Phylogeny is the study of the evolutionary relationships between a set of organisms. Phylogenetic trees, are a graphical representation of these relationships. However, sometimes we may seek to compare the phylogenetic trees of two species. We focus our research on project on the research presented by Brodal et al. (2008), “Efficient Algorithms for Computing the Triplet and Quartet Distance Between Trees of Arbitrary Degree”. According to Brodal et al. (2008), there exists several ways to measure the distance between two trees. One such method being the Robinson-Foulds method, with the paper focusing on the triplet and quartet distance. The paper states that these two measures of distance enumerates all subsets of three or four leaves respectively and then counts the number of topologies created from these subsets of leaves that are different between two given phylogenetic trees.

These topologies are created using the “Lowest Common Ancestor” as the root. The paper defines a triplet/quartet as a subset of three/four leaves. It also defines the triplet/quartet distance as the number of triplets/quartets whose topologies differ between two given input trees. Triplets and quartets are further divided into two classes: resolved and unresolved. The difference between a resolved triplet and an unresolved triplet is that in the unresolved case, all leaves have the same parent, whereas in a resolved triplet topology, two leaves may have the same parent and one leaf another. In an unresolved quartet topology, all leaves once again have the same parent, whereas in the resolved case, the four leaves are broken up into two partitions where one set of two leaves have a parent and the next set of two leaves have a different parent.

Within these two cases, the paper explains that between two trees there are five cases that are within the realm of possibility and that need to be counted. The first number is the count of resolved topologies in both trees where they agree, A. The second is the count of resolved topologies in both trees where they disagree, B. The third and fourth is the count where one topology is resolved in one tree and in the other unresolved, C and D. Finally, the fifth case is where both topologies are unresolved, E. The paper notes that in the third and fourth case, the topologies can never agree. While in the fifth case, it is also noted that they always agree.

The general algorithm focuses on the calculation of A+B+C, D+E, A+B+D, and C+E. By having certain variables, such as A and B, we can find C or D. The paper focuses on finding A, B, and E.