[WORK IN PROGRESS]
Report on Paper "Compressive Spectral Clustering" by Nicolas Tremblay, Gilles Puy, Rémi Gribnoval, Pierre Vandergheynst

Sourabh Antani

Chapter 1

Introduction

This is a paper report on the paper 'Compressive Spectral Clustering' by by Nicolas Tremblay, Gilles Puy, Rémi Gribnoval and Pierre Vandergheynst [1]. All the sections in this report are taken or paraphrased from [1] and not original work. The structure of this report also follows the structure of the paper so that it is easy to follow.

According to the authors, Spectral Clustering has three computational bottlenecks when N and/or k is large: Creation of the similarity matrix W, partial eigendecomposition of L and k-means.

In order to circumvent these, sevaral authors have published various ideas. Following are few examples:

Power method [Boutsidis and Gittens, 2015], [Lin and Cohen, 2010]

A careful optimisation of diagonalisation algorithms in the context of SC [Liu et al., 2007]

Matrix column-subsampling such as in the Nyström method (Fowlkes et al., 2004)

nSPEC and cSPEC methods (Wang et al., 2009), or in (Chen and Cai, 2011;Sakai and Imiya, 2009)

Reduce k-means complexity Jain, 2010

Line of work on coresets (Har-Peled and Maxmudar, 2004)

Reduction of Graph by successive aggregation of nodes (Dhillon et al. 2007) (Filippone et al., 2008)

Compressive clustering circumvents two last bottlenecks using $\mathcal{O}(\log(k))$ randomly fitered signals on the graph to serve as feature vectors instead of eigenvectors and by clustering random subset of $\mathcal{O}(k\log(k))$ nodes using random feature vectors and infering the cluster label of all N nodes. Thus, the complexity

of k-means is reduced from $\mathcal{O}(Nk^2)$ to $\mathcal{O}(k^2 \log^2(k))$

Chapter 2

Background

2.1 Graph Signal Processing

Graph Fourier Matrix

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be an undirected graph with $\mathcal{V}$ the set of N nodes, $\mathcal{E}$ the set of edges and W set weighted adjacency matrix of non-negative weights. The normalized Laplacian of $\mathcal{G}$ is given by $L = I - D^{1/2} W D^{-1/2}$. Here I is $N \times N$ identity matrix and D is a diagnoal matrix with $D_{ii} = \sum_{j \neq i} W_{ij}$. L is real, symmetric and positive semi-definite, hence diagnizable as $L = U \Lambda U^T$ with orthonormal set of eigenvectors $u_1, \ldots, u_N$ and eigenvalues $0 = \lambda_1 \leq \cdots \leq \lambda_N \leq 2$. By analogy to continuous Laplacian operator with classical fourier modes as eigenfunctions and squared frequencies as eigenvalues, the eigenvectors are considered Fourier modes and square-roots of eigenvalues are considered frequencies of the Graph.

Graph filtering

The graph Fourier transform $\hat{x}$ of signal x is $\hat{x} = U^T x$. Given a continuous filter function h defined on $[0, 2]$, its associated graph filter operator $H \in \mathbb{R}^{N \times N}$ is defined as $H := h(L) = U h(\Lambda) U^T$ where $h(\Lambda) := diag(h(\lambda_1), \ldots, h(\lambda_N))$. The filtered signal is then Hx. Now, considering an ideal low-pass filter h_{λ_c} defined as below, H_{λ_c} is the graph filter operator associated with h_{λ_c}.

$$h_{\lambda_c} = \begin{cases} 1, & \text{if } \lambda \leq \lambda_c \\ 0, & \text{otherwise} \end{cases}$$

Fast graph filtering

To filter the signal by h without doagonalizing L, we approximate h by a polynomial of degree p $\tilde{h}(\lambda) = \sum_{l=0}^{p} \alpha_l L^l \simeq h(\lambda), \forall \lambda \in [0, 2], \alpha_1, \ldots, \alpha_p \in \mathbb{R}$. Thus $\tilde{H} := \tilde{h}(L) = \sum_{l=0}^{p} \alpha_l L^l \simeq H$. Thus instead of computing dense $\tilde{H}$ we approximate $\tilde{H}x = \sum_{l=0}^{p} \alpha_l L^l x$ using successive matrix-vector products with L.

Thus the computational complexity is $\mathcal{O}(pe)$ where e is number of edges.

Chapter 3

Principles of Compressive Spectral Clustering

3.1 Ideal filtering of random signals

Definition 3.1.1 (Local cumulative coherence)**.** : Given a graph $\mathcal{G}$, the local cumulative coherence of order k at node i is $v_k(i) = ||U_k^T \delta_i|| = \sqrt{\sum_{=1}^{k} U_{ij}^2}$, i.e. ℓ_2-norm of ith row of U

Next, the authors define the diagonal matrix V_k such that $V_k(i,i) = 1/v_k(i)$, assuming $v_k > 0$. Consider matrix $R = (r_1|r_2|\cdots|r_d) \in \mathbb{R}^{N\times d}$ consisting of d random signals r_i whose components are independent, Bernouilli, Gaussian or sparse random variables with mean zero and variance $1/d$. Considering the coherence=normalized filtered version of R, $V_k H_{\lambda_k}, R \in \mathbb{R}^{N\times d}$, and define node i's new feature vector $f_i \in \mathbb{R}^d$ as transposed i-th row of the filtered matrix, i.e. $\tilde{f}_i := (V_k H_{\lambda_k}, R)^T \delta_i$. The following theorem shows that, for a large enough d,

$$\tilde{D_{ij}} := ||\tilde{f}_i - \tilde{f}_j|| = ||(V_k H_{\lambda_k}, R)^T (\delta_i - \delta_j)||$$

is a good estimation of D_{ij} with high probability.

Theorem 1. Let $\epsilon \in [0,1]$ and $\beta > 0$ be given. If d larger than

$$\frac{4+2\beta}{\epsilon^2/2 - \epsilon^3/3} \log N,$$

then with probability at least $1 - N^{-\beta}$, we have

$$(1-\epsilon)D_{ij} \leq \tilde{D}_i j \leq (1+\epsilon)D_{ij}, \forall (i,j) \in \{1,\ldots,N\}^2$$

Proof. To be filled in. □

3.2 Downsampling and interpolation

Let $c_j \in \mathbb{R}^N, j = 1,\ldots,k$ be the indicator vectors of clusters C_j. The authors propose to estimate c_j by running k-means on small subset of n feature vectors only using 1) low-dimentional model that captures the regularity of c_j, 2) make sure enough information is preserved after sampling, 3) algorithem that rapidly and accuraely estimates the vectors c_j.

3.2.1 The Low-Dimentional Model

For a simple regular graph (graph with nodes of same degree) with k disconnected clusters, one can see that the indicator vectors c_i form a set of k orthogonal eigenvectors of L with eigenvalue 0. Thus, al the indicator vectors live in $span(U_k)$. For general graph the authors assume that the indicator vectors live close to $span(U_k)$, i.e. there is a slight perturbation. In graph signal processing words, one can say that c_j is approximately k-bandlimited. i.e. the first k graph Fourier coefficients bear most of its energy.

3.2.2 Sampling and Interpolation

Bibliography

[1] Nicolas Tremblay, Gilles Puy, Remi Gribonval, and Pierre Vandergheynst. Compressive spectral clustering, arXiv:1602.02018, 2016.

[2] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[3] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.

[4] B. N. Parlett. *The Symmetric Eigenvalue Problem*. Number 20 in Classics in Applied Mathematics. SIAM, Philadelphia, PA, 1998.

[5] G. W. Stewart and J. G. Sun. *Matrix perturbation theory*. Academic Press, Boston, MA, 1990.

[6] G. Cheung, E. Magli, Y. Tanaka, and M. K. Ng. Graph spectral image processing. *Proceedings of the IEEE*, 106(5):907–930, 2018.