
Local Spectral Clustering of Density Upper Level Sets

Anonymous Authors¹

Abstract

1. Introduction

Clustering is a fundamental task in machine learning, which often involves splitting data into groups which satisfy some notion of within-group similarity and between-group difference. In particular, spectral clustering methods are a family of powerful non-parametric clustering algorithms.

Roughly speaking, spectral clustering techniques embed data points using the spectrum of (some form of) the graph Laplacian matrix and subsequently use this *spectral embedding* to find a clustering of the data. When applied to large graphs (or large point clouds) classical global spectral methods can be computationally cumbersome and can be insensitive to the local geometry of the distribution of the samples (Mahoney et al., 2012; Leskovec et al., 2010). This in turn has led to the investigation of local spectral algorithms (Spielman & Teng, 2013; Andersen et al., 2006; Leskovec et al., 2010) which leverage locally-biased spectra computed using random walks around a user-specified seed node.

A natural model to consider when analyzing point cloud data is based on the hypothesis that we receive samples $\{x_1, \dots, x_n\}$ independently drawn from a distribution f . In this case, we are interested in understanding what the output of a clustering algorithm on this finite sample reveals about the unknown distribution f . In particular, it is intuitive (Hartigan, 1981; Chaudhuri & Dasgupta, 2010) to define clusters in this context as *connected regions of high-density*, and study the ability of spectral methods to identify these clusters.

Given a matrix A with entries representing the similarity between pairs of data points, and $G = (V, E)$ a graph representation of A with vertices V and edges E , there are many (Yang & Leskovec, 2015; Fortunato, 2010) graph-theoretic measures which assess the cluster quality of a

subset $S \subseteq V$ (or more generally the quality of a partition $S_1 \cup \dots \cup S_m = V$, for $m \geq 2$.)

Arguably a natural way to assess cluster quality is via the bicriteria of *external* and *internal connectivity* of S , represented by the parameters Φ and Ψ , respectively. As the names suggest, external connectivity should relate to the number of edges between S and G/S (hereafter denoted S^c), while internal connectivity in turn measures the number of edges between subsets within S ¹. The clustering task then becomes to find a subset S (or, for global algorithms, a partition $S_1 \cup \dots \cup S_m = V$), which has both small external and large internal connectivity. We will refer to such a subset as a (Φ, Ψ) -cluster (or partition). Both local (Zhu et al., 2013) and global (Kannan et al., 2004) spectral algorithms can be understood as outputting clusters (or partitions) which provably satisfy approximations to the optimal (Φ, Ψ) -cluster (or partition).²

In the case where G encodes similarities between sampled point-cloud data, some understanding of how graph theoretic quantities such as Φ and Ψ relate to properties of the underlying distribution f has developed. Continuous analogues to many such graph measures of cluster quality have been defined, and, under appropriate graph construction, convergence of clusters output by spectral algorithm to approximate minimizers of these continuous analogues has been shown (von Luxburg et al., 2008; Trillos & Slepcev, 2018).

Relating these partitions to the arguably more simply defined high density clusters can be challenging in general. Intuitively, however, under the right conditions such a connection should exist. Given sufficient samples, and an appropriately specified notion of similarity, we can imagine that all points within a high density cluster will have many edges to other points within the cluster, and comparatively few edges to data points outside the cluster. Motivated by this intuition, one primary line of work in this article is to formally establish that, under a natural set of geometric conditions, a connection between certain graph-based criteria and density clusters does exist.

We also study the consequences of these facts as they re-

¹Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

¹We formally define our notions of external and internal connectivity in Section 2

²For slightly different definitions of the internal connectivity Ψ

late to local spectral clustering algorithms. We restrict our attention in particular to one such popular algorithm: *personalized PageRank* (PPR). We show that a variant of PPR, properly initialized inside a high-density cluster, performs well: not only does the output set of PPR approximate the cluster quality measures of the high-density cluster, the PPR algorithm also consistently recovers the population density cluster.

1.1. Related Work

1.2. Summary of Results

1. Theorem: Well-conditioned density clusters have low conductance.
2. Theorem: Well-conditioned density clusters have high internal connectivity.
3. Corollary: PPR local clustering algorithm, run on a well-conditioned density cluster and appropriately initialized, will output a low conductance set, which also has low symmetric set difference with the density cluster.
4. Theorem: PPR local clustering algorithm, run on a well-conditioned density cluster and appropriately initialized, exhibits successful density cluster recovery.

Graph notation. Let $G = (V, E)$ be an undirected, unweighted graph, with $S, S' \subseteq V$. The volume of S is given by $\text{vol}(S; G) = \sum_{v \in S} \deg(v; G)$ where $\deg(v; G) = \sum_{u \in V} 1((u, v) \in E)$ is the degree of v in G . For the random walk over G , denote the stationary distribution by π , where $\pi(S; G) = \frac{\text{vol}(S; G)}{\text{vol}(V; G)}$.

For the sample $\mathbf{X} := \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ and a set $A \subset \mathbb{R}^d$, define $A[\mathbf{X}] = A \cap \mathbf{X}$. We denote the uniform measure over \mathbb{R}^d by $\nu(\cdot)$.

2. Measures of cluster quality

Let P be a distribution supported on a compact set $\mathcal{X} \subset \mathbb{R}^d$, with continuous density function f (with respect to the uniform measure). The Euclidean distance is denoted by $\|\cdot\|$.

Density upper-level set. For a number $\tau \geq 0$, let $C_f(\tau)$ be the collection of collected components of the density upper-level set $\{x \in \mathcal{X} : f(x) \geq \tau\}$. Define a τ -density cluster to be one such connected component $A \in C_f(\tau)$. We will sometimes refer to $A[\mathbf{X}]$ as an *empirical density cluster*.

Graph bicriteria. For $G = (V, E)$ an undirected, unweighted graph and $S, S' \subset V$ as before, let $|E(S, S'; G)|$ denote the cut of S and S' , given by

$$|E(S, S'; G)| = \sum_{v \in S} \sum_{u \in S'} 1((v, u) \in E).$$

Define the balance $B(S; G)$ to be

$$B(S; G) = \min\{\text{vol}(S; G), \text{vol}(V \setminus S; G)\}.$$

We can now formally introduce our first criterion for assessing the quality of graph cuts: the **conductance**, Φ , given by

$$\Phi(S; G) \stackrel{\text{def}}{=} \frac{|E(S, V \setminus S; G)|}{B(S; G)}. \quad (1)$$

We typically seek a set $S^* \subset V$ such that $\Phi(S^*; G)$ is small.

Introduce and define the internal connectivity.

Neighborhood graph. Given $r \geq 0$, define the neighborhood graph to be $G_{n,r} = (\mathbf{X}, E_n)$, where for $x_i, x_j \in \mathbf{X}$, $(x_i, x_j) \in E_n$ if $\|x_i - x_j\| \leq r$. (By convention, we do not allow loops, meaning $(x_i, x_i) \notin E_n$). For ease of notation, for $S \subset \mathbf{X}$, let $\Phi_{n,r}(S) := \Phi(S; G_{n,r})$.

2.1. Well-conditioned density clusters.

In order to satisfy the **bicriteria**, we must introduce some assumptions on the density f . Let $B(x, r)$ be a closed ball with respect to Euclidean distance) of radius r around the point x . Given a set $A \subset \mathbb{R}^d$, and a number $\sigma > 0$, define the set $A_\sigma = A + B(0, \sigma) = \{y \in \mathbb{R}^d : \inf_{x \in A} \|y - x\| \leq \sigma\}$.

(A1) *Minimum density:* For numbers $\sigma, \tau_\sigma > 0$, a set $A \subset \mathcal{X}$ is said to be (σ, τ_σ) -regular if

$$\inf_{x \in A_\sigma} f(x) = \tau_\sigma \quad (2)$$

(A2) *Low noise density:* For numbers $\gamma, \sigma > 0$, and a set $A \subset \mathcal{X}$, the density function f is said to be (γ, σ) -low noise around A if there exists a constant C_1 such that for all $x \in \mathcal{X}$ with $0 < \rho(x, A_\sigma) \leq \sigma$,

$$\inf_{x' \in A_\sigma} f(x') - f(x) \geq C_1 \rho(x, A_\sigma)^\gamma.$$

where $\rho(x, A_\sigma) = \min_{x_0 \in A_\sigma} \|x - x_0\|$.

(A3) *Cluster separation:* For $\tau, \sigma > 0$, $C_f(\tau)$ the set of connected components of the density upper-level set is said to be σ -well separated if for all τ -density clusters $A, A' \in C_f(\tau)$,

$$\rho(A, A') > \sigma.$$

where $\rho(A, A') = \min_{x \in A, x' \in A'} \|x - x'\|$.

Assumptions are standard, provide references. We note that these assumptions are all local in nature, further motivating the study of a local algorithm. Assumptions (A1)-(A3) are used to upper bound **bicriteria 1**, whereas (A1) and **(A4)** are necessary to lower bound **bicriteria 2**.

Definition 1. For $\tau > 0$ and $A \in C_f(\tau)$ a τ -density cluster, we say that A is a $(\lambda, \sigma, \gamma)$ -well-conditioned cluster if A is (λ, σ) -regular, the density f is (γ, σ) -low noise around A^σ , and $C_f(\tau)$ is σ -well separated.

We note that σ plays dual roles here, both in effect precluding arbitrarily narrow and long clusters in (A1) and arbitrarily flat densities in (A2). **We give some examples of densities which satisfy these assumptions in Section 4.**

3. Local Clustering on Density Level Sets

In this section we show that a well-conditioned cluster A satisfies the bicriteria **criteria 1** and **criteria 2**, and therefore the **PPR algorithm** outputs a low conductance set which has small symmetric set difference with the empirical cluster.

Theorem 1. For some $\tau > 0$, let a τ -density cluster $A \in C_f(\tau)$ satisfy Assumptions (A1)-(A3) for some $\sigma, \tau_\sigma, \gamma > 0$. Then, for any $r < \sigma$ and $\delta > 0$, the following statements hold with probability at least $1 - \delta$:

- **Additive error bound.** Fix $\epsilon > 0$. Then, for

$$n \geq \log(2/\delta) \left(\frac{1 + \epsilon/2}{2\tau_\sigma^2 \nu(A_\sigma) \nu_d(r/2)^d} \right)^2 \quad (3)$$

we have

$$\frac{\Phi_{n,r}(A_\sigma[\mathbf{X}])}{r} \leq C_\sigma \frac{\tau}{\tau_\sigma} \frac{(\tau_\sigma - \frac{r^{\gamma+1}}{\gamma+1})}{\tau_\sigma} + \epsilon \quad (4)$$

- **Multiplicative error bound.**

where $C_0 = 2^{2d+1}d$ and $C_\sigma = C_0/\sigma$.

A few remarks are in order.

Remark 1. Note that the bound of (4) depends exponentially on d ; precisely, on $C_0 = d2^{2d+1}$. It is possible to improve this dependency to the order of $(1 + \frac{r}{\sigma})^{2d}$. However, in this setting, we think of r as being a constant radius (rather than $r = r_n \rightarrow 0$ as $n \rightarrow \infty$, and therefore even this improvement results in exponential dependency on the dimension d . **Make references suggesting this type of exponential dependency is not uncommon in non-parametric clustering problems.**

Remark 2. Other than the looseness implied by Remark 1, the error bound of (4) is almost tight. Specifically, choosing

$$A_\sigma = B(0, \sigma),$$

$$f(x) = \begin{cases} \tau & \text{for } x \in A_\sigma, \\ f(x) = \tau - \rho(x, A_\sigma)^\gamma & \text{for } 0 < \rho(x, A_\sigma) < r \end{cases}$$

we have

$$\frac{\Phi_{n,r}(A_\sigma[\mathbf{X}])}{r} \geq C_1 \frac{(\tau - \frac{r^{\epsilon+1}}{\epsilon+1})}{\tau} \quad (5)$$

for some constant C_1 , with probability at least $1 - \delta$. (Note that a factor of $1/\sigma$ is not (4) replicated in this lower bound.) **Provide justification in supplement, and cite it.**

4. Implications, extensions, and discussion

4.1. Examples

4.2. Experiments

References

- Andersen, R., Chung, F., and Lang, K. Local graph partitioning using pagerank vectors. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pp. 475–486, 2006.
- Balakrishnan, S., Narayanan, S., Rinaldo, A., Singh, A., and Wasserman, L. Cluster trees on manifolds. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS’13, pp. 2679–2687, USA, 2013. Curran Associates Inc. URL <http://dl.acm.org/citation.cfm?id=2999792.2999911>.
- Chaudhuri, K. and Dasgupta, S. Rates of convergence for the cluster tree. In Lafferty, J. D., Williams, C. K. I., Shawe-Taylor, J., Zemel, R. S., and Culotta, A. (eds.), *Advances in Neural Information Processing Systems 23*, pp. 343–351. Curran Associates, Inc., 2010.
- Fortunato, S. Community detection in graphs. *Physics Reports*, 486(3):75 – 174, 2010. ISSN 0370-1573. doi: <https://doi.org/10.1016/j.physrep.2009.11.002>. URL <http://www.sciencedirect.com/science/article/pii/S0370157309002841>.
- Hartigan, J. A. Consistency of single-linkage for high-density clusters. *Journal of the American Statistical Association*, 1981.
- Kannan, R., Vempala, S., and Vetta, A. On clusterings: Good, bad and spectral. *J. ACM*, 51(3):497–515, May 2004. ISSN 0004-5411. doi: 10.1145/990308.990313. URL <http://doi.acm.org/10.1145/990308.990313>.
- Leskovec, J., Lang, K. J., and Mahoney, M. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th International Conference on World Wide Web*, 2010.
- Mahoney, M. W., Orecchia, L., and Vishnoi, N. K. A local spectral method for graphs: with applications to improving graph partitions and exploring data graphs locally. *Journal of Machine Learning Research*, 13:2339–2365, 2012.
- Spielman, D. A. and Teng, S.-H. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM Journal on Computing*, 42(1):1–26, 2013.
- Trillos, N. G. and Slepcev, D. A variational approach to the consistency of spectral clustering. *Applied and Computational Harmonic Analysis*, 45(2):239 – 281, 2018. ISSN 1063-5203. doi: <https://doi.org/10.1016/j.acha.2016.09.003>. URL <http://www.sciencedirect.com/science/article/pii/S106352031630063X>.
- von Luxburg, U., Belkin, M., and Bousquet, O. Consistency of spectral clustering. *Ann. Statist.*, 36(2):555–586, 04 2008.
- Yang, J. and Leskovec, J. Defining and evaluating network communities based on ground-truth. *Knowledge and Information Systems*, 42(1):181–213, Jan 2015. ISSN 0219-3116. doi: 10.1007/s10115-013-0693-z. URL <https://doi.org/10.1007/s10115-013-0693-z>.
- Zhu, Z. A., Lattanzi, S., and Mirrokni, V. S. A local algorithm for finding well-connected clusters. In *ICML (3)*, pp. 396–404, 2013.