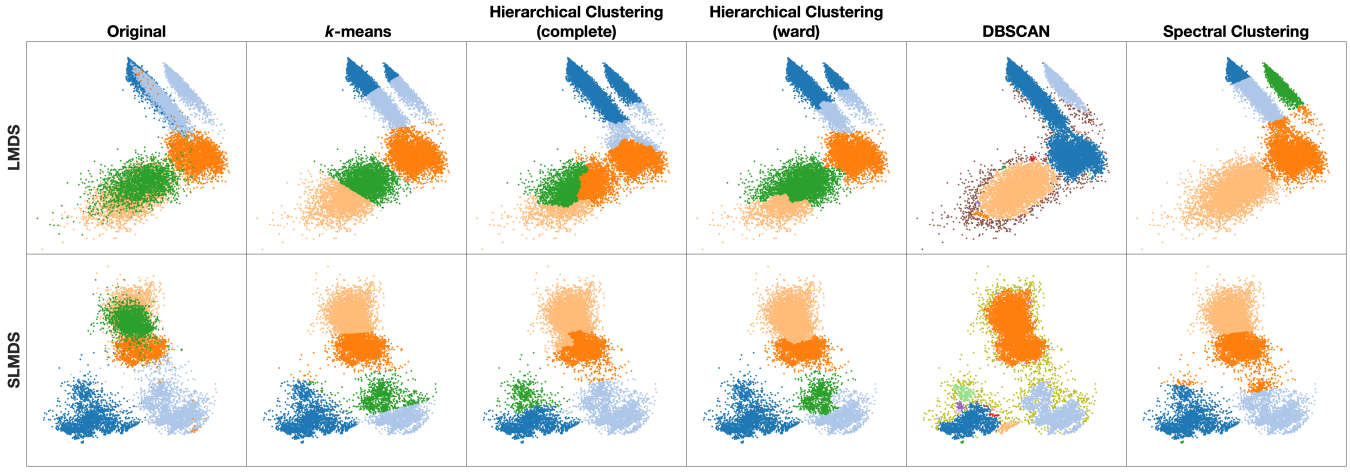
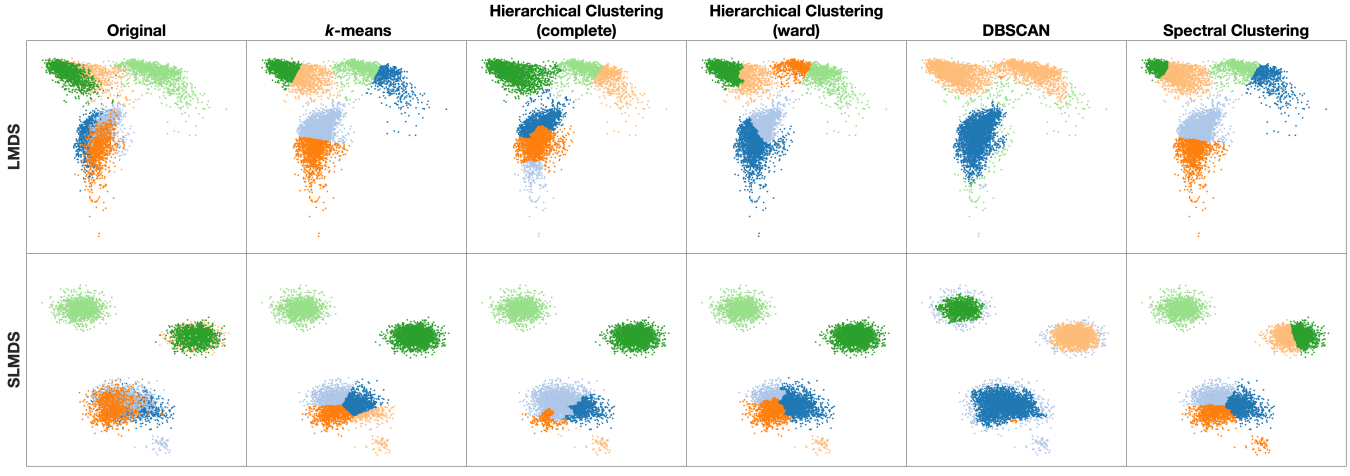
SUPPLEMENTAL MATERIALS

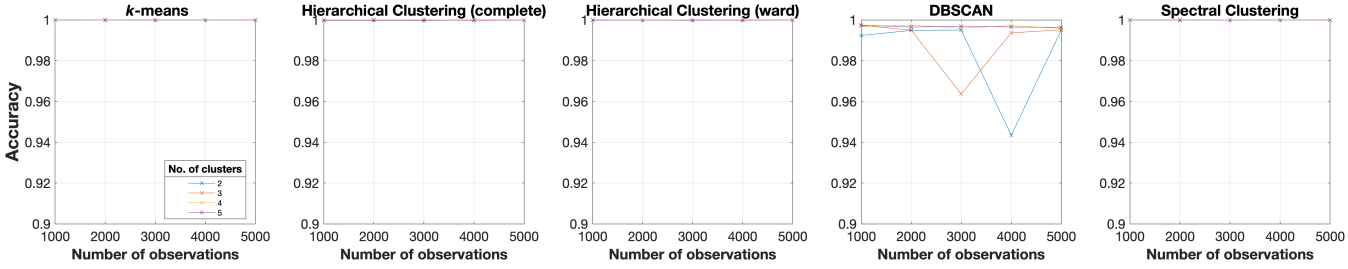


(a) Human activity data

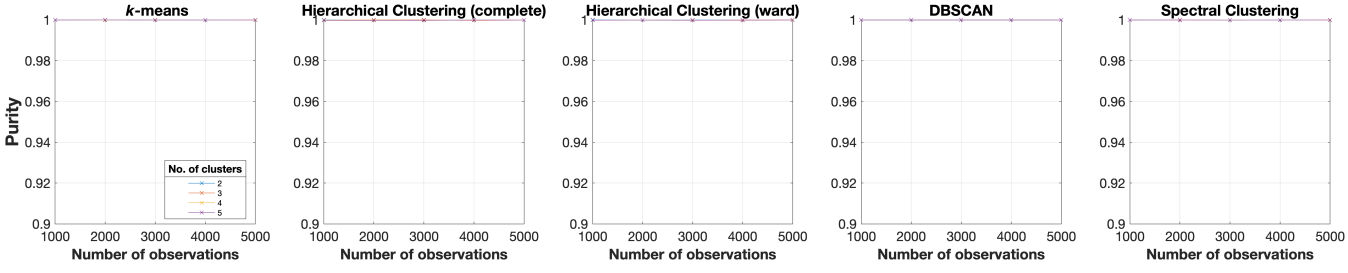


(b) Human activity recognition data

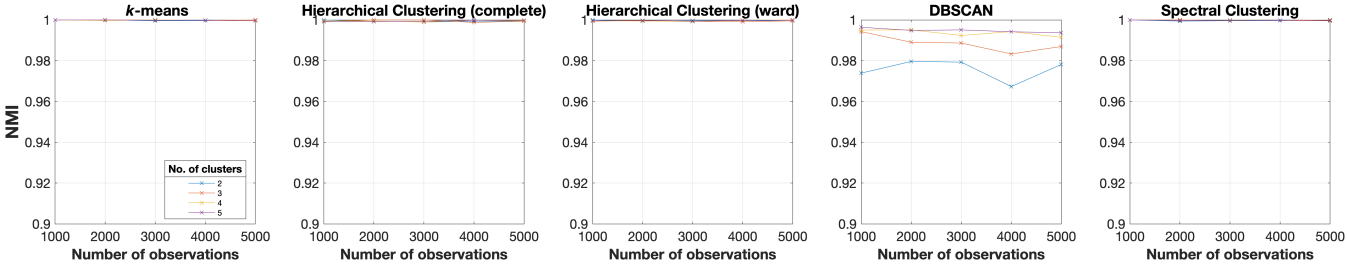
Fig. 1: Clusters obtained from k-means, hierarchical clustering (complete and ward linkage), DBSCAN, and spectral clustering on the original and SLMDS-processed versions of two human activity data: Human activity data ($n = 50$, $N = 24075$) and Human activity recognition data ($n = 10$, $N = 7352$). The HAD results show overlaps between two dynamic activities–running and dancing. The other three activities (sitting, standing, and walking) seem to have little overlap with different activities. The HAR results contain two overlapping regions. The first region contains two static human activities (sitting and standing) and the other consists of three dynamic walking motions (walking, walking downstairs and upstairs).



(a) Accuracy



(b) Purity



(c) NMI

Fig. 2: Accuracy, purity, and NMI of the clusters obtained from *k*-means, hierarchical clustering (complete and ward linkage), DBSCAN, and spectral clustering on 2D synthetic data. These results show that all five clustering algorithms yield nearly perfect accuracy, purity, and NMI for any number of clusters or observations. Hierarchical clustering (complete and ward linkage) and spectral clustering perform well especially without fluctuation of values. Although *k*-means displays a slight drop in accuracy, purity, and NMI for data with five clusters and 3000 observations, the values are still above 0.9.

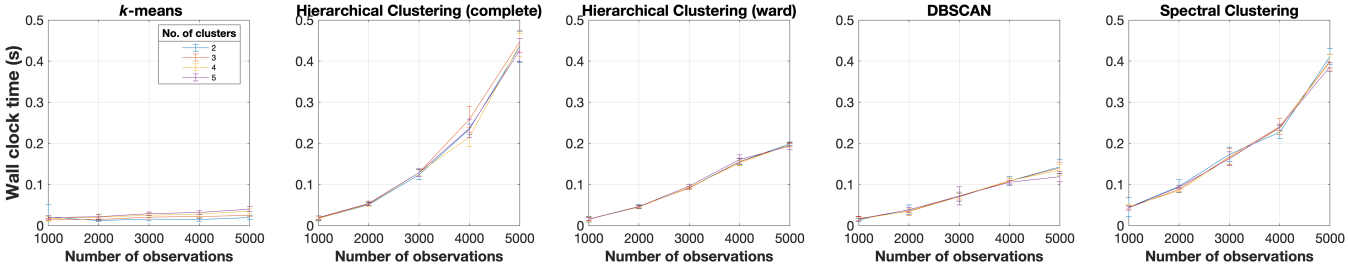
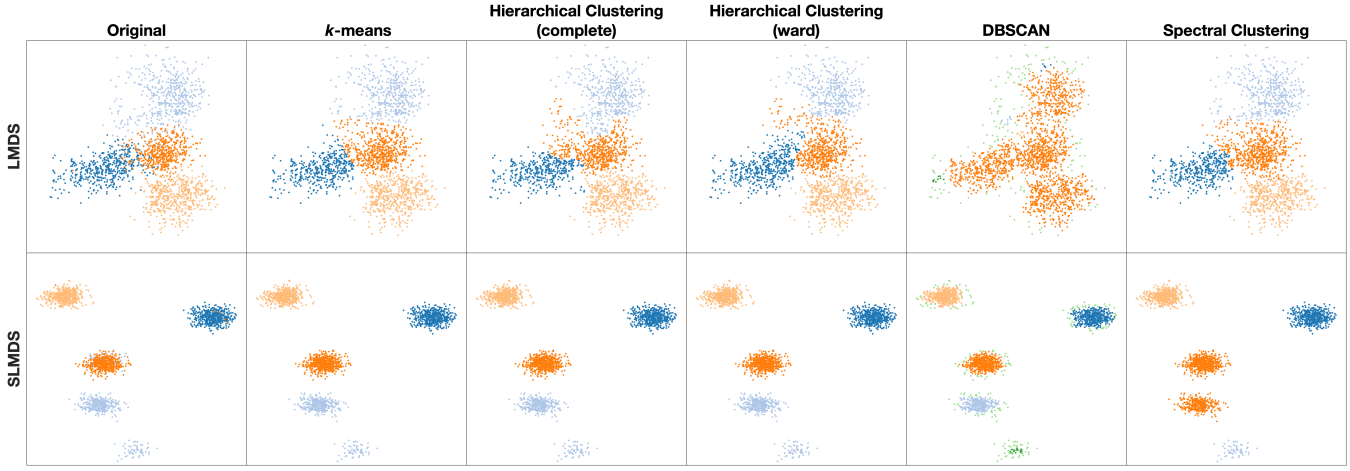
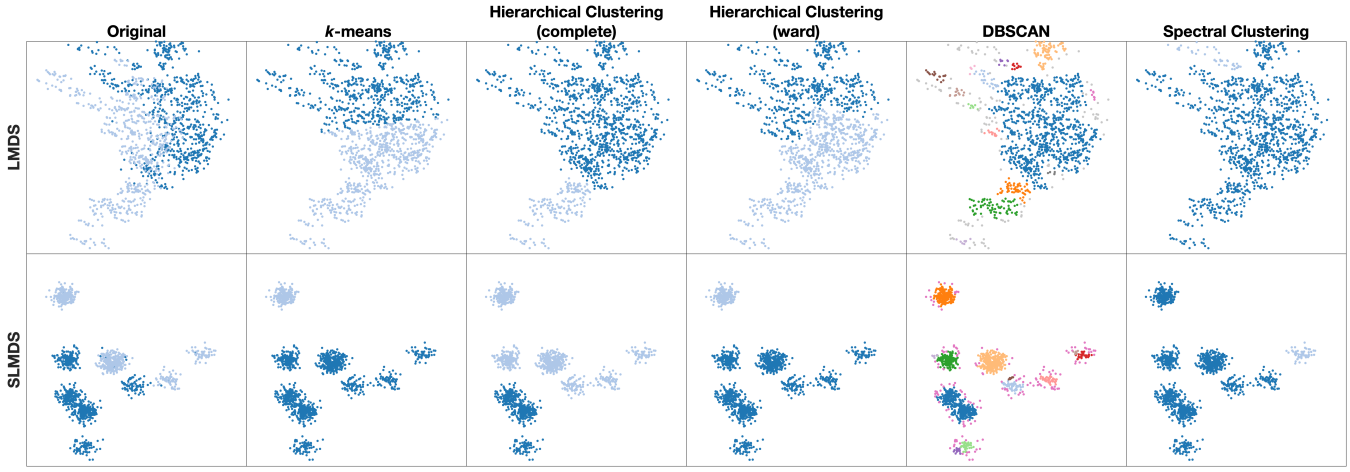


Fig. 3: Wall-clock timing of k-means, hierarchical clustering (complete and ward linkage), DBSCAN, and spectral clustering on 2D synthetic data with varying number of clusters and observations. We observe that k-means is the most fastest among other clustering algorithms.



(a) WiFi



(b) Banknote

Fig. 4: Clusters obtained from k-means, hierarchical clustering (complete and ward linkage), DBSCAN, and spectral clustering on the original and SLMDS-processed versions of two real-world data: WiFi ($n = 6$, $N = 2$K) and Banknote ($n = 4$, $N = 1327$).

Table 1: Average accuracy, purity, NMI scores for k-means, hierarchical clustering (complete and ward linkage), DBSCAN, and spectral clustering on SLMDS-processed HAD and HAR data sets

Measurement	Data	k-means	Hierarchical (complete)	Hierarchical (ward)	DBSCAN	Spectral
Accuracy	HAD	0.8137	0.8128	0.8039	0.6624	0.8488
	HAR	0.6249	0.6005	0.6340	0.5459	0.6301
Purity	HAD	0.8704	0.8396	0.8558	0.7205	0.8577
	HAR	0.6384	0.6079	0.6413	0.5468	0.6339
NMI	HAD	0.8161	0.7705	0.8059	0.7189	0.8149
	HAR	0.6451	0.6754	0.6722	0.7433	0.6159