Graph Convolutional Networks

Krishnendu Mukherjee

Advisor: Dr. Yamil J. Colón

ML Reading Group

August 2nd, 2021

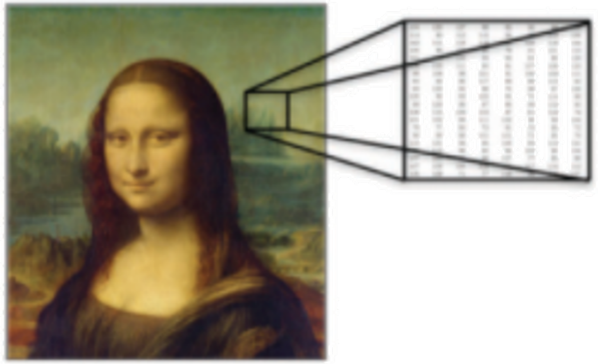
COLÓN GROUP

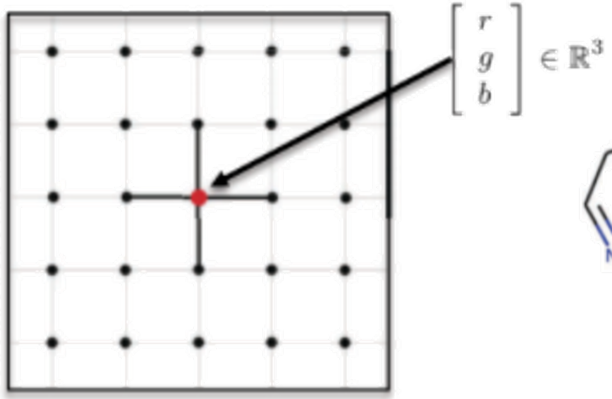
UNIVERSITY OF NOTRE DAME

Graph Domains

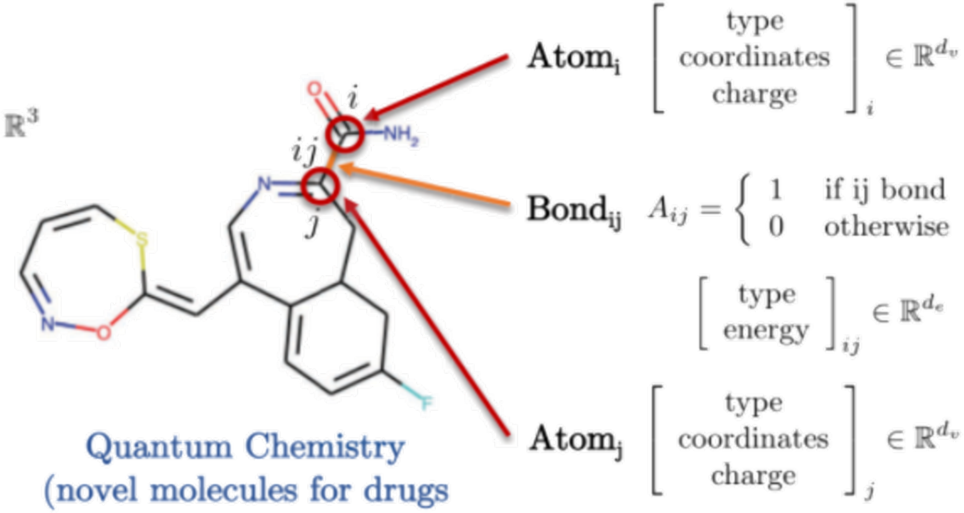
Data Domain

Graph Domain

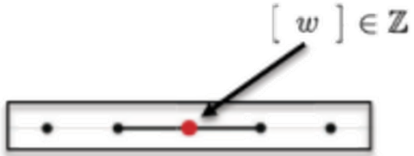




2D grid



Quantum Chemistry
(novel molecules for drugs and materials)

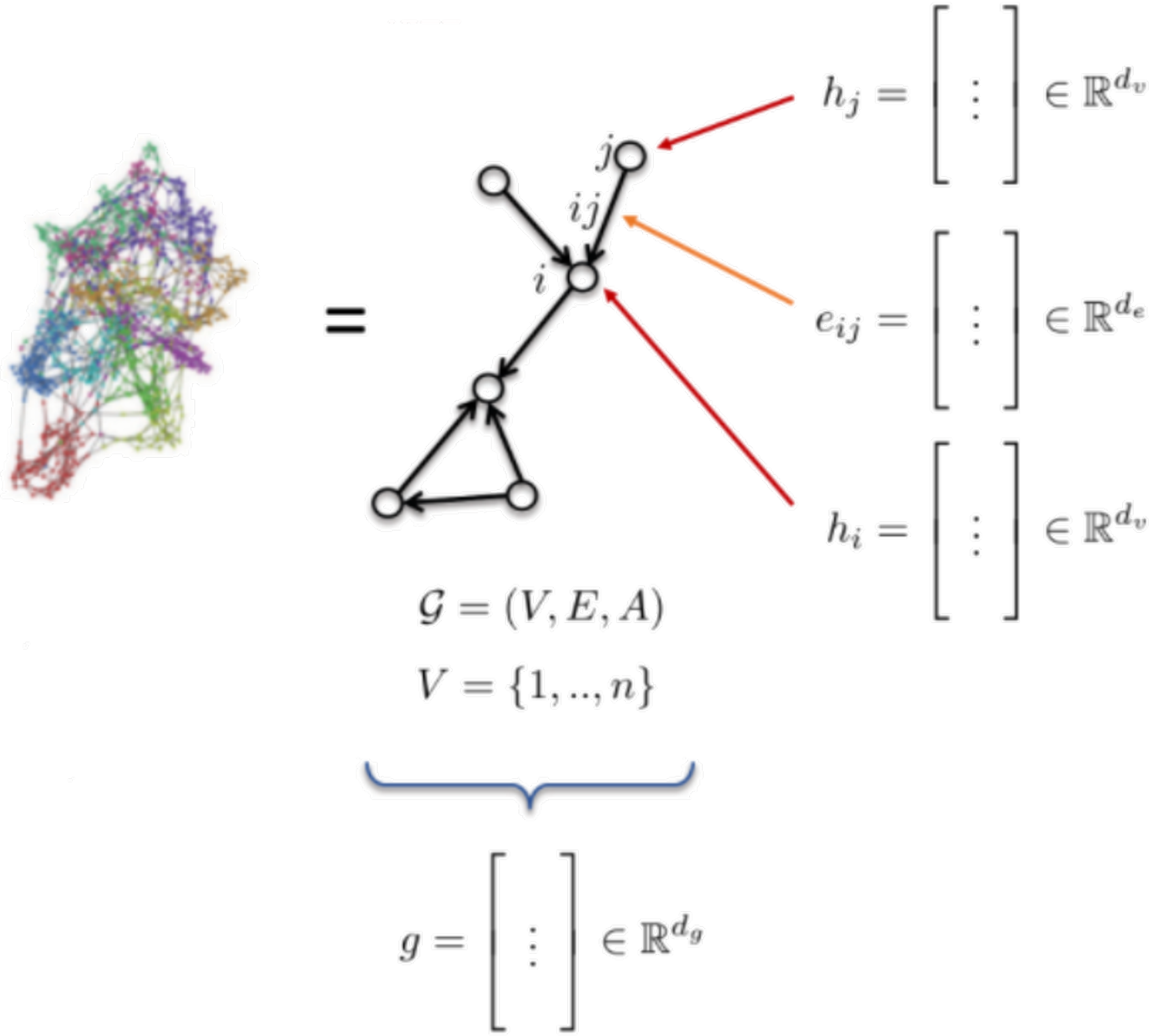


1D grid

Graph Definitions

- Graph G is defined as:
  - Vertices V
  - Edges E
  - Adjacency Matrix A

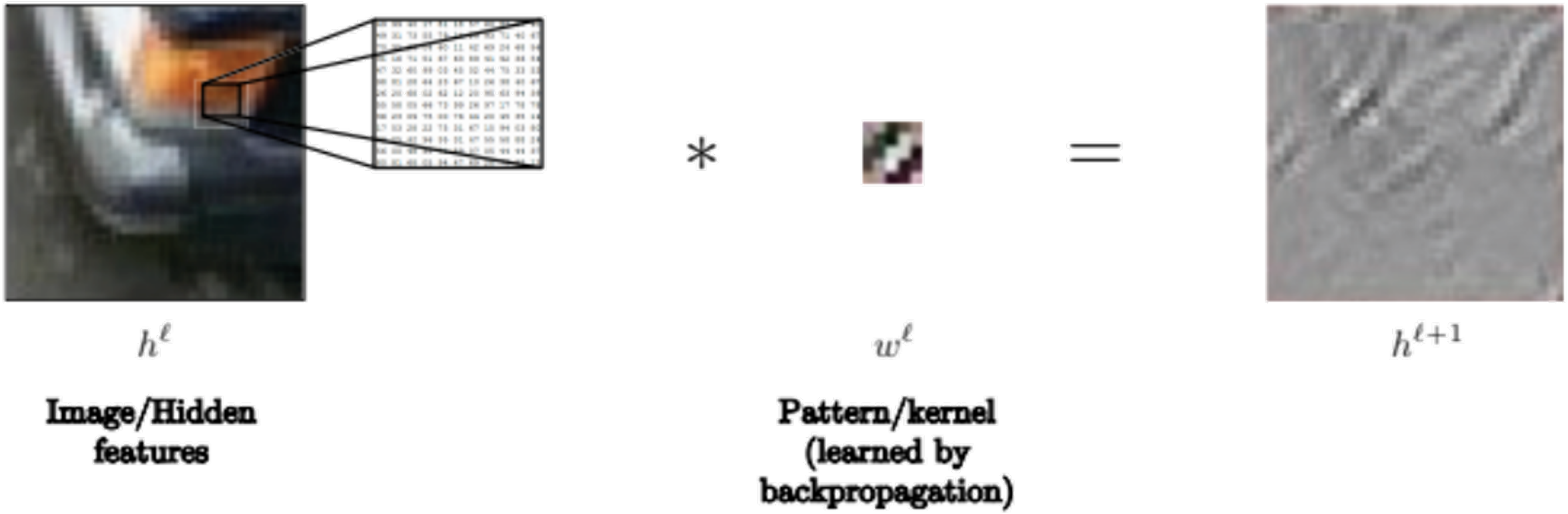
- Graph features are:
  - Node features: h_i, h_j (atom type)
  - Edge features: e_{ij}
  - Graph features: g (molecular energy)



What is a Convolution

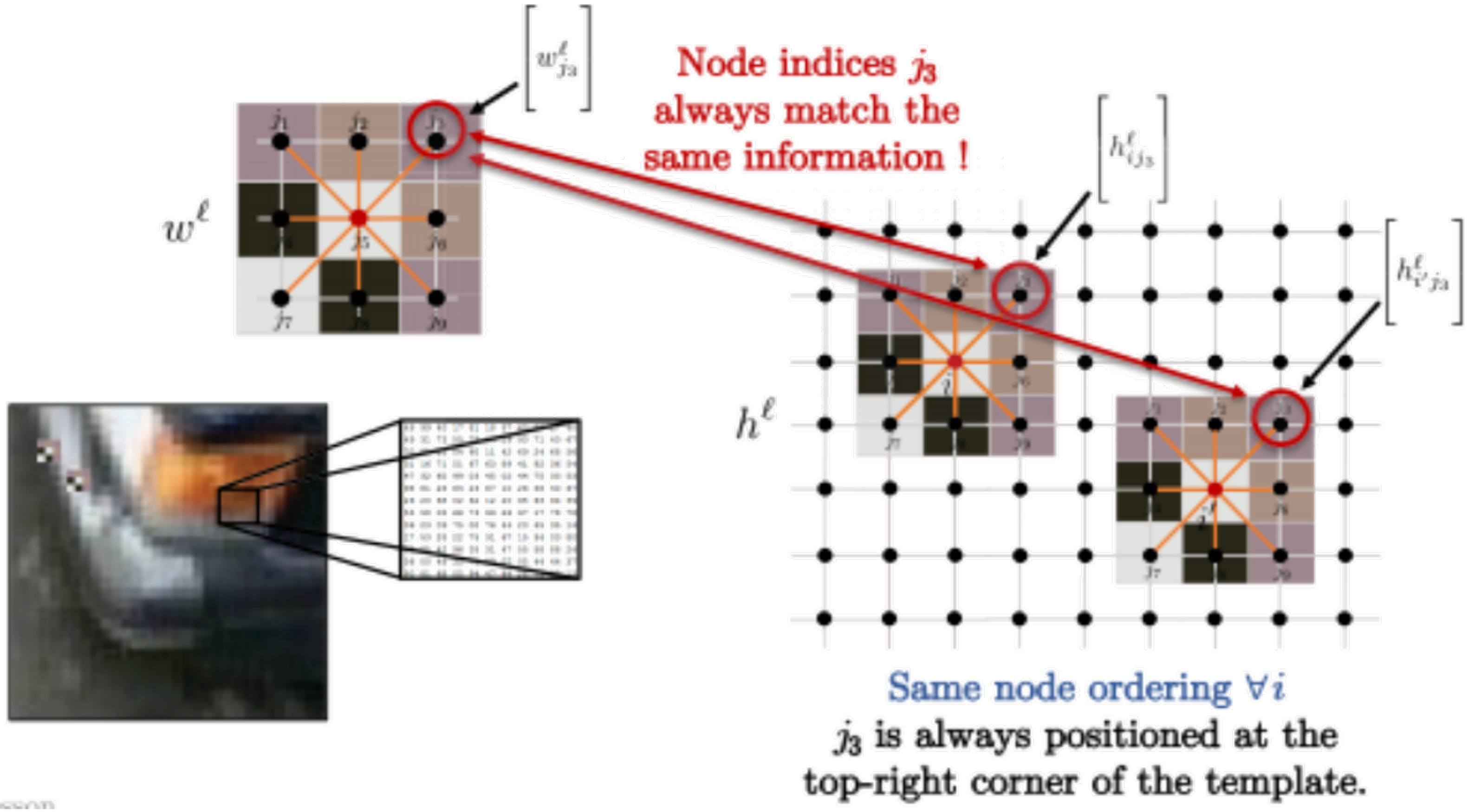
$$h^{l} \times w^{l} = h^{l+1}$$

Features at l^{th} layer	Pattern/Kernel (learned by backpropagation)	Convoluted feature at $(l+1)^{th}$ layer



Convolution as template matching

- The template/kernel w_l is overlayed (as defined by the strides) across the image from top-left to bottom right
- Different kernel convolutions are done to represent an image in terms of kernels/templates
- Kernels might become combination of kernels as we proceed just like ANNs

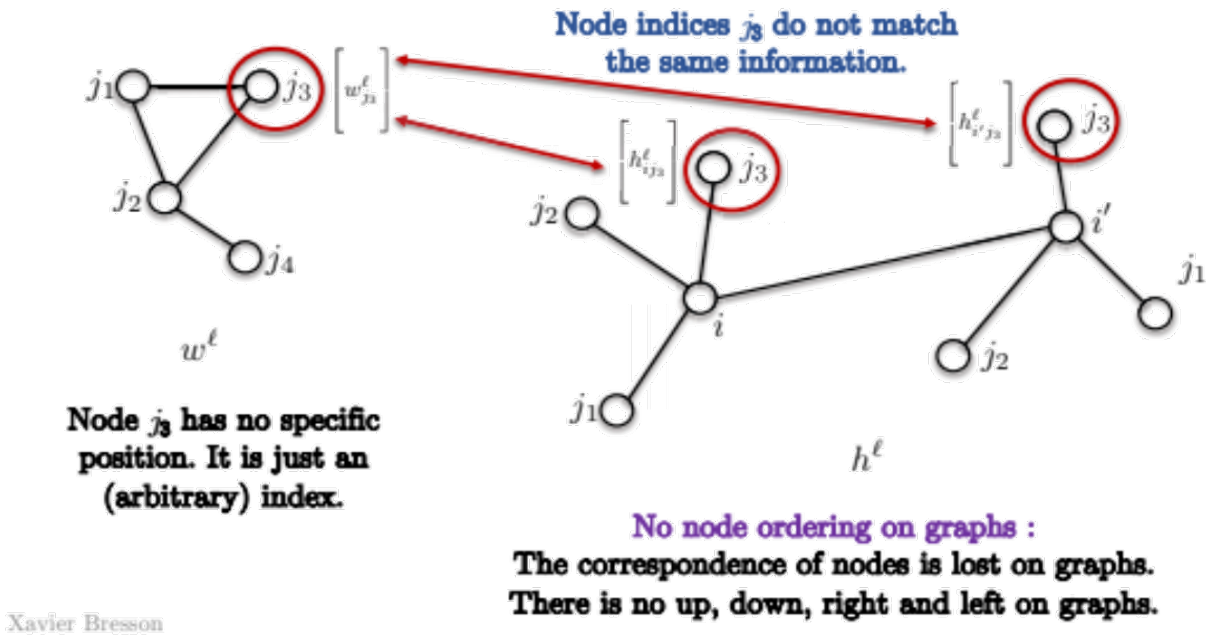


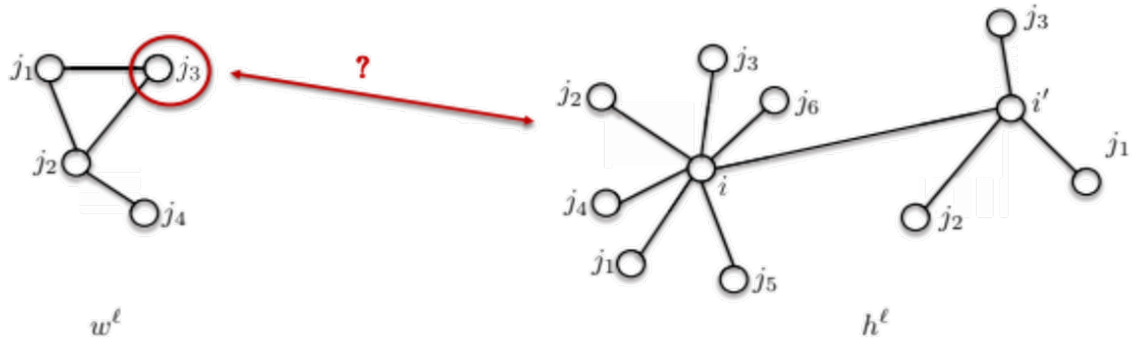
COLÓN GROUP

UNIVERSITY OF NOTRE DAME

Problems with extending template matching to Graphs

- No ordering of nodes
- Neighborhood sizes are different





Convolution theorem to graphs

$$\mathcal{F}(w * h) = \mathcal{F}\,(w) \odot \mathcal{F}\,(h)$$

- Use convolution theorem for spectral ConvNets

$$
\begin{aligned}
\underset{n\times 1}{w * h} &= \underbrace{\mathcal{F}^{-1}}_{\Phi}\big(\underbrace{\mathcal{F}(w)}_{\Phi^T w = \hat{w}} \odot \underbrace{\mathcal{F}(h)}_{\Phi^T h}\big) \\
&= \underset{n\times n}{\Phi}\left(\underset{n\times 1}{\hat{w}} \odot \underset{n\times 1}{\Phi^T h}\right) \\
&= \Phi\left(\underset{n\times n}{\hat{w}(\Lambda)}\underset{n\times 1}{\Phi^T h}\right) \\
&= \Phi\hat{w}(\Lambda)\Phi^T h \\
&= \hat{w}(\underbrace{\Phi\Lambda\Phi^T}_{\Delta})h \\
&= \underset{n\times n}{\hat{w}(\Delta)}\ \underset{n\times 1}{h}
\end{aligned}
$$

Convolution Theorem

$$w * h \ = \mathcal{F}^{-1}(\mathcal{F}\,(w) \odot \mathcal{F}\,(h))$$

$$\Delta\,(n \times n) = \ \Phi^T \Lambda\, \Phi$$

- Δ is the graph Laplacian operator
- Φ contains column vectors and those are also called Fourier functions
- Λ is a diagonal matrix with Laplacian Eigenvalues

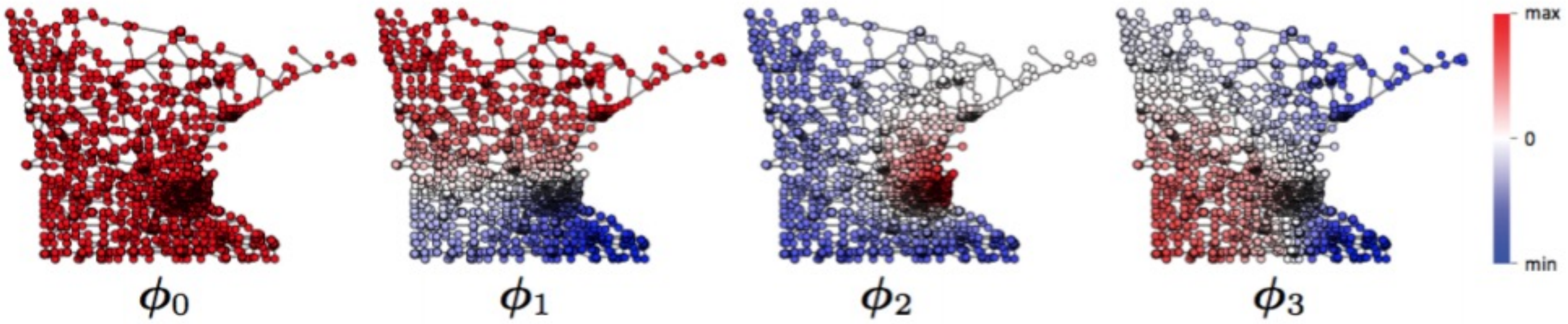
Fourier functions in Graph domain

- Use convolution theorem for spectral ConvNets

Convolution Theorem

$$w * h = \mathcal{F}^{-1}(\mathcal{F}(w) \odot \mathcal{F}(h))$$

$$\Delta\,(n \times n) = \Phi^T \Lambda\, \Phi$$



First Laplacian eigen vectors of a graph

Physical interpretation of terms in the Convolution Theorem

- Use convolution theorem for spectral ConvNets

$$\mathcal{G} = (V, E, A) \rightarrow \underset{n\times n}{\Delta} = \mathrm{I} - D^{-1/2} A D^{-1/2}$$

$$\text{where } \underset{n\times n}{D} = \mathrm{diag}\left(\sum_{j\neq i} A_{ij}\right)$$

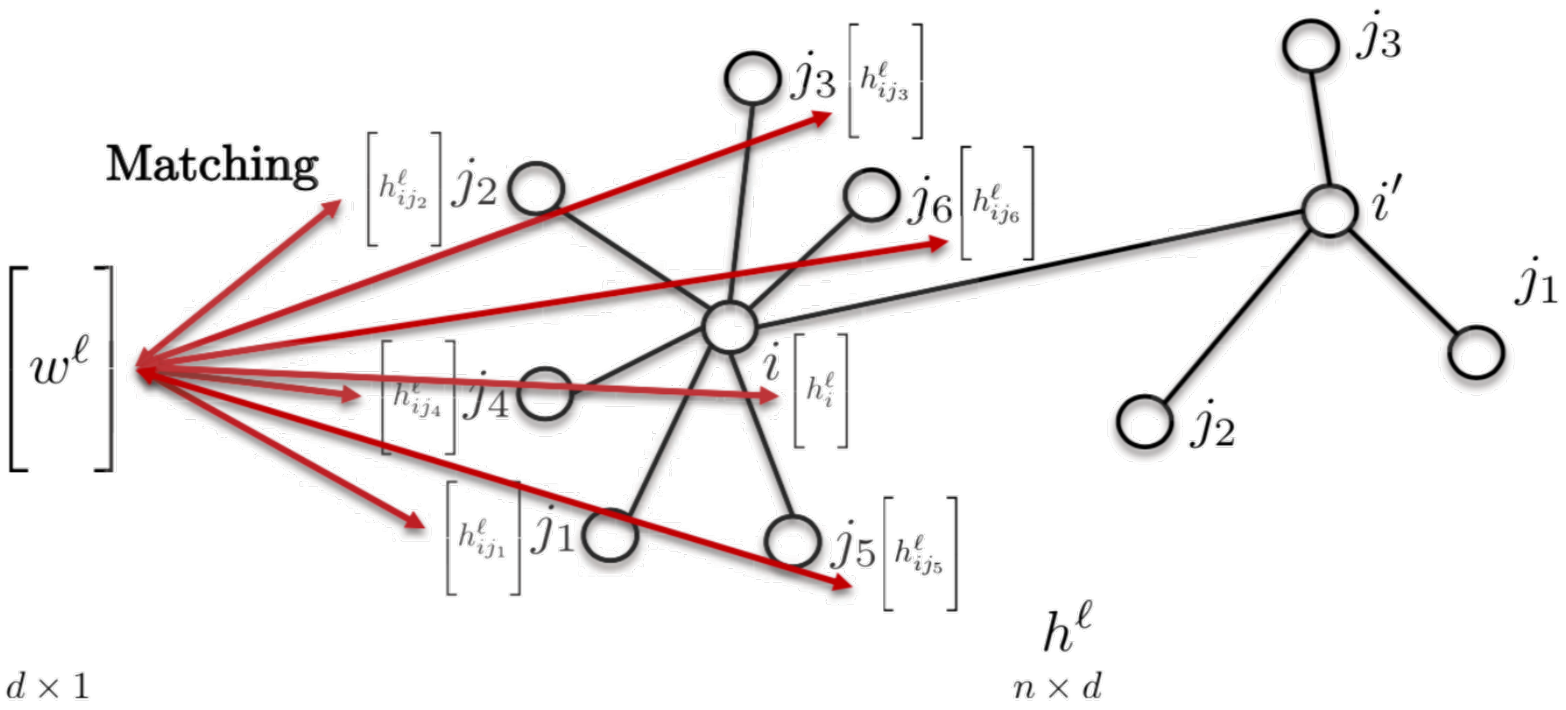
- Matrix A is the adjacency matrix.
- I is the identity matrix
- D is the diagonal matrix, and each element on the diagonal is the degree of the node

The Laplacian shows → the measurement of smoothness of graph or the different between local values h_i and its neighborhood values of h_j's.

COLÓN GROUP UNIVERSITY OF NOTRE DAME

Template matching for Spatial Graph ConvNets

- Using one template vector w_l instead of matrix (like 3 x 3)
- For each feature (in d) we would have a vector/template w



$$\underset{\text{scalar}}{h_i^{\ell+1}} = \eta\Big(\sum_{j \in \mathcal{N}_i} \underbrace{\langle \overset{d\times 1}{w^\ell}, \overset{d\times 1}{h_{ij}^\ell} \rangle}_{(h_{ij}^\ell)^T w^\ell} \Big)$$

One feature

$$\underset{d\times 1}{h_i^{\ell+1}} = \eta\Big(\sum_{j \in \mathcal{N}_i} \underset{d\times d}{W^\ell} \underset{d\times 1}{h_{ij}^\ell} \Big)$$

d features

$$\underset{n\times d}{h^{\ell+1}} = \eta\big(A h^\ell W^\ell \big)$$

$d \times d$
$n \times d$
$n \times n$

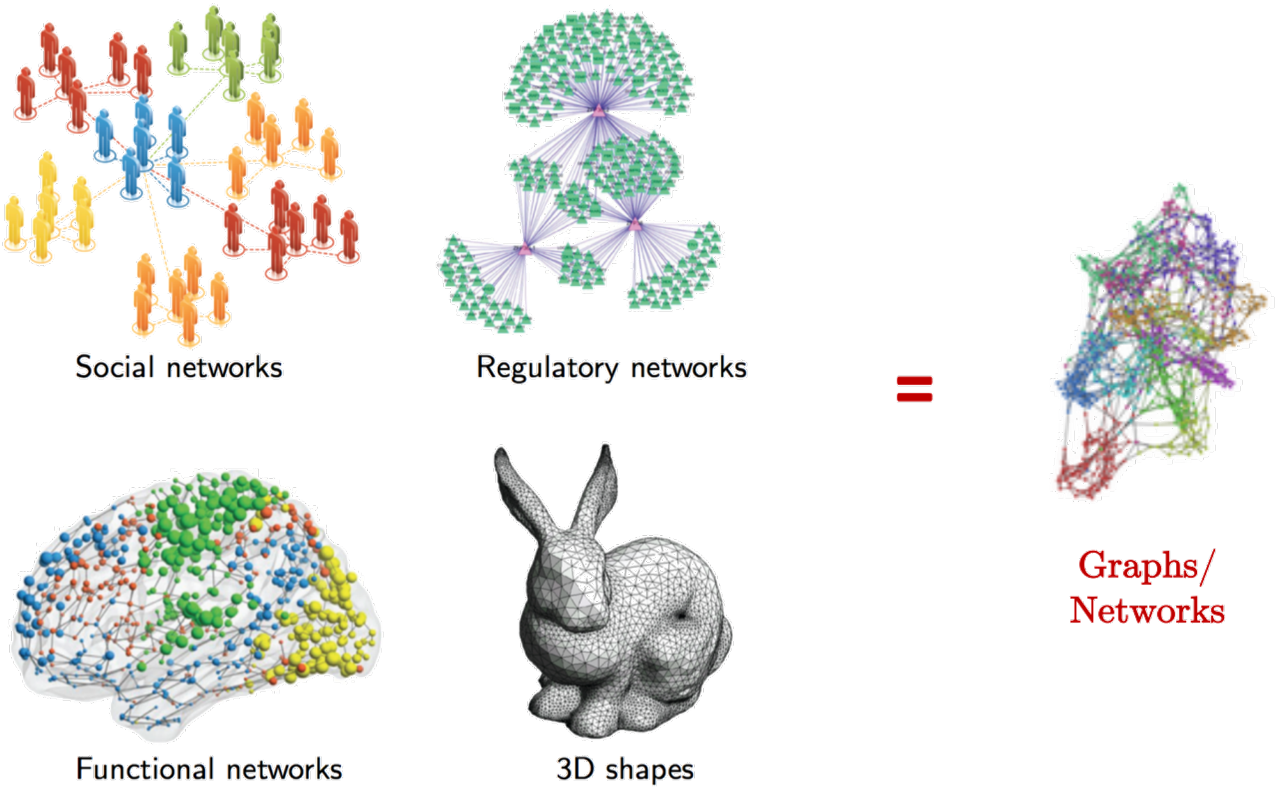
Vectorial representation

COLÓN GROUP

UNIVERSITY OF NOTRE DAME

Conclusion

- GCNs generalize CNNs to data on graphs
- The convolution operator needs to be redesigned for graphs → 2 ways, spectral GCNs (Convolution theorem) (cheaper) and Spatial GCNs (expensive)



THANKS FOR YOUR ATTENTION

&

ANY QUESTIONS?

COLÓN GROUP

UNIVERSITY OF NOTRE DAME