# Personalized PageRank Graph Attention Networks

Julie Choi

Amazon

## Summary


- We introduce a new graph attention network (GAT) based model that incorporates the global graph structure into the attention weights by using Personalized PageRank (PPR).
- Furthermore, this removes the need of multiple GAT layers, making GAT scalable and reducing over-smoothing.
- Our models outperform the GAT baseline models with lower computational cost. Our models also outperform the existing PPR baseline models.


## Background & Motivation

There has been a rising interest in graph neural networks (GNNs) for representation learning over the past few years. GNNs provide a general and efficient framework to learn from graph-structured data. Among the variants of GNNs, GATs [1] learn the importance weight of each neighbor to aggregate information, and it improved the performance of many graph learning tasks.

In GAT, the attention weights (importance weights) depend only on node features. However, the graph structure can be valuable information. For example, the nodes with large degrees could be more important. It's highly desirable to incorporate the graph structure into the attention weights for aggregation. In this work, we incorporate the limit distribution of Personalized PageRank into the attention weights to reflect the global graph structure. Our models outperform the baseline models on four widely used benchmark datasets. Our implementation is publicly available online at https://github.com/juliechoi12/pprgat.

## Personalized PageRank

### Background

In Personalized PageRank (also referred to as topic-specific PageRank), the random walk restarts at the root node with a teleport probability $\alpha$. The resulting stationary distribution is interpreted as the relative importance of nodes with respect to the root node, while PageRank measures the global importance of nodes in a graph. This relative importance is purely from the graph structure, making it valuable information in addition to the node features when computing the attention weights.

### Approximate Personalized PageRank

- We approximate the PageRank vector with ACL's algorithm [2].
- The algorithm is local (it needs only neighbors) and highly parallelizable.
- Furthermore, we select only the top k nodes, resulting in a sparse vector with PageRank scores of only the relevant nodes. This makes the batch training highly efficient by using fixed number of neighbors.

## Personalized PageRank Graph Attention Networks (PPRGAT)

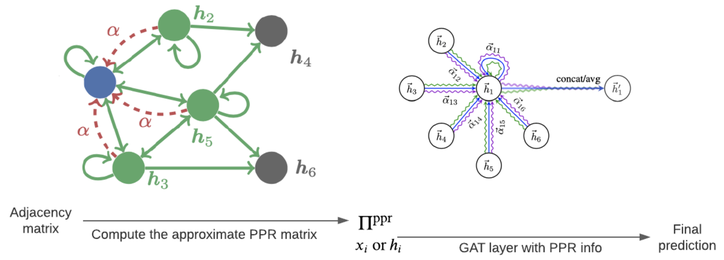


Figure 1. Illustration of PPRGAT. First, we precompute $\Pi^{\epsilon,k}$ from the adjacency matrix. Second, $\Pi^{\epsilon,k}_{ij}, x_i, x_j$ are used together to generate the attention weight from node $j$ to node $i$ in the GAT layer.

### Attention layer variants

PPRGAT

Modify the attention weight equation of GAT [1]

$$e(h_i, h_j) = \text{LeakyReLU}(a^T \cdot [Wh_i || Wh_j]) \quad (1)$$

to

$$e(h_i, h_j) = \text{LeakyReLU}(a^T \cdot [Wh_i \| Wh_j \| \Pi^{\epsilon,k}_{ij}]). \quad (2)$$

PPRGATv2

Modify the attention weight equation of GATv2 [3]

$$e(h_i, h_j) = a^T \text{LeakyReLU}(W \cdot [h_i || h_j]) \quad (3)$$

to

$$e(h_i, h_j) = a^T \text{LeakyReLU}(W \cdot [h_i \| h_j \| \Pi^{\epsilon,k}_{ij}]) \quad (4)$$

### Two neighbor aggregation methods

- **Default**: Aggregate over only the top-k nodes of $\pi(i)$
- **Local**: Aggregate over all neighbors defined by the original adjacency matrix (PPRGAT-local and PPRGATv2-local)

## Experiments

We evaluate the four variants described previously: PPRGAT, PPRGAT-local, PPRGATv2, and PPRGATv2-local. We compare our models against two GAT baseline models and two PPR baseline models. The GAT baseline models include GAT [1] and GATv2 [3], and the PPR baseline models include (A)PPNP [4] and PPRGo [5]. Our models outperform the baseline models on all datasets tested.

### Transductive learning

Table 1. Classification accuracies (in %) of different node classification algorithms on the citation datasets. Results are the averages of 10 runs.

| Model | Cora | Citeseer | Pubmed |
|---|---|---|---|
| GAT | 83.0 | 70.8 | 79.0 |
| GATv2 | 82.9 | 71.6 | 78.7 |
| APPNP | 83.3 | 71.8 | 80.1 |
| PPRGo | 74.2 | 65.6 | 70.7 |
| PPRGAT | 83.9 | **72.5** | 80.4 |
| PPRGAT-local | **84.0** | 72.2 | 80.1 |
| PPRGATv2 | 83.8 | 72.4 | **80.5** |
| PPRGATv2-local | 83.9 | 72.1 | 80.2 |

### Inductive learning

Table 2. Classification micro-F1 scores (in %) of different node classification algorithms on the PPI dataset. Results are the averages of 10 runs.

| Model | PPI |
|---|---|
| GAT | 96.5 |
| GATv2 | 96.3 |
| APPNP | 96.7 |
| PPRGo | 87.8 |
| PPRGAT | **97.5** |
| PPRGAT-local | 97.1 |
| PPRGATv2 | 97.4 |
| PPRGATv2-local | 97.1 |

## References


[1] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio, "Graph attention networks," in *International Conference on Learning Representations*, 2018.

[2] Reid Andersen, Fan Chung, and Kevin Lang, "Local graph partitioning using pagerank vectors," in *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*. IEEE, 2006.

[3] Shaked Brody, Uri Alon, and Eran Yahav, "How attentive are graph attention networks?," in *International Conference on Learning Representations*, 2022.

[4] Johannes Klicpera, Aleksandar Bojchevski, and Stephan Günnemann, "Predict then propagate: Graph neural networks meet personalized pagerank," in *International Conference on Learning Representations*, 2019.

[5] Aleksandar Bojchevski, Johannes Klicpera, Bryan Perozzi, Amol Kapoor, Martin Blais, Benedek Rózemberczki, Michal Lukasik, and Stephan Günnemann, "Scaling graph neural networks with approximate pagerank," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2020, ACM.




julichoi@amazon.com