Graph Neural Networks : an introduction

Benoit Gaüzère,
in close collaboration with Sébastien Adam, Pierre Héroux, Muhammet Balcilar, Guillaume Renton, Paul Honeine and Clément Gledel.

INSA Rouen Normandie - Laboratoire LITIS

18/02/2021

Outline I

The Deep Learning Era

The NN rise

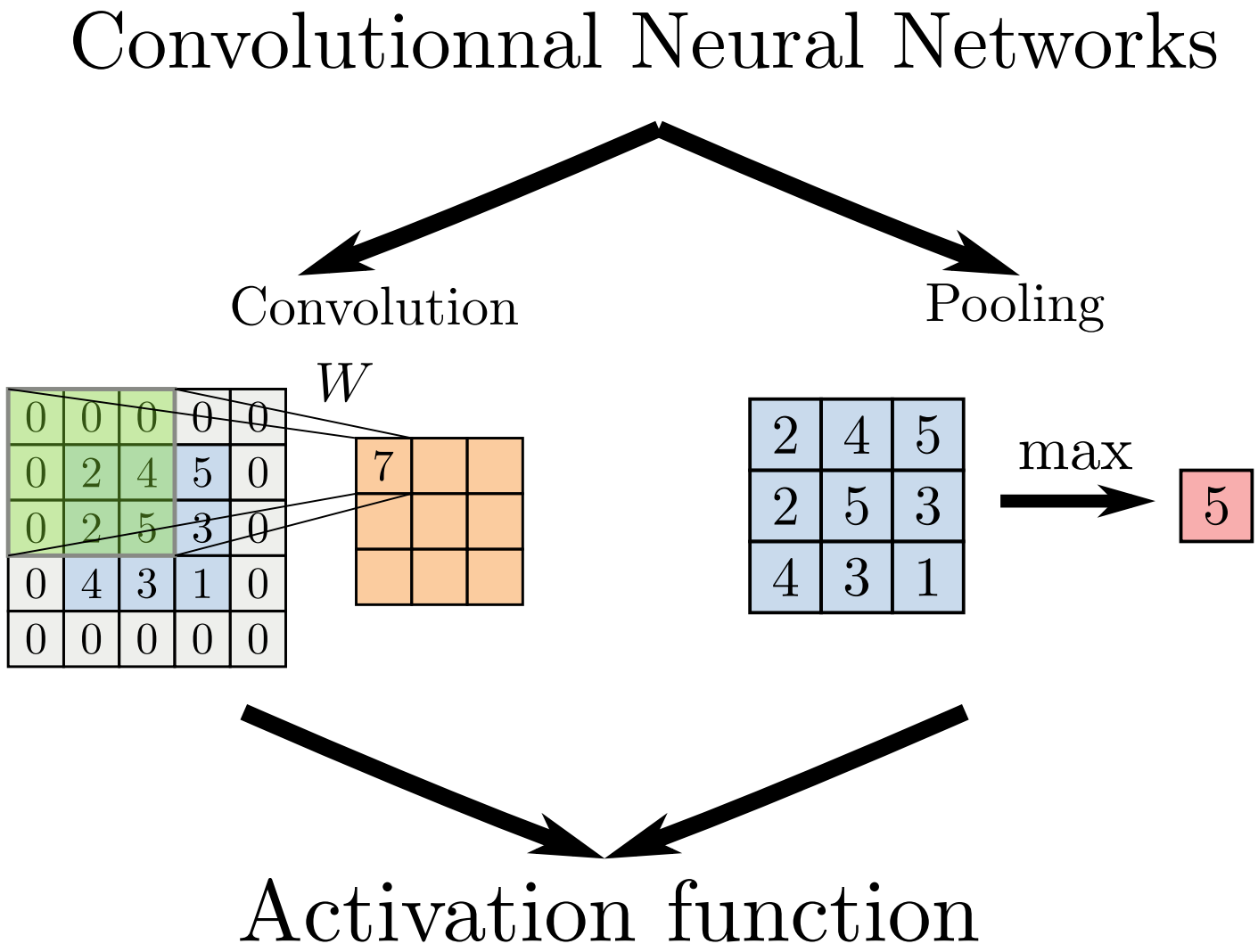
Since 2012, we observe the rise of NN based methods :

- Huge datasets
- High computationnal capacities (GPU, ...)
- Representation learning $\geq$ handcrafted features

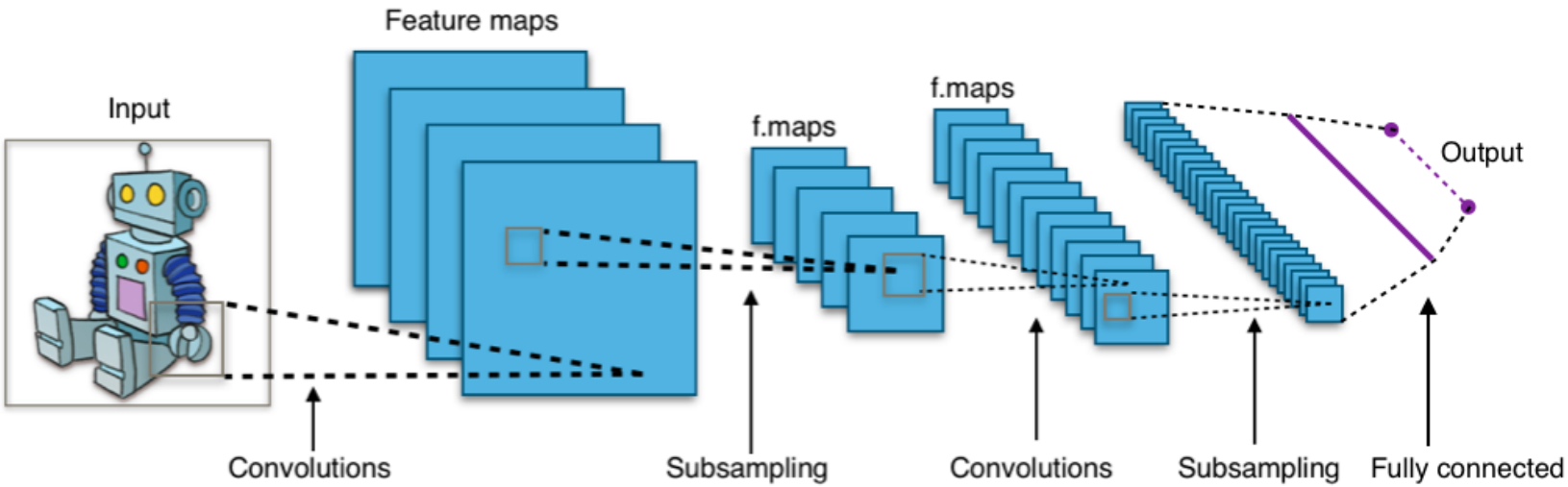
Successes

- Image/Object recognition
- Speech recognition
- Natural Language Processing
- Game theory (Go)
- ...

Deep learning on images : CNN



Deep learning on images : CNN



[Wikipedia]

Graph space

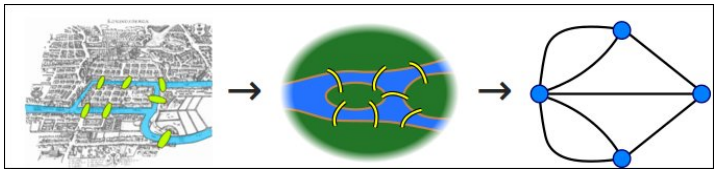
What is a graph ?

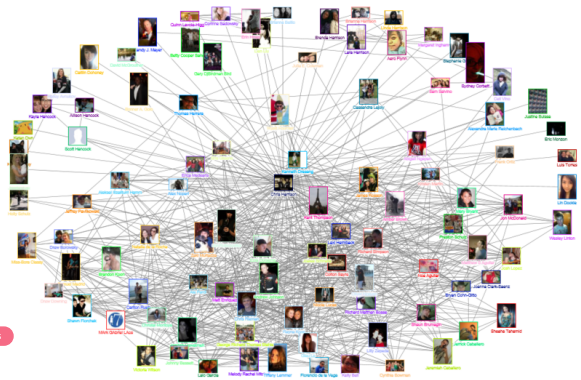
- $G = (V, E), E \in V \times V$
- Labels :
  - $l_v : V \to \mathbb{R}^{f_v}$
  - $l_e : E \to \mathbb{R}^{f_e}$
- degree : $d(v_i) = |\mathcal{N}(v_i)| \in \mathbb{N}^+$
- order : $|V|$, size : $|E|$.

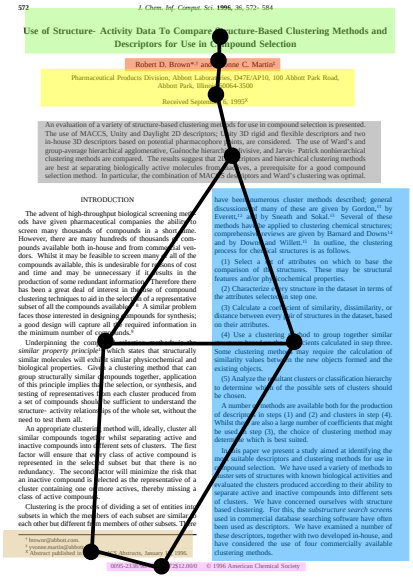
Graph representation (in ML)

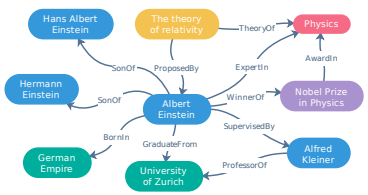
- Adjacency matrix $A \in \{0, 1\}^{n \times n}$, with $n = |V|$.
  - $A(i, j) = 1$ iff. $(v_i, v_j) \in E$
- Feature Matrix $X \in \mathbb{R}^{n \times f_v}$, $X(i, :) \Rightarrow$ features of node v_i.
- Laplacian : $D - A$, with $D(i, i) = d(v_i)$, else 0.

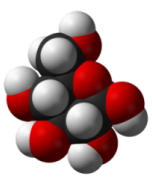
Graph space

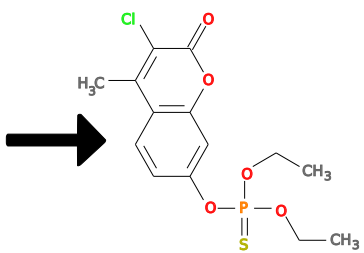




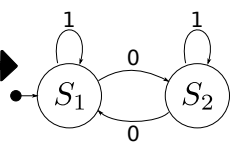








Current state \ Input	0	1
S_1	S_2	S_1
S_2	S_1	S_2

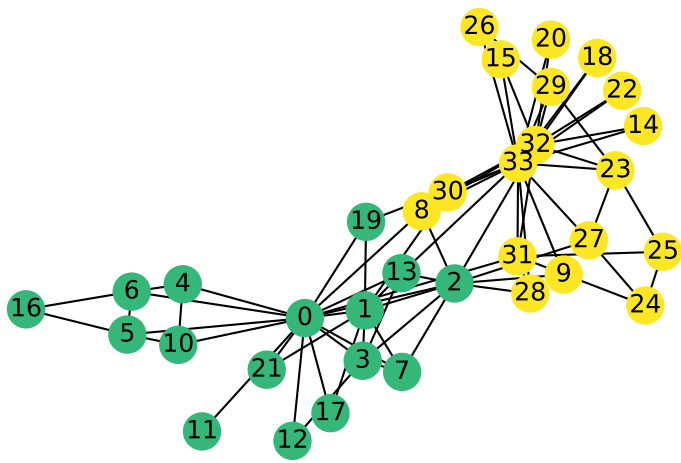


[from Linlin Jia]

Tasks on graphs I

Node level

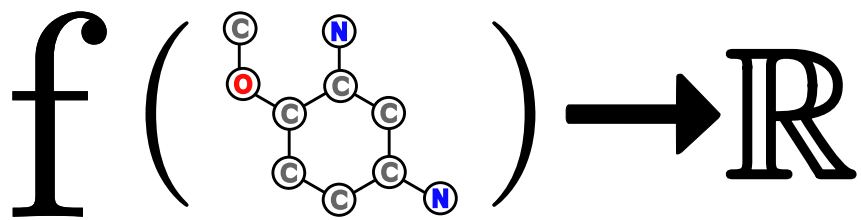
- Node label (i.e. property) prediction
  - Regression
  - Classification
- Transductive : predict unlabelled nodes on the same graph
- Inductive : predict node labels on a new graph
- Link (labels) prediction
- Clustering



Tasks on graphs II

Graph level

- Predict label (i.e. property) for a graph (e.g. toxicity of a molecule)
- Graph generation
- Metric learning



Why ML with graphs is particular ?

Graph problems

Graph space is not an Euclidean space

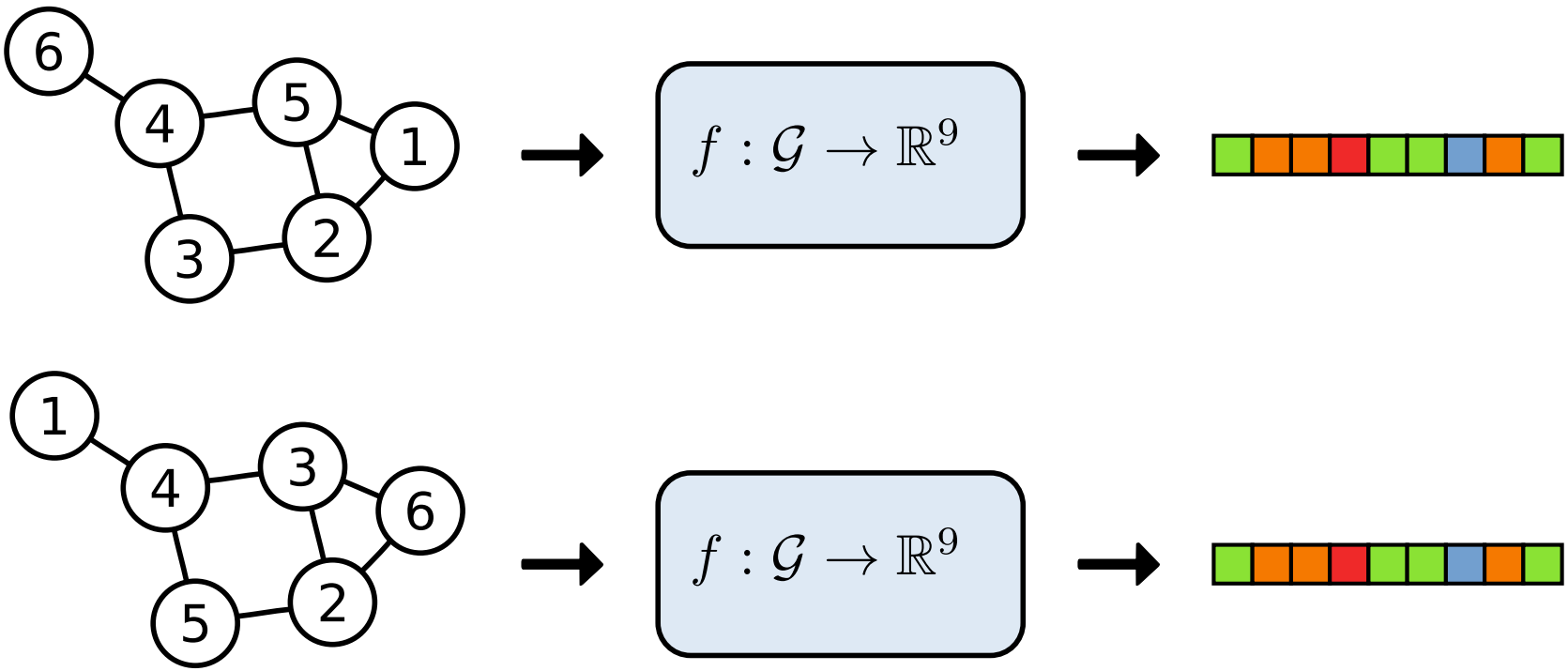
Variable number of nodes

- No fixed/limit number of nodes
- How to deal with a variable number of nodes/neighbours ?

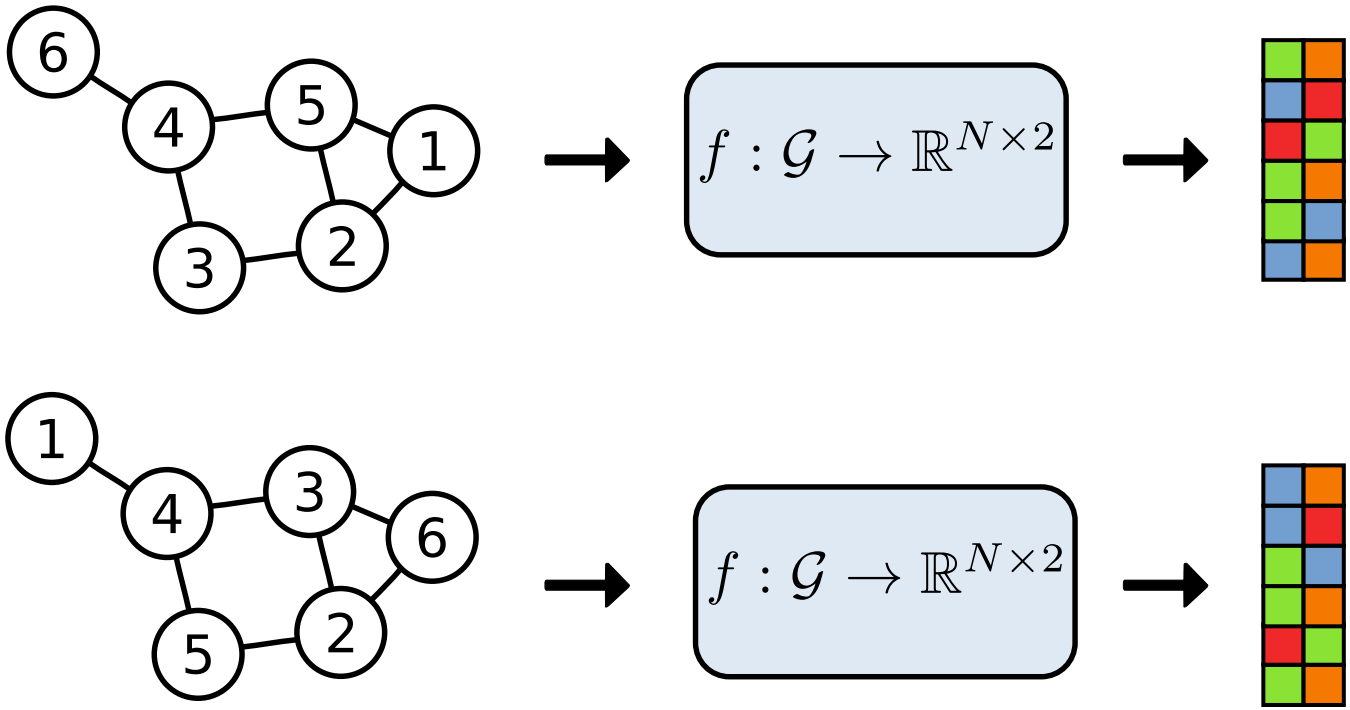
Permutation (equi/in)variance

- No predefined order of nodes
- $\Rightarrow$ No order on neighbours ($\neq$ images)

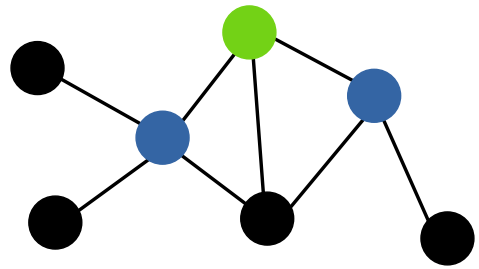
Permutation Invariance



Permutation Equivariance



Graphs versus Images



- Constant number of neighbours
- Fixed position of neighbours
- We want shift invariance

- Variable number of neighbours
- No predefined ordering of neighbours
- Permutation (equi/in)variance

Inspired by M. M. Bronstein

A first problem

Définition : Graph Isomorphism

$G_1 = (V_1, E_1) \simeq G_2 = (V_2, E_2)$ iff it exists a bijection $f : V_1 \to V_2$ s.t. $(u, v) \in E_1 \Leftrightarrow (f(u), f(v)) \in E_2$.

Remarks

- Notion of "Equality" between graphs.
- NP-Intermediate problem
- Labeled version : $l_v(u) = l_v(f(u)), \forall u \in V_1$

How to adapt CNN to Graphs ?

How to adapt convolution and pooling to variable dimension and permutation ?

A timeline

[Gori et al., 2005]

[Scarselli et al., 2009]

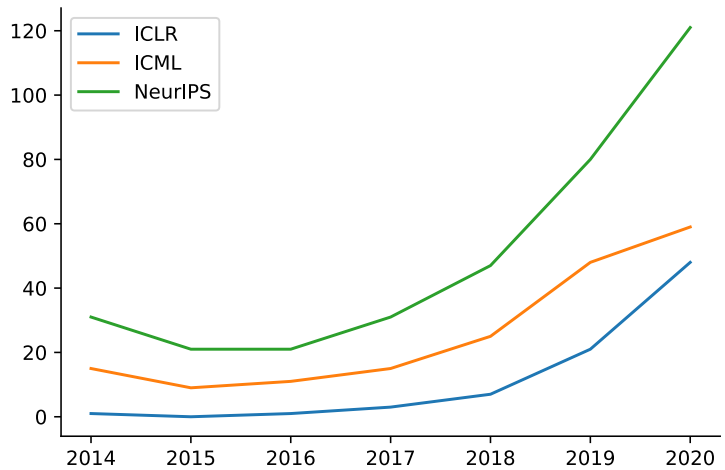
[Bruna et al. 2014] (Spectral)

[Kipf et al., 2016] (GCN)

[Wu et al., 2018] (Benchmark paper)

[Zhou et al., 2019] (First review)

- First attempt before Deep Era
- Explosion of papers since 2018
- See Sergey's analysis



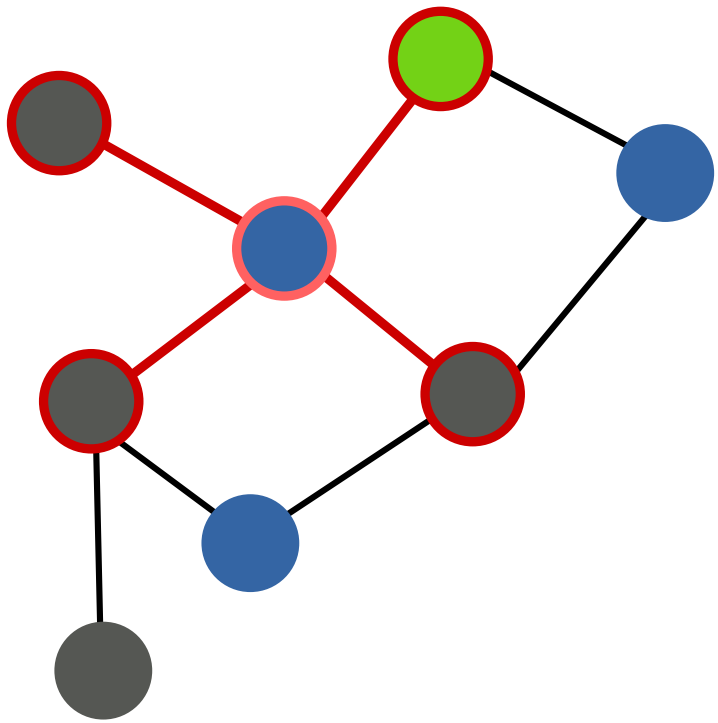
[data source: dplb]

Message passing framework

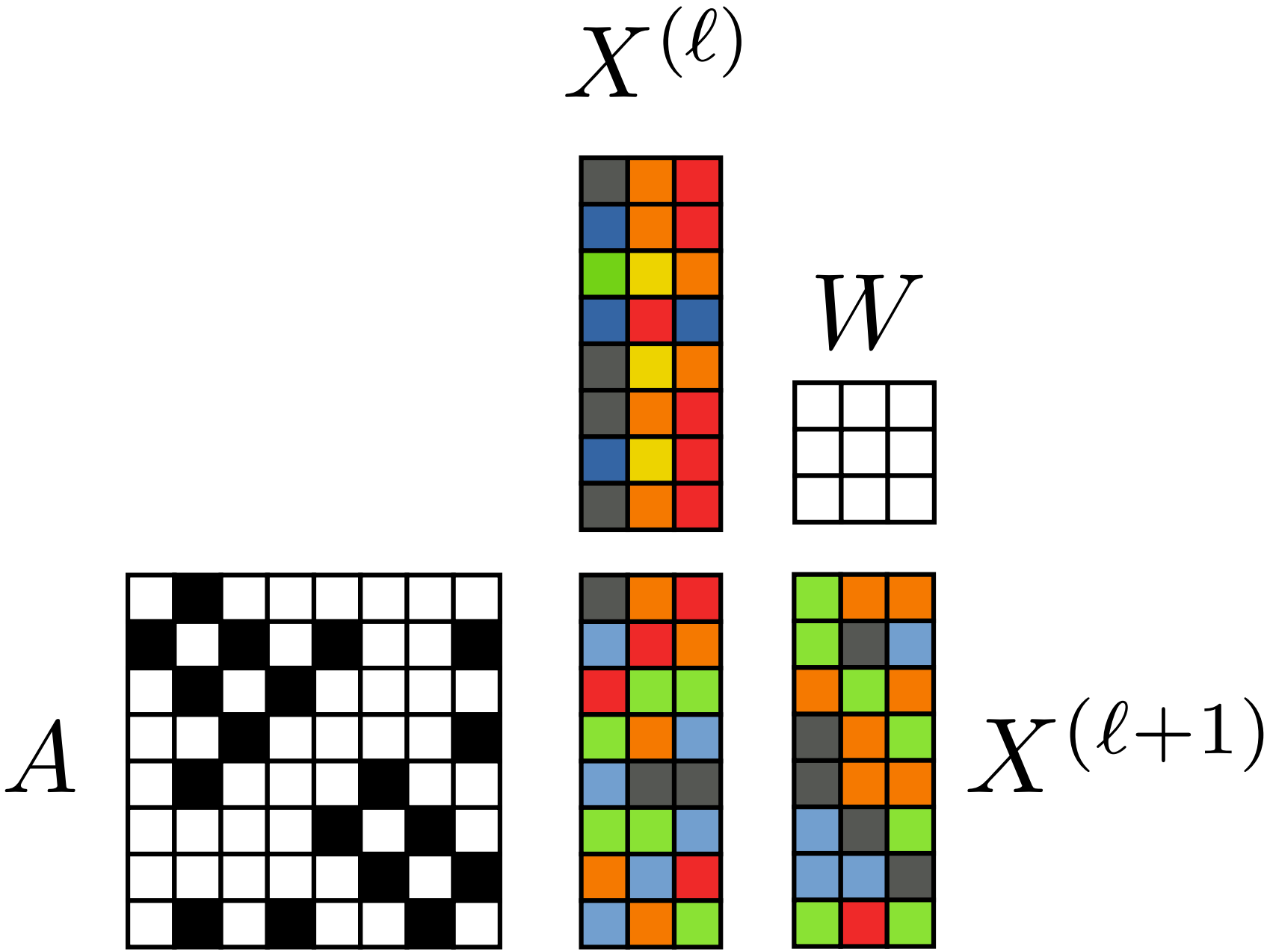
An introduction

Intuition

Update node representation according to neighbours

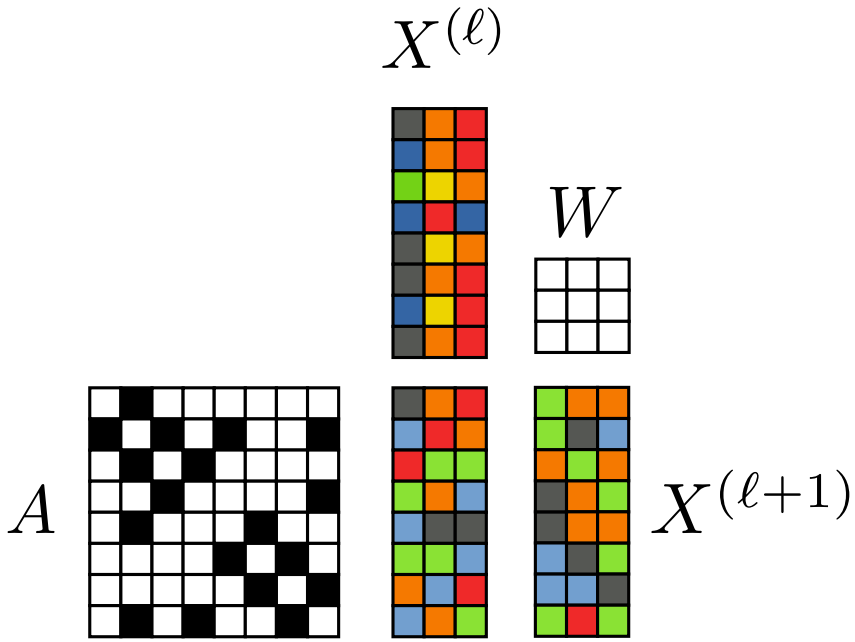


General principle



Convolutions on graphs

- A : Adjacency Matrix
- $AX^{(\ell)}(i,:)$: Sum all informations of $\mathcal{N}(v_i)$
- $AX^{(\ell)}(i,:)W = X^{(\ell+1)}$: "nearly" updated feature of v_i



Message passing framework

Message

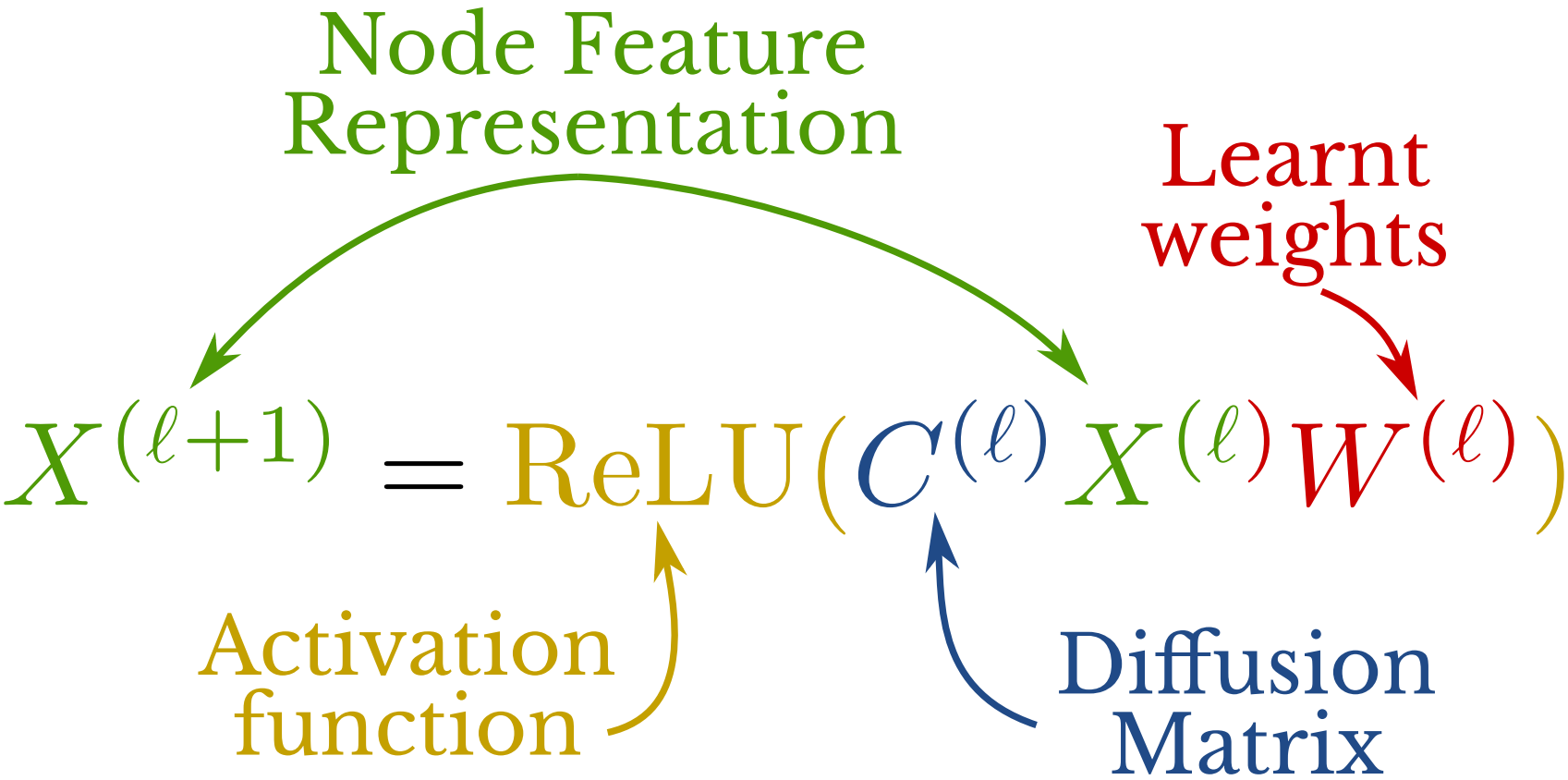
- Aggregate all information of neighboorhood
- $m_i^{(\ell+1)} = \sum_{v_j \in \mathcal{N}(v_i)} f(X_i^{(\ell)}, X_j^{(\ell)}, e_{i,j})$
- $AX(i,:) = \sum_{v_j \in \mathcal{N}(v_i)} X(j,:)$

Update

- Compute the new representation $X^{(\ell+1)}$
- $X_i^{(\ell+1)} = g(X_i^{(\ell)}, m_i^{(\ell+1)})$
- $X^{(\ell+1)}(i,:) = \sigma(AX^{(\ell)}(i,:)W + X^{(\ell)}(i,:))$
- $\sigma(\cdot)$: non linear activation function (ReLU).

[Gilmer et al., 2017]

MPNN Layer


Node Feature Representation
Learnt weights
$X^{(\ell+1)} = \mathrm{ReLU}(C^{(\ell)} X^{(\ell)} W^{(\ell)})$
Activation function
Diffusion Matrix

Graph Attention Networks

MPNN

- MPNN considers all neighboors in the same way
- Isotropic aggregation
- $\Rightarrow$ over-smoothing of information

Bringing attention to MPNN

- Weight each contribution of neighboor differently

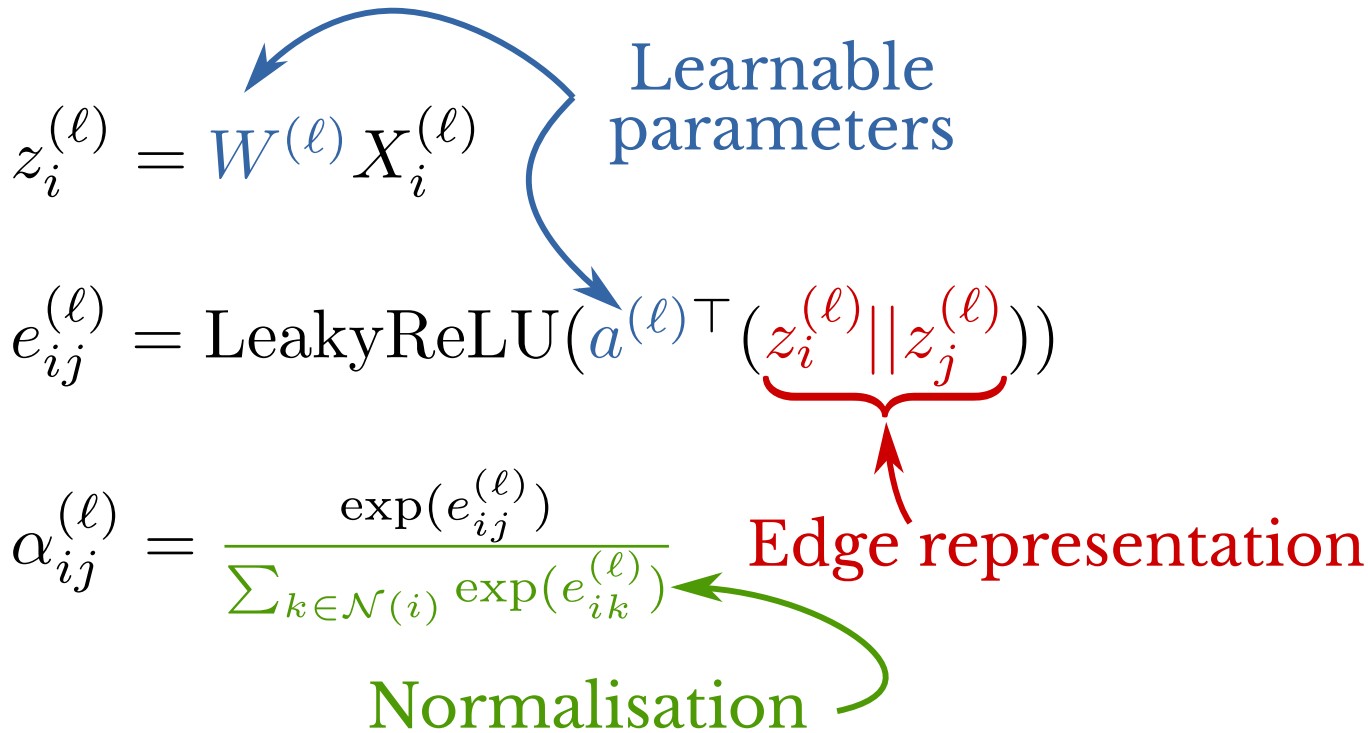
$$m_i^{\ell+1} = \sum_{v_j \in \mathcal{N}(v_i)} \alpha_{i,j} X(j,:)$$

- $\alpha_{i,j}$: attention coefficient

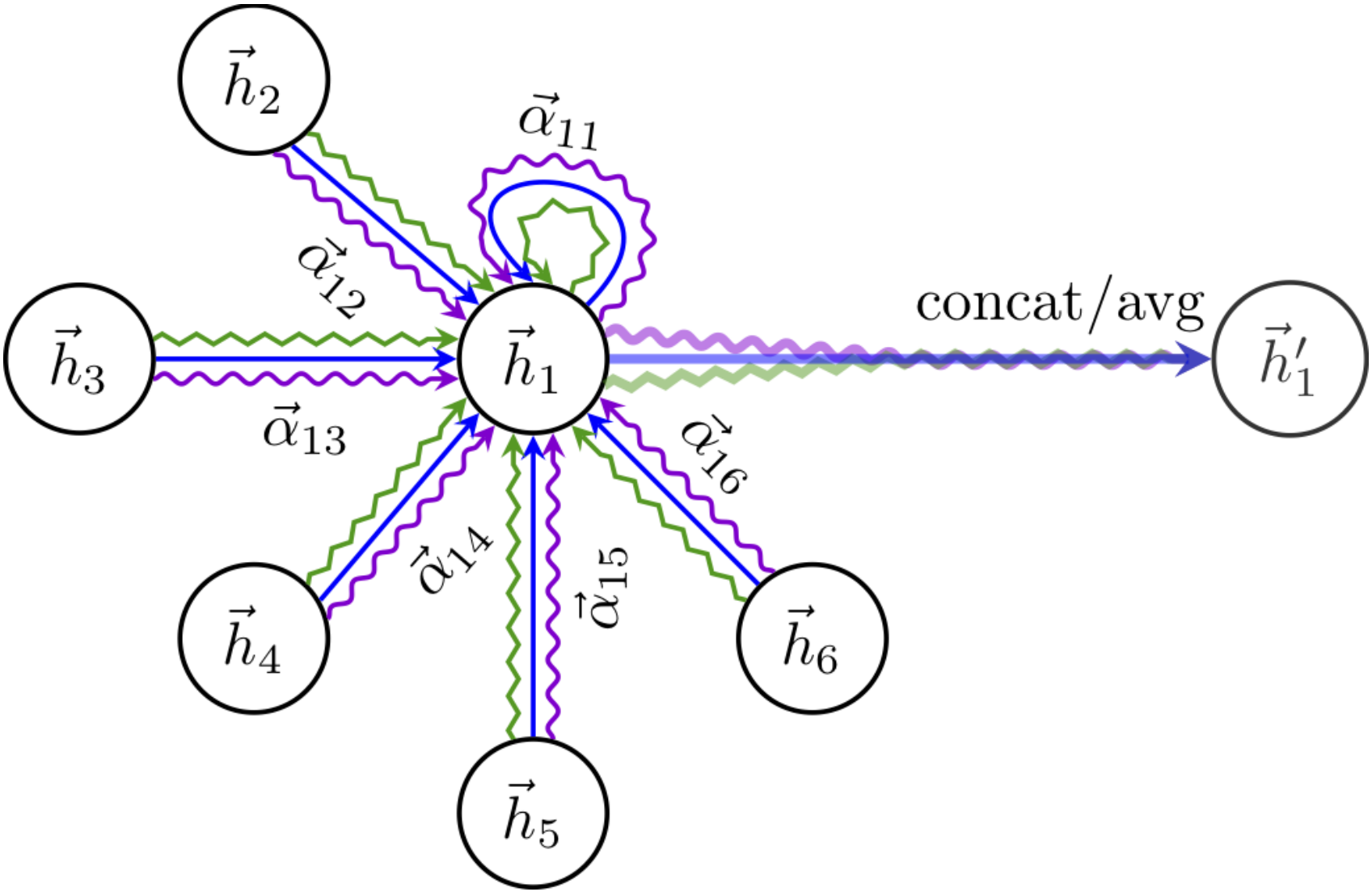
[Velicković et al., 2017]

Graph Attention Networks

Attention coefficient



Graph Attention Networks

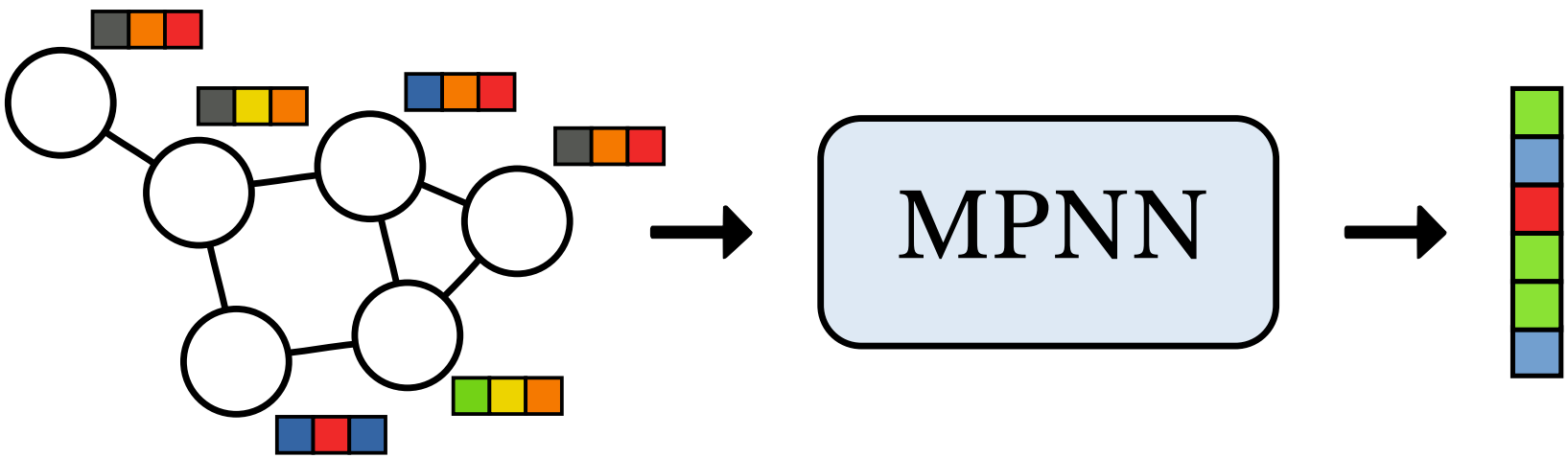


https://github.com/PetarV-/GAT

MPNN family

Node Representation Learning

- Build representation for nodes
- Useful for node level tasks
- Not complete for graph level



How to transform node representation to graph representation ?

Aggregation step

Readout function

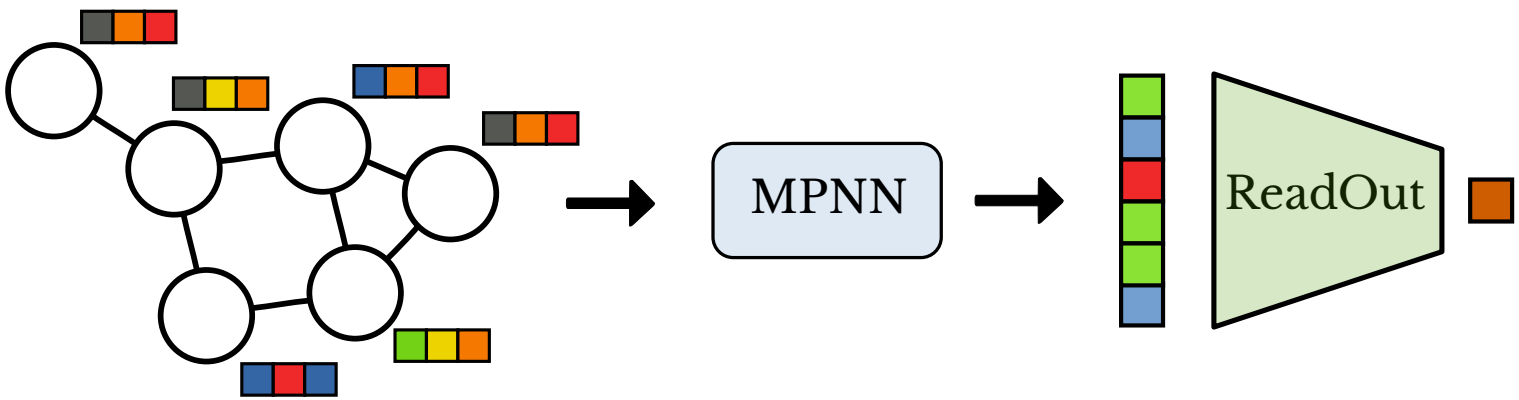
$$\hat{y} = R(\{X_i^{(\mathscr{L})} | v_i \in V\})$$

Aggregation step

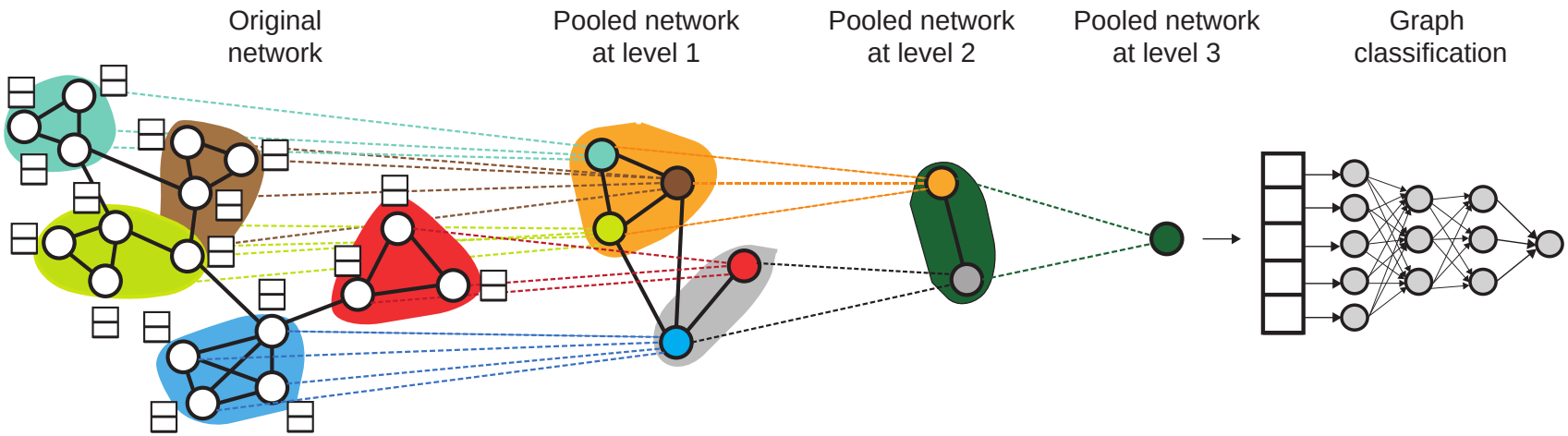
Readout function

$$\hat{y} = R(\{X_i^{(\mathscr{L})} | v_i \in V\})$$

- Differentiable
- Permutation invariant
- Simple statistics : mean, sum.
- Learnable : [Ying et al., 2018]



DiffPool



taken from [Ying et al., 2018]

Aggregation level

- Learnt Cluster Assignment Matrix $S^{(\ell)} \in \mathbb{R}^{n_l \times n_{l+1}}$
- Node representation : $X^{(\ell+1)} = S^{(\ell)\top} X^{(\ell)}$
- Adjacency matrix : $A^{(\ell+1)} = S^{(\ell)\top} A^{(\ell)} S^{(\ell)}$

[Ying et al., 2018]

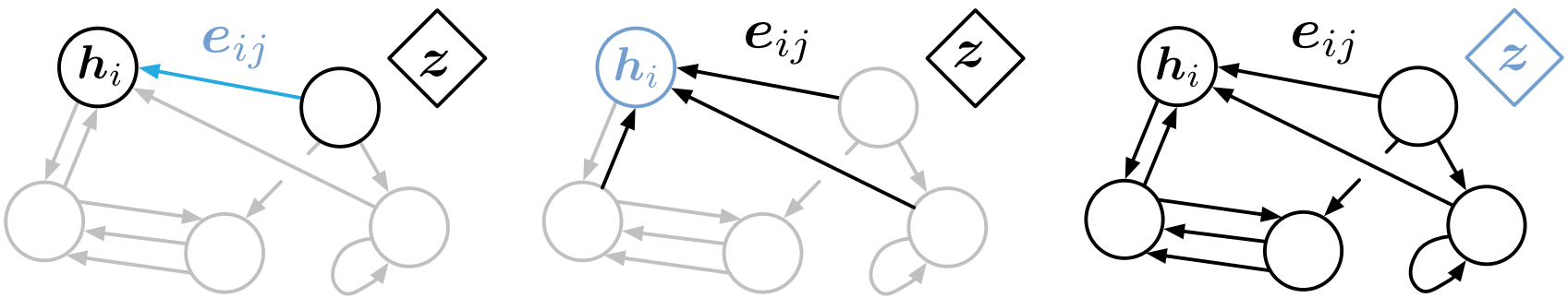
Graph Networks

Generalize MPNN

- Theoretical framework
- Three levels of representations:
  1. Edge $\mathbf{e}_{ij}$
  2. Node X_i or $\mathbf{h}_i$
  3. Graph $\boldsymbol{z}$

$\Rightarrow$ Three pairs of message/update functions

- Introduce edge representation learning



taken from [Battaglia et al., 2018]

[Battaglia et al., 2018]

Edge representation

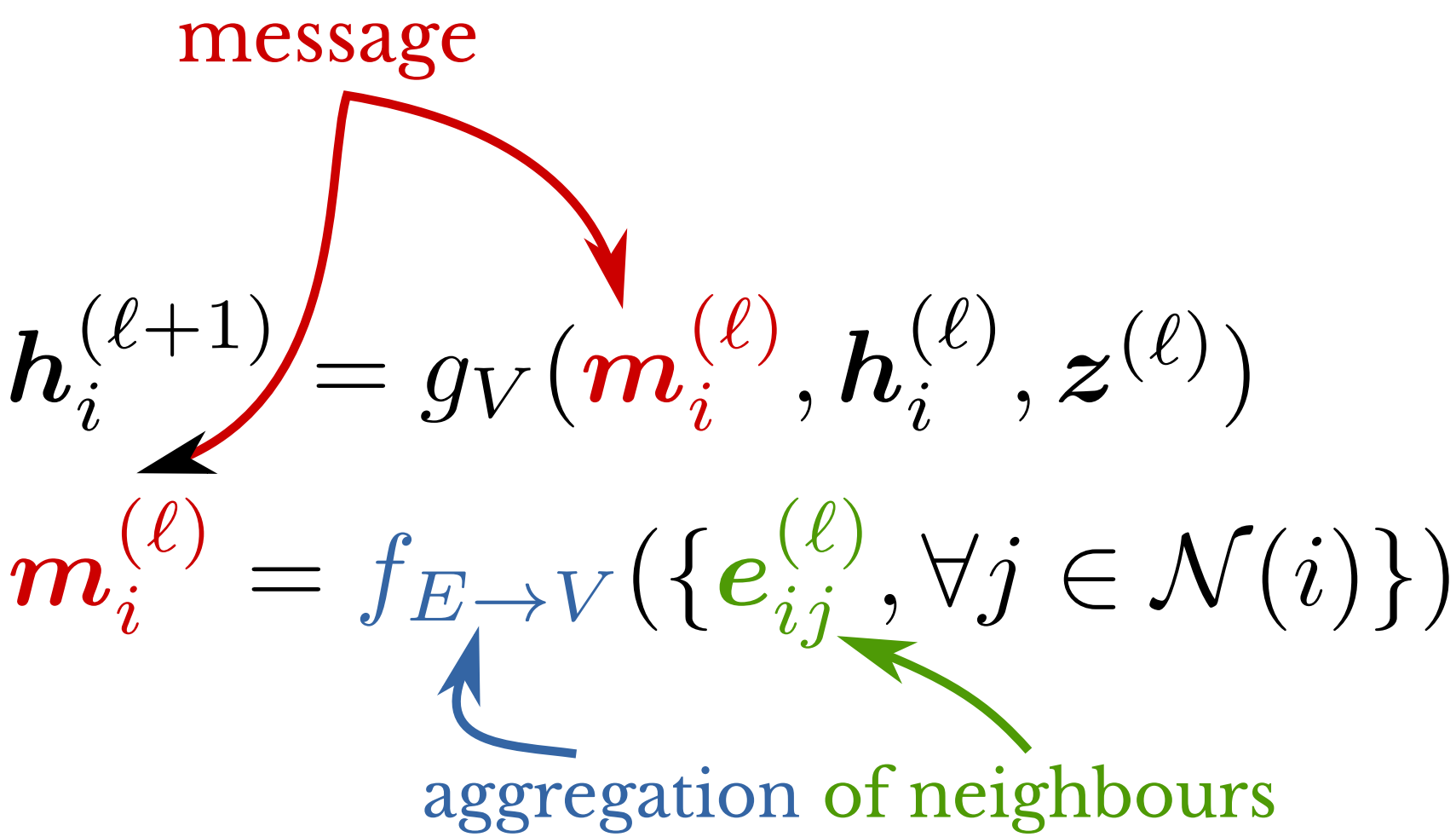
update function

$$\boldsymbol{e}_{ij}^{(\ell+1)} = g_E\left(\boldsymbol{e}_{ij}^{(\ell)}, \boldsymbol{h}_i^{(\ell)}, \boldsymbol{h}_j^{(\ell)}, \boldsymbol{z}^{(\ell)}\right)$$

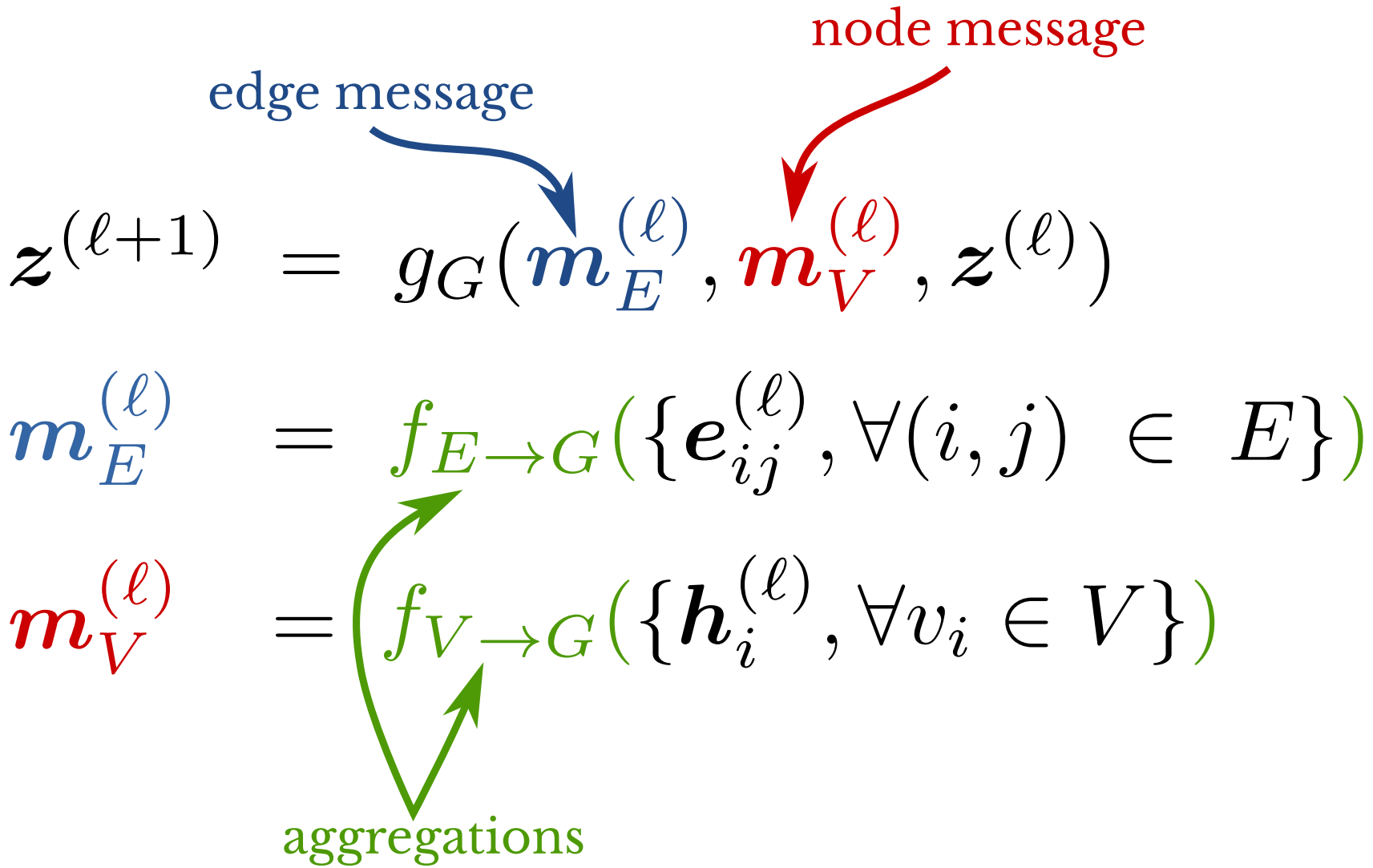
aggregation

→ See Guillaume's presentation for an implementation

Node representation



Graph representation



Theoretical aspects I

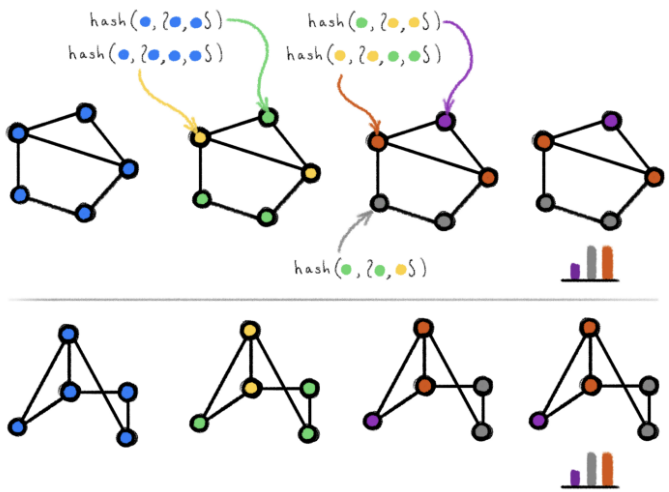
Interpretability: How to evaluate GNN ?

- Determine if a GNN can distinguish two graphs
- $GNN : \mathcal{G} \rightarrow \mathbb{R}^{d \times N}$
- $G_1 \simeq G_2 \Leftrightarrow GNN(G_1) = GNN(G_2)$?

Theoretical aspects II

Relationship with Weisfeler-Lehman test

- Iterative coloring process
- Polynomial approximation of isomorphism
- "classic" GCNs $\leq$ WL-Test
- Higher order of WL-test exist



M. Bronstein, medium blog

Limitations of GNNs

Low pass filtering

- Each iteration aggregates the neighboor's information
- Aggregation is (usually) isotropic
- Extend to not only low pass : spectral approaches

→ See Muhammet's talk

Over smoothing

- Adding layers increases smoothing
- At one point: all node's info is shared
- No real deep networks: generally 2 layers

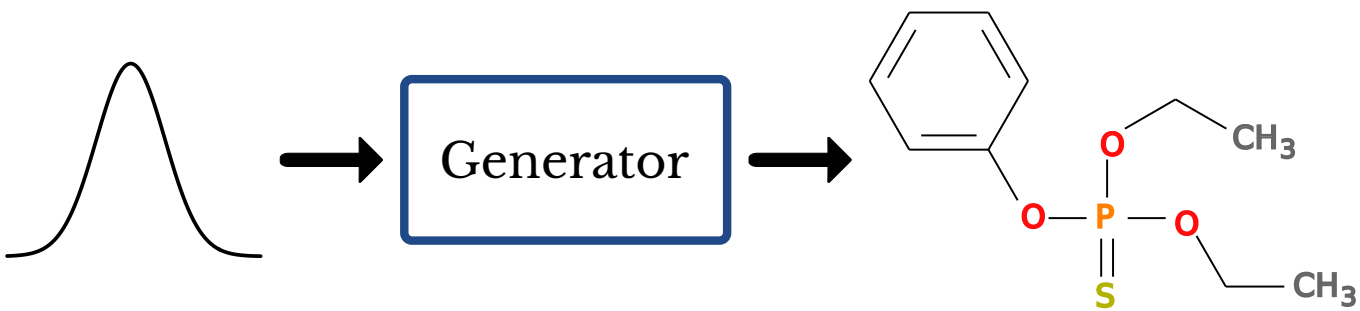
Graph generative models

Graph generation

- Create new graphs
- $f : \mathbb{R}^d \to \mathcal{G}$
- Explore latent euclidean space

Application

- Drug discovery
- Generate new molecules with particular properties



Graph Auto Encoders

Encoder

- GNN

$$Z = \mathrm{enc}_\Phi(A)$$

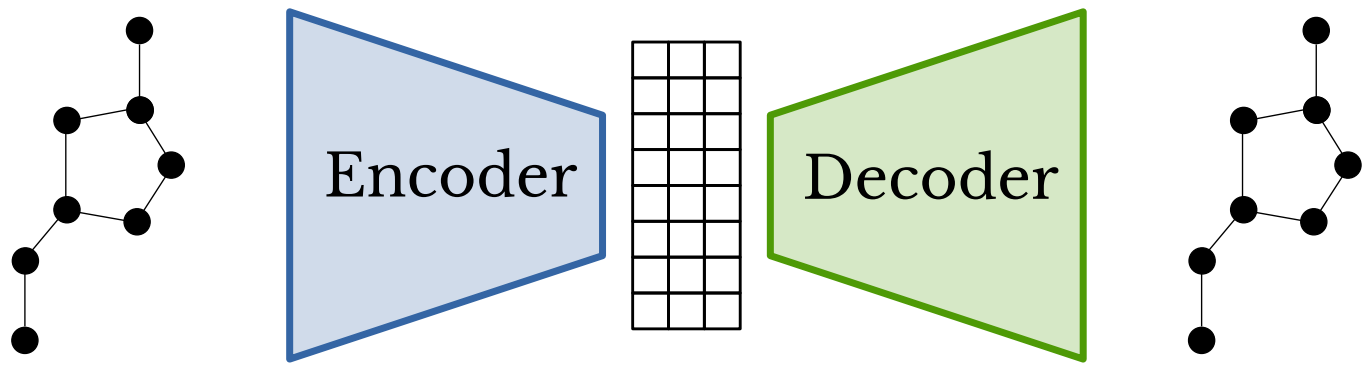
Decoder

- Reconstruct Adjacency matrix

$$\hat{A}(i,j) = \sigma(\boldsymbol{z}_i^\top \boldsymbol{z}_j)$$

Problem

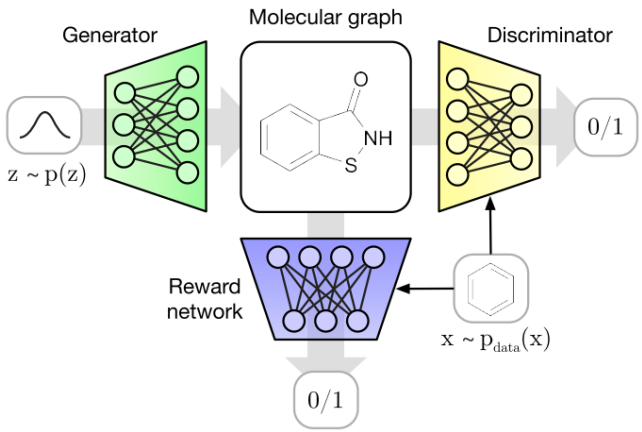
$$\min_\Phi \sum_{i=1}^{N} \|A(i,:) - \hat{A}(i,:)\|^2$$

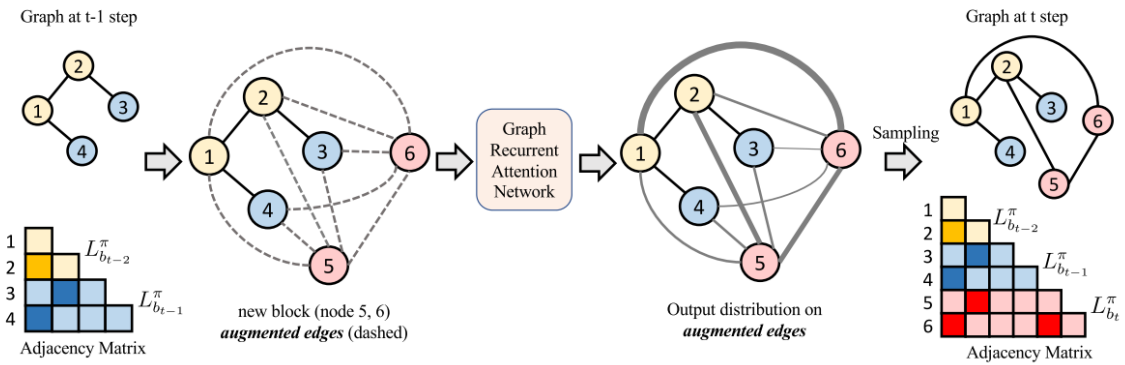


Other graph generation approaches

- Variationnal auto encoders (KL divergence)
- GANs
- Reinforcement learning
- Recursive processes

[De Cao and Kipf, 2018]





[Liao et al., 2019]

Application domains of GNN

Node level prediction

- Citation networks
- Node Clustering

Link prediction

- Collaboration graphs
- Knowledge graphs
- Temporal graphs
- Recommandation

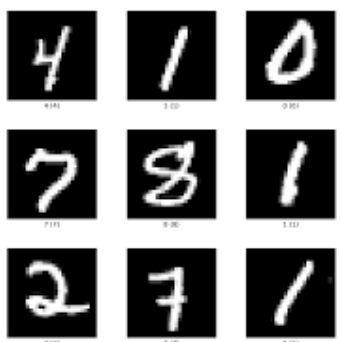
Graph prediction

- Molecular property prediction
  - Physiologic, toxicity, physical, quantum mechanics ...
- Protein-Protein interactions
- Programs, source code

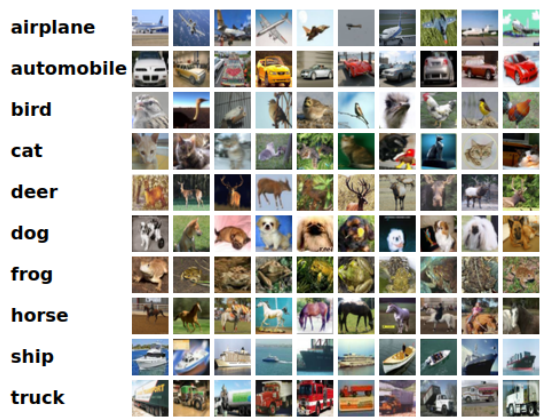
Emergence of datasets

The rise of Deep Learning in Computer Vision

- 3 causes :
  - Number of layers
  - Computationnal power
  - Datasets
- Standardized big datasets
- MNIST, ImageNet, CIFAR, ...



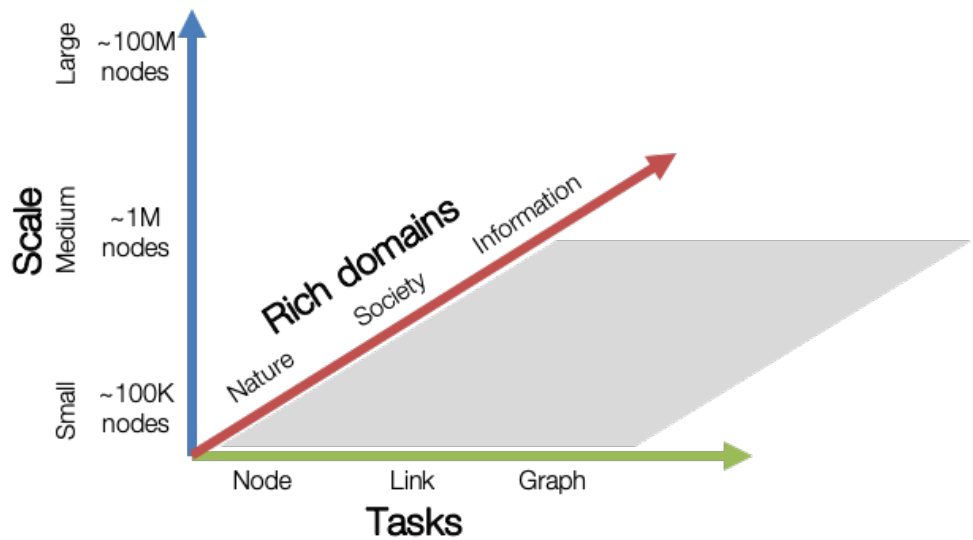
IMAGENET



Datasets for Graph Machine Learning I

OPEN GRAPH BENCHMARK

- ogb.stanford.edu
- 15 datasets
- Molecular graphs, citation networks,. . .



Datasets for Graph Machine Learning II

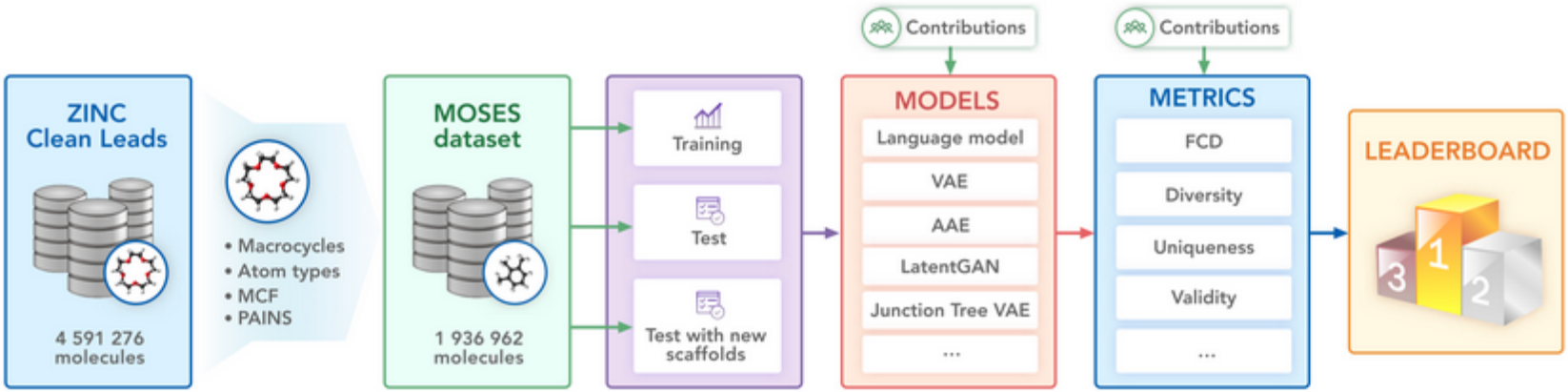
- MoleculeNet

Category	Dataset	Data Type	Task Type	# Tasks	# Compounds	Rec - Split[a]	Rec - Metric[b]
Quantum Mechanics	QM7	SMILES, 3D coordinates	Regression	1	7160	Stratified	MAE
	QM7b	3D coordinates	Regression	14	7210	Random	MAE
	QM8	SMILES, 3D coordinates	Regression	12	21786	Random	MAE
	QM9	SMILES, 3D coordinates	Regression	12	133885	Random	MAE
Physical Chemistry	ESOL	SMILES	Regression	1	1128	Random	RMSE
	FreeSolv	SMILES	Regression	1	642	Random	RMSE
	Lipophilicity	SMILES	Regression	1	4200	Random	RMSE
Biophysics	PCBA	SMILES	Classification	128	437929	Random	PRC-AUC
	MUV	SMILES	Classification	17	93087	Random	PRC-AUC
	HIV	SMILES	Classification	1	41127	Scaffold	ROC-AUC
	PDBbind	SMILES, 3D coordinates	Regression	1	11908	Time	RMSE
	BACE	SMILES	Classification	1	1513	Scaffold	ROC-AUC
Physiology	BBBP	SMILES	Classification	1	2039	Scaffold	ROC-AUC
	Tox21	SMILES	Classification	12	7831	Random	ROC-AUC
	ToxCast	SMILES	Classification	617	8575	Random	ROC-AUC
	SIDER	SMILES	Classification	27	1427	Random	ROC-AUC
	ClinTox	SMILES	Classification	2	1478	Random	ROC-AUC

moleculenet.ai

Datasets for Graph Machine Learning III

MOSES



- https://github.com/molecularsets/moses
- Dataset for generative models

Conclusion and Outlooks

GNNs

- Bring representation learning to graphs
- Dynamic and growning field
- Still in infancy, but supported by industry
- Still some limitations
  - Still computes a Euclidean embedding
  - Not deep yet

Outlooks

- Dynamic graphs
- Interpretability
- Real "Deep" Learning ?
- Improve readout ? is it useful ?

Questions and Discussion

- Possible use of GNN ?
- How to find our place ?
- Collaborations ?


Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gulcehre, C., Song, F., Ballard, A., Gilmer, J., Dahl, G., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., and Pascanu, R. (2018).
Relational inductive biases, deep learning, and graph networks.
arXiv, pages 1–40.


De Cao, N. and Kipf, T. (2018).
MolGAN: An implicit generative model for small molecular graphs.


Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. (2017).
Neural message passing for quantum chemistry.
arXiv.


Liao, R., Li, Y., Song, Y., Wang, S., Nash, C., Hamilton, W. L., Duvenaud, D., Urtasun, R., and Zemel, R. (2019).

Efficient Graph Generation with Graph Recurrent Attention Networks.
arXiv, (NeurIPS):1–13.


Velicković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2017).
Graph attention networks.
arXiv, pages 1–12.


Ying, R., Morris, C., Hamilton, W. L., You, J., Ren, X., and Leskovec, J. (2018).
Hierarchical graph representation learning with differentiable pooling.
Advances in Neural Information Processing Systems, 2018-December:4800–4810.