ETH Zürich **Metric Embeddings**
Institute of Theoretical Computer Science **Spring 2010**

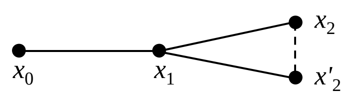
Problem Set 5

(to be handed in May 17 in class)

1. Given a string $u \in \{0,1\}^d$, prove that at most half of the strings $v \in \{0,1\}^d$ are at edit distance at most cd from u, for some constant $c > 0$. (Hint: Note that disallowing character replacements affects the edit distance metric only by a constant factor. Furthermore, without loss of generality we can assume that all deletions are performed first, and then all insertions.)

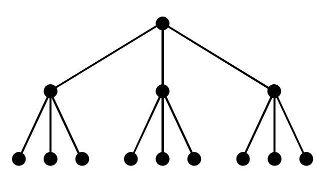
2. (Lower bound for embedding trees into ℓ_2.)

   (a) Show that for every $\varepsilon > 0$ there exists a $\delta > 0$ with the following property. Let x_0, x_1, x_2, $x_2' \in \ell_2$ be points such that $\|x_0 - x_1\|$, $\|x_1 - x_2\|$, $\|x_1 - x_2'\| \in [1, 1+\delta]$ and $\|x_0 - x_2\|$, $\|x_0 - x_2'\| \in [2, 2+\delta]$ (so all the distances are almost like the graph distances in the following tree, except possibly for the one marked by a dotted line).

   Then $\|x_2 - x_2'\| \le \varepsilon$; that is, the remaining distance must be very short.

   (b) Let $T_{k,m}$ denote the complete k-ary tree of height m; the following picture shows $T_{3,2}$:

   Show that for every r and m there exists a k such that whenever the leaves of $T_{k,m}$ are colored by r colors, there is a subtree of $T_{k,m}$ isomorphic to $T_{2,m}$ with all leaves having the same color.

   (c) Use (a), (b), and Question 6 of Problem Set 4 to prove that for every $D > 1$ there exist m and k such that the tree $T_{k,m}$ considered as a metric space with the shotest-path metric cannot be D-embedded into ℓ_2. (Hint: Asign to each leaf v of the tree a "color" which is a vector of length $\binom{m+1}{2}$, encoding information about the distortion between every pair of vertices in the path from v to the root.)