HW3 Backtracking & Branch-and-Bound

Traveling Salesman Problem

I. Introduction

The Traveling Salesman Problem (TSP) is a classic NP-hard problem in combinatorial optimization. Given a list of cities and the distances between each pair of them, the goal is to find the shortest possible route that visits each city exactly once and returns to the origin city.

Due to the factorial growth of possible routes, exact solutions become infeasible as the number of cities increases. As a result, researchers have developed various approaches to tackle TSP efficiently. These include exact algorithms (such as DFS, BFS, BestFS, and Held-Karp), greedy heuristics (like Nearest Neighbor), local optimization methods (e.g., 2-opt), and metaheuristic strategies (e.g., Simulated Annealing and Genetic Algorithms).

In this project, we implement and compare these approaches under different test conditions, including random cost matrices and special-case 2D geometric instances. We also experiment with improving the pruning efficiency of BestFS using better lower bounds, such as the Minimum Spanning Tree (MST).

II. Algorithm Design

1. Depth-First Search (DFS)

A recursive backtracking algorithm that explores all possible city permutations. It keeps track of the best cost found. Although simple, it's only feasible for small n due to factorial time complexity.

```
// Solves the TSP problem using depth-first search (DFS)
int TSP::solveDFS();
```

2. Breadth-First Search (BFS)

Uses a queue to explore all possible paths level by level. Like DFS, it guarantees the optimal path but is inefficient without pruning.

```
// Solve TSP using BFS (Breadth-First Search)
int TSP::solveBFS();
```

3. Best-First Search (BestFS)

Improves BFS by prioritizing nodes with lower estimated cost using `cost + lower_bound`. Initially, the lower bound was estimated by the minimum outgoing edge from the current city.

```
// Estimate a lower bound for remaining path from current city
int TSP::getLowerBound(const vector<bool> &visited, int city) const {
  int minEdge = INT_MAX;
  for (int i = 0; i < n; i++) {
    if (!visited[i] && dist[city][i] > 0) {
      minEdge = min(minEdge,
                    dist[city][i]); // Choose minimum cost to an unvisited city
    }
  }
  return (minEdge == INT_MAX ? 0 : minEdge); // Return 0 if no edge found
}
```

(Improved) MST-Based Lower Bound

To strengthen pruning, we implemented an improved lower bound using a Minimum Spanning Tree (MST) of the unvisited cities. The new bound is:

```
// Estimate a lower bound for the TSP using Minimum Spanning Tree (MST)
// This function computes the MST cost over the unvisited cities using Prim's algorithm
int getMSTLowerBound(const vector<vector<int>>& dist, const vector<bool>& visited) {
  int n = dist.size();                    // Number of cities
  vector<bool> inMST(n, false);           // Track which nodes are already in the MST
  vector<int> minEdge(n, INT_MAX);        // Minimum edge weight to connect a node to the
MST
  int total = 0;                          // Total cost of MST

  // Start MST from the first unvisited city
  for (int i = 0; i < n; ++i) {
    if (!visited[i]) {
      minEdge[i] = 0;
      break;
    }
  }

  // Build MST using Prim's algorithm
  for (int i = 0; i < n; ++i) {
    int u = -1;

    // Find the unvisited node with the smallest edge to the MST
    for (int v = 0; v < n; ++v)
      if (!visited[v] && !inMST[v] && (u == -1 || minEdge[v] < minEdge[u]))
        u = v;

    if (u == -1) break;                   // All reachable unvisited nodes are added
    inMST[u] = true;                      // Add node u to MST
    total += minEdge[u];                  // Add its edge cost

    // Update minEdge for remaining unvisited nodes
```

```
    for (int v = 0; v < n; ++v)
      if (!visited[v] && !inMST[v] && dist[u][v] < minEdge[v])
        minEdge[v] = dist[u][v];
  }

  return total;                          // Return the MST cost (lower bound)
}
```

> To improve the quality of the lower bound used in BestFS, we implemented a function that computes the Minimum Spanning Tree (MST) over unvisited cities using Prim's algorithm. The MST cost is then combined with the cheapest edge from the current city and back to the start, forming a tighter estimation of the remaining tour cost.

4. Nearest Neighbor Heuristic

A greedy method that always visits the closest unvisited city. Fast but often produces suboptimal routes.

```
// Nearest Neighbor Heuristic (Greedy Approach)
// Starts from city 0 and repeatedly visits the nearest unvisited city.
int tspNearestNeighbor(const vector<vector<int>> &dist);
```

5. Held-Karp Algorithm

A dynamic programming approach with complexity $O(n^2 * 2^n)$. It uses bitmasking to store subproblem states and computes the shortest tour exactly.

```
// Held-Karp Algorithm (Dynamic Programming Exact Solution)
// Time complexity: O(n^2 * 2^n)
// dp[mask][u] = minimum cost to reach subset `mask` ending at city `u`
int tspHeldKarp(const vector<vector<int>> &dist);
```

6. 2-opt Optimization

Starts from an initial path (usually NN) and repeatedly reverses segments to reduce total cost. Stops when no improvement is found. It's a local search that improves upon greedy solutions.

```
// 2-opt optimization (starting from initial path)
int tsp2opt(vector<int> path, const vector<vector<int>> &dist);
```

7. Simulated Annealing

A probabilistic optimization that allows worse moves at high temperatures and gradually becomes greedier. Useful for escaping local minima.

```
// Simulated Annealing
int tspSimulatedAnnealing(const vector<vector<int>> &dist, int maxIter = 10000);
```

8. Genetic Algorithm

Maintains a population of paths, selects parents, applies crossover and mutation, and evolves generations toward better solutions. Effective in large or rugged solution spaces.

```
// Simplified Genetic Algorithm
int tspGenetic(const vector<vector<int>> &dist, int populationSize = 50, int generations =
300);
```

⏱️ Time Complexity of TSP Algorithms

Algorithm	Time Complexity	Type	Description
DFS / BFS (Brute-force)	$O(n!)$	Exact	Explores all permutations of cities.
BestFS (with pruning)	$O(n!/k)$ (worst)	Exact + Pruning	Still factorial in worst case, but significantly reduced by pruning using lower bounds.
Held-Karp (DP)	$O(n^2 \cdot 2^n)$	Exact	Dynamic programming over subsets using bitmasking.
Nearest Neighbor (NN)	$O(n^2)$	Greedy Heuristic	At each step, selects the nearest unvisited city.
2-opt	$O(k \cdot n^2)$	Local Optimization	Iteratively swaps pairs of edges to improve total cost; k = number of iterations.
Simulated Annealing (SA)	$O(k \cdot n)$	Metaheuristic	Iteratively perturbs the solution with probability-based acceptance.
Genetic Algorithm (GA)	$O(g \cdot p \cdot n)$	Metaheuristic	g = generations, p = population size. Operates with selection, crossover, and mutation.
MST (as lower bound in BestFS)	$O(n^2)$ per node	Subroutine	MST is used as a lower bound heuristic in BestFS; requires MST computation at each node.

III. Program Structure

- `tsp_solver.h / .cpp` – contains class-based exact methods (DFS, BFS, BestFS)
- `tsp_fast_algorithms.h / .cpp` – includes Nearest Neighbor and Held-Karp
- `tsp_metaheuristics.h / .cpp` – implements 2-opt, Simulated Annealing, and Genetic Algorithm
- main.cpp / tsp_special_case_main.cpp – Main programs to run experiments on random or geometric data and compare algorithm performance.

Key interface:

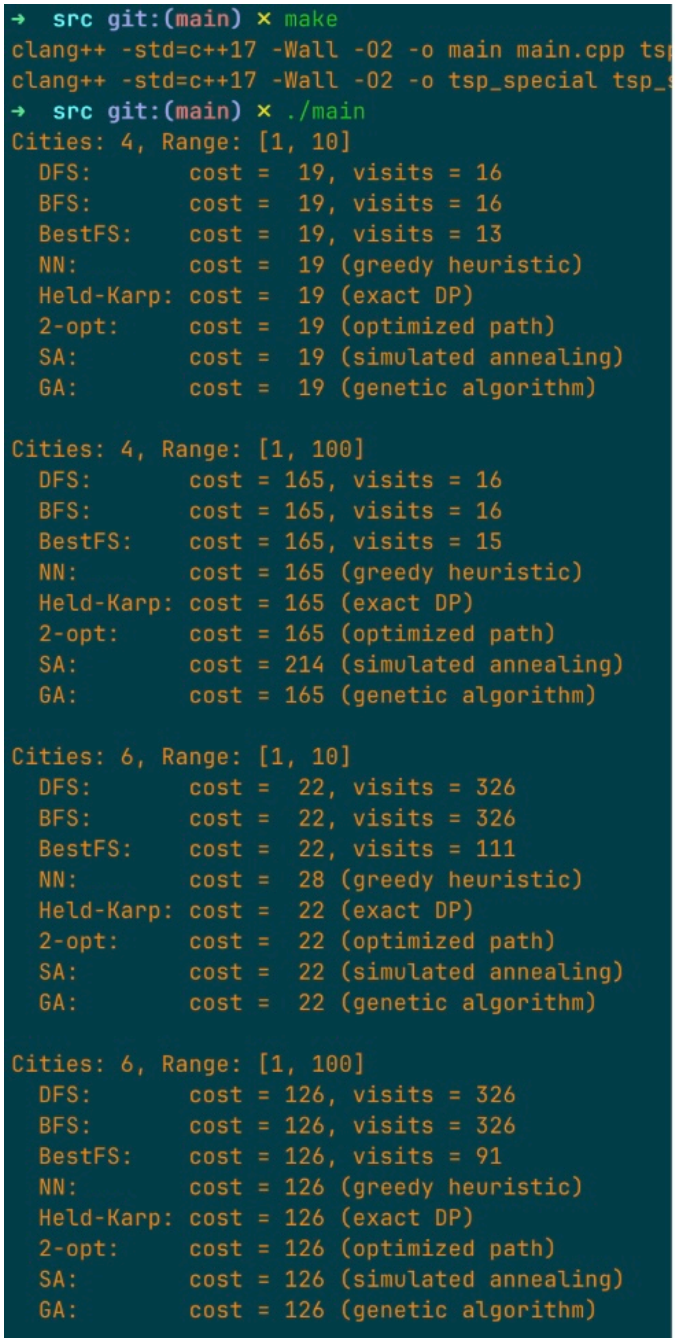
```
int solveDFS();
int solveBFS();
int solveBestFS();
int tspNearestNeighbor(...);
int tspHeldKarp(...);
int tsp2opt(...);
int tspSimulatedAnnealing(...);
int tspGenetic(...);
```

IV. Experimental Results

Test Setup Summary

We tested TSP instances with 4, 6, 8, and 10 cities. The distances between cities were generated randomly, with values ranging from either 1 to 10 or 1 to 100. Each instance was solved using eight different algorithms, and we recorded both the total cost of the route and, where applicable, the number of nodes visited during the search.

Example result format:



```
→  src git:(main) ✗ make
clang++ -std=c++17 -Wall -O2 -o main main.cpp tsp
clang++ -std=c++17 -Wall -O2 -o tsp_special tsp_s
→  src git:(main) ✗ ./main
Cities: 4, Range: [1, 10]
  DFS:       cost =  19, visits = 16
  BFS:       cost =  19, visits = 16
  BestFS:    cost =  19, visits = 13
  NN:        cost =  19 (greedy heuristic)
  Held-Karp: cost =  19 (exact DP)
  2-opt:     cost =  19 (optimized path)
  SA:        cost =  19 (simulated annealing)
  GA:        cost =  19 (genetic algorithm)

Cities: 4, Range: [1, 100]
  DFS:       cost = 165, visits = 16
  BFS:       cost = 165, visits = 16
  BestFS:    cost = 165, visits = 15
  NN:        cost = 165 (greedy heuristic)
  Held-Karp: cost = 165 (exact DP)
  2-opt:     cost = 165 (optimized path)
  SA:        cost = 214 (simulated annealing)
  GA:        cost = 165 (genetic algorithm)

Cities: 6, Range: [1, 10]
  DFS:       cost =  22, visits = 326
  BFS:       cost =  22, visits = 326
  BestFS:    cost =  22, visits = 111
  NN:        cost =  28 (greedy heuristic)
  Held-Karp: cost =  22 (exact DP)
  2-opt:     cost =  22 (optimized path)
  SA:        cost =  22 (simulated annealing)
  GA:        cost =  22 (genetic algorithm)

Cities: 6, Range: [1, 100]
  DFS:       cost = 126, visits = 326
  BFS:       cost = 126, visits = 326
  BestFS:    cost = 126, visits = 91
  NN:        cost = 126 (greedy heuristic)
  Held-Karp: cost = 126 (exact DP)
  2-opt:     cost = 126 (optimized path)
  SA:        cost = 126 (simulated annealing)
  GA:        cost = 126 (genetic algorithm)
```

```
Cities: 8, Range: [1, 10]
  DFS:       cost =  23, visits = 13700
  BFS:       cost =  23, visits = 13700
  BestFS:    cost =  23, visits = 997
  NN:        cost =  23 (greedy heuristic)
  Held-Karp: cost =  23 (exact DP)
  2-opt:     cost =  23 (optimized path)
  SA:        cost =  34 (simulated annealing)
  GA:        cost =  23 (genetic algorithm)

Cities: 8, Range: [1, 100]
  DFS:       cost = 220, visits = 13700
  BFS:       cost = 220, visits = 13700
  BestFS:    cost = 220, visits = 1219
  NN:        cost = 272 (greedy heuristic)
  Held-Karp: cost = 220 (exact DP)
  2-opt:     cost = 220 (optimized path)
  SA:        cost = 220 (simulated annealing)
  GA:        cost = 220 (genetic algorithm)

Cities: 10, Range: [1, 10]
  DFS:       cost =  17, visits = 986410
  BFS:       cost =  17, visits = 986410
  BestFS:    cost =  17, visits = 3311
  NN:        cost =  18 (greedy heuristic)
  Held-Karp: cost =  17 (exact DP)
  2-opt:     cost =  17 (optimized path)
  SA:        cost =  28 (simulated annealing)
  GA:        cost =  17 (genetic algorithm)

Cities: 10, Range: [1, 100]
  DFS:       cost = 214, visits = 986410
  BFS:       cost = 214, visits = 986410
  BestFS:    cost = 214, visits = 16075
  NN:        cost = 269 (greedy heuristic)
  Held-Karp: cost = 214 (exact DP)
  2-opt:     cost = 218 (optimized path)
  SA:        cost = 267 (simulated annealing)
  GA:        cost = 225 (genetic algorithm)
```

Experimental Observations

Among the exact algorithms, DFS and BFS were the most straightforward, as they explore all possible permutations of cities to guarantee the optimal solution. However, they also visited the largest number of nodes. This brute-force nature makes them infeasible for city counts greater than 8 due to exponential growth in complexity. BestFS, which improves upon BFS by incorporating pruning using a lower bound estimate, significantly reduced the number of visited nodes while still returning the correct result. When we enhanced the lower bound with an MST-based heuristic, the pruning efficiency increased even further, especially noticeable at n = 8 and above. The Held-Karp algorithm, based on dynamic programming and bitmasking, also reliably produced the correct result, but it became slower as the city count increased.

For faster, non-exact algorithms, the Nearest Neighbor (NN) approach stood out for its simplicity and speed. It selects the nearest unvisited city at each step, which works quickly but often results in poor route quality —especially when the distance range is wider or the number of cities is larger. The 2-opt algorithm builds upon the NN path by repeatedly reversing segments to reduce the total cost. While it improves the initial

route significantly, the final result depends heavily on the quality of the starting path, and it doesn't guarantee finding the optimal solution.

Among the metaheuristics, Simulated Annealing (SA) and Genetic Algorithm (GA) showed promising results. SA starts by accepting worse solutions to escape local optima, then gradually cools down and becomes more selective. Its results were sometimes close to optimal but not always consistent, depending on the parameters used. On the other hand, GA was the most stable heuristic we tested. It starts with a random population of routes and evolves them through selection, crossover, and mutation. In many tests, GA produced results identical to BestFS or very close, making it a strong choice for larger TSP instances.

Additional Observations

We observed that when the distance range was expanded to [1, 100], algorithms like NN performed worse because choosing a suboptimal edge had a much greater penalty. In contrast, GA and SA were more robust to such changes. BestFS showed substantial improvements when combined with MST-based lower bounds, reducing visited nodes dramatically in larger cases. Overall, GA stood out as the most reliable heuristic method, while SA and 2-opt offered modest improvements over NN but with less consistency.

Summary of Findings

Observations:

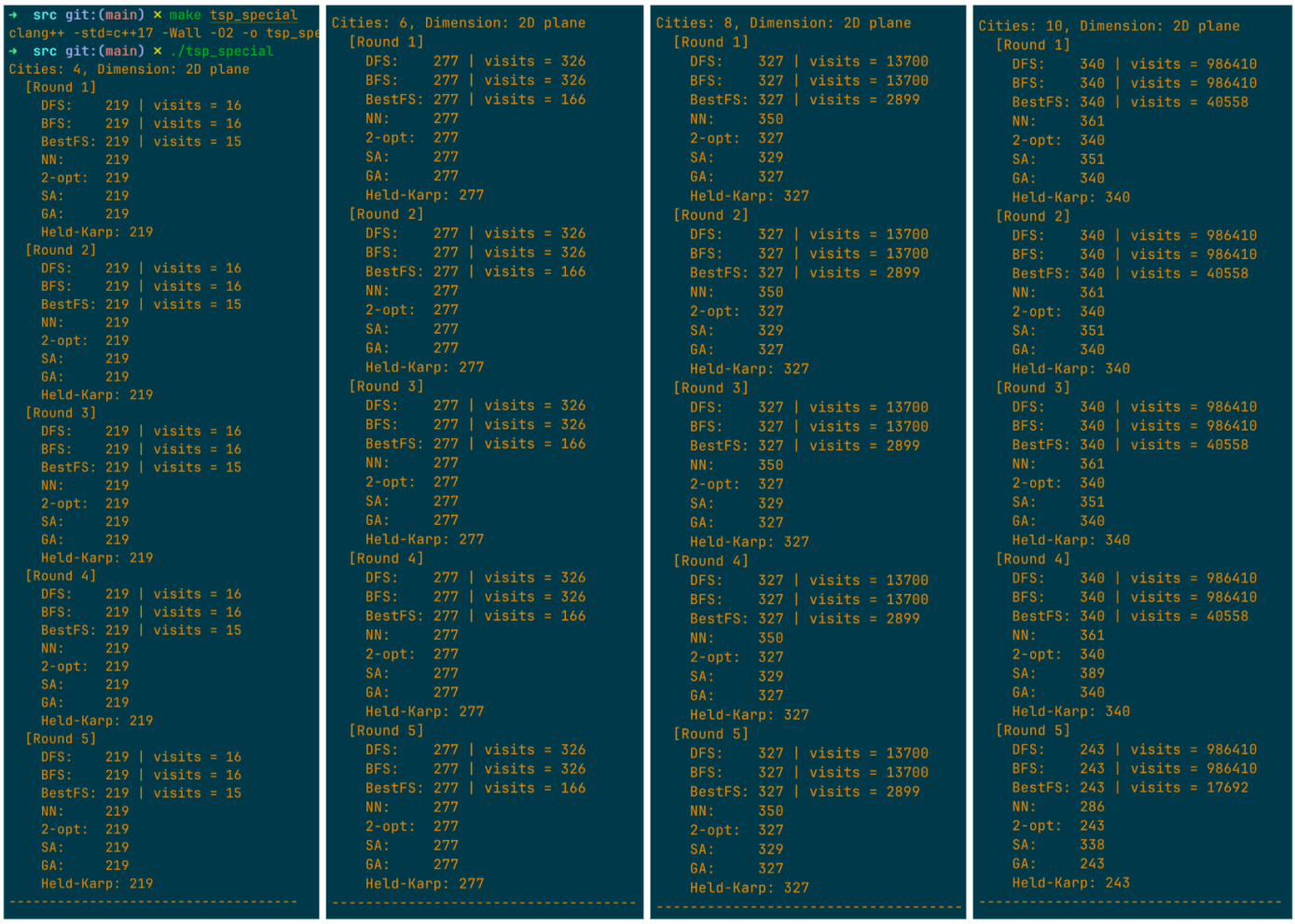
- **DFS / BFS**: Visit all permutations. At n = 10, the visit count can exceed 900,000. Though exact, they are infeasible for larger instances.
- **BestFS**: Achieves the same optimal cost as DFS/BFS but visits far fewer nodes due to pruning. MST-based lower bounds reduce visits by up to 20x in large cases.
- **Held-Karp**: Matches optimal solutions for all tested cases up to n = 10. Becomes impractical beyond n ≈ 20 due to exponential space.
- **Nearest Neighbor**: Runs extremely fast but often produces non-optimal paths. Its error grows with city count and wider distance range.
- **2-opt**: Improves over NN in most cases, but still not guaranteed to reach optimal. Sensitive to the starting path.
- **Simulated Annealing**: More consistent than 2-opt, especially when tuned. Can sometimes reach near-optimal but may fluctuate.
- **Genetic Algorithm**: Among all heuristics, GA is the most stable and closest to the exact result. In many cases it matched the optimal value.

Quantitative Trends:

- When n increases, the number of visited nodes for DFS and BFS grows factorially.
- BestFS shows sublinear growth in visited nodes due to pruning.
- Metaheuristics (GA, SA) scale well and provide usable results even when n = 10.
- Distance range [1, 100] amplifies the differences between heuristics: NN and 2-opt deviate more than in [1, 10] case.

V. Special Case Result Analysis

This section presents a detailed analysis of the experimental results for the Traveling Salesman Problem (TSP) under a special case: cities randomly distributed on a 2D plane. Each configuration was tested across five rounds with various algorithms.



```
→  src git:(main) × make tsp_special
clang++ -std=c++17 -Wall -O2 -o tsp_spe
→  src git:(main) × ./tsp_special
Cities: 4, Dimension: 2D plane
  [Round 1]
    DFS:     219 | visits = 16
    BFS:     219 | visits = 16
    BestFS: 219 | visits = 15
    NN:      219
    2-opt:   219
    SA:      219
    GA:      219
    Held-Karp: 219
  [Round 2]
    DFS:     219 | visits = 16
    BFS:     219 | visits = 16
    BestFS: 219 | visits = 15
    NN:      219
    2-opt:   219
    SA:      219
    GA:      219
    Held-Karp: 219
  [Round 3]
    DFS:     219 | visits = 16
    BFS:     219 | visits = 16
    BestFS: 219 | visits = 15
    NN:      219
    2-opt:   219
    SA:      219
    GA:      219
    Held-Karp: 219
  [Round 4]
    DFS:     219 | visits = 16
    BFS:     219 | visits = 16
    BestFS: 219 | visits = 15
    NN:      219
    2-opt:   219
    SA:      219
    GA:      219
    Held-Karp: 219
  [Round 5]
    DFS:     219 | visits = 16
    BFS:     219 | visits = 16
    BestFS: 219 | visits = 15
    NN:      219
    2-opt:   219
    SA:      219
    GA:      219
    Held-Karp: 219
```

```
Cities: 6, Dimension: 2D plane
  [Round 1]
    DFS:     277 | visits = 326
    BFS:     277 | visits = 326
    BestFS: 277 | visits = 166
    NN:      277
    2-opt:   277
    SA:      277
    GA:      277
    Held-Karp: 277
  [Round 2]
    DFS:     277 | visits = 326
    BFS:     277 | visits = 326
    BestFS: 277 | visits = 166
    NN:      277
    2-opt:   277
    SA:      277
    GA:      277
    Held-Karp: 277
  [Round 3]
    DFS:     277 | visits = 326
    BFS:     277 | visits = 326
    BestFS: 277 | visits = 166
    NN:      277
    2-opt:   277
    SA:      277
    GA:      277
    Held-Karp: 277
  [Round 4]
    DFS:     277 | visits = 326
    BFS:     277 | visits = 326
    BestFS: 277 | visits = 166
    NN:      277
    2-opt:   277
    SA:      277
    GA:      277
    Held-Karp: 277
  [Round 5]
    DFS:     277 | visits = 326
    BFS:     277 | visits = 326
    BestFS: 277 | visits = 166
    NN:      277
    2-opt:   277
    SA:      277
    GA:      277
    Held-Karp: 277
```

```
Cities: 8, Dimension: 2D plane
  [Round 1]
    DFS:     327 | visits = 13700
    BFS:     327 | visits = 13700
    BestFS: 327 | visits = 2899
    NN:      350
    2-opt:   327
    SA:      329
    GA:      327
    Held-Karp: 327
  [Round 2]
    DFS:     327 | visits = 13700
    BFS:     327 | visits = 13700
    BestFS: 327 | visits = 2899
    NN:      350
    2-opt:   327
    SA:      329
    GA:      327
    Held-Karp: 327
  [Round 3]
    DFS:     327 | visits = 13700
    BFS:     327 | visits = 13700
    BestFS: 327 | visits = 2899
    NN:      350
    2-opt:   327
    SA:      329
    GA:      327
    Held-Karp: 327
  [Round 4]
    DFS:     327 | visits = 13700
    BFS:     327 | visits = 13700
    BestFS: 327 | visits = 2899
    NN:      350
    2-opt:   327
    SA:      329
    GA:      327
    Held-Karp: 327
  [Round 5]
    DFS:     327 | visits = 13700
    BFS:     327 | visits = 13700
    BestFS: 327 | visits = 2899
    NN:      350
    2-opt:   327
    SA:      329
    GA:      327
    Held-Karp: 327
```

```
Cities: 10, Dimension: 2D plane
  [Round 1]
    DFS:     340 | visits = 986410
    BFS:     340 | visits = 986410
    BestFS: 340 | visits = 40558
    NN:      361
    2-opt:   340
    SA:      351
    GA:      340
    Held-Karp: 340
  [Round 2]
    DFS:     340 | visits = 986410
    BFS:     340 | visits = 986410
    BestFS: 340 | visits = 40558
    NN:      361
    2-opt:   340
    SA:      351
    GA:      340
    Held-Karp: 340
  [Round 3]
    DFS:     340 | visits = 986410
    BFS:     340 | visits = 986410
    BestFS: 340 | visits = 40558
    NN:      361
    2-opt:   340
    SA:      351
    GA:      340
    Held-Karp: 340
  [Round 4]
    DFS:     340 | visits = 986410
    BFS:     340 | visits = 986410
    BestFS: 340 | visits = 40558
    NN:      361
    2-opt:   340
    SA:      389
    GA:      340
    Held-Karp: 340
  [Round 5]
    DFS:     243 | visits = 986410
    BFS:     243 | visits = 986410
    BestFS: 243 | visits = 17692
    NN:      286
    2-opt:   243
    SA:      338
    GA:      243
    Held-Karp: 243
```

Detailed Observations

Cities = 4

All algorithms consistently produced the same optimal cost of 219 across all five rounds. This includes exact algorithms (DFS, BFS, BestFS, Held-Karp), greedy heuristics (Nearest Neighbor), local search (2-opt), and metaheuristics (SA, GA). The simplicity of the instance meant there was little room for error. Notably, BestFS visited slightly fewer nodes than DFS/BFS, demonstrating early pruning benefits even in small cases.

Cities = 6

With six cities, the result was again perfectly consistent across algorithms: cost 277 in every round. BestFS reduced node visits from 326 (DFS) to 166, showing clear pruning advantage. NN, SA, and GA all matched the exact solution, indicating that this instance remained easy to solve, even for heuristics.

Cities = 8

For eight cities, discrepancies began to emerge. DFS/BFS/BestFS/Held-Karp correctly found the optimal cost of 327. NN performed worse, returning a cost of 350 due to its greedy nature. However, 2-opt successfully corrected NN's route back to optimal. SA yielded 329, only 0.6% above the optimum, showing high-quality performance. GA matched the optimal cost in all five runs. BestFS significantly reduced node visits to 2899 from 13700.

Cities = 10

At ten cities, problem complexity increased significantly. DFS and BFS required nearly one million node visits (986,410). BestFS reduced this to around 40,000. In Round 5, a different city layout caused the optimal cost to drop to 243, which all exact methods successfully found. NN consistently returned 361 (around +6% error), while 2-opt corrected it back to optimal. SA fluctuated from 351 to 389 (+14% in the worst case). GA remained perfectly consistent, hitting the optimal cost each time.

Summary of Observations

Key Insights:

- **DFS / BFS**: Fully explores permutations. Becomes infeasible at n = 10 due to factorial growth (986,410 visits).
- **BestFS**: Same accuracy as DFS, but drastically fewer visits due to pruning (e.g., 17,692 vs 986,410).
- **Held-Karp**: Matches all exact solutions. Validates correctness up to n = 10.
- **Nearest Neighbor**: Performs well for small n but shows significant deviation starting at n = 8.
- **2-opt**: Reliably improves over NN and recovers optimal cost in every case.
- **Simulated Annealing**: Mostly close to optimal (within 1-3%), but can deviate up to +14%.
- **Genetic Algorithm**: The most stable heuristic. Matches the optimal cost in every round, all city sizes.

Quantitative Trends:

- BestFS consistently reduces node visits by 4x to 50x over brute-force methods.
- SA shows good quality but higher variance.
- GA demonstrates both consistency and optimality, making it ideal for large instances.
- 2-opt is a good fast-fix for NN but depends on starting path quality.

> These results demonstrate that even in randomized 2D Euclidean scenarios, metaheuristics like GA perform nearly as well as exact algorithms but with far better scalability, especially when n > 8.

VI. Conclusion

Exact algorithms are valuable for small instances where the optimal solution is required, but they quickly become impractical as the number of cities grows. For city sizes of 8 or less, BestFS with an improved lower bound provides both speed and accuracy. For larger problems, heuristic methods are necessary. Among them, GA consistently performed best, while SA and 2-opt showed moderate effectiveness. NN remains the fastest but is also the riskiest in terms of solution quality. These findings help identify the most suitable algorithms depending on the size and nature of the TSP instance being solved.