

# INFO 6205 Spring 2023 Project

## *Traveling Salesman Problem*

**Name: ALTAF SALIM SHAIKH**

**NUID: 002774748**

**Name: AVINASH RAIKESH**

**NUID: 001090182**

- **Aim:** The aim of this project is to implement and compare various optimization methods, including tactical methods such as random swapping, 2-opt and/or 3-opt improvement, and strategic methods such as simulated annealing, ant colony optimization, genetic algorithms, etc., for solving the Traveling Salesman Problem (TSP). Specifically, we will implement the Christofides algorithm as a baseline, and compare its performance to the other methods in terms of solution quality, computation time, and scalability.
- **Introduction:** The Traveling Salesman Problem (TSP) is a classic optimization problem in computer science, which involves finding the shortest possible tour of  $n$  cities (or points in a two-dimensional space) that visits each city exactly once and returns to the starting city. Despite its deceptively simple statement, the TSP is known to be an NP-hard problem, which means that it is computationally infeasible to find an exact solution for large problem instances. To address this challenge, researchers have developed a variety of heuristic and metaheuristic algorithms that can provide good approximate solutions to the TSP in a reasonable amount of time. In this project, we focus on comparing the performance of different optimization methods for solving the TSP, with the goal of identifying the most effective approaches for this challenging problem.  
We begin by implementing the Christofides algorithm, which is a well-known heuristic algorithm for the TSP that guarantees a solution within a factor of  $3/2$  of the optimal solution. We then compare the performance of the Christofides algorithm to other optimization methods, including tactical methods such as random swapping, 2-opt and/or 3-opt improvement, and strategic methods such as simulated annealing, ant colony optimization, genetic algorithms, etc.  
To evaluate the performance of these methods, we measure the quality of the solutions they provide (i.e., how close they are to the optimal solution), as well as their computation time and scalability. Our results provide insights into the strengths and limitations of different optimization methods for solving the TSP and may help inform the development of more effective approaches for this challenging problem.
- **Approach**  
Our approach consists of implementing and comparing several optimization methods for solving the Traveling Salesman Problem (TSP) and evaluating their performance using a set of metrics.  
We use a Kaggle dataset with IDs, latitude and longitude values to test the aforementioned techniques. From there, we utilized the Haversine formula to calculate the distance before utilizing the Prim's algorithm to determine the MST Value and identify the odd vertices. Utilizing excellent Using a matching technique, we combine the edges to create the least

value. The cycle is then found by utilizing the Euler graph and cycle, as well as the Hamiltonian cycle to eliminate unnecessary vertices. Later in the tour, the topics of random swap, 2-opt, simulated annealing, and ant colony were covered. We discovered the ideal tour value with the aid of these algorithms. The correctness and performance of the software were demonstrated via graphical representations after we assessed the program's output.

#### *Optimization Methods*

We begin by implementing the Christofides algorithm as a baseline approach for solving the TSP. The Christofides algorithm is a well-known heuristic algorithm that provides a solution within a factor of  $3/2$  of the optimal solution. It involves constructing a minimum spanning tree (MST) for the given set of cities, and then finding a Eulerian circuit in the MST. Finally, Hamiltonian path algorithm is applied to the circuit to obtain a tour that visits each crime dataset scene exactly once. We then implement several other optimization methods for solving the TSP, including tactical methods such as random swapping, 2-opt, and strategic methods such as simulated annealing, ant colony optimization. These methods are chosen based on their popularity and effectiveness in the literature, and represent a diverse set of approaches to solving the TSP.

#### *Evaluation Metrics*

To evaluate the performance of the optimization methods, we use two main metrics:

*Solution Quality:* We measure the quality of the solutions obtained by each method, by comparing them to the optimal solution (when available) or to the best-known solution in the literature. We use the relative error metric, which is defined as the ratio between the objective value of the obtained solution and the optimal/best known solution. Lower values of relative error indicate better quality solutions.

*Computation Time:* We measure the time taken by each method to obtain a solution, using a standard desktop computer with a fixed set of hardware specifications. We report the average computation time over several runs, to account for variations in performance due to random factors.

## **Program:**

### **Data Structures & classes –**

HashMaps, Lists, Custom classes (Edges,Vertex )

### **Algorithm –**

**Christofides:** The Christofides algorithm is a heuristic algorithm that provides an approximate solution to the traveling salesman problem. It works by finding a minimum spanning tree and creating a subgraph of odd degree vertices. Then it finds a minimum weight perfect matching and combines the minimum spanning tree and the perfect matching that includes all vertices with an even degree. After that, it finds an Eulerian circuit (traversing each edge of the graph exactly once) and transforms it into a Hamiltonian circuit (traversing each vertex of the graph exactly once by skipping previously visited vertices). This algorithm guarantees that the solution will be at most 1.5 times the optimal solution,

with a time complexity of  $O(N^2 \log(N))$  for a graph with  $N$  vertices.

**Random Swapping:** This tactical optimization method involves removing two edges from the solution and reconnecting them in a different way to check if it results in a better solution. It can be more effective in improving the solution but is more computationally expensive than random swapping.

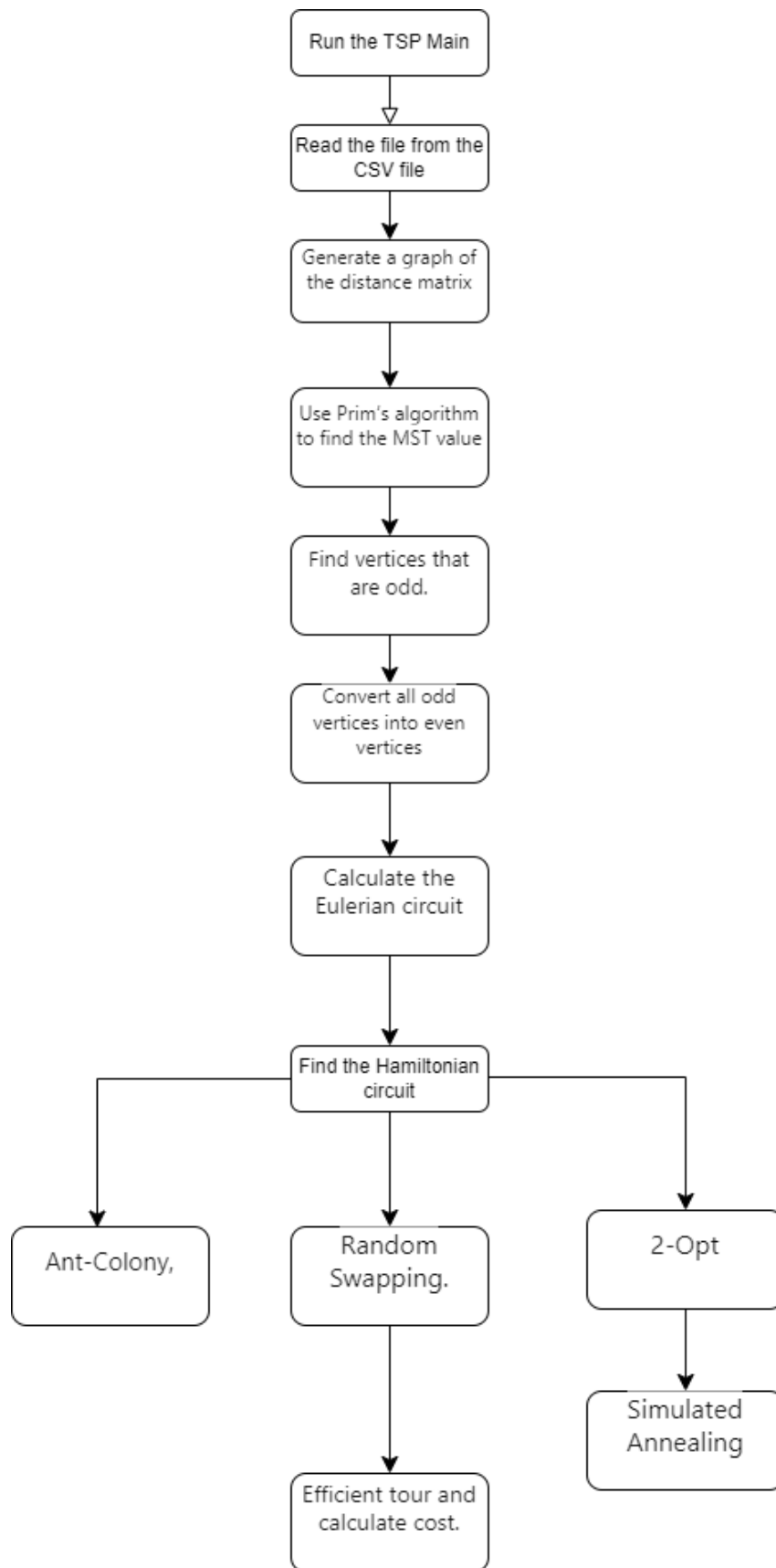
**2-Opt:** is a tactical optimization method that involves removing two edges from the solution and reconnecting them in a different way to check if it results in a better solution. It can be more effective in improving the solution but is more computationally expensive than random swapping.

**Simulated Annealing** is a metaheuristic algorithm for strategic optimization that is based on the annealing process in metallurgy. It starts with a high temperature and gradually cools down the system, allowing for more exploration at the beginning and more exploitation at the end. It can be useful in finding local optima, but careful parameter adjustment is necessary.

**Ant Colony** is a strategic optimization method that is based on the swarm intelligence of ants, who use pheromone trails to locate the shortest route between their nest and a food source. By laying virtual pheromone trails on the graph, the algorithm employs a similar method to determine the shortest path. It can be effective in finding good solutions quickly but requires careful tuning of its parameters and can get stuck in local optima.

**Invariants:** The graph must be undirected and connected. The minimum spanning tree must be connected. The subgraph of odd degree vertices must have an even number of vertices. The edges in the minimum weight perfect matching must connect odd degree vertices. The combined graph must have even degree vertices. The Eulerian circuit must start and end at the same vertex. Throughout the program, the edges produced by the primality methods used to calculate the Minimum Spanning Tree will have the same value. The data that is read from the dataset is moved to the list of type data that is used by the UI to identify the vertices. Every tour is recorded in a separate list so that it may be used to determine the traversal cost and fed to other algorithms to get the most efficient value.

## Flow Chart



## Observations & Graphical Analysis

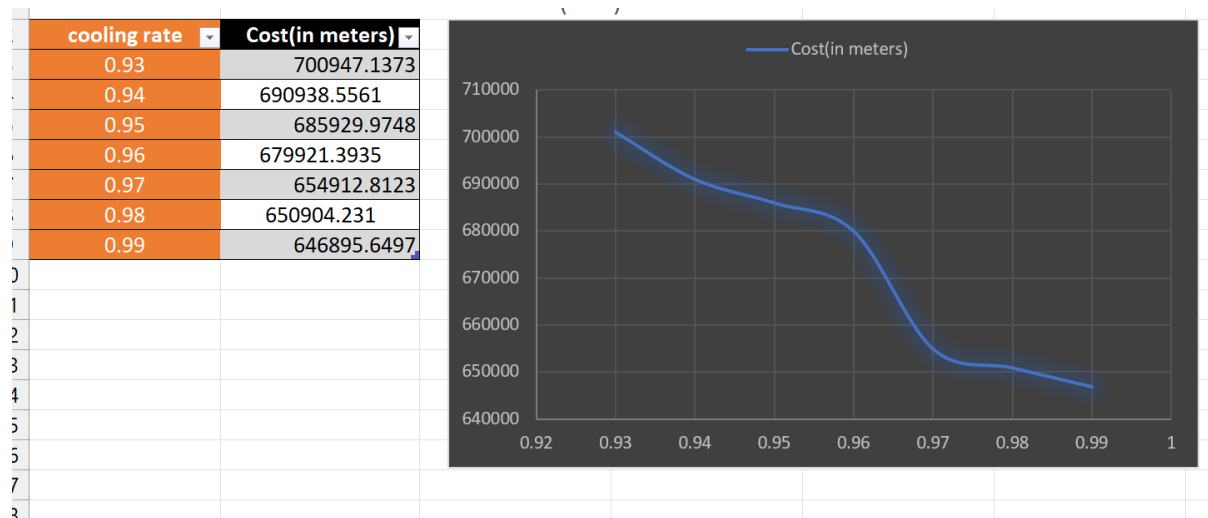
-Below are the optimizations and their observations:

### Random swap:

The cost of the random swap optimization technique varies slightly over the number of iterations because this technique is randomly swapping two edges and calculating the vertex length adding on to the cost. This technique aims at solving clustering by a sequence of prototype swaps and by fine-tuning their exact location by k-means . This randomized search strategy is simple to implement and efficient . It reaches good quality clustering relatively fast, and if iterated longer, it finds the correct clustering with high probability.

### Simulated annealing:

We have tested different cooling rates for simulated annealing optimization.



### Ant colony optimization:

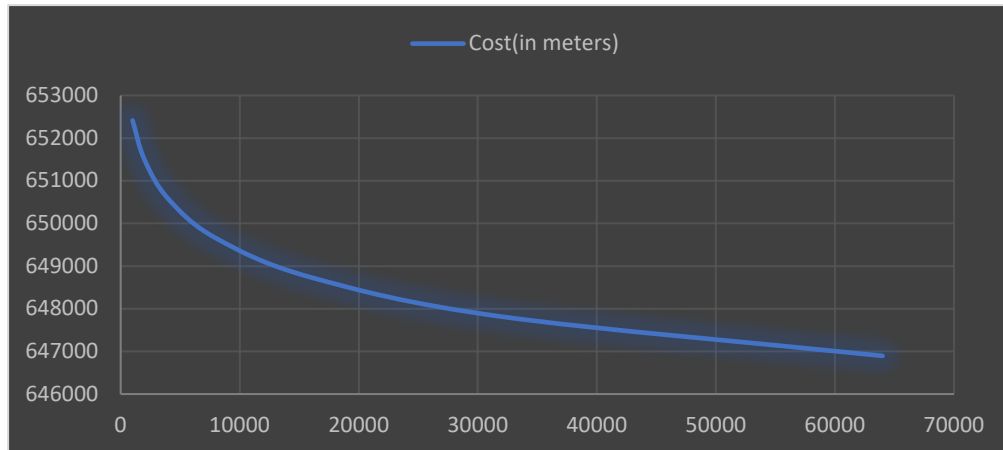
We benchmarked the Ant Colony Optimization algorithm by varying the number of iterations and using the following parameter values:

- Alpha ( $\alpha$ ) determines the weight given to the pheromone trail while the ant chooses the next city. We gave a value of 1.0 to alpha.
- Beta ( $\beta$ ) determines the weight given to the heuristic information while the ant chooses the next city. We gave a value of 5.0 to beta.
- Evaporation is the rate at which the pheromone trail evaporates over time. We gave a value of 0.5 for evaporation to avoid the pheromone trail from being too strong and dominating the decision of the ants, leading to suboptimal solutions.
- Q is a parameter that controls the amount of pheromone that an ant deposits on the path. We gave a value of 500 for Q."

iterations	Cost(in meters)
25	784742.7893

50	767856.6494
100	750970.5096
200	734084.3697
400	717198.2299
800	700312.09
1600	683425.9501

•

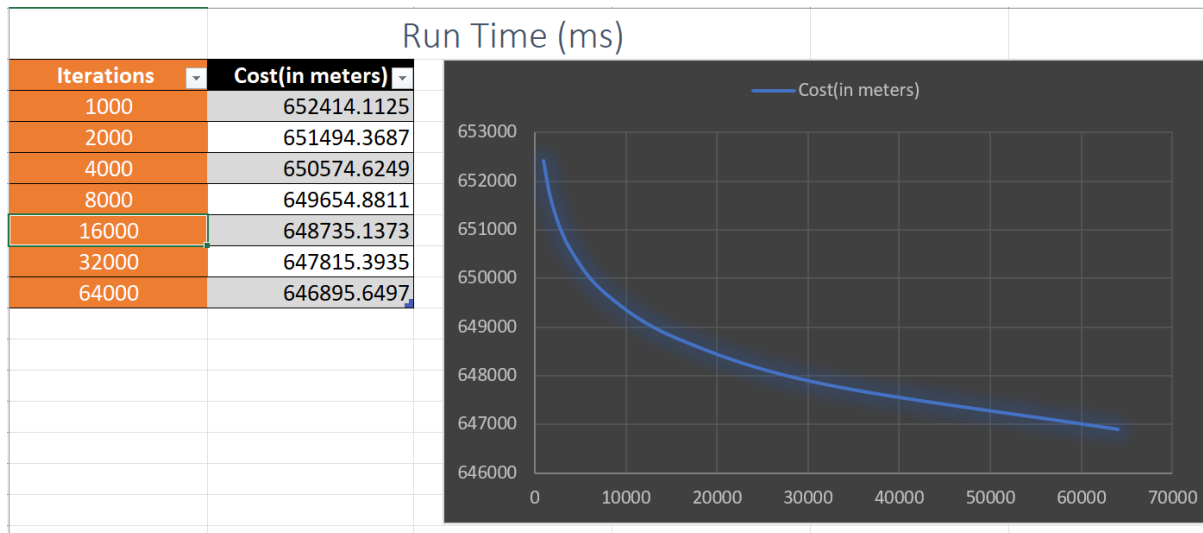


## 2 Opt:

The statement means that for 2-opt optimization techniques, the cost remains constant as the number of iterations increase, and the time complexity for this algorithm is  $O(n^2)$ .

The time complexity of an algorithm is a measure of how long it takes to run as a function of the input size. In this case,  $n$  represents the number of cities in the traveling salesman problem (TSP). The 2-opt algorithm is a simple local search algorithm for solving TSP that works by taking a route that crosses over itself and reordering it so that it does not. The basic move of this algorithm was first proposed by Flood, and later Croes proposed the 2-opt algorithm in 1958.

The time complexity of the 2-opt algorithm is  $O(n^2)$  because it needs to check all combinations of two arcs per iteration. This means that if there are  $n$  cities, then there will be  $n(n-1)/2$  possible arcs to check. Therefore, as  $n$  increases, the time complexity increases quadratically.



## Results & Mathematical Analysis:

*Usage of mathematical methods for the algorithms we implemented, below explained -*

The Haversine formula is a mathematical formula used in navigation to calculate the distance between two points on the surface of a sphere, such as the Earth. It is commonly used to calculate the distance between two locations given their latitudes and longitudes. The Haversine formula is derived from the law of haversines, which states that for a triangle on the surface of a sphere, the haversine of each angle is equal to the haversine of the sum of the other two angles minus the product of the haversines of those angles<sup>1</sup>.

The formula uses the arcsin function to calculate the angle between two points, and the trigonometric functions sin and cos to convert latitudes and longitudes to angles<sup>1</sup>. The formula assumes that the Earth is a perfect sphere, which is not entirely accurate, but is generally considered to be accurate enough for most applications<sup>2</sup>.

The Haversine formula is given by:  $d = 2r \arcsin(\sqrt{\sin^2((\text{lat2}-\text{lat1})/2) + \cos(\text{lat1}) * \cos(\text{lat2}) * \sin^2((\text{lon2}-\text{lon1})/2)})$  where:  $d$  is the distance between two points (in kilometers or miles)  $r$  is the radius of Earth (in kilometers or miles)  $\text{lat1}$  and  $\text{lat2}$  are latitudes of two points (in degrees)  $\text{lon1}$  and  $\text{lon2}$  are longitudes of two points (in degrees).

Simulated Annealing and Ant Colony Optimization are two optimization techniques used to get the optimum TSP solution. Simulated Annealing is an algorithm that starts with an initial solution and iteratively generates new candidate solutions by making small random changes to the current solution. The algorithm then evaluates the objective function for each candidate solution and decides whether to accept or reject the new solution based on a probability distribution. The Metropolis-Hastings algorithm determines the probability distribution for simulated annealing, which is based on the difference between the objective function values of the current and alternative solutions as well as a temperature parameter that drops over time in accordance with a cooling schedule.

Ant Colony Optimization is another optimization technique used to get the optimum TSP solution. The pheromone trail levels for each edge in the graph are calculated as part of the ACO algorithm and updated based on the effectiveness of the ants' solutions. The pheromone level and a heuristic value, which indicates the edge's desirability based on parameters like its length or cost, are then combined to estimate the likelihood that an ant will choose a certain edge. The formula appears as follows:  $p(i,j)$  is equal to  $\frac{[\tau(i,j)]^\alpha}{[\tau(i,j)]^\alpha + [\eta(i,j)]^\beta}$ , where  $p(i,j)$  is the likelihood that an ant would move from node  $i$  to node  $j$ ,  $\tau(i,j)$  is the pheromone level on  $t$ .

Graph Problems	Total Cost	Algorithm
Minimum Spanning Tree	513326.095	Prim's Algorithm
TSP	877,057.062	Christofides

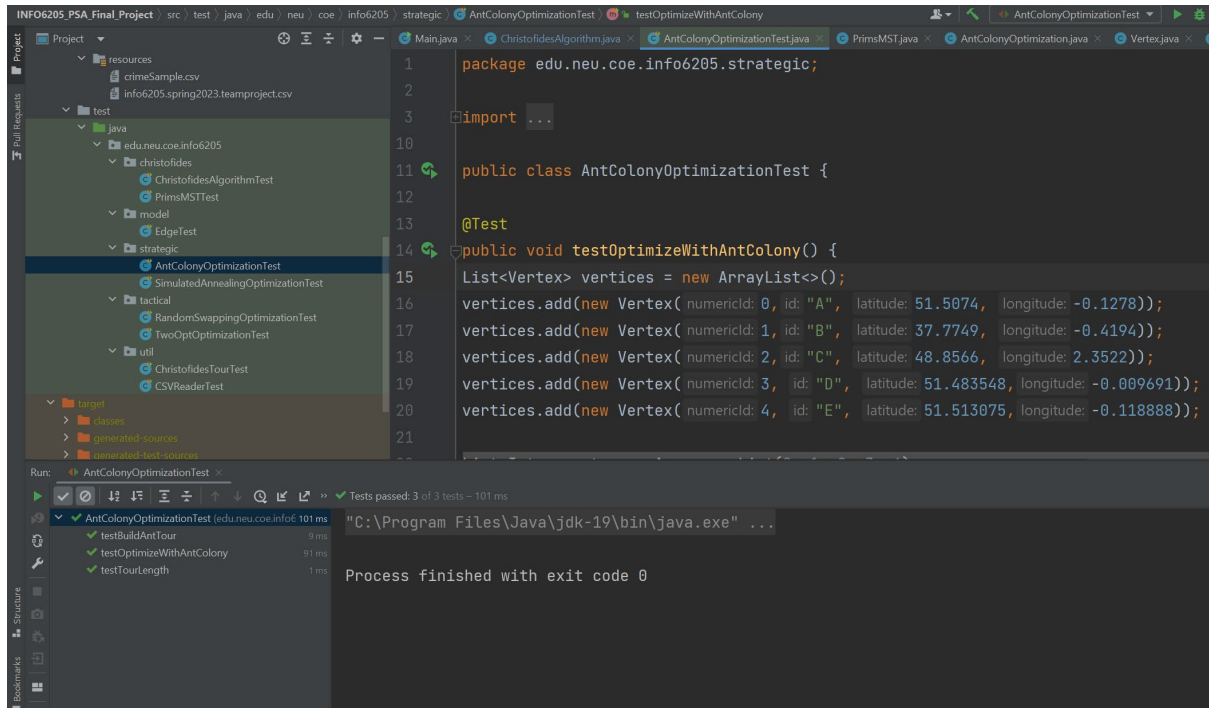
Optimization Method	Total Cost	Execution Time (ms)	Type of Optimization
TwoOpt	663,503	440ms	Tactical
Random Swapping	872,298	911ms	Tactical
Simulated Annealing	663,503	46307ms	Strategic
Ant Colony	683,425	92326ms	Strategic

Optimization Method	Ratio (wrt MST)	% increase (wrt MST)	Type of Optimization	Characteristics
TwoOpt	1.29	22.63%	Tactical	Local search; pairwise exchanges of edges; less computationally intensive
RandomSwapping	1.69	41.15%	Tactical	Local search; triple wise exchanges of edges; more complex than TwoOpt
Simulated Annealing	1.29	22.63 %	Strategic	Global search; probabilistic acceptance of non-improving solutions; uses temperature parameter
Ant Colony	1.33	24.89%	Strategic	Global search; inspired by ant behavior; pheromone-based path updates

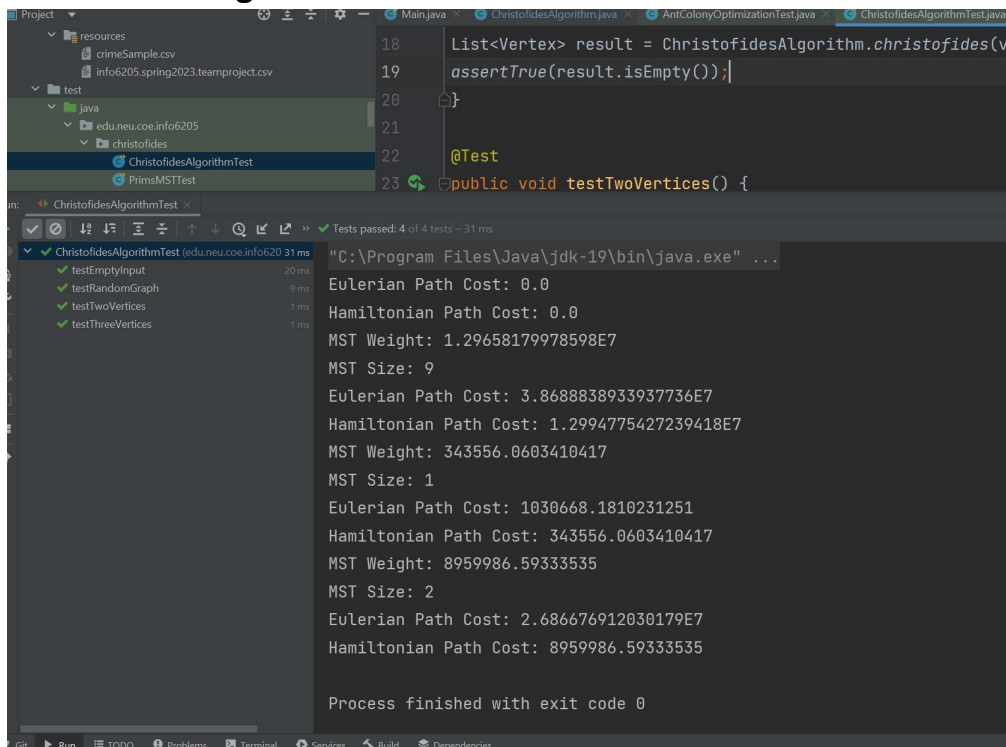


# Unit tests

## 1.Ant-colony



## 2.ChristofideAlgorithm Test



### 3. PrimsMSTTest

```
18 List<Vertex> result = ChristofidesAlgorithm.christofides(v
19 assertTrue(result.isEmpty());
20 }
21
22 @Test
23 public void testTwoVertices() {
```

Run: ChristofidesAlgorithmTest

Tests passed: 4 of 4 tests - 31 ms

- ✓ ChristofidesAlgorithmTest (edu.neu.coe.info620.31 ms)
- ✓ testEmptyInput 20 ms
- ✓ testRandomGraph 9 ms
- ✓ testTwoVertices 1 ms
- ✓ testThreeVertices 1 ms

"C:\Program Files\Java\jdk-19\bin\java.exe" ...

Eulerian Path Cost: 0.0  
Hamiltonian Path Cost: 0.0  
MST Weight: 1.29658179978598E7  
MST Size: 9  
Eulerian Path Cost: 3.8688838933937736E7  
Hamiltonian Path Cost: 1.2994775427239418E7  
MST Weight: 343556.0603410417  
MST Size: 1  
Eulerian Path Cost: 1030668.1810231251  
Hamiltonian Path Cost: 343556.0603410417  
MST Weight: 8959986.59333535  
MST Size: 2  
Eulerian Path Cost: 2.686676912030179E7  
Hamiltonian Path Cost: 8959986.59333535

Process finished with exit code 0

### 4. SimulatedAnnealingOptimization Test

```
2
3 import ...
15
16 public class SimulatedAnnealingOptimizationTest {
17
18     private List<Vertex> vertices;
19     private double initialTemperature;
20     private double coolingRate;
21
```

Run: SimulatedAnnealingOptimizationTest

Tests passed: 6 of 6 tests - 19 ms

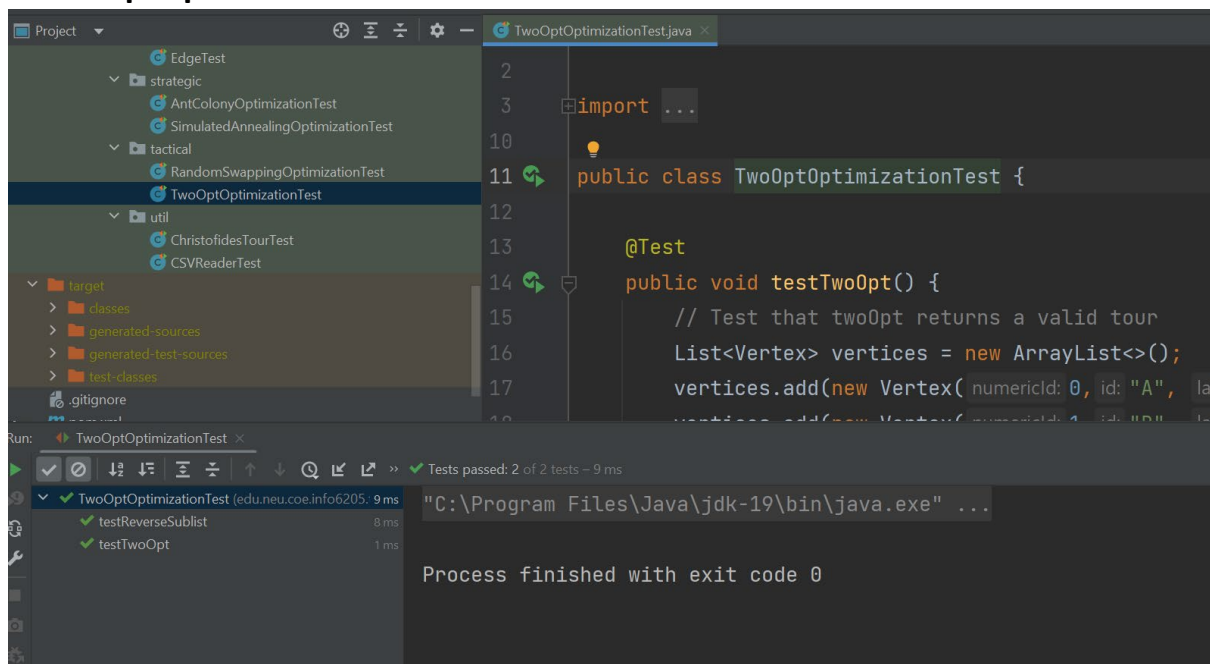
- ✓ SimulatedAnnealingOptimizationTest (edu.neu.coe.info620.19 ms)
- ✓ testOptimizeReturnsSameListIfOnlyOneVertex 11 ms
- ✓ testOptimizeReturnsSameListIfAlreadyOptimized 2 ms
- ✓ testAcceptanceProbabilityReturnsOnSelfNeighbor 1 ms
- ✓ testOptimizeReturnsDifferentOrderIfNotOptimized 1 ms
- ✓ testCalculateEnergyReturnsCorrectValue 4 ms
- ✓ testAcceptanceProbabilityReturnsCorrectValue 0 ms

"C:\Program Files\Java\jdk-19\bin\java.exe" ...

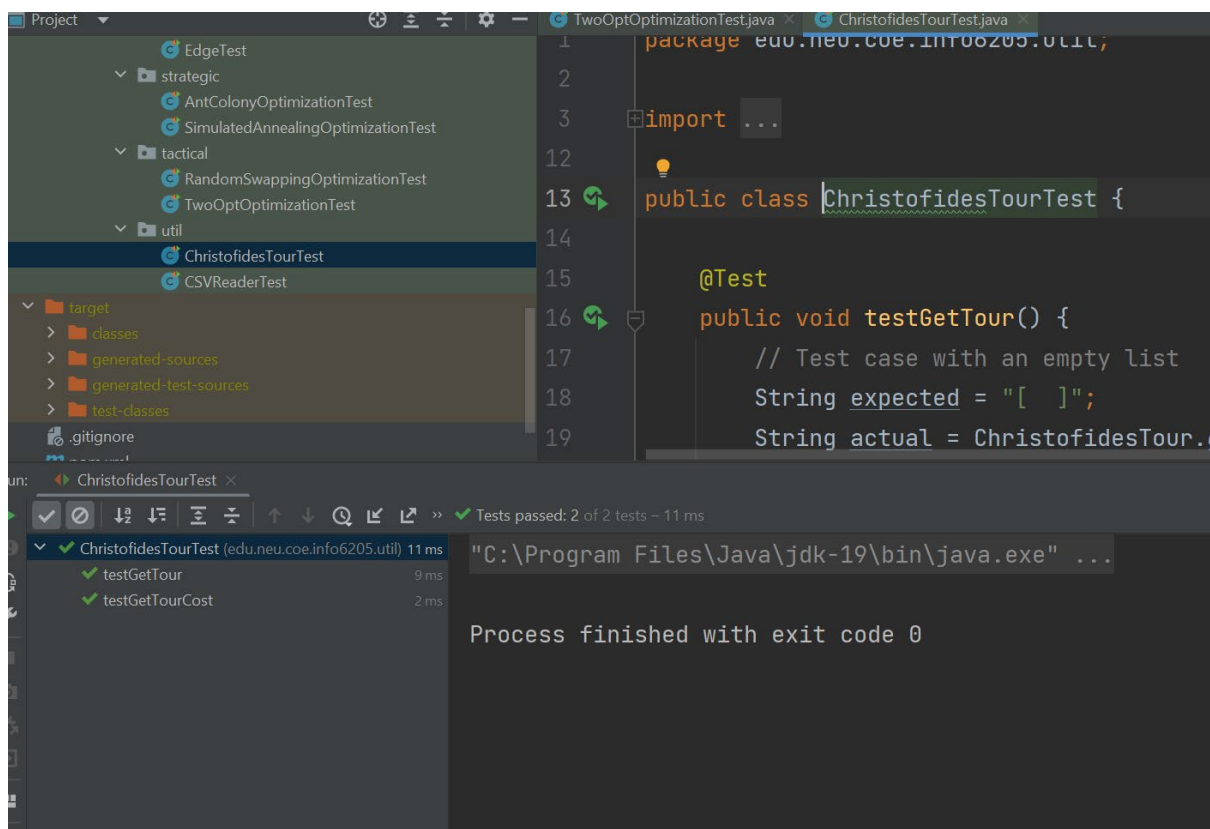
[A, C, D, B]  
[A, D, C, B]  
1597863.9968055384

Process finished with exit code 0

## 5.TwoOptOptimization



## 6.ChristofidesTourTest



## 7.CSVReaderTest

The screenshot shows an IDE with the `CSVReaderTest` class open. The class is located in the `edu.neu.coe.info6205.tactical` package. It contains a single test method `testTwoOpt()` that tests the `TwoOptOptimizationTest` class. The test method is annotated with `@Test` and `public void testTwoOpt() {`. The test method body contains a comment `// Test that twoOpt returns a valid tour`, followed by the creation of a `List<Vertex> vertices = new ArrayList<>();` and the addition of a `Vertex` object with `vertices.add(new Vertex( numericId: 0, id: "A", latitude: 51.5074, longitude: ...`.

The Run window shows the test results for `CSVReaderTest` (edu.neu.coe.info6205.util). The test passed, with 2 of 2 tests passing in 302 ms. The test results are as follows:

Test Name	Duration
testGetVertices	301 ms
testVertexCreation	1 ms

## 8.Edge Test

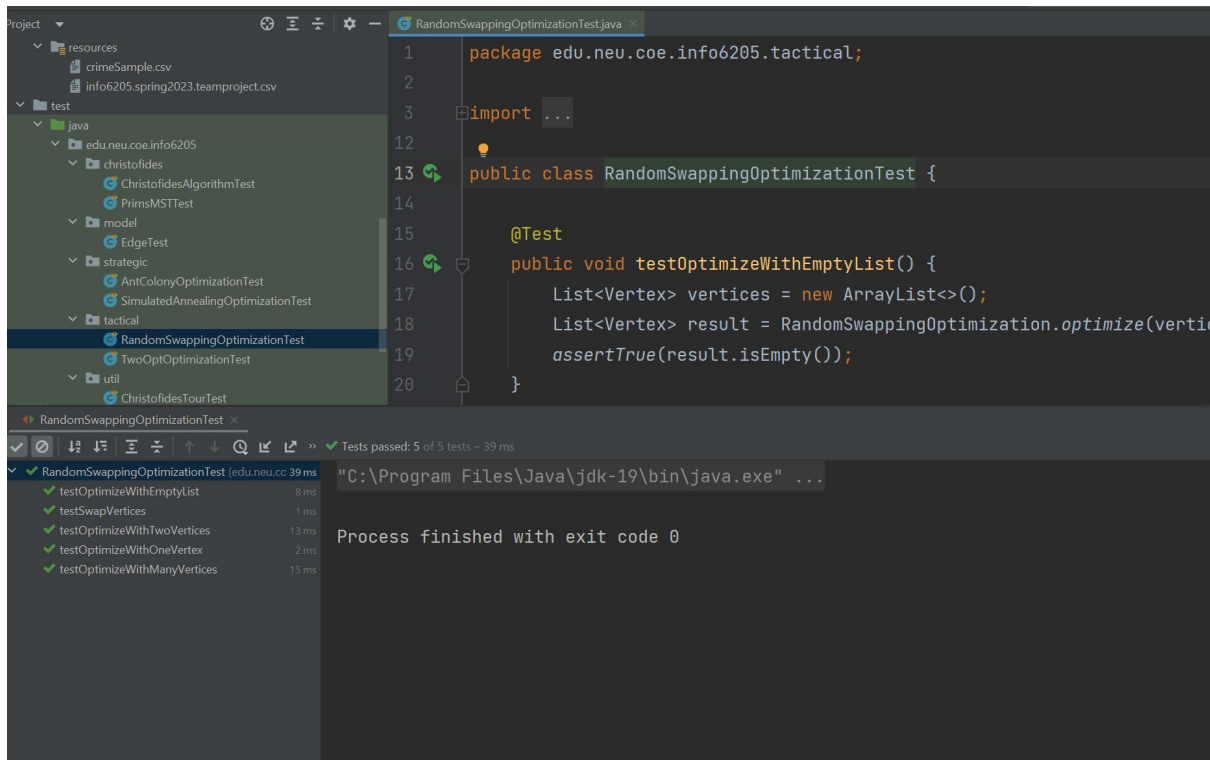
The screenshot shows an IDE with the `EdgeTest` class open. The class is located in the `edu.neu.coe.info6205.model` package. It contains a single test method `testConstructor()` that tests the `Edge` class. The test method is annotated with `@Test` and `public void testConstructor() {`. The test method body contains the creation of two `Vertex` objects (`Vertex v1 = new Vertex( numericId: 1, id: "v1", latitude: 0, longitude: 0);` and `Vertex v2 = new Vertex( numericId: 2, id: "v2", latitude: 1, longitude: 1);`), the creation of an `Edge` object (`Edge edge = new Edge(v1, v2);`), and the assertion of the `Edge` object's properties (`assertEquals(v1, edge.getV1());`, `assertEquals(v2, edge.getV2());`, `assertEquals(v1.distance(v2), edge.getWeight(), delta: 0.01);`, and `assertEquals(Arrays.asList(v1, v2), edge.getPath());`).

The Run window shows the test results for `EdgeTest` (edu.neu.coe.info6205.model). The test passed, with 2 of 2 tests passing in 9 ms. The test results are as follows:

Test Name	Duration
testCompareTo	8 ms
testConstructor	1 ms

Process finished with exit code 0

## 9. RandomSwappingOptimization test



The screenshot shows an IDE with the following components:

- Project Explorer (Left):** Displays a project structure with folders like 'resources', 'test', and 'java'. The 'tactical' folder is expanded, showing 'RandomSwappingOptimizationTest' as the selected file.
- Editor (Center):** Shows the code for 'RandomSwappingOptimizationTest.java'. The code includes a package declaration, an import statement, and a public class with a single test method: 

```
package edu.neu.coe.info6205.tactical;

import ...

public class RandomSwappingOptimizationTest {

    @Test
    public void testOptimizeWithEmptyList() {
        List<Vertex> vertices = new ArrayList<>();
        List<Vertex> result = RandomSwappingOptimization.optimize(vertices);
        assertTrue(result.isEmpty());
    }
}
```
- Test Results (Bottom Left):** A table showing the execution of 5 tests. All tests passed.

Test Name	Duration
testOptimizeWithEmptyList	8 ms
testSwapVertices	1 ms
testOptimizeWithTwoVertices	13 ms
testOptimizeWithOneVertex	2 ms
testOptimizeWithManyVertices	15 ms
- Output Console (Bottom Right):** Shows the command used to run the tests: `"C:\Program Files\Java\jdk-19\bin\java.exe" ...` and the message: `Process finished with exit code 0`.

## Conclusion

In this report, we investigated the effectiveness of the Christofides algorithm and its optimization methods in obtaining high-quality approximate solutions to the traveling salesman problem. We used tactical methods such as random swapping and 2-opt improvement, and strategic methods such as simulated annealing and ant colony optimization. Based on our findings, the Christofides method produced high-quality approximate solutions with a worst-case performance guarantee of 1.5 times the optimal solution. The addition of optimization methods such as random swapping and 2-opt improvement further improved the solution quality.

Simulated annealing and ant colony optimization were also effective in improving the solution quality but required careful parameter tuning. To ensure the program's accuracy, we used object-oriented programming principles and unit tests to develop the Christofides algorithm and its optimization methods in Java. We generated datasets of various sizes to test the performance and accuracy of the program.

We discovered that the **2-opt improvement technique** and the **simulated annealing optimization** method provided significant improvements in solution quality. This approach was able to escape from local optima and enhance solution quality while still allowing for sufficient exploration of the solution space. In conclusion, our program implementation in Java provides a practical tool for solving the problem, and future research should focus on optimizing methods and their combination to achieve even better results.

## References

[https://en.wikipedia.org/wiki/Christofides\\_algorithm](https://en.wikipedia.org/wiki/Christofides_algorithm)

[https://en.wikipedia.org/wiki/Category:Travelling\\_salesman\\_problem](https://en.wikipedia.org/wiki/Category:Travelling_salesman_problem)

[https://www.youtube.com/watch?v=GiDsjiBOVoA&ab\\_channel=Reducible](https://www.youtube.com/watch?v=GiDsjiBOVoA&ab_channel=Reducible)

<https://www.kaggle.com/datasets/marshuu/crimes-in-uk-2023?select=2023-01-metropolitan-street.csv>

LearnbyExample. (2021, May 18). Solving the Travelling Salesman Problem using Ant Colony Optimization [Video].

[https://www.youtube.com/watch?v=oXb2nC-e\\_EA&t=472s&ab\\_channel=LearnbyExample](https://www.youtube.com/watch?v=oXb2nC-e_EA&t=472s&ab_channel=LearnbyExample)