Notes/Questions (To Be Deleted) I

- Still unsure how to tie in PrimeKG paper
- Methods for edge selection/identification? Some instances we only want to preserve structure, but sometimes we want to predict/fit edges, how does that fit into message-passing/GCN methods or other extensions?
- Diffpool article as extension of general GNN, hierarchical pooling over global readout functions?
  https://papers.nips.cc/paper_files/paper/2018/file/e77dbaf6759253c7c6d0efc5690369c7-Paper.pdf

Downloadable Slides

Graphical Neural Networks

Dominic DiSanto

Department of Biostatistics, Harvard University

December 5, 2023

Downloadable Slides

Outline

1. Set-Up and Motivation 5ish minutes
2. General Construction 10-20ish minutes
3. Applications and Extensions 20-25ish minutes
   - Multimodal Graph Learning (MGL)
   - Knowledge-Graph Data
   - Structure-Based Drug Design (SBDD)

Downloadable Slides

Goals

- Provide a useful overview of Graphical Neural Networks (GNN)
  - Motivation for necessity of GNN's
  - Provide a general framework of fitting
- Describe applications and extensions of the general GNN model
  - Multimodal Physiological/Biomedical Data
  - Integration of Knowledge-Graph and EHR Data

Downloadable Slides

KG Application

Image here. Or maybe drop, 4 might be overkill

Downloadable Slides

Multi-modal Biomedical Data

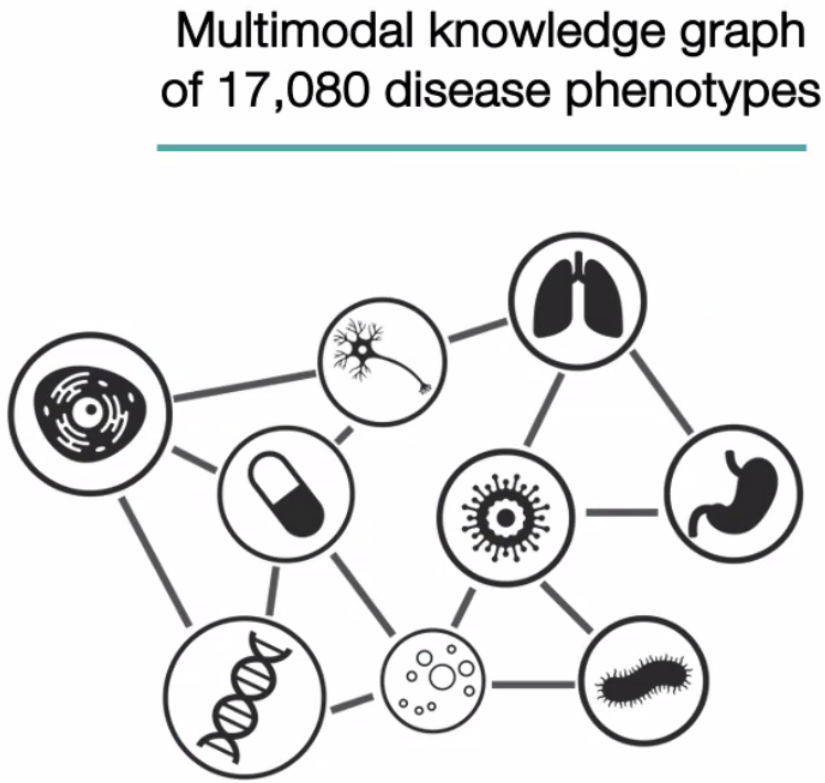
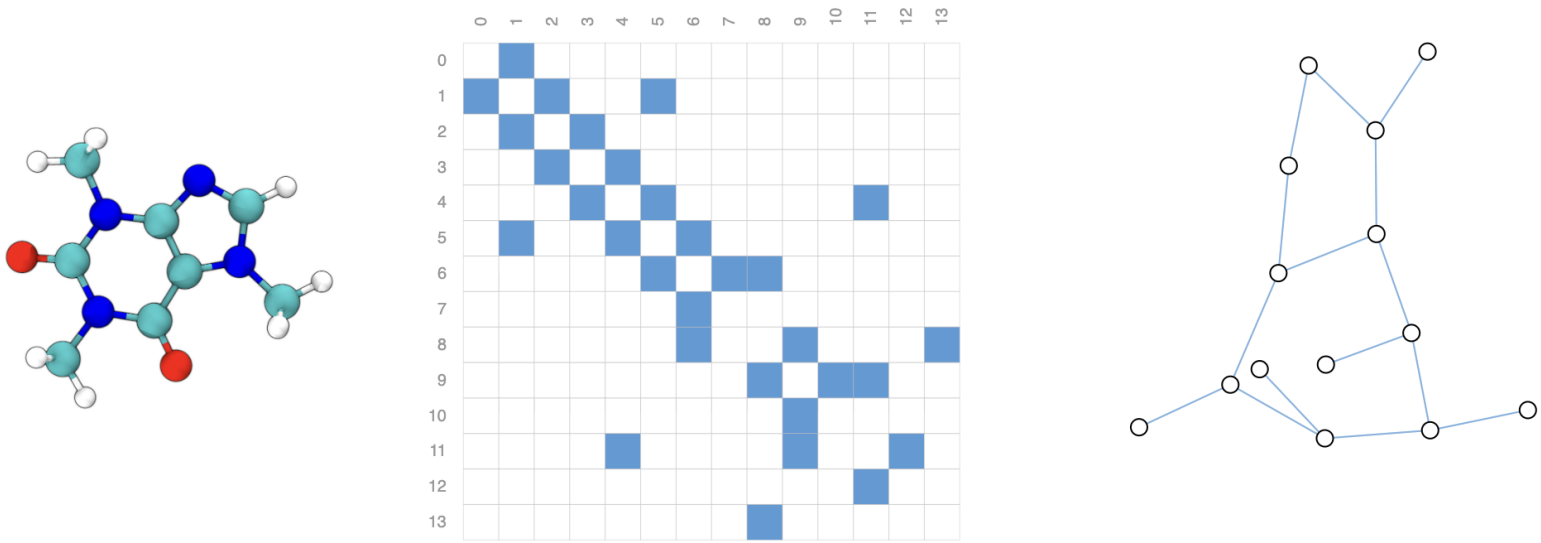


Image courtesy of partial figure from McDermott et al. *Structure-inducing pre-training* [10]

Downloadable Slides

Molecular/Biochemical

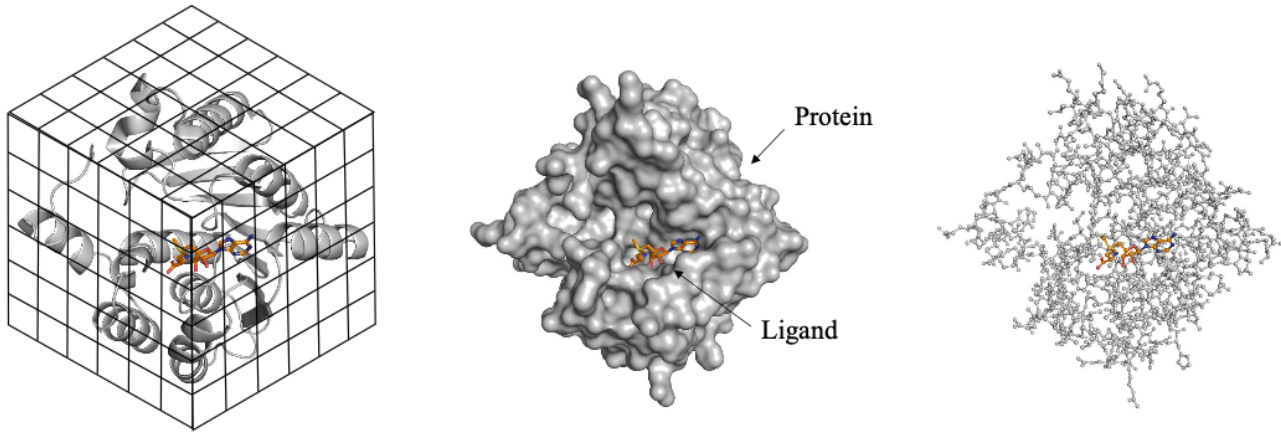


(Left) 3d representation of the Caffeine molecule (Center) Adjacency matrix of the bonds in the molecule (Right) Graph representation of the molecule.

Image courtesy of https://distill.pub/2021/gnn-intro/ [11]

Downloadable Slides

Protein Representation



(a) 3D grid (b) 3D surface (c) 3D graph

Fig. 2. 3D representations of proteins used for geometric deep learning: (a) 3D grid, (b) 3D surface, and (c) 3D graph, illustrated for PDB ID 2avd.

Fig 2. of Zhang (2023) *Geometric Deep Learning for Structure-Based Drug Design: A Survey* [15]

Downloadable Slides

Motivation

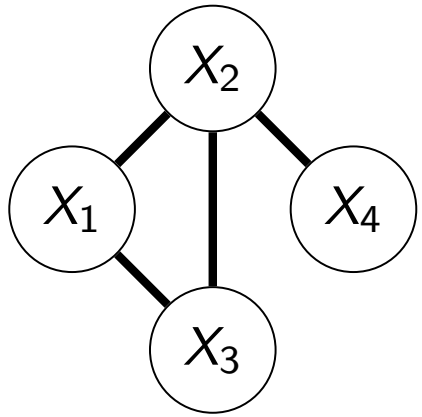
- Want to utilize the input structure of the graph
  - Respect/Maintain
  - Update/Estimate
- "Flattening" graphical data for DNN, CNN, etc. omits useful topology from our data
- Early methods attempting to retain topological info included recursive neural networks and random walk models, which GNN methodology extended [12]

Downloadable Slides

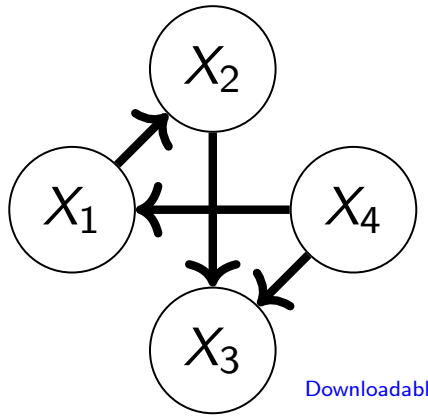
Notation/Set-Up

- Consider the graph $\mathcal{G} = (V, E), E \subseteq V \times V$, where any node v has a related "feature vector" $x_v \in \mathbb{R}^d$
  - Let $N = N$
- Let $\mathcal{N}_s(v)$ represent the s-hop neighborhood of any node v (and implicitly $\mathcal{N}(v) \equiv \mathcal{N}_1(v)$)
- Can construct adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ to capture structure of edge set E
  - $\mathbf{A}_{ij} = w_{ij} \mathbb{1}\{(i,j) \in E\}$ for scalar weight $w_{ij} \in \mathbb{R}$

Undirected Graph



Directed Graph



Downloadable Slides

Topology Representations

- Simplest/Naïve method is to use $\mathbf{A}$
- Consider also the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$
  - $\mathbf{D} = \text{diag}(\mathbf{A1_N})$
- Can use an eigenvalue-normalized Laplacian $\widetilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$
- Note that a given graph topology is represented equivalently by any permutation of its $\mathbf{A}$

Downloadable Slides

Outline

1. Set-Up and Motivation 5ish minutes
2. General Construction 10-20ish minutes
3. Applications and Extensions 20-25ish minutes
   - Multimodal Graph Learning (MGL)
   - Knowledge-Graph Data
   - Structure-Based Drug Design (SBDD)

Downloadable Slides

What do we estimate about graph structure?

$$\text{GNN learning can be} \begin{cases} \text{Node-wise } \Phi(\mathcal{G}, v) : (v \in V) \to \mathbb{R}^m \\ \text{Edge-wise } \Phi(\mathcal{G}, e) : (e \in E) \to \mathbb{R}^m \\ \text{Graph-level characteristics } \Phi(\mathcal{G}) \end{cases}$$

Downloadable Slides

What do we estimate about graph structure?

GNN learning can be

- Node-wise $\Phi(\mathcal{G}, x) : (x \in V) \to \mathbb{R}^m$
- Graph-level characteristics $\Phi(\mathcal{G}, e) : (e \in E) \to \mathbb{R}^m$
- Graph-level characteristics $\Phi(\mathcal{G})$

Downloadable Slides

General Framework[1]

We begin with the general Message Passing Neural Networks (MPNN) structure of a GNN:

Initialize $h^{(0)} \leftarrow x_v, \forall v \in V$
for $\kappa = 0, ..., K$ **do**:
  for $v \in \mathcal{G}$ **do**:
    $m_u^{(\kappa+1)} \leftarrow \text{Message}(h_v, h_u, e_{vu}), \forall u \in \mathcal{N}_v$
    $h_{agg}^{(\kappa+1)} \leftarrow \text{Aggregate}(\{m_u^{(\kappa)} \mid u \in \mathcal{N}_v\})$
    $h^{(\kappa+1)} \leftarrow \text{Update}(h^{(\kappa)}, h_{agg}^{(\kappa)})$
$\hat{y} \leftarrow \text{Transform}(\{h_v^K | v \in \mathcal{G}\})$ or $\text{Readout}(\cdot)$

[1] See [2, 5, 13]

General Framework

Can succinctly represent the κth layer as:

$$\mathbf{h}_v^{(\kappa+1)} = Up\left\{x_v^{(\kappa)}, Agg\left[Msg\left(h_v^{(\kappa)}, x_u^{(\kappa)}, e_{u,v}^{(\kappa)}\right)\right]\right\}$$

Choices of (differentiable) functions for Aggregate, Update, and Readout determine the architecture of your GNN

Trained end-to-end via backpropagation for problem-specific Transform function

Downloadable Slides

Aggregate & Update

- **Aggregate**$(\cdot)$ produces a representation of information from a node's neighborhood via **permutation invariant** function
- Can include weights (edge-wise or learned)
- Over later iterations, this includes information from further and further distant nodes to any one target node
- We then **Update**$(\cdot)$ our current state using this aggregated neighborhood-level information

Downloadable Slides

Transform/Readout

- **Transform**$(\cdot)$ translates our learned node representations to some desired outcome
  - Regression
  - Binary/Multi-class classification
  - MLP/DNN's
  - **Readout**$(\cdot)$ is common term for translating node-level output to graph-level
  - *Global Pooling* - Methods applied over entire graph (e.g. averaging, fitting "regular" deep neural network, etc.)

Downloadable Slides

Graph Convolutional Network

- Proposed in 2017 by Thomas Kipf, Max Welling [9], can consider one example of "Laplacian-based methods" [5]

$$\mathbf{H}^{(\kappa+1)} = \text{ReLu}\left(\widetilde{\mathbf{D}}^{-1/2}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-1/2}\mathbf{H}^{(\kappa)}\mathbf{\Theta}\right)$$

- Motivated by considering the graph convolution[2] $x \star g = UgU^T x$ as the message passing function
- Learned weight/parameter matrix $\mathbf{\Theta}$

[2] U the matrix of eigenvectors of $\mathbf{L}$

Downloadable Slides

Graph Convolutional Network

- Proposed in 2017 by Thomas Kipf, Max Welling [9], can consider one example of "Laplacian-based methods" [5]

$$\mathbf{H}^{(\kappa+1)} = \underbrace{\text{ReLu}}_{\text{Update}}\underbrace{\left(\widetilde{\mathbf{D}}^{-1/2}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-1/2}\mathbf{H}^{(\kappa)}\mathbf{\Theta}\right)}_{\text{Aggregate/Message}}$$

- Motivated by considering the graph convolution[3] $x \star g = UgU^T x$ as the message passing function
- Learned weight/parameter matrix $\mathbf{\Theta}$

[3] U the matrix of eigenvectors of $\mathbf{L}$

Downloadable Slides

Graph Convolutional Network

Maybe move to appendix

Intuitive "derivation":

$$\mathbf{H}^{(\kappa+1)} = \sigma\left(\mathbf{A}\mathbf{H}^{(\kappa)}\Theta\right)$$

$$\mathbf{H}^{(\kappa+1)} = \sigma\left(\mathbf{D}^{-1}\mathbf{A}\mathbf{H}^{(\kappa)}\Theta\right) \quad \text{Normalizing by degree}$$

$$\mathbf{H}^{(\kappa+1)} = \sigma\left(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\mathbf{H}^{(\kappa)}\Theta\right) \quad \text{Symmetric normalization}$$

$$\mathbf{H}^{(\kappa+1)} = \sigma\left(\widetilde{\mathbf{D}}^{-1/2}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-1/2}\mathbf{H}^{(\kappa)}\Theta\right) \quad \text{Adding self-loop}$$

where $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}, \widetilde{\mathbf{D}}_{ii} = \sum_j \widetilde{\mathbf{A}}_{ij}$, σ is any activation function, $\mathbf{H}$ is simply the matrix of $\mathbf{h}_v$ for all nodes

Downloadable Slides

General Framework (Revisited)

- **Input**
  - Node conceptualization
  - Node embeddings
- **Architecture**
  - $\mathbf{H}^{(\kappa+1)} = \text{ReLu}\left(\widetilde{\mathbf{D}}^{-1/2}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-1/2}\mathbf{H}^{(\kappa)}\mathbf{\Theta}\right)$
- **Output**
  - Target output (Readout, Transform functions)
  - Loss function

Downloadable Slides

Outline

1. Set-Up and Motivation 5ish minutes
2. General Construction 10-20ish minutes
3. Applications and Extensions 20-25ish minutes
   - Multimodal Graph Learning (MGL)
   - Knowledge-Graph Data
   - Structure-Based Drug Design (SBDD)

Downloadable Slides

Multimodal Graph Learning (MGL)

Downloadable Slides

Multimodal Graph Learning (MGL)

- Ektefaie (2023) *Multimodal learning with graphs* [2]
- Cite 14-16 in multimodal paper for multi improvement over uni VAE
- Topology is complicated by multimodality of input data
  - Modal collapse [8]
  - Differential data availability
  - Can mention Kipf algorithm does well in semi-supervised setting (even with limited labels available within a given "cluster"), so if we have some gold standard data (biomarker/strong proxy, etc.) that is difficult to obtain, can be VERY useful to include for MGL and lead to good performance, no bullet point here just mention

Downloadable Slides

Mutimodal

Clinical Data

- Clinical text
- -omics data
- Laboratory measurements
- Clinical imaging

Protein Structures

- 1° AA Sequence
- 2° Helix interactions
- 3° Folding, bridges

Downloadable Slides

Framework

Authors propose a four component "blueprint"

1. Identifying entitites (i.e. modalities)
2. Uncovering topology
   - *A priori*
   - Adaptively learned
3. Propagating information
4. Mixing representation

Downloadable Slides

Framework

Authors propose a four component "blueprint"

1. Identifying entitites (i.e. modalities) } Structure Learning
2. Uncovering topology
3. **Propagating information** } Learning On-Structure Phase
4. **Mixing representation**

Downloadable Slides

Structure Learning

- Consider patients as *nodes*
- Consider modalities as *entities* (colored nodes)
  - Clinical text/narrative data
  - Laboratory/Physiological measurements
  - Image/Video data
  - Patient reported measurements/symptoms
  - etc.

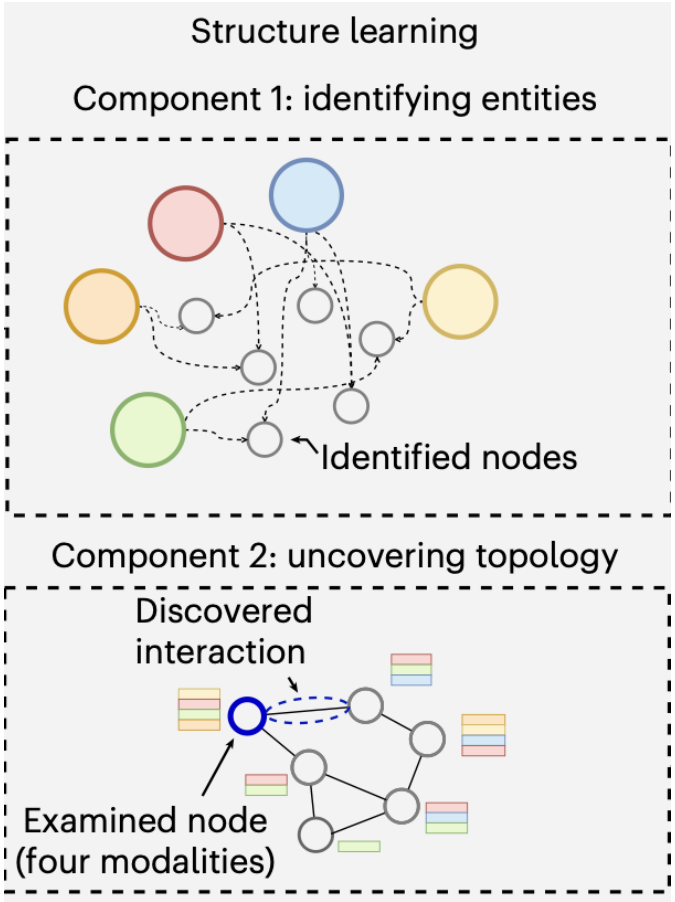


Figure: Subset of Fig 2c [2]

Downloadable Slides

Learning on Structure

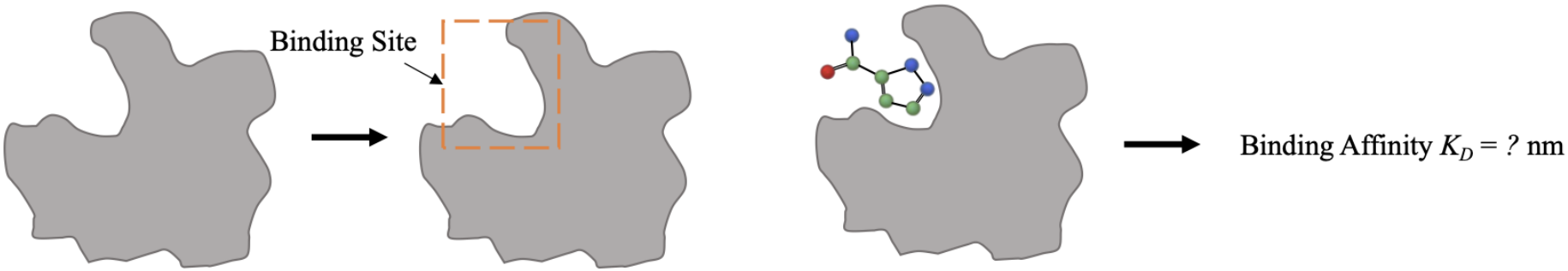
Downloadable Slides

Proteins and Structure-Based Drug Design (SBDD)

Downloadable Slides

Structure-Based Drug Design (SBDD)

- Zhang (2023) *Geometric Deep Learning for Structure-Based Drug Design: A Survey* [15]
- SBDD aims to improve drug-discovery by understanding 3D protein structures and predicting drug efficacy/behavior
- Problems of note for GNN's include **binding site prediction** and **binding affinity prediction**



Images from Fig. 1 of Zhang (2023) [15]

Downloadable Slides

Output

- **Binding site prediction** is binary categorization at the (surface) amino acid level
- **Binding affinity prediction** is a continuous measure of protein-ligand interaction strength
- Protein Databank (PDB), PDBBind, Dockground, CSAR-HiQ are existing data sets used to train most GNN's in these contexts
- Other characteristics are important for drug design and protein-ligand interactions but without current GNN methods to my knowledge
  - Binding pose
  - Ligand generation
  - Linker design

Downloadable Slides

Input

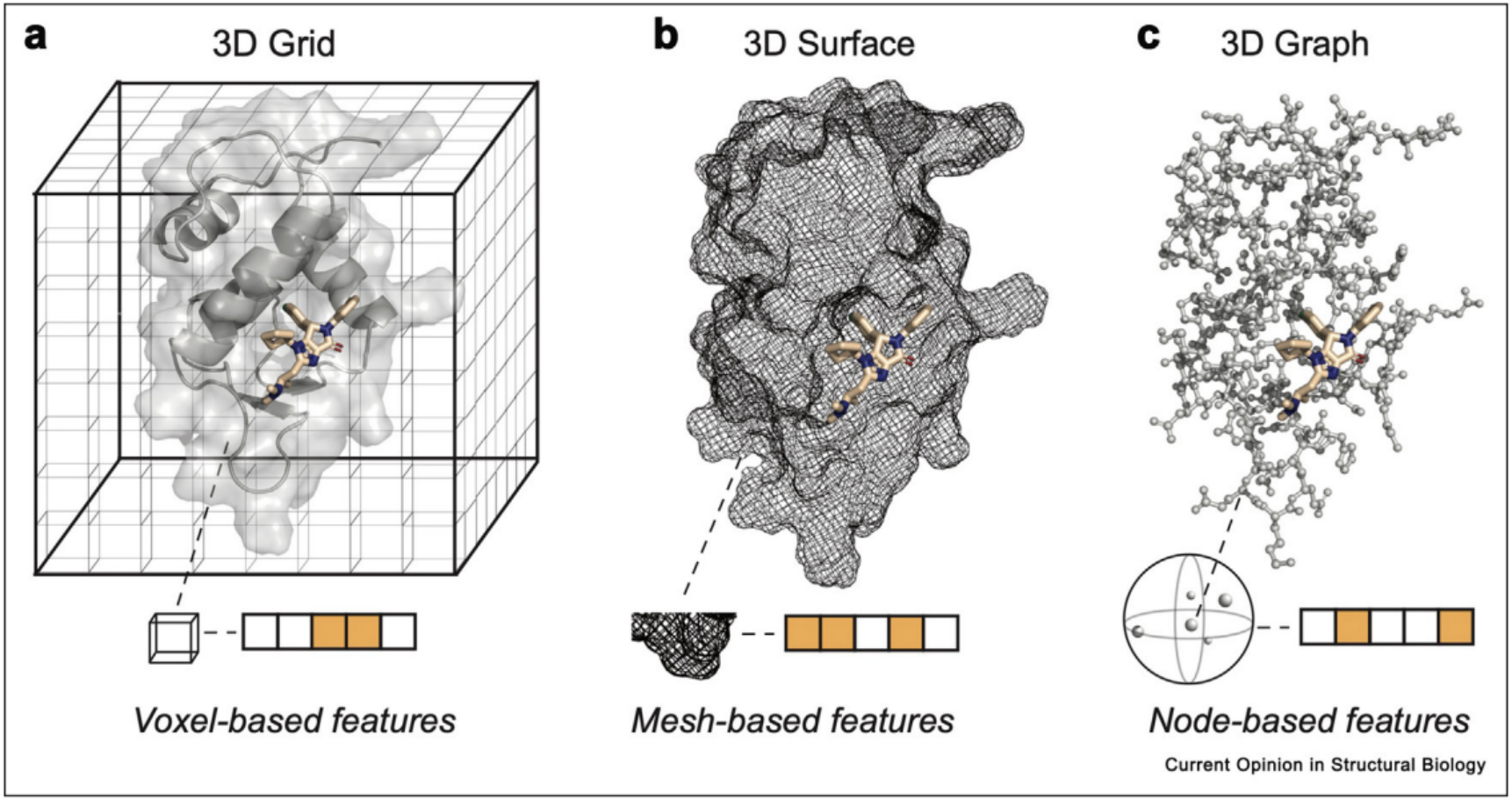
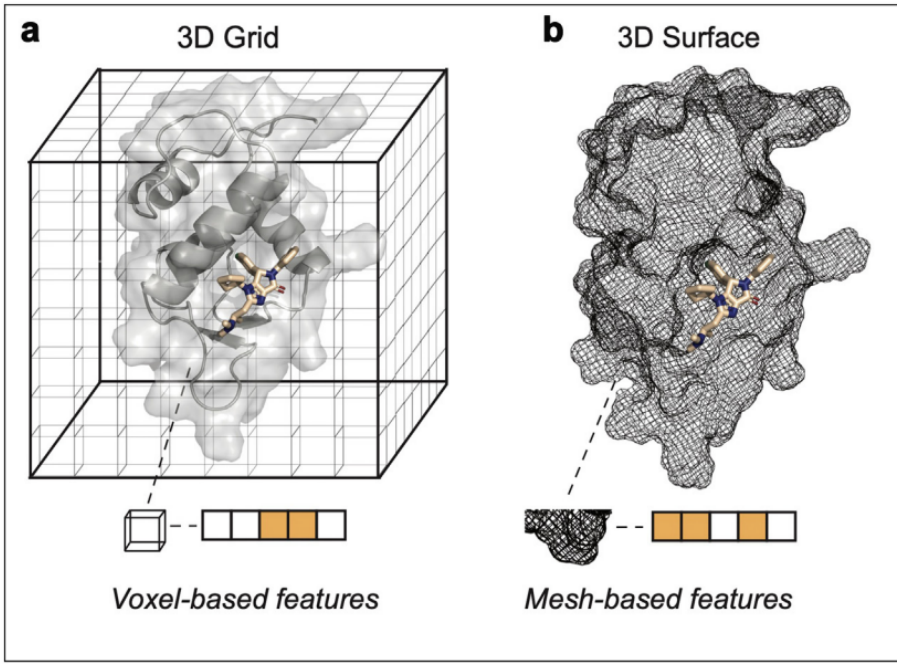


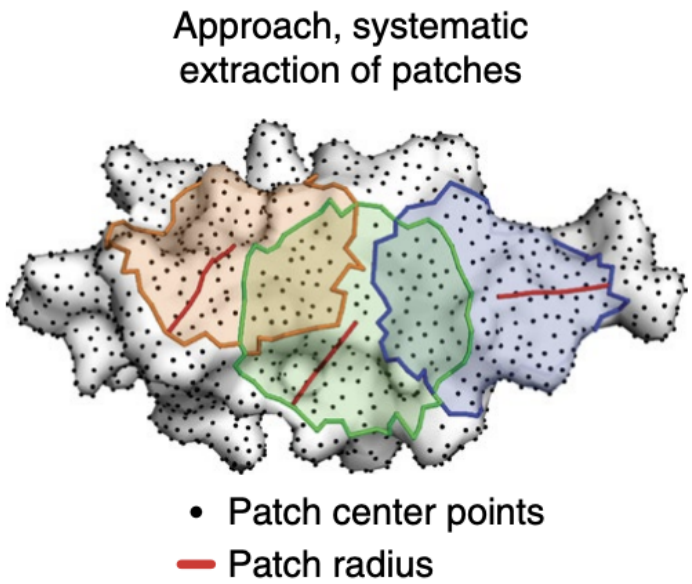
Figure from Isert (2023) [7]

Downloadable Slides

CNN Input

- Nodes and edges defined by radial patches about a discretization of the protein surface





Left: Fig 2(ab) from Isert (2023); Right: Fig 1a. Gainza (2020) [7, 4]

Downloadable Slides

GNN Input

- Atomic structure in Euclidean space
- Amino acid (also "residue") structure

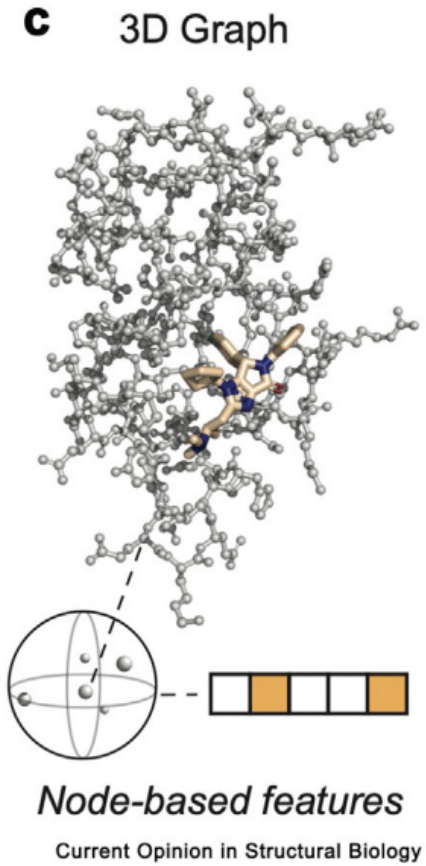


Figure: Figure 1c from Isert (2023) [7]

Downloadable Slides

Architecture

Consider some transformations T, T' (e.g. reflection, rotation, etc.) within the same symmetry group:

Invariance: $f(T(x)) = T'(f(x)) = f(x)$

Equivariance: $\forall T,\ \exists T' : f(T(x)) = T'(f(x))$

Downloadable Slides

Architecture

Forgoing the iteration supercripts, one message pass can be represented as such:

$$
\begin{aligned}
m_{ij} &= Msg_m(\mathbf{v_i}, \mathbf{v_j}, h_i, h_j, e_{ij}) \\
\mathbf{m}_{ij} &= Msg_{\mathbf{m}}(\mathbf{v_i}, \mathbf{v_j}, h_i, h_j, e_{ij}) \\
h_i' &= Update_h \left[h_i, Agg_h(\{m_{ij}\}_{j \in \mathcal{N}(v_i)})\right] \\
\mathbf{v_i'} &= Update_{\mathbf{v}} \left[\mathbf{v}_i Agg_{\mathbf{v}} \left(\{\mathbf{m}_{ij}\}_{j \in \mathcal{N}(v_i)}\right)\right]
\end{aligned}
$$

where $Msg_m, Update_h$ are geometrically **invariant**, scalar functions and $Msg_{\mathbf{m}}, Update_{\mathbf{v}}$ are geometrically **equivariant, vector functions**
Here $\mathbf{v} \in \mathbb{R}^3$ are 3-D coordinates (e.g. atom or amino acid location)

Downloadable Slides

Equivariant GNN's

Recreate message passing for equivariant GNN's with my notation for consistency, emphaseize new equivariant vector functions

$$h_{ij}^{(\kappa)} = \text{Aggregate}(\mathbf{v}_i,)$$

Downloadable Slides

Specific GNN Implementation

Review/cite one of the eGNN's from the Zhang paper

Conclusion

Downloadable Slides

References I

- Some diagrams generated in conjunction with ChatGPT 3.5

[1] David Duvenaud et al. *Convolutional Networks on Graphs for Learning Molecular Fingerprints*. arXiv:1509.09292 [cs, stat]. Nov. 2015.

[2] Yasha Ektefaie et al. "Multimodal learning with graphs". en. In: *Nature Machine Intelligence* 5.4 (Apr. 2023). Number: 4 Publisher: Nature Publishing Group, pp. 340–350.

[3] Oleksandr Ferludin et al. *TF-GNN: Graph Neural Networks in TensorFlow*. arXiv:2207.03522 [physics, stat]. July 2023.

[4] P. Gainza et al. "Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning". en. In: *Nature Methods* 17.2 (Feb. 2020). Number: 2 Publisher: Nature Publishing Group, pp. 184–192.

Downloadable Slides

References II

[5] Justin Gilmer et al. *Neural Message Passing for Quantum Chemistry*. arXiv:1704.01212 [cs]. June 2017.

[6] M. Gori, G. Monfardini, and F. Scarselli. "A new model for learning in graph domains". In: *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.* Vol. 2. ISSN: 2161-4407. July 2005, 729–734 vol. 2.

[7] Clemens Isert, Kenneth Atz, and Gisbert Schneider. "Structure-based drug design with geometric deep learning". en. In: *Current Opinion in Structural Biology* 79 (Apr. 2023), p. 102548.

[8] Adrián Javaloy, Maryam Meghdadi, and Isabel Valera. *Mitigating Modality Collapse in Multimodal VAEs via Impartial Optimization*. en. arXiv:2206.04496 [cs]. June 2022.

Downloadable Slides

References III

[9] Thomas N. Kipf and Max Welling. *Semi-Supervised Classification with Graph Convolutional Networks*. arXiv:1609.02907 [cs, stat]. Feb. 2017.

[10] Matthew B. A. McDermott et al. "Structure-inducing pre-training". en. In: *Nature Machine Intelligence* 5.6 (June 2023). Number: 6 Publisher: Nature Publishing Group, pp. 612–621.

[11] Benjamin Sanchez-Lengeling et al. "A Gentle Introduction to Graph Neural Networks". en. In: *Distill* 6.9 (Sept. 2021), e33.

[12] F. Scarselli et al. "The Graph Neural Network Model". en. In: *IEEE Transactions on Neural Networks* 20.1 (Jan. 2009), pp. 61–80.

[13] Keyulu Xu et al. *How Powerful are Graph Neural Networks?* arXiv:1810.00826 [cs, stat]. Feb. 2019.

Downloadable Slides

References IV

[14] Zhitao Ying et al. "Hierarchical Graph Representation Learning with Differentiable Pooling". In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc., 2018.

[15] Zaixi Zhang et al. *A Systematic Survey in Geometric Deep Learning for Structure-based Drug Design*. arXiv:2306.11768 [cs, q-bio]. Oct. 2023.

Downloadable Slides

Appendix Slides

Downloadable Slides

Implementations

- PyTorch Geometric with directed extension PyTorch Geometric Signed Directed
- TensorFlow GNN [3]
- GraphNeuralNetworks.jl
- Spektral, Keras-based Pythhon package
- Limited but some implementation in R
  - scapGNN, package GNN implementation but specific/narrow for single-cell -omics data

Downloadable Slides

Abbreviated History

- Graph Neural Network first(?) coined in Gori (2005) [6] and subqeuently in Scarselli (2009) *The Graph Neural Network Model* [12]
- Graph Convolutional Network by Kipf (2017) [9] but with similar convolutional message-passing algorithms (within GNN's) proposed in at least 2015 [1]
- Message passing GNN proposed in Gilmer (2017), applications in molecular chemistry [5]

Downloadable Slides

Additional Applications, Interesting Papers I

- Applications in travel time prediction, 2021 (https://arxiv.org/pdf/2108.11482.pdf)
- Someone has compiled graph-/GNN-relevant talks for NeurIPS 2023 at https://github.com/XiaoxinHe/neurips2023_learning_on_graphs

DiffPool

Figure 1 from Ying (2018) *Hierarchical Graph Representation Learning with Differentiable Pooling* [14]

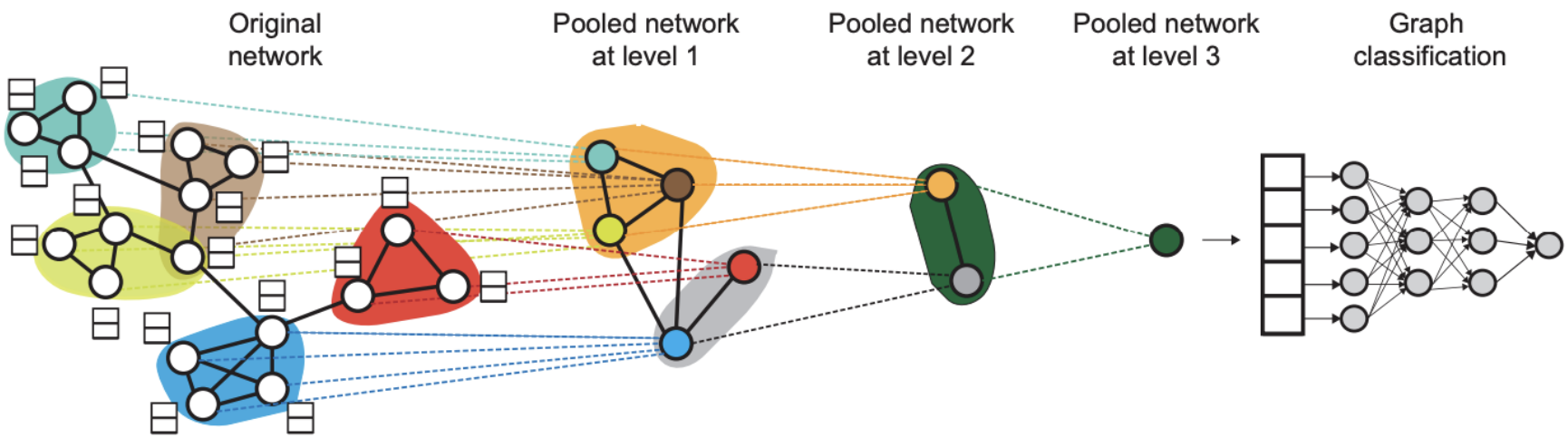
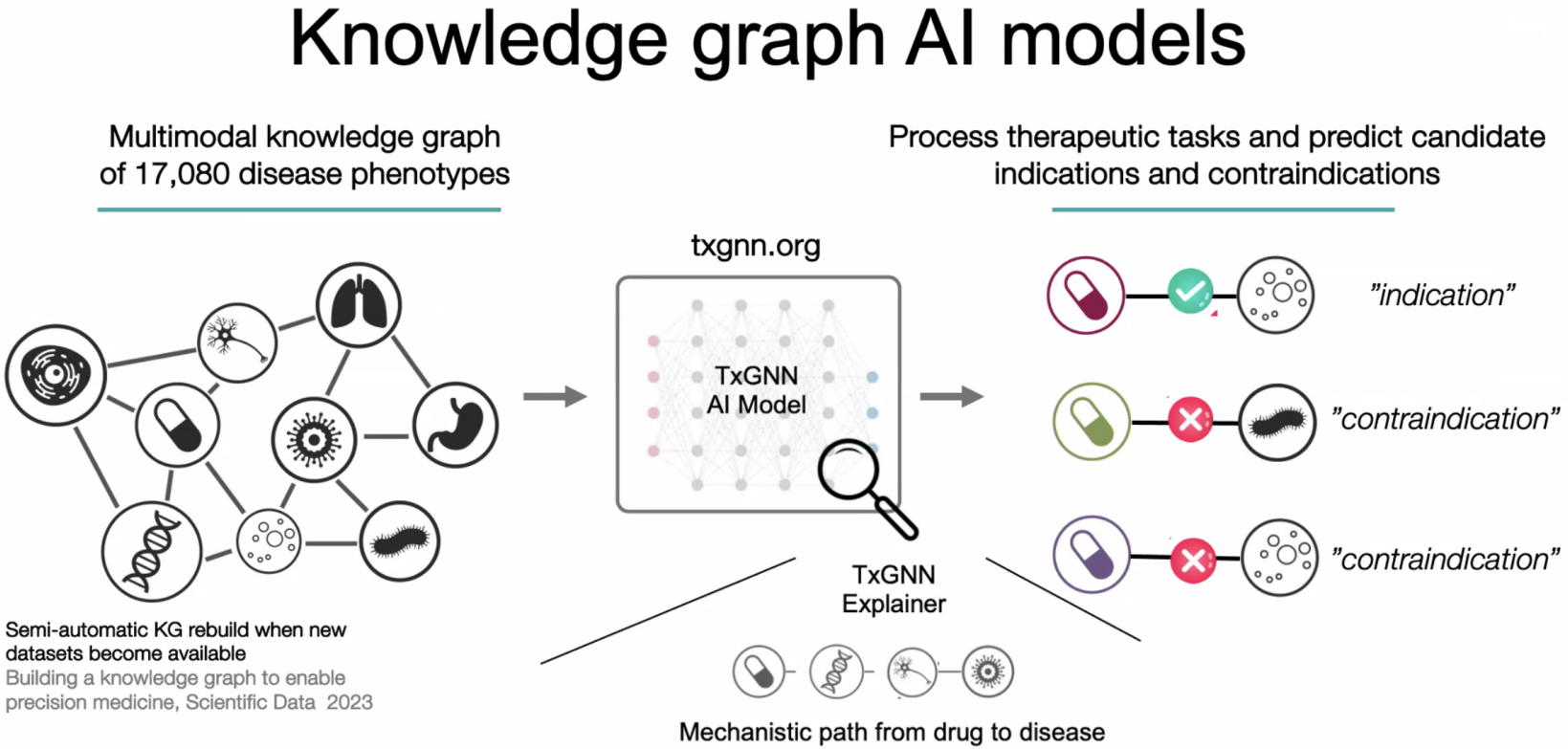


Figure 1: High-level illustration of our proposed method DIFFPOOL. At each hierarchical layer, we run a GNN model to obtain embeddings of nodes. We then use these learned embeddings to cluster nodes together and run another GNN layer on this coarsened graph. This whole process is repeated for L layers and we use the final output representation to classify the graph.

Downloadable Slides

KG AI Models

Full figure from McDermott et al. [10], cropped and presented in Introduction:



Structure-inducing pre-training, Nature Machine Intelligence 2023; Multimodal learning with graphs, Nature Machine Intelligence 2023; Zero-shot prediction of therapeutic use with geometric deep learning and clinician centered design, medRxiv, 2023

Downloadable Slides

Protein Data Bank File

```
HEADER    EXTRACELLULAR MATRIX                      22-JAN-98   1A3I
TITLE     X-RAY CRYSTALLOGRAPHIC DETERMINATION OF A COLLAGEN-LIKE
TITLE    2 PEPTIDE WITH THE REPEATING SEQUENCE (PRO-PRO-GLY)
...
EXPDTA    X-RAY DIFFRACTION
AUTHOR    R.Z.KRAMER,L.VITAGLIANO,J.BELLA,R.BERISIO,L.MAZZARELLA,
AUTHOR   2 B.BRODSKY,A.ZAGARI,H.M.BERMAN
...
REMARK 350 BIOMOLECULE: 1
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C
REMARK 350   BIOMT1   1  1.000000  0.000000  0.000000        0.00000
REMARK 350   BIOMT2   1  0.000000  1.000000  0.000000        0.00000
...
SEQRES   1 A    9  PRO PRO GLY PRO PRO GLY PRO PRO GLY
SEQRES   1 B    6  PRO PRO GLY PRO PRO GLY
SEQRES   1 C    6  PRO PRO GLY PRO PRO GLY
...
ATOM      1  N   PRO A   1       8.316  21.206  21.530  1.00 17.44           N
ATOM      2  CA  PRO A   1       7.608  20.729  20.336  1.00 17.44           C
ATOM      3  C   PRO A   1       8.487  20.707  19.092  1.00 17.44           C
ATOM      4  O   PRO A   1       9.466  21.457  19.005  1.00 17.44           O
ATOM      5  CB  PRO A   1       6.460  21.723  20.211  1.00 22.26           C
...
HETATM  130  C   ACY   401       3.682  22.541  11.236  1.00 21.19           C
HETATM  131  O   ACY   401       2.807  23.097  10.553  1.00 21.19           O
HETATM  132  OXT ACY   401       4.306  23.101  12.291  1.00 21.19           O
...
```