Benchmarking Graph Neural Networks

Xavier Bresson



School of Computer Science and Engineering
Data Science and AI Research Centre
Nanyang Technological University (NTU), Singapore

NATIONAL RESEARCH FOUNDATION
PRIME MINISTER'S OFFICE
SINGAPORE

Summer School on Data Science (SSDS 2020)

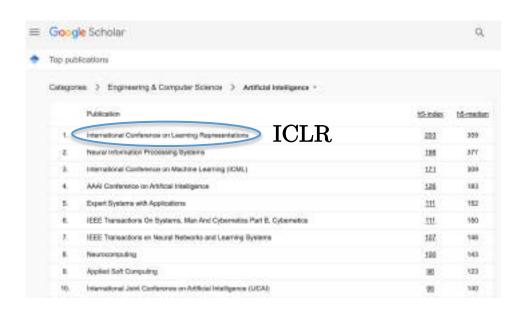
Sven Loncaric, Tomislav Smuc, Vinko Zlatic, Tomislav Lipic

Sept 11th 2020



Graph Neural Networks (GNNs)

- GNNs have become the standard toolkit for analyzing and learning from data on graphs.
- Hottest machine learning topics in 2020.







Graph Neural Networks (GNNs)

Applicative domains:

- Chemistry^[1,2] (generate new drugs and materials)
- Physics^[3,4,5] (learn and accelerate physics)
- Recommender systems^[6,7] (leverage consumer-product choices)
- Social sciences^[8,9] (predict future collaborators, identify fake news)
- Knowledge graphs^[10,11] (reasoning with entity-relationship)
- Neuroscience^[12] (understanding brain mechanisms and neuro-degenerative diseases)
- Computer Vision^[13] (scene understanding for visual reasoning)
- Natural Language Processing^[14] (transformers)
- Combinatorial Optimization^[15,16] (better/faster approximated solutions to NP-hard problems)
 - [1] Duvenaud, Maclaurin, Iparraguirre, Bombarell, Hirzel, Aspuru-Guzik, Adams, Convolutional networks on graphs for learning molecular fingerprints, 2015 [2] Gilmer, Schoenholz, Riley, Vinyals, Dahl, Neural message passing for quantum chemistry, 2017
 - [3] Battaglia, Pascanu, Lai, Rezende, Danilo, Interaction networks for learning about objects, relations and physics, 2016

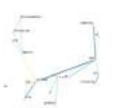
 - [4] Cranmer, Xu. Battaglia, Ho. Learning symbolic physics with graph networks, 2019
 - [5] Sanchez-Gonzalez, Godwin, Pfaff, Ying, Leskovec, Battaglia, Learning to simulate complex physics with graph networks, 2020
 - [6] Monti, Bronstein, Bresson, Geometric matrix completion with recurrent multi-graph neural networks, 2017
 - [7] Ying, He, Chen, Eksombatchai, Hamilton, Leskovec, Graph convolutional neural networks for web-scale recommender systems, 2018
 - [8] Kipf, Welling, Semi-supervised classification with graph convolutional networks, 2017
 - [9] Monti, Frasca, Evnard, Mannion, Bronstein, Fake news detection on social media using geometric deep learning, 2019
 - [10] Schlichtkrull, Kipf, Bloem, VanDenBerg, Titov, Welling, Modeling relational data with graph convolutional networks, 2018
 - [11] Chami, Wolf, Juan Sala Ravi, Re, Low-dimensional hyperbolic knowledge graph embeddings, 2020
 - [12] Parisot, Ktena, Ferrante, Lee, Guerrero, Glocker, Rueckert, Disease prediction using graph networks: Application to Autism Spectrum Disorder and Alzheimer's disease, 2018
 - [13] Johnson, Gupta, Fei-Fei, Image generation from scene graphs, 2018
 - [14] Vaswani, Shazeer, Parmar, Uszkoreit, Jones, N. Gomez, Kaiser, Polosukhin, Attention is all you need, 2017
 - [15] Vinyals, Fortunato, Jaitly, Pointer, 2015
 - [16] Bello, Pham, Le, Norouzi, Bengio, Neural combinatorial optimization with reinforcement learning, 2016

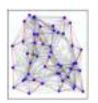






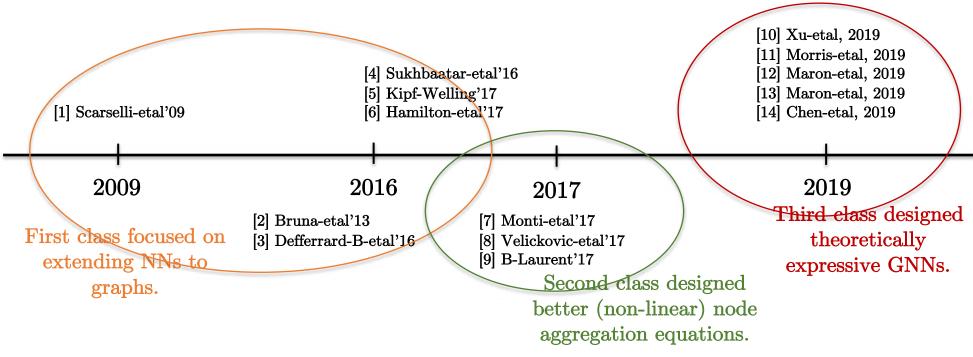








A Decade of GNNs



Developing powerful GNNs for real-world adoption of graph deep learning.

- [1] Scarselli, Gori, Tsoi, Hagenbuchner, Monfardini, The Graph Neural Network Model, 2009
- [2] Bruna, Zaremba, Szlam, LeCun, Spectral networks and locally connected networks on graphs, 2013
- [3] Defferrard, Bresson, Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, 2016
- [4] Sukhbaatar, Szlam, Fergus, Learning multiagent communication with backpropagation, 2016
- [5] Kipf, Welling, Semi-supervised classification with graph convolutional networks, 2017
- [6] Hamilton, Ying, Leskovec, Inductive representation learning on large graphs, 2017
- [7] Monti, Boscaini, Masci, Rodola, Svoboda, Bronstein, Geometric deep learning on graphs using mixture model cnns, 2017
- [8] Velickovic, Cucurull, Casanova, Romero, Lio, Bengio, Graph attention networks, 2017
- [9] Bresson, Laurent, Residual gated graph convnets, 2017
- [10] Xu, Hu, Leskovec, Jegelka, How powerful are graph neural networks?, 2019
- [11] Morris, Ritzert, Fey, Hamilton, Lenssen, Rattan, Grohe, Weisfeiler and leman go neural: Higher-order graph networks, 2019
- [12] Maron, Ben-Hamu, Shamir, Lipman, Invariant and equivariant graph networks, 2019
- [13] Maron, Ben-Hamu, Serviansky, Lipman, Provably powerful graph networks, 2019
- [14] Chen, Villar, Chen, Bruna, On the equivalence graph isomorphism testing and function approximation with gnns, 2019

Tracking Progress

- The field has grown extensively in the last three years ©
- Unfortunately, evaluating the effectiveness of new architectures/ideas has become difficult for two major reasons :
 - We have been evaluated progress on small datasets such as $Cora^{[1]}$, $Citeseer^{[2]}$ and $TU^{[3]}$.
 - Cora is a single graph of 2.7K nodes, TU-IMDB has 1.5K graphs with 13 nodes on average, and TU-MUTAG has 188 molecules with 18 nodes.
 - Simple or graph-agnostic architectures provide statistically same performance as complex architectures^[4,5].
 - We have not rigorously enforced standardized experimental settings for fair comparisons between models^[6].
- It has become critical to solve these issues.

^[1] McCallum, Nigam, Rennie, Seymore, Automating the construction of internet portals with machine learning, 2000

^[2] Getoor, Link-based classification, 2005

^[3] Kersting, Kriege, Morris, Mutzel, Neumann, Benchmark data sets for graph kernels, 2020

^[4] Hoang, Maehara, Revisiting graph neural networks, 2019

^[5] Chen, Bian, Sun, Are powerful graph neural nets necessary? a dissection on graph classification, 2019

^[6] Errica, Podda, Bacciu, Micheli, A fair comparison of graph neural networks for graph classification, 2019

Benchmarking GNNs

- Our motivation: Identify and quantify what types of architectures, first principles are universal, generalizable, and scalable to larger and more challenging datasets.
- Benchmarking has been beneficial for driving progress, identifying essential ideas, and solving domain-specific problems^[1].
 - The 2012 ImageNet challenge^[2] has provided a benchmark dataset that has triggered the deep learning revolution.
- Challenges:
 - Developing a rigorous experimental setting for fair comparisons, and being reproducible,
 - Using datasets that can statistically separate model performance,
 - Benchmarking distinct fundamental graph tasks.

^[1] Weber, Saelens, Cannoodt, Soneson, Hapfelmeier, Gardner, Boulesteix, Saeys, Robinson, Essential guidelines for computational method benchmarking, 2019

Outline

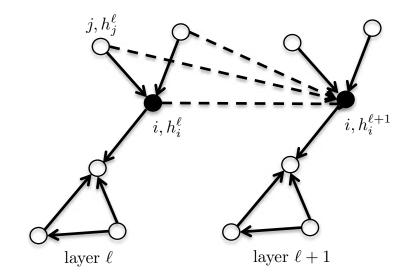
- Motivation
- Message-Passing GCNs
- Weisfeiler-Lehman GNNs
- Graph-Agnostic GNNs
- Datasets
- Infrastructure and Experimental Setting
- Benchmarking Results
- Laplacian Positional Encodings
- Link Prediction with Edge Representation
- Conclusion

Outline

- Motivation
- Message-Passing GCNs
- Weisfeiler-Lehman GNNs
- Graph-Agnostic GNNs
- Datasets
- Infrastructure and Experimental Setting
- Benchmarking Results
- Laplacian Positional Encodings
- Link Prediction with Edge Representation
- Conclusion

MP-GCNs

- Message-passing Graph Convolutional Networks
- What are the minimal inner structures to design a messagepassing node aggregation function f_{GCN} ?
 - Invariant by node permutation (equivariant/invariant layers^[1]).
 - Independent of graph size n and neighborhood size.
 - Locality (local reception field only neighbors are considered).
 - Graph convolution operator (weight sharing across graph).
 - Linear complexity O(E), E being the number of edges (reduces to O(n) for sparse graphs).



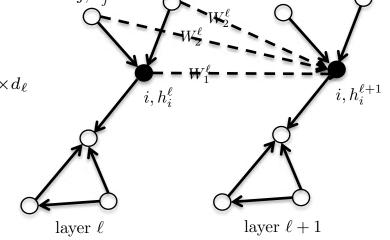
$$h_i^{\ell+1} = f_{\text{GCN}}\left(h_i^{\ell}, \{h_j^{\ell} : j \in \mathcal{N}_i\}\right)$$

Isotropic GCNs

• We consider a class of isotropic GCNs when the node update equation treats every "edge direction" equally, i.e. each neighbor contributes equally to the update of the central node:

$$\begin{split} \text{ReLU} & \sum_{h_i^{\ell+1} = \sigma \left(W_1^{\ell} \ h_i^{\ell} + \sum_{j \in \mathcal{N}_i} W_2^{\ell} h_j^{\ell} \right), \text{/max} \\ h^{\ell+1} \in \mathbb{R}^{n \times d_{\ell+1}}, \ h^{\ell} \in \mathbb{R}^{n \times d_{\ell}}, \ W_{1,2}^{\ell} \in \mathbb{R}^{d_{\ell+1} \times d_{\ell}} \end{split}$$

- Models :
 - ullet Vanilla GCNs $^{[1,2]}$
 - GraphSage^[3]
 - ChebNets^[4]



Isotropic GCNs are limited to learn template weights for the central node h_i (weight W_1) and the neighbors h_j (weight W_2).

^[1] Sukhbaatar, Szlam, Fergus, Learning multiagent communication with backpropagation, 2016

^[2] Kipf, Welling, Semi-supervised classification with graph convolutional networks, 2017

^[3] Hamilton, Ying, Leskovec, Inductive representation learning on large graphs, 2017

^[4] Defferrard, Bresson, Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, 2016

Anisotropic GCNs

- Standard ConvNets^[1] produce anisotropic filters because Euclidean grids have directional structure (up, down, left, right).
- 11
- GCNs such as vanilla GCNs, GraphSage, ChebNets compute isotropic filters as there is no notion of direction on arbitrary graphs.
- How to get anisotropy back?
 - Natural edge features^[2,3] if available (e.g. different bond connections between atoms in molecular graphs or distinct edge attributes in knowledge graphs).
 - Learn anisotropy with a mechanism invariant by index permutation and can treat neighbors differently:
 - MoNets^[4] with edge degrees
 - GatedGCNs [5] with edge gates
 - GAT^[6] with attention mechanism^[7]
- [1] LeCun Bottou Bengio, Haffner, Gradient-based learning applied to document recognition, 1998
- [2] Gilmer, Schoenholz, Riley, Vinyals, Dahl, Neural message passing for quantum chemistry, 2017
- [3] Bresson, Laurent, A Two-Step Graph Convolutional Decoder for Molecule Generation, 2019
- [4] Monti, Boscaini, Masci, Rodolà, J. Svoboda, M. Bronstein, Geometric deep learning on graphs and manifolds using mixture model CNNs, 2016
- [5] Bresson, Laurent, Residual gated graph convnets, 2017
- [6] Velickovic, Cucurull, Casanova, Romero, Lio, Bengio, Graph Attention Networks, 2018
- [7] Bahdanau, Cho, Bengio, Neural machine translation by jointly learning to align and translate, 2015

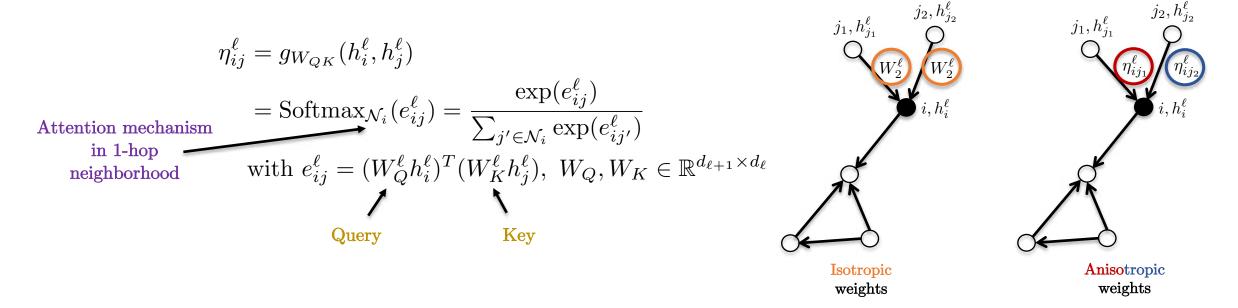
0

Anisotropic GCNs

• When the update equation treats every edge direction differently, we instantiate anisotropic GCNs with an aggregation equation of the form:

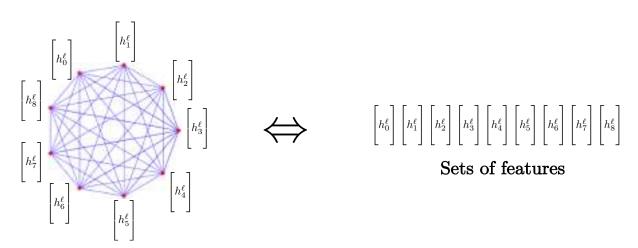
$$h_i^{\ell+1} = \sigma \Big(W_1^{\ell} \ h_i^{\ell} + \sum_{j \in \mathcal{N}_i} \eta_{ij}^{\ell} W_2^{\ell} h_j^{\ell} \Big), \ h^{\ell+1} \in \mathbb{R}^{n \times d_{\ell+1}}, \ h^{\ell} \in \mathbb{R}^{n \times d_{\ell}}, \ W_{1,2}^{\ell} \in \mathbb{R}^{d_{\ell+1} \times d_{\ell}},$$

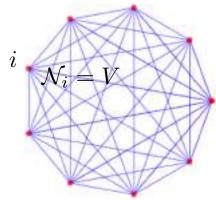
where $\eta_{ij}^{l} = g_W(h_i^{l}, h_j^{l})$ is a parameterized function which weights are learned during training.



Transformers

- Transformers is a special case of GCNs when the graph is fully connected.
 - The neighborhood \mathcal{N}_i is the whole graph.
- What does it mean to have a fully connected graph?
 - There is no particular graph structure that can be used. It becomes useless to talk about graphs (as each data point is connected to all other points).
 - It would be better to talk about sets.
 - Transformers are set neural networks, which analyze bags of features.





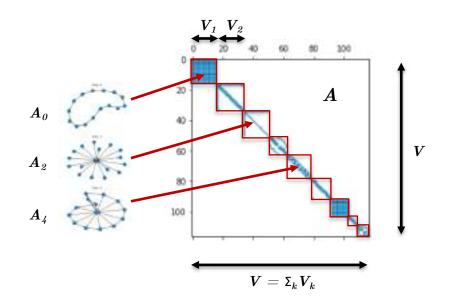
Fully connected graph

Batch Normalization and Residual Connections

- GCNs benefit from $BN^{[1,2]}$ and $RC^{[1,3]}$:
 - Speed up learning process.
 - Improve performance with better generalization.

$$h_i^{\ell+1} = h_i^{\ell} + \sigma \left(BN \left(\hat{h}_i^{\ell+1} \right) \right)$$
$$\hat{h}_i^{\ell+1} = f_{GCN} \left(h_i^{\ell}, \{ h_j^{\ell} : j \in \mathcal{N}_i \} \right)$$

- How to batch graphs of different sizes?
 - For images of size $B \times V_x \times V_y \times d$, BN is performed along the batch dimension, i.e. dim=0.
 - For graphs, a (big) sparse block diagonal matrix A of size $V \times d$ is first built from K matrices A_k and BN is carried out along the node direction, i.e. dim=0.



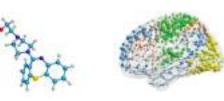
^[1] Bresson, Laurent, Residual gated graph convnets, 2017

^[2] Ioffe, Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015

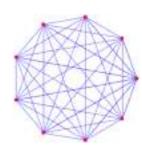
^[3] Li, Muller, Qian, Delgadillo, Abualshour, Thabet, Ghanem, Deepgcns: Making gcns go as deep as cnns, 2019

Sparsity and Local Computations

- GCNs leverage graph sparsity :
 - Sparsity is a good inductive bias for generalization.
 - Sparse vs full graphs.







• Local computations:

- Node update equation is local, only depending on neighborhood \mathcal{N}_i of node i, and independent of graph size n, making the space/time complexity O(n) for sparse graphs.
- GCNs are highly parallelizable on GPUs, and sparse matrix multiplications are efficiently implemented via GNN libraries s.a. DGL^[1], PyG^[2], Spektral^[3].

Sparse graphs

Fully connected graph

$$h_i^{\ell+1} = f_{\text{GCN}}\left(h_i^{\ell}, \{h_j^{\ell} : j \in \mathcal{N}_i\}\right)$$







^[1] Wang-etal, Deep graph library: Towards efficient and scalable deep learning on graphs, 2019

^[2] Fey, Lenssen, Fast graph representation learning with pytorch geometric, 2019

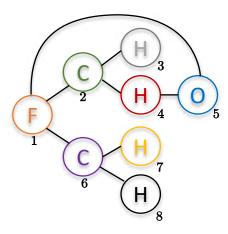
^[3] Grattarola, Alippi, Graph Neural Networks in TensorFlow and Keras with Spektral, 2020

Outline

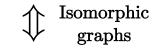
- Motivation
- Message-Passing GCNs
- Weisfeiler-Lehman GNNs
- Graph-Agnostic GNNs
- Datasets
- Infrastructure and Experimental Setting
- Benchmarking Results
- Laplacian Positional Encodings
- Link Prediction with Edge Representation
- Conclusion

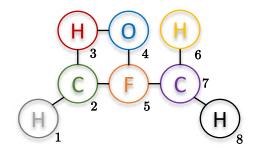
Weisfeiler-Lehman GNNs

- Motivation: Characterize expressivity power of GNNs with graph isomorphism.
- Graph isomorphism: Two graphs are isomorphic if there exists an index permutation between the nodes that preserves node adjacencies.
- Determining whether two graphs are isomorphic is NPintermediate: It is not known if a polynomial time algorithm exists, or the problem is NP-hard.
- Weisfeiler-Lehman (WL) proposed an algorithm^[1] to test if two graphs are not isomorphic. However, the WL test is not sufficient to guarantee that two graphs are isomorphic.



Graph 1



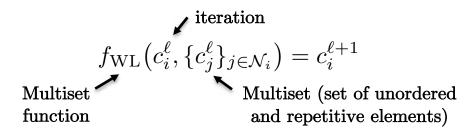


Graph 2

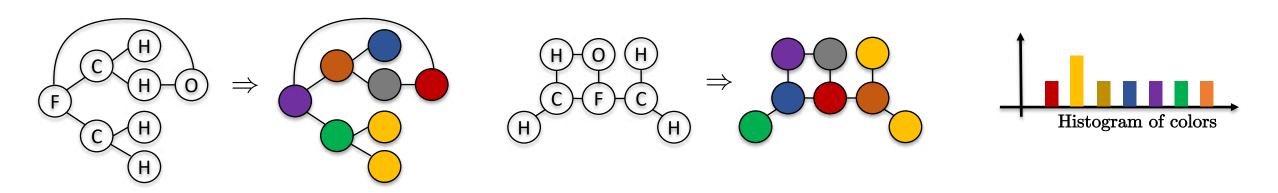
^[1] B Weisfeiler, A Lehman, A reduction of a graph to a canonical form and an algebra arising during this reduction, 1968

WL Test^[1]

• Idea: Design an injective "coloring" function f_{WL} :



- WL algorithm iteratively applies function f_{WL} until no new colors are created.
- This produces a canonical representation of a graph as a histogram of colors.
- If two graphs have different color histograms (up to all color permutations) then the two graphs are guaranteed to be non-isomorphic.



[1] B Weisfeiler, A Lehman, A reduction of a graph to a canonical form and an algebra arising during this reduction, 1968

Graph Isomorphism Networks (GINs)^[1]

- Goal: Design a GNN as expressive as the WL test to distinguish non-isomorphic graphs.
- Node aggregation of the form:

$$f_{\mathrm{NN}}\left(h_{i}^{\ell}, \{h_{j}^{\ell}\}_{j \in \mathcal{N}_{i}}\right) = (1 + \varepsilon)g(h_{i}^{\ell}) + \sum_{j \in \mathcal{N}_{i}} g(h_{j}^{\ell})$$
irrational injective sum

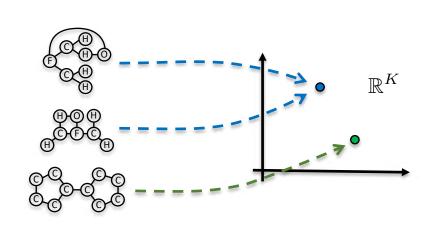
• It is difficult to design an analytical injective function \Rightarrow MLP is used to approximate g (existence guaranteed by the universal approximation theorem):

$$h_i^{\ell+1} = f_{GIN}(h_i^{\ell}, \{h_j^{\ell}\}_{j \in \mathcal{N}_i}) = MLP^{\ell}((1+\varepsilon)h_i^{\ell} + \sum_{j \in \mathcal{N}_i} h_j^{\ell})$$

• Graph readout function must also be injective :

$$h_{\mathcal{G}} = \mathrm{MLP}\Big(\sum_{i \in V} h_i^L\Big) \in \mathbb{R}^K$$
 Sum function

• Pioneer work with^[2] on GNN expressivity.



^[1] Xu, Hu, Leskovec, Jegelka, How powerful are graph neural networks?, 2019

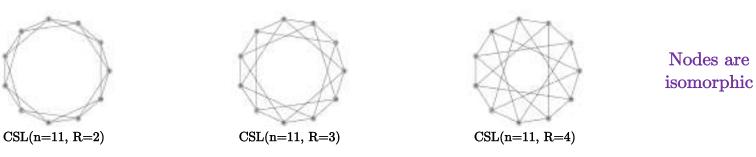
^[2] Morris, Ritzert, Fey, Hamilton, Lenssen, Rattan, Grohe, Weisfeiler and leman go neural: Higher-order graph networks, 2019

Limitation

- WL test^[1] is not a sufficient condition, it can fail to differentiate non-isomorphic graphs:
 - **Example** of two graphs with same color signatures but not isomorphic:



• Another example w/ Circular Skip Length (CSL) graphs^[2] (a circular graph with skip connections for each node to R-hop nodes):



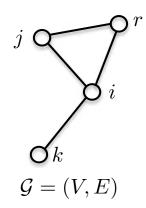
• Can we improve the expressivity of the original WL test?

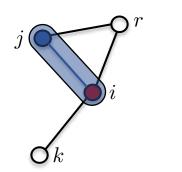
^[1] B Weisfeiler, A Lehman, A reduction of a graph to a canonical form and an algebra arising during this reduction, 1968

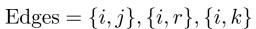
^[2] Murphy, Srinivasan, Rao, Ribeiro, Relational pooling for graph representations, 2019

k-WL Test

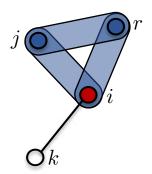
- Can we design better WL tests?
 - Original WL test uses 2-tuple of nodes to produce colors. To produce more colors and improve expressivity of WL tests, k-tuple of nodes with $k \ge 3$ can be used (higher-order interactions between nodes).







2-tuple of nodes can distinguish nonisomorphic graphs with the 1-WL/2-WL tests.



$$\begin{aligned} \text{Hyper-edges} = & \{i, j, r\}, \\ & \{i, k, r\}, \{i, j, k\} \end{aligned}$$

3-tuple of nodes can distinguish nonisomorphic graphs with the 3-WL test.

Equivariant GNNs^[1]

How to design GNNs with the same expressivity power as the k-WL test?

Permutation

• Let us define a k-order equivariant GNNs:

a k-order equivariant GNNs:
$$f_{W}(P \circ h) = P \circ f_{W}(h)$$
 Equivariant linear layers
$$y = m_{W^{L+1}} \circ g_{W^{L}} \circ \sigma \circ f_{W^{L-1}} \circ \sigma \circ f_{W^{L-2}} \circ ... \sigma \circ f_{W^{0}} \circ h_{0}$$
 with $k = \max_{\ell \in [0, L-1]} k_{\ell}$
$$g_{W}(P \circ h) = g_{W}(h) \in \mathbb{R}^{K}, K \geq 1$$
 Invariant linear layer and
$$f_{W^{\ell}} : \mathbb{R}^{n^{k_{\ell}} \times d_{\ell}} \to \mathbb{R}^{n^{k_{\ell+1}} \times d_{\ell+1}}$$

- Theorem^[1]: There exist k-order E-GNNs that can distinguish non-isomorphic graphs with the k-WL test.
 - However, k-order E-GNNs requite $O(n^k)$ memory/speed complexities.
 - Note that we need at least k=3 (hence $O(n^3)$) to be more powerful than GINs as GINs have the discriminative power of the 1-WL/2-WL tests^[2].

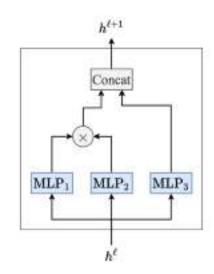
^[1] Maron, Ben-Hamu, Shamir, Lipman, Invariant and equivariant graph networks, 2019

^[2] Cai, Furer, Immerman, An optimal lower bound on the number of variables for graph identification, 1992

3-WL GNNs^[1]

- How to design GNNs that are 3-WL expressive but do not require $O(n^3)$ memory/speed complexities?
- 3-WL GNNs: To achieve higher interactions between nodes, it is sufficient to multiply second-order tensors feature-wise.
 - Theorem^[1]: There exist β -WL GNNs as expressive as the β -WL test.
 - Memory is quadratic $O(n^2)$ but matrix multiplication implies $O(n^3)$ speed.
 - Also, matrix multiplication densifies sparse matrix.
 - This is the simplest, most scalable and most expressive GNNs in terms of 3-WL test.
- 3-WL GNN update layer :

$$h^{\ell+1} = \operatorname{Concat}(m_{W_1^{\ell}}(h^{\ell}) \cdot m_{W_2^{\ell}}(h^{\ell}), \ m_{W_3^{\ell}}(h^{\ell}))$$
where $h^{\ell+1} \in \mathbb{R}^{n \times n \times d_{\ell+1}}, h^{\ell} \in \mathbb{R}^{n \times n \times d_{\ell}}, W_1, W_2, W_3 \in \mathbb{R}^{2 \times d_{\ell} \times d_{\ell+1}}$



RingGNNs^[1]

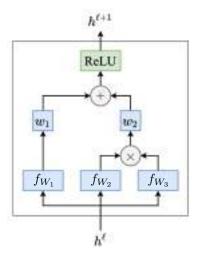
- Related method to 3-WL GNNs to design more expressive GNNs than GINs.
 - Higher-order interaction between nodes is produced by multiplying equivariant linear layers^[2].
 - RingGNN update layer :

$$h^{\ell+1} = \sigma \left(w_1^{\ell} f_{W_1^{\ell}}(h^{\ell}) + w_2^{\ell} f_{W_2^{\ell}}(h^{\ell}) \cdot f_{W_3^{\ell}}(h^{\ell}) \right),$$
where $h^{\ell+1} \in \mathbb{R}^{n \times n \times d_{\ell+1}}, h^{\ell} \in \mathbb{R}^{n \times n \times d_{\ell}}, w_{1,2}^{\ell} \in \mathbb{R}, W_1, W_2, W_3 \in \mathbb{R}^{d_{\ell} \times d_{\ell+1} \times 17},$

where f_W are the equivariant linear layers defined as:

$$(f_W(h))_{\cdot,\cdot,q'} = \sum_{k=1}^{15+2} \sum_{q=1}^{d_\ell} W_{k,q,q'} f_k(h_{\cdot,\cdot,q}) \in \mathbb{R}^{n \times n \times d_{\ell+1}}$$

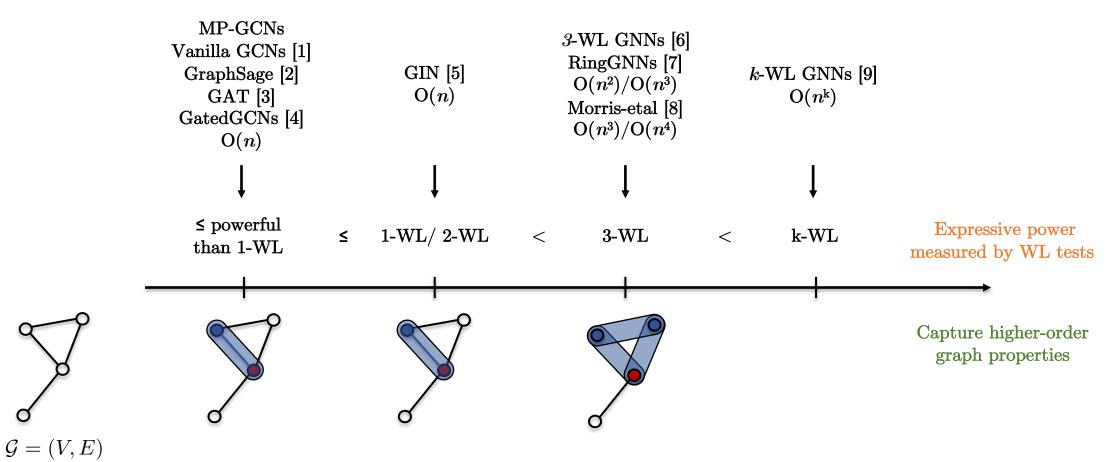
and f_k are all possible 15 equivariant linear functions $f_k : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ for a given tensor $h \in \mathbb{R}^{n \times n}$, and 2 bias functions.



^[1] Chen, Villar, Chen, Bruna, On the equivalence between graph isomorphism testing and function approximation with gnns, 2019

^[2] Maron, Ben-Hamu, Shamir, Lipman, Invariant and equivariant graph networks, 2019

Expressivity Hierarchy w.r.t. WL Test



- [1] Kipf, Welling, Semi-supervised classification with graph convolutional networks, 2017
- [2] Hamilton, Ying, Leskovec, Inductive representation learning on large graphs, 2017
- [3] Velickovic, Cucurull, Casanova, Romero, Lio, Bengio, Graph attention networks, 2017
- [4] Bresson, Laurent, Residual gated graph convnets, 2017
- [5] Xu, Hu, Leskovec, Jegelka, How powerful are graph neural networks?, 2019
- [6] Maron, Ben-Hamu, Serviansky, Lipman, Provably powerful graph networks, 2019
- [7] Chen, Villar, Chen, Bruna, On the equivalence between graph isomorphism testing and function approximation with gnns, 2019
- [8] Morris, Ritzert, Fey, Hamilton, Lenssen, Rattan, Grohe, Weisfeiler and leman go neural: Higher-order graph neural networks, 2019
- [9] Maron, Ben-Hamu, Shamir, Lipman, Invariant and equivariant graph networks, 2019

Outline

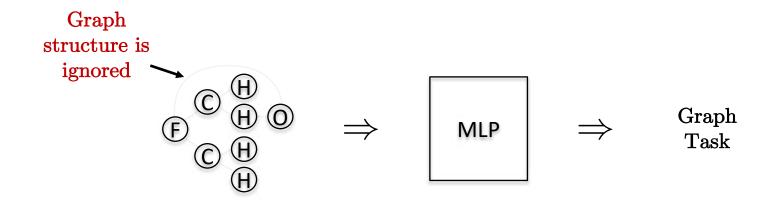
- Motivation
- Message-Passing GCNs
- Weisfeiler-Lehman GNNs
- Graph-Agnostic GNNs
- Datasets
- Infrastructure and Experimental Setting
- Benchmarking Results
- Laplacian Positional Encodings
- Link Prediction with Edge Representation
- Conclusion

MLP Baseline

• Graph-agnostic NNs: As a sanity check, we compare GNNs to a simple MLP network which updates each node independent of one-other:

$$h_i^{\ell+1} = \sigma(W^{\ell} \ h_i^{\ell}), \ h^{\ell+1} \in \mathbb{R}^{n \times d_{\ell+1}}, \ h^{\ell} \in \mathbb{R}^{n \times d_{\ell}}, \ W^{\ell} \in \mathbb{R}^{d_{\ell+1} \times d_{\ell}}$$

and passes these features to the task-based readout layer.



Outline

- Motivation
- Message-Passing GCNs
- Weisfeiler-Lehman GNNs
- Graph-Agnostic GNNs
- Datasets
- Infrastructure and Experimental Setting
- Benchmarking Results
- Laplacian Positional Encodings
- Link Prediction with Edge Representation
- Conclusion

Datasets

- Main issue with prevalent datasets : Small sizes
- Ideal dataset : Representative, realistic, and medium/large-scale size.
- Challenges:
 - What theoretical tool to define the quality of dataset or validate its statistical representativeness for a given task?
 - Several arbitrary choices when preparing graphs, such as node and edge features. For example, e-commerce product features.
 - Appropriate size may depend on the task complexity as well as the dimensionality and statistics of underlying data.
- The recent Open Graph Benchmark^[1] (OGB) project is a much needed initiative to tackle these challenges.
 - OGB offers a collection of medium/large-scale real-world graph datasets and evaluation protocols, with an emphasis on out-of-distribution generalization performance through meaningful data splits.

^[1] Hu, Fey, Zitnik, Dong, Ren, Liu, Catasta, Leskovec, Open graph benchmark: Datasets for machine learning on graphs, 2020

- Appropriate datasets :
 - Datasets that are able to statistically separate the performance of GNNs.
- Summary of the 7 medium-scale datasets :

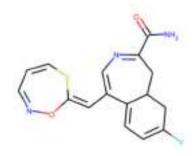
Domain & Construction	Dataset	#Graphs	#Nodes	Total #Nodes	Task
Chemistry: Real-world molecular graphs	ZINC	12K	9-37	277,864	Graph Regression
Mathematical Modelling: Artificial graphs generated from Stochastic Block Models	PATTERN CLUSTER	14K 12K	44-188 41-190	1,664,491 1,406,436	Node Classification
Computer Vision: Graphs constructed with SLIC super-pixels of images	MNIST CIFAR10	70K 60K	40-75 85-150	4,939,668 7,058,005	Graph Classification
Combinatorial Optimization: Uniformly generated artificial Euclidean graphs	TSP	12K	50-500	3,309,140	Edge Classification
Social Networks: Real-world citation graph	COLLAB	1	235,868	235,868	Edge Classification
Circular Skip Links: Isomorphic graphs with same degree	CSL	150	41	6,150	Graph Classification

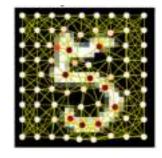
4 datasets are artificially generated, 2 datasets are semi-artificial, and 2 are real-world.

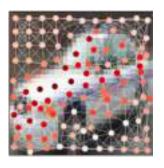
Sizes in terms of total number of nodes vary between 0.27M to 7M.

4 most fundamental graph tasks: graph regression, graph classification, node classification and link prediction.

- ZINC^[1]: A popular real-world molecular dataset of 250K graphs, out of which we randomly select 12K for efficiency. We consider the task of graph property regression for constrained solubility, an important chemical property for designing generative GNNs for molecules^[2].
 - Statistics: 10,000 train/1,000 validation/1,000 test graphs of sizes 9-37 nodes/heavy atoms.
- MNIST^[3]/CIFAR10^[4]: Classical image classification datasets converted into graphs using super-pixels^[5] and assigning node features as the super-pixel coordinates and mean intensity. These datasets are sanity-checks, as we expect most GNNs to perform close to 100% for MNIST and well enough for CIFAR10.
 - Statistics: MNIST has 55,000 train/5,000 validation/10,000 test graphs of sizes 40-75 nodes and CIFAR10 has 45,000 train/5,000 validation/10,000 test graphs of sizes 85-150 nodes.







^[1] Irwin, Sterling, Mysinger, Bolstad, Coleman, Zinc: a free tool to discover chemistry for biology, 2012

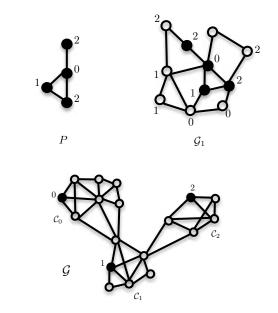
^[2] Jia, Lin, Ying, You, Leskovec, Aiken, Redundancy-free computation graphs for graph neural networks, 2019

^[3] LeCun, Bottou, Bengio, Haffner, Gradient-based learning applied to document recognition, 1998

^[4] Alex Krizhevsky etal, Learning multiple layers of features from tiny images, 2009

^[5] Achanta, Shaji, Smith, Lucchi, Fua, Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods, 2012

- PATTERN/CLUSTER: Node classification tasks generated with Stochastic Block Models^[1], which are widely used to model communities in social networks by modulating the intra- and extra-communities connections. PATTERN tests the fundamental task of recognizing specific predetermined subgraphs^[2].
 - Statistics: PATTERN has 10,000 train/2,000 validation/2,000 test graphs of sizes 50-180 nodes. CLUSTER has 10,000 train/1,000 validation/1,000 test graphs of sizes 40-190 nodes.
- CSL^[3]: Synthetic to test the expressivity of GNNs. Graphs are isomorphic if they have the same degree and the task is to classify non-isomorphic graphs.
 - Statistics: 5-fold cross-validation split of 150 graphs of sizes 41 nodes with train-val-test ratio 3:1:1.







CSL(n=11, R=2)

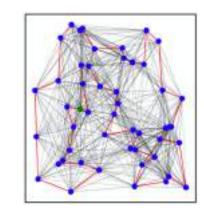
CSL(n=11, R=4)

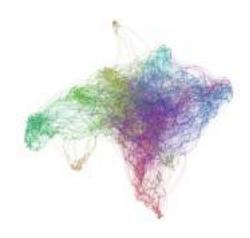
^[1] Abbe, Community detection and stochastic block models: recent developments, 2017

^[2] Scarselli, Gori, Tsoi, Hagenbuchner, Monfardini, The Graph Neural Network Model, 2009

^[3] Murphy, Srinivasan, Rao, Ribeiro, Relational pooling for graph representations, 2019

- Travelling Salesman Problem (TSP): Link prediction on 2D Euclidean graphs to identify edges belonging to the optimal TSP solution given by Concorde^[1]. TSP is the most studied NP-hard combinatorial problem with a growing body of literature on leveraging GNNs to learn better solvers^[2,3].
 - Statistics: 10,000 train TSPs / 1,000 validation TSPs / 1,000 test TSPs, where the number of nodes is randomly selected in [50, 500].
- OGB-COLLAB^[4]: Link prediction dataset proposed by OGB corresponding to a collaboration network between scientists. The task is to predict future author collaboration relationships given past collaboration links.
 - Statistics: A single large temporal graph of size 235K nodes with given train/validation/test edge splits.





^[1] Applegate, Bixby, Chvatal, Cook, Concorde tsp solver, 2006

^[2] Vinyals, Fortunato, Jaitly, Pointer, 2015

^[3] Bello, Pham, Le, Norouzi, Bengio, Neural combinatorial optimization with reinforcement learning, 2016

^[4] Hu, Fey, Zitnik, Dong, Ren, Liu, Catasta, Leskovec, Open graph benchmark: Datasets for machine learning on graphs, 2020

Outline

- Motivation
- Message-Passing GCNs
- Weisfeiler-Lehman GNNs
- Graph-Agnostic GNNs
- Datasets
- Infrastructure and Experimental Setting
- Benchmarking Results
- Laplacian Positional Encodings
- Link Prediction with Edge Representation
- Conclusion

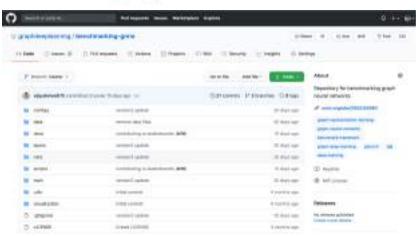
Benchmark Infrastructure

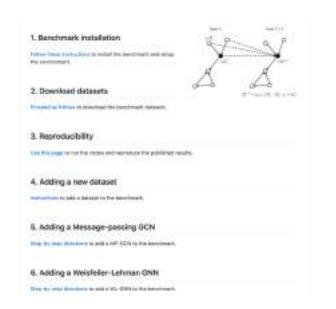


• GitHub Repo:

https://github.com/graphdeeplearning/benchmarking-gnns

- Objectives :
 - Ease-of-use and modular, enabling new users to experiment and study the building blocks of GNNs.
 - Experimentally rigorous and fair for all models being benchmarked.
 - Being comprehensive for tracking progress of new GNNs and novel dataset/task.
- Components:
 - Data pipelines
 - GNN layers and models
 - Training and evaluation functions
 - Network and hyperparameter configurations
 - Scripts for reproducibility





Experimental Setting

• Data splits: Given for ZINC, MNIST, CIFAR10, CSL, OGB-COLLAB, random for PATTERN, CLUSTER, TSP.

Training:

• Adam optimizer w/ learning rate decay strategy.

Initial learning rate is 1e-3/1e-4, reduced by half if validation loss does not improve after 5/10 epochs.

Training is stopped if learning rate \leq 1e-6 or computational time \geq 12 hours.

Statistics with 4 results using 4 different seeds are reported.

Parameter budgets :

- Our goal is not to find the optimal hyperparameters (computationally expensive) but to compare models within the same parameter budget and a maximal computational time.
- Two parameter budgets :
 - 100k (w/ L=4) and 500k (w/ L=16) parameters for each GNNs for all tasks.

- Motivation
- Message-Passing GCNs
- Weisfeiler-Lehman GNNs
- Graph-Agnostic GNNs
- Datasets
- Infrastructure and Experimental Setting
- Benchmarking Results
- Laplacian Positional Encodings
- Link Prediction with Edge Representation
- Conclusion

Benchmarking Results

- Result #1 : MP-GCNs outperformed WL-GNNs on all datasets.
 - Potential reasons:
 - $O(n^2)/O(n^3)$ memory/speed complexities of WL-GNNs do not scale to medium-scale datasets.
 - Best result for ZINC, which is the dataset with the smallest sizes, $n \in [9,37]$.
 - Early stage of developments of WL-GNNs.
- Result #2: MP-GCNs benefit from batch normalization and residual connection:
 - Boost performance
 - Accelerate training
 - WL-GNNs do not benefit from RC, BN or LN (performances degrade).
 - Batch implementation is carried out by gradient accumulation.

		I				NODE CLASS	SIFICATIO	N			
Model	L	#Param	Test Acc.±s.d.	PATTERN Train Acc.±s.d.	#Epoch		#Param	Test Acc.±s.d.	CLUSTER Train Acc.±s.d.	#Epoch	Epoch/Total
MLP	4		50 519+0 000	50.487±0.014	42.25	8 95s/0 11hr	106015	20.973±0.004	20.938±0.002	42.25	5 83s/0 07hr
GCN	4	105263	63 880±0.000	50.487±0.014 65.126±0.135	105.00	8.95s/0.11nr 118.85s/3.51hr	101655	53.445±2.029	54.041+2.197	70.00	
GCN	16	500823	71.892±0.334	78.409±1.592	81.50	492.19s/11.31hr	501687	53.445±2.029 68.498±0.976	71.729±2.212	79.75	65.72s/1.30hr 270.28s/6.08hr
GraphSage	4	101739	50.516±0.001	50.473±0.014	43.75	93.41s/1.17hr	102187	50.454±0.145	54.374±0.203	64.00	53.56s/0.97hr
	16	502842	50.492±0.001	50.487±0.005	46.50	391.19s/5.19hr	503350	63.844±0.110	86.710±0.167	57.75	225.61s/3.70hr
MoNet	4	103775	85.482±0.037	85.569±0.044	89.75	35.71s/0.90hr	104227	58.064±0.131	58.454±0.183	76.25	24.29s/0.52hr
GAT	16 4	511487 109936	85.582±0.038 75.824±1.823	85.720±0.068 77.883±1.632	81.75 96.00	68.49s/1.58hr 20.92s/0.57hr	511999 110700	66.407±0.540 57.732±0.323	67.727±0.649 58.331±0.342	77.75 67.25	47.82s/1.05hr 14.17s/0.27hr
GAI	16	526990	75.824±1.823 78.271±0.186	90.212±0.476	53.50	20.92s/0.5 /nr 50.33s/0.77hr	527874	70.587±0.447	76.074±1.362	73.50	35.94s/0.75hr
GatedGCN	4	104003	84.480±0.122	84.474±0.155	78.75	139.01s/3.09hr	104355	60.404±0.419	61.618±0.536	94.50	79.97s/2.13hr
GatedGCN-PE	16	502223 502457	85.568±0.088 86.508±0.085	86.007±0.123 86.801±0.133	65.25 65.75	644.71s/11.91hr 647.94s/12.08hr	502615 504253	73.840±0.326	87.880±0.908 88 919±0 720	60.00 57.75	400.07s/6.81hr 399.66s/6.58hr
							0.0.1000	76.082±0.196			
GIN	4 16	100884 508574	85.590±0.011 85.387±0.136	85.852±0.030 85.664±0.116	93.00 86.75	15.24s/0.40hr 25.14s/0.62hr	103544 517570	58.384±0.236 64.716±1.553	59.480±0.337 65.973±1.816	74.75 80.75	10.71s/0.23hr 20.67s/0.47hr
RingGNN	2 2	105206	86.245±0.013	86.118±0.034	75.00	573.37s/12.17hr	104746	42.418±20.063	42.520±20.212	74.50	428.24s/8.79hr
	2	504766	86.244±0.025	86.105±0.021	72.00	595.97s/12.15hr	524202	22.340±0.000	22.304±0.000	43.25	501.84s/6.22hr
3WLGNN	8	505749 103572	Diverged 85.661±0.353	Diverged 85.608±0.337	Diverged 95.00	Diverged 304.79s/7.88hr	514380 105552	Diverged 57.130±6.539	Diverged 57.404±6.597	Diverged 116.00	Diverged 219.51s/6.52hr
SWLGINN	3	502872	85.341±0.207	85.270±0.198	81.75	424.23s/9.56hr	507252	55.489±7.863	55.736±8.024	66.00	319.98s/5.79hr
	8	581716	Diverged	Diverged	Diverged	Diverged	586788	Diverged	Diverged	Diverged	Diverged
						GRAPH CLAS	SIFICATIO)N			
Model	L	#Param	Test Acc.±s.d.	MNIST Train Acc.±s.d.	#Epoch	Epoch/Total	#Param	Test Acc.±s.d.	CIFAR10 Train Acc.±s.d.	#Epoch	Epoch/Total
MLP	4	104044	95.340±0.138	97.432±0.470	232.25	22.74s/1.48hr	104380	56.340±0.181	65.113±1.685	185.25	29.48s/1.53hr
GCN	4	101365	90.705±0.218	97.196±0.223	127.50	83.41s/2.99hr	101657	55.710±0.381	69.523±1.948	142.50	109.70s/4.39hr
GraphSage	4	104337	97.312±0.097	100.000±0.000	98.25	113.12s/3.13hr	104517	65.767±0.308	99.719±0.062	93.50	124.61s/3.29hr
MoNet GAT	4	104049 110400	90.805±0.032 95.535±0.205	96.609±0.440 99.994±0.008	146.25 104.75	93.19s/3.82hr 42.26s/1.25hr	104229 110704	54.655±0.518 64.223±0.455	65.911±2.515 89.114±0.499	141.50 103.75	97.13s/3.85hr 55.27s/1.62hr
GatedGCN	4	104217	97.340±0.143	100.000±0.000	96.25	128.79s/3.50hr	104357	67.312±0.311	94.553±1.018	97.00	154.15s/4.22hr
GIN	4	105434	96.485±0.252	100.000±0.000	128.00	39.22s/1.41hr	105654	55.255±1.527	79.412±9.700	141.50	52.12s/2.07hr
RingGNN	4 2	105398	11.350 ± 0.000	11.235±0.000	14.00	2945.69s/12.77hr	105165	19.300±16.108	19.556±16.397	13.50	3112.96s/13.00h
	2	505182 506357	91.860±0.449	92.169±0.505	16.25	2575.99s/12.63hr	504949 510439	39.165±17.114	40.209±17.790	13.75	2998.24s/12.60h
3WLGNN	8	108024	Diverged 95.075±0.961	Diverged 95.830±1.338	Diverged 27.75	Diverged 1523.20s/12.40hr	108516	Diverged 59.175±1.593	Diverged 63.751±2.697	Diverged 28.50	Diverged 1506.29s/12.60h
SWEGITI	3	501690	95.002±0.419	95.692±0.677	26.25	1608.73s/12.42hr	502770	58.043±2.512	61.574±3.575	20.00	2091.22s/12.55hi
	8	500816	Diverged	Diverged	Diverged	Diverged	501584	Diverged	Diverged	Diverged	Diverged
				TSP		LINK PRE	DICTION		COLLAB		
Model	L	#Param	Test F1±s.d.	Train F1±s.d.	#Epoch	Epoch/Total	#Param (L = 3)	Test Hits±s.d.	Train Hits±s.d.	#Epoch	Epoch/Total
MLP	4	96956	0.544±0.001	0.544±0.001	164.25	50.15s/2.31hr	(L = 3) 39441	20.350±2.168	29.807±3.360	147.50	2.09s/0.09hr
GCN	4	95702	0.544±0.001	0.544±0.001	261.00	152.89s/11.15hr	40479	50.422±1.131	92.112±0.991	122.50	351 05s/12 04hr
GraphSage	4	99263	0.665±0.003	0.669 ± 0.003	266.00	157.26s/11.68hr	39856	51.618±0.690	99.949±0.052	152.75	277.93s/11.87hr
MoNet	4	99007	0.641±0.002	0.643±0.002	282.00	84.46s/6.65hr	39751	36.144±2.191	61.156±3.973	167.50	26.69s/1.26hr
GAT	4	96182	0.671 ± 0.002	0.673 ± 0.002	328.25	68.23s/6.25hr	42637	51.501±0.962	97.851±1.114	157.00	18.12s/0.80hr
GatedGCN atedGCN-PE	4	97858	0.791±0.003	0.793±0.003	159.00	218.20s/9.72hr	40965 41889	52.635±1.168 52.849±1.345	96.103±1.876 96.165±0.453	95.00 94.75	453.47s/12.09hr 452.75s/12.08hr
GatedGCN-E	4	97858	0.808 ± 0.003	0.811 ± 0.003	197.00	218.51s/12.04hr	40965	49.212±1.560	88.747±1.058	95.00	451.21s/12.03hr
GatedGCN-E	16	500770	0.838 ± 0.002	0.850 ± 0.001	53.00	807.23s/12.17hr			-		
									70.55514.444		
GIN	4	99002	0.656±0.003	0.660 ± 0.003	273.50	72.73s/5.56hr	39544	41.730±2.284	70.555±4.444	140.25	8.66s/0.34hr
GIN RingGNN	4 2	106862	0.643 ± 0.024	0.644 ± 0.024	2.00	17850.52s/17.19hr	39544	OOM	/0.555±4.444	140.25	8.66s/0.34hr
	2	106862 507938	0.643 ± 0.024 0.704 ± 0.003	0.644 ± 0.024 0.705 ± 0.003	2.00 3.00	17850.52s/17.19hr 12835.53s/16.08hr	39544	OOM	RingGNN	and 3WLGN	IN rely on
	2 8 3	106862 507938 506564 106366	0.643±0.024 0.704±0.003 Diverged 0.694±0.073	0.644±0.024 0.705±0.003 Diverged 0.695±0.073	2.00 3.00 Diverged 2.00	17850.52s/17.19hr 12835.53s/16.08hr Diverged 17468.81s/16.59hr	39544	OOM OOM OOM OOM	RingGNN dense tense	and 3WLGN	IN rely on
RingGNN	2 8 3 3	106862 507938 506564 106366 506681	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312	2.00 3.00 Diverged 2.00 2.00	17850.52s/17.19hr 12835.53s/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr	39544	OOM OOM OOM OOM	RingGNN dense tense	and 3WLGN	IN rely on
RingGNN 3WLGNN	2 8 3	106862 507938 506564 106366 506681 508832	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM	0.644±0.024 0.705±0.003 Diverged 0.695±0.073	2.00 3.00 Diverged 2.00	17850.52s/17.19hr 12835.53s/16.08hr Diverged 17468.81s/16.59hr	39544	OOM OOM OOM OOM	RingGNN dense tense	and 3WLGN	IN rely on
RingGNN 3WLGNN	2 8 3 3	106862 507938 506564 106366 506681	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312	2.00 3.00 Diverged 2.00 2.00	17850.52s/17.19hr 12835.53s/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr	39544 - - - - - - - - - - - - - -	OOM OOM OOM OOM	RingGNN dense tense	and 3WLGN	IN rely on
RingGNN 3WLGNN NN Heuristic Matrix Fact.	2 8 3 3 8	106862 507938 506564 106366 506681 508832 k =2	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM	2.00 3.00 Diverged 2.00 2.00 OOM	17850.52s/17.19hr 12835.53s/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM	60546561	OOM OOM OOM OOM OOM OOM	RingGNN dense tense on both GF 100.000±0.000	and 3WLGN ors which lea U and CPU 254.33	IN rely on ads to OOM memory.
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model	2 8 3 3 8	106862 507938 506564 106366 506681 508832 k =2	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d.	2.00 3.00 Diverged 2.00 2.00 OOM	17850.52s/17.19hr 12835.53s/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM	60546561	OOM OOM OOM OOM OOM OOM 44.206±0.452	RingGNN dense tense on both GF 100.000±0.000 er is better, except f	and 3WLGN ors which lea U and CPU 254.33 or ZINC)	IN rely on adds to OOM memory.
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP	2 8 3 3 8 0	106862 507938 506564 106366 506681 508832 k =2 #Param 108975	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 GRA Test MAE±s.d. 0.706±0.006	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005	2.00 3.00 Diverged 2.00 2.00 OOM -ZINC #Epoch 116.75	17850.52s/17.19hr 12835.53s/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr	60546561 Evaluat • CLUS	OOM OOM OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high	RingGNN dense tense on both GF 100.000±0.000 er is better, except f se weighted accura	and 3WLGN ors which lea U and CPU 254.33 or ZINC) cy w.r.t. the o	IN rely on ads to OOM memory. 2.66s/0.21hr
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model	2 8 3 3 8	106862 507938 506564 106366 506681 508832 k =2	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d.	2.00 3.00 Diverged 2.00 2.00 OOM	17850.52s/17.19hr 12835.53s/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM	60546561 Evaluat CLUS MNIS	OOM OOM OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high STER, PATTERN v	RingGNN dense tense on both GF 100.000±0.000 er is better, except f se weighted accura ulti-label classificat	and 3WLGN ors which lea U and CPU 254.33 or ZINC) by w.r.t. the orion accuracy	IN rely on ads to OOM memory. 2.66s/0.21hr
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP	2 8 3 3 8 0 <i>L</i> 4 16 4	106862 507938 506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 GRA Test MAE±s.d. 0.706±0.006 0.459±0.006 0.367±0.011 0.468±0.003	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.251±0.004	2.00 3.00 Diverged 2.00 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25	17850.52s/17.19hr 12835.53s/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s0.03hr 2.89s/0.16hr 12.78s/0.71hr 3.74s/0.15hr	60546561 Evaluat • CLUS • MNIS • TSP t	OOM OOM OOM OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high STER, PATTERN EST, CIFAR10 use ra sisses binary F1 score	RingGNN dense tense on both GF 100.000±0.000 er is better, except f se weighted accura util-label classificat e for the positive ed,	and 3WLGN ors which lea U and CPU 254.33 or ZINC) by w.r.t. the orion accuracy ges.	IN rely on adds to OOM memory. 2.66s/0.21hr
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP GCN GraphSage	2 8 3 3 8 0 <i>L</i> 4 16 4 16	106862 507938 506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505341	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 Test MAE±s.d. 0.706±0.006 0.459±0.006 0.367±0.011 0.468±0.003 0.398±0.002	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.251±0.004	2.00 3.00 Diverged 2.00 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25 145.50	17850.52s/17.19hr 12835.53s/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr 12.78s/0.71hr 3.74s/0.15hr 16.61s/0.68hr	60546561 Evaluat • CLUS • MNIS • TSP v	OOM OOM OOM OOM OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high STER, PATTERN to STER, CIFARIO use in Stees binary F1 score Stees binary F1 score AB uses Hits@50	RingGNN dense tense on both GF 100.000±0.000 er is better, except f see weighted accura ulti-label classificat for the positive ed, via the evaluator pr	and 3WLGN ors which lea U and CPU 254.33 or ZINC) by w.r.t. the orion accuracy ges.	IN rely on adds to OOM memory. 2.66s/0.21hr
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP GCN	2 8 3 3 8 0 <i>L</i> 4 16 4 16 4	106862 507938 506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 GRA Test MAE±s.d. 0.706±0.006 0.459±0.006 0.367±0.011 0.468±0.003	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±sd. 0.644±0.005 0.343±0.011 0.128±0.019 0.251±0.004 0.081±0.009	2.00 3.00 Diverged 2.00 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25 145.50	17850.52s/17.19hr 12835.53s/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr 12.78s/0.71hr 3.74s/0.15hr 16.61s/0.68hr	60546561 Evaluat CLUS MNIS TSP t COLL ZINC	OOM OOM OOM OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high STER, PATTERN v TT, CIFAR10 use n sees binary FI scon AB uses Hist@50 uses mean absolut	RingGNN dense tense on both GF 100.000±0.000 er is better, except f see weighted accura ulti-label classificat for the positive ed, via the evaluator pr	and 3WLGN ors which lea U and CPU 254.33 or ZINC) by w.r.t. the orion accuracy ges.	IN rely on adds to OOM memory. 2.66s/0.21hr
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP GCN GraphSage	2 2 8 3 3 8 0 0 L 4 4 16 4 16 4 16 4 16 4	106862 507938 506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505079 94977 505341 106002 504013 102385	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 Test MAE±s.d. 0.706±0.006 0.367±0.011 0.468±0.003 0.398±0.002 0.397±0.011 0.292±0.006 0.475±0.007	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.318±0.019 0.318±0.010 0.031±0.000 0.031±0.000	2.00 3.00 Diverged 2.00 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25 145.50 188.25 171.75	17850.52s/17.19h 12835.53s/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr 12.78s/0.71hr 16.61s/0.68hr 1.97s/0.00hr 10.82s/0.52hr 1.97s/0.01hr	60546561 Evaluat • CLUS • MNIS • TSP t • COLL • ZINC Notatioi	OOM OOM OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high STER, PATTERN to TT, CIFAR10 use n sees binary F1 scon AB uses Hits@50 uses mean absolut n:	RingGNN dense tense on both GF 100.000±0.000 er is better, except f se weighted accura f for the positive ed via the evaluator pr e regression error.	and 3WLGN rs which lea U and CPU 254.33 or ZINC) cy w.r.t. the cion accuracy ges. ovided by O	IN rely on dds to OOM memory. 2.66s/0.21hr class sizes.
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT	2 2 8 3 3 8 0 0 L 4 16 4 16 4 16 4 16 4 16	106862 507938 506564 106366 506681 508832 k =2 #Param 108977 505079 94977 505341 106002 504013 102385 531345	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 Test MAE±s.d. 0.706±0.006 0.459±0.006 0.367±0.011 0.408±0.002 0.397±0.010 0.398±0.002 0.397±0.010 0.398±0.002 0.397±0.010 0.398±0.002 0.397±0.010 0.398±0.002	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.251±0.004 0.081±0.009 0.318±0.016 0.093±0.014 0.317±0.006	2.00 3.00 Diverged 2.00 COM -ZINC -Epoch 116.75 196.25 197.00 147.25 145.25 171.75 137.50 144.00	17850.525/17.19th 12835.534/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr 1.278s/0.71hr 3.74s/0.15hr 1.97s/0.10hr 1.97s/0.00hr 1.97s/0.01	Evaluat CLUS MNIS TSP t COLL ZINC Notation	OOM OOM OOM OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high STER, PATTERN t TT, CIFAR10 use n sees binary F1 sees binary F1 sees AB uses Hits@50 uses mean absolut ns. Is with the suffix -1	RingGNN dense tense on both GF 100.000±0.000 er is better, except f se weighted accura ulti-label classificat for the positive via the evaluator pr e regression error. E use input edge fea	and 3WLGN rs which lea U and CPU 254.33 or ZINC) cy w.r.t. the cion accuracy ges. ovided by O tures to initia	iN rely on ds to OOM memory. 2.66s/0.21hr class sizes.
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN	2 2 8 3 3 8 0 0 L 4 4 16 4 16 4 16 4 16 4	106862 507938 506564 106366 506813 508832 k = 2 #Param 108975 103077 505079 94977 505341 106002 504013 102385 53415 105735	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 Test MAE±s.d. 0.706±0.006 0.367±0.011 0.468±0.003 0.398±0.002 0.397±0.010 0.292±0.006 0.475±0.007 0.348±0.007 0.397±0.010 0.292±0.006 0.475±0.007 0.348±0.007 0.348±0.007 0.348±0.007 0.348±0.007 0.348±0.007 0.348±0.007 0.348±0.007	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.318±0.019 0.318±0.010 0.031±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004	2.00 3.00 Diverged 2.00 2.00 OOM = ZINC #Epoch 116.75 196.25 197.00 147.25 145.50 188.25 171.75 144.00 173.50	17850 525/17.19th 12835.534/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr 1.74c0.15hr 16.74c0.15hr 16.74c0.15hr 16.74c0.15hr 16.91c0.08hr 1.97s/0.01hr 1.97s/0.01hr 1.97s/0.02hr	Evaluat CLUS MNIS TSP t COLLI ZINC Notation	OOM OOM OOM OOM OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high STER, PATTERN & TT, CIFAR10 use n sees binary F1 scon .AB uses Hits@50 uses mean absolut n: Is with the suffix J	RingGNN dense tenso on both GF 100.000±0.000 er is better, except f se weighted accura tulti-label classificate of or the positive ed via the evaluator pr e regression error. Z use input edge fee ond type, TSP. Euc	and 3WLGN rs which lea U and CPU 254.33 or ZINC) cy w.r.t. the cion accuracy ges. ovided by O tures to initia	iN rely on ds to OOM memory. 2.66s/0.21hr class sizes.
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN-E	2 2 8 3 3 8 0 0 L 4 4 16 4 4 16 4 16 4 4 16 4 4 16 4 16	106862 507938 506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505079 94977 5050413 102385 531345 105735 105875 504309	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 GRA Test MAE±s.d. 0.499±0.006 0.459±0.006 0.459±0.001 0.292±0.006 0.475±0.001 0.475±0.007 0.387±0.001 0.475±0.007 0.455±0.001 0.475±0.007 0.455±0.001 0.375±0.003 0.475±0.007 0.455±0.001 0.375±0.003 0.375±0.003 0.375±0.003 0.375±0.003 0.375±0.003 0.375±0.003 0.375±0.003 0.375±0.003 0.375±0.003 0.375±0.003 0.375±0.003	0.64±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.251±0.004 0.318±0.016 0.033±0.011 0.0318±0.016 0.075±0.014 0.356±0.007 0.375±0.014 0.256±0.017	2.00 3.00 Diverged 2.00 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25 145.50 188.25 171.75 137.50 144.00 173.50 194.75 166.75	17850.525/17.19th 12835.534/16.08hr Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr 12.78s/0.71hr 3.74s/0.15hr 19750.10hr 1.97s/0.10hr 1.97s/0.10hr 1.97s/0.10hr 1.97s/0.20hr 2.93s/0.11hr 1.2.9s/0.52hr 5.75s/0.29hr 5.75s/0.29hr 5.75s/0.29hr 5.75s/0.29hr 5.75s/0.29hr 5.75s/0.29hr	Evaluat CLUS MNIS TSP t COLLI ZINC Notation Mode repres	OOM OOM OOM OOM OOM OOM OOM OOM 44.206±0.452 Ion Metries: (high STER, PATTERN v. TT. CIFAR10 use n sees binary F1 scon uses mean absolut n: Is with the suffix -1 Is with the suffix -1 Is with the suffix -1	RingGNN dense tense on both GF 100.000±0.000 er is better, except f se weighted accura ulti-label classificat for the positive ed via the evaluator pr e regression error. E use input edge fea ond type, TSP: Euc and year).	and 3WLGN ars which lea U and CPU 254.33 or ZINC) cy w.r.t. the cion accuracy ges. ovided by O tures to initilidean distant	IN rely on dds to OOM memory. 2.66s/0.21hr class sizes. GB [?]. alize edge cee, COLLAB:
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN GatedGCN-E-PE	2 2 8 3 3 8 0 0 L 4 4 16 4 16 4 4 16 4 4 4 16 4 4 16 16 16 16 16 16 16 16 16 16 16 16 16	106862 507938 506564 106366 506681 508832 k =2 #Param 108975 103077 505079 94977 505341 106002 504013 102385 531345 102385 531345 105735 50575 50575 50575	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test FI: 0.693 GRA Test MAE±s.d. 0.706±0.006 0.459±0.006 0.367±0.011 0.408±0.003 0.398±0.002 0.397±0.010 0.398±0.007 0.355±0.011 0.355±0.011 0.375±0.003 0.252±0.015 0.214±0.013	0.64±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.318±0.009 0.318±0.009 0.318±0.000 0.057±0.004 0.067±0.000 0.074±0.016	2.00 3.00 Diverged 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25 145.25 171.75 188.25 171.75 188.25 171.75 184.00 173.50 174.75 186.75 186.75 186.75	17850.525/17.19th 12835.534/16.08hr Diverged 17468.814/16.59hr 17190.17s/16.51hr OOM 10.105/0.3hr 10.105/0.3hr 12.78s/0.71hr 16.61s/0.68hr 10.82s/0.52hr 10.82s/0.32hr 12.98s/0.1hr 12.78s/0.71hr 10.82s/0.32hr 12.98s/0.3hr 10.82s/0.32hr 10.80s/0.3hr 10.8	60546561 Evaluat CLUS MNIS TSP t COLL ZINC Notatiot Mode represe collab Mode encode	OOM OOM OOM OOM OOM OOM OOM OOM 44.206±0.452 ion Metries: (high STER, PATTERN 'c TT, CIFAR10 use n sees binary F1 scon uses mean absolut ns. Is with the suffix -1 use mine the suffix -1 uses mine	RingGNN dense tense on both GF 100.000±0.000 er is better, except f se weighted accura ulti-label classificat for the positive ed via the evaluator pr e regression error. E use input edge fea ond type, TSP: Euc and year). *E use Laplacian Ei *B (or ZINC, 2 for	and 3WLGN ars which lea U and CPU 254.33 or ZINC) sy w.r.t. the c ion accuracy ges. ovided by O tures to initia lidean distan genvectors as PATTERN 2	IN rely on ds to OOM memory. 2.66s/0.21hr 2.66s/0.21hr class sizes. GB [?]. alize edge tee, COLLAB: s node positional and 20 for others.
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN-E	2 2 8 3 3 8 0 L 4 4 16 4 16 4 4 16 4 4 16 4 4 16 4 4 4 16 4 4 16 4 16 4 4 4 4	106862 507938 506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505341 106002 504013 102385 531345 105875 50439 505011	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 Test MAE±s.d. 0.706±0.006 0.459±0.006 0.459±0.001 0.292±0.006 0.475±0.007 0.384±0.007 0.384±0.007 0.385±0.001 0.375±0.003 0.282±0.006 0.285±0.001 0.375±0.003 0.282±0.006 0.387±0.001 0.375±0.003 0.387±0.001	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.081±0.009 0.318±0.016 0.093±0.014 0.317±0.006 0.087±0.019 0.128±0.019 0.087±0.010 0.087±0.010 0.087±0.010	2.00 3.00 Diverged 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25 145.50 188.25 147.50 171.75 137.50 144.00 173.50 194.75 186.75 185.00	17850.525/17.1981 12835.534/16.08hr Diverged 17468.81 bi 16.59hr 17190.17816.51hr OOM Epoch/Total 1.01s0.03hr 12.7886.71hr 11.5786.77hr 11.5786.78hr	Evaluat CLUS MINIS TSP t COLL ZINC Notation Mode represcollat Mode encod	OOM OOM OOM OOM OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high STER, PATTERN v. T, CIFAR10 use n sees binary F1 sees binary F1 sees binary B1 sees binary	RingGNN dense tense on both GF 100.000±0.000 er is better, except f se weighted accura ulti-label classificat for the positive for the positive session error. 2. use input edge fea ond type, TSP. Euc and year). 22. Use input edge fea ond ype, TSP. Euc and year).	and 3WLGN srs which leave U and CPU 254.33 or ZINC) yw.r.t. the connection accuracy ges. ovided by O tures to initial idean distar genvectors as pATTERN ole and diver	IN rely on ds to OOM memory. 2.66s/0.21hr class sizes. GB [?]. dize edge cc, COLLAB: s node positional and 20 for others. gent runs across
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN-E- GatedGCN-E-PE	2 2 8 3 3 8 0 0 L 4 4 16 4 16 4 4 16 4 4 16 4 4 16 4 4 16 16 4 4 16 16 16 16 16 16 16 16 16 16 16 16 16	106862 507938 506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505341 106002 504013 102385 504309 504309 509549 97978	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.311 0CM Test F1: 0.693 Test MAE±s.d. 0.706±0.006 0.459±0.006 0.367±0.011 0.292±0.006 0.475±0.007 0.393±0.007 0.393±0.007 0.395±0.007	0.64±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.318±0.009 0.318±0.009 0.318±0.000 0.057±0.004 0.067±0.000 0.074±0.016	2.00 3.00 Diverged 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25 145.25 171.75 188.25 171.75 188.25 171.75 184.00 173.50 174.75 186.75 186.75 186.75	17850.525/17.1981 12835.534/16.08hr Diverged 17468.814/16.59hr 17190.179/16.51hr OOM Epoch/Total 1.01s/0.25hr 2.858/0.16hr 12.788/0.71hr 13.746/0.15hr 19.756.10hr 19.756.10hr 1.9756.10hr 1.9756.10hr 1.9756.10hr 1.9756.01hr 1.9756.01h	Evaluat CLUS MNIS TSP v COLLI ZINC Notation Mode repres collat Mode encod Resul	OOM OOM OOM OOM OOM OOM OOM OOM 44.206±0.452 Ion Metrics: (high STER, PATTERN v. rt. T. CIFAR10 uses Hisse950 Is with the suffix -1 Is with the suffix -1 Is with the suffix -1 Is migs, with dimens, w	RingGNN dense tens on both GF 100.000±0.000 er is better, except f se weighted accurac tulti-label classificat e for the positive ed via the evaluator pr e regression error. 2 use input edge fea ond type, TSP: Eu and year). **E use Laplacian Ei n 8 for ZINC, 2 for gred indicate unstal ming rate values.	and 3WLGN srs which lead U and CPU 254.33 or ZINC) cy w.r.t. the e ion accuracy ges. ovided by O tures to initia lidean distar genvectors as PATTERN: ele and diver 0-3,10-4,10-7,10-7,10-7,10-7,10-7,10-7,10-7,10-7	IN rely on dds to OOM memory. 2.66s/0.21hr 2.66s/0.21hr class sizes. GB [?]. alize edge nee, COLLAB: s node positional and 20 for others, gent runs across 0-5.
RingGNN 3WLGNN NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN-GatedGCN-E-PE GIN	2 2 8 3 3 8 0 0 L 4 4 16 4 16 4 4 16 4 4 16 4 4 16 4 4 16 16 4 4 16 16 16 16 16 16 16 16 16 16 16 16 16	106862 507938 506564 1106365 506681 508832 k = 2 #Param 108975 103077 505079 94977 505341 106002 504013 102385 531345 105735 105	0.643±0.024 0.704±0.003 Diverged 0.694±0.073 0.288±0.071 Test F1: 0.693 Test F1: 0.693 GRA Test MAE±s.d. 0.459±0.005 0.367±0.011 0.468±0.003 0.398±0.002 0.397±0.010 0.455±0.011 0.475±0.003 0.398±0.002 0.397±0.010 0.398±0.003 0.398±0	0.64±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM Train MAE±s.d. 0.64±0.003 0.128±0.019 0.128±0.019 0.051±0.009 0.318±0.016 0.093±0.014 0.287±0.014	2,00 3,00 Diverged 2,00 OOM -ZINC #Epoch 116,75 197,20 147,25 145,50 187,25 145,50 171,75 137,50 144,00 173,30 173,30 173,30 174,20 174,20 174,20 174,20 174,20 174,20 174,20 174,20 174,20 174,20 174,20 174,20 174,20 174,20 174,20 175,20 176,75 185,00 177,20 17	17850.523/17.1981. 12835.534/16.08hr Diverged 17468.814/16.59hr 17190.794/16.51hr OOM Epoch/Total 17190.038hr 2.8350.16hr 17190.038hr 1.9780.01hr 1.9780.10hr 1.9780.10hr 1.9780.00hr	Evaluat CLUS MNIS TSP COLI ZINC Notation Mode represcollat Mode encod Resul all 4 s Resul	OOM	RingGNN dense tense on both GF 100.000±0.000 er is better, except f se weighted accura utili-label classificat for the positive ed via the evaluator pr e regression error. © use input edge fea ond type, TSP: Eue and year). "Be use Laplacian Ei n 8 for ZINC, 2 for gred indicate unse i indicate runs s {\(\) i i i indicate runs s {\(\) i i i indicate runs s {\(\) i i i i indicate runs s {\(\) i i i i i indicate runs s {\(\) i i i i i i i i i i i i i i i i i i i	and 3WLGN srs which lead U and CPU 254.33 or ZINC) cy w.r.t. the e ion accuracy ges. ovided by O tures to initia lidean distar genvectors as PATTERN: ele and diver 0-3,10-4,10-7,10-7,10-7,10-7,10-7,10-7,10-7,10-7	IN rely on dds to OOM memory. 2.66s/0.21hr 2.66s/0.21hr class sizes. GB [?]. alize edge nee, COLLAB: s node positional and 20 for others, gent runs across 0-5.
RingGNN 3WLGNN 3WLGNN -NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN-E tedGCN-E-PE GIN RingGNN	2 2 8 3 3 8 0 0 L 4 16 4 16 4 16 4 4 16 4 4 16 16 16 16 16 16 16 16 16 16 16 16 16	106862 507938 506564 106366 506681 508832 k = 2 #Param 108975 705079 94977 505341 106002 504013 102385 531345 105875 504309 505011 103079 509549 97978 10403 509549 97978 10403 507549 5	0.643±0.024 0.704±0.003 Diverged 0.604±0.073 0.288±0.311 0.004 Test Fi: 0.699 CRN Test MAE:s.d. 0.706±0.006 0.367±0.001 0.459±0.006 0.367±0.001 0.367±0.001 0.375±0.001 0.375±0.001 0.375±0.001 0.375±0.003 0.384±0.007 0.375±0.003 0.375	0.64±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.051±0.004 0.071±0.004 0.072±0.014 0.072±0.014 0.072±0.014 0.073±0.017 0.073±0.016 0.073±0.017 0.0	2.00 3.00 Diverged 2.00 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25 145.50 188.25 171.75 137.50 144.00 173.50 194.75 185.20 144.00 90.25 95.00	17850.525/17.1981 12835.534/16.08hr Diverged 17468.81 ls/16.59hr 17190.179/16.51hr OOM Epoch/Total 1.018/0.03hr 1.018/0.03hr 12.7880.01hr 12.7880.01hr 16.6180.68hr 10.820.0.52hr 12.980.0.18hr 12.980.0.3hr 12.980.0.18hr 12.780.0.71hr 10.820.0.52hr 10.820.0.55hr 2.980.0.18hr 2.980.0.18hr 2.980.0.18hr 2.980.0.18hr 2.980.0.18hr 2.980.0.18hr 3.746.0.28hr 2.080.0.06hr 3.737.0.29hr 3.05.00.006hr 3.737.0.29hr 3.756.0.28hr	Evaluat CLUS MNIS TSP t COLLI ZINC Notatiot Mode represcollaid Mode encod Resul all 4 s Resul on ou	OOM OOM OOM OOM OOM OOM OOM OOM OOM 44.206±0.452 Ion Metrics: (high STER, PATTERN \cdot	RingGNN dense tense on both GF 100.000±0.000 er is better, except f se weighted accura utili-label classificat for the positive ed via the evaluator pr e regression error. © use input edge fea ond type, TSP: Eue and year). "Be use Laplacian Ei n 8 for ZINC, 2 for gred indicate unse i indicate runs s {\(\) i i i indicate runs s {\(\) i i i indicate runs s {\(\) i i i i indicate runs s {\(\) i i i i i indicate runs s {\(\) i i i i i i i i i i i i i i i i i i i	and 3WLGN srs which lead U and CPU 254.33 or ZINC) cy w.r.t. the e ion accuracy ges. ovided by O tures to initia lidean distar genvectors as PATTERN: ele and diver 0-3,10-4,10-7,10-7,10-7,10-7,10-7,10-7,10-7,10-7	IN rely on dds to OOM memory. 2.66s/0.21hr 2.66s/0.21hr class sizes. GB [?]. alize edge nee, COLLAB: s node positional and 20 for others, gent runs across 0-5.
RingGNN 3WLGNN Model MILP GCN GraphSage MoNet GAT GateGGCN GateGGCN GringGNN-E HongGNN RingGNN-E 3WLGNN	2 2 8 3 3 8 0 0 L 4 4 4 4 16 4 4 16 4 4 16 16 4 4 16 16 16 16 16 16 16 16 16 16 16 16 16	106862 507938 506564 100366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505341 106002 504013 102385 531345 105735 504013 102387 504013 103079 509549 97978 509549 97978 509549	0.643±0.024 0.704±0.003 Diverged 0.094±0.17 0.20±0.011 Test FIL: 0.004 0.005 CRA CRA CRA CRA CRA CRA CRA CR	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.012 PH REGRESSION Train MAE±s.d. 0.343±0.011 0.128±0.019 0.251±0.004 0.081±0.009 0.081±0.009 0.081±0.000 0.067±0.004 0.287±0.014 0.286±0.007 0.319±0.015 0.3	2,00 3,00 Diverged 2,00 COM 2,00 COM 2,00 COM 2,00 COM 4Epoch 116,75 196,25 197,00 147,25 147,25 145,50 144,00 173,50 194,75 166,75 185,00 153,25 147,00 90,25 147,00 90,25 150,00 179,75 Diverged 111,25 Diverged	17850.523/17.1981 17850.523/17.1981 17950.65881 17950.65881 17950.65881 17950.65881 17950.75881 17950.	Evaluat CLUS MNIS TSP t COLL ZINC Notation Mode represe collat Mode and Resul all 4 4 Resul on ou Extende	OOM	RingGNN dense tense on both GF 100,000±0,000 er is better, except f se weighted accura with-label classified accura for the positive ed via the evaluator per regression error. the input edge fen ond type. TSP. Euc and year. TSP. Euc and year. In findicate runs whice the control of the positive that in findicate runs whice the control of the positive that the that is the control of the positive that t	and 3WLGN srs which lea U and CPU 254.33 or ZINC) sy w.r.t. the cion accuracy ges. ovided by O tures to initial genvectors as PATTERN: lea and diver 0 - 3, 10 - 4, 11 h throw out of	in rely on the to OoM memory. 2.66w0.21hr 2.66w0.21hr 2.66w0.21hr 2.66w0.21hr 3.66w0.21hr 3.66w0.21hr
RingGNN 3WLGNN -NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN-E ttedGCN-E-PIN RingGNN-E	2 2 8 3 3 8 8 0 0 LL 4 4 4 16 4 16 16 4 16 16 16 2 2 2 8 3 3 3	106862 507938 506564 106366 506681 508832 **Param** 108975 103077 505079 94977 505341 106002 504013 102385 531345 105735 504309 505011 103079 509549 97978 104403 527283 510345 510345 510305 102150 104203	0.643±0.024 0.704±0.003 Deveged 0.694±0.071 0.288±0.311 0.088±0.031 0.088±0.031 0.088±0.031 0.088±0.000 0.408±0.000 0.408±0.000 0.408±0.000 0.408±0.000 0.398±0.000 0.408±0.000 0.398±0.0000 0.398±0.0000 0.398±0.0000 0.398±0.0000 0.398±0.0000 0.398±0.0000 0.398±0.0000 0.3	0.64±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.312 OOM Train MAE±s.d. 0.644±0.005 0.343±0.011 0.233±0.011 0.081±0.009 0.318±0.016 0.093±0.014 0.317±0.006 0.097±0.019 0.319±0.014 0.074±0.015 0.074±0.016 0.067±0.019 0.319±0.016 0.067±0.019 0.319±0.016 0.067±0.019 0.319±0.016 0.067±0.019 0.319±0.016 0.072±0.017 0.074±0.016 0.067±0.019 0.319±0.016 0.074±0.016 0.067±0.019 0.319±0.016 0.072±0.017 0.319±0.016 0.072±0.017 0.319±0.016 0.072±0.017 0.319±0.016 0.072±0.017 0.319±0.016 0.072±0.017 0.319±0.016 0.072±0.017 0.319±0.016 0.072±0.017 0.319±0.016 0.072±0.017 0.319±0.016 0.072±0.017 0.319±0.016 0.072±0.017 0.319±0.016 0.072±0.017 0.319±0.016 0.072±0.017 0.319±0.016 0.319±0.016 0.319±0.016 0.072±0.017 0.319±0.016 0.3	2.00 3.00 Diverged 2.00 OOM -ZINC #Epote 116.75 196.25 197.00 147.25 145.50 188.25 171.75 137.50 144.00 173.50 194.75 166.75 185.00 179.70 185.25 171.75 17	17850.525/17.1981 17850.525/17.1981 1790.179416.51br OOM Epoch/Total 1.0160.05br 1.7950.10br 1.7950.25br 1.7950.10br 1.7950.25br 1.7950.10br 1.7950.75br 1.7950.7	Evaluat CLUS MINIS TSP t COLL Mode represe collat Mode encod Resulu all 4 s Resulu on ou Extende	OOM OOM OOM OOM OOM 44.206±0.452 Ion Metrics: (highlighted) ATTER, PATTERN TERN TER, PATTERN WITHER IS Is with the suffix A is with th	RingGNN dense tense on both GF 100.000±0.000 100.000±0.000 is better, except for is better, except is the devaluation is for the positive of is the exclusion error. See input edge fee and year). Application Ei is for ENC. GR is for ENC. GR is for ENC. GR if indicate unstalled in	and 3WLGN ars which lea U and CPU 254.33 or ZINC) cy w.r.t. the cion accuracy ges. ovided by O tures to initial idean distar genvectors as PATTERN color and and or the and diver 0 - 3, 10 - 4, 1 h throw out of arameters accurated	IN rely on the to OOM memory. 2.66s/0.21hr 2.66s/0.21hr class sizes. GB [7]. alize edge tee, COLLAB: need of others, gent runs across of 50 fm emory errors thieved
RingGNN 3WLGNN Model MILP GCN GraphSage MoNet GAT GatedGCN-E-PE stedGCN-E-PE stedGCN-E-PE	2 2 8 3 3 8 0 0 L 4 4 4 4 16 4 4 16 4 4 16 16 4 4 16 16 16 16 16 16 16 16 16 16 16 16 16	106862 507938 506564 100366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505341 106002 504013 102385 531345 105735 504013 102387 504013 103079 509549 97978 509549 97978 509549	0.643±0.024 0.704±0.003 Diverged 0.094±0.17 0.20±0.011 Test FIL: 0.004 0.005 CRA CRA CRA CRA CRA CRA CRA CR	0.644±0.024 0.705±0.003 Diverged 0.695±0.073 0.290±0.012 PH REGRESSION Train MAE±s.d. 0.343±0.011 0.128±0.019 0.251±0.004 0.081±0.009 0.081±0.009 0.081±0.000 0.067±0.004 0.287±0.014 0.286±0.007 0.319±0.015 0.3	2,00 3,00 Diverged 2,00 COM 2,00 COM 2,00 COM 2,00 COM 4Epoch 116,75 196,25 197,00 147,25 147,25 145,50 144,00 173,50 194,75 166,75 185,00 153,25 147,00 90,25 147,00 90,25 150,00 179,75 Diverged 111,25 Diverged	17850.523/17.1981 17850.523/17.1981 17950.65881 17950.65881 17950.65881 17950.65881 17950.75881 17950.	Evaluat CLUS MINIS TSP t COLL Mode represe collat Mode encod Resulu all 4 s Resulu on ou Extende	OOM OOM OOM OOM OOM 44.206±0.452 Ion Metrics: (highlighted) ATTER, PATTERN TERN TER, PATTERN WITHER IS Is with the suffix A is with th	RingGNN dense tense on both GF 100,000±0,000 er is better, except f se weighted accura with-label classified accura for the positive ed via the evaluator per regression error. the input edge fen ond type. TSP. Euc and year. TSP. Euc and year. In findicate runs whice the control of the positive that in findicate runs whice the control of the positive that the that is the control of the positive that t	and 3WLGN ars which lea U and CPU 254.33 or ZINC) cy w.r.t. the cion accuracy ges. ovided by O tures to initial idean distar genvectors as PATTERN color and and or the and diver 0 - 3, 10 - 4, 1 h throw out of arameters accurated	IN rely on the to OOM memory. 2.66s/0.21hr 2.66s/0.21hr class sizes. GB [7]. alize edge tee, COLLAB: need of others, gent runs across of 50 fm emory errors thieved

Benchmarking Results

- Result #3: Anisotropic mechanism improve (isotropic) GCNs.
 - Sparse attention^[1,2] and dense attention^[3] are not injective (unlike f_{GIN}) but experiments showed that they are good at generalization. See also^[4].
 - Intuitively, softmax attention is flexible to represent max/mean/weighted mean w.r.t. contextual information.
 - Recent work^[5] studied effectiveness of global attention for generalization.

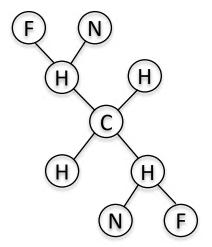
- [1] Bahdanau, Cho, Bengio, Neural machine translation by jointly learning to align and translate, 2015
- [2] Velickovic, Cucurull, Casanova, Romero, Lio, Bengio, Graph attention networks, 2017
- [3] Bresson, Laurent, Residual gated graph convnets, 2017
- [4] Abu-El-Haija, Perozzi, Al-Rfou, Alemi, Watch your step: Learning node embeddings via graph attention, 2018
- [5] Puny, Ben-Hamu, Lipman, From Graph Low-Rank Global Attention to 2-FWL Approximation, 2020

		1				NODE CLASS	SIFICATIO	N.			
	,	#P		PATTERN					CLUSTER		r 1m.1
Model	L	#Param	Test Acc.±s.d.	Train Acc.±s.d.	#Epoch		#Param	Test Acc.±s.d.	Train Acc.±s.d.	#Epoch	Epoch/Total
MLP	4	105263	50.519±0.000	50.487±0.014	42.25 105.00	8.95s/0.11hr	106015	20.973±0.004	20.938±0.002	42.25	5.83s/0.07hr
GCN	4 16	100923 500823	63.880±0.074 71.892±0.334	65.126±0.135 78.409±1.592	81.50	118.85s/3.51hr 492.19s/11.31hr	101655 501687	53.445±2.029 68.498±0.976	54.041±2.197 71.729±2.212	70.00 79.75	65.72s/1.30hr 270.28s/6.08hr
GraphSage	4	101739	50.516±0.001	50.473±0.014	43.75	93.41s/1.17hr	102187	50.454±0.145	54.374±0.203	64.00	53.56s/0.97hr
	16	502842	50.492±0.001	50.487±0.005	46.50	391.19s/5.19hr	503350	63.844±0.110	86.710±0.167	57.75	225.61s/3.70hr
MoNet	4	103775	85.482±0.037	85.569±0.044	89.75	35.71s/0.90hr	104227	58.064±0.131	58.454±0.183	76.25	24.29s/0.52hr
GAT	16	511487 109936	85.582±0.038 75.824±1.823	85.720±0.068 77.883±1.632	81.75 96.00	68.49s/1.58hr 20.92s/0.57hr	511999 110700	66.407±0.540 57.732±0.323	67.727±0.649 58.331±0.342	77.75 67.25	47.82s/1.05hr 14.17s/0.27hr
	16	526990	78.271 ± 0.186	90.212±0.476	53.50	50.33s/0.77hr	527874	70.587±0.447	76.074±1.362	73.50	35.94s/0.75hr
GatedGCN	16	104003 502223	84.480±0.122 85.568±0.088	84.474±0.155 86.007±0.123	78.75 65.25	139.01s/3.09hr 644.71s/11.91hr	104355 502615	60.404±0.419 73.840±0.326	61.618±0.536 87.880±0.908	94.50 60.00	79.97s/2.13hr 400.07s/6.81hr
GatedGCN-PE	16	502457	86.508±0.085	86.801±0.133	65.75	647.94s/12.08hr	504253	76.082±0.196	88.919±0.720	57.75	399.66s/6.58hr
GIN	4	100884	85.590±0.011	85.852±0.030	93.00	15.24s/0.40hr	103544	58.384±0.236	59.480±0.337	74.75	10.71s/0.23hr
n: can:	16 2	508574 105206	85.387±0.136 86.245±0.013	85.664±0.116 86.118±0.034	86.75 75.00	25.14s/0.62hr 573.37s/12.17hr	517570 104746	64.716±1.553 42.418±20.063	65.973±1.816 42.520±20.212	80.75 74.50	20.67s/0.47hr 428.24s/8.79hr
RingGNN	2	504766	86.244±0.025	86.105±0.021	72.00	595.97s/12.17hr	524202	22.340±0.000	22.304±0.000	43.25	501.84s/6.22hr
	8	505749	Diverged	Diverged	Diverged	Diverged	514380	Diverged	Diverged	Diverged	Diverged
3WLGNN	3	103572 502872	85.661±0.353 85.341±0.207	85.608±0.337 85.270±0.198	95.00 81.75	304.79s/7.88hr 424.23s/9.56hr	105552 507252	57.130±6.539 55.489±7.863	57.404±6.597 55.736±8.024	116.00 66.00	219.51s/6.52hr 319.98s/5.79hr
	8	581716	Diverged	Diverged	Diverged	Diverged	586788	Diverged	Diverged	Diverged	Diverged
						GRAPH CLAS	SIFICATIO	N			
Model	L	#Param	Test Acc.±s.d.	MNIST Train Acc.±s.d.	#Epoch	Epoch/Total	#Param	Test Acc.±s.d.	CIFAR10 Train Acc.±s.d.	#Epoch	Epoch/Total
MLP	4	104044	95.340±0.138	97.432±0.470	232.25	22.74s/1.48hr	104380	56.340±0.181	65.113±1.685	185.25	29.48s/1.53hr
GCN	4	101365	90.705±0.218	97.196±0.223	127.50	83.41s/2.99hr	101657	55.710±0.381	69.523±1.948	142.50	109.70s/4.39hr
GraphSage MoNet	4	104337	97.312±0.097 90.805±0.032	100.000±0.000 96.609±0.440	98.25 146.25	113.12s/3.13hr 93.19s/3.82hr	104517	65.767±0.308 54.655±0.518	99.719±0.062 65.911±2.515	93.50 141.50	124.61s/3.29hr 97.13s/3.85hr
GAT	4	110400	95.535±0.205	99.994±0.008	104.75	42.26s/1.25hr	110704	64.223±0.455	89.114±0.499	103.75	55.27s/1.62hr
GatedGCN	4	104217	97.340 ± 0.143	100.000 ± 0.000	96.25	128.79s/3.50hr	104357	67.312 ± 0.311	94.553±1.018	97.00	154.15s/4.22hr
GIN	4	105434	96.485±0.252	100.000 ± 0.000	128.00	39.22s/1.41hr	105654	55.255±1.527	79.412±9.700	141.50	52.12s/2.07hr
RingGNN	2	105398 505182	11.350±0.000 91.860±0.449	11.235±0.000 92.169±0.505	14.00 16.25	2945.69s/12.77hr 2575.99s/12.63hr	105165 504949	19.300±16.108 39.165±17.114	19.556±16.397 40.209±17.790	13.50 13.75	3112.96s/13.00hr 2998.24s/12.60hr
	8	506357	Diverged	Diverged	Diverged	Diverged	510439	Diverged	Diverged	Diverged	Diverged
3WLGNN	3	108024 501690	95.075±0.961 95.002±0.419	95.830±1.338 95.692±0.677	27.75 26.25	1523.20s/12.40hr 1608.73s/12.42hr	108516 502770	59.175±1.593 58.043±2.512	63.751±2.697 61.574±3.575	28.50 20.00	1506.29s/12.60hr 2091.22s/12.55hr
	8	500816	Diverged	Diverged	Diverged	Diverged	501584	Diverged	Diverged	Diverged	Diverged
		1				LINK PRE	DICTION				
				TSP			#Param		COLLAB		
Model	L	#Param	Test F1±s.d.	Train F1±s.d.	#Epoch	Epoch/Total	(L=3)	Test Hits±s.d.	Train Hits±s.d.	#Epoch	Epoch/Total
MLP	4	96956	0.544±0.001	0.544±0.001	164.25	50.15s/2.31hr	39441	20.350±2.168	29.807±3.360	147.50	2.09s/0.09hr
GCN GraphSage	4	95702 99263	0.630±0.001 0.665±0.003	0.631±0.001 0.669±0.003	261.00 266.00	152.89s/11.15hr 157.26s/11.68hr	40479 39856	50.422±1.131 51.618±0.690	92.112±0.991 99.949±0.052	122.50 152.75	351.05s/12.04hr 277.93s/11.87hr
MoNet	4	99007	0.641±0.002	0.643±0.002	282.00	84.46s/6.65hr	39751	36.144±2.191	61.156±3.973	167.50	26.69s/1.26hr
GAT	4	96182	0.671±0.002	0.673 ± 0.002	328.25	68.23s/6.25hr	42637	51.501±0.962	97.851±1.114	157.00	18.12s/0.80hr
GatedGCN	4	97858	0.791 ± 0.003	0.793 ± 0.003	159.00	218.20s/9.72hr	40965 41889	52.635±1.168 52.849±1.345	96.103±1.876	95.00	453.47s/12.09hr
GatedGCN-PE GatedGCN-E	4	97858	0.808±0.003	0.811±0.003	197.00	218.51s/12.04hr	41889	49.212±1.560	96.165±0.453 88.747±1.058	94.75 95.00	452.75s/12.08hr 451.21s/12.03hr
GatedGCN-E	16	500770	0.838 ± 0.002	0.850 ± 0.001	53.00	807.23s/12.17hr			-		
GIN	4	99002	0.656 ± 0.003	0.660 ± 0.003	273.50	72.73s/5.56hr	39544	41.730±2.284	70.555±4.444	140.25	8.66s/0.34hr
RingGNN	2	106862	0.643 ± 0.024	0.644 ± 0.024	2.00	17850.52s/17.19hr	-	OOM			
	2	507938		0.705 ± 0.003	3.00			OOM			
	2 2 8	507938 506564	0.704±0.003 Diverged	0.705±0.003 Diverged	3.00 Diverged	12835.53s/16.08hr Diverged	-	OOM OOM		and 3WLGN	
3WLGNN	8	506564 106366	0.704±0.003 Diverged 0.694±0.073	Diverged 0.695±0.073	Diverged 2.00	Diverged 17468.81s/16.59hr	-	OOM OOM	dense tenso	ors which lea	ds to OOM
3WLGNN	8	506564	0.704±0.003 Diverged	Diverged	Diverged	Diverged	-	OOM	dense tenso		ds to OOM
k-NN Heuristic	8 3 3 8	506564 106366 506681	0.704±0.003 Diverged 0.694±0.073 0.288±0.311	Diverged 0.695±0.073 0.290±0.312	Diverged 2.00 2.00	Diverged 17468.81s/16.59hr 17190.17s/16.51hr	- - - -	OOM OOM OOM	dense tense on both GF	ors which lea	ads to OOM memory.
	8 3 3	506564 106366 506681 508832	0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693	Diverged 0.695±0.073 0.290±0.312 OOM	Diverged 2.00 2.00 OOM	Diverged 17468.81s/16.59hr 17190.17s/16.51hr	60546561	OOM OOM OOM	dense tenso	ors which lea	ds to OOM
k-NN Heuristic	8 3 3 8	506564 106366 506681 508832	0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693	Diverged 0.695±0.073 0.290±0.312	Diverged 2.00 2.00 OOM	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM		OOM OOM OOM OOM 44.206±0.452	dense tenso on both GF 100.000±0.000	ors which lea PU and CPU 254.33	ads to OOM memory.
k-NN Heuristic Matrix Fact. Model MLP	8 3 8 0 <i>L</i>	506564 106366 506681 508832 k = 2 #Param 108975	0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 GRA Test MAE±s.d. 0.706±0.006	Diverged 0.695±0.073 0.290±0.312 OOM - PH REGRESSION Train MAE±s.d. 0.644±0.005	Diverged 2.00 2.00 OOM -ZINC #Epoch 116.75	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr	Evaluati	OOM OOM OOM OOM 44.206±0.452	dense tenso on both GF 100.000±0.000 er is better, except f	ors which lead CPU and CPU 254.33	ds to OOM memory. 2.66s/0.21hr
k-NN Heuristic Matrix Fact.	8 3 8 0 <i>L</i> 4	506564 106366 506681 508832 k = 2 #Param 108975 103077	0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 GRA Test MAE±s.d. 0.706±0.006 0.459±0.006	Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011	Diverged 2.00 2.00 OOM - ZINC #Epoch 116.75 196.25	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr	Evaluati	OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high TER, PATTERN u	dense tenso on both GF 100.000±0.000	254.33 for ZINC) cy w.r.t. the	ds to OOM memory. 2.66s/0.21hr
k-NN Heuristic Matrix Fact. Model MLP GCN	8 3 8 0 <i>L</i>	506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977	0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 GRA Test MAE±s.d. 0.706±0.006	Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.251±0.004	Diverged 2.00 2.00 OOM - ZINC #Epoch 116.75 196.25 197.00 147.25	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr	Evaluati • CLUS • MNIS	OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high TER, PATTERN u T, CIFAR10 use m	dense tenso on both GF 100.000±0.000 er is better, except f se weighted accura	254.33 for ZINC) by w.r.t. the stion accuracy	ds to OOM memory. 2.66s/0.21hr
k-NN Heuristic Matrix Fact. Model MLP GCN GraphSage	8 3 8 0 <i>L</i> 4 16 4 16	506564 106366 506681 508832 k =2 #Param 108975 103077 505079 94977 505341	0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 Test MAE±s.d. 0.706±0.006 0.459±0.006 0.367±0.011 0.468±0.003 0.398±0.002	Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.251±0.004 0.081±0.009	Diverged 2.00 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25 145.50	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr 12.78s/0.71hr 3.74s/0.15hr 16.61s/0.68hr	Evaluati CLUS MNIS TSP u	OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high ITER, PATTERN u T, CIFAR10 use m ses binary F1 score AB uses Hits @50	dense tense on both GF 100.000±0.000 er is better, except f se weighted accura- ulti-label classificat of the positive ed- via the evaluator pr	254.33 for ZINC) cy w.r.t. the cition accuracy ges.	2.66s/0.21hr
k-NN Heuristic Matrix Fact. Model MLP GCN	8 3 8 0 L 4 16 4 16 4	506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505341 106002	0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 GRA Test MAE±s.d. 0.706±0.006 0.459±0.006 0.357±0.011 0.468±0.003 0.398±0.002	Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.251±0.004 0.081±0.009 0.318±0.016	Diverged 2.00 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25 145.50 188.25	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr 12.78s/0.71hr 3.74s/0.15hr 16.61s/0.68hr 1.97s/0.10hr	Evaluati CLUS MNIS TSP u	OOM OOM OOM 44.206±0.452 ion Metrics: (high TER, PATTERN u T, CIFAR10 use m ses binary F1 score	dense tense on both GF 100.000±0.000 er is better, except f se weighted accura- ulti-label classificat of the positive ed- via the evaluator pr	254.33 for ZINC) cy w.r.t. the cition accuracy ges.	2.66s/0.21hr
k-NN Heuristic Matrix Fact. Model MLP GCN GraphSage	8 3 3 8 0 <i>L</i> 4 16 4 16 4 16 4	506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505341 106002 504013 102385	0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 GRA Test MAE±s.d. 0.706±0.006 0.459±0.006 0.459±0.006 0.397±0.011 0.292±0.006 0.397±0.010 0.292±0.006	Diverged 0.695±0.073 0.290±0.312 OOM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.251±0.004 0.081±0.009 0.318±0.0016 0.093±0.014 0.033±0.011	Diverged 2.00 2.00 OOM -ZINC #Epoch 116.75 196.25 197.00 147.25 145.50 188.25 171.75 137.50	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.8ss/0.16hr 12.78s/0.71hr 3.74s/0.15hr 16.61s/0.68hr 1.97s/0.10hr 10.82s/0.52hr 2.93s/0.10hr	Evaluati CLUS MNIS TSP u COLL ZINC Notation	OOM OOM OOM OOM 44.206±0.452 Ion Metrics: (high ITER, PATTERN u T, CIFAR10 use m ses binary F1 score AB uses Hits@50 uses mean absolut 1:	dense tense on both GF 100.000±0.000 er is better, except f se weighted accura- ulti-label classificat for the positive ed via the evaluator pr e regression error.	254.33 or ZINC) cy w.r.t. the tion accuracy ges. ovided by O	2.66s/0.21hr 2.66s/0.21hr class sizes.
k-NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT	8 3 3 8 0 L 4 16 4 16 4 16 4 16 4 16	506564 106366 506681 508832 k =2 #Param 108975 103077 505079 94977 505341 106002 504013 102385 531345	0.704±0.003 Diverged 0.694±0.073 0.288±0.31 0.288±0.31 OOM Test F1: 0.693 GRA 0.706±0.006 0.459±0.006 0.359±0.001 0.398±0.002 0.397±0.011 0.468±0.003 0.398±0.002 0.397±0.010 0.398±0.002 0.397±0.010 0.398±0.002 0.397±0.010	Diverged 0.695±0.073 0.290±0.312 0OM PH REGRESSION Train MAE±s.d. 0.644±0.005 0.333±0.011 0.128±0.019 0.251±0.004 0.081±0.009 0.318±0.010 0.093±0.014 0.317±0.006 0.093±0.014 0.317±0.006 0.007±0.004	Diverged 2.00 2.00 OOM ZINC #Epoch 116.75 196.25 197.00 147.25 145.50 188.25 171.75 137.50 144.00	Diverged 1746.8.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.289s/0.16hr 12.78s/0.71hr 3.74s/0.15hr 10.81s/0.03hr 1.97s/0.10hr 1.98s/0.52hr 1.97s/0.10hr 1.98s/0.53hr 1.97s/0.10hr 1.98s/0.53hr 1.9	Evaluati CLUS MNIS TSP u COLL ZINC Notation Mode	OOM OOM OOM OOM 44.206±0.452 Ion Metrics: (high ITER, PATTERN u T, CIFAR10 uses be innar P1 sees be innar P1 sees See binary P1 sees AB uses Hits@50 uses mean absolut 12: Is with the suffix -I	dense tense on both GF 100.000±0.000 er is better, except f se weighted accurately the decurate of the positive ed via the evaluator pre regression error. E use input edge fea	254.33 for ZINC) cy w.r.t. the rition accuracy ges. ovided by O	2.66s/0.21hr 2.66s/0.21hr class sizes.
k-NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN	8 3 3 8 0 <i>L</i> 4 16 4 16 4 16 4	506564 106366 506681 508832 k =2 #Param 108975 103077 505341 106002 504013 102385 531345 105735	0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 Test MAE±s.d. 0.706±0.006 0.459±0.006 0.359±0.001 0.367±0.011 0.468±0.003 0.398±0.002 0.397±0.010 0.292±0.006 0.475±0.007 0.384±0.007 0.384±0.007	Diverged 0.695±0.073 0.290±0.312 0.004 0.312 0.00 M PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.251±0.004 0.081±0.009 0.318±0.016 0.093±0.014 0.317±0.006 0.067±0.004 0.287±0.014	Diverged 2.00 COM	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.05hr 2.89s/0.16hr 12.78s/0.71hr 16.61s/0.68hr 16.9s/0.52hr 16.9s/0.05hr 10.82s/0.52hr 12.9s/0.01hr 12.9s/0.01hr 12.9s/0.05hr 15.9s/0.05hr 15.9	Evaluati CLUS MNIS TSP u COLL ZINC Notation Mode repres	OOM OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high ITER, PATTERN u T, CIFAR10 use m sees binary F1 scoro AB uses Hits@50 uses mean absolut It is with the suffix -I entations (ZINC: b	dense tens on both GF 100.000±0.000 er is better, except f se weighted accura ulti-label classificat e for the positive ed, via the evaluator pr e regression error. E use input edge fee ond type, TSP. Euc	254.33 for ZINC) cy w.r.t. the rition accuracy ges. ovided by O	2.66s/0.21hr 2.66s/0.21hr class sizes.
k-NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN GatedGCN-E	8 3 3 8 0 L 4 16 4 16 4 16 4 16 4 16 4 16 4 16 4	506564 106366 506681 508832 k =2 #Param 108975 103077 505079 94977 505301 106002 504013 102385 105735 105735 105735 105735	0.704±0.003 Diverged 0.694±0.073 Diverged 0.694±0.0731 OOM Test F1: 0.693 Test MAE±s.d. 0.706±0.006 0.367±0.011 0.468±0.003 0.398±0.002 0.397±0.010 0.292±0.006 0.475±0.007 0.348±0.007 0.348±0.007 0.357±0.010 0.292±0.006	Diverged 0.695±0073 0.290±0.312 0.004±0.312 0.00M PPH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.318±0.009 0.318±0.009 0.318±0.000 0.007±0.004 0.001±0.009 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.006 0.007±0.004 0.007±0.004 0.007±0.004 0.007±0.006 0.007±0.004 0.007±0.006 0.007±0.004 0.007±0.006 0.007±0.004 0.007±0.006 0.007±0.004 0.007±0.004 0.007±0.006 0.007±0.004 0.007±0.006 0.007±0.004 0.007±0.006 0.0000 0.000	Diverged 2.00 COM 2.0	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr 12.78s/0.71hr 3.7sk/0.15hr 10.82s/0.50hr 10.82s/0.50hr 12.98s/0.51hr 12.98s/0.53hr 12.98s/0.53hr 12.98s/0.53hr 12.98s/0.53hr 2.93s/0.50hr 2.93s/0.50hr 2.93s/0.50hr 2.93s/0.50hr 2.93s/0.50hr 2.93s/0.50hr 2.93s/0.50hr	Evaluati CLUS MNIS TSP u COLL ZINC Notation Mode repres collab Model	OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high TER, PATTERN u ses ses binary F1 score AB uses Hits@50 uses mean absolut ti sis with the suffix -1 entations (ZINC: b oration frequency1 s with the suffix -1	dense tens on both GF 100.000±0.000 er is better, except f se weighted accura- ulti-label classificat for the positive ed via the evaluator pr or ergression error. E use input edge fea ond type, TSP: Euc and year).	254.33 or ZINC) cy w.r.t. the tion accuracy ges. vovided by O tures to initilidean distar genvectors a	ds to OOM memory. 2.66s/0.21hr class sizes. GB [?]. alize edge tee, COLLAB: s node positional
k-NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN GatedGCN-E	8 3 3 8 0 L 4 16 4 16 4 16 4 16 4 16 4 4 16 4 4 16 4 4 16 4 4 4 16 4 4 4 4	506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505341 106002 504013 102385 531345 1057875	0.704±0.003 Diverged 0.694±0.073 0.288±0.31 0.288±0.31 0.288±0.31 0.288±0.31 0.706±0.006 0.479±0.006 0.459±0.006 0.367±0.011 0.468±0.003 0.398±0.002 0.397±0.010 0.492±0.006 0.475±0.007 0.354±0.017 0.354±0.017 0.354±0.017	Diverged 0.695±0.073 0.290±0.312 0.000 ±0.312 0.000 ±0.312 0.000 ±0.312 0.000 ±0.64±0.000 0.64±0.000 0.38±0.010 0.128±0.019 0.081±0.009 0.0318±0.016 0.093±0.014 0.317±0.006 0.093±0.014 0.317±0.006 0.067±0.004 0.287±0.014 0.236±0.007	Diverged 2.00 2.00 OOM ** - ZINC **Epoch** 116.75** 196.25** 197.00 147.25** 145.50** 188.25** 171.75** 137.50** 144.00** 173.50** 194.75	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr 12.78s/0.71hr 3.74s/0.15hr 1.97s/0.00hr	Evaluati CLUS MNIS TSP u COLL ZINC Notation Mode repres collab Model encod	OOM OOM OOM OOM 44.206±0.452 ION Metrics: (high TIER, PATTERN u T, CIFAR10 use m ses binary F1 score Susses mean absolut Is with the suffix -I st with the suffix -I s with the suffix -I	dense tens on both GF 100.000±0.000 er is better, except f se weighted accura- uthi-label classificate ovia the evaluator pre e regression error. E use input edge fea ond type, TSP: Euc and year). "Et use Laplacian Ei ns for ZINC, 2 for	254.33 or ZINC) cy w.r.t. the clion accuracy ges. tures to initilidean distar genvectors a PATTERN	ds to OOM memory. 2.66s/0.21hr class sizes. (GB [?]. alize edge tee, COLLAB: s node positional and 20 for others.
k-NN Heuristic Matrix Fact Model MLP GCN GraphSage MoNet GAT GatedGCN- GatedGCN-E-PE GIN	8 3 3 8 0 0 L 4 16 4 16 4 16 4 16 4 4 16 4 16 4 16	506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505341 106002 504013 102385 531345 105735 105875 504309 505011 103079 509549	0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test F1: 0.693 Test MAE±s.d. 0.706±0.006 0.367±0.001 0.489±0.006 0.367±0.010 0.292±0.006 0.367±0.010 0.292±0.006 0.375±0.000 0.395±0.0000 0.395±0.000 0.395±0.000 0.395±0.00	Diverged 0.695±0.073 0.290±0.312 0.00M PH REGRESSION Train M.E±s.d. 0.333±0.011 0.253±0.001 0.251±0.004 0.081±0.009 0.318±0.016 0.093±0.014 0.07±0.0067±0.004 0.087±0.004 0.097±0.004 0.004 0.004 0.004 0.004 0.004 0.004 0.004 0.004 0.004 0.004 0.004 0.004 0.0	Diverged 2.00 2.00 OOM 2.00 OO	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epach/Total 10130.03hr 12.28s0.016hr 12.28s0.016hr 12.78s0.071hr 16.61s0.68hr 16.97s0.00hr 12.98s0.03hr 15.97s0.00hr 10.82s0.52hr 15.97s0.00hr 10.82s0.52hr 10.97s0.00hr 10.97s0.00hr 10.97s0.00hr 10.97s0.00hr 10.97s0.00hr 10.97s0.00hr	Evaluati CLUS MNIS TSP u COLL ZINC Notation Mode repres collab Model encod Result	OOM OOM OOM OOM 44.206±0.452 ion Metrics: (high TER, PATTERN u TC HFAR10 use m ses binary F1 score AB uses His @50 uses mean absolut ti st with the suffix -1 entations (ZINC: b oration frequency to with dimensis to denoted by Dives	dense tens on both GF 100.000±0.000 er is better, except f se weighted accura- utifi-label classificated via the evaluator pt e regression error. Z use input edge fea ond type, TSP. Euc and year.). PE use Laplacian Ei na 8 for ZINC, 2 for gred indicate unstal	254.33 or ZINC) ey w.rt. the cition accuracy ges. ovided by O tures to initilidean distar genvectors a PATTERN	ads to OOM memory. 2.66s/0.21hr class sizes. GB [?]. alize edge cc, COLLAB: s node positional and 20 for others. gent runs across
k-NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN-E GatedGCN-E GatedGCN-E-PE GIN RingGNN	8 3 3 8 0 L 4 4 16 4 16 4 16 4 4 16 4 16 4 16 16 16 16 16 16 16 16 16 16 16 16 16	506564 106366 506681 508832 k = 2 #Param 108975 103077 505079 94977 505341 106002 504013 102385 504013 10375 504013 103875 504019 103079 509549 97978	0.704±0.003 Diverged 0.694±0.073 Diverged 0.694±0.073 C288±0.311 OOM Test F1: 0.693 GRA Test MAE±s.od. 0.705±0.006 0.439±0.006 0.359±0.006 0.359±0.006 0.398±0.002 0.397±0.010 0.408±0.003 0.398±0.002 0.397±0.010 0.475±0.007 0.435±0.011 0.484±0.003 0.375±0.005 0.214±0.013 0.375±0.015 0.356±0.051 0.356±0.051 0.356±0.051 0.356±0.051 0.512±0.003	Diverged 0.695±0.073 0.290±0.312 0.000 0.7	Diverged 2.00 2.00 OOM 2.00 OOM ** - ZINC **Epoch** 116.75 196.25 197.00 147.25 197.01 147.25 145.50 147.25 145.50 173.50 194.75 166.75 185.00 90.25 147.00 90.25	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM Epoch/Total 1.01s/0.03hr 2.89s/0.16hr 12.78s/0.71hr 16.61s/0.68hr 10.92s/0.52hr 16.61s/0.68hr 2.98s/0.11hr 10.92s/0.53hr 2.98s/0.31hr 2.98s/0.31hr 2.98s/0.35hr 1.052s/0.35hr 2.98s/0.35hr 1.052s/0.35hr 2.058s/0.36hr 2.058s/0.35hr 2.058s/0.36hr 2.058s/0.35hr 2.058s/0.36hr 2.058s/0.35hr 2.058s/0.36hr 2.058s/0.35hr	Evaluati CLUS MNIS TSP u COLL ZINC Notation Mode repress collab Model encod Result all 4 s	OOM OOM OOM OOM 44.206±0.452 TOM Metries: (high TIER, PATTERN u T, CIFAR10 use m sees binary F1 score AB uses Hists@50 uses mean absolut I: Is with the suffix -I entations (ZINC: b Is with the suffix -I ings, with dimens, with dimens, sis denoted by Dive eds and initial lea	dense tenss on both GF 100.000±0.000 er is better, except f se weighted accura- ulti-label classificate of or the positive ed via the evaluator pr e regression error. E use input edge fea ond type, TSP: Euc and year). "E use Laplacian Ei ns 8 for ZINC, 2 for rged indicate unuss {	254.33 or ZINC) cy w.r.t. the tion accuracy ges. ovided by O tures to initilidean distar genvectors a part accuracy genvectors a part accuracy genvectors a part accuracy genvectors a part accuracy genvectors and of the genvectors are genvectors and of the genvectors and of the genvectors are genvectors and of the genvectors and of the genvectors are genvectors and genvectors are genvectors are genvectors.	2.66s/0.21hr 2.66s/0.21hr class sizes. GB [?]. alize edge cec, COLLAB: s node positional and 20 for others. gent runs across 0^-5}.
k-NN Heuristic Matrix Fact Model MLP GCN GraphSage MoNet GAT GatedGCN- GatedGCN-E-PE GIN	8 3 3 8 0 L 4 16 4 16 4 16 4 16 4 16 16 4 16 16 2 2 2 2 2	506564 106366 506681 508832 k = 2 #Param 108975 103077 505379 94977 505371 106002 504013 102385 531345 105875 504309 505011 103079 509549 97978 104403 527283	0.704±0.003 Diverged 0.694±0.073 OCM Test F1: 0.693 Test MAE±s.4. 0.706±0.006 0.499±0.007 0.498±0.003 0.398±0.002 0.397±0.010 0.292±0.006 0.475±0.007 0.398±0.002 0.397±0.010 0.292±0.006 0.375±0.003 0.388±0.002 0.375±0.003 0.388±0.002 0.375±0.003 0.388±0.002 0.375±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003 0.388±0.003	Diverged 0.695±0.073 0.290±0.312 0.000±0.312 0.000 ±0.312 0.000 ±0.312 0.000 ±0.312 0.000 ±0.313±0.001 0.138±0.019 0.251±0.004 0.081±0.009 0.318±0.010 0.317±0.006 0.007±0.004 0.287±0.014 0.037±0.001 0.007±0.001 0.007±0.010 0.007±0.007	Diverged 2.00 OOM 2.0	Diverged 17468.8 1st 316.59hr 17190.17st 16.51hr OOM 17st 16.51hr 17st 16.51hr 16.51hr 16.51hr 18st 17st 16.51hr 18st 17st 18st 17st 18st 17st 18st 17st 18st 18st 18st 18st 18st 18st 18st 18	Evaluati CLUS MNIS TSP u COLL ZINC Notation Mode repres collabal Mode encod Result al 4 s Result	OOM OOM OOM OOM 44.206±0.452 TOM Metries: (high TIER, PATTERN u T, CIFAR10 use m sees binary F1 score AB uses Hists@50 uses mean absolut I: Is with the suffix -I entations (ZINC: b Is with the suffix -I ings, with dimens, with dimens, sis denoted by Dive eds and initial lea	dense tens on both GF 100.000±0.000 er is better, except f se weighted accura ulti-label classificat for the positive ed via the evaluator pr e regression error. E use input edge fea ond type, TSP: Euc and year.) E'E use Laplacian Ei no 8 for ZINC, 2 for gred indicate unsel	254.33 or ZINC) cy w.r.t. the tion accuracy ges. ovided by O tures to initilidean distar genvectors a part accuracy genvectors a part accuracy genvectors a part accuracy genvectors a part accuracy genvectors and of the genvectors are genvectors and of the genvectors and of the genvectors are genvectors and of the genvectors and of the genvectors are genvectors and genvectors are genvectors are genvectors.	2.66s/0.21hr 2.66s/0.21hr class sizes. GB [?]. alize edge cec, COLLAB: s node positional and 20 for others. gent runs across 0^-5}.
k-NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN-E GatedGCN-E GatedGCN-E-CIBC GIN RingGNN-E	8 3 3 8 0 0 L 4 16 4 16 4 16 4 16 4 16 16 4 16 16 2 2 2 2 2 8	506564 106366 506681 508832 k = 2 #Param 108975 103077 5053041 106002 505341 102385 531345 105735 105735 504309 505011 103079 509549 97978 104403 527283 510345	0.704-0.003 Diverged 0.694-0.073 0.288-0.311 OOM Test F1: 0.693 Test MaE+s.d. 0.786-0.006 0.459-0.006 0.459-0.006 0.359-0.006	Diverged 0.695±0.073 0.290±0.312 0.000±0.312 0.000 0.312 0.000 0.312 0.000 0.343±0.011 0.128±0.010 0.003±0.010 0.003±0.010 0.007±0.010 0.00000 0.00000 0.00000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.00000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.00000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.000000	Diverged 2.00 2.00 OOM 2.00 OO	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM 17s/16.51hr OOM 1.00 17s/16.51hr OOM 1.10s/0.25hr 12.78s/0.16hr 12.78s/0.16hr 12.78s/0.71hr 12.78s/0.71hr 12.78s/0.71hr 12.78s/0.71hr 19.78s/0.78hr 1.97s/0.10hr 1.97s/0.78hr 1.9	Evaluati CLUS MNIS TSP u COLL ZINC Notation Mode repres collabal Mode encod Resultal 4 s Result on onu	OOM	dense tense on both GF 100.000±0.000 er is better, except f se weighted accura utilit-label classificat for the positive ed via the evaluator pr e regression error. 2 use input edge fea cond type, TSF use Laplacian Ei on 8 for ZINC, 2 for ged indicate unit pred indicate unit pred indicate unit pred indicate unit which is the condition of the	254.33 or ZINC) 254.33 or ZINC) cy w.r.t. the consequence of the c	ds to OOM memory. 2.66s/0.21hr class sizes. GB [7]. alize edge cce, COLLAB: s node positional and 20 for others. gent runs across 0^-5].
k-NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN-E GatedGCN-E-PE GIN RingGNN RingGNN-E 3WLGNN	8 3 3 8 0 0 L 4 4 16 4 16 4 16 4 16 4 16 16 4 16 16 16 4 16 16 16 16 16 16 16 16 16 16 16 16 16	506564 106366 506681 508832 k = 2 #Param 108975 103077 505341 106002 504013 102385 531345 105875 504309 505511 103079 509549 97978 509549 97978 509549 50954	0.704±0.003 Diverged 0.694±0.073 0.288±0.311 OOM Test FI: 0.693 Test MAE±sd. 0.706±0.000 0.495±0.006 0.495±0.006 0.495±0.006 0.495±0.006 0.398±0.002 0.398±0.002 0.398±0.002 0.398±0.002 0.392±0.001 0	Diverged 0.695±0.073 0.290±0.312 0.00M PH REGRESSION Train MAE±s.d. 0.644±0.005 0.343±0.011 0.128±0.019 0.318±0.004 0.081±0.009 0.318±0.016 0.093±0.014 0.087±0.004 0.087±0.0	Diverged 2.00 OOM 2.0	Diverged 17468.8 1st 316.59hr 17190.17st 16.51hr OOM Epoch/Total 1.01st/0.35hr 1.01st	Evaluati CLUS MNIS MNIS TSP u COLL ZINC Notation Mode repres collab Mode encod Result all 4 s Result on on our	OOM	dense tense on both GF 100.000±0.000 er is better, except If se weighted accurate for the positive ed visit head examinate is the evaluator pr ergession error. Euse input edge fea ond type, TSP. Euc and year. Euse Input edge fre greating the evaluator is for TSP. Euc and year). Euse Laplacian Ei s for ZINC, ZiOn ingra tev aluses [III] II indicate unsus which the company to the company to the company of the company to the company to the company of the company to the company to the company of the company to the company to the company of the company to the company to the company of the company to the company to the company of the company to the company to the company of the company to the company to the company to the company of the company to	254.33 or ZINC) 254.33 or ZINC) cy w.r.t. the control of the c	das to OOM memory. 2.66s/0.21hr class sizes. GB [?]. alize edge tec, COLLAB: s node positional and 20 for other of the control of the con
k-NN Heuristic Matrix Fact. Model MLP GCN GraphSage MoNet GAT GatedGCN-E GatedGCN-E GatedGCN-E-PE GIN RingGNN-E	8 3 3 8 0 0 L 4 16 4 16 4 16 4 16 4 16 16 4 16 16 2 2 2 2 2 8	506564 106366 506681 508832 k = 2 #Param 108975 103077 5053041 106002 505341 102385 531345 105735 105735 504309 505011 103079 509549 97978 104403 527283 510345	0.704-0.003 Diverged 0.694-0.073 0.288-0.311 OOM Test F1: 0.693 Test MaE+s.d. 0.786-0.006 0.459-0.006 0.459-0.006 0.359-0.006	Diverged 0.695±0.073 0.290±0.312 0.000±0.312 0.000 0.312 0.000 0.312 0.000 0.343±0.011 0.128±0.010 0.003±0.010 0.003±0.010 0.007±0.010 0.00000 0.00000 0.00000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.00000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.00000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.000000	Diverged 2.00 2.00 OOM 2.00 OO	Diverged 17468.81s/16.59hr 17190.17s/16.51hr OOM 17s/16.51hr OOM 1.00 17s/16.51hr OOM 1.10s/0.25hr 12.78s/0.16hr 12.78s/0.16hr 12.78s/0.71hr 12.78s/0.71hr 12.78s/0.71hr 12.78s/0.71hr 19.78s/0.78hr 1.97s/0.10hr 1.97s/0.78hr 1.9	Evaluati CLUS MNIS MNIS TSP u COLL ZINC Notation Mode repres collab Mode encod Result all 4 s Result on on our	OOM	dense tense on both GF 100.000±0.000 er is better, except f se weighted accura utilit-label classificat for the positive ed via the evaluator pr e regression error. 2 use input edge fea cond type, TSF use Laplacian Ei on 8 for ZINC, 2 for ged indicate unit pred indicate unit pred indicate unit pred indicate unit which is the condition of the	254.33 or ZINC) 254.33 or ZINC) cy w.r.t. the control of the c	das to OOM memory. 2.66s/0.21hr class sizes. GB [?]. alize edge tec, COLLAB: s node positional and 20 for other of the control of the con

- Motivation
- Message-Passing GCNs
- Weisfeiler-Lehman GNNs
- Graph-Agnostic GNNs
- Datasets
- Infrastructure and Experimental Setting
- Benchmarking Results
- Laplacian Positional Encodings
- Link Prediction with Edge Representation
- Conclusion

Structural (GCNs) vs Positional Encodings

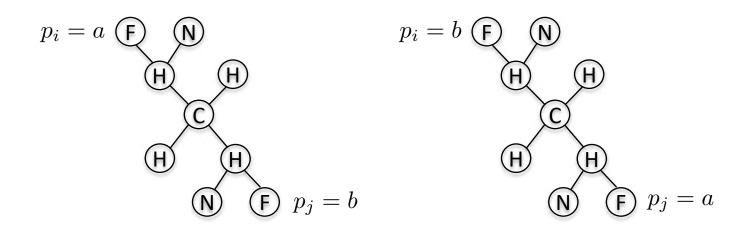
- GCNs like GINs^[1] are not able to differentiate isomorphic nodes, i.e. nodes with the same neighborhood structure :
 - All F atoms have the same representation.
 - All N atoms have the same representation.
 - Two H atoms have the same representation.
 - Two H atoms have the same representation.
- This is a limitation of the expressivity of GCNs.
 - Can we break this structural symmetry?
 - This can be done either by
 - Higher-order WL-GNNs, but require $O(n^2)/O(n^3)$.
 - Positional encodings of nodes with O(n).



[1] Xu, Hu, Leskovec, Jegelka, How powerful are graph neural networks?, 2019

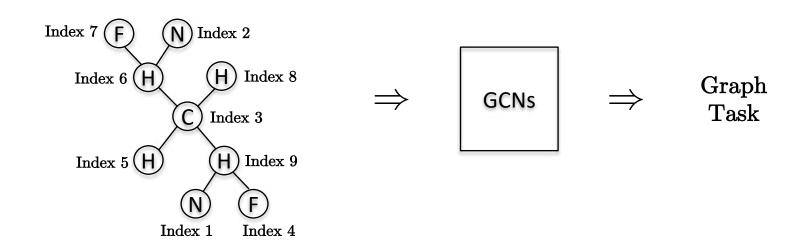
Graph Positional Encodings

- Properties of PEs:
 - Unique representation for each node.
 - Distance-sensitive: Nodes far apart on the graph should have different positional features whereas nodes nearby have similar positional features.
- Graph symmetries prevent assigning canonical representation of PEs.
 - if nodes i and j are structurally symmetric and we have PEs $p_i=a$, $p_j=b$, then it is also possible to arbitrary choose $p_i=b$, $p_j=a$.
 - PEs are always arbitrary up to the number of graph symmetries.



Index Positional Encodings

- \circ Simplest PEs are (arbitrary) ordering to the nodes, among n! possible orderings.
- Theorem^[1]: GCNs s.a. GINs^[2] are more expressive than the 1-WL test when considering (one-hot encoding of) node indices as PE features.
- During training, orderings are uniformly sampled from the n! possible choices in order for the network to learn to be independent to these arbitrary choices.



^[1] Murphy, Srinivasan, Rao, Ribeiro, Relational pooling for graph representations, 2019

^[2] Xu, Hu, Leskovec, Jegelka, How powerful are graph neural networks?, 2019

Laplacian Positional Encodings

• Laplacian PEs:

- Eigen-decomposition : $\Delta = I D^{-1/2}AD^{-1/2} = U^T\Lambda U$
- Hybrid positional and structural encodings, invariant by index permutation.
- Unique and distance-sensitive: two nodes far away on a graph have large PEs distance, and inversely (two one-encoding vectors of different indices are equally distant).
- LapPEs have also natural symmetries with the arbitrary sign of eigenvectors.
 - The number of possible sign flips is 2^k , k being the number of eigenvectors.
 - In practice, we choose $k \ll n$, and therefore 2^k is much smaller than n! (the number of possible ordering of the nodes).
 - During the training, sign of eigenvectors will be uniformly sampled at random between the 2^k possibilities.
- Lap PEs are graph generalizations of Transformer's PEs^[2].

^[1] Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, Polosukhin, Attention is all you need, 2017

Numerical Results

PE type

EigVecs-20

Abs(EigVecs)

Rand sign(EigVecs)

Fixed node ordering

Rand node ordering

Rand sign(EigVecs)

Fixed node ordering

Rand node ordering

Rand sign(EigVecs)

Fixed node ordering

Rand node ordering

Rand node ordering

ÅB

No PE

No PE

No PE

No PE

EigVecs-20

Abs(EigVecs)

EigVecs-2

Abs(EigVecs)

16

16

16

16

16

16

16

16

16

16

16

16

16

#Param

104007

105407

105407

105407

106807

106807

502223

505421

502457

505421

516887

516887

502615

504253

504253

504253

517435

517435

40965

507195

Test Acc.±s.d.

 10.000 ± 0.000

 68.633 ± 7.143

99.767±0.394

 99.433 ± 1.133

 10.533 ± 4.469

 11.133 ± 2.571

 85.605 ± 0.105

 86.029 ± 0.085

 86.508 ± 0.085

 86.393 ± 0.037

 80.133 ± 0.202

 85.767 ± 0.044

 73.684 ± 0.348

 75.520 ± 0.395

 76.082 ± 0.196

 73.796 ± 0.234

 69.232 ± 0.265

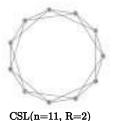
 74.656 ± 0.314

 52.635 ± 1.168

• Results: PEs with GatedGCN^[1].

Performance reported on the test sets of CSL, ZINC,

PATTERN, CLUSTER and COLLAB (higher is better, except for ZINC).







CSL(n=11, R=3)

EigVecs-20	3	41889	52.326 ± 0.678	96.700 ± 1.296	95.00	452.40s/12.10hr
Rand sign(EigVecs)	3	41889	52.849 ± 1.345	96.165 ± 0.453	94.75	452.75s/12.08hr
Abs(EigVecs)	3	41889	51.419 ± 1.109	95.984 ± 1.157	95.00	451.36s/12.07hr
PE type	L	#Param	Test MAE \pm s.d.	Train MAE±s.d.	#Epochs	Epoch/Total
No PE	16	504153	0.354 ± 0.012	0.095 ± 0.012	165.25	10.52s/0.49hr
EigVecs-8	16	505011	0.319 ± 0.010	0.038 ± 0.007	143.25	10.62s/0.43hr
Dandaian (Eightean)	1.	505011	0.01410.010	0.06510.010	10500	40 -0 10 -01
Rand sign(EigVecs)	16	505011	0.214 ± 0.013	0.067 ± 0.019	185.00	10.70s/0.56hr
Abs(EigVecs)	16	505011	$0.214\pm0.013 \ 0.214\pm0.009$	0.067 ± 0.019 0.035 ± 0.011	185.00 167.50	10.70s/0.56hr 10.61s/0.50hr
	Rand sign(EigVecs) Abs(EigVecs) PE type No PE EigVecs-8	Rand sign(EigVecs) 3 Abs(EigVecs) 3 PE type L No PE 16 EigVecs-8 16	Rand sign(EigVecs) 3 41889 Abs(EigVecs) 3 41889 PE type L #Param No PE 16 504153 EigVecs-8 16 505011	Rand sign(EigVecs) 3 Abs(EigVecs) 41889 3 41889 52.849±1.345 51.419±1.109 PE type L #Param Test MAE±s.d. No PE 16 504153 0.354±0.012 0.319±0.010 EigVecs-8 16 505011 0.319±0.010	Rand sign(EigVecs)3 41889 52.849 ± 1.345 96.165 ± 0.453 Abs(EigVecs)3 41889 51.419 ± 1.109 95.984 ± 1.157 PE typeL#ParamTest MAE \pm s.d.Train MAE \pm s.d.No PE16 504153 0.354 ± 0.012 0.095 ± 0.012 EigVecs-816 505011 0.319 ± 0.010 0.038 ± 0.007	Rand sign(EigVecs) 3 A1889 52.849±1.345 96.165±0.453 94.75 Abs(EigVecs) 3 41889 51.419±1.109 95.984±1.157 95.00 PE type L #Param Test MAE±s.d. Train MAE±s.d. #Epochs No PE 16 504153 0.354±0.012 0.095±0.012 165.25 EigVecs-8 16 505011 0.319±0.010 0.038±0.007 143.25

 0.321 ± 0.015

Train Acc.±s.d.

 10.000 ± 0.000

 99.811 ± 0.232

 99.689 ± 0.550

 100.000 ± 0.000

 76.056 ± 14.136

 10.944 ± 2.106

 85.999 ± 0.145

 86.955 ± 0.227

 86.801 ± 0.133

 87.011 ± 0.172

 98.416 ± 0.141

 85.998 ± 0.063

 88.356 ± 1.577

 89.332 ± 1.297

 88.919 ± 0.720

 91.125 ± 1.248

 92.298 ± 0.712

 82.940 ± 1.718

 96.103 ± 1.876

 0.177 ± 0.015

Epoch/Total

0.58s/0.05hr

0.59s/0.09hr

0.59s/0.16hr

0.60s/0.12hr

0.59s/0.05hr

0.60s/0.08hr

646.03s/11.36hr

645.36s/11.94hr

647.94s/12.08hr

645.90s/11.41hr

643.23s/8.27hr

645.09s/11.79hr

399.44s/6.97hr

400.50s/5.70hr

399.66s/6.58hr

398.97s/6.68hr

400.40s/5.82hr

397.75s/6.88hr

453.47s/12.09hr

10.55s/0.55hr

#Epochs

54.00

107.16

188.76

143.64

60.56

91.60

62.00

65.00

65.75

62.00

45.00

64.50

61.50

49.75

57.75

58.75

51.00

61.00

95.00

184.75

^[1] Bresson, Laurent, Residual gated graph convnets, 2017

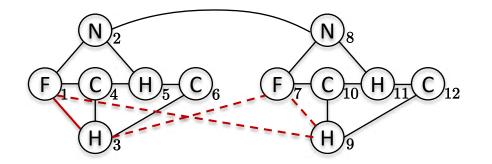
- Motivation
- Message-Passing GCNs
- Weisfeiler-Lehman GNNs
- Graph-Agnostic GNNs
- Datasets
- Infrastructure and Experimental Setting
- Benchmarking Results
- Laplacian Positional Encodings
- Link Prediction with Edge Representation
- Conclusion

Link Prediction

- GCNs may fail the link prediction task.
 - Apply any GCN to this molecule composed of two identical compounds:

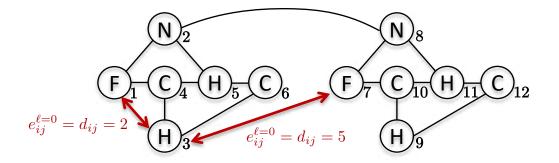
We have
$$h_1=h_7\in\mathbb{R}^d \\ h_3=h_9\in\mathbb{R}^d \qquad \boxed{\mathsf{F_1}} \ \ \mathsf{C_4} \ \ \mathsf{H_5} \ \ \mathsf{C_6} \qquad \boxed{\mathsf{F_7}} \ \ \mathsf{C_{10}} \ \ \mathsf{H_{11}} \ \mathsf{C_{12}}$$

- Perform link prediction (by transfer learning):
 - Suppose there exist a bond $(F_{(item 1)}, H_{(item 3)})$.
 - Can we predict the link $(F_{(\text{item }7)}, H_{(\text{item }9)})$? Yes, but the network will also predict (wrong) links between $(F_{(\text{item }7)}, H_{(\text{item }3)})$ and between $(F_{(\text{item }1)}, H_{(\text{item }9)})$.



Link Prediction with Edge Representation

- How to design expressive GCNs for the link prediction task?
- Theorem $^{[1]}$:
 - Link prediction is maximally expressive with a joint representation of nodes (intuitively this encodes the similarity/distance between pair of nodes).
 - This requires $O(n^2)$ edge representations.



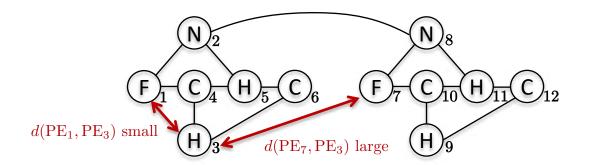
Input edge representation is $e_{ij}^{l=0} = d_{ij}$ (e.g. shortest distances).

[1] Srinivasan, Ribeiro, On the Equivalence between Positional Node Embeddings and Structural Graph Representations, 2019

Link Prediction with Edge Representation

• Alternate solution:

- Represent links as joint representations of nodes.
- Add node positional encodings that are unique and distance-sensitive to differentiate links, like Laplacian PEs.
- Complexity is O(E), E is the number of edges.



Numerical Results

- We study the impact of edge representation by instantiating three variants:
 - No edge feature/representation : Isotropic models (like vanilla GCNs^[1])

$$h_i^{\ell+1} = \sigma\left(\sum_{j \in \mathcal{N}_i} W^{\ell} h_j^{\ell}\right)$$
, identified by (E.Feat,E.Repr= x,x)

• Edge feature/no edge representation: Anisotropic models (like GAT^[2])

$$h_i^{\ell+1} = \sigma\left(\sum_{j \in \mathcal{N}_i} f_{V^{\ell}}(h_i^{\ell}, h_j^{\ell}) \cdot W^{\ell} h_j^{\ell}\right), \text{ with (E.Feat,E.Repr} = \checkmark, x)$$

• Edge feature and representation: Anisotropic models (like GatedGCNs^[3])

$$h_i^{\ell+1} = \sigma\left(\sum_{j \in \mathcal{N}_i} e_{ij}^{\ell} \cdot W^{\ell} h_j^{\ell}\right), \ e_{ij}^{\ell+1} = f_{V^{\ell}}\left(h_i^{\ell}, h_j^{\ell}, e_{ij}^{\ell}\right), \ \text{with (E.Feat, E.Repress, Text Assarch Text)}$$

	Model	E.Feat.	E.Repr.	L	#Param	Test Acc.±s.d.	Train Acc.±s.d.	#Epochs	Epoch/Total
		Х	Х	4	99026	0.646±0.002	0.648±0.002	197.50	150.83s/8.34hr
TSP	GatedGCN	\checkmark	X	4	98174	0.757 ± 0.009	0.760 ± 0.009	218.25	197.80s/12.06hr
		\checkmark	\checkmark	4	97858	0.791 ± 0.003	0.793 ± 0.003	159.00	218.20s/9.72hr
	GatedGCN-E	√	√	4	97858	$0.808 {\pm} 0.003$	0.811 ± 0.003	197.00	218.51s/12.04hr
L		X	Х	4	95462	0.643 ± 0.001	0.644 ± 0.001	132.75	325.22s/12.10hr
	GAT	\checkmark	X	4	96182	0.671 ± 0.002	0.673 ± 0.002	328.25	68.23s/6.25hr
		\checkmark	\checkmark	4	96762	0.748 ± 0.022	0.749 ± 0.022	93.00	462.22s/12.10hr
	GAT-E	√	√	4	96762	0.782 ± 0.006	0.783 ± 0.006	98.00	438.37s/12.11hr
		Х	Х	3	26593	35.989 ± 1.549	60.586 ± 4.251	148.00	263.62s/10.90h
	GatedGCN	\checkmark	X	3	26715	50.668 ± 0.291	96.128 ± 0.576	172.00	384.39s/18.44hr
AB		\checkmark	\checkmark	3	27055	51.537 ± 1.038	96.524 ± 1.704	188.67	376.67s/19.85hr
3	GatedGCN-E	√	√	3	27055	47.212 ± 2.016	85.801 ± 0.984	156.67	377.04s/16.49hr
COLL		Х	Х	3	28201	41.141 ± 0.701	70.344 ± 1.837	153.50	371.50s/15.97hr
\ddot{c}	GAT	\checkmark	X	3	28561	50.662 ± 0.687	96.085 ± 0.499	174.50	403.52s/19.69hr
		\checkmark	\checkmark	3	26676	49.674 ± 0.105	92.665 ± 0.719	201.00	349.19s/19.59hr
	GAT-E	√	√	3	26676	44.989±1.395	82.230±4.941	120.67	328.29s/11.10hr

^[1] Kipf, Welling, Semi-supervised classification with graph convolutional networks, 2017

^[2] Velickovic, Cucurull, Casanova, Romero, Lio, Bengio, Graph attention networks, 2017

^[3] Bresson, Laurent, Residual gated graph convnets, 2017

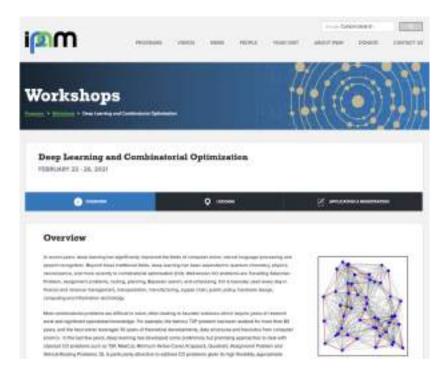
- Motivation
- Message-Passing GCNs
- Weisfeiler-Lehman GNNs
- Graph-Agnostic GNNs
- Datasets
- Infrastructure and Experimental Setting
- Benchmarking Results
- Laplacian Positional Encodings
- Link Prediction with Edge Representation
- Conclusion

Conclusion

- Take-home $messages^{[1]}$:
 - MP-GCNs outperformed WL-GNNs for GINs, 3WL-GNNs, RingGNNs on the 8 datasets used in this benchmark.
 - MP-GCNs benefit from graph sparsity, and universal building blocks w/ batch normalization and residual connection.
 - Anisotropic mechanism improve (isotropic) GCNs.
 - Laplacian eigenvectors improve index positional encodings.
 - Edge representation with positional encodings enhance link prediction.
- Recent works have focused on more efficient WL-inspired techniques^[2,3,4,5,6].
- Benchmarking will bridge the gap between theory and practical performances.

- [1] Dwivedi, Joshi, Laurent, Bengio, Bresson, Benchmarking graph neural networks, 2020, https://arxiv.org/pdf/2003.00982.pdf
- [2] Bouritsas, Frasca, Zafeiriou, Bronstein, Improving Graph Neural Network Expressivity via Subgraph Isomorphism Counting, 2020
- [3] Puny, Ben-Hamu, Lipman, From Graph Low-Rank Global Attention to 2-FWL Approximation, 2020
- [4] Morris, Rattan, Mutzel, Weisfeiler and Leman go sparse: Towards scalable higher-order graph embeddings, 2020
- [5] Corso, Cavalleri, Beaini, Lio, and Velickovic, Principal neighbourhood aggregation for graph nets, 2020
- [6] Vignac, Loukas, Frossard, Building powerful and equivariant graph neural networks with message-passing, 2020

Workshop Announcement



 $\frac{\text{https://www.ipam.ucla.edu/programs/workshops/deep-}}{\text{learning-and-combinatorial-optimization}}$

ORGANIZING COMMITTEE Peter Battaglia (DeepMind Technologies)

Xavier Bressen (Nanyang Technological University, Singapore)
Stefanie Jegelka (Massachusetts Institute of Technology)

Yann LeCun (New York University, Canadian Institute for Advanced Research)

Andres Lodi (École Polytechnique de Montréal)

Stanley Osher (University of California, Los Angeles (UCLA), Mathematics)

Oriol Vinyals (DeepMind Technologies)

Max Welling (University of Amsterdam)

Speaker List

Shipra Agrawal (Columbia University, Computer Science)

Sanjeev Arora (Princeton University)

Peter Battaglia (DeepMind Technologies)

Xavier Bresser (Nanyang Technological University, Singapore)

Joan Bruna (New York University)

Laurent Charlin (HEC Montreal)

Kyle Cranmer (New York University)

Sanjeeb Dash (IBM Watson Research Center)

Santanu Day (Georgia Institute of Technology, School of Industrial and Systems Engineering)

Bistra Dilkina (University of Southern California (USC))

Tina Eliassi-Rad (Northeastern University, Computer Science & Network Science)

Emma Frejinger (University of Montreal)

Maxime Gasse (École Polytechnique de Montréal)

Stefano Gualandi (Università di Pavia)

Oktay Gunluk (Cornell University)

Joey Huchette (Rice University)

Stefanie Jegelka (Massachusetts Institute of Technology)

Ron Kimmel (Technion - Israel Institute of Technology, Intel Perceptual Computing)

Zico Koller (Carnegie Mellon University)

Vladien Koltun (Intel Corporation)

Wouter Kool (University of Amsterdam)

Andrea Lodi (École Polytechnique de Montréal)

Azalia Mirhoseini (Google Inc.)

Starriey Osher (University of California, Los Angeles (UCLA), Mathematics)

Sebastian Pokutta (Korrad-Zuse-Zentrum für Informationstechnik (ZIB), Department of Mathematics)

Louis-Martin Rousseau (École Polytechnique de Montréal)

Thiago Serra (Bucknell University)

Le Song (Georgia institute of Technology)

Petar Velickovic (DeepMind Technologies)

Oriol Virgais (DeepMind Technologies)

Ellen Vitercik (Carnegie Mellon University)

Collaborators



Y. Bengio MILA



M. Bronstein Imperial/Twitter



F. Monti Twitter



C. Joshi NTU



V.P. Dwivedi NTU



Y.Y. Leow NTU



T. Laurent LMU



A. Szlam Facebook AI



R. Levie TAU



M. Defferrard EPFL



P. Vandergheynst EPFL



P. Hagmann UNIL



Xavier Bresson

xbresson@ntu.edu.sg

- http://www.ntu.edu.sg/home/xbresson
- https://github.com/xbresson
- https://twitter.com/xbresson
- If https://www.facebook.com/xavier.bresson.1
- in https://www.linkedin.com/in/xavier-bresson-738585b