LREC-COLING 2024

浙江大学 计算机科学与技术学院
COLLEGE OF COMPUTER SCIENCE AND TECHNOLOGY
ZHEJIANG UNIVERSITY

ZJUKG

DET: A Dual-Encoding Transformer for Relational Graph Embedding

Lingbing Guo*, Zhuo Chen, Jiaoyan Chen,
Qiang Zhang, and Huajun Chen

*: Corresponding author
github.com/zjukg/NCGNN

Motivation

AlphaFold uses the proteins with high MSA scores as augmented data to predict the 3D structures. Its input is not in the form of protein sequence but the alignment results produced by the MSA algorithms.

Even closely related proteins may have different lengths, encoding and non-encoding regions.

Different amino acids can be also replaced with each other safely in certain circumstances.

The alignment algorithms (e.g. Smith-Waterman) aim to find an alignment path with maximal score to support the comparison between proteins.

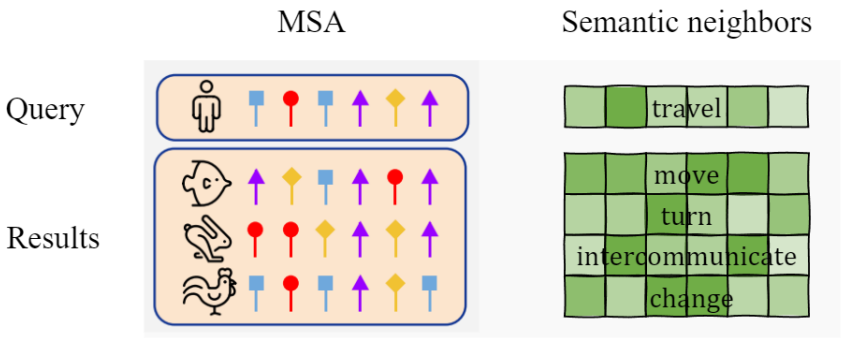
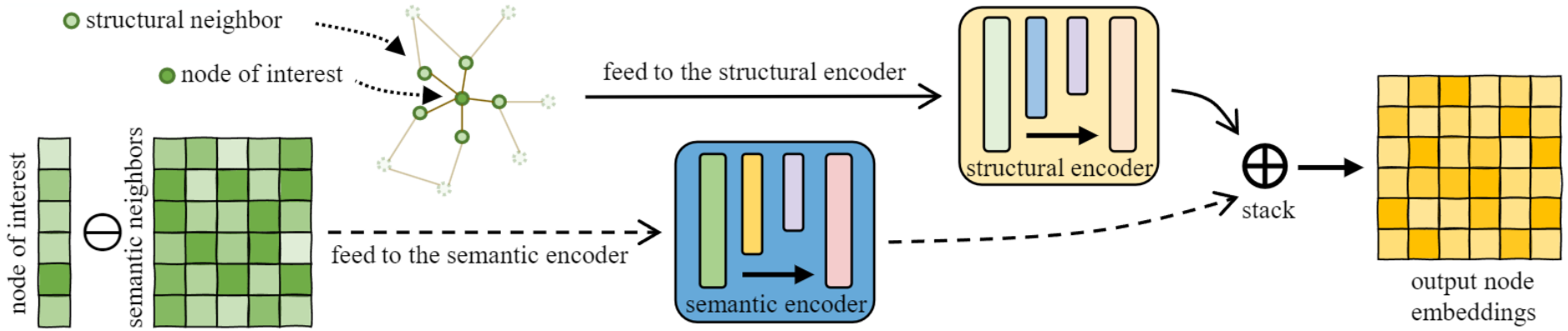


Figure 2: A comparison between MSA and semantic neighbors. The left figure is sliced from AlphaFold (Jumper et al., 2021). The right figure is an example from WordNet (Miller, 1995).

Method



For structural encoding, we use the standard Transformer layer to encode the structural neighbors.

For semantic encoding, we use the semantic operator to find and encode the semantic neighbors.

The dual encoding ensures both local aggregation and global connection, and also enables them to benefit from each other through back propagation.

Method

DET block starts from a structural encoding layer whose output embeddings will contain the local neighborhood information, functioning like encoding the amino acid sequences of proteins.

Then, the following semantic encoding layer will estimate the importance of the `` family members'' by their local context information.

By alternative stacking these two layers, these two types of encoding layers can support and enrich each other.

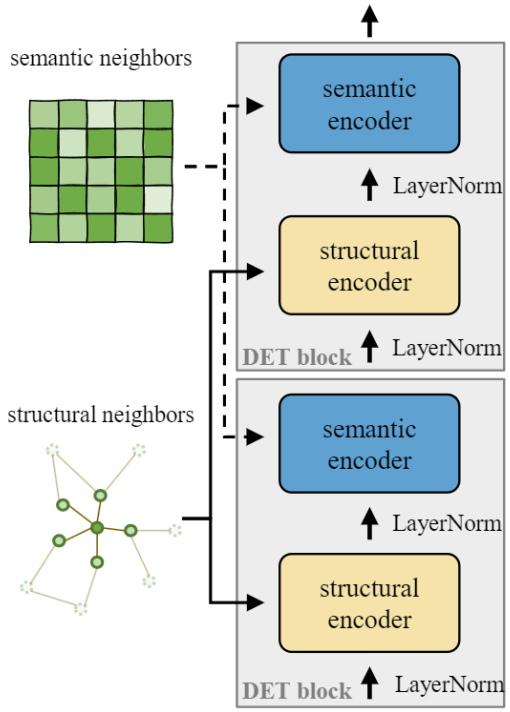


Figure 3: Example of a two-layer DET. The structural encoder and semantic encoder are stacked alternatively and feed with their neighborhood information respectively.

Two Hypothesizes

Hypothesis 1. *The local neighbors are the most informative features to identify and represent the node of interest.*

Hypothesis 2. *The distant nodes with high mutual information scores are important features to identify and represent the node of interest.*

Table 1: The occurrence frequency of entities in FB15K-237 and WN18RR, in term of hops.

Dataset	1-hop	2-hop	3-hop	5-hop
WN18RR	2.7	8.9	30.5	483.8
FB15K-237	20.3	1781.4	64,774.9	-

Algorithm 1 Dual-encoding Transformer

Input: the input graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the main prediction loss $\mathcal{L}_{\text{main}}$, the DET model $\mathcal{M}$;
Initialize all parameters;
repeat
  Update the semantic neighbors if necessary;
  for each batch data $(\boldsymbol{X}, \boldsymbol{Y})$ **do**
    $\boldsymbol{H} \leftarrow \boldsymbol{X}$;
    for each DET block $(\mathcal{M}^{st}, \mathcal{M}^{se})$ **do**
      $\boldsymbol{H} \leftarrow \mathcal{M}^{st}(\boldsymbol{H})$ (Equation (3));
      $\boldsymbol{H} \leftarrow \mathcal{M}^{se}(\boldsymbol{H})$ (Equation (7));
    end for
    $\mathcal{L} \leftarrow \mathcal{L}_{\text{main}}(\boldsymbol{H}, \boldsymbol{Y}) + \mathcal{L}_{\text{sn}}(\boldsymbol{H})$;
    Update the parameters of $\mathcal{M}$;
  end for
until the loss $\mathcal{L}$ converges;

Experiments

Graph property prediction

Table 2: Graph property prediction results on the PCQM4M-LSC dataset.

Model	#param.	train MAE	validate MAE
GCN	2.0M	0.1318	0.1691
DeeperGCN	25.5M	0.1059	0.1398
GraphSage	-	-	-
GIN	3.8M	0.1203	0.1537
GT	83.2M	0.0955	0.1408
Graphormer	47.1M	0.0582	0.1234
DET	47.1M	**0.0546**	**0.1212**

Table 3: Graph property prediction results on the ZINC dataset.

Model	#param.	test MAE
GCN	505,079	0.367
GraphSage	505,341	0.398
GIN	509,549	0.526
GAT	531,345	0.384
GT	588,929	0.226
Graphormer	489,321	0.122
DET	489,562	**0.113**

KG completion

Table 2: The entity prediction results on FB15K-237 and WN18RR. The results of the baselines are extracted from (Bi et al., 2022). The best and second-best results are **boldfaced** and underlined, respectively. ↑: higher is better; ↓: lower is better. -: unavailable entry.

Model	FB15K-237				WN18RR			
	MRR↑	MR↓	Hits@1↑	Hits@10↑	MRR↑	MR↓	Hits@1↑	Hits@10↑
TransE (Bordes et al., 2013)	.310	199	.218	.495	.232	5,249	.061	.522
RotatE (Sun et al., 2019)	.338	177	.241	.533	.476	3,340	.428	.571
TuckER (Balazevic et al., 2019)	.358	-	.266	.544	.470	-	.443	.526
RGCN (Schlichtkrull et al., 2018)	.273	221	.182	.456	.402	2,719	.345	.494
CoKE (Wang et al., 2019)	.364	-	.272	.549	.484	-	.450	.553
CompGCN (Vashishth et al., 2020)	.355	197	.264	.535	.479	3,533	.443	.546
Relphormer (Bi et al., 2022)	.371	-	**.314**	.481	.495	-	.448	**.591**
DET	**.376**	**150**	.281	**.560**	**.507**	**2,255**	**.465**	.585

Table 3: The accuracy results of node classification on five benchmarks.

Node classification

Model	Cora↑	Citeseer↑	Pumbed↑	Computer↑	Photo↑
GCN (Kipf and Welling, 2017)	87.33±0.38	79.43±0.26	84.86±0.19	89.65±0.52	92.70±0.20
GraphSage (Hamilton et al., 2017)	86.90±0.94	79.23±0.53	86.19±0.18	90.22±0.15	91.72±0.13
GAT (Velickovic et al., 2018)	86.29±0.53	80.13±0.62	84.40±0.05	90.78±0.13	93.87±0.11
GT (Dwivedi and Bresson, 2021)	71.84±0.62	67.38±0.76	82.11±0.39	91.18±0.17	94.74±0.13
SuperGAT (Kim and Oh, 2021)	82.70±0.60	72.50±0.80	81.30±0.50	77.44±0.26	84.53±0.32
SAN (Kreuzer et al., 2021)	74.02±1.01	70.64±0.97	86.22±0.43	89.83±0.16	94.86±0.10
Graphormer (Ying et al., 2021)	72.85±0.76	66.21±0.83	82.76±0.24	OOM	92.74±0.14
Gophormer (Zhao et al., 2021)	87.85±0.10	**80.23**±0.09	89.40±0.14	90.72±0.24	95.39±0.18
NAGphormer (Chen et al., 2023)	88.15±0.22	80.12±0.23	89.70±0.19	91.22±0.14	95.49±0.11
DET	**90.64**±0.27	80.14±0.35	**89.96**±0.20	**92.15**±0.11	**95.81**±0.13

Analysis

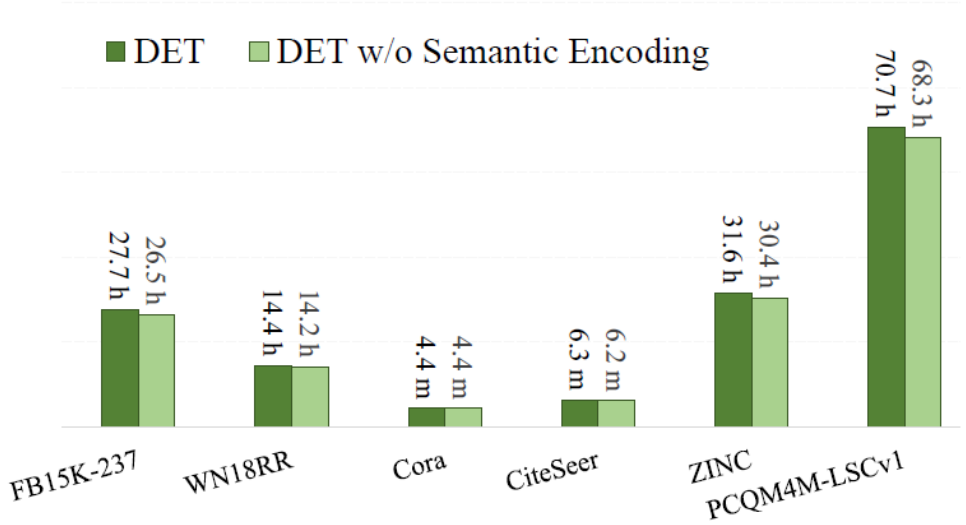


Figure 5: The training time (hours/minutes) of DET and DET w/o semantic encoding on six datasets.

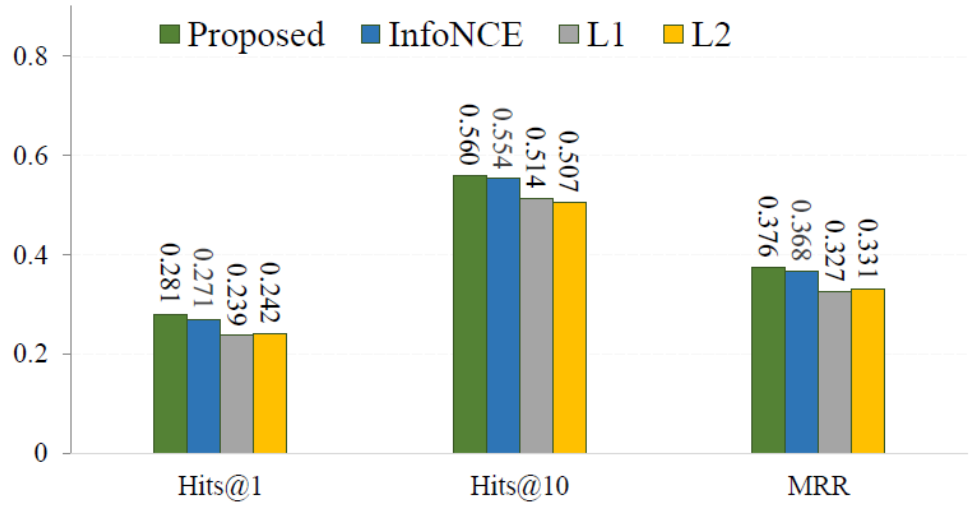
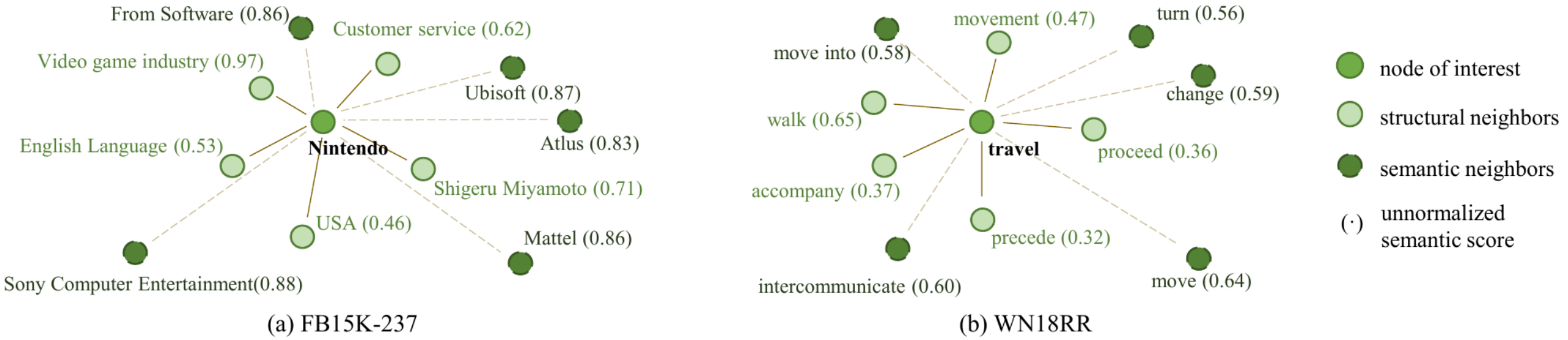


Figure 6: The performance of DET with different semantic estimators on FB15K-237.

Analysis



(a) FB15K-237 (b) WN18RR

Figure 4: Examples of the semantic attention scores to different types of neighbors.

Conclusion

- In this paper, we propose **DET** which achieves state-of-the-art performance across **9** different datasets.
- In DET, the **structural encoder** aggregates local nodes while the semantic encoder seeks for the remote nodes. Inspired by recent advances in biological sciences, DET finds the **semantic neighbors** with a **mutual-information-based operator** and stacks the two encoders alternatively.
- Bring more insights and inspirations in developing new Transformer architectures.

Thanks for your attention!

- **Code and datasets are available at https://github.com/zjukg/DET**

- This work is funded by New Generation AI Development Plan for 2030 of China (2023ZD0120802);
- National Natural Science Foundation of China (NSFC62302433, USFCU23A20496);
- Zhejiang Provincial Natural Science Foundation of China (No.LQ24F020007).