Extending Graph Neural Networks with Global Features

Andrei Dragos Brasoveanu TU Wien
Fabian Jogl TU Wien
Pascal Welke TU Wien
Maximilian Thiessen TU Wien

{dragos-andrei.brasoveanu,fabian.jogl,
pascal.welke,maximilian.thiessen}@tuwien.ac.at

Abstract

A common approach to boost the predictive performance of message passing graph neural networks (MPNNs) is to attach additional features to nodes. In contrast, we propose to use expressive *global* graph features. This is motivated by the limited expressivity of MPNNs resulting in an inability to compute certain global graph properties, like the Wiener index and Hosoya index. Such global graph features are well known in fields like chemoinformatics but seem to be overlooked by the GNN community. We propose an architecture which extends graph embeddings learned by MPNNs with global graph features, for example, topological indices describing the entire graph. Analyzing different global graph features, we show that certain global features like the Wiener index increase the expressivity of MPNNs, while others like the Zagreb indices do not. Our first experiments indicate that adding global graph features improves the performance of MPNNs on molecular benchmark datasets.

1 Introduction

We propose to combine global graph features with learned graph embeddings to increase the expressivity of graph neural networks (GNNs) and their predictive performance. GNNs have made significant progress in learning with graph-based data, for example, in domains such as chemoinformatics [Gilmer et al., 2017] and bioinformatics [Gainza et al., 2020]. Still, in-depth exploration of GNNs has revealed a fundamental challenge limiting their expressivity. This challenge arises from the networks' message passing approach, which is bounded in expressivity by the Weisfeiler-Leman graph isomorphism test (1-WL) [Xu et al., 2019; Morris et al., 2019].

Our goal is to tackle this challenge with global graph features. In chemoinformatics there exists a large number of specialized graph features called *topological indices* [Balaban, 1982] or *molecular fingerprints* [Cereto-Massagué et al., 2015; Muegge and Mukherjee, 2016; Capecchi et al., 2020] that are under-explored by the GNN community. These indices are known to capture important chemical properties on the *global graph* level, e.g., of the whole molecule. This is in contrast to most positional encodings [Dwivedi et al., 2023] studied so far, which capture local information and hence make sense as node features. Thus, we propose to combine global graph features with the graph embeddings learned by a GNN (see Figure 1). In this paper, we investigate a diverse set of global graph features (see Table 1). We show whether they improve the expressivity of MPNNs and empirically investigate their impact on the predictive performance.

Table 1: Expressivity of the global graph features

Global feature	Increases expressivity
Wiener index	✓
Maximum matching	✓
Hosoya index	✓
Independence number	✓
Second eigenvalue	✓
Circuit rank	✓
Spectral radius	✗
Zagreb M_1	✗
Zagreb M_2	✗

A. Brasoveanu et al., Extending Graph Neural Networks with Global Features (Extended Abstract). Presented at the Second Learning on Graphs Conference (LoG 2023), Virtual Event, November 27–30, 2023.

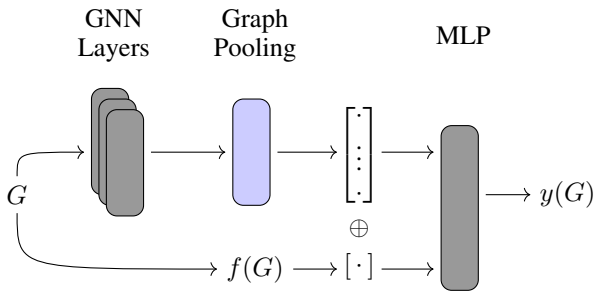


Figure 1: Our proposed GNN architecture. We concatenate global graph features to the output of the graph pooling layer after multiple GNN layers. Figure adapted from Welke et al. [2023].

2 GNNs with Global Graph Features

A common assumption for GNNs is that predictions should remain unchanged regardless of the order in which the nodes appear in the graph. This is formalized as *permutation invariance*: Let ϕ be a function mapping graphs to real-valued vectors. Then, ϕ is permutation invariant if for every graph G and permutation P of its vertices it holds that $\phi(P(G)) = \phi(G)$. This means that permutation invariant functions map isomorphic graphs to the same embedding. The ability of a permutation invariant function ϕ to map non-isomorphic graphs G, H to different embeddings $\phi(G) \neq \phi(H)$ is called *expressivity*. Most GNNs are designed to be differentiable permutation invariant functions $\mathrm{GNN} : \mathcal{G} \to \mathbb{R}^d$ where $\mathcal{G}$ is the set of all graphs. MPNNs in particular can be seen as a neural variant of the Weisfeiler-Leman algorithm (1-WL test) and are bounded by its expressivity [Xu et al., 2019; Morris et al., 2019]. As a result, there exist graphs that cannot be distinguished by any MPNN.

We propose an architecture that combines the learned graph embedding function of a GNN with another permutation invariant function on graphs by concatenating their outputs. In particular, we investigate existing numerical or ordinal properties of graphs which can be computed in practice.

Definition 1. *(Global Graph Feature) Let f be a permutation invariant function with $f : \mathcal{G} \to \mathbb{R}$. Then we call f a* global graph feature*.*

The concatenation of a global graph feature f and a learned graph embedding GNN will never decrease the expressivity of GNN or f. For 1-WL there exist global graph features that provably increase the expressivity, see Table 1 and Appendix B. A permutation invariant function on graphs is (strictly) more expressive than 1-WL if it can distinguish every pair of graphs that 1-WL can distinguish and at least one pair of graphs that 1-WL cannot distinguish. We call a global graph feature f *expressivity-increasing* if for a pair 1-WL indistinguishable graph G and H it holds $f(G) \neq f(H)$.

Proposition 1. *Let f be an expressivity-increasing global graph feature. Then, (1-WL, f) is strictly more expressive than 1-WL.*

For MPNNs this implies that we can increase their expressivity with global graph features: We concatenate a global graph feature to their output and send the result through a multilayer perceptron (MLP), which generates the final prediction. If the MLP computes and injective function, the resulting function is strictly more expressive than the MPNN without the concatenated global graph feature. Figure 1 shows our proposed architecture.

Our selection of global graph features. We investigate a set of global graph features based on node degrees, connectivity, distances, or spectral properties. We choose a representative set of indices that (1) consider different aspects of the graph and (2) have been used before in the field of computational chemistry. For this initial work, we decided to test features individually rather than using composite features. For example, the Wiener index W is computed using the graph distance matrix D by $W = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} D_{ij}$. Similarly, the related Balaban index J [Balaban, 1982, 1988] is computed as $J = \frac{m}{\gamma+1}\sum_{(i,j)\in E(G)}(D_{i\cdot}^T D_{j\cdot})^{-\frac{1}{2}}$, where γ is the circuit rank and m is the number of edges. Thus, we select only the Wiener index and the circuit rank in this case. Another well-known molecular index is the Hosoya index, which counts the number of matchings in the graph. We also use the size of the maximum matching. Both M_1 and M_2 Zagreb indices are defined over the node degrees. For graph connectivity, we choose the circuit rank and the second smallest eigenvalue of

the Laplacian matrix. Finally, we choose the independence number and the spectral radius (of the adjacency matrix) as global graph features. For a more detailed overview and a discussion whether each global graph features increases expressivity, see Appendix B. In particular, we explore the expressivity of the global graph features on 3 pairs of 1-WL indistinguishable graphs (Table 7,8,9).

3 Discussion and Related Work

Attaching features on the global level is a simple and versatile approach to making GNNs more expressive and providing a boost to their predictive performance. Graph level features can be attached to the embedding of *any* GNN without making changes to the GNN architecture. Furthermore, they can be used to increase the expressivity of a GNN *after* training it, by attaching the global graph features and only fine tuning the final MLP layers. In contrast, it is also possible to attach features to the nodes or edges, as has been done by Barceló et al. [2021] and Bouritsas et al. [2022]. However, for many global graph features there is no trivial way of *localizing* the graph-level feature to the node-level. For example, it is possible to obtain a local version of the Wiener index by summing all shortest path distances with the starting point fixed to a specific node. However, this is not possible, for example, with the eigenvalues of the Laplacian as they are inherently non-local; instead we can only attach the same value to all nodes. This would add redundant information and thus it is more appropriate to use such features globally.

Global graph features in the wild. Some previous GNN architectures have implicitly or explicitly used global features. Battaglia et al. [2018] discussed global features as part of a general learning on graphs framework. However, they do not mention global features in the context of expressivity. More recently, Lim et al. [2023] proposed SignNet, where eigenvalues can be added in each layer. They discuss the expressivity of SignNet with and without added eigenvalues, relying on common results concerning Weisfeiler-Leman and spectral graph theory, see, e.g., Rattan and Seppelt [2023]. While most pattern counting based approaches attach the counts to the nodes [Barceló et al., 2021; Bouritsas et al., 2022], Welke et al. [2023] consider homomorphism counts as global features. Finally, the circuit rank, also called first Betti number, was used for topology-based GNNs [Horn et al., 2022].

4 Experiments

In this extended abstract, we present initial experiments to evaluate the potential benefit of global graph features. Our experiments[1] are designed to measure the performance gain of MPNNs when attaching global graph features and not to beat the state-of-the-art.

Baseline. We experiment with the Graph Isomorphism Network (GIN) [Xu et al., 2019]. We tune hyperparameters for the GIN baseline with no global graph features attached. When evaluating GIN with global graph features we use the best hyperparameters of the baseline. Hyperparameter grids can be found in Appendix A. We also introduce an one-dimensional uniform noise feature in the range $[0, 1)$ for an ablation study. We treat it as a global graph feature to investigate if practical performance increases are due to additional finetuning and additional weights in the MLP.

Datasets. We use the datasets `ZINC` [Irwin and Shoichet, 2005] and `MOLHIV` [Hu et al., 2020] to evaluate the predictive performance of our proposed architecture on molecular data. `ZINC` is a regression task dataset of molecular graphs (we use the variant [Dwivedi et al., 2023] with 12000 graphs) with the aim to predict the constrained solubility value for each molecule. `MOLHIV` is a large molecular dataset (containing approximately 41000 graphs) from OGB [Hu et al., 2020], with the task to predict whether a molecule inhibits HIV.

Setup. For each global graph feature in Table 1, we train and evaluate 10 times on each dataset to compute the mean and standard deviation of the target test metrics in the epoch with the highest validation metric. On `ZINC` we use a batch size of 128 and an initial learning rate of 10^{-3}. We halve the learning rate whenever the validation metric does not improve for 20 epochs and stop the training after 1000 epochs or after the learning rate dips below 10^{-5}. On `MOLHIV` it is common [Bevilacqua et al., 2022] to only train for 100 epochs with a fixed learning rate and a batch size of 32. However,

[1] `https://github.com/andreibrasoveanu97/gnn-global-features`

Table 2: Test results for ZINC and MOLHIV datasets. The evaluation metric is written in brackets and ↓ / ↑ denotes that smaller / larger values are favorable. **Bold** values are better by at least one standard deviation for a given global-feature and dataset. Red values are the best result for a given dataset.

	ZINC (MAE ↓)		MOLHIV (ROC-AUC ↑)	
Global feature f	GIN	(GIN, f)	GIN	(GIN, f)
Wiener	0.187 ± 0.004	$\mathbf{0.177 \pm 0.004}$	0.7659 ± 0.015	0.7662 ± 0.015
Circuit rank	0.182 ± 0.006	0.179 ± 0.004	0.7620 ± 0.014	0.7673 ± 0.012
Independence	0.187 ± 0.005	$\mathbf{0.175 \pm 0.004}$	0.7719 ± 0.009	0.7705 ± 0.009
Hosoya	0.184 ± 0.002	$\mathbf{0.18 \pm 0.004}$	—	
Max. matching	0.186 ± 0.004	$\mathbf{0.175 \pm 0.003}$	0.7600 ± 0.017	0.7602 ± 0.013
Second eigenv.	0.187 ± 0.004	$\mathbf{0.174 \pm 0.004}$	0.7686 ± 0.013	0.7705 ± 0.010
Spectral radius	0.184 ± 0.003	$\mathbf{0.181 \pm 0.003}$	0.7685 ± 0.015	0.7707 ± 0.016
Zagreb M_1	0.188 ± 0.004	$\mathbf{0.177 \pm 0.003}$	0.7676 ± 0.015	0.7683 ± 0.014
Zagreb M_2	0.183 ± 0.004	$\mathbf{0.177 \pm 0.004}$	0.7714 ± 0.019	0.7761 ± 0.018
All features	0.186 ± 0.006	$\mathbf{0.164 \pm 0.003}$	0.7696 ± 0.018	0.7718 ± 0.014
Noise	0.185 ± 0.005	0.182 ± 0.004	0.7687 ± 0.017	0.7614 ± 0.018

we have found that in this setup the training is very noisy. Thus, we use a learning rate decay for MOLHIV: we use an initial learning rate of 10^{-3} which we half whenever the validation metric does not improve for 5 epochs. We train with a batch size of 32 and stop training after 100 epochs.

Results. Table 2 shows our experimental results. For MOLHIV, global features only yield slightly improved performance in 7 out of 9 datasets. However, the differences are much smaller than the standard deviation. For ZINC, global graph features give consistent and strong performance improvements over the GIN baseline. Combining all global graph features together ('all features') achieves the best results. This indicates that while the performance gain from a single global graph feature is not too large, combining multiple such features can give significant performance improvements without changing the GNN architecture. Interestingly, the expressivity of the global graph feature has no direct impact on the performance. The non-expressivity-increasing Zagreb features performing competitively to the expressivity-increasing features.

Finally, we can see that the noise feature only yield small improvements on ZINC and are harmful for MOLHIV. This implies that the improvement from global graph features is due to the additional information provided by the feature and not due to extra fine-tuning in training.

5 Conclusion

In this work, we have started the investigation of *expressivity-increasing* global graph features. We have proven that many common global graph features (e.g., Wiener and Hosoya index) from chemoinformatics increase the expressivity of MPNNs, while other features do not (e.g., Zagreb M_1,M_2). Our initial experiments indicate that adding global graph features to GNNs consistently improves the performance on ZINC and does not harm performance on MOLHIV. Interestingly, even the non-expressivity-increasing global features improved the performance (on ZINC). Combining all 9 studied global graph features simultaneously increased the performance even more. Our approach is simple to implement and requires no changes to an existing GNN architecture. Indeed, we can combine a pre-trained GNN with a global graph feature and only train the final MLP.

As future work, we will consider more recent molecular fingerprints [Capecchi et al., 2020] and applications to other domains. In particular, it would be interesting to see whether molecular descriptors are also useful for, e.g., social networks. A different direction might be to use the global graph features as a pre-training step before the actual training on the predictive task. Indeed, by first training the GNN to be able to predict global graph features (in a self-supervised manner), assuming they are related to the downstream task, we might be able to improve the actual predictive performance. Finally, inspecting the selection of the global graph features from a feature engineering and/or boosting perspective (viewing the global features as weak learners) is promising.

References

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *ICML*, 2017. 1

P. Gainza, F. Sverrisson, F. Monti, E. Rodolà, D. Boscaini, M. M. Bronstein, and B. E. Correia. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, 17(2):184–192, 2020. 1

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *ICLR*, 2019. 1, 2, 3

Christopher Morris, Martin Ritzert, Matthias Fey, William Hamilton, Jan Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and Leman go neural: Higher-order graph neural networks. In *AAAI*, 2019. 1, 2

Alexandru T Balaban. Highly discriminating distance-based topological index. *Chemical Physics Letters*, 89(5):399–404, 1982. 1, 2

Adrià Cereto-Massagué, María José Ojeda, Cristina Valls, Miquel Mulero, Santiago Garcia-Vallvé, and Gerard Pujadas. Molecular fingerprint similarity search in virtual screening. *Methods*, 71: 58–63, 2015. 1

Ingo Muegge and Prasenjit Mukherjee. An overview of molecular fingerprint similarity search in virtual screening. *Expert Opinion on Drug Discovery*, 11(2):137–148, 2016. 1

Alice Capecchi, Daniel Probst, and Jean-Louis Reymond. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. *Journal of Cheminformatics*, 12(1):1–15, 2020. 1, 4

Vijay Prakash Dwivedi, Chaitanya K. Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research*, 24 (43):1–48, 2023. 1, 3

Pascal Welke, Maximilian Thiessen, Fabian Jogl, and Thomas Gärtner. Expectation complete graph representations using graph homomorphisms. In *ICML*, 2023. 2, 3

Alexandru Balaban. Topological indices based on topological distances in molecular graph. *Pure and Applied Chemistry*, 55:199–206, 01 1988. 2

Pablo Barceló, Floris Geerts, Juan L. Reutter, and Maksimilian Ryschkov. Graph neural networks with local graph parameters. In *NeurIPS*, 2021. 3

Giorgos Bouritsas, Fabrizio Frasca, Stefanos Zafeiriou, and Michael M Bronstein. Improving graph neural network expressivity via subgraph isomorphism counting. *Transactions on Pattern Analysis and Machine Intelligence*, 45(1):657–668, 2022. 3

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Caglar Gulcehre, Francis Song, Andrew Ballard, Justin Gilmer, George Dahl, Ashish Vaswani, Kelsey Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matt Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks. *ArXiv 1806.01261*, 2018. 3

Derek Lim, Joshua David Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. In *ICLR*, 2023. 3

Gaurav Rattan and Tim Seppelt. Weisfeiler-leman and graph spectra. In *SODA*, 2023. 3

Max Horn, Edward De Brouwer, Michael Moor, Yves Moreau, Bastian Rieck, and Karsten Borgwardt. Topological graph neural networks. In *ICLR*, 2022. 3

John J Irwin and Brian K Shoichet. ZINC–a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*, 45(1):177–182, 2005. 3

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. In *NeurIPS*, 2020. 3

Beatrice Bevilacqua, Fabrizio Frasca, Derek Lim, Balasubramaniam Srinivasan, Chen Cai, Gopinath Balamurugan, Michael M. Bronstein, and Haggai Maron. Equivariant subgraph aggregation networks. In *ICLR*, 2022. 3

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017. 6

Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yu Guang Wang, Pietro Liò, Guido Montúfar, and Michael Bronstein. Weisfeiler and Lehman go cellular: CW networks. In *NeurIPS*, 2021. 6

Fabian Jogl, Maximilian Thiessen, and Thomas Gärtner. Weisfeiler and Leman return with graph transformations. In *International Workshop on Mining and Learning with Graphs at ECMLPKDD*, 2022. 6

Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. Moleculenet: A benchmark for molecular machine learning, 2018. 6

Dennis H. Rouvray and R. Bruce King. *Topology in Chemistry: Discrete Mathematics of Molecules*. Horwood series in chemical science. Elsevier, 2002. 8

Jack Edmonds. Paths, trees, and flowers. *Canadian Journal of Mathematics*, 17:449–467, 1965. 8

Haruo Hosoya. Topological index. a newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bulletin of the Chemical Society of Japan*, 44: 2332–2339, 1971. 8

Mark Jerrum. Two-dimensional monomer-dimer systems are computationally intractable. *Journal of Statistical Physics*, 48(1):121–134, 1987. 8

Danail Bonchev. *Chemical Graph Theory: Introduction and Fundamentals*. Chemical Graph Theory. Taylor & Francis, 1991. 8

J. A. Makowsky, Udi Rotics, Ilya Averbouch, and Benny Godlin. Computing graph polynomials on graphs of bounded clique-width. In *Graph-Theoretic Concepts in Computer Science*, 2006. 8

M. R. Garey and D. S. Johnson. "Strong" NP-completeness results: Motivation, examples, and implications. *Journal of the ACM*, 25(3):499–508, 1978. 8

Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2): 298–305, 1973. 8

Arthur T. White. Chapter 6 - imbedding problems in graph theory. In *Graphs of Groups on Surfaces*, volume 188 of *North-Holland Mathematics Studies*, pages 49–72. North-Holland, 2001. 8

I. Gutman and N. Trinajstić. Graph theory and molecular orbitals. total ϕ-electron energy of alternant hydrocarbons. *Chemical Physics Letters*, 17(4):535–538, 1972. 9

Batmend Horoldagva and Kinkar Chandra Das. On zagreb indices of graphs. *ArXiv 2305.05878*, 2023. 9

Hongbo Hua. Zagreb m1 index, independence number and connectivity in graphs. *MATCH Communications in Mathematical and in Computer Chemistry*, 60:45–56, 2008. 9

E.R. Scheinerman and D.H. Ullman. *Fractional Graph Theory: A Rational Approach to the Theory of Graphs*. Dover books on mathematics. Dover Publications, 2011. 9

V. Arvind, Frank Fuhlbrück, Johannes Köbler, and Oleg Verbitsky. On weisfeiler-leman invariance: Subgraph counts and related graph properties. *Journal of Computer and System Sciences*, 113: 42–59, 2020. 9

A Experimental Details

In this section we add results for two additional GNNs: Graph Convolutional Network (GCN) [Kipf and Welling, 2017] and CW Network (CWN) [Bodnar et al., 2021]. CWN is strictly more expressive than 1-WL and implemented through a graph transformation [Jogl et al., 2022] combined with GIN. We tune hyperparameters for GCN exactly as we tune them for GIN. We do not tune the hyperparameters for CWN and instead use hyperparameters which we know will give us a strong baseline.

In this section we also add results for an additional dataset `QM9` [Wu et al., 2018], a large molecular dataset (containing approximately 130000 molecules) with 19 regression targets, corresponding to different quantum chemical properties of small organic molecules. For the experimental setup, we use a batch size of 128 and a total number of epochs of 150. The stopping criteria is similar to the

Table 3: Hyperparameter grids for different datasets.

Parameter	GIN GCN On ZINC	GIN GCN On MOLHIV
Message passing layers	1, 2, 3, 4, 5	1, 2, 3, 4, 5
Final MLP layers	2	2
Pooling operation	mean, sum	mean, sum
Embedding dimension	64, 256	64, 256
Dropout rate	0, 0.5	0, 0.5
Batch size	128	32
Epochs	1000	100

Table 4: Experimental results of GIN and GCN on QM9 dataset.

	QM9 (MAE ↓)		QM9 (MAE ↓)	
Global feature f	GIN	(GIN, f)	GCN	(GCN, f)
Wiener	0.0614 ± 0.0020	$\mathbf{0.0598 \pm 0.0016}$	0.0771 ± 0.0016	$\mathbf{0.0749 \pm 0.0017}$
Circuit rank	0.0599 ± 0.0012	0.0591 ± 0.0011	0.0764 ± 0.0016	0.0756 ± 0.0015
Independence	0.0609 ± 0.0005	$\mathbf{0.0598 \pm 0.0005}$	0.0766 ± 0.0018	0.0756 ± 0.0014
Hosoya	0.0598 ± 0.0017	0.0589 ± 0.0015	0.0775 ± 0.0016	0.0765 ± 0.0014
Max. matching	0.0612 ± 0.0014	$\mathbf{0.0598 \pm 0.0010}$	0.0770 ± 0.0014	$\mathbf{0.0756 \pm 0.0012}$
Second eigenv.	0.0613 ± 0.0011	$\mathbf{0.0593 \pm 0.0008}$	0.0773 ± 0.0021	$\mathbf{0.0729 \pm 0.0018}$
Spectral radius	0.0609 ± 0.0014	0.0598 ± 0.0012	0.0763 ± 0.0014	0.0754 ± 0.0012
Zagreb M_1	0.0609 ± 0.0016	0.0598 ± 0.0015	0.0764 ± 0.0009	$\mathbf{0.0752 \pm 0.0007}$
Zagreb M_2	0.0620 ± 0.0019	0.0609 ± 0.0018	0.0772 ± 0.0010	$\mathbf{0.0760 \pm 0.0010}$
All features	0.0604 ± 0.0018	$\mathbf{0.0576 \pm 0.0013}$	0.0759 ± 0.0011	$\mathbf{0.0708 \pm 0.0010}$
Constant (1's)	0.0613 ± 0.0012	0.0604 ± 0.0011	0.0776 ± 0.0019	0.0765 ± 0.0014

ZINC setup, we only reduce the patience of the learning rate to 5. There are 19 regression tasks for QM9, we calculated the Mean Absolute Error (MAE) for each target individually and average them.

Furthermore, we also add the results of new ablation study experiments, where we introduced an one-dimensional global feature containing only values of 1 for each graph.

Table 3 shows the used hyperparameter grids. Tables 4, 5 and 6 show additional experimental results.

Table 5: Experimental results for GCN and CWN on ZINCdataset.

	ZINC (MAE ↓)		ZINC (MAE ↓)	
Global feature f	GCN	(GCN, f)	CWN	(CWN, f)
Wiener	0.217 ± 0.010	$\mathbf{0.206 \pm 0.005}$	0.128 ± 0.004	$\mathbf{0.103 \pm 0.003}$
Circuit rank	0.215 ± 0.007	0.214 ± 0.008	0.125 ± 0.005	$\mathbf{0.109 \pm 0.002}$
Independence	0.221 ± 0.013	$\mathbf{0.207 \pm 0.007}$	0.125 ± 0.007	$\mathbf{0.106 \pm 0.002}$
Hosoya	0.218 ± 0.009	$\mathbf{0.210 \pm 0.006}$	0.128 ± 0.004	$\mathbf{0.120 \pm 0.002}$
Max. matching	0.222 ± 0.011	$\mathbf{0.212 \pm 0.008}$	0.125 ± 0.004	$\mathbf{0.105 \pm 0.002}$
Second eigenv.	0.211 ± 0.008	0.204 ± 0.008	0.125 ± 0.003	$\mathbf{0.106 \pm 0.002}$
Spectral radius	0.220 ± 0.016	0.215 ± 0.012	0.128 ± 0.004	$\mathbf{0.121 \pm 0.003}$
Zagreb M_1	0.213 ± 0.009	0.208 ± 0.008	0.126 ± 0.003	$\mathbf{0.104 \pm 0.001}$
Zagreb M_2	0.218 ± 0.008	0.211 ± 0.006	0.128 ± 0.002	$\mathbf{0.105 \pm 0.001}$
All features	0.221 ± 0.004	$\mathbf{0.203 \pm 0.005}$	0.127 ± 0.005	$\mathbf{0.101 \pm 0.002}$
Constant (1's)	0.218 ± 0.007	0.213 ± 0.007	0.126 ± 0.005	0.122 ± 0.004

Table 6: Experimental results of CWN on MOLHIV dataset.

	MOLHIV (ROC-AUC ↑)	
Global feature f	CWN	(CWN, f)
Wiener	0.7781 ± 0.016	$\mathbf{0.7895 \pm 0.011}$
Circuit rank	0.7838 ± 0.014	0.7890 ± 0.014
Independence	0.7824 ± 0.020	0.7893 ± 0.023
Hosoya	—	
Max. matching	0.7850 ± 0.014	0.7930 ± 0.015
Second eigenv.	0.7880 ± 0.007	0.7899 ± 0.007
Spectral radius	0.7829 ± 0.012	$\mathbf{0.7945 \pm 0.007}$
Zagreb M_1	0.7801 ± 0.012	0.7911 ± 0.012
Zagreb M_2	0.7902 ± 0.014	0.7983 ± 0.009
All features	0.7830 ± 0.015	0.7885 ± 0.012
Constant (1's)	0.7840 ± 0.010	0.7912 ± 0.010

B Definitions of Global Graph Features

In this section, we give more details on our selected global graph features and discuss whether they improve expressivity compared to 1-WL.

Wiener Index represents the sum of the lengths of the shortest paths between all pairs of vertices in a graph [Rouvray and King, 2002]:

$$W = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} D_{ij},$$

where D is the shortest path distance matrix.

Maximum Matching. A matching is a set of edges with no common vertices. We take the cardinality of a maximum matching as a global parameter. It can be computed by the algorithm of Edmonds [1965].

Hosoya Index is the total number of matchings in a graph. As mentioned by Hosoya [1971], it is correlated to the boiling points of alkane isomers. Jerrum [1987] showed that this is a #P-complete problem for planar graphs, but for relatively smaller graphs it can be computed using the coefficients of the matching polynomial evaluated at 1 [Bonchev, 1991]. Using the latter, the Hosoya index can also be calculated in polynomial time for graphs of bounded clique-width [Makowsky et al., 2006].

$$Z = \sum_{k=0}^{n} |a_k|,$$

where n is the number of vertices in the graph, a_k is the kth coefficient of the matching polynomial and $|x|$ represents the absolute value of x.

Independence Number $\alpha(G)$ is the cardinality of the largest vertex set in a graph, such that no two vertices in this set represent an edge. Computing such a set is NP-hard [Garey and Johnson, 1978]. This number is equal to the largest exponent in the independence polynomial of graph G.

Second Eigenvalue For a given graph G, the Laplacian matrix is equal to

$$L = D - A,$$

where D is a diagonal matrix whose elements are the degrees of each node and A is the adjacency matrix. The value of the second smallest eigenvalue of matrix L, λ_2, reflects how well connected the graph is [Fiedler, 1973]. Smaller values correspond to less connected graphs, while for fully connected graphs, the value is relatively larger.

Circuit Rank is the smallest number of edges that must be removed from a graph so that are no graph cycles remaining. Also known as the first Betti number [White, 2001], it is given by:

$$\gamma = m - n + c,$$

where m is the number of edges in the graph, n is the number of nodes, and c is the number of connected components.

Spectral Radius is the largest absolute value among the graph's eigenvalues of the adjacency matrix. For a graph G, it is given by:

$$\rho(G) = \max_{1 \leq i \leq n} |\lambda_i|,$$

where λ_i are the eigenvalues of the adjacency matrix of the graph G.

Zagreb M_1 / M_2 The Zagreb indices were first introduced by Gutman and Trinajstić [1972] and are defined based on the values of the nodes' degrees. The first index is equal to the sum of squares of the degrees of the nodes, while the second is equal to the sum of the products of the degrees of pairs of adjacent nodes of the graph [Horoldagva and Das, 2023; Hua, 2008]. For a graph G, with node set $V(G)$ and edge set $E(G)$, they are given by:

$$M_1(G) = \sum_{i \in V(G)} d_i^2,$$
$$M_2(G) = \sum_{i,j \in E(G)} d_i d_j,$$

where d_i represents the degree of node i.

C Expressivity Results

As mentioned in 2, an architecture is said to be (strictly) more expressive than 1-WL if it can distinguish every pair of graphs that 1-WL can distinguish and at least one pair of graphs that 1-WL cannot distinguish. In this context, we propose an investigation designed to measure the expressivity of the proposed global features in regards to the 1-WL test.

For the experiments, we selected three pairs of graphs that are indistinguishable by the 1-WL test. For each pair, a comparison is made between the value of the global feature applied to each graph. A global feature is then said to be more expressive than 1-WL if it can distinguish between two graphs, having different values for each of them, that the 1-WL cannot distinguish.

Global features not increasing expressivity. As proven by Scheinerman and Ullman [2011], if two graphs are isomorphic, then the maximum eigenvalues of their adjacency matrices (spectral radius) have the same value. Arvind et al. [2020] shows that two graphs that are indistinguishable by 1-WL (color refinement) have the same degree sequence. The Zagreb M_1 and M_2 indices are computed using the degree sequence of the graphs, so their values are the same for a pair of graphs that are indistinguishable by 1-WL. Hence, we do not include them in the following exploration on pairs of graphs that are not distinguishable by 1-WL.

For a pair of connected graphs circuit rank γ will have the same value (number of connected components stays the same). We show for the third pair of graphs 4, G_5 (connected) and G_6 (disconnected), that the circuit rank is more expressive than 1-WL.

In the case of disconnected graphs, we set the value for Wiener index to ∞, as there is no shortest path between all pairs of vertices.

Table 7: Global feature values for G_1 and G_2

	G_1	G_2	**increases expressivity**
Wiener	25	27	✓
Hosoya	22	20	✓
Circuit rank	2	2	✗
Independence	3	2	✓
Max. matching	3	3	✗
Second eigenv.	1	0.438	✓

(a) G_1 **(b)** G_2

Figure 2: First pair of graphs, G_1 and G_2, indistinguishable by 1-WL

Table 8: Global feature values for G_3 and G_4

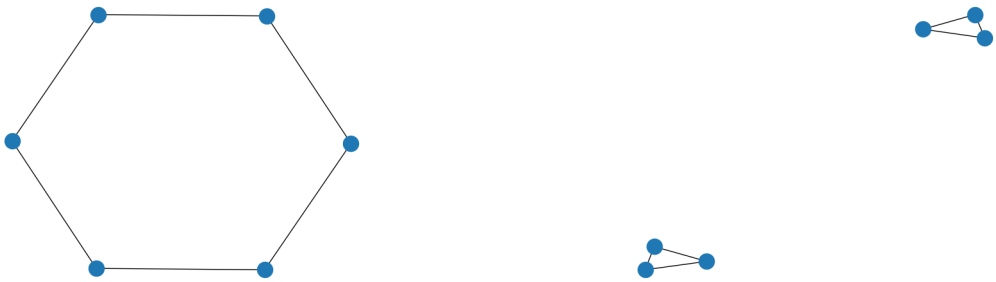
	G_3	G_4	**increases expressivity**
Wiener	281	261	✓
Hosoya	1280	1480	✓
Circuit rank	4	4	✗
Independence	5	6	✓
Max. matching	6	7	✓
Second eigenv.	0.14	0.14	✗

(a) G_3 **(b)** G_4

Figure 3: Second pair of graphs, G_3 and G_4, indistinguishable by 1-WL

Table 9: Global feature values for G_5 and G_6

	G_5	G_6	**increases expressivity**
Wiener	27	∞	✓
Hosoya	18	16	✓
Circuit rank	1	2	✓
Independence	3	2	✓
Max. matching	3	2	✓
Second eigenv.	1	0	✓



(a) G_5 **(b)** G_6

Figure 4: Third pair of graphs, G_5 and G_6, indistinguishable by 1-WL