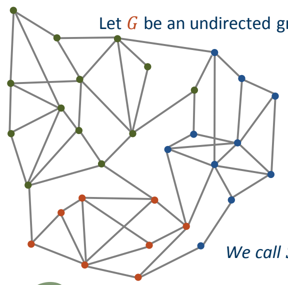
Partitioning Well-Clustered Graphs: Spectral Clustering Works!

Richard Peng (MIT) He Sun (University of Bristol) Luca Zanetti (University of Bristol)

Talk on Saturday 10:20AM, Session 9

1 Problem



Let G be an undirected graph with n vertices and m edges. Define the conductance of set S by

$$\phi_G(S) = \frac{|E(S, V\setminus S)|}{\mathrm{vol}(S)},$$

where $\mathrm{vol}(S) = \sum_{u\in S} d_u$, and define the k-way expansion by

$$\rho(k) = \min_{\text{partition } A_1,\cdots,A_k} \max_{1\le i\le k} \phi_G(A_i).$$

We call $S_1,\cdots,S_k$ achieving $\rho(k)$ an optimal partition.

? Partition G into k clusters such that each cluster is close to its correspondence in the optimal partition.

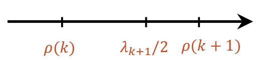
This problem is closely related to minimizing the normalized cut of a graph, and has comprehensive applications in community detection, image segmentation, as well as designing fast algorithms.

2 Well-clustered graphs

Let L be the normalized Laplacian matrix of G, with eigenvalues $0 = \lambda_1 \le \cdots \le \lambda_n$ and the corresponding eigenvectors $f_1,\cdots,f_n$.

Higher-Order Cheeger Inequality
Lee et al., 2012

$$\frac{\lambda_k}{2} \le \rho(k) \le O(k^3)\sqrt{\lambda_k}$$



By the higher-order Cheeger inequality, a large value of $\lambda_{k+1}/\rho(k)$ implies
(1) existence of a k-way partition with bounded $\rho(k)$, and
(1) any $(k+1)$-way partition contains a set of high conductance

Gap assumption: a lower bound on the ratio $\Upsilon = \lambda_{k+1}/\rho(k)$

We call a graph well-clustered if it naturally exhibits a k-cluster structure, i.e. any $(k+1)$-way partition will introduce a subset of high conductance. A well-clustered graph can be captured by the gap assumption, which is also closely related to the gap between λ_{k+1} and λ_k. The value of k for which there is a gap between λ_{k+1} and λ_k has been observed empirically as an indication of the correct number of clusters.

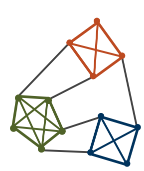
3 Provable guarantees for spectral clustering

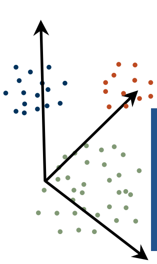
A widely used approach for graph partitioning is spectral clustering, which can be described as follows:

(1) map every vertex u to $F(u) \in \mathbb{R}^k$ where

$$F(u) = \frac{1}{\mathrm{NormalizationFactor}(u)} \cdot (f_1(u),\cdots,f_k(u))$$

(2) apply a k-means algorithm on embedded points in $\mathbb{R}^k$.





This approach has been the subject of extensive experimental studies for more than 20 years. Prior to our work, rigorous analysis was known only for graphs generated from stochastic models.

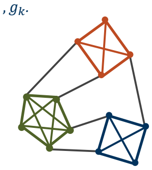
Main Result: Let $\Upsilon = \Omega(k^3)$, and $A_1,\cdots,A_k$ be the output of spectral clustering. For any $1\le i\le k$, the following holds:
(1) Symmetric difference between A_i and S_i is bounded: $\mathrm{vol}(A_i \triangle S_i) = O(k^3\cdot \mathrm{vol}(S_i)/\Upsilon)$;
(2) $\phi_G(A_i) = O(\phi_G(S_i) + k^3/\Upsilon)$.

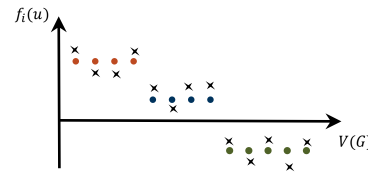
4 The structure theorem

Let S_i be a cluster, and define its indicator vector g_i by

$$g_i(u) = \begin{cases} \frac{\sqrt{d_u}}{\mathrm{vol}(S_i)} & \text{if } u\in S_i, \\ 0 & \text{otherwise.}\end{cases}$$

We prove that for well-clustered graphs every eigenvector f_i for $1\le i\le k$, represented by starred points below, can be approximated by a linear combination of the indicator vectors $g_1,\cdots,g_k$.





The structure theorem: Let $\Upsilon = \Omega(k^2)$. Then the following statements hold for any $1\le i\le k$:
- There is a linear combination $\widehat{f}_i$ of $\{f_\ell\}_{\ell=1}^k$ such that $\|g_i - \widehat{f}_i\|^2 \le 1/\Upsilon$.
- There is a linear combination $\widehat{g}_i$ of $\{g_\ell\}_{\ell=1}^k$ such that $\|f_i - \widehat{g}_i\|^2 \le k/\Upsilon$.

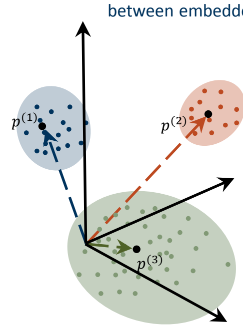
5 Well-separation properties of spectral embedding

Map vertex u to d_u identical points $F(u)$, where $F(u) = \frac{1}{\sqrt{d_u}}\cdot(f_1(u),\cdots,f_k(u))$.

The structure theorem and the gap assumption imply the following facts:

(1) Every cluster S_i is concentrated around a point $p^{(i)}$, and the total distance between embedded points to their closest $p^{(i)}$ is proportional to $1/\Upsilon$:

$$\sum_{i=1}^{k}\sum_{u\in S_i} d_u \left\|F(u) - p^{(i)}\right\|^2 \le k^2/\Upsilon$$



(2) Distance between different clusters is inversely proportional to the **smaller** cluster:

$$\left\|p^{(i)} - p^{(j)}\right\|^2 \ge \frac{1}{k\cdot \min\{\mathrm{vol}(S_i), \mathrm{vol}(S_j)\}}$$

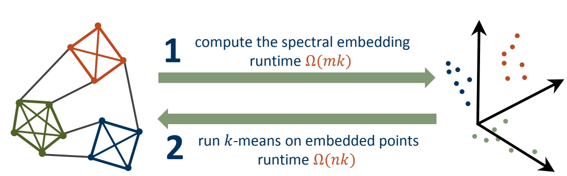
(3) Bigger clusters are closer to the origin:

$$\left\|p^{(i)}\right\|^2 \approx \frac{1}{\mathrm{vol}(S_i)}$$

These properties show that an optimal solution of the k-means problem is close to an optimal partition of the graph, and a reasonable approximation works as well. This leads to our main result.

6 A faster algorithm

We present an $\tilde{O}(m)$ time algorithm for partitioning G into k clusters. Notice that spectral clustering runs in super-linear time for large k due to the following reasons:



To break these two barriers, (1) we show that pairwise distances of any two vertices can be approximately computed in nearly-linear time via the heat kernel embedding, and (2) we develop an ad-hoc version of k-means algorithms, which runs in nearly-linear time.