International Journal of Education and Science Research Review
Volume-12, Issue-2 March-April-2025 E-ISSN 2348-6457
(Peer-Reviewed, Open Access, Fully Refereed International Journal) P-ISSN 2349-1817
www.ijesrr.org Impact Factor - 7.02 Email- editor@ijesrr.org

A NOVEL HIGHER-ORDER WEISFEILER-LEHMAN GRAPH CONVOLUTION

Dr. Preeti Panwar, Associate Professor in Mathematics

Guru Nanak Khalsa College,Karnal

Abstract

Graph Neural Networks (GNNs) have achieved remarkable success in learning from graph-structured data. However, standard message-passing GNNs are limited in their expressive power, being at most as powerful as the 1-dimensional Weisfeiler-Lehman (1-WL) graph isomorphism test. This paper introduces a novel higher-order Weisfeiler-Lehman Graph Convolution (HO-WL-GC) that significantly enhances the representational capacity of GNNs. Our approach systematically incorporates higher-order structural information while maintaining computational efficiency through a hierarchical message-passing scheme. Theoretical analysis demonstrates that HO-WL-GC achieves the discriminative power of k-WL tests for k > 1 without incurring the full computational complexity typically associated with higher-order methods. Extensive experiments on synthetic and real-world datasets show that our model consistently outperforms state-of-the-art GNNs on tasks requiring structural understanding, including molecular property prediction, social network analysis, and graph classification benchmarks. Our findings bridge the gap between theoretical expressiveness and practical performance in graph representation learning, offering a powerful yet efficient approach for capturing complex graph structures.

Keywords*:* Graph neural networks, Weisfeiler-Lehman test, higher-order convolution, graph representation learning, graph isomorphism

Introduction

Graphs are universal data structures that naturally represent a wide range of complex systems, from social networks and chemical compounds to biological structures and knowledge graphs. Effectively learning from graph-structured data has thus become a central challenge in machine learning research. Graph Neural Networks (GNNs) have emerged as powerful tools for this purpose, achieving impressive results across numerous domains and applications (Bronstein et al., 2017; Hamilton et al., 2017; Wu et al., 2020).

Despite their success, standard GNNs based on message-passing frameworks face fundamental limitations in their expressive power. Recent theoretical work has established that these models are at most as powerful as the 1-dimensional Weisfeiler-Lehman (1-WL) graph isomorphism test (Xu et al., 2019; Morris et al., 2019), meaning they cannot distinguish certain non-isomorphic graph structures. This limitation becomes particularly problematic in domains where fine-grained structural patterns determine the properties of interest, such as molecular chemistry or complex social network analysis.

Several approaches have been proposed to enhance the expressive power of GNNs, including higher-order graph networks (Morris et al., 2019), substructure-aware GNNs (Bouritsas et al., 2020), and spectral methods (Balcilar et al., 2021). However, these methods often face a critical trade-off between expressiveness and computational efficiency. Higher-order methods that simulate k-WL tests typically incur O(n^k) computational complexity, making them impractical for large-scale applications.

In this paper, we introduce a novel Higher-Order Weisfeiler-Lehman Graph Convolution (HO-WL-GC) that addresses this fundamental challenge. Our approach systematically incorporates higher-order structural information while maintaining computational tractability through a hierarchical message-passing scheme. Specifically, we make the following contributions:

1. We develop a novel graph convolution operation that captures higher-order structural information while preserving computational efficiency.
2. We provide theoretical analysis demonstrating that our approach achieves the discriminative power of k-WL tests for k > 1 without incurring the full O(n^k) computational complexity.
3. We propose an adaptive sampling strategy that further improves efficiency while preserving the most informative higher-order patterns.
4. We conduct extensive experiments on both synthetic and real-world datasets, demonstrating that our model consistently outperforms state-of-the-art GNNs on tasks requiring structural understanding.

Our work bridges the gap between theoretical expressiveness and practical performance in graph representation learning, offering a powerful yet efficient approach for capturing complex graph structures. The proposed methods are particularly effective for applications where fine-grained structural differences determine the properties of interest, such as molecular property prediction, social network analysis, and graph classification.

Methodology

Preliminaries

Graph Notation

Let G = (V, E) be a graph with vertex set V = {v_1, v_2, ..., v_n} and edge set E ⊆ V × V. Each node v_i may have a feature vector x_i∈ℝ^d. For simplicity, we focus on undirected graphs, though our approach generalizes to directed and weighted graphs.

Weisfeiler-Lehman Test

The 1-dimensional Weisfeiler-Lehman (1-WL) test, also known as color refinement, is a graph isomorphism test that proceeds as follows:

1. Initialize node colors (labels) c^(0)(v) based on node features.
2. Iteratively update colors: c^(t+1)(v) = HASH(c^(t)(v), {c^(t)(u) | u ∈ N(v)}), where N(v) denotes the neighbors of node v.

3. If the resulting multisets of colors differ between two graphs, they are not isomorphic.

The k-dimensional WL test (k-WL) extends this by considering k-tuples of nodes instead of individual nodes, offering strictly more discriminative power for k > 1.

Message-Passing GNNs

Standard message-passing GNNs update node representations through:

$h_v^{(t+1)} = \text{UPDATE}(h_v^{(t)}, \text{AGGREGATE}(\{h_u^{(t)} \mid u \in N(v)\}))$

where $h_v^{(t)}$ is the feature vector of node v at iteration t, and AGGREGATE and UPDATE are differentiable functions.

Higher-Order Weisfeiler-Lehman Graph Convolution

We now introduce our Higher-Order Weisfeiler-Lehman Graph Convolution (HO-WL-GC) framework. The key insight of our approach is to hierarchically build higher-order representations while maintaining computational efficiency.

Hierarchical Structure Representation

Instead of directly operating on k-tuples (which would incur $O(n^k)$ complexity), we build representations hierarchically:

1. First-order representations capture node-level information.
2. Second-order representations capture edge and local neighborhood patterns.
3. Higher-order representations capture increasingly complex structural motifs.

Formally, for a node v, we define its r-order contextual representation $h_v^{(r,t)}$ at iteration t as follows:

First-order (r=1): $h_v^{(1,t+1)} = \varphi_1(h_v^{(1,t)}, \text{AGGREGATE}_1(\{h_u^{(1,t)} \mid u \in N(v)\}))$

Second-order (r=2): $h_v^{(2,t+1)} = \varphi_2(h_v^{(2,t)}, \text{AGGREGATE}_2(\{h_v^{(1,t+1)}, h_u^{(1,t+1)}, h_v^{(2,t)}, h_u^{(2,t)} \mid u \in N(v)\}))$

Higher-order (r>2): $h_v^{(r,t+1)} = \varphi_r(h_v^{(r,t)}, \text{AGGREGATE}_r(\{h_v^{(r-1,t+1)}, h_u^{(r-1,t+1)}, h_v^{(r,t)}, h_u^{(r,t)} \mid u \in N(v)\}))$

where φ_r and AGGREGATE_r are order-specific transformation and aggregation functions, typically implemented as neural networks.

Efficient Higher-Order Message Passing

To ensure efficiency, we implement the higher-order aggregation operations through a novel tensor formulation. For each order r, we maintain a representation tensor $H^{(r)}$ of appropriate dimensions. The key innovation is our message-passing scheme that computes higher-order interactions without explicitly materializing all possible tuples.

For example, the second-order update can be efficiently implemented as:

$H^{(2,t+1)} = \varphi_2(H^{(2,t)}, A \odot (H^{(1,t+1)} \otimes H^{(1,t+1)}))$

where A is the adjacency matrix, ⊙ represents element-wise multiplication, and ⊗ denotes the appropriate tensor product operation.

Adaptive Sampling Strategy

To further improve efficiency while preserving expressive power, we introduce an adaptive sampling strategy that selectively processes the most informative higher-order patterns:

1. We maintain an importance score S_v^(r) for each node at each order r.
2. Based on these scores, we sample a subset of nodes for higher-order processing.
3. Importance scores are updated based on the gradients flowing through the model during training.

This approach reduces the computational burden while focusing computational resources on the most discriminative structural patterns.

Model Architecture

The complete HO-WL-GC architecture consists of:

1. Multiple layers of hierarchical structure representation learning (as described above).
2. A readout function that combines information across different orders for the final prediction:

OUTPUT(G) = MLP(POOL({h_v^(r,T) | v ∈ V, r ∈ {1,...,R}}))

where POOL is a pooling operation (e.g., mean or sum pooling), T is the final iteration, and R is the maximum order considered.

Theoretical Analysis

We now present theoretical results on the expressive power of our HO-WL-GC model.

Theorem 1: For any k ≥ 2, there exists an instance of HO-WL-GC with maximum order R = k that can distinguish all graph pairs that are distinguishable by the k-WL test.

***Proof Sketch*:** We show that our hierarchical representation can simulate the k-WL test by inductively building representations that capture the same information as the k-WL colorrefinement process. The key insight is that our higher-order updates combine information from lower-order representations in a way that preserves the discriminative power of k-WL.

Theorem 2: The computational complexity of HO-WL-GC with maximum order R is $O(n^2 \cdot R)$, compared to O(n^R) for direct simulation of R-WL.

***Proof Sketch*:** Our hierarchical approach processes each order sequentially, with the most expensive operations being the second-order updates that involve all pairs of connected nodes. The adaptive sampling strategy further reduces this complexity in practice.

Theorem 3: HO-WL-GC with maximum order R = 3 can count all connected subgraphs with up to 4 nodes.

***Proof Sketch*:** We demonstrate that the third-order representations capture sufficient information to distinguish all possible configurations of connected 4-node subgraphs. This is achieved through the hierarchical combination of lower-order structural information.

EXPERIMENTAL RESULTS

Experimental Setup

Datasets

We evaluate our method on both synthetic and real-world datasets:

Synthetic Datasets:

- SR-GRAPHS: Strongly regular graphs with the same degree, number of common neighbors for adjacent vertices, and number of common neighbors for non-adjacent vertices.
- CFI-GRAPHS: Cai-Fürer-Immerman graphs that are hard instances for the 1-WL test.

Real World Datasets:

- Molecular datasets: ZINC, QM9, and MUTAG for molecular property prediction.
- Social network datasets: COLLAB, IMDB-BINARY, and REDDIT-BINARY for graph classification.
- Citation networks: CORA, CITESEER, and PUBMED for node classification.

Table 1 summarizes the statistics of these datasets.

Table 1: Dataset Statistics

Dataset	Type	#Graphs	Avg. #Nodes	Avg. #Edges	#Classes	Task
SR-GRAPHS	Synthetic	500	50	350	5	Graph
CFI-GRAPHS	Synthetic	300	40	120	2	Graph
ZINC	Molecular	12,000	23.2	24.9	-	Regression
QM9	Molecular	133,885	18.0	18.6	-	Regression
MUTAG	Molecular	188	17.9	19.8	2	Graph
COLLAB	Social	5,000	74.5	2,457.8	3	Graph
IMDB-BINARY	Social	1,000	19.8	96.5	2	Graph
REDDIT-BINARY	Social	2,000	429.6	497.8	2	Graph
CORA	Citation	1	2,708	5,429	7	Node
CITESEER	Citation	1	3,327	4,732	6	Node
PUBMED	Citation	1	19,717	44,338	3	Node

Baselines

We compare our HO-WL-GC with the following state-of-the-art methods:

Standard GNNs: GCN (Kipf& Welling, 2017), GAT (Veličković et al., 2018), GIN (Xu et al., 2019)

Higher-Order GNNs: 3-GNN (Morris et al., 2019), PPGN (Maron et al., 2019), Ring-GNN (Chen et al., 2019)

Substructure-Based GNNs: GSN (Bouritsas et al., 2020), ID-GNN (You et al., 2021)

Implementation Details

We implement HO-WL-GC with the following configuration:

- Maximum order R = 3
- Hidden dimension = 128
- Number of layers = 4
- Optimizer: Adam with learning rate 0.001
- Batch size: 32
- Early stopping with patience 30
- Adaptive sampling budget: 40% of nodes for second order, 20% for third order

For all methods, we perform hyperparameter tuning using grid search on validation sets. We report the mean and standard deviation of performance metrics over 10 runs with different random seeds.

Results on Synthetic Datasets

We first evaluate the models on synthetic datasets designed to test the discriminative power of graph learning algorithms.

Table 2: Results on Synthetic Datasets (Accuracy %)

Method	SR-GRAPHS	CFI-GRAPHS
GCN	51.2 ± 1.7	50.6 ± 2.1
GAT	52.8 ± 2.0	51.3 ± 1.8
GIN	53.5 ± 1.4	52.1 ± 1.9
3-GNN	89.4 ± 2.3	91.2 ± 1.7
PPGN	82.7 ± 2.5	88.5 ± 2.4
Ring-GNN	85.8 ± 2.2	87.7 ± 2.6
GSN	65.3 ± 2.1	72.6 ± 2.3
ID-GNN	68.2 ± 1.9	74.5 ± 2.5
HO-WL-GC (Ours)	**92.1 ± 1.5**	**93.4 ± 1.6**

As shown in Table 2, standard GNNs (GCN, GAT, GIN) perform only slightly better than random guessing on these datasets, confirming their limited expressive power. Higher-order methods (3-GNN, PPGN, Ring-GNN) achieve significantly better results, demonstrating the importance of capturing higher-order structures. Our HO-WL-GC method outperforms all baselines, achieving the highest accuracy on both datasets.

Results on Molecular Datasets

Table 3 presents results on molecular property prediction tasks, where capturing structural patterns is crucial for accurate predictions.

Table 3: Results on Molecular Datasets

Method	ZINC (MAE↓)	QM9 (MAE↓)	MUTAG (Acc.↑)
GCN	0.469 ± 0.002	0.128 ± 0.001	76.3 ± 2.1

GAT	0.463 ± 0.003	0.119 ± 0.002	78.5 ± 1.9
GIN	0.387 ± 0.004	0.115 ± 0.002	89.4 ± 1.6
3-GNN	0.342 ± 0.003	0.093 ± 0.002	92.0 ± 1.7
PPGN	0.329 ± 0.004	0.095 ± 0.003	91.2 ± 1.9
Ring-GNN	0.353 ± 0.003	0.097 ± 0.002	91.5 ± 1.8
GSN	0.338 ± 0.004	0.106 ± 0.002	90.8 ± 1.6
ID-GNN	0.341 ± 0.003	0.102 ± 0.002	90.2 ± 1.7
HO-WL-GC (Ours)	**0.321 ± 0.003**	**0.089 ± 0.002**	**93.4 ± 1.4**

Our HO-WL-GC model achieves the lowest Mean Absolute Error (MAE) on both ZINC and QM9 regression tasks and the highest accuracy on MUTAG classification. The performance gap between standard GNNs and higher-order methods is particularly significant on these datasets, highlighting the importance of capturing higher-order structural information for molecular property prediction.

Results on Social Network Datasets

Table 4 shows the results on social network datasets, where the goal is to classify graphs based on social structures.

Table 4: Results on Social Network Datasets (Accuracy %)

Method	COLLAB	IMDB-BINARY	REDDIT-BINARY
GCN	73.9 ± 0.5	70.3 ± 0.8	85.2 ± 0.9
GAT	74.2 ± 0.6	71.6 ± 0.9	86.3 ± 0.7
GIN	80.2 ± 0.5	75.1 ± 0.7	92.4 ± 0.5
3-GNN	81.3 ± 0.6	74.8 ± 0.8	92.6 ± 0.6
PPGN	81.5 ± 0.5	75.5 ± 0.7	93.0 ± 0.5
Ring-GNN	81.0 ± 0.7	75.7 ± 0.9	92.8 ± 0.6
GSN	81.9 ± 0.6	76.2 ± 0.8	92.9 ± 0.7
ID-GNN	81.6 ± 0.5	76.0 ± 0.7	93.2 ± 0.5
HO-WL-GC (Ours)	**83.2 ± 0.4**	**77.4 ± 0.6**	**94.1 ± 0.4**

On social network datasets, all methods perform relatively well, but our HO-WL-GC still achieves the best results across all three datasets. The performance gap is smaller compared to molecular datasets, suggesting that capturing higher-order structures might be less critical for these particular social network tasks.

Results on Node Classification

Table 5 presents results on citation network datasets for node classification tasks.

Table 5: Results on Citation Networks (Accuracy %)

Method	CORA	CITESEER	PUBMED
GCN	81.5 ± 0.5	70.3 ± 0.7	79.0 ± 0.3

GAT	83.0 ± 0.7	72.5 ± 0.7	79.0 ± 0.3
GIN	80.2 ± 0.6	69.8 ± 0.8	78.7 ± 0.4
3-GNN	80.7 ± 0.6	70.2 ± 0.7	79.3 ± 0.3
PPGN	81.9 ± 0.5	71.0 ± 0.6	79.5 ± 0.4
Ring-GNN	81.2 ± 0.7	70.8 ± 0.8	79.2 ± 0.3
GSN	82.5 ± 0.6	71.3 ± 0.7	79.7 ± 0.3
ID-GNN	82.8 ± 0.5	71.5 ± 0.6	79.8 ± 0.4
HO-WL-GC (Ours)	**83.7 ± 0.4**	**73.2 ± 0.6**	**80.2 ± 0.3**

For node classification tasks, the performance differences between methods are less pronounced. Our HO-WL-GC still achieves the best results on all three datasets, but the margins are smaller compared to graph-level tasks. This is consistent with theoretical expectations, as node classification may rely more on local neighborhood information rather than global graph structure.

Efficiency Analysis

We analyze the computational efficiency of different methods in Table 6, reporting both the theoretical complexity and empirical runtime on the ZINC dataset.

Table 6: Efficiency Analysis on ZINC Dataset

Method	Theoretical Complexity	Training Time (s/epoch)	Memory Usage (GB)
GCN	O(	E	d)
GAT	O(	E	d^2)
GIN	O(	E	d)
3-GNN	O(	V	^{3}d)
PPGN	O(	V	^{2}d)
Ring-GNN	O(	V	^{2}d)
GSN	O(	E	d +
ID-GNN	O(	E	d·log
HO-WL-GC (Full)	O(	V	^{2}·R·d)
HO-WL-GC (Adaptive)	O(	E	d + s

where |V| is the number of nodes, |E| is the number of edges, d is the hidden dimension, R is the maximum order, and s is the sampling ratio.

The results show that standard GNNs (GCN, GAT, GIN) are the most efficient, while direct higher-order methods (3-GNN) are the most computationally demanding. Our HO-WL-GC with adaptive sampling achieves a favorable trade-off, with computational requirements closer to substructure-based methods while maintaining the expressiveness of higher-order approaches.

Ablation Studies

We conduct ablation studies to analyze the contribution of different components of our HO-WL-GC model. Table 7 presents results on the ZINC dataset with various configurations.

Table 7: Ablation Studies on ZINC Dataset (MAE↓)

Model Variant	MAE
HO-WL-GC (Full)	0.321 ± 0.003
HO-WL-GC (R=1, first-order only)	0.388 ± 0.003
HO-WL-GC (R=2, up to second-order)	0.329 ± 0.003
HO-WL-GC (R=3, no adaptive sampling)	0.323 ± 0.004
HO-WL-GC (R=3, random sampling)	0.342 ± 0.003
HO-WL-GC (R=3, degree-based sampling)	0.330 ± 0.004
HO-WL-GC (R=3, attention-based aggregation)	0.318 ± 0.003

These results demonstrate several important findings:

1. Higher-order representations significantly improve performance (comparing R=1 vs. R=2 vs. R=3).
2. Adaptive sampling is crucial for maintaining performance while improving efficiency (comparing "no adaptive sampling" vs. "random sampling" vs. "adaptive sampling").
3. Alternative aggregation schemes like attention-based aggregation can further improve performance.

DISCUSSION

Expressiveness vs. Efficiency Trade-off

Our experiments demonstrate that HO-WL-GC successfully addresses the fundamental trade-off between expressiveness and efficiency in graph representation learning. The proposed hierarchical approach achieves the discriminative power of higher-order WL tests without incurring the full computational complexity, making it applicable to real-world datasets of practical sizes.

The adaptive sampling strategy further improves efficiency by focusing computational resources on the most informative higher-order patterns. This is particularly important for large graphs where processing all possible higher-order interactions would be prohibitively expensive.

Domain-Specific Insights

Our results across different domains reveal interesting patterns:

1. **Molecular Datasets**: The significant performance gains on molecular datasets highlight the importance of capturing higher-order structural motifs for molecular property prediction. This aligns with chemical intuition, as functional groups and substructures often determine molecular properties.
2. **Social Networks**: The moderate improvements on social network datasets suggest that while higher-order structures are beneficial, they might be less critical for the specific tasks evaluated. This

could be because the social network datasets in our study may rely more on community structure than on specific motifs.

3. **Citation Networks**: The relatively small improvements on node classification tasks indicate that local neighborhood information might be sufficient for these tasks, with higher-order structures providing marginal benefits.

Limitations and Future Work

Despite its strong performance, our approach has several limitations that point to directions for future research:

1. **Scalability to Very Large Graphs**: While more efficient than direct higher-order methods, HO-WL-GC still faces challenges with very large graphs (millions of nodes). Future work could explore more aggressive approximation techniques or distributed computing approaches.

2. **Heterogeneous Graphs**: Our current formulation focuses on homogeneous graphs. Extending the approach to heterogeneous graphs with multiple node and edge types is an important direction for future work.

3. **Interpretability**: Higher-order representations can be difficult to interpret. Developing visualization and explanation techniques for higher-order patterns would enhance the practical utility of our approach.

4. **Dynamic Graphs**: Adapting our method to dynamic graphs, where the structure evolves over time, presents interesting challenges and opportunities for capturing temporal higher-order patterns.

CONCLUSIONS

In this paper, we introduced Higher-Order Weisfeiler-Lehman Graph Convolution (HO-WL-GC), a novel graph neural network architecture that significantly enhances the representational capacity of GNNs while maintaining computational efficiency. Our approach systematically incorporates higher-order structural information through a hierarchical message-passing scheme, achieving the discriminative power of k-WL tests for $k > 1$ without incurring the full computational complexity typically associated with higher-order methods.

Theoretical analysis demonstrated that our model can distinguish graph structures that are indistinguishable by standard message-passing GNNs. Through extensive experiments on both synthetic and real-world datasets, we showed that HO-WL-GC consistently outperforms state-of-the-art GNNs across a variety of tasks, including molecular property prediction, social network analysis, and graph classification.

Our adaptive sampling strategy further improves efficiency by focusing computational resources on the most informative higher-order patterns. This makes our approach applicable to large-scale real-world problems where capturing complex structural patterns is crucial for accurate predictions.

The success of HO-WL-GC highlights the importance of bridging the gap between theoretical expressiveness and practical performance in graph representation learning. By developing models that can capture higher-order structural information efficiently, we open new possibilities for applications in chemistry, biology, social network analysis, and beyond.

Future work will focus on extending our approach to heterogeneous and dynamic graphs, improving scalability to very large graphs, and developing better interpretation techniques for higher-order structural patterns.

REFERENCES

Balcilar, M., Heroux, P., Gauzere, B., Vasseur, P., Adam, S., &Honeine, P. (2021). Breaking the limits of message passing graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning* (pp. 599-608). PMLR.

Bouritsas, G., Frasca, F., Zafeiriou, S., & Bronstein, M. M. (2020). Improving graph neural network expressivity via subgraph isomorphism counting. arXiv preprint arXiv:2006.09252.

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., &Vandergheynst, P. (2017). Geometric deep learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine, 34*(4), 18-42.

Chen, Z., Chen, L., Villar, S., &Bruna, J. (2019). Can graph neural networks count substructures? In *Advances in Neural Information Processing Systems* (pp. 10947-10957).

Defferrard, M., Bresson, X., &Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems* (pp. 3844-3852).

Duvenaud, D. K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., & Adams, R. P. (2015). Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems* (pp. 2224-2232).

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 1263-1272). PMLR.

Grohe, M. (2017). *Descriptive complexity, canonisation, and definable graph structure theory*. Cambridge University Press.

Hamilton, W. L., Ying, R., &Leskovec, J. (2017). Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems* (pp. 1024-1034).

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*.

Kriege, N. M., Johansson, F. D., & Morris, C. (2020). A survey on graph kernels. *Applied Network Science, 5*(1), 1-42.

Li, Y., Tarlow, D., Brockschmidt, M., &Zemel, R. (2015). Gated graph sequence neural networks. arXiv preprint arXiv:1511.05493.

Maron, H., Ben-Hamu, H., Shamir, N., &Lipman, Y. (2019). Invariant and equivariant graph networks. In *International Conference on Learning Representations*.

Morris, C., Ritzert, M., Fey, M., Hamilton, W. L., Lenssen, J. E., Rattan, G., & Grohe, M. (2019). Weisfeiler and Leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 4602-4609).

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., &Monfardini, G. (2008). The graph neural network model. *IEEE Transactions on Neural Networks, 20*(1), 61-80.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., &Bengio, Y. (2018). Graph attention networks. In *International Conference on Learning Representations*.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems, 32*(1), 4-24.

Xu, K., Hu, W., Leskovec, J., &Jegelka, S. (2019). How powerful are graph neural networks? In *International Conference on Learning Representations*.

Yanardag, P., &Vishwanathan, S. V. N. (2015). Deep graph kernels. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1365-1374).

You, J., Gomes, J., Xiao, L., &Leskovec, J. (2021). Identity-aware graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 12, pp. 10737-10745).