

BSc Thesis Applied Mathematics



Analysis of Heuristics with Random Shortest Path Metrics on Sparse Graphs

L.M. Wackerle Garcia

Supervisor: Dr. B. Manthey, S. Klootwijk, Dr. P.T. de Boer

July, 2020

Department of Applied Mathematics
Faculty of Electrical Engineering,
Mathematics and Computer Science

Acknowledgements

This thesis marks the completion of my Bachelor's degree in Applied Mathematics and Technical Computer Science. Reflecting back on the last three years, I can only feel grateful to have been provided with the tools and knowledge that have made this research possible.

I would like to thank my primary supervisors: Dr. B. Manthey for his constructive criticism and continuous guidance and S. Klootwijk for providing me with detailed explanations of the many challenging concepts in his paper and for his enthusiastic support. They helped me overcome many difficulties and pointed me to new, critical insights. I have no doubt that this thesis would not have been possible without our weekly meetings.

I also wish to express my gratitude to the many people who helped keep my spirit up during the period of lockdown, allowing me to remain passionate about my thesis. In particular, I thank my parents and sister who supported me then, and throughout the many other challenges of my Bachelor program.

Analysis of Heuristics with Random Shortest Path Metrics on Sparse Graphs

L.M. Wackerle Garcia*

July, 2020

Abstract

Classical optimisation problems such as the minimum-distance perfect matching and travelling salesman problem can be impractical to solve to optimality for large instances. Simple heuristics have shown to perform strikingly well, producing near-optimal solutions with only a fraction of the time-complexity. Worst-case analysis of these heuristics suggests a far poorer performance than what we see in practice and so efforts to explain this behaviour have shifted to a probabilistic framework.

Withing a probabilistic framework, the instances of the two aforementioned optimisation problems can be considered as a discrete metric space drawn from a particular distribution. Much analysis has already been done for heuristics on instances drawn from Euclidean space. While these instances have useful mathematical properties, they are not always representative of realistic networks. Recent efforts to shift the analysis to a more realistic distribution has seen results produced for random shortest path metrics generated from dense graphs and a small subset of sparse graphs.

This thesis extends these finding to a wider class of sparse graphs by generalising the results from Klootwijk and Manthey [9]. We consider the optimisation problems on random shortest path metrics generated from sparse graphs with a fast growing cut size. The performance of three simple heuristics is analysed: the greedy heuristic for the minimum-distance perfect matching problem and the nearest neighbour and insertion heuristic for the travelling salesman problem. Applied within a probabilistic framework, it can be shown that all three heuristics achieve a constant expected approximation ratio. This is indicative of the empirical performance of these heuristics.

Keywords: Random shortest paths, Random metrics, Approximation algorithms, Combinatorial optimisation, Greedy heuristics.

*Email: l.m.wackerle-garcia@student.utwente.nl

1 Introduction

Combinatorial optimisation problems have a wide range of industrial applications. The minimum-distance perfect matching problem and travelling salesman problem (TSP), or variations of these problems, are useful for many problems from scheduling to logistics. For large scale instances these problems can often not be solved to optimality within a reasonable amount of time and resources. Instead, approximation algorithms are used which can compute near-optimal (or approximate) solutions with a significantly lower time complexity. Heuristic algorithms are one such class of approximation algorithms which, despite having a very poor worst-case performance, often perform extremely well in practice.

Heuristic algorithms for the TSP and minimum-distance perfect matching problem have well established results for their worst case performance. However, these results do not match the performance that we observe in practice. In order to obtain a theoretical understanding of these empirical observations, recent analysis has been done within a probabilistic framework. One of the main challenges with this approach has been choosing a probability distribution on the set of possible instances on which to model and analyse the problem. Earlier research started by using simple ‘well-behaved’ distributions as these were more tractable, but they are often not very representative of real instances. For example, instances embedded in Euclidean space have been extensively analysed. They offer a convenient mathematical structure which can be exploited to analyse the problem. However, many practical instances, like travelling times between cities, are metric but not Euclidean. Therefore, it is interesting to broaden the scope of the underlying models used to analyse heuristic algorithms.

The performance of an approximation algorithm can be quantified by its approximation ratio: the ratio of the value of the approximate solution produced by the algorithm to the value of the optimal solution. As mentioned before, finding a lower bound for the approximation ratio is not very indicative of an algorithm’s performance in practice. Instead, we analyse the algorithm in a probabilistic framework to determine the expected approximation ratio. For generality, we consider the algorithm applied to a distribution of graphs of any arbitrary size. Thus, rather than finding a specific value, we determine the order (time complexity) of the expected approximation ratio. This gives us an indication of how well the approximation algorithm scales.

Related Work. In the last few years, there have been several papers using random shortest path metrics as the underlying model for analysing heuristic algorithms. A random shortest path metric is constructed as follows: given an undirected graph with edge weights drawn independently at random, we generate a new graph where the distance between any two vertices is defined as the total weight of the shortest path between those two vertices with respect to the original graph. This metric, originally proposed by Karp and Steele [8], has been utilised by Bringmann et al. [5] and Klootwijk et al. [10] to obtain results for several underlying graph classes. Klootwijk and Manthey [9] extended these results to random shortest path metrics generated from sparse graphs, specifically, square grid graphs. While the approach of using sparse graphs provides a more realistic underlying model, it can be made more applicable by generalising the ideas beyond the rather restrictive scope of square grid graphs.

Main Result. By employing the same methods as Klootwijk and Manthey [9], we extend their results for three greedy-like heuristics. These heuristics are the greedy heuristic for the minimum-distance perfect matching problem and the nearest neighbour and insertion heuristics for the travelling salesman problem. We show that all three achieve a constant expected approximation ratio on instances drawn from a random shortest path metric generated from sparse graphs with a fast growing cut size (see Definition 1). Thus our results apply to a much broader subset of sparse graphs as the underlying model.

2 Notation and Model

We start by outlining the notation that we use throughout this thesis. We write $[n]$ as a shorthand notation for $\{1, 2, \dots, n\}$. We write $\exp(\cdot)$ to denote the exponential function and $\text{Exp}(\lambda)$ to denote the exponential distribution with parameter λ . For a random variable X distributed according to a probability distribution P we use $X \sim P$. We write $X \sim \sum_{i=1}^n \text{Exp}(\lambda_i)$ if X is the sum of n independent exponentially distributed random variables having parameters $\lambda_1, \dots, \lambda_n$, for this distribution we have that $\mathbb{E}[X] = \sum_{i=1}^n \frac{1}{\lambda_i}$. We indicate that a random variable X_1 is stochastically dominated by a random variable X_2 by writing $X_1 \lesssim X_2$, (this means that $F_1(x) \geq F_2(x)$ for all x , where F_i is the cumulative distribution function of X_i). Finally, we remind the reader that $f(x) = O(g(x))$ means that there exists positive constants x_0 and c such that $|f(x)| \leq cg(x)$ for all $x \geq x_0$. Note that while $O(\cdot)$ is really a set we adopt the convention of using '=' to denote set inclusion when used in this context.

To construct our random shortest path metric we consider simple connected undirected graphs $G = (V, E)$. Edge weights $w(\cdot)$ are drawn independently at random from the exponential distribution with parameter 1. The size of the graph is denoted by $n = |V|$. The shortest path distance function $d : V \times V \rightarrow \mathbb{R}_{\geq 0}$ maps each pair of vertices $u, v \in V$ to the total weight of the edges in the lightest u, v -path in G (w.r.t. to the random weights $w(\cdot)$). It follows immediately that $d(v, v) = 0$ 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 $u, s, v \in V$ i.e. $d(\cdot)$ is a metric. We consider the distance function $d(\cdot)$ as the random shortest path metric generated from G .

Furthermore, we define $\Delta_{\max} := \max_{u, v} d(u, v)$ as the diameter of the random metric. We define the Δ -ball around a vertex v as $B_{\Delta}(v) := \{u \in V \mid d(u, v) \leq \Delta\}$, i.e. the set of vertices within distance Δ of v . We denote the distance from a vertex v to the k th closest vertex from it (including v itself) as $\tau_k(v) := \min\{\Delta \mid |B_{\Delta}(v)| \geq k\}$. It follows that $B_{\tau_k(v)}(v)$ denotes the set of the k closest vertices to v including v itself. Observe that by these definitions we have that $\tau_1(v) = 0$ and $|B_0(v)| = 1$. For our analysis we are interested in the cut size of a subset of vertices. We denote the cut size of the ball containing the k closest vertices of v by $\chi_k(v) := |\delta(B_{\tau_k(v)}(v))|$ where $\delta(U) := \{\{u, v\} \in E \mid u \in U, v \notin U\}$ denotes the cut induced by a $U \subset V$.

Our analysis is performed for sparse graphs with a fast growing cut size. A family of graphs is sparse if $|E| = \Theta(|V|) = \Theta(n)$, that is, as n grows the average degree of its vertices is bounded by some constant. We define what we mean by a fast growing cut size as follows:

Definition 1. A family of sparse graphs is said to have a fast growing cut size if there exists constants $\alpha > 0$, $\epsilon \in (0, 1)$ and $\gamma \in (0, 1/2]$ such that for any size n and $U \subset V$

$$|\delta(U)| \geq \alpha|U|^\epsilon \quad \text{if } |U| \leq \gamma n.$$

We also assume that $n > 1/\gamma$ and often use $\beta := 1 - \epsilon \in (0, 1)$ instead of writing out $1 - \epsilon$. Naturally, this definition constrains the type of sparse graphs that we can analyse. However, choosing this constraint allows us to utilise the techniques used by Klootwijk and Manthey [9] for our analysis in order to obtain useful results. With some consideration it can be seen that this property holds for a wide range of classes of sparse graphs. In particular several types of sparse graphs that may be found in real world situations should satisfy the property (at least with high probability). To illustrate this, the following lemma shows that d -dimensional grid graphs satisfy the property given in Definition 1.

Lemma 2. *The family of d -dimensional grid graphs has a fast growing cut size for any integer $d > 1$.*

Proof. The grid graph is the graph on $[k]^d$ in which $x = (x_i)_1^d$ is joined to $y = (y_i)_1^d$ if for some i we have $|x_i - y_i| = 1$ and $x_j = y_j$ for all $j \neq i$. Observe that this graph has $n = k^d$ vertices. Bollabás and Leader [3, Thm. 3] proved that for any $U \subset [k]^d$ with $|U| \leq n/2$ we have

$$|\delta(U)| \geq \min \left\{ r n^{\frac{1}{r} - \frac{1}{d}} |U|^{1 - \frac{1}{r}} : r = 1, \dots, d \right\}.$$

We may bound this result by exploiting that $|U| \leq n/2$:

$$\begin{aligned} |\delta(U)| &\geq \min_{r \in [d]} \left\{ r n^{\frac{1}{r} - \frac{1}{d}} |U|^{1 - \frac{1}{r}} \right\} \geq \min_{r \in [d]} \left\{ r (2|U|)^{\frac{1}{r} - \frac{1}{d}} |U|^{1 - \frac{1}{r}} \right\} \\ &= \min_{r \in [d]} \left\{ r 2^{\frac{1}{r} - \frac{1}{d}} \right\} |U|^{1 - \frac{1}{d}} = 2^{1 - \frac{1}{d}} \cdot |U|^{1 - \frac{1}{d}} \end{aligned}$$

From the last line we get that a d -dimensional grid graph must have a fast growing cut size as given by Definition 1 with $\alpha = 2^{(1-1/d)}$, $\epsilon = 1 - 1/d$ and $\gamma = 1/2$. \square

Lattice graphs and sparse random geometric graphs, amongst many others, also have a fast growing cut size (at least with high probability). Throughout this thesis we apply our analysis to graphs that satisfy Definition 1 as underlying graphs for the distance metric.

3 Structural Properties

Our probabilistic analysis of the three greedy-like heuristics is based on the structural properties that we provide in this section. The first three lemmas are known results. These are used for the technical analysis as well as to bound the values of the optimal solution: this is necessary to be able to say something about the expected approximation ratio. The latter results are derived specifically for sparse graphs with a fast growing cut size as given by Definition 1.

Technical Lemma. The following tail bound for the sum of exponential variables is used frequently.

Lemma 3 ([7, Thm. 5.1(i,iii)]). *Let $X \sim \sum_{i=1}^m \text{Exp}(a_i)$. Let $\mu = \mathbb{E}[X] = \sum_{i=1}^m 1/a_i$ and $a_* = \min_i a_i$. Then*

$$\mathbb{P}(X \geq \lambda\mu) \leq \begin{cases} \lambda^{-1} \exp(-a_*\mu(\lambda - 1 - \ln(\lambda))) & \text{for any } \lambda \geq 1, \\ \exp(-a_*\mu(\lambda - 1 - \ln(\lambda))) & \text{for any } \lambda \leq 1. \end{cases}$$

Bounding the Optimal Solution for Sparse Graphs. In order to obtain an approximation ratio we need to establish a bound on the optimal solution for the minimum distance perfect matching problem and the travelling salesman problem. The following results hold for any arbitrary sparse graph G .

Lemma 4 ([9, Lem. 6]). *Let S_m denote the sum of the m lightest edge weights in G . Then*

$$\mathbb{P}(S_m \leq cn) \leq \exp\left(m \left(2 + \ln\left(\frac{c|E|n}{m^2}\right)\right)\right).$$

Lemma 5 ([9, Lem. 7]). *Let S_m denote the sum of the m lightest edge weights in G . Then we have $\text{TSP} \geq \text{MM} \geq S_{n/2}$, where TSP and MM are the total distance of a shortest TSP tour and a minimum-distance perfect matching, respectively.*

To make this thesis more self-contained we provide the proofs of Lemma 4 and 5 by Klootwijk and Manthey [9] as an Appendix.

Generalised Random Growth Process. We work with the random shortest path metric generated from a graph G on n vertices with a fast growing cut size. To prove our results for sparse random shortest path metrics it is important to analyse the distribution of $\tau_k(v)$. This distribution depends heavily on the exact position of v within G . Given the generality of G we do not have sufficient information to find an exact distribution of $\tau_k(v)$. Instead we use a stochastic upper bound to analyse the distribution which we derive from Definition 1. The following corollary establishes this result.

Lemma 6. *For any fixed $v \in V$ and any $k \leq \gamma n$ we have*

$$\tau_k(v) \lesssim \sum_{m=1}^{k-1} \text{Exp}(\alpha m^\epsilon).$$

Proof. The values of $\tau_k(v)$ are generated by the following a birth process (which has already been analysed by several others [5, 6, 9, 10]). By definition for $k = 1$ we have $\tau_k(v) = 0$ and also $\sum_{m=1}^{k-1} \text{Exp}(\alpha m^\epsilon) = 0$. For $k \geq 2$, we can determine $\tau_k(v)$ inductively by considering the outgoing edges from the ball $B_{\tau_{k-1}(v)}(v)$. These are all the edges (u, x) with $u \in B_{\tau_{k-1}(v)}(v)$ and $x \notin B_{\tau_{k-1}(v)}(v)$, and they are conditioned to have a weight $w(u, x)$ greater than $\tau_{k-1}(v) - d(v, u)$ (otherwise we have that $w(u, x) + d(v, u) \not\geq \tau_{k-1}(v)$ which implies $x \in B_{\tau_{k-1}(v)}(v)$). By definition we have $\chi_{k-1}(v)$ such edges. It follows from the memorylessness property of the exponential distribution that $\tau_k(v) - \tau_{k-1}(v)$ is the

minimum of $\chi_{k-1}(v)$ exponential random variables (with parameter 1). This implies that $\tau_k(v) - \tau_{k-1}(v) \sim \text{Exp}(\chi_{k-1}(v))$. By induction it follows that

$$\tau_k(v) \sim \sum_{m=1}^{k-1} \text{Exp}(\chi_m(v)).$$

It is rather complicated to work directly with $\chi_{k-1}(v)$ since it is a stochastic variable and depends on the choice of v . For this reason we only consider sparse graphs with a fast growing cut size, as given by Definition 1, as this allows us to bound $\chi_m(v) \geq \alpha m^\epsilon$ for all $m \leq \gamma n$. The result follows by stochastic dominance. \square

This stochastic upper bound is necessary to determine a bound on the cumulative distribution functions of $\tau_k(v)$ and $|B_\Delta(v)|$. These bounds are derived in the following lemma and corollary and are the foundation for the clustering result of the next section. All the methods used in this section closely follow the techniques used by Klootwijk and Manthey [9].

Lemma 7. *For any $\Delta > 0$, $v \in V$, and $k \in [n]$ such that $k \leq \min\{\gamma n, (\alpha\beta\Delta)^{1/\beta} + 1\}$, we have*

$$\mathbb{P}(\tau_k(v) \leq \Delta) \geq 1 - \frac{(k-1)^\beta}{\alpha\beta\Delta} \exp\left(-\frac{k^\beta - 1}{\beta} \left(\frac{\alpha\beta\Delta}{(k-1)^\beta} - 1 - \ln\left(\frac{\alpha\beta\Delta}{(k-1)^\beta}\right)\right)\right).$$

Proof. From Lemma 6 we can see that

$$\mathbb{P}(\tau_k(v) \leq \Delta) \geq \mathbb{P}\left(\sum_{m=1}^{k-1} \text{Exp}(\alpha m^\epsilon) \leq \Delta\right) = 1 - \mathbb{P}\left(\sum_{m=1}^{k-1} \text{Exp}(\alpha m^\epsilon) \geq \Delta\right).$$

Next, we want to apply the result of Lemma 3 for $\lambda \geq 1$. For this purpose, set

$$\mu := \mathbb{E}\left[\sum_{m=1}^{k-1} \text{Exp}(\alpha m^\epsilon)\right] = \sum_{m=1}^{k-1} \frac{1}{\alpha m^\epsilon} \quad \text{and} \quad \lambda := \frac{\Delta}{\mu},$$

We can bound μ from above by considering it as a Riemann sum:

$$\sum_{m=1}^{k-1} \frac{1}{\alpha m^\epsilon} \leq \int_0^{k-1} \frac{1}{\alpha} x^{-\epsilon} dx = \frac{1}{\alpha(1-\epsilon)} x^{1-\epsilon} \Big|_0^{k-1} = \frac{(k-1)^{(1-\epsilon)}}{\alpha(1-\epsilon)}.$$

Taking that $\beta := 1 - \epsilon$ and shifting the bounds of the integral to also bound μ from below, we get that

$$\frac{k^\beta - 1}{\alpha\beta} \leq \mu \leq \frac{(k-1)^\beta}{\alpha\beta} \quad \text{and} \quad \lambda = \frac{\Delta}{\mu} \geq \frac{\alpha\beta\Delta}{(k-1)^\beta}. \quad (1)$$

Observe that for $k \leq (\alpha\beta\Delta)^{1/\beta} + 1$ we have $\lambda \geq 1$. Lemma 3 now yields

$$1 - \mathbb{P}\left(\sum_{m=1}^{k-1} \text{Exp}(\alpha m^\epsilon) \geq \Delta\right) \geq 1 - \lambda^{-1} \exp(-\alpha\mu(\lambda - 1 - \ln(\lambda))).$$

It can now be seen that this final expression is increasing in both μ and λ . Therefore, we can apply the appropriate inequalities from Equation (1) to obtain the desired result. \square

Corollary 8. *Let n be sufficiently large. There exists a constant c_1 such that for any $\Delta > 0$ and $v \in V$ we have*

$$\mathbb{P}\left(|B_\Delta(v)| < \min\left\{\gamma(\alpha\beta\Delta)^{1/\beta}, \gamma n\right\}\right) \leq \frac{c_1}{\Delta^{1/\beta}}.$$

Proof. First we observe that $|B_\Delta(v)| \geq k$ if and only if $\tau_k(v) \leq \Delta$. Let $s_\Delta := \min\{\gamma(\alpha\beta\Delta)^{1/\beta}, \gamma n\}$ and consider n to be large: $n^\beta \geq \frac{\gamma^{-\beta}-\beta}{1-\gamma^\beta}$ is sufficient (this also implies that $n^\beta > \gamma^{-\beta}$). Using Lemma 7 with $k = s_\Delta$ we obtain

$$\mathbb{P}(|B_\Delta(v)| < s_\Delta) \leq \frac{(s_\Delta - 1)^\beta}{\alpha\beta\Delta} \exp\left(-\frac{s_\Delta^\beta - 1}{\beta} \left(\frac{\alpha\beta\Delta}{(s_\Delta - 1)^\beta} - 1 - \ln\left(\frac{\alpha\beta\Delta}{(s_\Delta - 1)^\beta}\right)\right)\right).$$

We need to show that there exists a constant c_1 such that for any $\Delta > 0$ we have

$$\frac{\Delta^{1/\beta}(s_\Delta - 1)^\beta}{\alpha\beta\Delta} \exp\left(-\frac{s_\Delta^\beta - 1}{\beta} \left(\frac{\alpha\beta\Delta}{(s_\Delta - 1)^\beta} - 1 - \ln\left(\frac{\alpha\beta\Delta}{(s_\Delta - 1)^\beta}\right)\right)\right) \leq c_1.$$

To do this we first consider the case $n^\beta \geq \alpha\beta\Delta$. In this case we have $s_\Delta = \gamma(\alpha\beta\Delta)^{1/\beta}$. Recall from the proof of Lemma 7 that the left hand side in the expression above can be written in terms of λ and μ as

$$f(\Delta) := \lambda^{-1} \Delta^{1/\beta} \exp(-\alpha\mu(\lambda - 1 - \ln(\lambda))),$$

which is decreasing in λ and in μ for $\lambda > 1$. This allows us to use the inequality,

$$\lambda = \frac{\alpha\beta\Delta}{(s_\Delta - 1)^\beta} > \frac{\alpha\beta\Delta}{s_\Delta^\beta} = \frac{1}{\gamma^\beta} > 1,$$

to bound $f(\Delta)$ as follows

$$f(\Delta) \leq \gamma^\beta \Delta^{1/\beta} \exp\left(-\frac{\gamma^\beta \alpha\beta\Delta - 1}{\beta} \left(\frac{1}{\gamma^\beta} - 1 - \ln\left(\frac{1}{\gamma^\beta}\right)\right)\right) \leq c_1.$$

Since $\gamma^{-\beta} > 1$ the term $(\gamma^{-\beta} - 1 - \ln(\gamma^{-\beta}))$ is just some positive constant, so we can rewrite this inequality as $f(\Delta) \leq p\Delta^{1/\beta}e^{-q\Delta}$ for some positive nonzero constants p and q . It can easily be verified that this function has a global maximum for $\Delta > 0$. Let c_1 denote this maximum value. Note that c_1 depends on α , β and γ but not on n , Δ or v . The proof for the second case $n^\beta < \alpha\beta\Delta$ is slightly more involved. In this case we have $s_\Delta = \gamma n$. To simplify some of the computations in this proof let

$$a_n := \frac{\lambda}{\Delta} = \frac{\alpha\beta}{(\gamma n - 1)^\beta} \quad \text{and} \quad b_n := \alpha\mu = \frac{(\gamma n)^\beta - 1}{\beta}.$$

Observe that these are constants with respect to Δ (but may vary with n) and that both are positive. In terms of this new notation our task is to show that

$$g(\Delta, n) := a_n^{-1} \Delta^{1/\beta-1} \exp(-b_n(a_n\Delta - 1 - \ln(a_n\Delta))) \leq c_1.$$

We can do this by showing that $g(\Delta, n) \leq g(\frac{n^\beta}{\alpha\beta}, n) \leq c_1$ for all $\Delta > \frac{n^\beta}{\alpha\beta}$. In order to do so we compute the partial derivative of g with respect to Δ and show that it is nonpositive

for all $\Delta > \frac{n^\beta}{\alpha\beta}$. The partial derivative is given by

$$\begin{aligned}\frac{\partial g(\Delta, n)}{\partial \Delta} &= \frac{\partial}{\partial \Delta} \left(a_n^{-1} \Delta^{1/\beta-1} e^{-b_n(a_n \Delta - 1 - \ln(a_n \Delta))} \right) \\ &= a_n^{-1} (1/\beta - 1) \Delta^{1/\beta-2} e^{-b_n(a_n \Delta - 1 - \ln(a_n \Delta))} \\ &\quad + a_n^{-1} \Delta^{1/\beta-1} (-b_n (a_n - 1/\Delta)) e^{-b_n(a_n \Delta - 1 - \ln(a_n \Delta))} \\ &= a_n^{-1} \Delta^{1/\beta-2} (1/\beta - 1 - b_n(a_n \Delta - 1)) e^{-b_n(a_n \Delta - 1 - \ln(a_n \Delta))}\end{aligned}$$

Notice that the terms $a_n^{-1} \Delta^{1/\beta-2}$ and e^x are both positive (for any x), so it remains to show that the term $1/\beta - 1 - b_n(a_n \Delta - 1)$ is nonpositive for $\Delta > \frac{n^\beta}{\alpha\beta}$. To do this we use the conditions $\Delta > \frac{n^\beta}{\alpha\beta}$ and $n^\beta \geq \frac{\gamma^{-\beta} - \beta}{1 - \gamma^\beta}$.

$$\begin{aligned}1/\beta - 1 - b_n(a_n \Delta - 1) &= \frac{1 - \beta}{\beta} - \frac{(\gamma n)^\beta - 1}{\beta} \left(\frac{\alpha\beta}{(\gamma n - 1)^\beta} \Delta - 1 \right) \\ &= \frac{1 - \beta}{\beta} + \frac{(\gamma n)^\beta - 1}{\beta} - \frac{\alpha((\gamma n)^\beta - 1)}{(\gamma n - 1)^\beta} \Delta \\ &= \frac{(\gamma n)^\beta - \beta}{\beta} - \frac{(\gamma n)^\beta - 1}{\beta \gamma^\beta} \cdot \frac{\alpha\beta\gamma^\beta}{(\gamma n - 1)^\beta} \Delta \\ &\leq \frac{(\gamma n)^\beta - \beta}{\beta} - \frac{(\gamma n)^\beta - 1}{\beta \gamma^\beta} \\ &= \frac{(\gamma n)^\beta - \beta}{\beta} - \frac{n^\beta - \gamma^{-\beta}}{\beta} \\ &= \frac{(\gamma^\beta - 1)n^\beta + \gamma^{-\beta} - \beta}{\beta} \\ &\leq \frac{(\beta - \gamma^{-\beta}) + \gamma^{-\beta} - \beta}{\beta} = 0.\end{aligned}$$

The first inequality sign in the equation above follow from considering that

$$\Delta > \frac{n^\beta}{\alpha\beta} > \frac{(\gamma n - 1)^\beta}{\alpha\beta\gamma^\beta} \implies -\frac{\alpha\beta\gamma^\beta}{(\gamma n - 1)^\beta} \Delta \leq -1.$$

Similarly, the second inequality sign follows from

$$n^\beta \geq \frac{\gamma^{-\beta} - \beta}{1 - \gamma^\beta} \iff (\gamma^\beta - 1)n^\beta \leq (\beta - \gamma^{-\beta}).$$

This completes the argument that the partial derivative of $g(\Delta, n)$ with respect to Δ is nonpositive. We can deduce from this that

$$g(\Delta, n) \leq g\left(\frac{n^\beta}{\alpha\beta}, n\right) = f\left(\frac{n^\beta}{\alpha\beta}\right) \quad \text{for all } \Delta > \frac{n^\beta}{\alpha\beta}.$$

Now, it can be easily established that the bounding function has a global maximum. This follows immediately from the first part of the proof by noticing that we can write it in terms of f . To see why this is true, recall that we defined f and g in the same way but with different values for s_Δ . However, since we set $\Delta = \frac{n^\beta}{\alpha\beta}$ (in the bounding function) we get that $s_\Delta = \gamma(\alpha\beta\Delta)^{1/\beta}$ in the first case is the same as $s_\Delta = \gamma n$ in the second case. Therefore, the same function is produced which is bounded by the constant c_1 . \square

Clustering. We can partition the vertices of a sparse random shortest path metric into clusters of any size and use the previous lemmas to determine a probabilistic bound for the number of these clusters as a function of their diameter. The following theorem based on the ideas of Bringmann et al. [5] establishes this result.

Theorem 9. *Let G be a finite sparse graph on n vertices, consider a sparse random shortest path metric generated using this graph, and let $\Delta > 0$. There exists a partition of vertices into clusters, each of diameter at most 4Δ , such that the expected number of clusters needed is bounded from above by $O(1 + n/\Delta^{1/\beta})$.*

Proof. Let n be sufficiently large and let $s_\Delta := \min\{\gamma(\alpha\beta\Delta)^{1/\beta}, \gamma n\}$, as in Corollary 8. We call vertex v Δ -dense if $|B_\Delta(v)| \geq s_\Delta$ and Δ -sparse otherwise. Let the random variable $X_v = 1$ if v is Δ -sparse and be zero otherwise. Using Corollary 8 we can bound the expected number of Δ -sparse vertices by

$$\mathbb{E}\left[\sum_{v \in V} X_v\right] = \sum_{v \in V} \mathbb{P}(|B_\Delta(v)| < s_\Delta) \leq \sum_{v \in V} \frac{c_1}{\Delta^{1/\beta}} = \frac{c_1 n}{\Delta^{1/\beta}} = O(n/\Delta^{1/\beta}).$$

We put each Δ -sparse vertex in its own cluster (of size 1 and diameter $0 \leq 4\Delta$).

In order to group the Δ -dense vertices into clusters we consider an auxiliary graph H consisting of only the Δ -dense vertices. Two vertices u and v are connected in H by an edge if and only if $B_\Delta(u) \cap B_\Delta(v) \neq \emptyset$ in G . We construct our clusters by starting with an arbitrary maximal independent set S in H . By construction each vertex $w \in S$ is not connected to any other vertex $t \in S \setminus \{w\}$ in H so we have $B_\Delta(w) \cap B_\Delta(t) = \emptyset$. For each vertex w in S there are at least s_Δ vertices that belong to $B_\Delta(w)$ so we can deduce that $|S| \leq n/s_\Delta$. For each $s \in S$ we can construct initial non-intersecting clusters equal to $B_\Delta(s)$ of diameter at most 2Δ .

For any Δ -dense vertex v not yet belonging to any cluster there must be a $w \in S$ such that $A := B_\Delta(w) \cap B_\Delta(v) \neq \emptyset$ by the maximality of S (otherwise v would be in S). We now add v to the initial cluster corresponding to w . Observe that for any $x \in A$ we have that $d(v, w) \leq d(v, x) + d(x, w) \leq \Delta + \Delta = 2\Delta$. Therefore, after repeating this procedure for all vertices not belonging to any initial cluster the diameter of each cluster is extended to at most 4Δ . This follows from considering any two vertices u, v in a cluster that initially corresponded to $w \in S$, then we have $d(u, v) \leq d(u, w) + d(w, v) \leq 2\Delta + 2\Delta = 4\Delta$.

This procedure yields in expectation at most $O(n/\Delta^{1/\beta})$ clusters containing one (Δ -sparse) vertex each, and at most n/s_Δ clusters containing at least s_Δ (Δ -dense) vertices each, all with diameter at most 4Δ . We can write that

$$\begin{aligned} \frac{n}{s_\Delta} &= \begin{cases} \gamma^{-1}(\alpha\beta)^{-1/\beta} \cdot n/\Delta^{1/\beta} & \text{if } \alpha\beta\Delta^\beta \leq n, \\ \gamma^{-1} & \text{if } \alpha\beta\Delta^\beta \geq n \end{cases} \\ &\leq \max\{\gamma^{-1}(\alpha\beta)^{-1/\beta}, \gamma^{-1}\} \left(1 + n/\Delta^{1/\beta}\right) = O\left(1 + n/\Delta^{1/\beta}\right). \end{aligned}$$

Thus, the expected number of clusters needed is bounded from above by $O(n/\Delta^{1/\beta}) + O(1 + n/\Delta^{1/\beta})$. The result follows immediately. \square

Bound for the Diameter. Theorem 9 provides us with a useful bound on the number of cluster with diameter at most 4Δ , however for large Δ the tightest bound this yields is 1 since there is always at least one cluster. For the proofs of the next section we require a tighter bound for $\Delta > O(n)$. We address this by considering that the distance between any two vertices is at most the diameter of the metric: Δ_{\max} . The following Lemma shows that $\Delta_{\max} \leq O(n)$ with high probability.

Lemma 10. *Let $x \geq 6n$. Then we have $\mathbb{P}(\Delta_{\max} \geq x) \leq ne^{-x/2}$.*

Proof. We first show that $\mathbb{P}(\tau_n(v) \geq x) \leq e^{-x/2}$. In the proof of Lemma 6 we established that for a fixed arbitrary $v \in V$ we have

$$\tau_k(v) \sim \sum_{m=1}^{k-1} \text{Exp}(\chi_m(v)).$$

By connectedness of the underlying sparse graph we have that $\chi_k(v) \geq 1$ for all k . It follows that

$$\tau_n(v) \gtrsim \sum_{m=1}^{n-1} \text{Exp}(1), \quad \text{and thus} \quad \mathbb{P}(\tau_n(v) \geq x) \leq \mathbb{P}\left(\sum_{m=1}^{n-1} \text{Exp}(1) \geq x\right).$$

We again use Lemma 3 to bound this probability. First we set

$$\mu := \mathbb{E}\left[\sum_{m=1}^{n-1} \text{Exp}(1)\right] = n-1 \leq n \quad \text{and thus} \quad \lambda = \frac{x}{\mu} \geq \frac{x}{n} \geq 6.$$

Lemma 3 now yields

$$\mathbb{P}(\tau_n(v) \geq x) \leq \lambda^{-1} e^{-\mu(\lambda-1-\ln(\lambda))} \leq e^{-\mu(\lambda/2)} = e^{-x/2},$$

where we use that $\lambda - 1 - \ln(\lambda) \geq \lambda/2$ (which is true for all $\lambda \geq 5.36$) for the second inequality. The final result follows by using Boole's inequality:

$$\begin{aligned} \mathbb{P}(\Delta_{\max} \geq x) &= \mathbb{P}\left(\max_{v \in V} \{\tau_n(v)\} \geq x\right) = \mathbb{P}\left(\bigcup_{v \in V} \tau_n(v) \geq x\right) \\ &\leq \sum_{v \in V} \mathbb{P}(\tau_n(v) \geq x) \leq \sum_{v \in V} e^{-x/2} = ne^{-x/2} \end{aligned}$$

□

Notice that this proof works for any connected graph not just graphs with a fast growing cut size. It is possible to obtain a slightly stronger version of this result by using the property given by Definition 1 which is stronger than connectedness (for $|U| \leq \gamma n$).

4 Analysis of Heuristics

In this section we determine a bound for the expected approximation ratio of the greedy heuristic for the minimum-distance perfect matching problem as well as the nearest neighbour heuristic and insertion heuristic for the TSP.

Greedy Heuristic for Minimum-Distance Perfect Matching. The minimum-distance perfect matching problem tends to be slightly easier to analyse than other problems, like the TSP, and is usually considered first. This optimisation problem is not too ‘difficult’ as it can be solved to optimality in polynomial time: $O(n^3)$. However, it remains useful to analyse heuristic algorithms as they provide an attractive alternative for very large instances of the problem. It also forms a stepping stone for analysing heuristic algorithms for other problems, such as for the TSP.

Arguably the simplest heuristic which performs remarkably well in practice is the greedy heuristic. It starts with a set of unmatched vertices and adds the pair of vertices in the metric which have the minimal distance. This pair is then removed from the set and the next closest pair of vertices is considered. The process is repeated until all vertices have been matched. Notice that this requires n to be even, so we quietly assume that this is the case whenever we consider the minimum-distance perfect matching problem.

Let GR denote the total distance of the matching produced by the greedy heuristic, and let MM denote the total distance of an optimal matching. To compare the performance of the algorithm with the optimal solution we consider the approximation ratio $\frac{\text{GR}}{\text{MM}} \geq 1$. The worst-case approximation ratio for this heuristic on metric instances is known to be $O(n^{\log_2(3/2)})$ [11]. For a variety of metrics it has been shown that the expected approximation ratio is $O(1)$. These include random Euclidean instances [1], and random shortest path metrics generated from complete graphs, Erdős–Rényi random graphs [5, 10] and (sparse) square grid graphs [9]. We generalise the last of these findings to show that a similar result holds for random shortest path metrics generated from sparse graphs with a fast growing cut size.

Theorem 11. $\mathbb{E}[\text{GR}] = O(n)$.

Proof. We partition the run of the greedy heuristic into phases, starting from phase 0. During phase i pairs of vertices $\{u, v\}$ are added to the matching which satisfy $d(u, v) \in (4i, 4(i+1)]$. Let X_i denote the number of vertex pairs that are added during phase i . It will also become useful to consider the number of unmatched vertex pairs at the start of phase i which is given by $Y_i := \sum_{j=i}^{\infty} X_j$. This is the number of vertex pairs $\{u, v\}$ in the final greedy matching which satisfy $d(u, v) > 4i$. Observe that we always consider pairs of vertices, so $Y_0 = n/2$. By Theorem 9, we can cluster the vertices during phase i into an expected number of $O(1 + n/i^{1/\beta})$ clusters each of diameter at most $4i$. Just before the greedy heuristic adds the first vertex pair in phase i it must be the case that each of these clusters contains at most one unmatched vertex. Otherwise there exists a vertex pair $\{u, v\}$ within a cluster, but this pair would have been chosen by the greedy heuristic in the previous phase since $d(u, v) \leq 4i$. Therefore, the expected number of unmatched vertices at the start of phase i is bounded from above: $\mathbb{E}[Y_i] \leq O(1 + n/i^{1/\beta})$ for $i > 0$. For any phase where $4i \geq 6n$, it follows from Lemma 10 that $\mathbb{E}[Y_i] \leq (n/2) \cdot \mathbb{P}(\Delta_{\max} \geq 4i) \leq n^2 e^{-2i}$. We are now ready to sum over all phases. Since the distance between any vertex pair that

is added in phase i is at most $4(i+1)$ we can bound $\mathbb{E}[\text{GR}]$ as follows:

$$\begin{aligned}
\mathbb{E}[\text{GR}] &\leq \sum_{i=0}^{\infty} 4(i+1) \cdot \mathbb{E}[X_i] \\
&= \sum_{i=1}^{\infty} 4i \cdot \mathbb{E}[X_{i-1}] = \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} 4 \cdot \mathbb{E}[X_{i-1}] = \sum_{i=1}^{\infty} 4 \cdot \mathbb{E}[Y_{i-1}] \\
&= 4 \cdot \mathbb{E}[Y_0] + \sum_{i=1}^{3n/2} 4 \cdot \mathbb{E}[Y_i] + \sum_{3n/2+1}^{\infty} 4 \cdot \mathbb{E}[Y_i] \\
&\leq 2n + \sum_{i=1}^{3n/2} O\left(1 + \frac{n}{i^{1/\beta}}\right) + \sum_{3n/2+1}^{\infty} 4n^2 e^{-2(i-1)}.
\end{aligned}$$

Here, the second equality sign follows from Fubini's theorem. Notice that the last summation term quickly converges by the ratio test and approaches 0 as $n \rightarrow \infty$, hence it becomes $o(1)$. It remains to explicitly show that

$$\sum_{i=1}^{3n/2} O\left(1 + \frac{n}{i^{1/\beta}}\right) = O(n).$$

By definition of the O -operator there exists constants c , n_0 and i_0 such that whenever $n \geq n_0$ and $i \geq i_0$ we have

$$\sum_{i=1}^{3n/2} O\left(1 + \frac{n}{i^{1/\beta}}\right) \leq \sum_{i=1}^{3n/2} c\left(1 + \frac{n}{i^{1/\beta}}\right)$$

We may assume without loss of generality that $n_0 = i_0 = 1$. The last term reduces to $O(n)$ as follows:

$$\begin{aligned}
\sum_{i=1}^{3n/2} c\left(1 + \frac{n}{i^{1/\beta}}\right) &= \frac{3}{2}cn + cn \sum_{i=1}^{3n/2} \frac{1}{i^{1/\beta}} = \frac{5}{2}cn + cn \sum_{i=2}^{3n/2} \frac{1}{i^{1/\beta}} \\
&\leq \frac{5}{2}cn + \frac{\beta}{1-\beta} \left(1 - \left(\frac{3}{2}n\right)^{1-1/\beta}\right) cn \\
&\leq \frac{5}{2}cn + \frac{\beta}{1-\beta} cn = \left(\frac{5-3\beta}{2(1-\beta)}c\right)n = O(n)
\end{aligned}$$

Here, the first inequality sign follows by considering the summation as a Riemann sum which is bounded by the integral

$$\int_1^{3n/2} \frac{1}{x^{1/\beta}} dx = \frac{\beta x^{-1/\beta+1}}{\beta-1} \Big|_1^{3n/2} = \frac{\beta}{1-\beta} \left(1 - \left(\frac{3}{2}n\right)^{1-1/\beta}\right).$$

Notice that we use the condition that $\beta \in (0, 1)$. Combining everything we get that $\mathbb{E}[\text{GR}] = O(n) + O(n) + o(1)$ from which the result immediately follows. \square

Theorem 12. *For random shortest path metrics generated from a finite sparse graph with a fast growing cut size we have $\mathbb{E} \left[\frac{\text{GR}}{\text{MM}} \right] = O(1)$.*

Proof. Let $c > 0$ be a sufficiently small constant. By conditioning the expectation on whether $\text{MM} \geq cn$ or $\text{MM} < cn$ we can use the result from Theorem 11 together with the worst-case approximation ratio of the greedy heuristic on metric instances to get the desired result. The expected approximation ratio for sparse graphs with a fast growing cut size is given by

$$\begin{aligned} \mathbb{E} \left[\frac{\text{GR}}{\text{MM}} \right] &= \mathbb{P}(\text{MM} < cn) \cdot \mathbb{E} \left[\frac{\text{GR}}{\text{MM}} \mid \text{MM} < cn \right] + \mathbb{P}(\text{MM} \geq cn) \cdot \mathbb{E} \left[\frac{\text{GR}}{\text{MM}} \mid \text{MM} \geq cn \right] \\ &\leq \mathbb{P}(\text{MM} < cn) \cdot \mathbb{E} \left[\frac{\text{GR}}{\text{MM}} \mid \text{MM} < cn \right] + \mathbb{P}(\text{MM} \geq cn) \cdot \mathbb{E} \left[\frac{\text{GR}}{cn} \mid \text{MM} \geq cn \right] \\ &\leq \mathbb{P}(\text{MM} < cn) \cdot \mathbb{E} \left[\frac{\text{GR}}{\text{MM}} \mid \text{MM} < cn \right] + \mathbb{E} \left[\frac{\text{GR}}{cn} \right]. \end{aligned} \quad (2)$$

Similarly to the first step, the last inequality also follows from conditioning:

$$\mathbb{E} \left[\frac{\text{GR}}{cn} \right] = \mathbb{P}(\text{MM} \geq cn) \cdot \mathbb{E} \left[\frac{\text{GR}}{cn} \mid \text{MM} \geq cn \right] + \mathbb{P}(\text{MM} \leq cn) \cdot \mathbb{E} \left[\frac{\text{GR}}{cn} \mid \text{MM} \leq cn \right].$$

By combining Lemmas 4 and 5 we get that the first term in Equation (2) satisfies

$$\mathbb{P}(\text{MM} \leq cn) \leq \mathbb{P}(S_{n/2} \leq cn) \leq \exp \left(\frac{n}{2} \left(2 + \ln \left(\frac{4c|E|}{n} \right) \right) \right).$$

Since the worst-case approximation ratio of the greedy heuristic on metric instances is $O(n^{\log_2(3/2)})$ [11], we can bound the second term in Equation (2):

$$\mathbb{E} \left[\frac{\text{GR}}{\text{MM}} \mid \text{MM} < cn \right] \leq O(n^{\log_2(3/2)})$$

Finally from Theorem 11 it follows that the third term in Equation (2) is simply $O(1)$. Together this yields

$$\mathbb{E} \left[\frac{\text{GR}}{\text{MM}} \right] \leq \exp \left(n \left(1 + \frac{1}{2} \ln \left(\frac{4c|E|}{n} \right) \right) \right) \cdot O \left(n^{\log_2(3/2)} \right) + O(1).$$

We can choose c sufficiently small to ensure $1 + (1/2) \ln(4c|E|/n) < 0$ (in particular $c \in (0, n/(4e^2|E|))$), which is a finite nonempty interval since $|E| = \Theta(n)$. Since we have that, for growing n , $e^{-pn} \rightarrow 0$ much faster than $n^q \rightarrow \infty$ for any positive constants p and q , the product of the first two terms reduces to $o(1)$. This completes the proof. \square

Nearest Neighbour Heuristic for TSP. The nearest neighbour heuristic for the TSP is a greedy-like heuristics that can be analysed in a very similar way to the greedy heuristic for the minimum-distance perfect matching problem. It works by starting with an arbitrary vertex which it adds to the TSP tour. From there it iteratively adds the closest unvisited vertex until all vertices have been added to the tour, at which point the tour is closed by going back to the starting vertex. For each successive pair of vertices $\{u, v\}$ visited by the tour the distance $d(u, v)$ is added to the tour length. We use **NN** to denote the total

tour length computed by the nearest neighbour heuristic and let TSP be the length of the optimal tour.

For this heuristic it is known that the worst-case approximation ratio on metric instances is $O(\ln(n))$ [12]. For a variety of metrics it has been shown that the expected approximation ratio is $O(1)$. These include random Euclidean instances [2], and random shortest path metrics generated from complete graphs, Erdős–Rényi random graphs [5, 10] and (sparse) square grid graphs [9]. We generalise the last of these findings to show that a similar result holds for random shortest path metrics generated from sparse graphs with a fast growing cut size.

Theorem 13. $\mathbb{E}[\text{NN}] = O(n)$.

Proof. We group vertex pairs $\{u, v\}$ that are being added to the TSP tour by the nearest neighbour heuristic according to their distance $d(u, v)$. (Note that we avoid calling these edges as they are not edges in the original sparse graph but ‘edges’ in the random metric). Let group i contain the set of vertex pairs $\{u, v\}$ that satisfy $d(u, v) \in (4i, 4(i + 1)]$ and let X_i denote the number of vertex pairs that are added to group i . We also consider the number of vertex pairs $\{u, v\}$ in the final tour which satisfy $d(u, v) > 4i$ which is given by $Y_i := \sum_{j=i}^{\infty} X_j$. Observe that $Y_0 = n$.

Consider that the run of the nearest neighbour heuristic is currently at some vertex u and it adds a vertex v to the tour with $d(u, v) \in (4i, 4(i + 1)]$. By Theorem 9, we can partition all vertices into clusters of diameter at most $4i$. It must be the case that $u \in C_j$ and $v \in C_k$ for some arbitrary clusters $C_j \neq C_k$ of diameter $4i$ and that all the vertices in C_j have already been visited by the tour (otherwise the tour would have added an unvisited vertex from C_j instead). It follows that a vertex pair with distance exceeding $4i$ will be added to the tour as many times as there are clusters of diameter at most $4i$. By Theorem 9, there are an expected number of $O(1 + n/i^{1/\beta})$ such clusters. This yields that $\mathbb{E}[Y_i] \leq O(1 + n/i^{1/\beta})$ for $i > 0$. For edges with distance $4i \geq 6n$, it follows from Lemma 10 that $\mathbb{E}[Y_i] \leq n \cdot \mathbb{P}(\Delta_{\max} \geq 4i) \leq n^2 e^{-2i}$. Notice that we have derived exactly the same bounds as in the proof of Theorem 11. Therefore, we can use the same calculations to arrive at $\mathbb{E}[\text{NN}] = O(n)$. \square

Theorem 14. *For random shortest path metrics generated from a finite sparse graphs with a fast growing cut size we have $\mathbb{E} \left[\frac{\text{NN}}{\text{TSP}} \right] = O(1)$.*

Proof. The proof of this theorem is nearly identical to the proof of Theorem 12. The only difference being that the worst-case approximation ratio of the nearest neighbour heuristic on metric instances is $O(\ln(n))$ [12]. Let $c > 0$ be a sufficiently small constant. Then the approximation ratio of the nearest neighbour heuristic on random shortest path metrics generated from a sparse graph with a fast growing cut size is

$$\mathbb{E} \left[\frac{\text{NN}}{\text{TSP}} \right] \leq \mathbb{P}(\text{TSP} < cn) \cdot O(\ln(n)) + \mathbb{E} \left[\frac{\text{NN}}{cn} \right],$$

since the worst-case approximation ratio of the nearest neighbour heuristic on metric instances is $O(\ln(n))$. Combining Lemmas 4 and 5, the first term can be bounded from above by $\exp(n(1 + \frac{1}{2} \ln(c \cdot \Theta(1)))) \cdot O(\ln(n)) = o(1)$ since c is sufficiently small. By Theorem 13 the second term is $O(1)$. \square

Insertion Heuristics for TSP. The insertion heuristics for the TSP is another greedy-like heuristics that can be analysed in a similar manner to the other two heuristics. The heuristic works by starting with some optimal tour on a small subset of the vertices and iteratively inserts a vertex which is not yet in the tour. The starting tour can be chosen arbitrarily or according to some predefined rule. The inserted vertex is chosen according to some rule which we denote R . To illustrate one such rule consider nearest insertion which inserts the vertex whose minimal distance to a vertex already in the tour is minimal. Random insertion, Farthest insertion and Cheapest insertion are other such rules. We use IN_R to denote the total tour length computed by the insertion heuristic using rule R . Let TSP be the length of the optimal tour.

For this heuristic it is known that the worst-case approximation ratio on metric instances is $O(\ln(n))$ [12]. For a variety of metrics it has been shown that the expected approximation ratio is $O(1)$. These include random shortest path metrics generated from complete graphs, Erdős–Rényi random graphs [5, 10] and (sparse) square grid graphs [9]. We generalise the last of these findings to show that a similar result holds for random shortest path metrics generated from sparse graphs with a fast growing cut size.

Theorem 15. $\mathbb{E}[\text{IN}_R] = O(n)$.

Proof. We group all the steps of the insertion heuristic according to the distance which they add to the tour. Let group i contain the set of vertices whose addition to the tour is in the range $(8i, 8(i+1)]$ and let X_i denote the number of vertices that are added to group i . We also define $Y_i := \sum_{j=i}^{\infty} X_j$. Observe that $Y_0 = n$.

Consider that the insertion heuristic adds some vertex v to the tour whose contribution to the tour length is in the range $(8i, 8(i+1)]$. By Theorem 9, we can partition all vertices into clusters of diameter at most $4i$. It must be the case that v is part of some arbitrary cluster which contains no vertex that is already part of the tour, otherwise the contribution made to the tour by adding v would be less than $8i$. It follows that a vertex whose contribution to the tour is in the range $(8i, 8(i+1)]$ will be added as many times as there are clusters of diameter at most $4i$. By Theorem 9, there are an expected number of $O(1 + n/i^{1/\beta})$ such clusters. This yields that $\mathbb{E}[Y_i] \leq O(1 + n/i^{1/\beta})$ for $i > 0$. For vertices whose contribution to the tour is $8i \geq 6n$, it follows from Lemma 10 that $\mathbb{E}[Y_i] \leq n \cdot \mathbb{P}(\Delta_{\max} \geq 8i) \leq n^2 e^{-4i}$. We are now ready to sum over all groups. Since the additional contribution of any vertex corresponding to group i is at most $8(i+1)$ we can bound $\mathbb{E}[\text{GR}]$ as follows, where the contribution of the initial tour is denoted as $\mathbb{E}[T_{\text{init}}]$:

$$\begin{aligned}
\mathbb{E}[\text{IN}_R] &\leq \mathbb{E}[T_{\text{init}}] + \sum_{i=0}^{\infty} 8(i+1) \cdot \mathbb{E}[X_i] = \mathbb{E}[T_{\text{init}}] + \sum_{i=1}^{\infty} 8 \cdot \mathbb{E}[Y_{i-1}] \\
&= \mathbb{E}[T_{\text{init}}] + 8 \cdot \mathbb{E}[Y_0] + \sum_{i=1}^{3n/4} 8 \cdot \mathbb{E}[Y_i] + \sum_{3n/4+1}^{\infty} 8 \cdot \mathbb{E}[Y_i] \\
&\leq O(n) + 8n + \sum_{i=1}^{3n/4} O\left(1 + \frac{n}{i^{1/\beta}}\right) + \sum_{3n/4+1}^{\infty} 8n^2 e^{-4(i-1)} \\
&= O(n) + O(n) + O(n) + o(1) = O(n).
\end{aligned}$$

Here, we used Theorem 13 to bound the expected contribution of the initial tour by $\mathbb{E}[T_{\text{init}}] \leq \mathbb{E}[\text{TSP}] \leq \mathbb{E}[\text{NN}] = O(n)$. Notice that the order in which we add closer or

farther vertices does not matter, consequently the proof is independent of the choice of rule R . Some of the details of the proof are omitted as they closely resemble those of Theorem 11.

□

Theorem 16. *For random shortest path metrics generated from a finite sparse graph with a fast growing cut size we have $\mathbb{E} \left[\frac{\text{IN}_R}{\text{TSP}} \right] = O(1)$.*

Proof. We omit a detailed proof of this theorem. Since the worst-case approximation ratio of the insertion heuristic on metric instances is also $O(\ln(n))$ [12], the proof is mathematically identical to that of Theorem 14. The only difference would be that we replace NN by IN_R and utilise Theorem 15 instead of Theorem 13. □

5 Conclusion

As proposed by Klootwijk and Manthey [9] it is indeed possible to generalise the analyses of greedy-like heuristics for minimum-distance perfect matching and the TSP on random shortest path metrics. Utilising the same techniques we were able to generalise their result for square grid graphs to the broader class of sparse graphs with a fast growing cut size. We showed that three heuristics achieve a constant expected approximation ratio on metrics generated from these graphs.

While the class of sparse graphs with a fast growing cut size is quite encompassing, there are still classes of sparse graphs that do not satisfy this property. For example the family of ‘line’ graphs: any d -dimensional grid graph that grows only along a single dimension as n grows and therefore has $\epsilon = 0$ by Definition 1. We believe that a similar result to the one we have given should hold for all sparse graphs. Evidently, some of the techniques we have used would fail if for example, we set $\epsilon = 0$, so a new approach may be necessary to show this. It remains an open question whether it is possible to extend our findings to arbitrary sparse graphs.

All in all, the analysis done in this thesis contributes non-trivial results about the expected performance of three heuristic algorithms. It takes us one step closer to consolidating empirical observations of these heuristics with a rigorous mathematical understanding thereof. Nevertheless, more research is valuable and necessary to obtain more profound theoretical insights into the behaviour of heuristic algorithms.

References

- [1] 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(2):143–156, 1988.
- [2] Jon Louis Bentley and James Benjamin Saxe. An analysis of two heuristics for the Euclidean traveling salesman problem. In *Proceedings of the Eighteenth Annual Allerton Conference on Communication, Control, and Computing*, pages 41–49, 1980.
- [3] Béla Bollobás and Imre Leader. Edge-isoperimetric inequalities in the grid. *Combinatorica*, 11(4):299–314, Dec 1991.
- [4] Jean-Louis Bon and Eugen Păltănea. Ordering properties of convolutions of exponential random variables. *Lifetime Data Analysis*, 5(2):185–192, Jun 1999.
- [5] Karl Bringmann, Christian Engels, Bodo Manthey, and B. V. Raghavendra Rao. Random shortest paths: Non-euclidean instances for metric optimization problems. *Algorithmica*, 73(1):42–62, Sep 2015.
- [6] Robert Davis and Armand Prieditis. The expected length of a shortest path. *Information Processing Letters*, 46(3):135 – 141, 1993.
- [7] Svante Janson. Tail bounds for sums of geometric and exponential variables. *Statistics & Probability Letters*, 135:1 – 6, 2018.
- [8] Richard M. Karp and J. Michael Steele. Probabilistic analysis of heuristics. *The traveling salesman problem*, pages 181–205, 1985.
- [9] Stefan Klootwijk and Bodo Manthey. Probabilistic Analysis of Optimization Problems on Sparse Random Shortest Path Metrics. In Michael Drmota and Clemens Heuberger, editors, *31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2020)*, volume 159 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 19:1–19:16, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.
- [10] Stefan Klootwijk, Bodo Manthey, and Sander K. Visser. Probabilistic analysis of optimization problems on generalized random shortest path metrics. In Gautam K. Das, Partha S. Mandal, Krishnendu Mukhopadhyaya, and Shin-ichi Nakano, editors, *WALCOM: Algorithms and Computation*, pages 108–120, Cham, 2019. Springer International Publishing.
- [11] Edward M. Reingold and Robert E. Tarjan. On a greedy heuristic for complete matching. *SIAM Journal on Computing*, 10(4):676–681, 1981.
- [12] 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.

A Proof of Lemmas 4 and 5

The proofs of Lemma 4 and Lemma 5 by Klootwijk and Manthey [9] utilise four prior results which we have also included in this section. The first corollary follows directly from Lemma 3.

Corollary A.1. *Let $X \sim \sum_{i=1}^m \text{Exp}(a_i)$. Let $\mu = \mathbb{E}[X] = \sum_{i=1}^m 1/a_i$ and $a_* = \min_i a_i$. For any x ,*

$$\mathbb{P}(X \leq x) \leq \exp(a_* \mu (1 + \ln(x/\mu))).$$

Proof. Let $\lambda := x/\mu$. If $\lambda \leq 1$, the result is a weaker version of Lemma 3. If $\lambda > 1$, then $1 + \ln(x/\mu) > 0$ and hence $\mathbb{P}(X \leq x) \leq 1 < \exp(a_* \mu (1 + \ln(x/\mu)))$. \square

Lemma A.2 ([4, Thm. 2(ii)]). *Let $X \sim \sum_{i=1}^m \text{Exp}(\lambda_i)$ and $Y \sim \sum_{i=1}^m \text{Exp}(\eta)$. Then*

$$X \lesssim Y \quad \text{if and only if} \quad \prod_{i=1}^m \lambda_i \leq \eta^m.$$

Lemma A.3 ([9, Lem. 4]). *Let S_m denote the sum of the m lightest edge weights in G . Then*

$$\sum_{i=0}^{m-1} \text{Exp}\left(\frac{e|E|}{m}\right) \lesssim S_m \lesssim \sum_{i=0}^{m-1} \text{Exp}\left(\frac{|E|}{m}\right).$$

Proof. Let σ_k denote the k th lightest edge weight in G . Since all edge weights are independent and standard exponentially distributed, we have $\sigma_1 = S_1 \sim \text{Exp}(|E|)$. Using the memorylessness property of the exponential distribution, it follows that $\sigma_2 \sim \sigma_1 + \text{Exp}(|E| - 1)$, i.e., the second lightest edge weight is equal to the lightest edge weight plus the minimum of $|E| - 1$ standard exponential distributed random variables. In general, we get $\sigma_{k+1} \sim \sigma_k + \text{Exp}(|E| - k)$. The definition $S_m = \sum_{k=1}^m \sigma_k$ yields

$$S_m \sim \sum_{i=0}^{m-1} (m-i) \cdot \text{Exp}(|E| - i) \sim \sum_{i=0}^{m-1} \text{Exp}\left(\frac{|E| - i}{m - i}\right).$$

Now, the first stochastic dominance relation follows from Lemma A.2 by observing that

$$\prod_{i=0}^{m-1} \frac{|E| - i}{m - i} = \frac{|E|!}{m!(|E| - m)!} = \binom{|E|}{m} \leq \left(\frac{e|E|}{m}\right)^m,$$

where the inequality follows from applying the well-known inequality $\binom{n}{k} \leq (en/k)^k$. The second stochastic dominance relation follows by observing that $|E| \geq m$, which implies that $(|E| - i)/(m - i) \geq |E|/m$ for all $i = 0, \dots, m - 1$. \square

Corollary A.4 ([9, Cor. 5]). *Let S_m denote the sum of the m lightest edge weights in G . Then $\mathbb{E}[S_m] = \Theta(m^2/n)$.*

Proof. From Lemma A.3 we can immediately see that

$$\mathbb{E} \left[\sum_{i=0}^{m-1} \text{Exp} \left(\frac{e|E|}{m} \right) \right] \leq \mathbb{E} [S_m] \leq \mathbb{E} \left[\sum_{i=0}^{m-1} \text{Exp} \left(\frac{|E|}{m} \right) \right].$$

The result follows by observing that

$$\mathbb{E} \left[\sum_{i=0}^{m-1} \text{Exp} \left(\frac{e|E|}{m} \right) \right] = \frac{m^2}{e|E|} \quad \text{and} \quad \mathbb{E} \left[\sum_{i=0}^{m-1} \text{Exp} \left(\frac{|E|}{m} \right) \right] = \frac{m^2}{|E|},$$

and recalling that $|E| = \Theta(n)$ by our restrictions imposed on G . \square

We are now ready to give the proofs of Lemma 4 and Lemma 5 respectively.

Lemma 4 ([9, Lem. 6]). *Let S_m denote the sum of the m lightest edge weights in G . Then we have*

$$\mathbb{P}(S_m \leq cn) \leq \exp \left(m \left(2 + \ln \left(\frac{c|E|n}{m^2} \right) \right) \right).$$

Proof. First of all, Lemma A.3 yields

$$S_m \lesssim \sum_{i=0}^{m-1} \text{Exp} \left(\frac{e|E|}{m} \right).$$

Now, we apply Corollary A.1 with $\mu = m^2/e|E|$, $a_* = e|E|/m$, and $x = cn$ to obtain

$$\mathbb{P}(S_m \leq cn) \leq \mathbb{P} \left(\sum_{i=0}^{m-1} \text{Exp} \left(\frac{e|E|}{m} \right) \leq cn \right) \leq \exp \left(m \left(1 + \ln \left(\frac{ce|E|n}{m^2} \right) \right) \right).$$

The result follows immediately. \square

Lemma 5 ([9, Lem. 7]). *Let S_m denote the sum of the m lightest edge weights in G . Then we have $\text{TSP} \geq \text{MM} \geq S_{n/2}$, where TSP and MM are the total distance of a shortest TSP tour and a minimum-distance perfect matching, respectively.*

Proof. The first inequality is trivial. For the second one, consider a minimum-distance perfect matching in G , and take the union of the shortest paths between each pair of matched vertices. This union H must contain at least $n/2$ different edges of G , since H is a forest with n vertices, all of which are non-isolated in H . These $n/2$ different edges together have a weight of at least $S_{n/2}$ and at most MM . So, the second inequality follows. \square