May 30, 2018

MASTER THESIS – APPLIED MATHEMATICS

PROBABILISTIC ANALYSIS OF OPTIMIZATION PROBLEMS IN RANDOM SHORTEST PATH METRICS APPLIED TO ERDŐS–RÉNYI RANDOM GRAPHS

Sander Visser

Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS)
Discrete Mathematics and Mathematical Programming (DMMP)

Graduation committee:

S. Klootwijk, MSc
Dr. B. Manthey
Dr. ir. W.R.W. Scheinhardt
Prof. dr. M.J. Uetz

UNIVERSITY OF TWENTE.

Abstract

Probabilistic analysis for metric optimization problems has mostly been conducted on instances from the Euclidean space. Little analysis has been done on metric instances drawn from other distributions. We want to extend the probabilistic analysis of optimization problems to more general metrics, since these might provide a better resemblance to real-world instances of problems. In this thesis we use random shortest path metrics applied to Erdős–Rényi random graphs. An Erdős–Rényi random graph is constructed by including an edge between every pair of vertices independently with probability p. A random shortest path metric is then constructed by drawing independent random edge weights for each edge in the graph and setting the distance between every pair of vertices to the length of a shortest path between them with respect to the drawn weights. For these metrics, we prove that the expected approximation ratios of the greedy heuristic for the minimum distance maximum matching problem, of the nearest neighbor and insertion (independent of the insertion rule) heuristics for the traveling salesman problem, and of the trivial heuristic for the k-median problem all have a constant upper bound. Additionally, we show an upper bound of $O(n^8 \ln^3(n))$ for the expected number of iterations of the 2-opt heuristic for the traveling salesman problem.

Contents

1 Introduction

Networks can be found all around us. Examples are the internet, social media and rail transport. In these networks, a variety of optimization problems arises. In order to solve these optimization problems, people have tried to formulate mathematical models that can be used to describe real-world instances of these problems. This is mostly done by describing a network as a graph $G = (V, E)$, where every edge of the graph is assigned a weight $w(e)$. In the example of rail transport, vertices could be stations, edges could be rails connecting stations and the weight of the edges could be the distances or travel times between the stations.

These mathematical models are used to analyze large-scale optimization problems. However, many of these problems are NP-hard. An example is the traveling salesman problem, for which the goal is to find a route of minimal length through all vertices of a graph. A consequence of problems being NP-hard is that optimal solutions can probably not be found within a reasonable amount of time for large-scale instances. In order to still find reasonable results in short time, heuristics and their worst-case performance are often researched for (NP-hard) optimization problems.

Even though many simple heuristics have a poor worst-case performance, in practice they often return much better or even near-optimal solutions. The worst-case performance tends to give a distorted view, because instances that lead to the worst-case performance are often unlikely to appear in practical situations. Since the performance of these heuristics cannot be explained by their worst-case performance, probabilistic analysis is used as an alternative. Probabilistic analysis starts from an assumption about a probability distribution over the set of all possible instances. The goal is to find the average-case performance of the heuristic, which might give a better explanation of the performance of the heuristics in practice.

Probabilistic analysis of optimization problems has mostly been conducted on instances with independent random edge lengths (the triangle inequality does not necessarily need to hold in this case) and instances from the Euclidean space. These instances are relatively easy to analyze from a mathematical point of view, because of the independence of the edge lengths in the first case and the structure of the Euclidean space in the second case. However, these instances often do not represent real-world instances very well. Therefore, we want to apply probabilistic analysis to more general metric instances. Bringmann et al. [4] have used a model of random shortest path metrics applied to a complete graph for this. A random shortest path metric is constructed by drawing independent random edge weights from a probability distribution for all edges of a graph. The distance between every pair of vertices is then given by the shortest path between those vertices with respect to the edge weights. Such metrics are harder to analyze, since distances between vertices are no longer independent.

In this thesis we generalize the model used by Bringmann et al. [4] to non-complete graphs. We use Erdős–Rényi random graphs ($G(n, p)$ model) instead of complete graphs. An instance in this model is a graph with n vertices, and between every pair of vertices there exists an edge independently with probability p. This results in a complete graph if $p = 1$. A random shortest path metric based on this graph is created analogous to the the way described above.

A probabilistic analysis for simple heuristics has been executed for several optimization problems in random shortest path metrics applied to complete graphs. Among these problems are the minimum distance perfect matching problem, the traveling salesman problem and the k-median problem. In this thesis we extend known results for these problems from random shortest path metrics applied to complete graphs to random shortest path metrics applied to Erdős–Rényi random graphs.

1.1 Related work

The model for random shortest path metrics is also known as first passage percolation. First passage percolation was first introduced by Hammersley and Welsh as a model of fluid flow through a randomly porous material [6]. When first passage percolation was introduced, it was applied to to the multidimensional grid $\mathbb{Z}^d$ [8]. Since then other graphs have also been analyzed.

For first passage percolation in complete graphs, the expected distance between two fixed vertices is approximately $\ln(n)/n$ and the expected distance from a fixed vertex to the vertex that is most distant is approximately $2\ln(n)/n$ [4, 9]. Furthermore, the expected longest distance in the metric is approximately $3\ln(n)/n$ [7, 9]. There are also some known structural properties of first passage percolation on the Erdős–Rényi random graph. Bhamidi et al. [3] have shown asymptotics for both the minimal weight of the path between uniformly chosen vertices in the giant component and for the hopcount, the number of edges, on this path.

In this thesis we consider three optimization problems. The first one is the minimum distance perfect matching problem. The goal of this problem is to find a perfect matching in a graph that minimizes the total distance between all matched vertices. The best known algorithm for this problem has a running time of $O(n^3)$, which is slow for very large instances. Reingold and Tarjan [15] analyzed a greedy heuristic for the minimum distance perfect matching problem on general metric instances. They have shown that this heuristic has a worst-case approximation ratio of $O(n^{\log_2(3/2)}) \approx O(n^{0.59})$ on general metric instances. Avis et al. [2] analyzed the performance of the greedy heuristic for random Euclidean instances. They have found an upper bound for the approximation ratio of $O(1)$ in such metrics with high probability.

Secondly, we consider the traveling salesman problem. The goal of this problem is to find a route of minimal distance that passes all vertices of a graph and starts and ends at the same vertex. The traveling salesman problem is known to be NP-hard [11]. Because of this, Rosenkrantz et al. [16] analyzed several heuristics for the traveling salesman problem. For general metric spaces, they have proven worst-case approximation ratios of $O(\ln(n))$ for the nearest neighbor heuristic and the insertion heuristic, independent of the insertion rule.

The 2-opt heuristic for the traveling salesman problem has been analyzed by Englert et al. [5]. They were able to bound the number of iterations the heuristic takes before the algorithm terminates by $O(n^{4+1/3} \cdot \phi^{8/3})$, where ϕ depends on the probability distributions used. If an initial tour is constructed with an insertion heuristic before 2-opt is applied, then the bound improves.

Finally we consider the k-median problem. The goal of this problem is to find a set of k vertices in a graph such that the total distance from all other vertices to this set is minimized. Megiddo and Supowit [12] have shown that the k-median problem is NP-hard. A heuristic for this problem was analyzed by Arya et al. [1]. This heuristic is a swap local search where p vertices are removed from the chosen set and p others are added simultaneously. They have shown this heuristic has a worst-case approximation ratio of $3 + 2/p$ in general metric spaces.

The probabilistic analysis of the behavior of simple heuristics in random shortest path metrics applied to complete graphs has been initiated by Bringmann et al. [4]. They have proven several structural properties of such metrics. Additionally, for such random metrics they have proven an upper bound of $O(1)$ for the expected approximation ratios of the greedy heuristic for the minimum distance perfect matching problem, of the nearest neighbor and insert heuristic for the traveling salesman problem and of the trivial heuristic for the k-median problem. Additionally, they have bounded the expected number of iterations of 2-opt from above by $O(n^8 \ln(n))$.

1.2 Outline of this thesis

The structure of the remainder of this thesis is as follows. In Section 2 we first give a proper definition of the Erdős–Rényi random graph model and random shortest path metrics. Afterwards, basic notation used throughout this thesis is introduced. Section 3 contains several technical lemmas that provide no new structural insights, but they are needed for the proofs of lemmas and theorems in subsequent sections.

Sections 4–6 are the core of this thesis. We begin Section 4 by introducing a property of graphs. We call this property the α-β-cut-property. The section continues with proofs of several structural properties of random shortest path metrics applied to graphs that satisfy this α-β-cut-property.

In Section 5 we introduce the optimization problems and corresponding heuristics that we analyze. We introduce the heuristic `greedy` for the minimum distance perfect matching problem, the heuristics `nearest neighbor`, `insert` and `2-opt` for the traveling salesman problem and the heuristic `trivial` for the k-median problem. Probabilistic analysis is done for these heuristics in random shortest path metrics applied to graphs that satisfy the α-β-cut-property. For such metrics, we find an upper bound for the expected number of iterations `2-opt` takes before it terminates and upper bounds for the expected approximation ratios of `greedy`, `nearest neighbor`, `insert` and `trivial`.

The main results of this thesis are derived in Section 6. We first show that instances of the Erdős–Rényi random graph model satisfy the α-β-cut-property for suitable values of α and β with high probability. We use this to prove the existence of a constant upper bound for the expected approximation ratios of `greedy`, `nearest neighbor`, `insert` and `trivial` and we prove an upper bound of $O(n^8 \ln^3(n))$ for the expected number of iterations of `2-opt`.

In Section 7 we summarize the results of this thesis and compare them to the previous results. We conclude with a discussion and some final remarks in Section 8.

2 Model and notation

In this section we first introduce the Erdős Rényi Random Graph model (ERRG model) and random shortest path metrics (RSPM), including an example. Afterwards, the notation used throughout the rest of this thesis is introduced, where we follow the notation introduced by Bringmann et al. [4].

2.1 Erdős–Rényi Random Graph

The Erdős–Rényi Random Graph model is a model that is used for generating random graphs. There are two closely related variations of the ERRG model. The model used in this thesis is usually denoted as the $G(n, p)$ model. A graph $G = (V, E)$ constructed by this model has n vertices ($|V| = n$) and between every pair of vertices there independently exists an undirected edge with probability p (and hence it does not exist with probability $1 - p$). A possible outcome of a graph created by the $G(5, 1/2)$ model is given in Figure 1. This graph will be used for further examples in this section.

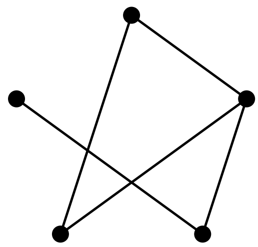

Figure **1**: Possible instance created by the $G(5, 1/2)$ model

2.2 Random Shortest Path Metrics

We define random shortest path metrics in a few steps. First of all, we need an undirected graph $G = (V, E)$. For every edge in this graph an independent edge weight $w(e)$ is drawn from an exponential distribution with parameter 1. For every $u, v \in V$ we define the distance $d(u, v)$ by the length of the shortest u, v-path with respect to the drawn edge weights. If no such path exists, we set $d(u, v) = \infty$. Then the following conditions hold:

- Identity of indiscernibles: $d(u, v) = 0$ if and only if $u = v$ for all $u, v \in V$ with probability 1.
- Symmetry: $d(u, v) = d(v, u)$ for all $u, v \in V$.
- Triangle Inequality: $d(u, w) \leq d(u, v) + d(v, w)$ for all $u, v, w \in V$.

The first condition follows from the fact that all drawn edge weights have a value of exactly 0 with probability 0. Therefore all shortest paths using at least one edge have a positive value with probability 1. The second condition holds, because the graph has undirected edges, which means that every (shortest) path can be traversed in both directions. The final condition follows directly from the definition of a shortest path. If the triangle inequality does not hold, a shorter path exists and $d(u, v)$ would not be the shortest path length.

We create our metric $M = (V, d)$ by taking all vertices from G and assigning the distance $d(u, v)$ to every pair $u, v \in V$. By slight abuse of notation, we can see this metric as a complete graph where the edge weights correspond to the metric distances between vertices.

An example of a graph G and the visualization of the corresponding metric M is given in Figure 2. Figure 2a shows a graph G with random edge weights. Figure 2b shows the visualization of the corresponding random shortest path metric M. In this metric $d(a, e) = 11$, because in graph G the shortest path from a to e goes through c. Similarly, $d(a, b) = 20$ by the shortest path *a*-*c*-*e*-*d*-*b*. All other distances follow with similar reasoning.

(a) Graph G

(b) Visualization of metric M

Figure **2**

2.3 Notation

So far we introduced the ERRG model and given a graph $G = (V, E)$, we defined the edge weights $w(e)$, shortest path distances $d(u, v)$ and random shortest path metric $M = (V, d)$. Further notation used for the analysis of these metrics is defined in this section.

First of all, we define $B_\Delta(v)$ as the set of vertices whose distance from v is at most Δ. In mathematical notation this becomes $B_\Delta(v) = \{u \in V \mid d(u, v) \leq \Delta\}$. Additionally, we define the shortest distance to the k-th closest vertex from v, including v, by $\tau_k(v, G)$. This equals the minimal value for Δ such that there are at least k vertices within distance Δ from v and thus $\tau_k(v, G) = \min\{\Delta \mid |B_\Delta(v)| \geq k\}$. Furthermore, since $\tau_k(v, G)$ is the shortest distance from v to some vertex $u \in V$, we must have $\tau_k(v, G) = d(v, u)$ for some $u \in V$. If we take Figure 2a as example with starting vertex e, we find $\tau_2(e, G) = 2$, since the vertex closest to e is d at distance 2. Then $\tau_3(e, G) = 8$, since the next vertex closest to e is c at distance 8. $\tau_4(e, G) = 9$, since b is at distance 9 from e (by passing through d) and finally $\tau_5(e, G) = 11$, since the shortest path from a to e goes through c and has distance 11.

Next, let a cut of a graph $G = (V, E)$ induced by $U \subset V$ be the set of edges with one endpoint in U and the other endpoint in $U \backslash V$. Then, for every $v \in V$, we define $\chi_k(v, G)$ as the number of edges in the cut induced by the k vertices closest to v. So $\chi_1(v, G)$ equals the number of edges that have v as an endpoint, $\chi_2(v, G)$ is the number of edges with v or its closest neighbor as one endpoint and some other vertex as other endpoint and so on. For example, in Figure 2a, we see $\chi_1(d, G) = 2$, since d has two neighbors. The closest neighbor of d is e. Therefore $\chi_2(d, G) = 3$, since there are 3 edges going from either d or e to the other three vertices. Continuing this process we find $\chi_3(d, G) = 2$ and $\chi_4(d, G) = 2$.

Finally, we define some general notation. We use $\mathrm{Exp}(\lambda)$ to denote an exponential distribution with parameter λ and we denote the n-th harmonic number by $H_n = \sum_{i=1}^{n} 1/i$. Furthermore, if a random variable X is distributed according to a probability distribution P, we write $X \sim P$ and if a random variable X is stochastically dominated by a random variable Y, this is denoted by $X \preceq Y$. We conclude by defining $[n] = \{1, \ldots, n\}$.

3 Technical lemmas

This section contains several technical lemmas that will be used in later sections. The lemmas in this section do not provide any new properties or insights, but are needed for the proofs of lemmas and theorems in later sections.

First we need to rewrite a summation using harmonic numbers.

Lemma 3.1. *For all $n \in \mathbb{N}$ and $k \in [n]$, we have*

$$\sum_{i=1}^{k-1} \frac{1}{i(n-i)} = \frac{H_{k-1} + H_{n-1} - H_{n-k}}{n}.$$

Proof. See Bringmann et al. [4, Lemma 3.1]. □

We need an upper bound for the cumulative distribution function of the sum of independent exponential distributions.

Lemma 3.2. *Let $X \sim \sum_{i=1}^{n} \operatorname{Exp}(ci)$. Then, for $a \geq 0$,*

$$\mathbb{P}(X \leq a) = \left(1 - e^{-ca}\right)^n.$$

Proof. See Bringmann et al. [4, Lemma 3.2]. □

The following summation is used several times throughout this thesis.

Lemma 3.3. *We have*

$$\sum_{i=1}^{O(\ln(n))} O\left(\frac{i}{n} + \frac{i}{e^{(i-1)/5}}\right) = O(1).$$

Proof. The first part of the summation can be bounded by substituting the highest value for i. This yields

$$\sum_{i=1}^{O(\ln(n))} O\left(\frac{i}{n}\right) \leq \sum_{i=1}^{O(\ln(n))} O\left(\frac{\ln(n)}{n}\right) = O\left(\frac{\ln^2(n)}{n}\right) = o(1).$$

For the second part of the summation, by the geometric series, we know for $|x| < 1$ that

$$\sum_{i=0}^{\infty} i \cdot x^{i-1} = \frac{1}{(1-x)^2}.$$

We observe $1/\exp(1/5) < 1$ to obtain

$$\sum_{i=1}^{O(\ln(n))} O\left(\frac{i}{e^{(i-1)/5}}\right) \leq \sum_{i=0}^{\infty} O\left(\frac{i}{e^{(i-1)/5}}\right) = O\left(\frac{1}{(1-e^{-1/5})^2}\right) = O(1).$$

Combining these results yields

$$\sum_{i=1}^{O(\ln(n))} O\left(\frac{i}{n} + \frac{i}{e^{(i-1)/5}}\right) = o(1) + O(1) = O(1).$$

□

We need to bound the expected ratio of non-negative random variables. This is done in the following lemma.

Lemma 3.4. *For all non-negative random variables A, B and constant $c > 0$, we have*

$$\mathbb{E}\left(\frac{A}{B}\right) \leq \mathbb{E}\left(\frac{A}{c}\right) + \mathbb{E}\left(\frac{A}{B} \,\middle|\, B < c\right) \cdot \mathbb{P}\left(B < c\right).$$

Proof. By conditioning on the outcome of B, we obtain

$$\mathbb{E}\left(\frac{A}{B}\right) = \mathbb{E}\left(\frac{A}{B} \,\middle|\, B \geq c\right) \cdot \mathbb{P}\left(B \geq c\right) + \mathbb{E}\left(\frac{A}{B} \,\middle|\, B < c\right) \cdot \mathbb{P}\left(B < c\right).$$

We give an upper bound for the first expectation by observing that $B \geq c$ implies $A/B \leq A/c$. This yields

$$\mathbb{E}\left(\frac{A}{B}\right) \leq \mathbb{E}\left(\frac{A}{c} \,\middle|\, B \geq c\right) \cdot \mathbb{P}\left(B \geq c\right) + \mathbb{E}\left(\frac{A}{B} \,\middle|\, B < c\right) \cdot \mathbb{P}\left(B < c\right).$$

Similarly, by conditioning on the outcome of B, we have

$$\begin{aligned}\mathbb{E}\left(\frac{A}{c}\right) &= \mathbb{E}\left(\frac{A}{c} \,\middle|\, B \geq c\right) \cdot \mathbb{P}\left(B \geq c\right) + \mathbb{E}\left(\frac{A}{c} \,\middle|\, B < c\right) \cdot \mathbb{P}\left(B < c\right) \\ &\geq \mathbb{E}\left(\frac{A}{c} \,\middle|\, B \geq c\right) \cdot \mathbb{P}\left(B \geq c\right),\end{aligned}$$

where the inequality follows from A being non-negative. Substituting this into our earlier result yields

$$\begin{aligned}\mathbb{E}\left(\frac{A}{B}\right) &\leq \mathbb{E}\left(\frac{A}{c} \,\middle|\, B \geq c\right) \cdot \mathbb{P}\left(B \geq c\right) + \mathbb{E}\left(\frac{A}{B} \,\middle|\, B < c\right) \cdot \mathbb{P}\left(B < c\right) \\ &\leq \mathbb{E}\left(\frac{A}{c}\right) + \mathbb{E}\left(\frac{A}{B} \,\middle|\, B < c\right) \cdot \mathbb{P}\left(B < c\right).\end{aligned}$$

□

At some point we need a different bound for the cumulative distribution of the sum of independent exponential distributions than the one found in Lemma 3.2. This other bound is give in the next lemma.

Lemma 3.5. *Let $X = \sum_{i=1}^{n} X_i$ with $X_i \sim \mathrm{Exp}(a_i)$ independent. Let $\mu = \mathbb{E}(X) = \sum_{i=1}^{n} 1/a_i$ and $a_* = \min_i a_i$. For any $\lambda \leq 1$,*

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

Proof. See Janson [10, Theorem 5.1 (iii)]. □

In the next lemma, we bound a non-trivial summation from above.

Lemma 3.6. *Let $0 < \alpha \leq \beta \leq 1$. Then we have*

$$\sum_{x=1}^{n!} O\left(n^{-\Omega(\ln(x))}\right) + \sum_{x=1}^{n!} O\left(\frac{\beta}{\alpha} n^6 \ln(n) \frac{\ln(x)}{x}\right) = O\left(\frac{\beta}{\alpha} n^8 \ln^3(n)\right).$$

Proof. Let $\log(x)$ be the base 10 logarithm of x. Then we have

$$\sum_{x=1}^{n!} n^{-\Omega(\ln(x))} = \sum_{x=1}^{n!} n^{-\Omega(\log(x))} \leq \sum_{x=1}^{n!} n^{-\Omega(\lfloor \log(x) \rfloor)}.$$

Since $1 \leq x \leq n!$, we have $0 \leq \lfloor \log(x) \rfloor \leq n \log(n)$. We split the summation to obtain

$$\begin{aligned}
\sum_{x=1}^{n!} n^{-\Omega(\ln(x))} &\leq \sum_{x=1}^{n!} n^{-\Omega(\lfloor \log(x) \rfloor)} \\
&\leq \sum_{i=0}^{n\log(n)} \sum_{j=10^i}^{10^{i+1}-1} n^{-\Omega(i)} \\
&\leq \sum_{i=0}^{n\log(n)} \left(10^{i+1} - 10^i\right) n^{-\Omega(i)} \\
&= \sum_{i=0}^{n\log(n)} 9 \left(\frac{10}{n^{\Omega(1)}}\right)^i.
\end{aligned}$$

For n sufficiently large we get

$$\begin{aligned}
\sum_{x=1}^{n!} n^{-\Omega(\ln(x))} &\leq \sum_{i=0}^{n\log(n)} 9 \left(\frac{10}{n^{\Omega(1)}}\right)^i \\
&\leq \sum_{i=0}^{n\log(n)} 9 \left(\frac{10}{n^{\Omega(1)}}\right) \\
&= 90(n\log(n)+1) \cdot n^{-\Omega(1)} \\
&= O(n\ln(n)).
\end{aligned}$$

Therefore we have

$$\sum_{x=1}^{n!} O\left(n^{-\Omega(\ln(x))}\right) = O\left(n\ln(n)\right).$$

For the second part we use

$$\begin{aligned}
\sum_{x=1}^{n!} \frac{\log(x)}{x} &\leq \sum_{x=1}^{n!} \frac{\lceil \log(x) \rceil}{x} \\
&= \sum_{x=2}^{n!} \frac{\lceil \log(x) \rceil}{x}
\end{aligned}$$

Since $2 \leq x \leq n!$, we have $1 \leq \lceil \log(x) \rceil \leq n \log(n)$. We split the summation to obtain

$$\begin{aligned}
\sum_{x=1}^{n!} \frac{\log(x)}{x} &\leq \sum_{x=2}^{n!} \frac{\lceil \log(x) \rceil}{x} \\
&\leq \sum_{i=1}^{n\log(n)} \sum_{j=10^{i-1}+1}^{10^i} \frac{i}{j}.
\end{aligned}$$

The summation over j equals the difference between two harmonic numbers. This yields

$$\sum_{x=1}^{n!} \frac{\log(x)}{x} \leq \sum_{i=1}^{n\log(n)} \sum_{j=10^{i-1}+1}^{10^i} \frac{i}{j}$$
$$= \sum_{i=1}^{n\log(n)} \left(H_{10^i} - H_{10^{i-1}}\right) i.$$

Since $\ln(i) \leq H_i \leq \ln(i) + 1$ holds for all i, we have

$$\begin{aligned}
\sum_{x=1}^{n!} \frac{\log(x)}{x} &= \sum_{i=1}^{n\log(n)} \left(H_{10^i} - H_{10^{i-1}}\right) i \\
&\leq \sum_{i=1}^{n\log(n)} \left(\ln(10^i) + 1 - \ln(10^{i-1})\right) i \\
&= \sum_{i=1}^{n\log(n)} (1 + \ln(10)) i \\
&= (1 + \ln(10)) \cdot \frac{1}{2} \cdot n\log(n) \cdot (n\log(n) + 1) \\
&= O\left(n^2 \ln^2(n)\right).
\end{aligned}$$

Combining our observations yields

$$\begin{aligned}
\sum_{x=1}^{n!} O\left(n^{-\Omega(\ln(x))}\right) + \sum_{x=1}^{n!} O\left(\frac{\beta}{\alpha} n^6 \ln(n) \frac{\ln(x)}{x}\right) &= \sum_{x=1}^{n!} O\left(n^{-\Omega(\ln(x))}\right) + \sum_{x=1}^{n!} O\left(\frac{\beta}{\alpha} n^6 \ln(n) \frac{\log(x)}{x}\right) \\
&= O\left(n\ln(n)\right) + O\left(\frac{\beta}{\alpha} n^6 \ln(n) n^2 \ln^2(n)\right) \\
&= O\left(\frac{\beta}{\alpha} n^8 \ln^3(n)\right).
\end{aligned}$$

□

If $X_1, \cdots, X_m$ are m random variables, then $X_{(1)}, \cdots, X_{(m)}$ are the order statistics corresponding to $X_1, \cdots, X_m$ if $X_{(i)}$ is the i-th smallest value among $X_1, \cdots, X_m$ for all $i \in [m]$. We need a link between the distribution of order statistics and the distribution of the sum of exponential distributions. We use the following lemma.

Lemma 3.7. *Let $X_1, \ldots, X_n$ be independent identically distributed random variables with distribution $\mathrm{Exp}(\lambda)$. Let $X_{(1)} \leq \cdots \leq X_{(n)}$ be the corresponding order statistics from this sample. We have for all $i \in [n]$*

$$X_{(i)} \sim \sum_{j=1-i+n}^{n} \mathrm{Exp}(\lambda j).$$

Proof. From Nagaraja [14] we have Rényi's representation,

$$X_{(i)} \sim \frac{1}{\lambda} \sum_{k=1}^{i} \frac{\mathrm{Exp}(1)}{n-k+1}.$$

This can be rewritten to

$$X_{(i)} \sim \sum_{k=1}^{i} \mathrm{Exp}(\lambda(n-k+1)).$$

Taking $j = k - i + n$ yields

$$X_{(i)} \sim \sum_{j=1-i+n}^{n} \mathrm{Exp}(\lambda j).$$

□

Finally, we need a two-sided chernoff bound. The bound we use is given in the following lemma.

Lemma 3.8. *Let $X_1, \cdots, X_n$ be independent Poisson trial such that $\mathbb{P}(X_i = 1) = p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}(X)$. For $0 < \varepsilon < 1$,*

$$\mathbb{P}\left(\left|X - \mu\right| \geq \varepsilon\mu\right) \leq 2e^{-\mu\varepsilon^2/3}.$$

Proof. See Mitzenmacher and Upfal [13, Corollary 4.6]. □

4 Structure of graphs with the α-β-cut-property

Before we look into instances of the ERRG model, we first consider arbitrary simple graphs. In this section we prove several structural properties of connected simple graphs with independent random edge weights drawn from Exp(1) and of corresponding random shortest path metrics. These properties will be used in the subsequent sections for the probabilistic analysis of heuristics in random shortest path metrics.

4.1 α-β-cut-property

In order to find structural properties for arbitrary simple graphs, we use a property based on the cuts of a graph. We first give a formal definition for a cut of a graph.

Definition 4.1. *In a simple graph $G = (V, E)$, a cut $c(U)$ induced by $U \subset V$ is the set of edges with one endpoint in U and the other endpoint in $V \backslash U$. The size of this cut, $|c(U)|$, is the number of edges in this cut. Define $\mu_U = |U|(n - |U|)$, the size of the cut in a complete simple graph.*

We use this definition for a cut of a graph to define the following property for graphs. This property will be used to analyze arbitrary simple graphs.

Definition 4.2 (**α-β**-cut-property)**.** *Let $0 < \alpha \leq \beta \leq 1$. A simple graph $G = (V, E)$ has the α-β-cut-property if the following holds true for all $U \subset V$:*

$$\alpha \cdot \mu_U \leq |c(U)| \leq \beta \cdot \mu_U.$$

Since α is strictly larger than 0 in this definition, all cuts of a graph that satisfies the α-β-cut-property must have a size of at least 1. This implies that the property holds for some values of α and β if and only if the graph is connected. Therefore, as long as we analyze graphs that satisfy the α-β-cut-property for some values of α and β, we know the graphs are connected.

4.2 Distribution of $\tau_k(v, G)$

In this section we assume we have a simple graph $G = (V, E)$, with independent random edge weights drawn from Exp(1), that satisfies the α-β-cut-property for some values of α and β. We give lower and upper bounds for several structural probabilistic properties of this graph and corresponding random shortest path metric $M = (V, d)$. We derive bounds for the probabilistic distribution of $\tau_k(v, G)$, which is the shortest distance to the k-th closest vertex from v, and for the diameter of the metric.

For the analysis of the distribution of $\tau_k(v, G)$ we start at vertex v and we analyze how the distances from v to other vertices develop through first passage percolation. The idea of first passage percolation is that we start percolating a fluid through our graph at rate 1, starting at the source v. The travel time of every edge equals the corresponding edge weight. The development of $\tau_k(v, G)$ can then be seen as follows. For $k = 1$, we have $\tau_k(v, G) = 0$, since v is the vertex closest to itself. Now assume that for $k \geq 2$, the fluid has reached exactly $k - 1$ vertices. We look for the time it takes until the k-th vertex is reached by the fluid.

Using our notation from Section 2.3, the number of edges going from the $k - 1$ already reached vertices to all other vertices of the graph equals $\chi_{k-1}(v, G)$. Fluid is percolating through all these edges and all edge weights are independent and drawn from Exp(1). For every edge in this set, by the memoryless property of the exponential distribution, the expected time until the fluid reaches the end of the edge equals 1. This implies $\tau_{k+1}(v,G) - \tau_k(v,G)$ is the minimum of χ_k exponential random variables with parameter 1. This has the same distribution as an exponential random variable with parameter $\chi_k(v,G)$ and thus $\tau_{k+1}(v,G) - \tau_k(v,G) \sim \mathrm{Exp}(\chi_k(v,G))$ [17, page 302]. We use this result to find bounds for the distribution of $\tau_k(v,G)$.

Lemma 4.3. *Let $0 < \alpha \leq \beta \leq 1$ and let $G = (V,E)$ be a graph that satisfies the α-β-cut-property. We have for all $k \in [n]$ and $v \in V$,*

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

Proof. By definition, $\chi_k(v,G)$ is the size of a cut of graph G induced by a set of k vertices. By assumption the α-β-cut property (Definition 4.2) holds. The result follows immediately from this property. □

Lemma 4.4. *Let $0 < \alpha \leq \beta \leq 1$ and let $G = (V,E)$ be a graph that satisfies the α-β-cut-property. We have for all $k \in [n]$ and $v \in V$,*

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

Proof. As previously stated, $\tau_{i+1}(v,G) - \tau_i(v,G) \sim \mathrm{Exp}(\chi_i(v,G))$. A summation over i yields

$$\tau_k(v,G) \sim \sum_{i=1}^{k-1} \mathrm{Exp}(\chi_i(v,G)).$$

Since an exponential distribution with parameter λ is stochastically dominated by an exponential distribution with parameter $\mu \leq \lambda$, we substitute the upper and lower bounds from Lemma 4.3 for $\chi_k(v,G)$ to obtain

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

□

By definition, $\tau_k(v,G)$ is the shortest distance from v to some other vertex, thus it corresponds to the distance between two vertices in the random shortest path metric. We use this to find bounds for the expected average distance and the expected longest distance from a given vertex.

Corollary 4.5. *Let $0 < \alpha \leq \beta \leq 1$ and let $G = (V,E)$ be a graph that satisfies the α-β-cut-property. We have for all $k \in [n]$ and $v \in V$,*

$$\mathbb{E}(\tau_k(v,G)) \in \left[\frac{H_{k-1} + H_{n-1} - H_{n-k}}{\beta n}, \frac{H_{k-1} + H_{n-1} - H_{n-k}}{\alpha n}\right].$$

Proof. By Lemma 4.4 we have

$$\sum_{i=1}^{k-1} \mathrm{Exp}\left(\beta i(n-i)\right) \leq \tau_k(v,G) \leq \sum_{i=1}^{k-1} \mathrm{Exp}\left(\alpha i(n-i)\right).$$

Taking the expectation of all three parts of this inequality yields

$$\sum_{i=1}^{k-1} \frac{1}{\beta i(n-i)} \leq \mathbb{E}\left(\tau_k(v,G)\right) \leq \sum_{i=1}^{k-1} \frac{1}{\alpha i(n-i)}.$$

Using Lemma 3.1, we obtain

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

□

Corollary 4.6. *Let $0 < \alpha \leq \beta \leq 1$ and let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. For all $u, v \in V$ we have*

$$\mathbb{E}(d(u,v)) \in \left[\frac{H_{n-1}}{\beta(n-1)}, \frac{H_{n-1}}{\alpha(n-1)}\right].$$

Proof. For any vertex v, $\tau_k(v,G)$ equals the distance to the k-th closest vertex to v and thus equals $d(u,v)$ for some $u \in V$. Averaging over all distances from v to other vertices yields

$$\mathbb{E}(d(u,v)) = \frac{1}{n-1}\sum_{k=2}^{n} \mathbb{E}\left(\tau_k(v,G)\right).$$

Substitution of the result of Corollary 4.5 yields

$$\sum_{k=2}^{n} \frac{H_{k-1} + H_{n-1} - H_{n-k}}{\beta n(n-1)} \leq \mathbb{E}(d(u,v)) \leq \sum_{k=2}^{n} \frac{H_{k-1} + H_{n-1} - H_{n-k}}{\alpha n(n-1)},$$

which can be simplified to

$$\frac{H_{n-1}}{\beta(n-1)} \leq \mathbb{E}(d(u,v)) \leq \frac{H_{n-1}}{\alpha(n-1)}.$$

□

Corollary 4.7. *Let $0 < \alpha \leq \beta \leq 1$ and let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. For all $v \in V$ we have*

$$\mathbb{E}\left(\max_{u\in V}\{d(u,v)\}\right) \in \left[\frac{2H_{n-1}}{\beta n}, \frac{2H_{n-1}}{\alpha n}\right].$$

Proof. Given vertex v, the distance to the vertex that is the furthest away from v is $\tau_n(v,G)$. This implies that $\tau_n(v,G)$ equals $\max_{u\in V}\{d(u,v)\}$. Therefore

$$\mathbb{E}\left(\max_{u\in V}\{d(u,v)\}\right) = \mathbb{E}(\tau_n(v,G)).$$

Substitution of the result of Corollary 4.5 yields

$$\frac{2H_{n-1}}{\beta n} \leq \mathbb{E}\left(\max_{u\in V}\{d(u,v)\}\right) \leq \frac{2H_{n-1}}{\alpha n}.$$

□

It is also possible to find bounds for the cumulative distribution function of $\tau_k(v, G)$. In order to find these bounds, we loosen the bounds for $\tau_k(v, G)$ found in Lemma 4.4. For the following proofs, we define $F_k(x) = \mathbb{P}(\tau_k(v, G) \leq x)$ for some fixed vertex $v \in V$.

Lemma 4.8. *Let $0 < \alpha \leq \beta \leq 1$ and let $G = (V, E)$ be a graph that satisfies the α-β-cut-property. We have for all $k \in [n]$ and $v \in V$,*

$$\left(1 - e^{-\alpha(n-k)x}\right)^{k-1} \leq F_k(x) \leq \left(1 - e^{-\beta nx}\right)^{k-1}.$$

Proof. By Lemma 4.4 we have

$$\sum_{i=1}^{k-1} \mathrm{Exp}\left(\beta i(n-i)\right) \leq \tau_k(v, G) \leq \sum_{i=1}^{k-1} \mathrm{Exp}\left(\alpha i(n-i)\right).$$

Since $\mathrm{Exp}\left(\beta ni\right) \leq \mathrm{Exp}\left(\beta i(n-i)\right)$ for $i \in [1, k-1]$, we obtain

$$\sum_{i=1}^{k-1} \mathrm{Exp}\left(\beta ni\right) \leq \sum_{i=1}^{k-1} \mathrm{Exp}\left(\beta i(n-i)\right) \leq \tau_k(v, G).$$

Similarly, since $\mathrm{Exp}\left(\alpha i(n-i)\right) \leq \mathrm{Exp}\left(\alpha(n-k)i\right)$ for $i \in [1, k-1]$, we obtain

$$\tau_k(v, G) \leq \sum_{i=1}^{k-1} \mathrm{Exp}\left(\alpha i(n-i)\right) \leq \sum_{i=1}^{k-1} \mathrm{Exp}\left(\alpha(n-k)i\right).$$

Combining these results yields

$$\sum_{i=1}^{k-1} \mathrm{Exp}\left(\beta ni\right) \leq \tau_k(v, G) \leq \sum_{i=1}^{k-1} \mathrm{Exp}\left(\alpha(n-k)i\right).$$

This immediately implies

$$\mathbb{P}\left(\sum_{i=1}^{k-1} \mathrm{Exp}\left(\alpha(n-k)i\right) \leq x\right) \leq F_k(x) \leq \mathbb{P}\left(\sum_{i=1}^{k-1} \mathrm{Exp}\left(\beta ni\right) \leq x\right).$$

Applying Lemma 3.2 to the above inequality gives the desired result. $\square$

We have found an upper and lower bound for the cumulative distribution function of $\tau_k(v, G)$. Using a different approach, we are able to make two improvements on the lower bound found in the previous lemma. These improvements are given in the next lemma and corollary.

Lemma 4.9. *Let $0 < \alpha \leq \beta \leq 1$ and let $G = (V, E)$ be a graph that satisfies the α-β-cut-property. We have for all $k \in [n]$,*

$$F_k(x) \geq \left(1 - e^{-\alpha(n-1)x/4}\right)^{n+1}.$$

Proof. First of all, note that $\tau_k(v, G) \leq \tau_{k+1}(v, G)$ for all k and thus $\tau_k(v, G)$ is monotonically increasing. This implies $F_{k+1}(x) \leq F_k(x)$. Because of this, we only need to prove our claim for the case $k = n$. In this case, by Lemma 4.4, we have $\tau_n(v, G) \leq \sum_{i=1}^{n-1} \mathrm{Exp}\left(\alpha i(n-i)\right)$.

Define $\lambda_i := \alpha i(n-i)$. We use this notation to obtain

$$\begin{aligned}
\tau_n(v,G) &\leq \sum_{i=1}^{n-1} \operatorname{Exp}\left(\alpha i(n-i)\right) \\
&= \sum_{i=1}^{n-1} \operatorname{Exp}\left(\lambda_i\right) \\
&\leq \sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\lambda_i\right) + \sum_{i=\lfloor n/2\rfloor}^{n-1} \operatorname{Exp}\left(\lambda_i\right).
\end{aligned}$$

Note that λ_i is symmetric around $i = n/2$, which implies $\lambda_i = \lambda_{n-i}$. This yields

$$\tau_n(v,G) \leq \sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\lambda_i\right) + \sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\lambda_i\right).$$

This approximation enables us to find a lower bound for $F_n(x)$ as follows:

$$\begin{aligned}
F_n(x) = \mathbb{P}(\tau_n(v,G) \leq x) &\geq \mathbb{P}\left(\sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\lambda_i\right) + \sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\lambda_i\right) \leq x\right) \\
&\geq \mathbb{P}\left(\sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\lambda_i\right) \leq \frac{x}{2} \text{ and } \sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\lambda_i\right) \leq \frac{x}{2}\right) \\
&\geq \mathbb{P}\left(\sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\lambda_i\right) \leq x/2\right)^2 \\
&= \mathbb{P}\left(\sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\alpha i(n-i)\right) \leq x/2\right)^2.
\end{aligned}$$

An exponential distribution with parameter λ is stochastically dominated by an exponential distribution with parameter $\mu \leq \lambda$. Therefore $\operatorname{Exp}(\alpha i(n-i))$ is stochastically dominated by $\operatorname{Exp}(\alpha i(n-\lceil n/2\rceil))$ for all $i \leq \lceil n/2\rceil$. Substituting this value for i yields

$$\begin{aligned}
F_n(x) &\geq \mathbb{P}\left(\sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\alpha i(n-i)\right) \leq x/2\right)^2 \\
&\geq \mathbb{P}\left(\sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\alpha i(n-\lceil n/2\rceil)\right) \leq x/2\right)^2 \\
&= \mathbb{P}\left(\sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\alpha i\lfloor n/2\rfloor\right) \leq x/2\right)^2.
\end{aligned}$$

Applying Lemma 3.2 to our previous inequality yields

$$\begin{aligned}
F_n(x) &\geq \mathbb{P}\left(\sum_{i=1}^{\lceil n/2\rceil} \operatorname{Exp}\left(\alpha i\lfloor n/2\rfloor\right) \leq x/2\right)^2 \\
&= \left(1 - e^{-\alpha\lfloor n/2\rfloor\cdot x/2}\right)^{2\lceil n/2\rceil}.
\end{aligned}$$

Finally, we use the inequalities $\lfloor n/2 \rfloor \geq (n-1)/2$ and $\lceil n/2 \rceil \leq (n+1)/2$ to obtain

$$\begin{aligned} F_n(x) &\geq \left(1 - e^{-\alpha \lfloor n/2 \rfloor \cdot x/2}\right)^{2\lceil n/2 \rceil} \\ &\geq \left(1 - e^{-\alpha(n-1)x/4}\right)^{n+1}. \end{aligned}$$

□

Corollary 4.10. *Let $0 < \alpha \leq \beta \leq 1$ and let $G = (V, E)$ be a graph that satisfies the α-β-cut-property. We have for all $k \in [n]$,*

$$F_k(x) \geq \left(1 - e^{-\alpha(n-1)x/4}\right)^{\frac{2}{3}(2k+1)}.$$

Proof. This proof is split into two cases. If $k-1 \geq 3(n-1)/4$, then it also holds that $2(2k+1)/3 \geq n+1$. Applying this to the result of Lemma 4.9 yields

$$\begin{aligned} F_k(x) &\geq \left(1 - e^{-\alpha(n-1)x/4}\right)^{n+1} \\ &\geq \left(1 - e^{-\alpha(n-1)x/4}\right)^{2/3(2k+1)}. \end{aligned}$$

On the other hand, if $k-1 < 3(n-1)/4$, then $n-k > (n-1)/4$. Combined with the result of Lemma 4.8 this yields

$$\begin{aligned} F_k(x) &\geq \left(1 - e^{-\alpha(n-k)x}\right)^{k-1} \\ &\geq \left(1 - e^{-\alpha(n-1)x/4}\right)^{k-1} \\ &\geq \left(1 - e^{-\alpha(n-1)x/4}\right)^{\frac{2}{3}(2k+1)}, \end{aligned}$$

where the last inequality follows since $k-1 < 2(2k+1)/3)$. □

The improved bound for the cumulative distribution function of $\tau_k(v, G)$ from Lemma 4.9 enables us to analyze the diameter of a random shortest path metric. We find the tail bound given in the following theorem.

Theorem 4.11. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. Let $\Delta_{\max} = \max_{u,v \in V}\{d(u, v)\}$. Then we have*

$$\mathbb{P}\left(\Delta_{\max} > \frac{c\ln(n)}{n\alpha}\right) = O\left(n^{2-c/8}\right).$$

Proof. Since $\tau_n(v, G)$ equals the distance to the vertex that is furthest away from v, we have $\tau_n(v, G) = \max_{u \in V}\{d(u, v)\}$. We obtain

$$\begin{aligned} \mathbb{P}\left(\Delta_{\max} > \frac{c\ln(n)}{n\alpha}\right) &= \mathbb{P}\left(\max_{v \in V}\{\tau_n(v, G)\} > \frac{c\ln(n)}{n\alpha}\right) \\ &= \mathbb{P}\left(\exists\, v \in V \left(\tau_n(v, G) > \frac{c\ln(n)}{n\alpha}\right)\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right), \end{aligned}$$

where A_i is the event that $\tau_n(v_i, G) > c\ln(n)/n\alpha$. Taking a union bound yields

$$\begin{aligned}\mathbb{P}\left(\Delta_{\max} > \frac{c\ln(n)}{n\alpha}\right) &= \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \\ &\leq n \cdot \mathbb{P}\left(\tau_n(v, G) > \frac{c\ln(n)}{n\alpha}\right) \\ &= n\left(1 - F_n\left(\frac{c\ln(n)}{n\alpha}\right)\right).\end{aligned}$$

Lemma 4.9 yields

$$\begin{aligned}\mathbb{P}\left(\Delta_{\max} > \frac{c\ln(n)}{n\alpha}\right) &\leq n\left(1 - F_n\left(\frac{c\ln(n)}{n\alpha}\right)\right) \\ &\leq n\left(1 - \left(1 - e^{-\alpha(n-1)c\ln(n)/4n\alpha}\right)^{n+1}\right) \\ &= n\left(1 - \left(1 - e^{-c\ln(n)(n-1)/4n}\right)^{n+1}\right).\end{aligned}$$

Since $(n-1)/n \geq 1/2$ for $n \geq 2$ and the Bernoulli inequality states $(1+x)^y \geq 1 + xy$ for $x \geq -1$ and $y \notin (0,1)$, we obtain

$$\begin{aligned}\mathbb{P}\left(\Delta_{\max} > \frac{c\ln(n)}{n\alpha}\right) &\leq n\left(1 - \left(1 - e^{-c\ln(n)(n-1)/4n}\right)^{n+1}\right) \\ &\leq n\left(1 - \left(1 - (n+1)e^{-c\ln(n)/8}\right)\right). \\ &= n(n+1) \cdot n^{-c/8} \\ &= n^{2-c/8} + n^{1-c/8} \\ &= O\left(n^{2-c/8}\right).\end{aligned}$$

□

4.3 Clustering

In this section, we partition the vertices of the random shortest path metric in a small number of subsets with a given maximum diameter. We call such a partitioning a clustering and the subsets are its clusters. The goal is to, given a maximum diameter for the clusters, find an upper bound for the number of clusters needed. This upper bound will be used for the probabilistic analysis of several heuristics in Section 5.

Before we prove an upper bound for the number of clusters needed, we first need a tail bound for the number of vertices that are within a given distance from a given vertex v, i.e. we need a tail bound for $|B_\Delta(v)|$.

Lemma 4.12. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. For $n \geq 5$ we have,*

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

Proof. By definition we have $|B_\Delta(v)| \geq k$ if and only if $\tau_k(v) \leq \Delta$. Therefore, $\mathbb{P}(B_\Delta(v) < k) = 1 - F_k(\Delta)$. Lemma 4.8 yields

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

We apply the Bernoulli inequality $(1+x)^y \geq 1 + xy$, which holds for $x \geq -1$ and $y \notin (0,1)$. We obtain

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

For $n \geq 5$ we have $(n-1)/4 \geq n/5$. Therefore, for $n \geq 5$, we can replace $(n-1)/4$ by $n/5$ to obtain

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

□

We use the previous lemma to prove our final structural property for random shortest path metrics. We show that a random shortest path metric instance can, given any maximum diameter, be partitioned into a number of clusters dependent on this diameter. This is shown in the next theorem.

Theorem 4.13. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. For every $\Delta \geq 0$, if we partition the vertices of M into clusters, each of diameter at most 6Δ, then the expected number of clusters needed to make this partition is bounded from above by $O(1 + n/\exp(\alpha\Delta n/5))$.*

Proof. Let $s_\Delta = \min\{\exp(\alpha\Delta n/5, (n+1)/2\}$. We call a vertex a dense center if the number of vertices within distance Δ of v is at least s_Δ. Thus v is a dense center if $|B_\Delta(v)| \geq s_\Delta$. We call the set of vertices within distance Δ of v the Δ-ball of v. The diameter of such a ball is at most 2Δ. If a vertex is not a dense center, we call it a sparse center.

In the remainder of this proof, we first show that the expected number of sparse centers is at most $O(n/s_\Delta)$. Then we show that the dense centers can be divided into at most $O(n/s_\Delta)$ clusters of diameter at most 6Δ. These results are then combined to obtain a partitioning into the number of clusters that we want to prove.

For the first part we use Lemma 4.12. This lemma implies that a vertex is a sparse center with a probability of at most $O(\exp(-\alpha\Delta n/5))$. Since there are n vertices, we find

$$\begin{aligned}\mathbb{E}(\text{number of sparse centers}) &= \sum_{v\in V}\mathbb{P}(v \text{ is a sparse center})\\ &\leq n\cdot O\left(e^{-\alpha\Delta n/5}\right)\\ &= O\left(\frac{n}{s_\Delta}\right).\end{aligned}$$

This leaves us with the dense centers. We cluster these as follows: Given all dense centers with corresponding Δ-balls, consider any maximal independent set of these Δ-balls. Since each Δ-ball contains at least s_Δ vertices, this independent set exists out of at most $t \leq n/s_\Delta$ Δ-balls. These balls form the initial clusters $C_1, ..., C_t$. Since we took a maximal independent set, the Δ-ball of every remaining dense center v has at least one vertex in one of these initial clusters. We add all remaining vertices of the corresponding ball $B_\Delta(v)$ to such a cluster to form final clusters $C'_1, ..., C'_t$.

Consider any two vertices u and v in some cluster C'_i. The distance from u towards its nearest neighbor in the initial cluster C_i is at most 2Δ by construction. The same holds for v. Finally, the diameter of the initial cluster was also 2Δ. Therefore, the diameter of the final clusters is at most 6Δ.

We obtain a clustering of all vertices as follows. Every sparse center is put in a cluster of its own. In expectation there are at most $O(n/s_\Delta)$ of these clusters, each of size 1 and diameter $0 \leq 6\Delta$. By our previous construction, the dense centers can be put in $t \leq n/s_\Delta$ clusters of size at least s_Δ and diameter at most 6Δ. Therefore, the total expected number of clusters needed is

$$O(n/s_\Delta) = O\left(n/e^{\alpha\Delta n/5} + n(n+1)/2\right) = O(1 + n/\exp(\alpha\Delta n/5)).$$

□

The general idea of the proof of Theorem 4.13 is visualized in Figure 3. The red vertices are sparse centers. The blue balls are Δ-balls which have been added together to one final cluster with diameter at most $3\cdot 2\Delta = 6\Delta$.

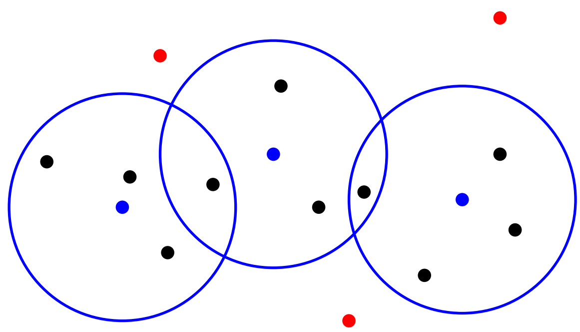

Figure **3**: Example clustering, sparse centers in red, combined cluster in blue

5 Analysis of Heuristics

In the previous section we have proven several structural properties of graphs that satisfy the α-β-cut-property and of corresponding random shortest path metrics. We use these properties for the probabilistic analysis of heuristics on these random metrics. We first look at the heuristic `greedy` for the minimum distance perfect matching problem. Afterwards we analyze three heuristics for the traveling salesman problem: `nearest neighbor`, `insert` and `2-opt`. This section concludes with an analysis of the heuristic `trivial` for the k-Median Problem.

5.1 Greedy for the Minimum Distance Perfect Matching Problem

We first consider a heuristic for the minimum distance perfect matching problem. We define this problem as follows.

Definition 5.1 (Minimum Distance Perfect Matching Problem). *Given a metric $M = (V, d)$, a perfect matching of M is a partitioning of the vertices into pairs, such that every vertex is part of exactly one pair. We say vertex u is matched to vertex v if they form a pair in the matching. The goal of the minimum distance perfect matching problem is to find a perfect matching that minimizes the total distance between all matched pairs of vertices.*

This problem has been widely analyzed throughout history. We do for instance know the worst-case running-time for finding a minimum distance perfect matching is $O(n^3)$, which is high when considering a large number of vertices. Because of this, heuristics are often used. We analyze one heuristic for this problem, which is called `greedy`. This heuristic follows a simple rule:

1. Start with an empty matching.
2. At every iteration, add a pair of unmatched vertices to the matching such that the distance between the added pair of vertices is minimized.
3. Stop when all vertices are matched (perfect matching).

The approximation ratio of `greedy` has already been analyzed for arbitrary metric instances. The worst-case approximation ratio on such instances is known to be $O(n^{\log_2(3/2)})$ [15]. Furthermore, for random shortest path metrics on complete graphs ($\alpha = \beta = 1$ in Definition 4.2), the expected approximation ratio has an upper bound of $O(1)$ [4]. We extend this last result to general values for α and β and give an upper bound for the expected approximation ratio of $O(\beta/\alpha)$.

From now on, let GR denote the cost of the matching computed by `greedy` and let MM denote the value of an optimal matching. In the following two theorems, we show the expected value of GR is $O(1/\alpha)$.

Lemma 5.2. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. Let $f(n)$ be an arbitrary function such that $f(n) = \omega(1)$. If we define $\delta_{f(n)}$ as the sum of all distances in the metric greater than or equal to $f(n)\ln(n)/\alpha n$, then*

$$\mathbb{E}\left(\delta_{f(n)}\right) = o\left(\frac{1}{\alpha}\right).$$

Proof. First of all, for arbitrary $u, v \in V$, we split the expectation as follows:

$$\mathbb{E}\left(\delta_{f(n)}\right) = \mathbb{E}\left(\text{number of distances} \geq \frac{f(n)\ln(n)}{\alpha n}\right) \cdot \mathbb{E}\left(d(u,v) \;\middle|\; d(u,v) \geq \frac{f(n)\ln(n)}{\alpha n}\right).$$

For the first part, since we have n vertices, there are n^2 distances defined. We obtain

$$\mathbb{E}\left(\text{number of distances} \geq \frac{f(n)\ln(n)}{\alpha n}\right) = n^2 \cdot \mathbb{P}\left(d(u,v) \geq \frac{f(n)\ln(n)}{\alpha n}\right).$$

The probability that an arbitrary distance exceeds some value is smaller than the probability that the largest distance exceeds this value. Combining this with Theorem 4.11 yields

$$\begin{aligned}\mathbb{E}\left(\text{number of distances} \geq \frac{f(n)\ln(n)}{\alpha n}\right) &= n^2 \cdot \mathbb{P}\left(d(u,v) \geq \frac{f(n)\ln(n)}{\alpha n}\right)\\ &\leq n^2 \cdot \mathbb{P}\left(\Delta_{\max} \geq \frac{f(n)\ln(n)}{\alpha n}\right)\\ &= n^2 \cdot O\left(n^{2-f(n)/4}\right) = O\left(n^{4-f(n)/4}\right).\end{aligned}$$

For the second part, since $d(u,v)$ equals the total weight of some shortest path in a graph on n vertices where all edge weights are drawn independently from Exp(1), we find

$$\mathbb{E}\left(d(u,v) \;\middle|\; d(u,v) \geq \frac{f(n)\ln(n)}{\alpha n}\right) \leq \frac{f(n)\ln(n)}{\alpha n} + n \cdot \mathbb{E}(\mathrm{Exp}(1)) = \frac{f(n)\ln(n)}{\alpha n} + n.$$

Combining these results yields

$$\begin{aligned}\mathbb{E}\left(\delta_{f(n)}\right) &= \mathbb{E}\left(\text{number of distances} \geq \frac{f(n)\ln(n)}{\alpha n}\right) \cdot \mathbb{E}\left(d(u,v) \mid d(u,v) \geq \frac{f(n)\ln(n)}{\alpha n}\right)\\ &\leq O\left(n^{4-f(n)/4}\right) \cdot \left(\frac{f(n)\ln(n)}{\alpha n} + n\right)\\ &= O\left(n^{5-f(n)/4}\right) + O\left(n^{3-f(n)/4} \cdot \frac{f(n)\ln(n)}{\alpha}\right).\end{aligned}$$

Since $f(n) = \omega(1)$, we have

$$O\left(n^{5-f(n)/4}\right) = o(1)$$

and

$$O\left(n^{3-f(n)/4} \cdot \frac{f(n)\ln(n)}{\alpha}\right) = O\left(\frac{n^{4+\ln(f(n))-f(n)/4}}{\alpha}\right) = o\left(\frac{1}{\alpha}\right).$$

Since $\alpha \leq 1$, this leads us to conclude that

$$\begin{aligned}\mathbb{E}\left(\delta_{f(n)}\right) &= O\left(n^{5-f(n)/4}\right) + O\left(n^{3-f(n)/4} \cdot \frac{f(n)\ln(n)}{\alpha}\right)\\ &= o(1) + o\left(\frac{1}{\alpha}\right) = o\left(\frac{1}{\alpha}\right).\end{aligned}$$

□

Theorem 5.3. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. For such random metrics, we have*

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

Proof. Let $\Delta_i = \frac{i}{\alpha n}$. We divide a run of `greedy` in phases as follows: the algorithm is in phase i if a pair (u, v) is added to the matching such that $d(u, v) \in (6\Delta_{i-1}, 6\Delta_i]$. We look at the contribution of each phase.

First of all, by Theorem 5.2, the expected sum of all distances greater than or equal to $\Delta_{\omega(\ln(n))}$ is $o(1/\alpha)$. Therefore we bound the contribution of all phases with $i \geq \omega(\ln(n))$ by $o(1/\alpha)$. What remains is to find the contribution of the first $O(\ln(n))$ phases.

Since `greedy` always chooses the shortest possible distance, it goes through all phases in increasing order. Now assume the algorithm has finished Phase $i - 1$. By Theorem 4.13 we can partition the vertices in an expected number of $O(1 + n/\exp((i-1)/5))$ clusters, each of diameter at most $6\Delta_{i-1}$. By definition, the longest distance in each cluster is at most $6\Delta_{i-1}$. This implies that every cluster has at most one unmatched vertex (otherwise the algorithm could have added an extra pair of vertices before reaching Phase i). This leaves a total expected number of at most $O(1+n/\exp((i-1)/5))$ unmatched vertices that might be matched in Phase i. Thus in expectation at most $O(1 + n/\exp((i-1)/5))$ pairs of vertices can be added in Phase i, each contributing a distance of at most $6\Delta_i$. This yields an upper bound for the total expected contribution of Phase i of $O(i/\alpha n + i/\alpha \exp((i-1)/5))$.

Let c_i be the contribution of Phase i. Summing over all phases yields

$$\begin{aligned}
\mathbb{E}(\mathsf{GR}) &= \sum_{i=1}^{O(\ln(n))} \mathbb{E}(c_i) + \sum_{i=\omega(\ln(n))}^{\infty} \mathbb{E}(c_i) \\
&= \sum_{i=1}^{O(\ln(n))} O\left(\frac{i}{\alpha n} + \frac{i}{\alpha e^{(i-1)/5}}\right) + o\left(\frac{1}{\alpha}\right) \\
&= \frac{1}{\alpha} \sum_{i=1}^{O(\ln(n))} O\left(\frac{i}{n} + \frac{i}{e^{(i-1)/5}}\right) + o\left(\frac{1}{\alpha}\right).
\end{aligned}$$

We use Lemma 3.3 to find an upper bound for the summation. This yields

$$\begin{aligned}
\mathbb{E}(\mathsf{GR}) &= \frac{1}{\alpha} \sum_{i=1}^{O(\ln(n))} O\left(\frac{i}{n} + \frac{i}{e^{(i-1)/5}}\right) + o\left(\frac{1}{\alpha}\right) \\
&= \frac{1}{\alpha} \cdot O(1) + o\left(\frac{1}{\alpha}\right) \\
&= O\left(\frac{1}{\alpha}\right).
\end{aligned}$$

□

Before we can prove an upper bound for the expected approximation ratio of `greedy`, we need one more lemma. The next lemma gives a tail bound for the set of lightest edge weights in a graph that satisfies the α-β-cut-property. This lemma is followed directly by a proof of an upper bound of $O(\beta/\alpha)$ for the approximation ratio of `greedy`.

Lemma 5.4. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. Let S_m bet the sum of the m lightest edge weights in G. For all $\phi \leq (n-1)/n$ and $c \in [0, 2\phi^2]$ we have*

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

Furthermore, $\mathsf{MM} \geq S_{n/2}$, *where* MM *is the total distance of a minimum distance perfect matching in* M.

Proof. G has $|E|$ edges. Every edge has an independent weight drawn from a standard exponential distribution. Therefore, the lightest edge weight is equal to the minimum of $|E|$ exponential distributions with parameter 1. This has the same distribution as an exponential distribution with parameter $|E|$ [17, page 302]. The second lightest edge weight equals the lightest edge weight plus the remaining weight of the second lightest edge. Since all weights are drawn from Exp(1), by the memoryless property of the exponential distribution, the remaining weight equals the minimum of $|E|-1$ exponential distributions with parameter 1. Thus the second lightest edge weight has as distribution $\text{Exp}(|E|) + \text{Exp}(|E|-1)$. By continuing this process for the lightest ϕn edge weights we obtain

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

Since the α-β-cut-property holds, G is connected and $|E| \geq n-1 \geq \phi n$. This implies $(|E| - i)/(\phi n - i) \leq |E|/\phi n$. We obtain

$$S_{\phi n} \sim \sum_{i=0}^{\phi n-1} \text{Exp}\left(\frac{|E|-i}{\phi n - i}\right) \geq \sum_{i=0}^{\phi n-1} \text{Exp}\left(\frac{|E|}{\phi n}\right).$$

Since the α-β-cut-property holds, we have $|E| \leq \beta n(n-1)/2$. This yields

$$S_{\phi n} \geq \sum_{i=0}^{\phi n-1} \text{Exp}\left(\frac{|E|}{\phi n}\right) \geq \sum_{i=0}^{\phi n-1} \text{Exp}\left(\frac{\beta(n-1)}{2\phi}\right).$$

This immediately implies

$$\mathbb{P}\left(S_{\phi n} \leq \frac{c}{\beta}\right) \leq \mathbb{P}\left(\sum_{i=0}^{\phi n-1} \text{Exp}\left(\frac{\beta(n-1)}{2\phi}\right) \leq \frac{c}{\beta}\right).$$

We apply Lemma 3.5 to bound this probability from above. In our case, $\mu = 2\phi^2 n/\beta(n-1)$, $a_* = \beta(n-1)/2\phi$ and $\lambda\mu = c/\beta$. This implies $\lambda = c(n-1)/2\phi^2 n$. We can apply the tailbound if $\lambda = c(n-1)/2\phi^2 n \leq 1$. Since $c(n-1)/2\phi^2 n < c/2\phi^2$, this certainly holds for $c \in [0, 2\phi^2]$. The tailbound yields

$$\begin{aligned}\mathbb{P}\left(S_{\phi n} \leq \frac{c}{\beta}\right) &\leq \mathbb{P}\left(\sum_{i=0}^{\phi n-1} \text{Exp}\left(\frac{\beta(n-1)}{2\phi}\right) \leq \frac{c}{\beta}\right) \\ &\leq \exp\left(-\phi n\left(\frac{c(n-1)}{2\phi^2 n} - 1 - \ln\left(\frac{c(n-1)}{2\phi^2 n}\right)\right)\right).\end{aligned}$$

Since $c(n-1)/2\phi^2 n \leq c/2\phi^2$, we obtain

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

For the second part of this lemma, we need to show $\mathsf{MM} \geq S_{n/2}$. In order to do this, consider a minimum distance perfect matching in a random shortest path metric. We replace every distance between matched pairs of vertices by their corresponding paths in G. Since every vertex is part of at least one of these paths, the union of the paths uses at least $n/2$ different edges of G. These $n/2$ edges have a total weight of at most MM and at least $S_{n/2}$. Therefore indeed $\mathsf{MM} \geq S_{n/2}$. □

Theorem 5.5. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. For such random metrics, we have*

$$\mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}}\right) = O\left(\frac{\beta}{\alpha}\right).$$

Proof. Let $c > 0$ be a sufficiently small constant. We use Lemma 3.4 to obtain

$$\mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}}\right) \leq \mathbb{E}\left(\frac{\beta}{c}\mathsf{GR}\right) + \mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}} \,\middle|\, \mathsf{MM} < \frac{c}{\beta}\right) \cdot \mathbb{P}\left(\mathsf{MM} < \frac{c}{\beta}\right).$$

We bound the first part by applying Theorem 5.3 and noting that c is a constant. We obtain

$$\mathbb{E}\left(\frac{\beta}{c}\mathsf{GR}\right) = O\left(\frac{\beta}{\alpha}\right).$$

We bound the second expectation from above by the worst-case approximation ratio of `greedy`, which is known to be $O(n^{\log_2(3/2)})$ [15]. Finally, since c is sufficiently small, Lemma 5.4 with $\phi = 1/2$ yields

$$\mathbb{P}\left(\mathsf{MM} < \frac{c}{\beta}\right) \leq \mathbb{P}\left(S_{n/2} < \frac{c}{\beta}\right) \leq e^{\frac{n}{2}(1+\ln(2c))}.$$

Combining these observations yields

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}}\right) &\leq \mathbb{E}\left(\frac{\beta}{c}\mathsf{GR}\right) + \mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}} \,\middle|\, \mathsf{MM} < \frac{c}{\beta}\right) \cdot \mathbb{P}\left(\mathsf{MM} < \frac{c}{\beta}\right) \\ &\leq O\left(\frac{\beta}{\alpha}\right) + O\left(n^{\log_2(3/2)}\right) \cdot e^{\frac{n}{2}(1+\ln(2c))}.\end{aligned}$$

Since c is sufficiently small, the second part becomes $o(1)$. We conclude

$$\mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}}\right) = O\left(\frac{\beta}{\alpha}\right).$$

□

5.2 Nearest Neighbor for the Traveling Salesman Problem

The second heuristic that we consider is a heuristic for the Traveling Salesman Problem. We define this problem as follows.

Definition 5.6 (Traveling Salesman Problem)**.** *Given is a metric $M = (V, d)$. The traveling salesman problem involves finding a tour through all vertices in V using the shortest possible route, with respect to the distance function d, such that every vertex is visited exactly once and the tour starts and ends at the same vertex.*

The traveling salesman problem is known to be an NP-hard problem [11]. This means that there does not exist an algorithm that can solve this problem in polynomial time, unless P = NP. Since no efficient algorithm for the traveling salesman problem is known, heuristics are often used to approximate the optimal solution to this problem. We analyze a heuristic that is called `nearest neighbor`. This heuristic follows the following rules:

1. Start with some starting vertex v_0 as current vertex v.
2. At every step, choose the nearest unvisited neighbor u of v as the next vertex in the tour and move to the next iteration with the new vertex u as current vertex v.
3. Go back to v_0 if all vertices are visited.

Since all vertices are visited exactly once and the algorithm terminates at the starting vertex, the heuristic does indeed find a feasible tour.

The approximation ratio of `nearest neighbor` has already been analyzed for arbitrary metric instances. The worst-case approximation ratio on such instances is known to be $O(\ln(n))$ [16]. Furthermore, for random shortest path metrics on complete graphs ($\alpha = \beta = 1$ in Definition 4.2), the expected approximation ratio has an upper bound of $O(1)$ [4]. We extend this last result to general values for α and β and show an upper bound for the expected approximation ratio of $O(\beta/\alpha)$.

From now on, let NN denote the cost of the tour computed by `nearest neighbor` and let TSP denote the value of an optimal tour. In the following theorem, we show that the expected value of NN is $O(1/\alpha)$.

Theorem 5.7. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. For such random metrics, we have*

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

Proof. Let $\Delta_i = \frac{i}{\alpha n}$. By Theorem 4.13 we can partition the vertices in an expected number of $O(1 + n/\exp(in/5))$ clusters, each of diameter at most $6\Delta_i$. We refer to these clusters as the i-clusters.

Suppose that the distance to the neighbor chosen by `nearest neighbor` for a vertex v is in the interval $(6\Delta_i, 6\Delta_{i+1}]$. Since all i-clusters have a diameter of at most $6\Delta_i$, the chosen neighbor cannot be in the same i-cluster as v. This is only possible if all other vertices of the i-cluster have already been visited. Otherwise, there would be another unvisited vertex u in the same i-cluster within a distance of at most $6\Delta_i$ to v. Since there is an expected number of $O(1 + n/\exp(i/5))$ clusters, the expected number of vertices contributing a distance in the interval $(6\Delta_i, 6\Delta_{i+1}]$ to the tour is at most $O(1+n/\exp(i/5))$. An upper bound for the total contribution of these vertices is then $O((i+1)/\alpha n + (i+1)/\alpha \exp(i/5)$.

The contribution that we found can be used if i is not too large. For $i \geq \omega(\ln(n))$, we use Theorem 5.2 to show that the expected sum of all distances greater than or equal to $\Delta_{\omega(\ln(n))}$ is $o(1/\alpha)$. Therefore, we bound the contribution of vertices contributing a distance greater than or equal to $\Delta_{\omega(\ln(n))}$ to the tour by $o(1/\alpha)$.

Let c_i be the contribution of all vertices added by `nearest neighbor` that contribute a distance in the interval $(6\Delta_i, 6\Delta_{i+1}]$ to the tour. Summing over all values of i yields

$$\begin{aligned}\mathbb{E}(\mathsf{NN}) &= \sum_{i=0}^{O(\ln(n))} \mathbb{E}(c_i) + \sum_{i=\omega(\ln(n))}^{\infty} \mathbb{E}(c_i) \\ &= \sum_{i=0}^{O(\ln(n))} O\left(\frac{i+1}{\alpha n} + \frac{i+1}{\alpha e^{i/5}}\right) + o\left(\frac{1}{\alpha}\right) \\ &= \frac{1}{\alpha} \sum_{i=1}^{O(\ln(n))} O\left(\frac{i}{n} + \frac{i}{e^{(i-1)/5}}\right) + o\left(\frac{1}{\alpha}\right).\end{aligned}$$

We use Lemma 3.3 to give an upper bound for the summation. This yields

$$\begin{aligned}\mathbb{E}(\mathsf{NN}) &= \frac{1}{\alpha} \sum_{i=1}^{O(\ln(n))} O\left(\frac{i}{n} + \frac{i}{e^{(i-1)/5}}\right) + o\left(\frac{1}{\alpha}\right) \\ &= \frac{1}{\alpha} \cdot O(1) + o\left(\frac{1}{\alpha}\right) \\ &= O\left(\frac{1}{\alpha}\right).\end{aligned}$$

□

Before we can prove the expected approximation ratio of `nearest neighbor`, we need one more lemma. This lemma states that the value of TSP can be bounded from below by the $n/2$ lightest edge weights in the graph. This lemma is followed directly by a proof of an upper bound of $O(\beta/\alpha)$ for the approximation ratio of `nearest neighbor`.

Lemma 5.8. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. Let $S_{n/2}$ be the sum of the $n/2$ lightest edge weights in G. Then we have $\mathsf{TSP} \geq S_{n/2}$, where TSP is the total distance of the optimal traveling salesman tour.*

Proof. By Lemma 5.4, we already know $\mathsf{MM} \geq S_{n/2}$, where MM is the distance of a minimum distance perfect matching. If we can prove $\mathsf{TSP} \geq \mathsf{MM}$, we are done. Indeed we have $\mathsf{TSP} \geq \mathsf{MM}$, since every tour through all vertices is the combination of two matchings. Both matchings can never be smaller than MM. Therefore we have $\mathsf{TSP} \geq 2 \cdot \mathsf{MM} \geq \mathsf{MM}$. □

Theorem 5.9. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. For such random metrics, we have*

$$\mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\right) = O\left(\frac{\beta}{\alpha}\right).$$

Proof. Let $c > 0$ be a sufficiently small constant. We use Lemma 3.4 to obtain

$$\mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\right) \leq \mathbb{E}\left(\frac{\beta}{c}\mathsf{NN}\right) + \mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}} \,\middle|\, \mathsf{TSP} < \frac{c}{\beta}\right) \cdot \mathbb{P}\left(\mathsf{TSP} < \frac{c}{\beta}\right).$$

We bound the first part by applying Theorem 5.7 and noting that c is a constant. We obtain

$$\mathbb{E}\left(\frac{\beta}{c}\mathsf{NN}\right) = O\left(\frac{\beta}{\alpha}\right).$$

We bound the second expectation from above by the worst-case approximation ratio of `nearest neighbor`, which is known to be $O(\ln(n))$ [16]. Finally, since c is sufficiently small, by Lemmas 5.8 and 5.4 with $\phi = 1/2$, we have

$$\mathbb{P}\left(\mathsf{TSP} < \frac{c}{\beta}\right) \leq \mathbb{P}\left(S_{n/2} < \frac{c}{\beta}\right) \leq e^{\frac{n}{2}(1+\ln(2c))}.$$

Combining these observations yields

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\right) &\leq \mathbb{E}\left(\frac{\beta}{c}\mathsf{NN}\right) + \mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}} \,\middle|\, \mathsf{TSP} < \frac{c}{\beta}\right) \cdot \mathbb{P}\left(\mathsf{TSP} < \frac{c}{\beta}\right) \\ &\leq O\left(\frac{\beta}{\alpha}\right) + O\left(\ln(n)\right) \cdot e^{\frac{n}{2}(1+\ln(2c))}.\end{aligned}$$

Since c is sufficiently small, the second part becomes $o(1)$. We conclude

$$\mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\right) = O\left(\frac{\beta}{\alpha}\right).$$

□

5.3 Insert for the Traveling Salesman Problem

In this section we consider another heuristic for the traveling salesman problem, which is given in Definition 5.6. We analyze the heuristic called `insert`. This heuristic takes the following steps:

1. Start with an initial optimal tour on a few vertices chosen according to some predefined rule R.
2. At every step, choose a vertex according to the same predefined rule R.
3. Insert the vertex in the current tour such that the total distance increases the least.

The performance of `insert` depends on the insertion rule R. However, for any rule R, the worst-case approximation ratio is at most $O(\ln(n))$ [16]. Furthermore, for random shortest path metrics on complete graphs ($\alpha = \beta = 1$ in Definition 4.2), for every rule R, the expected approximation ratio has an upper bound of $O(1)$ [4]. We extend this last result to general values for α and β and show an upper bound for the expected approximation ratio for every rule R of $O(\beta/\alpha)$.

From now on, let IN_R denote the cost of the tour computed by `insert` with rule R and let TSP denote the value of an optimal tour. In the following theorem, we show the expected value of IN_R is $O(1/\alpha)$.

Theorem 5.10. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. For such random metrics, we have for every rule R*

$$\mathbb{E}(\mathsf{IN}_R) = O\left(\frac{1}{\alpha}\right).$$

Proof. First of all, the expected cost of the initial tour T is the cost of an optimal tour through the initial vertices. Since we have a metric, we have $\mathsf{T} \leq \mathsf{TSP}$. We bound the expected cost of

TSP from above by the approximation ratio that we found for `nearest neighbor`. By Theorem 5.7 we have

$$\mathbb{E}(\mathsf{T}) \leq \mathbb{E}(\mathsf{TSP}) \leq \mathbb{E}(\mathsf{NN}) = O\left(\frac{1}{\alpha}\right).$$

Let $\Delta_i = \frac{i}{\alpha n}$. By Theorem 4.13 we can partition the vertices in an expected number of $O(1 + n/\exp(in/5))$ clusters, each of diameter at most $6\Delta_i$. We refer to these clusters as the i-clusters.

Suppose we have a partial tour and the next vertex to be inserted is v. If the tour has a vertex u that is in the same i-cluster as v, then the total distance of the tour is increased by at most $2d(u,v) \leq 12\Delta_i$ (triangle inequality). Therefore, for every i, only the insertion of the first vertex of an i-cluster can increase the distance by more than $12\Delta_i$. Since there is an expected number of $O(1+n/\exp(in/5))$ i-clusters, the number of vertices whose insertion would increase the distance by an amount in the range $(12\Delta_i, 12\Delta_{i+1}]$ is at most $O(1+n/\exp(in/5))$. Thus the contribution of insertions increasing the distance by an amount in the range $(12\Delta_i, 12\Delta_{i+1}]$ is at most $O(i/\alpha n + i/\alpha \exp(in/5))$

The contribution that we found can be used if i is not too large. For $i \geq \omega(\ln(n))$, we use Theorem 5.2 to show that the expected sum of all distances greater than or equal to $\Delta_{\omega(\ln(n))}$ is $o(1/\alpha)$. Therefore, we bound the contribution of insertions contributing a distance to the tour greater than or equal to $\Delta_{\omega(\ln(n))}$ by $o(1/\alpha)$.

Let c_i be the contribution of insertions increasing the tour distance by an amount in the range $(12\Delta_i, 12\Delta_{i+1}]$. Summing over all values of i yields

$$\begin{aligned}\mathbb{E}(\mathsf{IN}_R) &= \mathbb{E}(\mathsf{T}) + \sum_{i=0}^{O(\ln(n))} \mathbb{E}(c_i) + \sum_{i=\omega(\ln(n))}^{\infty} \mathbb{E}(c_i)\\ &= O\left(\frac{1}{\alpha}\right) + \sum_{i=0}^{O(\ln(n))} O\left(\frac{i+1}{\alpha n} + \frac{i+1}{\alpha e^{i/5}}\right) + o\left(\frac{1}{\alpha}\right)\\ &= O\left(\frac{1}{\alpha}\right) + \frac{1}{\alpha}\sum_{i=1}^{O(\ln(n))} O\left(\frac{i}{n} + \frac{i}{e^{(i-1)/5}}\right) + o\left(\frac{1}{\alpha}\right).\end{aligned}$$

We use Lemma 3.3 to give an upper bound for the summation. This yields

$$\begin{aligned}\mathbb{E}(\mathsf{IN}_R) &= O\left(\frac{1}{\alpha}\right) + \frac{1}{\alpha}\sum_{i=1}^{O(\ln(n))} O\left(\frac{i}{n} + \frac{i}{e^{(i-1)/5}}\right) + o\left(\frac{1}{\alpha}\right)\\ &= O\left(\frac{1}{\alpha}\right) + \frac{1}{\alpha}\cdot O(1) + o\left(\frac{1}{\alpha}\right)\\ &= O\left(\frac{1}{\alpha}\right).\end{aligned}$$

□

With this, we have everything we need to give an upper bound for the expected approximation ratio of `insert` independent of rule R.

Theorem 5.11. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V,E)$ be a graph that satisfies the α-β-cut-property and let $M = (V,d)$ be a corresponding random shortest path metric. For such random metrics, we have for every rule R*

$$\mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\right) = O\left(\frac{\beta}{\alpha}\right).$$

Proof. Let $c > 0$ be a sufficiently small constant. We use Lemma 3.4 to obtain

$$\mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\right) \leq \mathbb{E}\left(\frac{\beta}{c}\mathsf{IN}_R\right) + \mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}} \,\middle|\, \mathsf{TSP} < \frac{c}{\beta}\right) \cdot \mathbb{P}\left(\mathsf{TSP} < \frac{c}{\beta}\right).$$

We bound the first part by applying Theorem 5.10 and noting that c is a constant. We obtain

$$\mathbb{E}\left(\frac{\beta}{c}\mathsf{IN}_R\right) = O\left(\frac{\beta}{\alpha}\right).$$

We bound the second expectation from above by the worst-case approximation ratio of `insert`, which is known to be $O(\ln(n))$ [16]. Finally, since c is sufficiently small, by Lemmas 5.8 and 5.4 with $\phi = 1/2$, we have

$$\mathbb{P}\left(\mathsf{TSP} < \frac{c}{\beta}\right) \leq \mathbb{P}\left(S_{n/2} < \frac{c}{\beta}\right) \leq e^{\frac{n}{2}(1+\ln(2c))}.$$

Combining these observations yields

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\right) &\leq \mathbb{E}\left(\frac{\beta}{c}\mathsf{IN}_R\right) + \mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}} \mid \mathsf{TSP} < \frac{c}{\beta}\right) \cdot \mathbb{P}\left(\mathsf{TSP} < \frac{c}{\beta}\right)\\ &\leq O\left(\frac{\beta}{\alpha}\right) + O\left(\ln(n)\right) \cdot e^{\frac{n}{2}(1+\ln(2c))}.\end{aligned}$$

Since c is sufficiently small, the second part becomes $o(1)$. We conclude

$$\mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\right) = O\left(\frac{\beta}{\alpha}\right).$$

□

5.4 2-OPT for the Traveling Salesman Problem

The third and last heuristic that we cover for the traveling salesman problem, which is given in Definition 5.6, is called `2-opt`. This algorithm takes the following steps:

1. Start with an initial tour on all vertices.
2. Improve the tour by 2-exchanges, which are explained below.
3. Stop if no improvement can be made anymore.

A 2-exchange goes as follows. The algorithm takes vertices v_1, v_2, v_3, v_4 such that the vertices are visited in this order in the tour. Furthermore, v_2 is visited immediately after v_1 and v_4 is visited immediately after v_3. Then it replaces the order $\{v_1, v_2\}$ and $\{v_3, v_4\}$ by $\{v_1, v_3\}$ and $\{v_2, v_4\}$ if and only if this creates a shorter tour. We only have a 2-exchange if an improvement is made. An example of a 2-exchange is given in Figure 4

In contrast to the analysis of previous heuristics, we do not give a bound for the approximation ratio of `2-opt`. Instead, we give an upper bound for the expected number of iterations `2-opt` takes until the algorithm terminates. In the worst-case scenario, the number of iterations is exponential. However, for random shortest path metrics on complete graphs ($\alpha = \beta = 1$ in Definition 4.2), an upper bound of $O(n^8 \ln^3(n))$ is known for the expected number of iterations [4]. We extend

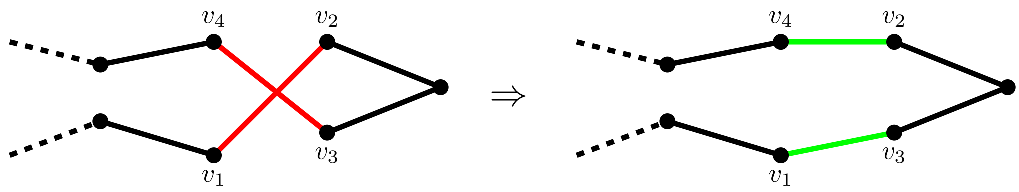

Figure **4**: Example of a 2-exchange

this result to general values for α and β and show an upper bound for the expected number of iterations of $O(n^8 \ln^3(n)\beta/\alpha)$.

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

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

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

Define $\zeta^{e,I} = \zeta_{12}^{e,I} + \zeta_{34}^{e,I} - \zeta_{13}^{e,I} - \zeta_{24}^{e,I}$. We use this information for the following two lemmas.

Lemma 5.12. *Let $G = (V, E)$ be a graph and let $M = (V, d)$ be a corresponding random shortest path metric. Then there exists an edge e and a set I such that $\zeta = \zeta^{e,I} + \gamma w(e)$, where $\gamma \in \{-2, -1, 1, 2\}$ is determined by e and I.*

Proof. Consider the four shortest paths from the 2-exchange. As previously stated there exists some edge e with non-zero value γ_e. Choose this e, the corresponding set I and take $\gamma = \gamma_e$. If $e \notin P_{ij}$, then by definition $\zeta_{ij}^{e,I} = d(v_i, v_j)$. If $e \in P_{ij}$, then by definition $\zeta_{ij}^{e,I} = d(v_i, v_j) - w(e)$. Adding this all together gives

$$\zeta^{e,I} + \gamma w(e) = d(v_1, v_2) + d(v_3, v_4) - d(v_1, v_3) - d(v_2, v_4) - \gamma_e w(e) + \gamma_e w(e) = \zeta.$$

$\square$

Lemma 5.13. *Let $G = (V, E)$ be a graph and let $M = (V, d)$ be a corresponding random shortest path metric. Let e and I be given with with $\gamma = \gamma_e \neq 0$. Then $\mathbb{P}(\zeta^{e,I} + \gamma w(e) \in (0, x]) \leq x$.*

Proof. Fix all edge weights except for $w(e)$. Then the value of $\zeta^{e,I}$ is known. Therefore we have $\zeta^{e,I} + \gamma w(e) \in (0, x]$ if and only if $w(e) \in (-\zeta^{e,I}/\gamma, (-\zeta^{e,I} + x)/\gamma]$, an interval of length $x/|\gamma| \leq x$. Since $w(e)$ is drawn from Exp(1) and the density function of this distribution does not exceed 1, we have

$$\mathbb{P}\left(\zeta^{e,I} + \gamma w(e)) \in (0, x]\right) = \mathbb{P}\left(w(e) \in (-\zeta^{e,I}/\gamma, (-\zeta^{e,I} + x)/\gamma]\right) \leq x/|\gamma| \leq x.$$

□

The previous two lemmas enable us to give an upper bound for the probability that a 2-exchange decreases the length of the tour by at most x. This is done in the next lemma, which is followed directly by a proof of an upper bound of $O(n^8 \ln^3(n)\beta/\alpha)$ for the number of iterations of `2-opt`.

Lemma 5.14. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. We have*

$$\mathbb{P}(\zeta \in (0, x]) = O\left(\beta n^2 x\right)$$

Proof. Using Lemma 5.12 we find

$$\mathbb{P}(\zeta \in (0, x]) \leq \mathbb{P}\left(\exists\, e, I \,\middle|\, \zeta^{e,I} + \gamma_e w(e) \in (0, x]\right).$$

The number of possible choices for edge e and corresponding set I is bounded by the number of edges in the graph. Thus the number of possible choices is bounded from above by $\beta n(n-1)/2$. Taking a union bound over all choices of e and I yields

$$\begin{aligned}
\mathbb{P}(\zeta \in (0, x]) &\leq \mathbb{P}\left(\exists\, e, I \,\middle|\, \zeta^{e,I} + \gamma_e w(e) \in (0, x]\right) \\
&\leq O\left(\beta n^2\right) \cdot \mathbb{P}\left(\zeta^{e,I} + \gamma_e w(e) \in (0, x]\right) \\
&= O\left(\beta n^2 x\right),
\end{aligned}$$

where the last inequality is a result of Lemma 5.13. □

Theorem 5.15. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. The expected number of iterations of* `2-opt` *until a local optimum is found is bounded from above by $O(n^8 \ln^3(n)\beta/\alpha)$.*

Proof. Let $\zeta_{\min} > 0$ be the minimum improvement that can be made by a 2-exchange. The total number of different 2-exchanges is bounded from above by $|V|^2 \cdot |V|^2 = n^4$. Therefore, by applying a union bound over all possible 2-exchanges, we obtain

$$\mathbb{P}\left(\zeta_{\min} \leq y\right) \leq n^4 \cdot \mathbb{P}(\zeta \in (0, y]).$$

Applying Lemma 5.14 yields

$$\begin{aligned}
\mathbb{P}\left(\zeta_{\min} \leq y\right) &\leq n^4 \cdot \mathbb{P}(\zeta \in (0, y]) \\
&= O(\beta n^6 y).
\end{aligned}$$

The initial tour has a length of at most $n\Delta_{\max}$, where $\Delta_{\max}$ is the maximum distance between two vertices in the metric. Let T be the number of iterations taken by `2-opt`. Since the length of an optimal tour is non-negative we have $T \leq n\Delta_{\max}/\zeta_{\min}$. This yields that $T > x$ implies $\Delta_{\max}/\zeta_{\min} > x/n$. This is contained in the union of the events $\Delta_{\max} > c\ln(x)\ln(n)/\alpha n$ and $\zeta_{\min} < c\ln(x)\ln(n)/\alpha x$, where c is a sufficiently large constant. We obtain

$$\mathbb{P}(T > x) \leq \mathbb{P}\left(\Delta_{\max} > c\ln(x)\frac{\ln(n)}{\alpha n}\right) + \mathbb{P}\left(\zeta_{\min} < c\ln(x)\frac{\ln(n)}{\alpha x}\right).$$

By Theorem 4.11 we have

$$\mathbb{P}\left(\Delta_{\max} > c\ln(x)\frac{\ln(n)}{\alpha n}\right) = O\left(n^{2-c\ln(x)/8}\right).$$

Since c is sufficiently large, this becomes $O(n^{-\Omega(\ln(x))})$. Furthermore, we have

$$\mathbb{P}\left(\zeta_{\min} < c\ln(x)\frac{\ln(n)}{\alpha x}\right) = O\left(\frac{\beta}{\alpha}n^6\ln(n)\frac{\ln(x)}{x}\right).$$

Since a total of $n!$ different tours through n vertices exist, we know that T is bounded from above by $n!$. Since c is sufficiently large, we obtain

$$\begin{aligned}
\mathbb{E}(T) &= \sum_{x=1}^{n!}\mathbb{P}(T > x) \\
&\le \sum_{x=1}^{n!}\mathbb{P}\left(\Delta_{\max} > c\ln(x)\frac{\ln(n)}{\alpha n}\right) + \sum_{x=1}^{n!}\mathbb{P}\left(\zeta_{\min} < c\ln(x)\frac{\ln(n)}{\alpha x}\right) \\
&= \sum_{x=1}^{n!}O\left(n^{-\Omega(\ln(x))}\right) + \sum_{x=1}^{n!}O\left(\frac{\beta}{\alpha}n^6\ln(n)\frac{\ln(x)}{x}\right) \\
&= O\left(\frac{\beta}{\alpha}n^8\ln^3(n)\right),
\end{aligned}$$

where the last equality follows from Lemma 3.6. □

5.5 Trivial for the k-Median Problem

The final problem that we look into is the k-median problem. This problem is defined as follows.

Definition 5.16 (k-Median Problem)**.** *Let a metric $M = (V, d)$ be given. The goal of the k-median problem is to find a set $U \subset V$ of size k such that the sum of the distances from all nodes to U is minimized. Thus the goal is to find a set U of size k such that $\sum_{v\in V}\min_{u\in U} d(v, u)$ is minimized.*

We analyze one heuristic for the k-median problem. This heuristic is called `trivial`. This is an algorithm that picks k vertices independently of the metric space M. The worst-case approximation ratio of `trivial` is unbounded. However, for random shortest path metrics on complete graphs ($\alpha = \beta = 1$ in Definition 4.2), the expected approximation ratio has an upper bound of $O(1)$ and even $1 + o(1)$ for k sufficiently small [4]. We extend this result to general values for α and β and give an upper bound for the expected approximation ratio of $O(\beta/\alpha)$ for k large and $\beta/\alpha + o(\beta/\alpha)$ for k sufficiently small.

For our analysis, let $U = \{v_1, \ldots, v_k\}$ be an arbitrary set of k vertices. Define $\mathsf{Cost}(U)$ as the total distance from U to all vertices in V. Furthermore, sort all other vertices $\{v_{k+1}, \ldots, v_n\}$ in increasing distance from U. For $k + 1 \le i \le n$, let $\rho_i = d(v_i, U)$ equal the distance from U to the $(i - k)$-th closest vertex to U. Finally, let TR denote the cost of the solution created by `trivial` and let ME be the cost of an optimal solution to the k-median problem. We start by bounding the expected value of TR.

Lemma 5.17. *Let $0 < \alpha \le \beta \le 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property, let $M = (V, d)$ be a corresponding random shortest path metric and let U be an arbitrary set of k*

vertices. Then we have

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

Proof. We analyze the distribution of ρ_i in the same way as we analyzed the distribution of $\tau_k(v, G)$ in Section 4.2. The only difference is that we start with k vertices instead of one vertex. We obtain for $k+1 \leq i \leq n$

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

Since ρ_i equals the distance to the $(i-k)$-th closest vertex to U, we have

$$\mathsf{Cost}(U) = \sum_{i=k+1}^{n} \rho_i.$$

We combine these results to obtain

$$\sum_{j=k}^{n-1} (n-j)\,\mathrm{Exp}(\beta j(n-j)) \leq \mathsf{Cost}(U) \leq \sum_{j=k}^{n-1} (n-j)\,\mathrm{Exp}(\alpha j(n-j)),$$

which can be simplified to

$$\sum_{j=k}^{n-1} \mathrm{Exp}(\beta j) \leq \mathsf{Cost}(U) \leq \sum_{j=k}^{n-1} \mathrm{Exp}(\alpha j).$$

□

Lemma 5.18. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. For such random metrics, we have*

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

Proof. Since `trivial` picks k vertices, we have

$$\mathbb{E}(\mathsf{TR}) = \mathbb{E}(\mathsf{Cost}(U)).$$

Applying Lemma 5.17 yields

$$\sum_{i=k}^{n-1} \frac{1}{\beta i} \leq \mathbb{E}(\mathsf{TR}) \leq \sum_{i=k}^{n-1} \frac{1}{\alpha i}.$$

Using the harmonic numbers we can rewrite this to

$$\frac{1}{\beta}\left(H_{n-1} - H_{k-1}\right) \leq \mathbb{E}(\mathsf{TR}) \leq \frac{1}{\alpha}\left(H_{n-1} - H_{k-1}\right).$$

Since $\ln(i) \leq H_i \leq \ln(i) + 1$ holds for all i, we obtain

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

□

In order to bound the expected value of the ratio TR/ME, we need to bound the probability that ME is very small and the probability that TR is very large. In the following three lemmas we give a bound from above for the probability that ME is very small for sufficiently small values of k and for large values of k. Afterwards we give a bound from above for the probability that TR is very large.

Lemma 5.19. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property, let $M = (V, d)$ be a corresponding random shortest path metric and let U be an arbitrary set of k vertices. Then the probability density function of $\mathsf{Cost}(U)$ is bounded from above by*

$$f(x) \leq \beta k \binom{n-1}{k} e^{-\beta k x} \left(1 - e^{-\beta x}\right)^{n-k-1}.$$

Proof. By Lemma 5.17 we have

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

According to Lemma 3.7, the right hand side has the same distribution as the $(n-k)$-th order statistic out of $n-1$ random variables with distribution $\mathrm{Exp}(\beta)$. The density of this order statistic is known to be the density we want to proof [17, Example 2.38]. The result follows. □

Lemma 5.20. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. Let $c > 0$ be sufficiently large and let $k \leq c'n$ for $c' = c'(c) > 0$ sufficiently small. Then we have*

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

Proof. By Lemma 5.19, for every fixed $U \subset V$ of size k, we have the following upper bound for the density function of $\mathsf{Cost}(U)$:

$$f(x) \leq \beta k \binom{n-1}{k} e^{-\beta k x} \left(1 - e^{-\beta x}\right)^{n-k-1}.$$

We want to bound this density function from above at $x = \ln((n-1)/ak)/\beta$ for sufficiently large a with $1 \leq a \leq (n-1)/k$. For this particular value of x we have

$$f(x) \leq \beta k \binom{n-1}{k} \left(\frac{ak}{n-1}\right)^k \left(1 - \frac{ak}{n-1}\right)^{n-k-1}.$$

Applying the inequality $\binom{a}{b} \leq (ae/b)^b$ yields

$$\begin{aligned} f(x) &\leq \beta k \frac{(n-1)^k e^k}{k^k} \left(\frac{ak}{n-1}\right)^k \left(1 - \frac{ak}{n-1}\right)^{n-k-1} \\ &\leq \beta k (ae)^k \left(1 - \frac{ak}{n-1}\right)^{n-k-1}. \end{aligned}$$

Since $1 + x \leq \exp(x)$, we obtain

$$f(x) \leq \beta k (ae)^k e^{ak(n-k-1)/(n-1)}.$$

Since k is sufficiently small, we have $(n-k-1)/(n-1) = \Omega(1)$. We bound the exponent to obtain

$$\begin{aligned} f(x) &\le \beta k(ae)^k e^{-\Omega(ak)} \\ &= \beta e^{\ln(k)+k\ln(a)+k} \cdot e^{-\Omega(ak)}. \end{aligned}$$

As a is sufficiently large, we have $\ln(k) + k\ln(a) + k = o(ak)$. This yields

$$\begin{aligned} f(x) &\le \beta e^{\ln(k)+k\ln(a)+k} \cdot e^{-\Omega(ak)} \\ &= \beta e^{o(ak)} \cdot e^{-\Omega(ak)} \\ &= \beta e^{-\Omega(ak)}. \end{aligned}$$

Earlier we chose $x = \ln((n-1)/ak)/\beta$. This can be rewritten to $a = (n-1)\exp(-\beta x)/k$. Substituting this value for a yields

$$f(x) \le \beta e^{-\Omega((n-1)\exp(-\beta x))},$$

which holds for $x \in [0, \ln((n-1)/bk)/\beta]$ for $b \ge 1$ sufficiently large. This constraint ensures that a is sufficiently large.

We use the density function to bound the probability that $\mathsf{Cost}(\mathsf{U})$ is small. We obtain

$$\begin{aligned} \mathbb{P}\left(\mathsf{Cost}(U) < \frac{1}{\beta}\ln\left(\frac{n-1}{bk}\right)\right) &= \int_0^{\frac{1}{\beta}\ln\left(\frac{n-1}{bk}\right)} f(x)\mathrm{d}x \\ &= \int_0^{\frac{1}{\beta}\ln\left(\frac{n-1}{bk}\right)} f\left(\frac{1}{\beta}\ln\left(\frac{n-1}{bk}\right) - x\right)\mathrm{d}x \\ &\le \int_0^{\frac{1}{\beta}\ln\left(\frac{n-1}{bk}\right)} \beta e^{-\Omega(bk\exp(\beta x))}\mathrm{d}x. \end{aligned}$$

Substituting $y = \beta x$ yields

$$\begin{aligned} \mathbb{P}\left(\mathsf{Cost}(U) < \frac{1}{\beta}\ln\left(\frac{n-1}{bk}\right)\right) &\le \int_0^{\frac{1}{\beta}\ln\left(\frac{n-1}{bk}\right)} \beta e^{-\Omega(bk\exp(\beta x))}\mathrm{d}x \\ &= \int_0^{\ln\left(\frac{n-1}{bk}\right)} e^{-\Omega(bk\exp(y))}\mathrm{d}y. \end{aligned}$$

Since $1 + x \le \exp(x)$, we obtain

$$\begin{aligned} \mathbb{P}\left(\mathsf{Cost}(U) < \frac{1}{\beta}\ln\left(\frac{n-1}{bk}\right)\right) &\le \int_0^{\ln\left(\frac{n-1}{bk}\right)} e^{-\Omega(bk\exp(y))}\mathrm{d}y \\ &\le \int_0^{\ln\left(\frac{n-1}{bk}\right)} e^{-\Omega(bk(1+y))}\mathrm{d}y \\ &\le e^{-\Omega(bk)} \int_0^{\infty} e^{-\Omega(bky)}\mathrm{d}y \\ &\le e^{-\Omega(bk)}, \end{aligned}$$

where the last step follows from the fact that $\int_0^\infty \exp(-\Omega(bky))\mathrm{d}y = O(1/bk) \le 1$ as b is sufficiently large.

In order for ME to be small, there must exist a subset $U \subset V$ of size k that has a small cost. We bound the probability that ME is small by taking a a union bound over all subsets $U \subset V$ of size k. This yields

$$\begin{aligned}\mathbb{P}\left(\mathsf{ME} < \frac{1}{\beta}\ln\left(\frac{n-1}{bk}\right)\right) &= \mathbb{P}\left(\exists U \subset V, |U| = k \;\middle|\; \mathsf{Cost}(U) < \frac{1}{\beta}\ln\left(\frac{n-1}{bk}\right)\right)\\ &\le \binom{n}{k}\cdot\mathbb{P}\left(\mathsf{Cost}(U) < \frac{1}{\beta}\ln\left(\frac{n-1}{bk}\right)\right)\\ &\le \binom{n}{k}\cdot e^{-\Omega(bk)}.\end{aligned}$$

Set $b = c\ln(n/k)$ for sufficiently large $c \geq 1$. Then we fulfill the condition that $b \geq 1$ and sufficiently large. This yields

$$\begin{aligned}\mathbb{P}\left(\mathsf{ME} < \frac{\ln\left(\frac{n-1}{ck\ln\left(\frac{n}{k}\right)}\right)}{\beta}\right) &= \mathbb{P}\left(\mathsf{ME} < \frac{\ln\left(\frac{n-1}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln(c)}{\beta}\right)\\ &\le \binom{n}{k}\cdot e^{-\Omega(ck\ln(n/k))}\\ &= \binom{n}{k}\cdot\left(\frac{n}{k}\right)^{-\Omega(ck)}.\end{aligned}$$

Since $\binom{a}{b} \leq (ae/b)^b$, we obtain

$$\begin{aligned}\mathbb{P}\left(\mathsf{ME} < \frac{\ln\left(\frac{n-1}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln(c)}{\beta}\right) &\le \binom{n}{k}\cdot\left(\frac{n}{k}\right)^{-\Omega(ck)}\\ &\le \left(\frac{en}{k}\right)^{k}\cdot\left(\frac{n}{k}\right)^{-\Omega(ck)}.\end{aligned}$$

Since k is sufficiently smaller than n, we have $en/k \leq (n/k)^2$. As c is sufficiently large, we obtain

$$\begin{aligned}\mathbb{P}\left(\mathsf{ME} < \frac{\ln\left(\frac{n-1}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln(c)}{\beta}\right) &\le \left(\frac{en}{k}\right)^{k}\cdot\left(\frac{n}{k}\right)^{-\Omega(ck)}\\ &\le \left(\frac{n}{k}\right)^{-\Omega(ck)}.\end{aligned}$$

Finally, since $k \geq 1$ and k is sufficiently smaller than n, we have $(n/k)^k \geq n$. This yields

$$\begin{aligned}\mathbb{P}\left(\mathsf{ME} < \frac{\ln\left(\frac{n-1}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln(c)}{\beta}\right) &\le \left(\frac{n}{k}\right)^{-\Omega(ck)}\\ &\le n^{-\Omega(c)}.\end{aligned}$$

□

Lemma 5.21. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. Let $k \leq (1-\varepsilon)n$ for some constant $\varepsilon > 0$. For every $c \in [0, 2\varepsilon^2)$, we have*

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

Proof. The value of ME is the sum of $n-k$ distances in M and therefore the sum of $n-k$ shortest path lengths in G. Since all these paths connect a vertex to one of the vertices in the optimal solution of k-median, the union of the paths uses at least $n-k$ different edges from G. Let S_m be the sum of the m lightest edge weights in G. We obtain $\mathsf{ME} \geq S_{n-k}$. Since $k \leq (1-\varepsilon)n$ for some constant $\varepsilon > 0$, we have $n-k \geq \varepsilon n$ and thus $S_{n-k} \geq S_{\varepsilon n}$. This yields

$$\mathbb{P}\left(\mathsf{ME} < \frac{c}{\beta}\right) \leq \mathbb{P}\left(S_{\varepsilon n} < \frac{c}{\beta}\right).$$

We apply Lemma 5.4 with $\phi = \varepsilon$ to the right-hand side to obtain

$$\begin{aligned}\mathbb{P}\left(\mathsf{ME} < \frac{c}{\beta}\right) &\leq e^{\varepsilon n\left(1+\ln\left(\frac{c}{2\varepsilon^2}\right)\right)} \\ &= e^{\Omega(\ln(c)n)} \\ &= c^{\Omega(n)}.\end{aligned}$$

□

Lemma 5.22. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. For any $c \geq 4$ we have*

$$\mathbb{P}\left(\mathsf{TR} > n^c\right) \leq e^{-n^{c/4}}.$$

Proof. We can roughly bound TR by n times the diameter, $\Delta_{\max}$, of the metric. Similarly, the diameter of the metric is the sum of the weights of at most n edges from G. This gives us the following bound:

$$\mathsf{TR} \leq n \cdot \Delta_{\max} \leq n^2 \cdot \max_{e \in E}\{w(e)\}.$$

Since the α-β-cut-property holds, G has at most $\beta\binom{n}{2}$ edges. Therefore, $\max_{e\in E} w(e)$ is the maximum of at most $\beta\binom{n}{2}$ independent exponentially distributed random variables with parameter 1. Each of these has a cumulative distribution function of $1 - \exp(-x)$. This yields

$$\begin{aligned}\mathbb{P}\left(\mathsf{TR} \leq n^c\right) &\geq \mathbb{P}\left(\max_{e \in E}\{w(e)\} \leq n^{c-2}\right) \\ &= \left(1 - e^{-n^{c-2}}\right)^{\beta\binom{n}{2}}.\end{aligned}$$

We apply the Bernoulli inequality $(1+x)^y \geq 1 + xy$, which holds for $x \geq -1$ and $y \notin (0, 1)$, to obtain

$$\begin{aligned}\mathbb{P}\left(\mathsf{TR} \leq n^c\right) &\geq \left(1 - e^{-n^{c-2}}\right)^{\beta\binom{n}{2}} \\ &\geq 1 - \beta\binom{n}{2}e^{-n^{c-2}}.\end{aligned}$$

The inequality $\binom{n}{2}\exp(-n^{c-2}) \leq \exp(-n^{c-3})$ combined with $\beta \leq 1$ yields

$$\begin{aligned}\mathbb{P}\left(\mathsf{TR} \leq n^c\right) &\geq 1 - \beta\binom{n}{2}e^{-n^{c-2}} \\ &\geq 1 - e^{-n^{c-3}} \\ &\geq 1 - e^{-n^{c/4}},\end{aligned}$$

where the last inequality follows from the assumption $c \geq 4$. The desired result follows. □

With this, we have everything we need to prove our main result for `trivial`, an upper bound for its expected approximation ratio.

Theorem 5.23. *Let $0 < \alpha \leq \beta \leq 1$, let $G = (V, E)$ be a graph that satisfies the α-β-cut-property and let $M = (V, d)$ be a corresponding random shortest path metric. Let $k \leq (1-\varepsilon)n$ for some constant $\varepsilon > 0$. Then*

$$\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}}\right) = O\left(\frac{\beta}{\alpha}\right).$$

Moreover, if we have $k \leq c'n$ for some fixed $c' \in (0,1)$ sufficiently small, then

$$\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}}\right) = \frac{\beta}{\alpha}\left(1 + O\left(\frac{\ln\ln\left(\frac{n}{k}\right)}{\ln\left(\frac{n}{k}\right)}\right)\right).$$

Proof. By Lemma 3.4, we have for all constants $m > 0$

$$\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}}\right) \leq \mathbb{E}\left(\frac{\beta}{m}\mathsf{TR}\right) + \mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) \cdot \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, \mathsf{ME} < \frac{m}{\beta}\right)$$

Case 1 ($k \leq c'n$, c' sufficiently small): According to Lemma 5.20 we can pick a constant $c > 0$ sufficiently large such that

$$\mathbb{P}\left(\mathsf{ME} \leq \frac{\ln\left(\frac{n-1}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln(c)}{\beta}\right) \leq n^{-9}. \tag{1}$$

Given this value of c, choose $m = \ln((n-1)/k) - \ln\ln(n/k) - \ln(c)$. By Lemma 5.18, we have

$$\begin{aligned}
\mathbb{E}\left(\frac{\beta}{m}\mathsf{TR}\right) &\leq \frac{\beta}{\alpha} \cdot \frac{\ln\left(\frac{n-1}{k-1}\right) + 1}{\ln\left(\frac{n-1}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln(c)} \\
&= \frac{\beta}{\alpha} \cdot \frac{\ln\left(\frac{n-1}{k}\right) + O(1)}{\ln\left(\frac{n-1}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln(c)} \\
&= \frac{\beta}{\alpha} \cdot \left(1 + \frac{\ln\ln\left(\frac{n}{k}\right) + \ln(c) + O(1)}{\ln\left(\frac{n-1}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln(c)}\right).
\end{aligned}$$

Since c is a constant, the dominating term in the numerator is $\ln\ln(n/k)$ and in the denominator $\ln((n-1)/k) = O(\ln(n/k))$. This yields

$$\begin{aligned}
\mathbb{E}\left(\frac{\beta}{m}\mathsf{TR}\right) &\leq \frac{\beta}{\alpha} \cdot \left(1 + \frac{\ln\ln\left(\frac{n}{k}\right) + \ln(c) + O(1)}{\ln\left(\frac{n-1}{k}\right) - \ln\ln\left(\frac{n}{k}\right) - \ln(c)}\right) \\
&= \frac{\beta}{\alpha} \cdot \left(1 + O\left(\frac{\ln\ln\left(\frac{n}{k}\right)}{\ln\left(\frac{n}{k}\right)}\right)\right).
\end{aligned}$$

For the second part we have

$$\mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) \cdot \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, \mathsf{ME} < \frac{m}{\beta}\right) = \mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) \cdot \int_0^\infty \mathbb{P}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \geq x \,\middle|\, \mathsf{ME} < \frac{m}{\beta}\right) \mathrm{d}x.$$

We use (1) to obtain

$$
\begin{aligned}
\mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) \cdot \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, \mathsf{ME} < \frac{m}{\beta}\right) &= \mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) \cdot \int_0^\infty \mathbb{P}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \geq x \,\middle|\, \mathsf{ME} < \frac{m}{\beta}\right) \mathrm{d}x \\
&\leq n^8 \cdot \mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) + \int_{n^8}^\infty \mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) \cdot \mathbb{P}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \geq x \,\middle|\, \mathsf{ME} < \frac{m}{\beta}\right) \mathrm{d}x \\
&\leq O\left(\frac{1}{n}\right) + \int_{n^8}^\infty \mathbb{P}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \geq x \text{ and } \mathsf{ME} < \frac{m}{\beta}\right) \mathrm{d}x \\
&\leq O\left(\frac{1}{n}\right) + \int_{n^8}^\infty \mathbb{P}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \geq x\right) \mathrm{d}x.
\end{aligned}
$$

Since $\mathsf{TR}/\mathsf{ME} \geq x$ implies $\mathsf{TR} \geq \sqrt{x}$ or $\mathsf{ME} \leq 1/\sqrt{x} \leq 1/\beta\sqrt{x}$, we can apply Lemma 5.21 and Lemma 5.22 and obtain

$$
\begin{aligned}
\mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) \cdot \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, \mathsf{ME} < \frac{m}{\beta}\right) &\leq O\left(\frac{1}{n}\right) + \int_{n^8}^\infty \mathbb{P}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \geq x\right) \mathrm{d}x \\
&\leq O\left(\frac{1}{n}\right) + \int_{n^8}^\infty 2\max\left\{\mathbb{P}\left(\mathsf{TR} \geq \sqrt{x}\right),\ \mathbb{P}\left(\mathsf{ME} \leq \frac{1}{\beta\sqrt{x}}\right)\right\} \mathrm{d}x \\
&\leq O\left(\frac{1}{n}\right) + \int_{n^8}^\infty 2\max\left\{\exp\left(-x^{1/8}\right),\ \left(\frac{1}{\sqrt{x}}\right)^{\Omega(n)}\right\} \mathrm{d}x \\
&= O\left(\frac{1}{n}\right).
\end{aligned}
$$

Combining these results yields

$$
\begin{aligned}
\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}}\right) &\leq \mathbb{E}\left(\frac{\beta}{m}\mathsf{TR}\right) + \mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) \cdot \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, \mathsf{ME} < \frac{m}{\beta}\right) \\
&\leq \frac{\beta}{\alpha} \cdot \left(1 + O\left(\frac{\ln\ln\left(\frac{n}{k}\right)}{\ln\left(\frac{n}{k}\right)}\right)\right) + O\left(\frac{1}{n}\right) \\
&= \frac{\beta}{\alpha} \cdot \left(1 + O\left(\frac{\ln\ln\left(\frac{n}{k}\right)}{\ln\left(\frac{n}{k}\right)}\right)\right).
\end{aligned}
$$

Case 2 ($c'n < k \leq (1-\varepsilon)n$, $\varepsilon > 0$): We repeat the proof for the previous part, but this time we choose m as a sufficiently small constant. Then, by Lemma 5.21, we have

$$
\mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) = m^{\Omega(n)} \leq n^{-9}. \tag{2}
$$

Furthermore, since m is a constant and $(n-1)/(k-1) = O(1)$, by Lemma 5.18, we have

$$
\mathbb{E}\left(\frac{\beta}{m}\mathsf{TR}\right) \leq \frac{\beta}{\alpha} \cdot \frac{\ln\left(\frac{n-1}{k-1}\right) + 1}{m} = O\left(\frac{\beta}{\alpha}\right).
$$

With identical reasoning as for case 1, this time using (2) instead of (1), we can prove

$$
\mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) \cdot \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, \mathsf{ME} < \frac{m}{\beta}\right) = O\left(\frac{1}{n}\right).
$$

Combining these results yields

$$
\begin{aligned}
\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}}\right) &\leq \mathbb{E}\left(\frac{\beta}{m}\mathsf{TR}\right) + \mathbb{P}\left(\mathsf{ME} < \frac{m}{\beta}\right) \cdot \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \;\middle|\; \mathsf{ME} < \frac{m}{\beta}\right) \\
&\leq O\left(\frac{\beta}{\alpha}\right) + O\left(\frac{1}{n}\right) \\
&= O\left(\frac{\beta}{\alpha}\right).
\end{aligned}
$$

□

6 Application to Erdős–Rényi Random Graphs

So far, we have analyzed random shortest path metrics applied to graphs that satisfy the α-β-cut-property (Definition 4.2). For these random metrics, the performance of several heuristics has been analyzed. In this section, we first show that instances of the Erdős–Rényi random graph model satisfy the α-β-cut-property with high probability for suitable values of α and β. We use this to prove our main results; upper bounds for the performance of the analyzed heuristics in random shortest path metrics applied to Erdős–Rényi random graphs.

6.1 ERRG cut property

In the $G(n,p)$ model, every edge exists with an independent probability p. For instances of the $G(n,p)$ model, we introduce the following specification of the α-β-cut-property, where α and β depend on the value of p.

Definition 6.1 (ERRG cut property)**.** *Let $G=(V,E)$ be an instance of the $G(n,p)$ model and let $\varepsilon \in (0,1)$. Then G has the ERRG cut property if G satisfies the α-β-cut-property for $\alpha=(1-\varepsilon)p$ and $\beta=(1+\varepsilon)p$. Thus G has the ERRG cut property if the following holds true for all $U \subset V$:*

$$\Big| |c(U)| - p\mu_U \Big| \leq \varepsilon p \mu_U,$$

where $|c(U)|$ is the size of the cut induced by U and $\mu_U = |U|(n-|U|)$.

When p is sufficiently large, the ERRG cut property holds with high probability. This is shown in the following lemma.

Lemma 6.2. *Let $G=(V,E)$ be an instance of the $G(n,p)$ model. For constant $\varepsilon \in (0,1)$ and $p \geq C\frac{\ln(n)}{n}$, where $C > \frac{9}{\varepsilon^2}$ is constant, the probability that the ERRG cut property does not hold is at most $o\left(\frac{1}{n^2}\right)$.*

Proof: Using Definition 6.1, we find that the probability that the ERRG cut property does not hold is given by

$$\mathbb{P}\left(\exists U \subset V \ : \ \Big||c(U)| - p\mu_U\Big| > \varepsilon p \mu_U\right).$$

In order to bound this probability from above, we observe that for all U we have that U and $V \backslash U$ induce the same cut. Because of this, we can restrict ourselves to the case where $|U| \leq n/2$.

Let B_U be the event that $\Big||c(U)| - p\mu_U)\Big| > \varepsilon p \mu_U$. Then, by applying a union bound, we obtain

$$\mathbb{P}\left(\exists U \subset V \ : \ \Big||c(U)| - p\mu_U\Big| > \varepsilon p \mu_U\right) = \mathbb{P}\left(\bigcup_{U \subset V} B_U\right) \leq \sum_{\substack{U \subset V \\ |U| \leq n/2}} \mathbb{P}(B_U). \tag{3}$$

Note that $p\mu_U = p|U|(n-|U|)$ equals the expected size of the cut induced by U in the $G(n,p)$ model. We use this to apply the two-sided chernoff bound from Lemma 3.8. For every $U \subset V$ and $\varepsilon \in (0,1)$, we obtain:

$$\mathbb{P}(B_U) = \mathbb{P}\left(\Big||c(U)| - p\mu_U\Big| > \varepsilon p \mu_U\right) \leq 2e^{-p\mu_U \cdot \varepsilon^2/3}.$$

Substituting this result in (3) and observing that there exist exactly $\binom{n}{k}$ subsets U of size k in a graph with n nodes yields

$$\begin{aligned}\mathbb{P}\left(\exists U \subset V \ : \ \Big||c(U)| - p\mu_U\Big| > \varepsilon p\mu_U\right) &\leq \sum_{\substack{U\subset V\\ |U|\leq n/2}} P(B_U)\\ &\leq \sum_{k=1}^{n/2} \binom{n}{k} \cdot 2e^{-k(n-k)p\varepsilon^2/3}.\end{aligned}$$

In each summand, the exponent is negative and monotonically decreasing in p. We get an upper bound by substituting the minimal value of p. This yields

$$\begin{aligned}\mathbb{P}\left(\exists U \subset V \ : \ \Big||c(U)| - p\mu_U\Big| > \varepsilon p\mu_U\right) &\leq \sum_{k=1}^{n/2} \binom{n}{k} \cdot 2e^{-k(n-k)p\varepsilon^2/3}\\ &\leq \sum_{k=1}^{n/2} \binom{n}{k} \cdot 2e^{-k(n-k)C\ln(n)\varepsilon^2/3n}.\end{aligned}$$

Choose $\xi > 0$ such that $1 - (1-\xi)C\varepsilon^2/3 < -2$. Note that ξ can only be a positive real number when $C > 9/\varepsilon^2$. Furthermore, we must have $\xi < 1$. We use ξ to split our summation as follows:

$$\begin{aligned}\mathbb{P}\left(\exists U \subset V \ : \ \Big||c(U)| - p\mu_U\Big| > \varepsilon p\mu_U\right) &\leq \sum_{k=1}^{n/2} \binom{n}{k} \cdot 2e^{-k(n-k)C\ln(n)\varepsilon^2/3n}\\ &\leq 2\sum_{k=1}^{\xi n} \binom{n}{k} \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n} + 2\sum_{k=\xi n}^{n/2} \binom{n}{k} \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n}.\end{aligned}$$

For the first summation we first apply the inequality $\binom{n}{k} \leq n^k$ to find

$$\begin{aligned}\sum_{k=1}^{\xi n} \binom{n}{k} \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n} &\leq \sum_{k=1}^{\xi n} n^k \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n}\\ &= \sum_{k=1}^{\xi n} e^{k\ln(n)(1-(1-k/n)C\varepsilon^2/3)}.\end{aligned}$$

The exponent is monotonically decreasing in $1 - k/n$. Substituting the maximum value of k in $1 - k/n$ yields

$$\begin{aligned}\sum_{k=1}^{\xi n} \binom{n}{k} \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n} &\leq \sum_{k=1}^{\xi n} e^{k\ln(n)(1-(1-k/n)C\varepsilon^2/3)}\\ &\leq \sum_{k=1}^{\xi n} e^{k\ln(n)(1-(1-\xi)C\varepsilon^2/3)}.\end{aligned}$$

This is a geometric series without the term for $k = 0$. This implies that our sum equals

$$\begin{aligned}
\sum_{k=1}^{\xi n} \binom{n}{k} \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n} &\leq \sum_{k=1}^{\xi n} e^{k\ln(n)(1-(1-\xi)C\varepsilon^2/3)} \\
&= \frac{1 - e^{(\xi n+1)\ln(n)(1-(1-\xi)C\varepsilon^2/3)}}{1 - e^{\ln(n)(1-(1-\xi)C\varepsilon^2/3)}} - 1 \\
&= e^{\ln(n)(1-(1-\xi)C\varepsilon^2/3)} \cdot \frac{1 - e^{\xi n\ln(n)(1-(1-\xi)C\varepsilon^2/3)}}{1 - e^{\ln(n)(1-(1-\xi)C\varepsilon^2/3)}} \\
&= n^{(1-(1-\xi)C\varepsilon^2/3)} \cdot \frac{1 - n^{\xi n(1-(1-\xi)C\varepsilon^2/3)}}{1 - n^{(1-(1-\xi)C\varepsilon^2/3)}}.
\end{aligned}$$

By construction we have $1 - (1-\xi)C\varepsilon^2/3 < -2$. This yields

$$\begin{aligned}
\sum_{k=1}^{\xi n} \binom{n}{k} \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n} &\leq n^{(1-(1-\xi)C\varepsilon^2/3)} \cdot \frac{1 - n^{\xi n(1-(1-\xi)C\varepsilon^2/3)}}{1 - n^{(1-(1-\xi)C\varepsilon^2/3)}} \\
&= o\left(\frac{1}{n^2}\right) \cdot O(1) \\
&= o\left(\frac{1}{n^2}\right).
\end{aligned}$$

For the second summation we apply the inequality $\binom{n}{k} \leq 2^n$ to find

$$\begin{aligned}
\sum_{k=\xi n}^{n/2} \binom{n}{k} \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n} &\leq \sum_{k=\xi n}^{n/2} 2^n \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n} \\
&= \sum_{k=\xi n}^{n/2} e^{n\ln(2)-k(n-k)C\ln(n)\varepsilon^2/3n} \\
&= \sum_{k=\xi n}^{n/2} e^{n\ln(2)-\frac{k}{n}(1-\frac{k}{n})Cn\ln(n)\varepsilon^2/3}.
\end{aligned}$$

The exponent is monotonically decreasing in k. Substituting the minimal value for k yields

$$\begin{aligned}
\sum_{k=\xi n}^{n/2} \binom{n}{k} \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n} &\leq \sum_{k=\xi n}^{n/2} e^{n\ln(2)-\frac{k}{n}(1-\frac{k}{n})Cn\ln(n)\varepsilon^2/3} \\
&\leq \sum_{k=\xi n}^{n/2} e^{n\ln(2)-\xi(1-\xi)Cn\ln(n)\varepsilon^2/3} \\
&\leq \frac{n}{2} \cdot e^{n\ln(2)-\xi(1-\xi)Cn\ln(n)\varepsilon^2/3} \\
&\leq e^{n(\ln(2)+1)-\xi(1-\xi)Cn\ln(n)\varepsilon^2/3}.
\end{aligned}$$

The dominating term in the exponent is $n\ln(n)$. By construction we have that $\xi \in (0,1)$ is constant and $1-(1-\xi)C\varepsilon^2/3 < -2$. This implies $-\xi(1-\xi)C\varepsilon^2/3 < -3\xi$. We conclude that

$$\begin{aligned}
\sum_{k=\xi n}^{n/2} \binom{n}{k} \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n} &\leq e^{n(\ln(2)+1)-\xi(1-\xi)Cn\ln(n)\varepsilon^2/3} \\
&\leq e^{n(\ln(2)+1)-3\xi n\ln(n)} \\
&= O\left(e^{-2\xi n\ln(n)}\right) \\
&= O\left(n^{-2\xi n}\right).
\end{aligned}$$

The results for the two individual summations can be used to obtain the desired result:

$$\begin{aligned}
&\mathbb{P}\left(\exists U \subset V \,:\, \Big||c(U)| - p\mu_U\Big| > \varepsilon p\mu_U\right) \\
&\qquad\qquad \leq 2\sum_{k=1}^{\xi n} \binom{n}{k} \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n} + 2\sum_{k=\xi n}^{n/2} \binom{n}{k} \cdot e^{-k(n-k)C\ln(n)\varepsilon^2/3n} \\
&\qquad\qquad = 2\cdot o\left(\frac{1}{n^2}\right) + 2\cdot O\left(\frac{1}{n^{2\xi n}}\right) = o\left(\frac{1}{n^2}\right).
\end{aligned}$$

□

6.2 Performance of heuristics

In this section, we give the main results of this thesis. We use the results from Section 5 and Lemma 6.2 to analyze the performance of several heuristics in random shortest path metrics applied to Erdős–Rényi random graphs. Constant upper bounds for the expected approximation ratios are proven of `greedy` for the minimum distance perfect matching problem, of `nearest neighbor` and `insert` for the traveling salesman problem and of `trivial` for the k-median problem. Additionally, an upper bound for the expected number of iterations of `2-opt` for the traveling salesman problem will be given.

When a graph $G=(V,E)$ is created by the $G(n,p)$ model, there is a non-zero probability of G being disconnected. In a corresponding random shortest path metric this results in $d(u,v)=\infty$ for any two vertices $u,v \in V$ that are in different components of G. This problem does not occur if we limit ourselves to graphs that satisfy the ERRG cut property (Definition 6.1), since all graphs satisfying this property are connected. Lemma 6.2 can then be used to prove that constant expected approximation ratios hold with high probability.

In this thesis however we also take all instances for which the ERRG cut property does not hold into account. In order to take all instances into account, we use the following observation: If a random shortest path metric contains $u,v \in V$ such that $d(u,v)=\infty$, then the identity of indiscernibles, symmetry and triangle inequality still hold. Thus we still have a metric and we can bound the expected approximation ratio for such graphs from above by the worst-case approximation ratio for metric instances.

We use the previous observation to prove a constant upper bound for the approximation ratio of `greedy` for the minimum distance perfect matching problem. For this, let GR denote the cost of the matching computed by `greedy` and let MM denote the value of an optimal matching.

Theorem 6.3. *Let $\varepsilon \in (0,1)$ be constant and $p \geq C\frac{\ln(n)}{n}$, where $C > \frac{9}{\varepsilon^2}$. For this p, let $G=(V,E)$ be an instance of the $G(n,p)$ model and let $M=(V,d)$ be a corresponding random shortest path* *metric. For such random metrics, we have*

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

Proof. Let A be the event that the ERRG cut property holds for G. Then we have

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}}\right) &= \mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}} \;\middle|\; A\right) \cdot \mathbb{P}(A) + \mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}} \;\middle|\; \overline{A}\right) \cdot \mathbb{P}\left(\overline{A}\right) \\ &\le \mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}} \;\middle|\; A\right) + \mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}} \;\middle|\; \overline{A}\right) \cdot \mathbb{P}\left(\overline{A}\right).\end{aligned}$$

Event A implies that G satisfies the α-β-cut-property for $\alpha = (1-\varepsilon)p$ and $\beta = (1+\varepsilon)p$. Theorem 5.5 yields

$$\mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}} \;\middle|\; A\right) = O\left(\frac{(1+\varepsilon)p}{(1-\varepsilon)p}\right) = O(1).$$

If the ERRG cut property does not hold, then we bound the expected approximation ratio of `greedy` from above by its worst-case approximation ratio for metric instances. Since the worst-case approximation ratio of `greedy` for metric instances is known to be $O(n^{\log_2(3/2)})$ [15], where $\log_2(3/2) < 0.59$, we obtain

$$\mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}} \;\middle|\; \overline{A}\right) = O\left(n^{\log_2(3/2)}\right).$$

Since the probability that the ERRG cut property does not hold for G is bounded from above by Lemma 6.2, we obtain

$$\mathbb{P}\left(\overline{A}\right) = o\left(\frac{1}{n^2}\right).$$

Combining these results yields

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}}\right) &\le \mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}} \;\middle|\; A\right) + \mathbb{E}\left(\frac{\mathsf{GR}}{\mathsf{MM}} \;\middle|\; \overline{A}\right) \cdot \mathbb{P}\left(\overline{A}\right) \\ &= O(1) + O\left(n^{\log_2(3/2)}\right) \cdot o\left(\frac{1}{n^2}\right) \\ &= O(1).\end{aligned}$$

□

In the following two theorems, we prove a constant upper bound for the approximation ratio of `nearest neighbor` and `insert` (independent of the insertion rule R) for the traveling salesman problem. For this, let NN denote the cost of the tour created by `nearest neighbor`, let IN_R denote the cost of the tour created by `insert` using insertion rule R and let TSP denote the cost of an optimal tour.

Theorem 6.4. *Let $\varepsilon \in (0,1)$ be constant and $p \ge C\frac{\ln(n)}{n}$, where $C > \frac{9}{\varepsilon^2}$. For this value of p, let $G = (V, E)$ be an instance of the $G(n,p)$ model and let $M = (V, d)$ be a corresponding random shortest path metric. For such random metrics, we have*

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

Proof. Let A be the event that the ERRG cut property holds for G. Then we have

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\right) &= \mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\;\middle|\; A\right)\cdot\mathbb{P}(A) + \mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\;\middle|\;\overline{A}\right)\cdot\mathbb{P}\left(\,\overline{A}\,\right)\\ &\le \mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\;\middle|\; A\right) + \mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\;\middle|\;\overline{A}\right)\cdot\mathbb{P}\left(\,\overline{A}\,\right).\end{aligned}$$

Event A implies that G satisfies the α-β-cut-property for $\alpha = (1-\varepsilon)p$ and $\beta = (1+\varepsilon)p$. Theorem 5.9 yields

$$\mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\;\middle|\; A\right) = O\left(\frac{(1+\varepsilon)p}{(1-\varepsilon)p}\right) = O(1).$$

If the ERRG cut property does not hold, then we bound the expected approximation ratio of `nearest neighbor` from above by its worst-case approximation ratio for metric instances. Since the worst-case approximation ratio of `nearest neighbor` for metric instances is known to be $O(\ln(n))$ [16], we obtain

$$\mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\;\middle|\;\overline{A}\right) = O\left(\ln(n)\right).$$

Since the probability that the ERRG cut property does not hold for G is bounded from above by Lemma 6.2, we obtain

$$\mathbb{P}\left(\,\overline{A}\,\right) = o\left(\frac{1}{n^2}\right).$$

Combining these results yields

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\right) &\le \mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\;\middle|\; A\right) + \mathbb{E}\left(\frac{\mathsf{NN}}{\mathsf{TSP}}\;\middle|\;\overline{A}\right)\cdot\mathbb{P}\left(\,\overline{A}\,\right)\\ &= O(1) + O\left(\ln(n)\right)\cdot o\left(\frac{1}{n^2}\right)\\ &= O(1).\end{aligned}$$

□

Theorem 6.5. *Let $\varepsilon \in (0,1)$ be constant and $p \ge C\frac{\ln(n)}{n}$, where $C > \frac{9}{\varepsilon^2}$. For this value of p, let $G = (V,E)$ be an instance of the $G(n,p)$ model and let $M = (V,d)$ be a corresponding random shortest path metric. For such random metrics, we have for any insertion rule R*

$$\mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\right) = O(1).$$

Proof. Let A be the event that the ERRG cut property holds for G. Then we have

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\right) &= \mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\;\middle|\; A\right)\cdot\mathbb{P}(A) + \mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\;\middle|\;\overline{A}\right)\cdot\mathbb{P}\left(\,\overline{A}\,\right)\\ &\le \mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\;\middle|\; A\right) + \mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\;\middle|\;\overline{A}\right)\cdot\mathbb{P}\left(\,\overline{A}\,\right).\end{aligned}$$

Event A implies that G satisfies the α-β-cut-property for $\alpha = (1-\varepsilon)p$ and $\beta = (1+\varepsilon)p$. Theorem 5.11 yields

$$\mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\;\middle|\; A\right) = O\left(\frac{(1+\varepsilon)p}{(1-\varepsilon)p}\right) = O(1).$$

If the ERRG cut property does not hold, then we bound the expected approximation ratio of `insert` from above by its worst-case approximation ratio for metric instances. Since the worst-case approximation ratio of `insert` for metric instances is known to be $O(\ln(n))$ [16], we obtain

$$\mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}} \,\middle|\, \overline{A}\right) = O\left(\ln(n)\right).$$

Since the probability that the ERRG cut property does not hold for G is bounded from above by Lemma 6.2, we obtain

$$\mathbb{P}\left(\,\overline{A}\,\right) = o\left(\frac{1}{n^2}\right).$$

Combining these results yields

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}}\right) &\le \mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}} \,\middle|\, A\right) + \mathbb{E}\left(\frac{\mathsf{IN}_R}{\mathsf{TSP}} \,\middle|\, \overline{A}\right) \cdot \mathbb{P}\left(\,\overline{A}\,\right) \\ &= O(1) + O\left(\ln(n)\right) \cdot o\left(\frac{1}{n^2}\right) \\ &= O(1).\end{aligned}$$

□

The next theorem is a result for `2-opt` for the traveling salesman problem. In this case no upper bound for the expected approximation ratio is given. Instead, we give an upper bound for the expected number of iterations of `2-opt` if G is connected.

Theorem 6.6. *Let $\varepsilon \in (0,1)$ be constant and $p \ge C\frac{\ln(n)}{n}$, where $C > \frac{9}{\varepsilon^2}$. For this value of p, let $G = (V,E)$ be an instance of the $G(n,p)$ model and let $M = (V,d)$ be a corresponding random shortest path metric. If G is connected, then the expected number of iterations of `2-opt` until a local optimum is found is bounded from above by $O(n^8 \ln^3(n))$.*

Proof. Let T be the number of iterations of `2-opt` needed to find a local optimum, then we want to find

$$\mathbb{E}\left(T \mid G \text{ is connected}\right).$$

Let A be the event that the ERRG cut property holds for G. Then we have

$$\begin{aligned}\mathbb{E}\left(T \mid G \text{ is connected}\right) &= \mathbb{E}\left(T \mid G \text{ is connected, } A\right) \cdot \mathbb{P}(A) + \mathbb{E}\left(T \,\middle|\, G \text{ is connected, } \overline{A}\,\right) \cdot \mathbb{P}\left(\,\overline{A}\,\right) \\ &\le \mathbb{E}\left(T \mid G \text{ is connected, } A\right) + \mathbb{E}\left(T \,\middle|\, G \text{ is connected, } \overline{A}\,\right) \cdot \mathbb{P}\left(\,\overline{A}\,\right)\end{aligned}$$

Event A implies that G satisfies the α-β-cut-property for $\alpha = (1-\varepsilon)p$ and $\beta = (1+\varepsilon)p$. Theorem 5.15 yields

$$\mathbb{E}\left(T \mid G \text{ is connected, } A\right) = O\left(\frac{(1+\varepsilon)p}{(1-\varepsilon)p} n^8 \ln^3(n)\right) = O\left(n^8 \ln^3(n)\right).$$

If the ERRG cut property does not hold and G is connected, then the α-β-cut-property still holds for some values of α and β. The maximal value for α that ensures that every cut has at least size 1, but does not enforce any cut to be larger than this, is $\Theta(1/n^2)$. The minimal value for β that does not enforce constraints on the size of any cut is 1. With these values of α and β, Theorem 5.15 yields

$$\mathbb{E}\left(T \,\middle|\, G \text{ is connected, } \overline{A}\,\right) = O\left(n^{10} \ln^3(n)\right).$$

Since the probability that the ERRG cut property does not hold for G is bounded from above by Lemma 6.2, we obtain

$$\mathbb{P}\left(\,\overline{A}\,\right) = o\left(\frac{1}{n^2}\right).$$

Combining these results yields

$$\begin{aligned}\mathbb{E}\left(T \mid G \text{ is connected}\right) &\leq \mathbb{E}\left(T \mid G \text{ is connected},\, A\right) + \mathbb{E}\left(T \,\middle|\, G \text{ is connected},\, \overline{A}\,\right) \cdot \mathbb{P}\left(\,\overline{A}\,\right) \\ &= O\left(n^8 \ln^3(n)\right) + O\left(n^{10} \ln^3(n)\right) \cdot o\left(\frac{1}{n^2}\right) \\ &= O\left(n^8 \ln^3(n)\right).\end{aligned}$$

□

Our final proof concerns the heuristic `trivial` for the k-median problem. Since the approximation ratio of `trivial` can become unbounded for graphs that are not connected, we only look at connected graphs. For connected graphs we prove an expected approximation ratio of $O(1)$ for large k and $(1+\varepsilon)/(1-\varepsilon)+o(1)$ for sufficiently small k. For this, let TR denote the cost of the solution computed by `trivial` and let ME denote the value of an optimal solution to the k-median problem.

Theorem 6.7. *Let $\varepsilon \in (0,1)$ be a sufficiently small constant and $p \geq C\frac{\ln(n)}{n}$, where $C > \frac{9}{\varepsilon^2}$. For this value of p, let $G = (V,E)$ be an instance of the $G(n,p)$ model and let $M = (V,d)$ be a corresponding random shortest path metric. Let $k \leq (1-\varepsilon)n$ for some constant $\varepsilon > 0$, then we have*

$$\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \textit{ is connected}\right) = O(1).$$

Moreover, if we have $k \leq c'n$ for $c' \in (0,1)$ sufficiently small, then

$$\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \textit{ is connected}\right) = \frac{1+\varepsilon}{1-\varepsilon} + o(1).$$

Proof. Let A be the event that the ERRG cut property holds for G. Then we have

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected}\right) &= \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected},\, A\right) \cdot \mathbb{P}(A) + \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected},\, \overline{A}\,\right) \cdot \mathbb{P}\left(\,\overline{A}\,\right) \\ &\leq \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected},\, A\right) + \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected},\, \overline{A}\,\right) \cdot \mathbb{P}\left(\,\overline{A}\,\right).\end{aligned}$$

Event A implies that G satisfies the α-β-cut-property for $\alpha = (1-\varepsilon)p$ and $\beta = (1+\varepsilon)p$. For $k \leq (1-\varepsilon)n$ for some constant $\varepsilon > 0$, Theorem 5.23 yields

$$\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected},\, A\right) = O\left(\frac{(1-\varepsilon)p}{(1+\varepsilon)p}\right) = O(1).$$

Moreover, if we have $k \leq c'n$ for $c' \in (0,1)$ sufficiently small, Theorem 5.23 yields

$$\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected},\, A\right) = \frac{(1+\varepsilon)p}{(1-\varepsilon)p} \cdot \left(1 + O\left(\frac{\ln\ln\left(\frac{n}{k}\right)}{\ln\left(\frac{n}{k}\right)}\right)\right) = \frac{1+\varepsilon}{1-\varepsilon} + O\left(\frac{\ln\ln\left(\frac{n}{k}\right)}{\ln\left(\frac{n}{k}\right)}\right).$$

If the ERRG cut property does not hold and G is connected, then the α-β-cut-property still holds for some values of α and β. The maximal value for α that ensures that every cut has at least size 1, but does not enforce any cut to be larger than this, is $\Theta(1/n^2)$. The minimal value for β that does not enforce constraints on the size of any cut is 1. With these values for α and β, if $k \leq (1-\varepsilon)n$ for some constant $\varepsilon > 0$, Theorem 5.23 yields

$$\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected}, \overline{A}\right) = O\left(n^2\right).$$

Moreover, if we have $k \leq c'n$ for $c' \in (0,1)$ sufficiently small, Theorem 5.23 yields

$$\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected}, \overline{A}\right) = n^2 \cdot \left(1 + O\left(\frac{\ln\ln\left(\frac{n}{k}\right)}{\ln\left(\frac{n}{k}\right)}\right)\right).$$

Since the probability that the ERRG cut property does not hold for G is bounded from above by Lemma 6.2, we obtain

$$\mathbb{P}\left(\overline{A}\right) = o\left(\frac{1}{n^2}\right).$$

Combining these results yields the desired result:

Let $k \leq (1-\varepsilon)n$ for some constant $\varepsilon > 0$, then we have

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected}\right) &\leq \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected}, A\right) + \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected}, \overline{A}\right) \cdot \mathbb{P}\left(\overline{A}\right)\\ &= O(1) + O\left(n^2\right) \cdot o\left(\frac{1}{n^2}\right) = O(1).\end{aligned}$$

Moreover, if we have $k \leq c'n$ for $c' \in (0,1)$ sufficiently small, then

$$\begin{aligned}\mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected}\right) &\leq \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected}, A\right) + \mathbb{E}\left(\frac{\mathsf{TR}}{\mathsf{ME}} \,\middle|\, G \text{ is connected}, \overline{A}\right) \cdot \mathbb{P}\left(\overline{A}\right)\\ &= \frac{1+\varepsilon}{1-\varepsilon} + O\left(\frac{\ln\ln\left(\frac{n}{k}\right)}{\ln\left(\frac{n}{k}\right)}\right) + n^2 \cdot \left(1 + O\left(\frac{\ln\ln\left(\frac{n}{k}\right)}{\ln\left(\frac{n}{k}\right)}\right)\right) \cdot o\left(\frac{1}{n^2}\right)\\ &= \frac{1+\varepsilon}{1-\varepsilon} + o(1).\end{aligned}$$

□

7 Overview of results

This section gives a brief overview of the results of this thesis. Table 1 summarizes the approximation ratios found for the heuristics `greedy`, `nearest neighbor`, `insert` and `trivial`.

In the second column of this table, we find the worst-case approximation ratios for the heuristics on metric instances. The third column contains known upper bounds for the expected approximation ratios on metric instances where the vertices are positioned independent identically distributed in the euclidean space. As can be seen, we are only aware of such a result for `greedy` [2].

The fourth column contains our results on random shortest path metrics applied to graphs satisfying the α-β-cut-property (Definition 4.2). The upper bounds we found are good when the graph is nearly regular, i.e. when $\beta = O(\alpha)$. In this case, the upper bounds for the expected approximation ratios are constant. This also corresponds to the results from Bringmann et al. [4], which we obtain by setting $\alpha = \beta = 1$. However, when the graph is not very regular, the value of $O(\beta/\alpha)$ can become at most $O(n^2)$, because the smallest value for α is $\Theta(1/n^2)$ and the largest possible value for β is 1. In this case the found upper bounds are useless for the first three heuristics, since they are worse than the worst-case scenarios.

The final column contains our results for random shortest path metrics applied to Erdős–Rényi random graphs. According to Lemma 6.2, these graphs are nearly regular with very high probability. Because of this, we were able to proof constant upper bounds for the expected approximation ratios of the heuristics, which shows the performance of these heuristics is much better than the worst-case makes us assume.

Algorithm	**Worst Case**	**Eucl.**	**RSPM with α-β-cut-property**	**RSPM on ERRG ($G(n,p)$)**†
Greedy Matching	$O\left(n^{\log_2(3/2)}\right)$	$O(1)$	$O\left(\frac{\beta}{\alpha}\right)$ (Th. 5.5)	$O(1)$ (Th. 6.3)
N.N. for TSP	$O(\log(n))$	-	$O\left(\frac{\beta}{\alpha}\right)$ (Th. 5.9)	$O(1)$ (Th. 6.4)
Insertion for TSP	$O(\log(n))$	-	$O\left(\frac{\beta}{\alpha}\right)$ (Th. 5.11)	$O(1)$ (Th. 6.5)
k-Median (k large)	∞	-	$O\left(\frac{\beta}{\alpha}\right)$ (Th. 5.23)	$O(1)$ (Th. 6.7)††
k-Median (k suff. small)	∞	-	$\frac{\beta}{\alpha}(1+o(1))$ (Th. 5.23)	$\frac{1+\varepsilon}{1-\varepsilon}+o(1)$ (Th. 6.7)††

Table **1**: Overview of approximation ratios

Our final result concerns the heuristic `2-opt` for the traveling salesman problem. For this heuristic, no upper bound for the expected approximation is given. Instead, we have proven an upper bound for the expected number of iterations of `2-opt`. In random shortest path metrics applied to connected graphs that satisfy the α-β-cut-property (Definition 4.2) this becomes $O(n^8 \ln^3(n)\beta/\alpha)$. For random shortest path metrics applied to connected instances of the ERRG model, this upper bound becomes $O(n^8 \ln^3(n))$.

†Results in this column hold for constant $\varepsilon \in (0,1)$ and $p \geq C\frac{\ln(n)}{n}$, where $C > \frac{9}{\varepsilon^2}$.
††Result only holds for connected graphs.

8 Discussion and remarks

This thesis provides a probabilistic analysis of heuristics for several optimization problems in random shortest path metrics applied to Erdős–Rényi random graphs. We have analyzed the expected approximation ratios of `greedy` for the minimum distance perfect matching problem, of `nearest neighbor` and `insert` for the traveling salesman problem and of `trivial` for the k-median problem. We have shown that the expected approximation ratio of these heuristics is constant in random shortest path metrics applied to Erdős–Rényi random graphs. Furthermore, for these metrics we have found a polynomial bound for the expected number of iterations of `2-opt` for the traveling salesman problem.

The results of this thesis extend the probabilistic analysis of heuristics in random shortest path metrics applied to complete graphs (Bringmann et al. [4]) to the probabilistic analysis of heuristics in random shortest path metrics applied to Erdős–Rényi random graphs. The found upper bounds for the expected approximation ratios of these heuristics show that the heuristics are expected to perform well in practice, even though they have a poor worst-case approximation ratio.

A few open questions remain. First of all, this thesis focused on analyzing random shortest path metrics applied to Erdős–Rényi random graphs. Of course many other models exist to create random graphs. This raises the question if it is also possible to achieve similar results for random shortest path metrics applied to other random graphs. Is it for instance possible for other random graphs to find values for α and β such that $\beta = O(\alpha)$ and the α-β-cut-property holds with high probability? Or can we find some other method to prove good approximation ratios?

Secondly, for most of the heuristics we have found a $O(1)$ approximation ratio. Since no lower bound is yet known for these expected approximation ratios, we wonder if it is possible to improve these ratios to $1 + o(1)$. Similarly, since we did not prove an upper bound for the expected approximation ratio of `2-opt`, we wonder what bounds can be proven for this heuristic.

Furthermore, we created our random shortest path metrics by giving every edge in our graph an independent weight from Exp(1). It is interesting to see if our results also carry over to other distributions. The general idea for this is to use a coupling argument given by Janson [9]. He has shown that his results for the distance between two fixed vertices and the diameter in random shortest path metrics applied to a complete graph carry over as long as the probability distribution for the edge weights has the following property: the probability that an edge weight in the graph is smaller than x is $x + o(x)$. Both the exponential distribution with parameter 1 and the uniform distribution on $[0, 1]$ satisfy this condition. In the proof of this coupling argument, Janson exploits the fact that, when drawing edge weights from Exp(1), the expected distance between two fixed vertices in random shortest path metrics applied to complete graphs is $o(1)$. For random shortest path metrics applied to Erdős–Rényi random graphs this holds if $\alpha = \omega(\ln(n)/n)$ (consequence of Corollary 4.6). Therefore, we believe our results carry over to general distributions with probability distribution satisfying $F(x) = x + o(x)$ when $\alpha = \omega(\ln(n)/n)$. This claim should however be analyzed in more detail.

We can take the last remark a step further and wonder what happens if the edge weights are not all drawn from the same probability distribution, but from several probability distributions. This would make the analysis of the metric instances harder. However, we believe our reasoning for the coupling argument still holds true as long as all possible probability distributions for the edge weights satisfy $F(x) = x + o(x)$. It would be interesting to research the case where the probability distributions do not satisfy this property, but we believe this is quite challenging.

Acknowledgements

I would like to thank my supervisors Bodo Manthey and Stefan Klootwijk for introducing the topic of this thesis to me. I enjoyed our discussions and am grateful for their helpful feedback and for guiding me from start to end through this final project.

References

[1] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local search heuristics for k-median and facility location problems. *SIAM Journal on Computing*, 33(3): 544–562, 2004. doi:10.1137/S0097539702416402.

[2] D. Avis, B. Davis, and J. M. Steele. Probabilistic analysis of a greedy heuristic for euclidean matching. *Probability in the Engineering and Informational Sciences*, 2(2):143156, 1988. doi:10.1017/S0269964800000711.

[3] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. First passage percolation on the Erdős-Rényi Random Graph. *Combinatorics, Probability and Computing*, 20:683–707, 2011. doi:10.1017/S096354831100023X.

[4] K. Bringmann, C. Engels, B. Manthey, and B. V. R. Rao. Random shortest paths: Non-euclidean instances for metric optimization problems. *Algorithmica*, 73:42–62, 2015. doi:10.1007/s00453-014-9901-9.

[5] M. Englert, H. Röglin, and B. Vöcking. Worst case and probabilistic analysis of the 2-opt algorithm for the TSP. *Algorithmica*, 68:190–264, 2014. doi:10.1007/s00453-013-9801-4.

[6] J. M. Hammersley and D. J. A. Welsh. First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. In J. Neyman and L. M. Le Cam, editors, *Bernoulli 1713 Bayes 1763 Laplace 1813: Anniversary Volume Proceedings of an International Research Seminar Statistical Laboratory University of California, Berkeley 1963*, pages 61–110. Springer Berlin Heidelberg, 1965. ISBN 978-3-642-49750-6. doi:10.1007/978-3-642-49750-6_7.

[7] R. Hassin and E. Zemel. On shortest paths in graphs with random weights. *Mathematics of Operations Research*, 10(4):557–564, 1985. doi:10.1287/moor.10.4.557.

[8] C. D. Howard. Models of first-passage percolation. In H. Kesten, editor, *Probability on Discrete Structures*, pages 125–173. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. ISBN 978-3-662-09444-0. doi:10.1007/978-3-662-09444-0_3.

[9] S. Janson. One, two and three times log n/n for paths in a complete graph with random weights. *Combinatorics, Probability and Computing*, 8(4):347–361, 1999. doi:10.1017/S0963548399003892.

[10] S. Janson. Tail bounds for sums of geometric and exponential variables. *Statistics & Probability Letters*, 135:1–6, 2018. doi:10.1016/j.spl.2017.11.017.

[11] R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller, J. W. Thatcher, and J. D. Bohlinger, editors, *Complexity of Computer Computations: Proceedings of a symposium on the Complexity of Computer Computations, held March 20–22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, and sponsored by the Office of Naval Research, Mathematics Program, IBM World Trade Corporation, and the IBM Research Mathematical Sciences Department*, pages 85–103. Springer US, 1972. ISBN 978-1-4684-2001-2. doi:10.1007/978-1-4684-2001-2_9.

[12] N. Megiddo and K. J. Supowit. On the complexity of some common geometric location problems. *SIAM Journal on Computing*, 13(1):182–196, 1984. doi:10.1137/0213014.

[13] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005. ISBN 0-521-83540-2.

[14] H. N. Nagaraja. Order statistics from independent exponential random variables and the sum of the top order statistics. In N. Balakrishnan, J. M. Sarabia, and E. Castillo, editors, *Advances in Distribution Theory, Order Statistics, and Inference*, pages 173–185. Birkhäuser Boston, Boston, MA, 2006. ISBN 978-0-8176-4487-1. doi:10.1007/0-8176-4487-3_11.

[15] E. M. Reingold and R. E. Tarjan. On a greedy heuristic for complete matching. *SIAM Journal on Computing*, 10(4):676–681, 1981. doi:10.1137/0210050.

[16] D. J. Rosenkrantz, R. E. Stearns, and P. M. Lewis. An analysis of several heuristics for the traveling salesman problem. In S. S. Ravi and S. K. Shukla, editors, *Fundamental Problems in Computing: Essays in Honor of Professor Daniel J. Rosenkrantz*, pages 45–69. Springer Netherlands, Dordrecht, 2009. ISBN 978-1-4020-9688-4. doi:10.1007/978-1-4020-9688-4_3.

[17] S. M. Ross. *Introduction to Probability Models.* Elsevier, 10th edition, 2010. ISBN 78-0-12-375686-2.