# 08-global-convexity

January 13, 2025

Made by: Andrei Kulchyk (155489) and Fiodar Piatrovich (155174)

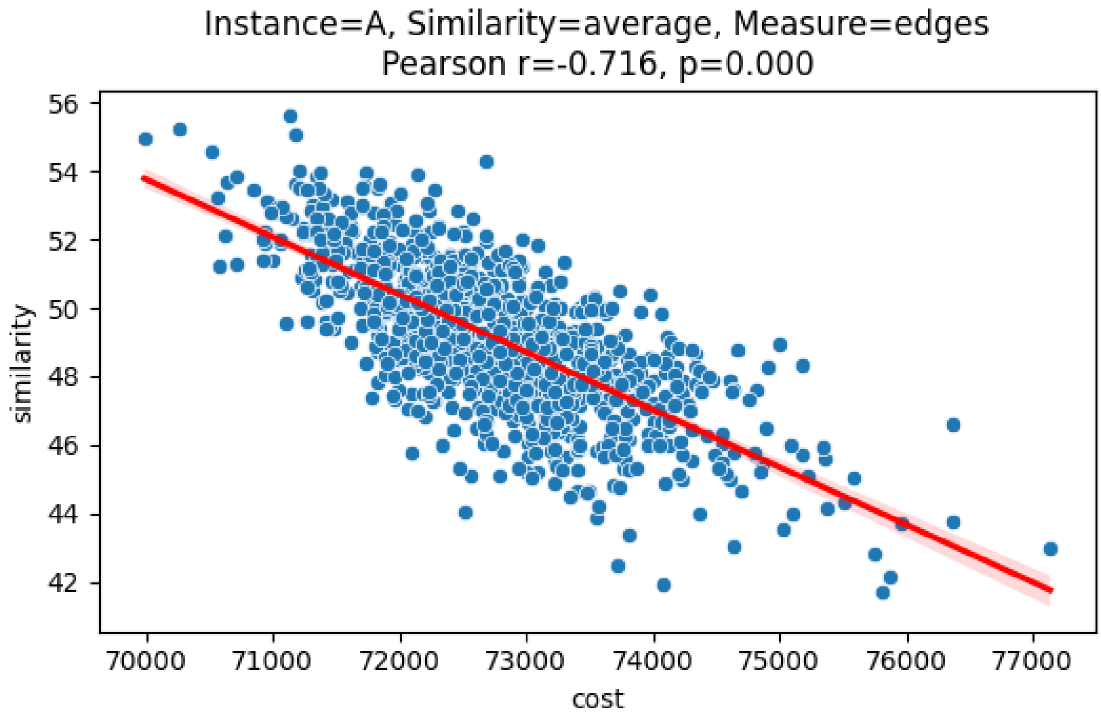
Github

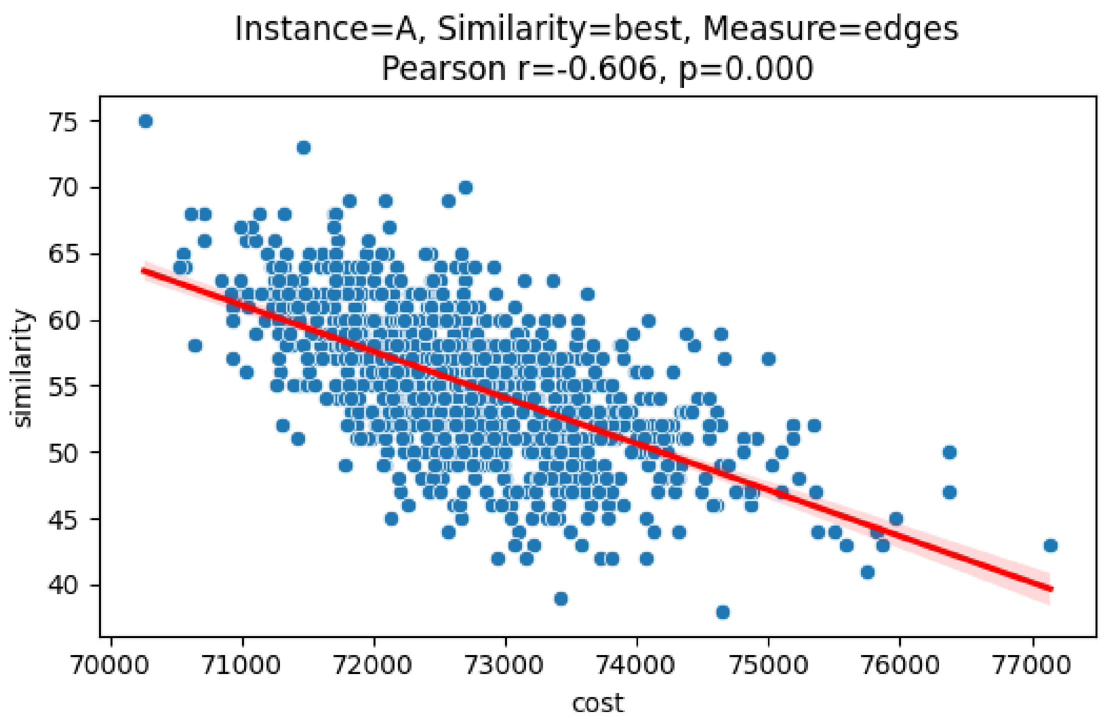
## 1 Description of a problem

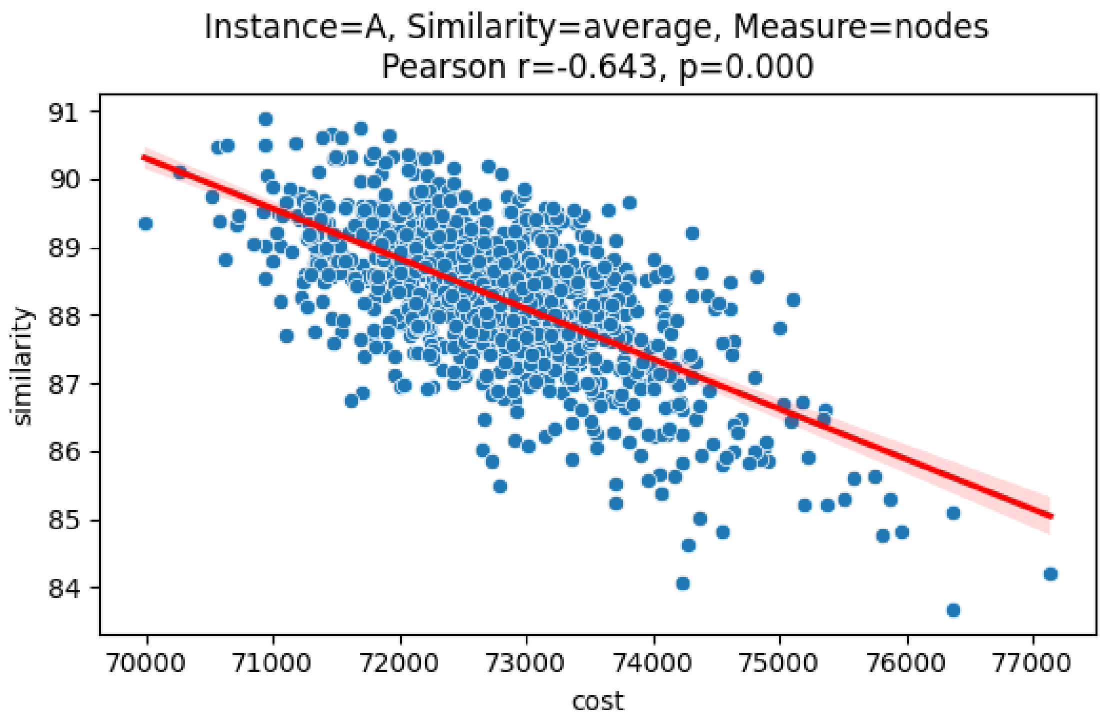
We are given three columns of integers with a row for each node. The first two columns contain x and y coordinates of the node positions in a plane. The third column contains node costs. The goal is to select exactly 50% of the nodes (if the number of nodes is odd we round the number of nodes to be selected up) and form a Hamiltonian cycle (closed path) through this set of nodes such that the sum of the total length of the path plus the total cost of the selected nodes is minimized.

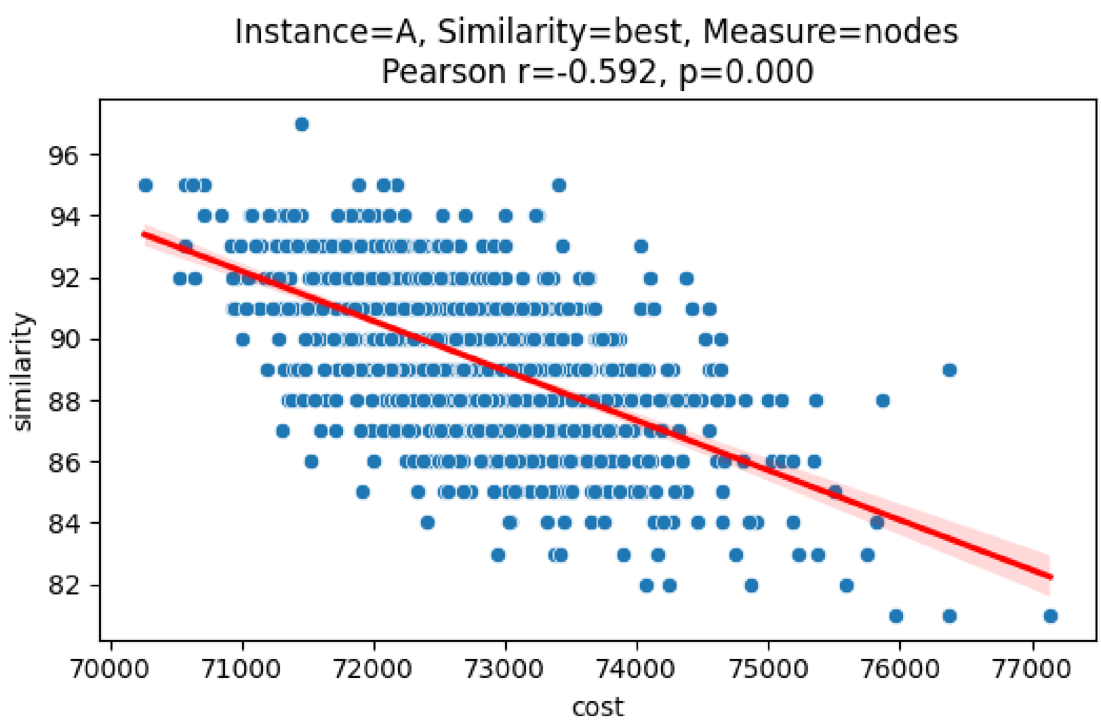
The distances between nodes are calculated as Euclidean distances rounded mathematically to integer values. The distance matrix should be calculated just after reading an instance and then only the distance matrix (no nodes coordinates) should be accessed by optimization methods to allow instances defined only by distance matrices.

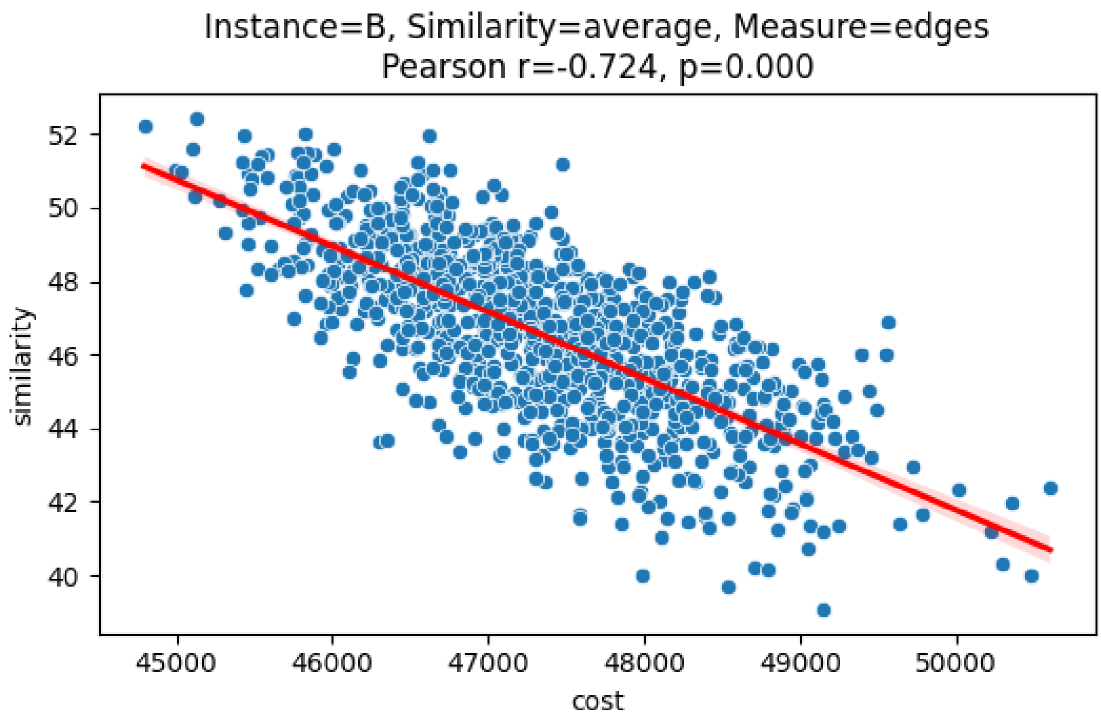
## 2 Charts

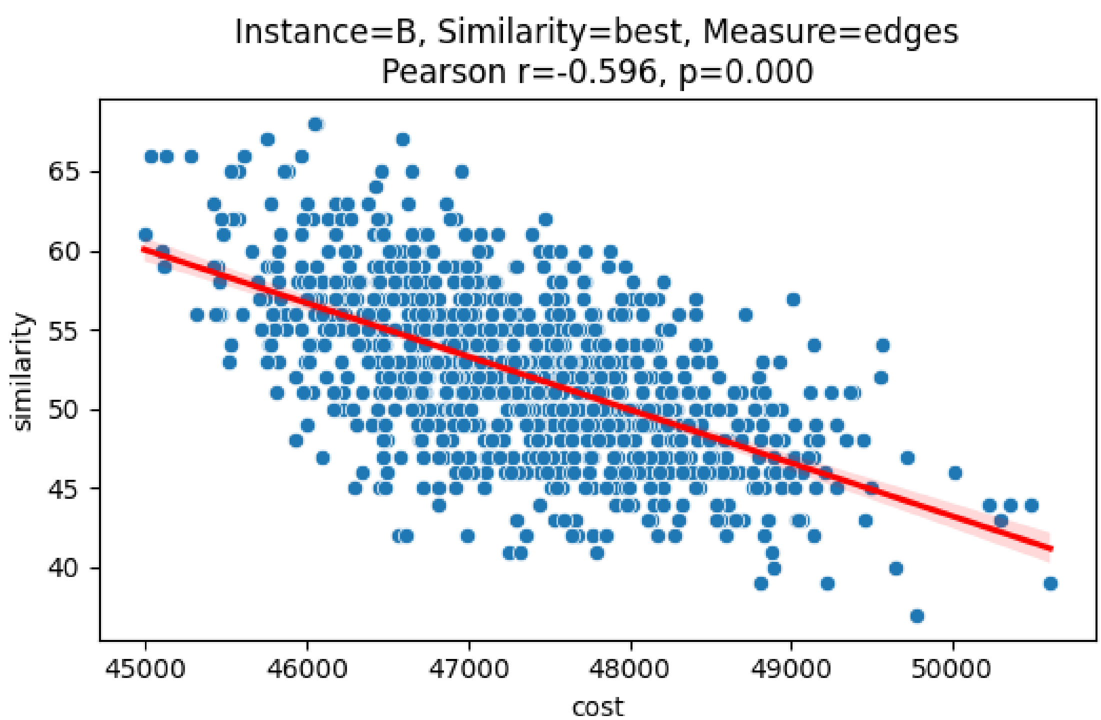

Instance=A, Similarity=average, Measure=edges
Pearson r=-0.716, p=0.000
similarity
cost
56
54
52
50
48
46
44
42
70000
71000
72000
73000
74000
75000
76000
77000

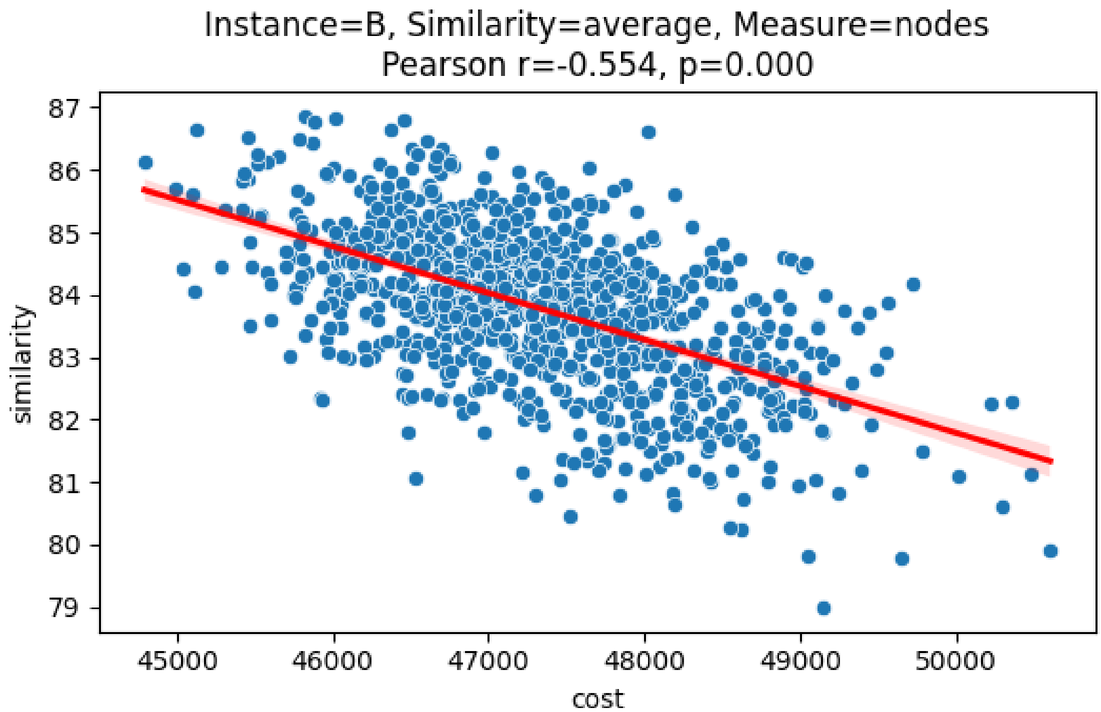

Instance=A, Similarity=best, Measure=edges
Pearson r=-0.606, p=0.000
similarity
cost
75
70
65
60
55
50
45
40
70000
71000
72000
73000
74000
75000
76000
77000

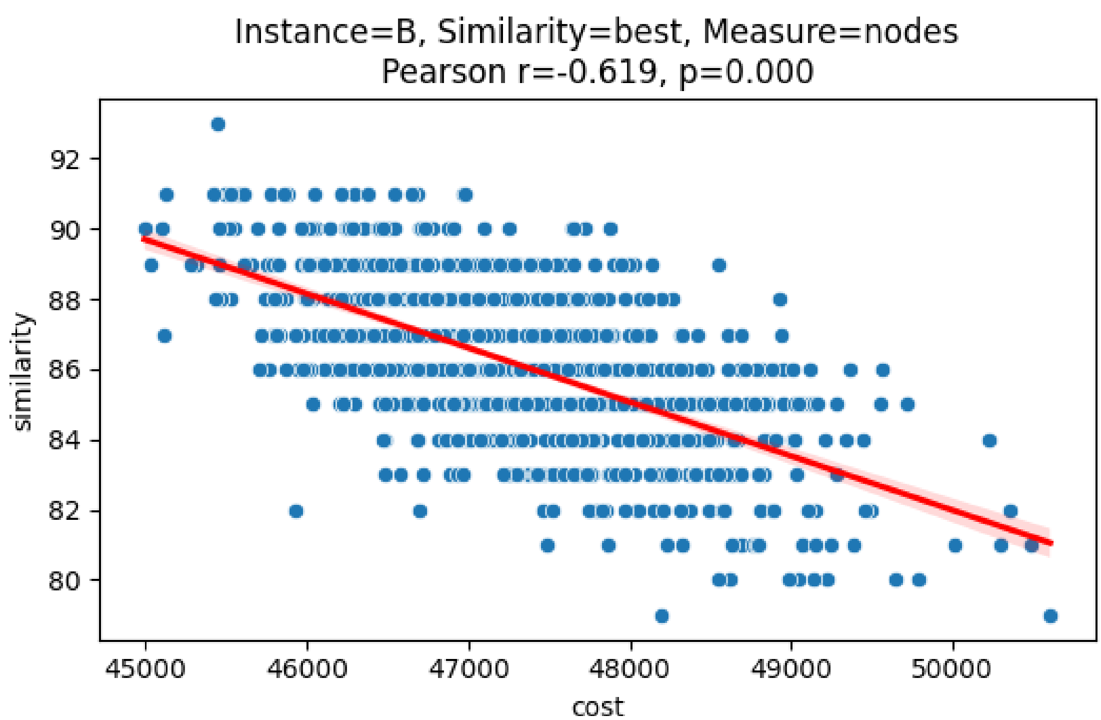

Instance=A, Similarity=average, Measure=nodes
Pearson r=-0.643, p=0.000
similarity
cost
91
90
89
88
87
86
85
84
70000
71000
72000
73000
74000
75000
76000
77000


Instance=A, Similarity=best, Measure=nodes
Pearson r=-0.592, p=0.000
similarity
cost
96
94
92
90
88
86
84
82
70000
71000
72000
73000
74000
75000
76000
77000


Instance=B, Similarity=average, Measure=edges
Pearson r=-0.724, p=0.000
similarity
cost
52
50
48
46
44
42
40
45000
46000
47000
48000
49000
50000


Instance=B, Similarity=best, Measure=edges
Pearson r=-0.596, p=0.000
similarity
cost
65
60
55
50
45
40
45000
46000
47000
48000
49000
50000


Instance=B, Similarity=average, Measure=nodes
Pearson r=-0.554, p=0.000
similarity
cost
87
86
85
84
83
82
81
80
79
45000
46000
47000
48000
49000
50000


Instance=B, Similarity=best, Measure=nodes
Pearson r=-0.619, p=0.000
similarity
cost
92
90
88
86
84
82
80
45000
46000
47000
48000
49000
50000

# 3 Conclusion

1. Correlations presented by the plots are representative, due to small `p` value emerged from statistical significance tests. In other words, 1000 runs of LS were enough.
2. There is a negative correlation between cost and similarity for all the kind of comparisons.
3. Similarity of the local optimas belong to fixed interval, depending on solution and comparison it would be [20%, 50%] of the maximal possible similarity (200). In other words, we could state that local optimas are 20% to 50% similar.
4. Node similarities are higher than edge similarity, for it's (to have similar nodes in solutions) a necessary condition for edge similarity (to have edge between them).
5. Generarly, both nodes and edges similarities are are higher when compared to the best solution.