

# Augmenting Graphs to Minimize the Diameter

Fabrizio Frati<sup>1</sup>, Serge Gaspers<sup>2,3</sup>,  
Joachim Gudmundsson<sup>1,3</sup>, and Luke Mathieson<sup>4</sup>

<sup>1</sup> University of Sydney, Australia,  
brillo@it.usyd.edu.au, joachim.gudmundsson@sydney.edu.au

<sup>2</sup> The University of New South Wales, Australia, sergeg@cse.unsw.edu.au

<sup>3</sup> NICTA, Australia

<sup>4</sup> Macquarie University, Australia, luke.mathieson@mq.edu.au

**Abstract.** We study the problem of augmenting a weighted graph by inserting edges of bounded total cost while minimizing the diameter of the augmented graph. Our main result is an FPT 4-approximation algorithm for the problem.

## 1 Introduction

We study the problem of minimizing the diameter of a weighted graph by the insertion of edges of bounded total cost. This problem arises in practical applications [2,4] such as telecommunications networks, information networks, flight scheduling, protein interactions, and it has also received considerable attention from the graph theory community, see for example [1,7,11].

We introduce some terminology. Let  $G = (V, E)$  be an undirected weighted graph. Let  $[V]^2$  be the set of all possible edges on the vertex set  $V$ . A *non-edge* of  $G$  is an element of  $[V]^2 \setminus E$ . The *weight* of a path in  $G$  is the sum of its edge weights. For any  $u, v \in V$ , the *shortest  $u$ - $v$  path* in  $G$  is the path connecting  $u$  and  $v$  in  $G$  with minimum weight. The weight of this path is said to be the *distance* between  $u$  and  $v$  in  $G$ . Finally, the *diameter* of  $G$  is the largest distance between any two vertices in  $G$ . The problem we study in this paper is formally defined as follows.

**PROBLEM:** Bounded Cost Minimum Diameter Edge Addition (BCMD)

**INPUT:** An undirected graph  $G = (V, E)$ , a weight function  $w : [V]^2 \rightarrow \mathbb{N}$ , a cost function  $c : [V]^2 \rightarrow \mathbb{N}^*$ , and an integer  $B$ .

**GOAL:** A set  $F$  of non-edges with  $\sum_{e \in F} c(e) \leq B$  such that the diameter of the graph  $G_B = (V, E \cup F)$  with weight function  $w$  is minimized. We say that  $G_B$  is a *B-augmentation* of  $G$ .

The main result of this paper is a fixed parameter tractable (FPT) 4-approximation algorithm for BCMD with parameter  $B$ . FPT approximation algorithms are surveyed by Marx [14]. For background on parameterized complexity we refer to [6,8,15] and for background on approximation algorithms to [17].

Several papers in the literature already dealt with the BCMD problem. However, most of them focused on restricted versions of the problem, namely the one in which all costs and all weights are identical [3,5,12,13], and the one in which all the edges have unit costs and the weights of the non-edges are all identical [2,4].

The BCMD problem can be seen as a bicriteria optimization problem where the two optimization criteria are: (1) the cost of the edges added to the graph and (2) the diameter of the augmented graph. As is standard in the literature, we say that an algorithm is an  $(\alpha, \beta)$ -approximation algorithm for the BCMD problem, with  $\alpha, \beta \geq 1$ , if it computes a set  $F$  of non-edges of  $G$  of total cost  $\alpha \cdot B$  such that the diameter of  $G' = (V, E \cup F)$  is at most  $\beta \cdot D_{opt}^B$ , where  $D_{opt}^B$  is the diameter of an optimal  $B$ -augmentation of  $G$ .

We survey some known results about the BCMD problem. Note that all the algorithms discussed below run in polynomial time.

**Unit weights and unit costs.** The restriction of BCMD to unit costs and unit weights was first shown to be NP-hard in 1987 by Schoone et al. [16]; see also the paper by Li et al. [13]. Bilò et al. [2] showed that, as a consequence of the results in [3,5,13], there exists no  $(c \log n, \delta < 1 + 1/D_{opt}^B)$ -approximation algorithm for BCMD if  $D_{opt}^B \geq 2$ , unless P=NP. For the case where  $D_{opt}^B \geq 6$ , they proved a stronger lower bound, namely that there exists no  $(c \log n, \delta < \frac{5}{3} - \frac{7 - (D_{opt}^B + 1) \bmod 3}{3D_{opt}^B})$ -approximation algorithm, unless P=NP.

Dodis and Khanna [5] gave an  $(O(\log n), 2 + 2/D_{opt}^B)$ -approximation algorithm (see also [12]). Li et al. [13] showed a  $(1, 4 + 2/D_{opt}^B)$ -approximation algorithm. The analysis of the latter algorithm was later improved by Bilò et al. [2], who showed that it gives a  $(1, 2 + 2/D_{opt}^B)$ -approximation. In the same paper they also gave a  $(O(\log n), 1)$ -approximation algorithm.

**Unit costs and restricted weights.** Some of the results from the unweighted setting have been extended to a restricted version of the weighted case, namely the one in which the edges of  $G$  have arbitrary non-negative integer weights, however all the non-edges of  $G$  have cost 1 and uniform weight  $\omega \geq 0$ .

Bilò et al. [2] showed how two of their algorithms can be adapted to this restricted weighted case. In fact, they gave a  $(1, 2 + 2\omega/D_{opt}^B)$ -approximation algorithm and a  $(2 - 1/B, 2)$ -approximation algorithm. Similar results were obtained by Demaine and Zadimoghaddam in [4].

Bilò et al. [2] also showed that, for every  $D_{opt}^B \geq 2\omega$  and for some constant  $c$ , there is no  $(c \log n, \delta < 2 - 3\omega/D_{opt}^B)$ -approximation algorithm for this restriction of the BCMD problem, unless P=NP.

**Arbitrary costs and weights.** To the best of our knowledge, there is only one theory paper that has considered the general BCMD problem. In 1999, Dodis and Khanna [5] presented an  $O(n \log D_{opt}^B, 1)$ -approximation algorithm, assuming that all weights are polynomially bounded. Their result is based on a multicommodity flow formulation of the problem.

**Our results.** In this paper we study the BCMD problem with arbitrary integer costs and weights. Our main result is a  $(1, 4)$ -approximation algorithm with running time  $O((3^B B^3 + n + \log(Bn))Bn^2)$ . We also prove that, considering  $B$  as a parameter, it is  $W[2]$ -hard to compute a  $(1 + c/B, 3/2 - \epsilon)$ -approximation, for any constants  $c$  and  $\epsilon > 0$ . Further, we present polynomial-time  $((k + 1)^2, 3)$ -,  $(k, 4)$ -, and  $(1, 3k + 2)$ -approximation algorithms for the unit-cost restriction of the BCMD problem. Due to space constraints, some proofs have been moved to the appendix.

## 2 Shortest Paths with Bounded Cost

Let  $(G = (V, E), w, c, B)$  be an instance of the BCMD problem and let  $K$  denote the complete graph on the vertex set  $V$ . The edges of  $K$  have the same weights and costs as they have in  $G$  (observe that an edge  $e$  of  $K$  is either an edge or a non-edge of  $G$ ). For technical reasons, we add self-loops with weight 0 and cost 1 at each vertex of  $K$ .

For any  $0 \leq \beta \leq B$ , a path in  $K$  is said to be a  $\beta$ -bounded-cost path if it uses non-edges of  $G$  of total cost at most  $\beta$ . We consider the problem of computing, for every integer  $0 \leq \beta \leq B$  and for every two vertices  $u, v \in V$ , a  $\beta$ -bounded-cost shortest path connecting  $u$  and  $v$ . We call this problem the *All-Pairs  $B$ -Shortest Paths* ( $APSP_B$ ) problem. We will prove the following.

**Theorem 1.** *The  $APSP_B$  problem can be solved in  $O(Bn^3 + Bn^2 \log(Bn))$  time using  $O(Bn^2)$  space.*

In order to prove **Theorem 1**, we construct a directed graph  $H = (U, F)$  as follows. First, consider  $G$  as a directed graph, i.e., replace every undirected edge  $\{u, v\}$  with two arcs  $(u, v)$  and  $(v, u)$  with the same weight and cost as the edge  $\{u, v\}$ . Then,  $H = (U, F)$  contains  $B + 1$  copies of  $G$ , denoted by  $G_0, \dots, G_B$ . For any  $1 \leq i \leq B$ , we denote by  $(v, i)$  the copy of vertex  $v \in V$  in  $G_i = (V_i, E_i)$ . The arc set  $F$  contains the union of  $E'$  and  $F'$ , where  $E' = \bigcup_{0 \leq i \leq k} E_i$ , and

$$F' = \left\{ ((u, i), (v, i + c(\{u, v\}))) : 0 \leq i \leq B - c(\{u, v\}), \{u, v\} \in [V]^2 \setminus E \right\}.$$

For each  $((u, i), (v, j)) \in F'$ , the weight and the cost of  $((u, i), (v, j))$  are  $w(\{u, v\})$  and  $c(\{u, v\}) = j - i$ , respectively.

**Observation 1** *The number of vertices in  $U$  is  $(B + 1)n$  and the number of arcs in  $F$  is  $O(Bn^2)$ .*

We will use directed graph  $H$  to efficiently compute  $\beta$ -bounded-cost shortest paths in  $K$ . This is possible due to the following two lemmata.

**Lemma 1.** *Suppose that  $H$  contains a directed path  $P_H$  with weight  $W$  connecting vertices  $(u, i)$  and  $(v, j)$ , for some  $j \geq i$ . Then, there exists a  $(j - i)$ -bounded-cost path  $P_K$  in  $K$  with weight  $W$  connecting  $u$  and  $v$ .*

**Lemma 2.** *Suppose that there exists a  $\beta$ -bounded-cost path  $P_K$  in  $K$  with weight  $W$  connecting vertices  $u$  and  $v$  and not containing any self-loop. Then, there exists a directed path  $P_H$  in  $H$  with weight  $W$  connecting vertices  $(u, 0)$  and  $(v, \beta')$  with  $\beta' \leq \beta$ .*

We have the following.

**Corollary 1.** *There is a  $\beta$ -bounded-cost path connecting vertices  $u$  and  $v$  in  $K$  with weight  $W$  if and only if there is a directed path in  $H$  connecting vertices  $(u, 0)$  and  $(v, \beta)$  with weight  $W$ .*

We are now ready to prove **Theorem 1**. Consider any vertex  $u$  in  $K$ . We first mark every vertex that can be reached from  $(u, 0)$  in  $H$  with the weight of its shortest path from  $(u, 0)$ . By **Observation 1**,  $H$  has  $O(Bn)$  vertices and  $O(Bn^2)$  edges, hence this can be

done in  $O(Bn^2 + Bn \log(Bn))$  time [9]. For every  $0 \leq \beta \leq B$  and for every vertex  $v \neq u$ , by [Corollary 1](#) the weight of a  $\beta$ -bounded cost shortest path in  $K$  is the same as the weight of a shortest directed path from  $(u, 0)$  to  $(v, \beta)$ . Hence, for every  $0 \leq \beta \leq B$  and for every vertex  $v \neq u$ , we can determine in total  $O(Bn^2 + Bn \log(Bn))$  time the weight of a  $\beta$ -bounded cost shortest path in  $K$  connecting  $u$  and  $v$ . Thus, for every  $0 \leq \beta \leq B$  and for every pair of vertices  $u$  and  $v$  in  $K$ , we can determine in total  $O(Bn^3 + Bn^2 \log(Bn))$  time the weight of a  $\beta$ -bounded cost shortest path in  $K$  connecting  $u$  and  $v$ . This concludes the proof of [Theorem 1](#).

### 3 Arbitrary Costs and Weights

Our algorithms, as many afore-mentioned approximation algorithms for the BCMD problem, use a clustering approach as a first phase to find a set  $C$  of  $B+1$  cluster centers. The idea of the algorithm is to create a minimum height rooted tree  $T = (U \subseteq V, D)$ , so that  $C \subseteq U$ , by adding a set of edges of total cost at most  $B$  to  $G$ . We will prove that such a tree approximates an optimal  $B$ -augmentation.

#### 3.1 Clustering

We start by defining the clustering approach used to generate the  $B + 1$  cluster centers. Whereas a costly binary search is used in [4] to guess the radius of the clusters, we adapt the approach of [2] to our more general setting.

For two vertices  $u, v$ , we denote by  $\text{dist}_G(u, v)$  the distance between  $u$  and  $v$  in  $G$ . For a vertex  $u$  and a set of vertices  $S$ , we denote by  $\text{dist}_G(u, S)$  the minimum distance between  $u$  and any vertex from  $S$  in  $G$ , i.e.,  $\text{dist}_G(u, S) = \min_{v \in S} \{\text{dist}_G(u, v)\}$ . For a set of vertices  $S$ , we denote by  $\text{dist}_G(S)$  the minimum distance between any two distinct vertices from  $S$  in  $G$ , i.e.,  $\text{dist}_G(S) = \min_{u \in S} \{\text{dist}_G(u, S \setminus \{u\})\}$ .

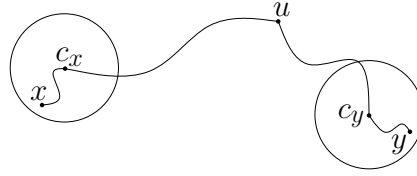
The clustering phase computes a set  $C = \{c_1, \dots, c_{B+1}\}$  of  $B + 1$  cluster centers as follows. Vertex  $c_1$  is an arbitrary vertex in  $V$ ; for  $2 \leq i \leq B + 1$ , vertex  $c_i$  is chosen so that  $\text{dist}_G(c_i, \{c_1, \dots, c_{i-1}\})$  is maximized. Ties are broken arbitrarily.

**Lemma 3.** *The clustering phase computes in  $O(Bn^2)$  time a set  $C \subseteq V$  of size  $B + 1$  such that  $\text{dist}_G(v, C) \leq D_{opt}^B$  for every vertex  $v \in V$ .*

**Proof.** First, note that the above described algorithm can easily be implemented in  $O(Bn^2)$  time using  $B$  iterations of Dijkstra's algorithm with Fibonacci heaps [9]. Let  $c_{B+2}$  denote a vertex maximizing  $\text{dist}_G(c_{B+2}, C)$ , and denote this distance by  $R$ . By definition,  $\text{dist}_G(v, C) \leq R$  for every  $v \in V$ . To prove the lemma it remains to show that  $R \leq D_{opt}^B$ . For the sake of contradiction, assume  $D_{opt}^B < R$ . Then,  $C \cup \{c_{B+2}\}$  is a set of  $B + 2$  vertices with pairwise distance larger than  $D_{opt}^B$  in  $G$ . We prove the following claim.

**Claim 1** *Let  $G'$  be a weighted graph and let  $C'$  be a set of vertices in  $G'$  such that  $\text{dist}_{G'}(C') > D$ . Then, for every graph  $G''$  obtained from  $G'$  by adding a single non-edge of  $G'$  with non-negative weight, there is a set  $C'' \subset C'$  with  $|C''| = |C'| - 1$  and with  $\text{dist}_{G''}(C'') > D$ .*

Now, since  $C \cup \{c_{B+2}\}$  is a set of  $B + 2$  vertices with pairwise distance larger than  $D_{opt}^B$  in  $G$ , by iteratively using the claim we have that in any graph  $G_B$  obtained from



**Fig. 1.** Illustrating the path defined in the proof of [Lemma 5](#).

$G$  by adding a set of edges  $F \subseteq [V]^2 \setminus E$  with total cost at most  $B$  (that is, in any  $B$ -augmentation  $G_B$  of  $G$ ), we have a set of  $B + 2 - |F| \geq 2$  vertices with pairwise distance greater than  $D_{opt}^B$ , thus contradicting the definition of  $D_{opt}^B$ . This concludes the proof of the lemma.  $\square$

### 3.2 A minimum height tree

Let  $C$  be a set of  $B + 1$  cluster centers such that the  $B + 1$  clusters with centers at  $C = \{c_0, \dots, c_B\}$  and radius  $D_{opt}^B$  cover the vertices of  $G$ . This set can be computed as described in the previous section.

**Definition 1.** Let  $G = (V, E)$  be a graph together with a weight function  $w : [V]^2 \rightarrow \mathbb{N}$ . Let  $C \subseteq V$  and let  $u$  be a vertex in  $V$ . A Shortest Path Tree of  $G$ ,  $C$ , and  $u$ , denoted by  $\text{SPT}(G, C, u)$ , is a tree  $T$  rooted at  $u$ , spanning  $C$ , whose vertices and edges belong to  $V$  and  $E$ , respectively, and such that, for every vertex  $c \in C$ , it holds  $d_T(u, c) = d_G(u, c)$ .

The *height* of a weighted rooted tree  $T$ , which is denoted by  $h(T)$ , is the maximum weight of a path from the root to a leaf.

**Definition 2.** Let  $G = (V, E)$  be a graph together with a weight function  $w : [V]^2 \rightarrow \mathbb{N}$  and a cost function  $c : [V]^2 \rightarrow \mathbb{N}^*$ . Let  $C \subseteq V$ , let  $u$  be a vertex in  $V$ , and let  $B \geq 0$  be an integer. A Minimum Height $_B$  SPT of  $G$ ,  $C$ , and  $u$ , denoted by  $\text{MH}_B\text{SPT}(G, c, u)$ , is a  $\text{SPT}(G_B, C, u)$  of minimum height over all  $B$ -augmentations  $G_B$  of  $G$ .

Let  $G_B$  be a  $B$ -augmentation of  $G$  with diameter  $D_{opt}^B$ .

**Lemma 4.** The height of a  $\text{MH}_B\text{SPT}(G, C, u)$  is at most  $D_{opt}^B$ .

**Proof.** We have (A)  $h(\text{MH}_B\text{SPT}(G, C, u)) \leq h(\text{SPT}(G_B, C, u))$ . Since  $G_B$  is a  $B$ -augmentation of  $G$  with diameter  $D_{opt}^B$ , we have (B)  $h(\text{SPT}(G_B, C, u)) \leq D_{opt}^B$ . Inequalities (A) and (B) together prove the lemma.  $\square$

We now present a relationship between the BCMD problem and the problem of computing a  $\text{MH}_B\text{SPT}(G, C, u)$ .

**Lemma 5.** Let  $G'_B$  be a  $B$ -augmentation of  $G$  such that it holds  $h(\text{SPT}(G'_B, C, u)) = h(\text{MH}_B\text{SPT}(G, C, u))$ , for any  $u \in V$ . Then, the diameter of  $G'_B$  is at most  $4 \cdot D_{opt}^B$ .

**Proof.** Consider two vertices  $x$  and  $y$  in  $V$ , see Fig. 1. Let  $c_x$  and  $c_y$  be centers of the clusters  $x$  and  $y$  belong to, respectively. Then, we have  $\text{dist}_{G'_B}(x, y) \leq \text{dist}_G(x, c_x) + \text{dist}_{G'_B}(c_x, u) + \text{dist}_{G'_B}(u, c_y) + \text{dist}_G(c_y, y)$ . By Lemma 3,  $\text{dist}_G(x, c_x), \text{dist}_G(c_y, y) \leq D_{opt}^B$ . Since  $h(\text{SPT}(G'_B, C, u)) = h(\text{MH}_B\text{SPT}(G, C, u))$  and by Lemma 4, it holds  $\text{dist}_{G'_B}(c_x, u), \text{dist}_{G'_B}(u, c_y) \leq D_{opt}^B$ . Hence,  $\text{dist}_{G'_B}(x, y) \leq 4 \cdot D_{opt}^B$ .  $\square$

### 3.3 Constructing a minimum height tree

In this section, we show an algorithm to compute a  $\text{MH}_B\text{SPT}(G, C \setminus \{c_1\}, c_1)$ .

We introduce some notation and terminology. Let  $C' = C \setminus \{c_1\}$ . Denote by  $d_K^j(u, v)$  the minimum weight of a  $j$ -bounded cost path connecting  $u$  and  $v$  in  $K$ . For any  $u \in V$ , for any  $S \subseteq C'$ , and for any  $j, 0 \leq j \leq B$ , let  $\gamma(u, S, j)$  denote the height of a  $\text{MH}_j\text{SPT}(G, S, u)$ . Hence, the height of a  $\text{MH}_B\text{SPT}(G, C', c_1)$  is  $\gamma(c_1, C', B)$ .

The following main lemma gives a dynamic programming recurrence for computing  $\gamma(c_1, C', B)$ .

**Lemma 6.** *For any  $u \in V$ , any  $S \subseteq C'$ , and any  $0 \leq j \leq B$ , the following hold: If  $|S| = 1$ , then  $\gamma(u, S, j) = d_K^j(u, c_i)$  where  $S = \{c_i\}$ . If  $|S| > 1$ , then*

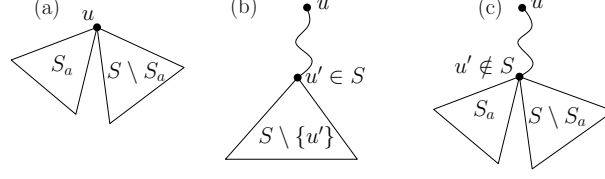
$$\gamma(u, S, j) = \min_{\substack{v \in V \\ S' \subsetneq S \\ j=j_1+j_2+j_3}} d_K^{j_1}(u, v) + \max\{\gamma(v, S', j_2), \gamma(v, S \setminus S', j_3)\}.$$

**Proof.** If  $|S| = \{c_i\}$ , then  $\text{MH}_j\text{SPT}(G, \{c_i\}, u)$  is a minimum-weight path connecting  $u$  and  $c_i$  and having total cost at most  $j$ . Hence,  $\gamma(u, S, j) = d_K^j(u, c_i)$ . In particular, notice that, if  $u = c_i$ , then  $\gamma(u, \{u\}, j) = d_K^j(u, u) = 0$ .

If  $|S| = m > 1$ , then suppose that the lemma holds for each  $\gamma(u', S', j')$  with  $|S'| \leq m - 1$  by induction. Denote by  $T$  any  $\text{MH}_j\text{SPT}(G, S, u)$ . Denote by  $P(v, w)$  the unique path in  $T$  connecting two vertices  $v$  and  $w$  of  $T$ . We distinguish three cases, based on the structure of  $T$ . In Case (a), the degree of  $u$  in  $T$  is at least two (see Fig. 2(a)). In Case (b), the degree of  $u$  in  $T$  is one and there exists a vertex  $u' \in S$  such that every internal vertex of  $P(u, u')$  has degree 2 in  $T$  and does not belong to  $S$  (see Fig. 2(b)). Finally, in Case (c), the degree of  $u$  in  $T$  is one and there exists a vertex  $u' \notin S$  such that every internal vertex of  $P(u, u')$  has degree 2 in  $T$  and does not belong to  $S$ , and such that the degree of  $u'$  is greater than two (see Fig. 2(c)).

First, we prove that one of the three cases always applies. If the degree of  $u$  in  $T$  is at least two, then Case (a) applies. Otherwise, the degree of  $u$  is 1. Traverse  $T$  from  $u$  until a vertex  $v'$  is found such that  $v' \in S$  or the degree of  $v'$  is at least 3. If  $v' \in S$ , then every internal vertex of  $P(u, u')$  has degree 2 in  $T$  and does not belong to  $S$ , hence Case (b) applies. If  $v' \notin S$ , then the degree of  $v'$  is at least 3, and every internal vertex of  $P(u, u')$  has degree 2 in  $T$  and does not belong to  $S$ , hence Case (c) applies. We now discuss the three cases.

In Case (a),  $T$  is composed of two subtrees  $\text{MH}_x\text{SPT}(G, S_a, u)$  and  $\text{MH}_y\text{SPT}(G, S \setminus S_a, u)$ , only sharing vertex  $u$ , with  $\emptyset \subsetneq S_a \subsetneq S$ . The height of  $T$  is the maximum of the heights of  $\text{MH}_x\text{SPT}(G, S_a, u)$  and  $\text{MH}_y\text{SPT}(G, S \setminus S_a, u)$ ; also the cost of  $T$  is at most  $x + y$ . By induction, the heights of  $\text{MH}_x\text{SPT}(G, S_a, u)$  and  $\text{MH}_y\text{SPT}(G, S \setminus S_a, u)$  are  $\gamma(u, S_a, x)$  and  $\gamma(u, S \setminus S_a, y)$ , respectively. Thus, the height of  $T$



**Fig. 2.** Illustration for the proof of **Lemma 6**.

is  $\max\{\gamma(v, S_a, x), \gamma(v, S \setminus S_a, y)\}$  and hence  $\gamma(u, S, j) = \max\{\gamma(u, S_a, x), \gamma(u, S \setminus S_a, y)\}$ . Such a value is found by the recursive definition of  $\gamma(u, S, j)$  with  $v = u$ ,  $S' = S_a$ ,  $j_1 = 0$ ,  $j_2 = x$ , and  $j_3 = y$ .

In Case (b),  $T$  is composed of a path from  $u$  to  $u'$  with cost  $x$  and weight  $d_K^x(u, u')$ , and of a  $\text{MH}_y\text{SPT}(G, S \setminus \{u'\}, u')$ . The height of  $T$  is the sum of  $d_K^x(u, u')$  and the height of  $\text{MH}_y\text{SPT}(G, S \setminus \{u'\}, u')$ ; also the cost of  $T$  is at most  $x + y$ . By induction, the height of  $\text{MH}_y\text{SPT}(G, S \setminus \{u'\}, u')$  is  $\gamma(u', S \setminus \{u'\}, y)$ . Thus, the height of  $T$  is  $d_K^x(u, u') + \gamma(u', S \setminus \{u'\}, y)$  and hence  $\gamma(u, S, j) = d_K^x(u, u') + \gamma(u', S \setminus \{u'\}, y)$ . Such a value is found by the recursive definition of  $\gamma(u, S, j)$  with  $v = u'$ ,  $S' = S \setminus \{u'\}$ ,  $j_1 = x$ ,  $j_2 = y$ , and  $j_3 = 0$ .

In Case (c),  $T$  is composed of a path from  $u$  to  $u'$  with cost  $x$  and weight  $d_K^x(u, u')$ , of a  $\text{MH}_y\text{SPT}(G, S_a, u')$ , and of a  $\text{MH}_z\text{SPT}(G, S \setminus S_a, u')$  with  $\emptyset \subsetneq S_a \subsetneq S$ . The height of  $T$  is the sum of  $d_K^x(u, u')$  and the maximum between the heights of  $\text{MH}_y\text{SPT}(G, S_a, u')$  and  $\text{MH}_z\text{SPT}(G, S \setminus S_a, u')$ ; also the cost of  $T$  is at most  $x + y + z$ . By induction, the heights of  $\text{MH}_y\text{SPT}(G, S_a, u')$  and  $\text{MH}_z\text{SPT}(G, S \setminus S_a, u')$  are  $\gamma(u', S_a, y)$  and  $\gamma(u', S \setminus S_a, z)$ , respectively. Thus, the height of  $T$  is  $d_K^x(u, u') + \max\{\gamma(u', S_a, y), \gamma(u', S \setminus S_a, z)\}$  and hence  $\gamma(u, S, j) = d_K^x(u, u') + \max\{\gamma(u', S_a, y), \gamma(u', S \setminus S_a, z)\}$ . Such a value is found by the recursive definition of  $\gamma(u, S, j)$  with  $v = u'$ ,  $S' = S_a$ ,  $j_1 = x$ ,  $j_2 = y$ , and  $j_3 = z$ .

This concludes the induction and hence the proof of the lemma.  $\square$

We are now ready to give the FPT (1, 4)-approximation algorithm.

**Theorem 2.** *There exists a (1, 4)-approximation algorithm for the BCMD problem with  $O((3^B B^3 + n + \log(Bn))Bn^2)$  running time.*

#### 4 Unit Costs and Arbitrary Weights

For the special case in which each edge has unit cost and arbitrary weight, our techniques lead to several results, that are described in the following. Observe that, in this case we are allowed to insert in  $G$  exactly  $k$  non-edges of  $G$ , where  $k = B = O(n^2)$ . We remark that **Theorem 2** gives a (1, 4)-approximation algorithm running in  $O((3^k k^3 + n)kn^2)$  time for this special case.

In all subsequent algorithms, we denote by  $C$  a clustering with  $k + 1$  clusters constructed as described in **Subsection 3.1**. We first show a  $((k + 1)^2, 3)$ -approximation algorithm.

**Theorem 3.** *Given an instance of the BCMD problem with unit costs, there exists a  $((k+1)^2, 3)$ -approximation algorithm with  $O(kn^3)$  running time.*

**Proof.** For every pair of cluster centers  $c_i, c_j \in C$  compute a shortest path in  $K$  between  $c_i$  and  $c_j$  that contains at most  $k$  non-edges of  $G$ . Add those edges to  $F$  and let  $G' = (V, E \cup F)$ . By [Theorem 1](#) and since  $k = O(n^2)$ ,  $G'$  can be constructed in  $O(kn^3)$  time. Observe that, for each pair of cluster centers, the algorithm adds at most  $k$  non-edges of  $G$  to  $F$ , thus at most  $k(k+1)^2$  non-edges in total. We prove that, for every  $v_i, v_j \in V$ , there exists a path in  $G'$  connecting  $v_i$  and  $v_j$  whose weight is at most  $3 \cdot D_{opt}^k$ . Denote by  $c_i$  and  $c_j$  the centers of the clusters  $v_i$  and  $v_j$  belong to, respectively. We have  $\text{dist}_{G'}(v_i, v_j) \leq \text{dist}_{G'}(v_i, c_i) + \text{dist}_{G'}(c_i, c_j) + \text{dist}_{G'}(c_j, v_j)$ . By [Lemma 3](#),  $\text{dist}_{G'}(v_i, c_i), \text{dist}_{G'}(v_j, c_j) \leq D_{opt}^k$ ; also, by construction,  $\text{dist}_{G'}(c_i, c_j) \leq D_{opt}^k$ , and the theorem follows.  $\square$

Next, we give a  $(k, 4)$ -approximation algorithm.

**Theorem 4.** *Given an instance of the BCMD problem with unit costs, there exists a  $(k, 4)$ -approximation algorithm with  $O(kn^2)$  running time.*

**Proof.** Pick an arbitrary cluster center, say  $c_1$ . For every cluster center  $c_j \in C \setminus \{c_1\}$ , compute a shortest path in  $K$  between  $c_1$  and  $c_j$  containing at most  $k$  non-edges of  $G$ . Add those edges to  $F$  and let  $G' = (V, E \cup F)$ . As described in [Subsection 3.1](#),  $G'$  can be constructed in  $O(kn^2)$  time. Observe that, for each cluster different from  $c_1$ , the algorithm adds at most  $k$  non-edges of  $G$  to  $F$ , thus at most  $k^2$  non-edges in total. We prove that, for every  $v_i, v_j \in V$ , there exists a path in  $G'$  connecting  $v_i$  and  $v_j$  whose weight is at most  $4 \cdot D_{opt}^k$ . Denote by  $c_i$  and  $c_j$  the centers of the clusters  $v_i$  and  $v_j$  belong to, respectively. We have  $\text{dist}_{G'}(v_i, v_j) \leq \text{dist}_{G'}(v_i, c_i) + \text{dist}_{G'}(c_i, c_1) + \text{dist}_{G'}(c_1, c_j) + \text{dist}_{G'}(c_j, v_j)$ . By [Lemma 3](#),  $\text{dist}_{G'}(v_i, c_i), \text{dist}_{G'}(v_j, c_j) \leq D_{opt}^k$ ; by construction,  $\text{dist}_{G'}(c_i, c_1), \text{dist}_{G'}(c_1, c_j) \leq D_{opt}^k$ , and the theorem follows.  $\square$

Finally, we present a  $(1, 3k+2)$ -approximation algorithm.

**Theorem 5.** *Given an instance of the BCMD problem with unit costs, there exists a  $(1, 3k+2)$ -approximation algorithm with  $O(n^2 + k^2)$  running time.*

**Proof.** For every pair of clusters  $C_i$  and  $C_j$ , with  $1 \leq i < j \leq k+1$ , let  $e_{ij}$  be the edge of minimum weight connecting a vertex in  $C_i$  with a vertex in  $C_j$ . The set of these edges is denoted  $F'$ . For a subset  $F$  of  $F'$ , we say that  $F$  spans  $C$  if the graph representing the adjacencies between clusters via the edges of  $F$  is connected. Let  $F$  be a minimum-weight set of  $k$  edges from  $F'$  spanning  $C$ . Let  $G' = (V, E \cup F)$ . The set  $F'$ , and hence the graph  $G'$ , can be constructed in  $O(n^2 + k^2)$  time as follows. For each pair of edges, consider all the edges of  $K$  and keep, for each pair of clusters, the edge with smallest weight. This can be done in  $O(n^2)$  time. Finally, compute in  $O(k^2)$  time a minimum spanning tree of the resulting graph [10], that has  $O(k)$  vertices and  $O(k^2)$  edges. Observe that the algorithm adds at most  $k$  non-edges of  $G$  to  $F$ .

We prove that, for every  $v_i, v_j \in V$ , there exists a path in  $G'$  connecting  $v_i$  and  $v_j$  whose weight is at most  $(3k+2)D_{opt}^k$ . Denote by  $P_C$  the (unique) subset of  $F$  connecting the clusters  $v_i$  and  $v_j$  belong to. Let  $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$  be the



edges of  $P_C$  in order from  $v_i$  to  $v_j$ . Then,  $\text{dist}_{G'}(v_i, v_j) \leq \text{dist}_G(v_i, x_1) + w(x_1, y_1) + \text{dist}_G(y_1, x_2) + \dots + w(x_m, y_m) + \text{dist}_G(y_m, v_j)$ . By **Lemma 3**,  $\text{dist}_G(y_i, x_{i+1}) \leq 2D_{opt}^k$ , and  $\text{dist}_G(v_i, x_1), \text{dist}_G(y_m, v_j) \leq D_{opt}^k$ . Also,  $w((x_i, y_i)) \leq D_{opt}^k$ , and the theorem follows.  $\square$

## 5 Hardness Results

The main theorem of this section provides a parameterized intractability result for BCMD with unit weights and unit costs, and some related problems. The U-BCMD problem has as input an unweighted graph  $G = (V, E)$  and two integers  $k$  and  $d$ , and the question is whether there is a set  $F \subseteq [V]^2 \setminus E$ ,  $|F| \leq k$ , such that the graph  $(V, E \cup F)$  has diameter at most  $d$ . The parameter is  $k$ . We will show that U-BCMD is  $W[2]$ -hard. We will also provide refinements to the minimum conditions required for intractability, namely U-BCMD remains NP-complete for graphs of diameter 3 with target diameter 2. We note that although Dodis and Kanna [5] provide an inapproximability reduction from SET COVER, they begin with a disconnected graph, and expand the instance with a series of size 2 sets, which does not preserve the size of the optimal solution, and therefore their reduction cannot be used to show parameterized complexity lower bounds.

**Theorem 6.** *SET COVER is polynomial-time reducible to U-BCMD. Moreover, the reduction is parameter preserving and creates an instance with diameter 3 and target diameter 2.*

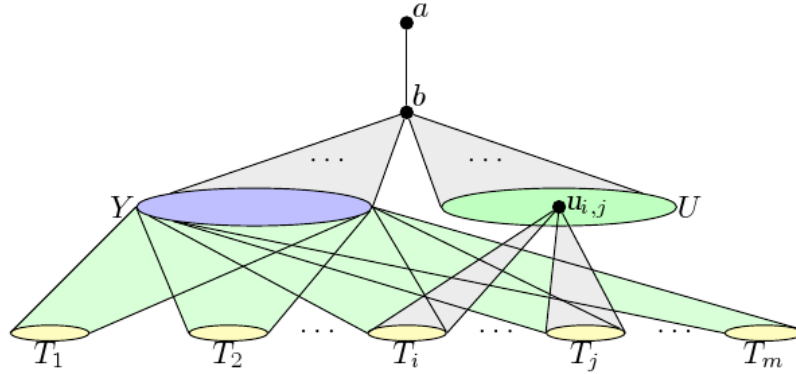
**Proof.** Let  $(X, S, k)$  be an instance of SET COVER where  $S$  is the base set and  $X \subset \mathcal{P}(S)$  is the set from which we must pick the set cover of  $S$  with size at most  $k$ . We construct an instance  $(G = (V, E), k, d)$  of U-BCMD as follows.

Let  $m = |X| \cdot k$ . The vertex set  $V$  is the disjoint union of 5 sets:

- a set  $Y$  corresponding to the set  $X$  where for each  $x \in X$  we have a vertex  $y \in Y$ ,
- a set  $T = \biguplus_{i \in [m]} T_i$  corresponding to  $S$  where for each  $s \in S$  we have a vertex  $t_i \in T_i$  for each  $i \in [m]$  (i.e we have  $m$  copies of a set of vertices corresponding to  $S$ ),
- a set  $U$  with  $\binom{m}{2}$  vertices  $u_{ij}$ , one for each pair of disjoint subsets  $T_i, T_j$  of  $T$  (where  $i \neq j$ ),
- the set  $\{a\}$ , and
- the set  $\{b\}$ .

The edge set  $E$  consists of the following edges:

- $ab$ ,
- $by$  for each vertex  $y \in Y$ ,
- $bu_{ij}$  for each vertex  $u_{ij} \in U$ ,
- $yy'$  for each pair of vertices  $y, y' \in Y$ ,
- $yt_i$  for each pair of vertices  $y \in Y$  and  $t_i \in T_i$  for each  $i \in [m]$  where the corresponding element  $s \in S$  is in the corresponding set  $x \in X$  in the SET COVER instance,
- $t_i u_{jl}$  for each pair of vertices  $t_i \in T_i$  and  $u_{jl} \in U$  such that  $i \in \{j, l\}$ , and



**Fig. 3.** Sketch of the construction for the SET COVER to U-BCMD reduction. The edge sets represented in gray are complete, the edge sets represented in light green correspond to the set membership from the SET COVER instance. The vertex sets  $Y$  and  $U$  are cliques. The vertex sets  $T_i$  are independent sets for all  $i \in [m]$ .

- $u_{ij}u_{lp}$  for each pair of vertices  $u_{ij}, u_{lp} \in U$ .

We set  $d = 2$ . Note that  $k$  in the U-BCMD instance is the same  $k$  as for the SET COVER instance. The construction is sketched in Fig. 3.

**Claim 2** For all  $v, v' \in V \setminus \{a\}$  we have  $\text{dist}(v, v') \leq 2$ .

**Claim 3** For all  $v \in V$  we have  $\text{dist}(a, v) \leq 3$ . If  $v \in T$  then  $\text{dist}(a, v) = 3$ .

Thus we are concerned only with reducing the distance between  $a$  and the vertices of  $T$ .

**Claim 4**  $(X, S, k)$  is a YES-instance of SET COVER if and only if  $(G, k, d)$  is a YES-instance of U-BCMD.

We note that the reduction is obviously polynomial-time computable, and the parameter  $k$  is preserved. The theorem now follows from the previous claims.  $\square$

**Corollary 2.** U-BCMD is NP-complete even for graphs of diameter three with target diameter two.

As SET COVER is  $W[2]$ -hard with parameter  $k$ , combined with Corollary 2 we also have the following result.

**Corollary 3.** U-BCMD is  $W[2]$ -hard even for graphs of diameter three with target diameter two.

We note additionally that as the initial graph has diameter 3 and the target diameter is 2, it is even NP-hard and  $W[2]$ -hard to decide if there is a set of  $k$  new edges that improves the diameter by one. Furthermore by taking  $a$  as source vertex, the results transfer immediately to the single-source version as discussed by Demaine & Zadimoghaddam [4].

The construction of Theorem 6 can even be extended to give a parameterized inapproximability result for U-BCMD.

**Theorem 7.** *It is  $W[2]$ -hard to compute a  $(1 + \frac{c}{k}, \frac{3}{2} - \varepsilon)$ -approximation for U-BCMD for any constants  $c$  and  $\varepsilon > 0$ .*

### Acknowledgments

SG acknowledges support from the Australian Research Council (grant DE120101761). JG is funded by the Australian Research Council FT100100755. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

### References

1. N. Alon, A. Gyarfas, and M. Ruzsinko. Decreasing the diameter of bounded degree graphs. *Journal of Graph Theory*, 35:161–172, 1999.
2. D. Bilo, L. Guala, and G. Proietti. Improved approximability and non-approximability results for graph diameter decreasing problems. *Theoretical Computer Science*, 417:12–22, 2012.
3. V. Chepoi and Y. Vaxes. Augmenting trees to meet biconnectivity and diameter constraints. *Algorithmica*, 33(2):243–262, 2002.
4. E. D. Demaine and M. Zadimoghaddam. Minimizing the diameter of a network using short-cut edges. In *Proceedings of the 12th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT)*, pages 420–431, 2010.
5. Y. Dodis and S. Khanna. Designing networks with bounded pairwise distance. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing (STOC)*, pages 750–759, 1999.
6. R. G. Downey and M. R. Fellows. *Parameterized Complexity*. Monographs in Computer Science. Springer, New York, 1999.
7. P. Erdos, A. Renyi, and V. T. Sos. On a problem of graph theory. *Studia Sci. Math. Hungar.*, 1:215–235, 1966.
8. J. Flum and M. Grohe. *Parameterized Complexity Theory*, volume XIV of *Texts in Theoretical Computer Science. An EATCS Series*. Springer, Berlin, 2006.
9. M. L. Fredman and R. E. Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the Association for Computing Machinery*, 34(3):596–615, 1987.
10. M. L. Fredman and D. E. Willard. Trans-dichotomous algorithms for minimum spanning trees and shortest paths. *Journal of Computer and System Sciences*, 48(3):533–551, 1994.
11. E. Grigorescu. Decreasing the diameter of cycles. *Journal of Graph Theory*, 43(4):299–303, 2003.
12. S. Kapoor and M. Sarwat. Bounded-diameter minimum-cost graph problems. *Theory of Computing Systems*, 41(4):779–794, 2007.
13. C.-L. Li, S. T. McCormick, and D. Simchi-Levi. On the minimum-cardinality-bounded-diameter and the bounded-cardinality-minimum-diameter edge addition problems. *Operations Research Letters*, 11(5):303–308, 1992.
14. D. Marx. Parameterized complexity and approximation algorithms. *The Computer Journal*, 51(1):60–78, 2008.
15. R. Niedermeier. *Invitation to Fixed-Parameter Algorithms*. Oxford Lecture Series in Mathematics and its Applications. Oxford University Press, Oxford, 2006.
16. A. A. Schoone, H. L. Bodlaender, and J. van Leeuwen. Diameter increase caused by edge deletion. *Journal of Graph Theory*, 11:409–427, 1997.
17. V. V. Vazirani. *Approximation algorithms*. Springer, 2001.

## A Omitted Proofs

**Lemma 1.** *Suppose that  $H$  contains a directed path  $P_H$  with weight  $W$  connecting vertices  $(u, i)$  and  $(v, j)$ , for some  $j \geq i$ . Then, there exists a  $(j - i)$ -bounded-cost path  $P_K$  in  $K$  with weight  $W$  connecting  $u$  and  $v$ .*

**Proof.** Consider a directed path  $P_H$  in  $H$  with weight  $W$  connecting vertices  $(u, i)$  and  $(v, j)$ , for some  $j \geq i$ . We define a path  $P_K$  in  $K$  as follows. Path  $P_K$  has the same number of vertices of  $P_H$ . Also, for any  $1 \leq j \leq |P_H|$ , if the  $j$ -th vertex of  $P_H$  is a vertex  $(w, i)$ , then the  $j$ -th vertex of  $P_K$  is vertex  $w$ . Observe that  $P_K$  connects vertices  $u$  and  $v$ . We prove that  $P_K$  is a  $(j - i)$ -bounded-cost path with weight  $W$ . Every edge of  $P_H$  connecting vertices  $(x, a)$  and  $(y, a)$  corresponds to an edge  $(x, y)$  of  $K$  that is an edge of  $G$  with the same weight. Moreover, every edge of  $P_H$  connecting vertices  $(x, a)$  and  $(y, b)$  with  $b > a$  either corresponds to an edge  $(x, y)$  of  $K$  that is a non-edge of  $G$  with cost  $b - a$  and with the same weight, or it corresponds to a self-loop in  $K$  with the same weight (that is 0). Hence,  $P_K$  uses non-edges of  $G$  of total cost at most  $j - i$  and total weight  $W$ . Thus,  $P_K$  is a  $(j - i)$ -bounded-cost path with weight  $W$  connecting  $u$  and  $v$ , and the lemma follows.  $\square$

**Lemma 2.** *Suppose that there exists a  $\beta$ -bounded-cost path  $P_K$  in  $K$  with weight  $W$  connecting vertices  $u$  and  $v$  and not containing any self-loop. Then, there exists a directed path  $P_H$  in  $H$  with weight  $W$  connecting vertices  $(u, 0)$  and  $(v, \beta')$  with  $\beta' \leq \beta$ .*

**Proof.** Consider a path  $P_K = \langle v_1, v_2, \dots, v_m \rangle$  in  $K$  with weight  $W$  not containing any self-loop. Set  $(v_1, 0)$  to be the first vertex of  $P_H$ . Suppose that path  $P_H$  has been defined until a vertex  $(v_h, j)$ , corresponding to vertex  $v_h$  of  $P_K$ . If edge  $(v_h, v_{h+1})$  of  $P_K$  is an edge of  $G$ , then let  $(v_{h+1}, j)$  be the vertex corresponding to  $v_{h+1}$ . If edge  $(v_h, v_{h+1})$  of  $P_K$  is a non-edge of  $G$ , then let  $(v_{h+1}, j + c(\{v_h, v_{h+1}\}))$  be the vertex corresponding to  $v_{h+1}$ . Observe that  $P_H$  is a directed path connecting  $(u, 0)$  and  $(v, \beta')$ . We prove that  $\beta' \leq \beta$  and that  $P_H$  has weight  $W$ . Every edge  $(x, y)$  of  $P_K$  that is an edge of  $G$  corresponds to an edge of  $H$  connecting vertices  $(x, a)$  and  $(y, a)$  with the same weight. Moreover, every edge  $(x, y)$  of  $P_K$  that is a non-edge of  $G$  corresponds to an edge of  $H$  connecting vertices  $(x, a)$  and  $(y, b)$ , with  $c\{x, y\} = b - a$  and with the same weight. By definition,  $P_K$  uses non-edges of  $G$  of total cost at most  $\beta$ . Hence,  $\beta' \leq \beta$ ; also,  $P_H$  has weight exactly  $W$  and the lemma follows.  $\square$

**Corollary 1.** *There is a  $\beta$ -bounded-cost path connecting vertices  $u$  and  $v$  in  $K$  with weight  $W$  if and only if there is a directed path in  $H$  connecting vertices  $(u, 0)$  and  $(v, \beta)$  with weight  $W$ .*

**Proof.** For the necessity, if there is a  $\beta$ -bounded-cost path connecting  $u$  and  $v$  in  $K$  with weight  $W$ , then there is also a  $\beta$ -bounded-cost path connecting  $u$  and  $v$  in  $K$  with weight  $W$  and not containing any self-loop. By Lemma 2, there is a path  $P_H$  in  $H$  connecting vertices  $(u, 0)$  and  $(v, \beta')$  with weight  $W$ , for some  $\beta' \leq \beta$ . Also, edges with weight 0 connecting  $(v, j)$  and  $(v, j + 1)$  exist for every  $v \in V$  and every  $\beta' \leq j \leq \beta - 1$ , and the necessity follows. For the sufficiency, if there is a path in  $H$  connecting vertices  $(u, 0)$  and  $(v, \beta)$  with weight  $W$ , then by Lemma 1 there is a  $\beta$ -bounded-cost path connecting  $u$  and  $v$  in  $K$  with weight  $W$ .  $\square$

**Claim 1.** *Let  $G'$  be a weighted graph and let  $C'$  be a set of vertices in  $G'$  such that  $\text{dist}_{G'}(C') > D$ . Then, for every graph  $G''$  obtained from  $G'$  by adding a single non-edge of  $G'$  with non-negative weight, there is a set  $C'' \subset C'$  with  $|C''| = |C'| - 1$  and with  $\text{dist}_{G''}(C'') > D$ .*

**Proof.** Let  $(u, v)$  denote the edge that is added to  $G'$  to obtain  $G''$ . For the sake of contradiction, assume that there is no vertex  $w \in C'$  such that  $\text{dist}_{G''}(C' \setminus \{w\}) > D$ . That is, every set  $C'' \subset C'$  with  $|C''| = |C'| - 1$  contains two vertices whose distance is at most  $D$ . Then, there are four vertices  $w_1, w_2, w_3, w_4 \in C$  such that  $\text{dist}_{G''}(w_1, w_2) \leq D$  and  $\text{dist}_{G''}(w_3, w_4) \leq D$ , or there are three vertices  $w_1, w_2, w_3 \in C$  such that  $\text{dist}_{G''}(w_1, w_2) \leq D$ ,  $\text{dist}_{G''}(w_1, w_3) \leq D$ , and  $\text{dist}_{G''}(w_2, w_3) \leq D$ .

In the first case, since  $\text{dist}_{G''}(w_1, w_2) < \text{dist}_{G'}(w_1, w_2)$  and  $\text{dist}_{G''}(w_3, w_4) < \text{dist}_{G'}(w_3, w_4)$ , we have that  $(u, v)$  is an edge of any shortest path  $P_{1,2}$  from  $w_1$  to  $w_2$  and of any shortest path  $P_{3,4}$  from  $w_3$  to  $w_4$ . Assume, without loss of generality, that  $u$  is encountered before  $v$  when traversing  $P_{1,2}$  starting at  $w_1$  and when traversing  $P_{3,4}$  starting at  $w_3$  (otherwise swap  $w_1$  and  $w_2$  and/or  $w_3$  and  $w_4$ ). Therefore, we get (1A)  $\text{dist}_{G'}(w_1, u) + \text{dist}_{G'}(v, w_2) \leq D$ , and (1B)  $\text{dist}_{G'}(w_3, u) + \text{dist}_{G'}(v, w_4) \leq D$ . However, since  $\text{dist}_{G'}(C') > D$ , we have (1C)  $\text{dist}_{G'}(w_1, u) + \text{dist}_{G'}(u, w_3) > D$ , and (1D)  $\text{dist}_{G'}(w_2, v) + \text{dist}_{G'}(v, w_4) > D$ . Denote  $K := \text{dist}_{G'}(w_1, u) + \text{dist}_{G'}(v, w_2) + \text{dist}_{G'}(w_3, u) + \text{dist}_{G'}(v, w_4)$ . Inequalities (1A) and (1B) give  $K \leq 2D$ , while inequalities (1C) and (1D) give  $K > 2D$ , a contradiction.

In the second case, denote by  $P_{1,2}$ ,  $P_{1,3}$ , and  $P_{2,3}$  three paths in  $G''$  with weight at most  $D$  connecting  $w_1$  and  $w_2$ , connecting  $w_1$  and  $w_3$ , and connecting  $w_2$  and  $w_3$ , respectively. Since  $\text{dist}_{G'}(\{w_1, w_2, w_3\}) > D$ , all these paths use edge  $(u, v)$ . Without loss of generality, assume  $\text{dist}_{G'}(w_1, u) \leq \text{dist}_{G'}(w_1, v)$ . Hence, both  $P_{1,2}$  and  $P_{1,3}$  reach  $u$  before  $v$  when traversing such paths starting at  $w_1$ . Without loss of generality, assume that  $P_{2,3}$  reaches  $u$  before  $v$  when traversing such path starting at  $w_2$  (otherwise, swap  $w_2$  and  $w_3$ ). Therefore, we get (2A)  $\text{dist}_{G'}(w_1, u) + \text{dist}_{G'}(v, w_2) \leq D$ , and (2B)  $\text{dist}_{G'}(w_2, u) + \text{dist}_{G'}(v, w_3) \leq D$ . However, since  $\text{dist}_{G'}(\{w_1, w_2, w_3\}) > D$ , we have (2C)  $\text{dist}_{G'}(w_2, v) + \text{dist}_{G'}(v, w_3) > D$ , and (2D)  $\text{dist}_{G'}(w_1, u) + \text{dist}_{G'}(u, w_2) > D$ . Denote  $L := \text{dist}_{G'}(w_1, u) + \text{dist}_{G'}(v, w_2) + \text{dist}_{G'}(w_2, u) + \text{dist}_{G'}(v, w_3)$ . Inequalities (2A) and (2B) give  $L > 2D$ , while inequalities (2C) and (2D) give  $L \leq 2D$ , a contradiction. This concludes the proof of the claim.  $\square$

**Theorem 2.** *There exists a  $(1, 4)$ -approximation algorithm for the BCMD problem with  $O((3^B B^3 + n + \log(Bn))Bn^2)$  running time.*

**Proof.** Given an instance  $(G, w, c, B)$  of the BCMD problem, by **Theorem 1** we can determine, for every pair of vertices  $u, v \in V$  and for every  $1 \leq j \leq B$ , the minimum weight of a  $j$ -bounded cost path connecting  $u$  and  $v$  in total  $O((n + \log(Bn))Bn^2)$  time. By **Lemma 3**, a clustering of  $G$  can be computed in  $O(Bn^2)$  time. Due to **Lemma 6**, the problem of computing a  $\text{MH}_{B\text{SPT}}(G, C \setminus \{c_1\}, c_1)$  can be solved by dynamic programming over the triples  $(u, S, j)$  (there are  $O\left(B\binom{B+1}{s}n\right)$  such triples with  $|S| = s$ ); the computation of the value for such a triple requires to take a minimum over  $j^3 2^{|S|} n$  values, hence the dynamic programming running time is  $O\left(nB \sum_{s=0}^B \binom{B+1}{s} B^3 2^s n\right) = O(B^4 3^B n^2)$ . By **Lemma 5**, augmenting  $G$  with the non-edges that are present in a

$\text{MH}_B\text{SPT}(G, C \setminus \{c_1\}, c_1)$  yields a  $B$ -augmentation  $G_B$  whose diameter is at most  $4 \cdot D_{opt}^B$ .  $\square$

**Claim 2.** For all  $v, v' \in V \setminus \{a\}$  we have  $\text{dist}(v, v') \leq 2$ .

**Proof.** The vertices of  $U$  are at distance one from each other. The vertices of  $Y$  are at distance one from each other.

If one of the vertices is  $b$ , clearly  $b$  is at distance 1 from the vertices of  $Y$  and  $U$ . Therefore the vertices of  $U$  and  $Y$  are at most distance 2 from each other via the path through  $b$ .

Each vertex  $t \in T$  is at distance one from some vertex  $y \in Y$ . As  $Y$  is a clique,  $t$  is at distance at most two from all the vertices in  $Y$ .

Each vertex  $t \in T$  is at distance one from some vertex  $u \in U$ . As  $U$  is also a clique,  $t$  is at distance at most two from all vertices of  $U$ .

For each pair of vertices  $t_i \in T_i$  and  $t_j \in T_j$  there is a vertex  $u_{ij} \in U$  such that  $t_i u_{ij} \in E$  and  $t_j u_{ij} \in E$ . If  $i = j$  then any vertex  $u_{ik} \in U$  will suffice. Thus all the vertices of  $T$  are at most distance 2 from each other.  $\square$

**Claim 3.** For all  $v \in V$  we have  $\text{dist}(a, v) \leq 3$ . If  $v \in T$  then  $\text{dist}(a, v) = 3$ .

**Proof.** As the distance from  $b$  to all other vertices is at most 2, the distance from  $a$  to all other vertices is at most 3. Moreover, as the distance from  $b$  to the vertices of  $U$  and  $Y$  is one, the distance from  $a$  to these vertices is two. Therefore the only vertices at distance three from  $a$  are the vertices of  $T$ .  $\square$

**Claim 4.**  $(X, S, k)$  is a YES-instance of SET COVER if and only if  $(G, k, d)$  is a YES-instance of U-BCMD.

**Proof.** Let  $X' \subseteq X$  be the set cover that witnesses that  $(X, S, k)$  is a YES-instance of SET COVER. Let  $Y' \subseteq Y$  be the set of vertices that corresponds to  $X'$ . We have  $|Y'| = |X'| \leq k$ . If we add the edges  $ay$  for all  $y \in Y'$ , then  $a$  is at distance at most 2 from all vertices  $t \in T$ . As  $X'$  is a set cover of  $S$ , for each  $s \in S$  there is at least one set  $x \in X'$  such that  $s \in x$ . Then there is an edge from  $a$  to the vertex  $y$  corresponding to  $x$ , and by the construction,  $y$  is adjacent to  $t \in T$  if and only if the corresponding element  $s$  is in  $S$ , thus we have a path  $a \rightsquigarrow y \rightsquigarrow t$ .

Now, assume  $(G, k, d)$  is a YES-instance of U-BCMD. First consider the case where all the edges are added between  $a$  and the vertices of  $Y$ . Then the set  $Y' \subseteq Y$  of vertices newly adjacent to  $a$  corresponds to a set cover  $X' \subseteq X$  in the same way as before.

We must demonstrate that we may only (productively) add edges between  $a$  and  $Y$ . Clearly we cannot add the edge  $ab$ , as it already exists, and clearly adding edges from  $b$  to other vertices is not necessary, as we could simply add the edges directly to  $a$ . Suppose we add edges between  $a$  and  $T$  directly, as there are  $|S| \cdot m$  vertices in  $T$ , we clearly cannot reduce the distance between  $a$  and all the vertices in  $T$ , so some edges must still be added elsewhere. If we add edges between  $a$  and  $U$  we can reduce the distance between  $a$  and at most  $2k$  vertices of  $T$ . Thus even were we to add such edges, there is at least one  $T_i \subset T$  still at distance 3 from  $a$ . Therefore we must add edges from  $a$  to  $Y$  such that  $T_i$  is dominated by this subset of  $Y$ . Clearly this corresponds to a set cover of  $S$ .  $\square$

**Corollary 2.** U-BCMD is NP-complete even for graphs of diameter three with target diameter two.

**Proof.** As it is already known that U-BCMD is in NP [5], the result for U-BCMD follows from Theorem 6.  $\square$

**Theorem 7.** It is  $W[2]$ -hard to compute a  $(1 + \frac{c}{k}, \frac{3}{2} - \varepsilon)$ -approximation for U-BCMD for any constants  $c$  and  $\varepsilon > 0$ .

**Proof.** We repeat the construction of Theorem 6, except that we introduce  $c + 1$  copies of the  $Y$  and  $T$  components and set  $k' = k \cdot (c + 1)$ . Let  $Y_i$  with  $1 \leq i \leq c + 1$  be the copies of the  $Y$  components and  $T_{i,j}$  with  $1 \leq i \leq c + 1$  and  $1 \leq j \leq m$  be the copies of the  $T$  components. The edges are similar to the previous construction; we highlight the differing edges:

- $yy'$  for each  $y, y' \in \bigcup_i Y_i$ ,
- $y_i t_{i,j}$  for each  $y_i \in Y_i$  and  $t_{i,j} \in T_{i,j}$  where the corresponding element  $x \in X$  is in the corresponding set  $S$ ,
- $by$  for all  $y \in \bigcup_i Y_i$ , and
- $t_{a,i} u_{i,j}$  for each vertex  $t_{a,i} \in \bigcup_a T_{a,i}$  and each vertex  $u_{i,j} \in U$ .

Then apart from  $a$ , all vertices remain at pairwise distance 2, with  $a$  at distance 3 from vertices in  $\bigcup_i Y_i$ . Clearly to reduce the diameter to 2 we require the addition of edges from  $a$  to vertices of the  $Y$  component copies as before, furthermore we require edges to each copy, otherwise there is some  $T_{i,j}$  component that remains at distance 3 from  $a$ . Thus if the SET COVER instance has a solution of size  $k$ , the U-BCMD instance has a solution of size  $k'$ .

Let  $F$  be a set of  $(1 + \frac{c}{k'})k' = k' + c$  edges such that the diameter of  $G' = (V, E \cup F)$  is at most  $(\frac{3}{2} - \varepsilon) \cdot 2$ . Since the diameter of  $G'$  is integral, it has diameter at most 2. Since there are  $c + 1$  copies of  $Y$ , at least one of them,  $Y_i$ , has at most  $k'$  vertices adjacent to  $a$ , giving a size  $k'$  set cover as before.  $\square$