Towards Scalable Spectral Clustering via Spectrum-Preserving Sparsification

Yongyu Wang, Zhuo Feng

Michigan Tech
Create the Future

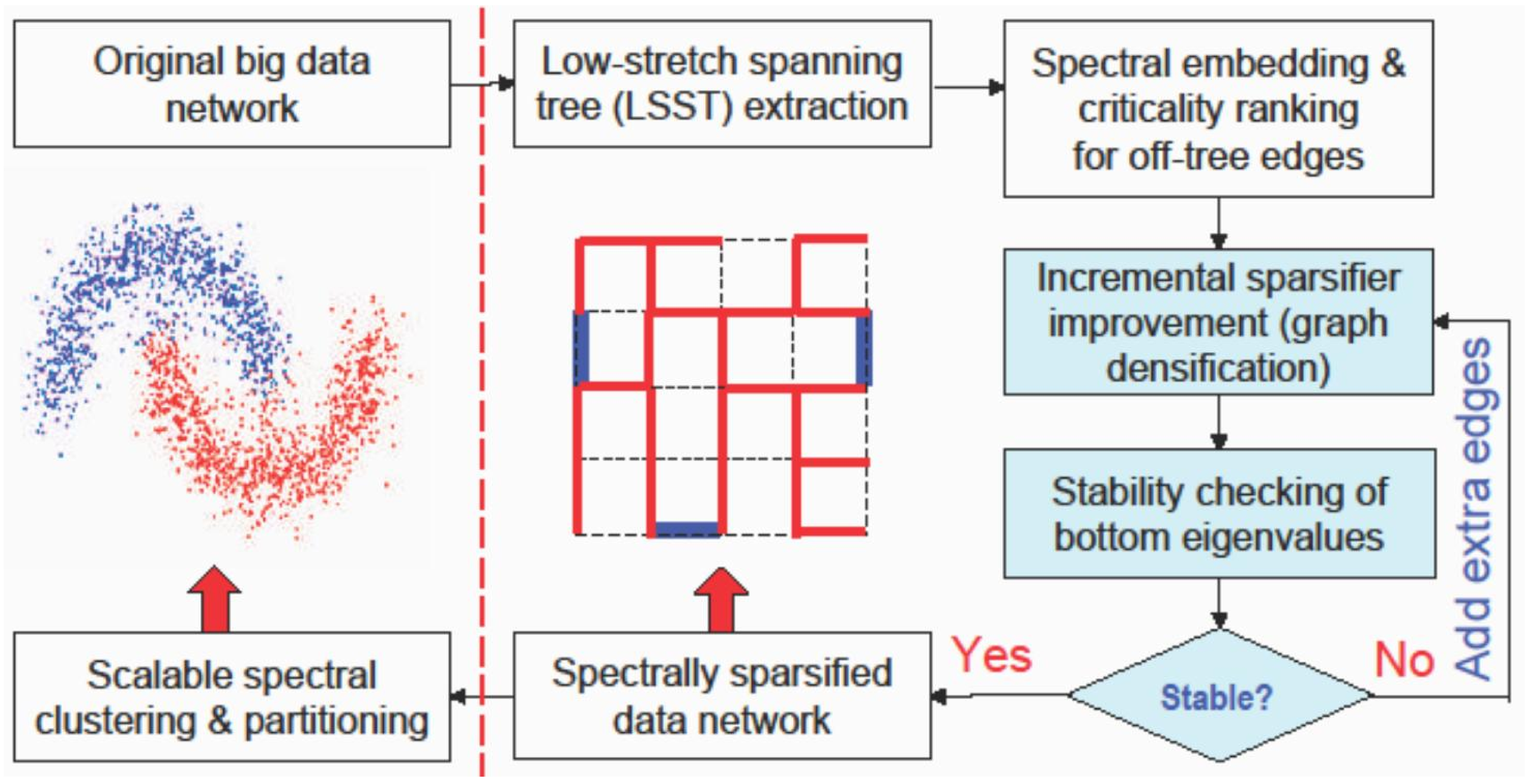
(I) Motivation

- Solving the main computational bottleneck in spectral clustering.
- Achieving scalable spectral clustering of large data networks without sacrificing solution quality.

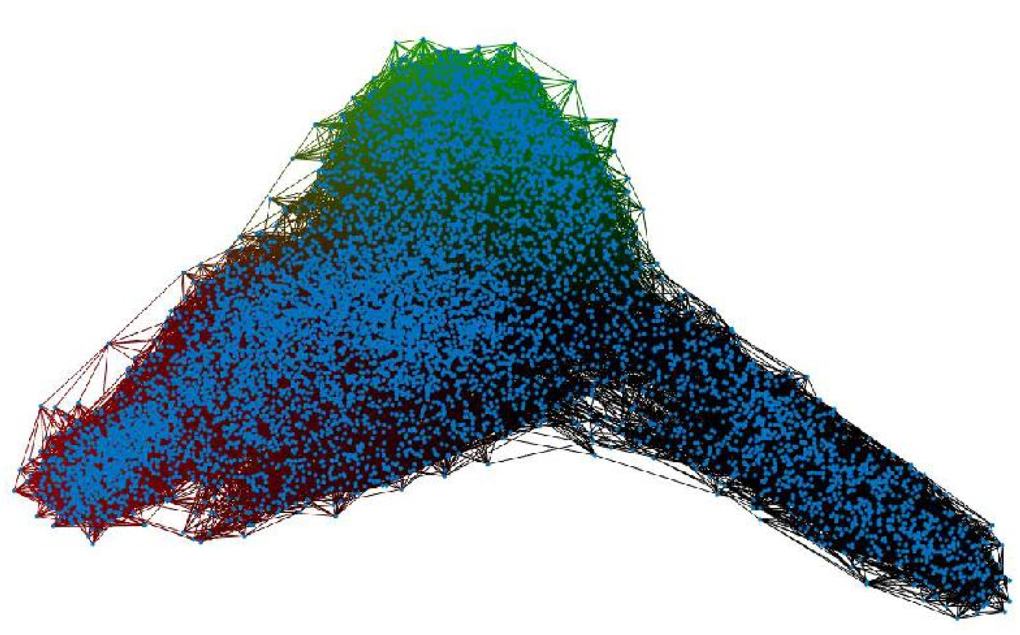
(II) Keys steps of spectral clustering

- Construct k-NN graph to represent the relationship among data set.
- Calculate the Laplacian martix of the k-NN graph.
- Calculate the eigenvectors corresponding to the bottom eigenvales of the Laplacian matrix.
- Embed the data points into spectral feature space with the calculated eigenvectors.
- Peform k-means on the spectral feature space to obtain the clustering result.

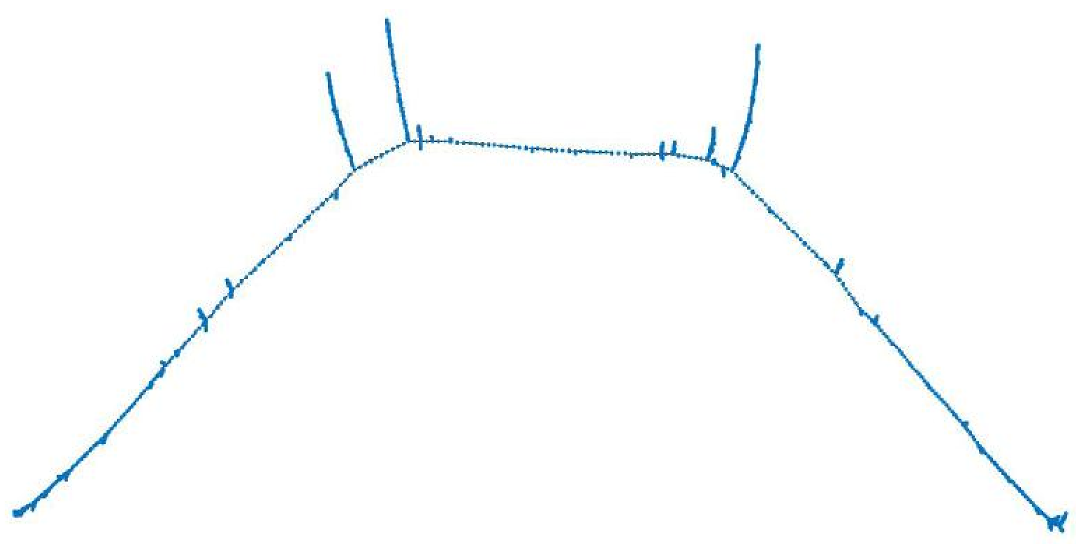
(III) Flowchart of the proposed method



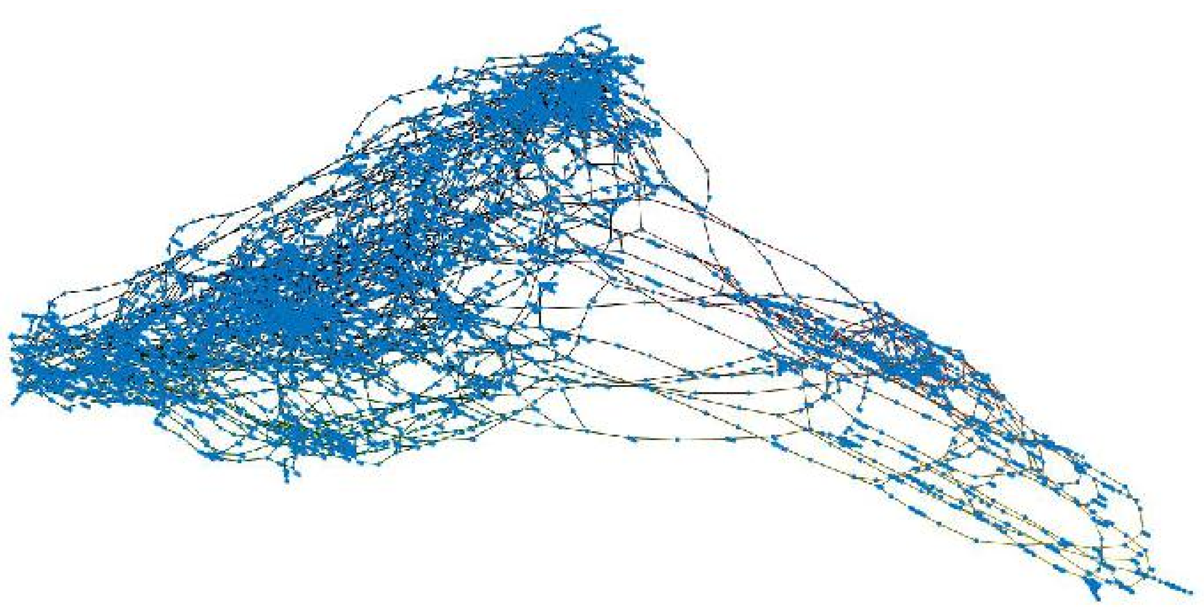
(III) Visualization of spectral sparsification



The original graph corresponding to the original affinity matrix (USPS)



The spanning tree of the original graph (USPS).



The sparsified graph corresponding to the affinity matrix (USPS)

(VI) Experimental Results

	Clustering Accuracy (ACC)							Spectral Clustering Time							
Data Set	Orig	Nyström	KASP	LSCK	LSCR	CSC	Ours	Orig	Nyström	KASP	LSCK	LSCR	CSC	Ours	λ_{max}
COIL-20	78.80	67.44	58.83	72.41	68.45	75.83	76.27	0.37	0.46	2.74	2.44	0.23	1.57	0.28 (1.32X)	138
PenDigits	81.12	68.70	75.83	80.77	77.89	47.09	83.26	0.47	0.28	1.00	0.81	0.23	6.03	0.36 (1.30X)	230
USPS	68.22	68.83	72.61	77.54	66.22	66.53	70.74	1.02	0.40	6.88	7.08	0.24	7.02	0.30 3.40X)	437
MNIST	71.95	53.27	68.03	69.88	57.24	29.86	72.27	6785	0.80	754	722	0.81	174.29	5.40 (**1,256X**)	569
Covtype	48.83	24.78	27.11	22.80	22.79	32.74	48.86	91,504	18.51	1,165	1,154	7.23	594.82	20.33 (**4,500X**)	456