1010-68-109 **Moses Charikar*** (moses@cs.princeton.edu) and **Nir Ailon** (nailon@cs.princeton.edu).

Fitting tree metrics: Hierarchical clustering and Phylogeny.

Given pairwise dissimilarity data, we study the problem of fitting a tree metric to this data so as to minimize additive error (i.e. the difference between the tree metric and the data).

This problem arises in constructing an M-level hierarchical clustering of objects (or an ultrametric) – a basic problem in statistics. This is a generalization of correlation clustering (which corresponds to $M = 1$). We give a very simple randomized combinatorial algorithm for this problem that achieves an approximation ratio of $M + 2$.

The problem of fitting tree metrics also arises in phylogeny where the objective is to infer the evolution tree by fitting a tree to dissimilarity data on taxa. The quality of fit is measured by taking the ℓ_p norm of the difference between the tree metric and the given data.

No non-trivial approximation for general ℓ_p norms was known before. We present a novel LP formulation for this problem and obtain an $O((\log n \log\log n)^{1/p})$ approximation using this. En route, we obtain an $O((\log n \log\log n)^{1/p})$ approximation for the closest ultrametric under the ℓ_p norm. Our techniques are based on representing and viewing an ultrametric as a hierarchy of clusterings, and may be useful in other contexts. (Received August 22, 2005)