Benchmarking in cluster analysis: a study on Spectral Clustering, DBSCAN, and K-means

Nivedha Murugesan · Irene Cho · Cristina Tortora

N. Murugesan ● I. Cho ● C. Tortora
Department of Mathematics and Statistics
San José State University
One Washington Square
San José, CA 95192
E-mail: nivedha.murugesan@sjsu.edu, sisang.cho@sjsu.edu, cristina.tortora@sjsu.edu

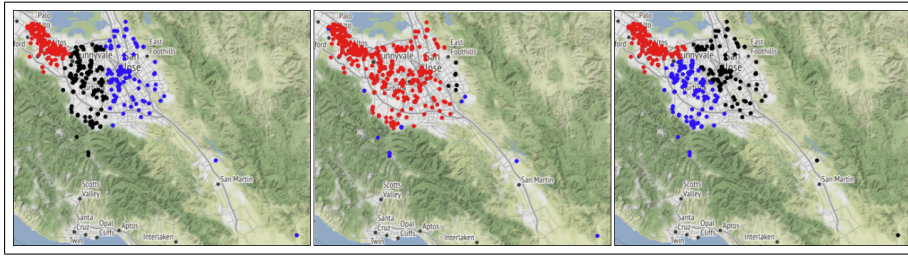| Data set	Method	Index	G = 2	G = 3	G = 4	G = 5
Bike	K-Means	Silhouette	0.23	0.22	0.18	0.18
		CH Index	8.32	7.03	6.69	5.90
	SC	Silhouette	0.23	0.11	0.18	-0.01
		CH Index	8.32	5.14	6.69	2.38
	DBSCAN	Silhouette	—	0.16	—	—
		CH Index	—	3.67	—	—
Forest Fire	K-Means	Silhouette	0.55	0.43	0.40	0.39
		S-Dbw Index	1.12	1.61	1.79	2.10
	SC	Silhouette	0.29	0.36	0.37	0.29
		S-Dbw Index	0.79	2.18	4.39	4.05
	DBSCAN	Silhouette	0.47	—	—	—
		S-Dbw Index	2.91	—	—	—
Airbnb	K-Means	Silhouette	0.52	0.46	0.47	0.46
		S-Dbw Index	1.10	1.48	1.86	4.31
	SC	Silhouette	0.52	0.46	0.35	0.38
		S-Dbw Index	1.08	1.48	2.72	5.52
	DBSCAN	Silhouette	—	0.34	—	—
		S-Dbw Index	—	2.66	—	—

Table 1: Internal Indices for Empirical data sets

Fig. 1: Airbnb Clustering Results on Geographical Plot, colors represent the K-means (left), DBSCAN (middle), and SC partition (right)