

# Spanning trees that preserve network distances

Lovro Šubelj

*University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, Ljubljana, Slovenia*

Spanning tree of a network or a graph is a subgraph connecting all the nodes with the minimum number of edges. Spanning tree retains the connectivity of a network and possibly other structural properties, and is one of the simplest techniques for network simplification or sampling, and for revealing its backbone or skeleton. The Prim's algorithm and the Kruskal's algorithm are well known algorithms for computing a spanning tree of a weighted network. We study the performance of these algorithms on unweighted networks, and compare them to other alternatives. We show that the distances between the nodes and the diameter of a network are better preserved by an alternative algorithm based on the breadth-first search node traversal. Spanning trees computed with the breadth-first search algorithm retain short distances and small diameter of small-world networks, with a heavy-tailed node degree distribution. We support our results by experiments on various synthetic graphs and more than a thousand real networks, and demonstrate different practical applications of computed spanning trees. We conclude that, if a spanning tree is supposed to retain the distances between the nodes of an unweighted network, then the breadth-first search algorithm should be the preferred choice.