Check for updates

The MIT Press

METHODS

Guided graph spectral embedding: Application to the *C. elegans* connectome

Miljan Petrovic[1,2], Thomas A. W. Bolton[1,2], Maria Giulia Preti[1,2], Raphaël Liégeois[1,2], and Dimitri Van De Ville[1,2]

[1]Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Campus Biotech, Geneva, Switzerland
[2]Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland

ABSTRACT

Graph spectral analysis can yield meaningful embeddings of graphs by providing insight into distributed features not directly accessible in nodal domain. Recent efforts in graph signal processing have proposed new decompositions—for example, based on wavelets and Slepians—that can be applied to filter signals defined on the graph. In this work, we take inspiration from these constructions to define a new guided spectral embedding that combines maximizing energy concentration with minimizing modified embedded distance for a given importance weighting of the nodes. We show that these optimization goals are intrinsically opposite, leading to a well-defined and stable spectral decomposition. The importance weighting allows us to put the focus on particular nodes and tune the trade-off between global and local effects. Following the derivation of our new optimization criterion, we exemplify the methodology on the *C. elegans* structural connectome. The results of our analyses confirm known observations on the nematode's neural network in terms of functionality and importance of cells. Compared with Laplacian embedding, the guided approach, focused on a certain class of cells (sensory neurons, interneurons, or motoneurons), provides more biological insights, such as the distinction between somatic positions of cells, and their involvement in low- or high-order processing functions.

INTRODUCTION

Many aspects of network science relate to graph partitioning—the grouping of nodes in subgraphs—and graph embedding—their representation in a low-dimensional space that accounts for graph topology (Von Luxburg, 2007). Spectral graph theory motivates analytical methods based on the eigenvectors of fundamental graph operators, such as the adjacency and the Laplacian operators (Chung, 1997). For instance, the well-known graph cut problem can be convexly relaxed and solved by thresholding of the Laplacian eigenvector with the smallest nonzero eigenvalue, known as the Fiedler vector (Fiedler, 1989). More recently, new approaches in graph signal processing have taken advantage of the Laplacian eigenvectors to define the graph Fourier transform, which can then be used to process (i.e., filter) graph signals in the spectral domain (Ortega, Frossard, Kovačević, Moura, & Vandergheynst, 2018; Shuman, Narang, Frossard, Ortega, & Vandergheynst, 2013); the spectral graph wavelet transform by Hammond, Vandergheynst, & Gribonval (2011) is one such example.

The Laplacian eigenvectors also provide a meaningful embedding by mapping nodes onto a line, or higher dimensional representation, that minimizes distances between connected nodes

Graph:
Mathematical model of networks including vertices (nodes) and edges connecting them.

Eigenvector:
A vector whose multiplication with the given matrix produces a scaled version of itself.

Laplacian:
Matrix describing a graph and containing information on smoothness of signals defined on that graph.

Eigenvalue:
The value of the scaling of an eigenvector due to its multiplication with a matrix.

Embedding:
Procedure of reducing dimensionality of the data; describing a dataset (graph) with less memory requirements while preserving important information.

Slepians:
Signals with maximally localized structure in both time/vertex and frequency domain.

(Belkin & Niyogi, 2003). Other well-known embedding techniques use different metrics for distance in order to assess local graph properties, ranging from simple Euclidean distance in locally linear embedding (Roweis, 2000), to shortest path in Isomap (Tenenbaum, 2000), transition probability (Shen & Meyer, 2008), or conditional probability of an edge in *t*-distributed stochastic neighbor embedding (van der Maaten & Hinton, 2008). A time-dependent dynamical similarity measure has also been introduced (Schaub, Delvenne, Lambiotte, & Barahona, 2018). In addition, efforts have been made to employ global properties of the graph, such as in Sammon mapping (Sammon, 1969), where a cost function including all pairwise distances is optimized. In this manner, embedding is performed while taking in consideration both local (neighborhood) and global (distant nodes) properties of the graph. However, these techniques are not aware of the network at the mesoscale: One cannot guide the embedding by giving a certain subgraph more importance while still preserving local features and global topology characteristics.

In essence, the most powerful feature of graph spectral embedding is to effectively summarize local structure across the graph into low-dimensional global patterns. This is achieved, for instance, with the recently introduced concept of graph Slepians; that is, graph signals that are bandlimited and take into account a subset of selected nodes. Specifically, two types of Slepian designs that respectively optimize for energy concentration and modified embedded distance have been introduced (Van De Ville, 2016; Van De Ville, Demesmaeker, & Preti, 2017b).

In this work, we further build on this framework by providing a simple way to guide analyses with additional flexibility. *Guidance* includes the selection of a given subgraph or group of nodes to study, and the ability to specify the intensity of the focus set on these selected nodes. With respect to graph Slepians, we hereby provide several extensions. First, we allow the selection process to be weighted, so that the importance of a node can be gradually changed. Second, we propose a new criterion that meaningfully combines the two existing ones; that is, we want to maximize energy concentration and minimize modified embedded distance at the same time. Third, as we detail below, these two criteria are counteracting, and hence, we obtain stable solutions even at full bandwidth, where the original Slepian designs degenerate numerically. Fourth, we show how this criterion can be rewritten as an eigenvalue problem of an easy modification of the adjacency matrix, which can be interpreted as reweighting paths in the graph, and thus significantly simplifies the whole Slepian concept. The solution of the eigendecomposition then defines the guided spectral domain, spanned by its eigenvectors. We illustrate the proposed approach with a proof-of-concept on the *Caenorhabditis elegans* (*C. elegans*) connectome. Through spectral embedding-based visualization, we observe the effects of focusing on a specific cellular population made of sensory neurons, interneurons, or motoneurons, and we reveal trajectories of these neurons as a function of focus strength.

METHODS

Essential Graph Concepts

We consider an undirected graph with N nodes, labeled $1, 2, \ldots, N$. The edge weights are contained in the symmetric weighted adjacency matrix $\tilde{\mathbf{A}}$ with nonnegative real-valued elements $\tilde{a}_{i,j}$, $i, j = 1, \ldots, N$. We also assume that the graph contains no self-loops; that is, all diagonal elements $\tilde{a}_{i,i}$ are zero. The degree matrix $\mathbf{D}$ is the diagonal matrix with elements $d_{i,i} = \sum_{j=1}^{N} \tilde{a}_{i,j}$. The graph Laplacian is defined as $\tilde{\mathbf{L}} = \mathbf{D} - \tilde{\mathbf{A}}$ and can be interpreted as a second-order derivative operator on the graph. Here, we use the symmetrically normalized variants of the adjacency $\tilde{\mathbf{A}}$ and graph Laplacian $\tilde{\mathbf{L}}$ defined as $\mathbf{A} = \mathbf{D}^{-1/2} \tilde{\mathbf{A}} \mathbf{D}^{-1/2}$ and $\mathbf{L} = \mathbf{I} - \mathbf{A}$.

This normalization is often used in applications to emphasize the changes in topology and not in nodal degree (De Lange, De Reus, & Van den Heuvel, 2014).

Let us define a graph signal as a vector of length N that associates a value with each node (Shuman et al., 2013). One way to recognize the importance of the Laplacian and its eigen-decomposition is to consider the smoothness of a graph signal $\mathbf{x}$ as

$$\mathbf{x}^\top \mathbf{L} \mathbf{x} = \sum_{i,j=1}^{N} a_{i,j}(x_i - x_j)^2, \tag{1}$$

which sums squared differences between signal values on nodes that are connected, proportionally to their link strength $a_{i,j}$. The eigenvectors of $\mathbf{L}$ minimize this distance that is reflected by the eigenvalues, sorted by convention increasingly as $\lambda_1 = 0 \leq \lambda_2 \leq \ldots \leq \lambda_N$. Therefore, considering the eigenvectors associated with the smallest nonzero eigenvalues provides the Laplacian embedding of the nodes that minimizes distance in a lower dimensional space (Belkin & Niyogi, 2003). The eigenvector with the smallest nonzero eigenvalue is also known as the Fiedler vector (Fiedler, 1989), which relates to the solution of the convex relaxation of the graph cut problem (Von Luxburg, 2007).

Therefore, the eigendecomposition $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ of the graph Laplacian is the cornerstone of spectral methods for graphs, as the eigenvectors $\{\mathbf{u}_k\}$, $k = 1, \ldots, N$ (columns of $\mathbf{U}$) play the role of graph Fourier components, and the associated eigenvalues $\{\lambda_k\}$, $k = 1, \ldots, N$, of frequencies (Chung, 1997). The graph Fourier transform (GFT) then provides the link between a graph signal $\mathbf{x}$ and its spectral coefficients given by vector $\hat{\mathbf{x}}$:

$$\mathbf{x} = \mathbf{U}\hat{\mathbf{x}}, \text{ and } \hat{\mathbf{x}} = \mathbf{U}^\top \mathbf{x}.$$

Graph Slepians

In earlier work, the combination of the concepts of selectivity and bandwidth for graph signals has been used to define "graph Slepians" (Tsitsvero, Barbarossa, & Di Lorenzo, 2016; Van De Ville, 2016; Van De Ville et al., 2017b); that is, bandlimited graph signals with maximal energy concentration in the subset of nodes $\mathcal{S}$—a generalization of prolate spheroidal wave functions that were proposed 50 years ago on regular domains to find a trade-off between temporal and spectral energy concentrations (Slepian, 1978; Slepian & Pollak, 1961). The presence or absence of a node in $\mathcal{S}$ is encoded by the diagonal elements of the selection matrix $\mathbf{S}$; that is, we have $S_{i,i} = \delta_{i \in \mathcal{S}}$, $i = 1, \ldots, N$, where δ is the Kronecker delta. The Slepian design then boils down to finding the linear combination of Laplacian eigenvectors, encoded by spectral coefficients $\hat{\mathbf{g}}$, within the bandlimit W with maximal energy in $\mathcal{S}$, reverting to the Rayleigh quotient

$$\mu = \frac{\hat{\mathbf{g}}^\top \mathbf{W}^\top \mathbf{U}^\top \mathbf{S} \mathbf{U} \mathbf{W} \hat{\mathbf{g}}}{\hat{\mathbf{g}}^\top \hat{\mathbf{g}}}, \tag{2}$$

where $\mathbf{W}$ is a spectral selection matrix that has W ones on its diagonal followed by $N - W$ zeros. This problem can be solved by the eigendecomposition of the concentration matrix $\mathbf{C} = \mathbf{W}^\top \mathbf{U}^\top \mathbf{S} \mathbf{U} \mathbf{W}$ as $\mathbf{C}\hat{\mathbf{g}}_k = \mu_k \hat{\mathbf{g}}_k$, $k = 1, \ldots, W$. The graph Slepians $\mathbf{g}_k = \mathbf{U}\hat{\mathbf{g}}_k$, $k = 1, \ldots, W$, are orthonormal over the entire graph as well as orthogonal over the subset $\mathcal{S}$; that is, we have $\mathbf{g}_k^\top \mathbf{g}_l = \delta_{k-l}$ as well as $\mathbf{g}_k^\top \mathbf{S} \mathbf{g}_l = \mu_k \delta_{k-l}$.

For the purpose of this work, we introduce the set of bandlimited graph signals

$$\mathcal{B}_W = \{\mathbf{x} | \hat{\mathbf{x}} = \mathbf{W}\hat{\mathbf{x}}\},$$

such that we can then rewrite the Slepian criterion of Equation (2) directly in the vertex domain as

$$\mu = \frac{\mathbf{g}^\top \mathbf{S} \mathbf{g}}{\mathbf{g}^\top \mathbf{g}} \quad \text{s.t. } \mathbf{g} \in \mathcal{B}_W. \tag{3}$$

An alternative Slepian design was also proposed in Van De Ville et al. (2017b)—see also Huang et al. (2018), modifying the Laplacian embedded distance of Equation (1) as follows:

$$\xi = \frac{\mathbf{g}^\top \mathbf{L}^{1/2} \mathbf{S} \mathbf{L}^{1/2} \mathbf{g}}{\mathbf{g}^\top \mathbf{g}} \quad \text{s.t. } \mathbf{g} \in \mathcal{B}_W. \tag{4}$$

The Laplacian embedded distance $\mathbf{x}^\top \mathbf{L} \mathbf{x}$ is a measure of smoothness of the vector $\mathbf{x}$ over the graph, which is why eigenvectors of $\mathbf{L}$ with increasing eigenvalues are ordered according to smoothness. Imposing the modification with the selection matrix $\mathbf{S}$ focuses the smoothness on a certain subgraph, notwithstanding how the signal behaves outside it. Equation (4) can also be seen as a generalization of Laplacian embedding, since $\mathbf{L}^{1/2} \mathbf{S} \mathbf{L}^{1/2}$ reverts to $\mathbf{L}$ for the special case of $\mathbf{S} = \mathbf{I}$.

It is important to realize that the eigenvalues $\{\mu_k\}$ of the original design reflect the energy concentration in the subset $\mathcal{S}$, while the eigenvalues $\{\xi_k\}$ of the alternative design correspond to a modified embedded distance that can be interpreted as a "frequency value" localized in $\mathcal{S}$, in analogy to the global GFT case. Consequently, "interesting" eigenvectors correspond to those with high μ_k, concentrated in the subset $\mathcal{S}$, or low ξ_k, showing the main localized low-frequency trends, respectively. However, the eigendecompositions, taken individually, do not necessarily lead to eigenvectors that combine both virtues.

Guiding Spectral Embedding Using a New Criterion

We hereby propose to further generalize the Slepian design in a number of ways. First, we relax the selection matrix $\mathbf{S}$ to a cooperation matrix $\mathbf{M}$ with diagonal elements that can take any nonnegative real values $m_l \geq 0, l = 1, \ldots, N$. This allows to gradually change the impact of a node on the analysis, between an enhanced ($m_l > 1$), an unmodified ($m_l = 1$), and a reduced ($m_l < 1$) importance with respect to the selection matrix case. Second, we combine the criteria of both already existing Slepian designs by subtracting the modified embedded distance from the energy concentration:

$$\zeta = \mu - \xi = \frac{\mathbf{g}^\top \mathbf{M} \mathbf{g} - \mathbf{g}^\top \mathbf{L}^{1/2} \mathbf{M} \mathbf{L}^{1/2} \mathbf{g}}{\mathbf{g}^\top \mathbf{g}} \quad \text{s.t. } \mathbf{g} \in \mathcal{B}_W. \tag{5}$$

Third, we remove the bandlimit constraint and allow $\mathbf{g}$ to be any graph signal, which is an operational choice because of the joint optimization of both criteria, as will be illustrated and discussed later.

Using the Taylor series approximation of the square root function, we derive $\mathbf{L}^{1/2}$ in terms of the adjacency matrix $\mathbf{A}$:

$$\mathbf{L}^{1/2} = (\mathbf{I} - \mathbf{A})^{1/2} = \mathbf{I} - \frac{1}{2}\mathbf{A} - \frac{1}{8}\mathbf{A}^2 - \frac{1}{16}\mathbf{A}^3 - \ldots \tag{6}$$

$$= \mathbf{I} - \sum_{k=1}^{\infty} c_k \mathbf{A}^k, \tag{7}$$

with $c_k = \frac{(2k)!}{2^{2k}(k!)^2(2k-1)}$. Details on the series expansion are discussed in the Taylor Series section. We can then further rewrite the internal part of the Criterion 5 as

$$\mathbf{M} - (\mathbf{I} - \mathbf{A})^{1/2}\mathbf{M}(\mathbf{I} - \mathbf{A})^{1/2} = \sum_{k=1}^{\infty} c_k \left(\mathbf{M}\mathbf{A}^k + \mathbf{A}^k\mathbf{M}\right)$$
$$- \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} \left(c_{k_1}c_{k_2}\right)\mathbf{A}^{k_1}\mathbf{M}\mathbf{A}^{k_2}. \tag{8}$$

By convention, the associated eigenvalues are sorted in decreasing order. Based on the fact that eigenvalues of the symmetric normalized Laplacian are greater or equal to 0 and lower or equal to 2, one can derive $m_{\max} \geq \zeta_1 \geq \zeta_2 \geq \ldots \geq -2m_{\max}$, where $m_{\max}$ is the highest cooperation value appearing in $\mathbf{M}$, using bounds from Corollary 2.4 in Lu and Pearce (2000).

In what follows, we will be considering the linear and quadratic approximations of the new criterion's eigenvalues:

$$\zeta_{\text{lin}} = \frac{\mathbf{g}^\top \left(\frac{\mathbf{M}\mathbf{A}+\mathbf{A}\mathbf{M}}{2}\right)\mathbf{g}}{\mathbf{g}^\top\mathbf{g}}, \tag{9}$$

$$\zeta_{\text{quad}} = \frac{\mathbf{g}^\top \left(\frac{\mathbf{M}\mathbf{A}+\mathbf{A}\mathbf{M}}{2} + \frac{\mathbf{M}\mathbf{A}^2+\mathbf{A}^2\mathbf{M}}{8} - \frac{\mathbf{A}\mathbf{M}\mathbf{A}}{4}\right)\mathbf{g}}{\mathbf{g}^\top\mathbf{g}}. \tag{10}$$

Interestingly, the combination of both existing Slepian criteria leads to the emergence of the adjacency matrix $\mathbf{A}$ as the key player in our new formalism. In fact, when the cooperation matrix is the identity matrix, the criterion reverts to the eigendecomposition of $\mathbf{A}$ itself.

Let us now interpret the impact of the cooperation weights: Obviously, an element $a_{i,j}$ of the adjacency matrix contains the weight of a direct path from i to j. The linear approximation ζ_{lin} reweights such a direct path with the average $(m_i + m_j)/2$ of the cooperation weights that are attributed to nodes i and j, as illustrated in Figure 1A (left half). Notice that paths where only one node has a cooperation weight equal to 0 are still possible, as the other cooperation weight is then simply divided by 2.

As for the quadratic approximation, it takes into account length-2 paths between nodes i and j. For instance, the sum of all length-2 paths between i and j can be read out from the squared adjacency matrix:

$$[\mathbf{A}^2]_{i,j} = \sum_{l=1}^{N} a_{i,l}a_{l,j} = \left\langle \mathbf{a}_{i,\cdot}, \mathbf{a}_{\cdot,j}\right\rangle,$$

where the inner product reveals the kernel interpretation of the length-2 walk matrix. Therefore, as illustrated in Figure 1A (right half), the term

$$[\mathbf{M}\mathbf{A}^2 + \mathbf{A}^2\mathbf{M}]_{i,j} = (m_i + m_j)\sum_{l=1}^{N} a_{i,l}a_{l,j}$$

reweights all length-2 paths by the summed cooperation weight between the start and end nodes, while subtracting the term

$$[\mathbf{A}\mathbf{M}\mathbf{A}]_{i,j} = \sum_{l=1}^{N} m_l a_{i,l}a_{l,j}$$

penalizes the path according to the cooperation weight of node l through which it passes.
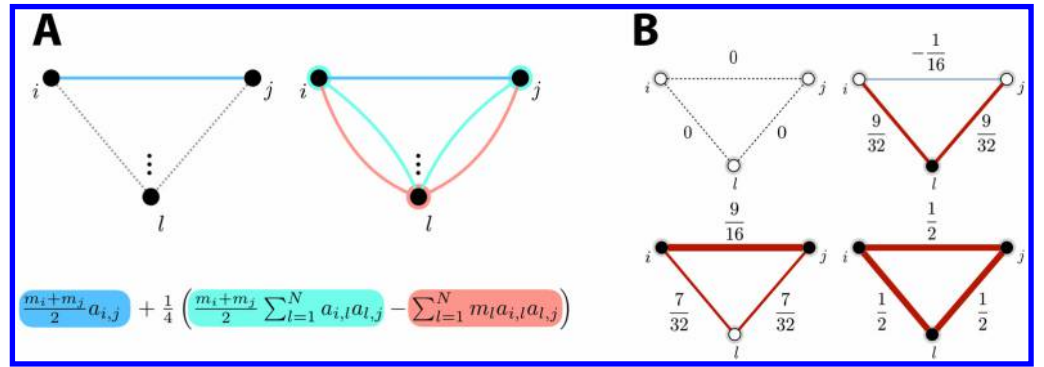
Figure 1. (A) In the case of two nodes i and j, the average of their cooperation weights yields the multiplying factor for $a_{i,j}$ (blue term). When a third node l is added, the difference between average cooperation weight between nodes i and j (light blue term), and the cooperation weight of node l (salmon term), multiplies the length-2 path and then also contributes to the output entry. (B) In an example three-node network, output entries for different examples where cooperation weights are either set to 0 (white nodes) or to 1 (black nodes). Edge thickness is proportional to the output entry weight. Red strokes denote positive edge values, while blue strokes highlight negative edge values. All nonzero entries of the normalized adjacency matrix of the example network equal 1/2.

Analogously, the term $\mathbf{A}^k$ in the criterion introduces modifications of k-length paths in the graph. However, for $k > N$, reweighting reduces to modifications of lower length paths. The Cayley-Hamilton theorem implies that for every matrix $\mathbf{A}$ of size $N \times N$, the matrix $\mathbf{A}^N$ can be written as a linear combination of matrices $\mathbf{A}^k$ for $k = 0, 1, \dots N - 1$. By induction, it holds that $\mathbf{A}^k$ for every $k > N$ can also be written as a linear combination of the same set of N matrices. Hence, modifications of paths longer than $N - 1$ can be seen as a linear combination of additional modifications of paths of length 0 to $N - 1$.

MATHEMATICAL CONSIDERATIONS

This section provides mathematical foundations supporting the methods and the results presented in this work. We start by discussing the link between the selection matrix and the eigenspectrum associated with the energy concentration criterion, and the relationship with the modified embedded distance criterion, using full bandwidth. Then, we provide a formal justification of the Taylor series approximation of the square root matrix function used in Equation 6, and discuss the error associated with this approximation.

Eigenspectrum Associated With the Energy Concentration Criterion

For full bandwidth, the concentration matrix is defined as $\mathbf{C} = \mathbf{U}^\top \mathbf{S} \mathbf{U}$, where $\mathbf{U}$ is the matrix whose columns are eigenvectors of the graph Laplacian, and $\mathbf{S}$ is a diagonal selection matrix. Hence, the eigendecomposition of $\mathbf{C}$ is trivial: Its eigenvectors are the rows of $\mathbf{U}$, and the eigenvalues of $\mathbf{C}$ correspond to the diagonal entries of $\mathbf{S}$, as can be seen from Figure 3A for $W = 279$.

Eigenspectrum Associated With the Modified Embedded Distance Criterion

We show that for full bandwidth, the number of zero eigenvalues of the modified embedded distance matrix, denoted z_λ, is lower bounded by the number of zeros on the diagonal of the

selection matrix, denoted z_S. To see this, consider the following decomposition of the modified embedded distance matrix $\mathbf{C}_{emb}$:

$$\mathbf{C}_{emb} = \mathbf{L}^{1/2}\mathbf{S}\mathbf{L}^{1/2} = \sum_{k=1}^{N-z_S} s_{i_k} l_{i_k} l_{i_k}^{\top},$$

where i_k is the index of the kth nonzero entry of the selection matrix $\mathbf{S}$, and l_{i_k} denotes the $i_k{}^{\text{th}}$ column vector of the matrix $\mathbf{L}^{1/2}$. From this expression, it can be seen that the rank of $\mathbf{C}_{emb}$ is at most $N - z_S$ and hence, $z_\lambda \geq z_S$. Equality holds when the set of vectors $\{l_{i_k}\}$ corresponding to the nonzero entries of $\mathbf{S}$ are linearly independent. This is the case for connected graphs, as any subset (with cardinality strictly less than N) of the columns of $\mathbf{L}^{1/2}$ is linearly independent. This relationship is observed in Figure 3B for $W = 279$.

Taylor Series of Matrix-Valued Functions

The Taylor expansion of $\mathbf{L}^{1/2}$ proposed in Equation 6 is derived using the scalar Taylor series of $f(x) = \sqrt{x}$ evaluated around the point $a = 1$:

$$\sqrt{x} = 1 + \sum_{k=1}^{\infty} t_k (x-1)^k,$$

where $t_k = \frac{(-1)^{k-1}(2k)!}{2^{2k}(k!)^2(2k-1)}$ and $x \in \mathbb{R}, x > 0$. The square root matrix of $\mathbf{L}$ then writes

$$\mathbf{L}^{1/2} = \mathbf{U}_L \mathbf{\Lambda}_L^{1/2} \mathbf{U}_L^{\top}$$

$$= \mathbf{U}_L \begin{bmatrix} 1 + \sum_{k=1}^{\infty} t_k(\lambda_1 - 1)^k & & \\ & \ddots & \\ & & 1 + \sum_{k=1}^{\infty} t_k(\lambda_N - 1)^k \end{bmatrix} \mathbf{U}_L^{\top}$$

$$= \mathbf{U}_L \left(\mathbf{I} + \sum_{k=1}^{\infty} t_k(\mathbf{\Lambda}_L - \mathbf{I})^k \right) \mathbf{U}_L^{\top}.$$

Since the Laplacian and adjacency matrices are normalized, their eigenvalues verify $\mathbf{\Lambda}_L = \mathbf{I} - \mathbf{\Lambda}_A$ and their eigenvectors are equal ($\mathbf{U}_L = \mathbf{U}_A$) when ordered following increasing and decreasing eigenvalues, respectively. The previous equation finally reduces to

$$\mathbf{L}^{1/2} = \mathbf{I} + \mathbf{U}_A \left(\sum_{k=1}^{\infty} t_k(-\mathbf{\Lambda}_A)^k \right) \mathbf{U}_A^{\top}$$

$$= \mathbf{I} + \sum_{k=1}^{\infty} (-1)^k t_k \mathbf{U}_A \mathbf{\Lambda}_A^k \mathbf{U}_A^{\top}$$

$$= \mathbf{I} - \sum_{k=1}^{\infty} c_k \mathbf{A}^k,$$

where $c_k = \frac{(2k)!}{2^{2k}(k!)^2(2k-1)}$, which is the expression used in Equation 6.

Truncation of the Taylor series of a function $f(x)$ to a finite upper bound on $k \leq K$ leads to an approximation error that can be estimated by the Lagrange form of the remainder

$$R_K(x) = \frac{f^{(K+1)}(y)}{(K+1)!}(x-1)^{K+1},$$

where the $(K+1)^{th}$ derivative is evaluated at the point y found between x and 1. On the other hand, since the eigenvectors forming $\mathbf{U}_L$ are unit-norm vectors, the distance d_K between a finite sum approximation of $\mathbf{L}^{1/2}$ and the true square root of the matrix is bounded as

$$d_K = ||\mathbf{L}^{1/2} - (\mathbf{I} - \sum_{k=1}^{K} c_k \mathbf{A}^k)||_F \leq \sum_{i=1}^{N} |R_K(\lambda_i)|,$$

where $|| \cdot ||_F$ denotes the Frobenius norm. In the case of a first-order Taylor approximation ($K = 1$), we get

$$d_1 \leq \sum_{i=1}^{N} \frac{|f^{(2)}(y_i)|}{2!} (\lambda_i - 1)^2.$$

The eigenvalues λ_i range from 0 to 2, and all contribute to the total approximation error d_1, with eigenvalues further from 1 contributing more. Since the second-order derivative of the square root function increases as its argument approaches 0, the most contributing factors of the error derive from Taylor approximation terms with near-zero eigenvalues. Hence, graphs whose Laplacian spectrum exhibits higher eigengaps in the lower band tend to have lower approximation error.

Finally, the Frobenius distance $d_{K,M}$ between the true proposed criterion $\mathbf{M} - \mathbf{L}^{1/2}\mathbf{M}\mathbf{L}^{1/2}$ and its approximation using a Kth-order Taylor approximation of $L^{1/2}$ verifies

Frobenius distance:
Square root of the sum of pointwise squared differences between two matrices.

$$d_{K,M} \leq d_K ||\mathbf{M}||_F d_K,$$

where $||\mathbf{M}||_F$ corresponds to the Frobenius norm of the cooperation matrix. Hence, the upper bound on $d_{K,M}$ reduces as the nodes are given less importance; that is, when the cooperation values get closer to 0.

RESULTS

The *C. elegans* worm is an intensely studied model organism in biology. In particular, the wiring diagram of its 302 neurons has been carefully mapped during a long and effortful study (White, Southgate, Thomson, & Brenner, 1986). Here, we use the graph that summarizes data from 279 somatic neurons (unconnected and pharyngeal neurons were excluded from the full diagram of 302 neurons), and combined connectivity from chemical synapses and gap junctions (Chen, Hall, & Chklovskii, 2006). The binary adjacency matrix $\mathbf{A}_{bin}$ with edge weights 0 or 1 has been symmetrically normalized with the degree matrix $\mathbf{D}$ into $\mathbf{A} = \mathbf{D}^{-1/2}\mathbf{A}_{bin}\mathbf{D}^{-1/2}$, as described in the Essential Graph Concepts section. We retrieved the type of each neuron (sensory neuron, interneuron, or motoneuron) from the WormAtlas database (http://www.wormatlas.org/).

In their modeling work, Varshney, Chen, Paniagua, Hall, & Chklovskii (2011) studied network properties of the worm connectome using different approaches, including Laplacian embedding. In particular, the topological view generated by mapping nodes on the first two eigenvectors with smallest nonzero eigenvalues already reveals interesting network organization (see Figure 2). The horizontal dimension ($\mathbf{u}_2$) mainly distinguishes the motoneurons from the head (right green circles) and from the ventral cord (left green circles). The vertical dimension ($\mathbf{u}_3$) reflects information flow from sensory neurons and interneurons of the animal's head (top) to the nerve ring and ventral cord circuitries (bottom).

Eigenvalues of Different Criteria

To illustrate the eigenvalues obtained with the existing Slepian designs, as well as the newly proposed criterion, we considered the 128 motoneurons and "unselected" them by setting their
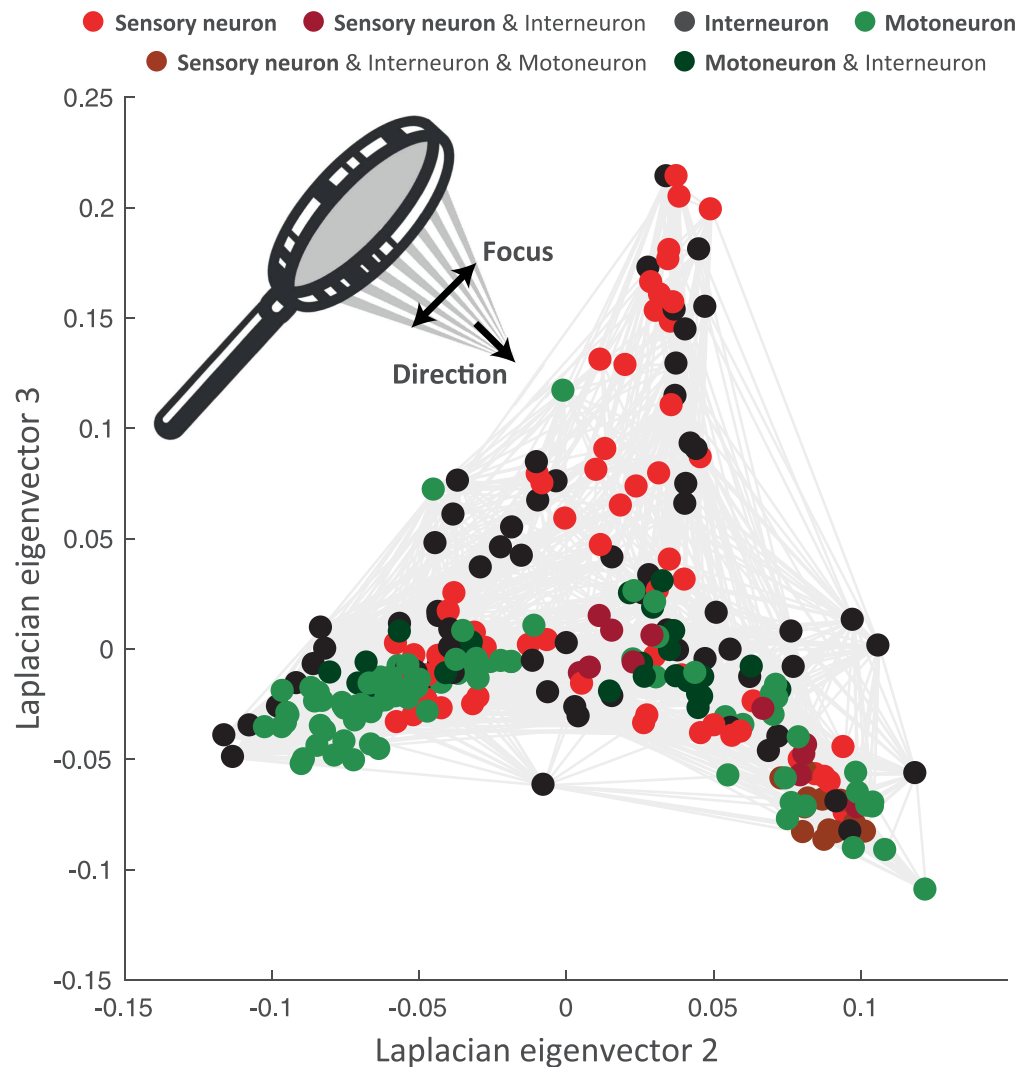
Figure 2. Spectral embedding of the *C. elegans* connectome according to the eigenvectors of the Laplacian matrix with second and third smallest eigenvalues. The purpose of this work is to introduce guided spectral analysis; that is, to indicate direction by selecting a subset of nodes, and to adjust the strength of the focus set on this subset. Each colored circle in the figure depicts one *C. elegans* neuron. Light gray strokes link the cells that are connected by gap junctions or chemical synapses. Labels and connectivity were retrieved from Varshney et al. (2011).

respective entries in **S** to 0. We applied the original, concentration-based Slepian design for different bandwidths $W = 100, 150, 200, 279$, the latter corresponding to full bandwidth. The eigenvalues μ_k, which reflect energy concentration in the 151 remaining neurons, are shown in Figure 3A. The characteristic behavior of classical Slepians is preserved for the graph variant; that is, eigenvalues cluster around 1 and 0 for well- and poorly concentrated eigenvectors, respectively, and the phase transition occurs more abruptly at higher bandwidth. For full bandwidth, perfect concentration becomes possible, and the problem degenerates in retrieving two linear subspaces of 151 and 128 dimensions spanned by eigenvectors with concentration 1 and 0, respectively (see the Mathematical Considerations section for a proof on the number of distinct eigenvalues). In practical terms, for high but not full bandwidth, the "interesting"
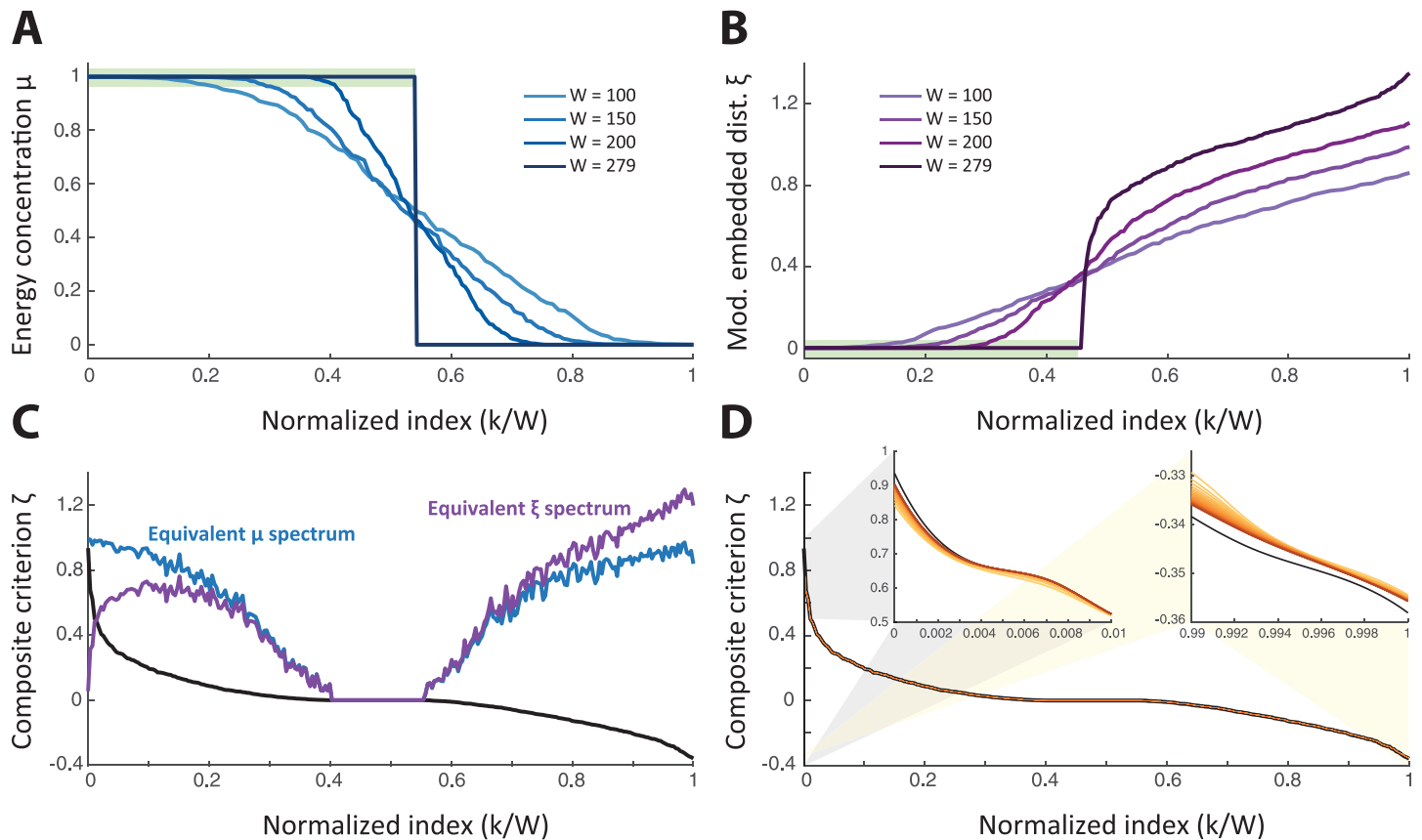
Figure 3. Plots of eigenvalues obtained using different Slepian criteria: (A) energy concentration μ, (B) modified embedded distance ξ, and (C) our new proposed criterion ζ. For the first two cases, in which the design depends on a bandwidth parameter, eigenvalue spectra are plotted for $W = 100$, 150, 200, and 279 with increasingly lighter blue or purple shades, respectively. In the full-bandwidth case, the shaded green areas highlight eigenvalues linked to optimal solutions of the respective criteria (see the Mathematical Considerations section for the associated mathematical derivations). In the third case, equivalent μ and ξ eigenspectra are plotted in blue and purple on top of the ζ one. The full ζ eigenspectrum is also compared with approximations obtained through Taylor series of increasing order (D), from linear to order 20, as depicted by increasingly darker brown curves. The two smaller plots are insets sampled at the start and at the end of the main plot, respectively.

eigenvectors with large concentration correspond to the part indicated by the green area on the plot, and become numerically indistinguishable. A few indicative examples of Slepian vectors across bandwidths are displayed in Supplementary Figure S1C (Supporting Information).

Next, we applied the modified Slepian design inspired by the Laplacian embedded distance. As shown in Figure 3B, the eigenvalues ξ_k reflect the modified embedded distance, which we now want to minimize. For increasing bandwidth (darker curves), its smallest values can be made lower; however, the subset of nodes with $S_{i,i}$ entries set to 0 is also described by eigenvectors with small eigenvalues. This becomes even clearer at full bandwidth, a case for which a subspace of 128 dimensions spanned by eigenvectors with a modified embedded distance of 0 is retrieved, as indicated by the green area in Figure 3B and explicitly demonstrated in the Mathematical Considerations section. Some examples of Slepians across bandwidths can be seen in Supplementary Figure S1D (Supporting Information).

The degeneracies of the Slepian designs at full bandwidth are instructive about the opposing effects of maximizing energy concentration and minimizing modified embedded distance; that

is, the subspaces indicated by the green areas in Figures 3A and 3B, which are optimal for the corresponding criteria, are actually different ones, representing signals on sensory and interneurons (151 nodes) on the one hand, and on motoneurons (128 nodes) on the other hand (compare Supplementary Figures S1C and S1D, first rows; Supporting Information). This leads us to the eigenvalues ζ_k of the proposed criterion, as shown in Figure 3C (black curve).

The maximum eigenvalue peaks close to 1, a case reflecting jointly high equivalent μ_k (blue curve) and low equivalent ξ_k (purple curve); that is, a high energy concentration at the same time as a low modified embedded distance (low localized graph frequency) within $\mathcal{S}$. The low amount of such solutions shows that it is difficult to conceal high energy concentration and small modified embedded distance.

As values of ζ_k decrease, we first observe a rise in modified embedded distance (eigenvectors remain reasonably concentrated within $\mathcal{S}$, but rapidly exhibit a larger localized graph frequency), and then a decrease of both μ_k and ξ_k, which indicates that eigenvectors become less concentrated within the subset of interest. Afterwards, we observe a regime in which both quantities are null at the same time; that is, a subspace spanned by eigenvectors that are fully concentrated outside $\mathcal{S}$. Notice that this set of eigenvectors is now "pushed away" from the meaningful low ξ_k ones, and lie in the middle of the spectrum. Finally, the sign of ζ_k switches, and the right-hand side of Figure 3C denotes eigenvectors of increasing concentration within $\mathcal{S}$ and localized graph frequency, the latter effect dominating over the former.

Interestingly, computing the eigenspectrum using a linear approximation of the criterion matrix (Figure 3D, light brown curve) leads to very similar results, which only slightly vary for the largest eigenvalues. When the approximation order is increased up to 20 (increasingly dark brown curves), this low error further diminishes, although a mild difference remains with the ground truth. Inspection of the Slepian vectors related to several locations of the eigenspectrum (Supplementary Figure S2, Supporting Information) confirmed that the only salient differences actually involved the first Slepian vector (largest eigenvalue one).

Topology Revealed by Guided Spectral Analysis

We now guide the spectral analysis to focus on the three different types of neurons. For instance, when focusing on the role of the sensory neurons, we gradually decrease the cooperation weights of interneurons and motoneurons from 1 to 0. For each setting, we then visualize the topology revealed by the guided analysis by projecting the nodes on the eigenvectors with the second and third largest eigenvalues. We build the trajectory of each node through this two-dimensional embedding, after applying the Procrustes transform (Schönemann, 1966) to compensate for any irrelevant global transformations. As a complementary visualization, note that we provide the start, intermediate, and end points of each trajectory as separate figures in Supplementary Figure S3 (Supporting Information). Finally, k-means clustering was performed on the nodes in focus at the end point embedding of trajectories, producing sets of clusters given in Supplementary Figure S4 and Supplementary Tables 1–3 (see Supporting Information for details). Example visualizations when resorting to different Slepian vectors are provided in Supplementary Figure S5 (Supporting Information).

In Figures 4A and 4B, the trajectories are depicted when focusing on the sensory neurons by attributing cooperation weights to the other types of neurons ranging from 1 to 0.5, and from 0.5 to 0, respectively. During the first half (Figure 4A), the network organization is only slightly altered with respect to the initial view of Figure 2; that is, the sensory neurons move slightly more to the periphery, while the interneurons and motoneurons move to the origin. In

Procrustes transform:
Geometric transform that changes only the size and the overall position of an object but not its shape.
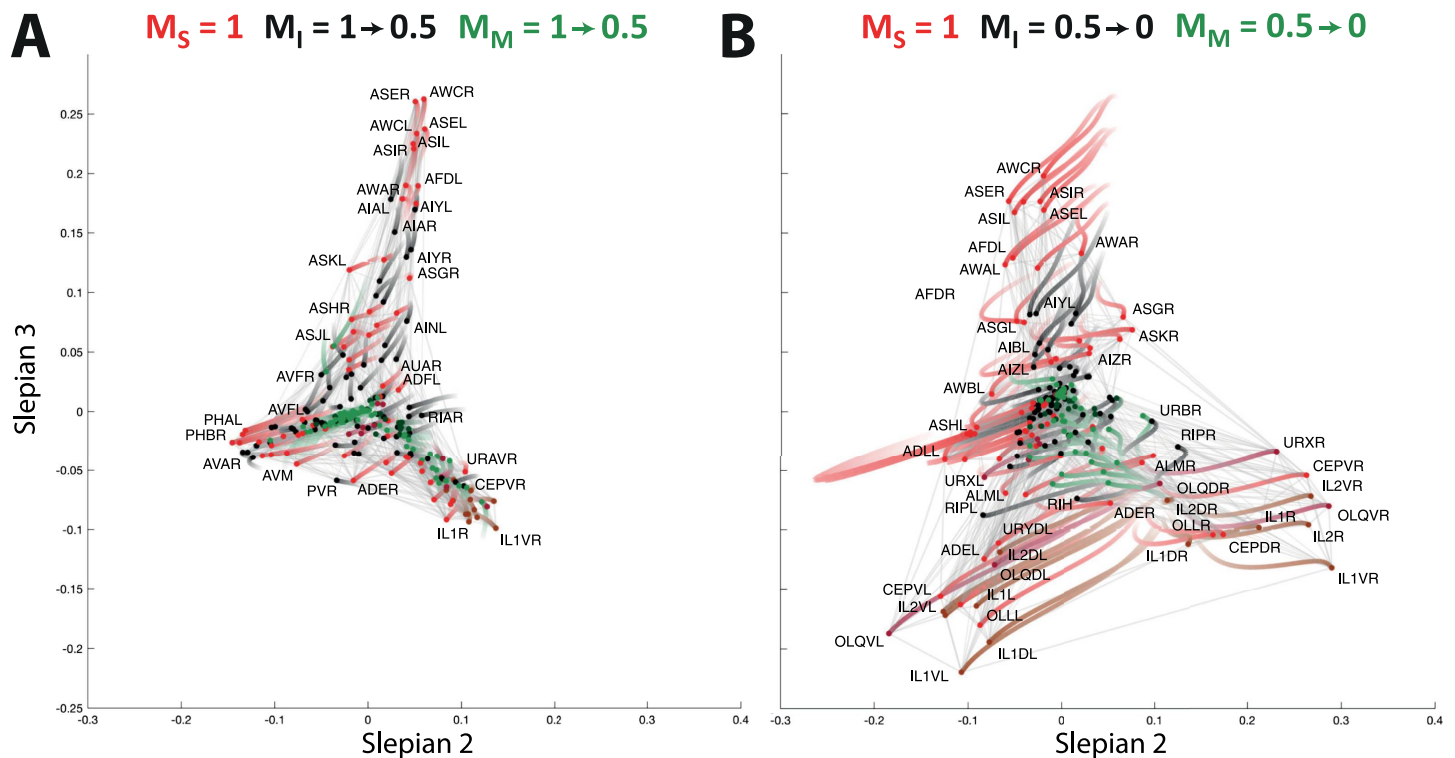
Figure 4. Focusing on the sensory neurons by reducing the cooperation weights of the interneurons and motoneurons (A) from 1 to 0.5, and (B) from 0.5 to 0. The trajectory of a neuron is represented by a color change from light to dark tones, and dots represent final positions. Note that the starting configuration in (A) is identical to the representation in Figure 2. Cells are labeled according to Varshney et al. (2011).

the second part of the trajectory (Figure 4B), a major split occurs in the bottom right branch of Figure 4A between the left and right versions of a whole series of neurons, while the bottom left branch neurons move back to the center of the coordinate frame. The cell types found in the top branch are amphid neurons, whereas the rest of the sensory neurons split into their left and right counterparts located in the left and right bottom branches. The clusters found by the k-means approach (see Supplementary Table 1, Supporting Information) include a group of five bilateral amphid neurons (AWA, AWC, ASE, ASI, and AFD; cluster C_3) and six other clusters, two of which span the bottom left and right subbranches (clusters C_5 and C_2).

As described in the Guiding Spectral Embedding section, since paths through nodes with cooperation weights set to 0 are still considered by the proposed criterion, the embedding focusing on a particular subtype of neurons can still include functionally distinct cells as clearly standing out in the visualization. For instance, in addition to the above clustering of sensory neurons in Figure 4B, we notice the segregation of the bilateral RIP interneurons towards the left and the right branch. This shows that the embedding does not neglect nodes outside the focus, even when their cooperation weight is set to 0.

In Figures 5A and 5B, we then focus on the interneurons by reducing the cooperation weights of sensory neurons and motoneurons in two steps. As expected, the interneurons move towards the periphery. Their organization does not seem to be dominated by left *versus* right variants, as we found for sensory neurons, but rather by a set of well-defined clusters related to their functional involvement in the *C. elegans* neuronal circuitry (see Supplementary Table 2, Supporting Information): In the first quadrant, we find the isolated AIA bilateral pair (cluster C_4).
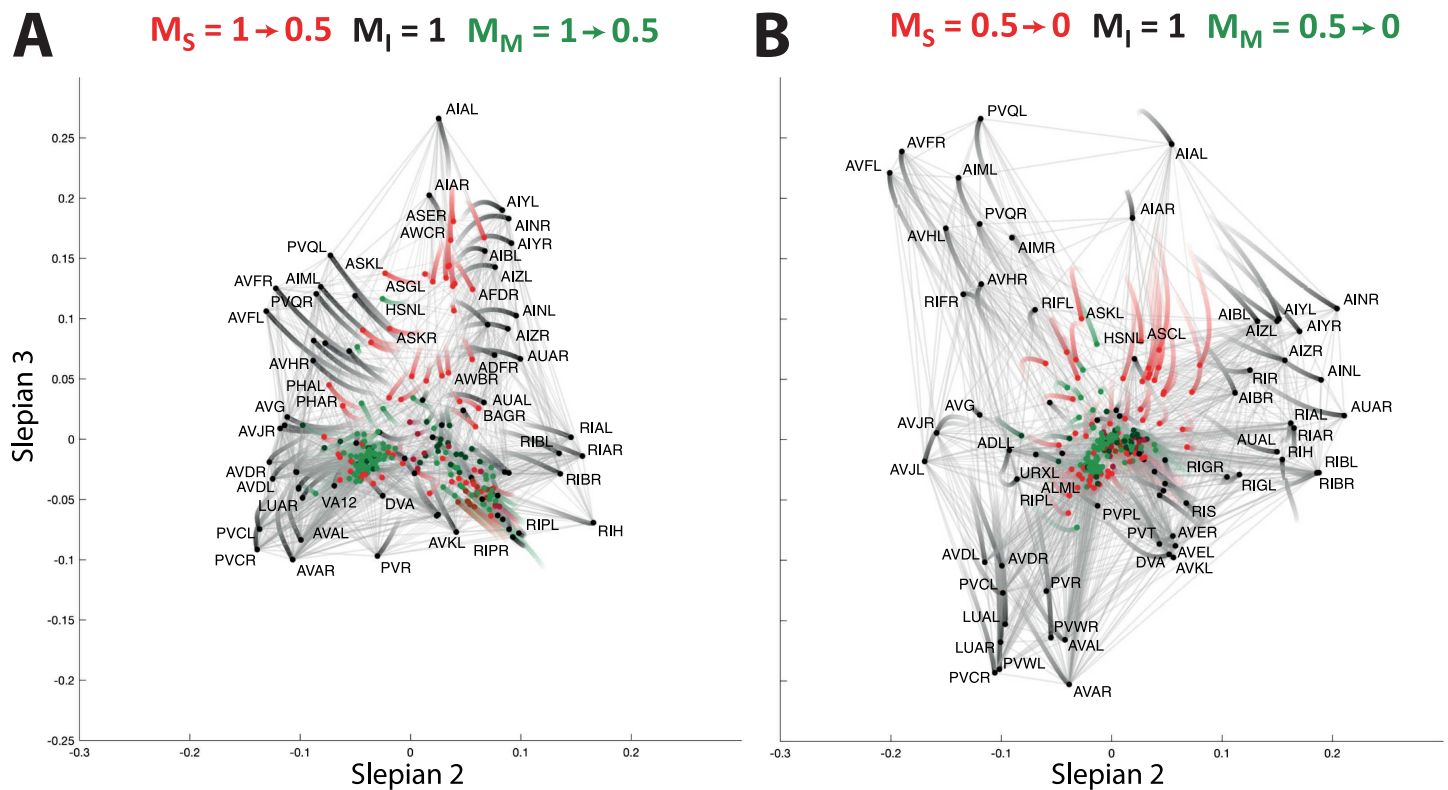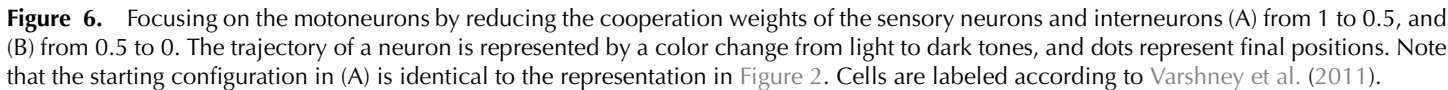
Figure 5. Focusing on the interneurons by reducing the cooperation weights of the sensory neurons and motoneurons (A) from 1 to 0.5, and (B) from 0.5 to 0. The trajectory of a neuron is represented by a color change from light to dark tones, and dots represent final positions. Note that the starting configuration in (A) is identical to the representation in Figure 2. Cells are labeled according to Varshney et al. (2011).

Moving clockwise, a larger cluster of neurons includes the bilateral AIY, AIZ, AIN, AIB, RIA, RIB, AUA, and the single neurons RIR and RIH (cluster C_3). Next we find a cluster including AVE, AVK, RIG, PVT, DVA, and other neurons located closer to the origin of Figure 5 (cluster C_5), before reaching another large ensemble of neurons including the bilateral AVA, AVD, LUA, PVC, PVW, and the single neuron PVR (cluster C_6). Moving back upwards, cluster C_1 contains the bilateral AVB, AVJ, BDU, the single neuron AVG, and PVPR, whose left counterpart PVPL belongs to cluster C_5, thus standing as the only bilateral pair of neurons split into different clusters. Finally, we reach the last group of cells containing the bilateral RIF, AVH, AIM, PVQ, and AVF (cluster C_2).

Finally, in Figures 6A and 6B, the organization of motoneurons is examined. Already in the first step (Figure 6A), when reducing the cooperation weights of the sensory and interneurons from 1.0 to 0.5, we observe much stronger changes than in the previous cases. In particular, the initial organization completely collapses and the left branch of the motoneurons spreads out. This branch then develops into a peripheral organization when further decreasing the cooperation weights (Figure 6B), with three main subsets of neurons and ambiguous positioning of the cell DVB between the left and the right bottom branches. K-means clustering into optimal cell groups captured this architecture into seven smaller clusters (Supplementary Table 3, Supporting Information): Clusters C_4 and C_7 spanned top neurons, clusters C_2 and C_3 included the bottom left branch neurons, and clusters C_5 and C_6 contained the bottom right branch cells.

Figure 6. Focusing on the motoneurons by reducing the cooperation weights of the sensory neurons and interneurons (A) from 1 to 0.5, and (B) from 0.5 to 0. The trajectory of a neuron is represented by a color change from light to dark tones, and dots represent final positions. Note that the starting configuration in (A) is identical to the representation in Figure 2. Cells are labeled according to Varshney et al. (2011).

DISCUSSION

Beyond Original Slepian Designs

The originality of our approach lies in providing a new and simple way to guide graph spectral analysis. Inspired by graph Slepians, we propose a novel criterion that combines energy concentration and modified embedded distance, taking into account cooperation weights that can gradually increase or decrease the importance of selected nodes. The new criterion lets the adjacency matrix emerge as the central graph operator, instead of the Laplacian, and is operational at full bandwidth.

This is surprising at first sight, because neither of the conventional Slepian criteria is practical without the bandlimit constraint. For the energy concentration with binary cooperation weights, as shown in Figure 3A for an illustrative example on the *C. elegans* connectome, full bandwidth leads to two eigenvalues (1 and 0), the dimensionality of the corresponding subspaces being the number of nodes with cooperation weight 1 and 0, respectively. For the modified embedded distance, as shown in Figure 3B, full bandwidth creates a subspace with eigenvalue 0 of dimensionality equal to the number of nodes with cooperation weight 0. Therefore, subtracting both criteria leads to opposing objectives; that is, at full bandwidth, an energy concentration of 1 encodes the subspace for nodes with weight 1, while a modified embedded distance of 0 encodes the subspace for nodes with weight 0.

The obtained eigenspectrum for the new criterion, shown in Figure 3C, illustrates that only a few eigenvectors are able to combine high energy concentration with low modified embedded

distance, a counterbalance that can be further revealed by measuring μ and ζ separately for these new eigenvectors. Such a large eigengap is also good news for numerical computation of the leading eigenvectors for large graphs when relying upon efficient large-scale solvers (Lehoucq & Sorensen, 1996) implemented in widely available software libraries such as ARPACK.

Intriguingly, the approximation error was already low using a linear approximation, and did not noticeably decrease further, except for the first Slepian vector, when resorting to higher order terms (see Figure 3 and Supplementary Figure S2, Supporting Information). Modifying the importance of a node *via* the corresponding cooperation value affects all-length paths through that node according to the series expansion from Equation 8, where the power of **A** in each term corresponds to the affected path length. Once we restrict the criterion to a linear approximation, the only paths whose importance is changed are those of length 1. This does not mean that other paths are not included in the graph analysis, but rather that they are included with their original (unmodified) effect on the topology. Low error of linear approximation suggests that the highest percentage of topological importance of a node falls into the importance of its length-1 paths. Further, a slightly higher error at eigenvectors with the highest ζ may be explained similarly: Not modifying higher order paths produces greater error at these eigenvectors because of their increased relative importance, because high ζ eigenvectors tend to be very smooth (even approaching a constant signal); thus, in order to even out the values at all nodes in the process, one needs to "reach" far enough.

The proposed criterion should not be confused with the Sobolev norm that is sometimes used to regularize graph signals (Mahadevan & Maggioni, 2006). Specifically, in the case of $\mathbf{M} = \mathbf{I}$, our criterion of Equation 5 applied to $\mathbf{g}$ reverts to $\mathbf{g}^\top \mathbf{g} - \mathbf{g}^\top \mathbf{L} \mathbf{g}$, whereas the Sobolev norm of $\mathbf{g}$ reads $\mathbf{g}^\top \mathbf{g} + \mathbf{g}^\top \mathbf{L} \mathbf{g}$. The difference in the sign of the second term introduces significantly distinct optimization goals regardless of the apparent similarity of the two expressions.

As for future extensions of our approach, one could envisage to dig into the relationship with graph uncertainty principles (Agaskar & Lu, 2013; Teke & Vaidyanathan, in press; Tsitsvero et al., 2016), to consider statistical resampling for graphs (Pirondini, Vybornova, Coscia, & Van De Ville, 2016), or to focus on the discovery of hierarchical graph structure (Arenas, Fernández, & Gómez, 2008; Irion & Saito, 2014) by gradual refinement of the subgraph. The design could also be extended to directed graphs using recent extensions of spectral decompositions in this context (Mhaskar, 2018; Sandryhaila & Moura, 2013).

Gaining Insights on C. elegans

The application of our newly developed approach to the *C. elegans* connectome enabled us to confirm past findings from the literature, and to shed light on additional cellular targets and groupings that may deserve further experimental analyses. At the level of sensory neurons (Figure 4, Supplementary Figure S3A, and Supplementary Figure S4A; Supporting Information), seven clusters were extracted, collectively accounting for the three branches evident in Figure 4: the top branch made of 12 (including the thermosensor AFD) pairs of amphid neurons (at *y*-coordinate greater than 0.04), and other cells split into the left and right bottom branches. Interestingly, one of the clusters found by k-means included five pairs of bilateral amphid neurons: AWA and AWC involved in odortaxis (Bargmann, Hartwieg, & Horvitz, 1993; Li et al., 2012), the thermosensor AFD (Mori & Ohshima, 1995), and ASE and ASI implicated in chemotaxis (Bargmann & Horvitz, 1991; Luo et al., 2014). These neurons act as low-order sensors,

whose extraction as a separate cluster inside the amphid group may suggest new information worth further exploration.

The lower branches in Figure 4 split the neurons into their right and left counterparts, thus extracting relevant somatic information. These neurons act as higher order sensing apparatus as compared with amphid neurons: IL1 and OLQ have jointly been implicated in the worm foraging response (Hart, Sims, & Kaplan, 1995); CEP and ADE are involved in the response upon food sensing (Sawin, Ranganathan, & Horvitz, 2000); URX, URY, and OLL are linked to the reproductive drive (Barrios, Ghosh, Fang, Emmons, & Barr, 2012), and so on. The split between low- and high-order sensing is summarized in Figure 7A.

Further inspection of the branches (Supplementary Figure S6A, Supporting Information) showed that the left-right segregation involved chemical synapses, but not gap junctions. Also, Supplementary Figure S5 (second row, Supporting Information) shows that for higher order Slepian vectors (fourth and fifth), additional contributors emerge, such as the bilateral PHA/PHB. This suggests that the approach finds different subgroups of higher order sensory neurons depending on the choice of the embedding eigenvectors. The biological/functional intepretation of the exact clusters asks for a more detailed analysis of the subgroups of neurons. Finally, the emergence of RIP interneurons in the embedding (Figure 4) points towards an important role of the sensory neurons yet to be explained, possibly in connection with their presynaptic inputs from IL1 (White et al., 1986).

Turning to interneurons (Figure 5, Supplementary Figure S3B, and Supplementary Figure S4B; Supporting Information), we notice a trend of grouping neurons at the same command-chain level. Starting from the top of Figure 5, we find AIA, AIB, AIY, and AIZ jointly known for their role on locomotory behavior and acting as a first-relay drives (Gray, Hill, & Bargmann, 2005; Wakabayashi, Kitagawa, & Shingai, 2004). Moving clockwise, we find RIA and RIB acting as second-layer intermediates, and further on, neurons such as AVE, and in the next cluster AVAL and AVD, all being command interneurons (Haspel, O'Donovan, & Hart, 2010; Hobert, 2003; Kawano et al., 2011). The trend of following the locomotory pathway clockwise in the embedding space suggests that the approach targets relevant information about the neural system. However, the exact compact clusters in Supplementary Figure S4B (Supporting Information) need further elaboration. Some of the interesting findings worth exploring would be the unexplained grouping of the scarcely studied RIR neuron (Hobert, Johnston, & Chang, 2002) with the cluster of cells including AIB and AIY, or the grouping of PVR and LUA (Chalfie et al., 1985; Wicks & Rankin, 1995) with locomotion-regulating neurons such as AVD and AVA.

Considering motoneurons (Figure 6, Supplementary Figure S3C, and Supplementary Figure S4C), the embedding positions fit somatic location (see Supplementary Figure S7): a spiral beginning at the origin, turning right, then moving clockwise and ending in the top branch follows the postero-anterior direction (Supporting Information). This confirms that the approach has extracted meaningful information. However, the exact split between the three branches as well as the k-means clustering into the seven ensembles remains unclear, since, from preliminary explorations, we find both A-type and B-type cholinergic motoneurons and the inhibitory D-type motoneurons in all clusters. Finally, DVB deserves further attention (Schuske, Beg, & Jorgensen, 2004) because of its isolated location between the two bottom branches.

In Figure 6B, two sensory nodes stick out the furthest away from the center; that is, towards the lower left and right branches of motoneurons. These are PVD and PHC neurons, responsible for nociceptive mechano- and thermosensation, respectively. The locations of these nodes in the embedding may be linked to the fact that harmful nociceptive stimuli induce a locomotory
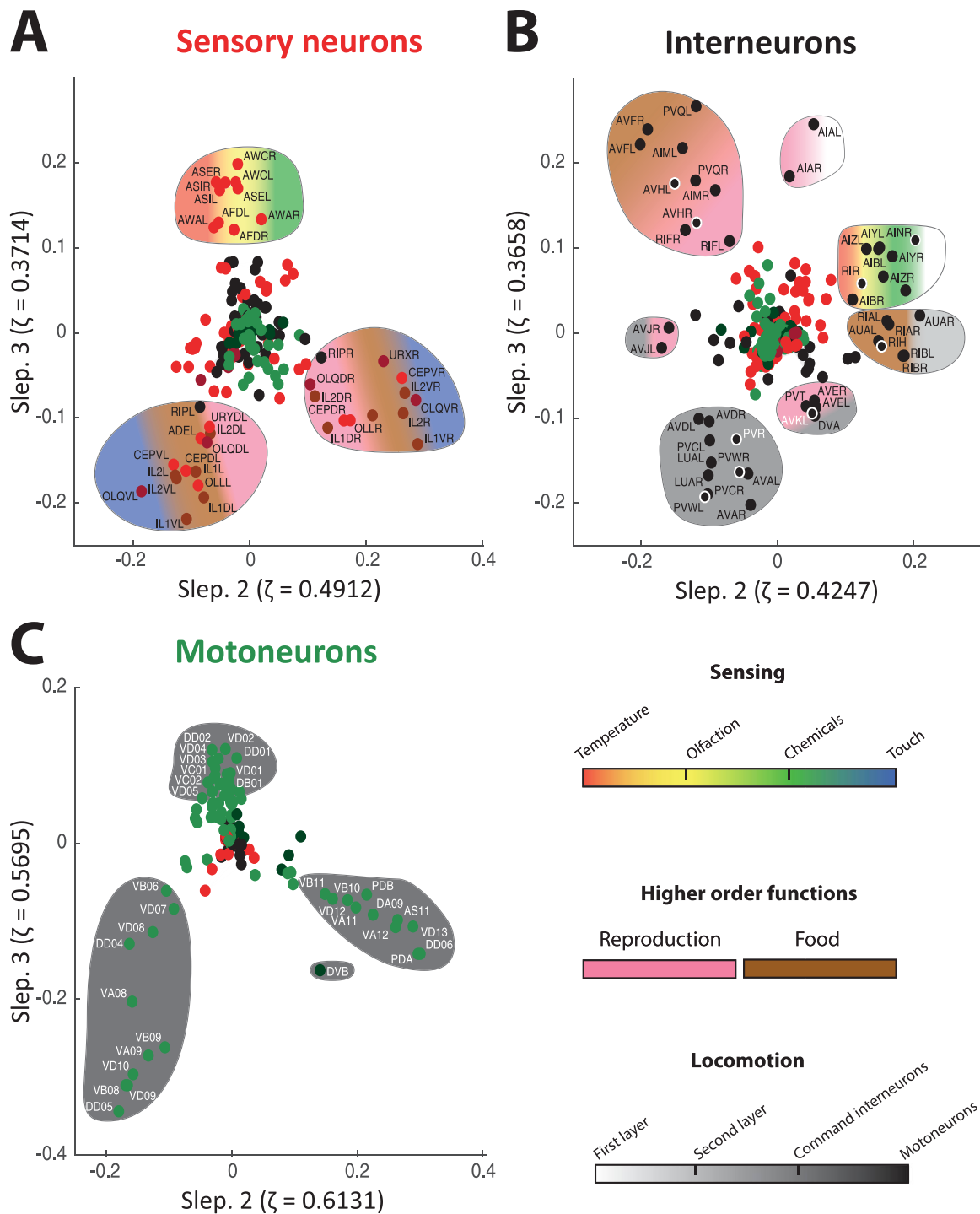
Figure 7. Summary of the main functions operated by the sensory neurons (A), interneurons (B), and motoneurons (C) unraveled by guided spectral analysis. Clusters of neurons discussed in the Discussion section are delineated and color coded according to their main roles: This may be in *sensing* (thermosensation in red, olfactory sensation in yellow, chemosensation in green, and mechanosensation in blue), *higher order functions* (reproduction in pink, food responses in brown), or *locomotion* (from first cellular relays to effector motoneurons in increasingly darker shades of gray). A gradient in the color coding indicates that more than one function is performed by neurons from a given cluster. Neurons that could not be clearly related to the rest of the unraveled circuitry are encircled in white.

response. As in the case of RIP interneurons emerging in the focused embedding of sensory neurons, we once again confirm the ability of the proposed approach to extract important nodes even when their cooperation weight was initially set to 0.

In summary, as illustrated in Figure 7, all three types of neurons found in the *C. elegans* nematode could be arranged in a meaningful hierarchy thanks to the introduced guided graph spectral embedding. Sensory neurons were separated between first-order and higher order sensors. Different levels of processing of motor functions were distinguished (see the gradient from white to dark gray tones going clockwise in Figure 7B), with the eventual recruitment of motoneurons, which have been separated on the basis of somatic location. Future analyses will allow the study of different types of neurons through more elaborate combinations of focused nodes. In addition, it will be interesting to see whether future experimental work can shed light on some of the neurons that were extracted here without being yet extensively documented in the literature, such as AVKL or RIR.

Perspectives for Future Uses

The proposed graph embedding provides a simple, yet powerful approach to visualization and, if combined with clustering techniques, to the extraction of meaningful subgraphs from any graph-modeled dataset. In neuroimaging, focusing on a specific subgraph of interest (by setting the appropriate cooperation values) can direct research onto clinically relevant concepts, such as the medial temporal lobe and limbic structures for human brain imaging studies comparing healthy controls and Alzheimer patients (Krasuski et al., 1998). Be it using the structural or the functional connectome for analyses (Contreras, Goñi, Risacher, Sporns, & Saykin, 2015), features such as cluster size and/or the inclusion of specific nodes (brain regions) in a cluster may become biomarkers for an early diagnosis or prediction of the disease.

Furthermore, graph modeling of the human brain is frequently employed to extract important nodes/brain regions and to identify their topological roles, such as provincial/connector hubs suggesting clinically significant functional roles (van den Heuvel & Sporns, 2013). Doing so requires the use of diverse node centrality measures, such as degree or betweenness centrality. On the other hand, entries of the proposed Slepian eigenvectors may be interpreted as higher order spectral centrality measures relative to the focused subgraph, and for the special case $M = I$, the eigenvector corresponding to the highest positive eigenvalue reverts to the eigenvector centrality (M. Newman, 2010). Hence, if clustering of a dataset based on the proposed embedding coordinates reveals nodes distant from the rest of the graph, it is suggested that those nodes exhibit a hub-like role when the focused subgraph is considered more important than the rest of the graph. For example, the AIA pair in the discussed *C. elegans* example emerges as a separate cluster in Figure 5 and Supplementary Figure S4B (Supporting Information), where the focus is set on interneurons. Its role as a hub can be confirmed by the high number of connections to the set of amphid neurons, and a small number of connections to the other cells, as compared with the rest of the interneurons. Identification of hubs and/or peripheral nodes with respect to other similar type nodes may lead to a better understanding of the functional role of both neurons and brain regions, depending on the inspected dataset.

AUTHOR CONTRIBUTIONS

Miljan Petrovic: Investigation; Methodology; Validation; Writing – Review & Editing. Thomas Bolton: Formal analysis; Visualization; Writing – Review & Editing. Maria Giulia Preti: Writing – Review & Editing. Raphaël Liégeois: Methodology; Validation; Writing – Review & Editing. Dimitri Van De Ville: Conceptualization; Formal analysis; Writing – Original Draft; Writing – Review & Editing.

REFERENCES

Agaskar, A., & Lu, Y. M. (2013). A spectral graph uncertainty principle. *IEEE Transactions on Information Theory*, *59*(7), 4338–4356.

Arenas, A., Fernández, A., & Gómez, S. (2008). Analysis of the structure of complex networks at different resolution levels. *New Journal of Physics*, *10*(5), 053039.

Bargmann, C. I., Hartwieg, E., & Horvitz, H. R. (1993). Odorant-selective genes and neurons mediate olfaction in *C. elegans*. *Cell*, *74*(3), 515–527.

Bargmann, C. I., & Horvitz, H. R. (1991). Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in *C. elegans*. *Neuron*, *7*(5), 729–742.

Barrios, A., Ghosh, R., Fang, C., Emmons, S. W., & Barr, M. M. (2012). PDF-1 neuropeptide signaling modulates a neural circuit for mate-searching behavior in *C. elegans*. *Nature Neuroscience*, *15*(12), 1675.

Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, *15*(6), 1373–1396.

Chalfie, M., Sulston, J. E., White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1985). The neural circuit for touch sensitivity in *Caenorhabditis elegans*. *Journal of Neuroscience*, *5*(4), 956–964.

Chen, B. L., Hall, D. H., & Chklovskii, D. B. (2006). Wiring optimization can relate neuronal structure and function. *Proceedings of the National Academy of Sciences*, *103*(12), 4723–4728.

Chung, F. R. K. (1997). *Spectral graph theory*. Providence: American Mathematical Society.

Contreras, J. A., Goñi, J., Risacher, S. L., Sporns, O., & Saykin, A. J. (2015). The structural and functional connectome and prediction of risk for cognitive impairment in older adults. *Current Behavioral Neuroscience Reports*, *2*(4), 234–245. https://doi.org/10.1007/s40473-015-0056-z

De Lange, S., De Reus, M., & Van den Heuvel, M. (2014). The laplacian spectrum of neural networks. *Frontiers in Computational Neuroscience*, *7*(189), 1–12.

Fiedler, M. (1989). Laplacian of graphs and algebraic connectivity. *Combinatorics and Graph Theory*, *25*, 57–70.

Gray, J. M., Hill, J. J., & Bargmann, C. I. (2005). A circuit for navigation in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences*, *102*(9), 3184–3191.

Hammond, D. K., Vandergheynst, P., & Gribonval, R. (2011). Wavelets on graphs via spectral graph theory. *Applied and Computational Harmomic Analysis*, *30*(2), 129–150.

Hart, A. C., Sims, S., & Kaplan, J. M. (1995). Synaptic code for sensory modalities revealed by *C. elegans* GLR-1 glutamate receptor. *Nature*, *378*(6552), 82.

Haspel, G., O'Donovan, M. J., & Hart, A. C. (2010). Motoneurons dedicated to either forward or backward locomotion in the nematode *Caenorhabditis elegans*. *Journal of Neuroscience*, *30*(33), 11151–11156.

Hobert, O. (2003). Behavioral plasticity in *C. elegans*: Paradigms, circuits, genes. *Journal of Neurobiology*, *54*(1), 203–223.

Hobert, O., Johnston, R. J., Jr., & Chang, S. (2002). Left–right asymmetry in the nervous system: The *Caenorhabditis elegans* model. *Nature Reviews Neuroscience*, *3*(8), 629.

Huang, W., Bolton, T. A., Medaglia, J. D., Bassett, D. S., Ribeiro, A., & Van De Ville, D. (2018). A graph signal processing perspective on functional brain imaging. *Proceedings of the IEEE*.

Irion, J., & Saito, N. (2014). Hierarchical graph Laplacian eigen transforms. *JSIAM Letters*, *6*, 21–24. https://doi.org/10.14495/jsiaml.6.21

Kawano, T., Po, M. D., Gao, S., Leung, G., Ryu, W. S., & Zhen, M. (2011). An imbalancing act: Gap junctions reduce the backward motor circuit activity to bias *C. elegans* for forward locomotion. *Neuron*, *72*(4), 572–586.

Krasuski, J. S., Alexander, G. E., Horwitz, B., Daly, E. M., Murphy, D. G., Rapoport, S. I., & Schapiro, M. B. (1998). Volumes of medial temporal lobe structures in patients with alzheimer's disease and mild cognitive impairment (and in healthy controls). *Biological Psychiatry*, *43*(1), 60–68. https://doi.org/10.1016/s0006-3223(97)00013-9

Lehoucq, R. B., & Sorensen, D. C. (1996). Deflation techniques for an implicitly restarted Arnoldi iteration. *SIAM Journal on Matrix Analysis and Applications*, *17*(4), 789–821.

Li, Z., Li, Y., Yi, Y., Huang, W., Yang, S., Niu, W., . . . Xu, T. (2012). Dissecting a central flip-flop circuit that integrates contradictory sensory cues in *C. elegans* feeding regulation. *Nature Communications*, *3*, 776.

Lu, L.-Z., & Pearce, C. E. M. (2000). Some new bounds for singular values and eigenvalues of matrix products. *Annals of Operations Research*, *98*(1–4), 141–148.

Luo, L., Wen, Q., Ren, J., Hendricks, M., Gershow, M., Qin, Y., . . . Zhang, Y. (2014). Dynamic encoding of perception, memory, and movement in a *C. elegans* chemotaxis circuit. *Neuron*, *82*(5), 1115–1128.

Mahadevan, S., & Maggioni, M. (2006). Value function approximation with diffusion wavelets and Laplacian eigenfunctions. In Y. Weiss, B. Schölkopf, & J. C. Platt (Eds.), *Advances in neural information processing systems 18* (pp. 843–850). Cambridge, MA: MIT Press.

Mhaskar, H. (2018). A unified framework for harmonic analysis of functions on directed graphs and changing data. *Applied and*

Computational Harmonic Analysis, *44*(3), 611–644. https://doi.org/10.1016/j.acha.2016.06.007

Mori, I., & Ohshima, Y. (1995). Neural regulation of thermotaxis in *Caenorhabditis elegans*. *Nature*, *376*(6538), 344.

Newman, M. (2010). *Networks: An introduction*. New York, NY: Oxford University Press.

Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, *74*(3). https://doi.org/10.1103/physreve.74.036104

Ortega, A., Frossard, P., Kovačević, J., Moura, J. M., & Vandergheynst, P. (2018). Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE*, *106*(5), 808–828.

Pirondini, E., Vybornova, A., Coscia, M., & Van De Ville, D. (2016). Spectral method for generating surrogate graph signals. *IEEE Signal Processing Letters*, *23*(9), 1275–1278. https://doi.org/10.1109/LSP.2016.2594072

Rousseeuw, P. J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1126/science.290.5500.2323

Roweis, S. T. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, *290*(5500), 2323–2326. https://doi.org/10.1016/0377-0427(87)90125-7

Sammon, J. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, *C-18*(5), 401–409. https://doi.org/10.1109/t-c.1969.222678

Sandryhaila, A., & Moura, J. M. F. (2013). Discrete signal processing on graphs. *IEEE Transactions on Signal Processing*, *61*(7), 1644–1656.

Sawin, E. R., Ranganathan, R., & Horvitz, H. R. (2000). *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. *Neuron*, *26*(3), 619–631.

Schaub, M. T., Delvenne, J.-C., Lambiotte, R., & Barahona, M. (2018). Multiscale dynamical embeddings of complex networks.

Schönemann, P. H. (1966). A generalized solution of the orthogonal procrustes problem. *Psychometrika*, *31*(1), 1–10.

Schuske, K., Beg, A. A., & Jorgensen, E. M. (2004). The GABA nervous system in *C. elegans*. *Trends in Neurosciences*, *27*(7), 407–414.

Shen, X., & Meyer, F. G. (2008). Low-dimensional embedding of fMRI datasets. *NeuroImage*, *41*(3), 886–902. https://doi.org/10.1016/j.neuroimage.2008.02.051

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, *30*(3), 83–98.

Slepian, D. (1978). Prolate spheroidal wave-functions, Fourier-analysis, and uncertainty. Discrete case. *Bell System Technical Journal*, *57*(5), 1371–1430.

Slepian, D., & Pollak, H. O. (1961). Prolate spheroidal wave functions, Fourier analysis, and uncertainty. *Bell System Technical Journal*, *40*(1), 43–63.

Teke, O., & Vaidyanathan, P. P. (in press). Uncertainty principles and sparse eigenvectors of graphs. *IEEE Transactions on Signal Processing*.

Tenenbaum, J. B. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, *290*(5500), 2319–2323. https://doi.org/10.1126/science.290.5500.2319

Tsitsvero, M., Barbarossa, S., & Di Lorenzo, P. (2016). Signals on graphs: Uncertainty principle and sampling. *IEEE Transactions on Signal Processing*, *64*(18), 4845–4860.

van den Heuvel, M. P., & Sporns, O. (2013). Network hubs in the human brain. *Trends in Cognitive Sciences*, *17*(12), 683–696. https://doi.org/10.1016/j.tics.2013.09.012

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605. Retrieved from http://www.jmlr.org/papers/v9/vandermaaten08a.html

Van De Ville, D. (2016). Steering macro-scale network community structure by micro-scale features. arXiv: 1603.05623

Van De Ville, D., Demesmaeker, R., & Preti, M. G. (2017a). Guiding network analysis using graph Slepians: An illustration for the *C. elegans* connectome. In *Wavelets and sparsity xvii* (Vol. 10394, p. 103941Y).

Van De Ville, D., Demesmaeker, R., & Preti, M. G. (2017b). When Slepian meets Fiedler: Putting a focus on the graph spectrum. *IEEE Signal Processing Letters*, *24*, 1001–1004. https://doi.org/10.1109/LSP.2017.2704359

Varier, S., & Kaiser, M. (2011). Neural development features: Spatiotemporal development of the *Caenorhabditis elegans* neuronal network. *PLoS Computational Biology*, *7*(1), e1001044.

Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H., & Chklovskii, D. B. (2011). Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Computational Biology*, *7*(2), e1001066.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, *17*(4), 395–416.

Wakabayashi, T., Kitagawa, I., & Shingai, R. (2004). Neurons regulating the duration of forward locomotion in *Caenorhabditis elegans*. *Neuroscience Research*, *50*(1), 103–111.

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *314*(1165), 1–340.

Wicks, S. R., & Rankin, C. H. (1995). Integration of mechanosensory stimuli in *Caenorhabditis elegans*. *Journal of Neuroscience*, *15*(3), 2434–2444.