Group Meeting - November 19, 2020

Paper review & Research progress

Truong Son Hy *

*Department of Computer Science
The University of Chicago

Ryerson Physical Lab

Software Packages

For **Compositional Covariant Networks** (CCNs):

1. `https://github.com/HyTruongSon/LibCCNs`:
   - C++ core, TensorFlow/PyTorch extensions
   - Examples: node classification on citation graphs, graph classification on graph kernel benchmark, etc.
2. `https://github.com/HyTruongSon/GraphFlow`:
   - A generic Deep Learning framework built from sratch in C++/CUDA
   - Supports dynamic computation graph, symbolic automatic differentation.

For slides (today – November 19):

`http://people.cs.uchicago.edu/~hytruongson/Discussions-2020/`

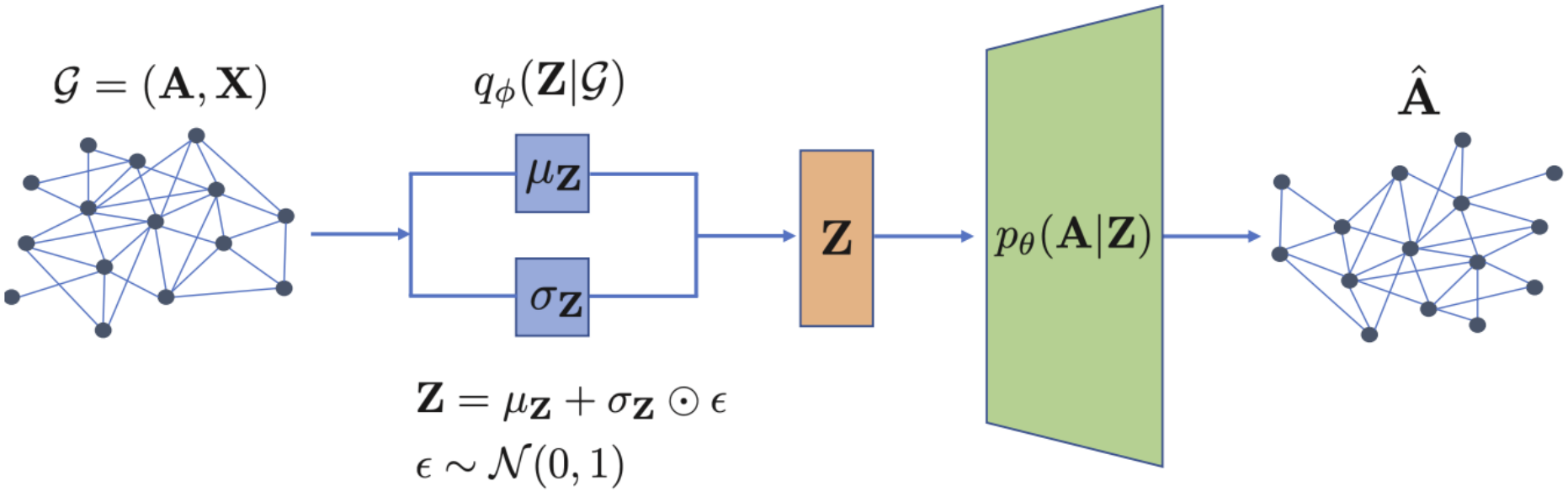
Paper

The general theory of permutation equivarant neural networks and higher order graph variational encoders

Erik Henning Thiede, Truong Son Hy, Risi Kondor

`https://arxiv.org/pdf/2004.03990.pdf`

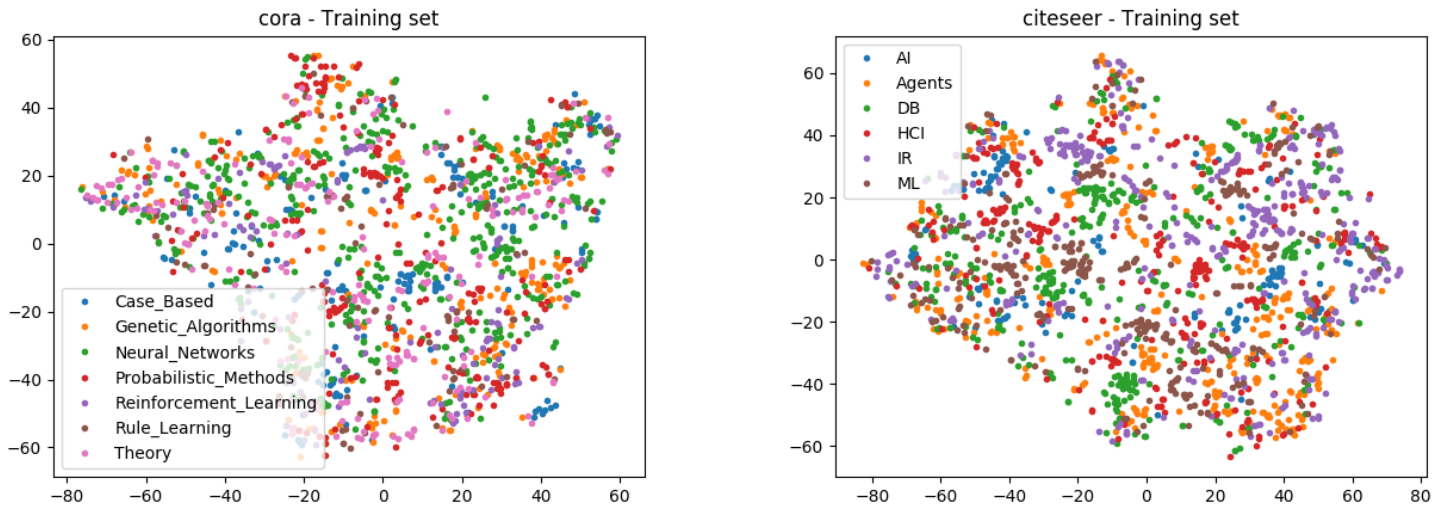
Variational Autoencoder Approaches



Graph Representation Learning, William L. Hamilton (McGill University, 2020) `https://www.cs.mcgill.ca/~wlh/grl_book/`

Citation graphs

- Each node/vertex is a paper. The node feature is the word frequency.
- An directed edge happens between papers A and B if paper A cites paper B.
- **Tasks:** (supervised) node classification, (unsupervised) link prediction.



t-SNE visualization on the learned node/vertex representation by CCNs on cora and citeceer

Link prediction

Remove 20% of edges out of the graph, keep 80% rest:

- **Training:** Apply the graph (variational) autoencoder to reconstruct the **partial** adjacency matrix of 80% of edges.
- **Testing:** Take the latent for each node/vertex z_i. Predict the missed edge $\sigma(z_i^T z_j)$.

Method	AUC	AP
SC	84.6 ± 0.01	88.5 ± 0.00
DW	83.1 ± 0.01	85.0 ± 0.00
GAE	91.0 ± 0.02	92.0 ± 0.03
VGAE	91.4 ± 0.01	92.6 ± 0.01
GraphStar	$95.65 \pm ?$	$96.15 \pm ?$
2nd order VGAE (our method)	$\mathbf{96.1} \pm 0.07$	$\mathbf{96.4} \pm 0.06$

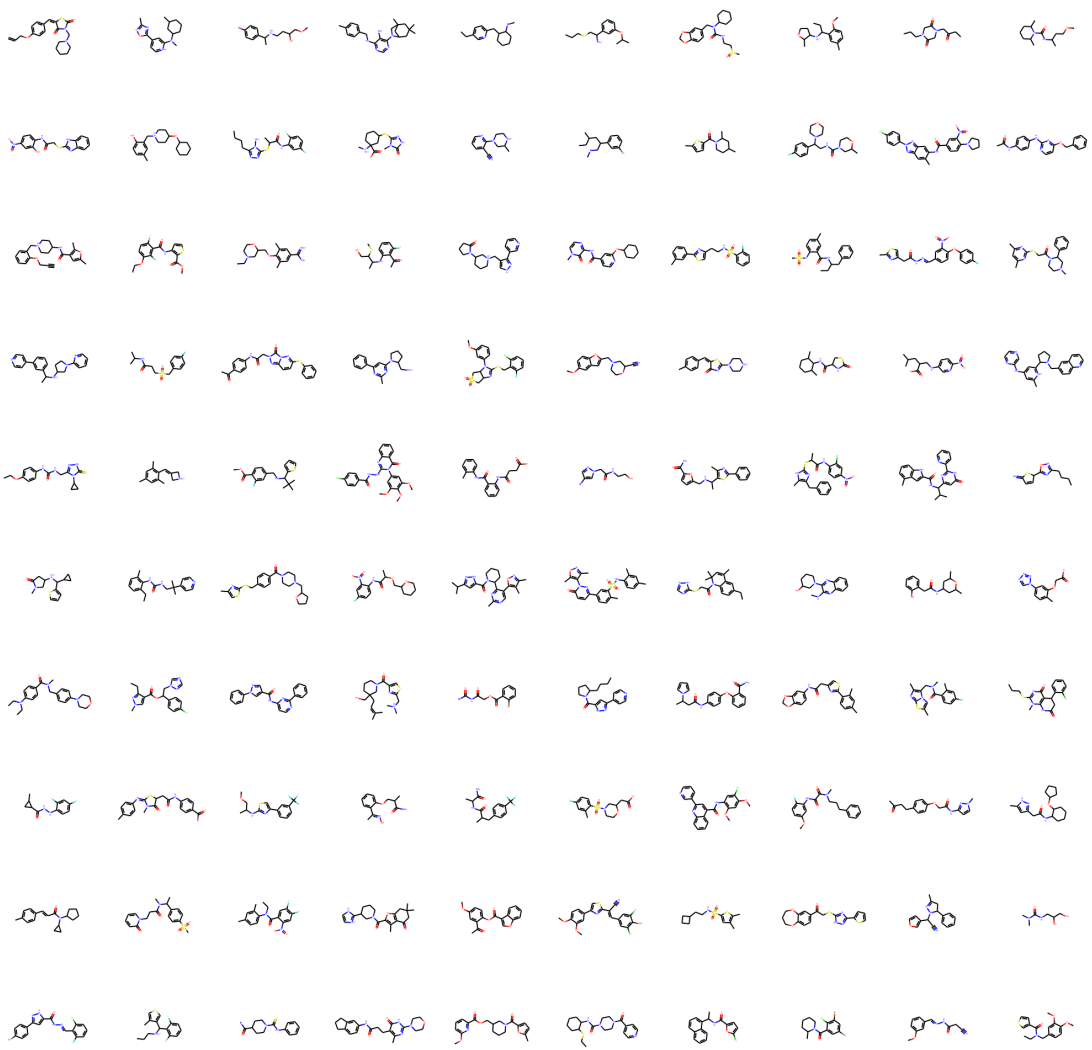
Table 1: Cora link prediction results (AUC & AP)

Method	AUC	AP
SC	80.5 ± 0.01	85.0 ± 0.01
DW	80.5 ± 0.02	83.6 ± 0.01
GAE	89.5 ± 0.04	89.9 ± 0.05
VGAE	90.8 ± 0.02	92.0 ± 0.02
GraphStar	$\mathbf{97.47} \pm ?$	$\mathbf{97.93} \pm ?$
2nd order VGAE (our method)	95.3 ± 0.02	94.3 ± 0.02

Table 2: Citeseer link prediction results (AUC & AP)

SC = Spectral Clustering, DW = Deep Walk, GAE = Graph Autoencoder
VGAE = Variational Graph Autoencoder

Molecular generation (1)



100 samples from ZINC dataset

Molecular generation (2)

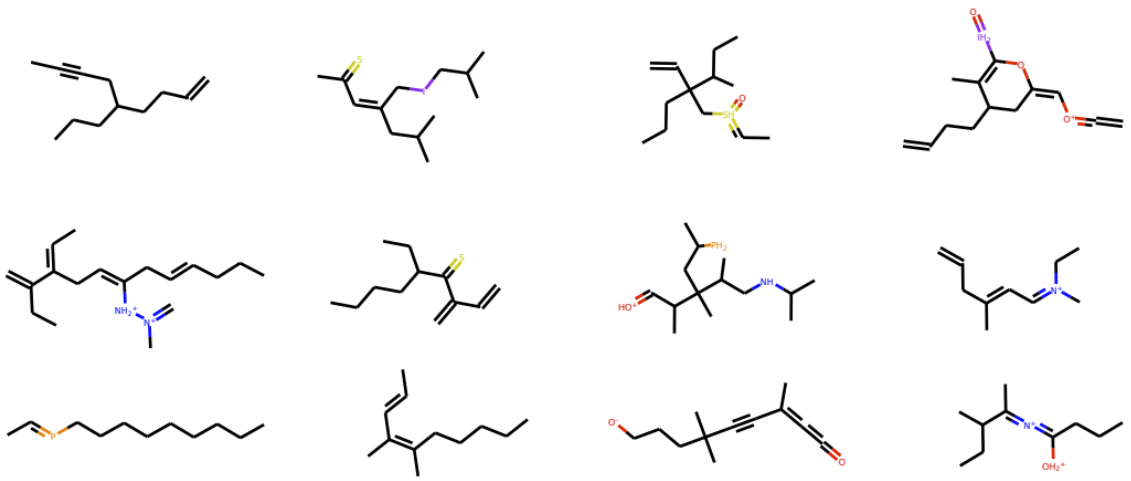
- **Training:** Train the graph VAE to reconstruct the adjacency and the atomic features.
- **Generation:** Sample from the prior and use the trained decoder to decode that latent.

Method	Accuracy	Validity
CVAE	44.6%	0.7%
GVAE	53.7%	7.2%
SD-VAE	76.2%	43.5%
GraphVAE	-	13.5%
Atom-by-Atom LSTM	-	89.2%
JT-VAE	76.7 %	**100.0%**
2nd order $\mathbb{S}_n$-Conv VAE (our method)	**98.4%**	33.4%

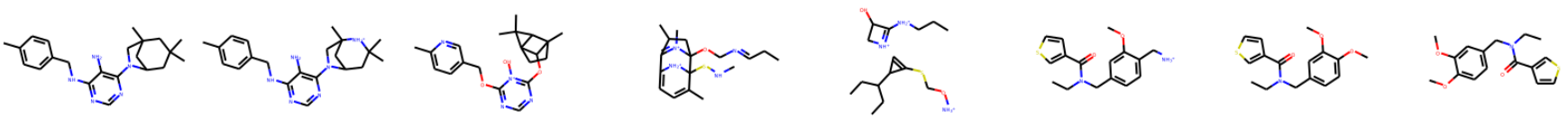
Reconstruction accuracy & Validity on ZINC dataset

Molecular generation (3)

Generated examples (model trained on ZINC):

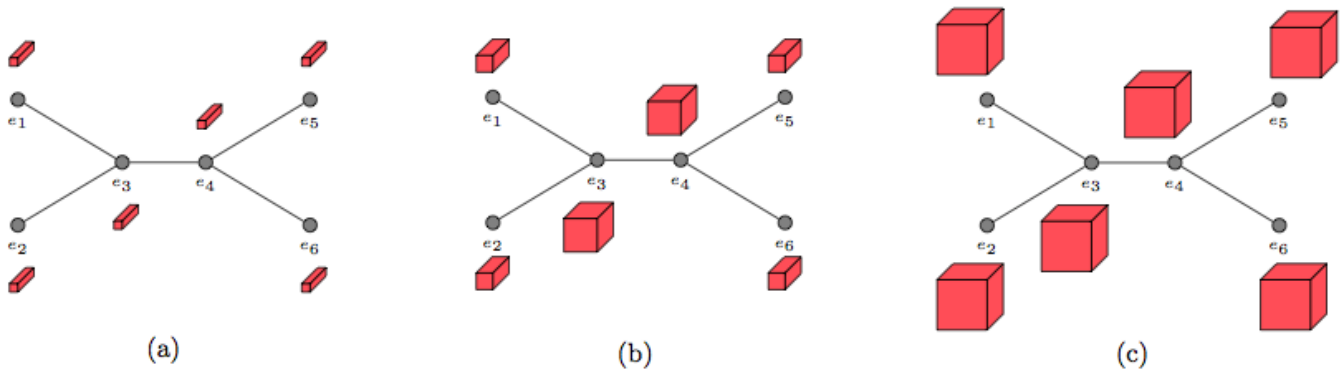


Interpolation on the latent:



Brainstorm (1)

Problem: How input vertex/atom featurization affects the adjacency reconstruction. Let consider the Ethylene (C_2H_4) molecular graph for example example.



- The input atom features of 4 hydrogen atoms are **identical** (the same). The input atom features of 2 carbon atoms are also the same.
- There is an edge between e1 (hydrogen) and e3 (carbon). But there is no edge between e5 (hydrogen) and e3 (carbon). Therefore, there is **NO** function can classify edge/non-edge for pairs (e1, e3) and (e5, e3).

Brainstorm (2)

Other existing problems:

1. The model can generate something-similar-to molecules.
2. Generated examples are still **not** chemical useful.
3. Some sub-structures don't exist in Nature. Why?
4. Computationally expensive.

Potential future direction:

1. Goal-oriented molecular generation to optimize a specific molecular property.
2. Drug discovery datasets? Self-supervised learning?
3. **Looking forward for your suggestions! Thanks.**