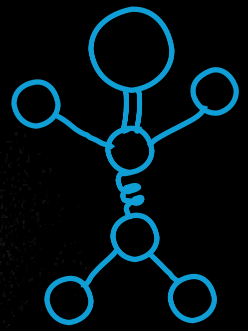
Learning a Molecule with GNNs



An interactive understanding of expressive GNNs for molecular representation, and how to scale them to infinity 🚀

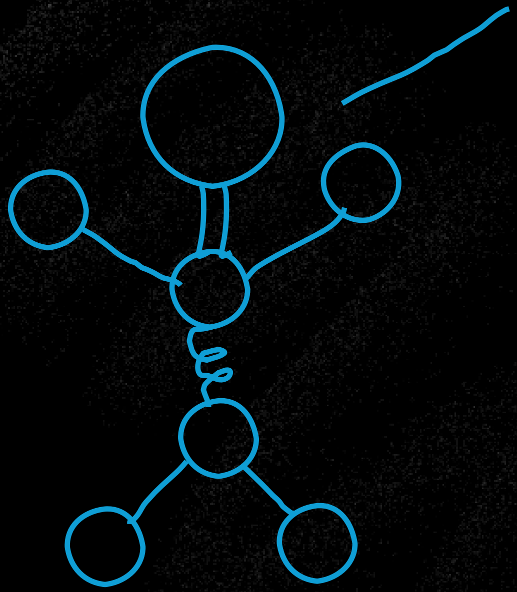
A presentation by Dominique Beaini

Research lead at Valence Labs

Adjunct Prof at Université de Montreal

Associate Industry member at Mila – Quebec AI institute

Meeting Graphy 👋



Hello everyone 👋! I'm Dom's assistant for today!

Let's visit the molecular graph world together!

We'll first learn what are graphs and how to manipulate them

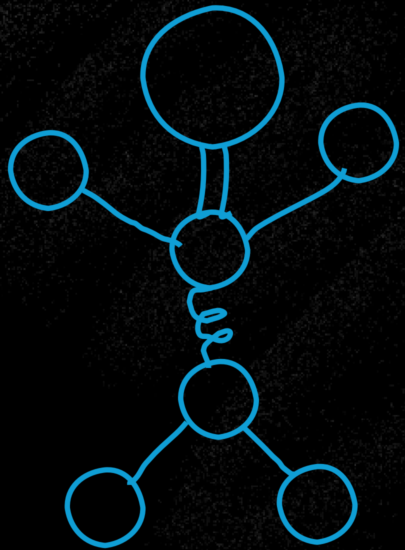
Then we'll look into standard GNNs graph neural networks

And how to build more expressive GNNs for molecules

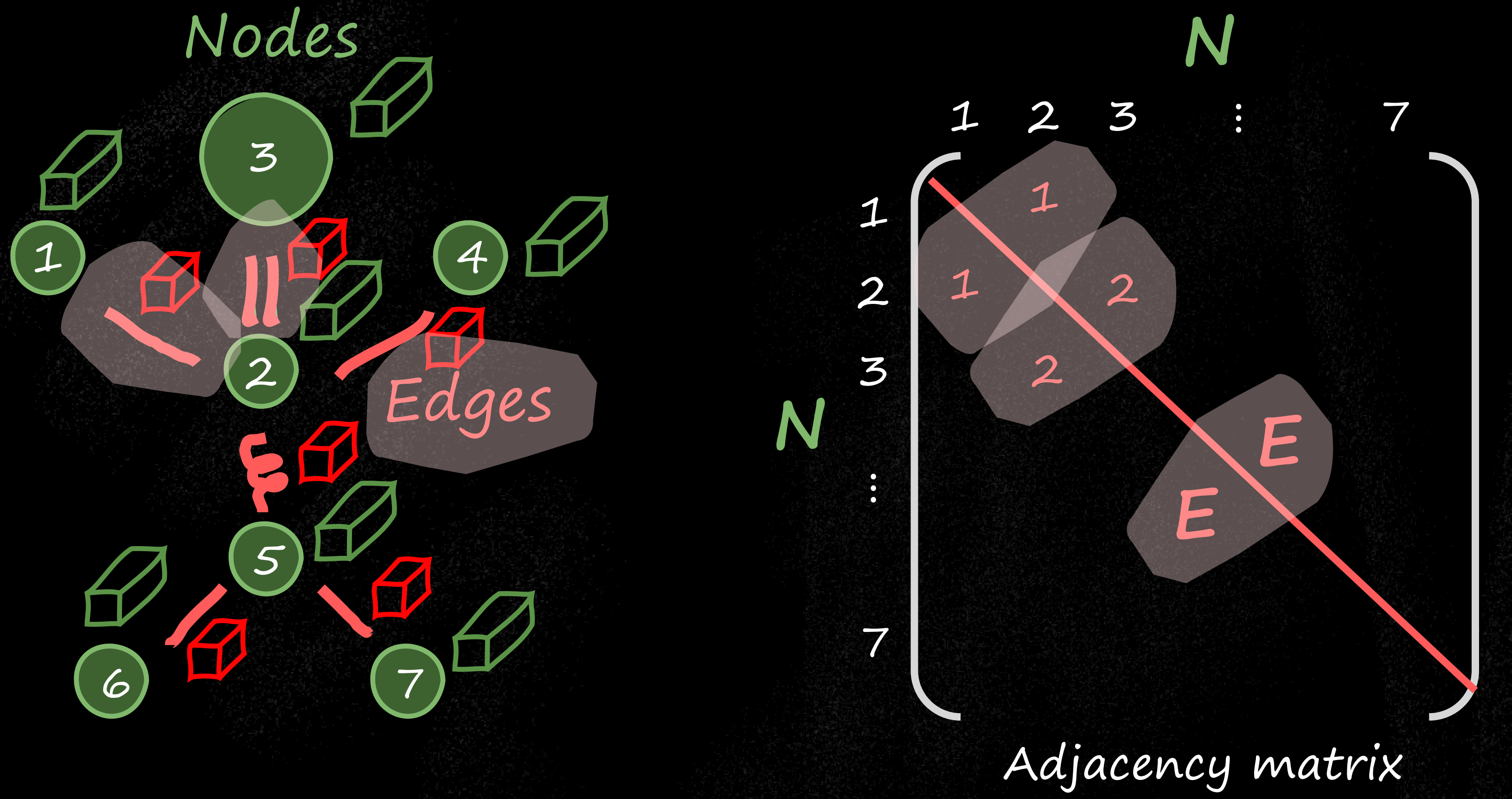
Then, we will scale a Graph Transformer together

Finally, we'll work through some applications

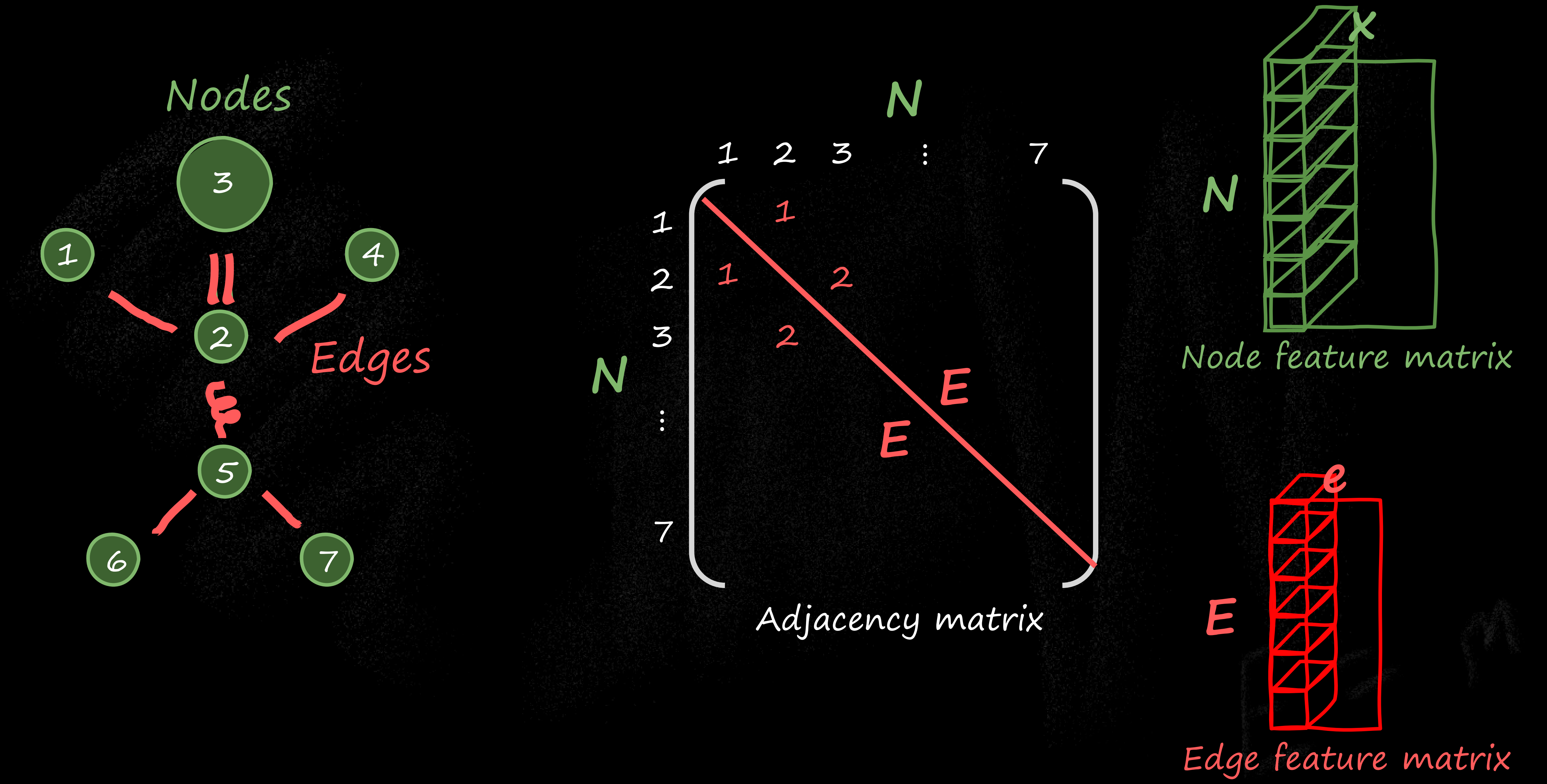
Anatomy of Graphy 🦴



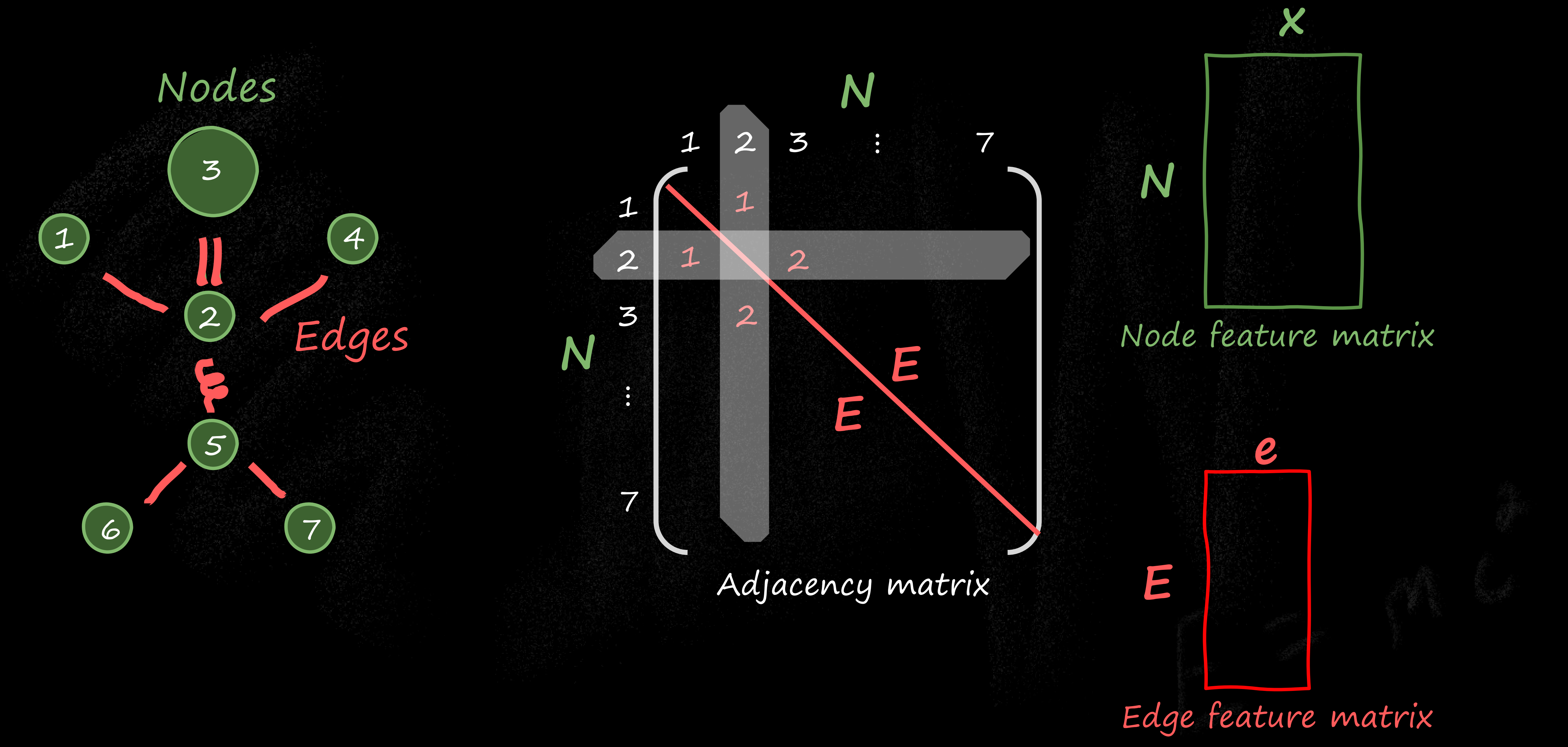
Anatomy of Graphy 🦴



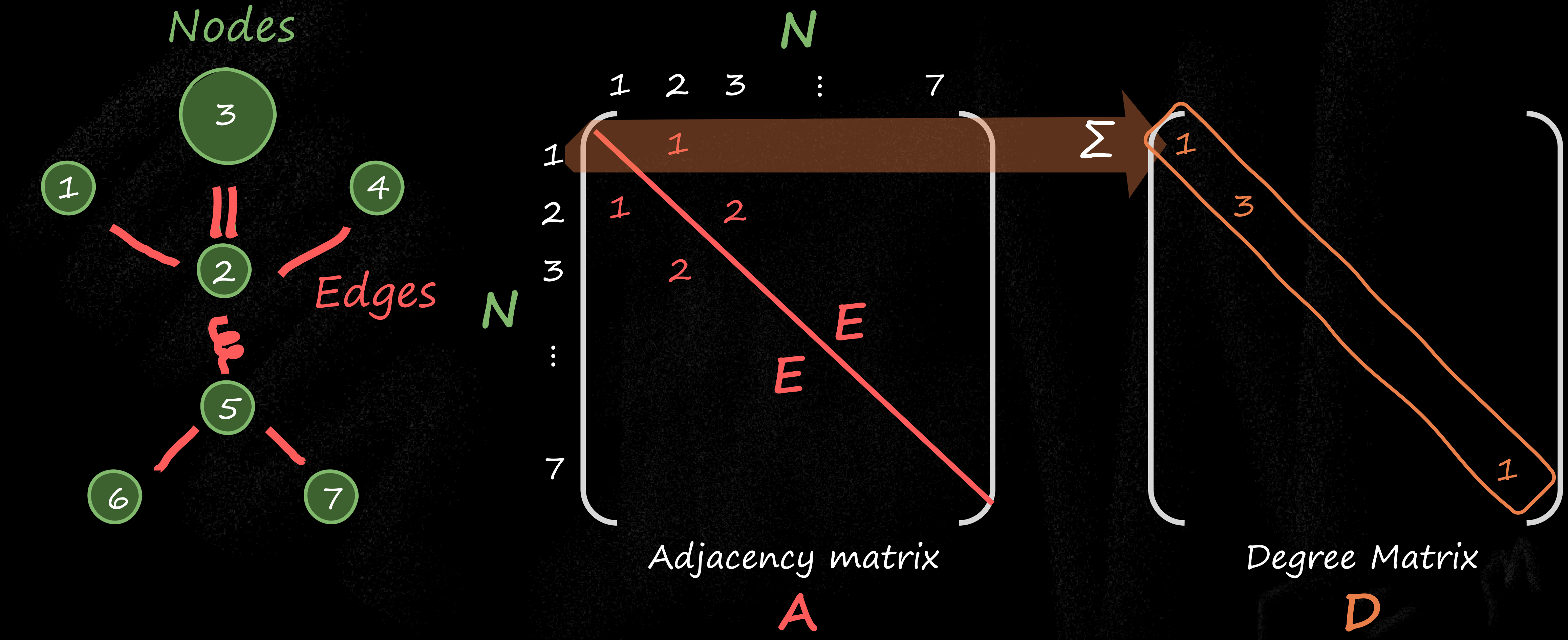
Anatomy of Graphy 🦴



Permutation invariance

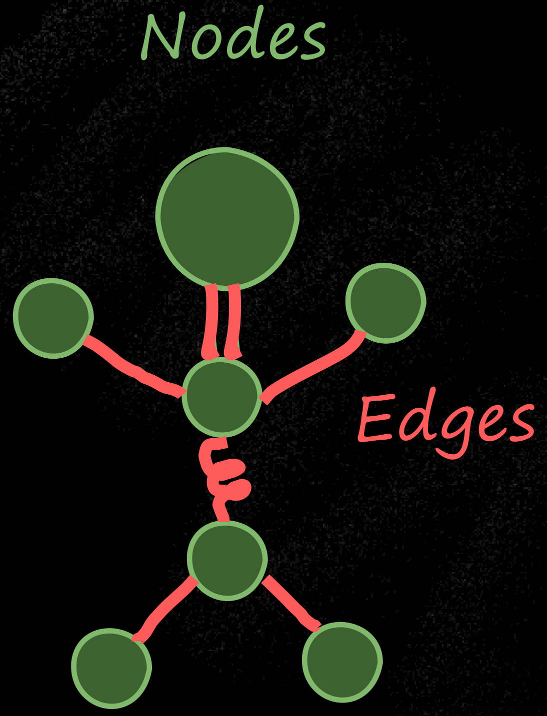


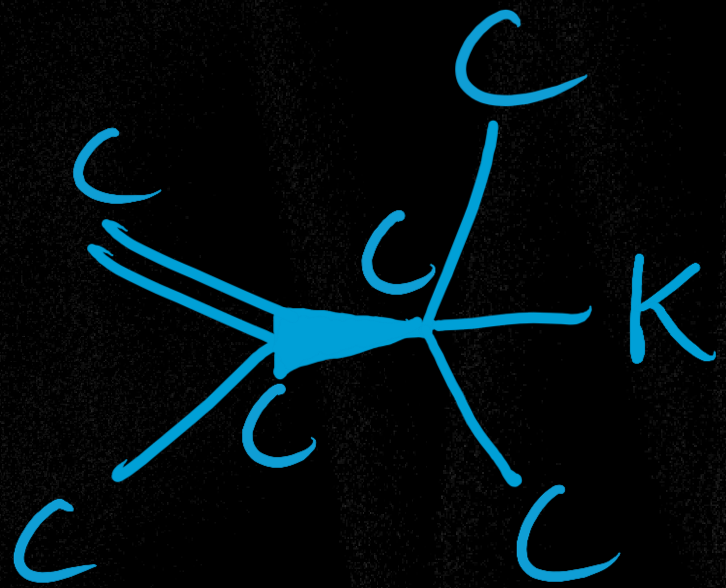
Laplacian matrix



$$L = D - A$$

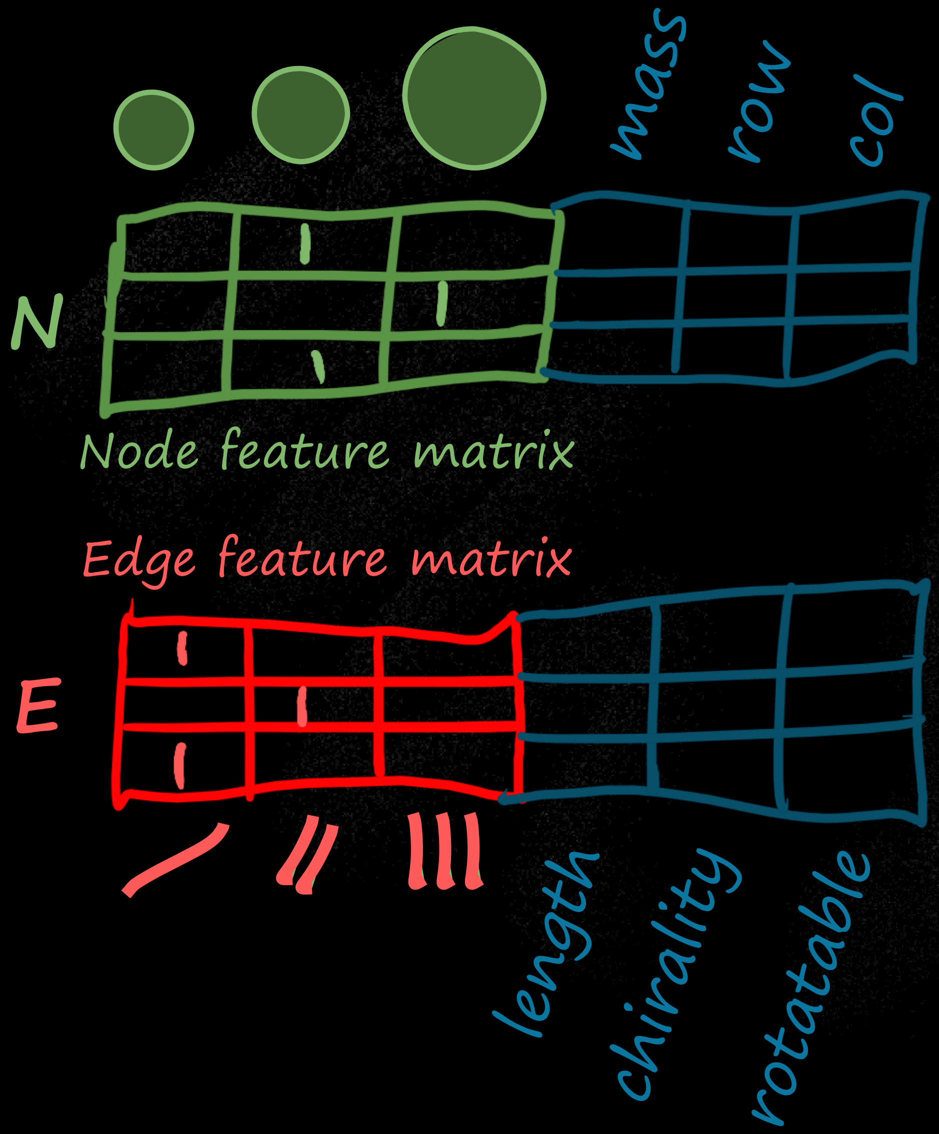
Anatomy of a molecule 💀🦴

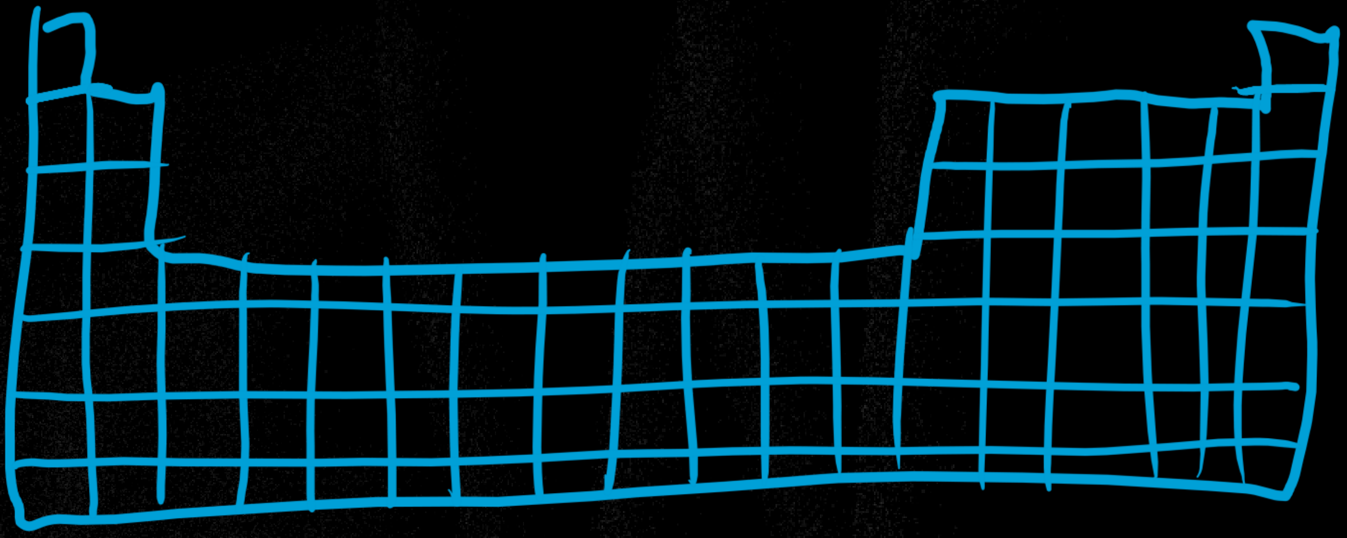




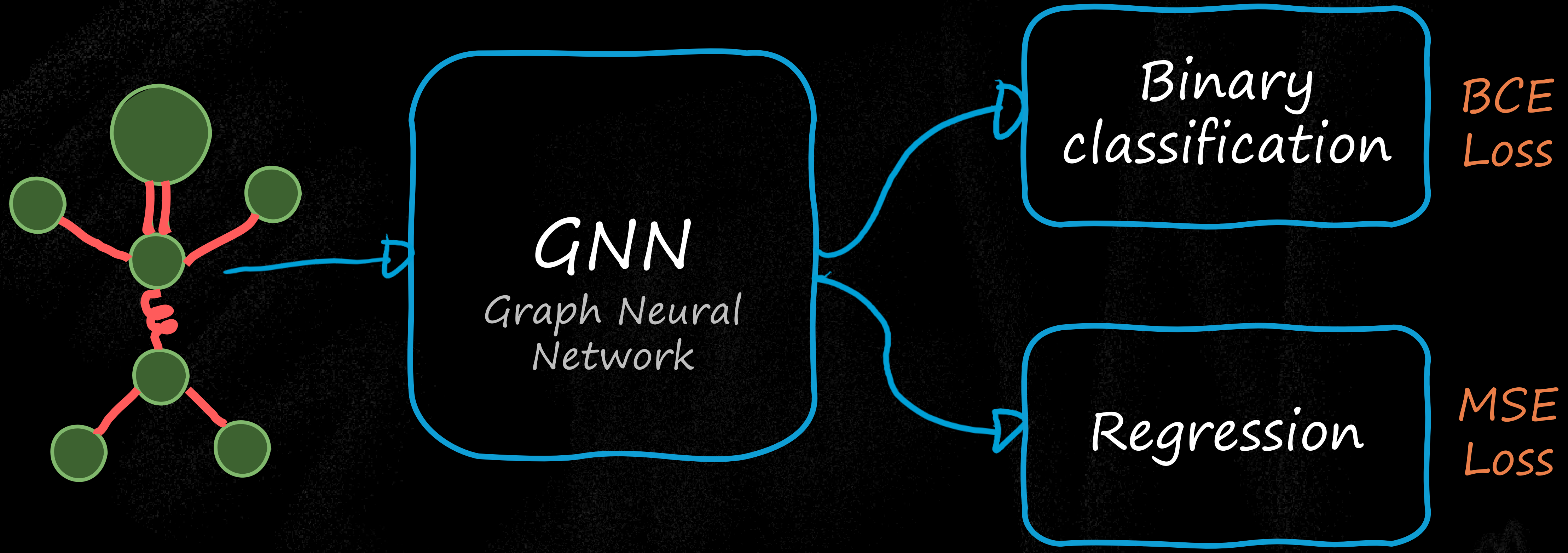
Molecules are <u>**defined**</u> as graphs!

Anatomy of atoms and bonds

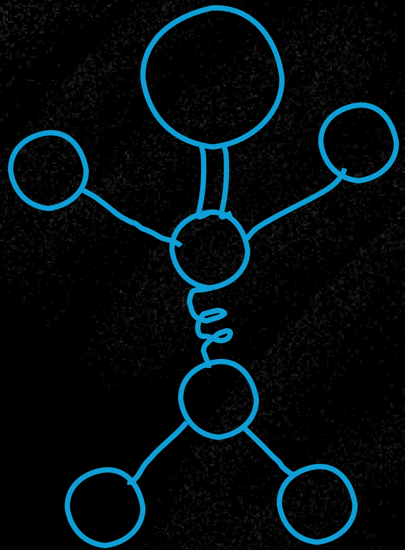




Molecular Property Prediction

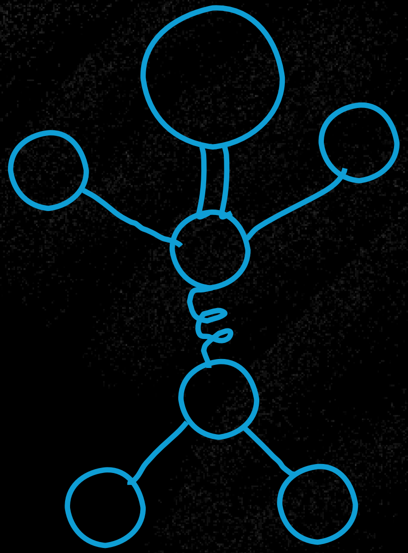


Designing a GNN 💡



- How do we deal with permutation invariance?
- And the varying #nodes / neighbors?
- And the isotropy or lack of direction?
- And the expressivity?

Designing a GNN 💡



permutation invariance

- Apply MLPs on each node
- Pass the features on the edges

varying #nodes

- Aggregation (mean/max/sum)
- Pooling (mean/max/sum)

lack of direction

- ~~Ignore that it is a problem 🤐~~

expressivity

- ~~Make some weird proofs~~

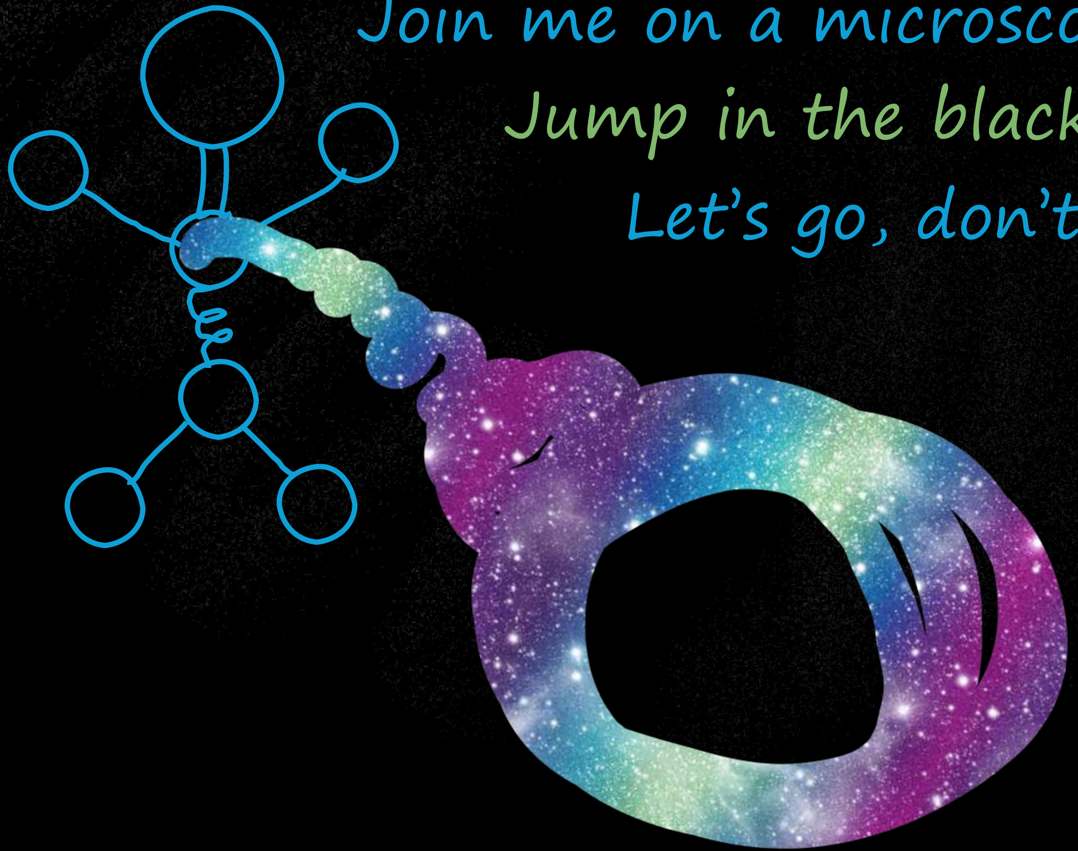
Don't worry, I'll show you a better way

A Microscopic Adventure 🔬

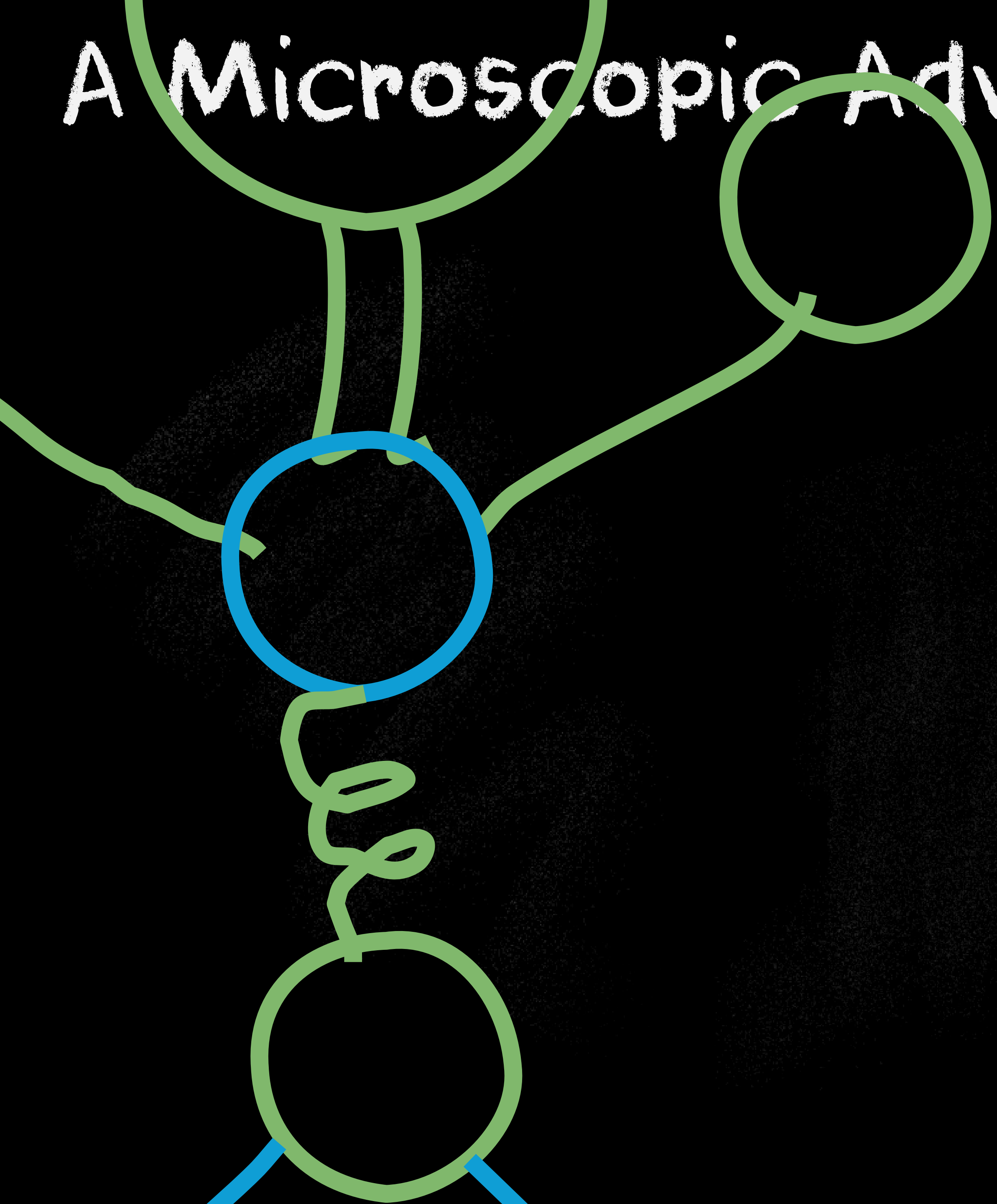
Join me on a microscopic adventure!

Jump in the black hole to see the perspective of a node

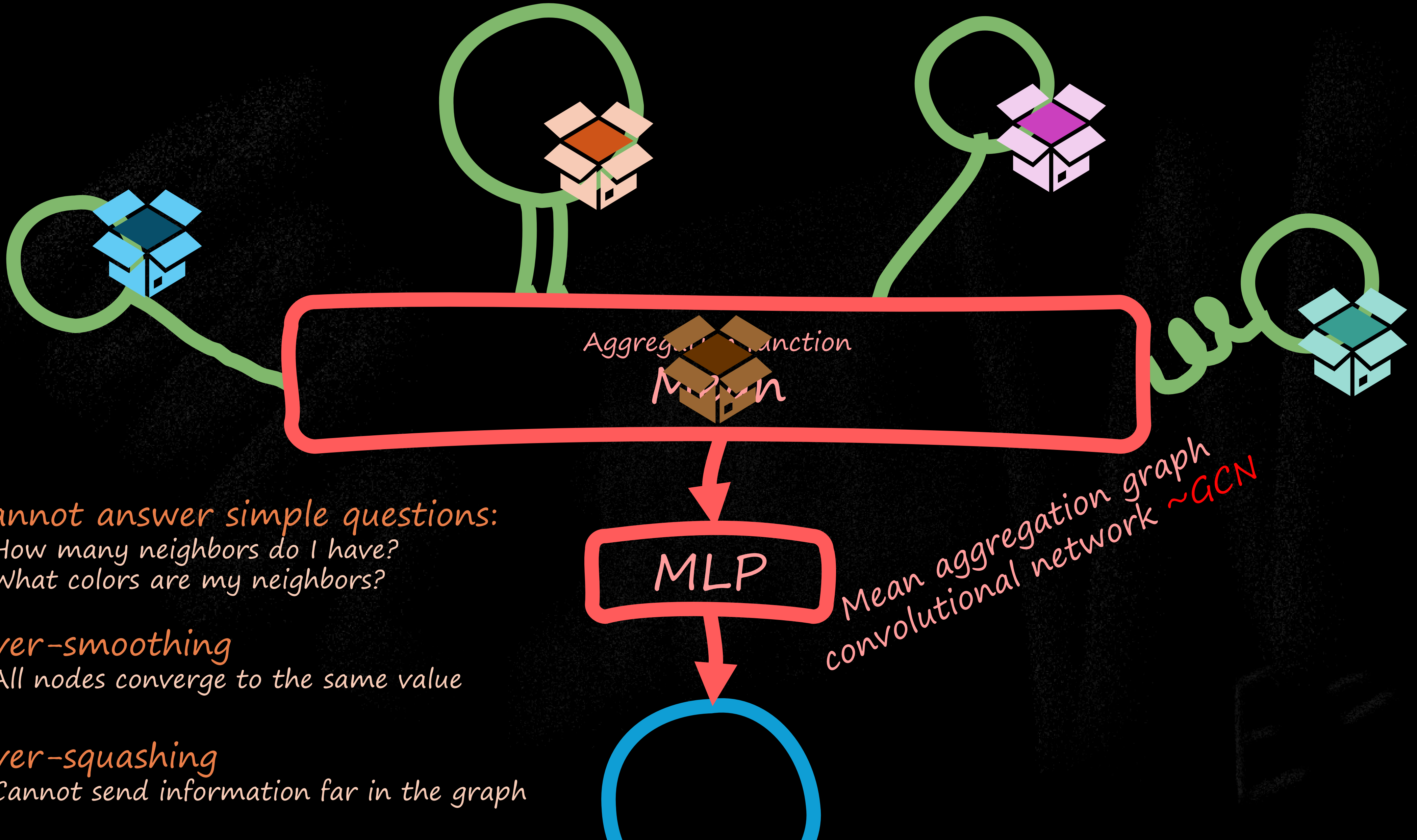
Let's go, don't be afraid!



A Microscopic Adventure 🔬



Mean Aggregation Conv 💩



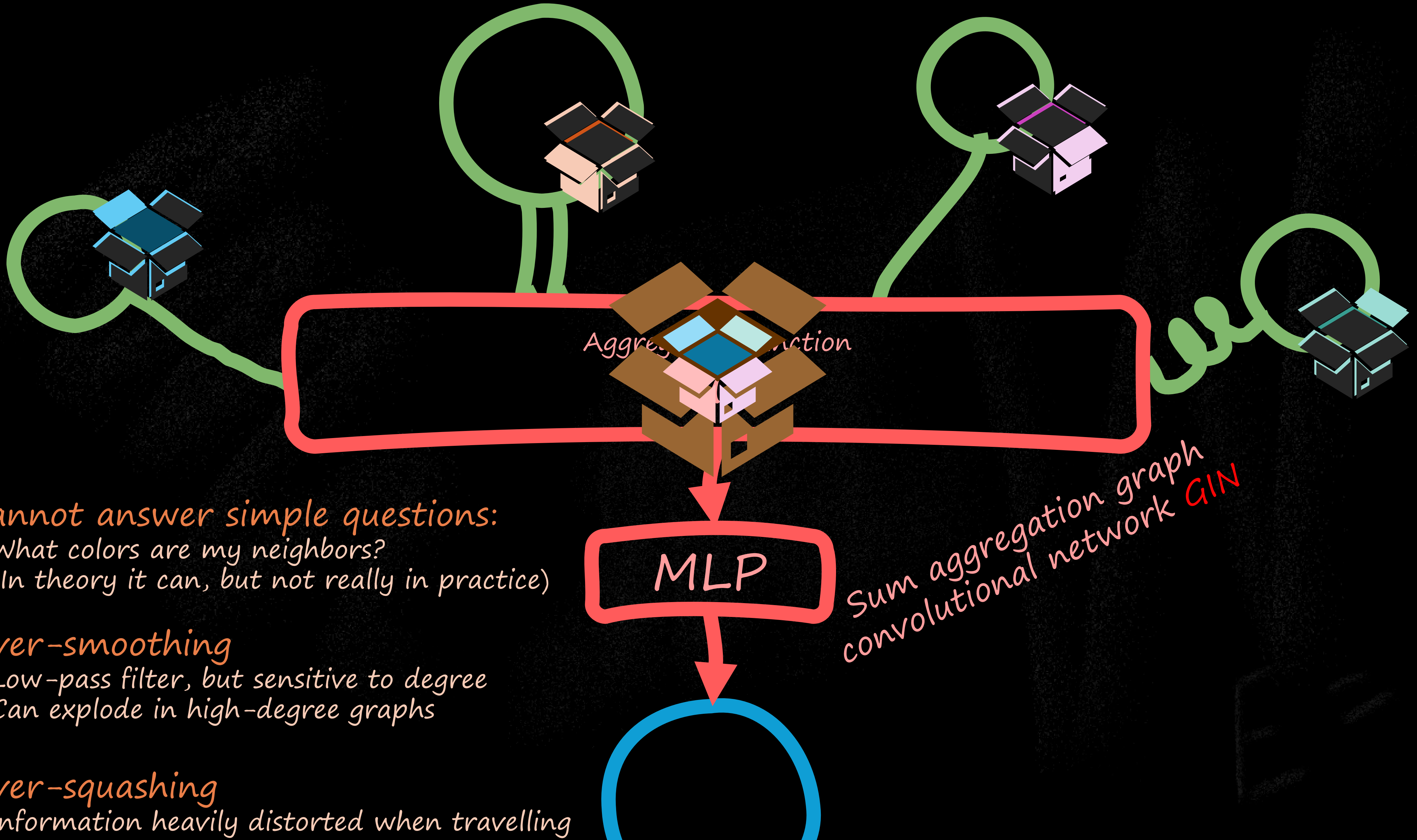
Mean aggregation graph convolutional network ~GCN

Cannot answer simple questions:
- How many neighbors do I have?
- What colors are my neighbors?

Over-smoothing
- All nodes converge to the same value

Over-squashing
- Cannot send information far in the graph

Sum Aggregation Conv 🧨

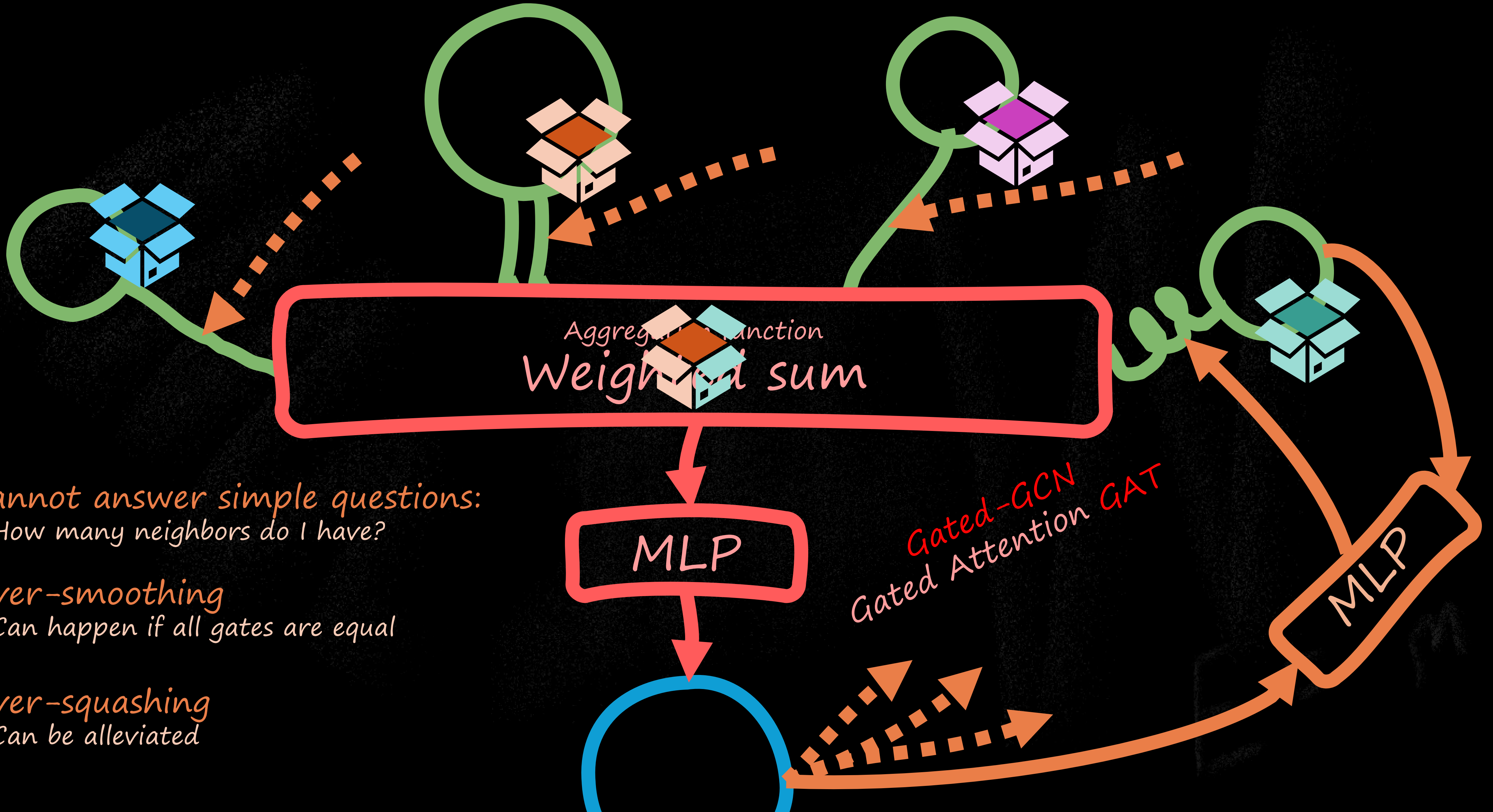


Cannot answer simple questions:
What colors are my neighbors?
(In theory it can, but not really in practice)

Over-smoothing
Low-pass filter, but sensitive to degree
Can explode in high-degree graphs

Over-squashing
Information heavily distorted when travelling

Gating 🚪

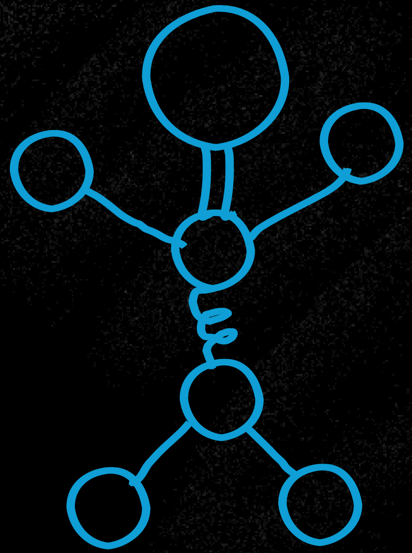


Cannot answer simple questions:
How many neighbors do I have?

Over-smoothing
Can happen if all gates are equal

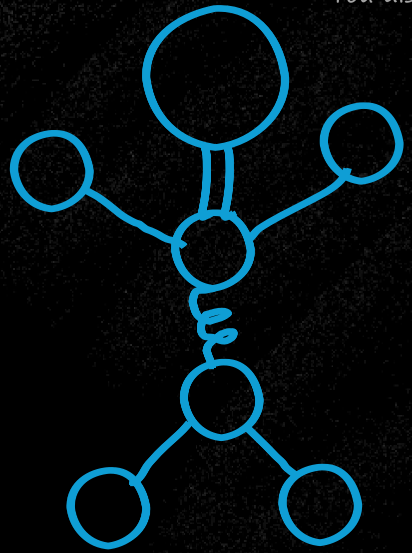
Over-squashing
Can be alleviated

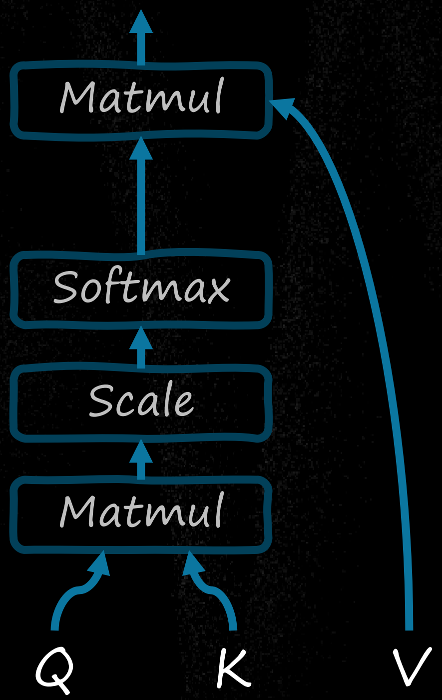
Attention is all you need



Attention is all you need

- ATTENTION IS ALL YOU NEED
  - Certain conditions apply
    - Read the fineprints for more details
      - You also need good positional and structural encodings, ideally a biased attention, lots and lots of long-range data, …

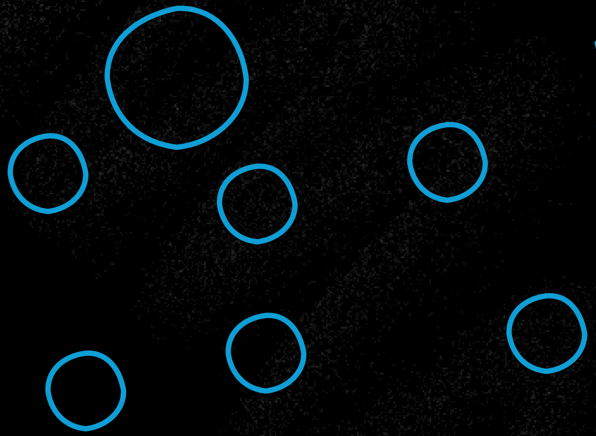




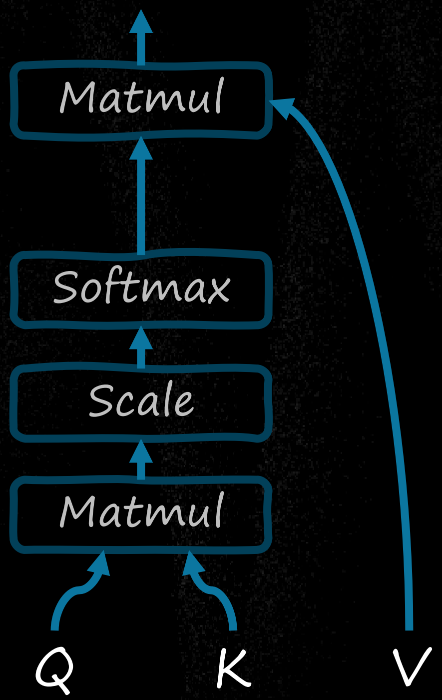
Transformer

Attention is all you need

- **ATTENTION IS ALL YOU NEED**
  - *Certain conditions apply*
    - *Read the fineprints for more details*
      - You also need good positional and structural encodings, ideally a biased attention, lots and lots of long-range data, …



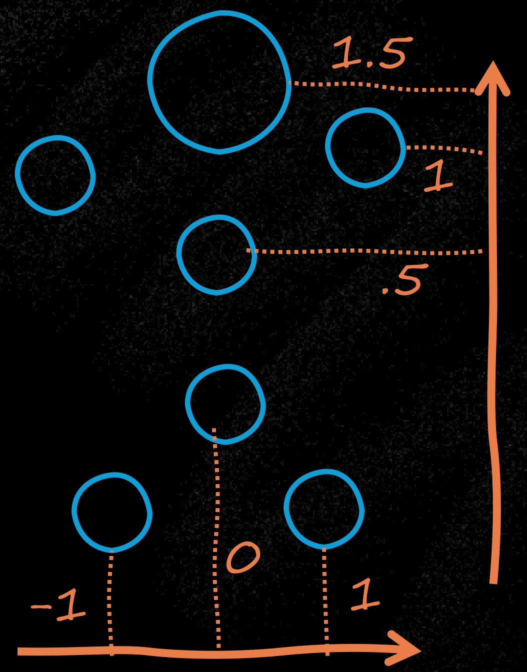
What's happening to me!!!
I'm being permuted!

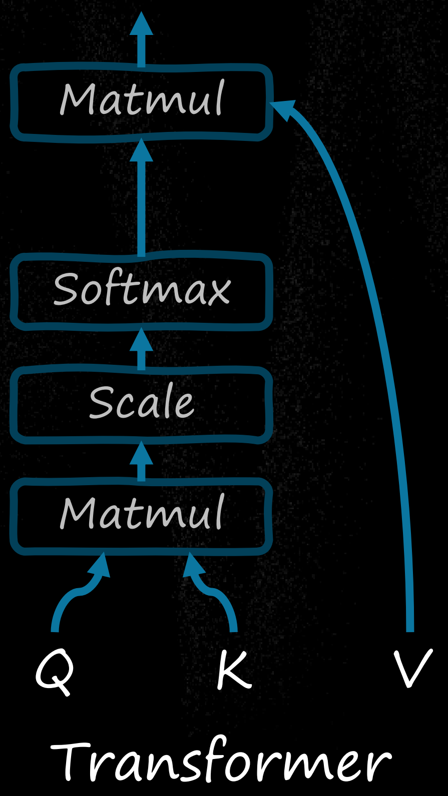


Transformer

Fineprints of Attention - Position

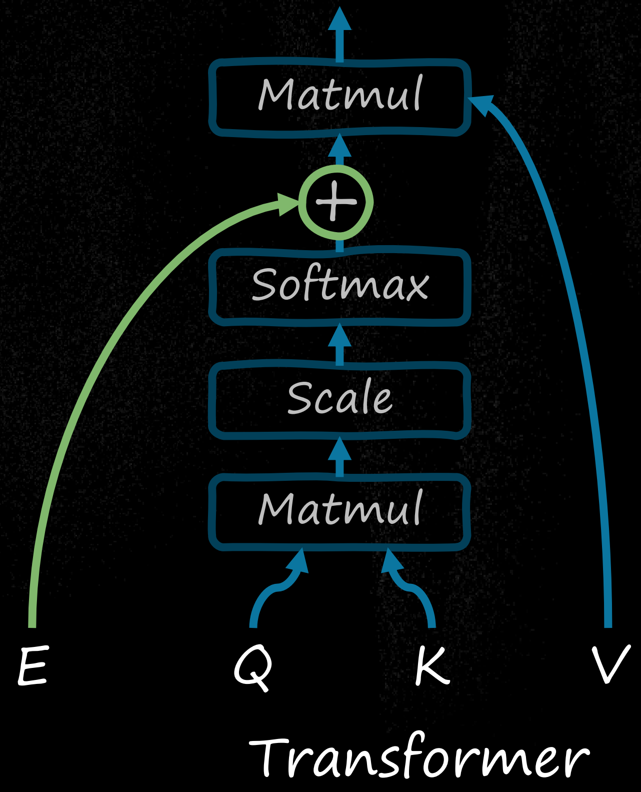
- positional and structural encodings
- biased attention
- lots and lots of long-range data





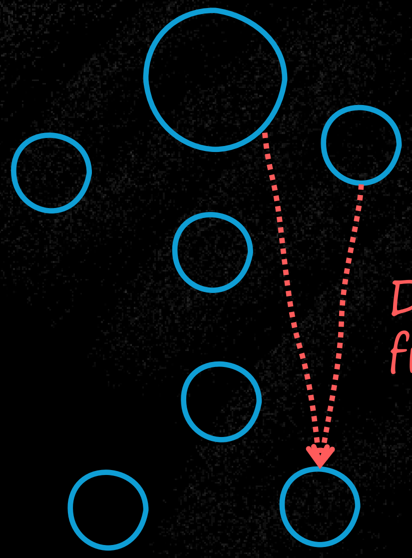
Fineprints of Attention – Bias

- positional and structural encodings
- biased attention
- lots and lots of long-range data

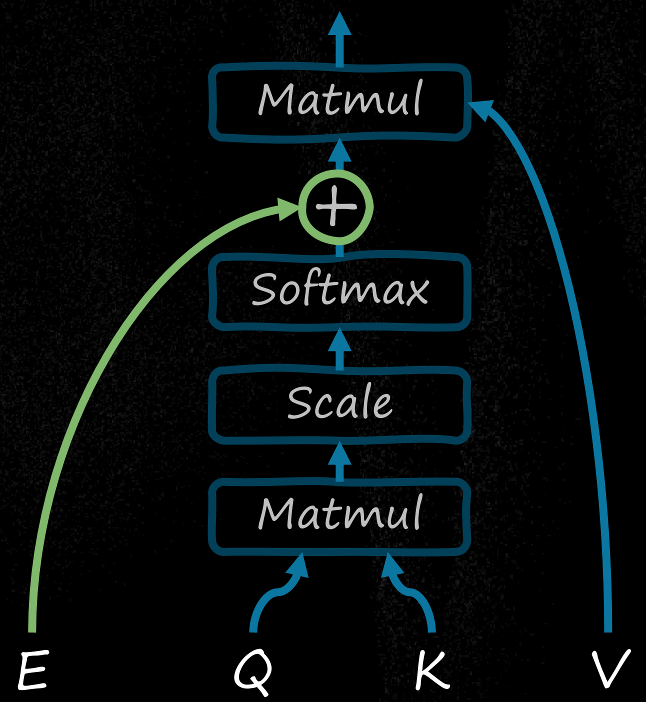


Fineprints of Attention – Data

- positional and structural encodings
- biased attention
- lots and lots of long-range data

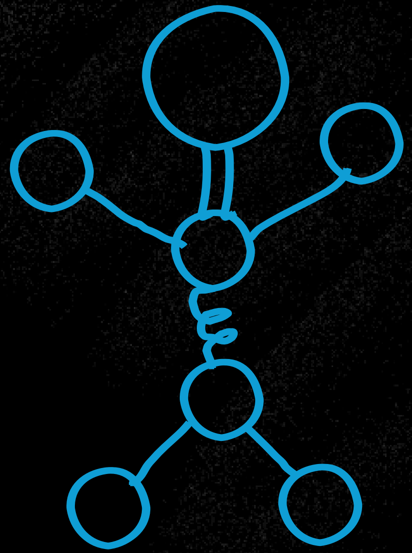


Do I need to aggregate from distant neighbors?



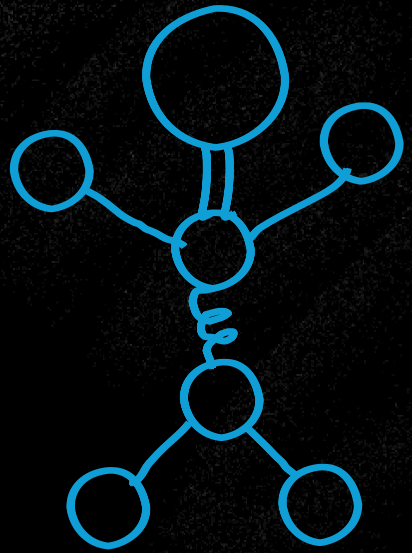
Transformer

Attention – Position and Structure



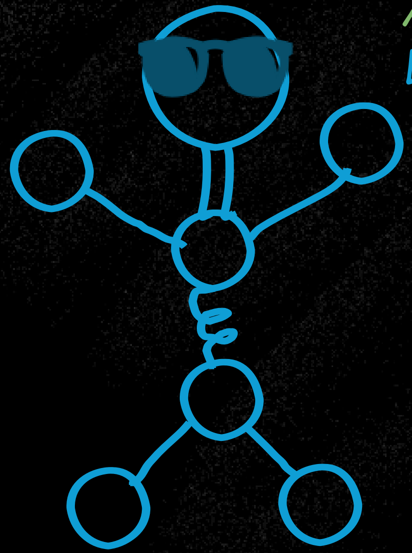
- WL Expressivity – why simple GNNs are not enough
- Positional encodings via eigenvectors
- Structural encodings via random walks
- Relative positions via distances and heat kernels

WL Expressivity

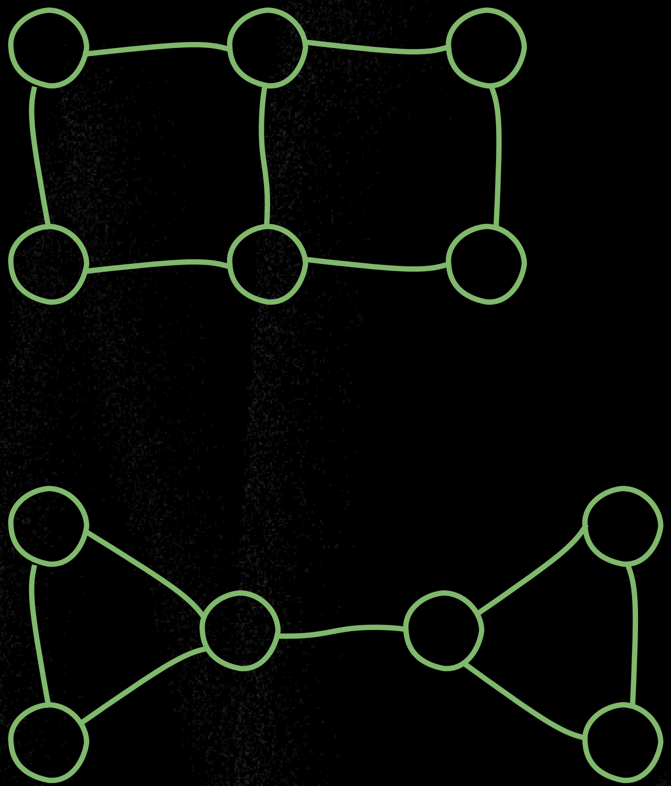


Weisfeiler-Lehman

WL Expressivity

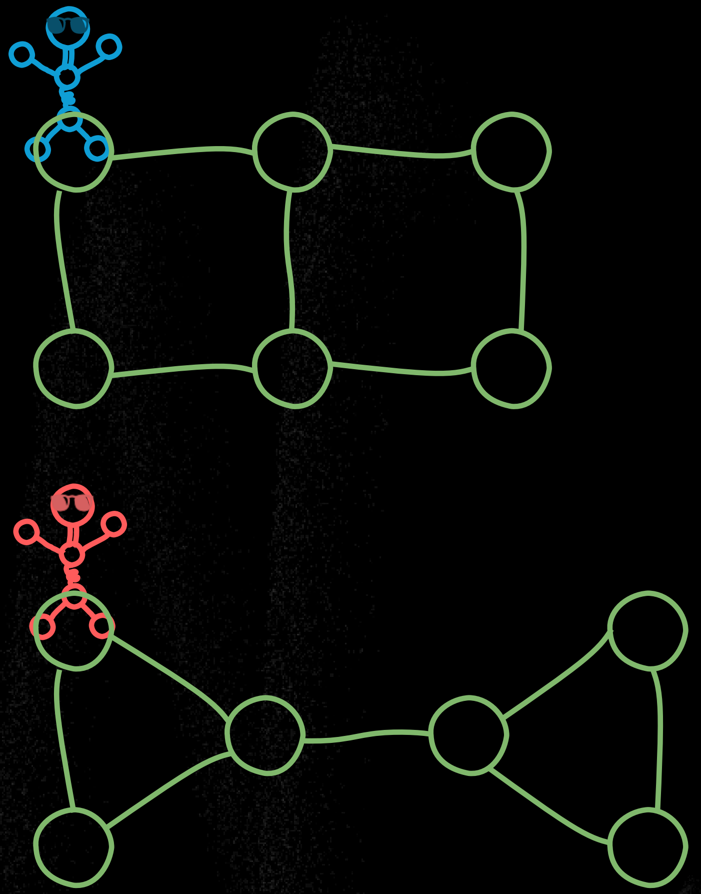


Let's play a game!
Are these graphs the same?
Let's shrink again and count the neighbors

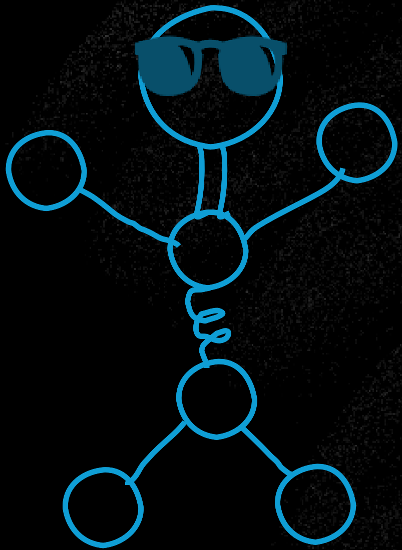


WL Expressivity 🕶

2	2
3	3
2	2
2	2
3	3
2	2
3	3
3	3
2	2



WL Expressivity



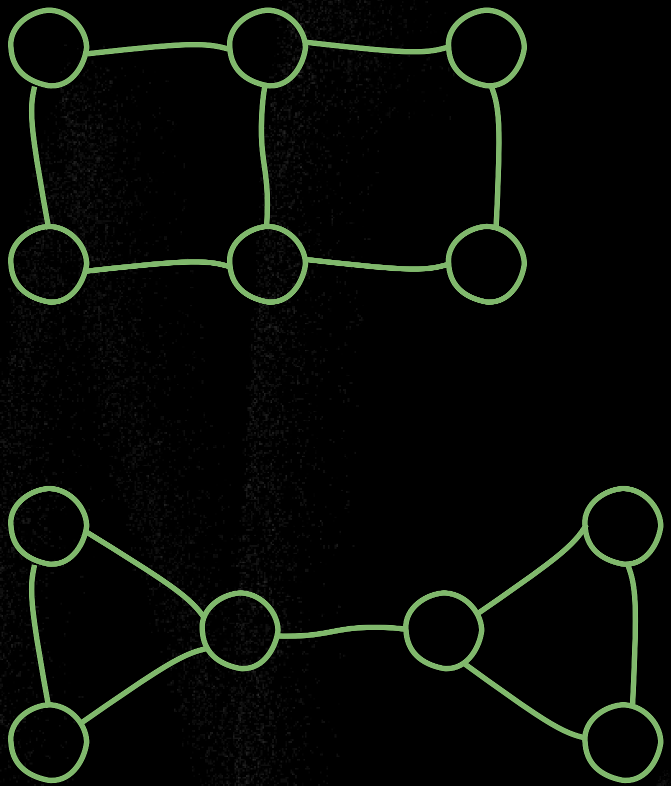
They're the same!
Wait... That's not right.

There are things we cannot see from inside because we do not have position or direction!

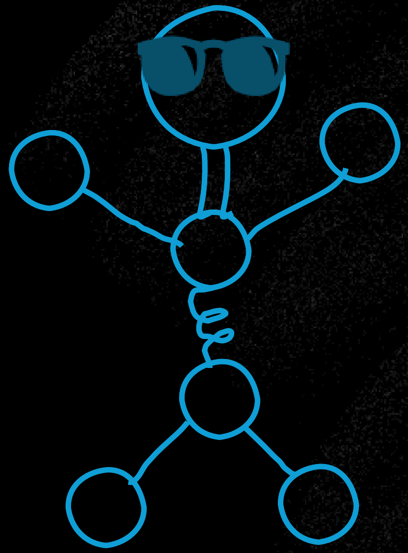
I know what you're thinking.
Graphs have no direction

We'll circle back on that...

2	2
3	3
2	2
2	2
3	3
2	2
3	3
3	3
2	2



Higher order features

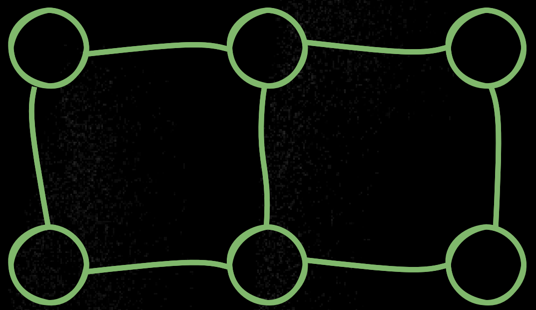


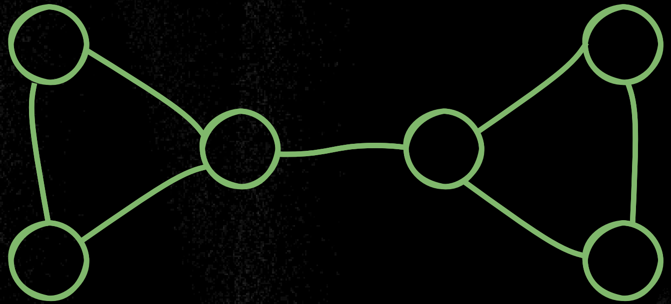
Can we find some higher-order features?

Features that are permutation invariant but that can be computed?

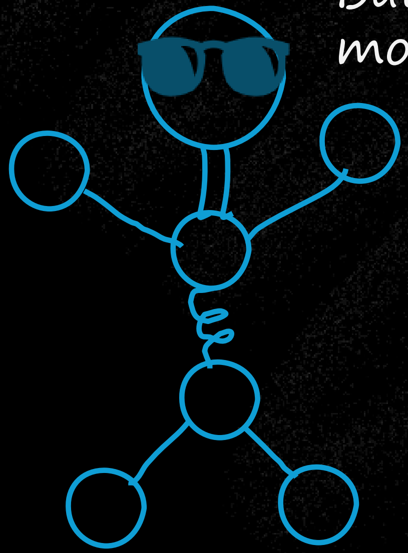
Perhaps features inspired by higher order WL-tests by walking around the graph?

Let's look at random walks and motif detection





Structural encodings



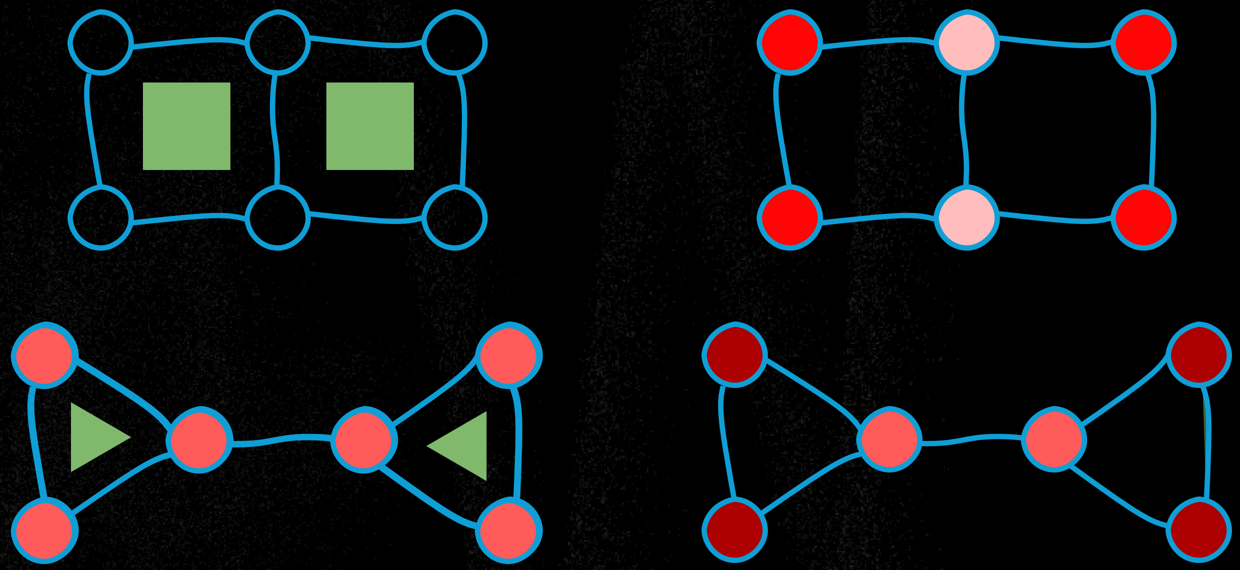
These are nice local encodings!

But is there anything more global?

Structure encoding by identifying motifs and random walks

Random walk 3-step

Random walk 4-step

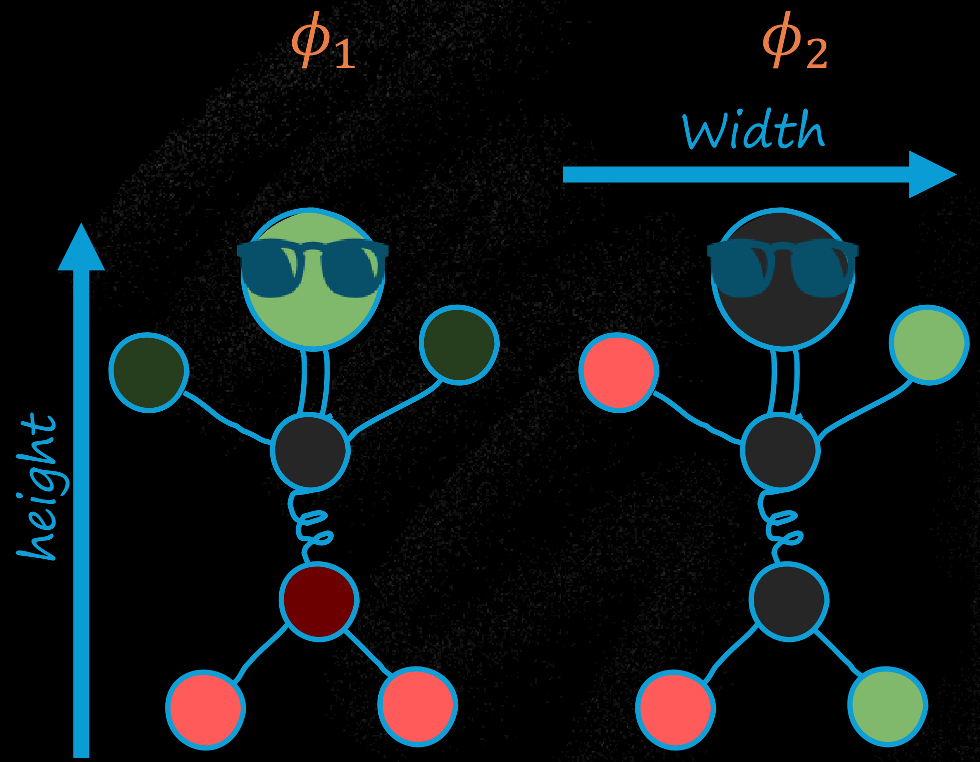


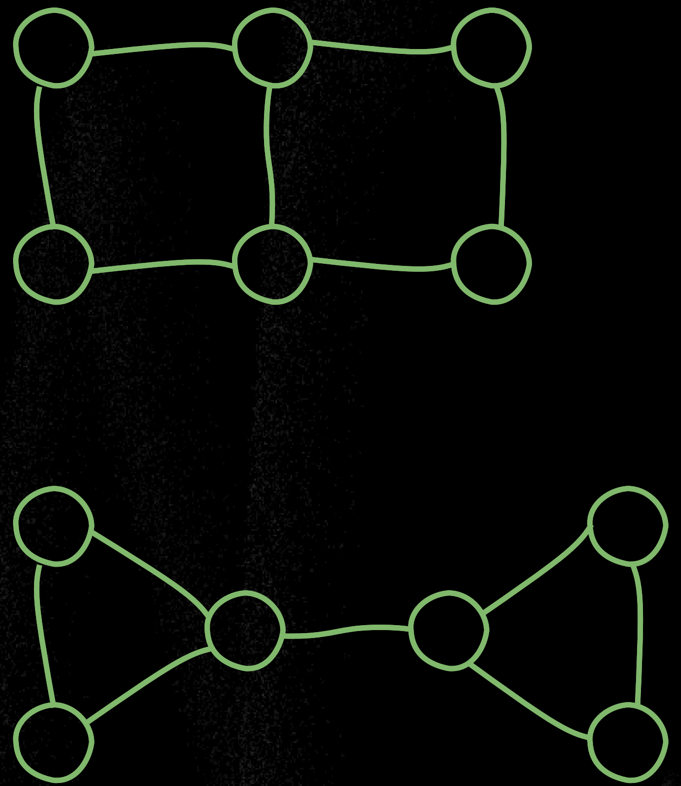
Now we can distinguish the graphs and nodes!
We can concatenate them to node features
We can bias the connectivity of the message passing
We are again more expressive

Positional encodings

Low-frequency eigenvectors of the Laplacian

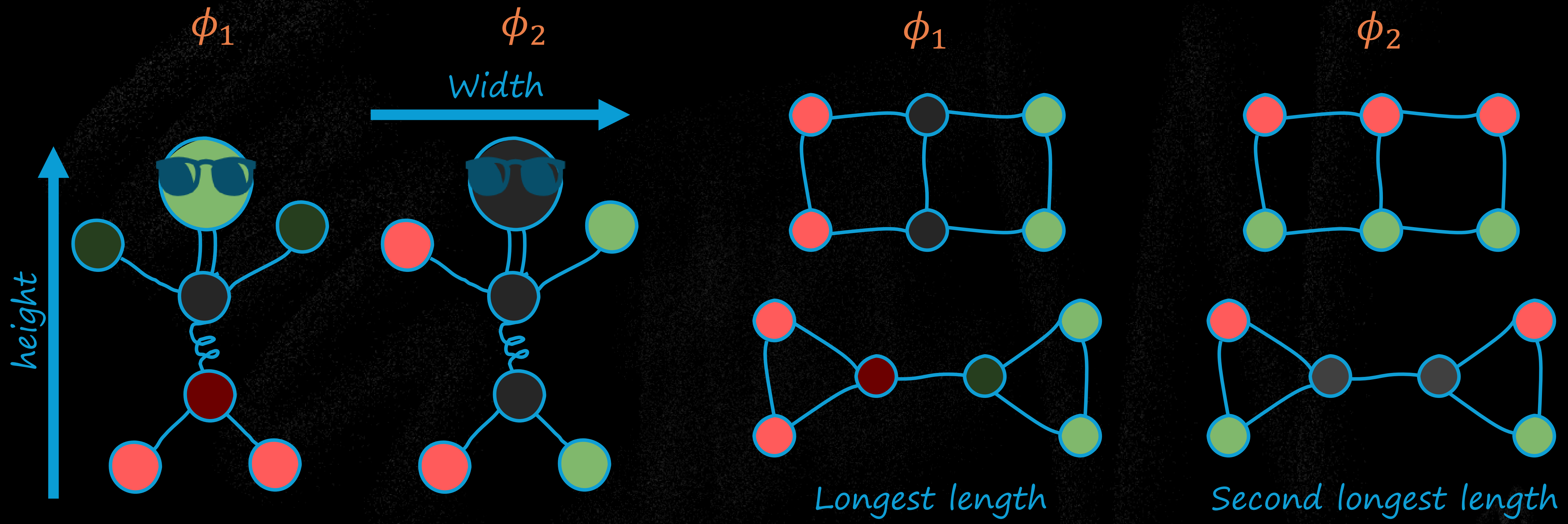
(lowest non-0 eigenvalue)





Positional encodings

Low-frequency eigenvectors of the Laplacian
(lowest non-0 eigenvalue)

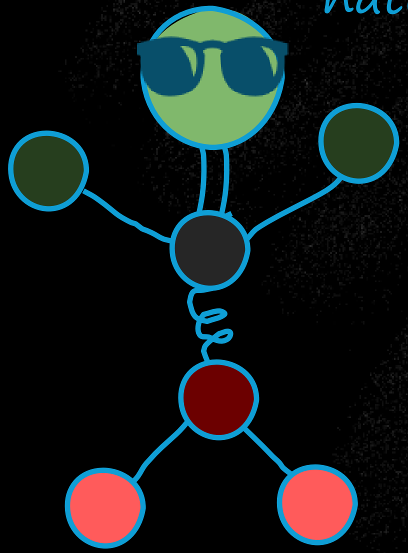


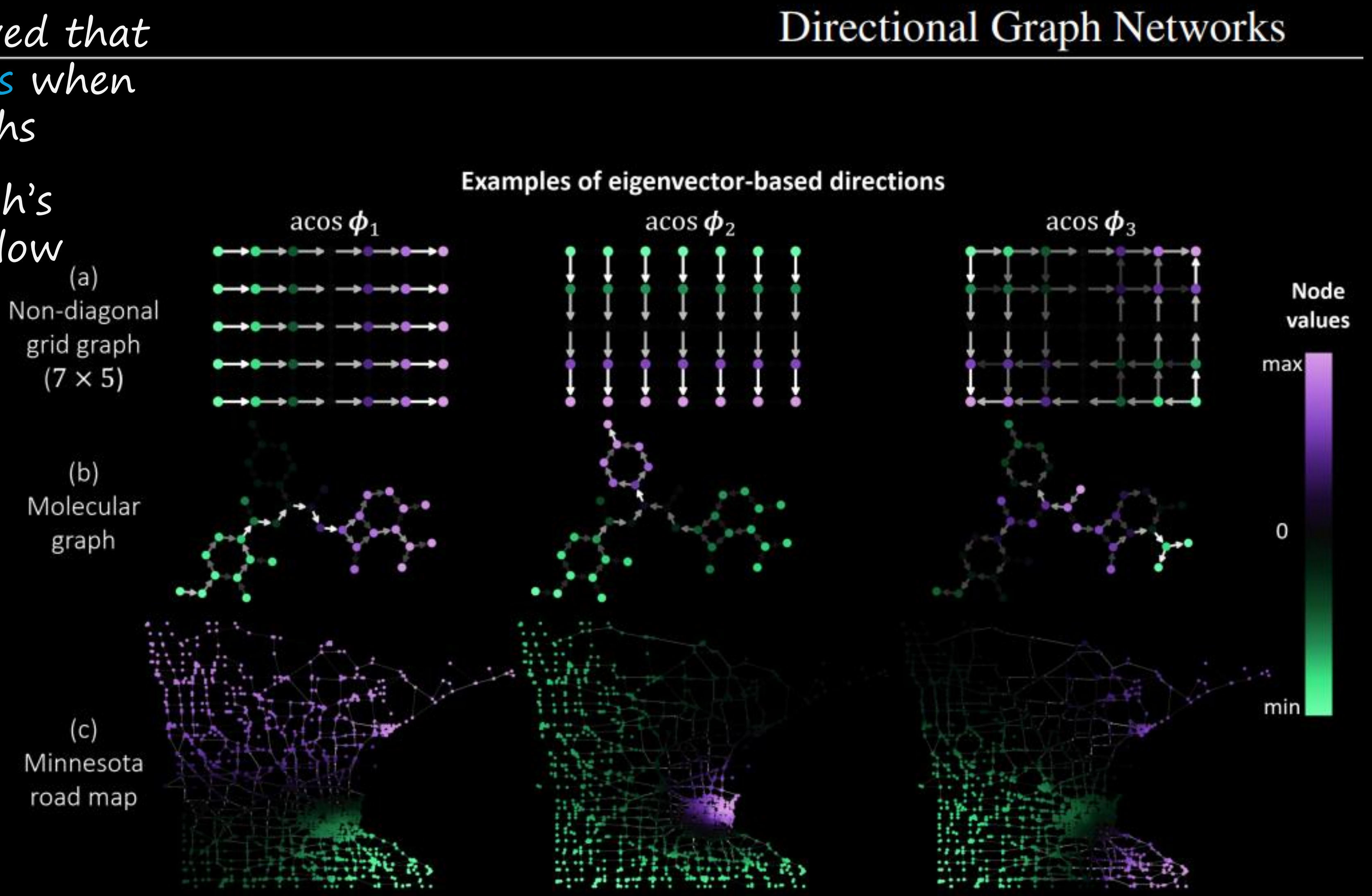
Now we can distinguish the graphs and nodes!
We can concatenate them to node features
We can bias the direction of the message passing
We are more expressive

Low-frequency eigenvectors

The DGN work showed that they generalize CNNs when applied to grid graphs

They retrieve a graph's natural directional flow

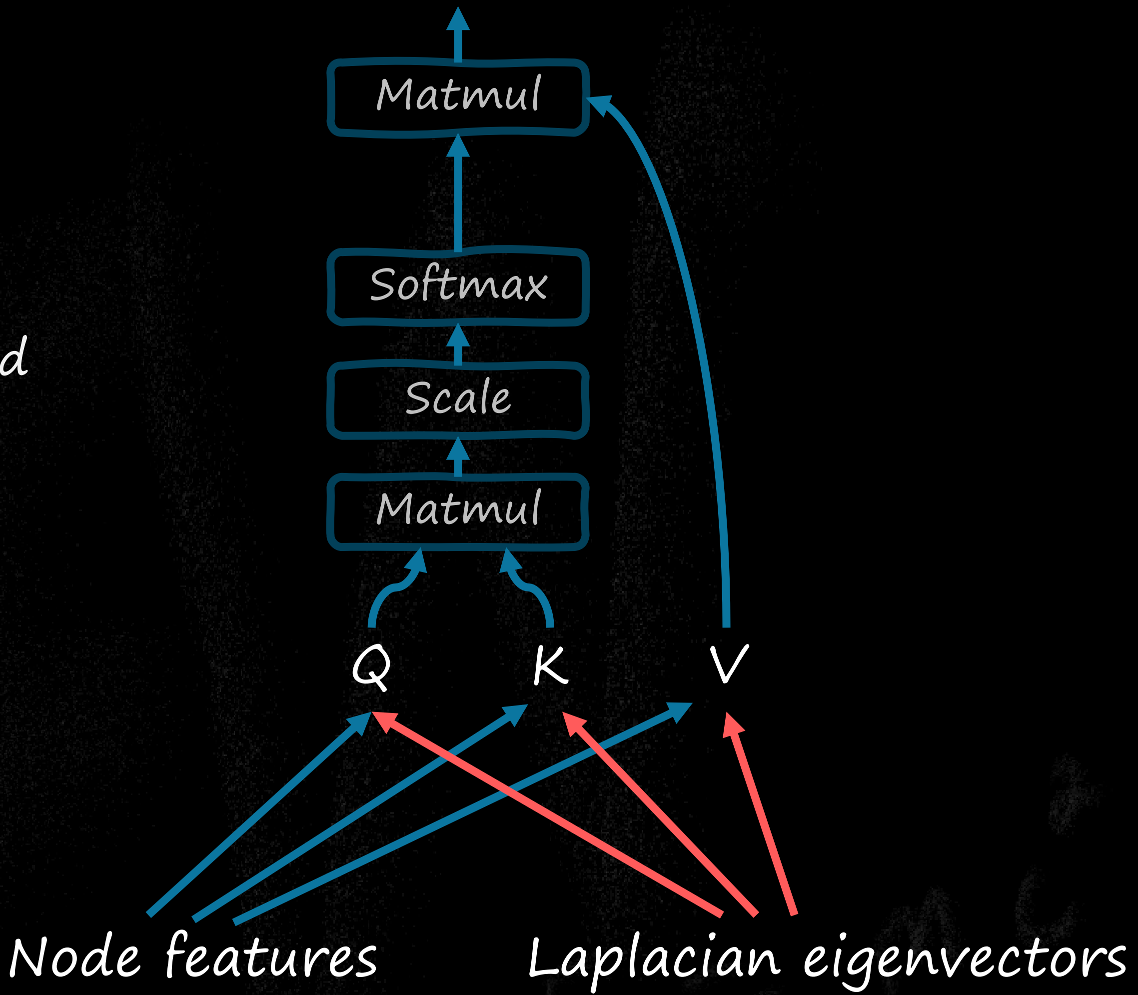




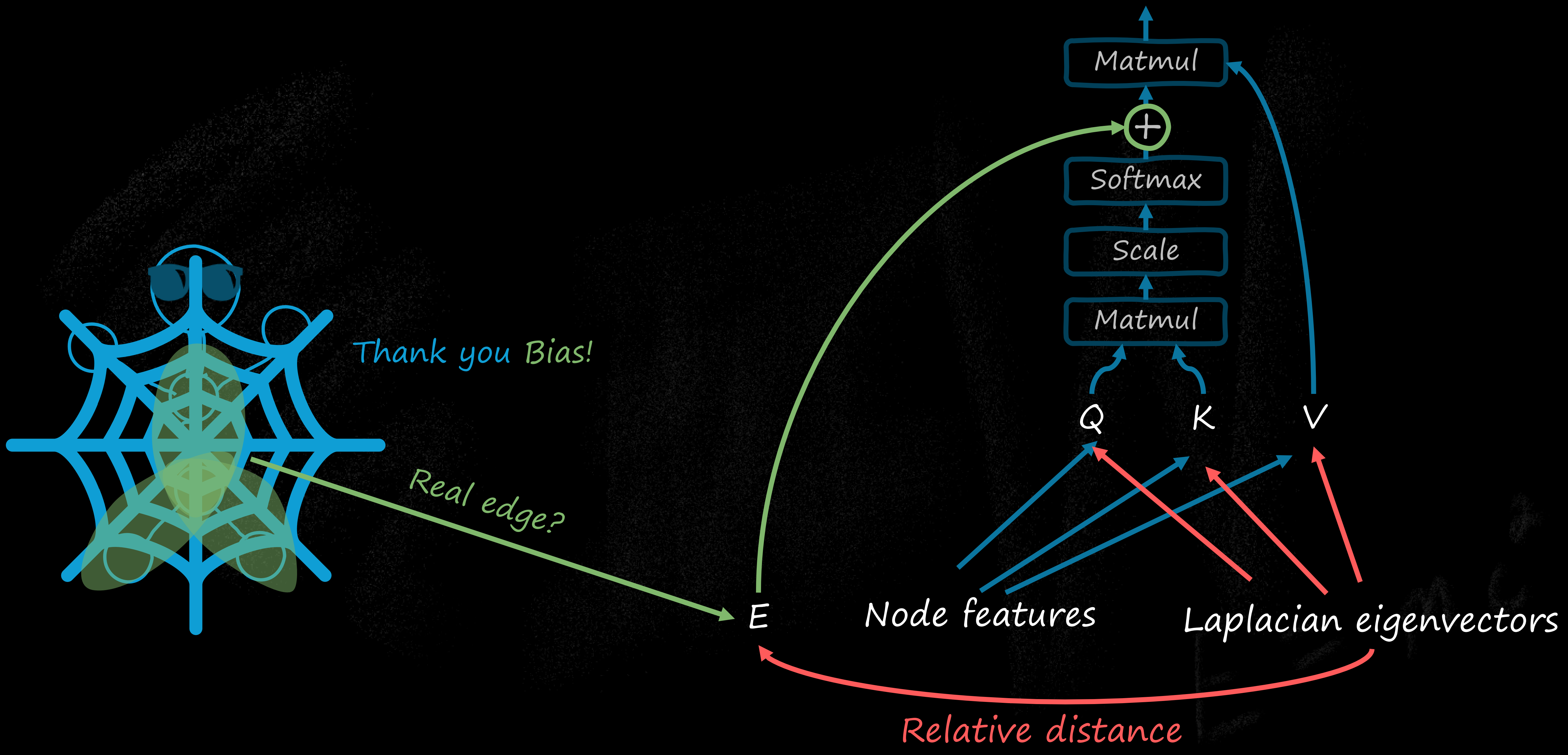
Basic graph Transformers

Basic graph Transformers have very poor results

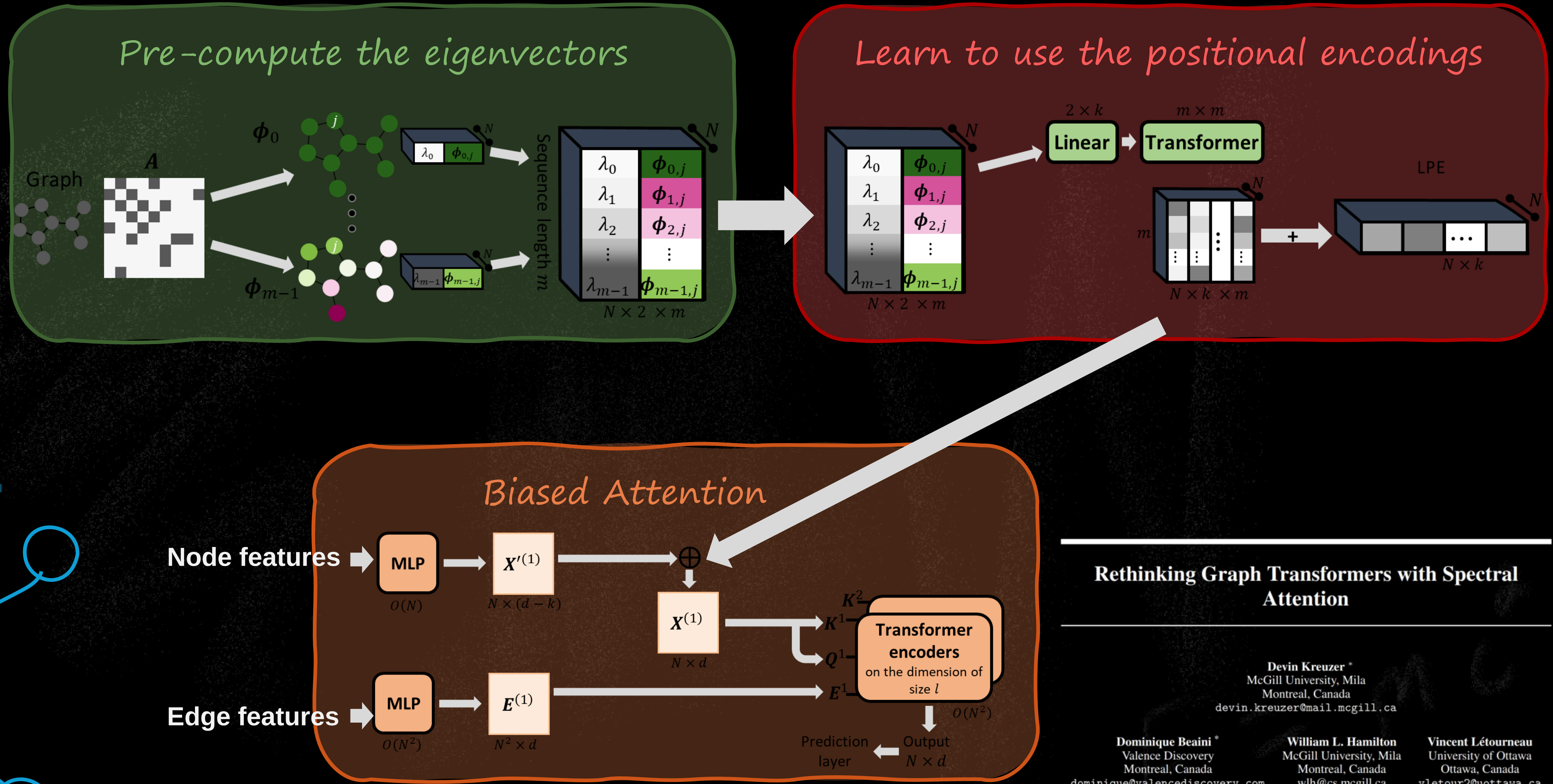
- The connectivity is a strong inductive bias
- The eigenvectors are noisy and hard to understand for the network
- Edge features are missing

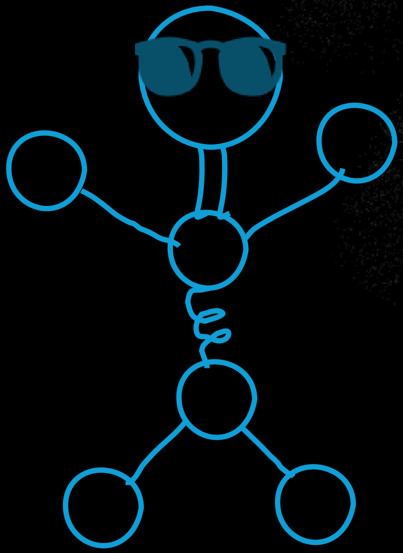


Biased FULL-Attention



SAN - Spectral Attention Network





Rethinking Graph Transformers with Spectral Attention

Devin Kreuzer *
McGill University, Mila
Montreal, Canada
devin.kreuzer@mail.mcgill.ca

Dominique Beaini *
Valence Discovery
Montreal, Canada
dominique@valencediscovery.com

William L. Hamilton
McGill University, Mila
Montreal, Canada
wlh@cs.mcgill.ca

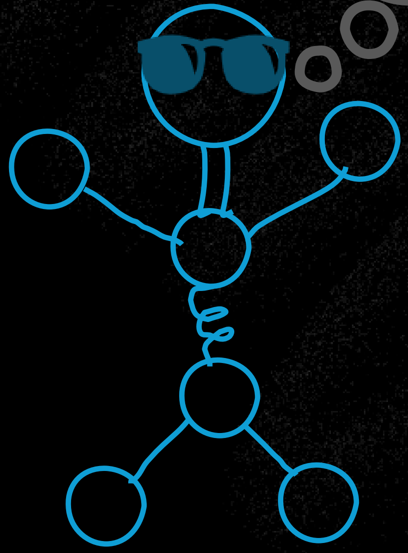
Vincent Létourneau
University of Ottawa
Ottawa, Canada
vletour2@uottawa.ca

Prudencio Tossou
Valence Discovery
Montreal, Canada
prudencio@valencediscovery.com

Pre-training

Since we need
lots and lots of data,
Let's do pre-training.

How do we pre-train a
molecular representation?



Biology

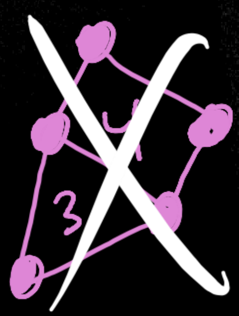
Cell assays
Transcriptomics

Chemistry

Protein assays
Physicochemical (solubility, etc.)

Quantum mechanics

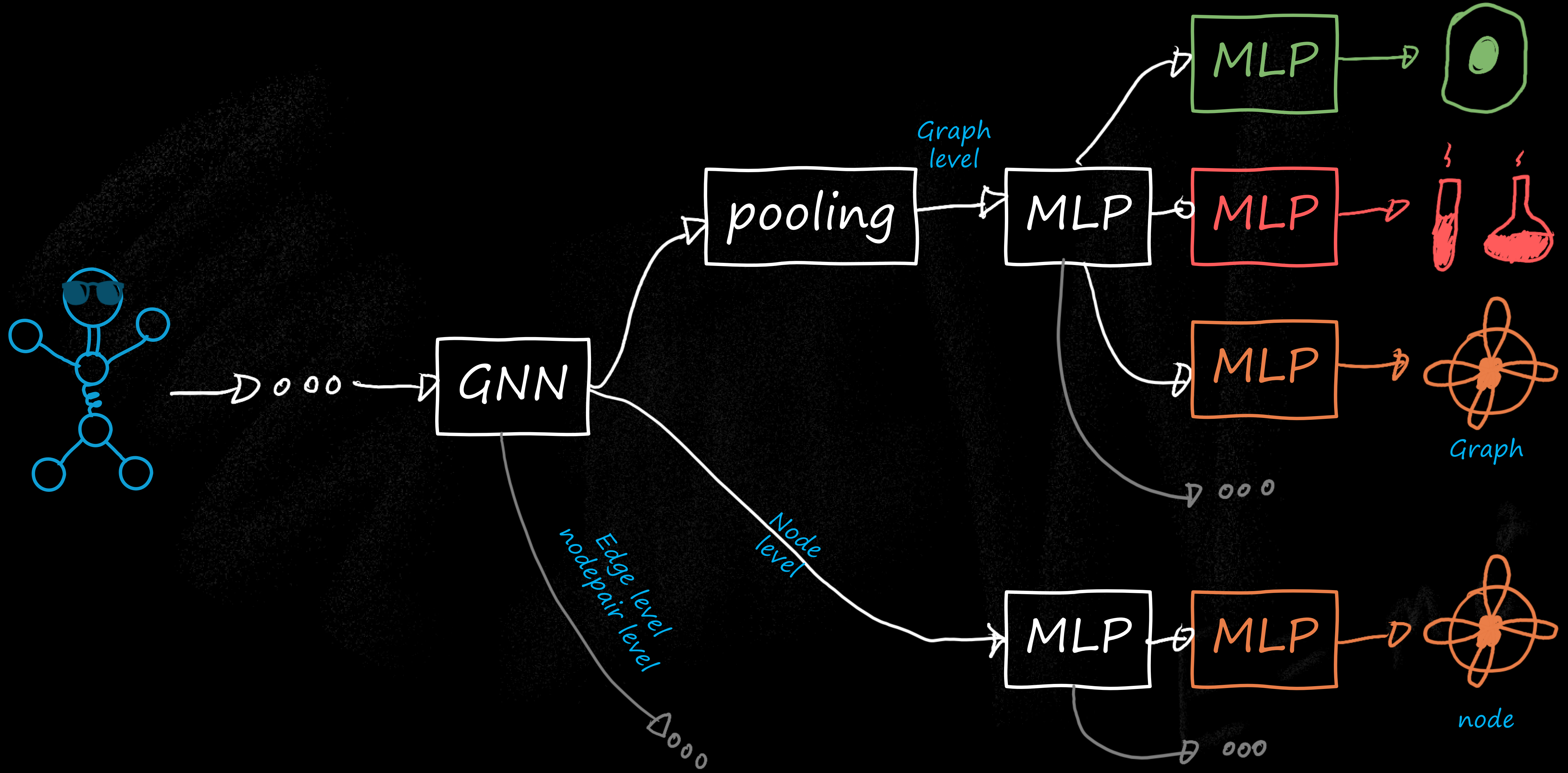
HOMO-LUMO gap
Partial charges



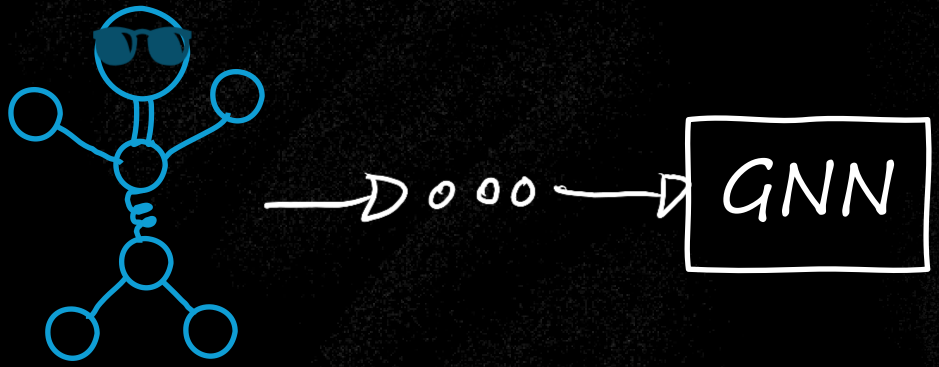
~~Self-supervision~~

~~Finding missing atoms~~
~~Enumerating structures~~
~~SMILES reconstruction~~

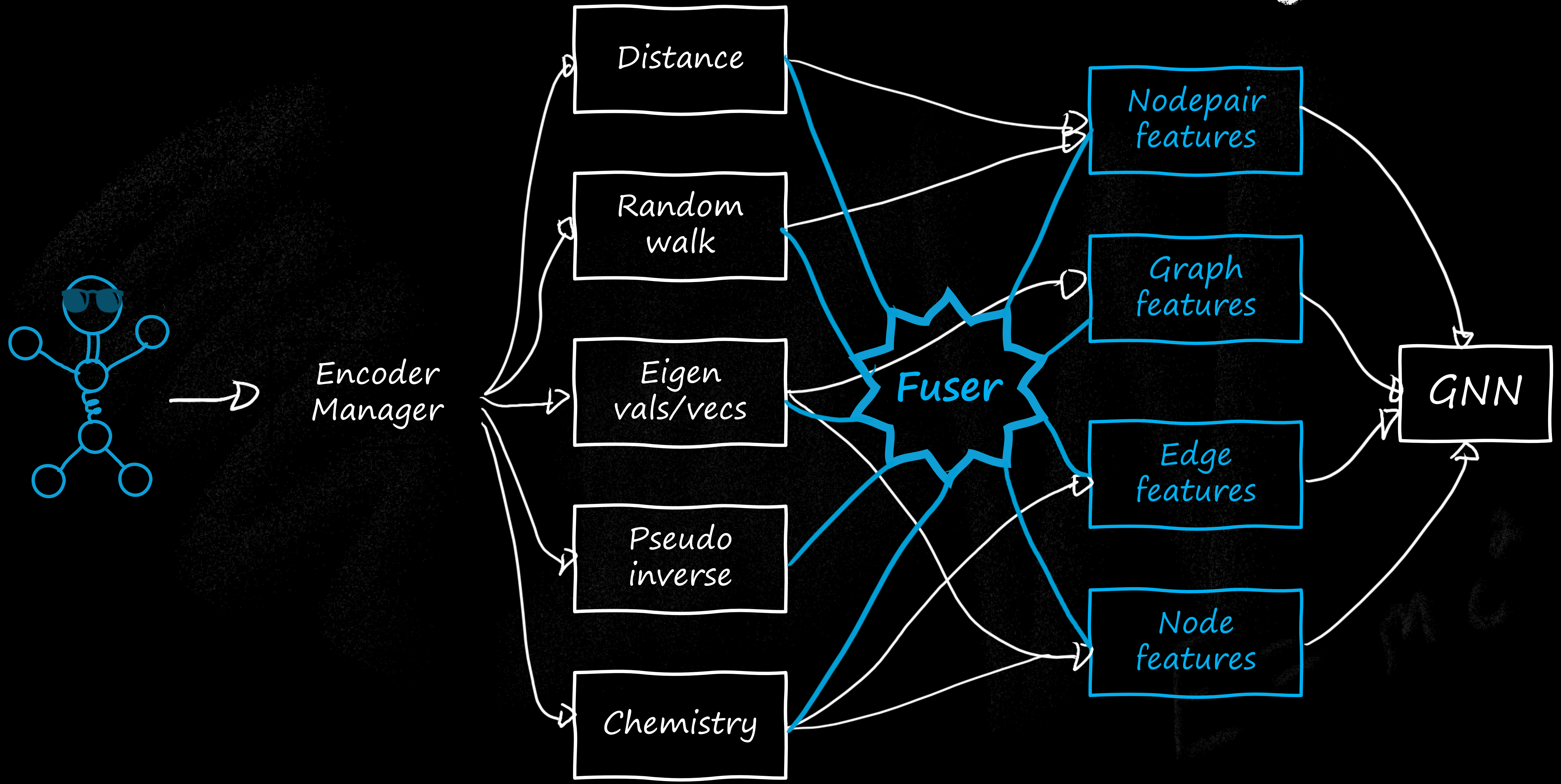
Multi-level multi-tasking



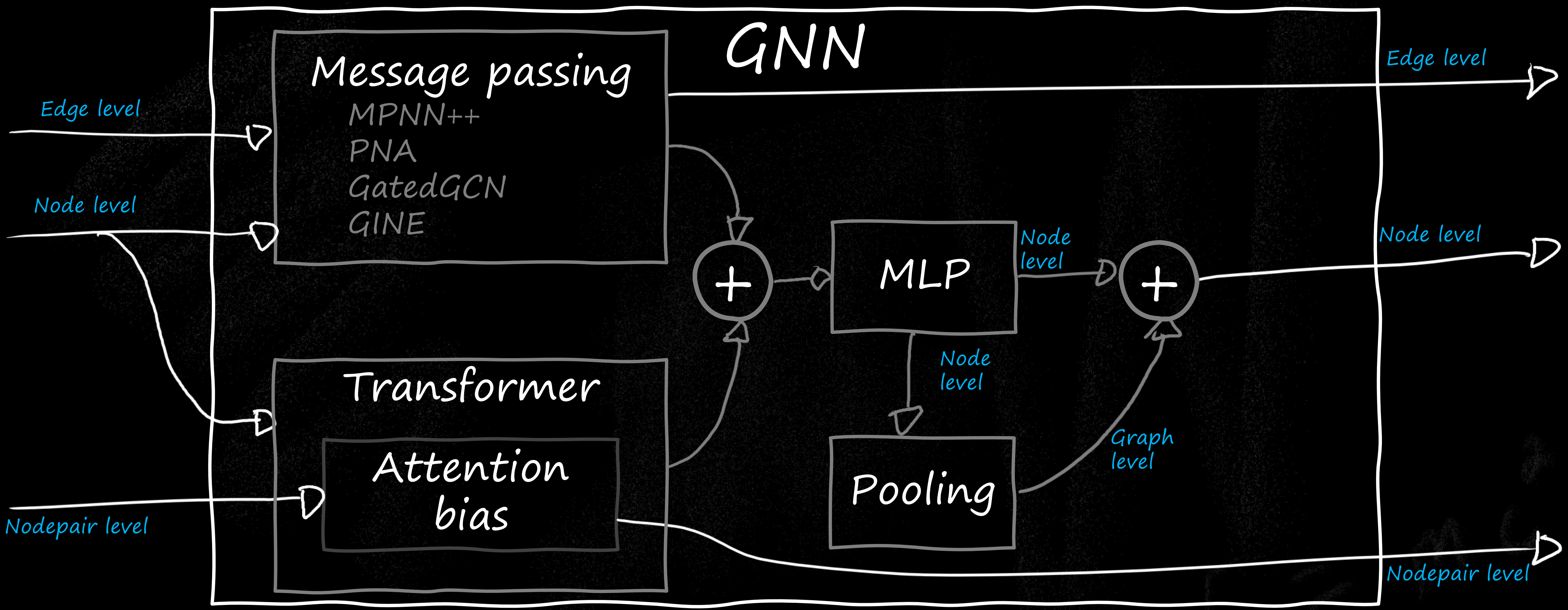
Multi-level multi-tasking



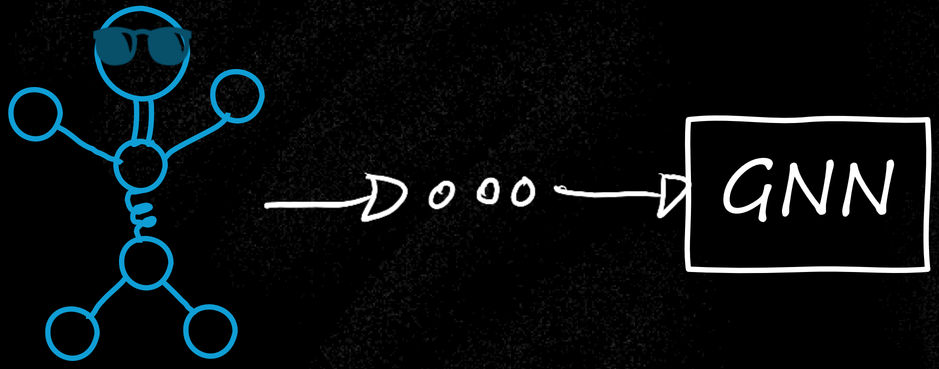
Multi-level positional encoding



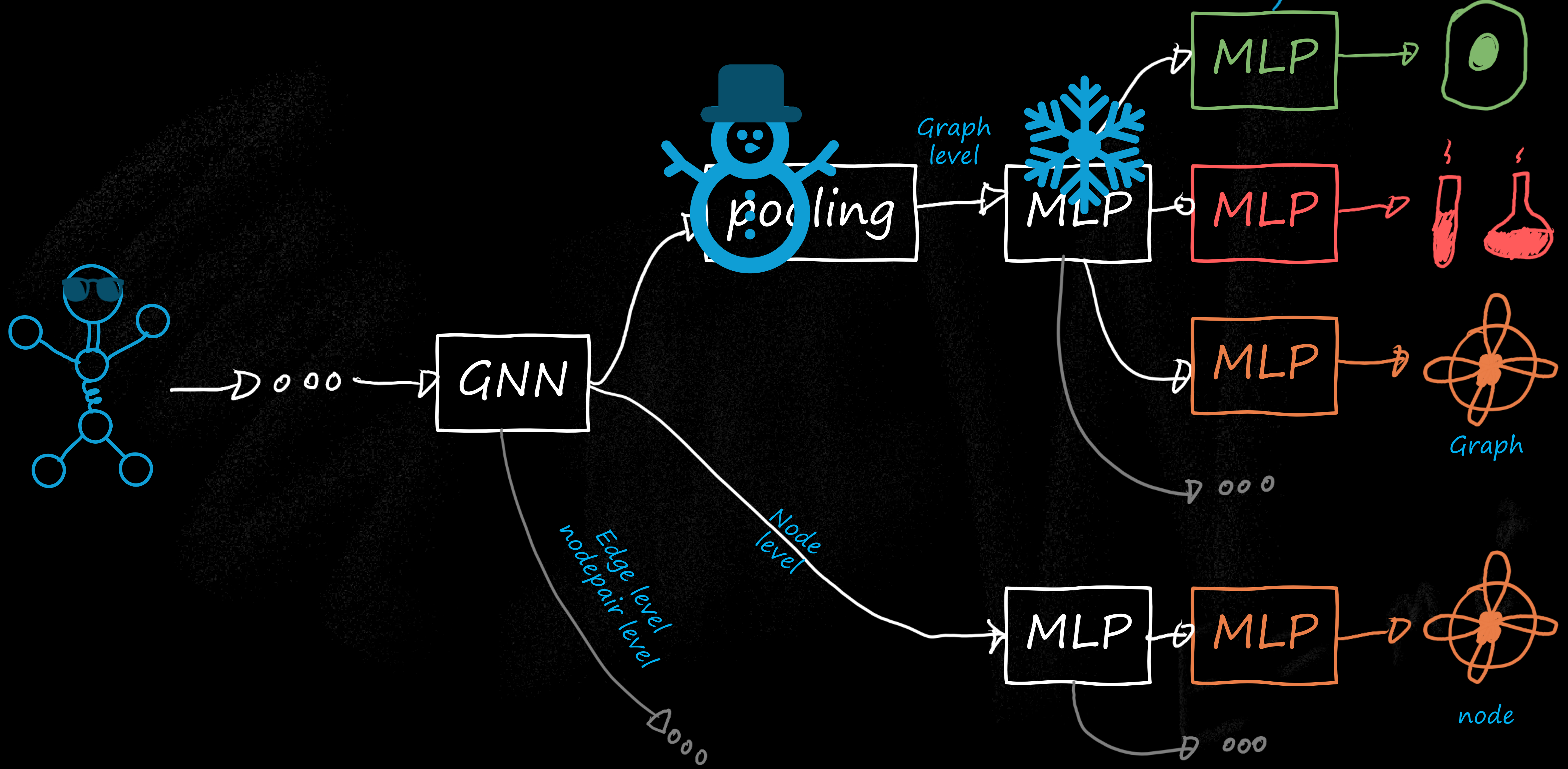
Multi-Level graph Transformer 🗼



Finetuning 🏹



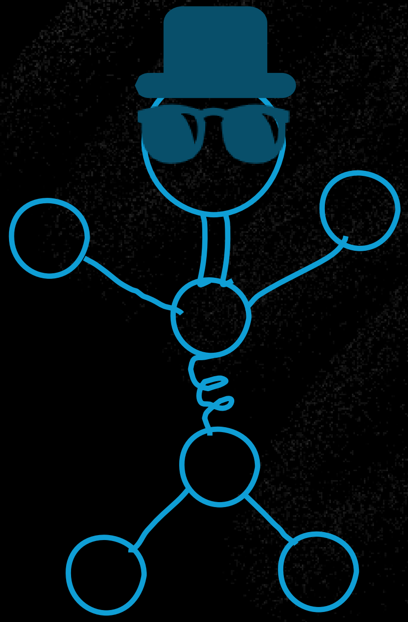
Finetuning ⛄

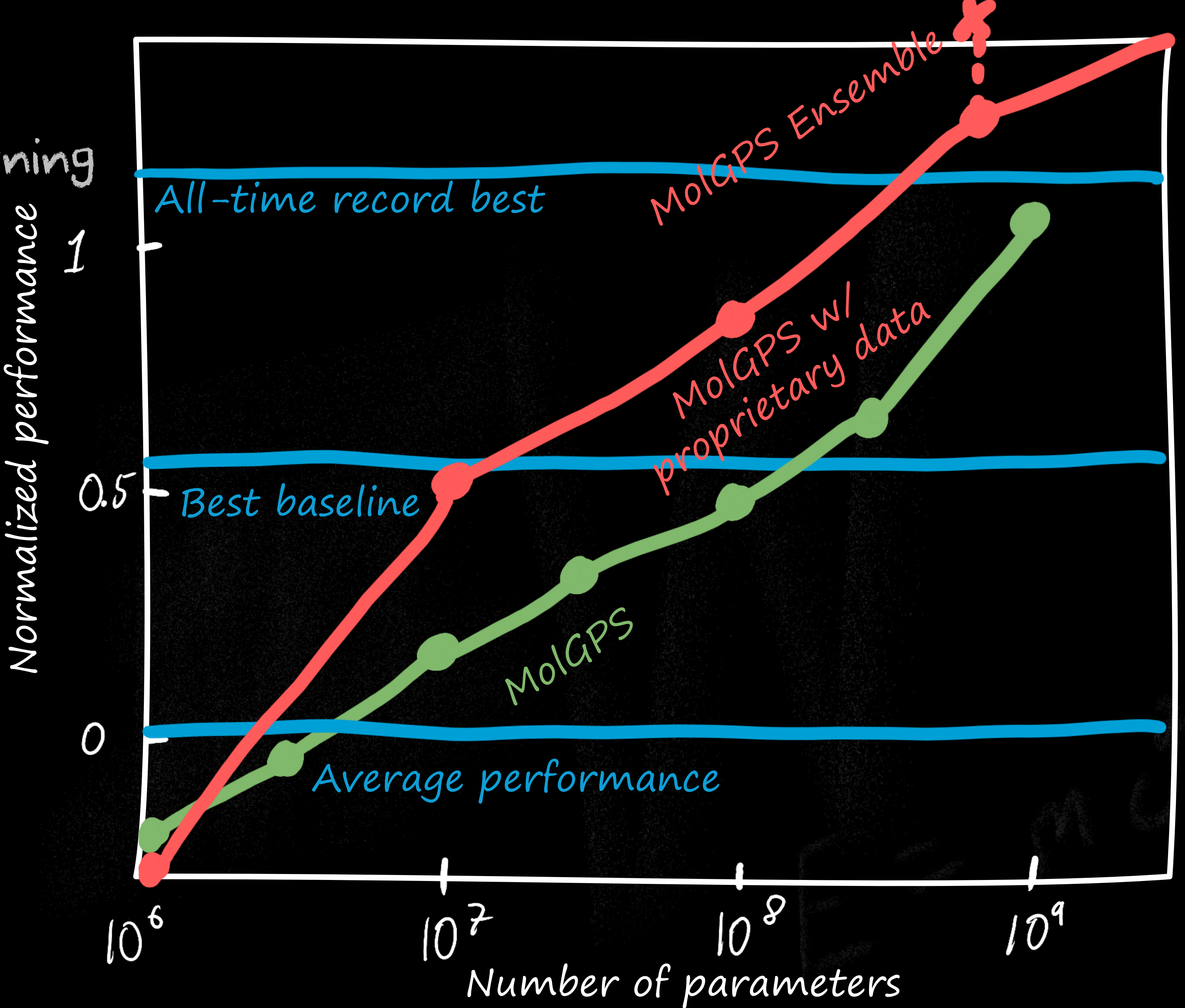

GNN
pooling
Graph level
MLP
MLP
MLP
MLP
New task
Graph
Node level
Edge level
nodepair level
MLP
MLP
node

Scaling

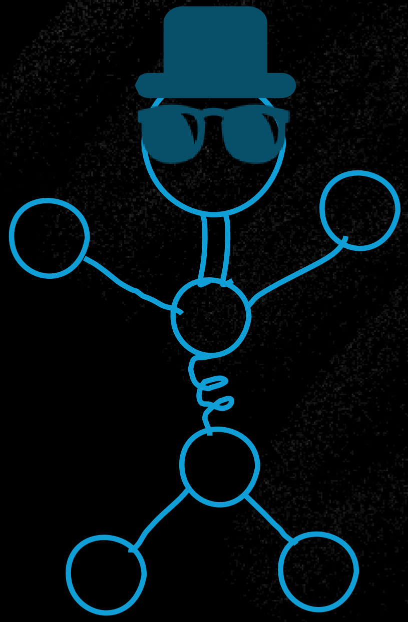
ON downstream finetuning

To infinity, and Beyond!





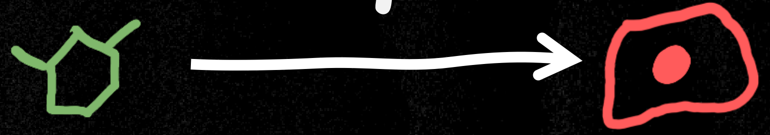
Limitation in no-context finetuning



- Even a model that perfectly understand physics and biology will overfit without context of the task
- Tasks can be encoded as
  - natural language
  - protein sequences
  - Cellular context
- Multimodality allows to encode context, and will make the GNNs much much more powerful

MolPhenix

How molecules impact cells

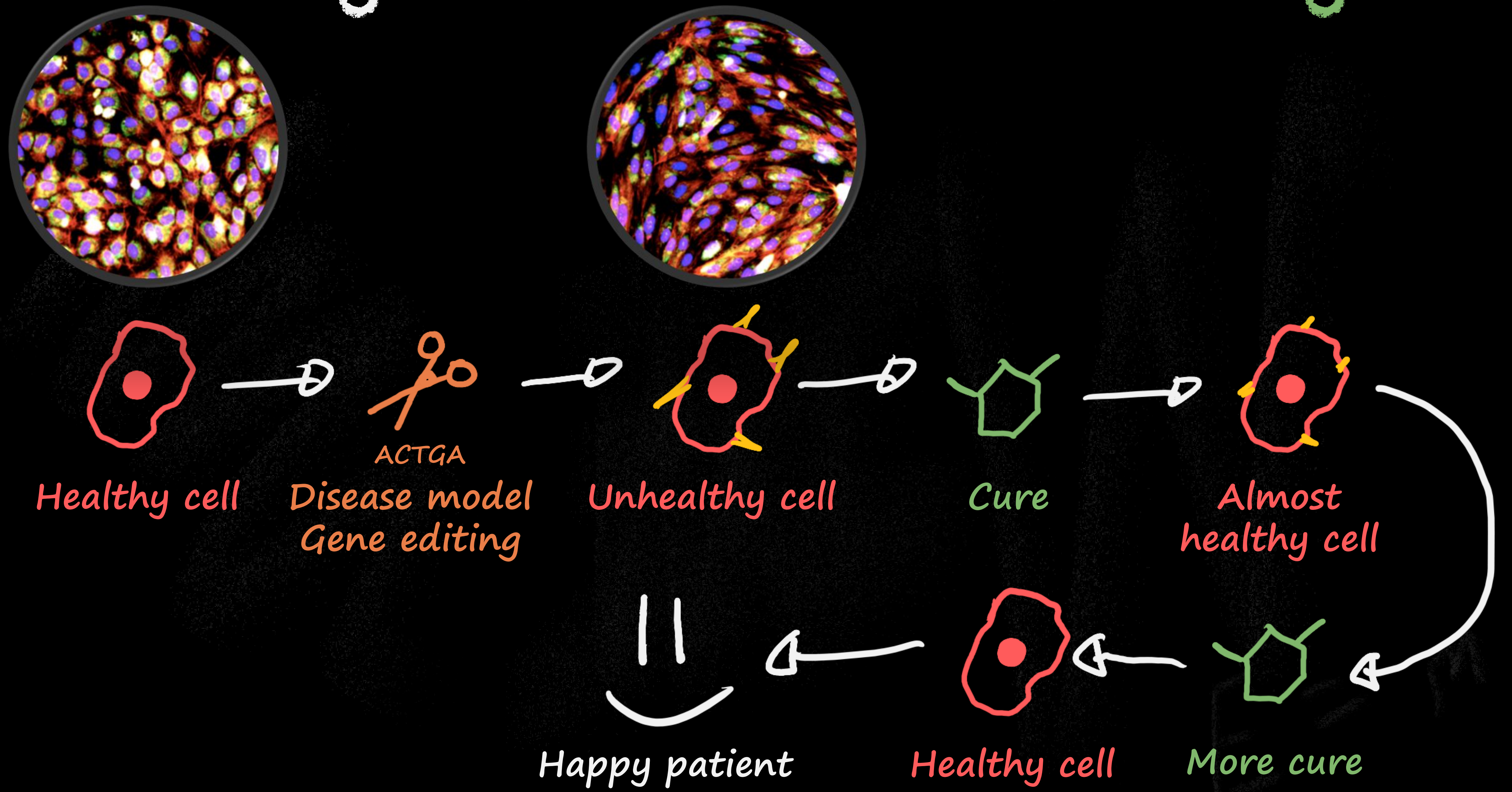


Unlocking phenomolecular retrieval

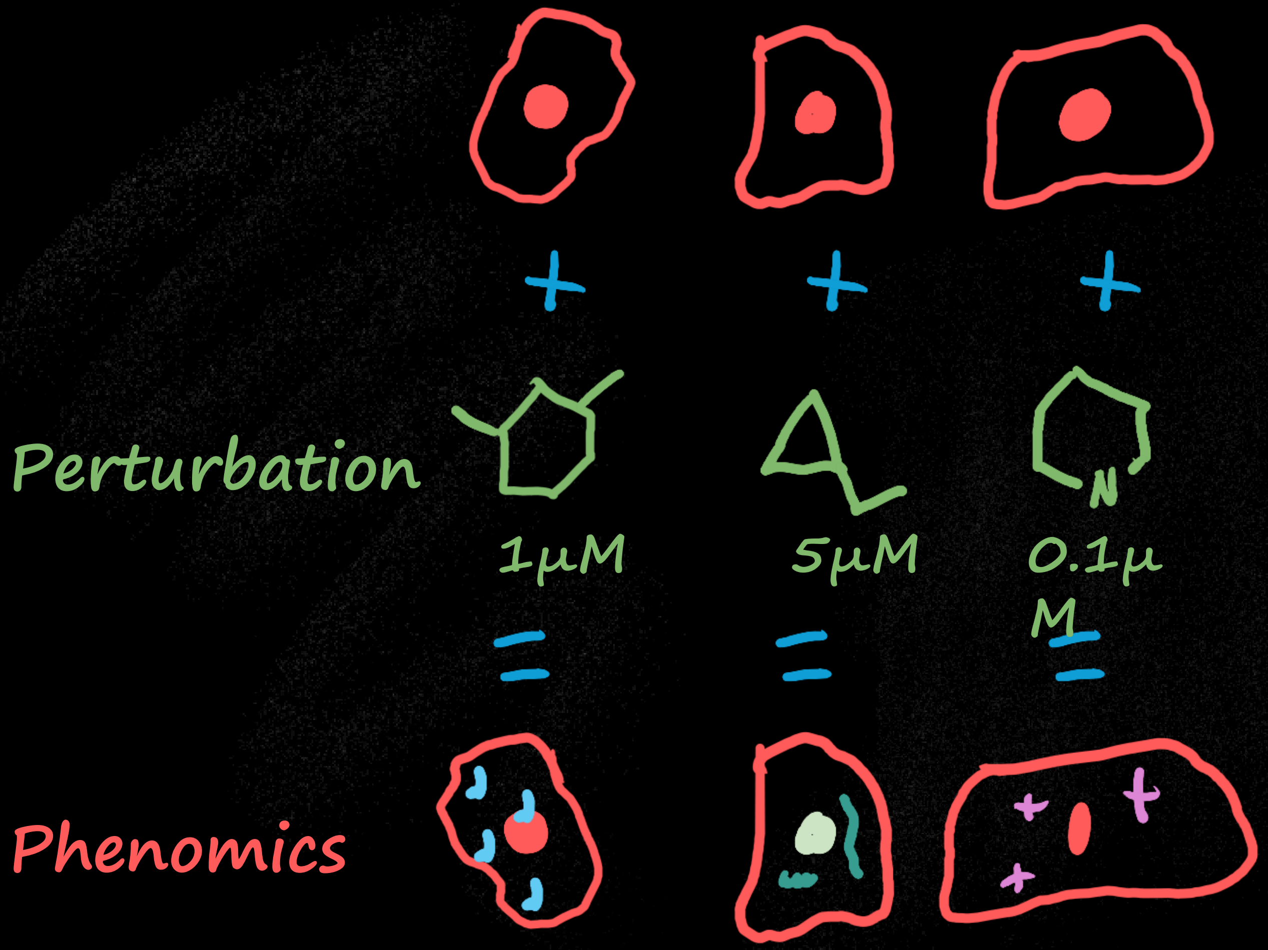
A presentation by Dominique Beaini
From Valence Labs / Recursion

Valence Labs
Powered by Recursion

Modeling a disease... And curing it

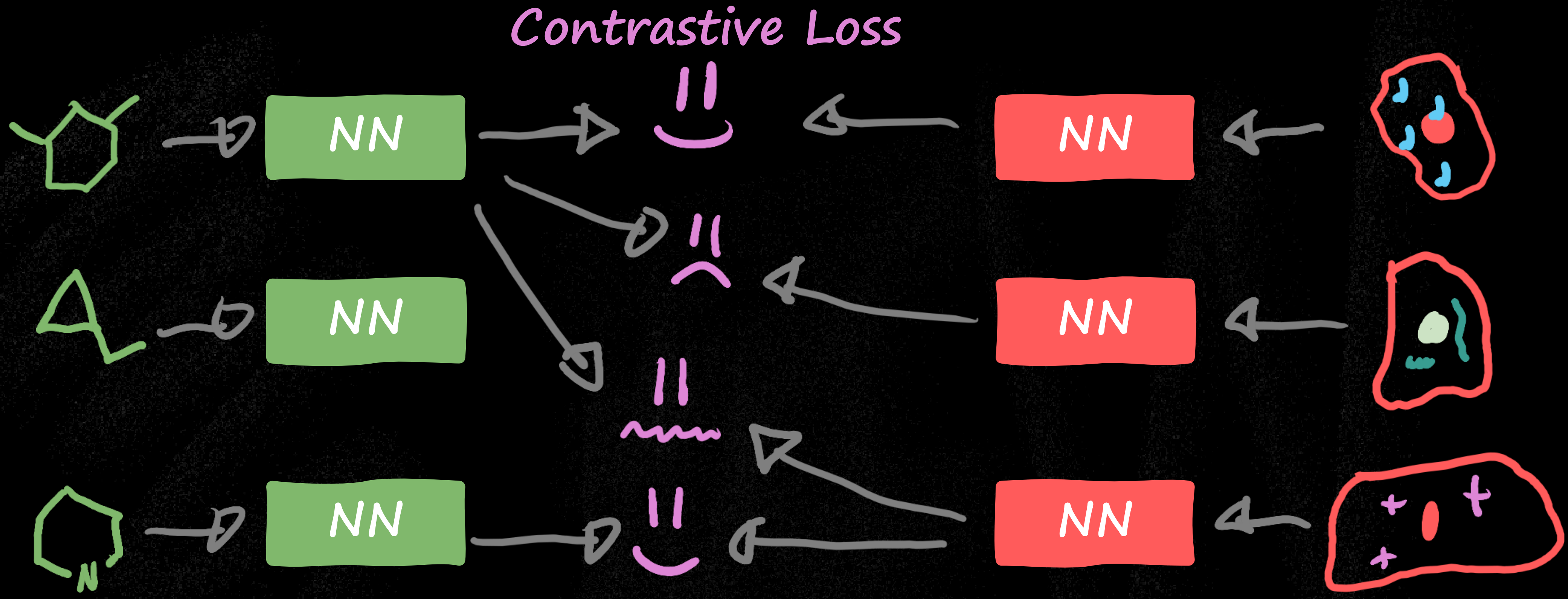


Phenomics screening of molecules



How to build a model for this equation to understand how molecules impact cells?

Let's try some contrastive learning

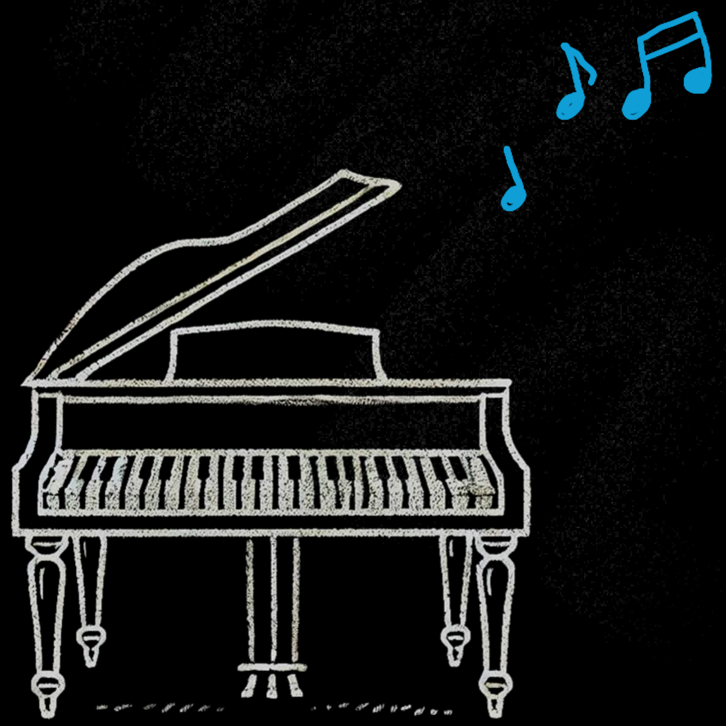


Prior methods have not succeeded, achieving only 8% recall

The 3 Challenges

and The Piano in New York

Which ~~molecule~~ is playing?
music



The 3 Challenges
and The Piano in New York

Is it visually ~~inactive~~?
Background noise

Which ~~molecule~~ is playing?
music

Increase the ~~concentration~~
volume

7pm

4pm

The 3 Challenges

Natural variations: Batch effects are the largest source of variation in phenomics images.

Can we ignore it?

90% inactives: Most molecules have no visible effect.

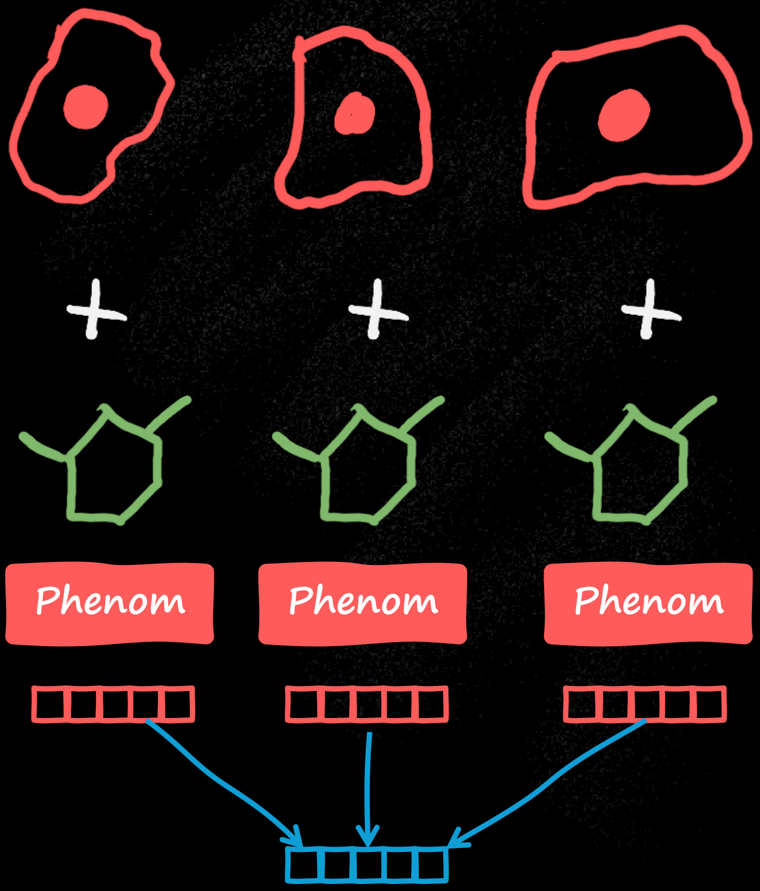
How to handle this source of random noise?

Concentration
Too low: Nothing happens
Too high: Everyone dies

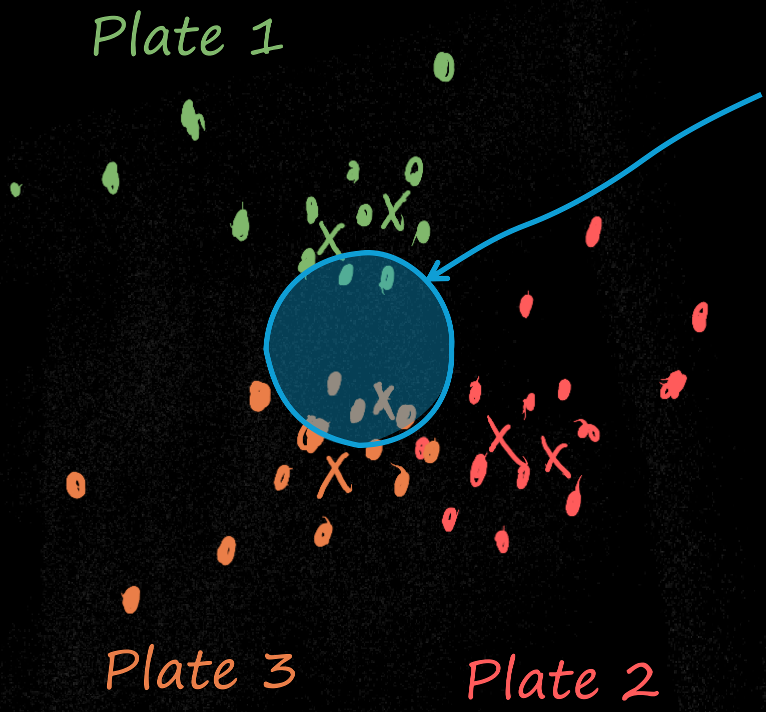
How to model this non-linear relationship?

Pre-train your vision encoder

Average experimental replicates



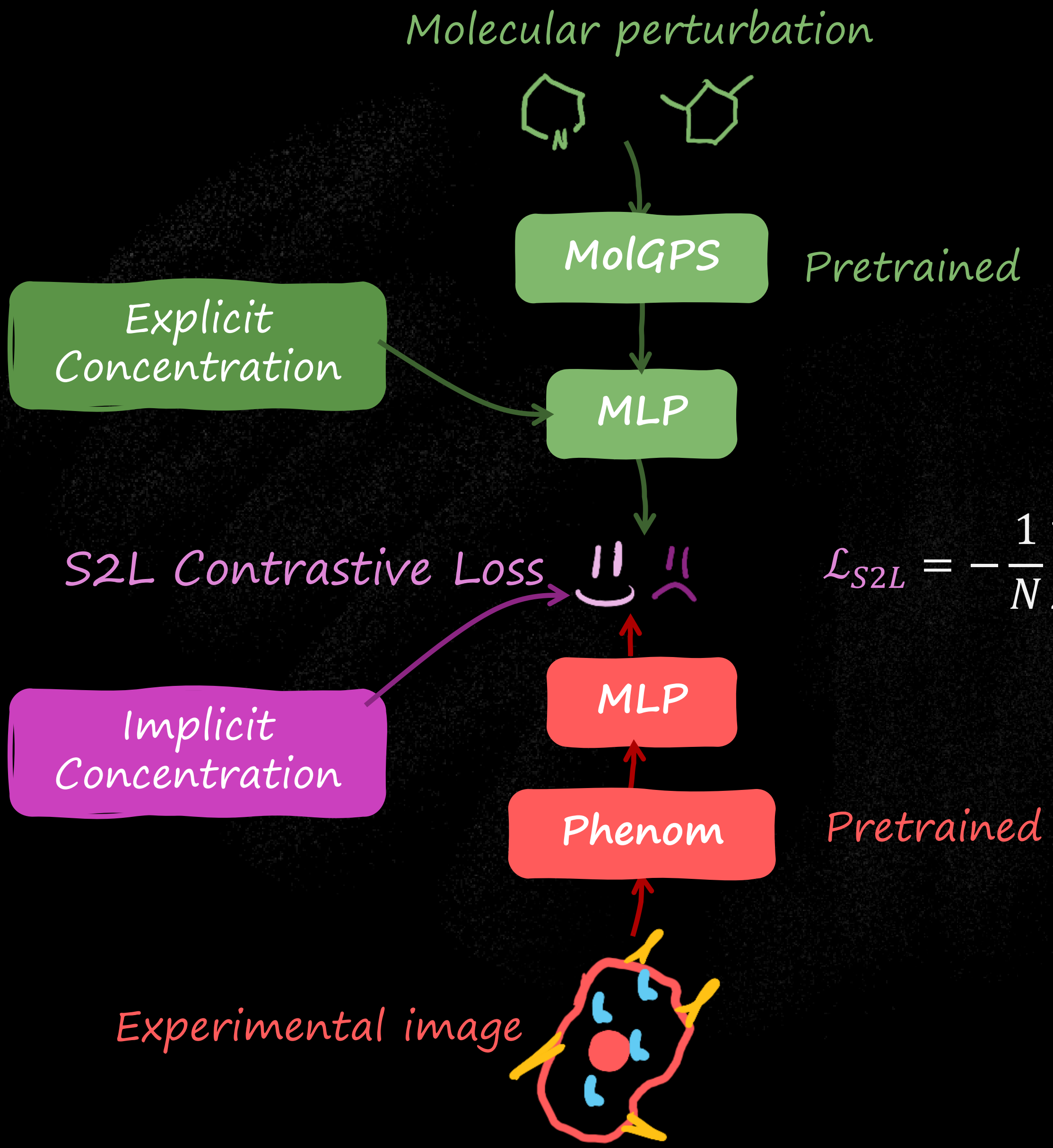
Re-center to the controls x



Discard/reweight inactives

Use statistical analysis to find inactive compounds from similarity to the controls

Better contrastive Learning with S2L Loss

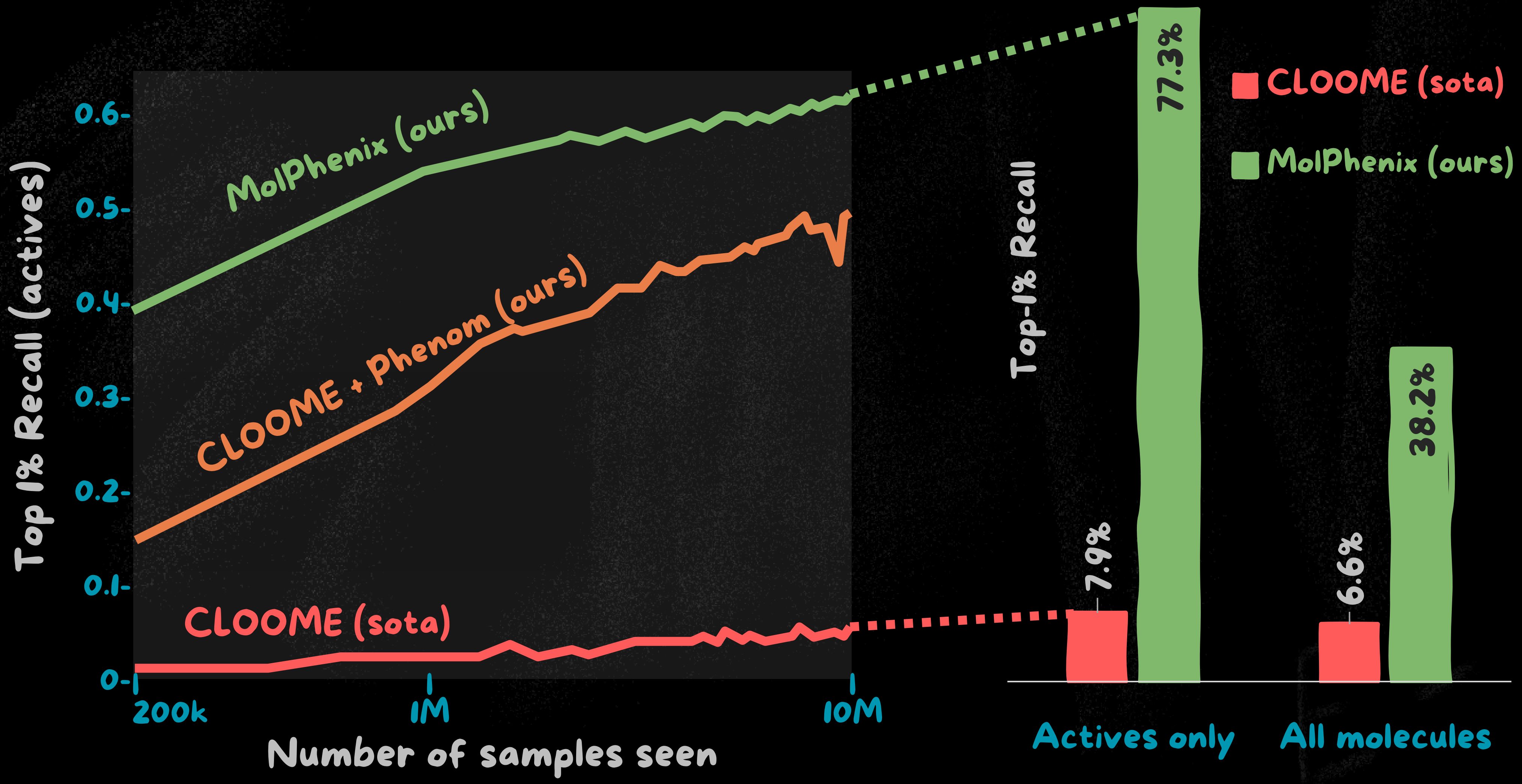


Re-weight samples based on Phenomics embedding

$$\mathcal{L}_{S2L} = -\frac{1}{N}\sum_{i-1}^{N}\sum_{j-1}^{N}\log\left[\frac{w_{i,j}^{x}}{1+\exp(-\alpha\langle \boldsymbol{z}_X, \boldsymbol{x}_m\rangle + b)} + \frac{1-w_{i,j}^{x}}{1+\exp(\alpha\langle \boldsymbol{z}_X, \boldsymbol{x}_m\rangle + b)}\right]$$

Use sigmoid to reduce the effect of false negatives

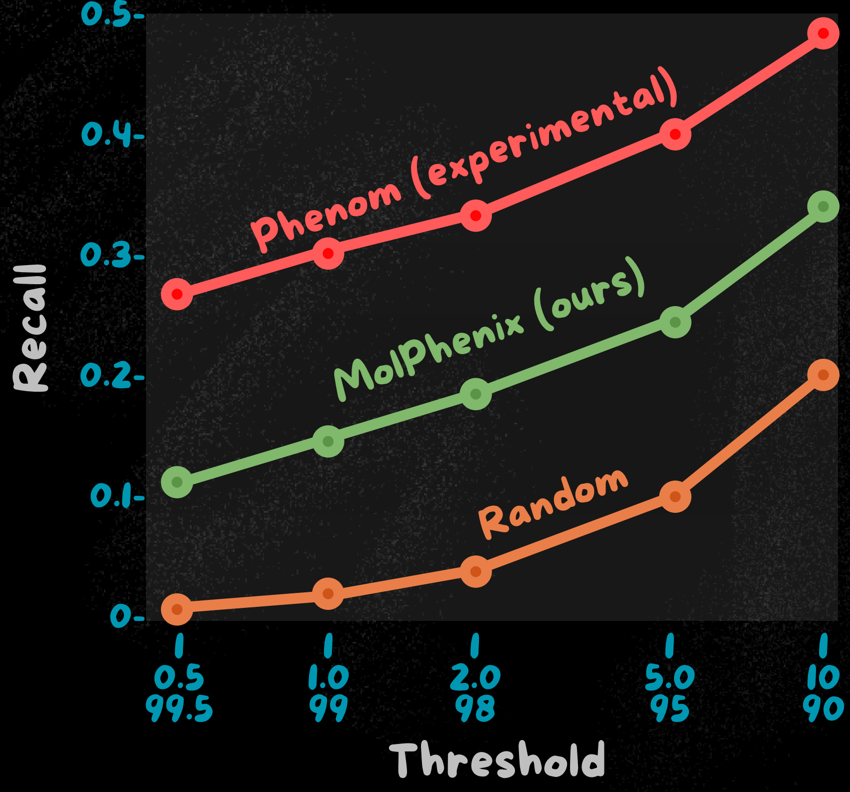
10x recall compared to previous SOTA



Downstream applications

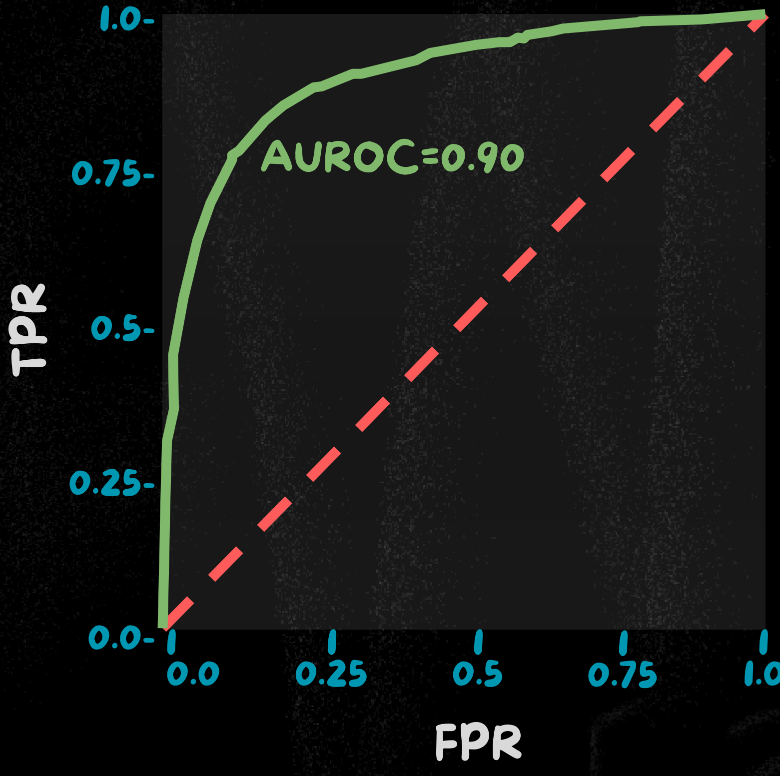
Can we find gene/mol relations?

Half as good as experiments
without even training on genes



Is a molecule pheno-active?

Single linear layer achieves 90% AUROC



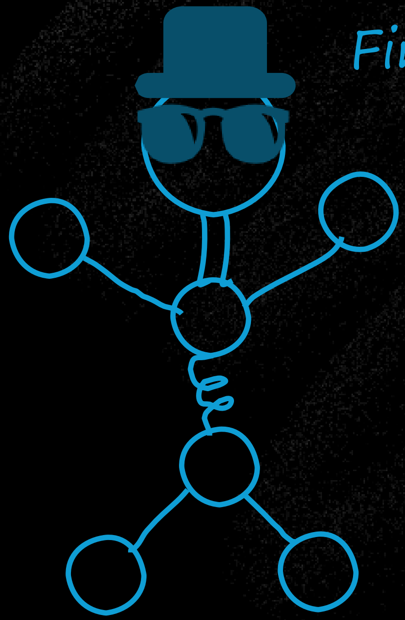
MolPhenix

- MolPhenix opens a completely new direction for ML in drug discovery with 10x improvement

- We can model how molecules impact cells, not just do some predictive assays

- A first step towards Virtual Cells, to industrialize drug discovery in the age of AI

Thank you Graphy!! And Dom



Finally, Dom will stop talking!

But if you're not tired of him, you can follow him on ~~Twitter~~ X
@Dom_Beaini

Thanks to a thousand co-authors!