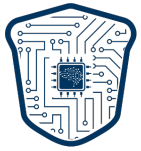
ICLR
International Conference On Learning Representations



TMLR
TRUSTWORTHY MACHINE LEARNING AND REASONING

Tencent AI Lab

# Neural Atoms: Propagating Long-Range Interaction In Molecular Graphs Through Efficient Communication Channel


Xuan Li
Hong Kong Baptist University

*with Zhanke Zhou, Jiangchao Yao, Yu Rong, Lu Zhang, Bo Han*


paper

code

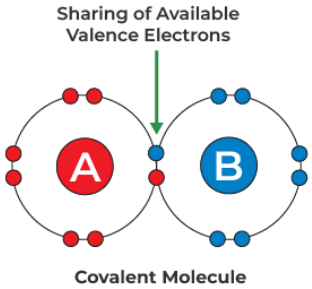
[ICLR'24 Paper]: https://openreview.net/pdf?id=CUfSCwcgqm [Code]: https://github.com/tmlr-group/NeuralAtom

# Outlines

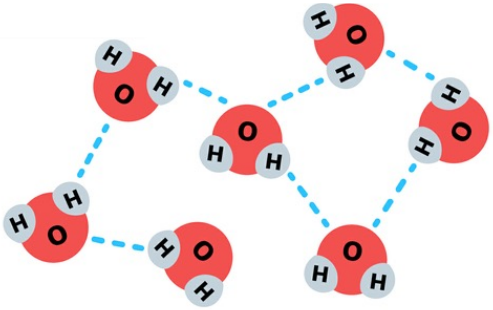
- Background
- Existed works
- Method
- Experiments
- Summary

# Background | atom interaction

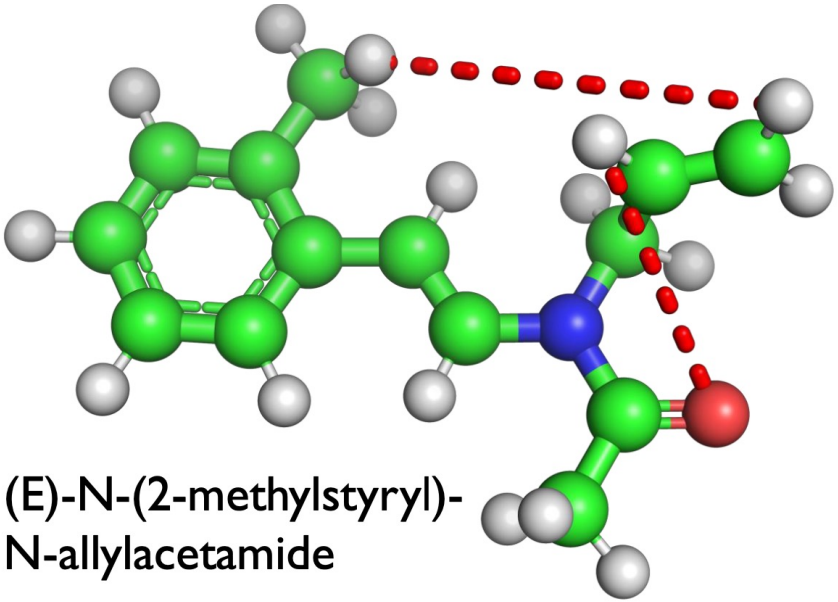
Molecular graphs consist of different types of atom interaction.




Covalent-bond



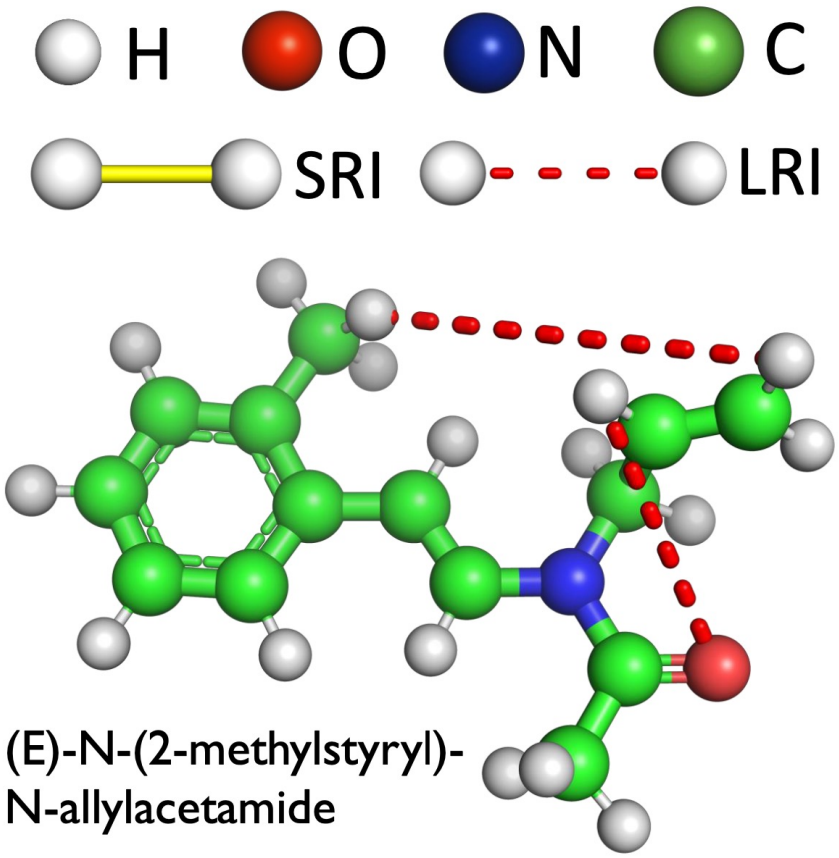
Hydrogen bond



(E)-N-(2-methylstyryl)-
N-allylacetamide

https://www.geeksforgeeks.org/covalent-bond/
https://theory.labster.com/hydrogen_bond/

# Background | SRI and LRI




The short-range interaction (**SRI**) forms the structure of the molecular graph.

The long-range interaction (**LRI**) could determine both the physical and chemical properties.



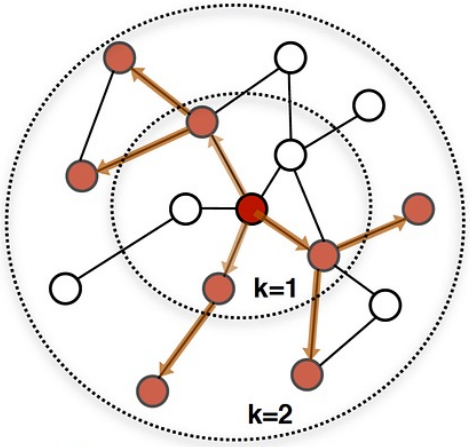
Treatment of electrostatic effects in macromolecular modeling. In Proteins: Structure, Function, and Bioinformatics, 1989.

# Outlines

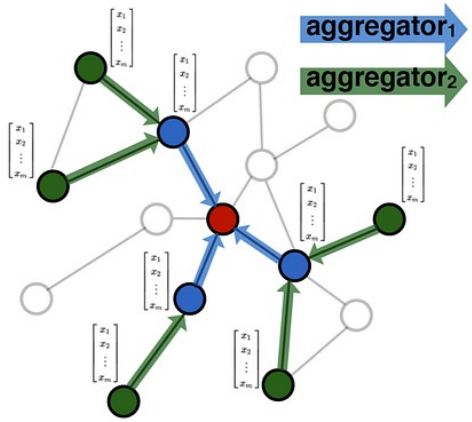
- Background
- Existed works
- Method
- Experiments
- Summary

# Existed works | Graph Neural Networks (GNNs)

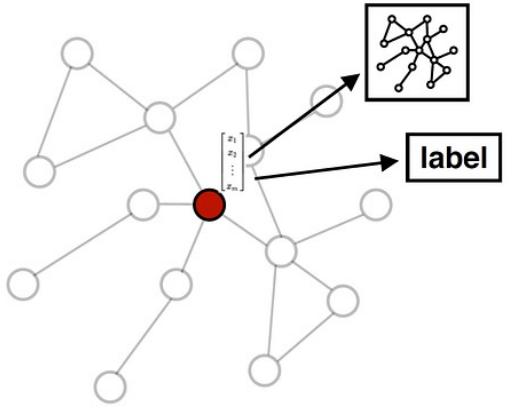
GNNs learn information by performing neighbor aggregation, each layer corresponds to one additional hop neighbors.




1. Sample neighborhood




2. Aggregate feature information from neighbors



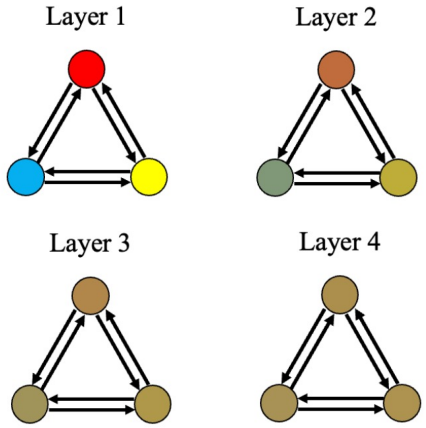

3. Predict graph context and label using aggregated information

To **capture** LRI, we need to **stack multiple GNN layers** for aggregating neighbor information.



https://snap.stanford.edu/graphsage/

# Existed works | Graph Neural Networks (GNNs)

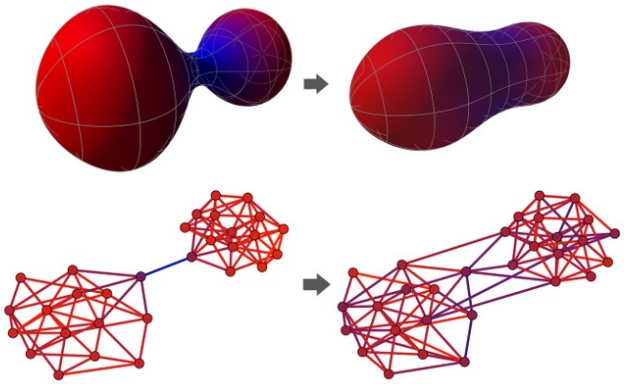



## Over-smoothing

Interacting nodes converge to indistinguishable representations as the number of GNN layers increases.

## Over-squashing

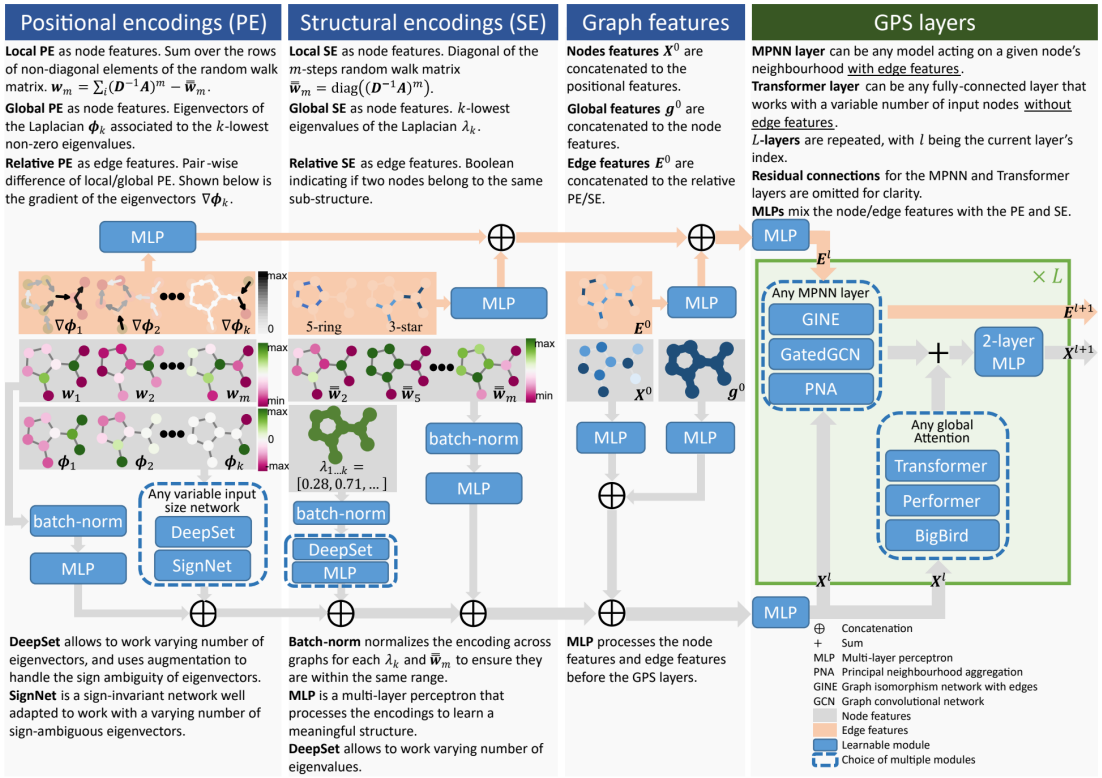
Over-squashing occurs when an exponentially growing amount of information is squashed into a fixed-size vector.



Understanding over-squashing and bottlenecks on graphs via curvature. In ICLR, 2022.
https://minyoungg.github.io/MIT-deeplearning-blogs/2021/12/09/oversquashing-in-gnns/

# Existed works | graph transformer

Node pair-wise attention for transformer block.



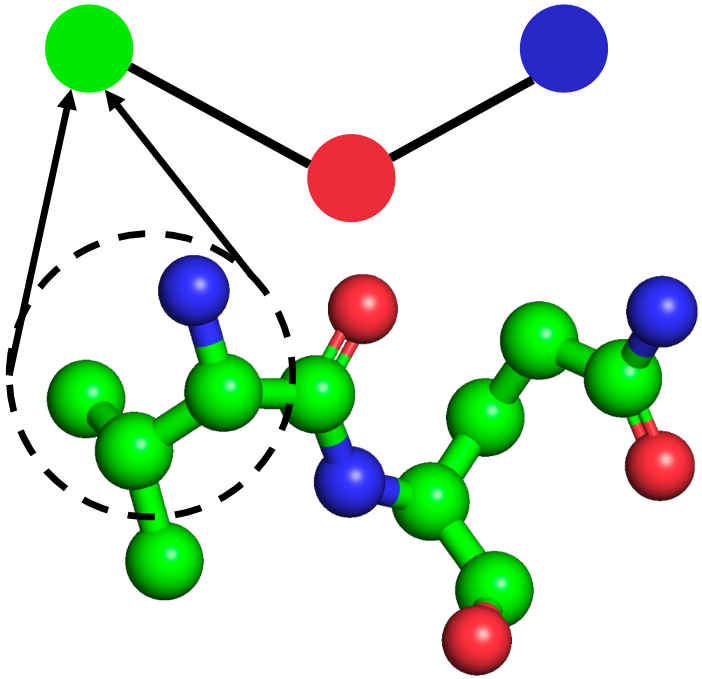

- **Irrelevant interactions**
  With self-attention, a node may attend to many nodes with no direct edge connection.
- **Additional computation**
  LRIs are usually sparse, and the node pair-wise attention might not be necessary.

Recipe for a General, Powerful, Scalable Graph Transformer. In NeurIPS, 2022.

# Outlines

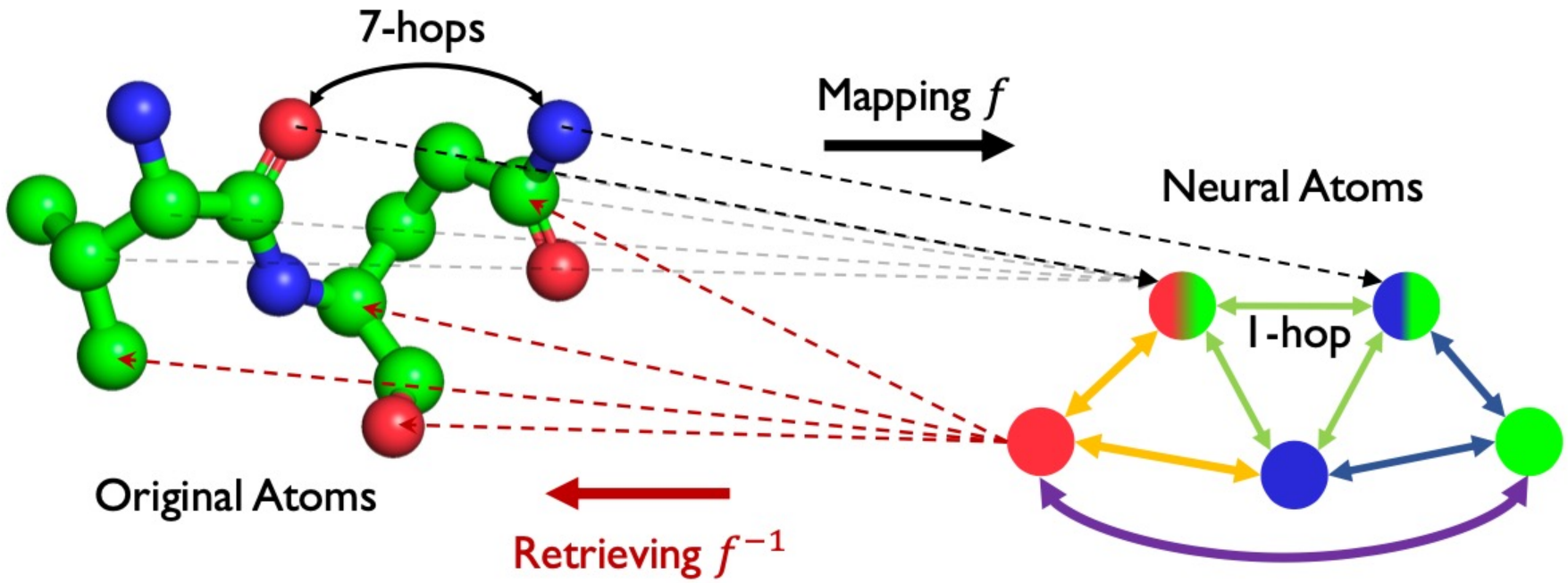
- Background
- Existed works
- Method
- Experiments
- Summary

# Change the interaction space.



Can we capture the LRI in a more manageable and computationally efficient space?

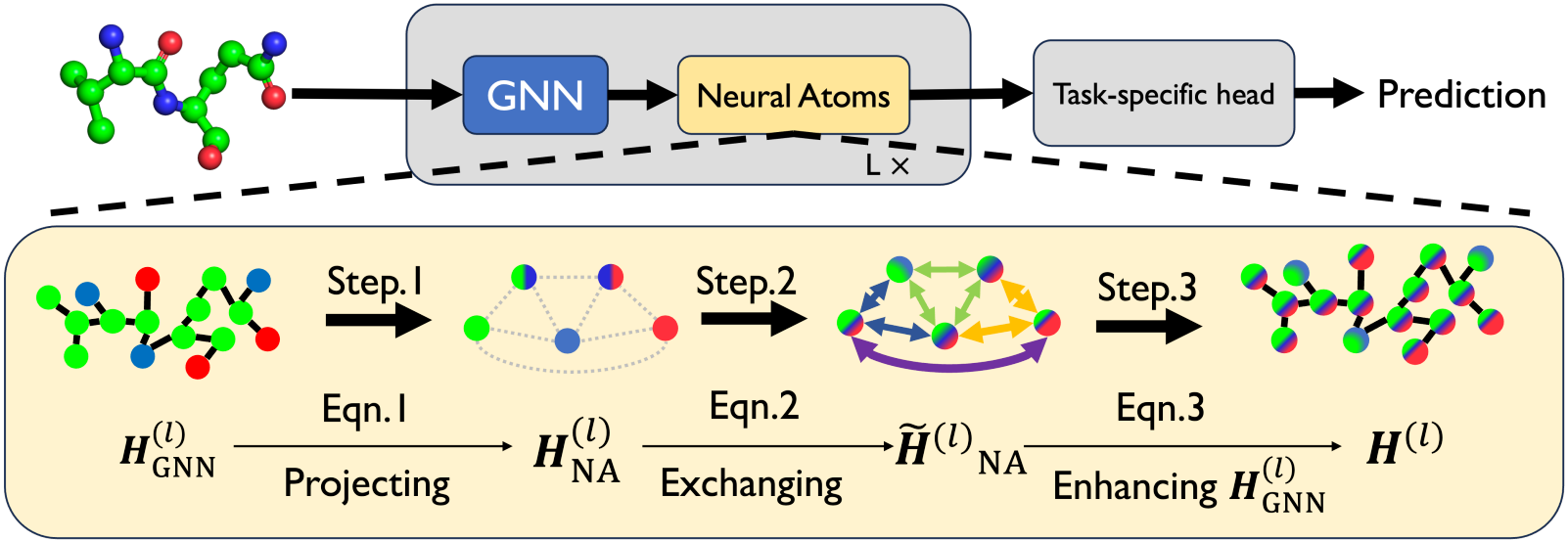
# General idea




## Aim

Learning to **project** all the original atoms into a few **neural atoms** that **abstract** the collective information of atomic groups in a molecule.

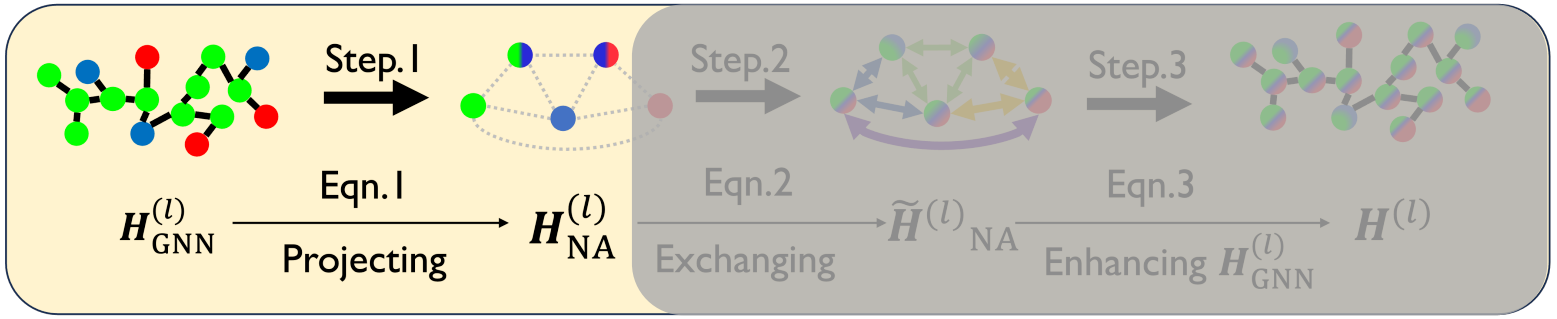
# Overview




The neural atom has several advantages:

- **Learnable projection** from atoms to neural atoms.
- **Reducing** the multi-hop long-range interaction to single-hop.
- GNN-agnostic and **plug-in-and-play.**

# Neural Atoms – step 1

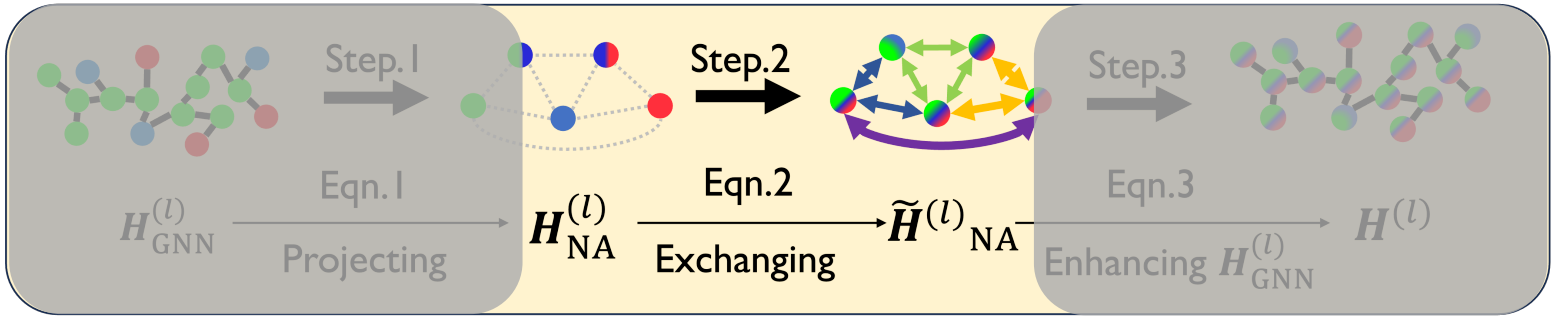



Step-1. Project atom representations $\boldsymbol{H}_{\boldsymbol{GNN}}^{(\ell)}$ to neural atom representations $\boldsymbol{H}_{\mathbf{NA}}^{(\ell)}$.

$$\boldsymbol{H}_{\mathrm{NA}}^{(\ell)} = \mathrm{LayerNorm}\left(\boldsymbol{Q}_{\mathrm{NA}}^{(\ell)} \oplus \mathrm{MultiHead}(\boldsymbol{Q}_{\mathrm{NA}}^{(\ell)}, \boldsymbol{H}_{\mathrm{GNN}}^{(\ell)}, \boldsymbol{H}_{\mathrm{GNN}}^{(\ell)})\right)$$

Projecting N atoms into **K (<< N)** neural atoms by multi-head attention.

# Neural Atoms – step 2

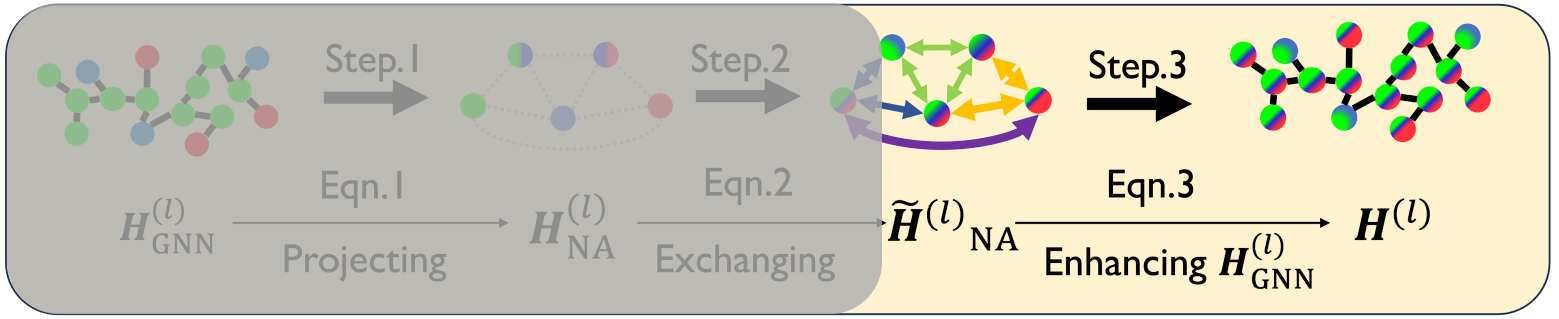



Step-1. Step-2. Exchange information among neural atoms $\boldsymbol{H}^{(\ell)}_{\mathbf{NA}} \mapsto \tilde{\boldsymbol{H}}^{(\ell)}_{\mathbf{NA}}$.

$$\tilde{\boldsymbol{H}}^{(\ell)}_{\text{NA}} = \text{LayerNorm}\left(\boldsymbol{H}^{(\ell)}_{\text{NA}} \oplus \text{MultiHead}(\boldsymbol{H}^{(\ell)}_{\text{NA}}, \boldsymbol{H}^{(\ell)}_{\text{NA}}, \boldsymbol{H}^{(\ell)}_{\text{NA}})\right)$$

Exchanging information among neural atoms

# Neural Atoms – step 3




Step-3. Project neural atoms back and enhance the atoms' representation $(\boldsymbol{H}^{(\ell)}_{\boldsymbol{GNN}}, \tilde{\boldsymbol{H}}^{(\ell)}_{\mathbf{NA}}) \mapsto \boldsymbol{H}^{(\ell)}$.

$$\boldsymbol{H}^{(\ell)} = \boldsymbol{H}^{(\ell)}_{\text{GNN}} \oplus \tilde{\boldsymbol{A}}^{(\ell)}_{\text{NA}} \tilde{\boldsymbol{H}}^{(\ell)}_{\text{NA}}, \quad \text{s.t.} \quad \tilde{\boldsymbol{A}}^{(\ell)}_{\text{NA}} = \text{Aggregate}\left(\{\hat{\boldsymbol{A}}_m\}_{m=1}^{M}\right)^{\top} \in \mathbb{R}^{N \times K}$$

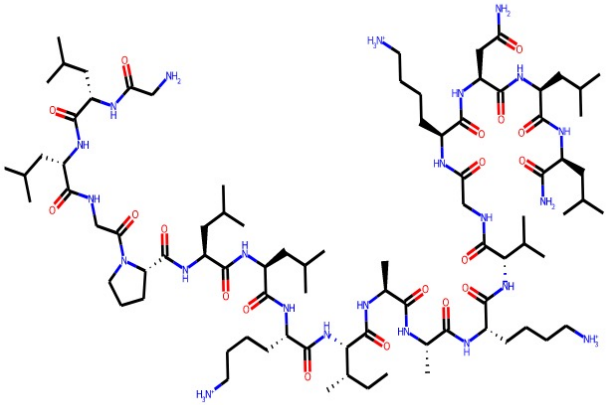
Enhancing the atom representation $\boldsymbol{H}^{(\ell)}_{\boldsymbol{GNN}}$ by $\tilde{\boldsymbol{H}}^{(\ell)}_{\mathbf{NA}}$

# Outlines

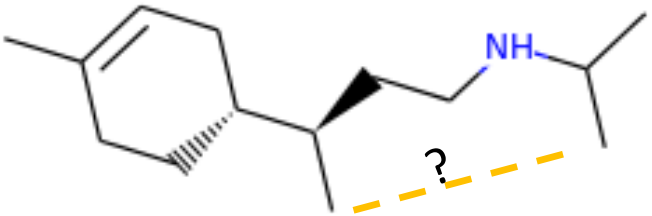
- Background
- Existed works
- Method
- Experiments
  - 2D & 3D
- Summary

# 2D setups

| Dataset | Total Graphs | Total Nodes | Avg Nodes | Mean Deg. | Total Edges | Avg Edges | Avg Short.Path. | Avg Diameter |
|---|---|---|---|---|---|---|---|---|
| pcqm-contact | 529,434 | 15,955,687 | 30.14 | 2.03 | 32,341,644 | 61.09 | 4.63±0.63 | 9.86±1.79 |
| pepfunc | 15,535 | 2,344,859 | 150.94 | 2.04 | 4,773,974 | 307.30 | 20.89±9.79 | 56.99±28.72 |
| pepstruct | 15,535 | 2,344,859 | 150.94 | 2.04 | 4,773,974 | 307.30 | 20.89±9.79 | 56.99±28.72 |



**Graph prediction** and **regression** on peptides-func and peptides-struct.




**Link prediction** on PCQM-Contact.

# 2D experiments

Neural atoms can **boost** the performance of various GNNs up to **27.32%**.

Table 1: Test performance on three LRGB datasets. Shown is the mean ± s.d. of 4 runs.

| Model | Peptides-func | Peptides-struct | PCQM-Contact |
|---|---|---|---|
| | AP ↑ | MAE ↓ | MRR ↑ |
| Transformer+LapPE | 0.6326 ± 0.0126 | 0.2529 ± 0.0016 | 0.3174 ± 0.0020 |
| SAN+LapPE | 0.6384 ± 0.0121 | 0.2683 ± 0.0043 | 0.3350 ± 0.0003 |
| GraphGPS | 0.6535 ± 0.0041 | 0.2500 ± 0.0005 | 0.3337 ± 0.0006 |
| GCN | 0.5930 ± 0.0023 | 0.3496 ± 0.0013 | 0.2329 ± 0.0009 |
| **+ Neural Atoms** | **0.6220 ± 0.0046** | **0.2606 ± 0.0027** | **0.2534 ± 0.0200** |
| GINE | 0.5498 ± 0.0079 | 0.3547 ± 0.0045 | 0.3180 ± 0.0027 |
| **+ Neural Atoms** | **0.6154 ± 0.0157** | **0.2553 ± 0.0005** | **0.3126 ± 0.0021** |
| GCNII | 0.5543 ± 0.0078 | 0.3471 ± 0.0010 | 0.3161 ± 0.0004 |
| **+ Neural Atoms** | **0.5996 ± 0.0033** | **0.2563 ± 0.0020** | **0.3049 ± 0.0006** |
| GatedGCN | 0.5864 ± 0.0077 | 0.3420 ± 0.0013 | 0.3218 ± 0.0011 |
| **+ Neural Atoms** | **0.6562 ± 0.0075** | **0.2585 ± 0.0017** | **0.3258 ± 0.0003** |
| GatedGCN+RWSE | 0.6069 ± 0.0035 | 0.3357 ± 0.0006 | 0.3242 ± 0.0008 |
| **+ Neural Atoms** | **0.6591 ± 0.0050** | **0.2568 ± 0.0005** | **0.3262 ± 0.0010** |

# 2D running time

Neural atoms bring acceptable additional computation.

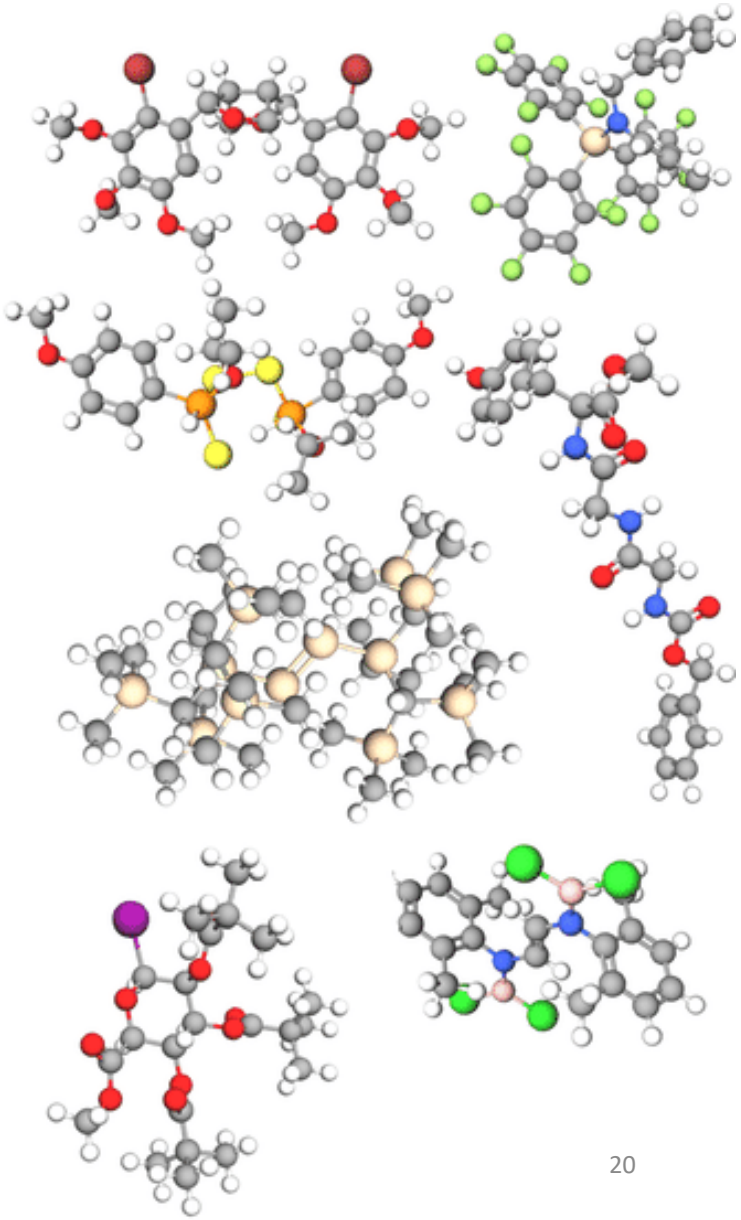
Table 14: Wall-clock run times. Average epoch time (average of 5 epochs, including validation performance evaluation) is shown for each model and dataset combination.

| avg. time / epoch | **Peptides-func** | **Peptides-struct** | **PCQM-Contact** |
|---|---|---|---|
| GCN | 2.6s | 2.5s | 56.9s |
| + Neural Atom | 5.5s | 4.9s | 65.1s |
| GINE | 2.6s | 2.6s | 56.7s |
| + Neural Atom | 4.8s | 4.2s | 66.8s |
| GCNII | 2.5s | 2.3s | 56.9s |
| + Neural Atom | 4.7s | 5.1s | 59.4s |
| GatedGCN | 3.3s | 3.2s | 56.5s |
| + Neural Atom | 6.1s | 5.5s | 61.6s |
| GatedGCN+RWSE | 3.4s | 4.1s | 59.4s |
| + Neural Atom | 6.4s | 5.2s | 65.0s |
| Transformer+LapPE | 6.4s | 6.2s | 59.2s |
| SAN+LapPE | 60s | 57.5s | 205s |
| GraphGPS | 6.5s | 6.5s | 61.5s |

# 3D setups

OE62 contains **61,489** organic molecular graphs, each consisting of up to **174** atoms with **16** different elements (H, Li, B, C …)

Molecular energy calculation (in eV) measured by **Energy MAE** and **Energy MSE** compares to DFT-computed energies.



# 3D experiments

**10** neural atoms **without 3D information** and **half the hidden dimension** achieve competitive performance compared to the SOTA, Ewald-based approach.

Table 2: Validation energy MAE and MSE comparison on OE62 dataset.

| | Energy MAE ↓ | Energy MSE ↓ | Number of Params. |
|---|---|---|---|
| SchNet (Schütt et al., 2017) | 0.1351 | 0.0658 | 2.75 M |
| + Ewald Block | **0.0811** | **0.0301** | 12.21 M |
| **+ Neural Atoms** | **0.0834** | **0.0309** | 2.63 M |
| PaiNN (Schütt et al., 2021) | 0.6049 | 0.0133 | 12.52 M |
| + Ewald Block | **0.0590** | **0.0134** | 15.68 M |
| **+ Neural Atoms** | **0.0558** | **0.0122** | 6.05 M |
| DimeNet++ (Gasteiger et al., 2020) | 0.0501 | 0.0117 | 2.76 M |
| + Ewald Block | **0.0479** | **0.0107** | 4.75 M |
| **+ Neural Atoms** | 0.0551 | 0.0129 | 1.97 M |

# Outlines

- Background
- Existed works
- Method
- Experiments
- Summary

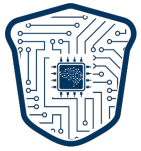
## Conclusion

We propose neural atoms to enhance GNNs capturing LRIs. Our method boost GNNs by transforming original atoms into neural atoms, facilitating information exchange, and then projecting the improved information back to atomic representations.

## Future directions

- Leveraging the atomic coordinate information better to capture the LRI.
- Instilling expert knowledge in the grouping strategy.

ICLR
International Conference On Learning Representations



TMLR
TRUSTWORTHY MACHINE LEARNING AND REASONING



Tencent AI Lab

# THANKS

Email: csxuanli@comp.hkbu.edu.hk

paper

code

[ICLR'24 Paper]: https://openreview.net/pdf?id=CUfSCwcgqm [Code]: https://github.com/tmlr-group/NeuralAtom