Distributed, Egocentric Representations of Graphs for Detecting Critical Structures

Ruo-Chun Tzeng[1] Shan-Hung Wu[2]

[1]Microsoft Inc, Taiwan

[2]National Tsing Hua University, Taiwan

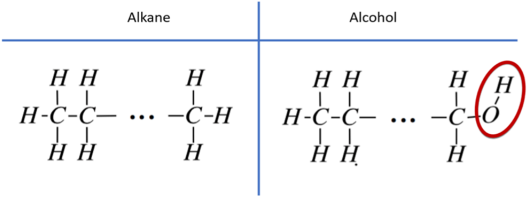
36th International Conference on Machine Learning

Goal

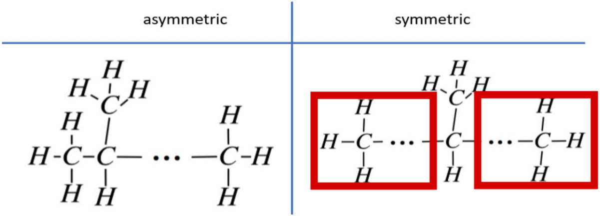
- To learn representations of graphs by generalizing convolutions
- while keeping all the nice properties of CNNs being able to
  - detect shift-invariant graph patterns by the filters
  - enlarge the receptive fields by multi-layer architecture
  - identify the critical parts (**critical structures**) most important to the jointly learned task

What are the critical structures?

- Local-Scale Critical Structures: Alkane vs Alcohol



- Global-Scale Critical Structures: Symmetric vs Asymmetric



STOA: Graph Attention Networks (Bengio et al. ICLR'18)

- The **1-head self-attentional network (1-head GAT)** is the state-of-the-art solution to the problem
- **1-head GAT** learns the attention score α_{ij} for each edge (i, j)

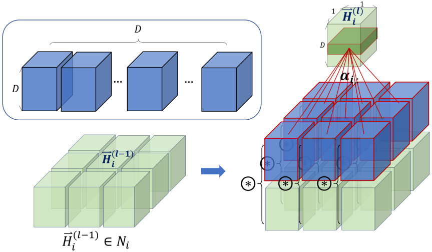
$$\overrightarrow{h_i'} = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W \overrightarrow{h_j}\right)$$

- In supervised tasks, α **are the critical structures**
  $\because \alpha_{ij}$ represents the contribution of edge (i, j) to the model prediction

Drawback: limited learning ability

- However, learning α sacrifices the learning ability

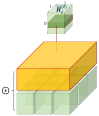
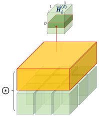
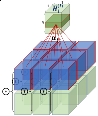
$$\overrightarrow{h'_i} = \sigma\left(\sum_{j\in\mathcal{N}_i} \alpha_{ij} W \overrightarrow{h_j}\right)$$



- It is not obvious in node classification, but severely affects the performance in graph classification

A fix: Ego-CNN

- Our idea: learning critical structures by the filters just like CNNs

Traditional Convolution		$H_{ij,d}^{(l)} = \sigma\left(E_{ij}^{(l)} \circledast W^{(l,d)} + b_d^{(l)}\right)$
Ego-Convolution (ours)		$H_{i,d}^{(l)} = \sigma\left(E_{i}^{(l)} \circledast W^{(l,d)} + b_d^{(l)}\right)$
1-head GAT (ICLR'18)		$H_{i,d}^{(l)} = \sigma\left(E_{i}^{(l)} \circledast \left(\alpha_{i,:} \otimes w^{(l,d)}\right) + b_d^{(l)}\right)$

- The filter of Ego-CNN captures the interaction of nodes in $\mathcal{N}_i$

Challenge: variable-sized $\mathcal{N}_i$ makes W ill-defined

- $H_{i,d}^{(l)} = \sigma\left(E_i^{(l)} \circledast W^{(l,d)} + b_d^{(l)}\right)$, where $E_i^{(l)} = \|_{j \in \mathcal{N}_i} H_j^{(l-1)}$
- How to define $\mathcal{N}_i$?
- In Ego-CNN, we define $\mathcal{N}_i$ as the top K nodes in the L-hop **ego**-networks at the L-th layer

Improved learning ability on graph classification

- Graph classification benchmark datasets

Table: Bioinformatic Datasets

Dataset	MUTAG	PTC	PROTEINS	NCI1
WL kernel	82.1	57.0	73.0	82.2
DGK	82.7	57.3	71.7	62.5
Subgraph2vec	87.2	60.1	73.4	80.3
MLG	84.2	63.6	**76.1**	80.8
Structure2vec	88.3			**83.7**
DCNN	67.0	56.6		62.6
Patchy-San	92.6	60.0	75.9	78.6
1-head GAT	81.0	57.0	72.5	74.3
Ego-CNN	**93.1**	**63.8**	73.8	80.7

Table: Social Network Datasets

Dataset	IMDB (B)	IMDB (M)	REDDIT (B)	COLLAB
DGK	67.0	44.6	78.0	73.0
Patchy-San	71.0	45.2	86.3	72.6
1-head GAT	70.0	-	78.8	-
Ego-CNN	**72.3**	**48.1**	**87.8**	**74.2**

- With $K = 16$, Ego-CNN is comparable to the state-of-the-arts

Ego-CNN can learng critical structure WITHOUT α

- Backtracking W with CNN visualization techniques shows the identified critical structures

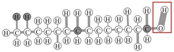
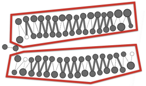
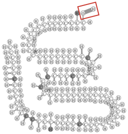
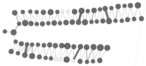
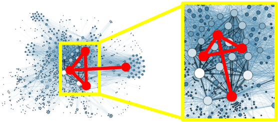
Local-Scale: Alkane vs Alcohol	Global-Scale: Symmetric vs Asymmetric
 (a) $C_{14}H_{29}OH$	 (c) Symmetric Isomer
 (b) $C_{82}H_{165}OH$	 (d) Asymmetric Isomer

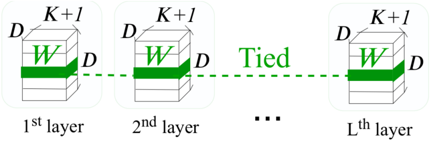
Table: Visualization of the critical structures detected by Ego-CNN

More benefits... and let's chat at Pacific Ballroom#22

- Ego-CNN can detect self-similar patterns
  - i.e., same ptterns that exist at different zoom levels
  - commonly exist in social networks



- How?
  - By simply tying the weights (W's) across different layers



- For more details, let's chat at Pacific Ballroom#22 6:30-9:00 PM