Decomposing cubic graphs into isomorphic linear forests

Gal Kronenberg[*]   Shoham Letzter[†]   Alexey Pokrovskiy[‡]   Liana Yepremyan[§]

Abstract

A common problem in graph colouring seeks to decompose the edge set of a given graph into few similar and simple subgraphs, under certain divisibility conditions. In 1987 Wormald conjectured that the edges of every cubic graph on $4n$ vertices can be partitioned into two isomorphic linear forests. We prove this conjecture for large connected cubic graphs. Our proof uses a wide range of probabilistic tools in conjunction with intricate structural analysis, and introduces a variety of local recolouring techniques.

1   Introduction

Many problems in graph theory seek to decompose the edges of a given graph into simpler pieces. A fundamental example seeks for a decomposition of the edges into matchings, that is, a proper edge-colouring of a graph. According to a well-known result of Vizing [32] from 1964, the *chromatic index* of a graph G, denoted $\chi'(G)$ and defined to be the minimum number of matchings needed to decompose the edges of a simple graph G, is either $\Delta(G)$ or $\Delta(G) + 1$. For certain applications, $\Delta(G)$ is too large, and thus there was interest in finding decompositions into a significantly smaller number of pieces, which are still relatively simple. Perhaps the most famous work in this direction is the Nash-Williams theorem [28] from 1964 regarding the *arboricity* of a graph G, namely the minimum number of forests needed in order to decompose the edges of G. As an interesting interpolation between decompositions into matchings and into forests, in 1970 Harary [20] suggested to study the minimum number of linear forests needed to decompose the edges of a given graph G, where a *linear forest* is a forest whose components are paths. This parameter is called the *linear arboricity* of G and is denoted $\mathrm{la}(G)$. Clearly, $\mathrm{la}(G) \leq \chi'(G)$ for every graph G, as a matching is also a linear forest. A well known conjecture by Akiyama, Exoo, and Harary from 1980 [4] states that by using linear forests instead of matchings, one can reduce the number of pieces needed by about a half, namely, that $\mathrm{la}(G) \leq \lceil \frac{\Delta(G)+1}{2} \rceil$. This conjecture is known in the literature as the *linear arboricity conjecture*. It is easy to check that the upper bound is tight. Also,

---

since every graph of maximum degree Δ can be embedded in a Δ-regular graph, an equivalent statement of the conjecture, which is more commonly studied, is the following.

Conjecture 1.1 (Linear arboricity conjecture; Akiyama, Exoo, and Harary [4])**.** *Every Δ-regular graph G satisfies* $\mathrm{la}(G) = \lceil \frac{\Delta+1}{2} \rceil$.

The case of cubic graphs (3-regular graphs), where the conjecture predicts $\mathrm{la}(G) = 2$, was verified already in the paper of Akiyama, Exoo, and Harary [4], and later a shorter proof was presented by Akiyama and Chvátal [3] in 1981. In 1988 Alon [7] showed that the linear arboricity conjecture holds asymptotically, meaning that if G has maximum degree at most Δ then $\mathrm{la}(G) = \Delta/2 + o(\Delta)$; his $o(\Delta)$ term was of order $O\left(\Delta \frac{\log \log \Delta}{\log \Delta}\right)$. In the same paper Alon also showed that the conjecture holds for graphs with high girth, that is, when the girth of the graph is $\Omega(\Delta)$. Subsequently, using Alon's approach, the linear arboricity conjecture was proved to hold for random regular graphs by Reed and McDiarmid [26] in 1990. Alon and Spencer [8] in 1992 improved Alon's asymptotic error term to $O(\Delta^{2/3}(\log \Delta)^{1/3})$. After almost thirty years, this error term was very recently improved; at first to $O(\Delta^{2/3-\varepsilon})$, for some absolute $\varepsilon > 0$ by Ferber, Fox and Jain [15], and finally to $O(\sqrt{\Delta}(\log \Delta)^4)$ by Lang and Postle [23]. All of these approximate results make use of classical probabilistic tools, specifically, Rödl's nibble and Lovász's local lemma. The result of Lang and Postle also apply to yet another folklore conjecture in graph theory, known as the *list colouring conjecture*, tying the two conjectures together. This conjecture asserts that if L assigns a list of $\Delta(G) + 1$ colours to each edge of G, then there is a proper edge-colouring of G where each edge uses a colour from its list. This conjecture emerged in 1970-80's, and was researched by prominent mathematicians, including Kahn [22] and Molloy and Reid [27].

Apart from approximate results and random regular graphs, the linear arboricity conjecture was verified for $d \in \{3, 4\}$ [3, 4, 5], $d \in \{5, 6\}$ [13, 29, 31], $d = 8$ [14], $d = 10$ [19], as well as for various other families of graphs such as complete graphs, complete bipartite graphs, trees, planar graphs, some k-degenerate graphs and binomial random graphs [4, 5, 11, 14, 18, 19, 34, 35]. However, in its full generality the conjecture still remains open.

As we have seen, the linear arboricity conjecture is known to hold for cubic graphs. It was thus asked whether it can be strengthened for such graphs. One such strengthening is the following conjecture, made by Wormald [33] in 1987, asking not only for a decomposition into two linear forests, but also requires them to be "balanced".

Conjecture 1.2 (Wormald)**.** *The edges of any cubic graph, whose number of vertices is divisible by 4, can be 2-edge-coloured such that the two colour classes are isomorphic linear forests.*

Prior to our work, Wormald's Conjecture was known to be true only for some very specific cubic graphs. It was proved for Jaeger graphs in work of Bermond, Fouquet, Habib and Péroche [16] and Wormald [33], and for some further classes of cubic graphs by Fouquet, Thuillier, Vanherpe and Wojda [17].

In our current work, we essentially settle this conjecture.

Theorem 1.3. *Let G be a connected cubic graph on n vertices, where n is large and divisible by 4. Then there is a red-blue colouring of the edges of G whose colour classes span isomorphic linear forests.*

We note that our proof can be modified to relax the requirement of G being connected to G having at least one large connected component, meaning a component of size at least a certain (large) universal

constant. This strengthening is described in the concluding section, Section 9. In addition, our proof can be modified to show that if the number of vertices is not divisible by 4 then we can guarantee that the two linear forests are isomorphic up to removing or adding an edge. Specifically, we can guarantee that the only difference in component structure is that the red graph has two extra edge components, and the blue graph has one extra component which is a path of length 3.

It is worth mentioning that, although our proof of Conjecture 1.2 is quite long and involved, we can prove the following approximate version of Warmald's conjecture with a fairly short and neat proof (note that here we do not require divisibility or connectivity).

Theorem 1.4. *Let G be a cubic graph on n vertices, where n is large. Then G can be red-blue coloured so that all monochromatic components are paths of length $O(\log n)$, and the numbers of red and blue components which paths of length t differ by at most $n^{2/3}$, for every $t \geq 1$.*

This approximate version is proved in Section 3 (this is a special case of Lemma 3.3), where the rest of the paper (from Section 4 onwards) is dedicated to proving Theorem 1.3 using Lemma 3.3.

We sketch the proof of Theorem 1.3 in great detail in the next section, but here is a quick summary. The starting point of the proof of Theorem 1.3 is Theorem 1.4, and as such, we now briefly sketch the proof of the latter, approximate theorem. One natural way to split a cubic graph into almost isomorphic parts is to colour each edge either red or blue uniformly at random and independently. However, the two colour classes will have many vertices of degree 3, and will thus be far from being a linear forest. To get a colouring which is both balanced with high probability, and whose colour classes are "close" to being linear forests, instead of using the random colouring described above, we use a semi-random colouring scheme whose colour classes have maximum degree at most 2. Before describing how we do so, we point out that this is not the end of the road; we still need to eliminate monochromatic cycles and long monochromatic paths. Doing so requires some clever tricks, which achieve these goals using local steps, without harming the nice properties we have achieved, and without introducing too many dependencies.

For the semi-random colouring, we start with a 2-colouring of G where each monochromatic component is a path, and recolour each path randomly and independently, reminiscently of Kempe-changes. For technical reasons (related to the concentration inequality which we apply at the end), we need every monochromatic component in the initial colouring to be a *short* path. This can be achieved through a well-known result of Thomassen [30].

In 1984, Bermond, Fouquet, Habib, and Péroche [10] conjectured that not only can every cubic graph be decomposed into two linear forests, but it can be done in such a way that every path in each of the two linear forests has length at most 5. In 1996 Jackson and Wormald [21] proved this conjecture with the constant 18 instead of 5. Later this was improved by Aldred and Wormald [6] to 9, and the conjecture was finally resolved in 1999 by Thomassen [30].

Theorem 1.5 (Thomassen). *Any cubic graph can be 2-edge-coloured such that every monochromatic component is a path of length at most five.*

We will, in fact, prove a more general statement than the one in Theorem 1.4 (see Lemma 3.3) that is applicable also for cubic graph with a given partial colouring with nice properties. This will allow us to pre-colour small parts of the graph in advance and obtain an "almost balanced" colouring which preserves large parts of the pre-colouring. The pre-coloured subgraph will contain many "gadgets",

3

which are small subgraphs with two colourings with the property that, by replacing one colouring by the other, the difference between the number of red and blue components which are paths of a given length ℓ decreases by 1, and the numbers of longer monochromatic components does not change. Since Lemma 3.3 guarantees that many gadgets for each relevant ℓ survive, we can use them straightforwardly to correct the imbalance between red and blue component counts.[1]

Finding gadgets that interact well with the rest of the graph is a pretty subtle process, as we need to consider not only the various different forms the structure of the gadget itself can take, but we also need to consider its neighbourhood, e.g. to make sure that its colouring can be extended to a colouring of the whole graph where monochromatic components are paths. This is particularly challenging when the graph contains many short cycles and to overcome this we need sophisticated ways of tracking a certain colouring process around a geodesic (see Section 8).

We believe that the tools we developed here could provide a new avenue for progress towards the Linear Arboricity Conjecture.

We will end this section with a short discussion on some related works.

One may wonder about possible vertex analogues of the problems mentioned here. Indeed, this has been considered before. In 1990 Ando conjectured that the vertices of any cubic graph can be two-coloured such that the two colour classes induce isomorphic subgraphs. Ando's conjecture was first mentioned in the paper of Abreu, Goedgebeur, Labbate and Mazzuoccolo [2] where they made an even stronger conjecture, adding the requirement that the two colour classes induce linear forests. Recently, Das, Pokrovskiy and Sudakov [12] proved Ando's conjecture for large connected cubic graphs. In fact, their proof also verifies the stronger conjecture for cubic graphs of large girth.

Ban and Linial [9] stated an even stronger conjecture (but under some further restrictions on G): the vertices of every bridgeless cubic graph, except for the Petersen graph, can be 2-vertex-coloured such that the two colour classes induce isomorphic matchings. This conjecture was proved for some specific cubic graphs (see [1, 9]), but it is still widely open in general. We discussed this in further details in Section 9.

2 Proof overview

We now give a detailed sketch of our proof. The main idea is to first colour a small part of the graph in a very structured way, so that it can later be used to make small fixes to the full colouring, and then colour the rest of the graph in a random way, while guaranteeing that the monochromatic components are (not too long) paths. Using the randomness, we show that the two colour classes are almost isomorphic. We then use the pre-coloured graph to fix the imbalance and make the colour classes isomorphic, thus completing the proof.

The structure of the paper will be as follows. In Section 3 we will state and prove Lemma 3.3 about the existence of an almost-balanced colouring. In section Section 4 we will state Lemma 4.2 about the existence of a good partial colouring, and in Section 4.1 will show how to use Lemma 3.3 and Lemma 4.2 in order to prove Theorem 1.3. In Sections 5 to 8 we will prove Lemma 4.2.

---

[1]Actually, since we can only guarantee gadgets of length $o(\log n)$, and the components resulting from Lemma 3.3 can be longer, we need a separate argument to balance intermediate lengths.

2.1 Notation

Given a graph G and an edge-colouring χ of G, will denote by $b_{G,\chi}(P_t)$ and $r_{G,\chi}(P_t)$ the number of blue and red paths of length t in G. For two paths P, Q in a graph G, we denote by $d_G(P, Q)$ the length of the shortest path between them. When G or χ are clear from the context, we will skip the corresponding subscript. In a graph G, a *geodesic* is defined to be the shortest path between some two vertices. Given an edge-coloured graph G, we call a vertex *monochromatic* if all edges incident to it have the same colour. In this paper $\log n$ is the natural logarithm. I am actually not sure think it is of base 2. In a graph G, we say a subgraph $H \subseteq G$ *touches* an edge e if e has exactly one endpoint in H.

2.2 The approximate result

While this is not the first step in the process, we now describe an approximate solution of Wormald's conjecture, and later explain how to obtain an appropriate partial pre-colouring. For the purpose of this explanation, our task is to red-blue colour a (large) given cubic graph such that the colour classes are "almost" isomorphic, that is, the difference between the number of red and blue components isomorphic to a path of length t is small, for all t. For this, we wish to colour the graph randomly, while maintaining certain properties such as the monochromatic components being paths.

Our random colouring will consist of three steps. For the first step, we use Thomassen's result (Theorem 1.5) about the existence of a 2-colouring where each monochromatic component is a path of length at most 5[2]; we denote the two colours here by purple and green. The first random step colours each purple or green component by one of the two possible alternating red-blue colourings, chosen uniformly at random and independently. Notice that this random red-blue colouring of G has no *monochromatic* vertices meaning a vertex who is incident to three edges of the same colour. Moreover, the symmetry between the colours and the bound on the lengths of purple and green paths would allow us to show that the colours are, in some sense, close to being isomorphic. However, there is nothing preventing the appearance of cycles, and we could not rule out the existence of very long monochromatic paths. This is the more technical part for applying concentration inequalities and doing the final re-balancing.

This brings us to the second random step, which will be broken into two parts, and whose purpose is to eliminate monochromatic cycles. Here, we first do something very intuitive: we simply flip the colour of one edge e_C of each monochromatic cycle C, choosing the edge e_C uniformly at random and independently. Unsurprisingly, while this breaks all monochromatic cycles that existed before the first step, new monochromatic cycles can appear. Luckily, a small fix saves us and eliminates all monochromatic cycles. The fix essentially consists of re-swapping the colour of e_C for some of the originally monochromatic cycles C, and swapping the colour of one of the two neighbouring edges of e_C in C, while again making choices randomly and independently.

The next and final random step is designed to break "long" paths. After this process monochromatic paths have length of order $O(\log n)$. Here the idea is less intuitive. We let each monochromatic path P choose one of the possibly four *boundary edges*, that is the edges of the opposite colour that touch an end of P uniformly at random and independently. Then, for each edge e that was chosen by two paths, we flip the colour of e with probability $1/2$, independently. This somewhat strange process has several benefits: first, with high probability, it swaps an edge of each monochromatic path of length at

---

[2]The exact constant here is not important. We could even make do with a polylogarithmic bound on the lengths, and, additionally, we can allow for even cycles, but not odd ones.

least $1000 \log n$; second, it creates no monochromatic cycles; and, third, it does not allow more than two monochromatic paths to join up (more precisely, monochromatic paths in the new colouring have at most one edge whose colour was swapped).

Finally, we analyse the resulting random colouring, and show that its colour classes are almost isomorphic. The proof of this is conceptually simple: observing that the distribution of the red and blue graphs is identical, thanks to all steps being performed simultaneously for red and blue, all we need to do is to prove that the number of red components isomorphic to P_t is concentrated, for every $t \geq 1$. We accomplish this goal via McDiarmid's inequality (Theorem 3.1), using the independence of the various random decisions, as well as the fact that each decision has a small impact on the resulting graph.

2.3 From approximate to exact result

Going from the approximate result to the exact result we wish to balance the number of red and blue components isomorphic to P_t, for every t. For this, the main idea is to pre-colour a small part of the graph, thereby creating many *gadgets* that can later be used for balancing the number of paths in each length. This is done before finding the approximate partition.

We define a blue ℓ-*gadget* in a cubic graph G to be a subgraph $H \subseteq G$ together with two red-blue edge-colourings χ, χ' such that for every red-blue colouring of G that extends (H, χ) and whose monochromatic components are paths, the monochromatic component counts change as follows if we switch the colouring of H from χ to χ': $b_{\chi',G}(P_\ell) = b_{\chi,G}(P_\ell) - 1$; $b_{\chi,G}(P_t) = b_{\chi',G}(P_t)$ with $t > \ell$; $b_{\chi,G}(P_t)$ with $t < \ell$ changes only slightly in χ'; and $r(P_t)$ with any t does not change. Such a gadget (and its counterpart with roles of colours reversed) will be used to equalise $r(P_t)$ and $b(P_t)$ iteratively, and we balance the path counts from the longest to the shortest, ensuring that the process terminates. Due to a simple counting argument, we only need to run the process until $\ell = 3$.

The most difficult and lengthy part of this paper is to find gadgets. We will not describe here the concrete structure of the gadgets we shall find (this can be found in Section 6), but let us just say that these gadgets consist of a long path along with some pending edges and paths. As such, it makes sense to look for gadgets in the neighbourhood of a "long" geodesic. This is indeed what we do in Sections 7 and 8. We start off with a geodesic P, whose length is sufficiently larger than ℓ, and then colour it and its neighbourhood appropriately. The difficulty is that, in addition to obtaining the gadget structure, we also need to make sure that other desirable properties are maintained, an obvious one being the non-existence of monochromatic vertices. For this, we define a colouring algorithm that colours the geodesic and some ball around it to guarantee this property. The fact that we started from a geodesic (rather than any arbitrary path) will allow us to claim that our colouring algorithm ends after a small number of steps while maintaining the desirable properties. The main obstruction to obtaining these desired property are short cycles. Indeed, our proof when no two vertices in the geodesic P have a common neighbour outside of P is much simpler (we treat it separately in Section 7), and this property holds automatically if the girth of G is at least 5. If the girth is required to be at least 7, we get an even simpler proof; indeed, with this girth assumption only the first case in Section 7 can occur. For the rest of this overview, let us assume that we know how to find a gadget (with some desirable properties) in a small ball around a geodesic of length $c \cdot \ell$, for some constant c.

There are three things to keep in mind now. First, we need to find many well-separated geodesics of length $c \cdot \ell$, for all $\ell = 3$ up to $\ell = O(\sqrt{\log n})$. Second, our strategy for the approximate result should be

adapted so as to allow the underlying graph being partially pre-coloured because of the gadgets. And, third, we need many of the gadgets to survive at the end of the random process that we run to prove the approximate result.

For the first part, it is easy enough to show in any n-vertex cubic graph there are $n^{1+o(1)}$ gadgets of length L that are far enough from each other, for any $L = o(\log n)$. This already presents a small hurdle: our strategy for the approximate result yields colourings whose monochromatic components are paths of length at most $1000 \log n$, a bound which is too large for our gadgets to be applicable for balancing. Luckily, we can balance paths of length at least $c\sqrt{\log n}$, for some constant c, using a separate argument, not involving gadgets. Suppose there are x many more blue paths of length t than red ones, we let $\mathcal{I}$ to be the collection of these x blue P_t-paths and break each one of them into segments of length $\Theta(\sqrt{\log n})$; denote the collection of segments by $\mathcal{I}$. Now, using the Lovász Local Lemma, we show that there is a choice of an edge e_I from each segment $I \in \mathcal{I}$, such that if the colours of these edges are flipped, then we create no cycles and no new monochromatic paths of length at least $c'\sqrt{\log n}$, for some appropriate constant c'.

Regarding the second point, luckily the strategy we described in Section 2.2 is amenable to the underlying graph being partially pre-coloured, as long as the following hold:

- vertices incident to two pre-coloured edges see edges of both colours,

- the coloured components in the pre-colouring are quite small,

- and a version of Thomassen's result holds for the graph of uncoloured edges.

The first property is the main property that we address when building gadgets. The second one holds naturally since we only colour edges in small neighbourhoods of geodesics of length $O(\sqrt{\log n})$. The third requires an additional argument, which can be found in Section 5. With these properties at hand, the only modification of the process described in Section 2.2 is in the first random step. Now we have some short purple and green paths, but also various small red-blue coloured components. We colour the purple and green components as before, and for each coloured component, we either keep it as is, or swap the colours of all its edges with the decision being made uniformly at random and independently. All other steps are carried out exactly as before. Notice in particular that, importantly, for any graph H, the red and blue copies of H will have the same distribution.

Finally, for the third point it suffices to make sure that each gadget survives the whole random process intact, with decent probability. This is easy to check for the first and third steps, making the second step the only real obstruction. Recall that in the second step, for each monochromatic cycle C, we choose an edge e_C in it, uniformly at random, and at the end of the second step either it or one of its neighbouring edges in C has its colour swapped (and all other edges remain as they were). Thus, to guarantee that a gadget H has a decent survival chance, we need to make sure that every monochromatic cycle C in some red-blue colouring of G extending H has two consecutive edges not in H. This property is quite annoying to deal with; having to maintain it influences the arguments in Sections 5, 7 and 8.

3 Proof of approximate version

In this section we prove a weaker version of Conjecture 1.2; see Lemma 3.3 below. Roughly speaking, we show, using a probabilistic approach, that the edges of the graph can be 2-coloured such that each

monochromatic component is a linear forest, and the two graphs are "almost isomorphic".

For the proof we will use McDiarmid's Inequality to show concentration.

Theorem 3.1 (McDiarmid's Inequality [25]). *Let $X = (X_1, \ldots, X_n)$ be a family of independent random variables, with X_i taking values in a set A_i, for $i \in [n]$. Suppose that $c > 0$ and f is a real-valued function defined on $\prod_{i \in [n]} A_i$ that satisfies $|f(x) - f(x')| \leq c$ whenever x and x' differ in a single coordinate. Then, for any $m \geq 0$,*

$$\mathbb{P}\left(f(X) - \mathbb{E}[f(X)] \geq m\right) \leq 2 \exp\left(\frac{-2m^2}{c^2 n}\right).$$

In order to use the "almost balanced" colouring and turn into a decomposition into two isomorphic linear forests, we will pre-colour a small part of the graph that will be used later to balance the colouring. However, we need to make sure that the pre-coloured graph will combine nicely with the almost balanced colouring. For this purpose we define the following.

Definition 3.2. Let G be a cubic graph of order n. A subgraph $G_0 \subseteq G$ whose edges are red-blue coloured is called *extendable* if all of the following hold. Let $G_0(R)$ and $G_0(B)$ be the red and blue subgraphs of G_0, respectively.

(E1) Every vertex v with $d_{G_0}(v) \geq 2$ satisfies $d_{G_0(R)}(v), d_{G_0(B)}(v) \geq 1$.

(E2) The size of each connected component in G_0 is $O(\sqrt{\log n})$ and every two components are of distance at least 10 from each other.

(E3) Every cycle C in G which does not contain both a red and a blue edge from the same component of G_0 has two consecutive edges outside of G_0.

(E4) The graph $G \setminus G_0$ has an edge-decomposition into graphs H_1, H_2, H_3 such that H_2 and H_3 are vertex disjoint and have the following properties.

- H_1 is a subdivision of a simple graph F all of whose vertices have degrees 1 or 3, where each edge in F is subdivided $O(\sqrt{\log n})$ times to create H_1.

- H_2 is a disjoint union of even cycles of length $O(\sqrt{\log n})$, each cycle touches at most two edges in H_1.

- H_3 is vertex-disjoint collection of subdivisions of multi-edges with multiplicity 3, where each edge is subdivided at most $O(\sqrt{\log n})$ times.

We note that $G_0 = \emptyset$ is always extendable. The purpose of this section is to prove the following lemma, showing that any extendable partial pre-colouring G_0 in a cubic graph G can be extended to a red-blue colouring of G where the two colour classes are almost isomorphic linear forests.

Lemma 3.3. *Given a cubic graph G of order n, where n is sufficiently large, and a red-blue coloured subgraph G_0 which is extendable according to Definition 3.2, there exists a red-blue colouring of G such that.*

(Q1) *Every monochromatic component is a path of length $O(\log n)$.*

(Q2) *There is no red path of length at most $(\log n)^{2/3}$ that intersects at least $100\sqrt{\log n}$ blue paths of length at least $100\sqrt{\log n}$, and similarly with the roles of red and blue reversed.*

(Q3) $|r(P_t) - b(P_t)| \leq n^{2/3}$ for every $t \geq 1$.

(Q4) *For every red-blue coloured subcubic graph H, if G_0 has at least $n^{0.99}$ components that are isomorphic to H, then at least $n^{0.9}$ of them appear in G with colours unchanged.*

Note that G_0 can be taken to be the empty graph, and then Lemma 3.3 yields a red-blue colouring of G for which the two colours span *almost* isomorphic linear forests . Indeed, this follows from Items (Q1) and (Q3); Items (Q2) and (Q4) are only needed for proving the exact result. In particular, this gives a short probabilistic proof for an approximate version of Conjecture 1.2, as written in Theorem 1.4. As mentioned above, the more involved part will be to prove the conjecture in its full power, which will be the content of the next sections of this paper.

We now turn to the proof of Lemma 3.3.

Proof. We will define four random red-blue colourings $\chi_1, \ldots, \chi_4$ of G, with the purpose of showing that χ_4 satisfies (Q1) to (Q4), with positive probability. We will start our colouring process with a deterministic colouring χ_0.

Defining χ_0. Write $G_1 := G \setminus G_0$. Let H_1, H_2, H_3 be an edge-decomposition of G_1 satisfying (E4). We first define a purple-green colouring of G_1. Let F be the graph with degrees 1 and 3 that H_1 is a subdivision of. By Theorem 1.5, there is a purple-green colouring of F whose monochromatic components are paths of length at most 5. This defines a purple-green colouring of H_1 in a natural way: if an edge e in F is subdivided into a path P_e in H_1 then colour the edges of P_e by the colour of e. Now colour H_2 as follows. Fix a component C in H_2; so C is an even cycle with at most two edges of $G_1 \setminus C$ touching it. If C has no edges of $G_1 \setminus C$ touching it, colour C purple; if C has exactly one such edge e touching it, colour C by the opposite colour to e; finally if there are two such edges e and f, if they both have the same colour then colour C by the opposite colour, and if they have opposite colours, colour one segment of C between e and f purple and the other green. Finally, we colour H_3. Let H' be a component in H_3; then H_3 consists of three paths P_1, P_2, P_3 whose interiors are mutually vertex-disjoint and which have the same ends points. Without loss of generality, the lengths of P_1 and P_2 have the same parity. Colour $P_1 \cup P_2$ purple, and colour P_3 green.

Denote the resulting colouring by χ_0.

Claim 3.4. *χ_0 is a purple-green colouring of G_1, whose monochromatic components are even cycles or paths, both of length $O(\sqrt{\log n})$. Additionally, for every vertex with degree at least 2 in G_1, there are two edges incident to it which have the same colour in G_1.*

Proof. Note that the vertices who have degree two either are part of even cycles of H_2 or internal vertices of the subdivision paths of H_3, so they are adjacent to two edges of the same colour. Similarly, by considering the vertices in H_1, H_2, H_3 separately, it is easy to verify that vertices of degree three have at least two edges adjacent to them of the same colour. $\qquad\square$

Defining χ_1. We now define χ_1 to be the random red-blue colouring of G as follows. For each monochromatic component in (G_1, χ_0) we uniformly at random colour it alternating starting red, or starting blue (this colouring step is similar to the first colouring step in [12]). Then for each monochromatic component of G_0 we either with probability $1/2$ keep the colouring of the component or swap

the colours of all of its edges (red becomes blue, blue becomes red). All these colourings are made independently.

We claim that every vertex in (G, χ_1) is incident with two red edges and one blue, or vice versa. Indeed, this holds for vertices with at least two neighbours in G_0 by (E1), and by the properties of the green-purple colouring which imply that every vertex with at least two neighbours in G_1 is incident to two edges of the same colour (see the second part of Claim 3.4).

Our next goal is to get rid of monochromatic cycles. In order to later prove concentration, we would like all monochromatic cycles to be broken simultaneously (as opposed to iteratively, like in Akiyama–Chvátal's proof [3]), as this makes it easy to analyse the random process. We will break all monochromatic cycles in the next two steps.

Defining χ_2. For every monochromatic cycle C in (G, χ_1), choose, uniformly at random and independently, an edge e_C in C and flip its colour.

Note that it is possible to create new monochromatic cycles in χ_2. In the next step we get rid of them, and this process does not propagate any further.

Defining χ_3. Let C be a monochromatic cycle in χ_2 and let $e_1, \ldots, e_k$ be the edges of C whose colour was changed in the previous step; notice that there is at least one such edge. Note that if C and hence $e_1, \ldots e_k$ are blue, then each e_i closes a red cycle C_i (C_i was the red cycle in χ_1 which prompted the colour change of e_i), and similarly with the roles of red and blue reversed. We call these k formerly red cycles the *petals* of C. We call a monochromatic cycle C together with its petals $C_1, \ldots, C_k$ a *monochromatic cycle-petal configuration* (see Figure 1). It is easy to see that all monochromatic cycles together their petals are vertex-disjoint from each others.

A monochromatic cycle-petal configuration   Its recolouring at e_2 and e_2'
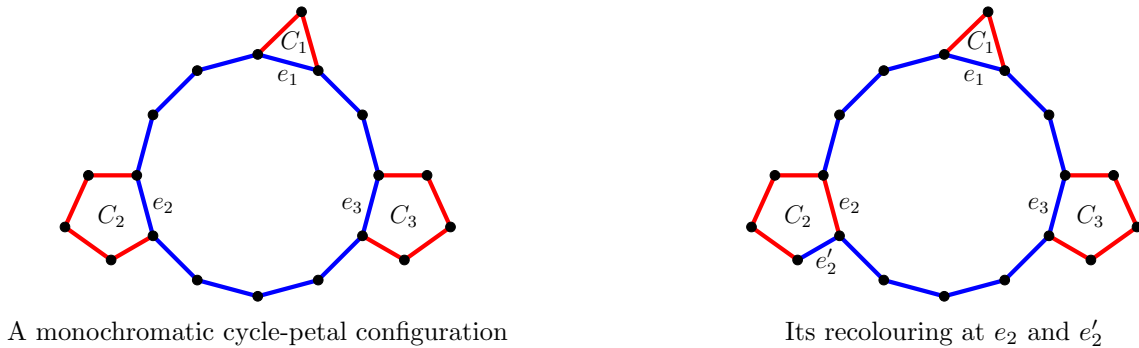
Figure 1: Getting rid of monochromatic cycles

It is easy to check that the following is true.

Claim 3.5. *Any two distinct monochromatic cycles-petals configurations in χ_2 are vertex-disjoint.*

For each monochromatic cycle C in χ_2 with petals $C_1, \ldots, C_k$ and petal edges $e_1, \ldots, e_k$, pick $i \in [k]$ uniformly at random and then pick one of the two edges next to e_i on C_i uniformly at random, call it e_i', and swap the colours of e_i and e_i' (see Figure 1). This clearly destroys the monochromatic cycle C; moreover, it does not create vertices of red degree 3 or blue degree 3. We do this for all cycles simultaneously, and Claim 3.5 implies that the following holds.

Claim 3.6. *The monochromatic components in (G, χ_3) are paths.*

Proof. Note that using Claim 3.5 we know that no edge changes their colour more than once while defining χ_3, thus it is enough to show that after a swap for one particular cycle C, e_i and e_i' are not involved in new monochromatic cycles. Let us assume the edge e_i which was swapped for cycle C was blue before the swap. That means the cycle C was blue in χ_2 and e_i' was red. Let v be the vertex common to e_i and e_i' and write $e_i = uv$, and assume $e_i' = vw$. Notice that v has blue degree 2 in χ_2 and so e_i which is red in χ_3 is not in a monochromatic cycle. For the monochromatic cycles involving e_i', there are two options, either they use e_i' and the vertex u or they use another vertex of cycle C; the point being every such cycle has to use w. For the first case, $C \setminus \{e_i\} \cup \{e_i'\}$ is a blue path ending with the vertex u which has red degree 2, showing that e_i' does not belong to a monochromatic cycle which uses the vertex u. For the second case, e_i' cannot be involved in a blue cycle using some other vertex of C because the blue graph has maximum degree two. $\qquad\square$

Defining χ_4. For each maximal monochromatic path P in (G, χ_3), pick, uniformly at random and independently, an edge e_P of the opposite colour which is incident to an end of P (so e_P is chosen out of a set of at most four edges). For each edge $e = xy$, if there are two distinct paths P and Q such that $e = e_P = e_Q$ (so x is an end of one of exactly one of these paths and y is an end of only the other path), swap the colour of e with probability $1/2$; if no such paths exist, e remains unchanged. Call the resulting colouring χ_4.

Define $\chi_3^{r \to b}$ to be the colouring of G obtained from χ_3 by swapping the colour of edges whose colour was swapped from red to blue when going from χ_3 to χ_4 (and keeping the colour of all other edges), and define $\chi_3^{b \to r}$ analogously with the roles of red and blue reversed.

Claim 3.7. *Every red path P in χ_3 touches at most one edge whose colour changed from blue to red when going from χ_3 to $\chi_3^{b \to r}$. Moreover, such an edge must touch an endpoint of P and contains no other vertices of P.*

Proof. The only blue edges touching P whose colour can be swapped are those that touch the ends of P, and of those only e_P can be swapped. This proves the first part of the claim. To see the second part, write $e_P = xy$ and suppose that x is an end of P and that the colour of e_P was swapped. Notice that, by definition of χ_4, if e_P is indeed swapped then there is another red path Q for which $e_Q = e_P$. But this means that y is an end of Q, and so $y \notin V(P)$. $\qquad\square$

The next claim easily follows.

Claim 3.8. *All monochromatic components in χ_4 are paths, and contain at most one edge that was of opposite colour in χ_3.*

Claim 3.9. *With probability at least $4/5$, there are no monochromatic paths of length at least $1000 \log n$ in χ_4.*

Proof. We will show that, with probability at least $9/10$, every blue path of length $400 \log n$ in χ_3 contains at least one edge whose colour in $\chi_3^{b \to r}$ is red. Fix a path P of length $400 \log n$ which is blue in χ_3. Let cand(P) be the set of edges in the interior of P that touch two distinct red components. These are *candidates* for possibly flipping their colour when going from χ_3 to χ_4. Notice that out of any

11

two consecutive edges in the interior of P, at least one is in $\operatorname{cand}(P)$. This is because monochromatic components are paths and they can only touch P at their ends. Thus $|\operatorname{cand}(P)| \geq (|P|-2)/2$. Consider an edge $e = xy$ in $\operatorname{cand}(P)$, and let Q_1 and Q_2 be the maximal red paths touching x and y, respectively. Then, with probability at least $1/16$ we have $e_{Q_1} = e_{Q_2} = e$ because there are at most four choices for e_{Q_i} and the choices are made independently. Thus, the probability that the colour of e is swapped is at least $1/32$.

Consider an auxiliary graph with vertices $\operatorname{cand}(P)$ and where ef is an edge if and only if there is a red path touching both e and f. This graph has maximum degree at most 4 thus, by Turán's theorem, it has an independent set, let's call it $i(P) \subseteq \operatorname{cand}(P)$ of size at least $|\operatorname{cand}(P)|/5 \geq (|P|-2)/10$. So the properties of edges in $i(P)$ is that no two edges in $i(P)$ touch the same red path. This means that the events $\{e$ is blue in $\chi_3^{b \to r}\}$, for $e \in i(P)$, are independent. It follows that the probability that none of the edges of P are blue in $\chi_3^{b \to r}$ is at most $(1 - 1/32)^{(|P|-2)/10}$.

The number of blue paths of length $400 \log n$ in χ_3 is at most $2n$ (there are n ways to pick an end point, and at most two possible directions). Hence the probability that some red path of length $400 \log n$ in χ_3 survives is at most

$$2n \cdot \left(1 - \frac{1}{32}\right)^{(400 \log n - 2)/10} \leq 2n \cdot \exp\left(-\frac{400 \log n - 1}{320}\right) < 1/10.$$

In particular, with probability at least $9/10$, there are no blue paths of length at least $400 \log n$ in $\chi_3^{b \to r}$. By Claim 3.8, this shows that there are no red paths in χ_4 of length at least $800 \log n + 1$. Repeating the same argument with the roles of red and blue reversed, the claim is proved. $\square$

Claim 3.10. *With probability at least $4/5$, there is no red path of length $(\log n)^{2/3}$ in χ_4 which is intersected by at least $100\sqrt{\log n}$ blue paths of length at least $100\sqrt{\log n}$, and vice verse.*

Proof. Let $\mathcal{H}$ be the collection of subgraphs H of χ_3 with the following structure: H is the union of a red path P of length at most $(\log n)^{2/3}$ with $100\sqrt{\log n}$ pairwise vertex-disjoint paths Q of length $50\sqrt{\log n}$, each of which is either a blue path that touches P, or the union of a maximal blue path Q_1 with one end in P, a red edge touching the other end of Q_1, and another blue path Q_2 that touches e and is disjoint of Q_1.

We claim that $|\mathcal{H}| \leq 2n(\log n)^{2/3} \cdot \binom{(\log n)^{2/3}}{100\sqrt{\log n}} \cdot 4^{100\sqrt{\log n}}$. Indeed, there are at most $2n(\log n)^{2/3}$ ways to choose a red path P of length at most $\log n^{2/3}$, then there are at most $\binom{(\log n)^{2/3}}{100\sqrt{\log n}}$ ways to choose the vertices on P where the red paths in $\mathcal{Q}$ touch P. Finally, for every such interior vertex v of P either there exists a maximal blue path Q of length at least $100 \log n$ which touches P at v and we add it to $\mathcal{Q}$, or the maximal path has length less than $100 \log n$ and at the end of it there are two red edges, each of which is possibly adjacent to two blue edges so we have four possible choices for this specific Q_2, thus resulting overall $4^{100\sqrt{\log n}}$ choices. What is the probability that such an H did not have any of its blue edges swapped to red colour in χ_4? This can be phrased as follows: let $\mathcal{P}$ be a collection of at most $200\sqrt{\log n}$ pairwise disjoint blue paths in χ_3 whose total length is at least $5000 \log n$. Similarly to the proof of Claim 3.9, the probability that all edges in paths in $\mathcal{P}$ remain blue in χ_4 is at most $(1 - 1/32)^{(1000 \log n - 400\sqrt{\log n})/10}$. Putting all this together, we get that the probability that there exists a

graph $H \in \mathcal{H}$ has none of its blue edges swapped is at most

$$2n \cdot (\log n)^{2/3} \cdot \left(\frac{(\log n)^{2/3}}{100\sqrt{\log n}} \right) \cdot 2^{100\sqrt{\log n}} \cdot \left(1 - \frac{1}{32} \right)^{(1000 \log n - 400\sqrt{\log n})/10} \leq \frac{1}{10}.$$

Notice that Claim 3.8 shows that any red path of length at most $(\log n)^{2/3}$ in χ_4 which is intersected by at least $100\sqrt{\log n}$ blue paths of length at least $100\sqrt{\log n}$ gives rise to a graph in $\mathcal{H}$ whose blue edges did not change. The claim thus follows. $\qquad \square$

Claim 3.11. *Property* (Q3) *holds, with probability at least* $4/5$.

Proof. In order to show concentration, we will first describe how the random colouring described in this section can be obtained as a function of $9n$ independent random variables. For this, fix a permutation σ of the edges of G. We define families of sets A_i, for $i \in [6]$, as follows.

- Let $A_1 = A_6 = \{0, 1\}$.

- Let $A_2, \ldots, A_5$ be the sets of permutations of $E(G)$ (so the sets $A_2, \ldots, A_5$ are identical).

For each edge e and $i \in [6]$, let $X_i(e)$ be the random variable obtained by picking an item from A_i, uniformly at random and independently.

For a permutation τ of $E(G)$ and a subgraph $F \subseteq G$, let $\tau(F)$ be the edge of F that appears first in τ. Note that if τ is a uniformly random permutation of $E(G)$ then $\tau(H)$ is a uniformly random edge in H.

Now define random variables $Y_i(H)$, as follows.

- For $i \in \{1, 6\}$, define $Y_i(H) := X_i(\sigma(H))$ (here σ is the permutation we fixed above). So $Y_i(H)$ is either 0 or 1, each with probability $1/2$, and the choice depends on $X_i(e)$, for the first edge e in H to appear in σ.

- For $i \in \{2, 3, 4\}$, define $Y_i(H) := \tau(H)$, where $\tau = X_i(\sigma(H))$. In other words, $Y_i(H)$ is an edge in H, chosen uniformly at random (indeed, $\tau = X_i(\sigma(H))$ is a uniformly random permutation of $E(G)$, and thus $\tau(H)$ is an edge of H, chosen uniformly at random), and this choice depends on $X_i(e)$, where e is the first edge in H to appear in σ.

- For a path P, let E_P be the set of edges outside of P that touch an end of P, and define $Y_5(P) := \tau(E_P)$, where $\tau = X_5(\sigma(P))$. So $Y_5(P)$ picks an edge from E_P uniformly at random, and the choice depends on $X_5(e)$, where e is the first edge in P to appear in σ.

We will show that the random colourings $\chi_1, \ldots, \chi_4$, defined above, can be defined as functions of the sequence of random variables $\bar{X} := (X_i(e))_{i,e}$. Given an assignment $\bar{x}$ of values to $\bar{X}$, write $\chi_i(\bar{x})$ to be the red-blue colouring χ_i determined by $\bar{x}$, for $i \in [4]$.

We will then show that, for every $t \geq 1$, if $\bar{x}$ and $\bar{y}$ are two assignments of values to $\bar{X}$ that differ on exactly one coordinate, then the number of red (blue) components in $\chi_4(\bar{x})$ and $\chi_4(\bar{y})$ that are isomorphic to P_t differs by $O(\sqrt{\log n})$; equivalently, $|r_{\chi_4(\bar{x})}(P_t) - r_{\chi_4(\bar{y})}(P_t)| = O(\sqrt{\log n})$ (with the analogous statement for blue also holding).

13

Apply McDiarmid's inequality (Theorem 3.1), with $f(\bar{X}) = r_{\chi_4(\bar{x})}(P_t)$, $m = \frac{1}{2}n^{2/3}$, $c := C\sqrt{\log n}$ for a large enough constant C, and $n_{3.1} = 6e(G)$. It follows that $|r_{\chi_4(\bar{X})}(P_t) - \mathbb{E}(r_{\chi_4(\bar{X})})| \le \frac{1}{2}n^{2/3}$, with probability at least $1 - 2\exp(-\Omega(n^{1/4}))$, for every t. By symmetry, the same holds for $b_{\chi_4(\bar{X})}(P_t)$. Noting that $\mathbb{E}(r_{\chi_4(\bar{X})}) = \mathbb{E}(b_{\chi_4(\bar{X})})$, again by symmetry, we have $|r_{\chi_4(bX)}(P_t) - b_{\chi_4(bX)}| \le n^{2/3}$, with probability $1 - 2\exp(-\Omega(n^{1/4}))$. Taking a union bound over all $t \le n$ implies that (Q3) holds.

Let us now explain how to obtain χ_i, with $i \in [4]$, from $\bar{x}$. Fix a colouring χ_0, obtained as explained at the beginning of the proof of Lemma 3.3. To define χ_1, for each purple or green component H, colour H by one of its two proper red-blue edge-colourings, according to $Y_1(H)$ (so we think of 0 as signifying one of these colourings, and 1 as signifying the other). Similarly, for a connected component H in the red-blue subgraph of G, leave H unchanged if $Y_1(H) = 0$, and otherwise swap red and blue on H. Note that these choices are independent, as the subgraphs in question are pairwise edge-disjoint and thus depend on different variables $X_1(e)$.

Next, to define χ_2, for each monochromatic cycle C swap the colour of $Y_2(C)$; as before, these choices are independent. For χ_3, consider a monochromatic (say red) cycle C in χ_2 and let E_C be the set of edges in C whose colour was swapped in the previous step. Write $e := Y_3(E_C)$, let f_1, f_2 be the two blue edges touching e, and let $f := Y_4(\{f_1, f_2\})$. Swap the colours of e and f to get χ_3. We again have independence, as each pair f_1, f_2 originates from a different cycle that was monochromatic in χ_1.

Finally, if there are two distinct maximal monochromatic paths P and Q such that $X_5(P) = X_5(Q)$, write $e := X_5(P)$ and swap the colour of e if $X_6(e) = 1$ (and do nothing otherwise). As usual, since the relevant paths are pairwise edge-disjoint, we have independence. It should be easy to check that $\chi_1, \dots, \chi_4$ defined in this way have the same distribution as their counterparts defined above.

In what follows we show that if $\bar{x}$ and $\bar{y}$ differ on at most one coordinate then the colourings $\chi_4(\bar{x})$ and $\chi_4(\bar{y})$ differ on $O(\sqrt{\log n})$ edges. Notice that this means that the collections of monochromatic components differ on $O(\sqrt{\log n})$ elements, as claimed above.

- If $\bar{x}$ and $\bar{y}$ differ on exactly one coordinate, with $i \in \{5, 6\}$, then the colourings $\chi_4(\bar{x})$ and $\chi_4(\bar{y})$ differ on at most one edge.

  Suppose that $\bar{x}$ and $\bar{y}$ agree on coordinates with $i \ge 5$ and $\chi_3(\bar{x})$ and $\chi_3(\bar{y})$ differ on at most a edges. Then the collections of maximal monochromatic paths in $\chi_3(\bar{x})$ and $\chi_3(\bar{y})$ differ on at most $6a$ elements. Suppose that e is an edge which is red in $\chi_4(\bar{x})$ and blue in $\chi_4(\bar{y})$. Then either e's colour in $\chi_3(\bar{x})$ and $\chi_3(\bar{y})$ was different, or, without loss of generality, its colour was swapped by $\chi_4(\bar{x})$ but not $\chi_4(\bar{y})$. In particular, in $\chi_3(\bar{x})$ the edge e was blue, and touched the ends of two distinct maximal red paths P and Q. If this were true for $\chi_3(\bar{y})$ as well, with P and Q unchanged, then e's colour would be swapped by $\chi_4(\bar{y})$ too (using that $\bar{x}$ and $\bar{y}$ agree on coordinates with $i \ge 5$). Thus one of P and Q is not a monochromatic component in $\chi_3(\bar{y})$. In summary, an edge e can have a different colour in $\chi_4(\bar{x})$ and $\chi_4(\bar{y})$ only if its colour in $\chi_3(\bar{x})$ and $\chi_3(\bar{y})$ is not the same (this can happen at most a times), or it touches the end of a path which is a monochromatic component in exactly one of $\chi_3(\bar{x})$ and $\chi_3(\bar{y})$ (this can happen at most $4 \cdot 6a$ times). Thus $\chi_4(\bar{x})$ and $\chi_4(\bar{y})$ differ on at most $25a$ edges.

- If $i = 4$ then $\chi_3(\bar{x})$ and $\chi_3(\bar{y})$ differ on at most two edges, and if $i = 5$ then $\chi_3(\bar{x})$ and $\chi_3(\bar{y})$ differ on at most four edges. By the previous item, this shows that $\chi_4(\bar{x})$ and $\chi_4(\bar{y})$ differ on at most 100 edges.

Suppose that $\bar{x}$ and $\bar{y}$ agree on coordinates with $i \geq 4$ and $\chi_j(\bar{x})$ and $\chi_j(\bar{y})$ differ on at most a edges for $j \in [2]$. Thus the collections of monochromatic cycles-and-petals, defined according to $(\chi_1(\bar{x}), \chi_2(\bar{x}))$ and $(\chi_1(\bar{x}), \chi_2(\bar{x}))$, differ on at most $4a$ such structures. Indeed, if H is a cycle-and-petals structure as defined by $(\chi_1(\bar{x}), \chi_2(\bar{x}))$ but not by $(\chi_2(\bar{y}), \chi_2(\bar{y}))$, then one of its edges has different colours in $\chi_j(\bar{x})$ and $\chi_j(\bar{y})$, for some $j \in [2]$. Since each such H gives rise to two colour swaps, and edges not in such structures do not change colours, it follows that $\chi_3(\bar{x})$ and $\chi_3(\bar{y})$ differ on at most $9a$ edges. The previous item implies that $\chi_4(\bar{x})$ and $\chi_4(\bar{y})$ differ on at most $225a$ edges.

- If $i = 2$ then $\chi_2(\bar{x})$ and $\chi_2(\bar{y})$ differ on at most two edges. By the previous item (using that $\chi_1(\bar{x}) = \chi_1(\bar{y})$), the colourings $\chi_4(\bar{x})$ and $\chi_4(\bar{y})$ differ on at most 450 edges.

  Suppose that $\bar{x}$ and $\bar{y}$ agree on coordinates with $i \geq 2$ and that $\chi_1(\bar{x})$ and $\chi_1(\bar{y})$ differ on at most a edges. Then the collections of monochromatic cycles in $\chi_1(\bar{x})$ and $\chi_1(\bar{y})$ differ by at most $2a$ cycles, implying that $\chi_2(\bar{x})$ and $\chi_2(\bar{y})$ differ on at most $3a$ edges. By the previous item, this shows that $\chi_4(\bar{x})$ and $\chi_4(\bar{y})$ differ on at most $675a$ edges.

- If $i = 1$ then $\chi_1(\bar{x})$ and $\chi_2(\bar{y})$ differ on $O(\sqrt{\log n})$ edges. Thus $\chi_4(\bar{x})$ and $\chi_4(\bar{y})$ differ on $O(\sqrt{\log n})$ edges.

To summarise, we have shown that if $\bar{x}$ and $\bar{y}$ differ on at most one coordinate then $|r_{\chi_4(\bar{x})}(P_t) - r_{\chi_4(\bar{y})}| = O(\sqrt{\log n})$, for every $t \geq 1$. It follows from McDiarmid's inequality Theorem 3.1 that, with probability at least $9/10$, we have $\left| r_{\chi_4(\bar{X})}(P_t) - \mathbb{E}[r_{\chi_4(\bar{X})}(P_t)] \right| \leq \frac{1}{2}n^{2/3}$, for every $t \geq 1$.

Observe that red and blue were completely symmetric throughout the random process, and in all steps all actions were performed simultaneously. Thus, $\mathbb{E}[r_{\chi_4(\bar{X})}(P_t)] = \mathbb{E}[b_{\chi_4(\bar{X})}(P_t)]$. It follows from the previous paragraph that $\left| r_{\chi_4(\bar{X})}(P_t) - b_{\chi_4(\bar{X})}(P_t) \right| \leq n^{2/3}$ for every $t \geq 1$, with probability at least $4/5$. This proves the claim. $\qquad \square$

Claim 3.12. *Property* (Q4) *holds for χ_4 with probability at least* $4/5$.

Proof. Let H be a red-blue subcubic graph, and let $\mathcal{H}_0$ be the collection of components in G_0 that are isomorphic to H. We assume that $|\mathcal{H}| \geq n^{0.99}$, as otherwise there is nothing to prove. It is easy to see that, with probability at least 0.99, in (G_0, χ_1) there are at least $n^{0.9}$ components that are isomorphic to H; denote the set of such components by $\mathcal{H}_1$. The main challenge of this claim is to show that many graphs in $\mathcal{H}_1$ remain unchanged in χ_2 and χ_3, and for this we will use Property (E3). It is then quite easy to show that such graphs have decent probability of remaining the same in χ_4, too.

For this claim it will be convenient to think of the process defining χ_2 and χ_3, as follows. Fix a direction for each monochromatic cycle in χ_1. Each such cycle chooses, uniformly at random, one of its edges e_C^1. Denote by e_C^2 the edge next to e_C^1 (according to the direction we fixed). Now, the colour of one of e_C^1 and e_C^2 (chosen uniformly at random) is swapped. This yields χ_2. Recall that in χ_2, we consider monochromatic cycles C along with their edges $e_1, \ldots, e_{k_C}$, that were swapped previously, and the corresponding petals $C_1, \ldots, C_{k_C}$ (where C_i is a cycle that was monochromatic in χ_1 and elected to swap e_i). Now, i is chosen uniformly at random from $[k_C]$. Notice that $e_i = e_{C_i}^j$ for some $j \in [2]$. Finally, the colour of e_i is swapped back, and the colour of $e_{C_i}^{3-j}$ is swapped. It is easy to see that this emulates the processes used to obtain χ_3 and χ_4, with the slight difference that e_i' is chosen pre-emptively.

The point of this discussion is that for some $H \in \mathcal{H}_1$ to survive, it suffices to make sure that e_C^1 is chosen so that it and its successor in C are not in H, for every monochromatic cycle C that intersects $E(H)$. While it is not hard to show that this is true with not-too-small probability for any $H \in \mathcal{H}_1$, we need to work a little harder to define events which are independent.

For each $H \in \mathcal{H}_1$, we define a collection $\mathcal{C}_H$ of subgraphs of monochromatic cycles as follows. For a monochromatic cycle C, let $H_1, \ldots, H_k$ be the graphs in $\mathcal{H}_1$ that contain edges of C. Let E_i be the collection of edges in C that are at distance at most 2 from H_i (notice that the sets E_i are pairwise disjoint by the assumption that the graphs in $\mathcal{H}_1$ are at distance at least 10 from each other). Add E_i to $\mathcal{C}_{H_i}$, for each i. Add $E_{C_0} := E(C) \setminus (E_{C,1} \cup \ldots \cup E_{C,k})$ to a "reserve collection" $\mathcal{C}_0$. To choose e_C^1, we let each set E_i choose an edge $e_{E_i} \in E_i$ uniformly at random and then pick $i \in [0, k]$ with probability $\frac{|E_i|}{|C|}$, and take $e_C^1 = e_{E_i}$. Observe that this process picks e_C^1 uniformly at random from C.

We claim that, for every $H \in \mathcal{H}_1$, every $E \in \mathcal{C}_H$ contains two consecutive edges outsider of H. Indeed, if a monochromatic cycle C intersects some $H' \in \mathcal{H}_1$ other than H, then this holds due to the distance assumption on H and H', and otherwise it follows from (E3). Moreover, by (E2), every $E \in \mathcal{C}_H$ has size $O(\sqrt{\log n})$. Thus, with probability at least $\Omega(1/\sqrt{\log n})$, both e_E and its successor in its monochromatic cycle C are not in H. Importantly, if this is the case then it is guaranteed that e_C^1 and its successor in C are not in H. Denote by A_H the event that e_E and its successor in its monochromatic cycle are not in H, for every $E \in \mathcal{C}_H$. Because H has $O(\sqrt{\log n})$ edges, and the elements in $\mathcal{C}_H$ are pairwise edge-disjoint, $\mathbb{P}[A_H] \geq (\sqrt{\log n})^{-O(\sqrt{\log n})} \geq n^{-0.01}$. Since the events $(A_H)_{H \in \mathcal{H}_1}$ are independent, it follows that with probability at least 0.99, at least $n^{0.8}$ of them hold, showing that the collection $\mathcal{H}_3$ of graphs $H \in \mathcal{H}_1$ whose colours remain unchanged in χ_3 has size at least $n^{0.8}$.

To finish, notice that, with probability at least $\exp(-O(\sqrt{\log n})) \geq n^{-0.01}$, all edges in H decide to veto being swapped, for each $H \in \mathcal{H}_3$. Since these events are independent, it follows that with probability at least 0.99, at least $n^{0.7}$ graphs in $\mathcal{H}_3$ remain unchanged in χ_4. This proves the claim. □

The proof of Lemma 3.3 now follows very easily. Indeed, with probability at least $1/5$, Properties (Q1) to (Q4) all hold, using Claims 3.8 to 3.12. □

4   Exact solution - proof of the main theorem

In this section we will state the main lemmas needed to turn the approximate version (Lemma 3.3) into an exact solution of Wormald's conjecture. We will then show how to prove our main theorem (Theorem 1.3) using these lemmas.

An important element of these lemmas is the notion of *gadgets*, defined here. We only give an abstract definition here, and later (in Section 6) we give a concrete definition, that will be shown to satisfy the abstract one.

Definition 4.1. A *blue ℓ-gadget* is a red-blue subcubic graph H, satisfying the following. There is another red-blue colouring H' of H, such that for every red-blue cubic graph G that contains H, and whose monochromatic components are paths, the graph G', obtained from G by replacing H by H', satisfies the following properties.

- the monochromatic components in G' are paths,

- $r_G(P_t) = r_{G'}(P_t)$ for every t,

- $b_G(P_t) - b_{G'}(P_t) = \begin{cases} 0 & \text{if } t > \ell \\ 1 & \text{if } t = \ell. \end{cases}$

- H and H' differ on at most two edges,

- H contains a blue path of length exactly ℓ, whose ends are incident with two red edges.

A *red ℓ-gadget* is defined analogously, with the roles of red and blue replaced.

We remark that the last two properties are not crucial, but it will later be convenient to have them, and they are satisfied naturally by the explicit gadgets that we use.

We now state a key lemma. It will give us a partial colouring of a cubic graph G with properties that make it amenable to an application of Lemma 3.3.

Lemma 4.2. *Let G be a connected cubic graph on n vertices, where n is large. Then there exists a red-blue colouring of a subgraph $G_0 \subseteq G$, such that the following holds.*

1. *G_0 is extendable according to Definition 3.2,*

2. *for every $\ell \in [3, 10^{10}\sqrt{\log n}]$ there are at least $n^{0.999}$ components in G_0 that contain a blue ℓ-gadget.*

Most of the work towards the proof of Lemma 4.2 will go into finding a partial colouring satisfying (E1) and (E3) from Definition 3.2. Later, we will show that this colouring also satisfies (E2) and (E4). To obtain such a colouring, we find many geodesic paths that are sufficiently far apart from each other (see Claim 5.1), and colour small balls around so that the desired properties hold. The latter is the content of the next lemma.

Lemma 4.3. *Let P be a geodesic of length at least $5\ell + 200$ in a cubic graph G. Then there is a partial colouring ϕ_P, such that*

1. *the edges coloured by ϕ_P are at distance at most 4 from P and form a connected subgraph,*

2. *the edges coloured by ϕ_P contains a blue ℓ-gadget,*

3. *the partial colouring ϕ_P satisfies (E1) and (E3) from Definition 3.2.*

The case that where no two vertices in the geodesic have a common neighbour outside of P is easier to handle, so we consider it separately in the next lemma. Given this lemma, the proof of Lemma 4.3 will focus on geodesics with (many) pairs of vertices having a common neighbour outside.

Lemma 4.4. *Let P be a geodesic of length at least $\ell + 36$ in a cubic graph G. Assume that no two vertices of P have a common neighbour outside of P. Then there is a partial colouring ϕ_P, such that*

1. *the edges coloured by ϕ_P are at distance at most 3 from P and form a connected subgraph,*

2. *the edges coloured by ϕ_P contains a blue ℓ-gadget,*

3. *the partial colouring ϕ_P satisfies (E1) and (E3) from Definition 3.2.*

The rest of the paper is organised as follows. In the next subsection (Section 4.1) we show how to prove the main theorem (Theorem 1.3) using Lemma 3.3 and Lemma 4.2. In Section 6 we present the explicit gadgets that we find in the proofs of Lemmas 4.3 and 4.4. We prove Lemma 4.4 and Lemma 4.3 in Sections 7 and 8, respectively. Finally, in Section 5 we show how to combine everything to prove Lemma 4.2.

4.1 Proof of the main theorem

In this section we prove our main result, restated here, from previously stated lemmas.

Theorem 1.3. *Let G be a connected cubic graph on n vertices, where n is large and divisible by 4. Then there is a red-blue colouring of the edges of G whose colour classes span isomorphic linear forests.*

The proof proceeds as follows. We start by fixing a partial colouring G_0, as guaranteed by Lemma 4.2. We then swap the colours of some of the component of G_0, to ensure that there are many blue and red ℓ-gadgets, for all relevant ℓ. We then consider a full red-blue colouring of G, as guaranteed by Lemma 3.3. We now equalise $r(P_t)$ and $b(P_t)$, for all $t \geq 1$, in three steps. First, we equalise $r(P_t)$ and $b(P_t)$ for all $t \geq 4000\sqrt{\log n}$. For this, we use that all components in G are monochromatic paths of length $O(\log n)$, and that G has no short red paths with many long blue paths touching it. We then equalise the number of red and blue edges. Finally, we equalise $r(P_t)$ and $b(P_t)$, one by one, from the largest t for which they differ, to $t = 3$, using gadgets. It is not hard to see that if $r(P_t) = b(P_t)$ for $t \geq 3$ and the number of red and blue edges is the same, then also $r(P_2) = b(P_2)$ and $r(P_1) = b(P_1)$, so we are done.

Proof of Theorem 1.3 using Lemma 4.2. Apply Lemma 4.2 to obtain a subgraph $G_0 \subseteq G$ along with a red-blue colouring which is extendable (recall Definition 3.2) and contains at least $n^{0.999}$ components containing blue ℓ-gadgets, for each $\ell \in [3, 10^{10}\sqrt{\log n}]$. Swap the colours of half of the blue ℓ-gadgets to obtain a new extendable colouring of G_0, denoted G_0', for which there are at least $\frac{1}{2}n^{0.999}$ components containing blue (respectively red) ℓ-gadgets. It is easy to see that this new colouring is also extendable.

Notice that the number of red-blue coloured subcubic graphs of order at most $10^{10}\sqrt{\log n}$ is at most $2^{3 \cdot 10^{10}\sqrt{\log n}} \left(10^{10}\sqrt{\log n}\right)^{3 \cdot 10^{10}\sqrt{\log n}} \leq n^{0.001}$, using that n is large. Thus, for every $\ell \in [3, 10^{10}\sqrt{\log n}]$, there are at least $n^{0.998}$ isomorphic components in G_0' containing a blue (respectively red) ℓ-gadget.

Apply Lemma 3.3 to obtain a red-blue colouring χ_1 of G satisfying properties (Q1) to (Q4). In the next claim we equalise the number of red and blue components that are isomorphic to P_t, for all $t \geq 4000\sqrt{\log n}$, while changing few edges.

Claim 4.5. *There is a red-blue colouring χ_2 of G such that*

(A1) *the monochromatic components in χ_2 are paths of length $O(\log n)$,*

(A2) *χ_1 and χ_2 differ on at most $n^{3/4}$ edges of G,*

(A3) *$r_{\chi_2}(P_t) = b_{\chi_2}(P_t)$ for every $t \geq 4000\sqrt{\log n}$.*

Proof. Let $\mathcal{P}_r$ and $\mathcal{P}_b$ be collections of, respectively, red and blue (maximal) paths in χ_1, obtained as follows. For $t \geq 4000\sqrt{\log n}$ write $x_t = r(P_t) - b(P_t)$; if $x_t > 0$ add x_t (maximal) red paths of length t to

$\mathcal{P}_r$, and if $x_t < 0$ add $-x_t$ (maximal) blue paths of length t to $\mathcal{P}_b$. Form $\mathcal{J}_r$ by decomposing each path in $\mathcal{P}_r$ into subpaths of length in $[1000\sqrt{\log n}, 2000\sqrt{\log n}]$, and define $\mathcal{J}_b$ analogously.

For a path $J \in \mathcal{J}_r$, let $\operatorname{cand}(J)$ be the set of edges xy in J such that the blue components containing x and y are distinct paths of length at most $c\sqrt{\log n}$. Observe that if x, y, z are three consecutive vertices in J, and x and y are in the same blue path P, then $z \notin P$. Thus at least $(|J| - 3)/2$ edges xy in J are such that x and y belong to distinct blue components. By (Q2), at most $100\sqrt{\log n}$ vertices x in the interior of J are incident to blue paths of length at least $100\sqrt{\log n}$. It follows that there are at most $200\sqrt{\log n} \leq |J|/5$ edges xy in the interior of J which touch blue paths of length at least $100\sqrt{\log n}$. Altogether, it follows that $|\operatorname{cand}(J)| \geq |J|/5 \geq 200\sqrt{\log n}$.

Form a graph F_r on vertex set $I_r = \bigcup_{J \in \mathcal{J}_r} \operatorname{cand}(J)$, where ef is an edge whenever e and f belong to distinct paths in $\mathcal{J}_r$ and there is a blue path that touches both e and f. We claim that F_r has an *independent transversal*, namely an independent set $\{e_J : J \in \mathcal{J}_r\}$, where $e_J \in \operatorname{cand}(J)$. This follows from Lovász's local lemma and the observation that F_r has maximum degree at most 4 (in fact, Loh and Sudakov [24] prove a similar but stronger result).

Define $\mathcal{J}_b$ and F_b analogously, and let $I_b = \{e_J : J \in \mathcal{J}_b\}$ be an independent transversal in F_b.

Now obtain χ_2 from by swapping the colours of edges in $I_r \cup I_b$. We claim that χ_2 satisfies the requirements of the claim.

Indeed, for (A1), notice that we only swap the colours of edges in the interior of a monochromatic path, and thus we do not form vertices with monochromatic degree 3. Moreover, for every maximal monochromatic path P, we swap the colour of at most one edge touching P, and only if it touches only one vertex of P (which is an end of P).

Notice that by (Q3), $|r_{\chi_1}(P_t) - b_{\chi_1}(P_t)| \leq n^{2/3}$ for every t, and thus $|\mathcal{P}_r|, |\mathcal{P}_b| = O(n^{2/3} \log n)$, implying that $|\mathcal{J}_r|, |\mathcal{J}_b| = O(n^{2/3}(\log n)^2)$. Since the number of edges whose χ_1 and χ_2 colours differ is $|\mathcal{J}_r| + |\mathcal{J}_b|$, property (A2) follows, using that n is large.

To see (A3), notice that maximal red paths in χ_1, whose length is at least $100\sqrt{\log n}$ and which are not in $\mathcal{P}_r$, remain maximal red paths in χ_2, whereas for every path $P \in \mathcal{P}_r$, at least one among any $4000\sqrt{\log n}$ consecutive edges in P becomes blue in χ_2. Moreover, denoting by χ_1' the colouring obtained by swapping the colour of edges in I_b (but not in I_r), the maximal red paths in χ_1' are either maximal red paths in χ_1, or consist of two maximal red paths in χ_1, of length at most $100\sqrt{\log n}$ each, and a single additional edge. Thus the maximal red paths in χ_2 whose length is at least $4000\sqrt{\log n}$ are exactly the maximal red paths in χ_1 whose length is at least $4000\sqrt{\log n}$ and which are not in $\mathcal{P}_r$. An analogous reasoning holds for maximal blue paths, and yields (A3). $\qquad\square$

Claim 4.6. *There is a red-blue colouring χ_3 of G such that*

(B1) *the monochromatic components in χ_3 are paths of length $O(\log n)$,*

(B2) *χ_1 and χ_3 differ on at most $n^{4/5}$ edges,*

(B3) *$r_{\chi_3}(P_t) = b_{\chi_3}(P_t)$ for $t \geq 5000\sqrt{\log n}$.*

(B4) *the number of blue edges in χ_3 equals the number of red edges in χ_3.*

Proof. We use an argument similar to the proof of Claim 4.5. Let x_r and x_b be the number of, respectively, red and blue edges in χ_2. By (Q3) and (A2), $|x_r - x_b| \leq n^{2/3} + n^{3/4} \leq 2n^{3/4}$. Without loss of generality,

assume that $x_r \geq x_b$. Recall that by (Q4), χ_1 has at least $n^{0.9}$ red ℓ_0-gadgets, where $\ell_0 := 4000\sqrt{\log n}$. Since at most $n^{3/4}$ of them are destroyed when going from χ_1 to χ_2, this means in particular that there is a collection $\mathcal{P}$ of exactly $(x_r - x_b)/2$ distinct maximal red paths of length ℓ_0 (here we use the assumption that n is divisible by 4). Following the reasoning in the proof of Claim 4.5, there is a set $I = \{e_P : P \in \mathcal{P}\}$, where e_P is an edge in the interior of P which is incident to two distinct blue components, none of which is a path of length at least $100\sqrt{\log n}$, and moreover no two edges from P touch the same blue component. Define χ_3 to be the colouring obtained from χ_2 by swapping the colour of the edges in I. The proof that items (B1) to (B4) hold is very similar to the end of the proof of Claim 4.5; we omit the details. $\hfill\square$

Let χ_3 be a red-blue colouring of G satisfying (B1) to (B4) above. Our plan now is to equalise the number of maximal red and blue paths of length ℓ, starting with the maximal ℓ for which they differ, using ℓ-gadgets. Let $\ell_{\max}$ be the maximal ℓ for which $r_{\chi_3}(P_t) \neq b_{\chi_3}(P_t)$; by (B3), $\ell_{\max} \leq 5000\sqrt{\log n}$.

Claim 4.7. *For $\ell \in [3, \ell_{\max} + 1]$ there is a red-blue colouring ψ_ℓ of G, such that*

(C1) *all monochromatic components of ψ_ℓ are paths of length $O(\log n)$,*

(C2) *$r_{\psi_\ell}(P_t) = b_{\psi_\ell}(P_t)$ for $t \geq \ell$,*

(C3) *the number of blue edges in ψ_ℓ equals the number of edges edges,*

(C4) *ψ_ℓ differs from χ_1 on at most $f(\ell)$ edges, where $f(\ell) = n^{4/5} \cdot 6^{\ell_{\max} - \ell - 1}$.*

Proof. We prove the claim by induction on ℓ. Notice that, by taking $\psi_{\ell_{\max}+1} := \chi_3$, the claim holds for $\ell = \ell_{\max}+1$. Suppose that $\psi_{\ell+1}$ is a suitable for $\ell+1$, where $\ell \in [3, \ell_{\max}]$. Write $x = r_{\psi_{\ell+1}}(P_\ell) - b_{\psi_{\ell+1}}(P_\ell)$. Because $\psi_{\ell+1}$ and χ_1 differ on at most $f(\ell+1)$ edges, and by (Q3), $|x| \leq n^{2/3} + 2f(\ell+1) \leq n^{5/6}$. Without loss of generality, assume $x \geq 0$. By (Q4), we may pick a collection $\mathcal{H}$ of x red ℓ-gadgets in $\psi_{\ell+1}$ (that are a distance at least 10 away from each other). For each gadget $H \in \mathcal{H}$, let H' be its alternate colouring, which satisfies the properties in Definition 4.1. Obtain ψ_ℓ from $\psi_{\ell+1}$ by replacing H by H' for each $H \in \mathcal{H}$. It is easy to see that items (C1) and (C2) hold for ψ_ℓ. Notice that by the second item in Definition 4.1, the number of red edges in $\psi_{\ell+1}$ equals the number of red edges in ψ_ℓ, and thus (C3) holds. Finally, notice that $\psi_{\ell+1}$ and ψ_ℓ differ only on edges in gadgets in $\mathcal{H}$, and in each gadget they differ on at most two edges. Thus, they differ on at most $2x$ edges, implying that ψ_ℓ and χ_1 differ on at most the following number of edges

$$2x + f(\ell+1) \leq 2n^{2/3} + 4f(\ell+1) \leq 6f(\ell+1) = f(\ell),$$

using $f(\ell+1) \geq n^{2/3}$. Thus (C4) holds and the claim is proved. $\hfill\square$

We claim that ψ_3 is a red-blue colouring of G whose colour classes are isomorphic linear forests. Indeed, by (C1), the two colour classes are linear forests. By (C2), $r(P_t) = b(P_t)$, for $t \geq 3$. By (C3), the numbers of red and blue edges are the same. Because G is a cubic graph all of whose vertices are incident to two edges of one colour and one edge of the other colour, this means that the number of red P_2's is the same as the number of blue P_2's. Since we already know that the number of red P_2's that are contained in a longer red path is the same as the number of blue P_2's contained in a longer blue path, we find that $r(P_2) = b(P_2)$. Finally, a similar argument shows that $r(P_1) = b(P_1)$. $\hfill\square$

5 Completing gadgets to an extendable pre-colouring

In this section we prove Lemma 4.2, which asserts that every large connected cubic graph has an extendable (according to Definition 3.2) red-blue subgraph G_0 which has many gadgets of all relevant lengths. The proof will be conditional on Lemma 4.3, which finds a gadget within a small-radius neighbourhood of any long geodesic. Thus, after this section our only remaining task will be to prove Lemma 4.3.

We will apply Lemma 4.3 to a large collection of geodesics that are sufficiently far apart, and then modify the resulting partial colouring to an extendable one. The first step is an easy claim that shows that there are many long geodesics that are far apart from each other. Here, given a graph G, we write $d(u, v)$ for the distance in G between the vertices u and v, and $d(H, F)$ for the distance between subgraphs H and F, namely the minimum of $d(u, v)$ over $u \in V(H)$, $v \in V(F)$.

Claim 5.1. *Let $0 < \varepsilon < 1$ and let n be large. Suppose that G is a connected cubic graph on n vertices. Then there exists a collection $\mathcal{P}$ of at least $\frac{1}{6}n^{1-\varepsilon}$ geodesics, each of length $m := 10^{10}\sqrt{\log n}$, such that $d(P, Q) \geq 50$ for every distinct $P, Q \in \mathcal{P}$.*

Proof. We will show that there are at least $k = \frac{1}{6}n^{1-\varepsilon}$ vertices $v_1, \ldots, v_k$ in G such that $d(v_i, v_j) \geq \varepsilon \log n$ for every $i \neq j$. Note that every subpath of a geodesic is also a geodesic, thus if we take one subpath of length m of the geodesic touching v_i and some v_k, for each i, the collection of these paths will be as desired.

Let k_0 be the maximum such number of vertices. Then since the graph is cubic, the number of vertices in each ball of radius $\varepsilon \log n$ is at most $3 \cdot 2^{\varepsilon \log n}$. Thus, $\left| \bigcup_{i \geq k_0} B(v_i, \varepsilon \log n) \right| \leq k_0 \cdot 2^{\varepsilon \log n + 1} = 2k_0 \cdot n^{\varepsilon}$. On the other hand, by the maximality of k_0, we have that $\left| \bigcup_{i \geq k_0} B(v_i, \varepsilon \log n) \right| \geq n$. This gives $k_0 \geq \frac{1}{6}n^{1-\varepsilon}$. $\square$

Now we can use Lemma 4.3. Recall that Lemma 4.3 guarantees that for every geodesic P of length $5\ell + 200$ in a cubic graph G (where $\ell \in [3, 10^{10}\sqrt{\log n}]$), there is a partial colouring ϕ_P, such that

1. the edges coloured by ϕ_P are at distance at most 4 from P,

2. the edges coloured by ϕ_P contain a blue ℓ-gadget,

3. the partial colouring ϕ_P satisfies (E1) and (E3) from Definition 3.2.

We recall that (E1) says that every vertex with coloured degree at least 2 has both red and blue neighbours, and (E3) says that every cycle with only blue edges or only red edges has two consecutive uncoloured edges.

We next unite all of these colouring around paths into one partial colouring ϕ.

Definition 5.2. Let G be a cubic graph, let $\mathcal{P}$ be a collection of geodesics as in Claim 5.1 and, for $P \in \mathcal{P}$, let ϕ_P be a partial colouring satisfying 1, 2 and 3 above. Let ϕ be the colouring obtained by the union of all of those partial colourings $\phi := \bigcup_{P \in \mathcal{P}} \phi_P$ (notice that there are no conflicts, using Property 1 above and the assumption that any two geodesics in $\mathcal{P}$ are a distance at least 50 apart).

Claim 5.3. *The colouring ϕ satisfies (E1) and (E3). In addition, the size of each connected component of the coloured graph is $O(\sqrt{\log n})$ and every two components are a distance at least 20 apart.*

Proof. Since the edges coloured by ϕ_P are at distance at most 4 from P, for every $P \in \mathcal{P}$, and the geodesics in $\mathcal{P}$ are at distance at least 50 from each other, any two coloured components in ϕ are at distance at least 20 from each other. In particular, the edges coloured by distinct colourings ϕ_P are pairwise vertex-disjoint. Thus, because (E1) holds for every ϕ_P, it also holds for ϕ.

For (E3), let C be a cycle with no red edges or no blue edges. If C contains edges coloured by at most one ϕ_P then C has two consecutive uncoloured edges, due to ϕ_P satisfying (E3). Otherwise, it has edges from two coloured components, so by the fact that they are at distance at least 20 from each other, C has two consecutive uncoloured edges. So (E3) holds for ϕ.

Finally, since $|P| = O(\sqrt{\log n})$ and the edges coloured by ϕ_P are at distance at most 4 from P, every coloured component has size $O(\sqrt{\log n})$. $\qquad\square$

We next show that the colouring ϕ can be extended to a colouring that also satisfies (E4). The proof of the following lemma will be postponed to the next subsection.

Lemma 5.4. *Let G be a cubic graph and χ a partial colouring satisfying (E1) and (E3).*

Then there is a partial colouring χ' which extends χ, satisfies (E1), (E3) and (E4) and, additionally, every χ'-coloured edge is within distance 2 of a χ-coloured edge, and connected to it via coloured edges.

It is now easy to prove Lemma 4.2.

Proof of Lemma 4.2 using Lemmas 4.3 and 5.4. Let $\mathcal{P}$ be a collection of $\Omega(n^{0.9999})$ geodesics of length $10^{11}\sqrt{\log n}$ that are at distance at least 50 from each other; such a collection exists by Claim 5.1. For each $\ell \in [3, 10^{10}\sqrt{\log n}]$, pick $n^{0.999}$ geodesics in $\mathcal{P}$ (such that each geodesic is picked at most once), and apply Lemma 4.3 to P to find a partial colouring Φ_P that forms a blue ℓ-gadget, its coloured edges are at distance at most 4 from P, and which satisfies (E1) and (E3). Consider the union ϕ of these partial colourings, like in Definition 5.2; so ϕ satisfies (E1) and (E3), its coloured components have size $O(\sqrt{\log n})$, and are at distance at least 50 from each other. Now apply Lemma 5.4 to find a partial colouring ϕ' that extends ϕ, satisfies (E1), (E3), and (E4), and whose newly coloured edges are at distance at most 2 from previously coloured edges. Thus coloured components in ϕ' have size $O(\sqrt{\log n})$ and are at distance at least 10 from each other, as required for (E2). The partial colouring ϕ' thus satisfies the requirements of the lemma. $\qquad\square$

5.1 Getting property (E4)

The goal of this section is to prove Lemma 5.4, which allows us to extend an arbitrary colouring satisfying (E1) and (E3) in order to get a new colouring satisfying (E4). Getting property (E4) turns out to be essentially equivalent to asking for a colouring without cycles satisfying the following definition:

Definition 5.5. Let χ be a partial colouring of a graph G. Say that a cycle C is χ-(E4)-bad if it is odd, its edges are uncoloured and it contains at most two vertices of uncoloured degree 3 (and the remaining vertices have uncoloured degree 2).

We will also use the following definition for maintaining (E3).

Definition 5.6. Let χ be a partial colouring of a graph G. Say that a cycle C is χ-(E3)-bad if it contains only one colour, and does not contain two consecutive uncoloured edges.

We will often use the useful fact that χ-(E3)-bad cycles cannot pass through vertices with uncoloured degree 3 (as if they did, they would have consecutive uncoloured edges).

The following three lemmas allow us to get rid of particular kinds of bad cycles. Throughout this section, when we say that a colouring χ' extends χ, we mean that χ' is a partial colouring that agrees with χ, and each χ'-coloured connected component contains a χ-coloured connected component.

Lemma 5.7. *Let G be a cubic graph and χ a partial colouring satisfying* (E1) *and* (E3). *Let C be a χ-(E4)-bad cycle with two adjacent vertices of uncoloured degree 3.*

Then there is a partial colouring χ' which extends χ, satisfies (E1) *and* (E3), *has that all χ'-(E4)-bad cycles are χ-(E4)-bad, and has C not χ'-(E4)-bad. Additionally, every χ'-coloured edge touches C.*

Proof. Let $C = (c_1 \ldots c_k)$ with c_2, c_3 having uncoloured degree 3. For each i, let c_i' be the unique neighbour of c_i, different from c_{i-1}, c_{i+1}. Since C is χ-(E4)-bad, we know that c_i, with $i \notin \{2,3\}$ has uncoloured degree 2, so $c_i c_i'$ is coloured. Without loss of generality, suppose that $c_1 c_1'$ is red. We construct a colouring χ' that will satisfy the lemma as follows:

(a) Suppose that c_2' has uncoloured degree 3. Colour $c_1 c_2$ blue, $c_2 c_3$ red. For $i = 4, \ldots, k$, colour $c_i c_{i+1}$ by the opposite colour of $c_i c_i'$.

(b) Suppose that c_2' touches a red edge $c_2' c_2''$. Colour $c_1 c_2$ blue, $c_2 c_2'$ blue, $c_2 c_3$ red. For $i = 4, \ldots, k$, colour $c_i c_{i+1}$ by the opposite colour of $c_i c_i'$.

(c) Suppose that c_2' has touches a blue edge $c_2' c_2''$. Colour $c_1 c_2$ blue, $c_2 c_2'$ red.

To see that χ' satisfies (E1), we need to check the property at all vertices whose edges changed colour: In case (a), $c_1, c_2, c_4, \ldots, c_k$ all see both colours, while c_3 has uncoloured degree 2. In case (b), $c_2', c_1, c_2, c_4, \ldots, c_k$ all see both colours, while c_3 has uncoloured degree 2. In case (c), c_1, c_2, c_2' all see both colours.

To see that χ' satisfies (E3), suppose for contradiction that we have a χ'-(E3)-bad cycle B. Then B must contain some edge whose colour changed in χ' (since χ satisfied (E3)). Note that B contains $c_i c_{i+1}$ for some i (the only other edge whose colour can change between χ and χ' is $c_2 c_2'$. But then B would also contain another edge through c_2 i.e. the edge $c_2 c_1$ or $c_2 c_3$). It is impossible that $B = C$ (in (a) and (b), C contains both colours on $c_1 c_2, c_2 c_3$. In (c), C contains adjacent uncoloured edges $c_2 c_3, c_3 c_4$), so B must leave C at some point i.e. for some i, B contains both $c_i c_{i+1}$ and $c_i c_i'$. This must mean that $i = 2$ since for all other i, we either have that $c_i c_{i+1}$ and $c_i c_i'$ have opposite colours or $c_i c_{i+1}$ and $c_i c_i'$ are both uncoloured. Finally, for $i = 2$ note that in case (a) B cannot go through c_2' due to it having uncoloured degree 3, in case (b) C' we cannot have $c_2 c_3, c_2 c_2' \in B$ due to them having opposite colours, and in case (c) B cannot go through c_3 due to it having uncoloured degree 3.

Suppose that there is some χ'-(E4)-bad cycle B which is not χ-(E4)-bad. Then it would have to go through some vertex x whose uncoloured degree is 3 in χ and 2 in χ'. It is easy to check that the only such vertex is c_3, which, in cases (a), (b), has the edge $c_2 c_3$ coloured and $c_3 c_3', c_3 c_4$ uncoloured (and in Case (c) still has uncoloured degree 3 in χ'). Then B goes through $c_3 c_3', c_3 c_4$ also. However, in cases (a), (b) the vertex c_4 has uncoloured degree 1, so it cannot be part of a χ'-(E4)-bad cycle. □

Lemma 5.8. *Let G be a cubic graph and χ a partial colouring satisfying* (E1), (E3). *Let C be a χ-(E4)-bad cycle whose vertices touch both red and blue edges.*

Then there is a partial colouring χ' which extends χ, satisfies (E1) *and* (E3), *has that all χ'-(E4)-bad cycles are χ-(E4)-bad, and has C not χ'-(E4)-bad. Additionally, every χ'-coloured edge touches C.*

Proof. If C contains two adjacent vertices of uncoloured degree 3, then we are done by Lemma 5.7, so suppose that this does not happen. Let $C = (c_1 \ldots c_k)$. For each i, let c'_i be the unique neighbour of c_i which is not c_{i-1} or c_{i+1}. Define the following colouring χ' that will satisfy the lemma: for each i with $c_i c'_i$ coloured in χ, colour $c_i c_{i+1}$ by the opposite colour of $c_i c'_i$.

For (E1), note that since χ satisfied (E1), the property can only fail at some c_i with $c_i c_{i+1}$ or $c_{i-1} c_i$ coloured. If $c_i c_{i+1}$ was coloured, then $c_i c_{i+1}, c'_i c_i$ have different colours. If $c_{i-1} c_i$ was coloured and $c_i c_{i+1}$ was not coloured, then $c'_i c_i$ is uncoloured and so we have that $c_{i-1} c_i$ is the only χ'-coloured edge through c_i.

For (E3), suppose that we have some χ'-(E3)-bad cycle B. Since χ satisfied (E3), B must contain some edge $c_i c_{i+1}$ which was coloured in χ'. Note that it is impossible that $B = C$, since C contains both a red and a blue edge (to see this first recall that by the lemma's assumptions we have some i, j with $c'_i c_i$ red and $c'_j c_j$ blue. By definition of χ', we have $c_i c_{i+1}$ blue and $c_j c_{j+1}$ red). Thus B must contain the sequence $c'_a c_a c_{a+1} \ldots c_b c'_b$ for some a, b. But then either $c'_a c_a, c_a c_{a+1}$ are both uncoloured, or they have opposite colours. In both cases, we get a contradiction to B being χ'-(E3)-bad.

Notice that the only new vertices of uncoloured degree 2 are c_i with $c'_i c_i$ uncoloured and $c'_{i-1} c_{i-1}$ coloured in χ. Any χ'-(E4)-bad cycle C' that is not χ-(E4)-bad must pass through such a vertex and so contain the sequence $c'_i c_i c_{i+1}$. If $c'_{i+1} c_{i+1}$ is coloured in χ, then we have that c_{i+1} has uncoloured degree 1 in χ' (and so could not be contained in the uncoloured cycle C). On the other hand, if $c'_{i+1} c_{i+1}$ is uncoloured in χ, then we have two consecutive vertices c_i, c_{i+1} of uncoloured degree 3 in χ (which we have assumed does not happen). $\qquad\square$

Lemma 5.9. *Let G be a cubic graph and χ a partial colouring satisfying* (E1) *and* (E3). *Let C be a χ-(E4)-bad cycle with two consecutive red edges.*

Then there is a partial colouring χ' which extends χ, satisfies (E1), (E3), *has that all χ'-(E4)-bad cycles are χ-(E4)-bad, and has C not χ'-(E4)-bad. Additionally, every χ'-coloured edge touches C.*

Proof. By Lemma 5.8, we can assume that C touches only red edges. Let $C = (c_1 \ldots c_k)$. For each i, let c'_i be the unique neighbour of c_i outside C. Colour $c_1 c_2$ blue to get a colouring χ'. We claim that χ' satisfies the lemma.

For (E1), note that since χ satisfied (E1), the property can only fail at c_1 or c_2. But both c_1, c_2 touch edges of both colours.

For (E3), suppose that we have some χ'-(E3)-bad cycle B. Since χ satisfied (E3), B must contain $c_1 c_2$. Note that since $c'_1 c_1, c'_2 c_2$ are red, and $c_1 c_2$ is blue, B must contain the sequence $c_1 c_2 c_3$ and so contain one of the sequences $c_1 c_2 c_3 c_4$ or $c_1 c_2 c_3 c'_3$. The sequence $c_1 c_2 c_3 c_4$ contains two consecutive uncoloured edges. If $c'_3 c_3$ is uncoloured, then $c_1 c_2 c_3 c'_3$ contains two consecutive uncoloured edges. If $c'_3 c_3$ is coloured, then it must be red, and so $c_1 c_2 c_3 c'_3$ contains both red and blue edges.

To see that all χ'-(E4)-bad cycles are χ-(E4)-bad, notice that there are no new vertices of uncoloured degree 2. $\qquad\square$

Combining the above three lemmas allows us to eliminate any given bad cycle.

Lemma 5.10. *Let G be a cubic graph and χ a partial colouring satisfying* (E1) *and* (E3). *Let C be a χ-*(E4)*-bad cycle.*

Then there is a partial colouring χ' which extends χ, satisfies (E1) *and* (E3)*, has that all χ'-*(E4)*-bad cycles are χ-*(E4)*-bad, and has C not χ'-*(E4)*-bad. Additionally, every χ'-coloured edge touches C.*

Proof. If C touches both colours, then we are done by Lemma 5.8. So C touches just one colour. Without loss of generality, C only touches red edges. If $|C| \geq 5$, then it will have two consecutive vertices touching red edges (since in a χ-(E4)-bad cycle there are at most two vertices which do not touch coloured edges). Thus, when $|C| \geq 5$, Lemma 5.9 applies to give what we want. The remaining case, not covered by Lemma 5.9, is when $|C| = 3$ and C has only one vertex touching red edges (recalling that a χ-(E3)-bad cycle is odd by definition). Then the remaining two vertices are adjacent vertices of uncoloured degree 2, and so Lemma 5.7 gives what we want. $\square$

By iterating the previous lemma, we can get rid of all χ-(E4)-bad cycles by extending a colouring.

Lemma 5.11. *Let G be a cubic graph and χ_0 a partial colouring satisfying* (E1) *and* (E3).

Then there is a partial colouring χ' which extends χ, satisfies (E1) *and* (E3)*, and has no χ'-*(E4)*-bad cycles. Additionally, every χ'-coloured edge is within distance 2 of a χ-coloured edge.*

Proof. Let $D_1, \ldots, D_t$ be the χ_1-(E4)-bad cycles listed in some order. Construct colourings $\chi_2, \chi_3, \ldots$ as follows. For each i, if D_i is χ_i-(E4)-bad, then apply Lemma 5.10 to D_i, χ_i to get a colouring χ_{i+1} extending χ_i. Otherwise, if D_i is not χ_i-(E4)-bad, then set $\chi_{i+1} = \chi_i$. In either case we have D_i is not χ_{i+1}-(E4)-bad and that all χ_{i+1}-(E4)-bad cycles are χ_i-(E4)-bad. Putting these together gives that $D_1, \ldots, D_t$ are not χ_{t+1}-(E4)-bad and that all χ_{t+1}-(E4)-bad cycles are χ_1-(E4)-bad i.e. there are no χ_{t+1}-(E4)-bad cycles at all.

For the "additionally" part, note that all χ_{t+1}-coloured edges that are not χ_1-coloured touch some D_i. In each D_i, there are at most two vertices of uncoloured degree 3 in χ_1, and so every vertex is within distance 2 of a χ_1-coloured edge. $\square$

The following lemma characterises graphs which do not contain odd cycles with at most three degree 3 vertices.

Lemma 5.12. *Let G be a subcubic graph in which every odd cycle contains at least three vertices of degree 3. Then we can edge-decompose G into H_1, H_2, H_3 where H_2, H_3 are vertex disjoint and:*

(1) *H_1 is a subdivision of a simple graph with all degrees $\in \{1, 3\}$.*

(2) *H_2 is a union of disjoint even cycles, each of which has at most two vertices in H_1.*

(3) *H_3 is a collection of subdivisions of multiedges with multiplicity 3, each of which has no vertices in H_1.*

Proof. Repeatedly contract vertices of degree 2 for as long as possible without creating loops. The result is a subcubic multigraph H in which the only degree 2 vertices occur as the endpoint of some multiedge (of multiplicity 2). Let $G_2, G_3 \subseteq G$ be the unions of multiedges of multiplicity 2 and 3, respectively. Since $\Delta(G) \leq 3$, we have that the multiedges in $G_2 \cup G_3$ are all vertex-disjoint. Let $G_1 = G \setminus E(G_2 \cup G_3)$, noting that G_1 is a simple graph. Also, each multiedge in G_2 can trivially intersect at most two vertices of G_1. We claim that G_1 has no vertices of degree 2. Indeed suppose that v is such a vertex. If $d_G(v) = 2$, then v is part of some multiedge which shows that $d_{G_1}(v) = 0$. Otherwise, if $d_G(v) = 1$ or 3, note that we either have $d_{G_1}(v) = d_G(v)$ or $d_{G_1}(v) \leq d_G(v) - 2$. So in all cases, we cannot get $d_{G_1}(v) = 2$.

Uncontract G_1, G_2, G_3 to get H_1, H_2, H_3 decomposing G. Properties (1) and (3) are immediate from the construction of G_1, G_3. For (2), it is immediate that H_2 is a union of cycles each of has at most vertices in H_1. The property "every odd cycle contains at least three vertices of degree 3" of G shows that all the cycles in H_2 are even. □

We are now ready to prove Lemma 5.4.

Proof of Lemma 5.4. Let χ' be the colouring given by Lemma 5.11. Properties (E1) and (E3) are immediate from that lemma, as is "every χ'-coloured edge is within distance 2 of a χ-coloured edge". For (E4), recall that Lemma 5.11 and the definition of "χ'-(E4)-bad cycle", tell us that all uncoloured odd cycles contains at least three vertices of uncoloured degree 3. By Claim 5.3, each coloured component in χ' has size $O(\sqrt{\log n})$, and by Lemma 5.12, the χ'-uncoloured edges can be partitioned into graph H_1, H_2, H_3 having the three properties in Lemma 5.12. Notice that each component K in H_2 or H_3 has all but at most two of its vertices incident to a coloured edge in χ. Because coloured components in χ are far away from each other, all the coloured edges touching K belong to the same coloured component, and thus $|K| = O(\sqrt{\log n})$. □

6 Explicit structure of gadgets

From this section onward, we will focus on the proof of Lemma 4.3, which will be the final element of the proof of the main theorem. Recall that in Section 4 we gave an abstract definition of a gadget. In this section, we will explicitly describe the gadgets that we find in our graph. In the two subsequent sections, we show that such gadgets can be found in small balls around sufficiently long geodesics: in Section 7 we will prove Lemma 4.3 for geodesics whose vertices have no common neighbours (which is actually Lemma 4.4), and in Section 8 we will prove Lemma 4.3 for the remaining case. Throughout the next three sections, we describe the gadgets both in words and in a figure, with the expectation that the figure will be much easier the understand.

Definition 6.1 (Type I gadgets)**.** Let $\ell \geq 3$. An *ℓ-gadget of Type I* is a graph H, along with two red-blue colourings χ_1 and χ_2, defined as follows. The graph H consists of paths Q_0, Q_1, Q_2 (where the ends of Q_i are denoted Q_i^+, Q_i^-) and edges $e_1, \ldots, e_5, f_1, \ldots, f_4$, satisfying the following properties (see Figure 2).

- $Q_0 = (q_0, \ldots, q_{\ell+2})$, so Q_0 has length $\ell + 2$,

- Q_1 and Q_2 are vertex-disjoint and have the same length,

- $Q_1^- = q_1$, $Q_2^- = q_3$, and $Q_i \setminus \{Q_i^-\}$ is vertex-disjoint of Q_0, for $i \in [2]$,

- the edges e_i and f_j are edge-disjoint of Q_0, Q_1, Q_2, for $i \in [5]$ and $j \in [4]$,

- e_1, e_2 are distinct edges that contain q_0; e_3 contains q_2; e_4, e_5 are distinct edges containing $q_{\ell+2}$; f_1, f_2 are distinct edges containing Q_1^+; and f_3, f_4 are distinct edges containing Q_2^+.

We remark that an edge e_i or f_j could intersect vertices of $Q_0 \cup Q_1 \cup Q_2$ unless specified otherwise (e.g. e_1 cannot intersect q_1), and similarly two elements in $(e_1, \ldots, e_5, f_1, \ldots, f_4)$ could be the same edge or be intersecting edges, unless specified otherwise.

Let χ_1 to be the red-blue colouring of H, defined as follows.

$$\chi_1(e) = \begin{cases} \text{red} & e \in Q_1 \cup Q_2 \text{ or } e \in \{e_1, \ldots, e_5, q_1 q_s\}, \\ \text{blue} & e \in Q_0 \setminus \{q_1 q_2\} \text{ or } e \in \{f_1, \ldots, f_4\}. \end{cases}$$

Let χ_2 be the red-blue colouring of H, obtained from χ_1 by swapping the colours of $q_1 q_2$ and $q_2 q_3$. Namely,

$$\chi_2(e) = \begin{cases} \text{red} & e \in Q_1 \cup Q_2 \text{ or } e \in \{e_1, \ldots, e_5, q_2 q_3\}, \\ \text{blue} & e \in Q_0 \setminus \{q_2 q_3\} \text{ or } e \in \{f_1, \ldots, f_4\}. \end{cases}$$

The colouring χ_1 The colouring χ_2
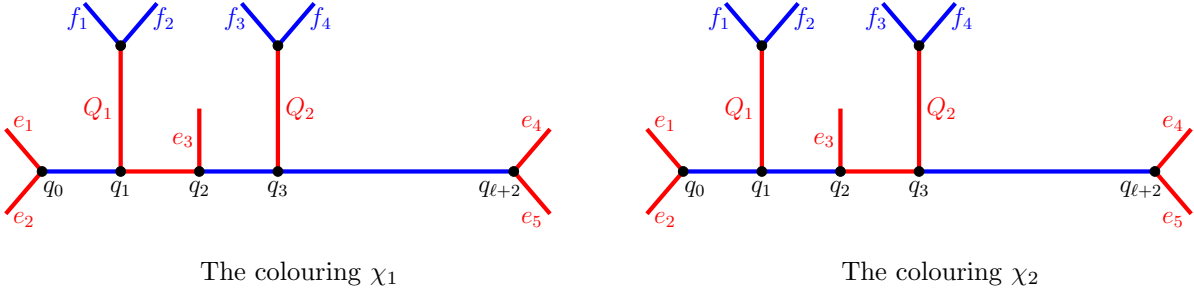
Figure 2: The two colourings of a Type I gadget

Observation 6.2. *An ℓ-gadget of Type I is a blue ℓ-gadget, for $\ell \geq 3$.*

Proof. Consider an ℓ-gadget of Type I, defined by H, χ_1, χ_2. Let G be a red-blue cubic graph, whose monochromatic components are paths, and suppose that G contains a copy H' of (H, χ_1). Consider the red-blue coloured graph G' obtained from G by replacing the colouring of H' by (H, χ_2). Let $\mathcal{P}_b, \mathcal{P}_r$ be the collections of blue and red components in G, and let $\mathcal{P}'_b, \mathcal{P}'_r$ be the collections of blue and red components in G'. It is easy to check that

$$\mathcal{P}'_b = (\mathcal{P}_b \setminus \{q_0 q_1, \, q_2 \ldots q_{\ell+2}\}) \cup \{q_0 q_1 q_2, \, q_3 \ldots q_{\ell+2}\}$$
$$\mathcal{P}'_r = (\mathcal{P}_r \setminus \{Q_2, \text{ the red component in } G \text{ containing } Q_1 q_1 q_2 e_3\})$$
$$\cup \{Q_1, \text{ the red component in } G' \text{ containing } Q_2 q_3 q_2 e_3\}.$$

Since Q_1 and Q_2 have the same length, the red component in G containing $Q_1 q_1 q_2 e_3$ has the same length

27

as the red component in G' containing $Q_2 q_3 q_2 e_3$. Thus $r_G(P_k) = r_{G'}(P_k)$ for every k, and

$$b_{G'}(P_k) = \begin{cases} b_G(P_k) & k \neq \ell, \ell - 1, 1, 2 \\ b_G(P_k) + 1 & k \in \{\ell - 1, 2\} \\ b_G(P_k) - 1 & k \in \{\ell, 1\}. \end{cases}$$

Notice also that (H, χ_1) and (H, χ_2) differ on exactly two edges, and that (H, χ_1) has a maximal blue path of length ℓ whose ends are incident with two red edges. It follows that (H, χ_1, χ_2) is a blue ℓ-gadget. □

Definition 6.3 (Type II gadgets). Let $\ell \geq 3$. An ℓ-gadget of Type II is a graph H, along with two red-blue colourings χ_1 and χ_2 of H, defined as follows. The graph H consists of paths Q_0, Q_1 (where the ends of Q_1 are denoted Q_1^+, Q_1^-) and edges $e_1, \ldots, e_5$, such that

- $Q_0 = (q_0 \ldots q_{\ell+2})$, so Q_0 has length $\ell + 2$,

- $Q_1^- = q_1$ and $Q_1^+ = q_3$, and $Q_1 \setminus \{q_1^+, q_1^-\}$ is vertex disjoint of Q_0,

- the edges e_i are edge-disjoint of Q_0, Q_1, for $i \in [5]$,

- e_1, e_2 are distinct edges containing q_0; e_3 contains q_2; and e_4, e_5 are distinct edges containing $q_{\ell+2}$.

Let χ_1 be the red-blue colouring of H, defined as follows.

$$\chi_1(e) = \begin{cases} \text{red} & e \in Q_1 \text{ or } e \in \{e_1, \ldots, e_5, q_1 q_2\}, \\ \text{blue} & e \in Q_0 \setminus \{q_1 q_2\}. \end{cases}$$

Let χ_2 be the red-blue colouring of H, obtained from χ_1 by swapping the colours of $q_1 q_2$ and $q_2 q_3$. Namely,

$$\chi_2(e) = \begin{cases} \text{red} & e \in Q_1 \text{ or } e \in \{e_1, \ldots, e_5, q_2 q_3\}, \\ \text{blue} & e \in Q_0 \setminus \{q_2 q_3\}. \end{cases}$$
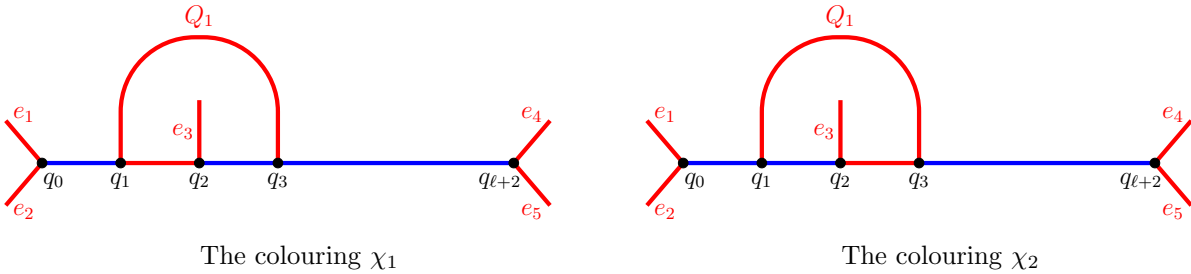
Figure 3: The two colourings of a Type II gadget

The following observation can be proved similarly to Observation 6.2; we omit the details.

Observation 6.4. *An ℓ-gadget of Type II is a blue ℓ-gadget, for $\ell \geq 3$.*

28

7 Geodesic with no common neighbours (Lemma 4.4)

In this section we prove Lemma 4.4, which finds a gadget in a small ball around a sufficiently long geodesic whose vertices do not have common neighbours (see Lemma 7.3).

Recall that here we assume that P is a geodesic of length $\ell + 36$, no two of whose vertices have a common neighbour outside of P. Throughout this section, fix such a geodesic P. Write $L = \ell + 36$ and $P = (p_0 \ldots p_L)$, and denote by p'_i the unique neighbour of p_i outside of P, for $i \in [L-1]$; so the p_i's are distinct.

Claim 7.1. *One of the following holds.*

    I. *There exists $s \in [10, L - \ell - 10]$ such that p'_{s-1} and p'_{s+1} have no common neighbours.*

    II. *For every $i \in [12, L - \ell - 12]$, the vertices p'_{i-1} and p'_{i+1} have a common neighbour in $\{p'_{i-2}, p'_i, p'_{i+2}\}$, and $\{p'_j : j \in [14, L - \ell - 14]\}$ induces no cycles.*

    III. *There exists $s \in [12, L - \ell - 10]$ such that p'_{s-1} and p'_{s+1} have a unique common neighbour w, and there is no i such that $w = p'_i$.*

Proof. Assume that Case **I** does not hold; then p'_{i-1} and p'_{i+1} have a common neighbour for all $i \in [10, L - \ell - 10]$. We claim that if there is a pair p'_{i-1}, p'_{i+1}, with $i \in [12, L - \ell - 12]$, with two common neighbours, then Case **III** must hold. Indeed, suppose that for some $i \in [12, L - \ell - 12]$ the vertices p'_{i-1} and p'_{i+1} have two common neighbours x, y. Then, since p'_{i+1} and p'_{i+3} also have a common neighbour, it must be either x or y, as otherwise p'_{i+1} would have degree 4. The same holds for p'_{i-3} and p'_{i-1}. Without loss of generality, x is joined to p'_{i+3}. But then $x \notin \{p'_j : j \in [L-1]\}$, because otherwise x would have degree at least 4. Moreover, p_{i+1} and p_{i+3} have no common neighbour other than x (because that neighbour would have to be y, making y's degree 4). Thus Case **III** holds for $s = i + 2$.

So now we are left with the case where all pairs of vertices p'_{i-1}, p'_{i+1}, with $i \in [12, L - \ell - 12]$, have exactly one common neighbour. If for one such i the common neighbour is not in $\{p'_j : j \in [L-1]\}$, then Case **III** holds, so suppose otherwise. Namely, for $i \in [12, L - \ell - 12]$, the unique common neighbour of p'_{i-1} and p'_{i+1} is p'_j, for some $j \in [L-1]$. Clearly, $j \notin \{i-1, i+1\}$. Also, j is not larger than $i + 2$ or smaller than $i - 2$ because P is a geodesic. Thus $j \in \{i - 2, i, i + 2\}$. It follows that, for every $i \in [14, L - \ell - 14]$, the vertex p'_i has exactly two neighbours in $\{p'_{i-3}, p'_{i-1}, p'_{i+1}, p'_{i+3}\}$ (because it has a common neighbour with both p'_{i+2} and p'_{i-2}, and that cannot be the same neighbour because such a neighbour would have degree at least 4). We show that $\{p'_j : j \in [14, L - \ell - 14]\}$ has no cycles. Suppose to the contrary that C is a cycle in this set, and let j be the least index of a vertex in C. Then p'_j is joined to both p'_{j+1} and p'_{j+3}, and p'_{j+1} has is joined to either p'_{j+2} or p'_{j+4}. It follows that p'_{j-2} and p'_j have no common neighbour (otherwise some vertex would have degree larger than 3), a contradiction. Thus Case **II** holds. $\square$

We now show how to proceed in each of the above three cases.

Algorithm 7.2. *Let P be as above.*

    I. *If Case **I** of Claim 7.1 holds, fix s for which it holds, and do the following.*

- *Colour the edges $p_{s-3}p_{s-2}$, $p_{s-1}p_s$ and $p_{s+\ell}p_{s+\ell+1}$ red, and colour all other edges in the path $(p_{s-10}\ldots p_{s+\ell+10})$ blue.*

- *Colour the edges $p_{s-3}p'_{s-3}$ and $p_{s+\ell}p'_{s+\ell}$ blue, and colour all other edges $p_ip'_i$ with $i \in [s-9, s+\ell+9]$ red.*

- *Denote the two uncoloured neighbours of p'_{s-1} by u, u' and the two uncoloured neighbours of p'_{s+1} by w, w'. Colour the edges $p'_{s-1}u, p'_{s-1}u', p_{s+1}w, p_{s+1}w'$ blue (note that these are three or four edges, depending on whether or not $p_{s-1}p_{s+1}$ is an edge).*

- *Colour all uncoloured edges with both end points in $\{u, u', w, w'\}$ red, unless they contain a cycle (of length 3 or 4), in which case we colour one edge in a cycle, with one end in $\{u, u'\}$ and one end in $\{w, w'\}$, blue, and the rest red.*

- *This gives a blue ℓ-gadget of type II, with $Q_0 = (p_{s-2}\ldots p_{s+\ell})$, $Q_1 = p_{s-1}p_{s-1'}$, and $Q_2 = p_{s+1}p'_{s+1}$; $e_1 = p_{s-3}p_{s-2}$, $e_2 = p_{s-2}p'_{s-2}$, $e_3 = p_sp'_s$, $e_4 = p_{s+\ell}p_{s+\ell+1}$, and $e_5 = p_{s+\ell}p'_{s+\ell}$; $f_1 = p'_{s-1}u$, $f_2 = p'_{s-1}u'$, $f_3 = p'_{s+1}w$ and $f_4 = p'_{s+1}w'$.*
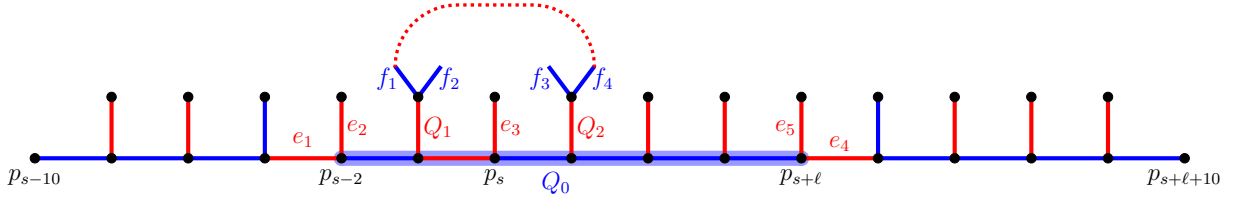
Figure 4: Case **I**

II. *If Case **II** of Claim 7.1 holds, take $s = 18$, and do the following.*

- *Colour the edges $p_{s-3}p_{s-2}$, $p_{s-1}p_s$ and $p_{s+\ell}p_{s+\ell+1}$ red, and colour all other edges in the path $(p_{s-2}\ldots p_{s+\ell})$ blue.*

- *Colour all edges $p_ip'_i$, with $i \in [s-2, s+\ell]$, red.*

- *Colour the four edges touching p'_{s-1} and p'_{s+1} and not touching P blue.*

- *This gives a blue ℓ-gadget of type II, with $Q_0 = (p_{s-2}\ldots p_{s+\ell})$, $Q_1 = p_{s-1}p'_{s-1}$, and $Q_2 = p_{s+1}p'_{s+1}$; $e_1 = p_{s-3}p_{s-2}$, $e_2 = p_{s-2}p'_{s-2}$, $e_3 = p_sp'_s$, $e_4 = p_{s+\ell}p_{s+\ell+1}$, and $e_5 = p_{s+\ell}p'_{s+\ell}$; f_1, f_2 are the two blue edges at p'_{s-1}, and f_3, f_4 are the two blue edges at p'_{s+1}.*
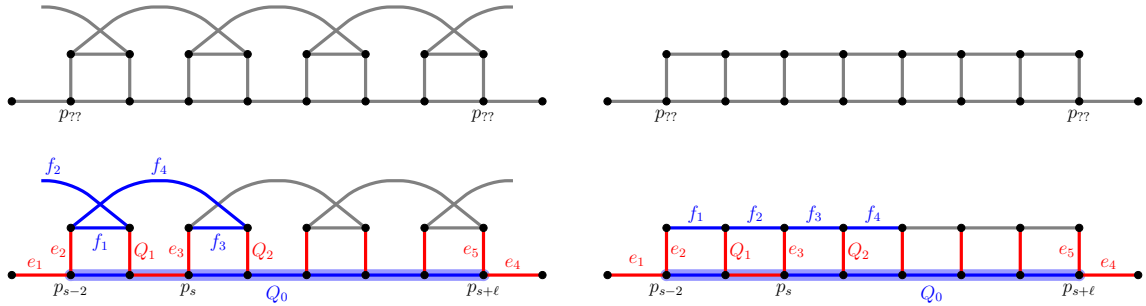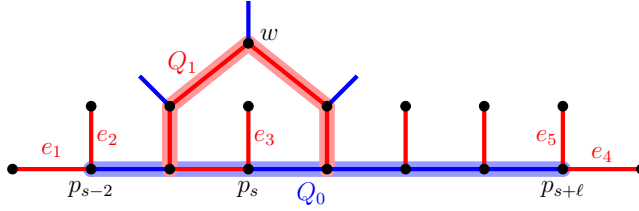

Figure 5: Case **II**

30

III. *If Case **III** of Claim 7.1 holds, fix s for which it does, and do the following.*

- *Colour the edges $p_{s-3}p_{s-2}$, $p_{s-1}p_s$ and $p_{s+\ell}p_{s+\ell+1}$ red, and colour all other edges in the path $(p_{s-2} \ldots p_{s+\ell})$ blue.*

- *Colour all edges $p_i p_i'$, with $i \in [s-2, s+\ell]$, red.*

- *Colour the edges $p_{s-1}'w$ and $p_{s+1}'w$ red.*

- *Repeat the following steps as long as possible:*

  - *Colour all uncoloured edges incident with a vertex, whose red degree is 2, blue.*
  - *Colour all uncoloured edges incident with a vertex, whose blue degree is 2, red.*

- *This yields a blue ℓ-gadget of type 2, with $Q_0 = (p_{s-2} \ldots p_{s+\ell})$ and $Q_1 = (p_{s-1}p_{s-1}'wp_{s+1}'p_{s+1})$; $e_1 = p_{s-3}p_{s-2}$, $e_2 = p_{s-2}p_{s-2}'$, $e_3 = p_s p_s'$, $e_4 = p_{s+\ell}p_{s+\ell+1}$ and $e_5 = p_{s+\ell}p_{s+\ell}'$.*

Figure 6: Case **III**

Lemma 7.3. *The partial colouring ϕ obtained by Algorithm 7.2 satisfies the following, and thus satisfies the requirements of Lemma 4.4.*

(1) *all coloured edges are at distance at most 3 from the geodesic P and form a connected subgraph,*

(2) *if a vertex u is incident to at least two coloured edges, then at least one of them is red and at least one of them is blue,*

(3) *if C is a cycle in G containing no red edges or no blue edges, then C has two consecutive edges that are not coloured by ϕ_3.*

Proof. We call a cycle that has no red edges or no blue edges a *single-colour cycle* (note that such a cycle can have uncoloured edges, and so this is not the same as a monochromatic cycle). We call a vertex with at least two coloured edges, all in the same colour, a *single-colour* vertex.

We first point out that the last step in Part **III** of Algorithm 7.2 does not last very long, leaves no single-colour vertices, and the edges it colours are not involved in single-colour cycles. The key observation is that initially three blue edges are added, not all sharing a vertex (if they did all share a vertex, then p_{s-1}' and p_{s+1}' would have two common neighbours, contrary to the specification of the Case **III**), and from then onward, there is always at most one single-colour vertex. Since each iteration in this step colours exactly one edge, this means that such a single-colour vertex has coloured degree exactly 2. Thus, at the end of the step, there are no single-colour vertices. In addition, since every newly coloured edge has one endpoint with degree 2 in the opposite colour, none of them is involved in a single-colour cycle. It remains to show that the step has few iterations. If no two of the initial three blue edges share a vertex

31

then the process stops immediately. Otherwise, since we started with only two vertices with blue degree at least 1 and colour degree at most 2, it is easy to check that during the last step of **III** we will repeat the second item at most twice. This means that the process has at most four iterations. Since each iteration involves at least one edge which was already coloured in previous steps of the algorithm, all edges coloured in this step are at distance at most 3 from P.

Items (1) and (2) are immediate in Case **I** and **II** and follow from the last paragraph otherwise.

It remains to prove (3). For this, first note that if an endpoint of a monochromatic path has both of its other edges coloured in the opposite colour, then this path cannot participate in a single-coloured cycle. Another important observation is that, since P is a geodesic, there are no edges of the form $p'_j p'_{j+r}$ when $r \geq 4$. So connecting two such vertices will require more than one edge. We consider each case separately.

Case **I**.   All red edges, except for those with two ends in $\{u, u', w, w'\}$, belong to a red path with an end with blue degree 2. Since all the edges in $\{u, u', w, w'\}$ are coloured, and do not contain a red cycle (since this is a graph with maximum degree 2 and four vertices, there is at most one cycle and such a cycle is not red due to the fourth step in **I** in Algorithm 7.2. Thus any C with no blue edges has all its red edges in $\{u, u', w, w'\}$ and has at least one vertex $v \notin \{u, u', w, w'\}$, showing that it has two consecutive uncoloured edges. Therefore, it remains to consider cycles with no red edges. Suppose that C is such a cycle. If it contains the left-most blue path (starting from p'_{s-3}), then, because p_{s-10} is at distance at least 5 from all other blue edges which might participate in a single-colour cycle, C contains at least 5 consecutive uncoloured edges. A similar argument works if C contains the right-most path. So it remains to consider the case where C contains $f_1 f_2$ or $f_3 f_4$ and not other blue edges. But there is no cycle with vertices in $\bigcup_{i \in [4]} V(f_i)$ that has only blue edges (due to the fourth step in this case).

Case **II**.   If C is a cycle with blue edges but no red edges, then its blue edges are in $\{f_1, f_2, f_3, f_4\}$. By the specifications of Item **II**, there are no cycles in the set $\bigcup_{i \in [4]} V(f_i)$, and so C must contain a vertex outside of this set, showing that C has two consecutive uncoloured edges. If C has only red edges then it contains one (or both) of the paths $e_1 e_2$ and $e_3 e_4$ (and no other red edges). But if it contains only one of them, then it has length at least 4 (using the assumption on P), and if it contains both, then it has length at least 10 (because p_{s-2} and $p_{s+\ell}$ are at distance at least 5 from each other). Either way, it has two consecutive uncoloured edges.

Case **III**.   There is no cycle with no red edges. If C is a cycle with red edges but not blue edges, then it contains one or both of $e_1 e_2$ and $e_3 e_4$ (any red edge added in the last step has an end with two blue edges). As in the previous case, C contains two consecutive uncoloured edges. $\square$

8   Geodesic with common neighbours (Lemma 4.3)

This section will be devoted to proving Lemma 4.3, which finds a gadget in a small ball around a sufficiently long geodesic (see Lemma 8.18). Given Lemma 4.4, which does the same for geodesics whose vertices have no common neighbours, we only need to prove the former lemma for a geodesic P many pairs of its vertices have a common neighbour outside of P. For this, we define a partial colouring process of the gadget and its neighbourhood. The partial colouring process is divided into three stages, each will be described in a different subsection.

- **Stage I** (Section 8.1). Initial partial colouring. In this stage we partially colour a small ball around them according to Algorithm 8.1, yielding a long blue path and avoiding vertices with red degree 2 and blue degree 0 or vice versa. This colouring will be denoted ϕ_1. A lot of this work in this section will be devoted to the analysis of the colouring ϕ_1, particularly to showing that monochromatic degree 3 vertices only arise in a specific structure.

- **Stage II** (Section 8.2). Fixing monochromatic vertices. In this stage we modify the colouring ϕ_1 (in Algorithm 8.14) so as to avoid vertices of monochromatic degree 3, by recolouring some of the edges around each vertex of monochromatic degree 3 in ϕ_1. This colouring will be denoted ϕ_2.

- **Stage III** (Section 8.3). Creating a gadget. In this stage we will further modify the colouring ϕ_2 (in Algorithm 8.17) to create a copy of a type II gadget. This partial colouring will be denoted ϕ_3.

We will finish this section with the proof of Lemma 4.3.

8.1 Stage I. Initial partial colouring (Algorithm 8.1)

The purpose of this section is to present Algorithm 8.1 that generates a partial colouring ϕ_1 around a geodesic P, and prove key properties of ϕ_1. The algorithm starts in (S1) by colouring the edges of P blue, and then, in (S2) and for technical reasons, modifying this colouring slightly for any pair of consecutive vertices in P that have a common neighbour. Next, in (S3), (S4) and (S5), whenever we encounter a vertex with two edges of the same colour and a third edge uncoloured, we coloured the third edge with the opposite colour. The purpose of this should be clear: we want to make sure that each vertex with coloured degree at least 2 is incident to red and blue edges. At the end of the process, namely in (S6), we recolour some of the edges touching the ends of P, again for the same purpose but also keeping in mind an additional property (namely, (2) below).

Throughout the process we orient all edges as they are coloured, except for edges of P and the edges coloured at the last step of the process. These orientations are instrumental when proving various properties of the colouring. Additionally, we output a blue path P', modified from P in (S2). This path P' (or, more precisely, a modification of P' obtained in the next section) will be the basis of the gadget that we will eventually construct.

**Algorithm 8.1** (Initial partial colouring).

Input. *A geodesic P in a cubic graph G.*

Output. *A partial red-blue colouring ϕ_1 of G; an oriented graph D whose underlying graph is a subgraph of G; and a path P'.*
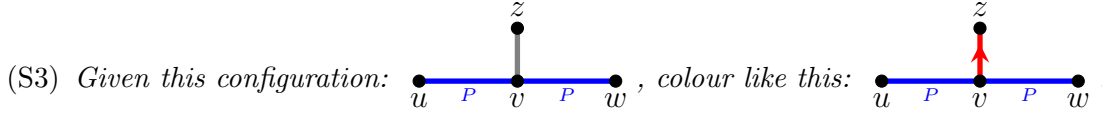
Initialisation. *Set $P' = P$; no edges of G are coloured; D is empty.*

Procedure. *First, perform (S1) below. Then, repeat (S2) as long as possible, and, afterwards, repeat (S3) as long as possible. Next, repeat (S4) and (S5) as long as possible, prioritising the former. Finally, repeat (S6).*
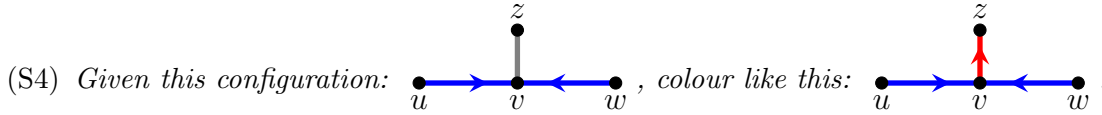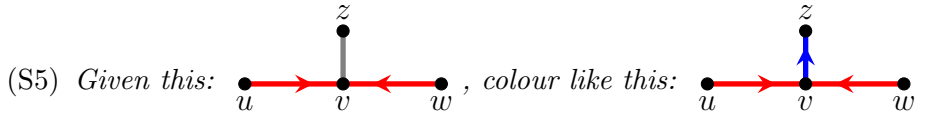
(S1) *Colour the edges in P blue.*

(S2) *Given this configuration:* *, recolour like this:* *.*
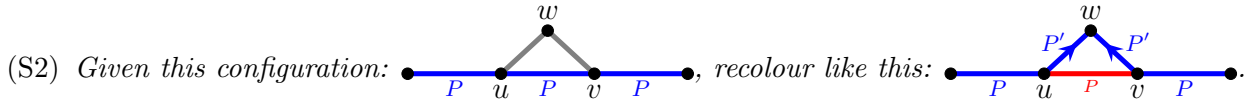
Namely, suppose: uv is a blue edge in the interior of P; uw, wv are uncoloured edges; $w \notin V(P)$.

Then: recolour uv red and uw, wv blue; add uw, vw to D; update $P' \leftarrow (P' \setminus \{uv\}) \cup \{uw, vw\}$.
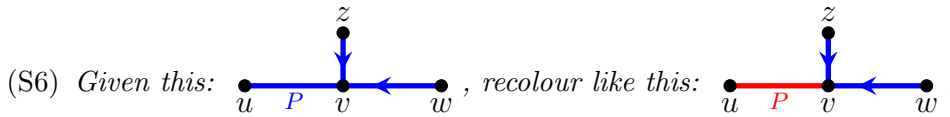
(S3) *Given this configuration:* *, colour like this:* *.*

Namely, suppose: uv, vw are blue edges in P; vz is uncoloured.

Then: colour vz red; add vz to D.

(S4) *Given this configuration:* *, colour like this:* *.*

Namely, suppose: uv, wv are directed blue edges; vz uncoloured.

Then: colour vz red; add vz to D.

(S5) *Given this:* *, colour like this:* *.*

Namely, suppose: uv, wv directed red edges; vz uncoloured.

Then: colour vz blue, add vz to D.

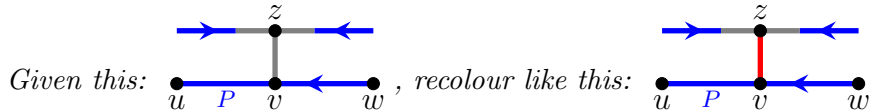(S6) *Given this:* *, recolour like this:* *.*

Namely, suppose: uv is a blue edge in P; wv and zv two directed blue edges.

Then: recolour uv red and set $P' \leftarrow P' \setminus \{uv\}$.

Given this: *, recolour like this:* *.*
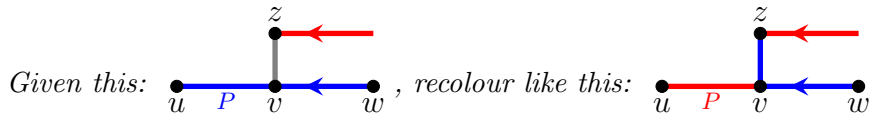
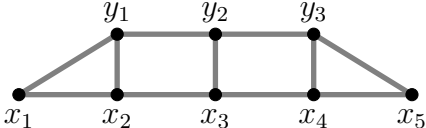*Namely, suppose: uv blue edge in P; wv directed blue edge; vz uncoloured edge; z has a red in-neighbour.*

Then: colour vz blue, recolour uv red and set $P' \leftarrow P' \setminus \{uv\}$.
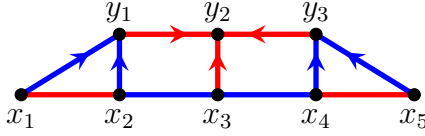
Given this: *, recolour like this:* *.*

Namely, suppose: uv blue edge in P; wv directed blue edge; vz uncoloured edge; each of the other two edges at z are either blue and directed into z or uncoloured.

Then: colour vz red.

In Lemma 8.3 below we state various properties of any partial colouring obtained by running Algorithm 8.1. One important property considers vertices with monochromatic degree 3. It turns out that they only appear given one specific structure, which we describe now.

Definition 8.2. Let $F_{\mathrm{exc}} =$. Namely, F is the graph on vertices $\{x_1, \ldots, x_5, y_1, y_2, y_3\}$ and edges $\{x_1x_2, x_2x_3, x_3x_4, x_4x_5, y_1y_2, y_2y_3, x_1y_1, x_2y_1, x_3y_2, x_4y_3, x_5y_3\}$.

Given a path P in a graph G, a copy of F_{exc} is *parallel* to P if $x_1, \ldots, x_5$ are consecutive vertices in P and $y_1, y_2, y_3 \notin V(P)$.

Let $F_{\mathrm{exc}}^* =$. Namely, F_{exc}^* is a red-blue coloured and partially oriented copy of F_{exc}, where the edges y_1y_2, x_3y_2, y_3y_2 are red and directed; $x_1y_1, x_2y_1, x_4y_3, x_5y_3$ are blue and directed; x_1x_2, x_4x_5 are red and undirected; and x_2x_3, x_3x_4 are blue and undirected.

Here is the main lemma of the subsection.

Lemma 8.3. *Let G be a cubic graph and let P be a geodesic in G. Consider the colouring and orientation process described in Algorithm 8.1, applied with P. Then*

(1) *if a vertex is incident to exactly two coloured edges then one of them is red and the other blue,*

(2) *every maximal monochromatic path, except for possibly P', has an end touching two edges of the opposite colour,*

(3) *there are no monochromatic cycles,*

(4) *all coloured edges are at distance at most 3 from P,*

(5) *a vertex has monochromatic degree 3 if and only if it is the vertex y_2 in a copy of F_{exc}^* which is parallel to P.*

Proof of Lemma 8.3. We start by proving (1), namely that every vertex which is incident to exactly two coloured edges, is incident to one red edge and one blue edge. To see this, notice that interior vertices of P have all three edges touching them coloured in (S1), (S2) and (S3). Moreover, (S6) prevents the ends of P from being incident to exactly two coloured edges, both of the same colour. It remains to consider vertices that are not in P. Let v be such a vertex, and suppose that it is incident to exactly two colour edges. Then they are both directed (because the only undirected edges are either in P or introduced in (S6) and do not create vertices of red degree 2 and blue degree 0 or vice versa) and they are directed into v (because whenever a directed edge xy is introduced and coloured, x is already incident to two edges of the other colour). Thus, these edges have different colours (otherwise, the third edges would be coloured in (S4) or (S5)), as required for (1).

Next, we note that right before (S6) is performed, every coloured edge which is not in P' has a vertex which is incident to two edges of the opposite colour; in particular, if xy is directed and not in P' then x is incident to two edges of the opposite colour. Now, taking u, v, w, z to be as in (S6), if uv is recoloured

red then v is incident with two blue edges, and so the other red edge at u is in a red path with an end incident to two blue edges. Similarly, if vz is coloured in this step then either it is coloured red and v has two blue neighbours, or it is coloured blue and intersects a blue edge (namely vw) with an end with two red neighbours. We conclude that every maximal monochromatic path, except for possibly P', has an end with two neighbours of the opposite colour, as stated in (2).

Now, by the last paragraph, no edges outside of P' are contained in a monochromatic cycle. Since P' is a path, it follows that there are no monochromatic cycles, namely (3) holds.

For the remaining items, we will look at the digraph D of edges directed by Algorithm 8.1. We will show that this subgraph has out-degree at most 1, and that, ignoring directions, it is a forest. This forest will turn out to have a very specific structure: its leaves are vertices of P and all its vertices, except for (maybe) the roots, have degree 1 or 3. In addition, the diameter of each tree in the forest is at least the number of leaves in the tree minus 1. Using these structural properties, we will show that the only "problematic" trees are "caterpillars", and among them only the one defined in Definition 8.2 yields a monochromatic vertex.

Claim 8.4. *D has maximum out-degree 1, and its underlying graph is a forest.*

Proof. Suppose that xy is a directed edge in D. Then when Algorithm 8.1 directed this edge, the other edges touching x were already coloured, and they were either undirected or directed towards x. Thus x cannot have an additional out-neighbour besides y, showing that D has maximum out-degree (at most) 1.

Suppose that $(x_1 x_2 \ldots x_t x_1)$ is a cycle in the underlying graph of D. By the maximum out-degree being 1, this is, in fact, a directed cycle in D. Assume that the direction of the edges follows the direction of the cycle, namely $x_1 x_2 \in E(D)$, etc. (if this is not the case, we can look at $(x_t x_{t-1} \ldots x_1 x_t)$ instead). Without loss of generality, $x_1 x_2$ was the last edge to be coloured. Then $x_2 x_3$ was directed before $x_1 x_2$ was coloured, a contradiction. $\square$

Claim 8.5. *Let T be a component of D. Then*

(i) *T is a tree whose edges are directed towards a single vertex, namely, a sink (which we think of as the root of the tree),*

(ii) *every non-root vertex in D has in-degree 0 or 2,*

(iii) *all non-root leaves of D lie in the interior of P,*

(iv) *if the root is in P then it is an end of P and incident to both red and blue edges.*

Proof. Note that if the minimum out-degree of T is 1, then T contains a cycle, contradicting Claim 8.4. Thus T has a vertex with out-degree 0, namely, a sink s. Let S be the set of vertices for which there is a directed path from them to s. Assume that $S \cup \{s\} \neq V(T)$ and let $U := V(T) \setminus (S \cup \{s\})$. Since T is connected, there is an edge connecting a vertex u in U with a vertex v in $S \cup \{s\}$. We claim that this edge must be directed from u to v. Indeed, this is true if $v = s$, because s is a sink. Suppose instead that $v \in S$. By definition of S, all vertices in S have out-degree at least 1 in $T[S \cup \{s\}]$, and since the maximum out-degree of T is 1, the edges uv must again be directed towards v, as claimed. This means that $u \in S$, a contradiction. We conclude that $V(T) = S \cup \{s\}$, proving (i).

36

Now, let $v \in S$ and let u be the only out-neighbour of v in T. Then vu was directed in one of the steps (S2) to (S5) of Algorithm 8.1. If vu was directed in (S2) or (S3) then v has in-degree 0, and otherwise it has in-degree 2, proving (ii). In particular, v is a non-root leaf if and only if it was directed in (S2) or (S3), in which case v is in the interior of P, as required for (iii).

In fact, all vertices in the interior of P are non-root leaves, because if v is in the interior of P, it is incident to an edge vz which is not in P. Because P is a geodesic, $z \notin V(P)$. It follows that vz is directed in (S2) or (S3).

Suppose that the sink s is in P. Then by the previous paragraph, s is an end of P, proving the first part of (iv). Let u be the neighbour of s in P. We now show that s is incident to both red and blue edges. Indeed, if s has two in-neighbours w_1 and w_2 then either at least one of them is red, so (S6) does nothing and s ends up incident with both red and blue edges, or s has two blue in-neighbours and then after (S6) it is incident to both blue and red edges. Otherwise, s has exactly one in-neighbour w. If ws is red, then (S6) does nothing and s has both blue and red neighbours, and if ws is blue then a quick case analysis shows that after (S6), s has both red and blue edges. $\qquad\square$

Given a rooted tree (directed or not), let $L(T)$ be the set of non-root leaves of T and let $\ell(T) := |L(T)|$. We say that a rooted tree T is *good* if $\operatorname{diam}(T) \geq \ell(T) - 1$.

Claim 8.6. *Every component T of D is good.*

Proof. By Claim 8.5, we have that $L(T) \subseteq V(P)$. Let x be the left-most non-root leaf of T in P and y the right-most vertex. Then $\operatorname{dist}_P(x, y) \geq \ell(T) - 1$. As P is a geodesic, $\operatorname{dist}_P(x, y) \leq \operatorname{dist}_T(x, y) \leq \operatorname{diam}(T)$. $\qquad\square$

We say that a rooted (undirected) tree T is *x-binary* if its root has degree x and its other vertices have degree either 1 or 3 (i.e. non-leaf non-roots have degree 3, and thus have two children). We say that a rooted tree is *almost binary* if it is x-binary for some $x \in [3]$, and that it is *binary* if it is 2-binary. A big part of the remainder of this proof will analyse the structure of good almost binary trees, ignoring directions.

Claim 8.7. *Let T be an almost binary tree, and let T' be an almost binary subtree of T. If T is good then so is T'.*

Proof. We prove the claim by induction on $|T| - |T'|$. If $|T| - |T'| = 0$, then $T = T'$, and the claim holds trivially. So suppose $|T| > |T'|$. We may also assume $|T'| \geq 2$, because a singleton is a good tree. Denote the roots of T and T' by r and r', respectively.

We claim that one of the following holds.

(a) $r' \neq r$ and $\deg(r') = 2$,

(b) $r' \neq r$ and $\deg(r') = 1$,

(c) $r' = r$ and $\deg(r') < \deg(r)$,

(d) there is a non-root leaf v in T' which is not a leaf in T.

37

Indeed, if $r' \neq r$ then $\deg(r') \in \{1, 2\}$ (because every neighbour of r' in T' is a child of r' in T and so $\deg(r') \leq 2$, and moreover $\deg(r') \geq 1$ because $|T'| \geq 2$). If $r' = r$ and $\deg(r') = \deg(r)$, then there is a non-root vertex v in T' which has a smaller degree in T' than in T. But then, by the trees being almost-binary, v is a leaf in T' but not in T.

In all cases, we will modify T' slightly to obtain an almost binary tree T'' which is a rooted subtree of T, apply induction to deduce that T'' is good, and then conclude that the same holds for T'.

Let us first define T'' in each of the four cases: if (a) holds, form T'' by adding a new vertex r'' to be the root of T'', and attaching r'' to r'; if (b) or (c) holds, form T'' by adding a new child to r'; and if (d) holds, add two new children to v.

Notice that in all cases T'' is an almost binary subtree of T, which is larger than T'. Thus, by induction, T'' is good, and so $\mathrm{diam}(T'') \geq \ell(T'') - 1$. In case (a), $\mathrm{diam}(T') = \mathrm{diam}(T'')$ and $\ell(T') = \ell(T'')$. In all other cases, $\mathrm{diam}(T') \geq \mathrm{diam}(T'') - 1$ and $\ell(T') = \ell(T'') - 1$. Either way, it follows that $\mathrm{diam}(T') \geq \ell(T') - 1$, so T' is good, as required. $\qquad\square$

Claim 8.8. *Let T be an almost binary good tree. Then every vertex in T has a non-root leaf at distance at most 3 from it.*

Proof. Assume that there is a non-root non-leaf vertex u with no non-root leaf vertex at distance at most 3 from it. Consider the subtree T_u of T rooted at u. It is easy to check that T_u is a binary tree, and by Claim 8.7 it is good. By assumption on u, the tree T_u has depth at least 3. Let $T' \subseteq T_u$ be the subtree of T_u rooted in u with depth 3; by Claim 8.7, T' is good. But T' has eight non-root leaves and diameter 6, a contradiction. So every non-root non-leaf has a non-root leaf at distance at most 2. Thus the root has a non-root leaf at distance at most 3. $\qquad\square$

Recall that by Claim 8.5 (i) and (ii) and by Claim 8.6, every component of D is a good almost binary tree. Thus, by Claim 8.8, every edge coloured by Algorithm 8.17 is at distance at most 3 from P, proving (4). It remains to prove (5).

Observe that in every copy of F_{exc} which is parallel to P, with vertices labelled as in Definition 8.2, Algorithm 8.1 colours it according to F_{exc}^*. In particular, y_2 has red degree 3. Thus, to prove (5), we need to show that every vertex of monochromatic degree 3 is the vertex y_2 in some copy of F_{exc} which is parallel to P.

Definition 8.9. A caterpillar T is a tree with $V(T) = \{y_0, \ldots, y_s, z_1, \ldots, z_{s-1}\}$ and $E(T) = \{y_{i-1}y_i : i \in [s]\} \cup \{y_i z_i : i \in [s-1]\}$. In other words, T consists of a path $(y_0 \ldots y_s)$, along with a pendant edge at each vertex in the interior of the path.
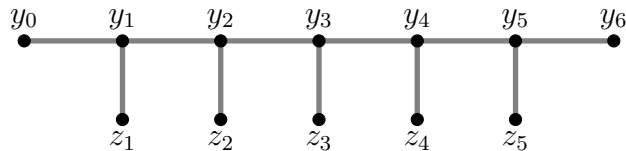

Figure 7: A caterpillar

Claim 8.10. *All good 3-binary trees are caterpillars.*

Proof. Suppose not, and take T to be a minimal good 3-binary tree that it is not a caterpillar. Let $P = (x_0 \ldots x_t)$ be a longest path in T; so the ends x_0 and x_t are leaves in T. If $t \leq 2$, then, using that T is 3-binary, T is a star and thus a caterpillar, a contradiction.

So $t \geq 3$. Since x_0 and x_t are leaves, and T is 3-binary, x_1 and x_{t-1} have degree 3. Let x_0' and x_t' be the neighbours of x_1 and x_{t-1} that are not in P, respectively. By possibly relabelling the vertices of P, we may assume that x_1 is not the root of T. Let $T' = T \setminus \{x_0, x_0'\}$. Then T' is a 3-binary tree with the same root as T, and $T' \subseteq T$. By Claim 8.7, T' is good, and by minimality of T we get that T' is a caterpillar.

Write $V(T') = \{y_0, \ldots, y_s, z_1, \ldots, z_{s-1}\}$ and $E(T') = \{y_{i-1}y_i : i \in [s]\} \cup \{y_i z_i : i \in [s-1]\}$. We now want to understand how to obtain T from T'. To do so, we need to determine which of the vertices of T' is x_1. We consider three cases: $x_1 \in \{y_1, \ldots, y_{s-1}\}$; $x_1 \in \{y_0, y_s, z_1, z_{s-1}\}$; and $x_1 \in \{z_2, \ldots, z_{s-2}\}$. In the first case, T has a vertex of degree 5, contradicting T being 3-binary. In the second case, T is a caterpillar, contradicting the choice of T. And in the third case, T has diameter s and $s + 2$ non-root leaves, contradicting T being good. We have thus reached a contradiction and thereby proved the claim. □

In the following claim we identify the only possible monochromatic vertices. Note that up to this point we did not care if the edges are coloured red or blue, only that they were coloured during Algorithm 8.1. In the sequel we look at the precise colouring.

For the remainder of the proof, let u be a vertex in D with monochromatic degree 3, and let T be the component of D that contains u. We will show that T is a copy of F_{exc}^* which is parallel to P, and that u is the vertex y_2 in F_{exc}^*.

Claim 8.11. *T is a caterpillar, whose leaves are in P and whose interior does not intersect P, and u is a non-leaf root in T.*

Proof. By Claim 8.5, u is not in P and the leaves in T are in the interior of P. Thus all edges incident to u are directed. Moreover, these edges are directed into u, because if there is an edge directed from u, it arises in (S4) or (S5), which ensure that u is not monochromatic. Recall that, by Claim 8.5, T is an almost binary tree, directed towards its root. Since u is a sink with in-degree 3, this means that T is a 3-binary tree rooted at u. By Claim 8.6, T is good, and thus by Claim 8.10, T is a caterpillar. To complete the proof, notice that non-root non-leaves of T have in-degree 1 and out-degree 2 and thus are not in P (the steps involving P, namely (S1), (S2) and (S6) do not create such a vertex). □

By the previous claim, T is a caterpillar with a non-leaf root u. We may thus write $V(T) = \{y_0, \ldots, y_s, z_1, \ldots, z_{s-1}\}$, for some $s \geq 2$. The claim implies that $u \in \{y_1, \ldots, y_{s-1}\}$, because all other vertices are leaves of T. Notice that the vertices of T which are in P form an interval in P (though possibly the order of their appearance in P does not correspond to their order in T), because the distance between any two of them in T is at most s, and if they do not form an interval then the distance between some two of them in P is at least $s + 1$, a contradiction to P being a geodesic. In a similar vain, we may assume that the left-most vertex is y_0 and the right-most vertex is y_s.

Claim 8.12. *The vertex u has exactly one neighbour in P.*

Proof. Notice that by Claim 8.11, u has at least one neighbour in P. Thus we need to rule out the possibility that u has either two or three neighbours in P. Suppose first that u has three neighbours in

P. This means $s = 2$ and $u = y_1$. Since y_0, z_1, y_2 form an interval, it follows from (S2) and (S3) that u has two blue in-edges and one red in-edge, contradicting its being monochromatic.

Now suppose u has two neighbours in P and $s \geq 3$. Without loss of generality, this means $u = y_1$. Notice that its neighbours y_0, z_1 in P are either consecutive or at distance 2 in P, as P is a geodesic. In the former case, u gets two blue in-edges in (S2) and its third edge is directed out of u and coloured red in (S4), a contradiction.
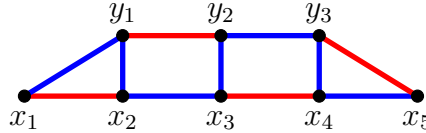
So suppose that y_0 and z_1 are at distance 2 in P. We claim that z_2 must appear between y_0 and z_1. Indeed, recalling that y_0 is the left-most vertex in P and y_s is the right-most vertex, otherwise the vertex w between y_0 and z_1 is at distance $s - 1$ from y_s in T but at distance $s - 2$ from y_s in T, a contradiction to P being a geodesic. It follows that u receives two red in-edges in (S2) and y_1 receives at least one red in-edge in the same step. But this means that the edge $y_1 u$ is directed in (S5) and is blue, a contradiction. □

By the previous claim, $u = y_i$ for some $i \in \{2, \ldots, s - 2\}$. In particular, $s \geq 4$. Notice that if $s = 4$ then $u = y_2$, as required for (5). So suppose that $s \geq 5$. Without loss of generality, $u \neq y_{s-2}$. Notice that both $u = y_i$ and y_{i+1} have exactly one neighbour in P, and thus they receive a red in-edge in (S3). But then if $y_{i+1} y_i$ is ever directed and coloured, this happens in (S5) and the colour is blue, preventing u from being monochromatic. □

8.2 Stage II: Fixing monochromatic vertices (Algorithm 8.14)

The purpose of this section is to fix the colouring given by Algorithm 8.1 to have no monochromatic vertices. Recall Definition 8.2 and Lemma 8.3.
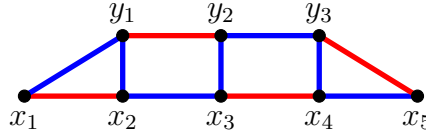
Definition 8.13. Let $F_{\text{exc}}^{**} = $. Namely, F_{exc}^{**} is a red-blue colouring of a F_{exc}, with vertices and edges labelled as in Definition 8.2, where the red edges are $\{x_1 x_2, y_1 y_2, y_2 x_3, x_3 x_4, x_4 y_3, y_3 y_5\}$ and the blue edges are $\{x_1 y_1, y_1 x_2, x_2 x_3, y_2 y_3, x_4 x_5\}$.

Algorithm 8.14 (Fixing monochromatic vertices). *Consider the partial colouring ϕ_1 (the outcome of Algorithm 8.1). Obtain a new partial colouring ϕ_2 from ϕ_1, as follows.*

- *For each copy F of F_{exc}^* (with vertices labelled as in Definition 8.2), do:*
  - *Recolour F according to F_{exc}^{**}. (Notice that, unlike F_{exc}^*, the coloured graph F_{exc}^{**} is not symmetric, so if we relabelled x_i to x_{6-i} and y_j to y_{4-j} in F_{exc}^*, we would get a different colouring. We choose one of these colourings arbitrarily.)*
  - *Update $P' \leftarrow (P' \setminus \{x_1 x_2, x_3 x_4\} \cup \{x_1 y_1 x_2, x_3 y_2 y_3 x_4\})$.*
- *Delete all directions.*

Lemma 8.15. *The partial colouring ϕ_2 has the following properties.*

(1) *if a vertex is incident to at least two coloured edges, then at least one of them is red and at least one is blue,*

(2) *every maximal monochromatic path, except for possibly P', has an end touching two edges of the opposite colour,*

(3) *there are no monochromatic cycles,*

(4) *if u is contained in some coloured edge, then its distance from P is at most 3.*

Proof. For (1), notice that every vertex which is in a copy of F_{exc}^* after Algorithm 8.1 is incident to at least one edge of each colour after Algorithm 8.14, and by Lemma 8.3, every other vertex which is incident to at least two coloured edges, is incident to at least one of each colour.

Suppose that Q is a maximal monochromatic path which is not P'. If it does not intersect the vertices of any copy of F_{exc} parallel to P then Q has an end with two neighbours of the opposite colour, by Lemma 8.3 (2). Now, if Q is red and intersects the vertices of a copy F of F_{exc} parallel to P, then it contains a red edge of F, which has a vertex incident to two blue edges. If Q is blue and intersects such a copy F then $Q = P'$. Item (2) follows.

Item (2) shows that there is no monochromatic cycle with edges outside of P'. Since P' is a path, there are no monochromatic cycles at all, proving (3). Item (4) follows directly from Lemma 8.3. ☐

8.3 Stage III: Creating gadgets (Algorithm 8.17)

In this section we will show that a small modification of the partial colouring ϕ_2, defined in Section 8.2, contains a blue ℓ-gadget.

Let P' and ϕ_2 be the path and partial colouring obtained by running Algorithm 8.1 and then Algorithm 8.14. Write $P' = (p_0 \ldots p_L)$, and note that $L \geq 5\ell + 200$.

Claim 8.16. *One of the following cases holds (or the lemma holds).*

  I. *There is a copy of F_{exc}^* in ϕ_1 which is parallel to P and whose distance on P' from both end points of P' is at least $\ell + 20$.*

 II. *Case I does not hold, and there is a vertex $p_s \in V(P') \setminus V(P)$, with $s \in [2\ell + 40, L - 2\ell - 40]$.*

III. *Cases I and II do not hold, and there exists $s \in [2\ell + 100, L - 2\ell - 100]$ with such that p_{s-1} and p_s have a common neighbour w outside of P'.*

Proof. Suppose that I and II do not hold. Then, writing $P = (q_0 \ldots q_{L'})$, we claim that the vertices q_s and q_{s+1}, with $s \in [2\ell + 60, L' - 2\ell - 60]$ do not have a common neighbour. Indeed, suppose that q_s and q_{s+1} have a common neighbour w, and write $w = p_{s'}$. Notice that $s' \in [2\ell + 60, L - 2\ell - 60]$. Assuming that I does not hold, we find that w is not in a copy of F_{exc}^* in ϕ_1, and thus its edges are not recoloured in ϕ_2, showing that II holds for s'.

It follows that step (S2) does not occur with $uv_{\text{(S2)}} = q_s q_{s+1}$ and $s \in [2\ell + 60, L' - 2\ell - 60]$, and so $P'' := (p_{2\ell+80} \ldots p_{L-2\ell-80})$ is a subpath of P. This implies that P'' is a geodesic (of length at least $\ell + 36$). If no two vertices in P'' have a common neighbour outside of P'', then we can apply Lemma 4.4 to complete the proof of this lemma. So, we assume that there are two vertices in P'' that have a common neighbour w outside of P''. These vertices are not consecutive in P'', by the previous paragraph, so they

41

are at distance 2 in P''. In other words, there exists $s \in [2\ell + 81, L - 2\ell - 81]$ such that p_{s-1} and p_{s+1} have a common neighbour w outside of P''. We claim that w is also outside of P'. Indeed, suppose not. Then w is at distance at most 1 from some vertex in $P \setminus \{p_{s-1}, p_s, p_{s+1}\}$; denote this vertex by p_t. So p_{s-1} and p_{s+1} are at distance at most 2 from p_t, a contradiction to P being a geodesic. $\square$

Algorithm 8.17 (Creating a gadget). *Throughout this description, if v is an interior vertex of P', then we denote by v' the (unique) neighbour of v such that vv' is not an edge of P' (so vv' is a red edge in ϕ_2).*

We explain how to proceed in each of the three cases given in Claim 8.16.

I. *There is a copy of F^*_{exc} in ϕ_1 which is parallel to P and whose distance on P' from both end points of P' is at least $\ell + 20$.*

Label the vertices and edges of such a copy of F^*_{exc} as in Definition 8.2. Thus $(x_1 \ldots x_5)$ is a subpath of P, and $(x_1 y_1 x_2 x_3 y_2 y_3 x_4 x_5)$ is a subpath of P'. Let s be such that $x_4 = p_s$ and suppose that $p_{s+1} = x_5$; notice that $s + \ell + 2 \leq L - 1$. Recolour the edges $p_s p_{s-1}$, $p_{s-2} p_{s-3}$ and $p_{s+\ell} p_{s+\ell+1}$ red. If $p'_{s+\ell} p'_{s+\ell+1}$ is an edge, colour it blue (see Figure 8).

In this way we get an ℓ-gadget of Type II as follows: $Q_0 = (p_{s-2} \ldots p_{s+\ell})$, $Q_1 = (p_{s-1} p_{s+1})$, $e_1 = p_{s-5} p_{s-2}$, $e_2 = p_{s-3} p_{s-2}$, $e_3 = p_{s-3} p_s$, $e_4 = p_{s+\ell} p_{s+\ell+1}$, and $e_5 = p_{s+\ell} p'_{s+\ell}$.
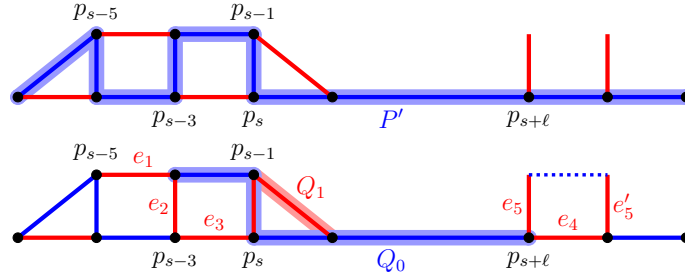
Figure 8: Case I: before and after Algorithm 8.17

II. *Case I does not hold, and there is a vertex $p_s \in V(P') \setminus V(P)$, with $s \in [2\ell + 40, L - 2\ell - 40]$.*

Recolour the edges $p_{s-3} p_{s-2}$, $p_{s-1} p_s$, and $p_{s+\ell} p_{s+\ell+1}$ red. If $p'_{s-2} p'_{s-3}$ or $p'_{s+\ell} p'_{s+\ell+1}$ are edges, colour them blue (see Figure 9).

This gives a gadget of Type II as follows: $Q_0 = (p_{s-2} \ldots p_{s+\ell})$, $Q_1 = (p_{s-1} p_{s+1})$, $e_1 = p_{s-3} p_{s-2}$, $e_2 = p_{s-2} p'_{s-2}$, $e_3 = p_s p'_s$, $e_4 = p_{s+\ell} p_{s+\ell+1}$, $e_5 = p_{s+\ell} p'_{s+\ell}$. Note that the edges e_3 and e_2 may intersect or be equal.

III. *Cases I and II do not hold, and there exists $s \in [2\ell + 100, L - 2\ell - 100]$ with such that p_{s-1} and p_s have a common neighbour w outside of P'.*

Recolour the edges $p_{s-3} p_{s-2}$, $p_{s-1} p_s$ and $p_{s+\ell} p_{s+\ell+1}$ red. If $p'_{s-2} p'_{s-3}$ or $p'_{s+\ell} p'_{s+\ell+1}$ are edges, colour them blue (see Figure 10).

This gives a gadget of Type II as follows: $Q_0 = (p_{s-2} \ldots p_{s+\ell})$, $Q_1 = p_{s-1} w p_{s+1}$, $e_1 = p_{s-3} p_{s-2}$, $e_2 = p_{s-2} p'_{s-2}$, $e_3 = p_s p'_s$, $e_4 = p_{s+j} p_{s+j+1}$, and $e_5 = p_{s+j+1} p'_{s+j+1}$.
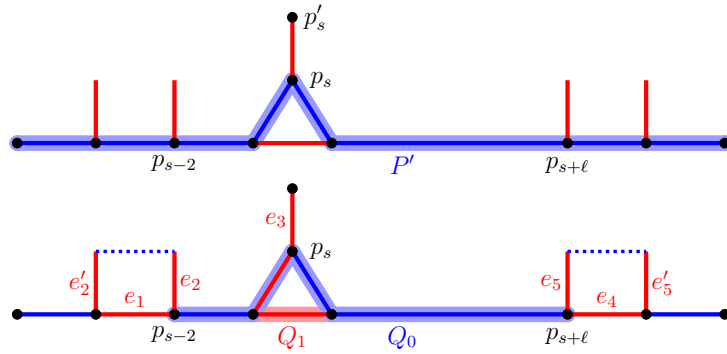
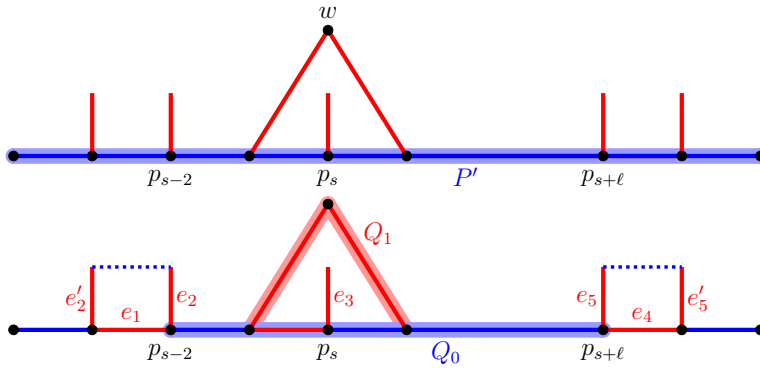**Figure 9:** Case II: before and after Algorithm 8.17

Figure 10: Case III: before and after Algorithm 8.17

Denote the partial colouring at the end of this process by ϕ_3 and write $e_5' := p_{s+\ell+1}p_{s+\ell+1}'$ and $e_2' := p_{s-3}p_{s-3}'$.

Lemma 8.18. *The partial colouring ϕ_3 satisfies the following properties, and thus satisfies the requirements of Lemma 4.3.*

(1) *all coloured edges are at distance at most 4 from the geodesic P and form a connected subgraph,*

(2) *if a vertex u is incident to at least two coloured edges, then at least one of them is red and at least one is blue,*

(3) *if C is a cycle in G containing no red edges or no blue edges, then C has two consecutive edges that are not coloured by ϕ_3.*

Proof. By Lemma 8.15, before applying Algorithm 8.17 all coloured edges are at distance at most 3 from P. Note that in Algorithm 8.17 all newly coloured edges (if any) intersect an already coloured edge, and thus are at distance at most 4 from P; therefore, (1) holds.

By Lemma 8.15, before applying Algorithm 8.17, Item (2) above holds, and thus it still holds after applying Algorithm 8.17, from the way the algorithm is defined. It remains to prove (3).

Consider the blue edges $f_1 := p_{s-3}'p_{s-2}'$ and $f_2' = p_{s+\ell}'p_{s+\ell+1}'$ (if exist). We claim that f_1 and f_2 were not red in ϕ_2. Indeed, notice that every red edge in both ϕ_1 and ϕ_2 has a vertex with two blue neighbours.

43

It follows that ϕ_2 has no red path of length 3, and thus f_1 and f_2 cannot be red in ϕ_2. Thus, either f_1 is in a blue path in ϕ_3 which has an end with two red neighbours, or both edges of f_1 have an uncoloured neighbour.

We now show that there are no monochromatic cycles. Suppose C is a monochromatic cycle. Then by Lemma 8.15 (3), C intersects an edge that was coloured (or recoloured) in Algorithm 8.17. By the previous paragraph, C does not contain f_1 or f_2. Notice that all maximal monochromatic paths in ϕ_3, that intersect an edge coloured in this algorithm, have an end with two edges of the opposite colour, except for possibly the maximal red paths containing $e_2 e_1 e_2'$ or $e_5 e_4 e_5'$ and for f_1 or f_2 if they do not intersect other blue paths. So we need to rule out the case that C contains one of e_5 and e_2.

We explain why C does not contain e_5. Indeed, suppose otherwise, and recall that $e_5 = p_{s+\ell} p_{s+\ell}'$. Then $p_{s+\ell} \notin P'$ (if $p_{s+\ell}'$ is in P' then it has two blue neighbours). In particular, e_5 is not in a copy of F_{exc}^{**}. Thus e_5 was coloured in (S3) and was not recoloured afterwards. Since e_5 is at distance at least 10, say, from the ends of P', the edge e_5 does not touch any edges coloured in (S6) of Algorithm 8.1. It follows that, in ϕ_1, the vertex $p_{s+\ell}'$ either has no red neighbours other than $p_{s+\ell}$, or it has both a red and a blue one. In the former case, in ϕ_3 the vertex $p_{s+\ell}'$ still has red degree 1. In the latter case, denoting by u the red neighbour of $p_{s+\ell}'$ which is not $p_{s+\ell}$, we find that u has two blue neighbours, and this remains the case in ϕ_3 (observe that $u \neq p_{s+\ell+1}$). Either way, e_5 is not contained in a red cycle, and similarly for e_2.

Suppose that C is a cycle containing edges of at most one colour. Because there are no monochromatic cycles, C contains at least one uncoloured edge, and we may assume that it contains at least one coloured edge (otherwise (3) holds trivially). Let $R_1, \ldots, R_k$ be the maximal monochromatic subpaths of C. We claim that R_i is a maximal monochromatic path also in ϕ_3, for $i \in [k]$. Indeed, write $R_i = (r_0 \ldots r_\rho)$ and suppose that r_0 is incident to an edge $e \neq r_0 r_1$ coloured with the colour of R_i. Then the third edge at r_0 has the opposite colour, by (2), so by assumption on C it cannot contain this third edge and thus it contains e, contradicting the maximality of R_i. Recalling our analysis of maximal monochromatic paths with an end incident with two edges of the same colour, the only candidates for $R_1, \ldots, R_k$ are f_1, f_2 (if their vertices are not incident with other blue edges), and the maximal red paths containing $e_5 e_4 e_5'$ or $e_2 e_1 e_2'$. In other words, either C contains one or both of f_1 and f_2 and no other coloured edges, or (and this follows similarly to the argument about there being no red cycles) C contains one of $e_5 e_4 e_5'$ and $e_2 e_1 e_2'$ and no other coloured edges.

Claim 8.19. *In Cases* II *and* III, *the vertices* $u := p_{s-3}$ *and* $v := p_{s+\ell+1}$ *are at distance at least 5 from each other.*

Proof. Notice that in both cases u and v are at distance at least $\ell + 4 \geq 7$ in P'. In Case III the vertices u and v are at distance at least 7 also in P and thus in G. In Case II both e_1 and e_4 are edges in P. Indeed, if say e_1 is not in P, then, using that Case I does not hold and recalling (S2) from Algorithm 8.1, one of the edges e_2 and e_2' has a vertex touching two blue edges, a contradiction. It follows that u and v are both in P and are at distance at most 5 in P and thus in G. Either way, u and v are at distance at least 5 from each other in G, so $|C| \geq 10$ and C has two consecutive uncoloured edges. $\square$

We obtain a lower bound on $|C|$ in each of the possible cases, using Claim 8.19 above.

- C contains exactly one of f_1, f_2. Then $|C| \geq 3$.

- C contains both f_1 and f_2. Then $|C| \geq 6$ as p_{s-3}' and $p_{s+\ell+1}'$ are a distance at least 3 apart.

- C contains exactly one of $e_2 e_1 e_2'$ and $e_5 e_4 e_5'$. Then $|C| \geq 5$, because f_1 and f_2, if exist, are blue.

- C contains both $e_2 e_1 e_2'$ and $e_5 e_4 e_5'$. Then $|C| \geq 10$, because C contains u and v which are a distance at least 5 apart.

In all cases, it easily follows that C has two consecutive uncoloured edges, proving (3). $\qquad\square$

9 Concluding Remarks

In this paper we essentially prove Wormald's conjecture, showing that large connected cubic graphs, whose number of vertices is divisible by 4, can be decomposed into two isomorphic linear forests (Theorem 1.3).

In fact, our proof can be tweaked to prove the following stronger statement.

Theorem 9.1. *There exists k_0 such that the following holds. Suppose that G is a cubic graph on n vertices, where n is divisible by 4, and G has a component on at least k_0 vertices. Then there is a red-blue edge-colouring of G whose colour classes span isomorphic linear forests.*

Proof sketch. We now sketch how this strengthening can be proved. First, we note that our only use of the connectivity assumption on G was in Claim 5.1, where we find $\Omega(n^{1-\varepsilon})$ geodesics of length $m := 10^{10}\sqrt{\log n}$ which are at distance at least 50 from each other. For this, it actually suffices to have at least say $n/2$ vertices in G be in components of order at least $3 \cdot 2^m$. Indeed, running the same proof on the union of components of at least this order yields the same result.

For the proof of Theorem 9.1, let G be a cubic graph as in the theorem. For each k let $\mathcal{C}_k$ be the collection of components of G on exactly k vertices. Apply Theorem 1.5 (Thomassen's theorem) to each component in $\mathcal{C}_k$ to get a red-blue colouring where every monochromatic component is a path of length at most 5. To each such colouring, associate a 10-tuple $(r_1, \ldots, r_5, b_1, \ldots, b_5)$, where r_i and b_i denote the numbers of red and blue components which are paths of length i. Notice that the number of possible 10-tuples is at most the number of non-negative integer solutions to the equation $x_1 + \ldots + x_5 + y_1 + \ldots + y_5 = 3k/2$, where x_i and y_i count the number of edges in red and blue components which are paths of length i. This number is $\binom{3k/2+9}{9} = O(k^9)$. This implies that we can pair up the components in $\mathcal{C}_k$, leaving at most $O(k^9)$ unpaired components, such that the 10-tuples corresponding to each pair (C_1, C_2) are identical. Now reverse the colouring in C_2 to get a red-blue colouring of $C_1 \cup C_2$ where the red and blue graphs are isomorphic linear forests (with paths of length at most 5). It thus suffices to consider the graph G' which is the union of unpaired graphs, and this is a graph where the number of components on k vertices is $O(k^9)$, for every k.

We now show that G' can be dealt with via the above paragraph. Indeed, denote by n the number of vertices in G' and by K the number of vertices in the largest component of G'. Then $K \geq k_0$ (technically we need to make sure we leave at least one component on at least k_0 unpaired; this is clearly possible) and $n \leq \sum_{i \leq K} O(i^9) = O(K^{10})$. Writing $m := 10^{10}\sqrt{\log n}$, as above, and noting that $m \leq n^{1/20}$, say (using that n is large, which follows from k_0 being large), we get that there are at most $\sum_{i \leq m} O(i^9) = O(m^{10}) = O(\sqrt{n})$ components on fewer than m vertices. Thus the first paragraph above can be applied to find a red-blue colouring of G' where the red and blue graphs are isomorphic linear forests, completing the proof. $\qquad\square$

Notice that the assumption that the number of vertices is divisible by 4 in Wormald's conjecture is necessary, as otherwise the number of edges is not divisible by 2. Nevertheless, we can address the case when the number of vertices is not divisible by 2, as follows.

Theorem 9.2. *Let G be a connected cubic graph on n vertices, where n is large and $n \equiv 2 \pmod 4$. Then there is a red-blue edge-colouring of G and a linear forest F, such that the blue graph is isomorphic to the disjoint union of F with a path of length 3, and the red graph is isomorphic to the disjoint union of F with a matching of size 2. In particular, the two colour classes are linear forests, and the red graph is isomorphic to the blue one minus one edge.*

Proof sketch. This can be proved by tweaking our proof of the main result, Theorem 1.3 very slightly, as follows. In Claim 4.6, instead of requiring the number of red and blue edges to be same, make sure that the number of blue edges exceeds the number of red edges by exactly 1 (this can be done analogously). Next, follow the proof Claim 4.7 to make sure that the number of red and blue components which are paths of length t, with $t \geq 4$, is the same, and that the number of blue components which are paths of length 3 exceeds the number of such red components by exactly 1. A counting argument like the one that appears after the proof of Claim 4.7 shows that the number of red and blue components which are paths of length 2 is the same. It thus follows that there are two more red than blue components which are edges, proving Theorem 9.2. □

As our paper is pretty long, we would like to point out that weaker versions of our main result can be proved with considerably less effort. Indeed, first, Section 3 on its own proves an approximate version of our main result (which applies for any cubic graph). Additionally, proving Conjecture 1.2 for (large connected) cubic graphs with high girth is also a mush easier task. Indeed, if the girth is at least 5, then there are no geodesics with two vertices having a common neighbour outside the geodesic, making Section 8 irrelevant, and if the girth is at least 6, then Cases **II** and **III** in Section 7 also become irrelevant; either way, imposing a high girth condition would shorten our paper by about 15 pages.

We now mention some directions for future research. It would be interesting to see if our result could be combined with Thomassen's result (that every cubic graph can be 2-edge-coloured so that monochromatic components are paths of length at most 5). That is, is it true that every cubic graph can be 2-edge-coloured such that the two colour classes are isomorphic linear forests with bounded component sizes?

Problem 9.3. Is there a constant c such that every cubic graph, whose number of vertices is divisible by 4, can be 2-edge-coloured such that the two colour classes are isomorphic, and every monochromatic component is a path of length at most c?

We note that the proof of Thomassen (as well as those of Jackson and Wormald [21] and Aldred and Wormald [6] with larger constants) is very technical and requires a lot of case analysis. However, we can prove a version of his result, where we bound components' length by a constant larger than 5, in a much simpler way, following the argument in Claim 4.5 (this is also similar to an argument from [7]). Therefore, it is possible that some of the ideas from this paper could be used in order to answer Problem 9.3.

Let us briefly sketch a simpler proof of a weakening of Thomassen's result here. Given a cubic graph G, pick a red-blue colouring of G whose colour classes are linear forests (see [3] for a very short proof of the existence of such a colouring). For an appropriate constant c, let $\mathcal{P}_r$ be the set of red components that are

paths of length at least c. Now decompose each path in $\mathcal{P}_r$ into segments of length in $[c/2, c]$, letting $\mathcal{I}_r$ be the family of such segments. For each segment $I \in \mathcal{I}_r$, let $\mathrm{cand}(I)$ be the set of edges in I's interior that do not close a blue cycle if their colour is swapped; recall that $|\mathrm{cand}(I)| \geq (|I|-3)/2 \geq (c/2-2)/2 = c/4-1$. Now form a graph H_r by joining an edge $e \in \mathrm{cand}(I)$ with an edge $f \in \mathrm{cand}(J)$, for distinct $I, J \in \mathcal{I}_r$, if they touch the same blue component. Then H_r is a $|\mathcal{I}_r|$-partite graph, with parts of size at least $c/4 - 1$, whose maximum degree is at most 4. An application of the Lovász local lemma (or the result in [24]) proves the existence of an independent J_r transversal (namely, an independent set consisting of one vertex from each part) in H_r, provided that c is large enough. Defining $\mathcal{P}_b$ etc. analogously, and flipping the colours of edges in $J_r \cup J_b$, yields a red-blue colouring where the monochromatic components are paths of length at most $2c + 1$.

Another interesting problem in this direction relates to the tightness of Thomassen's result. While it is known that the result is tight, namely that there is a cubic graph that cannot be 2-edge-coloured with monochromatic components being paths of length at most 4, the only known minimal examples are graphs on six vertices. It is thus plausible that for larger connected cubic graphs a bound of 4 on the length of monochromatic components is enough.

Problem 9.4. Can every large enough connected cubic graph be 2-edge-coloured so that all monochromatic components are paths of length at most 4?

As mentioned in the introduction, one can also consider vertex versions of the above problems. For example, Abreu, Goedgebeur, Labbate and Mazzuoccolo [2] , conjectured a vertex version of Wormald's conjecture, asserting that every cubic graph can be 2-vertex-coloured so that the two colour classes induce isomorphic linear forest. This was recently proved for large connected cubic graph with large girth [12], but it is open in general. Note that in [12] it is also proved that for large connected cubic graphs, with unrestricted girth, there is a 2-vertex-colouring where the two colours induce isomorphic graphs, whose components are paths and cycles.

Conjecture 9.5 (Conjecture 2.20 in [2]). *Every cubic graph admits two-colouring of vertex set such that the two colour classes induce isomorphic linear forests.*

Ban and Linial also considered related problems while studying so-called internal partitions of regular graphs (see [9] for more details). They noticed a nice connection between internal and external partitions and cubic graphs, and made the following conjecture.

Conjecture 9.6 (Conjecture 1 in [9]). *Every bridgeless cubic graph, with the exception of the Petersen graph, can be two-vertex-coloured so that the two colour classes induce isomorphic graphs that consist of a matching and isolated vertices.*

This was proved for some specific cubic graphs [1, 9] but it is still widely open in general, and seems hard.

Turning back to Wormald's conjecture, one can consider a similar statement for higher degrees. Let $\mathrm{la}(d)$ be the minimum k such that every d-regular graph can be decomposed into at most k linear forests; so the linear arboricity conjecture asserts that $\mathrm{la}(d) = \lceil (d+1)/2 \rceil$. Is it true that there is such a decomposition which is "balanced"?

Question 9.7. *Let $d \geq 3$ be an integer. Is it true that every large connected d-regular graph, whose number of edges is divisible by $\mathrm{la}(d)$, can be decomposed into $\mathrm{la}(d)$ isomorphic linear forests?*

In fact, one can ask a similar (and perhaps easier) question about a balanced decomposition into graphs with maximum degree at most 2. A simple consequence of Vizing's theorem is that every d-regular graph can be decomposed into $\lceil (d+1)/2 \rceil$ graphs with maximum degree 2. Here we ask if they can also be isomorphic.

Question 9.8. *Let $d \geq 3$ be an integer. Is it true that for every large connected d-regular graph, whose number of edges is divisible by $\lceil (d+1)/2 \rceil$, can be decomposed into $\lceil (d+1)/2 \rceil$ isomorphic graph with maximum degree 2?*

In this paper we showed that the answer is yes (to both questions) for cubic graphs. We believe this direction might be a potential avenue of further progress towards a solution of the linear arboricity conjecture. Some of our methods may be applied towards attacking this conjecture, however, there are certain obstacles which would need to be resolved. One of them is to find good "switchings" which would allow us to break monochromatic cycles or monochromatic "long" paths. In a cubic graph, it is easy to flip a colour of an edge and still maintain the property of monochromatic components being linear forests, but that is no longer the case when d grows.

References

[1] M. Abreu, J. Goedgebeur, D. Labbate, and G. Mazzuoccolo, *A note on 2-bisections of claw-free cubic graphs*, Discr. Appl. Math. **244** (2018), 214–217. 4, 47

[2] _____, *Colourings of cubic graphs inducing isomorphic monochromatic subgraphs*, J. Graph Theory **92** (2019), no. 4, 415–444. 4, 47

[3] J. Akiyama and V. Chvátal, *A short proof of the linear arboricity for cubic graphs*, Bull. Liber. Arts Sci. NMS **2** (1981), 1–3. 2, 10, 46

[4] J. Akiyama, G. Exoo, and F. Harary, *Covering and packing in graphs. III: Cyclic and acyclic invariants*, Mathematica Slovaca **30** (1980), no. 4, 405–417. 1, 2

[5] _____, *Covering and packing in graphs IV: Linear arboricity*, Networks **11** (1981), no. 1, 69–72. 2

[6] R. E. L. Aldred and N. C. Wormald, *More on the linear k-arboricity of regular graphs*, Australas. J. Combin. **18** (1998), 97–104. 3, 46

[7] N. Alon, *The linear arboricity of graphs*, Israel J. Math. **62** (1988), no. 3, 311–325. 2, 46

[8] N. Alon and J. H. Spencer, *The probabilistic method*, John Wiley & Sons, 2016. 2

[9] A. Ban and N. Linial, *Internal partitions of regular graphs*, J. Graph Theory **83** (2016), no. 1, 5–18. 4, 47

[10] J.-C. Bermond, J.-L. Fouquet, M. Habib, and B. Peroche, *On linear k-arboricity*, Discr. Math. **52** (1984), no. 2-3, 123–132. 3

[11] G. Chen, Y. Hao, and G. Yu, *Linear arboricity of degenerate graphs*, Journal of Graph Theory, 1–12. 2

[12] S. Das, A. Pokrovskiy, and B. Sudakov, *Isomorphic bisections of cubic graphs*, J. Combin. Theory Ser. B **151** (2021), 465–481. 4, 9, 47

[13] H. Enomoto, *The linear arboricity of 5-regular graphs*, Technical report, Dep. of Information Sci., Univ. of Tokyo, 1981. 2

[14] H. Enomoto and B. Péroche, *The linear arboricity of some regular graphs*, J. Graph Theory **8** (1984), no. 2, 309–324. 2

[15] A. Ferber, J. Fox, and V. Jain, *Towards the linear arboricity conjecture*, J. Combin. Theory Ser. B **142** (2020), 56–79. 2

[16] J. L. Fouquet, H. Thuillier, J. M. Vanherpe, and A. P. Wojda, *On linear arboricity of cubic graphs*, LIFO Univ. d'Orléans-Research Report **13** (2007), 1–28. 2

[17] J.-L. Fouquet, H. Thuillier, J.-M. Vanherpe, and A. P. Wojda, *On isomorphic linear partitions in cubic graphs*, Discr. Math. **309** (2009), no. 22, 6425–6433. 2

[18] 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. 2

[19] F. Guldan, *The linear arboricity of 10-regular graphs*, Mathematica Slovaca **36** (1986), no. 3, 225–228. 2

[20] F. Harary, *Covering and packing in graphs, I.*, Annals of the New York Academy of Sciences **175** (1970), no. 1, 198–205. 1

[21] B. Jackson and N. C. Wormald, *On the linear k-arboricity of cubic graphs*, Discr. Math. **162** (1996), no. 1-3, 293–297. 3, 46

[22] J. Kahn, *Asymptotically good list-colorings*, J. Combin. Theory Ser. A **73** (1996), no. 1, 1–59. 2

[23] R. Lang and L. Postle, *An improved bound for the linear arboricity conjecture*, arXiv:2008.04251 (2020). 2

[24] P.-S. Loh and B. Sudakov, *Independent transversals in locally sparse graphs*, J. Combin. Theory Ser. B **97** (2007), no. 6, 904–918. 19, 47

[25] C. McDiarmid, *On the method of bounded differences*, Surveys in combinatorics **141** (1989), no. 1, 148–188. 8

[26] C. McDiarmid and B. A. Reed, *Linear arboricity of random regular graphs*, Random Struct. Algorithms **1** (1990), no. 4, 443–447. 2

[27] M. Molloy and B. Reed, *Near-optimal list colorings*, Random Struct. Algo. **17** (2000), no. 3-4, 376–402. 2

[28] C. St. J. A. Nash-Williams, *Decomposition of finite graphs into forests*, J. London Math. Soc. **39** (1964), 12. 1

[29] B. Peroche, *On partition of graphs into linear forests and dissections*, manuscript (1982). 2

[30] C. Thomassen, *Two-coloring the edges of a cubic graph such that each monochromatic component is a path of length at most 5*, J. Combin. Theory Ser. B **75** (1999), no. 1, 100–109. 3

[31] P. Tomasta, *Note on linear arboricity*, Mathematica Slovaca **32** (1982), no. 3, 239–242. 2

[32] V. G. Vizing, *On an estimate of the chromatic class of a p-graph*, Discret. Analiz. **3** (1964), 25–30. 1

[33] N. Wormald, *Problem 13*, Ars Combinatoria **23** (1987), 332–334. 2

[34] J.-L. Wu, *On the linear arboricity of planar graphs*, J. Graph Theory **31** (1999), no. 2, 129–134. 2

[35] J.-L. Wu and Y.-W. Wu, *The linear arboricity of planar graphs of maximum degree seven is four*, J. Graph Theory **58** (2008), no. 3, 210–220. 2