

Unbalanced spanning subgraphs in edge labeled complete graphs*

Stéphane Bessy

LIRMM
Univ Montpellier, CNRS
Montpellier, France

`stephane.bessy@lirmm.fr`

Johannes Pardey

Institute of Optimization and OR
Ulm University
Ulm, Germany

`johannes.pardey@uni-ulm.de`

Lucas Picasarri-Arrieta

Université Côte d'Azur
Inria, CNRS, I3S
Sophia-Antipolis, France

`lucas.picasarri-arrieta@inria.fr`

Dieter Rautenbach

Institute of Optimization and OR
Ulm University
Ulm, Germany

`dieter.rautenbach@uni-ulm.de`

Submitted: Nov 11, 2021; Accepted: Jan 24, 2023; Published: Feb 24, 2023

© The authors. Released under the CC BY-ND license (International 4.0).

Abstract

Let K be a complete graph of order n . For $d \in (0, 1)$, let c be a ± 1 -edge labeling of K such that there are $d \binom{n}{2}$ edges with label $+1$, and let G be a spanning subgraph of K of maximum degree at most Δ and with $m(G)$ edges. We prove the existence of an isomorphic copy G' of G in K such that the number of edges with label $+1$ in G' is at least $\left(d + \frac{\min\{2-d-2\sqrt{1-d}, \sqrt{d}-d\}}{2\Delta+1} - O\left(\frac{1}{n}\right)\right) m(G)$, that is, this number visibly exceeds its expected value $d \cdot m(G)$ when considering a uniformly random copy of G in K . For $d = \frac{1}{2}$, and $\Delta \leq 2$, we present more detailed results.

Mathematics Subject Classifications: 05C22

1 Introduction

Let K be a complete graph with vertex set $[n] = \{1, \dots, n\}$, and let $c : E(K) \rightarrow \{\pm 1\}$ be a ± 1 -edge labeling of K . The edge-labeling c of K is *balanced* if there are equally many *plus-edges* and *minus-edges*, that is, edges with label $+1$ and -1 , respectively. For a spanning subgraph G of K , let $c(G) = c(E(G)) = \sum_{e \in E(G)} c(e)$, and let $m(G)$, $m^+(G)$,

*Research supported by research grant DIGRAPHS ANR-19-CE48-0013.

and $m^-(G)$ denote the number of edges, the number of plus-edges, and the number of minus-edges of G , respectively. Note that $c(G) = m^+(G) - m^-(G)$. For a permutation π from S_n , let G_π be the isomorphic copy of G in K with edge set $\{\pi(u)\pi(v) : uv \in E(G)\}$.

In the present paper, we study the structure of the set

$$\sigma_{(K,c)}(G) = \{m^+(G_\pi) : \pi \in S_n\}.$$

Our research is inspired by recent beautiful work of Caro, Hansberg, and Montejano [6] on so-called *omnitonal* graphs. Roughly speaking, a graph G is said to be omnitonal if for every pair (K, c) , where the order n of K is sufficiently large and there are sufficiently many plus-edges and minus-edges in K , and for every two non-negative integers m^+ and m^- with $m(G) = m^+ + m^-$, there is an isomorphic copy G' of G in K with $m^+(G') = m^+$ and $m^-(G') = m^-$. The key difference to the problems we study here is that the order of K is necessarily much bigger than the order of G , that is, the graph G is far from being a spanning subgraph of K . Quite surprisingly, exploiting recent strong results from Ramsey theory [8, 13], Caro et al. [6] achieve a very concise characterization of omnitonal graphs. While being inspired by their work, requiring that G is a spanning subgraph of K drastically changes the nature of the problem. The higher the density of a spanning graph G is, the more every isomorphic copy of G in K is forced to reproduce the density of plus- and minus-edges in (K, c) . Therefore, as a natural hypothesis excluding dense spanning subgraphs, we consider graphs of bounded maximum degree.

Another perspective on our results is that they correspond to relaxed versions of classical extremal problems, which ask how many edges suffice to ensure the existence of a specific subgraph. In order to force a Hamiltonian cycle in a graph G of order n , for instance, one needs to require at least $\binom{n-1}{2} + 2$ edges in G , that is, the graph has to be almost complete with a density $m(G)/\binom{n}{2}$ tending to 1. Our Theorem 2(i) below can be rephrased to say that while a density of $1/2$ does not force the existence of a Hamiltonian cycle, it forces the existence of 58% of it, more precisely, of a Hamiltonian cycle in the complete graph on the same vertex set in which 58% of the edges belong to the original graph, which is best possible. Also our main result, Theorem 1, can be rephrased in such a way.

A third motivation for our results is their relation to *graph discrepancy* notions originating in work of Erdős et al. [11] and recently considered in [2, 3, 12]. In these works, the authors mainly focus on the minimum degree threshold ensuring high discrepancy; considering a much simpler setting, we obtain better estimates, and illustrate the relation to our results below in Corollary 3.

If $d \in [0, 1]$ is such that $m^+(K) = d\binom{n}{2}$, that is, d is the density of the plus-edges in (K, c) , then, since, by symmetry, every edge of K belongs to the same number of subgraphs G_π , we obtain

$$\frac{1}{n!} \sum_{\pi \in S_n} m^+(G_\pi) = d \cdot m(G). \quad (1)$$

Furthermore, transposition arguments as in [7, 16] imply that if $\sigma_{(K,c)}(G) = \{m_1^+, \dots, m_k^+\}$

for $m_1^+ < \dots < m_k^+$, then

$$m_{i+1}^+ - m_i^+ \leq \Delta(G) + \delta(G) \leq 2\Delta(G) \text{ for every } i \in [k-1], \quad (2)$$

that is, if the maximum degree $\Delta(G)$ of G is small, then there are no big gaps in $\sigma_{(K,c)}(G)$. This motivates to consider $\max \sigma_{(K,c)}(G)$ and $\min \sigma_{(K,c)}(G)$.

In the case that c is balanced, that is, $d = \frac{1}{2}$, the existence of copies G_π for which $|c(G_\pi)|$ is small, or equivalently, $m^+(G_\pi)$ is close to $\frac{m(G)}{2}$ has been studied under the term *zero sum problems* or *zero sum Ramsey theory*; see [4, 5, 6, 7, 9, 14, 15, 16, 17] as well as the references therein pointing out connections to further research directions. The observations (1) and (2) are based on common arguments from this area, and together they imply the existence of some permutation π from S_n with

$$|m^+(G_\pi) - d \cdot m(G)| \leq \Delta(G),$$

that is, the averaging arguments (1) and transformation arguments (2) imply the existence of some permutation π from S_n for which $m^+(G_\pi)$ is close to its expected value $d \cdot m(G)$, when choosing π uniformly at random from S_n .

Our first result in this paper is that, for maximum degree $\Delta(G)$ at most some constant Δ , and $d \in (0, 1)$, the value $\max \sigma_{(K,c)}(G)$, and, by symmetry, also $\min \sigma_{(K,c)}(G)$, visibly deviates from $d \cdot m(G)$. Combined with (2), this implies that $\sigma_{(K,c)}(G)$ stretches in bounded discrete steps over a non-trivial interval depending on d and $\Delta(G)$.

Theorem 1. *If K is a complete graph of order n with $n \geq 4$, $c : E(K) \rightarrow \{\pm 1\}$ is a ± 1 -edge labeling of K such that $m^+(K) = d \binom{n}{2}$, and G is a spanning subgraph of K of maximum degree at most Δ , then there is a permutation π from S_n with*

$$m^+(G_\pi) \geq \left(d + \frac{\min \left\{ 2 - d - 2\sqrt{1-d}, \sqrt{d} - d \right\}}{2\Delta + 1} - \frac{3}{n-3} \right) m(G).$$

All proofs are given in the second section.

We now focus in more detail on the balanced case, that is, in (K, c) there are equally many plus-edges and minus-edges, or, equivalently, $d = \frac{1}{2}$. Note that K necessarily has an even number of edges in this case, which implies that n is equivalent to 0 or 1 modulo 4. Let $\mathcal{G}(n, \Delta)$ be the set of all triples (K, c, G) such that K is a complete graph of order n , c is a balanced ± 1 -edge labeling of K , and G is a spanning subgraph of K of maximum degree at most Δ .

If

$$c_\Delta = \liminf_{n \rightarrow \infty} \left(\min \left\{ \max \left\{ \frac{m^+(G_\pi)}{m(G)} : \pi \in S_n \right\} : (K, c, G) \in \mathcal{G}(n, \Delta) \right\} \right),$$

then Theorem 1 implies

$$c_\Delta \geq \frac{1}{2} + \frac{3 - 2\sqrt{2}}{4\Delta + 2}.$$

The following construction shows that there is little room for improvements of Theorem 1: Let n be a multiple of 4 and $\Delta + 1$, let G be the disjoint union of $\frac{n}{\Delta+1}$ copies of $K_{\Delta+1}$, and let c be such that the plus-edges of K form the graph that arises by removing a matching of size $\frac{n}{4}$ from the complete bipartite graph $K_{\frac{n}{2}, \frac{n}{2}}$. Note that c is balanced by construction. Now, since each copy of $K_{\Delta+1}$ within K contains at most $(\frac{\Delta+1}{2})^2$ many plus-edges, we obtain

$$c_{\Delta} \leq \frac{\left(\frac{\Delta+1}{2}\right)^2}{\binom{\Delta+1}{2}} = \frac{1}{2} + \frac{1}{2\Delta}.$$

Recall that the *matching number* of a graph G is the maximum number of edges in a spanning subgraph of G that is of maximum degree at most 1. A classical result of Erdős and Gallai, Theorem 4.1 in [10], states that a graph G with n vertices, m edges, and matching number ν satisfies

$$\nu \geq \begin{cases} n - \frac{1}{2} - \sqrt{n^2 - 2m - n + \frac{1}{4}} & , \text{ if } m \leq \frac{8n^2 - 14n + 3}{25} \\ \frac{1}{4}(\sqrt{8m + 1} - 1) & , \text{ otherwise.} \end{cases} \quad (3)$$

If K is a complete graph of sufficiently large order n and c is a balanced edge-labeling of K , then $m^+(K) = \frac{1}{2}\binom{n}{2} \leq \frac{8n^2 - 14n + 3}{25}$, and (3) implies that K contains a matching consisting of

$$n - \frac{1}{2} - \sqrt{n^2 - \binom{n}{2} - n + \frac{1}{4}} = \left(2 - \sqrt{2}\right) \frac{n}{2} - o(n)$$

plus-edges. Since every spanning subgraph of K of maximum degree at most 1 has at most $\frac{n}{2}$ edges, this implies $c_1 \geq 2 - \sqrt{2}$. Constructions given by Erdős and Gallai to show that (3) is best possible imply that

$$c_1 = 2 - \sqrt{2} \approx 0.58.$$

Our second result concerns c_2 .

Theorem 2. *Let K be a complete graph of order n and let $c : E(K) \rightarrow \{\pm 1\}$ be a balanced ± 1 -edge labeling of K .*

(i) *If $n \geq 4$, then there is a Hamiltonian cycle C of K with*

$$m^+(C) \geq \left(2 - \sqrt{2}\right)n - o(n) \approx 0.58n - o(n).$$

(ii) *If $n \equiv 0 \pmod{3}$, then there is a C_3 -factor F of K with*

$$m^+(F) \geq \left(\frac{3\sqrt{2}}{4} - \frac{1}{2}\right)n - o(n) \approx 0.56n - o(n).$$

In the setting of Theorem 2, the inequality (3) easily implies the existence of some spanning 2-regular subgraph H of K , that is, of a 2-factor of K , with $m^+(H) \geq (2 - \sqrt{2})n - o(n)$, obtained by extending the union of two disjoint matchings of size $(1 - \frac{1}{\sqrt{2}})n - o(n)$ in the spanning subgraph of K formed by the plus-edges. Nevertheless, this argument does not allow any control of the structure of H . It is conceivable that c_2 equals c_1 , that is, $c_2 = 2 - \sqrt{2} \approx 0.58$. Choosing G as the disjoint union of copies of K_4 , and choosing c such that the minus-edges essentially form a clique of order $\frac{n}{\sqrt{2}}$, which corresponds to one of the extremal configurations for the estimate (3) of Erdős and Gallai, it follows that $c_3 \leq 1 - \frac{\sqrt{2}}{3} \approx 0.53$.

Before we proceed to the proofs, we illustrate the relation of our results to the discrepancy notions mentioned above. Following Balogh et al. [2], the term (4) below can be considered the *discrepancy of Hamiltonian cycles in the complete graph K* . In the setting considered in Corollary 3, their Theorem 1 from [2] implies that (4) is at least $\frac{1}{128}n - o(n) \approx 0.0078n - o(n)$.

Corollary 3. *If K is a complete graph of order n with $n \geq 4$, then*

$$\min \left\{ \max \left\{ |c(C)| : C \text{ is a Hamiltonian cycle in } K \right\} : c \text{ is a } \pm 1\text{-edge labeling of } K \right\} \quad (4)$$

is at least $(3 - 2\sqrt{2})n - o(n) \approx 0.17n - o(n)$.

Proof. Let the ± 1 -edge labeling c_0 of K minimize the maximum value of $|c_0(C)|$, where C is a Hamiltonian cycle in K , that is, (4) equals

$$\max \{|c_0(C)| : C \text{ is a Hamiltonian cycle in } K\}.$$

Clearly, we may assume that the number of plus-edges under c_0 is at least the number of minus-edges, that is, $|c_0^{-1}(1)| \geq |c_0^{-1}(-1)|$.

For simplicity, we first assume that $n \bmod 4 \in \{0, 1\}$, which implies that K has an even number of edges. Let the ± 1 -edge labeling c_1 of K arise from c_0 by changing $\frac{1}{2}(|c_0^{-1}(1)| - |c_0^{-1}(-1)|)$ of the $+1$ -labels on edges to -1 -labels, which implies that c_1 is balanced. By Theorem 2(i), there is a Hamiltonian cycle C in K with $m_{c_1}^+(C) \geq (2 - \sqrt{2})n - o(n)$, where the index indicates with respect to which labeling we count the plus-edges. By construction,

$$c_0(C) \geq c_1(C) = m_{c_1}^+(C) - m_{c_1}^-(C) = 2m_{c_1}^+(C) - n \geq (3 - 2\sqrt{2})n - o(n),$$

and, hence, (4) is at least $(3 - 2\sqrt{2})n - o(n)$.

If $n \bmod 4 \notin \{0, 1\}$, then removing one or two vertices yields a complete graph K' of order n' with $n' \bmod 4 \in \{0, 1\}$. As above, we obtain the existence of a Hamiltonian cycle C' in K' with $c_0(C') \geq (3 - 2\sqrt{2})n' - o(n')$. Replacing one edge of C' with two or three edges including the removed vertices into C' yields a Hamiltonian cycle C in K with $c_0(C) \geq (3 - 2\sqrt{2})n' - o(n') - 4 = (3 - 2\sqrt{2})n - o(n)$. Hence, also in this case (4) is at least $(3 - 2\sqrt{2})n - o(n)$. \square

Similarly, Theorem 1 implies that the suitably defined *discrepancy of a fixed spanning subgraph with m edges and maximum degree at most Δ in a complete graph of order n* is at least $\left(\frac{3-2\sqrt{2}}{2\Delta+1} - O\left(\frac{1}{n}\right)\right)m$.

2 Proofs

For an even integer n at least 4, let $d^*(n) = \left(\frac{8n^2-14n+3}{25}\right) / \binom{n}{2}$. For an odd integer n at least 5, let $d^*(n) = d^*(n-1)$. The value $d^*(n)$ is non-decreasing as a function of n and tends to $\frac{16}{25}$.

Proof of Theorem 1. Since $n \geq 4$, we have $\frac{1}{2} \leq d^*(n) \leq \frac{16}{25}$.

First, we assume that n is even.

For $d \leq d^*(n)$, we have

$$\begin{aligned} n - \frac{1}{2} - \sqrt{n^2 - 2d\binom{n}{2} - n + \frac{1}{4}} &= n - \frac{1}{2} - \sqrt{(1-d)n^2 - \underbrace{\left((1-d)n - \frac{1}{4}\right)}_{\geq 0}} \\ &\geq n - \frac{1}{2} - \sqrt{(1-d)n^2}. \end{aligned}$$

For $d > d^*(n)$, we have

$$\frac{1}{4} \left(\sqrt{8d\binom{n}{2} + 1} - 1 \right) \geq \frac{\sqrt{d}}{2}n - \frac{1}{2};$$

in fact, subtracting the term on the right side of this inequality from the term on the left side yields a term that is decreasing with respect to d for d in $[d^*(n), 1]$, and that equals 0 for $d = 1$. Therefore, by (3) applied to the spanning subgraph of K containing all plus-edges of K , there is a perfect matching M_K in K such that

$$|M_K \cap c^{-1}(1)| \geq \begin{cases} (1 - \sqrt{1-d})n - \frac{1}{2} & , \text{ if } d \leq d^*(n), \text{ and} \\ \frac{\sqrt{d}}{2}n - \frac{1}{2} & , \text{ otherwise.} \end{cases}$$

For $p = \frac{|M_K \cap c^{-1}(1)|}{|M_K|}$, this implies

$$p \geq \begin{cases} 2 - 2\sqrt{1-d} - \frac{1}{n}, & \text{ if } d \leq d^*(n), \text{ and} \\ \sqrt{d} - \frac{1}{n}, & \text{ otherwise.} \end{cases} \quad (5)$$

Since G has maximum degree at most Δ , a simple greedy argument implies that a maximum matching M_G^0 in G satisfies

$$|M_G^0| \geq \frac{m(G)}{2\Delta - 1}. \quad (6)$$

Let M_G be a perfect matching in $K = G \cup \overline{G}$ containing M_G^0 , that is, we extend M_G^0 by adding edges from \overline{G} .

Now, instead of choosing π from S_n uniformly at random, which leads to $\mathbb{E}[m^+(G_\pi)] \stackrel{(1)}{=} d \cdot m(G)$, we change the random choice of π as follows in order to exploit M_K and M_G :

- We bijectively assign the $\frac{n}{2}$ edges in M_G uniformly at random to the $\frac{n}{2}$ edges in M_K , that is, each of the $(\frac{n}{2})!$ assignments is equally likely.
- If an edge uv from M_G is assigned to an edge xy from M_K , then we choose π from S_n such that $(\pi(u), \pi(v)) = (x, y)$ or $(\pi(u), \pi(v)) = (y, x)$ equally likely. Considering all $\frac{n}{2}$ edges of the perfect matchings, this leads to $2^{\frac{n}{2}}$ many possibilities.

Altogether, we choose the permutation π uniformly at random from a subset of $(\frac{n}{2})! 2^{\frac{n}{2}}$ permutations from S_n .

For $uv \in M_G^0$, we obtain

$$\mathbb{P}[c(\pi(u)\pi(v)) = 1] = p,$$

because the fraction of plus-edges in M_K is exactly p .

For $uv \in E(G) \setminus M_G^0$, note that there are exactly $(d\binom{n}{2} - \frac{pn}{2})$ plus-edges in $K - M_K$, and that $\pi(u)\pi(v)$ equals each of these with probability $\frac{2(\frac{n}{2}-2)!2^{\frac{n}{2}-2}}{(\frac{n}{2})!2^{\frac{n}{2}}}$. In fact, for $\pi(u)\pi(v)$ to equal some plus-edge xy in $K - M_K$, the two edges from M_G containing u and v have to be assigned to the two edges from M_K containing x and y , and $\{\pi(u), \pi(v)\}$ has to equal $\{x, y\}$. There are exactly two possibilities for this. The remaining $\frac{n}{2} - 2$ edges from M_G can be mapped onto the remaining $\frac{n}{2} - 2$ edges from M_K without any further restriction. There are exactly $(\frac{n}{2} - 2)! 2^{\frac{n}{2}-2}$ possibilities for this.

We obtain that

$$\mathbb{P}[c(\pi(u)\pi(v)) = 1] = \frac{2(\frac{n}{2}-2)!2^{\frac{n}{2}-2}}{(\frac{n}{2})!2^{\frac{n}{2}}} \left(d\binom{n}{2} - \frac{pn}{2} \right) \stackrel{p \leq 1}{\geq} d - \frac{1-d}{n-2} \stackrel{d \geq 0}{\geq} d - \frac{1}{n-2}$$

Note that

$$d \leq \begin{cases} 2 - 2\sqrt{1-d} & , \text{ if } d \leq d^*(n), \text{ and} \\ \sqrt{d} & , \text{ otherwise.} \end{cases} \quad (7)$$

If $d \leq d^*(n)$, then linearity of expectation implies

$$\begin{aligned} \mathbb{E}[m^+(G_\pi)] &\geq p|M_G^0| + \left(d - \frac{1}{n-2} \right) (m(G) - |M_G^0|) \\ &\stackrel{(5)}{\geq} \left(2 - 2\sqrt{1-d} - \frac{1}{n} \right) |M_G^0| + \left(d - \frac{1}{n-2} \right) (m(G) - |M_G^0|) \\ &\geq \left(2 - 2\sqrt{1-d} \right) |M_G^0| + d(m(G) - |M_G^0|) - \frac{m(G)}{n-2} \end{aligned}$$

$$\begin{aligned}
& \stackrel{(6),(7)}{\geq} \left(2 - 2\sqrt{1-d}\right) \frac{m(G)}{2\Delta+1} + d \left(m(G) - \frac{m(G)}{2\Delta+1}\right) - \frac{m(G)}{n-2} \\
& = \left(d + \frac{2-d-2\sqrt{1-d}}{2\Delta+1} - \frac{1}{n-2}\right) m(G).
\end{aligned}$$

Similarly, if $d > d^*(n)$, then linearity of expectation and a similar estimation as above imply

$$\begin{aligned}
\mathbb{E}[m^+(G_\pi)] & \stackrel{(5),(6),(7)}{\geq} \sqrt{d} \frac{m(G)}{2\Delta+1} + d \left(m(G) - \frac{m(G)}{2\Delta+1}\right) - \frac{m(G)}{n-2} \\
& = \left(d + \frac{\sqrt{d}-d}{2\Delta+1} - \frac{1}{n-2}\right) m(G),
\end{aligned}$$

which completes the proof in the case that n is even.

Now, let n be odd. There is a vertex x of K such that $m^+(K-x) \geq d \binom{n-1}{2}$. Possibly replacing G by an isomorphic copy, we may assume that x is a vertex of minimum degree in G , that is, $m(G-x) = m(G) - \delta(G) \geq \left(1 - \frac{2}{n}\right) m(G)$. Therefore, applying the above estimates to $(K-x, c|_{E(K-x)})$ and $G-x$, and using $d^*(n) = d^*(n-1)$, we obtain

$$\begin{aligned}
\mathbb{E}[m^+(G_\pi)] & \geq \begin{cases} \left(d + \frac{2-d-2\sqrt{1-d}}{2\Delta+1} - \frac{1}{n-3}\right) (m(G) - \delta(G)), & \text{if } d \leq d^*(n), \text{ and} \\ \left(d + \frac{\sqrt{d}-d}{2\Delta+1} - \frac{1}{n-3}\right) (m(G) - \delta(G)), & \text{otherwise} \end{cases} \\
& \geq \begin{cases} \left(d + \frac{2-d-2\sqrt{1-d}}{2\Delta+1} - \frac{1}{n-3}\right) \left(1 - \frac{2}{n}\right) m(G), & \text{if } d \leq d^*(n), \text{ and} \\ \left(d + \frac{\sqrt{d}-d}{2\Delta+1} - \frac{1}{n-3}\right) \left(1 - \frac{2}{n}\right) m(G), & \text{otherwise} \end{cases} \\
& \geq \begin{cases} \left(d + \frac{2-d-2\sqrt{1-d}}{2\Delta+1} - \frac{3}{n-3}\right) m(G), & \text{if } d \leq d^*(n), \text{ and} \\ \left(d + \frac{\sqrt{d}-d}{2\Delta+1} - \frac{3}{n-3}\right) m(G), & \text{otherwise,} \end{cases}
\end{aligned}$$

which completes the proof. \square

Proof of Theorem 2(i). Let G be the spanning subgraph of K formed by the plus-edges, that is, the graph G equals $(V(K), c^{-1}(1))$. In view of the desired statement, we may assume that $n \geq 10$. It is easy to see that the desired statement is equivalent (up to the specific choice of the $o(n)$ term) to the existence of non-trivial disjoint paths P_1, \dots, P_k in G such that

$$m(P_1) + \dots + m(P_k) \geq 2n + 3 - \sqrt{2n^2 + 14n + 1}.$$

In fact, removing the minus-edges from a Hamiltonian cycle C as in the statement yields such paths (as well as some isolated vertices), and, conversely, such paths (and the remaining isolated vertices) can easily be concatenated with edges from K to form a Hamiltonian cycle of K . Therefore, let the non-trivial disjoint paths P_1, \dots, P_k in G be chosen such that

- $m(H)$ is as large as possible, where $H = P_1 \cup \dots \cup P_k$, and
- subject to the first condition, the number k of paths is as small as possible.

For a contradiction, we suppose that $m(H) < 2n + 3 - \sqrt{2n^2 + 14n + 1}$. For $i \in \{1, 2\}$, let V_i be the set of vertices that have degree i in H , and, let $n_i = |V_i|$. Let $V_0 = V(G) \setminus (V_1 \cup V_2)$, and $n_0 = |V_0|$. Let P_i have the endvertices x_i and y_i for every $i \in [k]$, that is, $V_1 = \{x_1, y_1, \dots, x_k, y_k\}$. Note that $n_1 = 2k$, $n_2 = m(H) - k$, and $n_0 = n - n_1 - n_2 = n - m(H) - k$.

Claim 4. $m(G[V_0 \cup V_1]) \leq k$.

Proof. If uv is an edge of $G[V_0 \cup V_1]$ that does not belong to $\{x_i y_i : i \in [k]\}$, then $H + uv$ is the union of non-trivial disjoint paths in G with more edges than H , which is a contradiction. Therefore, the edge set of $G[V_0 \cup V_1]$ is contained in $\{x_i y_i : i \in [k]\}$. \square

The components of $H - V_1$ are paths Q_1, \dots, Q_ℓ with $\ell \leq k$, where we may assume that $Q_i = P_i - x_i - y_i$. Note that a Q_i may be trivial, that is, consist of just one vertex only. If $\ell = 0$, then $V(K) = V_0 \cup V_1$, and Claim 4 implies $m(G) = k \leq \frac{n}{2} < \frac{1}{2} \binom{n}{2}$, which contradicts the hypothesis that c is balanced. Since G contains ℓ disjoint paths P_1, \dots, P_ℓ of order at least 3, we have $\ell \leq \frac{n}{3}$.

Claim 5. For every vertex u from V_0 , there are at most $\frac{1}{2}(n_2 - \ell + 2)$ edges in G between u and V_2 , and u is adjacent to an endvertex in at most one of the paths Q_1, \dots, Q_ℓ .

Proof. If u is adjacent to two consecutive vertices v and w of some Q_i , then $H - vw + vu + uw$ is the union of non-trivial disjoint paths in G with more edges than H , which is a contradiction. Hence, the vertex u has at most $\frac{n(Q_i)+1}{2}$ many neighbors in $V(Q_i)$ for every i in $[\ell]$. If there are two distinct paths Q_i and Q_j such that u is adjacent to an endvertex x'_i in Q_i as well as an endvertex x'_j in Q_j , then, by symmetry, we may assume that x_i is a neighbor of x'_i in H , and x_j is a neighbor of x'_j in H , and $H - x_i x'_i - x_j x'_j + x'_i u + x'_j u$ is the union of less than k non-trivial disjoint paths in G with the same number of edges as H , which is a contradiction. So, assuming that u is adjacent to an endvertex of Q_i , for every $i' \in [\ell] \setminus \{i\}$, the vertex u is adjacent to at most $\frac{n(Q_{i'})-1}{2}$ vertices of $Q_{i'}$. This implies that the number of neighbors of u in V_2 is at most

$$\frac{n(Q_i) + 1}{2} + \sum_{j \in [\ell] \setminus \{i\}} \frac{n(Q_j) - 1}{2} = \frac{n_2 - \ell + 2}{2}$$

Clearly, if u is not adjacent to any endvertex of a path Q_i , this upper bound also holds. \square

Let d be the average number of neighbors in V_2 of the vertices in V_0 , that is, there are dn_0 edges in G between V_0 and V_2 altogether. Claim 5 implies

$$d \leq \frac{n_2 - \ell + 2}{2} \leq \frac{n_2 + 1}{2}. \quad (8)$$

Claim 6. There are at least $\binom{d-1}{2}$ non-edges in $G[V_2]$.

Proof. By the definition of d , there is a vertex u from V_0 that has at least d neighbors in V_2 . By Claim 5, the vertex u is adjacent to an endvertex in at most one of the paths Q_1, \dots, Q_ℓ . Therefore, removing at most one of the neighbors of u in V_2 , which is an endvertex of some Q_i , yields a set N of at least $d - 1$ neighbors of u in V_2 such that there is an orientation of $Q_1 \cup \dots \cup Q_\ell$ for which the set of in-neighbors N^- of the vertices from N satisfies $|N^-| = |N|$, that is, every vertex in N has exactly one in-neighbor in $Q_1 \cup \dots \cup Q_\ell$. If v^-w^- is an edge in G for two vertices v^- and w^- from N^- that are the in-neighbors of v and w from N , respectively, then $H - vv^- - ww^- + v^-w^- + vu + uw$ is the union of non-trivial disjoint paths in G with more edges than H , which is a contradiction. Hence, the set N^- is independent, which implies the claim. \square

Claim 7. *If $k \geq 2$, then G contains at most $\frac{1}{2}n_1(n_2 + \ell)$ edges between V_1 and V_2 .*

Proof. Let $i, j \in [k]$. Let $P_j : x_j u_1 \dots u_p y_j$, that is, Q_j is $u_1 \dots u_p$. If there is some $q \in [p - 1]$ such that x_i is adjacent to u_{q+1} and y_{i+1} is adjacent to u_q , where we identify y_{k+1} with y_1 , then $H - u_q u_{q+1} + x_i u_{q+1} + y_{i+1} u_q$ is the union of non-trivial disjoint paths in G with more edges than H , which is a contradiction. Note that this is also true even if $j = i$ or $j = i + 1$. Therefore, there are at most $p + 1$ edges in G between $\{x_i, y_{i+1}\}$ and $V(Q_j)$, which easily implies the statement. \square

We are now in a position to estimate the total number $m(G)$ of edges of G in order to derive a contradiction.

First, we assume that $k = 1$. In this case, the claims imply

$$m(G) \leq 1 + dn_0 + \binom{n_2}{2} - \binom{d-1}{2} + n_1 n_2.$$

Considered as a quadratic function of d , the right hand side is maximized for $d = n + \frac{1}{2} - m(H) > \frac{m(H)}{2}$, where the last inequality uses the upper bound on $m(H)$. Hence, the function is increasing for $d \stackrel{(8)}{\leq} \frac{n_2+1}{2} = \frac{m(H)}{2}$, and substituting $d = \frac{m(H)}{2}$ yields

$$m(G) \leq -\frac{1}{8}m(H)^2 + \left(\frac{4n+6}{8}\right)m(H) - 1,$$

which, for $m(H) < 2n + 3 - \sqrt{2n^2 + 14n + 1}$, is strictly less than $\frac{1}{2}\binom{n}{2}$, a contradiction.

Next, let $k \geq 2$. In this case, the claims imply

$$m(G) \leq k + dn_0 + \binom{n_2}{2} - \binom{d-1}{2} + \frac{1}{2}n_1 \left(n_2 + \frac{n}{3}\right),$$

where the last term $n_1 n_2$ has been improved using Claim 7 and $\ell \leq \frac{n}{3}$.

We consider two cases according to the value of k .

First, we suppose that $k \leq 2n - 3m(H) + 2$. In this case, considered as a quadratic function of d , the upper bound on $m(G)$ is maximized for $d = \frac{n_2+1}{2} = \frac{m(H)-k+1}{2}$, where we used (8). Substituting this value for d , we obtain

$$m(G) \leq -\frac{1}{8}k^2 - \left(\frac{2n - 3m(H) - 6}{12}\right)k + \frac{(4n - m(H) - 3)(m(H) + 1)}{8}.$$

Now, considered as a quadratic function of k , this upper bound is maximized for $k = 2$. Substituting this value for k , we obtain

$$m(G) \leq \frac{1}{6}n + \frac{1}{8} - \frac{1}{8}m(H)^2 + \frac{1}{2}m(H)n,$$

which, for $m(H) < 2n + 3 - \sqrt{2n^2 + 14n + 1}$ and $n \geq 10$, is strictly less than $\frac{1}{2}\binom{n}{2}$, a contradiction.

Next, we suppose that $k \geq 2n - 3m(H) + 3$. In this case, considered as a quadratic function of d , the upper bound on $m(G)$ is maximized for $d = n - m(H) - k + \frac{3}{2}$. Substituting this value for d , we obtain

$$m(G) \leq \frac{3}{2}n + km(H) + m(H)^2 - nm(H) - \frac{2}{3}nk + \frac{1}{2}n^2 + \frac{1}{8} - 2m(H).$$

Now, considered as a linear function of k , this upper bound is decreasing in k for $m(H) < 2n + 3 - \sqrt{2n^2 + 14n + 1}$. Hence, substituting $k = 2n - 3m(H) + 3$ yields

$$m(G) \leq -\frac{5}{6}n^2 + \frac{1}{2}(6m(H) - 1)n - 2m(H)^2 + m(H) + \frac{1}{8},$$

which, for $m(H) < 2n + 3 - \sqrt{2n^2 + 14n + 1}$, is strictly less than $\frac{1}{2}\binom{n}{2}$, a contradiction.

This completes the proof. \square

Proof of Theorem 2(ii). For (K, c) as in the statement, let C_1, \dots, C_k with $3k = n$ be the components of a C_3 -factor F of K . For $j \in \{0, 1, 2, 3\}$, let t_j be such that $t_j n$ is the number of i in $[k]$ with $m^+(C_i) = j$. In particular,

$$t_0 + t_1 + t_2 + t_3 = \frac{1}{3}. \quad (9)$$

We assume that F is chosen in such a way that

- $m^+(F) = m^+(C_1) + \dots + m^+(C_k)$ is as large as possible, and
- subject to the first condition, the value of t_2 is as large as possible.

The choice of F allows to upper bound the number of plus-edges between any two of the triangles in F . If C_i and C_j satisfy $m^+(C_i) = m^+(C_j) = 0$, then any plus-edge between $V(C_i)$ and $V(C_j)$ allows to replace C_i and C_j within F by two triangles C'_i and C'_j such that $m^+(C_i) + m^+(C_j) \geq 1$, which would contradict the choice of F . If C_i and C_j satisfy $m^+(C_i) = 0$ and $m^+(C_j) = 3$, and there are at least 4 plus-edges between these two triangles, then there are two such edges, say e and f , that are disjoint, and replacing C_i and C_j within F by two triangles C'_i and C'_j such that C'_i contains e and exactly one edge from C_i , and C'_j contains f and exactly one edge from C_j , yields a contradiction either to the first condition or to the second condition within the choice of F . If $C_i : xyzx$ and $C_j : x'y'z'x'$ satisfy $m^+(C_i) = 2$, $c(xz) = -1$, and $m^+(C_j) = 3$, and there are at least 8 plus-edges between these two triangles, then, by symmetry, we may assume that all

minus-edges between these two triangles are between $\{x, y\}$ and $\{x', y'\}$, and replacing C_i and C_j within F by the two triangles $x'y'zx'$ and $xyz'x$, contradicts the choice of F . Now, suppose that $C_i : xyzx$ and $C_j : x'y'z'x'$ satisfy $m^+(C_i) = 1$, $c(xz) = 1$, $m^+(C_j) = 2$, and $c(x'z') = -1$. Considering the two triangles $xy'zx$ and $x'y'z'x'$ implies that at most two of the four edges xy' , $y'z$, $x'y$, and yz' are plus-edges. Hence, if there are at least 4 plus-edges between the two triangles, then, by symmetry, we may assume that xx' is a plus-edge. Considering the two triangles $xx'zx$ and $yz'y'y$ implies that all the three edges $x'z$, yy' , and yz' are minus-edges. Considering the two triangles $xx'y'x$ and $yzzy$ implies that one of the two edges xy' and zz' is a minus-edges. Finally, considering the two triangles $xx'yx$ and $y'z'zy'$ implies that one of the two edges $x'y$ and $y'z$ is a minus-edges. Altogether, these observations yield at least $3 + 1 + 1$ minus-edges between $V(C_i)$ and $V(C_j)$, which implies that there are at most 4 plus-edges between these two triangles.

The following table summarizes the upper bounds on the number of plus-edges between the different types of triangles in F . Since verifying the correctness of these values is straightforward, we leave the remaining details to the reader.

(i, j)	$m^+(C_j) = 0$	$m^+(C_j) = 1$	$m^+(C_j) = 2$	$m^+(C_j) = 3$
$m^+(C_i) = 0$	0	0	3	3
$m^+(C_i) = 1$	0	1	4	6
$m^+(C_i) = 2$	3	4	5	7
$m^+(C_i) = 3$	3	6	7	9

Since c is balanced, we obtain

$$\begin{aligned}
\left(\frac{1}{4} - o(1)\right) n^2 &= \frac{1}{2} \binom{n}{2} = m^+(K) \\
&\leq t_1 n + 2t_2 n + 3t_3 n + \binom{t_1 n}{2} + 5 \binom{t_2 n}{2} + 9 \binom{t_3 n}{2} \\
&\quad + 3t_0 n(t_2 n + t_3 n) + 4t_1 n t_2 n + 6t_1 n t_3 n + 7t_2 n t_3 n \\
&= \left(\frac{1}{2} t_1^2 + \frac{5}{2} t_2^2 + \frac{9}{2} t_3^2 + 3t_0(t_2 + t_3) + 4t_1 t_2 + 6t_1 t_3 + 7t_2 t_3 + o(1)\right) n^2,
\end{aligned}$$

and, hence,

$$\begin{aligned}
h(t_0, t_1, t_2, t_3) &:= \frac{1}{2} t_1^2 + \frac{5}{2} t_2^2 + \frac{9}{2} t_3^2 + 3t_0(t_2 + t_3) + 4t_1 t_2 + 6t_1 t_3 + 7t_2 t_3 \\
&\geq \frac{1}{4} - o(1).
\end{aligned} \tag{10}$$

Since $m^+(F) = (t_1 + 2t_3 + 3t_3)n$, in order to complete the proof, it suffices to show that the optimum value of the following optimization problem is at least $\frac{3\sqrt{2}}{4} - \frac{1}{2} - o(1)$:

$$\begin{aligned}
\min \quad & t_1 + 2t_2 + 3t_3 \\
s.th. \quad & t_0 + t_1 + t_2 + t_3 = \frac{1}{3} \\
& h(t_0, t_1, t_2, t_3) \geq \frac{1}{4} - o(1) \\
& t_0, t_1, t_2, t_3 \geq 0
\end{aligned} \tag{11}$$

Since $\frac{\partial}{\partial t_0} h(t_0, t_1, t_2, t_3)$ is strictly less than $\frac{\partial}{\partial t_i} h(t_0, t_1, t_2, t_3)$ for every $i \in [3]$ and every feasible solution (t_0, t_1, t_2, t_3) of (11), every optimum solution (t_0, t_1, t_2, t_3) of (11) satisfies the inequality (10) with equality; otherwise, slightly increasing t_0 and decreasing $t_1 + t_2 + t_3$ by the same amount would yield a better feasible solution. Note that $\frac{\partial}{\partial t_3} h(t_0, t_1, t_2, t_3) - \frac{\partial}{\partial t_0} h(t_0, t_1, t_2, t_3) = 3t_0 + 6t_1 + 4t_2 + 6t_3 \geq 1$ for every feasible solution (t_0, t_1, t_2, t_3) of (11), that $\frac{\partial}{\partial t_3} h(t_0, t_1, t_2, t_3) - \frac{\partial}{\partial t_1} h(t_0, t_1, t_2, t_3) = 3t_0 + 5t_1 + 3t_2 + 3t_3 \geq 1$ for every feasible solution (t_0, t_1, t_2, t_3) of (11), and that $\frac{\partial}{\partial t_3} h(t_0, t_1, t_2, t_3) - \frac{\partial}{\partial t_2} h(t_0, t_1, t_2, t_3) = 2t_1 + 2t_2 + 2t_3 \geq \frac{2}{3}$ for every feasible solution (t_0, t_1, t_2, t_3) of (11) with $t_0 = 0$. This implies that, for every optimum solution (t_0, t_1, t_2, t_3) of (11), possibly increasing t_3 by $o(1)$ and decreasing $t_0 + t_1 + t_2$ by the same amount without violating the condition $t_0, t_1, t_2 \geq 0$, yields a feasible solution for the following optimization problem (12), whose objective function value is larger by at most $o(1)$.

$$\begin{aligned}
\min \quad & t_1 + 2t_2 + 3t_3 \\
s.th. \quad & t_0 + t_1 + t_2 + t_3 = \frac{1}{3} \\
& h(t_0, t_1, t_2, t_3) = \frac{1}{4} \\
& t_0, t_1, t_2, t_3 \geq 0
\end{aligned} \tag{12}$$

Since the optimum value of (12) is at least the optimum value of (11), this implies that the two optimal values differ only by $o(1)$. Hence, in order to complete the proof, it suffices to show that the optimum value of (12) is at least $\frac{3\sqrt{2}}{4} - \frac{1}{2}$.

The equation $h(t_0, t_1, t_2, t_3) = \frac{1}{4}$ allows to express t_0 in terms of t_1, t_2 , and t_3 , and substituting the corresponding expression into $t_0 + t_1 + t_2 + t_3 = \frac{1}{3}$, allows to express t_2 in terms of t_1 and t_3 . Substituting these expressions for t_0 and t_2 , we obtain

$$t_1 + 2t_2 + 3t_3 = 3t_1 + 5t_3 + 2 - \sqrt{8t_1^2 + (32t_3 + 8)t_1 + 16t_3^2 + 16t_3 + 2} =: f(t_1, t_3).$$

It follows that the optimum value of (12) is at least the optimum value of the following optimization problem, where we implicitly relax the conditions “ $t_0 \geq 0$ ” and “ $t_2 \geq 0$ ”:

$$\min \left\{ f(t_1, t_3) : t_1, t_3 \geq 0 \text{ and } t_1 + t_3 \leq \frac{1}{3} \right\} \tag{13}$$

Since there is no point (t_1, t_3) in the interior of $\{(x, y) : x, y \geq 0 \text{ and } x + y \leq \frac{1}{3}\}$ for which $\frac{\partial}{\partial t_1} f(t_1, t_3) = \frac{\partial}{\partial t_3} f(t_1, t_3) = 0$, the minimum (13) is assumed on the boundary.

Since $f(t_1, 0) = (3 - 2\sqrt{2})t_1 + 2 - \sqrt{2}$, we obtain

$$\min \left\{ f(t_1, 0) : 0 \leq t_1 \leq \frac{1}{3} \right\} = f(0, 0) = 2 - \sqrt{2} > \frac{3\sqrt{2}}{4} - \frac{1}{2}.$$

Since $\frac{\partial}{\partial t_3} f(0, t_3) = 0$ for $t_3 \in [0, \frac{1}{3}]$ only if $t_3 = \frac{5\sqrt{2}}{12} - \frac{1}{2} \approx 0.08$, we obtain

$$\begin{aligned} \min \left\{ f(0, t_3) : 0 \leq t_3 \leq \frac{1}{3} \right\} &= \min \left\{ f(0, 0), f\left(0, \frac{5\sqrt{2}}{12} - \frac{1}{2}\right), f\left(0, \frac{1}{3}\right) \right\} \\ &= \frac{3\sqrt{2}}{4} - \frac{1}{2}. \end{aligned}$$

Finally, since $\frac{\partial}{\partial t_1} f(t_1, \frac{1}{3} - t_1) = 0$ for $t_1 \in [0, \frac{1}{3}]$ only if $t_1 = \frac{5\sqrt{6}}{18} - \frac{1}{2}$, we obtain

$$\begin{aligned} &\min \left\{ f\left(t_1, \frac{1}{3} - t_1\right) : 0 \leq t_1 \leq \frac{1}{3} \right\} \\ &= \min \left\{ f\left(0, \frac{1}{3}\right), f\left(\frac{5\sqrt{6}}{18} - \frac{1}{2}, \frac{5}{6} - \frac{5\sqrt{6}}{18}\right), f\left(\frac{1}{3}, 0\right) \right\} \\ &= \frac{14}{3} - \frac{5\sqrt{6}}{3} > \frac{3\sqrt{2}}{4} - \frac{1}{2}. \end{aligned}$$

Altogether, it follows that the optimum value of (13) is $\frac{3\sqrt{2}}{4} - \frac{1}{2}$, which implies that the optimum value of (12) is also at least this value. This completes the proof. \square

It seems possible to apply a similar approach to other graphs whose components are all isomorphic, such as, for instance, K_4 -factors or $K_{1,3}$ -factors.

3 Conclusion

An obvious task motivated by our results is to determine values c_Δ more precisely; at least, for small values of Δ . Furthermore, it seems straightforward to generalize Theorem 2 to not necessarily balanced edge labelings c .

There seem to be no immediate directed analogues of our results. If D is the complete digraph with vertex set $[2n]$, that is, between every two vertices of D , there are both possible arcs, and $c : A(D) \rightarrow \{\pm 1\}$ is such that $c((u, v)) = 1$ for all arcs (u, v) with $u \in [n]$, and $c((u, v)) = -1$ for all arcs (u, v) with $u \in [2n] \setminus [n]$, then there are equally many plus- and minus-arcs, but every directed Hamiltonian cycle has exactly n plus- and n minus-arcs. Similarly, there are tournaments T with a unique directed Hamiltonian cycle, which allows to force all plus-arcs or all minus-arcs in the directed Hamiltonian cycle even though T has equally many plus- and minus-arcs.

The problems studied in this paper clearly relate to classical and new results concerning extremal graph theory, Ramsey theory, and Hamiltonicity. Further directions that could be pursued may be inspired by [1, 18].

References

- [1] J. Bang-Jensen and G. Gutin, Alternating cycles and paths in edge-coloured multi-graphs: A survey, *Discrete Math.*, 165/166:39–60, 1997.
- [2] J. Balogh, B. Csaba, Y. Jing, and A. Pluhár, On the discrepancies of graphs, *Electron. J. Combin.*, 27:Paper No. 2.12, 2020.
- [3] J. Balogh, B. Csaba, A. Pluhár, and A. Treglown, A discrepancy version of the Hajnal-Szemerédi theorem, *Combin. Probab. Comput.*, 30:444–459, 2021.
- [4] Y. Caro, Zero-sum problems - a survey, *Discrete Math.*, 152:93–113, 1996.
- [5] Y. Caro, A. Hansberg, J. Lauri, and C. Zarb, On zero-sum spanning trees and zero-sum connectivity, *Electron. J. Combin.*, 29:#P1.9, 2022.
- [6] Y. Caro, A. Hansberg, and A. Montejano, Unavoidable chromatic patterns in 2-colorings of the complete graph, *J. Graph Theory*, 97:123–147, 2021.
- [7] Y. Caro and R. Yuster, On zero-sum and almost zero-sum subgraphs over \mathbb{Z} , *Graphs Combin.*, 32:49–63, 2016.
- [8] J. Cutler and B. Montágh, Unavoidable subgraphs of colored graphs, *Discrete Math.*, 308:4396–4413, 2008.
- [9] S. Ehard, E. Mohr, and D. Rautenbach, Low weight perfect matchings, *Electron. J. Combin.*, 27:#P4.49, 2020.
- [10] P. Erdős and T. Gallai, On maximal paths and circuits of graphs, *Acta Math. Acad. Sci. Hungar.*, 10:337–356, 1959.
- [11] P. Erdős, Z. Füredi, M. Loeb, and V.T. Sós, Discrepancy of trees, *Studia Sci. Math. Hungar.*, 30:47–57, 1995.
- [12] A. Freschi, J. Hyde, J. Lada, and A. Treglown, A note on color-bias Hamilton cycles in dense graphs, *SIAM J. Discrete Math.*, 35:970–975, 2021.
- [13] J. Fox and B. Sudakov, Unavoidable patterns, *J. Combin. Theory Ser. A*, 115:1561–1569, 2008.
- [14] W. Gao and A. Geroldinger, Zero-sum problems in finite abelian groups: A survey, *Expo. Math.*, 24:337–369, 2006.
- [15] T. Kittipassorn and P. Sinsap, On the existence of zero-sum perfect matchings of complete graphs, *Ars Math. Contemp.*, 23:Paper No. 3.07, 2023.
- [16] E. Mohr, J. Pardey, and D. Rautenbach, Zero-sum copies of spanning forests in zero-sum complete graphs, *Graphs Combin.*, 38(5):Paper No. 132, 2022.
- [17] J. Pardey and D. Rautenbach, Efficiently finding low-sum copies of spanning forests in zero-sum complete graphs via conditional expectation, *Discrete Appl. Math.*, 328:108–116, 2023.
- [18] S. Zhang, X. Li, and H. Broersma, Heavy paths and cycles in weighted graphs, *Discrete Math.*, 223:327–336, 2000.