A Spectral Framework for a Class of Undirected Hypergraphs

Thomas Ricatte, Gemma Garriga, Rémi Gilleron, Marc Tommasi

A Spectral Framework for a Class of Undirected Hypergraphs

Thomas Ricatte thomas.ricatte@sap.com
SAP France
157-159 rue Anatole France
92300 Levallois-Perret, France

Gemma Garriga gemma.garriga@inria.fr
Rémi Gilleron remi.gilleron@univ-lille3.fr
Marc Tommasi marc.tommasi@univ-lille3.fr
INRIA Lille Nord Europe
Parc Scientifique de la Haute Borne
40, Av. Halley
59650 Villeneuve d'Ascq - France

Editor:

Abstract

We extend the graph spectral framework to a new class of undirected hypergraphs with bipartite hyperedges. A bipartite hyperedge is a pair of disjoint nodesets in which every node is associated with a weight. A bipartite hyperedge can be viewed as a relation between two teams in which every node has a weighted contribution. Undirected hypergraphs generalize over undirected graphs. Consistently with the case of graphs, we define the notions of hypergraph Laplacian and hypergraph kernel (as the Moore-Penrose pseudoinverse of a hypergraph Laplacian). Therefore, smooth labeling of nodes and hypergraph regularization methods can be performed. Contrary to the graph case, we show that the class of hypergraph Laplacians is closed by pseudoinversion (thus it is also the class of hypergraphs kernels) and is closed by convex linear combination. Closure properties allow us to define (hyper)graph combinations and operations that remain in the class of undirected hypergraphs. We exhibit a subclass of signed graphs that can be associated with hypergraphs in a constructive way and, thus, can inherit the properties and semantics of the hypergraphs. We claim that undirected hypergraphs open the way to solve new learning tasks and model new problems based on set similarity or dominance.

Keywords: Graph Spectral Theory, Undirected Hypergraphs, Graph Laplacian, Graph Kernel, Resistance Distance.

1. Introduction

In most of real-life problems, data are linked by various relationships giving large and complex information networks. This is particularly the case for data in the Web, social networks or databases. Mathematical abstraction of such networks are graphs and many problems that exploit the network structure of data like classification, finding groups, propagation, or relation prediction are therefore problems on graphs. Graph modeling provides also readable ways to represent or visualize large amount of data and their interactions. Moreover, any set of data instances can be easily transformed into a graph (which exhibits a manifold on data instances): each node represents a data point and edges correspond to relationships of similarity among the points. From the machine learning point of view, algorithms have been devised for graph-like data (or manifolds) and many of them rely on graph spectral theory, see Fouss et al. (2007); Von Luxburg (2007); Zhou et al. (2005); Zhu et al. (2003) among others.

Graph spectral theory allows to define a linear cost model for real-valued node functions such as scoring functions or labeling functions, characterizing the *smoothness* of real-valued node functions f, i.e., the propensity of f to take close values on nodes that are linked in the graph. This is done by using the *smoothness semi-norm* defined by $\|f\|_\Delta = f^T \Delta f$, where Δ is the Laplacian matrix of the graph defined by $\Delta = D - W$ where W and D are respectively the adjacency matrix and the diagonal degree matrix of the graph. Note that Δ can be equivalently put in the form $\Delta = G^T G$ where G is the gradient matrix of the graph. The regularization model introduced by the spectral framework is also closely related to the theory of positive semidefinite kernels. Indeed, as described by Saerens et al. (2004), the Moore-Penrose pseudoinverse of a graph Laplacian Δ, denoted by $\Delta^\dagger$, defines a graph kernel (i.e., a kernel on the nodes of the graph) and the associated kernel distance is equal, when the graph is connected, to the commute-time distance (time taken by a random walker to go from a starting node i to a target node j and come back). Thus, $\Delta^\dagger$ is intricately linked with the notion of random walk on the graph. Note that $\Delta^\dagger$ is also a reproducing kernel of the space $\mathrm{Null}(\Delta)^\perp$ (space orthogonal to the null space of Δ), equipped with the scalar product $f, g \to f^T \Delta g$ (see Herbster et al., 2005).

However, real networks are complex networks with data content associated with nodes or with different relationships between nodes. Modeling such networks towards graph based learning raises important and difficult challenges. One of this important challenges is to deal efficiently with the heterogeneity of the relationships. A first solution is to consider the problem from a semantic point of view and forget the notion of distance based on walks between nodes (see for example Shervashidze et al. (2011) that considers the semantic neighbourhood of the nodes). Another solution is to regard the network as the aggregation of several information layers or views and build algorithms that are able to learn from these multiple views (see for example Zhou et al., 2007b, 2008; Wang and Zhou, 2010; He and Lawrence, 2011).

In the specific case where the goal is to learn from several graphs that share the same set of vertices, an interesting option is to combine the different layers in a new one and use it to apply graph-based learning methods. Thus, the problem is solved by looking for the most efficient combination. At the end, the semantic of this combination will provide interesting interpretations of the role of each layer.

A well-studied approach is to work on the Laplacian matrices and leverage the fact that linear convex combination of Laplacians is a Laplacian. This idea was used with success in Tsuda et al. (2005) and Sindhwani et al. (2005). Zhou and Burges (2007) proposes a slightly different approach based on the convex combination of markov chains which reduces however to a convex combination of graph Laplacians in the undirected case.

Another line of research that we consider in this paper is to combine graph kernels. This idea promoted by Herbster et al. (2005) and Argyriou et al. (2006) allows us to leverage efficient algorithms coming from the theory of multiple kernel learning (see for example Rakotomamonjy et al., 2008; Xu et al., 2010; Tian et al., 2011). Moreover, a combination of graph kernels can be interpreted as a combination of the commute-time distances in the original graphs. Note that many kernel-based algorithms are looking for the best convex linear combination of kernels (see Rakotomamonjy et al., 2008). However, the convex linear combination of graph kernels is not a graph kernel in the general case, i.e., the pseudoinverse of this combination is not a graph Laplacian. Because of this issue, there may be no direct way to define a graph semantic for the combined kernel.

In this paper, we follow an original line of research based on the claim that the convex linear combination of several graph kernels always possesses a graph-like semantic. Indeed, a convex linear combination of graph kernel K is still a positive semi-definite matrix and its null space still contains the space of constant functions. For these two reasons, K can be still be used to define a smoothness semi-norm $\|f\| = f^T K^{\dagger} f$ (constant functions are perfectly smooth with this definition). Following this idea, we define a larger class of kernels which is closed by the pseudoinverse operation and is closed by convex linear combination. Moreover, and this is the main contribution of the paper, we attach a comprehensive semantic to this larger class of kernels. Indeed, we introduce a class of undirected hypergraphs with bipartite hyperedges. A bipartite hyperedge (simply called hyperedge in the rest of this paper) is an unordered pair of disjoint sets of nodes. Additionally, every node participates to a hyperedge with a weight so that an equilibrium property is established between both ends of the hyperedge. A hyperedge can be viewed as a relation between two teams of nodes in which every node has a weighted contribution to its team. It can be noted that an undirected graph is a special case of undirected hypergraph. And, similarly to the case of graphs, we introduce a smoothness measure in order to model similarities between teams (sets of nodes). For this, we extend the spectral framework to our class of hypergraphs defining the notions of hypergraph gradient, hypergraph Laplacian and hypergraph kernel. We show that the class of hypergraph kernels coincides with the extended class of kernels defined above, and thus, we prove that the class of hypergraph kernels is closed by the pseudoinverse operation and convex linear combination. Also, we embed a specific class of hypergraphs (strongly connected hypergraphs generalizing connected graphs) in a natural Riemannian structure. The Riemannian structure opens new opportunities to compute means, medians and centroids on sets of hypergraphs and thus permits to define new combination methods.

Hypergraphs have close relationships with signed graphs. We show how to define weights for node pairs in a hypergraph and we are able to prove that the hypergraph Laplacian can be defined with the classical formula $D - W$ where W is the weight matrix for node pairs (non-positive in general). This allows to define a signed graph from a hypergraph and to define a new Laplacian for (a subclass of) signed graphs (see Kunegis et al., 2010, for Laplacians of signed graphs). We also define a distance for hypergraphs which coincide with the usual commute-time distance when the hypergraph is a graph.

Note that many graph applications consider the normalized version of the Laplacian operator where the input function is normalized by the square root of the node degrees, i.e., the normalized gradient is defined by $GD^{-1/2}$ and the normalized Laplacian is defined by $D^{-1/2}\Delta D^{-1/2} = I - D^{-1/2}WD^{-1/2}$ where G is the (unnormalized) gradient and D the degree matrix. The normalized gradient and the normalized Laplacian can be defined for hypergraphs since the degree matrix is always positive. We do not investigate here the properties of normalized gradient and Laplacian and keep this line open for future research.

Related work on hypergraphs The most studied class of hypergraphs has been introduced by Berge (1989) in order to represent interactions between multiple entities. In this model, an hyperedge is a set of nodes. This model, also known as *clique hypergraph* model, has been extensively studied from an algorithmic point of view (connectivity, colorability, ...) and has led more recently to extensions of the spectral methods for graphs (Zhou et al., 2007a). But, it should be noted that the spectral extension for this class of hypergraphs can be obtained from the reduction of a hypergraph into a graph replacing every hyperedge by an uniform clique of edges where every edge weight in the clique is based on the weight and the size of the original hyperedge.

An other line of research popularized by Gallo et al. (1993) considers the notion of directed hypergraphs. These hypergraphs were first introduced to formalize the notion of functional dependency between objects and are also known as AND/OR hypergraphs. The main idea is to formalize a AND dependency $x_1 \wedge x_2 \rightarrow x_3$ by a directed link between the sets $\{1, 2\}$ and $\{3\}$. Conversely a directed link between $\{3\}$ and $\{1, 2\}$ is used to represent an OR expression $x_3 \rightarrow x_1 \vee x_2$. In the most general version of this framework, a directed hyperedge is a directed link between two disjoint subsets of vertices called respectively *tail* and *head*. This can be regarded as the concatenation of a AND edge followed by an OR edge (see for example Gallo et al., 1993). This model has been extensively studied from an algorithmic perspective (see for instance Cambini et al., 1997) and has led to multiple applications in the field of Boolean satisfiability (see Gallo et al., 1998). In terms of structure, our hypergraphs can be viewed as an undirected version of these objects even if we attach to the hyperedges a quite different semantic based on the notion of set similarity. As far as we know, this class of directed hypergraphs has not been studied from the machine learning point of view and no attempt was made to define a spectral framework for these objects.

2. Generalizing the Classes of Graph Kernels and Graph Laplacians

In this section, we recall definitions for undirected graphs and briefly review important concepts such as graph Laplacians and graph kernels. Graph Laplacians allow to define a global smoothness model for graphs that is the foundation of the so-called graph spectral framework (see Von Luxburg, 2007). Then, we discuss properties and limitations of the classes of graph Laplacians and graph kernels. Last, we propose a relaxation of these classes with important properties such as the closure by the pseudoinverse operation and the closure by convex linear combination.

2.1 Graph Kernels and Graph Laplacians

An *undirected graph* $\mathbf{g} = (N, E)$ is a set of nodes N together with a set of undirected edges E where an undirected edge is an unordered pair of nodes. The number of nodes is denoted by n and we suppose an arbitrary numbering of nodes, that is we fix N to be the set $\{1, \ldots, n\}$. The number of edges is denoted by $p = |E|$. In this paper we consider *undirected weighted graphs* in which each edge carries a weight. Undirected weighted graphs are represented by their symmetric weighted *adjacency matrix* W where $W_{i,j} = W_{j,i}$

is the weight between the two nodes i and j. When there is no edge between i and j then $W_{i,j} = 0$ and we assume that there is no self-loop in the graph, that is $\forall i \in N$, $W_{i,i} = 0$. Every adjacency matrix of an undirected weighted graph belongs to the half space $\mathcal{A} = \{W \in \mathbb{R}^{n\times n} \mid W \text{ is symmetric ; } W \geq 0\}$.

We define the *unnormalized graph gradient* to be the linear operator that maps every real-valued node function f to a real-valued edge function $\mathrm{grad}(f)$ defined by

$$\forall i,j \in N \quad \mathrm{grad}(f)(i,j) = \sqrt{W_{i,j}}(f(j) - f(i)) \ , \tag{1}$$

where an arbitrary orientation of all undirected edges has been chosen. We denote by $G \in \mathbb{R}^{p\times n}$ the matrix of $\mathrm{grad}(\cdot)$. The unnormalized gradient of every constant node function is the zero valued edge function. This can be written as $\mathrm{Span}(\mathbf{1}) \subseteq \mathrm{Null}(G)$ where $\mathrm{Span}(\mathbf{1})$ is the linear space spanned by the vector $\mathbf{1} = (1,\dots,1)^T$ and $\mathrm{Null}(G)$ is the null space of the unnormalized gradient, i.e., the set of so-called harmonic functions on the graph. $\mathrm{Null}(G)$ is spanned by the indicator functions of the graph components (see Von Luxburg, 2007). Consequently the number of components of the graph G is $\dim(\mathrm{Null}(G))$. And, when the graph is connected, then $\mathrm{Null}(G) = \mathrm{Span}(\mathbf{1})$.

The term $|\,\mathrm{grad}(f)(i,j)|^2$ measures the *smoothness* of f over the edge $\{i,j\}$: it is small when $f(i)$ and $f(j)$ are close and it is related to the strength of the link $\{i,j\}$ through the term $\sqrt{W_{i,j}}$. The *global smoothness* of a real-valued node function f over a graph $\mathbf{g}$ is defined by $\|Gf\|^2 = (Gf)^T(Gf) = f^TG^TGf$.

The square $n \times n$ real valued matrix $\Delta = G^TG$ is defined to be the *unnormalized Laplacian*[1] of the graph $\mathbf{g}$. Then, the smoothness of a real-valued node function f over a graph $\mathbf{g}$ can be written as $f^T\Delta f$. It should be noted that the Laplacian Δ does not depend on the arbitrary orientation of the edges. Also, it is easy to prove that, for every graph $\mathbf{g}$, the Laplacian Δ is symmetric ($\Delta = \Delta^T$) and positive semidefinite (denoted by $\Delta \succeq 0$). Let $\mathbf{g}$ be an undirected weighted graph with adjacency matrix W, the diagonal degree matrix of the graph $\mathbf{g}$ is denoted by D and defined by $D_{i,i} = \sum_{1\leq j\leq n} W_{i,j}$ for every $1 \leq i \leq n$ and $D_{i,j} = 0$ for every $i \neq j$. Then, it can be shown that the Laplacian matrix, as defined above, can equivalently be defined by $\Delta = D - W$ (see for example Von Luxburg, 2007). Note that $f \to f^T\Delta f$ only defines a semi-norm since $\mathbf{1}^T\Delta\mathbf{1} = 0$.

Example 1 *We consider the graph* $\mathbf{g}$ *over* $N = \{1,2,3\}$ *with edges* $\{1,2\}$ *of weight 1 and* $\{2,3\}$ *of weight 2. The matrices* W*,* D*,* G *(for the arbitrary orientation* $(1,2)$ *and* $(3,2)$ *for the two edges), and* Δ *are:*

$$W = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix} ; \ D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix} ; \ G = \begin{pmatrix} -1 & 1 & 0 \\ 0 & \sqrt{2} & -\sqrt{2} \end{pmatrix} ; \ \Delta = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 3 & -2 \\ 0 & -2 & 2 \end{pmatrix} \ .$$

Let us consider an undirected weighted graph $\mathbf{g}$ with Laplacian Δ, and let us consider the Moore-Penrose pseudoinverse, denoted by $\Delta^\dagger$, of the Laplacian matrix Δ. From the properties of the pseudoinverse operator and because Δ is symmetric positive semidefinite, $\Delta^\dagger$ is also symmetric positive semidefinite. The matrix $\Delta^\dagger$ is defined to be the *graph kernel*

1. the matrix G^T can be seen as the Hermitian adjoint of the gradient matrix G, and it is the matrix of the divergence operator $-div$. Thus, the Laplacian matrix is the matrix of $-div(grad)$

of $\mathbf{g}$ and it is related with random walks in $\mathbf{g}$ associated with the probability transition matrix $D^{-1}W$. Indeed, let us recall that the *commute-time distance* c_{ij} between two nodes i and j of a graph $\mathbf{g}$ is defined by: $c_{ij} = m(i|j) + m(j|i)$, where $m(i|j)$ is the expected time taken by a random walker to travel from node i to j in $\mathbf{g}$. Then, a classical result from spectral graph theory from Klein and Randić (1993) (see also Fouss et al., 2007) states that, for any connected undirected weighted graph $\mathbf{g}$ with kernel matrix $\Delta^\dagger$, we have

$$c_{i,j} = \mathrm{Vol}(G)\left(\Delta^\dagger_{i,i} + \Delta^\dagger_{j,j} - 2\Delta^\dagger_{i,j}\right) \quad , \tag{2}$$

where $vol(G) = \sum_i D_{i,i}$ is the volume of the graph. The graph kernel allows to embed the nodes in an Euclidean space in which the Euclidean distance is the commute-time between nodes in the graph.

2.2 The classes of Graph Kernels and Graph Laplacians

Let us consider the class

$$\mathcal{GL} = \{M \in \mathbb{R}^{n\times n} \mid M = M^T,\ \mathbf{1} \in \mathrm{Null}(M),\ \mathrm{extradiag}\,(M) \leq 0\} \quad , \tag{3}$$

where $\mathrm{extradiag}\,(M)$ is the matrix M with the diagonal removed.

Proposition 1 *The class $\mathcal{GL}$ is the set of unnormalized graph Laplacians.*

Proof For $M \in \mathcal{GL}$, let us consider $W = -\mathrm{extradiag}\,(M)$. By definition of $\mathcal{GL}$, $W \geq 0$, thus W is a graph adjacency matrix. Let us denote by D the degree matrix of W. For every $i \in N$, $W_{i,i} = 0$ and we have

$$D_{i,i} = \sum_{\substack{1\leq j\leq n\\ j\neq i}} W_{i,j} = -\sum_{\substack{1\leq j\leq n\\ j\neq i}} M_{i,j} \quad .$$

Since $M\mathbf{1} = 0$ by definition of $\mathcal{GL}$, then, for every $i \in N$, $\sum_{1\leq j\leq n} M_{i,j} = 0$, and thus $D_{i,i} = M_{i,i}$. That is $M = D - W$, and therefore M is a graph Laplacian.

Conversely if $M = D - W$ is a graph Laplacian, then $M = M^T$ and $M\mathbf{1} = D\mathbf{1} - W\mathbf{1} = 0$ by definition of the degree matrix. Finally, $\mathrm{extradiag}\,(M) = -W \leq 0$, which concludes the proof. ■

The *class $\mathcal{GK}$ of graph kernels* is the set of all matrices which are pseudoinverse of some matrix in $\mathcal{GL}$. Since the pseudoinverse operator is involutive, we can write equivalently $\mathcal{GK} = \{M \in \mathbb{R}^{n\times n} \mid M^\dagger \in \mathcal{GL}\}$. Then, the classes $\mathcal{GL}$ and $\mathcal{GK}$ satisfy

Proposition 2

(i) the set $\mathcal{GL}$ of graph Laplacians is closed by convex linear combination but is not closed by linear combination,

(ii) the set $\mathcal{GK}$ of graph kernels is not closed by convex linear combination,

(iii) the sets $\mathcal{GL}$ and $\mathcal{GK}$ are not closed under pseudoinverse, and

(iv) $\mathcal{GL} \cap \mathcal{GK} \neq \emptyset$

Proof

(i) Let us consider $\Delta_1, \Delta_2 \in \mathcal{GL}$ and let us denote by W_1 and W_2 the corresponding adjacency matrices. For every $\alpha \in [0,1]$, $W = \alpha W_1 + (1-\alpha) W_2$ is symmetric and satisfies $W \geq 0$ and W is therefore the adjacency matrix of an undirected graph. Then $\Delta = \alpha\Delta_1 + (1-\alpha)\Delta_2$ is the unnormalized Laplacian associated with W and thus $\Delta \in \mathcal{GL}$. Hence, $\mathcal{GL}$ is closed by convex linear combination.

When $\alpha \notin [0,1]$, then W, defined as above, can have negative weights, thus $\Delta \notin \mathcal{GL}$. Therefore, $\mathcal{GL}$ is not closed by linear combination.

(ii) Let us consider the matrices Δ_1 and Δ_2 in $\mathcal{GL}$ defined by

$$\Delta_1 = \begin{pmatrix} 2 & 0 & -1 & -1 \\ 0 & 1 & -1 & 0 \\ -1 & -1 & 2 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_2 = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{pmatrix}.$$

The matrices $\Delta_1^\dagger$ and $\Delta_2^\dagger$ are in $\mathcal{GK}$, but, let us consider the matrix $\frac{1}{2}(\Delta_1^\dagger + \Delta_2^\dagger)$, its pseudoinverse is

$$\left[\frac{1}{2}(\Delta_1^\dagger + \Delta_2^\dagger)\right]^\dagger = \frac{1}{13}\begin{pmatrix} 16 & 2 & -12 & -6 \\ 2 & 10 & -8 & -4 \\ -12 & -8 & 22 & -2 \\ -6 & -4 & -2 & 12 \end{pmatrix},$$

which is not in $\mathcal{GL}$ since some extra-diagonal elements are non-negative. Hence $\mathcal{GK}$ is not closed by convex linear combination.

(iii) Let us consider the Laplacian $\Delta_1 \in \mathcal{GL}$ from (ii). Its pseudoinverse is

$$\Delta_1^\dagger = \frac{1}{8}\begin{pmatrix} 3 & -3 & -1 & 1 \\ -3 & 7 & 1 & -5 \\ -1 & 1 & 3 & -3 \\ 1 & -5 & -3 & 7 \end{pmatrix},$$

which is not in $\mathcal{GL}$ since some extra-diagonal elements are non-negative. Hence $\mathcal{GL}$ is not closed by pseudoinverse.

The class $\mathcal{GK}$ is not closed by pseudoinverse. Indeed, let us consider $\Delta_1^\dagger$ in $\mathcal{GK}$, its pseudoinverse is $(\Delta_1^\dagger)^\dagger = \Delta_1$. And, Δ_1 does not belong to $\mathcal{GK}$ since its pseudoinverse $\Delta_1^\dagger \notin \mathcal{GL}$.

(iv) Let us consider the matrices Δ in $\mathcal{GL}$ and $\Delta^\dagger$ in $\mathcal{GK}$ defined by

$$\Delta = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 3 & -2 \\ 0 & -2 & 2 \end{pmatrix} \quad \Delta^\dagger = \frac{1}{6}\begin{pmatrix} 3 & -1 & -2 \\ -1 & 1 & 0 \\ -2 & 0 & 2 \end{pmatrix}.$$

We notice that $\Delta^\dagger \in \mathcal{GL}$, which concludes the proof: $\mathcal{GL} \cap \mathcal{GK} \neq \emptyset$. ∎

2.3 A Generalized Class of Graph Kernels and Graph Laplacians

As said in the introduction, our objective is to define a class which is closed by convex linear combination. For this, we replace in the definition of the class $\mathcal{GL}$ (see Equation ((3))) the condition extradiag$(M) \leq 0$ by the weaker condition that M is positive semidefinite. Let us recall that a matrix $M \in \mathbb{R}^{n\times n}$ is positive semidefinite, denoted by $M \succeq 0$, if for any $x \in \mathbb{R}^n$ we have $x^T M x \geq 0$. This leads to the definition of a new class denoted by $\mathcal{HL}$ and defined by

$$\mathcal{HL} = \{M \in \mathbb{R}^{n\times n} \mid M = M^T,\ \mathbf{1} \in \mathrm{Null}(M),\ M \succeq 0\}\ . \tag{4}$$

We will show in the next section that the class $\mathcal{HL}$ is the class of Laplacians (and kernels) of objects that we call undirected hypergraphs. We prove here that the class $\mathcal{HL}$ satisfies:

Proposition 3

(i) the class $\mathcal{HL}$ is closed under pseudoinverse,

(ii) the class $\mathcal{HL}$ contains the class of graph Laplacians $\mathcal{GL}$ and the class of graph kernels $\mathcal{GK}$,

(iii) the class $\mathcal{HL}$ is closed by convex linear combination but is not closed by linear combination.

Proof (i) The pseudoinverse operation preserves the symmetry and the semidefiniteness property since it does not modify the sign of the eigenvalues. Moreover, for every real-valued symmetric matrix M, $\mathrm{Null}(M^\dagger) = \mathrm{Null}(M)$ so $\mathbf{1} \in \mathrm{Null}(M^\dagger)$.

(ii) $\mathcal{HL}$ contains $\mathcal{GL}$ because graph Laplacians are positive semidefinite. From (i), $\mathcal{HL}$ is closed by pseudoinverse then it contains also $\mathcal{GK} = \{M \in \mathbb{R}^{n\times n} \mid M^\dagger \in \mathcal{GL}\}$.

(iii) Let us consider $M_1, M_2 \in \mathcal{HL}$ and a real α, and let us denote by M_α the matrix $\alpha M_1 + (1-\alpha)M_2$. M_α is symmetric and $M_\alpha \mathbf{1} = 0$. Now, when $\alpha \in [0,1]$, for every $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^T M_\alpha \mathbf{x} = \alpha \mathbf{x}^T M_1 \mathbf{x} + (1-\alpha)\mathbf{x}^T M_2 \mathbf{x} \geq 0$ because $M_1 \succeq 0$ and $M_2 \succeq 0$. That is $M_\alpha \succeq 0$, hence $\mathcal{HL}$ is closed by convex linear combination. However, it is easy to see that there exist $M_1, M_2, \mathbf{x}$ and $\alpha \in \mathbb{R}$ such that $\mathbf{x}^T M_\alpha \mathbf{x} < 0$. Thus, $\mathcal{HL}$ is not closed by linear combination, which concludes the proof. ∎

Because of the relation between graph kernels and the commute-time distance in connected graphs (see Equation ((2))), we also consider the extended class of connected graph Laplacians defined by

$$\mathcal{HL}^+ = \{M \in \mathbb{R}^{n\times n} \mid M = M^T,\ \mathrm{Null}(M) = \mathrm{Span}(\mathbf{1}),\ M \succeq 0\}\ .$$

It is worth noting that $\mathcal{HL}^+$ contains the Laplacians of all connected graphs and can be embedded in a complete Riemannian structure as shown in the next section.

2.4 A Riemannian Structure for $\mathcal{HL}^+$

As mentioned above, the class $\mathcal{HL}$ is closed by convex linear combination. This is not the case for the class $\mathcal{HL}^+$ which is a strict subspace of $\mathcal{HL}$. The geodesic in the Euclidean space $\mathbb{R}^{n\times n}$ is a "straight line" but the Euclidean geometry does not fit the class $\mathcal{HL}^+$. This is illustrated with a simple metaphor for $\mathbb{R}^2$ in Figure 1 below where the class $\mathcal{HL}^+$ is a curved space which can be seen as a boundary space of $\mathcal{HL}$.

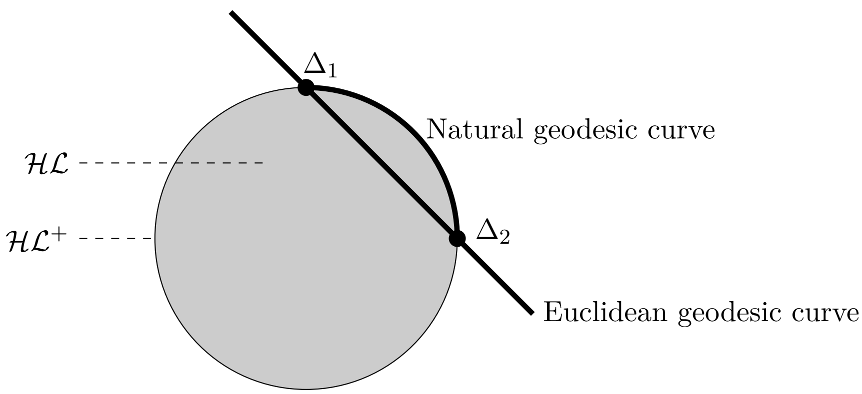


Figure 1: Convex space $\mathcal{HL}$ and non-convex space $\mathcal{HL}^+$. Euclidean geometry of $\mathbb{R}^{n\times n}$ is not suited for $\mathcal{HL}$ and $\mathcal{HL}^+$.

Thus, our goal is to exhibit a Riemannian geometry in which $\mathcal{HL}^+$ is geodesic complete. Geodesic completeness generalizes the notion of closure by convex linear combination. The notion of shortest route in a curved space derives from the notion of metric tensor that generalizes the inner product of the Euclidean space. At any point of a given manifold, a tangent space can be defined and the metric tensor defines an inner product for all tangent spaces, which leads to the notion of Riemannian metric. In order to define the Riemannian geometry over $\mathcal{HL}^+$, we use the property that, for every Δ in $\mathcal{HL}^+$, we have $\mathrm{Null}(\Delta) = \mathrm{Span}(\mathbf{1})$. We also define a smooth mapping between $\mathcal{HL}^+$ and the space of symmetric positive definite matrices $\mathcal{P}_{n-1} = \{R \in \mathbb{R}^{n-1\times n-1} \mid M = M^T,\ M \succ 0\}$ which can be embedded in a Riemannian geometry.

Formally, as for every $\Delta \in \mathcal{HL}^+$, we have $\mathrm{Null}(\Delta) = \mathrm{Span}(\mathbf{1})$, we deduce that the restriction of Δ to the vector space $\mathrm{Span}(\mathbf{1})^\perp$ is positive definite, where $\mathrm{Span}(\mathbf{1})^\perp$ denotes the vector space orthogonal to $\mathrm{Span}(\mathbf{1})$. It is important to note that the space $\mathrm{Span}(\mathbf{1})^\perp$ does not depend on Δ. In order to define the mapping between $\mathcal{P}_{n-1}$ and $\mathcal{HL}^+$, let us denote by $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ the canonical basis of $\mathbb{R}^{n\times n}$ (the i-th component of $\mathbf{e}_i$ is one, the other components are zeros) and let us consider the orthogonal basis $\mathcal{B}' = (\mathbf{1}, \mathbf{e}'_2, \ldots, \mathbf{e}'_n)$ where $(\mathbf{e}'_2, \ldots, \mathbf{e}'_n)$ is an orthogonal basis of $\mathrm{Span}(\mathbf{1})^\perp$. Let us now consider P, the change-of-coordinates operator $\mathcal{B} \to \mathcal{B}'$. We define the mapping f between $\mathcal{P}_{n-1}$ and $\mathcal{HL}^+$ by

$$f : A \to P^T \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & A \end{pmatrix} P \quad . \tag{5}$$

We show some important properties of f that will allow us to transfer the Riemannian structure from $\mathcal{P}_{n-1}$ to $\mathcal{HL}^+$.

Proposition 4 *f is a C^∞-diffeomorphism between $\mathcal{P}_{n-1}$ and $\mathcal{HL}^+$. For any $A \in \mathcal{P}_{n-1}$, we have $f(A^{-1}) = f(A)^\dagger$ and* $\mathrm{Tr}(f(A)) = \mathrm{Tr}(A)$.

Proof Let us consider $A \in \mathcal{P}_{n-1}$, $\mathrm{Null}(f(A)) = \mathrm{Span}(\mathbf{1})$ by construction. $f(A)$ is symmetric positive semidefinite since A is symmetric positive definite so we have $f(A) \in \mathcal{HL}^+$. Conversely, let us consider $M \in \mathcal{HL}^+$ and $U = PMP^T$. Since $\mathbf{1} \in \mathrm{Null}(M)$, we can write

$$U\mathbf{e}_1 = PMP^T\mathbf{e}_1 = PM\mathbf{1} = \mathbf{0} \ . \tag{6}$$

Thus, U can be written as

$$U = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & Q \end{pmatrix} \ , \tag{7}$$

with $Q \in \mathbb{R}^{n-1 \times n-1}$. Moreover $Q \in \mathcal{P}_{n-1}$ since the spectrum of Q is equal to the spectrum of M without the null eigenvalue associated with $\mathbf{1}$. Since $\mathrm{Null}(M) = \mathrm{Span}(\mathbf{1})$, the spectrum of Q is strictly positive. Finally, we have $M = f(Q)$ and f is a bijection between $\mathcal{P}_{n-1}$ and $\mathcal{HL}^+$. f and f^{-1} are infinitely differentiable as change-of-coordinates operators.

It remains to show the properties of f. First, let us consider $A \in \mathcal{P}_{n-1}$, then

$$\begin{aligned}
f(A)^\dagger &= \left(P^T \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & A \end{pmatrix} P \right)^\dagger \\
&= P^\dagger \left(P^T \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & A \end{pmatrix} \right)^\dagger \quad \text{since } P \text{ is orthogonal} \\
&= P^\dagger \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & A \end{pmatrix}^\dagger (P^T)^\dagger \quad \text{since } P^T \text{ is orthogonal} \\
&= P^T \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & A^{-1} \end{pmatrix} P \\
&= f(A^{-1}) \ .
\end{aligned}$$

Second, the property over traces can be proved by

$$\mathrm{Tr}(f(A)) = \mathrm{Tr}\left(P^T \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & A \end{pmatrix} P \right) = \mathrm{Tr}\left(PP^T \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & A \end{pmatrix} \right) = \mathrm{Tr}(A) \ .$$

∎

We will use Proposition 4 to transfer the geometrical structure of $\mathcal{P}_{n-1}$ to $\mathcal{HL}^+$. But, before let us review some results concerning the geometrical structure of $\mathcal{P}_{n-1}$. The set $\mathcal{P}_{n-1}$ is a space of matrices which can also be viewed as a space of multivariate Gaussian distributions with null mean (via the covariance matrix). These two views can be used to derive a Riemannian metric on $\mathcal{P}_{n-1}$. The first one is based on a pure geometrical approach and the second one is based on Fisher information theory. It is worth noting that both approaches produce the same metric tensor (see for instance Bonnabel and Sepulchre, 2009) defined by, for every $A \in \mathcal{P}_{n-1}$,

$$g_A^{\mathcal{P}_{n-1}}(D_1, D_2) = \mathrm{Tr}(D_1 A^{-1} D_2 A^{-1}) \ , \tag{8}$$

where D_1 and D_2 are in the tangent space at point A in $\mathcal{P}_{n-1}$ denoted by $\mathcal{T}_{\mathcal{P}_{n-1}}(A)$. A distance $d_{\mathcal{P}_{n-1}}$ over $\mathcal{P}_{n-1}$ can be derived from the definition of the metric tensor (8) by, for every A_1, A_2 in $\mathcal{P}_{n-1}$,

$$d_{\mathcal{P}_{n-1}}(A_1, A_2) = \| \log(A_1^{-1/2} A_2 A_1^{-1/2}) \|_2 \ . \tag{9}$$

and, also the metric tensor defined in (8) allows defining a geodesic curve from a matrix A_1 in $\mathcal{P}_{n-1}$ to a matrix A_2 in $\mathcal{P}_{n-1}$ by

$$\gamma_{A_1,A_2}^{\mathcal{P}_{n-1}}(\alpha) = A_1^{1/2} \exp(\alpha \log(A_1^{-1/2} A_2 A_1^{-1/2})) A_1^{1/2} \ .$$

It should be noted the geodesic from A_1 to A_2 is not, in general, equal the geodesic from A_2 to A_1. As noted in Bonnabel and Sepulchre (2009), $\mathcal{P}_{n-1}$ embedded with this natural geometry is geodesic complete, i.e., every geodesic can be extended to a maximal geodesic defined for $\alpha \in \mathbb{R}$. This property allows to use efficient short-step methods to solve complex optimization problems.

We are ready to define a Riemannian geometry over $\mathcal{HL}^+$ using Proposition 4 and the Riemannian geometry over $\mathcal{P}_{n-1}$. First, the mapping f between $\mathcal{P}_{n-1}$ and $\mathcal{HL}^+$ defined in Equation (5) can be extended to a C^∞-diffeomorphism between $\mathrm{Span}(\mathcal{P}_{n-1})$ and $\mathrm{Span}(\mathcal{HL}^+)$. This allows to define the tangent space at point M in $\mathcal{HL}^+$, denoted by $\mathcal{T}_{\mathcal{HL}^+}(M)$, by

$$\mathcal{T}_{\mathcal{HL}^+}(M) = f(\mathcal{T}_{\mathcal{P}_{n-1}}(f^{-1}(M))) \tag{10}$$

and to define the metric tensor for M in $\mathcal{HL}^+$, denoted by $g_M^{\mathcal{HL}^+}$. Let D_1 and D_2 be in the tangent space $\mathcal{T}_{\mathcal{HL}^+}(M)$. Then,

$$g_M^{\mathcal{HL}^+}(D_1, D_2) = g_{f^{-1}(M)}^{\mathcal{P}_{n-1}}(f^{-1}(D_1), f^{-1}(D_2)) \ . \tag{11}$$

Proposition 5 $g_M^{\mathcal{HL}^+}(D_1, D_2)$ *can be expressed in function of* M*,* D_1 *and* D_2 *using the formula*

$$g_M^{\mathcal{HL}^+}(D_1, D_2) = \mathrm{Tr}(D_1 M^\dagger D_2 M^\dagger) \ .$$

Proof First, note that the extension of the mapping f into a mapping between $\text{Span}(\mathcal{P}_{n-1})$ and $\text{Span}(\mathcal{HL}^+)$ can still be expressed using Equation (5). Thus, we get from Equation (11),

$$\begin{aligned}
g_M^{\mathcal{HL}^+}(D_1, D_2) &= g_{f^{-1}(M)}^{\mathcal{P}_{n-1}}(f^{-1}(D_1), f^{-1}(D_2)) \\
&= \text{Tr}\left[f^{-1}(D_1)(f^{-1}(M))^{-1}f^{-1}(D_2)(f^{-1}(M))^{-1}\right] \quad \text{(see Eq. (8))} \\
&= \text{Tr}\left[P\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & f^{-1}(D_1)(f^{-1}(M))^{-1}f^{-1}(D_2)(f^{-1}(M))^{-1}\end{pmatrix}P^T\right] \\
&= \text{Tr}\left[P\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & f^{-1}(D_1)\end{pmatrix}P^T P\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & (f^{-1}(M))^{-1}\end{pmatrix}P^T\right. \\
&\qquad \left. P\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & f^{-1}(D_2)\end{pmatrix}P^T P\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & (f^{-1}(M))^{-1}\end{pmatrix}P^T\right] \\
&= \text{Tr}\left[(f\circ f^{-1}(D_1))(f\circ f^{-1}(M))^{\dagger}(f\circ f^{-1}(D_2))(f\circ f^{-1}(M))^{\dagger}\right] \quad \text{(see Prop. 4)} \\
&= \text{Tr}(D_1 M^{\dagger} D_2 M^{\dagger}) \ .
\end{aligned}$$

■

We can now define the Riemannian structure of $\mathcal{HL}^+$. The Riemannian distance over $\mathcal{HL}^+$ is defined for every M_1, M_2 in $\mathcal{HL}^+$ by

$$d_{\mathcal{HL}^+}(M_1, M_2) = d_{\mathcal{P}_{n-1}}(f^{-1}(M_1), f^{-1}(M_2))$$

and the geodesic curve from M_1 to M_2 in $\mathcal{HL}^+$ is defined by

$$\gamma_{M_1,M_2}^{\mathcal{HL}^+}(\alpha) = f(\gamma_{f^{-1}(M_1),f^{-1}(M_2)}^{\mathcal{P}_{n-1}}(\alpha)) \ .$$

Proposition 6 *The Riemannian distance over $\mathcal{HL}^+$ between M_1, M_2 in $\mathcal{HL}^+$ can be expressed as*

$$d_{\mathcal{HL}^+}(M_1, M_2) = \| \log\left(\mathbf{1}\mathbf{1}^T + (M_1^{\dagger})^{1/2} M_2 (M_1^{\dagger})^{1/2}\right) \|_2$$

and the Riemannian distance is invariant by pseudoinverse.

Proof Let M_1, M_2 in $\mathcal{HL}^+$, then

$$
\begin{aligned}
d_{\mathcal{HL}^+}(M_1, M_2) &= d_{\mathcal{P}_{n-1}}(f^{-1}(M_1), f^{-1}(M_2)) \quad \text{by definition of } d_{\mathcal{HL}^+} \\
&= \left\| \begin{pmatrix} 0 & 0 \\ 0 & \log(f^{-1}(M_1)^{-1/2} f^{-1}(M_2) f^{-1}(M_1)^{-1/2}) \end{pmatrix} \right\|_2 \quad \text{by Equation (9)} \\
&= \left\| \log \left[\begin{pmatrix} 1 & 0 \\ 0 & f^{-1}(M_1)^{-1/2} f^{-1}(M_2) f^{-1}(M_1)^{-1/2} \end{pmatrix} \right] \right\|_2 \\
&= \left\| P^T \log \left[\begin{pmatrix} 1 & 0 \\ 0 & f^{-1}(M_1)^{-1/2} f^{-1}(M_2) f^{-1}(M_1)^{-1/2} \end{pmatrix} \right] P \right\|_2 \quad \text{since } P \text{ is orthogonal} \\
&= \left\| \log \left[P^T \begin{pmatrix} 1 & 0 \\ 0 & f^{-1}(M_1)^{-1/2} f^{-1}(M_2) f^{-1}(M_1)^{-1/2} \end{pmatrix} P \right] \right\|_2 \\
&= \left\| \log \left[\mathbf{1}\mathbf{1}^T + P^T \begin{pmatrix} 0 & 0 \\ 0 & f^{-1}(M_1)^{-1/2} \end{pmatrix} P P^T \begin{pmatrix} 0 & 0 \\ 0 & f^{-1}(M_2) \end{pmatrix} P P^T \begin{pmatrix} 0 & 0 \\ 0 & f^{-1}(M_1)^{-1/2} \end{pmatrix} P \right] \right\|_2 \\
&= \left\| \log \left(\mathbf{1}\mathbf{1}^T + (M_1^\dagger)^{1/2} M_2 (M_1^\dagger)^{1/2} \right) \right\|_2 \quad .
\end{aligned}
$$

The distance $d_{\mathcal{HL}^+}$ is invariant by pseudoinverse since the distance $d_{\mathcal{P}_{n-1}}$ is invariant by matrix inversion (see for example Bonnabel and Sepulchre, 2009), which concludes the proof. ■

It should be noted that, since f is a C^∞-diffeomorphism, the geodesic curves of $\mathcal{HL}^+$ can also be extended for $\alpha \in \mathbb{R}$. Thus, $\mathcal{HL}^+$ embedded with our new metric is geodesic complete. Moreover, because of the pseudoinverse invariance property, the Riemannian distance between two graph Laplacians in $\mathcal{HL}^+$ is equal to the Riemannian distance between the corresponding two graph kernels. The Riemannian structure of $\mathcal{P}_{n-1}$ with the Riemannian metric, also called the natural metric, has been used in many efficient applications in various fields (object detection in radar processing, bio medical imaging, kernel optimization). The Riemannian structure over $\mathcal{HL}^+$ introduced in this section should open new algorithmic perspectives such as Weiszfeld's algorithm for mean and median computation, usage of complete geodesics to express dissimilarities for graph kernels. We did not investigate in this paper this line of research but left it open for future work.

3. Undirected Hypergraphs and Hypergraphs Laplacians

In this section, we associate a complete semantic to the spaces $\mathcal{HL}$ and $\mathcal{HL}^+$. For instance, we show that every matrix M in $\mathcal{HL}$ is the Laplacian (also the kernel) of some undirected hypergraph.

3.1 Undirected Hypergraphs

We define formally an *undirected hypergraph* $\mathbf{h} = (N, H)$ as a set of nodes $N = (1, \ldots, n)$ and a set of hyperedges $H = \{h_1, \ldots, h_p\}$. A hyperedge $h = \{s_h, t_h\}$ in H is an unordered pair of two non empty and disjoint subsets of N. The subsets s_h and t_h are said to be the ends of the hyperedge $h = \{s_h, t_h\}$. If for every hyperedge, the two ends contain only one node, the hypergraph is an undirected graph without self-loops.

A *weighted undirected hypergraph* consists of a hypergraph $\mathbf{h} = (N, H)$ and, for every hyperedge $h = \{s_h, t_h\}$, a function w_h mapping every node i in $s_h \cup t_h$ to a positive weight $w_h(i)$ (for $i \notin s_h \cup t_h$, we define $w_h(i) = 0$). In order to define consistently the unnormalized gradient and the unnormalized Laplacian, the weight functions satisfy the *Equilibrium Condition.* For every hyperedge $h = \{s_h, t_h\}$ in H with weight function w_h, we have

$$\sum_{i \in t_h} \sqrt{w_h(i)} = \sum_{i \in s_h} \sqrt{w_h(i)} \ . \qquad \text{(Equilibrium Condition)}$$

We will say that a node i *belongs to* a hyperedge h ($i \in h$) if $w_h(i) \neq 0$. We define the degree of a node i by

$$d(i) = \sum_h w_h(i) \ . \qquad (12)$$

The degree of a node is positive when it participates in at least one hyperedge. In the following, we will always assume that the degrees are positive. We define the diagonal degree matrix by $D = \mathrm{diag}(d(1), \ldots, d(n))$ and the volume of the hypergraph by $\mathrm{Vol}(\mathbf{h}) = \sum_{i \in N} d(i)$.

Example 2 *Examples of hypergraphs are given in Figure 2. The hypergraph $\mathbf{h}_1$ has four nodes which can be viewed as tennis players. The first hyperedge connects the teams $\{1, 3\}$ and $\{2, 4\}$. The second hyperedge connects the teams $\{1, 4\}$ and $\{2, 3\}$. All weights have been chosen equal to 1 which can be interpreted as an uniform contribution of the players in the teams. The hypergraph $\mathbf{h}_2$ has only one hyperedge. As before, the nodes 1 to 4 may be viewed as tennis players. The nodes s_1 and s_2 are introduced for modeling the results of the games. The weights w_1 to w_4 can be, for example, defined according to the individual affinity of the tennis players with the type of court used in the game (grass, clay, concrete, ...).*

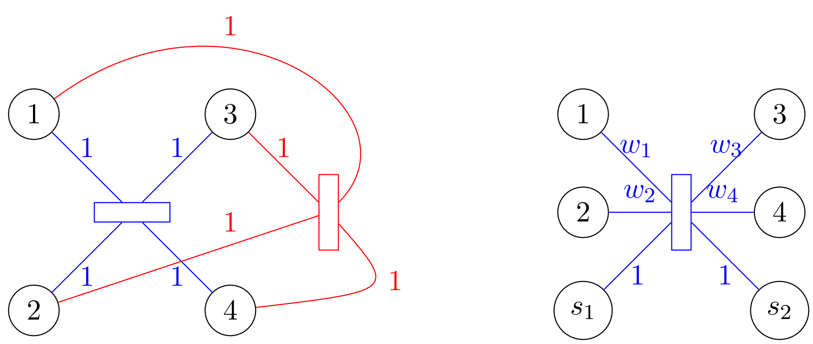


Figure 2: Hypergraph $\mathbf{h}_1$ (left) and hypergraph $\mathbf{h}_2$

When a hyperedge h is an unordered pair $\{\{i\}, \{j\}\}$ of two nodes i, j, the Equilibrium Condition states that the weights $w_h(i)$ and $w_h(j)$ are equals. Therefore, every undirected hypergraph such that all hyperedges are unordered pairs of singleton nodes can be viewed as an undirected graph, and we will say that the hypergraph is a graph. In this case, we define the adjacency matrix of the (equivalent) graph to be the matrix W defined by $W_{i,j} = W_{j,i} = w_h(i) = w_h(j)$ for every hyperedge $\{\{i\}, \{j\}\}$, and 0 otherwise. In Example 4, the hypergraph $\mathbf{h}_3$ is a graph with edges $\{1, 2\}$ and $\{1, 3\}$, and the two edge weights set to 1. Conversely, every undirected graph can be viewed as an undirected hypergraph.

3.2 Hypergraph Laplacians

The graph Laplacian allows to define a smoothness semi-norm (see Section 2.1) which models a similarity between connected nodes. Indeed, when assigning labels or scores to the nodes of a graph using a real-valued function f, the smoothness operator allows to ensure that $f(i)$ is close to $f(j)$ when i and j are connected. Note that higher is the edge weight connecting i and j, closer should be the values $f(i)$ and $f(j)$. We extend the notion of similarity between nodes to a notion of similarity between node sets for hyperedges by defining a smoothness operator for hypergraphs. For this, we will consider the weighted sums over all nodes in every end of the hyperedges and we will introduce formally the notions of hypergraph gradient, hypergraph Laplacian, hypergraph kernel and the hypergraph smoothness operator. But before, let us consider an example.

Example 3 *(Example 2 continued)*

Let us consider the hypergraph $\mathbf{h}_1$ and let us consider a real-valued node function f. The smoothness of f on $\mathbf{h}_1$ will express that $f(1) + f(3)$ should be close to $f(2) + f(4)$ and $f(1) + f(4)$ should be close to $f(2) + f(3)$.

Let us now consider the hypergraph $\mathbf{h}_2$ and let us consider a real-valued node function f. The smoothness of f on $\mathbf{h}_2$ will express that the weighted sum of $f(1)$, $f(2)$ and $f(s_1)$ should be close to the weighted sum of $f(3)$, $f(4)$ and $f(s_2)$. Let us suppose that we know that the team $\{1, 2\}$ has won the match against the team $\{3, 4\}$. We can assign values to $f(s_1)$ and $f(s_2)$ satisfying $f(s_1) < f(s_2)$ based on the score of the match. Then, the smoothness condition will allow to express that the weighted sum of $f(1)$ and $f(2)$ will be greater than the weighted sum of $f(3)$ and $f(4)$. Let us now suppose that we are given a hypergraph with many hyperedges, each of them modeling a tennis match. With the knowledge of some results, we can assign values to the nodes of type s. Then, a semi-supervised learning algorithm using the smoothness condition on the hypergraph should output a scoring function for the tennis players according to the known results.

Formally, let $\mathbf{h} = (N, H)$ be a hypergraph and f be a real-valued node function, we extend f by defining f over the ends of a hyperedge $h = \{s_h, t_h\}$ in H by

$$f(s_h) = \sum_{i \in s_h} f(i)\sqrt{w_h(i)};\ f(t_h) = \sum_{i \in t_h} f(i)\sqrt{w_h(i)}\ .$$

It should be noted that, when f is a constant function over N, then $f(t_h)$ is equal to $f(s_h)$ because of the Equilibrium Condition. Then, the *(hypergraph) unnormalized gradient* of a hypergraph $\mathbf{h} = (N, H)$ is a linear application, denoted by grad, that maps every real-valued node function f into a real-valued hyperedge function grad(f) defined for every

$h = \{s_h, t_h\}$ in H by

$$\text{grad}(f)(h) = f(t_h) - f(s_h) = \sum_{i \in t_h} f(i)\sqrt{w_h(i)} - \sum_{i \in s_h} f(i)\sqrt{w_h(i)} \ , \tag{13}$$

where an arbitrary orientation of the hyperedges has been chosen. As expected, $|\,\text{grad}(f)(h)|^2$ is small when the total value $f(s_h)$ is close to the total value $f(t_h)$. We denote by $G \in \mathbb{R}^{p \times n}$ the matrix of grad$(\cdot)$. We note that, for every hyperedge and its arbitrary orientation $h = (s_h, t_h)$ and every node i, $G_{h,i} = \epsilon_h(i)\sqrt{w_h(i)}$ where ϵ_h is the *orientation function* defined by

$$\epsilon_h(i) = \begin{cases} 1 \text{ if } i \in t_h \ , \\ -1 \text{ if } i \in s_h \ , \\ 0 \text{ otherwise} \ . \end{cases}$$

We present examples of gradient matrices in Example 4 and we discuss the relations between graph gradients and hypergraph gradients. Recall that a hypergraph in which every hyperedge is a pair of singleton nodes is a graph and conversely. It is easy to show that, for every graph $\mathbf{g}$, the graph gradient described in Section 2.1 coincides with the hypergraph gradient for the equivalent hypergraph, and conversely.

Because of the Equilibrium Condition, the gradient of every constant node function is the zero-valued hyperedge function. This can be written as $\mathbf{1} \in \text{Null}(G)$, where $\text{Null}(G)$ is the set of so-called harmonic functions. As in the graph case, the *global smoothness* of a real-valued node function f over a hypergraph $\mathbf{h}$ is defined by $\|Gf\|^2 = (Gf)^T(Gf) = f^T G^T G f$.

Let $\mathbf{h}$ be an undirected hypergraph with unnormalized gradient G, the square $n \times n$ real valued matrix $\Delta = G^T G$ is defined to be the *unnormalized Laplacian* of the hypergraph $\mathbf{h}$. When the hypergraph is a graph, the unnormalized hypergraph Laplacian coincides with the unnormalized Laplacian described in Section 2.1. It should be noted that the Laplacian Δ does not depend on the arbirary orientation of the hyperedges. The Laplacian hypergraph shares several important properties with the graph Laplacian.

Proposition 7 *Let* $\mathbf{h}$ *be an undirected hypergraph with unnormalized gradient* G *and with Laplacian* Δ*,* Δ *is symmetric positive semidefinite and* $\text{Null}(\Delta) = \text{Null}(G)$*. As direct consequences,* $\mathbf{1} \in \text{Null}(\Delta)$ *and* $f \to f^T \Delta f$ *is a (*smoothness*) semi-norm.*

Proof We have $\Delta^T = (G^T G)^T = G^T G = \Delta$ so Δ is symmetric. For any $x \in \mathbb{R}^n$, we have

$$x^T \Delta x = x^T G^T G x = \|Gx\|^2 \geq 0 \ .$$

So Δ is positive semidefinite. Moreover, if $\Delta x = 0$, then $\|Gx\|^2 = 0$ and $Gx = 0$. Conversely, if $Gx = 0$, then $\Delta x = 0$ so $\text{Null}(\Delta) = \text{Null}(G)$. ∎

Example 4 *We consider three hypergraphs over* $N = \{1, 2, 3\}$ *depicted in Figure 3. The hypergraph* $\mathbf{h}_4$ *shown in the middle of Figure 3 has two hyperedges* $h = \{\{1\}, \{3\}\}$ *and* $h' = \{\{1, 3\}, \{2\}\}$*. The hypergraph* $\mathbf{h}_3$ *shown on the left has two hyperedges* h *and* $h'' = \{\{1\}, \{2\}\}$*. The hypergraph* $\mathbf{h}_5$ *shown on the right has only one hyperedge* h'*. The hyperedge weights in* *the three graphs are defined to be $w_h(1) = w_h(3) = 1$; $w_{h'}(1) = w_{h'}(3) = 0.5$, $w_{h'}(2) = 2$ and $w_{h''}(1) = w_{h''}(2) = 1$. For each of the three hypergraphs, we compute a gradient matrix (based on an arbitrary orientation of the hyperedges) and the Laplacian.*

$$G_3 = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}; \ \Delta_3 = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

$$G_4 = \begin{pmatrix} \frac{-\sqrt{2}}{2} & \sqrt{2} & \frac{-\sqrt{2}}{2} \\ \frac{-\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{pmatrix}; \ \Delta_4 = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

$$G_5 = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}; \ \Delta_5 = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 2 & 0 \\ -1 & -1 & 0 & 2 \end{pmatrix}$$

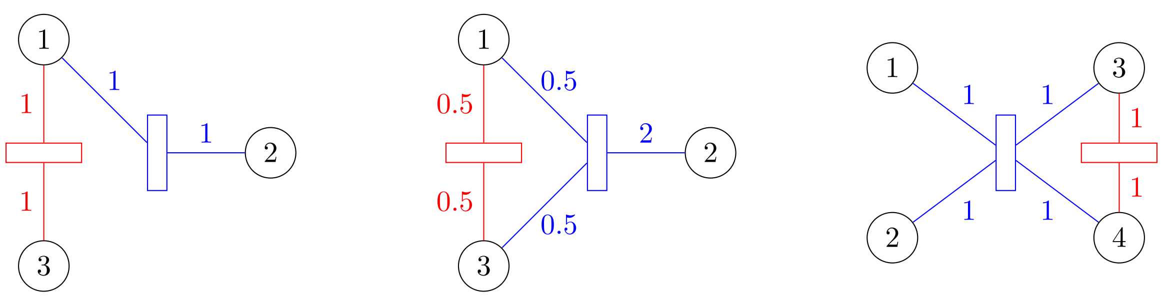


Figure 3: Hypergraphs $\mathbf{h}_3$ (left), $\mathbf{h}_4$ (middle) and $\mathbf{h}_5$ (right).

As in the graph case, the Moore-Penrose pseudoinverse of the Laplacian matrix $\Delta^\dagger$ of a hypergraph $\mathbf{h}$ is symmetric and positive semidefinite. We define the *hypergraph kernel* of a hypergraph $\mathbf{h}$ to be the pseudoinverse $\Delta^\dagger$ of its Laplacian.

We now prove that the class of undirected hypergraphs allows us to define a graph-like semantic for the class of kernels $\mathcal{HL}$ introduced in the previous section.

Theorem 8 *The set of hypergraph Laplacians is equal to the class $\mathcal{HL}$.*

Proof First, let us consider a square $n \times n$ matrix M in $\mathcal{HL}$. As M is symmetric positive semidefinite, there exists a square root decomposition $M = G^T G$ where $G \in \mathbb{R}^{p \times n}$, and G is defined to be a square root of M. We have $\mathbf{1} \in \text{Null}(G)$ since $\text{Null}(G) = \text{Null}(\Delta)$. Using the decomposition, we can define a hypergraph using the following procedure

Input: $M = G^T G \in \mathcal{HL}$
Let $N = \{1, \dots, n\}$ and $H = \emptyset$
for each line $i = 1 \dots p$ of G **do**
  Let $s = \{j \in N \mid G_{i,j} > 0\}$, let $t = \{j \in N \mid G_{i,j} < 0\}$, and let $h = \{s, t\}$
  Let w_h be defined by $w_h(j) = G_{i,j}^2$ for j in $s \cup t$
  Let $H = H \cup \{h\}$

6: **end for**
7: **return** $\mathbf{h} = (N, H)$ and the weight functions.

Note that every hyperedge $h = \{s, t\}$ computed along the algorithm satisfies the Equilibrium Condition since $\mathbf{1} \in \mathrm{Null}(G)$. One can easily verify that G is a gradient matrix of $\mathbf{h}$. Therefore, the matrix M is the hypergraph Laplacian of $\mathbf{h}$.

Conversely, a hypergraph Laplacian Δ is a symmetric positive semidefinite matrix and $\mathbf{1} \in \mathrm{Null}(\Delta)$ as stated in Proposition 7. Thus, a hypergraph Laplacian Δ is in $\mathcal{HL}$ by definition of $\mathcal{HL}$. ■

In the proof of Proposition 8, the hypergraph $\mathbf{h}$ has been defined from a decomposition $\Delta = G^T G$. As the square root of a symmetric positive semidefinite matrix is not unique, several hypergraphs with the same Laplacian Δ can be defined. Based on this idea, we define the notion of hypergraph equivalence.

Definition 9 *Two hypergraphs are said to be* equivalent *if they have the same Laplacian matrix.*

To illustrate this definition, let us consider the Laplacian Δ_4 of the hypergraph $\mathbf{h}_4$ (see Example 4), we can also consider the decomposition $\Delta_4 = G_6^T G_6$ where G_6 is defined by

$$G_6 = \begin{pmatrix} -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & \sqrt{2} & 0 \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 & \sqrt{2} \end{pmatrix} .$$

From G_6, we can define the hypergraph $\mathbf{h}_6$ presented in Figure 4 below.

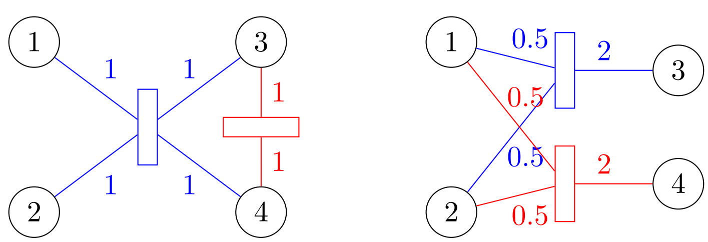


Figure 4: Equivalent hypergraphs $\mathbf{h}_5$ and $\mathbf{h}_6$

Equivalent hypergraphs have the same Laplacian and therefore have the same smoothness operator. We can note that the smoothness constraints expressed by the hyperedges of $\mathbf{h}_6$ are consistent with the smoothness constraints expressed by the hyperedges of $\mathbf{h}_5$. Indeed, any smooth function f on $\mathbf{h}_5$ must have $f(1) + f(2)$ close to $f(3) + f(4)$ and $f(3)$ close to $f(4)$. In $\mathbf{h}_6$, a smooth function should have $f(1) + f(2)$ close to $2f(3)$ and to $2f(4)$. Thus, the constraints from $\mathbf{h}_6$ are a linear combination of the ones from $\mathbf{h}_5$. From an algebraical perspective, we can show the gradient matrix of $\mathbf{h}_6$, G_6, can be written under the form QG_5 where Q is an isometry ($Q^T Q = I$) that express the linear relations between the constraints of both hypergraphs.

Note that the equivalence relation denotes a very strong link between two hypergraphs that is actually more specific than a simple linear relation between the set of constraints expressed by the hyperedges. Indeed, we can easily build two non-equivalent hypergraphs

$\mathbf{h}$ and $\mathbf{h}'$ such that the smoothness constraints of the one are linear combination of the smoothness constraints of the other. Indeed, let us consider $G' = RG$ where G and G' are respectively the gradient matrix of two hypergraphs $\mathbf{h}$ and $\mathbf{h}'$, and R is an invertible but non orthogonal matrix. In the general case $G'^T G' = G^T R^T R G \neq G^T G$ but the constraints of $\mathbf{h}$ are a linear combination of the constraints of $\mathbf{h}'$ (and vice-versa).

We conclude the section by giving properties of hypergraph Laplacians and hypergraph kernels, that is

Corollary 10

- *The class of hypergraph Laplacians and the class of hypergraph kernels are equal.*
- *The class of hypergraph kernels is closed by convex linear combination.*
- *The convex linear combination of graph kernels is a hypergraph kernel.*

These properties are direct consequences from Theorem 8 and of Proposition 3.

3.3 Pairwise Weight Matrix

Let us recall that, in the graph case, the Laplacian Δ of a graph $\mathbf{g} = (N, E)$ can be computed from the (edge) weight matrix and the degree matrix with the equation $\Delta = D - W$. The objective of this section is to prove an analog result for hypergraphs. For this, we introduce *pairwise weights* for node pairs in weighted undirected hypergraphs.

First, we define, for every node pair (i, j) and every hyperedge h, the hyperedge pairwise weight to be

$$w_h(i,j) = \delta_{i \neq j} P_h(i,j) \sqrt{w_h(i)} \sqrt{w_h(j)} \ ,$$

where

$$P_h(i,j) = \begin{cases} \ \ 1 \text{ if } i \text{ and } j \text{ belong to different ends of } h \ , \\ -1 \text{ if } i \text{ and } j \text{ belong to the same end of } h \ , \\ \ \ 0 \text{ if } i \text{ or } j \text{ does not belong to } h \ , \end{cases}$$

and where δ stands for the Kronecker delta ($\delta_{i \neq j} = 1$ if $i \neq j$ and 0 otherwise). The quantity $P_h(i, j)$ can be interpreted as a type of electrical polarity which depends on whether the two nodes belong to the same end of the hyperedge. It should be noted that $P_h(i, j)$ is independent of the arbitrary orientation of the hyperedges. Also, it could be noted that, for every orientation function ϵ_h and every node pair (i, j), we have

$$P_h(i,j) = -\epsilon_h(i)\epsilon_h(j) \ . \tag{14}$$

In the definition of the hyperedge pairwise weight, the quantity $\sqrt{w_h(i)}$ can be viewed as the cost of entering the hyperedge h at node i and $\sqrt{w_h(j)}$ as the cost of exiting the hyperedge h at node j. When a hyperedge h is a pair of two singletons $\{i\}$ and $\{j\}$, $w_h(i)$ and $w_h(j)$ are equal because of the equilibrium condition, and the hyperedge pairwise edge satisfies $w_h(i,j) = \sqrt{w_h(i)}\sqrt{w_h(j)} = w_h(i) = w_h(j)$.

Second, the *pairwise (hypergraph) weight matrix* W of a hypergraph $\mathbf{h} = (N, H)$ is defined by

$$\forall i,j \in N, W_{i,j} = \sum_{h\in H} w_h(i,j) = \delta_{i\neq j} \sum_{h\in H} P_h(i,j)\sqrt{w_h(i)}\sqrt{w_h(j)} \ . \tag{15}$$

It can be noted that the diagonal terms of W are zero because all $w_h(i,i)$ are equal to zero. It is easy to see that the pairwise weight matrix of a graph $\mathbf{g}$ considered as a hypergraph is equal to the adjacency matrix of the graph $\mathbf{g}$. It can also be noted that the diagonal degree matrix of a hypergraph defined in Equation (12) can be computed from the pairwise weight matrix because

$$\begin{aligned} \forall i \in N,\ \sum_j W_{i,j} &= \sum_h \sqrt{w_h(i)} \sum_{j\neq i} P_h(i,j)\sqrt{w_h(j)} \\ &= \sum_h w_h(i) \quad \text{(Equilibrium Condition)} \\ &= d(i) \ . \end{aligned} \tag{16}$$

We can now prove that the hypergraph Laplacian of a hypergraph can be computed from its pairwise weight matrix and its degree matrix.

Proposition 11 *Let $\mathbf{h} = (N,H)$ be a hypergraph, let W be the pairwise weight matrix of $\mathbf{h}$, and let D be the diagonal degree matrix of $\mathbf{h}$. Then, the unnormalized Laplacian of $\mathbf{h}$ is $\Delta = D - W$.*

Proof Let $\mathbf{h} = (N,H)$ be an undirected weighted hypergraph with $H = (h_1,\ldots,h_p)$. Let G be a gradient matrix of $\mathbf{h}$ for some arbitrary orientation ϵ of the hyperedges. By definition of G, we have $G_{i,j} = \epsilon_{h_i}(j)\sqrt{w_{h_i}(j)}$. Now, the unnormalized (hypergraph) Laplacian Δ is defined by $\Delta = G^T G$. This leads to

$$\forall 1 \le i,j \le n,\ \Delta_{i,j} = \sum_{k=1}^{p} G_{k,i}G_{k,j} = -\sum_{k=1}^{p} P_{h_k}(i,j)\sqrt{w_{h_k}(i)}\sqrt{w_{h_k}(j)}$$

and it is easy to verify that $\Delta = D - W$. ∎

Example 5 *(Example 4 continued) For each hypergraph presented in Example 4, we compute the pairwise weight matrix W and the corresponding degree matrix D. One can verify that $D - W$ is equal to the Laplacian matrices presented in Example 4.*

$$W_3 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \ ; \ D_3 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \ ;$$

$$W_4 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \ ; \ D_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} ;$$

$$W_5 = \begin{pmatrix} 0 & -1 & 1 & 1 \\ -1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \;;\; D_5 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} .$$

As a consequence of Proposition 11, we can leverage the pairwise weight matrix to characterize equivalent hypergraphs and give a property of degree matrices as

Corollary 12 *Two hypergraphs are equivalent if and only if they have the same pairwise weight matrix. Two equivalent hypergraphs have the same degree matrix.*

Proof Proposition 11 states that a hypergraph Laplacian Δ can be written in a unique way as $D - W$ with D a diagonal matrix and W a matrix with diagonal terms equal to 0. Thus, equivalent hypergraphs will necessarily share the same pairwise matrix W and the same degree matrix D, which concludes the proof. ■

4. Properties of Undirected Hypergraphs

In this section, we discuss some important properties of the undirected hypergraphs. We first show relations between hypergraphs, graphs and a specific class of signed graphs. Then we propose an extension of the notion of commute-time distance that we call potential distance. Finally, we study whether the notions of path and of connected component can be extended to the case of undirected hypergraphs.

4.1 Hypergraphs, Graphs and Signed Graphs

In Section 3.3, we have observed that the pairwise weight matrix of a hypergraph can have negative weights (see for example the pairwise weight matrix W_5 of the graph $\mathbf{h}_5$ in Example 5). Thus, the pairwise weight matrix can not, in general, be interpreted as an adjacency matrix of a graph. However, it can be interpreted as the adjacency matrix of a *signed graph*, i.e., an undirected graph with possibly negative weights. Following this idea, we define the notion of *reduced signed graph.*

Definition 13 *The* reduced signed graph *of a hypergraph* $\mathbf{h}$ *is the signed graph* $\tilde{h}$ *with adjacency matrix* W*, where* W *is the pairwise weight matrix of the hypergraph* $\mathbf{h}$.

Examples of reduced signed graphs are shown in Example 6. Since equivalent hypergraphs share the same pairwise weight matrix (see Corollary 12), they also share the same reduced signed graph. It is easy to see that the reduced signed graph $\tilde{g}$ of a graph g is equal to g. Consequently, we can easily characterize the hypergraphs that are equivalent to a graph.

Proposition 14 *A hypergraph* $\mathbf{h}$ *is equivalent to a graph* g *if and only if the reduced signed graph* $\tilde{h}$ *of* $\mathbf{h}$ *is a graph. And then,* $g = \tilde{h}$ *is the unique graph equivalent to* $\mathbf{h}$.

Proof Let us assume that g is a graph equivalent to $\mathbf{h}$. Necessarily, the pairwise weight matrix W of $\mathbf{h}$ is also the pairwise weight matrix of g which is equal to the adjacency matrix of g. Then $\tilde{h}$ is equal to the graph g. Conversely, if $\tilde{h}$ is a graph, then its pairwise weight matrix is equal to the adjacency matrix of a graph $\mathbf{g}$. Since $\mathbf{g}$ and $\mathbf{h}$ share the same pairwise weight matrix, Corollary 12 allows us to state that $\mathbf{h}$ and $\mathbf{g}$ are equivalent. ■

Example 6 *(Example 5 continued) The hypergraph* $\mathbf{h}_3$ *is a graph and is equal to its reduced signed graph. The reduced signed graphs for the hypergraphs* $\mathbf{h}_4$ *and* $\mathbf{h}_5$ *are shown in Figure 5 and Figure 6.*

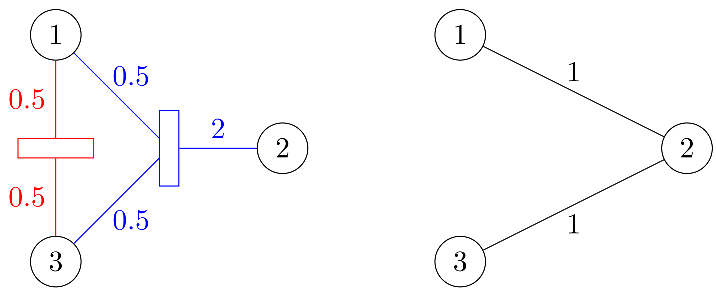


Figure 5: Hypergraph $\mathbf{h}_4$ from Example 4 and its reduced signed graph $\tilde{h}_4$

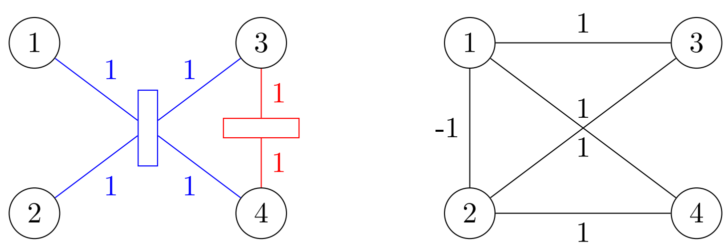


Figure 6: Hypergraph $\mathbf{h}_5$ from Example 4 and its reduced signed graph $\tilde{h}_5$

For instance, the hypergraph $\mathbf{h}_4$ of Example 6 is equivalent to the graph $\tilde{h}_4$ which is the unique graph in the equivalence class of $\mathbf{h}_4$. But, the hypergraph $\mathbf{h}_5$ of Example 6 has no equivalent graph since $\tilde{h}_5$ is a signed graph which is not a graph.

We can note that the interpretation of the smoothness on a hypergraph is still consistent with the smoothness on its equivalent graph, when it exists. For instance, let us consider the hypergraph $\mathbf{h}_1$ and the graph $\mathbf{g}_1$ presented in Figure 7. One can verify that they are equivalent, i.e., they have the same Laplacian. The smoothness on $\mathbf{h}_1$ states that, for any real-valued node function f, $f(1) + f(3)$ should be close to $f(2) + f(4)$ and $f(1) + f(4)$ should be close to $f(2) + f(3)$. We can observe that these two conditions reduce to "$f(1)$ should be close to $f(2)$" and "$f(3)$ should be close to $f(4)$", which is the interpretation of the smoothness over the graph $\mathbf{g}_1$.

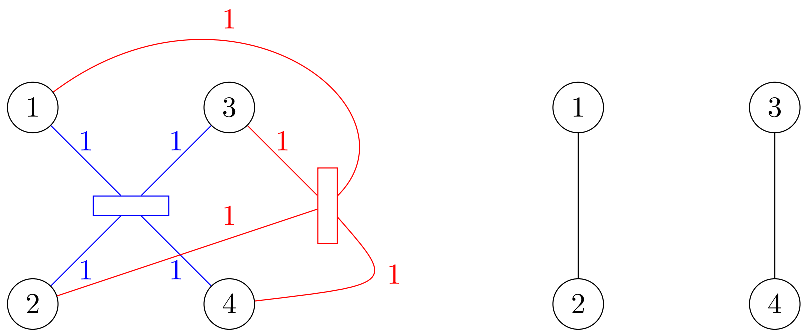


Figure 7: Hypergraph $\mathbf{h}_1$ and equivalent graph $\mathbf{g}_1$

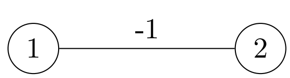
So far we have shown that every hypergraph can be reduced to a signed graph. That is, every hypergraph can be associated to a signed graph through its pairwise weight matrix. Note that the converse is not true: for a given signed graph, there may be no hypergraph that can be reduced to it. Indeed, let us consider a signed graph $\tilde{g}$ with adjacency matrix $W \in \mathbb{R}^{n \times n}$, we define the diagonal degree matrix D to be the unique diagonal matrix such that $W\mathbf{1} = D\mathbf{1}$ (same as in the classic graph case). We can easily show that

Proposition 15 *$\tilde{g}$ is equivalent to a hypergraph if and only if the matrix $D - W$ is positive semidefinite.*

Proof Since $W\mathbf{1} = D\mathbf{1}$, we always have $\mathbf{1} \in \mathrm{Null}(D - W)$. Moreover, since W is symmetric, $D - W$ is also symmetric. Let us recall that $\mathcal{HL} = \{M \in \mathbb{R}^{n \times n} \mid M = M^T,\ \mathbf{1} \in \mathrm{Null}(M),\ M \succeq 0\}$. If $D - W$ is positive semidefinite, then we have directly $D - W \in \mathcal{HL}$. Conversely, if W is the pairwise weight matrix of a hypergraph, then $D - W$ is a hypergraph Laplacian and is thus positive semidefinite. ■

Example 7 below presents a simple counterexample with the signed graph $\tilde{\mathbf{g}}_7$.

Example 7 *Let us consider the signed graph $\tilde{\mathbf{g}}_7$ with $N = \{1, 2\}$ and $W_{1,2} = -1$ (see Figure 8). The matrix $D - W$ is not positive semidefinite so the signed graph $\tilde{\mathbf{g}}_7$ is not the reduced graph of a hypergraph.*



$$D - W = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

Figure 8: Signed graph $\tilde{\mathbf{g}}_7$

Several approaches have been proposed to define Laplacian matrices for signed graphs (see Hou, 2005; Kunegis et al., 2010). An important reason to reject $D - W$ as a valid Laplacian is that it produces in general an indefinite matrix and thus cannot be used to define a meaningful smoothness semi-norm. As stated in Proposition 15, the class of pairwise weight matrices corresponds to the case where the matrix $D - W$ is positive semidefinite, and such a matrix has a semantic defined by an undirected hypergraph.

4.2 Potential distance in Hypergraphs

One reason to use graph kernels for learning in graphs is their relation with random walks because the commute-time distance can be computed from the graph kernel (see Section 2.1). We study whether a similar result exists for hypergraphs. For this, let us consider, throughout the section, a hypergraph $\mathbf{h}$ and its Laplacian Δ, and let $\mathbf{e}_i$ be the i^{th} vector of the identity matrix. By analogy with Equation (2), we define the potential distance $\Omega(i,j)$ between two nodes i and j by

$$\Omega(i,j) = \mathrm{Vol}(\mathbf{h})(\mathbf{e}_i - \mathbf{e}_j)^T \Delta^\dagger (\mathbf{e}_i - \mathbf{e}_j) = \mathrm{Vol}(\mathbf{h}) \left(\Delta^\dagger_{i,i} + \Delta^\dagger_{j,j} - 2\Delta^\dagger_{i,j} \right) \quad , \tag{17}$$

where $\Delta^\dagger$ is the Moore-Penrose pseudoinverse of the Laplacian Δ of $\mathbf{h}$.

Proposition 16 *The potential distance Ω is a pseudo-distance (or pseudo-metric) on the node set N, i.e., it is positive, symmetric and satisfies the triangle inequality. The pseudo distance Ω is a distance (or metric) when the Laplacian Δ of $\mathbf{h}$ is in $\mathcal{HL}^+$, i.e., when Δ satisfies* $\mathrm{Null}(\Delta) = \mathrm{Span}(\mathbf{1})$.

Proof Ω is positive since $\Delta^\dagger$ is positive semidefinite. It is easy to check in Equation (17) that, for every i,j, $\Omega(i,j) = \Omega(j,i)$. We show that it satisfies the triangle inequality. Indeed, For every $i, j, k \in N$, we have

$$\begin{aligned}
\frac{\Omega(i,j)}{\mathrm{Vol}(\mathbf{h})} &= \Delta^\dagger_{i,i} + \Delta^\dagger_{j,j} - 2\Delta^\dagger_{i,j} \quad \text{by definition of } \Omega(i,j) \\
&= \Delta^\dagger_{i,i} + \Delta^\dagger_{k,k} - 2\Delta^\dagger_{i,k} + \Delta^\dagger_{k,k} + \Delta^\dagger_{j,j} - 2\Delta^\dagger_{k,j} + 2(\Delta^\dagger_{i,k} + \Delta^\dagger_{k,j} - \Delta^\dagger_{i,j} - \Delta^\dagger_{k,k}) \\
&= \Omega(i,k) + \Omega(k,j) + 2(\mathbf{e}_i - \mathbf{e}_k)^T \Delta^\dagger (\mathbf{e}_k - \mathbf{e}_j) \\
&\le \Omega(i,k) + \Omega(k,j) \quad \text{since } \Delta^\dagger \text{ is positive semidefinite.}
\end{aligned}$$

So Ω is a pseudo distance on N. Let us now assume that $\Delta \in \mathcal{HL}^+$. We show that in this case, Ω is a distance. For this purpose, we consider $i, j \in N$ such that $\Omega(i,j) = 0$ and we show that $i = j$. We first write $\Delta^\dagger = V\Lambda^\dagger V^T$ where Λ is a diagonal matrix containing the eigenvalues $\lambda_1, \ldots, \lambda_n$ of Δ and V is an orthogonal matrix. Since $\mathrm{Null}(\Delta) = \mathrm{Span}(\mathbf{1})$, we can assume without loss of generality that $\lambda_1 = 0$ and $\lambda_k > 0$ for $k \ge 2$. By definition of $\Omega(i,j)$, we get $\Omega(i,j) = \sum_k \lambda_k^\dagger \left(V_{i,k}^2 + V_{j,k}^2 - 2V_{i,k}V_{j,k} \right)$, and thus $\Omega(i,j) = \sum_k \lambda_k^\dagger (V_{i,k} - V_{j,k})^2$. Consequently, $V_{i,k} = V_{j,k}$ for every $k \ge 2$.

Moreover, we also have $V_{i,1} = V_{j,1}$ since the column vector $V_{.,1} \in \mathrm{Span}(\mathbf{1})$ (eigenvector associated to $\lambda_1 = 0$). Consequently, $V_{i,k} = V_{j,k}$ for all k: the line vectors $V_{i,.}$ and $V_{j,.}$ are equals, which is only possible when $i = j$ since V is an orthogonal matrix ($V_{i,.}^T V_{j,.} = \delta_{i=j}$). ■

When the Laplacian Δ of $\mathbf{h}$ is not in the class $\mathcal{HL}^+$, which is the case when $\mathrm{Rank}(\Delta) < N - 1$, we can have $\Omega(i,j) = 0$ with $i \ne j$. For instance, let us consider the hypergraph $\mathbf{h}_3$ from Example 4, we have $\mathrm{Rank}(\Delta_3) = 1 < 2$ and the potential distance between nodes 1 and 3 in the hypergraph $\mathbf{h}_3$ is zero.

We will now express the potential distance in term of a diffusion function in hypergraphs mimicking the graph case. For this, we first define the diffusion function V from a potential function f solution of a Poisson equation modeling the diffusion according to a hypergraph Laplacian. Let $\mathbf{h} = (N, H)$ be a hypergraph with $\Delta \in \mathcal{HL}^+$. We first consider the Poisson equation $\Delta f = \mathbf{In}$ that models the diffusion of an input charge $\mathbf{In}$ through a system associated with the Laplacian operator Δ. Let us consider a node $j \in N$ called *sink* node, we consider the input function $\mathbf{In}_j$ defined by

$$\mathbf{In}_j(i) = \begin{cases} d(i) \text{ if } i \neq j \ , \\ d(j) - \mathrm{Vol}(\mathbf{h}) \text{ if } i = j \ , \end{cases} \tag{18}$$

where $d(i)$ denotes the degree of the node i. We define the set $\mathcal{S}_j$ as the set of functions $f \in \mathbb{R}^n$ which are solutions of the equation $\Delta f = \mathbf{In}_j$. For every node i in N, we define $\mathrm{V}(i, j) = f(i) - f(j)$, i.e., $\mathrm{V}(i, j)$ is the difference of potential between node i and the fixed sink node j. However, to make the definition of V consistent, we first have to prove that $\mathrm{V}(i, j)$ does not depend on the choice of a solution f in the set $\mathcal{S}_j$.

Lemma 17 *For every sink node j, the solutions of $\Delta f = \mathbf{In}_j$ are the functions $f = \mu\mathbf{1} + \Delta^\dagger \mathbf{In}_j$ where $\mu \in \mathbb{R}$.*

Proof Since $\Delta \in \mathcal{HL}^+$, we have $\mathrm{Null}(\Delta) = \mathrm{Span}(\mathbf{1})$. Therefore, $\mathbb{R}^n$ is the direct sum of the space $\mathrm{Span}(\mathbf{1})$ and of the space $\mathrm{Null}(\Delta)^\perp$. Since $f \in \mathbb{R}^n$, it can be written $f = \mu\mathbf{1} + g$ where $g \in \mathrm{Null}(\Delta)^\perp$. Thus, we have $\Delta f = \Delta(\mu\mathbf{1} + g) = \Delta g$.

Let us suppose that f satisfies $\Delta f = \mathbf{In}_j$, we deduce that $\Delta g = \mathbf{In}_j$. Now, from properties of the Moore-Penrose pseudoinverse, we know that the operator $\Delta^\dagger\Delta$ is the orthogonal projector operator on $\mathrm{Null}(\Delta)^\perp$. Since $g \in \mathrm{Null}(\Delta)^\perp$, we have $g = \Delta^\dagger\Delta g$. Consequently, using the fact that $\Delta g = \mathbf{In}_j$, we obtain $g = \Delta^\dagger \mathbf{In}_j$. Hence we can write f under the form $\mu\mathbf{1} + \Delta^\dagger\mathbf{In}_j$.

Conversely, let us consider $f = \mu\mathbf{1} + \Delta^\dagger\mathbf{In}_j$ with $\mu \in \mathbb{R}$. We have $\Delta f = \Delta\left(\mu\mathbf{1} + \Delta^\dagger\mathbf{In}_j\right) = \mu\Delta\mathbf{1} + \Delta\Delta^\dagger\mathbf{In}_j$. From the definition of $\mathbf{In}_j$, we deduce that $\mathbf{In}_j \in \mathrm{Null}(\Delta)^\perp$. Since Δ is symmetric, $\Delta\Delta^\dagger$ is also the orthogonal projector on $\mathrm{Null}(\Delta)^\perp$ and thus $\Delta\Delta^\dagger\mathbf{In}_j = \mathbf{In}_j$. Since $\Delta\mathbf{1} = 0$, we get $\Delta f = \mathbf{In}_j$ which concludes the proof. ■

As a consequence, we can write $\mathrm{V}(i, j) = f(i) - f(j) = (\mathbf{e}_i - \mathbf{e}_j)^T\Delta^\dagger\mathbf{In}_j$ which is independent of the choice of f in the set of solutions of the equation $\Delta f = \mathbf{In}_j$. This leads to the equivalent definition of the diffusion function

$$\mathrm{V}(i, j) = (\mathbf{e}_i - \mathbf{e}_j)^T\Delta^\dagger\mathbf{In}_j$$

and we can now give the main proposition relating the potential distance and the diffusion function.

Proposition 18 *Let $\mathbf{h} = (N, H)$ be a hypergraph such that $\Delta \in \mathcal{HL}^+$. For every i, j in N, we have $\Omega(i, j) = \mathrm{V}(i, j) + \mathrm{V}(j, i)$, and $\mathrm{V}(i, j)$ satisfies*

$$\begin{cases} \mathrm{V}(i, j) = \displaystyle\sum_{h|i\in h} \frac{w_h(i)}{d(i)} \left[1 + \sum_{k\in h, k\neq i} P_h(i, k)\sqrt{\frac{w_h(k)}{w_h(i)}}\mathrm{V}(k, j)\right] & \text{if } i \neq j \ , \\ \mathrm{V}(i, i) = 0 \ . \end{cases} \tag{19}$$

Proof First, we show that $\Omega(i,j) = \frac{\mathrm{V}(i,j)+\mathrm{V}(j,i)}{\mathrm{Vol}(\mathbf{h})}$. For that, let us develop the expression of $\mathrm{V}(i,j)$ obtained above

$$\begin{aligned}\mathrm{V}(i,j) =& (\mathbf{e}_i - \mathbf{e}_j)^T \Delta^\dagger \mathbf{In}_j \\ =& \sum_{k\neq j} d(k)(\Delta^\dagger_{k,i} - \Delta^\dagger_{k,j}) - (\mathrm{Vol}(\mathbf{h}) - d(j))(\Delta^\dagger_{i,j} - \Delta^\dagger_{j,j}) \\ =& \sum_{k\neq i,j} d(k)(\Delta^\dagger_{k,i} - \Delta^\dagger_{k,j}) + d(i)(\Delta^\dagger_{i,i} - \Delta^\dagger_{i,j}) + (\mathrm{Vol}(\mathbf{h}) - d(j))(\Delta^\dagger_{j,j} - \Delta^\dagger_{i,j}) \\ =& \mathrm{Vol}(\mathbf{h})(\Delta^\dagger_{j,j} - \Delta^\dagger_{i,j}) + R(j,i) \ \ ,\end{aligned}$$

where

$$R(j,i) = \sum_{k\neq i,j} d(k)(\Delta^\dagger_{k,i} - \Delta^\dagger_{k,j}) + d(i)(\Delta^\dagger_{i,i} - \Delta^\dagger_{i,j}) - d(j)(\Delta^\dagger_{j,j} - \Delta^\dagger_{i,j}) \ \ .$$

We can observe that, for every node i and j in N, we have $R(i,j) + R(j,i) = 0$. Thus, we can write

$$\begin{aligned}\mathrm{V}(i,j) + \mathrm{V}(j,i) =& \mathrm{Vol}(\mathbf{h})(\Delta^\dagger_{i,i} + \Delta^\dagger_{j,j} - \Delta^\dagger_{i,j} - \Delta^\dagger_{j,i}) \\ =& \mathrm{Vol}(\mathbf{h})(\Delta^\dagger_{i,i} + \Delta^\dagger_{j,j} - 2\Delta^\dagger_{i,j}) \\ =& \Omega(i,j) \ \ ,\end{aligned}$$

which concludes the first part of the proof.

It remains to show that V satisfies Equation (19). By definition of V, we get that, for every node i, $\mathrm{V}(i,i) = 0$. Let us now consider $i \neq j$ and let us consider f in $\mathcal{S}_j$, i.e., a solution of $\Delta f = \mathbf{In}_j$. As $i \neq j$, we have $d(i) = \mathbf{e}_i^T \mathbf{In}_j$. Since $f \in \mathcal{S}_j$, we have $\mathbf{In}_j = \Delta f = G^T G f$ so we can rewrite the previous equality as $d(i) = \mathbf{e}_i^T G^T G f = (G\mathbf{e}_i)^T (Gf)$. Now, because of the Equilibrium Condition, we have $G\mathbf{1} = \mathbf{0}$, thus $Gf = G(f - f(j)\mathbf{1})$. This leads to

$$\begin{aligned}d(i) &= (G\mathbf{e}_i)^T(G(f - f(j)\mathbf{1}) \\ &= \sum_h \left(G\mathbf{e}_i\right)(h) \cdot \left(G(f - f(j)\mathbf{1}\right)(h) \\ &= \sum_h \left(\sqrt{w_h(i)}\epsilon_h(i)\right) \cdot \left(\sum_{k\in h} \sqrt{w_h(k)}\epsilon_h(k)(f(k) - f(j))\right) \quad \text{(using Equation (13))} \\ &= \sum_{h|i\in h} \sqrt{w_h(i)} \left\{\sum_{k\in h} (-P_h(k,i))\sqrt{w_h(k)}(f(k) - f(j))\right\} \quad \text{(using Equation (14))} \\ &= \sum_{h|i\in h} \sqrt{w_h(i)} \left\{\sum_{k\in h} (-P_h(k,i))\sqrt{w_h(k)}\mathrm{V}(k,j)\right\} \quad \text{(by Def. of V, } \mathrm{V}(k,j) = f(k) - f(j)\text{)} \\ &= \mathrm{V}(i,j) \sum_{h|i\in h} w_h(i)(-P_h(i,i)) + \sum_{h|i\in h} \sqrt{w_h(i)} \left\{\sum_{k\in h, k\neq i} (-P_h(k,i))\sqrt{w_h(k)}\mathrm{V}(k,j)\right\} \ \ .\end{aligned}$$

We have for all i, $P_h(i,i) = -1$ and $\sum_h w_h(i) = d(i)$ so the previous equality can be rewritten under the form

$$d(i) = \mathrm{V}(i,j)d(i) + \sum_{h|i\in h} \sqrt{w_h(i)} \left\{ \sum_{k\in h, k\neq i} (-P_h(k,i))\sqrt{w_h(k)}\mathrm{V}(k,j) \right\} \ .$$

Hence, we get the linear system

$$\begin{aligned}
\mathrm{V}(i,j) &= 1 + \sum_{h|i\in h} \frac{\sqrt{w_h(i)}}{d(i)} \left\{ \sum_{k\in h, k\neq i} P_h(k,i)\sqrt{w_h(k)}\mathrm{V}(k,j) \right\} \\
&= \sum_{h|i\in h} \frac{w_h(i)}{d(i)} \left\{ 1 + \sum_{k\in h, k\neq i} P_h(k,i)\sqrt{\frac{w_h(k)}{w_h(i)}}\mathrm{V}(k,j) \right\} \ ,
\end{aligned}$$

which concludes the proof. ■

It should be noted that the above proof generalizes over the classic proof for graphs based on electrical equivalence. Indeed, the proof from Chandra et al. (1996) considers an electrical network where each edge of the original graph is replaced by a one Ohm resistor. He shows that, when we inject $d(r)$ unit of current in each node r and remove an equivalent quantity from a specific sink node j, the difference of potential between a random node i and the sink node j is proportional to the hitting-time distance from i to j. Please note that this definition of the input current is equivalent to our input function **In**. Such a system can be seen as a density of charge that diffuses through an electrical network. Chandra et al. (1996) leverages the classic laws of electrostatic to solve this problem (Ohm's law and Kirchoff's law) but from a more general perspective, the diffusion of a charge in a continuous system can be described by Poisson's equation for electrostatics

$$\Delta V = \frac{-\rho}{\epsilon} \ ,$$

where ρ describes the charges brought from outside (free charge density) and ϵ is a constant depending on the material. This equation is similar to the diffusion equation $\Delta f = \mathbf{In}$ that links an input function **In** with a function f which can be seen as the potential function of the system. Thus, our definition of $\mathrm{V}(i,j) = f(i) - f(j)$ is compliant with the one of Chandra et al. (1996) since it denotes the difference of potential between a node i and the system sink node j (i.e., $V_i - V_j$).

When the hypergraph is a graph, all hyperedges h that contain i are simple edges $\{\{i\},\{k\}\}$ with $k \in N$. In Equation (19), $w_h(i)$ and $w_h(k)$ reduces to $W_{i,k}$ and the linear system reduces to

$$\begin{cases}
\mathrm{V}(i,j) = \sum_{k\in N} \frac{W_{i,k}}{d(i)}(1 + \mathrm{V}(k,j)) & \text{if } i \neq j \ , \\
\mathrm{V}(i,i) = 0 \ .
\end{cases}$$

Consequently, $\mathrm{V}(i,j)$ can be interpreted as the hitting-time distance from i to j (average number of steps needed by a random walker to travel from i to j). Therefore, the potential distance $\Omega(i,j)$ coincides with the commute-time distance divided by the overall volume in the case of graphs (see also Klein and Randić, 1993; Chandra et al., 1996; Fouss et al., 2007).

In the general hypergraph case, the situation is more elaborated. Indeed, let us define $p(h|i) = \frac{w_h(i)}{d(i)}$ and $p(k|h,i) = P_h(i,k)\sqrt{\frac{w_h(k)}{w_h(i)}}$. Then, we can rewrite the Equation (19) as

$$\begin{cases} \mathrm{V}(i,j) = \sum_{h|i\in h} p(h|i) \left[1 + \sum_{k\in h, k\neq i} p(k|h,i)\mathrm{V}(k,j)\right] & \text{if } i \neq j \ , \\ \mathrm{V}(i,i) = 0 \ . \end{cases}$$

Notice that $p(h|i)$ is non-negative and that $\sum_h p(h|i) = 1$. Thus, $p(h|i)$ can be interpreted as a jumping probability from i to the hyperedge h. We also have $\sum_n p(k|h,i) = 1$ but $p(k|h,i)$ is negative as soon as i and k belong to the same end of h. This prevents us from interpreting this quantity as a jumping probability from i to k by h. Therefore, in the general case, we do not have a random walk interpretation of the potential distance. Notice however that several theoretical approaches coming from the world of quantum physics have been built to take into consideration quantities like $p(k|h,i)$. (See for example Burgin, 2010).

Another way to apprehend the potential distance in hypergraphs is to rewrite Equation (19) as

$$\begin{aligned} \mathrm{V}(i,j) &= 1 + \sum_{h|i\in h} \sum_{\substack{k\in h \\ k\neq i}} \frac{W_{i,k}}{d(i)} \mathrm{V}(k,j) \\ &= 1 + \sum_{k\in N} P_{i,k} \mathrm{V}(k,j) \ , \end{aligned}$$

where W is the pairwise weight matrix of $\mathbf{h}$ and P is the *transition matrix* defined by $P = D^{-1}W$. The rows of P sums to 1 but P is not a stochastic matrix since it can contain negative values. However, P can still be seen as a transition matrix associated with the reduced signed graph of $\mathbf{h}$. Let us now briefly discuss the existence of a stationary distribution based on the transition matrix P. In the graph case, the transition matrix P is a stochastic matrix. For connected graphs we can apply Perron-Frobenius Theorem to show the uniqueness of a stationary distribution of nodes. In the hypergraph case, this is not possible in general but we can still exhibit a stationary distribution based on the degrees and we study the connectivity notion in the next section.

Proposition 19 *Let* $\mathbf{h} = (N,H)$ *be a hypergraph with transition matrix* P. *The vector* $\pi = \frac{1}{\mathrm{Vol}(\mathbf{h})}(d(0),\ldots,d(n))^T$ *is a probability vector and satisfies*

$$\pi P = \pi \ .$$

Proof It is easy to see that $\sum_{i\in N} \pi_i = 1$ since $\mathrm{Vol}(\mathbf{h}) = \sum_{i\in N} d(i)$. Let us consider a node $i \in N$. From the definition of degrees, $d(i) > 0$, thus $\pi_i > 0$. So π is a probability vector and:

$$(\pi P)_i = \sum_{k \in N} \pi_k P_{k,i} = \frac{1}{\mathrm{Vol}(\mathbf{h})} \sum_{k \in N} d(k) \frac{W_{k,i}}{d(k)} = \frac{1}{\mathrm{Vol}(\mathbf{h})} d(i) = \pi_i \enspace .$$

■

4.3 Paths, connected components and independent components

We define the notion of path in hypergraphs as a path in the reduced signed graph. It should be noted that this definition of path is consistent with hypergraph equivalence since two equivalent hypergraphs share the same reduced signed graph. When the hypergraph is a graph, the definition coincides with the definition of path in undirected graphs since the pairwise matrix is equal to the adjacency matrix.

Definition 20 *Let $\mathbf{h} = (N, H)$ be a hypergraph with pairwise weight matrix W. A* path *between two nodes i and j is a sequence of nodes $u_1 = i, u_1, \ldots, u_{m-1}, u_m = j$ such that $W_{\ell,\ell+1} \neq 0$ for $1 \leq \ell < m$.*

Definition 20 allows us to define the notion of connected components in hypergraphs. Recall that in the graph case, connected components can be characterized using the null space of the Laplacian matrix (see Section 2.1). We now study whether such properties can be derived in the hypergraph case.

Definition 21 *A connected component of a hypergraph is a maximal connected set, i.e., a maximal set of nodes such that there exists a path between any two nodes.*

As in the graph case, we now show that any constant function defined on a connected component is in the null space of the Laplacian matrix. This property can be interpreted as an independence property since we can define independently a constant label for each connected component, without modifying the global smoothness.

Proposition 22 *Let $C_1, \ldots, C_l$ be l connected components of a hypergraph $\mathbf{h}$. Let Δ be the unnormalized Laplacian of $\mathbf{h}$. We have*

$$\mathrm{Span}(\mathbf{1}_{C_1}, \ldots, \mathbf{1}_{C_p}) \subseteq \mathrm{Null}(\Delta) \enspace . \tag{20}$$

Proof Let $\mathbf{h} = (N, H)$ be a hypergraph and let $S \subset N$ be a connected component. Let i be a node in S. By definition S is a maximal connected set, hence $\forall j \in N \backslash S, \;\; W_{i,j} = 0$ and we have

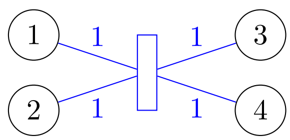
$$(\Delta \mathbf{1}_S)(i) = d(i) - \sum_{j \in S} W_{i,j} = d(i) - \sum_{j \in N} W_{i,j} = 0 \enspace .$$

For $i \notin S$, we have

$$(\Delta \mathbf{1}_S)(i) = 0 - \sum_{j \notin S} W_{i,j} = 0 \enspace .$$

Finally, $\Delta \mathbf{1}_S = 0$, which concludes the proof. ■

The reader should note that contrarily to the graph case where the inclusion in Equation (20) is an equality, the dimension of the null space of the hypergraph Laplacian does not define the number of disjoint components in the hypergraph. From the flow point of view, it is even possible to find non disjoint sets of nodes on which any constant function nullify the hypergraph Laplacian. As an example, consider the hypergraph depicted in Figure 9. The application of the Laplacian is null on a constant function applied on $\{1, 3\}$ (for instance the vector $(1, 0, 1, 0)^T$).



$$\Delta = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix}$$

Figure 9: Hypergraph $\mathbf{h}_8$ and its Laplacian

Let $\mathbf{h} = (N, H)$ be a hypergraph and W its edge weight matrix. Let $S \subseteq N$ be a subset of nodes, a node i in N is said to be *independent* of S if the contribution of the nodes in S to the degree of i is 0, i.e., $\sum_{j \in S} W_{i,j} = 0$. The set S is an *independent component* of $\mathbf{h}$ if every node in $N \backslash S$ is independent of S and if every node in S is independent of $N \backslash S$. Let us consider the hypergraph in Figure 9, its independent components are $\{1, 2, 3, 4\}$,$\{1, 4\}$, $\{1, 3\}$, $\{2, 3\}$ and $\{2, 4\}$.

Proposition 23 *A set S of nodes is an independent component of a hypergraph $\mathbf{h}$ if and only if its indicator function $\mathbf{1}_S$ is in $\mathrm{Null}(\Delta)$. As a consequence, if $C_1, \ldots, C_p$ are the independent components of $\mathbf{h}$, we have*

$$\mathrm{Span}(\mathbf{1}_{C_1}, \ldots, \mathbf{1}_{C_p}) \subseteq \mathrm{Null}(\Delta) \ .$$

Proof Let $\mathbf{h} = (N, H)$ be a hypergraph and let W be its edge weight matrix. For any $i \in N$, we have

$$(\Delta \mathbf{1}_S)(i) = d(i) \mathbf{1}_S(i) - \sum_{j \neq i} W_{i,j} \mathbf{1}_S(j) \ .$$

We have $\Delta \mathbf{1}_S = 0$ if and only if for every node $i \in N$, $(\Delta \mathbf{1}_S)(i) = 0$. If $i \notin S$ then$(\Delta \mathbf{1}_S)(i) = \sum_{j \in S} W_{i,j}$. Otherwise, if $i \in S$ then

$$(\Delta \mathbf{1}_S)(i) = d(i) - \sum_{j \in S \backslash \{i\}} W_{i,j} = \sum_{j \notin S} W_{i,j} \ .$$

Therefore we have $\Delta \mathbf{1}_S = 0$ if and only if for all $i \in S$, $\sum_{j \notin S} W_{i,j} = 0$ and for all $i \notin S$, $\sum_{j \in S} W_{i,j} = 0$, which turns out to be the exact definition of an independent component. ■

If S is a connected component, Proposition 22 implies that $\Delta \mathbf{1}_S = 0$ so according to Proposition 23, S is also an independent component. When $\mathrm{Null}(\Delta) = \mathrm{Span}(\mathbf{1})$ or equivalently when $\Delta \in \mathcal{HL}^+$, the hypergraph $\mathbf{h}$ is said to be *strongly connected*.

Proposition 24 *Let $\mathbf{h}(N, H)$ be a hypergraph with Laplacian matrix Δ. Then*

1. *if $\mathbf{h}$ is strongly connected then N is the only connected component of $\mathbf{h}$,*
2. *the class of strongly connected hypergraphs can be embedded with a complete Riemannian structure,*
3. *if $\mathbf{h}$ is strongly connected then the potential distance Ω is a distance on N.*

Proof If $\mathbf{h}$ is strongly connected then $\mathrm{Null}(\Delta) = \mathrm{Span}(\mathbf{1}_N)$ so, following Proposition 23, N is the only independent component of $\mathbf{h}$. Since a connected component is always an independant component, we only have to prove that $\mathbf{h}$ contains at least a connected component. This result is direct since the node singletons are always connected. (2) follows from (1) and the propositions of Section 2.4. (3) is a consequence of Proposition 16. ■

It should be noted that, unfortunately, there exists hypergraphs where N is the only one connected component and where $\mathrm{Span}(\mathbf{1}) \subsetneq \mathrm{Null}(\Delta)$. As an example, consider the hypergraph $\mathbf{h}_9$ in Figure 10. This hypergraph has only one independent component ($N = \{1, 2, 3\}$) but has two null eigenvalues ($G \in \mathbb{R}^{1\times 3}$ so $\mathrm{Rank}(\Delta) = 1$). Thus $\mathbf{h}_9$ is not strongly connected.

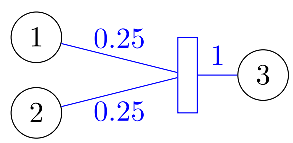


Figure 10: Hypergraph $\mathbf{h}_9$ is connected but not strongly connected.

5. Conclusion

We have introduced, both from a geometrical perspective and from an algebraic perspective, a novel notion of undirected hypergraphs that generalizes the notion of undirected graphs. We have defined a complete spectral framework for this class of undirected hypergraphs through the notion of hypergraph Laplacian. This will allow to naturally extend learning algorithms, e.g., spectral clustering algorithms or semi-supervised learning algorithms based on Laplacian harmonicity, from the graph case to the hypergraph case. Undirected hypergraphs and their spectral theory allow us to encode similarity and dominance between sets of collaborating nodes and, therefore, we strongly believe that it opens the way to solve new learning tasks. Following this idea, we are currently investigating the modeling of games between teams of players in order to infer player-level information such as ranking players or quatifying the complementarity of players.

We have highlighted some interesting relations between our hypergraphs and signed graphs. However, much remain to be done in this area to understand more deeply the semantic links between these concepts. Similarly, the generalized notion of potential distance offers new perspectives to interpret the smoothness on an undirected hypergraph but still lacks of a constructive definition of random walks in undirected hypergraphs. These important challenges are part of the topics we want to tackle in the future.

References

Andreas Argyriou, Mark Herbster, and Massimiliano Pontil. Combining graph laplacians for semi-supervised learning. In *Proceedings of the 19th conference Annual Conference on Neural Information Processing Systems (NIPS-05)*, pages 67–74, Cambridge, MA, 2006. MIT Press.

Claude Berge. *Hypergraphs: combinatorics of finite sets*. North-Holland Mathematical Library. Elsevier, Burlington, MA, 1989.

Silvere Bonnabel and Rodolphe Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1055–1070, 2009.

Mark Burgin. Interpretations of negative probabilities. *arXiv preprint arXiv:1008.1287*, 2010. URL `http://arxiv.org/abs/1008.1287`.

Riccardo Cambini, Giorgio Gallo, and MariaGrazia Scutella. Flows on hypergraphs. *Mathematical Programming*, 78(2):195–217, 1997.

Ashok K. Chandra, Prabhakar Raghavan, Walter L. Ruzzo, Roman Smolensky, and Prasoon Tiwari. The electrical resistance of a graph captures its commute and cover times. *Computational Complexity*, 1996.

Francois Fouss, Alain Pirotte, Jean-Michel Renders, and Marco Saerens. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *Knowledge and Data Engineering, IEEE Transactions on*, 19(3):355–369, 2007.

Giorgio Gallo, Giustino Longo, Stefano Pallottino, and Sang Nguyen. Directed hypergraphs and applications. *Discrete Applied Mathematics*, 42(2-3):177–201, 1993.

Giorgio Gallo, Claudio Gentile, Daniele Pretolani, and Gabriella Rago. Max horn sat and the minimum cut problem in directed hypergraphs. *Mathematical Programming*, 80(2): 213–237, 1998.

Jingrui He and Rick Lawrence. A graph-based framework for multi-task multi-view learning. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 25–32, 2011.

Mark Herbster, Massimiliano Pontil, and Lisa Wainer. Online learning over graphs. In *Proceedings of the 22nd International Conference on Machine Learning (ICML-05)*, pages 305–312. ACM, 2005.

Yao Ping Hou. Bounds for the least laplacian eigenvalue of a signed graph. *Acta Mathematica Sinica*, 21(4):955–960, 2005.

Douglas J. Klein and M. Randić. Resistance distance. *Journal of Mathematical Chemistry*, 12(1):81–95, 1993.

Jérôme Kunegis, Stephan Schmidt, Andreas Lommatzsch, Jürgen Lerner, Ernesto William De Luca, and Sahin Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proceedings of the 10th SIAM International Conference on Data Mining (SDM-10)*, volume 10, pages 559–559, 2010.

Alain Rakotomamonjy, Francis Bach, Stéphane Canu, Yves Grandvalet, et al. Simplemkl. *Journal of Machine Learning Research*, 9:2491–2521, 2008.

Marco Saerens, Francois Fouss, Luh Yen, and Pierre Dupont. The principal components analysis of a graph, and its relationshipsto spectral clustering. In *Proceedings of the 15th European Conference on Machine Learning (ECML-04)*, pages 371–383. Springer, 2004.

Nino Shervashidze, Pascal Schweitzer, Erik Jan Van Leeuwen, Kurt Mehlhorn, and Karsten M Borgwardt. Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research*, 12: 2539–2561, 2011.

Vikas Sindhwani, Partha Niyogi, and Mikhail Belkin. A co-regularization approach to semi-supervised learning with multiple views. In *Proceedings of ICML Workshop on Learning with Multiple Views*, pages 74–79. Citeseer, 2005.

Xilan Tian, Gilles Gasso, and S Canu. A multi-kernel framework for inductive semi-supervised learning. In *Proceedings of the 18th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN-11)*, pages 27–29, 2011. ISBN 9782874190445.

Koji Tsuda, Hyunjung Shin, and Bernhard Schölkopf. Fast protein classification with multiple networks. *Bioinformatics*, 21(suppl 2):ii59–ii65, 2005.

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

Wei Wang and Zhi-Hua Zhou. A new analysis of co-training. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 1135–1142, 2010.

Zenglin Xu, Rong Jin, Haiqin Yang, Irwin King, and Michael R Lyu. Simple and efficient multiple kernel learning by group lasso. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 1175–1182, 2010.

Dengyong Zhou and Christopher JC Burges. Spectral clustering and transductive learning with multiple views. In *Proceedings of the 24th International conference on Machine learning (ICML-07)*, pages 1159–1166. ACM, 2007.

Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning from labeled and unlabeled data on a directed graph. In *Proceedings of the 22nd International conference on Machine learning (ICML-05)*, pages 1036–1043. ACM, 2005.

Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *Proceedings of the 20th conference Annual Conference on Neural Information Processing Systems (NIPS-06)*, pages 1601–1608, Cambridge, MA, 2007a. MIT Press.

Ding Zhou, Sergey A Orshanskiy, Hongyuan Zha, and C Lee Giles. Co-ranking authors and documents in a heterogeneous network. In *Proceedings of the 7th IEEE International Conference on Machine Learning (ICDM-07)*, pages 739–744. IEEE, 2007b.

Ding Zhou, Shenghuo Zhu, Kai Yu, Xiaodan Song, Belle L Tseng, Hongyuan Zha, and C Lee Giles. Learning multiple graphs for document recommendations. In *Proceedings of the 17th international conference on World Wide Web (WWW-08)*, pages 141–150. ACM, 2008.

Xiaojin Zhu, Zoubin Ghahramani, John Lafferty, et al. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International conference on Machine learning (ICML-03)*, volume 3, pages 912–919, 2003.