Global Convexity Tests for the Hamiltonian Cycle Problem

Evolutionary Computation Laboratory: Assignment 8

151927 Samuel Janas, samuel.janas@student.put.poznan.pl
151955 Bruno Urbaniak, bruno.urbaniak@student.put.poznan.pl
151960 Patryk Maciejewski, patryk.maciejewski.1@student.put.poznan.pl

1 Introduction

In this assignment, we investigate the **global convexity** of the Hamiltonian Cycle Problem (HCP) search landscape. The concept of global convexity relates the similarity of solutions to their fitness values. If a landscape is globally convex, we would expect that as we move closer to an optimal solution (in terms of similarity), the fitness improves in a more predictable manner.

As a reference for generating high-quality solutions, we employed the **Large Neighborhood Search (LNS) with Local Search (LS)** method described and evaluated in **Assignment 7**. In that earlier work, we found that LNS combined with local search yielded some of the best results among all tested approaches (including MSLS, ILS, and various greedy heuristics). Thus, we use the best solutions obtained from LNS-LS as a benchmark to compare other local optima against. This allows us to interpret similarity to these top-quality solutions in the context of global convexity analysis.

To test global convexity, we:

1. Generate 1000 random local optima obtained by applying a greedy local search to 1000 random initial solutions.
2. Measure each local optimum's similarity to either:
   - The best local optimum found among the 1000 runs (excluding the best one itself when plotting).
   - The average similarity to all other local optima.
3. Consider two similarity measures:
   - The number of common edges.
   - The number of common selected nodes.
4. Plot the relationship between fitness (objective function value) and similarity for two instances of the problem (TSPA and TSPB).

This results in 8 charts in total:

- 2 instances (TSPA, TSPB)
- 2 versions of similarity (to the best solution, and average similarity to all others)
- 2 similarity measures (common edges, common nodes)

We also compute and report the correlation coefficients between fitness and similarity for each chart.

2 Problem Description and Implementation

We deal with a variant of the Hamiltonian Cycle Problem, where the goal is to select a subset (e.g., 50%) of nodes to form a cycle that minimizes the combined objective of node costs and edge distances.

In **Assignment 7**, we explored the Large Neighborhood Search (LNS) approach, which works as follows:

1. Generate an initial solution x and optionally improve it with local search.
2. Repeat until a stopping condition:
   - $y := Destroy(x)$ by removing a relatively large fraction (20%) of edges or nodes.
   - $y := Repair(y)$ using a strong greedy heuristic.
   - $y := LocalSearch(y)$.
   - If $f(y) > f(x)$, then $x := y$.

The LNS variant with local search was found to produce some of the best solutions. These top solutions form the reference baseline for our global convexity analysis here.

2.1 Global Convexity Concept

If the landscape is globally convex, solutions that are structurally similar to a known good (or best) solution tend to have better fitness on average. Thus, a negative correlation between similarity and fitness supports the notion of global convexity.

2.2 Parameters

- **Number of local optima generated**: 1000
- **Local search method for generating local optima**: Greedy local search from random initial solutions.
- **Instances**: TSPA and TSPB.
- **Similarity measures**:
  - Common edges.
  - Common selected nodes.
- **Similarity references**:
  - Best local optimum (obtained from the pool of 1000 local optima).
  - Average similarity to all other local optima.

3 Methodology

3.1 Procedure

1. **Generate Solutions:** For each instance (TSPA, TSPB), generate 1000 random solutions.
2. **Local Search:** Apply a greedy local search to each random solution to obtain a local optimum.
3. **Collect Local Optima:** Store the 1000 local optima and their fitness values.
4. **Identify the Best Local Optimum:** Find the local optimum with the best (lowest) fitness, which is compared against others to measure global convexity.
5. **Calculate Similarities:** For each local optimum:
   - Compute similarity to the best local optimum (in terms of edges and nodes).
   - Compute average similarity to all other local optima (in terms of edges and nodes).
6. **Plot and Correlate:** Plot fitness (x-axis) vs. similarity (y-axis) and calculate Pearson correlation coefficients.

3.2 Pseudocode

Problem 1: Global Convexity Testing

1. **Input**:
   - Cost matrix $costMatrix$
   - Number of random solutions $N = 1000$
2. **Output**:
   - Arrays/lists of $(fitness, similarity)$ pairs for each scenario (common edges/nodes $\times$ best/average)
   - Correlation coefficients for each scenario
3. **Procedure**:
   (i) Initialize $localOptima \leftarrow []$
   (ii) For $i \leftarrow 1$ to N:
      - $sol \leftarrow RandomSolution(costMatrix)$
      - $localOpt \leftarrow GreedyLocalSearch(sol, costMatrix)$
      - $f \leftarrow Fitness(localOpt, costMatrix)$
      - Append $(localOpt, f)$ to $localOptima$
   (iii) $bestOpt \leftarrow \min_{x \in localOptima} Fitness(x)$
   (iv) For each opt_i in $localOptima$:
      - $f_i \leftarrow Fitness(opt_i)$
      - $sim^{edges}_{best}(i) \leftarrow SimilarityEdges(opt_i, bestOpt)$
      - $sim^{nodes}_{best}(i) \leftarrow SimilarityNodes(opt_i, bestOpt)$
      - $sim^{edges}_{avg}(i) \leftarrow \frac{1}{N-1} \sum_{j \neq i} SimilarityEdges(opt_i, opt_j)$
      - $sim^{nodes}_{avg}(i) \leftarrow \frac{1}{N-1} \sum_{j \neq i} SimilarityNodes(opt_i, opt_j)$
   (v) Exclude $bestOpt$ from the sim_{best} charts to avoid the trivial outlier.
   (vi) Calculate correlation coefficients:
      - $corr(sim^{edges}_{best}, fitness)$
      - $corr(sim^{nodes}_{best}, fitness)$
      - $corr(sim^{edges}_{avg}, fitness)$
      - $corr(sim^{nodes}_{avg}, fitness)$

4 Computational Experiment

We ran the global convexity tests on two instances: TSPA and TSPB. We generated 1000 local optima from random solutions using a greedy local search. The best local optima obtained (comparable to or inspired by the quality seen with LNS-LS from Assignment 7) served as our reference solutions.

5 Results

Below we present the correlation coefficients for each scenario. The results show a strong negative correlation between similarity and fitness, indicating a globally convex tendency of the landscape.

- **TSPA - Common Edges, Best**: $r = -0.6152$, $p \approx 4.3 \times 10^{-105}$
- **TSPA - Common Edges, Average**: $r = -0.6619$, $p \approx 4.39 \times 10^{-127}$
- **TSPA - Common Nodes, Best**: $r = -0.6298$, $p \approx 1.7 \times 10^{-111}$
- **TSPA - Common Nodes, Average**: $r = -0.5306$, $p \approx 1.08 \times 10^{-73}$
- **TSPB - Common Edges, Best**: $r = -0.6679$, $p \approx 4.61 \times 10^{-130}$
- **TSPB - Common Edges, Average**: $r = -0.7452$, $p \approx 8.31 \times 10^{-178}$
- **TSPB - Common Nodes, Best**: $r = -0.6593$, $p \approx 1.33 \times 10^{-125}$
- **TSPB - Common Nodes, Average**: $r = -0.5162$, $p \approx 3.53 \times 10^{-69}$

5.1 Visualizations

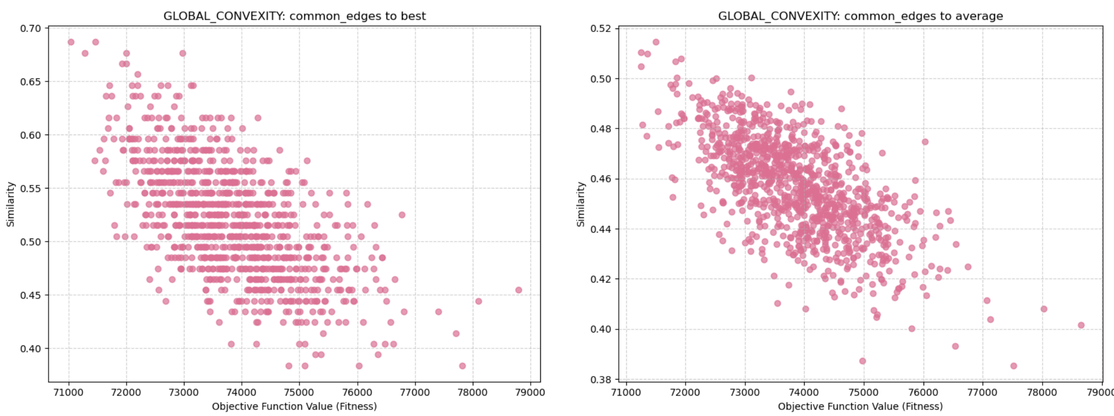

Figure 1: TSPA: Fitness vs. similarity (common edges) to best (left) and average (right).

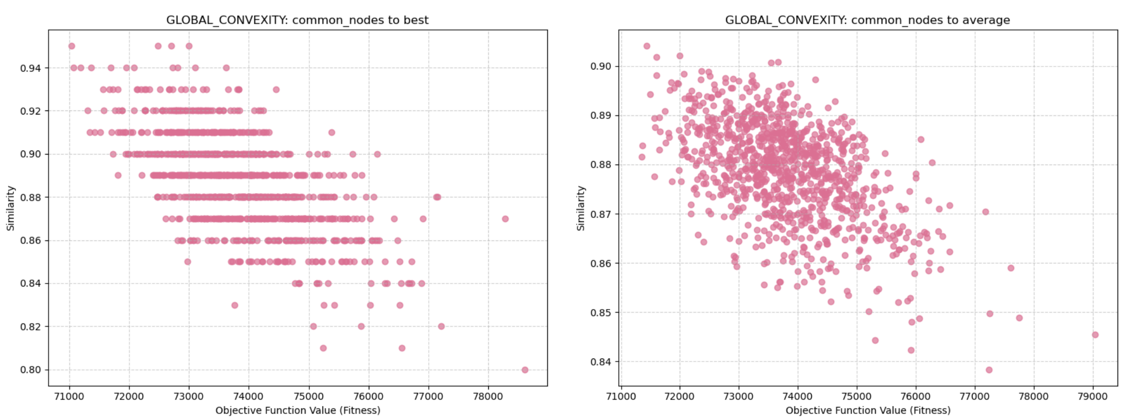

Figure 2: TSPA: Fitness vs. similarity (common nodes) to best (left) and average (right).

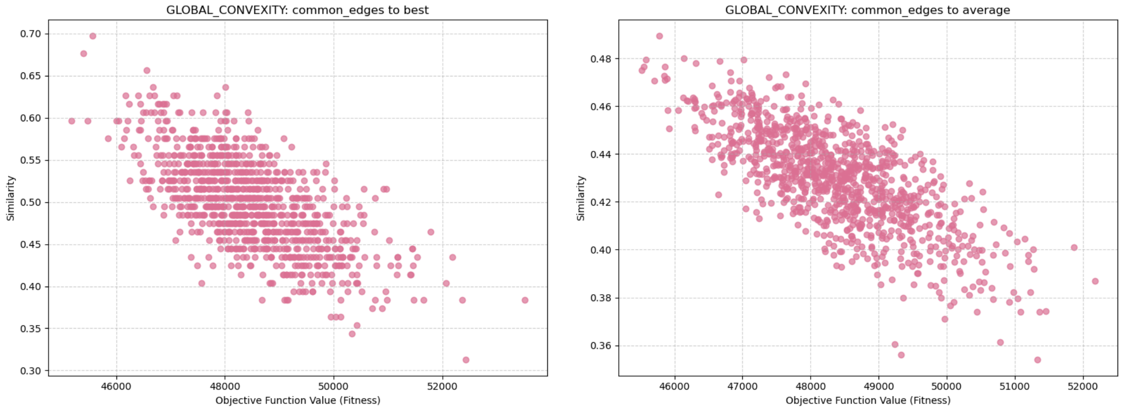

Figure 3: TSPB: Fitness vs. similarity (common edges) to best (left) and average (right).

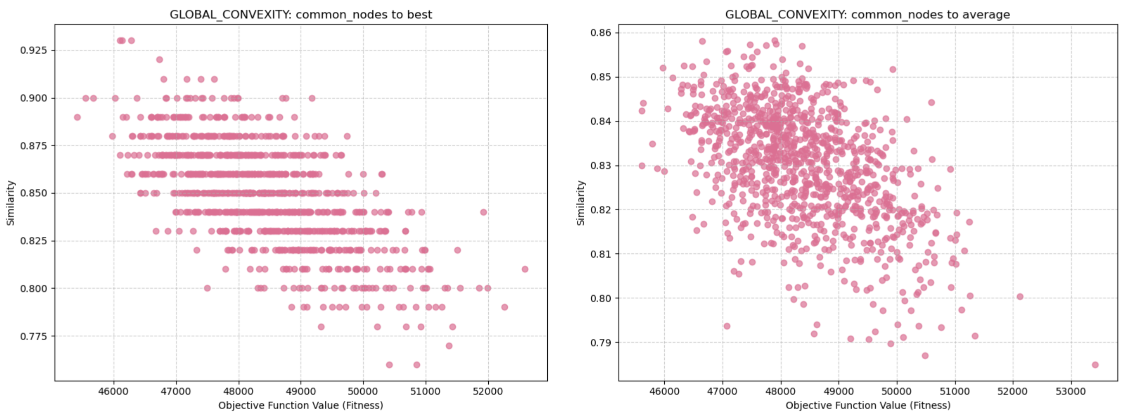

Figure 4: TSPB: Fitness vs. similarity (common nodes) to best (left) and average (right).

6 Discussion

These results demonstrate that:

- All correlations are negative, meaning that higher similarity to good solutions (best or average) is associated with lower (better) fitness.
- For TSPA, correlations range roughly from about -0.53 to -0.66, indicating a moderate to strong negative relationship.
- For TSPB, correlations are even stronger in some cases, with $r = -0.7452$ for common edges to average similarity, indicating a strong linear relationship.
- Similarity to the best solution (particularly in terms of common edges) tends to be a strong predictor of good fitness, suggesting that being structurally closer to a known best solution yields better solutions.
- Average similarity also correlates negatively with fitness, but generally not as strongly as similarity to the best solution.
- Edge-based similarity consistently shows equal or higher correlations (in absolute value) than node-based similarity. This suggests that preserving the particular edges of a high-quality cycle is more influential on fitness than just matching the set of selected nodes.

Overall, the presence of a strong negative correlation in all scenarios implies that using similarity information could guide metaheuristics towards better regions of the solution space.

7 Conclusion

The global convexity tests confirm that for the TSPA and TSPB instances, solutions that are more similar to the best local optimum or the average structure of multiple local optima tend to have better fitness. The strong negative correlations and extremely low p-values emphasize a meaningful relationship, indicating that similarity can be a valuable cue in guiding search strategies.

As a result, incorporating similarity-based heuristics or restarting strategies that push solutions toward the structure of known good solutions could improve the efficiency of metaheuristics, potentially leading to faster convergence or better solutions for the Hamiltonian Cycle Problem.