

Editing to Eulerian Graphs^{*}

Konrad K. Dabrowski¹, Petr A. Golovach², Pim van 't Hof², and
Daniël Paulusma¹

¹ School of Engineering and Computing Sciences, Durham University,
Science Laboratories, South Road, Durham DH1 3LE, United Kingdom
{konrad.dabrowski,daniel.paulusma}@durham.ac.uk

² Department of Informatics, University of Bergen,
PB 7803, 5020 Bergen, Norway
{petr.golovach,pim.vanthof}@ii.uib.no

Abstract. The Eulerian Editing problem asks, given a graph G and an integer k , whether G can be modified into an Eulerian graph using at most k edge additions and edge deletions. We show that this problem is polynomial-time solvable for both undirected and directed graphs. We generalize these results for problems with degree parity constraints and degree balance constraints, respectively. We also consider the variants where vertex deletions are permitted. Combined with known results, this leads to full complexity classifications for both undirected and directed graphs and for every subset of the three graph operations.

Keywords: Eulerian graphs, graph editing, polynomial algorithm

1 Introduction

Graph modification problems play a central role in algorithmic graph theory, partly due to the fact that they naturally arise in numerous practical applications. A graph modification problem takes as input a graph G and an integer k , and asks whether G can be modified into a graph belonging to a prescribed graph class \mathcal{H} , using at most k operations of certain allowed types. The most common operations that are considered in this context are edge additions (\mathcal{H} -COMPLETION), edge deletions (\mathcal{H} -EDGE DELETION), vertex deletions (\mathcal{H} -VERTEX DELETION), and a combination of edge additions and edge deletions (\mathcal{H} -EDITING). The intensive study of graph modification problems has produced a plethora of classical and parameterized complexity results (see e.g. [2–8,10,13,15–17,19,21–23,25,26]).

^{*} The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 267959 and from EPSRC Grant EP/K025090/1. An extended abstract of this paper appeared in the proceedings of FSTTCS 2014 [9].

An undirected (resp. directed) graph G is Eulerian if it contains a walk that begins and ends with the same vertex and that uses every edge (resp. arc) exactly once. As an immediate consequence, an undirected graph is Eulerian if and only if it is connected and every vertex has even degree. Similarly, a directed graph is Eulerian if it is strongly connected³ and balanced, i.e. the in-degree of every vertex equals its out-degree. Eulerian graphs form a well-known graph class both within algorithmic and structural graph theory.

Several groups of authors have investigated the problem of deciding whether or not a given undirected graph can be made Eulerian using a small number of operations. Boesch et al. [2] presented a polynomial-time algorithm for EULERIAN COMPLETION, and Cai and Yang [5] showed that the problems EULERIAN VERTEX DELETION and EULERIAN EDGE DELETION are NP-complete [5]. When parameterized by the number k of allowed operations, it is known that EULERIAN VERTEX DELETION is W[1]-hard [5], while EULERIAN EDGE DELETION is fixed-parameter tractable [8]. Cygan et al. [8] showed that the classical and parameterized complexity results for EULERIAN VERTEX DELETION and EULERIAN EDGE DELETION also hold for the directed variants of these problems. Recently, Goyal et al. [17] improved the fixed-parameter tractability results of Cygan et al. [8] for the directed and undirected variants of EULERIAN EDGE DELETION by giving algorithms with running times that are single-exponential in k . The same authors also proved that the UNDIRECTED CONNECTED ODD EDGE DELETION problem, which asks whether it is possible to obtain a connected graph in which all vertices have odd degree by deleting at most k edges, is fixed-parameter tractable when parameterized by k .

Another problem that can be seen as involving editing to an Eulerian multigraph is the CHINESE POSTMAN problem, also known as the ROUTE INSPECTION problem [20]. In this problem a connected graph G , together with an integer k , is given and the question is whether or not there exists a closed walk in G that uses every edge of G at least once and that has length at most $|E(G)| + k$. In other words, can we add a total of at most k copies of existing edges to G in order to modify G into an Eulerian multigraph? Edmonds and Johnson [12] showed that both the undirected and directed variant of this problem can be solved in polynomial time. The RURAL POSTMAN problem generalizes the CHINESE POSTMAN problem,

³ Replacing “strongly connected” by “weakly connected” yields an equivalent definition of Eulerian digraphs, as it is well-known that a balanced digraph is strongly connected if and only if it is weakly connected (see e.g. [8]).

as it requires that only every edge of some subset of $E(G)$ needs to be used at least once in the closed walk. Dorn et al. [10] proved that the RURAL POSTMAN is fixed-parameter tractable for directed multigraphs when parameterized by the number of arcs that may be added.

Our Contribution We generalize, extend and complement known results on graph modification problems dealing with Eulerian graphs and digraphs. The main contribution of this paper consists of two non-trivial polynomial-time algorithms: one for solving the EULERIAN EDITING problem, and one for solving the directed variant of this problem. Given the aforementioned NP-completeness result for EULERIAN EDGE DELETION and the fact that \mathcal{H} -EDITING is NP-complete for many graph classes \mathcal{H} [3, 26], we find it particularly interesting that EULERIAN EDITING turns out to be polynomial-time solvable. To the best of our knowledge, the only other natural non-trivial graph class \mathcal{H} for which \mathcal{H} -EDITING is known to be polynomial-time solvable is the class of split graphs [18].

In fact, our polynomial-time algorithms are implications of two more general results. In order to formally state these results, we need to introduce some terminology. Let ea , ed and vd denote the operations edge addition, edge deletion and vertex deletion, respectively. For any set $S \subseteq \{\text{ea}, \text{ed}, \text{vd}\}$ and non-negative integer k , we say that a graph G can be (S, k) -modified into a graph H if H can be obtained from G by using at most k operations from S . We define the following problem for every $S \subseteq \{\text{ea}, \text{ed}, \text{vd}\}$:

CDPE(S): CONNECTED DEGREE PARITY EDITING(S)
Instance: A (simple) graph G , an integer k
and a function $\delta: V(G) \rightarrow \{0, 1\}$.
Question: Can G be (S, k) -modified into a connected graph H
with $d_H(v) \equiv \delta(v) \pmod{2}$ for each $v \in V(H)$?

Inspired by the work of Cygan et al. [8] on directed Eulerian graphs, we also study a natural directed variant of the CDPE(S) problem. Denoting the in- and out-degree of a vertex v in a digraph G by $d_G^{\text{in}}(v)$ and $d_G^{\text{out}}(v)$, respectively, we define the following problem for every $S \subseteq \{\text{ea}, \text{ed}, \text{vd}\}$:

CDBE(S): CONNECTED DEGREE BALANCE EDITING(S)

Instance: A (simple) digraph G , an integer k
and a function $\delta: V(G) \rightarrow \mathbb{Z}$.

Question: Can G be (S, k) -modified into a weakly connected digraph H with $d_H^{out}(v) - d_H^{in}(v) = \delta(v)$ for each $v \in V(H)$?

In Section 3, we prove that CDPE(S) can be solved in polynomial time when $S = \{\text{ea}\}$ and when $S = \{\text{ea}, \text{ed}\}$. The first of these two results extends the result by Boesch et al. [2] on EULERIAN COMPLETION and the second yields the first polynomial-time algorithm for EULERIAN EDITING, as these problems are equivalent to CDPE($\{\text{ea}\}$) and CDPE($\{\text{ea}, \text{ed}\}$), respectively, when we set $\delta \equiv 0$ (i.e. when $\delta(v) = 0$ for every $v \in V(G)$). The complexity of the problem drastically changes when vertex deletion is allowed: we prove that for every subset $S \subseteq \{\text{ea}, \text{ed}, \text{vd}\}$ with $\text{vd} \in S$, the CDPE(S) problem is NP-complete and W[1]-hard with parameter k , even when $\delta \equiv 0$. This complements results by Cai and Yang [5] stating that CDPE(S) is NP-complete and W[1]-hard with parameter k when $S = \{\text{vd}\}$ and $\delta \equiv 0$ or $\delta \equiv 1$. Our results, together with the aforementioned results due to Cygan et al. [8]⁴ and Cai and Yang [5], yield a complete classification of both the classical and the parameterized complexity of CDPE(S) for all $S \subseteq \{\text{ea}, \text{ed}, \text{vd}\}$; see the middle column of Table 1.

In Section 4, we use different and more involved arguments to classify the classical and parameterized complexity of the CDBE(S) problem for all $S \subseteq \{\text{ea}, \text{ed}, \text{vd}\}$. Interestingly, the classification we obtain for CDBE(S) turns out to be identical to the one we obtained for CDPE(S). In particular, our proof of the fact that CDBE(S) is polynomial-time solvable when $S = \{\text{ea}\}$ and $S = \{\text{ea}, \text{ed}\}$ implies that the directed variants of EULERIAN COMPLETION and EULERIAN EDITING are not significantly harder than their undirected counterparts. All results on CDBE(S) are summarized in the right column of Table 1.

We would like to emphasize that there are no obvious hardness reductions between the different problem variants. The parameter k in the problem definitions represents the budget for all operations in total; adding a new operation to S may completely change the problem, as there is no way of forbidding its use. Hence, our polynomial-time algorithms for

⁴ The FPT-results by Cygan et al. [8] only cover CDPE($\{\text{ed}\}$) and CDBE($\{\text{ed}\}$) when $\delta \equiv 0$, but it can easily be seen that their results carry over to CDPE($\{\text{ed}\}$) and CDBE($\{\text{ed}\}$) for any function δ .

S	CDPE(S)	CDBE(S)
ea, ed	P	P
ea	P	P
ed	FPT [8]	FPT [8]
vd	W[1]-hard [5]	W[1]-hard [8]
ea, vd	W[1]-hard	W[1]-hard
ed, vd	W[1]-hard	W[1]-hard
ea, ed, vd	W[1]-hard	W[1]-hard

Table 1. A summary of the results for CDPE(S) and CDBE(S). All results are new except those for which a reference is given. The number of allowed operations k is the parameter in the parameterized results, and if a parameterized result is stated, then the corresponding problem is NP-complete.

CDPE($\{\text{ea, ed}\}$) and CDBE($\{\text{ea, ed}\}$) do not generalize the polynomial-time algorithms for CDPE($\{\text{ea}\}$) and CDBE($\{\text{ea}\}$), and as such require significantly different arguments. In particular, our main result, stating that EULERIAN EDITING is polynomial-time solvable, is not a generalization of the fact that EULERIAN COMPLETION is polynomial-time solvable and stands in no relation to the FPT-result by Cygan et al. [8] for EULERIAN EDGE DELETION.

We end this section by mentioning two similar graph modification frameworks in the literature that formed a direct motivation for the framework defined in this paper. Mathieson and Szeider [23] considered the DEGREE CONSTRAINT EDITING(S) problem, which is that of testing whether a graph G can be (S, k) -modified into a graph H in which the degree of every vertex belongs to some list associated with that vertex; recently some new results for this problem were obtained by Froese et al. [13] and Golovach [16]. Golovach [15] performed a similar study to that of Mathieson and Szeider [23], but with the additional condition that the resulting graph must be connected.

2 Preliminaries

We consider finite graphs $G = (V, E)$ that may be undirected or directed; in the latter case we will always call them digraphs. All our undirected graphs will be simple, that is, without loops or multiple edges; in particular, this is the case for both the input and the output graph in every undirected problem we consider. Similarly, for every directed problem that we consider, we only consider simple digraphs, that is, we do not allow the input or output digraph to contain multiple arcs (but we do allow

opposite arcs). In our proofs we will also make use of *directed multigraphs*, which are digraphs that are permitted to have multiple arcs.

The complete graph on n vertices is denoted K_n and the complete bipartite graph with classes of size s and t is denoted $K_{s,t}$. The *length* of a path or cycle is equal to the number of edges that it contains.

We denote an edge between two vertices u and v in a graph by uv . We denote an arc between two vertices u and v by (u, v) , where u is the *tail* of (u, v) and v is the *head*. We say that a set A of arcs or edges has *size* $|A|$. The disjoint union of two graphs G_1 and G_2 is denoted $G_1 + G_2$.

Let $G = (V, E)$ be a graph or a digraph. Throughout the paper we assume that $n = |V|$ and $m = |E|$. For $U \subseteq V$, we let $G[U]$ be the graph (digraph) with vertex set U and an edge (arc) between two vertices u and v if and only if this is the case in G ; we say that $G[U]$ is *induced by* U . We write $G - U = G[V \setminus U]$. For $E' \subseteq E$, we let $G(E')$ be the graph (digraph) with edge (arc) set E' whose vertex set consists of the end-vertices of the edges in E' ; we say that $G(E')$ is *edge-induced by* E' . Let S be a set of (ordered) pairs of vertices of G . We let $G - S$ be the graph (digraph) obtained by deleting all edges (arcs) of $S \cap E$ from G , and we let $G + S$ be the graph (digraph) obtained by adding all edges (arcs) of $S \setminus E$ to G . We may write $G - e$ or $G + e$ if $S = \{e\}$.

Let $G = (V, E)$ be a graph. A *component* of G is a maximal connected subgraph of G . The *complement* of G is the graph $\bar{G} = (V, \bar{E})$ with vertex set V and an edge between two distinct vertices u and v if and only if $uv \notin E$. For a vertex $v \in V$, we let $N_G(v) = \{u \mid uv \in E\}$ denote its (*open*) *neighbourhood*. The *degree* of v is denoted $d_G(v) = |N_G(v)|$. The graph G is *even* if all its vertices have even degree, and it is *Eulerian* if it is even and connected. We say that a set $D \subseteq E$ is an *edge cut-set* in G if $G - D$ has more components than G . An edge cut-set of size 1 is said to be a *bridge*.

A *matching* of a graph G is a set of edges, in which no two edges have a common end-vertex; it is called a *maximum* matching if its number of edges is maximum over all matchings of G . We need the following lemma due to Micali and Vazirani.

Lemma 1 ([24]). *A maximum matching of an n -vertex graph can be found in $O(n^{5/2})$ time.*

Let $G = (V, E)$ be a digraph. If (u, v) is an arc, then (v, u) is the *reverse* of this arc. For a subset $F \subseteq E$, we let $F^R = \{(u, v) \mid (v, u) \in F\}$ denote the set of arcs whose reverse is in F . The *underlying* graph of G is the undirected graph with vertex set V where two vertices $u, v \in V$

are adjacent if and only if (u, v) or (v, u) is an arc in G . We say that G is *(weakly) connected* if its underlying graph is connected. A *component* of G is a connected component of its underlying graph. An arc $a \in E$ is a *bridge* in G if it is a bridge in the underlying graph of G . A vertex u is an *in-neighbour* or *out-neighbour* of a vertex v if $(u, v) \in E$ or $(v, u) \in E$, respectively. Let $N_G^{in}(v) = \{u \mid (u, v) \in E\}$ and $N_G^{out}(v) = \{u \mid (v, u) \in E\}$, where we call $d_G^{in}(v) = |N_G^{in}(v)|$ and $d_G^{out}(v) = |N_G^{out}(v)|$ the *in-degree* and *out-degree* of v , respectively. A vertex $v \in V$ is *balanced* if $d_G^{out}(v) = d_G^{in}(v)$, or equivalently, its *degree balance* $d_G^{out}(v) - d_G^{in}(v) = 0$. Recall that G is *Eulerian* if it is connected and *balanced*, that is, the out-degree of every vertex is equal to its in-degree.

Let $G = (V, E)$ be a graph and let $T \subseteq V$. A subset $J \subseteq E$ is a *T -join* if the set of odd-degree vertices in $G(J)$ is precisely T . If G is connected and $|T|$ is even then G has at least one T -join. In Section 3 we need to find a *minimum T -join*, that is, one of minimum size. We use the following result of Edmonds and Johnson [12] to do so.

Lemma 2 ([12]). *Let $G = (V, E)$ be a graph, and let $T \subseteq V$. Then a minimum T -join (if one exists) can be found in $O(n^3)$ time.*

Lemma 2 was used by Cygan et al. [8] to solve \mathcal{H} -EDGE DELETION in polynomial time when \mathcal{H} is the class of even graphs. It would immediately yield a polynomial-time algorithm for CDPE($\{\text{ed}\}$) if we dropped the connectivity condition.

We need a variant of Lemma 2 for digraphs in Section 4. Let $G = (V, E)$ be a directed multigraph and let $f : T \rightarrow \mathbb{Z}$ be a function for some $T \subseteq V$. A multiset $E' \subseteq E$ with $T \subseteq V(G(E'))$ is a *directed f -join* in G if the following two conditions hold: $d_{G(E')}^{out}(v) - d_{G(E')}^{in}(v) = f(v)$ for every $v \in T$ and $d_{G(E')}^{out}(v) - d_{G(E')}^{in}(v) = 0$ for every $v \in V(G(E')) \setminus T$. A directed f -join is *minimum* if it has minimum size. The next lemma was used by Cygan et al. [8] to solve \mathcal{H} -EDGE DELETION in polynomial time when \mathcal{H} is the class of balanced digraphs; it would also yield a polynomial-time algorithm for CDBE($\{\text{ed}\}$) if we dropped the connectivity condition.

Lemma 3 ([8]). *Let $G = (V, E)$ be a directed multigraph and $f : T \rightarrow \mathbb{Z}$ be a function for some $T \subseteq V$. A minimum directed f -join F (if one exists) can be found in $O(nm \log n \log \log m)$ time. Moreover, F consists of mutually arc-disjoint directed paths from vertices u with $f(u) > 0$ to vertices v with $f(v) < 0$.*

3 Connected Degree Parity Editing

Let $S \subseteq \{\text{ea}, \text{ed}, \text{vd}\}$. In Section 3.1 we will show that $\text{CDPE}(S)$ is polynomial-time solvable if $S = \{\text{ea}\}$ or $S = \{\text{ea}, \text{ed}\}$ and in Section 3.2 we will show that it is NP-complete and W[1]-hard with parameter k if $\text{vd} \in S$.

3.1 The Polynomial-Time Solvable Cases

First, let $\{\text{ea}\} \subseteq S \subseteq \{\text{ea}, \text{ed}\}$. Let (G, δ, k) be an instance of $\text{CDPE}(S)$ with $G = (V, E)$. Let A be a set of edges not in G , and let D be a set of edges in G , with $D = \emptyset$ if $S = \{\text{ea}\}$. We say that (A, D) is a *solution* for (G, δ, k) if its *size* $|A| + |D| \leq k$, the congruence $d_H(u) \equiv \delta(u) \pmod{2}$ holds for every vertex u and the graph $H = G + A - D$ is connected; if we drop the condition that H is connected then (A, D) is a *semi-solution* for (G, δ, k) . If $S = \{\text{ea}\}$ we may denote the solution by A rather than (A, D) (since $D = \emptyset$). We consider the optimization version for $\text{CDPE}(S)$. The input is a pair (G, δ) , and we aim to find the minimum k such that (G, δ, k) has a solution (if one exists). We call such a solution *optimal* and denote its size by $\text{opt}_S(G, \delta)$. We say that a (semi)-solution for (G, δ, k) is also a (semi)-solution for (G, δ) . If (G, δ, k) has no solution for any value of k , then (G, δ) is a *no-instance* of $\text{CDPE}(S)$ and $\text{opt}_S(G, \delta) = +\infty$.

Let $T = \{v \in V \mid d_G(v) \not\equiv \delta(v) \pmod{2}\}$. Define $G_S = K_n$ if $S = \{\text{ea}, \text{ed}\}$ and $G_S = \overline{G}$ if $S = \{\text{ea}\}$. Note that if $S = \{\text{ea}\}$ then G_S contains no edges of G , so in this case any T -join in G_S can only contain edges in $E(\overline{G})$. The following key lemma is an easy observation.

Lemma 4. *Let $\{\text{ea}\} \subseteq S \subseteq \{\text{ea}, \text{ed}\}$. Let (G, δ) be an instance of $\text{CDPE}(S)$ and $A \subseteq E(\overline{G})$, $D \subseteq E(G)$. Then (A, D) is a semi-solution of $\text{CDPE}(S)$ if and only if $A \cup D$ is a T -join in G_S .*

We recall that Boesch et al. [2] proved that EULERIAN COMPLETION can be solved in polynomial time, that is, $\text{CDPE}(S)$ is polynomial time solvable if $S = \{\text{ea}\}$ and $\delta \equiv 0$. We extend this result to arbitrary δ . Our proof is based around similar ideas but we also had to do some further analysis. The main difference in the two proofs is the following. If $\delta \equiv 0$ then none of the added edges in a solution will be a bridge in the modified graph (as the number of vertices of odd degree in a graph is always even). However this is no longer true for arbitrary δ and extra arguments are needed. We therefore present a full proof of our result.

Theorem 1. *Let $S = \{\text{ea}\}$. Then $\text{CDPE}(S)$ can be solved in $O(n^3)$ time.*

Proof. Let $S = \{ea\}$ and let (G, δ) be an instance of CDPE(S). We first use Lemma 2 to check in $O(n^3)$ time whether G_S has a T -join. If not then (G, δ) has no semi-solution by Lemma 4, and thus no solution either. We may therefore assume that $|T|$ is even and F is a minimum T -join in G_S . (Recall that Lemma 2 states that we can find F in $O(n^3)$ time if it exists.) We also assume that either $T \neq \emptyset$ or G is not connected, otherwise the trivial solution $A = \emptyset$ is clearly optimal. Let p be the number of components of G that do not contain any vertex of T and let q be the number of components of G that contain at least one vertex of T .

We will prove the following series of statements. Under the assumptions made in the previous paragraph, these statements give necessary and sufficient conditions for (G, δ) to be a yes-instance, and if (G, δ) is a yes-instance tell us the exact size of an optimal solution for (G, δ) . We recall that $G_1 + G_2$ denotes the disjoint union of two graphs G_1 and G_2 .

- (G, δ) is a no-instance if $p = 2, q = 0$ and $G = K_1 + K_t$ for $t \geq 1$.
- $opt_S(G, \delta) = 4$ if $p = 2, q = 0$ and $G = K_s + K_t$ for $s, t \geq 2$.
- $opt_S(G, \delta) = 3$ if $p = 2, q = 0$ and G has a component that is not complete.
- $opt_S(G, \delta) = p$ if $p \geq 3, q = 0$.
- $opt_S(G, \delta) = \max\{|F|, p + q - 1, p + \frac{1}{2}|T|\}$ if $q > 0$.

We split our proof into two parts depending on the value of q .

Case 1: $q = 0$.

In this case $T = \emptyset$, so by Lemma 4 for any semi-solution A , every vertex in $G_S(A)$ must have even degree in $G_S(A)$. In other words, every vertex of G must be incident to an even number of edges in A . Since $T = \emptyset$, we assumed above that G was disconnected, so $p \geq 2$ and any solution A must be non-empty. This means that $G_S(A)$ must contain a cycle, so $opt_S(G, \delta) \geq 3$. Recall that $G_S(A)$ is a subgraph of \overline{G} .

Suppose $p = 2$. If $G = K_1 + K_t$ for $t \geq 2$ then $\overline{G} = K_{1,t}$, which does not contain a cycle. Therefore (G, δ) is a no-instance in this case. If $G = K_s + K_t$ for $s, t \geq 2$ then $\overline{G} = K_{s,t}$, which contains no cycles of length 3. Therefore $opt_S(G, \delta) \geq 4$ in this case. Indeed, if u, v are vertices in the K_s component of G and u', v' are vertices in the K_t component, then $A = \{uu', u'v, vv', v'u\}$ is a solution of size 4 and this solution must therefore be optimal. Finally, suppose G contains exactly two components, at least one of which is not a clique. Let x, y be non-adjacent vertices in this component and let z be a vertex in the other component. Then $A = \{xy, yz, zx\}$ is a solution of size 3, which must therefore be optimal.

Finally, suppose that $p \geq 3$. Since $G + A$ must be connected for any solution A , every component in G must contain at least one vertex incident to an edge of A . By Lemma 4, this vertex must be incident to an even number of edges of A , meaning that it must be incident to at least two such edges. Therefore $\text{opt}_S(G, \delta) \geq p$. Indeed, if we choose vertices v_1, \dots, v_p , one from each component of G then $A = \{v_1v_2, v_2v_3, \dots, v_{p-1}v_p, v_pv_1\}$ is a solution of size p , which is therefore optimal.

This concludes the $q = 0$ case.

Case 2: $q > 0$.

In this case $T \neq \emptyset$. We first show that $\text{opt}_S(G, \delta) \geq \max\{|F|, p + q - 1, p + \frac{1}{2}|T|\}$. Since F is a minimum T -join in G_S , Lemma 4 implies that $\text{opt}_S(G, \delta) \geq |F|$. Since G has $p + q$ components, any solution A must contain at least $p + q - 1$ edges to ensure that $G + A$ is connected, so $\text{opt}_S(G, \delta) \geq p + q - 1$. Finally, let G_1, \dots, G_p be the components of G that do not contain any vertices of T . If A is a solution then every component G_i must contain a vertex incident to some edge in A . By Lemma 4, this vertex must be incident to an even number of edges of A , meaning that it must be incident to at least two such edges. By Lemma 4, every vertex of T must be incident to some edge in A . Therefore A must contain at least $p + \frac{1}{2}|T|$ edges, so $\text{opt}_S(G, \delta) \geq p + \frac{1}{2}|T|$.

Next we show that we can always construct a solution of size $\max\{|F|, p + q - 1, p + \frac{1}{2}|T|\}$. To do this, we try to replace edges of F in such a way that F remains a minimum T -join in G_S , but the number of components in $G + F$ is reduced. After we have finished this process, if $G + F$ is connected then setting $A = F$ gives a solution of size $|F|$, which is therefore optimal. Otherwise, we will be able to use the structure of F to construct a solution of size either $p + q - 1$ or $p + \frac{1}{2}|T|$.

Consider the graph $G_S(F)$. Since F is a minimum T -join, $G_S(F)$ cannot contain any cycles (otherwise the edges in the cycle could be removed from F to give a smaller T -join). We prove the following claim.

Claim 1. The graph $G_S(F)$ is a forest that contains no path of length at least 3 (in other words $G_S(F)$ is a forest of stars).

Suppose, for contradiction, that there is such a path with edge set P and end-vertices u and v . Note that u and v are in the same component of $G + F$. Since $G + F$ is not connected (otherwise $A = F$ would be an optimal solution of size $|F|$), there must be a vertex $x \in V(G)$ which is in a different component of $G + F$ from the one containing u and v . In this case $ux, xv \in E(G_S)$. Let $F' = (F \setminus P) \cup \{ux, xv\}$. Then F' is also a T -join in G_S , since the degree parity of any vertex in $G + F'$ is the same

as its degree parity in $G + F$. However, $|F'| < |F|$, which contradicts the fact that F is a minimum T -join. This proves Claim 1.

Now suppose that u, v, u', v' are four distinct vertices in F with $uv, u'v' \in F$, such that uv is not a bridge in $G + F$ and the vertices u and u' are in different components of $G + F$. Let $F' = (F \setminus \{uv, u'v'\}) \cup \{u'v, uv'\}$. Then F' is also a minimum T -join in G_S . However, $G + F'$ has one component less than $G + F$. Indeed, since uv is not a bridge in $G + F$, the vertices u, u', v, v' must all be in the same component of $G + F'$. Therefore, if such edges $uv, u'v' \in F$ exist, we replace F by F' . We do this exhaustively until no further such pairs of edges exist. At this point either every edge in F must be a bridge or every edge in F is in the same component of $G + F$. We consider these possibilities separately.

First suppose that every edge in F is a bridge. Choose $uv \in F$ and let G_1, \dots, G_k be the components of $G + F$, with $u, v \in V(G_1)$. Note that since every edge in F is a bridge, $k = p + q - |F|$. Now let $v_i \in V(G_i)$ for $i \in \{2, \dots, k\}$. Let $A = F$ if $k = 1$ and $A = (F \setminus \{uv\}) \cup \{uv_2, v_2v_3, \dots, v_{k-1}v_k, v_kv\}$ otherwise. Now every vertex in $G + A$ has the same degree parity as in $G + F$, so A is a T -join in G_S . The graph $G + A$ is connected, so A is a solution. However, $|A| = |F| - 1 + p + q - |F| = p + q - 1$. Therefore A is an optimal solution.

We may now assume that every edge in F is in the same component of $G + F$. If $G + F$ is connected, then $A = F$ is a solution of size $|F|$ and is therefore optimal, so we may assume that $G + F$ is not connected.

Suppose $uv, vw \in F$. Then $uw \in E(G)$, as otherwise we could replace uv, vw in F by uw to get a smaller T -join in G_S . Suppose that uv, vw do not form an edge cut-set in $G + F$. In other words, we suppose that u and v are in the same component of $G + (F \setminus \{uv, vw\})$. Let x be a vertex in a different component of $G + F$ from the one containing u, v, w . Then $ux, xw \in E(G_S)$. Let $F' = (F \setminus \{uv, vw\}) \cup \{ux, xw\}$. Then F' must also be a minimum T -join in G_S . However, $G + F'$ has one less component than $G + F$. Indeed, x is in the same component of $G + F'$ as u, v, w . In this case we may replace F by F' . Again, we apply this replacement exhaustively until it can no longer be applied. This process ends when either $G + F$ becomes connected (in which case $A = F$ is an optimal solution of size $|F|$) or, for every pair of edges of the form $uv, vw \in F$, we find that $\{uv, vw\}$ is an edge cut-set in $G + F$. We may assume the latter is the case.

We will prove the following claim.

Claim 2. Let $uv, vw \in F$. Then the component C of $G + (F \setminus \{uv, vw\})$ that contains v contains no vertices of T . Moreover, $d_{G_S(F)}(v) = 2$ and v is the unique vertex of $G_S(F)$ in C .

We prove Claim 2 as follows. We first show that C contains no vertices of T . Suppose, for contradiction, that $x \in T \cap V(C)$ (x is not necessarily distinct from v). Then by Lemma 4, x must be the end-vertex of some edge in $F \setminus \{uv, vw\}$, say xy (again y is not necessarily distinct from v). Note that x and y are in the same component of $G + (F \setminus \{uv, vw\})$, which is different from the component containing u and w . Let $F' = (F \setminus \{xy, uv, vw\}) \cup \{ux, wy\}$. Then F' is also a T -join in G_S , but $|F'| = |F| - 1$, contradicting the minimality of F . Hence, C contains no vertices of T .

By Claim 1 and the definition of a T -join, every vertex of $G_S(F)$ that is not in T must have a neighbour in T . In particular, this means that in the graph $G + (F \setminus \{uv, vw\})$, no vertex of $V(C) \setminus \{v\}$ has a neighbour in T and the vertex v has no neighbours in $T \setminus \{u, w\}$, otherwise in both cases such a neighbour would be in C , a contradiction. This completes the proof of Claim 2.

Recall, that by Claim 1, $G_S(F)$ is a forest in which each component is a star. Then, by Claim 2, each component of $G_S(F)$ is in fact a path of length 1 or 2. Hence, $G_S(F)$ consists of $\frac{1}{2}|T|$ vertex-disjoint paths of length at most 2 with their ends in T . Since $G_S(F)$ has $|F|$ edges, $G_S(F)$ consists of $|T| - |F|$ paths of length 1 and $|F| - \frac{1}{2}|T|$ paths of length 2. By Claim 2, the middle vertex of every path of length 2 lies in a different component of those p components of G that do not contain any vertices of T . Let G_0, G_1, \dots, G_k be the components of $G + F$ such that G_0 is the only component containing vertices of T . Note that $k = p - (|F| - \frac{1}{2}|T|)$. Let $v_i \in V(G_i)$ for $i \in \{1, \dots, k\}$. Choose $uv \in F$ and let $A = (F \setminus \{uv\}) \cup \{uv_1, v_1v_2, \dots, v_{k-1}v_k, v_kv\}$. Then every vertex in $G + A$ has the same degree parity as in $G + F$ and the graph $G + A$ is connected, so A is a solution. Furthermore, $|A| = |F| + p - (|F| - \frac{1}{2}|T|) = p + \frac{1}{2}|T|$, so A is an optimal solution. This concludes the proof of Case 2.

Note that p and q can be computed and T can be found in $O(n + m)$ time. Recall that a minimum T -join in G_S can be found in $O(n^3)$ time by Lemma 2, so the value of $opt_S(G, \delta)$ can be computed in $O(n^3)$ time. Note that the constructive proofs for Cases 1 and 2 can be turned into algorithms. For Case 1, if $p = 2$, then we can check in $O(n + m)$ time whether or not G is the disjoint union of two cliques and either return a no-answer or find a solution. If $p \geq 3$, then we can find a solution in $O(n + m)$ time. For Case 2, observe that for a given F , we can find the

components and the bridges of $G + F$ in $O(n^2)$ time. Hence, as $G + F$ has at most n components, glueing these components of $G + F$ via non-bridge edges $uv \in F$ can be done in $O(n^3)$ time. If every edge of F is a bridge of $G + F$, then the rest of the construction of a solution can be done in $O(n)$ time. Otherwise we do as follows. We partition the edges of $G_S(F)$ into paths of length 1 and paths of length 2 with their ends in T . This can be done in $O(n)$ time. The replacements of paths of length 2 consisting of two edges uv, vw that do not form an edge cut-set of $G + F$ by pairs of edges ux, xw that do can be done in $O(n^3)$ time. After this, we can find an optimal solution in $O(n)$ time. We conclude that an optimal solution A can be found in $O(n^3)$ total time. \square

We are now ready to present the main result of this section. Proving this result requires significantly different arguments than the ones used in the proof of Theorem 1. Let $S = \{\text{ea}, \text{ed}\}$ and let (G, δ) be an instance of $\text{CDPE}(S)$. If F is a T -join in $G_S = K_n$, let $D = F \cap E(G)$ and $A = F \setminus D$. Then by Lemma 4, (A, D) is a semi-solution. Note that if F is a minimum T -join in G_S then it is a matching in which every vertex of T is incident to precisely one edge of F , so $|F| = \frac{1}{2}|T|$. We will show how this allows us to calculate $\text{opt}_S(G, \delta)$ directly from the structure of G , without having to find a T -join. Note that there is no connected graph on exactly two vertices in which both vertices have odd degree. Similarly, no graph can contain an odd number of vertices of odd degree. This means that not every instance of $\text{CDPE}(S)$ is a yes-instance. However, we will show that all no-instances are trivial, that is, they occur when $|T|$ is odd or G contains only two vertices.

Theorem 2. *Let $S = \{\text{ea}, \text{ed}\}$. Then $\text{CDPE}(S)$ can be solved in $O(n+m)$ time and an optimal solution (if one exists) can be found in $O(n^3)$ time.*

Proof. Let $S = \{\text{ea}, \text{ed}\}$ and let (G, δ) be an instance of $\text{CDPE}(S)$. By Lemma 4, we may assume that $|T|$ is even, otherwise (G, δ) is a no-instance. If $G = K_2$ and $T = V(G)$, or $G = K_1 + K_1$ and $T = \emptyset$, then (G, δ) is a no-instance. If $G = K_2$ and $T = \emptyset$ then, trivially, $\text{opt}_S(G, \delta) = 0$, and if $G = K_1 + K_1$ and $T = V(G)$ then $\text{opt}_S(G, \delta) = 1$. To avoid these trivial instances, we therefore assume that G contains at least three vertices. Under these assumptions we will show that $\text{opt}_S(G, \delta)$ is always finite and give exact formulas for the value of $\text{opt}_S(G, \delta)$. Let p be the number of components of G that do not contain any vertex of T and let q be the number of components of G that contain at least one vertex of T . We prove the following series of statements.

- $opt_S(G, \delta) = 0$ if $p = 1, q = 0$,
- $opt_S(G, \delta) = \max\{3, p\}$ if $p \geq 2, q = 0$,
- $opt_S(G, \delta) = \frac{1}{2}|T| + 1$ if $p = 0, q = 1, G[T] = K_{1,r}$, for some $r \geq 1$, and each edge of $G[T]$ is a bridge of G ,
- $opt_S(G, \delta) = \max\{p + q - 1, p + \frac{1}{2}|T|\}$ in all other cases.

Note that if $p = 1, q = 0$, then the first statement applies and the trivial solution $(A, D) = (\emptyset, \emptyset)$ is optimal. We now consider the remaining three cases separately.

Case 1: $p \geq 2$ and $q = 0$.

Then $T = \emptyset$, so by Lemma 4 for any semi-solution (A, D) , every vertex in $G_S(A \cup D)$ must have even degree in $G_S(A \cup D)$. In other words, every vertex of G must be incident to an even number of edges in $A \cup D$. Since $p \geq 2$, the graph G is disconnected, so any solution (A, D) is non-empty. This means that $G_S(A \cup D)$ must contain a cycle, so $opt_S(G, \delta) \geq 3$ if a solution exists. Suppose $p = 2$. As G has at least three vertices, it contains a component containing an edge xy . Let z be a vertex in its other component. We set $A = \{xz, yz\}$ and $D = \{xy\}$ to obtain a solution for (G, δ) . Since $|A| + |D| = 3$, this solution is optimal. Suppose $p \geq 3$. Since $G + A - D$ must be connected for any solution (A, D) , every component in G must contain at least one vertex incident to an edge of A . By Lemma 4, this vertex must be incident to an even number of edges of $A \cup D$, meaning that it must be incident to at least two such edges. Therefore $opt_S(G, \delta) \geq p$. Indeed, if we choose vertices v_1, \dots, v_p , one from each component of G , then setting $A = \{v_1v_2, v_2v_3, \dots, v_{p-1}v_p, v_pv_1\}$ and $D = \emptyset$ gives a solution of size p , which is therefore optimal. This concludes Case 1.

Case 2: $p = 0, q = 1, G[T] = K_{1,r}$ for some $r \geq 1$ and each edge of $G[T]$ is a bridge of G .

Then G is connected. Let v_0 be the central vertex of the star $G[T]$ and let v_1, \dots, v_r be the leaves. By Lemma 4, in any semi-solution (A, D) , every vertex of T must be incident to an odd number of edges in $A \cup D$, so $opt_S(G, \delta) \geq \frac{1}{2}|T|$.

We claim that $opt_S(G, \delta) \geq \frac{1}{2}|T| + 1$. Suppose, for contradiction, that (A, D) is a semi-solution of size $|A| + |D| = \frac{1}{2}|T|$. Then $A \cup D$ must be a matching with each edge joining a pair of vertices of T . However, then $v_0v_i \in A \cup D$ for some i . Since $v_0v_i \in E(G)$, we must have $v_0v_i \in D$. However, since v_0v_i is a bridge of G , v_0 and v_i must then be in different components of $G + A - D$, so $G + A - D$ is not connected and (A, D) is not a solution, a contradiction. Therefore $opt_S(G, \delta) \geq \frac{1}{2}|T| + 1$.

Next we show how to find a solution of size $\frac{1}{2}|T| + 1$. Since $|T|$ is even, r must be odd. First suppose that $r = 1$. Since G is connected and v_0v_1 is a bridge, $G \setminus \{v_0v_1\}$ has exactly two components. Since G contains at least three vertices, one of these components contains another vertex x . Without loss of generality assume $xv_0 \in E(G)$, in which case $xv_1 \notin E(G)$. Then setting $A = \{xv_1\}$ and $D = \{xv_0\}$ gives a semi-solution. Since x, v_0, v_1 are all in the same component of $G + A - D$, the graph $G + A - D$ must be connected, so (A, D) is a solution. Since $|A| + |D| = 2 = \frac{1}{2}|T| + 1$, this solution is optimal. Now suppose $r \geq 3$. Let $A = \{v_1v_2, v_2v_3\} \cup \{v_{2i}v_{2i+1} \mid 2 \leq i \leq \frac{1}{2}(r-1)\}$ and $D = \{v_0v_2\}$. Then (A, D) is a semi-solution and since v_0, \dots, v_r are all in the same component of $G + A - D$, we find that (A, D) is a solution. Since $|A| + |D| = 2 + \frac{1}{2}(r-1) - 1 + 1 = \frac{1}{2}|T| + 1$, this solution is optimal. This concludes Case 2.

Case 3: $q \geq 1$ and Case 2 does not hold.

Then $T \neq \emptyset$. Let G_1, \dots, G_p be the components of G without vertices of T and let $G' = G - (V(G_1) \cup \dots \cup V(G_p))$. Note that $G' = G$ if $p = 0$ and that G' is not the empty graph, as $q > 0$. Choose $v_i \in V(G_i)$ for $i \in \{1, \dots, p\}$.

We first show that $\text{opt}_S(G, \delta) \geq \max\{p+q-1, p+\frac{1}{2}|T|\}$. Since G has $p+q$ components, any solution (A, D) must contain at least $p+q-1$ edges in A to ensure that $G + A - D$ is connected, so $\text{opt}_S(G, \delta) \geq p+q-1$. If (A, D) is a solution then every component G_i must contain a vertex incident to some edge in A . By Lemma 4, this vertex must be incident to an even number of edges of $A \cup D$, meaning that it must be incident to at least two such edges. By Lemma 4, every vertex of T must be incident to some edge in $A \cup D$. Therefore $A \cup D$ must contain at least $p + \frac{1}{2}|T|$ edges, so $\text{opt}_S(G, \delta) \geq p + \frac{1}{2}|T|$.

We now show how to find a solution of size $\max\{p+q-1, p+\frac{1}{2}|T|\}$. We start by finding a maximum matching M in $\overline{G[T]}$. Let U be the set of vertices in T that are not incident to any edge in M . We divide the argument into two cases, depending on the size of U .

Case 3a: $U = \emptyset$.

In this case, by Lemma 4, setting $A = M$ and $D = \emptyset$ gives a semi-solution. Now suppose that $uv, u'v' \in M$, such that uv is not a bridge in $G + M$ and the vertices u and u' are in different components of $G + M$. Let $M' = (M \setminus \{uv, u'v'\}) \cup \{u'v, uv'\}$. Then M' is also a maximum matching in $\overline{G[T]}$. However, $G + M'$ has one component less than $G + M$. Indeed, since uv is not a bridge in $G + M$, the vertices u, u', v, v' must all

be in the same component of $G + M'$. Therefore, if such edges $uv, u'v' \in M$ exist, we replace M by M' . We do this exhaustively until no further such pairs of edges exist. At this point either every edge in M is a bridge in $G + M$ or every edge in M is in the same component of $G + M$. We consider these possibilities separately.

First suppose that every edge in M is a bridge in $G + M$. Choose $uv \in M$ and let Q_1, \dots, Q_k be the components of $G + M$, with $u, v \in V(Q_1)$. Note that since every edge in M is a bridge, $k = p + q - |M|$. Now let $x_i \in V(Q_i)$ for $i \in \{2, \dots, k\}$. Let $D = \emptyset$ and let $A = M$ if $k = 1$ and $A = (M \setminus \{uv\}) \cup \{ux_2, x_2x_3, \dots, x_{k-1}x_k, x_kv\}$ otherwise. Now every vertex in $G + A - D$ has the same degree parity as in $G + M$, so (A, D) is a semi-solution by Lemma 4. The graph $G + A - D$ is connected, so (A, D) is a solution. As $|A| + |D| = |M| - 1 + p + q - |M| + 0 = p + q - 1$, we find that (A, D) is an optimal solution.

Now suppose that every edge in M is in the same component of $G + M$. Note that G_1, \dots, G_p are the remaining components of $G + M$. Choose $uv \in M$. Let $D = \emptyset$ and let $A = M$ if $p = 0$ and $A = (M \setminus \{uv\}) \cup \{uv_1, v_1v_2, \dots, v_{p-1}v_p, v_pv\}$ otherwise. Then every vertex in $G + A - D$ has the same parity as in $G + M$ and $G + A - D$ is connected, so by Lemma 4 (A, D) is a solution. Since $|A| + |D| = \frac{1}{2}|T| - 1 + p + 1 = p + \frac{1}{2}|T|$, this solution is optimal. This concludes Case 3a.

Case 3b: $U \neq \emptyset$.

Note that $z = |U|$ must be even since $|T|$ is even. Every pair of vertices in U must be non-adjacent in \overline{G} , as otherwise M would not be maximum. Therefore $G[U]$ is a clique. Let $U = \{u_1, \dots, u_z\}$.

Recall that $G' = G - (V(G_1) \cup \dots \cup V(G_p))$. We claim that $Q = G' + M$ is connected. Clearly every vertex of the clique U must be in the same component of $Q = G' + M$. Suppose for contradiction that Q_1 is a component of Q that does not contain U . Then Q_1 must contain some edge $w_1w_2 \in M$. However, in this case $M' = (M \setminus \{w_1w_2\}) \cup \{u_1w_1, u_2w_2\}$ is a larger matching in $\overline{G[T]}$ than M , which contradicts the maximality of M . Therefore Q is connected.

First suppose that $z \geq 4$ (recall that z is even). Let $M' = \{u_1u_2, u_3u_4, \dots, u_{z-1}u_z\}$. Since U is a clique, $G' + M - M'$ is connected. If $p = 0$ set $A = M$ and $D = M'$. If $p > 0$ set $A = M \cup \{u_1v_1, v_1v_2, \dots, v_{p-1}v_p, v_pv_2\}$ and $D = M' \setminus \{u_1u_2\}$. Then $G + A - D$ is connected, so (A, D) is a solution by Lemma 4. This solution has size $|A| + |D| = p + \frac{1}{2}|T|$, so it is optimal.

Now suppose that $z \leq 3$. Then $z = 2$. If $p > 0$, let $A = M \cup \{u_1v_1, v_1v_2, \dots, v_{p-1}v_p, v_pv_2\}$ and $D = \emptyset$. Then $G + A - D$ is connected, so (A, D) is a solution by Lemma 4. This solution has size $|A| + |D| = p + \frac{1}{2}|T|$, so

it is optimal. Assume that $p = 0$, so $G + M$ contains only one component. If u_1u_2 is not a bridge in $G + M$, let $A = M$ and $D = \{u_1u_2\}$. Then $G + M$ is connected, so (A, D) is a solution. This solution has size $|A| + |D| = p + \frac{1}{2}|T|$, so it is optimal.

Now assume that u_1u_2 is a bridge in $Q = G + M$. Let Q_1 and Q_2 denote the components of $Q - \{u_1u_2\}$ with $u_1 \in V(Q_1)$ and $u_2 \in V(Q_2)$. Note that u_1u_2 is also a bridge in G . We claim that the edges of M are either all in Q_1 or all in Q_2 . Suppose for contradiction that $y_1z_1 \in E(Q_1) \cap M$ and $y_2z_2 \in E(Q_2) \cap M$. Then $M' = (M \setminus \{y_1z_1, y_2z_2\}) \cup \{u_1y_2, u_2y_1, z_1z_2\}$ would be a larger matching in $\overline{G[T]}$ than M , contradicting the maximality of M . Without loss of generality, we may therefore assume that all edges of M are in Q_1 .

Let $M = \{x_1y_1, \dots, x_ry_r\}$, where $r = \frac{1}{2}|T| - 1$. We claim that u_1 must be adjacent in G to all vertices of $T \setminus \{u_1\}$. Suppose for contradiction that u_1 is non-adjacent in G to some vertex of $T \setminus \{u_1\}$. Since $u_1u_2 \in E(G)$, this vertex would have to be incident to some edge in M . Without loss of generality, assume $u_1x_1 \notin E(G)$. Then $M' = (M \setminus \{x_1y_1\}) \cup \{u_1x_1, u_2y_1\}$ would be a larger matching in $\overline{G[T]}$ than M , contradicting the maximality of M . Therefore u_1 is adjacent in G to every vertex of $T \setminus \{u_1\}$. In particular, since $p = 0$, it follows that $q = 1$ and G is connected.

Suppose that every edge between u_1 and $T \setminus \{u_1\}$ is a bridge in G . Then no two vertices of $T \setminus \{u_1\}$ can be adjacent, and $G[T] = K_{1,r}$. However, then Case 2 applies, which we assumed was not the case. Without loss of generality, we may therefore assume that u_1x_1 is not a bridge in G . Let $A = (M \setminus \{x_1y_1\}) \cup \{y_1u_2\}$ and $D = \{u_1x_1\}$. Then $G + A - D$ is connected, so (A, D) is a solution. Since $|A| + |D| = \frac{1}{2}|T| - 1 - 1 + 1 + 1 = p + \frac{1}{2}|T|$, this solution is optimal. This concludes Case 3b and therefore also concludes Case 3.

It is clear that $opt_S(G, \delta)$ can be computed in $O(n + m)$ time. We also observe that the above proof is constructive. Hence, we not only solve the decision variant of CDPE(ea, ed) but we can also find an optimal solution in polynomial time. To do so, we first observe that we can check in $O(n + m)$ time whether Case 1, 2, or 3 applies. Moreover, if we are in Case 1 or 2 then we can also construct an optimal solution in $O(n + m)$ time. Suppose that we are in Case 3. Then we must find a maximum matching in $\overline{G[T]}$. This takes $O(n^{5/2})$ time by Lemma 1. However, the bottleneck is in Case 3a, where we are glueing components by replacing two matching edges by two other matching edges, which takes $O(n^2)$ time. As the total number of times we may need to do this is $O(n)$, this pro-

cedure may take $O(n^3)$ time in total. Hence, we can obtain an optimal solution in $O(n^3)$ time. \square

3.2 The W[1]-Hard Cases

We first describe the problem used in our W[1]-hardness construction. A *red/blue graph* is a bipartite graph $G = (\mathcal{R}, \mathcal{B}, E)$ whose vertices are partitioned into independent sets \mathcal{R} (the red vertices) and \mathcal{B} (the blue vertices). A set $R \subseteq \mathcal{R}$ is an *odd set* if every vertex in \mathcal{B} has an odd number of neighbours in R . The ODD SET problem takes as input a red/blue graph $G = (\mathcal{R}, \mathcal{B}, E)$ and an integer $k > 0$, and asks whether there is an odd set $R \subseteq \mathcal{R}$ of size at most k . This problem is known to be NP-complete as well as W[1]-hard when parameterized by k [11]. We are now ready to prove the hardness results of this section.

Theorem 3. *Let $\{\text{vd}\} \subseteq S \subseteq \{\text{vd}, \text{ed}, \text{ea}\}$. Then CDPE(S) is NP-complete and W[1]-hard when parameterized by k , even if $\delta \equiv 0$.*

Proof. The CDPE(S) problem clearly belongs to NP. To prove that the problem is NP-complete and W[1]-hard when parameterized by k , even if $\delta \equiv 0$, we reduce from ODD SET. Recall that ODD SET is NP-complete as well as W[1]-hard when parameterized by k [11], and this clearly remains true when we assume that every vertex in \mathcal{R} has at least one neighbour in \mathcal{B} .

Let (G, k) be an instance of ODD SET, where $G = (\mathcal{R}, \mathcal{B}, E)$ is a red/blue graph with $\mathcal{R} = \{r_1, \dots, r_p\}$ and $\mathcal{B} = \{b_1, \dots, b_q\}$, and where every vertex in \mathcal{R} has at least one neighbour in \mathcal{B} . We construct a graph G^* from G as follows:

- Keep the set $\mathcal{R} = \{r_1, \dots, r_p\}$.
- Replace \mathcal{B} by $2(k+1)$ copies of \mathcal{B} denoted by $\mathcal{B}_1, \dots, \mathcal{B}_{2(k+1)}$, and for $i \in \{1, \dots, 2(k+1)\}$, let $\mathcal{B}_i = \{b_1^i, \dots, b_q^i\}$, where b_h^i is a copy of b_h for $h \in \{1, \dots, q\}$.
- For $i \in \{1, \dots, 2(k+1)\}$, $j \in \{1, \dots, p\}$ and $h \in \{1, \dots, q\}$, add the edge $r_j b_h^i$ if and only if $r_j b_h \in E(G)$.
- For $h \in \{1, \dots, q\}$, if $d_G(b_h)$ is odd then add the edge $b_h^{2i-1} b_h^{2i}$ for $i \in \{1, \dots, k+1\}$.
- Introduce $2k$ new vertices $x_1, \dots, x_k, y_1, \dots, y_k$ and add the edge $x_i y_i$ for $i \in \{1, \dots, k\}$.
- Add a new vertex z and make it adjacent to all vertices of $\mathcal{B}^* = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_{2(k+1)}$ and to all vertices $x_1, \dots, x_k, y_1, \dots, y_k$.

This completes the construction of G^* . We define a parity function $\delta : V(G^*) \rightarrow \{0, 1\}$ by setting $\delta(v) = 0$ for every $v \in V(G^*)$.

We will show that (G^*, k, δ) is a yes-instance of CDPE(S) if and only if (G, k) is a yes-instance of ODD SET. We first observe that in G^* , each vertex of $\mathcal{R} \cup \{x_1, \dots, x_k\} \cup \{y_1, \dots, y_k\} \cup \{z\}$ has even degree and that each vertex of \mathcal{B}^* has odd degree.

Suppose that (G, k) is a yes-instance of ODD SET. Then there exists an odd set $R \subseteq \mathcal{R}$ in G such that $|R| \leq k$. Note that R is also a subset of the vertices of G^* . Let $H = G^* - R$. Because R is an odd set, the vertices of \mathcal{B}^* have even degree in H . As the degree of the other vertices of H is the same as in G^* , all the vertices of H have even degree. Since every vertex of \mathcal{R} has at least one neighbour in each $\mathcal{B}_1, \dots, \mathcal{B}_{2(k+1)}$ and z is adjacent to all the vertices of \mathcal{B}^* , H is connected. We conclude that (G^*, k, δ) is a yes-instance of CDPE(S).

Now suppose that (G^*, k, δ) is a yes-instance of CDPE(S). Then there is a sequence L of at most k operations from S transforming G^* into an Eulerian graph H . Let R be the set of vertices of \mathcal{R} that are deleted by L . Clearly, $|R| \leq k$. We will show that R is an odd set in G .

We claim that z cannot be deleted by an operation from L . To obtain a contradiction, suppose that z is deleted. Then the obtained graph has at least $k + 1$ components. Notice that it is impossible to obtain a connected graph by the remaining at most $k - 1$ operations, as each operation can reduce the number of components by at most one.

Because L contains at most k operations, it follows that there is an index $i \in \{1, \dots, 2(k + 1)\}$ such that

- (i) no vertex of \mathcal{B}_i is deleted by an operation of L ,
- (ii) no edge incident to a vertex of \mathcal{B}_i is deleted,
- (iii) no edge incident to a vertex of \mathcal{B}_i is added.

Recall that the vertices of \mathcal{B}_i have odd degree in G^* . Since performing the operations of L causes these vertices to have even degree, but the vertex z is not deleted, it follows that deleting the vertices of R makes every vertex of \mathcal{B}_i have even degree. Therefore, R is an odd set in G . \square

4 Connected Degree Balance Editing

Let $S \subseteq \{\text{ea}, \text{ed}, \text{vd}\}$. In Section 4.1 we will show that CDBE(S) is polynomial-time solvable if $\{\text{ea}\} \subseteq S \subseteq \{\text{ea}, \text{ed}\}$ and in Section 4.2 we will show that it is NP-complete and W[1]-hard with parameter k if $\text{vd} \in S$.

4.1 The Polynomial-Time Solvable Cases

Let $\{\mathbf{ea}\} \subseteq S \subseteq \{\mathbf{ea}, \mathbf{ed}\}$. Let (G, δ, k) be an instance of $\text{CDBE}(S)$ with $G = (V, E)$. Let A be a set of arcs not in G , and let D be a set of arcs in G , with $D = \emptyset$ if $S = \{\mathbf{ea}\}$. We say that (A, D) is a *solution* for (G, δ, k) if its *size* $|A| + |D| \leq k$, the equation $d_H^{\text{out}}(u) - d_H^{\text{in}}(u) = \delta(u)$ holds for every vertex u and the graph $H = G + A - D$ is connected; if we drop the condition that H is connected then (A, D) is a *semi-solution* for (G, δ, k) . Just as in Section 3.1 we consider the optimization version for $\text{CDBE}(S)$ and we use the same terminology.

Let (G, δ) be an instance of (the optimization version) of $\text{CDBE}(S)$ where $G = (V, E)$. Let $T = T_{(G, \delta)}$ be the set of vertices v such that $d_G^{\text{out}}(v) - d_G^{\text{in}}(v) \neq \delta(v)$. Define a function $f_{(G, \delta)} : T \rightarrow \mathbb{Z}$ by $f(v) = f_{(G, \delta)}(v) = \delta(v) - d_G^{\text{out}}(v) + d_G^{\text{in}}(v)$ for every $v \in T$.

We construct a directed multigraph G_S with vertex set V and arc set determined as follows. If $\{\mathbf{ea}\} \subseteq S \subseteq \{\mathbf{ea}, \mathbf{ed}\}$, for each pair of distinct vertices u and v in G , if $(u, v) \notin E$, add the arc (u, v) to G_S (these arcs are precisely those that can be added to G). If $S = \{\mathbf{ea}, \mathbf{ed}\}$, for each pair of distinct vertices u and v , if $(u, v) \in E$, add the arc (v, u) to G_S (these arcs are precisely those whose reverse can be deleted from G). Note that adding a (missing) arc has the same effect on the degree balance of the vertices in a digraph as deleting the reverse of the arc (if it exists). Also observe that G_S becomes a directed multigraph rather than a digraph only if $S = \{\mathbf{ea}, \mathbf{ed}\}$ and there are distinct vertices u and v such that $(u, v) \in E$ and $(v, u) \notin E$ applies. Moreover, G_S contains at most two copies of any arc, and if there are two copies of (u, v) then (v, u) is not in G_S .

Let F be a minimum directed f -join in G_S (if one exists). Note that F may contain two copies of the same arc if G_S is a directed multigraph. Also note that for any pair of vertices u, v , either $(u, v) \notin F$ or $(v, u) \notin F$, otherwise $F' = F \setminus \{(u, v), (v, u)\}$ would be a smaller f -join in G_S , contradicting the minimality of F .

We define two sets A_F and D_F which, as we will show, correspond to a semi-solution (A_F, D_F) of (G, δ) . Initially set $A_F = D_F = \emptyset$. Consider the arcs in F . If F contains (u, v) exactly once then add (u, v) to A_F if $(u, v) \notin E$ and add (v, u) to D_F if $(u, v) \in E$ (in this case $(v, u) \in E$ holds). If F contains two copies of (u, v) then add (u, v) to A_F and (v, u) to D_F ; note that by definition of F and G_S , in this case $S = \{\mathbf{ea}, \mathbf{ed}\}$, $(u, v) \notin E$ and $(v, u) \in E$. Observe that the sets A_F and D_F are not multisets. We need the following lemma, which consists of seven easy observations.

Lemma 5. *Let $\{\mathbf{ea}\} \subseteq S \subseteq \{\mathbf{ea}, \mathbf{ed}\}$. Let (G, δ) be an instance of $\text{CDBE}(S)$ where $G = (V, E)$. Let F be a minimum directed f -join. The following statements hold.*

- (i) *If $(u, v) \in A_F$ then $(u, v) \notin E$.*
- (ii) *If $(u, v) \in D_F$ then $(u, v) \in E$.*
- (iii) *$A_F \cap D_F = \emptyset$ and moreover, $(u, v) \in F$ if and only if $(u, v) \in A_F$ or $(v, u) \in D_F$.*
- (iv) *There are two copies of (u, v) in F if and only if $(u, v) \in A_F$ and $(v, u) \in D_F$.*
- (v) *If $S = \{\mathbf{ea}\}$, then $D_F = \emptyset$.*
- (vi) *If vertices u and v are joined by an arc in G then they are joined by an arc in $G + A_F - D_F$.*
- (vii) *If $(u, v) \in F$ then u and v are connected by an arc in $G + A_F - D_F$.*

Proof. Statements (i) and (ii) follow directly from the definitions of A_F and D_F , respectively. The fact that $A_F \cap D_F = \emptyset$ follows directly from Statements (i) and (ii). The second part of Statement (iii) follows directly from the definitions of A_F and D_F . Statement (iv) follows directly from the definition of A_F and D_F .

To prove Statement (v), suppose for contradiction that $S = \{\mathbf{ea}\}$ and $(u, v) \in D_F$. By Statement (ii), $(u, v) \in E$. Since $S = \{\mathbf{ea}\}$, F can contain at most one copy of (v, u) . By definition of A_F and D_F , it follows that $(v, u) \in F$ and $(v, u) \in E$. However, since $(u, v), (v, u) \in E$ and $S = \{\mathbf{ea}\}$, (v, u) is not an arc in G_S by definition of G_S . Therefore F cannot be an f -join in G_S , which is a contradiction.

Next we consider Statement (vi). First suppose that $(u, v), (v, u) \in E$. If u and v are not connected by an arc in $G + A_F - D_F$, then $(u, v), (v, u) \in D_F$. Then, by Statement (iii), $(v, u), (u, v) \in F$. However, as stated earlier, this cannot happen, since F is minimum. Now suppose $(u, v) \in E$ and $(v, u) \notin E$. If u and v are not connected by an arc in $G + A_F - D_F$, then $(u, v) \in D_F$. By Statement (iii), $(v, u) \in F$. Since $(v, u) \notin E$, we find that F must contain two copies of (v, u) . Hence $(v, u) \in A_F$. However in this case u and v are connected by an arc in $G + A_F - D_F$. This completes the proof of Statement (vi).

Finally, we consider Statement (vii). Suppose $(u, v) \in F$. If $(u, v) \in A_F$ then by Statement (iii), (u, v) is an arc in $G + A_F - D_F$. Otherwise, by Statement (iii), $(v, u) \in D_F$, so $(v, u) \in E$ by Statement (ii). However, in this case Statement (vi) implies that u and v are connected by an arc in $G + A_F - D_F$. \square

If X and Y are sets, then $X \uplus Y$ is the multiset that consists of one copy of each element that occurs in exactly one of X and Y and two copies of each element that occurs in both.

The next lemma provides the starting point for our algorithm. Recall that $D^R = \{(u, v) \mid (v, u) \in D\}$ denotes the set of arcs whose reverse is in D .

Lemma 6. *Let $\{\text{ea}\} \subseteq S \subseteq \{\text{ea}, \text{ed}\}$. Let (G, δ) be an instance of CDBE(S) where $G = (V, E)$. The following holds:*

- (i) *If F is a minimum directed f -join in G_S , then (A_F, D_F) is a semi-solution for (G, δ) of size $|F|$.*
- (ii) *If (A, D) is a semi-solution for (G, δ) , then $A \uplus D^R$ is a directed f -join in G_S of size $|A| + |D|$.*

Proof. First consider Statement (i). Suppose F is a minimum directed f -join in G_S . By Lemma 5 (iii) and (iv), (A_F, D_F) has size $|A_F| + |D_F| = |F|$.

Let $H = G + A_F - D_F$. Let $u \in V$. Let $A^{\text{out}}(u)$ and $A^{\text{in}}(u)$ be the sets of arcs in F with u as tail or head, respectively, that were put into A_F . Let $D^{\text{out}}(u)$ and $D^{\text{in}}(u)$ be the set of arcs in F with u as tail or head, respectively, whose reverse was put into D_F . Define $d_{G_S(F)}^{\text{out}}(u) = d_{G_S(F)}^{\text{in}}(u) = 0$ if u is not in $G(F)$ and $f(u) = 0$ if $u \notin T$. Then by the definition of a directed f -join, we have

$$\begin{aligned} \delta(u) - (d_G^{\text{out}}(u) - d_G^{\text{in}}(u)) &= f(u) \\ &= d_{G_S(F)}^{\text{out}}(u) - d_{G_S(F)}^{\text{in}}(u) \\ &= |A^{\text{out}}(u)| + |D^{\text{out}}(u)| - |A^{\text{in}}(u)| - |D^{\text{in}}(u)|. \end{aligned}$$

If $(u, v) \in A_F$ then $(u, v) \notin E$ by Lemma 5 (i). If $(u, v) \in D_F$ then $(u, v) \in E$ by Lemma 5 (ii). Moreover, in that case, $(v, u) \in F$. Consequently, we find that

$$\begin{aligned} d_H^{\text{out}}(u) - d_H^{\text{in}}(u) &= d_G^{\text{out}}(u) - d_G^{\text{in}}(u) + |A^{\text{out}}(u)| - |A^{\text{in}}(u)| + |D^{\text{out}}(u)| - |D^{\text{in}}(u)| \\ &= d_G^{\text{out}}(u) - d_G^{\text{in}}(u) + \delta(u) - (d_G^{\text{out}}(u) - d_G^{\text{in}}(u)) \\ &= \delta(u). \end{aligned}$$

We conclude that (A_F, D_F) is a semi-solution for (G, δ) .

Now consider Statement (ii). Suppose (A, D) is a semi-solution for (G, δ) . Let $A^{out}(u)$ and $A^{in}(u)$ be the sets of arcs in A with u as tail or head, respectively. Let $D^{out}(u)$ and $D^{in}(u)$ be the set of arcs in D with u as tail or head, respectively. Let $H = G + A - D$. Let $u \in T$ (recall that T consists of every vertex u with $d_G^{out}(u) - d_G^{in}(u) \neq \delta(u)$). Because (A, D) is a semi-solution, we have

$$\begin{aligned}
d_G^{out}(u) - d_G^{in}(u) + |A^{out}(u)| - |A^{in}(u)| - (|D^{out}(u)| - |D^{in}(u)|) \\
&= d_H^{out}(u) - d_H^{in}(u) \\
&= d_G^{out}(u) - d_G^{in}(u) + \delta(u) - (d_G^{out}(u) - d_G^{in}(u)) \\
&= d_G^{out}(u) - d_G^{in}(u) + f(u),
\end{aligned}$$

where we define $f(u) = 0$ if $u \notin T$. This leads to

$$f(u) = |A^{out}(u)| - |A^{in}(u)| - (|D^{out}(u)| - |D^{in}(u)|).$$

Let $F = A \uplus D^R$. Suppose (u, v) appears once in F . Let $(u, v) \in A$. Then $(u, v) \notin E$. By definition, G_S contains (u, v) . Let $(u, v) \in D^R$. Then $S = \{ea, ed\}$, so $(v, u) \in E$. By definition, G_S contains (u, v) . Suppose (u, v) appears twice in F . Then $(u, v) \in A$ and $(u, v) \in D^R$. Hence, $(u, v) \notin E$ and $(v, u) \in E$, and moreover, $S = \{ea, ed\}$. Then (u, v) appears twice in G_S . We conclude that F is a subset of the arcs in G_S . Let $D^{out}(u)^R$ and $D^{in}(u)^R$ be the set of arcs in D^R with u as tail or head, respectively. Then $|D^{out}(u)^R| = |D^{in}(u)|$ and $|D^{in}(u)^R| = |D^{out}(u)|$. We find that, for all $u \in V$,

$$\begin{aligned}
d_{G_S(F)}^{out}(u) - d_{G_S(F)}^{in}(u) &= |A^{out}(u)| - |A^{in}(u)| + |D^{out}(u)^R| - |D^{in}(u)^R| \\
&= |A^{out}(u)| - |A^{in}(u)| - (|D^{out}(u)| - |D^{in}(u)|) \\
&= f(u).
\end{aligned}$$

Hence, F is a directed f -join. It follows from the corresponding definitions that the size of (A, D) is $|A| + |D| = |A| + |D^R| = |A \uplus D^R| = |F|$. This completes the proof of Lemma 6. \square

Let (G, δ) be an instance of CDBE(S). Let $p = p_{(G, \delta)}$ be the number of components of G that contain no vertex of T . Let $q = q_{(G, \delta)}$ be the number of components of G that contain at least one vertex of T . Let $t = t_{(G, \delta)} = \sum_{u \in T} |f(u)|$.

We now state the following lemma; its proof is based on Lemmas 3, 5 and 6.

Lemma 7. *Let $\{\text{ea}\} \subseteq S \subseteq \{\text{ea}, \text{ed}\}$. Let (G, δ) be an instance of $\text{CDBE}(S)$ with $q \geq 1$. If F is a (given) minimum directed f -join in G_S , then (G, δ) has a solution that has size at most $\max\{|F|, p+q-1, p+\frac{1}{2}t\}$, which can be found in $O(n^3)$ time.*

Proof. Let F be a minimum directed f -join in G_S . If $H = G + A_F - D_F$ is connected, then the statement of the theorem holds by Lemma 6. Suppose H is not connected. We will try to replace arcs in F to obtain a different minimum directed f -join F' such that $H' = G + A_{F'} - D_{F'}$ will have fewer components. Either this will eventually cause the graph to be connected (in which case the corresponding solution will still have size $|F|$), or else the structure of this directed f -join will enable us to find a solution for $\text{CDBE}(S)$ of size either $p+q-1$ or $p+\frac{1}{2}t$. Our changes to F will be such that no additional arcs are ever added to the corresponding set D_F . Thus, if $S = \{\text{ea}\}$, then the property $D_F = \emptyset$ will be preserved.

By Lemma 3, $G_S(F)$ must only consist of mutually arc-disjoint directed paths from vertices u with $f(u) > 0$ to vertices v with $f(v) < 0$. We claim that all such paths must be of length at most 2. Suppose, for contradiction, that there is a directed path of length at least 3 in $G_S(F)$ from some vertex u to some vertex v . Note that u and v are in the same component of H . Since H is not connected, there must be a vertex x in some other component of H . By Lemma 5 (vi), this means that x is not in the same component of G as u or v , so (u, x) and (x, v) are arcs in G_S . Replacing the directed path from u to v in F by the arcs $(u, x), (x, v)$ would yield a smaller directed f -join in G_S , which is a contradiction. Therefore all directed paths in $G_S(F)$ must be of length at most 2.

Let (u, v) and (u', v') be arcs in F . Note that by Lemma 5 (vii), u and v are in the same component of H and u' and v' are in the same component of H . Suppose that (u, v) and (u', v') are chosen such that u and v are in a different component of H from the one containing u' and v' and that one of the following situations holds:

- (i) either $(u, v) \in A_F$ and (u, v) is not a bridge in H , or
- (ii) $(v, u) \in D_F$.

By Lemma 5 (vi), vertex u is not in the same component of G as v' and vertex v is not in the same component of G as u' . Hence, by the definition of G_S , the arcs (u, v') and (u', v) are in G_S . As such, we may replace (u, v) and (u', v') in F by (u, v') and (u', v) . This yields another minimum directed f -join in G_S which, as we explain below, reduces the number of components in H by one. Because u and v are not in the same components of G as u' or v' , adding (u, v) and (u', v') to F means that these two arcs

will be put into A_F . Suppose (i) holds. Then the vertices in the original component of H that contained u and v will still be connected, whereas the vertices in the original component of H that contained u' and v' will still be connected as well (if necessary via a path that uses the new arcs (u, v) and (u', v')). Thus, H has one component less. Suppose (ii) holds. Then removing (v, u) from F means removing it from D_F . Hence, in H , the arc (v, u) is restored and we can apply the same arguments.

We apply the above replacement operation exhaustively. At termination, we have modified F into a minimum directed f -join of G_S , in which either every arc in A_F will be a bridge in H and $D_F = \emptyset$, or the end-vertices of every arc in F will all be in the same component of H . We discuss these two cases separately.

Case 1: *Every arc in A_F is a bridge in H and $D_F = \emptyset$.*

Then $F = A_F$. We claim that every directed path in $G_S(F)$ has length 1. For contradiction, suppose (u, v) and (v, w) are two arcs in F . Since both (u, v) and (v, w) are bridges in H , we must have that (u, w) is not an arc in H . Then replacing (u, v) and (v, w) in F by (u, w) would yield a smaller directed f -join in G_S , which would contradict the minimality of F .

As every directed path in $G_S(F)$ has length 1, every arc $(u, v) \in F$ must be such that $f(u) > 0$ and $f(v) < 0$. Hence, $F = A_F$ contains exactly $\frac{1}{2}t$ arcs.

Let H_1, \dots, H_k be the components of H . Because every arc in A_F is a bridge in H and $D_F = \emptyset$, we find that $k = p + q - \frac{1}{2}t$. Suppose $k = 1$. Then H is connected, so $p = 0$. Hence we have a solution for $\text{CDBE}(S)$ that uses $p + \frac{1}{2}t$ arcs. Suppose $k \geq 2$. Choose an arc $(u, v) \in A_F$ arbitrarily and assume without loss of generality that u and v are in H_1 . Next, choose a vertex v_i in H_i for $i \in \{2, \dots, k\}$. Replace the arc (u, v) in A_F by the arcs $(u, v_2), (v_2, v_3), \dots, (v_{k-1}, v_k), (v_k, v)$. This gives a solution for $\text{CDBE}(S)$ that uses $\frac{1}{2}t - 1 + k = \frac{1}{2}t - 1 + p + q - \frac{1}{2}t = p + q - 1$ arcs.

Case 2: *The end-vertices of each arc in $A_F \cup D_F$ are all in the same component of H .*

Suppose H has at least one other component; let x be a vertex in such a component. Suppose that (u, v) and (v, w) are two distinct arcs in F such that the following situation holds: u and v are in the same component of the graph obtained from H after removing (u, v) and (v, w) . Because F is a minimum directed f -join, u and w are distinct vertices. By Lemma 5 (vi), vertices u and w are not in the component of G that contains x . Hence, by the definition of G_S , the arcs (u, x) and (x, w) are in G_S . As such, we may replace (u, v) and (v, w) in F by (u, x) and (x, w) . This yields another

minimum directed f -join in G_S which, as we explain below, reduces the number of components in H by one.

Because u and w are not in the component of G that contains x , we find that (u, x) and (x, w) will be put into A_F . Because F is a minimum directed f -join, (u, w) must be in H already, so $(u, w) \in E$ or $(u, w) \in F$. By Lemma 5 (vi) and (vii), u and w are still in the same component after our replacement. Consequently, all vertices u, v, w, x will be in the same component. Hence, the number of components in H is reduced by one.

We apply the above replacement operation exhaustively. If H becomes connected, then since F is (still) a minimum directed f -join, we have found a solution of size $|F|$. Assume H does not become connected. Then, at termination of our procedure, we have obtained the following situation. For every two distinct arcs (u, v) and (v, w) , we have that u and v are in different components of the graph H' obtained from H after removing (u, v) and (v, w) . Moreover, w is in the same component of H' as u (by our earlier arguments, we have that $(u, w) \in H$).

Let H'_v be the component of H' that contains v . We claim that $(u, v) \in A_F$ and $(v, w) \in A_F$, and that H'_v contains no vertices incident to arcs in $F \setminus \{(u, v), (v, w)\}$. This can be seen as follows. Because H'_v does not contain u or w , we find that (u, v) and (v, w) are both in A_F due to Lemma 5 (vii). If H'_v contains a vertex incident to some arc in $F \setminus \{(u, v), (v, w)\}$, then this component must also contain the other end-vertex of this arc by Lemma 5 (vii). Suppose u', v' are in H'_v and $(u', v') \in F \setminus \{(u, v), (v, w)\}$. (Note that we do not insist that $u' \neq v$ or $v' \neq v$.) Then we find a smaller directed f -join of G_S by replacing (u, v) , (v, w) and (u', v') in F by the arcs (u, v') and (u', w) (which are already not in $F \setminus \{(u, v), (v, w)\}$ due to Lemma 5 (vi)). This contradicts the minimality of F .

We now do as follows. Recall that every directed path in F has length at most 2. Hence, we can partition F into r arcs (u, w) with $f(u) > 0$ and $f(w) < 0$ and $\frac{1}{2}t - r$ pairs of arcs (u, v) , (v, w) with $f(u) > 0$ and $f(w) < 0$. We deduced above that every directed path (u, v) , (v, w) reduces the number of components in H by one. Hence, the number of components in H is $1 + p - (\frac{1}{2}t - r)$.

Let G_1, \dots, G_k be the components of H that do not contain any vertex v with $f(v) \neq 0$. Note that $k = p - (\frac{1}{2}t - r)$. Because H is not connected and every vertex v with $f(v) \neq 0$ belongs to the same component of H , we find that $k \geq 1$. Choose an arbitrary arc (u, v) from F and for $i \in \{1, \dots, k\}$, choose an arbitrary vertex v_i in G_i . Remove (u, v) from H if $(u, v) \in A_F$ or add (v, u) to H otherwise (by Lemma 5 (iii) $(v, u) \in D_F$

if $(u, v) \notin A_F$). Add the arcs $(u, v_1), (v_1, v_2), \dots, (v_{k-1}, v_k), (v_k, v)$ to A_F . This gives a solution for CDBE(ea) that uses $r + 2(\frac{1}{2}t - r) + p - (\frac{1}{2}t - r) = p + \frac{1}{2}t$ arcs.

We now analyze the running time of an algorithm that can be derived from the above construction of a solution. As long as we find can two arcs (u, v) and (u', v') in F , such that u and v are in a different component of H than u' and v' , and moreover, Condition (i) or (ii) holds, we replace the arcs (u, v) and (u', v') in F by the arcs (u, v') and (u', v) . Note that we can identify all arcs of F that satisfy Condition (i) or (ii) in $O(n^2)$ time. We showed that every time we apply the above operation we reduce the number of components in H by at least one. Hence we apply this operation at most n times. Consequently, this step takes $O(n^3)$ time in total.

By checking whether the condition in Case 2 holds, we can find out whether Case 1 or 2 applies in $O(n + m)$ time. If Case 1 applies, we can then construct an optimal solution in $O(n^2)$ time (as described).

Suppose Case 2 applies. Recall that every directed path in $G_S(F)$ has length at most 2. If we have arcs $(u, v), (v, w) \in F$, then by minimality of F it follows that $(u, w) \in E(H)$. This means that in the graph H every vertex of $N_{G_S(F)}^{in}(v)$ is connected by an arc to every vertex of $N_{G_S(F)}^{out}(v)$. Thus if a vertex v has at least two in-neighbours or at least two out-neighbours in $G_S(F)$, it immediately follows that $\{(u, v), (v, w)\}$ is not an edge-cut set in H and we can do the replacement as described in Case 2, after which H will have one fewer component. Hence, we need to search for such a vertex v at most n times, so this procedure stops in $O(n^2)$ time. If there is no such vertex v (anymore) with at least two in-neighbours or at least two out-neighbours in $G_S(F)$, that is, if all middle vertices of the 3-vertex paths in $G_S(F)$ are distinct, then there are at most $O(n)$ paths composed of two arcs $(u, v), (v, w) \in F$. All such paths can be found in $O(n^2)$ time. For each of these, we can check whether or not $\{(u, v), (v, w)\}$ is an edge-cut set in H and do the replacement if appropriate in $O(n^2)$ time. After this, we can construct a solution in $O(n^2)$ time (as described in Case 2). The total runtime is therefore $O(n^3)$. This completes the proof of Lemma 7. \square

The next result is our first main result of this section. We prove it by showing that the upper bound in Lemma 7 is also a lower bound for (almost) any instance of CDBE(S) with $\{\text{ea}\} \subseteq S \subseteq \{\text{ea}, \text{ed}\}$ that has a semi-solution.

Theorem 4. For $\{\text{ea}\} \subseteq S \subseteq \{\text{ea}, \text{ed}\}$, $\text{CDBE}(S)$ can be solved in time $O(n^3 \log n \log \log n)$.

Proof. Let $\{\text{ea}\} \subseteq S \subseteq \{\text{ea}, \text{ed}\}$, and let (G, δ) be an instance of $\text{CDBE}(S)$. We first use Lemma 3 to check whether G_S has a directed f -join. Because G_S has at most $2n^2$ arcs, this takes $O(n^3 \log n \log \log n)$ time. If G_S has no directed f -join then (G, δ) has no semi-solution by Lemma 6, and thus no solution either. Assume that G_S has a directed f -join, and let F be a minimum directed f -join that can be found in time $O(n^3 \log n \log \log n)$ by Lemma 3. As before, p denotes the number of components of G that do not contain any vertex of T , while q is the number of components of G that contain at least one vertex of T , and $t = \sum_{u \in T} |f(u)|$.

We will prove the following series of statements.

- $\text{opt}_S(G, \delta) = 0$ if $p \leq 1, q = 0$,
- $\text{opt}_S(G, \delta) = p$ if $p \geq 2, q = 0$,
- $\text{opt}_S(G, \delta) = \max(|F|, p + q - 1, p + \frac{1}{2}t)$ if $q > 0$.

If $p \leq 1$ and $q = 0$ then $A = D = \emptyset$ is an optimal solution. If $p \geq 2$ and $q = 0$, to ensure connectivity and preserve degree balance, for every component of G there must be at least one arc whose head is in this component and at least one arc whose tail is in this component, thus any solution must contain at least p arcs. Let G_1, \dots, G_p be the components of G and arbitrarily choose vertices $v_i \in V(G_i)$ for $i \in \{1, \dots, p\}$. Let $A = \{(v_1, v_2), (v_2, v_3), \dots, (v_{p-1}, v_p), (v_p, v_1)\}$ and $D = \emptyset$. Then (A, D) is a solution which has size p and is therefore optimal.

Suppose $q \geq 1$. By Lemma 7 we find a solution (A, D) for (G, δ) of size at most $\max\{|F|, p + q - 1, p + \frac{1}{2}t\}$ in $O(n^3)$ time. Hence, the total running time is $O(n^3 \log n \log \log n)$, and it remains to show that any solution has size at least $\max(|F|, p + q - 1, p + \frac{1}{2}t)$.

Let (A, D) be an arbitrary solution. Then (A, D) is also semi-solution. Every semi-solution has size at least $|F|$ by Lemma 6 (ii). Therefore (A, D) has size at least $|F|$.

Since there are $p + q$ components in G , we must add at least $p + q - 1$ arcs to ensure $G + A - D$ is connected. Therefore (A, D) has size at least $p + q - 1$.

Finally, for every vertex u with $f(u) > 0$ (resp. $f(u) < 0$) we find that (A, D) must be such that at least $|f(u)|$ arcs are either in A and have u as a tail (resp. head) or else are in D and have u as a head (resp. tail). For every component containing only vertices v with $f(v) = 0$, there must be

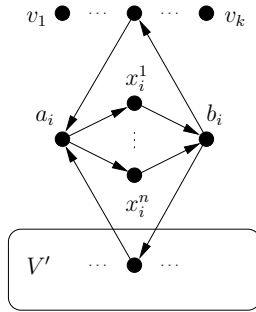


Fig. 1. The gadget used in the construction of G' .

at least one arc in A whose head is in this component and at least one arc in A whose tail is in this component (to ensure connectivity and to ensure that the degree balance is not changed for any vertex in this component). Therefore we have that (A, D) has size at least $p + \frac{1}{2}t$. This completes the proof of Theorem 4. \square

4.2 The $W[1]$ -Hard Cases

Recall that Cygan et al. [8] proved that $CDBE(\{\text{vd}\})$ is NP-complete and $W[1]$ -hard when parameterized by k , even when $\delta \equiv 0$. Our next results shows that this remains true if we allow not only vertex deletions, but also edge deletions and/or edge additions.

Theorem 5. *Let $\{\text{vd}\} \subseteq S \subseteq \{\text{vd}, \text{ed}, \text{ea}\}$. Then $CDBE(S)$ is NP-complete and $W[1]$ -hard when parameterized by k , even if $\delta \equiv 0$.*

Proof. Let $\{\text{vd}\} \subseteq S \subseteq \{\text{vd}, \text{ed}, \text{ea}\}$. The $CDBE(S)$ problem trivially belongs to NP. To prove hardness, we describe a parameterized reduction from DIRECTED BALANCED NODE DELETION. This problem takes as input a digraph G and an integer $k > 0$, and asks whether there exists a set A of at most k vertices whose deletion yields a balanced digraph. This problem is known to be NP-complete and $W[1]$ -hard with parameter k [8].

Let (G, k) be an instance of DIRECTED BALANCED NODE DELETION, and let $n = |V(G)|$. We construct a digraph G' as follows (see also Fig. 1). We start with a copy of G , where for every $v \in V(G)$, we denote the copy of v in G' by v' . Let $V' = \{v' \mid v \in V(G)\}$. We add k isolated vertices v_1, \dots, v_k . For each $i \in \{1, \dots, 2k + 1\}$, we construct a gadget G_i consisting of vertices $a_i, b_i, x_i^1, \dots, x_i^n$ and arcs (a_i, x_i^j) and (x_i^j, b_i) for every

$j \in \{1, \dots, n\}$. We make every vertex $v \in V' \cup \{v_1, \dots, v_k\}$ adjacent to each of the gadgets by adding arcs (v, a_i) and (b_i, v) for every $i \in \{1, \dots, 2k+1\}$. This completes the construction of G' . We define a function $\delta : V(G') \rightarrow \mathbb{Z}$ by setting $\delta(v) = 0$ for every $v \in V(G')$.

We claim that (G', k, δ) is a yes-instance of $\text{CDBE}(S)$ if and only if (G, k) is a yes-instance of $\text{DIRECTED BALANCED NODE DELETION}$.

First suppose (G, k) is a yes-instance of $\text{DIRECTED BALANCED NODE DELETION}$. Then there is a set $A \subseteq V(G)$ of size at most k such that $G - A$ is balanced. We define a set $A' \subseteq V(G')$ of size k as follows. If $|A| = k$, then we set $A' = \{a' \mid a \in A\}$. If $|A| < k$, then we set $A' = \{a' \mid a \in A\} \cup \{v_1, \dots, v_{k-|A|}\}$. We claim that $G' - A'$ is Eulerian. Since the gadgets are connected and every vertex outside the gadgets is adjacent to each of the gadgets, it is clear that $G' - A'$ is connected. It remains to show that every vertex in $G' - A'$ is balanced. In G' , the in- and out-degrees of each vertex a_i equal $n+k$ and n , respectively, while the in- and out-degrees of each vertex b_i equal n and $n+k$, respectively. Since each of the k vertices in A' is an in-neighbour of a_i and an out-neighbour of b_i , it holds that $d_{G'-A'}^{\text{out}}(a_i) = d_{G'-A'}^{\text{in}}(a_i) = d_{G'-A'}^{\text{out}}(b_i) = d_{G'-A'}^{\text{in}}(b_i) = n$ for each $i \in \{1, \dots, 2k+1\}$. All other vertices in the gadgets, already balanced in G' , remain balanced in $G' - A'$. The same holds for the vertices in $\{v_1, \dots, v_k\} \setminus A'$; the in- and out-degree of each of these vertices, both in G' and in $G' - A'$, equals $2k+1$. For every vertex $v' \in V' \setminus A'$, it holds that $d_{G'-A'}^{\text{out}}(v') = d_{G-A}^{\text{out}}(v) + 2k+1$ and $d_{G'-A'}^{\text{in}}(v') = d_{G-A}^{\text{in}}(v) + 2k+1$. Since $d_{G-A}^{\text{out}}(v) = d_{G-A}^{\text{in}}(v)$ for every $v \in V(G) \setminus A$ due to the assumption that $G - A$ is balanced, it holds that every $v' \in V' \setminus A'$ is balanced in $G' - A'$. We conclude that $G' - A'$ is Eulerian.

For the reverse direction, suppose there exists a sequence L of operations from S that transforms G' into an Eulerian digraph. We first argue that L deletes exactly k vertices from $V' \cup \{v_1, \dots, v_k\}$. As we mentioned before, the in- and out-degrees of each vertex a_i in G' equal $n+k$ and n in G' , respectively, while the in- and out-degrees of each vertex b_i in G' equal n and $n+k$, respectively. Since $k > 0$ by assumption, this means that the operations in L need to either delete or balance each of the $4k+2$ vertices in the set $Z = \{a_1, \dots, a_{2k+1}, b_1, \dots, b_{2k+1}\}$. Since $|L| = k$ and each edge deletion or edge addition changes the degree of at most two vertices in Z , there is a gadget G_j such that L neither deletes a vertex of G_j nor adds or deletes an edge incident with any of the vertices of G_j . The fact that the vertices of G_j , and a_j and b_j in particular, are balanced after applying the operations in L implies that L deletes exactly k

in-neighbours of a_j (all of which are out-neighbours of b_j). We conclude that L deletes exactly k vertices from $V' \cup \{v_1, \dots, v_k\}$.

Let $A' \subseteq V'$ be the set of at most k vertices that are deleted from V' by L , and let $A = \{v \in V(G) \mid v' \in A'\}$ be the corresponding set of vertices in G . Let $v \in V(G) \setminus A$. From the construction of G' , it holds that $d_{G-A}^{out}(v) = d_{G'-A'}^{out}(v) - (2k+1)$ and $d_{G-A}^{in}(v) = d_{G'-A'}^{in}(v') - (2k+1)$. Since $d_{G'-A'}^{out}(v') = d_{G'-A'}^{in}(v')$, we have that $d_{G-A}^{out}(v) = d_{G-A}^{in}(v)$. This shows that $G - A$ is balanced, and hence (G, k) is a yes-instance of DIRECTED BALANCED NODE DELETION. \square

5 Conclusions

By extending previous work [2,5,8] we completely classified both the classical and parameterized (with respect to parameter k) complexity of the problems CDPE(S) and CDBE(S), as summarized in Table 1. Our work followed the framework used [15,23] for (CONNECTED) DEGREE CONSTRAINT EDITING(S). Our study was motivated by Eulerian graphs. As such, the variants DPE(S) and DBE(S) of CDPE(S) and CDBE(S), respectively, in which the graph H is no longer required to be connected, were beyond the scope of this paper. It follows from results of Cai and Yang [5] and Cygan et al. [8], respectively, that for $S = \{\mathbf{vd}\}$, DPE(S) and DBE(S) are NP-complete and, when parameterized by k , W[1]-hard, whereas they are polynomial-time solvable for $S = \{\mathbf{ed}\}$ as a result of Lemmas 2 and 3, respectively. The problems DPE(S) and DBE(S) are also polynomial-time solvable if $\{\mathbf{ea}\} \subseteq S \subseteq \{\mathbf{ea}, \mathbf{ed}\}$; this is in fact proven by combining Lemmas 2 and 4 for the undirected case, and Lemmas 3 and 6 for the directed case. We expect the remaining (hardness) results of Table 1 to carry over as well.

Let ℓ be an integer. Here is a natural generalization of CDPE(S).

ℓ -CDME(S): CONNECTED DEGREE MODULO- ℓ -EDITING(S)
Instance: A graph G , integer k and
a function $\delta: V(G) \rightarrow \{0, \dots, \ell - 1\}$.
Question: Can G be (S, k) -modified into a connected graph H
with $d_H(v) \equiv \delta(v) \pmod{\ell}$ for each $v \in V(H)$?

Note that 2-CDME(S) is CDPE(S). The following theorem shows that the complexity of 3-CDME(S) may differ from 2-CDME(S).

Theorem 6. 3-CDME($\{\mathbf{ea}, \mathbf{ed}\}$) is NP-complete even if $\delta \equiv 2$.

Proof. Reduce from the HAMILTONICITY problem, which is NP-complete for connected cubic graphs [14]. Let G be a connected cubic graph. Let $\delta(v) = 2$ for every $v \in V(G)$, and take $k = |E(G)| - |V(G)|$. Then G has a Hamiltonian cycle if and only if G can be (S, k) -modified into a connected graph H with $d_H(v) = 2 \pmod{3}$ for all $v \in V(H)$. \square

It is natural to ask whether $3\text{-CDME}(\{\text{ea, ed}\})$ is fixed-parameter tractable with parameter k .

Finally, another direction for future research is to investigate how the complexity of $\text{CDPE}(S)$ and $\text{CDBE}(S)$ changes if we permit other graph operations, such as edge contraction, to be in the set S . For instance, Belmonte et al. [1] considered this operation and obtained the first results extending the work of Mathieson and Szeider [23] in this direction.

Acknowledgements. We are grateful to three anonymous reviewers for a number of comments that improved the presentation of our paper.

References

1. R. Belmonte, P. A. Golovach, P. van 't Hof, and D. Paulusma. Parameterized complexity of three edge contraction problems with degree constraints. *Acta Informatica*, 51(7):473–497, 2014.
2. F. T. Boesch, C. L. Suffel, and R. Tindell. The spanning subgraphs of Eulerian graphs. *Journal of Graph Theory*, 1(1):79–84, 1977.
3. P. Burzyn, F. Bonomo, and G. Durán. NP-completeness results for edge modification problems. *Discrete Applied Mathematics*, 154(13):1824–1844, 2006.
4. L. Cai. Fixed-parameter tractability of graph modification problems for hereditary properties. *Information Processing Letters*, 58(4):171–176, 1996.
5. L. Cai and B. Yang. Parameterized complexity of even/odd subgraph problems. *Journal of Discrete Algorithms*, 9(3):231–240, 2011.
6. K. Cechlárová and I. Schlotter. Computing the deficiency of housing markets with duplicate houses. In *IPEC 2010*, volume 6478 of *Lecture Notes in Computer Science*, pages 72–83. Springer, 2010.
7. R. Crowston, G. Gutin, M. Jones, and A. Yeo. Parameterized Eulerian strong component arc deletion problem on tournaments. *Information Processing Letters*, 112(6):249–251, 2012.
8. M. Cygan, D. Marx, M. Pilipczuk, M. Pilipczuk, and I. Schlotter. Parameterized complexity of Eulerian deletion problems. *Algorithmica*, 68(1):41–61, 2014.
9. K. K. Dabrowski, P. A. Golovach, P. van 't Hof, and D. Paulusma. Editing to Eulerian graphs. In *FSTTCS 2014*, volume 29 of *LIPICs*, pages 97–108. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2014.
10. F. Dorn, H. Moser, R. Niedermeier, and M. Weller. Efficient algorithms for Eulerian extension and rural postman. *SIAM Journal on Discrete Mathematics*, 27:75–94, 2013.
11. R. G. Downey, M. R. Fellows, A. Vardy, and G. Whittle. The parametrized complexity of some fundamental problems in coding theory. *SIAM Journal on Computing*, 29:545–570, 1999.

12. J. Edmonds and E. L. Johnson. Matching, Euler tours and the Chinese postman. *Mathematical Programming*, 5:88–124, 1973.
13. V. Froese, A. Nichterlein, and R. Niedermeier. Win-win kernelization for degree sequence completion problems. In *SWAT 2014*, volume 8503 of *Lecture Notes in Computer Science*, pages 194–205. Springer, 2014.
14. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.
15. P. A. Golovach. Editing to a connected graph of given degrees. In *MFCS Part II*, volume 8635 of *Lecture Notes in Computer Science*, pages 324–335. Springer, 2014.
16. P. A. Golovach. Editing to a graph of given degrees. *Theoretical Computer Science*, 591:72–84, 2015.
17. P. Goyal, P. Misra, F. Panolan, G. Philip, and S. Saurabh. Finding even subgraphs even faster. *CoRR*, abs/1409.4935, 2014.
18. P. L. Hammer and B. Simeone. The splittance of a graph. *Combinatorica*, 1(3):275–284, 1981.
19. W. Höhn, T. Jacobs, and N. Megow. On Eulerian extensions and their application to no-wait flowshop scheduling. *Journal of Scheduling*, 15(3):295–309, 2012.
20. M.-K. Kwan. Graphic programming using odd or even points. *Chinese Mathematics*, 1:273–277, 1962.
21. L. Lesniak and O. R. Oellermann. An Eulerian exposition. *Journal of Graph Theory*, 10(3):277–297, 1986.
22. J. M. Lewis and M. Yannakakis. The node-deletion problem for hereditary properties is NP-complete. *Journal of Computer and System Sciences*, 20(2):219–230, 1980.
23. L. Mathieson and S. Szeider. Editing graphs to satisfy degree constraints: A parameterized approach. *Journal of Computer and System Sciences*, 78(1):179–191, 2012.
24. S. Micali and V. V. Vazirani. An $O(\sqrt{|V|}|E|)$ algorithm for finding maximum matching in general graphs. In *FOCS 1980*, pages 17–27. IEEE Computer Society, 1980.
25. H. Moser and D. M. Thilikos. Parameterized complexity of finding regular induced subgraphs. *Journal of Discrete Algorithms*, 7(2):181–190, 2009.
26. A. Natanzon, R. Shamir, and R. Sharan. Complexity classification of some edge modification problems. *Discrete Applied Mathematics*, 113(1):109–128, 2001.