Edge Latents Guided Graph Attention Networks

STA-695 Final Project

Jiaying Weng, Liyu Gong

Department of Statistics
University of Kentucky

Outline

1. Introduction to graph and graph learning
2. Introduction to attention mechanism
3. Edge latents guided graph attention networks
4. Results
5. Future dirctions

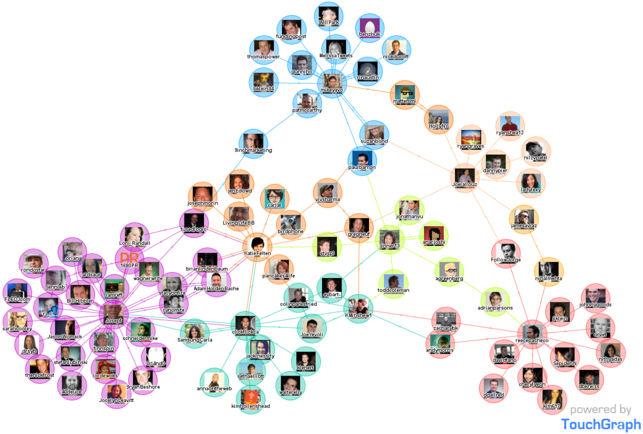
Outline

1. Introduction to graph and graph learning
2. Introduction to attention mechanism
3. Edge latents guided graph attention networks
4. Results
5. Future dirctions

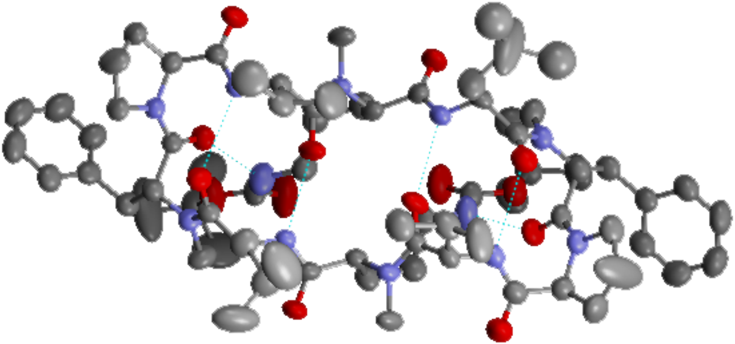
Graph

- A graph is a structure amounting to a set of objects in which some pairs of the objects are in some sense "related".
- Components
  - Nodes (vertices, points)
  - Edges (arcs, lines)
- Graph learning
  - Node classification (regression)
  - Edge classification (regression)
  - Graph classification (regression)
  - Node clustering
  - ...

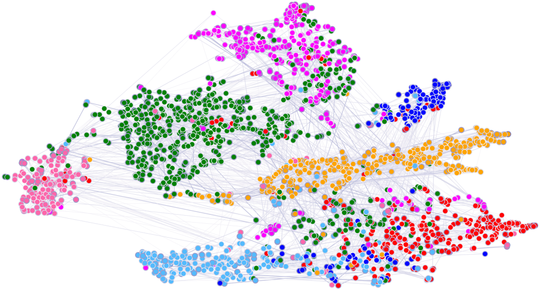
Social network



Protein-protein interaction



Citation graph



- Node: Paper
- Edge: Citation

Outline

1 Introduction to graph and graph learning

2 Introduction to attention mechanism

3 Edge latents guided graph attention networks

4 Results

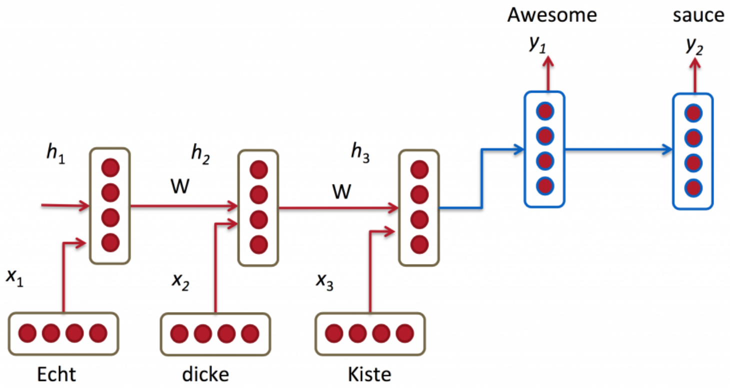
5 Future dirctions

Attention mechanism

- Hot and new topic in deep learning
- Successfully applied to machine translation problem
- A learnable mechanism to aggregate features
- Aggregate by weighted average
- The weights for aggregating is a function of features
- different content of features will have different weights (attention)

RNN based machine translation

- Input sequence are encoded into a single vector by RNN
- Output sequence are generated from this vector by RNN



Why attention mechanism?

- The RNN (LSTM) based method surprisingly works well!
- But, obviously, there is space for improvement

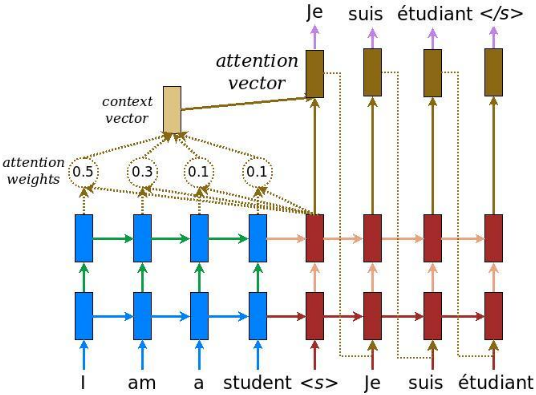
Problem

The dependency between output sequence and input sequence is complex which could not be captured by the encoding RNN

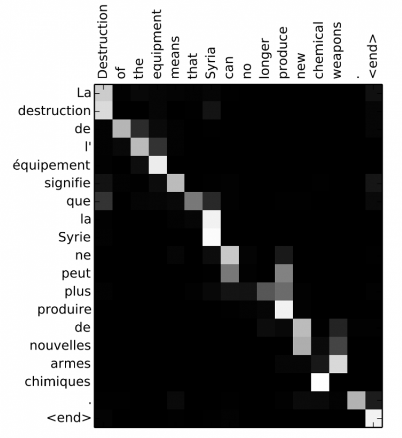
Attention mechanism

- Instead of using the last encoded vector, aggregate a feature vector from all the previous hidden states each time generating an output
- Aggregating could be a weighted average
- Weights are a trainable function of the contents

Machine translation with attention mechanism



Interpreting attention weights



Outline

1. Introduction to graph and graph learning
2. Introduction to attention mechanism
3. Edge latents guided graph attention networks
4. Results
5. Future dirctions

Graph node classification with attention mechanism

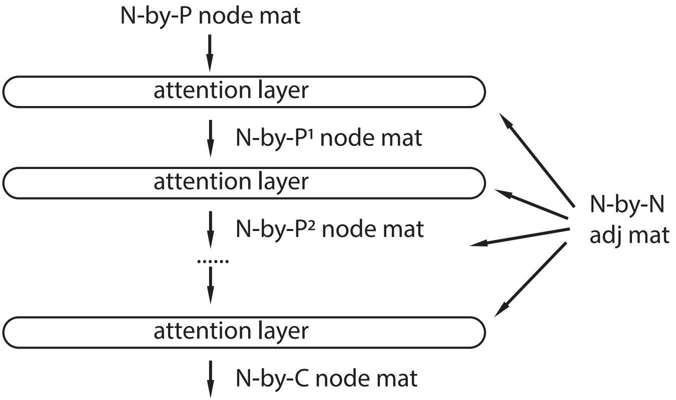
Useful information to classify a node

- features of the node
- features of its neighbor nodes
- label or prediction of its neighbor nodes

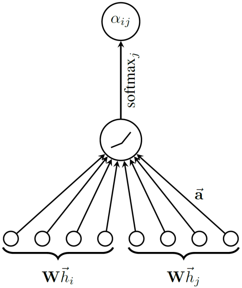
Graph attention network

- aggregate node features using multiple attention layers
- simutaneously classify all nodes of a graph
- neighbor label or prediction are aggregated through back-propagation

Graph attention network

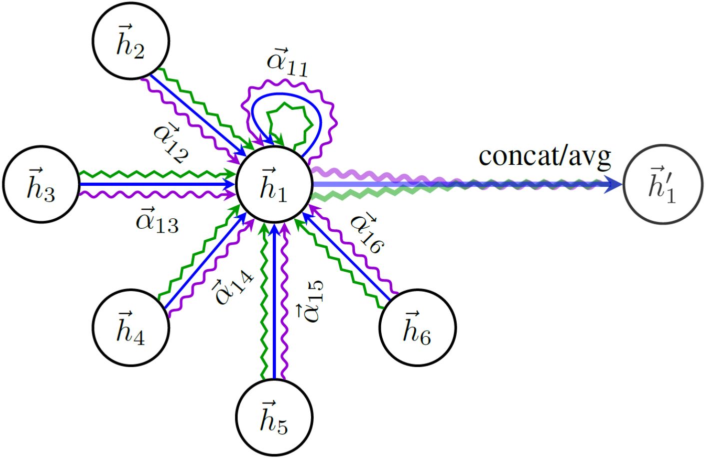


Graph attention mechanism

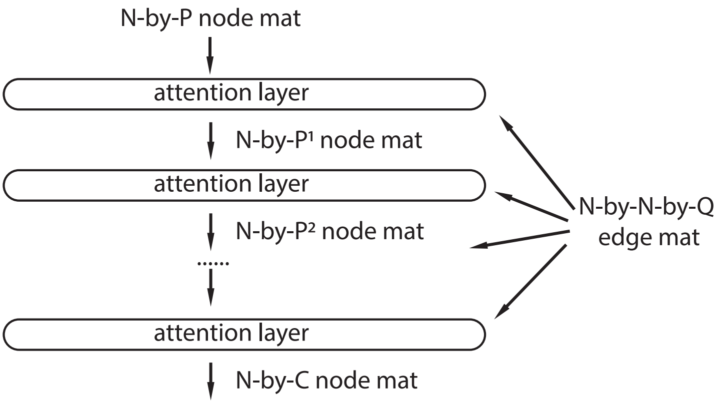


$$\alpha_{ij} = \frac{\exp\left(\mathrm{LeakyReLU}\left(\boldsymbol{a}^T\left[\boldsymbol{W}\boldsymbol{h}_i \,\|\, \boldsymbol{W}\boldsymbol{h}_j\right]\right)\right)}{\sum_{j\in\mathcal{N}_i}\exp\left(\mathrm{LeakyReLU}\left(\boldsymbol{a}^T\left[\boldsymbol{W}\boldsymbol{h}_i \,\|\, \boldsymbol{W}\boldsymbol{h}_j\right]\right)\right)}$$

Multi-head attention



Edge guided graph attention network



Edge latents guided attention

- For different type of edges, attention functions are different
- Edge types are represented by edge latent variables

$$\alpha_{ij} = \frac{\exp\left(\mathrm{LeakyReLU}\left(\sum_l \left(\boldsymbol{a}_l^T \left[\boldsymbol{W}\boldsymbol{h}_i \parallel \boldsymbol{W}\boldsymbol{h}_j\right] + \boldsymbol{v}_l \boldsymbol{g}_{ij}\right)\right)\right)}{\sum_{j \in \mathcal{N}_i} \exp\left(\mathrm{LeakyReLU}\left(\sum_l \left(\boldsymbol{a}_l^T \left[\boldsymbol{W}\boldsymbol{h}_i \parallel \boldsymbol{W}\boldsymbol{h}_j\right] + \boldsymbol{v}_l \boldsymbol{g}_{ij}\right)\right)\right)}$$

Outline

1. Introduction to graph and graph learning
2. Introduction to attention mechanism
3. Edge latents guided graph attention networks
4. Results
5. Future dirctions

Cora dataset

# Nodes	2708
# Edges	5429
# Features/Node	1433
# Features/Edge	2
# Classes	7
# Training Nodes	140
# Validation Nodes	500
# Test Nodes	1000

Table: Summary of cora dataset

Experimental Results

Method	Accuracy
MLP	55.1%
ManiReg [1]	59.5%
SemiEmb [2]	59.0%
LP [3]	68.0%
DeepWalk [4]	67.2%
ICA [5]	75.1%
Planetoid [6]	75.7%
Chebyshev [7]	81.2%
GCN [8]	81.5%
MoNet [9]	$81.7 \pm 0.5\%$
GAT	$83.0 \pm 0.7\%$
EGAT	$83.4 \pm 0.6\%$

Table: Summary of results

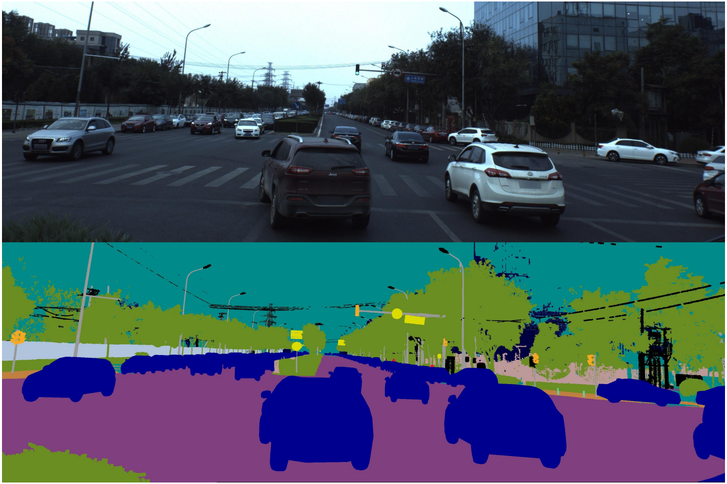
Outline

1. Introduction to graph and graph learning
2. Introduction to attention mechanism
3. Edge latents guided graph attention networks
4. Results
5. Future dirctions

Future directions

- Apply to problems with more edge information
- Address problems with large graphs
- Combine with GCN
- Apply to image semantic segmentation (FCN/MS-D + EGAT)

Semantic segmentation



Semantic segmentation as node classification

Current state-of-the-art method

- Convolution based
- Fully CNN or MS-D (Mixed scale dense CNN)
- Convolution weights are shared by all pixel position

Promising EGAT based method

- Model the problem as a graph node classification
- Fully CNN or MS-D for feature learning
- EGAT for feature aggregation and pixel classification

References I

- M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *Journal of Machine Learning Research*, vol. 7, pp. 2399–2434, 2006.
- J. Weston, F. Ratle, H. Mobahi, and R. Collobert, "Deep learning via semi-supervised embedding," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7700, pp. 639–655, 2012.
- X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions," in *ICML*, p. 8, 2003.
- B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online Learning of Social Representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, (New York, NY, USA), pp. 701–710, ACM, 2014.
- L. Getoor, "Link-based Classification," in *Advanced Methods for Knowledge Discovery from Complex Data*, Advanced Information and Knowledge Processing, pp. 189–207, Springer, London, 2005.
- Z. Yang, W. W. Cohen, and R. Salakhutdinov, "Revisiting Semi-Supervised Learning with Graph Embeddings," *arXiv:1603.08861 [cs]*, Mar. 2016.
- M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," in *NIPS*, p. 9, 2016.
- T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," *arXiv:1609.02907 [cs, stat]*, Sept. 2016.
- F. Monti, D. Boscaini, J. Masci, E. Rodol, J. Svoboda, and M. M. Bronstein, "Geometric Deep Learning on Graphs and Manifolds Using Mixture Model CNNs," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5425–5434, July 2017.