Probabilistic Analysis of Optimization Problems on Generalized Random Shortest Path Metrics

Stefan Klootwijk, Bodo Manthey, Sander K. Visser

University of Twente, Department of Applied Mathematics, Enschede, The Netherlands

Abstract

A graph $G = (V, E)$ satisfies the α, β-cut-property if the fraction of edges present in each cut of the graph lies between α and β. The Erdős-Rényi random graph $G(n, p)$ satisfies this property w.h.p. for $\alpha = (1 - \varepsilon)p$ and $\beta = (1 + \varepsilon)p$ whenever p is sufficiently large and ε is a suitably chosen constant.

We study the behavior of random shortest path metrics applied to graphs G that satisfy the α, β-cut-property. These random metrics are defined as follows: Let $w(e)$ be independently drawn random edge weights for all $e \in E$, and define $d(u, v)$ to be the shortest path distance between u and v in G with respect to the weights w.

Using the ideas of Bringmann et al. (*Algorithmica, 2015*), who studied random shortest path metrics on the complete graph, i.e., the graph that satisfies the $1, 1$-cut-property, we derive some properties of the metric and obtain a clustering of the vertices. Using this, we conduct a probabilistic analysis of some simple heuristics on these random shortest path metrics.

Keywords : *Random shortest paths, Random metrics, Approximation algorithms, Erdős-Rényi random graph*

1 Introduction

Large-scale optimization problems, such as the traveling salesman problem, show up in many applications and domains all around us. Often it is not possible to solve those problems exactly, since that would take too much time. In practice often ad-hoc heuristics are being used that provide solutions that come quite close to the optimal solutions. In many cases those, often simple, heuristics show a remarkable performance, even though the theoretical results about those heuristics are way more pessimistic.

In order to explain this difference, the method of probabilistic analysis has been widely used over the last decades. In this method, the performance of the heuristics is analysed with respect to a random instance. However, it is not trivial to come up with a good model for such random instances. So far, in almost all cases either Euclidean space has been used to generate instances, or independent, identically distributed edge lengths were used.

However, both approaches fall short of explaining the average-case performance of heuristics on general metric instances. In order to overcome this, Bringmann et al. [2] used the following model for generating random metric spaces, which had been proposed by Karp and Steele [6]. Given an undirected complete graph, start by drawing random edge weights for each edge independently and then define the distance between any two vertices as the total weight of the shortest path between them, measured with respect to the random weights.

Bringmann et al. called this model *random shortest path metrics*. This model is also known as *first-passage percolation*, introduced by Hammersley and Welsh as a model for fluid flow through a (random) porous medium [4, 5].

Our goal is to further broaden the knowledge of metric spaces that can be used to generate random instances, with special interest to the probabilistic analysis of algorithms on those random metric spaces. By making extensive use of the ideas of Bringmann et al. [2] we will extend their results to random shortest path metrics on a class of graphs that naturally arises when considering the Erdős-Rényi random graph $G(n, p)$. To be more precise, we consider arbitrary (fixed) graphs that satisfy what we called the α, β-cut-property.

Definition 1 *Let $0 < \alpha \leq \beta \leq 1$. A finite simple graph $G = (V, E)$ satisfies the α, β-cut-property if for all $\varnothing \subset U \subset V$ it holds that $\alpha\mu_U \leq |\delta(U)| \leq \beta\mu_U$, where $\mu_U := |U|(|V| - |U|)$ is the maximum number of possible edges in the cut defined by U.*

Note that the complete graph satisfies the α, β-cut-property for $\beta = 1$ and any value of $\alpha \leq 1$, so in particular for $\alpha = \beta = 1$. Moreover, for sufficiently large p and suitably chosen constant ε, it follows that the Erdős-Rényi random graph $G(n, p)$ w.h.p. satisfies the α, β-cut-property for $\alpha = (1 - \varepsilon)p$ and $\beta = (1 + \varepsilon)p$. So, the Erdős-Rényi random graph $G(n, p)$ satisfies the α, β-cut-property w.h.p. for α, β with $\beta/\alpha = O(1)$.

2 Model and Clustering

Let $G = (V, E)$ be a graph on n vertices that satisfies the α, β-cut-property. Let $w : E \to \mathbb{R}_{\geq 0}$ denote the random weights of the edges, independently drawn from the standard exponential distribution. And let $d : V \times V \to \mathbb{R}_{\geq 0}$ denote the shortest path distances with respect to G and w. Note that d defines a metric on V. Also observe that having $\alpha > 0$ ensures that G is connected; this allows us to define d as we did, without having to be careful in case G is not connected and some shortest paths do not exist.

Using of the ideas of Bringmann et al. [2] we find a clustering of the vertices in generalized random shortest path metrics. This clustering is a very helpful tool for the probabilistic analysis of the simple heuristics in the next section.

Lemma 1 *Consider a random shortest path metric on a graph G that satisfies the α, β-cut-property and let $\Delta \geq 0$. The expected number of clusters needed to partition the instance into clusters, each of diameter at most 6Δ, is $O(1 + n/\exp(\beta\Delta n/5))$.*

3 Results

Greedy matching. The greedy algorithm for the minimum-weight perfect matching problem[1] iteratively adds the edge between the two closest unmatched vertices to the matching. Reingold and Tarjan showed that the worst-case approximation ratio for greedy matching on metric instances is $\Theta(n^{\log_2(3/2)})$ [7]. We show that the greedy matching outputs a matching with expected costs at most $O(\beta/\alpha)$. On the other hand, the optimum matching has a total length of $\Omega(1)$. So, we obtain an expected approximation ratio of $O(\beta/\alpha)$.

Theorem 1 *The greedy algorithm for minimum-length perfect matching has an expected approximation ratio of $O(\beta/\alpha)$ on generalized random shortest path metrics.*

The main idea of the proof is to divide the algorithm into several phases defined by the length of the edges that are added to the matching. For the first phases we can bound the maximum number of edges hat is being added in that phase by making use of the clustering property, and we can show that the total expected contribution of the last phases is negligible, in order to obtain the result.

Nearest-neighbor algorithm for TSP. The nearest-neighbor algorithm for the traveling salesman problem starts building a tour from an arbitrary vertex, and then iteratively adds the edge that connects the last added vertex to its closest unvisited vertex (and finally closes the tour by adding the edge that connects the last vertex to the first vertex). Ausiello et

[1] Within our terminology, minimum-distance perfect matching might be a better name for this problem.

al. showed that the worst-case approximation ratio for nearest-neighbor on metric instances is $O(\log(n))$ [1]. We show that nearest-neighbor outputs a tour of length at most $O(\beta/\alpha)$ in expectation. On the other hand, the optimum tour has a total length of $\Omega(1)$. So, we obtain an expected approximation ratio of $O(\beta/\alpha)$.

Theorem 2 *The nearest-neighbor algorithm for the traveling salesman problem has an expected approximation ratio of $O(\beta/\alpha)$ on generalized random shortest path metrics.*

The main idea of the proof is to partition the used edges into several groups defined by the length of the edges. For the groups with the shortest edges we can bound the maximum number of edges in those groups by making use of the clustering property, and we can show that the total expected contribution of the groups with the longest edges is negligible, in order to obtain the result.

Insertion heuristics for TSP. An insertion heuristic for the traveling salesman problem starts building a tour from a small initial tour on a few vertices and then iteratively adds the other vertices to this tour according to some rule. Rosenkrantz et al. showed that the worst-case approximation ratio for any insertion heuristic on metric instances is $O(\log(n))$ [8]. We show that all insertion heuristics output a tour of length at most $O(\beta/\alpha)$ in expectation. On the other hand, the optimum tour still has a total length of $\Omega(1)$. So, we obtain an expected approximation of $O(\beta/\alpha)$.

Theorem 3 *Any insertion heuristic heuristic for the traveling salesman problem has an expected approximation ratio of $O(\beta/\alpha)$ on generalized random shortest path metrics.*

The main idea of the proof is to show that the initial tour has small expected costs and then to partition the partition the added vertices into several groups defined by the additional length added to the tour when the vertex was added to it. We bound the number of vertices in each group, in order to obtain the result.

Running time of 2-opt for TSP. The 2-opt heuristic for the traveling salesman problem starts with an initial tour and then improves this tour by so-called 2-exchanges until no further improvement is possible. A 2-exchange consists of selecting two edges $\{u, v\}$ and $\{x, y\}$ from the tour, removing those edges from the tour and replacing them by the edges $\{u, x\}$ and $\{v, y\}$ to obtain a shorter tour. Even on metric instances, the number of iterations that 2-opt needs to terminate can be $\Omega(2^n)$, as was shown by Englert et al. [3]. We show that the expected number of iterations of 2-opt is polynomially bounded for random shortest path metrics on graphs with the α, β-cut-property.

Theorem 4 *The expected number of iteration that 2-opt needs to find a local optimum is bounded by $O(n^{10}\ln^3(n))$ on generalized random shortest path metrics.*

The main idea of the proof is to use tail bounds for the maximum length of the initial tour and for the minimum improvement made by any 2-exchange. The number of iterations of 2-opt is bounded by the ratio of these values.

***k*-Median.** The k-median problem is to find a subset $U \subseteq V$ with $|U| = k$ such that $\sum_{v\in V} \min_{u\in U} d(u, v)$ is minimized. We consider the (trivial) heuristic that selects the k vertices of U randomly, independent of the metric space, e.g., $U = \{v_1, \dots, v_k\}$. We call the solution that this heuristic outputs the *trivial solution* to the k-median problem. In general this trivial solution can be arbitrarily bad, even if we only consider metric instances. However, we show that this trivial heuristic yields an $O(\beta/\alpha)$ approximation in expectation whenever k is not too large.

Theorem 5 *If $k \le (1-\varepsilon)n$ for some constant $\varepsilon > 0$, then the trivial solution to the k-median problem has an expected approximation ratio of $O(\beta/\alpha)$ on generalized random shortest path metrics. Moreover, for sufficiently small k this expected approximation ratio is given by $(\beta/\alpha)(1 + O(\ln\ln(n/k)/\ln(n/k)))$.*

References

[1] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi. *Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties*. Springer-Verlag Berlin Heidelberg, 1999. ISBN 978-3-642-63581-6. doi: 10.1007/978-3-642-58412-1.

[2] K. Bringmann, C. Engels, B. Manthey, and B.V.R. Rao. Random shortest paths: Non-euclidean instances for metric optimization problems. *Algorithmica*, 73(1):42–62, 2015. doi: 10.1007/s00453-014-9901-9.

[3] M. Englert, H. Röglin, and B. Vöcking. Worst case and probabilistic analysis of the 2-opt algorithm for the TSP. *Algorithmica*, 68(1):190–264, 2014. doi: 10.1007/s00453-013-9801-4.

[4] 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. doi: 10.1007/978-3-642-49750-6_7.

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

[6] R.M. Karp and J.M. Steele. Probabilistic analysis of heuristics. In *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, chapter 6, pages 181–205. John Wiley & Sons Ltd., 1985. ISBN 978-0-471-90413-7.

[7] 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.

[8] D.J. Rosenkrantz, R.E. Stearns, and P.M. Lewis II. An analysis of several heuristics for the traveling salesman problem. *SIAM Journal on Computing*, 6(3):563–581, 1977. doi: 10.1137/0206041.