



Community detection for directed networks revisited using bimodularity

Alexandre Cionca^{a,b} , Chun Hei Michael Chan^{a,b}, and Dimitri Van De Ville^{a,b,1}

Edited by David Donoho, Stanford University, Stanford, CA; received February 11, 2025; accepted July 14, 2025

Community structure is a key feature omnipresent in real-world network data. Plethora of methods have been proposed to reveal subsets of densely interconnected nodes using criteria such as the modularity index. These approaches have been successful for undirected graphs but directed edge information has not yet been dealt with in a satisfactory way. Here, we revisit the concept of directed communities as a mapping between sending and receiving communities. This translates into a definition that we term bimodularity. Using convex relaxation, bimodularity can be optimized with the singular value decomposition of the directed modularity matrix. Subsequently, we propose an edge-based clustering approach to reveal the directed communities including their mappings. The feasibility of the framework is illustrated on a synthetic model and further applied to the neuronal wiring diagram of the *Caenorhabditis elegans*, for which it yields meaningful feedforward loops of the head and body motion systems. This framework sets the ground for the understanding and detection of community structures in directed networks.

community structure | modularity | directed graphs | spectral clustering

Networks are powerful models to represent interactions within almost any type of structured data. Nodes and edges can symbolize relationships between agents in social networks, metabolic interactions between cells, or information flow between neuronal populations in the human brain. The study of the emerging network properties provides critical insights into the observed data (1–6). One predominant attribute of many real-world networks is the presence of community structure where subsets of nodes are more densely connected between them than expected by the degree distribution (7). Community detection is about finding the best graph partitioning \mathcal{C}_k , $k = 1, \dots, K$ (8–10). Mathematically, the modularity index Q for an undirected graph captures the excess proportion of edges running within the communities (11, 12):

$$Q = \frac{1}{2m} \sum_{k=1}^K \sum_{i,j \in \mathcal{C}_k} [A_{ij} - \mathbb{E}(A_{ij}|\mathcal{H}_0)], \quad [1]$$

where m is the total number of edges, the elements of adjacency matrix A_{ij} contain the edge weights between nodes i and j , and $\mathbb{E}(A_{ij}|\mathcal{H}_0)$ expresses the expected proportion under the null hypothesis. The configuration model is a common null hypothesis that redistributes the edges weights evenly over the nodes; i.e., $k_i k_j / (2m)$, where k_i is the degree of node i . Modularity can thus be seen as a statistical measure of the unexpectedness in edges arrangement and of how graph partitions are exploiting such modular structures to form densely connected (assortative) or bipartite (disassortative) communities (13–15). Many approaches have been proposed to optimize modularity, such as the Louvain algorithm which operates in the graph domain (16, 17) or the spectral method that eigendecomposes the modularity matrix \mathbf{B} , where $B_{ij} = A_{ij} - k_i k_j / (2m)$ (18, 19). Such optimizations can be interpreted as embedding problems for the graph nodes or edge, which in the latter may provide overlapping groups of nodes (20, 21).

At first sight, the definition of modularity is amendable for directed graphs. First, the adjacency matrix considered in Q can be asymmetric. Second, the null model can be changed to $k_i^{\text{out}} k_j^{\text{in}} / m$ to account for out- and in-degree of the nodes (22, 23). However, this type of directed modularity does not consider edge direction (24, 25); i.e., the contribution to modularity of a specific edge between nodes i and j remains constant as long as $A_{ij} + A_{ji}$ and the in- and out-degrees are the same (25). In addition, the spectral method cannot be applied to nonsymmetric matrices.

Considerable work has been carried out to go beyond these limitations for community detection in directed networks (9, 23, 25–29). First, the spectral method has been

Significance

The art of finding patterns or communities plays a central role in the analysis of structured data such as networks. Community detection in graphs has become a field on its own. Real-world networks, however, tend to describe asymmetric, directed relationships, and community detection methods have not yet reached consensus on how to define and retrieve communities in this setting. This work introduces a framework for the interpretation of directed graph partitions and communities, for which we define the bimodularity index and provide an optimization method to retrieve the embedding and detection of directed communities. The application of our approach to the worm neuronal wiring diagram highlights the importance of directed information that remains hidden from conventional community detection.

Author affiliations: ^aNeuro-X Institute, École Polytechnique Fédérale de Lausanne, Geneva 1202, Switzerland; and ^bDepartment of Radiology and Medical Informatics, University of Geneva, Geneva 1202, Switzerland

Author contributions: A.C. and D.V.D.V. designed research; A.C., C.H.M.C., and D.V.D.V. performed research; A.C., C.H.M.C., and D.V.D.V. analyzed data; and A.C., C.H.M.C., and D.V.D.V. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2025 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹To whom correspondence may be addressed. Email: dimitri.vandeville@epfl.ch.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2500571122/-/DCSupplemental>.

Published August 25, 2025.

applied to the symmetrized matrix $\mathbf{B} + \mathbf{B}^T$. However, because this approach obfuscates edge direction (23), the directed modularity has instead been redefined with random walk priors in the LinkRank method (25). It is defined as the difference between the fraction of time a random walker will spend within communities and the expected value of this fraction. While this approach accounts for edge direction and is coherent with the undirected modularity, the matrix definition of the LinkRank modularity is, however, asymmetric and the authors therefore present an optimization process that again requires symmetrization similar to the one in ref. 23. Alternative symmetrization methods have also been considered, such as the weighted label propagation algorithm to convert a directed network into an undirected weighted yet signed graph (26). Second, recent approaches proposed a workaround to matrix symmetrization through spectral clustering based on the singular value decomposition (SVD) of directed graph operators (9, 27–29). The left and right singular vectors indeed capture meaningful node embeddings that are tied to the row and column spaces, respectively, of the asymmetric normalized Laplacian (27), regularized Laplacian (9, 28), and adjacency (29) matrices. While the SVD offers relevant adaptations of spectral methods to directed graphs, none of these frameworks combines the SVD with the notion of a null model that is proper to modularity, nor leverage the natural correspondence between left- and right-singular vectors in their clustering scheme. In sum, there is not yet a satisfactory, intuitive, and straightforward extension of community detection to directed graphs.

Here, we revisit the definition of modularity in a subtle but essential way. In particular, instead of considering a single partitioning in terms of communities, we propose the concept of bimodularity that identifies both sending and corresponding receiving communities; i.e., two partitions that are not necessarily overlapping. The directed nature of the graph can therefore be reflected in terms of community structure. We also derive an efficient algorithm that jointly optimizes both partitions and detects directed communities. The feasibility and interpretation are first demonstrated on synthetic examples. Finally, insightful results are obtained on the directed neuronal wiring diagram of the nematode *Caenorhabditis elegans* and indicate the potential of this method for real-world applications.

Results

Bimodularity. Communities in undirected graphs are characterized by more edges than expected by the null model, however, in directed graphs, these edges do not necessarily map to the same set of nodes. Therefore, we introduce K sending communities as $C^{\text{out}} = (C_1^{\text{out}}, C_2^{\text{out}}, \dots, C_K^{\text{out}})$ and assume a mapping \mathcal{M} that relates each sending community to a corresponding receiving one such that $C^{\text{in}} = (C_1^{\text{in}}, C_2^{\text{in}}, \dots, C_K^{\text{in}})$, where $\mathcal{M}(C_k^{\text{out}}) = C_k^{\text{in}}$. The sending and receiving communities can also be overlapping between them. We now define bimodularity Q_{bi} as the difference between the fraction of edges from sending communities to their respective receiving communities and the expected value of this fraction:

$$Q_{\text{bi}} = \frac{1}{m} \sum_{k=1}^K \sum_{\substack{i \in C_k^{\text{out}} \\ j \in C_k^{\text{in}}}} [A_{ij} - \mathbb{E}(A_{ij} | \mathcal{H}_0)], \quad [2]$$

where edges (i, j) run from a sending to the corresponding receiving community. Bimodularity reverts to conventional

modularity when the graph is undirected and the two partitions coincide (see *Materials and Methods* for details).

Let us consider the canonical directed community structure in Fig. 1A; i.e., we have a cycle between four sets of nodes that are not only densely connected within themselves but also between them according to the directional pattern that is indicated. Two types of communities can be recognized: first, the conventional community within the sets, and second, the directed communities between the sets according to the cycle structure. The adjacency matrix for an instantiation of this model is shown in Fig. 1B. Each set contains 50 nodes, connected with density $\gamma = 30\%$ with randomly picked nodes within the set, and then connected to other sets with the same edge density and following the directions of the model. The modularity matrix \mathbf{B} is defined as $B_{ij} = A_{ij} - k_i^{\text{out}} k_j^{\text{in}} / m$. The configuration model for this graph as well as the modularity matrix are illustrated in *SI Appendix, Fig. S1*.

Spectral Method for Optimizing Bimodularity. Finding the directed communities can be elegantly solved by a spectral method that maximizes the bimodularity index Q_{bi} . If we revert for now to the bipartitioning problem where only two sending/receiving communities are to be identified, then this information can be encoded in two indicator vectors \mathbf{s}^{out} and \mathbf{s}^{in} , respectively, of which the elements take values $+1/-1$ to indicate the partition. The bimodularity index can then be rewritten as

$$Q_{\text{bi}}(\mathbf{s}^{\text{out}}, \mathbf{s}^{\text{in}}) = \frac{1}{m} \sum_{\substack{(i,j) \text{ s.t.} \\ \mathbf{s}^{\text{out}}[i] = \mathbf{s}^{\text{in}}[j]}} B_{ij} = \frac{1}{2m} (\mathbf{s}^{\text{out}})^T \mathbf{B} \mathbf{s}^{\text{in}}. \quad [3]$$

Convex relaxation of the vectors \mathbf{s}^{out} and \mathbf{s}^{in} does not restrict their values to $+1/-1$ and instead imposes a unit norm, which allows to obtain the solution by the SVD: $\mathbf{B} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$; see *Materials and Methods* for the full derivation. The columns of \mathbf{U} contain the left singular vectors, those of \mathbf{V} the right singular vectors, and the diagonal elements of $\mathbf{\Sigma}$ are the corresponding singular values that relate to bimodularity up to a factor m .

Akin to Laplacian spectral embedding, different pairs of singular vectors $(\mathbf{u}_k, \mathbf{v}_k)$ provide representations of the network by decreasing bimodularity (1, 30). For the running example, the singular values are plotted in Fig. 1C and reveal two components with large bimodularity. The visualization of the first component in Fig. 1D essentially unveils the diagram of the underlying model, with the four sets clearly separated and edges running counterclockwise.

Detection of Directed Communities. Bimodularity-based embedding of the graph nodes provides insightful representations of network organization. However, an additional step is needed to effectively detect pairs of sending C_k^{out} and receiving communities C_k^{in} , termed bicomunities.

The key to recover the mapping between sending and receiving communities from the embedding is to perform clustering of the edges instead of the nodes. An edge is represented by the projection of its source node i onto the sending partition and, likewise, the projection of its target node j onto the receiving partition. For instance, the first component of the SVD provides for every edge (i, j) the 2-D feature vector $(\mathbf{u}_1[i], \mathbf{v}_1[j])$. For N components, the complete feature vector for an edge (i, j) can be obtained by concatenation and scaling by the singular values for

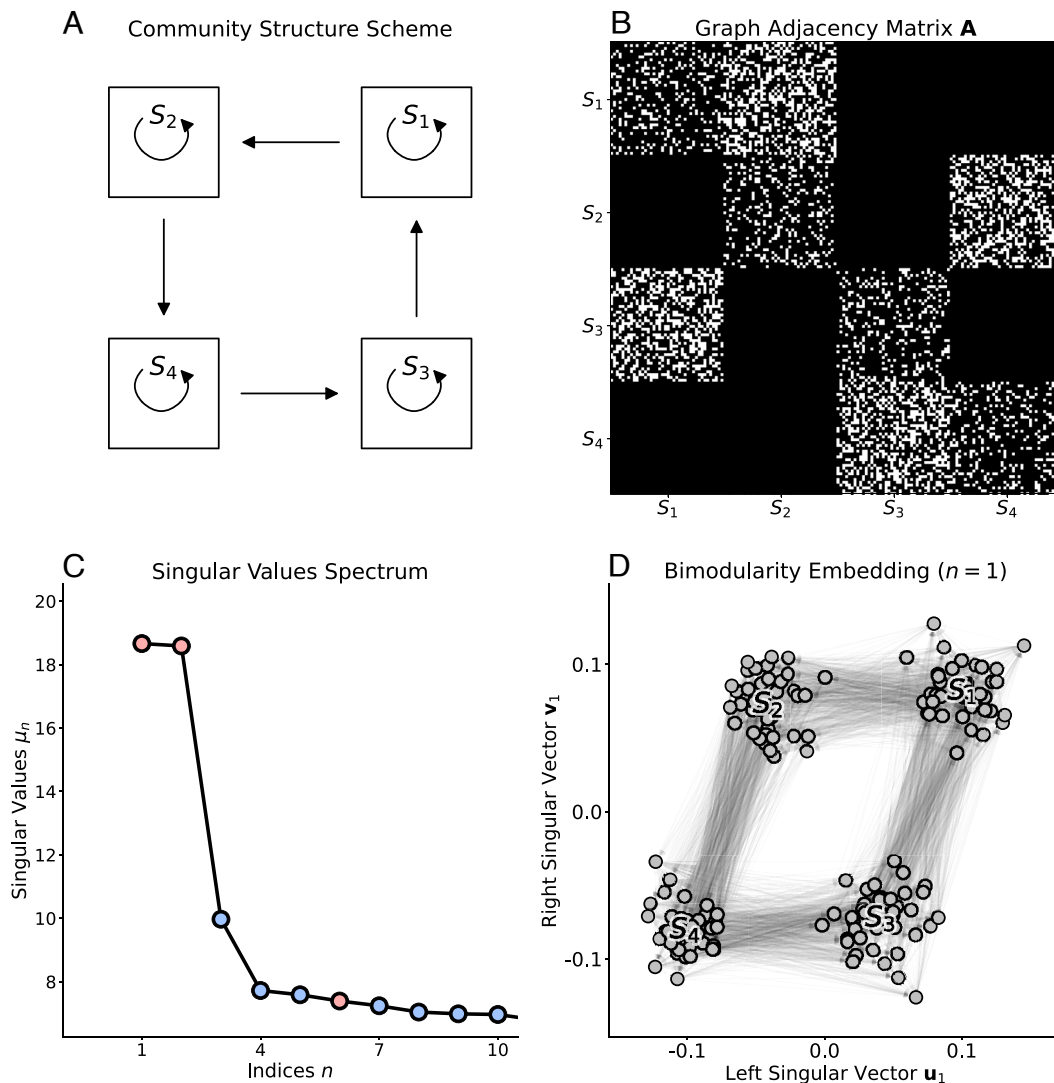


Fig. 1. Bimodular community structure. (A) Model of conventional communities (S_1, S_2, S_3, S_4) that additionally project onto one another according to a cycle structure. (B) Adjacency matrix \mathbf{A} of a stochastic block graph of 200 nodes (50 per community) with the structure of A. (C) Spectrum of the 10 highest singular values of the singular value decomposition (SVD) of **B**. Colors indicate the sign of the singular value associated to assortative (red, $\mu > 0$) or disassortative (blue, $\mu < 0$) community structures. (D) Projection of the graph nodes on the two-dimensional space given by the first component of the SVD of **B**; i.e., the left and right singular vectors associated to the highest singular value μ_1 .

stability purpose (SI Appendix, Fig. S2) (31):

$$\mathbf{f} = (\mu_1 \mathbf{u}_1[i], \mu_1 \mathbf{v}_1[j], \dots, \mu_N \mathbf{u}_N[i], \mu_N \mathbf{v}_N[j]).$$

The set of all feature vectors is then fed into a clustering algorithm (i.e., k-means) to identify groups of edges that represent the bicomunity mappings. The sending and receiving parts of such a bicomunity then emerge by considering the set of source and target nodes, respectively. Note that, with this detection method, a sending community can overlap with its corresponding receiving partition (i.e., self-community) or with the sending or receiving part of other bicomunities. The flowchart in Fig. 2 illustrates the process for the running example. The clustering of the edges leads to 8 communities, 4 conventional ones, and 4 bicomunities, thus retrieving the complete original model (Fig. 2E).

Bicomunities of the *C. elegans* Neuronal Network. We present bimodularity maximization to highlight directed communities in experimental data, in particular the neuronal connectome of the hermaphrodite nematode *C. elegans* (Materials and Methods).

Fig. 3A shows that there are several singular vectors with high bimodularity index. The embedding of the neurons (nodes) onto the first bimodular component is displayed in Fig. 3B, which shows a distinct segregation between the head and body motion systems (SI Appendix, Fig. S3). The head motion circuit appears in the lower-left quadrant ($\mathbf{u}_1 < 0$ and $\mathbf{v}_1 < 0$) of the embedding space with sensory (*OLL*, *CEPD*) neurons projecting to intermediate ones (*RIA*, *RIH*, *RIS*) and then reaching motor neurons (*RMD*, *SMD*). Similarly, neurons of the body motion circuit are embedded in the upper-right quadrant ($\mathbf{u}_1 > 0$ and $\mathbf{v}_1 > 0$). Body motion neurons are distributed along the mid-upper part of the vertical axis ($\mathbf{u}_1 = 0$) and receive mostly from sensory (*PHB*, *PLM*) and inter- (*PVC*, *LUA*) neurons of the body. Some notable head interneurons (*AVA*, *AVB*, *AVD*) stand out within the body motion circuit and seem to provide major projections to both body inter- and motor neurons. The *AVE* interneurons seem to play a similar mediation role. This complex interaction fully emerges from the analysis of directed connectivity through the lens of bimodularity.

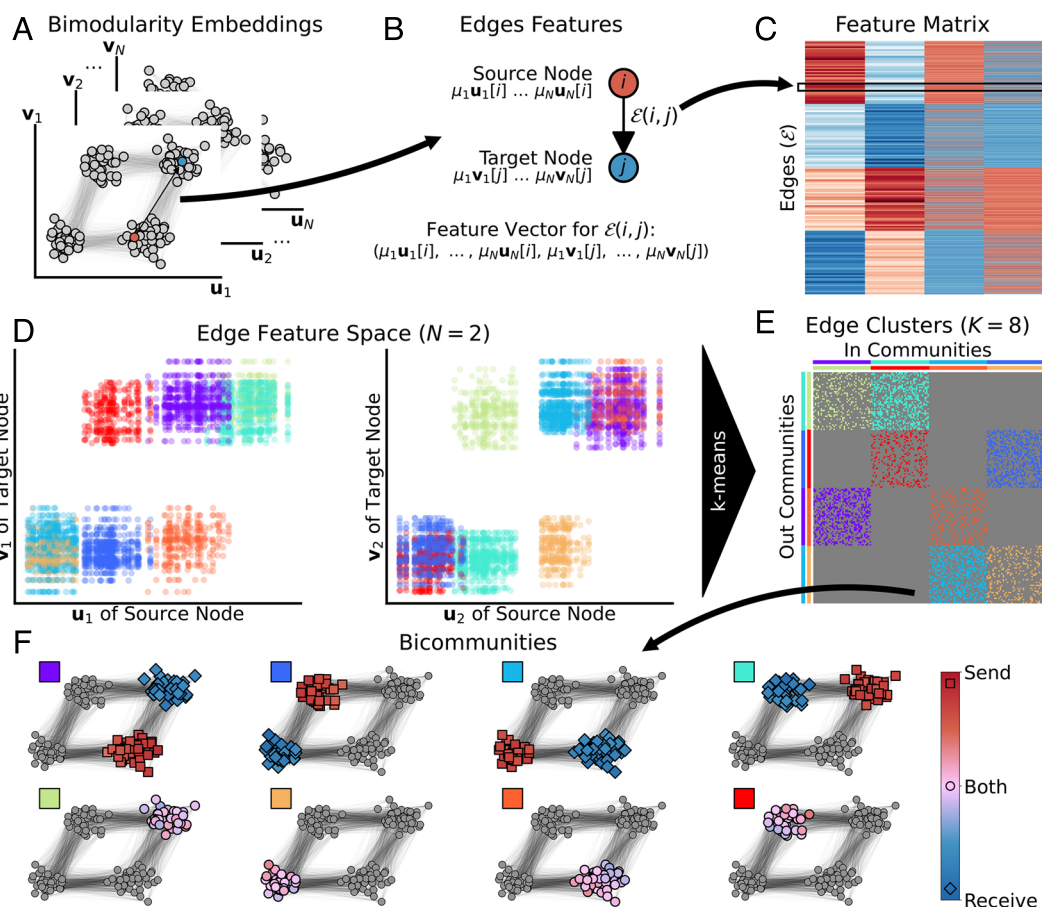


Fig. 2. Detection of bicomunities. (A) N Left u and Right v singular vectors of the SVD of the modularity matrix B (bimodularity embeddings). (B) Aggregation of embeddings at the level of graph nodes to build a feature vector that represents the sending and receiving partition of the source (i) and target (j) node respectively. (C) Feature matrix with features of all graph edges. (D) Edge feature space (first two SVD components). Each dot represents an edge of the graph colored based on its corresponding cluster. (E) Graph adjacency matrix where each edge cluster is colored. Sending (Left) and receiving (Top) nodes are shown with colors of the corresponding cluster. (F) Bicomunities of the block-cycle graph with colors corresponding to the edge cluster. Colors and markers of nodes indicate whether a node belongs to the sending (red, square), receiving (blue, diamond), or both (circle, purple) community.

To further interpret these findings in terms of information pathways, we first denote the difference between sensory and motor neurons that tend to be more distributed on the horizontal and vertical axes of the embedding, respectively. This is coherent with the idea that sensory (motor) neurons may have more edges going out (in) thus they will be aligned along the axis that characterizes the sending u_1 (receiving v_1) behavior. With this in mind, we describe three information pathways (SI Appendix, Fig. S4): the head motion pathway in the lower left quadrant; the body motion pathway in the upper right quadrant; and a last, more complex, pathway from head sensory to body motor neurons by passing through the key *AVE* interneurons. Specifically, these interneurons play a central part in conveying signals from the head system to the body system and finally back to motor neurons of the head and neck SI Appendix, Fig. S4.

Fig. 4 shows the 3 bicomunities of the *C. elegans* with highest bimodularity index (out of 5) extracted from the 5 first components of the SVD of B . These clusters highlight specific information pathways in the worm's motion system. In detail, C_3 identifies communication from body sensory processes to key interneurons in the worm's brain (*AVA*, *AVB*, *AVD*). C_4 describes feed-forward transmissions of information from these interneurons to motor neurons of the body and interneurons of the tail (*PVC*). C_2 finally shows interconnectivity between motor neurons of the body and the neck as well as their communication

with interneurons of the worm's tail. Finer definitions of the *C. elegans* bicomunities are achieved with a larger number of edge clusters; see SI Appendix, Fig. S5.

Discussion

We investigated community structure in directed graphs and how the concept of sending and receiving communities is key to provide insight into such network organization. Specifically, we revisited the definition of modularity for directed graphs to introduce bimodularity being tailored to account for the directed nature of edges. The main idea behind bimodularity is to provide pairs of sending and receiving communities that are not necessarily mutually exclusive. We presented a spectral method for maximization of bimodularity through the SVD of the modularity matrix. This optimization revealed embeddings based on bimodular structure. We then proposed the appropriate clustering approach to combine these embeddings and obtain the bicomunities in terms of sending/receiving communities and their mappings. We demonstrated the feasibility of the approach on synthetic models as well as to unravel the direct communication pathways in the *C. elegans*. Neurons and connections which are central in the worm's sensory and motion systems are made visible in the first dominant bimodularity embedding.

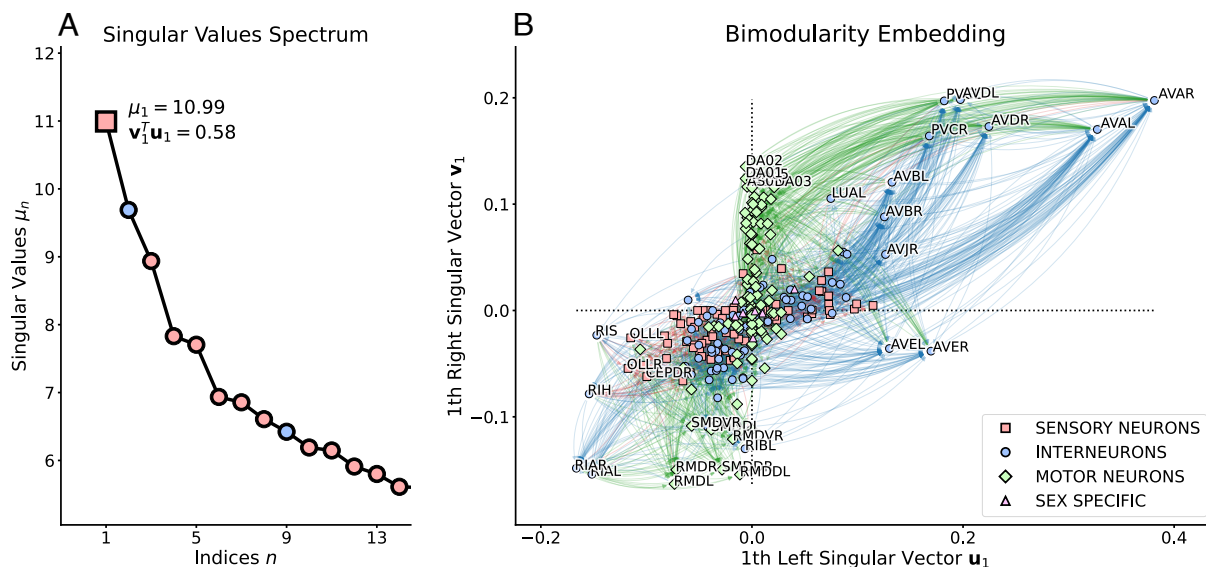


Fig. 3. Singular value decomposition (SVD) of the modularity matrix built from the *C. elegans* wiring network of chemical synapses. (A) Spectrum of the singular values of the SVD. Colors indicate the sign of the singular value associated to assortative (red, $\mu > 0$) or disassortative (blue, $\mu < 0$) community structures. (B) Projection of the graph nodes (neurons) onto the *Left* (\mathbf{u}_1 , horizontal axis) and *Right* (\mathbf{v}_1 , vertical axis) singular vectors of the bimodular component associated with the largest singular value ($\mu_1 = 10.99$). Colors and shape of nodes indicate the type of the neurons (sensory: red square; interneuron: blue circle; motor: green diamond; or sex specific: purple triangle). Edge colors indicate the type of the target node with the same color code as for the neuron types. Neuron labels are shown for graph nodes that are projected far from the origin (0, 0) of the axes.

Bicommunities dissect the sensory to motor feed-forward loop through key interneurons of the head and the tail.

Methods previously tackled the embedding of directed graph nodes through the SVD of graph operators (e.g. adjacency or Laplacian) (9, 27–29). These approaches however usually fail to provide an interpretation of the embedding space. This limitation is further emphasized as they do not leverage the intrinsic correspondence between the left and right singular vectors through the associated singular value. The proposed bimodularity perspective intuitively describes the relationship between a sending and its corresponding receiving projection. Then, the bimodular edge embedding is a natural approach to identify edges mapping from nodes with similar sending patterns to nodes with similar receiving ones. In comparison to conventional approaches, the proposed edge embedding is unique to bimodularity in the interpretability of edge features (20, 21).

Advances in machine learning have recently also contributed to directed community detection (32–34). Specifically, graph neural networks that incorporate the adjacency line graph (i.e., edges of the original graph are turned into nodes and then connected when sharing an original node) (32), or a spectral convolution layer via the magnetic Laplacian that is Hermitian (33, 35), have proven to be promising avenues for learning node embeddings. However, learning-based methods rely on training processes in high-dimensional parameter spaces that limit their generalizability and interpretability. In addition, it is not straightforward to understand how directionality is exploited to represent communities in terms of sending or receiving constituents. To overcome this shortcoming, a shallower neural architecture has been proposed to learn node embeddings that represent the source and target roles (34). Specifically, single-layer neural models were used to extract embeddings from features of an alternating random walk. Such identification of source and target roles is similar to our concept of sending and receiving communities, albeit there is no correspondence between

sending and receiving patterns, as is the case for bimodularity. These frameworks primarily focus on representation learning instead of unsupervised discovery. In contrast, SVD-based edge embedding and detection of bicommunities flow naturally from the definition of bimodularity. Nevertheless, future work could integrate these operations into learning frameworks as well.

We investigated the properties of bimodularity, such as the fact that it coincides with conventional modularity when the graph is undirected (*Materials and Methods*). The separation of a graph into sending and receiving communities essentially augments the single dimension per component to two dimensions. This allows to distinguish between nodes with high degree imbalance (sources and sinks) and to accurately place them in their respective sending/receiving partition (*SI Appendix, Fig. S6*). Specifically, we observe that only few source and sink nodes are required to create and identify subcommunities that receive or send more (for example followers and followed in a social network). The combination of bimodularity embeddings into bicommunities is able to capture encoded structures and to accurately separate pre-encoded clusters in nonideal examples that mimic real-world networks.

The application of the proposed scheme to the *C. elegans* neuronal wiring diagram further confirmed its relevance. In particular, the embedding according to the first component of the SVD encompasses the organization of the worm's head and body locomotion system and untangles information pathways such as head-to-head, body-to-body, and head-to-body sensory-motor flows. Specifically, this latter communication scheme is mainly driven by the *AVE* neurons that could integrate information from the sensory/control neurons of the head system and further transmit signals to the body control and motor neurons. We also observed that most body motor neurons also send information to more anterior motor neurons in the neck and the head, which confirms previous findings of intricate feedforward loop between the head and body motion systems (36–38). In short,

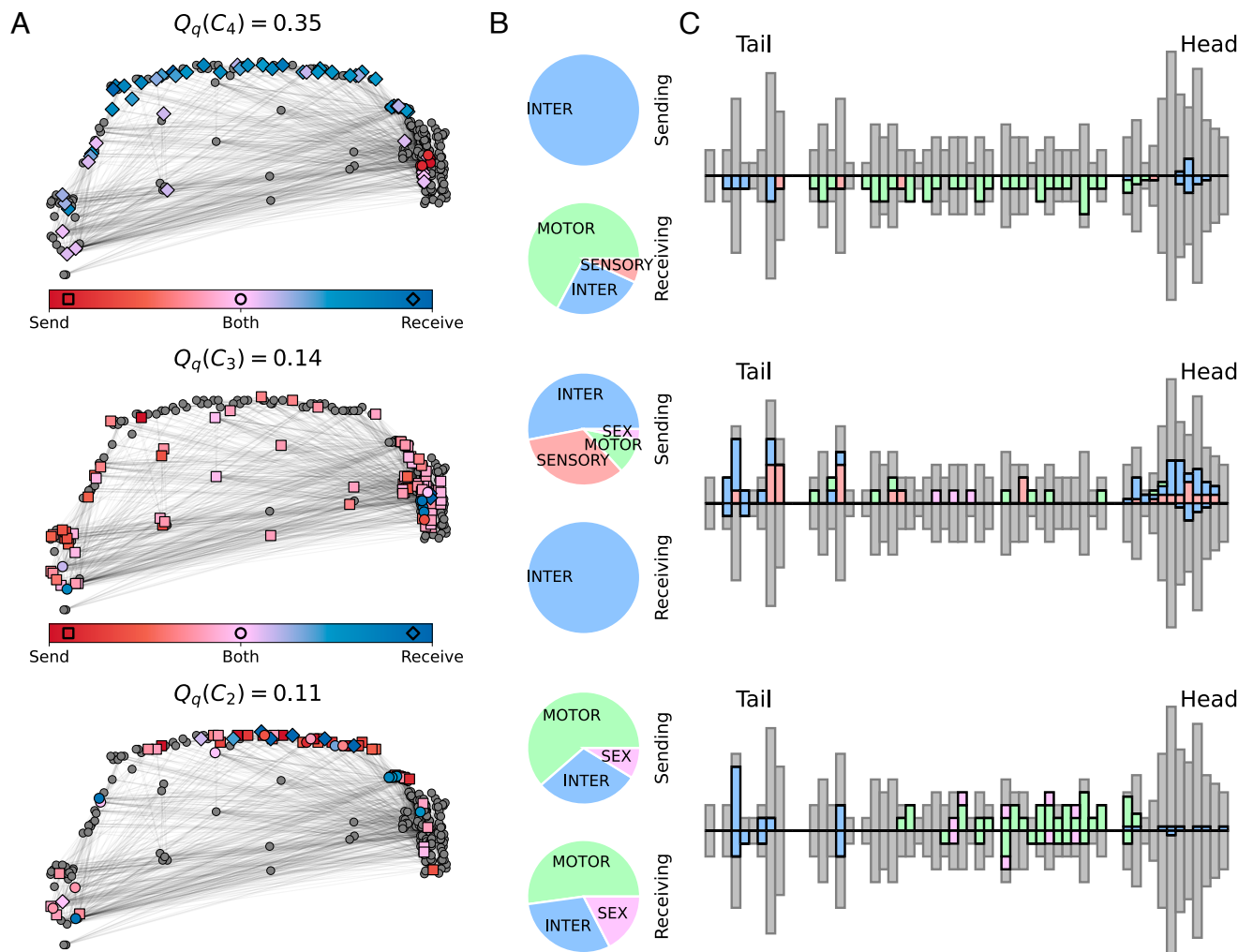


Fig. 4. Bicomunity of the *C. elegans* extracted from the $N = 5$ first bimodularity embeddings (components of the SVD). Only the clusters with top 3 highest bimodularity index (out of $K = 5$) are displayed. (A) The sending and receiving parts of each bicomunity are highlighted on the spatial location of neurons. Markers indicate whether a neuron belongs to the sending (red square), the receiving (blue diamond), or both (pink circle) parts of a bicomunity. The colors represent the difference between the proportion of edges in the sending and in the receiving partitions with red and blue for nodes that tend to send and receive more respectively. (B) The distribution of neuron types (sensory, inter, motor, and sex-specific neurons) in each sending or receiving community is summarized in the pie charts (Middle). (C) The spatial distribution of bicomunities along the tail to head axis is detailed in the histograms (Right) as a proportion of the total number of neurons (gray bars). The sending and receiving patterns are separated in the Upper and Lower bars respectively. Colors indicate the type of neuron following the color scheme of the pie charts. The body and tail histograms are made bigger (twice their original size/number of neurons) for visualization purpose.

the innovative aspect of bimodularity is that key clusters of neurons and their projections are captured in a single bimodular component (i.e., by a left/sending and right/receiving singular vectors) of the directed modularity matrix.

There is one methodological aspect that needs to be addressed when dealing with the SVD; i.e., the sign ambiguity singular values and vectors (39). Indeed, in the $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ decomposition, the sign of a singular value and one singular vector: $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{U}(-\mathbf{\Sigma})(-\mathbf{V}^T)$ can be freely swapped, which raises the need for a clear sign convention for univocal interpretation. In the case of undirected graph, the modularity matrix is symmetric and the singular value (eigenvalue) hints whether a community structure is assortative (positive) or rather disassortative (negative, also known as antimodular) (14, 15, 40). Here, we propose a sign convention associated to the dot product between the two singular vectors (i.e., cosine of the angle between them) and show that it is related to the assortative

(or disassortative) nature of the partitions; see *SI Appendix, Fig. S7*. In particular, we observe that the bimodularity index of a bicomunity is negative when there are more edges between the sending and receiving sets of nodes than within them. Therefore, this sign convention provides additional information about the observed community structure. In the *C. elegans* wiring network, the second component with the largest singular value is disassortative and supports the findings of the first component (Fig. 3) by highlighting head interneurons that densely send information to body motor neurons without necessarily being connected (see *SI Appendix, Fig. S8* for detailed description).

While bimodularity is a suitable metric for highlighting community structure in directed networks, as an extension of modularity it shares some of its limitations such as the resolution limit (41). Future work should investigate how this limit is influenced by directed information. In *SI Appendix, Fig. S9*, we present

a number of variations on the generated graphs that confirm the relevance and robustness of the bimodularity and bicomunity detection. The presented method for bicomunity detection is parameterized by the number of bimodular components N and the number of desired clusters K . While K is here tied to the clustering algorithm (i.e., k-means), we show that influence of N is attenuated by multiplying entries in the edge feature (singular vectors \mathbf{u}_k or \mathbf{v}_k) by their corresponding singular value μ_k (SI Appendix, Fig. S2). This is a common approach in spectral clustering to incorporate the spectrum of eigen-/singular values to weight the matrix of eigen-/singular vectors that is then fed to clustering (30, 31).

To conclude, we provide a theoretical framework based on bimodularity that extends and generalizes the community detection for directed graphs. The joint optimization of the sending and the corresponding receiving communities unveils distinct pathways within networks that would remain invisible to conventional community detection. Therefore, the bimodularity index provides the foundation to better describe community structure in directed graphs and expands the range of applications to networked data.

Materials and Methods

All code and data used in this article are available as open-source code at: <https://github.com/MIPLabCH/Bimodularity>.

Synthetic Graphs. The synthetic graphs built and used in this article (such as the block cycle graph) have binary edge weights generated in a probabilistic manner following the stochastic block model (14). We construct self-community blocks by creating edges with a probability p_{self} and then assigning a direction to them with outgoing probability $p_{\text{dir}} = 0.5$. The result is a block-diagonal matrix of size $n_{\text{blocks}} \times n_{\text{node per block}}$. Then these blocks are connected with the same approach, but by creating edges between self-communities with probability p_{con} and assigning directions based on the desired structure (counterclockwise for the block cycle graph). Such models offered the possibility to freely adapt within- and between-communities edge densities (p_{self} and p_{con} respectively). We emphasize that the model assigns a unique direction to an edge, thus preventing bidirectional relationships. We however demonstrate that the presented approach is valid for networks composed of both uni- and bidirectional edges (as in the *C. elegans*). This constraint is made so that the symmetrized version of such a synthetic graph $\mathbf{A}_{\text{undir}} = \mathbf{A} + \mathbf{A}^T$ will be a binary undirected graph with p_{self} density of edges within communities and p_{con} edge density between communities. This means that if all self-communities are connected with density $p_{\text{con}} = p_{\text{self}}$ the symmetrized graph will be an unstructured random graph.

C. elegans Wiring Network. A directed binary graph was built from the wiring diagram of the *C. elegans*' nervous system using information about 2194 chemical synapses between 279 neurons (graph nodes) (1). Out of all these connections, 1961 (89%) were asymmetric while only 233 (11%) were bidirectional. The undirected electric gap junction network was not used in the main analyses, but the bimodularity optimization and singular vector embedding of the joint (chemical synapses and gap junctions) graph are shown in SI Appendix, Fig. S10. Details about the type of each neuron (sensory, inter-, or motor neurons) and their position in the worm's body have been gathered from refs. 38 and 42 and are summarized in Dataset S1.

Bimodularity Index. The definition of bimodularity is a measure of deviation of the edges configuration (and direction) compared to a null model. Specifically, we leverage the idea that directed graphs may have different community structures when observing incoming or outgoing edges. Hence, graphs with bimodular structure are expected to have more connections from sending communities

to their corresponding receiving communities than expected "on average." We make use of the directed configuration model to express the null probability of having a directed edge between two nodes. We then develop the expression of the bimodularity index:

$$\begin{aligned} Q_{\text{bi}} &= \frac{1}{m} \sum_{k=1}^K \sum_{\substack{i \in C_k^{\text{out}} \\ j \in C_k^{\text{in}}}} [A_{ij} - \mathbb{E}(A_{ij} | \mathcal{H}_0)] \\ &= \frac{1}{m} \sum_{k=1}^K \sum_{\substack{i \in C_k^{\text{out}} \\ j \in C_k^{\text{in}}}} \left[A_{ij} - \frac{k_i^{\text{out}} k_j^{\text{in}}}{m} \right] \\ &= \frac{1}{m} \sum_{i,j} \left[A_{ij} - \frac{k_i^{\text{out}} k_j^{\text{in}}}{m} \right] \delta_{C^{\text{out}}, C^{\text{in}}}(i, j), \end{aligned}$$

where $\delta_{C^{\text{out}}, C^{\text{in}}}$ is 1 if (i, j) runs between corresponding sending and receiving communities, and 0 otherwise.

We consider the graph partition problem to separate the graph into two communities of nodes (sending community) that have common targets (receiving community). Let \mathbf{s}^{out} be a separator vector that takes value $s_i^{\text{out}} = 1$ if node i is in one sending partition and $s_i^{\text{out}} = -1$ if it is in the other. Similarly, we define \mathbf{s}^{in} as a separator vector for the receiving partition corresponding to the sending one. Therefore, $\delta_{C^{\text{out}}, C^{\text{in}}}(i, j)$ can be rewritten as

$$\frac{s_i^{\text{out}} s_j^{\text{in}} + 1}{2}$$

that is 1 if $s_i^{\text{out}} = s_j^{\text{in}}$ and 0 otherwise. We then develop the expression of bimodularity as

$$\begin{aligned} Q_{\text{bi}} &= \frac{1}{m} \sum_{i,j} \left[A_{ij} - \frac{k_i^{\text{out}} k_j^{\text{in}}}{m} \right] \frac{s_i^{\text{out}} s_j^{\text{in}} + 1}{2} \\ &= \frac{1}{2m} \sum_{i,j} \left(\left[A_{ij} - \frac{k_i^{\text{out}} k_j^{\text{in}}}{m} \right] s_i^{\text{out}} s_j^{\text{in}} + \left[A_{ij} - \frac{k_i^{\text{out}} k_j^{\text{in}}}{m} \right] \right). \end{aligned}$$

Knowing that $\sum_{i,j} A_{ij} = m$, that $\sum_i k_i^{\text{out}} = m$ and that $\sum_j k_j^{\text{in}} = m$, the right part of the sum is equal to zero and the bimodularity index can be expressed as a matrix multiplication

$$\begin{aligned} Q_{\text{bi}} &= \frac{1}{2m} \sum_{i,j} \underbrace{\left[A_{ij} - \frac{k_i^{\text{out}} k_j^{\text{in}}}{m} \right]}_{=B_{ij}} s_i^{\text{out}} s_j^{\text{in}} \\ &= \frac{1}{2m} (\mathbf{s}^{\text{out}})^T \mathbf{B} \mathbf{s}^{\text{in}} = Q_{\text{bi}}(\mathbf{s}^{\text{out}}, \mathbf{s}^{\text{in}}). \end{aligned}$$

For the undirected case, the definition of bimodularity falls back to the one of modularity. Indeed, the sending and receiving partitions \mathbf{s}^{out} and \mathbf{s}^{in} are the same when edges have no directions and the number of directed edges m is twice the number of undirected edges $2n = m$:

$$Q_{\text{bi}}(\mathbf{s}^{\text{out}}, \mathbf{s}^{\text{in}}) = \frac{1}{2m} (\mathbf{s}^{\text{out}})^T \mathbf{B} \mathbf{s}^{\text{in}} = \frac{1}{4n} \mathbf{s}^T \mathbf{B} \mathbf{s} = Q(\mathbf{s}).$$

Optimization of Bimodularity. Maximization of bimodularity $Q_{\text{bi}}(\mathbf{s}^{\text{out}}, \mathbf{s}^{\text{in}})$ is achieved by convex relaxation of the separator vectors $\mathbf{s}^{\text{out}}, \mathbf{s}^{\text{in}}$. In essence,

the separator vectors are not anymore restricted to $+1/-1$ but under the constraint to be unitary vectors such that $(\mathbf{s}^{\text{out}})^T \mathbf{s}^{\text{out}} = 1$ and $(\mathbf{s}^{\text{in}})^T \mathbf{s}^{\text{in}} = 1$. We then write the unconstrained optimization problem using Lagrange multipliers μ^{out} and μ^{in} :

$$\max_{\mathbf{s}^{\text{out}}, \mathbf{s}^{\text{in}}} (\mathbf{s}^{\text{out}})^T \mathbf{B} \mathbf{s}^{\text{in}} + \mu^{\text{out}} (1 - (\mathbf{s}^{\text{out}})^T \mathbf{s}^{\text{out}}) + \mu^{\text{in}} (1 - (\mathbf{s}^{\text{in}})^T \mathbf{s}^{\text{in}}).$$

Partial derivatives with respect to \mathbf{s}^{out} and \mathbf{s}^{in} leads to:

$$\begin{cases} (\mathbf{s}^{\text{in}})^T \mathbf{B}^T - \mu^{\text{out}} (\mathbf{s}^{\text{out}})^T = 0 \\ (\mathbf{s}^{\text{out}})^T \mathbf{B} - \mu^{\text{in}} (\mathbf{s}^{\text{in}})^T = 0 \end{cases} \rightarrow \begin{cases} \mu^{\text{out}} \mathbf{s}^{\text{out}} = \mathbf{B} \mathbf{s}^{\text{in}} \\ \mu^{\text{in}} \mathbf{s}^{\text{in}} = \mathbf{B}^T \mathbf{s}^{\text{out}} \end{cases}.$$

Multiplying the first expression by \mathbf{B}^T , we then obtain $\mu^{\text{out}} \mathbf{B}^T \mathbf{s}^{\text{out}} = \mathbf{B}^T \mathbf{B} \mathbf{s}^{\text{in}}$, where we can substitute the second expression to obtain

$$\mu^{\text{out}} \mu^{\text{in}} \mathbf{s}^{\text{in}} = \mathbf{B}^T \mathbf{B} \mathbf{s}^{\text{in}}.$$

Here, we recognize that \mathbf{s}^{in} is an eigenvector of the symmetric matrix $\mathbf{B}^T \mathbf{B}$ with the corresponding eigenvalue $\lambda^{\text{in}} = \mu^{\text{out}} \mu^{\text{in}}$. Similarly, applying the same approach to the second (lower) expression and substituting the upper one, we obtain

$$\mu^{\text{out}} \mu^{\text{in}} \mathbf{s}^{\text{out}} = \mathbf{B} \mathbf{B}^T \mathbf{s}^{\text{out}},$$

which shows that \mathbf{s}^{out} is an eigenvector of $\mathbf{B} \mathbf{B}^T$ with the eigenvalue $\lambda^{\text{out}} = \mu^{\text{out}} \mu^{\text{in}}$.

We observe that the eigenvectors obtained from the eigendecomposition of $\mathbf{B} \mathbf{B}^T$ and $\mathbf{B}^T \mathbf{B}$ are equivalent to the left and right singular vectors of the SVD of \mathbf{B} , respectively, with the eigenvalues λ_i being the squared singular values μ_i^2 . In particular, from the SVD

$$\mathbf{B} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T,$$

we can derive

$$\begin{cases} \mathbf{B} \mathbf{B}^T = \mathbf{U} \mathbf{\Sigma}^T \mathbf{V} \mathbf{\Sigma} \mathbf{U}^T = \mathbf{U} \mathbf{\Sigma}^2 \mathbf{U}^T = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \\ \mathbf{B}^T \mathbf{B} = \mathbf{V} \mathbf{\Sigma} \mathbf{U}^T \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \mathbf{V} \mathbf{\Sigma}^2 \mathbf{V}^T = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \end{cases},$$

meaning that \mathbf{s}^{out} and \mathbf{s}^{in} can be identified as the columns of \mathbf{U} and \mathbf{V} , respectively. Therefore, when choosing $\mathbf{s}^{\text{out}} = \mathbf{u}_i$ and $\mathbf{s}^{\text{in}} = \mathbf{v}_i$ the bimodularity index is proportional to corresponding to the singular value (up to a normalization $\frac{1}{2m}$):

$$Q_{\text{bi}}(\mathbf{u}_i, \mathbf{v}_i) = \frac{1}{2m} \mathbf{u}_i^T \mathbf{B} \mathbf{v}_i = \frac{1}{2m} \mathbf{u}_i^T \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \mathbf{v}_i = \frac{\mu_i}{2m}.$$

Detection of Bicomunity. We present an elegant method to extract a set of sending and their corresponding receiving communities (i.e., bicomunities) by combining $N \geq 1$ components of the SVD of \mathbf{B} . For each graph edge, we aggregate the value of the N leading left singular vector for the source node and the value of the N leading right singular vectors for the target node. The feature vector for edge $\mathcal{E}(i, j)$ from node i to node j thus is

$$\mathbf{f} = (\mu_1 \mathbf{u}_1[i], \mu_1 \mathbf{v}_1[j], \dots, \mu_N \mathbf{u}_N[i], \mu_N \mathbf{v}_N[j]).$$

Clustering edges with such a feature vector will group edges that go from nodes in similar sending partition(s) to nodes in similar receiving partition(s). This reminds the idea of a mapping \mathcal{M} from a sending community C_k^{out} to its corresponding receiving community $C_k^{\text{in}} = \mathcal{M}(C_k^{\text{out}})$. Intuitively, the clustering approach separates edges into a sequence of mappings $\mathcal{M}(C_k^{\text{out}}) = C_k^{\text{in}}$, $k = 1, \dots, K$, from which we can derive the sending and receiving part of each bicomunity. While it is clear that the clusters of edges are not overlapping, the nodes corresponding to the sending/receiving part of a bicomunity can be overlapping.

Data, Materials, and Software Availability. Previously published data were used for this work [All code and data used in this article are openly available at: <https://github.com/MIPLabCH/Bimodularity> (1, 38, 42)]. All other data are included in the manuscript and/or [supporting information](#).

ACKNOWLEDGMENTS. This work was supported in part by the Swiss National Science Foundation, Sinergia project "Precision mapping of electrical brain network dynamics with application to epilepsy," grant No. 209470.

1. L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, D. B. Chklovskii, Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* **7**, e1001066 (2011).
2. A. L. Barabási, N. Gulbahce, J. Loscalzo, Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
3. P. E. Vértés et al., Simple models of human brain functional networks. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5868–5873 (2012).
4. A. Fornito, A. Zalesky, M. Breakspear, The connectomics of brain disorders. *Nat. Rev. Neurosci.* **16**, 159–172 (2015).
5. D. S. Bassett, O. Sporns, Network neuroscience. *Nat. Neurosci.* **20**, 353–364 (2017).
6. M. M. Li, K. Huang, M. Zitnik, Graph representation learning in biomedicine and healthcare. *Nat. Biomed. Eng.* **6**, 1353–1369 (2022).
7. M. Girvan, M. E. J. Newman, Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821–7826 (2002).
8. S. Fortunato, D. Hric, Community detection in networks: A user guide. *Phys. Rep.* **659**, 1–44 (2016).
9. K. Rohe, T. Qin, B. Yu, Co-clustering directed graphs to discover asymmetries and directional communities. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12679–12684 (2016).
10. A. Pathak, S. N. Menon, S. Sinha, A hierarchy index for networks in the brain reveals a complex entangled organizational structure. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2314291121 (2024).
11. E. A. Bender, E. R. Canfield, The asymptotic number of labeled graphs with given degree sequences. *J. Comb. Theory Ser. A* **24**, 296–307 (1978).
12. M. E. J. Newman, The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003).
13. P. W. Holland, K. B. Laskey, S. Leinhardt, Stochastic blockmodels: First steps. *Soc. Netw.* **5**, 109–137 (1983).
14. T. A. Snijders, K. Nowicki, Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classif.* **14**, 75–100 (1997).
15. T. P. Peixoto, "Bayesian stochastic blockmodeling" in *Advances in Network Clustering and Blockmodeling*, P. Doreian, V. Batagelj, A. Ferligoj, Eds. (John Wiley & Sons, Ltd.) (2019), pp. 289–332, <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119483298.ch11>.
16. V. D. Blondel, J. L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
17. R. Lambiotte, P. Panzarasa, Communities, knowledge creation, and information diffusion. *J. Informet.* **3**, 180–190 (2009).
18. M. E. J. Newman, Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8577–8582 (2006).
19. M. E. J. Newman, Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **74**, 036104 (2006).
20. T. S. Evans, R. Lambiotte, Line graphs, link partitions, and overlapping communities. *Phys. Rev. E* **80**, 016105 (2009).
21. Y. Y. Ahn, J. P. Bagrow, S. Lehmann, Link communities reveal multiscale complexity in networks. *Nature* **466**, 761–764 (2010).
22. A. Arenas, J. Duch, A. Fernández, S. Gómez, Size reduction of complex networks preserving modularity. *New J. Phys.* **9**, 176 (2007).
23. E. A. Leicht, M. E. J. Newman, Community structure in directed networks. *Phys. Rev. Lett.* **100**, 118703 (2008).
24. F. D. Malliaros, M. Vazirgiannis, Clustering and community detection in directed networks: A survey. *Phys. Rep.* **533**, 95–142 (2013).
25. Y. Kim, S. W. Son, H. Jeong, Finding communities in directed networks. *Phys. Rev. E* **81**, 016103 (2010).
26. M. Hosseini-Pozveh, M. Ghorbanian, M. Tabaiyan, A label propagation-based method for community detection in directed signed social networks. *Phys. A Stat. Mech. Appl.* **604**, 127875 (2022).
27. T. D. Dang, D. H. Do, T. H. D. Phan, Community detection in directed graphs using stationary distribution and hitting times methods. *Soc. Netw. Anal. Min.* **13**, 80 (2023).
28. Z. Wang, Y. Liang, P. Ji, Spectral algorithms for community detection in directed networks. *J. Mach. Learn. Res.* **21**, 1–45 (2020).
29. D. L. Sussman, M. Tang, D. E. Fishkind, C. E. Priebe, A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Am. Stat. Assoc.* **107**, 1119–1128 (2012), <https://doi.org/10.1080/01621459.2012.699795>.
30. U. von Luxburg, A tutorial on spectral clustering. *Stat. Comput.* **17**, 395–416 (2007).
31. A. Ng, M. Jordan, Y. Weiss, "On spectral clustering: Analysis and an algorithm" in *Advances in Neural Information Processing Systems*, T. Dietterich, S. Becker, Z. Ghahramani, Eds. (MIT Press, 2001), vol. 14.
32. Z. Chen, X. Li, J. Bruna, Supervised community detection with line graph neural networks. *arXiv [Preprint]* (2020), <https://arxiv.org/abs/1705.08415v6> (Accessed 13 June 2025).
33. X. Zhang et al., "MagNet: A neural network for directed graphs" in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, J. W. Vaughan, Eds. (Curran Associates, Inc., 2021), vol. 34, pp. 27003–27015.
34. M. Khosla et al., "Node representation learning for directed graphs" in *Machine Learning and Knowledge Discovery in Databases*, U. Brefeld et al., Eds. (Springer International Publishing, Cham, 2020), pp. 395–411.
35. S. Furutani, T. Shibahara, M. Akiyama, K. Hato, M. Aida, "Graph signal processing for directed graphs based on the Hermitian Laplacian" in *Machine Learning and Knowledge Discovery in Databases*, U. Brefeld et al., Eds. (Springer International Publishing, Cham, 2020), pp. 447–463.

36. F. Morone, H. A. Makse, Symmetry group factorization reveals the structure-function relation in the neural connectome of *Caenorhabditis elegans*. *Nat. Commun.* **10**, 4961 (2019).
37. Y. Sohn, M. K. Choi, Y. Y. Ahn, J. Lee, J. Jeong, Topological cluster analysis reveals the systemic organization of the *Caenorhabditis elegans* connectome. *PLoS Comput. Biol.* **7**, e1001139 (2011).
38. S. J. Cook *et al.*, Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature* **571**, 63–71 (2019).
39. R. Bro, E. Acar, T. G. Kolda, Resolving the sign ambiguity in the singular value decomposition. *J. Chemom.* **22**, 135–140 (2008), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/cem.1122>.
40. L. Zhang, T. P. Peixoto, Statistical inference of assortative community structures. *Phys. Rev. Res.* **2**, 043271 (2020).
41. S. Fortunato, M. Barthélemy, Resolution limit in community detection. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 36–41 (2007).
42. M. Kaiser, C. C. Hilgetag, Nonoptimal Component Placement, but Short Processing Paths, due to Long-Distance Projections in Neural Systems. *PLoS Comput. Biol.* **2**, e95 (2006).