Random Shortest Paths: Non-Euclidean Instances for Metric Optimization Problems*

Karl Bringmann†[1], Christian Engels[2], Bodo Manthey[3], and B. V. Raghavendra Rao[4]

[1]Max Planck Institute for Informatics, Saarbrücken, Germany, kbringma@mpi-inf.mpg.de
[2]Saarland University, Saarbrücken, Germany, engels@cs.uni-saarland.de
[3]University of Twente, Enschede, Netherlands, b.manthey@utwente.nl
[4]Indian Institute of Technology Madras, Chennai, India, bvrr@cse.iitm.ac.in

May 16, 2014

Probabilistic analysis for metric optimization problems has mostly been conducted on random Euclidean instances, but little is known about metric instances drawn from distributions other than the Euclidean. This motivates our study of random metric instances for optimization problems obtained as follows: Every edge of a complete graph gets a weight drawn independently at random. The distance between two nodes is then the length of a shortest path (with respect to the weights drawn) that connects these nodes.

We prove structural properties of the random shortest path metrics generated in this way. Our main structural contribution is the construction of a good clustering. Then we apply these findings to analyze the approximation ratios of heuristics for matching, the traveling salesman problem (TSP), and the k-median problem, as well as the running-time of the 2-opt heuristic for the TSP. The bounds that we obtain are considerably better than the respective worst-case bounds. This suggests that random shortest path metrics are easy instances, similar to random Euclidean instances, albeit for completely different structural reasons.

1 Introduction

For large-scale optimization problems, finding optimal solutions within reasonable time is often impossible, because many such problems, like the traveling salesman problem (TSP), are NP-hard. Nevertheless, we often observe that simple heuristics succeed surprisingly quickly in finding close-to-optimal solutions. Many such heuristics perform well in practice but have a

*An extended abstract of this work has appeared in the *Proceedings of the 38th Int. Symp. on Mathematical Foundations of Computer Science (MFCS 2013)*.

†Karl Bringmann is a recipient of the *Google Europe Fellowship in Randomized Algorithms*, and this research is supported in part by this Google Fellowship.

poor worst-case performance. In order to explain the performance of such heuristics, probabilistic analysis has proved to be a useful alternative to worst-case analysis. Probabilistic analysis of optimization problems has been conducted with respect to arbitrary instances (without the triangle inequality) [19, 26] or instances embedded in Euclidean space. In particular, the limiting behavior of various heuristics for many of the Euclidean optimization problems is known precisely [40].

However, the average-case performance of heuristics for general metric instances is not well understood. This lack of understanding can be explained by two reasons: First, independent random edge lengths (without the triangle inequality) and random geometric instances are relatively easy to handle from a technical point of view – the former because of the independence of the lengths, the latter because Euclidean space provides a structure that can be exploited. Second, analyzing heuristics on random metric spaces requires an understanding of random metric spaces in the first place. While Vershik [39] gave an analysis of a process for obtaining random metric spaces, using this directly to analyze algorithms seems difficult.

In order to initiate systematic research of heuristics on general metric spaces, we use the following model, proposed by Karp and Steele [27, Section 3.4]: given an undirected complete graph, we draw edge weights independently at random according to exponential distributions with parameter one. The distance between any two vertices is the total weight of the shortest path between them, measured with respect to the random weights. We call such instances *random shortest path metrics*.

This model is also known as *first-passage percolation*, and has been introduced by Broadbent and Hammersley as a model for passage of fluid in a porous medium [10, 11]. More recently, it has also been used to model shortest paths in networks such as the Internet [16]. The appealing feature of random shortest path metrics is their simplicity, which enables us to use them for the analysis of heuristics.

1.1 Known and Related Results

There has been significant study of random shortest path metrics or first-passage percolation. The expected length of an edge is known to be $\ln n/n$ [13, 24]. Asymptotically the same bound holds also for the longest edge almost surely [21, 24]. These results hold not only for the exponential distribution, but for every distribution with distribution function F satisfying $F(x) = x + o(x)$ for small values of x [24]. (See also Section 6.) This model has been used to analyze algorithms for computing shortest paths [20, 21, 34]. Kulkarni and Adlakha have developed algorithmic methods to compute distribution and moments of several optimization problems [30–32]. Beyond shortest path algorithms, random shortest path metrics have been applied only rarely to analyze algorithms. Dyer and Frieze [15], answering a question raised by Karp and Steele [27, Section 3.4], analyzed the patching heuristic for the asymmetric TSP (ATSP) in this model. They showed that it comes within a factor of $1 + o(1)$ of the optimal solution with high probability. Hassin and Zemel [21] applied their findings to the 1-center problem.

From a more structural point of view, first-passage percolation has been analyzed in the area of complex networks, where the hop-count (the number of edges on a shortest path) and the length of shortest path trees have been analyzed [23]. These properties have also been studied on random graphs with random edge weights in various settings [7–9, 22, 29]. Addario-Berry et al. [1] have shown that the number of edges in the longest of the shortest paths is $O(\log n)$ with high probability, and hence the shortest path trees have depth $O(\log n)$.

1.2 Our Results

As far as we are aware, simple heuristics such as greedy heuristics have not been studied in this model yet. Understanding the performance of such algorithms is particularly important as they are easy to implement and used in many applications.

We provide a probabilistic analysis of simple heuristics for optimization under random shortest path metrics. First, we provide structural properties of random shortest path metrics (Section 3). Our most important structural contribution is proving the existence of a good clustering (Lemma 3.9). Then we use these structural insights to analyze simple algorithms for minimum weight matching and the TSP to obtain better expected approximation ratios compared to the worst-case bounds. In particular, we show that the greedy algorithm for minimum-weight perfect matching (Theorem 4.2), the nearest-neighbor heuristic for the TSP (Theorem 4.4), and every insertion heuristic for the TSP (Theorem 4.6) achieve constant expected approximation ratios. We also analyze the 2-opt heuristic for the TSP and show that the expected number of 2-exchanges required before the termination of the algorithm is bounded by $O(n^8 \log^3 n)$ (Theorem 4.7). Investigating further the structural properties of random shortest path metrics, we then consider the k-median problem (Section 5), and show that the most trivial procedure of choosing k arbitrary vertices as k-median yields a $1 + o(1)$ approximation in expectation, provided $k = O(n^{1-\varepsilon})$ for some $\varepsilon > 0$ (Theorem 5.2).

2 Model and Notation

We consider undirected complete graphs $G = (V, E)$ without loops. First, we draw *edge weights* $w(e)$ independently at random according to the exponential distribution[1] with parameter 1.

Second, let the *distances* $d : V \times V \to [0, \infty)$ be given as follows: the distance $d(u, v)$ between u and v is the minimum total weight of a path connecting u and v. In particular, we have $d(v, v) = 0$ for all $v \in V$, $d(u, v) = d(v, u)$ because G is undirected, and the triangle inequality: $d(u, v) \leq d(u, x) + d(x, v)$ for all $u, x, v \in V$. We call the complete graph with distances d obtained from random weights w a *random shortest path metric*.

We use the following notation: Let $\Delta_{\max} = \max_{u,v} d(u, v)$ denote the *diameter* of the random shortest path metric. Let $B_\Delta(v) = \{u \in V \mid d(u, v) \leq \Delta\}$ be the ball of radius Δ around v, i.e., the set of all nodes whose distance to v is at most Δ.

We denote the minimal Δ such that there are at least k nodes within a distance of Δ of v by $\tau_k(v)$. Formally, we define $\tau_k(v) = \min\{\Delta \mid |B_\Delta(v)| \geq k\}$.

By $\mathrm{Exp}(\lambda)$, we denote the exponential distribution with parameter λ. If a random variable X is distributed according to a probability distribution P, we write $X \sim P$. In particular, $X \sim \sum_{i=1}^{m} \mathrm{Exp}(\lambda_i)$ means that X is the sum of m independent exponentially distributed random variables with parameters $\lambda_1, \ldots, \lambda_m$.

By exp, we denote the exponential function. For $n \in \mathbb{N}$, let $[n] = \{1, \ldots, n\}$ and let $H_n = \sum_{i=1}^{n} 1/i$ be the n-th harmonic number.

[1] Exponential distributions are technically the easiest to handle because they are memoryless. We will discuss other distributions in Section 6.

3 Structural Properties of Shortest Path Metrics

3.1 Random Process

To understand random shortest path metrics, it is convenient to fix a starting vertex v and see how the lengths from v to the other vertices develop. In this way, we analyze the distribution of $\tau_k(v)$.

The values $\tau_k(v)$ are generated by a simple birth process as follows. (The same process has been analyzed by Davis and Prieditis [13], Janson [24], and also in subsequent papers.) For $k = 1$, we have $\tau_k(v) = 0$.

For $k \geq 1$, we are looking for the closest vertex to any vertex in $B_{\tau_k(v)}(v)$ in order to obtain $\tau_{k+1}(v)$. This conditions all edges (u, x) with $u \in B_{\tau_k(v)}(v)$ and $x \notin B_{\tau_k(v)}(v)$ to be of length at least $\tau_k(v) - d(v, u)$. The set $B_{\tau_k(v)}(v)$ contains k vertices. Thus, there are $k \cdot (n-k)$ edges to the rest of the graph. Consequently, the difference $\delta_k = \tau_{k+1}(v) - \tau_k(v)$ is distributed as the minimum of $k(n-k)$ exponential random variables (with parameter 1), or, equivalently, as $\operatorname{Exp}(k \cdot (n-k))$. We obtain that $\tau_{k+1}(v) \sim \sum_{i=1}^{k} \operatorname{Exp}\big(i \cdot (n-i)\big)$. Note that these exponential distributions as well as the random variables $\delta_1, \ldots, \delta_n$ are independent.

Exploiting linearity of expectation and that the expected value of $\operatorname{Exp}(\lambda)$ is $1/\lambda$ we obtain the following lemma.

Lemma 3.1. *For any $k \in [n]$ and any $v \in V$, we have $\mathbb{E}\big(\tau_k(v)\big) = \frac{1}{n} \cdot \big(H_{k-1} + H_{n-1} - H_{n-k}\big)$ and $\tau_k(v) \sim \sum_{i=1}^{k-1} \operatorname{Exp}\big(i \cdot (n-i)\big)$.*

Proof. The proof is by induction on k. For $k = 1$, we have $\tau_k(v) = 0$ and $H_{k-1} + H_{n-1} - H_{n-k} = H_0 + H_{n-1} - H_{n-1} = 0$. Now assume that the lemma holds for k for some $k \geq 1$. In the paragraph preceding this lemma we have seen that $\tau_{k+1}(v) - \tau_k(v) \sim \operatorname{Exp}(k(n-k))$. Thus, $\mathbb{E}(\tau_{k+1}(v) - \tau_k(v)) = \frac{1}{k(n-k)}$. Plugging in the induction hypothesis yields

$$\begin{aligned}\mathbb{E}\big(\tau_{k+1}(v)\big) &= \mathbb{E}\big(\tau_k(v)\big) + \frac{1}{k \cdot (n-k)} = \frac{1}{n} \cdot \left(H_{k-1} + H_{n-1} - H_{n-k} + \frac{1}{k} + \frac{1}{n-k}\right) \\ &= \frac{1}{n} \cdot \big(H_k + H_{n-1} - H_{n-(k+1)}\big).\end{aligned}$$

□

From this result, we can easily deduce two known results: averaging over k yields that the expected distance of an edge is $\frac{H_{n-1}}{n-1} \approx \ln n/n$ [13, 24]. The longest distance from v to any other node is $\tau_n(v)$, which is $2H_{n-1}/n \approx 2 \ln n/n$ in expectation [24]. For completeness, let us mention that the diameter $\Delta_{\max}$ is approximately $3 \ln n/n$ [24]. However, this does not follow immediately from Lemma 3.1.

3.2 Distribution of $\tau_k(v)$

Let us now have a closer look at cumulative distribution function of $\tau_k(v)$ for fixed $v \in V$ and $k \in [n]$. To do this, the following lemma is very useful.

Lemma 3.2. *Let $X \sim \sum_{i=1}^{n} \operatorname{Exp}(ci)$. Then $\mathbb{P}(X \leq \alpha) = (1 - e^{-c\alpha})^n$.*

Proof. The random variable X has the same distribution as $\max_{i=1}^{n} Y_i$, where $Y_i \sim \operatorname{Exp}(c)$. We have $X \leq \alpha$ if and only if $Y_i \leq \alpha$ for all $i \in \{1, \ldots, n\}$. □

In the following, let F_k denote the cumulative distribution function of $\tau_k(v)$ for some fixed vertex $v \in V$, i.e., $F_k(x) = \mathbb{P}(\tau_k(v) \leq x)$.

Lemma 3.3. *For every $\Delta \geq 0$, $v \in V$, and $k \in [n]$, we have*

$$\big(1 - \exp(-(n-k)\Delta)\big)^{k-1} \leq F_k(\Delta) \leq \big(1 - \exp(-n\Delta)\big)^{k-1}.$$

Proof. Lemma 3.1 states that $\tau_k(v) \sim \sum_{i=1}^{k-1} \operatorname{Exp}(i(n-i))$. We approximate the parameters by ci for $c \in \{n-k, n\}$. The distribution with $c = n$ is stochastically dominated by the true distribution, which is in turn dominated by the distribution obtained for $c = n - k$. We apply Lemma 3.2 with $c = n$ and $c = n - k$. □

Lemma 3.4. *Fix $\Delta \geq 0$ and a vertex $v \in V$. Then*

$$\big(1 - \exp(-(n-k)\Delta)\big)^{k-1} \leq \mathbb{P}\big(|B_\Delta(v)| \geq k\big) \leq \big(1 - \exp(-n\Delta)\big)^{k-1}.$$

Proof. We have $|B_\Delta(v)| \geq k$ if and only if $\tau_k(v) \leq \Delta$. The lemma follows from Lemma 3.3. □

We can improve Lemma 3.3 slightly in order to obtain even closer upper and lower bounds. For $n, k \geq 2$, combining Lemmas 3.3 and 3.5 yields tight upper and lower bounds if we disregard the constants in the exponent, namely $F_k(\Delta) = \big(1 - \exp(-\Theta(n\Delta))\big)^{\Theta(k)}$.

Lemma 3.5. *For all $v \in V$, $k \in [n]$, and $\Delta \geq 0$, we have*

$$F_k(\Delta) \geq \big(1 - \exp(-(n-1)\Delta/4)\big)^{n-1}$$

and

$$F_k(\Delta) \geq \big(1 - \exp(-(n-1)\Delta/4)\big)^{\frac{4}{3}(k-1)}.$$

Proof. As $\tau_k(v)$ is monotonically increasing in k, we have $F_k(\Delta) \geq F_{k+1}(\Delta)$ for all k. Thus, we have to prove the claim only for $k = n$. In this case, $\tau_n(v) \sim \sum_{i=1}^{n-1} \operatorname{Exp}(\lambda_i)$, with $\lambda_i = i(n-i) = \lambda_{n-i}$. Setting $m = \lceil n/2 \rceil$ and exploiting the symmetry around m yields

$$\tau_n(v) \leq \sum_{i=1}^{m} \operatorname{Exp}(\lambda_i) + \sum_{i=1}^{m} \operatorname{Exp}(\lambda_i) = \tau_m(v) + \tau_m(v).$$

Here, "$\leq$" means stochastic dominance, "$=$" means equal distribution, and "$+$" means adding up two independent random variables. Hence,

$$F_n(\Delta) = \mathbb{P}\big(\tau_n(v) \leq \Delta\big) \geq \mathbb{P}\big(\tau_m(v) + \tau_m(v) \leq \Delta\big) \geq \mathbb{P}\big(\tau_m(v) \leq \Delta/2\big)^2.$$

By Lemma 3.3, and using $m \leq (n+1)/2$, this is bounded by

$$F_n(\Delta) \geq (1 - \exp(-(n-m)\Delta/2))^{2(m-1)} \geq (1 - \exp(-(n-1)\Delta/4))^{n-1}.$$

For the second inequality, we use the first inequality of Lemma 3.5 for $k - 1 \geq \frac{3}{4}(n-1)$ and Lemma 3.3 for $k - 1 < \frac{3}{4}(n-1)$ as then $n - k \geq (n-1)/4$. □

3.3 Tail Bounds for $|B_\Delta(v)|$ and $\Delta_{\max}$

Our first tail bound for $|B_\Delta(v)|$, which is the number of vertices within distance Δ of a given vertex v, follows directly from Lemma 3.3. From this lemma we derive the following corollary, which is a crucial ingredient for the existence of good clusterings and, thus, for the analysis of heuristics in the remainder of this paper.

Corollary 3.6. *Let $n \geq 5$ and fix $\Delta \geq 0$ and a vertex $v \in V$. Then we have*

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

Proof. Lemma 3.4 yields

$$\begin{aligned}\mathbb{P}\left(|B_\Delta(v)| < \min\left\{\exp\left(\Delta\frac{n-1}{4}\right), \frac{n+1}{2}\right\}\right) &\leq 1 - \left(1 - \exp\left(-\frac{n-1}{2}\Delta\right)\right)^{\exp(\Delta(n-1)/4)} \\ &\leq \exp\left(-\Delta\frac{n-1}{4}\right),\end{aligned}$$

where the last inequality follows from $(1-x)^y \geq 1-xy$ for $y \geq 1$, $x \geq 0$. Using $(n-1)/4 \geq n/5$ for $n \geq 5$ completes the proof. □

Corollary 3.6 is almost tight according to the following result.

Corollary 3.7. *Fix $\Delta \geq 0$, a vertex $v \in V$, and any $c > 1$. Then*

$$\mathbb{P}\big(|B_\Delta(v)| \geq \exp(c\Delta n)\big) < \exp\big(-(c-1)\Delta n\big).$$

Proof. Lemma 3.4 with $k = c\Delta n$ yields

$$\mathbb{P}\big(|B_\Delta(v)| \geq \exp(c\Delta n)\big) \leq \big(1 - \exp(-n\Delta)\big)^{\exp(c\Delta n)-1}.$$

Using $1 + x \leq e^x$, we get

$$\mathbb{P}\big(|B_\Delta(v)| \geq \exp(c\Delta n)\big) \leq \exp\left(\exp(-n\Delta) - \exp\big((c-1)\cdot\Delta n\big)\right).$$

Now, we bound $\exp(-n\Delta) \leq 1$ and $\exp\big((c-1)\cdot\Delta n\big) \geq 1 + (c-1)\cdot\Delta n$, which yields the claimed inequality. □

Janson [24] derived the following tail bound for the diameter $\Delta_{\max}$. A qualitatively similar bound can be proved using Lemma 3.4 and can also be derived from Hassin and Zemel's analysis [21]. However, Janson's bound is stronger with respect to the constants in the exponent.

Lemma 3.8 (Janson [24, p. 352])**.** *For any fixed $c > 3$, we have $\mathbb{P}(\Delta_{\max} > c\ln(n)/n) \leq O(n^{3-c}\log^2 n)$.*

3.4 Balls and Clusters

In this section, we show our main structural contribution, which is a global property of random shortest path metrics. We show that such instances can be divided into a small number of clusters of any given diameter.

From now on, let $s_\Delta = \min\{\exp(\Delta n/5), (n+1)/2\}$, as in Corollary 3.6. If $|B_\Delta(v)|$, the number of vertices within distance Δ of v, is at least s_Δ, then we call the vertex v a *dense* Δ*-center*, and we call the set $B_\Delta(v)$ of vertices within distance Δ of v (including v itself) the Δ*-ball of* v. Otherwise, if $|B_\Delta(v)| < s_\Delta$, and v is not part of any Δ-ball, we call the vertex v a *sparse* Δ*-center*. Any two vertices in the same Δ-ball have a distance of at most 2Δ because of the triangle inequality.

If Δ is clear from the context, then we also speak about centers and balls without parameter. We can bound, by Corollary 3.6, the expected number of sparse Δ-centers to be at most $O(n/s_\Delta)$.

We want to partition the graph into a small number of clusters, each of diameter at most 6Δ. For this purpose, we put each sparse Δ-center in its own cluster (of size 1). Then the diameter of each such cluster is 0, which is trivially upper-bounded by 6Δ, and the number of these clusters is expected to be at most $O(n/s_\Delta)$.

We are left with the dense Δ-centers, which we cluster using the following algorithm: Consider an auxiliary graph whose vertices are all dense Δ-centers. We draw an edge between two dense Δ-centers u and v if $B_\Delta(u) \cap B_\Delta(v) \neq \emptyset$. Now consider any maximal independent set of this auxiliary graph (for instance, a greedy independent set), and let t be the number of its vertices. Then we form initial clusters $C'_1, \ldots, C'_t$, each containing one of the Δ-balls corresponding to the vertices in the independent set. By the independence, all these t Δ-balls are disjoint, which implies $t \leq n/s_\Delta$. The ball of every remaining center v has at least one vertex in one of the C'_i. We add all remaining vertices of $B_\Delta(v)$ to such a C'_i to form the final clusters $C_1, \ldots, C_t$. By construction, the diameter of each C_i is at most 6Δ: Consider any two vertices $u, v \in C_i$. The distance of u towards its closest neighbor in the initial ball C'_i is at most 2Δ. The same holds for v. Finally, the diameter of the initial ball C'_i is also at most 2Δ.

With this partitioning, we have obtained the following structure: We have an expected number of $O(n/s_\Delta)$ clusters of size 1 and diameter 0, and a number of $O(n/s_\Delta)$ clusters of size at least s_Δ and diameter at most 6Δ. Thus, we have $O(n/s_\Delta) = O(1 + n/\exp(\Delta n/5))$ clusters in total. We summarize these findings in the following lemma. This lemma is the crucial ingredient for bounding the expected approximation ratios of the greedy, nearest-neighbor, and insertion heuristics.

Lemma 3.9. *Consider a random shortest path metric and let $\Delta \geq 0$. If we partition the instance into clusters, each of diameter at most 6Δ, then the expected number of clusters needed is $O(1 + n/\exp(\Delta n/5))$.*

4 Analysis of Heuristics

4.1 Greedy Heuristic for Minimum-Length Perfect Matching

Finding minimum-length perfect matchings in metric instances is the first problem that we consider. This problem has been widely considered in the past and has applications in, e.g., optimizing the speed of mechanical plotters [35, 38]. The worst-case running-time of $O(n^3)$ for finding an optimal matching is prohibitive if the number n of points is large. Thus, simple

heuristics are often used, with the greedy heuristic being probably the simplest one: at every step, choose an edge of minimum length incident to the unmatched vertices and add it to the partial matching. Let GREEDY denote the cost of the matching output by this greedy matching heuristic, and let MM denote the optimum value of the minimum-length perfect matching. The worst-case approximation ratio for greedy matching on metric instances is $\Theta(n^{\log_2(3/2)})$ [35], where $\log_2(3/2) \approx 0.58$. In the case of Euclidean instances, the greedy algorithm has an approximation ratio of $O(1)$ with high probability on random instances [5]. For independent random edge weights (without the triangle inequality), the expected weight of the matching computed by the greedy algorithm is $\Theta(\log n)$ [14] whereas the optimal matching has a weight of $\Theta(1)$ with high probability, which gives an $O(\log n)$ approximation ratio.

We show that greedy matching finds a matching of constant expected length on random shortest path metrics.

Theorem 4.1. $\mathbb{E}(\mathsf{GREEDY}) = O(1)$.

Proof. Let $\Delta_i = \frac{i}{n}$. We divide the run of GREEDY in phases as follows: we say that GREEDY is in phase i if edges $\{u, v\}$ are inserted such that $d(u, v) \in (6\Delta_{i-1}, 6\Delta_i]$. Lemma 3.8 allows to show that the expected sum of all edges longer than $\Delta_{\Omega(\log n)}$ is $o(1)$, so we can ignore them.

GREEDY goes through phases i with increasing i (phases can be empty). We now estimate the contribution of phase i to the matching computed by GREEDY. Using Lemma 3.9, after phase $i-1$ we can find a clustering into clusters of diameter at most $6\Delta_{i-1}$ using an expected number of $O(1+n/e^{(i-1)/5})$ clusters. Each such cluster can have at most one unmatched vertex. Thus, we have to add at most $O(1 + n/e^{(i-1)/5})$ edges in phase i. Each such edge connects vertices at a distance of at most $6\Delta_i$. Hence, the contribution of phase i is $O(\frac{i}{n} \cdot (1+n/e^{(i-1)/5}))$ in expectation. Summing over all phases yields the desired bound:

$$\mathbb{E}\big(\mathsf{GREEDY}\big) = o(1) + \sum_{i=1}^{O(\log n)} O\left(\frac{i}{e^{(i-1)/5}} + \frac{i}{n}\right) = O(1).$$

□

Careful analysis allows us to bound the expected approximation ratio.

Theorem 4.2. *The greedy algorithm for minimum-length perfect matching has constant approximation ratio on random shortest path metrics, i.e.,* $\mathbb{E}\left(\frac{\mathsf{GREEDY}}{\mathsf{MM}}\right) = O(1)$.

We will use the following tail bound to estimate the approximation ratios of the greedy heuristic for matching as well as the nearest-neighbor and insertion heuristics for the TSP.

Lemma 4.3. *Let $\alpha \in [0, 1]$. Let S_m be the sum of the lightest m edge weights, where $m \geq \alpha n$. Then, for all $c \in [0, 1]$, we have*

$$\mathbb{P}(S_m \leq c) \leq \left(\frac{e^2 c}{2\alpha^2}\right)^{\alpha n}.$$

Furthermore, $\mathsf{TSP} \geq \mathsf{MM} \geq S_{n/2}$, *where* TSP *and* MM *denote the length of the shortest TSP tour and the minimum-weight perfect matching, respectively, in the corresponding shortest path metric.*

Proof. Let $X \sim \sum_{i=1}^{m} \mathrm{Exp}(1)$, and let Y be the sum of m independent random variables drawn uniformly from $[0,1]$. The random variable X stochastically dominates Y, and $\mathbb{P}(Y \le c) = c^m/m!$.

The probability that $S_m \le c$ is at most the probability that there exists a subset of the edges of cardinality m whose total weight is at most c. By a union bound and using $\binom{a}{b} \le (ae/b)^b$, $\binom{n}{2} \le n^2/2$, and $a! > (a/e)^a$, we obtain

$$\mathbb{P}(S_m \le c) \le \binom{\binom{n}{2}}{m} \cdot \frac{c^m}{m!} \le \left(\frac{n^2 e^2 c}{2m^2}\right)^m \le \left(\frac{e^2 c}{2\alpha^2}\right)^m.$$

We can replace m by its lower bound αn in the exponent [2, Fact 2.1] to obtain the first claim.

It remains to prove $\mathsf{TSP} \ge \mathsf{MM} \ge S_{n/2}$. The first inequality is trivial. For the second inequality, consider a minimum-weight perfect matching in a random shortest path metric. We replace every edge by the corresponding paths. If we disregard multiple edges, then we are still left with at least $n/2$ edges whose length is not shortened by taking shortest paths. The sum of the weights of these $n/2$ edges is at most MM and at least $S_{n/2}$. □

Proof of Theorem 4.2. The worst-case approximation ratio of GREEDY for minimum-weight perfect matching is $n^{\log_2(3/2)}$ [35]. Let $c > 0$ be a sufficiently small constant. Then the approximation ratio of GREEDY on random shortest path instances is

$$\mathbb{E}\left(\frac{\mathsf{GREEDY}}{\mathsf{MM}}\right) \le \mathbb{E}\left(\frac{\mathsf{GREEDY}}{c}\right) + \mathbb{P}(\mathsf{MM} < c) \cdot n^{\log_2(3/2)}.$$

By Theorem 4.1, the first term is $O(1)$. Since c is sufficiently small, Lemma 4.3 shows that the second term is $o(1)$. □

4.2 Nearest-Neighbor algorithm for the TSP

A greedy analogue for the traveling salesman problem (TSP) is the *nearest neighbor* heuristic: (1) Start with some starting vertex v_0 as the current vertex v. (2) At every iteration, choose the nearest yet unvisited neighbor u of the current vertex v (called the successor of v) as the next vertex in the tour, and move to the next iteration with the new vertex u as the current vertex v. (3) Go back to the first vertex v_0 if all vertices are visited. Let NN denote both the nearest-neighbor heuristic itself and the cost of the tour computed by it. Let TSP denote the cost of an optimal tour. The nearest-neighbor heuristic NN achieves a worst-case ratio of $O(\log n)$ for metric instances and also an average-case ratio (for independent, non-metric edge lengths) of $O(\log n)$ [4]. We show that NN achieves a constant approximation ratio on random shortest path instances.

Theorem 4.4. *For random shortest path instances we have* $\mathbb{E}(\mathsf{NN}) = O(1)$ *and* $\mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\right) = O(1)$.

Proof. The proof is similar to the proof of Theorem 4.2. Let $\Delta_i = i/n$ for $i \in \mathbb{N}$. Let $Q = O(\log n/n)$ be sufficiently large.

Consider the clusters obtained with parameter Δ_i as in the discussion preceding Lemma 3.9. These clusters have diameters of at most $6\Delta_i$. We refer to these clusters as the *i-clusters*. Let v be any vertex. We call v *bad at i*, if v is in some i-cluster and NN chooses a vertex at a distance of more than $6\Delta_i$ from v for leaving v. Hence, if v is bad at i, then the next vertex

lies outside of the cluster to which v belongs. (Note that v is not bad at i if the outgoing edge at v leads to a neighbor outside of the cluster of v but at a distance of at most $6\Delta_i$ from v.)

In the following, let the cost of a vertex v be the distance from v to its successor u. The length of the tour produced by NN is equal to the sum of costs over all vertices.

Claim 4.5. *The expected number of vertices with costs in the range* $(6\Delta_i, 6\Delta_{i+1}]$ *is at most* $O(1 + n/\exp(i/5))$.

Proof of Claim 4.5. Suppose that the cost of the neighbor chosen by NN for a vertex v is in the interval $(6\Delta_i, 6\Delta_{i+1}]$. Then v is bad at i. This happens only if all other vertices of the i-cluster containing v have already been visited. Otherwise, there would be another vertex u in the same i-cluster with a distance of at most $6\Delta_i$ to v. By Lemma 3.9, the number of i-clusters is at most $O(1 + n/\exp(i/5))$. □

If $\Delta_{\max} \leq Q$, then it suffices to consider i for $i \leq O(\log n)$. If $\Delta_{\max} > Q$, then we bound the value of the tour produced by NN by $n\Delta_{\max}$. This failure event, however, contributes only $o(1)$ to the expected value by Lemma 3.8. For the case $\Delta_{\max} \leq Q$, the contribution to the expected length of the NN tour is bounded from above by

$$\sum_{i=0}^{O(\log n)} 6\Delta_{i+1} \cdot O\left(1 + \frac{n}{\exp(i/5)}\right) = \sum_{i=0}^{O(\log n)} O\left(\frac{i+1}{n} + \frac{i+1}{\exp(i/5)}\right) = O(1).$$

Using the fact that the worst-case approximation ratio of NN is $O(\log n)$, the proof of the constant expected approximation ratio is similar to the proof of Theorem 4.2. □

4.3 Insertion Heuristics

An insertion heuristic for the TSP is an algorithm that starts with an initial tour on a few vertices and extends this tour iteratively by adding the remaining vertices. In every iteration, a vertex is chosen according to some rule, and this vertex is inserted at the place in the current tour where it increases the total tour length the least. The approximation ratio achieved depends on the rule used for selecting the next node to insert. Certain insertion heuristics such as nearest neighbor insertion (which is different from the nearest neighbor algorithm from the previous section) achieve constant approximation ratios [36]. The random insertion algorithm, where the next vertex is chosen uniformly at random from the remaining vertices, has a worst-case approximation ratio of $\Omega(\log\log n/\log\log\log n)$, and there are insertion heuristics with a worst-case approximation ratio of $\Omega(\log n/\log\log n)$ [6].

A rule R that specifies an insertion heuristic can be viewed as follows: depending on the distances d, it (1) chooses a set R_V of vertices for computing an initial tour and (2) given any tour of vertices $V' \supseteq R_V$, describes how to choose the next vertex. Let INSERT_R denote the length of the tour produced with rule R.

For random shortest path metrics, we show that any insertion heuristic produces a tour whose length is expected to be within a constant factor of the optimal tour. This result holds irrespective of which insertion strategy we actually use.

Theorem 4.6. *For every rule* R*, we have* $\mathbb{E}(\mathsf{INSERT}_R) = O(1)$ *and* $\mathbb{E}\left(\frac{\mathsf{INSERT}_R}{\mathsf{TSP}}\right) = O(1)$.

Proof. Let $\Delta_i = i/n$ for $i \in \mathbb{N}$ and $Q = O(\log n/n)$ be sufficiently large. Assume that $\Delta_{\max} \leq Q$. If $\Delta_{\max} > Q$, then we bound the length of the tour produced by $n \cdot \Delta_{\max}$. This contributes only $o(1)$ to the expected value of length of the tour produced by Lemma 3.8.

Suppose we have a partial tour T and v is the vertex that we have to insert next. If T has a vertex u such that v and u are in a common i-cluster, then the triangle inequality implies that the costs of inserting v into T is at most $12\Delta_i$ because the diameters of i-clusters are at most $6\Delta_i$ [36, Lemma 2]. For each i, only the insertion of the first vertex of each i-cluster can possibly cost more than $12\Delta_i$. Thus, the number of vertices whose insertion would incur costs in the range $(12\Delta_i, 12\Delta_{i+1}]$ is at most $O\big(1 + \frac{n}{\exp(i/5)}\big)$ in expectation. Note that we only have to consider i with $i \leq O(\log n)$ since $\Delta_{\max} \leq Q$. The expected costs of the initial tour are at most $\mathsf{TSP} = O(1)$ [19]. Summing up the expected costs for all i plus the costs of the initial tour, we obtain that the expected costs of the tour obtained by an insertion heuristic is bounded from above by

$$\mathbb{E}(\mathsf{INSERT}_R) = O(1) + \sum_{i=0}^{O(\log n)} \Delta_i \cdot O\left(1 + \frac{n}{\exp(i/5)}\right) = O(1).$$

Note that the above argument is independent of the rule R used.

The proof for the approximation ratio is similar to the proof of Theorem 4.2 and uses the worst-case ratio of $O(\log n)$ for insertion heuristics for any rule R [36, Theorem 3]. □

4.4 Running-Time of 2-Opt for the TSP

The 2-opt heuristic for the TSP starts with an initial tour and successively improves the tour by so-called 2-exchanges until no further refinement is possible. In a 2-exchange, a pair of edges $e_{12} = \{v_1, v_2\}$ and $e_{34} = \{v_3, v_4\}$, where v_1, v_2, v_3, v_4 appear in this order in the Hamiltonian tour, are replaced by a pair of edges $e_{13} = \{v_1, v_3\}$ and $e_{24} = \{v_2, v_4\}$ to get a shorter tour. The 2-opt heuristic is easy to implement and widely used. In practice, it usually converges quite quickly to close-to-optimal solutions [25]. To explain its performance in practice, probabilistic analyses of its running-time on geometric instances [18, 28, 33] and its approximation performance on geometric instances [18] and with independent, non-metric edge lengths [17] have been conducted. We prove that for random shortest path metrics, the expected number of iterations that 2-opt needs is bounded by a polynomial.

Theorem 4.7. *The expected number of iterations that 2-opt needs to find a local optimum is bounded by $O(n^8 \log^3 n)$.*

Proof. The proof is similar to the analysis of 2-opt by Englert et al. [18]. Consider a 2-exchange where edges e_1 and e_2 are replaced by edges f_1 and f_2 as described above. The improvement obtained from this exchange is given by $\delta = \delta(v_1, v_2, v_3, v_4) = d(v_1, v_2) + d(v_3, v_4) - d(v_1, v_3) - d(v_2, v_4)$.

We estimate the probability $\mathbb{P}(\delta \in (0, \varepsilon])$ of the event that the improvement is at most ε for some $\varepsilon > 0$. The distances $d(v_i, v_j)$ correspond to shortest paths with respect to the exponentially distributed edge weights w. Assume for the moment that we know these paths. Then we can rewrite the improvement as

$$\delta = \sum_{e \in E} \alpha_e w(e) \tag{1}$$

for some coefficients $\alpha_e \in \{-2,-1,0,1,2\}$. If the exchange considered is indeed a 2-exchange, then $\delta > 0$. Thus, in this case, there exists at least on edge $e = \{u, u'\}$ with $\alpha_e \neq 0$. Let $I \subseteq \{e_{12}, e_{34}, e_{13}, e_{24}\}$ be the set of edges of the 2-exchange such that the corresponding paths use e.

For all combinations of I and e, let $\delta_{ij}^{I,e}$ be the following quantity:

- If $e_{ij} \notin I$, then $\delta_{ij}^{I,e}$ is the length of a shortest path from v_i to v_j without using e.
- If $e_{ij} \in I$, then $\delta_{ij}^{I,e}$ is the minimum of
  - the length of a shortest path from v_i to u without e plus the length of a shortest path from u' to v_j without e and
  - the length of a shortest path from v_i to u' without e plus the length of a shortest path from u to v_j without e.

Let $\delta^{e,I} = \delta_{12}^{e,I} + \delta_{34}^{e,I} - \delta_{13}^{e,I} - \delta_{24}^{e,I}$.

Claim 4.8. *For every outcome of the random edge weights, there exists an edge e and a set I such that $\delta = \delta^{e,I} + \alpha w(e)$, where $\alpha \in \{-2,-1,1,2\}$ is determined by e and I.*

Proof of Claim 4.8. Fix the edge weights arbitrarily and consider any four shortest paths. Then there exists some edge e with non-zero α_e in (1). We choose this e, an appropriate set I, and we choose $\alpha = \alpha_e$. Then the claim follows from the definition of $\delta^{e,I}$. □

Claim 4.8 yields that $\delta \in (0, \varepsilon]$ implies that there are an e and an I with $\delta^{e,I} + \alpha w(e) \in (0, \varepsilon]$.

Claim 4.9. *Let e and I be arbitrary with $\alpha = \alpha_e \neq 0$. Then $\mathbb{P}(\delta^{e,I} + \alpha w(e) \in (0, \varepsilon]) \leq \varepsilon$.*

Proof of Claim 4.9. We fix the edge weights of all edges except for e. This determines $\delta^{e,I}$. Thus, $\delta^{e,I} + \alpha w(e) \in (0, \varepsilon]$ if and only of $w(e)$ assumes a value in a now fixed interval of size $\varepsilon/\alpha \leq \varepsilon$. Since the density of the exponential distribution is bounded from above by 1, the claim follows. □

The number of possible choices for e and I is $O(n^2)$. Thus, $\mathbb{P}(\delta \in (0, \varepsilon]) = O(n^2\varepsilon)$.

Let $\delta_{\min} > 0$ be the minimum improvement made by any 2-exchange. Since there are at most n^4 different 2-exchanges, we have $\mathbb{P}(\delta_{\min} \leq \varepsilon) = O(n^6\varepsilon)$.

The initial tour has a length of at most $n\Delta_{\max}$. Let T be the number of iterations that 2-opt takes. Then $T \leq n\Delta_{\max}/\delta_{\min}$. Now, $T > x$ implies $\Delta_{\max}/\delta_{\min} > x/n$. The event $\Delta_{\max}/\delta_{\min} > x/n$ is contained in the union of the events $\Delta_{\max} > \log x \ln n/n$, and $\delta_{\min} < \ln n \cdot \log x/x$. The first happens with a probability of at most $n^{-\Omega(\log(x))}$ by Lemma 3.8. The second happens with a probability of at most $O(n^6 \log(x)/x)$. Thus, we obtain

$$\mathbb{P}(T > x) \leq n^{-\Omega(\log(x))} + O\big(n^6 \ln n \cdot \log(x)/x\big).$$

Since the number of iterations is at most $n!$, we obtain an upper bound of

$$\mathbb{E}(T) \leq \sum_{x=1}^{n!} \left(n^{-\Omega(\log(x))} + O(n^6 \ln n \log(x)/x)\right).$$

The sum of the $n^{-\Omega(\log(x))}$ is negligible. The sum of the $O(n^6 \ln n \log(x)/x)$ contributes $O(n^6 \ln n \log(n!)^2) = O(n^8 \log^3 n)$. □

5 k-Median

In the (metric) k-median problem, we are given a finite metric space (V, d) and should pick k points $U \subseteq V$ such that $\sum_{v \in V} \min_{u \in U} d(v, u)$ is minimized. We call the set U a k-median. Regarding worst-case analysis, the best known approximation algorithm for this problem achieves an approximation ratio of $3 + \varepsilon$ [3].

In this section, we consider the k-median problem in the setting of random shortest path metrics. In particular we examine the approximation ratio of the algorithm TRIVIAL, which picks k points independently of the metric space, e.g., $U = \{1, \dots, k\}$ or k random points in V. We show that TRIVIAL yields a $(1 + o(1))$-approximation for $k = O(n^{1-\varepsilon})$. This can be seen as an algorithmic result since it improves upon the worst-case approximation ratio, but it is essentially a structural result on random shortest path metrics. It means that any set of k points is, with high probability, a very good k-median, which gives some knowledge about the topology of random shortest path metrics. For larger, but not too large k, i.e., $k \leq (1 - \varepsilon)n$, TRIVIAL still yields an $O(1)$-approximation.

The main insight comes from generalizing the growth process described in Section 3.2. Fixing $U = \{v_1, \dots, v_k\} \subseteq V$ we sort the vertices $V \setminus U$ by their distance to U in ascending order, calling the resulting order $v_{k+1}, \dots, v_n$. Now we consider $\delta_i = d(v_{i+1}, U) - d(v_i, U)$ for $k \leq i < n$. These random variables are generated by a simple growth process analogous to the one described in Section 3.2. This shows that the δ_i are independent and $\delta_i \sim \operatorname{Exp}(i \cdot (n - i))$. Since $a \operatorname{Exp}(b) \sim \operatorname{Exp}(b/a)$, we have

$$\mathsf{cost}(U) = \sum_{i=k}^{n-1} (n - i) \cdot \delta_i \sim \sum_{i=k}^{n-1} (n - i) \cdot \operatorname{Exp}(i \cdot (n - i)) \sim \sum_{i=k}^{n-1} \operatorname{Exp}(i).$$

From this, we can read off the expected cost of U immediately, and thus the expected cost of TRIVIAL.

Lemma 5.1. *Fix $U \subseteq V$ of size k. We have*

$$\mathbb{E}(\mathsf{TRIVIAL}) = \mathbb{E}\big(\mathsf{cost}(U)\big) = H_{n-1} - H_{k-1} = \ln(n/k) + \Theta(1).$$

Proof. We have $\mathbb{E}(\mathsf{cost}(U)) = \sum_{i=k}^{n-1} \frac{1}{i} = H_{n-1} - H_{k-1}$. Using $H_n = \ln(n) + \Theta(1)$ yields the last equality. □

By closely examining the random variable $\sum_{i=k}^{n-1} \operatorname{Exp}(i)$, we can show good tail bounds for the probability that the cost of U is lower than expected. Together with the union bound this yields tail bounds for the optimal k-median MEDIAN, which implies the following theorem. In this theorem, the approximation ratio becomes $1 + O\big(\frac{\ln \ln(n)}{\ln(n)}\big)$ for $k = O(n^{1-\varepsilon})$.

Theorem 5.2. *Let $k \leq (1 - \varepsilon)n$ for some constant $\varepsilon > 0$. Then*

$$\mathbb{E}\left(\frac{\mathsf{TRIVIAL}}{\mathsf{MEDIAN}}\right) = O(1).$$

If we have $k \leq \kappa n$ for some sufficiently small constant $\kappa \in (0, 1)$, then

$$\mathbb{E}\left(\frac{\mathsf{TRIVIAL}}{\mathsf{MEDIAN}}\right) = 1 + O\left(\frac{\ln \ln(n/k)}{\ln(n/k)}\right). \tag{2}$$

We need the following lemmas to prove Theorem 5.2.

Lemma 5.3. *The density f of $\sum_{i=k}^{m} \mathrm{Exp}(i)$ is given by*

$$f(x) = k \cdot \binom{m}{k} \cdot \exp(-kx) \cdot \big(1 - \exp(-x)\big)^{m-k}.$$

Proof. The distribution $\sum_{i=k}^{m} \mathrm{Exp}(i)$ corresponds to the k-th largest element of a set of m independent, exponentially distributed random variables with parameter 1. The density of such order statistics is known [37, Example 2.38]. □

Lemma 5.4. *Let $c > 0$ be sufficiently large, and let $k \le c'n$ for $c' = c'(c) > 0$ be sufficiently small. Then*

$$\mathbb{P}\left(\mathsf{MEDIAN} < \ln\left(\frac{n}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln c\right) = n^{-\Omega(c)}.$$

Proof. Fix $U \subseteq V$ of size k and consider $\mathsf{cost}(U) \sim \sum_{i=k}^{n-1} \mathrm{Exp}(i)$. In the following we set $m := n - 1$ to shorten notation. We now want to bound $f(x)$ from above at $x = \ln\big(\frac{m}{ak}\big)$ for a sufficiently large a with $1 \le a \le m/k$ (such an a exists since k is small enough). Plugging in this particular x and using $\binom{m}{k} \le m^k e^k / k^k$ yields

$$f(x) = k \cdot \binom{m}{k} \cdot \frac{a^k k^k (m - ak)^{m-k}}{m^m} \le k(ea)^k \left(1 - \frac{ak}{m}\right)^{m-k}.$$

Using $1 + x \le e^x$ and $m - k = \Omega(m)$, so that $(m-k)/m = \Omega(1)$, yields

$$f(x) \le k(ea)^k \exp(-\Omega(ak)).$$

Since a is sufficiently large, the first two factors are lower order terms that we can hide by the Ω. Thus, we can simplify this further to

$$f(x) \le \exp(-\Omega(ak)).$$

Rearranging this using $a = \frac{m}{k} e^{-x}$ yields

$$f(x) = \exp(-\Omega(m \exp(-x)), \tag{3}$$

which holds for any $x \in [0, \ln\big(\frac{m}{\alpha k}\big)]$ for any sufficiently large $\alpha \ge 1$. Now we can bound the probability that $\mathsf{cost}(U) < \ln\big(\frac{m}{\alpha k}\big)$. This probability is equal to

$$\begin{aligned}
\int_0^{\ln(\frac{m}{\alpha k})} f(x)\,\mathrm{d}x &= \int_0^{\ln(\frac{m}{\alpha k})} f\left(\ln\left(\frac{m}{\alpha k}\right) - x\right)\,\mathrm{d}x \\
&= \int_0^{\ln(\frac{m}{\alpha k})} \exp\big(-\Omega(\alpha k \exp(x))\big)\,\mathrm{d}x && \text{using (3)} \\
&\le \int_0^{\infty} \exp\big(-\Omega(\alpha k(1+x))\big)\,\mathrm{d}x \le \exp\big(-\Omega(\alpha k)\big)
\end{aligned}$$

since $\int_0^\infty \exp(-\Omega(\alpha k x))\,\mathrm{d}x = O(1/(\alpha k)) \le 1$ as α is sufficiently large.

In order for MEDIAN to be less than $\ln\left(\frac{m}{\alpha k}\right)$, one of the subsets $U \subseteq V$ of size k has to have cost less than $\ln\left(\frac{m}{\alpha k}\right)$. We bound the probability of the latter using the union bound and get

$$\begin{aligned}\mathbb{P}\left(\mathsf{MEDIAN} < \ln\left(\frac{m}{\alpha k}\right)\right) &= \mathbb{P}\left(\exists U \subseteq V, |U| = k\colon\ \mathsf{cost}(U) < \ln\left(\frac{m}{\alpha k}\right)\right)\\ &\leq \binom{n}{k} \cdot \mathbb{P}\left(\mathsf{cost}(U) < \ln\left(\frac{m}{\alpha k}\right)\right)\\ &\leq \binom{n}{k} \cdot \exp\bigl(-\Omega(\alpha k)\bigr).\end{aligned}$$

By setting $\alpha = c\ln\left(\frac{n}{k}\right)$ for sufficiently large $c \geq 1$, we fulfill all conditions on α. This yields

$$\mathbb{P}\left(\mathsf{MEDIAN} < \ln\left(\frac{n}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln c\right) \leq \left(\frac{en}{k}\right)^k \cdot \left(\frac{n}{k}\right)^{-\Omega(ck)}.$$

Since k is sufficiently smaller than n, we have $\frac{en}{k} \leq (\frac{n}{k})^2$. Thus, for sufficiently large c, the right hand side simplifies to $(\frac{n}{k})^{-\Omega(ck)}$. Since k is at least 1 and sufficiently smaller than n, we have $(\frac{n}{k})^k \geq n$. Thus, the probability is bounded by $n^{-\Omega(c)}$, which finishes the proof. □

To bound the expected value of the quotient TRIVIAL / MEDIAN, we further need to bound the probabilities that TRIVIAL is much too large or MEDIAN is much too small. This is achieved by the following two lemmas.

Lemma 5.5. *Let $k \leq (1-\varepsilon)n$ for some constant $\varepsilon > 0$. Then, for any $c > 0$, we have*

$$\mathbb{P}(\mathsf{MEDIAN} < c) = O(c)^{\Omega(n)}.$$

Proof. Since $n - k$ vertices have to be connected to the k-median, the cost of the k-median is the sum of $n - k$ shortest path lengths. Thus, the cost of the minimal k-median is at least the sum of the smallest $n - k$ edge weights $w(e)$. We use Lemma 4.3 with $\alpha = \varepsilon$. □

Lemma 5.6. *For any $c \geq 3$, we have $\mathbb{P}(\mathsf{TRIVIAL} > n^c) \leq \exp(-n^{c/3})$.*

Proof. We can bound very roughly $\mathsf{TRIVIAL} \leq n \max_e\{w(e)\}$. As $\max_e\{w(e)\}$ is the maximum of $\binom{n}{2}$ independent exponentially distributed random variables, we have

$$\begin{aligned}\mathbb{P}\bigl(\mathsf{TRIVIAL} \leq n^c\bigr) &\geq (1 - \exp(-n^{c-1}))^{\binom{n}{2}} \geq 1 - \binom{n}{2} \cdot \exp(-n^{c-1})\\ &\geq 1 - \exp\bigl(-n^{c-2}\bigr) \geq 1 - \exp\bigl(-n^{c/3}\bigr).\end{aligned}$$

□

Proof of Theorem 5.2. Let $T = \mathsf{TRIVIAL}$ and $C = \mathsf{MEDIAN}$ for short. We have for any $m \geq 0$

$$\mathbb{E}\left(\frac{T}{C}\right) \leq \mathbb{E}\left(\frac{T}{m}\right) + \mathbb{P}(C < m) \cdot \mathbb{E}\left(\frac{T}{C} \mid C < m\right). \tag{4}$$

Case 1 ($k \leq c'n$, c' sufficiently small): Using Lemma 5.4, we can pick $c > 0$ such that

$$\mathbb{P}\left[C < \ln\left(\frac{n}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln c\right] \leq n^{-7}.$$

Set $m = \ln\left(\frac{n}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln c$. Then, by Lemma 5.1, we have

$$\mathbb{E}\left(\frac{T}{m}\right) \leq \frac{\ln(n/k) + O(1)}{m} \leq 1 + O\left(\frac{\ln\ln(n/k)}{\ln(n/k)}\right).$$

We show that the second summand of inequality (4) is $O(1/n)$ in the current situation, which shows the claim. We have

$$\begin{aligned}
\mathbb{P}(C < m) \cdot \mathbb{E}\left(\frac{T}{C} \mid C < m\right) &= \mathbb{P}(C < m) \cdot \int_0^\infty \mathbb{P}\left(\frac{T}{C} \geq x \mid C < m\right) \mathrm{d}x \\
&\leq \mathbb{P}(C < m) \cdot \left(n^6 + \int_{n^6}^\infty \mathbb{P}\left(\frac{T}{C} \geq x \mid C < m\right) \mathrm{d}x\right) \\
&\leq n^{-1} + \int_{n^6}^\infty \mathbb{P}\left(\frac{T}{C} \geq x \text{ and } C < m\right) \mathrm{d}x \\
&\leq n^{-1} + \int_{n^6}^\infty \mathbb{P}\left(\frac{T}{C} \geq x\right) \mathrm{d}x \\
&\leq n^{-1} + \int_{n^6}^\infty 2\max\left\{\mathbb{P}\left(T \geq \sqrt{x}\right), \mathbb{P}\left(C \leq \frac{1}{\sqrt{x}}\right)\right\} \mathrm{d}x
\end{aligned}$$

since $T/C \geq x$ implies $T \geq \sqrt{x}$ or $C \leq 1/\sqrt{x}$. Using Lemmas 5.5 and 5.6, this yields

$$\mathbb{P}(C < m) \cdot \mathbb{E}\left(\frac{T}{C} \mid C < m\right) \leq n^{-1} + \int_{n^6}^\infty 2\max\left\{\exp\left(-x^{1/6}\right), O\left(\frac{1}{\sqrt{x}}\right)^{\Omega(n)}\right\} \mathrm{d}x = O(1/n).$$

Case 2 ($c'n < k \leq (1-\varepsilon)n$): We repeat the proof above, now choosing m to be a sufficiently small constant. Then $\mathbb{P}(C < m) = O(m)^{\Omega(n)} \leq O(n^{-7})$ by Lemma 5.5, and we have

$$\mathbb{E}\left(\frac{T}{m}\right) = \frac{\ln(n/k) + O(1)}{m} = O(1),$$

since $k > c'n$. Together with the first case, this shows the first claim. □

6 Concluding Remarks

6.1 General Probability Distributions

Using a coupling argument, Janson [24, Section 3] proved that the results about the length of a fixed edge and the longest edge carry over if the exponential distribution is replaced by a probability distribution with the following property: the probability that an edge weight is smaller than x is $x + o(x)$. This property is satisfied, e.g., by the exponential distribution with parameter 1 and by the uniform distribution on the interval $[0, 1]$. The intuition is that, because the longest edge has a length of $O(\log n/n) = o(1)$, only the behavior of the distribution in a small, shrinking interval $[0, o(1)]$ is relevant and the $o(x)$ term becomes irrelevant.

We believe that also all of our results carry over to such probability distributions. In fact, we started our research using the uniform distribution and only switched to exponential distributions because they are technically easier to handle. However, we decided not to carry out the corresponding proofs because, first, they seem to be technically very tedious and, second, we feel that they do not add much.

6.2 Open Problems

To conclude the paper, let us list the open problems that we consider most interesting:

1. While the distribution of distances in asymmetric instances does not differ much from the symmetric case, an obstacle in the application of asymmetric random shortest path metrics seems to be the lack of clusters of small diameter (see Section 3). Is there an asymmetric counterpart for this?

2. Is it possible to prove an $1 + o(1)$ approximation ratio (like Dyer and Frieze [15] for the patching algorithm) for any of the simple heuristics that we analyzed?

3. What is the approximation ratio of 2-opt in random shortest path metrics? In the worst case on metric instances, it is $O(\sqrt{n})$ [12]. For independent, non-metric edge lengths drawn uniformly from the interval $[0, 1]$, the expected approximation ratio is $O(\sqrt{n} \cdot \log^{3/2} n)$ [17]. For d-dimensional geometric instances, the smoothed approximation ratio is $O(\phi^{1/d})$ [18], where ϕ is the perturbation parameter.

   We easily get an approximation ratio of $O(\log n)$ based on the two facts that the length of the optimal tour is $\Theta(1)$ with high probability and that $\Delta_{\max} = O(\log n/n)$ with high probability. Can we prove that the expected ratio of 2-opt is $o(\log n)$?

References

[1] Louigi Addario-Berry, Nicolas Broutin, and Gábor Lugosi. The longest minimum-weight path in a complete graph. *Combinatorics, Probability and Computing*, 19(1):1–19, 2010.

[2] David Arthur, Bodo Manthey, and Heiko Röglin. Smoothed analysis of the k-means method. *Journal of the ACM*, 58(5), 2011.

[3] Vijay Arya, Naveen Garg, Rohit Khandekar, Adam Meyerson, Kamesh Munagala, and Vinayaka Pandit. Local search heuristic for k-median and facility location problems. *SIAM Journal on Computing*, 33(3):544–562, 2004.

[4] Giorgio Ausiello, Pierluigi Crescenzi, Giorgio Gambosi, Viggo Kann, Alberto Marchetti-Spaccamela, and Marco Protasi. *Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties*. Springer, 1999.

[5] David Avis, Burgess Davis, and J. Michael Steele. Probabilistic analysis of a greedy heuristic for Euclidean matching. *Probability in the Engineering and Informational Sciences*, 2:143–156, 1988.

[6] Yossi Azar. Lower bounds for insertion methods for TSP. *Combinatorics, Probability and Computing*, 3:285–292, 1994.

[7] Shankar Bhamidi, Remco van der Hofstad, and Gerard Hooghiemstra. First passage percolation on random graphs with finite mean degrees. *Annals of Applied Probability*, 20(5):1907–1965, 2010.

[8] Shankar Bhamidi, Remco van der Hofstad, and Gerard Hooghiemstra. First passage percolation on the Erdős-Rényi random graph. *Combinatorics, Probability & Computing*, 20(5):683–707, 2011.

[9] Shankar Bhamidi, Remco van der Hofstad, and Gerard Hooghiemstra. Universality for first passage percolation on sparse random graphs. Technical Report 1210.6839 [math.PR], arXiv, 2012.

[10] Nathaniel D. Blair-Stahn. First passage percolation and competition models. Technical Report 1005.0649v1 [math.PR], arXiv, 2010.

[11] S. R. Broadbent and J.M. Hammersley. Percolation processes. I. Crystals and mazes. *Proceedings of the Cambridge Philosophical Society*, 53(3):629–641, 1957.

[12] Barun Chandra, Howard J. Karloff, and Craig A. Tovey. New results on the old k-opt algorithm for the traveling salesman problem. *SIAM Journal on Computing*, 28(6):1998–2029, 1999.

[13] Robert Davis and Armand Prieditis. The expected length of a shortest path. *Information Processing Letters*, 46(3):135–141, 1993.

[14] Martin Dyer, Alan Frieze, and Boris Pittel. The average performance of the greedy matching algorithm. *Annals of Applied Probability*, 3(2):526–552, 1993.

[15] Martin E. Dyer and Alan M. Frieze. On patching algorithms for random asymmetric travelling salesman problems. *Mathematical Programming*, 46:361–378, 1990.

[16] Maren Eckhoff, Jesse Goodman, Remco van der Hofstad, and Francesca R. Nardi. Short paths for first passage percolation on the complete graph. *Journal of Statistical Physics*, 151(6):1056–1088, 2013.

[17] Christian Engels and Bodo Manthey. Average-case approximation ratio of the 2-opt algorithm for the TSP. *Operations Research Letters*, 37(2):83–84, 2009.

[18] Matthias Englert, Heiko Röglin, and Berthold Vöcking. Worst case and probabilistic analysis of the 2-Opt algorithm for the TSP. *Algorithmica*, 68(1):190–264, 2014.

[19] Alan M. Frieze. On random symmetric travelling salesman problems. *Mathematics of Operations Research*, 29(4):878–890, 2004.

[20] Alan M. Frieze and G. R. Grimmett. The shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics*, 10:57–77, 1985.

[21] Refael Hassin and Eitan Zemel. On shortest paths in graphs with random weights. *Mathematics of Operations Research*, 10(4):557–564, 1985.

[22] Remco van der Hofstad, Gerard Hooghiemstra, and Piet van Mieghem. First passage percolation on the random graph. *Probability in the Engineering and Informational Sciences*, 15(2):225–237, 2001.

[23] Remco van der Hofstad, Gerard Hooghiemstra, and Piet van Mieghem. Size and weight of shortest path trees with exponential link weights. *Combinatorics, Probability and Computing*, 15(6):903–926, 2006.

[24] Svante Janson. One, two, three times $\log n/n$ for paths in a complete graph with edge weights. *Combinatorics, Probability and Computing*, 8(4):347–361, 1999.

[25] David S. Johnson and Lyle A. McGeoch. Experimental analysis of heuristics for the STSP. In Gregory Gutin and Abraham P. Punnen, editors, *The Traveling Salesman Problem and its Variations*, chapter 9. Kluwer, 2002.

[26] Richard M. Karp. Probabilistic analysis of partitioning algorithms for the traveling-salesman problem in the plane. *Mathematics of Operations Research*, 2(3):209–224, 1977.

[27] Richard M. Karp and J. Michael Steele. Probabilistic analysis of heuristics. In Eugene L. Lawler, Jan Karel Lenstra, Alexander H. G. Rinnooy Kan, and David B. Shmoys, editors, *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, pages 181–205. Wiley, 1985.

[28] Walter Kern. A probabilistic analysis of the switching algorithm for the TSP. *Mathematical Programming*, 44(2):213–219, 1989.

[29] István Kolossváry and Júlia Komjáthy. First passage percolation on inhomogeneous random graphs. Technical Report 1201.3137v1 [math.PR], arXiv, 2012.

[30] V. G. Kulkarni and V. G. Adlakha. Maximum flow in planar networks in exponentially distributed arc capacities. *Communications in Statistics. Stochastic Models*, 1(3):263–289, 1985.

[31] Vidyadhar G. Kulkarni. Shortest paths in networks with exponentially distributed arc lengths. *Networks*, 16(3):255–274, 1986.

[32] Vidyadhar G. Kulkarni. Minimal spanning trees in undirected networks with exponentially distributed arc weights. *Networks*, 18(2):111–124, 1988.

[33] Bodo Manthey and Rianne Veenstra. Smoothed analysis of the 2-Opt heuristic for the TSP: Polynomial bounds for Gaussian noise. In Leizhen Cai, Siu-Wing Cheng, and Tak-Wah Lam, editors, *Proc. 24th Int. Symp. on Algorithms and Computation (ISAAC)*, volume 8283 of *Lecture Notes in Computer Science*, pages 579–589. Springer, 2013.

[34] Yuval Peres, Dmitry Sotnikov, Benny Sudakov, and Uri Zwick. All-pairs shortest paths in $O(n^2)$ time with high probability. *Journal of the ACM*, 60(4):26, 2013.

[35] Edward M. Reingold and Robert Endre Tarjan. On a greedy heuristic for complete matching. *SIAM Journal on Computing*, 10(4):676–681, 1981.

[36] Daniel J. Rosenkrantz, Richard E. Stearns, and Philip M. Lewis II. An analysis of several heuristics for the traveling salesman problem. *SIAM Journal on Computing*, 6(3):563–581, 1977.

[37] Sheldon M. Ross. *Introduction to Probability Models*. Academic Press, 10th edition, 2010.

[38] Kenneth J. Supowit, David A. Plaisted, and Edward M. Reingold. Heuristics for weighted perfect matching. In *Proc. of the 12th Annual ACM Symposium on Theory of Computing (STOC)*, pages 398–419. ACM, 1980.

[39] A. M. Vershik. Random metric spaces and universality. *Russian Mathematical Surveys*, 59(2):259–295, 2004.

[40] Joseph E. Yukich. *Probability Theory of Classical Euclidean Optimization Problems*, volume 1675 of *Lecture Notes in Mathematics*. Springer, 1998.