

Ph.D. Dissertation of Engineering

Graph Neural Networks for Molecular Property Prediction

분자 물성 예측을 위한 그래프 신경망

February 2026

Graduate School of Seoul National University
College of Engineering
Department of Computer Science and Engineering

Ji-woo Park

Unofficial template, not affiliated with or endorsed by Seoul National University. The author, committee and results in this sample are fictional. Delete this line before you submit.

Ph.D. Dissertation of Engineering

Graph Neural Networks for Molecular Property Prediction

분자 물성 예측을 위한 그래프 신경망

February 2026

Graduate School of Seoul National University
College of Engineering
Department of Computer Science and Engineering

Ji-woo Park

Graph Neural Networks for Molecular Property Prediction

Advisor: Professor Min-jun Kang

Submitting a Ph.D. Dissertation of Engineering
October 2025

Graduate School of Seoul National University
Department of Computer Science and Engineering

Ji-woo Park

Confirming the Ph.D. Dissertation written by Ji-woo Park
December 2025

Chair	_____	Seo-yeon Lim
Vice Chair	_____	Min-jun Kang
Examiner	_____	Hyun-woo Jeong
Examiner	_____	Soo-ah Yoon
Examiner	_____	Tae-ho Baek

Abstract

Predicting the properties of a molecule from its structure is central to drug discovery and materials design, and experimental measurements are slow and costly. Graph neural networks treat a molecule as a graph of atoms and bonds and learn representations directly from it. This dissertation studies how to make such models accurate when labelled data are scarce. It proposes a message-passing network that encodes bond angles and ring membership, a pre-training objective that predicts masked substructures on ten million unlabelled molecules, and a calibration method for its uncertainty estimates. On eight property benchmarks the pre-trained model reduces mean absolute error by 17% on average compared with the same architecture trained from scratch.

Keywords: graph neural networks, molecular property prediction, self-supervised pre-training, uncertainty calibration

Student Number: 2020-31234

Contents

Abstract	i
1 Introduction	1
2 Background	2
3 Geometry-Aware Pre-Training	3
4 Results	4
5 Conclusion	5
A Pre-Training Data	7
국문초록	8
Acknowledgements	9

List of Figures

3.1	Pre-training on unlabelled molecules before property fine-tuning.	3
-----	---	---

List of Tables

4.1	Mean absolute error on property benchmarks (lower is better).	4
-----	---	---

Chapter 1

Introduction

Screening candidate molecules by experiment is slow, and simulating them from first principles is expensive. Machine-learning models that predict properties from structure can narrow the search. Graph neural networks have become the standard architecture for this task because they operate directly on the molecular graph [1].

Chapter 2

Background

A message-passing neural network updates the representation h_v of each atom v from its neighbours $N(v)$ over T rounds,

$$h_v^{t+1} = U\left(h_v^t, \sum_{w \in N(v)} M(h_v^t, h_w^t, e_{vw})\right), \quad (2.1)$$

and a readout function pools the atoms into a molecular representation [1]. Graph convolutional networks are a special case with a fixed, normalized message function [2].

Chapter 3

Geometry-Aware Pre-Training

Our messages include the angles between adjacent bonds and a flag for ring membership, and the network is pre-trained to reconstruct masked substructures before it is fine-tuned on each property.

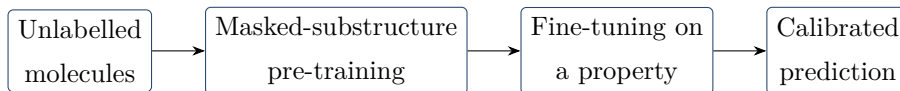


Figure 3.1: Pre-training on unlabelled molecules before property fine-tuning.

Chapter 4

Results

Table 4.1 reports mean absolute error on three of the eight benchmarks.

Table 4.1: Mean absolute error on property benchmarks (lower is better).

Model	Solubility	Lipophilicity	Band gap (eV)
From scratch	0.71	0.66	0.142
Pre-trained	0.58	0.55	0.119

Chapter 5

Conclusion

Encoding molecular geometry in messages and pre-training on unlabelled molecules both help when labels are scarce, and calibrated uncertainty tells a chemist when to trust a prediction.

Bibliography

- [1] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1263–1272, 2017.
- [2] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Appendix A

Pre-Training Data

The pre-training set contained ten million molecules drawn from a public database, filtered to at most 50 heavy atoms.

국문초록

분자의 구조로부터 물성을 예측하는 일은 신약 개발과 소재 설계의 핵심이지만, 실험적 측정은 느리고 비용이 많이 든다. 그래프 신경망은 분자를 원자와 결합으로 이루어진 그래프로 보고 그로부터 직접 표현을 학습한다. 본 논문은 레이블된 데이터가 부족할 때에도 정확한 모델을 만드는 방법을 연구한다. 결합 각과 고리 구조를 반영하는 메시지 전달 신경망, 천만 개의 레이블 없는 분자에서 가려진 부분 구조를 예측하는 사전 학습 목표, 그리고 불확실성 추정의 보정 방법을 제안한다. 여덟 가지 물성 벤치마크에서 사전 학습된 모델은 처음부터 학습한 같은 구조의 모델에 비해 평균 절대 오차를 평균 17% 줄였다.

주요어: 그래프 신경망, 분자 물성 예측, 자기지도 사전 학습, 불확실성 보정

학번: 2020-31234

Acknowledgements

I thank my advisor, Professor Min-jun Kang, for his guidance, and the members of my dissertation committee for their advice. I am grateful to my lab-mates and to my family for their support.