Clustering the Nodes of Sparse Edge-Weighted Graphs via Non-Backtracking Spectra

Marianna Bolla[1], Hannu Reittu[2], and Fatma Abdelkhalek[3]

[1] Department of Stochastics, Budapest University of Technology and Economics, Budapest, Hungary
[2] VTT Technical Research Centre of Finland, Espoo, Finland
[3] Faculty of Commerce, Assiut University, Assiut, Egypt
marib@math.bme.hu, Hannu.Reittu@vtt.fi, fatma.said@aun.edu.eg

Abstract. Theoretically supported techniques are given for clustering the nodes of edge-weighted graphs via non-backtracking spectra when the number of nodes is large and the skeleton graph is sparse. If the graph comes from a sparse stochastic block model, the structural real eigenvalues, out of the bulk of the spectrum, of the non-backtracking matrix are aligned with those of the expected adjacency matrix if it is of low rank. However, only the unweighted or weighted non-backtracking matrix is at our disposal. We show how the corresponding eigenvectors of the non-backtracking matrix and lower order companion matrices can be used to find assortative clusters of the nodes even in the case, when the expected adjacency matrix does not have a reduced rank, but it has a low-rank approximation. The paper gives the theoretical background and tools for sparse spectral clustering in very general frameworks. Application to sparse quantum chemistry networks is also presented.

Keywords: Weighted non-backtracking matrix, Stochastic block models, K-means clustering, Non-backtracking eigenvectors

1 Introduction

The purpose of this paper is to illustrate the usage of the non-backtracking matrix for spectral clustering purposes in case of sparse graphs. Classical spectral clustering uses some (k) structural (outstanding) eigenvalues of the unweighted or weighted adjacency, Laplacian, or modularity matrices of the observed simple or edge-weighted graph together with eigenvectors; in case of dense graphs, they mirror the structural eigenvalues and eigenvectors of the expectations of these matrices, see [1–3]. Therefore, by the Weyl's and Davis–Kahan type perturbation theorems for the structural eigenvalues and the corresponding eigensubspaces, the eigenvector-based representatives of the nodes are applicable for metric clustering purposes. The idea behind these methods, though it is not always emphasized, is that there is a dense stochastic block-model in the background, and the reduced rank of the expected adjacency matrix results in some structural eigenvalues of the observed adjacency (or normalized Laplacian or

modularity) matrix. This gives a hint for the number of clusters, and the corresponding eigenvectors are used for the clustering itself; in [4] this is supported with theorems.

However, in the sparse case, the structural eigenvalues of the expected adjacency matrix of a random graph, coming from a percolated k-cluster model, are with high probability (w.h.p.) closer to the structural eigenvalues of the non-backtracking matrix $\boldsymbol{B}$ than to those of the adjacency matrix $\boldsymbol{A}$ of a randomly generated graph from this model, under some balancing conditions for the cluster sizes and average degrees of the nodes. This seems to contradict to the laws of large numbers, but in the sparse case (possessing nearly constant average degrees) it is supported by computations, simulations, and also by theoretical considerations, see, e.g., [5, 6]. Here behavior of deformed Wigner matrices and Bauer–Fike type perturbation results are applicable, see, e.g., [6, 7]. More generally, e.g., for poly-$\log n$ order average degrees there are similar results in [8], in the two-cluster case only, but the author says that it can be extended to the more than two clusters situation too. However, here the reciprocals of the structural real eigenvalues also appear in the bulk of the non-backtracking spectrum.

The structure of the non-backtracking spectrum for simple graphs is throughly analyzed recently. For example, in [9] it is proved that two simple graphs are isomorphic if and only if they have the same non-backtracking matrix. In [10] it is proved that the non-backtracking eigenvalues of unit modulus (in the complex plane) are, in fact, roots of the unity. The authors of [11] investigated when the largest eigenvalue of the non-backtracking matrix is increased dramatically under node removal, causing an abrupt decrease in the percolation threshold for the critical probability of the transmission of an epidemic (they called it node immunization).

In very general sparse stochastic block models (obtained by bond percolation), we show how to find so-called assortative clusters of the nodes with spectral techniques. We also apply the method to smaller parts of networks from quantum chemistry, and with randomized methods elaborated for spectral decomposition of sparse matrices (see, e.g., [12]), the technique can be extended to larger graphs too. For the reduction of the molecular Hamiltonian matrix, the paper [13] uses the modularity matrix, whereas we use the non-backtracking one.

Summarizing, the novelty of the paper is that it deals not only with the eigenvalues, but also with the eigenvectors, corresponding to the structural eigenvalues of the non-backtracking matrix of an edge-weighted graph, so that to cluster the nodes by means of the transformed smaller size vectors.

The organization of the paper is as follows. In Section 2, the notion and properties of non-backtracking matrices are discussed for simple, edge-weighted, and even more generally, for directed graphs. The size of this matrix is $2m$, where m is the number of edges in the skeleton graph. Albeit we have sparse graphs, m is usually much larger than n (the number of nodes), but we use smaller size, so-called companion matrices, the eigenvalues and eigenvectors of which are closely related to those of the non-backtracking matrix. We show, how to find these connections via inflation–deflation techniques. In Section 3, sparse

(percolated) stochastic block models are discussed. A common feature of them is that the expected adjacency matrix of the random graphs coming from these models (after correcting the diagonal) is a low rank ($k \ll n$) matrix with step-vectors as eigenvectors. As the inflated versions of them are close to the leading eigenvectors of the non-backtracking matrix, the k-means algorithm is applicable to the node representatives obtained by the leading k eigenvectors. In Section 4, these possibilities are discussed and supported by theoretical facts. In Section 5, applications in quantum chemistry are presented, whereas Section 6 is devoted to conclusions and further perspectives.

2 Preliminaries

2.1 Non-backtracking Matrix of an Edge-weighted, Undirected or Directed Graph

First the *non-backtracking matrix* $\boldsymbol{B}$ of an edge-weighted, undirected graph $G = (V, \boldsymbol{W})$ is introduced, where V is the set of nodes, $|V| = n$ and $\boldsymbol{W}$ is the $n \times n$ symmetric edge-weight matrix of nonnegative entries and zero diagonal. Since a zero weight means no edge, the number m of edges is the number of the positive upper-diagonal entries of $\boldsymbol{W}$. In [6], the general entry of the $2m \times 2m$ non-backtracking matrix $\boldsymbol{B}$ of G is defined by

$$b_{ef} = W_f \delta_{e \to f} \delta_{f \neq e^{-1}},$$

where $e = \{i \to j\}$ and $f = \{s \to l\}$ are directed edges, $e^{-1} = \{j \to i\}$ is the reverse of edge e and $W_f = w_{sl} = w_{ls}$, so $W_f = W_{f^{-1}}$; further, the shorthand $e \to f$ with $e = (e_1, e_2)$ and $f = (f_1, f_2)$ means that $e_2 = f_1$, while δ is the Kronecker-delta and also the indicator of the event in its lower index. So equivalently,

$$b_{i \to j,\, s \to l} = w_{sl} \delta_{js} (1 - \delta_{il}), \tag{1}$$

and $b_{i \to j,\, s \to l} = w_{sl}$ if and only if for the quadruple in the lower indices, $i \to j = s \to l$ holds, where $l \neq i$; otherwise, it is 0. It may seem that the edges are directed, but each edge is considered in both possible directions, and making the edges bidirected just facilitates the definition of $\boldsymbol{B}$ (it is also related to non-backtracking random walks, see [14] for details).

Another equivalent definition of $\boldsymbol{B}$ is via the unweighted non-backtracking matrix $\boldsymbol{N}$ of the skeleton graph and the $2m \times 2m$ diagonal matrix $\boldsymbol{D}$ containing the positive edge-weights in its main diagonal. (Actually, the diagonal entries of $\boldsymbol{D}$ are the lower- and upper-diagonal non-zero entries of $\boldsymbol{W}$, and the first m diagonal entries are the same as the second m ones.) The entries of $\boldsymbol{N}$ only indicate the connection of the artificially bidirected edges, see [15], and are defined with Equation (1), where the usual 0-1 adjacency matrix $\boldsymbol{A}$ plays the role of $\boldsymbol{W}$. With them,

$$\boldsymbol{B} = \boldsymbol{N}\boldsymbol{D} \quad \text{and} \quad \boldsymbol{B}^* = \boldsymbol{D}\boldsymbol{N}^*, \tag{2}$$

where * denotes the adjoint of a matrix; in case of real matrices, akin to $\boldsymbol{N}$ or $\boldsymbol{B}$, this is the usual transposition, but complex matrices and vectors will also come

into existence in the sequel. Historically (e.g., in [14]), it is the above $\boldsymbol{B}^*$ that is defined as the weighted non-backtracking matrix and $\boldsymbol{N}^*$ as the unweighted one. However, the eigenvalues of $\boldsymbol{B}$ and $\boldsymbol{B}^*$ are the same, just the left and right eigenvectors are interchanged, which does not make too much difference in the subsequent discussion. Also, in [15], only the non-backtracking matrix of an unweighted graph is discussed, and it is denoted by $\boldsymbol{B}$. Here we use the notation $\boldsymbol{N}$ for the non-backtracking matrix of the skeleton of G, and the notation $\boldsymbol{B}$ is preserved for the edge-weighted case.

Actually, $\boldsymbol{N}$ is the adjacency matrix of the line-graph of the directed graph with nodes which are the bidirected edges of the original graph. In this context, the non-backtracking Laplacian is also defined in [9]. As a further perspective, we want to generalize it to edge-weighted graphs, and define the normalized non-backtracking modularity matrix too.

More generally, in [16], to any $n \times n$ matrix $\boldsymbol{W}$ of complex or real entries (even if real, it is not necessary symmetric), the $n^2 \times n^2$ non-backtracking matrix $\boldsymbol{B}$ of entries indexed by the pairs (i, j), (s, l), is introduced akin to in Eq. (1):

$$b_{(i,j),(s,l)} = w_{sl}\delta_{js}(1 - \delta_{il}).$$

In particular, if $\boldsymbol{W}$ is the adjacency matrix of a simple graph, then the above definition of $\boldsymbol{B}$ gives $\boldsymbol{N}$, except of the zero rows/columns. If $\boldsymbol{W}$ is the symmetric edge-weights matrix of zero diagonal, then the maximum size of $\boldsymbol{B}$ is $n(n-1) \times n(n-1)$. In the aforementioned cases, both $\boldsymbol{N}$ and $\boldsymbol{B}$ have the so-called parity time invariance (see [15, 16] and the forthcoming discussion). More generally, if $\boldsymbol{W}$ is the non-symmetric adjacency matrix of a directed graph with possible loops, then the maximum size of $\boldsymbol{B}$ is $n^2 \times n^2$ and it has no parity time invariance.

However, our main concerns are sparse graphs, where the adjacency matrix $\boldsymbol{A}$ (in the unweighted case) or the edge-weight matrix $\boldsymbol{W}$ (in the edge-weighted case) have $o(n)$ non-zero entries in each of their rows, and so, the size of $\boldsymbol{B}$ is much smaller: it is $2m \times 2m$, where m is the number of existing edges (with non-zero weights).

Now let $G = (V, \boldsymbol{W})$ be a directed edge-weighted graph, $|V| = n$ and $\boldsymbol{W}$ is the $n \times n$, usually not symmetric edge-weight matrix of non-negative entries and zero diagonal. Here w_{ij} for $i \neq j$ denotes the weight of the $i \to j$ edge. Consequently, for $i < j$, the $i \to j$ edge-weight (w_{ij}) appears above the main diagonal of $\boldsymbol{W}$, whereas the $j \to i$ edge-weight (w_{ji}) appears below the main diagonal of it (it can occur that one is zero, and the other is not). In the sparse case, we can also reduce the size of $\boldsymbol{B}$ as follows. Let m and m' denote the number of strictly positive weights above and below the main diagonal of $\boldsymbol{W}$, respectively. Also, these positive weights are placed into the first m and m' diagonal positions of the $(m + m') \times (m + m')$ (positive definite) diagonal matrix $\boldsymbol{D}$. Sometimes we refer to these diagonal entries as $W_e := w_{ij} > 0$. The $i < j$ entries appear in the first m, whereas the $i > j$ ones in the next m' positions of the diagonal of $\boldsymbol{D}$. Then $\boldsymbol{B} = \boldsymbol{N}\boldsymbol{D}$, where the non-backtracking matrix $\boldsymbol{N}$ is calculated from the skeleton, and

$$b_{(i,j),(s,l)} = w_{sl}a_{js}(1 - a_{il}),$$

where $\boldsymbol{A} = (a_{ij})$ is the, usually not symmetric, 0-1 adjacency matrix of the directed graph's skeleton ($a_{ij} = 1$ if $w_{ij} > 0$ and 0, otherwise). The companion matrices, to be introduced in Section 2.2, can be defined similarly in the directed case, but here there are in- and out-degrees instead of the usual node-degrees. However, directed graphs are not further treated in this paper.

In [9] it is proved that two simple graphs are isomorphic if and only if their corresponding non-backtracking graphs are isomorphic (non-backtracking matrices are the same if we consider the bidirected edges in the same succession). Under non-backtracking graph we understand the graph on $2m$ nodes with adjacency relation corresponding to the definition of the non-backtracking matrix of the directed graph.

In [15], it is also discussed that if the skeleton of G is a connected graph that is not a cycle and the minimum node degree is at least 2, then $\boldsymbol{N}$, and so, $\boldsymbol{B}$ is irreducible, see also [17]. Therefore, the Perron–Frobenius theory is applicable to $\boldsymbol{B}$, and under the above conditions, its largest absolute value eigenvalue is a single positive real one, with corresponding eigenvector having positive real coordinates. Since the characteristic polynomial of $\boldsymbol{B}$ has real coefficients, its complex eigenvalues occur in conjugate pairs in the bulk of its spectrum.

Also, the following, so-called parity time invariance is true. Let $\boldsymbol{V}$ denote the following involution on $\mathbb{R}^{2m}$ ($\boldsymbol{V} = \boldsymbol{V}^{-1}$, $\boldsymbol{V}^2 = \boldsymbol{I}$; further, $\boldsymbol{V}$ is orthogonal and symmetric at the same time):

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{I}_m \\ \boldsymbol{I}_m & \boldsymbol{O} \end{pmatrix}.$$

Then in [15] we proved that $\boldsymbol{N}^* = \boldsymbol{VNV}$ and $(\boldsymbol{NV})^* = \boldsymbol{VN}^* = \boldsymbol{VVNV} = \boldsymbol{NV}$, so $\boldsymbol{NV}$ and $\boldsymbol{VN}$ are symmetric matrices. Consequently,

$$\boldsymbol{VB}^* = \boldsymbol{VDN}^* = \boldsymbol{DVN}^* = \boldsymbol{DNV} = \boldsymbol{NDV} = \boldsymbol{BV},$$

and $(\boldsymbol{VB}^*)^* = \boldsymbol{BV}$ too. So, $\boldsymbol{VB}^*$ and $\boldsymbol{BV}$ are symmetric matrices, they are diagonalizable in an orthogonal basis, and so, their eigenvalues give the singular values of $\boldsymbol{B}$; however, they only depend on the node-degrees and have noting to do with the eigenvalues of $\boldsymbol{B}$ (same as of $\boldsymbol{B}^*$).

Since $\boldsymbol{B}^* = \boldsymbol{VBV}$, it follows that if $\mathbf{x}$ is a right eigenvector of $\boldsymbol{B}$, then $\breve{\mathbf{x}} := \boldsymbol{V}\mathbf{x}$ is a a right eigenvector of $\boldsymbol{B}^*$ with the same eigenvalue μ. Indeed,

$$\boldsymbol{B}^*\breve{\mathbf{x}} = \boldsymbol{VBV}\breve{\mathbf{x}} = \boldsymbol{VB}\mathbf{x} = \boldsymbol{V}(\mu\mathbf{x}) = \mu\breve{\mathbf{x}}.$$

It is easy to see that $\breve{\mathbf{x}} \in \mathbb{R}^{2m}$ is the swapping of $\mathbf{x} \in \mathbb{R}^{2m}$, i.e., the first and second m coordinates are interchanged. This also implies that the left and right eigenvectors of $\boldsymbol{B}$ are the swappings of each other.

Still, it easily follows by Eq. (2) that the rank of $\boldsymbol{B}$ is the same as that of $\boldsymbol{N}$; further, the number of the structural real eigenvalues of $\boldsymbol{B}$ is inherited from $\boldsymbol{N}$ that is calculated from the skeleton of G. In this way, the edge-weights, included along the diagonal of $\boldsymbol{D}$ will not substantially change the number of clusters under some uniform boundedness conditions for them, if the graph G comes from a sparse stochastic block model to be introduced in Section 3.

In [14], the relation between the eigenvalues of $\boldsymbol{B}^* = \boldsymbol{D}\boldsymbol{N}^*$ and the multivariate *edge zeta function* $\zeta_G(\operatorname{diag}\boldsymbol{D})$ of the underlying edge-weighted graph G is discussed, and it is proved that

$$\zeta_G^{-1}(\operatorname{diag}\boldsymbol{D}) = |\boldsymbol{I}_{2m} - \boldsymbol{D}\boldsymbol{N}^*|,$$

i.e., it is the product of the transformed eigenvalues $1 - \lambda_i(\boldsymbol{B}^*) = 1 - \lambda_i(\boldsymbol{B})$. In the unweighted case (all W_e's are equal to 1), the Ihara's formula (see [15, 17]) further expands the above determinant with $n \times n$ matrices. A similar expansion is proved in Theorem 2 of [14] in the weighted case with a Laplacian type operator that will be revisited later, in Proposition 1.

Note that in case of an unweighted graph, the Ihara's formula implies that $\boldsymbol{N}$ has $m-n$ eigenvalues equal to 1 and $m-n$ eigenvalues equal to -1, whereas its further eigenvalues are those of the $2n \times 2n$ matrix

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{F} - \boldsymbol{I}_n \\ -\boldsymbol{I}_n & \boldsymbol{A} \end{pmatrix}, \tag{3}$$

where $\boldsymbol{A}$ is the adjacency matrix and the diagonal entries of the diagonal matrix $\boldsymbol{F}$ are the usual node-degrees.

2.2 Eigenvectors and Companion Matrices

To ease the discussion, some notation and assumptions will be introduced. We assume that the non-zero weights are uniformly bounded from above and from below: there are constants C_1 and C_2 (independent of n) such that

$$C_1 \le w_{ij} \le C_2, \quad \text{for} \quad w_{ij} \neq 0. \tag{4}$$

Without hurting the generality, $0 < C_1 \le C_2 = 1$ can be assumed. (For fixed n, in a real-life edge-weighted graph we can tailor the weights, by making the "small" weights equal to 0 so that to get a sparse graph). Further, we assume that, with the possibly manipulated edge-weights, the node degrees

$$d_i = |\{i : w_{ij} > 0, \quad j = 1, \dots, n\}|, \quad i = 1, \dots, n \tag{5}$$

are of order $o(n)$ (usually of constant order), with increasing n. This is the case in the k-cluster stochastic block models, where the weights are defined to be 1, but edges come into existence with probability that is abruptly decreasing with n.

Let $\boldsymbol{D}^{\boldsymbol{W}}$ denote the $n \times n$ diagonal matrix of diagonal entries

$$d_i^{\boldsymbol{W}} = \sum_{j=1}^{n} w_{ij}, \quad i = 1, \dots, n,$$

that are the so-called generalized degrees. In the unweighted case (0-1 weights), $d_i^{\boldsymbol{W}} = d_i$, and d_i's are the usual node degrees; further, $C_1 = C_2 = 1$. In general,

$$C_1 d_i \le d_i^{\boldsymbol{W}} \le C_2 d_i, \quad i = 1, \dots, n. \tag{6}$$

Two auxiliary matrices, defined in [6], will also be used: the $2m \times n$ *end matrix* $\boldsymbol{End}$ has entries $end_{ei} = 1$ if i is the end-node of the (directed) edge e and 0, otherwise; the $2m \times n$ *start matrix* $\boldsymbol{Start}$ has entries $start_{ei} = 1$ if i is the start-node of the (directed) edge e and 0, otherwise. Then for any vector $\mathbf{u} \in \mathbb{R}^n$ and for any edge $e = \{i \to j\}$ the following holds:

$$(\boldsymbol{End}\,\mathbf{u})_e = u_j \quad \text{and} \quad (\boldsymbol{Start}\,\mathbf{u})_e = u_i.$$

Consequently, $\boldsymbol{End}\,\mathbf{u}$ is the $2m$-dimensional inflated version of the n-dimensional vector $\mathbf{u}$, where the coordinate u_j of $\mathbf{u}$ is repeated as many times, as many edges (with positive weight) have end-node j; $\boldsymbol{Start}\,\mathbf{u}$ is the $2m$-dimensional inflated version of the n-dimensional vector $\mathbf{u}$, where the coordinate u_j of $\mathbf{u}$ is repeated as many times, as many edges (with positive weight) have start-node j; as each edge is considered in both possible directions, this number is just d_j. Trivially,

$$\boldsymbol{End}^*\,\boldsymbol{D}\,\boldsymbol{End} = \boldsymbol{Start}^*\,\boldsymbol{D}\,\boldsymbol{Start} = \boldsymbol{D}^{\boldsymbol{W}} \quad \text{and} \quad \boldsymbol{Start}^*\,\boldsymbol{D}\,\boldsymbol{End} = \boldsymbol{W}.$$

In [18], the following n-dimensional (column) vectors $\mathbf{x}^{out}$ and $\mathbf{x}^{in}$ are introduced in the unweighted situation. Now they are extended to the edge-weighted case.

For any vector $\mathbf{x} \in \mathbb{R}^{2m}$, the following n-dimensional vectors are defined:

$$\mathbf{x}^{out} := \boldsymbol{Start}^*\boldsymbol{D}\mathbf{x} \quad \text{and} \quad \mathbf{x}^{in} := \boldsymbol{End}^*\boldsymbol{D}\mathbf{x}. \tag{7}$$

Coordinatewise, for $i = 1, \ldots, n$,

$$x_i^{out} = \sum_{j:\, j\sim i} w_{ij}x_{i\to j} = \sum_{e:\, e_1=i} W_e x_e \quad \text{and} \quad x_i^{in} = \sum_{j:\, j\sim i} w_{ij}x_{j\to i} = \sum_{e:\, e_2=i} W_e x_e.$$

In the general edge-weighted situation, unfortunately, we cannot trace back the problem to the eigenvalue-eigenvector decomposition of a $2n \times 2n$ matrix, akin to the Ihara-formula in the unweighted case, see [15]. However, by a technique similar to that of [6], if we know a real eigenvalue μ of $\boldsymbol{B}$, we are able to find a linear system of equations for the *out*-transform of the corresponding eigenvector that is necessary for spectral clustering in Section 4. Then, with a Lapacian type equation, μ can also be concluded.

We will use Proposition 1 of [6] for the leading real eigenvalues and corresponding eigenvectors of $\boldsymbol{B}$.

Proposition 1. *Let* $\mathbf{x}$ *be a (right) eigenvector of* $\boldsymbol{B}$ *corresponding to a single positive real eigenvalue* μ *such that* $\mu \neq w_{ij}$, $\forall i, j \in \{1, \ldots, n\}$. *Let* $\mathbf{y} := \mathbf{x}^{out}$. *Then* $\mathbf{y}$ *satisfies the homogeneous system of linear equations*

$$[\boldsymbol{I}_n - \tilde{\boldsymbol{A}}(\mu) + \tilde{\boldsymbol{D}}(\mu)]\mathbf{y} = \mathbf{0}, \tag{8}$$

with a Laplacian type coefficient matrix, where the matrix $\tilde{\boldsymbol{A}}(\mu)$ *and the diagonal matrix* $\tilde{\boldsymbol{D}}(\mu)$ *have entries*

$$\tilde{\boldsymbol{A}}(\mu)_{ij} = \frac{\mu w_{ij}}{\mu^2 - w_{ij}^2} \quad \textit{and} \quad \tilde{\boldsymbol{D}}(\mu)_{ii} = \sum_{j=1}^{n} \frac{w_{ij}^2}{\mu^2 - w_{ij}^2}, \tag{9}$$

with the understanding that $w_{ij} = 0$ *whenever* $i \not\sim j$.

This result is also supported by [14, 16, 19] and will be substantially used in the next section. To be self-contained, we include a simplified proof.

Proof. If $\mathbf{x}$ is an eigenvector of $\boldsymbol{B}$ with corresponding eigenvalue μ, then

$$\mu x_e = \sum_{e\to f,\, f\neq e^{-1}} W_f x_f = \sum_{f:f_1=e_2} W_f x_f - W_{e^{-1}} x_{e^{-1}} = y_{e_2} - W_e x_{e^{-1}}. \qquad (10)$$

Likewise,

$$\mu x_{e^{-1}} = \sum_{e^{-1}\to f,\, f\neq e} W_f x_f = \sum_{f:f_1=e_1} W_f x_f - W_e x_e = y_{e_1} - W_e x_e.$$

From here,

$$\mu^2 x_e = \mu y_{e_2} - \mu W_e x_{e^{-1}} = \mu y_{e_2} - W_e y_{e_1} + W_e^2 x_e,$$

and so,

$$x_e = \frac{\mu y_{e_2} - W_e y_{e_1}}{\mu^2 - W_e^2}$$

that shows that $\mathbf{y} \neq \mathbf{0}$ as $\mathbf{x} \neq 0$. Substituting this formula for x_e in Eq. (10), we get that for any edge $e = \{j \to i\}$,

$$\frac{\mu^2 y_i - \mu w_{ij} y_j}{\mu^2 - w_{ij}^2} = \sum_{l:l\sim i,\, l\neq j} w_{li} \frac{\mu y_l - w_{li} y_i}{\mu^2 - w_{li}^2}.$$

Further developing, we get that

$$\frac{\mu^2 y_i}{\mu^2 - w_{ij}^2} - \frac{\mu w_{ij} y_j}{\mu^2 - w_{ij}^2} = \sum_{l:l\sim i} \frac{\mu w_{li}}{\mu^2 - w_{li}^2} y_l - \sum_{l:l\sim i} \frac{w_{li}^2}{\mu^2 - w_{li}^2} y_i - w_{ji} \frac{\mu y_j - w_{ji} y_i}{\mu^2 - w_{ji}^2},$$

which provides

$$\frac{\mu^2 y_i}{\mu^2 - w_{ij}^2} - \frac{w_{ij}^2 y_i}{\mu^2 - w_{ij}^2} = \sum_{l:l\sim i} \frac{\mu w_{li}}{\mu^2 - w_{li}^2} y_l - \sum_{l:l\sim i} \frac{w_{li}^2}{\mu^2 - w_{li}^2} y_i.$$

This proves Eq. (8).

Eq. (8) is a system of homogeneous linear equations for the coordinates of $\mathbf{y}$, and to get a non-trivial solution, the determinant equation

$$|\boldsymbol{I}_n - \tilde{\boldsymbol{A}}(\mu) + \tilde{\boldsymbol{D}}(\mu)| = 0 \qquad (11)$$

should hold. This is not a polynomial (characteristic) equation, but it is a rational function of μ. Since, by the assumptions of Proposition 1, the denominators are not zeros, we can multiply the determinant equations with them, and we obtain an at most n^2-degree polynomial of μ. The leading positive real solutions $\mu_1 \geq \cdots \geq \mu_k$ are considered that, by Proposition 1, should be the same as the structural eigenvalues of $\boldsymbol{B}$. Their number will be denoted by k. The corresponding $\mathbf{y}_1, \ldots, \mathbf{y}_k$ can be obtained by solving the system of homogeneous linear Eq. (8), and so, providing the vectors $\mathbf{x}_1^{out}, \ldots, \mathbf{x}_k^{out}$ for clustering purposes in Section 4.

3 Sparse Stochastic Block Models

The general discussion of Stephan and Massoulié [6] is concentrated on models where the expected adjacency matrix is of reduced rank, and so, $\boldsymbol{B}$ has some outstanding real eigenvalues too. The authors introduce $\boldsymbol{P} = (p_{ij})$ as a probability matrix containing the edge-probabilities, and $\mathbb{W} = (W_{ij})$ as the matrix of random weights of edges. The number of nodes is increasing. At the instance, when the this number is n, both $\boldsymbol{P}$ and $\mathbb{W}$ are symmetric real matrices of size $n \times n$. Their theory is applicable if $\bar{\boldsymbol{A}} := \boldsymbol{P} \circ \mathbb{E}\mathbb{W}$ is a low rank matrix, where $\circ$ denotes the Hadamard (entry-wise) product of matrices, and the so obtained graph is sparse enough. A constant average degree can be guaranteed if, in the instance of n nodes, the p_{ij}'s are proportional to $\frac{1}{n}$. The authors of [6] only require for the average degree to be of order $o(\log n)$. In the classical literature, for the average degrees, the order $o(n)$ is considered as sparse. Nowadays the notion of intermediate density is introduced, e.g., for $\log n$ or poly-$\log n$ order average degrees, which is the case in the subsequent quantum chemistry examples. For the general treatment of this situation, see also [8], where the bulk of the spectrum of $\boldsymbol{N}$ and its eigenvalues inside and outside the bulk are completely characterized.

However, the stochastic block models, discussed in [6, 15] and briefly introduced below, are special cases, where the weights are constantly 1. Another treatable case is when $\mathbb{W}$ has Gaussian entries, highly concentrated to their expectation and with variances decreasing with n. In this case the matrix $\boldsymbol{P} \circ \mathbb{E}\mathbb{W}$ is the expected adjacency matrix, and $\boldsymbol{P} \circ \boldsymbol{W}$ is close to it, where $\boldsymbol{W}$ contains the observed weights.

We consider the practical situation when the entry W_{ij} of the $n \times n$ random weighted adjacency matrix $\mathbb{W}$ is $\widetilde{w}_{ij}$ times a Bernoulli distributed random variable with parameter p_{ij}, for $1 \leq i < j \leq n$; these entries above the diagonal are independent of each other, while those below the diagonal are identical to them. So the parameters of this distribution are contained in the symmetric matrices $\boldsymbol{P}$ and $\widetilde{\boldsymbol{W}} = (\widetilde{w}_{ij})$ of real entries in (0,1]. In this way, the expected adjacency matrix is $\bar{\boldsymbol{A}} = \boldsymbol{P} \circ \widetilde{\boldsymbol{W}}$, with approximate matrix of variances $\boldsymbol{P} \circ \widetilde{\boldsymbol{W}} \circ \widetilde{\boldsymbol{W}}$ (if the entries of $\boldsymbol{P}$ are of order $\frac{1}{n}$), and so, the theory of [6] is applicable to it. However, for given (large) n, we only observe a realization $\boldsymbol{W}$ from the the distribution $\mathbb{W}$, in the non-zero positions of which the entries are equal to those of $\widetilde{\boldsymbol{W}}$.

Proposition 2 below suggests that only $\boldsymbol{B}$ (calculated by the observed $\boldsymbol{W}$) is at our disposal. If there is a remarkable spectral gap after its kth largest (real) eigenvalue (or else, the number of positive real eigenvalues that are isolated from the bulk is k) and the corresponding eigenvector based representatives well cluster into k parts, this is an indication that the unknown expected adjacency matrix has a good k-rank approximation, see Section 4. (In Section 6, it will be forecasted that a block-matrix approximation is as well possible.) The point is that even if the observed $\boldsymbol{W}$ is sparse, the expected weighted adjacency matrix $\bar{\boldsymbol{A}}$ is full, and usual matrix techniques of spectral clustering are applicable with it theoretically. But in practice, we treat only $\boldsymbol{B}$.

We shall use the following general (informal) statement of [6]. The formal statement uses many parameters, the definition of which is quite not trivial; further, the proof of it needs several auxiliary theorems and propositions, so this will be waived.

Proposition 2 (Based on Theorem 1 of [6]). *Assume that the rank of the matrix $\bar{\boldsymbol{A}} = \boldsymbol{P} \circ \mathbb{E}\mathbb{W}$ is $k = n^{o(1)}$, the graph is sparse enough, and the eigenvectors, corresponding to the non-zero eigenvalues of the matrix $\bar{\boldsymbol{A}}$, are sufficiently delocalized. Let k_0 denote the number of eigenvalues of $\bar{\boldsymbol{A}}$ whose absolute value is larger than $\sqrt{\rho}$, where ρ is the spectral radius of the matrix $\boldsymbol{P} \circ \mathbb{E}(\mathbb{W} \circ \mathbb{W})$: these are $\nu_1 \geq \cdots \geq \nu_{k_0}$ with corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_{k_0}$ (they form an orthonormal system as $\bar{\boldsymbol{A}}$ is a real symmetric matrix). Then for $i \leq k_0 \leq k$, the ith largest eigenvalue μ_i of $\boldsymbol{B}$ is asymptotically (w.h.p. as $n \to \infty$) equals to ν_i and all the other eigenvalues of $\boldsymbol{B}$ are constrained to the circle (in the complex plane) of center 0 and radius $\sqrt{\rho}$. Further, if $i \leq k_0$ is such that ν_i is a sufficiently isolated eigenvalue of $\bar{\boldsymbol{A}}$, then the standardized eigenvector of $\boldsymbol{B}$ corresponding to μ_i has inner product close to 1 with the standardized inflated version of $\mathbf{u}_i$, namely, with $\frac{\boldsymbol{End}\,\mathbf{u}_i}{\|\boldsymbol{End}\,\mathbf{u}_i\|}$.*

Unfortunately, we do not know the matrix $\bar{\boldsymbol{A}}$ and its leading eigenvalues; but if our graph is from a certain block model, then we can conclude for those via the non-backtracking spectrum. The observed adjacency or edge-weight matrix do not help much in this issue, as the leading eigenvalues of those are farther from their expectation as the corresponding quantities of the expectation are from those of the non-backtracking matrix, see [6, 14, 20]. Now the statement of Proposition 2 is applied to specific models, where the matrix $\bar{\boldsymbol{A}}$ is not only the diagonal-corrected expected adjacency matrix, but it is approximately the variance matrix of the random adjacency matrix $\boldsymbol{A}$ too, in the k-block model.

In [5, 15, 20–22] the sparse SBM_k model is considered with the following parameters. The $k \times k$ *probability matrix* $\boldsymbol{P}$ has entries

$$p_{ab} = p_{ba} = \frac{c_{ab}}{n}, \quad 1 \leq a \leq b \leq k,$$

where the $k \times k$ symmetric *affinity matrix* $\boldsymbol{C} = (c_{ab})$ stays constant as $n \to \infty$. If a random graph G_n on node-set $V = \{1, \ldots, n\}$ comes from this model, then an edge between $i < j$ comes into existence, independently of the others, with probability p_{ab} if $i \in V_a$ and $j \in V_b$, where $(V_1, \ldots, V_k)$ is a partition of V into k disjoint clusters. This will produce the upper-diagonal part of the $n \times n$ random adjacency matrix $\boldsymbol{A}$, and $a_{ji} := a_{ij}$. It can be extended to the $i = j$ case when self-loops are allowed, or else, the diagonal entries of the adjacency matrix are zeros.

In this unweighted case the weights are 1's, and $\bar{\boldsymbol{A}}$ is the $n \times n$ inflated matrix of the $k \times k$ matrix $\boldsymbol{P}$: $\bar{a}_{ij} = p_{ab}$ if $i \in V_a$ and $j \in V_b$. When loops are allowed, then $\mathbb{E}(a_{ij}) = \bar{a}_{ij}$ for all $1 \leq i, j \leq n$. In the loopless case, the expected adjacency matrix $\mathbb{E}\boldsymbol{A}$ differs from $\bar{\boldsymbol{A}}$ with respect to the main diagonal, but the diagonal entries are negligible, see [15] for details. In addition to $\boldsymbol{C}$, another

fixed model parameter is introduced that governs the growth of the clusters. Let $n_1, \dots, n_k$ be the cluster sizes (positive integers with $\sum_{a=1}^{k} n_a = n$). Then the $k \times k$ diagonal matrix $\boldsymbol{R} := \operatorname{diag}(r_1, \dots, r_k)$, where $r_a = \frac{n_a}{n}$ is the relative size of cluster a ($a = 1, \dots, k$), is also a model parameter ($\sum_{a=1}^{k} r_a = 1$). Usually, r_a's are fixed, or else, the diagonal of $\boldsymbol{R}$ tends to a probability vector as $n \to \infty$.

The expected average degree of the random graph $G_n \in SBM_k$ is

$$c = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \bar{a}_{ij} = \frac{1}{n} \sum_{a=1}^{k} \sum_{b=1}^{k} n_a n_b p_{ab} = \frac{1}{n^2} \sum_{a=1}^{k} \sum_{b=1}^{k} n_a n_b c_{ab} = \sum_{a=1}^{k} r_a c_a, \quad (12)$$

where $c_a = \sum_{b=1}^{k} r_b c_{ab}$ is the average degree of cluster a. In [5], the case when $c_a = c$, for all a, is considered. (The authors say that this assumption is not too restrictive, as otherwise the clusters could be distinguished by sorting the node-degrees.) Also note that in the SBM_k model, when $c_a = c$ ($a = 1, \dots, k$), then in Proposition 2: $\rho = c$ and $k_0 = k$; also, $\boldsymbol{B} = \boldsymbol{N}$.

Sometimes $c_{ab} = c_{in}$ is the within-cluster ($a = b$) and $c_{ab} = c_{out}$ is the between-cluster ($a \neq b$) affinity. In [2], the network is called *assortative* if $c_{in} > c_{out}$, and *disassortative* if $c_{in} < c_{out}$. Of course, remarkable difference is needed between the two, if one wants to recognize the clusters.

If there are only c_{in} and c_{out} affinities and $c_1 = \cdots = c_k$, then $r_1 = \cdots = r_k = \frac{1}{k}$ too. In this case, the model SBM_k is called *symmetric*. Then

$$c = \frac{c_{in} + (k-1)c_{out}}{k},$$

and the separation of the clusters only depends on the c_{in}, c_{out} relation. The detectability threshold (Kesten–Stigum threshold) in the symmetric case is

$$|c_{in} - c_{out}| > k\sqrt{c}, \quad (13)$$

see [5, 20, 23].

Note that if our random graph comes from the SBM_k model with $k \times k$ parameter matrices $\boldsymbol{R}$ and $\boldsymbol{C}$, then $\bar{\boldsymbol{A}}$ has rank $k_0 \leq k$ (so Proposition 2 is applicable), and its nonzero eigenvalues are identical to the real eigenvalues of the matrix $\boldsymbol{RC}$ (same as those of $\boldsymbol{R}^{\frac{1}{2}}\boldsymbol{C}\boldsymbol{R}^{\frac{1}{2}}$), and the eigenvectors of it are inflated versions of those of $\boldsymbol{R}^{\frac{1}{2}}\boldsymbol{C}\boldsymbol{R}^{\frac{1}{2}}$, see [15]. Consequently, they are step-vectors on k different steps and so, the k-means algorithm is applicable to the k_0-dimensional representatives of the nodes constructed with the k_0 normalized eigenvectors of $\boldsymbol{N}$, corresponding to its leading eigenvalues $\mu_1, \dots, \mu_{k_0}$, see Section 4. The point is that we do not need the model parameters and the spectral decomposition of $\bar{\boldsymbol{A}}$ itself for the k-means clustering, see [15] for details.

Even if the condition $c_1 = \cdots = c_k = c$ does not hold, in [15] we applied the method of belief propagation so that to conclude for the number of clusters; see also [23, 24]. Assume that the number of clusters is k and a graph on n nodes comes from the SBM_k model. For $a = 1, \dots, k$, let $\psi^a_{j \to i}$, denote the marginal (in other words, state or membership) of node j if i were not there (more precisely, if

we did not know whether or not there is an edge between i and j). If we assume that our neighbors are correlated only through us, the overall state of the graph can be modeled by having each node j send a message to its neighbors, and the messages together must give the overall truth. The conditional probabilities

$$\psi^a_{j\to i} := \mathbb{P}(j \text{ is in cluster } a \text{ when } i \text{ is not present})$$

are computed through the neighbors of j that are different from i (in the realization of the random graph coming from the SBM_k model) as follows:

$$\psi^a_{j\to i} = C^{ij}_a r_a \prod_{l\sim j,\, l\neq i} \sum_{b=1}^{k} \psi^b_{l\to j}\, p_{ab}, \quad a = 1, \dots, k, \tag{14}$$

where C^{ij}_a is a normalizing factor, see [15].

This message-passing equation is a system of $2mk$ non-linear equations with the same number of unknowns. It can be solved by initializing messages randomly, then repeatedly updating them via the fixed point iteration $\boldsymbol{\varepsilon}^{(t+1)} = \boldsymbol{M}\boldsymbol{\varepsilon}^{(t)}$ $(t = 1, 2, \dots)$, where $\varepsilon^a_{j\to i} = \psi^a_{j\to i} - r_a$ and the $2mk \times 2mk$ matrix $\boldsymbol{M}$ corresponds to the linear approximation of the system of which the zero vector is a trivial fixed point (it gives the normalizing factors too). In [15] it is shown that $\boldsymbol{M}$ is a Kronecker-product, $\boldsymbol{M} = \boldsymbol{N} \otimes \boldsymbol{T}$, where $\boldsymbol{T} = \boldsymbol{GRC}$ is the *transmission matrix* with $\boldsymbol{G} = \mathrm{diag}\,(\frac{1}{c_1}, \dots, \frac{1}{c_k})$. We can find a fixed point other than the trivial $\mathbf{0}$ when $\mathbf{0}$ is not a globally asymptotically stable solution. For this, a sufficient condition is that the spectral radius of the matrix $\boldsymbol{M}$ is greater than 1.

Note that in [23] only the symmetric case is treated. In [5], the special case $c_1 = \cdots = c_k = c$ is considered when the matrix $\boldsymbol{T}$ becomes $\boldsymbol{T} = \frac{1}{c}\boldsymbol{RC}$; then the leading eigenvalues of $\boldsymbol{N}$ and $\boldsymbol{RC}$ are w.h.p. "close" to each other. Also, the largest eigenvalue of $\boldsymbol{RC}$ is c, which is trivially the case if $k = 1$ and we have the Erdős–Rényi random graph [25]. The authors of [5] also allow "small" fluctuations of the cluster membership proportions that causes the same order of fluctuations in the average degrees of the clusters. Even in this case they prove that the leading eigenvalues of $\boldsymbol{N}$ and $\boldsymbol{RC}$ are aligned (w.h.p.), and so, the eigenvalues of their Kronecker product $\boldsymbol{N} \otimes \boldsymbol{RC}$ should satisfy $\lambda(\boldsymbol{N})\lambda(\boldsymbol{RC}) > c$. Consequently, the eigenvalues of $\boldsymbol{N}$ greater than $\sqrt{c}$ should be taken into consideration.

4 Spectral Clustering of Sparse Edge-weighted, Undirected Graphs

We will use that, by [6], if $\mathbf{x}$ is a unit-norm eigenvector of $\boldsymbol{B}$, corresponding to a structural real eigenvalue μ which is close to an eigenvalue ν of the expected adjacency matrix with corresponding unit-norm eigenvector $\mathbf{u} \in \mathbb{R}^n$, then $\mathbf{x}$ is also "close" to $\mathbf{u}$ (in the sense that their inner product is approaching 1 for "large" n). It is valid if the expected adjacency matrix has low rank k (or at least it has a good k-rank approximation), where k does not depend on n.

If our graph is from the SBM_k model, then (without knowing its parameters) we know that $\mathbf{u}$ is a step-vector with k different coordinates, and in [15] we estimate the objective function of the k-means algorithm with a term that tends to 0 w.h.p. as $n \to \infty$.

More generally, under the conditions of Proposition 2, the relation

$$\left\langle \mathbf{x}, \frac{\boldsymbol{End}\mathbf{u}}{\|\boldsymbol{End}\mathbf{u}\|} \right\rangle \geq \sqrt{1-\varepsilon} \geq 1 - \frac{1}{2}\varepsilon$$

holds for any eigenvector $\mathbf{x}_i$ corresponding to an eigenvalue μ_i of $\boldsymbol{B}$ which is separated from the bulk, and the signal to noise ratio, $\max_{\mu_1<\mu_i^2} \frac{\mu_1}{\mu_i^2}$, is "small". In this case, ε can be arbitrarily "small" with increasing n. Here we use the k structural eigenvalues of the weighted non-backtracking matrix $\boldsymbol{B}$. Furthermore, if the matrix $\bar{\boldsymbol{A}}$ has step-vectors as leading eigenvectors (as in the SBM_k model), then the k-means algorithm is applicable.

Consequently,

$$\left\| \mathbf{x} - \frac{\boldsymbol{End}\mathbf{u}}{\|\boldsymbol{End}\,\mathbf{u}\|} \right\|^2 \leq 2 - 2(1 - \frac{1}{2}\varepsilon) = \varepsilon,$$

which in view of $\boldsymbol{Start}^*\boldsymbol{D}\mathbf{x} = \mathbf{x}^{out}$ and $\boldsymbol{Start}^*\boldsymbol{D}\boldsymbol{End} = \boldsymbol{W}$ gives rise to the inequality

$$\left\| \mathbf{x}^{out} - \boldsymbol{W}\frac{\mathbf{u}}{\|\boldsymbol{End}\,\mathbf{u}\|} \right\|^2 = \left\| \boldsymbol{Start}^*\boldsymbol{D}\left(\mathbf{x} - \frac{\boldsymbol{End}\,\mathbf{u}}{\|\boldsymbol{End}\,\mathbf{u}\|} \right) \right\|^2 \leq \|\boldsymbol{Start}^*\boldsymbol{D}\|^2\varepsilon.$$

Therefore, $\boldsymbol{W}^{-1}\mathbf{x}^{out}$ is close to the appropriately normalized $\mathbf{u}$:

$$\left\| \boldsymbol{W}^{-1}\mathbf{x}^{out} - \frac{\mathbf{u}}{\|\boldsymbol{End}\,\mathbf{u}\|} \right\|^2 \leq \|\boldsymbol{W}^{-1}\boldsymbol{Start}^*\boldsymbol{D}\|^2\varepsilon.$$

Indeed, the largest eigenvalue of $\boldsymbol{Start}^*\boldsymbol{D}^2\boldsymbol{Start} = (\boldsymbol{D}^{\boldsymbol{W}\circ\boldsymbol{W}})^2$ is $max_{ij}w_{ij}^2$, so the largest singular value, i.e., the spectral norm of $\boldsymbol{Start}^*\boldsymbol{D}$ is $max_{ij}w_{ij} \leq C_2$. The spectral norm of $\boldsymbol{W}^{-1}$ is at most $\frac{1}{C_1}$. So $\|\boldsymbol{W}^{-1}\boldsymbol{Start}^*\boldsymbol{D}\|^2 \leq \frac{C_2^2}{C_1^2}$.

Now we apply this to the k leading normalized eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ of $\boldsymbol{B}$ and so, the following theorem is proved.

Theorem 1. *Assume that the expected adjacency matrix of the underlying random graph on n nodes and m edges has rank k with k single non-zero eigenvalues and corresponding unit-norm eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathbb{R}^n$. Assume that the non-backtracking matrix $\boldsymbol{B}$ of the random graph has k structural eigenvalues (aligned with those of the expected adjacency matrix) with eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^{2m}$ such that*

$$\left\langle \mathbf{x}_j, \frac{\boldsymbol{End}\,\mathbf{u}_j}{\|\boldsymbol{End}\,\mathbf{u}_j\|} \right\rangle \geq \sqrt{1-\varepsilon}, \quad j = 1, \ldots, k. \tag{15}$$

Then for the transformed vectors $\boldsymbol{W}^{-1}\mathbf{x}_j^{out} \in \mathbb{R}^n$*, the relation*

$$\sum_{j=1}^{k} \left\| \boldsymbol{W}^{-1}\mathbf{x}_j^{out} - \frac{\mathbf{u}_j}{\|\boldsymbol{End}\,\mathbf{u}_j\|} \right\|^2 \le k\varepsilon \frac{C_2^2}{C_1^2}. \tag{16}$$

holds.

If our graph comes from the stochastic block model with k clusters (k is the number of structural eigenvalues of $\boldsymbol{B}$), then $\mathbf{u}_j$'s are step-vectors with k different coordinates on the same k steps and the objective function of the k-means algorithm is less than or equal to the above sum of the squares. Without knowing the $\mathbf{u}_j$'s, we minimize the k-means objective with the k-dimensional node representatives

$$\left(\boldsymbol{W}^{-1}\mathbf{x}_1^{out}, \dots \boldsymbol{W}^{-1}\mathbf{x}_k^{out}\right)_i, \quad i = 1, \dots, n.$$

Therefore, based on the spectral gap in $\boldsymbol{B}$ (k is the number of its structural eigenvalues) and its k leading eigenvectors, if the representatives constructed from them by the above formula "well" cluster into k parts, then without knowing $\boldsymbol{P}$ or $\boldsymbol{W}$, we can cluster the nodes of the original edge-weighted graph. For this, only the $\mathbf{x}_1^{out}, \dots, \mathbf{x}_k^{out}$ vectors are used.

5 Application

As an example, we present a network from quantum chemistry, where the molecular Hamiltonian matrix is considered in the so-called Slater determinant basis. Slater determinants describe wave functions for a collection of electrons, each with a wave function known as spin-orbital, see [13, 26–28].

In our small sample example we consider a graph on $n = 133$ nodes, obtained from the water molecule. The connectivity between the Slater determinants are the edge-weights of the graph, whereas the skeleton just detects these connectivities.

In the subsequent illustrations $n = 133$ and $m = 3032$. In Fig. 1, the 22 real eigenvalues of the the $2n \times 2n$ matrix $\boldsymbol{K}$ are plotted, which together with number $m - n$ of $+1$'s and $m - n$ of -1's constitute the real eigenvalues of the $2m \times 2m$ matrix $\boldsymbol{N}$, the non-backtracking matrix corresponding to the unweighted skeleton (the other eigenvalues are complex numbers). Note that the largest eigenvalue of $\boldsymbol{K}$ is near to the average degree $c = 45.594$ and the smallest one is a single 1 (indicating that the graph is connected, i.e., has one connected component). Positive real eigenvalues indicate assortative clusters. The number of the real eigenvalues larger than $\sqrt{c} = 6.752$ is 11, so the number of clusters will be $k = 11$.

Fig. 2 plots the first fifty largest absolute value eigenvalues (they are real positive) of the weighted non-backtracking matrix $\boldsymbol{B}$, obtained by the edge-weight matrix $\boldsymbol{W}$. (The weights are transformed into the [0,1] interval, and "small" weights are made equal to zero.) A gap can be inspected after the fourth and

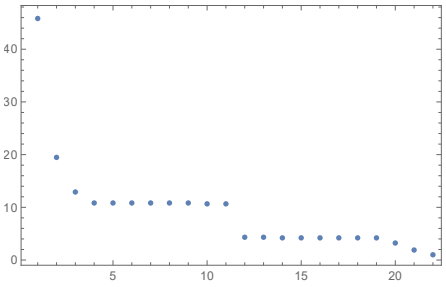


Fig. 1. 22 real eigenvalues of the $2n \times 2n$ matrix $\boldsymbol{K}$, $n = 133$

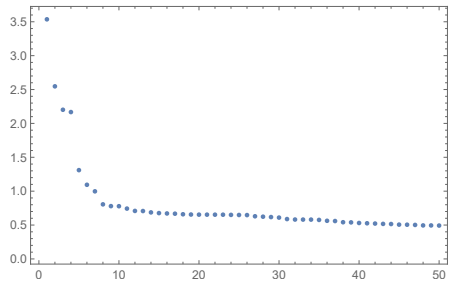
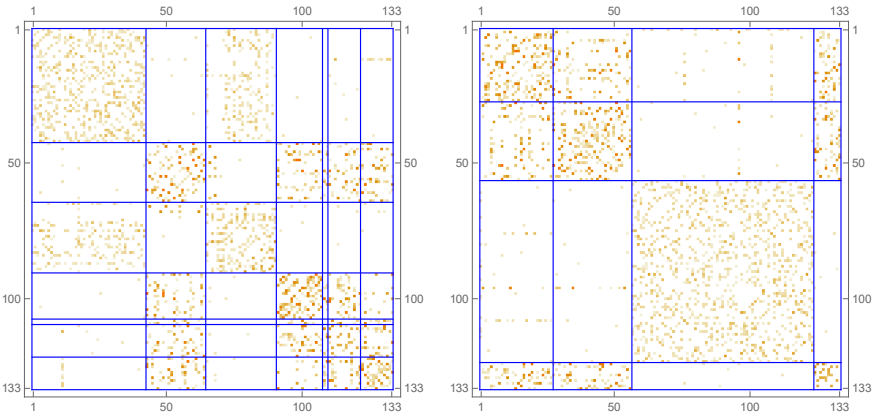


Fig. 2. Leading 50 real eigenvalues of the $2m \times 2m$ matrix $\boldsymbol{B}$, $m = 3032$

seventh eigenvalues in decreasing order. Therefore 7 and 4 clusters are obtained by the k-means algorithm applied to the out-vectors corresponding to the leading $\boldsymbol{B}$-eigenvectors, see Fig. 3.



(a) Number of clusters = 7. One cluster is very "small"

(b) Number of clusters = 4

Fig. 3. Number of clusters in the appropriate labeling of the nodes, obtained by the k-means algorithm applied to the out-vectors corresponding to the leading $\boldsymbol{B}$-eigenvectors. The clusters are sketched by coloring the weights of the matrix $\boldsymbol{W}$, where darker colors correspond to larger edge-weights.

Fig. 4 shows the 11 clusters in the appropriate labeling of the nodes, obtained by the k-means algorithm applied to the in-vectors corresponding to the leading $\boldsymbol{K}$-eigenvectors. This clustering technique uses the unweighted skeleton graph only, see [15]. However, the clusters themselves are sketched by coloring the weights of the matrix $\boldsymbol{W}$, where darker colors correspond to larger edge-weights.

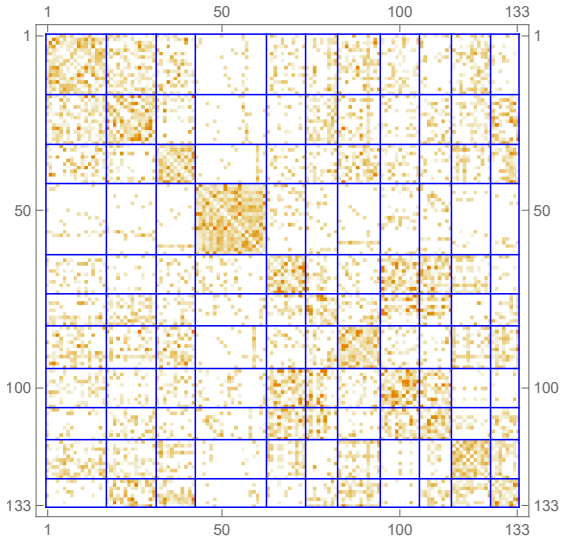


Fig. 4. 11 clusters in the appropriate labeling of the nodes, obtained by the k-means algorithm applied to the in-vectors corresponding to the leading $\boldsymbol{K}$-eigenvectors

In high precision computations of larger molecules, one has extremely large matrices, say, in multi-billion range. However, these matrices are sparse, each row having at most poly-logarithmic number of non-zero elements. As a result, corresponding weighted networks are also sparse with most weights equal to zero. Clustering of such networks could be a valuable tool in understanding molecular structures and aiding computations which nowadays require huge computational resources. Perhaps one could learn clusters from smaller scale samples of the network and then just "blow-up" clusters as redundant structures and use them for simplified large scale computations.

6 Conclusions and Further Directions

We considered the non-backtracking matrix based spectral clustering for sparse graphs that usually come from a percolated stochastic block model. We introduced estimates for the leading eigenvalues of the non-backtracking matrix and companion matrices of it. We recommended the usage of the k-means algorithm for the node representatives. Further, we used the corresponding eigenvectors for spectral clustering. Our results are supported with real life data processing.

Proposition 2 postulates that $\bar{\boldsymbol{A}}$ is a matrix of low rank. This is a strong condition, but with the help of a theorem of [3], under some assumptions for its structural eigenvalues and eigenvectors, we are able to give a construction for a low rank approximation of $\bar{\boldsymbol{A}}$. Moreover, the approximating matrix has a block structure, and therefore, it also has stepwise constant eigenvectors for which clustering techniques of Section 4 are applicable.

Assume that the matrix $\bar{\boldsymbol{A}} = \boldsymbol{P} \circ \mathbb{E}\mathbb{W}$ has $k = n^{o(1)}$ structural eigenvalues that are sufficiently delocalized and whose absolute value is larger than $\sqrt{\rho}$, where ρ is the spectral radius of the matrix $\boldsymbol{P} \circ \mathbb{E}(\mathbb{W} \circ \mathbb{W})$: these are $\nu_1 \geq \cdots \geq \nu_k$ with corresponding orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$. Also assume that the other eigenvalues are smaller than $\sqrt{\rho}$ in absolute value; further, the sum of the inner variances of the node representatives based on $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is $O(\frac{1}{\rho})$. This means that the squared distance between the eigensubspace spanned by $\mathbf{u}_1, \ldots, \mathbf{u}_k$ and the subspace F spanned by step-vectors (over the k-partition of the nodes given by the k-means algorithm applied to the node representatives $(\mathbf{u}_1, \ldots, \mathbf{u}_k)_i$, $i = 1, \ldots, n$) is $O(\frac{1}{\rho})$, but usually, it is much smaller.

Then by the construction of Theorem 3.1.17 [3], there are (unit-norm) step-vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ such that

$$\bar{\boldsymbol{A}} = \sum_{j=1}^{k} \nu_j \mathbf{v}_j \mathbf{v}_j^* + \boldsymbol{E} = \boldsymbol{M} + \boldsymbol{E},$$

where $\boldsymbol{M}$ is a matrix of rank at most k and it has a block structure; further, $\|\boldsymbol{E}\| = O(\sqrt{\rho})$. With them, the conditions of Proposition 2 still hold.

In classical spectral clustering, Laplacian, normalized Laplacian, and modularity matrices are also used in the degree-corrected cases. Analogously, the authors of [9] recommend the usage of the non-backtracking Laplacian, which is defined by the transition probabilities of the non-backtracking random walk. As a future direction, we plan to extend it to the non-backtracking matrices of edge-weighted graphs, and define non-backtracking modularity matrices too. We also plan to consider directed graphs.

Acknowledgment

This research was supported by the NKFIH project Dynamical systems and fractals, no.142169.

We are indebted to László Lovász for valuable conversations on this topic. We also thank Tamás Móri and Katalin Friedl for their useful remarks; further, Werner Dobrautz for helping to use PySCF 2.3 and for creating the test data. The research of the second author was supported by the BF_MatterQ project 342373.

Data Availability Statement

We generated the matrix using the open-source code PySCF 2.4 (the Python-based Simulations of Chemistry Framework), available at https://pyscf.org/https://pyscf.org/. The data can be provided by the second author.

References

1. Chung, F. (1997), *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics **92**. American Mathematical Society, Providence RI.
2. Newman, M. E. J. (2010), *Networks, An Introduction*. Oxford University Press.
3. Bolla, M. (2013), Spectral Clustering and Biclusterig, Wiley.
4. Bolla, M. (2020), Generalized quasirandom properties of expanding graph sequences, *Nonlinearity* **33** (4), 1405–1424.
5. Bordenave, C., Lelarge, M. and Massoulié, L. (2015), Non-backtracking spectrum of random graphs: Community detection and non-regular Ramanujan graphs. In: *Proc. 56th Annual Symposium on Foundations of Computer Science, FOCS*, pp. 1347–1357. More detailed version: arXiv:1501.06087.
6. Stephan, L. and Massoulié (2022), Non-backtracking spectra of inhomogeneous random graphs, *Mathematical Statistics and Learning* **5**(3), 201–271.
7. Capitaine, M., Donati–Martin, C. and Féral, D. (2009), The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations, *The Annals of Probability* **37** (1), 1–47.
8. Coste, S. and Zhu,Y. (2021), Eigenvalues of the non-backtracking operator detached from the bulk, *Random Matrices: Theory and Applications* **10**(03):2150028.
9. Mulas, R., Zhang, D., and Zucal, G. (2024), There is no going back: Properties of the non-backtracking Laplacian, *Linear Algebra and its Applications* **680**, 341–370.
10. Torres, L. (2022), Geometric multiplicity of unitary non-backrtacking eigenvalues. Preprint, arXiv:2205.02004v1 [math.CO].
11. Torres, L., Chan, K. S., Tong, H., and Eliassi-Rad, T. (2021), Nonbacktracking eigenvalues under node removal: X-centrality and targeted immunisation, *SIAM Journal on Mathematics of Data Science* **3**(2), 656–675.
12. Achlioptas, D. and McSherry, F. (2007), Fast computation of low-rank matrix approximations, *J. ACM* **54** (2), Article 9.
13. Mniszewski, S. M. et al. (2021), Reduction of the molecular hamiltonian matrix using quantum community detection (2021), Scientific Reports, *Nature* 11:4099.
14. Watanabe, Y. and Fukumizu, K. (2010), Graph Zeta Function in the Bethe Free Energy and Loopy Belief Propagation. In Y. Bengio, D. Schuurmans, J.D. Lafferty, C.K.I. Williams, A. Culotta eds, *Advances in Neural Information Processing Systems 22*, pp. 2017–2025, Curran Associates, Inc.
15. Bolla, M. and Zhou, D. (2023), Percolated stochastic block model via EM algorithm and belief propagation with non-backtracking spectra (2023). Preprint, arXiv:2307.16502v5 [math.CO].
16. Benaych–Georges, F., Bordenave, C. and Knowles, A. (2020), Spectral radii of sparse random matrices (2020), *Ann. Inst. Henri Poincaré Probab. Stat.* **56**(3), 2141–2161.
17. Glover, C. and Kempton, M. (2021), Some spectral properties of the non-backtracking matrix of a graph, *Linear Algebra and its Applications* **618**, 37–57.
18. Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L. and Zhang, P. (2013), Spectral redemption in clustering sparse networks, *Proc. Nat. Acad. Sci. USA* **110**, 20935–20940.
19. Bass, H. (1992), The Ihara–Selberg Zeta Function of a Tree Lattice, *International Journal of Mathematics* **03(06)**, 717–797.
20. Decelle, A., Krzakala, F., Moore, C. and Zdeborová, L. (2011), Inference and phase transitions in the detection of modules in sparse networks, *Phys. Rev. Lett.* **107**, 065701.

21. Abbe, E. and Sandon, C. (2015), Community detection in general stochastic block models: fundamental limits and efficient algorithms for recovery, 2015 IEEE 56th Annual Symposium on FOCS.
22. Abbe, E. (2018), Community detection in general stochastic block models: recent developments, *Journal of Machine Learning Research* **18** (177), 1–86.
23. Moore, C. (2017), The computer science and physics of community detection: Landscapes, phase transitions, and hardness, *Bull. EATCS* **121**, Computer Science and Physics.
24. Newman, M. E. J. (2023), Message passing methods on complex networks, *Proc. R. Soc. London A* **479**, 20220774.
25. Erdős, P. and Rényi, A. (1960), On the evolution of random graphs, *Publication of the Mathematical Institute of HAS* **5**, 17–61.
26. Toloui, B. and Love, P. (2013), Quantum algorithms for quantum chemistry based on the sparsity of the CI-matrix. Preprint, arXiv:1312.2579v2 [quant-ph].
27. Árendás, P., Furtenbacher, T. and Császár A. G. (2016), On spectra of spectra, *J. Math. Chem.* **54**, 806–822.
28. Wang, H., Huang, Y. and Chen, A. (2015), Eigenvalue problem of Hamiltonian operator matrices, *Journal of Inequalities and Applications*, 115.