

Scalable Force Scheme – Supplementary Material

1 Hyperparameters for FS, GFS and SFS

Table 1 shows the differences between FS, GFS, and SFS, as well as the specific hyperparameter values used for the results reported in the paper. Note that we set a conservative value of $Iter_{max}$ according to our observations: for FS, all datasets had already “converged” (i.e., the error was no longer decreasing for the last iterations) before 50 iterations, while for GFS and SFS, it would converge due to the early-stopping criterion well before 200 iterations. Note also that, when reporting the time needed to run a single iteration of GFS on the large datasets, the distance matrix was not precomputed (due to excessive memory requirements) and thus distances were also not normalized.

Parameter	FS	GFS	SFS
$Iter_{max}$	50	200	200
η_0	0.1	0.1	0.1
γ	1	0.9	0.9
W	No error window	10	10
Permute points	No	Yes	Yes
Normalize	Yes	Yes	No
Precompute distance matrix	Yes	Yes	No
Number of anchors points	N	N	\sqrt{N}

Table 1: Comparison of the hyperparameters and main algorithmic differences between FS, GFS and SFS.

For the results where 10 seeds have been used for random number generation, these seeds have been integers 0-9. The only exception is the estimation of the time it takes FS / GFS to run large data sets, where only 1 repetition was done; in that case, `seed=0`.

2 Global and local quality metrics

Table 2 shows the values quantifying the goodness of a projection with respect to the original data. The values reported are averaged over 10 iterations. Note that continuity and trustworthiness are local metrics, used mostly to measure the quality of local methods, and might not reflect user perception when interpreting a global projection.

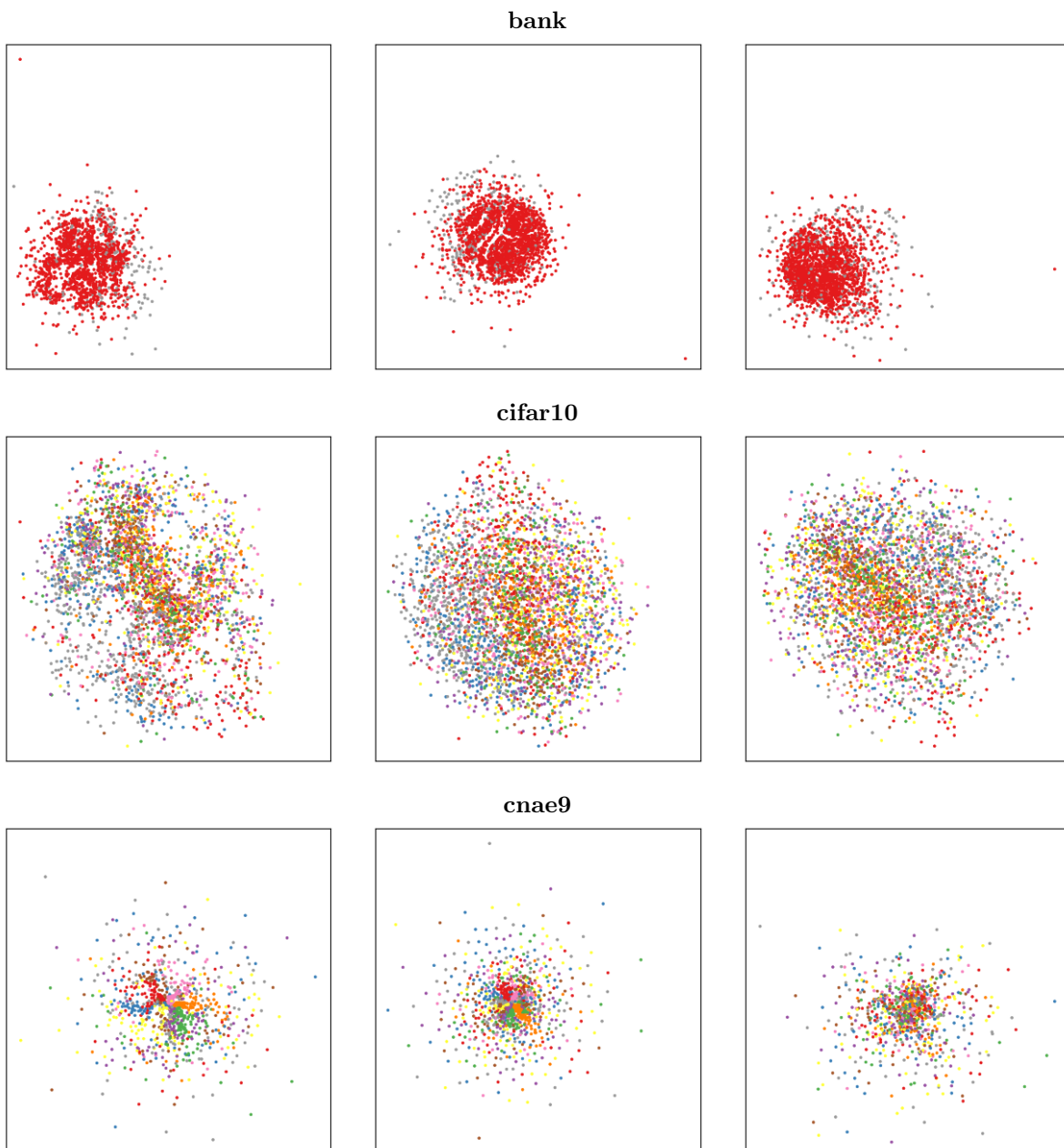
We have chosen to show continuity and trustworthiness with $k = 7$, but our observations show that other values of k offer similar results (can be found in the code).

Dataset	N	D	Stress ↓			Continuity ↑ ($k = 7$)			Trustworth. ↑ ($k = 7$)		
			FS	GFS	SFS	FS	GFS	SFS	FS	GFS	SFS
bank	2059	63	0.18	0.16	0.20	0.87	0.87	0.62	0.80	0.82	0.59
cifar10	3250	1024	0.93	0.93	0.94	0.92	0.92	0.84	0.81	0.81	0.72
cnae9	1080	856	0.71	0.72	0.72	0.81	0.79	0.71	0.57	0.56	0.51
coil20	1440	400	0.53	0.48	0.50	0.94	0.93	0.76	0.87	0.90	0.77
epileptic	5750	178	0.63	0.63	0.63	0.87	0.87	0.77	0.65	0.65	0.60
fashion_mnist	3000	784	0.57	0.56	0.60	0.95	0.96	0.81	0.89	0.90	0.80
fmd	997	1536	0.82	0.83	0.84	0.81	0.82	0.69	0.74	0.74	0.64
har	735	561	0.66	0.67	0.72	0.91	0.90	0.83	0.90	0.90	0.85
hatespeech	3221	100	0.32	0.33	0.34	0.93	0.92	0.62	0.69	0.75	0.54
hiva	3076	1617	0.78	0.77	0.77	0.87	0.88	0.66	0.65	0.66	0.52
imdb	3250	700	0.45	0.46	0.47	0.70	0.71	0.70	0.54	0.54	0.51
orl	400	396	0.81	0.82	0.83	0.90	0.94	0.66	0.85	0.89	0.65
secom	1567	590	0.77	0.71	0.71	0.77	0.83	0.64	0.64	0.67	0.53
seismic	646	24	0.29	0.26	0.30	0.97	0.96	0.72	0.93	0.94	0.74
sentiment	2748	200	0.35	0.35	0.36	0.90	0.89	0.61	0.63	0.66	0.51
sms	835	500	0.52	0.53	0.52	0.75	0.74	0.71	0.55	0.54	0.51
spambase	4601	57	0.28	0.27	0.28	0.90	0.90	0.67	0.70	0.73	0.55
svhn	732	1024	0.92	0.92	0.92	0.91	0.91	0.88	0.84	0.84	0.82

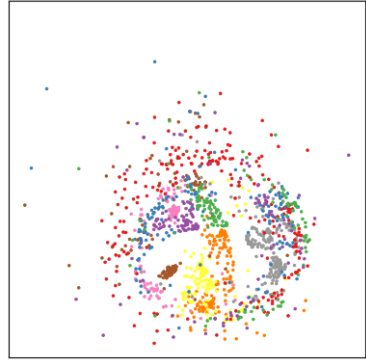
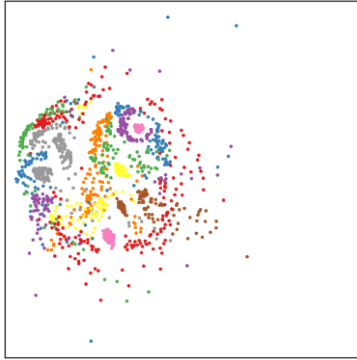
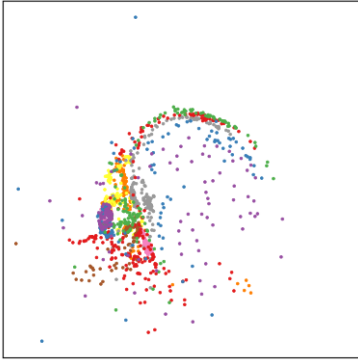
Table 2: Values for global (stress) and local (continuity and trustworthiness) quality metrics, comparing FS, GFS and SFS. The reported values are the average over 10 iterations, using different seeds for the random number generation.

3 Projections of FS, GFS and SFS

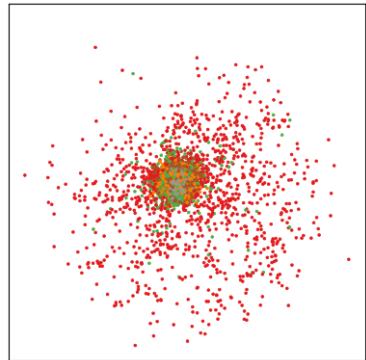
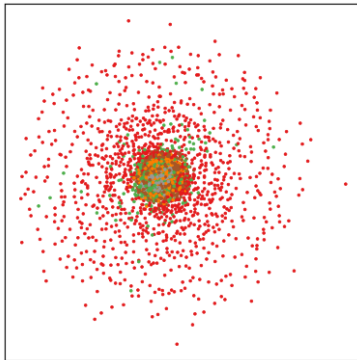
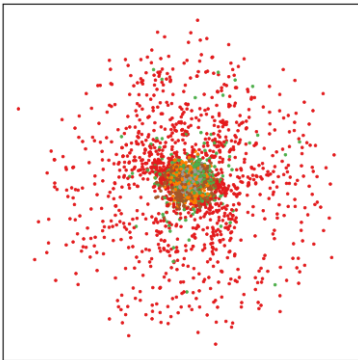
There follow the layouts produced with FS (left), GFS (middle), and SFS (right) for the 18 different datasets (taken from the survey of Espadoto et al.) reported in the quantitative results. These correspond to a single run, with `seed=0`. The coloring of the points indicates the class, as reported in the original dataset.



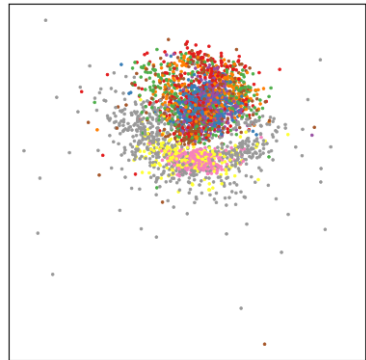
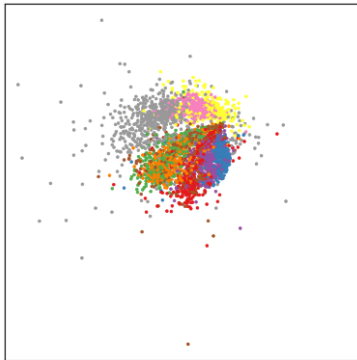
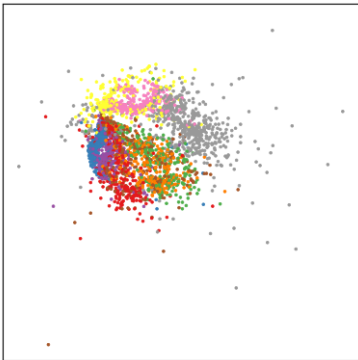
coil20



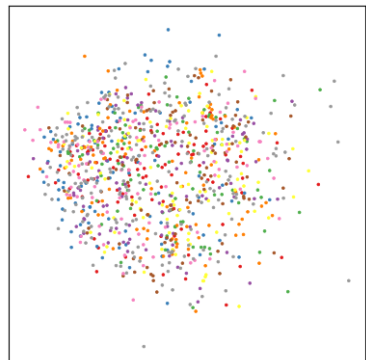
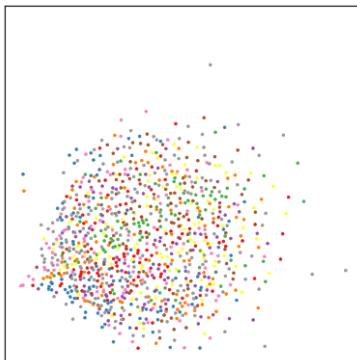
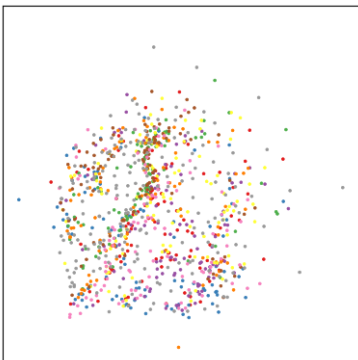
epileptic



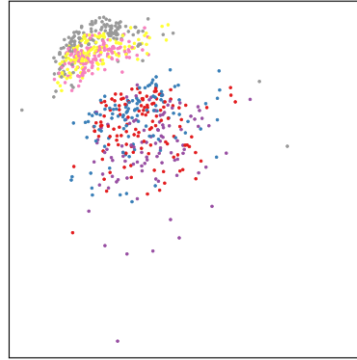
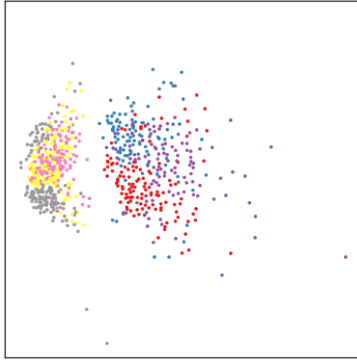
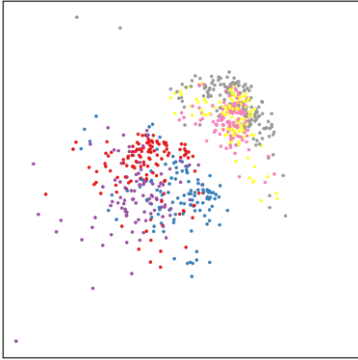
fashion_mnist



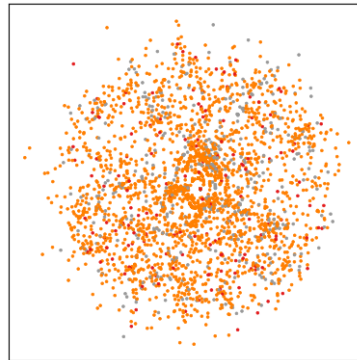
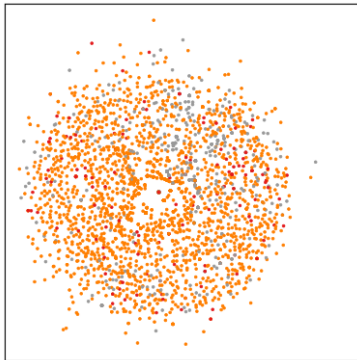
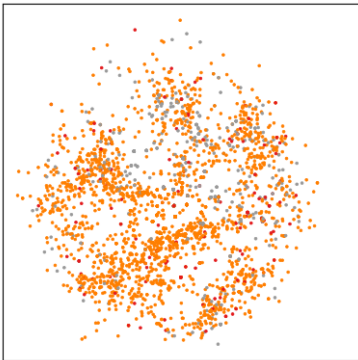
fmd



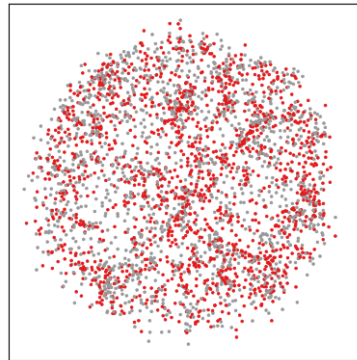
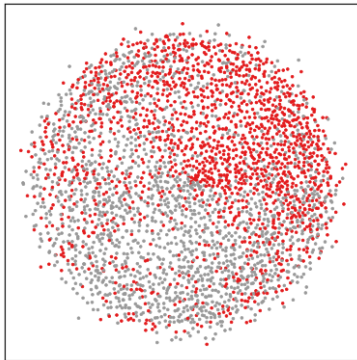
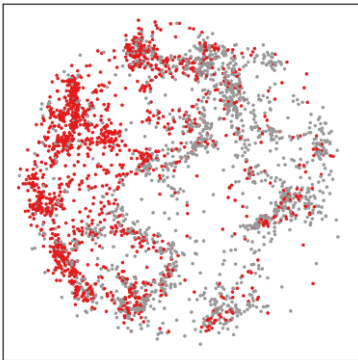
har



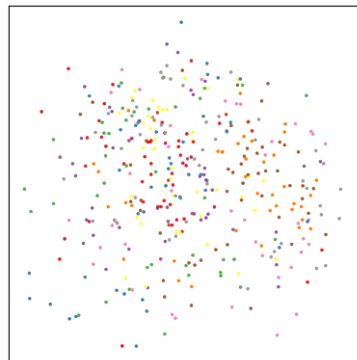
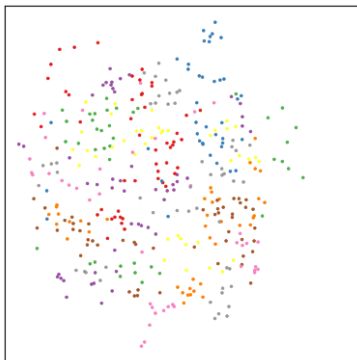
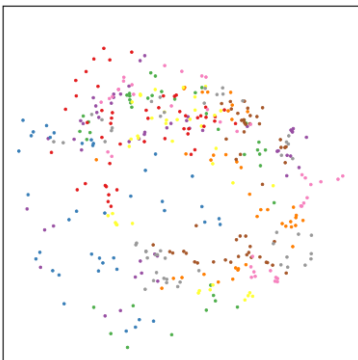
hatespeech



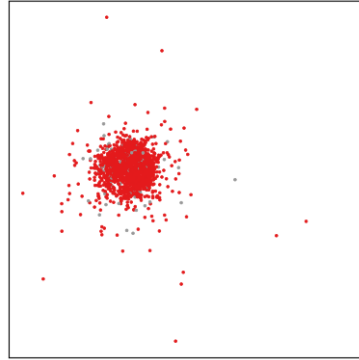
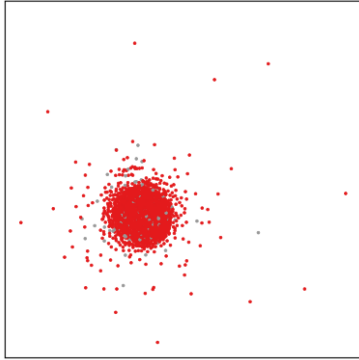
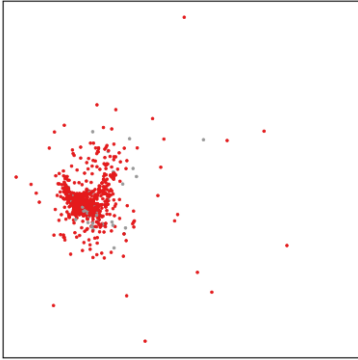
imdb



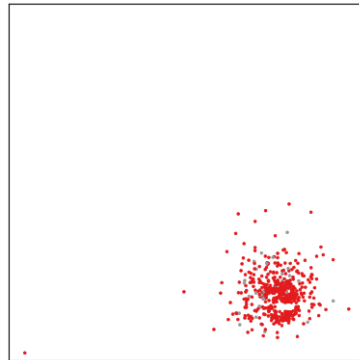
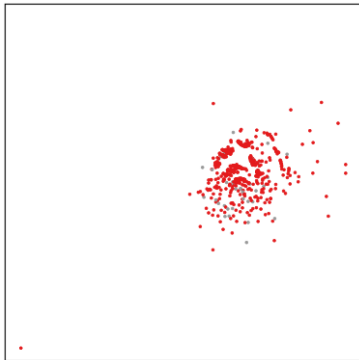
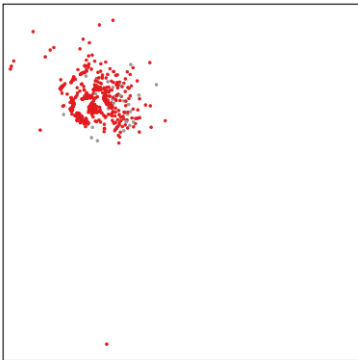
orl



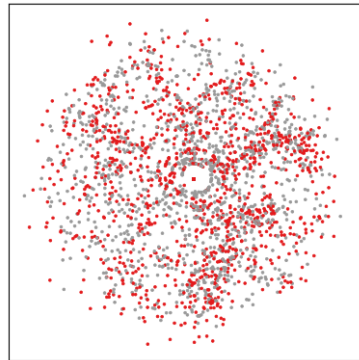
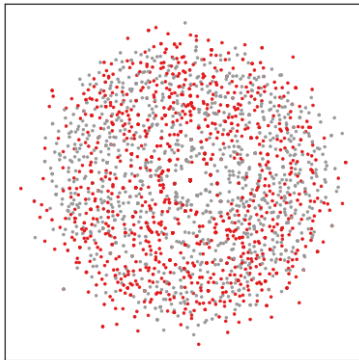
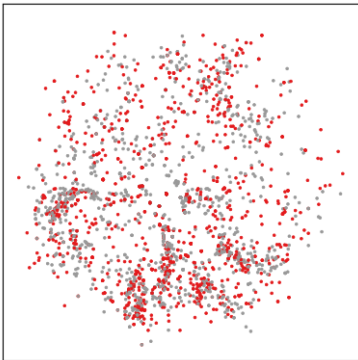
secom



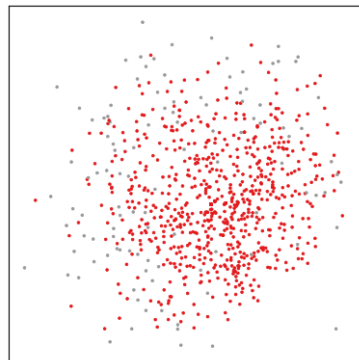
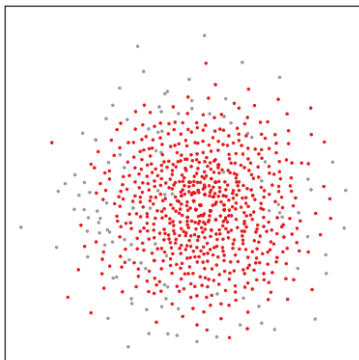
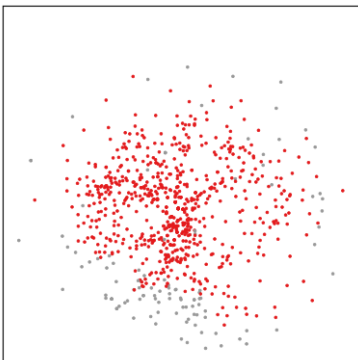
seismic



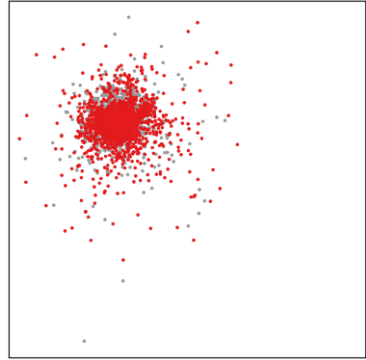
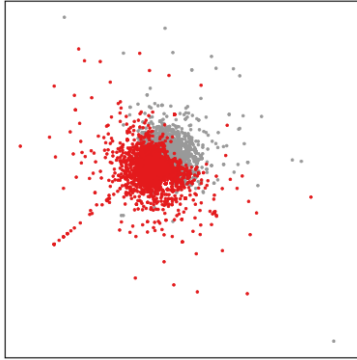
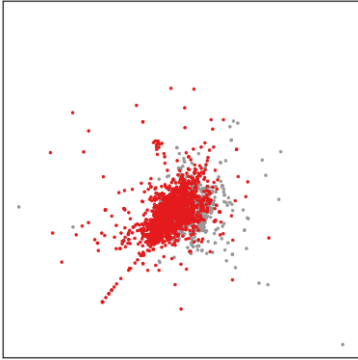
sentiment



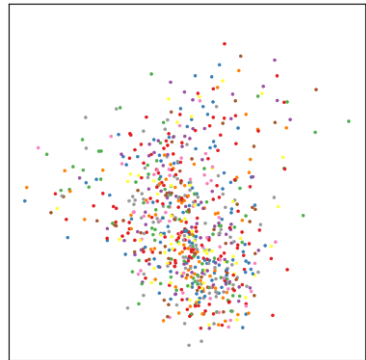
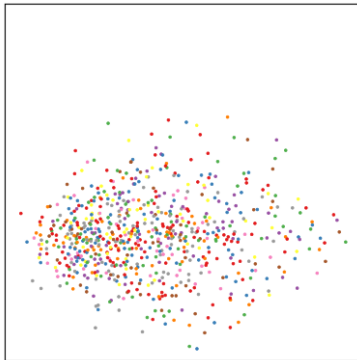
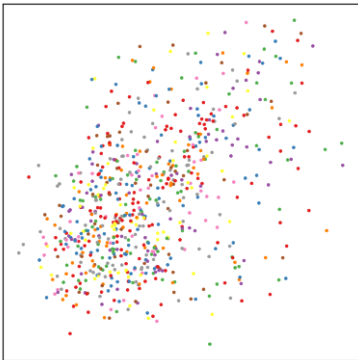
sms



spambase



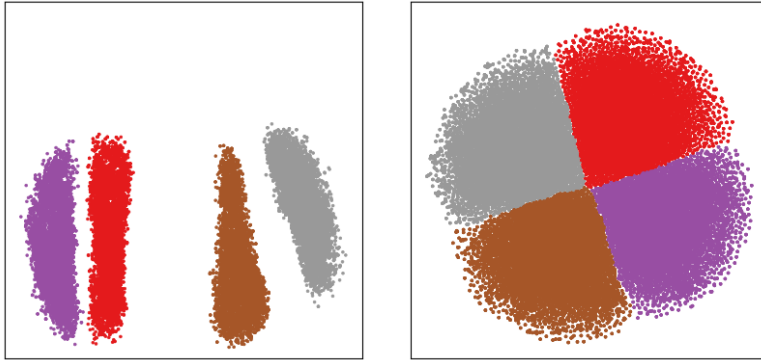
svhn



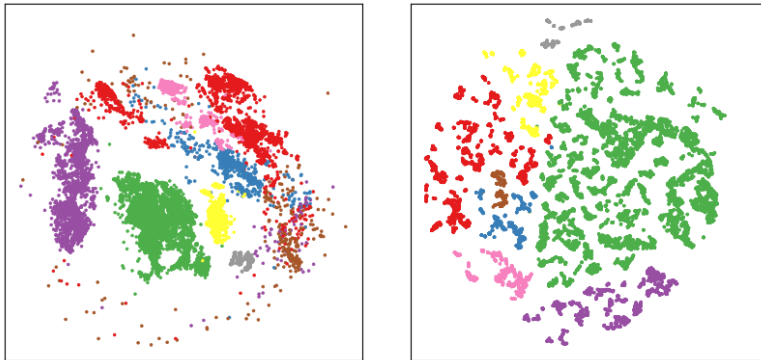
4 Examples of SFS on large datasets

In the following we include projections for large datasets produced by SFS, which would require several hours on conventional hardware for FS and GFS. For each row, on the left we show SFS projection while on the right, for comparison, we put the projection obtained with t-SNE ($perplexity = 30$), one of the most popular nonlinear local DR methods. Local methods are able to scale to much larger datasets, but the results below show that the layouts obtained are different, since local methods faithfully reproduce neighborhoods of points, but are unable to capture the global structure of the data.

Mammals ($N = 50000$, $D = 71$)



Fourier ($N = 19029$, $D = 38$)



Fiber ($N = 250000$, $D = 30$)

