Introduction to Geometric Deep Learning

Nick Ioannidis

June 14, 2021

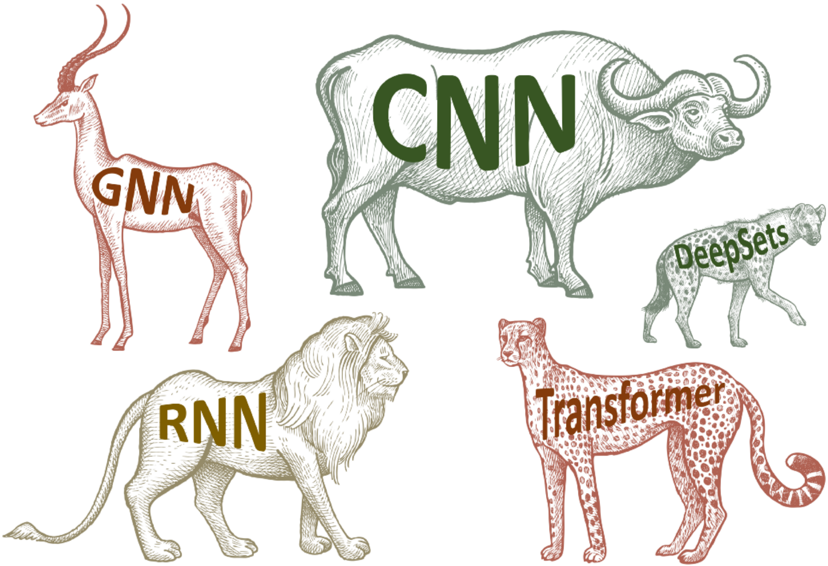
Historical Background of Geometry

- Geometry synonymous with Euclid until early 19th century
- Lobachevsky, Bolyai, Gauss, Riemann construct first examples of non-Euclidean geometry
- In 1872 Felix Klein publishes the Erlangen program that formalizes geometries into groups

Symmetry in Physics

Conservation laws derived from the first principles of symmetry (Noether's Theorem).

Unification on NN architectures

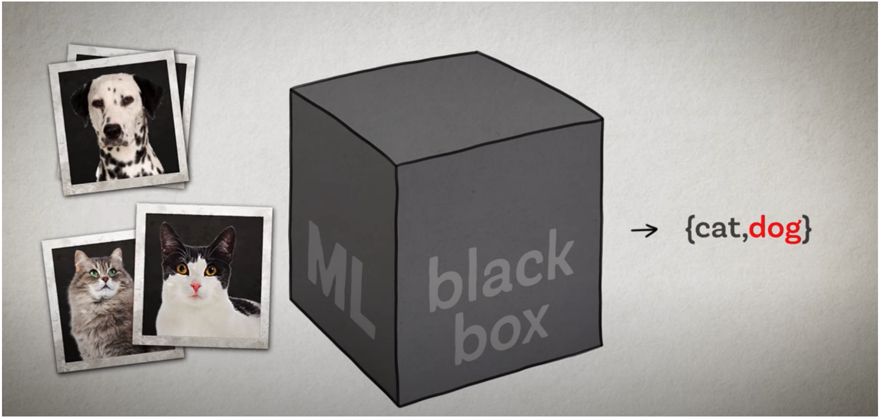


Goal of Geometric Deep Learning 2017

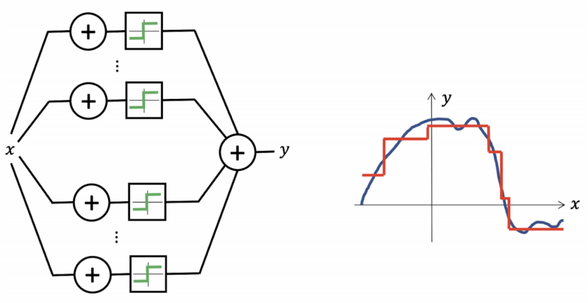
- Introduce an umbrella term for emerging techniques attempting to generalize deep neural models to non-Euclidean domains

Goal of Geometric Deep Learning 2021

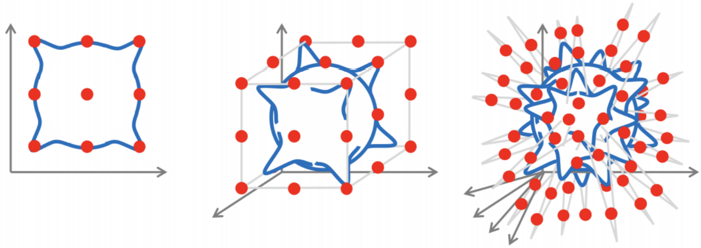
- Provide a common mathematical framework to derive some of the successful NN architectures
- Create a constructive procedure to build future architectures


ML
black
box
→ {cat,dog}

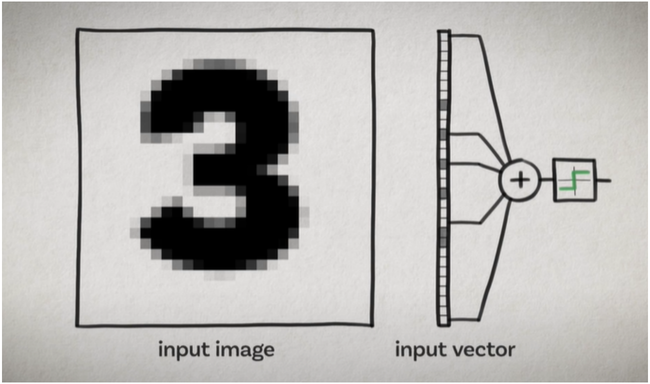
Neural Networks as Universal Approximatros



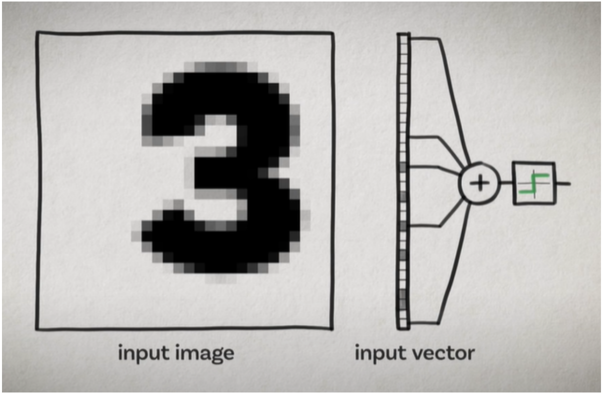
Curse of Dimensionality



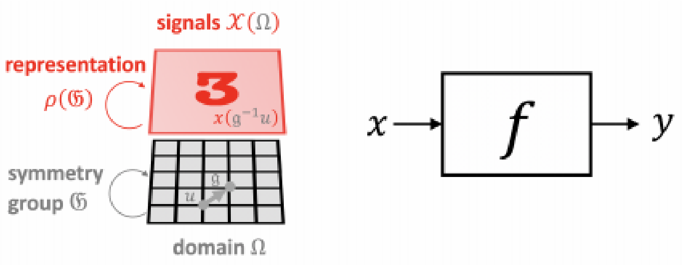
Curse of Dimensionality



Curse of Dimensionality



Geometric Priors



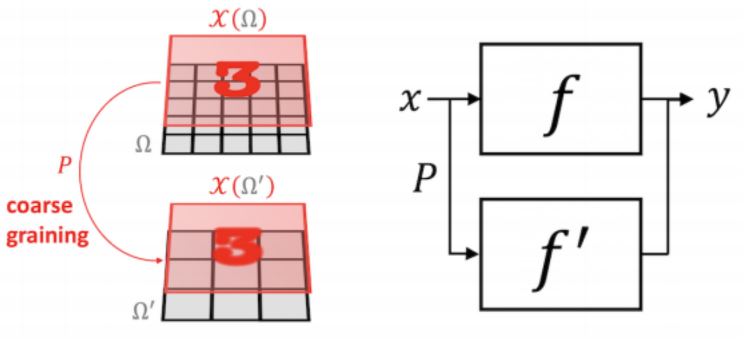
Group Invariant

A function $f : \mathcal{X}(\Omega) \to \mathcal{Y}$ is $\mathfrak{G}$-*invariant* if $f(\rho(\mathfrak{g})x) = f(x)$ for all $\mathfrak{g} \in \mathfrak{G}$ and $x \in \mathcal{X}(\Omega)$, i.e., its output is unaffected by the group action on the input.

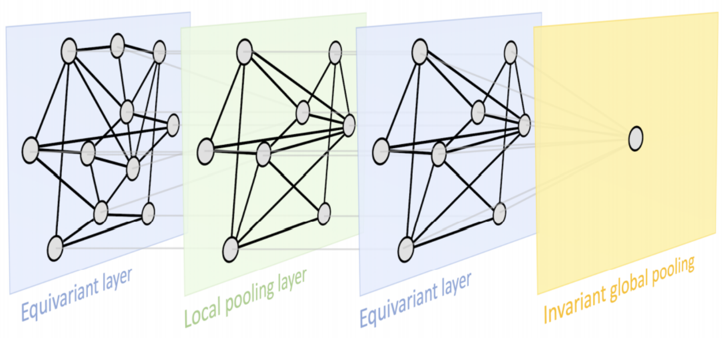
Group Equivariant

A function $f : \mathcal{X}(\Omega) \rightarrow \mathcal{X}(\Omega)$ is $\mathfrak{G}$-*equivariant* if $f(\rho(\mathfrak{g})x) = \rho(\mathfrak{g})f(x)$ for all $\mathfrak{g} \in \mathfrak{G}$, i.e., group action on the input affects the output in the same way.

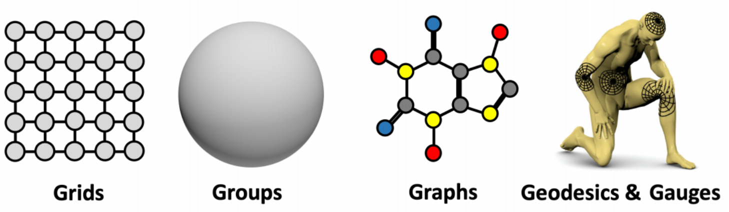
Scale seperation

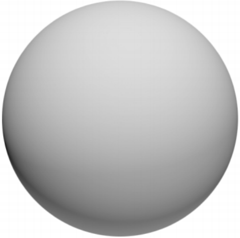


Geometric Deep Learning Blueprint



The "5G" of Geometric Deep Learning

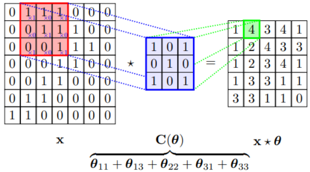




Groups

Convolution in Euclidean Space

$$(x \star \theta)(u) = \langle x, S_u \theta \rangle = \int_{\mathbb{R}} x(v)\theta(u+v)\mathrm{d}v$$

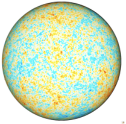


Group Convolution

General case of group convolution
$(x \star \theta)(\mathfrak{g}) = \langle x, \rho(\mathfrak{g})\theta \rangle = \int_\Omega x(u)\theta\left(\mathfrak{g}^{-1}u\right) \mathrm{d}u$

Group equivariance
$(x \star \theta)(\mathfrak{g}) = \langle x, \rho(\mathfrak{g})\theta \rangle = \int_\Omega x(u)\theta\left(\mathfrak{g}^{-1}u\right) \mathrm{d}u$

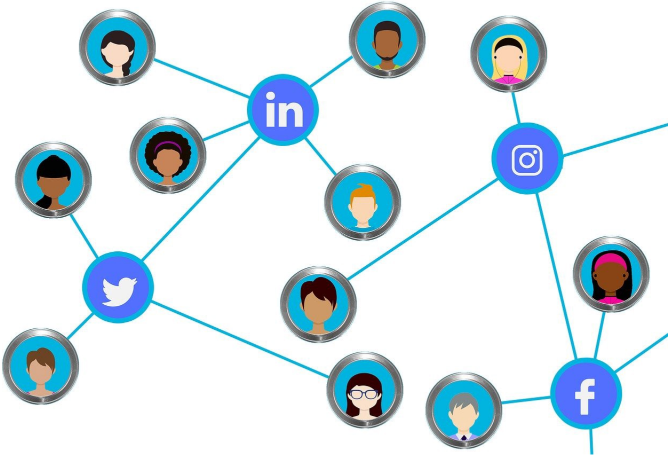
Convolution applied on signal domain
$(x \star \theta)(\mathrm{R}) = \int_{\mathbb{S}^2} x(\mathrm{u})\theta\left(\mathrm{R}^{-1}\mathrm{u}\right) \mathrm{d}\mathrm{u}$



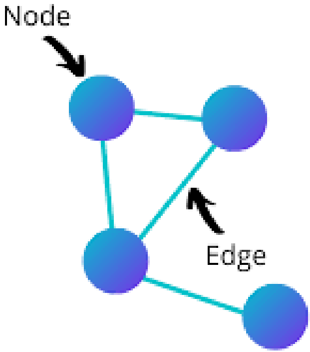
Cosmic microwave background radiation. Signal on $\mathbb{S}^2$.

Convolution applied on $SO(3)$ group
$((x \star \theta) \star \phi)(\mathrm{R}) = \int_{\mathrm{SO}(3)} (x \star \theta)(\mathrm{Q})\phi\left(\mathrm{R}^{-1}\mathrm{Q}\right) \mathrm{d}\mathrm{Q}$

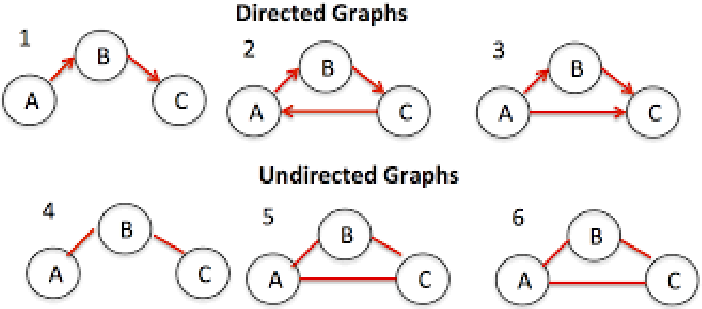
Graphs



Graphs

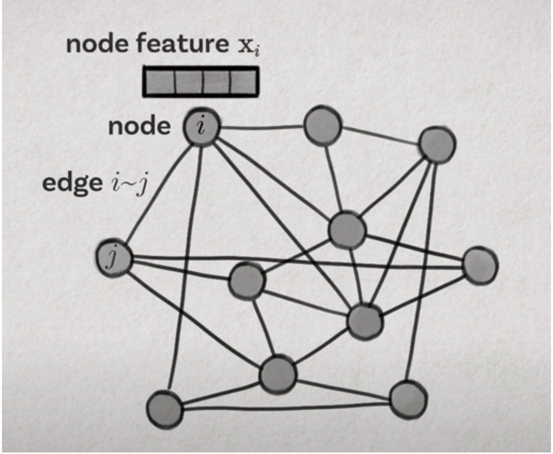


Graphs

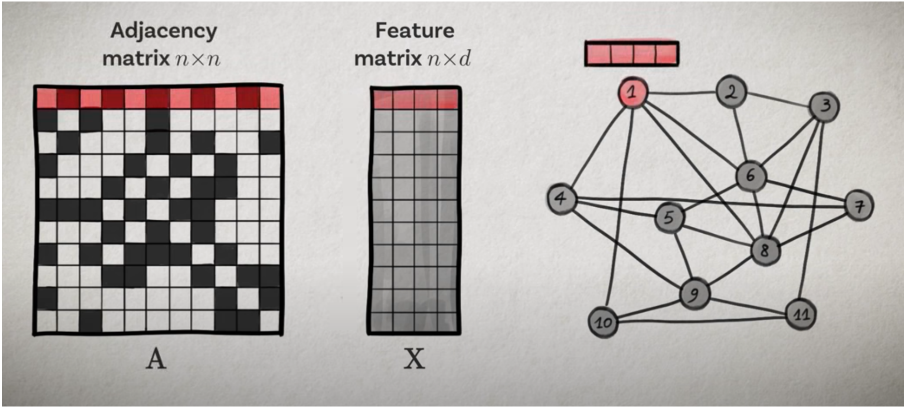


Directed Graphs
1
B
A
C
2
B
A
C
3
B
A
C
Undirected Graphs
4
B
A
C
5
B
A
C
6
B
A
C

Graphs

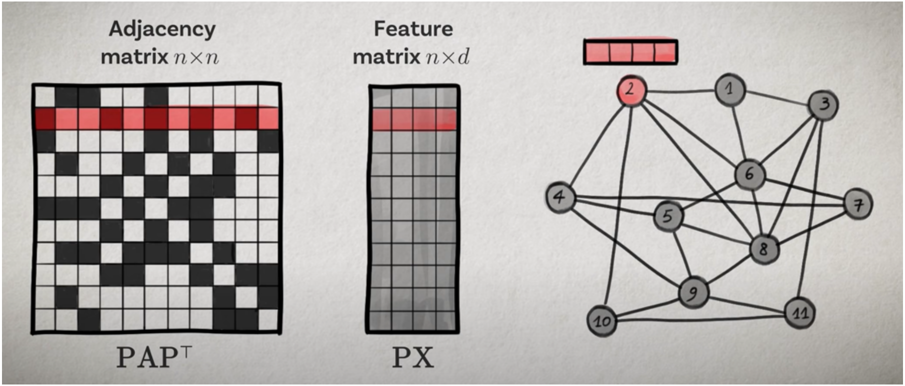


Graphs



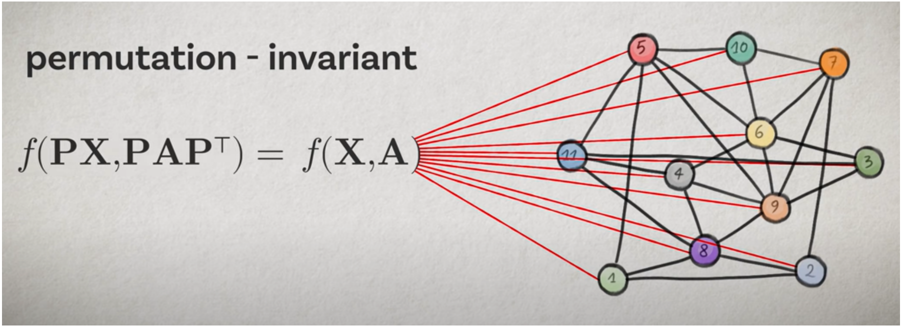
Adjacency
matrix $n \times n$
A
Feature
matrix $n \times d$
X
1
2
3
4
5
6
7
8
9
10
11

Graphs

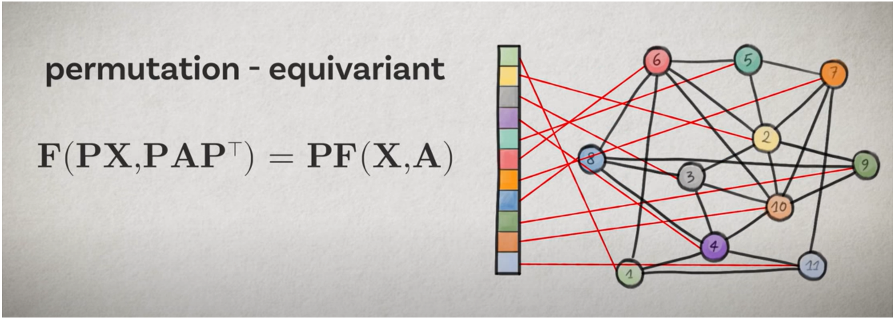


Adjacency
matrix $n \times n$
Feature
matrix $n \times d$
2
1
3
6
4
7
5
8
9
10
11
$PAP^\top$
PX

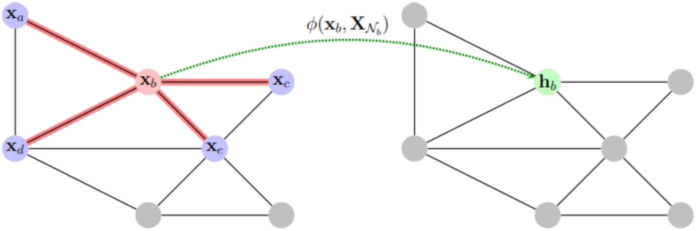
Graphs


permutation - invariant
$f(\mathbf{PX},\mathbf{PAP}^\top) = f(\mathbf{X},\mathbf{A})$
5
10
7
6
11
3
4
9
8
1
2

Graphs

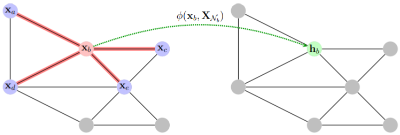

permutation - equivariant
$\mathbf{F}(\mathbf{PX},\mathbf{PAP}^\top) = \mathbf{PF}(\mathbf{X},\mathbf{A})$
6
5
7
2
8
9
3
10
4
1
11

Graphs

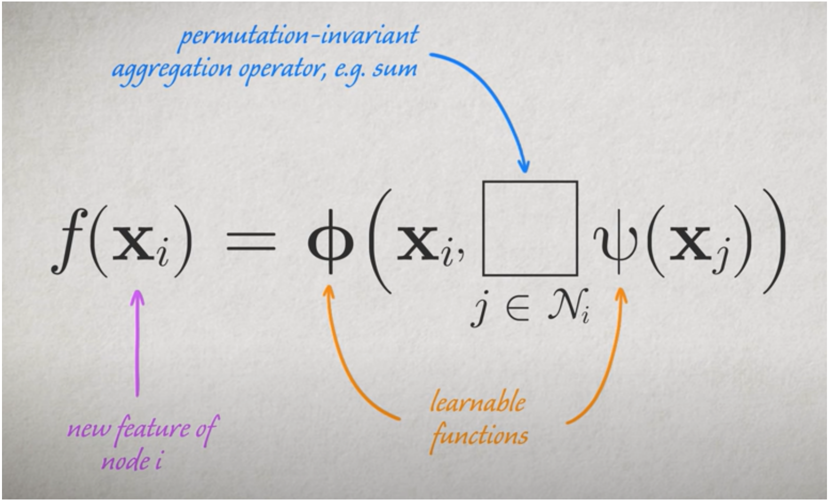

$\mathbf{x}_a$
$\mathbf{x}_b$
$\mathbf{x}_c$
$\mathbf{x}_d$
$\mathbf{x}_e$
$\phi(\mathbf{x}_b, \mathbf{X}_{\mathcal{N}_b})$
$\mathbf{h}_b$

Graphs

$$F(X, A) = \begin{bmatrix} - & \phi\left(x_1, X_{\mathcal{N}_1}\right) & - \\ - & \phi\left(x_2, X_{\mathcal{N}_2}\right) & - \\ \vdots & & \\ - & \phi\left(x_n, X_{\mathcal{N}_n}\right) & - \end{bmatrix}$$

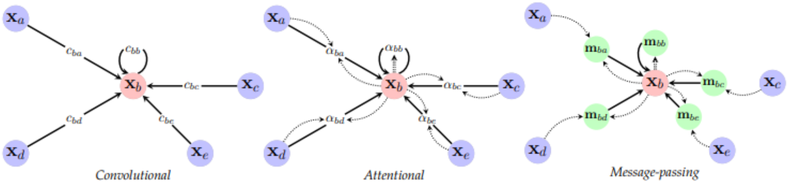


GNNs



permutation-invariant
aggregation operator, e.g. sum
$f(\mathbf{x}_i) = \phi\left(\mathbf{x}_i, \square_{j \in \mathcal{N}_i} \psi(\mathbf{x}_j)\right)$
new feature of
node i
learnable
functions

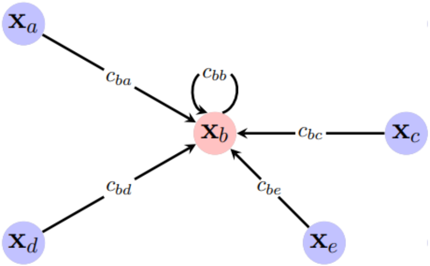
GNNs



$\mathbf{x}_a$
c_{ba}
c_{bb}
$\mathbf{x}_b$
c_{bc}
$\mathbf{x}_c$
c_{bd}
c_{be}
$\mathbf{x}_d$
$\mathbf{x}_e$
Convolutional
$\mathbf{x}_a$
α_{ba}
α_{bb}
$\mathbf{x}_b$
α_{bc}
$\mathbf{x}_c$
α_{bd}
α_{be}
$\mathbf{x}_d$
$\mathbf{x}_e$
Attentional
$\mathbf{x}_a$
$\mathbf{m}_{ba}$
$\mathbf{m}_{bb}$
$\mathbf{x}_b$
$\mathbf{m}_{bc}$
$\mathbf{x}_c$
$\mathbf{m}_{bd}$
$\mathbf{m}_{be}$
$\mathbf{x}_d$
$\mathbf{x}_e$
Message-passing

GNN - Convolutional

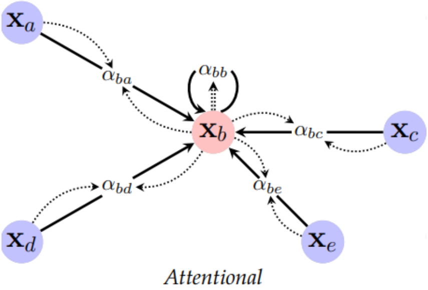
$$\mathsf{f}_i = \phi\left(\mathsf{x}_i, \bigoplus_{j \in \mathcal{N}_i} c_{ij} \psi\left(\mathsf{x}_j\right)\right)$$



Convolutional

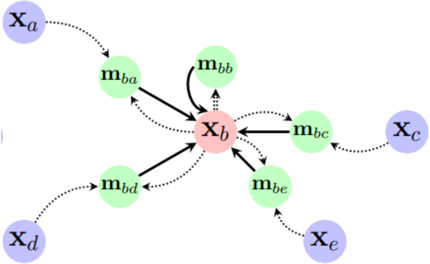
GNN - Attentional

$$\mathsf{f}_i = \phi\left(\mathsf{x}_i, \bigoplus_{j \in \mathcal{N}_i} a\left(\mathsf{x}_i, \mathsf{x}_j\right) \psi\left(\mathsf{x}_j\right)\right)$$



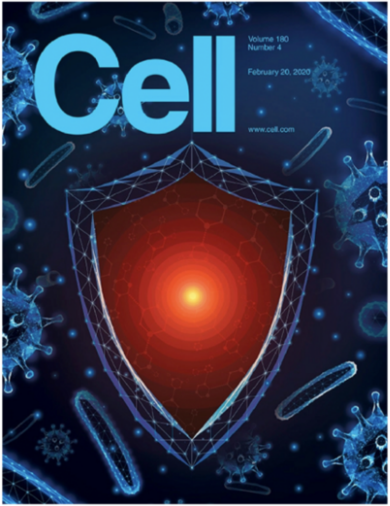
GNN - Message-Passing

$$f_i = \phi\left(x_i, \bigoplus_{j \in \mathcal{N}_i} \psi\left(x_i, x_j\right)\right)$$

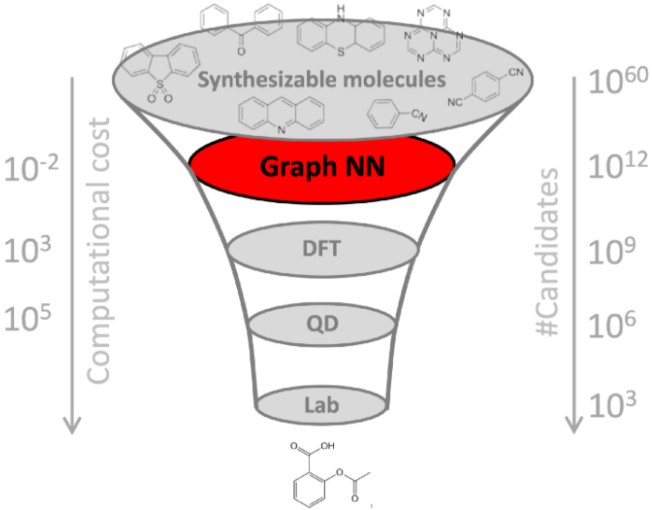


Message-passing

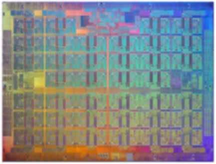
Applications of Graphs in Natural Sciences



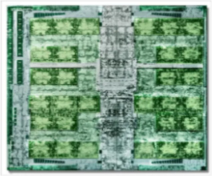
GNNs in Drug Discovery



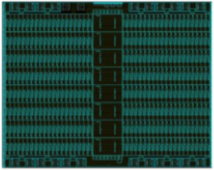
Hardware Advances



CPU
Scalar



GPU
Vector



IPU
Graph

Thank you for your time