Learning Long Range Dependencies on Graphs via Random Walks

Dexiong Chen, Till Hendrik Schulz, Karsten Borgwardt

Max Planck Institute of Biochemistry

ICLR 2025

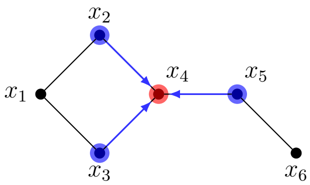
MAX PLANCK INSTITUTE
OF BIOCHEMISTRY

MLSB

Graph Neural Networks: Two Approaches

Message-passing GNNs

- Strong at capturing local relationships
- Struggle with long-range dependencies
- Issues: over-smoothing and over-squashing

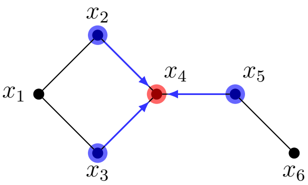


[Xu et al., 2019, Chen et al., 2022, Rampášek et al., 2022]

Graph Neural Networks: Two Approaches

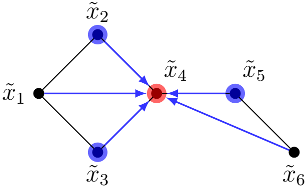
Message-passing GNNs

- Strong at capturing local relationships
- Struggle with long-range dependencies
- Issues: over-smoothing and over-squashing



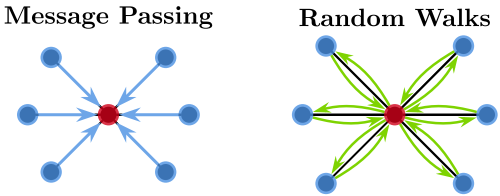
Graph transformers

- Enable global information exchange
- Oversimplify graph structure with fixed-length vectors
- Sacrifice structural richness for global reach



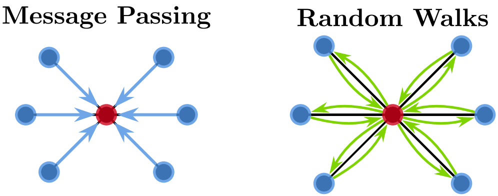
[Xu et al., 2019, Chen et al., 2022, Rampášek et al., 2022]

The Challenge



- **Message passing:** breadth-first, struggle with long-range dependencies.
- **Random walks:** naturally depth-first, effective for long-range relationships, inefficient for local relationships.

The Challenge



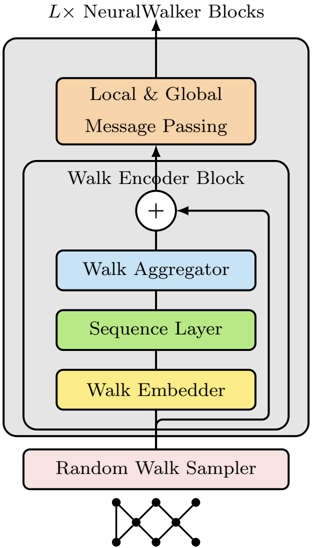
- **Message passing:** breadth-first, struggle with long-range dependencies.
- **Random walks:** naturally depth-first, effective for long-range relationships, inefficient for local relationships.

Our Insight: Combine the strengths of both approaches.

NeuralWalker: Our Approach

Key components

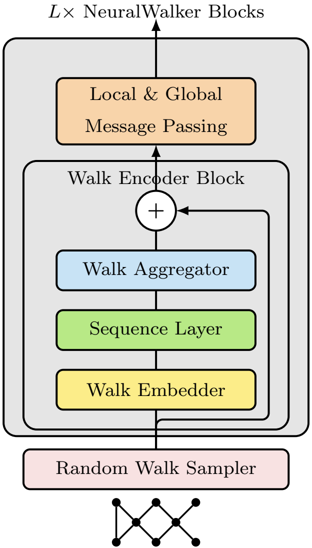
- Random Walk Sampler: Samples walks with positional encodings



NeuralWalker: Our Approach

Key components

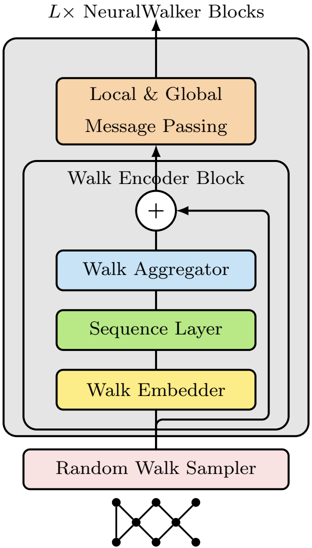
- Random Walk Sampler: Samples walks with positional encodings
- Walk Embedder: Computes walk embeddings from node/edge features



NeuralWalker: Our Approach

Key components

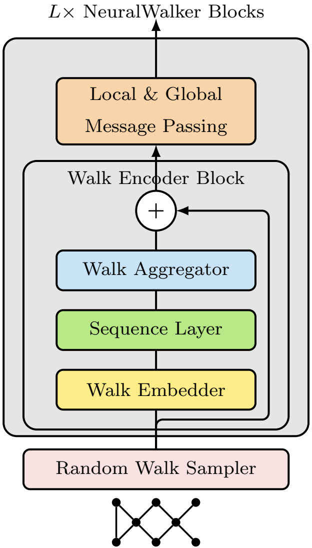
- Random Walk Sampler: Samples walks with positional encodings
- Walk Embedder: Computes walk embeddings from node/edge features
- Sequence Layer: Processes walks using advanced sequence models



NeuralWalker: Our Approach

Key components

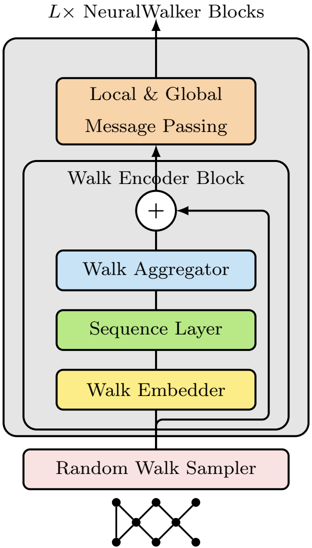
- Random Walk Sampler: Samples walks with positional encodings
- Walk Embedder: Computes walk embeddings from node/edge features
- Sequence Layer: Processes walks using advanced sequence models
- Walk Aggregator: Pools features across walks into nodes



NeuralWalker: Our Approach

Key components

- Random Walk Sampler: Samples walks with positional encodings
- Walk Embedder: Computes walk embeddings from node/edge features
- Sequence Layer: Processes walks using advanced sequence models
- Walk Aggregator: Pools features across walks into nodes
- Local & Global Message Passing: Complements walk information



Key Contributions

- Explicit walk sequences rather than compressed structural encodings

Key Contributions

- Explicit walk sequences rather than compressed structural encodings
- Leveraging sequence models for random walks:
  - CNNs
  - Transformers
  - State Space Models (Mamba, S4)

Key Contributions

- Explicit walk sequences rather than compressed structural encodings
- Leveraging sequence models for random walks:
  - CNNs
  - Transformers
  - State Space Models (Mamba, S4)
- Complementary message passing to address limitations of random walks alone

Key Contributions

- Explicit walk sequences rather than compressed structural encodings
- Leveraging sequence models for random walks:
  - CNNs
  - Transformers
  - State Space Models (Mamba, S4)
- Complementary message passing to address limitations of random walks alone
- Theoretical guarantees:
  - More expressive than 1-WL and k-subgraph isomorphism tests
  - Lipschitz continuity for stability
  - Linear complexity in graph size

Experimental Results

Dataset	SOTA	CRaWL	NeuralWalker
ZINC ↓	0.070	0.085	**0.053**
MNIST ↑	98.39	97.944	**98.692**
CIFAR10 ↑	76.853	69.013	**76.903**
PATTERN ↑	**87.196**	–	86.977
CLUSTER ↑	**80.026**	–	78.189
PascalVOC-SP ↑	0.4440	–	**0.4912**
COCO-SP ↑	0.3974	–	**0.4398**
Peptides-func ↑	**0.7133**	0.7074	0.7096
Peptides-struct ↓	**0.2455**	0.2506	0.2463
PCQM-Contact ↑	0.4703	–	**0.4707**
Pokec ↑	86.10	–	**86.46**

Check out our paper for more results!

Key results:

- SOTA on multiple datasets.
- Up to +13% improvement on PascalVOC-SP and COCO-SP.
- Significantly outperforms previous random walk-based model (CRaWL).
- Successfully scales to large graphs (> 1M nodes).

[Tönshoff et al., 2023]

Ablation Studies: Model Components

Sequence layer performance

- SSM $>$ CNN $>$ Transformer

Message passing impact

- Consistent improvement with local message passing
- Variable benefits from global message passing

Random walk parameters

- Performance vs. computational cost tradeoff
- Can be explicitly controlled through the walk sampling number and length

[Rosenbluth et al., 2024]

Masked Positional Encoding Pretraining

Novel pretraining strategy:

- Randomly mask 15% of positions in positional encodings
- Train model to recover original binary encoding vectors
- Combine with attribute masking pretraining

Masked Positional Encoding Pretraining

Novel pretraining strategy:

- Randomly mask 15% of positions in positional encodings
- Train model to recover original binary encoding vectors
- Combine with attribute masking pretraining

Result: Significant performance boost on ZINC dataset (0.063 $\rightarrow$ 0.053)

Conclusions

Contributions

- Novel framework combining random walks with message passing
- Leveraging powerful sequence models for walk encoding
- Theoretical guarantees of expressiveness
- SOTA performance on multiple benchmarks

References I

D. Chen, L. O'Bray, and K. Borgwardt. Structure-aware transformer for graph representation learning. In *International Conference on Machine Learning (ICML)*, 2022.

L. Rampášek, M. Galkin, V. P. Dwivedi, A. T. Luu, G. Wolf, and D. Beaini. Recipe for a general, powerful, scalable graph transformer. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

E. Rosenbluth, J. Tönshoff, M. Ritzert, B. Kisin, and M. Grohe. Distinguished in uniform: Self-attention vs. virtual nodes. In *International Conference on Learning Representations (ICLR)*, 2024.

J. Tönshoff, M. Ritzert, H. Wolf, and M. Grohe. Walking out of the weisfeiler leman hierarchy: Graph learning beyond message passing. *Transactions on Machine Learning Research (TMLR)*, 2023.

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