$ and

(C1) $\kappa = 3N^{2/5}$,

(C2) $np \geq 365N^{2/5}$, and

(C3) $\lambda = \min\{np/3, \kappa^2/12\}$, where $N := c'n \log n$ and c' is the constant from Theorem 2.3.4. Then, any (n, p, κ, λ)-digraph D admits a perfect decomposition.

The same conclusion holds if D is an (n, p, κ, λ)-pseudorandom digraph and (C1) and (C2) are replaced by

(C′1) $\kappa = 6(N^2 p)^{1/5}$, and

(C′2) $p \geq n^{-1/3} \log^4 n$.

Observe that, by Remark 2.4.2, we can extend Theorem 2.4.3 to any (n, p, κ, λ)-(pseudorandom) digraph where κ is larger than the value given in (C1) or (C′1), respectively, and λ is smaller than the value given in (C3).

We further remark that the constants in Theorem 2.4.3 as well as in Definition 2.4.1 are not optimal. In fact, there is a trade-off between some of them: by making one worse, others can be improved. In order to ease readability, we refrain from stating the most general result possible, and simply note that a host of similar statements, with different constants, can be obtained by going through the proofs of the lemmas in this section. Furthermore, we note that some of the conditions in Definition 2.4.1 can be relaxed; in particular, (P3) is only used in the proof of Lemma 2.4.11, where only one of the two bounds stated in (P3) is required. Thus, as long as all vertices in $A^+ \cup A^-$ satisfy one (and the same) of the two bounds, Theorem 2.4.3 still holds, so it can be applied to a larger class of digraphs than stated in Definition 2.4.1.

Assuming Theorem 2.4.3, we give the proof of Theorem 2.1.3.

Proof of Theorem 2.1.3. We set n_0 as in Theorem 2.4.3 and $c := 500(c')^{2/5}$, where c' is the constant from Theorem 2.3.4. Then, properties (i)–(iv) of Theorem 2.1.3 and our choice of A^+, A^- and A^0 correspond to (P1)–(P4) with $t := c(n \log n)^{2/5}$, d, and $\min\{d/3, t^2/10^6\}$ playing the roles of 155κ, np, and λ, respectively, so D is an (n, p, κ, λ)-digraph. By Remark 2.4.2 and our choice of t and d, we then conclude that D is also an $(n, p, \kappa', \lambda')$-digraph which satisfies properties (C1)–(C3) of Theorem 2.4.3[1]. Thus, we may apply Theorem 2.4.3 and D is consistent. □

2.4.1 *Finding absorbing structures*

The next definition describes the absorbing structure that we will find in (n, p, κ, λ)-digraphs D. It will be used to absorb the majority of edges of

D into a set of (A^+, A^-)-paths that will end up being part of our perfect decomposition. We will essentially show that, when we take an edge-disjoint union of our absorbing structure with any Eulerian subdigraph of D, the resulting digraph has a perfect decomposition.

Definition 2.4.4. Let D be an (n, p, κ, λ)-digraph, and let $Z \subseteq V(D)$ and $t \in \mathbb{N}$. A (Z, t)-*absorbing structure* is a pair $\mathcal{A} = (E^{\mathrm{ab}}, f)$, where $E^{\mathrm{ab}} \subseteq E(D)$ and $f\colon Z \to 2^{E^{\mathrm{ab}}}$, such that

(A1) if $z \in Z \cap A^+$, then $f(z)$ contains exactly t edges from $E_D(z, A^-)$;

(A2) if $z \in Z \cap A^-$, then $f(z)$ contains exactly t edges from $E_D(A^+, z)$;

(A3) if $z \in Z \cap A^0$, then $f(z)$ contains exactly t edges from $E_D(A^+, z)$ and exactly t edges from $E_D(A^+, z)$, and

(A4) the collection $\{f(z)\}_{z \in Z}$ is a partition of E^{ab}; in particular, the sets $f(z)$ are disjoint.

Note that, for convenience, for $z \in A^+ \cup A^-$, we often think of the t edges in $f(z)$ as t edge-disjoint (A^+, A^-)-paths of length 1. For $z \in A^0$, we arbitrarily pair up the in- and outedges in $f(z)$ to create t edge-disjoint (A^+, A^-)-paths of length 2 through z.

The following lemmas show the existence of absorbing structures in (n, p, κ, λ)-digraphs.

Lemma 2.4.5. Let D be an (n, p, κ, λ)-digraph with $100 \log n < \kappa \leq np/120$. Then, D contains an $(\dot{A}, 12\kappa)$-absorbing structure which contains at most 150κ edges incident to each $v \in \dot{A}$.

Proof. Consider $D[A^+, A^-]$. We define D_q as a random subdigraph of $D[A^+, A^-]$ by including each of the edges of $E_D(A^+, A^-)$ with probability $q \coloneqq 120\kappa/(np)$, independently of each other. For each $v \in A^+$, let $\mathcal{B}_v$ be the event that $d^+_{D_q}(v) \notin [25\kappa, 150\kappa]$. Similarly, for each $v \in A^-$, let $\mathcal{B}_v$ be the event that $d^-_{D_q}(v) \notin [25\kappa, 150\kappa]$. By (P1), (P2) and Lemma 2.3.6, it follows that, for each $v \in \dot{A}$, we have $\mathbb{P}[\mathcal{B}_v] \leq e^{-\kappa/50}$[2]. Then, by a union bound over all $v \in \dot{A}$ and the lower bound on κ, we conclude that there exists a digraph $D' \subseteq D[A^+, A^-]$ such that, for each $v \in A^+$, it holds that $d^+_{D'}(v) \in [25\kappa, 150\kappa]$, and for each $v \in A^-$, it holds that $d^-_{D'}(v) \in [25\kappa, 150\kappa]$.

We are now going to randomly split the edges of D' into two sets E^+ and E^-, and then prove that, with positive probability, E^+ contains an $(A^+, 12\kappa)$-absorbing structure $\mathcal{A}^+$, and E^- contains an $(A^-, 12\kappa)$-absorbing structure $\mathcal{A}^-$. It then immediately follows that $\mathcal{A}^+ \cup \mathcal{A}^-$ is the desired $(\dot{A}, 12\kappa)$-absorbing structure.

For each $e \in E(D')$, with probability $1/2$ and independently of all other edges, we assign e to E^+, and otherwise we assign it to E^-. Let $D^+ := (\dot{A}, E^+)$ and $D^- := (\dot{A}, E^-)$ (so, in particular, $D' = D^+ \cup D^-$). Now, for each $v \in A^+$, let $\mathcal{B}'_v$ be the event that $d^+_{D^+}(v) < 12\kappa$, and for each $v \in A^-$, let $\mathcal{B}'_v$ be the event that $d^-_{D^-}(v) < 12\kappa$. In particular, by Lemma 2.3.6, it follows that, for each $v \in \dot{A}$, we have $\mathbb{P}[\mathcal{B}'_v] \leq e^{-\kappa/100}$[3]. By a union bound, we conclude that there exists a partition of $E(D')$ into E^+ and E^- such that, for each $v \in A^+$, we have $d^+_{D^+}(v) \geq 12\kappa$, and for each $v \in A^-$ we have $d^-_{D^-}(v) \geq 12\kappa$[4].

In order to obtain the desired absorbing structure, for each $v \in A^+$ let $f(v)$ be an arbitrary set of 12κ of the edges of E^+ which contain v, and for each $v \in A^-$ let $f(v)$ be an arbitrary set of 12κ of the edges of E^- which contain v. □

Lemma 2.4.6. Let D be an (n, p, κ, λ)-digraph with $8\log(4n) < \kappa \leq np/12$, $\lambda \leq np/3$ and $\kappa\lambda \geq 4np\log(2n)$. Then, D contains an $(A^0, 3\kappa)$-absorbing structure which contains at most 5κ edges incident to each $v \in \dot{A}$.

Proof. Let $D' := D[A^+, A^0] \cup D[A^0, A^-]$, and let D_q be a random subdigraph of D' obtained by adding each edge of D' with probability $q := 12\kappa/(np)$ and independently of each other. For each $v \in A^+$, let $\mathcal{B}_v$ be the event that $d^+_{D_q}(v) > 5\kappa$. Similarly, for each $v \in A^-$, let $\mathcal{B}_v$ be the event that $d^-_{D_q}(v) > 5\kappa$. Finally, for each $v \in A^0$, let $\mathcal{B}^+_v$ and $\mathcal{B}^-_v$ be the events that $d^-_{D_q}(v) < 3\kappa$ and $d^+_{D_q}(v) < 3\kappa$, respectively.

It follows from (P3) and Lemma 2.3.6 that, for each $v \in \dot{A}$, we have $\mathbb{P}[\mathcal{B}_v] \leq e^{-\kappa\lambda/(4np)}$[5]. Similarly, by (P4) and Lemma 2.3.6, for each $v \in A^0$ we have that $\mathbb{P}[\mathcal{B}^+_v], \mathbb{P}[\mathcal{B}^-_v] \leq e^{-\kappa/8}$[6]. By a union bound (the trivial bound is given by $2ne^{-\kappa/8} + ne^{-\kappa\lambda/(4np)}$, and this is < 1 by the assumptions in the statement), we conclude that there exists $D^* \subseteq D'$ such that, for each $v \in A^+$, we have $d^+_{D_q}(v) \leq 5\kappa$; for each $v \in A^-$, we have $d^-_{D_q}(v) \leq 5\kappa$, and for each $v \in A^0$, we have $d^+_{D_q}(v), d^-_{D_q}(v) \geq 3\kappa$.

In order to obtain the absorbing structure, for each $v \in A^0$, let $f(v)$ be the union of an arbitrary subset of $E_{D^*}(A^+, v)$ of size 3κ and an arbitrary subset of $E_{D^*}(v, A^-)$ of size 3κ. $\square$

2.4.2 *Using absorbing structures*

In this subsection, we show how to use absorbing structures to obtain perfect decompositions, and we use this to prove Theorem 2.4.3. As mentioned earlier, the idea will be to use these absorbing structures to absorb Eulerian digraphs. The Eulerian digraphs will be decomposed into cycles, using Theorem 2.3.4, and absorbed one cycle at a time.

Given an (n, p, κ, λ)-digraph D, we set $N := c'n \log n$, where c' is the constant given by Theorem 2.3.4, so any Eulerian subdigraph of D can be decomposed into at most N cycles. We call a cycle $C \subseteq D$ *short* if $|V(C) \cap \dot{A}| \leq \kappa$, *long* if $|V(C) \cap \dot{A}| \geq N/\kappa$, and *medium* otherwise. We will need a different strategy to absorb the set of cycles of each type. We will show how to absorb long, medium and short cycles in Lemmas 2.4.9, 2.4.11 and 2.4.14, respectively.

The following lemma shows how to absorb a single long or medium cycle, under suitable conditions, and will be used in Lemmas 2.4.9 and 2.4.11.

Lemma 2.4.7. Let D be an (n, p, κ, λ)-digraph. Let $C \subseteq D$ be a cycle with $\ell := |V(C) \cap \dot{A}| > \kappa$ and $S \subseteq V(C) \cap \dot{A}$ with $|S| \geq \ell/\kappa + 1$. Let $\mathcal{A} = (E^{\text{ab}}, f)$ be an $(S, \kappa + 2)$-absorbing structure such that $E(C) \cap E^{\text{ab}} = \emptyset$. Then, there exist distinct vertices $v_1, v_2 \in S$ and edges $e_1 \in f(v_1)$ and $e_2 \in f(v_2)$ such that $E(C) \cup \{e_1, e_2\}$ can be decomposed into two (A^+, A^-)-paths.

Proof. Assume first that there are two distinct vertices $v_1, v_2 \in S$ such that, for each $i \in [2]$, there is an edge $e_i \in f(v_i)$ whose other vertex is not contained in $V(C)$. Observe that the definition of $\mathcal{A}$ ensures that $e_1 \cup e_2$ is not a path of length 2[7]. Now, for each $i \in [2]$, if $e_i = v_i x_i$, let $P_i^+ := v_i x_i$ and $P_i^- := v_i$, and if $e_i = x_i v_i$, let $P_i^+ := v_i$ and $P_i^- := x_i v_i$. Let P be the (v_1, v_2)-subpath of C, and let P' be the (v_2, v_1)-subpath of C. The paths described in the statement are now given by $P_1 := P_1^- P P_2^+$ and $P_2 := P_2^- P' P_1^+$. Since $e_1 \cup e_2$ is not a path of length 2, these two structures must indeed be paths and in all cases they are (A^+, A^-)-paths since the paths have the same start- and endpoints as e_1 and e_2. See Figure 5 for a visual representation of two of the four possible outcomes.

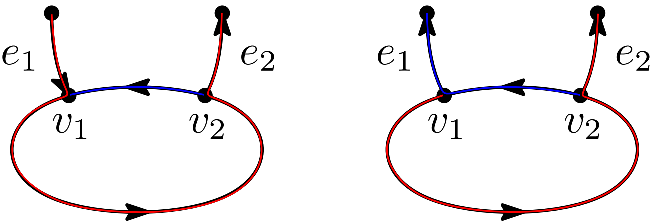


Figure 5: A representation of the path decomposition of a cycle and two edges as proposed in Lemma 2.4.7, in the case where we can find said edges with their endpoints outside $V(C)$.

Therefore, we may assume that there are at least $\ell/\kappa > 1$ vertices $v \in S$ such that all $e \in f(v)$ have both endpoints in $V(C)$. Let us denote the set of these vertices by S'. For each $v \in S'$, let P_v be the shortest subpath of C which does not contain v and contains all other endpoints of the edges $e \in f(v)$ (recall that all said endpoints lie in $\dot{A}$). In particular, $|V(P_v) \cap \dot{A}| \geq \kappa + 2$. Now label the vertices of $V(C) \cap \dot{A}$ as $y_1, \ldots, y_\ell$ in such a way that, when traversing C, they are visited in this (cyclic) order. A simple counting argument shows the following.

Claim 2.4.8. There exist two distinct vertices $v_1, v_2 \in S'$ such that P_{v_1} and P_{v_2} share at least two consecutive vertices of $V(C) \cap \dot{A}$.

Proof of Claim 2.4.8. Assume the statement does not hold. Then, any two paths from $\{P_v \mid v \in S'\}$ can intersect only at their endpoints, and any vertex of $V(C) \cap \dot{A}$ can be an endpoint of at most two paths. This means

$$\sum_{v \in S'} |V(P_v) \cap \dot{A}| \leq \ell + |S'|.$$

However, using the bounds we have obtained so far, we can confirm that

$$\sum_{v \in S'} |V(P_v) \cap \dot{A}| \geq |S'|(\kappa + 2) \geq \ell + 2|S'| > \ell + |S'|. \quad \blacktriangleleft$$

By Claim 2.4.8, we can choose two edges $e_1 \in f(v_1)$ and $e_2 \in f(v_2)$ which form a ‘crossing configuration’, that is, such that the vertices of e_1 and e_2 alternate when traversing C (e.g., wy and zx are crossing edges in Figure 6). In order to complete the proof, label the vertices of e_1 and e_2 as w, x, y, z in such a way that, when traversing the cycle, they appear in this (cyclic) order and such that the edges are oriented towards x and

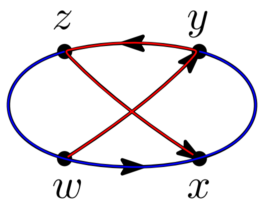


Figure 6: A representation of the path decomposition of a cycle and two edges as proposed in Lemma 2.4.7, in the case where we can find a 'crossing configuration'.

towards y, respectively (note that in any crossing configuration there exist two consecutive vertices into which the edges are directed). The two paths of the statement are now given by $P_1 := wyCzx$ and $P_2 := zCy$, and these are (A^+, A^-)-paths since they have the same start- and endpoints as e_1 and e_2. See Figure 6 for a visual representation. $\square$

We now prove Lemma 2.4.9, which shows how an absorbing structure can be used to absorb a collection of long cycles.

Lemma 2.4.9. Let D be an (n, p, κ, λ)-digraph with $10 \leq \kappa < N^{1/2}$. Let $\mathcal{C}_1$ be a collection of edge-disjoint cycles in D with $|\mathcal{C}_1| \leq 2N$ and such that, for each $C \in \mathcal{C}_1$, we have $|V(C) \cap \dot{A}| \geq N/\kappa$. Let $\mathcal{A} = (E^{\mathrm{ab}}, f)$ be an $(\dot{A}, 7\kappa - 1)$-absorbing structure with $E(\mathcal{C}_1) \cap E^{\mathrm{ab}} = \emptyset$. Then, the digraph with edge set $E(\mathcal{C}_1) \cup E^{\mathrm{ab}}$ has a perfect decomposition in which each path is an (A^+, A^-)-path.

Proof. For each $C \in \mathcal{C}_1$, we are going to use Lemma 2.4.7 to find two edges $e_1, e_2 \in E^{\mathrm{ab}}$ such that $E(C) \cup \{e_1, e_2\}$ can be decomposed into two (A^+, A^-)-paths. We proceed iteratively as follows.

Assume that, for some of the cycles in $\mathcal{C}_1$, we have already found two edges as described above, and we now wish to do this for the next cycle $C \in \mathcal{C}_1$. Let $\ell := |V(C) \cap \dot{A}| \geq N/\kappa$. We say that an edge $e \in E^{\mathrm{ab}}$ is *available* if it has not been used to absorb any of the earlier cycles. We say that a vertex $v \in V(C) \cap \dot{A}$ is *available* if at least $\kappa + 2$ edges of $f(v)$ are available, and we say that it is *unavailable* otherwise. Let $S_C \subseteq V(C) \cap \dot{A}$ be the set of available vertices. Then, we can define an $(S_C, \kappa + 2)$-absorbing structure $\mathcal{A}_C$ using edges from E^{ab} by selecting, for each $v \in S_C$, any set of $\kappa + 2$ available edges from $f(v)$.

Note that the total number of edges assigned to cycles so far is at most $2|\mathcal{C}_1| \leq 4N$. On the other hand, for each $v \in V(C) \cap \dot{A}$ which is unavailable, at least $5\kappa^{[8]}$ edges of $f(v)$ have already been assigned to cycles. Therefore, the total number of unavailable vertices is at most $4N/(5\kappa)$, so $|S_C| \geq \ell - 4N/(5\kappa) \geq \ell/5 \geq \ell/\kappa + 1^{[9]}$. Therefore (noting that $\ell \geq N/\kappa > \kappa$), we can apply Lemma 2.4.7 (with S_C and $\mathcal{A}_C$ playing the roles of S and $\mathcal{A}$, respectively) to obtain two (available) edges $e_1, e_2 \in E^{\mathrm{ab}}$ such that $E(C) \cup \{e_1, e_2\}$ can be decomposed into two (A^+, A^-)-paths.

After each cycle has been handled in this way and, together with two edges, decomposed into two (A^+, A^-)-paths, we are left with some edges in E^{ab}, which we treat as (A^+, A^-)-paths. We therefore have a decomposition of $E(\mathcal{C}_1) \cup E^{\mathrm{ab}}$ into (A^+, A^-)-paths, which is a perfect decomposition by Proposition 2.3.3. □

We will use flow problems in order to prove Lemmas 2.4.11 and 2.4.14. All our flow problems will follow a similar structure, so we introduce the following definition in addition to the common definitions given in Subsection 2.3.3.

Definition 2.4.10. Let D be a multidigraph and $\mathcal{C}$ be a set of edge-disjoint cycles of D. Set $B := V(\mathcal{C})$. We define a flow network (F, w, s, t) as follows. We define a digraph $F = F(\mathcal{C})$ on vertex set $\{s\} \,\dot\cup\, \mathcal{C} \,\dot\cup\, B \,\dot\cup\, \{t\}$, where s and t are the source and sink of the flow problem, respectively. We set $E_1 := \{sC \mid C \in \mathcal{C}\}$, $E_2 := \{Cb \mid C \in \mathcal{C}, b \in V(C)\}$, $E_3 := \{bt \mid b \in B\}$ and $E(F) := E_1 \cup E_2 \cup E_3$. Given any two functions $g\colon \mathcal{C} \to \mathbb{R}$ and $h\colon B \to \mathbb{R}$, we will write $FP(\mathcal{C}; g, h)$ to denote the maximum flow problem on the digraph $F = F(\mathcal{C})$ defined above where each edge $sC \in E_1$ has capacity $w(sC) = g(C)$, each edge $Cb \in E_2$ has capacity $w(Cb) = 1$, and each edge $bt \in E_3$ has capacity $w(bt) = h(b)$. If g or h are constant functions, we will simply replace them by the corresponding constant in the notation.

The following lemma shows how an absorbing structure can be used to absorb a collection of medium cycles.

Lemma 2.4.11. Let D be an (n, p, κ, λ)-digraph D with $\kappa \geq \max\{12, (12\lambda)^{1/2}, (72N^2)^{1/5}\}$, or an (n, p, κ, λ)-pseudorandom digraph D with $\kappa \geq \max\{12, (12\lambda)^{1/2}, (7200N^2p)^{1/5}, \sqrt{12/(25p)}\log n\}$. Let $\mathcal{C}_2$ be a collection of at most $2N$ edge-disjoint cycles in D such that, for each $C \in \mathcal{C}_2$, we have

$$\kappa < |V(C) \cap \dot{A}| < N/\kappa. \tag{2.4.1}$$

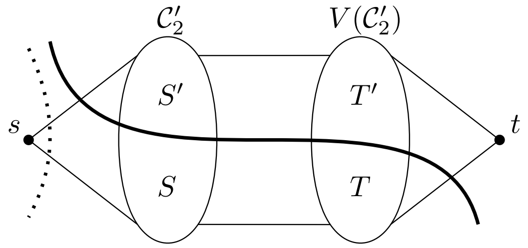


Figure 7: The graph $F(\mathcal{C}_2')$. The thick dotted line illustrates the cut-set M_0. The regular thick line illustrates the cut-set M.

Let $\mathcal{A} = (E^{\mathrm{ab}}, f)$ be an $(\dot{A}, 2\kappa+1)$-absorbing structure with $E(\mathcal{C}_2) \cap E^{\mathrm{ab}} = \emptyset$. Then, the digraph with edge set $E(\mathcal{C}_2) \cup E^{\mathrm{ab}}$ has a perfect decomposition in which each path is an (A^+, A^-)-path.

Proof. Given any digraph H with $V(H) \subseteq V(D)$, we define $g(H) := \lceil |V(H) \cap \dot{A}|/\kappa \rceil + 1$. We use a flow problem to assign, to each $C \in \mathcal{C}_2$, a set of $g(C)$ vertices of $V(C) \cap \dot{A}$ in such a way that no vertex is assigned to more than κ cycles. We will then use Lemma 2.4.7 to find two edges in E^{ab} with which to absorb C. To this end, we construct a multiset of auxiliary cycles $\mathcal{C}_2'$ as follows. We obtain $\mathcal{C}_2'$ from $\mathcal{C}_2$ by replacing each cycle $C \in \mathcal{C}_2$ by the auxiliary cycle $i(C)$ with vertices $V(C) \setminus A^0$ and whose cyclic vertex order is inherited from C. Note that the cycles in $\mathcal{C}_2'$ are not necessarily cycles of D and, indeed, the set $E(\mathcal{C}_2')$ (which forms a multidigraph) includes all the edges of $E(\mathcal{C}_2)$ inside $\dot{A}$ as well as an extra edge every time a cycle in $\mathcal{C}_2$ leaves and reenters $\dot{A}$. We note for later that, since $e_D(v, A^0) \leq \lambda$ by (P3), the number of these extra edges contained in any $T \subseteq \dot{A}$ is at most $\lambda|T|$. Consider $FP(\mathcal{C}_2'; g, \kappa)$.

Claim 2.4.12. *$FP(\mathcal{C}_2'; g, \kappa)$ has a flow ϕ with $val(\phi) = \sum_{C \in \mathcal{C}_2'} g(C)$.*

Proof of Claim 2.4.12. Throughout this proof we use the notation set up in Definition 2.4.10 and Subsection 2.3.3. As $M_0 := \{sC \mid C \in \mathcal{C}_2'\}$ is the cut-set of a cut of $F = F(\mathcal{C}_2')$ of capacity $\sum_{C \in \mathcal{C}_2'} g(C)$, by Theorem 2.3.5 it remains to show that this is a minimum cut. We assume the existence of a cut-set M of F with smaller capacity and will show that this contradicts our assumption on the value of κ. Let $T \subseteq V(\mathcal{C}_2') \subseteq \dot{A}$ be the set of vertices that are separated from t by M and $T' := V(\mathcal{C}_2') \setminus T$. Let $S \subseteq \mathcal{C}_2'$ be the set of cycles which are not separated from s by M, and $S' := \mathcal{C}_2' \setminus S$. These sets are illustrated in Figure 7. Let D_S be the multidigraph that is the union of the cycles in S. We have that

$$w(M) = \sum_{C \in S'} g(C) + e_F(S, T') + |T|\kappa < \sum_{C \in \mathcal{C}_2'} g(C) = w(M_0),$$

which is equivalent to

$$\sum_{C \in S} g(C) > e_F(S, T') + |T|\kappa. \tag{2.4.2}$$

(Note that we may assume $T \neq \emptyset$, as otherwise (2.4.2) cannot hold[10].) Now observe that[11]

$$\sum_{C \in S} g(C) = \sum_{C \in S} \left(\left\lceil \frac{|V(C)|}{\kappa} \right\rceil + 1 \right) < \frac{e(D_S)}{\kappa} + 2|S|. \tag{2.4.3}$$

By (2.4.1), we have $|V(C)| > \kappa$ for all $C \in \mathcal{C}_2'$, so it follows that

$$|S| < \sum_{C \in S} |V(C)|/\kappa = e(D_S)/\kappa. \tag{2.4.4}$$

Combining (2.4.2), (2.4.3) and (2.4.4), it follows that

$$\frac{3e(D_S)}{\kappa} > e_F(S, T') + |T|\kappa. \tag{2.4.5}$$

Next, since $|V(C)| \leq N/\kappa$ for all $C \in \mathcal{C}_2'$ by (2.4.1) and $|\mathcal{C}_2'| = |\mathcal{C}_2| \leq 2N$, we have

$$e(D_S) < 2N^2/\kappa, \tag{2.4.6}$$

Furthermore, since $e(D_S) = e_{D_S}(T) + e_{D_S}(T') + e_{D_S}(T', T) + e_{D_S}(T, T')$, we have[12]

$$\begin{aligned} e_F(S, T') &= \sum_{v \in T'} \frac{1}{2}(d^+_{D_S}(v) + d^-_{D_S}(v)) \\ &\geq \frac{1}{2}(e_{D_S}(T') + e_{D_S}(T', T) + e_{D_S}(T, T')) = \frac{1}{2}(e(D_S) - e_{D_S}(T)). \end{aligned}$$

Combining this with (2.4.5), we have

$$\frac{6e(D_S)}{\kappa} > 2e_F(S, T') \geq e(D_S) - e_{D_S}(T),$$

which implies

$$e_{D_S}(T) \geq \left(1 - \frac{6}{\kappa}\right) e(D_S) \geq \frac{1}{2} e(D_S). \tag{2.4.7}$$

By the discussion before the claim concerning the construction of $\mathcal{C}_2'$, we have $e_{D_S}(T) \leq e_D(T) + \lambda|T|$. This implies that either $e_{D_S}(T) \leq 2e_D(T)$ or $e_{D_S}(T) \leq 2\lambda|T|$. If $e_{D_S}(T) \leq 2\lambda|T|$, then using (2.4.7) we obtain that $|T| \geq e(D_S)/(4\lambda)$, and combining this with (2.4.5) we have

$$\frac{3e(D_S)}{\kappa} > |T|\kappa \geq \frac{\kappa e(D_S)}{4\lambda},$$

so that $\kappa^2 < 12\lambda$, contradicting our choice of κ. Therefore, we may assume

$$e_{D_S}(T) \leq 2e_D(T). \tag{2.4.8}$$

Now we distinguish between the two cases in the statement of the lemma, i.e., when D is an (n, p, κ, λ)-digraph and when D is an (n, p, κ, λ)-pseudorandom digraph.

Case 1: D is an (n, p, κ, λ)-digraph. By (2.4.8) we have $e_{D_S}(T) \leq 2e_D(T) \leq 2|T|^2$. Combined with (2.4.7), we conclude that $|T| \geq \sqrt{e(D_S)/4}$. By (2.4.5), we have

$$\frac{3e(D_S)}{\kappa} > |T|\kappa \geq \kappa\sqrt{e(D_S)/4}$$

Combining this with (2.4.6), we obtain that

$$2N^2/\kappa > e(D_S) > \kappa^4/36,$$

contradicting our choice of $\kappa \geq (72N^2)^{1/5}$.

Case 2: D is an (n, p, κ, λ)-pseudorandom digraph. We further split this into two cases. Assume first that $|T| \geq \log n/(50p)$, so by (P5) and (2.4.8) we have that $e_{D_S}(T) \leq 2e_D(T) \leq 200|T|^2 p$. Combined with (2.4.7), we have that $|T| \geq \sqrt{e(D_S)/(400p)}$. By (2.4.5), we have

$$\frac{3e(D_S)}{\kappa} > |T|\kappa \geq \kappa\sqrt{e(D_S)/(400p)}$$

Combining this with (2.4.6), we obtain that

$$2N^2/\kappa > e(D_S) > \kappa^4/(3600p),$$

contradicting our choice of $\kappa \geq (7200N^2p)^{1/5}$.

We may thus assume that $|T| < \log n/(50p)$. In this case, we may consider any superset of T of size $\log n/(50p)$ and, by applying (P5) to this superset and considering (2.4.8), we have that $e_{D_S}(T) \leq 2e_D(T) \leq 2\log^2 n/(25p)$. Then, by (2.4.7),

$$e(D_S) \leq 4\log^2 n/(25p).$$

Now, using (2.4.5) and the fact that $T \neq \emptyset$, we also have that

$$e(D_S) > |T|\kappa^2/3 \geq \kappa^2/3.$$

But these two bounds on $e(D_S)$ lead to a contradiction on our choice of $\kappa \geq \sqrt{12/(25p)}\log n$. ◀

We interpret the flow given by Claim 2.4.12 as follows. As all capacities are integers, there exists an integer flow with value $\sum_{C\in\mathcal{C}_2'} g(C)$, so assume ϕ is such an integer flow. For each cycle $C \in \mathcal{C}_2$, writing $C' = i(C)$, let $V_C := \{v \in V(C') \subseteq V(C) \mid \phi(C'v) = 1\}$ be the vertices assigned to C. As ϕ saturates all edges sC', we have $|V_C| = g(C') = g(C)$. The capacity κ of the edges vt with $v \in \dot{A}$ ensures that no vertex is assigned to more than κ cycles of $\mathcal{C}_2$.

We will now iteratively assign two edges e_1, e_2 to each cycle $C \in \mathcal{C}_2$ so that $E(C) \cup \{e_1, e_2\}$ can be decomposed into two (A^+, A^-)-paths, where $e_1 \in f(v_1)$, $e_2 \in f(v_2)$ and $v_1, v_2 \in V_C$. We do this as follows using Lemma 2.4.7. Assume that, for some of the cycles in $\mathcal{C}_2$, we have already found two edges as described above, and assume that we next want to do this for $C \in \mathcal{C}_2$. We say that an edge $e \in E^{\mathrm{ab}}$ is *available* if it has not been assigned to any of the previous cycles. Then, for each $v \in V_C$, the number of edges $e \in f(v)$ that are available is at least $\kappa + 2$ (since no vertex is assigned to more than κ cycles and $\mathcal{A} = (E^{\mathrm{ab}}, f)$ is an $(\dot{A}, 2\kappa + 1)$-absorbing structure). Thus, we may define a $(V_C, \kappa + 2)$-absorbing structure $\mathcal{A}_C$ using available edges from E^{ab} by selecting, for each $v \in V_C$, any set of $\kappa + 2$ available edges at v. Then, with V_C and $\mathcal{A}_C$ playing the roles of S and $\mathcal{A}$, respectively, Lemma 2.4.7 gives two edges $e_1 \in f(v_1)$ and $e_2 \in f(v_2)$ with $v_1, v_2 \in V_C$ such that $E(C) \cup \{e_1, e_2\}$ can be decomposed into two (A^+, A^-)-paths. After repeating this for every cycle $C \in \mathcal{C}_2$ and treating each of the remaining edges in E^{ab} as an (A^+, A^-)-path, we have an edge decomposition of $E(\mathcal{C}_2) \cup E^{\mathrm{ab}}$ into (A^+, A^-)-paths, which is a perfect decomposition by Proposition 2.3.3. □

We have seen earlier in Lemma 2.4.7 how a single long or medium cycle can be absorbed using an absorbing structure. The following lemma shows how to absorb a single short cycle using our absorbing structure. In fact, it is slightly more general: it shows how to absorb a short Eulerian digraph, namely one that is the union of two short edge-disjoint paths. Another difference is that now we must work with vertices in A^0. As before, in order to absorb a cycle C, we take two suitable vertices $v_1, v_2 \in V(C)$. For long and medium cycles, both v_1 and v_2 had been in $\dot{A}$, and we used a single edge in $f(v_1)$ and a single edge in $f(v_2)$ for absorption. For short cycles, if $v_1, v_2 \in \dot{A}$, we do the same, but here one or both may be in A^0. If, for instance, $v_1 \in A^0$, we use a pair of edges from $f(v_1)$ (which should be thought of as an (A^+, A^-)-path of length two through v_1) for absorption.

Lemma 2.4.13. Let D be a (n, p, κ, λ)-digraph and $v_1, v_2 \in V(D)$. Let $P_1 \subseteq D$ be a (v_1, v_2)-path and $P_2 \subseteq D$ be a (v_2, v_1)-path which are edge-disjoint. Let $k \geq \max_{i\in[2]} |V(P_i) \cap \dot{A}|$. Let $\mathcal{A} = (E^{\mathrm{ab}}, f)$ be a $(\{v_1, v_2\}, k+1)$-absorbing structure such that, for each $i \in [2]$, it holds that $E(P_i) \cap E^{\mathrm{ab}} = \emptyset$. Then, for each $i \in [2]$ there exists a set $E_i \subseteq f(v_i)$, where $|E_i| = 1$ if $v_i \in \dot{A}$ and $|E_i| = 2$ otherwise, such that the digraph with edge set $E(P_1) \cup E(P_2) \cup E_1 \cup E_2$ can be decomposed into two (A^+, A^-)-paths.

Proof. For each $i \in [2]$, we consider three cases. If $v_i \in A^+$, by our choice of k, there is some edge $v_iy_i \in f(v_i)$ with $y_i \notin V(P_{3-i})$[13]. In such a case, we let $P_i^+ := v_iy_i$ and $P_i^- := v_i$. If $v_i \in A^-$, similarly, there is some edge $x_iv_i \in f(v_i)$ with $x_i \notin V(P_i)$, and we let $P_i^+ := v_i$ and $P_i^- := x_iv_i$. Otherwise, we have $v_i \in A^0$ and, again by assumption, there must be two edges $x_iv_i, v_iy_i \in f(v_i)$ such that $x_i \notin V(P_i)$ and $y_i \notin V(P_{3-i})$. In this case, we let $P_i^+ := v_iy_i$ and $P_i^- := x_iv_i$. In all cases we set $E_i := E(P_i^+) \cup E(P_i^-)$.

Now let $P := P_1^- P_1 P_2^+$ and $P' := P_2^- P_2 P_1^+$. Clearly, P and P' decompose $E(P_1) \cup E(P_2) \cup E_1 \cup E_2$. Furthermore, both P and P' are (A^+, A^-)-paths by the definition of $\mathcal{A}$ and our choice of x_i, y_i. Indeed, for each $i \in [2]$, by definition we have that the first vertex of P_i^- lies in A^+, and the last vertex of P_i^+ lies in A^-, which immediately yields the result. □

The following lemma shows how an absorbing structure can be used to absorb a collection of short cycles.

Lemma 2.4.14. Let D be an (n, p, κ, λ)-digraph with $\kappa \geq 4N/n$. Let $\mathcal{C}_3$ be a collection of at most N edge-disjoint cycles such that, for each

$C \in \mathcal{C}_3$, we have $|V(C) \cap \dot{A}| \leq \kappa$. Let $\mathcal{A} = (E^{\mathrm{ab}}, f)$ be an $(A^0 \cup \dot{A}, 3\kappa)$-absorbing structure with $E(\mathcal{C}_3) \cap E^{\mathrm{ab}} = \emptyset$. Then, the digraph with edge set $E(\mathcal{C}_3) \cup E^{\mathrm{ab}}$ can be decomposed into a set of cycles $\mathcal{C}^*$ and a digraph Q such that

(S1) $E(\mathcal{C}^*) \subseteq E(\mathcal{C}_3)$;

(S2) for all $C \in \mathcal{C}^*$ we have $|V(C) \cap \dot{A}| > \kappa$, and

(S3) Q has a perfect decomposition in which each path is an (A^+, A^-)-path.

Proof. We will construct Q and $\mathcal{C}^*$ over multiple rounds. We start with a set of cycles $\mathcal{C} := \mathcal{C}_3$ and a set of edges $F^{\mathrm{ab}} := E^{\mathrm{ab}}$, and we set $Q := (V(D), \emptyset)$ and $\mathcal{C}^* := \emptyset$. In each round, we will update $\mathcal{C}$, F^{ab}, Q and $\mathcal{C}^*$ by moving some edges from $E(\mathcal{C}) \cup F^{\mathrm{ab}}$ to $E(Q) \cup E(\mathcal{C}^*)$. In particular, in each round, we will combine edges from F^{ab} with some Eulerian subdigraph of $E(\mathcal{C})$ to form (A^+, A^-)-paths (by using Lemma 2.4.13) and move the (edges of these) paths into Q. Since we only ever add (A^+, A^-)-paths to Q, then Q always has a perfect decomposition by Proposition 2.3.3. (Throughout we will also maintain that F^{ab} can be decomposed into (A^+, A^-)-paths.) After these paths have been added to Q, what remains of $E(\mathcal{C})$ will be Eulerian and reside on a significantly smaller number of vertices. We will then apply Theorem 2.3.4 to decompose what remains of $E(\mathcal{C})$ into cycles: any medium or long cycle in this decomposition (i.e., those that have more than κ vertices in $\dot{A}$) will be added to $\mathcal{C}^*$, while the remaining cycles in the decomposition form the set $\mathcal{C}$ for the next round. Since $|V(\mathcal{C})|$ decreases in each round, this process will stop after a finite number of rounds. At that point, we add any remaining edges from F^{ab}, decomposed into (A^+, A^-)-paths, into Q, which will have a perfect decomposition.

It is important that we use edges/paths from our absorbing structure carefully in each round so that there are sufficiently many choices available at each vertex in future rounds. By solving a suitable flow problem, we will make sure that, over the course of all rounds, we use at most κ edges/paths from E^{ab} at each vertex. This will ensure there are always at least 2κ choices of edges/paths available in F^{ab} at every vertex in every round, which will allow us to construct suitable absorbing (sub)structures in order to apply Lemma 2.4.13.

Let us now give the details of this iterative process. At the start of each round we are given a digraph Q, a set of edges $F^{\mathrm{ab}} \subseteq E^{\mathrm{ab}}$ and two sets of cycles $\mathcal{C}$ and $\mathcal{C}^*$, which have been updated in previous rounds and satisfy the following properties:

(a) $E(\mathcal{C}_3) \cup E^{\mathrm{ab}}$ is the disjoint union of $E(Q)$, F^{ab}, $E(\mathcal{C})$, and $E(\mathcal{C}^*)$;

(b) Q can be decomposed into (A^+, A^-)-paths;

(c) writing $n' := |V(\mathcal{C})|$, we have $|\mathcal{C}| \leq c'n' \log n'$ (where c' is the constant from Theorem 2.3.4) and $|V(C) \cap \dot{A}| \leq \kappa$ for all $C \in \mathcal{C}$, and

(d) $|V(C) \cap \dot{A}| > \kappa$ for all $C \in \mathcal{C}^*$.

The digraph Q and the sets F^{ab} and $\mathcal{C}$ are updated several times throughout each round, and the notation will always refer to their updated form.

Recall that, as stated in Definition 2.4.4, we may think of $\mathcal{A} = (E^{\mathrm{ab}}, f)$ as a set of edge-disjoint paths of length 1 or 2. In the same way, we also think of the edges of F^{ab} as paths of length 1 or 2. For any $v \in A^+ \cup A^-$, we think of each edge in $F^{\mathrm{ab}} \cap f(v)$ as an (A^+, A^-)-path of length 1. Because of the way we use edges for absorption (i.e., by using Lemma 2.4.13), for any $v \in A^0$, the set $F^{\mathrm{ab}} \cap f(v)$ will always contain the same number of edges from A^+ to v as from v to A^-, and these will be (implicitly) paired up arbitrarily and thought of as (A^+, A^-)-paths of length 2. Note that the pairing is updated (arbitrarily) every time F^{ab} is updated. For each vertex v, let $a(v)$ denote the current number of available paths in $F^{\mathrm{ab}} \cap f(v)$, that is,

$$a(v) = \begin{cases} d^+_{F^{\mathrm{ab}} \cap f(v)}(v) & \text{if } v \in A^+, \\ d^-_{F^{\mathrm{ab}} \cap f(v)}(v) & \text{if } v \in A^-, \\ d^+_{F^{\mathrm{ab}} \cap f(v)}(v) = d^-_{F^{\mathrm{ab}} \cap f(v)}(v) & \text{if } v \in A^0. \end{cases}$$

As we want to use at most κ paths at each vertex $v \in V(\mathcal{C})$, we define the number of *ready paths* at v as $r(v) := a(v) - 2\kappa$. Throughout, we implicitly update the values of $a(v)$ and $r(v)$ each time we update F^{ab}.

We further assume the following property about $\mathcal{C}$ at the start of the round:

(e) for all $v \in V(\mathcal{C})$ we have at least one of $d^+_{\mathcal{C}}(v) \leq r(v)$, or $r(v) = \kappa$ (i.e., the number of cycles in $\mathcal{C}$ passing through v is bounded above by $r(v)$ or $r(v) = \kappa$).

Note that, at the start of the first round, we have $Q = (V(D), \emptyset)$, $F^{\mathrm{ab}} = E^{\mathrm{ab}}$, $\mathcal{C} = \mathcal{C}_3$ and $\mathcal{C}^* = \emptyset$, so (a)–(e) hold.

We now show how to update Q, $\mathcal{C}$, and $\mathcal{C}^*$ and check that (a)–(e) hold at the end of the round. Consider the flow problem $FP(\mathcal{C}; 2, \kappa)$ and let ϕ be a maximum integer flow. Let F_ϕ be the residual digraph of $F = F(\mathcal{C})$ under ϕ. Set $T := \{v \in V(\mathcal{C}) \mid F_\phi \text{ contains an } (s, v)\text{-path}\}$ and $T' := V(\mathcal{C}) \setminus T$.

We establish a bound on $|T|$ for later. Since the cut-set $M_0 := \{sC \mid C \in \mathcal{C}\}$, by (c), has capacity $2|\mathcal{C}| \leq 2c'n' \log n'$, the max-flow min-cut theorem (Theorem 2.3.5) implies that $val(\phi) \leq 2c'n' \log n'$. Furthermore, all vertices in T must have κ units of flow going through them in ϕ, as otherwise we would immediately be able to increase the flow. Therefore,

$$\kappa|T| \leq val(\phi) \leq 2c'n' \log n',$$

which implies

$$|T| \leq \frac{2c'n' \log n'}{\kappa} \leq \frac{n'}{2}, \tag{2.4.9}$$

as $\kappa \geq 4N/n \geq 4c' \log n'$.

We use ϕ to assign vertices to cycles as follows. First, we greedily decompose ϕ into single-unit flows. As each single-unit flow goes through one cycle $C \in \mathcal{C}$ and one vertex $v \in V(C)$, we understand this as assigning v to C. Note that for every $v \in V(\mathcal{C})$, the flow $\phi(vt)$ through the edge vt satisfies

$$\phi(vt) \leq \min\{d^+_{\mathcal{C}}(v), \kappa\} \leq r(v), \tag{2.4.10}$$

where the last inequality holds by (e).

We partition $\mathcal{C}$ into three sets $\mathcal{C} = \mathcal{C}^0 \cup \mathcal{C}^1 \cup \mathcal{C}^2$, where $\mathcal{C}^i$ is the set of cycles $C \in \mathcal{C}$ that are assigned exactly i vertices from T'. Recall that we decomposed the flow ϕ into single-unit flows. For each $i \in [2]_0$, let ϕ^i be the flow that is given by the sum of the single-unit flows of the decomposition that pass through cycles in $\mathcal{C}^i$. In particular, this means that $\phi = \phi^0 + \phi^1 + \phi^2$ and, for each $v \in V(\mathcal{C})$, the number of cycles in $\mathcal{C}^i$ to which v is assigned is $\phi^i(vt)$. We next show how to process the cycles in each $\mathcal{C}^i$, but first we need the following claim.

Claim 2.4.15. For all cycles $C \in \mathcal{C}^1$, we have $|V(C) \cap T'| = 1$ (and so the unique vertex in $V(C) \cap T'$ must be assigned to C). In particular, for all $v \in T'$ we have $\phi^1(vt) = d^+_{\mathcal{C}^1}(v)$.

For all cycles $C \in \mathcal{C}^0$, we have $|V(C) \cap T'| = 0$.

Proof of Claim 2.4.15. For all $C \in \mathcal{C}^0 \cup \mathcal{C}^1$, note first that there is a path from s to C in F_ϕ. Indeed, if C is assigned fewer than two vertices, then the path is immediate, while if C is assigned two vertices, at least one of them, say u, is in T, and so the (s,u)-path in F_ϕ (which exists by the definition of T) can be extended to C. Now any vertices $v \in V(C)$ that are not assigned to C must lie in T by definition, as we can extend the (s,C)-path in F_ϕ to v. Therefore, for each $i \in \{0,1\}$ and all $C \in \mathcal{C}^i(v)$ we must have $|V(C) \cap T'| = i$.

Now, any vertex $v \in T'$ that belongs to a cycle $C \in \mathcal{C}^1$ is also assigned to it, establishing that $\phi^1(vt) = d^+_{\mathcal{C}^1}(v)$ for all $v \in T'$. ◀

We start by processing the cycles in $\mathcal{C}^2$. For each cycle $C \in \mathcal{C}^2$, let $v_1, v_2 \in T'$ be such that $\phi(Cv_i) = 1$ for each $i \in [2]$, i.e., these are the vertices assigned to C. We split C into a (v_1,v_2)-path P_{12} and a (v_2,v_1)-path P_{21}. We select any $\kappa+1$ available paths at each v_i from F^{ab} to define a $(\{v_1,v_2\},\kappa+1)$-absorbing structure $\mathcal{A}_C$ (we show below that this is always possible). We then apply Lemma 2.4.13 to the paths P_{12}, P_{21} and the absorbing structure $\mathcal{A}_C$ with $k=\kappa$. Thus, for each $i \in [2]$ we obtain an available path $E_i \subseteq f(v_i) \cap F^{\mathrm{ab}}$ such that $E(P_{12}) \cup E(P_{21}) \cup E_1 \cup E_2$ can be decomposed into two (A^+,A^-)-paths P'_1 and P'_2. For each $i \in [2]$, we add the edges of P'_i to Q, remove E_i from F^{ab} and remove C from $\mathcal{C}$. We repeat this for all cycles in $\mathcal{C}^2$.

We now check that it is always possible to find the desired absorbing structure $\mathcal{A}_C$. Notice that, in order to process $\mathcal{C}^2$, the number of available paths that we use at any vertex v is the number of cycles of $\mathcal{C}^2$ to which v is assigned, which at the start of the round is $\phi^2(vt) \le \min\{d^+_{\mathcal{C}}(v), \kappa\} \le r(v) = a(v) - 2\kappa$ (by (2.4.10)). This means there are always 2κ available paths at every vertex each time we apply Lemma 2.4.13.

After processing $\mathcal{C}^2$, for any vertex $v \in T'$, we have used at most $\phi^2(vt)$ available paths from $f(v)$. Recalling that we always update $a(v)$, we now have for any $v \in T'$ that

$$a(v) \ge \phi(vt) + 2\kappa - \phi^2(vt) = \phi^1(vt) + 2\kappa = d^+_{\mathcal{C}^1}(v) + 2\kappa, \qquad (2.4.11)$$

where we have used (2.4.10) for the first inequality and Claim 2.4.15 for the last equality. The first equality holds as $\phi^0(vt) = 0$ by definition, since $v \in T'$. Note that (e) holds for the current value of $a(v)$ and the current set of cycles $\mathcal{C} = \mathcal{C}^0 \cup \mathcal{C}^1$, since $a(v)$ is unchanged for $v \in T$, and that (2.4.11) confirms (e) for $v \in T'$ (by using Claim 2.4.15 to note that $d^+_{\mathcal{C}^0}(v) = 0$ for all $v \in T'$).

Next we process cycles in $\mathcal{C}^1$. Recall that, by Claim 2.4.15, such cycles contain exactly one vertex of T'. Let R be an empty set of edges; this set will be updated while processing $\mathcal{C}^1$ and will always form an Eulerian digraph. We say a pair of cycles $C_1, C_2 \in \mathcal{C}^1$ is T-intersecting if $\emptyset \neq V(C_1) \cap V(C_2) \subseteq T$ (and thus their unique vertices in T' are distinct). Whenever we have a T-intersecting pair of cycles $C_1, C_2 \in \mathcal{C}^1$, we process them as follows. Let $v_1 \neq v_2$ be the vertices of C_1 and C_2 in T', respectively. Starting from v_1, let v_1' be the first vertex along C_1 in $V(C_1) \cap V(C_2)$ and define $P_{12} \coloneqq v_1 C_1 v_1' C_2 v_2$. Define v_2' analogously, and let $P_{21} \coloneqq v_2 C_2 v_2' C_1 v_1$. It is easy to see that P_{12} and P_{21} are edge-disjoint. Again, we construct a $(\{v_1, v_2\}, 2\kappa + 1)$-absorbing structure $\mathcal{A}_{C_1 C_2}$ by taking $2\kappa + 1$ available paths at each v_i from F^{ab}; this is always possible by (2.4.11), as we find an absorbing structure for v_i at most $d^+_{\mathcal{C}^1}(v_i)$ times. We apply Lemma 2.4.13 to the paths P_{12}, P_{21} and the absorbing structure $\mathcal{A}_{C_1 C_2}$ with $k = 2\kappa$ to obtain available paths $E_i \subseteq f(v_i) \cap F^{\mathrm{ab}}$, for $i \in [2]$, such that $E(P_{12}) \cup E(P_{21}) \cup E_1 \cup E_2$ can be decomposed into two (A^+, A^-)-paths P_1' and P_2'. For each $i \in [2]$, we add the edges of P_i' to Q and remove the edges of E_i from F^{ab}. The remaining edges of the cycles C_1 and C_2, namely $(E(C_1) \cup E(C_2)) \setminus (E(P_{12}) \cup E(P_{21}))$, are then added to the residual digraph R. Notice that the set of edges added to R is Eulerian, so R remains Eulerian. Furthermore, note that all edges added to R have both endpoints in T. Finally, we remove C_1 and C_2 from $\mathcal{C}$ (and from $\mathcal{C}^1$).

We repeat this as long as we can find a T-intersecting pair of cycles in $\mathcal{C}^1$. When no such pair can be found, then, among the remaining cycles of $\mathcal{C}^1$, any two either share a vertex in T' or are vertex-disjoint. This implies that, at this stage, the set $\overline{T} \coloneqq V(\mathcal{C}^1) \cap T'$ satisfies $|\overline{T}| \leq |T|/2$. (To see this, for each vertex $v \in \overline{T}$, pick a cycle $C_v \in \mathcal{C}^1$ containing v. Notice that each such cycle has all its (at least two) remaining vertices in T and, furthermore, the cycles C_v are vertex-disjoint.) We move all the remaining cycles of $\mathcal{C}$ (i.e., all that remain in $\mathcal{C}^1$ and all in $\mathcal{C}^0$) to R. Then, R is Eulerian and $V(R) \subseteq T \cup \overline{T}$ (recall any cycle in $\mathcal{C}^0$ has all its vertices in T by Claim 2.4.15). Then,

$$n'' \coloneqq |V(R)| \leq |T| + |\overline{T}| \leq 3|T|/2 \leq 3n'/4, \tag{2.4.12}$$

where the last inequality follows by (2.4.9).

Now we decompose R into at most $c'n'' \log n''$ cycles using Theorem 2.3.4; any resulting cycles with more than κ vertices in $\dot{A}$ are added to $\mathcal{C}^*$, while all other cycles are added to (the currently empty) $\mathcal{C}$. This completes the round and the description of the sets $\mathcal{C}$, $\mathcal{C}^*$, F^{ab}, and Q ready for the next round. Notice that at the end of the round $V(\mathcal{C})$ is smaller than at the start, by (2.4.12). It remains to check that (a)–(e) hold.

It immediately follows by construction that (a)–(d) hold ((a) holds because we only move edges between the sets, and (b) holds because we only add (A^+, A^-)-paths to Q). Finally, we prove that (e) holds too. As noted after (2.4.11), we know (e) holds after $\mathcal{C}^2$ is processed. After that, when processing $\mathcal{C}^1$, whenever an application of Lemma 2.4.13 reduces $a(v)$ by 1, it also reduces $d^+_{\mathcal{C}}(v)$ by 1, so condition (e) is maintained to the end of the round.

Thus, we may iterate the described process through the rounds, until we obtain the final sets Q, $\mathcal{C} = \emptyset$, $\mathcal{C}^*$ and F^{ab} satisfying (a)–(e). (Recall that the process must terminate since, by (2.4.12), the set of cycles that is considered for each subsequent round is contained in a smaller set of vertices than the previous.) The remaining paths of F^{ab} are (A^+, A^-)-paths; these paths are removed from F^{ab} and added to Q.

It is straightforward to check that Q and $\mathcal{C}^*$ now satisfy the conclusion of the lemma. Indeed, over the course of all rounds, we moved all edges from $E(\mathcal{C}_3) \cup E^{\mathrm{ab}}$ to $E(Q) \cup E(\mathcal{C}^*)$. At every stage, Q was updated by adding (A^+, A^-)-paths (which gives a perfect decomposition of Q by Proposition 2.3.3), and $\mathcal{C}^*$ was updated by adding cycles that have more than κ vertices in $\dot{A}$. □

We are finally ready to prove the main result.

Proof of Theorem 2.4.3. Recall that D is either an (n, p, κ, λ)-digraph satisfying (C1)–(C3) or an (n, p, κ, λ)-pseudorandom digraph satisfying (C′1), (C′2), and (C3), with $n \geq n_0$ (for a suitably large choice of n_0). We work with both cases simultaneously.

First, one can easily check that the conditions (C1)–(C3) together with $n \geq n_0$, for a sufficiently large n_0, imply the conditions (a)–(c) below, which are precisely the parameter conditions required in order to apply Lemmas 2.4.5, 2.4.6, 2.4.9, 2.4.11 and 2.4.14 to an (n, p, κ, λ)-digraph:

(a) $\max\{100 \log n, 12, (12\lambda)^{1/2}, (72N^2)^{1/5}, 4N/n\}$
$\leq \kappa < \min\{np/120, N^{1/2}\}$,

(b) $\lambda \leq np/3$,

(c) $4np\log(2n) \leq \kappa\lambda$,

where $N := c'n\log n$ and c' is the constant from Theorem 2.3.4. [14] Similarly, one can easily check that the conditions (C′1), (C′2), and (C3) together with $n \geq n_0$, for a sufficiently large n_0, imply the conditions (a′), (b), and (c) (with (a′) given below), which are precisely the parameter conditions required in order to apply Lemmas 2.4.5, 2.4.6, 2.4.9, 2.4.11 and 2.4.14 to an (n, p, κ, λ)-pseudorandom digraph:

(a′) $\max\{100\log n, 12, (12\lambda)^{1/2}, (7200N^2p)^{1/5}, \frac{4N}{n}, \sqrt{12/(25p)}\log n\} \leq \kappa < \min\{\frac{np}{120}, N^{1/2}\}$.

The pseudorandom case only makes a difference for Lemma 2.4.11. [15]

For the (n, p, κ, λ)-(pseudorandom) digraph D, let $A^+ \cup A^- \cup A^0$ be the associated partition of $V(D)$. Write B^+, B^-, and B^0 for the set of vertices $v \in V(D)$ such that $\mathrm{ex}_D(v) > 0$, $\mathrm{ex}_D(v) < 0$, and $\mathrm{ex}_D(v) = 0$, respectively. From Definition 2.4.1, clearly $A^+ \subseteq B^+$ and $A^- \subseteq B^-$.

Let $\dot{\mathcal{A}} = (\dot{E}^{\mathrm{ab}}, \dot{f})$ be an $(\dot{A}, 12\kappa)$-absorbing structure contained in D, which exists by Lemma 2.4.5, and let $\mathcal{A}^0 = (E_0^{\mathrm{ab}}, f_0)$ be an $(A^0, 3\kappa)$-absorbing structure contained in D, which exists by Lemma 2.4.6. Note that these two absorbing structures must be edge-disjoint by definition. We next split up $\dot{\mathcal{A}}$ into an $(\dot{A}, 7\kappa - 1)$-absorbing structure $\dot{\mathcal{A}}_1 = (\dot{E}_1^{\mathrm{ab}}, \dot{f}_1)$, an $(\dot{A}, 2\kappa + 1)$-absorbing structure $\dot{\mathcal{A}}_2 = (\dot{E}_2^{\mathrm{ab}}, \dot{f}_2)$, and an $(\dot{A}, 3\kappa)$-absorbing structure $\dot{\mathcal{A}}_3 = (\dot{E}_3^{\mathrm{ab}}, \dot{f}_3)$. To do so, for each $v \in \dot{A}$, we arbitrarily split the 12κ edges in $\dot{f}(v)$ into sets of size $7\kappa - 1$, $2\kappa + 1$ and 3κ and set these to be $\dot{f}_1(v)$, $\dot{f}_2(v)$ and $\dot{f}_3(v)$, respectively, and set $\dot{E}_i^{\mathrm{ab}} := \bigcup_{v\in\dot{A}} \dot{f}_i(v)$ for each $i \in [3]$. Lastly, we combine $\dot{\mathcal{A}}_3$ and $\mathcal{A}^0$ into an $(\dot{A} \cup A^0, 3\kappa)$-absorbing structure $\mathcal{A}_3 = (\dot{E}_3^{\mathrm{ab}} \cup E_0^{\mathrm{ab}}, f_3)$, where $f_3|_{\dot{A}} = \dot{f}_3$ and $f_3|_{A^0} = f_0$.

Consider a set of paths which consists of every individual edge in $\dot{E}^{\mathrm{ab}}$ and a partition of the edges in E_0^{ab} into paths of length two. Each path is an (A^+, A^-)-path and, therefore, a (B^+, B^-)-path. Moreover, note that, by Lemmas 2.4.5 and 2.4.6, $\dot{E}^{\mathrm{ab}} \cup E_0^{\mathrm{ab}}$ contains at most $150\kappa + 5\kappa = 155\kappa$ edges incident to each $v \in \dot{A}$. This means (by (P1) and (P2)) that removing all these paths from D will not change the sign of the excess of any vertex $v \in V(D)$, that is, if we write $D' := D \setminus (\dot{E}^{\mathrm{ab}} \cup E_0^{\mathrm{ab}})$, then a vertex of positive (resp. negative) excess in D' belongs to B^+ (resp. B^-).

Next, we greedily remove paths from D' that start in vertices with positive excess in D' and end in vertices with negative excess in D' until this is no longer possible. We call the set of these paths $\mathcal{P}'$ (so every path in $\mathcal{P}'$ is a (B^+, B^-)-path) and set $D^* := D' \setminus E(\mathcal{P}')$ (so that $\mathrm{ex}(D^*) = 0$ by Proposition 2.3.2).

We apply Theorem 2.3.4 to every component of D^* and obtain a decomposition $\mathcal{C}$ of the edges of D^* into at most $N := c'n \log n$ cycles. Let

$$\begin{aligned} \mathcal{C}_1 &:= \{C \in \mathcal{C} : |V(C) \cap \dot{A}| \geq N/\kappa\}, \\ \mathcal{C}_2 &:= \{C \in \mathcal{C} : \kappa < |V(C) \cap \dot{A}| < N/\kappa\}, \text{ and} \\ \mathcal{C}_3 &:= \{C \in \mathcal{C} : |V(C) \cap \dot{A}| \leq \kappa\}. \end{aligned}$$

At this point, we have

$$\begin{aligned} E(D) &= E(D') \cup \dot{E}^{\mathrm{a}b} \cup E_0^{\mathrm{a}b} \\ &= E(D') \cup \dot{E}_1^{\mathrm{a}b} \cup \dot{E}_2^{\mathrm{a}b} \cup \dot{E}_3^{\mathrm{a}b} \cup E_0^{\mathrm{a}b} \\ &= E(\mathcal{P}') \cup E(D^*) \cup \dot{E}_1^{\mathrm{a}b} \cup \dot{E}_2^{\mathrm{a}b} \cup \dot{E}_3^{\mathrm{a}b} \cup E_0^{\mathrm{a}b} \\ &= E(\mathcal{P}') \cup \left(E(\mathcal{C}_1) \cup \dot{E}_1^{\mathrm{a}b}\right) \cup \left(E(\mathcal{C}_2) \cup \dot{E}_2^{\mathrm{a}b}\right) \cup \left(E(\mathcal{C}_3) \cup \dot{E}_3^{\mathrm{a}b} \cup E_0^{\mathrm{a}b}\right). \end{aligned}$$

Noting that $|\mathcal{C}_3| \leq N$, we apply Lemma 2.4.14 to $\mathcal{C}_3$ and $\mathcal{A}_3$ to decompose the edges of $E(\mathcal{C}_3) \cup \dot{E}_3^{\mathrm{a}b} \cup E_0^{\mathrm{a}b}$ into a set of cycles $\mathcal{C}_3^*$ and a digraph Q, where Q has a perfect decomposition $\mathcal{P}_3$ into (A^+, A^-)-paths, $|\mathcal{C}_3^*| \leq |\mathcal{C}_3|$, and for all $C \in \mathcal{C}_3^*$ we have $|V(C) \cap \dot{A}| > \kappa$. (Indeed, the fact that $|\mathcal{C}_3^*| \leq |\mathcal{C}_3|$ follows from conclusions (S1) and (S2) of Lemma 2.4.14. To see this, note that any cycle $C \subseteq D$ satisfies $|\{uv \in E(C) \mid v \in \dot{A}\}| = |V(C) \cap \dot{A}|$. Thus, by (S2), for each $C \in \mathcal{C}_3^*$ and each $C' \in \mathcal{C}_3$ we must have $|\{uv \in E(C) \mid v \in \dot{A}\}| > |\{uv \in E(C') \mid v \in \dot{A}\}|$, so by (S1), $\mathcal{C}^*$ must have fewer cycles than $\mathcal{C}_3$.) Let $\mathcal{C}_1^* := \{C \in \mathcal{C}_3^* : |V(C) \cap \dot{A}| \geq N/\kappa\}$ and $\mathcal{C}_2^* := \{C \in \mathcal{C} \mid \kappa < |V(C) \cap \dot{A}| < N/\kappa\}$ and note that, as $|\mathcal{C}_3^*| \leq |\mathcal{C}_3|$, we have $|\mathcal{C}_1^*|, |\mathcal{C}_2^*| \leq N$.

Next, we apply Lemma 2.4.9 to $\mathcal{C}_1 \cup \mathcal{C}_1^*$ and $\dot{\mathcal{A}}_1$; this shows that the digraph with edge set $E(\mathcal{C}_1 \cup \mathcal{C}_1^*) \cup \dot{E}_1^{\mathrm{ab}}$ has a perfect decomposition $\mathcal{P}_1$ into (A^+, A^-)-paths.

In the same way, applying Lemma 2.4.11 to $\mathcal{C}_2 \cup \mathcal{C}_2^*$ and $\dot{\mathcal{A}}_2$ shows that the digraph with edge set $E(\mathcal{C}_2 \cup \mathcal{C}_2^*) \cup \dot{E}_2^{\mathrm{ab}}$ has a perfect decomposition $\mathcal{P}_2$ into (A^+, A^-)-paths.

Now it is easy to check that $\mathcal{P}' \cup \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{P}_3$ is a decomposition of $E(D)$ into paths, and every path is a (B^+, B^-)-path, so this is a perfect decomposition of D by Proposition 2.3.3. □

2.5 PATH DECOMPOSITIONS OF RANDOM DIGRAPHS

In this section we derive Theorem 2.1.2. This will follow immediately as a consequence of Theorem 2.4.3 and the following result.

Theorem 2.5.1. Let $13\log^2 n/\sqrt{n} \leq p \leq 1 - 150\log^4 n/n$. Let $\kappa := \sqrt{np(1-p)}/(155\log^{3/4} n)$ and $\lambda := 5\sqrt{n/(1-p)}\log^2 n$. Then, a.a.s. $D_{n,p}$ is an (n,p,κ,λ)-pseudorandom digraph.

Now we prove Theorem 2.1.2.

Proof of Theorem 2.1.2. Let $\log^4 n/n^{1/3} \leq p \leq 1 - \log^{5/2} n/n^{1/5}$, (so within the range stated in the theorem), and let n be sufficiently large. As usual, let $N := c'n\log n$, where c' is the constant from Theorem 2.3.4.

If we let $D = D_{n,p}$, then by Theorem 2.5.1 we have that a.a.s. D is an (n,p,κ,λ)-pseudorandom digraph, where $\kappa = \sqrt{np(1-p)}/(155\log^{3/4} n)$ and $\lambda = 5\sqrt{n/(1-p)}\log^2 n$. As mentioned in Remark 2.4.2, D is also an (n,p,κ',λ')-pseudorandom digraph for any $\kappa' \leq \kappa$ and any $\lambda' \geq \lambda$. Taking $\kappa' = 6(N^2p)^{1/5}$ and $\lambda' = \min\{np/3, (\kappa')^2/12\}$, and checking that $\kappa' < \kappa$ and $\lambda' > \lambda$ for the given range of p [16] and n sufficiently large, we have that D is an (n,p,κ',λ')-pseudorandom digraph, so we can apply Theorem 2.4.3 to conclude that D has a perfect decomposition (that is, it is consistent). □

In order to prove Theorem 2.5.1, we will show that each of the properties of Definition 2.4.1 holds a.a.s. First, we require some properties about the edge distribution in $D_{n,p}$.

Lemma 2.5.2. There exists a constant $C > 0$ such that, for all $p \geq C\log n/n$, a.a.s. the digraph $D = D_{n,p}$ satisfies that, for all $A \subseteq V(D)$ with $|A| \geq \log n/(50p)$, we have

$$e_D(A) < 100|A|^2p.$$

Proof. Fix some $\log n/(50p) \leq i \leq n$, and fix a set $A \subseteq V(D)$ with $|A| = i$. Let $X := e_D(A)$, so $\mathbb{E}[X] = (1-1/i)i^2p$. A direct application of Lemma 2.3.7 shows that, for sufficiently large n,[17]

$$\mathbb{P}[X \geq 100i^2p] \leq (e/100)^{99i^2p}.$$

Now consider all sets A with $|A| = i$, and let $\mathcal{E}_i$ be the event that at least one of these sets induces at least $100i^2p$ edges. By a union bound, it follows that[18]

$$\mathbb{P}[\mathcal{E}_i] \leq \binom{n}{i}\left(\frac{e}{100}\right)^{99i^2p} \leq \left(\frac{en}{i}\right)^i\left(\frac{e}{100}\right)^{99i^2p} \leq \frac{1}{n^3},$$

where one can check the final inequality using the lower bound on i. The conclusion follows by a union bound over all values of i. □

Lemma 2.5.3. There exist constants $C, c > 0$ such that, for all $C\log n/n \leq p \leq 1 - C\log n/n$, with probability at least $1 - o(1/n^3)$ the digraph $D = D_{n,p}$ satisfies that, for all $v \in V$, we have

$$d^+_D(v) = np \pm c\sqrt{np(1-p)\log n} \quad \text{and} \quad d^-_D(v) = np \pm c\sqrt{np(1-p)\log n}.$$

Proof. We split the proof into two cases. Assume first that $p \leq 1/2$. Fix a vertex $v \in V(D)$ and a symbol $* \in \{+, -\}$. Then, $\mathbb{E}[d^*_D(v)] = (n-1)p$ and, if C and n are sufficiently large (we need C to be sufficiently large so that the value of δ in Lemma 2.3.6 satisfies $\delta \in (0, 1)$), by Lemma 2.3.6 we conclude that[19]

$$\mathbb{P}\left[d^*_D(v) \neq np \pm c\sqrt{np(1-p)\log n}\right] \leq e^{-c^2\log n/50}.$$

Now, by a union bound over all choices of v and $*$, it follows that the probability that the statement fails is at most[20] $2ne^{-c^2\log n/50} = o(1/n^3)$ (where this equality holds for sufficiently large c; $c \geq 15$ suffices).

For the second case, assume $p > 1/2$, and consider the complement digraph $\overline{D} \sim D_{n,1-p}$. We have that $1 - p < 1/2$, so we can apply the same argument as above to obtain that, for each $v \in V(D)$ and $* \in \{+, -\}$,

$$\mathbb{P}\left[d^*_{\overline{D}}(v) \neq n(1-p) \pm c\sqrt{np(1-p)\log n}\right] \leq e^{-c^2\log n/50}.$$

The conclusion follows by a union bound over all $v \in V(D)$ and $* \in \{+, -\}$ and going back to D[21]. □

Our next aim is to show that most vertices will have ‘high’ excess, meaning that its absolute value is ‘close’ to the maximum possible value (around $\sqrt{np(1-p)}$, up to a polylog factor) that follows from Lemma 2.5.3. The following remark will come in useful.

Remark 2.5.4. Let $p \in [0,1]$ and $n \in \mathbb{Z}$ with $n \geq 0$. Let $X \sim \mathrm{Bin}(n,p)$. For each $i \in \mathbb{Z}$, let $p_i := \mathbb{P}[X = i]$. Let D be a digraph and $v \in V(D)$ be such that $d^+(v) = d^-(v) = n$. Let D_p be a random subdigraph of D obtained by deleting each edge of D with probability $1-p$ independently of all other edges. Then, $\mathrm{ex}_{D_p}(v)$ follows a probability distribution which, for each $i \in \{-n, \dots, n\}$, satisfies that

$$\mathbb{P}[\mathrm{ex}_{D_p}(v) = i] = \sum_{j=0}^{n} p_j p_{j-i}.$$

In particular, the probability function is symmetric around $i = 0$.

Lemma 2.5.5. Consider the setting described in Remark 2.5.4, and assume $n \geq 2$. Then, there exists an absolute constant K such that

$$\mathbb{P}[\mathrm{ex}_{D_p}(v) = 0] \leq K\sqrt{\frac{\log n}{np(1-p)}}.$$

Proof. First note that, by adjusting the value of K, we may assume that n is larger than any fixed n_0 (by making the right hand side above greater than 1)[22]; we choose a sufficiently large n_0 so that all subsequent claims hold. By similarly adjusting the value of K, for any given constant C_0 we may assume that $C_0 \log n/n \leq p \leq 1 - C_0 \log n/n$[23].

So assume $C \log n/n \leq p \leq 1 - C \log n/n$, for a constant C defined below. One can readily check that $p^* := \max_{i \in [n]_0} p_i$ is achieved for $i = np \pm 2$[24] (where the p_i are as defined in Remark 2.5.4). By using Stirling's approximation, it follows that $p^* \leq 1/\sqrt{np(1-p)}$[25]. On the other hand, by an application of Lemma 2.3.6, there exist constants $c, C > 0$ such that for all $C \log n/n \leq p \leq 1 - C \log n/n$ we have that[26]

$$\sum_{i=0}^{np - c\sqrt{np(1-p)\log n}} p_i + \sum_{i=np+c\sqrt{np(1-p)\log n}}^{n} p_i \leq e^{-c^2 \log n/50}.$$

Combining the above with Remark 2.5.4, it follows that[27]

$$\mathbb{P}[\mathrm{ex}_{D_p}(v) = 0] \leq 2c\sqrt{np(1-p)\log n} \cdot (p^*)^2 + e^{-c^2 \log n/50} \leq K\sqrt{\frac{\log n}{np(1-p)}}. \qquad \square$$

Lemma 2.5.6. There exists a constant $C > 0$ such that, for all $C \log n/n \leq p \leq 1 - C \log n/n$, a.a.s. the digraph $D = D_{n,p}$ contains at most $n/\log^{1/8} n$ vertices v such that $|\mathrm{ex}_D(v)| \leq \sqrt{np(1-p)}/\log^{3/4} n$.

Proof. Take some vertex $v \in V(D)$. For each $i \in \mathbb{Z}$, let $p_i := \mathbb{P}[d_D^+(v) = i] = \mathbb{P}[d_D^-(v) = i]$. Now, by Remark 2.5.4 we have that $q_0 := \mathbb{P}[\mathrm{ex}_D(v) = 0] = \sum_{j=0}^{n-1} p_j^2$ and, for all $i \in [n-1]$, we have that $q_i := \mathbb{P}[|\mathrm{ex}_D(v)| = i] = \sum_{j=0}^{n-1} p_j(p_{j-i} + p_{j+i})$. In particular, by Lemma 2.3.9, it follows that

$$q_i \leq 2q_0 \tag{2.5.1}$$

for all $i \in [n-1]$. By combining this with Lemma 2.5.5 (with $n-1$ playing the role of n), it follows that

$$\mathbb{P}[|\mathrm{ex}_D(v)| \leq \sqrt{np(1-p)}/\log^{3/4} n] = \mathcal{O}(1/\log^{1/4} n). \tag{2.5.2}$$

Let $Y := |\{v \in V(D) : |\mathrm{ex}_D(v)| \leq \sqrt{np(1-p)}/\log^{3/4} n\}|$. The statement follows by applying Markov's inequality to this random variable[28]. □

We consider a partition of the vertices of $D_{n,p}$ into those with high excess, low excess, and the rest. In general, given $D = D_{n,p}$, we write

$$\begin{aligned} A^+ &= A^+(D) := \{v \in V(D) \mid \mathrm{ex}_D(v) \geq \sqrt{np(1-p)}/\log^{3/4} n\}, \\ A^- &= A^-(D) := \{v \in V(D) \mid \mathrm{ex}_D(v) \leq -\sqrt{np(1-p)}/\log^{3/4} n\} \text{ and} \\ A^0 &= A^0(D) := V(D) \setminus (A^+ \cup A^-). \end{aligned}$$

Corollary 2.5.6 shows that $|A^0| = o(n)$, and it is reasonable to expect that A^+ and A^- have roughly the same size. Even more, we will need the property that, a.a.s., all vertices have roughly the expected number of neighbors in the sets A^+ and A^-, as we show next.

Lemma 2.5.7. There exists a constant $C > 0$ such that, for all $C \log n/n \leq p \leq 1 - C\log n/n$, a.a.s. the graph $D = D_{n,p}$ satisfies that, for all $v \in V(D)$,

$$e(v, A^+), e(v, A^-), e(A^+, v), e(A^-, v) = np/2 \pm 2\sqrt{n/(1-p)}\log^2 n.$$

Proof. Let $V := V(D)$, and let $E := \{uv \mid u, v \in V, u \neq v\}$. Let $N := \binom{n}{2} = |E|/2$. For each $k \in [n-1]_0$, let $Z_k \sim \mathrm{Bin}(k, p)$ and, for each $j \in \mathbb{Z}$, let $p_j^{(k)} := \mathbb{P}[Z_k = j]$.

We begin by setting some notation. Consider any labelling $e_1, \ldots, e_N$ of all (unordered) pairs of distinct vertices $e = \{u, u'\}$ with $u, u' \in V$. We will later reveal the edges in succession following one such labelling. For each $i \in$

$[N]$, let $e_i = \{u_i, u_i'\}$, define $e_i^1 := u_iu_i'$ and $e_i^2 := u_i'u_i$ (the choice of e_i^1 and e_i^2 is arbitrary), and consider the random variable $X_i := (X_i^1, X_i^2)$, where X_i^1 and X_i^2 are indicator random variables for the events $\{e_i^1 \in E(D)\}$ and $\{e_i^2 \in E(D)\}$, respectively. For each $i \in [N]_0$, let $D^i := (V, E^i)$, where $E^i := \bigcup_{j\in[i]}\{e_j^1, e_j^2\}$. We set $D^i_{\text{cond}} := (V, E^i_{\text{cond}})$ to be the subdigraph of D^i with $E^i_{\text{cond}} := \{e_j^1 \mid j \in [i], X_j^1 = 1\} \cup \{e_j^2 \mid j \in [i], X_j^2 = 1\}$. (That is, without conditioning, D^i_{cond} is a random subdigraph of D^i where each edge is retained with probability p independently of all other edges, and it becomes a deterministic graph after conditioning on the outcomes of $X_1, \ldots, X_i$.) We also define $D_p^i := (V, E_{i,p})$, where $E_{i,p} \subseteq E \setminus E^i$ is obtained by adding each edge of $E \setminus E^i$ with probability p, independently of all other edges. In particular, for any $i \in [N]_0$ and any digraph F on V such that $D^i_{\text{cond}} \subseteq F$, we have that $\mathbb{P}[D_{n,p} = F \mid X_1, \ldots, X_i] = \mathbb{P}[D_p^i = F \setminus D^i_{\text{cond}}]$. For each $i \in [N]_0$ and each $v \in V$, we define $k_i(v) := n - 1 - |\{u \in V \mid uv \in E^i\}|$. This is the number of (pairs of) edges incident to v which have not been revealed after revealing $X_1, \ldots, X_i$. Thus, by Remark 2.5.4, the variable $\mathrm{ex}_{D_p^i}(v)$ follows a probability distribution which, for each $j \in \mathbb{Z}$, satisfies that

$$\mathbb{P}[\mathrm{ex}_{D_p^i}(v) = j] = \sum_{\ell=0}^{k_i(v)} p_\ell^{(k_i(v))} p_{\ell-j}^{(k_i(v))}. \tag{2.5.3}$$

Observe that, by Lemma 2.3.9 (in a similar way to (2.5.1)), for all $i \in [N]_0$, $v \in V$ and $j \in \mathbb{Z}$ we have that

$$\mathbb{P}[\mathrm{ex}_{D_p^i}(v) = j] \leq q_0^{(k_i(v))} := \mathbb{P}[\mathrm{ex}_{D_p^i}(v) = 0] = \sum_{\ell=0}^{k_i(v)} \left(p_\ell^{(k_i(v))}\right)^2. \tag{2.5.4}$$

Furthermore, observe the following. Choose a vertex $v \in V$ and an index $i \in [N-1]_0$ such that $d^+_{D^{i+1}}(v) - d^+_{D^i}(v) = 1$, and let $a \in \mathbb{Z}$. Then,[29]

$$\begin{aligned}\mathbb{P}[\mathrm{ex}_{D_p^i}(v) \geq a+1 \mid X_1, \ldots, X_i] &\leq \mathbb{P}[\mathrm{ex}_{D_p^{i+1}}(v) \geq a \mid X_1, \ldots, X_{i+1}] \\ &\leq \mathbb{P}[\mathrm{ex}_{D_p^i}(v) \geq a-1 \mid X_1, \ldots, X_i].\end{aligned}$$

(Note that the events above are actually independent from the variables upon which we condition. This notation, however, makes the statement more intuitive and is what we will require later in the proof.) In particular, this means that[30]

$$\left|\mathbb{P}[\mathrm{ex}_{D_p^i}(v) \geq a \mid X_1, \ldots, X_i] - \mathbb{P}[\mathrm{ex}_{D_p^{i+1}}(v) \geq a \mid X_1, \ldots, X_{i+1}]\right| \leq q_0^{(k_i(v))}.$$

(2.5.5)

(Indeed, we may bound $\mathbb{P}[\mathrm{ex}_{D_p^{i+1}}(v) \geq a \mid X_1, \dots, X_{i+1}]$ by one of the two terms in the previous expression, which gives us two cases to consider. In either of the cases, the difference becomes equal to the probability that $\mathrm{ex}_{D_p^i}(v)$ takes a specific value, which is in turn bounded by (2.5.4).)

Fix a vertex $v \in V$ and reveal all of its in- and outneighbors. Label all pairs of distinct vertices e as $e_1, \dots, e_N$ in such a way that, first, we have all pairs containing v, and then the rest, in any arbitrary order. In particular, we have already revealed the outcome of $X_1, \dots, X_{n-1}$. Let $\mathcal{E}$ be the event that $d_D^+(v), d_D^-(v) = np \pm c\sqrt{np(1-p)\log n}$, where c is the constant from the statement of Lemma 2.5.3. By Lemma 2.5.3, we have that $\mathbb{P}[\mathcal{E}] \geq 1 - 1/n^3$. Condition on this event. We will denote probabilities in this conditional space by $\mathbb{P}'$, and expectations by $\mathbb{E}'$. Observe that the variables $X_n, \dots, X_N$ are independent of $\mathcal{E}$, so for all events that only involve these variables we have that $\mathbb{P}' = \mathbb{P}$. Then, for all $u \in N_D^+(v)$ we have that $\mathrm{ex}_D(u) = \mathrm{ex}_{D_{\mathrm{cond}}^{n-1}}(u) + \mathrm{ex}_{D_p^{n-1}}(u)$, where $\mathrm{ex}_{D_{\mathrm{cond}}^{n-1}}(u) = 0$ if $u \in N_D^-(v)$ and $\mathrm{ex}_{D_{\mathrm{cond}}^{n-1}}(u) = -1$ otherwise, and $\mathrm{ex}_{D_p^{n-1}}(u)$ follows a probability distribution which, by (2.5.3), for each $j \in \{2-n, \dots, n-2\}$ satisfies that

$$\mathbb{P}'[\mathrm{ex}_{D_p^{n-1}}(u) = j] = \sum_{\ell=0}^{n-2} p_\ell^{(n-2)} p_{\ell-j}^{(n-2)}.$$

By a similar argument as the one used to obtain (2.5.2), i.e., combining (2.5.4) and the above with Lemma 2.5.5 (with $n-2$ playing the role of n), it follows that, for all $u \in V \setminus \{v\}$[31],

$$\mathbb{P}'[|\mathrm{ex}_{D_p^{n-1}}(u)| \geq \sqrt{np(1-p)}/\log^{3/4} n] = 1 - \mathcal{O}(1/\log^{1/4} n)$$

and, therefore, one easily deduces (by symmetry and conditioning on the event that $|\mathrm{ex}_{D_p^{n-1}}(u)| \geq \sqrt{np(1-p)}/\log^{3/4} n$) that[32]

$$\mathbb{P}'[u \in A^+] = 1/2 - \mathcal{O}(1/\log^{1/4} n). \tag{2.5.6}$$

Consider the edge-exposure martingale given by the variables $Y_i :=$ $\mathbb{E}'[|A^+ \cap N_D^+(v)| \mid X_1, \dots, X_i]$, for $i \in [N] \setminus [n-2]$. By (2.5.6), it follows that $Y_{n-1}, \dots, Y_N$ is a Doob martingale with $Y_{n-1} = \mathbb{E}'[|A^+ \cap N_D^+(v)|] =$ $(1 \pm 2c\sqrt{(1-p)\log n/(np)} - \mathcal{O}(1/\log^{1/4} n))np/2$ and $Y_N = |A^+ \cap N_D^+(v)|$. In order to prove that Y_N is concentrated around Y_{n-1}, we need to bound the martingale differences with a view to applying Lemma 2.3.8. Observe that, for all $i \in [N] \setminus [n-2]$, we have that $Y_i = \sum_{u \in N_D^+(v)} \mathbb{P}'[u \in A^+ \mid X_1, \dots, X_i]$.

For all $i \in [N-1] \setminus [n-2]$ such that $e_{i+1} \cap N_D^+(v) = \emptyset$, we have that $Y_{i+1} = Y_i$, and we set

$$c_i \coloneqq |Y_{i+1} - Y_i| = 0. \tag{2.5.7}$$

Consider now any $i \in [N-1] \setminus [n-2]$ such that $e_{i+1} = \{u, u'\}$ satisfies that $e_{i+1} \cap N_D^+(v) = \{u\}$. Then,

$$\begin{aligned}
&Y_{i+1} - Y_i \\
&\quad = \mathbb{P}'[u \in A^+ \mid X_1, \dots, X_{i+1}] - \mathbb{P}'[u \in A^+ \mid X_1, \dots, X_i] \\
&\quad = \mathbb{P}'[\mathrm{ex}_D(u) \geq \sqrt{np(1-p)}/\log^{3/4} n \mid X_1, \dots, X_{i+1}] \\
&\qquad - \mathbb{P}'[\mathrm{ex}_D(u) \geq \sqrt{np(1-p)}/\log^{3/4} n \mid X_1, \dots, X_i] \\
&\quad = \mathbb{P}'[\mathrm{ex}_{D_p^{i+1}}(u) \geq \sqrt{np(1-p)}/\log^{3/4} n - \mathrm{ex}_{D_{\mathrm{cond}}^{i+1}}(u) \mid X_1, \dots, X_{i+1}] \\
&\qquad - \mathbb{P}'[\mathrm{ex}_{D_p^{i}}(u) \geq \sqrt{np(1-p)}/\log^{3/4} n - \mathrm{ex}_{D_{\mathrm{cond}}^{i}}(u) \mid X_1, \dots, X_i],
\end{aligned}$$

so by (2.5.4) and (2.5.5), and using the fact that $|\mathrm{ex}_{D_{\mathrm{cond}}^{i+1}}(u) - \mathrm{ex}_{D_{\mathrm{cond}}^{i}}(u)| \leq 1$, we conclude that[33]

$$|Y_{i+1} - Y_i| \leq 2q_0^{(k_i(u))} \eqqcolon c_i. \tag{2.5.8}$$

Finally, for any $i \in [N-1] \setminus [n-2]$ such that $e_{i+1} = \{u, u'\} \subseteq N_D^+(v)$, one can similarly show that[34]

$$|Y_{i+1} - Y_i| \leq 2(q_0^{(k_i(u))} + q_0^{(k_i(u'))}) \eqqcolon c_i. \tag{2.5.9}$$

This covers all the range of $i \in [N-1] \setminus [n-2]$.

By combining (2.5.7)–(2.5.9), we observe that, for each $u \in N^+(v)$ and each $k \in [n-2]$, the value $q_0^{(k)}$ appears as part of c_i for exactly one value of $i \in [N-1] \setminus [n-2]$. Then, we have

$$\sum_{i=n-1}^{N-1} c_i^2 \leq \sum_{u \in N_D^+(v)} \sum_{k=1}^{n-2} 8 \left(q_0^{(k)}\right)^2,$$

where we have used the fact that $(x+y)^2 \leq 2x^2 + 2y^2$. Now, by applying Lemma 2.5.5 and the conditioning on $\mathcal{E}$, we have that[35]

$$\sum_{i=n-1}^{N-1} c_i^2 \leq \left(1 \pm c\sqrt{\frac{(1-p)\log n}{np}}\right) 8K^2 np \left(1 + \sum_{k=2}^{n-2} \frac{\log k}{kp(1-p)}\right) = \mathcal{O}\left(\frac{n\log^2 n}{(1-p)}\right).$$

Therefore, we can apply Lemma 2.3.8 to conclude that[36]

$$\mathbb{P}'[|A^+ \cap N_D^+(v)| \neq np/2 \pm 2\sqrt{n/(1-p)}\log^2 n] = e^{-\Omega(\log^2 n)}. \quad (2.5.10)$$

By similar arguments, we can show that

$$\mathbb{P}'[|A^- \cap N_D^+(v)| \neq np/2 \pm 2\sqrt{n/(1-p)}\log^2 n] = e^{-\Omega(\log^2 n)}, \quad (2.5.11)$$

$$\mathbb{P}'[|A^+ \cap N_D^-(v)| \neq np/2 \pm 2\sqrt{n/(1-p)}\log^2 n] = e^{-\Omega(\log^2 n)}, \quad (2.5.12)$$

$$\mathbb{P}'[|A^- \cap N_D^-(v)| \neq np/2 \pm 2\sqrt{n/(1-p)}\log^2 n] = e^{-\Omega(\log^2 n)}. \quad (2.5.13)$$

Let $\mathcal{E}'$ be the event that $|A^+ \cap N_D^+(v)|, |A^- \cap N_D^+(v)|, |A^+ \cap N_D^-(v)|, |A^- \cap N_D^-(v)| = np/2 \pm 2\sqrt{n/(1-p)}\log^2 n$. By combining (2.5.10)–(2.5.13) with a union bound, it follows that $\mathbb{P}'[\mathcal{E}'] = 1 - e^{-\Omega(\log^2 n)}$. Therefore, $\mathbb{P}[\mathcal{E}'] \geq 1 - 2/n^3$[37]. Finally, the statement follows by a union bound over all vertices $v \in V$. □

Proof of Theorem 2.5.1. We condition on the event that the statements of Lemmas 2.5.2, 2.5.3, 2.5.6 and 2.5.7 hold, which occurs a.a.s. Then, Lemma 2.5.2 directly implies (P5) holds. We may partition the vertices by defining $A^+ := \{v \in V(D) \mid \mathrm{ex}_D(v) \geq 155\kappa\}$, $A^- := \{v \in V(D) \mid \mathrm{ex}_D(v) \leq -155\kappa\}$ and $A^0 := V(D) \setminus (A^+ \cup A^-)$. In particular, by Corollary 2.5.6 we have that $|A^0|$ is sublinear. The condition on the excess in (P1) and (P2) holds now by definition. The conditions on the edge distribution in (P1) and (P2) as well as (P4) follow by Lemma 2.5.7 in the given range of p[38]. Finally, (P3) holds by combining Lemma 2.5.3 and Lemma 2.5.7[39]. □

2.6 CONCLUSION

We have shown in Theorem 2.1.2 that, for p in the range $n^{-1/3}\log^4 n \leq p \leq 1 - n^{-1/5}\log^{5/2} n$, a.a.s. $D_{n,p}$ is consistent. Of course, we should expect to be able to improve this range, particularly the lower bound, and perhaps even no lower bound is necessary. Indeed, when $p \ll 1/n$, we know $D_{n,p}$ is acyclic, and it is easy to see that acyclic digraphs are consistent (simply iteratively remove maximal length paths and observe that the excess decreases by 1 each time).

The bottleneck in our current approach is in Lemma 2.4.11 where we process medium length cycles. An improvement in the bounds there would lead to an improvement in the range of p in Theorem 2.1.2. However this alone can only achieve a lower bound for p of approximately $n^{-1/2}$: beyond that one needs to improve other aspects of the argument and new ideas are necessary.

We remark that both the process of selecting an appropriate absorbing structure and the process of finding a path decomposition as described in Theorem 2.4.3 can be made algorithmic. One can easily check that all steps can be completed in time polynomial in n, including the procedure described in the proof of Theorem 2.3.4.

Finally, we saw that our methods can be used to show that a fairly broad class of digraphs (that are far from pseudorandom) are consistent; see Theorem 2.1.3 and Theorem 2.4.3. It would be interesting to find other classes of digraphs that are consistent.

NOTES

[1]We have that $\kappa = t/155 = 500N^{2/5}/155 > 3N^{2/5}$, so (C1) holds. (C2) holds since $d \geq t = 500N^{2/5} > 365N^{2/5}$. Finally,

$$\min\left\{\frac{d}{3}, \frac{t^2}{10^6}\right\} = \min\left\{\frac{np}{3}, \frac{500^2 N^{4/5}}{10^6}\right\} \leq \min\left\{\frac{np}{3}, \frac{9N^{4/5}}{12}\right\},$$

so (C3) holds too.

[2]Assume $v \in A^+$ (the other case is proved analogously). We apply Lemma 2.3.6 twice to obtain the two bounds on the resulting degree. Note that $30\kappa \leq \mu := \mathbb{E}[d^+_{D_q}(v)] \leq 120\kappa$. From the lower bound, by an application of Lemma 2.3.6 we have that

$$\mathbb{P}[d^+_{D_q}(v) < 25\kappa] \leq \mathbb{P}[d^+_{D_q}(v) < 25\mu/30] \leq e^{-\mu/72} \leq e^{-5\kappa/12}$$

(we apply Lemma 2.3.6 with $\delta = 1/6$, and in the last bound we use the lower bound on μ). Similarly, from the upper bound we have that

$$\mathbb{P}[d^+_{D_q}(v) > 150\kappa] \leq \mathbb{P}[d^+_{D_q}(v) > 150\mu/120] \leq e^{-\mu/48} \leq e^{-5\kappa/8}$$

(we apply Lemma 2.3.6 with $\delta = 1/4$). The claim follows by adding these two terms; the bound we claim is very rough.

[3]Assume $v \in A^+$ (the other case is analogous). By the property we are assuming on D' (i.e., all vertices have degree at least 25κ), we have $\mu := \mathbb{E}[d^+_{D^+}(v)] \geq 25\kappa/2$. Then, by Lemma 2.3.6,

$$\mathbb{P}[\mathcal{B}'_v] = \mathbb{P}[d^+_{D^+}(v) < 12\kappa] \leq \mathbb{P}[d^+_{D^+}(v) < 24\mu/25] \leq e^{-\mu/2\cdot 625} \leq e^{-\kappa/100}.$$

[4]This is the event that none of the $\mathcal{B}'_i$ occur, so it suffices to check that this happens with positive probability. By the union bound, the probability that any of the $\mathcal{B}'_i$ occur is at most $ne^{-\kappa/100} < 1$, so we are done.

[5]Say $v \in A^+$ (the other case is analogous). By (P3), we have $\mathbb{E}[d^+_{D_q}(v)] \leq \mu := q\lambda \leq 4\kappa$ (this follows by substituting the value of q and using the bound on λ from the statement). Observe further that the variable $d^+_{D_q}(v)$, which is a binomial variable $\mathrm{Bin}(d^+_{D'}(v), q)$, is stochastically dominated by

a variable $X \sim \text{Bin}(\lambda, q)$ (this means that, for every t, we have $\mathbb{P}[d^+_{D_q}(v) \geq t] \leq \mathbb{P}[X \geq t]$). Then, using this fact and Lemma 2.3.6,

$$\mathbb{P}[\mathcal{B}_v] = \mathbb{P}[d^+_{D_q}(v) > 5\kappa] \leq \mathbb{P}[X > 5\kappa] \leq \mathbb{P}[X > 5\mu/4] \leq e^{-\mu/48} = e^{-\frac{1}{4}\frac{\kappa\lambda}{np}}.$$

[6]Consider $\mathcal{B}^+_v$ (the other case is analogous). By (P4) we have that $\mu := \mathbb{E}[d^-_{D_q}(v)] \geq npq/3 = 4\kappa$. Now, by Lemma 2.3.6,

$$\mathbb{P}[\mathcal{B}^+_v] = \mathbb{P}[d^-_{D_q}(v) < 3\kappa] \leq \mathbb{P}[d^-_{D_q}(v) < 3\mu/4] \leq e^{-\mu/32} \leq e^{-\kappa/8}.$$

[7]All edges of $\mathcal{P}$ are oriented either towards A^- or from A^+. By having the extra vertex outside $V(C)$, if e_1 and e_2 share a vertex, then they are both oriented towards this vertex or away from this vertex. So they do not form a path of length 2.

[8]There are at most $\kappa + 1$ available edges, so at least $7\kappa - 1 - \kappa - 1 = 6\kappa - 2 \geq 5\kappa$ have been assigned.

[9]By the upper bound on κ, we have that $\ell \geq N/\kappa > \kappa$, so in particular $\ell > 10$. The first inequality holds since $4N/(5\kappa) \leq 4\ell/5$. Now it suffices to check that $\ell/5 \geq \ell/10 + 1$, which holds.

[10]Assume $T = \emptyset$, so $T' = V(\mathcal{C}'_2)$. Then, for each $C \in S$, we have $e_F(\{C\}, T') = |V(C)| = |V(C) \cap \dot{A}| > g(C)$.

[11]The term $e(D_S)/\kappa$ appears by ignoring the ceilings, and just looking at the definition of D_S. Because of the ceilings, then, we could be adding much more, but at most $|S|$ more (and the final $|S|$ is by adding 1 each time).

[12]Recall T and T' are just sets of vertices. Each edge between a vertex in T' and a cycle in S corresponds simply to this vertex lying on the cycle, which means it contributes to two edges in D_S (one towards the vertex, and one away from it). By considering all the cycles it belongs to, we recover precisely the indegree and the outdegree of that vertex in D_S. This gives the first equality. Now, $\sum_{v \in T'} d^+_{D_S}(v) + d^-_{D_S}(v) = 2e_{D_S}(T') + e_{D_S}(T', T) + e_{D_S}(T, T')$, so the inequality is trivial.

[13]By pigeonhole, we have $y_i \notin V(P_{3-i}) \cap \dot{A}$. Here we are implicitly using the definition of $\mathcal{A}$ in the sense that $y_i \in \dot{A}$ to derive the conclusion.

[14] Recall that the conditions (C1)–(C3) are (C1) $\kappa = 3N^{2/5}$, (C2) $np \geq 365N^{2/5}$, (C3) $\lambda = \min\{np/3, \kappa^2/12\}$.
All the following hold for $n \geq n_0$ for n_0 suitably large. For (a), note that the maximum on the left hand side is dominated by $\max((12\lambda)^{1/2}, (72N^2)^{1/5})$, which is at most κ by (C1) and (C3). The upper bound in (a) holds by noting $\kappa \leq np/120$ by (C1) and (C2), and $\kappa < N^{1/2}$ by (C1). (b) holds by (C3). For (c), if $\lambda = np/3$, then $\kappa\lambda \geq 4np\log(2n)$ by (C1), and if $\lambda = \kappa^2/12$, then $\kappa\lambda = \kappa^3/12 \geq N^{6/5} \geq 4np\log(2n)$.

[15] Recall that the conditions (C′1), (C′2), and (C3) are (C′1) $\kappa = 6(N^2p)^{1/5}$, (C′2) $p \geq n^{-1/3}\log^4 n$, and (C3) $\lambda = \min\{np/3, \kappa^2/12\}$.
All the following hold for $n \geq n_0$ for n_0 suitably large. For (a′), the maximum on the left hand side is dominated by $\max((12\lambda)^{1/2}, (7200N^2p)^{1/5})$ (where we can exclude $\sqrt{12/(25p)}\log n$ by (C′2)). We see κ is bigger than this by (C3) and (C′1). For the upper bound we have $\kappa \leq N^{1/2}$ by (C′1) and $\kappa \leq np/120$ by (C′1) and (C′2). (b) holds by (C3) again. For (c), if $\lambda = np/3$, then $\kappa\lambda \geq 4np\log(2n)$ (using (C′1) and (C′2)), and if $\lambda = \kappa^2/12$ then $\kappa\lambda = \kappa^3/12 \geq N^{6/5}p^{3/5}$ by (C′1), which is at least $4np\log(2n)$.

[16] Note that the first inequality is equivalent after rearrangement to $C\log^{23/6} n n^{-1/3} \leq p(1-p)^{5/3}$ for a suitable constant C. If $p < 1/2$ then the RHS is at least $(n^{-1/3}\log^4 n)/4$ so the inequality is satisfied, and if $p \geq 1/2$ then the RHS is at least $(n^{-1/3}\log^{25/6} n)/2$ so the inequality is satisfied.
For the second inequality, we must show that

$$\frac{5n^{1/2}\log^2 n}{(1-p)^{1/2}} \leq np/3 \quad \text{and} \quad \frac{5n^{1/2}\log^2 n}{(1-p)^{1/2}} \leq \kappa'^2/12 = 100(N^2p)^{2/5}/12.$$

After rearrangement, the first of these is equivalent to $Cn^{-1/2}\log^2 n \leq p(1-p)^{1/2}$, which holds in our range of p. After rearrangement, the second of these is equivalent to $Cn^{-3/10}\log^{6/5} n \leq p^{2/5}(1-p)^{1/2}$, which also holds in our range of p.

[17] Since $i^2p \to \infty$ with n, we have that

$$\begin{aligned}\mathbb{P}[X \geq 100i^2p] &\leq \mathbb{P}[X \geq 100\mathbb{E}[X]] \leq (e/100)^{100\mathbb{E}[X]} \\ &= (e/100)^{100(1-1/i)i^2p} \leq (e/100)^{99i^2p},\end{aligned}$$

where the last inequality holds since $1 - 1/i = 1 - o(1)$.

[18]In order to see the last inequality, let us take logarithms. Clearly, the inequality is equivalent to

$$\begin{aligned} &i(1+\log n-\log i)+99i^2p(1-\log 100)\le -3\log n\\ &\qquad\iff i(1+\log n-\log i)+3\log n\le 99i^2p(\log 100-1). \end{aligned}$$

Now clearly, if i is sufficiently large, $i(1+\log n-\log i)+3\log n\le 2i\log n$ and $99i^2p(\log 100-1)\ge 100i^2p$, so it would suffice to check that

$$2i\log n\le 100i^2p,$$

and this holds by the bound on i in the statement.

[19]Let $X:=d_D^*(v)\sim \mathrm{Bin}(n-1,p)$. We have

$$\begin{aligned} &\mathbb{P}\left[X\neq np\pm c\sqrt{np(1-p)\log n}\right]\\ &\le \mathbb{P}\left[X\neq (n-1)p\pm c\sqrt{(n-1)p(1-p)\log n}/2\right]\\ &\le \mathbb{P}\left[X\neq (n-1)p\pm c\sqrt{(n-1)p\log n}/4\right]\\ &= \mathbb{P}\left[X\neq \left(1\pm\frac{c}{4}\sqrt{\frac{\log n}{(n-1)p}}\right)(n-1)p\right]\\ &\le 2e^{-\frac{c^2}{16}\frac{\log n}{(n-1)p}(n-1)p/3}\le e^{-c^2\log n/50}. \end{aligned}$$

For the second inequality we are using the fact that $1-p\ge 1/2$. The last inequality holds for n sufficiently large. Observe that the condition that $p\ge C\log n/n$ is needed to guarantee that the δ with which we apply Lemma 2.3.6 lies in $(0,1)$. Indeed, in our application of the Chernoff bound we have

$$\delta=\frac{c}{4}\sqrt{\frac{\log n}{(n-1)p}}<1\iff p>\frac{c^2}{16}\frac{\log n}{n-1},$$

so it suffices to have $p>c^2\log n/8n$.

[20]$2ne^{-c^2\log n/50}=e^{\log 2+(1-c^2/50)\log n}=o(1/n^3)$

[21]We have that

$$\begin{aligned} d_D^*(v)&=n-1-d_{\overline{D}}^*=n-1-n(1-p)\pm c\sqrt{np(1-p)\log n}\\ &=np-1\pm c\sqrt{np(1-p)\log n}, \end{aligned}$$

and this is what we want (by making c slightly worse).

[22]For every $n \geq 1$ we have that $\log n/(np(1-p)) \geq \log n/n$. Simply, for any given n_0, we can set $K \geq \sqrt{n_0/\log n_0}$ (note this function is increasing for $n_0 \geq 3$, and soon overtakes the value of $n_0 = 2$), which means that, for $n \leq n_0$, the statement is satisfied by the trivial upper bound:

$$\begin{aligned}\mathbb{P}[\mathrm{ex}_{D_p}(v) = 0] \leq 1 = \sqrt{\frac{n}{\log n}\frac{\log n}{n}} \leq \sqrt{\frac{n_0}{\log n_0}\frac{\log n}{n}} \\ \leq K\sqrt{\frac{\log n}{n}} \leq K\sqrt{\frac{\log n}{np(1-p)}}.\end{aligned}$$

[23]Indeed, if we assume $p < C_0 \log n/n$, by adjusting K we may guarantee that

$$K\sqrt{\frac{\log n}{np(1-p)}} > K/\sqrt{C_0} \geq 1,$$

and the case when $p > 1 - C_0 \log n/n$ is proved analogously.

[24]Consider the ratio p_{i+1}/p_i, for $i \in [n-1]_0$. We have that

$$\begin{aligned}\frac{p_{i+1}}{p_i} = \frac{\binom{n}{i+1}p^{i+1}(1-p)^{n-i-1}}{\binom{n}{i}p^i(1-p)^{n-i}} &= \frac{i!(n-i)!}{(i+1)!(n-i-1)!}\frac{p}{1-p} \\ &= \frac{n-i}{i+1}\frac{p}{1-p}.\end{aligned}$$

We want to know when this ratio changes from greater than 1 (which means the ratio is increasing) to when it is less than 1 (decreasing). By setting $p_{i+1}/p_i = 1$ and isolating, we have that

$$\begin{aligned}\frac{p_{i+1}}{p_i} = 1 \iff \frac{n-i}{i+1}\frac{p}{1-p} = 1 \iff (n-i)\frac{p}{1-p} = i+1 \\ \iff \frac{np}{1-p} = i + 1 + \frac{p}{1-p}i = \frac{i}{1-p} + 1 \iff i = np - 1 + p.\end{aligned}$$

Since this is the only solution, we know that the maximum must be achieved for either $\lfloor np-1+p \rfloor$ or $\lceil np-1+p \rceil$, and both of these lie in $np \pm 2$.

[25]Let us write p_{np}, and assume n is sufficiently large (smaller values of n are hidden by K; it is easy to check that, for p in the given range, ± 2 does not affect the asymptotic statements, and we will increase the final constant here to avoid issues). We have that

$$p_{np} = \binom{n}{np} p^{np}(1-p)^{n(1-p)}.$$

We now use the bounds $\sqrt{2\pi} n^{n+1/2} e^{-n} \leq n! \leq e n^{n+1/2} e^{-n}$, similar to Stirling's approximation and valid for all n, to conclude that

$$p_{np} \leq \frac{e\sqrt{n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi}\sqrt{np}\left(\frac{np}{e}\right)^{np} \sqrt{2\pi}\sqrt{(1-p)n}\left(\frac{(1-p)n}{e}\right)^{(1-p)n}} p^{np}(1-p)^{(1-p)n}$$
$$= \frac{e}{2\pi} \frac{1}{\sqrt{np(1-p)}}.$$

Clearly the constant in front is less than 1, so what we claim must hold true by considering the small changes by ± 2.

[26]This sum of probabilities is the same as the probability that the outcome of a binomial variable deviates from its mean by at least $c\sqrt{np(1-p)\log n}$. Check the proof of Lemma 2.5.3 for the details of the calculation.

[27]We have that (setting $r = c\sqrt{np(1-p)\log n}$ for clarity)

$$\begin{aligned}\mathbb{P}[\mathrm{ex}_D(v) = 0] &= \sum_{i=0}^{n-1} p_i^2 \\ &= \sum_{i=0}^{np-r} p_i^2 + \sum_{i=np-r}^{np+r} p_i^2 + \sum_{i=np+r}^{n-1} p_i^2 \\ &\le \sum_{i=0}^{np-r} p_i + 2r(\max_{i\in[n]_0} p_i)^2 + \sum_{i=np+r}^{n-1} p_i \\ &\le e^{-c^2\log n/50} + \frac{2r}{np(1-p)} \le K\sqrt{\frac{\log n}{np(1-p)}}.\end{aligned}$$

Note that we need c and C to be large enough so that we can apply Lemma 2.3.6, and also c to be large enough so that the second term above dominates.

[28]We have that $\mathbb{E}[Y] = \mathcal{O}(n/\log^{1/4} n)$. By Markov's inequality, it follows that

$$\mathbb{P}[Y \ge n/\log^{1/8} n] \le \mathbb{E}[Y]\log^{1/8} n/n = \mathcal{O}(1/\log^{1/8} n) = o(1).$$

[29]Indeed, consider the joint distribution of D_p^i and D_p^{i+1} where we first reveal D_p^i and then reveal the last pair of edges needed to obtain D_p^{i+1}. If $\mathrm{ex}_{D_p^{i+1}}(v) \ge a$, then we are guaranteed that $\mathrm{ex}_{D_p^i}(v) \ge a-1$, since the last pair of edges we reveal can only increase the excess by at most 1. Similarly, we cannot have $\mathrm{ex}_{D_p^i}(v) \ge a+1$ unless $\mathrm{ex}_{D_p^{i+1}}(v) \ge a$, as the excess cannot decrease by more than 1 when revealing the last pair of edges.

[30]Indeed, we have that either

$$\begin{aligned}&\mathbb{P}[\mathrm{ex}_{D_p^i}(v) \ge a+1 \mid X_1,\dots,X_i] \\ \le\, &\mathbb{P}[\mathrm{ex}_{D_p^{i+1}}(v) \ge a \mid X_1,\dots,X_{i+1}] \\ \le\, &\mathbb{P}[\mathrm{ex}_{D_p^i}(v) \ge a \mid X_1,\dots,X_i]\end{aligned}$$

or

$$\begin{aligned}&\mathbb{P}[\mathrm{ex}_{D_p^i}(v) \ge a \mid X_1,\dots,X_i] \\ \le\, &\mathbb{P}[\mathrm{ex}_{D_p^{i+1}}(v) \ge a \mid X_1,\dots,X_{i+1}] \\ \le\, &\mathbb{P}[\mathrm{ex}_{D_p^i}(v) \ge a-1 \mid X_1,\dots,X_i].\end{aligned}$$

Now assume that the first of the two cases holds (the other follows similarly). The claim follows since

$$\mathbb{P}[\mathrm{ex}_{D_p^i}(v) \geq a \mid X_1, \ldots, X_i] - \mathbb{P}[\mathrm{ex}_{D_p^i}(v) \geq a+1 \mid X_1, \ldots, X_i] \leq q_0^{(k_i(v))}.$$

[31]By (2.5.4), each $\mathbb{P}'[\mathrm{ex}_{D_p^{n-1}}(u) = i]$ (recall that here $\mathbb{P} = \mathbb{P}'$) is at most $q_0^{(n-2)}$, so

$$\begin{aligned}
&\mathbb{P}'[|\mathrm{ex}_{D_p^{n-1}}(u)| \geq \sqrt{np(1-p)}/\log^{3/4} n] \\
&\geq \mathbb{P}'[|\mathrm{ex}_{D_p^{n-1}}(u)| > \sqrt{np(1-p)}/\log^{3/4} n] \\
&= 1 - \mathbb{P}'[|\mathrm{ex}_{D_p^{n-1}}(u)| \leq \sqrt{np(1-p)}/\log^{3/4} n] \\
&= 1 - \sum_{i=-\sqrt{np(1-p)}/\log^{3/4} n}^{\sqrt{np(1-p)}/\log^{3/4} n} \mathbb{P}[\mathrm{ex}_{D_p^{n-1}}(u) = i] \\
&\geq 1 - 2\sqrt{np(1-p)}/\log^{3/4} n \cdot q_0^{(n-2)} = 1 - \mathcal{O}(1/\log^{1/4} n),
\end{aligned}$$

where the last inequality uses Lemma 2.5.5 (the change from n to $n-2$ does not change the asymptotics).

[32]To see this, observe the following. Let $\mathcal{E}_1$ be the event that $|\mathrm{ex}_D(u)| \geq \sqrt{np(1-p)}/\log^{3/4} n$. Then, we have that $\mathbb{P}'[u \in A^+] = \mathbb{P}'[\mathcal{E}_1]\,\mathbb{P}'[\mathrm{ex}_D(u) > 0 \mid \mathcal{E}_1]$. We have that $\mathbb{P}'[\mathcal{E}_1] = 1 - \mathcal{O}(1/\log^{1/4} n)$ Indeed, by the discussion above we have that $\mathrm{ex}_D(u) = \mathrm{ex}_{D_p^{n-1}}(u) \pm 1$, and

$$\begin{aligned}
&\mathbb{P}'[|\mathrm{ex}_D(u)| \geq \sqrt{np(1-p)}/\log^{3/4} n] \\
&\geq \mathbb{P}'[|\mathrm{ex}_{D_p^{n-1}}(u)| \geq 2\sqrt{np(1-p)}/\log^{3/4} n] = 1 - \mathcal{O}(1/\log^{1/4} n).
\end{aligned}$$

After conditioning on this, the vertex has positive or negative excess only depending on whether $\mathrm{ex}_{D_p^{n-1}}(u)$ is positive or negative, and the probability distribution for this random variable is symmetric around 0, so each must have probability a half.

[33]Here we use the fact that $\mathrm{ex}_{D_{\mathrm{cond}}^{i+1}}(u) = \mathrm{ex}_{D_{\mathrm{cond}}^{i}}(u) \pm 1$. Thus, we have three cases. Let us show here the case when $\mathrm{ex}_{D_{\mathrm{cond}}^{i+1}}(u) = \mathrm{ex}_{D_{\mathrm{cond}}^{i}}(u) + 1$; the

other two are done similarly. Let $a := \sqrt{np(1-p)}/\log^{3/4} n - \mathrm{ex}_{D^i_{\mathrm{cond}}}(u)$. Using (2.5.5), we conclude that

$$\begin{aligned}
&|\mathbb{P}'[\mathrm{ex}_{D_p^{i+1}}(u) \geq a-1 \mid X_1,\dots,X_{i+1}] - \mathbb{P}'[\mathrm{ex}_{D_p^i}(u) \geq a \mid X_1,\dots,X_i]| \\
\leq &|\mathbb{P}'[\mathrm{ex}_{D_p^{i+1}}(u) \geq a-1 \mid X_1,\dots,X_{i+1}] - \mathbb{P}'[\mathrm{ex}_{D_p^i}(u) \geq a-1 \mid X_1,\dots,X_i]| \\
&+ |\mathbb{P}'[\mathrm{ex}_{D_p^i}(u) \geq a-1 \mid X_1,\dots,X_i] - \mathbb{P}'[\mathrm{ex}_{D_p^i}(u) \geq a \mid X_1,\dots,X_i]| \\
\leq &2q_0^{(k_i(u))}.
\end{aligned}$$

Here, the first inequality holds by the triangle inequality. Then, the first difference is bounded by (2.5.5), and the second, by (2.5.4).

[34]We have that

$$\begin{aligned}
&Y_{i+1} - Y_i \\
&\quad = \mathbb{P}'[u \in A^+ \mid X_1,\dots,X_{i+1}] + \mathbb{P}'[u' \in A^+ \mid X_1,\dots,X_{i+1}] \\
&\qquad - \mathbb{P}'[u \in A^+ \mid X_1,\dots,X_i] - \mathbb{P}'[u' \in A^+ \mid X_1,\dots,X_i] \\
&\quad = \mathbb{P}'[\mathrm{ex}_{D_p^{i+1}}(u) \geq \sqrt{np(1-p)}/\log^{3/4} n - \mathrm{ex}_{D^{i+1}_{\mathrm{cond}}}(u) \mid X_1,\dots,X_{i+1}] \\
&\quad - \mathbb{P}'[\mathrm{ex}_{D_p^i}(u) \geq \sqrt{np(1-p)}/\log^{3/4} n - \mathrm{ex}_{D^i_{\mathrm{cond}}}(u) \mid X_1,\dots,X_i] \\
&\quad + \mathbb{P}'[\mathrm{ex}_{D_p^{i+1}}(u') \geq \sqrt{np(1-p)}/\log^{3/4} n - \mathrm{ex}_{D^{i+1}_{\mathrm{cond}}}(u') \mid X_1,\dots,X_{i+1}] \\
&\quad - \mathbb{P}'[\mathrm{ex}_{D_p^i}(u') \geq \sqrt{np(1-p)}/\log^{3/4} n - \mathrm{ex}_{D^i_{\mathrm{cond}}}(u') \mid X_1,\dots,X_i].
\end{aligned}$$

Observe that the variables above are not independent, but the bound given by (2.5.5) still holds. By applying that, letting $a := \sqrt{np(1-p)}/\log^{3/4} n - \mathrm{ex}_{D^i_{\mathrm{cond}}}(u)$ and $b := \sqrt{np(1-p)}/\log^{3/4} n - \mathrm{ex}_{D^i_{\mathrm{cond}}}(u')$, we have that (again, here we only write one case, there are others that work in the same way)

$$\begin{aligned}
&|Y_{i+1} - Y_i| \\
&\leq |\mathbb{P}'[\mathrm{ex}_{D_p^{i+1}}(u) \geq a-1 \mid X_1,\dots,X_{i+1}] - \mathbb{P}'[\mathrm{ex}_{D_p^i}(u) \geq a \mid X_1,\dots,X_i]| \\
&+ |\mathbb{P}'[\mathrm{ex}_{D_p^{i+1}}(u') \geq b-1 \mid X_1,\dots,X_{i+1}] - \mathbb{P}'[\mathrm{ex}_{D_p^i}(u') \geq b \mid X_1,\dots,X_i]| \\
&\leq |\mathbb{P}'[\mathrm{ex}_{D_p^{i+1}}(u) \geq a-1 \mid X_1,\dots,X_{i+1}] - \mathbb{P}'[\mathrm{ex}_{D_p^i}(u) \geq a-1 \mid X_1,\dots,X_i]| \\
&+ |\mathbb{P}'[\mathrm{ex}_{D_p^i}(u) \geq a-1 \mid X_1,\dots,X_i] - \mathbb{P}'[\mathrm{ex}_{D_p^i}(u) \geq a \mid X_1,\dots,X_i]| \\
&+ |\mathbb{P}'[\mathrm{ex}_{D_p^{i+1}}(u') \geq b-1 \mid X_1,\dots,X_{i+1}] - \mathbb{P}'[\mathrm{ex}_{D_p^i}(u') \geq b-1 \mid X_1,\dots,X_i]| \\
&+ |\mathbb{P}'[\mathrm{ex}_{D_p^i}(u') \geq b-1 \mid X_1,\dots,X_i] - \mathbb{P}'[\mathrm{ex}_{D_p^i}(u') \geq b \mid X_1,\dots,X_i]| \\
&\leq 2(q_0^{(k_i(u))} + q_0^{(k_i(u'))}).
\end{aligned}$$

[35]In order to see the last equality, observe that, since $\frac{\log k}{kp(1-p)}$ is a function decreasing in k for $k \geq 3$ (we may treat p as a constant, for this sum), we have that

$$\begin{aligned} 1 + \sum_{k=2}^{n-2} \frac{\log k}{kp(1-p)} &= 1 + \frac{1}{p(1-p)} \sum_{k=2}^{n-2} \frac{\log k}{k} \\ &\leq 1 + \frac{1}{p(1-p)} \left(C + \int_3^{n-2} \frac{\log x}{x} \, \mathrm{d}x \right) = \mathcal{O}\left(\frac{\log^2 n}{p(1-p)} \right), \end{aligned}$$

where $C < 1$ is the sum of the first two terms (which is a constant), and the last inequality follows since $\int_4^{n-2} \frac{\log x}{x} \, \mathrm{d}x = C_1 + \log^2 n/2$.

[36]We have that

$$\mathbb{P}'[|Y_N - Y_{n-1}| \geq t] \leq 2e^{-\frac{t^2}{2\sum_{i=n-1}^{N-1} c_i^2}} = e^{-\Omega(t^2(1-p)/n \log^2 n)}.$$

Thus, if $t = \sqrt{n/(1-p)} \log^2 n$, we have that

$$\mathbb{P}'[|Y_N - Y_{n-1}| \geq \sqrt{n/(1-p)} \log^2 n] \leq e^{-\Omega(\log^2 n)}.$$

The last bound follows by observing that

$$\begin{aligned} &\mathbb{P}[|A^+ \cap N_D^+(v)| \neq np/2 \pm 2\sqrt{n/(1-p)} \log^2 n] \\ \leq\ &\mathbb{P}[|A^+ \cap N_D^+(v)| \neq np/2 \pm (c\sqrt{np(1-p) \log n} + \sqrt{n/(1-p)} \log^2 n)] \\ =\ &\mathbb{P}[|Y_N - Y_{n-1}| \geq \sqrt{n/(1-p)} \log^2 n]. \end{aligned}$$

[37]By the law of total probability, we have that

$$\begin{aligned} \mathbb{P}[\mathcal{E}'] &= \mathbb{P}'[\mathcal{E}'] \, \mathbb{P}[\mathcal{E}] + \mathbb{P}[\mathcal{E}' \mid \overline{\mathcal{E}}] \, \mathbb{P}[\overline{\mathcal{E}}] \\ &\geq \mathbb{P}'[\mathcal{E}'] \, \mathbb{P}[\mathcal{E}] \geq (1 - e^{-\Omega(\log^2 n)})(1 - 1/n^3) \geq 1 - 2/n^3 \end{aligned}$$

(where the last inequality holds for n sufficiently large).

[38]Let us prove this for one of the cases of (P4) (the other cases are analogous or give slightly weaker bounds). By Lemma 2.5.7, for each $v \in A^+$ we have that

$$e_D(v, A^-) \geq np/2 - 2\sqrt{n/(1-p)} \log^2 n,$$

so it suffices to check that this is at least $np/3$. But this is equivalent to having

$$\frac{np}{6} \geq 2\sqrt{\frac{n}{1-p}}\log^2 n \iff p^2(1-p) \geq 144\log^4 n/n.$$

This clearly holds for all constant $p \in (0,1)$ (for sufficiently large n), so we may assume that $p = o(1)$ or $p = 1 - o(1)$. In the first case, the inequality becomes $p^2 \geq (1+o(1))144\log^4 n/n$, which holds for $p \geq 13\log^2 n/n^{1/2}$ (and n large enough). In the second case, we get $(1-p) \geq (1+o(1))144\log^4 n/n$, which holds for $p \leq 1 - 150\log^4 n/n$ (and n large enough).

[39] We have that (we only write one case, the other is analogous)

$$\begin{aligned} e_D(v, A^0) &= d^+_D(v) - e_D(v, A^+) - e_D(v, A^-) \\ &\leq np + c\sqrt{np(1-p)\log n} - np + 4\sqrt{n/(1-p)}\log^2 n \\ &\leq 5\sqrt{n/(1-p)}\log^2 n \end{aligned}$$

(where the last inequality holds for sufficiently large n).

3 A POLYNOMIAL-TIME ALGORITHM TO DETERMINE (ALMOST) HAMILTONICITY OF DENSE REGULAR GRAPHS

3.1 INTRODUCTION

The study of Hamilton cycles in graphs is a classical part of graph theory. Hamilton cycles have been studied intensely from structural, extremal and algorithmic perspectives and they are especially relevant due to their connection with the traveling salesman problem. This chapter is concerned with the algorithmic question of determining whether a dense regular graph contains an (almost) Hamilton cycle. Dense in this chapter means that the minimum degree is linear in the number of vertices.

Dirac's theorem (1.2.1) guarantees the existence of a Hamilton cycle in any n-vertex graph of minimum degree at least $n/2$, so this immediately gives a (trivial) algorithm to determine existence in such graphs (and its proof also gives a polynomial-time algorithm for finding a Hamilton cycle). On the other hand, for each $\varepsilon > 0$, the problem of determining Hamiltonicity in n-vertex graphs of minimum degree $(\frac{1}{2} - \varepsilon)n$ is $\mathcal{NP}$-complete [19] (see also Proposition 3.1.2). Our main result, given below, shows that the situation is quite different if we also insist the graphs are regular: we show that determining almost Hamiltonicity in dense regular graphs is polynomial-time solvable.

Theorem 3.1.1. For every $\alpha \in (0, 1]$, there exists $c = c(\alpha)$ and a (deterministic) polynomial-time algorithm that, given an n-vertex D-regular graph G with $D \geq \alpha n$ as input, determines whether G contains a cycle on at least $n - c$ vertices. In fact, we can take $c(\alpha) = 100\alpha^{-2}$. Furthermore there is a (randomized) polynomial-time algorithm to find such a cycle if it exists.

Note that the problem of determining the existence of a very long cycle (as in the result above) becomes $\mathcal{NP}$-complete if we drop either the density or the regularity condition on G; see Proposition 3.1.2. The question of whether Theorem 3.1.1 holds for $c = c(\alpha) = 0$ (i.e. the Hamilton cycle problem) remains open and is discussed in Section 3.5. Also, see Remark 3.4.17 for a discussion of the explicit running time of the algorithm.

Arora, Karger, and Karpinski [5, 6] initiated the systematic study of $\mathcal{NP}$-hard problems on dense graphs and this continues to be an active area of research. The closest result to ours (to the best of our knowledge) is an approximation algorithm for the longest cycle problem in dense (not necessarily regular) graphs that is due to Csaba, Karpinski and Krysta [16]. For each $\alpha \in (0, 1/2)$, they give a polynomial-time algorithm which, given an n-vertex graph G of minimum degree αn, finds a cycle of length at least $(\frac{\alpha}{1-\alpha})\ell$, where ℓ is the length of the longest cycle in G.[1] They also show one cannot replace $(\frac{\alpha}{1-\alpha})$ with $(1 - \varepsilon_0(1 - 2\alpha))$ where $\varepsilon_0 = 1/742$ unless $\mathcal{P} = \mathcal{NP}$. The two algorithms are not directly comparable: while theirs works on all dense graphs, ours achieves a much better approximation ratio for dense regular Hamiltonian graphs. In Section 3.5, we discuss how our methods can be used for the longest cycle problem to achieve an approximation ratio very close to 1 for general dense regular graphs.

Our algorithm is inspired by questions and results about Hamiltonicity in extremal graph theory. Here one is interested in various types of conditions that guarantee Hamiltonicity such as in Dirac's theorem; see e.g. the surveys [11, 34, 59]. There are two extremal examples that show $n/2$ is tight in Dirac's theorem: a slightly imbalanced complete bipartite graph and a graph consisting of two disjoint cliques. One might hope to eliminate these barriers to Hamiltonicity by imposing some connectivity and regularity conditions. A graph is connected if for any two vertices u, v, there is a path from u to v. A graph is t-connected if the graph remains connected after removing any set of up to $t-1$ vertices. In this direction, Bollobás [8] and Häggkvist (see [42]) independently conjectured that a t-connected regular graph with degree at least $n/(t+1)$ is Hamiltonian. Jackson [42] proved the conjecture for $t = 2$, while Jung [47] and Jackson, Li, and Zhu [43] gave an example showing the conjecture does not hold for $t \geq 4$. Finally, Kühn, Lo, Osthus, and Staden [54, 55] resolved the conjecture by proving the case $t = 3$ asymptotically. Although the conjecture does not hold in general, it suggests that questions of Hamiltonicity (and long cycles) might be easier in some sense for (dense) regular graphs, and our result seems to confirm this.

1 The actual approximation ratio here is $(\frac{\alpha}{1-\alpha}) - \varepsilon$ for arbitrarily small ε. As mentioned, for $\alpha \geq 1/2$, Dirac's theorem gives a trivial algorithm for the longest cycle problem.

Our algorithm relies heavily on the notion of robust expansion, a notion of expansion for dense (directed) graphs introduced and applied by Kühn and Osthus together with several co-authors to resolve and make progress on a number of long-standing conjectures in extremal graph theory; see for example [17, 56, 57, 58]. In particular, Kühn, Lo, Osthus and Staden [54, 55], in their proof of the $t = 3$ case of the Bollobás-Häggkvist conjecture, showed that all dense regular graphs have a vertex partition into a small number of parts where each part induces a (bipartite) robust expander. This decomposition is central to our algorithm, and by combining their argument with some spectral partitioning techniques, we are able to construct such a partition algorithmically in polynomial time; this may be of independent interest. A further by-product of our algorithm is that we can partially answer a question of Kühn and Osthus [58] about algorithms to check whether a graph is a robust expander in polynomial time; this result and its background are presented in Section 3.3.3 after robust expansion has been formally defined.

Once we have the algorithm for constructing the robust expander partition, we will also require a result of Letzter and Gruslys [35] for finding certain structures between the parts in this partition. Combining all of this with some further algorithmic ingredients will yield the desired algorithm.

Below we give a more detailed account of our algorithm as well as the proof of the hardness results (Proposition 3.1.2) mentioned above. In Section 3.2 we give some general notation and we formally define robust expansion, as well as stating some of the results from spectral graph theory that we will need in later sections. In Section 3.3, we give the algorithm for finding the robust expander partition mentioned above, and Section 3.4 is about using the structure of a robust partition to find a long cycle. This is where the proof of Theorem 3.1.1 is given.

3.1.1 *Proof outline*

We now present further details about our algorithm. The first step of the algorithm, given in Section 3.3, is to obtain a so-called robust partition of our graph. This is a vertex partition in which each part induces a robust expander or a bipartite robust expander and where there are few edges between parts. We give the precise definitions below, but informally we can think of (bipartite) robust expanders as dense (bipartite) graphs with good connectivity properties that are resilient to small alterations. In [54], Kühn, Lo, Osthus and Staden show that such a robust partition exists for dense regular graphs, and crucially, the number of parts is independent of the number of vertices and depends only on the density. The idea of the proof in [54] is to iteratively refine the vertex partition as follows. Given a vertex partition $\mathcal{P} = \{U_1, \dots, U_k\}$, if some U_i is not a (bipartite) robust expander, then it is shown there exists a partition $U_i = A \cup B$ of U_i where there are few edges between A and B; subsequently U_i is replaced with A, B in $\mathcal{P}$ and this is repeated with the new partition. This process must end after a finite number of steps since the density inside parts increases at each step (since there were not many edges between A and B). We follow this argument closely, except that the existence of A, B is not enough for us: we need a polynomial-time algorithm to find A and B. We make use of spectral algorithms to achieve this.

In the second step, given in Section 3.4, we make use of the robust partition to decide whether a very long cycle exists. Using further results from [54], we will see that a very long cycle exists if and only if a certain type of structure exists between the parts of our robust partition. With the help of a result from [35], we give a fast algorithm to determine whether such a structure is present in our graph and to find it if it is. We will give a more detailed sketch of this at the start of Section 3.4.

We end this subsection by proving the simple hardness results mentioned earlier in the introduction.

Proposition 3.1.2. For each fixed integer $C \geq 0$ and each real $\alpha \in (0, 1/2)$ the following holds.

(i) The problem of deciding whether a regular n-vertex graph has a cycle of length at least $n - C$ is $\mathcal{NP}$-complete.
(ii) The problem of deciding whether an n-vertex graph of minimum degree at least αn has a cycle of length at least $n - C$ is $\mathcal{NP}$-complete.

Proof. For part (i), it is known that the problem of determining Hamiltonicity of 3-regular graphs is $\mathcal{NP}$-complete [31], which takes care of the case $C = 0$. Fix $C \in \{1, 2, 3\}$. For a 3-regular graph G, let G' be the 3-regular graph on $3|V(G)|$ vertices obtained from G by replacing each vertex of G with a triangle in such a way that we recover G by contracting each triangle to a vertex. The following are equivalent:

G has a Hamilton cycle;

G'' has a cycle of length $|V(G'')| - 1$;

G'' has a cycle of length $|V(G'')| - 2$;

G'' has a cycle of length $|V(G'')| - 3$.

We defer the proof of this to the appendix at the end of this chapter. Let $H = G''$. Then H has a cycle of length at least $n - C$ if and only if G has a Hamilton cycle.

Now fix C even with $C \geq 4$. Given a 3-regular graph G, where without loss of generality we assume $|G| > C$, let H be the disjoint union of G with an arbitrary 3-regular graph on C vertices. Thus H and G have n and $n - C$ vertices, respectively. It follows that G has a Hamilton cycle if and only if H has a cycle of length at least $n - C$.

Finally, fix $C \geq 5$ odd. Given a 3-regular graph G (again with $|G| > C$), let H be the disjoint union of G'' with an arbitrary 3-regular graph on $C - 1$ (even) vertices. Again, H has a cycle of length at least $n - C$ if and only if G has a Hamilton cycle. In each of these cases a polynomial-time algorithm for deciding the problem in part (i) would give a polynomial-time algorithm for deciding Hamiltonicity in 3-regular graphs.

(ii) We reduce to the problem of deciding the existence of a Hamilton path in general graphs, which is known to be $\mathcal{NP}$-complete [30]. Given a graph G on k vertices, construct the graph H as follows. Start by taking a complete bipartite graph with bipartition $V(H) = A \cup B$ where $|A| = 1 + r$ and $|B| = (C + 1)k + r$ and r is an integer greater than $\frac{\alpha((C+1)k+1)}{1-2\alpha} - 1$ so that $|A|/(|A| + |B|) > \alpha$. Now we insert $C + 1$ disjoint copies of G into B to form H. Note that $\delta(H) \geq r + 1$ and by choice of r we have $\delta(H) \geq \alpha|V(H)|$. It is easy to see that H has a cycle of length at least $|V(H)| - C$ if and only if G has a Hamilton path. This gives the desired reduction since $|V(H)|$ is linear in $|V(G)|$. □

3.2 PRELIMINARIES

We follow general graph theory notation found e.g. in [20].

Given a graph G, we denote its vertex and edge sets by $V(G)$ and $E(G)$ respectively. For a vertex $v \in V(G)$, we write $N(v)$ for the neighbors of v in G and write $d_G(v) := |N(v)|$ for the degree of v. Given $S \subseteq V(G)$, we also write $d_S(v) := |N(v) \cap S|$ for the degree of v in S. We denote with $\delta(G)$ the smallest degree among vertices in G.

We write $H \subseteq G$ to mean that H is a subgraph of G, i.e. $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. We define $E_G(S) := \{ab \in E(G) \mid a, b \in S\}$ and we write $G[S]$ for the graph induced by G on S, i.e. the graph with vertex set S and edge set $E_G(S)$. For $S, T \subseteq V(G)$, we define $E_G(S,T) := \{xy \in E(G) \mid x \in S, y \in T\}$ and $e_G(S,T) := |E_G(S,T)|$. We will sometimes omit the subscript if it is clear. For $S, T \subseteq V(G)$ disjoint, we write $G[S,T] := (S \cup T, E_G(S,T))$ for the bipartite graph induced by G between S and T. We often denote the complement of $S \subseteq V(G)$ by $\overline{S}$ i.e. $\overline{S} := V(G) \setminus S$.

We write $a \ll b$ to mean that $a \leq f(b)$ for some implicitly given non-decreasing function $f : (0,1] \to (0,1]$. Informally, this is understood to mean that a is small enough in relation to b. We sometimes also write $a \ll_f b$ when we wish to be specific about the function f.

3.2.1 *Spectral partitioning*

Given a graph G and $S \subseteq V(G)$, the *conductance* of S, written $\Phi(S) = \Phi_G(S)$, is given by

$$\Phi(S) := \frac{e_G(S, \overline{S})}{\min(\mathrm{vol}_G(S), \mathrm{vol}_G(\overline{S}))},$$

where $\mathrm{vol}_G(S) = \mathrm{vol}(S) := \sum_{i \in S} d_G(i)$ refers to the volume of S. The *edge expansion* $\Phi(G)$ of G is defined by $\Phi(G) := \min_{S \subseteq V(G)} \Phi(S)$.

We write $A_G \in \mathbb{R}^{V(G) \times V(G)}$ for the adjacency matrix of G, where A_G is the matrix whose rows and columns are indexed by vertices of G and is defined by

$$(A_G)_{uv} := \begin{cases} 1 & \text{if } uv \in E(G); \\ 0 & \text{otherwise.} \end{cases}$$

We write

$$L_G := I - D^{-\frac{1}{2}} A_G D^{-\frac{1}{2}}$$

for the normalized Laplacian of G, where $I \in \mathbb{R}^{V(G) \times V(G)}$ is the identity matrix and D is the diagonal matrix of degrees (where $D_{uu} = d_G(u)$ for each $u \in V(G)$ and $D_{uv} = 0$ for $u \neq v$).

Suppose the eigenvalues of L_G are ordered $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Theorem 3.2.1 gives an algorithm for approximating the expansion of G and gives a corresponding partition of the vertices.

Theorem 3.2.1 ([1], [73])**.** For any graph G, we have $\frac{\lambda_2}{2} \leq \Phi(G) \leq \sqrt{2\lambda_2}$ and there is an algorithm that finds $S \subseteq V$ such that $\Phi(S) \leq \sqrt{2\lambda_2}$ in time polynomial in $n = |V(G)|$. In particular, $\Phi(G) \geq \Phi(S)^2/4$.

The inequality $\frac{\lambda_2}{2} \leq \Phi(G) \leq \sqrt{2\lambda_2}$ is often referred to as Cheeger's inequality. There is an analogue of Cheeger's inequality for the largest eigenvalue λ_n and the *bipartiteness ratio* $\beta(G)$. For $y \in \{-1,0,1\}^{V(G)} \setminus \{\mathbf{0}\}$ we define

$$\beta(y) := \frac{\sum_{uv\in E(G)} |y_u + y_v|}{\sum_{v\in V(G)} d_G(v)|y_v|}$$

and $\beta(G) := \min_{y\in\{-1,0,1\}^V\setminus\{\mathbf{0}\}} \beta(y)$. We can think of a small value $\beta(G)$ to mean that G is close to bipartite. In particular, if we set $A = \{v \in V(G) \mid y_v = 1\}$ and $B = \{v \in V(G) \mid y_v = -1\}$ then

$$\beta(y) = \frac{2e_G(A) + 2e_G(B) + e_G(A\cup B, V(G)\setminus(A\cup B))}{\mathrm{vol}_G(A\cup B)}. \tag{3.2.1}$$

Theorem 3.2.2 ([73, 74])**.** For any graph G, we have $\frac{2-\lambda_n}{2} \leq \beta(G) \leq \sqrt{2(2-\lambda_n)}$ and there is an algorithm that finds $y \in \{-1,0,1\}^{V(G)}$ such that $\beta(y) \leq \sqrt{2(2-\lambda_n)}$ in time polynomial in $n = |V(G)|$. In particular, $\beta(G) \geq \beta(y)^2/4$

Remark 3.2.3. The algorithms from both Theorem 3.2.1 and 3.2.2 run in time $O(|E(G)| + |V(G)|\log|V(G)|)$.

3.2.2 *Robust expanders*

The following definitions follow closely those in [54]. Throughout, assume G is an n-vertex graph.

Robust expanders and bipartite robust expanders - Given an n-vertex graph G, and $S \subseteq V(G)$ and parameters $0 < \nu \leq \tau < 1$, we define the *ν-robust neighborhood* of S to be $\mathrm{RN}_{\nu,G}(S) := \{v \in G \mid d_S(v) \geq \nu n\}$. We say G is a *robust (ν,τ)-expander* if for all $S \subseteq V(G)$ with $\tau n \leq |S| \leq (1-\tau)n$ we have $|\mathrm{RN}_{\nu,G}(S)| \geq |S| + \nu n$. We say G is a *bipartite robust (ν,τ)-expander* with bipartition A, B if A, B is a partition of $V(G)$ and for every $S \subseteq A$ with $\tau|A| \leq |S| \leq (1-\tau)|A|$ we have $|\mathrm{RN}_{\nu,G}(S)| \geq |S| + \nu n$. Note that the order of A and B matters here.

Robust expander components and bipartite robust expander components - Given $0 < \rho < 1$ and an n-vertex graph G, we say that $U \subseteq V(G)$ is a *ρ-component* if $|U| \geq \sqrt{\rho}n$ and $e_G(U, \overline{U}) \leq \rho n^2$, where as usual $\overline{U} := V(G) \setminus U$. We say that U is *ρ-close to bipartite* with bipartition A, B if A, B is a partition of U, $|A|, |B| \geq \sqrt{\rho}n$, $||A| - |B|| \leq \rho n$, and $e_G(A, \overline{B}) + e_G(B, \overline{A}) \leq \rho n^2$. We will sometimes call a graph a ρ-component or ρ-close to bipartite if $V(G)$ is itself a ρ-component resp. ρ-close to bipartite. We say that $G[U]$ is a *(ρ, ν, τ)-robust expander component* of G if U is a ρ-component and $G[U]$ is a robust (ν, τ)-expander. We say that $G[U]$ is a *bipartite (ρ, ν, τ)-robust expander component* with bipartition A, B if U is ρ-close to bipartite with bipartition A, B and $G[U]$ is a bipartite robust (ν, τ)-expander with bipartition A, B.

We now introduce the concept of a robust partition, which is central to our result.

Robust partitions - Let $k, \ell, D \in \mathbb{N}$ and $0 < \rho \leq \nu \leq \tau < 1$. Given an n-vertex, D-regular graph G, we say that $\mathcal{V}$ is a *robust partition of G with parameters ρ, ν, τ, k, ℓ* if the following hold:

(D1) $\mathcal{V} = \{V_1, \ldots, V_k, W_1, \ldots, W_\ell\}$ is a partition of $V(G)$;
(D2) for all $1 \leq i \leq k$, $G[V_i]$ is a (ρ, ν, τ)-robust expander component of G;
(D3) for all $1 \leq j \leq \ell$, there exists a partition A_j, B_j of W_j such that $G[W_j]$ is a bipartite (ρ, ν, τ)-robust expander component with bipartition A_j, B_j;
(D4) for all $X, X' \in \mathcal{V}$ and all $x \in X$, we have $d_X(x) \geq d_{X'}(x)$; in particular, $d_X(x) \geq D/m$, where $m := k + \ell$;
(D5) for all $1 \leq j \leq \ell$, we have $d_{B_j}(u) \geq d_{A_j}(u)$ for all $u \in A_j$ and $d_{A_j}(v) \geq d_{B_j}(v)$ for all $v \in B_j$; in particular, $\delta(G[A_j, B_j]) \geq D/2m$;
(D6) $k + 2\ell \leq \lfloor (1 + \rho^{1/3})n/D \rfloor$;
(D7) for all $X \in \mathcal{V}$, all but at most ρn vertices $x \in X$ satisfy $d_X(x) \geq D - \rho n$.

For technical reasons, we also introduce weak robust subpartitions. We will use this definition and Lemma 3.2.4 below only in Section 3.4. A weak robust subpartition differs from a robust partition mainly in that the disjoint subsets need not be a partition of the vertices. Let $k, \ell \in \mathbb{N}_0$ and $0 < \rho \leq \nu \leq \tau \leq \eta < 1$. Given a graph G on n vertices, we say that $\mathcal{U}$ is a *weak robust subpartition* of G with parameters $\rho, \nu, \tau, \eta, k, \ell$ if the following conditions hold:

(D1′) $\mathcal{U} = \{U_1, \ldots, U_k, Z_1, \ldots, Z_\ell\}$ is a collection of disjoint subsets of $V(G)$;

(D2′) for all $1 \le i \le k$, $G[U_i]$ is a (ρ, ν, τ)-robust expander component of G;

(D3′) for all $1 \le j \le \ell$, there exists a partition A_j, B_j of Z_j such that $G[Z_j]$ is a bipartite (ρ, ν, τ)-robust expander component with bipartition A_j, B_j;

(D4′) $\delta(G[X]) \ge \eta n$ for all $X \in \mathcal{U}$;

(D5′) for all $1 \le j \le \ell$, we have $\delta(G[A_j, B_j]) \ge \eta n/2$.

Lemma 3.2.4 (Proposition 6.1 in [54])**.** Let $k, \ell, D \in \mathbb{N}_0$ and suppose that $0 < 1/n \ll \rho \le \nu \le \tau \le \eta \le \alpha^2/2 < 1$. Suppose that G is a D-regular graph on n vertices where $D \ge \alpha n$. Let $\mathcal{V}$ be a robust partition of G with parameters ρ, ν, τ, k, ℓ. Then $\mathcal{V}$ is a weak robust subpartition of G with parameters $\rho, \nu, \tau, \eta, k, \ell$.

3.3 ROBUST PARTITIONS

3.3.1 *Statements of algorithms*

In this section we present an algorithm (Theorem 3.3.20) that we use to find robust partitions (see previous section for the definition) of regular graphs. As mentioned earlier, the main algorithm and its analysis are obtained by combining the robust expander decomposition of regular graphs from [54] together with spectral algorithms for graph partitioning from [73, 74].

We begin by presenting four algorithms in the following lemmas that will eventually be used together to obtain the main algorithm. The proofs appear in Subsection 3.3.2.

Lemma 3.3.1. For each fixed choice of parameters $1/n_0 \ll \rho \ll \nu \ll \rho' \ll \tau \ll \alpha < 1$ there exists a polynomial-time algorithm that does the following. Given a D-regular n-vertex graph $G = (V, E)$ and $U \subseteq V$ as input, where $D \ge \alpha n$, $n \ge n_0$ and $G[U]$ is a ρ-component of G that is not ρ'-close to bipartite, the algorithm determines that either

(i) $G[U]$ is a robust (ν, τ)-expander, or

(ii) U has a partition U_1, U_2 such that U_1, U_2 are ρ'-components,

and in the case of (ii) identifies the partition U_1, U_2. We call this Algorithm 1.

Lemma 3.3.2. For each fixed choice of parameters $1/n_0 \ll \rho \ll \rho' \ll \alpha < 1$ there is a polynomial time algorithm that does the following. Given a D-regular, n-vertex graph $G = (V, E)$ and $U \subseteq V$ as input, where $D \geq \alpha n$, $n \geq n_0$, and $G[U]$ is a ρ-component of G, the algorithm determines that either

(i) $G[U]$ is not ρ-close to bipartite, or
(ii) $G[U]$ is ρ'-close to bipartite,

and in the case of (ii) identifies the corresponding bipartition. We call this Algorithm 2.

Lemma 3.3.3. For each fixed choice of parameters $1/n_0 \ll \rho \ll \nu \ll \rho' \ll \tau \ll \alpha < 1$ there is a polynomial-time algorithm that does the following. Given a D-regular, n-vertex graph $G = (V, E)$ and $U \subseteq V$ and a partition A, B of U as input, where $D \geq \alpha n$, $n \geq n_0$, and $G[U]$ is ρ-close to bipartite with bipartition A, B, the algorithm determines that either

(i) $G[U]$ is a bipartite robust (ν, τ)-expander with bipartition A, B, or
(ii) U has a partition U_1, U_2 such that $G[U_1]$, $G[U_2]$ are ρ'-components,

and in the case of (ii) identifies the partition U_1, U_2 of U. We call this Algorithm 3.

Lemma 3.3.4. For each fixed choice of parameters $1/n_0 \ll \rho \ll \nu \ll \rho' \ll \tau \ll \alpha < 1$ there exists a polynomial-time algorithm that does the following. Given a D-regular n-vertex graph $G = (V, E)$ and $U \subseteq V$ as input, where $D \geq \alpha n$, $n \geq n_0$, and $G[U]$ is a ρ-component, the algorithm determines that either

(i) $G[U]$ is a robust (ν, τ)-expander, or
(ii) $G[U]$ is a bipartite robust (ν, τ)-expander, or
(iii) U has a partition U_1, U_2 such that $G[U_1]$, $G[U_2]$ are ρ'-components,

and in the case of (ii) and (iii) identifies the corresponding partition. We call this Algorithm 4.

Remark 3.3.5. In each of the four lemmas above, the algorithm distinguishes between various cases. It may be that more than one of these cases hold for the given input graph; if so then the algorithm will output any one case that holds for the given graph.

The running time of each of the algorithms is $O(n^3)$, where n is the number of vertices of the input graph. The running time does not depend at all on the fixed parameters (not even as hidden constants in the 'Big O' notation). However in each lemma, the hierarchy is necessary for the fixed parameters in order to guarantee that at least one of the outcomes occurs in the conclusion of the lemma.

3.3.2 *Proofs of correctness of algorithms*

We now give the proofs of Lemmas 3.3.1–3.3.4. We begin with a simple proposition.

Proposition 3.3.6. Let G be an n-vertex D-regular graph with $D \geq \alpha n$ and let U be a ρ-component of G. Then

(i) $|U| \geq D - \sqrt{\rho}n \geq (\alpha - \sqrt{\rho})n$
(ii) There are at most $\frac{2\rho}{\alpha(\alpha-\sqrt{\rho})}|U|$ vertices of degree at most $\frac{1}{2}\alpha n$ in $G[U]$.

Proof. (i) Since G is D-regular and U is a ρ-component, we have $\frac{1}{2}|U|^2 \geq e_G(U) \geq \frac{1}{2}D|U| - \rho n^2$, from which we obtain $|U| \geq D - \frac{\rho n^2}{|U|} \geq D - \sqrt{\rho}n$, where the second inequality uses that $|U| \geq \sqrt{\rho}n$ since it is a ρ-component.

(ii) If the number of vertices of degree at most $\frac{1}{2}\alpha n$ is $\gamma|U|$, then we have

$$(D/2)\gamma|U| + D(1-\gamma)|U| \geq 2e_G(U) \geq D|U| - \rho n^2,$$

from which we get $\gamma \leq \frac{2\rho n^2}{D|U|} \leq \frac{2\rho}{\alpha(\alpha-\sqrt{\rho})}$ using part (i) and $D \geq \alpha n$ for the final inequality. □

Remark 3.3.7. A similar calculation shows that if U is σ-close to bipartite with bipartition A, B, we have $|A|, |B| \geq D - 2\sqrt{\sigma}n \geq (\alpha - 2\sqrt{\sigma})n$.

Proof of Lemma 3.3.1. We will use the algorithm in Theorem 3.2.1 to iteratively find subgraphs of $G[U]$ that are not well connected to the rest of U and remove them until this is no longer possible. If this process continues to a point where the removed parts are large enough then we can show both the removed part and the remaining part each form a ρ'-component. If the process stops before the removed part becomes large then we can show $G[U]$ is a robust expander.

Let $G = (V, E)$ and, in this proof, for any subset $S \subseteq U$ we will use $\overline{S}$ to denote $U \setminus S$ rather than our usual convention where it denotes $V \setminus S$.

Let $n' = |U|$ so that $n' \geq (\alpha - \sqrt{\rho})n \geq \frac{1}{2}\alpha n$ (by Proposition 3.3.6). Let U_0 be the vertices of degree at most $\frac{1}{2}\alpha n$ in $G[U]$ so that $|U_0| \leq \frac{2\rho}{\alpha(\alpha-\sqrt{\rho})}n' \leq \alpha\nu n'/2$ also by Proposition 3.3.6. Note for later that

$$\mathrm{vol}_G(U_0) \leq n|U_0| \leq (2n'/\alpha)(\alpha\nu n'/2) \leq \nu n'^2. \tag{3.3.1}$$

Set $U' := U \setminus U_0$ and choose ϕ such that $\nu \ll \phi \ll \rho'$. We apply Theorem 3.2.1 to $G[U']$ as follows to construct $U_1, U_2, \ldots$. Given U_i, set $\overline{U_i} := U \setminus U_i$ and $G_i := G[\overline{U_i}]$. Apply the algorithm of Theorem 3.2.1 to G_i to output some $S_i \subseteq \overline{U_i}$. By replacing S_i with $U_i \setminus S_i$ if necessary, assume $|S_i| \leq |U_i \setminus S_i|$. If

$$\phi_i := \Phi_{G_i}(S_i) > \phi \text{ or } |U_i| \geq \tfrac{1}{3}|U|$$

then stop. Otherwise set $U_{i+1} = U_i \cup S_i$ and repeat. In this way we obtain sets $S_0, \ldots, S_{t-1}$ and $U_0, \ldots, U_t$ in polynomial time. Note that $|U_{t-1}| < \frac{1}{3}|U|$, so

$$|U_t| = |U_{t-1}| + |S_{t-1}| \leq |U_{t-1}| + \frac{1}{2}(|U| - |U_{t-1}|) \leq \frac{2}{3}|U|. \tag{3.3.2}$$

There are two cases to consider:

(a) $|U_t| > \frac{1}{4}\rho' n'$ and
(b) $|U_t| \leq \frac{1}{4}\rho' n'$.

Claim 3.3.8. In case (a), U_t, $\overline{U_t}$ are ρ'-components.

Claim 3.3.9. In case (b), $G[U]$ is a robust (ν, τ)-expander.

Since we can output $U_t, \overline{U_t}$ in polynomial time, these two claims prove Lemma 3.3.1.

Proof of Claim 3.3.8. Since we are in case (a), note that $\Phi_{G_i}(S_i) \leq \phi$ for all $i = 1, \ldots, t$ and so

$$e_G(S_i, U_i \setminus S_i) \leq \phi \mathrm{vol}_{G_i}(S_i) \leq \mathrm{vol}_G(S_i). \tag{3.3.3}$$

Recall also that $U_t = U_0 \cup (\bigcup_{i=0}^{t-1} S_i)$. Using that volume is additive, i.e. $\mathrm{vol}_G(U_t) = \mathrm{vol}_G(U_0) + \sum_{i=0}^{t-1} \mathrm{vol}_G(S_i)$, we have

$$\begin{aligned} e_G(U_t, \overline{U_t}) = e_G(U_0, \overline{U_t}) + \sum_{i=0}^{t-1} e_G(S_i, \overline{U_t}) &\leq \mathrm{vol}_G(U_0) + \sum_{i=0}^{t-1} e_G(S_i, U_i \setminus S_i) \\ &\overset{(3.3.1),(3.3.3)}{\leq} \nu n'^2 + \sum_{i=0}^{t-1} \phi \ \mathrm{vol}_G(S_i) \\ &\leq \nu n'^2 + \phi \ \mathrm{vol}_G(U_t) \leq \nu n'^2 + \phi |U_t| n. \end{aligned}$$

Therefore

$$e_G(U_t,\overline{U_t}) \le \nu n'^2 + \phi|U_t|n \overset{(3.3.2)}{\le} \nu n'^2 + \frac{2}{3}\phi|U|n \overset{\text{Prop 3.3.6}}{\le} \nu n'^2 + \frac{\phi}{\alpha-\sqrt{\rho}}n'^2 \overset{\nu,\phi\ll\rho'}{\le} \frac{1}{2}\rho' n'^2.$$

Hence $e_G(U_t, V\setminus U_t) \le \frac{1}{2}\rho' n'^2 + \rho n^2 \le \rho' n^2$ since $U_t \subseteq U$ and U is a ρ-component. Similarly $e_G(\overline{U_t}, V\setminus \overline{U_t}) \le \rho' n^2$. Also, $|U_t|$, $|\overline{U_t}| \ge \frac{1}{4}\rho' n$ by (a) and (3.3.2). However, by Proposition 3.3.6, we in fact have $|U_t|$, $|\overline{U_t}| \ge (\alpha - \rho'^2)n \ge \sqrt{\rho'}n$, so U_t and $\overline{U_t}$ are ρ'-components. ◀

Proof of Claim 3.3.9. First some observations. Since case (b) holds, $|U_t| \le \frac{1}{4}\rho' n' \le \frac{1}{2}\tau n' \le \frac{1}{3}|U|$ and $\phi_t = \Phi_{G_t}(S_t) > \phi$.

Also, $\delta(G_t) = \delta(G[\overline{U_t}]) \ge \min_{x\in\overline{U_t}} d_U(x) - |U_t| \ge \frac{1}{2}\alpha n - \frac{1}{2}\tau n' \ge \frac{1}{3}\alpha n$, where the penultimate inequality follows from our choice of U_0. By Theorem 3.2.1, for all $R \subseteq V(G_t) = U\setminus U_t$ we have $\Phi_{G_t}(R) \ge \Phi(G_t) \ge \phi_t^2/4 \ge \phi^2/4$, i.e.

$$e_{G_t}(R,R') \ge \frac{\phi^2}{4}\min(\text{vol}(R),\text{vol}(R')) \ge \frac{1}{12}\phi^2\alpha n \min(|R|,|R'|)$$

where $R' = \overline{U_t}\setminus R = U\setminus(U_t\cup R)$. Furthermore, for $R \subseteq U$ and recalling $\overline{R} := U\setminus R$, we have

$$\begin{aligned} e_{G[U]}(R,\overline{R}) \ge e_{G[U]}(R\setminus U_t, \overline{R}\setminus U_t) &\ge \frac{1}{12}\phi^2\alpha n \min(|R\setminus U_t|, |\overline{R}\setminus U_t|) \\ &\ge \frac{1}{12}\phi^2\alpha n\left(\min(|R|,|\overline{R}|) - \frac{1}{4}\rho' n'\right). \end{aligned} \tag{3.3.4}$$

We will now show that $G[U]$ is a (ν,τ)-expander by assuming that $G[U]$ does not expand and deducing that $G[U]$ is ρ'-close to bipartite, contradicting the premise of the lemma.

Suppose there exists $S \subseteq U$ with $\tau n' \le |S| \le (1-\tau)n'$ such that $N = \text{RN}_{\nu,G[U]}(S)$ satisfies $|N| < |S| + \nu n$. Since $\tau n' \le |S| \le (1-\tau)n'$, we have $\frac{1}{4}\rho' n' \le \frac{1}{2}\tau n' \le \frac{1}{2}\min(|S|,|\overline{S}|)$ so by (3.3.4), we have

$$e_{G[U]}(S,\overline{S}) \ge \frac{1}{24}\phi^2\alpha n \min(|S|,|\overline{S}|). \tag{3.3.5}$$

Claim 3.3.10. We may assume $\frac{1}{4}\alpha n \le |S| \le |U| - \frac{1}{4}\alpha n$.

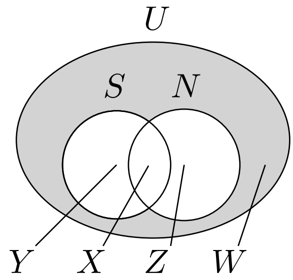


Figure 8: Overview of subsets mentioned in the coming section.

Proof of Claim 3.3.10. If $|S| < \frac{1}{4}\alpha n$ then $e_G(S,\overline{S}) \geq |S|(\alpha n - |S|) - \rho n^2$ and $e_G(S,\overline{S}) \leq |N||S| + |U \setminus N|\nu n \leq |N||S| + \nu n^2$, so combining these inequalities and rearranging, we obtain

$$|N| \geq \alpha n - |S| - (\rho+\nu)\frac{n^2}{|S|} \geq \alpha n - |S| - (\rho+\nu)\frac{n^2}{\tau n'}$$
$$\overset{Prop\,3.3.6}{\geq} \alpha n - \frac{1}{4}\alpha n - (\rho+\nu)\frac{n'}{\tau(\alpha-\sqrt{\rho})^2} \geq \frac{1}{2}\alpha n' \geq |S| + \nu n,$$

contradicting our choice of S.

Similarly if $|S| > |U| - \frac{1}{4}\alpha n$ recall that by Proposition 3.3.6 that all but the $\gamma n'$ vertices in U_0 have degree at least $\frac{1}{2}\alpha n$ in U and so for all $x \in U \setminus U_0$, we have

$$d_S(x) \geq \frac{1}{2}\alpha n - |U \setminus S| \geq \frac{1}{4}\alpha n \geq \nu n.$$

Hence $N \supseteq U \setminus U_0$ and so $|N| \geq |U| - |U_0| \geq (1-\nu)n' \geq |S| + \nu n'$, a contradiction. This proves Claim 3.3.10. ◀

We continue with the proof of Claim 3.3.9. We define $Y = S \setminus N$, $X = S \cap N$, $Z = N \setminus S$, $W = U \setminus (S \cup N)$; see Figure 8. Since each vertex in Y has at most νn neighbors in S and since G is D-regular and U is a ρ-component we have $e_G(Y,\overline{S}) \geq D|Y| - \rho n^2 - \nu n^2$. Using this, we obtain

$$e_G(Y,Z) = e_G(Y,\overline{S}) - e_G(Y,W) \geq D|Y| - \rho n^2 - \nu n^2 - |W|\nu n$$
$$\geq D|Y| - 3\nu n^2. \tag{3.3.6}$$

On the other hand $e_G(Z,Y) \leq D|Z|$, which together with (3.3.6) implies after rearranging that $|Z| \geq |Y| - \frac{3\nu}{\alpha}n$. Also $|Z| \leq |Y| + \nu n$; otherwise S does not violate (ν,τ)-expansion. Hence we have shown

$$|Y| - \frac{3\nu}{\alpha}n \leq |Z| \leq |Y| + \nu n. \tag{3.3.7}$$

Considering W (and taking $\overline{W} := U \setminus W$), we see

$$e_G(W,\overline{W}) = e_G(W,S) + e_G(Z,W) \le e_G(W,S) + (D|Z| - e_G(Z,Y))$$

$$\overset{(3.3.7),(3.3.6)}{\le} \nu n^2 + D(|Y| + \nu n) - (D|Y| - 3\nu n^2) \le 5\nu n^2, \tag{3.3.8}$$

as well as

$$\frac{1}{12}\phi^2 \alpha n \min(|W|,|\overline{W}|) - \frac{1}{48}\phi^2\alpha\rho' n n' \overset{(3.3.4)}{\le} e_G(W,\overline{W}) \overset{(3.3.8)}{\le} 5\nu n^2.$$

Since $|\overline{W}| \ge |S| \ge \tau n' > 2\rho' n'$, we must have

$$|W| \le \frac{60\nu n}{\phi^2\alpha} + \frac{1}{4}\rho' n' \le \frac{1}{2}\rho' n'. \tag{3.3.9}$$

Now consider $Y \cup Z$ (and recall $\overline{Y \cup Z} := U \setminus (Y \cup Z)$). We have

$$e_G(Y\cup Z, \overline{Y\cup Z}) \le D|Y \cup Z| - 2e_G(Y,Z)$$

$$\overset{(3.3.7),(3.3.6)}{\le} D(2|Y| + \nu n) - 2(D|Y| - 3\nu n^2) \le 7\nu n^2. \tag{3.3.10}$$

Combining this with an application of (3.3.4)

$$\frac{1}{12}\phi^2\alpha n(\min(|Y\cup Z|, |\overline{Y\cup Z}|) - \frac{1}{4}\rho' n') \overset{(3.3.4)}{\le} e_G(Y\cup Z, \overline{Y\cup Z}) \overset{(3.3.10)}{\le} 7\nu n^2,$$

and hence

$$\min(|Y\cup Z|, |\overline{Y\cup Z}|) \le 84\frac{\nu n}{\phi^2\alpha} + \frac{1}{4}\rho' n' \le \frac{1}{2}\rho' n'.$$

If $|Y \cup Z| \le \frac{1}{2}\rho' n'$, then

$$|S| = |U| - |W| - |Z| \ge |U| - |W| - |Y \cup Z|$$

$$\overset{(3.3.9)}{\ge} n' - \frac{1}{2}\rho' n' - \frac{1}{2}\rho' n' \ge (1-\tau)n',$$

a contradiction. So we have

$$|\overline{Y\cup Z}| \le \frac{1}{2}\rho' n'. \tag{3.3.11}$$

Finally we show that Y,$\overline{Y}$ gives a partition that shows $G[U]$ is ρ'-close to bipartite, giving a contradiction. Note that $|\overline{Y}| = |Z| + |\overline{Y \cup Z}|$, so

$$|Y| - \frac{3\nu}{\alpha}n \overset{(3.3.7)}{\leq} |Z| \leq |\overline{Y}| = |Z| + |\overline{Y \cup Z}| \overset{(3.3.7),(3.3.11)}{\leq} |Y| + \nu n + \frac{1}{2}\rho' n' \leq |Y| + \frac{3}{4}\rho' n'.$$

Therefore,

$$||Y| - |\overline{Y}|| \leq \frac{3}{4}\rho' n'. \tag{3.3.12}$$

If ρ' is small enough, e.g. $\rho' \leq \frac{1}{10}$, this also gives us $|Y|$, $|\overline{Y}| \geq \sqrt{\rho'}n'$. Also

$$\begin{aligned} e_G(Y, V \setminus \overline{Y}) + e_G(\overline{Y}, V \setminus Y) &\leq D|Y \cup \overline{Y}| - 2e_G(Y, \overline{Y}) \\ &\leq D|U| - 2e_G(Y, Z) \overset{(3.3.6)}{\leq} Dn' - 2(D|Y| - 3\nu n^2) \\ &\overset{(3.3.12)}{\leq} Dn' - D|Y| - D\left(|\overline{Y}| - \frac{3}{4}\rho' n'\right) + 6\nu n^2 \leq \frac{4}{5}D\rho' n' \leq \rho' n'^2. \end{aligned}$$

So Y,$\overline{Y}$ is a partition of U showing $G[U]$ is ρ'-close to bipartite, a contradiction, completing the proof of the claim and the lemma. ◀

□

Proof of Lemma 3.3.2. The idea is to repeatedly apply the algorithm in Theorem 3.2.2 and iteratively remove vertices that are assigned to bipartite parts until the remaining induced graph is either small or far from bipartite.

We choose β such that $\rho \ll \beta \ll \rho'$. Set $U_0 = \emptyset$, and given U_i, let $G_i = G[U \setminus U_i]$. Let y be obtained from running the algorithm in Theorem 3.2.2 on G_i. We set $U_{i+1} = U_i \cup A_i \cup B_i$, where $A_i := \{v \mid y_v = 1\}$ and $B_i := \{v \mid y_v = -1\}$ and we set $\beta_i = \beta(y)$. Note that $G_{i+1} \subset G_i$. We continue until either

(a) $|G_i| \leq \rho' n$ or
(b) $\beta_i \geq \beta$.

Let t be the first index where (a) or (b) occurs.

Claim 3.3.11. If $|G_t| \leq \rho' n$, then $G[U]$ is ρ'-close to bipartite.

Claim 3.3.12. If $\beta_t > \beta$ and $|G_t| \geq \rho' n$, then $G[U]$ is not ρ-close to bipartite.

Note that these two claims together prove the lemma since we can compute the β_i and the G_i in polynomial time (and for the first claim, the proof will show how to compute the corresponding partition).

Proof of Claim 3.3.11. Let $R = U \setminus U_t$, i.e. the set of vertices that are not part of some A_j or B_j for $j \le t$. Note that $|R| \le \rho' n$. For each $j \le t$, using the definition of A_j, B_j and (3.2.1), we have

$$E_j := 2e_{G_j}(A_j) + 2e_{G_j}(B_j) + e_{G_j}(A_j \cup B_j, U \setminus U_{j+1}) \le \beta \text{ vol}_{G_j}(A_j \cup B_j).$$

First we note that for each $j \le t$, we have

$$\begin{aligned} e_G(U_j, U \setminus U_j) \le \sum_{i=0}^{j-1} e_G(A_i \cup B_i, U \setminus U_{i+1}) &\le \beta \sum_{i=0}^{j-1} \text{vol}_{G_i}(A_i \cup B_i) \\ &\le \beta \text{vol}_G(U_j) \le \frac{1}{10}\rho' Dn, \end{aligned} \tag{3.3.13}$$

where the final inequality follows by our choice of $\beta \ll \rho'$ and $\text{vol}_G(U_j) \le Dn$. In particular, for each $j < t$, we have

$$e_G(A_j, U_j) \le e_G(U_j, U \setminus U_j) \le \frac{1}{10}\rho' Dn.$$

Next, we claim that for each j, it holds that $||A_j| - |B_j|| \le \rho' n$. Assume for a contradiction that $|A_j| - |B_j| \ge \rho' n$ for some j. First we note that

$$\begin{aligned} e_{G_j}(A_j, \overline{B_j}) &\ge (|A_j| - |B_j|)D - e_G(U, \overline{U}) - e_G(A_j, U_j) \\ &\ge \rho' Dn - \rho n^2 - \frac{1}{10}\rho' Dn \ge \frac{1}{2}\rho' Dn, \end{aligned}$$

where we use $\rho \ll \rho'$ for the last inequality. On the other hand we have $e_{G_j}(A_j, \overline{B_j}) \le e_G(A_j, U_j) \le \frac{1}{10}\rho' Dn$, a contradiction.

By the preceding claim, we can form a partition A, B of U such that (i) for each $j < t$, either $A_j \subseteq A$ and $B_j \subseteq B$, or $A_j \subseteq B$ and $B_j \subseteq A$ and (ii) $||A| - |B|| \le \rho' n$. Indeed we can start with an arbitrary partition satisfying (i) and then iteratively swap suitable A_j and B_j if this reduces the value of $||A| - |B||$. (Note that A and B also contain vertices of R (i.e. vertices not belonging to any A_j or B_j) that can be freely moved to reduce $||A| - |B||$). It is easy to see A, B can be computed in polynomial time and we shall see below that this partition demonstrates that $G[U]$ is ρ'-close to bipartite.

To see this, we count edges not in $E_G(A, B)$. We have

$$\begin{aligned} e_G(A) + e_G(B) + e_G(A \cup B, \overline{U}) &\leq \sum_{j=0}^{i-1} E_j + \mathrm{vol}_{G[R]}(R) + e_G(U, \overline{U}) \\ &\leq \underbrace{\beta}_{\ll \rho'} \underbrace{\mathrm{vol}_{G[U]}(U \setminus R)}_{\leq n^2} + (\rho' n)^2 + \rho n^2 \leq \rho' n^2. \end{aligned}$$

◀

Proof of Claim 3.3.12. Define

$$\begin{aligned} \beta'(G) &:= \min_{y \in \{-1,1\}^{V(G)}} \frac{\sum_{uv \in E(G)} |y_u + y_v|}{\sum_{v \in V(G)} d_G(v)|y_v|} \geq \beta(G), \\ \overline{\beta}(G) &:= \min_{A,B \text{ bipartition of } G} e_G(A) + e_G(B). \end{aligned}$$

Then we have $\overline{\beta}(G[U]) \geq \overline{\beta}(G_t)$ and recalling that $V(G_t) = U \setminus U_t$, we have

$$2\frac{\overline{\beta}(G_t)}{\mathrm{vol}_{G_t}(U \setminus U_t)} = \beta'(G_t) \geq \beta(G_t) \geq \frac{\beta_t^2}{4} \geq \frac{\beta^2}{4}, \tag{3.3.14}$$

where we use the definition of β_i and Theorem 3.2.2. Then we have

$$\begin{aligned} \mathrm{vol}_{G_t}(U \setminus U_t) &\geq D|U \setminus U_t| - \rho n^2 - e_G(U_t, U \setminus U_t) \\ &\geq \rho' D n - \rho n^2 - \frac{1}{10}\rho' D n \geq \frac{1}{2}\rho' D n, \end{aligned}$$

where we have used that U is a ρ-component, (3.3.13), and $\rho \ll \rho'$. Combining with (3.3.14) we see

$$\overline{\beta}(G) \geq \frac{\beta^2}{8} \mathrm{vol}_{G_t}(U \setminus U_t) \geq \frac{\beta^2}{16}\rho' D n > \rho n^2.$$

◀

This completes the proof of the lemma. □

Proof of Lemma 3.3.3. Fix ϕ such that $\nu \ll \phi \ll \rho'$. As in Lemma 3.3.1, we use the algorithm in Theorem 3.2.1 to iteratively find poorly connected subgraphs of $G[U]$ and remove them.

In polynomial time, we can find $S_0, \dots, S_{t-1}$, $U_0, \dots, U_t$, and $\phi_1, \dots, \phi_t$, which are defined and found in exactly the same way as in the proof of Lemma 3.3.1, so again, we have $\phi_t > \phi$ or $|U_t| \geq \frac{1}{3}|U|$. There are two cases:

(a) $|U_t| > \frac{1}{4}\rho' n'$ and
(b) $|U_t| \leq \frac{1}{4}\rho' n'$.

Claim 3.3.13. In case (a), U_t, $\overline{U_t} := U \setminus U_t$ are ρ'-components.

Noting that $G[U]$ is a ρ-component, the proof of Claim 3.3.8 holds here as well.

Claim 3.3.14. In case (b), $G[U]$ is a robust bipartite (ν, τ)-expander with bipartition A, B.

Once again, the two claims together prove the lemma since we can compute $U_t, \overline{U}_t$ (which give the partition U_1, U_2 in the statement of the lemma) in polynomial time.

Proof of Claim 3.3.14. As in (3.3.4) in the proof of Claim 3.3.9, for $S \subseteq U$ and $\overline{S} = U \setminus S$ we have

$$e_{G[U]}(S, \overline{S}) \geq \frac{1}{12}\phi^2 \alpha n \left(\min(|S|, |\overline{S}|) - \frac{1}{4}\rho' n'\right) \tag{3.3.15}$$

We will show that $G[U]$ is a bipartite robust expander by assuming the existence of a non-expanding set and finding a contradiction.

Suppose $A^* \subseteq A$ with $\tau|A| \leq |A^*| \leq (1-\tau)|A|$, let $B^* := \mathrm{RN}_{G[U]}(A^*) \cap B$ and assume $|B^*| < |A^*| + \nu n$. Define $\hat{A} := A \setminus A^*$ and $\hat{B} := B \setminus B^*$. We will give an upper bound on $e_G(A^* \cup B^*, \hat{A} \cup \hat{B})$ that contradicts (3.3.15). Indeed, we have (suppressing the subscript G)

$$\begin{aligned} e(A^* \cup B^*, \hat{A} \cup \hat{B}) &\leq e(A^*, \hat{A}) + e(B^*, \hat{B}) + e(A^*, \hat{B}) + e(B^*, \hat{A}) \\ &\leq \rho n^2 + \nu n^2 + e(B^*, \hat{A}), \end{aligned}$$

where we used that $e(A^*, \hat{A}) + e(B^*, \hat{B}) \leq \rho n^2$ (since G is ρ-close to bipartite) and $e(A^*, \hat{B}) < \nu n^2$ (since every vertex in $\hat{B}$ has at most νn neighbors in A^*). In order to bound $e(B^*, \hat{A})$, we have

$$\begin{aligned} e(B^* \hat{A}) &\leq |B^*|D - e(B^*, A^*) \\ &\leq (|A^*| + \nu n)D - [|A^*|D - e(A^*, \hat{A}) - e(A^* \hat{B}) - e(A^*, \overline{U})] \\ &\leq \nu n|D| + \rho n^2 + \nu n^2 \leq \rho n^2 + 2\nu n^2, \end{aligned}$$

where we used that $e(A^*, \hat{B}) \leq \nu n^2$ (as above) and $e(A^*, \hat{A}) + e(A^*, \overline{U}) \leq \rho n^2$ (since U is ρ-close to bipartite). Combining, we obtain

$$e_G(A^* \cup B^*, \hat{A} \cup \hat{B}) \leq 2\rho n^2 + 3\nu n^2 \leq 5\nu n^2. \tag{3.3.16}$$

However, as $\min(|A^* \cup B^*|, |\hat{A} \cup \hat{B}|) \geq \tau|A| \geq \tau\frac{1}{3}|U|$ (using Remark 3.3.7), with (3.3.15) we have

$$e_G(A^* \cup B^*, \hat{A} \cup \hat{B}) \geq \frac{1}{12}\phi^2\alpha n\left(\frac{1}{3}\tau|U| - \frac{1}{4}\rho'|U|\right) > 5\nu n^2,$$

using $|U| \geq \frac{1}{2}\alpha n$ by Proposition 3.3.6 and our choice of parameters, which contradicts (3.3.16). ◀

This completes the proof of the lemma. □

Proof of Lemma 3.3.4. Fix ρ_1, ρ_2, ν_2 such that $\rho \ll \nu \ll \rho_1 \ll \rho_2 \ll \nu_2 \ll \rho'$. We run Algorithm 2 on U with (ρ_1, ρ_2) playing the roles of (ρ, ρ'). The algorithm determines either that

- $G[U]$ is not ρ_1-close to bipartite, or
- $G[U]$ is ρ_2-close to bipartite (and outputs a bipartition A, B of U that demonstrates this).

In the first case, we apply Algorithm 1 with (ρ, ν, ρ_1) playing the roles of (ρ, ν, ρ') and the algorithm either concludes that $G[U]$ is a robust (ν, τ)-expander, or it outputs a partition U_1, U_2 of U such that U_1 and U_2 are ρ_1-components and hence are also ρ'-components.

In the second case, we apply Algorithm 3 with (ρ_2, ν_2, ρ') playing the roles of (ρ, ν, ρ') and the algorithm either concludes that $G[U]$ is a bipartite robust (ν_2, τ)-expander and hence also a bipartite robust (ν, τ)-expander (and it outputs a bipartition A, B of U to demonstrate this) or it outputs a partition U_1, U_2 of U such that U_1 and U_2 are ρ'-components. □

3.3.3 *Recognizing robust expanders*

In this subsection, we make a small digression to partially address a question of Kühn and Osthus from [58]; the result of this subsection will not be needed in the remainder of the chapter. Using the Szemerédi Regularity Lemma, Kühn and Osthus [58] give a polynomial time algorithm for deciding whether a graph[2] is a robust (ν, τ)-expander or whether it is not a (ν', τ)-expander (provided $\nu \ll \nu'$, which is the case in all applications). They asked whether the use of the Szemerédi Regularity Lemma can be avoided, and we answer this affirmatively for regular graphs.

2 In fact, their algorithm works more generally for digraphs.

Corollary 3.3.15. For each fixed choice of parameters $0 \leq \nu \ll \nu' \ll \tau \ll \alpha < 1$ there exists a polynomial-time algorithm that does the following. Given a D-regular n-vertex graph $G = (V, E)$, where $D \geq \alpha n$, the algorithm determines that either

(i) G is a robust (ν, τ)-expander, or
(ii) G is not robust (ν', τ)-expander,

and in case (ii) the algorithm finds a set $S \subseteq V$ such that $\tau n \leq |S| \leq (1-\tau)n$ and $|\mathrm{RN}_{\nu',G}(S)| \leq |S| + \nu' n$.

Proof. The proof is a variation of Lemma 3.3.4. First choose parameters $1/n_0 \ll \rho \ll \nu \ll \rho_1 \ll \rho_2 \ll \nu' \ll \tau \ll \alpha \ll 1$. If $n \leq n_0$ then we check whether (i) or (ii) holds by exhaustive search in constant time.

If $n \geq n_0$, we apply Algorithm 2 to G with (ρ_1, ρ_2, V) playing the roles of (ρ, ρ', U) (and thinking of $G = G[V]$ as a ρ_1-component of G). The algorithm determines that either

(a) G is ρ_2-close to bipartite (and gives a partition A, B of V showing this), or
(b) G is not ρ_1-close to bipartite.

In case (b) we apply Algorithm 1 with (ρ, ν, ρ_1, V) playing the roles of (ρ, ν, ρ', U) (and thinking of $G = G[V]$ as a ρ-component of G), and the algorithm determines that either

(bi) $G = G[V]$ is a robust (ν, τ)-expander;
(bii) $U = V$ has a partition U_1, U_2 such that U_1, U_2 are ρ_1-components.

In case (bi), we are done. In case (a) and (bii), we show G is not a robust (ν', τ)-expander. Indeed, in case (a), assume that $|A| \leq |B|$. We have $|A|, |B| \geq \frac{1}{2}\alpha n \geq 2\tau n$ by Remark 3.3.7, so $\tau n \leq |B| \leq (1-\tau)n$. We cannot have that $|\mathrm{RN}_{\nu',G}(B)| \geq |B| + \nu' n$, for otherwise $|\mathrm{RN}_{\nu',G}(B) \cap B| \geq \nu' n$ and therefore $e_G(B, \overline{A}) = e_G(B) \geq \frac{1}{2}\nu'^2 n^2 > \rho_2 n^2$, contradicting that G is ρ_2-close to bipartite. So G is not a robust (ν', τ)-expander in this case and the algorithm outputs $S = B$.

Similarly in case (bii) we know that $|U_1|, |U_2| \geq \frac{1}{2}\alpha n \geq 2\tau n$ by Proposition 3.3.6 and so $\tau n \leq |U_1| \leq (1-\tau)n$. Also, we cannot have that $|\mathrm{RN}_{\nu',G}(U_1)| \geq |U_1| + \nu' n$, for otherwise $|\mathrm{RN}_{\nu',G}(U_1) \cap U_2| \geq \nu' n$ and therefore $e_G(U_1, U_2) \geq \nu'^2 n^2 > \rho_1 n^2$, contradicting that U_1 is a ρ_1-component. So G is not a robust (ν', τ)-expander in this case and the algorithm outputs $S = U_1$. □

3.3.4 *Assembling the robust partition*

We begin with several basic facts from [54]. The first two, Lemmas 3.3.16 and 3.3.17, are basic facts about (bipartite) robust expanders, which are taken from [54] unchanged and their proofs are included for completeness.

Lemma 3.3.16. Let $0 < \nu \ll \tau < 1$. Suppose that G is a graph and U, $U' \subseteq V(G)$ are such that $G[U]$ is a robust (ν, τ)-expander and $|U \triangle U'| \leq \nu|U|/2$. Then $G[U']$ is a robust $(\nu/2, 2\tau)$-expander.

Proof. The statement immediately follows by considering a set $S \subseteq U'$ with $2\tau|U'| \leq |S| \leq (1-2\tau)|U'|$ and considering its robust neighborhood. As $\tau|U| \leq |S \cap U| \leq (1-\tau)|U|$, we have $|\mathrm{RN}_{\nu,U}(S \cap U)| \geq |S \cap U| + \nu|U| \geq |S| - |U \setminus U'| + \nu|U|$. With $|\mathrm{RN}_{\nu,U}(S \cap U) \cap U'| \geq |\mathrm{RN}_{\nu,U}(S \cap U)| - |U' \setminus U|$ it follows that $|\mathrm{RN}_{\nu/2,U'}(S)| \geq |S| + \nu/2|U'|$. □

Lemma 3.3.17. Let $0 < 1/n \ll \rho \leq \gamma \ll \nu \ll \tau \ll \alpha < 1$ and suppose that G is a D-regular graph on n vertices where $D \geq \alpha n$.

(i) Suppose that $G[A \cup B]$ is a bipartite (ρ, ν, τ)-robust expander component of G with bipartition A, B. Let $A', B' \subseteq V(G)$ be such that $|A \triangle A'| + |B \triangle B'| \leq \gamma n$. Then $G[A' \cup B']$ is a bipartite $(3\gamma, \nu/2, 2\tau)$-robust expander component of G with bipartition A', B'.

(ii) Suppose that $G[U]$ is a bipartite (ρ, ν, τ)-robust expander component of G. Let $U' \subseteq V(G)$ be such that $|U \triangle U'| \leq \gamma n$. Then $G[U']$ is a bipartite $(3\gamma, \nu/2, 2\tau)$-robust expander component of G.

Proof. We start with (i). To see that $G[A' \cup B']$ is 3γ-close to bipartite, we see that $|A'|, |B'| \geq D - 2\sqrt{\rho} \geq \sqrt{3\gamma}n$ by Remark 3.3.7. We have that $||A'| - |B'|| \leq ||A| - |B|| + \gamma n \leq 3\gamma n$ and $e(A', \overline{B'}) + e(B', \overline{A'}) \leq e(A, \overline{B}) + e(B, \overline{A}) + 2(|A' \triangle A| + |B' \triangle B|)n \leq 3\gamma n$. $G[A' \cup B']$ is a bipartite $(\nu/2, 2\tau)$-robust expander by a straightforward calculation as in the proof of Lemma 3.3.16. It is easy to see that part (ii) follows from (i). □

The non-algorithmic versions of the next two lemmas can be found in [54]; we use a simple greedy procedure to make them algorithmic. These lemmas will be used later to ensure conditions (D4), (D5), and (D7) when constructing our robust partition.

Lemma 3.3.18. Let $m, n, D \in \mathbb{N}$ and $0 < 1/n_0 \ll \rho \ll \alpha, 1/m \leq 1$. Let G be a D-regular graph on n vertices where $n \geq n_0$ and $D \geq \alpha n$. Suppose that

$\mathcal{U} := \{U_1, \dots, U_m\}$ is a partition of $V(G)$ such that U_i is a ρ-component for each $1 \le i \le m$. Then G has a vertex partition $\mathcal{V} := \{V_1, \dots, V_m\}$ such that

(i) $|U_i \triangle V_i| \le \rho^{1/3} n$;
(ii) V_i is a $\rho^{1/3}$-component for each $1 \le i \le m$;
(iii) if $x \in V_i$, then $d_{V_i}(x) \ge d_{V_j}(x)$ for all $1 \le i, j \le m$. In particular, $d_V(x) \ge D/m$ for all $x \in V$ and all $V \in \mathcal{V}$;
(iv) for all but at most $\rho^{1/3} n$ vertices $x \in V_i$ we have $d_{V_i}(x) \ge D - 2\sqrt{\rho} n$.

Furthermore, (for fixed n_0, ρ, α, m satisfying the hierarchy above) there is an algorithm that finds such a vertex partition $\mathcal{V}$ in time polynomial in n.

Proof. For each $1 \le i \le m$, let X_i be the collection of vertices $y \in U_i$ with $d_{\overline{U_i}}(x) \ge \sqrt{\rho} n$. Since U_i is a ρ-component, we have $|X_i| \le \sqrt{\rho} n$ (otherwise $e(U_i, \overline{U_i}) \ge \rho n^2$). Let $W_i := U_i \setminus X_i$. Then each $x \in W_i$ satisfies

$$d_{W_i}(x) = D - d_{\overline{U_i} \cup X_i}(x) \ge D - \sqrt{\rho} n - |X_i| \ge D - 2\sqrt{\rho} n. \qquad (3.3.17)$$

We now redistribute the vertices of $X := \cup_{1 \le i \le m} X_i$ as follows: Iteratively move any $x \in X \cap U_i$ to U_j where $j = \arg\max_i d_{U_i}(x)$ until this is no longer possible (where $\arg\max_i d_{U_i}(x)$ denotes the value of i that maximises $d_{U_i}(x)$). This process terminates, as the number of edges crossing the partition is reduced with each step. It is easy to see that this redistribution can be done in time polynomial in n. Call the resulting partition $\mathcal{V} := \{V_1, \dots, V_m\}$, (so $V_i = W_i \cup X_i'$ for some $X_i' \subseteq X$ and $X = \sqcup X_i'$).

We show that $\mathcal{V}$ fulfils (i)-(iv). It is easy to see that (iii) holds by our choice of $\mathcal{V}$ for all $x \in X$. For $x \in W_i$, (3.3.17) implies $d_{V_i}(x) \ge d_{W_i}(x) \ge D - 2\sqrt{\rho} n \ge D/2$, so (iii) holds. Next, since each step of our procedure reduces the number of edges crossing the partition, we have

$$\sum_{1 \le i \le m} e(V_i, \overline{V_i}) \le \sum_{1 \le i \le m} e(U_i, \overline{U_i}) \le \rho m n^2 \le \rho^{1/3} n^2$$

and therefore each V_i is a $\rho^{1/3}$-component, so (ii) holds. We have $|U_i \triangle V_i| \le |X| \le m\sqrt{\rho} n \le \rho^{1/3} n$ for all i, so (i) holds as well. To see (iv), note that for all $x \in W_i$ we have $d_{V_i}(x) \ge D - 2\sqrt{\rho} n$ by (3.3.17) and $|V(G) \setminus \cup_{i=1}^m W_i| = |X| \le \rho^{1/3} n$. □

Lemma 3.3.19. Let $0 < 1/n_0 \ll \rho \ll \nu \ll \tau \ll \alpha < 1$ and let G be a D-regular graph on n vertices where $n \ge n_0$ and $D \ge \alpha n$. Suppose that U

is a bipartite (ρ, ν, τ)-robust expander component of G with bipartition A, B. Then there exists a bipartition A', B' of U such that

(i) U is a bipartite $(3\sqrt{\rho}, \nu/2, 2\tau)$-robust expander component with partition A' ,B';
(ii) $d_{B'}(u) \geq d_{A'}(u)$ for all $u \in A'$, and $d_{A'}(v) \geq d_{B'}(v)$ for all $v \in B'$.

Furthermore, (for fixed $n_0, \rho, \nu, \tau, \alpha$ satisfying the hierarchy above) there is an algorithm that finds such a partition in time polynomial in n.

Proof. This proof is similar to that of Lemma 3.3.18. Let $A_0 := \{x \in A \mid d_{\overline{B}}(x) \geq 2\sqrt{\rho}n\}$ and define B_0 similarly. The fact that U is a ρ-component implies that

$$\begin{aligned}\rho n^2 \geq & e(A, \overline{B}) + e(B, \overline{A}) \geq \frac{1}{2}\left(\sum_{x \in A} d_{\overline{B}}(x) + \sum_{x \in B} d_{\overline{A}}(x)\right) \\ \geq & \frac{1}{2}\left(\sum_{x \in A_0} d_{\overline{B}}(x) + \sum_{x \in B_0} d_{\overline{A}}(x)\right) \geq (|A_0| + |B_0|)\sqrt{\rho}n\end{aligned}$$

and therefore $|A_0| + |B_0| \leq \sqrt{\rho}n$. Define $\hat{A} := A \setminus A_0$ and $\hat{B} := B \setminus B_0$. For all $x \in \hat{A}$ we have $d_{\hat{B}}(x) \geq D - d_{\overline{B}}(x) - |B_0| \geq D - 3\sqrt{\rho}n$ and an analogous statement holds for $x \in \hat{B}$. We iteratively move vertices between A_0 and B_0 as follows: for $x \in A_0$ if $d_A(x) > d_B(x)$ then move x from A_0 to B_0 and for $y \in B_0$ if $d_B(y) > d_A(y)$ then move y from B_0 to A_0 (and update A, B, A_0, B_0 accordingly). Continue this until it is no longer possible. This process terminates, as the number of edges not crossing the partition is reduced at each step. It is easy to see that this redistribution can be done in time polynomial in n. Call the resulting parts A', B'. We show that A', B' fulfil (i) and (ii).

The choice of A', B' implies that all $x \in A_0 \cup B_0$ fulfil (ii). For $x \in \hat{A}$ we have $d_{B'}(x) \geq d_{\hat{B}}(x) \geq D - 3\sqrt{\rho}n \geq d_U(x)/2$. A similar statement holds for all $x \in \hat{B}$, by our choice of vertex redistribution, completing the proof of (ii). For (i), note that $|A \triangle A'| + |B \triangle B'| \leq |A_0| + |B_0| \leq \sqrt{\rho}n$. Now Lemma 3.3.17(i) with $\rho, \sqrt{\rho}, \nu, \tau, A, B, A', B'$ playing the roles of $\rho, \gamma, \nu, \tau, A, B, A', B'$ shows that U is a bipartite $(3\sqrt{\rho}, \nu/2, 2\tau)$-robust expander component with bipartition A', B', which completes the proof of (i). □

Finally, we can prove the existence of a polynomial-time algorithm to find a robust partition in regular graphs. Again, we follow the proof from [54] closely, but must suitably apply the algorithms developed in this section.

Theorem 3.3.20. For every $0 < \tau < \alpha < 1$ and every non-decreasing function $f : (0,1) \to (0,1)$ there is a n_0 and a polynomial-time algorithm that does the following. Given an n-vertex D-regular graph G as input with $n \geq n_0$ and $D \geq \alpha n$, the algorithm finds a robust partition $\mathcal{V}$ with parameters ρ, ν, τ, k, ℓ with $1/n_0 < \rho < \nu < \tau$; $\rho < f(\nu)$, and $1/n_0 < f(\rho)$.

Proof. Set $t = \lceil 2/\alpha \rceil$. Define constants satisfying

$$0 < 1/n_0 \ll \rho_1 \ll \nu_1 \ll \rho_2 \ll \nu_2 \ll \cdots \ll \rho_t \ll \nu_t \ll \tau' \ll \tau \leq \alpha.$$

We start with the following claim:

Claim 3.3.21. There is some $1 \leq h < t$ and a partition $\mathcal{U}$ of $V(G)$ such that, for each $U \in \mathcal{U}$, U is a (ρ_h, ν_h, τ')-robust expander component or a bipartite (ρ_h, ν_h, τ')-robust expander component. Furthermore, we can find $\mathcal{U}$ in polynomial time (and we can determine those $U \in \mathcal{U}$ that are bipartite robust expander components together with a corresponding bipartition).

Proof of Claim 3.3.21. We will iteratively construct (in polynomial time) a partition $\mathcal{U}_i$ of $V(G)$ such that U is a ρ_i-component for all $U \in \mathcal{U}_i$.

We know $V(G)$ is a ρ_1-component for any choice of $\rho_1 > 0$ and we set $\mathcal{U}_1 = \{V(G)\}$.

Assume that for some $1 \leq i \leq t$ we have constructed such a partition $\mathcal{U}_i$ of $V(G)$. We apply Algorithm 4 to each $U \in \mathcal{U}_i$ with $\rho_i, \nu_i, \rho_{i+1}, \tau'$ playing the roles of ρ, ν, ρ', τ. If the algorithm finds some $U \in \mathcal{U}_i$ for which it returns U_1, U_2, a partition of U in which U_1 and U_2 are ρ_{i+1}-components, then we set $\mathcal{U}_{i+1} := (\mathcal{U}_i \setminus \{U\}) \cup \{U_1, U_2\}$ and we continue. Otherwise the algorithm determines that $G[U]$ is a robust (ν_i, τ')-expander or a bipartite robust (ν_i, τ')-expander for all $U \in \mathcal{U}_i$ and so each $U \in \mathcal{U}_i$ is a (ρ_i, ν_i, τ')-robust expander component or a bipartite (ρ_i, ν_i, τ')-robust expander component (and Algorithm 4 is able to determine which $U \in \mathcal{U}_i$ are bipartite robust expander components and to determine a corresponding bipartition A, B of any such U). In this case we are done with the claim provided $i < t$, which we now show.

By induction $|\mathcal{U}_{i+1}| = i + 1$ and all $U \in \mathcal{U}_{i+1}$ are ρ_{i+1}-components whenever $\mathcal{U}_{i+1}$ is defined. To see that the process terminates before $\mathcal{U}_t$, assume for the sake of contradiction that $\mathcal{U}_t$ is defined. Since every $U \in \mathcal{U}_t$ is a ρ_t-component, $|U| \geq (\alpha - \sqrt{\rho_t})n$ for all $U \in \mathcal{U}_t$ by Proposition 3.3.6, and so

$$n = |V(G)| \geq t(\alpha - \sqrt{\rho_t})n \geq \frac{2}{\alpha}(\alpha - \sqrt{\rho_t})n > n,$$

a contradiction, proving the claim. ◀

So in polynomial time, we can find $\mathcal{U} = \{U_1, \dots, U_k, Z_1, \dots, Z_\ell\}$ for some $k, \ell \in \mathbb{N}$, where U_i is a (ρ', ν', τ')-robust expander component for all $1 \leq i \leq k$ and Z_j is a bipartite (ρ', ν', τ')-robust expander component for all $1 \leq j \leq \ell$, where $\rho' = \rho_h, \nu' = \nu_h$ for some $h < t$. Furthermore our algorithm determines which $U \in \mathcal{U}$ are bipartite robust expander components and gives corresponding bipartitions for them.

From Proposition 3.3.6 and Remark 3.3.7 we know that $|U_i| \geq (D - \sqrt{\rho'}n)$ for $1 \leq i \leq k$ and $|Z_j| \geq 2(D - 2\sqrt{\rho'}n)$ for $1 \leq j \leq \ell$. Therefore

$$n = \sum_{1 \leq i \leq k} |U_i| + \sum_{1 \leq j \leq l} |W_j| \geq (D - 2\sqrt{\rho'}n)(k + 2\ell)$$

and so

$$k + 2\ell \leq \left\lfloor \frac{n}{D - 2\sqrt{\rho'}n} \right\rfloor \leq \left\lfloor (1 + \rho'^{1/3}) \frac{n}{D} \right\rfloor. \tag{3.3.18}$$

In particular $m := k + \ell \leq (k + 2\ell) \leq 2n/D \leq 2\alpha^{-1}$. Now we apply the algorithm of Lemma 3.3.18 (with ρ' playing the role of ρ) to $\mathcal{U}$ to obtain (in polynomial time) the partition $\mathcal{V} = \{V_1, \dots, V_k, W_1, \dots, W_\ell\}$ of $V(G)$ satisfying (i)-(iv) so that in particular

$$|U_i \triangle V_i|, |Z_i \triangle W_i| \leq \rho'^{1/3} n \leq \nu' n$$

for all applicable i and j. We now show that $\mathcal{V}$ is a (ρ, ν, τ)-robust partition of G, where $\rho = 3^{3/2}\rho'^{1/6}$, $\nu = \nu'/4$. Note that $\rho \leq f(\nu)$ by making a suitable choice of $\rho_i \ll \nu_i$ for each i at the start. Similarly, a suitable choice of ρ_1 guarantees that $1/n_0 \leq f(\rho)$.

Obviously (D1) holds. For (D2), note that V_i is a $\rho'^{1/3}$-component by Lemma 3.3.18(ii). As $\rho'^{1/3} \leq \rho$ and $|V_i| \geq D/2 \geq \sqrt{\rho}n$ (by Proposition 3.3.6), V_i is a ρ-component. By Lemma 3.3.18(i) and Lemma 3.3.16 with ν', τ', U_i, V_i playing the roles of ν, τ, U, U', we have that $G[V_i]$ is a robust $(\nu'/2, 2\tau')$-expander and thus also a robust (ν, τ)-expander. This shows (D2). To show (D3), recall that $G[Z_j]$ is a bipartite (ρ', ν', τ')-robust expander component and our algorithm gives us a partition A'_j, B'_j of Z_j demonstrating this. We obtain a partition A''_j, B''_j of W_j by taking $A''_j = A'_j \cap W_j$ and $B''_j = W_j \setminus A''_j$ so that $|A''_j \triangle A'_j| + |B''_j \triangle B'_j| \leq |Z_j \triangle W_j| \leq \rho'^{1/3} n$. Then Lemma 3.3.18(ii) together with Lemma 3.3.17(i) where $\rho', \rho'^{1/3}$,

ν', τ', Z_j, W_j play the roles of $\rho, \gamma, \nu, \tau, U, U'$ imply that $G[W_j]$ is a bipartite $(3\rho'^{1/3}, \nu'/2, 2\tau')$-robust expander component. Next we apply (the algorithm of) Lemma 3.3.19 with $(3\rho'^{1/3}, \nu'/2, 2\tau', W_j, A''_j, B''_j)$ playing the roles of $(\rho, \nu, \tau, U, A, B)$ to obtain a bipartition A_j, B_j of W_j (in polynomial time). Now (D3) follows from Lemma 3.3.19(i). We find that (D4) follows from Lemma 3.3.18(iii) and (D5) follows from Lemma 3.3.19(ii). Lastly, (D6) follows from (3.3.18) and (D7) follows from Lemma 3.3.18(iv). □

Remark 3.3.22. The running time of the algorithm of Theorem 3.3.20 is bounded by $O(n^3\alpha^{-2})$ where $n = |V(G)|$. Indeed, examining the proof of Theorem 3.3.20, the algorithm in Claim 3.3.21 makes $O(t^2) = O(\alpha^{-2})$ calls to Algorithm 4. Algorithm 4 makes a single call to each of Algorithms 1,2,3, and each of these algorithms requires at most n applications of either Theorem 3.2.1 or Theorem 3.2.2, i.e. a total running time of $O(\alpha^{-2}) \cdot n \cdot O(n^2) = O(\alpha^{-2}n^3)$. This dominates the running time as the application of the (greedy) algorithms in Lemma 3.3.18 and Lemma 3.3.19 runs in time $O(n^3)$.

3.4 FINDING ALMOST-HAMILTON CYCLES

In this section we show how to determine algorithmically whether a dense, regular graph G has a very long cycle (missing at most a constant number of vertices) and how to construct such a cycle if it exists. The idea is that we first use the algorithm of Theorem 3.3.20 to find a robust partition $\mathcal{U} = \{U_1, \ldots, U_m\}$ of our input dense regular graph. Then we try to find a *path system* $\mathcal{P}$ (defined below) that supplies all the edges of our desired cycle between the U_i.[3] What properties should the edges in such a path system have? For any (almost) Hamilton cycle H of G, the edges of H between the U_i should connect the U_i's in some sense; thus the path system $\mathcal{P}$ should be *connecting*, which we define precisely below. The path system should also be balancing in some sense: if U_i is a bipartite component with parts A_i and B_i then the edges of $H \cap G[A_i, B_i]$ hit an equal number of vertices from A_i and B_i, so the remaining edges of H (namely those of $\mathcal{P}$) should counter any imbalance in the sizes of A_i and B_i. It was established in [54] that G has a Hamilton cycle if and only if there is a connecting, balancing path system (with respect to $\mathcal{U}$); see Lemma 3.4.1 below, which

3 If U_i is a bipartite robust component with bipartition A_i, B_i then $\mathcal{P}$ may contain edges from $G[A_i]$ or $G[B_i]$ but will not contain edges from $G[A_i, B_i]$.

uses robust expansion to connect a connecting, balancing path system into a Hamilton cycle. Furthermore, it was shown in [35] that a balancing path system always exists for dense, regular graphs.

We show how to determine the existence of a connecting path system in polynomial time. We then show it is possible to combine a connecting path system (if it exists) with the (guaranteed) balancing path system to obtain a path system that is connecting and almost balancing. An almost Hamilton cycle exists if and only if such a connecting, almost balancing path system exists.

Note that a dense regular graph may have a connecting path system and a balancing path system, but no connecting and balancing path system, see the example given in Section 3.5. We have been unable to find an efficient algorithm that determines whether a connecting and balancing path system exists.

3.4.1 *Preliminaries*

In this subsection, we recall some definitions and results that will be used later. We begin by defining the structure required between the parts of our robust partition that ensures a Hamilton cycle.

A *path system* $\mathcal{P} = \{P_1, \dots, P_k\}$ in a graph G is a collection of vertex-disjoint paths $P_1, \dots, P_k$ in G. We also think of $\mathcal{P}$ as a subgraph $\mathcal{P} = \cup P_i \subseteq G$, so that $V(\mathcal{P})$ and $E(\mathcal{P})$ make sense.

Reduced graphs - Let G be a graph and $\mathcal{U}$ a partition of $V(G)$. For a path system $\mathcal{P} \subseteq E(G)$ we define the *reduced multigraph* $R_{\mathcal{U}}(\mathcal{P})$ of $\mathcal{P}$ with respect to $\mathcal{U}$ to be the multigraph with vertex set $\mathcal{U}$ and where there is an edge between $U, U' \in \mathcal{U}$ for each path in $\mathcal{P}$ whose endpoints are in U and U'. We also define the *reduced edge multigraph* $R'_{\mathcal{U}}(\mathcal{P})$ of $\mathcal{P}$ with respect to $\mathcal{U}$ as the multigraph with vertex set $\mathcal{U}$ and where there is an edge between $U, U' \in \mathcal{U}$ for each *edge* in $\mathcal{P}$ with endpoints in U, U'. Note that both $R_{\mathcal{U}}(\mathcal{P})$ and $R'_{\mathcal{U}}(\mathcal{P})$ may contain loops and multiedges. We will often identify edges in $R_{\mathcal{U}}(\mathcal{P})$ (resp. $R'_{\mathcal{U}}(\mathcal{P})$) with their corresponding paths (resp. edges) in $\mathcal{P}$. We sometimes write $R(\mathcal{P})$ or $R'(\mathcal{P})$ if $\mathcal{U}$ is clear from the context.

Connecting and balancing path systems - Let G be a graph and $\mathcal{U}$ a partition of $V(G)$. A path system $\mathcal{P} \subseteq G$ is called $\mathcal{U}$*-connecting* if $R_{\mathcal{U}}(\mathcal{P})$ is Eulerian, that is if $R_{\mathcal{U}}(\mathcal{P})$ is connected and all vertices have even degree.

Let $A, B \subseteq V(G)$ be two disjoint sets. We say $\mathcal{P}$ is *r-almost (A,B)-balancing* if

$$\Big|(|A| - e_{\mathcal{P}}(A, \overline{A \cup B}) - 2e_{\mathcal{P}}(A)) - (|B| - e_{\mathcal{P}}(B, \overline{A \cup B}) - 2e_{\mathcal{P}}(B))\Big| \leq r$$

and we say $\mathcal{P}$ is *(A,B)-balancing* if it is 0-almost (A,B)-balancing. The significance of this is that, given any cycle C of G that covers all vertices of $A \cup B$, if we delete from C all edges of $E_G(A,B)$, the resulting path system will be (A,B)-balancing.

For a robust partition $\mathcal{V} = \{V_1, \dots, V_k, W_1, \dots, W_\ell\}$ of G where A_j, B_j is the corresponding bipartition of W_j for $1 \leq j \leq \ell$, we say $\mathcal{P}$ is $\mathcal{V}$-balancing if it is (A_i, B_i)-balancing for $1 \leq i \leq \ell$, and we say $\mathcal{P}$ is r-almost $\mathcal{V}$-balancing if it is r_i-almost (A_i, B_i)-balancing for $1 \leq i \leq \ell$ and $\sum_{i=1}^{\ell} r_i \leq r$. The *$\mathcal{V}$-imbalance* of $\mathcal{P}$ is the smallest r for which $\mathcal{P}$ is r-almost $\mathcal{V}$-balancing. We will omit $\mathcal{V}$ if it is clear from context.

The definitions introduced so far have been for $\mathcal{U}$ a partition of $V(G)$, but they extend in the obvious way when $\mathcal{U}$ is a subpartition of $V(G)$, i.e. where $\mathcal{U}$ consists of disjoint subsets of vertices that do not necessarily cover all of $V(G)$ (and where it is implicitly assumed that $V(\mathcal{P}) \subseteq \cup_{U \in \mathcal{U}} U$).

Lemma 3.4.1 (Lemmas 7.8 and 6.2 in [54])**.** Let $n, k, \ell \in \mathbb{N}_0$ and $0 < 1/n \ll \rho \ll \nu \ll \tau \ll \eta < 1$. Let G be a graph on n vertices and suppose that $\mathcal{V} := \{V_1, \dots, V_k, W_1, \dots, W_\ell\}$ is a weak robust subpartition of G with parameters $\rho, \nu, \tau, \eta, k, \ell$. For each $1 \leq j \leq \ell$, let A_j, B_j be the bipartition of W_j. If $\mathcal{P}$ is a $\mathcal{V}$-connecting, $\mathcal{V}$-balancing path system such that $|V(\mathcal{P}) \cap X| \leq \rho n$ for all $X \in \mathcal{V}$ then there is a cycle C in G that contains every vertex in $\cup_{U \in \mathcal{V}} U$. Furthermore there is a polynomial-time algorithm for constructing such a cycle.

Remark 3.4.2. Lemma 3.4.1 follows directly from Lemmas 7.8 and 6.2 in [54]. We do not state these results because their statements involve extraneous definitions not required for our purposes. Instead we briefly discuss the relevant results informally and how to make them algorithmic.

In this chapter, our definition of $\mathcal{V}$-balancing is different from that used in [54]. Lemma 7.8 from [54] is used to show that a path system $\mathcal{P}$ satisfying the conditions of Lemma 3.4.1 can be used to construct a so-called $\mathcal{V}$-tour, which satisfies their stronger definition of balance. The proof is constructive and easily gives a polynomial-time algorithm for constructing such a $\mathcal{V}$-tour. Lemma 6.2 in [54] then shows how, given a $\mathcal{V}$-tour, one can construct a cycle

C as in Lemma 3.4.1. The proof shows explicitly how to reduce this problem to that of finding a Hamilton cycle in a robust (ν, τ)-expander. While in [54], finding the Hamilton cycle is done by appealing to Theorem 6.7 there, we can do this in polynomial time by appealing to Theorem 5 in [12].

Next we will state the results from [35] that allow one to find balancing path systems in dense regular graphs. Their setup is different from [54], so we now introduce the necessary definitions.

α-sparse and α-far from bipartite - Let G be a graph on n vertices. A *cut* of a set $A \subseteq V(G)$ is a partition X, Y of A, where X and Y are both non-empty. We say that a cut X, Y is *α-sparse* if $e_G(X,Y) \leq \alpha|X||Y|$. We say that a set $A \subseteq V(G)$ is *α-almost-bipartite* if there exists a partition X, Y of A such that $G[A]$ has at most αn^2 edges that are not in $E_G(X,Y)$. Otherwise, we say that A is *α-far-from-bipartite.*

Clustering - Let $c_{min} \in (0,1)$ and let G be a D-regular graph on n vertices with $D \geq c_{min}n$. A *clustering of* G with parameters $\zeta, \delta, \gamma, \beta, \eta$ is a partition $\{A_1, \ldots, A_r\}$ of $V(G)$ into non-empty sets satisfying the following properties:

(a) G has at most ηn^2 edges with ends in different A_i's;
(b) for each $i \in [r]$, the minimum degree of $G[A_i]$ is at least δn;
(c) for each $i \in [r]$, A_i has no ζ-sparse cuts;
(d) for each $i \in [r]$, A_i is either β-almost bipartite or γ-far from bipartite. If A_i is β-almost-bipartite, we also give an appropriate partition X_i, Y_i.

We will always choose the parameters such that $1/n \ll \eta \ll \beta \ll \gamma \ll \zeta \ll \delta$. Theorem 3.4.3 below states that a clustering always has a balancing path system. Here we think of a path system as a subgraph of G.

Theorem 3.4.3 (Lemma 5 in [35])**.** Let $1/n \ll \eta \ll \beta \ll \xi, \gamma \ll \zeta \ll \delta < 1$. Suppose G is an n-vertex, D-regular graph with $D \geq c_{min}n$ and $\mathcal{A} = \{A_1, \ldots, A_r\}$ is a clustering of G with parameters $\zeta, \delta, \gamma, \beta, \eta$, and assume that whenever A_i is β-almost-bipartite the corresponding partition of A_i is X_i, Y_i. Then there exists a path system $H \subseteq G$ with the following properties:

(a) For each $i \in [r]$ such that A_i is β-almost-bipartite, we have

$$2e_H(X_i) - 2e_H(Y_i) + e_H(X_i, \overline{A_i}) - e_H(Y_i, \overline{A_i}) = 2(||A_j| - |B_j||);$$

(b) The number of leaves (i.e. vertices of degree 1) of H in A_i is even for all $1 \leq i \leq r$;
(c) $|V(H)| \leq \xi n$.

Furthermore, there is a randomized algorithm that finds H with probability $p > \frac{3}{4}$ and runs in time polynomial in n.

Remark 3.4.4. Note firstly that (a) says that H is an $\mathcal{A}$-balancing path system. We shall see in Lemma 3.4.5 that a robust partition is a clustering, so this gives us a way of obtaining balancing path systems for robust partitions.

Theorem 3.4.3 is not stated to be algorithmic in [35], but in fact their probabilistic proof essentially gives a (randomized) polynomial-time algorithm. Also, their proof requires that the probability p of success be positive, but the analysis can easily be modified to show a lower bound of e.g. $p > \frac{3}{4}$.

As Theorem 3.4.3 uses the concept of a clustering, we use Lemma 3.4.5 to show that a robust partition is also a clustering. This allows us to apply Theorem 3.4.3 to a robust partition.

Lemma 3.4.5. For every non-decreasing function $f : (0,1) \to (0,1)$ there is a non-decreasing function $f' : (0,1) \to (0,1)$ satisfying $f'(x) < f(x)$ for all $x \in (0,1)$ such that the following holds. For any choice of parameters $\rho, \nu, \tau, \alpha, n, k, \ell$ satisfying $1/n \leq \rho \ll_{f'} \nu \leq \tau \ll_{f'} \alpha$ and $n, k, \ell \in \mathbb{N}$ there exist parameters $\zeta, \delta, \gamma, \beta, \eta$ satisfying $\rho \ll_f \eta \ll_f \beta \ll_f \gamma \ll_f \zeta \ll_f \nu$ and $\tau < \delta < \alpha$ such that if G is an n-vertex D-regular graph with $D \geq \alpha n$ and $\mathcal{V}$ is a robust partition of G with parameters ρ, ν, τ, k, ℓ then $\mathcal{V}$ is also a clustering with parameters $\zeta, \delta, \gamma, \beta, \eta$.

Remark 3.4.6. A proof of the above lemma is provided in the appendix for completeness.

3.4.2 *Path systems and long cycles*

The first lemmas in this subsection, 3.4.7 to 3.4.11 show how to find connecting path systems. The rest of the chapter shows how to combine all the elements. Lemma 3.4.13 allows us to combine balancing and connecting path systems into a single path system that is connecting and almost balancing, and Lemma 3.4.15 allows us to extend this path system into a very long cycle (by applying Lemma 3.4.1). At the end of the section comes the proof of Theorem 3.4.16, which describes the whole algorithm.

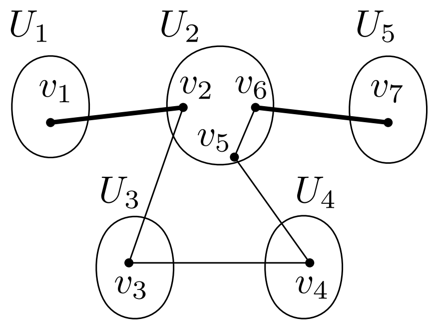


Figure 9: Example: The path $v_1 \dots v_7$ from U_1 to U_5 has the edges between v_2 and v_6 pruned, resulting in two paths (thick lines), one from U_1 to U_2 and one from U_2 to U_5. Note that this ensures that $\mathcal{C}'$ contains no edges inside components.

Lemma 3.4.7. Let G be a graph, let $\mathcal{U} = \{U_1, \dots, U_m\}$ be a partition of $V(G)$, and let $\mathcal{C}$ be a $\mathcal{U}$-connecting path system in G. Then there exists a $\mathcal{U}$-connecting path system $\mathcal{C}'$ such that

(a) $E(\mathcal{C}') \cap E(G[V_i]) = \emptyset$ for all $i = 1, \dots, m$ and
(b) $|E(\mathcal{C}') \cap E_G(V_i, V_j)| \leq 2$ for all $1 \leq i < j \leq m$.

Proof. For any path $P = v_1 v_2 \cdots v_j$ in $\mathcal{C}$, if two vertices of P belong to the same component $U \in \mathcal{U}$, let v_a and v_b be the first and last vertices of P that belong to U and replace P with the paths $v_1 P v_a$ and $v_b P v_j$; it is easy to see that the resulting path system is $\mathcal{U}$-connecting (see Figure 9). We make replacements as described above until no paths contain multiple vertices from the same component and we call the resulting $\mathcal{U}$-connecting path system $\mathcal{C}^*$. Next we show how to reduce the number of edges between components.

Claim 3.4.8. Let $\mathcal{D}$ be a $\mathcal{U}$-connecting path system (i.e. $R_{\mathcal{U}}(\mathcal{D})$ is Eulerian). For $X, Y \in \mathcal{U}$ such that $E_{\mathcal{D}}(X, Y) > 2$, it is possible to find two edges $e, f \in E_{\mathcal{D}}(X, Y)$ such that $\mathcal{D}' = \mathcal{D} \setminus \{e, f\}$ is a $\mathcal{U}$-connecting path system. (Here deleting e, f from $\mathcal{D}$ may create isolated vertices which we remove to form $\mathcal{D} \setminus \{e, f\}$.)

Proof of Claim 3.4.8. We first note that if $e \in E_{\mathcal{D}}(X, Y)$, then the effect of deleting e from $\mathcal{D}$ is to keep all degrees of $R(\mathcal{D})$ unchanged except that the degrees of X and Y will increase or decrease by 1. (Note that we only get a decrease by 1 if e is the first or last edge of a path in $\mathcal{D}$.) Therefore removing two edges of $E_{\mathcal{D}}(X, Y)$ from $\mathcal{D}$ preserves the parity of all vertices of $R(\mathcal{D})$.

Next suppose that $R(\mathcal{D})$ is Eulerian (and hence connected). Hence $R(\mathcal{D})$ is in fact 2-edge connected (since an Eulerian graph can be decomposed into cycles but a cut edge cannot belong to a cycle). Therefore by Menger's theorem there are two edge-disjoint paths Q_1 and Q_2 between X and Y in $R(\mathcal{D})$. Given any three edges of $E_{\mathcal{D}}(X, Y)$, we can find two, say e, f, that miss either Q_1 or Q_2, say Q_1 (where we think of Q_i as a disjoint union of paths in G).

Let P_e be the path of $\mathcal{D}$ containing e. The effect on $R(\mathcal{D})$ of removing e from $\mathcal{D}$ is to replace some edge AB with two edges AX, BY.[4] Therefore A and B are still connected in $R(\mathcal{D} \setminus \{e\})$ via the path AXQ_1YB. Similarly, deleting f keeps the reduced graph connected. Therefore $R(\mathcal{D} \setminus \{e, f\})$ is connected with all degree parities preserved, so is Eulerian, i.e. $\mathcal{D}' = \mathcal{D} \setminus \{e, f\}$ is a connecting path system. ◀

We construct $\mathcal{C}'$ from $\mathcal{C}^*$ by iteratively applying Claim 3.4.8 whenever possible. By construction $\mathcal{C}'$ is a $\mathcal{U}$-connecting path system satisfying (a) and (b). □

Lemma 3.4.9 will be useful in our algorithm for detecting graphs that do not have very long cycles. It essentially says that the absence of a $\mathcal{U}$-connecting path system implies the absence of a very long cycle.

Lemma 3.4.9. Let G be a graph and $\mathcal{U} = \{U_1, \ldots, U_m\}$ be a partition of $V(G)$. If there exists a cycle K in G that contains at least $r > 2m$ vertices from each $U \in \mathcal{U}$, then there also exists a $\mathcal{U}$-connecting path system $\mathcal{C}$ with at most $m^2 - m$ edges. Further, $\mathcal{C}$ contains at most two edges between any two $U_i, U_j \subseteq \mathcal{U}$.

Proof. We start by deleting edges from K to form a path system $\mathcal{C}^*$ such that $R_{\mathcal{U}}(\mathcal{C}^*)$ is a Hamilton cycle on $\mathcal{U}$.

Claim 3.4.10. There exist vertex-disjoint paths $P_1, \ldots, P_m \subseteq K$ such that the endpoints of P_i are in U_i.

Proof of Claim 3.4.10. Suppose, by induction, we have found vertex-disjoint paths $P_1, \ldots, P_{k-1}$ (with $k \leq m$) such that

4 It does not affect what follows, but strictly speaking, if e is the first (resp. last) edge of P_e then AX (resp. BY) is a loop and is not present in $R(\mathcal{D} \setminus \{e\})$.

(a) each P_i (with $i \leq k-1$) has its endpoints in U_i (after relabelling of indices);
(b) $K \setminus (\cup_{i=1}^{k-1} V(P_i))$ is a union of paths that visits U_i at least $r-(k-1) > m$ times for each $i \geq k$.

Any vertex of $\cup_{i=k}^{m} U_i$ is called *untreated*. We know that since K is a cycle, $K \setminus (\cup_{i=1}^{k-1} V(P_i))$ is a disjoint union of $k-1$ paths, which we denote by $Q_1, \dots, Q_{k-1}$. At least one of these paths, say Q_1 must contain at least $(r-k+1)(m-k+1)/(k-1) > m-k+1$ untreated vertices. Pick two untreated vertices $a, b \in V(Q_1)$ that are as close together as possible and belong to the same U_j for some $j \geq k$. In particular, no two internal untreated vertices of aQ_1b belong to the same U_i and so aQ_1b contains at most $m-k+1$ untreated vertices. Then we swap the indices of U_j and U_k and set $P_k = aQ_1b$. It is clear that (a) holds with $k-1$ replaced by k. Since, for each $i \geq k+1$, the path P_k visits each U_i at most once, part (b) also holds. (It is easy to see that a slight variant of the above argument allows us to pick the first path.) ◀

Let $\mathcal{C}^*$ be the set of non-trivial paths of $K \setminus \cup_{i=1}^{m} E(P_i)$; it is easy to see that $\mathcal{C}^*$ is a Hamilton cycle on $\mathcal{U}$ and so is a $\mathcal{U}$-connecting path system. Then, by Lemma 3.4.7 applied to $\mathcal{C}^*$, there exists a $\mathcal{U}$-connecting path system $\mathcal{C}$ that has no edges inside any $U \in \mathcal{U}$ and that has at most 2 edges between any distinct $U_i, U_j \in \mathcal{U}$ (and therefore has at most $m(m-1)$ edges). □

Lemma 3.4.11 gives an algorithm for deciding whether a graph with vertex partition $\mathcal{V}$ has a $\mathcal{V}$-connecting path system.

Lemma 3.4.11. Let G be a graph on n vertices and $\mathcal{V}$ a partition of $V(G)$ with $|\mathcal{V}| = m$. There exists an algorithm that determines whether there exists a $\mathcal{V}$-connecting path system in G, and if one does, then the algorithm finds one with at most $m^2 - m$ edges. This algorithm runs in time $m^{O(m^2)} + O(m^2 n^{5/2})$.

Proof. The algorithm proceeds by first preselecting a small number of plausible edges and then using brute force to find a connecting path system as a subset of these edges. The preselected edges are chosen such that if a $\mathcal{V}$-connecting path system exists, then one exists amongst the preselected edges.

Assume $\mathcal{V} = \{V_1, \dots, V_m\}$. For each $1 \leq i < j \leq m$, let $E_{i,j} \subseteq E_G(V_i, V_j)$ be defined as follows. If the bipartite graph $G[V_i, V_j]$ contains a matching of size $4m$, let $E_{i,j}$ be the edges in any such matching. If not then $G[V_i, V_j]$ has a dominating set $F_{i,j}$ of size at most $8m$ (taking the vertices incident to a maximum matching). For each vertex v in $F_{i,j}$, select any set $E^v_{i,j}$ of $\min(d_{G[V_i,V_j]}(v), 2m)$ edges incident to v in $G[V_i, V_j]$ and take $E_{i,j} = \cup_{v \in F_{i,j}} E^v_{i,j}$. Finally our preselected edge set is defined to be $E' := \cup_{i<j} E_{i,j}$.

Next we show that if a $\mathcal{V}$-connecting path system $\mathcal{C}$ exists, then also a $\mathcal{V}$-connecting path system $\mathcal{D} \subseteq E'$ exists. By Lemma 3.4.7 we may assume that $\mathcal{C}$ has no edges inside any V_i and has at most two edges between each pair V_i, V_j (so in particular there are at most $2(m-1)$ edges of $E(\mathcal{C})$ incident with V_i (and V_j)).

Claim 3.4.12. Let $\mathcal{C}$ be any $\mathcal{V}$-connecting path system as described above, i.e. $\mathcal{C}$ has no edges inside any V_i and has at most two edges between each pair V_i, V_j. Then for any $e \in E(\mathcal{C})$, we can find $r(e) \in E'$ such that

(R1) if e has its endpoints in V_i and V_j, then so does $r(e)$;

(R2) for all $f \in E(\mathcal{C}) \setminus \{e\}$, if $e \cap f = \emptyset$, then $r(e) \cap f = \emptyset$.

We will repeatedly apply this claim to replace edges $e \in \mathcal{C}$ with edges $r(e) \in E'$ to obtain $\mathcal{D}$.

Proof of Claim 3.4.12. In order to find $r(e)$ satisfying (R1) and (R2), assume e has endpoints in V_i and V_j. If $e \in E'$ then set $R(e) = e$ and note that (R1) and (R2) clearly hold. If not, then we have two cases to consider.

If $E_{i,j}$ is a matching of size $4m$ then at least one edge of $E_{i,j}$ is not incident with any edge in $E(\mathcal{C})$ (since there are at most $2(m-1)$ edges of $\mathcal{C}$ incident with any V_i) and this is the edge we choose as $r(e)$; clearly (R1) and (R2) hold in this case.

If $E_{i,j}$ is not a matching of size $4m$, then e is incident to some vertex $v \in F_{i,j}$, so assume $e = vv'$ and that $v \in V_i$ and $v' \in V_j$. Since $e \notin E_{i,j}$, then $E_{i,j}$ has $2m$ edges incident to v, and so there is at least one edge $vv^* \in E_{i,j}$ such that v^* is not incident to any edge in $E(\mathcal{C})$ (again since there are at most $2(m-1)$ edges of $\mathcal{C}$ incident to V_j), and we choose $r(e) = vv^*$. Again (R1) and (R2) follow by construction. ◀

We now apply the above claim to $\mathcal{C}$, replacing each edge $e \in E(\mathcal{C})$ with $r(e)$ one at a time (each time updating $\mathcal{C}$ before the next application of the claim). Denote the resulting set of edges by $\mathcal{D}$. Note that $E(\mathcal{D}) \subseteq E'$ and

(a) if $e \in E(\mathcal{C})$ has its endpoints in V_i and V_j, then so does $r(e) \in \mathcal{D}$;

(b) if $e, f \in \mathcal{C}$ are independent (i.e. $e \cap f = \emptyset$) then so are $r(e)$ and $r(f)$.

Here (b) holds because (R2) guarantees we never introduce any new incidences during the process of replacing edges.

It is easy to see from (b) that $\mathcal{D}$ is a path system, and we now check that $\mathcal{D}$ is $\mathcal{V}$-connecting. By (a) and (b), for any path $P \in \mathcal{C}$, the set of edges $\{r(e) \mid e \in E(P)\}$ is a union of vertex-disjoint paths $P_1, \ldots, P_t$ with $P_i = a_i P_i b_i$ and a_{i+1} and b_i belong to the same $V \in \mathcal{V}$. Therefore each edge $e = VV' \in R(\mathcal{C})$ corresponds to a path from V to V' in $R(\mathcal{D})$ (with edges $e_1, \ldots, e_t$ corresponding to the paths $P_1, \ldots, P_t$). This shows that $R(\mathcal{D})$ can be obtained from $R(\mathcal{C})$ by replacing each edge with a path having the same endpoints as the edge: it is now clear that if $R(\mathcal{C})$ is Eulerian then so is $R(\mathcal{D})$ and so $\mathcal{D}$ is $\mathcal{V}$-connecting.

We have now shown that if a $\mathcal{V}$-connecting path system exists, then one exists inside E' (and we have seen that it uses at most 2 edges between each V_i, V_j, so at most $m^2 - m$ edges in total). For the algorithm to find such a path system, we first construct each $E_{i,j}$; the running time here is dominated in searching for a maximum matching in each $G[V_i, V_j]$, which takes total time $\binom{m}{2} n^{2.5}$ (using e.g. the Hopcroft-Karp algorithm [40]). We then check every possible way of selecting at most two edges from each $E_{i,j}$; since $E_{i,j}$ has size at most $(8m)(2m) = 16m^2$, there are $\left(\binom{16m^2}{2} + 16m^2 + 1\right)^{\binom{m}{2}} = m^{O(m^2)}$ possibilities. If a $\mathcal{V}$-connecting path system exists, then one of these possibilities will give us such a path system and it takes time $m^{O(m^2)} + O(m^2 n^{2.5})$-time to determine this. □

The next lemma allows us to combine a connecting path system with a balancing path system into a path system that is connecting and almost-balancing.

Lemma 3.4.13. Given a graph G on n vertices with a robust partition $\mathcal{V} = \{V_1, \ldots, V_k, W_1, \ldots, W_\ell\}$, a $\mathcal{V}$-balancing path system $\mathcal{B}$ and a $\mathcal{V}$-connecting path system $\mathcal{C}$, there exists a connecting, $(5|E(\mathcal{C})| + m - 1)$-almost balancing path system $\mathcal{P}$, where $m := k + \ell$ is the number of components in $\mathcal{V}$, and $\mathcal{P} \subseteq \mathcal{B} \cup \mathcal{C}$ (when thought of as sets of edges). Furthermore, $\mathcal{P}$ can be constructed in time polynomial in n. (Note that we suppress the parameters of the robust partition as they are irrelevant for this lemma.)

Proof. We begin by constructing $\mathcal{B}' \subseteq \mathcal{B}$ as follows: First delete any edge from $\mathcal{B}$ that shares a vertex with an edge from $\mathcal{C}$ to obtain $\mathcal{B}^*$. As each edge in $\mathcal{C}$ is incident to at most four edges in $\mathcal{B}$, we delete at most $4|E(\mathcal{C})|$ edges here.

Claim 3.4.14. There exists $\mathcal{B}' \subseteq \mathcal{B}^*$ such that $|E(\mathcal{B}^*) \setminus E(\mathcal{B}')| \leq m - 1$ and every vertex of $R_{\mathcal{V}}(\mathcal{B}')$ has even degree.

Proof of Claim 3.4.14. Consider a connected component X of the multigraph $\mathcal{R}'_{\mathcal{V}}(\mathcal{B}^*)$. As in any graph, there are an even number of vertices with odd degree in X. For each component of $\mathcal{R}'_{\mathcal{V}}(\mathcal{B}^*)$, pair up these vertices arbitrarily and find paths (not necessarily disjoint) between each pair within $\mathcal{R}'_{\mathcal{V}}(\mathcal{B}^*)$ (which is possible since each pair belongs to the same connected component of $\mathcal{R}'_{\mathcal{V}}(\mathcal{B}^*)$); call these paths $P_1, \dots, P_t$. Set $Q = \triangle_{i=1}^{t} P_i$ as the symmetric difference of the edge sets of $P_1, \dots, P_t$. Note that removing all edges in Q from $\mathcal{R}'_{\mathcal{V}}(\mathcal{B}^*)$ will result in a graph with even degree in each vertex. Next, construct Q' from Q by iteratively removing edges that form cycles, where we count a double edge as a cycle. Do this until no cycles remain, i.e. Q' is a forest so has at most $m - 1$ edges. Again, removing the edges in Q' from $\mathcal{R}'_{\mathcal{V}}(\mathcal{B}^*)$ results in a graph with even degree in each vertex. The edges in Q' correspond to edges in $\mathcal{B}^*$ that we delete to construct $\mathcal{B}'$, and so $\mathcal{R}'_{\mathcal{V}}(\mathcal{B}')$ has even degree in every vertex. As the parity of each degree in $\mathcal{R}_{\mathcal{V}}(\mathcal{B}')$ and $\mathcal{R}'_{\mathcal{V}}(\mathcal{B}')$ are the same, $\mathcal{R}_{\mathcal{V}}(\mathcal{B}')$ has even degree in each vertex. ◀

We construct $\mathcal{P}$ as the union of $\mathcal{B}'$ and $\mathcal{C}$. Both $R_{\mathcal{V}}(\mathcal{B}')$ and $R_{\mathcal{V}}(\mathcal{C})$ have even degree for every vertex and so this also holds for $R_{\mathcal{V}}(\mathcal{P})$. Since $R_{\mathcal{V}}(\mathcal{C})$ is connected so is $R_{\mathcal{V}}(\mathcal{P})$ and so $\mathcal{R}_{\mathcal{V}}(\mathcal{P})$ is Eulerian, i.e. $\mathcal{P}$ is $\mathcal{V}$-connecting. By construction $\mathcal{P}$ arises from $\mathcal{B}$ by at most $5|E(\mathcal{C})| + m - 1$ additions or deletions of edges, each of which contributes at most 1 to the $\mathcal{V}$-imbalance of $\mathcal{P}$. It is straightforward to see that $\mathcal{P}$ can be constructed in time polynomial in n given $G, \mathcal{B}, \mathcal{C}$. □

If we have a connecting, almost balancing path system (as provided by Lemma 3.4.13) with respect to a robust partition, then we can use Lemma 3.4.1 to construct a very long cycle, as described below.

Lemma 3.4.15. Let $0 < 1/n_0 \ll \rho \leq \gamma \ll \nu \ll \tau \leq \alpha < 1$ and $t \leq \rho n$. There is an algorithm that, given an n-vertex, D-regular graph G with $n \geq n_0$ and $D \geq \alpha n$ and a robust partition $\mathcal{V} = \{V_1, \dots, V_k, W_1, \dots, W_\ell\}$ of G with parameters ρ, ν, τ, k, ℓ and a $\mathcal{V}$-connecting t-almost balancing path system $\mathcal{P}$ with $|V(\mathcal{P}) \cap V| \leq \gamma n$ for all $V \in \mathcal{V}$, constructs a cycle through all but at most t vertices of G. It does this in time polynomial in n.

Proof. We use Lemma 3.2.4 to see that $\mathcal{V}$ is also a weak robust subpartition with parameters $\rho, \nu, \tau, \eta, k, \ell$ where we set $\eta = \alpha^2/2$.

For $1 \leq j \leq \ell$, let t_j be such that $\sum t_j = t$ and such that $\mathcal{P}$ is t_j-almost (A_j, B_j)-balancing, where A_j, B_j is the bipartition corresponding to W_j. By selecting t_j vertices T_j from either $A_j \setminus V(\mathcal{P})$ or $B_j \setminus V(\mathcal{P})$, we can ensure that $\mathcal{P}$ is $(A_j \setminus T_j, B_j \setminus T_j)$-balancing. Set $T = \cup T_j$ so that $|T| = t \leq \rho n$ and define $\mathcal{V}' = \{V'_1, \dots, V'_k, W'_1, \dots, W'_\ell\}$ with $V'_i = V_i \setminus T = V_i$ and $W'_j = W_j \setminus T$ with $A_j \setminus T, B_j \setminus T$ as the bipartition of W'_j.

Next we show that $\mathcal{V}'$ is a weak robust subpartition of G with parameters $3\gamma, \nu/2, 2\tau, \alpha^2/4, k, \ell$.

First we apply Lemma 3.3.17(ii) to each W_j with $W_j \setminus T$ playing the role of U'. As $|W_j \triangle W'_j| \leq \rho n \leq \gamma n$, we see that each W_j is a bipartite $(3\gamma, \nu/2, 2\tau)$-robust expander component of G (with bipartition $A_j \setminus T, B_j \setminus T$ by Lemma 3.3.17(i)). Clearly each $V'_i = V_i$ remains a (ρ, ν, τ)-robust expander component and so is a $(3\gamma, \nu/2, 2\tau)$-robust expander component as well. This shows that (D2′) and (D3′) hold. (D1′) obviously holds, and as $|T| \leq \rho n$, it is easy to see that (D4′) and (D5′) also hold.

To construct the desired cycle (i.e. one that contains every vertex of $V(G) \setminus T$), we apply Lemma 3.4.1 with $G, 3\gamma, \nu/2, 2\tau, \alpha^2/4, n, k, \ell, \mathcal{V}', \mathcal{P}$ playing the roles of $G, \rho, \nu, \tau, \eta, n, k, \ell, \mathcal{V}, \mathcal{P}$. We obtain a cycle C that contains all vertices in $\cup_{X \in \mathcal{V}'} X = V(G) \setminus T$. Moreover, this cycle can be found in time polynomial in n since we can find T in polynomial time and apply Lemma 3.4.1 in polynomial time. □

Finally, we prove the main result, which we repeat here for convenience.

Theorem 3.4.16. For every $\alpha \in (0, 1]$, there exists $c = c(\alpha)$ and a (deterministic) polynomial-time algorithm that, given an n-vertex D-regular graph G with $D \geq \alpha n$ as input, determines whether G contains a cycle on at least $n - c$ vertices. In fact, we can take $c(\alpha) = 100\alpha^{-2}$. Furthermore there is a (randomized) polynomial-time algorithm to find such a cycle if it exists.

Proof. We are given α in the statement of the theorem. We will choose non-decreasing functions $f_1, f_2, f_3, f_4 : (0, 1) \to (0, 1)$ with $f_i(x) \leq x$ for all $x \in (0, 1), i \in [4]$ as follows. Let f_1 be the function governing the hierarchy in the statement of Lemma 3.4.15 and let f_2 be the function governing the hierarchy of Theorem 3.4.3. Define $f_3 : (0,1) \to (0,1)$ as $f_3(x) = \min\{f_1(x), f_2(x), \alpha^2 x^2/100\}$. Applying Lemma 3.4.5 with f_3 playing the role of f, let f_4 be the function we obtain (i.e. $f_4 := f'$) and note that $f_4(x) \leq f_3(x)$ for all $x \in (0,1)$.

We define $\tau = f_4(\alpha)$ and apply Theorem 3.3.20 with τ, α, f_4 playing the roles of τ, α, f to obtain a number $n_0 \in \mathbb{N}$. Define $c := 100\alpha^{-2}$. So far we have defined $f_1, \dots, f_4, \tau, \alpha, n_0, c$.

Given an n-vertex D-regular graph G with $D \geq \alpha n$, then if $n \leq \max(n_0, 1000\alpha^{-3})$ we can use brute force to determine in polynomial time if there exists a cycle in G on at least $n - c$ vertices. So we assume that $n \geq \max(n_0, 1000\alpha^{-3})$.

By applying Theorem 3.3.20 to G (with τ, α, n_0 as above and $f = f_4$), we obtain a robust partition $\mathcal{V}$ of G with parameters ρ, ν, τ, k, ℓ satisfying

$$1/n_0 \ll_{f_4} \rho \ll_{f_4} \nu \leq \tau \ll_{f_4} \alpha. \tag{3.4.1}$$

Set $m := k + \ell = |\mathcal{V}|$ and note that $m \leq (1 + \rho^{1/3})/\alpha \leq 2\alpha^{-1}$.

We claim that G contains a cycle with at least $n - c$ vertices if and only if G has a $\mathcal{V}$-connecting path system. The claim proves the first part of the Theorem because, by applying the algorithm of Lemma 3.4.11, we can determine in time polynomial in n whether G has a $\mathcal{V}$-connecting path system (and if it does, we can find one in time polynomial in n with at most m^2 edges).

So let us prove the claim. First assume G has no $\mathcal{V}$-connecting path system. Then by Lemma 3.4.9, for every cycle K of G, there is some $U \in \mathcal{V}$ such that K contains at most $2m$ vertices of U; in particular K misses at least

$$|U| - 2m \geq (\alpha - \sqrt{\rho})n - 2m \geq (\alpha/2)n - 2m \geq c$$

vertices, where the first inequality is by Proposition 3.3.6, the second since $\rho \ll_{f_4} \alpha$ with $f_4(x) \leq f_3(x) \leq x^2/4$, and the third by our choice of n large and c.

Now suppose G contains a $\mathcal{V}$-connecting path system. Then we know there exists a $\mathcal{V}$-connecting path system $\mathcal{P}$ with at most m^2 edges. By Lemma 3.4.5 with f_3, f_4 playing the roles of f, f' and using (3.4.1), we see that $\mathcal{V}$ is a clustering with parameters $\zeta, \delta, \gamma, \beta, \eta$ where

$$1/n \ll_{f_3} \rho \ll_{f_3} \eta \ll_{f_3} \beta \ll_{f_3} \gamma \ll_{f_3} \zeta \ll_{f_3} \nu \leq \tau \leq \delta \leq \alpha. \tag{3.4.2}$$

Set $\xi := \gamma$. In particular $n, \eta, \beta, \gamma, \xi, \zeta, \delta$ satisfy the hierarchy needed to apply Theorem 3.4.3 to G (with $\mathcal{V}, \alpha$ playing the roles of $\mathcal{A}, c_{\min}$). Thus there exists $H \subseteq G$ that is $\mathcal{V}$-balancing (by part (a)) and such that $|V(H)| \leq \xi n = \gamma n$ (by part (c)). Now applying Lemma 3.4.13 with $G, \mathcal{V}, H, \mathcal{P}$ playing the roles of $G, \mathcal{V}, \mathcal{B}, \mathcal{C}$, there exists a $\mathcal{V}$-connecting, r-almost balancing path system $\mathcal{P}' \subseteq \mathcal{P} \cup H$ where $r \leq 5|E(\mathcal{P})| + m - 1 \leq 5m^2 + m \leq c$ (hence $\mathcal{P}'$ is also c-almost balancing). Note that for each $U \in \mathcal{V}$, we have $|V(\mathcal{P}') \cap U| \leq |V(H) \cap U| + |V(\mathcal{P})| \leq \xi n + 2m^2 \leq 2\xi n$. By Lemma 3.4.15 with $G, \mathcal{V}, \mathcal{P}', \rho, 2\xi, \nu, \tau, \alpha, c$ playing the role of $G, \mathcal{V}, \mathcal{P}, \rho, \gamma, \nu, \tau, \alpha, t$, we see there exists a cycle C in G with at least $n - c$ vertices. We note that the required hierarchy for applying Lemma 3.4.15 follows from (3.4.2) and our choice of f_3 and it is also easy to see that $c \leq \rho n$ (since $1/n \ll_{f_3} \rho$ and our choice of f_3). This proves the claim.

Finally, if our algorithm determines that there exists a cycle in G with at least $n - c$ vertices then there is also a randomized polynomial-time algorithm to construct such a cycle. Indeed repeating the argument above with the corresponding algorithms, in polynomial time we can construct $\mathcal{P}$ (Lemma 3.4.11) and H (Theorem 3.4.3 and Remark 3.4.4) and therefore also $\mathcal{P}'$ (Lemma 3.4.13) and hence also C (Lemma 3.4.15). □

Remark 3.4.17. The algorithm in Theorem 3.4.16 (for determining the existence of the cycle) has a crude running time upper bound of $O(\alpha^{-2}n^3) + O(\alpha^{-4}n^{5/2}) + g(\alpha)$, for some function g. Indeed $O(\alpha^{-2}n^3)$ comes from the application of Theorem 3.3.20 and Lemma 3.4.11. The contribution of $g(\alpha)$ comes from using brute force when $n \leq \max(n_0, 100\alpha^{-3})$ and the application of Lemma 3.4.11.

We do not give an explicit running time for finding the desired cycle (when it exists) because this algorithm is based on other polynomial-time algorithms in the literature where no explicit running time bound was given.

3.5 CONCLUSION

The most obvious question that arises from this work is whether we can take $c = 0$ in Theorem 3.4.16, i.e. whether the Hamilton cycle problem is polynomial-time solvable for dense, regular graphs. Our work shows that to answer this affirmatively, it is enough to give a polynomial-time algorithm to decide whether there exists a path system that is both $\mathcal{V}$-connecting and $\mathcal{V}$-balancing when given a dense regular graph together with a robust partition $\mathcal{V}$.

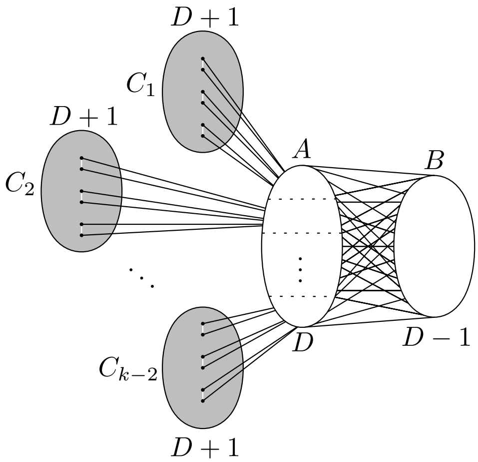


Figure 10: The graph G above has $n = kD + k - 3$ vertices (and we assume k divides D for simplicity). A and B are independent sets with all edges between them present. There are D/k independent edges from A to each C_i so that these edges together form a matching. Then we delete a matching from each C_i so that the resulting graph is D-regular. The graph has no cycle on $n - (k-4)$ vertices because deleting D vertices from G would then yield at most $D + (k-4)$ components in G (at most D from the cycle and at most $k-4$ from the missed vertices), but deleting A from G yields $D + k - 3$ components.

One important aspect of Theorem 3.4.16 is that it shows that the circumference (the length of a longest cycle) of an n-vertex, D-regular graph G with $D \geq \alpha n$ cannot take values between roughly $(1-\alpha)n$ and $n - c$, where $c = c(\alpha) = 100\alpha^{-2}$. For our algorithm, this gives some slack to play with. On the other hand, for the Hamiltonicity problem, there is no such slack: by an easy generalization of the example of Jung [47] and Jackson-Li-Zhu [43] (see Figure 10) there are regular graphs of degree roughly n/k whose circumference is $n - (k-3)$.

If Hamiltonicity turns out to be $\mathcal{NP}$-complete for dense, regular graphs then the question remains as to the smallest value of c for which Theorem 3.4.16 holds. This may turn out to be closely related to the smallest c for which the the circumference cannot take values between roughly $(1-\alpha)n$ and $n - c$. It is also worth noting that the example in Figure 10 has a large independent set (roughly of size αn) and one can in fact show that any non-Hamiltonian dense regular graph with long cycles (say of length at least $(1-(\alpha/2))n$) must have a large independent set (of size at least $(\alpha-\varepsilon)n$).

Finally, we expect that the algorithm given in Theorem 3.4.16 can be modified to give an approximation algorithm for the longest path/cycle problems in dense regular graphs. The idea would be to search for (similarly to Lemma 3.4.11) a connecting path system that maximises the number of vertices in the parts it connects together; write S for this union of parts. We would then combine it with a balancing path system (guaranteed by Theorem 3.4.3) and use the resulting path system together with (a variant of) Lemma 3.4.15 to produce a cycle passing through all but a fixed number c of vertices in S. We should not expect any paths/cycles of length bigger than $|S|$ so this would give a $(1-\frac{c}{n})$-approximation for the longest path/cycle.

APPENDIX

Proof (of equivalence claim in Proposition 3.1.2). For $v \in V(G)$, let $f(v)$ be the 9 vertices in G'' that arise from applying the replacement operation twice, first on v, then on the three vertices that we replace v with; see Figure 11.

First assume G contains a Hamilton cycle C. Note that C can be easily extended to a Hamilton cycle of G'' by tracing a path as in Figure 11, bottom left, through each $f(v)$ for all $v \in V(G)$. The claimed cycles in G'' can be constructed by replacing one, two or three such subpaths with e.g. the path in Figure 11, bottom right.

Now let C' be a cycle of length $|V(G'')|-i$ in G'' with $i \in \{1,2,3\}$. Then C' induces a cycle C on G by contracting $f(v)$ back to a single vertex for all $v \in V(G)$. Clearly C' contains at least one vertex in $f(v)$ for all $v \in V(G)$, so C is a Hamilton cycle of G. □

Proof (of Lemma 3.4.5). We define $f^*, f' : (0,1) \to (0,1)$ as $f^*(x) = \min\{x^2/4, f(x)\}$, and $f'(x) = f_5^*(x)$, where $f_5^*(x)$ denotes composing f^* with itself five times. Note that $f^*(x) < x$ and $f^*(x) \leq f(x)$ for all $x \in (0,1)$, so (by induction) $f_5^*(x) < f(x)$ for all $x \in (0,1)$.

We choose $\zeta, \delta, \gamma, \beta, \eta$ such that $\delta = f^*(\alpha), \zeta = f^*(\nu), \gamma = f^*(\zeta), \beta = f^*(\gamma), \eta = f^*(\beta)$. Note that this also implies $\tau \leq f^*(\delta)$ and $\rho \leq f^*(\eta)$. Writing $x \ll_{f^*} y$ to mean that $x \leq f^*(y)$, one easily checks that

$$\rho \ll_{f^*} \eta \ll_{f^*} \beta \ll_{f^*} \gamma \ll_{f^*} \zeta \ll_{f^*} \nu \leq \tau \ll_{f^*} \delta \ll_{f^*} \alpha.$$

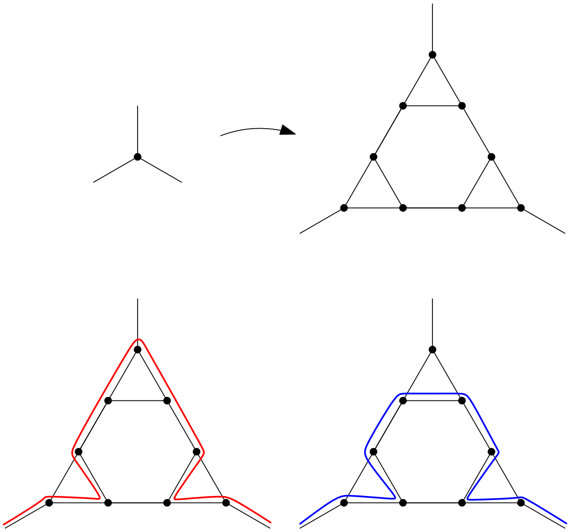


Figure 11: Top: replacing a vertex v in G with $f(v)$ in G''. Bottom left: a (red) path through all vertices of $f(v)$. Bottom right: a (blue) path through 8 vertices of $f(v)$.

Furthermore (D6) implies $m := k + \ell \leq 2n/D \leq 2\alpha^{-1}$ and so $D/m \geq \alpha n/m \geq \alpha^2 n/2 \geq \delta n$.

Property (a) follows from (D2), (D3) and $\rho \ll \eta$. Property (b) follows from (D4) and $\alpha/m \geq 2\alpha^2 \geq \delta$.

For property (c), let X, Y be a non-trivial partition of A_i. We will show $e_G(X, Y) > \zeta|X||Y|$.

First we consider the case that A_i is a robust expander component. Assume without loss of generality that $|X| \leq |Y|$. If $|X| < \tau|A_i|$, each vertex in $|X|$ sends at least $D/m - |X|$ edges to $|Y|$ by (D4). Then $\frac{D}{m} - |X| \geq \delta n - \tau|A_i| \geq \zeta n \geq \zeta|Y|$, so $e_G(X, Y) \geq \zeta|X||Y|$. If $|X| \geq \tau|A_i|$, then since $|X| \leq |Y|$, we have $|X| \leq |A_i|/2 \leq (1-\tau)|A_i|$. Therefore $|\mathrm{RN}_{\nu,A_i}(X)| \geq |X| + \nu|A_i|$, so $|\mathrm{RN}_{\nu,A_i}(X) \cap Y| \geq \nu|A_i|$, and so $e_G(X, Y) \geq \nu^2|A_i|^2 \geq \zeta|X||Y|$.

Now consider the case that A_i is a bipartite robust expander component with parts U_1, U_2. Let X be such that $|X \cap U_1| \leq |Y \cap U_1|$, so we also have $|X \cap U_1| \leq |U_1|/2$.

If $|X \cap U_1| < \tau|U_1|$ and $|X \cap U_2| < \tau|U_1|$, we have

$$\begin{aligned} e_G(X \cap U_1, Y \cap U_2) \geq & |X \cap U_1|(D/2m - |X \cap U_2|) \\ & \geq |X \cap U_1|(\delta n/2 - \tau|U_1|) \geq \zeta n|X \cap U_1|. \end{aligned}$$

By the same argument $e_G(Y \cap U_1, X \cap U_2) \geq \zeta n|X \cap U_2|$, and taking the sum of these inequalities gives $e_G(X, Y) \geq \zeta|X||Y|$.

If $|X \cap U_1| < \tau|U_1|$ and $|X \cap U_2| \geq \tau|U_1|$, we have $e_G(Y \cap U_1, X \cap U_2) \geq (D/2m - |X \cap U_1|)|X \cap U_2| \geq 2\zeta n|X \cap U_2| \geq \zeta n|X| \geq \zeta|X||Y|$.

If $|X \cap U_1| \geq \tau|U_1|$, then since $|X \cap U_1| \leq |Y \cap U_1|$, we have that

$$\tau|U_1| \leq |X \cap U_1|, |Y \cap U_1| \leq (1 - \tau)|U_1|.$$

Therefore (dropping subscripts in RN),

$$\begin{aligned} |\mathrm{RN}(X \cap U_1) \cap U_2| + |\mathrm{RN}(Y \cap U_1) \cap U_2| &\geq |U_1| + 2\nu|A_i| \\ &\geq |U_2| + 2\nu|A_i| - \rho n \\ &\geq |U_2| + \nu|A_i|, \end{aligned} \tag{3.5.1}$$

using Proposition 3.3.6(i) and $\rho \ll \nu$ for the last inequality. This implies that $|\mathrm{RN}(X \cap U_1) \cap (Y \cap U_2)| > \nu|A_i|/2$ or $|\mathrm{RN}(Y \cap U_1) \cap (X \cap U_2)| > \nu|A_i|/2$ since if both fail then we have

$$\begin{aligned} &|\mathrm{RN}(X \cap U_1) \cap U_2| < (\nu/2)|A_i| + |X \cap U_2| \text{ and} \\ &\qquad\qquad |\mathrm{RN}(Y \cap U_1) \cap U_2| < (\nu/2)|A_i| + |Y \cap U_2|, \end{aligned}$$

which when summed contradict (3.5.1). Without loss of generality, we assume $|\mathrm{RN}(X \cap U_1) \cap (Y \cap U_2)| > (\nu/2)|A_i|$, so that $e_G(X, Y) \geq e_G(X \cap U_1, Y \cap U_2) \geq \nu^2|A_i|^2/4 \geq \zeta|X||Y|$.

For property (d), if A_i is a bipartite robust expander component with bipartition U_1, U_2 then the number of non-U_1-U_2 edges is at most $e_G(U_1, \overline{U_2}) + e_G(U_2, \overline{U_1}) \leq \rho n^2 \leq \beta n^2$, showing that A_i is β-almost-bipartite with partition U_1, U_2. If instead A_i is a robust expander component, we claim that A_i is γ-far from bipartite. Let X, Y be a non-trivial partition with $|X| \leq |Y|$, so $|X| \leq |A_i|/2 \leq (1 - \tau)|A_i|$. If $|X| < \tau|A_i|$, then $e_G(X, Y) \leq |X|D$, so

$$\begin{aligned} e(X) + e(Y) \geq (D/2m)|A_i| - D|X| \geq \alpha n|A_i|((2m)^{-1} - \tau) \geq &\ (\alpha^3/16)n^2 \\ &\geq \gamma|X||Y|, \end{aligned}$$

where the penultimate inequality follows since $|A_i| \geq \alpha n/2$ by (D3) and Remark 3.3.7, and $m \leq k + 2\ell \leq 2\alpha^{-1}$ by (D6). If $|X| \geq \tau|A_i|$, then recalling $|X| \leq (1-\tau)|A_i|$, we also have $\tau|A_i| \leq |Y| \leq (1-\tau)|A_i|$ so $\mathrm{RN}_{\nu,A_i}(Y) \geq |Y| + \nu|A_i|$. Therefore, since $|Y| \geq |A_i|/2$, we have $|\mathrm{RN}(Y) \cap Y| \geq |Y| + \nu|A_i|$, so $e(Y) \geq \nu^2|A_i|^2/2 \geq \gamma|X||Y|$. □

4 RECONFIGURATION OF HAMILTON CYCLES AND APPLICATIONS

This chapter is divided into two sections. The first section is devoted to the reconfiguration of Hamilton cycles under k-switches, and in the second section we discuss an application of the results of the first section to computational counting and sampling.

4.1 RECONFIGURATION OF HAMILTON CYCLES UNDER k-SWITCHES

In this section we present results on reconfiguration of Hamilton cycles. We begin by introducing the problem and stating our results and some general context (Subsection 4.1.1). Then some additional definitions (Subsection 4.1.2) and the proofs of the main results (Subsections 4.1.3 and 4.1.4) are presented.

4.1.1 *Introduction*

Throughout this section, let G be an n-vertex graph and denote its minimum degree by $\delta(G)$. Recall that a Hamilton cycle of G is a simple cycle of G that includes every vertex. Given a graph G, let $\mathcal{H}_G$ denote the set of Hamilton cycles of G. We say that $H' \in \mathcal{H}_G$ can be obtained from $H \in \mathcal{H}_G$ by a k-switch if $|E(H) \triangle E(H')| \leq 2k$, that is, a k-switch is an operation for transforming one Hamilton cycle into another by altering at most $2k$ of its edges.[1] Note that the k-switch operation is symmetric, meaning that if H can be obtained from H' by a k-switch, then H' can be obtained from H by a k-switch. See Figure 12 for an example. In this section, we consider reconfigurations of the set of Hamilton cycles of a graph under k-switches.

Given a graph G, we say $\mathcal{H}_G$ is *k-switch irreducible* if for every $H, H' \in \mathcal{H}_G$, we can obtain H' from H by a sequence of k-switches, i.e. there exists

1 Such operations are also widely used, for example, in heuristics for the traveling salesman problem; see e.g., [61].

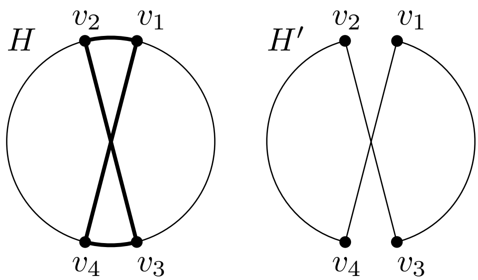


Figure 12: Example of a k-switch for $k = 2$. Left side: The Hamilton cycle H is the circle. We have $E(H) \triangle E(H') = \{v_1v_2, v_2v_3, v_3v_4, v_4v_1\}$ these edges are shown thick. Right side: The modified graph H', which is also a Hamilton cycle.

a sequence $H = H_1, \ldots, H_q = H' \in \mathcal{H}_G$ where H_i can be obtained from H_{i-1} by a k-switch for $i = 2, \ldots, q$. Here we employ the language of Markov chains for later convenience; in the language of reconfiguration problems, the notion of k-switch irreducibility might be referred to as the connectivity of the reconfiguration graph of $\mathcal{H}_G$ under k-switches as introduced in Chapter 1.

Our contributions The main goal of Section 4.1 is to provide the first k-switch irreducibility results for $\mathcal{H}_G$. Our results are as follows.

(i) We prove that $\mathcal{H}_G$ is 10-switch irreducible if $\delta(G) \geq \frac{1}{2}n + 7$. (See Theorem 4.1.1.)

(ii) For each $k \geq 4$, we give examples of graphs G satisfying $\delta(G) \geq \frac{n-3k-4}{2}$ for which $\mathcal{H}_G$ is not k-switch irreducible. (See Example 4.1.5.)

(iii) We give examples of graphs G with $\delta(G) \geq \frac{2}{3}n - 1$ for which $\mathcal{H}_G$ is not 2-switch irreducible. (See Example 4.1.6.)

Theorem 1.2.1 guarantees the existence of Hamilton cycles in graphs G whenever $\delta(G) \geq n/2$. Item (i) shows that very slightly above this threshold, we obtain 10-switch irreducibility, allowing us to move throughout $\mathcal{H}_G$ with small local operations. Item (ii) shows that very slightly below the threshold for Hamiltonicity, there are examples of graphs with Hamilton cycles, but where k-switch irreducibility is lost for $k \geq 4$. So (i) and (ii) essentially establish a threshold in terms of minimum degree for k-switch irreducibility for $k \geq 10$. The degree threshold in 1.2.1 and (i) and (ii) of $n/2$ appears because it allows us to find a set of neighbors with useful properties for any two vertices. It is perhaps surprising that one can lose 2-switch irreducibility quite far above the threshold for Hamiltonicity, as shown in item (iii). We will see later that because of item (iii) certain Markov chains cannot be used to sample Hamilton cycles.

Related work For general background on reconfiguration problems, we refer the reader to the surveys of van den Heuvel [38] and Nishimura [65]. For Hamilton cycles, Takaoka [71] has considered the complexity of deciding whether $\mathcal{H}_G$ is 2-switch irreducible when G belongs to particular structural graph classes. This includes a hardness result for chordal bipartite graphs, but also a result establishing the 2-switch irreducibility of Hamilton cycles in unit interval graphs and monotone graphs.[2] A slightly different Hamilton reconfiguration problem is considered by Lignos [60].

4.1.2 *Preliminaries*

We begin by recalling some common definitions. Let G be a simple undirected graph with vertex set V. We use the shorthand notation uv to denote an edge $\{u, v\} \in E$. Given two graphs $G = (V, E)$ and $G' = (V, E')$ on the same vertex set V, their symmetric difference is denoted by $G \triangle G' = (V, E \triangle E') = (V, (E \setminus E') \cup (E' \setminus E))$. We often write $|G \triangle G'|$ in place of $|E(G) \triangle E(G')|$. We use $N_G(v) = \{w \mid vw \in E\}$ to denote the set of neighbors of $v \in V$ in G and we write $d_G(v) = |N(v)|$ for the degree of v, dropping subscripts when the graph is clear. A 2-factor of G is a subgraph F in which every vertex $v \in V$ has degree precisely $d_F(v) = 2$. We use $\mathcal{F}_G$ to denote the set of all 2-factors of G. We use $\mathcal{H}_G$ to denote the set of all Hamilton cycles of G.

We have defined k-switches for Hamilton cycles, but let us define them more generally. For a given $k \geq 2$ and (finite) set $\mathcal{A}$ of graphs on some vertex set V, we say $F' = (V, E') \in \mathcal{A}$ is obtained from $F = (V, E) \in \mathcal{A}$ by a switch of size k if $|E(F) \triangle E(F')| = 2k$. We say that such a switch of size k (with respect to $\mathcal{A}$) transforms F into F'. We then define a k-switch

2 Chordal bipartite graphs are bipartite graphs in which every cycle on at least 6 vertices contains a chord. Unit interval graphs have as vertices some unit intervals of the real line, with any overlapping vertices connected by an edge. They do not appear in the rest of this thesis. Monotone graphs will be defined below.

(with respect to $\mathcal{A}$) to be a switch of size at most k. In this work we are mostly interested in $\mathcal{A} = \mathcal{H}_G$ or $\mathcal{A} = \mathcal{F}_G$ for a given undirected graph G.

Fix a graph G and consider a k-switch transforming F into F' with respect to $\mathcal{F}_G$ or $\mathcal{H}_G$. It is easy to see that every vertex of the graph $S = F \triangle F'$ must have even degree (since all graphs in $\mathcal{F}_G$ or $\mathcal{H}_G$ are regular of degree 2). Moreover, every connected component of S can be thought of as an alternating circuit, i.e. a circuit whose edges alternate between edges in $E \setminus E'$ and $E' \setminus E$. Recall that a circuit in $G = (V, E)$ is a sequence of $v_1 e_1 v_2 e_2 \cdots v_{k-1} e_{k-1} v_k$ of vertices and edges where $e_i = v_i v_{i+1} \in E$, the edges e_i are distinct, and $v_1 = v_k$.[3]

k-switch irreducibility. For a given graph G and integer k, we say that $\mathcal{H}_G$ is (weakly) k-switch irreducible if for every $H, H' \in \mathcal{H}_G$, there exists a sequence $H = H_1, \ldots, H_q = H'$ of Hamilton cycles in $\mathcal{H}_G$ such that for every consecutive pair of Hamilton cycles (H_i, H_{i+1}), the Hamilton cycle H_{i+1} can be obtained from H_i by a k-switch. Moreover, for a given *class* of graphs $\mathcal{G}$ and integer k, we say that $\mathcal{G}$ is *strongly k-switch irreducible for Hamilton cycles* if there exists a function $\phi : \mathbb{N} \to \mathbb{N}$ with the following property: For all $G \in \mathcal{G}$, whenever $H, H' \in \mathcal{H}_G$ with $|E(H) \triangle E(H')| \leq x$, there exists a sequence of Hamilton cycles $H = H_1, \ldots, H_q = H'$ in $\mathcal{H}_G$ such that for every consecutive pair of Hamilton cycles (H_i, H_{i+1}), the Hamilton cycle H_{i+1} can be obtained from H_i by a k-switch and $q \leq \phi(x)$.

Roughly speaking, strong irreducibility states that if two Hamilton cycles are somewhat 'close' to each other in terms of symmetric difference, then we should be able to transform one into the other with a 'small' number of k-switches. We note that the notion of strong irreducibility will be important in Section 4.2.

Similarly we define (strong) irreducibility for 2-factors. For a given graph G, we say that $\mathcal{F}_G$ (the set of 2-factors of G) is (weakly) k-switch irreducible if for every $F, F' \in \mathcal{F}_G$, there exists a sequence $F = F_1, \ldots, F_q = F'$ of 2-factors in $\mathcal{F}_G$ such that for every consecutive pair of 2-factors (F_i, F_{i+1}), the 2-factor F_{i+1} can be obtained from F_i by a k-switch. For a given *class* of graphs $\mathcal{G}$ and integer k, we say that $\mathcal{G}$ is *strongly k-switch irreducible for 2-factors* if there exists a function $\phi : \mathbb{N} \to \mathbb{N}$ with the following property: For all $G \in \mathcal{G}$, whenever $F, F' \in \mathcal{F}_G$ with $|E(F) \triangle E(F')| \leq x$, there exists a sequence $F = F_1, \ldots, F_q = F'$ of 2-factors in $\mathcal{F}_G$ such that for every

3 The key difference between circuits and cycles is that circuits may repeat vertices and cycles may not.

consecutive pair of 2-factors (F_i, F_{i+1}), the 2-factor F_{i+1} can be obtained from F_i by a k-switch and $q \leq \phi(x)$.

4.1.3 *Strong 10-switch irreducibility*

In this section we prove various results regarding the (non)-irreducibility of the k-switch irreducibility of $\mathcal{H}_G$. The main result of this section is Theorem 4.1.1 below. Afterwards, we provide various examples of non-irreducibility for certain combinations of $\delta(G)$ and k.

Theorem 4.1.1. If a graph G satisfies $\delta(G) \geq \frac{1}{2}n + 7$, then the set $\mathcal{H}_G$ of all Hamilton cycles of G is 10-switch irreducible. Moreover, the class of graphs G for which $\delta(G) \geq \frac{1}{2}n + 7$ is strongly 10-switch irreducible for Hamilton cycles.

Remark 4.1.2 (Bipartite case)*.* Theorem 4.1.1 remains true if we restrict ourselves to balanced bipartite graphs $G = (A \cup B, E)$ on $2n$ vertices, where $|A| = |B| = n$, and $\delta(G) \geq \frac{1}{2}n + 7$. The proofs are almost identical, so we make remarks in footnotes where the proofs differ.

In order to prove Theorem 4.1.1, we rely on Lemma 4.1.3 below. It allows us to quickly reconfigure a 2-factor T into a Hamilton cycle H' without increasing the symmetric difference with respect to some fixed Hamilton cycle H.

Lemma 4.1.3 (Reconnecting lemma)**.** Let $G = (V, E)$ be an undirected graph with minimum degree $\delta(G) \geq \frac{1}{2}n + 1$, and let H be a fixed Hamilton cycle in G. Let T be an arbitrary 2-factor of G with t components.

Then there exists a Hamilton cycle H', so that T can be transformed into H' with at most $t - 1$ switches of size at most 3, and for which

$$|H' \triangle H| \leq |T \triangle H|. \tag{4.1.1}$$

Proof. Let t be the number of components of T. We will prove the statement in the lemma using induction. If $t = 1$ then T is a Hamilton cycle and we are done as we may take $H' = T$. Suppose $t > 1$. Let $C_1, \ldots, C_t$ denote the cyclic components of T. Since H is a Hamilton cycle, there must be some edge $vw \in E(H)$ connecting two components of T (see Figure 13). We assume without loss of generality that vw connects C_1 and C_2, i.e that

$v \in V(C_1)$ and $w \in V(C_2)$ (by renumbering if necessary). Moreover, since v has degree two in H and $vw \in E(H)$, it must be that there exists an $a \in V(C_1)$ (one of the two neighbors of v in T) so that $va \in E(T)$, but $va \notin E(H)$. Similarly, there is a $b \in V(C_2)$ so that $wb \in E(T)$, but $wb \notin E(H)$.

We assign orientations to $C_1, \ldots, C_t$. For any vertex u the vertex following u in the appropriate orientation will be called u^+ and the preceding vertex will be called u^-. We choose the orientations on C_1 and C_2 such that $v = a^+$ and $b = w^+$, see Figure 13, and we assign arbitrary orientations on

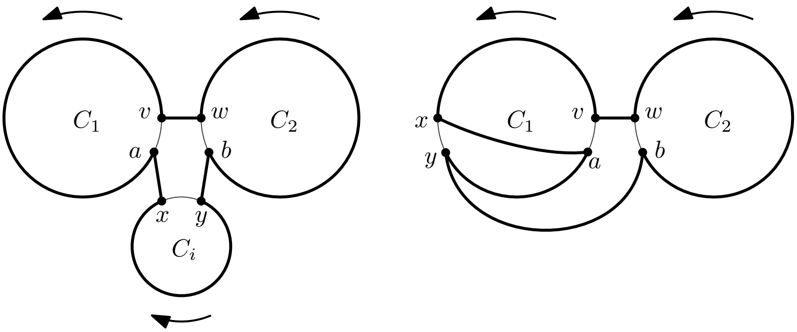


Figure 13: Two situations in the general case. The thick black line shows the cycle after the switch, the arrows show our chosen orientations. Left: xy is on a third cycle. Right: xy is on C_1.

$C_3, \ldots, C_t$. Consider $X := \{u^+ \mid u \in N(a)\}$. As $\delta(G) \geq \frac{1}{2}n + 1$, we have that $|X| \geq \frac{1}{2}n + 1$. Also consider $N(b)$, and note that we have $|N(b)| \geq \frac{1}{2}n + 1$. Therefore $|X \cap N(b)| \neq \emptyset$. Select $y \in X \cap N(b)$ and set $x = y^-$ noting that $ax \in E(G)$.[4] If $y \notin \{a, b^+, w, v^+\}$, the general case, we now switch along the cycle $vaxybwv$; see Figure 13. Note that the edge xy may lie on C_1, C_2 or a different cycle C_i. In all these cases, we do not increase $|T \triangle H|$, as $vw \in E(H)$ and $va, bw \notin E(H)$. If $xy \notin E(C_1 \cup C_2)$, we decrease the number of cycles by two, otherwise by one. For the special cases $y \in \{a, b^+, w, v^+\}$, we switch along different cycles as follows; see Figure 14. If $y = v^+$, we switch along the cycle $vybwv$. If $y = w$, we switch along the cycle $vaxwv$. If $y \in \{a, b^+\}$, then $ab \in E(G)$, and we switch along the cycle

4 In the case of bipartite graphs (see Remark 4.1.2), we note that $avwb$ is a path of G so a and b are in different parts, say $a \in A$ and $b \in B$. Then $X \subseteq A$ with $|X| \geq \frac{1}{2}n + 1$ and $N(b) \subseteq A$ with $|N(b)| \geq \frac{1}{2}n + 1$, so $X \cap N(b) \neq \emptyset$ and we continue.

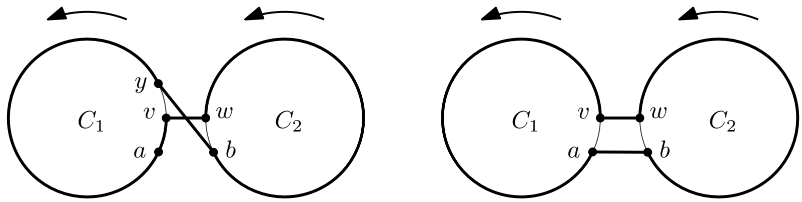


Figure 14: Two situations from the special cases. Left: Case $y = v^+$, Right: Cases $y = a$ and $y = b^+$

$vabwv$. It is easy to see that in these cases we decrease $|T \triangle H|$ by at least two and we decrease the number of cycles by one.

In any case, the resulting 2-factor has fewer components and the symmetric difference is not larger. Repeated application of this procedure proves the statement of the lemma. □

We now continue with the proof of Theorem 4.1.1.

Proof of Theorem 4.1.1. We claim that for two given Hamilton cycles H_1 and H_2 there is a switch of size at most 4 that transforms H_1 into a 2-factor T with at most 3 components such that $|T \triangle H_2| < |H_1 \triangle H_2|$. The theorem then follows from Lemma 4.1.3 since with two switches of size at most 3, we can transform T into some Hamilton cycle H' satisfying

$$|H' \triangle H_2| \leq |T \triangle H_2| < |H_1 \triangle H_2|.$$

In particular we can transform H_1 to H' with a switch of size at most $4 + 2 \times 3 = 10$, and repeating this we can transform H_1 into H_2 with at most $x = |H_1 \triangle H_2|$ switches of size 10, proving the theorem (where we take $\phi(x) = x$ in the definition of strong irreducibility).

We now prove the claim. Note that the symmetric difference of H_1 and H_2 is the vertex-disjoint union of circuits in which edges alternate between H_1 and H_2 and the circuits visit each vertex zero, one, or two times. If the symmetric difference of H_1 and H_2 contains such alternating circuits with four or six edges (corresponding to switches of size 2 or 3), the claim obviously holds, so assume otherwise. In this case it is not hard to see that we can find an H_1, H_2-alternating walk $P = a_1a_2a_3a_4a_5a_6$ (here the a_i are vertices and a_1 and a_6 are distinct) such that the a_1a_2, a_3a_4, a_5a_6 are edges of H_1, and a_2a_3, a_4a_5 are edges of H_2.

We try to find vertices b and c that are neighbors on H_1 such that $b \in N(a_1)$ and $c \in N(a_6)$. Then the circuit $C := a_1a_2a_3a_4a_5a_6cba_1$ is a 4-switch for H_1. Deleting the edges a_1a_2, a_3a_4, a_5a_6 and cb divides H_1 into four paths and adding a_2a_3, a_4a_5, a_6c and ba_1 can connect some of these paths again.

Therefore, switching H_1 along C can produce at most 4 connected components, and this only happens if the four edges a_2a_3, a_4a_5, a_6c and ba_1 connect each path into a cycle (see Figure 15, left side). If one of the paths is just an isolated vertex, it cannot be connected to itself in this way. It is easy to check that 4 components are produced if and only if the vertices $a_1, a_2, \ldots, a_6, c, b$ are distinct and appear in that order along H_1 (as in Figure 15, left side). To prevent this, we choose b and c as follows: orient H_1 so

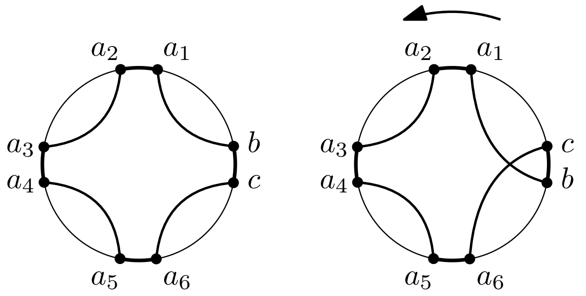


Figure 15: Left side: The circle is H_1. The only way that a 4-switch (thick lines) leads to four components is the shown configuration. Right side: Choosing c to follow b leads to at most three components. Note that in general the edges a_3a_4 and a_5a_6 could appear in different places and orientations.

that a_2 follows a_1. We call the vertex following a vertex v in this orientation v^+ and the previous vertex v^-. Set $M = \{v^- \mid v \in N(a_6)\}$ and consider $N(a_1) \cap M$. As both $|N(a_1)|, |M| \geq n/2 + 7$ we have $|N(a_1) \cap M| \geq 2 \cdot 7 = 14$.[5] Select $b \in (N(a_1) \cap M) \setminus \{a_i^+, a_i^-, i = 1, \ldots, 6\}$ and set $c = b^+$. This ensures that the resulting 4-switch (along the circuit $C := a_1a_2a_3a_4a_5a_6cba_1$) produces at most three components.

Finally, if T is the 2-factor produced by switching H_1 along C, then compared to H_1, the 2-factor T contains at least two new edges of H_2 (namely a_2a_3, a_4a_5) but T may have lost one edge of H_2 (namely bc if it

5 In the case of bipartite graphs (see Remark 4.1.2), we note that $a_1a_2a_3a_4a_5a_6$ is a walk in G and so a_1 and a_6 are in different parts; say $a_1 \in A$ and $a_6 \in B$. Then $N(a_1), M \subseteq B$, so since $|N(a_1)|, |M| \geq \frac{1}{2}n + 7$, so $|N(a_1) \cap M| \geq 14$, and we continue as before.

was in fact an edge of H_2), giving a net gain of one. Since T and H_1 have the same number of edges, we see that $|T \triangle H_2| \leq |H_1 \triangle H_2| - 1$, as required. $\square$

We also give a version of Theorem 4.1.1 for 2-factors, instead of Hamilton cycles, that we will need later. The proof is a simplification of Theorem 4.1.1 and we give it for completeness.

Proposition 4.1.4. The class of graphs G for which $\delta(G) \geq \frac{1}{2}n + 7$ is strongly 4-switch irreducible for 2-factors.

For bipartite graphs the following holds. The class of bipartite graphs $G = (A \cup B, E)$ with bipartition $A \cup B$, where $|A| = |B| = n$, and $\delta(G) \geq \frac{1}{2}n + 7$ is strongly k-switch irreducible for 2-factors.

Proof. We claim that given $F_1, F_2, \in \mathcal{F}_G$, there is a $T \in \mathcal{F}_G$ that can be obtained from F_1 by a 4-switch such that $|T \triangle F_2| < |F_1 \triangle F_2|$. Applying this repeatedly proves the proposition, taking $\phi(k) = k$.

Let $F_1, F_2 \in \mathcal{F}_G$. Note that the symmetric difference of F_1 and F_2 is the vertex-disjoint union of circuits in which edges alternate between F_1 and F_2 and the circuits visit each vertex zero, one, or two times. If the symmetric difference of F_1 and F_2 contains such alternating circuits with four or six edges (corresponding to switches of size 2 or 3), then switching along such a circuit reduces the symmetric difference, so assume otherwise.

In this case it is not hard to see that we can find an H_1, H_2-alternating walk $P = a_1a_2a_3a_4a_5a_6$ (here the a_i are vertices and a_1 and a_6 are distinct) such that a_1a_2, a_3a_4, a_5a_6 are edges of F_1, and a_2a_3, a_4a_5 are edges of F_2.

We try to find vertices b and c that are neighbors on F_1 such that $b \in N(a_1)$ and $c \in N(a_6)$. Then the circuit $C := a_1a_2a_3a_4a_5a_6cba_1$ is a 4-switch for F_1. We choose b and c as follows. Orient the cycles of F_1 arbitrarily. We call the vertex following a vertex v in this orientation v^+ and the previous vertex v^-. Set $M = \{v^+ \mid v \in N(a_6)\}$ and consider $N(a_1) \cap M$. As both $|N(a_1)|, |M| \geq n/2 + 7$ we have $|N(a_1) \cap M| \geq 2 \cdot 7 = 14$.[6] Select $c \in (N(a_1) \cap M) \setminus \{a_i^+, a_i^-, i = 1, \ldots, 6\}$ and set $b = c^-$. For T, the 2-factor produced by switching F_1 along $C := a_1a_2a_3a_4a_5a_6cba_1$, we see that compared to F_1, T contains at least two new edges of F_2 (namely a_2a_3, a_4a_5) but T may have lost one edge of F_2 (namely bc if it was in fact an edge of F_2), giving a net gain of one. Since T and F_1 have the same number of edges, we see that $|T \triangle F_2| \leq |F_1 \triangle F_2| - 1$, as required. $\square$

6 In the case of bipartite graphs, we note that $a_1a_2a_3a_4a_5a_6$ is a walk in G and so a_1 and a_6 are in different parts; say $a_1 \in A$ and $a_6 \in B$. Then $N(a_1), M \subseteq B$, so since $|N(a_1)|, |M| \geq \frac{1}{2}n + 7$, so $|N(a_1) \cap M| \geq 14$, and we continue as before.

4.1.4 *Counterexamples*

We continue with examples showing non-irreducibility under certain assumptions on $\delta(G)$ and k, as stated in contributions (ii) and (iii) in Section 4.1.1.

Example 4.1.5 (The case $\delta(G) = \frac{2n}{3} - 1$ and $k = 2$)**.** Construct $G = (V, E)$ as follows: Set $V = A_1 \cup A_2 \cup A_3$, where $|A_i| = n/3 =: m$. For convenience, we select n such that m is odd and $m \geq 3$. We denote the vertices of A_i by $v_{i,j}$ for $j = 1, \ldots, m$. Take as edge set E all edges between vertices in A_1, all edges between vertices in A_3, and all edges from vertices in A_i to vertices in A_{i+1} for $i = 1, 2$ (see Figure 16).

We color edges as follows: All edges incident to a vertex in A_1 are colored blue, and all other edges red. Note that all cycles of length 4 contain an even number of red and blue edges. This means that any switch along a 4-cycle preserves the parity of red and blue edges.

We will finish the construction by describing two Hamilton cycles H_1 and H_2 that have different parities of blue edges. As any 2-switches preserve the parity of blue edges, H_1 cannot be converted to H_2 via 2-switches.

The blue edges in H_1 are $v_{2,1}v_{1,1}$, $v_{1,k}v_{1,k+1}$ for $k = 1, \ldots, m-1$ and $v_{1,m}v_{2,m}$. The red edges in H_1 are $v_{2,k}v_{3,k}$, $v_{3,k}v_{2,k+1}$ for $k = 1, m-2$ and $v_{2,m-1}v_{3,m-1}$, $v_{3,m-1}v_{3,m}$, $v_{3,m}v_{2,m}$, see Figure 16. There are an even number of blue edges and an odd number of red edges in H_1. The Hamilton cycle H_2 is constructed by swapping the roles of the blue and red edges.

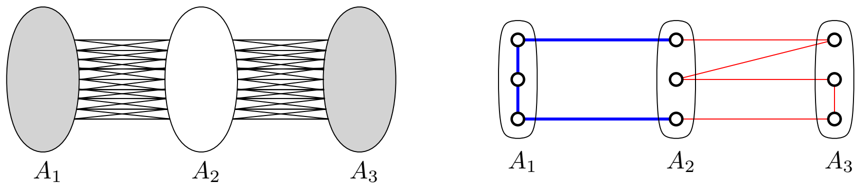


Figure 16: Left: The graph G. Right: The Hamilton cycle H_1 in G with $n = 9$. There are an even number of (thick) blue edges and an odd number of (thin) red edges.

Example 4.1.6 (The case $\delta(G) \approx \frac{n}{2}$ for each fixed k.)**.** For k fixed and $n \geq 3k + 5$, there is a graph G with $\delta(G) \geq (n - 3k - 4)/2$ for which $\mathcal{H}_G$ is not k-switch irreducible. Our construction relies on the following lemma.

Lemma 4.1.7. For any ℓ, there is a graph X with $3\ell+1$ vertices that has exactly two Hamilton paths H_1 and H_2. Moreover, these two paths satisfy $|H_1 \triangle H_2| = 2\ell$.

Proof. Without loss of generality let ℓ be odd, and set $n = 3\ell+1$. Let $X = (V, E)$ with $V := \{v_1, \ldots, v_n\}$ and $E := E_1 \cup E_2$, where $E_1 = \{v_i v_{i+1} \mid 1 \leq i \leq n-1\}$, and $E_2 = \{v_j v_{j+4} \mid j \equiv 2 (\text{mod } 3) \text{ and } j \leq n-5\} \cup \{v_3 v_{n-2}\}$; see left side of Figure 17. As vertices v_1 and v_n have degree 1, they must be the ends of any Hamilton path in X. Vertices v_i with $i \equiv 1 (\text{mod } 3)$ and $4 \leq i \leq n-3$ have degree 2 in X, so both of their incident edges must be part of any Hamilton path; call the set of these 2ℓ edges F and call the remaining edges F'. Note that the edges of F' form a cycle $C \subseteq X$. In F, every vertex of V has degree 1 or 2 and those vertices of degree 1 (except for v_1 and v_n) are precisely the vertices in the cycle C. Therefore we can only extend F to a Hamilton path by adding a perfect matching from C, and it is easy to see that adding either perfect matching from C results in a Hamilton path. These Hamilton paths have symmetric difference of size $|E(C)| = |F'| = 2\ell$. □

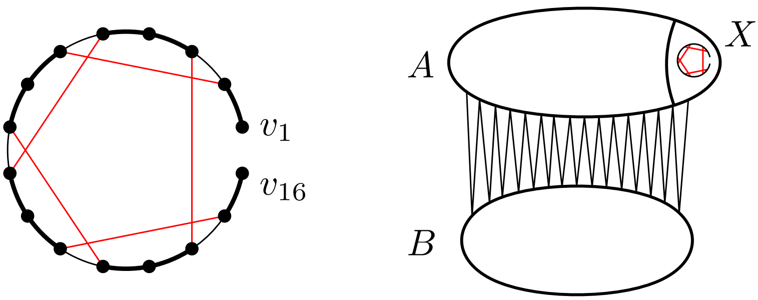


Figure 17: Left: Example of X for $\ell = 5$ (edges in E_1 are black, edges in E_2 are red, edges in F are heavy, edges in C are thin). Right: The example graph G

For the example we begin by applying Lemma 4.1.7 with $\ell = k+1$ to obtain the graph X of order $r := 3\ell + 1$. For any n such that $n+r$ is odd, we construct our example G by taking an (unbalanced) complete bipartite graph with parts A and B of size $\frac{n+(r-1)}{2}$ and $\frac{n-(r-1)}{2}$ respectively and adding a copy of X inside A. See Figure 17, right side.

As there are no edges inside B, any Hamilton cycle of G must use $r-1$ edges inside A, and so these must be within X. Since X has r vertices, any Hamilton cycle of G must induce a Hamilton path on X. By construction, X has exactly two Hamilton paths H_1 and H_2, and they have a symmetric difference of $2k+2$. It is easy to see that G has Hamilton cycles that use each of the two Hamilton paths in X, but it is impossible to perform a sequence of k-switches to transform a Hamilton cycle that uses H_1 into one that uses H_2; indeed if such a sequence existed, examining its restriction to X would yield a sequence of switches of size at most k that transforms H_1 into H_2 but maintaining a Hamilton path in X at each stage; this is impossible since X has only two Hamilton paths and their symmetric difference has size $2\ell = 2(k+1) > k$.

4.1.5 *Concluding remarks*

Overall, several interesting new questions arise in light of our work and we hope our results will stimulate more work in the area. In particular, what is the smallest k for which $\mathcal{H}_G$ is (strongly) k-switch irreducible for graphs with $\delta(G) \geq \frac{n}{2} + c$, where c is a (small) constant? Furthermore, given the interest in the 2-switch irreducibility for other combinatorial objects (see Subsection 4.2.1), what is the smallest[7] constant $\frac{2}{3} \leq \gamma \leq 1$ such that $\mathcal{H}_G$ is 2-switch irreducible for all graphs with $\delta(G) \geq \gamma n + c$ for some (small) constant c?

4.2 RAPID MIXING FOR DENSE MONOTONE GRAPHS

In this section we apply our results from Section 4.1 to analyze certain Markov chains. Again the section starts with the problem and our results, followed by general context (Subsection 4.2.1). We include an informal introduction to computational counting and sampling (Subsection 4.2.2) and continue with preliminaries (Subsection 4.2.3) and then the proofs of the main results (Subsections 4.2.4 and 4.2.5) and some concluding remarks (Subsection 4.2.6).

7 It is not hard to argue that the result is true for complete graphs G where $\gamma = c = 1$.

4.2.1 *Introduction*

For each $t \in \mathbb{N}_0$ let X_t be a random variable with state space Ω. The family of random variables $(X_t)_{t=0}^{\infty}$ is a *Markov chain* if

$$\mathbb{P}[X_{n+1} = x_{n+1} \mid X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n] \\ = \mathbb{P}[X_{n+1} = x_{n+1} \mid X_n = x_n]$$

for all choices $n \in \mathbb{N}_0$ and $x_0, \ldots, x_{n+1} \in \Omega$. This property means that the state of the next random variable only depends on the state of the current random variable. We will only consider Markov chains where the transition probability is independent of t. This allows us to write

$$P(x, y) \coloneqq \mathbb{P}[X_{n+1} = y \mid X_n = x],$$

where P is the transition matrix of the Markov chain. More generally for $t \in \mathbb{N}$ we may also define

$$P^t(x, y) \coloneqq \mathbb{P}[X_{n+t} = y \mid X_n = x].$$

A *stationary distribution* of a Markov chain is a probability distribution $\pi : \Omega \to [0, 1]$ such that

$$\pi(y) = \sum_{x \in \Omega} \pi(x) P(x, y).$$

A Markov chain is *irreducible* if for all $x, y \in \Omega$, we have $P^t(x, y) > 0$ for some $t \in \mathbb{N}$. A Markov chain is *aperiodic* if $\gcd(t \mid P^t(x, x) > 0) = 1$ for all $x \in \Omega$. If π is the stationary distribution of a Markov chain, that Markov chain is *time-reversible* if, for all $x, y \in \Omega$, we have $\pi(x)P(x, y) = \pi(y)P(y, x)$. A Markov chain is *lazy* if $P(x, x) > 0$ for all $x \in \Omega$.

We follow Jerrum [45] for an introduction to the concepts discussed below. Let $\mathcal{M}$ be an aperiodic, irreducible and time-reversible Markov chain on a finite state space Ω with transition matrix P. Note that $\mathcal{M}$ has a unique stationary distribution π, and if P is symmetric, then π is the uniform distribution on Ω. For two probability distributions π and π' on Ω, define the *total variation distance* between π and π' as

$$||\pi - \pi'||_{\mathrm{TV}} \coloneqq \frac{1}{2} \sum_{\omega \in \Omega} |\pi(\omega) - \pi'(\omega)|.$$

The total variation distance of the distribution $P^t(x,\cdot)$ from the (unique) stationary distribution π at time t with initial state x is defined as

$$\Delta_x(t) := ||P^t(x,\cdot) - \pi||_{\mathrm{TV}} = \frac{1}{2}\sum_{y\in\Omega} \left|P^t(x,y) - \pi(y)\right|,$$

and the *mixing time* of $\mathcal{M}$ is defined as

$$\tau(\varepsilon) := \max_{x\in\Omega} \min\{t \mid \Delta_x(t') \leq \varepsilon \text{ for all } t' \geq t\}.$$

Informally, $\tau(\varepsilon)$ is the number of steps until the Markov chain is guaranteed to be 'ε-close' to its stationary distribution given any starting state. We only consider Markov chains that have uniform stationary distributions. In that context, a set of Markov chains is said to be rapidly mixing if there exists a polynomial p such that for each Markov chain $\mathcal{M}$ in the set, its mixing time $\tau(\varepsilon)$ can be upper bounded by $p(\ln(|\Omega|/\varepsilon))$, where Ω is the state space of $\mathcal{M}$. The set of Markov chains will always be clear from context; for example we often discuss the k-switch Markov chain on $\mathcal{H}_G$ (defined below) for all graphs G in some graph class, which gives a set of Markov chains (with different state spaces).

We will be concerned with switch Markov chains. They are arguably the simplest and most natural Markov chains on the set of Hamilton cycles of a graph. Given a graph G recall the definitions of $\mathcal{H}_G$, k-switches and strong k-switch irreducibility from Section 4.1. For a given constant $k \in \mathbb{N}$, the *k-switch Markov chain* on $\mathcal{H}_G$ is defined as follows. Given that the Markov chain is currently in state $H \in \mathcal{H}_G$, we first pick $\ell \in \{1,\dots,k\}$ uniformly at random, and then select a set $L \subseteq E(G)$ with $|L| = 2\ell$ uniformly at random. If the graph H' with edge set

$$E(H') = E(H) \triangle L$$

is again in $\mathcal{H}_G$, i.e., a Hamilton cycle of G, then we transition to H'. Otherwise, we do nothing and stay in the state H. Note that the k-switch Markov chain on $\mathcal{H}_G$ is aperiodic (since it is lazy) and time-reversible. Further, the transition matrix of the k-switch Markov chain is symmetric and so its unique stationary distribution is the uniform distribution. As we have seen in Section 4.1, these Markov chains are not always irreducible.

We will consider the k-switch Markov chain on $\mathcal{H}_G$ for monotone graphs G (also known as bipartite permutation graphs). A bipartite graph $G =$

$(A \cup B, E)$, with $|A| = |B| = n$, is *monotone* if there exists a permutation $(a_1, \ldots, a_n)$ of the vertices in A and a permutation $(b_1, \ldots, b_n)$ of the vertices in B, such that the adjacency matrix C of G, with rows indexed by $a_1, \ldots, a_n$ and columns indexed by $b_1, \ldots, b_n$, has *monotone rows and columns*. This means that for each i, there exists $1 \leq r_i \leq t_i \leq n$ such that $C(a_i, b_j) = 1$ if and only if $r_i \leq j \leq t_i$ and the sequences $(r_i)_{i=1}^n$ and $(t_i)_{i=1}^n$ are non-decreasing. Intuitively, this means that the 1-entries in every row and column are contiguous. Note that although the definition does not immediately appear to be symmetric in A and B, one can easily check that it is. An example of such an adjacency matrix of a monotone graph is

$$C = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

We will call a monotone graph G *dense* if $\delta(G) \geq n/2 = |G|/4$.

Our contributions Our main contribution in this section is as follows:

Theorem 4.2.1. Let $\mathcal{D}$ be a set of monotone graphs G with $\delta(G) \geq n/2$ where $2n$ is the number of vertices in G. If $\mathcal{D}$ is strongly k-switch irreducible for Hamilton cycles for some $k \in \mathbb{N}$ (this is the case for $k = 10$ by Remark 4.1.2 if $\delta(G) \geq n/2 + 7$), then the set of k-switch Markov chains on $\mathcal{H}_G$ for each $G \in \mathcal{D}$ is rapidly mixing.[8]

Strong k-switch irreducibility for monotone graphs for $k = 10$ (see Subsection 4.1.3) plays a key role in our proof, which we give later in Subsection 4.2.5.

Related work Dyer, Frieze, and Jerrum [22] consider the question of counting and sampling Hamilton cycles in graphs G with $\delta(G) \geq \alpha n$ for $1/2 < \alpha \leq 1$. For the sampling problem, they take a two-step approach. First, based on a result of Jerrum and Sinclair [46], they show that there

8 To be precise: there exists a polynomial p such that the mixing time $\tau(\varepsilon)$ of the k-switch Markov chain on $\mathcal{H}_G$ for $G \in \mathcal{D}$ is bounded by $p(|G| + \ln(\varepsilon^{-1}))$ and we note $|G| = O(\ln |\mathcal{H}_G|)$, as can be seen by adapting methods from [18, 68].

is a rapidly mixing Markov chain on the set $\mathcal{F}_G$ of all 2-factors of G (recall these are all subgraphs of G in which every vertex has degree 2). Then it is shown that the number of 2-factors in G is at most a polynomial factor larger than the number of Hamilton cycles in G. This then implies (roughly speaking) that if one takes a polynomial number of samples from the Markov chain that samples 2-factors approximately uniformly, most likely one of those samples will be a Hamilton cycle. This sample is then also an approximately uniform sample from the set of all Hamilton cycles in G.

At the end of their paper, Dyer, Frieze and Jerrum [22] ask if there is a rapidly mixing Markov chain on the set of Hamilton cycles, and possibly 'near-Hamilton cyles' on graphs with $\delta(G) \geq \alpha n$ for $1/2 < \alpha \leq 1$, that mixes rapidly.[9] The main result of this section answers this in the affirmative for the 10-switch Markov chain on dense monotone graphs. Moreover, item (*iii*) in 4.1.1 shows that the 2-switch Markov chain (arguably the simplest Markov chain on Hamilton cycles) cannot be used to address the question of Dyer, Frieze and Jerrum for all graphs with $\delta(G) \geq n/2$. This is because item (iii) in 4.1.1 shows the 2-switch Markov chain (for graphs of minimum degree bigger than $n/2$) is not always irreducible and therefore cannot converge to the uniform distribution on $\mathcal{H}_G$.

The mixing time of switch-based Markov chains have been studied extensively for sampling subgraphs of K_n with a given degree sequence, see, e.g., [4, 15, 48, 63]. It is well known, see e.g. [72], that every two graphs (thought of as subgraphs on K_n) with the same degree sequence can be transformed into each other with switches of size 2 (in K_n). This remains true if one restricts oneself to the class of all connected subgraphs of K_n with a fixed degree sequence [72]. In particular, relevant to our setting, Feder et al. [27] (implicitly) show that the 2-switch chain is rapidly mixing on the set of all Hamilton cycles in case G is the complete graph. There are more direct ways to obtain this result, but we mention it here as we rely on some of their ideas.

Monotone graphs, also known as bipartite permutation graphs, have been widely studied from the structural graph theory perspective, perhaps most notably in their characterization [70]. Monotone graphs are also considered in the context of switch-based Markov chains for the sampling of perfect matchings: in particular, Dyer, Jerrum and Müller [23] show that the 2-

9 To be precise, in [22] they ask: "Second, is there a random walk on Hamilton cycles and (in some sense) "near-Hamilton cycles" which is rapidly mixing?"

switch Markov chain for sampling perfect matchings is rapidly mixing on monotone graphs. We refer the reader to [23] for further results in this direction.

We mentioned in the previous section that Takaoka [71] shows that the set of all Hamilton cycles in a given monotone graph is 2-switch irreducible. We remark that in [71] this is established in the weak sense by showing that every Hamilton cycle can be transformed, by switches of size 2, into a fixed *canonical* Hamilton cycle. However, we need the stronger notion of irreducibility for our rapid mixing proof for dense monotone graphs to go through.

4.2.2 *A brief digression on sampling and counting*

A reader who is unfamiliar with rapid mixing might wonder how the mixing time of certain Markov chains is related to approximate counting. In this subsection we informally describe the ideas behind sampling and counting via Markov chains. We stress that none of the material in this subsection is required for the rest of the chapter and the reader may safely skip ahead to the next subsection.

Sampling: Sampling in this context refers to the problem of finding fast algorithms for selecting some object x out of a set Ω according to a desired distribution, often the uniform distribution. This is relatively simple if one can easily count and enumerate all of the objects in Ω, but if Ω is large (like $\mathcal{H}_G$), one needs a different strategy. Typically we want to select $x \in \Omega$ in time $\text{poly}(\ln |\Omega|)$.

One such strategy is simulating a Markov chain with state space Ω. If a Markov chain with stationary distribution π runs through 'enough' steps, the probability of ending up in state $x \in \Omega$ is close enough to $\pi(x)$, independent of the starting state. This naturally gives an algorithm for sampling from Ω with distribution π: simply simulate the Markov chain for 'enough' steps, then output the current state. In order for this algorithm to be fast, it is necessary that 'enough' steps do not take too long to simulate. This is where rapid mixing comes in useful. If a chain is rapidly mixing, roughly, this means we only need to simulate the chain for a small number of steps relative to $|\Omega|$ (typically logarithmic in $|\Omega|$). For example, in our case of sampling subgraphs of a graph of order n, we want to bound the mixing time by a polynomial in n. Note that in general we cannot sample from π precisely this way, merely from a distribution that is close enough to π.

Counting: Sampling and counting are closely related, and for many problems we can transform an (approximate) sampling algorithm into an (approximate) counting algorithm [45]. We give a rough sketch on how we may apply this principle to our example of approximately counting the Hamilton cycles $|\mathcal{H}_G|$ of a graph G on n vertices. Assume we can uniformly sample from $\mathcal{H}_G$. Fix an arbitrary edge $e \in E(G)$. We take a large (but polynomial in n) number of samples from $\mathcal{H}_G$ and compute the proportion that contain e. This proportion gives (with good probability) a good approximation for $|\mathcal{H}_{G-e}|/|\mathcal{H}_G|$ (provided this ratio is not too small or too large). In the same way, we may add any edge e_1 to G and approximate $|\mathcal{H}_G|/|\mathcal{H}_{G+e_1}|$. Let $e_1, e_2, \dots, e_k$ be the edges not in G. Then we may sample Hamilton cycles in $G + e_1$, $G + e_1 + e_2$,..., $K_n - e_k$ and use these to approximate the telescoping product

$$|\mathcal{H}_G| = \frac{|\mathcal{H}_G|}{|\mathcal{H}_{G+e_1}|} \frac{|\mathcal{H}_{G+e_1}|}{|\mathcal{H}_{G+e_1+e_2}|} \cdots \frac{|\mathcal{H}_{K_n-e_k}|}{|\mathcal{H}_{K_n}|} |\mathcal{H}_{K_n}|.$$

For a more detailed introduction, we refer the reader to [7].

4.2.3 *Preliminaries*

Markov chains and mixing times.

It is known that for time-reversible Markov chains, such as the ones we study, the transition matrix P only has real eigenvalues, which we denote by $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|\Omega|-1} > -1$. We can always replace the transition matrix P of the Markov chain by $(P + I)/2$, to make the chain lazy[10], and, hence, guarantee that all its eigenvalues are non-negative. It then follows that the second-largest eigenvalue in absolute value of (the new transition matrix) P is λ_1. In this work we always consider the lazy versions of the Markov chains involved, but we do not always mention this explicitly. It follows directly from Proposition 1 in [69] that

$$\tau(\varepsilon) \leq \frac{1}{1 - \lambda_1} \Big(\ln(1/\pi_*) + \ln(1/\varepsilon) \Big),$$

10 I.e. all diagonal entries of P are non-zero.

where $\pi_* = \min_{x\in\Omega} \pi(x)$. When π is the uniform distribution, the above bound reduces to

$$\tau(\varepsilon) \leq \frac{1}{1-\lambda_1}(\ln(|\Omega|) + \ln(1/\varepsilon)).$$

The quantity $(1-\lambda_1)^{-1}$ can be upper bounded using the *multicommodity flow method* of Sinclair [69].

We define the state space graph of the chain $\mathcal{M}$ as the directed graph $\mathbb{G}$ with vertex set Ω that contains exactly the arcs $(x, y) \in \Omega \times \Omega$ for which $P(x, y) > 0$ and $x \neq y$. Let $\mathcal{P} = \cup_{x\neq y}\mathcal{P}_{xy}$, where $\mathcal{P}_{xy}$ is the set of simple paths between x and y in $\mathbb{G}$. A *flow* f in Ω is a function $\mathcal{P} \to [0, \infty)$ with the property $\sum_{p\in\mathcal{P}_{xy}} f(p) = \pi(x)\pi(y)$ for all $x, y \in \Omega$, where $x \neq y$. The flow f can be extended to a function on oriented edges of $\mathbb{G}$ by setting $f(e) = \sum_{p\in\mathcal{P}:e\in p} f(p)$, so that $f(e)$ is the total flow routed through the edge $e \in E(\mathbb{G})$. We call $f(e)$ the *congestion* of e and $\max_{e\in E(\mathbb{G})} f(e)$ the congestion of f. Let $\ell(f) = \max_{p\in\mathcal{P}:f(p)>0} |p|$ be the length of a longest flow carrying path, and let $\rho(e) = f(e)/Q(e)$ be the *load* of the edge e, where $Q(e) = \pi(x)P(x, y)$ for $e = (x, y)$. The maximum load of the flow is then given by $\rho(f) = \max_{e\in E(\mathbb{G})} \rho(e)$. Sinclair, in Corollary $6'$ of [69], shows that

$$(1-\lambda_1)^{-1} \leq \rho(f)\ell(f).$$

We use the following (by now standard) technique for bounding the maximum load of a flow in the case that the chain $\mathcal{M}$ has uniform stationary distribution π. Suppose θ is the smallest positive transition probability of the Markov chain between two distinct states in Ω. If b is such that $f(e) \leq b/|\Omega|$ for all $e \in E(\mathbb{G})$, then it follows that $\rho(f) \leq b/\theta$. This implies that

$$\tau(\varepsilon) \leq \frac{\ell(f)\cdot b}{\theta} \ln(|\Omega|/\varepsilon)\,. \tag{4.2.1}$$

Now, if $\ell(f), b$ and $1/\theta$ can be bounded by a function polynomial in $\ln(|\Omega|)$ for some (set of) Markov chains, it follows that the Markov chains are rapidly mixing. In this case, we say that f is an *efficient* flow. Note that in this approach the transition probabilities do not play a role as long as $1/\theta$ is polynomially bounded.

For a more detailed introduction to this concept, we refer the reader to [45]. We remark that most importantly, we seek to bound the congestion by a polynomial in n; this will be enough to ensure rapid mixing.

4.2.4 *Rapid mixing on 2-factors*

We first present a result for the sampling of 2-factors using switch-based Markov chains, which will be used later on, and that might be of independent interest. Given a graph G, recall the k-switch Markov chain on $\mathcal{H}_G$ defined in the introduction. Replacing $\mathcal{H}_G$ with $\mathcal{F}_G$ (the set of all 2-factors of G) everywhere in that definition defines the k-switch Markov chain on $\mathcal{F}_G$. Here is the explicit definition for the reader's convenience.

For a given constant $k \in \mathbb{N}$, the *k-switch Markov chain* on $\mathcal{F}_G$ is defined as follows. Given that the Markov chain is currently in state $F \in \mathcal{F}_G$, we first pick $\ell \in \{1, \dots, k\}$ uniformly at random, and then select a set $L \subseteq E(G)$ with $|L| = 2\ell$ uniformly at random. If the graph F' with edge set

$$E(F') = E(F) \triangle L$$

is again in $\mathcal{F}_G$, i.e., a 2-factor of G, then we transition to F'. Otherwise, we do nothing and stay in the state F.

Theorem 4.2.2. Let $\mathcal{G}$ be the class of all graphs G with $\delta(G) \geq |V(G)|/2$. If $\mathcal{G}$ is strongly k-switch irreducible for 2-factors for some $k \in \mathbb{N}$ (this is the case for $k = 4$ by Proposition 4.1.4) then there is an efficient multicommodity flow for the k-switch Markov chain on $\mathcal{F}_G$ for each $G \in \mathcal{G}$. In particular, the set of k-switch Markov chains on $\mathcal{F}_G$ for all $G \in \mathcal{G}$ is rapidly mixing.

Moreover, Theorem 4.2.2 remains true for the bipartite case of the problem, where we are given a bipartite graph $G = (A \cup B, E)$ with both $|A| = |B| = n$, and where every vertex in $A \cup B$ has degree at least $n/2$.

The JS chain, which we detail below, is known to have an efficient multicommodity flow. Its state space contains $\mathcal{F}_G$, but also subgraphs of G which are only nearly 2-factors. The main idea behind the proof of Theorem 4.2.2 is to use the flow f on the JS chain in order to obtain such a flow g for $\mathcal{F}_G$. We obtain g from f by first restricting f to paths that only go between states in $\mathcal{F}_G$ and then making further adjustments, being careful not to increase the load on any edge by more than a factor polynomial in n. This is an example of the Markov chain comparison technique.

The proof of Theorem 4.2.2 is based on the embedding argument introduced in [4] for the switch Markov chain that samples graphs with a given degree sequence. It is perhaps interesting to note that it seems much harder to prove Theorem 4.2.2 by using other approaches for that problem, such as [15, 63]. These approaches do have the advantage that they get better mixing time bounds than those in [4].

The proof of Theorem 4.2.2 is a modification of certain parts in [4]. We will tailor all definitions to the notion of 2-factors for sake of readability. Let $\mathbf{2} = (2, 2, \ldots, 2)$ be the all-twos sequence of length n. Let $G \in \mathcal{G}$ be a given undirected n-vertex graph G with $\delta(G) \geq n/2$ and let $\mathcal{F}_G$ be the set of all 2-factors of G.

We write $G(d')$ for the set of all subgraphs of G with degree sequence d'. Let $\mathcal{F}'_G = \cup_{d'} G(d')$ with d' ranging over the set

$$\left\{ d' \mid d'_j \leq 2 \text{ for all } j, \text{ and } \sum_{i=1}^{n} |2 - d'_i| \leq 2 \right\}.$$

In other words, $\mathcal{F}'_G$ is the set of almost 2-factors, that is, subgraphs of G with degree sequence d' where (i) $d' = \mathbf{2}$, or (ii) there exist distinct κ, λ such that $d'_i = 1$ if $i \in \{\kappa, \lambda\}$ and $d'_i = 2$ otherwise, or (iii) there exists a κ so that $d'_i = 0$ if $i = \kappa$ and $d'_i = 2$ otherwise. In the case (ii) we say that d' has two vertices with degree deficit one, and in the case (iii) we say that d' has one vertex with degree deficit two.

Jerrum and Sinclair [46] define a Markov chain that, tailored to 2-factors, works as follows.

Let $F \in \mathcal{F}'_G$ be the current 2-factor of the JS chain. Choose an ordered pair of vertices (i, j) uniformly at random:

1. if $F \in \mathcal{F}_G$ and ij is an edge of F, delete ij from G (*Type 0 transition*),

2. if $F \notin \mathcal{F}_G$ and the degree of i in G is less than 2, and ij is not an edge of F, add ij to F if this edge is in G; if this causes the degree of j to exceed 2, select an edge jk uniformly at random from F and delete it (*Type 1 transition*).

In case the degree of j does not exceed 2 in the second case, we call this a *Type 2 transition*.

The graphs $F, F' \in \mathcal{F}'_G$ are *JS adjacent* if F can be obtained from F' with positive probability in one transition of the JS chain and note this relation is symmetric. The properties of the JS chain, stated in Theorem 4.2.3 below, are easy to check [46].

Theorem 4.2.3. The JS chain on $\mathcal{F}'_G$ is irreducible, aperiodic and symmetric, and, hence, has uniform stationary distribution over $\mathcal{F}'_G$. Moreover, $P(F, F')^{-1} \leq 2n^3$ for all JS adjacent $F, F' \in \mathcal{F}'_G$, and also the maximum in- and out-degrees of the state space graph of the JS chain are bounded by n^3.

We say that two graphs $F, F' \in \mathcal{F}'_G$ are *within distance r in the JS chain* if there exists a path of length at most r from F to F' in the state space graph of the JS chain. By $\text{dist}(F', \mathbf{2})$ we denote the minimum distance of $F' \in \mathcal{F}'_G$ to an element in $\mathcal{F}$. The following parameter will play a central role in this work. Let

$$k_{JS}(G) = \max_{F' \in \mathcal{F}'_G} \text{dist}(F', \mathbf{2})\,. \tag{4.2.2}$$

Based on the parameter k_{JS}, we define the notion of *strong stability* [4].

Definition 4.2.4 (Strong stability)**.** A family of graphs $\mathcal{D}$ is called *strongly stable* if there exists a constant ℓ such that $k_{JS}(G) \leq \ell$ for all $G \in \mathcal{D}$.

It is shown by Jerrum and Sinclair [46], that if $\mathcal{D}$ is the set of all graphs G with $\delta(G) \geq n/2$, then $\mathcal{D}$ is strongly stable for $\ell = 3$.[11] (This gives rise to the condition on the minimum degree in the statement of Theorem 4.2.2.)

We now have all the ingredients for the proof of Theorem 4.2.2. It uses essentially the same argument as that in [4], where it is shown that the switch Markov chain for sampling graphs with given degrees is rapidly mixing for certain strongly stable classes of degree sequence, i.e., for the notion of strong stability in that setting which corresponds to Definition 4.2.4 in our setting.

Proof of Theorem 4.2.2. The high-level idea is to use an embedding argument which states that an efficient multi-commodity flow for the JS chain can be transformed into an efficient flow for the k-switch Markov chain on $\mathcal{F}_G$.

11 See Theorem 4.1 there. This is implicitly shown in the proof.

The fact that there exists an efficient multi-commodity flow for the JS chain can be shown using exactly the same arguments as in Theorem 3.2 in [4].[12]

Without going into all the details, we will give a sketch of this argument. Recall that Sinclair's multi-commodity flow method asks us to define a flow f in the state space graph of the JS chain that routes a fraction $\pi(X)\pi(Y)$ of flow from X to Y for every $X, Y \in \mathcal{F}'_G$. Here,

$$\pi(Z) = \frac{1}{|\mathcal{F}'_G|}$$

for every $Z \in \mathcal{F}'_G$.

The notion of strong stability allows us to take a shortcut here: Instead of defining a flow between every two states in $\mathcal{F}'_G$, one can first define a flow between any two 2-factors $F, F' \in \mathcal{F}_G$. Then, roughly speaking, in order to define a flow between any two states in $\mathcal{F}'_G$, we use the fact that every 'almost 2-factor' $X \in \mathcal{F}'_G \setminus \mathcal{F}_G$ is close to some actual 2-factor in the state space graph, because of strong stability. These short paths between states in $\mathcal{F}'_G \setminus \mathcal{F}_G$ and $\mathcal{F}_G$ can be exploited to define the desired flow between any two states in $\mathcal{F}'_G$.

In order to define the flow between two 2-factors F and F', we decompose the symmetric difference $F \triangle F'$ into a collection of alternating circuits.[13] We then use the operations defining the JS chain in order to transform F into F' by 'flipping' edges on an alternating circuit in order to move from F to F'; see Figure 18 for a short example and [4] for a more detailed explanation.

In particular, all these flow-carrying paths will have polynomial length. Moreover, all these operations only use edges in $F \triangle F'$ and so the approach taken in the proof of Theorem 3.2 in [4] can be used here as well (when G is not a complete graph) to give Lemma 4.2.5 below.

12 That theorem essentially shows the result in the case where the graph G is complete and strong irreducibility for $k = 2$, but the analysis remains true when G is not a complete graph, and when $k > 2$ (still assuming the notion of strong stability of the given class of degree sequences).

13 To be more precise, the flow is spread out over all possible ways in which the symmetric difference can be decomposed.

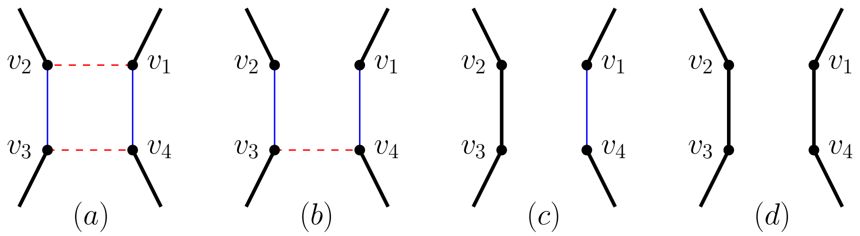


Figure 18: An example of how to process one circuit in the symmetric difference of two 2-factors F and F'. (a): The alternating circuit $v_1v_2v_3v_4v_1$ is in $F\triangle F'$. Black thick edges are in both F and F', red dashed edges are only in F, blue (thin, normal) edges are only in F'. We present operations in the JS chain that remove edges in $F \setminus F'$ and add edges in $F' \setminus F$. Position (b) occurs after a type 0 transition on v_1v_2. Position (c) occurs after a type 1 transition, adding v_2v_3 and removing v_3v_4. Finally, (d) is achieved with a type 2 transition on v_3v_4. The order of these operations is obtained by considering a fixed total order on the edges.

Lemma 4.2.5. Let $\mathcal{D}$ be the collection of graphs with $\delta(G) \geq n/2$. Then there exist polynomials $p(n)$ and $q(n)$ such that for any $G \in \mathcal{D}$ there exists an efficient multi-commodity flow f for the JS chain on $\mathcal{F}'_G$ satisfying

$$\max_e f(e) \leq p(n) \quad \text{and} \quad \ell(f) \leq q(n),$$

where $f(e)$ is the total amount of flow routed over edge e in the state space graph, and $\ell(f)$ the maximum length of a flow-carrying path.

The next step entails transforming the flow f in Lemma 4.2.5 into an efficient multi-commodity flow for the k-switch Markov chain on $\mathcal{F}_G$ (assuming strong irreducibility). First note that the flow f above is a flow between any two states in $\mathcal{F}'_G$, whereas we are interested in defining a flow, let us call it g, between any two states in $\mathcal{F}_G$. Therefore, the first step will be to restrict ourselves to the flow routed in f between states in $\mathcal{F}_G$, which we call $\tilde{f}$.

A subtlety here is that we route a flow of $1/|\mathcal{F}'_G|^2$ between any two states in $\mathcal{F}_G$ in $\tilde{f}$ (and also f), whereas we need to route $1/|\mathcal{F}_G|^2$ between two such states in the desired (final) flow g. This is not a problem as replacing $|\mathcal{F}'_G|$ by $|\mathcal{F}_G|$ in the definition of $\tilde{f}$ only blows up the congestion $f(e)$ on a given edge e, by at most a polynomial factor, using the fact that

$$\frac{|\mathcal{F}'_G|}{|\mathcal{F}_G|} \leq s(n)$$

for some polynomial s, since $\delta(G) \geq \frac{1}{2}n$.[14] Let us call the resulting (intermediate) flow $\bar{f}$, which now routes a fraction $1/|\mathcal{F}_G|^2$ of flow between any two states in $\mathcal{F}_G$ in the JS chain, and that has polynomially bounded congestion.[15]

We next continue with transforming the flow $\bar{f}$ into the desired flow g. We do this by a sequence of reductions.

We first identify for every $X \in \mathcal{F}'_G \setminus \mathcal{F}_G$ some 2-factor $\psi(X) \in \mathcal{F}_G$ that is within $k_{JS} = 3$ moves (in the JS chain) away from X. All X that map onto the same 2-factor $F = \psi(X)$ are merged with F into a supervertex that we identify with F. If this procedure gives rise to parallel (directed) edges, we replace them by one edge and route all flow over that edge; self-loops are removed. It is not hard to see $|\psi^{-1}(F)|$ has size polynomial in n, as we only merge vertices that are close to each other (in the original JS chain) and the maximum degrees are bounded by n^3. Moreover, it is not hard to see that this procedure will only give rise to at most a polynomial number of parallel edges between two given vertices in $\mathcal{F}_G$ (for the same reason). Let us call the resulting (simple) graph $\mathbb{J} = (\mathcal{F}_G, A)$. As $X \in \mathcal{F}'_G \setminus \mathcal{F}_G$ are merged into $\psi(X)$, any edge (X, Y) in $\mathcal{F}'_G$ corresponds to an edge in $\mathbb{J}$, and so every path in $\mathcal{F}'_G$ corresponds to a path in $\mathbb{J}$. Here the original edge (X, Y) corresponds to the edge $(\psi(X), \psi(Y))$ in $\mathbb{J}$, and loops that occur after merging are ignored. The flow $\bar{f}$ induces a flow f^* on $\mathbb{J}$. Since $\bar{f}$ sends a flow through a path, we define f^* as sending the same flow through the corresponding path in $\mathbb{J}$. We now see that for $e = (F, F') \in A$ we have

$$f^*(e) = \sum_{X \in \psi^{-1}(F), Y \in \psi^{-1}(F')} \bar{f}(X, Y).$$

By what is said above, we have $\max_e f^*(e) \leq p'(n)$ for some polynomial p', i.e., the congestion of f^* is at most a polynomial factor larger than that of $\bar{f}$.

The final problem, before we obtain the desired flow g, is that the graph $\mathbb{J}$ contains edges (possibly with flow) between 2-factors $F, F' \in \mathcal{F}_G$ that might be more than a k-switch away from each other. Said differently, these edges

---

14 Given $F \in \mathcal{F}'_G \setminus \mathcal{F}_G$, let x, y be vertices of degree 1 or $x = y$ the vertex of degree 0. Find $z \in N(x) \cap N(y)^+$ and replace zz^- with xz, yz^- to obtain $\sigma(F) \in \mathcal{F}_G$ with $|F \triangle \sigma(F)| \leq 3$. Thus $|\sigma^{-1}(F)| \leq n^3 =: s(n)$.

15 The flows $\tilde{f}$ and $\bar{f}$ are not efficient multi-commodity flows for Markov chains, but 'auxiliary flows'.

do not represent transitions in the k-switch Markov chain. Let us partition the edge set $A = A_{\text{switch}} \cup A_{\text{infeasible}}$ where A_{switch} contains all edges of A that represent a transition in the k-switch Markov chain, and $A_{\text{infeasible}}$ all those edges that do not.

We argue that for every edge $a = (F, F') \in A_{\text{infeasible}}$, we can always find a short 'detour' in the graph $\mathbb{J}$ using only edges in A_{switch}. To see this, fix some $a \in A_{\text{infeasible}}$. Suppose that X and Y are adjacent in the JS chain and that $F = \psi(X)$ and $F' = \psi(Y)$ (these X and Y exist by existence of the infeasible edge a). Since $k_{JS} = 3$, it can be shown that

$$|F \triangle F'| \leq 12.$$

This follows from the fact that in the JS chain, $F = \psi(X)$ is close to X, which is close to Y, which is in turn close to $\psi(Y) = F'$.[16] Recall that since the graph class $\mathcal{G}$ in Theorem 4.2.2 is strongly k-switch irreducible and $G \in \mathcal{G}$, there exists a function ϕ such that for any $F, F' \in \mathcal{F}_G$ with $|F \triangle F'| \leq t$, there exists a sequence of at most $\phi(t)$ k-switches transforming F into F'. It follows that we can find a detour from F to F' of length at most $\phi(12)$, and this detour only uses edges in A_{switch}.

Since all these detours take place on a 'local' level, the congestion of the resulting multi-commodity flow for the k-switch Markov chain on $\mathcal{F}_G$, that we get from rerouting the flow of infeasible edges over their respective detour, increases at most by a polynomial factor on every *fixed* feasible edge in $\mathbb{J}$. That is, for a fixed edge $b = (F_0, F_0') \in A_{\text{switch}}$, the total number of edges $a = (F, F') \in A_{\text{infeasible}}$ that use b in their detour is at most poly(n), as (roughly speaking) F_0 is at most $\phi(12)$ transitions away from F by construction (and $\phi(12)$ is constant).

This yields the desired flow g. For a precise and detailed outline of this idea, we refer the reader to [4]. □

4.2.5 *Hamilton cycles in dense monotone graphs*

In this section we will describe a rapid mixing result for sampling Hamilton cycles from dense monotone graphs that is based on Theorem 4.2.2. We repeat the main theorem of this section.

16 We have $|\psi(X) \triangle X| \leq 5$ and $|X \triangle Y| \leq 2$ as they are 3 resp. 1 step in the JS chain, and at least one of the transitions from $\psi(X)$ to X is of type 0 or 2.

Theorem 4.2.1. Let $\mathcal{D}$ be a set of monotone graphs with $\delta(G) \geq n/2$ where $2n$ is the number of vertices in G. If $\mathcal{D}$ is strongly k-switch irreducible for Hamilton cycles for some $k \in \mathbb{N}$, then the set of k-switch Markov chains on $\mathcal{H}_G$ for $G \in \mathcal{D}$ is rapidly mixing.

As mentioned earlier, the set of all Hamilton cycles for (not necessarily dense) monotone graphs is connected under switches of size two [71] in the weak sense as defined in the preliminaries. Takaoka shows that every Hamilton cycle can be transformed into a 'canonical' Hamilton cycle using switches of size two. This is, however, not enough for the argument we will give below. For our argument we need the strong sense of irreducibility. The proof of Theorem 4.2.2 uses a Markov chain comparison, this time between the k-switch Markov chains on $\mathcal{H}_G$ and $\mathcal{F}_G$. We know that the 4-switch Markov chain on $\mathcal{F}_G$ is rapidly mixing by Theorem 4.2.2 and Proposition 4.1.4.

Proof of Theorem 4.2.1. The proof relies on an embedding argument similar to that in [27], but technically somewhat different. While the argument in [27] corresponds to the case where G is a complete bipartite graph (which is indeed monotone), here we relax the argument so that it extends to monotone graphs.

Let $G \in \mathcal{D}$ be given. In particular, our goal is to show, for every $G \in \mathcal{D}$, the existence of a function $\phi : \mathcal{F}_G \to \mathcal{H}_G$ with the properties

i) $|\phi^{-1}(H)| \leq \text{poly}(n)$ for every $H \in \mathcal{H}_G$, and,

ii) there exists a function $f : \mathbb{N} \to \mathbb{N}$ such that whenever $F, F' \in \mathcal{F}_G$ with $|F \triangle F'| \leq k$, we have $|\phi(F) \triangle \phi(F')| \leq f(k)$.

If such a function exists, one can argue exactly as in [27] that every efficient multi-commodity flow for the k-switch Markov chain on the set of all 2-factors $\mathcal{F}_G$ can be transformed into an efficient multi-commodity flow for the k-switch Markov chain on the set of all Hamilton cycles $\mathcal{H}_G$.[17] The embedding argument from [27] that we refer to here is essentially the same as that used to prove Theorem 4.2.2.

The differences are as follows. We use Theorem 4.2.2 and Proposition 4.1.4 to establish an efficient multicommodity flow on the k-switch

17 In [27], it is shown that any efficient flow for the 2-switch Markov chain for sampling subgraphs of K_n with a given degree sequence can be turned into an efficient flow for the 2-switch Markov chain for sampling *connected* graphs with a given degree sequence.

Markov chain on $\mathcal{F}_G$. We restrict the flow to paths that go between states in $\mathcal{H}_G$. Then we adjust the flow to accommodate the difference in size of the state space, using i) in order to show that adjusting the flows does not blow up the congestion by more than a polynomial factor. We then contract the graph by merging each state F with $\phi(F)$, which induces a flow f on the k-switch chain on $\mathcal{H}_G$ (corresponding to g in the proof of Theorem 4.2.2). When arguing that the congestion of f is not too large, ii) shows that two 2-factors that produce an infeasible edge map to two Hamilton cycles that have symmetric difference at most $f(2k)$, and the strong k-switch irreducibility with associated function ψ then shows that the detour due to infeasible edges has length at most $\psi(f(2k))$, a constant.

The remainder of the proof is dedicated to showing the existence of such a function ϕ for each $G \in \mathcal{D}$, which we will do in three claims. Let $G = (A \cup B, E) \in \mathcal{D}$ be a monotone graph with $|A| = |B| = n$ where we assume that n is even for simplicity.[18] Let $a_1, \ldots, a_n$ (resp. $b_1, \ldots, b_n$) be the vertices of A (resp. B) in order as given in Subsection 4.2.1. Set $A_1 = \{a_1, \ldots, a_{n/2}\}$ with $A_2 = A \setminus A_1$ and $B_1 = \{b_1, \ldots, b_{n/2}\}$ with $B_2 = B \setminus B_1$.

Claim 4.2.6. With the setup above, the graphs $G[A_1 \cup B_1]$ and $G[A_2 \cup B_2]$ are complete bipartite.

Claim 4.2.7. Given $G \in \mathcal{D}$, let $\mathcal{P}_G$ be the set of all subgraphs $K \subseteq G$ such that K is the union of three vertex-disjoint paths that together cover all vertices of G. Then there exists an injective function $\phi_1 : \mathcal{F}_G \to \mathcal{P}_G$ and a function $g : \mathbb{N} \to \mathbb{N}$ such that whenever $F, F' \in \mathcal{F}_G$ with $|F \triangle F'| \leq k$, we have $|\phi_1(F) \triangle \phi_1(F')| \leq g(k)$.

Claim 4.2.8. Given $G \in \mathcal{D}$, there is a function $\phi_2 : \mathcal{P}_G \to \mathcal{H}_G$ such that for every $K \in \mathcal{P}_G$, we have that $|K \triangle \phi_2(K)| \leq 9$; in particular, for each $H \in \mathcal{H}_G$, we have $|\phi_2^{-1}(H)| \leq |E(G)|^9 = \text{p}oly(n)$.

The function ϕ is the composition of ϕ_1 and ϕ_2 and can easily be seen to satisfy the desired properties (taking $f(k) = g(k) + 18$). Therefore it remains only to prove the claims.

Proof of Claim 4.2.6. Note that $a_1 b_1$ must be an edge of G. If this is not the case, then b_1 can never have positive degree, because of monotonicity of the rows of the adjacency matrix. As both a_1 and b_1 have degree at least $n/2$, we

18 When n is odd, one can work with $\lceil n/2 \rceil$ instead of $n/2$ throughout the proof.

can conclude that all edges of the form $a_i b_j$ with $1 \leq i, j \leq n/2$ are present (again because of monotonicity) so $G[A_1 \cup B_1]$ is complete bipartite. A similar argument holds for the edge $a_n b_n$ that yields $G[A_2 \cup B_2]$ is complete bipartite. ◀

Proof of Claim 4.2.7. We use a similar idea as in [27]. We fix the total orderings

$$a_{\frac{n}{2}+1} < a_{\frac{n}{2}+2} < \cdots < a_n < a_1 < a_2 < \cdots < a_{\frac{n}{2}}$$

on the vertices in A and

$$b_{\frac{n}{2}+1} < b_{\frac{n}{2}+2} < \cdots < b_n < b_1 < b_2 < \cdots < b_{\frac{n}{2}}$$

on the vertices of B.

Fix $F \in \mathcal{F}_G$ and let $C_1, \ldots, C_q$ be the cycles (or connected components) of F. For a given cycle C_r, we use a^r to denote the highest ordered vertex of A in C_r, and we use b^r to denote the highest ordered vertex of B in C_r. We first group the cycles in three sets depending on the vertices a^r and b^r. We define

$$Q_{A_1} = \{C_r \mid a^r \in A_1\}, \qquad Q_{B_1} = \{C_r \mid a^r \in A_2 \text{ and } b^r \in B_1\}$$

and $Q_{A_2 \cup B_2}$ as the set of all remaining cycles not in Q_{A_1} or Q_{B_1}. Note that the cycles in $Q_{A_2 \cup B_2}$ are fully contained in $A_2 \cup B_2$. For each cycle C^r in Q_{A_1} and $Q_{A_2 \cup B_2}$, let c^r be an arbitrary neighbor of a^r in C^r and for each cycle C^r in Q_{B_1} let d^r be an arbitrary neighbor of b^r on C^r (in each case there are two choices). We delete the edges $a^r c^r$ and $b^r d^r$ from F to create paths; we will connect the paths in each group together to build the three paths which will define $\phi_1(F) \in \mathcal{P}_G$.

We first explain the idea (of Feder et al. [27]) on how to glue together the paths from $Q_{A_2 \cup B_2}$ in such a way that we can uniquely recover the original paths from the single glued path: this case is easiest because we know from Claim 4.2.6 that the graph $G[A_2 \cup B_2]$ is complete bipartite.

After renaming the cycles, let us assume the cycles in $Q_{A_2 \cup B_2}$ are $C^1, \ldots, C^q$ where $a^1 < a^2 \cdots < a^q$. Let P_r be the path obtained by deleting the edge $a^r c^r$ from the cycle C_r. As all the cycles lie entirely within $A_2 \cup B_2$ and $G[A_2 \cup B_2]$ is complete bipartite, we know that all the edges $c^r a^{r+1}$ are present in G for $r = 1, \ldots, q-1$. Adding these edges to the graph consisting of $P_1, \ldots, P_q$, results in a path that we call $P_{A_2 \cup B_2}$.

Note that, given $P_{A_2 \cup B_2}$, (without knowing the paths $P_1, \ldots, P_q$), we can uniquely recover these $P_1, \ldots, P_q$ as follows. We know that the endpoint of

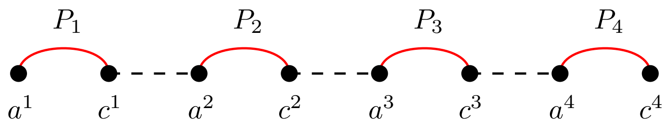


Figure 19: Sketch of path P from the paths $P_1, \dots, P_q$ for the case $q = 4$.

$P_{A_2 \cup B_2}$ that is contained in A is the first vertex of P_1, i.e., the vertex a^1 (the other endpoint is necessarily in B). In order to recover P_1 we start following the path $P_{A_2 \cup B_2}$, starting from a^1, until we reach the first vertex in A that is ordered higher than a^1; this is the first vertex of P_2, i.e., the vertex a^2. Continuing in this fashion we can uniquely recover all the paths P_i.

We apply a similar procedure to the paths obtained from Q_{A_1} and Q_{B_1} to form paths P_{A_1} and P_{B_1}, respectively. The problem here is that the underlying graph is not complete bipartite so we do not apriori know if the edges to 'glue' the paths together are all present: we argue that they are in fact present. The proof for Q_{A_1} that we will give below also holds for Q_{B_1} by symmetry of monotonicity (the case of Q_{B_1} is essentially a slightly more restrictive setting in which some of the cases below cannot occur).

Assume that the cycles in Q_{A_1} are $C_1, \dots, C_p$ labeled so that $a^1 < a^2 < \cdots < a^p$. By means of a case distinction, depending on whether $c^r \in B_1$ or $c^r \in B_2$ for $r = 1, \dots, p-1$, we will show that the edges $c^r a^{r+1}$ always exist.

Case 1: $c^r \in B_1$. As we know that $a^{r+1} \in A_1$, by definition of Q_{A_1} it follows that $c^r a^{r+1}$ is in G, since $G[A_1 \cup B_1]$ is complete bipartite by Claim 4.2.6.

Case 2: $c^r \in B_2$. Since $a^r < a^{r+1} =: a_j$ by assumption, monotonicity tells us that the neighborhood $N(a^{r+1}) \subseteq B$ ends at either c^r or to the right of c^r. Furthermore, we know $a_j b_j \in E(G)$, again since $G[A_1 \cup B_1]$ is bipartite by Claim 4.2.6. Since $b_j \in B_1$, it lies to the left of $c^r \in B_2$ so, in particular, the neighborhood $N(a^{r+1})$ starts before c^r. Monotonicity then tells us that the edge $c^r a^{r+1}$ is also present in G.

We have shown how to construct the paths P_{A_1}, P_{B_1}, and $P_{A_2 \cup B_2}$, which together clearly cover all vertices of G. We define $\phi_1(F) = P_{A_1} \cup P_{B_1} \cup P_{A_2 \cup B_2} \in \mathcal{P}_G$.

In order to see that ϕ_1 is injective, note first that if $K \in \mathcal{P}_G$ is the image of some (unknown) $F \in \mathcal{F}_G$ under ϕ_1, then one of the paths in K has all its vertices in $A_2 \cup B_2$ (we call this path $P_{A_2 \cup B_2}$), one has all its vertices from

A in A_2 and some vertices from B_1 (we call this path P_{B_1}), and we call the remaining path P_{A_1}. As described earlier, we can then easily identify the constituent paths that were glued together to form P_{A_1}, P_{B_1}, and $P_{A_2 \cup B_2}$. Finally we can complete each constituent path to a cycle to uniquely recover F. Therefore ϕ_1 is injective.

Finally, suppose $F, F' \in \mathcal{F}_G$ with $|F \triangle F'| \leq k$. In particular, there are at most k cycles that belong to one of F or F' but not both. In constructing $\phi_1(F)$ (resp. $\phi_1(F')$), we first delete one edge from each cycle of F (resp. F') to obtain a union of paths, which we call J (resp. J'). Then $|J \triangle J'| \leq k$ and there are at most k paths that belong to one of J or J' but not both. When gluing paths of J (resp. J') together to form $\phi_1(F)$ (resp. $\phi_1(F')$) there are at most $2k$ gluing edges that are used for one of J or J' but not both (at most two such edges for each differing path). This shows that $|\phi_1(F) \triangle \phi_1(F')| \leq k + 2k = 3k$, showing ϕ_1 has the desired property (taking $g(k) = 3k$). ◀

Proof of Claim 4.2.8. This claim follows immediately from Lemma 4.2.9 below. ◀

Lemma 4.2.9. Suppose $G = (V, E)$ is an n-vertex graph with $\delta(G) > n/2$. If $P_1, \ldots, P_k$ are k vertex-disjoint paths in G that together cover all vertices V, then there exists a Hamilton cycle H of G such that $E(H) \triangle E(P_1 \cup \cdots \cup P_k) \leq 3k$.

For bipartite graphs, we have the following. Suppose $G = (V, E)$ is a bipartite graph with bipartition $V = A \cup B$ with $|A| = |B| = n$ and $\delta(G) \geq n/2$. If $P_1, \ldots, P_k$ are k vertex-disjoint paths in G that together cover all vertices V, then there exists a Hamilton cycle H of G such that $E(H) \triangle E(P_1 \cup \cdots \cup P_k) \leq 3k$.

We prove the lemma for graphs; an almost identical proof works for bipartite graphs and we indicate where the proofs differ.

Proof. We will inductively modify the system of paths, at each step modifying at most 3 edges and reducing the number of paths by 1.

Let x_i and y_i be the endpoints of P_i and orient the path P_i from x_i to y_i. For any vertex x, let x^+ (resp. x^-) be the successor (resp. predecessor) of x on its path (note that these exist except possibly at the $2k$ endpoints of the paths). For any set $S \subseteq V(G)$, we define $S^+ := \{x^+ \mid s \in S\}$.

Assuming $k \geq 2$, take any two paths, say P_1 and P_2. [If G is bipartite, we choose P_2 s.t. x_1 and y_2 are in different parts, say $x_1 \in A$ and $x_2 \in B$. Note that this is always possible, renaming paths if necessary.] If x_1 is adjacent to any of $x_2, \ldots, x_k$, say to x_i, then we can reduce the number of paths by replacing P_1 and P_i by $y_1 P_1 x_1 x_i P_i y_i$ as required (only modifying one edge) and we continue. Therefore we may assume that x_1 is not adjacent to any of $x_2, \ldots, x_k$, and in particular, $|N(x_1)^-| = |N(x_1)| > n/2$. Then since $|N(y_2)| > n/2$, we must have that $N(x_1)^- \cap N(y_2)$ is non-empty. [Note that for G bipartite $N(x_1)^-, N(y_2) \subseteq A$ and therefore $N(x_1)^- \cap N(y_2)$ also holds.] Let $z \in N(x_1)^- \cap N(y_2)$ and assume $z \in V(P_i)$ for some $i = 1, \ldots, k$. If $i \neq 1, 2$ then we can replace P_1, P_2, P_i with the two paths $y_1 P_1 x_1 z^+ P_i y_i$ and $x_i P_i z y_2 P_2 x_2$, which together cover all the vertices of $V(P_1) \cup V(P_2) \cup V(P_i)$ (see Figure 20 (a)). If $i = 1$, we replace P_1, P_2 with the path $y_1 P_1 z^+ x_1 P_1 z y_2 P_2 x_2$ (see Figure 20 (b)) and if $i = 2$, we replace P_1, P_2 with $y_1 P_1 x_1 z^+ P_2 y_2 z P_2 x_2$. In all three of these cases, we delete one edge and add two (i.e. we modify three edges) and reduce the number of paths by 1.

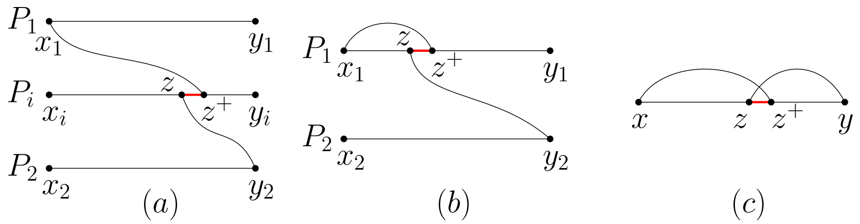


Figure 20: (a) and (b): Reducing the number of paths, cases $i \neq 1, 2$ and $i = 1$. Case $i = 2$ is similar. (c): completing the Hamilton cycle. In all cases, the thick, red edge is removed, and the curvy edges are introduced.

By iterating this, we obtain a Hamilton path P by modifying at most $3(k-1)$ edges. We can then complete this to a Hamilton cycle in the standard way. Let x and y be the endpoints of P and pick $z \in N(x)^- \cap N(y)$ (which exists as before since $|N(x)^-|, |N(y)| > n/2$). Then we obtain a Hamilton cycle $H = xPzyPz^+x$ (see Figure 20 (c)), where again we have added two edges and removed one. [In the case of G being bipartite, P has its endpoints in different parts, so that again $N(x)^-$, $N(y)$ are subsets of the same part, so again $N(x)^- \cap N(y) \neq \emptyset$.] $\square$

This completes the proof of the three claims and hence of the theorem. $\square$

4.2.6 *Concluding remarks*

It is perhaps interesting to note that, in general, it is necessary to make some kind of assumption on the minimum degree of the monotone graph for the argument in the proof of Theorem 4.2.1 to work. Without it, it is not necessarily true that the number of 2-factors is at most a polynomial factor larger than the number of Hamilton cycles of a given graph G. See the matrix and explanation below for an indication of the family of instances that illustrate this.

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Let the rows be indexed by $A = (a_1, \ldots, a_n)$ and the columns by $B = (b_1, \ldots, b_n)$. As a_1 only has two neighbors, any Hamilton cycle must contain the edges a_1b_1 and a_1b_2. This is indicated in the matrix below.

$$\begin{pmatrix} \underline{\mathbf{1}} & \underline{\mathbf{1}} & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Now, the vertex a_2 cannot also have neighbors b_1 and b_2, as this creates a cycle of length four. So we have $N(a_2) = \{b_1, b_3\}$ or $N(a_2) = \{b_2, b_3\}$; see the matrices below.

$$\begin{pmatrix} \underline{\mathbf{1}} & \underline{\mathbf{1}} & 0 & 0 & 0 & 0 \\ \underline{\mathbf{1}} & 1 & \underline{\mathbf{1}} & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \underline{\mathbf{1}} & \underline{\mathbf{1}} & 0 & 0 & 0 & 0 \\ 1 & \underline{\mathbf{1}} & \underline{\mathbf{1}} & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Note that in both the matrices above, there is now one vertex in B that has two neighbors already (and therefore cannot be chosen as neighbor in any later step). By repeating this argument, one can show that for every row $i = 2, \ldots, n-1$ there are two possible choices of extending the current Hamilton path, and so the number of Hamilton cycles equals 2^{n-2}.

However, the number of 2-factors is at least $(n/4)!$. To see this, first note that this is a lower bound on the number of Hamilton cycles in the (complete) subgraph induced by the vertices $\{a_{3n/4+1}, \ldots, a_n\}$ and $\{b_1, \ldots, b_{n/4}\}$ (assuming that n is divisible by four). It is not hard to see that any Hamilton cycle on this induced subgraph can be extended to a 2-factor of the original bipartite graph.[19]

Nevertheless, we believe that our result can be generalized to monotone graphs with minimum degree γn for any $\gamma \in (0,1)$. However, this comes at the expense of many more technicalities that (in our opinion) do not offer any additional insights. Remember that in Claim 4.2.6, we show that the nodes of G can be partitioned into two complete bipartite graphs whenever $\gamma \geq 1/2$. More generally, for a given $\gamma \in (0,1)$, it should be possible to partition the nodes of G into a constant $c = c(\gamma)$ number of complete bipartite graphs. The analogue of Claim 4.2.7 would then be to show that all cycles in a given 2-factor can be broken up, and glued together again, into a constant $d(\gamma)$ number of (vertex-disjoint) paths, after which one would need to argue that the resulting collection of paths is close, in terms of symmetric difference, to a Hamilton cycle in the monotone graph.

19 One can give a sharper bound here than $(n/4)!$, but this is not needed for our purposes.

BIBLIOGRAPHY

[1] N. Alon, Eigenvalues and expanders. *Combinatorica* 6.2 (1986), 83–96.

[2] N. Alon and J. H. Spencer, *The probabilistic method.* Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons and Inc., 4th ed. (2016).

[3] B. Alspach, D. W. Mason and N. J. Pullman, Path numbers of tournaments. *J. Comb. Theory, Ser. B* 20.3 (1976), 222–228.

[4] G. Amanatidis and P. Kleer, Rapid mixing of the switch Markov chain for strongly stable degree sequences and 2-class joint degree matrices. *Proceedings of the 2019 Annual ACM-SIAM Symposium on Discrete Algorithms* (2019) 966–985.

[5] S. Arora, D. Karger and M. Karpinski, Polynomial time approximation schemes for dense instances of *NP*-hard problems. *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing* (1995) 284–293.

[6] S. Arora, D. Karger and M. Karpinski, Polynomial time approximation schemes for dense instances of NP-hard problems. *J. Comput. System Sci.* 58.1 (1999), 193–210.

[7] E. Behrends, *Introduction to Markov chains.* Advanced Lectures in Mathematics, Friedr. Vieweg & Sohn, Braunschweig (2000).

[8] B. Bollobás, *Extremal graph theory*, *London Mathematical Society Monographs*, vol. 11. Academic Press (1978).

[9] B. Bollobás, *Random graphs*, *Cambridge Studies in Advanced Mathematics*, vol. 73. Cambridge University Press, 2nd ed. (2001).

[10] B. Bollobás and A. G. Thomason, Threshold functions. *Combinatorica* 7.1 (1987), 35–38.

[11] H. Broersma, How tough is toughness? *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* 117 (2015), 28–52.

[12] D. Christofides, P. Keevash, D. Kühn and D. Osthus, Finding Hamilton cycles in robustly expanding digraphs. *J. Graph Algorithms Appl.* 16.2 (2012), 335–358.

[13] F. Chung and X. Peng, Decomposition of random graphs into complete bipartite graphs. *SIAM J. Discrete Math.* 30.1 (2016), 296–310.

[14] W. J. Cook, *In pursuit of the traveling salesman.* Princeton University Press (2012).

[15] C. Cooper, M. E. Dyer and C. S. Greenhill, Sampling regular graphs and a peer-to-peer network. *Combin. Probab. Comput.* 16.4 (2007), 557–593.

[16] B. Csaba, M. Karpinski and P. Krysta, Approximability of dense and sparse instances of minimum 2-connectivity, TSP and path problems. *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (2002) 74–83.

[17] B. Csaba, D. Kühn, A. Lo, D. Osthus and A. Treglown, Proof of the 1-factorization and Hamilton decomposition conjectures. *Mem. Amer. Math. Soc.* 244.1154 (2016), v+164.

[18] B. Cuckler and J. Kahn, Hamiltonian cycles in Dirac graphs. *Combinatorica* 29.3 (2009), 299–326.

[19] W. F. De La Vega and M. Karpinski, On the approximation hardness of dense TSP and other path problems. *Inform. Process. Lett.* 70.2 (1999), 53–55.

[20] R. Diestel, *Graph theory*, *Graduate Texts in Mathematics*, vol. 137. Springer, 5th ed. (2018).

[21] G. A. Dirac, Some theorems on abstract graphs. *Proc. London Math. Soc. (3)* 2 (1952), 69–81.

[22] M. Dyer, A. Frieze and M. Jerrum, Approximately counting Hamilton paths and cycles in dense graphs. *SIAM J. Comput.* 27.5 (1998), 1262–1272.

[23] M. Dyer, M. Jerrum and H. Müller, On the switch Markov chain for perfect matchings. *J. ACM* 64.2 (2017), 12:1–12:33.

[24] J. Edmonds, Paths, trees, and flowers. *Canadian J. Math.* 17 (1965), 449–467.

[25] P. Erdős and R. J. Wilson, On the chromatic index of almost all graphs. *J. Combinatorial Theory Ser. B* 23.2-3 (1977), 255–257.

[26] A. Espuny Díaz, V. Patel and F. Stroh, Path decompositions of random directed graphs (2021). arXiv: 2109.13565.

[27] T. Feder, A. Guetz, M. Mihail and A. Saberi, A local switch Markov chain on given degree graphs with application in connectivity of peer-to-peer networks. *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science* (2006) 69–76.

[28] L. R. Ford and D. R. Fulkerson, Maximal flow through a network. *Canadian J. Math.* 8 (1956), 399–404.

[29] A. M. Frieze, B. Jackson, C. J. H. McDiarmid and B. Reed, Edge-colouring random graphs. *J. Combin. Theory Ser. B* 45.2 (1988), 135–149.

[30] M. R. Garey and D. S. Johnson, *Computers and intractability.* A Series of Books in the Mathematical Sciences, W. H. Freeman and Co. (1979).

[31] M. R. Garey, D. S. Johnson and R. E. Tarjan, The planar Hamiltonian circuit problem is NP-complete. *SIAM J. Comput.* 5.4 (1976), 704–714.

[32] A. Girão, B. Granet, D. Kühn, A. Lo and D. Osthus, Path decompositions of tournaments. *arXiv e-prints* (2020). arXiv: 2010.14158.

[33] S. Glock, D. Kühn and D. Osthus, Optimal path and cycle decompositions of dense quasirandom graphs. *J. Combin. Theory Ser. B* 118 (2016), 88–108.

[34] R. J. Gould, Recent advances on the Hamiltonian problem: Survey III. *Graphs Combin.* 30.1 (2014), 1–46.

[35] V. Gruslys and S. Letzter, Cycle partitions of regular graphs. *Combin. Probab. Comput.* 30.4 (2021), 526–549.

[36] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities.* Cambridge Mathematical Library, Cambridge University Press (1988).

[37] P. Haxell, M. Krivelevich and G. Kronenberg, Goldberg's conjecture is true for random multigraphs. *J. Combin. Theory Ser. B* 138 (2019), 314–349.

[38] J. van den Heuvel, The complexity of change. *Surveys in Combinatorics 2013*, *London Mathematical Society Lecture Note Series*, vol. 409, 127–160, Cambridge University Press (2013).

[39] I. Holyer, The NP-Completeness of edge-coloring. *SIAM J. Comput.* 10.4 (1981), 718–720.

[40] J. E. Hopcroft and R. M. Karp, An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput.* 2.4 (1973), 225–231.

[41] H. Huang, J. Ma, A. Shapira, B. Sudakov and R. Yuster, Large feedback arc sets, high minimum degree subgraphs, and long cycles in Eulerian digraphs. *Comb. Probab. Comput.* 22.6 (2013), 859–873.

[42] B. Jackson, Hamilton cycles in regular 2-connected graphs. *J. Combin. Theory Ser. B* 29.1 (1980), 27–46.

[43] B. Jackson, H. Li and Y. J. Zhu, Dominating cycles in regular 3-connected graphs. *Discrete Math.* 102.2 (1992), 163–176.

[44] S. Janson, T. Łuczak and A. Ruciński, *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience (2000).

[45] M. Jerrum, *Counting, sampling and integrating: algorithms and complexity.* Lectures in Mathematics ETH Zürich, Birkhäuser Verlag (2003).

[46] M. Jerrum and A. Sinclair, Fast uniform generation of regular graphs. *Theoret. Comput. Sci.* 73.1 (1990), 91–100.

[47] H. A. Jung, Longest circuits in 3-connected graphs. *Finite and infinite sets, Vol. I, II*, *Colloq. Math. Soc. János Bolyai*, vol. 37, 403–438, North-Holland, Amsterdam (1984).

[48] R. Kannan, P. Tetali and S. Vempala, Simple Markov-chain algorithms for generating bipartite graphs and tournaments. *Random Structures Algorithms* 14.4 (1999), 293–308.

[49] R. M. Karp, Reducibility among combinatorial problems. *Complexity of computer computations*, 85–103, Springer (1972).

[50] P. Kleer, V. Patel and F. Stroh, Switch-based markov chains for sampling Hamiltonian cycles in dense graphs. *Electron. J. Combin.* 27.4 (2020), Paper No. 4.29, 25.

[51] C. Knierim, M. Larcher, A. Martinsson and A. Noever, Long cycles, heavy cycles and cycle decompositions in digraphs. *J. Comb. Theory, Ser. B* 148 (2021), 125–148.

[52] J. Komlós and E. Szemerédi, Limit distribution for the existence of hamiltonian cycles in a random graph. *Discrete Math.* 43.1 (1983), 55–63.

[53] M. Krivelevich, Triangle factors in random graphs. *Comb. Probab. Comput.* 6.3 (1997), 337–347.

[54] D. Kühn, A. Lo, D. Osthus and K. Staden, The robust component structure of dense regular graphs and applications. *Proc. Lond. Math. Soc. (3)* 110.1 (2015), 19–56.

[55] D. Kühn, A. Lo, D. Osthus and K. Staden, Solution to a problem of Bollobás and Häggkvist on Hamilton cycles in regular graphs. *J. Combin. Theory Ser. B* 121 (2016), 85–145.

[56] D. Kühn, R. Mycroft and D. Osthus, A proof of Sumner's universal tournament conjecture for large tournaments. *Proc. Lond. Math. Soc. (3)* 102.4 (2011), 731–766.

[57] D. Kühn and D. Osthus, Hamilton decompositions of regular expanders: a proof of Kelly's conjecture for large tournaments. *Adv. Math.* 237 (2013), 62–146.

[58] D. Kühn and D. Osthus, Hamilton decompositions of regular expanders: applications. *J. Comb. Theory, Ser. B* 104 (2014), 1–27.

[59] D. Kühn and D. Osthus, A survey on Hamilton cycles in directed graphs. *European J. Combin.* 33.5 (2012), 750–766. EuroComb '09.

[60] I. Lignos, Reconfigurations of combinatorial problems: graph colouring and Hamilton cycle. *PhD thesis* (2017).

[61] S. Lin and B. W. Kernighan, An effective heuristic algorithm for the traveling-salesman problem. *Operations Res.* 21.2 (1973), 498–516.

[62] A. Lo, V. Patel, J. Skokan and J. Talbot, Decomposing tournaments into paths. *Proc. Lond. Math. Soc. (3)* 121.2 (2020), 426–461.

[63] I. Miklós, P. L. Erdős and L. Soukup, Towards random uniform sampling of bipartite graphs with given degree sequence. *Electron. J. Combin.* 20.1 (2013), Paper 16, 51.

[64] J. W. Moon, *Topics on tournaments.* Holt, Rinehart and Winston (1968).

[65] N. Nishimura, Introduction to reconfiguration. *Algorithms (Basel)* 11.4 (2018).

[66] V. Patel and F. Stroh, A polynomial-time algorithm to determine (almost) Hamiltonicity of dense regular graphs (2020). arXiv: 2007.14502.

[67] V. Rödl, A. Ruciński and E. Szemerédi, A Dirac-type theorem for 3-uniform hypergraphs. *Comb. Probab. Comput.* 15.1-2 (2006), 229–251.

[68] G. N. Sárközy, S. M. Selkow and E. Szemerédi, On the number of Hamiltonian cycles in Dirac graphs. *Discrete Math.* 265.1-3 (2003), 237–250.

[69] A. Sinclair, Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.* 1.4 (1992), 351–370.

[70] J. Spinrad, A. Brandstädt and L. Stewart, Bipartite permutation graphs. *Discrete Appl. Math.* 18.3 (1987), 279–292.

[71] A. Takaoka, Complexity of Hamiltonian cycle reconfiguration. *Algorithms (Basel)* 11.9 (2018).

[72] R. Taylor, Constrained switchings in graphs. *Combinatorial Mathematics, VIII*, *Lecture notes in Math.*, vol. 884, 314–336, Springer (1981).

[73] L. Trevisan, Max cut and the smallest eigenvalue. *SIAM J. Comput.* 41.6 (2012), 1769–1786.

[74] L. Trevisan, Lecture Notes CS294 (2016).
URL https://lucatrevisan.wordpress.com/2016/02/09/cheeger-type-inequalities-for-λn/.

[75] V. G. Vizing, On an estimate of the chromatic class of a p-graph. *Discret Analiz* 3 (1964), 25–30.

[76] T. de Vos, *Decomposing directed graphs into paths.* Master's thesis, Universiteit van Amsterdam (2020).

SUMMARY

Hamilton cycles and algorithms

This thesis presents three results in graph theory, united by the themes of Hamilton cycles and algorithms. A Hamilton cycle in a graph is a cycle that contains every vertex of the graph. The first chapter introduces the most important concepts and gives an overview of the main results. Chapters 2, 3 and 4 each concern a separate topic and can be read in any order.

Chapter 2 considers path decompositions of digraphs, specifically an extension of a conjecture due to Alspach, Mason, and Pullman. There is a natural lower bound for the number paths needed in an edge decomposition of a directed graph in terms of its degree sequence; the conjecture in question states that this bound is correct for tournaments of even order. (This conjecture is actually a vast generalization of a conjecture due to Kelly that states that every regular tournament can be decomposed into Hamilton cycles.) The conjecture of Alspach, Mason, and Pullman was recently resolved for large tournaments, and here we investigate to what extent the conjecture holds for directed graphs in general. In particular, we prove that the conjecture asymptotically almost surely holds for the random directed graph $D_{n,p}$ for a large range of p. The proof consists of two parts: in the first we show that the conjecture holds for directed graphs satisfying certain (deterministic) properties, and in the second part we show that the random directed graph satisfies these properties asymptotically almost surely for our range of p.

In Chapter 3 we give a polynomial-time algorithm for detecting almost-Hamilton cycles in dense regular graphs. Specifically, we show that, given $\alpha \in (0, 1)$, there exists a $c = c(\alpha)$ such that the following holds: there is a polynomial-time algorithm that, given a D-regular graph G on n vertices with $D \geq \alpha n$, determines whether G contains a cycle on at least $n - c$ vertices. If such a cycle exists, we give a (randomized) polynomial-time algorithm to find it. The problem becomes NP-complete if we drop either the density or the regularity condition. The algorithm uses spectral partitioning to construct a robust expander decomposition, a structure introduced by Kühn and Osthus, as well as some further algorithmic ingredients.

In Chapter 4, we consider switch-based Markov chains for the approximate uniform sampling of Hamiltonian cycles in graphs of high minimum degree. These are Markov chains on the space of all Hamilton cycles of a given graph, where transitions are between Hamilton cycles that differ on a bounded number k of edges (such a transition is called a k switch). As our main result, we show that every pair of Hamiltonian cycles in a graph with minimum degree at least $n/2+7$ can be transformed into each other by 10-switches, implying that the 10-switch Markov chain is irreducible on such graphs. We show that $n/2+7$ cannot be significantly reduced in this result. Using a strengthening of our irreducibility result, we prove that the 10-switch Markov chain is rapidly mixing (i.e. converges quickly to its stationary distribution) on the class of dense monotone graphs.

SAMENVATTING

Hamiltoncircuits en algoritmes

Dit proefschrift presenteert drie resultaten in de grafentheorie, verbonden door twee gemeenschappelijke thema's: Hamiltoncircuits en algoritmen. Een Hamiltoncircuit in een graaf is een circuit dat elk punt van de graaf bevat. Het eerste hoofdstuk introduceert de belangrijkste concepten en geeft een overzicht van de belangrijkste resultaten. Hoofdstukken 2, 3 en 4 hebben elk een apart onderwerp en kunnen in willekeurige volgorde worden gelezen.

Hoofdstuk 2 behandelt paddecomposities van gerichte grafen, in het bijzonder een uitbreiding van een vermoeden van Alspach, Mason en Pullman. Er is een natuurlijke ondergrens voor het aantal paden dat nodig is in een kantdecompositie van een gerichte graaf in termen van de graadrij; het vermoeden in kwestie stelt dat deze grens correct is voor toernooien van even orde. (Dit vermoeden is eigenlijk een uitgebreide generalisatie van een vermoeden van Kelly dat stelt dat elk regulier toernooi kan worden opgedeeld in Hamiltoncircuits.) Het vermoeden van Alspach, Mason en Pullman is onlangs opgelost voor grote toernooien, en hier onderzoeken we in welke mate het vermoeden geldt voor algemene gerichte grafen. In het bijzonder bewijzen we dat het vermoeden asymptotisch vrijwel zeker geldt voor de willekeurig gerichte graaf $D_{n,p}$ voor een groot bereik van p. Het bewijs bestaat uit twee delen: in het eerste laten we zien dat het vermoeden geldt voor gerichte grafen die aan bepaalde (deterministische) eigenschappen voldoen, en in het tweede deel laten we zien dat de willekeurige gerichte graaf vrijwel zeker asymptotisch aan deze eigenschappen voldoet voor ons bereik van p.

In Hoofdstuk 3 geven we een polynomiale tijd algoritme voor het detecteren van bijna-Hamiltoncircuits in dichte reguliere grafen. Concreet laten we zien dat, gegeven $\alpha \in (0,1)$, er een $c = c(\alpha)$ bestaat zodat het volgende geldt: er is een polynomiale tijd algoritme dat, gegeven een D-reguliere graaf G op n punten met $D \geq \alpha n$, bepaalt of G een circuit bevat op minimaal $n - c$ punten. Als zo'n circuit bestaat, geven we een (gerandomiseerd) polynomiale tijd algoritme om het te vinden. Het probleem wordt NP-compleet als we ofwel de dichtheids- ofwel de regulariteitsvoorwaarde laten vallen. Het algoritme maakt gebruik van spectrale partitionering om een robuuste expander-decompositie te construeren, een structuur geïntroduceerd door Kühn en Osthus, evenals enkele andere algoritmische ingrediënten.

In Hoofdstuk 4 beschouwen we Markovketens gebaseerd op 'switches' voor het bij benadering uniform trekking van Hamiltoncircuits in grafen met een hoge minimumgraad. Dit zijn Markovketens, gedefinieerd op de ruimte van alle Hamiltoncircuits van een gegeven graaf, waarbij een overgang (met positieve kans) tussen twee Hamiltoncircuits mogelijk is indien zij verschillen op een begrensd aantal k kanten. (een dergelijke overgang wordt een k-switch genoemd). Als ons belangrijkste resultaat laten we zien dat elk paar Hamiltoncircuits in een graaf met een minimale graad van ten minste $n/2 + 7$ in elkaar kan worden omgezet door 10-switches, wat impliceert dat de 10-switch Markov-keten irreducibel is voor dergelijke grafen. We laten zien dat $n/2 + 7$ niet significant gereduceerd kan worden in dit resultaat. Gebruikmakend van een versterking van ons irreducibiliteitsresultaat, bewijzen we dat de 10-switch Markovketen snel convergeert naar zijn stationaire verdeling in de klasse van dichte monotone grafen.

ACKNOWLEDGMENTS

I am thankful to the people that supported me during my time at the KdVI. Without their help, this thesis would not exist.

Viresh, thank you for accepting me as a PhD candidate. You took on the role of my supervisor and helped me grow as a mathematician. You supported me a great deal in making this thesis what it is. Your door is always open when I want to talk and I always feel better after talking to you. I enjoyed our discussions about puzzles. Thank you for all the trust, guidance and encouragement that you gave me and for all you have done.

Jo, thank you for being very helpful and supportive even though circumstances sadly made that cumbersome.

Pieter and Alberto, collaborating on our problems was fun and interesting. I enjoyed working together and I am proud of our results. Thank you.

Thank you Maria and Torsten, for nurturing my interest in graph theory during my Bachelor's and Master's studies and for supporting me in finding a PhD position.

Thank you Lorenzo and David for being my paranymphs with everything that entails. Our friendship means a lot to me.

Thank you to everyone at the KdVI for fun and interesting discussions over lunch, the great seminars, the time we got to spend together, and the board game nights. I wish we could have seen each other more over the past two years. Evelien, Marieke and Tiny, thank you for everything that you do.

I would like to thank everyone at NETWORKS for the interesting talks, lectures, and conversations. I much enjoyed the time we spent together at the events.

Wessel, Reinier, Madelon, Lorenzo, Kayed: Thank you for the shared evenings at the institute, the adventures, the laughs, and the friends we made along the way.

Maran, Betty, Marin, thanks for making my time in Amsterdam great. I enjoy spending time with you.

Kaspar, Ruben, Ruben, Tristan, Eisi, Bero, Larisa, Richard, Babsi, Marten, Marcel, Fabian: Vielen Dank für die vielen schönen Abende. Ihr habt nicht nur meine letzten vier Jahre bereichert, sondern auch vorher und offline. Ich hoffe, dass wir noch viele weitere Abende miteinander verbringen werden.

Simon, du bringst mich dazu, hin und wieder Urlaub zu machen, danke dafür. Vielen Dank fürs Korrekturlesen. Simon und Laura, vielen Dank für eure Gastfreundschaft und für die Zeit, die wir miteinander verbringen.

Mama und Papa: Ihr seid immer für mich da, ihr unterstützt mich in allen Angelegenheiten, ihr nehmt die größten Strapazen für mich auf. Ich habe euch regelmäßig als die besten Eltern der Welt vorgestellt, und ich glaube fest, dass ich nicht übertreibe. Ich könnte mir keine besseren Eltern wünschen. Ihr gebt mir Wurzeln und Flügel. Danke für eure Liebe und Unterstützung.