symmetry

MDPI

Article

A Further Study on the Degree-Corrected Spectral Clustering under Spectral Graph Theory

Fangmeng Liu †, Wei Li *,† and Yiwen Zhong

College of Computer and Information Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China
* Correspondence: liwei@fafu.edu.cn; Tel.: +86-136-5509-6690
† These authors contributed equally to this work.

Abstract: Spectral clustering algorithms are often used to find clusters in the community detection problem. Recently, a degree-corrected spectral clustering algorithm was proposed. However, it is only used for partitioning graphs which are generated from stochastic blockmodels. This paper studies the degree-corrected spectral clustering algorithm based on the spectral graph theory and shows that it gives a good approximation of the optimal clustering for a wide class of graphs. Moreover, we also give theoretical support for finding an appropriate degree-correction. Several numerical experiments for community detection are conducted in this paper to evaluate our method.

Keywords: spectral clustering; graphs; degree-corrected Laplacian; partition; eigenvalue

1. Introduction

Due to the growing availability of datasets of large-scale networks, community detection has attracted significant consideration. The community detection problem is to discover a community structure by dividing the network into multiple clusters according to the affinity between nodes. Because the spectral clustering method is easy to implement and can detect non-convex clusters, it is widely used for detecting clusters in networks. Compared to the traditional algorithms, spectral clustering performs well and has many fundamental advantages [1–4].

In the spectral clustering algorithm, the similarity between the data points is reflected by the weights on the edges in the graph. The data points are mapped to a lower-dimensional space through the Laplacian matrix of the graph, and finally, the non-convex datasets in the obtained low-dimensional space are clustered by traditional clustering algorithms.

Let $G = (V, E)$ be an undirected and unweighted simple graph with n nodes, where V and E are the set of nodes and edges, respectively. The adjacency matrix of graph G, denoted by $W = (w_{ij})$, is a 0–1 symmetric matrix of order n, where the (i, j)-th and (j, i)-th element is 1 if there is an edge between two nodes i and j, and 0 otherwise. Let $d_i = \sum_{j=1}^{n} w_{ij}$, which is defined as the degree of node i. Moreover, $d_{\max} = \max_{i \in V} d_i$ and $d_{\min} = \min_{i \in V} d_i$ are called the maximal degree and minimal degree of G, respectively. Denote $\bar{d}$ as the average degree of graph G, which equals $\frac{1}{n}\sum_{i=1}^{n} d_i$. The degree matrix is defined by $D = diag(d_1, \cdots, d_n)$. The symmetric matrix $D - W$ is called an unnormalized Laplacian of G, each of whose row sum is zero. The normalized Laplacian $L = I - D^{-1/2}WD^{-1/2}$ has zero as the smallest eigenvalue and plays an very important role in the spectral clustering algorithm. It is well defined only in case D^{-1} exists, i.e., there are no isolated nodes.

In 2002, Ng et al. [5] proposed a version of spectral clustering (NJW) under the normalized Laplacian matrix. Moreover, the authors in [5] analyzed their algorithm using matrix perturbation theory and gave the conditions for the algorithm performing well when nodes from different clusters are well-separated. However, when dealing with a sparse network with a strong degree of heterogeneity, i.e., the minimum degree of the graph

check for updates

Citation: Liu, F.; Li, W.; Zhong, Y. A Further Study on the Degree-Corrected Spectral Clustering under Spectral Graph Theory. *Symmetry* **2022**, *14*, 2428. https://doi.org/10.3390/sym14112428

Academic Editor: Alice Miller

Received: 31 October 2022
Accepted: 12 November 2022
Published: 16 November 2022

is low, and NJW cannot concentrate well. To resolve this issue, Chaudhuri and Chung [6] introduced the notion of a degree-corrected random-walk Laplacian $I-(D+\tau I)^{-1}W$ and demonstrated that it outputs the correct partition under a wide-range graph generated from extended planted partition (EPP) model. Instead of doing the spectral decomposition on the entire matrix, Chaudhuri and Chung [6] divided the nodes into two random subsets and only used the induced subgraph on one of those random subsets to compute the spectral decomposition. Qin and Rohe [7] investigated the spectral clustering algorithm using the degree-corrected normalized Laplacian $L_\tau = I-(D+\tau I)^{-1/2}W(D+\tau I)^{-1/2}$ under the degree-corrected stochastic blockmodel, where $\tau=\bar{d}$. This method extended the previous statistical estimation results to the more canonical spectral clustering algorithm, which is called the regularized spectral clustering (RSC). Recently, Qing and Wang [8] proposed an improved spectral clustering under the degree-corrected stochastic blockmodel also, where $\tau = 0.1\frac{d_{min}+d_{max}}{2}$, (ISC). Unlike NJW and RSC, which use the top k eigenvectors to construct the mapping matrix, ISC uses the top $k+1$ eigenvectors and the corresponding eigenvalues instead and outperforms especially in the weak signal networks, where k is the number of clusters.

Actually, previous works for spectral clustering with the degree-corrected Laplacian were mostly applied to graphs generated from stochastic blockmodels. Moreover, the optimal τ has a complex dependence on the degree of distribution of the graph and $\tau=\bar{d}$ provides good results [6,7]. In [7], the authors claimed that when $\tau=\bar{d}$, it could be adjusted by a multiplicative constant and the results are not sensitive to such adjustments. However, some numerical experiments show that an appropriate τ could be found for a better performance.

This paper investigates the spectral clustering algorithm using the degree-corrected Laplacian in view of spectral graph theory [9] and shows that it also works for a wide class of graphs. Moreover, we also provide theoretical guidance on the choice of the parameter τ. Finally, six real-world datasets are used to test the performance of our method for an appropriate τ. The results are roughly equivalent to that of RSC, or even better.

The rest of this paper is organized as follows. In Section 2, we list some relative definitions and useful lemmas in the analysis of our main results in Section 3. In Section 4, some numerical experiments are conducted for the real-world datasets. Moreover, some artificial networks are generated to analyze the effect of our method in terms of some related parameters. The conclusion and future work are provided in Section 5.

2. Preliminary

Let $G=(V,E)$ be a graph. The symmetric difference of two subsets S and T of V is defined as $S\Delta T = (S\backslash T)\cup(T\backslash S)$. For a subset S of V, $E(S,V\backslash S)=\{(u,v)\in E: u\in S, v\in V\backslash S\}$. The symbol $\mu(S)$ denotes the volume of S that is given by the sum of degree of all notes in S, i.e., $\mu(S)=\sum_{v\in S} d_v$. If k disjoint subsets $S_1,\cdots,S_k$ of V satisfy $\cup_{i=1}^{k} S_i = V$, we call $\{S_1,\cdots,S_k\}$ a k-way partition of V. Kolev and Mehlhorn [10] introduced the minimal average conductance denoted by

$$\bar{\phi}_k(G) = \min_{\{S_1,\cdots,S_k\}\in U} \frac{1}{k}(\phi(S_1)+\cdots+\phi(S_k)),$$

where U is a set of containing every k way partition of the points set of G, and $\phi(S)=\frac{|E(S,V\backslash S)|}{\mu(S)}$. A partition $\{S_1,\cdots,S_k\}$ is called optimal, if it satisfies that $\frac{1}{k}(\phi(S_1)+\cdots+\phi(S_k))=\bar{\phi}_k(G)$. In this paper, we denote $\{A_1,\ldots,A_k\}$ as the actual partition returned by the RSC algorithm, where k is the number of classes of the graph.

Let $\|\cdot\|_2$ denote the 2-norm for a vector and $\|\cdot\|_F$ denote the Frobenius norm for a matrix.

The k-means algorithm tends to find a set of k centers $\boldsymbol{c}_1,\cdots,\boldsymbol{c}_k$ to minimize the sum of the squared-distance between the points and the center to which it is assigned.

Let F be a spectral embedding map from V to a vector space. Given any k-way partition of G and a set of vectors, say $\{S_1, \cdots, S_k\}$ and $\boldsymbol{w}_1, \dots, \boldsymbol{w}_k$, respectively, the cost function of partition $\{S_1, \cdots, S_k\}$ of V, mentioned in [11], is defined as

$$g(S_1, \cdots, S_k, \boldsymbol{w}_1, \cdots, \boldsymbol{w}_k) = \sum_{i=1}^{k} \sum_{v \in S_i} d_v \|F(v) - \boldsymbol{w}_i\|_2^2. \tag{1}$$

The main idea of this function is to expand each element $F(v)$ of V by making d_v copies of $F(v)$ and form a set with $2|E(G)|$ nodes. Then, it acquires a partition by using k-means algorithm. The "trick" is to copy every node u to d_u identical nodes. This method can efficiently deal with the networks, which have the overlap between clusters. For convenience, it is necessary to assume that the k-means clustering algorithm outputs of the expansion of vertices V satisfying the following condition.

(A) For every $v \in V$, all d_v copies of $F(v)$ are contained in one part.

Suppose that $\{Y_1, \cdots, Y_k\}$ is the partition of V with centers $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_k$, which is the output of the k-means clustering algorithm, the value of the clustering cost function is denoted by "COST", i.e.,

$$\text{COST} = g(Y_1, \cdots, Y_k, \boldsymbol{z}_1, \cdots, \boldsymbol{z}_k).$$

Then, we will introduce the traditional NJW and RSC Algorithm 1.

Algorithm 1 The traditional NJW and RSC algorithm

```
Input: $W, k$, ( $\tau$ for RSC)
Calculate the normalized Laplacian matrix $L = D^{-1/2} W D^{-1/2}$.
    ($L_\tau = (D + \tau I)^{-1/2} W (D + \tau I)^{-1/2}$ for RSC).
Find the eigenvectors $f_1, \cdots, f_k$ corresponding to the $k$ largest eigenvalues of $L$. Form
    $X = [f_1, \cdots, f_k]$ by putting the eigenvectors into the columns.
Normalize each row of $X$ to get matrix $Y$, i.e., $Y_{ij} = X_{ij} / (\sum_{j=1}^{k} X_{ij}^2)^{1/2}$, where
    $i = 1, \cdots, n$ and $j = 1, \cdots, k$.
Apply $k$-means method to $Y$ to get the label of each node.
Output: labels for all nodes
```

3. Analysis of RSC Algorithm

Our method for analyzing the RSC algorithm follows the strategy developed by Peng et al. [11], Kolve et al. [10], and Mizutani [12]. Let $\{S_1, \cdots, S_k\}$ be a partition of the nodes set of V. Define $\boldsymbol{g}_i \in \mathbb{R}^n$ is the normalized indicator of S_i. That means, if $v \in S_i$, the v-th element of $\boldsymbol{g}_i$ is one, or else is zero. The normalized indicator $\bar{\boldsymbol{g}}_i$ of S_i is given as

$$\bar{\boldsymbol{g}}_i = \frac{D^{1/2} \boldsymbol{g}_i}{\|D^{1/2} \boldsymbol{g}_i\|_2} = \begin{cases} \sqrt{\frac{d_v}{\mu(S_i)}} & v \in S_i \\ 0 & v \notin S_i. \end{cases}$$

It is obvious that $\|\bar{\boldsymbol{g}}_i\|_2 = 1$.

The following result is called the structure theoremwhich plays a very important role to examine the performance of the spectral clustering. It shows that there is a linear combination $\hat{\boldsymbol{f}}_i$ of $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_{k+1}$ such that $\hat{\boldsymbol{f}}_i$ and $\boldsymbol{g}_i$ are close.

Theorem 1 (Structure Theorem). *Let*

$$\Psi = \frac{1}{1 - \lambda_{k+1}(\tau)} \left(1 - \frac{d_{min}}{d_{max} + \tau} + \bar{\phi}_k(G) \frac{d_{min}}{d_{max} + \tau}\right),$$

where $\lambda_{k+1}(\tau)$ (λ_{k+1} for short) is the $(k+1)$-th largest eigenvalue of L_τ, and $\{S_1, \cdots, S_k\}$ be the $\bar{\phi}_k(G)$-optimal partition of G, $\bar{\boldsymbol{G}} = [\bar{\boldsymbol{g}}_1, \ldots, \bar{\boldsymbol{g}}_k] \in \mathbb{R}^{n\times k}$, $\bar{\boldsymbol{F}} = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k] \in \mathbb{R}^{n\times k}$. If $k\Psi < 1$, then there exists a $k \times k$ orthogonal matrix $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]$, such that

$$||\bar{\boldsymbol{F}}\boldsymbol{U} - \bar{\boldsymbol{G}}||_F \leq 2\sqrt{k\Psi}.$$

Proof. Denote $\bar{g}_i^u$ as the element in $\bar{\boldsymbol{g}}_i$ corresponding to the vertex u. Moreover, let

$$\bar{\boldsymbol{g}}_i = \sum_{j=1}^{n} h_{i,j}\boldsymbol{f}_j,\ \hat{\boldsymbol{f}}_i = \sum_{j=1}^{k} h_{i,j}\boldsymbol{f}_j.$$

First,

$$\begin{aligned}
\bar{\boldsymbol{g}}_i^T L_\delta \bar{\boldsymbol{g}}_i &= \sum_{\{u,v\}\in E(G)} \left[\left(\frac{1}{\sqrt{d_u}}\bar{g}_i^u\right)^2 - \frac{2}{\sqrt{d_u+\tau}\sqrt{d_v+\tau}}\bar{g}_i^u\bar{g}_i^v + \left(\frac{1}{\sqrt{d_v}}\bar{g}_i^v\right)^2 \right] \\
&= \sum_{u\in S_i; v\in \bar{S}_i} \frac{1}{\mu(S_i)} + \sum_{u,v\in S_i} \frac{2}{\mu(S_i)}\left(1 - \frac{\sqrt{d_u d_v}}{\sqrt{d_u+\tau}\sqrt{d_v+\tau}}\right) \\
&\leq \phi(S_i) + \frac{2E(S_i)}{\mu(S_i)}\left(1 - \frac{d_{min}}{d_{max}+\tau}\right) \\
&= 1 - \frac{2E(S_i)}{\mu(S_i)}\frac{d_{min}}{d_{max}+\tau} \\
&= 1 - \frac{d_{min}}{d_{max}+\tau} + \phi(S_i)\frac{d_{min}}{d_{max}+\tau} < 1.
\end{aligned} \tag{2}$$

On the other hand,

$$\begin{aligned}
\bar{\boldsymbol{g}}_i^T L_\delta \bar{\boldsymbol{g}}_i &= \left(\sum_{j=1}^{n} h_{i,j}\boldsymbol{f}_j\right)^T L_\delta \left(\sum_{j=1}^{n} h_{i,j}\boldsymbol{f}_j\right) \\
&= \left(\sum_{j=1}^{n} h_{i,j}\boldsymbol{f}_j\right)^T \left(\sum_{j=1}^{n} h_{i,j}(1-\lambda_j)\boldsymbol{f}_j\right) \\
&= \sum_{j=1}^{n} h_{i,j}^2(1-\lambda_j) \\
&\geq \sum_{j=k+1}^{n} h_{i,j}^2(1-\lambda_j) \\
&\geq (1-\lambda_{k+1}) \sum_{j=k+1}^{n} h_{i,j}^2.
\end{aligned}$$

Then,

$$||\hat{\boldsymbol{f}}_i - \bar{\boldsymbol{g}}_i||_2^2 = \sum_{j=k+1}^{n} h_{i,j}^2 \leq \frac{1}{1-\lambda_{k+1}}\left(1 - \frac{d_{min}}{d_{max}+\tau} + \phi(S_i)\frac{d_{min}}{d_{max}+\tau}\right),$$

and

$$||\hat{\boldsymbol{F}} - \bar{\boldsymbol{G}}||_F^2 = \sum_{i=1}^{k} ||\hat{\boldsymbol{f}}_i - \bar{\boldsymbol{g}}_i||_2^2 \leq \frac{1}{1-\lambda_{k+1}}\left(1 - \frac{d_{min}}{d_{max}+\tau} + \bar{\phi}_k(G)\frac{d_{min}}{d_{max}+\tau}\right).$$

Let $\boldsymbol{h}_i = [h_{i,1}, \ldots, h_{i,k}]^T$, $i = 1, \cdots, k$, and $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_k] \in \mathbb{R}^{k\times k}$. Considering the singular value decomposition of $\boldsymbol{H}$, given as $\boldsymbol{H} = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^T$, where $\boldsymbol{A} \in \mathbb{R}^{k\times k}$ and $\boldsymbol{B} \in \mathbb{R}^{k\times k}$ are orthogonal matrices, and $\boldsymbol{\Sigma}$ is a $k \times k$ diagonal matrix.

Let $\boldsymbol{U} = \boldsymbol{A}\boldsymbol{B}^T$ and $\boldsymbol{R} = \boldsymbol{U} - \boldsymbol{H} \in \mathbb{R}^{k\times k}$. Then, $\boldsymbol{U}$ is an orthogonal matrix. According to the proof of Theorem 4 in [12], it obtains

$$\|\boldsymbol{R}\|_F \leq k\Psi \text{ and } \|\bar{\boldsymbol{F}}\boldsymbol{U} - \bar{\boldsymbol{G}}\|_F \leq k\Psi + \sqrt{k\Psi}.$$

When $k\Psi < 1$, we have

$$\|\bar{\boldsymbol{F}}\boldsymbol{U} - \bar{\boldsymbol{G}}\|_F \leq 2\sqrt{k\Psi}. \tag{3}$$

This completes the proof. □

Given k vectors, say $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k \in \mathbb{R}^k$, we suppose that $||\boldsymbol{c}_i - \boldsymbol{c}_j||_2^2$ is lower bounded by some real numbers $\zeta_{i,j} \geq 0$ and $g(S_i, \ldots, S_k, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_k)$ is upper bound by a real number $\omega \geq 0$, i.e.,

$$||\boldsymbol{c}_i - \boldsymbol{c}_j||_2^2 \geq \zeta_{ij}\ (i \neq j) \text{ and } g(S_i, \ldots, S_k, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_k) \leq \omega. \tag{4}$$

We are now ready to derive the bounds of ζ_{ij} and ω shown in (4) for the RSC algorithm. Let $\bar{\boldsymbol{F}} = [\boldsymbol{f}_1, \cdots, \boldsymbol{f}_k]$ and $\boldsymbol{p}_v$ be the v-th row of $\bar{\boldsymbol{F}}$, corresponding to the node v. Since U is an orthogonal matrix, the inequality (3) can be rewritten as

$$\|\bar{\boldsymbol{F}}U - \bar{\boldsymbol{G}}\|_F = \|\bar{\boldsymbol{F}} - \bar{\boldsymbol{G}}\boldsymbol{U}^T\|_F = \|\bar{\boldsymbol{F}}^T - \boldsymbol{U}\bar{\boldsymbol{G}}^T\|_F = \sum_{i=1}^{k}\sum_{v\in S_i}\left\|\boldsymbol{p}_v - \sqrt{\frac{d_v}{\mu(S_i)}}\boldsymbol{u}_i\right\|_2^2. \tag{5}$$

The spectral embedding map in the RSC algorithm, denoted by $F_{RSC}(v)$, is given as

$$F_{RSC}(v) = \frac{1}{\|\boldsymbol{p}_v\|_2}\boldsymbol{p}_v.$$

Hence, according to the discussion in [12], it is easy to obtain the upper bound of "COST". The discussion needs the following inequality.

Lemma 1 ([12])**.** *The following inequality holds for a vector $\boldsymbol{a} \in \mathbb{R}^k$ and a vector $\boldsymbol{u} \in \mathbb{R}^k$ with $\|\boldsymbol{u}\|_2 = 1$,*

$$\left\|\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2} - \boldsymbol{u}\right\|_2 \leq 2\|\boldsymbol{a} - \boldsymbol{u}\|_2.$$

Theorem 2. *Let a partition $\{S_1, \cdots, S_l\}$ of G be an optimal achieving $\bar{\phi}_k(G)$ and F_{RSC} be the spectral embedding map in RSC algorithm. Define the center of S_i, $\boldsymbol{c}_i = \boldsymbol{u}_i$ for $i = 1, \cdots, k$, then*

- $\|\boldsymbol{c}_i - \boldsymbol{c}_j\|_2^2 = 2$.
- $g(S_1, \ldots, S_k, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_k) \leq 16k\mu_{max}\Psi$,

where $\mu_{max} = \max\{\mu(S_i)|i = 1, 2, \cdots, k\}$.

Proof. First, since $\boldsymbol{c}_i = \boldsymbol{u}_i$, we have that

$$\|\boldsymbol{c}_i - \boldsymbol{c}_j\|_2^2 = (\boldsymbol{u}_i - \boldsymbol{u}_j)^T(\boldsymbol{u}_i - \boldsymbol{u}_j) = 2.$$

On the other hand, let $F(v) = \sqrt{\frac{\mu(S_i)}{d_v}}\boldsymbol{p}_v$. Then,

$$\begin{aligned}
& g(S_1, \cdots, S_k, \boldsymbol{c}_1, \cdots, \boldsymbol{c}_k) \\
=& \sum_{i=1}^{k} \sum_{v \in S_i} d_v \|F_{RSC}(v) - \boldsymbol{u}_i\|_2^2 \\
=& \sum_{i=1}^{k} \sum_{v \in S_i} d_v \left\| \frac{\boldsymbol{p}(v)}{\|\boldsymbol{p}(v)\|} - \boldsymbol{u}_i \right\|_2^2 \\
=& \sum_{i=1}^{k} \sum_{v \in S_i} d_v \left\| \frac{F(v)}{\|F(v)\|} - \boldsymbol{u}_i \right\|_2^2 \\
\leq& 4 \sum_{i=1}^{k} \sum_{v \in S_i} d_v \left\| \sqrt{\frac{\mu(S_i)}{d_v}} \boldsymbol{p}_v - \boldsymbol{u}_i \right\|_2^2 \quad \text{(by Lemma 1)} \\
=& 4 \sum_{i=1}^{k} \sum_{v \in S_i} \mu(S_i) \left\| \boldsymbol{p}_v - \sqrt{\frac{d_v}{\mu(S_i)}} \boldsymbol{u}_i \right\|_2^2 \quad \text{(by Equation (5) and Theorem 1)} \\
\leq& 16k\mu_{max}\Psi.
\end{aligned}$$

The result holds. □

Assume that OPT stands for the optimal clustering cost of graph G, then it is obvious that COST $\leq \alpha \cdot$ OPT, where α is an approximation ratio. Moreover, OPT $\leq g(S_1, \cdots, S_k, \boldsymbol{c}_1, \cdots, \boldsymbol{c}_k)$. Therefore, we can obtain the upper bound of COST.

Theorem 3. *Let $\{S_1, \cdots, S_k\}$ be a $\bar{\phi}_k(G)$-optimal partition of G. Then*

$$COST \leq 16k\alpha\mu_{max}\Psi.$$

Lemma 2 ([12])**.** *Assume that, for every permutation $\pi : \{1, \ldots, k\} \rightarrow \{1, \ldots, k\}$, there is an index l such that $\mu(A_l \Delta S_{\pi(l)}) \geq 2\epsilon\mu(S_{\pi(l)})$ for a real number $0 \leq \epsilon \leq 1/2$. Then, the following inequality holds,*

$$COST \geq \frac{1}{8} \sum_{i \in H} \xi_i \zeta_{i,p} \min\{\mu(S_i), \mu(S_l)\} - \omega,$$

where H is a subset of $\{1, \ldots, k\}$, p is an element of $\{1, \ldots, k\}$ and $\xi_i \geq 0$ is a non-negative real number satisfying $\sum_{i \in H} \xi_i \geq \epsilon$, and ω is the upper bound of $g(S_1, \cdots, S_k, \boldsymbol{c}_1, \cdots, \boldsymbol{c}_k)$ in (4).

By setting $\zeta_{i,j} = 2$ and $\omega = 16k\alpha\mu_{max}\Psi$, then we obtain the following result.

Theorem 4. *Suppose that the assumption of Lemma 2 holds. Then,*

$$COST \geq \frac{1}{4}\epsilon\mu_{min} - 16k\mu_{max}\Psi.$$

Theorem 5 (Main result)**.** *Given a graph G = (V,E) and a positive integer k, let a partition $\{S_1, \ldots, S_n\}$ of G be $\bar{\phi}_k(G) - optimal$ and $A_1, \ldots, A_n$ be a partition of G returned by the RSC clustering algorithm. Assume that a k-means clustering algorithm has an approximation ratio of α and satisfies assumption (A). If $\Psi \leq \frac{\mu_{min}}{264 * 2k\alpha\mu_{max}}$, then, after a suitable renumbering of $A_1, \ldots, A_k$, the following holds for $i = 1, \ldots, k$,*

$$\mu(A_i \Delta S_i) \leq \left(\frac{264k\alpha\mu_{max}}{\mu_{min}} \Psi \right) \mu(S_i).$$

Proof. Choose a real number

$$\epsilon = \frac{132k\alpha\mu_{max}}{\mu_{min}}\Psi < \frac{1}{4}.$$

Assume that, for every permutation $\pi : \{1,\ldots,k\} \to \{1,\ldots,k\}$, there is an index l such that $\mu(A_l \Delta S_{\pi(l)}) \geq 2\epsilon\mu(S_{\pi(l)})$ for a real number ϵ. Hence, applying Theorems 3 and 4, we can obtain the following

$$\begin{aligned} COST &\geq \frac{1}{4}\epsilon\mu_{min} - 16k\mu_{max}\Psi \\ &\geq 33k\alpha\mu_{max}\Psi - 16k\alpha\mu_{max}\Psi \\ &= 17k\alpha\mu_{max}\Psi \\ &> 16k\alpha\mu_{max}\Psi, \end{aligned}$$

which contradicts Theorem 3. That means, after a suitable renumbering of $A_1,\ldots,\ A_n$, we have

$$\mu(A_i\Delta S_i) \leq 2\epsilon\mu(S_i) = \left(\frac{264k\alpha\mu_{max}}{\mu_{min}}\Psi\right)\mu(S_i),$$

for every $i = 1,2,\cdots,k$. □

4. Finding an Appropriate τ and Numerical Experiment

The main theorem gives an upper bound of $\mu(A_i\Delta S_i)$ in RSC algorithm. It tells us that the performance would vary while the term Ψ decreases with increasing τ. In this section, we will try to find an appropriate τ for the good partitioning, according to this main theorem.

Before our analysis, we may make some reasonable assumptions as (B) to (D).

(B) $2|E(S_i)|/\mu(S_i) > 1/\bar{d}$

(C) $\tau \leq k\bar{d}$

(D) $\mu_{min}/\mu_{max} \leq \frac{2\bar{d}}{n}$.

Firstly, $\frac{2|E(S_i)|}{\mu(S_i)}$ stands for the ratio of the edges in S_i to the degree summation of all points in S_i. We may assume that $\frac{2|E(S_i)|}{\mu(S_i)} > 1/\bar{d}$, since S_i is one of clusters in the optimal partitioning. Second, as mentioned in [6,7], the choice of τ is very important. If τ is too small, there is insufficient regularization. If τ is too large, it washes out significant eigenvalues. Then, it is reasonable to assume that $\tau \leq k\bar{d}$. Moreover, μ_{min} and μ_{max} stand for the edges in the responding cluster and the ratio of them stands for the relative density. Hence, we can assume that $\frac{\mu_{min}}{\mu_{max}} \leq \frac{2\bar{d}}{n}$.

Then,

$$\Psi \leq \frac{1}{1-\lambda_{k+1}(\tau)}\left(1 - \frac{d_{min}}{\bar{d}(d_{max}+\tau)}\right) \leq \frac{\bar{d}(d_{max}+k\bar{d}) - d_{min}}{(1-\lambda_{k+1}(\tau))(\bar{d}(d_{max}+\tau))},$$

where $\lambda_{k+1}(\tau)$ is the $(k+1)$-th largest eigenvalue of L_τ. Furthermore, the theoretical analysis in [7] shows that $\tau = \bar{d}$ provides good results and one could adjust this by a multiplicative constant. For these reasons, we set $\tau = \delta\bar{d}$ and attend to find an appropriate δ to refine the algorithm.

Six real datasets are used to test our method. These datasets can be downloaded directly from http://zke.fas.harvard.edu/software.html, accessed on 10 September 2022. Table 1 shows the detail information of six real datasets, including the source of the dataset, the number of data points (n), the number of communities included (k), the minimum degree of data points (d_{min}), and the maximum degree of data points (d_{max}).

Table 1. The information of six real datasets.

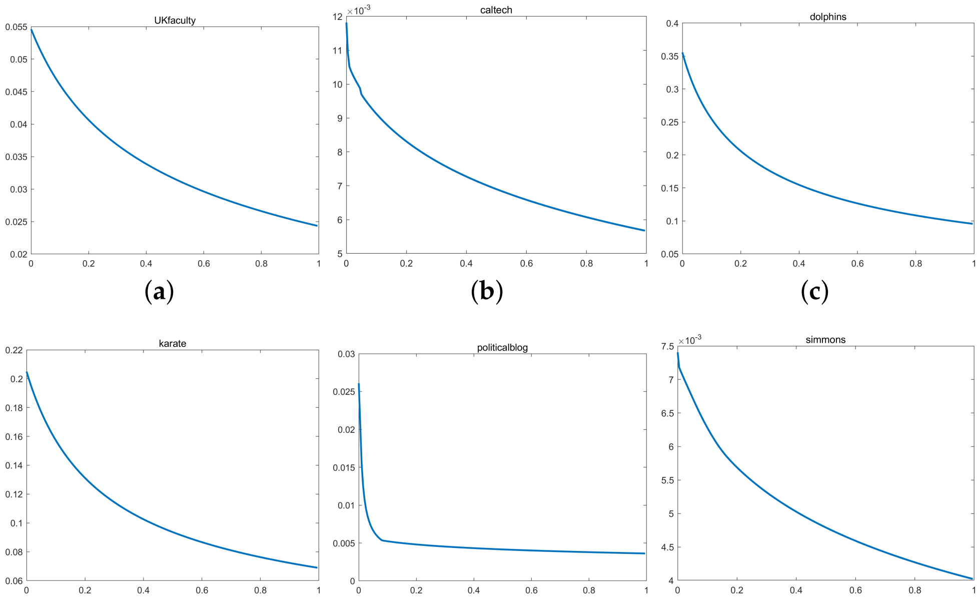
DataSet	Source	n	k	d_{min}	d_{max}	$\bar{d}$
UKfaculty	Nepusz et al. (2008) [13]	79	3	2	39	13.97
caltech	Traud et al. (2011) [14]	590	8	1	179	43.46
dolphins	Lusseau (2003) [15]	62	2	1	12	5.12
karate	Zachary (1977) [16]	34	2	1	17	4.6
politicalblog	Adamic and Glance (2005) [1]	1222	2	1	351	27.35
simmons	Traud et al. (2011) [14]	1137	4	1	293	42.66

4.1. Find an Appropriate δ

Let

$$UB(\delta) = \frac{1}{(1 - \lambda_{k+1}(\delta\bar{d}))(d_{max} + \delta\bar{d})}.$$

Figure 1 plots the variation of UB(δ) when δ varies between 0 and 1 in six real datasets. It is obvious that $UB(\delta)$ is decreasing with increasing δ. The following theorem (often called the Geršgorin disc theorem) makes this observation true.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

Figure 1. Plots of the UB(δ) in six real datasets: x axis: δ and y axis: UB(δ). (**a**) UB for different values of δ on UKfaculty; (**b**) UB for different values of δ on caltech; (**c**) UB for different values of δ on dolphins; (**d**) UB for different values of δ on karate; (**e**) UB for different values of δ on politicalblog; (**f**) UB for different values of δ on simmons;

Theorem 6 (Geršgorin Disk Theorem)**.** *Let* $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ *and*

$$R_i(A) = \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|, \quad 1 \leq i \leq n$$

denote the deleted absolute row sums of A. Then, all the eigenvalues of A are located in the union of n discs

$$\bigcup_{i=1}^{n} \{z \in \mathbf{C} : |z - a_{ii}| \leq R_i(A)\} \equiv G(A).$$

It tells us that, for all $i = 1, 2, \cdots, n$, $R_i(L_{\delta\bar{d}})$ decreases with increasing δ. Then, $\lambda_{k+1}(\delta\bar{d})$ and the term $UB(\delta)$ will decrease as well. It is easy to see that $\lim_{\delta\to\infty} UB(\delta) = 0$. Therefore, we would like to find an appropriate δ, such that the upper bound $UB(\delta)$ will not vary too much, when δ varies small.

According to Theorem 5, we may assume that

$$\Psi \leq \frac{\bar{d}(d_{max} + k\bar{d}) - d_{min}}{(1 - \lambda_{k+1}(\tau))(\bar{d}(d_{max} + \tau))} \leq \frac{\mu_{min}}{264 * 2k\alpha\mu_{max}}.$$

Then

$$\frac{1}{(1 - \lambda_{k+1}(\tau))(\bar{d}(d_{max} + \tau))} \leq \frac{\bar{d}}{264k^2n},$$

follows from the assumption $\frac{\mu_{min}}{\mu_{max}} \leq \frac{2\bar{d}}{n}$ and $\bar{d}(d_{max} + k\bar{d}) - d_{min} \geq k$.

Define $UBD(\delta)$ as the absolute difference of UB when δ increases 0.005, i.e., $UBD(\delta) = |UB(\delta + 0.005) - UB(\delta)|$. We would like to find that δ_0 satisfies the following conditions:

$$\begin{cases} \forall\delta \geq \delta_0, & UBD_\delta \leq \frac{\bar{d}}{264k^2n} \\ \forall\delta < \delta_0, & UBD_\delta > \frac{\bar{d}}{264k^2n}. \end{cases} \tag{6}$$

In the rest of this paper, three indices, namely RI, NMI, and error rate, are used to evaluate the effectiveness.

Evaluation Indices

Rand Index For a dataset with n data points, the total number of sample pairs is $\frac{n(n-1)}{2}$. If two sample points belong to the same class are classified into the same class, we denote the number of such sample pairs as a. If two sample points belong to different classes are divided into different classes, we denote the number of such sample pairs as b. The calculation formula of RI is as follows:

$$RI = \frac{a + b}{n(n-1)/2}.$$

The RI value represents the proportion of correctly clustered sample pairs in all sample pairs and is often used to measure the similarity between two datasets. Obviously, RI is between 0 and 1. If $RI = 1$, the clustering is completely correct, and if $RI = 0$, it is completely wrong.

Normalized Mutual Information We use U and V to denote the true label vector and predicted label vector, respectively. Let U_i represent the elements belonging to class i in U and V_j represent the elements belonging to class j in V. $H(U)$ represents the information entropy of U, that could be calculated by

$$H(U) = -\sum_{i=1}^{n} p_i \log p_i,$$

where the base of the logarithmic function is usually 2 and p_i represents the ratio of the number of nodes belonging to class i to the total amount of nodes, i.e., $p_i = \frac{|U_i|}{n}$. Now, we can obtain the formula for calculating mutual information (MI):

$$MI(U, V) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} log\left(\frac{p_{ij}}{p_i \times p_j}\right),$$

where $p_{ij} = \frac{|U_i \cap V_j|}{n}$. Based on the information entropy and the mutual information, we can obtain the normalized mutual information as

$$NMI(U,V) = 2\frac{MI(U,V)}{H(U) \times H(V)}.$$

Error Rate Error rate is defined by

$$min_{\{\pi:permutation\ over\ \{1,2,\cdots,k\}\}}\frac{1}{n}\sum_{i=1}^{n}1\left\{\pi\left(\hat{l}_i \neq l_i\right)\right\},$$

where $\hat{l}_i$ and l_i are the true and predicted labels of node i, respectively.

4.2. Real Networks Experiments

After some pre-processing, these six real datasets are all networks containing k non-overlapping communities and are labeled. We will use RSC-δ to stand for the RSC algorithm when $\delta = \delta_0$ which satisfies the condition in (6). Actually, NJW, RSC, and RSC-δ are three different cases in RSC algorithm, when δ takes different values. When $\delta = 0$, it is NJW algorithm. When $\delta = 1$, it is the RSC algorithm. When $\delta = \delta_0$ in (6), it is RSC-δ. Table 2 shows the experimental results of these three cases. Furthermore, the best performance in each dataset is indicated by the bold-type letter. The last row in Table 2 shows the corresponding δ_0 in RSC-δ.

Table 2. Results on six real datasets.

DataSet		UKfaculty	Caltech	Dolphins	Karate	Politicalblog	Simmons
RI	NJW	0.9834	**0.9091**	**1**	0.9412	0.5003	**0.8596**
	RSC	0.9834	**0.9091**	**1**	0.9412	0.5003	**0.8596**
	RSC-δ	**1**	0.8936	0.9677	**1**	**0.9095**	0.8550
NMI	NJW	0.9502	**0.6138**	**1**	0.8365	0.0006	**0.6796**
	RSC	**1**	0.5881	0.8904	**1**	0.7133	0.6143
	RSC-δ	**1**	0.5867	0.8904	**1**	**0.7317**	0.6187
Error rate	NJW	1/79	**149/590**	**0/62**	1/34	586/1222	284/1137
	RSC	**0/79**	170/590	1/62	**0/34**	64/1222	244/1137
	RSC-δ	**0/79**	174/590	1/62	**0/34**	**58/1222**	**238/1137**
δ_0		0.71	2.155	2.435	1.205	0.15	0.625

As can be seen from the table, RSC-δ is fully clustered correctly on UKfaculty and karate dataset. Moreover, RSC-δ achieves the best clustering results on the politicalblog dataset, with only 58 clustering errors.

Table 3 shows the items of the upper bound for $\mu(S_i \Delta A_i)$ proposed in Theorem 5. From the observation, the performance of the RSC-δ algorithm is effected by the two parameters of $\frac{\mu_{max}}{\mu_{min}}$ and $\bar{\phi}_k(G)$. For example, the RSC-δ does not perform well in caltech and dolphins. All networks except caltech have the minimal average conductance smaller than 0.4 and that of caltech is larger than 0.5. Although dolphins has a small $\bar{\phi}_k(G)$, $\frac{\mu_{max}}{\mu_{min}}$ is larger than 2.

4.3. Synthetic Data Experiments

In this section, we will use artificial networks to evaluate the performance of the RSC-δ algorithm in terms of the average degree, mixing parameter, and the number of nodes in the largest community. We generate artificial networks using the LFR benchmark, which is considered as a standard test network for community detection, characterized by a non-uniform distribution of node degrees and community sizes.

Table 3. The $\bar{\phi}_k(G)\frac{k\mu_{max}}{\mu_{min}}$ of six real datasets.

DataSet	μ_{min}	μ_{max}	$\bar{\phi}_k(G)$	$\frac{\mu_{max}}{\mu_{min}}$	$\bar{\phi}_k(G)\frac{k\mu_{max}}{\mu_{min}}$
UKfaculty	189	519	0.1909	2.7460	1.5724
caltech	1443	4821	0.5062	3.3410	13.5302
dolphins	94	224	0.0453	2.3830	0.2159
karate	76	80	0.1283	1.0526	0.2701
politicalblog	16,175	17,253	0.0943	1.0666	0.2012
simmons	8796	15,592	0.2946	1.7726	2.0890

The test artificial networks are generated with the following parameters: the number of nodes (n), the average degree ($\bar{d}$), the maximum degree(maxd), the mixing parameter (μ), the number of nodes in the smallest community (minc), and the number of nodes in the largest community (maxc). The value of the mixing parameter, denoted by μ, is between 0.1 and 0.9. Low amounts of μ give a clear community structure where the intra-cluster link is much more than the inter-cluster link [17].

4.3.1. The Ratio of the Average Degree to the Maximum Degree

In this experiment, we generate nine artificial networks consisting of 500 nodes. To evaluate the performance of RSC-δ in terms of the average degree, we fix the parameter $\mu = 0.5$, minc = 100, maxc = 300, maxd = 220, respectively, and the average degreevaries from 10 to 170, i.e., 10, 30, 50, 70, 90, 110, 130, 150, and 170, respectively. Then, the ratio of the average degree to the maximum degree varies from 0.0455 to 0.7727. The performance comparison is summarized in Figure 2.

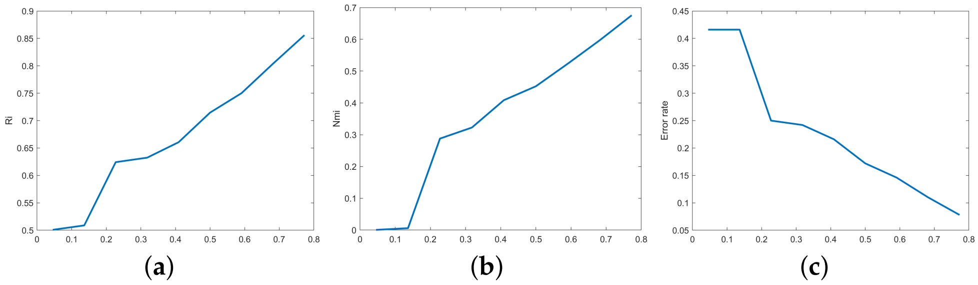


Figure 2. Ri, Nmi, and Error rate for different average degrees: x axis: the ratio of the average degree to the maximum degree; and the y axis: Ri, Nmi, Error rate. (**a**) Ri for different values of $\bar{d}$; (**b**) Nmi for different values of $\bar{d}$; (**c**) Error rate for different values of $\bar{d}$.

From our observation, we can understand that the performance of RSC-δ is highly dependent on the average degree of the network. With the average degree increasing, RI and NMI increase, and the error rate decreases significantly. Actually, this phenomenon is verified by the inequality (2), since the equality holds when the graph is regular.

4.3.2. Mixing Parameter

In this experiment, we also generate nine artificial networks with 500 nodes and fix the parameter $\bar{d} = 15$, minc = 100, maxc = 300, maxd = 220, respectively. In order to study the effect of the mixing parameter on RSC-δ, μ varies from 0.1 to 0.9. The experimental results are shown in Figure 3.

From the observation, we understand that RSC-δ performs excellently when μ is between 0 and 0.3. However, it drops sharply when μ is varying from 0.3 to 0.5. This phenomenon coincides with the result for the real datasets, that RSC-δ does not perform well when $\phi_k(G)$ is larger than 0.5. However, the performance of RSC-δ remains stable when $\mu \geq 0.5$. This shows that RSC-δ is less affected by $\mu \geq 0.5$.

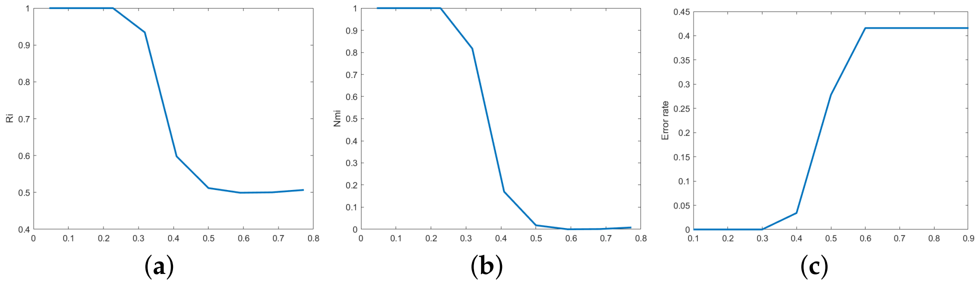


Figure 3. Ri, Nmi, and Error rate for different mixing parameters: x axis: μ and y axis: Ri, Nmi, Error rate. (**a**) Ri for different values of μ; (**b**) Nmi for different values of μ; (**c**) Error rate for different values of μ.

4.3.3. The Number of Nodes in the Largest Community

In this experiment, we generate 13 artificial networks consisting of 1700 nodes. To evaluate the performance of RSC-δ in terms of the number of nodes in the largest community, we fix the parameter $\mu = 0.5$, $\bar{d}$ = 30, minc = 300, maxd = 500, respectively, and the number of nodes in the largest community is varying from 300 to 900, step size is 50. The experimental results are shown in Figure 4.

Since both the degree and the community size distributions, in the graph generated by the LFR benchmark, are power laws, this experiment uses $\frac{minc}{maxc}$ to simulate $\frac{\mu_{max}}{\mu_{min}}$, and the experiment result shows that the RSC-δ algorithm performs well when the network is "balanced", which also verifies the results in the real datasets.

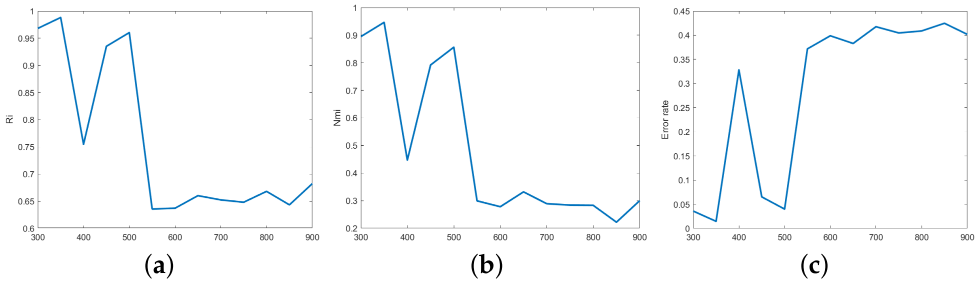


Figure 4. Ri, Nmi, and Error rate for different numbers of nodes in the largest community: x axis: the number of nodes; and y axis: Ri, Nmi, Error rate. (**a**) Ri for different values of maxc; (**b**) Nmi for different values of maxc; (**c**) Error rate for different values of maxc.

5. Conclusions

Traditional spectral clustering algorithms such as NJW have poor performance in sparse networks with a strong degree of heterogeneity. The RSC algorithm improves the performance of spectral clustering in sparse networks through degree correction. Based on the spectral graph theory, this paper investigates the degree correction method of RSC, and shows that the RSC algorithm works for a wide class of networks. Moreover, we also provide a method to find an appropriate degree-correction τ to refine the RSC algorithm. Some numerical experiments are conducted to evaluate the performance of our method. By comparing the experimental results on the six real datasets, RSC-δ performs well on the datasets named karate, politicalblog, and simmons. Finally, the experimental results on the artificial networks show that RSC-δ performs well when the average degree is much smaller than the maximum degree. Furthermore, the performance of RSC-δ algorithm is less affected by the mixing parameter $\mu \geq 0.5$. At last, the numerical experiments also show that the algorithm is affected by two parameters of $\bar{\phi}_k(G)$ and $\frac{\mu_{max}}{\mu_{min}}$.

6. Discussion

The RSC algorithm uses a constant τ for the degree-correction. Can we use different degree-corrections for different nodes? We try to use the information of the neighbor nodes of each node as follows.

Let $N(i)$ be the set of nodes adjacent to node i. Denote $d^i_{max} = \max\{d_j : v_j \in N(i)\}$, $d^i_{min} = \min\{d_j : v_j \in N(v_i)\}$, $d^i_{mid} = \frac{1}{2}(d^i_{max} + d^i_{min})$ and $d^i_{mean} = \sum_{j \in N(i)} d_j / d_i$.

Let $\Pi = diag(\pi_1, \cdots, \pi_n)$ be a diagonal matrix of order n. The modified normalized Laplacian matrix is

$$L_\Pi = (D + \Pi)^{-1/2} W (D + \Pi)^{-1/2},$$

We used RSC-max, RSC-min, RSC-mean, and RSC-mid to denote the method when π_i equals to d^i_{max}, d^i_{min}, d^i_{mean}, d^i_{mid}, and $i = 1, 2, \cdots, n$, respectively. Table 4 shows the experimental results of these methods. We can see that the RSC-min algorithm is a bit better than RSC. The RSC-min algorithm performs better than RSC in five datasets, and only misclassifies two nodes on UKfaculty. Therefore, using a different degree-correction for each node might improve the performance of the RSC algorithm. We will leave this to our future work.

Table 4. Different methods of degree correction.

DataSet		UKfaculty	Caltech	Dolphins	Karate	Politicalblog	Simmons
RI	RSC	**1**	0.8967	0.9677	**1**	0.9007	0.8521
	RSC-min	0.9646	0.9008	**1**	**1**	**0.9065**	**0.8590**
	RSC-max	0.9834	**0.9018**	**1**	**1**	0.5104	0.8525
	RSC-mean	0.9646	0.8976	**1**	**1**	0.5002	0.8504
	RSC-mid	0.9834	0.9005	**1**	**1**	0.5002	0.8539
NMI	RSC	**1**	0.5881	0.8904	**1**	0.7133	0.6143
	RSC-min	0.8985	0.5953	**1**	**1**	**0.7243**	**0.6228**
	RSC-max	0.9502	**0.6016**	**1**	**1**	0.0227	0.6172
	RSC-mean	0.8985	0.5933	**1**	**1**	0.0019	0.6073
	RSC-mid	0.9502	0.6006	**1**	**1**	0.0019	0.6189
Error rate	RSC	**0/79**	170/590	1/62	**0/34**	64/1222	244/1137
	RSC-min	2/79	**162/590**	**0/62**	**0/34**	**60/1222**	**222/1137**
	RSC-max	1/79	163/590	**0/62**	**0/34**	521/1222	242/1137
	RSC-mean	2/79	170/590	**0/62**	**0/34**	586/1222	240/1137
	RSC-mid	1/79	164/590	**0/62**	**0/34**	586/1222	237/1137

Author Contributions: Conceptualization, W.L.; data curation, W.L. and F.L.; formal analysis, W.L.; funding acquisition, W.L.; investigation, F.L.; methodology, W.L.; project administration, W.L.; resources, W.L. and F.L.; software, F.L.; supervision, W.L.; validation, W.L., F.L. and Y.Z.; visualization, W.L. and F.L.; writing—original draft preparation, W.L.; writing—review and editing, W.L., F.L. and Y.Z.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (No. 11901094).

Data Availability Statement: The data presented in this study are openly available at http://zke.fas.harvard.edu/software.html, accessed on 10 September 2022.

Conflicts of Interest: The authors declare that there are no conflict of interest.

References

1. Adamic, L.A.; Glance, N. The political blogosphere and the 2004 US election: Divided they blog. In Proceedings of the 3rd International Workshop on Link Discovery, Chicago, IL, USA, 21–25 August 2005; pp. 36–43.
2. Hamad, D.; Biela, P. Introduction to spectral clustering. In Proceedings of the 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, Damascus, Syria, 7–11 April 2008; pp. 1–6.
3. Khan, B.S.; Niazi, M.A. Network community detection: A review and visual survey. *arXiv* **2017**, arXiv:1708.00977.
4. Von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **2007**, *17*, 395–416. [CrossRef]

5. Ng, A.; Jordan, M.; Weiss, Y. On spectral clustering: Analysis and an algorithm. *Adv. Neural Inf. Process. Syst.* **2001**, 14, 849–856.
6. Chaudhuri, K.; Chung, F.; Tsiatas, A. Spectral clustering of graphs with general degrees in the extended planted partition model. In Proceedings of the Conference on Learning Theory. JMLR Workshop and Conference Proceedings, Edinburgh, UK, 25–27 June 2012; pp. 1–35.
7. Qin, T.; Rohe, K. Regularized spectral clustering under the degree-corrected stochastic blockmodel. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 3120–3128.
8. Qing, H.; Wang, J. An improved spectral clustering method for community detection under the degree-corrected stochastic blockmodel. *arXiv* **2020**, arXiv:2011.06374.
9. Chung, F.R.K. *Spectral Graph Theory*; CBMS. Reg. Conf. Ser. Math. 92; AMS: Providence, RI, USA, 1997.
10. Kolev, P.; Mehlhorn, K. A Note on Spectral Clustering. In Proceedings of the 24th Annual European Symposium on Algorithms (ESA 2016), Aarhus, Denmark, 22–26 August 2016; Volume 57, pp. 57:1–57:14.
11. Peng, R.; Sun, H.; Zanetti, L. Partitioning well-clustered graphs: Spectral clustering works! In Proceedings of the Conference on Learning Theory, Paris, France, 3–6 July 2015; pp. 1423–1455.
12. Mizutani, T. Improved analysis of spectral algorithm for clustering. *Optim. Lett.* **2021**, *15*, 1303–1325. [CrossRef]
13. Nepusz, T.; Petróczi, A.; Négyessy, L.; Bazsó, F. Fuzzy communities and the concept of bridgeness in complex networks. *Phys. Rev. E* **2008**, *77*, 016107. [CrossRef] [PubMed]
14. Red, V.; Kelsic, E.D.; Mucha, P.J.; Porter, M.A. Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev.* **2011**, *53*, 526–543. [CrossRef]
15. Lusseau, D. The emergent properties of a dolphin social network. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **2003**, *270*, S186–S188. [CrossRef] [PubMed]
16. Zachary, W.W. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **1977**, *33*, 452–473. [CrossRef]
17. Yang, C.; Liu, Z.; Zhao, D.; Sun, M.; Chang, E. Network representation learning with rich text information. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.