



Online and Approximate Network Construction from Bounded Connectivity Constraints

Jesper Jansson¹(✉), Christos Levcopoulos², and Andrzej Lingas²

¹ The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
jesper.jansson@polyu.edu.hk

² Department of Computer Science, Lund University, 22100 Lund, Sweden
{Christos.Levcopoulos, Andrzej.Lingas}@cs.lth.se

Abstract. The *Network Construction problem*, studied by Angluin et al., Hodosa et al., and others, asks for a minimum-cost network satisfying a set of *connectivity constraints* which specify subsets of the vertices in the network that have to form connected subgraphs. More formally, given a set V of vertices, construction costs for all possible edges between pairs of vertices from V , and a sequence $S_1, S_2, \dots \subseteq V$ of connectivity constraints, the objective is to find a set E of edges such that each S_i induces a connected subgraph of the graph (V, E) and the total cost of E is minimized. First, we study the online version where every constraint must be satisfied immediately after its arrival and edges that have already been added can never be removed. We give an $O(B^2 \log n)$ -competitive and $O((B + \log r) \log n)$ -competitive polynomial-time algorithms along with an $\Omega(B)$ -competitive lower bound, where B is an upper bound on the size of constraints, while r , n denote the number of constraints and the number of vertices, respectively. In the cost-uniform case, we provide an $\Omega(\sqrt{B})$ -competitive lower bound and an $O(\sqrt{n}(\log n + \log r))$ -competitive upper bound with high probability, when constraints are unbounded. All our randomized competitive bounds are against an adaptive adversary, except for the last one which is against an oblivious adversary. Next, we discuss a hybrid approximation method for the (offline) Network Construction problem combining an approximation algorithm of Hosoda et al. with one of Angluin et al. and an application of the hybrid method to bioinformatics. Finally, we consider a natural strengthening of the connectivity requirements in the Network Construction problem, where each constraint is supposed to induce a subgraph (of the constructed graph) of diameter at most d . Among other things, we provide a polynomial-time $((\binom{B}{2} - B + 2)\binom{B}{2})$ -approximation algorithm for the Network Construction problem with the d -diameter requirements.

Keywords: Network optimization · Induced subgraph · Connectivity · Approximation algorithm · Online algorithm

Research supported in part by VR grant 2017-03750 (Swedish Research Council).

© Springer Nature Switzerland AG 2021

T. Calamoneri and F. Corò (Eds.): CIAC 2021, LNCS 12701, pp. 314–325, 2021.

https://doi.org/10.1007/978-3-030-75242-2_22

1 Introduction

Korach and Stern introduced the problem of interconnecting possibly overlapping groups of users by a network such that the users in the same group do not need to use connections outside the group [9]. The optimization objective is to minimize the total cost of the pairwise connections. Angluin et al. and Chockler et al. studied this problem in [2] and [5], respectively.

Angluin et al. showed in [2] that if $P \neq NP$ and n is the number of vertices in the network then the problem cannot be approximated within a factor that is sublogarithmic in n , even in the uniform edge cost case. On the other hand, they proved that a greedy heuristic can approximate the optimal solution within a factor of $O(\log r)$, where r is the number of constraints. As observed in [2], the lower bound matches the upper bound in case r is polynomial in n .

Angluin et al. also studied the online version of this problem where each constraint has to be satisfied directly after its arrival [2]. Their motivation for this problem variant was to help infer the structure of a social network describing the spread of diseases in a community and to decide where to allocate resources to fight an epidemic efficiently. They assumed that the individuals affected by each outbreak of a disease are specified by a connectivity constraint, that the outbreaks occur over time, and that resources that have been committed cannot be released. They provided an $O(n \log n)$ -competitive online algorithm for the online version along with an $\Omega(n)$ -competitive lower bound. They also considered the uniform cost case of this online version, providing an $O(n^{2/3} \log^{2/3} n)$ -competitive algorithm against an oblivious adversary and an $\Omega(\sqrt{n})$ -competitive lower bound against an adaptive adversary.

Hosoda et al. studied a *B-constraint-bounded* variant of the Network Construction problem, where the cardinality of each connectivity constraint S_i does not exceed B [8]. This corresponds to constructing a minimum overlay network for a topic-based peer-to-peer pub/sub system where users (represented by vertices) who are interested in a common topic (represented by connectivity constraints) form connected subgraphs, and moreover, the number of users following each topic is bounded by a constant due to the publisher of that topic having a limited number of available slots for users. Hosoda et al. provided a polynomial-time approximation algorithm for this variant and proved its APX-completeness in [8].

A natural generalization of the Network Construction problem, where some pairwise connections are given a priori has applications in bioinformatics [11]. The purpose is to infer protein-protein interactions that are missing from a database based on a collection of known, overlapping protein complexes (see Sect. 4.1 for details).

1.1 The Structure of the Paper and Our New Results

The next section defines the Network Construction problem and its *B-constraint-bounded* variant, where each constraint includes at most B vertices. We also recall the Minimum Weight Set Cover problem and some facts about its

approximability. In Sect. 3, we study the online version of the B -constraint-bounded Network Construction problem. We present $O(B^2 \log n)$ -competitive and $O((B + \log r) \log n)$ -competitive polynomial-time algorithms for the online B -constraint-bounded Network Construction problem, where r , n stand for the number of constraints and the number of vertices in the network, respectively. In the cost-uniform case when constraints are unbounded, we provide an $O(\sqrt{n}(\log n + \log r))$ -competitive upper bound with high probability. All our randomized competitive bounds are against an adaptive adversary but for the last one which is against an oblivious adversary. We also provide a $(B-1)$ -competitive lower bound in case of arbitrary edge costs and a \sqrt{B} -competitive lower bound in case of uniform edge costs. In Sect. 4, we study approximation algorithms for the offline Network Construction problem and its extensions. First, we discuss a hybrid approximation method combining the approximation algorithm of Hosoda et al. from [8] with that of Angluin et al. from [2] in the context of the application to bioinformatics. Next, we consider a natural strengthening of the connectivity requirements in the Network Construction problem. Each constraint is supposed to induce a subgraph (of the constructed graph) of diameter at most d , where d is given a priori. We provide a polynomial-time $\binom{B}{2} - B + 2$ approximation algorithm for the aforementioned problems with the d -diameter requirements, when each constraint has at most B vertices. Also, we present a polynomial-time algorithm achieving a non-trivial approximation ratio in the general case of the d -diameter variant, where the size of constraints is unbounded. We conclude with final remarks.

Our approximate or online solutions to the aforementioned variants with bounded constraints can be used to solve approximately or online the corresponding variants with unbounded constraints by splitting the constraints into small and large ones (Sects. 3, 4).

2 Preliminaries

For a positive integer r , the term $[r]$ will denote $\{1, \dots, r\}$, and for sets S , V , $|S|$ will stand for the cardinality of S while V^2 for $\{\{v, u\} | v, u \in V\}$.

A *subgraph* of a graph (V, E) is a graph (V', E') such that $V' \subseteq V$ and $E' \subseteq E$. The subgraph of a graph (V, E) *induced* by a subset S of V is the graph $(S, E \cap S^2)$. A *perfect cut* of a graph (V, E) is a partition of V into subsets V' and V'' such that $E \cap \{\{v, u\} | v \in V' \ \& \ u \in V''\} = \emptyset$. The *diameter* of a graph (V, E) is the minimum number ℓ such that any pair of vertices in V can be connected by a path composed of at most ℓ edges in E . If the graph is disconnected, its diameter is undefined.

The *Network Construction* problem is as follows [2]. We are given a set V of vertices and for each possible edge $e = \{v_i, v_j\}$, the cost $c(e)$ of its construction. We are also given a collection of connectivity constraints $S = \{S_1, \dots, S_r\}$, where each S_i is a subset of V . The objective is to construct a set E of edges in V^2 such that for $i = 1, \dots, r$, the subgraph of the graph (V, E) induced by S_i is connected and the total cost of the edges in E is minimal. In the *uniform-cost* case of the

problem, we have $c(e) = 1$ for all edges in V^2 . We can naturally generalize the problem to include the *Network Extension* problem, where some subset E' of edges is already given (constructed) a priori. Note that when zero construction costs of edges are allowed the Network Construction problem is equivalent to that of Network Extension. Simply it is sufficient to set the construction costs of the edges given a priori to zero in order to obtain an equivalent version of the Network Construction problem. In order to avoid duplications in our statements, in the aforementioned situation we shall mention only the Network Construction problem.

Among other things, the following fact was established by Angluin et al. in [2].

Fact 1 (Theorem 2 in [2]). *There is a polynomial-time $O(\log r)$ -approximation algorithm for the Network Construction problem on r constraints.*

We shall also consider a *B-constraint-bounded* variant of the Network Construction problem, where the cardinality of each connectivity constraint S_i does not exceed B . It was studied by Hosoda et al. in [8]. They provided a polynomial-time approximation algorithm for this variant and showed its APX-completeness.

Fact 2 (Theorem 4 in [8]). *There is a polynomial-time $\lfloor B/2 \rfloor \lceil B/2 \rceil$ -approximation (i.e., $\approx B^2/4$ -approximation) algorithm for the B-constraint-bounded Network Construction problem.*

In the context of the Network Construction and Extension problems, we refer to two types of edges: those already constructed and the remaining ones that potentially could be constructed. For instance, when referring to a perfect cut, we consider the edges of the first type while when we refer to edges crossing a perfect cut we mean the edges of the second type.

Recall the definition of *Minimum Weight Set Cover* problem. The input to this problem is a universal set U on n elements and a family F of m subsets of U . Each subset in F is assigned a non-negative weight. A *set cover* is a sub-family of F whose union is equal to U . The objective is to find a set cover of minimum total weight. The decision version of this optimization problem is already NP-hard in the uniform-weight case [6]. The so-called *Minimum Weight Hitting Set* problem is an equivalent formulation of the Minimum Weight Set Cover problem with the roles of elements and subsets exchanged. Here, the input is a finite set S of weighted elements and a family C of subsets of S . The objective is to find a minimum weight subset of S that hits all the subsets in C , i.e., that has a non-empty intersection with each of the subsets in C . This problem is known to be equivalent to Minimum Weight Set Cover [3]. Consequently, approximation algorithms and inapproximability results for each of them carry over to the other one.

Hochbaum [7] used a relaxation of an integer linear programming formulation to obtain an approximation of the Minimum Weight Set Cover in cubic time. The same approximation ratio was obtained by Bar-Yehuda and Even [4] with a more direct, linear-time method. We summarize their results as follows.

Fact 3. *The Minimum Weight Set Cover problem (U, F) , where each element of the universal set U occurs in at most B subsets of U in F , can be approximated within multiplicative factor B in linear time. Consequently, the Minimum Weight Hitting set problem, where each subset in the given family has cardinality at most B , can be approximated within B in linear time.*

3 Online B -Constraint-Bounded Network Construction

In this section, we consider the online version of the Network Construction problem studied in [2]. It arises naturally in the situation when the knowledge about the relationships between the entities represented by vertices changes over time. In the online version, the collection of connectivity constraints is given one at a time. When a constraint S_i is presented, the online algorithm is in round i . The algorithm is now supposed to satisfy this constraint during this round by constructing, if necessary, additional edges before the start of the next round (no previously constructed edges may be removed). The next constraint is then presented in round $i + 1$. To study the worst-case performance of our online algorithms, we shall use an *adaptive adversary* that can wait with setting the next constraint until the online algorithm satisfied the previous one. We shall use competitive analysis of our online algorithms. An online algorithm is *c-competitive* if the cost of its solution does not exceed c times the cost of an optimal offline solution.

3.1 Upper Bounds

First consider the following online Fractional Network Construction problem: For a set V of vertices and edge costs $c(e)$ for $e \in V^2$, and sequence of connectivity constraints S_1, \dots, S_r , assign fractional capacities $w(e)$ to the edges e such that for each $i \in [r]$, for each pair of vertices in S_i , the maximum flow between them is at least 1. The optimization objective is to minimize $\sum_e w(e)c(e)$.

Fact 4 (Lemma 2 in [2]). *There is an $O(\log n)$ -competitive polynomial-time algorithm for the online Fractional Network Construction problem on n vertices.*

By using this fact, we obtain the following theorem.

Theorem 1. *There is an $O(B^2 \log n)$ -competitive polynomial-time algorithm for the online B -constraint-bounded Network Construction problem on n vertices.*

Proof. Run the online $O(\log n)$ -competitive algorithm for the online Fractional Network Construction problem from Fact 4. Disregard all edges that are assigned capacity smaller than B^{-2} by the online solution to the fractional problem and construct all the remaining edges. Note that after the edges of capacity smaller than B^{-2} are removed, for any pair of vertices in any B bounded constraint the maximum flow is still at least $1 - \binom{B}{2}B^{-2} \geq \frac{1}{2}$. Hence, there is a path composed of the constructed edges between such a pair. The cost of the constructed edges is at most B^2 times larger than the cost of the fractional solution, i.e., the sum of products of edge cost and edge capacity over all edges. \square

To derive another competitive upper bound for the online B -constraint-bounded Network Construction problem, we shall consider the online version of the Minimum Weight Set Cover problem. In this version, a family of subsets of the universal set is given a priori while the elements of a subset of the universal set are presented online one at a time [1]. A new element has to be covered before the arrival of the next one. Analogously, in the online version of the equivalent Minimum Weight Hitting set problem, the set of hitting elements is given a priori, and the sets to be hit arrive online one at a time. A new set has to be hit before the arrival of the next one. Alon et al. established the following fact in [1].

Fact 5. *There is an $O(\log n \log r)$ -competitive polynomial-time algorithm for the online Minimum Weight Set Cover problem, where n is the cardinality of the universal set and r is the cardinality of the given family of subsets of the universal set. Consequently, there is an $O(\log n \log r)$ -competitive polynomial-time algorithm for the online Minimum Weight Hitting set, where n is the cardinality of the family of sets to hit and r is the cardinality of the set of all possible hitting elements.*

By combining Fact 5 with the reduction of the Network Construction problem to the Minimum Weight Set Cover problem given by Angluin et al. in [2], we obtain another competitive upper bound for the online B -constraint-bounded Network Construction problem.

Theorem 2. *There is an $O((B + \log r) \log n)$ -competitive polynomial-time algorithms for the online B -constraint-bounded Network Construction problem with n vertices and r constraints.*

Proof. We shall reduce the online Network Construction problem to the online Minimum Weight Hitting Set problem, following the reduction of the former problem to the online Minimum Weight Set Cover problem from [2]. The set of the possible hitting elements given a priori is just the set of all possible edges. Each edge has weight equal to the cost of its construction. Next, each constraint upon its arrival online, for each perfect cut of the subgraph induced by the constraint, yields the set of all (additional potential) edges crossing the perfect cut, i.e., having endpoints in the two different parts of the bipartition. Note that the constraint is satisfied if and only if each perfect cut in the subgraph induced by it is crossed by some edge accounted to the online formed hitting set. It follows that the cost of an optimal solution to the resulting online Minimum Hitting Set problem is the same as that to the original Network Construction problem. Now it is sufficient to observe that the former problem has $O(n^2)$ possible hitting elements and at most $r2^B$ sets to hit, and then to apply Fact 5. \square

We can use Theorem 2 to derive a competitive upper bound for the uniform-cost variant of the Network Construction problem with unbounded constraints. Angluin et al. considered also the uniform cost variant of the Network Construction problem in [2], providing an $O(n^{2/3} \log^{2/3} n)$ -competitive algorithm against

an oblivious adversary and an $\Omega(\sqrt{n})$ -competitive lower bound against an adaptive adversary. Our upper bound in the uniform case is as that of Angluin et al. against an oblivious adversary, i.e., an adversary not knowing the randomized results of the algorithm. The key idea is to split the constraints into small and large ones, and use Theorem 2 to process the former ones.

Theorem 3. *The uniform cost online Network Construction problem on n vertices and r constraints admit an $O(kn^{0.5}(\log n + \log r))$ -competitive polynomial-time solution, for every positive k , with probability at least $1 - O((nr)^{-1}) - \frac{1}{k}$ provided that r is known in advance and the adversary is oblivious.*

Proof. Split the set of constraints into two sets, one consisting of all constraints of size $\leq n^{0.5}$ and one consisting of the rest. We can apply the $O((B + \log r) \log n)$ competitive algorithm from Theorem 2 to the small constraints obtaining an $O((n^{0.5} + \log r) \log n)$ competitive solution. To satisfy the large constraints with more than $n^{0.5}$ vertices we proceed as follows.

Let Q be the set of vertices involved in the large constraints, and let q stand for the cardinality of Q . We initialize an empty vertex set S . Upon an arrival of a new large constraint, each vertex v in the constraint that is outside of S is added to S with probability $q^{-0.5}(\ln n + 2 \ln r)$. We may assume w.l.o.g. that $\ln n + 2 \ln r < q^{0.5}$. Furthermore, if v is added to S then all missing edges incident to v are constructed. It follows that the expected total number, and hence, the expected total cost of the so constructed edges amounts to $q^{0.5}(q-1)(\ln n + 2 \ln r)$. Thus, the total cost is at most $kq^{0.5}(q-1)(\ln n + 2 \ln r)$ with probability at least $1 - \frac{1}{k}$ by Markov's inequality. For each large constraint, the probability that it does not contain any vertex from S is at most

$$\left(1 - \frac{q^{0.5}(\ln n + 2 \ln r)}{q}\right)^{n^{0.5}} \leq \left(1 - \frac{1}{n^{0.5}}\right)^{n^{0.5}(\ln n + 2 \ln r)} \leq O\left(\frac{1}{nr^2}\right)$$

Since there are at most r large constraints, the cost of an optimal solution is at least $q-1$ and $q \leq n$, we obtain an $kn^{0.5}(\ln n + 2 \ln r)$ competitive upper bound for the large constraints with probability at least $1 - O((nr)^{-1}) - \frac{1}{k}$. \square

3.2 Lower Bounds

We present two lower bounds on the competitiveness of algorithms for the online B -constraint-bounded Network Construction problem.

When the edge costs can be arbitrary, it is not possible to achieve a competitive ratio smaller than $B-1$.

Theorem 4. *For any $c < 1$, there is no $c(B-1)$ -competitive algorithm for the online B constraint-bounded Network Construction problem.*

Proof. We modify the proof of Theorem 6 in [2] for the competitive ratio in the general case of online Network Construction, in our case the optimal offline solution is not necessarily a path. Following [2], we set the cost of edges among

the first $n - 1$ vertices to zero, and the cost of all edges incident to the last vertex to 1. The adversary divides the first $n - 1$ vertices into blocks of $B - 1$ vertices. Assume first that $n - 1$ is divisible by $B - 1$. Then for $i = 1, \dots, (n - 1)/(B - 1)$, the adversary repetitively picks an l -tuple of vertices, $l \in [2, B]$, that includes the last vertex and all vertices from the i -th block that are not endpoints of already constructed edges incident to the last vertex. In this way, the algorithm is forced to construct all $B - 1$ edges connecting the vertices in the i -th block with the last vertex while in the optimal offline solution only such a last edge is needed while the vertices in the i -th block are connected in the order of their removal by a constructed path. Thus, the algorithm constructs $(B - 1) \times \frac{n-1}{B-1}$ edges of cost 1 while the optimal offline solution uses only $\frac{n-1}{B-1}$ edges of cost 1. It follows that the algorithm cannot be $c(B - 1)$ competitive.

If $n - 1$ is not divisible by $B - 1$ then the algorithm constructs at least $(B - 1) \times \lfloor \frac{n-1}{B-1} \rfloor + 1$ edges of cost 1 while the optimal offline solution uses only $\lceil \frac{n-1}{B-1} \rceil$ edges of cost 1. Hence, for enough large n , the algorithm cannot be $c(B - 1)$ -competitive. \square

For the uniform cost case, we can present a weaker lower bound. The proof of the following theorem can be found in the full version of this paper.

Theorem 5. *The online uniform cost B constraint-bounded Network Construction problem has an $\Omega(\sqrt{B})$ competitive lower bound.*

4 Offline Approximation Algorithms

In this section, we discuss first a hybrid approximation method for the offline Network Construction problem and its application to bioinformatics. It combines the approximation algorithm of Hosoda et al. from [8] with that of Angluin et al. from [2]. Next, we present approximation algorithms for a strengthened version of the Network Construction problem, where each constraint is supposed to induce a subgraph (of the constructed graph) of diameter at most d for a d given a priori.

4.1 A Hybrid Method with Biological Applications

An application of the Network Extension problem to bioinformatics was given in [11]. There, the goal was to infer protein-protein interactions (PPIs) that were missing from a database based on a collection of known, overlapping protein complexes. More precisely, the vertices V in the input graph were used to represent proteins, the set E' of a priori given edges represented PPIs already in the database, and each input connectivity constraint S_i consisted of the proteins belonging to a single protein complex. Using the assumption that each protein complex must induce a connected subgraph, solving instances of the Network Extension problem gave lower bounds on the number of missing PPIs in various widely used PPI databases. The overwhelming majority of complexes in the

existing PPI databases seem to be of small size, containing at most 10 proteins each, but a few larger ones with up to 100 proteins also occur (for details, see Table A3 in the Supporting Information file for [11]).

The aforementioned statistics suggest a hybrid method consisting of applying the approximation algorithm of Hosoda et al. from [8] to the constraints corresponding to small complexes and that of Angluin et al. from [2] to the constraints corresponding to larger complexes. We can express it in terms of the Network Construction problem by the equivalence observed in Sect. 2. The output is the union of the output of each of the two algorithms applied separately. Hence, by combining Fact 1 with Fact 2, we obtain the following theorem.

Theorem 6. *Consider an instance of the Network Construction problem. For $B \in [n] \setminus \{1\}$, let r_B be the number of constraints with more than B vertices in the instance. A solution to the instance (for the respective problem) of total cost not exceeding $\min_{B \in [n] \setminus \{1\}} \lfloor B/2 \rfloor \lceil B/2 \rceil + O(\log r_B)$ times the minimum can be found in polynomial time.*

The hybrid method will be useful when there is a relatively small $B \in [n] \setminus \{1\}$ such that the number r_B of large constraints including more than B vertices, i.e., the number of large complexes in the biological application, is small. More details about the hybrid method can be found in the full version of this paper.

4.2 Bounded Diameter Requirements

One can naturally strengthen the connectivity requirements in the Network Construction or Extension problems by demanding that each constraint should induce a subgraph of the constructed network of diameter at most d , where $d \in [n - 1]$ is given a priori (cf. [5]).

For instance, Chockler et al. studied the Network Construction problem in [5] using a different terminology. They considered the problem of constructing an optimal overlay (network) that for each topic (constraint) includes a dissemination tree composed of nodes interested in the topic (i.e., belonging to the constraint). One of the measures of the quality of such an overlay suggested on p. 116 of [5] is the diameter. Intuitively, having a low diameter is good because it means that two users interested in the same topic do not need to rely on many intermediate parties, which leads to more efficient communication and better performance.

We shall term the strengthened version of the Network Construction problem as the d -diameter Network Construction problem. In fact, the latter problem restricted to instances with a single constraint is already hard. The restriction can be simply rephrased as follows: given a vertex set V , edge costs $c(e)$ for potential edges in V^2 , find a cheapest graph spanning V with diameter not exceeding d .

The d -diameter Network Construction problem restricted to single constraint instances is known to be NP-hard already for $d = 2$ [10]. In contrast, when restricted to instances with uniform edge costs, this problem variant becomes trivial as any spanning star graph provides an optimal solution.

Analogously to the preceding sections, we can consider the d -diameter Network Construction problem with constraints of cardinality not exceeding B . By using an *auxiliary problem*, we can obtain a $((\binom{B}{2} - B + 2)\binom{B}{2})$ approximation in polynomial time for the B -constraint-bounded d -diameter Network Construction problem. The auxiliary problem is as follows.

For an instance of the B -constraint-bounded d -diameter Network Construction problem with a vertex set V , edge construction costs $c(e)$, a set E' of edges e with $c(e) = 0$, and connectivity constraints S_1, \dots, S_r find a minimum cost edge set $E'' \subseteq V^2 \setminus E'$ such that for $i = 1, \dots, r$, if the diameter of the subgraph of $G' = (V, E)$ induced by S_i is larger than d then $E'' \cap S_i^2 \neq \emptyset$.

The following lemma provides an approximation algorithm for the auxiliary problem.

Lemma 1. *The auxiliary problem can be approximated within $\binom{B}{2}$ in polynomial time.*

Proof. Consider an instance of the auxiliary problem with a vertex set V , edge construction costs $c(e)$, a set E' of edges with zero construction cost, and connectivity constraints S_1, \dots, S_r . We may assume w.l.o.g. that for $i = 1, \dots, r$, the diameter of the subgraph of the graph $G' = (V, E')$ induced by S_i is larger than d since otherwise the constraint S_i can be disregarded. To solve the auxiliary problem, for $i = 1, \dots, r$, form the set E_i of all edges in $S_i^2 \setminus E'$. The auxiliary problem is equivalent to finding a minimum weight subset of the set of all potential edges that hits all the sets E_1, \dots, E_r , where the weights of the edges are equal to their construction costs. By our assumptions, for $i = 1, \dots, r$, $|E_i| \leq \binom{B}{2}$ hold. Now it is sufficient to apply Fact 3 in order to obtain a $\binom{B}{2}$ approximation for the auxiliary problem in time linear in the total size of the family $\{E_1, \dots, E_r\}$ and $|V|^2$. The latter size is in turn polynomial in the size of the input instance of the auxiliary problem. \square

Now, in order to provide an approximate solution to an instance of the B -constraint-bounded d -diameter Network Construction problem, we iterate the method of Lemma 1 as shown in Fig. 1.

Theorem 7. *The B -constraint-bounded d -diameter Network Construction problem can be approximated within $((\binom{B}{2} - B + 2)\binom{B}{2})$ in polynomial time.*

Proof. We shall analyze the iterative method based on Lemma 1. Since for $i = 1, \dots, r$, $|S_i| \leq B$, the subgraph of the original graph $G' = (V, E')$ induced by S_i can be completed by at most $\binom{B}{2}$ edges. Hence, at most $\binom{B}{2}$ iterations of the while block are sufficient. In fact, already $\binom{B}{2} - B + 2$ iterations are sufficient since in a graph with B vertices and at least $\binom{B}{2} - (B - 2)$ edges each pair of non-adjacent vertices has a common neighbor. Note that the cost of an optimal solution to any of the at most $\binom{B}{2} - B + 2$ auxiliary problems approximately solved in consecutive iterations of the while block cannot be greater than that of an optimal solution to the original B -constraint-bounded d -diameter Network Construction problem. Hence, the upper bound $((\binom{B}{2} - B + 2)\binom{B}{2})$ on the approximation factor of the iterative method follows from Lemma 1. \square

Input: a vertex set V , edge construction costs $c(e)$, a set E' of edges e with $c(e) = 0$, and d -diameter constraints S_1, \dots, S_r of cardinality $\leq B$.

Output: A set of edges yielding an $\left(\binom{B}{2} - B + 2\right)\binom{B}{2}$ approximation to the d -diameter Network Construction problem for the input instance.

- 1: $E'' \leftarrow E'$
- 2: **while** there is a not yet satisfied constraint S_i **do**
- 3: Use the method of Lemma 1 to solve the auxiliary problem approximately by an edge set $E''' \subseteq V^2 \setminus E''$
- 4: $E'' \leftarrow E'' \cup E'''$ (E'' can be interpreted as the set of already constructed edges)
- 5: Remove all connectivity constraints S_i such that the subgraph of $G' = (V, E'')$ induced by S_i has diameter $\leq d$.
- 6: Set the construction costs of the edges in E''' to zero.
- 7: **end while**
- 8: **return** $E'' \setminus E'$

Fig. 1. The $\left(\binom{B}{2} - B + 2\right)\binom{B}{2}$ approximation algorithm for the B -constraint-bounded-diameter Network Construction problem.

In the general case with unbounded constraints, straightforward greedy approaches do not seem to work. However, if the edge costs are uniform, we can obtain a large but still a nontrivial approximation factor in polynomial time by splitting the constraints into small and large ones, and using Lemma 1 to obtain an approximation for the former. The proof is analogous to that of Theorem 3. It can be found in the full version of this paper.

Theorem 8. *The uniform cost d -diameter Network Construction problem with n vertices and r constraints admits an $O(n^{0.8}(\ln n + \ln r))$ approximation with probability at least $1 - (nr)^{-1}$ in polynomial time.*

5 Final Remarks

It would be useful to tighten the upper and lower competitiveness bounds on the online version of the B -constraint-bounded Network Construction problem. It would be especially interesting to know if the factor that is logarithmic in n can be removed from the upper bounds.

As mentioned in Sect. 4.2, straightforward greedy approaches do not seem to work for the d -diameter Network Construction problem with unbounded constraints. One reason for this is that natural candidates for potential functions in greedy methods seem to lack the submodularity property. It is an interesting question if it is possible to achieve a reasonable approximation factor for this problem in the general case, at least when edge costs are uniform.

Acknowledgments. We would like to thank Tatsuya Akutsu and Natsu Nakajima for introducing us to the problem studied in this paper.

References

1. Alon, N., Awerbuch, B., Azar, Y., Buchbinder, N., Naor, J.: The online set cover problem. *SIAM J. Comput.* **39**(2), 361–370 (2009)
2. Angluin, D., Aspnes, J., Reyzin, L.: Network construction with subgraph connectivity constraints. *J. Comb. Optim.* **29**(2), 418–432 (2013). <https://doi.org/10.1007/s10878-013-9603-2>
3. Ausiello, G., D’Atri, A., Protasi, M.: Structure preserving reductions among convex optimization problems. *J. Comput. Syst. Sci.* **21**, 136–153 (1980)
4. Bar-Yehuda, R., Even, S.: A linear-time approximation algorithm for the weighted vertex cover problem. *J. Algorithms* **2**(2), 198–203 (1981)
5. Chockler, G.V., Melamed, R., Tock, Y., Vitenberg, R.: Constructing scalable overlays for pub-sub with many topics. In: *Proceedings of the Twenty-Sixth Annual ACM Symposium on Principles of Distributed Computing (PODC 2007)*, pp. 109–118 (2007)
6. Garey, M.R., Johnson, D.S.: *Computers and Intractability - A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, New York (1979)
7. Hochbaum, D.S.: Approximation algorithms for the set covering and vertex cover problems. *SIAM J. Comput.* **11**(3), 555–556 (1982)
8. Hosoda, J., Hromkovič, J., Izumi, T., Ono, H., Steinová, M., Wada, K.: On the approximability and hardness of minimum topic connected overlay and its special instances. *Theor. Comput. Sci.* **429**, 144–154 (2012)
9. Korach, E., Stern, M.: The clustering matroid and the optimal clustering tree. *Math. Program.* **98**, 385–414 (2003). <https://doi.org/10.1007/s10107-003-0410-x>
10. Li, C., McCormick, S.T., Simchi-Levi, D.: On the minimum-cardinality-bounded-diameter and the bounded-cardinality-minimum-diameter edge addition problems. *Oper. Res. Lett.* **11**, 303–308 (1992)
11. Nakajima, N., Hayashida, M., Jansson, J., Maruyama, O., Akutsu, T.: Determining the minimum number of protein-protein interactions required to support known protein complexes. *PLOS ONE* **13**(4) (2018). Article e0195545