Valence
The Drug Design Company

Mila

UNIVERSITY OF CAMBRIDGE

Directional Graph Networks

ICML2021 – Long presentation

Valence

The Drug Design Company

Learn

Leverage diverse sources of data across your entire discovery organization for unparalleled predictive performance

Design

Unify generative methods with expert intuition to systematically explore new chemical space free from IP constraints

Optimize

Ensure rapid progress against critical design criteria using active learning and iterative optimization strategies

Integrate

Seamlessly integrate modern deep learning workflows into R&D organizations of all sizes

Directional Graph Networks

Dominique Beaini, Saro Passaro*, Vincent Létourneau, William L. Hamilton, Gabriele Corso, Pietro Liò*

Dominique Beaini*

Saro Passaro*

Follow us on Twitter

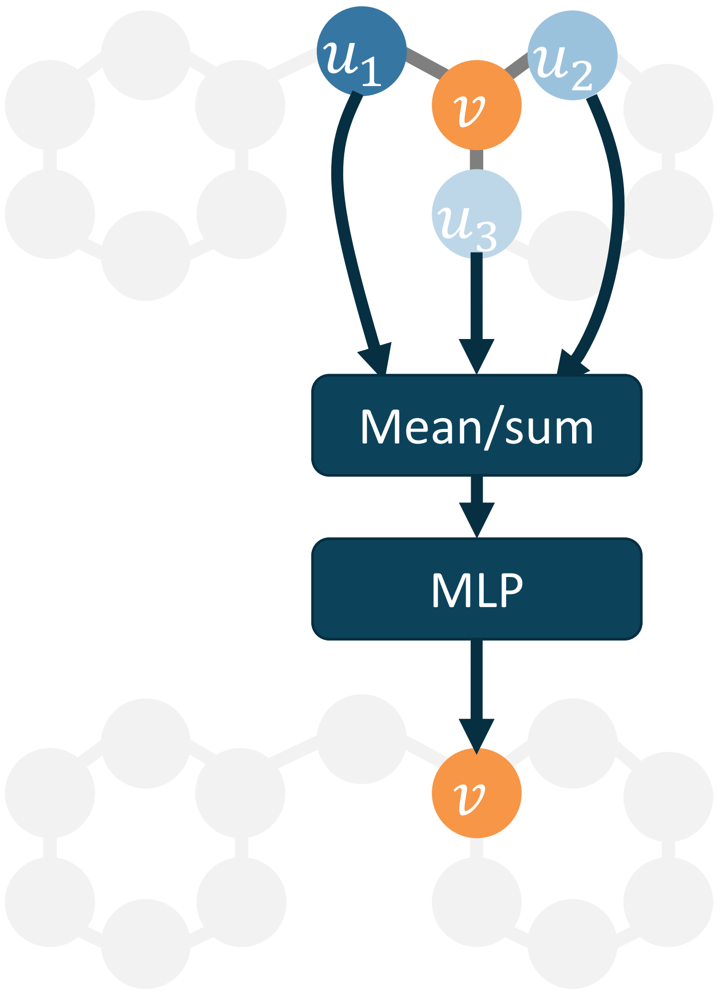
@valence_ai
@dom_beaini
@Saro2000
@vincentmillions

Valence

UNIVERSITY OF CAMBRIDGE

Mila

Graph neural networks (GNNs)



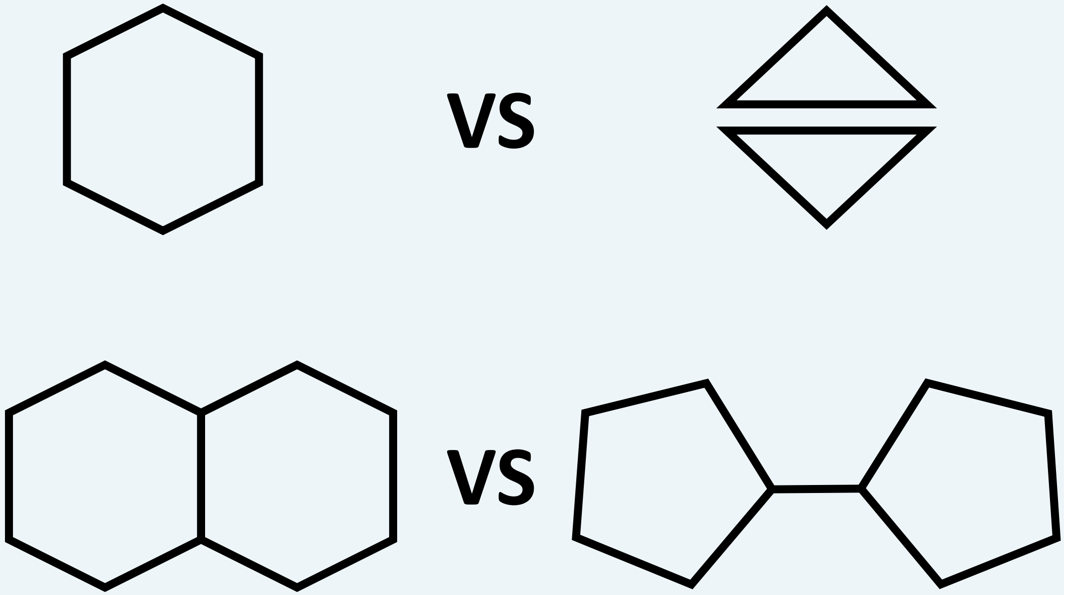
Low discriminative power of GNNs

- Most GNNs consider all neighbours equally
- They lack directional propagation → less powerful than CNNs (Convolutional neural networks)
- Deep GNNs tend to over-smooth and over-squash the information
- Anisotropic kernels are feature-based, and do not use the topology of the graph (e.g. GAT)

Why is this a problem for molecular graphs?

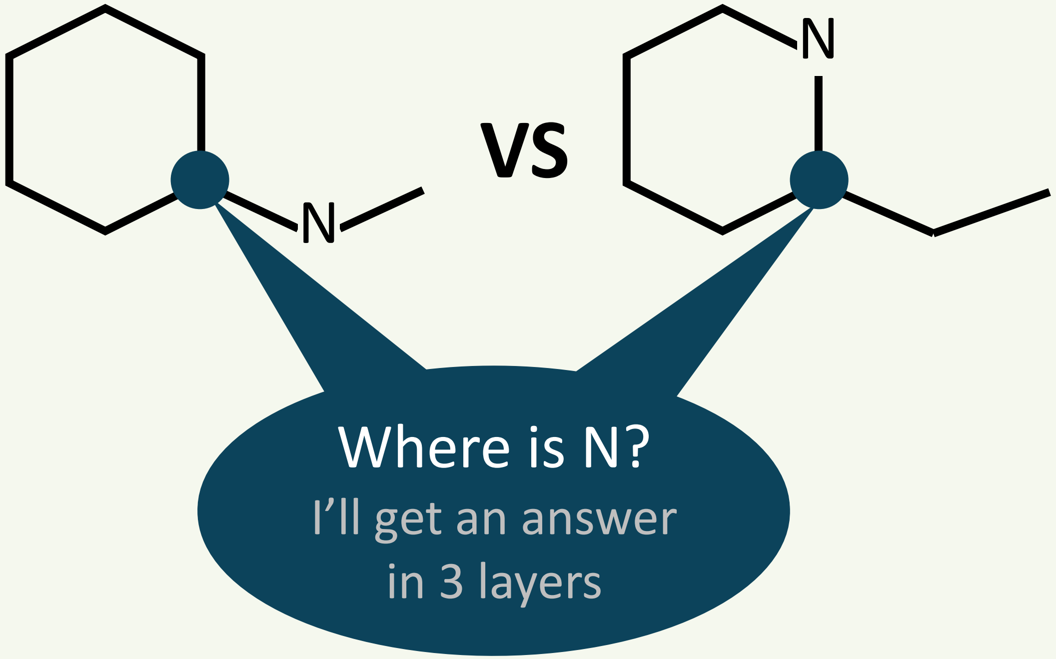
The isomorphism problem

- Isomorphic molecular graphs cannot be distinguished without directional information



Who is what?

- Multiple layers required to properly understand who sends a message



Proposed DGN

Projecting message passing on a directional vector field

- We define a set of vector fields in a graph
- We project the incoming messages on the vector field
  - **Directional smoothing:** average of forward and backward messages
  - **Directional derivative:** subtraction of forward and backward messages

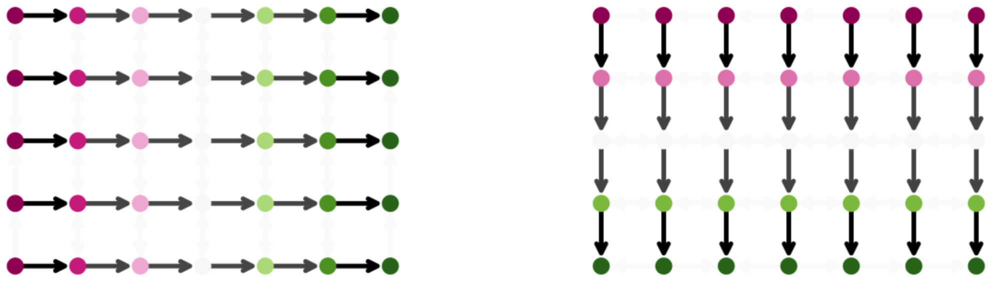
Eigenvectors give a directional flow

- The gradient of the low-frequency Laplacian eigenvectors flows in interpretable directions
- We theoretically reduces over-smoothing and over-squashing
- We generalize CNNs on grid graphs

Grid graph (7×5)

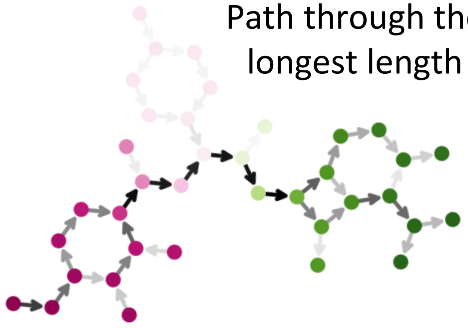
$\text{acos}\,\hat{\boldsymbol{\Phi}}_1$ Path through the longest length

$\text{acos}\,\hat{\boldsymbol{\Phi}}_2$ Path through the 2nd longest length

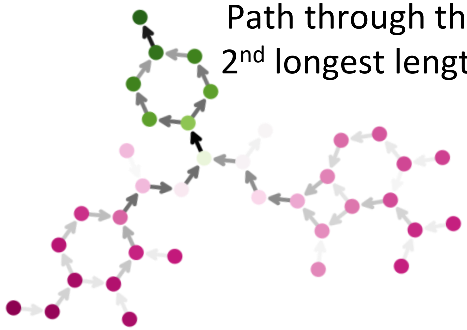


Molecular graph

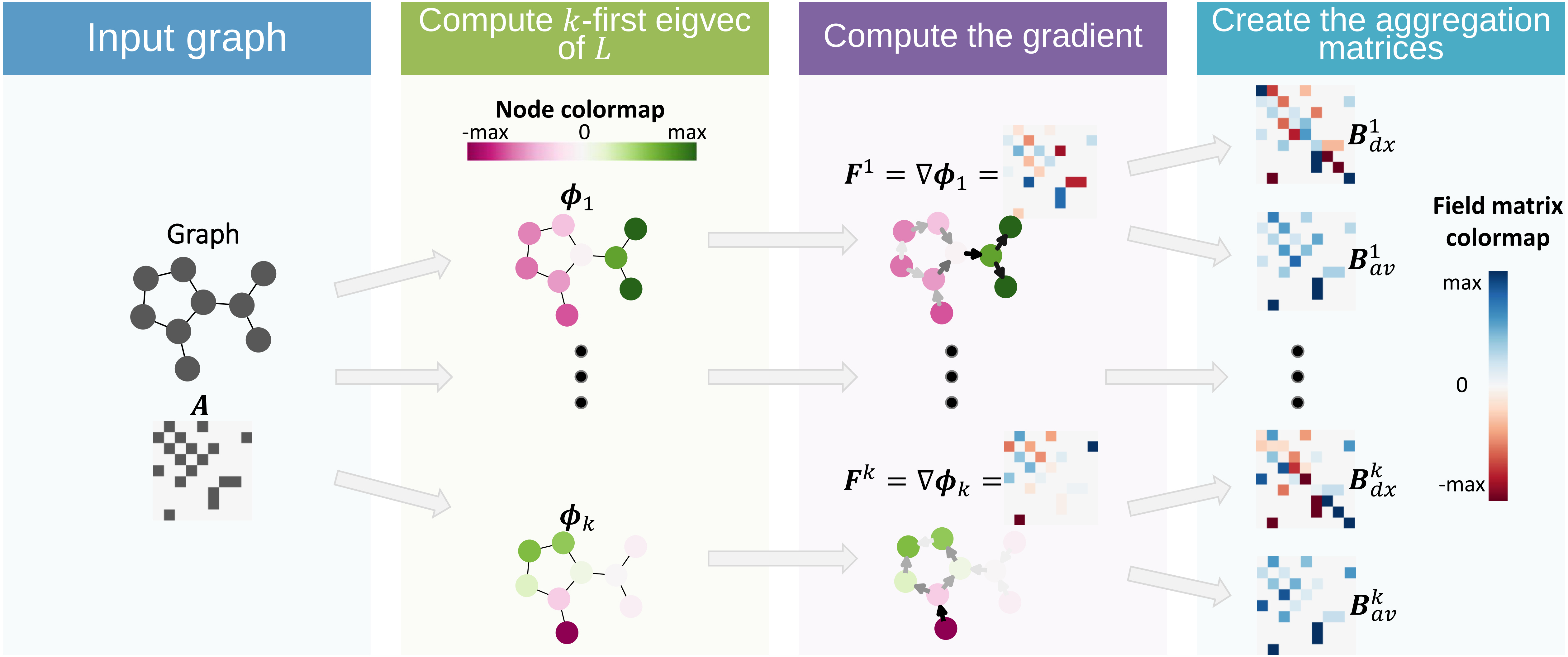
$\text{acos}\,\hat{\boldsymbol{\Phi}}_1$ Path through the longest length



$\text{acos}\,\hat{\boldsymbol{\Phi}}_2$ Path through the 2nd longest length

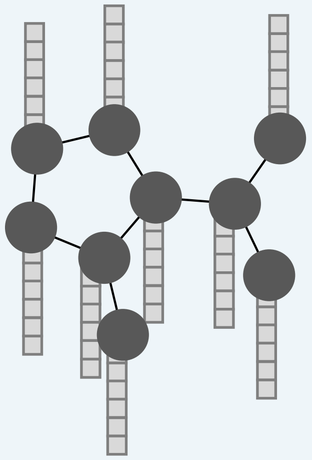


Building the directional matrices (pre-computed steps)



GNN architecture

Input graph with features



Pre-computed aggregation matrices

$$\begin{Bmatrix} \boldsymbol{B}^1_{dx} \\ \boldsymbol{B}^1_{av} \\ \vdots \\ \boldsymbol{B}^k_{dx} \\ \boldsymbol{B}^k_{av} \end{Bmatrix}$$

Aggregation of features $\boldsymbol{X}$

- Different directional aggregators are used
- The results of the aggregations are concatenated

$$\boldsymbol{Y}^{(t)} = \text{concat}\begin{Bmatrix} \boldsymbol{D}^{-1}\boldsymbol{A}\,\boldsymbol{X}^{(t)} \\ \left|\boldsymbol{B}^1_{dx}\,\boldsymbol{X}^{(t)}\right| \\ \boldsymbol{B}^1_{av}\,\boldsymbol{X}^{(t)} \\ \vdots \\ \left|\boldsymbol{B}^k_{dx}\,\boldsymbol{X}^{(t)}\right| \\ \boldsymbol{B}^k_{av}\,\boldsymbol{X}^{(t)} \end{Bmatrix}$$

MLP

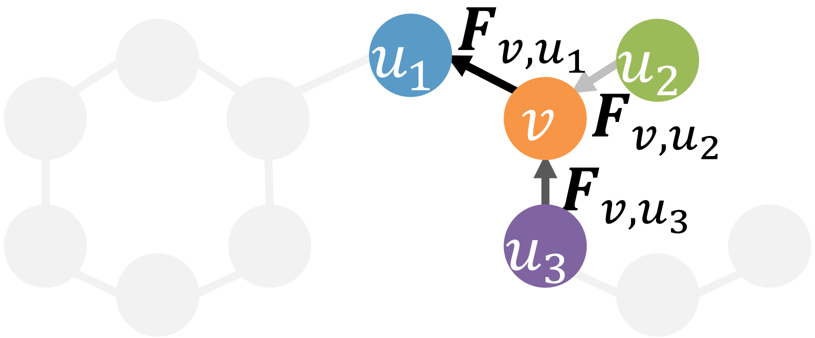
- This is the only step with learned parameters

$$\boldsymbol{X}^{(t+1)} = \text{MLP}\left(\boldsymbol{Y}^{(t)}\right)$$

Next GNN layer

$$t \rightarrow t+1$$
$$\boldsymbol{X}^{(t)} \rightarrow \boldsymbol{X}^{(t+1)}$$

The directional aggregation matrices



Directional smoothing

$$\boldsymbol{B}_{av} = \frac{|\boldsymbol{F}_{v,u_1}|u_1 + |\boldsymbol{F}_{v,u_2}|u_2 + |\boldsymbol{F}_{v,u_3}|u_3}{|\boldsymbol{F}_{v,u_1}| + |\boldsymbol{F}_{v,u_2}| + |\boldsymbol{F}_{v,u_3}|}$$

$$\frac{\text{Absolute weighted sum}}{\text{Sum of the absolute weights}}$$

Directional derivative

$$\boldsymbol{B}_{dx} = \frac{\boldsymbol{F}_{v,u_1}(u_1 - v) + \boldsymbol{F}_{v,u_2}(v - u_2) + \boldsymbol{F}_{v,u_3}(v - u_3)}{|\boldsymbol{F}_{v,u_1}| + |\boldsymbol{F}_{v,u_2}| + |\boldsymbol{F}_{v,u_3}|}$$

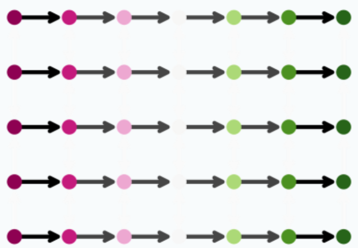
Weighted *forward* derivative with u_1 + Weighted *backward* derivative with u_2 + Weighted *backward* derivative with u_3

Sum of the absolute weights

Generalizing CNNs

Horizontal direction

$\nabla \boldsymbol{\phi}_1$



$$\boldsymbol{y} = 2\boldsymbol{B}_{av}^{1}\,\boldsymbol{x}$$

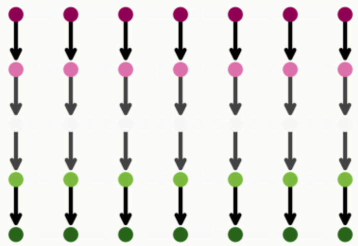
$$I_y = \left(I_x * \begin{bmatrix} & & \\ 1 & & 1 \\ & & \end{bmatrix} \right)$$

$$\boldsymbol{y} = 2\boldsymbol{B}_{dx}^{1}\,\boldsymbol{x}$$

$$I_y = \left(I_x * \begin{bmatrix} & & \\ -1 & & 1 \\ & & \end{bmatrix} \right)$$

Vertical direction

$\nabla \boldsymbol{\phi}_2$



$$\boldsymbol{y} = 2\boldsymbol{B}_{av}^{m}\,\boldsymbol{x}$$

$$I_y = \left(I_x * \begin{bmatrix} & 1 & \\ & & \\ & 1 & \end{bmatrix} \right)$$

$$\boldsymbol{y} = 2\boldsymbol{B}_{dx}^{m}\,\boldsymbol{x}$$

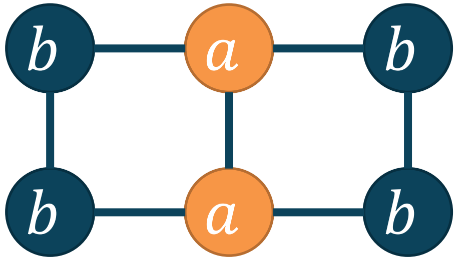
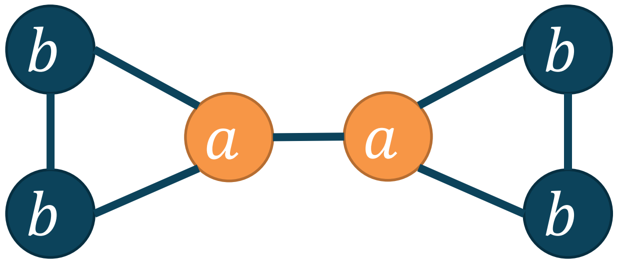
$$I_y = \left(I_x * \begin{bmatrix} & 1 & \\ & & \\ & -1 & \end{bmatrix} \right)$$

GNNs = CNNs

$$\boldsymbol{y} = \begin{pmatrix} w_1\boldsymbol{I} + 2w_2\boldsymbol{B}_{av}^{1} + 2w_3\boldsymbol{B}_{dx}^{1} \\ +2w_4\boldsymbol{B}_{av}^{m} + 2w_5\boldsymbol{B}_{dx}^{m} \end{pmatrix} \boldsymbol{x}$$

$$I_y = \left(I_x * \begin{bmatrix} & w_4 + w_5 & \\ w_2 - w_3 & w_1 & w_2 + w_3 \\ & w_4 - w_5 & \end{bmatrix} \right)$$

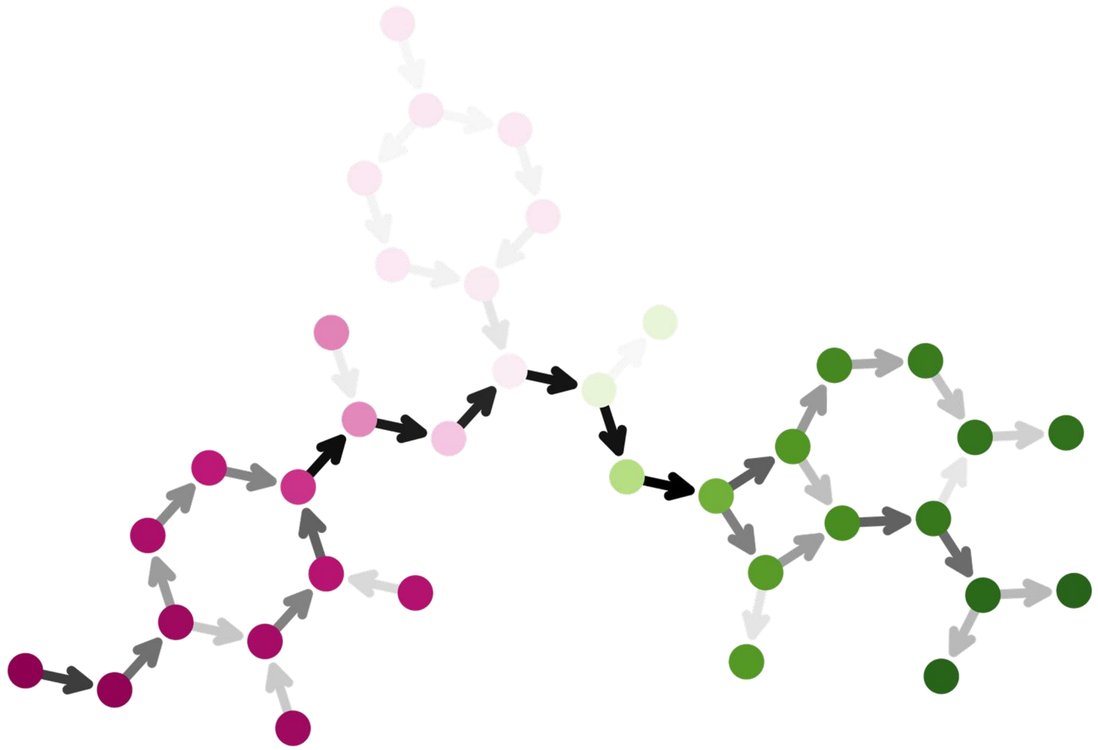
1-WL test for molecules

Aggregation matrix		
$\boldsymbol{A}$	$a + b \rightarrow b$ $a + 2b \rightarrow a$	$a + b \rightarrow b$ $a + 2b \rightarrow a$
$\boldsymbol{B}_{dx}^{1}$	$\lvert a - b \rvert \rightarrow b$ $0 \rightarrow a$	$\lvert a - b \rvert \rightarrow b$ $a + 2b \rightarrow a$
$\boldsymbol{B}_{av}^{1}$	$a \rightarrow b$ $b \rightarrow a$	$a \rightarrow b$ $.56a + .44b \rightarrow a$

Directional Graph Networks. Dominique Beaini*, Saro Passaro*, Vincent Létourneau, William L. Hamiton, Gabriele Corso, Pietro Liò
https://arxiv.org/abs/2010.02863

Reducing the over-smoothing and over-squashing

- By following the gradient of the eigenvectors (the arrows):
  - **No over-squashing.** The message can be sent efficiently from one side to the other side of the .
  - **No over-smoothing.** The message does not converge to a mean equilibrium



Generalization of data augmentation

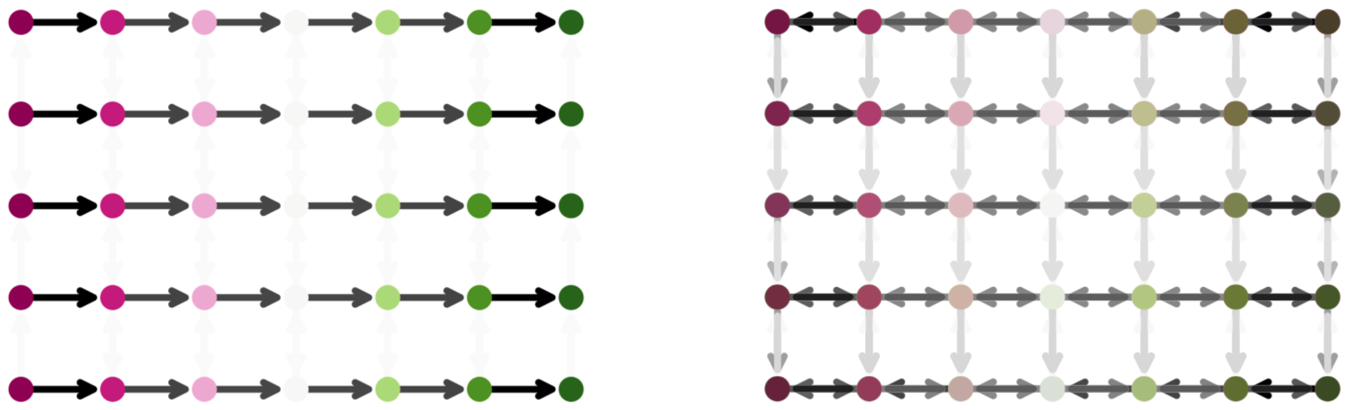
- Flipping: Changing the sign of the vector field

$$F_{flip} = -F$$

- Rotation: Linear combination of vector fields

$$\mathrm{F}_{\theta} = F_1 \cos\theta + F_2^{\perp} \sin\theta$$

- Distortion: Random fluctuation in the field or the potential



Ablation study

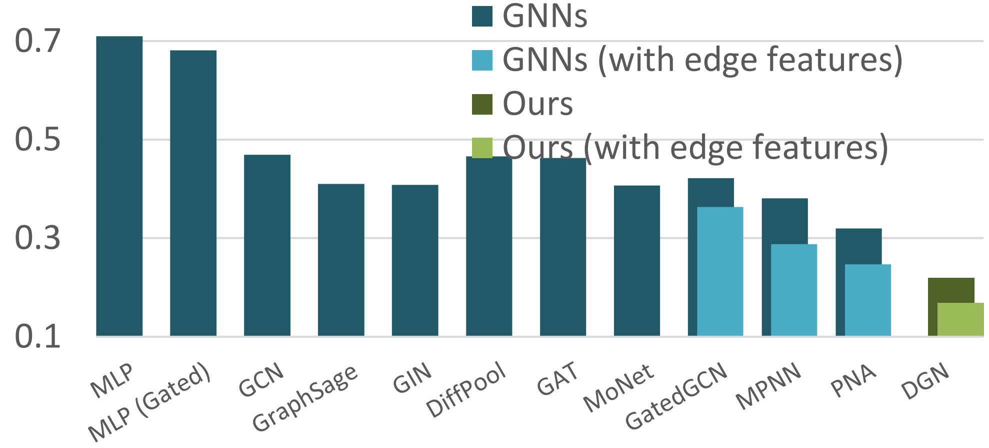
	ZINC			PATTERN		CIFAR10		MolHIV	MolPCBA	
Aggregators	Simple MAE	Complex MAE	Complex-E MAE	Simple % acc	Complex % acc	Simple % acc	Complex % acc	Simple % ROC-AUC	Complex % AP	Complex-E % AP
mean	0.316	0.353	0.262	80.77	83.34	55.9	62.8	75.1	26.04	26.38
mean pos_1	0.349	0.332	0.297	80.76	83.74			75.8	26.97	27.50
mean pos_1 pos_2	0.344	0.330	0.284	84.51	81.25			76.1	26.03	25.65
mean dx_1	0.296	**0.233**	**0.191**	84.22	83.44			78.0	26.79	**27.91**
mean dx_1 dx_2	0.337	0.271	0.205	81.61	86.62	52.9	**69.8**	76.5	**27.16**	26.55
mean av_1	0.317	0.332	0.276	84.54	83.21			78.4	25.97	26.66
mean av_1 av_2	0.367	0.332	0.260	85.12	85.38	**60.6**	65.1	77.1	25.61	26.67
mean dx_1 av_1	**0.290**	0.245	0.192	**85.17**	**86.68**			**79.0**	26.40	27.47

Best

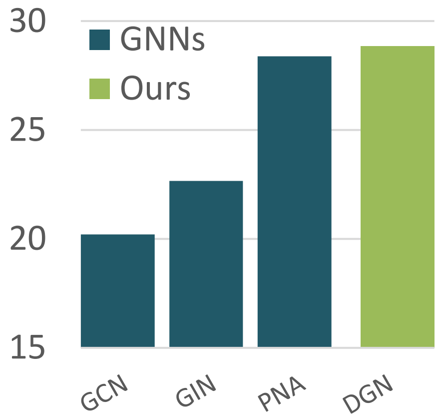
Worst

Results compared to the literature

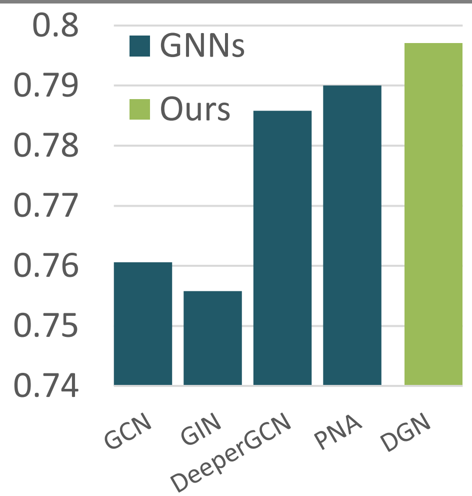
ZINC (MAE, lower is better)



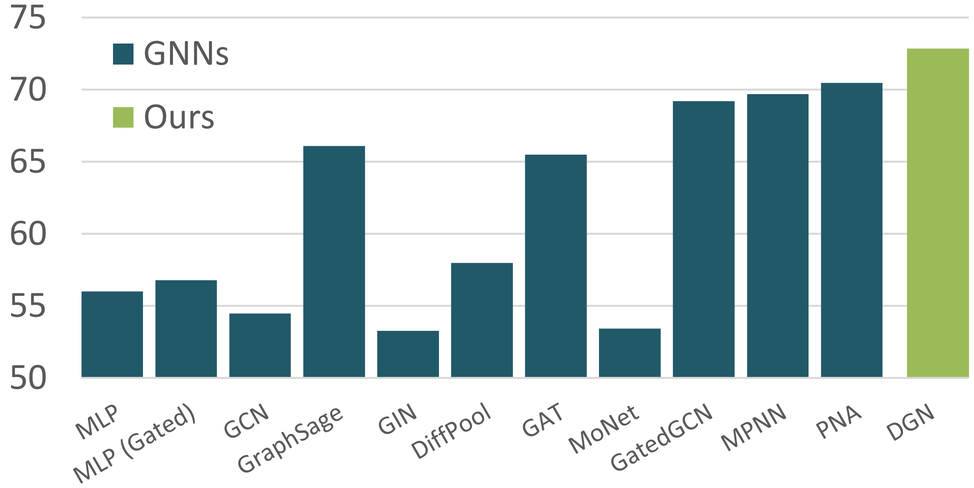
MolPCBA (Average Precision, higher is better)



MolHiv (ROC-AUC, higher is better)



CIFAR10 (Accuracy, higher is better)



Directional Graph Networks

Dominique Beaini, Saro Passaro*, Vincent Létourneau, William L. Hamilton, Gabriele Corso, Pietro Liò*

Dominique Beaini*

Saro Passaro*

Follow us on Twitter

@valence_ai
@dom_beaini
@Saro2000
@vincentmillions

Valence

UNIVERSITY OF CAMBRIDGE

Mila