Figure: Sensitivity analysis for the K hyperparameter for L2H-TURTLE on the CIFAR-100 datase. The K hyperparameter corresponds to number of leaves in the hierarchy and the number of clusters set for training the pre-trained flat model (TURTLE in this case). We perform an ablation for the value of K, varying its value across the range $\{85, 90, 95, 100, 105, 110, 115\}$, that is symmetric around the true number of classes/clusters equal to 100. Therefore, we both explore the case of over- and under-estimating the true number clusters at the leaf level of the hierarchy. We report the results for both flat (NMI, ARI, ACC, LP) and hierarchical (DP, LHD) metrics—standard deviations across runs with five different seeds are reported as shaded areas around the line indicating mean values. Best performance across all metrics is achieved when K is set to the true number of clusters, while the performance gracefully degrades when K is set to be an over- or under-estimated value. This demonstrates robustness and stability of our approach with respect to this hyperparameter, which is particularly important in practical settings where the exact true number of classes if not known a priori. Finally, we find log-normalized TURTLE model loss—reported in the rightmost plot in the bottom row—to be indicative of the true value of K, with the minimum value achieved when K equals the true number of clusters. In practical settings one can use this metric to select the value of K when a value/proxy for the true number of clusters is not available.