

Probabilistic Analysis of Facility Location on Random Shortest Path Metrics

Stefan Klootwijk¹ and Bodo Manthey¹

¹University of Twente, Enschede, The Netherlands

Abstract

The facility location problem is an \mathcal{NP} -hard optimization problem. Therefore, approximation algorithms are often used to solve large instances. Probabilistic analysis is a widely used tool to analyze such algorithms. Most research on probabilistic analysis of \mathcal{NP} -hard optimization problems involving metric spaces, such as the facility location problem, has been focused on Euclidean instances, and also instances with independent (random) edge lengths, which are non-metric, have been researched. However, we would like to extend this knowledge to other, more general, metrics.

We investigate the facility location problem using random shortest path metrics. We analyze some probabilistic properties for a simple heuristic which gives a solution to the facility location problem: opening a certain number of arbitrary facilities (with that certain number only depending on the facility opening cost). We show that, for almost any facility opening cost, this heuristic yields a $1 + o(1)$ approximation in expectation. In the remaining few cases we show that this heuristic yields an $O(1)$ approximation in expectation.

Keywords: Facility location, Random shortest paths, Random metrics, Approximation algorithm

1 Introduction

The (uncapacitated) facility location problem can be described as follows: given a (complete) graph $G = (V, E)$, facility opening cost f_i for each vertex $v_i \in V$ and a distance $d(u, v)$ between each pair of vertices $u, v \in V$, find a subset $U \subseteq V$ of vertices at which you open facilities such that the total cost is minimized. Here, the total cost is given by the sum of the opening cost f_i for all vertices $v_i \in U$ and the sum of the ‘connection’ cost $\min_{u \in U} d(v, u)$ for all vertices $v \in V \setminus U$.

This problem is known to be \mathcal{NP} -hard [2]. Therefore research on the facility location problem (and other \mathcal{NP} -hard problems) has been focused on different heuristics, ranging from straightforward to rather sophisticated, and their worst-case performance (for instance [4, 7]).

So far, probabilistic analysis of heuristics for optimization problems like the facility location problem has been focused on instances either using Euclidean space or on (non-metric)

instances with independent (random) edge lengths, since such instances are technically relatively easy to handle. However, we would like to apply probabilistic analysis to more general metric instances. To do so, we use so-called ‘random shortest path metrics’, which have also been used by Bringmann et al. [1], who initiated this research.

Random shortest path metrics are defined as follows. Consider an undirected complete graph $G = (V, E)$ on n vertices. For any edge $e \in E$, let $w(e) \sim \text{Exp}(1)$ be the weight of edge e , independently drawn from the standard exponential distribution. Then, the distances $d(u, v)$ between each pair of vertices $u, v \in V$ are defined as the minimum total weight of any u, v -path \mathcal{P} in G . The underlying model of random shortest path metrics is also known as first-passage percolation.

Many structural properties of random shortest path metrics are known, such as the expected shortest path length ($\ln(n)/n$ in expectation as $n \rightarrow \infty$) [1, 3, 6], and the number of edges on the shortest path between any two vertices [5].

We consider instances of the facility location problem for which the distances d are randomly generated using the principle of random shortest path metrics and for which every vertex has the same nonnegative facility opening cost f , i.e. $f_i = f \geq 0$ for all $i \in V$. This implies that the total cost of any solution $\emptyset \neq U \subseteq V$ is given by

$$\text{cost}(U) = f \cdot |U| + \sum_{v \in V \setminus U} \min_{u \in U} d(v, u).$$

Although we do not mention it explicitly, the facility opening cost f does depend on the size of the instance, i.e., we have $f = f(n)$. It makes sense to do this, since the expected distance between two arbitrary vertices also depends on n when using random shortest path metrics.

We show that the most trivial procedure of opening a fixed number of arbitrary facilities (with that fixed number only depending on the facility opening cost f) yields a $1 + o(1)$ approximation in expectation unless $f \in \Theta(1/n)$. If $f \in \Theta(1/n)$, then this procedure is shown to yield a $O(1)$ approximation in expectation.

2 An intuitive approach

Intuitively, we observe that the optimal solution for our problem will satisfy $|U| \approx n$ when the facility opening cost f is (almost) 0. On the other hand, the optimal solution will satisfy $|U| = 1$ when the facility opening cost is relatively large. And when the facility opening cost f are neither ‘relatively small’ nor ‘relatively large’, $|U|$ will be neither close to n nor close to 1 for the optimal solution.

Let OPT denote the total cost of the optimal solution to the facility location problem. Furthermore, for $k \in [n] := \{1, \dots, n\}$, let OPT_k denote the total cost of the optimal solution to the corresponding k -median problem. Then it follows that

$$\text{OPT} = \min_{\emptyset \neq U \subseteq V} \text{cost}(U) = \min_{\emptyset \neq U \subseteq V} \left(f \cdot |U| + \sum_{v \in V \setminus U} \min_{u \in U} d(v, u) \right) = \min_{k \in [n]} (f \cdot k + \text{OPT}_k).$$

Moreover, based on the results of Bringmann et al. [1, Sect. 5] we know that $U_k := \{v_1, \dots, v_k\}$ is a good approximation for the optimal solution to the k -median problem (whenever k is not too large), and that the expected cost of this solution is given by $\mathbb{E}[\text{cost}_k(U_k)] = \ln(n/k) + \Theta(1)$.

So, intuitively we see that

$$\begin{aligned} \mathbb{E}[\text{OPT}] &= \mathbb{E} \left[\min_{k \in [n]} (f \cdot k + \text{OPT}_k) \right] \approx \min_{k \in [n]} (f \cdot k + \mathbb{E}[\text{OPT}_k]) \\ &\approx \min_{k \in [n]} (f \cdot k + \mathbb{E}[\text{cost}_k(U_k)]) = \min_{k \in [n]} (f \cdot k + \ln(n/k) + \Theta(1)). \end{aligned}$$

Finally, we observe that the function $g(k) = f \cdot k + \ln(n/k)$ is minimal for $k = 1/f$, resulting in $\mathbb{E}[\text{OPT}] \approx 1 + \ln(nf) + \Theta(1)$. Combining this observation with the foregoing intuitive arguments, it seems likely that any arbitrary solution U to the facility location problem with $|U| \approx 1/f$ yields a good approximation for OPT .

3 Main results

We are interested in the expected approximation ratio of an algorithm that opens approximately $1/f$ arbitrary facilities. In order to analyze this, we use the rather trivial algorithm which opens exactly $k := \min\{\lceil 1/f \rceil, n\}$ randomly chosen facilities. Let TRIV denote the total cost of the solution computed by this algorithm, and let OPT denote the total cost of an optimal solution to the facility location problem. Then, using a result from Bringmann et al. [1, Sect. 5], we can derive the probability distribution of TRIV .

Lemma 1. *If $k = n$, then TRIV has a degenerate probability distribution with $\mathbb{P}(\text{TRIV} = nf) = 1$. Otherwise, the distribution of TRIV is given by*

$$\text{TRIV} \sim k \cdot f + \sum_{i=k}^{n-1} \text{Exp}(i),$$

where the $\text{Exp}(i)$ are independent exponentially distributed random variables with parameter i .

Observe that we can use our intuitive approach to show that $\mathbb{E}[\text{TRIV}] \approx \mathbb{E}[\text{OPT}]$. However, using a more thorough analysis, in which we combine this probability distribution with some bounds for OPT , we can show that TRIV yields either a constant or an asymptotically optimal approximation ratio. This is summarized in the following theorem.

Theorem 2. *Let OPT denote the total cost of the optimal solution to the facility location problem, and let TRIV denote the total cost of the solution which opens exactly $\min\{\lceil 1/f \rceil, n\}$ randomly chosen facilities. Then, it follows that*

$$\mathbb{E} \left[\frac{\text{TRIV}}{\text{OPT}} \right] = O(1).$$

Moreover, if either $f \in o(1/n)$ or $f \in \omega(1/n)$, then it follows that

$$\mathbb{E} \left[\frac{\text{TRIV}}{\text{OPT}} \right] = 1 + o(1).$$

In order to prove this theorem, we divide the range of possible (asymptotic) facility opening costs f in three (slightly overlapping) intervals, corresponding to the three intuitive cases mentioned above: opening all facilities ($f \leq (2 - \varepsilon)/n$), opening exactly one arbitrary facility ($f \geq 1/n^\varepsilon$), and opening some arbitrary facilities ($(1 + \varepsilon)/n \leq f \leq M/n^\varepsilon$). For each case we have found a threshold such that conditioning the expected approximation ratio on the events

OPT is larger relatively smaller than this threshold, allows us to prove the bounds mentioned in Theorem 2.

These thresholds are chosen in such a way that for the case in which OPT is larger than the threshold, it is relatively easy to bound the conditional expected approximation ratio. On the other hand, the thresholds are also chosen in such a way that the probability of OPT being smaller than the threshold becomes sufficiently small. By doing so, we are able to show that the (relatively) large conditional expected approximation ratio in this case becomes negligible when multiplied with that probability.

4 Final remarks

As far as we are aware, these results form only a second step into the research of the behavior of (combinatorial) optimization problems using random shortest path metrics (the first step being the results in by Bringmann et al. [1]). Even though random shortest path instances are more difficult to analyze than Euclidean instances or instances with independent random edge lengths, we were able to derive some good results when analyzing the facility location problem on it.

It would be interesting to see whether it is possible to prove similar results when using more sophisticated heuristics that aim to solve the facility location problem. Furthermore, there are many other \mathcal{NP} -hard (combinatorial) optimization problems involving metric spaces for which it would be interesting to know how they behave on random shortest path metrics.

References

- [1] K. Bringmann, C. Engels, B. Manthey, and B.V.R. Rao. Random shortest paths: Non-euclidean instances for metric optimization problems. *Algorithmica*, 73(1):42–62, 2015. doi: 10.1007/s00453-014-9901-9.
- [2] G. Cornuejols, G.L. Nemhauser, and L.A. Wolsey. The uncapacitated facility location problem. In Pitu B. Mirchandani and Richard L. Francis, editors, *Discrete Location Theory*, chapter 3, pages 119–171. Wiley-Interscience, New York, 1990. ISBN 978-0-471-89233-5.
- [3] R. Davis and A. Prieditis. The expected length of a shortest path. *Information Processing Letters*, 46(3):135–141, 1993. doi: 10.1016/0020-0190(93)90059-I.
- [4] A.D. Flaxman, A.M. Frieze, and J.C. Vera. On the average case performance of some greedy approximation algorithms for the uncapacitated facility location problem. *Combinatorics, Probability and Computing*, 16(5):713–732, 2007. doi: 10.1017/S096354830600798X.
- [5] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. First-passage percolation on the random graph. *Probability in the Engineering and Informational Science*, 15(2): 225–237, 2001. doi: 10.1017/S026996480115206X.
- [6] S. Janson. One, two and three times $\log n/n$ for paths in a complete graph with random weights. *Combinatorics, Probability and Computing*, 8(4):347–361, 1999. doi: 10.1017/S0963548399003892.
- [7] S. Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. *Information and Computation*, 222:45–58, 2013. doi: 10.1016/j.ic.2012.01.007.
- [8] J. Vygen. Approximation algorithms for facility location problems (lecture notes). Technical Report No. 05950, Research Institute for Discrete Mathematics, University of Bonn, 2005. URL <http://www.or.uni-bonn.de/~vygen/files/fl.pdf>.