ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

SCHOOL OF COMPUTER AND COMMUNICATION SCIENCES

Sephora MADJIHEUREM

Optional Semester Project

Learning with Hypergraphs

Supervisors:
Nathanaël PERRAUDIN
Nauman SHAHID

Responsible:
Prof. Pierre VANDERGHEYNST

June 10, 2016

Abstract

Hypergraphs have been shown to present several advantages over usual graphs to represent complex relationships among objects. In this study, we generalize the spectral graph theory to hypergraphs. We formally show how a hypergraph Laplacian can emerge from a gradient operator on hypergraphs, and use this formalism to build different Laplacians. We then use these operators to solve clustering and classification problems on different dataset and compare the results. The results suggest that the so-called normalized Laplacian leads to best performance for clustering and classification problems.

1 Introduction

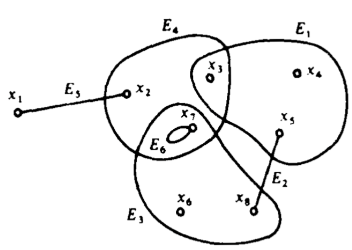
Previous studies have shown that spectral clustering methods are powerful techniques to partition simple graphs (Ng, Jordan, and Weiss, 2002; Shi and Malik, 2000). However, real-world data often present complex relationships among objects, and it is difficult to model these relationships with simple graphs. Hypergraphs, which are extensions of graphs, allow to represent such higher-order relations in data. They have been shown to perform well in several applications such as computational biology and computer vision (Tian, Hwang, and Kuang, 2009; Huang, Liu, and Metaxas, 2009; P.Ochs and T.Brox, 2012) and to perform better than simple graphs for classification on complex relational data (Zhou, Huang, and Schölkopf, 2007).

The aim of this work is to generalize the spectral graph theory to hypergaphs. To this end, we formalize the construction of the hypergraph Laplacian and explain how it can be used to solve partitioning problems. We finally apply our formalism in different clustering and classification problems.

The rest of this report is organized as follows: The notations used throughout the report are introduced in Section 2. Then, Section 3 presents the methodology to build the hypergraph Laplacian from a hyperedge derivative, and introduces the *non normalized*, *normalized* and *combinatorial* Laplacians. In Section 4, we show how the hypergraph Laplacian can be used to solve partitioning problems, and explain which Laplacian corresponds to which partitioning problem. Section 5 presents and discusses the results of two experiments; clustering and classification done on different datasets. Finally, we conclude and suggest future work in Section 6.

2 Preliminaries

A hypergraph is a generalization of a graph. While in simple graphs the edges are pairs of nodes, in hypergraphs the edges, know as hyperedges can connect any number of vertices. Figure 1 (from (Berge, 1984)) depicts an example of a hypergraph and its incidence matrix. They are proven to allow to better represent the complex relationships among objects than simple graphs (Zhou, Huang, and Schölkopf, 2007).



$$A = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{array} \begin{array}{c} \begin{array}{cccccc} E_1 & E_2 & E_3 & E_4 & E_5 & E_6 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Figure 1: Example of a hypergragh and its incidence matrix

We call $G = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ a *hypergraph* with the *vertex* set $\mathcal{V}$ and the *hyperedge* set $\mathcal{E}$ such that $\mathcal{E}$ is a family of subsets e of $\mathcal{V}$, and $\cup_{e\in\mathcal{E}} = \mathcal{V}$. The weight of a hyperedge $e \in \mathcal{E}$ is defined by $\mathcal{W}(e)$. The *degree of a vertex* $v \in \mathcal{V}$ is denoted

$d(v) = \sum_{e \in \mathcal{E} | v \in e} w(e)$. The *degree of a hyperedge* $e \in \mathcal{E}$ is denoted $\delta(e) = |e|$. The hypergraph $G = (\mathcal{V}, \mathcal{E}, W)$ can be represented by a $|\mathcal{V}| \times |\mathcal{E}|$ matrix H and a diagonal $|\mathcal{E}| \times |\mathcal{E}|$ matrix W such that the entry $h(v, e)$ of H is 1 if $v \in e$ and 0 otherwise and W is a matrix with the hyperedge weights on its diagonal.

Let D_v and D_e denote the diagonal matrices with the vertex and the hyperedges degrees respectively. The *adjacency matrix* A of G is then defined as $A = HWH^T - D_v$, where H^T denotes the transpose of H. We say that vertices u and v are *adjacent* if $u \in e$ and $v \in e$ for a hyperedge $e \in \mathcal{E}$.

A sequence of distinct vertices and hyperedges $v_1, e_1, v_2, e_2, \ldots, v_{k-1}, e_{k-1}$ such that $\{v_i, v_{i+1}\} \subseteq e_i$ for $1 \leq i \leq k-1$ is called a *hyperpath* from v_1 to v_k. In the remaining of this report, we assume that the considered hypergrahs are *connected*, that is, there is a hyperpath between every pair of vertices.

3 Laplacian of a hypergraph

Although we could directly define the Laplacian since only the Laplacian will be used in the following applications, we choose to emphasize the construction of the Laplacian more than the final result.

This section introduces a step-by-step construction of the hypergraph Laplacian. The Laplacian operator is defined as $\mathcal{L} : \mathbb{R}^{|\mathcal{V}|} \to \mathbb{R}^{|\mathcal{V}|} : \mathcal{L}f \mapsto -\text{div}\nabla f$, where ∇f is the gradient operator of a signal f on the vertices and $\text{div}g : \mathbb{R}^{|\mathcal{E}|} \to \mathbb{R}^{|\mathcal{V}|}$ is the divergence operator for a signal g on the hyperedges (the adjoint of the gradient). To construct the Laplacian of a hypergraph, we start by defining the gradient operator, then we derive the divergence of the hypergraph using the relationship $\langle \nabla f, g \rangle = \langle f, \text{div}g \rangle$. The Laplacian then simply follows from its definition.

When the edge derivative is a linear operator, the associated Laplacian is a symmetric positive semi-definite operator.

3.1 Non-normalized Laplacian

Using the formalism described previously we can construct the non-normalized Laplacian, by first defining the hyperedge derivative $\frac{\partial f}{\partial_e} : \mathcal{E} \to \mathbb{R}$ for a signal $f : \mathcal{V} \to \mathbb{R}$:

$$\frac{\partial f}{\partial_e}(e) = \sum_{u,v \in e} \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} (f(u) - f(v))$$

which is the weighted average of finite differences between the signal at pairs of adjacent vertices.

To get the gradient $\nabla f : \mathbb{R}^{|\mathcal{V}|} \to \mathbb{R}^{|\mathcal{E}|}$ of a hypergraph signal f, we stack all edge derivatives:

$$\nabla f = \left(\frac{\partial f}{\partial_e}(e) \right)_{e \in \mathcal{E}}$$

Using the relationship $\langle \nabla f, g \rangle = \langle f, \text{div}g \rangle$, the divergence operator $(\text{div}g)(u)$ is then defined as

$$\frac{1}{2} \sum_e \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \big(\sum_v \big(g(u, v) - g(v, u)\big)\big) h(e, v) h(e, u)$$

And the non-normalized Laplacian follows:

$$\mathcal{L} = D_v - HWD_e^{-1}H^T$$

3.2 Normalized Laplacian

Similarly, we can construct the normalized Laplacian using the following normalized hyperedge derivative: $\frac{\partial f}{\partial_e} : \mathcal{E} \to \mathbb{R}$ for a signal $f : \mathcal{V} \to \mathbb{R}$:

$$\frac{\partial f}{\partial_e}(e) = \sum_{u,v \in e} \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \right)$$

which corresponds to the weighted average of finite differences between the normalized signal at pairs of adjacent vertices.

The divergence operator $(\text{div}g)(u)$ is then defined as

$$\frac{1}{2\sqrt{d(u)}} \sum_e \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \big(\sum_v \big(g(u, v) - g(v, u)\big)\big) h(e, v) h(e, u)$$

And the normalized Laplacian follows:

$$\mathcal{L}_n = I - D_v^{-\frac{1}{2}} HWD_e^{-1}H^T D_v^{-\frac{1}{2}}$$

3.3 Combinatorial Laplacian

We also introduce an combinatorial Laplacian defined using the following hyperedge derivative: $\frac{\partial f}{\partial_e} : \mathcal{E} \to \mathbb{R}$ for a signal $f : \mathcal{V} \to \mathbb{R}$:

$$\frac{\partial f}{\partial_e}(e) = \sum_{u,v \in e} \sqrt{w(e)}(f(u) - f(v))$$

which corresponds to the sum of the weighted finite differences between the signal at pairs of adjacent vertices.
The divergence operator $(\text{div} g)(u)$ is then defined as

$$\frac{1}{2} \sum_e \sqrt{w(e)} \big(\sum_v \big(g(u,v) - g(v,u) \big) \big) h(e,v) h(e,u)$$

And the combinatorial Laplacian follows:

$$\mathcal{L}_A = D'_v - HWH^T$$

where D'_v is a diagonal matrix similar to D_v but with weights $w'(e) = w(e)\delta(e)$ instead.

The details of the computations of the different Laplacians can be found in appendix A.

4 Hypergraph partitioning

For a hypergraph $G = (\mathcal{V}, \mathcal{E}, W)$, a cut is a partition of $\mathcal{V}$ into two parts S and $\bar{S}$, where $\bar{S}$ denotes the complement of S. We say that a hyperedge e is cut if there exist vertices u and $v \in e$ such that $u \in S$ and $v \in \bar{S}$.

Given a non empty subset of vertices $S \subset \mathcal{V}$, define the hyperedge boundary ∂S of S to be a hyperedge set consisting of hyperedges which are cut, i.e. $\partial S := \{e \in \mathcal{E} | e \cap S \neq \emptyset, e \cap \bar{S} \neq \emptyset\}$, and the volume Vol$S$ of S to be the sum of the degrees of the vertices in S, i.e. Vol$S := \sum_{v \in S} d(v)$. The definition of the volume of ∂S is proposed by (Zhou, Huang, and Schölkopf, 2007) and is as follows:

$$\text{Vol}\partial S := \sum_{e \in \partial S} w(e) \frac{|e \cap S||e \cap \bar{S}|}{\delta(e)}$$

Using these definitions, we can define the hypergraph normalized cut Ncut(S) and the ratio cut Rcut(S) as

$$\text{Ncut}(S) := \text{Vol}\partial S \big(\frac{1}{\text{Vol}S} + \frac{1}{\text{Vol}\bar{S}} \big)$$

$$\text{Rcut}(S) := \text{Vol}\partial S \big(\frac{1}{|S|} + \frac{1}{|\bar{S}|} \big)$$

We can extend these definition to partitions and formalize the combinatorial optimization problems:

$$\underset{\emptyset \neq S_1, \ldots, S_K \subset \mathcal{V}}{\text{argmin}} \ \text{Npart}(S_1, S_2, \ldots, S_K) := \sum_{i=1}^{K} \frac{\text{Vol}\partial S_i}{\text{Vol}S_i} \tag{1}$$

$$\underset{\emptyset \neq S_1, \ldots, S_K \subset \mathcal{V}}{\text{argmin}} \ \text{Rpart}(S_1, S_2, \ldots, S_K) := \sum_{i=1}^{K} \frac{\text{Vol}\partial S_i}{|S_i|} \tag{2}$$

Where $\{S_1, S_2, \ldots, S_K\}$ form a partition of $\mathcal{V}$.

In their study, (Zhou, Huang, and Schölkopf, 2007) explain that minimizing the normalized cut (hence the normalized partition by extension) aims at finding a partition in which the connections among the vertices in the same cluster is dense while the connections between two clusters is sparse. It is yet to be determined the kind of application for which the ratio cut performs better.

4.1 Spectral partitioning

The optimization problems given by Eq. 1 and 2 are NP-complete (Berge, 1984) and can be relaxed into a real-valued problem. In this section, we propose a procedure to derive such relaxations using the Laplacian of the hypergraph.

For a hypergraph cut cut(S) and a hypergraph Laplacian $\mathcal{L}$ with associated derivative $\frac{\partial f}{\partial_e}$, we find that

$$\underset{f \in \mathbb{R}^{\mathcal{V}}}{\operatorname{argmin}} \frac{1}{2} \sum_{e \in \mathcal{E}} \big(\frac{\partial f}{\partial_e}(e)\big)^2 \tag{3}$$

Is the real-valued relaxation of a min hypergraph cut problem:

$$\underset{\emptyset \neq S \subset \mathcal{V}}{\operatorname{argmin}} \operatorname{cut}(S)$$

Moreover, it can be verified that

$$\sum_{e \in \mathcal{E}} \big(\frac{\partial f}{\partial_e}(e)\big)^2 = 2 f^T \mathcal{L} f \tag{4}$$

Thus, the solution to the optimization problem is an eigenvector of $\mathcal{L}$ associated with its smallest nonzero eigenvalue. We can use this result to solve the general partitioning problem:

$$\underset{\emptyset \neq S_1, \ldots, S_K \subset \mathcal{V}}{\operatorname{argmin}} \operatorname{part}(S_1, S_2, \ldots, S_k) := \underset{\emptyset \neq S_1, \ldots, S_K \subset \mathcal{V}}{\operatorname{argmin}} \sum_{i=1}^{K} \operatorname{cut}(S_i)$$

It can be achieved by constructing a matrix F containing k vectors f^i and checking that $(f^i)^T \mathcal{L} f^i = (F^T \mathcal{L} F)_{ii}$. Thus

$$\operatorname{part}(S_1, S_2, \ldots, S_k) = \sum_{i=1}^{K} (f^i)^T \mathcal{L} f^i = \sum_{i=1}^{K} (F^T \mathcal{L} F)_{ii} = Tr(F^T \mathcal{L} F) \tag{5}$$

where $Tr(M)$ denotes the trace of a matrix M. And the solution is then given by choosing a the first k eigenvectors of $\mathcal{L}$ with non-zero associated eigenvalues.

It appears that the relaxation of the normalized partitioning problem is obtained when $\mathcal{L}$ is the normalized Laplacian, whereas the relaxation of the ratio partitioning problem is obtained when $\mathcal{L}$ is the non normalized Laplacian. The details of these computations are given in B.

5 Experiments

We performed two learning experiments, clustering and classification, on two different types of datasets. In this section we first describe the datasets used, then present and discuss the results of each experiment.

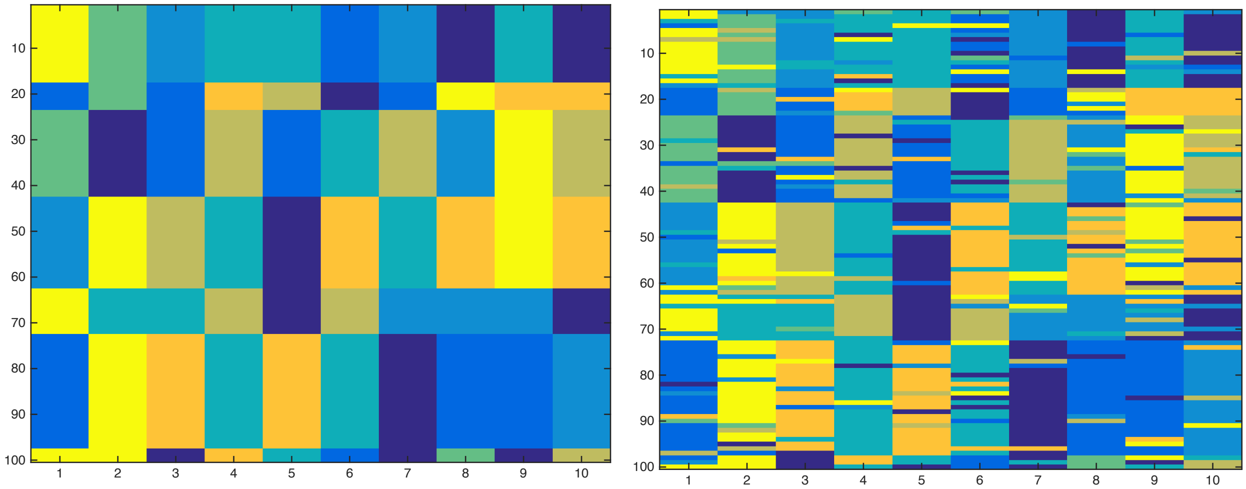
5.1 Datasets

Synthetic dataset: The first dataset D of size $N \times F$ was artificially created as follows: The rows of D are partitioned into K clusters of randomly chosen sizes (such that the sum of the size of the K clusters equals N). The value of all F attributes are set to be equal within a cluster (integers between 1 and $featureSpace$, also randomly selected). This results in a noiseless dataset consisting of K clusters. Figure 2(a) depicts such a dataset with $N = 100$, $K = 7$, $F = 10$ and $fetureSpace = 8$. We then added noise to this perfect data by randomly flipping some values. Figure 2(b) shows the dataset after addition of 20% noise.

The hypergraph was constructed as follows: for each attribute f, a hyperedge containing all the vertices that have the same value for feature f is created. For instance, in the dataset shown in Figure 2(a) for each column we construct a hyperedge for each color (e.g. 4 hyperedges were constructed for the first attribute: yellow, blue, light blue and green). We set the value of the weight of all hyperedges to 1.

We then construct the Laplacian using the matrices I, H, W, D_v and D_e. We embedded the points into Euclidean space by using the eigenvectors of the hypergraph Laplacian associated with the smallest eigenvalues. Figures 2(c) and 2(d) depict this embedding for the non normalized Laplacian whereas Figures 2(e) and 2(f) correspond to the embedding using the normalized Laplacian.

Zoo dataset: The second dataset used in the experiments is the *zoo dataset* (Lichman, 2013). It consists of 100 animals. Each instance in this dataset is described by 17 attributes. Each attribute takes only a small number of values, each corresponding to a specific category. The attributes include hair, feathers, eggs, milk, legs, tail, etc. The animals have been manually classified into 7 different categories. We used the same procedure as for the synthetic dataset to construct the hyperedges, and also embedded the points into Euclidean space by using the eigenvectors of the hypergraph Laplacian associated with the smallest eigenvalues. Figures 3(a) and 3(b) depict this embedding for the non normalized Laplacian whereas Figures 3(c) and 3(d) correspond to the embedding using the normalized Laplacian.



(a) Synthetic noiseless data

(b) Synthetic data with 10% noise.

	syn. data noiseless	syn. data 20% noise	syn. data 30% noise	zoo data
$\mathcal{L}$	0	0.0083	0.0465	0.3891
$\mathcal{L}_N$	0	0.0075	0.0404	0.1634
$\mathcal{L}_A$	0	0.1036	0.1880	0.5149

Table 1: Clustering error

5.2 Clustering

Using these embeddings (as well as the embedding using the combinatorial Laplacian), we performed clustering using the non normalized, normalized and combinatorial Laplacians by creating a matrix U having as columns the K (= number of clusters) eigenvectors with the smallest eigenvalues and clustering the rows of U into clusters $C_1, C_2, ..., C_K$ with *K-means* algorithm. Table 1 shows the average clustering error (over 20 trials) on different types of synthetic datasets and on the zoo dataset.

From Table 1, we notice that normalized clustering achieves systematically better results than the other methods. This observation is consistent with Figures 2 and 3, where we can visually notice that there is less overlapping of the clusters in the normalized embeddings. We can also note that the combinatorial clustering performs worst. The difference between the non normalized and the normalized clustering methods is far more significant on the zoo dataset. This suggests that when the clusters are very "clean", that is the graph is very regular and most vertices have approximately the same degree, such as in the artificially generated dataset (which is still very clean even with added noise), then there is not much difference between the ratio and the normalized partitioning. However, on "real" data, the normalized partitioning should be preferred over other methods.

5.3 Classification

We also used the hypergraphs and their Laplacian to solve a classification problem: given a subset denoted by an indicator matrix $M \in \mathbb{R}^{N \times F}$ of the data with known labels y (classes, represented by an integer), find the labels of the remaining of the dataset. This problem can be formulated using the Tikonov regularization (Golub, Hansen, and O'Leary, 1999):

$$\underset{x}{\operatorname{argmin}} \|Mx - y\|_2^2 + \tau x^T \mathcal{L} x$$

where τ is the regularization parameter.

For each of our datasets, we labelled each cluster by assigning an integer from 1 to K. We then randomly shuffle the data and remove the labels of 30% of the data. Using the remaining 70%, we apply the classification to recover the missing labels. In the zoo dataset, there is a class that is far more represented than the others (over 40% of the data is in the class "1 = *mammal*"), and the classification method resulted in poor results; namely all unlabelled data points where classified as this majority class. However, when the task is a simple binary classification (belongs to the majority class or not), then the classification method worked better. Tabel 2 reports the Balanced Error Rate (BER) (the average of the proportion of wrong classifications in each class), for the mutliclassification problem on synthetic data and the binary classification on the zoo data.

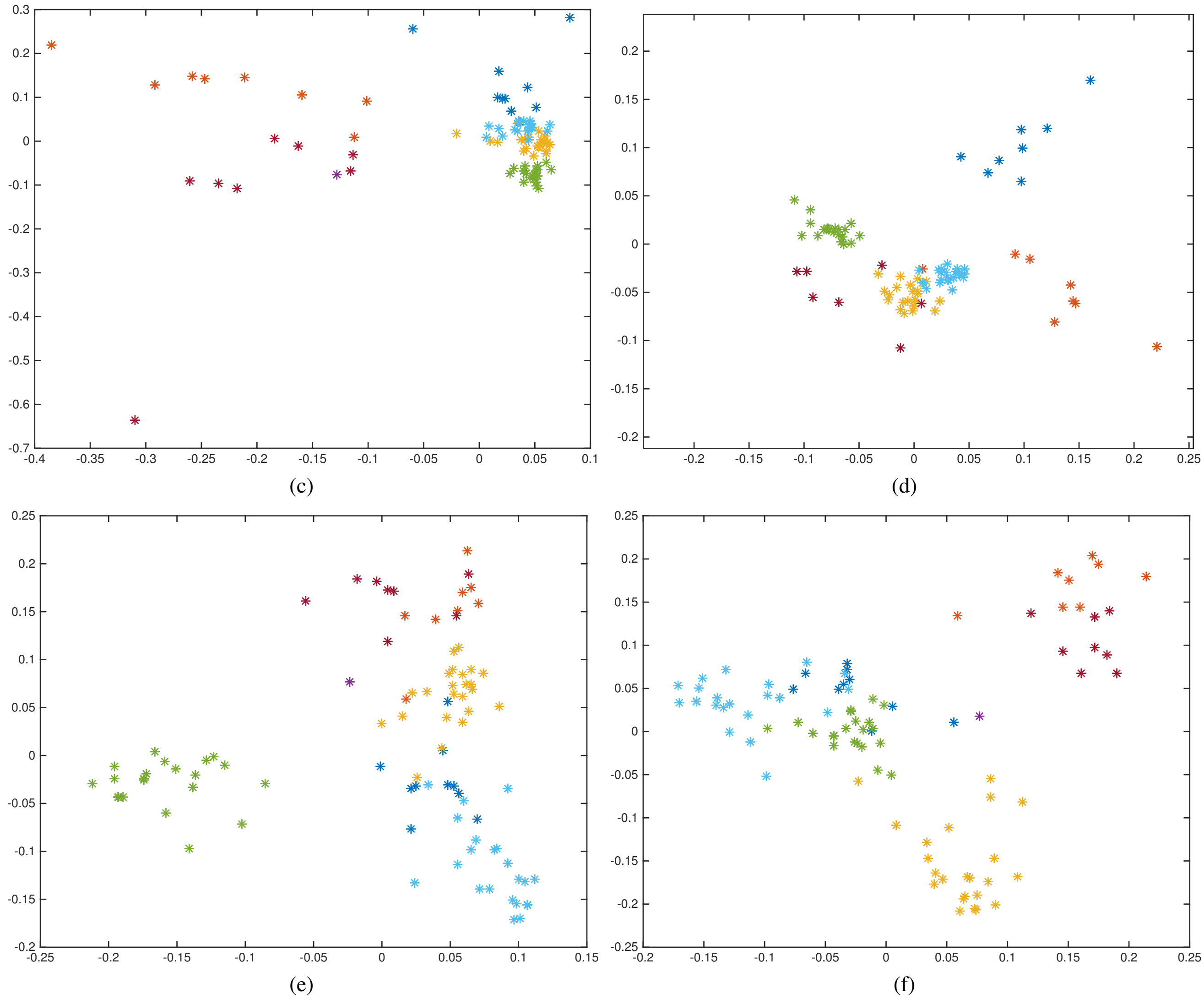


Figure 2: Embedding of the synthetic dataset (with 20% noise). The different colors correspond to different classes. (a) The eigenvectors of the the *non normalized Laplacian* with the 2nd and 3rd smallest eigenvalues . (b) The eigenvectors of the the *non normalized Laplacian* with the 3rd and 4th smallest eigenvalues. (c) The eigenvectors of the the *normalized Laplacian* with the 2nd and 3rd smallest eigenvalues . (d) The eigenvectors of the the *normalized Laplacian* with the 3rd and 4th smallest eigenvalues.

	syn. data noiseless	syn. data 20% noise	syn. data 30% noise	zoo data
$\mathcal{L}$	0	0.0892	0.1597	0.0304
$\mathcal{L}_N$	0	0.0878	0.1628	0.0304
$\mathcal{L}_A$	0.1	0.2404	0.2420	0.1208

Table 2: Classification BER (averaged over 20 trials)

Once again, we can notice that the method that uses the combinatorial Laplacian leads to higher error. It is however unclear from these experiments, which method should be preferred between the normalized and non normalized ones.

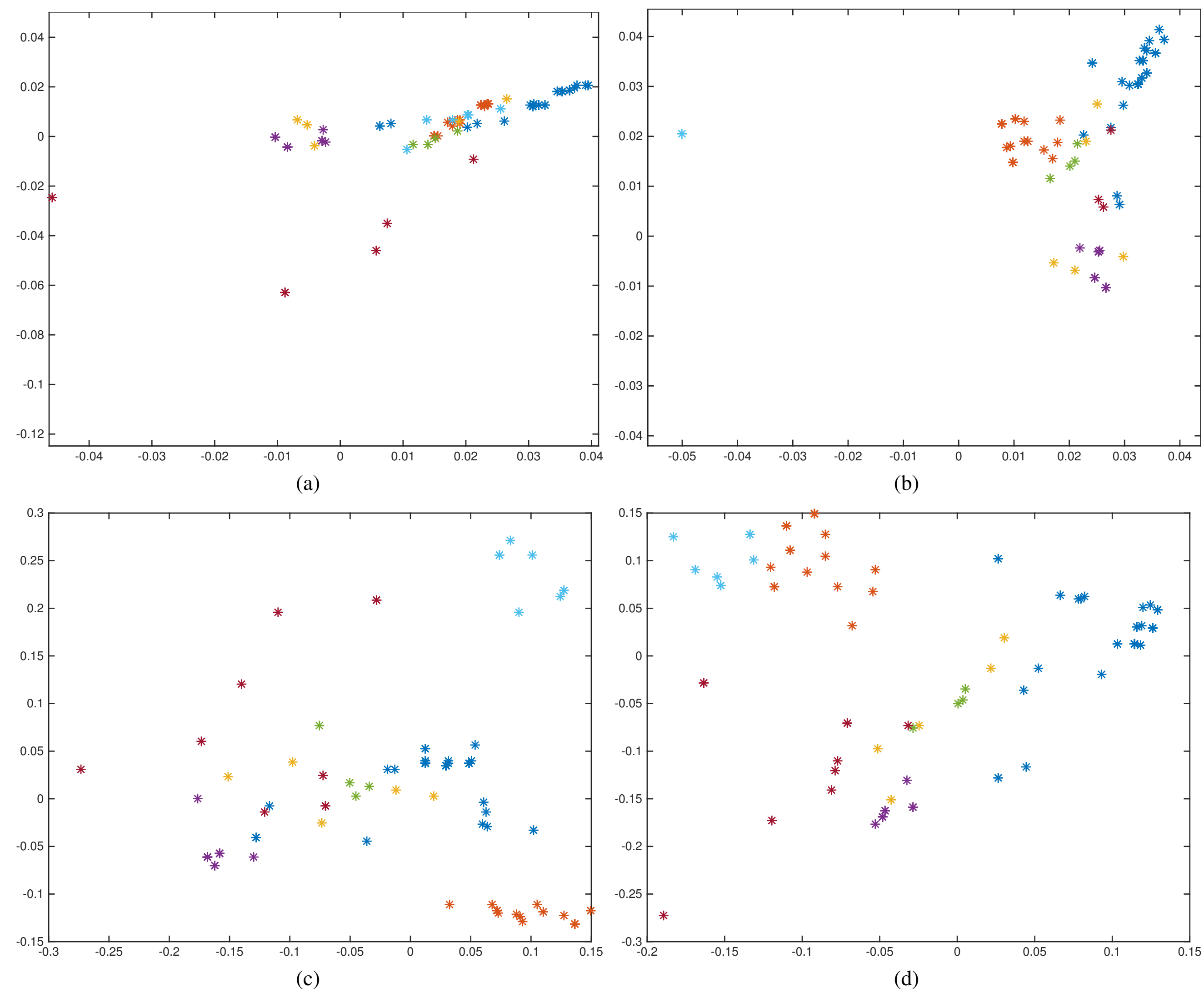


Figure 3: Embedding of the zoo dataset. The different colors correspond to different classes. (a) The eigenvectors of the the *non normalized Laplacian* with the 2nd and 3rd smallest eigenvalues . (b) The eigenvectors of the the *non normalized Laplacian* with the 3rd and 4th smallest eigenvalues. (c) The eigenvectors of the the *normalized Laplacian* with the 2nd and 3rd smallest eigenvalues . (d) The eigenvectors of the the *normalized Laplacian* with the 3rd and 4th smallest eigenvalues.

6 Conclusion and Future Work

Throughout this project, we have formalized a procedure to build Laplacians of hypergraphs from a linear derivative operator on the vertices. We then used this procedure to build the non normalized, normalized and combinatorial Laplacians for arbitrary hypergraphs. We then showed how two of these Laplacians can be used to solve NP-complete partitioning problems.

Using these defined operators, we suggested an embedding of the data represented by the hypergraphs and presented a visual representation of these embeddings on different types of datasets.

Finally, we ran several clustering and classification problems that make use of the Laplacians and compare the results. The results showed that for both clustering and classification, the normalized methods work best.

The topic of spectral clustering with hypergraphs can be further studied and explored. For example, future work could consist of investigating how to build hypergraphs from the data in an optimal manner. The question of which Laplacian to choose depending on the problem should also be further studied.

7 Acknowledgements

We thank Nathanaël Perraudin and Nauman Shahid for being very accessible supervisors and for providing guidance and expertise that greatly assisted the research.

References

Berge, C. 1984. *Hypergraphs: Combinatorics of Finite Sets*. North-Holland Mathematical Library. Elsevier Science.

Golub, G. H.; Hansen, P. C.; and O'Leary, D. P. 1999. Tikhonov regularization and total least squares. *SIAM J. Matrix Anal. Appl.* 21(1):185–194.

Huang, Y.; Liu, Q.; and Metaxas, D. N. 2009. Video object segmentation by hypergraph cut. In *CVPR*, 1738–1745. IEEE Computer Society.

Lichman, M. 2013. UCI machine learning repository.

Ng, A. Y.; Jordan, M. I.; and Weiss, Y. 2002. On spectral clustering: Analysis and an algorithm. In Dietterich, T. G.; Becker, S.; and Ghahramani, Z., eds., *Advances in Neural Information Processing Systems 14*. MIT Press. 849–856.

P.Ochs, and T.Brox. 2012. Higher order motion models and spectral clustering. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*.

Shi, J., and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(8):888–905.

Tian, Z.; Hwang, T.; and Kuang, R. 2009. A hypergraph-based learning algorithm for classifying gene expression and arraycgh data with prior knowledge. *Bioinformatics* 25(21):2831–2838.

Zhou, D.; Huang, J.; and Schölkopf, B. 2007. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in Neural Information Processing Systems 19*, 1601–1608. Cambridge, MA, USA: Max-Planck-Gesellschaft.

A Laplacian constructions

A.1 Non normalized Laplacian

Define the **hyperedge derivative** $\frac{\partial f}{\partial_e} : \mathcal{E} \to \mathbb{R}$ for a signal $f : \mathcal{V} \to \mathbb{R}$ as :

$$\frac{\partial f}{\partial_e}(e) = \sum_{u,v \in e} \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}(f(u) - f(v))$$

Then obtain the **gradient** $\nabla_{HG} f : \mathbb{R}^{|\mathcal{V}|} \to \mathbb{R}^{|\mathcal{E}|}$ of a hypergraph signal f by stacking all hyperedge derivatives:

$$\nabla_{HG} f = \left(\sum_{u,v \in e} \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}(f(u) - f(v)) \right)_{e \in \mathcal{E}}$$

The **divergence**, the adjoint of the gradient operator, satisfies $\langle \nabla f, g_e \rangle = \langle f, \text{div} g_e \rangle$ for a signal g on the hyperedges $g_e(e) = \big(g(u,v)\big)_{\{u,v\} \in e}$.

$$\begin{aligned}
\langle \nabla f, g \rangle &= \sum_e \sum_{\{u,v\} \subseteq e} \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}(f(u) - f(v)) g(u,v) \\
&= \frac{1}{2} \sum_e \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \sum_{u \in e} \sum_{v \in e} (f(u) g(u,v) - f(v) g(u,v)) \\
&= \frac{1}{2} \sum_e \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \big(\sum_{u \in e} \sum_{v \in e} f(u) g(u,v) - \sum_{u \in e} \sum_{v \in e} f(v) g(u,v) \big) \\
&= \frac{1}{2} \sum_e \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \big(\sum_{u \in e} \sum_{v \in e} f(u) g(u,v) - \sum_{v \in e} \sum_{u \in e} f(u) g(v,u) \big) \\
&= \frac{1}{2} \sum_{u \in e} f(u) \sum_e \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \big(\sum_{v \in e} g(u,v) - \sum_{v \in e} g(v,u) \big) \\
&= \frac{1}{2} \sum_u f(u) \sum_e \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \big(\sum_v g(u,v) h(e,v) - \sum_v g(v,u) h(e,v) \big) h(e,u)
\end{aligned}$$

Thus the divergence is defined as:

$$(\text{div} g)(u) = \frac{1}{2} \sum_e \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \big(\sum_v \big(g(u,v) - g(v,u) \big) \big) h(e,v) h(e,u)$$

We now derive the **hypergraph Laplacian**: $\mathcal{L} : \mathbb{R}^{|\mathcal{V}|} \to \mathbb{R}^{|\mathcal{V}|}$ defined as: $\mathcal{L} f \mapsto -\text{div} \nabla f$

$$\begin{aligned}
\mathcal{L}f(u) &= -\mathrm{div}\nabla f\\
&= \frac{1}{2}\sum_e \big(\sum_v \big(\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}(f(u)-f(v)) - \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}(f(v)-f(u))\big)\big)h(e,v)h(e,u)\\
&= \frac{1}{2}\big(\sum_e\sum_v \frac{w(e)}{\delta(e)}(f(u)-f(v) - \sum_e\sum_v \frac{w(e)}{\delta(e)}(f(v)-f(u))\big)h(e,v)h(e,u)\\
&= \frac{1}{2}\big(\sum_e\sum_v \frac{w(e)}{\delta(e)}(f(u)-f(v) + \sum_e\sum_v \frac{w(e)}{\delta(e)}(f(u)-f(v))\big)h(e,v)h(e,u)\\
&= \sum_e\sum_v \frac{w(e)}{\delta(e)}(f(u)-f(v))h(e,v)h(e,u)\\
&= \sum_e\sum_v \frac{w(e)}{\delta(e)}f(u)h(e,v)h(e,u) - \sum_e\sum_v \frac{w(e)}{\delta(e)}f(v)h(e,v)h(e,u)\\
&= f(u)\sum_e \frac{w(e)}{\delta(e)}h(e,u)\underbrace{\sum_v h(e,v)}_{\delta(e)} - \sum_v f(v)\sum_e h(e,v)\frac{w(e)}{\delta(e)}h(e,u)\\
&= f(u)\sum_e w(e)h(e,u) - \sum_v f(v)\sum_e h(e,v)\frac{w(e)}{\delta(e)}h(e,u)
\end{aligned}$$

And the corresponding Laplacian follows:

$$\mathcal{L} = D_{\mathcal{V}} - A'$$

where $A' = HWD_e^{-1}H^T$

A.2 Normalized Laplacian

Define the **hyperedge derivative** $\frac{\partial f}{\partial_e}:\mathcal{E}\to\mathbb{R}$ for a signal $f:\mathcal{V}\to\mathbb{R}$ as :

$$\frac{\partial f}{\partial_e}(e) = \sum_{u,v\in e}\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}\left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}}\right)$$

Then obtain the **gradient** $\nabla_{HG}f:\mathbb{R}^{|\mathcal{V}|}\to\mathbb{R}^{|\mathcal{E}|}$ of a hypergraph signal f by stacking all hyperedge derivatives:

$$\nabla_{HG}f = \left(\sum_{u,v\in e}\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}\left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}}\right)\right)_{e\in\mathcal{E}}$$

The **divergence**, the adjoint of the gradient operator, satisfies $\langle\nabla f, g_e\rangle = \langle f, \mathrm{div}g_e\rangle$ for a signal g on the hyperedges $g_e(e) = \big(g(u,v)\big)_{\{u,v\}\in e}$

$$\begin{aligned}
\langle\nabla f, g\rangle &= \sum_e\sum_{\{u,v\}\subseteq e}\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}\left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}}\right)g(u,v)\\
&= \frac{1}{2}\sum_e\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}\sum_{u\in e}\sum_{v\in e}\big(\frac{f(u)}{\sqrt{d(u)}}g(u,v) - \frac{f(v)}{\sqrt{d(v)}}g(u,v)\big)\\
&= \frac{1}{2}\sum_e\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}\big(\sum_{u\in e}\sum_{v\in e}\frac{f(u)}{\sqrt{d(u)}}g(u,v) - \sum_{u\in e}\sum_{v\in e}\frac{f(v)}{\sqrt{d(v)}}g(u,v)\big)\\
&= \frac{1}{2}\sum_e\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}\big(\sum_{u\in e}\sum_{v\in e}\frac{f(u)}{\sqrt{d(u)}}g(u,v) - \sum_{v\in e}\sum_{u\in e}\frac{f(u)}{\sqrt{d(u)}}g(v,u)\big)\\
&= \frac{1}{2}\sum_{u\in e}\frac{f(u)}{\sqrt{d(u)}}\sum_e\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}\big(\sum_{v\in e}g(u,v) - \sum_{v\in e}g(v,u)\big)\\
&= \frac{1}{2}\sum_u\frac{f(u)}{\sqrt{d(u)}}\sum_e\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}}\big(\sum_v g(u,v)h(e,v) - \sum_v g(v,u)h(e,v)\big)h(e,u)
\end{aligned}$$

Thus the divergence is then defined as:

$$(\text{div}g)(u) = \frac{1}{2\sqrt{d(u)}} \sum_e \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \Big(\sum_v \big(g(u,v) - g(v,u) \big) \Big) h(e,v)h(e,u)$$

We now derive the **hypergraph normalized Laplacian**: $\mathcal{L} : \mathbb{R}^{|\mathcal{V}|} \to \mathbb{R}^{|\mathcal{V}|}$ defined as usual by: $\mathcal{L}f \mapsto -\text{div}\nabla f$

$$\begin{aligned}
\mathcal{L}f(u) &= -\text{div}\nabla f \\
&= \frac{1}{2\sqrt{d(u)}} \sum_e \Big(\sum_v \Big(\frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \Big(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \Big) - \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \frac{\sqrt{w(e)}}{\sqrt{\delta(e)}} \Big(\frac{f(v)}{\sqrt{d(v)}} - \frac{f(u)}{\sqrt{d(u)}} \Big) \Big) \Big) h(e,v)h(e,u) \\
&= \frac{1}{2\sqrt{d(u)}} \Big(\sum_e \sum_v \frac{w(e)}{\delta(e)} \Big(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} - \sum_e \sum_v \frac{w(e)}{\delta(e)} \Big(\frac{f(v)}{\sqrt{d(v)}} - \frac{f(u)}{\sqrt{d(u)}} \Big) \Big) h(e,v)h(e,u) \\
&= \frac{1}{2\sqrt{d(u)}} \Big(\sum_e \sum_v \frac{w(e)}{\delta(e)} \Big(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} + \sum_e \sum_v \frac{w(e)}{\delta(e)} \Big(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(u)}} \Big) \Big) h(e,v)h(e,u) \\
&= \frac{1}{\sqrt{d(u)}} \sum_e \sum_v \frac{w(e)}{\delta(e)} \Big(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \Big) h(e,v)h(e,u) \\
&= \frac{1}{\sqrt{d(u)}} \sum_e \sum_v \frac{w(e)}{\delta(e)} \frac{f(u)}{\sqrt{d(u)}} h(e,v)h(e,u) - \frac{1}{\sqrt{d(u)}} \sum_e \sum_v \frac{w(e)}{\delta(e)} \frac{f(v)}{\sqrt{d(v)}} h(e,v)h(e,u) \\
&= \frac{f(u)}{d(u)} \sum_e \frac{w(e)}{\delta(e)} h(e,u) \underbrace{\sum_v h(e,v)}_{\delta(e)} - \sum_v \frac{f(v)}{d(v)} \sum_e \frac{w(e)}{\delta(e)} h(e,v)h(e,u) \\
&= \frac{f(u)}{d(u)} \sum_e w(e)h(e,u) - \sum_v f(v) \sum_e \frac{1}{\sqrt{d(v)}} h(e,u) \frac{w(e)}{\delta(e)} h(e,v) \frac{1}{\sqrt{d(v)}}
\end{aligned}$$

And the normalized Laplacian follows:

$$\mathcal{L}_N = I - \Theta$$

where I is the identity matrix and $\Theta = D_v^{-\frac{1}{2}} H W D_e^{-1} H^T D_{\mathcal{V}}^{-\frac{1}{2}}$

A.3 Alternative Laplacian

Define the **hyperedge derivative** $\frac{\partial f}{\partial_e} : \mathcal{E} \to \mathbb{R}$ for a signal $f : \mathcal{V} \to \mathbb{R}$ as :

$$\frac{\partial f}{\partial_e}(e) = \sum_{u,v \in e} \sqrt{w(e)} (f(u) - f(v))$$

Then obtain the associated **gradient** $\nabla_{HG} f : \mathbb{R}^{|\mathcal{V}|} \to \mathbb{R}^{|\mathcal{E}|}$ of a hypergraph signal f by stacking all hyperedge derivatives:

$$\nabla_{HG} f = \left(\sum_{u,v \in e} \sqrt{w(e)} (f(u) - f(v)) \right)_{e \in E}$$

The **divergence**, the adjoint of the gradient operator, satisfies $\langle \nabla f, g_e \rangle = \langle f, \text{div} g_e \rangle$ for a signal g on the hyperedges $g_e(e) = \big(g(u,v)\big)_{\{u,v\} \in e}$.

$$
\begin{aligned}
\langle \nabla f, g \rangle &= \sum_e \sum_{\{u,v\} \subseteq e} \sqrt{w(e)}(f(u) - f(v))g(u,v) \\
&= \frac{1}{2} \sum_e \sqrt{w(e)} \sum_{u \in e} \sum_{v \in e} (f(u)g(u,v) - f(v)g(u,v)) \\
&= \frac{1}{2} \sum_e \sqrt{w(e)} \big(\sum_{u \in e} \sum_{v \in e} f(u)g(u,v) - \sum_{u \in e} \sum_{v \in e} f(v)g(u,v) \big) \\
&= \frac{1}{2} \sum_e \sqrt{w(e)} \big(\sum_{u \in e} \sum_{v \in e} f(u)g(u,v) - \sum_{v \in e} \sum_{u \in e} f(u)g(v,u) \big) \\
&= \frac{1}{2} \sum_{u \in e} f(u) \sum_e \sqrt{w(e)} \big(\sum_{v \in e} g(u,v) - \sum_{v \in e} g(v,u) \big) \\
&= \frac{1}{2} \sum_u f(u) \sum_e \sqrt{w(e)} \big(\sum_v g(u,v)h(e,v) - \sum_v g(v,u)h(e,v) \big) h(e,u)
\end{aligned}
$$

Thus the divergence is defined as:

$$
(\text{div} g)(u) = \frac{1}{2} \sum_e \sqrt{w(e)} \big(\sum_v \big(g(u,v) - g(v,u)\big)\big) h(e,v)h(e,u)
$$

We now derive the **hypergraph alternative Laplacian**: $\mathcal{L} : \mathbb{R}^{|\mathcal{V}|} \to \mathbb{R}^{|\mathcal{V}|}$ defined as usual by: $\mathcal{L}f \mapsto -\text{div} \nabla f$

$$
\begin{aligned}
\mathcal{L}f(u) &= -\text{div} \nabla f \\
&= \frac{1}{2} \sum_e \big(\sum_v \big(\sqrt{w(e)}\sqrt{w(e)}(f(u) - f(v)) - \sqrt{w(e)}\sqrt{w(e)}(f(v) - f(u)) \big) \big) h(e,v)h(e,u) \\
&= \frac{1}{2} \big(\sum_e \sum_v w(e)(f(u) - f(v) - \sum_e \sum_v w(e)(f(v) - f(u)) \big) h(e,v)h(e,u) \\
&= \frac{1}{2} \big(\sum_e \sum_v w(e)(f(u) - f(v) + \sum_e \sum_v w(e)(f(u) - f(v)) \big) h(e,v)h(e,u) \\
&= \sum_e \sum_v w(e)(f(u) - f(v))h(e,v)h(e,u) \\
&= \sum_e \sum_v w(e)f(u)h(e,v)h(e,u) - \sum_e \sum_v w(e)f(v)h(e,v)h(e,u) \\
&= f(u) \sum_e w(e)h(e,u) \underbrace{\sum_v h(e,v)}_{\delta(e)} - \sum_v f(v) \underbrace{\sum_e w(e)h(e,v)h(e,u)}_{\mathbf{A}(u,v)}
\end{aligned}
$$

And the corresponding alternative Laplacian follows:

$$
\mathcal{L}_A = D'_v - A
$$

where D'_v is a diagonal matrix similar to D_v but using weights $w'(e) = w(e)\delta(e)$ instead and $A = HWH^T$

B Spectral Hypergraph Partitioning

Given a hypergraph $G = (\mathcal{V}, \mathcal{E}, w)$ with Laplacian $\mathcal{L}$ and corresponding hyperedge derivative $\frac{\partial f}{\partial_e}$,

$$
\underset{f \in \mathbb{R}^{\mathcal{V}}}{\text{argmin}} \frac{1}{2} \sum_{e \in \mathcal{E}} \big(\frac{\partial f}{\partial_e}(e) \big)^2 = \underset{f \in \mathbb{R}^{\mathcal{V}}}{\text{argmin}} f^T \mathcal{L} f \tag{6}
$$

Is the real-valued relaxation of a min hypergraph cut problem:

$$
\underset{S \subset \mathcal{V}}{\text{argmin}}\, c(S)
$$

1. **Claim:** When $\mathcal{L}$ is the normalized Laplacian $\mathcal{L}_N$ we have that

$$\underset{f\in\mathbb{R}^{\mathcal{V}}}{\operatorname{argmin}} \frac{1}{2}\sum_{e\in\mathcal{E}}\sum_{\{u,v\}\subseteq e}\frac{w(e)}{\delta(e)}\left(\frac{f(u)}{\sqrt{d(u)}}-\frac{f(v)}{\sqrt{d(v)}}\right)^2 = \underset{f\in\mathbb{R}^{\mathcal{V}}}{\operatorname{argmin}} f^T\mathcal{L}_N f \tag{7}$$

is the relaxation of the optimal normalized cut :

$$\underset{\emptyset\neq S\subset\mathcal{V}}{\operatorname{argmin}} \mathrm{Vol}\partial S\big(\frac{1}{\mathrm{Vol}S}+\frac{1}{\mathrm{Vol}\bar{S}}\big)$$

2. **Claim:** When $\mathcal{L}$ is the non normalized Laplacian $\mathcal{L}$ we have that

$$\underset{f\in\mathbb{R}^{\mathcal{V}}}{\operatorname{argmin}} \frac{1}{2}\sum_{e\in\mathcal{E}}\sum_{\{u,v\}\subseteq e}\frac{w(e)}{\delta(e)}(f(u)-f(v))^2 = \underset{f\in\mathbb{R}^{V}}{\operatorname{argmin}} f^T\mathcal{L}f \tag{8}$$

is the relaxation of the optimal ratio cut :

$$\underset{\emptyset\neq S\subset V}{\operatorname{argmin}} \mathrm{Vol}\partial S\big(\frac{1}{|S|}+\frac{1}{|\bar{S}|}\big)$$

Proof:

1. We first show that Eq. 6 holds for the non normalized Laplacian:

$$\begin{aligned}
&\frac{1}{2}\sum_{e\in\mathcal{E}}\sum_{\{u,v\}\subseteq e}\big(\frac{w(e)}{\delta(e)}(f(u)-f(v))^2\big)\\
&=\frac{1}{2}\sum_{e\in\mathcal{E}}\sum_{\{u,v\}\subseteq e}\big(\frac{w(e)}{\delta(e)}(f(u)^2+f(v)^2-2f(u)f(v))\big)\\
&=\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\big(\frac{w(e)}{\delta(e)}(f(u)^2-f(u)f(v)))\big)h(e,u)h(e,v)\\
&=\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\frac{w(e)}{\delta(e)}f(u)^2h(e,u)h(e,v)-\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\frac{w(e)}{\delta(e)}f(u)f(v)h(e,u)h(e,v)\\
&=\sum_{e\in\mathcal{E}}\sum_{u\in e}w(e)f(u)^2h(e,u)\sum_{v\in e}\frac{h(e,v)}{\delta(e)}-\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\frac{w(e)}{\delta(e)}f(u)f(v)h(e,u)h(e,v)\\
&=\sum_{e\in\mathcal{E}}\sum_{u\in e}w(e)f(u)^2h(e,u)-\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\frac{w(e)}{\delta(e)}f(u)f(v)h(e,u)h(e,v)\\
&=\sum_{u\in e}f(u)^2\sum_{e\in\mathcal{E}}w(e)h(e,u)-\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\frac{w(e)}{\delta(e)}f(u)f(v)h(e,u)h(e,v)\\
&=\sum_{u\in e}f(u)^2d(u)-\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}f(u)h(e,u)\frac{w(e)}{\delta(e)}h(e,v)f(v)\\
&=f^TD_{\mathcal{V}}f-f^THWD_e^{-1}H^Tf\\
&=f^T(D_v-HWD_e^{-1}H^T)f\\
&=f^t\mathcal{L}f
\end{aligned}$$

2. We now similarly show that Eq. 6 holds for the normalized Laplacian:

$$
\begin{aligned}
&\frac{1}{2}\sum_{e\in\mathcal{E}}\sum_{\{u,v\}\subseteq e}\big(\frac{w(e)}{\delta(e)}\Big(\frac{f(u)}{\sqrt{d(u)}}-\frac{f(v)}{\sqrt{d(v)}}\Big)^2\big)\\
&=\sum_{e\in\mathcal{E}}\sum_{\{u,v\}\subseteq e}\big(\frac{w(e)}{\delta(e)}\big(\frac{f(u)}{d(u)}^2-\frac{f(u)f(v)}{\sqrt{d(u)d(v)}}\big)\big)\\
&=\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\big(\frac{w(e)}{\delta(e)}\big(\frac{f(u)}{d(u)}^2-\frac{f(u)f(v)}{\sqrt{d(u)d(v)}}\big)\big)h(e,u)h(e,v)\\
&=\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\frac{w(e)}{\delta(e)}\frac{f(u)}{d(u)}^2h(e,u)h(e,v)-\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\frac{w(e)}{\delta(e)}\frac{f(u)f(v)}{\sqrt{d(u)d(v)}}\big)h(e,u)h(e,v)\\
&=\sum_{e\in\mathcal{E}}\sum_{u\in e}\frac{w(e)f(u)^2h(e,u)}{d(u))}\sum_{v\in e}\frac{h(e,v)}{\delta(e)}-\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\frac{w(e)f(u)f(v)h(e,u)h(e,v)}{\delta(e)\sqrt{d(u)d(v)}}\\
&=\sum_{u\in e}f(u)^2\sum_{e\in\mathcal{E}}\frac{w(e)h(e,u)}{d(u)}-\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\frac{w(e)f(u)f(v)h(e,u)h(e,v)}{\delta(e)\sqrt{d(u)d(v)}}\\
&=\sum_{u\in e}f(u)^2-\sum_{e\in\mathcal{E}}\sum_{u\in e}\sum_{v\in e}\frac{w(e)f(u)f(v)h(e,u)h(e,v)}{\delta(e)\sqrt{d(u)d(v)}}\\
&=f^Tf-f^TD_v^{-\frac{1}{2}}HWD_e^{-1}H^TD_v^{-\frac{1}{2}}f\\
&=f^T(I-D_v^{-\frac{1}{2}}HWD_e^{-1}H^TD_v^{-\frac{1}{2}})f\\
&=f^T\mathcal{L}_Nf
\end{aligned}
$$

3. To see that Eq. 2 and 8 are in fact the real-valued relaxations of the normalized cut and the ratio cut respectively, observe that if we represent the i^{th} cluster by an indicator vector f^i such that

$$
f^i_u=\begin{cases}a_i & \text{if } u\in S\\ 0 & \text{if } u\in\bar{S}\end{cases}
$$

for some $a_i\neq 0$ then

$$
(f^i)^T\mathcal{L}_Bf^i=\frac{1}{2}\sum_{e\in\mathcal{E}}\sum_{\{u,v\}\subseteq e}(f^i_u-f^i_u)^2\frac{w(e)}{\delta(e)}=a_i^2\text{Vol}(\partial S_i)
$$

$$
(f^i)^TD_vf^i=\frac{1}{2}\sum_{e\in\mathcal{E}}\sum_{\{u,v\}\subseteq e}f_u^2d(u)=a_i^2\text{Vol}(S_i)
$$

$$
(f^i)^Tf^i=\frac{1}{2}\sum_{e\in\mathcal{E}}\sum_{\{u,v\}\subseteq e}f_u^2=a_i^2|S_i|
$$

and thus we have in the normalized case that

$$
\sum_{i=1}^{K}\frac{\text{Vol}\partial\text{S}_i}{\text{VolS}_i}=\sum_{i=1}^{K}\frac{(f^i)^T\mathcal{L}f^i}{(f^i)^TD_vf^i}
$$

The relaxed problem is to minimize $\frac{(f^i)^T\mathcal{L}_Bf^i}{(f^i)^TD_vf^i}$ subject to $(f^i)^TD_vI=0$. Let $y^i=D_v^{\frac{-1}{2}}f^i$. Then,

$$
\frac{(f^i)^T\mathcal{L}f^i}{(f^i)^TD_vf^i}=\frac{(y^i)^TD_v^{\frac{-1}{2}}\mathcal{L}D_v^{\frac{-1}{2}}y^i}{(y^i)^Ty^i}
$$

And we know from resutls in linear algebra that the above expression is minimized when y^i is an eigenvector corresponding to the second smallest eigenvalue of $\mathcal{L}_N=D^{\frac{-1}{2}}\mathcal{L}D^{\frac{-1}{2}}$.
Similarly for the non normalized case we have,

$$\sum_{i=1}^{K} \frac{\mathrm{Vol}\partial \mathrm{S}_i}{|S_i|} = \sum_{i=1}^{K} \frac{(f^i)^T \mathcal{L} f^i}{(f^i)^T f^i}$$

which is minimized subject to $(f^i)^T I = 0$ when f^i is an eigenvector corresponding to the second smallest eigenvalue of $\mathcal{L}$.