Message Passing In Graph Neural Networks

Johannes Lutzeyer

Data Science and Mining Team, Computer Science Department,
École Polytechnique, Institut Polytechnique de Paris

Inaugural Paris Learning-on-Graphs Meetup

June 14, 2023

Today I present work that was done in collaboration with

Gaspard Michel
PhD Student Deezer

Ariel Ramos Vela
Master Student IPP

Michalis Chatzianastasis
PhD Student LIX

Dr. Guillaume Salha Galvan
Research Scientist Deezer

Dr. Changmin Wu
Researcher Huawei

Dr. George Dasoulas
Postdoc Harvard

Dr. Mohamed Seddik
Senior Researcher TII

Dr. Giannis Nikolentzos
Postdoc LIX

Dr. Anastasios Giovanidis
Data Science Researcher
Ericsson

Dr. Romain Hennequin
Lead Research Scientist
Deezer

Prof. Michalis Vazirgiannis
Distinguished Professor LIX

Graph Neural Networks

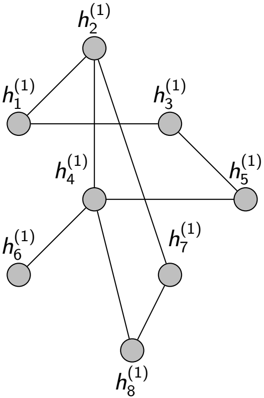
Graph Neural Networks (GNNs) are neural networks that take graph-structured data as input.

In this talk we will only see a specific type of GNN, the Message Passing Neural Networks.

$$
\begin{aligned}
m_v^{(k)} &= M^{(k)}\left(\left\{h_w^{(k-1)} : w \in \mathcal{N}(v)\right\}\right), \\
h_v^{(k)} &= U^{(k)}\left(h_v^{(k-1)}, m_v^{(k)}\right).
\end{aligned}
$$

E.g., the Graph Convolutional Network (GCN, Kipf and Welling, 2017)

$$
H^{(1)} = \mathrm{ReLU}\left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} X W^{(1)}\right).
$$



Iteratively performing the message-passing and update computations allows us to build 'deep' learning models, e.g., a 3-layer GCN

$$
\hat{y} = \sigma\left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \; \mathrm{ReLU}\left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \; \mathrm{ReLU}\left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} X W^{(1)}\right) W^{(2)}\right) W^{(3)}\right).
$$

Motivation & Outline

The message passing step is a defining component of GNNs.

> *"any function of interest we want to compute over graphs can, in all likelihood, be expressed using pairwise message passing – just over a potentially modified graph [...]"*

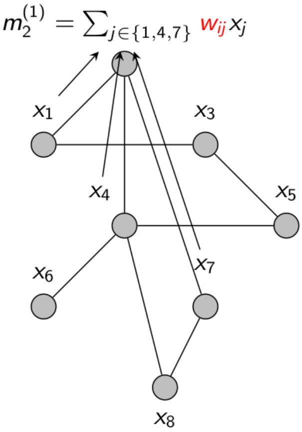
Petar Veličković (2022)

Topic of Today's Talk

Categorise several existing GNNs by their message passing step and place our recent work into that context.

Different Approaches to Message Passing

- Fixed Graph



Different Approaches to Message Passing

- Fixed Graph
  - GCN (Kipf and Welling, 2017)

    Message-Passing Operation: $D^{-\frac{1}{2}} A D^{-\frac{1}{2}} X.$

  - GIN (Xu et al., 2019)

    Message-Passing Operation: $(A + \epsilon I)\, X.$

  - PGSO-GNN (Dasoulas et al., 2021, ICLR)

    Message-Passing Operation: $(m_1 D_a^{e_1} + m_2 D_a^{e_2} A_a D_a^{e_3} + m_3 I_n)\, X,$

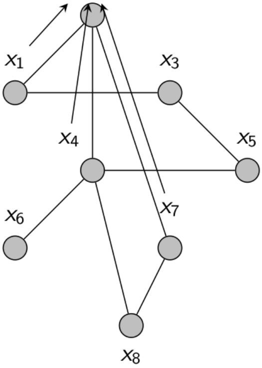
    where $A_a = A + aI_n$, $D_a = \mathrm{Diag}(A_a \mathbf{1}_n)$ and $(m_1, m_2, m_3, e_1, e_2, e_3, a)$ are scalar, *trainable* parameters.

$\mathcal{S} = (m_1,\ m_2,\ m_3,\ e_1,\ e_2,\ e_3,\ a)$	**Operator**	**Description**
$(0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0)$	A	Adjacency matrix and Summation Aggregation Operator of GNNs
$(1,\ -1,\ 0,\ 1,\ 0,\ 0,\ 0)$	$D - A$	Unnormalised Laplacian matrix L
$(1,\ 1,\ 0,\ 1,\ 0,\ 0,\ 0)$	$D + A$	Signless Laplacian matrix Q (Cvetkovic et al., 1997)
$(0,\ -1,\ 1,\ 0,\ -1,\ 0,\ 0)$	$I_n - D^{-1}A$	Random-walk Normalised Laplacian L_{rw}
$(0,\ -1,\ 1,\ 0,\ -\frac{1}{2},\ -\frac{1}{2},\ 0)$	$I_n - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$	Symmetric Normalised Laplacian L_{sym}
$(0,\ 1,\ 0,\ 0,\ -\frac{1}{2},\ -\frac{1}{2},\ 1)$	$D_1^{-\frac{1}{2}} A_1 D_1^{-\frac{1}{2}}$	Normalised Adjacency matrix of GCNs (Kipf and Welling, 2017)
$(0,\ 1,\ 0,\ 0,\ -1,\ 0,\ 0)$	$D^{-1}A$	Mean Aggregation Operator of GNNs (Xu et al., 2019)

Different Approaches to Message Passing

- Fixed Graph
- Feature-Dependent Reweighting of Edges

$$m_2^{(1)} = \sum_{j \in \{1,4,7\}} w_{ij}(X) x_j$$



Different Approaches to Message Passing

- Fixed Graph
- Feature-Dependent Reweighting of Edges
  - GAT (Veličković et al., 2018)
    Message-Passing Operation: $A_{att}X$,

$$\text{where } (A_{att})_{ij} = \begin{cases} 0, & \text{for } A_{ij} = 0; \\ \dfrac{\exp\left(\mathrm{LeakyReLU}\left(w_2^T \begin{bmatrix} W_1 h_i \\ W_1 h_j \end{bmatrix}\right)\right)}{\sum_{j \in \mathcal{N}(v_i)} \exp\left(\mathrm{LeakyReLU}\left(w_2^T \begin{bmatrix} W_1 h_i \\ W_1 h_j \end{bmatrix}\right)\right)}, & \text{for } A_{ij} \neq 0. \end{cases}$$

  - GATv2 (Brody et al., 2022)
    Message-Passing Operation: $A_{attv2}X$,

$$\text{where } (A_{attv2})_{ij} = \begin{cases} 0, & \text{for } A_{ij} = 0; \\ \dfrac{\exp\left(w_2^T \mathrm{LeakyReLU}\left(W_1 \begin{bmatrix} h_i \\ h_j \end{bmatrix}\right)\right)}{\sum_{j \in \mathcal{N}(v_i)} \exp\left(w_2^T \mathrm{LeakyReLU}\left(W_1 \begin{bmatrix} h_i \\ h_j \end{bmatrix}\right)\right)}, & \text{for } A_{ij} \neq 0. \end{cases}$$

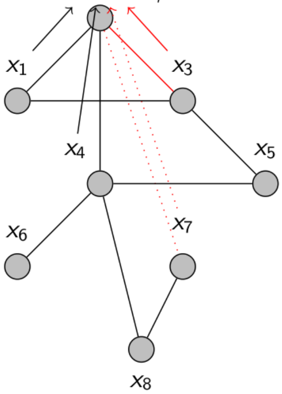
  - GCN-k (Seddik et al., 2022, AISTATS)
    Message-Passing Operation: $\left(\epsilon D^{-\frac{1}{2}} A D^{-\frac{1}{2}} + (1-\epsilon) D_K^{-\frac{1}{2}} K D_K^{-\frac{1}{2}}\right) X$,

$$\text{where } (K)_{ij} = \begin{cases} 0, & \text{for } A_{ij} = 0; \\ x_i^T x_j, & \text{for } A_{ij} \neq 0. \end{cases}$$

Different Approaches to Message Passing

- Fixed Graph
- Feature-Dependent Reweighting of Edges
- Adding and or Removing Edges

$$m_2^{(1)} = \sum_{j \in \{1,3,4,\cancel{7}\}} w_{ij} x_j$$

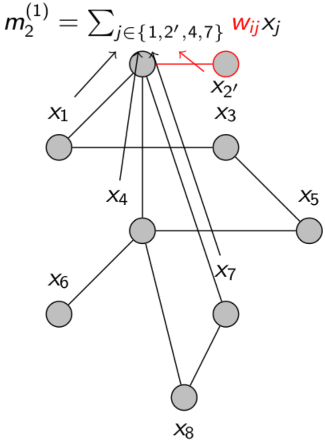


Different Approaches to Message Passing

- Fixed Graph
- Feature-Dependent Reweighting of Edges
- Adding and or Removing Edges
  - SDRF (Topping, Di Giovanni et al., 2022)
    Rewiring according to curvature metrics on graphs
  - PPRGo (Bojchevski et al., 2020)
    Rewiring according to thresholded Personalised PageRank Scores
  - CorePPR Ramos Vela et al. (2022, NeurIPS Workshop)
    Rewiring according to thresholded Personalised PageRank and CoreRank Scores
  - Modularity-Aware (V)GAE (Salha-Galvan et al., 2022, Neural Networks)
    Add edges based on Louvain Clustering

Different Approaches to Message Passing

- Fixed Graph
- Feature-Dependent Reweighting of Edges
- Adding and or Removing Edges
- Explicitly Representing Substructures of Graphs



Different Approaches to Message Passing

- Fixed Graph
- Feature-Dependent Reweighting of Edges
- Adding and or Removing Edges
- Explicitly Representing Substructures of Graphs
    - Subgraph GNNs (Frasca et al., 2022)
    - PathNNs (Michel et al., 2023, ICML)
        1) At layer k, a PathNN uses an LSTM to learn path representations of all paths emanating from a node of length k.
        2) Path Representations are subsequently aggregated.

Different Approaches to Message Passing

- Fixed Graph
- Feature-Dependent Reweighting of Edges
- Adding and or Removing Edges
- Explicitly Representing Substructures of Graphs
- Some GNNs are difficult categorise
  - GOAT (Chatzianastasis et al., 2023, AAAI)
    1) A self-attention mechanism is used to obtain a ranking of nodes in neighbourhoods.
    2) An LSTM processed the ordered neighbourhoods to produce updated node representation.

Conclusions

- If the original graph is sufficient for the performed learning task, then we should simply aggregate over the fixed graph.
- If the node features contain complementary information to the graph on the relevance of neighbours to a given node, then we should use a feature-dependent reweighting scheme on the edges.
- If the graph structure is insufficient for the learning task, e.g., necessary information presents itself as a long-range effect on the original graph, then we should pick a criterion to limit the search space of n^2 node pairs and rewire accordingly.
- If we are aware of certain substructures of particular relevance to our learning task or require a highly expressive model, then we should explicitly represent these substructures in the message passing scheme.

We are currently looking for Postdocs & Research Engineers!

More detailed job postings are on Twitter and our group's website.

Thank you for your attention!

@JLutzeyer

References

A. Bojchevski, J. Gasteiger, B. Perozzi, A. Kapoor, M. Blais, B. Rózemberczki, M. Lukasik & S. Günnemann, "Scaling graph neural networks with approximate pagerank," *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 2464-2473, 2020.

S. Brody, U. Alon & E. Yahav, "How Attentive Are Graph Attention Networks?," In: *International Conference on Learning Representations (ICLR)*, 2022.

M. Chatzianastasis, J. F. Lutzeyer, G. Dasoulas & M. Vazirgiannis, "Graph Ordering Attention Networks," *Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI),* 2023.

D. Cvetkovic, R. Rowlinson & S. Simic, *Eigenspaces of graphs*, Cambridge, UK: Cambridge University Press, 1997.

G. Dasoulas, J. F. Lutzeyer & M. Vazirgiannis, "Learning Parametrised Graph Shift Operators," In: *International Conference on Learning Representations (ICLR)*, 2021.

F. Frasca, B. Bevilacqua, M. Bronstein & H. Maron, "Understanding and Extending Subgraph GNNs by Rethinking Their Symmetries," *Proceedings of the 36st International Conference on Neural Information Processing Systems (NeurIPS),* 2022.

Thomas N. Kipf & M. Welling, "Semi-supervised classification with graph convolutional networks," *International Conference on Learning Representations (ICLR)*, 2017.

G. Michel, G. Nikolentzos, J. Lutzeyer & M. Vazirgiannis, "Path Neural Networks: Expressive and Accurate Graph Neural Networks," *Proceedings of the 40th International Conference on Machine Learning (ICML)*, 2023.

A. R. Ramos Vela, J. F. Lutzeyer, A. Giovanidis & M. Vazirgiannis, "Improving Graph Neural Networks at Scale: Combining Approximate PageRank and CoreRank," *NeurIPS New Frontiers in Graph Learning Workshop*, 2022.

G. Salha-Galvan, J. F. Lutzeyer, G. Dasoulas, R. Hennequin & M. Vazirgiannis, "Modularity-Aware Graph Autoencoders for Joint Community Detection and Link Prediction," *Neural Networks*, vol. 153, pp. 474–495, 2022.

M. E. A. Seddik, C. Wu, J. F. Lutzeyer & M. Vazirgiannis, "Node Feature Kernels Increase Graph Convolutional Network Robustness," *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

J. Topping, F. Di Giovanni, B. P. Chamberlain, X. Dong & M. M. Bronstein, "Understanding Over-Squashing and Bottlenecks on Graphs via Curvature," *International Conference on Learning Representations (ICLR)*, 2022.

P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò & Y. Bengio, "Graph Attention Networks," *6th International Conference on Learning Representations (ICLR),* 2018.

P. Veličković, "Message Passing All the Way Up," *ICLR Workshop on Geometrical and Topological Representation Learning,* 2022.

K. Xu, W. Hu, J. Leskovec & S. Jegelka. "How powerful are graph neural networks?", *International Conference on Learning Representations (ICLR)*, 2019.