Sparsest Cut Problem in Graphs and its Parallel Algorithms

By Saleh Ashkboos

Supervisor: Dr. Ramin Javadi - Dr. Masoud Reza Hashemi

Abstract

In this project we are studying the sparsest cut problem in graphs and its application.
We study:

- Formulation of the problem and it various generalization.
- Spectral algorithms and their applications in data clustering
- An efficient algorithms on weighted trees
- An implementation of the above algorithm in parallel using CUDA platform
- Experimental results on various data sets

Introduction

- The *edge expansion* ("sparsity" or "Cheeger constant") of a graph G is:

$$\phi(G) := \frac{E(S,\bar{S})}{\min(|S|,|\bar{S}|)}$$

The ***Sparsest Cut*** problem asks to find a cut $E(S,\bar{S})$ with smallest possible sparsity.

- For a subset of vertices A in a graph G, the normalized flow of A is:

$$\frac{E(S,\bar{S})}{|S|}$$

The mean version of the problem seeks for a cut which minimize the average normalized flows of the sets S and $\bar{S}$.

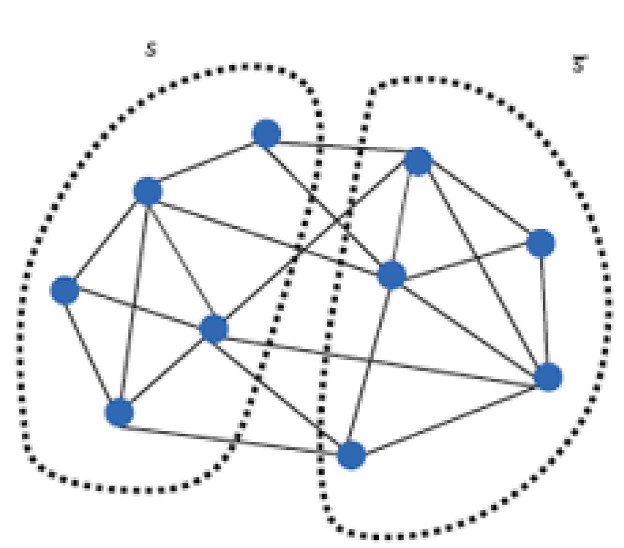

Figure 1: Sparsest Cut Solution

Problem Generalization

- k-Normalized Cut Problem:
Given a weighted graph $G = (V, E, \varphi, \omega)$ and an integer k. Find a k-partition of vertices G such that the sum (maximum) of their normalized flows is minimized.

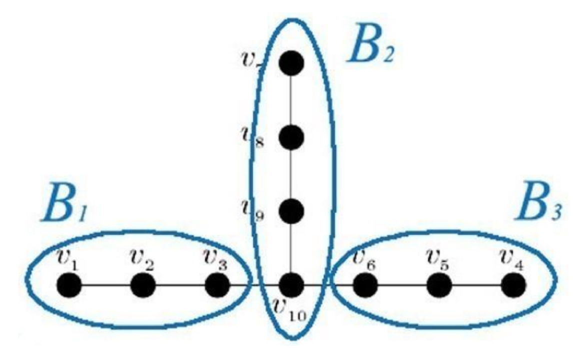

Figure 2: 3-Normalized Cut Solution

- k-Isopertimetric Problem:
The isoperimetric problem aims to minimize the normalized cut objective function over the space of subpartitions.

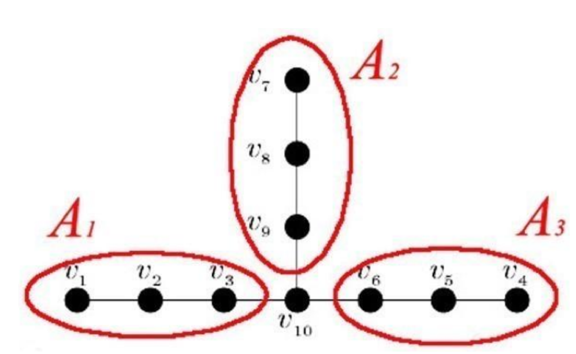

Figure 3: 3-Isoperimetric Solution

Relation to Laplacian Spectrum

Let $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{|V|}$ be the Laplacian eigenvalues of a graph G.

- *Cheeger's inequality* offers the following quantitative connection betwen $\phi(G)$ and λ_2:

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$$

- In [2], Lee et al. show the following relation for k-normalized cut problem:

$$\frac{\lambda_k}{2} \leq \phi^k(G) \leq O(k^2)\sqrt{\lambda_k}$$

- In [4], Shi and Malik, use the second Laplacian eigenvector to approximate the mean normalized cut problem.

Spectral Algorithm

- In [3], Ng et al. proposed the following approximation algorithm for k-Normalized cut problem:

1. Compute the first k eigenvectors $u_1, \ldots, u_k$ of Laplacian matrix.
2. Let $U \in R^{n*k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns.
3. For $i = 1, \ldots, n$, let $y_i \in R^k$ be the vector corresponding to the i-th row of U.
4. Cluster the points $(y_i)_{i=1,\ldots,n}$ in R^k with the k-means algorithm into clusters $C_1, \ldots, C_k$.

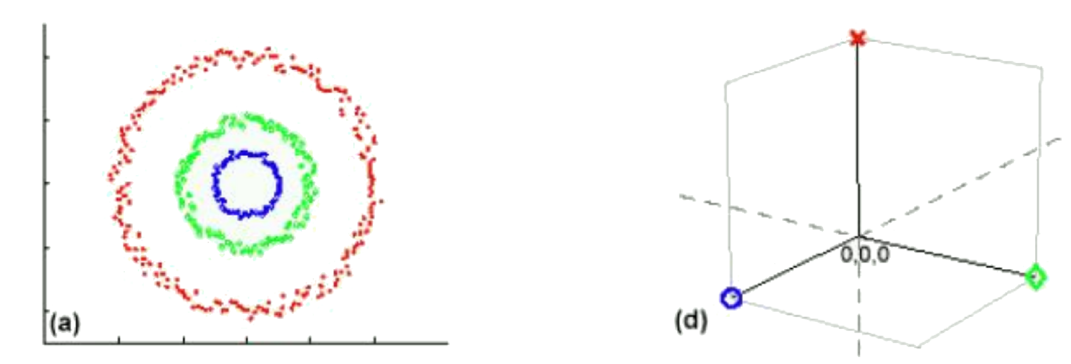

Figure 4: Spectral Projection of Data in the Algorithm

- The approximation factor of this algorithm can be bounded by higher order Cheeger's inequalities.
- The algorithm can be used as an alternative algorithm for data clustering tasks.

Algorithms on Trees

In [1], Daneshgar et al. show that the maximum k-Isoperimetric problem for weighted trees can be solved and an affirmative subpartition can be found (if exists) in time $O(n \log n)$ and proposed a clustering algorithm as follows:

1. Construct similarity graph G,
2. Compute minimum spanning tree of G,
3. Solve the k-Isoperimetric problem using the algorithm in [1].

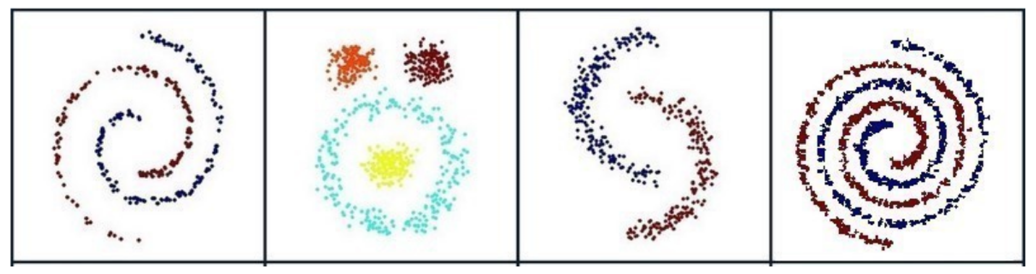

Figure 5: Experimental Results of Algorithm [1]

Parallel Implementation and Experimental Results

- We implement the first two steps of the above algorithm in parallel,
- The implementation is done using CUDA platform on Nvidia Geforce GTX 850M card.
- The sketch of the algorithm is:

1. Find all matrix elements independently in parallel,
2. Find the minimum spanning tree in parallel using algorithm in [5],
3. Solve the k-Isoperimetric problem using algorithm in [1].

Data Set	n	k	d	CPU(ms)	GPU (ms)	Speed-Up	Misclassification rate
Wine	178	3	13	12.51	58.40	0.21	0.281
Skin	2000	2	3	373	350	1.06	0.014
Yeast	1484	10	8	323	300	1.07	0.686
Wine Quality	4898	10	11	4161	956	4.35	0.844
Internet Ads	3279	2	1558	177291	9412	18.83	0.482

Figure 6: Experimental Results

References

[1] Amir Daneshgar, Ramin Javadi, and SB Shariat Razavi, *Clustering and outlier detection using isoperimetric number of trees*, Pattern Recognition **46** (2013), no. 12, 3371–3382.

[2] James R Lee, Shayan Oveis Gharan, and Luca Trevisan, *Multiway spectral partitioning and higher-order cheeger inequalities*, Journal of the ACM (JACM) **61** (2014), no. 6, 37.

[3] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al., *On spectral clustering: Analysis and an algorithm*, Advances in neural information processing systems **2** (2002), 849–856.

[4] Jianbo Shi and Jitendra Malik, *Normalized cuts and image segmentation*, Pattern Analysis and Machine Intelligence, IEEE Transactions on **22** (2000), no. 8, 888–905.

[5] Wei Wang, Yongzhong Huang, and Shaozhong Guo, *Design and implementation of GPU-based Prim's algorithm*, International Journal of Modern Education and Computer Science (IJMECS) **3** (2011), no. 4, 55.