Prediction of Interaction Energies of Molecules via Graph Neural Networks

Computational photonic: Advanced Seminar

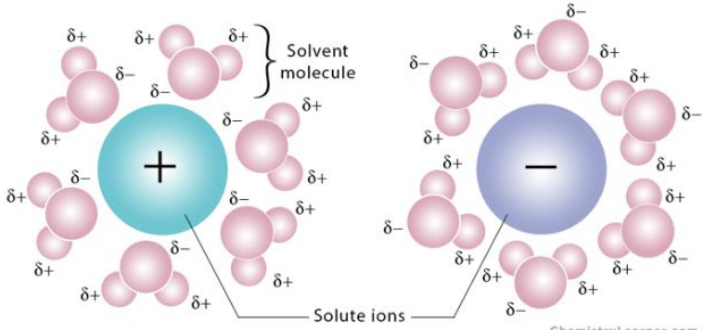
By-Utkarsh Saraswat
MSc Computational Science and Engineering
Technical University of Munich

Table of content

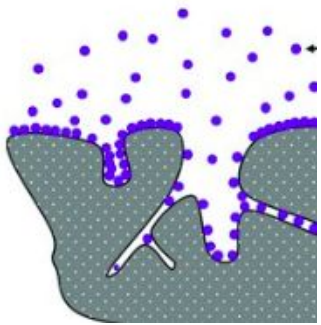
Motivation and Overview

Interaction energy

Solvation energy [1]



Adsorption energy [2]



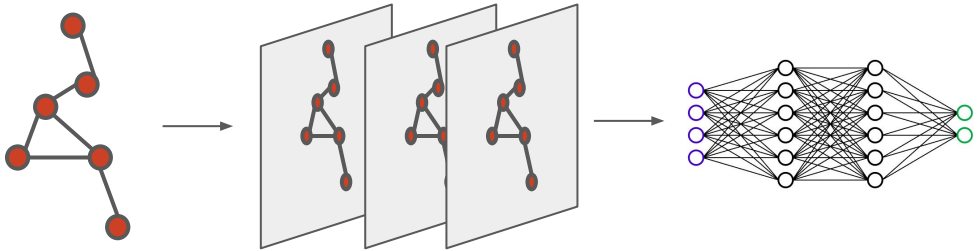
Comparison of different GNN approaches to predict

- Solvation energy
- Adsorption energy

Coming up with an **accurate**, **extensible** and **interpretable** GNN model for **nanoparticles**

1. Chemistrylearner.com
2. Thermochemical Sorption Heat Storage: A Review

Brief introduction to Graph Neural Networks (GNN)



1

Fig 1. GNN architecture

How molecules are transformed into graphs

- Atoms converted to nodes
- Bond converted to edges
- In some cases, molecules to nodes
- *Multiple features* in nodes and edges

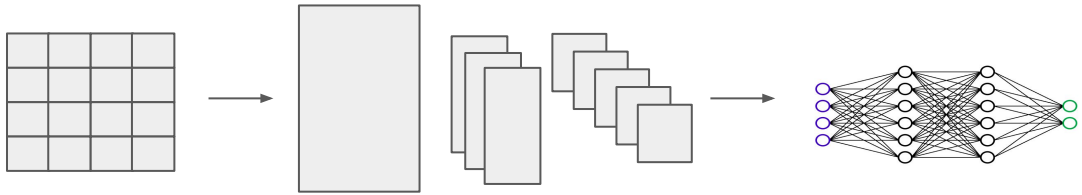
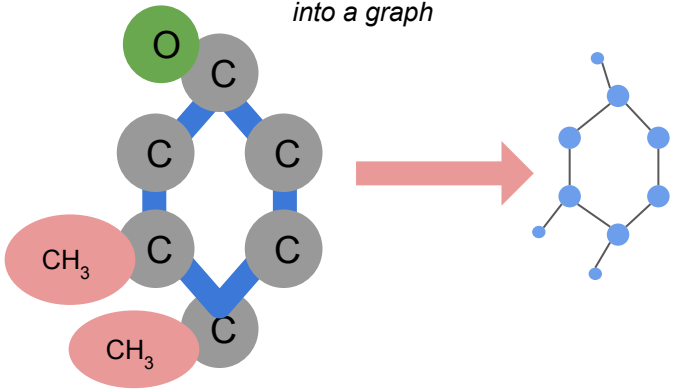


Fig 2. CNN architecture

Fig 3. Transformation of molecule into a graph



1. Victor Zhou: Neural network from scratch

Prediction of solvation energy and molecular properties in solutions

Study 1)*

- Features : Properties of Covalent bond
- GNN architecture: **Graph attention network (GAN)** and **Message Passing Neural Network (MPNN)**
- Pairwise inclusion of solvent and solute features

- RMSE of 0.13
- higher MAE in solvents with less presence in the training data

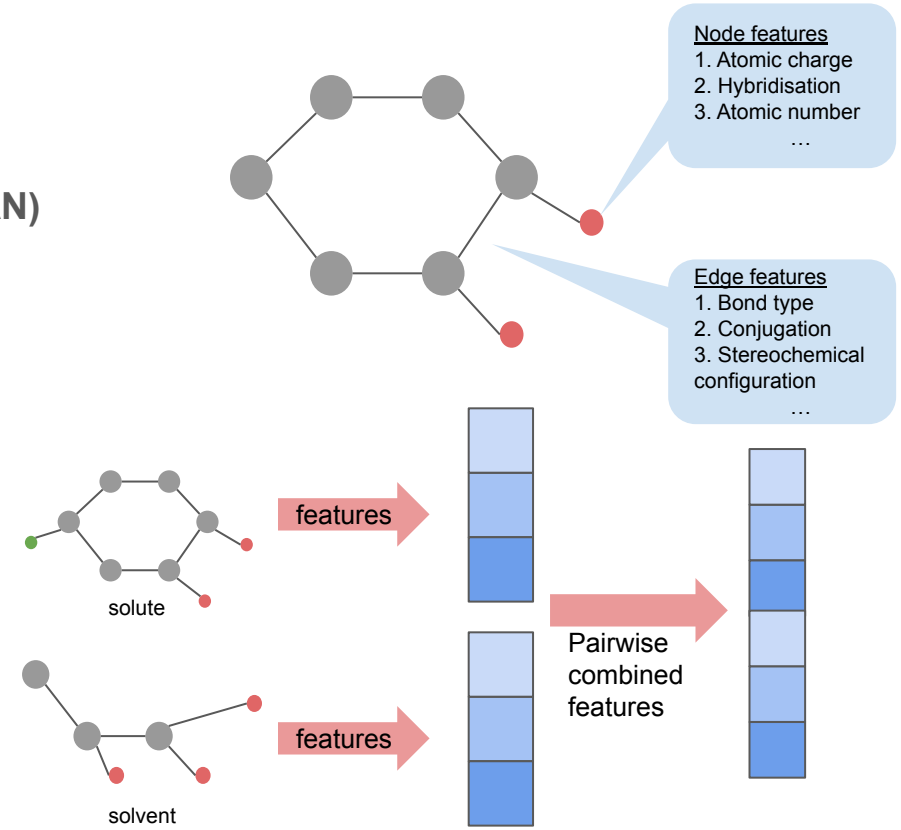


Fig 4. Node and edge features used in GNN in the study 1

** Ramin Ansari, Amirata Ghorbani "Accurate Prediction of Free Solvation Energy of Organic Molecules via Graph Attention Network and Message Passing Neural Network from Pairwise Atomistic Interactions"*

Prediction of solvation energy and molecular properties in solutions

Study 2)*

- Features: radical electrons, donor-acceptor electrons, acidic or basic nature
- GNN architecture: MPNN
- Intermolecular attraction : linear interaction phase

- result yielded An RMSE of 0.73 ± 0.01
- RMSE 0f 1.03 for test data (Not used in training)

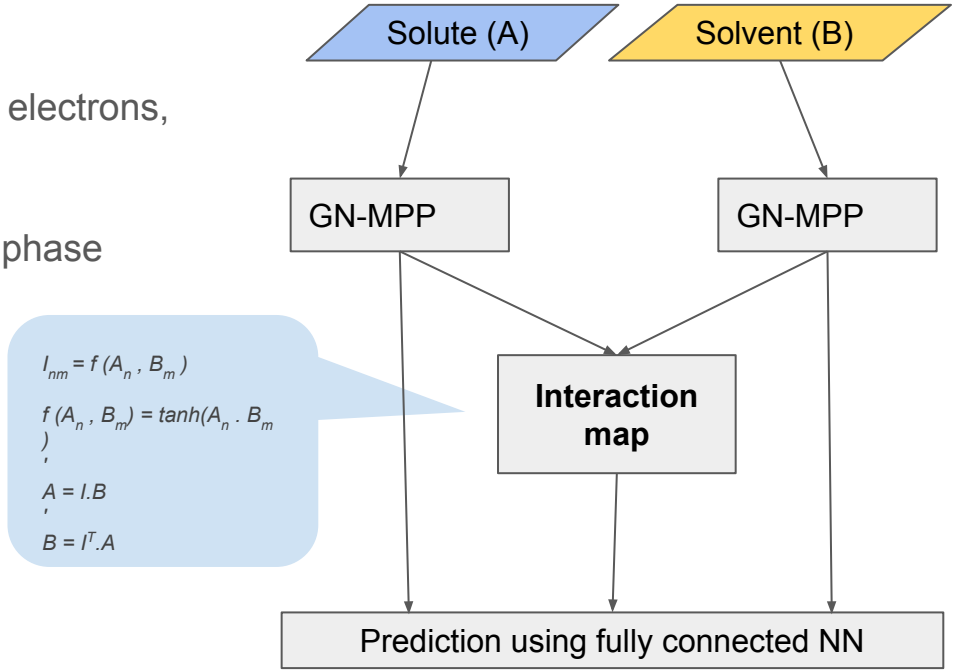


Fig 5. GNN architecture with solute-solvent interaction in the study 2

** Yashaswi Pathak, Siddhartha Laghuvarapu et el, "Chemically Interpretable Graph Interaction Network for Prediction of Pharmacokinetic Properties of Drug-Like Molecules"*

Prediction of solvation energy and molecular properties in solutions

Study 3)*

- Large database with **200,000 binary mixtures** and **160,000 ternary mixtures**
- Three GNN architectures
  - SolvCAT: intramolecular graph only
  - SolvGCN: intramolecular graph combined with inter-molecular graph
  - **SolvGNN: edge features global interaction were specified as a Hydrogen bonds**

- RMSE of 0.0032 (in *ln(activity coefficient)*)
- SolveGNN: inclusion of H-bond in graph leads to high accuracy

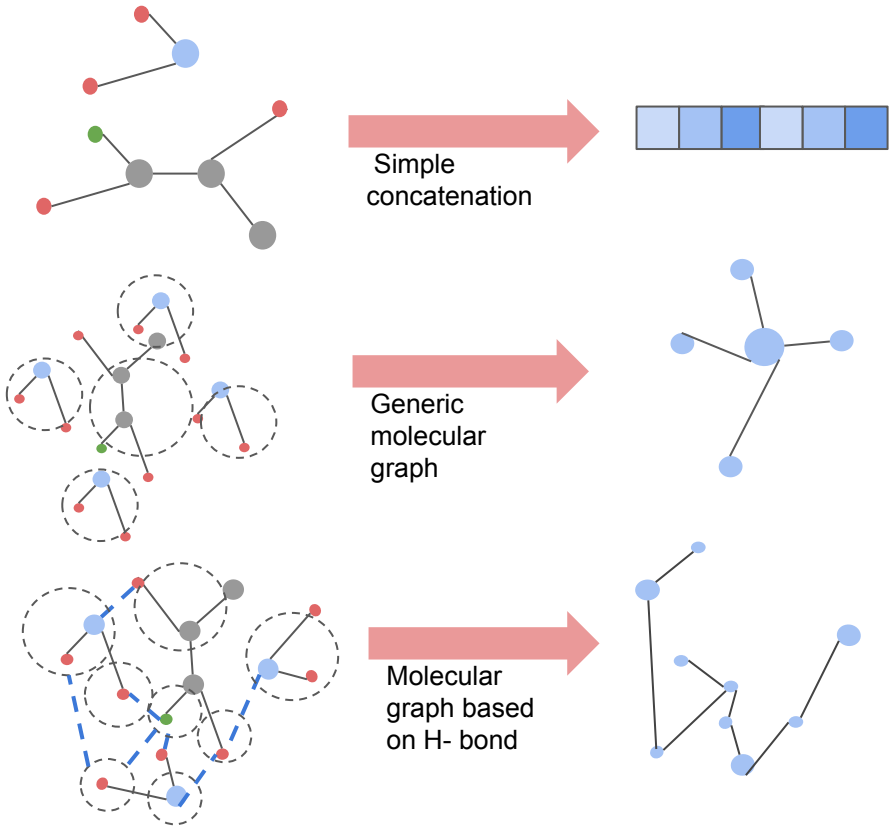


Fig 6. Different GNN architectures implemented in study 3

** Shiyi Qin Shengli Jiang, et el, "Capturing molecular interactions in graph neural networks: a case study in multi-component phase equilibrium"*

Prediction of interaction energies of multi-molecular systems

Study 4)*

- Features : **No molecular properties of CO_2 used**
- Multiple edges between the same pair of nodes
- Edge convolution was allowed along with node features
- GNN architecture: GCNN with Attention

- RMSE of 0.256 on test

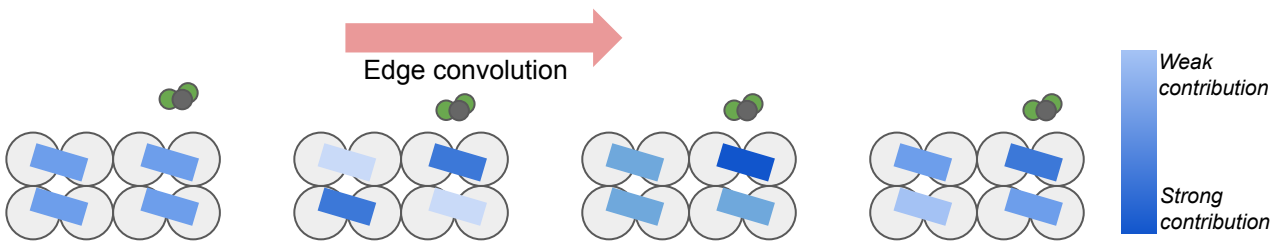


Fig 8. Transformation of edge features via edge convolution can help to filter out strong contributors to interaction

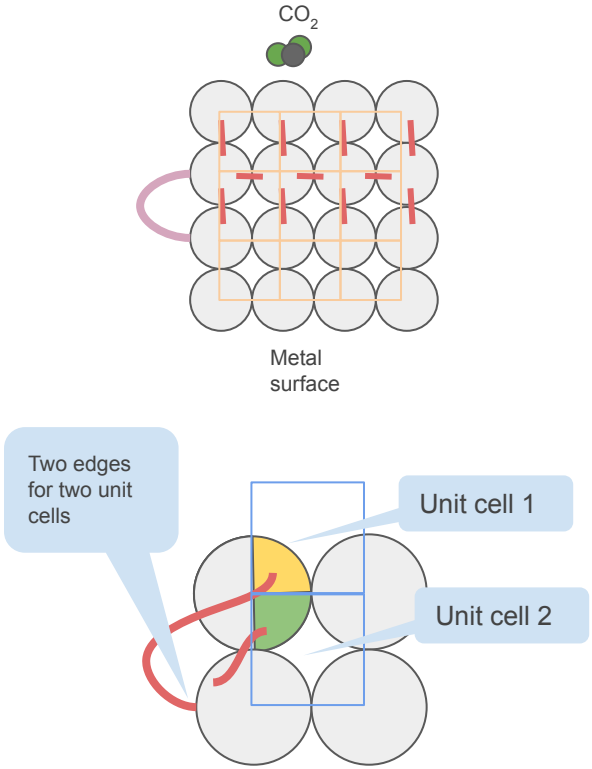


Fig 7. Multiple edges within same pair of atoms in study

* *Guojing Cong, Anshul Gupta et el, "Prediction of CO_2 Adsorption in Nano-Pores with Graph Neural Networks"*

Prediction of interaction energies of multi-molecular systems

Study 5)*

- Edge features : “type of bond” (metal-organic, organic-organic, or metal-metal)
- GNN architecture **GAME-Net** consisted of i) fully connected layers, ii) convolutional layers, and iii)pooling layer
- **Combined graph generated from adsorbent and adsorbate**
- **Energy of remaining system subtracted**

- RMSE 0f 0.18 for test data
- More error in aromatic compounds but still much accurate
- Model could be applied to larger molecules
- Can be used to predict IE of industrial molecules

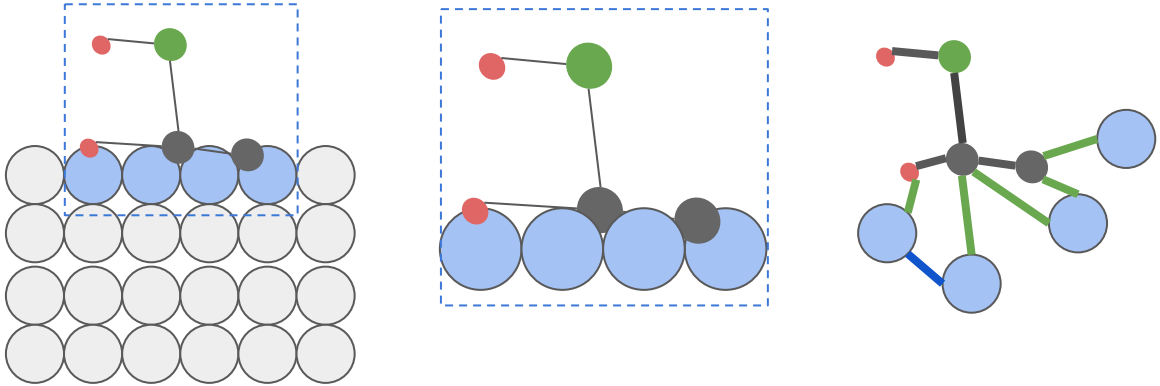


Fig 9. Different GNN architectures implemented in study 3

** Sergio Pablo-García, Santiago Morand et el, “Fast evaluation of the adsorption energy of organic molecules on metals via graph neural networks”*

Conclusion

- GNN can be
  - Difficult to implement for nanoparticles due to
    - large interaction graphs
    - Interactions within complex geometries
    - Ideally need to account all environmental effects
- But also, GNN are-
  - intuitive feature building,
  - lesser need for training data
  - fast learning time
  - Works well for solvation energy prediction
  - Can be used to predict IE for large molecules

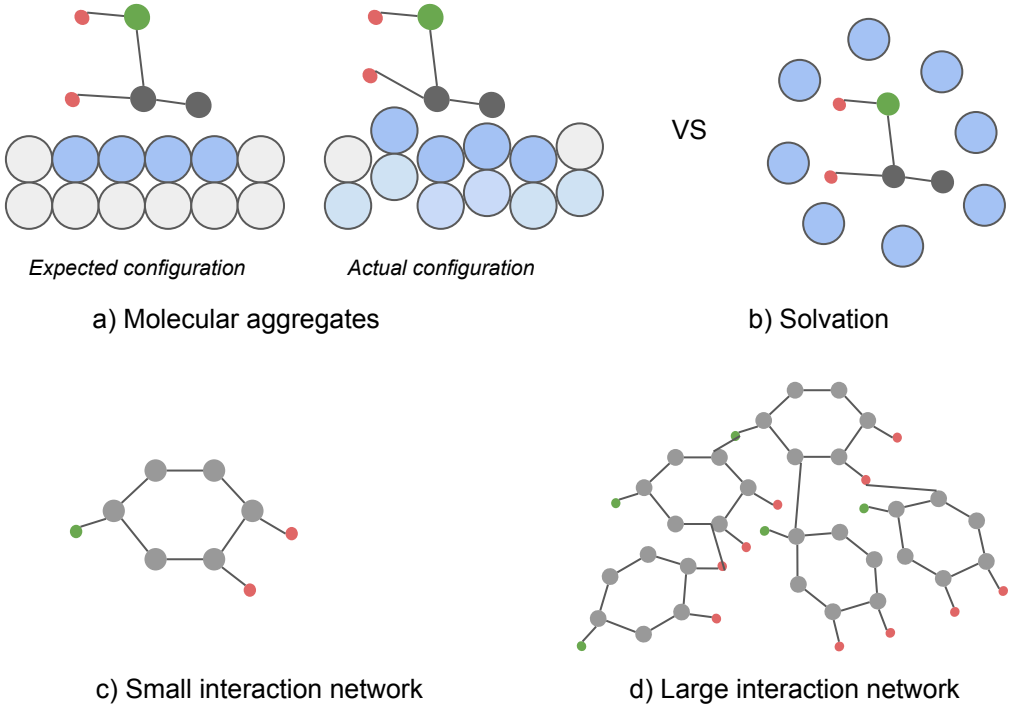


Fig 10. Diagrams depicting why it is difficult to implement GNN in molecular aggregates

Conclusion

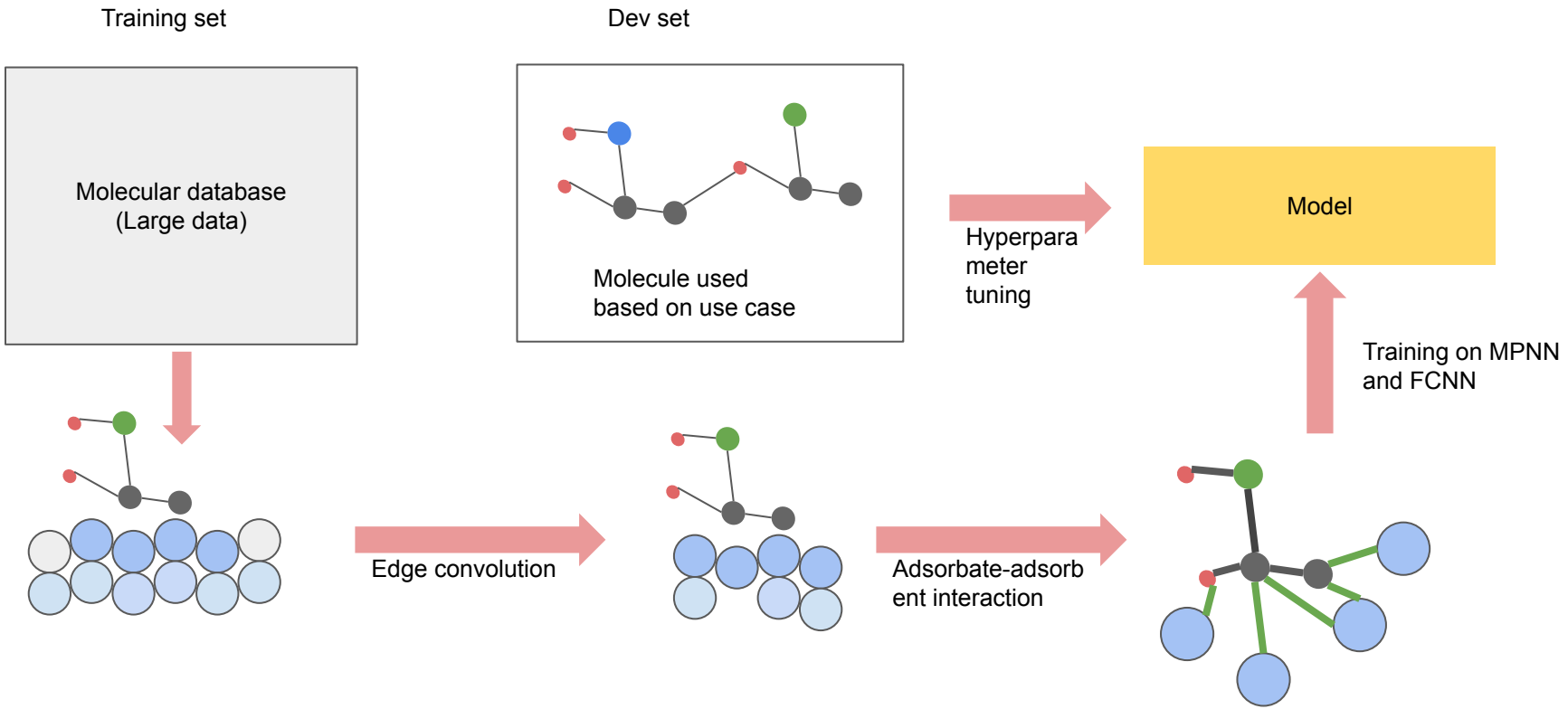


Fig 10. Proposed schematics for improved GNN architecture to predict interaction energy of Molecular aggregates and nanoparticles

Thankyou

Questions?