Graph Neural Net Applications for Natural Language Processing

Xipeng Qiu
Fudan University
xpqiu@fudan.edu.cn

Zhijing Jin & Xiangkun Hu
Amazon Shanghai AI Lab
zhijing.jin@connect.hku.hk
xiangkhu@amazon.com

Introduction

Graphs are widely used in natural language processing (NLP) research from the era of expert systems, such as WordNet (Fellbaum 1998), a graph-based representations of words' meanings through their relations with other words (Quillian 1968). Graph-based representation of a text item (document, passage or sentence) is usually called text graph. There is an annual workshop, TextGraphs,[1] intended to encourage the synergy between the fields of natural language processing (NLP) and graph theory.

Depending on the NLP application, the nodes and edges of text graph may represent a variety of language-related units and links. Nodes can represent text units of various sizes and characteristics, e.g., words, collocations, word senses, entities, sentences or even documents. Edges can encode relationships like co-occurrence (two words appearing together in a text unit), collocation (two words appearing next to each other or separated by a function word), syntactic structure (e.g., the parent and child in a syntactic dependency), semantic relations (e.g., anaphora), lexical similarity (e.g., cosine between the vector representations of two sentences), or links to a knowledge base.

Recently, graph neural networks (GNNs) have raised the great interests of researchers in NLP field, since GNNs effectively combine graph-based algorithms with deep learning architectures.

Structure of TextGraph

The graph structure is the basic architecture of GNNs. Therefore, when applying GNNs to NLP tasks, we need to answer the following questions: (1) what should be modeled as nodes, (2) what should be modeled as edges. We can categorize different text graphs into the following levels:

1. Statistical Level: The nodes may be any text unit, such as word, sentences or documents. The edges are usually constructed by statistic methods, such as co-occurrence, point-wise mutual information (PMI). This kind of text graphs usually is heterogeneous graph, such as Text GCN (Yao et al., 2019b), HeterGraph (Wang et al., 2020). The traditional word-document co-occurrence matrix is a naturally graph.
2. Syntactic Level: The nodes are usually words and the edges are syntactic relations between words. Since syntactic structure of text can represent the intrinsic relations among words, the graph-based methods can better integrate the interaction than sequence-based methods. This kind of text graphs includes Tree-LSTM (Tai et al., 2015b), BiLSTM+GCN (Marcheggiani & Titov, 2017), etc.
3. Semantic Level: This kind of text graphs usually represent the semantic relations between text units, such as words, entities, or documents. For example, the nodes in citation network are academic papers and the edges are citation relations. Some others graphs include word similarity graph where the edges are weighted by some similarity functions, coreference graph which links the different mentions of the same entity, AMR, and so on.
4. Knowledge Level: Since external knowledge is usually helpful for language understanding, it is natural to encode text with help of external knowledge. The external knowledge has

[1] http://www.textgraphs.org/

graph-based form, called knowledge graph (KG). To link the text and knowledge, we need recognize the entity mentions in text via the NER tool and associate them with the entities in KG.

Besides, there is also a special graph to encode text, fully-connected graph, which does not depend on any assumption and usually expects to find some latent graph via dynamic attention mechanism.

The above four categories of text graphs have their own advantages and weaknesses in practice. For example, the semantic and knowledge level graphs are largely reliant on external resources, so they perform well in scenarios with abundant, accurate external knowledge, but correspondinly are constrained in other cases. Another rising research question is whether it is better to use graph edges defined by hand-crafted features or learn the edges from data. Previously, most work constructs edges in a hand-crafted way, using linguistic features such as syntax trees, word co-occurrence, and PMI. A counter way is to learn from data, – for example, Transformer (Vaswani et al., 2017) is a successful instance of GNNs where each token is a node, and weights of edges are learned through attention. Recent advances have shown that Transformer, after large-scale pretraining Devlin et al. (2019), are powerful models, so it remains a research question whether data-driven learning or pre-defined graphs is better.

Graph Generation Besides as encoding representation, how generating the last three kinds of text graphs is also interesting. For example, current SOTA parser is a graph-based method, which predicts the relations among all the words and then generates the dependency tree using MST.

INFORMATION AGGREGATION OF TEXTGRAPH

Another key element of GNNs is the aggregator, namely how information is aggregated in the graph. There are several main methods to aggregate information: mean pooling as proposed in GCN (Kipf & Welling, 2017), weighted sum by multi-head attention as in GAT (Velickovic et al., 2018), and type-specific message passing as in Relational-GCN (Schlichtkrull et al., 2018). Sum or mean pooling is widely used in early applications of GNNs to NLP, and problem settings where the relations among nodes are simple. The majority of graph-based NLP models use attention to aggregate node informations, as the learnable attention weights can adapt to various scenarios. Edge-specific message passing is also useful in many applications such as graphs with heteorogeneous nodes and edge types. More complicated aggregators as in hypergraph neural networks (Feng et al., 2019), and sparse graph attention networks (Ye & Ji, 2019) are less frequent in NLP literature. We list the main GNN-based models, and their corresponding model details in Table 1.

Model	Node	Edge	Aggregator	Task
Text Encoding				
Transformer (Vaswani et al., 2017)	Word	Full connection	Multi-head attention	Any text
LeviGraphGRU (Beck et al., 2018)	Word, AMR/syntax elements	Levi graph	GRU-like gating	NMT, AMR-to-text
GAMT (Chen et al., 2020)	Word	Full connection	Attention	Utterance encoding
GP-GNN (Zhu et al., 2019)	Entity	Full connection	Sum	Relation extraction
AGGCN (Guo et al., 2019)	Word	Full connection	Attention, dense net	Relation extraction
Eo-GNN (Christopoulou et al., 2019)	Mention, entity, sentence	Co-occurrence, appearance, full connection for sentences	Weighted sum	Relation extraction
HGAT (Linmei et al., 2019)	Topic, document, entity	Heterogeneous graph, document-topic, entity-document	Attention	Semi-supervised text classification
Text GCN (Yao et al., 2019b)	Word, document	Word-document, co-occurrence in corpus	Same as GCN	Text classification
CogQA (Ding et al., 2019)	Entity	Dynamically built edges	Localized spectral filter	Multi-hop QA
HSG (Wang et al., 2020)	Word, sentence	Word-sentence	Attention	Summarization
Syntactic GCN				
BiLSTM+GCN (Marcheggiani & Titov, 2017)	Word	Syntax tree	Same as GCN	Semantic role labeling
BiRNN+GCN (Bastings et al., 2017)	Word	Syntax, semantic relations	Weighted sum	NMT
C-GCN (Zhang et al., 2018b)	Word	Syntax tree	Mean Pooling	Relation extraction
RESIDE (Vashishth et al., 2018b)	Word, entity	Syntax tree	Weighted sum	Relation extraction
BiLSTM+GCN (Sahu et al., 2019)	Word	Syntax tree, coreference, adjacent sentence, adjacent word	Same as GCN	Relation extraction
GraphRel (Fu et al., 2019)	Word	Dependency tree	Sum	Joint NER and RE
BiLSTM+GCN (Sun et al., 2019)	Entity	Entity-relation bipartite graph	Same as GCN	Joint NER and RE
MUL_GCN (Le et al., 2020)	Word	Dependency tree	Attention	Metaphor detection
JMEE (Liu et al., 2018b)	Word	Syntax tree	Weighted sum	Event detection
BiLSTM+GCN+AAP (Nguyen & Grishman, 2018b)	word	Syntax tree	Weighted sum	Event detection
Tree-LSTM (Tai et al., 2015b)	Word	Constituency parse tree	Sum and LSTM cell	Sentiment analysis
NeuralDater (Vashishth et al., 2018a)	Word	Syntax tree	Mean Pooling	Document timestamping
AD3 (Ray et al., 2018)	Word	Syntax tree	Attention	Document timestamping
SynGCN, SemGCN (Vashishth et al., 2019)	Word	Syntax tree	Weighted sum	Word embedding
Encoding Tabular Structure				
GraphIE (Qian et al., 2019)	Cell	Position (neighboring cells)	Attention	Encode tables
CFGNN (Zhang, 2020)	Cell	Data info flow	Attention	Encode tables (Text-to-SQL)
ER-GCN (Li et al., 2020)	Token, attribute	record-attribute, attribute-inclusive token, token co-occurrence	Attention	Entity resolution
GAMT (Chen et al., 2020)	Slot, domain	Slot-its domain, multiple tokens in a slot	Attention	DST
RGAT (Chen et al., 2020)	Slot	same-domain slots, slots with overlapping candidate values	Attention	DST
Knowledge Graphs				
CRAFT-NET (Sun et al., 2018)	KB entity, document	KB relation, entity mention	Sum	Open Domain QA
GraphWriter (Koncel-Kedziorski et al., 2019)	Entity	KG relation	Attention	Graph-to-Text
MuGNN (Cao et al., 2019)	Entity	KG relation	Attention	Entity Alignment
LAN (Wang et al., 2019)	Entity	KG relation	Attention	KG Embedding
Other				
RRN (Palm et al., 2018)	Entity, number	Action, Relative Position	Sum and RNN cell	Symbolic Reasoning (bAbI, Sudoku)

Table 1: Main GNN-based NLP models, with their corresponding node type, edge definition, aggregator, and task.

Categorization by Tasks

GNNs are widely used in recent natural language processing (NLP) research. They are first introduced to classification tasks (Hamilton et al., 2017; Defferrard et al., 2016; Kipf & Welling, 2017), then extended to other natural language understanding (NLU) tasks such as sequence labeling (Zhang et al., 2018a) and semantic role labeling (Marcheggiani & Titov, 2017), and finally to text generation such as neural machine translation (Bastings et al., 2017; Marcheggiani et al., 2018) and text generation from Abstract Meaning Representation (AMR) (Song et al., 2018b; Beck et al., 2018). We list NLP tasks with their corresponding GNN-based models as follows:

- Text classification: GCN (Hamilton et al., 2017; Atwood & Towsley, 2016; Defferrard et al., 2016; Kipf & Welling, 2017; Monti et al., 2017; Henaff et al., 2015), GAT (Velickovic et al., 2018), GraphCNN(Peng et al., 2018), TextGCN (Yao et al., 2019a), Graph-State LSTM (Zhang et al., 2018a), Tree LSTM (Tai et al., 2015a)
- Sequence labeling (POS, NER): Graph-State LSTM (Zhang et al., 2018a)
- Semantic role labeling: Syntactic GCN (Marcheggiani & Titov, 2017)
- Neural machine translation: Syntactic GCN (Bastings et al., 2017; Marcheggiani et al., 2018), GatedGNN (Beck et al., 2018)
- Relation extraction: Tree LSTM (Miwa & Bansal, 2016), Graph LSTM (Peng et al., 2017; Song et al., 2018c), Syntactic GCN (Zhang et al., 2018b)
- Event extraction: Syntactic GCN (Nguyen & Grishman, 2018a; Liu et al., 2018a)
- AMR to text generation: Graph-to-LSTM (Song et al., 2018b), GatedGNN (Beck et al., 2018)
- Multih-hop reading comprehension: Graph-to-LSTM (Song et al., 2018a)
- Relational reasoning: Relational Networks (RN) (Santoro et al., 2017), Recurrent RN (Santoro et al., 2018), Interaction Networks (Battaglia et al., 2016)

REFERENCES

James Atwood and Don Towsley. Diffusion-convolutional neural networks. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pp. 1993–2001, 2016. URL `http://papers.nips.cc/paper/6212-diffusion-convolutional-neural-networks`.

Jasmijn Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, and Khalil Sima'an. Graph convolutional encoders for syntax-aware neural machine translation. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel (eds.), *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pp. 1957–1967. Association for Computational Linguistics, 2017. doi: 10.18653/v1/d17-1209. URL `https://doi.org/10.18653/v1/d17-1209`.

Peter W. Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, and Koray Kavukcuoglu. Interaction networks for learning about objects, relations and physics. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pp. 4502–4510, 2016. URL `http://papers.nips.cc/paper/6418-interaction-networks-for-learning-about-objects-relations-and-physics`.

Daniel Beck, Gholamreza Haffari, and Trevor Cohn. Graph-to-sequence learning using gated graph neural networks. In Iryna Gurevych and Yusuke Miyao (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pp. 273–283. Association for Computational Linguistics, 2018. doi: 10.18653/v1/P18-1026. URL `https://www.aclweb.org/anthology/P18-1026/`.

Yixin Cao, Zhiyuan Liu, Chengjiang Li, Zhiyuan Liu, Juanzi Li, and Tat-Seng Chua. Multi-channel graph neural network for entity alignment. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1452–1461, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1140. URL `https://www.aclweb.org/anthology/P19-1140`.

Lu Chen, Boer Lv, Chi Wang, Su Zhu, Bowen Tan, and Kai Yu. Schema-guided multi-domain dialogue state tracking with graph attention neural networks. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 7521–7528. AAAI Press, 2020. URL `https://aaai.org/ojs/index.php/AAAI/article/view/6250`.

Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 4927–4938. Association for Computational Linguistics, 2019.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pp. 3837–3845, 2016. URL `http://papers.nips.cc/paper/6081-convolutional-neural-networks-on-graphs-with-fast-localized-spectral-filterin`

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT*

2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), pp. 4171–4186. Association for Computational Linguistics, 2019. doi: 10.18653/v1/n19-1423. URL https://doi.org/10.18653/v1/n19-1423.

Ming Ding, Chang Zhou, Qibin Chen, Hongxia Yang, and Jie Tang. Cognitive graph for multi-hop reading comprehension at scale. *arXiv preprint arXiv:1905.05460*, 2019.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pp. 3558–3565. AAAI Press, 2019. doi: 10.1609/aaai.v33i01.33013558. URL https://doi.org/10.1609/aaai.v33i01.33013558.

Tsu-Jui Fu, Peng-Hsuan Li, and Wei-Yun Ma. GraphRel: Modeling text as relational graphs for joint entity and relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1409–1418, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1136. URL https://www.aclweb.org/anthology/P19-1136.

Zhijiang Guo, Yan Zhang, and Wei Lu. Attention guided graph convolutional networks for relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 241–251, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1024. URL https://www.aclweb.org/anthology/P19-1024.

William L. Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pp. 1024–1034, 2017. URL http://papers.nips.cc/paper/6703-inductive-representation-learning-on-large-graphs.

Mikael Henaff, Joan Bruna, and Yann LeCun. Deep convolutional networks on graph-structured data. *CoRR*, abs/1506.05163, 2015. URL http://arxiv.org/abs/1506.05163.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Rik Koncel-Kedziorski, Dhanush Bekal, Yi Luan, Mirella Lapata, and Hannaneh Hajishirzi. Text generation from knowledge graphs with graph transformers. *arXiv preprint arXiv:1904.02342*, 2019.

Duong Le, My Thai, and Thien Nguyen. Multi-task learning for metaphor detection with graph convolutional neural networks and word sense disambiguation. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 8139–8146. AAAI Press, 2020. URL https://aaai.org/ojs/index.php/AAAI/article/view/6326.

Bing Li, Wei Wang, Yifang Sun, Linhan Zhang, Muhammad Asif Ali, and Yi Wang. Grapher: Token-centric entity resolution with graph convolutional neural networks. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 8172–8179. AAAI Press, 2020. URL https://aaai.org/ojs/index.php/AAAI/article/view/6330.

Hu Linmei, Tianchi Yang, Chuan Shi, Houye Ji, and Xiaoli Li. Heterogeneous graph attention networks for semi-supervised short text classification. In *Proceedings of the 2019 Conference* *on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 4821–4830, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1488. URL https://www.aclweb.org/anthology/D19-1488.

Xiao Liu, Zhunchen Luo, and Heyan Huang. Jointly multiple events extraction via attention-based graph information aggregation. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pp. 1247–1256. Association for Computational Linguistics, 2018a. doi: 10.18653/v1/d18-1156. URL https://doi.org/10.18653/v1/d18-1156.

Xiao Liu, Zhunchen Luo, and Heyan Huang. Jointly multiple events extraction via attention-based graph information aggregation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1247–1256, Brussels, Belgium, October-November 2018b. Association for Computational Linguistics. doi: 10.18653/v1/D18-1156. URL https://www.aclweb.org/anthology/D18-1156.

Diego Marcheggiani and Ivan Titov. Encoding sentences with graph convolutional networks for semantic role labeling. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel (eds.), *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pp. 1506–1515. Association for Computational Linguistics, 2017. doi: 10.18653/v1/d17-1159. URL https://doi.org/10.18653/v1/d17-1159.

Diego Marcheggiani, Jasmijn Bastings, and Ivan Titov. Exploiting semantics in neural machine translation with graph convolutional networks. In Marilyn A. Walker, Heng Ji, and Amanda Stent (eds.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers)*, pp. 486–492. Association for Computational Linguistics, 2018. doi: 10.18653/v1/n18-2078. URL https://doi.org/10.18653/v1/n18-2078.

Makoto Miwa and Mohit Bansal. End-to-end relation extraction using lstms on sequences and tree structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics, 2016. doi: 10.18653/v1/p16-1105. URL https://doi.org/10.18653/v1/p16-1105.

Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodolà, Jan Svoboda, and Michael M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pp. 5425–5434. IEEE Computer Society, 2017. doi: 10.1109/CVPR.2017.576. URL https://doi.org/10.1109/CVPR.2017.576.

Thien Huu Nguyen and Ralph Grishman. Graph convolutional networks with argument-aware pooling for event detection. In Sheila A. McIlraith and Kilian Q. Weinberger (eds.), *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 5900–5907. AAAI Press, 2018a. URL https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16329.

Thien Huu Nguyen and Ralph Grishman. Graph convolutional networks with argument-aware pooling for event detection. In *AAAI*, volume 18, pp. 5900–5907, 2018b.

Rasmus Palm, Ulrich Paquet, and Ole Winther. Recurrent relational networks. In *Advances in Neural Information Processing Systems*, pp. 3368–3378, 2018.

Hao Peng, Jianxin Li, Yu He, Yaopeng Liu, Mengjiao Bao, Lihong Wang, Yangqiu Song, and Qiang Yang. Large-scale hierarchical text classification with recursively regularized deep graph-cnn. In Pierre-Antoine Champin, Fabien L. Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis (eds.), *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018*, pp. 1063–1072. ACM, 2018. doi: 10.1145/3178876.3186005. URL https://doi.org/10.1145/3178876.3186005.

Nanyun Peng, Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen-tau Yih. Cross-sentence n-ary relation extraction with graph lstms. *Trans. Assoc. Comput. Linguistics*, 5:101–115, 2017. URL https://transacl.org/ojs/index.php/tacl/article/view/1028.

Yujie Qian, Enrico Santus, Zhijing Jin, Jiang Guo, and Regina Barzilay. Graphie: A graph-based framework for information extraction. In Jill Burstein, Christy Doran, and Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pp. 751–761. Association for Computational Linguistics, 2019. doi: 10.18653/v1/n19-1082. URL https://doi.org/10.18653/v1/n19-1082.

Swayambhu Nath Ray, Shib Sankar Dasgupta, and Partha Talukdar. AD3: Attentive deep document dater. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1871–1880, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1213. URL https://www.aclweb.org/anthology/D18-1213.

Sunil Kumar Sahu, Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. Inter-sentence relation extraction with document-level graph convolutional neural network. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4309–4316, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1423. URL https://www.aclweb.org/anthology/P19-1423.

Adam Santoro, David Raposo, David G. T. Barrett, Mateusz Malinowski, Razvan Pascanu, Peter W. Battaglia, and Tim Lillicrap. A simple neural network module for relational reasoning. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pp. 4967–4976, 2017. URL http://papers.nips.cc/paper/7082-a-simple-neural-network-module-for-relational-reasoning.

Adam Santoro, Ryan Faulkner, David Raposo, Jack W. Rae, Mike Chrzanowski, Theophane Weber, Daan Wierstra, Oriol Vinyals, Razvan Pascanu, and Timothy P. Lillicrap. Relational recurrent neural networks. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada*, pp. 7310–7321, 2018. URL http://papers.nips.cc/paper/7960-relational-recurrent-neural-networks.

Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam (eds.), *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, volume 10843 of *Lecture Notes in Computer Science*, pp. 593–607. Springer, 2018. doi: 10.1007/978-3-319-93417-4_38. URL https://doi.org/10.1007/978-3-319-93417-4_38.

Linfeng Song, Zhiguo Wang, Mo Yu, Yue Zhang, Radu Florian, and Daniel Gildea. Exploring graph-structured passage representation for multi-hop reading comprehension with graph neural networks. *arXiv preprint arXiv:1809.02040*, 2018a.

Linfeng Song, Yue Zhang, Zhiguo Wang, and Daniel Gildea. A graph-to-sequence model for amr-to-text generation. In Iryna Gurevych and Yusuke Miyao (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pp. 1616–1626. Association for Computational Linguistics, 2018b. doi: 10.18653/v1/P18-1150. URL https://www.aclweb.org/anthology/P18-1150/.

Linfeng Song, Yue Zhang, Zhiguo Wang, and Daniel Gildea. N-ary relation extraction using graph-state LSTM. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pp. 2226–2235. Association for Computational Linguistics, 2018c. doi: 10.18653/v1/d18-1246. URL `https://doi.org/10.18653/v1/d18-1246`.

Changzhi Sun, Yeyun Gong, Yuanbin Wu, Ming Gong, Daxin Jiang, Man Lan, Shiliang Sun, and Nan Duan. Joint type inference on entities and relations via graph convolutional networks. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1361–1370, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1131. URL `https://www.aclweb.org/anthology/P19-1131`.

Haitian Sun, Bhuwan Dhingra, Manzil Zaheer, Kathryn Mazaitis, Ruslan Salakhutdinov, and William W Cohen. Open domain question answering using early fusion of knowledge bases and text. *arXiv preprint arXiv:1809.00782*, 2018.

Kai Sheng Tai, Richard Socher, and Christopher D. Manning. Improved semantic representations from tree-structured long short-term memory networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pp. 1556–1566. The Association for Computer Linguistics, 2015a. doi: 10.3115/v1/p15-1150. URL `https://doi.org/10.3115/v1/p15-1150`.

Kai Sheng Tai, Richard Socher, and Christopher D. Manning. Improved semantic representations from tree-structured long short-term memory networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 1556–1566, Beijing, China, July 2015b. Association for Computational Linguistics. doi: 10.3115/v1/P15-1150. URL `https://www.aclweb.org/anthology/P15-1150`.

Shikhar Vashishth, Shib Sankar Dasgupta, Swayambhu Nath Ray, and Partha Talukdar. Dating documents using graph convolution networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1605–1615, Melbourne, Australia, July 2018a. Association for Computational Linguistics. doi: 10.18653/v1/P18-1149. URL `https://www.aclweb.org/anthology/P18-1149`.

Shikhar Vashishth, Rishabh Joshi, Sai Suman Prayaga, Chiranjib Bhattacharyya, and Partha Talukdar. RESIDE: Improving distantly-supervised neural relation extraction using side information. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1257–1266, Brussels, Belgium, October-November 2018b. Association for Computational Linguistics. doi: 10.18653/v1/D18-1157. URL `https://www.aclweb.org/anthology/D18-1157`.

Shikhar Vashishth, Manik Bhandari, Prateek Yadav, Piyush Rai, Chiranjib Bhattacharyya, and Partha Talukdar. Incorporating syntactic and semantic information in word embeddings using graph convolutional networks. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3308–3318, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1320. URL `https://www.aclweb.org/anthology/P19-1320`.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pp. 5998–6008, 2017. URL `http://papers.nips.cc/paper/7181-attention-is-all-you-need`.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *6th International Conference on Learning Representations,* *ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Danqing Wang, Pengfei Liu, Yining Zheng, Xipeng Qiu, and Xuanjing Huang. Heterogeneous graph neural networks for extractive document summarization. *arXiv preprint arXiv:2004.12393*, 2020.

Peifeng Wang, Jialong Han, Chenliang Li, and Rong Pan. Logic attention based neighborhood aggregation for inductive knowledge graph embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 7152–7159, 2019.

Liang Yao, Chengsheng Mao, and Yuan Luo. Graph convolutional networks for text classification. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pp. 7370–7377. AAAI Press, 2019a. doi: 10.1609/aaai.v33i01.33017370. URL https://doi.org/10.1609/aaai.v33i01.33017370.

Liang Yao, Chengsheng Mao, and Yuan Luo. Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 7370–7377, 2019b.

Yang Ye and Shihao Ji. Sparse graph attention networks. *CoRR*, abs/1912.00552, 2019. URL http://arxiv.org/abs/1912.00552.

Xuanyu Zhang. CFGNN: cross flow graph neural networks for question answering on complex tables. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 9596–9603. AAAI Press, 2020. URL https://aaai.org/ojs/index.php/AAAI/article/view/6506.

Yue Zhang, Qi Liu, and Linfeng Song. Sentence-state LSTM for text representation. In Iryna Gurevych and Yusuke Miyao (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pp. 317–327. Association for Computational Linguistics, 2018a. doi: 10.18653/v1/P18-1030. URL https://www.aclweb.org/anthology/P18-1030/.

Yuhao Zhang, Peng Qi, and Christopher D. Manning. Graph convolution over pruned dependency trees improves relation extraction. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pp. 2205–2215. Association for Computational Linguistics, 2018b. doi: 10.18653/v1/d18-1244. URL https://doi.org/10.18653/v1/d18-1244.

Hao Zhu, Yankai Lin, Zhiyuan Liu, Jie Fu, Tat-Seng Chua, and Maosong Sun. Graph neural networks with generated parameters for relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1331–1339, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1128. URL https://www.aclweb.org/anthology/P19-1128.