On Positional and Structural Node Features for Graph Neural Networks on Non-attributed Graphs

Presented by **Hejie CUI**

hejie.cui@emory.edu

https://hejiecui.com/

Hejie CUI[1,*], Zijie LU[2,*], Pan LI[3], Carl YANG[1,†]

[1] Emory University

[2] University of Illinois at Urbana-Champaign

[3] Purdue University

[*] These two authors made equal contributions to this work

[†] Corresponding Author

EMORY UNIVERSITY

Outline

- **Introduction**
- Two types of Artificial Node Features
- Experiments
- Conclusion

Privilege of GNNs on common graph tasks

- Various powerful GNNs demonstrate privilege on graph data.
- GNNs combine node features and graph structures by aggregating node features through links into low-dimensional vector representations.
- Superior performance is mainly established <u>when natural node features are available.</u>

Challenge from natural features missing

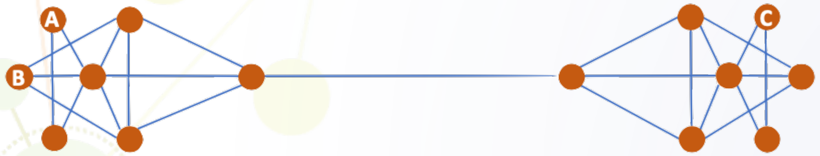
- A great number of graphs in the wild **do not** contain natural node features, due to privacy concerns and/or difficulties in data collection.
- Several intuitive methods have been commonly practiced to initialize node features (e. g. random, degree-based, etc.).

> **How to choose artificial node features for GNNs on non-attributed graphs?**

Outline

- Introduction
- **Two types of Artificial Node Features**
- Experiments
- Conclusion

Positional Node Features



Positional node features help GNNs capture node distance information regarding their relative positions in the graph.

In the Figure, nodes A and B are positional close.

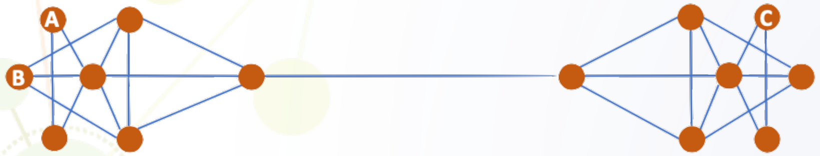
E.g. publication network, where two authors who cite each other and also cite / get cited by similar other authors should be recognized as sharing similar research interests considering their graph positions.

- ***random***: a feature vector following random distribution. The random feature of each node varies among runs with difference random seeds
- ***one-hot***: a unique one-hot feature vector is initialized for each node
- ***eigen***: eigen decomposition is performed on the normalized adjacency matrix and the top k eigen vectors are used to generate a k-dimensional feature vector for each node (k is decided by grid search)
- ***deepwalk***: the initial feature of a node is generated based on DeepWalk algorithm [1] with walk length set as 40. (deep walk features with walk length longer than 2 can capture higher-order positional information)

Specifically, eigen and deepwalk generate features by matrix decomposition, essentially dimension reduction.

1 [Perozzi et al., 2014] Deepwalk: Online learning of social representations. In SIGKDD.

Structural Node Features



Structural node features help GNNs capture local structural information of nodes, such as degree information and neighborhood connection patterns.

In the Figure, nodes A and C are positional close.

E.g. molecular network, where two nodes with similar degrees and connection patterns should be recognized as atoms with similar properties or functions.

- ***shared***: an initial feature vector is shared across all nodes (in the experiments we simply use a vector of all 1's)
- ***degree***: the degree value is converted to a one-hot degree vector for each node, where the vector dimension is selected based on the max degree of all nodes
- ***pagerank***: the original PageRank [2] score of a given node is calculated and then flattened into a vector, where the dimension of the extended vector is selected by grid-search. Pagerank can be viewed as a generalized higher-order node degree information

Byproduct: new SOTA for structural node classification

- ***degree+***: divide the degree values into several buckets, then map the degree values distributed in each bucket range into one class, and finally construct a unique one-hot vector for each class.

2 [Brin et al., 1997] The Anatomy of a Large-Scale Hypertex- tual Web Search Engine. Comput. Networks 30 (1998), 107–117.

Outline

- Introduction
- Two types of Artificial Node Features
- **Experiments**
- Conclusion

Positional Node Classification - 1

Aggr.	Type	Feature	Cora Acc.(%)	Pubmed Acc.(%)	Citeseer Acc.(%)
Mean	$\mathcal{P}$	random	56.1±1.6	42.3±1.4	36.0±1.0
		one-hot	58.2±4.0	51.4±3.1	37.3±2.5
		eigen	73.2±2.3	70.0±4.8	42.9±2.3
		deepwalk	**75.3±1.0**	**74.0±2.6**	**46.8±0.9**
	$\mathcal{S}$	shared	17.9±0.0	38.6±0.0	20.2±0.0
		degree	37.4±2.1	41.1±2.9	36.0±1.3
		pagerank	25.2±2.4	39.8±1.9	20.5±3.4
		real feat.	80.2±1.1	79.0±2.2	68.0±4.0
Sum	$\mathcal{P}$	random	45.2±3.9	41.7±2.7	32.8±2.7
		one-hot	47.0±3.7	46.4±4.4	33.0±1.8
		eigen	70.5±5.1	68.8±4.1	40.1±5.0
		deepwalk	70.0±2.3	72.5±2.2	43.7±2.7
	$\mathcal{S}$	shared	17.1±5.2	33.3±6.4	22.3±4.6
		degree	50.7±3.7	42.6±1.8	32.0±3.5
		pagerank	27.8±4.4	33.0±6.3	23.4±1.3
		real feat.	70.5±3.7	75.4±3.7	59.3±4.0

Table 1: Positional node classification results

- Definition and Datasets
  - The tasks of positional node classification target at predicting the “positional role” of each node
  - Datasets: Core, Citeseer and Pubmed, consisting of scientific publications as nodes, which can be classified into several content categories
- Protocols
  - Train and test the GraphSAGE model using the same split as [3]
  - Performance with real node features as baseline
  - Comprehensive grid search for the best hyper-parameter settings including the learning rate, number of epochs and neighborhood sample size
  - The final performance of each feature initialization method is averaged over five runs under the optimal hyper-parameter settings

3 [N. Kipf et al., 2017] Semi-Supervised Classification with Graph Convolutional Networks. In ICLR.

Aggr.	Type	Feature	Cora Acc.(%)	Pubmed Acc.(%)	Citeseer Acc.(%)
Mean	$\mathcal{P}$	random	56.1±1.6	42.3±1.4	36.0±1.0
		one-hot	58.2±4.0	51.4±3.1	37.3±2.5
		eigen	73.2±2.3	70.0±4.8	42.9±2.3
		deepwalk	**75.3±1.0**	**74.0±2.6**	**46.8±0.9**
	$\mathcal{S}$	shared	17.9±0.0	38.6±0.0	20.2±0.0
		degree	37.4±2.1	41.1±2.9	36.0±1.3
		pagerank	25.2±2.4	39.8±1.9	20.5±3.4
		real feat.	80.2±1.1	79.0±2.2	68.0±4.0
Sum	$\mathcal{P}$	random	45.2±3.9	41.7±2.7	32.8±2.7
		one-hot	47.0±3.7	46.4±4.4	33.0±1.8
		eigen	70.5±5.1	68.8±4.1	40.1±5.0
		deepwalk	70.0±2.3	72.5±2.2	43.7±2.7
	$\mathcal{S}$	shared	17.1±5.2	33.3±6.4	22.3±4.6
		degree	50.7±3.7	42.6±1.8	32.0±3.5
		pagerank	27.8±4.4	33.0±6.3	23.4±1.3
		real feat.	70.5±3.7	75.4±3.7	59.3±4.0

Table 1: Positional node classification results

Positional Node Classification - 2

- Observations
  - **Aggregation**:
    - mean aggregation > sum aggregation
  - **Cross Feature Type Comparison**:
    - Most positional node features achieve much better performance than structural node features
  - **Within Feature Type Comparison**:
    - Random and one-hot achieve comparable results
    - Among all positional features, deep-walk and eigen demonstrate the best performance across all the datasets

Aggr.	Type	Initial.	USA-air Acc.(%)	Brazil-air Acc.(%)	Europe-air Acc.(%)
Mean	$\mathcal{P}$	random	59.3±1.8	45.7±5.9	44.9±5.8
		one-hot	59.2±2.6	48.6±7.4	44.0±0.7
		eigen	55.3±1.5	40.0±6.9	31.6±2.1
		deepwalk	58.1±2.8	42.1±9.6	41.5±3.3
	$\mathcal{S}$	shared	25.0±0.0	25.0±0.0	25.0±0.0
		degree	53.8±1.9	48.6±4.1	42.7±2.7
		degree+	59.2±2.7	60.0±3.0	50.6±3.9
		pagerank	39.7±2.9	47.9±7.4	25.9±0.0
Sum	$\mathcal{P}$	random	60.7±3.2	47.9±7.4	48.9±5.1
		one-hot	59.2±3.3	50.7±8.5	48.9±5.4
		eigen	67.8±2.5	57.8±5.3	49.4±4.5
		deepwalk	68.8±3.0	65.0±6.4	54.1±2.8
	$\mathcal{S}$	shared	55.7±2.0	61.4±4.7	45.4±1.0
		degree	63.6±3.0	70.0±4.1	58.0±3.6
		degree+	**69.1±2.6**	**76.4±4.1**	**61.2±3.8**
		pagerank	58.8±2.0	73.6±5.4	45.9±1.0
SOTA		struc2vec	63.8±1.6	73.6±9.6	58.8±3.0

Table 2: Structural node classification results

Structural Node Classification - 1

- Definition and Datasets
  - The tasks of structural node classification target at predicting the "structural role" of each node
  - Datasets: American air-traffic network, Brazilian air-traffic network and European air-traffic network
  - Given an airport node in the air-traffic network, the target is to predict passenger flow level of that node solely based on the structural of air-traffic network
- Protocols
  - Following struc2vec [4]
  - To highlight the performance of our novel *degree+* method, adopt regression with L2 regularization to train the classifier using the representation learned by struc2vec, which demonstrates SOTA results on these datasets

4 [Ribeiro et al., 2017] struc2vec: Learning node representations from structural identity. In SIGKDD.

Aggr.	Type	Initial.	USA-air Acc.(%)	Brazil-air Acc.(%)	Europe-air Acc.(%)
Mean	$\mathcal{P}$	random	59.3±1.8	45.7±5.9	44.9±5.8
		one-hot	59.2±2.6	48.6±7.4	44.0±0.7
		eigen	55.3±1.5	40.0±6.9	31.6±2.1
		deepwalk	58.1±2.8	42.1±9.6	41.5±3.3
	$\mathcal{S}$	shared	25.0±0.0	25.0±0.0	25.0±0.0
		degree	53.8±1.9	48.6±4.1	42.7±2.7
		degree+	59.2±2.7	60.0±3.0	50.6±3.9
		pagerank	39.7±2.9	47.9±7.4	25.9±0.0
Sum	$\mathcal{P}$	random	60.7±3.2	47.9±7.4	48.9±5.1
		one-hot	59.2±3.3	50.7±8.5	48.9±5.4
		eigen	67.8±2.5	57.8±5.3	49.4±4.5
		deepwalk	68.8±3.0	65.0±6.4	54.1±2.8
	$\mathcal{S}$	shared	55.7±2.0	61.4±4.7	45.4±1.0
		degree	63.6±3.0	70.0±4.1	58.0±3.6
		degree+	**69.1±2.6**	**76.4±4.1**	**61.2±3.8**
		pagerank	58.8±2.0	73.6±5.4	45.9±1.0
SOTA		struc2vec	63.8±1.6	73.6±9.6	58.8±3.0

Table 2: Structural node classification results

Structural Node Classification - 2

- Observations
  - **Aggregation**:
    - sum aggregation > mean aggregation
  - **Cross Feature Type Comparison**:
    - In most cases structural node features demonstrate superiority compared with positional ones
    - Our proposed structural node feature degree+ manifests the most distinct advantage over other positional features, new SOTA
  - **Within Feature Type Comparison**:
    - Degree+ improves on degree by using a degree bucket, which alleviates the node degree sparsity and skewness problem
    - Shared performs rather poorly
    - Pagerank can be viewed as a generalized higher-order node degree. Its performance deterioration may arise from over-smoothing

Aggr.	Typ.	Initial.	MUTAG Acc.(%)	PROTEINS Acc.(%)	IMDB-B Acc.(%)	IMDB-M Acc.(%)
Mean	$\mathcal{P}$	random	64.9±4.1	67.2±4.2	58.0±2.9	36.1±1.9
		one-hot	65.8±7.0	67.8±2.6	56.9±3.4	36.8±3.2
		eigen	63.8±2.1	60.4±1.0	50.2±1.3	33.4±0.7
		deepwalk	65.1±8.3	68.1±4.0	52.1±3.4	35.7±1.9
	$\mathcal{S}$	shared	66.7±0.0	59.6±0.0	50.0±0.0	33.3±0.0
		degree	**84.4±7.7**	69.5±2.6	69.7±5.1	45.1± 2.6
		pagerank	66.5±1.9	68.0±5.5	54.4±4.0	35.5±1.7
		real feat.	71.4±4.4	74.0±4.2	-	-
Sum	$\mathcal{P}$	random	66.9±7.1	67.5±4.1	54.0±3.6	36.2±2.1
		one-hot	65.1±3.8	66.8±3.8	52.8±2.7	33.4±2.6
		eigen	65.4±7.7	69.0±4.1	69.3±4.6	42.4±3.4
		deepwalk	64.2±8.6	66.2±4.2	51.9±2.8	35.3±3.0
	$\mathcal{S}$	shared	79.9±6.7	69.1±4.5	67.9±2.8	43.3±4.6
		degree	84.0±8.4	69.3±3.3	68.9±2.5	44.9±4.1
		pagerank	77.3±7.6	**69.9±3.1**	**70.3±2.9**	**48.2±3.2**
		real feat.	83.0±6.3	73.8±2.6	-	-

Table 3: Graph classification results

Graph Classification - 1

- Definition and Datasets
  - Two datasets w. real node features from chemical domain: MUTAG and PROTEINS
  - Two datasets w/o real node features from social domain: IMDB-BINARY and IMDB-MULTI
- Protocols
  - Graph level experiments are conducted with artificial features of sizes ranging from 100 to 500 with step 100

Graph Classification - 2

Aggr.	Typ.	Initial.	MUTAG Acc.(%)	PROTEINS Acc.(%)	IMDB-B Acc.(%)	IMDB-M Acc.(%)
Mean	$\mathcal{P}$	random	64.9±4.1	67.2±4.2	58.0±2.9	36.1±1.9
		one-hot	65.8±7.0	67.8±2.6	56.9±3.4	36.8±3.2
		eigen	63.8±2.1	60.4±1.0	50.2±1.3	33.4±0.7
		deepwalk	65.1±8.3	68.1±4.0	52.1±3.4	35.7±1.9
	$\mathcal{S}$	shared	66.7±0.0	59.6±0.0	50.0±0.0	33.3±0.0
		degree	**84.4±7.7**	69.5±2.6	69.7±5.1	45.1± 2.6
		pagerank	66.5±1.9	68.0±5.5	54.4±4.0	35.5±1.7
		real feat.	71.4±4.4	74.0±4.2	-	-
Sum	$\mathcal{P}$	random	66.9±7.1	67.5±4.1	54.0±3.6	36.2±2.1
		one-hot	65.1±3.8	66.8±3.8	52.8±2.7	33.4±2.6
		eigen	65.4±7.7	69.0±4.1	69.3±4.6	42.4±3.4
		deepwalk	64.2±8.6	66.2±4.2	51.9±2.8	35.3±3.0
	$\mathcal{S}$	shared	79.9±6.7	69.1±4.5	67.9±2.8	43.3±4.6
		degree	84.0±8.4	69.3±3.3	68.9±2.5	44.9±4.1
		pagerank	77.3±7.6	**69.9±3.1**	**70.3±2.9**	**48.2±3.2**
		real feat.	83.0±6.3	73.8±2.6	-	-

Table 3: Graph classification results

- Observations
  - **Aggregation**:
    - sum aggregation > mean aggregation
  - **Cross Feature Type Comparison**:
    - Though the best performance is not consistently achieved on a particular feature, it always falls in the category of structural node features
  - **Within Feature Type Comparison**:
    - Among the structural node features, pagerank demonstrates better performance in most of the cases
    - Degree feature on MUTAG and pagerank feature on PROTEIN with sum aggregator surpass those with real features

Outline

- Introduction
- Two types of Artificial Node Features
- Experiments
- **Conclusion**

Conclusion

- A generic practical guideline on the choices between different artificial node features for GNNs on non-attributed graphs.
  - ***Types*** : categorize commonly used artificial node features into two groups, positional node features and structural node features, based on what kind of information they can help GNNs capture.
  - ***Tasks*** : conduct experiments across three graph mining tasks, positional node classification, structural node classification and graph classification.
- Insights: positional node features are more suitable for positional node classification, while structural node features benefit more for structural node classification and graph classification tasks.

Thanks