

(a) For each protein p , we define a set S_p consisting of all proteins similar to it; we do this by simply enumerating all proteins q for which $d(p, q) \leq \Delta$. With respect to these sets, a representative set $R \subseteq P$ is simply a set for which $\{S_p : p \in R\}$ is a set cover for P .

Thus, to approximate the size of the smallest representative set, we can use the approximation algorithm for Set Cover from this chapter, obtaining an approximation guarantee of $O(\log n)$.

(b) The problem with using the approximation algorithm for Center Selection is that we'd obtain a set R of proteins for which every protein is within distance 2Δ of some element of R . But this doesn't satisfy the requirements for a representative, which stipulated that every protein had to be within distance Δ of some element of R .

¹ex815.903.104