A Theoretically-Principled Sparse, Connected, and Rigid Graph Representation of Molecules

Shih-Hsin Wang*[1], Yuhao Huang*[1], Justin Baker[2], Yuan-En Sun[1], Qi Tang[3], Bao Wang[1]

1: University of Utah,

2: University of California, Los Angeles,

3: Georgia Institute of Technology

Motivation

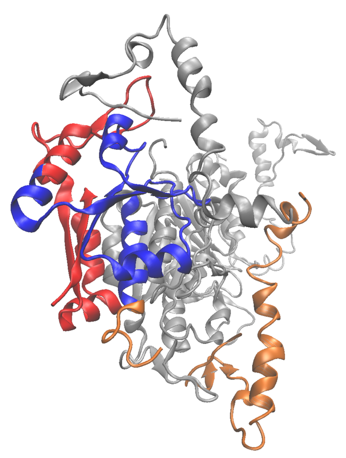
- GNNs need graphs to analyze 3D data.
- Graph structure impacts performance.

What Makes a 3D Graph Good?

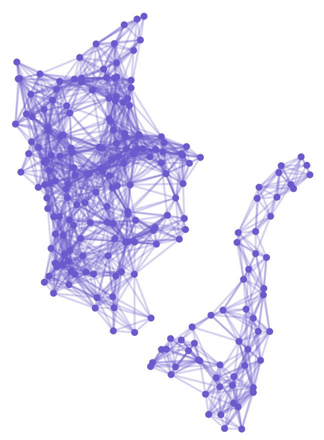
- **Sparsity** → Efficient computation
- **Connectedness** → Full message passing
- **Rigidity** → Unique 3D structure recovery

Challenges

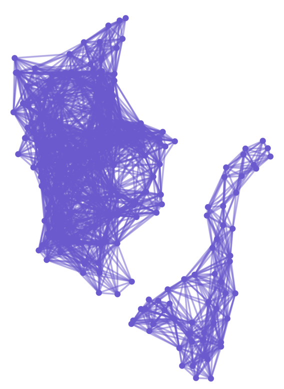
- Existing methods (e.g., radial cutoff) **trade off** between sparsity and connectedness.
- No method ensures all 3 properties **at once**.



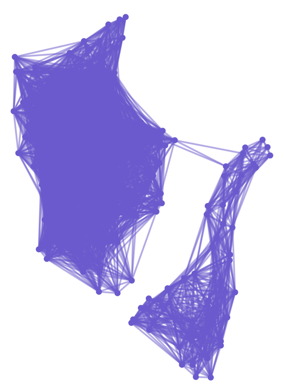
Protein



Cutoff = 8

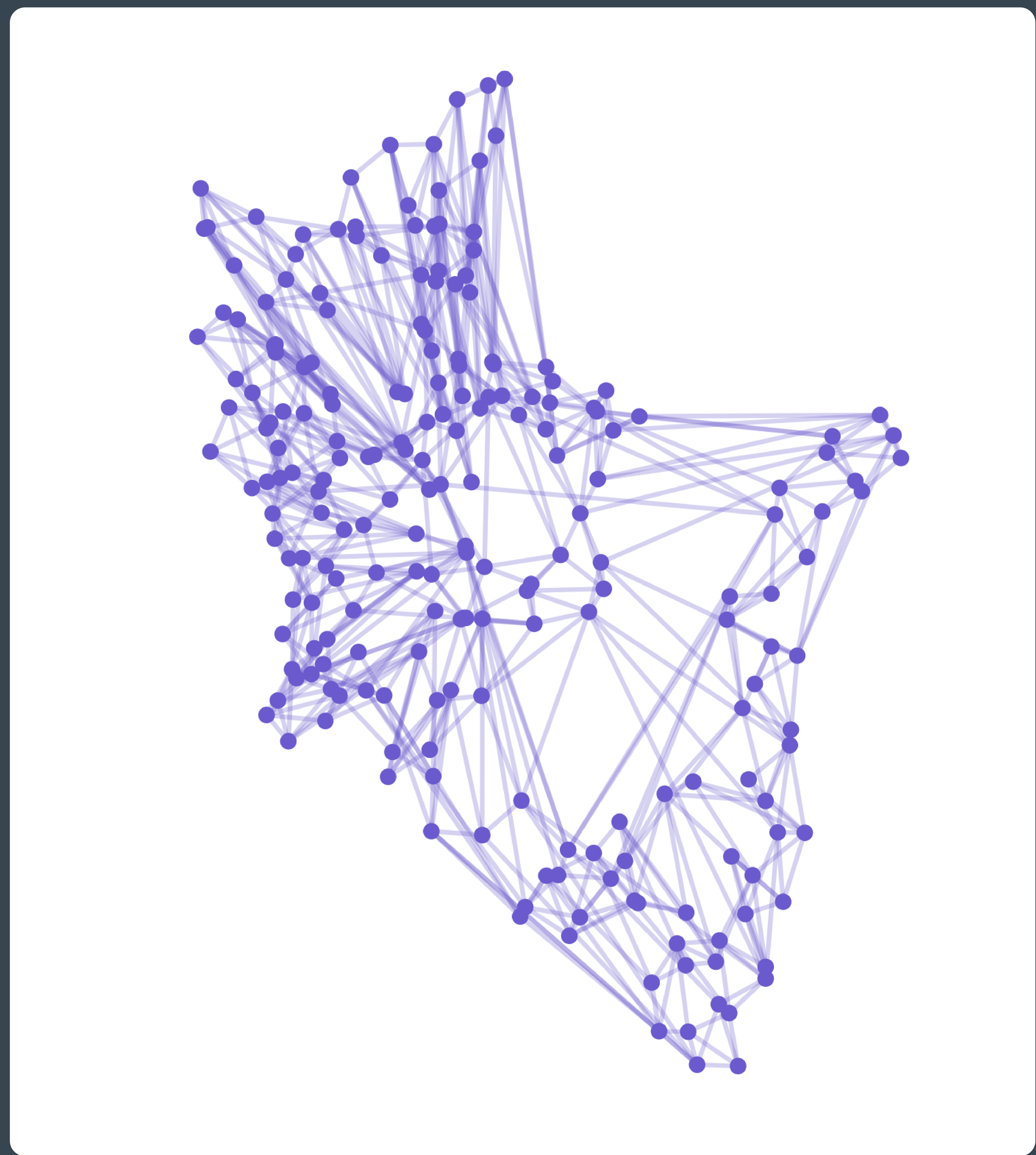


Cutoff = 12



Cutoff = 16

A **hyperparameter-free** method to construct **sparse, connected, and rigid** graphs for molecules, with **theoretical guarantees** and real-world performance.



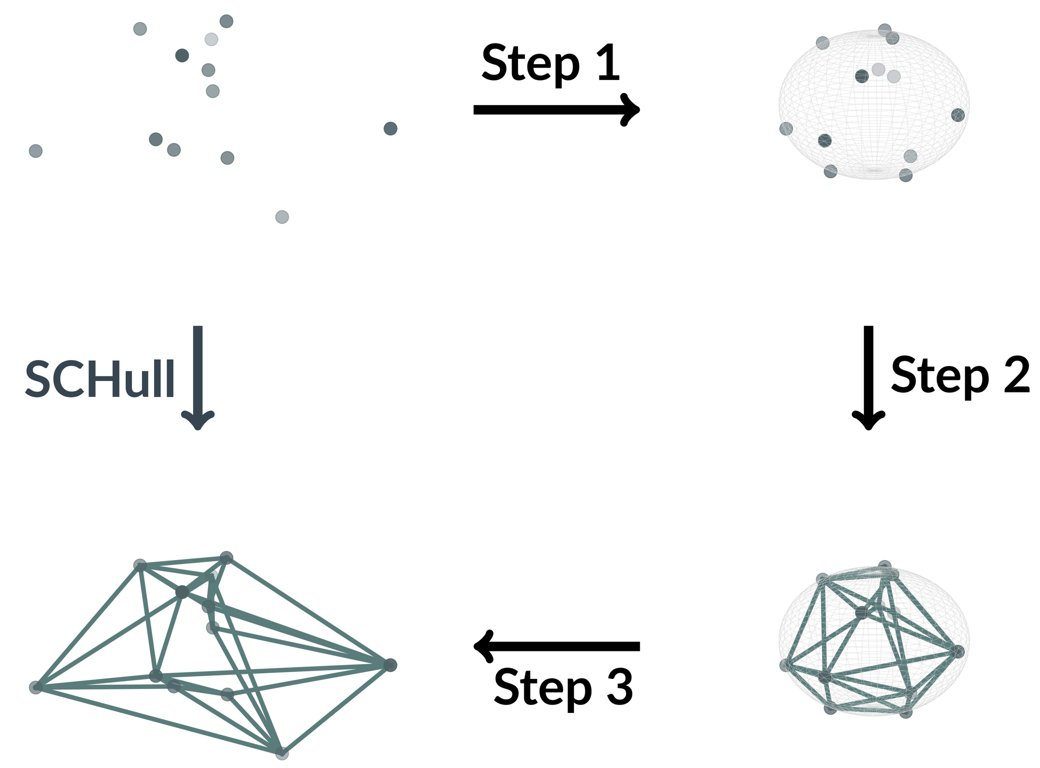
Scan to access paper, code, and slides

Our Solution: SCHull

Step 1: Project to unit sphere

Step 2: Compute spherical convex hull

Step 3: Use hull edges to define the graph



Why It Works

- **No hyperparameters.**
- **Sparse & Connected**: #Edges $<3\times$ #Nodes.
- **Rigid**: Dihedrals + lengths + distances to center.
- **Theoretical Guarantee:** 1-layer GNN can distinguish between non-isomorphic point clouds.

Selected Results

Fold and Reaction Classification (Accuracy %)

Method	React	Avg. Time	Fold				Avg. Time
			Fold	Super	Family	Avg.	
GCN	67.3		16.8	21.3	82.8	40.3	–
IEConv	87.2	–	45.0	69.7	98.9	71.2	–
DWNN	76.7	–	31.8	37.8	85.2	51.5	–
GearNet	79.4	–	28.4	42.6	95.3	55.4	–
HoloProt	78.9	–	–	–	–	–	–
MACE	–	–	$23.7_{\pm 0.5}$	$21.4_{\pm 0.5}$	$60.2_{\pm 0.2}$	35.1	$114_{\pm 0.5}$
MACE+SCHull	–	–	$27.0_{\pm 0.6}$	$23.1_{\pm 0.5}$	$65.0_{\pm 0.2}$	38.4	$135_{\pm 0.5}$
SEGNN	–	–	$28.8_{\pm 0.6}$	$30.3_{\pm 0.6}$	$77.1_{\pm 0.3}$	45.4	$121_{\pm 0.7}$
SEGNN+SCHull	–	–	$32.0_{\pm 0.4}$	$36.8_{\pm 0.7}$	$86.9_{\pm 0.3}$	51.9	$152_{\pm 0.5}$
GVP-GNN	65.5	$320_{\pm 5}$	16.0	22.5	83.8	40.8	$106.3_{\pm 0.5}$
GVP-GNN + SCHull	$77.1_{\pm 0.5}$	$345_{\pm 5}$	$24.5_{\pm 0.3}$	$27.1_{\pm 0.2}$	$88.6_{\pm 0.3}$	46.7	$111.5_{\pm 0.5}$
ProNet-Amino-Acid	86.0	$210_{\pm 5}$	51.5	69.9	99.0	73.5	$70.5_{\pm 0.5}$
ProNet-Amino Acid+SCHull	$87.9_{\pm 0.3}$	$221_{\pm 6}$	$55.2_{\pm 0.2}$	$73.9_{\pm 0.2}$	$99.1_{\pm 0.1}$	76.1	$73.8_{\pm 0.5}$
ProNet-Backbone	86.4	$213_{\pm 5}$	52.7	70.3	99.3	74.1	$71.4_{\pm 0.8}$
ProNet-Backbone+SCHull	$\mathbf{88.1}_{\pm 0.3}$	$230_{\pm 5}$	$\mathbf{56.1}_{\pm 0.3}$	$\mathbf{74.6}_{\pm 0.2}$	$\mathbf{99.4}_{\pm 0.1}$	**76.7**	$75.8_{\pm 0.5}$

Ligand Binding Affinity Prediction

Method	LBA				Avg. Time
	RMSE↓	Pearson↑	Spearman↑	Kendall↑	
IEConv	1.554	0.414	0.428	–	–
HoloProt-Full Surface	1.464	0.509	0.500	–	–
HoloProt-Superpixel	1.491	0.491	0.482	–	–
GVP-GNN	$1.529_{\pm 0.001}$	$0.441_{\pm 0.001}$	$0.432_{\pm 0.002}$	$0.301_{\pm 0.002}$	$48.6_{\pm 0.6}$
GVP-GNN + SCHull	$1.401_{\pm 0.001}$	$0.475_{\pm 0.001}$	$0.459_{\pm 0.001}$	$0.335_{\pm 0.002}$	$53.6_{\pm 0.6}$
ProNet-Amino-Acid	1.455	0.536	0.526	$0.465_{\pm 0.001}$	$31.7_{\pm 0.5}$
ProNet-Amino Acid+SCHull	$1.355_{\pm 0.002}$	$0.556_{\pm 0.001}$	$0.568_{\pm 0.001}$	$0.512_{\pm 0.001}$	$33.9_{\pm 0.5}$
ProNet-Backbone	1.458	0.546	0.550	$0.481_{\pm 0.001}$	$32.1_{\pm 0.5}$
ProNet-Backbone+SCHull	$\mathbf{1.321}_{\pm 0.002}$	$\mathbf{0.581}_{\pm 0.001}$	$\mathbf{0.578}_{\pm 0.1}$	$\mathbf{0.535}_{\pm 0.001}$	$34.4_{\pm 0.5}$

See more results — scan the QR code!