a4-writeup

Raul Rodriguez Castro

2025-07-18

CSC 226 Algorithms and Data Structures II Assignment 4

1. Introduction

Let G be a connected-edge weighted undirected graph with no negative cycles.

Let M_u be a minimum cost path (mcp) tree for G rooted at vertex u.

Let T be a minimum spanning tree (mst) of G.

Consider the following conjecture: If G is a connected edge-weighted undirected graph, the average value of TotalWeight(M_u) is at most 2 · TotalWeight(T)

2. Methods

2.1 The Environment

All of the code was written and executed in TypeScript on a Node v.23.2.0 environment.

2.2 Generating the graphs

The conjecture was tested on 30 procedurally generated graphs. The graphs as well as the algorithm's outputs were written to a .txt and .csv files respectively to ensure reproducibility.

The graphs generated were weighted and undirected. Let $V,\ E$ be the set of vertices and edges in G respectively. For any of the generated graphs $|V|\ \leq\ |E|\ \leq\ |V|^2$. Let w be the weight of some edge $e,\ e \in E$, the range of possible values for w is $|V|\ \leq\ w\ \leq\ |V|^2$. This weight range ensures that both, graphs with multiple mst's and mcp's are produced as well as graphs with unique trees, where all edge weights are distinct.

2.3 Collecting the Data

The data was collected by running implementations of Kruskal's and Dijkstra's algorithm for the mst's and mcp's respectively.

3. Hypothesis Testing

To determine whether the conjecture holds over graphs with the properties mentioned above, a null hypothesis was formulated using the paired data. The paired data consists of the mst's and mcp's because they were run on the same set of graphs. Let μ_{M_u} be the true mean of an mcp, Let μ_T be the true mean of an mst. Let the null hypothesis be:

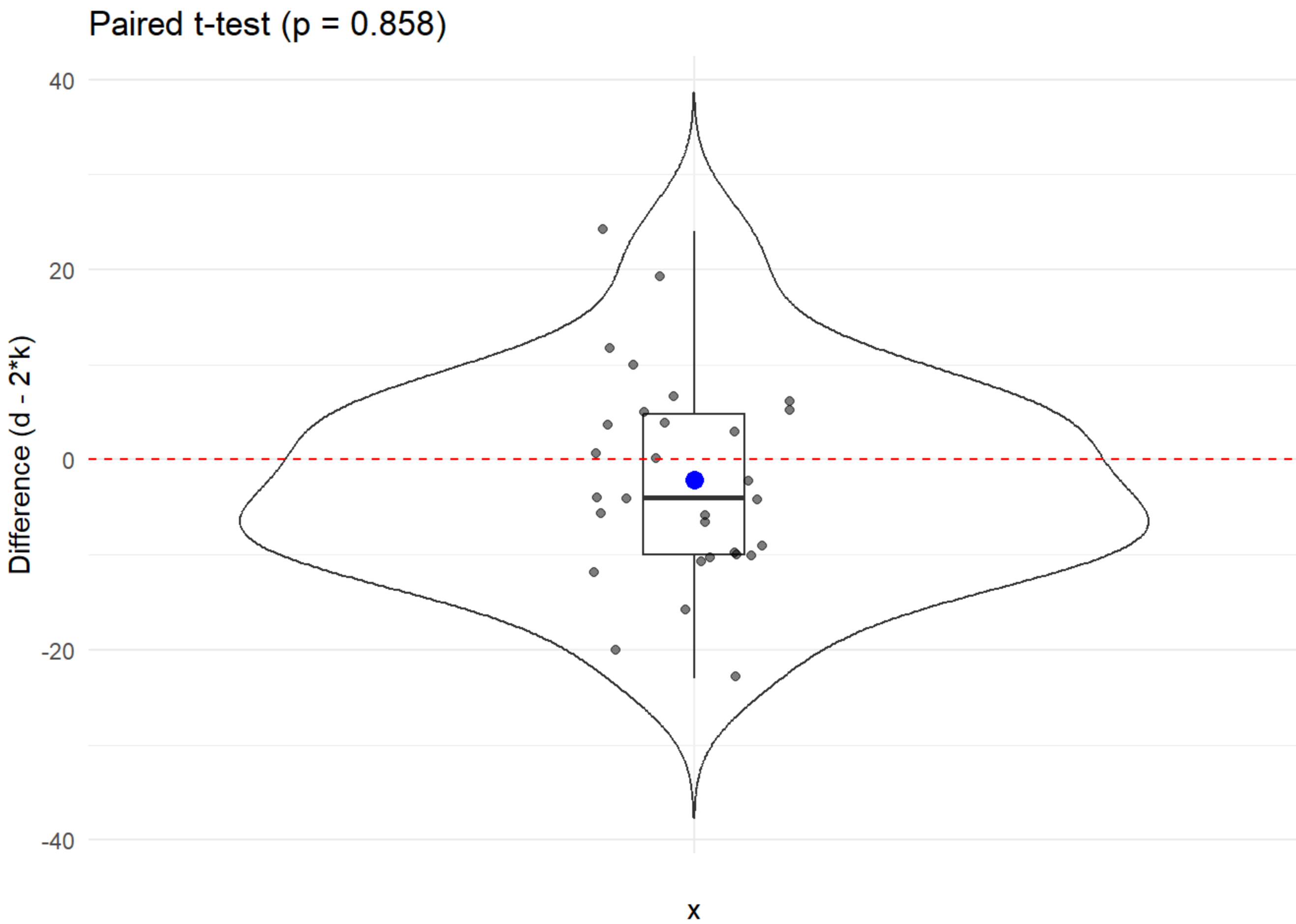
$$H_0 : \ \mu_{M_u} \ - \ 2\mu_T \leq 0$$

Since our sample space consists of observations across 30 different graphs we can apply the central limit theorem which states that the distribution of our observations converges to a standard normal random variable. This information is used to gather evidence against H_0.

Performing a paired t-test on the data suggests the mean difference was -2.13 (95% CI: [-5.46, ∞]), $t(29) = -1.09$, $p = 0.86$. The 95% confidence interval suggests the true mean difference is at least -5.46.

```
data <- read.csv("results.csv")
k <- data[[1]] # kruskal's results
d <- data[ , -1] # dijkstra's results

#difference data set
diff <- d - 2 * k
mu <- mean(diff)

# perform a t-test with d - 2*k
t_test <- t.test(d, 2*k, mu = 0, alternative = "greater", paired = TRUE)

library(ggplot2)

ggplot(data.frame(diff), aes(x = "", y = diff)) +
  geom_violin(trim = FALSE, alpha = 0.3) +
  geom_boxplot(width = 0.1, outlier.shape = NA) +
  geom_jitter(width = 0.1, alpha = 0.5) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  geom_point(aes(y = mu), color = "blue", size = 3) +
  labs(
    title = paste0("Paired t-test (p = ", signif(t_test$p.value, 3), ")"),
    y = "Difference (d - 2*k)"
  ) +
  theme_minimal()
```

```
## Warning in geom_point(aes(y = mu), color = "blue", size = 3): All aesthetics have length 1, b
ut the data has 30 rows.
## ℹ Please consider using `annotate()` or provide this layer with data containing
##   a single row.
```




```
mu
```

```
## [1] -2.133333
```

```
t_test
```

```
##
##  Paired t-test
##
## data:  d and 2 * k
## t = -1.0912, df = 29, p-value = 0.8579
## alternative hypothesis: true mean difference is greater than 0
## 95 percent confidence interval:
##  -5.455279       Inf
## sample estimates:
## mean difference
##       -2.133333
```

As we can observe from the violin-plot, most observations fall around the 0 range, having a few outliers above and below it. The violin plot outlines the tendency of the difference between an mcp and 2*mst to stay within the 0 range.

4. Discussion

This analysis aimed to test the conjecture on the relationship between the total weights of an mcp and mst. The high p-value (0.86) provides no evidence against the null hypothesis. The data suggests that the difference between the total weights of an mcp and 2*mst tends to stay close to zero which is consistent with the conjecture.

5. Conclusion

In conclusion, we fail to reject the null hypothesis. The results provide no evidence against the conjecture, consistent with the claim that TotalWeight(M_u) is at most 2 · TotalWeight(T)