



SHEEP, a Signed Hamiltonian Eigenvector Embedding for Proximity

Shazia'Ayn Babul ¹✉ & Renaud Lambiotte ^{1,2}

Signed network embedding methods allow for a low-dimensional representation of nodes and primarily focus on partitioning the graph into clusters, hence losing information on continuous node attributes. Here, we introduce a spectral embedding algorithm for understanding proximal relationships between nodes in signed graphs, where edges can take either positive or negative weights. Inspired by a physical model, we construct our embedding as the minimum energy configuration of a Hamiltonian dependent on the distance between nodes and locate the optimal embedding dimension. We show through a series of experiments on synthetic and empirical networks, that our method (SHEEP) can recover continuous node attributes showcasing its main advantages: re-configurability into a computationally efficient eigenvector problem, retrieval of ground state energy which can be used as a statistical test for the presence of strong balance, and measure of node extremism, computed as the distance to the origin in the optimal embedding.

¹Mathematical Institute, University of Oxford, Oxford, UK. ²Alan Turing Institute, London, UK. ✉email: shazia.babul@new.ox.ac.uk

Networks provide a powerful representation of complex systems, comprising of a collection of nodes joined by edges that represent interactions¹. Typical examples stem from an array of disciplines, including sociology (social contact networks) urban infrastructures (public transportation networks), or biological interactions (gene interaction networks). Edges represent diverse types of interactions, which can be distinguished by attributes on the edges, such as edge weights, to identify the nature and intensity of the relationship. In this paper, we consider signed networks, where the edges can have either positive or negative weights. A fundamental example is a social network, where positive edges represent friendship and negative edges represent animosity between individuals.

Despite their classical origin², signed networks represent a relatively new field of exploration. However, research into signed networks has been increasing in popularity due to a number of examples from online social networks such as Epinions³ and Slashdot⁴ where users can mark each other as friend or foe. More recently, signed social networks have been constructed by inferring the signed signals using sentiment analysis, using data from social media platforms like Reddit⁵ and Twitter⁶. We give an overview of the related literature on signed networks in the following sections, and direct the reader to⁷ for a more detailed explanation of their properties and principles.

In general, networks are highly complex, and the structure of signed (and unsigned) networks are often investigated through dimensionality-reduction techniques called network embedding. Network embedding refers to a method for obtaining a low-dimensional representation of nodes, taking into account multi-scale graph topology. Embedding methods have a range of applications from clustering to semi-supervised node classification, and are used in recommender systems to make recommendations based on proximity in a latent (embedding) space^{8,9}. For signed networks, embedding methods typically seek to place nodes close together in latent space if they are connected by positive edges and paths, while simultaneously separating nodes connected by negative edges^{10–14}.

In this article, we present a physically inspired signed network embedding method called SHEEP (Signed Hamiltonian Eigenvector Embedding for Proximity), which incorporates local, intermediate and global information into a continuous proximity measure between nodes. Modeling the nodes as a system of particles interacting with attractive forces (positive edges) and repulsive forces (negative edges), we construct a Hamiltonian which is dependant on the distance between nodes, such that the relative embedding distance results in a similarity metric. Unlike embedding methods which focus on partitioning the graph into clusters, and are useful for binary prediction, our measure is intended to retain proximal information between nodes. We propose two applications of the signed network embedding method; (1) a way to recover continuous node attributes based on their proximity in the embedding, and (2) using the node's distance to the origin as a measure of extremism or conflictuality, evaluating the method on both synthetic generated signed networks and empirical networks. In addition, the Hamiltonian admits a global minimum energy configuration, which can be re-configured as an eigenvector problem, and therefore is computationally efficient to compute. We also present an energy-based approach to finding the optimal embedding dimension for proximity. At the global level, the embedding generated by the minimum energy configuration is intrinsically related to the notion of structural balance, i.e., the existence of stable and unstable types of cycles in a signed graph, and the ground state energy can be used as a statistical test for bi-polarization.

Results and discussion

Notation. For a matrix $M \in \mathbb{R}^{n \times n}$ we denote the eigenvalues $\lambda_i(M)$ with associated eigenvectors $v_i(M)$. A symmetric matrix M has an eigenvalue ordering $\lambda_1(M) \geq \lambda_2(M) \dots \geq \lambda_n(M)$ with associated eigenvectors. We denote $\mathbf{1}$ as the all ones vector. The graph $G = (V, E)$ is the signed graph with node set V and edges E , with E^+ being the set of positive signed edges, and E^- being the set of negative signed edges, such that $E^+ \cup E^- = E$ and $E^+ \cap E^- = \emptyset$.

If A is the adjacency matrix of graph G , then $A_{ij} = 0, \forall i \in V$. In what follows, we focus on unweighted signed networks. For an edge $(i, j) \in E^+$, it follows that $A_{ij} = 1$ and likewise, an edge $(i, j) \in E^-$ corresponds to the matrix entry $A_{ij} = -1$. The positive adjacency matrix A^+ is defined by $A_{ij}^+ = 1$ if $A_{ij} = 1$, and 0 elsewhere. Similarly, the negative adjacency matrix A_{ij}^- is defined by $A_{ij}^- = -1$ if $A_{ij} = -1$, and 0 elsewhere, such that $A = A^+ + A^-$. We also have that $D^+ = \text{diag}(deg_1^+, deg_2^+ \dots deg_n^+)$, where $deg_i^+ = \sum_j A_{ij}^+$, the degree matrix of the positive sub-graph, and $D^- = \text{diag}(deg_1^-, deg_2^- \dots deg_n^-)$, where $deg_i^- = -\sum_j A_{ij}^-$, the degree matrix of the negative sub-graph, where in particular we note that $deg_i^- \geq 0$. We also use the convention that an n -complete graph is a complete graph with n nodes, such that every possible edge exists. In this work, we are looking at the possibility to associate each node i to a position in k dimensional space, which is encoded by the $n \times k$ matrix X with entries $X_{i\alpha}$ or, equivalently, the set of n vectors x_i . For further clarity, we use Latin indices for nodes, and Greek indices for dimensions. The literature does not have a consistent naming system for the different Laplacians we refer to here. What we call the opposing Laplacian after¹⁵ is also known as the signed Laplacian as in¹⁰. The opposing Laplacian is defined as $L_o = \bar{D} - A$ where \bar{D} is a diagonal matrix with entries $\bar{D}_{ii} = \sum_j |A_{ij}|$. Following¹⁵ we also consider the repelling Laplacian $L_r = D - A$ where D is a diagonal matrix defined $D_{ii} = \sum_j A_{ij}$. It is also known as the physics Laplacian in¹³, the net Laplacian in¹⁶, or simply the unsigned Laplacian due to its identical construction to the Laplacian on unsigned graphs.

Structural balance. Signed networks are most commonly understood through Heider's psychological theory of structural balance, developed to explain why "balanced" structures are more favorable in interpersonal relationships, following the adage *an enemy of my enemy is my friend* and that a friend of a friend is also a friend². Cartwright and Harary use the term *strong balance* to refer to a signed network where all cycles contain an even number of negative edges. They show this is equivalent *bi-polarization*, when a graph can be divided into two groups with positive edges inside, and negative edges connecting them^{17,18}. In¹⁹, Davis defines *weakly balanced* graphs as graphs where no cycle has only a single negative edge, proving that such graphs can be partitioned into k clusters with positive edges inside, and negative edges connecting them. Identifying the optimal partition of a signed graph into k clusters, where k is unknown, has been shown to be NP hard and various methods have been proposed to solve this problem numerically^{20,21}. A popular approach takes inspiration from spin glass literature, using simulated annealing to locate the optimal cluster assignment^{22,23}. Figure 1 shows examples strong balanced, weak balanced and unbalanced cycles, while Fig. 2 depicts a strongly balanced (2 faction) graph and a weakly balanced (3 faction) graph. These conditions are deeply related to the Ising spin glass model, where negative plaquettes introduce geometric frustration, analogous to the unbalanced k -cycles on signed graphs²⁴.

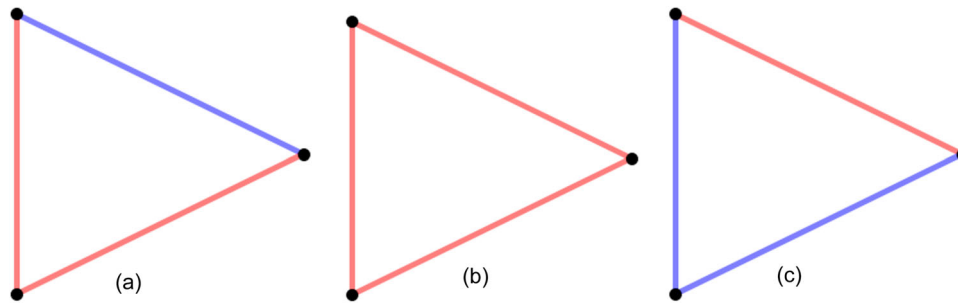


Fig. 1 Examples of balanced and unbalanced cycles. Strongly balanced cycles (a) have even numbers of negative edges. Weakly balanced cycles (b) have i negative edges, where $i \neq 1$. Unbalanced cycles (c) have a single negative edge. Red indicates a negative edge, while blue indicates a positive edge.

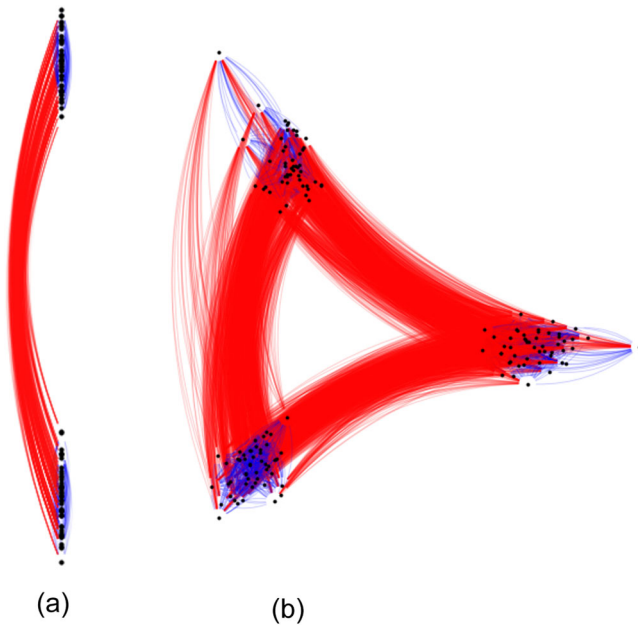


Fig. 2 Strong and weakly balanced signed networks. Strongly balanced signed networks (a) have two opposing factions with positive edges inside and negative edges between. Weakly balanced networks (b) can have three or more factions. Red indicates a negative edge, while blue indicates a positive edge.

While structural balance provides a valuable framework for understanding signed networks, it has been argued in the literature that empirical networks are often not perfectly balanced, and the tendency for signed social networks to become balanced may not even hold over time²⁵. In particular, signed networks may not have a natural cluster structure, and as proposed in²¹, their formation may be influenced by a combination of other processes, resulting in different patterns than structure balance alone. It is precisely when ground truth clusters do not exist that it may be more useful to consider the proximity measure provided by SHEEP, eliminating the need to make assumptions about faction numbers and memberships.

Spring/anti-spring signed graph Hamiltonian. Several network embedding methods are based on the minimization of a Hamiltonian associated to a network (See Supplementary Note 1 for an overview of spectral and energy-based embedding methods). Here, we draw inspiration from SpringRank, a physically inspired model generating hierarchical rankings of nodes for directed, unsigned networks²⁶, which we adapt to a signed setting. Suppose each node i has an associated position vector \mathbf{x}_i in metric space \mathbb{R}^k . Positive edges are modeled as spring attractive forces,

and negative edges are associated with an anti-spring repulsive force, which is similarly quadratic in distance, resulting in:

$$H = \sum_{ij} A_{ij}^+ |\mathbf{x}_i - \mathbf{x}_j|^2 + \sum_{ij} A_{ij}^- |\mathbf{x}_i - \mathbf{x}_j|^2. \quad (1)$$

The first term of the Hamiltonian is minimized when positively connected nodes have minimal distance between them, while the second term requires negatively connected nodes to have the largest possible distance between them. We seek to find the set of position vectors $\{\mathbf{x}_i\}$ that describe the minimum energy configuration. When a negative edge exists between two nodes, their interaction term may be minimized when the two nodes are pushed apart to infinity. Consequently, we introduce a constraint on Eq. (1) to control for the explosion of distance. As we will see, the optimal embedding defined by the position vectors $\{\mathbf{x}_i\}$ does not only depend on the number of positive and negative connections, but also on their specific arrangement between the nodes and on the constraints induced by the dimension k of the embedding.

Let us first consider a one-dimensional form where each node i is associated with a position π_i , such that $\boldsymbol{\pi} \in \mathbb{R}^n$. Then, our objective function to minimize is:

$$H(\boldsymbol{\pi}) = \sum_{ij} A_{ij}^+ |\pi_i - \pi_j|^2 + \sum_{ij} A_{ij}^- |\pi_i - \pi_j|^2, \quad (2)$$

where, $H(\boldsymbol{\pi}) = \boldsymbol{\pi}^T L \boldsymbol{\pi}$, the quadratic form of the repelling Laplacian (see Methods). As the minimization of Eq. (2) is computationally difficult, we follow a method proposed in²⁷ for ordering nodes in graphs, and impose the spherical constraint condition that $\sum \pi_i^2 = R$. This constraint can be written using a Lagrange multiplier, as:

$$\mathcal{L}(\boldsymbol{\pi}, \lambda) = H(\boldsymbol{\pi}) - \beta (\sum \pi_i^2 - R), \quad (3)$$

where β is the Lagrange multiplier. Taking the gradient and setting it to zero gives the following eigenvalue equation for the extrema:

$$L_r \boldsymbol{\pi} = \beta \boldsymbol{\pi}. \quad (4)$$

The extrema conditions generated by the constraint requires that $\boldsymbol{\pi}$ is an eigenvector of the repelling Laplacian, \mathbf{v} , which gives:

$$H(\mathbf{v}) = \mathbf{v}^T L_r \mathbf{v} = \lambda R, \quad (5)$$

where λ is the associated eigenvalue, and the eigenvector has been normalized such that $\mathbf{v}^T \mathbf{v} = R$.

From Eq. (5), we see that the minimum is obtained when \mathbf{v} is the eigenvector \mathbf{v}_n associated to the minimum eigenvalue λ_n . Finally, we have that minimum value of the Hamiltonian subject to the spherical constraint is the minimum eigenvalue of the repelling Laplacian. This minimum is a global minimum for the system defined in Eq. (3) with the spherical constraint, and the associated one-dimensional embedding is given by the eigenvector \mathbf{v}_n . The resulting embedding is invariant under reflection

about the origin. Furthermore, since we are most interested in the relative distances between nodes in this proximity-based embedding, the resulting normalized distance matrix generated from the embedding is also invariant under translations and re-scaling of the vector \mathbf{v}_n . In what follows we take $R = 1$.

In the Methods section, we provide a more thorough discussion of the one-dimensional case, its relation to strong balance, and show how the minimum eigenvalue can be statistical measure for the bi-polarization of the system, i.e., one faction on the left and one faction on the right. Note that unlike the other force based approaches which often employ non-convex optimization to locate local minima, finding the minimal value of the Hamiltonian with the spherical constraint reduces to a computationally efficient eigenvalue problem.

In the method proposed here, we only consider signed graphs which have at least one negative eigenvalue. For positive graphs, the repelling Laplacian is equivalent to the standard graph Laplacian. Since this matrix always has zero row sum, it has a 0 eigenvalue associated to the eigenvector $\mathbf{1}$, which is the smallest eigenvalue for the positive graph. From the physical interpretation of the system, this means the attractive forces have collapsed all the nodes onto a single point. It is also possible for signed networks to have no negative eigenvalues, for example if the positive sub-graph is fully connected²⁸. We consider only graphs that have at least one negative eigenvalue, which occurs when the negative edges are sufficiently dense or well positioned in the graph such that the proximity embedding information is non-trivial. We refer the reader to the Supplementary Note 2 for a more detailed explanation of this argument and a numerical experiment.

Weak balance and extension to higher dimensions. In a one-dimensional embedding, we can understand graphs in terms of strong balance, because the bi-partition can be projected along the one dimensional line. In contrast, weakly balanced graphs are k -clusterable, meaning they can be partitioned into k factions. This also implies that representing the relationships between nodes may not be possible using a one dimensional projection. For noisy and unbalanced graphs, this is even more important; more dimensions may be required to ensure that the distance between nodes in the embedding represents a proximity measure that takes into account the global graph structure.

When considering embedding in higher dimensions, finding the optimal dimension of the embedding is a challenging task. In the extreme case of a complete negative graph, intuition suggests that an optimal embedding should place these n nodes at equidistant positions, which requires a $n - 1$ -dimensional embedding (a detailed analysis of this case is discussed in Supplementary Note 5). To investigate the general case of non-complete graphs, we define, for each node i , a k -dimensional vector \mathbf{x}_i , which is the vector built using the i -th entry of each of the first k eigenvectors, i.e. $\mathbf{x}_i = (\nu_{ni}, \nu_{n-1i}, \dots, \nu_{(n-k+1)i})$. Let the distance matrix $D(k)$ correspond to the distances associated to the positions of the nodes found using the first k eigenvectors, such that:

$$D(k)_{ij} = |\mathbf{x}_i - \mathbf{x}_j|^2 = \sum_{\alpha=n}^{n-k} (\nu_{\alpha i} - \nu_{\alpha j})^2, \quad (6)$$

where $\nu_{\alpha i}$ is i -th entry of the α -th eigenvector of the repelling Laplacian. This implies that the distance matrix $D(k)$ is linear in dimension in the sense that if $D(\mathbf{v}_\alpha)_{ij} = (\nu_{\alpha i} - \nu_{\alpha j})^2$ are the entries of the distance matrix associated with the α -th

eigenvector, then:

$$D(k) = \sum_{\alpha=n}^{n-k} D(\mathbf{v}_\alpha). \quad (7)$$

Since the Hamiltonian in Eq. (1) can be written in terms of the distance matrix $D(k)$:

$$H(D(k)) = \sum_{ij} A_{ij}^+ D(k)_{ij} + \sum_{ij} A_{ij}^- D(k)_{ij}, \quad (8)$$

where in one dimension,

$$H(D(1)) = \sum_{ij} A_{ij}^+ D(\mathbf{v}_n)_{ij} + \sum_{ij} A_{ij}^- D(\mathbf{v}_n)_{ij} = \lambda_n, \quad (9)$$

it follows that the Hamiltonian is similarly linear in dimension:

$$H(D(k)) = \sum_{\alpha=n}^{n-k} H(D(\mathbf{v}_\alpha)). \quad (10)$$

With the spherical constraint given in Eq. (5), we have that $H(D(\mathbf{v}_\alpha))$ is simply λ_α , the eigenvalue associated with the α -th eigenvector. Then:

$$H(D(k)) = \sum_{\alpha=n}^{n-k} \lambda_\alpha. \quad (11)$$

As a consequence of linearity, the Hamiltonian as a function of distance trivially decreases with an increase in dimension as long as $\lambda_\alpha \leq 0$. Our experiments on artificial networks and real-life networks show that the number of negative eigenvalues is usually relatively large, in any case too large for an efficient embedding of the signed network. At the same time, one can observe that each new dimension allows the nodes to be separated by greater distances, as the norm of the distance matrix will grow with each added dimension. To compare the quality of embeddings at different dimensions, a natural choice is to combine these two quantities, and to normalize the Hamiltonian by the norm of the distance matrix, i.e., dividing by the term $\sqrt{(\sum_{ij} D(k)_{ij})^2}$. This procedure defines the higher dimensional generalized Hamiltonian:

$$\tilde{H}(D(k)) = \frac{\sum_{\alpha=n}^{n-k} \lambda_\alpha}{\sqrt{\sum_{ij} D(k)_{ij}^2}}. \quad (12)$$

Note that this choice is a heuristic, which we show to work well in practice in the following.

Based on these arguments, and on our understanding of the n -complete negative graph, we propose to minimize the higher dimensional generalized Hamiltonian given in Eq. (12) as a function of k , thus finding the embedding dimension that minimizes the ground state energy. Note that even if $\lambda_{k+1} < 0$, the addition of the $k + 1$ -th dimension may still increase the energy $\tilde{H}(D(k+1)) > \tilde{H}(D(k))$ due to the normalization factor. This is an alternative to the idea proposed in¹² of using the k smallest eigenvectors for the embedding, where k is chosen by looking for the largest eigen-gap, but formalizes this argument in term of a physical energy function. Let us emphasize again that, within our framework, the energy minimization is intended to find the dimension that produces the best proximity measure, and not the number of weak balance clusters (see Supplementary Note 5).

We test our results using signed stochastic block models (SSBMs) which are synthetic graphs of n communities with positive edges inside the communities, and negative edges between. Figure 3 shows the normalized energy versus dimension plots for various realizations of SSBMs with different community numbers, sizes and sign flip probabilities. Figure 3a shows the normalized energy as a function of dimension for three community SSBM with edge probability 0.5, and randomly

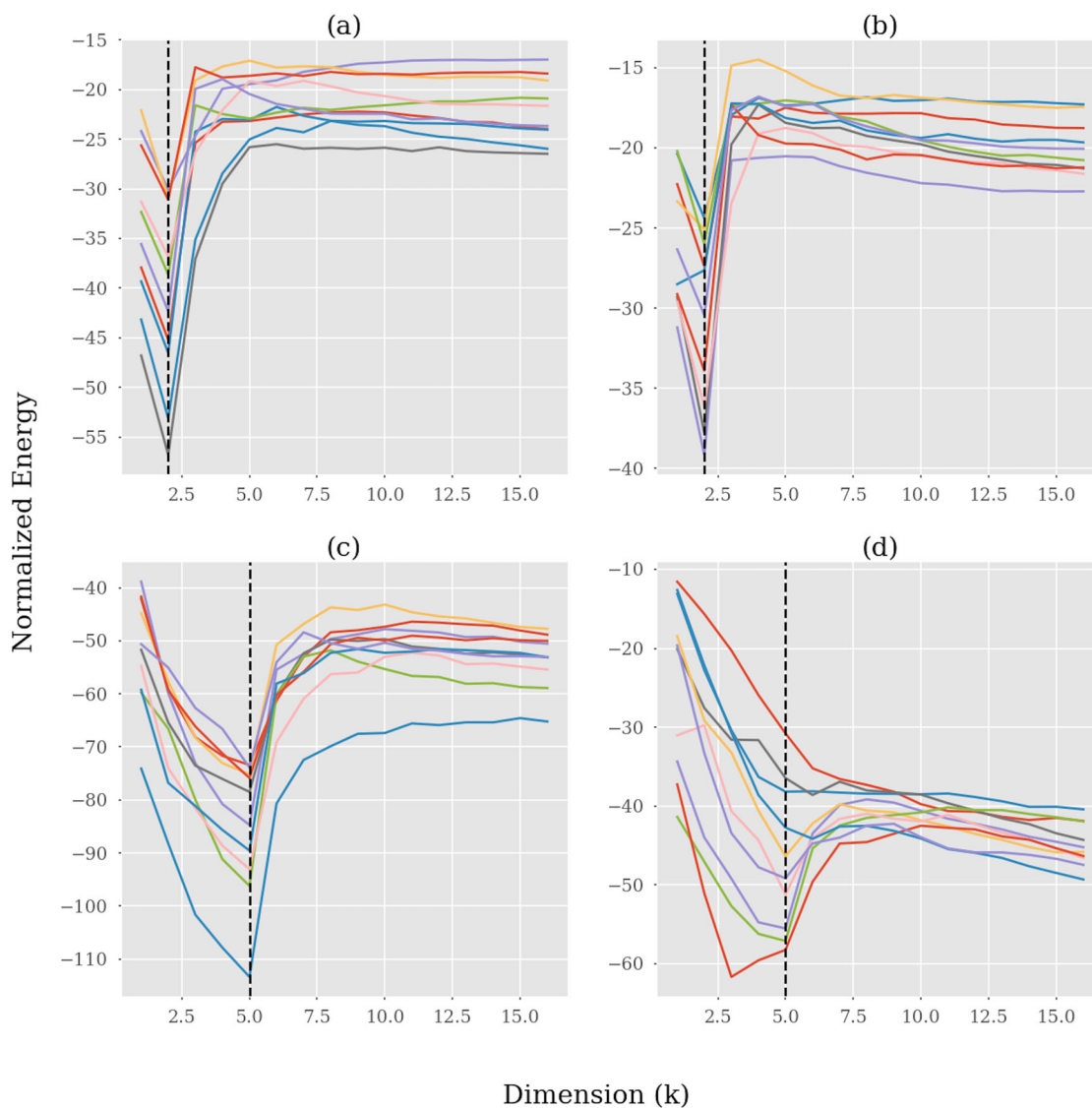


Fig. 3 Normalized energy vs dimension for various Signed Stochastic Block Models (SSBMs). Normalized Energy $\tilde{H}(D(k))$ vs dimension k for 10 realizations of different SSBMs. Each line color indicates a different SSBM realization, dashed line shows the energy minima. **a** 3 Community SSBM. The energy minimum is obtained at $k=2$ as expected. **b** Same as in (a), but with 0.1 probability of sign flip. The energy minimum is still obtained at $k=2$. **c** 6 Community SSBM. The energy minimum is obtained at $k=5$ as expected. **d** Same as in (c) with 0.1 probability of sign flip. The energy minimum is still obtained at $k=5$, for most realizations, but can vary depending on realization faction sizes and random variations.

generated community sizes between 20 and 50. For each of the ten realizations, the energy is minimized at $k=2$, the expected dimension for representing 3 communities. Likewise, in Fig. 3b, the energy versus dimension plots for the three community SSBMs with noise (0.1 probability of edge sign flip), shows the minimum is still obtained at $k=2$. Figure 4 shows a two-dimensional SHEEP embedding for two realizations of a three community SSBM, with and without noise on the edge signs. When the number of SSBM communities is increased to 6 as in Fig. 3c, the minimum energy is obtained at dimension $k=5$ for each of the ten realizations. For the noisy six community SSBMs show in Fig. 3d, the energy minimum is obtained at $k=5$ for most realizations, but can vary depending on the faction size and noise. When the probability of sign flip increases, the optimal dimension found will no longer be equal to the number of clusters. This is because our method is designed for finding the best embedding to represent proximity, not for clustering, and the noise will naturally effect the proximal relationships between

nodes. We explore the effect of noise on the optimal dimension in Supplementary Note 6.

Numerical experiments I: proximity-based node attributes.

Since our constructed Hamiltonian is a function of distance, SHEEP is designed to represent node proximity, taking into account both local interactions (link sign) and global structure (factions from structural balance). Here, we build on this insight to show how SHEEP can be used to recover continuous node attributes. We emphasize that embedding methods designed for clustering are particularly adapt at predicting binary variables (edge signs), while SHEEP provides a continuous metric for understanding node relationships and relative extremism. We compare our method to two other spectral embedding methods for signed networks, the opposing Laplacian as defined in ref. ¹⁰ and SPONGE (Signed Positive Over Negative Generalized Eigenproblem) ¹¹. Both of these methods introduce a matrix

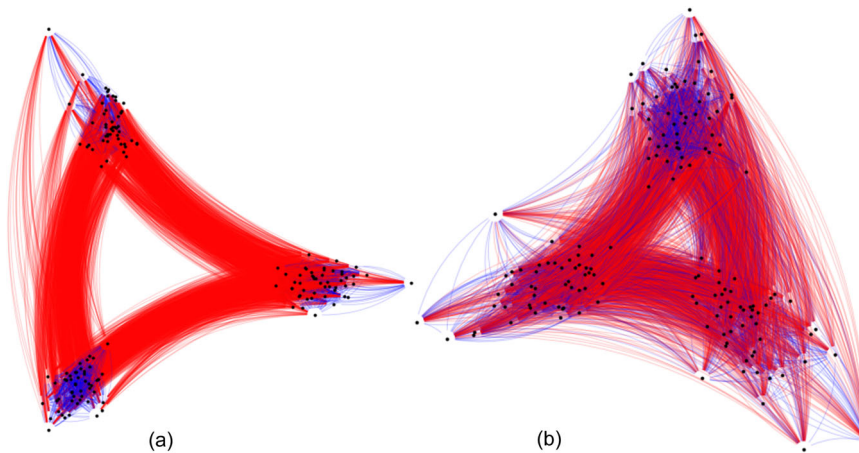


Fig. 4 SHEEP embedding of 3-community SSBMs. 3 Community SSBMs with 50 node communities and 0.5 edge probability, embedded using SHEEP (first two eigenvectors of repelling Laplacian). No edge sign flips (a) and 0.2 probability of sign flip (b). Red indicates a negative edge, while blue indicates a positive edge.

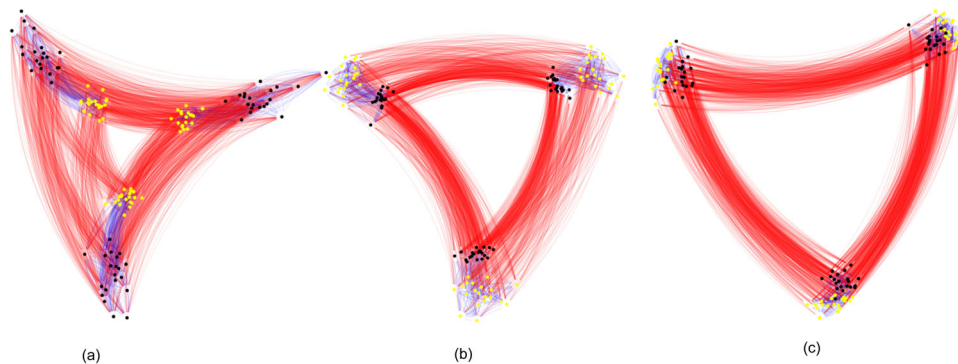


Fig. 5 Proximity vs clustering: embedding examples. Resulting embedding of grouped signed stochastic block model with 3-way partition, which obeys weak balance. The yellow nodes have fewer negative edges compared to the black nodes. SHEEP/Repelling Laplacian (a), opposing Laplacian (b), SPONGE (c). Red indicates a negative edge, while blue indicates a positive edge.

operator, and use the k smallest eigenvectors to embed the graph. Since these methods are designed for finding factions, the dimension k is chosen first by selecting the number of expected factions, $k + 1$. These factions are recovered by performing a k -means clustering on the embedding.

Figure 5 provides an illuminating example of the distinction between the proximity-based SHEEP embedding method and the two spectral methods designed for clustering. The figure shows various embeddings for a modified signed stochastic block model which obeys weak balance. Each embedding has been normalized such that the norm of both the x and y embedding vectors are equal to 1, which allows for visual comparison between the methods. Each signed community in the SSBM is separated into two groups of nodes: the yellow nodes have a lower density of negative edges to the other communities ($p = 0.3$), as compared to the black nodes which have a higher density of negative edges ($p = 0.7$). The two spectral methods defined for clustering, the opposing Laplacian and SPONGE, produce embeddings that can easily recover the 3 factions. In the embedding produced by SHEEP, the yellow nodes appear closer to the origin, indicating the weaker negative relationships between these groups, as compared to the black nodes. In the cases of the opposing Laplacian and SPONGE, the separation between the yellow and black nodes is smaller, and the embeddings actually place the less conflictual yellow nodes further away from the other factions, as compared to the black nodes. This indicates that the SHEEP embedding is better at representing the strength of node

relationships, taking into account the presence (or absence) of edges, as compared to the other two methods. See Supplementary Note 7 for further visual examples. In the following section, we refine this argument by quantitatively evaluating these methods' performance on both synthetic generated and empirical signed networks.

Rankings on generated synthetic graphs. Since the SHEEP Hamiltonian is a function of distance, the sign of an edge between two nodes is related to the distance between them in the resulting embedding. In one dimension, the positions generated by the SHEEP embedding resembles an ordering of the nodes, where nodes that are closer together in the ordering have more positive connections. Here, we present an experiment designed to test the ability of SHEEP to recover ordinal information, using a synthetic network where the edge signs correspond to a well-defined node order. This well-ordered synthetic graph is constructed such that nodes are assigned a random position from -1 to 1 , and connected by a positive (negative) edge if the difference in their position is smaller (larger) than a chosen threshold (see Methods).

Figure 6a shows the initial positions of 50 nodes, and the resulting edges signs with threshold of 0.2. We generate the one-dimensional SHEEP embedding using the first eigenvector of the repelling Laplacian, and compare the resulting node ordering to the initial positions by taking the Kendall Tau correlation, denoted as KT correlation from now on. Since multiplying the

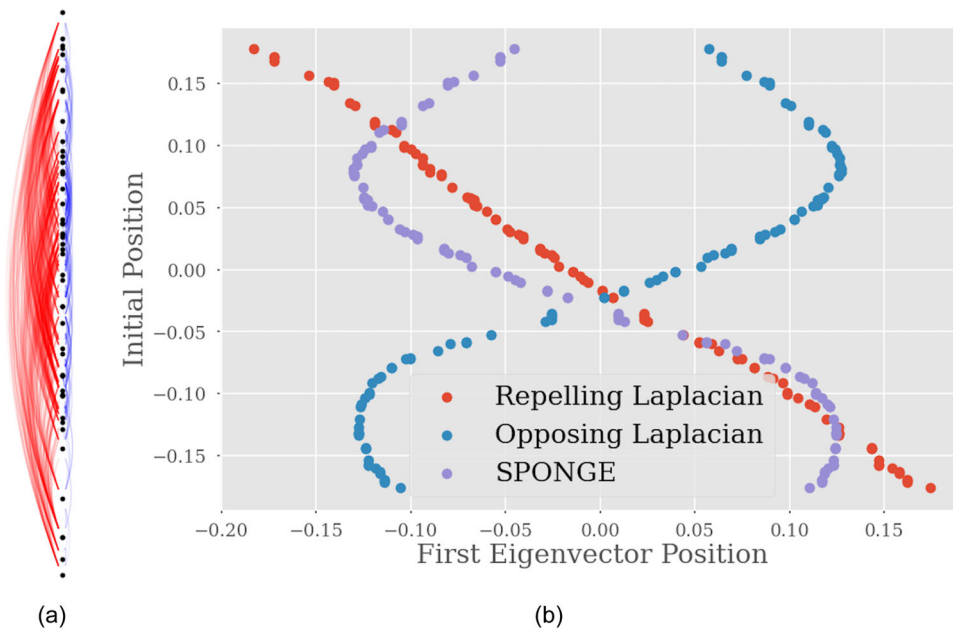


Fig. 6 Well-ordered synthetic graph embedding. Positions of 50 nodes with edge signs assigned according to threshold of 0.1 Red indicates a negative edge, while blue indicates a positive edge. **a** Generated node position (initial position) plotted against first eigenvector position from SHEEP/repelling Laplacian (red dots), the opposing Laplacian (blue dots) and SPONGE (purple dots) for a 100 node graph with threshold 0.1 **(b)**. There is a high correlation between the initial and SHEEP embedding position.

eigenvector by -1 returns the same eigenvalue, but with opposite node ordering, we take the absolute value of the KT correlation. We also compare SHEEP to the opposing Laplacian and SPONGE, two spectral methods designed for clustering, finding that our method is better at recovering ordinal information. Figure 6b shows a plot of the generated node position versus the position determined by the first eigenvector for a well-ordered synthetic graph with 100 nodes and threshold = 0.1. Unlike the opposing Laplacian and SPONGE, the SHEEP embedding recovers the initial node ordering. We test this over several realizations of generated well-ordered synthetic graph with various parameters (see Supplementary Note 8), finding that SHEEP is highly successful at recovering the resulting node ordering, unlike the other two spectral methods, demonstrating the usefulness of our proximity-based embedding measure.

Australian rainfall correlations. We test the method on an empirical graph with a ground-truth proximity measure. Using a time series of seasonal rainfalls across 305 different stations in Australia, we construct signed network from the Pearson correlation matrix (see Methods for details). Using the higher dimensional generalized Hamiltonian method, we find that the best embedding dimension is $k = 1$. Intuitively we might expect geographically embedded stations to require two dimensions, however, meteorological research has shown that the Australian climate is determined largely by latitude, due to a high-pressure belt which moves north and south over the year, influencing the rainfall patterns over the seasons²⁹. Since our embedding method can be used to understand proximal relationships and continuous node properties, we can compare the embedding positions to the station latitude, which has an approximate linear relationship to north-south distance. As in previous experiments, we take the absolute value of the KT correlation between the repelling Laplacian first eigenvector and the latitude. Again, we compare to the opposing Laplacian and SPONGE.

The first eigenvector of the repelling Laplacian (SHEEP) has the highest ordinal correlation with station latitude (0.744),

compared to the first eigenvector of both the opposing Laplacian (0.719) and SPONGE (0.178). We note that the second eigenvector of SPONGE has a higher correlation with latitude as compared to the first (0.733). As we will see with the following examples, the location of the best SPONGE eigenvector changes for different networks, making it difficult to know which eigenvector to choose without ground truth proximity information. Figure 7 shows the SHEEP embedding position versus latitude for each station, visually demonstrating the correlation between the embedding and the node attribute.

Numerical experiments II: node extremism measure

Signed configuration model. In this section, we argue that the distance from the node to the origin in the optimal embedding dimension gives a measure of the node's "extremism" or conflictuality. In general, the distance from the node to the origin in the embedding is dependent on both local measures, like the node's positive and negative degrees, and more global properties like the structure of the graph. For instance, as in Fig. 5, SHEEP places nodes with fewer negative edges to the other factions closer to the origin. Thus, the distance to the origin is determined by both the cluster structure of the graph, and by the intensity of the negative interactions between the nodes. When the graph structure is random, we expect that local and global measures will coincide. Using both a signed graph configuration model, and SBMs with various levels of noise, we investigate the relationship between the node's distance to the origin in the SHEEP embedding, and the net degree (negative degree minus positive degree).

The configuration model is a random graph model constructed from a given degree sequence, where each edge "stub" is matched with equal probability³⁰. We construct a signed configuration model and take 100 realizations of the graph with the same optimal dimension ($k = 9$). See Methods for a more in-depth description. We find for each node the net degree, and the distance to the origin, that is, the norm of its 9-dimensional position vector. The mean distance for each net degree is shown

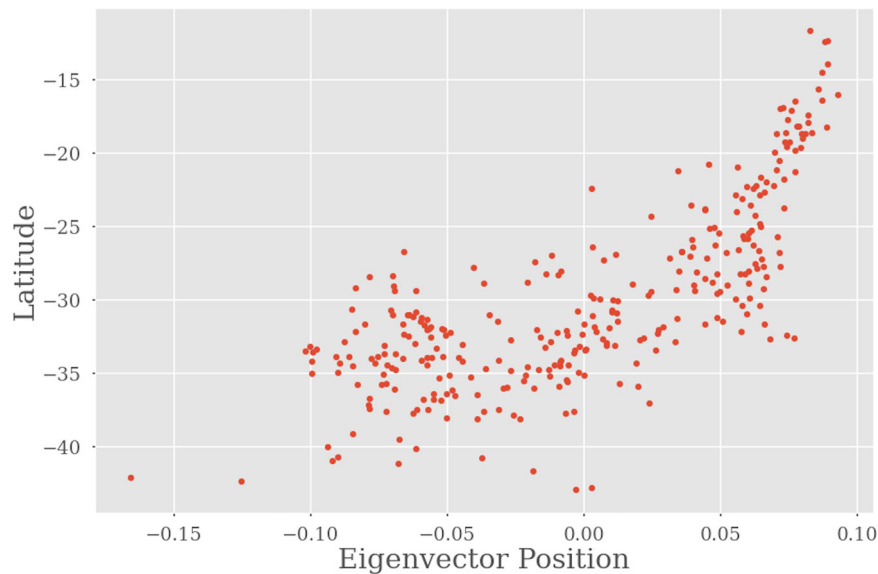


Fig. 7 Australian rainfall network latitude vs embedding. Station latitude versus SHEEP embedding positions, obtained from the first eigenvector of the repelling Laplacian. Edge weights are masked using the sign function, $(i, j \in \{-1, 1\})$. KT correlation is 0.744.

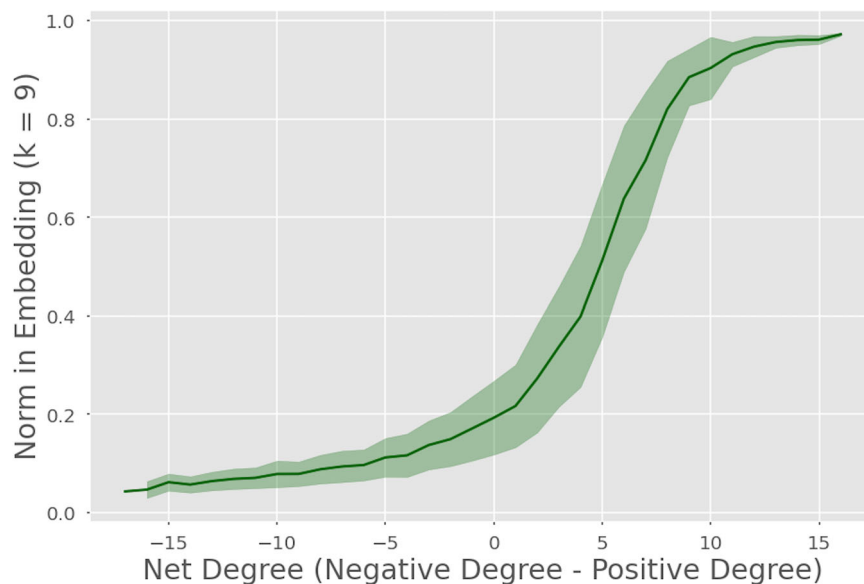


Fig. 8 Signed configuration model node norm vs net degree. Mean distance to origin in SHEEP embedding (the norm in $k = 9$ dimensions) versus the node's net degree over 100 realizations of the signed network configuration model, for which $k = 9$ was the optimal embedding dimension. Shaded area gives the standard deviation. Nodes having relatively more negative edges are further from the origin.

in Fig. 8, where the shaded area gives the standard deviation. We observe an increasing relationship, where nodes with a higher proportion of negative edges are more “extreme” and thus further from the origin.

In the case when the graph is not random and has structure (eg. factions), the norm gives a measure of extremism that cannot be explained by the net degree. In Supplementary Note 9, we use SSBMs with varying levels of noise, to show how for networks with strong bi-polarization, the norm in the embedding is determined primarily by the negative degree of the node. When the bi-polarization structure of the graph is perturbed, by noise or by other structural effects, the extremism measure is no longer determined just by the the negative degree. For graphs without a perfect polarization structure, both the positive and negative degree play a role in the extremism measure outputted by SHEEP,

as well as the structure of the graph, indicating that this measure incorporates more than just local information.

Continuous political ideology. Here, we consider an empirical network with ground truth continuous node attributes, which provides a better intuition for the node extremism measure. We take a signed network representing relationships between members of the USA House of Representatives for different congresses (here, we focus on sessions 110, and 144). Since we focus on the more recent congresses, which Aref et al. argue in ref. ²⁰ are highly bi-polarized along Democrat-Republican lines, it is not surprising that the optimal SHEEP embedding dimension is $k = 1$. Because SHEEP represents proximal relationships, we want to understand whether the embedding can recover the political ideology of the members on a continuous scale, and investigate

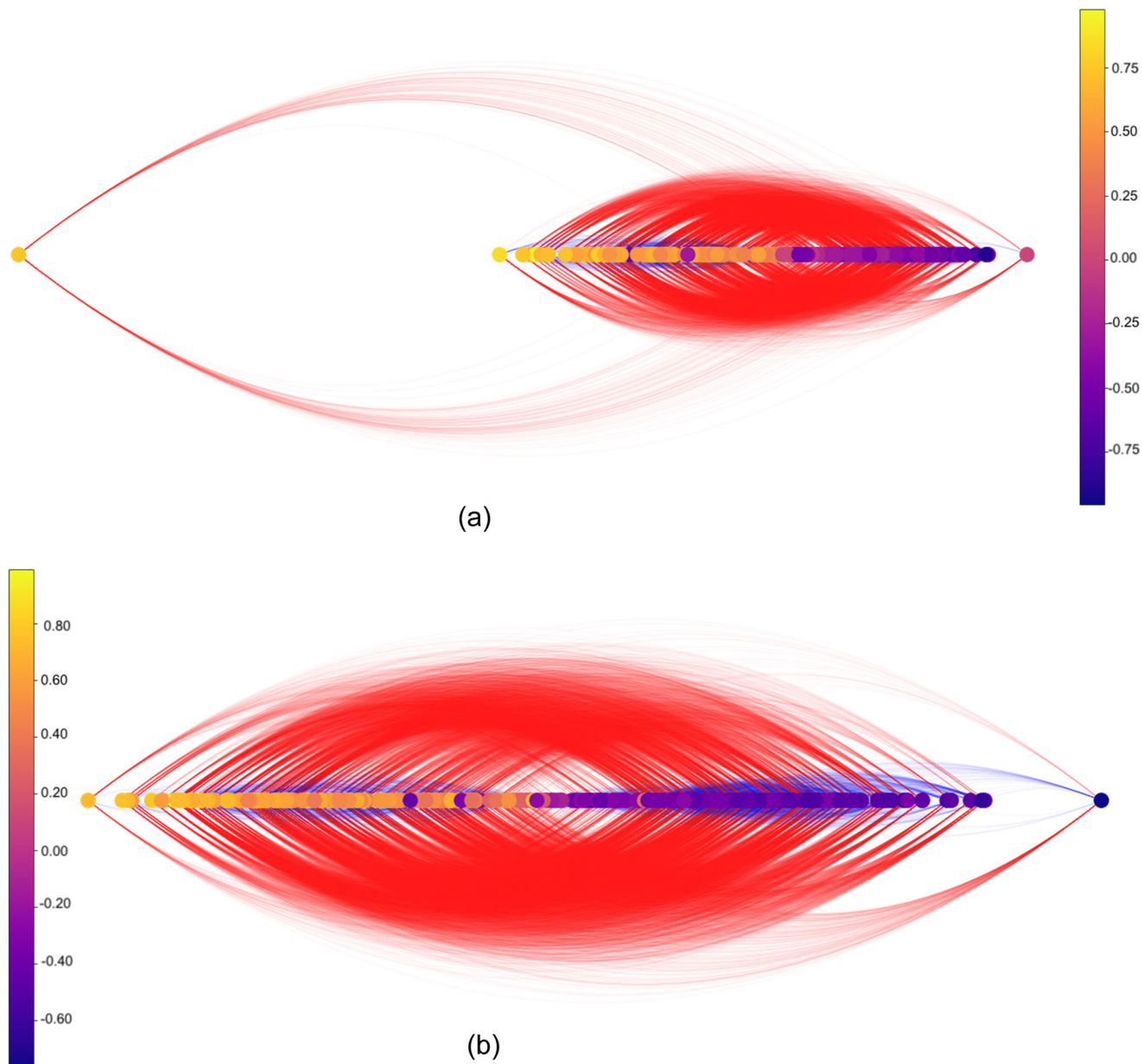


Fig. 9 House of representative network embeddings. One dimensional SHEEP embedding of Congress 110 (a) and Congress 114 (b) signed networks. Nodes colors indicate Nokken-Poole political ideology scores obtained from²⁰.

whether the distance to the origin in the embedding corresponds with political extremism. As suggested in ref.²⁰, we use Nokken-Poole (NP) ideology scores as our ground truth measure. See Methods for more details on the data. We compare the SHEEP embedding to the embedding obtained by the opposing Laplacian and SPONGE, noting that these methods are designed for clustering. In this example, the opposing Laplacian and SPONGE would likely outperform SHEEP at finding the two clusters associated to the Democrat and Republican parties, while SHEEP recovers the continuous ideology information contained in the NP scores.

For the 110th congress, the first first eigenvector of the repelling Laplacian (SHEEP) has the highest correlation with the NP scores (0.685), compared to the opposing Laplacian (0.432) and SPONGE (0.209). We note that the fourth eigenvector of SPONGE obtains a better correlation with the NP scores compared to the first (0.640). As we will see with the second

choice of congress, the location of the best SPONGE eigenvector changes as the network changes. Figure 9a shows the one-dimensional embedding of the network obtained by SHEEP, where the nodes are colored according to the NP scores.

For the 114th congress, SHEEP once again obtains the highest correlation with the NP scores (0.698), compared to the opposing Laplacian (0.382) and SPONGE (0.626). For this network, the first eigenvector of SPONGE has the highest correlation with the NP scores. We observe that the location of the best SPONGE eigenvector has changed for each graph. Knowing where this proximal information will occur in the set of SPONGE eigenvectors is non-trivial, especially when the ground truth scores do not exist. For this, SHEEP is superior, as the information is contained in first k eigenvectors deemed optimal by the embedding method. Figure 9b shows the one-dimensional SHEEP embedding of the network, where the nodes are colored according to the NP scores. Since the optimal embedding

dimension for the 114th congress is $k=1$, the KT correlation shows directly the relationship between the node's distance to the origin and its political extremism, using the NP scores.

Since this graph is neither random, like the signed configuration model, nor perfectly bi-polarized, like the SSBM, the node extremism measure given by the NP scores cannot be completely explained by the net degree or the negative degree alone. Neither of these measure are sufficient for determining the extremism of the node; the positive edges and the graph structure play a role, which is captured in the SHEEP embedding. Consequently, the SHEEP embedding gives a better correlation with the NP scores, compared to the negative degree alone. For a more detailed analysis, see Supplementary Note 9.

The node extremism measure outputted by SHEEP is a result of the interplay between the negative degree, positive degree and graph structure. For random graphs, like the signed configuration model, the the local and global measures coincide, and the norm in the embedding is determined by the net degree. For graphs with perfect faction structure, the norm is determined by the negative degree. SHEEP is most useful for investigating the cases in between, for which the graph structure is not completely random nor completely polarized. In these cases, the norm in the SHEEP embedding provides a meaningful way of understanding node extremism, as we see for the case of the political ideologies.

Conclusions

In this paper, we have presented SHEEP, a spectral embedding algorithm for finding proximal relationships between nodes in signed networks. The method is based on a physically inspired model: we construct a Hamiltonian that assigns attractive and repulsive forces to the positive and negative edges in the graph. We show that the Hamiltonian is intrinsically related to the graph's repelling Laplacian, and that finding the minimum energy configuration reduces to an eigenvector problem. We show that the resulting ground state energy, or minimum eigenvalue, is a statistical measure of graph bi-polarization structure. We extend our results to higher dimensions, presenting an energy-based approach to locating the optimal embedding dimension for the network. We propose an application of our measure to recovering proximity-based continuous node attributes, showing how the SHEEP embedding reproduces ordinal information on synthetic and empirical networks. We also show that the distance to the origin in the optimal embedding dimension gives a measure of node "extremism", which is related both to local information like the net degree, and to the graph's global structure. Overall, this work contributes to the growing body of literature on spectral methods for understanding signed networks, and characterizing node relationships by taking into account multi-scale information.

Future research perspectives include exploring application of SHEEP to signed social media datasets, for providing fresh/additional insight on the node proximity and extremism. With further investigation, the method could be extended quite naturally to weighted signed networks as well. Given the form of the Hamiltonian, a weighted adjacency matrix is equivalent to loosening the requirement that each spring has a unit-weight spring constant, incorporating the weights into the proximity-based embedding. The method could also be modified to perform better on sparse networks, perhaps through regularization³¹. Furthermore, extending the method to be robust to changes in number and density of edges would allow applications to temporal signed networks. In this case, understanding the dynamics of node-to-node proximity could be a rich area for future investigation. We also note that we made the choice of the spherical constraint as it seemed the most natural, but other constraints are possible, and

comparing this method with other constraints could also prove to be a fruitful research direction.

Methods

Relation between SHEEP and spectral methods. Here, we illustrate the relationship between SHEEP, and spectral approaches. As a graph with strong balance can be separated into two factions, we constrain ourselves to a one dimensional embedding for the remainder of this section. The vector $\boldsymbol{\pi}$ denotes the n -dimensional vector describing the positions π_i of the n nodes along a line.

In¹⁵, the authors introduce two types of signed graph Laplacian. The Laplacian of the positive part of the graph is $L^+ = D^+ - A^+$. There are two possibilities for the negative graph Laplacian, corresponding to two possibilities for the signed Laplacian. The opposing Laplacian:

$$L_o = D^+ - A^+ + D^- - A^-, \quad (13)$$

and the repelling Laplacian:

$$L_r = D^+ - A^+ - D^- - A^-. \quad (14)$$

Note that in the case of a positive graph, the repelling Laplacian is precisely the standard graph Laplacian. These two matrices induce quadratic forms on the vector $\boldsymbol{\pi} \in \mathbb{R}^n$. Note that π_i is a scalar value associated with the node i . The induced quadratic form of the opposing Laplacian is:

$$\boldsymbol{\pi}^T L_o \boldsymbol{\pi} = \sum_{i,j} A_{ij}^+ |\pi_i - \pi_j|^2 - \sum_{i,j} A_{ij}^- |\pi_i + \pi_j|^2, \quad (15)$$

and the repelling Laplacian:

$$\boldsymbol{\pi}^T L_r \boldsymbol{\pi} = \sum_{i,j} A_{ij}^+ |\pi_i - \pi_j|^2 + \sum_{i,j} A_{ij}^- |\pi_i - \pi_j|^2. \quad (16)$$

The quadratic form resulting from the repelling Laplacian is equivalent to the spring-inspired Hamiltonian proposed in Eq. (1), for one-dimensional position vectors.

Unlike the function arising from the inner product of the opposing Laplacian (Eq. (15)), the inner product of the repelling Laplacian is a function of distance for pairs of nodes connected by both positive or negative edges. Consequently, in the embedding generated by the opposing Laplacian, the presence of many or few negative links between positively connected factions does not effect the final positions, unlike the physically inspired method that we propose here. In particular, the opposing Laplacian identifies whether the graph is strongly balanced or not, but does not given an indication of the intensity of the negative interactions between the nodes, which is precisely the advantage of our proximity-based embedding.

The repelling Laplacian is symmetric and real, so its eigenvectors can be chosen to be orthonormal. Unlike the opposing Laplacian, which has been previously used for spectral embedding as in ref. ¹⁰ because it is positive semi-definite, the repelling Laplacian is indefinite. From a physical perspective, it describes a Hamiltonian that permits explosions of distances due to the quadratically increasing energy of repulsive forces, which is why we impose the spherical constraint.

Ground state energy and strong balance. In this section, we propose a test statistic for bi-polarization, which is inspired by the physical interpretation of our system, the ground state energy, denoted as E^0 . Analogously to the Ising model, where the lowest ground state energy is achieved when there is no geometric frustration introduced by negative cycles, we want to show that the ground state energy of our Hamiltonian is minimized when the graph exhibits strong balance. If a test statistic is significant, it should be highly improbable on a null model. As the ground state

energy depends on many aspects of the network structure, we focus on the following null model: fixing the graph topology, we randomly reshuffle the signs of the edges, while preserving the density of positive and negative edges, as in^{32,33}. If the real network has a lower ground state energy than the networks produced by the null model, we can conclude that the network has significant underlying bi-polarization structure. Since the ground state energy of our Hamiltonian is the minimum eigenvalue of the repelling Laplacian, we present here some results on the spectrum of the repelling Laplacian for complete graphs, and then generalize the results numerically to non-complete graphs.

Repelling Laplacian: spectral results. We consider the set of complete signed graphs with n nodes, subject to the condition that $E^+, E^- \neq \emptyset$, denoted by $\{G^n\}$. We present three results: (1) The minimal eigenvalue λ_n is bounded from below by $-n$. (2) The lower bound $\lambda_n = -n$ is achieved when the graph has a perfect bi-partition (strong balance). And (3) Introducing frustration by switching edge signs will strictly increase the value of the ground state energy. These results motivate the use of the ground state energy as a bi-polarization measure, or measure of strong balance.

Theorem 1. For an n -complete signed graph subject to the condition that $E^+, E^- \neq \emptyset$, the smallest eigenvalue of the repelling Laplacian $\lambda_n \geq -n$.

The proof is deferred to Supplementary Note 3, but the main steps involved are as follows. Beginning with the n -complete graph of all negative edges, one can easily show that the spectrum is $-n$ with multiplicity $n - 1$ and 0 with multiplicity 1 . Any other signed n -complete graph can be reached by successively changing an edge sign to positive, associated with the addition of a L^{shift} to the repelling Laplacian, which is positive semi-definite. Applying Weyl's inequality to the sum of the two Laplacians gives the required result.

Next, we show that the lower bound is reached when the graph is strongly balanced and admits a bi-partition.

Theorem 2. Consider a strongly balanced graph $\tilde{G} \in \{G^n\}$, such that the nodes of graph \tilde{G} can be partitioned into two sets V_1 and V_2 , where $V_1, V_2 \neq \emptyset$. Nodes inside each set are connected with positive edges, while the edges connecting the sets are negative. The smallest eigenvalue of the repelling Laplacian is $\lambda_n = -n$.

Again, we defer the proof to Supplementary Note 3. Broadly, we construct the eigenvector v associated to the minimum eigenvalue λ_n by placing nodes in the same set at the same point such that for node $k \in V_i$ we set $v_k = x_i$ where $i \in (1, 2)$. Using the orthonormality of the eigenvectors, we can show that associated eigenvalue is $-n$,

$$E_{bal}^0 = \lambda_n = -n. \quad (17)$$

By Theorem 1, this is the minimum eigenvalue. By Theorems 1 and 2, the ground state energy reaches the lower bound when the graph is non-frustrated, and has a perfect bi-partition in accordance with strong balance. When the graph is frustrated we want to show that the ground state energy, E_{frus}^0 is larger than the energy associated with the balanced graph E_{bal}^0 given in Eq. (17) as $-n$. We compare between graphs that have the same number of nodes, n , as well as the same density of positive and negative edges.

Theorem 3. Consider a strongly balanced graph $\tilde{G} \in \{G^n\}$, such that the nodes of graph \tilde{G} can be partitioned into two sets V_1 and V_2 , where $V_1, V_2 \neq \emptyset$. Introduce frustration into the graph by switching two edge signs. When the graph is sufficiently large such that $n > 4$, and $|V_1|, |V_2| > 2$, the ground state energy is strictly increased $E_{frus}^0 > E_{bal}^0$.

The proof involves writing the repelling Laplacian of the frustrated graph as the sum of the repelling Laplacian of the balanced graph $L_r^{\tilde{G}}$, and a perturbation, L^{switch} , associated to the switching of a positive and negative edge sign. The new graph is frustrated because the two terms $L_r^{\tilde{G}}$ and L^{switch} cannot be simultaneously minimized. Finding a bound on the second smallest eigenvalue of the repelling Laplacian of the balanced graph $L_r^{\tilde{G}}$, gives an approximation for the energy gap between the ground and first energy levels. When the graph is sufficiently large such that $n > 4$, and $|V_1|, |V_2| > 2$, we can use the energy gap to bound the ground state energy from below, obtaining the inequality we require. In practice, $n > 4$ is a weak constraint, as most graphs of interest contain more than four nodes, so

$$E_{frus}^0 > -n = E_{bal}^0. \quad (18)$$

Bi-polarization measure. Following the intuition provided by the spectrum of the repelling Laplacian on complete signed graphs, we numerically generalize to the case of non-complete graphs, showing that the ground state energy is a statistically significant measure of bi-polarization, as compared to the null model. The test correctly identifies the polarization structures on various synthetic signed graphs with strong balance. We propose a bi-polarization score using the z-score of the ground state energy of the graph, compared to the the null model energy distribution. We define:

$$Z(G) = \frac{E_0^G - \langle E_0 \rangle}{\sigma}, \quad (19)$$

where $\langle E_0 \rangle$ is the mean of the null model distribution, and σ is the standard deviation. A large negative value of $Z(G)$ indicates that the graph is significantly polarized.

There exist a number of proposed bi-polarization measures for signed networks, which typically focus on the local or global properties of the graph. For example,³⁴ proposes a balance measure that counts the number of unbalanced cycles in the graph, inversely weighted by length. A more global measure is the frustration index, which counts the number of frustrated edges associated with the best bi-partition³⁵. An intermediate scale measure, POLE, takes the correlation between a node's signed and unsigned random walk, based on the assumption that link sign should be related to unsigned community structure³⁶. Our measure incorporates both local and global structure, and allows for comparison between networks with fixed size and edge density. See Supplementary Note 4 for a numerical test of the bi-polarization measure on synthetic signed networks with polarization structure.

Well-ordered synthetic graph construction. The well-ordered synthetic graph is constructed as a complete signed graph. We place nodes randomly with uniform distribution between -1 and 1 , and normalizing the resulting position vector to 1 . The signed graph is constructed deterministically: nodes are connected by a positive edge if the distance between them is less than a chosen threshold value, otherwise the edge sign is negative. In comparing the ordinal information obtained by the embedding, we use the KT correlation is over the Pearson correlation because the proximal relationship obtained by the embedding may be monotonic, but non linear.

Signed configuration model. The signed configuration model is generated by first generating two 50 node (unsigned) configuration models using randomly generated degree sequences of integers between 1 and 20 . We take the graph obtained by subtracting the adjacency matrix of the first model from the second to obtain a random signed network. Over 5000 realizations of the

signed configuration model, the most frequently obtained optimal embedding dimension is $k=9$ using SHEEP. To ensure we compare across graphs with the same optimal dimension, we use 100 realizations of the signed configuration models for which $k=9$ was optimal in our analysis.

Australian rainfall network construction. The Australian Rainfall graph is constructed from the Pearson correlation matrix obtained from the seasonal rainfall time series. The rainfall is given in millimeters. This results in a complete signed graph on $n=305$ nodes, corresponding to the 305 stations. We mask the edge weight magnitude by taking the sign of the Pearson correlations such that for all $(i, j) \in E$, $(i, j) \in \{-1, 1\}$. Note that the sign of Pearson correlation is not transitive, depending non-trivially on the magnitude of the correlation. As a result, the resulting signed graph does not have a perfect bi-partition. The same dataset was studied in ref. ¹¹, where the authors used a $k=6$ and $k=10$ dimensional embedding to obtain clusters of stations, which corresponded with the ground truth geographical regions.

House of representatives data. The data-set for this analysis is obtained from ref. ²⁰. The signed network represents the relationships between members of the USA House of Representatives for two different congresses (sessions 110, and 144). In this signed network, a positive (negative) edge indicates that the two representatives have co-sponsored statistically significantly more (fewer) bills than expected by chance. The NP ideology scores place each member of congress on a continuous scale from +1 (conservative) to -1 (liberal). The score is frequently used in political science, and is the result of a three step multi-dimensional scaling method based on member's voting habits in a given congress, derived by maximizing their utility^{37,38}. The NP scores are two dimensional, but the first dimension is often taken to represent political ideology on a left/right spectrum. The 110th congress session forms a signed graph of $n=452$ nodes. Taking the largest connected component, results in a graph of $n=450$ nodes. The 114th congress session forms a signed graph of $n=446$ nodes. Taking the largest connected component results in a signed graph with $n=443$ nodes. When looking for the eigenvector of SPONGE and the opposing Laplacian that has a better correlation with the NP scores, we search through the ten smallest eigenvectors of each matrix respectively.

Data availability

The Australian rainfall data-set was obtained from the authors of ref. ¹¹, and is available upon request. The congress network data-sets were obtained from^{20,39,40} and can be downloaded: <https://osf.io/3qtfb/>.

Code availability

The code required to perform the SHEEP embedding is available at <https://github.com/saynabul/SHEEP>.

Received: 15 February 2023; Accepted: 14 December 2023;

Published online: 04 January 2024

References

- Newman, M. *Networks: An Introduction*. (Oxford University Press, Oxford, UK, 2010).
- Heider, F. Attitudes and cognitive organization. *J. Psychol.* **21**, 107–112 (1946).
- Richardson, M., Agrawal, R. & Domingos, P. Trust management for the semantic web. In *Proc. Semantic Web-ISWC 2003: Second International Semantic Web Conference, Sanibel Island, FL, USA, October 20–23, 2003*, 351–368 (Springer, 2003).
- Leskovec, J., Lang, K. J., Dasgupta, A. & Mahoney, M. W. Community structure in large networks: natural cluster sizes and the absence of large well-defined clusters. *Internet Math.* **6**, 29–123 (2009).
- Pougué-Biyong, J. et al. Debatement: A comment-reply dataset for (dis) agreement detection in online debates. *Proc. Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)* (2021).
- Pougué-Biyong, J., Gupta, A., Haghghi, A. & El-Kishky, A. Learning stance embeddings from signed social graphs. *Proc. Sixteenth ACM International Conference on Web Search and Data Mining*, 177–185 (2023).
- Tang, J., Chang, Y., Aggarwal, C. & Liu, H. A survey of signed network mining in social media. *ACM Comput. Surv.* **49**. <https://doi.org/10.1145/2956185> (2016).
- Zhang, S. et al. Where are we in embedding spaces? *Proc. of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2223–2231 (2021).
- Fouss, F., Saerens, M. & Shimbo, M. *Algorithms and models for network data and link analysis* (Cambridge University Press, 2016).
- Kunegis, J. et al. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proc. 2010 SIAM international conference on data mining*, 559–570 (SIAM, 2010).
- Cucuringu, M., Davies, P., Glielmo, A. & Tyagi, H. Sponge: A generalized eigenproblem for clustering signed networks. In *Proc. 22nd International Conference on Artificial Intelligence and Statistics*, 1088–1098 (PMLR, 2019).
- Knyazev, A. On spectral partitioning of signed graphs. In *Proc. Seventh SIAM Workshop on Combinatorial Scientific Computing*, 11–22 (SIAM, 2018).
- Fox, A., Manteuffel, T. & Sanders, G. Numerical methods for gremban's expansion of signed graphs. *SIAM J. Sci. Comput.* **39**, S945–S968 (2017).
- Kermarrec, A.-M. & Moin, A. Energy models for drawing signed graphs. [Research Report] 2011, pp. 29 inria-00605924v3.
- Shi, G., Altafini, C. & Baras, J. S. Dynamics over signed networks. *SIAM Rev.* **61**, 229–257 (2019).
- Ou, L., Hou, Y. & Xiong, Z. The net laplacian spectra of signed complete graphs. *Contem. Math.* **2**, 409–417 (2021).
- Harary, F. On the notion of balance of a signed graph. *Mich. Math. J.* **2**, 143–146 (1953).
- Cartwright, D. & Harary, F. Structural balance: a generalization of Heider's theory. *Psychol. Rev.* **63**, 277 (1956).
- Davis, J. A. Clustering and structural balance in graphs. *Hum. Relat.* **20**, 181–187 (1967).
- Aref, S. & Neal, Z. P. Identifying hidden coalitions in the US House of Representatives by optimally partitioning signed networks based on generalized balance. *Sci. Rep.* **11**, 1–9 (2021).
- Doreian, P. & Mrvar, A. Partitioning signed social networks. *Soc. Netw.* **31**, 1–11 (2009).
- Traag, V. A. & Bruggeman, J. Community detection in networks with positive and negative links. *Phys. Rev. E* **80**, 036115 (2009).
- He, X., Du, H., Xu, X. & Du, W. An energy function for computing structural balance in fully signed network. *IEEE Trans. Comput. Soc. Syst.* **7**, 696–708 (2020).
- Cao, J., Fan, Y. & Di, Z. Frustration of signed networks: how does it affect the thermodynamic properties of a system? arXiv preprint arXiv:1810.10481 (2018).
- Doreian, P. & Krackhardt, D. Pre-transitive balance mechanisms for signed networks. *J. Mathe. Sociol.* **25**, 43–67 (2001).
- De Bacco, C., Larremore, D. B. & Moore, C. A physical model for efficient ranking in networks. *Sci. Adv.* **4**, eaar8260 (2018).
- Kawamoto, T., Ochi, M. & Kobayashi, T. Consistency between ordering and clustering methods for graphs. *Physical Review Research* **5**, 023006 (2023).
- Bronski, J. C. & DeVille, L. Spectral theory for dynamics on graphs containing attractive and repulsive interactions. *SIAM J. Appl. Math.* **74**, 83–105 (2014).
- Deacon, E. Climatic change in australia since 1880. *Aust. J. Phys.* **6**, 209–218 (1953).
- Newman, M. *Networks* (Oxford University Press, 2018).
- Qin, T. & Rohe, K. Regularized spectral clustering under the degree-corrected stochastic blockmodel. *Advances in neural information processing systems* **26** (2013).
- Leskovec, J., Huttenlocher, D. & Kleinberg, J. Signed networks in social media. *Proc. SIGCHI Conference On Human Factors In Computing Systems*, 1361–1370 (2010).
- Szell, M., Lambiotte, R. & Thurner, S. Multirelational organization of large-scale social networks in an online world. *Proc. National Acad. Sci.* **107**, 13636–13641 (2010).
- Kirkley, A., Cantwell, G. T. & Newman, M. E. J. Balance in signed networks. *Phys. Rev. E* **99**. <https://doi.org/10.1103/physreve.99.012320> (2019).
- Aref, S. & Wilson, M. C. Balance and frustration in signed networks. *J. Complex Netw.* **7**, 163–189 (2019).

36. Huang, Z., Silva, A. & Singh, A. Pole: Polarized embedding for signed networks. *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 390–400 (2022).
37. Carroll, R., Lewis, J. B., Lo, J., Poole, K. T. & Rosenthal, H. Measuring bias and uncertainty in dw-nominate ideal point estimates via the parametric bootstrap. *Political Anal.* **17**, 261–275 (2009).
38. Nokken, T. P. & Poole, K. T. Congressional party defection in American history. *Legis. Stud. Q.* **29**, 545–568 (2004).
39. Neal, Z. The backbone of bipartite projections: inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Soc. Netw.* **39**, 84–97 (2014).
40. Domagalski, R., Neal, Z. P. & Sagan, B. Backbone: an R package for extracting the backbone of bipartite projections. *Plos One* **16**, e0244363 (2021).

Acknowledgements

S.A.B. was supported by EPSRC grant EP/W523781/1. The work of R.L. was supported by EPSRC grants EP/V03474X/1 and EP/V013068/1. We thank the authors of¹¹ for the access to the Australian Rainfall Correlations data set.

Author contributions

S.A.B. and R.L. designed the SHEEP method. S.A.B. ran the numerical experiments. S.A.B. and R.L. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42005-023-01504-6>.

Correspondence and requests for materials should be addressed to Shazia'Ayn Babul.

Peer review information *Communications Physics* thanks Homayoun Hamedmoghadam and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024