Probabilistic Analysis of Optimization Problems on Generalized Random Shortest Path Metrics

Stefan Klootwijk(✉), Bodo Manthey, and Sander K. Visser

Faculty of Electrical Engineering, Mathematics and Computer Science,
University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
{s.klootwijk,b.manthey}@utwente.nl, s.k.visser@alumnus.utwente.nl

Abstract. Simple heuristics often show a remarkable performance in practice for optimization problems. Worst-case analysis often falls short of explaining this performance. Because of this, "beyond worst-case analysis" of algorithms has recently gained a lot of attention, including probabilistic analysis of algorithms.

The instances of many optimization problems are essentially a discrete metric space. Probabilistic analysis for such metric optimization problems has nevertheless mostly been conducted on instances drawn from Euclidean space, which provides a structure that is usually heavily exploited in the analysis. However, most instances from practice are not Euclidean. Little work has been done on metric instances drawn from other, more realistic, distributions. Some initial results have been obtained by Bringmann et al. (*Algorithmica*, 2013), who have used random shortest path metrics on complete graphs to analyze heuristics.

The goal of this paper is to generalize these findings to non-complete graphs, especially Erdős–Rényi random graphs. A random shortest path metric is constructed by drawing independent random edge weights for each edge in the graph and setting the distance between every pair of vertices to the length of a shortest path between them with respect to the drawn weights. For such instances, we prove that the greedy heuristic for the minimum distance maximum matching problem, the nearest neighbor and insertion heuristics for the traveling salesman problem, and a trivial heuristic for the k-median problem all achieve a constant expected approximation ratio. Additionally, we show a polynomial upper bound for the expected number of iterations of the 2-opt heuristic for the traveling salesman problem.

1 Introduction

Large-scale optimization problems, such as the traveling salesman problem (TSP), show up in many applications. These problems are often computationally intractable. However, in practice often ad-hoc heuristics are successfully used

A full version containing all proofs is available at https://arxiv.org/abs/1810.11232.

G. K. Das et al. (Eds.): WALCOM 2019, LNCS 11355, pp. 108–120, 2019.
https://doi.org/10.1007/978-3-030-10564-8_9

that provide solutions that come quite close to optimal solutions. In many cases these, often simple, heuristics show a remarkable performance, even though the theoretical results about those heuristics are way more pessimistic.

In order to explain this difference, probabilistic analysis has been widely used over the last decades. However, the challenge in probabilistic analysis is to come up with a good probabilistic model: it should reflect realistic instances, but also be sufficiently simple to make the analysis tractable.

So far, in almost all cases, either Euclidean space has been used to generate instances of metric optimization problems, or independent, identically distributed edge lengths have been used (e.g. [1,6]). However, both approaches have considerable shortcomings to explain the average-case performance of heuristics on general metric instances: the structure of Euclidean space is heavily used in the probabilistic analysis, but realistic instances are often not Euclidean. The independent, identically distributed edge lengths do not even yield a metric in the first place. In order to overcome these shortcomings, Bringmann et al. [3] have proposed and analyzed the following model to generate random metric spaces, which had already been proposed by Karp and Steele in 1985 [12]: given an undirected complete graph, start by drawing random edge weights for each edge independently and then define the distance between any two vertices as the total weight of the shortest path between them, measured with respect to the random weights.

1.1 Related Work

Bringmann et al. called the model described above *random shortest path metrics*. This model is also known as *first-passage percolation*, introduced by Hammersley and Welsh as a model for fluid flow through a (random) porous medium [7,9].

For first passage percolation in complete graphs, the expected distance between two fixed vertices is approximately $\ln(n)/n$ and the expected distance from a fixed vertex to the vertex that is most distant is approximately $2\ln(n)/n$ [3,10]. Furthermore, it is known that the expected diameter of the metric is approximately $3\ln(n)/n$ [8,10]. There are also some known structural properties of first passage percolation on the Erdős–Rényi random graph. Bhamidi et al. [2] have shown asymptotics for both the minimal weight of the path between uniformly chosen vertices in the giant component and for the hopcount, the number of edges, on this path. Bringmann et al. [3] used this model on the complete graph to analyze heuristics for matching, TSP, and k-median.

1.2 Our Results

As far as we know, no heuristics have been studied in this model for non-complete graphs yet. However, we believe that random shortest path metrics on non-complete graphs will bring us a step further in the direction of realistic input model.

This paper provides a probabilistic analysis of some simple heuristics in the model of random shortest path metrics on non-complete graphs. First, we provide some structural properties of generalized random shortest path metrics (Sect. 3), which can be seen as a generalization of the structural properties found by Bringmann et al. [3]. Although this generalization might seem straightforward at first sight, it brings up some new difficulties that need to be overcome. Most notably, since we do not restrict ourselves to the complete graph, we cannot make use anymore of its symmetry and regularity. This problem is partially solved by introducing two graph parameters, which we call the cut parameters of a graph (Definition 1).

Then, we use these structural insights to perform a probabilistic analysis for some simple heuristics for combinatorial optimization problems (Sect. 4), where the results are still depending on the cut parameters of a graph. Finally, we use these results, to show our main results, namely that these simple heuristics achieve constant expected approximation ratios for random shortest path metrics applied to Erdős–Rényi random graphs (Sect. 5).

2 Notation and Model

We use $X \sim P$ to denote that a random variable X is distributed using a probability distribution P. $\text{Exp}(\lambda)$ is being used to denote the exponential distribution with parameter λ. In particular, we use $X \sim \sum_{i=1}^{n} \text{Exp}(\lambda_i)$ to denote that X is the sum of n independent exponentially distributed random variables having parameters $\lambda_1, \ldots, \lambda_n$.

For $n \in \mathbb{N}$, we use $[n]$ as shorthand notation for $\{1, \ldots, n\}$. We denote the nth harmonic number by $H_n = \sum_{i=1}^{n} 1/i$. Sometimes we use exp to denote the exponential function. Finally, if a random variable X is stochastically dominated by a random variable Y, i.e., we have $F_X(x) \geq F_Y(x)$ for all x (where $X \sim F_X$ and $Y \sim F_Y$), we denote this by $X \precsim Y$.

Generalized Random Shortest Path Metrics. Given an undirected graph $G = (V, E)$ on n vertices, we construct the corresponding generalized random shortest path metric as follows. First, for each edge $e \in E$, we draw a random edge weight $w(e)$ independently from an exponential distribution[1] with parameter 1. Second, we define the distances $d : V \times V \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ as follows: for every $u, v \in V$, $d(u, v)$ denotes the length of the shortest u, v-path with respect to the drawn edge weights. If no such path exists, we set $d(u, v) = \infty$. By doing so, the distance function d satisfies $d(v, v) = 0$ for all $v \in V$, $d(u, v) = d(v, u)$ for all $u, v \in V$, and $d(u, v) \leq d(u, s) + d(s, v)$ for all $u, s, v \in V$. We call the complete graph with distances d obtained from this process a generalized random shortest path metric. If $G = K_n$ (the complete graph on n vertices), then this generalized random shortest path metric is equivalent to the random shortest path metric as defined by Bringmann et al. [3]

[1] Exponential distributions are technically easiest to handle due to their memorylessness property. A (continuous, non-negative) probability distribution of a random variable X is said to be memoryless if and only if $\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s)$ for all $s, t \geq 0$. [15, p. 294].

We use the following notation within generalized random shortest path metrics: $\Delta_{\max} := \max_{u,v} d(u,v)$ denotes the diameter of the graph. Note that $\Delta_{\max} < \infty$ if and only if G is connected. $B_{\Delta}(v) := \{u \in V \mid d(u,v) \leq \Delta\}$ denotes the 'ball' of radius Δ around v, i.e., the set containing all vertices at distance at most Δ from v. $\tau_k(v) := \min\{\Delta \mid |B_{\Delta}(v)| \geq k\}$ denotes the distance to the kth closest vertex from v (including v itself). Equivalently, one can also say that $\tau_k(v)$ is equal to the smallest Δ such that the ball of radius Δ around v contains at least k vertices.

Now, $B_{\tau_k(v)}(v)$ denotes the set of the k closest vertices to v. During our analysis, we make use of the size of the cut induced by this set, which we denote by $\chi_k(v) := |\delta(B_{\tau_k(v)}(v))|$, where $\delta(U)$ denotes the cut induced by U.

Erdős–Rényi Random Graphs. The main results of this work consider random shortest path metrics applied to Erdős–Rényi random graphs. An undirected graph $G(n,p) := G = (V,E)$ generated by this model has n vertices ($V = \{1,\dots,n\}$) and between each pair of vertices an edge is included with probability p, independent of every other pair.

Working with the Erdős–Rényi random graph introduces an extra amount of stochasticity to the probabilistic analysis, since both the graph and the edge weights are random. In order to avoid this extra stochasticity as long as possible, in Sects. 3 and 4 we start our analysis using an arbitrary fixed (deterministic) graph G. Later on, in Sect. 5 we will consider Erdős–Rényi random graphs again.

3 Structural Properties

In order to analyze the structural properties of generalized random shortest path metrics, we first introduce the notion of what we call the cut parameters of a simple graph G.

Definition 1. *Let $G = (V,E)$ be a finite simple connected graph. Then we define the cut parameters of G by*

$$\alpha := \min_{\varnothing \neq U \subset V} \frac{|\delta(U)|}{\mu_U} \quad \textit{and} \quad \beta := \max_{\varnothing \neq U \subset V} \frac{|\delta(U)|}{\mu_U},$$

where $\mu_U := |U| \cdot (|V| - |U|)$ is the maximum number of possible edges in the cut defined by U.

It follows immediately from this definition that $0 < \alpha \leq \beta \leq 1$ for any finite simple connected graph G. Moreover, for any such graph the following holds for all $\varnothing \neq U \subset V : \alpha \cdot \mu_U \leq |\delta(U)| \leq \beta \cdot \mu_U$. We observe that the cut parameters of the complete graph are given by $\alpha = \beta = 1$.

Distribution of $\boldsymbol{\tau_k(v)}$. Now we have a look at the distribution of $\tau_k(v)$. For this purpose we use an arbitrary fixed undirected connected simple graph G (on n vertices) and let α and β denote its cut parameters.

The values of $\tau_k(v)$ are then generated by a birth process as follows. (Amongst others, a variant of this process for complete graphs has been analyzed by Davis and Prieditis [5] and Bringmann et al. [3].) For $k = 1$, we have $\tau_k(v) = 0$. For $k \geq 2$, we look at all edges (u, x) with $u \in B_{\tau_{k-1}(v)}(v)$ and $x \notin B_{\tau_{k-1}(v)}(v)$. By definition there are $\chi_{k-1}(v)$ such edges. Moreover the length of these edges is conditioned to be at least $\tau_{k-1}(v) - d(v, u)$. Using the memorylessness of the exponential distribution, we can now see that $\tau_k(v) - \tau_{k-1}(v)$ is the minimum of $\chi_{k-1}(v)$ (standard) exponential variables, or, equivalently, $\tau_k(v) - \tau_{k-1}(v) \sim \mathrm{Exp}(\chi_{k-1}(v))$. We use this result to find bounds for the distribution of $\tau_k(v)$.

Lemma 2. *For all $k \in [n]$ and $v \in V$ we have,*

$$\alpha k(n-k) \leq \chi_k(v) \leq \beta k(n-k).$$

Lemma 3. *For all $k \in [n]$ and $v \in V$ we have,*

$$\sum_{i=1}^{k-1} \mathrm{Exp}(\beta i(n-i)) \precsim \tau_k(v) \precsim \sum_{i=1}^{k-1} \mathrm{Exp}(\alpha i(n-i)).$$

Exploiting the linearity of expectation, the fact that the expected value of an exponentially distributed random variable with parameter λ is $1/\lambda$ and the fact that $\sum_{i=1}^{k-1} 1/(i(n-i)) = (H_{k-1} + H_{n-1} - H_{n-k})/n$, we obtain the following corollary.

Corollary 4. *For all $k \in [n]$ and $v \in V$ we have,*

$$\frac{H_{k-1} + H_{n-1} - H_{n-k}}{\beta n} \leq \mathbb{E}(\tau_k(v)) \leq \frac{H_{k-1} + H_{n-1} - H_{n-k}}{\alpha n}.$$

From this result, we can derive the following extensions of two known results. First of all, if we randomly pick two vertices $u, v \in V$, then averaging over k yields that the expected distance $\mathbb{E}[d(u, v)]$ between them is bounded between $\frac{H_{n-1}}{\beta(n-1)} \approx \ln(n)/\beta n$ and $\frac{H_{n-1}}{\alpha(n-1)} \approx \ln(n)/\alpha n$, which is in line with the known result for complete graphs, where we have $\mathbb{E}[d(u, v)] \approx \ln(n)/n$ [3,5,10]. Secondly, for any vertex v, the longest distance from it to another vertex is $\tau_n(v)$, which in expectation is bounded between $\frac{2H_{n-1}}{\beta n} \approx 2\ln(n)/\beta n$ and $\frac{2H_{n-1}}{\alpha n} \approx 2\ln(n)/\alpha n$, which also is in line with the known result for complete graphs, where we have an expected value of approximately $2\ln(n)/n$ [3,10].

It is also possible to find bounds for the cumulative distribution function of $\tau_k(v)$. To do so, we define $F_k(x) = \mathbb{P}(\tau_k(v) \leq x)$ for some fixed vertex $v \in V$.

Lemma 5. [3, Lemma 3.2] *Let $X \sim \sum_{i=1}^{n} \mathrm{Exp}(ci)$. Then, for any $a \geq 0$ we have $\mathbb{P}(X \leq a) = (1 - e^{-ca})^n$.*

Lemma 6. *For all $x \geq 0$ and $k \in [n]$ we have,*

$$(1 - \exp(-\alpha(n-k)x))^{k-1} \leq F_k(x) \leq (1 - \exp(-\beta n x))^{k-1}.$$

We can improve this result slightly.

Lemma 7. *For all $x \geq 0$ and $k \in [n]$ we have,*

$$F_k(x) \geq (1 - \exp(-\alpha n x/4))^n .$$

Using this improved bound for the cumulative distribution function of $\tau_k(v)$, we can derive the following tail bound for the diameter $\Delta_{\max}$.

Lemma 8. *Let $\Delta_{\max} = \max_{u,v \in V}\{d(u,v)\}$. For any fixed c we have $\mathbb{P}(\Delta_{\max} > c\ln(n)/\alpha n) \leq n^{2-c/4}$.*

Clustering. In this section we show that we can partition the vertices of generalized random shortest path metrics into a small number of clusters with a given maximum diameter. Before we prove this main result, we first provide a tail bound for $|B_\Delta(v)|$.

Lemma 9. *For $n \geq 5$ and for any fixed $\Delta \geq 0$ we have,*

$$\mathbb{P}\left(|B_\Delta(v)| < \min\left\{\exp(\alpha\Delta n/5), \frac{n+1}{2}\right\}\right) \leq \exp(-\alpha\Delta n/5).$$

We use the result of this lemma to prove our main structural property for generalized random shortest path metrics.

Theorem 10. *For any fixed $\Delta \geq 0$, if we partition the vertices into clusters, each of diameter at most 4Δ, then the expected number of clusters needed is bounded from above by $O(1 + n/\exp(\alpha\Delta n/5))$.*

4 Analysis of Heuristics

In this section we bound the expected approximation ratios of the greedy heuristic for minimum-distance perfect matching, the nearest neighbor and insertion heuristics for the traveling salesman problem, and a trivial heuristic for the k-median problem. For this purpose we still use an arbitrary fixed undirected connected simple graph G (on n vertices) and let α and β denote its cut parameters. The results in this section will depend on α and β.

Greedy Heuristic for Minimum-Distance Perfect Matching. The minimum-distance perfect matching problem has been widely analyzed throughout history. We do for instance know that the worst-case running-time for finding a minimum distance perfect matching is $O(n^3)$, which is high when considering a large number of vertices. Because of this, simple heuristics are often used, with the greedy heuristic probably being the simplest of them: at each step, add a pair of unmatched vertices to the matching such that the distance between the added pair of vertices is minimized. From now on, let GR denote the cost of the matching computed by this heuristic and let MM denote the value of an optimal matching.

The worst-case approximation ratio of this heuristic on metric instances is known to be $O(n^{\log_2(3/2)})$ [13]. Furthermore, for random shortest path metrics on complete graphs (for which the cut parameters are given by $\alpha = \beta = 1$) the heuristic has an expected approximation ratio of $O(1)$ [3]. We extend this last result to general values for α and β and show that the greedy matching heuristic has an expected approximation ratio of $O(\beta/\alpha)$.

Theorem 11. $\mathbb{E}[\mathsf{GR}] = O(1/\alpha)$.

Lemma 12. [11, Theorem 5.1(iii)] *Let* $X \sim \sum_{i=1}^{n} X_i$ *with* $X_i \sim \mathrm{Exp}(a_i)$ *independent. Let* $\mu = \mathbb{E}[X] = \sum_{i=1}^{n}(1/a_i)$ *and* $a_* = \min_i a_i$. *For any* $\lambda \leq 1$,

$$\mathbb{P}(X \leq \lambda\mu) \leq \exp(-a_*\mu(\lambda - 1 - \ln(\lambda))).$$

Lemma 13. *Let* S_m *denote the sum of the* m *lightest edge weights in* G. *For all* $\phi \leq (n-1)/n$ *and* $c \in [0, 2\phi^2]$ *we have*

$$\mathbb{P}\left(S_{\phi n} \leq \frac{c}{\beta}\right) \leq \exp\left(\phi n\left(1 + \ln\left(\frac{c}{2\phi^2}\right)\right)\right).$$

Furthermore, $\mathsf{TSP} \geq \mathsf{MM} \geq S_{n/2}$, *where* TSP *and* MM *are the total distance of a shortest TSP tour and a minimum-distance perfect matching, respectively.*

Theorem 14. *The greedy heuristic for minimum-distance perfect matching has an expected approximation ratio on generalized random shortest path metrics given by* $\mathbb{E}\left[\frac{\mathsf{GR}}{\mathsf{MM}}\right] = O(\beta/\alpha)$.

Nearest Neighbor Heuristic for TSP. The nearest-neighbor heuristic is a greedy approach for the TSP: start with some starting vertex v_0 as current vertex v; at every step, choose the nearest unvisited neighbor u of v as the next vertex in the tour and move to the next iteration with the new vertex u as current vertex v; go back to v_0 if all vertices are visited. From now on, let NN denote the cost of the TSP tour computed by this heuristic and let TSP denote the value of an optimal TSP tour.

The worst-case approximation ratio of this heuristic on metric instances is known to be $O(\ln(n))$ [14]. Furthermore, for random shortest path metrics on complete graphs (for which the cut parameters are given by $\alpha = \beta = 1$) the heuristic has an expected approximation ratio of $O(1)$ [3]. We extend this last result to general values for α and β and show that the nearest-neighbor heuristic has an expected approximation ratio of $O(\beta/\alpha)$.

Theorem 15. *For generalized random shortest path metrics, we have* $\mathbb{E}[\mathsf{NN}] = O(1/\alpha)$ *and* $\mathbb{E}\left[\frac{\mathsf{NN}}{\mathsf{TSP}}\right] = O(\beta/\alpha)$.

Insertion Heuristics for TSP. The insertion heuristics are another greedy approach for the TSP: start with an initial optimal tour on a few vertices chosen according to some predefined rule R; at every step, choose a vertex according to the same predefined rule R and insert this vertex in the current tour such that the total distance increases the least. From now on, let IN_R denote the cost of the TSP tour computed by this heuristic (with rule R) and let TSP still denote the value of an optimal TSP tour.

The worst-case approximation ratio of this heuristic for any rule R on metric instances is known to be $O(\ln(n))$ [14]. Furthermore, for random shortest path metrics on complete graphs (for which the cut parameters are given by $\alpha = \beta = 1$) the heuristic has an expected approximation ratio of $O(1)$ [3]. We extend this last result to general values for α and β and show that the insertion heuristic for any rule R has an expected approximation ratio of $O(\beta/\alpha)$.

Theorem 16. *For generalized random shortest path metrics, we have* $\mathbb{E}[\mathsf{IN}_R] = O\left(1/\alpha\right)$ *and* $\mathbb{E}\left[\frac{\mathsf{IN}_R}{\mathsf{TSP}}\right] = O\left(\beta/\alpha\right)$.

Running Time of 2-opt Heuristic for TSP. The 2-opt heuristic is an often used local search algorithm for the TSP: start with an initial tour on all vertices and improve the tour by 2-exchanges until no improvement can be made anymore. In a 2-exchange, the heuristic takes 'edges' $\{v_1, v_2\}$ and $\{v_3, v_4\}$, where v_1, v_2, v_3, v_4 are visited in this order in the tour, and replaces them by $\{v_1, v_3\}$ and $\{v_2, v_4\}$ to create a shorter tour.

We provide an upper bound for the expected number of iterations that 2-opt needs. In the worst-case scenario, this number is exponential. However, for random shortest path metrics on complete graphs (for which the cut parameters are given by $\alpha = \beta = 1$) an upper bound of $O(n^8 \ln^3(n))$ is known for the expected number of iterations [3]. We extend this result with a similar proof to general values for α and β and show an upper bound for the expected number of iterations of $O(n^8 \ln^3(n)\beta/\alpha)$.

We first define the improvement obtained from a 2-exchange. If $\{v_1, v_2\}$ and $\{v_3, v_4\}$ are replaced by $\{v_1, v_3\}$ and $\{v_2, v_4\}$, then the improvement made by the exchange equals the change in distance $\zeta = d(v_1, v_2) + d(v_3, v_4) - d(v_1, v_3) - d(v_2, v_4)$. These four distances correspond to four shortest paths (P_{12}, P_{34}, P_{13}, P_{24}) in the graph $G = (V, E)$. This implies that we can rewrite ζ as the sum of the weights on these paths. We obtain $\zeta = \sum_{e \in E} \gamma_e w(e)$, for some $\gamma_e \in \{-2, -1, 0, 1, 2\}$.

Since we are looking at the improvement obtained by a 2-exchange, we have $\zeta > 0$. This implies that there exists some $e = \{u, u'\} \in E$ such that $\gamma_e \neq 0$. Given this edge e, let $I \subseteq \{P_{12}, P_{34}, P_{13}, P_{24}\}$ be the set of all shortest paths of the 2-exchange that contain e. Then, for all combinations e and I, let $\zeta_{ij}^{e,I}$ be defined as follows:

- If $P_{ij} \notin I$, then $\zeta_{ij}^{e,I}$ is the length of the shortest path from v_i to v_j without using e.
- If $P_{ij} \in I$, then $\zeta_{ij}^{e,I}$ is the minimum of

- the length of a shortest path from v_i to u without using e plus the length of a shortest path from u' to v_j without using e and
- the length of a shortest path from v_i to u' without using e plus the length of a shortest path from u to v_j without using e.

Define $\zeta^{e,I} = \zeta_{12}^{e,I} + \zeta_{34}^{e,I} - \zeta_{13}^{e,I} - \zeta_{24}^{e,I}$.

Lemma 17. *For every outcome of the edge weights, there exists an edge e and a set I such that $\zeta = \zeta^{e,I} + \gamma w(e)$, where $\gamma \in \{-2, -1, 1, 2\}$ is determined by e and I.*

Lemma 18. *Let e and I be given with $\gamma = \gamma_e \neq 0$. Then $\mathbb{P}(\zeta^{e,I} + \gamma w(e) \in (0, x]) \leq x$. Moreover, $\mathbb{P}(\zeta \in (0, x]) = O(\beta n^2 x)$.*

Theorem 19. *The expected number of iterations of the 2-opt heuristic until a local optimum is found is bounded by $O(n^8 \ln^3(n)\beta/\alpha)$.*

Trivial Heuristic for *k*-Median. The goal of the (metric) k-median problem is to find a set $U \subseteq V$ of size k such that $\sum_{v \in V} \min_{u \in U} d(v, u)$ is minimized. The best known approximation algorithm for this problem achieves an approximation ratio of $2.675 + \varepsilon$ [4].

Here, we consider the k-median problem in the setting of generalized random shortest path metrics. We analyze a trivial heuristic for the k-median problem: simply pick k vertices independently of the metric space, e.g., $U = \{v_1, \dots, v_k\}$. The worst-case approximation ratio of this heuristic is unbounded, even if we restrict ourselves to metric instances. However, for random shortest path metrics on complete graphs (for which the cut parameters are given by $\alpha = \beta = 1$) the expected approximation ratio has an upper bound of $O(1)$ and even $1 + o(1)$ for k sufficiently small [3]. We extend this result to general values for α and β and give an upper bound for the expected approximation ratio of $O(\beta/\alpha)$ for 'large' k and $\beta/\alpha + o(\beta/\alpha)$ for k sufficiently small.

For our analysis, let $U = \{v_1, \dots, v_k\}$ be an arbitrary set of k vertices. Sort the remaining vertices $\{v_{k+1}, \dots, v_n\}$ in increasing distance from U. For $k + 1 \leq i \leq n$, let $\rho_i = d(v_i, U)$ equal the distance from U to the $(i - k)$-th closest vertex to U. Let TR denote the cost of the solution generated by the trivial heuristic and let ME be the cost of an optimal solution to the k-median problem.

Observe that the random variables ρ_i are generated by a simple growth process analogously to the one described in Sect. 3 for $\tau_k(v)$. Using this observation, we can see that

$$\sum_{j=k}^{i-1} \mathrm{Exp}(\beta j(n - j)) \precsim \rho_i \precsim \sum_{j=k}^{i-1} \mathrm{Exp}(\alpha j(n - j)),$$

which in turn implies that $\mathsf{cost}(U) = \sum_{i=k+1}^{n} \rho_i$ is stochastically bounded by

$$\sum_{i=k}^{n-1} \mathrm{Exp}(\beta i) \precsim \mathsf{cost}(U) \precsim \sum_{i=k}^{n-1} \mathrm{Exp}(\alpha i).$$

From this, we can immediately derive bounds for the expected value of the k-median returned by the trivial heuristic.

Lemma 20. *Fix $U \subseteq V$ of size k. Then, we have $\mathbb{E}[\mathsf{TR}] = \mathbb{E}[\mathsf{cost}(U)]$ and*

$$\frac{1}{\beta}\left(\ln\left(\frac{n-1}{k-1}\right)-1\right) \le \mathbb{E}[\mathsf{TR}] \le \frac{1}{\alpha}\left(\ln\left(\frac{n-1}{k-1}\right)+1\right).$$

Before we provide our result for the expected approximation ratio of the trivial heuristic, we first provide some tail bounds for the distribution of the optimal k-median ME and the trivial solution TR.

Lemma 21. *Fix $U \subseteq V$ of size k. Then the probability density function f of $\sum_{i=k}^{n-1} \mathrm{Exp}(\beta i)$ is given by*

$$f(x) = \beta k \cdot \binom{n-1}{k} \cdot \exp(-\beta k x) \cdot (1-\exp(-\beta x))^{n-k-1}.$$

Lemma 22. *Let $c > 0$ be sufficiently large and let $k \le c'n$ for $c' = c'(c) > 0$ sufficiently small. Then we have*

$$\mathbb{P}\left(\mathsf{ME} \le \left(\ln\left(\tfrac{n-1}{k}\right) - \ln\ln\left(\tfrac{n}{k}\right) - \ln(c)\right)/\beta\right) = n^{-\Omega(c)}.$$

Lemma 23. *Let $k \le (1-\varepsilon)n$ for some constant $\varepsilon > 0$. For every $c \in [0, 2\varepsilon^2)$, we have*

$$\mathbb{P}\left(\mathsf{ME} \le c/\beta\right) \le c^{\Omega(n)}.$$

Lemma 24. *For any $c \ge 4$ we have $\mathbb{P}\left(\mathsf{TR} > n^c\right) \le \exp(-n^{c/4})$.*

Now we have obtained everything needed to provide an upper bound for the expected approximation ratio of the trivial heuristic.

Theorem 25. *Let $k \le (1-\varepsilon)n$ for some constant $\varepsilon > 0$. For generalized random shortest path metrics, we have $\mathbb{E}\left[\frac{\mathsf{TR}}{\mathsf{ME}}\right] = O\left(\beta/\alpha\right)$. Moreover, if we have $k \le c'n$ for some fixed $c' \in (0,1)$ sufficiently small, then we have*

$$\mathbb{E}\left[\tfrac{\mathsf{TR}}{\mathsf{ME}}\right] = (\beta/\alpha) \cdot \left(1 + O\left(\tfrac{\ln\ln(n/k)}{\ln(n/k)}\right)\right).$$

5 Application to the Erdős-Rényi Random Graph Model

So far, we have analyzed random shortest path metrics applied to graphs based on their cut parameters (Definition 1). In this section, we first use a well-known result to show that instances of the Erdős–Rényi random graph model have 'nice' cut parameters with high probability. We then use this to prove our main results.

Lemma 26. *Let $G = (V, E)$ be an instance of the $G(n, p)$ model. For constant $\varepsilon \in (0,1)$ and for any $p \ge c\ln(n)/n$ (as $n \to \infty$), in which $c > 9/\varepsilon^2$ is constant, the cut parameters of G are bounded by $(1-\varepsilon)p \le \alpha \le \beta \le (1+\varepsilon)p$ with probability at least $1 - o\left(1/n^2\right)$.*

Recall that from the result of Corollary 4 we could derive (approximate) bounds for the expected distance $\mathbb{E}[d(u,v)]$ between two arbitrary vertices in a random shortest path metric. Combining this with the result of the foregoing lemma, we can see that, for the case of the application to the Erdős–Rényi random graph model, w.h.p. over the random graph $\mathbb{E}[d(u,v)]$ is approximately bounded between $\ln(n)/((1+\varepsilon)np)$ and $\ln(n)/((1-\varepsilon)np)$ for any constant $\varepsilon \in (0,1)$. This is in line with the known result $\mathbb{E}[d(u,v)] \approx \ln(n)/np$ for p sufficiently large [2].

5.1 Performance of Heuristics

In this section, we provide the main results of this work. We use the results from Sect. 4 and Lemma 26 to analyze the performance of several heuristics in random shortest path metrics applied to Erdős–Rényi random graphs.

When a graph $G=(V,E)$ is created by the $G(n,p)$ model, there is a non-zero probability of G being disconnected. In a corresponding random shortest path metric this results in $d(u,v)=\infty$ for any two vertices $u,v \in V$ that are in different components of G. Observe that, if this is the case, then the identity of indiscernibles, symmetry and triangle inequality still hold. Thus we still have a metric and we can bound the expected approximation ratio for such graphs from above by the worst-case approximation ratio for metric instances.

Using this observation, we can prove the following results.

Theorem 27. *Let $\varepsilon \in (0,1)$ be constant. Let $G=(V,E)$ be a random instance of the $G(n,p)$ model, for p sufficiently large ($p \geq c\ln(n)/n$ as $n \to \infty$ for a constant $c > 9/\varepsilon^2$ satisfies), and consider the corresponding random shortest path metric. Then, we have*

$$\mathbb{E}\left[\frac{\mathsf{GR}}{\mathsf{MM}}\right] = O(1).$$

Theorem 28. *Let $\varepsilon \in (0,1)$ be constant. Let $G=(V,E)$ be a random instance of the $G(n,p)$ model, for p sufficiently large ($p \geq c\ln(n)/n$ as $n \to \infty$ for a constant $c > 9/\varepsilon^2$ satisfies), and consider the corresponding random shortest path metric. Then, we have*

$$\mathbb{E}\left[\frac{\mathsf{NN}}{\mathsf{TSP}}\right] = O(1) \quad \text{and} \quad \mathbb{E}\left[\frac{\mathsf{IN}_R}{\mathsf{TSP}}\right] = O(1).$$

For the last two results, we need the assumption that G is connected.

Theorem 29. *Let $\varepsilon \in (0,1)$ be constant. Let $G=(V,E)$ be a random instance of the $G(n,p)$ model, for p sufficiently large ($p \geq c\ln(n)/n$ as $n \to \infty$ for a constant $c > 9/\varepsilon^2$ satisfies), and consider the corresponding random shortest path metric. If G is connected, then the expected number of iterations of the 2-opt heuristic for TSP is bounded by $O(n^8\ln^3(n))$.*

Theorem 30. *Let $\tilde{\varepsilon} \in (0,1)$ be constant. Let $G = (V, E)$ be a random instance of the $G(n,p)$ model, for p sufficiently large ($p \geq c\ln(n)/n$ as $n \to \infty$ for a constant $c > 9/\tilde{\varepsilon}^2$ satisfies), and consider the corresponding random shortest path metric. Let $\mathcal{E}'$ denotes the event that G is connected. Let $k \leq (1-\varepsilon')n$ for some constant $\varepsilon' > 0$, then we have $\mathbb{E}\left[\frac{\mathsf{TR}}{\mathsf{ME}} \mid \mathcal{E}'\right] = O(1)$. Moreover, if we have $k \leq c'n$ for $c' \in (0,1)$ sufficiently small, then $\mathbb{E}\left[\frac{\mathsf{TR}}{\mathsf{ME}} \mid \mathcal{E}'\right] = 1 + \varepsilon + o(1)$.*

6 Concluding Remarks

We have analyzed heuristics for matching, TSP, and k-median on random shortest path metrics on Erdős–Rényi random graphs. However, in particular for constant values of p, these graphs are still dense. Although our results hold for decreasing $p = \Omega(\ln n/n)$, we obtain in this way metrics with unbounded doubling dimension. In order to get an even more realistic model for random metric spaces, it would be desirable to analyze heuristics on random shortest path metrics on sparse graphs. Hence, we raise the question to generalize our findings to sparse random graphs or sparse (deterministic) classes of graphs.

References

1. Ahn, S., Cooper, C., Cornuéjols, G., Frieze, A.: Probabilistic analysis of a relaxation for the k-median problem. Math. Oper. Res. **13**(1), 1–31 (1988). https://doi.org/10.1287/moor.13.1.1
2. Bhamidi, S., van der Hofstad, R., Hooghiemstra, G.: First passage percolation on the Erdős-Rényi random graph. Comb. Probab. Comput. **20**(5), 683–707 (2011). https://doi.org/10.1017/S096354831100023X
3. Bringmann, K., Engels, C., Manthey, B., Rao, B.V.R.: Random shortest paths: non-Euclidean instances for metric optimization problems. Algorithmica **73**(1), 42–62 (2015). https://doi.org/10.1007/s00453-014-9901-9
4. Byrka, J., Pensyl, T., Rybicki, B., Srinivasan, A., Trinh, K.: An improved approximation for k-median, and positive correlation in budgeted optimization. In: Indyk, P. (ed.) Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2015), pp. 737–756 (2015). https://doi.org/10.1137/1.9781611973730.50
5. Davis, R., Prieditis, A.: The expected length of a shortest path. Inf. Process. Lett. **46**(3), 135–141 (1993). https://doi.org/10.1016/0020-0190(93)90059-I
6. Frieze, A.M., Yukich, J.E.: Probabilistic analysis of the TSP (Chap. 7). In: Gutin, G., Punnen, A.P. (eds.) The Traveling Salesman Problem and Its Variations, pp. 257–307. Springer, Boston (2007). https://doi.org/10.1007/0-306-48213-4_7
7. Hammersley, J.M., Welsh, D.J.A.: First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. In: Neyman, J., Le Cam, L.M. (eds.) Bernoulli 1713 Bayes 1763 Laplace 1813, pp. 61–110. Springer, Heidelberg (1965). https://doi.org/10.1007/978-3-642-49750-6_7
8. Hassin, R., Zemel, E.: On shortest paths in graphs with random weights. Math. Oper. Res. **10**(4), 557–564 (1985). https://doi.org/10.1287/moor.10.4.557
9. Howard, C.D.: Models of first-passage percolation. In: Kesten, H. (ed.) Probability on Discrete Structures, pp. 125–173. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-662-09444-0_3

10. Janson, S.: One, two and three times log n/n for paths in a complete graph with random weights. Comb. Probab. Comput. **8**(4), 347–361 (1999). https://doi.org/10.1017/S0963548399003892
11. Janson, S.: Tail bounds for sums of geometric and exponential variables. Stat. Probab. Lett. **135**, 1–6 (2018). https://doi.org/10.1016/j.spl.2017.11.017
12. Karp, R.M., Steele, J.M.: Probabilistic analysis of heuristics. In: Lawler, E.L., Lenstra, J.K., Rinnooy Kan, A.H.G., Shmoys, D.B. (eds.) The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization, pp. 181–205. Wiley, Hoboken (1985)
13. Reingold, E.M., Tarjan, R.E.: On a greedy heuristic for complete matching. SIAM J. Comput. **10**(4), 676–681 (1981). https://doi.org/10.1137/0210050
14. Rosenkrantz, D.J., Stearns, R.E., Lewis II, P.M.: An analysis of several heuristics for the traveling salesman problem. SIAM J. Comput. **6**(3), 563–581 (1977). https://doi.org/10.1137/0206041
15. Ross, S.M.: Introduction to Probability Models, 10th edn. Academic Press, Burlington (2010)