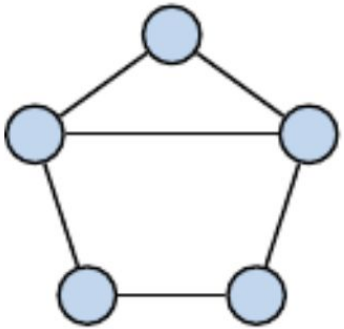
Graph Positional Encoding via Random Feature Propagation

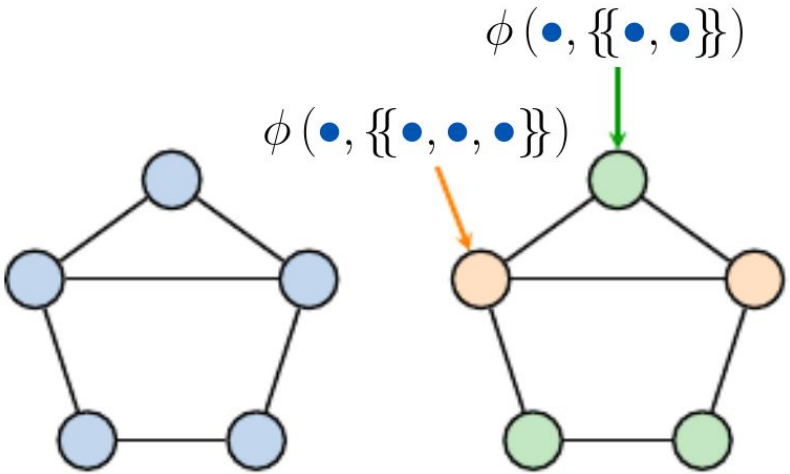
Peter Jolles

MPNN (Message Passing Neural Network)



Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges.* arXiv preprint arXiv:2104.13478.

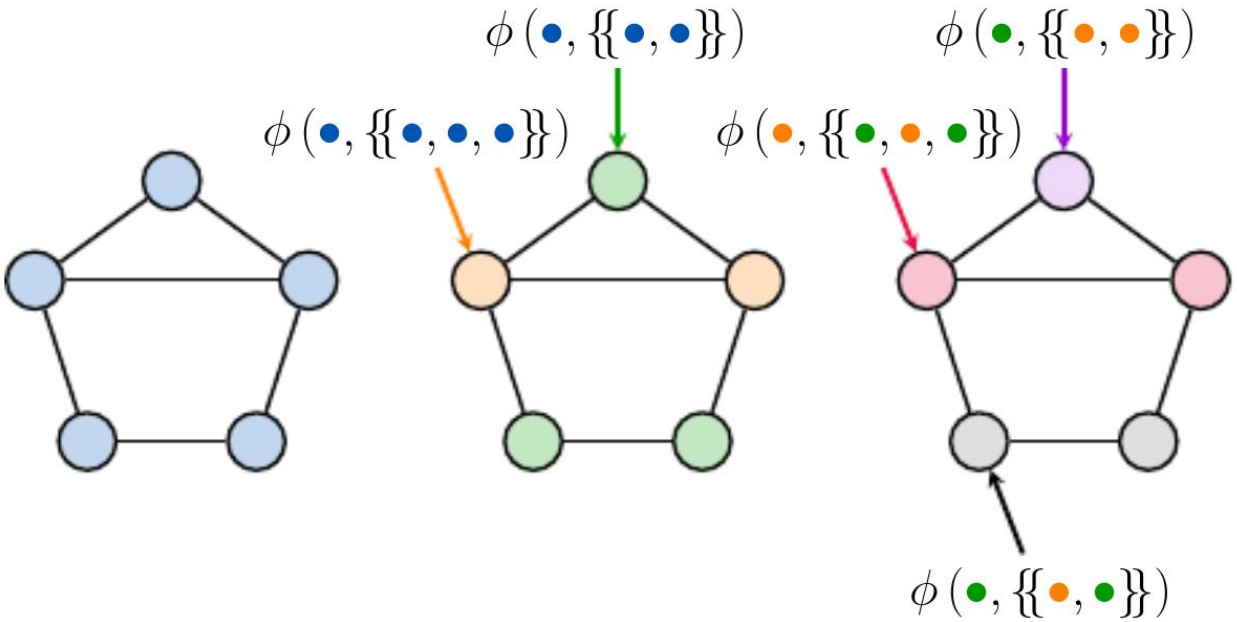
MPNN (Message Passing Neural Network)



Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges.* arXiv preprint arXiv:2104.13478.

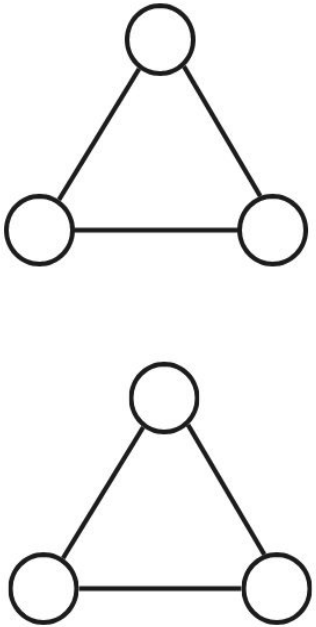
MPNN (Message Passing Neural Network)

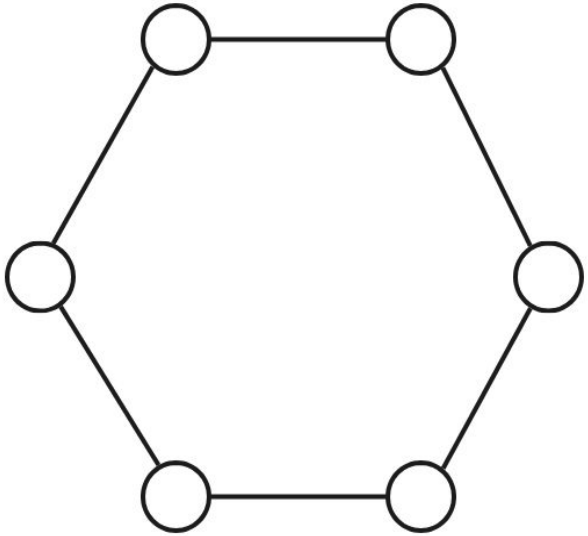
$$X^{l+1} = \sigma(AX^{(l)}W)$$



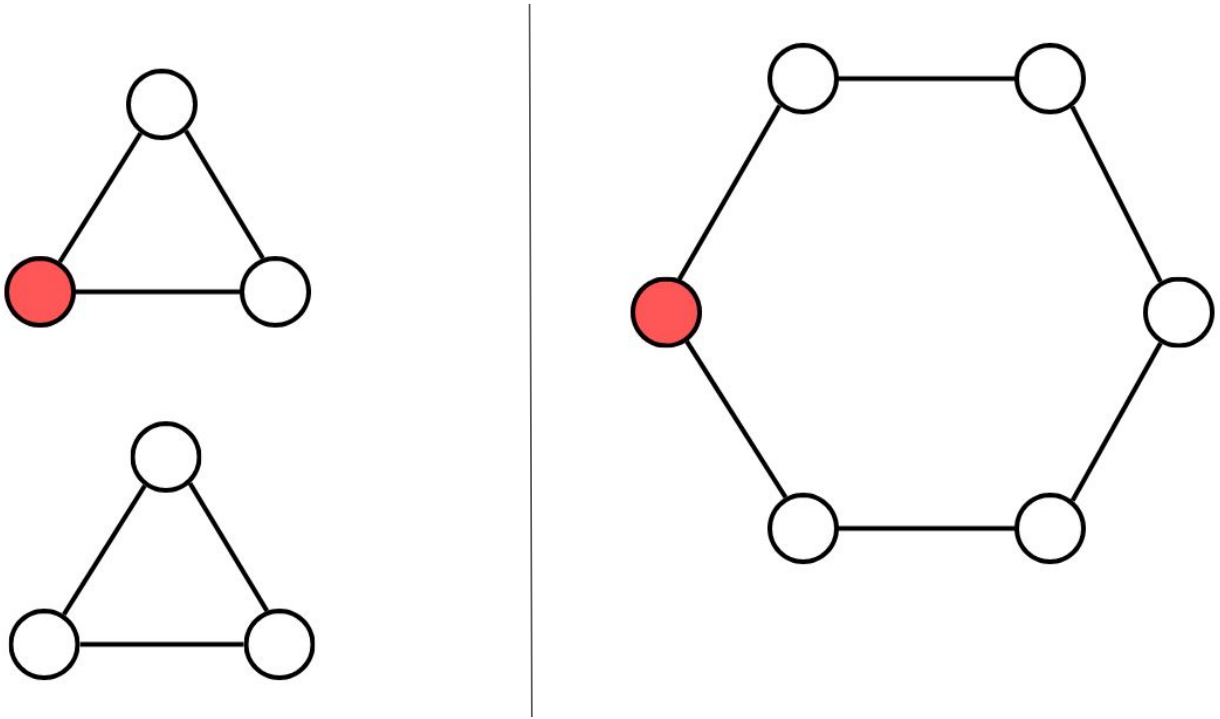
Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges.* arXiv preprint arXiv:2104.13478.

Limitations of MPNN



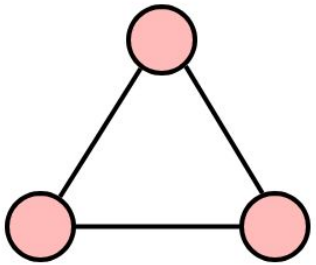


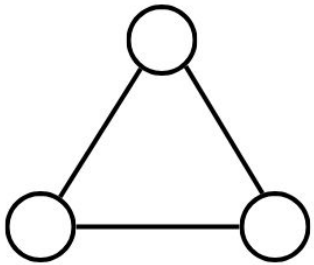
Solution: Random Node Features

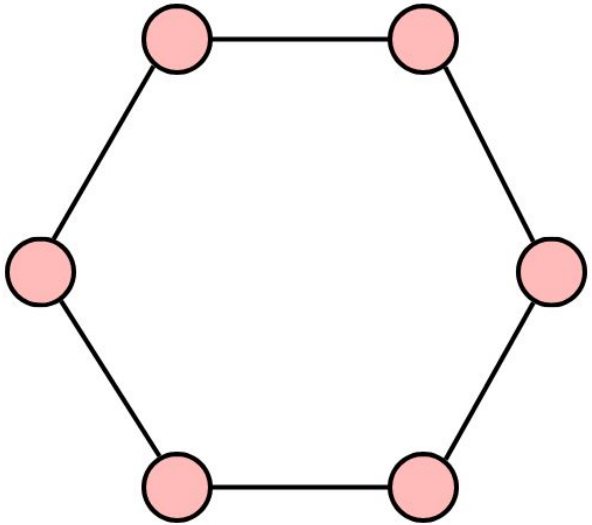


1. Abboud, R., Ceylan, İ. İ., Grohe, M., & Lukasiewicz, T. (2020). *The surprising power of graph neural networks with random node initialization.* arXiv:2010.01179.
2. Papp, P. A., & Wattenhofer, R. (2022). *A Theoretical Comparison of Graph Neural Network Extensions.* ICML.

Solution: Random Node Features





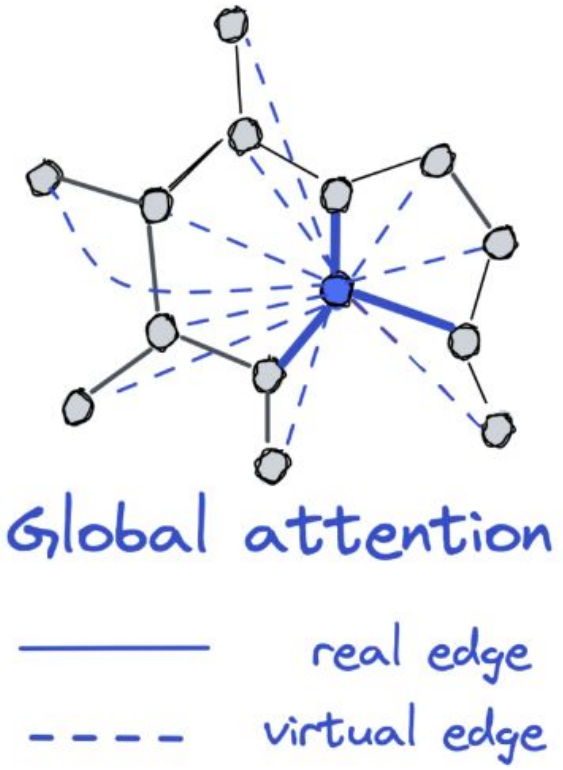


1. Abboud, R., Ceylan, İ. İ., Grohe, M., & Lukasiewicz, T. (2020). *The surprising power of graph neural networks with random node initialization.* arXiv:2010.01179.
2. Papp, P. A., & Wattenhofer, R. (2022). *A Theoretical Comparison of Graph Neural Network Extensions.* ICML.

Graph Transformer

- Requires input features to properly apply attention mechanism

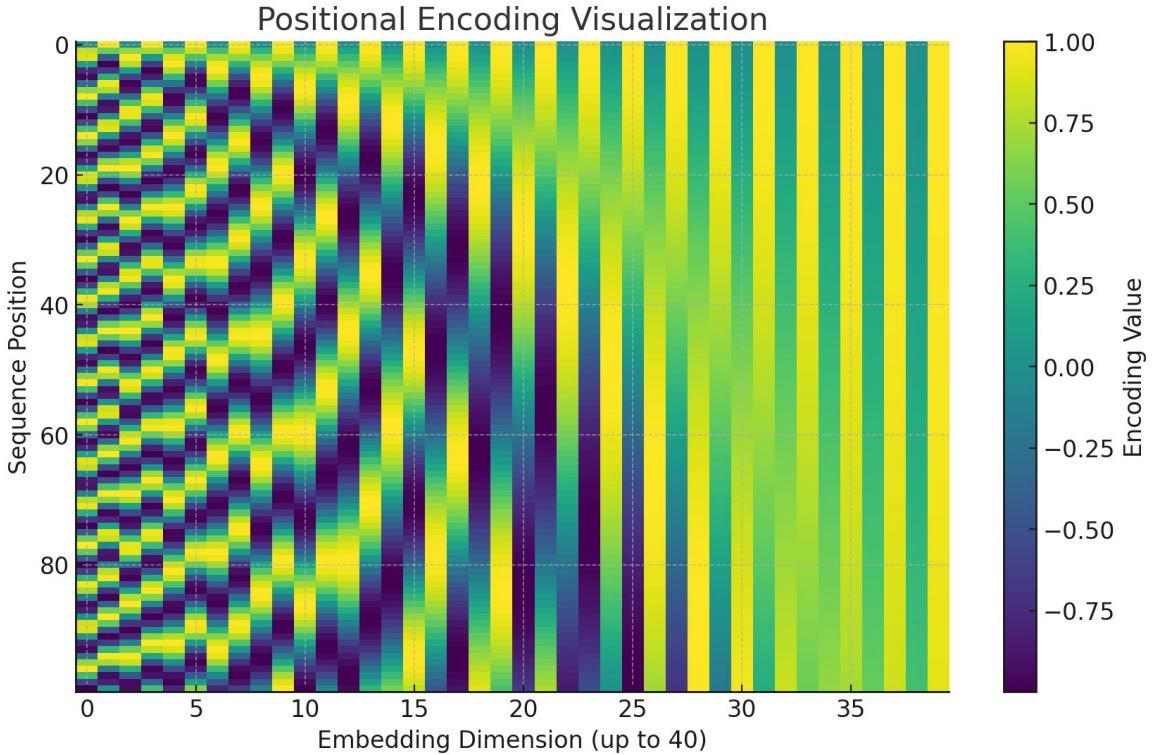
- Solution: Positional Encodings for nodes



https://towardsdatascience.com/graphgps-navigating-graph-transformers-c2cc223a051c/

Positional Encodings - LLMs

Inject sequence order into the model



https://geoffroypeeters.github.io/deeplearning-101-audiomir_book/bricks_architecture.html

Positional Encodings - Graphs

Unlike sequence models, graphs don’t have a natural order

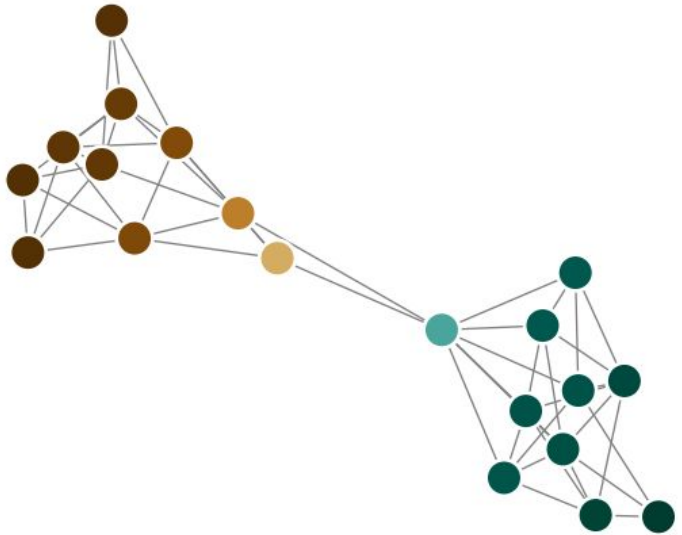
- Encoding based on absolute position isn’t possible

Commonly used: Eigenvectors of laplacian matrix

$$L = D - A$$

D is the diagonal degree matrix

A is the adjacency matrix



https://network-science-notes.github.io/chapters/03-components-laplacian.html

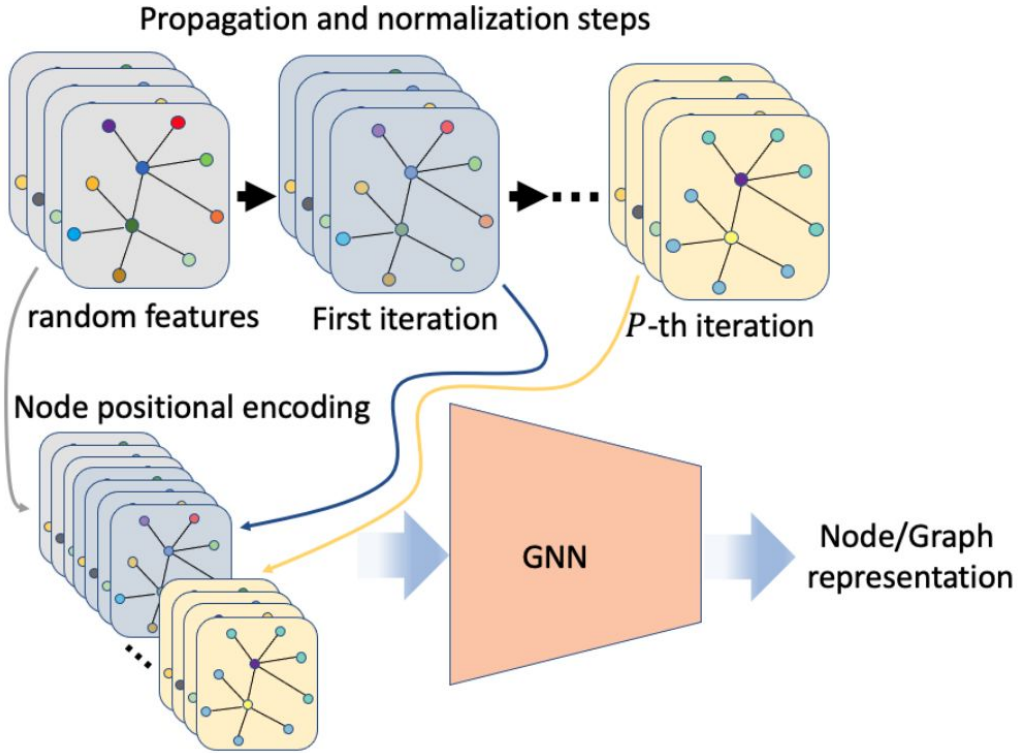
Downside: Computing Eigenvectors is expensive

Putting everything together

- Random node features

- Positional encoding

Putting everything together



Random Feature Propagation

Start with random features

$$\mathbf{a}^{(0)} = \mathbf{r}$$

For P iterations:

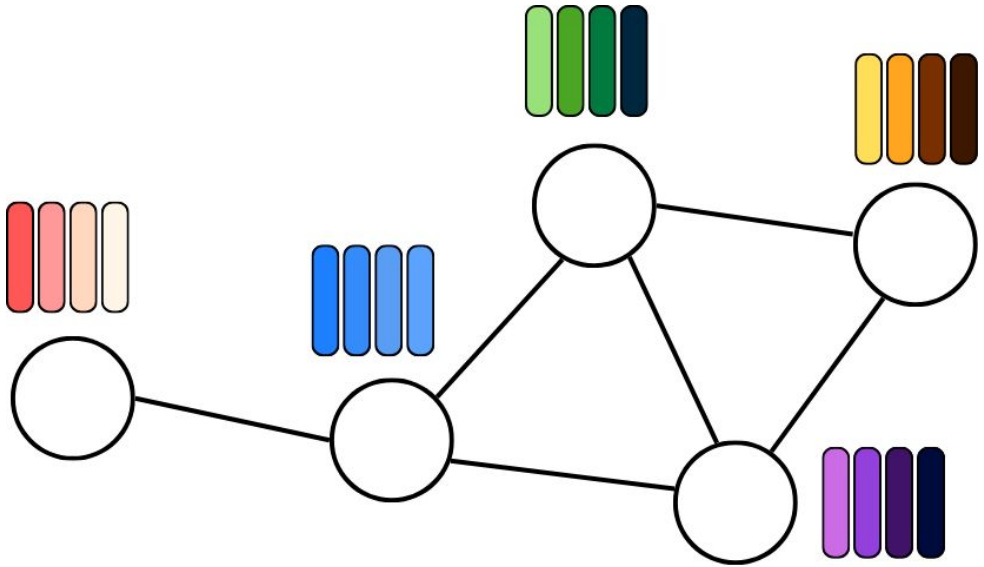
- Propagate features

$$\hat{\mathbf{a}}^{(p)} = \mathbf{S}\mathbf{a}^{(p-1)}$$

- Normalize (every few rounds)

$$\mathbf{a}^{(p)} = \mathbf{N}(\hat{\mathbf{a}}^{(p)})$$

Return: Full Trajectory with concatenated Features

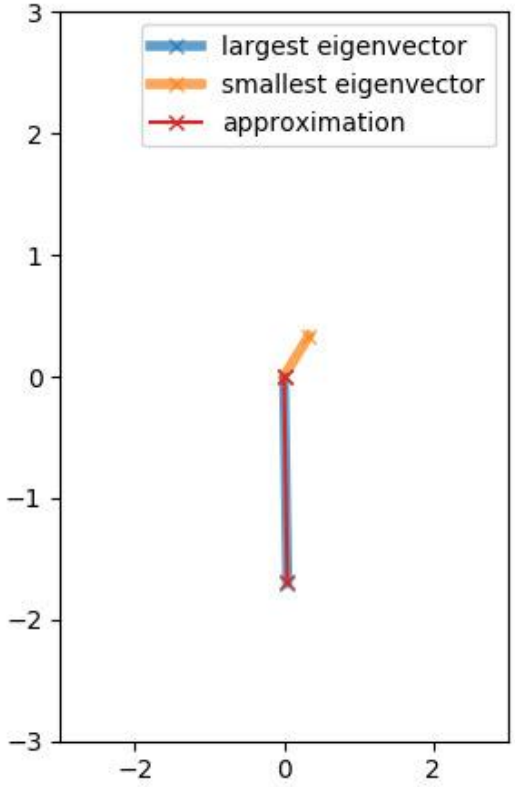


Power Iteration

$$b_{k+1} = \frac{Ab_k}{\|Ab_k\|}$$

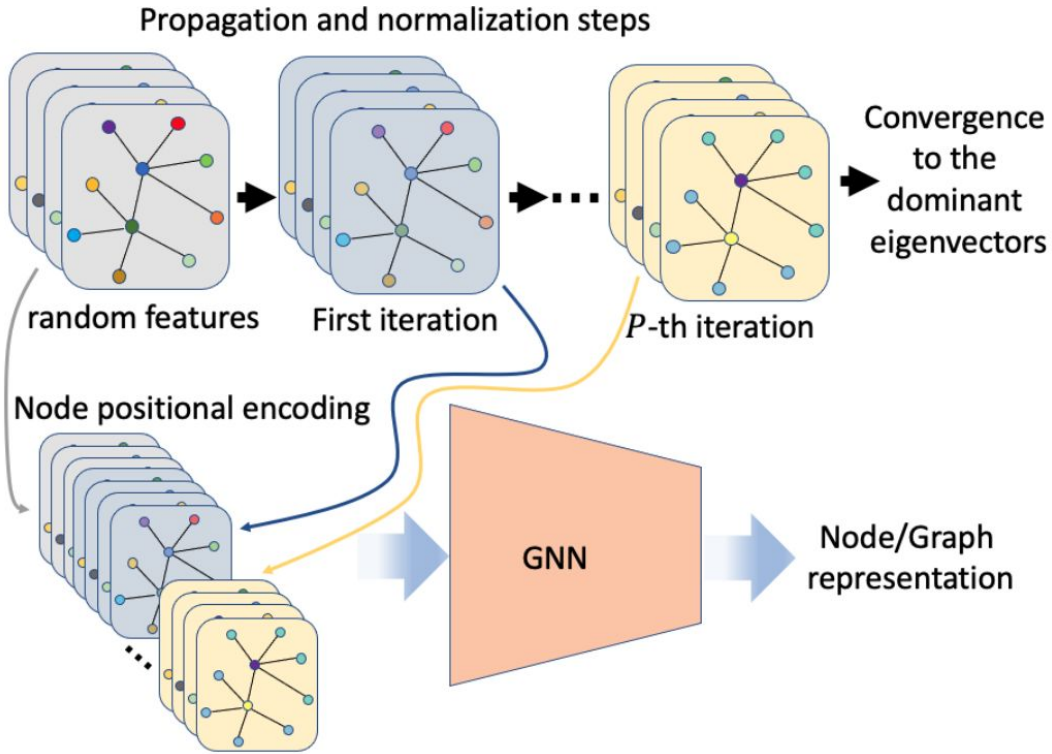
Sequence (b_k) converges to dominant eigenvector of A

Consequently, random features converge to largest eigenvector of propagation matrix



https://en.wikipedia.org/wiki/Power_iteration

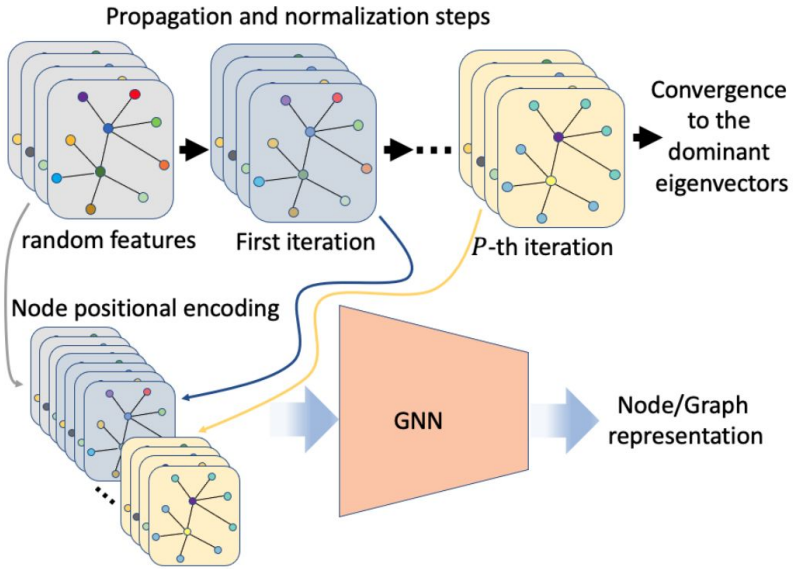
Random Feature Propagation



Normalization Function Options

- L2 Normalization

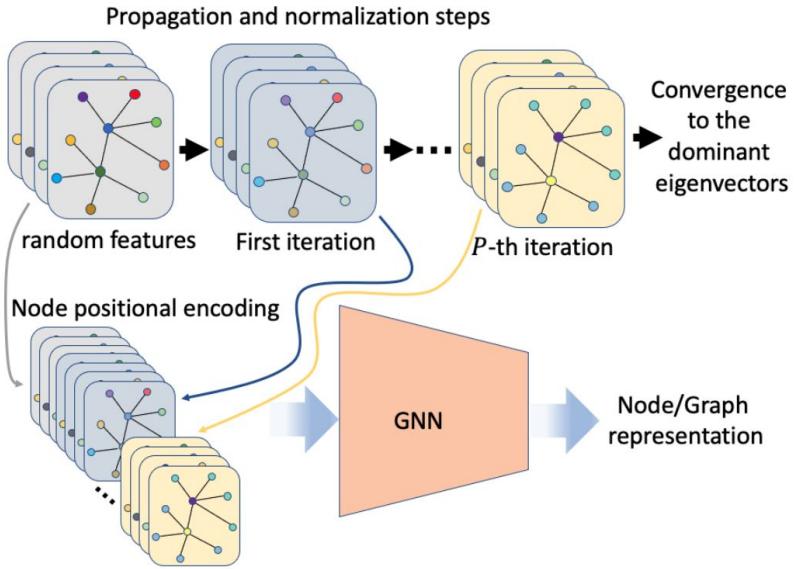
- Orthonormalization
  - Different random feature vectors will converge to different eigenvectors



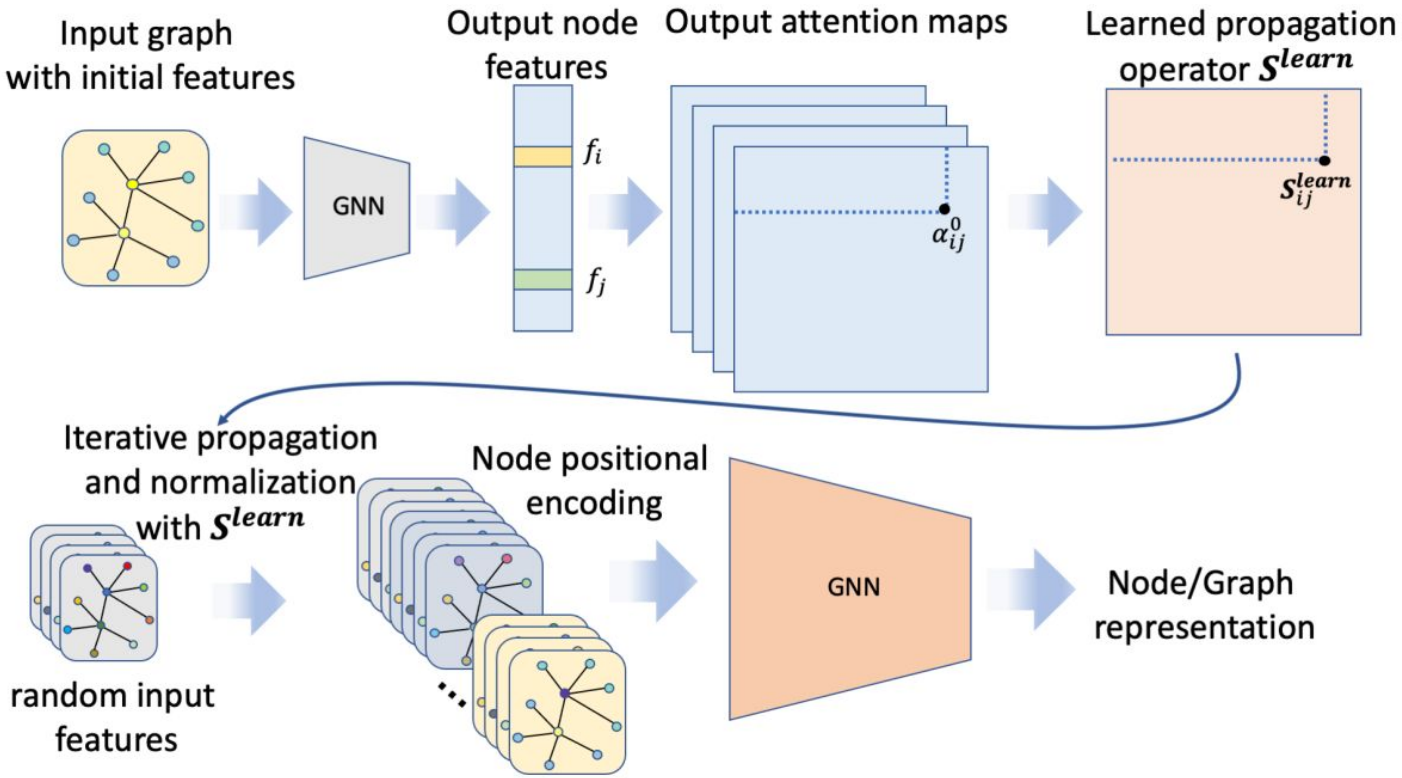
Propagation Operator Options

- Predefined Operators
  - Adjacency matrix A (with self loops and symmetrically normalized)
  - Laplacian matrix L = D - A (symmetrically normalized)

- Learnable Operator
  - Use a GNN to create propagation matrix



Extension: Learnable Propagation Operator



Results - Graph-level Tasks

Experiments with QR-Orthonormalization yield strong results

Table 1. ZINC-12k and OGBG-MOLHIV datasets. Backbone is GINE for all baselines and RFP variants. † denotes eigenvectors corresponding to eigenvalues in ascending order of magnitude. The top three models are highlighted by **First**, **Second**, **Third**.

Method	ZINC-12K MAE ↓	OGBG-MOLHIV ROC-AUC(%) ↑
Standard GNNs		
GCN	0.321 ± 0.009	76.06 ± 0.97
GIN	0.163 ± 0.004	75.58 ± 1.40
PNA	0.133 ± 0.011	79.05 ± 1.32
DGN	–	79.70 ± 0.97
Domain-aware		
HIMP	–	78.80 ± 0.82
GSN	**0.101 ± 0.010**	**80.39 ± 0.90**
CIN	**0.079 ± 0.006**	**80.94 ± 0.57**
Natural baselines		
$\hat{\mathbf{L}}$ EIGVECS †	0.1557 ± 0.012	77.88 ± 1.82
$\hat{\mathbf{A}}$ EIGVECS	0.1501 ± 0.018	78.13 ± 1.61
$\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$ EIGVECS	0.1434 ± 0.015	78.30 ± 1.47
RNF	0.1621 ± 0.014	75.98 ± 1.63
RNF & $\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$ EIGVECS	0.1408 ± 0.020	78.96 ± 1.33
RWPE	0.1279 ± 0.005	78.62 ± 1.13
RFP (ours)		
RFP - ℓ_2 - $\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$	0.1368 ± 0.010	77.91 ± 1.43
RFP - QR - $\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$	0.1152 ± 0.006	79.83 ± 1.16
RFP - QR - $\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$, $\mathbf{S}^{\text{learn}}$	0.1143 ± 0.008	79.94 ± 1.51
RFP - QR - $\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$ -DSS	0.1117 ± 0.009	80.53 ± 1.04
RFP - QR - $\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$, $\mathbf{S}^{\text{learn}}$ -DSS	**0.1106 ± 0.012**	**80.58 ± 1.21**

Results - Node-level Tasks

Experiments with QR-Orthonormalization yield strong results

Table 2. Node classification accuracy (%) ↑. * denotes the best result out of several variants. An MLP backbone is used for Cornell, Texas, and Wisconsin, and a GCN backbone for the rest of the datasets. The top three models are highlighted by **First**, **Second**, **Third**.

Method Homophily	Squirrel 0.22	Film 0.22	Cham. 0.23	Corn. 0.30	Texas 0.11	Wisc. 0.21	Citeseer 0.74	Pubmed 0.80	Cora 0.81
Standard baselines									
MLP	28.77	36.41	46.24	80.89	80.81	83.88	74.16	87.05	75.59
GCN	23.96	26.86	28.18	52.70	52.16	48.92	73.68	88.13	85.77
Node classification designated architectures									
WRGAT	48.85	36.53	65.24	81.62	83.62	86.98	76.81	88.52	**88.20**
GGCN	**55.17**	**37.81**	**71.14**	**85.68**	**84.86**	86.86	**77.14**	**89.15**	87.95
H2GCN	36.48	35.70	60.11	82.70	**84.86**	**87.65**	**77.11**	**89.49**	87.87
GPRGNN	31.61	34.63	46.58	80.27	78.38	82.94	**77.13**	87.54	**87.95**
G^2-GCN	46.48	37.09	55.83	**86.49**	84.86	87.06	–	–	–
G^2-GraphSAGE	**64.26**	**37.14**	**71.40**	86.22	**87.57**	**87.84**	–	–	–
Natural baselines									
$\hat{\mathbf{L}}$ 64 EIGVECS	35.85	34.06	50.87	74.10	74.03	78.17	74.42	88.17	86.31
$\hat{\mathbf{A}}$ 64 EIGVECS	35.15	34.05	51.44	75.40	72.97	77.25	75.01	88.41	86.37
$\hat{\mathbf{A}}$ ALL EIGVECS	34.78	34.32	50.99	72.70	73.24	80.18	73.89	88.03	85.91
RNF	34.40	35.34	51.20	78.90	80.03	81.93	73.70	88.61	85.43
RNF & 64 $\hat{\mathbf{A}}$ EIGVECS	44.89	34.68	58.21	79.46	78.37	82.25	75.29	88.96	85.67
Propagation-based PE									
SGC	37.07	–	55.11	75.68	75.68	75.29	73.39	–	84.89
SIGN	40.97	–	60.11	76.76	75.14	78.43	73.27	–	83.92
PowerEmbed*	53.53	–	64.98	78.30	79.19	78.43	73.27	–	85.03
NFP - QR - $\hat{\mathbf{A}}$	46.85	27.72	61.25	65.40	78.23	62.97	73.89	88.86	85.93
RFP (ours)									
RFP - ℓ_2 - $\hat{\mathbf{A}}$	48.89	35.11	62.13	82.13	81.18	81.66	75.59	88.62	87.10
RFP - QR - $\hat{\mathbf{A}}$	54.28	36.87	65.98	85.13	84.74	86.08	76.48	**90.01**	87.99
RFP - QR - $\hat{\mathbf{A}}$ -DSS	54.78	36.93	66.16	85.69	85.10	86.35	76.92	90.00	**88.09**
RFP - QR - $\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$, $\mathbf{S}^{\text{learn}}$ - DSS	**54.97**	**37.12**	**66.81**	**86.74**	**87.13**	**87.05**	76.53	89.96	87.99

Results - Counting

Table 3. Graph Counting dataset MAE ↓. Backbone is GIN for all baselines and RFP variants. † denotes eigenvectors corresponding to eigenvalues in ascending order of magnitude. The top three models are highlighted by **First**, **Second**, **Third**.

Method	Triangle	Tailed-tri	Star	Cyc4
Standard models				
GCN	0.4186	0.3248	0.1798	0.2822
GIN	0.3569	0.2373	0.0224	0.2185
PNA	0.3532	0.2648	0.1278	0.2430
PPGN	**0.0089**	**0.0096**	0.0148	**0.0090**
Subgraph GNNs				
GNN-AK	0.0934	0.0751	0.0168	0.0726
GNN-AK-CTX	0.0885	0.0696	0.0162	0.0668
GNN-AK+	**0.0123**	**0.0112**	0.0150	**0.0126**
DSS-GNN (EGO+)	0.0450	0.0393	0.0129	0.0550
SUN (EGO+)	**0.0079**	**0.0080**	**0.0064**	**0.0105**
Natural baselines				
$\hat{\mathbf{L}}$ EIGVECS †	0.1681	0.1728	0.0168	0.1572
$\hat{\mathbf{A}}$ EIGVECS	0.1904	0.1714	0.0140	0.1533
$\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$ EIGVECS	0.1694	0.1680	0.0121	0.1506
RNF	0.1775	0.1807	0.0189	0.1516
RNF & $\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$ EIGVECS	0.1512	0.1503	**0.0120**	0.1497
RFP (ours)				
RFP - ℓ_2 - $\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$ - GINE	0.1521	0.1573	0.0177	0.1439
RFP - QR - $\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$ - GINE	0.0923	0.0826	0.0118	0.0991
RFP - QR - $\hat{\mathbf{L}}$, $\hat{\mathbf{A}}$ - GINE-DSS	0.0820	0.0764	**0.0072**	0.0715

Results - Ablation

More random features is better

Table 4. Ablation study on the effect of the chosen number of eigenvectors k in RFP with fixed number of propagations $P = 16$ on ZINC-12k and OGBG-MOLHIV datasets.

Method	k	ZINC-12K MAE ↓	OGBG-MOLHIV ROC-AUC (%) ↑
GINE Backbone			
RFP - ℓ_2 - $\hat{\mathbf{L}}, \tilde{\mathbf{A}}$	64	0.1368 ± 0.010	77.91 ± 1.43
RFP - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}$	4	0.1305 ± 0.007	78.26 ± 1.51
RFP - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}$	8	0.1223 ± 0.009	79.43 ± 1.37
RFP - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}$	16	0.1152 ± 0.006	79.83 ± 1.16
RFP - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}, \mathbf{S}^{\text{learn}}$	16	0.1143 ± 0.008	80.29 ± 1.01
RFP - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}$ - DSS	16	0.1117 ± 0.009	80.53 ± 1.04
RFP - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}, \mathbf{S}^{\text{learn}}$ - DSS	16	0.1106 ± 0.012	80.58 ± 1.21
GraphGPS Backbone			
RFP - ℓ_2 - $\hat{\mathbf{L}}, \hat{\mathbf{A}}$	64	0.1322 ± 0.014	77.98 ± 1.57
RFP - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}$	16	0.1100 ± 0.005	79.95 ± 1.41
Combining RFP with domain-aware PE			
RFP+GSN - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}$	16	0.0734 ± 0.007	80.07 ± 1.19

Table 5. Ablation study on the effect of the chosen number of eigenvectors k in RFP with fixed number of propagations $P = 16$ on the Graph Counting dataset test MAE ↓. GIN is used as backbone.

Method	k	Triangle	Tailed-tri	Star	Cyc4
RFP_{ℓ_2}	16	0.1521	0.1573	0.0177	0.1439
RFP - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}$	4	0.1322	0.01289	0.0126	0.1401
RFP - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}$	8	0.1231	0.1143	0.0119	0.1170
RFP - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}$	16	0.0923	0.0826	0.0118	0.0991
RFP - QR - $\hat{\mathbf{L}}, \hat{\mathbf{A}}$ - DSS	16	0.0820	0.0764	0.0072	0.0715

Results - Ablation

More propagations is better

Table 6. Node classification accuracy (%) ↑ on homophilic and heterophilic datasets.

Method Homophily	OP –	k (vecs) –	# props –	Squirrel 0.22	Film 0.22	Cham. 0.23	Corn. 0.30	Texas 0.11	Wisc. 0.21	Cora 0.81	Cite 0.74	Pub 0.80
EIGVEC-PE	$\hat{\mathbf{L}}$	64	–	35.85	34.06	50.87	74.10	74.03	78.17	86.31	74.42	88.17
EIGVEC-PE	$\hat{\mathbf{A}}$ (TopAbs)	64	–	35.15	34.05	51.44	75.40	72.97	77.25	86.37	75.01	88.41
EIGVEC-PE ALL	–	$\lvert\mathcal{V}\rvert$	–	34.78	34.32	50.99	72.70	73.24	80.18	85.91	73.89	88.03
RNF	–	64	0	34.40	35.34	51.20	78.90	80.03	81.93	85.43	73.70	88.61
RNF	–	$\lvert\mathcal{V}\rvert$	0	34.36	34.74	53.28	79.30	80.00	81.97	84.56	72.29	88.51
RNF & $\hat{\mathbf{A}}$ EIGVECS	$\hat{\mathbf{A}}$	64	–	44.89	34.68	58.21	79.46	78.37	82.25	85.67	75.29	88.96
NFP - QR	$\hat{\mathbf{A}}$	64	16	46.85	27.72	61.25	65.40	78.23	62.97	73.89	88.86	85.93
RFP - ℓ_2	$\hat{\mathbf{L}}$	1	16	48.89	35.11	62.13	82.13	81.18	81.66	87.10	75.59	88.62
RFP - QR	$\hat{\mathbf{A}}$	$\lvert\mathcal{V}\rvert$	16	54.23	36.61	64.77	85.07	85.41	86.27	85.68	75.16	89.00
RFP - QR	$\hat{\mathbf{A}}$	64	2	53.87	36.40	64.50	83.78	82.70	83.13	86.26	75.22	88.68
RFP - QR	$\hat{\mathbf{A}}$	64	4	54.11	36.44	65.16	84.59	83.78	84.90	86.67	75.37	88.76
RFP - QR	$\hat{\mathbf{A}}$	64	8	54.34	36.56	65.83	84.86	84.05	85.88	87.35	75.94	89.95
RFP - QR	$\hat{\mathbf{A}}$	64	16	54.28	36.87	65.98	85.13	84.32	86.08	87.99	76.40	90.01
RFP - QR	$\hat{\mathbf{A}}$	64	32	54.25	36.92	65.80	85.04	84.59	85.88	87.81	76.43	90.03
RFP - QR	$\hat{\mathbf{A}}$	4	16	51.96	35.88	64.01	83.24	82.70	82.16	86.93	76.11	89.66
RFP - QR	$\hat{\mathbf{A}}$	8	16	52.72	35.92	64.63	83.78	83.51	82.43	87.32	76.27	89.93
RFP - QR	$\hat{\mathbf{A}}$	16	16	53.87	36.51	65.11	84.59	83.78	82.97	87.65	76.39	90.00
RFP - QR	$\hat{\mathbf{A}}$	32	16	54.33	36.42	65.66	84.86	84.74	84.05	87.84	76.48	89.97
RFP - QR	$\hat{\mathbf{A}}$	64	16	54.28	36.87	65.98	85.13	84.32	86.08	87.99	76.40	90.01
RFP - QR	$\hat{\mathbf{A}}$	128	16	54.40	36.70	65.90	85.07	84.59	86.12	87.96	76.42	89.83
RFP - QR	$\hat{\mathbf{L}}, \hat{\mathbf{A}}, \mathbf{S}^{\text{learn}}$	64/32	16/8	54.56	36.95	66.34	86.40	86.72	86.88	87.92	76.41	89.63
RFP - QR - DSS	$\hat{\mathbf{A}}$	64/32	16/8	54.78	36.93	66.16	85.69	85.10	86.35	88.09	76.92	90.00
RFP - QR - DSS	$\hat{\mathbf{L}}, \hat{\mathbf{A}}, \mathbf{S}^{\text{learn}}$	64/32	16/8	54.97	37.12	66.81	86.74	87.13	87.05	87.99	76.53	89.96

Table 7. The obtained node classification accuracy(%) on ogbn-arxiv. The exact computation of the Laplacian $\hat{\mathbf{L}}$ eigenvectors required several hours and therefore the result is not available.

GCN	RNF + GCN	GCN + $\hat{\mathbf{L}}$ EIGVECS †	GCN + RFP (Ours)
71.08	70.91	N/A	71.97

Opinion

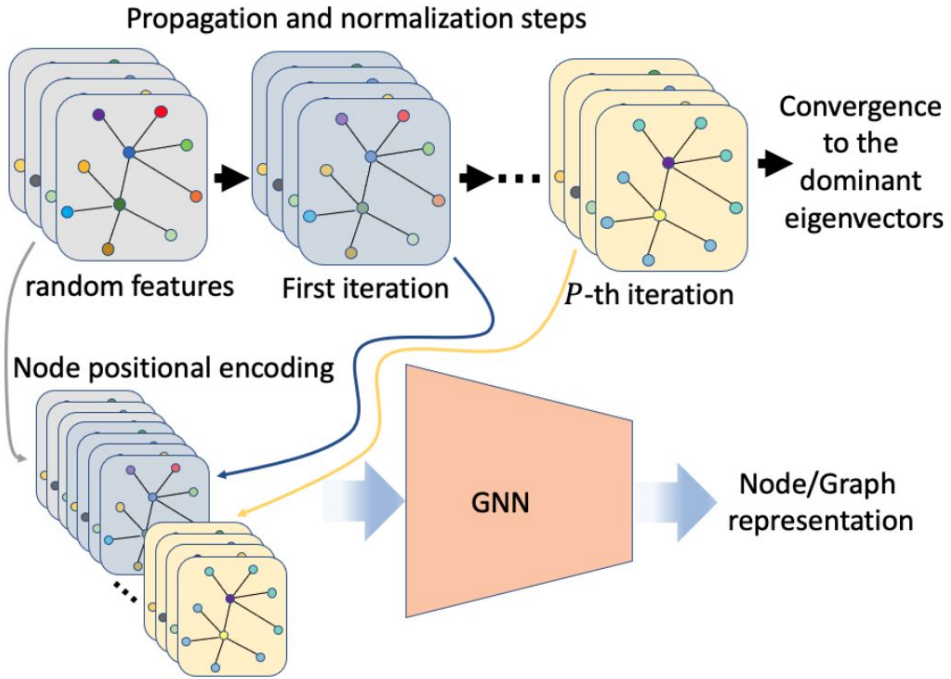
Pros:

+ Thorough theoretical analysis
+ Can be used to improve different GNN types
+ Experimental tests

Cons:

- Compute heavy (expensive normalization)
- Random features are not permutation invariant

Questions


Propagation and normalization steps
random features
First iteration
P-th iteration
Convergence to the dominant eigenvectors
Node positional encoding
GNN
Node/Graph representation