Maximum Entropy Summary Trees
Howard Karloff, Yahoo! Research

Given a very large, node-weighted, rooted tree on, say, $n$ nodes, if one has only enough space to display a $k$-node summary of the tree, what is the most informative way to draw the tree? We define a type of weighted tree that we call a "summary tree" of the original tree, that results from aggregating nodes of the original tree subject to certain constraints.

We suggest that the best choice of which summary tree to use (among those with a fixed number of nodes) is the one that maximizes the information-theoretic entropy of a natural probability distribution associated with the summary tree, and we provide a (pseudopolynomial-time) dynamic-programming algorithm to compute this maximum entropy summary tree, when the weights are integral.

The result is an automated way to summarize large trees and retain as much information about them as possible, while using (and displaying) only a fraction of the original node set. We also provide an additive approximation algorithm and a greedy heuristic that are faster than the optimal algorithm, and generalize to trees with real-valued weights.

This is joint work with Ken Shirley of ATT Labs and Richard Cole of NYU.