2024.2.26

Weisfeiler and Lehman Go Cellular: CW Networks

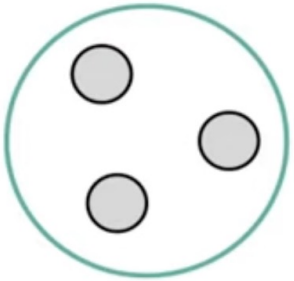
Liyan Tan

Electrical & Computer Engineering Department

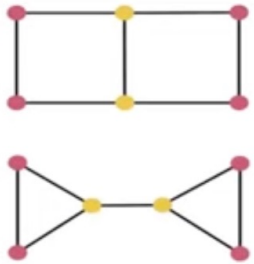
UC SANTA BARBARA

Motivation: Limitations of GNN

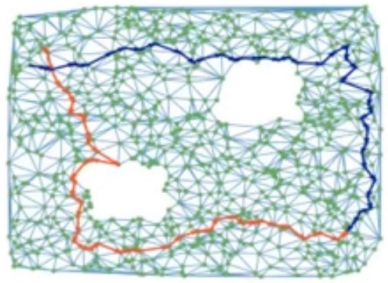
Groupwise interactions



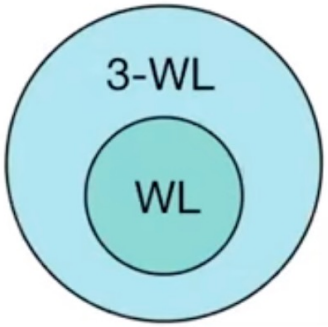
Higher-order structures



Higher-order signals



Expressive power

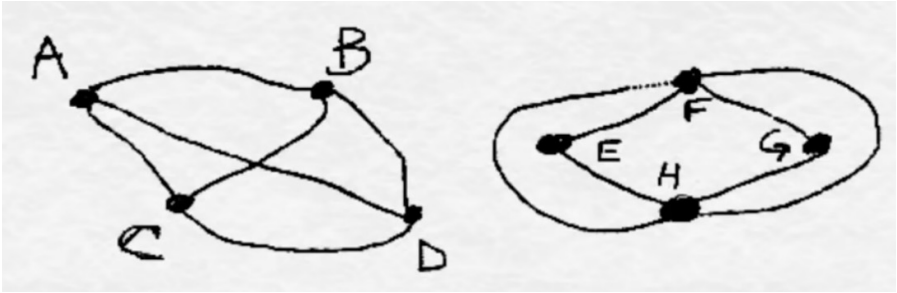


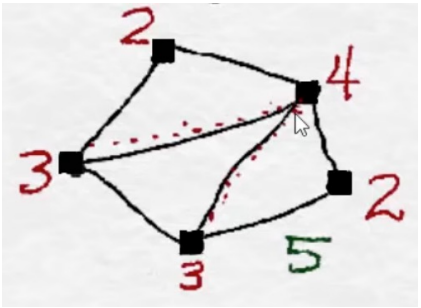
Long-range interactions

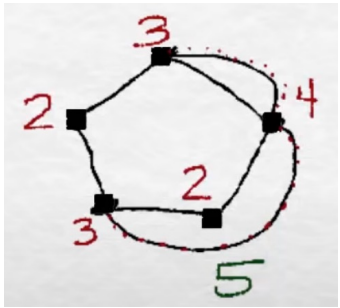
Isomorphic Graphs

Same number of:

- Vertices
- Degrees
- Shape

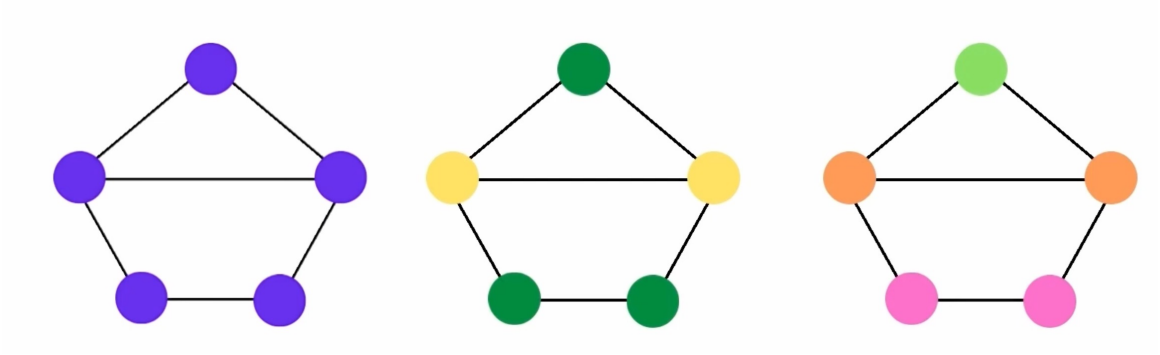


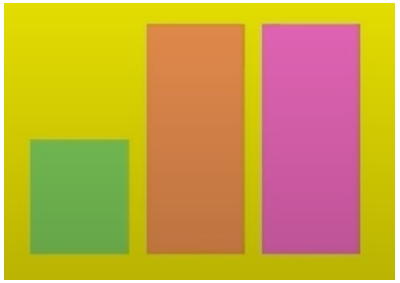




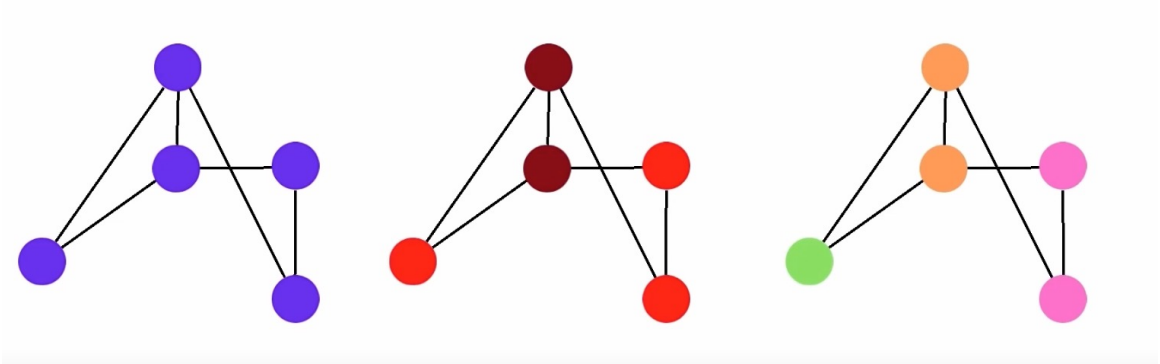
Weisfeiler-Lehman WL Test

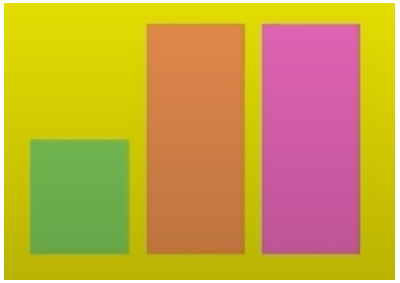
How to tell if two graphs are isomorphic or not? – By WL test





Bar chart is the same





What did they do?

The authors propose a new message passing technique that utilizes regular cell complexes, also known as CW complexes.

"Lifting" Transformation: The method involves a **"lifting"** transformation that augments the graph with higher-dimensional **"cells"**.
This transformation allows for a more **complex** and **hierarchical** message passing scheme over the input graph, enhancing the network's ability to handle complex structures.

Advantages over Simplicial Complexes: Traditional Message Passing Simplicial Networks (MPSNs) operate on simplicial complexes (SCs).
Disadvantages: rigid in their structure and limit the transformations that can be applied.
The proposed method uses cell complexes. Cell complexes generalize SCs and offer additional flexibility, allowing for new and improved ways to decouple input from computational graphs.

Main contributions

•They introduce a family of models named **CW Networks** (CWNs).

•They assess the **expressive power** of these models and show that they are at least as powerful as the WL test, a standard for measuring the ability of a GNN to distinguish between different graph structures.

•They provide proof that for **certain transformations**, CWNs are more powerful than both the WL test and the Simplicial WL (SWL) test, and not less powerful than the 3-WL test.

•They highlight the models' ability to respect the fundamental symmetries of cell complexes, functioning as generalized convolutional operators.

A regular cell complex

Local Finiteness: Around every point x in X, there is an open neighborhood that intersects only a finite number of cells.

Poset Structure: There is a partially ordered set (poset) structure defined by the cells, where for any two cells σ and τ, X_τ intersects X_σ nontrivially if and only if X_τ is contained in the closure of X_σ.

Homeomorphism to Euclidean Space: Each cell is homeomorphic to $\mathbb{R}^n$ for some integer n, meaning each cell can be continuously deformed to an n -dimensional Euclidean space.

Regularity: For each cell represented by σ in the poset P_X, there is a homeomorphism ϕ from a closed ball in $\mathbb{R}^n$ to the closure of the cell X_σ, and the restriction of ϕ to the interior of the ball is a homeomorphism onto the interior of X_σ.

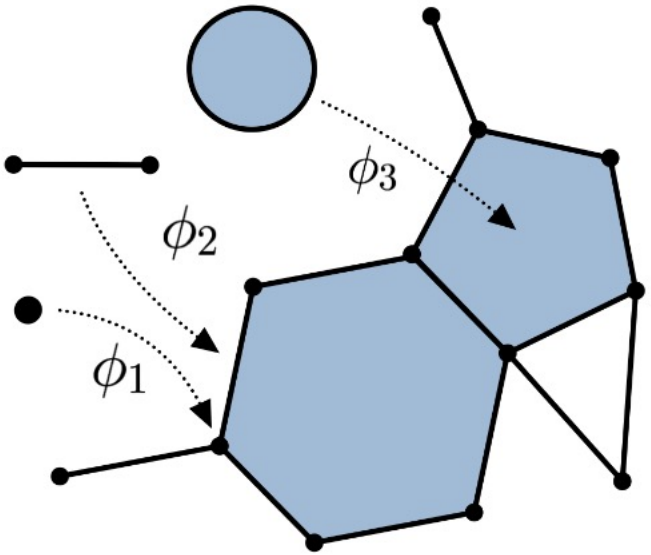


Figure 1: A cell complex X and the corresponding homeomorphisms to the closed balls for three cells of different dimensions in the complex.

vertices (0-cells)
edges (1-cells)
2-dimensional disks (2-cells)

Background

Definition 2. *The* **k-*skeleton*** *of a cell complex X, denoted $X^{(k)}$, is the subcomplex of X consisting of cells of dimension at most k.*

This concept is useful for discussing specific parts of a cell complex.

- For example: $X^{(0)}$ refers to the 0-skeleton, which includes just the vertices of the complex.

$X^{(1)}$ refers to the 1-skeleton, which includes both the vertices and the edges.

- The 1-skeleton is essentially the underlying graph if you were to ignore any higher-dimensional cells.

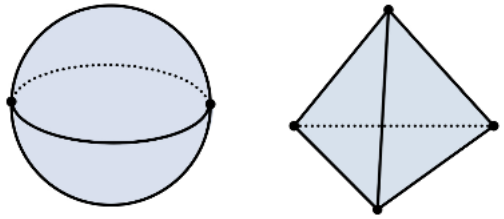


Figure 3: A sphere and an empty tetrahedron. The latter is also a simplicial complex.

Background

Definition 3. *We have the* ***boundary relation*** $\sigma \prec \tau$ *iff* $\sigma < \tau$ *and there is no cell* δ *such that* $\sigma < \delta < \tau$.

The boundary relation, denoted as $\sigma < \tau$, is defined such that a cell σ is considered to be a boundary of another cell τ if σ is of lower dimension than τ (i.e., $\sigma < \tau$), and there is no intermediate cell δ with $\sigma < \delta < \tau$. This means that σ is directly on the boundary of τ without any other cells in between them that are part of the boundary of τ.

This definition is used to characterize different types of adjacencies within cell complexes, which are crucial for understanding the topological structure of the complex.

Cellular Weisfeiler Lehman

The section discusses transforming graphs into higher-dimensional cell complexes, which makes it easier to test their isomorphism.

Definition 5. *Let c be a colouring of the cells in a complex X with c_σ denoting the colour assigned to cell $\sigma \in P_X$. Define $\mathcal{B}(\sigma,\tau) := \mathcal{B}(\sigma) \cap \mathcal{B}(\tau)$ and $\mathcal{C}(\sigma,\tau) := \mathcal{C}(\sigma) \cap \mathcal{C}(\tau)$. We define the following multi-sets of colours:*

1. *The colours of the boundary cells of σ: $c_{\mathcal{B}}(\sigma) = \{\!\{c_\tau \mid \tau \in \mathcal{B}(\sigma)\}\!\}$.*
2. *The colours of the co-boundary cells of σ: $c_{\mathcal{C}}(\sigma) = \{\!\{c_\tau \mid \tau \in \mathcal{C}(\sigma)\}\!\}$.*
3. *The lower adjacent colours of σ: $c_{\downarrow}(\sigma) = \{\!\{(c_\tau, c_\delta) \mid \tau \in \mathcal{N}_{\downarrow}(\sigma) \text{ and } \delta \in \mathcal{B}(\sigma,\tau)\}\!\}$.*
4. *The upper adjacent colours of σ: $c_{\uparrow}(\sigma) = \{\!\{(c_\tau, c_\delta) \mid \tau \in \mathcal{N}_{\uparrow}(\sigma) \text{ and } \delta \in \mathcal{C}(\sigma,\tau)\}\!\}$.*

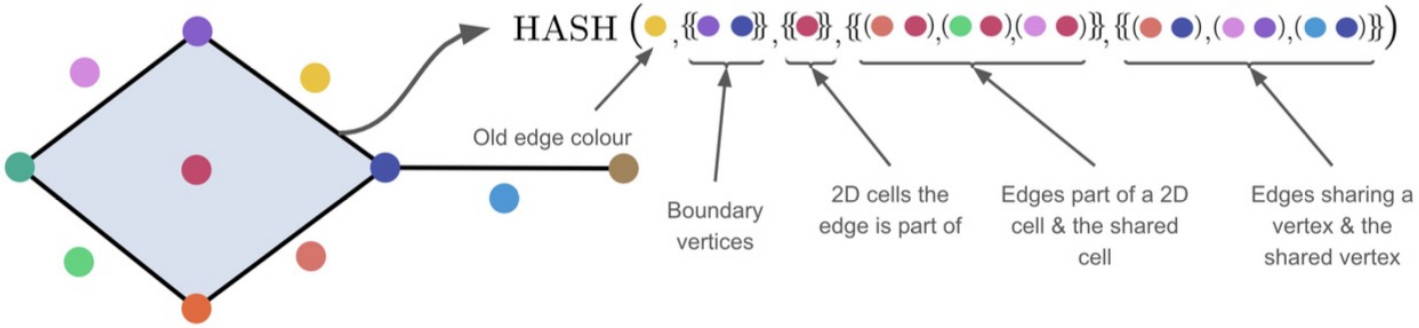


Figure 4: The CWL colouring procedure for the yellow edge of the cell complex. All cells have been assigned unique colours to aid the visualisation of the adjacencies. Note that the yellow edge aggregates long-range information from the light green edge.

Cellular Weisfeiler Lehman

This is a color refinement algorithm for cell complexes that extends the WL and SWL tests, which are used for graph isomorphism testing. The CWL works as follows:

- All cells in a cell complex are **initialized with the same color**.
- At each iteration, the color of a cell is updated based on a hash function that **combines the colors of its adjacent cells**.
- The algorithm stops when it reaches a stable coloring, where **no more color changes occur**. Two complexes are considered non-isomorphic if their color histograms differ after the algorithm concludes.

Theorem 6. *SWL without coboundary and lower-adjacencies has the same expressive power in distinguishing non-isomorphic simplicial complexes as SWL with the complete set of adjacencies.*

This suggests that the algorithm can effectively identify distinct simplicial complexes even when it ignores certain types of adjacencies

Theorem 7. *CWL without coboundary and lower-adjacencies has the same expressive power in distinguishing non-isomorphic cell complexes as CWL with the complete set of adjacencies.*

This demonstrates that even when simplifying the CWL by removing these two adjacency types, it can still effectively differentiate between non-isomorphic cell complexes.

Lifting map

Definition 8. *A **cellular lifting map** is a function $f : \mathcal{G} \to \mathcal{X}$ from the space of graphs $\mathcal{G}$ to the space of regular cell complexes $\mathcal{X}$ with the property that two graphs G_1, G_2 are isomorphic iff the cell complexes $f(G_1), f(G_2)$ are isomorphic.*

In other words, the function f is an isomorphism-preserving transformation from graphs to cell complexes. This ensures that testing for isomorphism between two **cell complexes** is **equivalent** to testing isomorphism between the **original graphs**.

Definition 11. *A lifting map is **skeleton-preserving** if for any graph G, the 1-skeleton of $f(G)$ and G are isomorphic as (multi) graphs.*

This means that the lifting map adds higher-dimensional cells (dimensions at least two) to the graph without altering the fundamental graph structure composed of 0-cells (vertices) and 1-cells (edges)

Molecular Message Passing with CW Networks

CW Networks apply a skeleton-preserving lifting transformation to molecular graphs. This process involves creating **higher-dimensional structures** (2-cells) for all induced cycles (which are chordless cycles) in the graph. These 2-cells are attached to the existing graph structure (the 1-skeleton, which includes edges and vertices) without altering it. Message passing then occurs across these structures, involving:

- Atoms (0-cells or vertices)
- Bonds between atoms (edges or 1-cells)
- Chemical rings (induced cycles or 2-cells)

The message passing process within a CW Network is divided into two types:
$m_B^{t+1}(\sigma)$: Messages from atoms to bonds and from bonds to rings. This message type captures the interactions from lower-dimensional cells (atoms) to their immediate higher-dimensional adjacencies (bonds and rings).
$m_\Upsilon^{t+1}(\sigma)$: Messages between atoms connected by a bond and messages between bonds that are part of the same ring. This message type captures interactions within the same level of dimensionality (e.g., within the same ring) and features of the bonds and rings involved in the communication.

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

Molecular Message Passing with CW Networks

Expressivity of CW Networks: CW Networks (CWNs) can distinguish non-isomorphic regular cell complexes with the **same level of expressiveness** as the Cellular Weisfeiler-Lehman (CWL) method, as long as they have a **sufficient number of layers** and use **injective local aggregators**. These multi-set aggregators are known to exist and can be used directly in the model.

Long-Range Interactions: CW Networks can capture long-range interactions between nodes effectively. They require fewer layers to achieve the same depth of message passing compared to traditional GNNs because the higher-dimensional cells (like 2-cells representing rings in molecules) **act as shortcuts** in the message passing scheme.

Computational Complexity: The computational complexity of the message passing scheme in CW Networks is **linear in the size of the input cell complex**, making it **comparable to GNNs**. The preprocessing step for lifting the graphs to cell complexes can be done efficiently, and the overall complexity of the lifting procedure is **almost linear in the size of the graph**.

Experiments

Synthetic Benchmarks

CSL (Circular Skip Link) Dataset:

- This dataset is used to test the **expressiveness** of GNNs.
- It consists of 150 4-regular graphs from 10 different isomorphism classes, where the goal is to predict the WL test and message passing approaches.
- CWNs, referred to as CIN (Ours), achieved 100% accuracy, validating their expressive power.
- Classic message-passing GNNs, like GAT and RingGNN, performed as random guessers, highlighting the superior performance of CWNs.

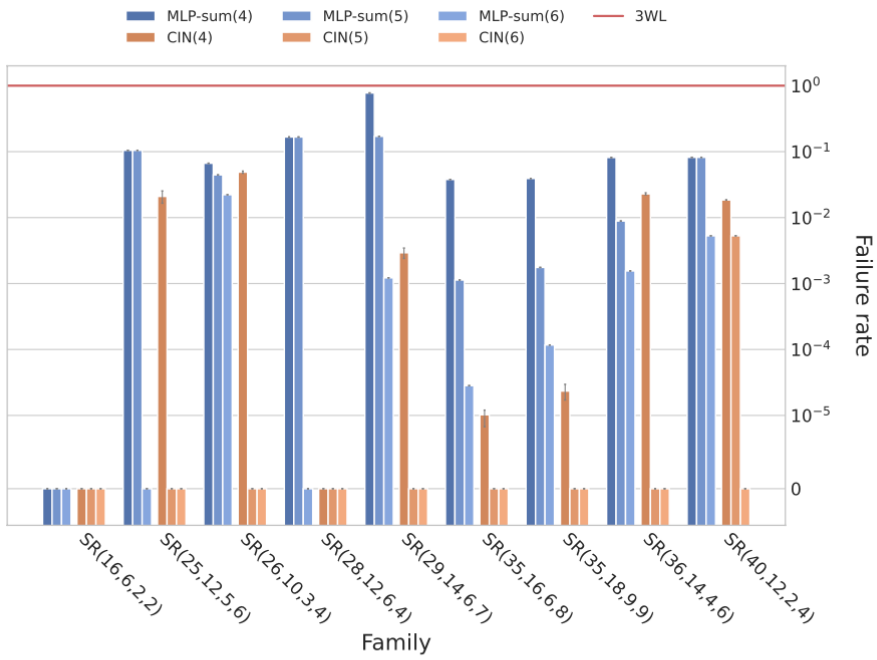
Method	Mean	Min	Max
MP-GNNs	10.000±0.000	10.000	10.000
RingGNN	10.000±0.000	10.000	10.000
3WLGNN	97.800±10.916	30.000	100.000
CIN (Ours)	100.000±0.000	100.000	100.000

Experiments

Synthetic Benchmarks

SR (Strongly Regular) Graphs:

- These graphs are non-isomorphic and challenging for GNNs to distinguish.
- The CWN model, when trained on cell complex lifting of each graph, performed well without training on the specific graph structure.
- Better performance was observed for larger graph sizes (denoted by k), with almost perfect discrimination for k = 6.



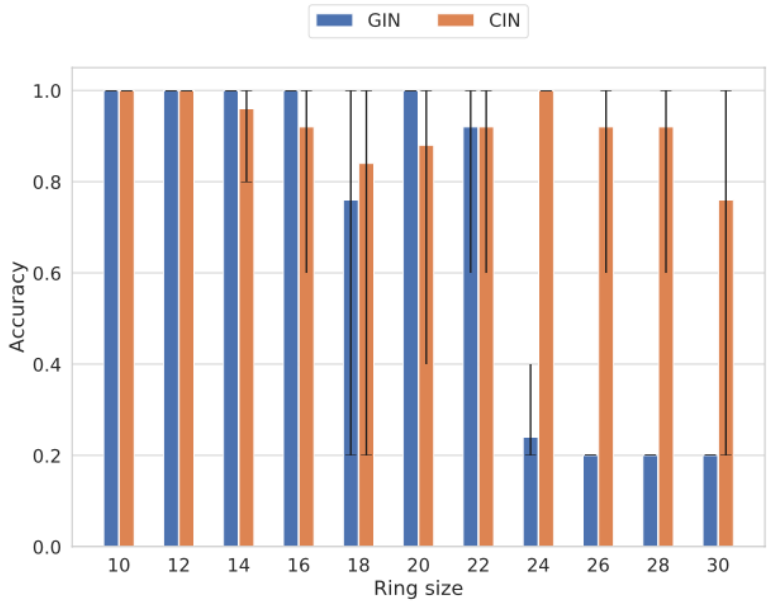
(b) Failure rates on the SR isomorphism task, *the smaller the better* (mean and std-error over 5 runs). In parantheses, for each model, the maximum size k of rings lifted to 2-cells.

Experiments

Synthetic Benchmarks

RingTransfer Benchmark:

- A synthetic benchmark was created to test the ability of CWNs to **capture long-range node dependencies**.
- The benchmark consists of chordless cycles (rings) where the task is to transfer a signal between two special nodes.
- CWNs showed good performance with only **3 computational steps**, regardless of the ring size.
- In contrast, GNNs like GIN showed performance degradation for **larger ring** sizes and complete failure for k > 24 due to difficulties in training such a deep network.



(a) RingTransfer Results. Accuracy is over 5 balanced classes. A score of 0.2 is equivalent to a random guess. Error bars show the min and max. Our model obtains high-scores in average even for large rings despite using only three layers.

Experiments

Real-World Graph Benchmarks: TUDataset benchmarks.
These benchmarks cover a range of domains from biology to social networks.

- CIN showed strong empirical performance, ranking top on four out of the eight datasets.
- The mean accuracy of CIN was particularly high on datasets relevant to biological and chemical domains where **rings play a significant role**.

Table 2: TUDatasets. The first section of the table includes the accuracy of graph kernel methods, while the second includes GNNs. The top three are highlighted by **First**, **Second**, **Third**.

Dataset	MUTAG	PTC	PROTEINS	NCI1	NCI109	IMDB-B	IMDB-M	RDT-B
RWK [29]	79.2±2.1	55.9±0.3	59.6±0.1	>3 days	N/A	N/A	N/A	N/A
GK ($k = 3$) [64]	81.4±1.7	55.7±0.5	71.4±0.3	62.5±0.3	62.4±0.3	N/A	N/A	N/A
PK [58]	76.0±2.7	59.5±2.4	73.7±0.7	82.5±0.5	N/A	N/A	N/A	N/A
WL kernel [65]	90.4±5.7	59.9±4.3	75.0±3.1	**86.0**±1.8	N/A	73.8±3.9	50.9±3.8	81.0±3.1
DCNN [3]	N/A	N/A	61.3±1.6	56.6±1.0	N/A	49.1±1.4	33.5±1.4	N/A
DGCNN [76]	85.8±1.8	58.6±2.5	75.5±0.9	74.4±0.5	N/A	70.0±0.9	47.8±0.9	N/A
IGN [52]	83.9±13.0	58.5±6.9	**76.6**±5.5	74.3±2.7	**72.8**±1.5	72.0±5.5	48.7±3.4	N/A
GIN [74]	89.4±5.6	64.6±7.0	76.2±2.8	82.7±1.7	N/A	75.1±5.1	52.3±2.8	**92.4**±2.5
PPGNs [53]	**90.6**±8.7	66.2±6.6	**77.2**±4.7	83.2±1.1	**82.2**±1.4	73.0±5.8	50.5±3.6	N/A
Natural GN [21]	89.4±1.6	**66.8**±1.7	71.7±1.0	82.4±1.3	N/A	73.5±2.0	51.3±1.5	N/A
GSN [10]	**92.2** ± 7.5	**68.2** ± 7.2	**76.6** ± 5.0	**83.5** ± 2.0	N/A	**77.8** ± 3.3	**54.3** ± 3.3	N/A
SIN [8]	N/A	N/A	76.4 ± 3.3	82.7 ± 2.1	N/A	**75.6** ± 3.2	**52.4** ± 2.9	**92.2** ± 1.0
CIN (Ours)	**92.7** ± 6.1	**68.2** ± 5.6	**77.0** ± 4.3	**83.6** ± 1.4	**84.0** ± 1.6	**75.6** ± 3.7	**52.7** ± 3.1	**92.4** ± 2.1

Conclusion

1. A message passing procedure on **cell complexes** has been proposed which is provably powerful.

2. The procedure is motivated by a **novel color refinement algorithm** used to test **graph isomorphism**.

3. The approach allows for flexible lifting operations on graphs to **create more expressive network architectures**.

4. These architectures benefit from the **separation of computational and input graphs**.

5. The methods presented have demonstrated excellent performance across diverse **synthetic** and **real-world molecular benchmarks**.

2024.1.10

Thank you!

Liyan Tan