Statistical Analysis and Data Mining: An ASA Data Science Journal

WILEY

Check for updates

RAPID PUBLICATION OPEN ACCESS

Recovering the Number of Clusters From a Laplacian Matrix by Nuclear Norm Penalization

Cinzia Di Nuzzo[1] | Matteo Farnè[2]

[1]Department of Economics and Business, University of Catania, Catania, Italy | [2]Department of Statistical Sciences, University of Bologna, Bologna, Italy

Correspondence: Cinzia Di Nuzzo (cinzia.dinuzzo@unict.it)

Received: 28 March 2025 | **Revised:** 15 May 2025 | **Accepted:** 3 June 2025

Funding: This work was supported by Cinzia di Nuzzo gratefully acknowledges the support of "European Union-NextGenerationEU," in the framework of "GRINS-Growing Resilient, INclusive and Sustainable" project (Project code PE00000018-CUP E63C22002120006). Matteo Farnè gratefully acknowledges the support of the Italian Ministry of University and Research, in the framework of PRIN project "Latent variable models and dimensionality reduction methods for complex data" (Project code 20224CRB9E-CUP B53C24006310006).

Keywords: internal cluster validation | Laplacian embedding | low-rank representation | nuclear norm

ABSTRACT

Spectral clustering is a powerful technique for data partitioning, but determining the optimal number of clusters remains challenging. This article introduces ALLE (ALgebraic Laplacian Estimator), an automatic method for estimating the number of clusters within the spectral clustering framework. By formulating the cluster recovery problem as a penalized minimization task, ALLE is able to systematically recover the number of clusters and the embedding space by assuming for the Laplacian matrix a low-rank plus sparse decomposition. Specifically, ALLE recovers the low-rank representation of the Laplacian matrix using nuclear norm plus ℓ_1-norm penalization. ALLE is computed via a proximal gradient algorithm alternating Singular Value Thresholding and Soft Thresholding, and it's very good performance is shown via a simulation study.

1 | Introduction and Background Theory

Spectral clustering is a technique that is employed to partition data into groups based on the spectral properties of a similarity matrix. This method is particularly effective in capturing complex and non-linear structures without assuming any predefined cluster shape [1, 2]. Unlike traditional clustering techniques such as k-means, spectral clustering relies on the eigenvector decomposition of the Laplacian matrix, projecting data into a lower-dimensional space where the underlying clustering structures are clearly highlighted. Once this feature space is obtained, a clustering method can be applied to identify groups. The integration of dimensionality reduction with clustering not only enhances computational efficiency but also improves the visualization of data representation.

More formally, given a set of objects $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathbb{R}^P$ and a similarity matrix $\mathbf{W} = (w_{ij})$ (for $i, j = 1, \ldots, N$), the goal of spectral clustering is to partition the data $\mathbf{X}$ into K disjoint clusters $\{\mathbf{G}_1, \ldots, \mathbf{G}_K\}$, maximizing between-cluster dissimilarity while minimizing within-cluster dissimilarity [3]. This objective function is formulated using the K *normalized cut (Ncut)*:

$$\mathrm{Ncut}\{\mathbf{G}_1, \ldots, \mathbf{G}_K\} = \sum_{k=1}^{K} \frac{\mathrm{cut}(\mathbf{G}_k, \mathbf{X}\backslash\mathbf{G}_k)}{\mathrm{vol}(\mathbf{G}_k)} \tag{1}$$

where $\mathrm{cut}(\mathbf{G}_k, \mathbf{X}\backslash\mathbf{G}_k) = \sum_{\mathbf{x}_i \in \mathbf{G}_k, \mathbf{x}_j \in \mathbf{X}\backslash\mathbf{G}_k} w_{ij}$ and $\mathrm{vol}(\mathbf{G}_k) = \sum_{i \in \mathbf{G}_k} \mathbf{D}_i$, with $\mathbf{D} = \mathrm{diag}(\mathbf{d}_1, \ldots, \mathbf{d}_N)$, where $\mathbf{d}_i = \sum_{j=1}^{N} w_{ij}$ (for $i, j = 1, \ldots, N$). Rewriting (1) with an indicator matrix $\mathbf{A}$ results

Abbreviation: ALLE, ALgebraic Laplacian Estimator.

Cinzia Di Nuzzo and Matteo Farnè are equally contributing authors.

in a Rayleigh quotient formulation, whose relaxed minimization is given by

$$\min_{\mathbf{A}\in\mathbb{R}^{N\times K}} \operatorname{tr}\left(\mathbf{A}^\top \mathbf{L}_{\text{sym}} \mathbf{A}\right), \quad \text{subject to}\, \mathbf{A}^\top\mathbf{A} = \mathbf{I}$$

where $\mathbf{L}_{\text{sym}} = \mathbf{I}_N - \mathbf{L}$, with $\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$, is the normalized Laplacian matrix. This minimization is equivalent to

$$\min_{\mathbf{A}\in\mathbb{R}^{N\times K}} \|\mathbf{L} - \mathbf{A}\mathbf{A}^\top\|_2^2, \quad \text{subject to } \mathbf{A}^\top\mathbf{A} = \mathbf{I} \tag{2}$$

which defines the *Laplacian embedding* $\mathbf{A}$, where $\|\cdot\|_2$ is the Euclidean norm. The matrix $\mathbf{A}$ represents the new feature space where each row corresponds to a data object, transformed according to the first K eigenvectors of $\mathbf{L}$ associated with the largest K eigenvalues. This transformation captures the geometric structure of the clusters. Finally, a clustering algorithm is applied to this reduced representation to assign data objects to clusters [1, 2, 4].

One of the fundamental challenges in spectral clustering—similar to most clustering algorithms—is the selection of the optimal number of clusters K. In this particular case, the challenge is to select K directly from the Laplacian matrix. This problem is often nontrivial, requiring either prior knowledge or heuristic approaches that may not always be reliable, see Di Nuzzo and Ingrassia [4], Zelnik-Manor and Perona [5], and John et al. [6]. In this article, we propose a new automatic method for estimating the number of clusters within the spectral clustering framework, called ALLE (ALgebraic Laplacian Estimator), which comes from ALCE (ALgebr aic Covariance Estimator, Farnè and Montanari [7, 8], a high-dimensional covariance matrix estimator based on nuclear norm plus ℓ_1-norm penalization. In this way, we are able to identify the low-dimensional embedding space, which is a low-rank representation of the Laplacian matrix associated to the clustering partition.

Our approach is based on a methodology originally introduced to recover high-dimensional exact [9] or perturbed covariance matrices [7, 10, 11], precision matrices [12], or spectral density matrices ([8]), by penalizing a nuclear norm plus ℓ_1-norm heuristics. Such a method is able to recover any matrix sum on which a proper low-rank plus sparse decomposition is imposed.

Nuclear norm penalization takes origin from the work of Fazel et al. [13], as an alternative to principal component analysis for low rank approximation. In Fazel [14], it is proved that nuclear norm penalization is the tightest convex relaxation of the original latent rank penalization, which is NP-hard. Similarly, in Donoho [15] it is showed that the ℓ_1-norm penalization is the tightest convex relaxation of the original ℓ_0-norm penalization, which also is NP-hard.

Relevant literature on nuclear norm plus ℓ_1-norm heuristics includes Candès et al. [16], which proposes a robust version of principal component analysis, where the low rank subspace is allowed to be perturbed by a random matrix. In the spectral clustering context, we impose a similar model: the nuclear norm of the low rank component serves to recover the latent rank (which corresponds here to the number of clusters K), while the ℓ_1-norm of the residual serves to account for the noise.

The remainder of the article is organized as follows. In Section 2, we introduce our approach for estimating the low-rank Laplacian embedding and the number of clusters. Section 3 presents a simulation study, while Section 4 discusses conclusions and future research directions within this framework.

2 | ALLE: Recovering the Number of Clusters

In the spectral clustering framework, the number of clusters is equal to the number of dimensions of the Laplacian embedding, because projection into the reduced space is carried out considering the first K eigenvectors of the Laplacian matrix $\mathbf{L}$. This implies that the structure of the clusters is directly determined by the effective rank of the Laplacian matrix, which governs the partitioning of the data in the latent Laplacian embedding space.

In our proposal, the low rank plus sparse assumption is imposed on the Laplacian matrix, which allows the dimensionality reduction step to be performed by the nuclear norm plus ℓ_1-norm heuristics. This assumption is imposed through the spectral decomposition of the $N \times N$ Laplacian matrix, producing an embedding of dimension $N \times K$, where K is the number of clusters. Traditionally, K is either assumed a priori or determined using heuristic techniques. In contrast, our approach proposes an adaptive estimation of K by employing the nuclear norm plus ℓ_1-norm penalization framework.

Given these premises, we can formulate the ALLE (ALgebraic Laplacian Estimator) optimization problem as in (4). Problem (4) is typically solved by a proximal gradient algorithm, originally proposed by Luo [11], which follows the general acceleration scheme of Nesterov [17]. The algorithm alternates Singular Value Thresholding [18] for the nuclear norm part, and the Soft-Thresholding algorithm, originally codified as FISTA in Daubechies et al. [19], for the ℓ_1-norm part.

Summing up, the seminal idea of this article is to employ the mentioned proximal gradient algorithm to recover the latent Laplacian embedding space where clusters are located (relying on the algebraic consistency property of the nuclear norm, see Chandrasekaran et al. [12] from a spectral clustering partition, where the noise random matrix is accounted for by the ℓ_1 norm. In this synthetic article, we just show the very good recovery of the number of clusters and the latent Laplacian embedding space by ALLE, while we leave to future research formal statements about the performance of ALLE both in these two mentioned capabilities, and in the subsequent cluster membership recovery.

Formally, given the normalized Laplacian matrix $\mathbf{L} \in \mathbb{R}^{N\times N}$, and since from (2) the statistical modeling of the Laplacian embedding can be interpreted as $\mathbf{A}\mathbf{A}^\top$ plus an error term, we seek a decomposition of the form:

$$\mathbf{L} = \mathbf{A}\mathbf{A}^\top + \mathbf{S}, \quad \mathbf{A} \in \mathbb{R}^{N\times K} \text{ such that } \mathbf{A}^\top\mathbf{A} = \mathbf{I}_K, \quad \text{and } \mathbf{S} \in \mathbb{R}^{N\times N} \tag{3}$$

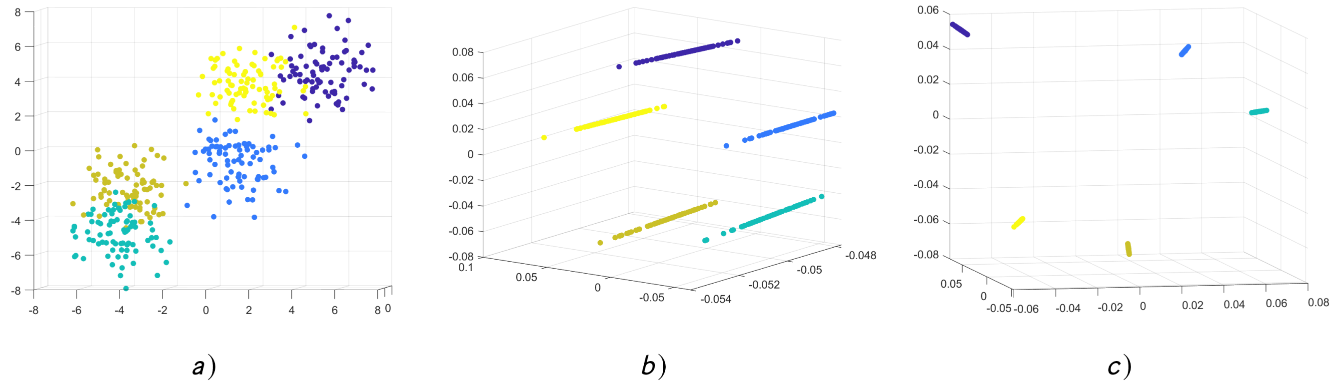

FIGURE 1 | (a) Original data $\mathbf{X}$. (b) Embedding space estimated by the eigenvectors of the Laplacian matrix. (c) Embedding space estimated by ALLE.

where $\mathbf{A}$ represents the low-rank spectral embedding (i.e., Laplacian embedding), capturing the clustering structure, while $\mathbf{S}$ constitutes the residual noise component. The goal is recovering the true rank K, which corresponds directly to the number of clusters in this case. To this end, we minimize the following penalized loss function:

$$(\hat{\mathbf{A}}, \hat{\mathbf{S}}) = \arg \min_{\mathbf{A} \in \mathbb{R}^{N \times K}, \mathbf{S} \in \mathbb{R}^{N \times N}} \|\mathbf{L} - \mathbf{A}\mathbf{A}^\top - \mathbf{S}\|_F^2 + \psi \|\mathbf{A}\mathbf{A}^\top\|_* + \rho \|\mathbf{S}\|_1 \tag{4}$$

which is a penalized version of (2), where $\|\cdot\|_F$ denotes the Frobenius norm, $\|\cdot\|_*$ is the nuclear norm (sum of singular values), $\|\cdot\|_1$ represents the element-wise ℓ_1 norm, while threshold parameters ψ and ρ control the trade-off between low-rank approximation and sparsity. By solving problem (4), we obtain an estimate of K, denoted as $\hat{K}$, without requiring manual tuning or heuristic selection. This approach only requires in input the similarity matrix $\mathbf{W}$, from which we can construct the Laplacian matrix $\mathbf{L}$, and then automatically determine the appropriate embedding dimension K and an estimation of Laplacian embedding. Following Farnè and Montanari [7], the optimal threshold pair $(\tilde{\psi}, \tilde{\rho})$ is selected by minimizing across a grid of initial thresholds for ψ and ρ the following criterion:

$$MC(\gamma) = \max \left\{ \frac{\hat{K} \|\hat{\mathbf{A}}\hat{\mathbf{A}}^\top\|_2}{\hat{\theta}}, \frac{\|\hat{\mathbf{S}}\|_{1,v}}{\gamma(1 - \hat{\theta})} \right\},$$
$$\text{where } \gamma = \rho/\psi, \text{ and } \hat{\theta} = \mathrm{tr}\left(\hat{\mathbf{A}}\hat{\mathbf{A}}^\top\right)/\mathrm{tr}(\mathbf{L})$$

where $\|_{1,v}$ is the column-sum norm. Initial thresholds are set up for ψ as a function of the mean eigenvalue of $\mathbf{L}$, and for ρ as $\psi/\sqrt{p}$. We refer to Section 3 in the Supplement of Farnè and Montanari [8] for more details. Finally, it should be noted that the estimate of the embedding space using the ALLE method is more interpretable than the same estimate derived by the standard eigenvectors of the Laplacian matrix, as shown in Figure 1.

3 | An Exhaustive Simulation Study

To evaluate the effectiveness of the proposed ALLE method in recovering the number of clusters and the Laplacian embedding structure, we performed a simulation study. The goal is to generate controlled synthetic datasets, to compute the associated Laplacian matrices and their Laplacian embeddings, and to introduce structured noise according to (3). The synthetic data $\mathbf{X} \subset \mathbb{R}^{N \times V}$ are constructed by generating cluster centroids of dimension V on the vertices of a regular polygon inscribed in a circle of radius $r = 5$. The number of vertices corresponds to the number of clusters, K. Each centroid is located on a circle, ensuring an evenly distributed cluster structure. Within each cluster, data are generated uniformly inside an ellipsoidal region centered at the corresponding centroid. The axes of the ellipsoid are denoted as the vector $\boldsymbol{\delta} \in \mathbb{R}^V$, representing the within-cluster dispersion around the centroid. Therefore, the overlap between clusters was determined by the ratio between the circle radius r and the within-cluster "radius" δ, controlling the separation between clusters.

The planned simulation design sets the overlap parameter $\boldsymbol{\delta}$ at three levels: **OL** low overlap (well-separated clusters, $\boldsymbol{\delta} = 2 \cdot \mathbf{1}_V$); **OM** (medium overlap, $\boldsymbol{\delta} = 3.5 \cdot \mathbf{1}_V$); **OH** high overlap (closely spaced clusters, $\boldsymbol{\delta} = 6.5 \cdot \mathbf{1}_V$). After generating the dataset $\mathbf{X}$, a similarity matrix $\mathbf{W}$ was computed using a self-tuning kernel function [5] by selecting the hyper-parameter of the kernel function equal to 7. From this, the corresponding normalized Laplacian matrix is constructed. The spectral decomposition of this matrix provided the first K eigenvectors, forming the (true) embedding matrix $\mathbf{A}$ that captured the underlying clustering structure. Then, the (true) within-cluster Laplacian matrix $\mathbf{A}\mathbf{A}^\top$ is built. To introduce structured noise represented by the matrix $\mathbf{S}$ in (3), specifically, we perturbed the eigenvalues of the (true) within-cluster Laplacian matrix $\mathbf{A}\mathbf{A}^\top$, which is exactly K-ranked, by a random noise matrix $\mathbf{S}$. Since the Laplacian cannot have eigenvalues larger than 1, the top K eigenvalues are negatively perturbed by an amount which has a maximum in $p \in (0,0.5)$, while the remaining $NK - K$ eigenvalues are positively perturbed by an amount which has a maximum in p. The perturbation parameter p was varied at three levels: **PL** low perturbation (minimal noise, $p = 0.05$); **PM** medium perturbation (intermediate noise, $p = 0.2$); **PH** high perturbation (strong noise effect, $p = 0.35$). The final noisy Laplacian matrix $\mathbf{L}$ is obtained by adding the perturbation matrix $\mathbf{S}$ so derived to the (true) within-cluster Laplacian structure $\mathbf{A}\mathbf{A}^\top$ according to (3), while ensuring symmetry and positive semi-definiteness (as the original eigenvectors are unvaried).

TABLE 1 | ALLE metrics (with standard errors) on the recovery of cluster number and embedding space by cluster size and number.

		$K = 2$	$K = 3$	$K = 4$	$K = 5$	$K = 6$
$n_k = 20$	$\mu(\hat{\pi}_K)$	0.138	0.342	0.349	0.293	0.298
	$\sigma(\hat{\pi}_K)$	0.282	0.147	0.143	0.229	0.283
	$\mu(\hat{\alpha}_K)$	0.089	0.105	0.09	0.095	0.109
	$\sigma(\hat{\alpha}_K)$	0.016	0.027	0.009	0.011	0.016
$n_k = 40$	$\mu(\hat{\pi}_K)$	0.007	0.024	0.036	0.067	0.091
	$\sigma(\hat{\pi}_K)$	0.038	0.085	0.072	0.127	0.14
	$\mu(\hat{\alpha}_K)$	0.084	0.081	0.081	0.082	0.081
	$\sigma(\hat{\alpha}_K)$	0.013	0.01	0.008	0.009	0.007
$n_k = 80$	$\mu(\hat{\pi}_K)$	0.016	0.08	0.071	0.127	0.184
	$\sigma(\hat{\pi}_K)$	0.068	0.137	0.134	0.162	0.166
	$\mu(\hat{\alpha}_K)$	0.083	0.084	0.082	0.08	0.077
	$\sigma(\hat{\alpha}_K)$	0.01	0.01	0.01	0.011	0.009

TABLE 2 | ALLE metrics (with standard errors) on the recovery of cluster number and embedding space by cluster size and simulation setting.

		OH			OM			OL		
		PL	PM	PH	PL	PM	PH	PL	PM	PH
$n_k = 20$	$\mu(\hat{\pi}_K)$	0.78	0.028	0.108	0.78	0.012	0.068	0.712	0.004	0.064
	$\sigma(\hat{\pi}_K)$	0.279	0.103	0.297	0.354	0.068	0.236	0.347	0.028	0.241
	$\mu(\hat{\alpha}_K)$	0.041	0.067	0.176	0.043	0.067	0.172	0.061	0.067	0.169
	$\sigma(\hat{\alpha}_K)$	0.01	0.009	0.028	0.023	0.009	0.024	0.007	0.008	0.021
$n_k = 40$	$\mu(\hat{\pi}_K)$	0	0	0.252	0	0	0.096	0	0	0.056
	$\sigma(\hat{\pi}_K)$	0	0	0.394	0	0	0.268	0	0	0.171
	$\mu(\hat{\alpha}_K)$	0.017	0.063	0.168	0.017	0.062	0.166	0.018	0.061	0.161
	$\sigma(\hat{\alpha}_K)$	0.002	0.009	0.022	0.002	0.008	0.018	0.002	0.007	0.015
$n_k = 80$	$\mu(\hat{\pi}_K)$	0	0	0.372	0	0	0.252	0	0	0.236
	$\sigma(\hat{\pi}_K)$	0	0	0.443	0	0	0.393	0	0	0.365
	$\mu(\hat{\alpha}_K)$	0.011	0.061	0.183	0.01	0.058	0.172	0.01	0.058	0.17
	$\sigma(\hat{\alpha}_K)$	0.002	0.007	0.028	0.001	0.005	0.02	0.001	0.004	0.019

The simulation study explored multiple conditions to assess ALLE; the main factors varied in the design are: number of clusters ($K = 2, \dots, 6$); number of data points in each cluster ($n_K = 20,40,80$); overlap levels (**OL**, **OM**, and **OH**); perturbation levels (**PL**, **PM**, and **PH**). Each combination of these factors defines an experimental cell of the simulation design. The total number of experimental cells is given by the product of all factor levels resulting in 135 cells of the experiment. For each cell, we generated 50 different trials. Over these 50 trials, we compute the proportion of incorrectly recovered number of clusters, $\hat{\pi}_K = \sum_{t=1}^{50} \mathbf{1}\left(\hat{K}_t \neq K\right)/50$, and the maximum principal angle between recovered and true eigenvectors for the Laplacian, $\hat{\alpha}_K = \max_{i=\{1,\dots,K\}, j=\{1,\dots,K\}} < \mathbf{u}_i, \mathbf{v}_j >$, where $\mathbf{u}_i$ is the i-th eigenvector of the recovered embedded space, and $\mathbf{v}_j$ is the j-th eigenvector of the true embedded space. Note that $\hat{\alpha}_K$ is computed only when $\hat{K}_t = K$.

Averaged simulation results for each K and each Overlap-Perturbation condition are summarized in Tables 1 and 2. Table 1 presents the error rate $\left(\mu\left(\hat{\pi}_K\right)\right)$ and standard deviation $\left(\sigma\left(\hat{\pi}_K\right)\right)$ of the ALLE recovered number of clusters across different values of K and sample sizes n_k. As K increases, the error rate $\mu\left(\hat{\pi}_K\right)$ tends to increase, although not monotonically, due to the increased complexity in recovering multiple cluster structures. However, as n_k grows, the method demonstrates improved accuracy, with a noticeable reduction in both the mean and variability of $\hat{\pi}_K$. The embedding recovery (angle between the true Laplacian embedding space and the estimated Laplacian embedding space), measured by $\mu\left(\hat{\alpha}_K\right)$, remains quite stable across different cluster numbers and sample sizes, indicating the good performance of the spectral embedding estimated by ALLE. The low values of $\sigma\left(\hat{\alpha}_K\right)$ confirm that the method maintains consistent performance in recovering the true low-rank structure.

Table 2 examines the effect of overlap (**OL**, **OM**, and **OH**) and perturbation (**PL**, **PM**, and **PH**) on clustering accuracy, aggregating results across all values of K. Higher overlap levels (**OH**) result

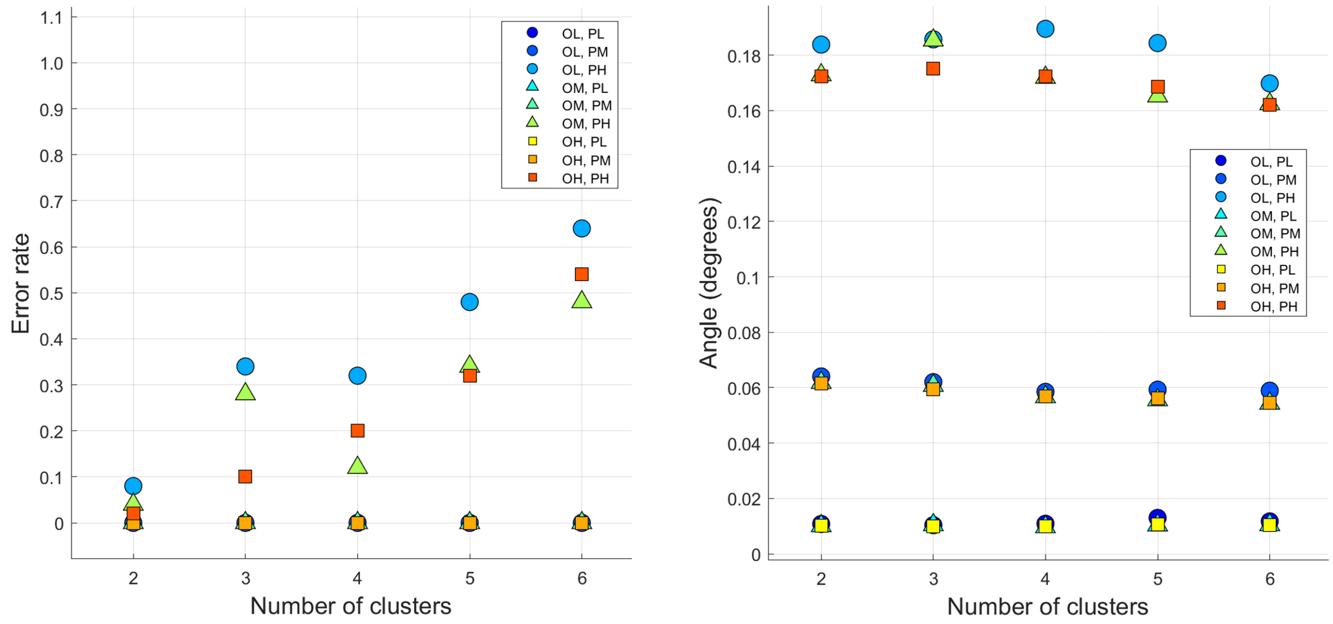

FIGURE 2 | ALLE metrics on the recovery of cluster number and embedding space by cluster size and simulation setting, with $n_K = 80$.

in increased clustering error $\mu(\hat{\pi}_K)$, reflecting the difficulty in correctly distinguishing clusters when their centroids are closer. Similarly, increasing perturbation levels (**PH**) decreases the accuracy of cluster recovery and increases variability in $\hat{\alpha}_K$. The results of Tables 1 and 2 indicate that ALLE effectively recovers the number of clusters and Laplacian embedding space, particularly in settings with moderate to low overlap and low perturbation. Larger sample sizes significantly enhance performance, reducing both recovery of rank errors and variability. Moreover, Figure 2 clearly shows that, even in the case $n_K = 80$, under the PH setting the error rate increases linearly with K, irrespective of the overlap condition. The angle metric is instead characterized by precise levels, which again depend almost completely on the perturbation level.

Overall, the simulation results confirm that ALLE effectively recovers both the number of clusters and the Laplacian embedding space in favorable conditions (low overlap, low noise, and sufficient sample size). However, performance decreases under high-overlap and high-perturbation settings.

4 | Conclusions and Future Research

In this article, we have proposed a new method to spot the unknown number of clusters and embedding space within the spectral clustering framework. The method is based on within-deviance minimization, penalized by a nuclear norm plus ℓ_1-norm penalty, which allows to systematically recover the number of clusters, which corresponds to the rank of the underlying Laplacian matrix. The new method is called ALLE (Algebraic Laplacian Estimator). An exhaustive simulation study has shown that ALLE achieves optimal performance, unless the data are not well clustered and the perturbation level is high. This study opens up the possibility to proceed with a larger inquiry about ALLE. In particular, formal statements about its capability to recover the number of groups and the embedding space, a wider presentation of simulation results under different and more challenging settings, and more importantly, to study the cluster membership recovery performance of ALLE in theory and in practice.

Author Contributions

Cinzia Di Nuzzo and Matteo Farnè are equally contributing authors.

Acknowledgment

Open access publishing facilitated by Universita degli Studi di Catania, as part of the Wiley - CRUI-CARE agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

1. A. Ng, M. Jordan, and Y. Weiss, "On Spectral Clustering: Analysis and an Algorithm," in *Advances in Neural Information Processing Systems*, vol. 14, ed. T. Dietterich, S. Becker, and Z. Ghahramani (MIT Press, 2002).

2. U. von Luxburg, "A Tutorial on Spectral Clustering," *Statistics and Computing* 17 (2007): 395–416.

3. L. Labiod and M. Nadif, "Efficient Regularized Spectral Data Embedding," *Advances in Data Analysis and Classification* 15 (2021): 99–119.

4. C. Di Nuzzo and S. Ingrassia, "A Mixture Model Approach to Spectral Clustering and Application to Textual Data," *Statistical Methods & Applications* 31 (2022): 1071–1097.

5. L. Zelnik-Manor and P. Perona, "Self-Tuning Spectral Clustering," in *Advances in Neural Information Processing Systems*, vol. 17, ed. L. Saul, Y. Weiss, and L. Bottou (MIT Press, 2004).

6. C. R. John, D. Watson, M. R. Barnes, C. Pitzalis, and M. J. Lewis, "Spectrum: Fast Density-Aware Spectral Clustering for Single and Multi-Omic Data," *Bioinformatics* 36 (2020): 1159–1166.

7. M. Farnè and A. Montanari, "A Large Covariance Matrix Estimator Under Intermediate Spikiness Regimes," *Journal of Multivariate Analysis* 176 (2020): 104577.

8. M. Farnè and A. Montanari, "Large Factor Model Estimation by Nuclear Norm Plus ℓ_1-Norm Penalization," *Journal of Multivariate Analysis* 199 (2024): 105244.

9. V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-Sparsity Incoherence for Matrix Decomposition," *SIAM Journal on Optimization* 21 (2011): 572–596.

10. A. Agarwal, S. Negahban, and M. J. Wainwright, "Noisy Matrix Decomposition via Convex Relaxation: Optimal Rates in High Dimensions," *Annals of Statistics* 40 (2012): 1171–1197.

11. X. Luo, "High Dimensional Low Rank and Sparse Covariance Matrix Estimation via Convex Minimization," 2011 Arxiv preprint.

12. V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Latent Variable Graphical Model Selection via Convex Optimization," *Annals of Statistics* 40 (2012): 1935–1967.

13. M. Fazel, H. Hindi, and S. P. Boyd, "A Rank Minimization Heuristic With Application to Minimum Order System Approximation," in *Proceedings of the 2001 American Control Conference.(Cat. No. 01CH37148)*, vol. 6 (IEEE, 2001), 4734–4739.

14. M. Fazel, "Matrix Rank Minimization With Applications," Ph.D. thesis Stanford University, 2002.

15. D. L. Donoho, "For Most Large Underdetermined Systems of Linear Equations the Minimal ℓ_1-Norm Solution Is Also the Sparsest Solution," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 59 (2006): 797–829.

16. E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust Principal Component Analysis?," *Journal of the ACM* 58 (2011): 1–37.

17. Y. Nesterov, "Gradient Methods for Minimizing Composite Functions," *Mathematical Programming* 140 (2013): 125–161.

18. T. Cai and W. Liu, "Adaptive Thresholding for Sparse Covariance Matrix Estimation," *Journal of the American Statistical Association* 106 (2011): 672–684.

19. I. Daubechies, M. Defrise, and C. De Mol, "An Iterative Thresholding Algorithm for Linear Inverse Problems With a Sparsity Constraint," *Communications on Pure and Applied Mathematics* 57 (2004): 1413–1457.