Thai Journal of Mathematics
Volume 18 Number 4 (2020)
Pages 1881–1894

http://thaijmath.in.cmu.ac.th

ISSN 1686-0209

Worst Case Analysis of Nearest Neighbour Algorithms for the Minimum Weighted Directed k-Cycle Problem

Piyashat Sripratak

Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
e-mail : psripratak@gmail.com

Abstract Given a weighted complete directed graph on n nodes, the asymmetric traveling salesman problem (ATSP) is to find a minimum weighted directed cycle of length n. This is a well-studied NP-hard problem. Sometimes, we require a cycle containing a specific number of nodes. Thus, we concentrate on finding a minimum weighted directed cycle of length k when k is a positive integer in $\{2, 3, \ldots, n\}$. The problem is called the minimum weighted directed k-cycle problem (MWDkCP), a generalization of ATSP. Nearest neighbour algorithm (NN) and repetitive nearest neighbour algorithm (RNN) for ATSP are known for good computational results on Euclidean ATSP, but poor performances on some graphs. We give instances to show that establishing an approximation ratio for NN is impossible, and a result from NN can be worse than average. We also prove that NN can output a unique maximum weighted directed k-cycle, and offer a sufficient condition to avoid that scenario. As for RNN, when $n \neq k$, it has no approximation ratio and can be worse than average. When $n \geq 4$, we obtain a lower bound and an upper bound for the domination number of RNN.

MSC: 90C27; 90C35; 90C59

Keywords: minimum weight k-cycle; worst case analysis; greedy heuristic

Submission date: 05.05.2020 / Acceptance date: 03.11.2020

1. Introduction

For positive integers n and k where $2 \leq k \leq n$, the *minimum weighted directed k-cycle problem* (MWDkCP) is to find a minimum weighted directed cycle of length k in a weighted complete directed graph. This problem bears a similarity to a famous NP-hard problem, the *asymmetric traveling salesman problem* (ATSP) - if a salesman wants to visit n houses, find the shortest route that allows him to visit all of these houses without passing through any houses twice and go back to the starting place. In the graph theoretical point of view, ATSP is to find a minimum cycle containing all nodes. Thus, it is a special case of the MWDkCP when $k = n$. By the NP-hardness of the TSP [1], the MWDkCP is NP-hard.

Published by The Mathematical Association of Thailand.

Due to the close relationship of these two problems, it is reasonable to construct heuristics for the MWDkCP by modifying those for ATSP. The most simple ones are greedy-type heuristics for ATSP. Gutin et. al [2] present the domination analyses for some of them, and it turns out that they are not good in term of domination number. A basic approach is to keep adding the cheapest available arcs unless the selected arc forms a cycle of length less than k or makes the indegree or outdegree of a chosen node greater than 1. Unlike ATSP, once $k-1$ arcs have been chosen, no one can guarantee that adding the k^{th} arc would give a directed k-cycle. To establish a greedy heuristic for the MWDkCP from this procedure, more modification is required. Another type of these greedy heuristics is *Nearest neighbour algorithm* (NN). For any nodes u and v in K_n, we use $w(u,v)$ to denote the weight of the arc from u to v. The process begins with choosing a node u_1 in K_n, adding the arc from u_1 to a nearest node u_2, that is $w(u_1,u_2) \leq w(u_1,u_j)$ for all $j \in \{2,\ldots,n\}$. Repeat this process until $k-1$ arcs are selected. Then add the arc (u_k,u_1) to form a k-cycle. *Repetitive nearest neighbour algorithm* (RNN) is to apply NN starting from every node and output the minimum one among the cycles constructed in each NN performance. The complexity of NN is $O(n^2)$ while that of RNN is $O(n^3)$, so both of them are polynomial-time algorithms. In this study, we concentrate on analyzing these two heuristics.

There are abundant computational results of NN and RNN for ATSP [3–5]. Even they work well on graphs satisfying the triangle inequality, the performance in general are significantly worse. Therefore, it is interesting to analyze these heuristics from a theoretical point of view. Approximation ratio is the most common method for analyzing the performance of a heuristic. However, for many combinatorial optimization problems, establishing a heuristic with a constant approximation ratio is an NP-hard problem [2]. Thus, we also consider other approaches on worst case analysis. We compare whether the objective function values of the outputs from our heuristics are worse than the average value of the objective function values of all solutions, another metric for worst cases analysis [6–9]. Besides, we study the domination numbers [10–12] and the domination ratio [10, 13], which consider the number of solutions that are not better than the results from the heuristics.

For any directed graph G, let $N(G)$ denote the node set of G and $A(G)$ denote the arc set of G. We use I_n to denote an instance of MWDkCP which is the complete directed graphs of size n with a *weight function* w from $A(I_n)$ to $\mathbb{R}$. For any instance I_n with a weight function w, we use the notation $C(I_n)$ for a k-cycle C in I_n. For any subgraph H of I_n, $w(H)$ is the total weight of all arcs in H. Note that for any instance I_n, we can construct a new weight function $w'(a) = w(a) - \min\{w(a) : a \in A(I_n)\} + 1$ for all $a \in A(I_n)$. We can see that $w'(a) > 0$ for all $a \in A(I_n)$ and for any k-cycles C^A and C^B in I_n, $w(C^A) \leq w(C^B)$ if and only if $w'(C^A) \leq w(C^B)$. Therefore, we can consider only instances whose weights for all arcs are positive. We use $\mathscr{K}_n$ to denote the collection of all complete directed graphs of size n whose weights for all arcs are positive.

Let $\mathscr{H}$ be a heuristic for MWDkCP. For $I_n \in \mathscr{K}_n$, let $C^M(I_n)$ denote a minimum k-cycle in I_n and $C^{\mathscr{H}}(I_n)$ a k-cycle with maximum weight among all k-cycles obtained by applying heuristic $\mathscr{H}$ on I_n. $\mathscr{H}$ has an *approximation ratio* of ϵ if for any integer $n \geq 2$ and instance $I_n \in \mathscr{K}_n$, $\frac{w(C^{\mathscr{H}}(I_n))}{w(C^M(I_n))} \leq \epsilon$.

Let $\mathscr{A}(I_n)$ be the average value of the weights of all k-cycles in I_n. $\mathscr{H}$ is said to be *not worse than average* if for all integer $n \geq 2$ and all instance I_n, $w(C^{\mathscr{H}}(I_n)) \leq \mathscr{A}(I_n)$. Otherwise, $\mathscr{H}$ is *worse than average*.

We use $d(I_n)$ to denote the number of k-cycles with weight at least $w(C^{\mathscr{H}}(I_n))$. The *domination number* of $\mathscr{H}$ is the maximum $d(n) \in \mathbb{Z}^+$ such that for every $I_n \in \mathscr{K}_n$, $d(I_n) \geq d(n)$. If C^A and C^B are two k-cycles in I_n, we says that C^A *dominates* C^B or C^A *is not worse than* C^B if $w(C^A) \leq w(C^B)$. In other words, we can say that $d(n)$ is the domination number of $\mathscr{H}$ if $d(n)$ is the maximum number such that for each $I_n \in \mathscr{K}_n$, any heuristic solutions obtained from $\mathscr{H}$ dominates at least $d(n)$ k-cycles in I_n. Let $c(I_n)$ denote the number of all k-cycles in I_n. The *domination ratio* of $\mathscr{H}$ is the maximum $dr(n) \in \mathbb{Z}^+$ such that for any instance $I_n \in \mathscr{K}_n$, $\frac{d(I_n)}{c(I_n)} \geq dr(n)$. Note that by definition, $d(n) \geq 1$ since in any instance I_n, a heuristic solution of the maximum weight is a k-cycle whose weight is at least $w(C^{\mathscr{H}}(I_n))$. On the other hand, $dr(n) \leq 1$ and the equality holds when $d(I_n) = c(I_n)$ for all instance I_n, which implies that $\mathscr{H}$ always gives an optimal solution.

Throughout this paper, we assume that the node set of a graph of size n is $\{1, 2, \ldots, n\}$. When $n = 2$, K_2 contains exactly one 2-cycle which is always a minimum 2-cycle in the graph. Hence, we only consider K_n with $n \geq 3$. Many of our constructed instances contain arcs of weight 0 or ϵ. We call an arc of weight 0 a *zero-arc* and an arc of weight ϵ an *ϵ-arc*. We represent a *directed walk* which is composed of the arcs $(u_1, u_2), (u_2, u_3), \ldots, (u_{w-1}, u_w)$ as $W = (u_1, u_2, \ldots, u_w)$. If $u_1, u_2, \ldots, u_w$ are all different, W is a directed path. If $u_1, u_2, \ldots, u_{w-1}$ are all different and $u_w = u_1$, W is a directed cycle.

2. Results

Our main interests in this work are the worst case analyses of two nearest neighbour heuristics, namely NN and RNN. For each type of nearest neighbour heuristic, we analyze the algorithm in three aspects, which are considering approximation ratio, average-based analysis and domination analysis.

2.1. Nearest Neighbour Algorithms (NN)

In this section, we show that giving an approximation ratio for NN is impossible, and the domination number of NN for the MWDkCP is 1, that is NN outputs the unique maximum weighted directed k-cycles. It follows that NN can be worse than average. Lastly, we propose a sufficient condition so that performing NN on the instances satisfying the condition does not result in the unique maximum k-cycles.

We begins with showing that, for any $n \geq 3$, NN is arbitrarily bad, that is, for any $N > 1$, there exists an instance I_n such that $w(C^{NN}(I_n))$ is greater than N times of $w(C^M(I_n))$.

Theorem 2.1. *Let $n, k \in \mathbb{Z}^+$ where $n \geq 3$, $2 \leq k \leq n$ and $N > 1$. There exists $I_n \in \mathscr{K}_n$ such that $\frac{w(C^{NN}(I_n))}{w(C^M(I_n))} > N$.*

Proof. For each $N > 1$, we consider the instance I_n where the weight for each arc (i, j) is defined as follows:

$$w(i,j) = \begin{cases} \epsilon & ; 1 \leq i \leq k-1, j = i+1 \\ 4N & ; i = k, j = 1 \\ 1 & ; \text{otherwise}, \end{cases}$$

where $0 < \epsilon < \frac{1}{k}$. Note that $\epsilon < k\epsilon < 1$. We use C^H to denote the output from applying NN on I_n starting at node 1. Then $C^H = (1, 2, \ldots, k, 1)$ with weight $w(C^H) = 4N + (k-1)\epsilon$.

Since the weight of any arc other than $(k, 1)$ is ϵ or 1, $w(C^M) = (k-a)\epsilon + a$ for some $a \in \mathbb{Z}$ where $0 \leq a \leq 4$. Note that since $\epsilon < 1$, when $a < b$, we have $(k-a)\epsilon + a < (k-b)\epsilon + b$. Since there are $k-1$ ϵ-arcs in I_n, there are no k-cycles of weight $k\epsilon$ in I_n. Suppose C^M has weight $(k-1)\epsilon + 1$. C^M must contain all $k-1$ ϵ-arcs and another arc of weight 1. However, all ϵ-arcs are $(1, 2), (2, 3), \ldots, (k-1, k)$ and the only k-cycle in I_n that contains all of them is C^H with weight $w(C^H) = 4N + (k-1)\epsilon > (k-1)\epsilon + 1$, a contradiction. At this point, we have $w(C^M) = (k-a)\epsilon + a$ for some integer $2 \leq a \leq 4$.

Case 1 $k < n$. If $k = 2$, the minimum k-cycle in I_n is composed of 2 arcs of weight 1, so $w(C^M) = 2 = (k-2)\epsilon + 2$. For $k \geq 3$, we examine the cycle $(2, 3, \ldots, k, k+1, 2)$. There are two arcs of weight 1 which are $(k, k+1)$ and $(k+1, 2)$, and the others are ϵ-arcs. Hence, the cycle has weight $(k-2)\epsilon + 2$ and is a minimum k-cycle in I_n.

Case 2 $k = n$. First, suppose that $w(C^M) = (k-2)\epsilon + 2$. The arc set of C^M must contain $k-2$ ϵ-arcs and other 2 arcs of weight 1. We consider 3 subcases depending on the ϵ-arcs included in C^M.

Case 2.1 The ϵ-arc $(1, 2)$ is not included. Then the arcs $(2, 3), (3, 4), \ldots, (n-1, n)$ form a directed path from node 2 to node n. The only remaining node is node 1. The only way to obtain all nodes in the graph is adding the arcs $(n, 1)$ and $(1, 2)$, a contradiction.

Case 2.2 The ϵ-arc $(n-1, n)$ is not included. Similar to Case 2.1, to get a Hamiltonian cycle, we are forced to include the arc $(n-1, n)$, leading to a contradiction.

Case 2.3 An ϵ-arc $(i, i+1)$, for some $2 \leq i \leq n-2$, is not included. Note that this case occurs only when $k = n \geq 4$. We get two subpaths in C^M, namely, $(1, 2, \ldots, i)$ and $(i+1, i+2, \ldots, n)$. They contain all nodes in I_n, and the arc $(i, i+1)$ must be in C^M, a contradiction.

So we can see that when $k = n$, $w(C^M) \geq (k-3)\epsilon + 3$. Consider $(2, 3, \ldots, k-1, 1, k, 2)$. It is composed of $k-3$ ϵ-arcs, namely, $(2, 3), (3, 4), \ldots, (k-2, k-1)$ and other 3 arcs of weight 1, namely, $(k-1, 1), (1, k)$ and $(k, 2)$. Thus, this is a k-cycle of weight $(k-3)\epsilon + 3$, and hence, a minimum k-cycle. Therefore, $w(C^H) = 4N + (k-1)\epsilon$ and $w(C^M)$ is $(k-2)\epsilon + 2$ when $k < n$ and is $(k-3)\epsilon + 3$ when $k = n$. Since $\epsilon < k\epsilon < 1$, we get $w(C^H) = 4N + (k-1)\epsilon > 4N$ and $w(C^M) \leq (k-3)\epsilon + 3 < 4$. Hence, $\frac{w(C^{NN}(I_n))}{w(C^M(I_n))} \geq \frac{w(C^H)}{w(C^M)} > \frac{4N}{4} = N$. ∎

According to Theorem 2.1, given a constant ϵ, we can establish an instance I_n where $\frac{w(C^{NN}(I_n))}{w(C^M(I_n))} > \epsilon$. It leads to the fact that an approximation ratio of NN for the MWDkCP does not exist.

We next perform the domination analysis for NN applied on MWDkCP. For any $n \geq 3$ and $2 \leq k \leq n$, we construct an instance of n nodes where NN produces the maximum k-cycle when a specific node is chosen to be the starting node. It implies that the domination number of NN for the MWDkCP is 1.

Theorem 2.2. *For any $n, k \in \mathbb{Z}^+$ where $n \geq 3$ and $2 \leq k \leq n$, there exists $I_n \in \mathscr{K}_n$ where* NN *gives the unique maximum k-cycle.*

Proof. Consider the instance I_n whose weight for each arc (i, j) is

$$w(i,j) = \begin{cases} ik & ; 1 \le i \le k-1, j = i+1 \\ ik+1 & ; 1 \le i \le k-1, j \ge i+2 \\ k^2(k-1)+k & ; i = k, j = 1 \\ \epsilon & ; \text{otherwise}, \end{cases}$$

where $0 < \epsilon < \frac{1}{k}$. When we start at node 1, we obtain the heuristic solution $C^H = (1, 2, \ldots, k, 1)$ of weight $w(C^H) = \frac{3}{2}k^2(k-1)+k$.

Suppose that there exists a weighted directed k-cycle C^X of weight $w(C^X) \ge w(C^H)$. If C^X does not contain the arc $(k, 1)$, $w(C^X) \le w(k, 1) < w(C^H)$. Thus, the arc $(k, 1)$ is in C^X. It follows that there exists a subpath $P = (1 = u_1, u_2, \ldots, u_k = k)$ of length $k-1$ in C^X. If the sequence of nodes in P is an increasing sequence, P is $(1, 2, \ldots, k)$ and we have $C^X = C^H$. Assume that P is not an increasing sequence of nodes. Hence, there exists an arc (u_i, u_j) where $u_i > u_j$ in P. This arc must be an arc of weight ϵ. Besides, note that for each i, $1 \le i \le k-1$, all arcs with weight ik or $ik+1$ shares the same tail i. Hence, for each $i, 1 \le i \le k-1$, at most one of arcs of weight ik or $ik+1$ can appear in P. Therefore,

$$w(C^X) \le \epsilon + \sum_{i=2}^{k-1}(ik+1) + w(k,1) = w(C^H) - (2-\epsilon) < w(C^H),$$

contradicting to the assumption that $w(C^X) \ge w(C^H)$. We can conclude that $w(C^H)$ is the only maximum k-cycle in I_n. ∎

Corollary 2.3. *Let $n, k \in \mathbb{Z}^+$ where $n \ge 3$ and $2 \le k \le n$. For the* MWDkCP *on instances of size n, $d(n)$ of* NN *is* 1.

Proof. From Theorem 2.2, in the given instance I_n, there exists a heuristic solution C^H that dominates only 1 weighted directed k-cycle in I_n which is C^H itself. As a result, $d(n)$ of NN for the MWDkCP is at most 1. Recall that $d(n)$ of any heuristic is at least 1. Therefore, $d(n)$ of NN for the MWDkCP is 1. ∎

Corollary 2.4. *Let $n, k \in \mathbb{Z}^+$ where $n \ge 3$ and $2 \le k \le n$. For the* MWDkCP *on instances of size n, $dr(n)$ of* NN *is $\frac{1}{\binom{n}{k}(k-1)!}$. Then we get $dr(n) = O(n^{-k})$.*

Gutin et al. [2] presented an instance that greedy heuristic for the ATSP yields the maximum Hamiltonian cycle and inferred, without providing a detailed proof, that the result from NN starting at node 1 is also the maximum Hamiltonian cycle. Our constructed graph can be used as an instance to prove that $d(n)$ of NN for the ATSP is 1.

Since there are at least two different k-cycle in the instance I_n constructed in the proof of Theorem 2.2, the weight of the heuristic solution, which is also the unique maximum k-cycle, is greater than $\mathscr{A}(I_n)$. So we obtain the following result.

Theorem 2.5. *Let $n, k \in \mathbb{Z}^+$ where $n \ge 3$ and $2 \le k \le n$.* NN *is worse than average.*

Theorem 2.2 shows that $d(n)$ of NN is 1 in general. To finish the investigation on NN, we propose a condition where NN does not give the unique maximum k-cycle on instances satisfying the condition.

For any node u, let $\mathbb{C}^u$ denote the set of all k-cycles we obtain when using NN starting at node u. Note that it is possible that there are two distinct nodes u and v such that $\mathbb{C}^u \cap \mathbb{C}^v \neq \emptyset$.

Lemma 2.6. *Let I_n be an instance of* MWDkCP *for $n, k \in \mathbb{Z}^+$ where $n \geq 4$ and $k \geq 4$. Let $C = (u_1, u_2, \ldots, u_k, u_{k+1} = u_1)$ be a cycle obtained from NN. If there exist nodes u_i and u_j that are non-adjacent in C and $C \in \mathbb{C}^{u_i} \cap \mathbb{C}^{u_j}$, C is not the unique maximum k-cycle in I_n.*

Proof. Let I_n be an instance of size n and $C = (u_1, u_2, \ldots, u_k, u_{k+1} = u_1)$ be a cycle resulted from applying NN. Assume that there exist non-adjacent nodes u_i and u_j in C such that $C \in \mathbb{C}^{u_i} \cap \mathbb{C}^{u_j}$. Since u_i and u_j are in C, without loss of generality, we can assume that $i = 1$. Thus, $3 \leq j \leq k - 1$.

Consider the cycle C' obtained from C by replacing the arcs (u_1, u_2), (u_{j-1}, u_j) and (u_j, u_{j+1}) by the arcs $(u_1, u_j), (u_j, u_2)$ and (u_{j-1}, u_{j+1}). Since $C \in \mathbb{C}^{u_1}$, $w(u_1, u_j) - w(u_1, u_2) \geq 0$ and $w(u_{j-1}, u_{j+1}) - w(u_{j-1}, u_j) \geq 0$. Similarly, $C \in \mathbb{C}^{u_j}$ leads to the fact that $w(u_j, u_2) - w(u_j, u_{j+1}) \geq 0$. Then

$$\begin{aligned} w(C') &= w(C) + w(u_1, u_j) - w(u_1, u_2) + w(u_{j-1}, u_{j+1}) \\ &\quad - w(u_{j-1}, u_j) + w(u_j, u_2) - w(u_j, u_{j+1}) \\ &\quad \geq w(C). \end{aligned}$$

Therefore, C is not the the unique maximum k-cycle. ■

The proof for the sufficiency of the condition directly follows.

Theorem 2.7. *Let I_n be an instance of* MWDkCP *for $n, k \in \mathbb{Z}^+$ where $n \geq 4$ and $k \geq 4$. If for any output C from NN, there exist nodes i and j that are non-adjacent in C and $C \in \mathbb{C}^i \cap \mathbb{C}^j$, NN does not give the unique maximum k-cycle in I_n.*

As for RNN, if RNN results in the unique maximum k-cycle C, we have $C \in \mathbb{C}^1 \cap \mathbb{C}^2 \cap \ldots \cap \mathbb{C}^n$. Without loss of generality, suppose that $C = (1, 2, \ldots, k, 1)$. When $k \geq 4$, node 1 and node 3 are non-adjacent in C and $C \in \mathbb{C}^1 \cap \mathbb{C}^3$. From Lemma 2.6, C is not the unique maximum k-cycle, a contradiction. Hence, when $k \geq 4$, for RNN, $d(n) \neq 1$. We offer a better lower bound for $d(n)$ of RNN in the next section.

2.2. Repetitive Nearest Neighbour Algorithms (RNN)

When $k \neq n$, we show that RNN is arbitrarily bad and can be worse than average. As for the domination analysis, $d(n)$ of RNN is 2 when $n = 3$, and we propose a lower bound and an upper bound of it when $n \geq 4$.

Theorem 2.8. *Let $n, k \in \mathbb{Z}^+$ where $n \geq 3$, $2 \leq k < n$ and $N > 1$. There exists $I_n \in \mathscr{K}_n$ such that $\frac{w(C^{RNN}(I_n))}{w(C^M(I_n))} > N$.*

Proof. Consider the instance I_n of size n where the positive weight for each arc (i, j) is defined as follows:

$$w(i,j) = \begin{cases} M & ; i = 1, 2 \leq j \leq n \\ \epsilon & ; 2 \leq i \leq n, j = 1 \\ 2\epsilon & ; 2 \leq i, j \leq n \end{cases}$$

where $\epsilon > 0$ and $M = 2k\epsilon N$. Since $k \geq 2$ and $N > 1$, $M > 4\epsilon$.

Performing NN starting at node 1, the procedure gives the k-cycle $(1, u_1, u_2, \ldots, u_{k-1}, 1)$ for some $u_1, u_2, \ldots, u_{k-1}$ in the set $\{2, 3, \ldots, n\}$. Its weight is $M + (2k-3)\epsilon$. When the process begins at $i \neq 1$, the result is the cycle $(i, 1, u_1, u_2, \ldots, u_{k-2}, i)$ for some $u_1, u_2, \ldots, u_{k-2}$ in the set $\{2, 3, \ldots, n\}$, which has weight $M + (2k-3)\epsilon$. Thus, any of these cycles constructed from NN can be the heuristic solution from RNN of weight $M + (2k-3)\epsilon > 2k\epsilon N$.

Consider the k-cycle $C = (2, 3, \ldots, k, k+1, 2)$ of weight $2k\epsilon$. Suppose that there exists a cycle C' with weight $w(C') < 2k\epsilon$. The only possibility is that C' contains an arc of weight ϵ. Observe that any arc of weight ϵ has node 1 as its terminal node, so any cycle can have at most one arc of weight ϵ, and there must be an arc of weight M in the cycle. Then $w(C') \geq M + (2k-3)\epsilon > 2k\epsilon$, a contradiction. It follows that $w(C^M(I_n)) = 2k\epsilon$. Therefore, $\frac{w(C^{RNN}(I_n))}{w(C^M(I_n))} \geq \frac{M+(2k-3)\epsilon}{2k\epsilon} > \frac{2k\epsilon N}{2k\epsilon} = N$. ∎

Similar to NN, we can conclude from Theorem 2.8 that when $k \neq n$, an approximation ratio of RNN for the MWDkCP does not exist.

We can modify the instance constructed in the proof of Theorem 2.8 so that it becomes an instance showing that RNN is worse than average when $k < n$. We start from finding the average value of the weights of all weighted directed k-cycles in each instance.

Lemma 2.9. *Let $n, k \in \mathbb{Z}^+$ where $n \geq 3$ and $I_n \in \mathscr{K}_n$. The average value of the weights of all k-cycles in I_n where $2 \leq k \leq n$ is*

$$\mathscr{A}(I_n) = \frac{k}{n(n-1)} \sum_{(u,v)\in A(I_n)} w(u,v).$$

Proof. Let $\mathscr{C} = \{C^X : C^X \text{ is a } k\text{-cycle in } I_n\}$. Consider an arc $(u,v) \in A(I_n)$. Any k-cycle containing (u,v) is composed of (u,v) and a directed vu-path from v to u of length $k-1$. There are $\binom{n-2}{k-2}(k-2)!$ directed paths satisfying the condition. Hence, $w(u,v)$ appears $\binom{n-2}{k-2}(k-2)!$ times in $\sum_{C^X\in\mathscr{C}} w(C^X)$. The number of all k-cycles in I_n is $\binom{n}{k}(k-1)!$. Therefore,

$$\mathscr{A}(I_n) = \frac{1}{|\mathscr{C}|} \sum_{C^X\in\mathscr{C}} w(C^X) = \frac{k}{n(n-1)} \sum_{(u,v)\in A(I_n)} w(u,v).$$

∎

Next, we show that RNN can be worse than average when $k < n$.

Theorem 2.10. *Let $n, k \in \mathbb{Z}^+$ where $n \geq 3$ and $2 \leq k < n$. RNN is worse than average*

Proof. We consider the instance I_n in the proof of Theorem 2.8 where the value of M is changed from $2k\epsilon N$ to $3\epsilon + 1$. We have $\mathscr{A}(I_n) = 2k\epsilon + \frac{k}{n} < M + (2k-3)\epsilon$, which is the weight of an outcome from RNN. Hence, RNN is worse than average. ∎

As for the domination analysis, we first show that $d(3) = 2$.

Theorem 2.11. RNN *for the* MWDkCP *has* $d(3) = 2$.

Proof. Let I_3 be an instance with node set $\{1,2,3\}$ and weight function w.

When $k=2$, there are exactly three 2-cycles, which are $C^1=(1,2,1)$, $C^2=(2,3,2)$, and $C^3=(3,1,3)$. We can assume without loss of generality that C^1 is an outcome of RNN. Applying NN starting at node 3 cannot give C^1, so we obtain at least 2 distinct 2-cycles after performing NN starting at all nodes in I_3, namely C^1 and C^3. Therefore, C^1 dominates C^1 itself and another cycle, that is $d(3)\geq 2$.

To show that $d(3)\leq 2$, consider an instance with weight function w where $w(1,2)=w(3,2)=1, w(2,1)=2, w(2,3)=2.1$, and $w(1,3)=w(3,1)=1.1$. NN produces C^1 with $w(C^1)=3$ when starting at node 1 and node 2, and produces C^2 with $w(C^2)=3.1$ when starting at node 3. So RNN outputs C^1 of weight 3 at the end, but the optimal solution is C^3 with weight 2.2. Therefore, C^1 dominates only two 2-cycles in this instance. In conclusion, when $k=2$, $d(3)=2$.

If $k=3$, the only 3-cycles in I_3 is $C^A=(1,2,3,1)$ and $C^B=(3,2,1,3)$. Without loss of generality, assume that the cycle C^A is the outcome from RNN. If there is a node u such that performing NN starting at u results in the cycle C^B, we have $w(C^A)\leq w(C^B)$ since RNN gives C^A instead of C^B at the end of the process. Consider the case when NN always outputs C^A for any starting node $u\in\{1,2,3\}$. We can see that $(1,2)$ is chosen when starting at node 1, $(2,3)$ is chosen when starting at node 2, and $(3,1)$ is chosen when starting at node 3. Hence, $w(1,2)\leq w(1,3)$, $w(2,3)\leq w(2,1)$, and $w(3,1)\leq w(3,2)$. Furthermore, $w(C^A)=w(1,2)+w(2,3)+w(3,1)\leq w(1,3)+w(2,1)+w(3,2)=w(C^B)$ in this situation. Therefore, C^A dominates 2 cycles: C^B and C^A itself. It follows that when $k=3$, we have $d(3)\geq 2$. Since there are only two 3-cycles in I_3, we have $d(3)=2$. ∎

Previously, we have pointed out that when $n,k\geq 4$, as a trivial result from Lemma 2.6, $d(n)$ of RNN is greater than 1. In the next theorem, we establish that when $n\geq 4$ and $k\geq 2$, $d(n)$ of RNN is at least $\left\lceil\frac{k}{2}\right\rceil$. The proof follows the idea of the proof for a lower bound of $d(n)$ of RNN for the ATSP by Gutin et al. [2], and hence, is omitted.

Theorem 2.12. *For any $n,k\in\mathbb{Z}^+$ where $n\geq 4$ and $2\leq k\leq n$, $d(n)$ of* RNN *for the* MWDkCP *on instances of size n is at least $\left\lceil\frac{k}{2}\right\rceil$.*

Note that we can construct an instance I_n with positive weight w' from an instance with any real-valued weight function w in the way that for any k-cycles C^A and C^B in I_n, $w'(C^A)\leq w'(C^B)$ if and only if $w(C^A)\leq w(C^B)$. Therefore, the weight function of the instance we give in the following proof allows weight zero on some arcs.

Theorem 2.13. *For $n,k\in\mathbb{Z}^+$ where $n\geq 4$ and $2\leq k\leq n$, $d(n)$ of* RNN *for the* MWDkCP *on instances of size n is $O(n^{\sqrt{2k-2}})$, when k is a constant.*

Proof. Consider the instance I_n of size n and the non-negative weight for each arc (i,j) is defined as follows:

$$w(i,j)=\begin{cases} iM & ;1\leq i\leq k-1, j=i+1\\ iM+1 & ;1\leq i\leq k-1, j=i+2\\ iM+2 & ;1\leq i\leq k-1, j\geq i+3\\ jM & ;i=k, 2\leq j\leq k-1\\ 0 & ;j=1\\ \epsilon & ;\text{otherwise,}\end{cases}$$

where $0 < \epsilon < \frac{1}{k}$ and $M = 4k + \epsilon$. Denote $N = \{1, 2, \ldots, n\}$ and $K = \{1, 2, \ldots, k\}$.

Performing NN starting at node 1 or k, the procedure gives the k-cycle $C^H = (1, 2, \ldots, k, 1)$ of weight $\frac{1}{2}k(k-1)M$. When the process begins at $j \in K - \{1, k\}$, the result is the cycle $(j, 1, 2, \ldots, j-1, j+1, j+2, \ldots, k, j)$ of weight $w(C^H) + 1$. If NN starts at $j \in N \backslash K$, the output cycle is $(j, 1, 2, \ldots, k-1, j)$ of weight $w(C^H) + 1$ as well. Thus, the heuristic solution from RNN is $C^H = (1, 2, \ldots, k, 1)$. Then we have $w(C^H) = \frac{1}{2}k(k-1)M$.

We consider the weight of k-cycle $C^X = (u_1, u_2, \ldots, u_k, u_1)$ in I_n in two cases, based on the node set of the cycle. For any $i \leq j$, we call a directed path $(u_i, u_{i+1}, \ldots, u_j)$ in C^X an *inner path* if the following properties hold

(1) $1 \leq u_i < u_{i+1} < \ldots < u_j \leq k$.
(2) Denote $u_{k+1} = u_1$ and $u_0 = u_k$. The cycles $(u_i, u_{i+1}, \ldots, u_j, u_{j+1})$ and $(u_{i-1}, u_i, u_{i+1}, \ldots, u_j)$ do not satisfy property 1.

In case that an inner path has only one node, we call it an *inner node*. We can see that all inner paths in C^X are pairwisely disjoint. Let r be the number of inner paths in C^X. Let all inner paths in C^X be denoted by inner path P_i for $i = 1, 2, \ldots, r$, appearing in order as one traverses from the smallest node. Let s_i and t_i be the first and the last node in P_i, respectively.

Case 1 The node set $N(C^X) = K$. Then each node in C^X is in an inner path. We consider whether 1 and k are in the same inner path. Then node 1 is in P_1 and $s_1 = 1$. Note that for every $i = 1, 2, \ldots, r-1$, $t_i > s_{i+1}$, otherwise, attaching the node s_{i+1} at the end of P_i satisfies property 1 in the definition of an inner path, a contradiction.

Case 1.1 Node k is in P_1. Then P_1 starts at node 1 and ends at node k. One of the cycle in this case is C^H, containing P_1 as the only inner path. The other cycle must contain more than one inner path.

First, we consider the case when $r = 2$, that is all inner paths in C^X are P_1 and P_2. If P_2 is an inner node j for some $2 \leq j \leq k-1$, then $C^X = (j, 1, 2, \ldots, j-1, j+1, \ldots, k, j)$. Recall that these C^X are the outputs from applying NN starting at the node $j \in \{2, 3, \ldots, k-1\}$ and are dominated by C^H. When P_2 is not an inner node, we have $s_2 < t_2$. Hence, $w(C^X) \leq w(C^H) + 2(k-2) + (s_2 - t_2)M < w(C^H)$.

Now let consider the case when $r \geq 3$. Since nodes 1 and k are in the same inner path P_1, $r \leq k-1$. Recall that for all $i = 1, 2, \ldots, r-1$, $t_i > s_{i+1}$, and hence, $w(t_i, s_{i+1}) = \epsilon$ for all $i \in \{2, 3, \ldots, r-1\}$. Thus,

$$\begin{aligned} w(C^X) &\leq w(C^H) + 2(k-r) + s_2M - \sum_{i=2}^{r} t_iM + (r-2)\epsilon \\ &\leq w(C^H) + 2(k-3) + (s_2 - t_2)M - t_3M + (k-3)\epsilon \\ &< w(C^H). \end{aligned}$$

Therefore, we obtain $k-1$ k-cycles dominated by C^H from this case, which are C^H itself and $k-2$ cycles from the case when $r = 2$ and P_2 is an inner node.

Case 1.2 Node k is not in P_1. Let P_j be the inner path ending with node k, that is $t_j = k$. Then P_1 and P_j are not inner nodes. Thus, $j \geq 2$ and $2 \leq r \leq k-2$. Recall that $w(t_i, s_{i+1}) = \epsilon$ for all $i \in \{1, 2, \ldots, r-1\} \backslash \{j\}$.

If $j \neq r$, we have

$$w(C^X) \leq w(C^H) + 2(k-r) + s_{j+1}M - \sum_{\substack{i=1 \\ i\neq j}}^{r} t_i M + (r-2)\epsilon$$
$$\leq w(C^H) + 2(k-2) + (s_{j+1} - t_{j+1})M - t_{j-1}M + (k-2)\epsilon$$
$$< w(C^H).$$

When $j = r$, $w(C^X) \leq w(C^H) + 2(k-r) - \sum_{i=1}^{r-1} t_i M + (r-1)\epsilon < w(C^H)$.

Hence, C^H dominates no weighted directed k-cycles in this case.

Case 2 The node set $N(C^X) \neq K$. Let the node set of C^X be $(K \backslash \{h_1, h_2, \ldots, h_m\}) \cup \{l_1, l_2, \ldots, l_m\}$ for some integer $m \geq 1$, $h_1, h_2, \ldots,\ h_m \in K$ and $l_1, l_2, \ldots, l_m \in N \backslash K$. Denote $H = \{h_1, h_2, \ldots, h_m\}$ and $L = \{l_1, l_2, \ldots, l_m\}$.

We call a path in C^X consisting only nodes in L an *outer path* if it is not contained in another longer path consisting only nodes in L. Hence, these outer paths are pairwisely disjoint. Let ρ be the number of outer paths in C^X. Let all outer paths in C^X be denoted by outer path Q_i for $i = 1, 2, \ldots, \rho$, appearing in order as one traverses from the smallest node. Let σ_i and τ_i be the first and the last node in Q_i, respectively. Note that $\rho \leq m$.

Denote $P_{r+1} = P_1$. For each pair of P_i and P_{i+1} when $i = 1, 2, \ldots, r$, they appear consecutively or are separated by an outer path. Consider a sequence $\mathbb{P}$ of consecutive inner paths, without any outer paths in between. We call $\mathbb{P}$ a *chain* if it is not a part of another longer sequence of consecutive inner paths. Then there are exactly ρ chains in C^X, and hence, $1 \leq \rho \leq r$. Let $\mathbb{P}_i$ starting from an inner path $P_{i,1}$ to an inner path $P_{i,r(i)}$ for $i = 1, 2, \ldots, \rho$ be all chains in C^X, appearing in order as one traverses from the smallest node. Thus, $P_{1,1} = P_1$ and $r(1) + r(2) + \ldots + r(\rho) = r$. Let $s_{i,j}$ and $t_{i,j}$ be the first and the last node in $P_{i,j}$, respectively. Then

$$w(C^X) = \sum_{i=1}^{\rho} \sum_{j=1}^{r(i)} w(P_{i,j}) + \sum_{i=1}^{\rho} \sum_{j=1}^{r(i)-1} w(t_{i,j}, s_{i,j+1}) +$$
$$\sum_{i=1}^{\rho} w(t_{i,r(i)}, \sigma_i) + (m-1)\epsilon + w(\tau_\rho, s_{1,1}).$$

Case 2.1 Node $k \in H$. Then we have $1 \leq m \leq k$ and $1 \leq \rho \leq r \leq k$. Assume that $m \neq 1$ or there exists a chain consisting of at least 2 inner paths. Then $h_m = k$. Observe that $w(t_{i,j}, s_{i,j}) = \epsilon$ for $i = 1, 2, \ldots, \rho$ and $j = 1, 2, \ldots, r(i) - 1$. By the assumption, we have $\sum_{i=1}^{m-1} h_i M \geq M$ or $\sum_{i=1}^{\rho} \sum_{j=1}^{r(i)-1} t_{i,j} M \geq M$. We obtain

$$w(C^X) \leq w(C^H) + 2(k - m - r + \rho) + (r + m - \rho)\epsilon$$
$$- \sum_{i=1}^{m-1} h_i M - \sum_{i=1}^{\rho} \sum_{j=1}^{r(i)-1} t_{i,j} M$$
$$< w(C^H) + 2k + 2k\epsilon - M$$
$$< w(C^H).$$

Now we consider the case when $m = 1$ and every chain is a single inner path. Then $H = \{k\}$. Let l be the only element in L. Note that $1 \leq \rho \leq m = 1$, so $\rho = 1$. Hence, C^X contains only one chain which is an inner path of $k-1$ nodes and one outer path which is a single node l. Thus, $C^X = (1, 2, \ldots, k-1, l, 1)$ with weight at least $w(C^H) + 1$. These $n-k$ k-cycles are dominated by C^H.

Case 2.2 Node $k \notin H$. Let $P_{\kappa,\lambda}$ be the inner path ending with node k, that is k is in chain $\mathbb{P}_\kappa$ and $t_{\kappa,\lambda} = k$.

Case 2.2.1: k is at the end of $\mathbb{P}_\kappa$, that is $\lambda = r(\kappa)$. Since $1 \leq m \leq k$ and $1 \leq \rho \leq r \leq k$,

$$\begin{aligned} w(C^X) \leq\ & w(C^H) + 2(k - m - r + \rho - 1) + (r - \rho + m + 1)\epsilon \\ & - \sum_{i=1}^{m} h_i M - \sum_{i=1}^{\rho} \sum_{j=1}^{r(i)-1} t_{i,j} M \\ & < w(C^H) + 2k + 2k\epsilon - M \\ & < w(C^H). \end{aligned}$$

No directed k-cycles in this case is dominated by the heuristic solution.

Case 2.2.2: k is not at the end of $\mathbb{P}_\kappa$, that is $\lambda \neq r(\kappa)$. Since k is always at the end of an inner path, we have a chain containing more than one inner path. Hence, $1 \leq \rho < r \leq k$ in this case. For any arc $a = (i, j)$ of weight $iM + \delta$ where $\delta = 0, 1$ or 2, denote $\delta_a = w(u, v) - iM$. Therefore, for any arc $a \in A(I_n)$, $\delta_a = 0, 1$ or 2. Recall that $w(t_{i,j}, s_{i,j}) = \epsilon$ for $i = 1, 2, \ldots, \rho$ and $j = 1, 2, \ldots, r(i) - 1$ when $t_{i,j} \neq k$. Let $T = \{t_{i,j(i)} : i = 1, 2, \ldots, \rho, j(i) = 1, 2, \ldots, r(i) - 1\} \backslash \{t_{\kappa,\lambda}\}$. Denote $\tau = |T| = r - \rho - 1$. Let $t_1 < t_2 < \ldots < t_\tau$ be all elements in T. It follows that

$$\begin{aligned} w(C^X) =\ & w(C^H) - dM + \sum_{i=1}^{\rho} \sum_{j=1}^{r(i)} \sum_{a \in A(P_{i,j})} \delta_a + (m + r - \rho - 2)\epsilon \\ & + \sum_{i=1}^{\rho} \delta_{(t_{i,r(i)}, \sigma_i)} + w(\tau_\rho, s_{1,1}), \end{aligned}$$

where $d = \sum_{i=1}^{m} h_i + \sum_{i=1}^{\tau} t_i M - s_{\kappa,\lambda+1}$. If $d \geq 1$, since $1 \leq m \leq k$ and $1 \leq \rho \leq r \leq k$, we have $w(C^X) \leq w(C^H) - M + 2(k - m - (r - \rho)) + (m + r - \rho - 1)\epsilon < w(C^H)$.

Otherwise, $d \leq 0$, that is

$$s_{\kappa,\lambda+1} - \sum_{i=1}^{m} h_i - \sum_{i=1}^{\tau} t_i \geq 0. \tag{2.1}$$

In case that $d < 0$, $w(C^X) \geq w(C^H) + M + (m + r - \rho - 2)\epsilon > w(C^H)$. It means that these C^X are dominated by C^H.

When (2.1) holds as an equation, C^X is not dominated by C^H only if

$$\sum_{i=1}^{\rho} \sum_{j=1}^{r(i)} \sum_{a \in A(P_{i,j})} \delta_a + \sum_{i=1}^{\rho} \delta_{(t_{i,r(i)}, \sigma_i)} + (m + r - \rho - 2)\epsilon + w(\tau_\rho, s_{1,1}) < 0.$$

It implies that $m + r - \rho < 2$. Since $m \geq 1$ and $\rho < r$, this inequality does not hold for any k-cycles in this case. Thus, all k-cycles satisfying inequality (2.1) in this case are dominated by C^H. To find the number of such k-cycles, we consider the process of constructing these required k-cycles, splitting into three steps of work.

Step 1: Choose nodes in H and T. Inequality (2.1) holds when $\sum_{i=1}^{m} h_i + \sum_{i=1}^{\tau} t_i \leq s_{\kappa,\lambda+1}$. For each $s_{\kappa,\lambda+1} \leq k - 1$, we consider all partitions Π_p of p, where p is an integer such that $1 \le p \le s_{\kappa,\lambda+1}$, into distinct summands. Observe that there are $\frac{1}{2}k(k-1)$ ways to choose a pair of p and $s_{\kappa,\lambda+1}$.

Let π_p be the number of partitions Π_p of p. For each p, $\pi_p \le \pi_{k-1}$ which is known to be $O\left(e^{\pi\sqrt{\frac{k-1}{3}}}k^{-\frac{3}{4}}\right)$ [14]. For each Π_p, we use n_p to denote the number of summands in Π_p. Then we have $1+2+\ldots+n_p \le p$. Hence, $n_p \le \frac{1}{2}(\sqrt{8p+1}-1) \le \sqrt{2p} \le \sqrt{2(k-1)}$. There are at most $2^{\sqrt{2(k-1)}}$ ways to put these n_p summands into H and T so that $H \ne \emptyset$. All k-cycles with sets H and T obtained in this way satisfy inequality (2.1). Moreover, we have $m+\tau = |H|+|T| \le \sqrt{2(k-1)}$.

In conclusion, the number of ways to select the elements in H and T is

$$O\left(k^2 \cdot e^{\pi\sqrt{\frac{k-1}{3}}}k^{-\frac{3}{4}} \cdot 2^{\sqrt{2(k-1)}}\right) = O\left(k^{\frac{5}{4}} \cdot (e^{\frac{\pi}{\sqrt{3}}} \cdot 2^{\sqrt{2}})^{\sqrt{k-1}}\right).$$

Step 2: Constructing the inner paths. Given H and T, we construct a k-cycles C^X. For $i = 1,2,\ldots,\tau$, let P_i be the inner path ended by t_i and for $j = 1,2,\ldots,\rho$, $P_{j,r(j)}$ is the inner path that is the end of chain $\mathbb{P}_j$. Recall that $\rho \le m$, so there are $\frac{1}{\rho}(\tau+\rho)! \le (\sqrt{2(k-1)}-m+\rho)! \le (\sqrt{2(k-1)})!$ ways to arrange these P_i's and $P_{j,r(j)}$'s in a directed cycle.

Consider the inner path starting with $s_{\kappa,\lambda+1}$. Since $t_i < p \le s_{\kappa,\lambda+1}$ for all $t_i \in T$, this path cannot be ended by a node in T. So this path must be a path $P_{j,r(j)}$ for some $1 \le j \le \rho$. The inner path ended by k must be placed right before this path. Thus, there are $\rho \le m \le \sqrt{2(k-1)}$ choices for the position of the inner path ended by k.

Now nodes $k, s_{\kappa,\lambda+1}$ and all nodes in T are already placed in inner paths. We consider nodes in $\{1,2,\ldots,k-1\}\backslash(H\cup\{s_{\kappa,\lambda+1}\}\cup T)$. There are $k-1-(m+1+\tau) = k-2-(\tau+m) \le k-3$ nodes in total. Note that for each P_i, any node included in P_i cannot exceed t_i. Each of these remaining nodes has at most $r = \tau+\rho+1 \le \tau+m+1 \le \sqrt{2(k-1)}+1$ inner paths in which it can be put. Once the node set of an inner path is given, all nodes must be ordered increasingly. Thus, there are at most $(\sqrt{2(k-1)}+1)^{k-3}$ ways to complete this process.

From the asymptotical approximation in [14], the number of choices to place the nodes in all inner paths is

$$\begin{aligned}
&O\left((\sqrt{2(k-1)})^{\sqrt{2(k-1)}+\frac{1}{2}}e^{-\sqrt{2(k-1)}}\sqrt{2(k-1)}(\sqrt{2(k-1)}+1)^{k-3}\right)\\
&= O\left((e^{-\sqrt{2}})^{\sqrt{k-1}}(\sqrt{2(k-1)}+1)^{k+\sqrt{2(k-1)}-\frac{3}{2}}\right).
\end{aligned}$$

Step 3: Constructing the outer paths. There are $\binom{n-k}{m}$ ways to choose the elements in L to be in the outer paths $Q_1, Q_2, \ldots, Q_\rho$. There are $m!$ ways to arrange all m nodes in a row. Since every outer path cannot be empty, there are $\binom{m-1}{\rho-1}$ ways to split these m nodes into ρ outer paths. Since $1 \le \rho \le m \le \sqrt{2(k-1)}$, the number of ways to finish this step is

$$\begin{aligned}
\binom{n-k}{m}\binom{m-1}{\rho-1}m! &\le \frac{(n-k)^m}{m!}\frac{(m-1)^{\rho-1}}{(\rho-1)!}m!\\
&\le (n-k)^{\sqrt{2(k-1)}}(\sqrt{2(k-1)}-1)^{\sqrt{2(k-1)}-1},
\end{aligned}$$

Summarily, the number of directed k-cycles satisfying inequality (2.1) in this case is $O(f(n,k))$ where $f(n,k)$ is

$$k^{\frac{5}{4}}(2^{\sqrt{2}} \cdot e^{\frac{\pi}{\sqrt{3}}-\sqrt{2}})^{\sqrt{k-1}}(\sqrt{2(k-1)}+1)^{k+2\sqrt{2(k-1)}-\frac{5}{2}}(n-k)^{\sqrt{2(k-1)}},$$

which is $O(n^{\sqrt{2k-2}})$ when k is a constant.

In conclusion, the number of k-cycles dominated by C^H is $k-1$ in case 1.1, 0 in case 1.2, $n-k$ in case 2.1 and $O(n^{\sqrt{2k-2}})$ in case 2.2. Thus, $d(n) = O(n^{\sqrt{2k-2}})$. ∎

Note that when $k = n$, case 2 in the proof does not occur. Hence, in the case that $k = n$, our upper bound is $k-1 = n-1$, which is the same as the result for ATSP from Gutin et al. [2].

3. Conclusions

In this study, we first show that the approximation ratio of NN for the MWDkCP does not exist. Then we find $d(n)$ and $dr(n)$ of NN for the MWDkCP. Theorem 2.2 shows that for any $n, k \in \mathbb{Z}^+$ where $n \geq 3$ and $2 \leq k \leq n$, there is an instance $I_n \in \mathscr{K}_n$ whose solution from NN can be the unique maximum k-cycle. It implies that for NN, $d(n) = 1$, and hence, NN is also worse than average.

Similar to NN, when $k \neq n$, an approximation ratio of RNN for the MWDkCP does not exist. We also provide a formula for the average value of the weights of all k-cycles in each instance, and prove that RNN can be worse than average if $k < n$.

On the domination analysis, $d(n)$ of RNN is 2 when $n = 3$. Then we show that for any $n, k \in \mathbb{Z}^+$ where $n \geq 4$ and $k \geq 2$, $d(n)$ of RNN is at least $\lceil \frac{k}{2} \rceil$ and is $O(n^{\sqrt{2k-2}})$, when k is a constant. The gap between the lower bound and the upper bound proposed in this article is considerably more narrow when $k = n$.

It is impossible to establish an approximation ratio for either NN or RNN, and both of them are worse than average. We can see from Theorem 2.2 that $d(n)$ of NN is 1. We offer a condition to guarantee that NN does not give the unique maximum k-cycle on an instance meeting the condition. Finding other sufficient conditions for avoiding these worst cases would be useful for those who apply NN and RNN to a specific collection of instances. For example, if one can find a sufficient condition for the existence of approximation ratio of NN or RNN for the MWDkCP, then an approximation ratio can be established when we perform that heuristic on the collection of instances under that condition.

Our analyses show that NN is not appropriate for MWDkCP in term of effectiveness, while RNN is just slightly better when k is close to or equal to n. However, the low complexities of these heuristics makes an output from NN or RNN a good choice for an initial solution of more complicated heuristics, especially when applied to instances satisfying sufficient conditions which ensure that we can avoid these worst scenarios and get some theoretical guarantee.

Acknowledgements

I would like to express my gratitude to my mentor, Chariya Uiyyasathian, for many useful suggestions in this study, and the Institute for the Promotion of Teaching Science and Technology who grants the research fund for my work on this problem.

References

[1] D. L. Applegate, R. E. Bixby, V. Chvtal, W. J. Cook, The Traveling Salesman Problem: A Computational Study, Princeton University Press, Princeton, New Jersey, 2006.

[2] G. Gutin, A. Yeo, A. Zverovich, Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP, Discrete Applied Mathematics 117 (2002) 81–86.

[3] J. Cirasella, D.S. Johnson, L.A. McGeoch, W. Zhang, The asymmetric traveling salesman problem: algorithms, instance generators and tests, Proceedings of Workshop on Algorithm Engineering and Experiments (ALENEX 2001), Lecture Notes in Computer Sciences 2153 (2016) 32–59.

[4] D.S. Johnson, L.A. McGeoch, The traveling salesman problem: a case study in local optimization, Local Search in Combinatorial Optimization, John Wiley and Sons, London (1997), 215–310.

[5] G. Reinelt, The Traveling Salesman Problem: Computational Solutions for TSP Applications, Springer-Verlag, Berlin, Heidelberg, 1994.

[6] V.I. Rublineckii, Estimates of the accuracy of procedures in the traveling salesman problem, Numerical Mathematics and Computer Technology 4 (1973) 18–23 (in Russian).

[7] V.I. Sarvanov, N. Doroshko, The approximate solution of the traveling salesman problem by a local algorithm that searches neighborhoods of factorial cardinality in cubic time, Software: Algorithms and Programs 31 (1981) 11–13 (in Russian).

[8] V.I. Sarvanov, The mean value of the functional in sampling problems, Vestsi Akademii Navuk BSSR. Seryya Fizika-Matematychnykh Navuk 139 (1978) 51–54.

[9] Y. Twitto, Dominance guarantees for above-average solutions, Discrete Optimization 5 (2008) 563–568.

[10] F. Glover, A.P. Punnen, The travelling salesman problem: new solvable cases and linkages with the development of approximation algorithms, Journal of the Operational Research Society 48 (1997) 502–510.

[11] A.P. Punnen, P. Sripratak, D. Karapetyan, Average value of solutions for the bipartite boolean quadratic programs and rounding algorithms, Theoretical Computer Science 565 (2015) 77–89.

[12] E. Zemel, Measuring the quality of approximate solutions to zero-one programming problems, Mathematics of Operations Research 6 (1981) 319–332.

[13] R. Hassin, S. Khuller, z-Approximations, Journal of Algorithms 41 (2001) 429–442.

[14] P. Flajolet, R. Sedgewick, Analytic Combinatorics, Cambridge University Press, New York, 2009.