

Minimum Spanning Trees in R

Read in dissimilarity matrix:

```
dm <- as.matrix(read.csv("~/Downloads/dissimilarity_matrix.csv", header = F))
```

Note that this matrix has many non-unique edges in it, thus implying that there is no unique minimum spanning tree. Since the matrix is 100x100, how many non-unique edges are there?

```
length(which(duplicated(as.vector(dm)) == T))
```

```
## [1] 5273
```

5273 edges are non-unique, overall not very many. Nonetheless, the sum of edge weights should be the same.

Load OPTREES library, reformat dissimilarity matrix as an edgelist:

```
suppressMessages(library("optrees", quietly = T))
arcs <- as.matrix(expand.grid(t1=1:nrow(dm), t2=1:nrow(dm)))
arcs <- arcs[order(arcs[,1], arcs[, 2]),]
arcs <- cbind(arcs, dist=as.vector(dm))
```

MST is non-unique, but the weight across the found MST should be the same regardless of algorithm used:

```
# Prims algorithm
mst_prims <- optrees::msTreePrim(nodes = 1:nrow(dm), arcs = arcs)
sum(mst_prims$tree.arcs[, 3])
```

```
## [1] 33.50638
```

Same answer for boruvkas:

```
# Boruvkas algorithm
mst_boruvka <- optrees::msTreeBoruvka(nodes = 1:nrow(dm), arcs = arcs)
sum(mst_boruvka$tree.arcs[, 3])
```

```
## [1] 33.50638
```

Same answer for Kruskals

```
# Kruskals algorithm
mst_kruskals <- optrees::msTreeBoruvka(nodes = 1:nrow(dm), arcs = arcs)
sum(mst_kruskals$tree.arcs[, 3])
```

```
## [1] 33.50638
```

What if I switch packages?

```
suppressMessages(library("igraph", quietly = T))
graph <- igraph::graph_from_edgelist(arcs[, 1:2])
E(graph)$weight <- dm
prims <- igraph::mst(graph)
el <- igraph::as_edgelist(prims)
sum(mapply(function(i, j) dm[i, j], el[, 1], el[, 2]))
```

```
## [1] 33.50638
```

What about the vegan package?

```
suppressMessages(library("vegan", quietly = T))
sum(vegan::spantree(dm)$dist)
```

```
## [1] 33.50638
```

What about the ape package? (requires division by two due to double-counting)

```
suppressMessages(library("ape", quietly = T))
row_indices <- apply(ape::mst(dm), MARGIN = 1, function(row) which(row == 1))
edge_weights <- c()
for (i in 1:length(row_indices)) {
  edge_weights <- c(edge_weights, sapply(row_indices[i], function(j) dm[i, j]))
}
sum(edge_weights)/2
```

```
## [1] 33.50638
```