Weisfeiler and Lehman Go Topological: Message Passing Simplicial Networks

Cristian Bodnar, Fabrizio Frasca, Yu Guang Wang , Nina Otter ,Guido Montúfar, Pietro Liò, Michael M. Bronstein

Presented by: Jerry Wei

UW Stats

Table of Contents

Table of Contents

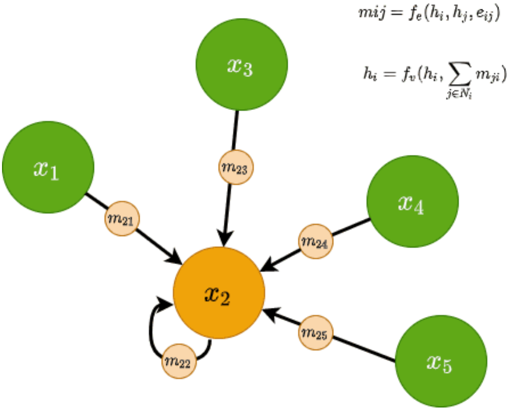
Graph Neural Network (GNN)

A graph $G = (V, E, \omega)$ is a set of triplets with vertices $V = \{v_i\}_{i=1}^{S_0}$, edges $E \subseteq V \times V$, and edge weight function $\omega : E \to \mathbb{R}$. The graph has an adjacency matrix A with the (i, j) th entry $a_{ij} = \omega(v_i, v_j)$. Each node has a d-dimensional feature, and we collect the feature vectors into a matrix $H^{\text{in}} \in \mathbb{R}^{S_0 \times d}$. A common GNN convolutional layer has the form

$$H^{\text{out}} = \psi\left(\mathcal{H}\left(A, H^{\text{in}}\right) W_0\right),$$

where ψ is the activation function, $\mathcal{H}(A, H)$ is an aggregation mapping, and $W_0 \in \mathbb{R}^{d \times m}$ are the trainable weights.

Graph Neural Networks (GNNs)



Expressive Power of GNN

Graph Isomorphism Test:

- Two graphs are considered isomorphic if there is a mapping between the nodes of the graphs that preserves node adjacencies.
- Expressive power is whether can detect isomorphic graphs.

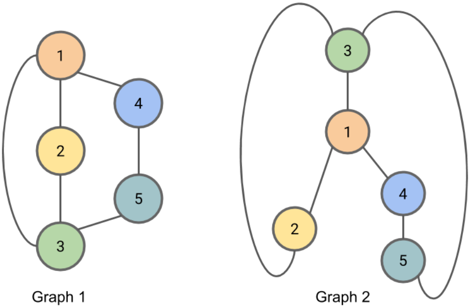


Figure: Image adopted from `https://davidbieber.com/post/2019-05-10-weisfeiler-lehman-isomorphism-test/`

Table of Contents

The Weisfeiler-Lehman (WL) Isomorphism Test

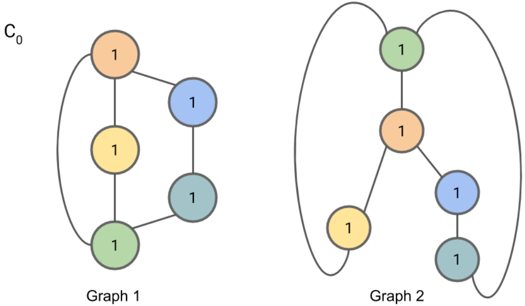
Precisely how hard the graph isomorphism problem is remains an open question in computer science.

The 1-WL Algorithm:

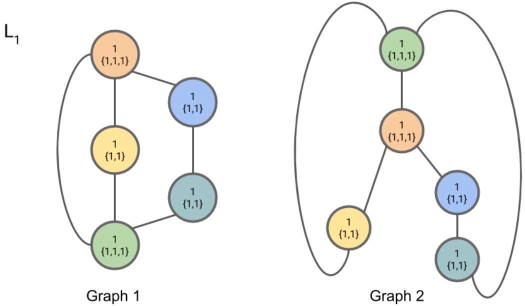
- For iteration i, we assign to each node a tuple $L_{i,n}$ containing the nodes old compressed label $C_{i-1,n}$ and a multiset of the nodes neighbors' previous labels.
- Hash $L_{i,n}$ to a new label $C_{i,n}$. Any two nodes with the same $L_{i,n}$ will get the same compressed label.
- Partition the nodes in the graph assigned label and repeat above for N (the number of nodes) iterations, or until there is no change in the partition of nodes.

It produces each graph a canonical form. If the canonical forms of two graphs are not equivalent, then the graphs are definitively not isomorphic. However, it is possible for two non-isomorphic graphs to share a canonical form.

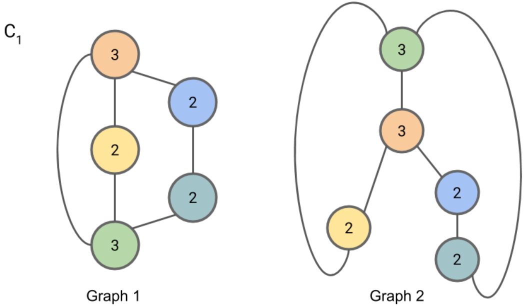
The Weisfeiler-Lehman (WL) Isomorphism Test



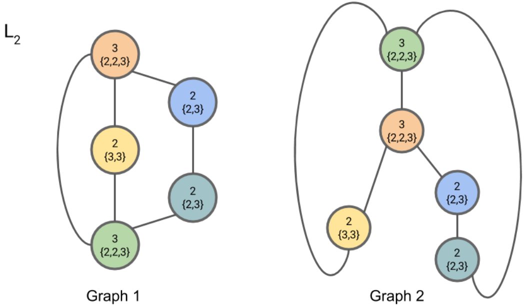
The Weisfeiler-Lehman (WL) Isomorphism Test



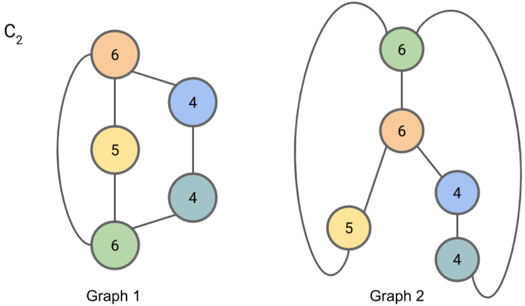
The Weisfeiler-Lehman (WL) Isomorphism Test



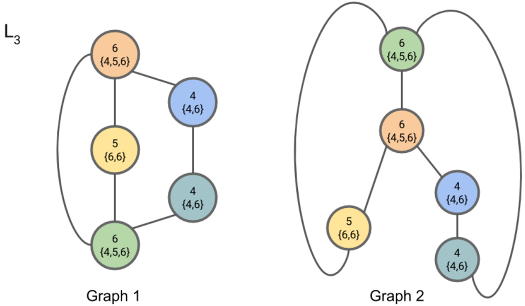
The Weisfeiler-Lehman (WL) Isomorphism Test



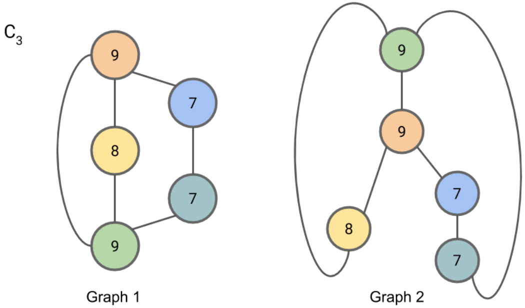
The Weisfeiler-Lehman (WL) Isomorphism Test



The Weisfeiler-Lehman (WL) Isomorphism Test



The Weisfeiler-Lehman (WL) Isomorphism Test



The canonical form of the 2 7s, 1 8, and 2 9s.

WL Test and GNN

GNN with injective aggregating function is equivalent in its expressive power to the WL test (Xue et al., 2019).
However, limited in their capability of detecting graph structures such as triangles or cliques (Chen et al., 2020).

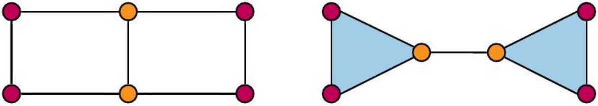


Figure: Two graphs that cannot be distinguished by 1-WL, but have distinct clique complexes (the second contains triangles).

Introduce simplicial complex for better expressive power.

Table of Contents

Background on Simplicial Complexes

Definition (simplicial complex)

Let V be a non-empty vertex set. A **simplicial complex** $\mathcal{K}$ is a collection of nonempty subsets of V that contains all the singleton subsets of V and is closed under the operation of taking subsets.

A member $\sigma = \{v_0, \ldots, v_k\} \in \mathcal{K}$ with cardinality $k+1$ is called a k-dimensional simplex or simply a k-simplex.
E.g. 0-simplices as vertices, 1-simplices as edges, 2-simplices as triangles, and so on.

Definition (Boundary incidence relation)

We say $\sigma \prec \tau$ iff $\sigma \subset \tau$ and there is no δ such that $\sigma \subset \delta \subset \tau$.

Definition (Orientation)

An oriented k-simplex is a k-simplex with a chosen order for its vertices. An oriented k-simplex is positively oriented if its vertices form an even permutation and negatively oriented otherwise.

Hodge Laplacian

The oriented boundary relations $\prec_+, \prec\_$ can be encoded by the signed boundary matrices (or incidence matrix) $B_k \in \mathbb{R}^{S_{k-1} \times S_k}$ has entries

$$B_k(i,j) = \begin{cases} 1, & \text{if } \tau_i \prec_+ \sigma_j \\ -1, & \text{if } \tau_i \prec -\sigma_j \\ 0, & \text{otherwise} \end{cases}$$

where $\dim(\sigma_j) = k, \dim(\tau_i) = k-1$, and S_k denotes the number of simplices of dimension k.

Definition (Hodge Laplacian)

The k-th Hodge Laplacian of the simplicial complex, a diffusion operator for signals over the oriented k-simplices, is defined as

$$L_k = B_k^\top B_k + B_{k+1} B_{k+1}^\top$$

Note that L_0 gives the well-known graph Laplacian.

Table of Contents

Simplicial WL (SWL) Test

Steps of SWL in the most general sense:

1. Given a complex $\mathcal{K}$, all the simplices $\sigma \in \mathcal{K}$ are initialized with the same colour.
2. Given the color c_σ^t of simplex σ at iteration t, we compute the color of simplex σ at the next iteration c_σ^{t+1}, by perfectly hashing the multi-sets of colors belonging to the **adjacent** simplices of σ.
3. The algorithm stops once a stable coloring is reached. Two simplicial complexes are considered nonisomorphic if the color histograms at any level of the complex are different.

Key: what simplices are considered to be adjacent

Types of Adjacencies

Definition (Adjacent Simplices)

Consider a simplex $\sigma \in \mathcal{K}$. Four types of adjacent simplices can be defined:

1. Boundary adjacencies: $\mathcal{B}(\sigma) = \{\tau \mid \tau \prec \sigma\}$.
2. Co-boundary adjacencies: $\mathcal{C}(\sigma) = \{\tau \mid \sigma \prec \tau\}$.
3. Lower-adjacencies: $\mathcal{N}_{\downarrow}(\sigma) = \{\tau \mid \exists \delta, \delta \prec \tau \wedge \delta \prec \sigma\}$
4. Upper-adjacencies: $\mathcal{N}_{\uparrow}(\sigma) = \{\tau \mid \exists \delta, \tau \prec \delta \wedge \sigma \prec \delta\}$

The boundary simplices of an edge: its vertices
The co-boundary simplices of a vertex: the edges connected with
The lower-adjacent edges: share a vertex
The upper adjacencies between vertices: part of the same line

The Multisets in SWL

Definition

Let c^t be the coloring of SWL for the simplices in $\mathcal{K}$ at iteration t. We define the following multi-sets of colors, corresponding to each type of adjacency:

1. $c^t_{\mathcal{B}}(\sigma) = \{\{c^t_\tau \mid \tau \in \mathcal{B}(\sigma)\}\}$.
2. $c^t_{\mathcal{C}}(\sigma) = \{\{c^t_\tau \mid \tau \in \mathcal{C}(\sigma)\}\}$.
3. $c^t_{\downarrow}(\sigma) = \{\{(c^t_\tau, c^t_{\sigma\cap\tau}) \mid \tau \in \mathcal{N}_{\downarrow}(\sigma)\}\}$.
4. $c^t_{\uparrow}(\sigma) = \{\{(c^t_\tau, c^t_{\sigma\cup\tau}) \mid \tau \in \mathcal{N}_{\uparrow}(\sigma)\}\}$.

SWL Update Rule

The SWL update rule with the complete set of adjacencies:

$$c_\sigma^{t+1} = \text{HASH}\left\{c_\sigma^t, c_\mathcal{B}^t(\sigma), c_\mathcal{C}^t(\sigma), c_\downarrow^t(\sigma), c_\uparrow^t(\sigma)\right\}$$

Theorem

SWL with $\text{HASH}\left(c_\sigma^t, c_\mathcal{B}^t(\sigma), c_\uparrow^t(\sigma)\right)$ *is as powerful as SWL with the generalized update rule* $\text{HASH}\left(c_\sigma^t, c_\mathcal{B}^t(\sigma), c_\mathcal{C}^t(\sigma), c_\downarrow^t(\sigma), c_\uparrow^t(\sigma)\right)$.

Reduces to WL when applied to graphs and only considers vertex coloring.
"this result comes from a (theoretical) color-refinement perspective and it does not imply that the pruned adjacencies cannot be useful in practice."

Link the Expressive Power of WL and SWL

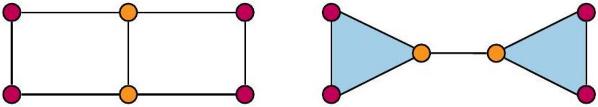
Definition

The **clique complex** of a graph G is the simplicial complex $\mathcal{K}$ with the property that if nodes $\{v_0, \ldots v_k\}$ form a clique in G, then simplex $\{v_0, \ldots v_k\} \in \mathcal{K}$. In other words, every $(k+1)$-clique in G becomes a k-simplex in $\mathcal{K}$. This is called a **lifting transformation**.

Theorem

SWL with a clique complex lifting is strictly more powerful than WL.

As illustrated by the previous example:



Link the Expressive Power of WL and SWL

Theorem

SWL is not less powerful than 3-WL.

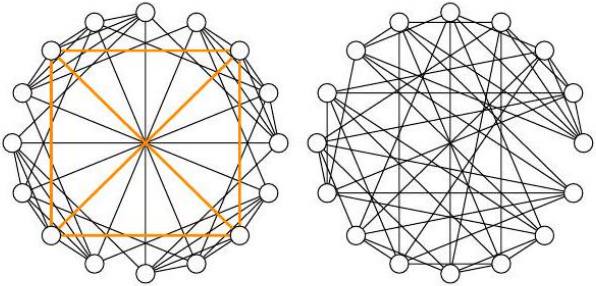


Figure: Rook's 4x4 graph and the Shrikhande graph: Strongly Regular non-isomorphic graphs with parameters SR(16, 6, 2, 2). SWL can distinguish them: only Rook's graph (left) possesses 4 -cliques (orange) and thus the two graphs are associated with distinct clique complexes. The 3-WL test fails to distinguish them.

Table of Contents

Message Passing Simplicial Networks (MPSN)

Message passing model using the following message passing operations based on the four types of multisets. For a simplex σ in a complex $\mathcal{K}$ we have:

$$
\begin{aligned}
m_{\mathcal{B}}^{t+1}(\sigma) &= \mathrm{AGG}_{\tau \in \mathcal{B}(\sigma)}\left(M_{\mathcal{B}}\left(h_\sigma^t, h_\tau^t\right)\right)\\
m_{\mathcal{C}}^{t+1}(\sigma) &= \mathrm{AGG}_{\tau \in \mathcal{C}(\sigma)}\left(M_{\mathcal{C}}\left(h_\sigma^t, h_\tau^t\right)\right)\\
m_{\downarrow}^{t+1}(\sigma) &= \mathrm{AGG}_{\tau \in \mathcal{N}_{\downarrow}(\sigma)}\left(M_{\downarrow}\left(h_\sigma^t, h_\tau^t, h_{\sigma\cap\tau}^t\right)\right)\\
m_{\uparrow}^{t+1}(\sigma) &= \mathrm{AGG}_{\tau \in \mathcal{N}_{\uparrow}(\sigma)}\left(M_{\uparrow}\left(h_\sigma^t, h_\tau^t, h_{\sigma\cup\tau}^t\right)\right).
\end{aligned}
$$

and the update operation takes into account these four types of incoming messages and the previous color of the simplex:

$$
h_\sigma^{t+1} = U\left(h_\sigma^t, m_{\mathcal{B}}^t(\sigma), m_{\mathcal{C}}^t(\sigma), m_{\downarrow}^{t+1}(\sigma), m_{\uparrow}^{t+1}(\sigma)\right).
$$

Expressive Power of MPSN

Theorem

MPSNs with sufficient layers and injective neighborhood aggregators are as powerful as SWL.

Corollary

There exists an MPSN that is more powerful than WL at distinguishing non-isomorphic graphs when using a clique-complex lifting.

In particular, based on previous Theorem, it is sufficient for such an MPSN to use boundary and upper adjacencies.

Relation to Spectral Convolutions

GNNs are known for their relationship with spectral convolution operators on graphs obtained via graph Laplacian (Hammond et al., 2011).
Analogously to this, MPSNs generalize certain spectral convolutions on graphs derived from the higher-order simplicial Hodge Laplacian

Theorem

Theorem 12. MPSNs generalize certain spectral convolution operators (Ebli et al., 2020; Bunch et al., 2020) defined over simplicial complexes.

Table of Contents

Expressive Power by Linear Regions

The number of linear regions using piece-wise linear activations has been used to study conventional neural networks.

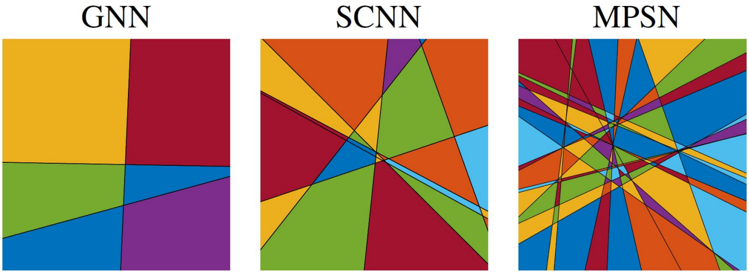


Figure: A 2D slice of the input feature spaces of GNN, SCNN, MPSN layers with $S_0 = S_1 = 3, S_2 = 1$ (the complex is a triangle), $d_0 = d_1 = d_2 = 1, m = 3$, colored by linear regions of the represented functions, for a random choice of the weights.

Graph Neural Network (GNN)

The graph has an adjacency matrix A. Each node has a d-dimensional feature, and we collect the feature vectors into a matrix $H^{\text{in}} \in \mathbb{R}^{S_0 \times d}$. We consider a GNN convolutional layer of the form

$$H^{\text{out}} = \psi\left(\mathcal{H}\left(A, H^{\text{in}}\right) W_0\right),$$

where ψ is the entry-wise ReLU, $\mathcal{H}(A, H)$ is an aggregation mapping, and $W_0 \in \mathbb{R}^{d \times m}$ are the trainable weights.

GNN Linear Regions

Theorem (Number of linear regions of a GNN layer)

Consider a graph G with S_0 nodes, node input features of dimension $d \geq 1$, and node output features of dimension m. Suppose the aggregation function $\mathcal{H}$ as function of H is linear and invertible. Then, the number of linear regions of the functions represented by a ReLU GNN layer has the optimal upper bound

$$R_{\text{GNN}} = \left(2\sum_{i=0}^{d-1}\binom{m-1}{i}\right)^{S_0}.$$

To compare: the optimal upper bound for a standard dense ReLU layer without biases with d inputs and m outputs is $2\sum_{i=0}^{d-1}\binom{m-1}{i}$

Simplicial Complex Neural Networks (SCNNs)

Consider a version of the model in Ebli et al. (2020) using only the first power of a Laplacian matrix, generically denoted here by M_n :

$$H_n^{\text{out}} = \psi\left(M_n H_n^{\text{in}} W_n\right), \quad n = 0, \ldots, p$$

where the features on simplices of different dimensions $n = 0, 1 \ldots, p$ are computed in parallel.

Definition (Number of linear regions for an SCNN layer)

Consider a p-dimensional simplicial complex with $S_n n$ simplicies for $n = 0, 1, \ldots, p$. Suppose that each M_n is invertible. Then the number of linear regions of the functions represented by a ReLU SCNN layer has the optimal upper bound

$$R_{\mathrm{SCNN}} = \prod_{n=0}^{p} \left(2 \sum_{i=0}^{d_n - 1} \binom{m_n - 1}{i}\right)^{S_n}$$

where, for each of the n-simplices, d_n is the input feature dimension and m_n is the number of output features.

MPSN Formulation

For a p-dimensional complex, denote the set of n-simplices by $\mathcal{S}_n$ with $S_n = |\mathcal{S}_n|$. Denote the n-simplex input feature dimension by d_n, and the output feature dimension by $m_n = m, n = 0, \ldots, p$.
We consider an MPSN layer with

- linear message functions
- sum aggregation for all messages
- an update function taking the sum of the messages followed by a ReLU activation

For each dimension n, the output feature matrix H_n^{out} equals:

$$\psi\left(M_n H_n^{\text{in}} W_n + U_n H_{n-1}^{\text{in}} W_{n-1} + O_n H_{n+1}^{\text{in}} W_{n+1}\right),$$

where ψ is an entry-wise activation, $W_n \in \mathbb{R}^{d_n \times m_n}$ are trainable weight matrices and $M_n \in \mathbb{R}^{S_n \times S_n}$, $U_n \in \mathbb{R}^{S_n \times S_{n-1}}$, and $O_n \in \mathbb{R}^{S_n \times S_{n+1}}$ are some choice of adjacency matrices for the simplicial complex.

MPSN in Standard Form

Roth's lemma (Roth, 1934) states

$$\operatorname{vec}\left(M_n H_n^{\text{in}} W_n\right) = \left(W_n^\top \otimes M_n\right) \cdot \operatorname{vec}\left(H_n^{\text{in}}\right)$$

where vec denotes column-by-column vectorization and $\otimes$ the Kronecker product. Using Roth's lemma and concatenating over n we can write as

$$H^{\text{out}} = \psi\left(W H^{\text{in}}\right),$$

$H^{\text{in}} = \operatorname{vec}\left(\left[H_0^{\text{in}} | H_1^{\text{in}} | \cdots | H_p^{\text{in}}\right]\right) \in \mathbb{R}^N$, $N = \sum_{n=0}^{p} S_n d_n$,
$H^{\text{out}} = \operatorname{vec}\left(\left[H_0^{\text{out}} | H_1^{\text{out}} | \cdots | H_p^{\text{out}}\right]\right) \in \mathbb{R}^M$, $M = \sum_{n=0}^{p} S_n m$, and

$$W = \begin{bmatrix} W_0^\top \otimes M_0 & W_1^\top \otimes O_0 & & \\ W_0^\top \otimes U_1 & W_1^\top \otimes M_1 & W_2^\top \otimes O_1 & \\ & W_1^\top \otimes U_2 & W_2^\top \otimes M_2 & W_3^\top \otimes O_2 \\ & & \ddots & \end{bmatrix} \in \mathbb{R}^{M \times N}$$

MPSN Linear Regions

Theorem (Number of linear regions of an MPSN layer)

With the above settings, the maximum number of linear regions of the functions represented by a ReLU MPSN layer is upper bounded by

$$R_{\mathrm{MPSN}} \leq \prod_{n=0}^{p} \left(2 \sum_{i=0}^{d_{n-1}+d_n+d_{n+1}-1} \binom{m-1}{i} \right)^{S_n},$$

where we set $d_{-1} = d_{p+1} = 0$*. We also note the 'trivial' upper bound, with* $N := \sum_{n=0}^{p} S_n d_n$ *and* $M := \sum_{n=0}^{p} S_n m$*,*

$$R_{\mathrm{MPSN}} \leq 2 \sum_{j=0}^{N-1} \binom{M-1}{j}.$$

Moreover, if $\operatorname{rank}\left((O_n)_{C:}\right) \geq \operatorname{rank}\left((M_n)_{C:}\right)$ *for any selection* C *of rows and* $d_{n+1} \geq d_n$*, for* $n = 0, \ldots, p-1$*, then for networks with outputs* $H_0^{out}, \ldots, H_{p-1}^{out}$ *we have*

$$R_{\mathrm{MPSN}} \geq R_{\mathrm{SCNN}}.$$

Table of Contents

Strong Regular Graphs

Strongly Regular (SR) graphs represent 'hard' instances of graph isomorphism, as pairs thereof cannot provably be distinguished by the 3-WL test

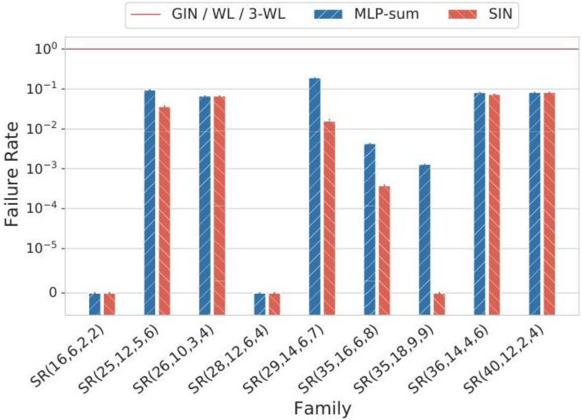


Figure: Failure rate on the task of distinguishing SR graphs; logscale, the smaller the better. GIN fails to distinguish all graph pairs in all families.

Edge-Flow Classification

Synthetic flow as shown in the figure below.

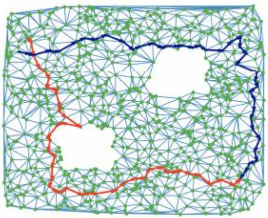


Figure: Two samples from the two different classes of trajectories. The two trajectories correspond to approximately orthogonal directions in the space of harmonic functions of L_1 associated with the two holes.

Ocean drifter data with trajectories around the island of Madagascar between years 2011-2018 with the task to distinguish between the clockwise and counter-clockwise flows around the island.

Edge-Flow Classification

Table 1. Trajectory classification accuracy. Models with triangle awareness and orientation equivariance generalise better.

Method	Synthetic Flow		Ocean Drifters	
	Train	Test	Train	Test
GNN L_0-inv	63.9±2.4	61.0±4.2	70.1±2.3	63.5±6.0
MPSN L_0-inv	88.2±5.1	85.3±5.8	84.6±4.0	**71.5±4.1**
MPSN - ReLU	**100.0±0.0**	50.0±0.0	**100.0±0.0**	46.5±5.7
MPSN - Id	88.0±3.1	82.6±3.0	94.6±0.9	**73.0±2.7**
MPSN - Tanh	97.9±0.7	**95.2±1.8**	**99.7±0.5**	**72.5±0.0**

Real-World Graph Classification

Table 2. Graph classification results on the TUDatasets benchmark. The table contains: dataset details *(top)*, graph kernel methods *(middle)*, and graph neural networks *(bottom)*.

Dataset	Proteins	NCI1	IMDB-B	IMDB-M	RDT-B	RDT-M5K
Avg △	27.4	0.05	392.0	305.9	24.8	21.8
Med △	21.0	0.0	119.5	56.0	11.0	11.0
RWK	59.6±0.1	>3 days	N/A	N/A	N/A	N/A
GK (k=3)	71.4±0.31	62.5±0.3	N/A	N/A	N/A	N/A
PK	73.7±0.7	82.5±0.5	N/A	N/A	N/A	N/A
WL kernel	75.0±3.1	86.0±1.8	73.8±3.9	50.9±3.8	81.0±3.1	52.5±2.1
DCNN	61.3±1.6	56.6±1.0	49.1±1.4	33.5±1.4	N/A	N/A
DGCNN	75.5±0.9	74.4±0.5	70.0±0.9	47.8±0.9	N/A	N/A
IGN	76.6±5.5	74.3±2.7	72.0±5.5	48.7±3.4	N/A	N/A
GIN	76.2±2.8	82.7±1.7	75.1±5.1	52.3±2.8	92.4±2.5	57.5±1.5
PPGNs	77.2±4.7	83.2±1.1	73.0±5.8	50.5±3.6	N/A	N/A
Natural GN	71.7±1.0	82.4±1.3	73.5±2.0	51.3±1.5	N/A	N/A
GSN	76.6 ± 5.0	83.5 ± 2.0	77.8 ± 3.3	54.3 ± 3.3	N/A	N/A
SIN (Ours)	76.5 ± 3.4	82.8 ± 2.2	75.6 ± 3.2	52.5 ± 3.0	92.2 ± 1.0	57.3 ± 1.6

Thanks for listening