Supplemental Material

A Graph Coarsening Algorithm for Compressing Representations of Single-Cell Data with Clinical or Experimental Attributes

Chi-Jane Chen[†], Emma Crawford[ʎ], and Natalie Stanley[‡]

Department of Computer Science and Computational Medicine Program
The University of North Carolina at Chapel Hill,
Chapel Hill, NC, 27599, USA
{[†]*chijane@cs.unc.edu*,[ʎ] *emmabc@email.unc.edu*,[‡] *natalies@cs.unc.edu*}

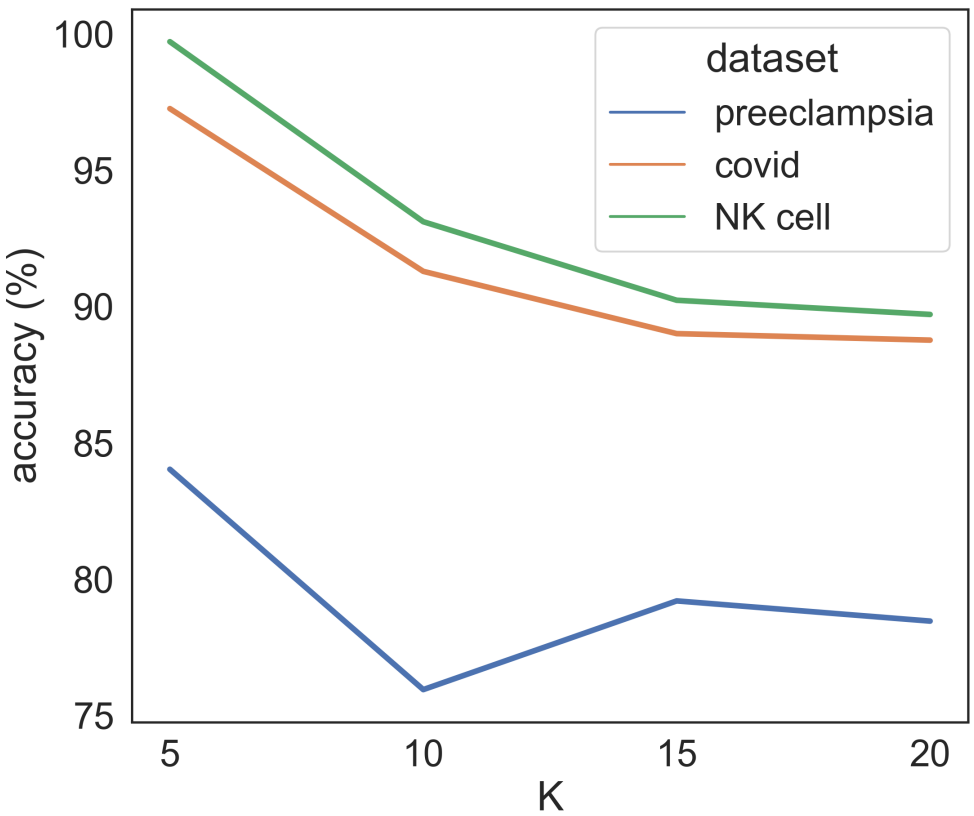

Fig. 1. **Accuracy as a function of k for the KNN graph.** To determine the optimal number of nearest neighbors, K for cytocoarsening, we ran it on three datasets, keeping all other parameters as their default values and varying K between 5 and 20. The accuracy measures the percentage of nodes for which the original attribute labels and coarsened attribute labels align. The default parameter values are $\alpha = 26$, The coarsening process terminates if the node of the particular passes less than seventy-five percent.

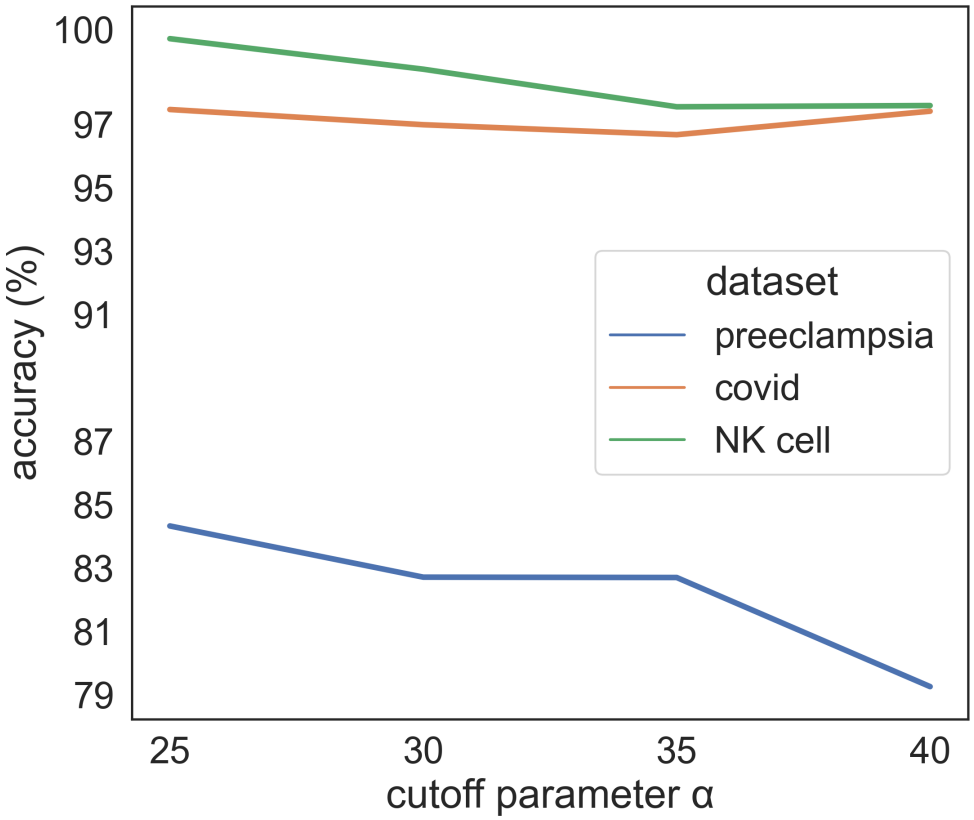

Fig. 2. **Accuracy as a function of cutoff percentile thresholds for the coarsening graph.** While setting the percentage of coarse nodes to 75% of the original nodes in the graph, we ran cytocoarsening on each of the three datasets for different cutoff percentiles, ranging between 25 and 40. The results show that higher accuracy is achieved for lower cutoff percentiles.