Algorithmica (2020) 82:2474–2501
https://doi.org/10.1007/s00453-020-00693-8

The Approximability of Multiple Facility Location on Directed Networks with Random Arc Failures

Refael Hassin[1] · R. Ravi[2] · F. Sibel Salman[3] · Danny Segev[4]

Received: 13 March 2019 / Accepted: 24 February 2020 / Published online: 11 March 2020

Abstract
We introduce and study the *maximum reliability coverage* problem, where multiple facilities are to be located on a network whose arcs are subject to random failures. Our model assumes that arcs fail independently with non-uniform probabilities, and the objective is to locate a given number of facilities, aiming to maximize the expected demand serviced. In this context, each demand point is said to be serviced (or covered) when it is reachable from at least one facility by an operational path. The main contribution of this paper is to establish tight bounds on the approximability of maximum reliability coverage on bidirected trees as well as on general networks. Quite surprisingly, we show that this problem is NP-hard on bidirected trees via a carefully-constructed reduction from the partition problem. On the positive side, we make use of approximate dynamic programming ideas to devise an FPTAS on bidirected trees. For general networks, while maximum reliability coverage is $(1 - 1/e + \epsilon)$-inapproximable as an extension of the max k-cover problem, even estimating its objective value is #P-complete, due to generalizing certain network reliability problems. Nevertheless, we prove that by plugging-in a sampling-based additive estimator into the standard greedy algorithm, a matching approximation ratio of $1 - 1/e - \epsilon$ can be attained.

Keywords Facility location · Random arc failures · FPTAS · Dynamic programming · Hardness

1 Introduction

In this paper, we introduce and study a multiple facility location problem on a network whose arcs are subject to random failures. Specifically, a given number of facilities should be located at the nodes of a directed graph, whose arcs may fail independently with prespecified probabilities. The objective is to maximize the

✉ Danny Segev
segevd@stat.haifa.ac.il

Extended author information available on the last page of the article

expected demand serviced, where the latter expectation is taken over the possible network realizations, and a demand point is said to be serviced (or covered) when it is reachable from at least one facility by an operational path. Problems of this nature find practical applications in computer and telecommunications networks, where arcs represent communication links. In these contexts, arc failures may occur due to a random disruption in communication or due to transmission equipment malfunctions, and service providers are to be located at selected nodes to guarantee the most reliable data services to the demand nodes.

As facility location problems subject to stochastic network failures have attracted a great deal of attention in the last two decades, it is beyond the scope of this paper to provide an exhaustive overview of previous work. To avoid an overly-lengthy exposition, we present a succinct summary of directly-related work in Sect. 1.2, and refer the reader to the references therein for a comprehensive literature review.

Problem formulation Formally, an instance of the *maximum reliability coverage* problem consists of a directed network $G = (V, E)$, where each node v is associated with a non-negative demand d_v. After a stochastic disruption, each arc may exist in either operational or non-operational state, which is referred to as the survival or failure of that arc. To capture the randomness in a disruption event, we assume that each arc e survives with probability p_e, independently of other arcs.

An additional ingredient of the input is a parameter k, specifying an upper bound on the number of facilities to be located. With respect to any realization of the surviving network, a facility located at node v covers the demand of all nodes reachable from v. For a set of facilities $F \subseteq V$, we use $\pi_v(F)$ to denote the probability that node v is covered by at least one facility in F. With this notation, the expected demand covered by F can be written as $\mathcal{E}(F) = \sum_{v \in V} d_v \cdot \pi_v(F)$. The objective of the maximum reliability coverage problem is to compute a set $F \subseteq V$ of at most k facilities that maximizes the expected demand covered $\mathcal{E}(F)$.

1.1 Our Results

The main contribution of this paper is to establish tight bounds on the approximability of maximum reliability coverage on bidirected trees as well as on general directed networks.

Bidirected trees Quite surprisingly, unlike many k-facility location problems on deterministic networks, random arc failures introduce a host of new computational difficulties. Specifically, an interesting reduction from the partition problem enables us to show in Sect. 2 that maximum reliability coverage is in fact NP-hard on bidirected trees. We remind the reader that such networks are obtained by substituting each edge of an undirected tree by two anti-parallel arcs, each potentially with a survival probability of 0. On the positive side, we make use of approximate dynamic programming ideas in Sect. 3 to devise a fully polynomial-time approximation scheme (FPTAS) on bidirected trees. We mention in passing that most of the technical novelty of this paper resides in deriving these two results.

General networks When the underlying network is arbitrary, it is easy to verify that maximum reliability coverage captures the max k-cover problem as a special case. Therefore, maximum reliability coverage cannot be approximated within a constant greater than $1 - 1/e$, unless P = NP [8]. On the positive side, one can also verify that our expected demand coverage function $\mathcal{E}$ is monotone and submodular, meaning that it can be approximated within factor $1 - 1/e$ subject to the k-cardinality constraint (see, e.g., [1, 16]), given an oracle access to $\mathcal{E}$. However, as explained later on, the problem of estimating $\mathcal{E}$ is in fact #P-complete due to generalizing certain network reliability problems [3, 17, 20]. For this reason, we dedicate Sect. 4 to showing that, by plugging-in a sampling-based estimator (with an additive error) into the standard greedy algorithm, the maximum reliability coverage problem can be approximated within factor $1 - 1/e - \epsilon$. Our Monte-Carlo algorithm is successful with constant probability and its running time is polynomial in n and $1/\epsilon$.

Further discussion and open questions In light of the above-mentioned findings, a natural question is whether our results carry over to undirected graphs. We conclude by elaborating on this question and pinpoint specific directions for future research:

- *Trees* While the FPTAS proposed in Sect. 3 is clearly applicable in undirected trees, our NP-hardness proof in Sect. 2 relies on associating anti-parallel arcs with different survival probabilities. It would be interesting to examine whether maximum reliability coverage remains NP-hard on undirected trees.
- *Undirected graphs* Similarly, on general directed networks, the straightforward reduction from max k-cover, that gives an inapproximability bound of $1 - 1/e$ unless P = NP [8], also crucially depends on having asymmetric survival probabilities. Interestingly, we show in "Appendix 2" that maximum reliability coverage remains APX-hard on undirected graphs via a reduction from minimum-cardinality vertex cover on cubic graphs [2]. A seemingly challenging direction for future work would be to significantly narrow the gap between the upper bound established by this reduction (which is very close to 1) and the approximation ratio of $1 - \frac{1}{e} - \epsilon$, attained by our sampling-based greedy algorithm in Sect. 4.

1.2 Related Work

Independent failures In the more widespread setting, edges or nodes are associated with individual survival probabilities, and the underlying assumption is that these components fail independently. Eiselt et al. [6] considered the setting of a single-edge failure for the problem of locating multiple facilities to minimize the total expected demand disconnected from these facilities. They provided an exact polynomial-time algorithm and showed that locating a single facility is equivalent to the 1-median problem. Following up on this work, Eiselt et al. [7] proposed an exact polynomial-time algorithm when either a single node or a single edge may fail.

Melachrinoudis and Helander [14] studied a single-facility location problem on an undirected tree where multiple edges may fail simultaneously and independently. Here, the objective is to maximize the expected number of demand nodes reachable by operational paths, for which the authors devised two exact polynomial-time algorithms. It is worth noting that this setting is a special case of our model, with a single facility, undirected tree, and unit demands. Interestingly, they stated that this problem becomes extremely challenging for multiple-facility location on general networks, which is indeed verified by our hardness and inapproximability results. Later on, Xue [22] gave a linear-time algorithm for the problem studied by Melachrinoudis and Helander [14] as well as for its maximin version. Additional work along these lines includes that of Ding and Xue [5], who considered the same problem on a tree with unreliable nodes and provided a linear-time algorithm based on dynamic programming ideas. This problem has also been studied by Colbourn and Xue [4], who proposed a linear-time algorithm on partial 2-trees. Santivanez et al. [18] also investigated a single-facility location problem under edge failures, with certain restrictions on the network topology. They focused on the maximin objective, stipulating that the most reliable route should be used. Taking advantage of the latter policy, which allows one to focus on a specific path between any two nodes, the authors presented an exact polynomial-time algorithm. It is worth noting that Nel and Colbourn [15] proved that the problem considered by Melachrinoudis and Helander [14] is NP-hard on general networks.

Correlated failures Hassin et al. [11] considered correlated edge failures, motivated by modeling how network links are affected by a disaster event. Their dependency model first sorts the edges by their reliability values and then assumes that the failure of an edge implies the failure of any lower-ranked edge. This linear ordering makes the number of possible network realizations linear in the number of edges, which enables one to avoid certain difficulties related to reliability computations. Hassin et al. studied a multiple-facility location problem under multiple reliability vectors, each inducing a separate linear ordering. Their objective function is that of maximizing the expected total demand covered by the selected facilities, which is similar to the one considered in the current paper. The authors provided exact algorithms for one and two linear orderings by dynamic programming and linear programming, respectively. From a hardness perspective, they proved that the problem becomes NP-hard in the 3-ordering case and $(1 - 1/e + \epsilon)$-inapproximable under an arbitrary number of orderings. To our knowledge, other than having a similar objective function, the probabilistic model considered by Hassin et al. is very different in nature than our independent-failures model, and in particular, algorithmic and hardness results do not seem to migrate from one problem to the other.

2 NP-Hardness for Bidirected Trees

In what follows, we propose a reduction from the partition problem to maximum reliability coverage, showing that the latter problem is NP-hard even when the underlying network is a bidirected tree. For simplicity of presentation, it is convenient to start off in Sect. 2.1 with an intermediate reduction, where the survival probabilities of certain arcs are either irrational or require a pseudo-polynomial number of bits to be specified. With the basic ideas in place, we explain in Sect. 2.2 how to "round" these probabilities into having only polynomially-many bits and how to resolve additional numerical issues, while still being able to distinguish between YES and NO instances of partition.

Theorem 2.1 *Maximum reliability coverage is NP-hard on bidirected trees.*

2.1 Reduction: Pseudo-Polynomial Time

Our proof is based on a reduction from the partition problem, which is well known to be NP-hard [9, 13]. Here, given a collection of positive integers $a_1, \dots, a_n$, whose sum is denoted by A, we wish to decide whether there exists a subset $I \subseteq [n]$ for which $\sum_{i \in I} a_i = \frac{A}{2}$. Given such an instance, we construct a corresponding instance of maximum reliability coverage as follows:

- The network is a bidirected tree on $3n + 1$ nodes, with a root r and n arms, where the top-to-bottom node order on each arm i is x_i, y_i, and z_i.
- The root r is connected to x_i using bidirected arcs, such that the survival probability of (r, x_i) is 1/2 whereas that of (x_i, r) is $1 - e^{-a_i/A}$. Then, x_i is connected to y_i by an arc (x_i, y_i) with survival probability 1, and similarly, z_i is also connected to y_i, again by an arc (z_i, y_i) with probability 1. We mention in passing that the arcs (y_i, x_i) and (y_i, z_i) do not exist.
- The respective demands of x_i, y_i, and z_i are a_i, μA, and λa_i, where $\lambda = \frac{2+3e^{-1/2}}{4}$ and $\mu = \lceil 2(\lambda + 2) \rceil$. The root r has a demand of 0.
- At most n facilities can be located.

It is worth emphasizing that the above reduction is generally not polynomial in $\log A$, due to having probabilities of the form $1 - e^{-a_i/A}$, which may require $\Omega(A)$ bits to be specified. In fact, since these probabilities (and λ) are irrational numbers, additional numerical complications are incurred when one wishes to ensure sufficient accuracy. As previously mentioned, we deal with this issue in Sect. 2.2. For the remainder of our analysis, we use $F^* \subseteq V$ to denote a fixed optimal set of facilities for the resulting instance of maximum reliability coverage. The next claim identifies an important structural property of F^* that will be useful later on.

Lemma 2.2 *F^* picks exactly one of the nodes x_i and z_i within each arm $i \in [n]$.*

Proof We begin by arguing that F^* picks exactly one of x_i, y_i, and z_i within each arm $i \in [n]$. To this end, suppose that the latter property does not hold, meaning that due to having $|F^*| \leq n$ there exists some $j \in [n]$ for which F^* does not pick any of x_j, y_j, and z_j. In this case, y_j can only be covered by facilities different from x_j, y_j, and z_j, implying that the probability for y_j being covered is $\pi_{y_j}(F^*) = \pi_r(F^*) \cdot p_{(r,x_j)} \cdot p_{(x_j,y_j)} \leq 1/2$, since $p_{(r,x_j)} = 1/2$. It follows that

$$
\begin{aligned}
\mathcal{E}(F^*) &= \sum_{i\in[n]} \left(a_i \cdot \pi_{x_i}(F^*) + \mu A \cdot \pi_{y_i}(F^*) + \lambda a_i \cdot \pi_{z_i}(F^*)\right) \\
&\le \sum_{i\in[n]\setminus\{j\}} \left(a_i + \mu A + \lambda a_i\right) + \left(a_j + \frac{\mu A}{2} + \lambda a_j\right) \\
&= n\mu A - \left(\frac{\mu}{2} - (\lambda+1)\right)A \\
&\le (n\mu - 1)A \ ,
\end{aligned}
$$

where the last inequality holds since $\mu = \lceil 2(\lambda+2) \rceil$. This contradicts the trivial lower bound of $\mathcal{E}(F^*) \ge n\mu A$, attained by picking $y_1, \dots, y_n$ as facilities.

We are now left with showing that F^* does not pick y_i within each arm $i \in [n]$. For this purpose, suppose that there exists some $j \in [n]$ for which F^* picks y_j. Then, let us create a modified solution $\tilde{F}^*$ where this facility is relocated from y_j to x_j. Clearly, $\pi_{y_j}(\tilde{F}^*) = \pi_{y_j}(F^*) = 1$, $\pi_{x_j}(\tilde{F}^*) = 1 > \frac{1}{2} \ge \pi_{x_j}(F^*)$, and $\pi_v(\tilde{F}^*) \ge \pi_v(F^*)$ for any other node. As a result, $\mathcal{E}(\tilde{F}^*) - \mathcal{E}(F) \ge a_j \cdot (\pi_{x_j}(\tilde{F}^*) - \pi_{x_j}(F^*)) > 0$, contradicting the optimality of F^*. $\square$

Given the particular structure of optimal sets of facilities, as stated in Lemma 2.2, we proceed by deriving a simple expression for the expected demand coverage of such sets.

Lemma 2.3 *Let $F \subseteq V$ be a feasible set of facilities that picks exactly one of x_i and z_i within each arm $i \in [n]$. Then, $\mathcal{E}(F) = n\mu A + \varphi(\sum_{i\in I_F} a_i)$, where $I_F = \{i : x_i \in F\}$ and $\varphi : [0, A] \to \mathbb{R}$ is the real-valued function defined by $\varphi(x) = x + (A - x) \cdot (\lambda + \frac{1}{2} \cdot (1 - e^{-x/A}))$.*

Proof When exactly one of x_i and z_i is picked within each arm $i \in [n]$, the expected demand coverage $\mathcal{E}(F)$ can be broken down into the following ingredients:

- *y-nodes*: A total demand of $n\mu A$ is covered over $y_1, \dots, y_n$ with probability 1.
- *x-nodes with I_F-indices*: A total demand of $\sum_{i\in I_F} a_i$ is covered over $\{x_i\}_{i\in I_F}$ with probability 1.
- *z-nodes with non-I_F-indices*: A total demand of $\lambda \cdot \sum_{i\notin I_F} a_i$ is covered over $\{z_i\}_{i\notin I_F}$ with probability 1.
- *x-nodes with non-I_F-indices*: The demand a_j of each x_j with $j \notin I_F$ is covered with probability $\frac{1}{2}(1 - e^{-\sum_{i\in I_F} a_i/A})$.

Therefore,

$$\begin{aligned}\mathcal{E}(F) &= n\mu A + \sum_{i \in I_F} a_i + \lambda \cdot \sum_{i \notin I_F} a_i + \frac{1}{2} \cdot \left(1 - e^{-\sum_{i \in I_F} a_i / A}\right) \cdot \sum_{i \notin I_F} a_i \\ &= n\mu A + \varphi\left(\sum_{i \in I_F} a_i\right).\end{aligned}$$

□

We are now ready to show that the optimal expected demand coverage $\mathcal{E}(F^*)$ can be used to distinguish between YES and NO instances of the partition problem:

- *YES instances are mapped to* $\mathcal{E}(F^*) \geq n\mu A + \varphi(\frac{A}{2})$. Suppose there exists a subset $I \subseteq [n]$ for which $\sum_{i \in I} a_i = \frac{A}{2}$. In this case, Lemma 2.3 implies that the set of facilities $F = \{x_i : i \in I\} \cup \{z_i : i \notin I\}$ has an expected demand coverage of $\mathcal{E}(F) = n\mu A + \varphi(\sum_{i \in I} a_i) = n\mu A + \varphi(\frac{A}{2})$. As a result, the optimality of F^* guarantees that $\mathcal{E}(F^*) \geq \mathcal{E}(F) \geq n\mu A + \varphi(\frac{A}{2})$.
- *NO instances are mapped to* $\mathcal{E}(F^*) \leq n\mu A + \varphi(\frac{A}{2}) - \frac{1}{25A}$. By combining Lemmas 2.2 and 2.3, it follows that the optimal expected demand coverage is $\mathcal{E}(F^*) = n\mu A + \varphi(\sum_{i \in I_{F^*}} a_i)$, where $I_{F^*} = \{i : x_i \in F^*\}$. However, one can easily verify by elementary calculus that the function φ is concave, with a unique maximizer at $\frac{A}{2}$. Moreover, as shown in Lemma 2.4 below, for any integer $a \neq \frac{A}{2}$ we actually have $\varphi(a) \leq \varphi(\frac{A}{2}) - \frac{1}{25A}$. Consequently, due to considering a NO instance, $\sum_{i \in I_{F^*}} a_i \neq \frac{A}{2}$, meaning that $\mathcal{E}(F^*) \leq n\mu A + \varphi(\frac{A}{2}) - \frac{1}{25A}$.

Lemma 2.4 $\varphi(a) \leq \varphi(\frac{A}{2}) - \frac{1}{25A}$, *for every integer* $a \neq \frac{A}{2}$.

Proof As previously mentioned, φ is concave, with a unique maximizer at $\frac{A}{2}$. For this reason, it is sufficient to show that $\max\{\varphi(\frac{A}{2} - 1), \varphi(\frac{A}{2} + 1)\} \leq \varphi(\frac{A}{2}) - \frac{1}{25A}$. First, in order to bound $\varphi(\frac{A}{2} - 1)$, note that the (differentiable and concave) function φ lies below all of its tangents, meaning in particular that for the tangent at $\frac{A}{2} - \frac{1}{2}$ we have $\varphi(x) \leq \varphi(\frac{A}{2} - \frac{1}{2}) + \varphi'(\frac{A}{2} - \frac{1}{2}) \cdot (x - (\frac{A}{2} - \frac{1}{2}))$. By substituting $x = \frac{A}{2} - 1$, it follows that

$$\begin{aligned}
\varphi\left(\frac{A}{2}-1\right) &\le \varphi\left(\frac{A}{2}-\frac{1}{2}\right)-\frac{1}{2}\cdot\varphi'\left(\frac{A}{2}-\frac{1}{2}\right)\\
&\le \varphi\left(\frac{A}{2}\right)-\frac{1}{2}\cdot\left(\frac{1}{2}-\lambda+e^{-x/A}\cdot\left(1-\frac{x}{2A}\right)\right)\bigg|_{x=\frac{A}{2}-\frac{1}{2}}\\
&= \varphi\left(\frac{A}{2}\right)-\frac{1}{2}\cdot\left(-\frac{3e^{-1/2}}{4}+e^{-\frac{1}{2}+\frac{1}{2A}}\cdot\left(\frac{3}{4}+\frac{1}{4A}\right)\right)\\
&= \varphi\left(\frac{A}{2}\right)+\frac{e^{-1/2}}{8}\cdot\left(\underbrace{3-3e^{\frac{1}{2A}}}_{\le 0}-\underbrace{\frac{1}{A}e^{\frac{1}{2A}}}_{\ge 1/A}\right)\\
&\le \varphi\left(\frac{A}{2}\right)+\frac{e^{-1/2}}{8}\cdot\frac{1}{A}\\
&\le \varphi\left(\frac{A}{2}\right)-\frac{7}{100A}\,,
\end{aligned}$$

where the second inequality holds since φ is maximized at $\frac{A}{2}$. Now, in order to bound $\varphi(\frac{A}{2}+1)$, we make use of the tangent at $\frac{A}{2}+\frac{1}{2}$, which provides the bound $\varphi(x) \le \varphi(\frac{A}{2}+\frac{1}{2})+\varphi'(\frac{A}{2}+\frac{1}{2})\cdot(x-(\frac{A}{2}+\frac{1}{2}))$. Here, by substituting $x=\frac{A}{2}+1$, it follows that

$$\begin{aligned}
\varphi\left(\frac{A}{2}+1\right) &\le \varphi\left(\frac{A}{2}+\frac{1}{2}\right)+\frac{1}{2}\cdot\varphi'\left(\frac{A}{2}+\frac{1}{2}\right)\\
&\le \varphi\left(\frac{A}{2}\right)+\frac{1}{2}\cdot\left(\frac{1}{2}-\lambda+e^{-x/A}\cdot\left(1-\frac{x}{2A}\right)\right)\bigg|_{x=\frac{A}{2}+\frac{1}{2}}\\
&= \varphi\left(\frac{A}{2}\right)+\frac{1}{2}\cdot\left(-\frac{3e^{-1/2}}{4}+e^{-\frac{1}{2}-\frac{1}{2A}}\cdot\left(\frac{3}{4}-\frac{1}{4A}\right)\right)\\
&= \varphi\left(\frac{A}{2}\right)-\frac{e^{-1/2}}{8}\cdot\left(\underbrace{3-3e^{-\frac{1}{2A}}}_{\ge 0}+\underbrace{\frac{1}{A}e^{-\frac{1}{2A}}}_{\ge e^{-1/2}/A}\right)\\
&\le \varphi\left(\frac{A}{2}\right)-\frac{1}{8e}\cdot\frac{1}{A}\\
&\le \varphi\left(\frac{A}{2}\right)-\frac{1}{25A}.
\end{aligned}$$

□

2.2 Reduction: Truly Polynomial Time

We now turn our attention to ensuring that the survival probability of each arc is sufficiently large, and moreover, can be specified using polynomially-many bits via a simple and efficient calculation. The technical idea for "rounding" these probabilities as well as the irrational parameter $\lambda = \frac{2+3e^{-1/2}}{4}$ is based on Maclaurin series approximation for the exponential function, which states that $|e^x - \sum_{t=0}^{T} \frac{x^t}{t!}| = O(\frac{x^{T+1}}{(T+1)!})$ for any $x \in \mathbb{R}$ (see, e.g., [19, Chap. 20]). In particular, specializing the latter bound to $x \in [-1, 1]$, we have $|e^x - \sum_{t=0}^{T} \frac{x^t}{t!}| = O(\frac{1}{2^{T/2}})$, with room to spare. As a result, by taking $T(K) = \Theta(\log K)$, one has $|e^x - \sum_{t=0}^{T(K)} \frac{x^t}{t!}| \le \frac{1}{K}$, which we instantiate with $K = 200nA^2$. Therefore, rather than associating each arc (x_i, r) with a survival probability of $1 - e^{-a_i/A}$, we use $1 - \mathcal{M}_{T(K)}(-\frac{a_i}{A})$ instead, where $\mathcal{M}_{T(K)}(x) = \sum_{t=0}^{T(K)} \frac{x^t}{t!}$. With this definition, the probability $\mathcal{M}_{T(K)}(-\frac{a_i}{A})$ requires only $O(\log(nA))$ bits to be specified and can be computed (by definition) in polynomial time, which makes our overall reduction polynomial. Similarly, rather than $\lambda = \frac{2+3e^{-1/2}}{4}$, we will be using $\tilde{\lambda} = \frac{2+3\cdot\mathcal{M}_{T(K)}(-1/2)}{4}$.

That said, we still have to show that one can indeed distinguish between YES and NO instances of partition based on the optimal expected demand coverage, which is precisely what Lemmas 2.5 and 2.7 below argue. To avoid confusion, we denote by $\tilde{\mathcal{E}}$ the expected demand coverage function with respect to the resulting instance (with rounded probabilities and $\tilde{\lambda}$), while keeping $\mathcal{E}$ for the analogous function with respect to the original (pre-rounding) instance. In addition, it is worth pointing out that F^* is now denoting an optimal facility set for the new objective function, $\tilde{\mathcal{E}}$.

Lemma 2.5 (YES instances) *Suppose there exists a subset $I \subseteq [n]$ for which $\sum_{i\in I} a_i = \frac{A}{2}$. Then, $\tilde{\mathcal{E}}(F^*) \ge n\mu A + \varphi(\frac{A}{2}) - \frac{1}{100A}$.*

Proof By repeating the proof of Lemma 2.3, where each of the survival probabilities $1 - e^{-a_i/A}$ are replaced by their rounded version $1 - \mathcal{M}_{T(K)}(-\frac{a_i}{A})$ and λ is replaced by $\tilde{\lambda}$, it follows that the set of facilities $F = \{x_i : i \in I\} \cup \{z_i : i \notin I\}$ has an expected demand coverage of

$$\begin{aligned}
\tilde{\mathcal{E}}(F) &= n\mu A + \sum_{i\in I} a_i + \tilde{\lambda} \cdot \sum_{i\notin I} a_i + \frac{1}{2} \cdot \left(1 - \prod_{i\in I} \mathcal{M}_{T(K)}\left(-\frac{a_i}{A}\right)\right) \cdot \sum_{i\notin I} a_i \\
&\ge n\mu A + \sum_{i\in I} a_i + \left(\lambda - \frac{1}{K}\right) \cdot \sum_{i\notin I} a_i + \frac{1}{2} \cdot \left(1 - \prod_{i\in I} \left(e^{-a_i/A} + \frac{1}{K}\right)\right) \cdot \sum_{i\notin I} a_i \\
&\ge n\mu A + \sum_{i\in I} a_i + \left(\lambda - \frac{1}{K}\right) \cdot \sum_{i\notin I} a_i + \frac{1}{2} \cdot \left(1 - e^{-\sum_{i\in I} a_i/A} - \frac{1}{100A^2}\right) \cdot \sum_{i\notin I} a_i \\
&\ge n\mu A + \varphi\left(\sum_{i\in I} a_i\right) - \frac{1}{100A} \\
&= n\mu A + \varphi\left(\frac{A}{2}\right) - \frac{1}{100A},
\end{aligned}$$

where the second inequality follows from the next technical claim, whose proof is provided in "Proof of Claim 2.6" section of "Appendix 1".

Claim 2.6 $\prod_{i \in I}(e^{-a_i/A} + \frac{1}{K}) \le e^{-\sum_{i \in I} a_i/A} + \frac{1}{100A^2}$.

As a result, the optimality of F^* guarantees that $\tilde{\mathcal{E}}(F^*) \ge \tilde{\mathcal{E}}(F) \ge n\mu A + \varphi(\frac{A}{2}) - \frac{1}{100A}$. □

Lemma 2.7 (NO instances) *Suppose that* $\sum_{i \in I} a_i \ne \frac{A}{2}$ *for every subset* $I \subseteq [n]$. *Then,* $\tilde{\mathcal{E}}(F^*) \le n\mu A + \varphi(\frac{A}{2}) - \frac{3}{100A}$.

Proof Once again, by repeating the proof of Lemma 2.3 with the rounded probabilities and $\tilde{\lambda}$ plugged-in, and by defining $I_{F^*} = \{i : x_i \in F^*\}$, it follows that

$$\begin{aligned}
\tilde{\mathcal{E}}(F^*) &= n\mu A + \sum_{i \in I_{F^*}} a_i + \tilde{\lambda} \cdot \sum_{i \notin I_{F^*}} a_i + \frac{1}{2} \cdot \left(1 - \prod_{i \in I_{F^*}} \mathcal{M}_{T(K)}\left(-\frac{a_i}{A}\right)\right) \cdot \sum_{i \notin I_{F^*}} a_i \\
&\le n\mu A + \sum_{i \in I_{F^*}} a_i + \left(\lambda + \frac{1}{K}\right) \cdot \sum_{i \notin I_{F^*}} a_i \\
&\quad + \frac{1}{2} \cdot \left(1 - \prod_{i \in I_{F^*}} \left(e^{-a_i/A} - \frac{1}{K}\right)\right) \cdot \sum_{i \notin I_{F^*}} a_i \\
&\le n\mu A + \sum_{i \in I_{F^*}} a_i + \left(\lambda + \frac{1}{K}\right) \cdot \sum_{i \notin I_{F^*}} a_i \\
&\quad + \frac{1}{2} \cdot \left(1 - e^{-\sum_{i \in I_{F^*}} a_i/A} + \frac{1}{100A^2}\right) \cdot \sum_{i \notin I_{F^*}} a_i \\
&\le n\mu A + \varphi\left(\sum_{i \in I_{F^*}} a_i\right) + \frac{1}{100A} \\
&= \mathcal{E}(F^*) + \frac{1}{100A} ,
\end{aligned}$$

where the second inequality follows from the next technical claim, whose proof is provided in "Proof of Claim 2.8" section of "Appendix 1".

Claim 2.8 $\prod_{i \in I}(e^{-a_i/A} - \frac{1}{K}) \ge e^{-\sum_{i \in I} a_i/A} - \frac{1}{100A^2}$.

Now, as shown at the end of Sect. 2.1, NO instances of the partition problem are mapped to $\mathcal{E}(F^*) \le n\mu A + \varphi(\frac{A}{2}) - \frac{1}{25A}$, and therefore, $\tilde{\mathcal{E}}(F^*) \le n\mu A + \varphi(\frac{A}{2}) - \frac{3}{100A}$. □

3 FPTAS for Bidirected Trees

In what follows, we utilize approximate dynamic programming ideas to devise an FPTAS for the maximum reliability coverage problem on bidirected trees. Here, the underlying tree $T = (V, E)$ can easily be transformed into a binary one, with a pair of anti-parallel arcs connecting every node to each of its children; these arcs are potentially associated with zero survival probabilities. Essentially, when a node has $m \geq 3$ children, we replace the connecting arcs by an $O(\log m)$-depth binary tree with m leaves, each corresponding to a different child. Arcs adjacent to such leaves are identical to those connecting the original child to its parent, whereas all other (interior) arcs survive with probability 1.

For ease of exposition, we present our algorithm and its analysis in an incremental way. First, Sect. 3.1 provides an exact dynamic program for the special case where T is a line. Then, Sect. 3.2 extends these ideas to arbitrary trees while making use of a continuous state space of coverage probabilities. This leads to a non-algorithmic characterization of optimal facility sets. Finally, the main novelty of our method is revealed in Sect. 3.3, where we show how to discretize the state space into polynomially-many coverage probabilities, while losing only a negligible factor in optimality and still preserving the correctness of our dynamic program.

3.1 Warm-Up: Exact Dynamic Program on the Line

Preliminaries We begin by rooting the path T at one of its endpoints r_{up}, where the opposite endpoint is designated by r_{down}; the underlying number of nodes in T will be denoted by n. The important property we exploit in the context of line networks is that, with respect to any set of facilities $F \subseteq V$, each node v has only $O(n)$ possible values for the probability to be covered by a facility located on the path $r_{\text{up}} \rightsquigarrow \psi(v)$, from the root r_{up} to $\psi(v)$, the parent of v. This probability is precisely $p_{f_{\text{up}} \rightsquigarrow \psi(v)} \cdot p_{(\psi(v),v)}$, where $f_{\text{up}} \in F$ is the maximum-depth facility located on $r_{\text{up}} \rightsquigarrow \psi(v)$ and $p_{\cdot \rightsquigarrow \cdot}$ stands for the survival probability of a given path; when no facilities are located on $r_{\text{up}} \rightsquigarrow \psi(v)$, the probability in question is clearly 0. As a result, by defining $\mathcal{P}_{\text{up}}^{v} = \{p_{u \rightsquigarrow \psi(v)} \cdot p_{(\psi(v),v)} : u \in V(r_{\text{up}} \rightsquigarrow \psi(v))\} \cup \{0\}$, we obtain an $O(n)$-sized set that contains any possible probability of v being covered by a facility located on $r_{\text{up}} \rightsquigarrow \psi(v)$, noting that this construction is independent of F. Similarly, one can easily construct an analogous $O(n)$-sized set $\mathcal{P}_{\text{down}}^{v}$ containing any possible probability of v being covered by a facility located on $r_{\text{down}} \rightsquigarrow v$. In this case, we also consider v itself as a potential node to locate a facility, meaning that $\mathcal{P}_{\text{down}}^{v} = \{p_{u \rightsquigarrow v} : u \in V(r_{\text{down}} \rightsquigarrow v)\} \cup \{0\}$.

The dynamic program For every node $v \in V$, number of facilities k, and pair of probabilities $(p_{\text{up}}, p_{\text{down}}) \in \mathcal{P}_{\text{up}}^{v} \times \mathcal{P}_{\text{down}}^{v}$, we define the function value $\mathcal{E}(v, k, p_{\text{up}}, p_{\text{down}})$ as the maximum expected demand coverage collected over all nodes in the subpath T_v rooted at v, given that the coverage probability of v by a facility located on $r_{\text{up}} \rightsquigarrow \psi(v)$ is p_{up}, under the following conditions: (1) The number of facilities located in T_v is at most k; and (2) The coverage probability of v by a facility located on $r_{\text{down}} \rightsquigarrow v$ is p_{down}. With these definitions, since the coverage probability of the root r_{up} by any set of facilities necessarily resides within $\mathcal{P}_{\text{down}}^{r_{\text{up}}}$, the optimal expected demand coverage corresponds to $\max\{\mathcal{E}(r_{\text{up}}, k, 0, p_{\text{down}}) : p_{\text{down}} \in \mathcal{P}_{\text{down}}^{r_{\text{up}}}\}$.

Recursive equation for $\boldsymbol{\mathcal{E}(v,k,p_{\mathrm{up}},p_{\mathrm{down}})}$ We begin by considering the scenario where $p_{\mathrm{down}} < 1$. Here, we must avoid locating a facility at the node v (or otherwise, $p_{\mathrm{down}} = 1$), which leads to an expected demand coverage of

$$(1-(1-p_{\mathrm{up}})\cdot(1-p_{\mathrm{down}}))\cdot d_v + \mathcal{E}\left(u,k,p_{\mathrm{up}}\cdot p_{(v,u)},\frac{p_{\mathrm{down}}}{p_{(u,v)}}\right), \tag{1}$$

where u is the single child node of v. This equation states that v is covered with probability $1-(1-p_{\mathrm{up}})\cdot(1-p_{\mathrm{down}})$, and it remains to decide how the k facilities are located in T_u. We note that states where $\frac{p_{\mathrm{down}}}{p_{(u,v)}} > 1$ are not considered.

Now, in the opposite scenario where $p_{\mathrm{down}} = 1$, the above-mentioned option is still feasible; by plugging in the value of p_{down}, we obtain an expected demand coverage of

$$d_v + \mathcal{E}\left(u,k,p_{\mathrm{up}}\cdot p_{(v,u)},\frac{1}{p_{(u,v)}}\right). \tag{2}$$

Yet another option is to locate a facility at v, leading to an expected demand coverage of

$$d_v + \max\left\{\mathcal{E}(u,k-1,p_{(v,u)},\hat{p}_{\mathrm{down}}) : \hat{p}_{\mathrm{down}} \in \mathcal{P}^u_{\mathrm{down}}\right\}. \tag{3}$$

In this case, the node v is covered with probability 1, resulting in a demand of d_v, and it remains to decide how the $k-1$ remaining facilities are located in T_u. As a result, $\mathcal{E}(v,k,p_{\mathrm{up}},1)$ is attained by taking the maximum of (2) and (3).

Termination Terminal states of our recursion involve the bottom endpoint r_{down}. In this case, when a facility is located at r_{down}, we obtain an expected demand coverage of $d_{r_{\mathrm{down}}}$. Otherwise, when one avoids locating a facility at r_{down}, the expected demand coverage is $p_{\mathrm{up}}\cdot d_{r_{\mathrm{down}}}$. Consequently, we have $\mathcal{E}(r_{\mathrm{down}},k,p_{\mathrm{up}},1) = d_{r_{\mathrm{down}}}$ for $k \ge 1$ and $\mathcal{E}(r_{\mathrm{down}},k,p_{\mathrm{up}},0) = p_{\mathrm{up}}\cdot d_{r_{\mathrm{down}}}$ for $k \ge 0$. All other states of the form $\mathcal{E}(r_{\mathrm{down}},\cdot,\cdot,\cdot)$ are impossible.

Summary As a preprocessing step, the sets $\mathcal{P}^v_{\mathrm{up}}$ and $\mathcal{P}^v_{\mathrm{down}}$ over all nodes can be constructed in time $O(n^2)$. In addition, based on the preceding discussion, the parameters p_{up} and p_{down} take $O(n)$ values each, while the current root and the remaining number of facilities take $O(n)$ and $O(k)$ values, respectively. Therefore, our dynamic program consists of $O(n^3k)$ states, each evaluated in $O(n)$ time, thus leading to the next theorem.

Theorem 3.1 *On bidirected lines, the maximum reliability coverage problem can be solved to optimality in* $O(n^4k)$*time.*

3.2 Exact Dynamic Program on Trees: Continuous State Space

Preliminaries We begin by rooting the tree T at an arbitrary node r. Unlike the special case of line networks, simple examples demonstrate that there are bidirected trees on n nodes where the probability of a given node to be covered takes $\Omega(2^{\Omega(n)})$ possible values. For instance, consider a complete binary tree on n leaves, where all down-going arcs have zero survival probability and all up-going arcs survive with probability 1, other than those adjacent to leaves, whose survival probabilities are $1 - \frac{1}{q_1}, \dots, 1 - \frac{1}{q_n}$. Here, $q_1 < \cdots < q_n$ is the sequence of n smallest prime numbers, noting that the binary representation of these probabilities is logarithmic in n, as $q_n \sim n \log n$ by the Prime Numbers Theorem [10, 21]. Clearly, picking different subsets of leaves as facilities results in different coverage probabilities for the root r. In particular, for $k = \lfloor \frac{n}{2} \rfloor$ facilities, the number of such probabilities is $\binom{n}{\lfloor n/2 \rfloor} = \Omega(\frac{2^n}{\sqrt{n}})$, where the latter transition follows from Stirling's approximation. Motivated by this observation, we operate for the time being while allowing a continuum of coverage probabilities, taking any value in [0, 1], which leads to a characterization of optimal facility sets by means of continuous dynamic programming. As previously mentioned, the algorithmic implications of this characterization are discussed in Sect. 3.3.

The dynamic program For every node $v \in V$, number of facilities k, and pair of probabilities $(p_{\text{up}}, p_{\text{down}}) \in [0,1]^2$, we define the function value $\mathcal{E}(v, k, p_{\text{up}}, p_{\text{down}})$ as the maximum expected demand coverage collected over all nodes in the subtree T_v rooted at v, given that the coverage probability of v by a facility located in $T \setminus T_v$ is p_{up}, under the following conditions: (1) The number of facilities located in T_v is at most k; and (2) The coverage probability of v by a facility located in T_v is p_{down}. With these definitions, the optimal expected demand coverage corresponds to $\max\{\mathcal{E}(r, k, 0, p_{\text{down}}) : p_{\text{down}} \in [0,1]\}$.

Recursive equation for $\mathcal{E}(v, k, p_{\text{up}}, p_{\text{down}})$ Let u_ℓ and u_r be the left and right children of v, respectively. When $k \geq 1$, the first option is to locate a facility at v, and to obtain an expected demand coverage of

$$\begin{aligned} & d_v + \max \left\{ \mathcal{E}(u_\ell, k_\ell, p_{(v,u_\ell)}, p_{\text{down},\ell}) + \mathcal{E}(u_r, k_r, p_{(v,u_r)}, p_{\text{down},r}) \right\} \\ & \text{such that: } (C_1^{(4)}) \; k_\ell + k_r = k - 1 \\ & \qquad\qquad\quad (C_2^{(4)}) \; p_{\text{down},\ell}, p_{\text{down},r} \in [0,1] \\ & \text{variables: } k_\ell, k_r, p_{\text{down},\ell}, p_{\text{down},r} \end{aligned} \tag{4}$$

Here, the node v is covered with probability 1, resulting in a demand of d_v, and it remains to decide how the $k-1$ remaining facilities are divided between the left and right subtrees, T_{u_ℓ} and T_{u_r}, using the decision variables k_ℓ and k_r. Clearly, this option is relevant only when $p_{\text{down}} = 1$. As a side note, due to having recursive equations with multiple constraints, each will be denoted by $(C_\cdot^{(\cdot)})$, where superscripts refer to the corresponding equation number and subscripts refer to the internal indexing between constraints.

The second option is to avoid locating a facility at v, which leads to an expected demand coverage of

$$\begin{aligned}
&\max \Big\{(1-(1-p_{\text{up}})\cdot(1-p_{\text{down},\ell}\cdot p_{(u_\ell,v)})\cdot(1-p_{\text{down},r}\cdot p_{(u_r,v)}))\cdot d_v \\
&\qquad + \mathcal{E}(u_\ell,k_\ell,p_{\text{up},\ell},p_{\text{down},\ell})+\mathcal{E}(u_r,k_r,p_{\text{up},r},p_{\text{down},r})\Big\} \\
&\text{such that: } (C_1^{(5)})\ k_\ell+k_r=k \\
&\qquad (C_2^{(5)})\ p_{\text{up},\ell}=p_{(v,u_\ell)}\cdot(1-(1-p_{\text{up}})\cdot(1-p_{(u_r,v)}\cdot p_{\text{down},r})) \\
&\qquad (C_3^{(5)})\ p_{\text{up},r}=p_{(v,u_r)}\cdot(1-(1-p_{\text{up}})\cdot(1-p_{(u_\ell,v)}\cdot p_{\text{down},\ell})) \\
&\qquad (C_4^{(5)})\ p_{\text{down}}=1-(1-p_{\text{down},\ell}\cdot p_{(u_\ell,v)})\cdot(1-p_{\text{down},r}\cdot p_{(u_r,v)}) \\
&\qquad (C_5^{(5)})\ p_{\text{up},\ell},p_{\text{up},r},p_{\text{down},\ell},p_{\text{down},r}\in[0,1] \\
&\text{variables: } k_\ell,k_r,p_{\text{up},\ell},p_{\text{up},r},p_{\text{down},\ell},p_{\text{down},r}
\end{aligned} \tag{5}$$

In this case, v is covered with probability $1-(1-p_{\text{up}})\cdot(1-p_{\text{down},\ell}\cdot p_{(u_\ell,v)})\cdot(1-p_{\text{down},r}\cdot p_{(u_r,v)})$, and it remains to decide how the k facilities are divided between the left and right subtrees. As a result, $\mathcal{E}(v,k,p_{\text{up}},p_{\text{down}})$ is attained by taking the maximum of (4) and (5). When $k=0$, the first option mentioned above is not possible, in which case $\mathcal{E}(v,0,p_{\text{up}},p_{\text{down}})$ is given by (5).

Termination Terminal states of our recursion involve leaves of the tree. Similarly to bidirected paths, when a facility is located at a leaf v, we obtain an expected demand coverage of d_v. Without a facility at v, the expected demand coverage is $p_{\text{up}}\cdot d_v$. Consequently, we have $\mathcal{E}(v,k,p_{\text{up}},1)=d_v$ for $k\geq 1$ and $\mathcal{E}(v,k,p_{\text{up}},0)=p_{\text{up}}\cdot d_v$ for $k\geq 0$. All other states of the form $\mathcal{E}(v,\cdot,\cdot,\cdot)$ are impossible whenever v is a leaf.

3.3 Approximate Dynamic Program on Trees: Discretized State Space

Unfortunately, the dynamic program proposed in Sect. 3.2 makes use of a continuous state space, due to allowing the probabilities p_{up} and p_{down} as well as the decision variables $p_{\text{up},\ell}$, $p_{\text{up},r}$, $p_{\text{down},\ell}$, and $p_{\text{down},r}$ to take any value in [0, 1]. In what follows, our objective is to discretize the latter state space in order to obtain a dynamic program that can be solved in polynomial time. For this purpose, note that for any path without arcs that fail with probability 1, its survival probability is trivially lower-bounded by $p_{\min}^{n-1}$, where $p_{\min}$ stands for the minimum non-zero survival probability, i.e., $p_{\min}=\min\{p_e : p_e>0, e\in E\}$. Therefore, any non-zero probability that would be incurred during the overall computation resides within $[p_{\min}^{n-1},1]$. We proceed by showing how to discretize this interval such that, as formally explained later on, when the true probabilities are "rounded" accordingly, the overall error accumulated throughout the recursive calls would be provably negligible.

Construction of $\mathcal{P}$ Given an error parameter $\epsilon\in(0,\frac{1}{3})$, the discretized set of "probabilities" $\mathcal{P}$ consists all (non-positive) integer powers of $\mu=1+\frac{\epsilon}{n^2}$ within the interval $[\epsilon p^n_{\min}, 1]$; we also add 0 to this set. It is worth emphasizing that the left endpoint $\epsilon p^n_{\min}$ of the latter interval was purposely chosen to be smaller than the previously-mentioned lower bound of $p^{n-1}_{\min}$ on any non-zero probability incurred in Sect. 3.2. As we explain later on, this feature will be important to analyze the performance guarantee of our algorithm. Note that $|\mathcal{P}| = O(\log_\mu \frac{1}{\epsilon p^n_{\min}}) = O(\frac{n^2}{\epsilon} \cdot (\log \frac{1}{\epsilon} + n \log \frac{1}{p_{\min}}))$, which is polynomial in the input size and in $\frac{1}{\epsilon}$.

Approximate dynamic program For every node $v \in V$, number of facilities k, and pair of probabilities $(p_{\text{up}}, p_{\text{down}}) \in \mathcal{P}^2$, we explain how to compute $\tilde{\mathcal{E}}(v, k, p_{\text{up}}, p_{\text{down}})$, which constitutes a lower bound on $\mathcal{E}(v, k, p_{\text{up}}, p_{\text{down}})$. When $k \geq 1$, the former function is defined by taking the maximum of (6) and (7) below. The first of these subproblems is a restricted version of (4), with $p_{\text{up},\cdot}$ and $p_{\text{down},\cdot}$ variables for the left and right subtrees, forced to reside in $\mathcal{P}$:

$$\begin{aligned} & d_v + \max \left\{ \tilde{\mathcal{E}}(u_\ell, k_\ell, p_{\text{up},\ell}, p_{\text{down},\ell}) + \tilde{\mathcal{E}}(u_r, k_r, p_{\text{up},r}, p_{\text{down},r}) \right\} \\ & \text{such that: } (C^{(6)}_1)\ k_\ell + k_r = k - 1 \\ & \qquad (C^{(6)}_2)\ p_{\text{up},\ell} \leq p_{(v,u_\ell)},\ p_{\text{up},r} \leq p_{(v,u_r)} \\ & \qquad (C^{(6)}_3)\ p_{\text{up},\ell}, p_{\text{up},r}, p_{\text{down},\ell}, p_{\text{down},r} \in \mathcal{P} \\ & \text{variables: } k_\ell, k_r, p_{\text{up},\ell}, p_{\text{up},r}, p_{\text{down},\ell}, p_{\text{down},r} \end{aligned} \tag{6}$$

It is easy to verify that, when condition $(C^{(6)}_3)$ is replaced by $p_{\text{up},\ell}, p_{\text{up},r}, p_{\text{down},\ell}, p_{\text{down},r} \in [0, 1]$, an optimal solution would necessarily set $p_{\text{up},\ell} = p_{(v,u_\ell)}$ and $p_{\text{up},r} = p_{(v,u_r)}$, leading back to subproblem (4). Similarly, the second subproblem is a restricted version of (5), defined by:

$$\begin{aligned} & \max \Big\{ (1 - (1 - p_{\text{up}}) \cdot (1 - p_{\text{down},\ell} \cdot p_{(u_\ell,v)}) \cdot (1 - p_{\text{down},r} \cdot p_{(u_r,v)})) \cdot d_v \\ & \qquad + \tilde{\mathcal{E}}(u_\ell, k_\ell, p_{\text{up},\ell}, p_{\text{down},\ell}) + \tilde{\mathcal{E}}(u_r, k_r, p_{\text{up},r}, p_{\text{down},r}) \Big\} \\ & \text{such that: } (C^{(7)}_1)\ k_\ell + k_r = k \\ & \qquad (C^{(7)}_2)\ p_{\text{up},\ell} \leq p_{(v,u_\ell)} \cdot (1 - (1 - p_{\text{up}}) \cdot (1 - p_{(u_r,v)} \cdot p_{\text{down},r})) \\ & \qquad (C^{(7)}_3)\ p_{\text{up},r} \leq p_{(v,u_r)} \cdot (1 - (1 - p_{\text{up}}) \cdot (1 - p_{(u_\ell,v)} \cdot p_{\text{down},\ell})) \\ & \qquad (C^{(7)}_4)\ p_{\text{down}} \leq 1 - (1 - p_{\text{down},\ell} \cdot p_{(u_\ell,v)}) \cdot (1 - p_{\text{down},r} \cdot p_{(u_r,v)}) \\ & \qquad (C^{(7)}_5)\ p_{\text{up},\ell}, p_{\text{up},r}, p_{\text{down},\ell}, p_{\text{down},r} \in \mathcal{P} \\ & \text{variables: } k_\ell, k_r, p_{\text{up},\ell}, p_{\text{up},r}, p_{\text{down},\ell}, p_{\text{down},r} \end{aligned} \tag{7}$$

Here, the only differences are substituting [0, 1] by $\mathcal{P}$ in condition $(C^{(5)}_5)$ and using inequalities in conditions $(C^{(5)}_2)$–$(C^{(5)}_4)$ rather than equalities. When $k = 0$, the first option mentioned above is not possible, in which case $\tilde{\mathcal{E}}(v, 0, p_{\text{up}}, p_{\text{down}})$ is given by (7). Moreover, terminal states for the approximate value function $\tilde{\mathcal{E}}$ are treated identically to Sect. 3.2. Namely, when v is a leaf, $\tilde{\mathcal{E}}(v, k, p_{\text{up}}, 1) = d_v$ for $k \geq 1$ and $\tilde{\mathcal{E}}(v, k, p_{\text{up}}, 0) = p_{\text{up}} \cdot d_v$ for $k \geq 0$.

Intermediate summary Based on the preceding discussion, the resulting dynamic program consists of $O(nk \cdot |\mathcal{P}|^2)$ states, whereas the time to evaluate $\tilde{\mathcal{E}}$ for each state is $O(k \cdot |\mathcal{P}|^4)$. Since $|\mathcal{P}| = O(\frac{n^2}{\epsilon} \cdot (\log \frac{1}{\epsilon} + n \log \frac{1}{p_{\min}}))$, the overall running time is

$$O\left(nk^2 \cdot |\mathcal{P}|^6\right) = O\left(\frac{n^{19}k^2}{\epsilon^6} \cdot \log^6\left(\frac{1}{\epsilon p_{\min}}\right)\right),$$

which is polynomial in the input size and in $\frac{1}{\epsilon}$. Consequently, we can construct in polynomial time a feasible set of facilities with an expected demand coverage of at least $\max\{\tilde{\mathcal{E}}(r, k, 0, p_{\text{down}}) : p_{\text{down}} \in \mathcal{P}\}$. However, the interesting question is: Since the dynamic program $\tilde{\mathcal{E}}$ is restricted to using "probabilities" from the discrete set $\mathcal{P}$, while the exact continuous program operates without this restriction, why are we guaranteed that $\max\{\tilde{\mathcal{E}}(r, k, 0, p_{\text{down}}) : p_{\text{down}} \in \mathcal{P}\}$ nearly matches the optimal expected demand coverage $\max\{\mathcal{E}(r, k, 0, p_{\text{down}}) : p_{\text{down}} \in [0, 1]\}$? The remainder of this section is devoted to proving the next theorem.

Theorem 3.2 $\max\{\tilde{\mathcal{E}}(r, k, 0, p_{\text{down}}) : p_{\text{down}} \in \mathcal{P}\} \geq (1 - 52\epsilon) \cdot \max\{\mathcal{E}(r, k, 0, p_{\text{down}}) : p_{\text{down}} \in [0, 1]\}$.

Notation For a real number $x \in [p^n_{\min}, 1]$ and an integer $t \leq 2n$, we use $\lfloor x \rfloor_t$ to denote the value obtained by rounding $\frac{x}{\mu^t}$ down to the nearest (non-zero) number in $\mathcal{P}$. This operator is indeed well-defined, since $\mathcal{P}$ was constructed with respect to the interval $[\epsilon p^n_{\min}, 1]$, and since

$$\frac{p^n_{\min}}{\mu^{2n}} = \frac{p^n_{\min}}{(1 + \epsilon/n^2)^{2n}} \geq p^n_{\min} \cdot e^{-2\epsilon/n} \geq p^n_{\min} \cdot \left(1 - \frac{2\epsilon}{n}\right) \geq \epsilon p^n_{\min},$$

where the last inequality holds for any $\epsilon \leq 1/3$.

Defining the solution Let F^* be an optimal set of facilities, with an expected demand coverage of $\max\{\mathcal{E}(r, k, 0, p_{\text{down}}) : p_{\text{down}} \in [0, 1]\}$. For each node $v \in V$, let $p^*_{\text{down},v}$ be the probability that v is covered by a facility in $F^* \cap T_v$, and let $p^*_{\text{up},v}$ be the probability that v is covered by a facility in $F^* \cap (T \setminus T_v)$. Finally, let $L(v)$ be the level of v in the tree T, i.e., the root r is at level 0, its two children are at level 1, so on and so forth.

We create a candidate solution to the dynamic program $\tilde{\mathcal{E}}$ as follows. First, facilities are located at F^*. Given this decision, our goal is to define "rounded" probabilities in $\mathcal{P}$ that approximate the true ones, $\tilde{p}_{\text{down},v} \approx p^*_{\text{down},v}$ and $\tilde{p}_{\text{up},v} \approx p^*_{\text{up},v}$, in order to ensure two basic properties:

- *Feasibility*: $(F^*, \tilde{p})$ is a feasible solution to $\tilde{\mathcal{E}}$.
- *Near-optimality*: The expected demand coverage of $(F^*, \tilde{p})$ is within factor $1 - 52\epsilon$ of $\max\{\mathcal{E}(r, k, 0, p_{\text{down}}) : p_{\text{down}} \in [0, 1]\}$.

As it turns out later on, the difficult conditions to satisfy are $(C_2^{(7)})$–$(C_4^{(7)})$. A close inspection of condition $(C_4^{(7)})$ reveals that rounding errors would accumulate through the dynamic program $\tilde{\mathcal{E}}$ in a bottom-up way, meaning that our approximation $\tilde{p}_{\text{down},v}$ for $p^*_{\text{down},v}$ should be tighter as the distance from v to the root grows. Therefore, we will make use of $\tilde{p}_{\text{down},v} = \lfloor p^*_{\text{down},v} \rfloor_{n-L(v)}$. Having fixed this decision, by analyzing how the rounding error accumulates with respect to conditions $(C_2^{(7)})$ and $(C_3^{(7)})$, we would have the opposite trend, and will consequently use $\tilde{p}_{\text{up},v} = \lfloor p^*_{\text{up},v} \rfloor_{n+L(v)}$. Note that $n + L(v) \le 2n - 1$ for every $v \in V$, and therefore $\tilde{p}_{\text{down},v}$ and $\tilde{p}_{\text{up},v}$ are well-defined.

Analysis As shown below, these definitions are sufficient to establish feasibility, regardless of the value of μ, whose powers within $[\epsilon p^n_{\min}, 1]$ were used to define $\mathcal{P}$. In fact, any choice of the parameter $\mu > 1$ works in this context. However, for the expected demand coverage of $(F^*, \tilde{p})$ to be near-optimal, this parameter needs to be accurate enough, and our choice of $\mu = 1 + \frac{\epsilon}{n^2}$ will be shown to be sufficient. With respect to subproblems (6) and (7), the conditions $(C_1^{(6)})$, $(C_1^{(7)})$, $(C_2^{(6)})$, $(C_3^{(6)})$, and $(C_5^{(7)})$ are clearly satisfied. The more challenging conditions to prove are $(C_2^{(7)})$–$(C_4^{(7)})$, in addition to deriving a lower bound on the objective function.

Lemma 3.3 *The solution* $(F^*, \tilde{p})$ *satisfies condition* $(C_4^{(7)})$.

Proof Let us focus on some node $v \notin F^*$. Then,

$$
\begin{aligned}
1 &- (1 - \tilde{p}_{\text{down},\ell} \cdot p_{(u_\ell,v)}) \cdot (1 - \tilde{p}_{\text{down},r} \cdot p_{(u_r,v)}) \\
&= \tilde{p}_{\text{down},\ell} \cdot p_{(u_\ell,v)} + \tilde{p}_{\text{down},r} \cdot p_{(u_r,v)} - \tilde{p}_{\text{down},\ell} \cdot p_{(u_\ell,v)} \cdot \tilde{p}_{\text{down},r} \cdot p_{(u_r,v)} \\
&= \lfloor p^*_{\text{down},\ell} \rfloor_{n-L(u_\ell)} \cdot p_{(u_\ell,v)} + \lfloor p^*_{\text{down},r} \rfloor_{n-L(u_r)} \cdot p_{(u_r,v)} \\
&\quad - \lfloor p^*_{\text{down},\ell} \rfloor_{n-L(u_\ell)} \cdot p_{(u_\ell,v)} \cdot \lfloor p^*_{\text{down},r} \rfloor_{n-L(u_r)} \cdot p_{(u_r,v)} \\
&\ge \frac{p^*_{\text{down},\ell}}{\mu^{n-L(u_\ell)+1}} \cdot p_{(u_\ell,v)} + \frac{p^*_{\text{down},r}}{\mu^{n-L(u_r)+1}} \cdot p_{(u_r,v)} \\
&\quad - \frac{p^*_{\text{down},\ell}}{\mu^{n-L(u_\ell)}} \cdot p_{(u_\ell,v)} \cdot \frac{p^*_{\text{down},r}}{\mu^{n-L(u_r)}} \cdot p_{(u_r,v)} \\
&\ge \frac{1}{\mu^{n-L(u_\ell)+1}} \cdot \Big(p^*_{\text{down},\ell} \cdot p_{(u_\ell,v)} + p^*_{\text{down},r} \cdot p_{(u_r,v)} \\
&\quad - p^*_{\text{down},\ell} \cdot p_{(u_\ell,v)} \cdot p^*_{\text{down},r} \cdot p_{(u_r,v)} \Big) \\
&= \frac{p^*_{\text{down},v}}{\mu^{n-L(v)}} \\
&\ge \lfloor p^*_{\text{down},v} \rfloor_{n-L(v)} \\
&= \tilde{p}_{\text{down},v}.
\end{aligned}
$$

□

Lemma 3.4 *The solution $(F^*, \tilde{p})$ satisfies conditions $(C_2^{(7)})$ and $(C_3^{(7)})$.*

Proof For the left subtree, condition $(C_2^{(7)})$ is met since:

$$\begin{aligned}
&p_{(v,u_\ell)} \cdot (1 - (1 - \tilde{p}_{\mathrm{up},v}) \cdot (1 - p_{(u_r,v)} \cdot \tilde{p}_{\mathrm{down},r}))\\
&\quad= p_{(v,u_\ell)} \cdot \left(\tilde{p}_{\mathrm{up},v} + p_{(u_r,v)} \cdot \tilde{p}_{\mathrm{down},r} - \tilde{p}_{\mathrm{up},v} \cdot p_{(u_r,v)} \cdot \tilde{p}_{\mathrm{down},r}\right)\\
&\quad= p_{(v,u_\ell)} \cdot \Big(\lfloor p^*_{\mathrm{up},v} \rfloor_{n+L(v)} + p_{(u_r,v)} \cdot \lfloor p^*_{\mathrm{down},r} \rfloor_{n-L(u_r)}\\
&\qquad - \lfloor p^*_{\mathrm{up},v} \rfloor_{n+L(v)} \cdot p_{(u_r,v)} \cdot \lfloor p^*_{\mathrm{down},r} \rfloor_{n-L(u_r)} \Big)\\
&\quad\geq p_{(v,u_\ell)} \cdot \left(\frac{p^*_{\mathrm{up},v}}{\mu^{n+L(v)+1}} + p_{(u_r,v)} \cdot \frac{p^*_{\mathrm{down},r}}{\mu^{n-L(u_r)+1}} - \frac{p^*_{\mathrm{up},v}}{\mu^{n+L(v)}} \cdot p_{(u_r,v)} \cdot \frac{p^*_{\mathrm{down},r}}{\mu^{n-L(u_r)}} \right)\\
&\quad\geq \frac{p_{(v,u_\ell)}}{\mu^{n+L(v)+1}} \cdot \left(p^*_{\mathrm{up},v} + p_{(u_r,v)} \cdot p^*_{\mathrm{down},r} - p^*_{\mathrm{up},v} \cdot p_{(u_r,v)} \cdot p^*_{\mathrm{down},r}\right)\\
&\quad= \frac{p^*_{\mathrm{up},\ell}}{\mu^{n+L(u_\ell)}}\\
&\quad\geq \lfloor p^*_{\mathrm{up},\ell} \rfloor_{n+L(u_\ell)}\\
&\quad= p^*_{\mathrm{up},\ell}.
\end{aligned}$$

The argument for the right subtree and condition $(C_3^{(7)})$ is symmetrical. □

Concluding the proof of Theorem 3.2 In order to establish the theorem, it suffices to argue that the solution $(F^*, \tilde{p})$, whose feasibility has just been proven, attains an objective value of at least $(1 - 52\epsilon) \cdot \max\{\mathcal{E}(r, k, 0, p_{\mathrm{down}}) : p_{\mathrm{down}} \in [0, 1]\}$ with respect to our approximate dynamic program $\tilde{\mathcal{E}}$. We begin by observing that the latter quantity can be written as

$$\begin{aligned}
\max\{\mathcal{E}(r, k, 0, p_{\mathrm{down}}) : p_{\mathrm{down}} \in [0, 1]\} &= \mathcal{E}(F^*) = \sum_{v \in V} \pi_v(F^*) \cdot d_v\\
&= \sum_{v \in F^*} d_v + \sum_{v \notin F^*} \pi_v(F^*) \cdot d_v.
\end{aligned} \tag{8}$$

Now, for every $v \in F^*$, the solution $(F^*, \tilde{p})$ covers v with probability 1. In addition, for every $v \notin F^*$, the probability that v is covered by $(F^*, \tilde{p})$ can be lower bounded in terms of $\pi_v(F^*)$ as follows:

$$\begin{aligned}
&1-(1-\tilde{p}_{\mathrm{up},v})\cdot(1-\tilde{p}_{\mathrm{down},\ell}\cdot p_{(u_\ell,v)})\cdot(1-\tilde{p}_{\mathrm{down},r}\cdot p_{(u_r,v)})\\
&\quad=1-(1-\lfloor p^*_{\mathrm{up},v}\rfloor_{n+L(v)})\cdot(1-\lfloor p^*_{\mathrm{down},\ell}\rfloor_{n-L(u_\ell)}\cdot p_{(u_\ell,v)})\\
&\qquad\cdot(1-\lfloor p^*_{\mathrm{down},r}\rfloor_{n-L(u_r)}\cdot p_{(u_r,v)})\\
&\quad\geq 1-\left(1-\frac{p^*_{\mathrm{up},v}}{\mu^{n+L(v)+1}}\right)\cdot\left(1-\frac{p^*_{\mathrm{down},\ell}}{\mu^{n-L(u_\ell)+1}}\cdot p_{(u_\ell,v)}\right)\\
&\qquad\cdot\left(1-\frac{p^*_{\mathrm{down},r}}{\mu^{n-L(u_r)+1}}\cdot p_{(u_r,v)}\right)\\
&\quad\geq\frac{1}{\mu^{3n+L(v)-L(u_\ell)-L(u_r)+3}}\cdot\Big(1-\Big(\mu^{n+L(v)+1}-p^*_{\mathrm{up},v}\Big)\\
&\qquad\cdot\Big(\mu^{n-L(u_\ell)+1}-p^*_{\mathrm{down},\ell}\cdot p_{(u_\ell,v)}\Big)\\
&\qquad\cdot\Big(\mu^{n-L(u_r)+1}-p^*_{\mathrm{down},r}\cdot p_{(u_r,v)}\Big)\Big)\\
&\quad\geq\frac{1}{\mu^{3n+1}}\cdot\Big(1-\Big(\mu^{2n}-p^*_{\mathrm{up},v}\Big)\cdot\Big(\mu^{n}-p^*_{\mathrm{down},\ell}\cdot p_{(u_\ell,v)}\Big)\cdot\Big(\mu^{n}-p^*_{\mathrm{down},r}\cdot p_{(u_r,v)}\Big)\Big)\\
&\quad\geq(1-8\epsilon)\cdot\Big(1-\Big(1+\frac{4\epsilon}{n}-p^*_{\mathrm{up},v}\Big)\cdot\Big(1+\frac{2\epsilon}{n}-p^*_{\mathrm{down},\ell}\cdot p_{(u_\ell,v)}\Big)\\
&\qquad\cdot\Big(1+\frac{2\epsilon}{n}-p^*_{\mathrm{down},r}\cdot p_{(u_r,v)}\Big)\Big)\\
&\quad\geq(1-8\epsilon)\cdot\Big(1-\Big(1-p^*_{\mathrm{up},v}\Big)\cdot\Big(1-p^*_{\mathrm{down},\ell}\cdot p_{(u_\ell,v)}\Big)\\
&\qquad\cdot\Big(1-p^*_{\mathrm{down},r}\cdot p_{(u_r,v)}\Big)-\frac{44\epsilon}{n}\Big)\\
&\quad=(1-8\epsilon)\cdot\Big(\pi_v(F^*)-\frac{44\epsilon}{n}\Big).
\end{aligned}$$

Here, the fourth inequality hold since $\mu^n=(1+\frac{\epsilon}{n^2})^n\leq e^{\epsilon/n}\leq 1+\frac{2\epsilon}{n}$, where in the last transition we are using $e^x\leq 1+2x$ for $x\in[0,1]$; similarly, $\mu^{2n}\leq 1+\frac{4\epsilon}{n}$ and $\mu^{3n+1}\leq 1+\frac{6\epsilon}{n}+\frac{2\epsilon}{n^2}\leq 1+8\epsilon$. The term $\frac{44\epsilon}{n}$ in the fifth inequality results from using $\frac{\epsilon}{n}$ as an upper bound on both $(\frac{\epsilon}{n})^2$ and $(\frac{\epsilon}{n})^3$ in the expansion of $(1+\frac{4\epsilon}{n}-p^*_{\mathrm{up},v})\cdot(1+\frac{2\epsilon}{n}-p^*_{\mathrm{down},\ell}\cdot p_{(u_\ell,v)})\cdot(1+\frac{2\epsilon}{n}-p^*_{\mathrm{down},r}\cdot p_{(u_r,v)})$, noting that all $p_{\cdot}$ and $p^*_{\cdot}$ parameters are non-negative. The last equality is obtained by observing that $\pi_v(F^*)=1-(1-p^*_{\mathrm{up},v})\cdot(1-p^*_{\mathrm{down},\ell}\cdot p_{(u_\ell,v)})\cdot(1-p^*_{\mathrm{down},r}\cdot p_{(u_r,v)})$ for every $v\notin F^*$.

To summarize, we can now derive a lower bound on the optimal value of our dynamic program, $\max\{\tilde{\mathcal{E}}(r,k,0,p_{\mathrm{down}}) : p_{\mathrm{down}}\in\mathcal{P}\}$, via the expected demand coverage attained by $(F^*,\tilde{p})$, implying that

$$\begin{aligned}
&\max\{\tilde{\mathcal{E}}(r,k,0,p_{\text{down}}) : p_{\text{down}} \in \mathcal{P}\} \\
&\quad \geq \sum_{v\in F^*} d_v + (1-8\epsilon)\cdot \sum_{v\notin F^*}\left(\pi_v(F^*) - \frac{44\epsilon}{n}\right)\cdot d_v \\
&\quad \geq (1-8\epsilon)\cdot\left(\sum_{v\in F^*} d_v + \sum_{v\notin F^*}\pi_v(F^*)\cdot d_v\right) - \frac{44\epsilon}{n}\cdot\sum_{v\notin F^*} d_v \\
&\quad \geq (1-52\epsilon)\cdot \max\{\mathcal{E}(r,k,0,p_{\text{down}}) : p_{\text{down}} \in [0,1]\}\ ,
\end{aligned}$$

where the last inequality results from combining (8) with the observation that $d_v \leq \mathcal{E}(F^*) = \max\{\mathcal{E}(r,k,0,p_{\text{down}}) : p_{\text{down}} \in [0,1]\}$ for every node $v \in V$.

Theorem 3.5 *On bidirected trees, the maximum reliability coverage problem admits an FPTAS. The running time of our algorithm is* $O(\frac{n^{19}k^2}{\epsilon^6}\cdot\log^6(\frac{1}{\epsilon p_{\min}}))$.

4 General Networks

In what follows, we present a sampling-based greedy algorithm for approximating the maximum reliability coverage problem on general networks. This result is formally stated in the next theorem, where n and m respectively designate the number of nodes and arcs of the input graph, while k stands for the number of facilities to be located.

Theorem 4.1 *For any* $\epsilon > 0$, *there is a Monte-Carlo algorithm that computes a* $(1-1/e-\epsilon)$*-approximation for the maximum reliability coverage problem with probability at least* 1/2. *The running time of this algorithm is* $O(\frac{(nk)^3\cdot(n+m)\cdot\log n}{\epsilon^2})$.

It is worth pointing out that, as mentioned in Sect. 1.1, one can easily show that the expected demand coverage function $\mathcal{E} : 2^V \to \mathbb{R}_+$ is monotone and submodular. For completeness, we establish this claim in "Properties of ε" section of "Appendix 1". Consequently, a natural approach for maximizing this function subject to the k-cardinality constraint on the allowed number of facilities is to utilize the standard greedy algorithm [16]. The resulting approximation ratio, $1-1/e$, would be best-possible since maximum reliability coverage generalizes the max k-cover problem, that cannot be approximated within a constant greater than $1-1/e$, unless P = NP [8]. Nevertheless, to implement the greedy algorithm in a straightforward way, one should be equipped with an oracle access to $\mathcal{E}$. Unfortunately, the existence of an exact polynomial-time algorithm to evaluate $\mathcal{E}$ implies, in particular, that we can efficiently compute the probability that two given vertices, u and v, in an undirected graph remain connected subject to random independent edge failures; this estimation problem is known to be #P-complete [3, 17, 20]. The latter reduction works by setting the demand of v to 1 and that of any other node to 0. It is easy to verify that $\mathcal{E}(\{u\})$ is precisely the probability that u and v remain connected.

4.1 Estimating the Function $\mathcal{E}$

To go around the above-mentioned difficulties, we propose a sampling-based estimator for the expected demand coverage function $\mathcal{E}$ up to a certain additive error, which will be shown to be good enough for our purposes. To this end, given an accuracy level $\epsilon > 0$, we begin by setting the value of two parameters, $\delta = \frac{\epsilon}{2nk}$ and $M = \lceil \frac{\ln(4kn^2)}{2\delta^2} \rceil$. For a set of facilities $F \subseteq V$ and a node $v \in V$, let $\tilde{\pi}_v(F)$ be an estimator for the coverage probability $\pi_v(F)$, defined as follows: Out of M independently-generated surviving networks, $\tilde{\pi}_v(F)$ is the (random) proportion of those where v is covered by at least one facility in F. By Hoeffding's inequality [12],

$$\Pr\left[\left|\tilde{\pi}_v(F) - \pi_v(F)\right| \geq \delta\right] \leq 2e^{-2M\delta^2}. \tag{9}$$

Now, by plugging-in $\tilde{\pi}_v(F)$ instead of $\pi_v(F)$ into the definition $\mathcal{E}(F)$, we obtain our estimator $\tilde{\mathcal{E}}(F) = \sum_{v \in V} \tilde{\pi}_v(F) \cdot d_v$. It is easy to verify that the latter quantity can be computed in $O(M \cdot (n+m))$ time. The next claim provides a concentration bound for $\tilde{\mathcal{E}}(F)$ with respect to additive deviations in terms of the optimal expected demand coverage.

Lemma 4.2 *Let F^* be an optimal set of facilities. Then, for any $F \subseteq V$,*

$$\Pr\left[\left|\tilde{\mathcal{E}}(F) - \mathcal{E}(F)\right| \leq \delta n \cdot \mathcal{E}(F^*)\right] \geq 1 - 2ne^{-2M\delta^2}.$$

Proof By the union bound and inequality (9), with probability at least $1 - 2ne^{-2M\delta^2}$ we have $|\tilde{\pi}_v(F) - \pi_v(F)| \leq \delta$, simultaneously for all nodes $v \in V$. We now show that, given this event, $|\tilde{\mathcal{E}}(F) - \mathcal{E}(F)| \leq \delta n \cdot \mathcal{E}(F^*)$. For this purpose, note that

$$\begin{aligned}
\tilde{\mathcal{E}}(F) &= \sum_{v \in V} \tilde{\pi}_v(F) \cdot d_v \\
&\geq \sum_{v \in V} \left(\pi_v(F) - \delta\right) \cdot d_v \\
&= \mathcal{E}(F) - \delta \cdot \sum_{v \in V} d_v \\
&\geq \mathcal{E}(F) - \delta n \cdot \mathcal{E}(F^*) ,
\end{aligned}$$

where the last inequality holds since $\mathcal{E}(F^*) \geq d_v$ for every $v \in V$. The other direction, $\tilde{\mathcal{E}}(F) \leq \mathcal{E}(F) + \delta n \cdot \mathcal{E}(F^*)$, can be derived by nearly identical arguments. □

4.2 The Greedy Algorithm

In order to derive Theorem 4.1, we proceed by showing that the additive estimator $\tilde{\mathcal{E}}$ can be employed as an approximate oracle for $\mathcal{E}$ within the standard greedy algorithm. Specifically, starting with an empty set of facilities, this algorithm picks in each step a facility whose addition maximizes the estimated coverage function over all unpicked facilities:

- Initialize $F_0 = \emptyset$.
- For $\kappa \leftarrow 1$ to k:
  - For every $v \in V \setminus F_{\kappa-1}$, let $\tilde{\mathcal{E}}_{\kappa-1,v}$ be a realization of $\tilde{\mathcal{E}}(F_{\kappa-1} \cup \{v\})$.
  - Set $F_\kappa = F_{\kappa-1} \cup \{f_\kappa\}$, where $f_\kappa = \mathrm{argmax}_{v \in V \setminus F_{\kappa-1}} \tilde{\mathcal{E}}_{\kappa-1,v}$.
- Return F_k.

It is worth noting that, due to using a randomized oracle, the sequence of sets $F_0, \dots, F_k$ is obviously random as well. In addition, as far as running time is concerned, each of the estimators $\tilde{\mathcal{E}}_{\kappa-1,v}$ is computed in $O(M \cdot (n+m))$ time. For this reason, the greedy algorithm requires an overall running time of $O(Mnk \cdot (n+m)) = O(\frac{(nk)^3 \cdot (n+m) \cdot \log n}{\epsilon^2})$, by recalling that $M = \lceil \frac{\ln(4kn^2)}{2\delta^2} \rceil$ and $\delta = \frac{\epsilon}{2nk}$.

4.3 Analysis

To prove that, with constant probability, the resulting set of facilities F_k guarantees an expected demand coverage of at least $(1 - 1/e - \epsilon) \cdot \mathcal{E}(F^*)$, our arguments follow the classic analysis of Nemhauser, Wolsey, and Fisher [16] for maximizing a monotone submodular function subject to a cardinality constraint. However, suitable adaptations are required to account for using the randomized oracle $\tilde{\mathcal{E}}$ rather than $\mathcal{E}$.

Lemma 4.3 *Let $F \subseteq V$ be a set of facilities with $F^* \setminus F \neq \emptyset$. Then, there exists a node $v \in F^* \setminus F$* satisfying

$$\mathcal{E}(F \cup \{v\}) - \mathcal{E}(F) \geq \frac{\mathcal{E}(F^*) - \mathcal{E}(F)}{|F^* \setminus F|}.$$

Proof Letting $F^* \setminus F = \{v_1, \dots, v_t\}$, we have

$$\begin{aligned}
\mathcal{E}(F^*) - \mathcal{E}(F) &\leq \mathcal{E}(F^* \cup F) - \mathcal{E}(F) \\
&= \sum_{\tau=1}^{t} \left(\mathcal{E}(F \cup \{v_1, \dots, v_\tau\}) - \mathcal{E}(F \cup \{v_1, \dots, v_{\tau-1}\}) \right) \\
&\leq \sum_{\tau=1}^{t} \left(\mathcal{E}(F \cup \{v_\tau\}) - \mathcal{E}(F) \right),
\end{aligned}$$

where the first and second inequalities follow from the monotonicity and submodularity of $\mathcal{E}$, respectively. Therefore,

$$\max_{1 \leq \tau \leq t} \left(\mathcal{E}(F \cup \{v_\tau\}) - \mathcal{E}(F) \right) \geq \frac{1}{t} \cdot \sum_{\tau=1}^{t} \left(\mathcal{E}(F \cup \{v_\tau\}) - \mathcal{E}(F) \right) \geq \frac{\mathcal{E}(F^*) - \mathcal{E}(F)}{|F^* \setminus F|}.$$

□

Lemma 4.4 $\mathcal{E}(F_k) \geq (1 - \frac{1}{e} - \epsilon) \cdot \mathcal{E}(F^*)$ *with probability at least* 1/2.

Proof By the union bound and Lemma 4.2, the estimate $\tilde{\mathcal{E}}_{\kappa-1,v}$ resides within the interval $\mathcal{E}(F_{\kappa-1} \cup \{v\}) \pm \delta n \cdot \mathcal{E}(F^*)$, simultaneously for all steps $1 \leq \kappa \leq k$ and nodes $v \in V \setminus F_{\kappa-1}$, with probability at least $1 - 2kn^2 e^{-2M\delta^2} \geq 1/2$, where the last inequality holds since $M = \lceil \frac{\ln(4kn^2)}{2\delta^2} \rceil$. Given this event, we prove the desired claim by showing that, for every $0 \leq \kappa \leq k$,

$$\mathcal{E}(F_\kappa) \geq \left(1 - \left(1 - \frac{1}{k}\right)^\kappa - \frac{\kappa\epsilon}{k}\right) \cdot \mathcal{E}(F^*).$$

The proof works by induction on κ. The base case $\kappa = 0$ is trivial: Since $F_0 = \emptyset$, we have $\mathcal{E}(F_0) = 0$, and the right-hand-side of the above inequality evaluates to 0 as well. For the general case of $\kappa \geq 1$, we have

$$\begin{aligned}
\mathcal{E}(F_\kappa) &= \mathcal{E}(F_{\kappa-1} \cup \{f_\kappa\}) \\
&\geq \tilde{\mathcal{E}}_{\kappa-1,f_\kappa} - \delta n \cdot \mathcal{E}(F^*) \\
&= \max_{v \in V \setminus F_{\kappa-1}} \tilde{\mathcal{E}}_{\kappa-1,v} - \delta n \cdot \mathcal{E}(F^*) \\
&\geq \max_{v \in V \setminus F_{\kappa-1}} \mathcal{E}(F_{\kappa-1} \cup \{v\}) - 2\delta n \cdot \mathcal{E}(F^*) \\
&\geq \mathcal{E}(F_{\kappa-1}) + \frac{\mathcal{E}(F^*) - \mathcal{E}(F_{\kappa-1})}{|F^* \setminus F_{\kappa-1}|} - 2\delta n \cdot \mathcal{E}(F^*)\,,
\end{aligned}$$

where the third inequality follows from Lemma 4.3, instantiated with $F = F_{\kappa-1}$. Therefore, by the induction hypothesis,

$$\begin{aligned}
\mathcal{E}(F_\kappa) \geq &\left(\left(1 - \frac{1}{|F^* \setminus F_{\kappa-1}|}\right) \cdot \left(1 - \left(1 - \frac{1}{k}\right)^{\kappa-1} - \frac{(\kappa-1)\cdot\epsilon}{k}\right)\right. \\
&\left. + \frac{1}{|F^* \setminus F_{\kappa-1}|} - 2\delta n\right) \cdot \mathcal{E}(F^*) \\
= &\left(1 - \left(1 - \frac{1}{k}\right)^{\kappa-1} - \frac{(\kappa-1)\cdot\epsilon}{k}\right. \\
&\left. + \underbrace{\frac{1}{|F^* \setminus F_{\kappa-1}|} \cdot \left(\left(1 - \frac{1}{k}\right)^{\kappa-1} + \frac{(\kappa-1)\cdot\epsilon}{k}\right)}_{\geq 0} - 2\delta n\right) \cdot \mathcal{E}(F^*) \\
\geq &\left(1 - \left(1 - \frac{1}{k}\right)^{\kappa} - \frac{(\kappa-1)\cdot\epsilon}{k} - 2\delta n\right) \cdot \mathcal{E}(F^*) \\
= &\left(1 - \left(1 - \frac{1}{k}\right)^{\kappa} - \frac{\kappa\epsilon}{k}\right) \cdot \mathcal{E}(F^*)\,,
\end{aligned}$$

where the last equality is obtained by substituting $\delta = \frac{\epsilon}{2nk}$. □

Appendix 1: Additional Proofs

Proof of Claim 2.6

To obtain the desired inequality, note that

$$\begin{aligned}
\prod_{i\in I}\left(e^{-a_i/A}+\frac{1}{K}\right) &= e^{-\sum_{i\in I}a_i/A}+\sum_{S\subsetneq I}\frac{e^{-\sum_{i\in S}a_i/A}}{K^{|I|-|S|}}\\
&\le e^{-\sum_{i\in I}a_i/A}+\sum_{s=0}^{|I|-1}\binom{|I|}{s}\cdot\frac{1}{K^{|I|-s}}\\
&= e^{-\sum_{i\in I}a_i/A}+\left(1+\frac{1}{K}\right)^{|I|}-1\\
&\le e^{-\sum_{i\in I}a_i/A}+e^{|I|/K}-1\\
&\le e^{-\sum_{i\in I}a_i/A}+\frac{2\cdot|I|}{K}\\
&\le e^{-\sum_{i\in I}a_i/A}+\frac{1}{100A^2}.
\end{aligned}$$

Here, the first inequality is obtained by observing that $e^{-\sum_{i\in S}a_i/A}\le 1$. In the third inequality we are using $e^x\le 1+2x$ for $x\in[0,1]$. Finally, the last inequality holds since $|I|\le n$ and $K=200nA^2$.

Proof of Claim 2.8

The proof is similar to that of Claim 2.6, and we provide it for completeness. To this end, note that

$$\begin{aligned}
\prod_{i\in I}\left(e^{-a_i/A}-\frac{1}{K}\right) &= e^{-\sum_{i\in I}a_i/A}+\sum_{S\subsetneq I}\cdot\frac{e^{-\sum_{i\in S}a_i/A}}{(-K)^{|I|-|S|}}\\
&\ge e^{-\sum_{i\in I}a_i/A}-\sum_{S\subsetneq I}\frac{1}{K^{|I|-|S|}}\\
&= e^{-\sum_{i\in I}a_i/A}-\sum_{s=0}^{|I|-1}\binom{|I|}{s}\cdot\frac{1}{K^{|I|-s}}\\
&= e^{-\sum_{i\in I}a_i/A}-\left(1+\frac{1}{K}\right)^{|I|}+1\\
&\ge e^{-\sum_{i\in I}a_i/A}-e^{|I|/K}+1\\
&\ge e^{-\sum_{i\in I}a_i/A}-\frac{2\cdot|I|}{K}\\
&\ge e^{-\sum_{i\in I}a_i/A}-\frac{1}{100A^2}.
\end{aligned}$$

Properties of $\mathcal{E}$

Lemma 5.1 *The expected demand coverage function $\mathcal{E} : 2^V \to \mathbb{R}_+$ is monotone and submodular.*

Proof We begin by observing that it suffices to prove that each of the functions $\{\pi_v\}_{v \in V}$ is monotone and submodular, since $\mathcal{E}(F) = \sum_{v \in V} d_v \cdot \pi_v(F)$ is a non-negative weighted sum of these functions. To this end, recall that $\pi_v(F)$ stands for the probability that node v is covered by at least one facility in F. Put differently, letting $\mathcal{D}_{F \rightsquigarrow v}$ be the event where a least one of the directed paths connecting a facility in F to the node v survives, we have $\pi_v(F) = \Pr[\mathcal{D}_{F \rightsquigarrow v}]$. With this representation, we derive the desired properties as follows:

- Monotonicity of π_v: For two subsets of facilities $F_1 \subseteq F_2$, since $\mathcal{D}_{F_1 \rightsquigarrow v} \subseteq \mathcal{D}_{F_2 \rightsquigarrow v}$,

$$\pi_v(F_1) = \Pr\left[\mathcal{D}_{F_1 \rightsquigarrow v}\right] \leq \Pr\left[\mathcal{D}_{F_2 \rightsquigarrow v}\right] = \pi_v(F_2).$$

- Submodularity of π_v: For two subsets of facilities F_1 and F_2,

$$\begin{aligned} \pi_v(F_1 \cup F_2) &= \Pr\left[\mathcal{D}_{F_1 \cup F_2 \rightsquigarrow v}\right] \\ &= \Pr\left[\mathcal{D}_{F_1 \rightsquigarrow v} \cup \mathcal{D}_{F_1 \rightsquigarrow v}\right] \\ &= \Pr\left[\mathcal{D}_{F_1 \rightsquigarrow v}\right] + \Pr\left[\mathcal{D}_{F_2 \rightsquigarrow v}\right] - \Pr\left[\mathcal{D}_{F_1 \rightsquigarrow v} \cap \mathcal{D}_{F_2 \rightsquigarrow v}\right] \\ &\leq \Pr\left[\mathcal{D}_{F_1 \rightsquigarrow v}\right] + \Pr\left[\mathcal{D}_{F_2 \rightsquigarrow v}\right] - \Pr\left[\mathcal{D}_{F_1 \cap F_2 \rightsquigarrow v}\right] \\ &= \pi_v(F_1) + \pi_v(F_2) - \pi_v(F_1 \cap F_2)\,, \end{aligned}$$

  where the inequality above holds since $\mathcal{D}_{F_1 \cap F_2 \rightsquigarrow v} \subseteq (\mathcal{D}_{F_1 \rightsquigarrow v} \cap \mathcal{D}_{F_2 \rightsquigarrow v})$.

□

Appendix 2: APX-Hardness for Undirected Graphs

Theorem 6.1 *The maximum reliability coverage problem on undirected graphs is APX-hard.*

Proof We describe a gap-preserving reduction from the minimum-cardinality vertex cover problem on cubic graphs (henceforth, VCC), which is known to be APX-hard [2]. In other words, for some constant $\alpha > 0$, it is NP-hard to distinguish between graphs with $\tau(G) \leq k$ and those with $\tau(G) \geq (1 + \alpha)k$, where $\tau(G)$ stands for the minimum size of a vertex cover in G. Given an instance of VCC, consisting of a cubic graph $G = (V, E)$ on n vertices and a parameter $k \geq |E|/3$, we construct an instance of maximum reliability coverage on the same underlying graph as follows:

- Each vertex has a demand of 1.
- Each edge has a survival probability of 1/2.
- At most k facilities can be located.

Under this reduction, letting F^* be an optimal set of facilities, we proceed by proving the following claims:

1. If $\tau(G) \le k$ then $\mathcal{E}(F^*) \ge \frac{7n}{8} + \frac{k}{8}$.
2. If $\tau(G) \ge (1+\alpha)k$ then $\mathcal{E}(F^*) \le \frac{7n}{8} + \frac{k}{8} - \frac{\alpha n}{64}$.

These claims imply that, unless P = NP, maximum reliability coverage on undirected graphs cannot be approximated within factor greater than

$$\frac{\frac{7n}{8} + \frac{k}{8} - \frac{\alpha n}{64}}{\frac{7n}{8} + \frac{k}{8}} = 1 - \frac{\alpha n}{8 \cdot (7n+k)} \le 1 - \frac{\alpha}{64}.$$

Proof of Claim 1 Since $\tau(G) \le k$, there exists a vertex cover $U \subseteq V$ with $|U| = k$. Now, when we locate facilities at U, each $v \in U$ is covered with probability $\pi_v(U) = 1$, and each vertex $v \notin U$ is covered with probability $\pi_v(U) = 7/8$. To understand the latter claim, note that since U is a vertex cover, when $v \notin U$ its set of neighbors $N(v)$ is necessarily a subset of U, in which case $\pi_v(U) = 1 - (1/2)^{|N(v)|} = 1 - (1/2)^3$. As a result,

$$\mathcal{E}(F^*) \ge \mathcal{E}(U) = |U| \cdot 1 + |V \backslash U| \cdot \frac{7}{8} = k + (n-k) \cdot \frac{7}{8} = \frac{7n}{8} + \frac{k}{8}.$$

Proof of Claim 2 With respect to the optimal set of facilities F^*, as before, each $v \in F^*$ is covered with probability $\pi_v(F^*) = 1$. On the other hand, each $v \notin F^*$ has $|N(v) \cap F^*|$ facilities within its set of neighbors as well as $|N(v) \backslash F^*|$ facility-free neighbors. Therefore, to derive a simple bound on $\pi_v(F^*)$, note that this probability can only increase when we replace the two neighbors (different from v) of each $u \in N(v) \backslash F^*$ by two auxiliary vertices that are connected only to u, while locating facilities in both. In this setting, it is easy to verify that we obtain an upper bound of

$$\pi_v(F^*) \le 1 - \left(\frac{1}{2}\right)^{|N(v) \cap F^*|} \cdot \left(1 - \frac{1}{2} \cdot \frac{3}{4}\right)^{|N(v) \backslash F^*|} = 1 - \frac{1}{8} \cdot \left(\frac{5}{4}\right)^{|N(v) \backslash F^*|}.$$

Consequently,

$$\begin{aligned}
\mathcal{E}(F^*) &\le |F^*| \cdot 1 + \sum_{v \in V \backslash F^*} \left(1 - \frac{1}{8} \cdot \left(\frac{5}{4}\right)^{|N(v) \backslash F^*|}\right) \\
&\le |F^*| + (|V| - |F^*| - \alpha k) \cdot \frac{7}{8} + \alpha k \cdot \frac{27}{32} \\
&= \frac{7n}{8} + \frac{k}{8} - \frac{\alpha k}{32} \\
&\le \frac{7n}{8} + \frac{k}{8} - \frac{\alpha n}{64}.
\end{aligned}$$

The second inequality holds since the number of vertices $v \in V \setminus F^*$ for which $N(v) \setminus F^* \neq \emptyset$ is at least αk. Otherwise, by adding these vertices to F^* we obtain a vertex cover in G with cardinality strictly smaller than $|F^*| + \alpha k \leq (1+\alpha)k$, which contradicts the case hypothesis, $\tau(G) \geq (1+\alpha)k$. The last inequality follows by recalling that $k \geq |E|/3 = n/2$, as G is a cubic graph. □

References

1. Ageev, A.A., Sviridenko, M.: Pipage rounding: a new method of constructing algorithms with proven performance guarantee. J. Comb. Optim. **8**(3), 307–328 (2004)
2. Alimonti, P., Kann, V.: Some APX-completeness results for cubic graphs. Theor. Comput. Sci. **237**(1–2), 123–134 (2000)
3. Ball, M.O.: Computational complexity of network reliability analysis: an overview. IEEE Trans. Reliab. **35**(3), 230–239 (1986)
4. Colbourn, C.J., Xue, G.: A linear time algorithm for computing the most reliable source on a series–parallel graph with unreliable edges. Theor. Comput. Sci. **209**(1), 331–345 (1998)
5. Ding, W., Xue, G.: A linear time algorithm for computing a most reliable source on a tree network with faulty nodes. Theor. Comput. Sci. **412**(3), 225–232 (2011)
6. Eiselt, H.A., Gendreau, M., Laporte, G.: Location of facilities on a network subject to a single-edge failure. Networks **22**(3), 231–246 (1992)
7. Eiselt, H.A., Gendreau, M., Laporte, G.: Optimal location of facilities on a network with an unreliable node or link. Inf. Process. Lett. **58**(2), 71–74 (1996)
8. Feige, U.: A threshold of $\ln n$ for approximating set cover. J. ACM **45**(4), 634–652 (1998)
9. 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)
10. Hadamard, J.: Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. Bull. Soc. Math. Fr. **24**, 199–220 (1896)
11. Hassin, R., Ravi, R., Salman, F.S.: Multiple facility location on a network with linear reliability order of edges. J. Comb. Optim. **34**(3), 931–955 (2017)
12. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. **58**(301), 13–30 (1963)
13. Karp, R.M.: Reducibility among combinatorial problems. In: Miller, R.E., Thatcher, J.W., Bohlinger, J.D. (eds.) Complexity of Computer Computations, pp. 85–103. Springer, Berlin (1972)
14. Melachrinoudis, E., Helander, M.E.: A single facility location problem on a tree with unreliable edges. Networks **27**(3), 219–237 (1996)
15. Nel, L.D., Colbourn, C.J.: Locating a broadcast facility in an unreliable network. INFOR Inf. Syst. Oper. Res. **28**(4), 363–379 (1990)
16. Nemhauser, G.L., Wolsey, L.A., Fisher, M.L.: An analysis of approximations for maximizing submodular set functions—I. Math. Program. **14**(1), 265–294 (1978)
17. Provan, J.S., Ball, M.O.: The complexity of counting cuts and of computing the probability that a graph is connected. SIAM J. Comput. **12**(4), 777–788 (1983)
18. Santivanez, J., Melachrinoudis, E., Helander, M.E.: Network location of a reliable center using the most reliable route policy. Comput. Oper. Res. **36**(5), 1437–1460 (2009)
19. Spivak, M.: Calculus, 3rd edn. Cambridge University Press, Cambridge (1967)
20. Valiant, L.G.: The complexity of enumeration and reliability problems. SIAM J. Comput. **8**(3), 410–421 (1979)
21. Vallée-Poussin, C.: Recherches analytiques sur la théorie des nombres premiers. Ann. Soc. Sci. Brux. **20**, 183–256 (1896)
22. Xue, G.: Linear time algorithms for computing the most reliable source on an unreliable tree network. Networks **30**(1), 37–45 (1997)

Affiliations

Refael Hassin[1] · R. Ravi[2] · F. Sibel Salman[3] · Danny Segev[4]

Refael Hassin
hassin@post.tau.ac.il

R. Ravi
ravi@cmu.edu

F. Sibel Salman
ssalman@ku.edu.tr

1 Department of Statistics and Operations Research, Tel Aviv University, 69978 Tel Aviv, Israel

2 Tepper School of Business, Carnegie Mellon University, Pittsburgh, USA

3 College of Engineering, Koç University, Sariyer, Istanbul, Turkey

4 Department of Statistics, University of Haifa, 31905 Haifa, Israel