Theoretical Computer Science 417 (2012) 12–22

Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Improved approximability and non-approximability results for graph diameter decreasing problems☆

Davide Bilò[a], Luciano Gualà[b], Guido Proietti[a,c,*]

[a] Dipartimento di Informatica, Università di L'Aquila, 67010 L'Aquila, Italy
[b] Dipartimento di Matematica, Università di Tor Vergata, 00133 Roma, Italy
[c] Istituto di Analisi dei Sistemi ed Informatica, CNR, 00185 Roma, Italy

ARTICLE INFO

Keywords:
Approximation algorithms
Graphs
Edge addition
Diameter

ABSTRACT

In this paper, we study two variants of the problem of adding edges to a graph so as to reduce the resulting diameter. More precisely, given a graph $G = (V, E)$, and two positive integers D and B, the *Minimum-Cardinality Bounded-Diameter Edge Addition* (MCBD) problem is to find a minimum-cardinality set F of edges to be added to G in such a way that the diameter of $G + F$ is less than or equal to D, while the *Bounded-Cardinality Minimum-Diameter Edge Addition* (BCMD) problem is to find a set F of B edges to be added to G in such a way that the diameter of $G + F$ is minimized. Both problems are well known to be NP-hard, as well as approximable within $O(\log n \log D)$ and 4 (up to an additive term of 2), respectively. In this paper, we improve these long-standing approximation ratios to $O(\log n)$ and to 2 (up to an additive term of 2), respectively. As a consequence, we close, in an asymptotic sense, the gap on the approximability of MCBD, which was known to be not approximable within $c \log n$, for some constant $c > 0$, unless P = NP. Remarkably, as we further show in the paper, our approximation ratio remains asymptotically tight even if we allow for a solution whose diameter is optimal up to a multiplicative factor approaching $\frac{5}{3}$. On the other hand, on the positive side, we show that at most twice of the minimal number of additional edges suffices to get at most twice of the required diameter. Some of our results extend to the edge-weighted version of the problems.

1. Introduction

In this paper, we study two basic network design problems. In the first one, we are given a network and a distance requirement D. The goal is to find a minimum-cardinality set of links to be added to the network so that every pair of its nodes is connected by a path of at most D links. More formally

MINIMUM-CARDINALITY BOUNDED-DIAMETER EDGE ADDITION (MCBD)
Instance: an undirected graph $G = (V, E)$ and a value $D \in \mathbb{Z}^+$.
Goal: find a minimum-cardinality set F of edges to be added to G such that the *diameter* of $G + F = (V, E \cup F)$ is less than or equal to D.

☆ A preliminary version of this work has appeared in Proceedings of the *35th International Symposium on Mathematical Foundations of Computer Science (MFCS 2010)*, Lecture Notes in Computer Science 6281, Springer, pp. 150–161. This work was partially supported by the PRIN 2008 research project COGENT (*COmputational and GamE-theoretic aspects of uncoordinated NeTworks*), funded by the Italian Ministry of Education, University, and Research.

* Corresponding author at: Dipartimento di Informatica, Università di L'Aquila, 67010 L'Aquila, Italy. Tel.: +39 0862 433727.
E-mail addresses: davide.bilo@univaq.it (D. Bilò), guala@mat.uniroma2.it (L. Gualà), guido.proietti@univaq.it (G. Proietti).

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2011.05.014

Similarly, one can define the specular problem in which we are given a network and a *budget* B on the number of addable links, and the goal is to add such links so that the resulting network has minimum (in terms of number of links) diameter. More formally

Bounded-Cardinality Minimum-Diameter Edge Addition (BCMD)
Instance: an undirected graph G and a value $B \in \mathbb{Z}^+$.
Goal: find a set F of B edges to be added to G such that the diameter of $G + F$ is minimized.

These two problems arise in practical applications like telecommunication networks and airplane flights scheduling [7,12], but they also received a lot of attention in the graph theory community (see [1,8,11,14,23,25]).

Notice that the two defined problems are the optimization version of the same underlying decision problem. Therefore, for the sake of unifying the exposition, we will denote by B the cardinality of an optimal solution for MCBD, and by D the value of an optimal solution for BCMD.

Having this in mind, and following standard terminology on *bicriteria* optimization problems, for $\beta, \delta \geq 1$, a (β, δ)-approximation algorithm for BCMD will denote an algorithm which can select a set F of additional edges whose size is at most β times the budget B, and returns a graph $G + F$ of diameter less than or equal to δD (where D is the value of the diameter of an optimal solution for BCMD with budget B). Symmetrically, a (δ, β)-approximation algorithm for MCBD will denote an algorithm that returns a graph $G + F$ whose diameter is at most δ times the required value D, by using at most βB edges (where B is the size of an optimal solution for MCBD with required diameter D). Observe that a (β, δ)-approximation algorithm for BCMD is a (δ, β)-approximation algorithm for MCBD, and vice versa.[1]

1.1. Related work and our results

In the rest of the paper we will denote by n the number of vertices of G. For $D = 1$ MCBD is clearly in P, while in papers [10,20] it was proven that for $D = 2, 3, 4$, MCBD is not approximable within $c \log n$ for some constant $c > 0$, unless P = NP. This inapproximability result holds in fact for any fixed value of $D \geq 5$, by an easy extension of the NP-hardness result for MCBD given by Chepoi and Vaxès [6]. As a consequence, there exists no $(c \log n, \delta)$-approximation algorithm for BCMD for $\delta < 1 + 1/D$, unless P = NP.

On the positive side, BCMD admits a constant $(4 + \frac{2}{D})$-approximation algorithm [20]; the same algorithm guarantees a $(2 + \frac{2}{D})$-approximation for *forests*. Concerning positive results for MCBD, Dodis and Khanna [10] provide an $O(\log n \log D)$-approximation algorithm as well as both approximability and non-approximability results for a more general version of MCBD in which edges are associated with a cost and a length function, and B and D are redefined accordingly. Furthermore, some variants of MCBD have been studied by Chepoi and Vaxès [6]. Here, a 2-approximation algorithm for forests is given for even values of D, while for odd values an 8-approximation algorithm has been given by Ishii et al. [16]. This latter result has been improved by in paper [5], where a $(2 + \epsilon)$-approximation algorithm up to an additive constant of $O(\epsilon^{-5})$, for every $\epsilon > 0$, has been given. Establishing whether BCMD and MCBD restricted to trees/forests are in P is still an open problem. Furthermore, concerning bicriteria approximation algorithms, Dodis and Khanna [10] provide a polynomial time $\left(O(\log n), 2 + \frac{2}{D}\right)$-approximation algorithm for BCMD. The same results has been proved also by Kapoor and Sarwat [18]. Finally, Meyerson and Tagiku [21] study a problem related to BCMD on edge-weighted graphs, where the objective is that of minimizing the average distance.

Graph theory community addressed the problems for paths and cycles. Chung and Garey [8] provide lower and upper bounds on the value D of the diameter when B edges are added to a graph. For paths, they show that $\frac{n}{B+1} - 1 \leq D < \frac{n}{B+1} + 3$, while for cycles they show that $\frac{n}{B+1} - 1 \leq D \leq \frac{n}{B+1} + 3$ if B is odd, and $\frac{n}{B+2} - 1 \leq D < \frac{n}{B+2} + 3$ otherwise. Alon et al. [1] provide lower and upper bounds on the number B of edges to add to a cycle to obtain a graph of diameter at most D. They show that $\lfloor \frac{n}{D-1} \rfloor - 7 \leq B \leq \lfloor \frac{n}{D-1} \rfloor$ if D is even, and $\lfloor \frac{n}{D-2} \rfloor - 155 \leq B \leq \lfloor \frac{n}{D-2} \rfloor$ otherwise. All the above upper bounds are obtained via polynomial time algorithms. This implies better approximations for paths and cycles.

In this paper, we provide a different analysis of the algorithm of Li et al. [20], in order to show that it actually computes a $\left(2 + \frac{2}{D}\right)$-approximate solution for BCMD. Moreover, when the input instance is a forest, we achieve optimality up to small constant additive terms. More precisely, we get an approximation guarantee of $(1 + \frac{2}{D})$ for even values of D, and of $(1 + \frac{4}{D})$ for odd values of D. Concerning approximability of MCBD, we improve the result given by Dodis and Khanna [10], by providing an $O(\log n)$-approximation algorithm. Thus, we close in an asymptotic sense the approximability of MCBD. Notably, our algorithm extends to directed graphs, as well as to the case where we place the distance requirements $D_{u,v}$ for each pair u, v of vertices. We regard our result as a significative contribution for the comprehension of MCBD since, as we further show in the paper, our approximation ratio cannot be improved asymptotically, unless P = NP, even if we allow for a solution whose diameter is optimal up to a multiplicative factor of $\frac{5}{3} - \frac{7-(D+1) \bmod 3}{3D}$. Notice that this also implies a better inapproximability threshold for BCMD for any $D \geq 6$. On the other hand, on the positive side, we also show that if a doubling of the optimal diameter is tolerated, then MCBD admits a $\left(2 - \frac{1}{B}\right)$-approximation algorithm. Table 1 summarizes the currently best known results for the two problems (results are given in form of bicriteria ratios for BCMD, and our contributions are written in bold).

[1] Indeed, we can always guess the value of an optimal solution for the problem.

Table 1
Table of currently best known results for (β, δ)-approximation algorithms for BCMD. The non-approximability results hold for some constant $c > 0$ unless P = NP. The question mark means that the entry of the table is still an open problem.

Input instance	Approximability	Non-approximability
General	$\left(\mathbf{2 - \frac{1}{B}, 2}\right)$ $\left(\mathbf{1, 2 + \frac{2}{D}}\right)$ $\left(\boldsymbol{O(\log n), 1}\right)$	$\left(c \log n, \delta < 1 + \frac{1}{D}\right), \forall D \geq 2$ [6,10,20] $\left(\boldsymbol{c \log n, \delta < \frac{5}{3} - \frac{7-(D+1) \bmod 3}{D}}\right), \boldsymbol{\forall D \geq 6}$
Forests	(2, 1) for $D = 2h$ [6] $(2 + \epsilon, 1)$ for $D = 2h + 1$ [5] $\left(\mathbf{1, 1 + \frac{2}{D}}\right)$ **for** $\boldsymbol{D = 2h}$ $\left(\mathbf{1, 1 + \frac{4}{D}}\right)$ **for** $\boldsymbol{D = 2h + 1}$	?

Some of our results extend to the edge-weighted version of BCMD (WBCMD for short), where each edge e of the graph in input has a non-negative real weight $w(e)$ associated with it, all edges to be added to the graph have weight equal to $\omega \geq 0$, and distances are measured w.r.t. edge weights.[2] More precisely, we provide $\left(2 + \frac{2\omega}{D}\right)$- and $\left(2 - \frac{1}{B}, 2\right)$-approximation algorithms. Moreover, we prove an inapproximability result of $(c \log n, \delta)$ for some constant c and for every $\delta < 2 - \frac{3\omega}{D}$, unless P = NP. Observe that our approximation ratio matches the lower bound for $\omega = 0$. Weaker results for WBCMD have been found independently by Demaine and Zadimoghaddam [9]. Moreover, in the same paper the authors also give an $(O(\log n), 1 + \epsilon)$-approximation algorithm for the edge-weighted version of the MCBD.

The paper is organized as follows. We give preliminary definitions in Section 2. The $\left(2 + \frac{2}{D}\right)$-approximation algorithm for BCMD is described in Section 3, and Section 4 is devoted to the approximation for MCBD. Sections 3 and 4 also contain some conjectures that would imply better approximability results. Bicriteria approximability and non-approximability results are given in Section 5. Finally, in Section 6, we provide the results for WBCMD.

2. Basic definitions

A *graph* G is a pair $G = (V(G), E(G))$, where $V(G)$ is the set of *vertices* of G and $E(G)$ is the set of *edges* of G. The set $E(G)$ is a subset of all the unordered pairs of distinct vertices in G, i.e., $E(G) \subseteq \{(u, v) \mid u, v \in V(G), u \neq v\}$. Moreover, we say that u and v are the *endvertices* of the edge (u, v), as well as that (u, v) is *adjacent* to u and to v, respectively. Furthermore, for two given vertices $u, v \in V(G)$ we say that u is a *neighbor* of v if and only if $(u, v) \in E(G)$. We denote by $\bar{G}$ the *complement* of G, i.e., the graph with $V(\bar{G}) := V(G)$ and $E(\bar{G}) := \{(u, v) \mid u, v \in V(G), u \neq v\} \setminus E(G)$. A graph G' is a *subgraph* of G if $V(G') \subseteq V(G)$ and $E(G') \subseteq E(G)$. For every $F \subseteq \{(u, v) \mid u, v \in V(G), u \neq v\}$, we denote by $G + F$ the graph with vertex set equal to $V(G)$ and edge set equal to $E(G) \cup F$. We use $G + e$ instead of $G + \{e\}$. For a given subset $U \subseteq V(G)$ we denote by $G[U]$ the subgraph of G *induced* by U, i.e., the graph with vertex set $V(G[U]) := U$ and edge set $E(G[U]) := \{(u, v) \mid u, v \in U, (u, v) \in E(G)\}$.

A *path* P in a graph G is an alternating sequence of vertices and edges $(v_1, e_1, \ldots, v_h, e_h, v_{h+1})$ (not necessarily distinct) such that $e_i = (v_{i-1}, v_i) \in E(G)$ for every $i = 1, \ldots, h$. The vertices v_1 and v_{h+1} are the *endvertices* of the path P, or equivalently, P is path from v_1 to v_{h+1} in G. Moreover, the *length* of P is equal to h. A path is *simple* if $v_i \neq v_j$ for every $i \neq j$. A *shortest path* between two vertices u and v of a given graph G is a path from u to v in G of minimum length. By $d_G(u, v)$ we denote the *distance* between u and v in G, i.e., the length of a shortest path from u to v. The *diameter* of G will be denoted by $\mathtt{diam}(G) := \max_{u,v \in V(G)} d_G(u, v)$. Moreover, for a given $U \subseteq V(G)$, we denote by $r_G(U) := \min_{v \in V(G)} \max_{u \in U} d_G(v, u)$ the *radius* of U in G. A graph is *connected* if for every $u, v \in V(G)$ there exists a path in G having u and v as endvertices.

Let $\lambda \in \mathbb{Z}^+$. We denote by G^λ the graph with vertex set $V(G)$ and edge set $E(G^\lambda) = \{(u, v) \mid u, v \in V(G), d_G(u, v) \leq \lambda\}$. A set $\{v_1, \ldots, v_\ell\}$ of vertices of G is an *independent set* of G if and only if $(v_i, v_j) \notin E(G)$ for every $i, j = 1, \ldots, \ell$. An independent set $\{v_1, \ldots, v_\ell\}$ of G is maximum if for every independent set $\{v'_1, \ldots, v'_{\ell'}\}$ of G we have that $\ell \geq \ell'$. By $\alpha(G)$ we denote the cardinality of a maximum independent set of G. A graph G is a *clique* if $V(G)$ is an independent set in $\bar{G}$.

For a given positive integer $k > 0$, and a subset $U \subseteq V(G)$, we say that $\langle V_1, \ldots, V_k \rangle$ is a *k-clustering* of U if (i) $\forall i, V_i \subseteq V(G)$, and (ii) $U \subseteq \bigcup_{i=1}^k V_i$. We say that $z_i \in V(G)$ is *a center* of *cluster* V_i if $\max_{v \in V_i} d_G(z_i, v) = r_G(V_i)$. For two given $j, j' \in \mathbb{Z}^+, j \leq j'$, we denote by $[j, j']$ the set $\{j'' \mid j'' \in \mathbb{Z}^+, j \leq j'' \leq j'\}$ and by $(j, j']$ the set $\{j'' \mid j'' \in \mathbb{Z}^+, j < j'' \leq j'\}$.

3. Approximation algorithms for BCMD

We begin this section by describing the $\left(4 + \frac{2}{D}\right)$-approximation algorithm for BCMD of Li et al. [20] and show that it actually computes a $\left(2 + \frac{2}{D}\right)$-approximate solution. This algorithm uses an algorithm for the *k-center Problem* as a black

[2] The addition of parallel edges is allowed.

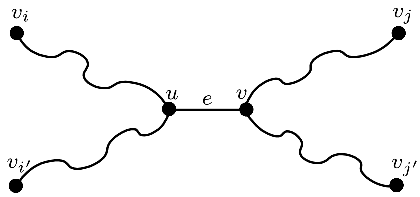

Fig. 1. Case (i) of the proof of Lemma 2.

box. The *k-center Problem* is the problem that takes a graph G and an integer $k \in \mathbb{Z}^*$ as inputs and asks for a k-clustering $\langle V_1, \dots, V_k \rangle$ of $V(G)$ that minimizes $\max_{i\in[1,k]} r_G(V_i)$. It is well known that this problem cannot be approximated within a factor better than 2, unless $\mathsf{P} = \mathsf{NP}$ [22]. On the other hand, 2-approximation algorithms are given in papers [13,15].

The algorithm of Li et al. [20] uses any algorithm $\mathcal{A}$ for the k-center problem on input G and $k = B + 1$ to find a $(B + 1)$-clustering $\langle V_0, \dots, V_B \rangle$ of $V(G)$. Then, it computes a center z_i for every cluster V_i and outputs the set $F = \{(z_0, z_i) \mid i \in [1, B]\}$. We prove that this algorithm computes a $\left(2 + \frac{2}{D}\right)$-approximate solution when $\mathcal{A}$ is the 2-approximation algorithm of Gonzalez [13]. In the following we provide the description of a variant of the Gonzalez algorithm (`Gonzalez` for short) that computes a k-clustering of a subset of vertices of the graph.

The algorithm `Gonzalez` takes as input a graph H, a set of vertices $U \subseteq V(H)$, and a value $k \in \mathbb{Z}^+$. The algorithm returns a k-clustering $\langle V_1, \dots, V_k \rangle$ of U which is computed in two steps. In the first step, k vertices $v_1, \dots, v_k \in U$ (not necessarily distinct) are computed, and in the second step, the k-clustering $\langle V_1, \dots, V_k \rangle$ of U is computed using information gathered in the first step. The first step goes as follows. Vertex $v_1 \in U$ is chosen arbitrarily. Then, for every $i = 2, \dots, k$, vertex $v_i \in U$ is a vertex that maximizes the minimum distance from $v_1, \dots, v_{i-1}$, i.e., $v_i \in \arg\max_{v\in U} \min_{j\in[1,i-1]} d_H(v_j, v)$. The second step goes as follows. At the beginning $V_i = \{v_i\}$ for every i. Then, every vertex $v \in U \setminus \{v_1, \dots, v_k\}$ is assigned to the cluster associated with the closest of the v_i's, i.e., vertex v is added to cluster V_i if $d_H(v_i, v) \le d_H(v_j, v)$, for every $j \in [1, k], j \neq i$. Ties are broken arbitrarily.

The subsequent lemmas are the key of our proof.

Lemma 1. *Let H be a graph, let $U \subseteq V(H)$, and let $\lambda \in \mathbb{Z}^+$. If $\alpha(H^\lambda[U]) \le k$,*[3] *then the algorithm* `Gonzalez` *on input H, U, and k computes a k-clustering $\langle V_1, \dots, V_k \rangle$ of U such that $r_H(V_i) \le \lambda$, $\forall i \in [1, k]$.*

Proof. We first show by contradiction that in H with $\alpha(H^\lambda[U]) \le k$, we have $\max_{v\in U} \min_{j\in[1,k]} d_H(v_j, v) \le \lambda$. For the sake of contradiction, assume that $\max_{v\in U} \min_{j\in[1,k]} d_H(v_j, v) > \lambda$. Then, by construction, $d_H(v_j, v_{j'}) > \lambda$ for every $j, j' \in [1, k], j \neq j'$. Let $u \in U$ be any vertex such that $u \in \arg\max_{v\in U} \min_{j\in[1,k]} d_H(v_j, v)$. We have that $\{v_1, \dots, v_k\} \cup \{u\}$ is an independent set of $H^\lambda[U]$ of cardinality $k + 1$, thus contradicting the assumption that $\alpha(H^\lambda[U]) \le k$. As a consequence of the above claim, we have that

$$r_H(V_i) \le \max_{v\in V_i} d_H(v_i, v) \le \max_{v\in U} \min_{j\in[1,k]} d_H(v_j, v) \le \lambda$$

for every $i \in [1, k]$. $\square$

Lemma 2. *Let H be a graph and let $\lambda \in \mathbb{Z}^+$. For every $e \in E(\bar{H})$, $\alpha((H + e)^\lambda) \ge \alpha(H^\lambda) - 1$.*

Proof. Let $v_1, \dots, v_\ell$ be an independent set in H^λ. For the sake of contradiction, assume there exists an edge $e = (u, v) \in E(\bar{H})$ such that $\alpha((H + e)^\lambda) < \ell - 1$. This implies that (i) there exist four distinct indexes $i, i', j, j' \le \ell$ such that $(v_i, v_j), (v_{i'}, v_{j'}) \in E((H + e)^\lambda)$ or (ii) there exist three distinct indexes $i, i', i'' \le \ell$ such that $(v_i, v_{i'}), (v_{i'}, v_{i''}), (v_{i''}, v_i) \in E((H + e)^\lambda)$.

We deal with case (i) first. Since $d_H(v_i, v_j), d_H(v_{i'}, v_{j'}) > \lambda$ and since $d_{H+e}(v_i, v_j), d_{H+e}(v_{i'}, v_{j'}) \le \lambda$, then, both the shortest path from v_i and v_j in $H + e$ and the shortest path from $v_{i'}$ and $v_{j'}$ in $H + e$ pass through edge e. Therefore, without loss of generality, we have that (see also Fig. 1)

$$\begin{cases} d_H(v_i, u) + 1 + d_H(v_j, v) \le \lambda \\ d_H(v_{i'}, u) + 1 + d_H(v_{j'}, v) \le \lambda \\ \lambda < d_H(v_i, u) + d_H(v_{i'}, u) \\ \lambda < d_H(v_j, v) + d_H(v_{j'}, v). \end{cases}$$

If we sum up all the inequalities we get $2 < 0$, a contradiction.

Now, we deal with case (ii). Since $d_H(v_i, v_{i'}), d_H(v_{i'}, v_{i''}), d_H(v_{i''}, v_i) > \lambda$ and since $d_{H+e}(v_i, v_{i'}), d_{H+e}(v_{i'}, v_{i''}), d_{H+e}(v_{i''}, v_i) \le \lambda$, any shortest path in $H + e$ between any pair of the three vertices $v_i, v_{i'}, v_{i''}$ has to go through the edge e. Thus, two

[3] Here, $H^\lambda[U]$ means the subgraph of H^λ induced by U.

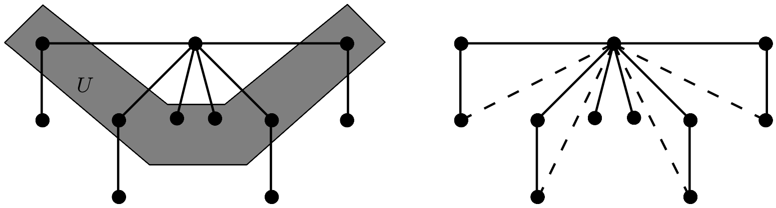

Fig. 2. Tight example for the upper bound on D'. On the left side, the input graph G is depicted. Let U be the set of vertices within the shaded area. The budget is $B = |U| - 2$. Observe that U is a maximum independent set of G, and thus, $\alpha(G) = B + 2$. On the right side, the addition of B edges to G (the dashed edges) induces a graph of diameter $D = 2$. Thus, $\alpha(G^{D'}) \leq B + 1$ if and only if $D' \geq 2$.

of the three vertices $v_i, v_{i'}, v_{i''}$ are closer to one endvertex of e than to the other one. Without loss of generality, assume that $d_H(v_{i'}, v) \leq d_H(v_{i'}, u)$, and $d_H(v_{i''}, v) \leq d_H(v_{i''}, u)$. As a consequence,

$$\begin{aligned} d_H(v_{i'}, v_{i''}) &\leq d_H(v_{i'}, v) + d_H(v_{i''}, v) \\ &< \min\{d_H(v_{i'}, v) + 1 + d_H(v_{i''}, u), d_H(v_{i'}, u) + 1 + d_H(v_{i''}, v)\} \\ &= d_{H+e}(v_{i'}, v_{i''}) \leq \lambda, \end{aligned}$$

and thus $(v_{i'}, v_{i''}) \in E(H^\lambda)$, a contradiction. □

Thanks to the above lemmata, we can prove the following.

Theorem 1. *There exists a $\left(2 + \frac{2}{D}\right)$-approximation algorithm for BCMD running in polynomial time.*

Proof. Let F^* be an optimal solution for the BCMD on input instance G and B, and let $D = \mathtt{diam}(G+F^*)$. As $\alpha((G+F^*)^D) = 1$ and $|F^*| \leq B$, by a repeated use of Lemma 2, we obtain $\alpha((G+F^*)^D) \geq \alpha(G^D) - |F^*|$, and thus $\alpha(G^D) \leq B + 1$. As a consequence, Lemma 1 implies that the algorithm Gonzalez computes a $(B+1)$-clustering $\langle V_0, \ldots, V_B \rangle$ of $V(G)$ such that $r_G(V_i) \leq D, \forall i \in [0, B]$. Let z_i be a center of cluster V_i and let $F = \{(z_0, z_i) \mid i \in [1, B]\}$ be the solution returned by the algorithm of Li et al. [20] when algorithm Gonzalez is used as the k-clustering procedure. Clearly, $|F| \leq B$. Moreover, every vertex $v \in V(G)$ is at distance at most $D + 1$ from z_0 in $G + F$. Therefore, $\mathtt{diam}(G + F) \leq 2D + 2$. The claim follows. □

Looking at the proof of Theorem 1, one can notice that the approximation ratio of the algorithm for BCMD is actually $\frac{2D'+2}{D}$, where $D' \in \mathbb{Z}^+$ is the minimum positive integer value such that $\alpha(G^{D'}) \leq B + 1$. Moreover, as a consequence of Lemma 2, we proved that $D' \leq D$. The example in Fig. 2 shows that this upper bound is essentially tight.

The following theorem shows a better approximability result for the class of forests.

Theorem 2. *For the BCMD restricted to forests, there exists a linear time algorithm which returns a $\left(1+\frac{2}{D}\right)$-approximate solution for even values of D and a $\left(1 + \frac{4}{D}\right)$-approximate algorithm for odd values of D.*

Proof. Let G be a forest and let $\{H_1, \ldots, H_\ell\}$ be a minimum partition of G^D in cliques.[4] As forests are perfect graphs[5] and because G^D is still a perfect graph (see [3]), we have that $\alpha(G^D) = \ell$. Moreover, analogously to the proof of Theorem 1, we have that $\ell \leq B+1$. Additionally, each H_i can be extended to a subtree T_i of G such that $V(H_i) \subseteq V(T_i)$ and $\mathtt{diam}(T_i) \leq D$. As an unweighted undirected tree of diameter λ has radius equal to $\lceil \lambda/2 \rceil$ [2], it follows that $\langle V(T_1), \ldots, V(T_\ell) \rangle$ is an ℓ-clustering of $V(G)$ such that $r_G(V(T_i)) \leq \lceil D/2 \rceil$, for every $i \in [1, \ell]$. As a consequence, an optimal solution for the k-center problem on input G and $k = B+1$ is a $(B+1)$-clustering $\langle V_0, \ldots, V_B \rangle$ such that $r_G(V_i) \leq \lceil D/2 \rceil$, for every $i \in [0, B]$. Furthermore, such a clustering can be found in linear time [4,19]. Let z_i be a center of V_i. Let $F = \{(z_0, z_i) \mid i \in [1, B]\}$. Clearly, $|F| \leq B$. Moreover, every vertex $v \in V(G)$ is at distance at most $\lceil D/2 \rceil + 1$ from z_0 in $G+F$. Therefore, $\mathtt{diam}(G+F) \leq 2\lceil D/2 \rceil + 2$. Thus, if D is even we get $\mathtt{diam}(G+F) \leq D+2$, while for odd values of D we obtain $\mathtt{diam}(G+F) \leq D+4$. This completes the proof. □

Open problems and conjectures. Despite all our efforts, we could not find an example in which the algorithm of Theorem 1 returns a solution having diameter strictly greater than $2D - 1$. Therefore we conjecture that

Conjecture 1. *The algorithm of Theorem 1 computes a $\left(2 - \frac{1}{D}\right)$-approximate solution for BCMD.*

As it might be difficult to either prove or disprove the above conjecture, we propose the following alternative conjecture which, as we will prove, implies a weaker approximability result.

Conjecture 2. *Let G be a graph and let B be the minimum number of edges whose addition to G induces a graph of diameter equal to D. Then, there exists a $(B+1)$-clustering $\langle U_0, U_1, \ldots, U_B \rangle$ of $V(G)$ such that* (i) $r_G(U_0) \leq D$ *and* (ii) $\mathtt{diam}(G[U_i]) \leq D-1, \forall i \in [1, B]$.

4 A *partition of G in cliques* is a collection $\{H_1, \ldots, H_\ell\}$ of vertex-disjoint subgraphs of G such that $\bigcup_{i=1}^{\ell} V(H_i) = V(G)$ and each H_i is a clique. A partition $\{H_1, \ldots, H_\ell\}$ of a graph G in cliques is minimum if and only if for every other partition $\{H'_1, \ldots, H'_{\ell'}\}$ of G in cliques we have that $\ell \leq \ell'$.

5 A graph G is *perfect* if and only if the size of a minimum partition of G in cliques equals $\alpha(G)$ [3].

Theorem 3. *If Conjecture 2 holds, then there exists a 2-approximation algorithm for BCMD running in polynomial time.*

Proof. The algorithm guesses D and the center z_0 of the cluster U_0 and builds cluster $V_0 = \{v \mid v \in V(G), d_G(z_0, v) \leq D\}$. Let $U = V(G) \setminus V_0$. Then, the algorithm computes a B-clustering $\langle V_1, \ldots, V_B\rangle$ of U using the algorithm `Gonzalez`. Next, it computes a center z_i for each cluster V_i and finally returns the set $F = \{(z_0, z_i) \mid i \in [1, B]\}$.

Clearly $U_0 \subseteq V_0$. Moreover, because $\mathtt{diam}(G[U_i]) \leq D - 1$ and $U \subseteq \bigcup_{i=1}^{B} U_i$, we have that $\alpha(G^{D-1}[U]) \leq B$ and thus, by Lemma 1, $r_G(V_i) \leq D - 1$. As a consequence, $\mathtt{diam}(G + F) \leq 2D$. This completes the proof. □

4. An $O(\log n)$-approximation algorithm for MCBD

In this section we describe an $O(\log n)$-approximation algorithm for MCBD. Without loss of generality, we can restrict our focus to the class of connected graphs. Indeed, if $D < \mathtt{diam}(G)$ and the number of connected components of G is equal to $\ell \geq 2$, then we can first add $\ell - 1$ edges to G to make it connected and then run our $O(\log n)$-approximation algorithm. Since the value of an optimal solution has to be greater than or equal to $\ell - 1$, then we are still guaranteed to compute an $O(\log n)$-approximate solution. In what follows, we provide an informal description of the algorithm. For a formal description, see Algorithm 1.

Algorithm 1 $O(\log n)$-approximation algorithm for MCBD

if $\mathtt{diam}(G) \leq D$ **then return** $\emptyset$ **endif**
if G is not connected **then**
  add to G the minimum number of edges to make it connected;
end if
— beginning of first phase —
$F_1 := \emptyset$;
$i := 0$;
$H_i := G$;
fix a vertex $s \in V(G)$;
while $\mathtt{cost}(H_i) > 0$ **do**
  $g := 0$;
  for all $e \in \{(s, v) \mid v \in V(G)\}$ **do**
    if $\mathtt{gain}(e, H_i) > g$ **then** $g := \mathtt{gain}(e, H_i)$; $\bar{e} := e$; **endif**
  end for
  $F_1 := F_1 \cup \{\bar{e}\}$;
  $H_{i+1} := H_i + \bar{e}$;
  $i := i + 1$;
end while
$G_1 := G + F_1$;
— end of first phase —
— beginning of second phase —
$Z := \big\{\{u, v\} \mid \{u, v\} \in \mathcal{I}(G_1)\big\}$;
for all $e \in E(\bar{G})$ **do**
  $S_e := \big\{\{u, v\} \mid \{u, v\} \in Z, d_{G_1+e}(u, v) \leq D\big\}$;
end for
$\mathcal{S} := \{S_e \mid e \in E(\bar{G})\}$;
compute a solution X for the Set Cover Problem on input Z and $\mathcal{S}$;
$F_2 := \{e \mid S_e \in X\}$;
— end of second phase —
return $F_1 \cup F_2$;

In the rest of the section, for any graph H, let $\mathcal{I}(H)$ be a set of unordered pairs of vertices defined as follows $\mathcal{I}(H) := \big\{\{u, v\} \mid u, v \in V(H), d_H(u, v) > D\big\}$. The algorithm uses a greedy approach and consists of two phases. In the first phase, the algorithm fixes a vertex $s \in V(G)$ and computes a set F_1 of edges incident to s such that for each pair $\{u, v\} \in \mathcal{I}(G)$, $d_{G+F_1}(s, u) + d_{G+F_1}(s, v) \leq D + 1$. Observe that this immediately implies that $\mathtt{diam}(G + F_1) \leq D + 1$. Moreover, for every pair $\{u, v\} \in \mathcal{I}(G)$ there exists a path from u to v in $G + F_1$ of length less than or equal to $D + 1$ which contains s. Let $G_1 = G + F_1$. In the second phase, the algorithm computes a set F_2 of edges such that $\mathtt{diam}(G_1 + F_2) \leq D$. More precisely, for every pair $\{u, v\} \in \mathcal{I}(G_1)$, there exists an edge $e \in F_2$ such that $d_{G_1+e}(u, v) \leq D$. We will prove that $|F_1|, |F_2| = O(B \log n)$.

For the rest of the section, let $s \in V(G)$ be fixed. We now describe the first phase of the algorithm. For any graph H and for every two vertices $u, v \in V(H)$, we define

$$\mu_H(u, v) := \max\big\{0, d_H(s, u) + d_H(s, v) - (D + 1)\big\}.$$

Furthermore, let

$$\texttt{cost}(H) := \sum_{\{u,v\}\in \mathcal{I}(H)} \mu_H(u,v).$$

Let $F_s = \{(s,v) \mid v \in V(G)\}$. For a given edge e, define $\texttt{gain}(e,H) := \texttt{cost}(H) - \texttt{cost}(H+e)$ to be the *gain* of edge e w.r.t. H. The algorithm first sets $F_1 := \emptyset$ and $H_0 := G$ and then proceeds in discrete steps. During step $i \geq 1$, the algorithm selects an edge $e = (s,v) \in F_s$ that maximizes $\texttt{gain}(e, H_{i-1})$, adds e to F_1, and sets $H_i := H_{i-1} + e$. The first phase of the algorithm ends when $\texttt{cost}(H_i) = 0$. We prove the following

Lemma 3. *At the end of the first phase, $\mu_{G+F_1}(u,v) = 0$ for every $u, v \in V(G)$. Moreover, $|F_1| = O(B \log n)$.*

Proof. Clearly, the first phase of the algorithm ends when $\mu_{H_i}(u,v) = \mu_{G+F_1}(u,v) = 0$, for every $\{u,v\} \in \mathcal{I}(G)$.

Let F^* be an optimal solution for MCBD on inputs G and D and let $U = \{v_1, \ldots, v_\ell\}$ be the set of endvertices of the edges in F^*. Clearly, $\ell \leq 2B$. Let $e_j^* = (s, v_j)$, for every $j \in [1,\ell]$, and let $\hat{F} = \{e_1^*, \ldots, e_\ell^*\}$. Observe that $\mu_{G+\hat{F}}(u,v) = 0$, for every $u, v \in \mathcal{I}(G)$. Moreover, $\hat{F} \subseteq F_s$.

Proposition 1. *For every i, $\sum_{j=1}^{\ell} \texttt{gain}(e_j^*, H_i) \geq \texttt{cost}(H_i)$.*

Proof. For each $j \in [1,\ell]$, let H_i^j denote the graph $H_i + \{e_1^*, \ldots, e_j^*\}$ and let $H_i^0 = H_i$. Observe that $\texttt{cost}(H_i^\ell) = 0$ as $G + \hat{F}$ is a subgraph of H_i^ℓ. As edges $e_1^*, \ldots, e_\ell^*$ are all incident to s, for all $j \in [1,\ell]$ and for all $v \in V(G)$, $d_{H_i^j}(s,v) < d_{H_i^{j-1}}(s,v)$ implies $d_{H_i^j}(s,v) = d_{H_i+e_j^*}(s,v)$. As a consequence, for all $j \in [1,\ell]$ and for all $v \in V(G)$, $d_{H_i}(s,v) - d_{H_i+e_j^*}(s,v) \geq d_{H_i^{j-1}}(s,v) - d_{H_i^j}(s,v)$. Then, by a simple case analysis, it follows that $\mu_{H_i}(u,v) - \mu_{H_i+e_j^*}(u,v) \geq \mu_{H_i^{j-1}}(u,v) - \mu_{H_i^j}(u,v)$ for all $j \in [1,\ell]$ and for all $u, v \in V(G)$, and thus $\texttt{gain}\big(e_j^*, H_i\big) \geq \texttt{gain}\big(e_j^*, H_i^{j-1}\big)$. Therefore

$$\sum_{j=1}^{\ell} \texttt{gain}\big(e_j^*, H_i\big) \geq \sum_{j=1}^{\ell} \texttt{gain}\big(e_j^*, H_i^{j-1}\big) = \sum_{j=1}^{\ell} \Big(\texttt{cost}\big(H_i^{j-1}\big) - \texttt{cost}\big(H_i^j\big)\Big)$$
$$= \texttt{cost}(H_i) - \texttt{cost}(H_i^\ell) = \texttt{cost}(H_i). \quad \square$$

As a consequence of the above proposition, for every $i \geq 1$, there exists an edge $e \in F_s$ such that $\texttt{gain}(e, H_{i-1}) \geq \max\left\{1, \frac{\texttt{cost}(H_{i-1})}{\ell}\right\} \geq \max\left\{1, \frac{\texttt{cost}(H_{i-1})}{2B}\right\}$. This implies that $\texttt{cost}(H_i) \leq \texttt{cost}(H_0)\big(1 - \frac{1}{2B}\big)^i = \texttt{cost}(G)\big(1 - \frac{1}{2B}\big)^i$ for every i. Moreover, at the beginning of the last step of the algorithm, say $\eta + 1$, we have that $\texttt{cost}(H_\eta) \geq 1$. Therefore, $1 \leq \texttt{cost}(H_\eta) \leq \texttt{cost}(G)\big(1 - \frac{1}{2B}\big)^\eta$. Taking the natural logarithm and simplifying, we finally get

$$\eta \leq 2B \ln \texttt{cost}(G) = O(B \log n),$$

where the equality comes from the fact that $\texttt{cost}(G) = O(n^3)$ as G is connected. $\square$

We now describe the second phase of the algorithm. Let $G_1 = G + F_1$, where F_1 is the set of edges computed by the algorithm in the first phase. We make a reduction to the *Set Cover Problem*. The *Set Cover Problem* takes as input a set of objects Z and a collection $\mathcal{S}$ of subsets of Z and asks for the minimum-cardinality subset $\mathcal{S}' \subseteq \mathcal{S}$ that *covers* Z, i.e., $\bigcup_{S\in\mathcal{S}'} S = Z$. The Set Cover Problem is well known to be approximable within $O(\log |Z|)$ [17]. We build an instance of the Set Cover Problem as follows. The objects in Z are the unordered pairs in $\mathcal{I}(G_1)$. There is a set S_e for every edge e in $\bar{G}$ which is defined as follows

$$S_e := \Big\{\{u,v\} \mid \{u,v\} \in Z, d_{G_1+e}(u,v) \leq D\Big\}.$$

Let $\mathcal{S}''$ be a solution computed by the $O(\log |Z|)$-approximation algorithm for Set Cover Problem and let $F_2 = \{e \mid S_e \in X\}$. We can prove the following

Lemma 4. *At the end of the second phase, $\texttt{diam}(G_1 + F_2) \leq D$. Moreover, $|F_2| = O(B \log n)$.*

Proof. Let $\{u,v\} \in \mathcal{I}(G_1)$. By construction, pair $\{u,v\} \in Z$ and there exists $S_e \in \mathcal{S}''$ such that $d_{G_1+e}(u,v) \leq D$. As $e \in F_2$, we have that $d_{G_1+F_2}(u,v) \leq d_{G_1+e}(u,v) \leq D$. Therefore, $\texttt{diam}(G_1 + F_2) \leq D$.

Let F^* be an optimal solution for MCBD on inputs G and D and let U be the set of endvertices of the edges in F^*. We claim that

$$\mathcal{S}^* = \Big\{S_{(s,u)} \mid u \in U\Big\} \cup \Big\{S_e \mid e \in F^*\Big\}$$

is a feasible solution for the instance of the Set Cover Problem defined above. Observe that this is enough to prove the claim, as $|\mathcal{S}^*| \leq 3B$ and $|Z| = |\mathcal{I}(G_1)| \leq n^2$ implies $|F_2| = O(|\mathcal{S}^*| \log |Z|) = O(B \log n)$.

Let $\{u,v\}$ be a pair in $\mathcal{I}(G_1)$ and let P be a shortest path from u to v in $G + F^*$. As $d_{G_1}(u,v) > D$ while $d_{G_1+F^*}(u,v) \leq d_{G+F^*}(u,v) \leq D$, P contains some edge of F^*. We traverse P from u to v. Let u' be the first vertex of P which is also a vertex of U, and let v' be the last vertex of P which is also a vertex of U. If P contains exactly one edge of F^*, i.e.,

$(u', v') \in F^*$, then pair $\{u, v\}$ is in set $S_{(u',v')}$ by construction and $S_{(u',v')} \in \mathcal{S}^*$. If P contains two or more edges of F^*, then $d_{G_1}(u, u') + d_{G_1}(v', v) \leq D - 2$. Moreover, as Lemma 3 implies that $d_{G_1}(s, u) + d_{G_1}(s, v) = D + 1$, we have $d_{G_1}(u, u') + d_{G_1}(v', v) \leq d_{G_1}(s, u) + d_{G_1}(s, v) - 3$. Therefore, $d_{G_1}(u, u') \leq d_{G_1}(s, u) - 2$ or $d_{G_1}(v', v) \leq d_{G_1}(s, v) - 2$. Without loss of generality, assume that $d_{G_1}(u, u') \leq d_{G_1}(s, u) - 2$. Then, $d_{G_1+(s,u')}(u, v) \leq d_{G_1+(s,u')}(s, u) + d_{G_1+(s,u')}(s, v) \leq d_{G_1}(u, u') + 1 + d_{G_1}(s, v) \leq d_{G_1}(s, u) + d_{G_1}(s, v) - 1 = D$. As a consequence, the pair $\{u, v\}$ is in the set $S_{(s,u')}$. As $S_{(s,u')} \in \mathcal{S}^*$, the proof is completed. □

Lemmas 3 and 4 yield the main theorem of this section.

Theorem 4. *Algorithm 1 is a polynomial time $O(\log n)$-approximation algorithm for MCBD.*

Remark 1. We already pointed out that our algorithm extends to directed graphs. It also extends to the case when we place the distance requirements D_{u_i,v_i} for ℓ pairs $\{u_1, v_2\}, \ldots, \{u_\ell, v_\ell\}$ of vertices of G (in particular the resulting graph need not be connected). The approximation ratio becomes $O(\log \ell + \log D_{\max})$, where $D_{\max} = \max_{i\in[1,\ell]} D_{u_i,v_i}$.

Open problems and conjectures for MCBD. The first natural open problem is that of determining whether MCBD is in P for the class of forests. In the following, we provide a conjecture which, as we will prove, guarantees the existence of a polynomial time $(1+1/B)$-approximation algorithm for the class of forests for all even values of D. To the best of our knowledge, 2 is the approximation factor of the best up to date algorithm for the MCBD problem on forests for even values of D (see paper [6]).

Conjecture 3. *Let G be a forest, let $D \in \mathbb{Z}^+$ be an even integer, and let B the minimum number of edges whose addition to G induces a graph of diameter less than or equal to D. Then, there exists a $(B + 1)$-clustering $\langle U_0, U_1, \ldots, U_B\rangle$ of $V(G)$ such that* (i) $\mathtt{diam}(G[U_0]) \leq D$, (ii) $\mathtt{diam}(G[U_1]) \leq D - 1$, *and* (iii) $\mathtt{diam}(G[U_i]) \leq D - 2, \forall i \in (1, B]$.

Theorem 5. *If Conjecture 3 holds, then there exists a polynomial time algorithm computing a $(1 + \frac{1}{B})$-approximate solution for MCBD on forests for the case in which D is an even number.*

Proof. We claim that a clustering satisfying the conditions of Conjecture 3 can be found in polynomial time. Let $\langle U_0, \ldots, U_B\rangle$ be a $(B+1)$-clustering of $V(G)$ satisfying the condition of Conjecture 3. First of all observe that $G[U_i]$ is a tree. It is well known that a tree T has exactly one center if $\mathtt{diam}(T)$ is even while it has exactly two adjacent centers if $\mathtt{diam}(T)$ is odd [2]. Let Z_i be the set of centers of $G[U_i]$. The algorithm guesses one center $z_0 \in Z_0$ and the set Z_1. Let E_0 be the set of edges of G incident to z_0. The algorithm builds a new tree G' from G as follows. First, it appends sufficiently many paths of length $D/2 - 1$ ($2n/D$ paths are enough) to z_0 and sufficiently many paths of length $D/2 - 1$ (again, $2n/D$ paths are enough) to all vertices in Z_1. Then it identifies all the edges in E_0.[6] Finally, if $|Z_1| = 2$, the algorithm identifies also the (unique) edge $\bar{e}$ linking the two vertices in Z_1. Observe that G' has a $(B + 1)$-clustering $\langle U'_0, \ldots, U'_B\rangle$ such that $r_{G'}(U'_i) \leq D/2 - 1$. Moreover, every such $(B + 1)$-clustering has to contain a cluster whose only center is x_{E_0} and, if $|Z_1| = 2$, a cluster whose only center is $x_{\{\bar{e}\}}$.

The algorithm then finds in linear time a $(B+1)$-clustering $\langle V'_0, \ldots, V'_B\rangle$ of $V(G')$ minimizing the maximum value among the cluster radii [4,19]. Moreover, each cluster V'_i can be extended to a cluster V''_i such that (i) $V'_i \subseteq V''_i$, (ii) $G[V''_i]$ is a tree, and (iii) $r_{G'}(V'_i) = r_{G'}(V''_i)$. Without loss of generality, assume that x_{E_0} is the center of V'_0. Moreover, if $|Z_1| = 2$, without loss of generality, assume that $x_{\{\bar{e}\}}$ is the center of V'_1. Let V_0 be equal to $V'_0 \setminus \{x_{E_0}\}$ plus z_0 and all vertices adjacent to z_0 in G. If $|Z_1| = 2$, then let V_1 be equal to $V'_1 \setminus \{x_{\{\bar{e}\}}\}$ plus the vertices in Z_1. For every other $i \in (1, B]$ let $V_i = V'_i$. The $(B + 1)$-clustering $\langle V_0, \ldots, V_B\rangle$ of $V(G)$ satisfies the conditions of Conjecture 3. Let z_i be a center of cluster V_i, $i \in (1, B]$, and let $F := \{(z_0, z_i) \mid i \in (1, B]\} \cup \{(z_0, z) \mid z \in Z_1\}$. Observe that each vertex is at distance of at most $D/2$ in the graph $G + F$. As a consequence $\mathtt{diam}(G + F) \leq D$. Moreover, $|F| \leq B + 1$.[7] The claim follows. □

We observe that the result of Theorem 5 implies the existence of a PTAS for the problem. In fact, for every $\epsilon > 0$ and assuming that Conjecture 3 holds, the algorithm of Theorem 5 computes a $(1 + \epsilon)$-approximate solution for every $B > 1/\epsilon$, while, for every $B \leq 1/\epsilon$, the brute-force-search algorithm that computes an optimal solution runs in time $n^{O(B)} = n^{O(1/\epsilon)}$.

5. On the existence of bicriteria approximation algorithms

In this section we first prove the existence of a $(2 - \frac{1}{B}, 2)$-approximation algorithm for BCMD and then we show that for every $D \geq 4$ there is no polynomial time algorithm with an approximation guarantee of $(c \log n, \delta)$, for some constant $c > 0$, and for every $\delta < \frac{5}{3} - \frac{7-(D+1) \bmod 3}{3D}$, unless P = NP.

We slightly modify the algorithm described in Section 3 to show the existence of a $\left(2 - \frac{1}{B}, 2\right)$-approximation algorithm for BCMD. The correctness proof follows from the subsequent two key lemmas.

Lemma 5. *Let $U \subseteq V(G)$. Let $\langle U_1, \ldots, U_k\rangle$ be a k-clustering of U and let $R = \max_{i\in[1,k]} r_G(U_i)$. Then the algorithm* Gonzalez *finds a k-clustering $\langle V_1, \ldots, V_k\rangle$ of U such that $r_G(V_i) \leq 2R, \forall i \in [1, k]$.*

---

6 Let H be a graph and let $F \subseteq E(H)$. Let U be the set of vertices which are not incident to any of the edges in F and let $\bar{F} := \{(u, v) \mid (u, v) \in E(H), u, v \in U\}$. The graph obtained from H by *identifying* the edges in F is the graph H' with vertex set $V(H') := U \cup \{x_F\}$ and edge set $E(H') := \bar{F} \cup \{(x_F, v) \mid (u, v) \in E(H), u \in V(H) \setminus U\}$.

7 Observe that if $|Z_1| = 1$, then $|F| \leq B$ and the algorithm is exact.

Proof. The proof immediately follows from Lemma 1 after observing that $\alpha(G^{2R}) \le k$ as the distance between every pair of vertices $u, v \in U_i$ for every i is $d_G(u, v) \le 2 \cdot r_G(U_i) \le 2R$. □

Lemma 6. *There exists a $(2B)$-clustering $\langle U_1, \dots, U_{2B}\rangle$ of $V(G)$ such that $r_G(U_1) \le D$ and $r_G(U_i) \le \frac{D-1}{2}$, for every $i \in (1, 2B]$.*

Proof. Let F^* be an optimal solution for BCMD on input G and B, and let $V' = \{v_1, \dots, v_\ell\}$ be the set of endvertices of the edges in F^*. Clearly, $\ell \le 2B$. Let $\langle U_1, \dots, U_\ell\rangle$ be an ℓ-clustering of $V(G)$ defined as follows. The cluster U_i contains vertex v_i and all the vertices $v \in V(G) \setminus \bigcup_{k=1}^{i-1} U_k$ such that $d_G(v, v_i) \le d_G(v, v_j)$, $\forall j \in (i, \ell]$. Clearly, $r_G(U_i) \le \max_{v \in U_i} d_G(v, v_i)$. As a consequence, we have to prove the claim for the case in which there exist $i \in [1, \ell]$ and $v \in U_i$ such that $d_G(v, v_i) > \frac{D-1}{2}$. Without loss of generality, let us assume that there exists $v \in U_1$ such that $d_G(v, v_1) > \frac{D-1}{2}$; let v^* denote an arbitrary but fixed vertex v with this property.

Let v, v' be two vertices of U_i. Every path from v to v' in $G+F^*$ passing through some edge of F^* has a length greater than or equal to $d_G(v, v_i) + 1 + d_G(v', v_i)$, while a shortest path in G from v to v' has a length of at most $d_G(v, v_i) + d_G(v', v_i)$. As a consequence, $d_G(v, v') = d_{G+F^*}(v, v') \le D$.

We modify the ℓ clusters by moving vertices from $U_2, \dots, U_\ell$ to U_1 as follows. As long as there exists a vertex $v \in U_i$, $i \in (1, \ell]$, such that $d_G(v, v_i) > \frac{D-1}{2}$, we remove v from U_i and we add v to U_1. Any path in $G+F^*$ from v^* to v passing through any edge in F^* has a cost greater than or equal to $d_G(v^*, v_1) + 1 + d_G(v, v_i) > D$. Therefore, $d_G(v^*, v) = d_{G+F^*}(v^*, v) \le D$. As a consequence, $r_G(U_1) \le \max_{v \in U_1} d_G(v^*, v) \le D$. Moreover, for every $i \in (1, \ell]$, $r_G(U_i) \le \max_{v \in V_i} d_G(v_i, v) \le \frac{D-1}{2}$ by construction. The claim follows. □

Let $\langle U_1, \dots, U_{2B}\rangle$ be a $(2B)$-clustering of $V(G)$ from Lemma 6. Let v_1 be a center of U_1, and let $D' = r_G(U_1)$. Our algorithm, whose formal description is given below (Algorithm 2), first guesses v_1 and D'. Then, it computes a cluster $V_1 = \{v \in V(G) \mid d_G(v_1, v) \le D'\}$ and uses the Gonzalez algorithm to find a $(2B-1)$-clustering $\langle V_2, \dots, V_{2B}\rangle$ of $V(G) \setminus V_1$. As $U_1 \subseteq V_1$, $\langle U_2, \dots, U_{2B}\rangle$ is a $(2B-1)$-clustering of $V(G) \setminus V_1$. Therefore, Lemma 5 implies that $r_G(V_i) \le D - 1$, for every $i \in (1, 2B]$. Let z_i be a center of cluster V_i. The algorithm outputs $F = \{(z_1, z_i) \mid i \in (1, 2B]\}$. Observe that every vertex v is at distance of at most D from z_1 in $G + F$. Therefore, $\mathtt{diam}(G + F) \le 2D$. Also observe that $|F| \le 2B - 1$. We have proved the following theorem.

Theorem 6. *Algorithm 2 returns a $\left(2 - \frac{1}{B}, 2\right)$-approximate solution for BCMD.*

Algorithm 2 Bicriteria approximation algorithm

$F := \emptyset$;
for all $v \in V(G)$ **do**
  for all $v' \in V(G)$ **do**
    $V_1 := \{u \in V(G) \mid d_G(v, u) \le d_G(v, v')\}$;
    compute a $(2B-1)$-clustering $\langle V_2, \dots, V_{2B}\rangle$ of $V(G) \setminus V_1$ using `Gonzalez`;
    compute center z_i for every cluster V_i, $i \in [1, 2B]$;
    $F' := \{(z_1, z_i) \mid i \in [1, 2B]\}$;
    if $\mathtt{diam}(G + F') < \mathtt{diam}(G + F)$ **then** $F := F'$ **endif**
  end for
end for
return F

Concerning inapproximability of BCMD we can prove the following.

Theorem 7. *For every fixed integer $D \ge 4$, there exists no $(c \log n, \delta)$-approximation algorithm for BCMD, for some constant $c > 0$ and for every $\delta < \frac{5}{3} - \frac{7-(D+1) \bmod 3}{3D}$, unless $\mathsf{P} = \mathsf{NP}$.*

Proof. The reduction is from the *Minimum Dominating Set Problem* (MDS for short), i.e., the problem of finding a minimum-cardinality set of vertices U of a given graph G' on $\hat{n}$ vertices such that every vertex of G' is in U or it is a neighbor of some vertex in U. The MDS is not approximable within $c' \log \hat{n}$, for some constant $c' > 0$, unless $\mathsf{P} = \mathsf{NP}$ [24].

Let G' be a graph with $\hat{n}$ vertices and let k^* be the size of a minimum dominating set in G'. We transform the instance of MDS to an instance of BCMD with n vertices and claim that the existence of a $(c \log n, \delta)$-approximation algorithm for BCMD, with $\delta < \frac{5}{3} - \frac{7-(D+1) \bmod 3}{3D}$, implies the existence of a $(c' \log \hat{n})$-approximation algorithm for MDS, for some $c' < 17c$. This would immediately lead to a contradiction by choosing c small enough.

For the sake of exposition, we prove the result for every $D = 6\rho + 4$, where $\rho \ge 0$ is a fixed integer. We build the input graph G in the following way (see Fig. 3). G contains 2 copies G_1, G_2 of G' plus a singleton vertex s such that $s \notin V(G')$. For every $u \in V(G')$, denote by u_i the copy of u in G_i. Replace each edge $(u_i, v_i) \in E(G_i)$ with a path P_{u_i, v_i} from u_i to v_i of length $2\rho + 1$ by adding 2ρ new vertices and $2\rho + 1$ new edges. For every vertex $u \in V(G')$, and for every $i = 1, 2$, append a path P_u^i to u_i of length ρ by adding ρ new vertices and ρ new edges. For every $i = 1, 2$, denote by ν_u^i the endvertex of P_u^i different from u_i (if $\rho = 0$ then $\nu_u^i = u_i$). Set $B = 2k^*$. Observe that $n \le 1 + 2\hat{n} + 2(2\rho\hat{n}^2 + \rho\hat{n}) \le 5\rho\hat{n}^2$, for $\hat{n}$ large enough.

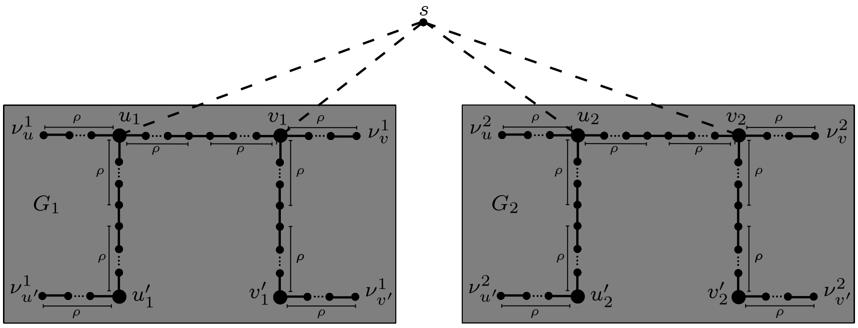

Fig. 3. The reduction for the case $D = 6\rho + 4$, where $\rho \geq 0$ is a fixed integer. The big vertices are the copies of the vertices contained in G'. The solid edges are the edges in the reduction. Observe that the addition of the dashed edges to the graph induces a graph of diameter less than or equal to D.

Let U^* be a minimum-cardinality dominating set in G'. By augmenting G with the $B = 2k^*$ edges from s to both the copies of each vertex in U^*, we obtain a graph having diameter less than or equal to D. Indeed, every vertex in P_{u_i,v_i} is at distance of at most ρ from either u_i or v_i. Furthermore, every vertex in P^i_u is at distance of at most ρ from u_i. Finally, every u_i is at distance of at most $2\rho + 1$ from a copy of some vertex in U^* in G, as U^* is a dominating set in G'. As a consequence, every vertex in G_i is at distance at most $3\rho + 2$ from s. Therefore, the diameter of the resulting graph is less than or equal to D.

Now, let F be the set of edges computed by any $\left(c \log n, \delta\right)$-approximation algorithm for BCMD, with $\delta < \frac{5}{3} - \frac{7-(D+1) \bmod 3}{3D} = \frac{5}{3} - \frac{5}{3D}$. Let X be the set of the endvertices of the edges in F. We have that $|X| \leq 2cB \log n$. Let Y be equal to X. We modify Y as follows. As long as there is an $x \in Y$ which is an internal vertex of P_{u_i,v_i}, then we remove x from Y and we add u and v to Y. Next, as long as there is an $x \in Y$ which is a vertex of $V(P^i_u) \setminus \{u_i\}$, then we remove x from Y and we add u_i to Y. Clearly, $|Y| \leq 2|X| \leq 4cB \log n$. Let U be the set of vertices in G' defined as follows: U contains a vertex u of G' if and only if $u_1 \in Y$ or $u_2 \in Y$. We have that $|U| \leq |Y| \leq 4cB \log n \leq 8ck^* \log\left(5\rho\hat{n}^2\right) < 17ck^* \log \hat{n}$, for large values of $\hat{n}$. To complete the proof, it is enough to show that U is a dominating set in G'. Let u be any vertex in $V(G')$ and consider the two vertices ν^1_u and ν^2_u in G; their distance in G is $+\infty$, while their distance in $G + F$ is upper bounded by

$$\delta D < \frac{5}{3}D - \frac{5}{3} = \frac{5}{3}(6\rho + 4) - \frac{5}{3} = 10\rho + 5.$$

As a consequence, $\delta D \leq 10\rho + 4$. As there is no edge between $V(G_1)$ and $V(G_2)$ in G, there exists a vertex $x \in X$ such that $d_G(x, \nu^1_u) \leq 5\rho + 1$ or $d_G(x, \nu^2_u) \leq 5\rho + 1$. Therefore, by construction, there exists a vertex v in Y such that $d_G(v, \nu^1_u) \leq 5\rho + 1$ or $d_G(v, \nu^2_u) \leq 5\rho + 1$, i.e., $d_G(v, u_1) \leq 4\rho + 1$ or $d_G(v, u_2) \leq 4\rho + 1$. Since each of the vertices in Y is a copy of some vertex of G' and because $d_G(u_i, v_i) \geq 4\rho + 2$ for every $u, v \in V(G')$, $(u, v) \notin E(G')$, it follows that U is a dominating set in G'.

To extend the proof for all the other values of $D \geq 4$ we do the following. Let $D = 6\rho + 3x + \theta_1 + \theta_2 + 2$, where ρ is a non-negative integer while $x, \theta_1, \theta_2 \in \{0, 1\}$. Observe that there exists a feasible choice of the variables for every $D \geq 4$. The algorithm builds an instance similar to the one we described above with the only difference that (i) the length of path P_{u_i,v_i} is equal to $2\rho + \theta_i + x$ and (ii) the length of the path P^i_u is $\rho + x$ if $i = 1$, ρ otherwise. The proof then goes along the same line of the above proof. □

6. Extension to edge-weighted graphs

In this section we extend some of our results for a generalized version of BCMD which is defined as follows.

Weighted Bounded-Cardinality Minimum-Diameter Edge Addition (WBCMD)
Instance: an undirected graph G with a real weight $w(e)$ associated with each edge $e \in E(G)$, a value $B \in \mathbb{Z}^+$, and a non-negative real value $\omega \geq 0$.
Goal: find a set F of edges of cardinality at most B such that the diameter of $G + F$ is minimized, where each edge e in F has a weight $w(e) = \omega$ associated with it.[8]

The following results hold.

Theorem 8. *There exists a* $\left(2 + \frac{2\omega}{D}\right)$*-approximation algorithm for WBCMD running in polynomial time.*

Proof. The proof is identical to the proof of Theorem 1 after observing that (i) the definition of G^λ can be naturally extended to every real value of $\lambda \geq 0$, (ii) the algorithm `Gonzalez` can be naturally extended to deal with edge-weighted graphs, (iii) Lemma 1 holds for the extension of `Gonzalez` to edge-weighted graphs, and (iv) Lemma 2 holds for the extension of G^λ for real values of λ. □

---

8 The length of a path is now measured w.r.t. its edge weights, i.e., the length of a path $(v_1, e_1, \ldots, v_h, e_h, v_{h+1})$ is equal to $\sum_{i=1}^{h} w(e_i)$. Thus the addition of parallel edges is allowed.

Theorem 9. *Algorithm* 2 *returns a* $\left(2 - \frac{1}{B}, 2\right)$*-approximate solution for WBCMD.*

Proof. The proof is identical to the proof of Theorem 6 after observing that Lemma 6 can be extended to prove that there exists a $(2B)$-clustering $\langle U_1, \dots, U_{2B} \rangle$ of $V(G)$ such that $r_G(U_1) \leq D$ and $r_G(U_i) \leq \frac{D-\omega}{2}$, for every $i \in (1, 2B]$. $\square$

Theorem 10. *For every* $\omega \geq 0$*, and for every* $D \geq 2\omega$*, there exists no* $(c \log n, \delta)$*-approximation algorithm for WBCMD, for some constant* $c > 0$ *and for every* $\delta < 2 - \frac{3\omega}{D}$*, unless* $\mathsf{P} = \mathsf{NP}$*.*

Proof. Let $\omega \geq 0$ and let $D \geq 2\omega$ be fixed. We build the input graph G in the following way. G contains 2 copies G_1, G_2 of G' plus a singleton vertex s such that $s \notin V(G')$. All the edges in G have a weight equal to $D/2 - \omega$. Set $B = 2k^*$. Observe that $n = 1 + 2\hat{n}$. The proof goes along the same line of the proof of Theorem 7. $\square$

Acknowledgements

The authors wish to thank the anonymous referees for their insightful comments.

References

[1] Noga Alon, András Gyárfás, Miklós Ruszinkó, Decreasing the diameter of bounded degree graphs, Journal of Graph Theory 35 (3) (2000) 161–172.
[2] R. Balakrishnan, K. Ranganathan, A Textbook of Graph Theory, Springer, 2000.
[3] Andreas Brandstädt, Van Bang Le, Jeremy Spinrad, Graph classes: a survey, SIAM Monographs on Discrete Mathematics and Applications (1999).
[4] R. Chandrasekaran, Andrew F. Daughety, Location on tree networks: p-centre and n-dispersion problems, Mathematics of Operations Research 6 (1) (1981) 50–57.
[5] Victor Chepoi, Bertrand Estellon, Karim Nouioua, Yann Vaxès, Mixed covering of trees and the augmentation problem with odd diameter constraints, Algorithmica 45 (2) (2006) 209–226.
[6] Victor Chepoi, Yann Vaxès, Augmenting trees to meet biconnectivity and diameter constraints, Algorithmica 33 (2) (2002) 243–262.
[7] F.R.K. Chung, Diameters of graph: old problems and new results, Congressus Numerantium 60 (1987) 295–317.
[8] F.R.K. Chung, M.R. Garey, Diameter bounds for altered graphs, Journal of Graph Theory 8 (4) (1984) 511–534.
[9] Erik D. Demaine, Morteza Zadimoghaddam, Minimizing the diameter of a network using shortcut edges, in: Haim Kaplan (Ed.), Proceedings of the 12th Scandinavian Symposium and Workshops on Algorithm Theory, SWAT, in: Lecture Notes in Computer Science, vol. 6139, Springer, 2010, pp. 420–431.
[10] Yevgeniy Dodis, Sanjeev Khanna, Designing networks with bounded pairwise distance, in: Proceedings of the 31st Annual ACM Symposium on Theory of Computing, STOC, 1999, pp. 750–759.
[11] Paul Erdös, András Gyárfás, Miklós Ruszinkó, How to decrease the diameter of triangle-free graphs, Combinatorica 18 (4) (1998) 493–501.
[12] Paul Erdös, Alfréd Rényi, On a problem in the theory of graphs, Publications of the Mathematical Institute of the Hungarian Academy of Sciences (series B) (7) (1963) 623–639.
[13] Teofilo F. Gonzalez, Clustering to minimize the maximum intercluster distance, Theoretical Computer Science 38 (1985) 293–306.
[14] Elena Grigorescu, Decreasing the diameter of cycles, Journal of Graph Theory 43 (4) (2003) 299–303.
[15] Dorit S. Hochbaum, David B. Shmoys, A unified approach to approximation algorithms for bottleneck problems, Journal of the ACM 33 (3) (1986) 533–550.
[16] Toshimasa Ishii, Shigeyuki Yamamoto, Hiroshi Nagamochi, Augmenting forests to meet odd diameter requirements, Discrete Optimization 3 (2) (2006) 154–164.
[17] David S. Johnson, Approximation algorithms for combinatorial problems, Journal of Computer and System Science 9 (3) (1974) 256–278.
[18] Sanjiv Kapoor, Mohammad Sarwat, Bounded-diameter minimum-cost graph problems, Theory of Computing Systems 41 (4) (2007) 779–794.
[19] O. Kariv, S.L. Hakimi, An algorithmic approach to network location problems, SIAM Journal on Applied Mathematics 37 (3) (1979) 513–538.
[20] Chung-Lun Li, S. Thomas McCormick, David Simchi-Levi, On the minimum-cardinality-bounded-diameter and the bounded-cardinality-minimum-diameter edge addition problem, Operations Research Letters 11 (1992) 303–308.
[21] Adam Meyerson, Brian Tagiku, Minimizing average shortest path distances via shortcut edge addition, in: Irit Dinur, Klaus Jansen, Joseph Naor, José D.P. Rolim (Eds.), Proceedings of the 12th International Workshop on Approximation, Randomization, and Combinatorial Optimization (APPROX-RANDOM), in: Lecture Notes in Computer Science, vol. 5687, Springer, 2009, pp. 272–285.
[22] Jan Plesník, On the computational complexity of centers locating in a graph. Aplikace Mat., 25.
[23] Jan Plesník, The complexity of designing a network with minimum diameter, Networks 11 (1) (1981) 77–85.
[24] Ran Raz, Shmuel Safra, A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP, in: Proceedings of the 29th Annual ACM Symposium on Theory of Computing, STOC, 1997, pp. 475–484.
[25] Anneke A. Schoone, Hans L. Bodlaender, Jan van Leeuwen, Diameter increase caused by edge deletion, Journal of Graph Theory 11 (3) (1987) 409–427.