



HAL
open science

The Currents Space of Graphs

James Benn, Anna Calissano, Stephen Marsland, Xavier Pennec

► **To cite this version:**

James Benn, Anna Calissano, Stephen Marsland, Xavier Pennec. The Currents Space of Graphs. 2022. hal-03910825

HAL Id: hal-03910825

<https://inria.hal.science/hal-03910825>

Preprint submitted on 22 Dec 2022

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

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

The Currents Space of Graphs

James Benn · Anna Calissano · Stephen Marsland · Xavier Pennec

Received: 22.12.2022

Abstract We represent simple graphs on n vertices by 1-forms over a complete graph K_n . It is shown that these 1-forms lie on a hypersphere in the Hilbert space $L^2(\Omega_{K_n}^1)$ of all 1-forms and the round metric induces a natural distance metric on graphs. The metric itself reduces to a global spherical Jaccard-type metric which can be computed from edge-list data alone. Lastly, we describe the global structure of the 1-forms representing graphs and how these can be exploited for the statistical analysis of a sample of graphs.

1 Introduction

Analysing sets of graphs is crucial in different applications such as social networks of different users, brain connectivity networks of different patients, protein-protein interactions, public transport systems, and stocks market correlations. When moving from a single graph to a set of graphs there may or may not be node correspondences

James Benn
INRIA Sophia Antipolis and University of Côte d'Azur
2004 Rte des Lucioles, 06902 Valbonne, France
E-mail: james.benn@inria.fr

Anna Calissano
INRIA Sophia Antipolis and University of Côte d'Azur
2004 Rte des Lucioles, 06902 Valbonne, France
E-mail: anna.calissano@inria.fr

Stephen Marsland
Victoria University of Wellington
PO Box 600, Wellington 6140, New Zealand
E-mail: stephen.marsland@vuw.ac.nz

Xavier Pennec
INRIA Sophia Antipolis and University of Côte d'Azur
2004 Rte des Lucioles, 06902 Valbonne, France
E-mail: xavier.pennec@inria.fr

across the graphs - these are referred to as labelled or unlabelled graphs, respectively, [2] - for example, social networks of different users on a social platform are treated as unlabelled graphs because different users have different friends which are not directly comparable; —. In all these phenomena, sets of graphs or networks arise naturally and both a graph representation and data analysis tools are required. Defining a common embedding space for a set of graphs is the first step in their statistical analysis and the majority of data analysis methods developed are based on graph matrix representations which are suited for graph classification or regression. However, if the final aim is to interpolate between graphs or develop dimensionality reduction methods which describe common topological substructures of graphs then a purely topological representation is preferable. In addition, a representation which allows for an analysis of graphs using the natural format in which graph data is presented - namely, edge lists - would have a pre-processing advantage and avoid the zero node padding problem which is necessary when graphs of different dimensions are compared. This article presents a new geometric representation of finite simple graphs, both labelled and unlabelled, and a new distance metric based on edge lists which can be used for graph comparison. The representation is a re-tooling of ideas developed for the measurement and statistical analysis of shapes in [1]. The construction and results are as follows.

Fix the complete graph K_n on n vertices and label the vertices V from 1 to n . The labelling gives rise to orientations of the edges $E = \{(i, j) : i, j \in V\}$, which we set to be $i < j$, and permits an integration of alternating functions $\alpha(i, j)$ on E (1-forms), and an integration of 1-forms over subgraphs of K_n . Equip the space of 1-forms on K_n with an L^2 inner product

$$\langle \alpha, \beta \rangle_{L^2} = \sum_{i < j} \alpha(i, j) \cdot \beta(i, j)$$

where the orientation $i < j$ ensures that each edge is counted exactly once. Since K_n is finite it is clear that all 1-forms are bounded in the L^2 metric but we will denote the resulting Hilbert space by $L^2(\Omega_{K_n}^1)$ to indicate the inner product in play.

Let G be any simple graph (a graph without double edges or self loops) on n vertices, containing at least one edge, and consider the set of all embeddings $\mathcal{E}(G, K_n)$ of G into K_n . By an embedding of G we mean an injective map φ carrying the vertices of G onto a subset of vertices of K_n and preserving the edge adjacency relations. The set $\mathcal{E}(G, K_n)$ functions as a proxy for the isomorphism class of G since, by definition, it encodes all the isomorphic images of G as subgraphs of K_n . To each embedding φ we associate a bounded linear functional on $L^2(\Omega_{K_n}^1)$ via the current map

$$C : \mathcal{E}(G, K_n) \rightarrow L^2(\Omega_{K_n}^1)^*$$

$$\varphi \mapsto \int_{\varphi(G)}$$

where the integration is understood as the summation of the values of 1-forms over the edges of $\varphi(G)$ according to the orientation of K_n so that each edge is counted exactly once. Just as in the continuous case ([6], [14]), this definition of integration uniquely characterizes the domain $\varphi(G)$ (Proposition 2.2), which, in conjunction with the

Riesz Representation Theorem, allows us to represent the image of each embedding $\varphi \in \mathcal{E}(G, K_n)$ by a unique L^2 1-form $\alpha_{\varphi(G)} \in L^2(\Omega_{K_n}^1)$. We have called the map C the current map because it is the discrete analogue of the integrable de-Rham currents ([4]).

Performing this construction for all graphs with at least one edge and n vertices - call this set \mathcal{G}_+^n - we find that the geometric structure of the set of resulting 1-forms is quite tractable. In particular, we show that

$$\alpha_{\varphi(G)} = c \cdot \alpha_{\psi(H)}$$

for some $0 \neq c \in \mathbb{R}$, if and only if $\varphi(G)$ and $\psi(H)$ are identical as subgraphs of K_n , in which case $c = 1$, and G and H are isomorphic. This gives rise to the Structure Theorem (Theorem 4.4) which provides us with an injective map from graphs into the positive orthant of the unit sphere in $L^2(\Omega_{K_n}^1)$

$$F : \bigcup_{\mathcal{G}_+^n} \mathcal{E}(G, K_n) \rightarrow \mathbb{S}_{L^2(\Omega_{K_n}^1)} \quad (1)$$

$$\varphi_G \mapsto \frac{\alpha_{\varphi(G)}}{\|\alpha_{\varphi(G)}\|_{L^2}},$$

where $\|\alpha_{\varphi(G)}\|_{L^2} = \sqrt{|E_G|}$, and $\alpha_{\varphi(G)}(i, j) = 1$ if $i < j$ and $(i, j) \in E_{\varphi(G)}$, and 0 otherwise.

A natural distance metric on non-degenerate isomorphism classes of graphs on n vertices is induced by the round metric on the unit sphere (Theorems 5.2):

$$d([G], [H]) = \min_{\varphi \in \mathcal{E}(G, K_n), \psi \in \mathcal{E}(H, K_n)} \arccos \left(\langle F(\varphi), F(\psi) \rangle_{L^2(\Omega_{K_n}^1)} \right), \quad (2)$$

which is non-negative, satisfies $d([G], [H]) = 0$ if and only if $[G] = [H]$, and satisfies the triangle inequality. Formula (2) is a geometric/functional-analytic expression for the distance metric but there is a familiar set-theoretic expression for it that does *not* require the enumeration and evaluation of graph embeddings - all one needs are the edge lists of the two graphs being compared. The minimisation of the arccos function is equivalent to the maximisation of its argument, and the maximisation of the L^2 inner-product between $F(\varphi)$ and $F(\psi)$ can be realized as the maximum number of edges G and H can have in common when "laid over the top of each other"; that is,

$$d([G], [H]) = \arccos \left(\max \frac{|E_G \cap E_H|}{\sqrt{|E_G|} \cdot \sqrt{|E_H|}} \right), \quad (3)$$

where the maximisation of the intersection cardinality is performed over vertex/edge relabellings. One can recognise this expression as a spherical "Jaccard-type" distance metric. A local variant of (3) can be found in [5] and is used to assess the probability that two vertices should be connected or not (see [8] for another application in link-prediction) - incidentally, this local variant had been used in text-mining since the 80s (see [10] for example). Classical Jaccard coefficients/metrics have also been used to

find communities in graphs ([9], [11], and [12]). Similarly, our metric (3) could be used to find substructures common to a collection of graphs by exploiting the image of the map F , and the geometry of the sphere.

Although discrete optimization is challenging, our construction represents graphs by points in a sphere - a continuous space with a well understood geometry. The discrete image of the map (1) has a particular geometric structure and forms a kaleidoscope of concentrically circumscribed polygons, all centered on the unique representative of K_n . The geometric structure of this image, as well as the continuous background geometry, could be used to perform constrained gradient ascent/descent to deal with the maximisation component of the metric (2). The development of this aspect of the dual space of graphs requires a somewhat separate analysis and, therefore, will be left for future research.

In section 2 we introduce the central tools used in our construction. These are the Riesz Representation Theorem, and the combinatorial analogues of orientation, differential forms, and integration on graphs. Section 3 assembles the current map and generates the duality between graphs and 1-forms on an ambient graph using the instruments of section 2. The geometric structure of the set of dual 1-forms is analyzed in section 4, which culminates in The Structure Theorem and the kaleidoscope map F . Section 5 describes the resulting distance metric on isomorphism classes of graphs on n vertices and an obvious extension to labelled graphs, its features, as well as a description of the image of the map F and how it could be used to improve the computation of the metric. In section 6 we describe some possible modifications to the framework and in section 7 we discuss some particular features of the framework which can be exploited for analysing populations of networks.

2 Notation and Preliminary Considerations

All graphs considered here will be simple graphs $G = (V, E)$ with a set of vertices, a set of edges, and a set of edge-vertex adjacency relations. In order to describe the main constructions, we first need to introduce the central functional analytic and graph theoretic tools.

2.1 The Riesz Representation Theorem

A real Hilbert space H is a complete linear space provided with an inner product, denoted $\langle \cdot, \cdot \rangle$, and satisfying:

$$\begin{aligned} \langle ax_1 + bx_2, y \rangle &= a\langle x_1, y \rangle + b\langle x_2, y \rangle \quad \forall a, b \in \mathbb{R} \\ \langle x, y \rangle &= \langle y, x \rangle \\ \langle x, x \rangle &> 0 \quad \text{unless } x = 0 \end{aligned}$$

If we fix an element $y \in H$, then the expression $F_y(x) = \langle y, x \rangle$ assigns to each $x \in H$ a real number. Observe that F_y is linear in x and is bounded. We call F_y a bounded linear functional on H . It is a Theorem of Riesz that these are the only bounded linear functionals on H , which we will denote by H^* :

Theorem 1 *For every bounded linear functional $F \in H^*$ there exists a unique $y \in H$ such that*

$$F(x) = \langle y, x \rangle, \quad \forall x \in H.$$

Furthermore, the Riesz map, given by $\mathcal{R} : H^ \rightarrow H$ and sending F to its corresponding y , is an isometric isomorphism.*

The particular Hilbert spaces considered here will be the Hilbert spaces of L^2 p -forms over a graph G ; the linear functionals of interest will be the combinatorial analogue of integrable currents defined by integration over subgraphs. The remainder of this section will introduce the elementary concepts of orientation and integration in, and over, a graph.

2.2 Orientation and Integration

Denote by G_p the set of complete K_{p+1} subgraphs of a fixed graph G - elements of G_p are often referred to as *cliques*; for example, $G_0 = V$, $G_1 = E$, and $G_2 = T$ - the set of all triangles in G . A 0-form on G is a function of its vertices; for example, the vertex degree function is a zero form on a graph G . A 1-form is a function on E which is antisymmetric in its 2 arguments - these being the vertices, or end-points, of the edge. A 2-form is a function on T which is antisymmetric in its 3 arguments - these being the vertices, or corners, of the triangle. The antisymmetry in the 1-forms and 2-forms is there to assign a direction, much like a vector field on a manifold. The set Ω_G^0 is the set of all 0-forms on G , Ω_G^1 the set of all 1-forms on G , and Ω_G^2 the set of all 2-forms on G , and, more generally, Ω_G^p is the set of p -forms on G . The sets of p -forms on a graph G form finite dimensional vector spaces.

Since 1-forms are alternating in their arguments, there is some ambiguity in terms of which sign we start with. To deal with this we introduce an orientation on G . This corresponds to labelling the vertices and then prescribing an order to this labelling. The simplest orientation one can give to the set of edges in a graph is to numerically label the vertices and put them increasing order:

$$\tilde{E} = \{(i, j) : i < j\}.$$

A 1-form on G is now completely determined by the values it takes on \tilde{E} . Similarly, the simplest orientation one can give to the set of triangles in a graph is to traverse it's edges in the direction of increasing order:

$$\tilde{T} = \{(i, j, k) : i < j < k\}.$$

A 2-form on G is now completely determined by the values it takes on \tilde{T} .

There is an obvious L^2 inner product on 0-forms of G , and using the orientations \tilde{E} and \tilde{T} , we may equip the space of 1-forms and 2-forms with L^2 inner products

aswell

$$\begin{aligned}\langle f, g \rangle_V &= \sum_V f(i) \cdot g(i) \\ \langle \alpha, \beta \rangle_E &= \sum_{\bar{E}} \alpha(i, j) \cdot \beta(i, j) \\ \langle \chi, \omega \rangle_T &= \sum_{\bar{T}} \chi(i, j, k) \cdot \omega(i, j, k)\end{aligned}$$

In complete analogy to the continuous setting, the orientations \bar{E} and \bar{T} ensure that each edge is counted exactly once in the sum. The Hilbert spaces of L^2 0-, 1-, and 2-forms on G are denoted by $L^2(\Omega_G^0)$, $L^2(\Omega_G^1)$ and $L^2(\Omega_G^2)$, respectively.

Similarly, one can define an orientation on $(p+1)$ -cliques

$$\bar{\Omega}_G^p = \{(i_1, \dots, i_{p+1}) : i_1 < i_2 < \dots < i_{p+1}\}$$

and define an L^2 metric by

$$\langle \alpha, \beta \rangle_{\bar{\Omega}_G^p} = \sum_{\bar{\Omega}_G^p} \alpha \cdot \beta.$$

Denote the space of L^2 p -forms by $L^2(\Omega_G^p)$.

The orientation of a graph gives a very natural notion of integration of 1-forms over a graph G :

$$\int_G : L^2(\Omega_G^1) \rightarrow \mathbb{R}$$

defined by

$$\int_G \alpha = \sum_{\bar{E}} \alpha.$$

If H is a subgraph of G we can also integrate over H , which inherits an orientation from G by restriction:

$$\int_{H < G} \alpha = \sum_{\bar{E}|_H} \alpha.$$

We say two subgraphs H and K are equal (write $H = K$) if they are exactly the same subgraphs of G , vertex-for-vertex and edge-for-edge. When two graphs (or subgraphs), H and K , are isomorphic, in the usual graph theoretic sense, we write $H \cong K$. In complete analogy with integration theory (cf. Whitney, Harrison), the definition of integration allows us to characterize subgraphs of G by the integration of 1-forms.

Proposition 1 *Suppose H and K are two subgraphs of an oriented graph G and let $L^2(\Omega_G^1)$ be the space of L^2 1-forms over G . Then*

$$\int_H \alpha = \int_K \alpha, \quad \forall \alpha \in L^2(\Omega_G^1)$$

if and only if $H = K$.

Proof If $H = K$ then clearly the integration of all 1-forms over H is identically equal to the integration of all 1-forms over K .

If $H \neq K$ then there exists an edge e contained in H that is not contained in L . Take a 1-form supported on e only (i.e. is non-zero on e and zero everywhere else on G) - the 1-integral of this form over H is non-zero while the 1-integral of this form over K is identically zero. That is, the two integrations are not equal.

3 The Current Space

3.1 The Dual Representation of a Graph

Let K_n be the complete graph on n vertices; fix an orientation of its edges, and let $L^2(\Omega_{K_n}^1)$ be the set of L^2 1-forms on K_n with $L^2(\Omega_{K_n}^1)^*$ its dual. The core idea of this section is predicated on the observation that $L^2(\Omega_{K_n}^1)$ encodes all topological graph structures on n vertices, viewed as subgraphs of K_n .

Let G be a graph on n vertices, with at least one edge, and consider the set of all possible embeddings $\mathcal{E}(G, K_n)$ of G into K_n - by an embedding we mean a map carrying the vertices of G bijectively onto the vertices of K_n and carrying the edges of G onto a subset of the edges of K_n in which edge-vertex adjacency relations are preserved. The set of embeddings $\mathcal{E}(G, K_n)$ represents the isomorphism class of G and the images of elements of $\mathcal{E}(G, K_n)$ are presentations of G as subgraphs of K_n .

Define a map

$$C : \mathcal{E}(G, K_n) \rightarrow L^2(\Omega_{K_n}^1)^*$$

by

$$\varphi \mapsto \int_{\varphi(G)}$$

- 1-integration over $\varphi(G) < K_n$. Since the image of C is a bounded linear functional on $L^2(\Omega_{K_n}^1)$ we can apply the Riesz Representation Theorem - Theorem 2.1:

Proposition 2 For a graph G on n -vertices, each embedding $\varphi \in \mathcal{E}(G, K_n)$ is represented by a unique L^2 1-form on K_n :

$$\alpha_{\varphi(G)} = \mathcal{R} \circ C(\varphi) \in L^2(\Omega_{K_n}^1),$$

where \mathcal{R} is the Riesz map of Theorem 2.1.

Theorem 2 Let G be a graph, and $\mathcal{E}(G, K_n)$ its embedding set. The unique 1-form $\alpha_{\varphi(G)} \in L^2(\Omega_{K_n}^1)$ representing $\varphi \in \mathcal{E}(G, K_n)$ is given by

$$\alpha_{\varphi(G)}(i, j) = \begin{cases} 1 & \text{if } (i, j) \subset E_{\varphi(G)} \text{ and } i < j. \\ -1 & \text{if } (i, j) \subset E_{\varphi(G)} \text{ and } i > j. \\ 0 & \text{otherwise.} \end{cases}$$

where $E_{\varphi(G)}$ denotes the edge set of $\varphi(G) < K_n$.

Proof The equation for $\alpha_{\varphi(G)}$ is defined by the Riesz representation theorem and integration:

$$\sum_{\bar{E}} \alpha_{\varphi(G)}(i, j) \cdot \beta(i, j) = \langle \alpha_{\varphi(G)}, \beta \rangle_E = \int_{\varphi(G)} \beta = \sum_{\bar{E}|_{\varphi(G)}} \beta(i, j).$$

The unique solution to this equation is the one stated in the Theorem; $\alpha_{\varphi(G)}$ is zero on edges not contained in $\varphi(G)$ and is one, or minus one depending on the orientation, on edges contained in $\varphi(G)$.

3.2 The Structure Theorem

Given a graph G we can associate to each presentation of it as a subgraph of K_n a unique 1-form in $L^2(\Omega_{K_n}^1)$; this was the content of section 3. Unfortunately, *each* presentation is assigned it's own 1-form, rather than the isomorphism class of G receiving a unique 1-form. Nevertheless, we are able to determine the geometric structure of the set of all 1-forms representing all possible presentations of all possible graphs on n -vertices, and with least one edge, and then use this structure to place a metric on isomorphism classes of graphs (section 5).

Determining the structure of the set of representing 1-forms will proceed in three stages. First, we fix a graph G and consider the set of all embeddings of G into K_n . From Proposition 2.2 it follows that no two 1-forms representing two distinct embeddings of G are collinear with the origin in $L^2(\Omega_{K_n}^1)$ - this is Lemma 4.1. Next, we fix two graphs G and H , and consider the two sets $\mathcal{E}(G, K_n)$ and $\mathcal{E}(H, K_n)$. Lemma 4.2 shows that the set of 1-forms representing $\mathcal{E}(G, K_n)$ and the set of 1-forms representing $\mathcal{E}(H, K_n)$ are equal as sets if and only if G and H are isomorphic; furthermore, when G and H are *not* isomorphic, the two sets of 1-forms have empty intersection. Lemma 4.3 combines the last two Lemmas and asserts that not two 1-forms representing two different embeddings of any two graphs G and H are collinear with the origin. These three Lemmas culminate in the Structure Theorem - Theorem 4.4 - which defines an injective map from the set of all embeddings of all graphs on n vertices, and having at least one edge, into the unit sphere of $L^2(\Omega_{K_n}^1)$.

Let \mathcal{G}_+^n be the set of all graphs on n vertices and containing at least one edge. Define

$$\mathcal{E}_+^n = \bigcup_{G \in \mathcal{G}_+^n} \mathcal{E}(G, K_n).$$

This is the set of all possible embeddings of all possible graphs on n vertices with at least one edge, into K_n . We now describe the geometric structure of the set $\mathcal{R} \circ C^1(\mathcal{E}_+^n)$, as outlined above.

Lemma 1 *Let φ and ψ belong to $\mathcal{E}(G, K_n)$. Then,*

$$\alpha_\varphi = c \cdot \alpha_\psi$$

for some $0 \neq c \in \mathbb{R}$ if and only if $\psi = \varphi$, in which case $c = 1$. In particular, no two distinct 1-forms α_φ and α_ψ in $\mathcal{R} \circ C^1(\mathcal{E}(G, K_n))$ are collinear with the origin.

Proof Clearly if $\psi = \varphi$ then $\alpha_\varphi = \alpha_\psi$, so suppose that there exists a $0 \neq c \in \mathbb{R}$ such that

$$\alpha_\varphi = c \cdot \alpha_\psi.$$

The Riesz representation Theorem implies

$$\int_{\varphi(G)} \beta = c \cdot \int_{\psi(G)} \beta$$

for all 1-forms $\beta \in L^2(\Omega^1)$. If $\psi \neq \varphi$ then $\psi(G) \neq \varphi(G)$ as subgraphs of K_n and consequently there exists a 1-form $\hat{\beta}$ supported on an edge appearing in $\psi(G)$ that is not in $\varphi(G)$. This means

$$0 \neq c \cdot \int_{\psi(G)} \hat{\beta} = \int_{\varphi(G)} \hat{\beta} = 0$$

- a contradiction. We conclude that $\varphi = \psi$ and $c = 1$ by Proposition 2.4.

Lemma 2 *Suppose G and H are graphs and consider their embedding sets $\mathcal{E}(G, K_n)$ and $\mathcal{E}(H, K_n)$, respectively. Then,*

$$C(\mathcal{E}(G, K_n)) = C(\mathcal{E}(H, K_n))$$

if and only if $G \cong H$; when $G \not\cong H$

$$C(\mathcal{E}(G, K_n)) \cap C(\mathcal{E}(H, K_n)) = \emptyset.$$

Proof Clearly, if $G \cong H$ then the sets $\mathcal{E}(G, K_n)$ and $\mathcal{E}(H, K_n)$, as they're defined, are equal as sets and hence their images under the map C^1 are equal as sets as well.

Now suppose $G \not\cong H$ but suppose, further, that there exists a $\varphi \in \mathcal{E}(G, K_n)$ and a $\psi \in \mathcal{E}(H, K_n)$ such that

$$\int_{\varphi(G)} = \int_{\psi(G)}.$$

By proposition 2.4 this means $\varphi(G) = \psi(H)$ as subgraphs of K_n and, in particular,

$$\psi^{-1} \circ \varphi(G) = H$$

is an isomorphism between G and H - a contradiction.

Lemma 3 *Suppose G and H are graphs and consider their embedding sets $\mathcal{E}(G, K_n)$ and $\mathcal{E}(H, K_n)$. Then*

$$\alpha_{\varphi(G)} = c \cdot \alpha_{\psi(H)}$$

for some $0 \neq c \in \mathbb{R}$ if and only if $\varphi(G) = \psi(H)$, in which case $c = 1$ and $G \cong H$. In particular, no two distinct 1-forms $\alpha_{\varphi(G)}$ and $\alpha_{\psi(H)}$ in $\mathcal{R} \circ C(\mathcal{E}(G, K_n))$ and $\mathcal{R} \circ C(\mathcal{E}(H, K_n))$, respectively, are collinear with the origin.

Proof Immediate consequence of Propositions 3.4 and 3.3

From these three Lemmas we obtain the Structure Theorem:

Theorem 3 *The map*

$$F : \mathcal{E}_n^+ \rightarrow \mathbb{S}_{L^2(\Omega_{K_n}^1)}$$

defined by

$$\varphi_G \mapsto \frac{1}{\|\alpha_{\varphi(G)}\|_{L^2(\Omega_{K_n}^1)}} \cdot \alpha_{\varphi(G)},$$

is an embedding of \mathcal{E}_n^+ into the L^2 unit sphere $\mathbb{S}_{L^2(\Omega_{K_n}^1)}$. The L^2 norm of $\alpha_{\varphi(G)}$ is exactly the square root of the number of edges in G : $\|\alpha_{\varphi(G)}\|_{L^2(\Omega_{K_n}^1)} = \sqrt{|E_G|}$. Moreover, the image of F lies in the positive orthant of $\mathbb{S}_{L^2(\Omega_{K_n}^1)}$.

Proof By Lemma 4.3, we know that no two 1-forms representing two distinct embeddings φ and $\psi \in \mathcal{E}$ are collinear with the origin $0 \in L^2(\Omega_{K_n}^1)$. Consequently, the map given in the statement of the Theorem is injective between the indicated spaces. The explicit description of a representing 1-form $\alpha_{\varphi(G)}$ given in Theorem 3.2 implies its L^2 norm is exactly the square root of the number of edges in G and the definition of the L^2 inner product with respect to the orientation \tilde{E} gives

$$\langle F(\varphi_G), F(\psi_H) \rangle \geq 0$$

so that the image of F is contained in the positive orthant of $\mathbb{S}_{L^2(\Omega_{K_n}^1)}$.

Remark 1 It is possible to extend this result to graphs with different numbers of vertices but it should be understood in a particular way. Let G be a graph on m vertices with $m \leq n$, and at least one edge, and consider the set of embeddings $\mathcal{E}(G, K_n)$ of G into K_n . Set

$$\mathcal{E}_+^m = \bigcup_{G \in \mathcal{G}_+^m} \mathcal{E}(G, K_n).$$

and

$$\mathcal{E}_n = \bigcup_{m=1}^n \mathcal{E}_+^m$$

- this is the set of all possible embeddings into K_n of all possible graphs on at most n vertices with at least one edge. Unfortunately, F is no longer injective on this larger set because F cannot see isolated vertices without any edges and the graphs on i vertices, $2 \leq i \leq n$, with exactly one edge all have the same image under F . So graphs on different numbers of vertices can be compared in the same hypersphere $\mathbb{S}_{L^2(\Omega_{K_n}^1)}$ but it should be understood that the graphs have, implicitly, a certain number of isolated vertices bringing the total order of each graph to n .

4 A Metric on Graphs

The Structure Theorem can be used to write down two distance metrics on graphs. One for isomorphism classes of graphs, and one for graphs where vertex labels matter.

4.1 A Metric on Isomorphism Classes of Graphs

The Structure Theorem gives an embedding of graphs into the L^2 unit sphere of 1-forms on K_n . This does not give us a unique representation of an isomorphism class of graphs (unlabelled graphs); rather, it gives us a unique representation of each representative of an isomorphism class, viewed as a subgraph of K_n . We can use the Structure Theorem to calculate quantitative differences between graphs with the spherical arc length distance - this is the arcsine of the L^2 inner product between the two 1-forms representing the two graphs of interest. Intuitively, the more edges two subgraphs have in common, the larger the L^2 inner product between the 1-forms representing them will be, and, consequently, the shorter the length of the spherical arc joining them will be. The fewer edges two subgraphs have in common, the smaller the L^2 inner product will be and the larger the length of the spherical arc joining them will be. The length of the shortest arc amongst all arcs joining elements of $F(\mathcal{E}(G, K_m))$ with elements of $F(\mathcal{E}(H, K_m))$ is, then, a distance between the isomorphism class of G and the isomorphism class of H , even if G and H have different numbers of vertices.

Let $\text{Iso}(n)$ denote the set of isomorphism classes of graphs on n vertices. So far we have only considered graphs with at least one edge; this is the set \mathcal{E}_+^n . The degenerate graph E_0^n on n vertices is a member of $\text{Iso}(n)$ but there is no reason to assume that the degenerate graph is closer to one graph than any other, and so we will take this graph to be at a constant distance from all other graphs - this is a canonical choice.

Definition 1 Define the following real-valued function on $\text{Iso}(n) \times \text{Iso}(n)$:

$$d_n([G], [H]) = \begin{cases} \frac{\pi}{2} & \text{if } G = E_0^n \vee H = E_0^n. \\ \min_{\substack{\varphi \in \mathcal{E}(G, K_m) \\ \psi \in \mathcal{E}(H, K_m)}} \arccos \left(\langle F(\varphi), F(\psi) \rangle_{L^2(\Omega_K^1)} \right) & \text{otherwise.} \end{cases}$$

The choice of $\frac{\pi}{2}$ is due to the fact that $\arccos 0 = \frac{\pi}{2}$ (0 formally representing the value of the inner product between the zero 1-form representing the degenerate graph and any other 1-form) and is consistent with the measurement units for non-degenerate graphs.

Theorem 4 *The set $\text{Iso}(n)$, equipped with the function d_n , is a discrete metric space $(\text{Iso}(n), d_n)$.*

Proof The function d_n is a distance metric:

1. Non-degeneracy follows directly from The Structure Theorem - Theorem 4.4;
2. Symmetry is obvious from symmetry of the L^2 inner product;
3. The L^2 unit sphere is, itself, a complete metric space with the spherical arc-length distance which satisfies the triangle inequality. The spherical arc-length function restricted to any (discrete) subset of the sphere continues to satisfy the triangle inequality and, consequently, d_n satisfies the triangle inequality as well. The distance between the degenerate graph and any other graph, as specified in the definition, ensures the triangle inequality is satisfied; this is because the image of F is contained in the positive orthant of $\mathbb{S}_{L^2(\Omega_K^1)}$.

By considering the graph of the cosine function on the interval $[0, \frac{\pi}{2}]$ it's straightforward to see that minimisation of arccos is equal to the maximisation of it's argument:

$$\min_{\substack{\varphi \in \mathcal{E}(G, K_m); \\ \psi \in \mathcal{E}(H, K_m)}} \arccos \left(\langle F(\varphi), F(\psi) \rangle_{L^2(\Omega_K^1)} \right) = \arccos \left(\max_{\substack{\varphi \in \mathcal{E}(G, K_m); \\ \psi \in \mathcal{E}(H, K_m)}} \langle F(\varphi), F(\psi) \rangle_{L^2(\Omega_K^1)} \right)$$

for two non-degenerate isomorphism classes of graphs $[G]$ and $[H]$. Using the particular expressions for the 1-forms $\alpha_{\varphi(G)}$ and $\alpha_{\psi(H)}$ given in Theorem 3.2, it becomes clear that maximising the L^2 -inner product between $F(\varphi(G))$ and $F(\psi(H))$ reduces to determining the largest number of edges two graphs, G and H , can have in common when "laid over the top of each other". From this perspective, the d_n metric can be expressed as a spherical Jaccard-type metric:

$$d_n([G], [H]) = \arccos \left(\max \frac{|E_G \cap E_H|}{\sqrt{|E_G|} \cdot \sqrt{|E_H|}} \right)$$

where E_G and E_H are the edge lists of G and H , and the maximisation is performed over edge relabellings. It's important to observe that one does not need to enumerate all embeddings of graphs in K_n to calculate this metric. Rather, one only needs to compare the edge lists of two given graphs and maximise the cardinality of their intersection over labellings.

4.2 A Metric on Labelled Graphs

Instead of performing graphs matching, one might be interested in comparing graphs that come with pre-assigned vertex labels. In this case there is only one embedding of each labelled graph: the labelled vertices of G are mapped to the corresponding vertices in K_n and the adjacency relations are preserved. A metric on labelled graphs, then, is simply the spherical arc length without the minimisation term. We state this using the set-theoretic expression derived above.

Let $L_G(n)$ denote the set of labelled graphs on n vertices.

Theorem 5 *The function $d_{L_G(n)}$ on $L_G(n) \times L_G(n)$, defined by*

$$d_{L_G(n)}(G, H) = \arccos \left(\frac{|E_G \cap E_H|}{\sqrt{|E_G|} \cdot \sqrt{|E_H|}} \right)$$

where E_G and E_H are the edge lists of the labelled graphs G and H , is a distance metric on $L_G(n)$.

4.3 Computing the Metric and the Image of F

In this paragraph we describe the image of F in the hypersphere as shown in Figure 1 for $n = 3$ and for a three dimensional slice of $F(\mathcal{E}_4^+)$. The edges of K_n are indexed by

the ordered set $E_{K_n} = \{(1, 2), (1, 3), \dots, (2, 3), (2, 4), \dots, (n, n+1), \dots, (\frac{n(n-1)}{2} - 1, \frac{n(n-1)}{2})\}$ and since a 1-form is completely determined by the values it takes on these edges, the vector space of 1-forms is $\frac{n(n-1)}{2}$ dimensional and can be visualised as $\mathbb{R}^{\frac{n(n-1)}{2}}$, where each 1-form is represented by a vector written with respect to the standard basis. The unit sphere in $L^2(\Omega_{K_n}^1)$ is then the standard $\frac{(n+1)(n-2)}{2}$ -sphere. By Proposition 2.2, there is only one representative of K_n in $\mathbb{S}_{L^2(\Omega_{K_n}^1)}$ which has coordinates $p_{K_n} = \left(\sqrt{\frac{2}{n(n-1)}}, \sqrt{\frac{2}{n(n-1)}}, \dots, \sqrt{\frac{2}{n(n-1)}}\right)$. If G is a graph with k edges then it comes with an edge list of length k , describing the graph adjacency. Augment this list with zeros so that it has the same length as E_{K_n} , and place the zeros in such a way so that the resulting list is ordered in the same way as E_{K_n} and a zero appears in the position of an edge *not* contained in G . Now replace each non-zero entry in the list with $\frac{1}{\sqrt{k}}$. The result is the list of components of the point p_G in the sphere $\mathbb{S}_{L^2(\Omega_{K_n}^1)}$ representing G with its given presentation $\varphi(G) < K_n$. The spherical arc-length distance between $p_{\varphi(G)}$ and p_{K_n} is

$$\arccos\left(\sqrt{\frac{2k}{n(n-1)}}\right).$$

This expression only depends on k and n , and is true for any graph with k edges. In particular, all points representing all graphs on k edges are at a constant distance from p_{K_n} and form a polygon centered at p_{K_n} . So, the image of $F(\mathcal{E}_n) \subset \mathbb{S}_{L^2(\Omega_{K_n}^1)}$ is a kaleidoscopic set of concentric polygons, all centered at p_{K_n} , and ordered by edge number.

Subgraphs of K_n with only one edge are mapped to the coordinate axes; subgraphs with two edges are mapped to the midpoints of the great arcs connecting points representing graphs with one edge (where those edges each appear in the subgraph with two edges); etc... See Figure 1 for $n = 3$ and Figure 2 for a three dimensional slice of $F(\mathcal{E}_4^+)$. Since graphs with exactly one edge are mapped by F to the coordinate axes, it's clear the closure of the convex hull of F (on the unit sphere) is the positive orthant $\mathbb{S}_{L^2(\Omega_{K_n}^1)}^+$.

Although we have connected our "geometry of graphs" with similarity measures of sets, this formulation of the metric suffers from the same drawback that affects the standard representation of unlabelled graphs: optimisation over vertex/edge re-labellings. However, the construction performed and analyzed over sections 3 and 4 represents graphs by points in the unit sphere of 1-forms over the ambient embedding graph K_m . The discrete image of the embedding F , described in the Structure Theorem, has a particular geometric form, which is interesting in and of itself, and, once fully understood, could be used to perform constrained gradient ascent/descent in the continuous unit sphere for the computation of the d_m metric.

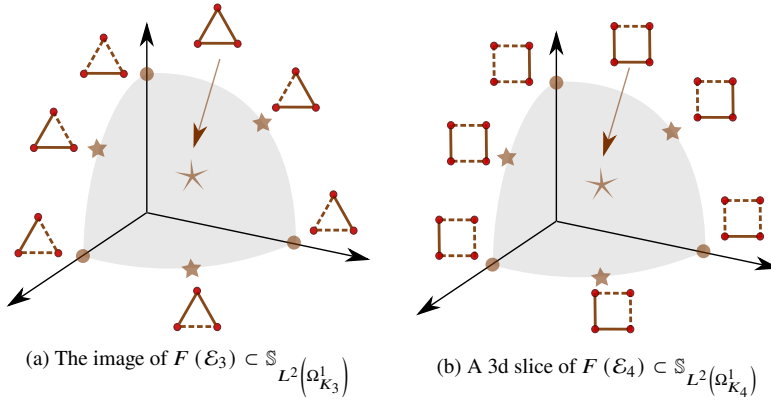


Fig. 1: Visualisations of the image of F for graphs on 3 and 4 vertices, respectively

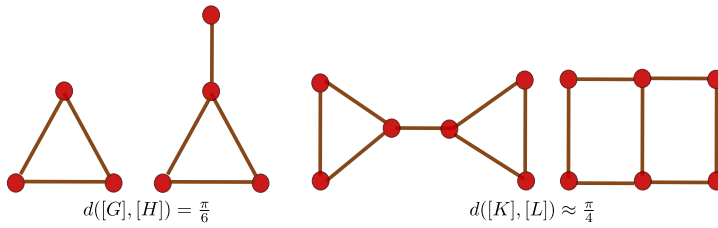


Fig. 2: Left: Two different, but topologically similar, graphs, which are "close" in the d_n metric. Right: Two topologically distinct graphs which are well separated in the d_m metric, but fail both the 1- and 2-WL tests

4.4 Metric Features

A good metric for the topology of graphs should be robust to minor additions or deletions of edges, yet sensitive to the presence of truly different topological structures of different dimensions. For example, the two graphs G and H , shown on the left in Figure 2, differ by an isthmus but basically have the same topological structure (see Remark 4.5 on how to interpret graphs with differing vertex orders). The d_n distance between them is

$$d_n([G], [H]) = \frac{\pi}{6}.$$

Compare this with the two graphs K and L shown on the right in Figure 2. These two graphs contain fundamentally different topological structures and the d_n distance between them is

$$d_n([K], [L]) = 0.775 \approx \frac{\pi}{4}.$$

While this distance doesn't appear "huge", it is worth remembering that the image of F spans the positive orthant of $\mathbb{S}_{L^2}(\Omega_k^1)$ and the d_m diameter of Iso_m is $\frac{\pi}{2}$ - the distance between $[K]$ and $[L]$ is the length of an arc spanning almost half the closure

of the convex hull of $F(\mathcal{E})$. It's interesting to note that both the 1 and 2-Weisfeiler-Lehman tests fail for the two graphs $[K]$ and $[L]$ [13], whereas the duality generated by treating the integral as a function of its domain, in combination with the Riesz representation Theorem, captures high order topological structures with 1-forms over K_n .

5 Further Modifications

It's also possible to bias certain edges in K_n so that these edges receive a greater bias than others when they appear in a subgraph $\varphi(G)$. We refer to b_{ij} as a bias of edge (i, j) so as not to confuse it with the terminology of edge attributes or weights which often represent features on the graph. This can be captured in our construction through a modification of the L^2 metric:

$$\langle \alpha, \beta \rangle_E^b = \sum_{\bar{E}} b_{ij} \alpha(i, j) \cdot \beta(i, j), \quad 0 < b_{ij} \in \mathbb{R}$$

where the strictly positive real numbers b_{ij} are fixed weights assigned to edges and can be thought of as defining a line element for integration. In this case, the unique 1-form $\alpha_{\varphi(G)} \in L_b^2(\Omega_{K_n}^1)$, where subscript b indicates measurement in the biased L^2 metric, is given by

$$\alpha_{\varphi(G)}(i, j) = \begin{cases} b_{ij} & \text{if } (i, j) \subset E_{\varphi(G)} \text{ and } i < j. \\ -b_{ij} & \text{if } (i, j) \subset E_{\varphi(G)} \text{ and } i > j. \\ 0 & \text{otherwise.} \end{cases}$$

The structure theorem can be proved just as before and we obtain an injective map

$$F : \mathcal{E}_n^+ \rightarrow \mathbb{S}_{L_b^2(\Omega_{K_n}^1)}$$

defined by

$$\varphi_G \mapsto \frac{1}{\|\alpha_{\varphi(G)}\|_{L_b^2(\Omega_K^1)}} \cdot \alpha_{\varphi(G)},$$

where the L_b^2 norm is the square root of the sum of the squared edge bias b_{ij} on edges appearing in $\varphi(G)$. A metric on isomorphism classes of graphs and labelled graphs can be defined as before but no longer has a simple Jaccard expression due to the incorporation of fixed edge weights on K_n :

$$d_n^b([G], [H]) = \arccos \left(\max \frac{\sum_{b_{ij} \in E_G \cap E_H} b_{ij}^2}{\sqrt{\sum_{b_{ij} \in E_G} b_{ij}^2} \cdot \sqrt{\sum_{b_{ij} \in E_H} b_{ij}^2}} \right)$$

where the maximisation is again performed over vertex-edge relabellings. The $\langle \cdot, \cdot \rangle_E^b$ -unit sphere appears as an ellipsoid in $\mathbb{R}^{\frac{n(n-1)}{2}}$ where the principle axes are determined by $\frac{1}{b_{ij}}$. Although the metric

6 Applications

Defining a common embedding space for a set of graphs is the first step in any statistical analysis of graphs. Analysing sets of graphs is crucial in different applications such as social networks of different users, brain connectivity network of different patients, protein-protein interactions, public transport systems, and stocks market correlation. In all these phenomena, sets of graphs or networks arise naturally and data analysis tools are required. The Currents Space of graphs addresses the problem of defining a more suitable and natural embedding for graphs. The majority of data analysis methods developed are based on graph matrix representations which are suited for graph classification or regression. However, if the final aim is to interpolate between graphs or develop dimensionality reduction methods an embedding based on the graph topology is preferable and the current space of graphs defines a suitable and natural embedding of graphs in which to analyze topology and topological differences. An implementation of the Currents Space embedding is based on the edge-list graph representation - which is a natural format in which graph data is presented - and overcomes the matrix representation and the zero padding problem which is necessary when graphs of different dimensions are compared.

To conduct a data analysis on the Currents Space of graphs, an efficient graph matching procedure should be defined by exploiting the geometrical structure of the space. As shown in Figure 1, graphs with the same topology represent vertices of polygons centered on the unique representative of K_n . Such geometrical structures could help find the optimal overlapping embedding of two graphs in K_n to compute the distance.

The spherical embedding in the Currents Space allows one to use the technique of principal symmetric space analysis [7, 3] for graphs. Such a method can be used for dimensionality reduction which involves finding a sub-sphere corresponding to a K_k which captures the majority of the variability in a set of graphs originally embedded in K_n . Other dimensionality reduction methods have been proposed, but all based on adjacency or laplacian matrix representations.

The Currents Space is also a suitable embedding space to study probability distributions over graphs as they generate binary graphs. A population of graphs prospective on graph distributions has been largely overlooked as there have been no suitable embedding. Computing the Fréchet mean, the Fréchet variance, and asymptotic behaviour of such distributions in the Currents Space can be of interest to capture novel geometrical insights.

7 Conclusion

We have constructed a representation of graphs by first embedding them in a complete graph and representing the resulting subgraph by a unique 1-form on K_n . The set of

all 1-forms representing all possible subgraphs turned out to have a fairly tractable geometry; that is, they could be arranged to lie on a hypersphere. This induced a natural distance metric on isomorphism classes of graphs and labelled graphs which reduced to a global Jaccard-type metric that can be computed from edge-list data alone. The global structure of the discrete image set of the map F has an interesting geometric structure in and of itself and we believe that a better understanding of this image will be of interest to the statistical analysis of graphs and could help improve computational efficiency of the metric.

Funding statement J.B., A.C., and X.P. are supported by the European Research Council (ERC) under the EU Horizon 2020 research and innovation program (grant agreement G- Statistics No. 786854). S.M. is supported by Te Pūnaha Matatini, a New Zealand Centre of Research Excellence in Complex Systems.

Competing interests None.

Author contributions J.B. and A.C. designed the study. J.B. wrote the first draft. A.C., S.M., and X.P. revised the manuscript and approved the final version.

References

1. J. Benn and S. Marsland. The measurement and analysis of shapes: An application of hydrodynamics and probability theory. *Annals of Global Analysis and Geometry*, 62:47–70, 2022.
2. A. Calissano, A. Feragen, and S. Vantini. Graph-valued regression: Prediction of unlabelled networks in a non-euclidean graph space. *Journal of Multivariate Analysis*, 190:104950, 2022.
3. C. Curry, S. Marsland, and R. I. McLachlan. Principal symmetric space analysis. *Journal of Computational Dynamics*, 6(2):251, 2019.
4. G. de Rham. *Differentiable Manifolds*, volume 266. Grundlehren der mathematischen Wissenschaften, 1984.
5. D. S. Goldberg and F. P. Roth. Assessing experimentally derived interactions in a small world. *Proceedings of the National Academy of Sciences of the United States of America*, pages 4372–4376, 2003.
6. J. Harrison. Geometric hodge star operator with applications to the theorems of gauss and green. *Mathematical Proceedings of the Cambridge Philosophical Society*, 2006.
7. S. Jung, I. L. Dryden, and J. S. Marron. Analysis of principal nested spheres. *Biometrika*, 99(3):551–568, 2012.
8. D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, pages 1019–1031, 2007.
9. G. Melanon and A. Sallaberry. Edge metrics for visual graph analytics: A comparative study. *Information Visualisation, 2008. IV'08. 12th International Conference*, pages 610–615, 2008.

10. G. Salton. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, 1989.
11. J. Scripps and C. Trefftz. Parallelizing an algorithm to find communities using the jaccard metric. *2015 IEEE International Conference on Electro/Information Technology (EIT)*, pages 370–372, 2015.
12. F. Y. Sun, Q. Meng, J. Hoffman, C. W. Huang, and J. Tang. vgraph: A generative model for joint community detection and node representation modelling. *Advances in Neural Information Processing Systems*, 32, 2019.
13. B. Weisfeiler and A. Leman. The reduction of a graph to canonical form and the algebra which appears therein. *NTI, Series*, 2(9):12–16, 1968.
14. H. Whitney. *Geometric Integration Theory*. Princeton University Press, 1957.