

Median, tree indexed random walks, and phylogeny reconstruction

Arash Jamshidpey^{*}

ajams015@uottawa.ca

Median problem has played an important role in phylogeny reconstruction. It has been studied by many people to approximate the true ancestor, and more generally, to reconstruct the ancestral tree. But the question is: When does the median reflect the true ancestor?

Suppose $A := \{y_1, \dots, y_k\}$ be points of a metric space (Y, d) . Then a median of A is a point of space, say y , whose total distance to A , $d(y, A) := \sum d(y, y_i)$ be minimum. We investigate when the median well approximates the true ancestor in the following model and some similar models. Consider a rooted tree T with finitely many leaves and denote the set of its leaves by ∂T . Suppose X be a T -indexed simple random walk, starting at identity, on the Cayley graph of the signed symmetric group S_n^\pm with respect to reversals as the set of generators. S_n^\pm represents the space of all unichromosomal genomes with n genes, where the signs indicate the orientation of genes. We study the median of the points of random walk occurring at times ∂T . Denote the set of all these random points by $X_{\partial T}$. There may exist finitely many medians for a set of signed permutations length n . Therefore, it will be appropriate to find some common properties for all medians of $X_{\partial T}$. In many cases we observe that medians loose credibility as an exact approximation of the true ancestor. Adding some conditions, we will discuss how we can, better, approximate the true ancestor, and, generally, reconstruct the phylogenetic tree.

^{*}Department of Mathematics and Statistics, University of Ottawa, 585 King Edward, Ottawa, ON K1N 6N5, CANADA.