



**HAL**  
open science

## Clustering with Simplicial Complexes

Thummaluru Siddartha Reddy, Sundeep Prabhakar Chepuri, Pierre. Borgnat

► **To cite this version:**

Thummaluru Siddartha Reddy, Sundeep Prabhakar Chepuri, Pierre. Borgnat. Clustering with Simplicial Complexes. EUSIPCO 2023, 31st European Signal Processing Conference, EURASIP, European Association for Signal Processing, Sep 2023, Helsinki, Finland. hal-04227080

**HAL Id: hal-04227080**

**<https://hal.science/hal-04227080>**

Submitted on 3 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Clustering with Simplicial Complexes

Thummaluru Siddhartha Reddy\*, Sundeep Prabhakar Chepuri\*, and Pierre Borgnat†

\*Indian Institute of Science, Bengaluru, India

†ENS de Lyon, CNRS, Lab. Physique, Lyon, France

**Abstract**—In this work, we propose a new clustering algorithm to group nodes in networks based on second-order simplices (aka filled triangles) to leverage higher-order network interactions. We define a simplicial conductance function, which on minimizing, yields an optimal partition with a higher density of filled triangles within the set while the density of filled triangles is smaller across the sets. To this end, we propose a simplicial adjacency operator that captures the relation between the nodes through second-order simplices. This allows us to extend the well-known Cheeger inequality to cluster a simplicial complex. Then, leveraging the Cheeger inequality, we propose the simplicial spectral clustering algorithm. We report results from numerical experiments on synthetic and real-world network data to demonstrate the efficacy of the proposed approach. \*

**Index Terms**—Cheeger inequality, clustering, higher-order cuts, simplicial complexes, triangle conductance.

## I. INTRODUCTION

Networks model complex interactions (as edges) between entities (as nodes). Networks often have community structures, and determining these communities is a topic of significant interest in network science [1], [2]. Clustering algorithms detect communities by partitioning the nodes in a network into sets with high edge density within a set while maintaining a low edge density between nodes of different sets. Such partitioning is promoted by cut criteria, such as modularity [3] or edge conductance [4], [5]. Spectral clustering [6], [7] is a well-known heuristic obtained by relaxing the edge conductance cut criterion and relating it to the second smallest eigenvalue of the graph Laplacian through the so-called Cheeger inequality. However, these methods only capture the pairwise relations between nodes. In many real-world networks [8], [9], *supra-pairwise* relations, i.e., beyond pairwise relationships, e.g., triadic or more, are common [10], and preserving such structures while clustering is paramount.

Higher-order spectral clustering algorithms account for higher-order interactions, wherein the cut of triangles (capturing triadic interactions) is minimized for network partitioning [2], [11], [12]. These methods can be broadly classified into two classes. The first class encodes higher-order structures in an affinity tensor (a higher-order generalization of a similarity matrix), then apply spectral clustering on it [11]. The second class directly constructs the so-called motif Laplacian matrix by counting higher-order structures in the network, then apply spectral clustering on it [2], [12]. An affinity tensor or motif adjacency matrix assumes that all pairwise interactions lead to a higher-order (e.g., triadic) relation. That is, it assumes

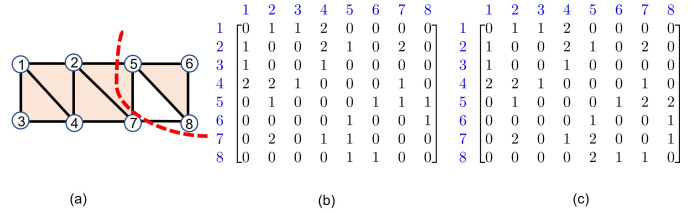


Fig. 1: (a) Simplicial complex with a 2-simplex cut. (b) Simplicial adjacency matrix. (c) Triangle motif adjacency matrix.

all triangles are filled, while we may have hollow triangles in some cases. As an example, the network in Fig. 1(a) has only pairwise relations between the nodes  $\{5, 7, 8\}$ , through  $\{5, 7\}$ ,  $\{7, 8\}$ ,  $\{8, 5\}$ , and the triangle  $\{5, 7, 8\}$  is hollow. This, in other words, means that the network does not contain a triadic interaction between nodes  $\{5, 7, 8\}$ . A triangle motif, on the contrary, counts  $\{5, 7, 8\}$  as a triangle and assumes it as equally important as other filled triangles, e.g.,  $\{1, 3, 4\}$ ,  $\{1, 2, 4\}$ ,  $\{2, 4, 7\}$  and  $\{5, 6, 8\}$  (see its adjacency matrix in Fig. 1(b)). To distinguish filled higher-order structures, we model networks as simplicial complexes (see its adjacency matrix in Fig. 1(c)). Simplicial complexes [13] are mathematical objects that model higher-order interactions in networks and are composed of simplices of different orders, such as nodes (0-simplices), edges (1-simplices), and filled triangles (2-simplices) as basic building blocks.

This work focuses on clustering a network modeled as a simplicial complex using 2-simplices. Specifically, we develop a simplicial conductance function, which, when minimized, outputs a partition with a higher density of 2-simplices within the set and a lower density of 2-simplices across the sets. To circumvent the combinatorial optimization problem when minimizing the simplicial conductance, we propose a new operator called *simplicial adjacency* to encode the similarity between the nodes through the 2-simplices. The simplicial adjacency matrix can be computed in closed form as a boundary of a boundary matrix associated with the simplicial complex. We then relate the spectrum of the simplicial Laplacian matrix (derived from the simplicial adjacency matrix) to simplicial conductance through the Cheeger inequality. This allows us to propose a simplicial spectral clustering algorithm.

## II. SIMPLICIAL ADJACENCY MATRIX

A simplex is a subset of the vertex set  $\mathcal{V} = \{v_1, \dots, v_N\}$  of a graph with a  $k$ -simplex (or simplex of order  $k$ ), denoted as

\*Supported by the IFCAM project MA/IFCAM/19/56; and the CHIST-ERA-19-XAI-006, for the GRAPHNEX ANR-21-CHR4-0009 project.

$\sigma_i^{(k)} = \{v_{i_1}, v_{i_2}, \dots, v_{i_{k+1}}\}$ , being a subset of  $\mathcal{V}$  of cardinality  $k + 1$ . For instance,  $\sigma_i^{(2)} = \{v_{i_1}, v_{i_2}, v_{i_3}\} \subseteq \mathcal{V}$  denotes a 2-simplex. An undirected simplicial complex  $\mathcal{X}$  is a finite collection of simplices such as nodes (0-simplices), edges (or 1-simplices), and filled-triangles (or 2-simplices). The order of a simplicial complex is the highest order of the simplices it contains. The boundary matrix  $\mathbf{B}_k \in \mathbb{R}^{N_{k-1} \times N_k}$  encodes the relation between  $(k-1)$ -simplices and  $k$ -simplices, i.e., it encodes which  $(k-1)$ -simplex is adjacent to which  $k$ -simplices. For an undirected simplicial complex  $\mathcal{X}$ , the boundary operator  $\mathbf{B}_k$  has entries as

$$[\mathbf{B}_k]_{ij} = \begin{cases} 1, & \text{if } \sigma_i^{(k-1)} \subset \sigma_j^{(k)}, \\ 0, & \text{otherwise.} \end{cases}$$

Since we do not account for any orientation in  $\mathbf{B}_k$ , the boundary of a boundary map is not equal to zero, i.e.,  $\mathbf{B}_k \mathbf{B}_{k+1} \neq \mathbf{0}$ .

We define the *simplicial adjacency matrix*  $\mathbf{A}_{0,2}$  that encodes the relations between 0-simplices through 2-simplices with its  $(i, j)$  entry  $[\mathbf{A}_{0,2}]_{ij}$  equal to the number of 2-simplices the nodes  $v_i$  and  $v_j$  appear in, i.e.,

$$[\mathbf{A}_{0,2}]_{ij} = \sum_{p=1}^{N_2} \mathbb{I}(\{v_i, v_j\} \in \sigma_p^{(2)}) \quad (1)$$

for  $i \neq j$  and  $[\mathbf{A}_{0,2}]_{ii} = 0$ . Here,  $\mathbb{I}(\cdot)$  is the indicator function that returns 1 when its argument is true. Let us define the diagonal degree matrix  $\mathbf{D}_{0,2} \in \mathbb{R}^{N_0 \times N_0}$  matrix with entries  $[\mathbf{D}_{0,2}]_{ii} = \sum_j [\mathbf{A}_{0,2}]_{ij}$  and the simplicial Laplacian matrix as  $\mathbf{L}_{0,2} = \mathbf{D}_{0,2} - \mathbf{A}_{0,2}$ . The normalized simplicial Laplacian is given by  $\tilde{\mathbf{L}}_{0,2} = \mathbf{D}_{0,2}^{-1/2} \mathbf{L}_{0,2} \mathbf{D}_{0,2}^{-1/2}$ . The simplicial adjacency matrix in (1) can also be obtained using the boundary operators as follows.

**Proposition.** *The simplicial adjacency matrix  $[\mathbf{A}_{0,2}]_{ij}$  can be computed from the boundary matrices  $\mathbf{B}_1$  and  $\mathbf{B}_2$  as*

$$[\mathbf{A}_{0,2}]_{i,j} = \frac{[\mathbf{B}_{12} \mathbf{B}_{12}^T]_{ij}}{4}. \quad (2)$$

where  $\mathbf{B}_{12} = \mathbf{B}_1 \mathbf{B}_2 \in \mathbb{R}^{N_0 \times N_2}$ .

Since each row of  $\mathbf{B}_1$  corresponds to a 0-simplex and has nonzero entries where 1-simplices are incident on the 0-simplex and each column of  $\mathbf{B}_2$  corresponds to a 2-simplex and has nonzero entries where 2-simplices are incident on 1-simplices. Therefore, the inner product between  $i$ th row and  $j$ th column gives a nonzero value if a 0-simplex is incident upon a 2-simplex. Hence  $\mathbf{B}_{12}$  has nonzero entries only if a 0-simplex is incident on a 2-simplex. In other words,  $[\mathbf{B}_{12}]_{ij} = 2\mathbb{I}(v_i \in \sigma_j^{(2)})$ . Here, the factor 2 appears as we consider undirected simplices. Now,  $[\mathbf{A}_{0,2}]_{i,j}$  is obtained by taking the inner product between the  $i$ th row and  $j$ th column of  $\mathbf{B}_{12}$  that counts the total number of 2-simplices the edge between nodes  $v_i$  and  $v_j$  appear in. The factor of 4 is due to undirected simplices.

### III. CLUSTERING USING 2-SIMPLICES

This section discusses the proposed algorithm for clustering 0-simplices (i.e., nodes) in a simplicial complex  $\mathcal{X}$  by leveraging the higher-order relation through 2-simplices.

#### A. Simplicial conductance

The 2-way partitioning of the undirected simplicial complex  $\mathcal{X}$  based on 2-simplices corresponds to finding a nodal partition such that 0-simplices within a set have a high density of 2-simplices, and 0-simplices across sets have a low density of 2-simplices. To obtain this, we define the following cut measure:

$$\phi_{0,2}(\mathcal{S}) = \frac{\text{cut}_{0,2}(\mathcal{S}, \bar{\mathcal{S}})}{\min\{\text{vol}_{0,2}(\mathcal{S}), \text{vol}_{0,2}(\bar{\mathcal{S}})\}}, \quad (3)$$

where we call  $\phi_{0,2}(\mathcal{S})$  the *simplicial conductance* induced by 2-simplices. In (3),  $\text{cut}_{0,2}(\mathcal{S}, \bar{\mathcal{S}})$  measures the number of 2-simplex cut of  $\mathcal{S}$ , i.e., the number of 2-simplices that have one vertex in  $\mathcal{S}$  and other vertices in  $\bar{\mathcal{S}}$ ;  $\text{vol}_{0,2}(\mathcal{S})$  measures the total number of 2-simplices having vertices in  $\mathcal{S}$ . For the optimal partitioning of the simplicial complex, we minimize  $\phi_{0,2}(\mathcal{S})$  as

$$\phi_{0,2}^*(\mathcal{S}) = \underset{\mathcal{S} \subset \mathcal{V}}{\text{minimize}} \phi_{0,2}(\mathcal{S}). \quad (4)$$

Solving (4) is NP-hard as it involves evaluating all the possible cuts. To circumvent the difficulty, we develop an algorithm similar to spectral clustering, namely, simplicial spectral clustering.

#### B. Cheeger inequality with simplicial adjacency

We now discuss the relationship between simplicial conductance to the second smallest eigenvalue of the simplicial Laplacian matrix  $\mathbf{L}_{0,2}$  in the following theorem.

**Theorem.** *For an undirected simplicial complex  $\mathcal{X}$  having the normalized simplicial Laplacian matrix  $\tilde{\mathbf{L}}_{0,2}$  with  $\lambda_2$  being its second smallest eigenvalue, we have*

$$\frac{\lambda_2}{2} \leq \phi_{0,2}^*(\mathcal{S}) \leq \sqrt{2\lambda_2}.$$

*Proof.* We prove the above Cheeger inequality for an undirected simplicial complex by relating the simplicial conductance  $\phi_{0,2}(\mathcal{S})$  to the quadratic form of simplicial Laplacian  $\mathbf{L}_{0,2}$ . Let us define  $z_i(\mathcal{S})$  as number of 2-simplices with exactly  $i$  vertices in  $\mathcal{S}$ . We can express the  $\text{vol}_{0,2}(\mathcal{S})$  that measures the total number of 2-simplices having vertices in  $\mathcal{S}$  as:

$$\text{vol}_{0,2}(\mathcal{S}) = 3z_3(\mathcal{S}) + 2z_2(\mathcal{S}) + z_1(\mathcal{S}).$$

The factors 3 and 2 are due to the undirectedness of the simplicial complex. Similarly, the cut function  $\text{cut}_{0,2}(\mathcal{S}, \bar{\mathcal{S}})$  can be expressed as  $\text{cut}_{0,2}(\mathcal{S}, \bar{\mathcal{S}}) = z_2(\mathcal{S}) + z_1(\mathcal{S})$ . Therefore the simplicial conductance is

$$\phi_{0,2}(\mathcal{S}) = \frac{z_2(\mathcal{S}) + z_1(\mathcal{S})}{\min(\text{vol}_{0,2}(\mathcal{S}), \text{vol}_{0,2}(\bar{\mathcal{S}}))}. \quad (5)$$

**Algorithm 1** Clustering 0-simplices based on 2-simplices

- 
- 1: Input: Simplicial complex  $\mathcal{X}$ , boundary matrices  $\mathbf{B}_1, \mathbf{B}_2$
  - 2: Output: Clusters  $(\mathcal{S}, \bar{\mathcal{S}})$  based on 2-simplices
  - 3: Compute  $\mathbf{A}_{0,2}$  from (2),  $\mathbf{D}_{0,2}$ , and  $\tilde{\mathbf{L}}_{0,2} = \mathbf{D}_{0,2}^{-1/2} \mathbf{L}_{0,2} \mathbf{D}_{0,2}^{-1/2}$
  - 4:  $\tilde{\mathbf{e}}_{\mathcal{S}} \leftarrow$  Eigenvector of  $\tilde{\mathbf{L}}_{0,2}$  corresponding to its second smallest eigenvalue
  - 5:  $\gamma_k \leftarrow$  Node index corresponding to the  $k$ th smallest entry of  $\mathbf{D}_{0,2}^{-1/2} \tilde{\mathbf{e}}_{\mathcal{S}}$
  - 6:  $\mathcal{S} \leftarrow \min_{1 \leq k \leq N_0} \phi_{0,2}(\mathcal{S}_k)$ , where  $\mathcal{S}_k = \{\gamma_1, \gamma_2, \dots, \gamma_k\}$ .
- 

Next we express the numerator in (5) using the quadratic form of the simplicial Laplacian matrix. Let us first define the  $3 \times 3$  symmetric matrix  $\mathbf{M}(\sigma_i)$  with the following entries:

$$[\mathbf{M}(\sigma_i^{(2)})]_{mm} = \begin{cases} 2, & \text{if } v_m \in \sigma_i^{(2)}, \\ 0, & \text{otherwise} \end{cases}$$

and

$$[\mathbf{M}(\sigma_i^{(2)})]_{mn} = \begin{cases} -1, & \text{if } \{v_m, v_n\} \in \sigma_i^{(2)}, \\ 0, & \text{otherwise.} \end{cases}$$

Let us also define the vector  $\mathbf{e}_{\mathcal{S}} \in \{0, 1\}^{N_0}$  with entries  $[\mathbf{e}_{\mathcal{S}}]_i = \mathbb{I}(v_i \in \mathcal{S})$ . For a 2-simplex, say  $\sigma_i^{(2)} = \{v_{i_1}, v_{i_2}, v_{i_3}\}$ , we define the  $3 \times 1$  vector  $\mathbf{c}(\sigma_i^{(2)}) = [\mathbb{I}(v_{i_1} \in \mathcal{S}), \mathbb{I}(v_{i_2} \in \mathcal{S}), \mathbb{I}(v_{i_3} \in \mathcal{S})]^T$ . Then the quadratic form can be expressed as

$$\begin{aligned} \frac{1}{2} \mathbf{e}_{\mathcal{S}}^T \mathbf{L}_{0,2} \mathbf{e}_{\mathcal{S}} &= \sum_{i=1}^{N_2} \mathbf{c}^T(\sigma_i^{(2)}) \mathbf{M}(\sigma_i^{(2)}) \mathbf{c}(\sigma_i^{(2)}) \\ &= (z_2(\mathcal{S}) + z_1(\mathcal{S})) \\ &= \text{cut}_{0,2}(\mathcal{S}, \bar{\mathcal{S}}). \end{aligned} \quad (6)$$

Similarly, we can express  $\text{vol}_{0,2}(\mathcal{S})$  in the quadratic form of the degree matrix as

$$\begin{aligned} \frac{1}{2} \mathbf{e}_{\mathcal{S}}^T \mathbf{D}_{0,2} \mathbf{e}_{\mathcal{S}} &= \sum_{i=1}^{N_2} \mathbf{c}^T(\sigma_i^{(2)}) \text{diag}(\mathbf{M}(\sigma_i^{(2)})) \mathbf{c}(\sigma_i^{(2)}) \\ &= 3z_3(\mathcal{S}) + 2z_2(\mathcal{S}) + z_1(\mathcal{S}) \\ &= \text{vol}_{0,2}(\mathcal{S}). \end{aligned} \quad (7)$$

Hence, (4) can be equivalently expressed as

$$\begin{aligned} &\underset{\mathcal{S}}{\text{minimize}} \quad \frac{\mathbf{e}_{\mathcal{S}}^T \mathbf{L}_{0,2} \mathbf{e}_{\mathcal{S}}}{\mathbf{e}_{\mathcal{S}}^T \mathbf{D}_{0,2} \mathbf{e}_{\mathcal{S}}} \\ &\text{s.t.} \quad \mathbf{1}^T \mathbf{D}_{0,2} \mathbf{e}_{\mathcal{S}} = 0, \end{aligned} \quad (8)$$

where  $\mathbf{1}$  is the all-one vector, the constraint ensures a non-trivial solution.

By defining  $\tilde{\mathbf{e}}_{\mathcal{S}} = \mathbf{D}_{0,2}^{1/2} \mathbf{e}_{\mathcal{S}}$ , (9) can be transformed to

$$\begin{aligned} &\underset{\mathcal{S}}{\text{minimize}} \quad \tilde{\mathbf{e}}_{\mathcal{S}}^T \tilde{\mathbf{L}}_{0,2} \tilde{\mathbf{e}}_{\mathcal{S}} \\ &\text{s.t.} \quad \mathbf{1}^T \mathbf{D}_{0,2}^{1/2} \tilde{\mathbf{e}}_{\mathcal{S}} = 0, \\ &\quad \quad \tilde{\mathbf{e}}_{\mathcal{S}}^T \tilde{\mathbf{e}}_{\mathcal{S}} = 1. \end{aligned} \quad (9)$$

TABLE I: DATASETS.

Dataset	# of nodes	# of edges	# of clusters
Zachary [15]	34	78	2
Polbooks [3]	105	441	3
Football [3]	115	613	12

Thus solving the above problem to minimize the simplicial conductance as in (4) is analogous to the well-known 2-way spectral clustering problem based on edge cuts and deriving the related Cheeger inequality [14], [7] remains the same, but with the main difference being that the simplicial conductance is now bounded by the second smallest eigenvalue of normalized simplicial Laplacian.  $\square$

Based on this theorem, we propose the simplicial spectral clustering algorithm as detailed in Algorithm 1, which extends classical spectral clustering to obtain an optimal 2-way partition of the simplicial complex using 2-simplices.

## IV. NUMERICAL EXPERIMENTS

Numerical experiments to test the proposed method are conducted on synthetic and real datasets. We compare the proposed approach with triangle motif-based [2] and graph-based [7] spectral clustering algorithms. As a performance metric, we use normalized mutual information (NMI) [16].

## A. Synthetic dataset

We generate the simplicial complex shown in Fig. 1(a), which has  $N_0 = 8$  0-simplices,  $N_1 = 13$  1-simplices, and  $N_2 = 5$ , 2-simplices. So the boundary matrices are  $\mathbf{B}_1 \in \mathbb{R}^{8 \times 13}$  and  $\mathbf{B}_2 \in \mathbb{R}^{13 \times 5}$ . For this curated simplicial complex, the 0-simplices are categorized into 2-classes as shown in Fig. 2(a). Ground truth class labels for 0-simplices are based on their relationship with neighboring 0-simplices through 2-simplices. We compute the simplicial adjacency matrix  $\mathbf{A}_{0,2} \in \mathbb{R}^{8 \times 8}$  using (2), and obtain the clusters using simplicial spectral clustering (Algorithm 1). They are shown in Fig. 2(d). Given that it relies on adjacency through 2-simplices, it recovers ground truth. As a comparison, motif-based spectral clustering [2] (Fig. 2(c)) and graph spectral clustering (Fig. 2(b)) do not recover it. For motifs, it assumes that every triangle is filled, while spectral clustering ignores them. NMI between the obtained clusters and ground truth is reported in Fig. 4. The proposed method is optimal with NMI = 1.

## B. Real datasets

We also apply the proposed method on the Zachary Karate Club network [15], the Polbooks network [3], and a football network [3]; for details about the datasets see Table I. Although information about filled triangles is not directly available in these datasets, we report observations about what happens when we assume that some triangles are not filled.

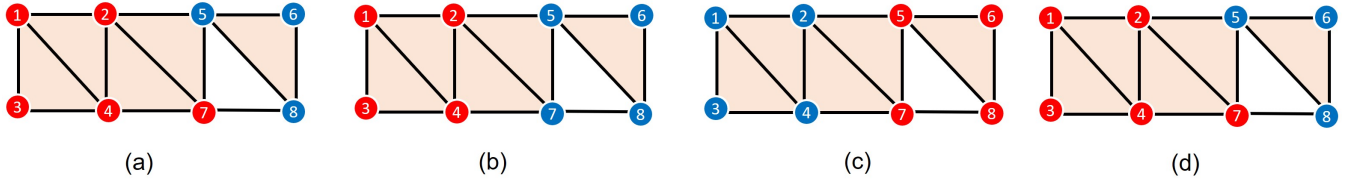


Fig. 2: Synthetic dataset. (a) Ground truth. (b) Communities from graph spectral clustering. (c) Communities from motif spectral clustering. (d) Communities from simplicial spectral clustering.

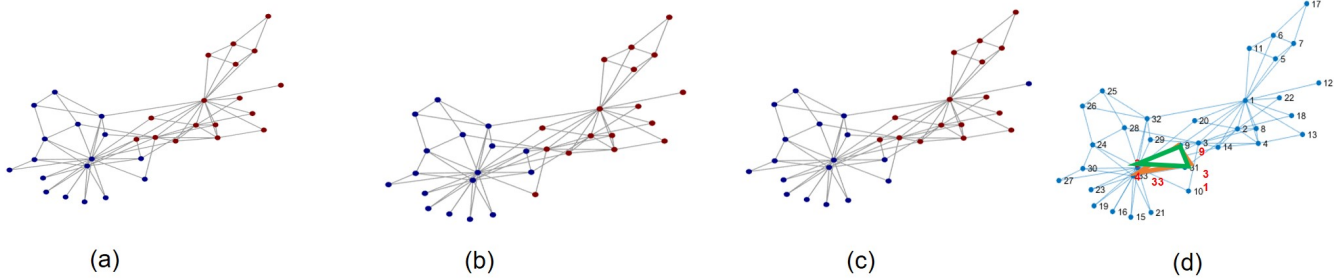


Fig. 3: Zachary Karate network data. (a) Ground truth network. (b) Communities obtained from motif spectral clustering. (c) Communities obtained from simplicial spectral clustering. (d) Removed triangles.

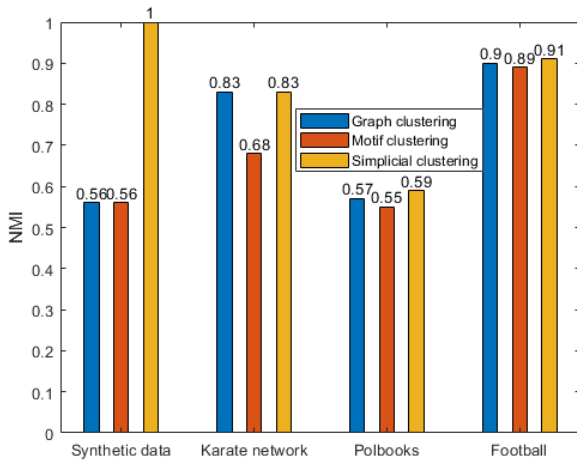


Fig. 4: Normalized mutual information.

1) *Zachary karate club network*: This dataset [15] is a well-known community detection dataset about the social relationships of members in a karate club, where the members belong to 2 groups within the club as shown by two different colors in Fig. 3(a). We follow the procedure from [2] to obtain the triangle motif adjacency matrix. The cluster assignments from motif spectral clustering are shown in Fig. 3(b), and NMI is reported in Fig. 4.

For the proposed simplicial spectral clustering, we study the impact of filled triangles with respect to (w.r.t.) the hollow ones. Towards that end, we conduct the following analysis: out of all triangles listed by the triangle adjacency matrix, we

remove a few triangles assuming they are open. We remove the triangles formed with edges  $\{9, 31\}$  and  $\{9, 34\}$  with members from the 2 ground truth groups. These triangles act as a bridge between the two groups (as shown in Fig. 3(d)). The proposed simplicial spectral clustering significantly improves NMI (see Fig. 4). This asserts our claim that assuming all the open triangles as filled overlooks the importance of filled triangles for graph partitioning.

2) *US Polbooks data* [3]: This network has 3 clusters. Hence we use motif clustering with multiple clusters as a baseline [2]. For simplicial spectral clustering, the method is: 1) Compute the eigenvectors of the normalized simplicial Laplacian matrix; 2) Collect the eigenvectors corresponding to the 3 smallest eigenvalues of  $\tilde{L}$ ; 3) Run  $K$ -means algorithm on obtained eigenvectors with  $K = 3$  to obtain the clusters.

The ground truth network results from motif clustering, and simplicial spectral clustering are in Figs. 5(a)-(c). As before, we assume a few triangles as hollow. The obtained NMI is reported in Fig. 4, where the proposed simplicial clustering outperforms the baselines.

3) *American Football network data* [3]: American football club network is a multicluster dataset with 12 communities, where we consider a few triangles as hollow. The simplicial spectral clustering for obtaining multicluster assignments is as before.

Figs. 6(a)-(c) show the ground truth network along with the cluster assignments from motif spectral clustering and simplicial spectral clustering. NMIs for these methods are reported in Fig. 4, where it can be seen that the proposed method achieves better performance at finding the communities.

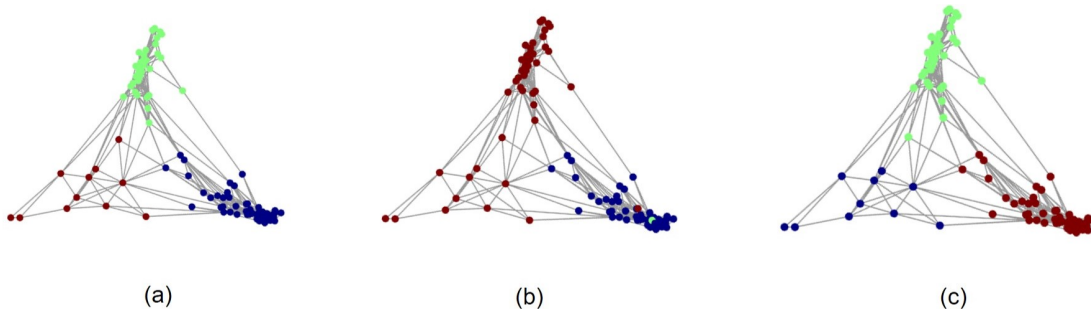


Fig. 5: US Polbooks data. (a) Ground truth network. (b) Communities obtained from motif spectral clustering. (c) Communities obtained from simplicial spectral clustering.

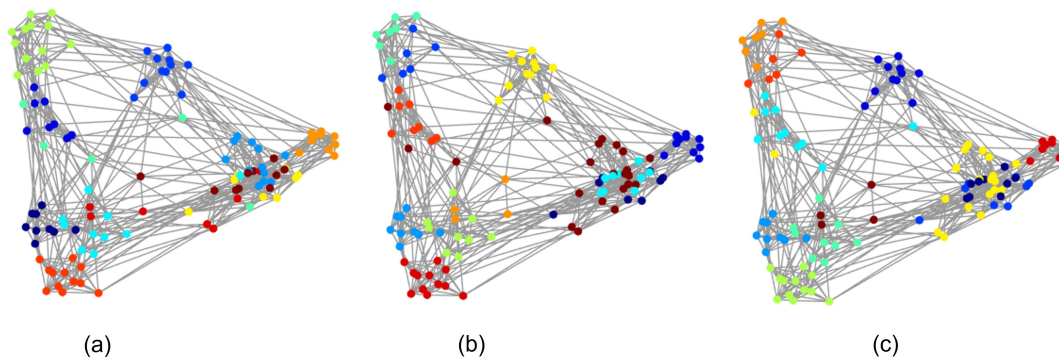


Fig. 6: Football network data. (a) Ground truth network. (b) Communities obtained from motif spectral clustering. (c) Communities obtained from simplicial spectral clustering.

## V. CONCLUSIONS

We proposed a simplicial conductance function to leverage higher-order network interactions while clustering. We defined a simplicial Laplacian operator that captures the relationship between the nodes through 2-simplices and developed a Cheeger inequality relating the second smallest eigenvalue of the proposed simplicial Laplacian matrix to the optimal simplicial conductance. Further, leveraging the Cheeger inequality, we developed a new simplicial spectral clustering algorithm, which was found to cluster networks better than edge cut-based and triangle motif-based spectral clustering methods while being able to distinguish filled and hollow triangles.

## REFERENCES

- [1] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010.
- [2] A. R. Benson, D. F. Gleich, and J. Leskovec, "Higher-order organization of complex networks," *Science*, vol. 353, no. 6295, pp. 163–166, 2016.
- [3] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.
- [4] Z. Lu, J. Wahlström, and A. Nehorai, "Community detection in complex networks via clique conductance," *Scientific reports*, vol. 8, no. 1, p. 5982, 2018.
- [5] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [6] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Advances in neural information processing systems*, vol. 14, 2001.
- [7] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, pp. 395–416, 2007.
- [8] M. E. Newman, "The structure of scientific collaboration networks," *Proceedings of the national academy of sciences*, vol. 98, no. 2, pp. 404–409, 2001.
- [9] M. S. Granovetter, "The strength of weak ties," *American journal of sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.
- [10] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, "Networks beyond pairwise interactions: Structure and dynamics," *Physics Reports*, vol. 874, pp. 1–92, 2020.
- [11] A. R. Benson, D. F. Gleich, and J. Leskovec, "Tensor spectral clustering for partitioning higher-order network structures," in *Proceedings of the 2015 SIAM International Conference on Data Mining*. SIAM, 2015, pp. 118–126.
- [12] C. E. Tsourakakis, J. Pachocki, and M. Mitzenmacher, "Scalable motif-aware graph clustering," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 1451–1460.
- [13] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, and J. Kleinberg, "Simplicial closure and higher-order link prediction," *Proceedings of the National Academy of Sciences*, vol. 115, no. 48, pp. E11 221–E11 230, 2018.
- [14] A. S. Bandeira, A. Singer, and D. A. Spielman, "A Cheeger inequality for the graph connection Laplacian," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 4, pp. 1611–1630, 2013.
- [15] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of anthropological research*, vol. 33, no. 4, pp. 452–473, 1977.
- [16] L. Ana and A. Jain, "Robust data clustering," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, 2003, pp. II–II.