# Workshop on Relational Representation Learning NeurIPS 2018

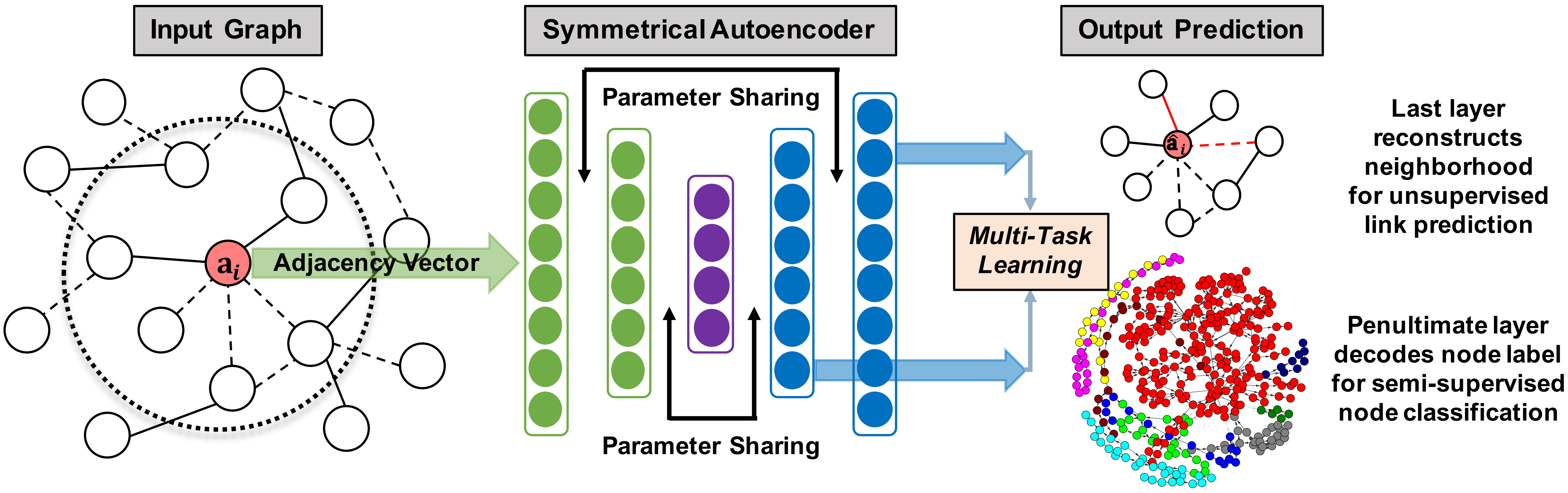
# Multi-Task Graph Autoencoders

*Phi Vu Tran | tran_phi@bah.com*

## Autoencoder Architecture for Link Prediction and Node Classification

We present the multi-task graph autoencoder (MTGAE) architecture, a ***simple, yet versatile and effective*** neural network with parameter sharing that can combine optional side node features for improved graph representation learning:

1. Learns from ***complex graph structures***: bipartite, sparse, weighted, directed;
2. Performs efficient ***end-to-end simultaneous, multi-task learning*** of link prediction and node classification;
3. Yields ***superior predictive performances*** over three strong baselines across five benchmark relational datasets.




### Unsupervised Link Prediction

- *Inference model*

Encoder $\mathbf{z}_i = \mathrm{ReLU}\left(\mathbf{W} \cdot \mathrm{ReLU}\left(\mathbf{V}\mathbf{a}_i + \mathbf{b}^{(1)}\right) + \mathbf{b}^{(2)}\right).$

Decoder $\hat{\mathbf{a}}_i = \mathbf{V}^{\mathrm{T}} \cdot \mathrm{ReLU}\left(\mathbf{W}^{\mathrm{T}}\mathbf{z}_i + \mathbf{b}^{(3)}\right) + \mathbf{b}^{(4)}.$

- *Learning*

$$\mathcal{L}_{\mathrm{BCE}} = -\mathbf{a}_i \log\left(\sigma\left(\hat{\mathbf{a}}_i\right)\right) \cdot \zeta - \left(1 - \mathbf{a}_i\right) \log\left(1 - \sigma\left(\hat{\mathbf{a}}_i\right)\right),$$

$$\mathcal{L}_{\mathrm{MBCE}} = \frac{\sum_i \mathbf{m}_i \odot \mathcal{L}_{\mathrm{BCE}}}{\sum_i \mathbf{m}_i}.$$

**Encoder Network**

**Code Layer**

**Decoder Network**

### Semi-Supervised Node Classification

- *Inference model*

$$\tilde{\mathbf{z}}_i = \mathbf{U} \cdot \mathrm{ReLU}\left(\mathbf{W}^{\mathrm{T}}\mathbf{z}_i + \mathbf{b}^{(3)}\right) + \mathbf{b}^{(5)},$$

$$\hat{\mathbf{y}}_i = \mathrm{softmax}(\tilde{\mathbf{z}}_i).$$

- *Learning*

$$\mathcal{L}_{\mathrm{SSC}} = -\mathrm{MASK}_i \sum_{c \in C} \mathbf{y}_{ic} \log(\hat{\mathbf{y}}_{ic}).$$

### Multi-Task Learning

$$\mathcal{L}_{\text{MULTI-TASK}} = \mathcal{L}_{\mathrm{MBCE}} + \mathcal{L}_{\mathrm{SSC}}$$

## Empirical Evaluation

Table 1: Summary statistics of datasets used in empirical evaluation.

| Dataset | Nodes | Average Degree | $\lvert O^+ \rvert : \lvert O^- \rvert$ Ratio | Node Features | Node Classes | Label Rate |
|---|---|---|---|---|---|---|
| Pubmed | 19,717 | 4.5 | 1 : 4384 | 500 | 3 | 0.003 |
| Citeseer | 3,327 | 2.8 | 1 : 1198 | 3,703 | 6 | 0.036 |
| Cora | 2,708 | 3.9 | 1 : 694 | 1,433 | 7 | 0.052 |
| Arxiv-GRQC | 5,242 | 5.5 | 1 : 947 | – | – | – |
| BlogCatalog | 10,312 | 64.8 | 1 : 158 | – | – | – |

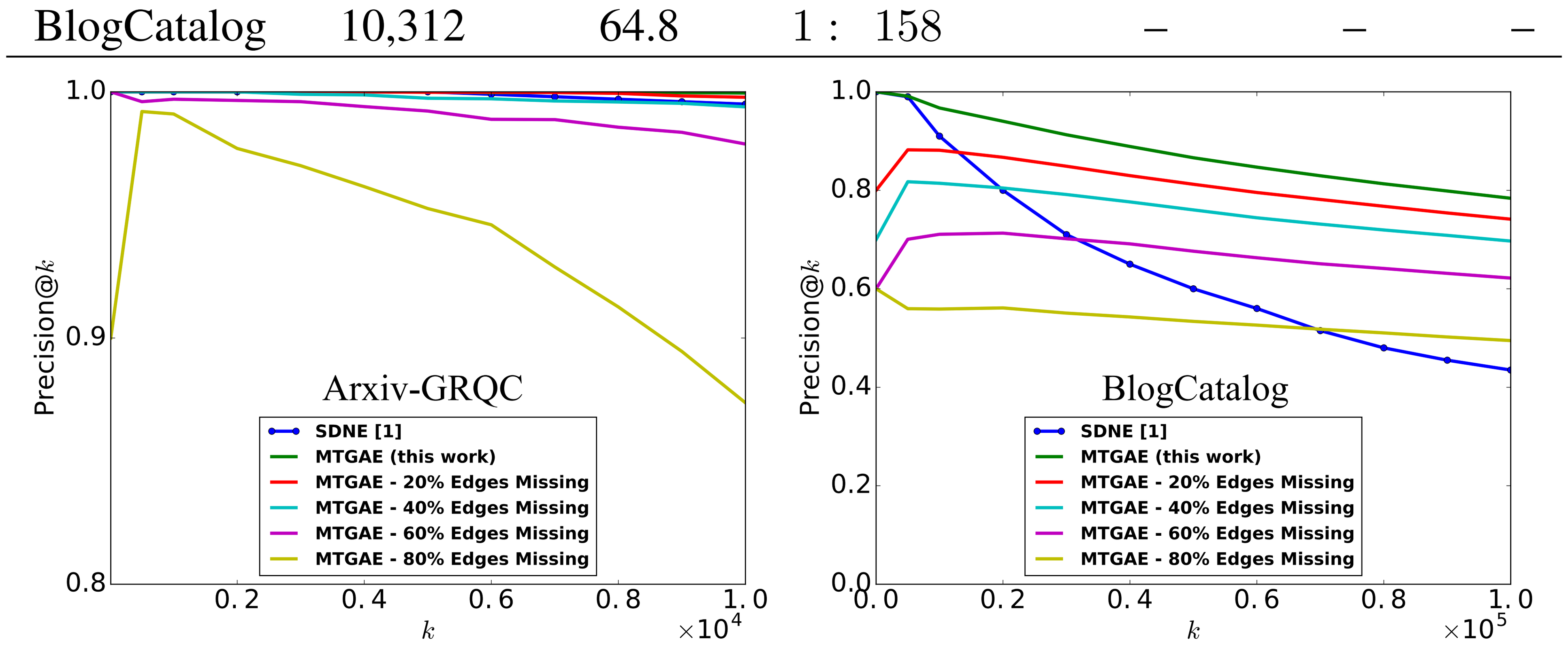



Figure 1: Comparison of network reconstruction performances.

Table 2: Related deep graph embedding baselines.

| Baseline | Evaluation Task | Metric |
|---|---|---|
| SDNE [1] | Reconstruction | Precision@$k$ |
| VGAE [2] | Link Prediction | AUC, AP |
| GCN [3] | Node Classification | Accuracy |

Table 3: Comparison of link prediction and node classification performances.

| Method | Cora | Citeseer | Pubmed |
|---|---|---|---|
| **Link Prediction** | | | |
| MTGAE | **0.946** | **0.949** | 0.944 |
| VGAE [2] | 0.920 | 0.914 | **0.965** |
| **Node Classification** | | | |
| MTGAE | 0.790 | **0.718** | **0.804** |
| GCN [3] | **0.815** | 0.703 | 0.790 |
| Planetoid [4] | 0.757 | 0.647 | 0.772 |

### Discussion

- MTGAE model produces non-linear node embeddings optimized for both link prediction and node classification tasks;
- Multi-task learning improves semi-supervised node classification on datasets with extremely low label rate;
- Parameter sharing between the encoder and decoder significantly boosts representation learning and generalization;
- Future work will explore inductive reasoning on out-of-network nodes and mitigate $\mathcal{O}(N)$ complexity for enhanced scalability on large, dynamic graphs.


**References**

[1] D. Wang, P. Cui, and W. Zhu. Structural deep network embedding. *KDD 2016.*

[2] T.N. Kipf and M. Welling. Variational graph auto-encoders. *Bayesian Deep Learning Workshop, NeurIPS 2016.*

[3] T.N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *ICLR 2017.*

[4] Z. Yang, W. Cohen, and R. Salakhutdinov. Revisiting semi-supervised learning with graph embeddings. *ICML 2016.*


**Booz | Allen | Hamilton**