Combinatorics, Probability and Computing (2025), **34**, pp. 565–584
doi:10.1017/S0963548325000057

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Tight Hamilton cycles with high discrepancy

Lior Gishboliner[1], Stefan Glock[2] and Amedeo Sgueglia[2]

[1]Department of Mathematics, University of Toronto, Toronto, ON, Canada and [2]Fakultät für Informatik und Mathematik, Universität Passau, Passau, Germany
Corresponding author: Amedeo Sgueglia; Email: amedeo.sgueglia@uni-passau.de

(Received 22 February 2024; revised 24 January 2025; accepted 28 January 2025)

Abstract

In this paper, we study discrepancy questions for spanning subgraphs of k-uniform hypergraphs. Our main result is that, for any integers $k \geq 3$ and $r \geq 2$, any r-colouring of the edges of a k-uniform n-vertex hypergraph G with minimum $(k-1)$-degree $\delta(G) \geq (1/2 + o(1))n$ contains a tight Hamilton cycle with high discrepancy, that is, with at least $n/r + \Omega(n)$ edges of one colour. The minimum degree condition is asymptotically best possible and our theorem also implies a corresponding result for perfect matchings. Our tools combine various structural techniques such as Turán-type problems and hypergraph shadows with probabilistic techniques such as random walks and the nibble method. We also propose several intriguing problems for future research.

Keywords: Discrepancy; tight Hamilton cycles; random walks; nibble method

2010 Mathematics Subject Classification: Primary: 05C35, Secondary: 05C65, 05C45

1. Introduction

In discrepancy theory, the basic question is whether a structure can be partitioned in a balanced way, or if there is always some 'discrepancy' no matter how the partition is made. Formally, let $\mathcal{H}$ be a hypergraph and let $f : V(\mathcal{H}) \to \{\text{red}, \text{blue}\}$ be a 2-colouring of its vertices. For an edge $e \in E(\mathcal{H})$ and a colour c, let $c(e) := \{x \in e : f(x) = c\}$. The discrepancy of e is defined as $D_f(e) := \big||\text{red}(e)| - |\text{blue}(e)|\big| = 2 \cdot \max_{c \in \{\text{red}, \text{blue}\}} \left(|c(e)| - \frac{|e|}{2}\right)$; the larger $D_f(e)$ is, the less balanced is the colouring of e. The discrepancy of $\mathcal{H}$ is then defined as $\min_f \max_e D_f(e)$. In other words, the discrepancy measures the maximum imbalance that is guaranteed to occur in every 2-colouring of $V(\mathcal{H})$. Discrepancy of hypergraphs is a classical topic in combinatorics; we refer the reader to [2, Chapter 13] for an introduction. The notion of discrepancy naturally generalises to more than 2 colours: For a hypergraph $\mathcal{H}$, an r-colouring $f : V(\mathcal{H}) \to [r]$ and an edge $e \in E(\mathcal{H})$, define the discrepancy of e as $D_f(e) := r \cdot \max_{c \in [r]} \left(|c(e)| - \frac{|e|}{r}\right)$. This coincides with the above definition of $D_f(e)$ for the case $r = 2$. The r-colour discrepancy of $\mathcal{H}$ is then defined as $\min_f \max_e D_f(e)$.

There are many works studying discrepancy problems for hypergraphs arising from graphs, namely, when $V(\mathcal{H})$ is the *edge set* of a graph G and $E(\mathcal{H})$ is a family of subgraphs of G. Two early results of this type are the theorem of Erdős and Spencer [12] on the discrepancy of cliques in the complete graph, and the work of Erdős, Füredi, Loebl and Sós [11] on the discrepancy of copies of a given spanning tree in the complete graph. In recent years there has been a lot of interest in discrepancy problems in general graphs, and there are by now many works studying conditions that guarantee the existence of high-discrepancy subgraphs of various types, such as perfect matchings and Hamilton cycles [3, 14, 16, 17], spanning trees [17], H-factors [4, 7] and

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

powers of Hamilton cycles [6]. See also [15, 18] for an oriented analogue. Many of the results study minimum degree thresholds for linear discrepancy, namely, they determine how large the minimum degree of G should (asymptotically) be so that G is guaranteed to contain a subgraph of a certain type with discrepancy $\Omega(n)$. For example, in [3, 14, 17], it is shown that for every $\varepsilon > 0$, every r-edge-colouring of an n-vertex graph with minimum degree $(\frac{r+1}{2r} + \varepsilon)n$ has a Hamilton cycle and a perfect matching with linear discrepancy, and the constant $\frac{r+1}{2r}$ is best possible. We will often say *high discrepancy* to mean linear discrepancy, i.e., discrepancy $\Omega(n)$.

In this work, we study discrepancy problems in k-uniform hypergraphs (short: *k-graphs*), in analogy to the aforementioned works for graphs. In this context, it is worth mentioning the seminal result of Alon, Frankl, and Lovász [1] in which they proved, using topological methods, the following conjecture of Erdős: If the edges of the complete n-vertex k-graph $K_n^{(k)}$ are coloured with r colours, and $n \geq (r-1)(s-1) + sk$, then there exists a monochromatic matching of size s. This generalises Kneser's conjecture which corresponds to the case $k = 2$ and was resolved by Lovász [25]. The Alon–Frankl–Lovász result implies in particular that any 2-edge-colouring of $K_n^{(k)}$ contains a monochromatic matching of size at least $\lfloor \frac{n}{k+1} \rfloor$ and, by arbitrarily adding edges, this can be extended to a perfect matching with high discrepancy (assuming $k \mid n$ of course).

Our main result is the determination of the minimum $(k-1)$-degree threshold for the discrepancy of perfect matchings and tight Hamilton cycles in k-graphs, thereby establishing a discrepancy version of the celebrated theorem of Rödl, Ruciński and Szemerédi [29]. Recall that a *tight Hamilton cycle* of a k-graph G is a cyclic ordering $v_1, \dots, v_n$ of the vertices of G such that $v_i v_{i+1} \dots v_{i+k-1}$ is an edge for every $1 \leq i \leq n$, with indices taken modulo n. Before stating our main result, let us give some background on perfect matchings and Hamilton cycles in hypergraphs of large minimum degree. For a k-graph G and a set $S \subseteq V(G)$, we say that the *degree* of S in G, denoted by $d_G(S)$, is the number of edges containing S. We use $\delta(G)$ to denote the *minimum $(k-1)$-degree*, which is the minimum of $d_G(S)$ over all $(k-1)$-sets $S \subseteq V(G)$. In their seminal paper which introduced the absorbing method systematically, Rödl, Ruciński, and Szemerédi [29] showed that for every $\varepsilon > 0$, any n-vertex k-graph G with $\delta(G) \geq (1/2 + \varepsilon)n$ contains a tight Hamilton cycle. Moreover, this is best possible, as there are k-graphs G with $\delta(G) = n/2 - O(1)$ and no tight Hamilton cycle (see [21, Theorem 3]).

Let us now consider discrepancy of tight Hamilton cycles. Mansilla Brito [27] showed that if a 3-graph G satisfies $\delta(G) \geq (5/6 + \varepsilon)n$, then it contains a tight Hamilton cycle with high discrepancy. We improve this to $\delta(G) \geq (1/2 + \varepsilon)n$ and show that this holds for any uniformity $k \geq 3$ and any (fixed) number of colours $r \geq 2$. This result is best possible since $\delta(G) \geq \frac{n}{2} - O(1)$ is needed even to guarantee the existence of a tight Hamilton cycle. Thus, for $k \geq 3$, the threshold for the discrepancy of Hamilton cycles is the same as the existence threshold, and does not depend on the number of colours r. This is in contrast to the graph case, where the discrepancy threshold is strictly larger than the existence threshold and decreases as r increases (see [3, 14, 17]).

Theorem 1.1. *For all $k, r \in \mathbb{N}$ with $k \geq 3$ and $r \geq 2$, and all $\varepsilon > 0$, there exists $\mu > 0$ such that the following holds for all sufficiently large n. Let G be an n-vertex k-graph with $\delta(G) \geq (1/2 + \varepsilon)n$ whose edges are r-coloured. Then there exists a tight Hamilton cycle in G which contains at least $(1 + \mu)\frac{n}{r}$ edges of the same colour.*

Note that if n is divisible by k, then a tight Hamilton cycle decomposes into k perfect matchings. So by using Theorem 1.1 and averaging, we obtain the following corollary, which was proved independently by Balogh, Treglown and Zárate-Guerén [5].

Corollary 1.2. *For all $k, r \in \mathbb{N}$ with $k \geq 3$ and $r \geq 2$, and all $\varepsilon > 0$, there exists $\mu > 0$ such that the following holds for all sufficiently large n divisible by k. Let G be an n-vertex k-graph with $\delta(G) \geq (1/2 + \varepsilon)n$ whose edges are r-coloured. Then there exists a perfect matching in G which contains at least $(1 + \mu)\frac{n}{rk}$ edges of the same colour.*

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

The constant $1/2$ in Corollary 1.2 is tight, as there exist n-vertex k-graphs with $\delta(G) = n/2 - O(1)$ and no perfect matching (see [24, 30]).

Similarly, Theorem 1.1 also implies an upper bound of $1/2$ for the discrepancy threshold of Hamilton ℓ-cycles, where, for $1 \le \ell \le k-1$ and for n divisible by $k-\ell$, a *Hamilton ℓ-cycle* on n vertices is a cyclic order $v_1, \dots, v_n$ of the vertices such that $v_i v_{i+1} \dots v_{i+k-1}$ is an edge for every i divisible by $k-\ell$ (so any two such consecutive edges intersect in exactly ℓ vertices). Note that the case $\ell = k-1$ corresponds to a tight Hamilton cycle. Observe indeed that if n is divisible by $k-\ell$, then a tight Hamilton cycle on n vertices decomposes into $k-\ell$ Hamilton ℓ-cycles. Thus, by using Theorem 1.1 and averaging, we get that in every r-edge-colouring of a k-graph G on n vertices, with n divisible by $k-\ell$ and $\delta(G) \ge (1/2+\varepsilon)n$, there is a Hamilton ℓ-cycle with at least $(1+\mu)\frac{n}{r(k-\ell)}$ edges of the same colour. Unlike Theorem 1.1 and Corollary 1.2, here we do not know whether the constant $\frac{1}{2}$ is tight, i.e., whether minimum degree $\frac{n}{2}$ is necessary. See the concluding remarks for more on this.

1.1 Organisation of the paper

In Section 2, we provide an overview of the key steps of our proof. In Section 3, we summerise some known tools. Section 4 contains our key structural lemma which is already sufficient to prove the case of perfect matchings, i.e., Corollary 1.2. In Sections 5–7, we use additional methods to deal with Hamilton cycles. In the final section, we collect various other problems concerning discrepancy of spanning structures in hypergraphs which seem very interesting for further research.

1.2 Notation

For a set V and a natural number m, we write $\binom{V}{m}$ to denote the set of all m-subsets of V. We write $(V)_m$ to denote the set of all ordered m-tuples of distinct elements of V. We use capital letters with arrows above to denote ordered tuples $\overrightarrow{S} \in (V)_m$. We shall subsequently drop the arrow to denote the unordered version of this m-tuple, so that if $\overrightarrow{S} := (v_1, v_2, \dots, v_m)$, then S denotes the set $\{v_1, \dots, v_m\}$. Moreover, we write $\overleftarrow{S}$ to denote the ordered m-tuple obtained by reversing the ordering of $\overrightarrow{S}$, so that $\overleftarrow{S} := (v_m, v_{m-1}, \dots, v_1)$.

Let G be a k-graph. For $v \in V(G)$ and a $(k-1)$-set $S \subseteq V(G)$, we say that v is a *neighbour* of S in G if $S \cup \{v\}$ is an edge in G, and we denote the set of neighbours of S in G by $N_G(S)$. The *shadow* of G is the $(k-1)$-graph on $V(G)$ whose edges are the $(k-1)$-sets which are contained in at least one edge of G.

Given a tight path $P = v_1 v_2 \dots v_\ell$ on $\ell \ge k-1$ vertices in a k-graph, we say that P *connects* the ordered $(k-1)$-sets $(v_1, v_2, \dots, v_{k-1})$ and $(v_\ell, v_{\ell-1}, \dots, v_{\ell-k+2})$, which we call the *ends* of P. The choice of taking $(v_\ell, v_{\ell-1}, \dots, v_{\ell-k+2})$ rather than $(v_{\ell-k+2}, v_{\ell-k+3}, \dots, v_\ell)$ as an end of P is intentional and due to the fact that P is an undirected path. We call ℓ the *order* of P.

Given a 2-edge-colouring of G where we allow edges to receive multiple colours, we call the edges receiving both colours *double-coloured*.

For $a, b, c \in (0,1]$, we write $a \ll b \ll c$ in our statements to mean that there are increasing functions $f, g : (0,1] \to (0,1]$ such that whenever $a \le f(b)$ and $b \le g(c)$, then the subsequent result holds. Moreover, when using the Landau symbols $O(\cdot)$, $\Omega(\cdot)$, subscripts denote variables that the implicit constant may depend on.

We say that an event holds *with high probability (w.h.p.)* if the probability that it holds tends to 1 as the number of vertices n tends to infinity.

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

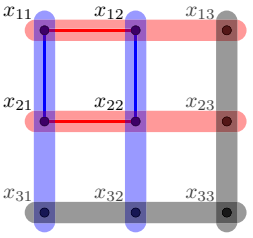


Figure 1. An alternating 2-grid on vertices $\{x_{11}, x_{12}, x_{21}, x_{22}\}$ and a near-alternating 3-grid on vertices $\{x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}, x_{31}, x_{32}, x_{33}\}$. The grey edges stand for edges whose colour is arbitrary.

2. Proof overview

Throughout this section, we let G be an r-edge-coloured n-vertex k-graph with $\delta(G) \geq (1/2 + \varepsilon)n$. We will first sketch a proof that G contains a perfect matching with high discrepancy. Subsequently we will discuss what more needs to be done in order to find a tight Hamilton cycle with high discrepancy.

2.1 Perfect matchings

We start by assuming that $r = 2$; we will handle the case of an arbitrary number of colours later on.

We aim to find an (edge-coloured) 'gadget' in G with the property that it contains a perfect matching where the majority colour is red and a perfect matching where the majority colour is blue. Such a gadget can then be used to 'push' the majority colour. A natural candidate is the alternating k-grid, which is defined as follows. A *k-grid* is the k-graph on vertices $\{x_{ij} : 1 \leq i, j \leq k\}$ and with edges $x_{i1} \cdots x_{ik}$ for each $i \in [k]$ (which we will call *horizontal* edges) and $x_{1j} \cdots x_{kj}$ for each $j \in [k]$ (which we will call *vertical* edges). An *alternating k-grid* is a 2-edge-coloured k-grid with all horizontal edges red and all vertical edges blue (cf. Figure 1). Observe that the horizontal edges form a red perfect matching and the vertical edges form a blue perfect matching. If we can find linearly many such vertex-disjoint gadgets, say $\varepsilon n/2k^2$ many, then, after removing them, the resulting k-graph G' still has $\delta(G') \geq (1/2 + \varepsilon/2)|V(G')|$ and hence has a perfect matching M' (cf. [24, 30]). Without loss of generality, assume that at least half of the edges of M' are red. Then, for each of the gadgets, take the red perfect matching. The union of all such matchings gives a perfect matching of G with at least $(1 - \frac{\varepsilon}{2})\frac{n}{2k} + \frac{\varepsilon}{2k}n = (1 + \frac{\varepsilon}{2})\frac{n}{2k}$ red edges, thus providing a perfect matching with high discrepancy.

Unfortunately, an alternating k-grid does not necessarily exist in G, not even if G is complete and has many edges of both colours. (For instance, if we choose a subset $A \subseteq V(G)$ and colour all edges which intersect A blue and all edges which do not intersect A red, then it is easily verified that there is no alternating k-grid.) However, our key lemma says that, unless the colouring is almost monochromatic, we can guarantee a *near-alternating k-grid*, namely, a 2-edge-coloured k-grid such that all horizontal edges but at most one are red, and all vertical edges but at most one are blue (cf. Figure 1).

(L1) If $\delta(G) \geq (1/2 + \varepsilon)n$, then either G contains a near-alternating k-grid, or G is almost monochromatic.

The formal statement of **(L1)** is offered by Lemma 4.1. We defer a proof sketch of this result to Section 4. Applying **(L1)** repeatedly gives that either G contains linearly many vertex-disjoint near-alternating k-grids, or its colouring is almost monochromatic. In the first case, one can apply the same argument as above, with alternating k-grids replaced by near-alternating k-grids: this

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

still works since, for each gadget, we can decide to cover its k^2 vertices with a perfect matching containing more red edges or more blue edges (as $k \geq 3$). In the second case, when almost all edges of G have the same colour, one can use standard methods to show that G contains an almost-monochromatic perfect matching. We remark this here for the benefit of the reader, but will actually not implement these steps since the result for perfect matchings will follow from the more general theorem about tight Hamilton cycles.

The case of an arbitrary number r of colours can be handled by identifying the first $r-1$ colours into one colour (say 'blue') and applying the 2-colour argument outlined above. The key point is that it suffices to find a perfect matching of G that either has at least $\frac{n}{rk} + \Omega(n)$ edges in the rth colour or at least $\frac{(r-1)n}{rk} + \Omega(n)$ edges in blue, because in the latter case, averaging gives that one of the first $r-1$ colours appears at least $\frac{n}{rk} + \Omega(n)$ times. Such a perfect matching can be found using the same strategy as above, consisting of first removing near-alternating grids and then covering the rest with a perfect matching M', and by applying a 'biased' case distinction for M'. Namely, in M', either a $(1-\frac{1}{r})$-fraction of edges is blue or a $\frac{1}{r}$-fraction has the rth colour, and whichever case holds, we can use the gadgets to 'push' the relevant colour(s).

2.2 Tight Hamilton cycles

We now discuss how to find a tight Hamilton cycle with high discrepancy. For hypergraphs with minimum degree above 1/2 (to which we refer informally as *Dirac hypergraphs*), there are well-known tools which allow one to connect a given set of disjoint tight paths into a tight Hamilton cycle (see Section 3.1). Therefore, if these paths are of high discrepancy and only 'miss' a small number of edges to close a Hamilton cycle, then we are done as, no matter which colours are used in the completion, the discrepancy cannot be ruined anymore. A crucial step of our argument is to find such paths (the formal statement is offered by Lemma 6.1).

(**L2**) If $\delta(G) \geq (1/2+\varepsilon)n$, then G contains a collection of vertex-disjoint tight paths, whose union contains $(1-o(1))n$ edges and has high discrepancy.

In order to show the above, we proceed as follows. We use our key lemma to find a *perfect fractional matching* $\mathbf{x}$ such that $\mathbf{x}$ is 'normal', i.e. each edge has weight $\Theta(n^{-k+1})$, and such that $\mathbf{x}$ has high discrepancy, in the sense that the total weight received by some colour class (say 'red') is significantly above the average, i.e. larger than $n/(rk) + \Omega(n)$ (see Lemma 5.2, and see Section 3.1 for the definition of a perfect fractional matching). We then use $\mathbf{x}$ to define a random walk $\mathcal{Y}$ on $V(G)$, such that a path of order t sampled according to the first t vertices of $\mathcal{Y}$ (conditioning on being self avoiding) has the following properties: Every vertex is approximately equally likely to be contained in the path and the probability of an edge e appearing in the path is roughly proportional to $\mathbf{x}(e)$ (see Lemma 5.3).

We sample N paths of order t independently, where t is a sufficiently large constant and $N := n^{t-1/2}$. Finally, we define an auxiliary t-uniform hypergraph $\mathcal{H}$ with vertex set $V(G)$ and edges corresponding to the vertex sets of the sampled paths. (The choice of N ensures that this hypergraph is rather dense, but still we do not expect too many parallel edges.) Owing to the first property of $\mathcal{Y}$ mentioned above, the t-graph $\mathcal{H}$ is almost-regular and, obviously, its maximum 2-degree is at most n^{t-2}. By a fundamental theorem of Frankl, Rödl [13], and Pippenger (see [28]), we can then establish that $\mathcal{H}$ contains an almost-perfect matching, which corresponds to a collection of vertex-disjoint tight paths of order t covering almost all the vertices. However, our goal is to find such a collection with high discrepancy. Owing to the second property of $\mathcal{Y}$ mentioned above and the fact that the weight of red edges is significantly above the average, the paths in G which correspond to the edges in $\mathcal{H}$ are likely to contain many red edges. A result concerning finding hypergraph matchings with pseudorandom properties due to Ehard, Glock, and Joos [9]

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

(see Theorem 3.5) then allows us to find an almost-perfect matching in $\mathcal{H}$ such that the corresponding collection of vertex-disjoint paths indeed has high discrepancy. This proves **(L2)** and establishes Theorem 1.1.

3. Preliminaries

In this section, we collect some preliminary results.

3.1 Dirac hypergraphs

In this subsection we state three results concerning hypergraphs with minimum $(k-1)$-degree above $n/2$. The first is the so-called 'Connecting Lemma' which allows to connect any two disjoint ordered $(k-1)$-tuples of vertices by a tight path of bounded length.

Lemma 3.1 (Lemma 2.4 in [29]). *Let $1/n \ll \varepsilon \ll 1/k$ with $k \in \mathbb{N}$ and $k \geq 3$. Let G be a k-graph on n vertices with $\delta(G) \geq (1/2+\varepsilon)n$. Then for every two disjoint ordered $(k-1)$-subsets of vertices of G, there is a tight path in G of order at most $2k/\varepsilon^2$ which connects them.*

The second result is the so-called 'Absorbing Lemma' and shows the existence of an absorber for tight paths in Dirac hypergraphs.

Lemma 3.2 (Lemma 2.1 in [29]). *Let $1/n \ll \beta \ll \mu \ll \varepsilon, 1/k$, with $k \in \mathbb{N}$ and $k \geq 3$. Let G be a k-graph on n vertices with $\delta(G) \geq (1/2+\varepsilon)n$. Then there exists a tight path P with $|V(P)| \leq \mu n$ such that for every subset $W \subseteq V(G) \setminus V(P)$ of size $|W| \leq \beta n$, there is a tight path $\tilde{P}$ in G with $V(\tilde{P}) = V(P) \cup W$ and such that $\tilde{P}$ has the same ends as P.*

The third result concerns the existence of 'balanced' perfect fractional matchings. Let us introduce the relevant definitions. Let G be an n-vertex k-graph. A *perfect fractional matching* of G is a function $\mathbf{x}\colon E(G) \to [0,1]$ such that, for every vertex $v \in V(G)$, we have $\sum_{e \ni v} \mathbf{x}(e) = 1$. Observe that this implies that $\sum_{e \in E(G)} \mathbf{x}(e) = n/k$, which we will use throughout without mention. For $\mu \in (0,1]$, a perfect fractional matching $\mathbf{x}$ is said to be *μ-normal* if $\mu n^{-k+1} \leq \mathbf{x}(e) \leq \mu^{-1} n^{-k+1}$ for all $e \in E(G)$. The following result states that Dirac hypergraphs have normal perfect fractional matchings.

Lemma 3.3 (Lemma 4.2 in [19]). *Let $1/n \ll \mu \ll \varepsilon, 1/k$ with $k \in \mathbb{N}$ and $k \geq 2$. Let G be an n-vertex k-graph with $\delta(G) \geq (1/2+\varepsilon)n$. Then G has a μ-normal perfect fractional matching.*

3.2 Probabilistic tools

We will often apply the following standard Chernoff-type concentration inequality (see [8, Theorem 1.1]).

Lemma 3.4 (Chernoff's inequality). *Let $X_1, \ldots, X_N$ be independent random variables taking values in $[0,1]$, and let $X = \sum_{i=1}^{N} X_i$. Then for every $0 < \beta < 1$,*

$$\mathbb{P}\Big[|X - \mathbb{E}[X]| \geq \beta \mathbb{E}[X]\Big] \leq 2\exp\left(-\frac{\beta^2}{3}\mathbb{E}[X]\right).$$

As explained at the end of Section 2, our proof will use a hypergraph matching argument and will need the matching to look random-like with respect to some properties. In order to achieve that, we use a nibble-type result due to Ehard, Glock, and Joos [9]. For a hypergraph $\mathcal{H}$, define $\Delta(\mathcal{H}) := \max_{v \in V(\mathcal{H})} d_{\mathcal{H}}(\{v\})$ and $\Delta^c(\mathcal{H}) := \max_{\{u,v\} \subseteq V(\mathcal{H})} d_{\mathcal{H}}(\{u,v\})$. We will consider edge weight functions $w\colon E(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ and, for a set $A \subseteq E(\mathcal{H})$, we use the notation $w(A) := \sum_{e \in A} w(e)$.

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

Theorem 3.5 (Theorem 1.2 in [9]). *Suppose $\delta \in (0,1)$ and $t \in \mathbb{N}$ with $t \geq 2$, and set $\gamma := \delta/50t^2$. Then there exists Δ_0 such that for all $\Delta \geq \Delta_0$, the following holds. Let $\mathcal{H}$ be a t-uniform hypergraph satisfying $\Delta(\mathcal{H}) \leq \Delta$, $\Delta^c(\mathcal{H}) \leq \Delta^{1-\delta}$ and $e(\mathcal{H}) \leq \exp(\Delta^{\gamma^2})$. Let $\mathcal{W}$ be a set of at most $\exp(\Delta^{\gamma^2})$ weight functions on $E(\mathcal{H})$ such that $w(E(\mathcal{H})) \geq \max_{e \in E(\mathcal{H})} w(e)\Delta^{1+\delta}$ for every $w \in \mathcal{W}$. Then there exists a matching $\mathcal{M}$ in $\mathcal{H}$ such that $w(\mathcal{M}) = (1 \pm \Delta^{-\gamma})w(E(\mathcal{H}))/\Delta$ for every $w \in \mathcal{W}$.*

4. Key lemma

In this section, we state and prove our key structural lemma. Recall from Section 2 that a *k-grid* is the k-graph on vertices $\{x_{ij} : 1 \leq i, j \leq k\}$ and with edges $x_{i1} \cdots x_{ik}$ for each $i \in [k]$ (which we will call *horizontal* edges) and $x_{1j} \cdots x_{kj}$ for each $j \in [k]$ (which we will call *vertical* edges). An *alternating k-grid* is a 2-edge-coloured k-grid with all the horizontal edges red and all vertical edges blue. A *near-alternating k-grid* is a 2-edge-coloured k-grid such that all horizontal edges but (at most) one are red, and all vertical edges but (at most) one are blue. Thus, the difference between a near-alternating k-grid and an alternating k-grid lies in not prescribing the colour of one horizontal and one vertical edge (cf. Figure 1).

Our key lemma exploits the specific structure of the given colouring and shows that either it contains a near-alternating k-grid or the colouring is almost monochromatic. While our main result applies to hypergraphs coloured with any number of colours, it suffices to handle here the case of only two colours.

Lemma 4.1 (Key lemma). *Let $1/n \ll \zeta \ll \varepsilon, \rho, 1/k$ with $k \in \mathbb{N}$ and $k \geq 3$. Let G be an n-vertex k-graph whose edges are 2-coloured. Assume that all but at most ζn^{k-1} $(k-1)$-subsets S of $V(G)$ satisfy $d_G(S) \geq (1/2 + \varepsilon)n$. Then either there exists a near-alternating k-grid, or one of the colour classes has size at most ρn^k.*

As explained in Section 2, Lemma 4.1 is enough to derive Corollary 1.2. Moreover, it is easy to see that the lemma also holds for $k = 2$, but then a near-alternating grid is not a suitable gadget, because such a grid (which is a 4-cycle) might only have perfect matchings with one red edge and one blue edge, so that no colour appears more often. Hence we ignore this case here.

We now sketch the proof of Lemma 4.1. It is helpful to first guarantee that if a $(k-1)$-set of vertices is contained in an edge of a certain colour, then it is actually contained in many edges of this colour. While this is not true for an arbitrary edge-colouring, we can make sure this is the case after deleting only few edges. The required cleaning procedure is given by the following standard tool. We provide the short proof for the convenience of the reader.

Proposition 4.2. *Let G be a k-graph on n vertices. Then by removing at most $t\binom{n}{k-1}$ edges, one can ensure that in the resulting subhypergraph, every $(k-1)$-set has degree either 0 or at least t.*

Proof. As long as there is a $(k-1)$-set S with degree between 1 and t, delete all edges containing S. Obviously, every $(k-1)$-set is considered at most once during this process, and when it is considered, we delete at most t edges. □

Once G has been cleaned (with respect to both colours and to a suitable choice of t), we obtain a subhypergraph G' with the desired property for both colours. Observe that, since we removed only few edges, for most of the $(k-1)$-subsets of $V(G)$, their degree in G' will still be linearly above $n/2$.

Let H be the $(k-1)$-shadow of G', and equip H with the following 2-colouring of its edges: Colour an edge of H with colour c if it is contained in at least one edge of G' of colour c (and hence at least t such edges). We note that an edge of H can receive both colours (if it is contained in an edge of G' of each of the colours). As we will show, by choosing t appropriately, it suffices to find an alternating $(k-1)$-grid in H in order to obtain a near-alternating k-grid in G (see Claim 1).

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

If H contains many double-coloured edges, then we can easily find an alternating $(k-1)$-grid. In fact, we can use the following classical result of Erdős [10], which states that the Turán density of k-partite k-graphs is 0.

Theorem 4.3 ([10]). *Let $1/n \ll \eta, 1/\ell$ with $\ell \in \mathbb{N}$. Let L be any k-partite k-graph with ℓ vertices. Then any k-graph with n vertices and at least ηn^k edges contains L as a subgraph.*

Call an edge of H bad if it is double-coloured or has small degree in G'. Then by Theorem 4.3 and what we observed above, we can assume that only few edges of H are bad. We would like to argue that all the edges of H which are not bad are then coloured with the same unique colour, which in turn will imply that G itself is almost-monochromatic.

We proceed as follows: Let $\overrightarrow{T_1}$ and $\overrightarrow{T_2}$ be arbitrary orderings of any two edges $T_1, T_2 \in E(H)$ which are not bad. The key observation we use is the following: Suppose there is a tight walk in G' connecting $\overrightarrow{T_1}$ and $\overrightarrow{T_2}$ such that any $(k-1)$ consecutive vertices of the walk give an edge of H which is not bad. The power of the cleaning procedure then comes in handy, as we can claim that each $(k-1)$-set along the walk is contained in edges of G' of the same unique colour. Therefore the colour information propagates from T_1 to T_2 and the tight walk must be monochromatic, implying that T_1 and T_2 must have the same colour.

To utilise the above observation, we want to be able to connect every (or almost every) pair of ordered edges $\overrightarrow{T_1}, \overrightarrow{T_2}$, as above. The standard tool to obtain this connection is Lemma 3.1. However, we cannot apply this lemma directly, as the required minimum degree does not hold in G'. Nevertheless, by randomly sampling a small set of vertices, we can avoid all the bad $(k-1)$-sets (as these are few) and thus apply Lemma 3.1 within the sample. To make this work, we perform another cleaning step which removes $(k-1)$-sets that intersect with too many bad sets.

We are now ready to prove the key lemma.

Proof of Lemma 4.1. Let $k \in \mathbb{N}$ with $k \geq 3$ and $\varepsilon, \rho > 0$ be given, and observe that we can assume ε to be small enough for Lemma 3.1 to hold. Let

$$1/n \ll \zeta \ll \eta \ll \varepsilon, \rho, 1/k\,.$$

Let G be a 2-edge-coloured (with red and blue) n-vertex k-graph on V such that all but at most ζn^{k-1} of the $(k-1)$-subsets of V satisfy $d_G(S) \geq (1/2+\varepsilon)n$. We apply Proposition 4.2 with parameter $t := \eta\varepsilon n$ twice, once to the subhypergraph induced by the red edges and once to the subhypergraph induced by the blue edges. This gives a subhypergraph G' of G with $e(G') \geq e(G) - 2\eta\varepsilon n^k$ such that each $(k-1)$-set S satisfies the following condition for each of the colours: either S is not contained in any edge of G' of this colour, or it is contained in at least $\eta\varepsilon n$ edges of G' of this colour.

Let H be the $(k-1)$-uniform shadow of G', equipped with the following 2-colouring of its edges: Colour an edge of H red (resp. blue) if it is contained in at least one red (resp. blue) edge of G' (and hence at least $\eta\varepsilon n$ such edges), noting that we allow edges of H to receive both colours. For the purpose of finding coloured structures in H, an edge that has both colours can be used either way. □

Claim 1. *If H contains an alternating $(k-1)$-grid, then G contains a near-alternating k-grid.*

Proof of Claim 1. Suppose H contains an alternating $(k-1)$-grid. Then there exists $W := \{x_{ij} : 1 \leq i,j \leq k-1\} \subseteq V$ such that $x_{i1}\cdots x_{i(k-1)}$ is a red edge of H for each $i \in [k-1]$ and $x_{1j}\cdots x_{(k-1)j}$ is a blue edge of H for each $j \in [k-1]$. Let $\mathcal{R}$ be the set of ordered tuples $(x_{1k}, x_{2k}, \ldots, x_{(k-1)k})$ such that $x_{1k}, x_{2k}, \ldots, x_{(k-1)k} \in V \setminus W$ are pairwise distinct, $\{x_{i1}\cdots x_{ik}\}$ is a red edge of G for each $i \in [k-1]$, and $d_G(\{x_{1k}, x_{2k}, \ldots, x_{(k-1)k}\}) \geq (1/2+\varepsilon)n$. Owing to the fact that a red edge of H can be extended to a red edge of G' (and hence G) in at least $\eta\varepsilon n$ ways and that for all but ζn^{k-1} of the $(k-1)$-subsets S of V it holds that $d_G(S) \geq (1/2+\varepsilon)n$, we have

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

$|\mathcal{R}| \geq (\eta\varepsilon n)^{k-1} - n^{k-2} - (k-1)! \cdot \zeta n^{k-1} \geq (\eta\varepsilon n)^{k-1}/2$, where we used $\zeta \ll \eta, \varepsilon, 1/k$ (and the n^{k-2} term accounts for the choices of $x_{1k}, \ldots, x_{(k-1)k}$ which are not pairwise distinct). Similarly, let $\mathcal{B}$ be the set of ordered tuples $(x_{k1}, x_{k2}, \ldots, x_{k(k-1)})$ such that $x_{k1}, x_{k2}, \ldots, x_{k(k-1)} \in V \setminus W$ are pairwise distinct, $\{x_{1j} \cdots x_{kj}\}$ is a blue edge of G for each $j \in [k-1]$, and $d_G(\{x_{k1}, x_{k2}, \ldots, x_{k(k-1)}\}) \geq (1/2 + \varepsilon)n$. As above for $\mathcal{R}$, we have $|\mathcal{B}| \geq (\eta\varepsilon n)^{k-1}/2$. Therefore, there exist vertex-disjoint $R \in \mathcal{R}$ and $B \in \mathcal{B}$. As $d_G(R), d_G(B) \geq (1/2 + \varepsilon)n$, we have $|N_G(R) \cap N_G(B)| \geq 2(1/2 + \varepsilon)n - n = 2\varepsilon n$. Hence, there exists $x_{kk} \in (N_G(R) \cap N_G(B)) \setminus W$, i.e. a vertex x_{kk} which forms an edge of G with both R and B (although we have no control of the colours of these edges). This gives a near-alternating k-grid in G, as desired. □

Owing to Claim 1, from now on we can assume that H does not contain an alternating $(k-1)$-grid. We show that then H (and thus G) must be almost monochromatic. A $(k-1)$-subset S of V is called *bad* if $d_{G'}(S) < (1/2 + \varepsilon/2)n$ or if, seen as an edge of H, S is coloured with both colours. We bound the number of bad sets as follows.

Claim 2. *There are at most $5k\eta n^{k-1}$ bad sets.*

Proof of Claim 2. We begin by bounding the number of $(k-1)$-sets S with $d_{G'}(S) < (1/2 + \varepsilon/2)n$. Observe that if $d_{G'}(S) < (1/2 + \varepsilon/2)n$, then either $d_G(S) < (1/2 + \varepsilon)n$ (i.e. S has small degree already in G), or we removed at least $\varepsilon n/2$ edges of G containing S during the cleaning process (i.e. when obtaining G' from G). The former case holds for at most ζn^{k-1} sets S, by assumption. Let us now bound the number of S in the latter case. Since removing an edge of G decreases (by one) the degree of precisely k $(k-1)$-sets, and as the total number of removed edges is at most $2\eta\varepsilon n^k$, the number of such sets S is at most $\frac{k \cdot 2\eta\varepsilon n^k}{\varepsilon n/2} = 4k\eta n^{k-1}$.

Next, we bound the number of double-coloured edges of H. Let $\tilde{H}$ be the subhypergraph of H on V induced by the double-coloured edges. Then $e(\tilde{H}) \leq \eta n^{k-1}$. Indeed, otherwise $\tilde{H}$ contains a copy of the $(k-1)$-grid by Theorem 4.3, since the $(k-1)$-grid is a $(k-1)$-partite $(k-1)$-graph (with each vertex class of size $k-1$). However, as each edge of $\tilde{H}$ receives both colours, this gives an alternating $(k-1)$-grid, which is a contradiction.

Summarising, the number of bad sets is at most $(\zeta + 4k\eta + \eta)n^{k-1} \leq 5k\eta n^{k-1}$, where we used $\zeta \ll \eta$. □

Given a $(k-1)$-subset S of V, we say that S is *clean* if, for every $0 \leq j \leq k-1$, the number of bad sets T with $|S \cap T| = j$ is at most $\eta^{1/2} n^{k-1-j}$. We remark that for $j = k-1$ the condition means that S itself is not bad.

Claim 3. *All but at most $\eta^{1/3} n^{k-1}$ of the $(k-1)$-subsets of V are clean.*

Proof of Claim 3. For each $0 \leq j \leq k-1$, we bound the number of $(k-1)$-sets S for which the number of bad sets T with $|S \cap T| = j$ is more than $\eta^{1/2} n^{k-1-j}$. If $j = 0$, this cannot happen by Claim 2, so we can assume $0 < j \leq k-1$. Given a bad set T, the number of $(k-1)$-sets S intersecting T in j vertices is at most $\binom{k-1}{j} n^{k-1-j}$. Therefore, using the bound in Claim 2, the number of $(k-1)$-subsets of V which are not clean is at most

$$\sum_{j=1}^{k-1} \frac{\binom{k-1}{j} n^{k-1-j} \cdot 5k\eta n^{k-1}}{\eta^{1/2} n^{k-1-j}} \leq \eta^{1/3} n^{k-1}.$$

□

Let H' be the $(k-1)$-subhypergraph of H consisting of the edges which are clean $(k-1)$-sets. Observe in particular that each edge of H' has a unique colour (because double-coloured edges are bad).

Claim 4. *All edges of H' have the same colour.*

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

Using Claim 4, we can easily complete the proof of the lemma. Indeed, suppose without loss of generality that the edges of H' are red. Then for an edge of G to be blue, it has to be either an edge of $G \setminus G'$, or an edge of G' none of whose $(k-1)$-subsets belongs to H'. Using that $e(G) - e(G') \leq 2\eta\varepsilon n^k$ and $e(H) - e(H') \leq \eta^{1/3} n^{k-1}$ (by Claim 3), we get that the number of blue edges of G is at most $2\eta\varepsilon n^k + n \cdot \eta^{1/3} n^{k-1} \leq \rho n^k$, as wanted. We are left to prove Claim 4.

Proof of Claim 4. Consider two arbitrary edges $T_1, T_2 \in E(H')$. We will show that they have the same colour. Fix C, β with $\eta \ll 1/C \ll \beta \ll \varepsilon, \rho, 1/k$. We claim that there exists a set $R \subseteq V$ such that

(i) $|R| \geq C/2$;

(ii) for $i = 1, 2$, for every $(k-1)$-set S which is contained in $R \cup T_i$ and is not bad, it holds that $|N_{G'}(S) \cap R| \geq (1/2 + \varepsilon/8)|R \cup T_i|$;

(iii) no $(k-1)$-subset contained in $R \cup T_1$ or $R \cup T_2$ is bad.

Let $R \subseteq V$ be obtained by independently including each vertex of V with probability C/n. We show that such R satisfies (i), (ii) and (iii) with positive probability.

Since $\mathbb{E}[|R|] = C$ and $1/C \ll \beta$, an easy application of Chernoff's inequality (Lemma 3.4) shows that $(1-\beta)C \leq |R| \leq (1+\beta)C$ with probability at least 0.9.

Fix $i \in \{1, 2\}$, let $S \subseteq V$ be a $(k-1)$-set which is not bad, and define the events $A_S := \{S \subseteq R \cup T_i\}$ and $B_S := \{X_S < (1/2 + \varepsilon/4)C\}$, where $X_S := |N_{G'}(S) \cap R|$. Furthermore, let $j = j_S := |S \cap T_i|$, so $0 \leq j \leq k-1$. We are going to show that with probability at least 0.9, the event $A_S \cap B_S$ does not hold for any such S. Note that A_S and B_S are independent, and $\mathbb{P}[A_S] = \left(\frac{C}{n}\right)^{k-1-j}$.

Since S is not bad, we have $\mathbb{E}[X_S] = |N_{G'}(S)| \cdot \frac{C}{n} \geq (1/2 + \varepsilon/2)C$ and, using Chernoff's inequality (Lemma 3.4), we conclude that $\mathbb{P}[B_S] \leq 2\exp\left(-\frac{(\varepsilon/4)^2}{3}\mathbb{E}[X_S]\right) = \exp(-\Omega_\varepsilon(C))$. Therefore, $\mathbb{P}[A_S \cap B_S] \leq \left(\frac{C}{n}\right)^{k-1-j} \cdot \exp(-\Omega_\varepsilon(C))$. By taking the union bound over the at most $2^{k-1} n^{k-1-j}$ $(k-1)$-sets $S \subseteq V$ with $j_S = j$, we see that the probability that $A_S \cap B_S$ holds for some such S is at most $\left(\frac{C}{n}\right)^{k-1-j} \cdot \exp(-\Omega_\varepsilon(C)) \cdot 2^{k-1} n^{k-1-j} \leq 0.1k^{-1}$, where we used that $1/C \ll \varepsilon, 1/k$. The conclusion now follows by taking another union bound over the k choices for j.

Next, let Y be the random variable counting the number of bad sets contained in $R \cup T_i$, and observe that $Y = \sum_{j=0}^{k-1} Y_j$, where Y_j counts the number of bad sets $T \subseteq R \cup T_i$ with $|T \cap T_i| = j$. Since T_i is a clean set, the number of bad sets T with $|T \cap T_i| = j$ is at most $\eta^{1/2} n^{k-1-j}$. Therefore, $\mathbb{E}[Y_j] \leq \eta^{1/2} n^{k-1-j} \cdot \left(\frac{C}{n}\right)^{k-1-j} \leq \eta^{1/2} C^{k-1}$ and hence $\mathbb{E}[Y] \leq k\eta^{1/2} C^{k-1} \leq 0.1$, using $\eta \ll 1/C$. By Markov's inequality, we have that $Y = 0$ with probability at least 0.9.

Therefore, with positive probability we have that $(1-\beta)C \leq |R| \leq (1+\beta)C$, the event $A_S \cap B_S$ does not hold for any $(k-1)$-set which is not bad (for both $i = 1, 2$), and no $(k-1)$-subset contained in $R \cup T_1$ or $R \cup T_2$ is bad. These in turn imply properties (i)-(iii). Indeed, (i) and (iii) follow directly from the above, and for (ii) it is enough to observe that if S is not bad and is contained in $R \cup T_i$ (i.e. A_S holds), then B_S cannot hold and thus $X_S \geq (1/2 + \varepsilon/4)C \geq (1/2 + \varepsilon/8)|R \cup T_i|$, where the first inequality follows from the definition of B_S, and the second inequality uses $|R| \leq (1+\beta)C$, $|T_i| = k-1$, and $1/C, \beta \ll \varepsilon, 1/k$.

We conclude that a set R satisfying (i), (ii) and (iii) does indeed exist. Note that (ii)–(iii) imply that $\delta(G'[R \cup T_i]) \geq (1/2 + \varepsilon/8)|R \cup T_i|$. Moreover, by (i), $|R|$ is large enough to apply Lemma 3.1 to $G'[R \cup T_1]$ and $G'[R \cup T_2]$. Now, fix two arbitrary orderings $\overrightarrow{T_1}$ and $\overrightarrow{T_2}$ of T_1, T_2, respectively, and let $\overrightarrow{T}$ be an arbitrary ordering of a $(k-1)$-set $T \subseteq R \setminus (T_1 \cup T_2)$. Using Lemma 3.1 twice, we find a tight path P_1 connecting $\overrightarrow{T_1}$ and $\overrightarrow{T}$ in $G'[R \cup T_1]$ and a tight path P_2 connecting $\overrightarrow{T_2}$ and $\overrightarrow{T}$ in $G'[R \cup T_2]$. Owing to (iii), we know that every $k-1$ consecutive vertices in P_1 form a set

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

which is not bad, which means that all edges of G' containing this set have the same colour. By the definition of the colouring of H, every such $(k-1)$-set has the same colour and, in particular, this holds for T_1 and T. By repeating the same argument for P_2, this holds for T_2 and T. We conclude that T_1 and T_2 have the same colour, as desired.

This concludes the proof of Lemma 4.1. □

5. Perfect fractional matchings with high discrepancy

In this section we focus on the random walk which we will use to sample a collection of paths of high discrepancy. But first, recalling the definition of perfect fractional matchings from Section 3.1, we introduce the following notation: For a perfect fractional matching $\mathbf{x}\colon E(G)\to[0,1]$ of a k-graph G, and for a set $S\subseteq V(G)$ with $|S|\le k$, define $\mathbf{x}(S):=\sum_{e\in E(G)\colon S\subseteq e}\mathbf{x}(e)$. Note that $\mathbf{x}(\{v\})=1$ for all $v\in V(G)$ and $\mathbf{x}(S)=0$ for all $S\subseteq V(G)$ with $|S|=k$ and $S\notin E(G)$.

We define the following random walk $\mathcal{Y}=(Y_1,Y_2,\dots)$ on $V:=V(G)$. It begins with an ordered $(k-1)$-tuple $(Y_1,\dots,Y_{k-1})\in(V)_{k-1}$ chosen according to the following *initial distribution* $\pi\colon(V)_{k-1}\to[0,1]$. Pick an ordered $(k-1)$-set $(Y_1,\dots,Y_{k-1})\in(V)_{k-1}$ at random with probability proportional to $\mathbf{x}(\,\cdot\,)$, that is, for any $\overrightarrow{S}\in(V)_{k-1}$,

$$\pi(\overrightarrow{S}):=\frac{\mathbf{x}(S)}{\sum_{\overrightarrow{S'}\in(V)_{k-1}}\mathbf{x}(S')}\,. \tag{5.1}$$

Observe that the denominator of (5.1) can be rewritten as

$$\sum_{\overrightarrow{S'}\in(V)_{k-1}}\mathbf{x}(S')=k!\cdot\sum_{e\in E(G)}\mathbf{x}(e)=(k-1)!\cdot n\,, \tag{5.2}$$

where we used that every edge contains $k!$ ordered $(k-1)$-sets. The *transition probability* will also be defined according to $\mathbf{x}$. For all $i\ge k-1$, conditional on the outcome of $Y_{i-(k-2)},\dots,Y_i$, we choose the next vertex Y_{i+1} as follows: Let $\overrightarrow{Z_i}:=(Y_{i-(k-2)},\dots,Y_i)$ be the ordered set of the last $k-1$ vertices in the sequence and choose Y_{i+1} with probability proportional to $\mathbf{x}(Z_i\cup\{\cdot\})$, that is, for any $v\in V\setminus Z_i$,

$$\mathbf{Pr}[Y_{i+1}=v\,|\,Y_{i-(k-2)},\dots,Y_i]=\frac{\mathbf{x}(Z_i\cup\{v\})}{\sum_{v'\in V\setminus Z_i}\mathbf{x}(Z_i\cup\{v'\})}=\frac{\mathbf{x}(Z_i\cup\{v\})}{\mathbf{x}(Z_i)}\,, \tag{5.3}$$

and for any $v\in Z_i$ the transition probability is 0. Observe that $\mathcal{Y}$ is equivalent to the random walk $\mathcal{Z}:=(\overrightarrow{Z_{k-1}},\overrightarrow{Z_k},\dots)$, and we can refer to both. In fact, with $(V)_{k-1}$ viewed as the state space, $\mathcal{Z}$ is a Markov chain and it is then easy to check that the distribution π defined in (5.1) is stationary (cf. [19, Proposition 5.5]). The important fact of defining the transition probabilities in terms of the perfect fractional matching $\mathbf{x}$ is that then the random walk behaves uniformly with respect to the visited vertices, in the sense that the distribution of the vertices for $\mathcal{Y}$ to visit at any step is uniform over $V(G)$ as proved below.

Fact 5.1. *For each integer $i\ge1$ and $v\in V(G)$, we have that $\mathbf{Pr}[Y_i=v]=1/n$. Moreover, for each $k\le i\le t$, with $e_i:=\{Y_{i-k+1},\dots,Y_i\}$ denoting the $(i-k+1)$-st edge of $\mathcal{Y}$, the following holds: For each $e\in E(G)$, we have that $\mathbf{Pr}[e_i=e]=\frac{k}{n}\mathbf{x}(e)$.*

Proof. Let $v\in V$. For $1\le i\le k-1$, observe the following identity where the first sum runs over all $\overrightarrow{S}\in(V)_{k-1}$ such that v is the i-th element of $\overrightarrow{S}$, and the second sum runs over all $S\in\binom{V}{k-1}$ with $v\in S$:

$$\sum_{\substack{\overrightarrow{S}=(v_1,\dots,v_{k-1})\\ v_i=v}}\mathbf{x}(S)=(k-2)!\cdot\sum_{S\in\binom{V}{k-1}\colon v\in S}\mathbf{x}(S)=(k-1)!\cdot\sum_{e\colon e\ni v}\mathbf{x}(e)=(k-1)!\,.$$

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

Together with (5.2), we get $\mathbf{Pr}[Y_i = v] = 1/n$.

Since π as defined in (5.1) is the stationary distribution of $\mathcal{Z}$, it holds that $\mathbf{Pr}[\overrightarrow{Z_i} = \overrightarrow{S}] = \pi(\overrightarrow{S})$ for each $i \geq k-1$. Let $i \geq k$. Then by the law of total probability we have

$$\begin{aligned}\mathbf{Pr}[Y_i = v] &= \sum_{\overrightarrow{S} \in (V)_{k-1}} \mathbf{Pr}\left[Y_i = v | \overrightarrow{Z_{i-1}} = \overrightarrow{S}\right] \cdot \mathbf{Pr}\left[\overrightarrow{Z_{i-1}} = \overrightarrow{S}\right] \\ &= \sum_{\overrightarrow{S} \in (V)_{k-1}} \mathbf{Pr}\left[Y_i = v | \overrightarrow{Z_{i-1}} = \overrightarrow{S}\right] \cdot \pi(\overrightarrow{S}) \\ &= \sum_{\substack{\overrightarrow{S} \in (V)_{k-1}: \\ v \notin S}} \frac{\mathbf{x}(S \cup \{v\})}{\cancel{\mathbf{x}(\mathcal{S})}} \cdot \frac{\cancel{\mathbf{x}(\mathcal{S})}}{(k-1)! \cdot n} = \frac{(k-1)! \cdot \sum_{e \ni v} \mathbf{x}(e)}{(k-1)! \cdot n} = \frac{1}{n},\end{aligned}$$

where we used (5.2), that for an edge e with $e \ni v$ there are $(k-1)!$ choices for $\overrightarrow{S} \in (V)_{k-1}$ such that $S \cup \{v\} = e$, and that $\mathbf{x}$ is a perfect fractional matching.

For the 'moreover'-part of Fact 5.1, let $k \leq i \leq t$ and $e \in E(G)$. Then

$$\begin{aligned}\mathbf{Pr}[e_i = e] &= \sum_{\overrightarrow{S} \in (V)_{k-1}} \mathbf{Pr}\left[e_i = e | \overrightarrow{Z_{i-1}} = \overrightarrow{S}\right] \cdot \mathbf{Pr}\left[\overrightarrow{Z_{i-1}} = \overrightarrow{S}\right] \\ &= \sum_{\overrightarrow{S} \in (V)_{k-1} \colon S \subseteq e} \frac{\mathbf{x}(e)}{\cancel{\mathbf{x}(\mathcal{S})}} \cdot \frac{\cancel{\mathbf{x}(\mathcal{S})}}{(k-1)! \cdot n} = \frac{k}{n} \cdot \mathbf{x}(e),\end{aligned}$$

where we used (5.2) and that there are $k!$ ways to choose $\overrightarrow{S} \in (V)_{k-1}$ such that $S \subseteq e$. □

Despite Fact 5.1, for an arbitrary perfect fractional matching $\mathbf{x}$, the behaviour of the random walk $\mathcal{Y}$ can still be quite trivial. For instance, suppose that $\mathbf{x}$ is indeed a perfect matching M, that is $\mathbf{x}(e) = \mathbb{1}(e \in M)$. Then, once the first ordered $(k-1)$-set $\overrightarrow{Z_{k-1}}$ is chosen, the walk is completely deterministic (and uses the same edge in each step). In order to avoid this in a robust way, we will assume that $\mathbf{x}$ is μ-normal for some (small) constant $\mu > 0$ (see Section 3.1 for the definition).

Ultimately, we would like to use the random walk $\mathcal{Y}$ to sample a collection of tight paths which cover almost all the vertices of G and have high discrepancy. As proved in Fact 5.1, the probability that $\mathcal{Y}$ sees a certain edge e is proportional to $\mathbf{x}(e)$. Therefore, in order to guarantee that $\mathcal{Y}$ sees a substantial number of edges of the same colour, it will be enough that $\mathbf{x}$ is a normal perfect fractional matching with high discrepancy. Recall that the total weight assigned by any perfect fractional matching is n/k, so, just by averaging, some colour will receive a total weight of at least $n/(rk)$. By combining Lemma 3.3 with our key lemma (Lemma 4.1), we are able to boost the discrepancy. Starting with a perfect fractional matching given by Lemma 3.3, if we can find a near-alternating k-grid (which we refer to as a *gadget* from now on), then by increasing the weight of the red matching, say, and decreasing the weight of the blue matching by the same amount, the total weight of each vertex remains unchanged, but the total weight of the red edges has increased. Obviously, one gadget will only allow us to perform an insignificant perturbation, but by applying the key lemma iteratively, we can find many edge-disjoint gadgets and together they allow us to perturb the initial perfect fractional matching by a significant amount.

Lemma 5.2. *Let $1/n \ll \mu \ll \varepsilon, 1/r, 1/k$ with $k, r \in \mathbb{N}$, $k \geq 3$ and $r \geq 2$. Let G be an n-vertex k-graph with $\delta(G) \geq (1/2 + \varepsilon)n$ whose edges are r-coloured. Then G has a μ-normal perfect fractional matching such that the total weight received by some colour class is at least $(1+\mu)\frac{n}{rk}$.*

Proof. Let

$$1/n \ll \mu \ll \eta \ll \zeta \ll \rho \ll \mu_0 \ll \varepsilon, 1/r, 1/k,$$

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

where μ_0 is small enough for Lemma 3.3 to hold on input k and ε, and ζ is small enough for Lemma 4.1 to hold on input $k, \varepsilon/2$ and $\rho/2$. Let $\mathbf{x}_0$ be a μ_0-normal perfect fractional matching as given by Lemma 3.3.

Let $c_1, \dots, c_r$ be the r colours used in the edge-colouring of G. Since Lemma 4.1 is only stated for 2-edge-coloured graphs, we will now consider the following 2-edge-colouring of G, obtained by identifying $r-1$ of the colours to a unique colour: Colour by 'red' any edge coloured by c_1, and by 'blue' every other edge.

We claim that either G contains ηn^k edge-disjoint gadgets or one of the colour classes of G has size at most ρn^k. This follows by applying Lemma 4.1 iteratively. Suppose indeed that a maximal collection of edge-disjoint gadgets has size $\ell < \eta n^k$, and let G' be the subhypergraph of G obtained by removing all the edges of such gadgets. Let $\mathcal{S}$ be the collection of the $(k-1)$-subsets S of V with $d_{G'}(S) < (1/2 + \varepsilon/2)n$. We now bound the size of $\mathcal{S}$. The total number of removed edges is $2k\ell < 2k\eta n^k$. In order to have $S \in \mathcal{S}$, we must have removed at least $\varepsilon n/2$ edges containing S, and each removed edge decreases the degree of k $(k-1)$-sets by one. Therefore $|\mathcal{S}| \leq \frac{2k^2\eta n^k}{\varepsilon n/2} \leq \zeta n^{k-1}$, where we used $\eta \ll \zeta, \varepsilon$ for the last inequality. Since the collection was maximal, invoking Lemma 4.1, we get that one of the colour classes of G' has size at most $\rho n^k/2$. Therefore one of the colour classes of G has size at most $\rho n^k/2 + 2k \cdot \ell < \rho n^k$, where we used $\ell < \eta n^k$ and $\eta \ll \rho$.

If one of the colour classes of G, say $\mathcal{C}$, has size at most ρn^k, then, since $\mathbf{x}_0$ is μ_0-normal, this colour class gets a total weight of at most $\sum_{e\in\mathcal{C}} \mathbf{x}_0(e) \leq |\mathcal{C}|\mu_0^{-1} n^{-k+1} \leq \mu_0^{-1}\rho n$. Then the other colour class gets a total weight of at least $(1/k - \mu_0^{-1}\rho)n$. By averaging, $\mathbf{x}_0$ assigns to one of the colour classes of the original r-edge-colouring of G a total weight of at least $\frac{(1/k-\mu_0^{-1}\rho)n}{r-1} \geq (1+\mu)\frac{n}{rk}$, where we used $\rho \ll \mu_0, 1/r, 1/k$. In particular, $\mathbf{x}_0$ is already a desired μ-normal perfect fractional matching with high discrepancy.

We are left with the case where there exists a collection $\mathcal{L} := \{L_i : i \in [\eta n^k]\}$ of ηn^k edge-disjoint gadgets. Now, either the total weight assigned by $\mathbf{x}_0$ to the red edges is at least $\frac{n}{rk}$, or the total weight assigned by $\mathbf{x}_0$ to the blue edges is at least $\frac{(r-1)n}{rk}$.

Suppose we are in the first case. We will modify $\mathbf{x}_0$ on the edges of each gadget in $\mathcal{L}$ to obtain a μ-normal perfect fractional matching $\mathbf{x}$ with high discrepancy in colour c_1. For $L \in \mathcal{L}$, let $e_1^L, \dots, e_k^L$ (resp. $f_1^L, \dots, f_k^L$) be the horizontal (resp. vertical) edges of L, and recall these are pairwise distinct. By the definition of a near-alternating k-grid, all edges $e_1^L, \dots, e_k^L$ but at most one are red, and all edges $f_1^L, \dots, f_k^L$ but at most one are blue. Define $\mathbf{x}$: $E(G) \to [0,1]$ as follows: $\mathbf{x}(e_i^L) = \mathbf{x}_0(e_i^L) + \mu_0 n^{-k+1}/2$ and $\mathbf{x}(f_i^L) = \mathbf{x}_0(f_i^L) - \mu_0 n^{-k+1}/2$ for each $i \in [k]$ and $L \in \mathcal{L}$. Moreover, set $\mathbf{x}(e) = \mathbf{x}_0(e)$ for any other edge $e \in E(G)$. In other words, $\mathbf{x}$ is obtained from $\mathbf{x}_0$ by decreasing the weight of each vertical edge by $\mu_0 n^{-k+1}/2$ and increasing the weight of each horizontal edge by the same quantity, in each of the gadgets in $\mathcal{L}$. Observe that if a vertex is contained in a gadget, then it belongs to precisely one horizontal edge and one vertical edge of this gadget. Therefore for every vertex $v \in V(G)$ we have that $\sum_{e\ni v} \mathbf{x}(e) = \sum_{e\ni v} \mathbf{x}_0(e) = 1$. Also, for every $e \in E(G)$ we have $\mu n^{-k+1} \leq \mathbf{x}_0(e) - \mu_0 n^{-k+1}/2 \leq \mathbf{x}(e) \leq \mathbf{x}_0(e) + \mu_0 n^{-k+1}/2 \leq \mu^{-1} n^{-k+1}$, where we used that $\mu_0 n^{-k+1} \leq \mathbf{x}_0(e) \leq \mu_0^{-1} n^{-k+1}$ and $\mu \ll \mu_0$. This shows that $\mathbf{x}$ is a μ-normal perfect fractional matching. Moreover, for each $L \in \mathcal{L}$, the weight given by $\mathbf{x}$ to the red edges in L is bigger by at least $\mu_0 n^{-k+1}/2$ than the weight given by $\mathbf{x}_0$ to these edges. This is because we increased (by $\mu_0 n^{-k+1}/2$) the weight of at least $k-1 \geq 2$ red edges, and decreased (by the same amount) the weight of at most one such edge. As the gadgets in $\mathcal{L}$ are edge-disjoint and $|\mathcal{L}| = \eta n^k$, we conclude that the total weight assigned by $\mathbf{x}$ to the red edges is at least $\frac{n}{rk} + \eta n^k \mu_0 n^{-k+1}/2 \geq (1+\mu)\frac{n}{rk}$, using that $\mu \ll \eta, \mu_0$. Therefore, the total weight assigned by $\mathbf{x}$ to the colour class of c_1 in the original edge-colouring of G is at least $(1+\mu)\frac{n}{rk}$.

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

Suppose now that we are in the second case, namely, that the total weight assigned by $\mathbf{x}_0$ to the blue edges is at least $\frac{(r-1)n}{rk}$. We then use the same argument as above to find a μ-normal perfect fractional matching $\mathbf{x}$ which assigns a total weight of at least $(r-1)(1+\mu)\frac{n}{rk}$ to the blue edges. Recall that the blue edges are precisely those coloured by $c_2, \dots, c_r$ in the original colouring of G, and thus, by averaging, there is $2 \le i \le r$ such that the total weight assigned by $\mathbf{x}$ to the colour class of c_i in the original colouring of G is at least $(1+\mu)\frac{n}{rk}$. □

Ideally, we would like to use the random walk $\mathcal{Y}$ and Lemma 5.2 to obtain long tight paths with high discrepancy. However, one remaining problem is that, if we let the random walk continue for too many steps, then with high probability, it will not be self-avoiding anymore. It might be possible to analyse the 'self-avoiding' version of this random walk and show that with high probability, it will cover almost all the vertices and will have high discrepancy. However, such an analysis might be very intricate. To circumvent this, we prove the following 'sampling lemma' which allows us, using the random walk, to sample from the set of all tight paths in G of order t, where t is a (large) constant, in such a way that every vertex appears in the chosen path with approximately the same probability, and we expect more edges of one specific colour. In Section 6, we will then produce a large collection of paths sampled from this distribution, and then use a nibble-type argument to select from this collection a large linear forest with high discrepancy.

Lemma 5.3. *Let $1/n \ll 1/t, \mu \ll \varepsilon, 1/k, 1/r$ with $k, r, t \in \mathbb{N}$, $k \ge 3$ and $r \ge 2$. Let G be an n-vertex k-graph with $\delta(G) \ge (1/2+\varepsilon)n$ whose edges are r-coloured. Let Ω be the set of all tight paths of order t in G. Then, there exists a colour 'red' and a probability distribution on Ω such that a randomly chosen element $P \in \Omega$ has the following properties:*

(1) for any given tight path Q, we have $\mathbb{P}[P = Q] \le O_{t,\mu}(n^{-t})$;

(2) for every $v \in V(G)$, we have $\left|\mathbb{P}[v \in V(P)] - \frac{t}{n}\right| \le O_{t,\mu}\left(n^{-2}\right)$;

(3) the expected number of red edges in P is at least $\frac{1+\mu}{r} \cdot (t-k+1) - O_{t,\mu}(n^{-1})$.

Proof. Let $\mathbf{x}$ be a μ-normal perfect fractional matching of G such that the total weight received by some colour class, say 'red', is at least $(1+\mu)\frac{n}{rk}$. This exists by Lemma 5.2. Let $\mathcal{Y} = (Y_1, Y_2, \dots)$ be the random walk defined via $\mathbf{x}$, with notation as introduced at the beginning of this section. (In particular, we use $\mathbf{Pr}$ as the probability measure corresponding to the random walk, whereas $\mathbb{P}$ will denote the desired probability measure on Ω.) Here, we will only be interested in the first t vertices of $\mathcal{Y}$ and we note that those form a tight walk of order t. Now, sample elements of Ω according to the first t vertices of $\mathcal{Y}$, conditioned on $\mathcal{Y}$ being self-avoiding up to Y_t. More precisely, let $\mathcal{B}$ be the event that $Y_1, \dots, Y_t$ are pairwise distinct, and denote by Q any tight path of order t and by $q_1, \dots, q_t$ the vertices of Q (appearing in this order). Then take the distribution on Ω where a randomly chosen element $P \in \Omega$ satisfies

$$\mathbb{P}[P = Q] = \mathbf{Pr}\Big[\{Y_1 = q_1, \dots, Y_t = q_t\} \cup \{Y_1 = q_t, \dots, Y_t = q_1\} \mid \mathcal{B}\Big],$$

where we considered both orders of Q since the elements of Ω are unordered. We claim that this distribution on Ω has the desired properties. Before proving that this is the case, we need some preliminary observations.

Since $\mathbf{x}$ is μ-normal, the probability that the walk starts with $\overrightarrow{S} = (q_1, \dots, q_{k-1})$ is

$$\pi(\overrightarrow{S}) = \frac{\mathbf{x}(S)}{\sum_{\overrightarrow{S'} \in (V)_{k-1}} \mathbf{x}(S')} \le \frac{n \cdot \mu^{-1} n^{-k+1}}{(k-1)! \cdot n} = O_\mu(n^{-k+1}),$$

where we used (5.2). By using in addition that $\delta(G) \ge (1/2+\varepsilon)n$, we can show that the transition probabilities of $\mathcal{Y}$ are $O_\mu(n^{-1})$. Indeed, with $\overrightarrow{Z_i} = (Y_{i-(k-2)}, \dots, Y_i)$ being the ordered set of the

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

last $k-1$ vertices of $\mathcal{Y}$, for $v \in V(G) \setminus Z_i$ we have

$$\mathbf{Pr}[Y_{i+1} = v | Y_{i-(k-2)}, \dots, Y_i] = \frac{\mathbf{x}(Z_i \cup \{v\})}{\mathbf{x}(Z_i)} \leq \frac{\mu^{-1} n^{-k+1}}{1/2 \cdot n \cdot \mu n^{-k+1}} = O_\mu(n^{-1}),$$

while for $v \in Z_i$ we have $\mathbf{Pr}[Y_{i+1} = v | Y_{i-(k-2)}, \dots, Y_i] = 0$. Therefore, by applying the chain rule, $\mathbf{Pr}[Y_1 = q_1, \dots, Y_t = q_t] = O_{t,\mu}(n^{-t})$. Moreover, the number of walks of order t which are not self-avoiding is $O_t(n^{t-1})$ and thus $\mathbf{Pr}[\mathcal{B}^c] = O_t(n^{t-1}) \cdot O_{t,\mu}(n^{-t}) = O_{t,\mu}(n^{-1})$.

Now Item (1) follows easily as we have

$$\mathbb{P}[P = Q] \leq \frac{\mathbf{Pr}[Y_1 = q_1, \dots, Y_t = q_t] + \mathbf{Pr}[Y_1 = q_t, \dots, Y_t = q_1]}{\mathbf{Pr}[\mathcal{B}]} = O_{t,\mu}(n^{-t}),$$

where we used that $\mathbf{Pr}[\mathcal{B}] = 1 - O_{t,\mu}(n^{-1}) \geq 1/2$.

Fix $v \in V(G)$ and $1 \leq i \leq t$. The number of walks of order t which are not self-avoiding and whose i-th vertex is v is $O_t(n^{t-2})$ and thus $\mathbf{Pr}[\{Y_i = v\} \cap \mathcal{B}^c] = O_t(n^{t-2}) \cdot O_{t,\mu}(n^{-t}) = O_{t,\mu}(n^{-2})$. Therefore,

$$\mathbb{P}[v \in V(P)] = \mathbf{Pr}\left[\bigcup_{i \in [t]} \{Y_i = v\} \mid \mathcal{B}\right] = \sum_{i \in [t]} \frac{\mathbf{Pr}\big[\{Y_i = v\} \cap \mathcal{B}\big]}{\mathbf{Pr}[\mathcal{B}]} \geq t n^{-1} - O_{t,\mu}(n^{-2}),$$

where we used that the events $\{Y_i = v\} \cap \mathcal{B}$ with $i \in [t]$ are pairwise disjoint, that $\mathbf{Pr}[\{Y_i = v\}] = n^{-1}$ by Fact 5.1, that

$$\mathbf{Pr}\big[\{Y_i = v\} \cap \mathcal{B}\big] = \mathbf{Pr}\big[Y_i = v\big] - \mathbf{Pr}\big[\{Y_i = v\} \cap \mathcal{B}^c\big] = n^{-1} - O_{t,\mu}(n^{-2}),$$

and that $\mathbf{Pr}[\mathcal{B}] \leq 1$. Similarly, by using $\mathbf{Pr}[\mathcal{B}] = 1 - O_{t,\mu}(n^{-1})$, we get that $\mathbb{P}[v \in V(P)] \leq t n^{-1} + O_{t,\mu}(n^{-2})$. This proves Item (2).

For a given $k \leq i \leq t$, denote by $e_i := \{Y_{i-k+1}, \dots, Y_i\}$ the $(i-k+1)$-st edge of $\mathcal{Y}$. Let $e \in E(G)$. Similarly as above, $\mathbf{Pr}[\{e_i = e\} \cap \mathcal{B}^c] = O_{t,\mu}(n^{-k-1})$ because there are $O_t(n^{t-k-1})$ tight walks which are not self-avoiding and satisfy $e_i = e$, and each such walk has probability $O_{t,\mu}(n^{-t})$. Thus,

$$\mathbf{Pr}\Big[e_i = e \mid \mathcal{B}\Big] = \frac{\mathbf{Pr}\big[e_i = e\big] - \mathbf{Pr}\big[\{e_i = e\} \cap \mathcal{B}^c\big]}{\mathbf{Pr}[\mathcal{B}]} \geq \frac{k}{n} \cdot \mathbf{x}(e) - O_{t,\mu}(n^{-k-1}),$$

where we used Fact 5.1 and $\mathbf{Pr}[\mathcal{B}] \leq 1$. Let $\mathcal{R}$ be the set of edges of G which are coloured red. Then

$$\mathbf{Pr}\Big[e_i \in \mathcal{R} \mid \mathcal{B}\Big] \geq \sum_{e \in \mathcal{R}} \left(\frac{k}{n} \cdot \mathbf{x}(e) - O_{t,\mu}(n^{-k-1})\right) \geq \frac{1+\mu}{r} - O_{t,\mu}(n^{-1}),$$

where we used that $|\mathcal{R}| \leq n^k$ and $\sum_{e \in \mathcal{R}} \mathbf{x}(e) \geq (1+\mu)\frac{n}{rk}$. Since P is a path of order t, it has $t-k+1$ edges, and thus Item (3) follows by linearity of expectation. □

6. Finding a linear forest with high discrepancy

The goal of this section is to prove the following result.

Lemma 6.1. *Let $1/n \ll 1/t \ll \beta \ll \mu \ll \varepsilon, 1/k, 1/r$ with $k, r, t \in \mathbb{N}$, $k \geq 3$, $r \geq 2$. Let G be an n-vertex k-graph with $\delta(G) \geq (1/2 + \varepsilon)n$ whose edges are r-coloured. Then G contains a collection of vertex-disjoint tight paths of order t such that their union covers all but at most βn vertices of G and there is some colour which appears on at least $(1+\mu)\frac{n}{r}$ edges in the paths.*

Proof. Let

$$1/n \ll \delta \ll 1/t \ll \beta \ll \mu \ll \varepsilon, 1/k, 1/r,$$

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

where 2μ is given by Lemma 5.3 on input ε, r, k, and set $\gamma := \delta/(50t^2)$ and $V := V(G)$.

Set $N := n^{t-1/2}$ and let $\mathcal{P} := \{P_i : i \in [N]\}$ be a collection of N tight paths of order t independently sampled according to the distribution given by Lemma 5.3. For a given $v \in V$, define $X_v := \{i \in [N] : v \in V(P_i)\}$. Using Item (2) in Lemma 5.3, we have $\mathbb{E}[|X_v|] = N \cdot tn^{-1} \cdot \left[1 \pm O_{t,\mu}(n^{-1})\right] = tn^{t-3/2} \cdot \left[1 \pm O_{t,\mu}(n^{-1})\right]$. Therefore, by Chernoff's inequality (Lemma 3.4) and a union bound over $v \in V$, w.h.p. we have

$$|X_v| = (1 \pm \beta/3)tn^{t-3/2} \tag{6.1}$$

for every $v \in V$. For a path $P \in \mathcal{P}$, let $\mathrm{red}(P)$ be the number of red edges of P, and note that $0 \le \mathrm{red}(P) \le t - k + 1$. Let $R := \sum_{i=1}^{N} \mathrm{red}(P_i)$. By Item (3) of Lemma 5.3, $\mathbb{E}[R] \ge N \cdot \left[(t-k+1) \cdot \frac{1+2\mu}{r} - O_{t,\mu}(n^{-1})\right]$ and by Chernoff's inequality (Lemma 3.4), we have w.h.p. that

$$R \ge (1-\beta) \cdot (t-k+1) \cdot \frac{1+2\mu}{r} \cdot n^{t-1/2}. \tag{6.2}$$

We now show that we can pass to a large subcollection $\mathcal{P}' \subseteq \mathcal{P}$ such that no two paths in $\mathcal{P}'$ have the same vertex set. Let Y be the set of pairs $1 \le i < j \le N$ such that $V(P_i) = V(P_j)$. We now bound $|Y|$. Given any t-subset S of V, observe that there are $t!/2$ (unordered) paths P with $V(P) = S$ and that, using Item (1) in Lemma 5.3, the probability that $P_i = P$ for a given $1 \le i \le N$ is $O_{t,\mu}(n^{-t})$. Therefore, the probability that a given pair i, j belongs to Y is $O_{t,\mu}(n^{-t})$, and thus $\mathbb{E}[|Y|] \le N^2 \cdot O_{t,\mu}(n^{-t}) = O_{t,\mu}(n^{t-1})$. It follows from Markov's inequality that w.h.p. $|Y| \le n^{t-2/3}$.

We now fix such a collection of paths that satisfies (6.1), (6.2) and $|Y| \le n^{t-2/3}$. Let $\mathcal{P}' \subseteq \mathcal{P}$ be obtained from $\mathcal{P}$ by deleting P_i, P_j for every pair $\{i, j\} \in Y$. Then $|\mathcal{P}'| \ge |\mathcal{P}| - 2|Y| \ge (1 - \beta/3)n^{t-1/2}$. Let $\mathcal{H}$ be the auxiliary t-graph on V with edge set $\{V(P) : P \in \mathcal{P}'\}$. By the definition of $\mathcal{P}'$, we have $V(P) \ne V(Q)$ for each $P, Q \in \mathcal{P}'$, and thus $e(\mathcal{H}) = |\mathcal{P}'|$ (i.e., $\mathcal{H}$ has no multiple edges). We now show that $\mathcal{H}$ is suitable for an application of Theorem 3.5, by establishing bounds on $\Delta(\mathcal{H})$ and $\Delta^c(\mathcal{H})$. Using (6.1), we have $d_{\mathcal{H}}(\{v\}) \le |X_v| \le (1 + \beta/3)tn^{t-3/2}$. Setting $\Delta := (1 + \beta/3)tn^{t-3/2}$, we have $\Delta(\mathcal{H}) \le \Delta$, and it trivially holds that $\Delta^c(\mathcal{H}) \le n^{t-2} \le \Delta^{1-\delta}$, as $\delta \ll 1/t$.

We would like the matching given by Theorem 3.5 to be almost-spanning and with large discrepancy. To this end, we define two weight functions $w_1, w_2 \colon E(\mathcal{H}) \to \mathbb{R}_{\ge 0}$ as follows: $w_1 \equiv 1$; and for an edge $e \in E(\mathcal{H})$ corresponding to a path $P \in \mathcal{P}'$, we define $w_2(e) := \mathrm{red}(P)$. We claim that $w_i(E(\mathcal{H})) \ge \max_{e \in E(\mathcal{H})} w(e)\Delta^{1+\delta}$ for each $i = 1, 2$. For $i = 1$, this is obvious as

$$w_1(E(\mathcal{H})) = e(\mathcal{H}) \ge (1 - \beta/3)n^{t-1/2} \ge \Delta^{1+\delta} = \max_{e \in E(\mathcal{H})} w_1(e)\Delta^{1+\delta},$$

where the last inequality uses that $\delta \ll 1/t$. And for $i = 2$, we have

$$\begin{aligned} w_2(E(\mathcal{H})) &= \sum_{P \in \mathcal{P}'} \mathrm{red}(P) = R - \sum_{P \in \mathcal{P} \setminus \mathcal{P}'} \mathrm{red}(P) \ge R - (t-k+1) \cdot 2n^{t-2/3} \\ &\ge (1-2\beta) \cdot (t-k+1) \cdot \frac{1+2\mu}{r} \cdot n^{t-1/2}, \end{aligned}$$

where the first inequality uses $|\mathcal{P}| - |\mathcal{P}'| \le 2|Y| \le 2n^{t-2/3}$ and that each path has at most $t - k + 1$ red edges, while the second inequality uses (6.2). Therefore

$$w_2(E(\mathcal{H})) = \Omega(n^{t-1/2}) \ge (t-k+1)\Delta^{1+\delta} \ge \max_{e \in E(\mathcal{H})} w_2(e)\Delta^{1+\delta},$$

using that $\Delta = O_t(n^{t-3/2})$ and $\delta \ll 1/t$.

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

Let $\mathcal{M}$ be the matching in $\mathcal{H}$ given by applying Theorem 3.5 with $\mathcal{W} := \{w_1, w_2\}$. Using $w_1(E(\mathcal{H})) = e(\mathcal{H}) = |\mathcal{P}'| \geq (1-\beta/3)n^{t-1/2}$ and the guarantees of Theorem 3.5, we have

$$|\mathcal{M}| = w_1(\mathcal{M}) \geq \left(1-\Delta^{-\gamma}\right) \cdot \frac{w_1(E(\mathcal{H}))}{\Delta} \geq \left(1-\Delta^{-\gamma}\right) \cdot \frac{(1-\beta/3)n^{t-1/2}}{(1+\beta/3)tn^{t-3/2}} \geq (1-\beta)\frac{n}{t}\,.$$

Similarly, for w_2 we have

$$\begin{aligned} w_2(\mathcal{M}) \geq \left(1-\Delta^{-\gamma}\right) \cdot \frac{w_2(E(\mathcal{H}))}{\Delta} &\geq (1-o(1)) \cdot \frac{(1-2\beta)\cdot(t-k+1)\cdot\frac{1+2\mu}{r}\cdot n^{t-1/2}}{(1+\beta/3)tn^{t-3/2}} \\ &= (1-o(1)) \cdot \frac{1-2\beta}{1+\beta/3} \cdot \frac{t-k+1}{t} \cdot \frac{1+2\mu}{r} \cdot n \\ &\geq (1-3\beta) \cdot \frac{1+2\mu}{r} \cdot n \geq (1+\mu)\frac{n}{r}\,, \end{aligned}$$

where we used the bound on $w_2(E(\mathcal{H}))$ established above, the definition of Δ, together with $1/t \ll \beta \ll \mu \ll 1/k$.

Therefore the matching $\mathcal{M}$ corresponds to a collection of vertex-disjoint paths of G of order t such that their union covers all but at most βn vertices and the colour red appears on at least $(1+\mu)\frac{n}{r}$ edges in the paths, as desired. □

7. Proof of the main theorem

We are now ready to prove our main result.

Proof of Theorem 1.1. Let

$$1/n \ll 1/t \ll \beta \ll \mu \ll \varepsilon \ll 1/k, 1/r\,,$$

where we have assumed without loss of generality that ε is sufficiently small. Let G be an n-vertex k-graph with $\delta(G) \geq (1/2+\varepsilon)n$ whose edges are r-coloured. By Lemma 3.2, there exists a tight path P_0 such that $|V(P_0)| \leq \mu n$ and for each $W \subseteq V(G) \setminus V(P_0)$ of size $|W| \leq 3\beta n$ there is a tight path covering $V(P_0) \cup W$ and with the same ends as P_0. Let R be a random subset of $V(G) \setminus V(P_0)$ obtained by including each vertex with probability β independently. Then w.h.p. it holds that $\beta n/2 \leq |R| \leq 2\beta n$ and that $|N_G(S) \cap R| \geq (1/2+\varepsilon/2)|R|$ for every $(k-1)$-subset $S \subseteq V(G)$. Indeed, this can be shown by a standard application of Chernoff's inequality (Lemma 3.4) and a union bound over S.

Let $G' := G \setminus (V(P_0) \cup R)$ and observe that $\delta(G') \geq \delta(G) - |V(P_0) \cup R| \geq (1/2+\varepsilon/2)n \geq (1/2+\varepsilon/2)|V(G')|$. Then, by Lemma 6.1 (with 4μ playing the role of μ), G' contains a collection $\mathcal{P} := \{P_i : i \in [N]\}$ of vertex-disjoint tight paths of order t such that their union covers all but at most βn vertices, and there is a colour, say red, which appears on at least $(1+4\mu)|V(G')|/r \geq (1+4\mu)(1-\mu-2\beta)n/r \geq (1+\mu)n/r$ edges in the paths. Therefore, if we manage to connect the paths in $\mathcal{P}$ into a tight cycle while covering all the vertices then we are done. This connection can be achieved using the standard tools collected in Section 3.1. The details follow.

Denote by $\overrightarrow{S_i}$ and $\overrightarrow{T_i}$ the ends of P_i for each $0 \leq i \leq N$. Add the uncovered vertices of G' to R to get a set R', and observe that $|R| \leq |R'| \leq |R| + \beta n$. Using multiple applications of the connecting lemma (Lemma 3.1), we can connect the path P_0 and the paths in $\mathcal{P}$ into an almost-spanning tight cycle using vertices in R', as proved by the following claim. □

Claim 5. *For each $0 \leq i \leq N$, there is a tight path Q_i of order at most $2k/\varepsilon^2$ connecting $\overleftarrow{T_i}$ and $\overleftarrow{S_{i+1}}$ (where indices are modulo $N+1$), such that $V(Q_i) \setminus (T_i \cup S_{i+1}) \subseteq R'$. Moreover, we can choose such paths to be pairwise vertex-disjoint.*

Proof. Suppose we can find vertex-disjoint connecting paths $Q_0, \ldots, Q_{m-1}$ as in the statement of the claim, and let m be as large as possible. If we are not done yet, then $m \leq N$. The union of

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

$Q_0, \dots, Q_{m-1}$ covers at most $m \cdot 2k/\varepsilon^2 \le 2kn/(\varepsilon^2 t)$ vertices of R', where we used that $N \le n/t$. Let R'' denote the subset of R' consisting of the uncovered vertices. Then for every $(k-1)$-subset $S \subseteq V(G)$ we have $|N_G(S) \cap R''| \ge (1/2+\varepsilon/2)|R| - |R' \setminus R''| \ge (1/2+\varepsilon/4)|R'' \cup T_{m+1} \cup S_{m+2}|$, where we used that $|R| \ge |R'| - \beta n$, $|R' \setminus R''| \le 2kn/(\varepsilon^2 t)$ and $1/t \ll \beta \ll \varepsilon, 1/k$. Therefore, we can apply Lemma 3.1 to $G[R'' \cup T_{m+1} \cup S_{m+2}]$ to get a tight path of order at most $2k/\varepsilon^2$ connecting $\overleftarrow{T_{m+1}}$ and $\overleftarrow{S_{m+2}}$, which is vertex-disjoint from $Q_0, \dots, Q_{m-1}$. This contradicts the maximality of m. □

Observe that $C := \bigcup_{0 \le i \le N} (P_i \cup Q_i)$ is a tight cycle. Now let $W \subseteq R'$ be the subset of vertices not covered by C and observe that clearly $|W| \le |R'| \le 3\beta n$. By the property of the absorbing path P_0, there exists a tight path $\tilde{P}_0$ which covers $V(P_0) \cup W$ and has the same ends as P_0, i.e. $\overrightarrow{S_0}$ and $\overrightarrow{T_0}$. It follows that $\tilde{C} := \tilde{P}_0 \cup \bigcup_{1 \le i \le N} (P_i \cup Q_i)$ is a tight Hamilton cycle of G and, since each edge of $P_1 \cup \dots \cup P_N$ is an edge of $\tilde{C}$ as well, then $\tilde{C}$ has at least $(1+\mu)n/r$ red edges, as desired.

8. Concluding remarks

The main result of this paper offers a discrepancy version of the celebrated result of Rödl, Ruciński, and Szemerédi [29], and determines the minimum $(k-1)$-degree threshold for high discrepancy of tight Hamilton cycles and perfect matchings; both of these thresholds equal $1/2$, which is also the existence threshold for these structures. In the following we discuss some natural open problems for further research.

- A very natural question is to study minimum-degree discrepancy thresholds for the j-degree[1] with $j < k-1$. We remark that for this question, the existence threshold for perfect matchings (i.e. the minimum j-degree guaranteeing the existence of a perfect matching) is mostly not known. We wonder if, for k-uniform hypergraphs with $k \ge 3$, there is some j such that the j-degree discrepancy threshold is strictly larger than the corresponding existence threshold (as is the case for graphs).
  Remark added. This problem has been solved for each $j \ne 1$ in a simultaneous work of Balogh, Treglown, and Zárate-Guerén [5], who also provided a construction showing that, for $j=1$ and $k=3$, the discrepancy threshold is significantly larger than the existence threshold. The case $j=1$ was then fully resolved by Lu, Ma and Xie [26], and, independently, by Hàn, Lang, Marciano, Pavez-Signé, Sanhueza-Matamala, Treglown, and Zárate-Guerén [20].
- Another natural question is to consider other notions of Hamilton cycles. As mentioned in the introduction, Theorem 1.1 implies that minimum $(k-1)$-degree $(1/2+\varepsilon)n$ guarantees the existence of Hamilton ℓ-cycles of high discrepancy, for every $1 \le \ell \le k-1$. However, unlike in the case of tight Hamilton cycles (namely, $\ell = k-1$), we do not have a matching lower bound. For example, it is known that the minimum $(k-1)$-degree threshold for the existence of loose Hamilton cycles in k-graphs is $\frac{1}{2(k-1)}$, see [22, 23]. We wonder if this is also the discrepancy threshold of loose Hamilton cycles.
- As mentioned in the introduction, $1/2$ is the minimum $(k-1)$-degree threshold for the existence of perfect matchings in k-graphs (see [24, 30]), which shows that Corollary 1.2 is tight. However, it is also known [24, 30] that a k-graph with minimum $(k-1)$-degree at least $(1+o(1))\frac{n}{k}$ contains a *near-perfect matching*, i.e. a matching of size $\frac{n}{k} - O_k(1)$. For $k=3$ and 2 colours, we have the following simple example showing that $1/2$ is the discrepancy threshold of near-perfect matchings (which is larger than the existence threshold

[1] The *minimum j-degree* of a hypergraph is the minimum of $d(S)$ over all sets S of j vertices.

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

of 1/3). Partition the vertices into two sets A and B of equal size and take G to be the hypergraph consisting of all edges which intersect both A and B. Colour in red the edges which intersect A in two vertices, and colour the remaining edges in blue. Every matching of size $n/3 - t$ must have at least $n/6 - 2t$ edges in each colour, meaning that there is no near-perfect matching with discrepancy $\Omega(n)$. It is therefore natural to ask, for general $k \geq 3$ and $r \geq 2$, what is the minimum $(k-1)$-degree threshold guaranteeing a near-perfect matching of high discrepancy in every r-edge-colouring.

- It would also be interesting to prove similar results for other spanning structures in hypergraphs, perhaps even of design-type, such as Steiner triple systems. We will return to this in a future work.
- Instead of studying minimum degree thresholds, one might also consider random hypergraphs. In the graph case, Gishboliner, Krivelevich, and Michaeli [16] showed that with high probability, the random graph $G(n,p)$ has the following property: in every r-edge-colouring, there exists a Hamilton cycle which has at least roughly $\frac{2n}{r+1}$ edges of the same colour and hence a perfect matching with at least roughly $\frac{n}{r+1}$ edges of the same colour. The respective constants $\frac{2}{r+1}$ and $\frac{1}{r+1}$ are best possible even in the complete graph. This raises the question of whether the same phenomenon holds in hypergraphs. For example, as mentioned in the introduction, the result of Alon–Frankl–Lovász [1] implies that every 2-edge-colouring of $K_n^{(k)}$ has a perfect matching with at least roughly $\frac{n}{k+1}$ edges of the same colour. Is the same true in a random k-graph (say, with edge probability above the existence threshold $(\log n)/n^{k-1}$)?
- Generalising the previous item, it would be very interesting to prove a general result relating the threshold for containing a structure (in a random graph/hypergraph) to the threshold for having high discrepancy for this structure. Namely, for a family $\mathcal{F}$ of graphs (or hypergraphs) on $[n]$, let p_0 be the threshold for the event that $G \sim G(n,p)$ contains a member from $\mathcal{F}$, and let p_1 be the threshold for the event that in every 2-edge-colouring of $G \sim G(n,p)$, there is a member $F \in \mathcal{F}$ with high discrepancy, say of order $\Theta(e(F))$. Note that p_1 is well-defined if (and only if) $\mathcal{F}$ has high discrepancy in K_n. Clearly $p_1 \geq p_0$. Is there a general upper bound on p_1 in terms of p_0? We wonder if the recent breakthroughs around the expectation-threshold conjecture are relevant to this question.
- Even more generally, one can ask about the discrepancy of random subhypergraphs of general hypergraphs (not necessarily those arising from graphs). Namely, we return to the original definition of discrepancy, where $\mathcal{H}$ is a hypergraph, and we colour the vertices of $\mathcal{H}$ with two colours. How 'robust' is discrepancy? For instance, suppose $\mathcal{H}$ has high discrepancy, and we take a random subset of vertices by including each vertex independently with probability p. Is the random induced subhypergraph likely to still have high discrepancy?

Acknowledgements

We thank the anonymous referees for their valuable comments. This research was initiated while AS was visiting the University of Passau; he would like to thank the University of Passau for the hospitality and the stimulating research environment.

When conducting this work, LG was supported by SNSF grant 200021_196965.

This research was conducted while AS was affiliated with University College London and supported by the Royal Society.

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press

References

[1] Alon, N., Frankl, P. and Lovász, L. (1986) The chromatic number of Kneser hypergraphs. *Trans. Amer. Math. Soc* **298** 359–370.
[2] Alon, N. and Spencer, J. H. (2016) The probabilistic method, 4th ed. *Wiley-Intersci. Ser. Discrete Math Optim.*, John Wiley & Sons.
[3] Balogh, J., Csaba, B., Jing, Y. and Pluhár, A. (2020) On the discrepancies of graphs. *Electron. J. Combin.* **27**(2.12) 14. Paper
[4] Balogh, J., Csaba, B., Pluhár, A. and Treglown, A. (2021) A discrepancy version of the Hajnal–Szemerédi theorem. *Combin. Probab. Comput.* **30** 444–459.
[5] Balogh, J., Treglown, A. and Zárate-Guerén, C. (2024) A note on color-bias perfect matchings in hypergraphs. *SIAM J. Discrete Math.* **38** 2543–2552.
[6] Bradač, D. (2022) Powers of Hamilton cycles of high discrepancy are unavoidable. *Electron. J. Combin.* **29** P3.22.
[7] Bradač, D., Christoph, M. and Gishboliner, L. (2024) Minimum degree threshold for H-factors with high discrepancy. *Electron. J. Combin.* **31**(3.33) 82. Paper.
[8] Dubhashi, D. P. and Panconesi, A. (2009) *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, Cambridge.
[9] Ehard, S., Glock, S. and Joos, F. (2020) Pseudorandom hypergraph matchings. *Combin. Probab. Comput* **29** 868–885.
[10] Erdős, P. (1964) On extremal problems of graphs and generalized graphs. *Israel J. Math.* **2** 183–190.
[11] Erdős, P., Füredi, Z., Loebl, M. and Sós, V. T. (1995) Discrepancy of trees. *Studia Sci. Math. Hungar.* **30** 47–57.
[12] Erdős, P. and Spencer, J. H. (1971) Imbalances in k-colorations. *Networks* **72** 379–385.
[13] Frankl, P. and Rödl, V. (1985) Near perfect coverings in graphs and hypergraphs. *European J. Combin.* **6** 317–326.
[14] Freschi, A., Hyde, J., Lada, J. and Treglown, A. (2021) A note on color-bias Hamilton cycles in dense graphs. *SIAM J. Discrete Math.* **35** 970–975.
[15] Freschi, A. and Lo, A. (2024) An oriented discrepancy version of Dirac's theorem. *J. Combin. Theory Ser. B.* **169** 338–351.
[16] Gishboliner, L., Krivelevich, M. and Michaeli, P. (2022) Color-biased Hamilton cycles in random graphs. *Random Struct. Algor.* **60** 289–307.
[17] Gishboliner, L., Krivelevich, M. and Michaeli, P. (2022) Discrepancies of spanning trees and Hamilton cycles. *J. Combin. Theory Ser. B.* **154** 262–291.
[18] Gishboliner, L., Krivelevich, M. and Michaeli, P. (2023) Oriented discrepancy of Hamilton cycles. *J. Graph Theory* **103** 780–792.
[19] Glock, S., Gould, S., Joos, F., Kühn, D. and Osthus, D. (2021) Counting Hamilton cycles in Dirac hypergraphs. *Combin. Probab. Comput.* **30** 631–653.
[20] Hàn, H., Lang, R., Marciano, J. P., Pavez-Signé, M., Sanhueza-Matamala, N., Treglown, A. and Zárate-Guerén, C. Colour-bias perfect matchings in hypergraphs, arXiv: https://arxiv.org/abs/2408.11016, 2024.
[21] Katona, G. Y. and Kierstead, H. A. (1999) Hamiltonian chains in hypergraphs. *J. Graph Theory* **30** 205–212.
[22] Keevash, P., Kühn, D., Mycroft, R. and Osthus, D. (2011) Loose Hamilton cycles in hypergraphs. *Discrete Math.* **311** 544–559.
[23] Kühn, D. and Osthus, D. (2006) Loose Hamilton cycles in 3-uniform hypergraphs of high minimum degree. *J. Combin. Theory Ser. B.* **96** 767–821.
[24] Kühn, D. and Osthus, D. (2006) Matchings in hypergraphs of large minimum degree. *J. Graph Theory* **51** 269–280.
[25] Lovász, L. (1978) Kneser's conjecture, chromatic number, and homotopy. *J. Combin. Theory Ser. A.* **25** 319–324.
[26] Lu, H., Ma, J. and Xie, S. Discrepancies of perfect matchings in hypergraphs, arXiv: https://arxiv.org/abs/2408.06020, 2024.
[27] Brito, C. J. M. (2023) Discrepancia de ciclos hamiltonianos en hipergrafos 3-uniformes, Master's thesis, Universidad de Concepción.
[28] Pippenger, N. and Spencer, J. (1989) Asymptotic behaviour of the chromatic index for hypergraphs. *J. Combin. Theory Ser. A.* **51** 24–42.
[29] Rödl, V., Ruciński, A. and Szemerédi, E. (2008) An approximate Dirac-type theorem for k-uniform hypergraphs. *Combinatorica* **28** 229–260.
[30] Rödl, V., Ruciński, A. and Szemerédi, E. (2009) Perfect matchings in large uniform hypergraphs with large minimum collective degree. *J. Combin. Theory Ser. A.* **116** 613–636.

Cite this article: Gishboliner L, Glock S, and Sgueglia A (2025). Tight Hamilton cycles with high discrepancy. *Combinatorics, Probability and Computing* **34**, 565–584. https://doi.org/10.1017/S0963548325000057

https://doi.org/10.1017/S0963548325000057 Published online by Cambridge University Press