Graph and geometric deep learning

Presenters: Simone Scardapane, Indro Spinelli

SAPIENZA UNIVERSITÀ DI ROMA

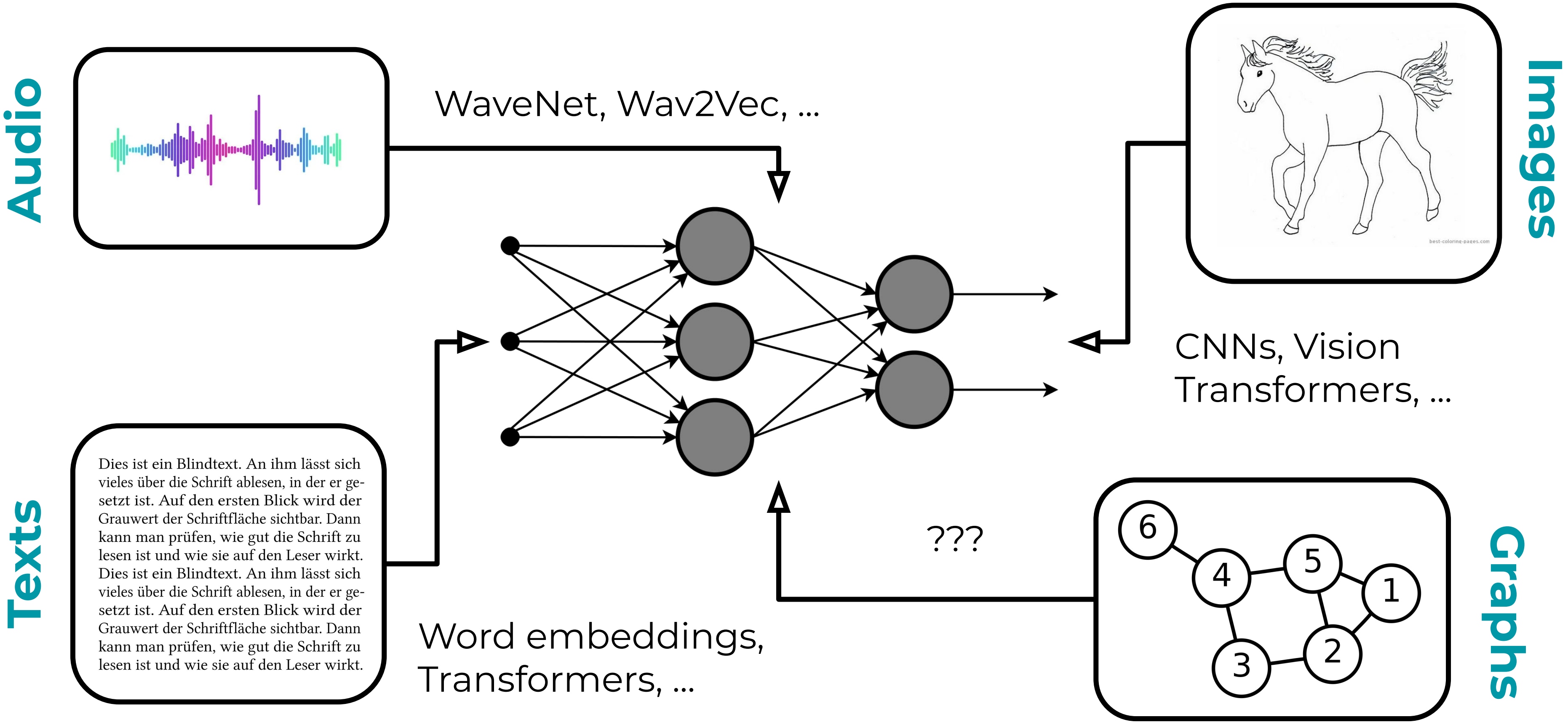
intelligent signal processing and multimedia lab

Enel Third Global Data Meet Up, April 22nd 2021

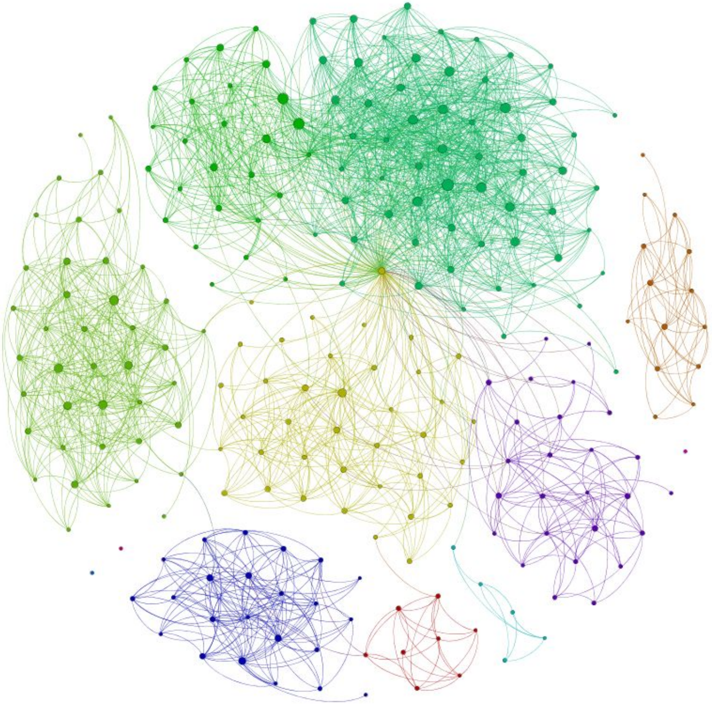
Introduction

On the importance of **graphs** in **deep learning**

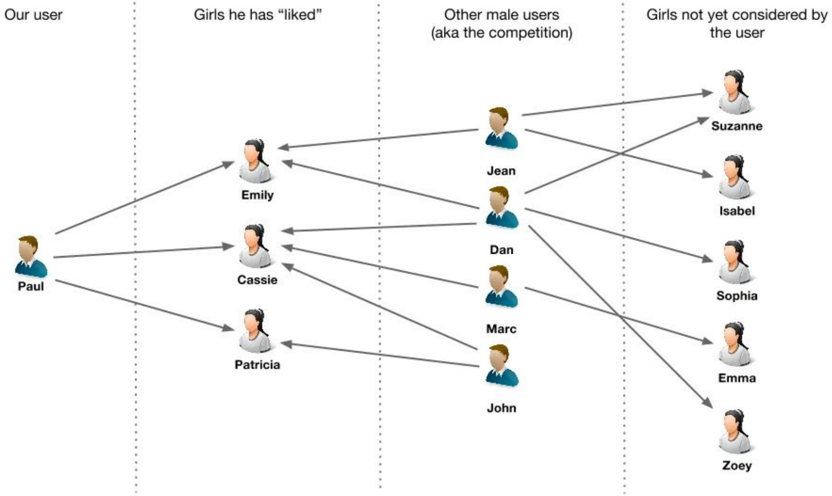
Data ingestion in deep learning



Graphs are everywhere (1/4)

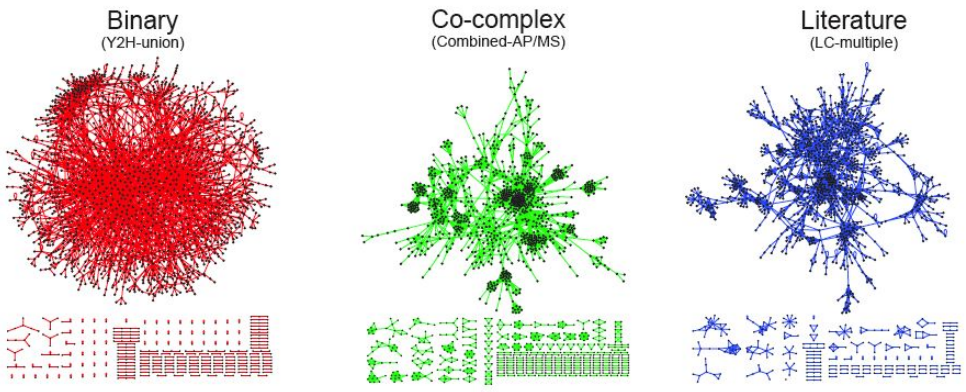


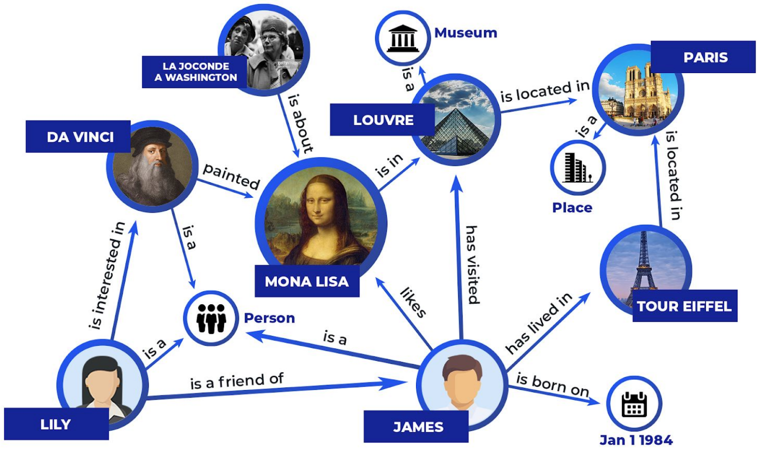
http://allthingsgraphed.com/2014/08/28/facebook-friends-network/



https://linkurio.us/blog/using-neo4j-to-build-a-recommendation-engine-based-on-collaborative-filtering/

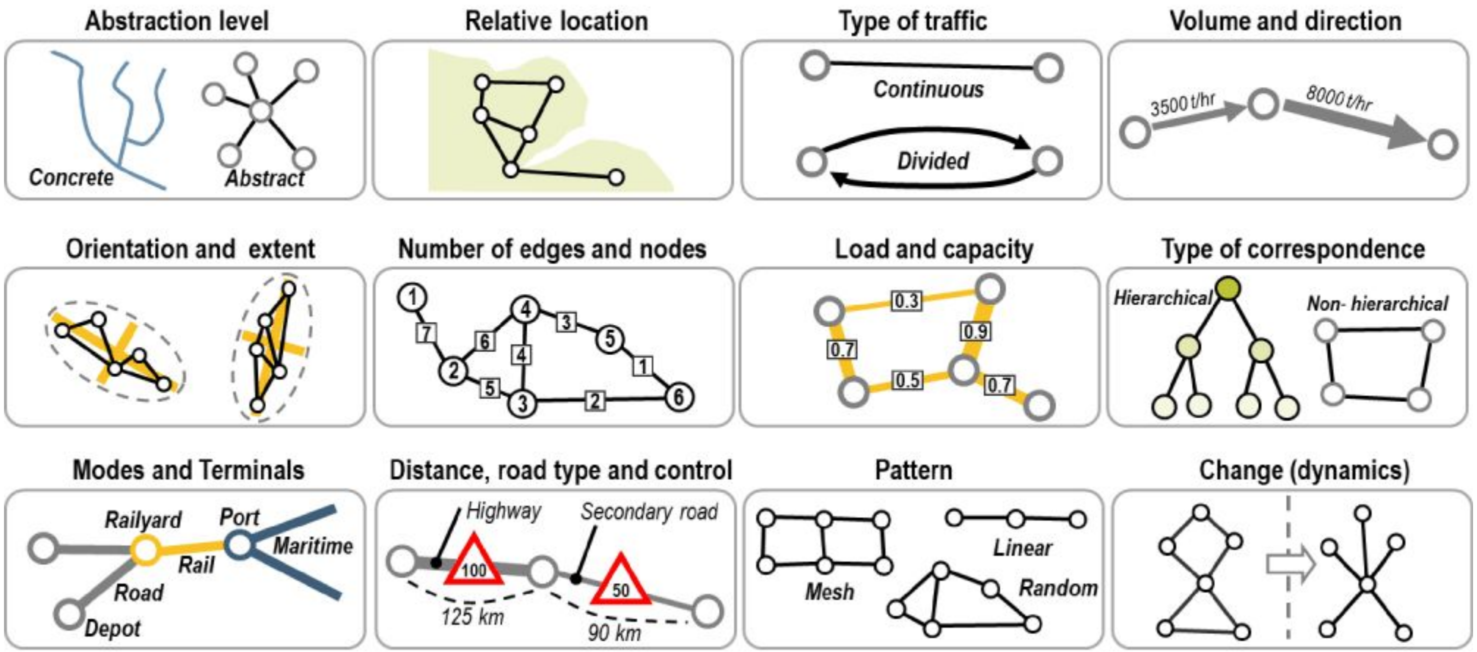
Graphs are everywhere (2/4)





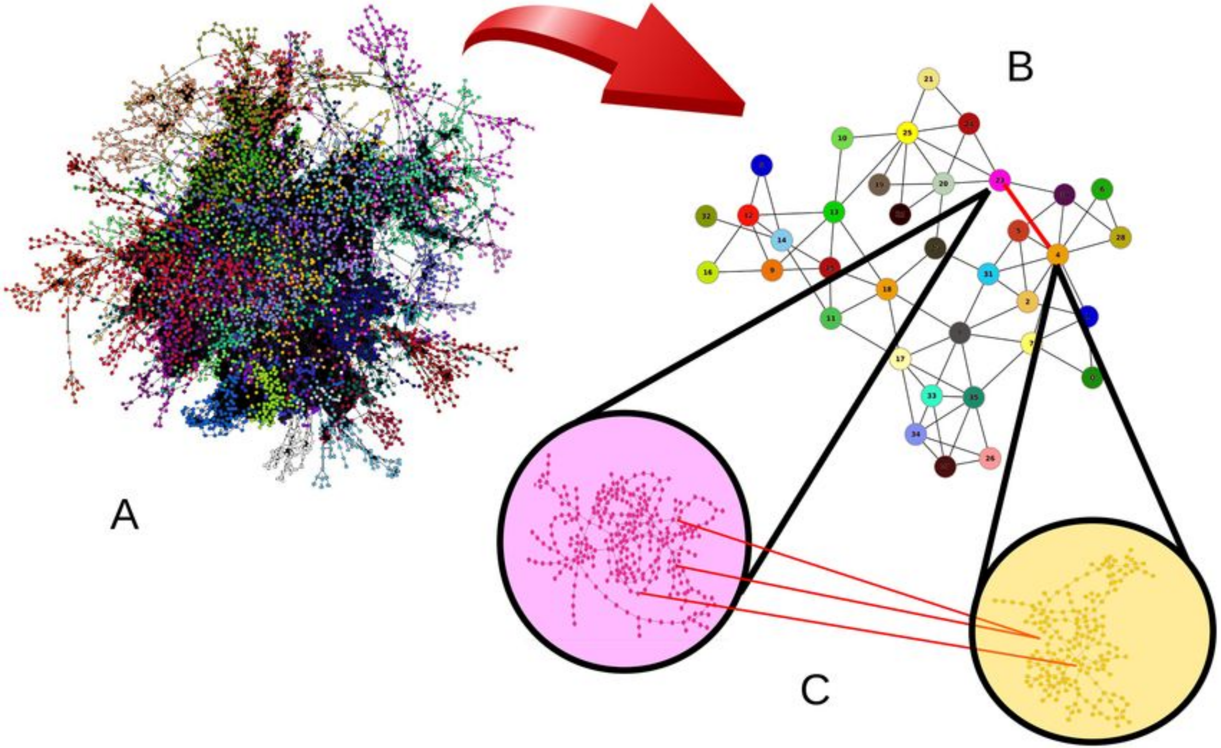
http://snap.stanford.edu/deepnetbio-ismb/ipynb/Graph+Convolutional+Prediction+of+Protein+Interactions+in+Yeast.html

Graphs are everywhere (3/4)



https://transportgeography.org/?page_id=719

Graphs are everywhere (4/4)

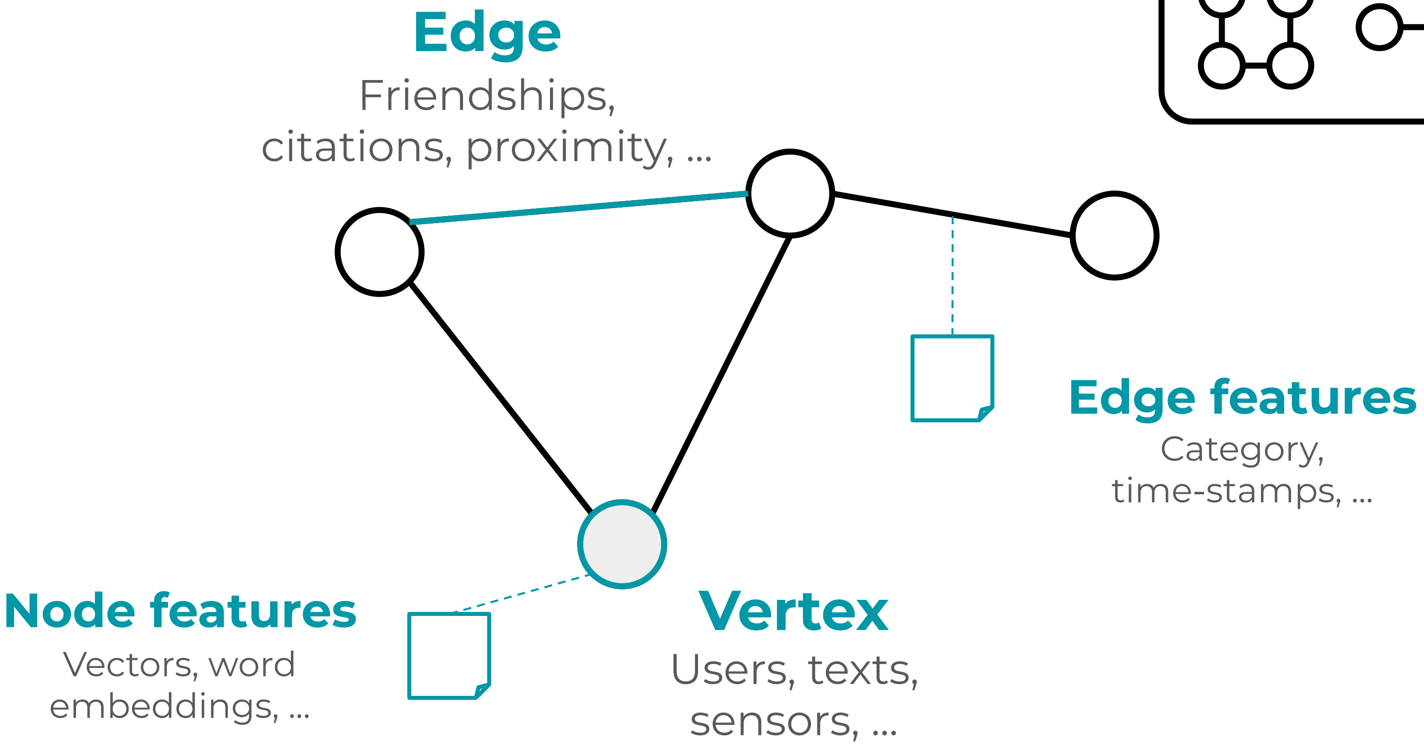


[Fast Fragmentation of Networks Using Module-Based Attacks]

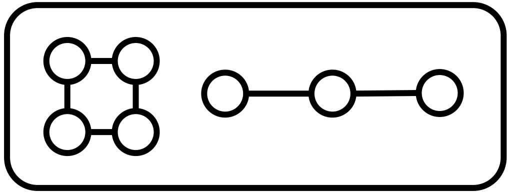
Graph and learning

Machine learning on graphs

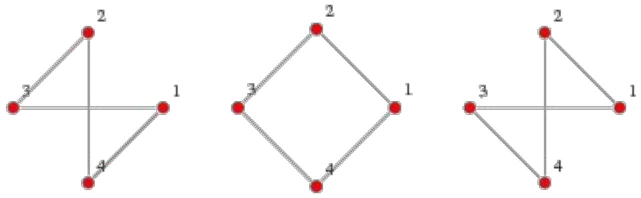
What is a graph?



Other graphs!



Graphs are (represented by?) matrices



$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Adjacency matrix

$$\mathbf{A} \in \mathbb{R}^{n \times n}$$

n vertices in the graph

Node features

$$\mathbf{X} \in \mathbb{R}^{n \times d}$$

each node has *d* features

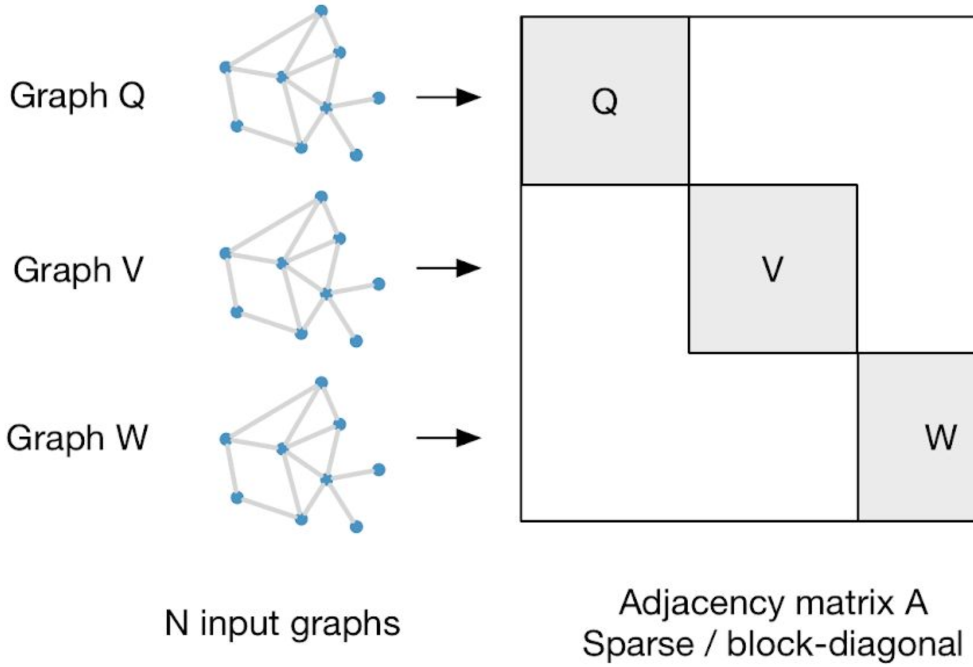
Edge features

$$\mathbf{E} \in \mathbb{R}^{e \times f}$$

each edge has *f* features

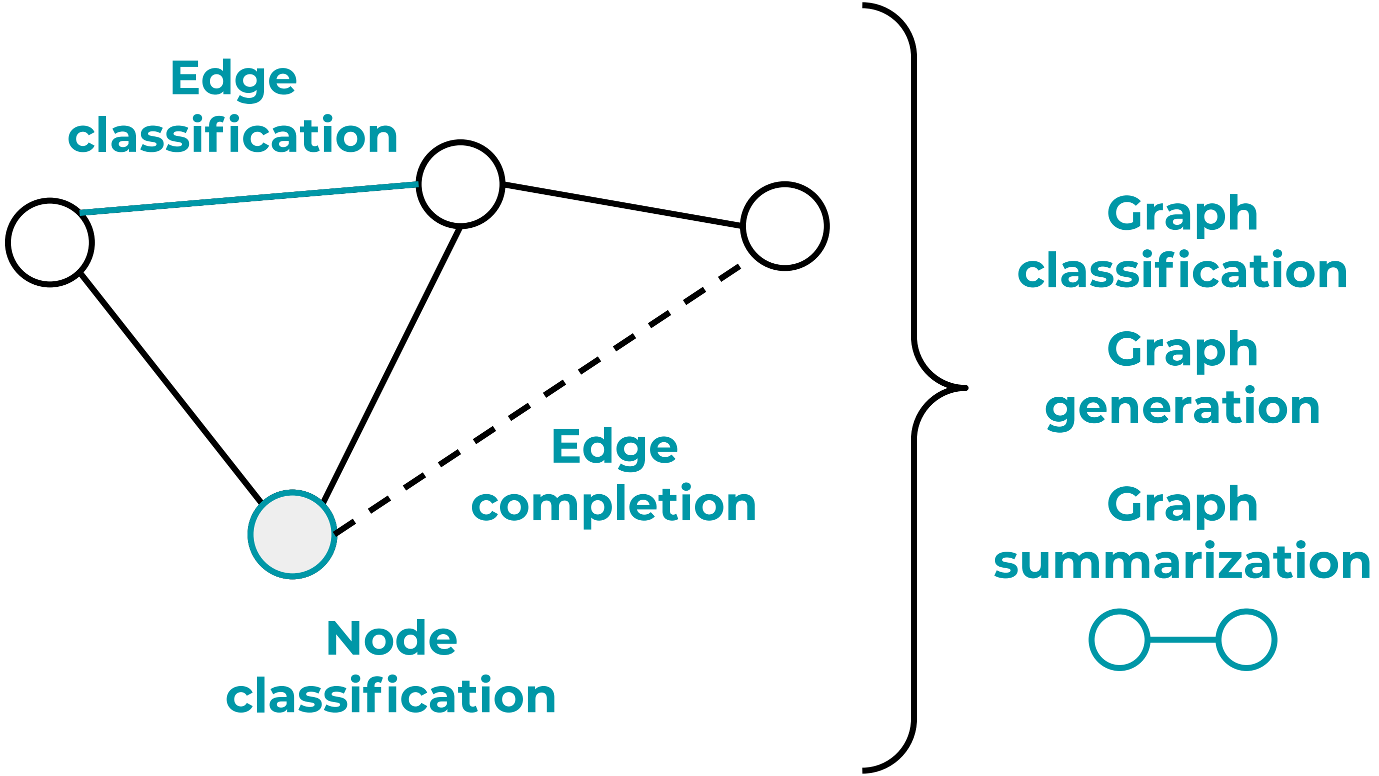
[Adjacency Matrix -- from Wolfram MathWorld]

Scaling up and batching

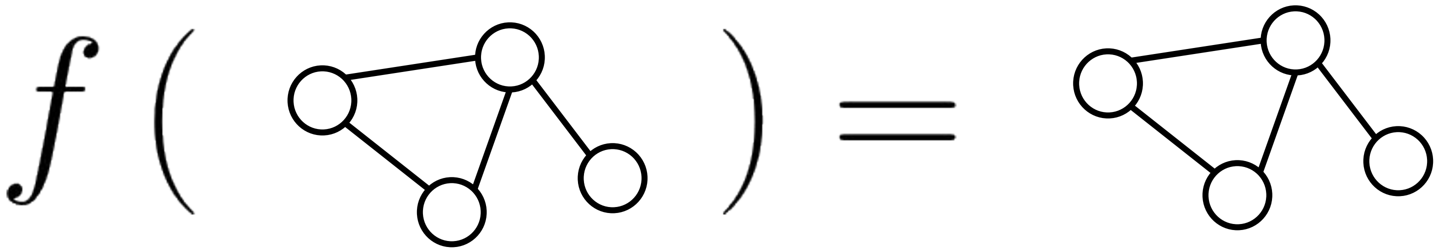


[tkipf/gcn: Implementation of Graph Convolutional Networks in TensorFlow]

What can we learn on a graph?



How can we learn on a graph?

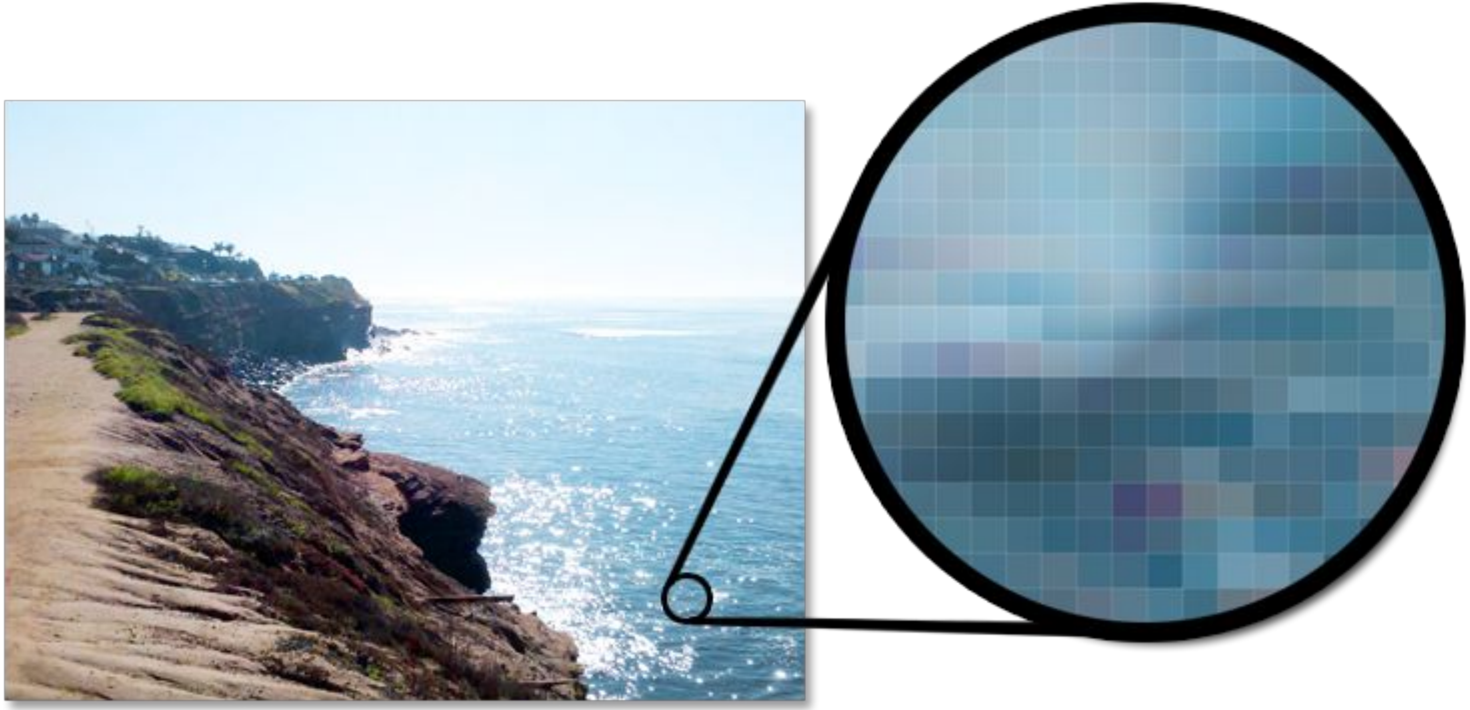


We want to do deep learning, hence *f* should be **differentiable**, **composable**, **scalable**.

Graph and learning

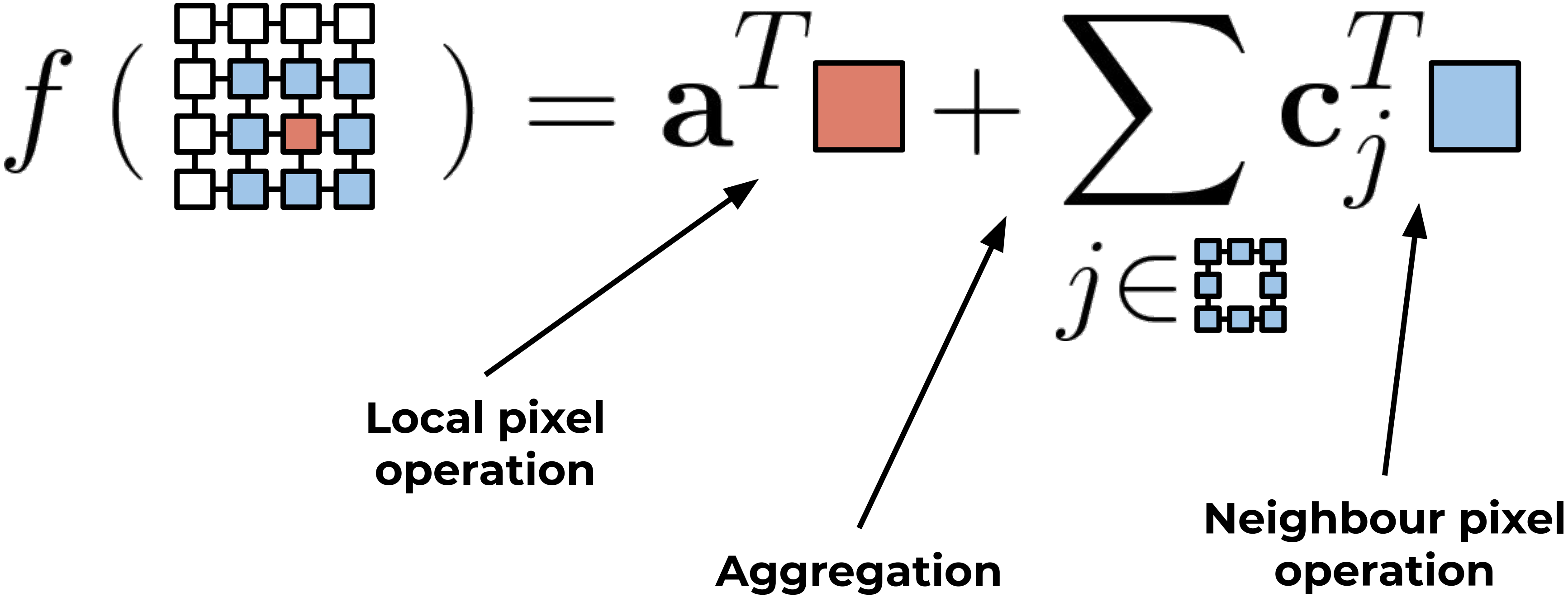
Defining a **graph layer**

Deep learning is about leveraging structure

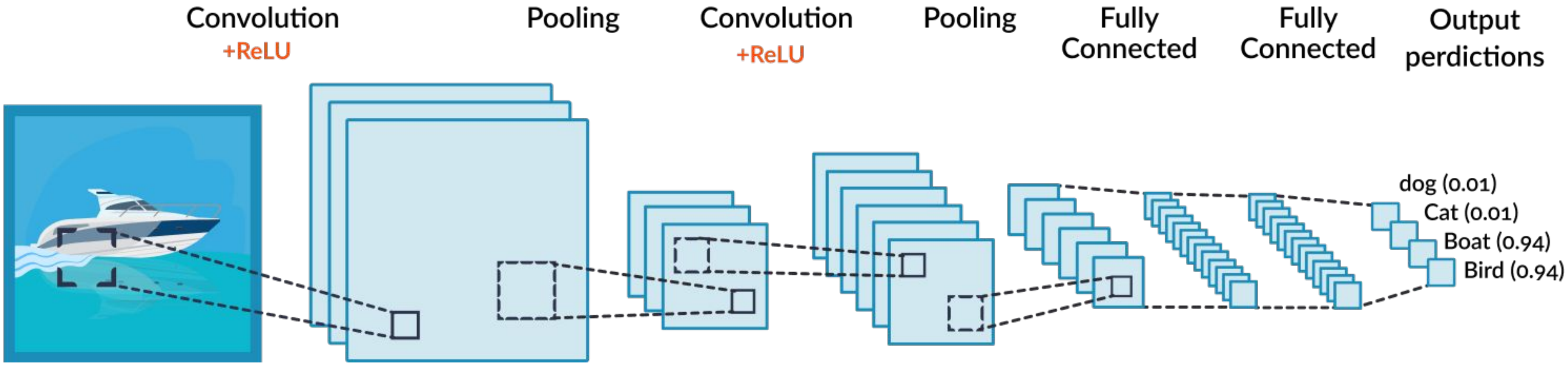


[Generate cross-stitch patterns from any image.]

Image convolutions



Convolutional neural networks



[Convolutional Neural Network Tutorial: From Basic to Advanced]

Images vs. graphs

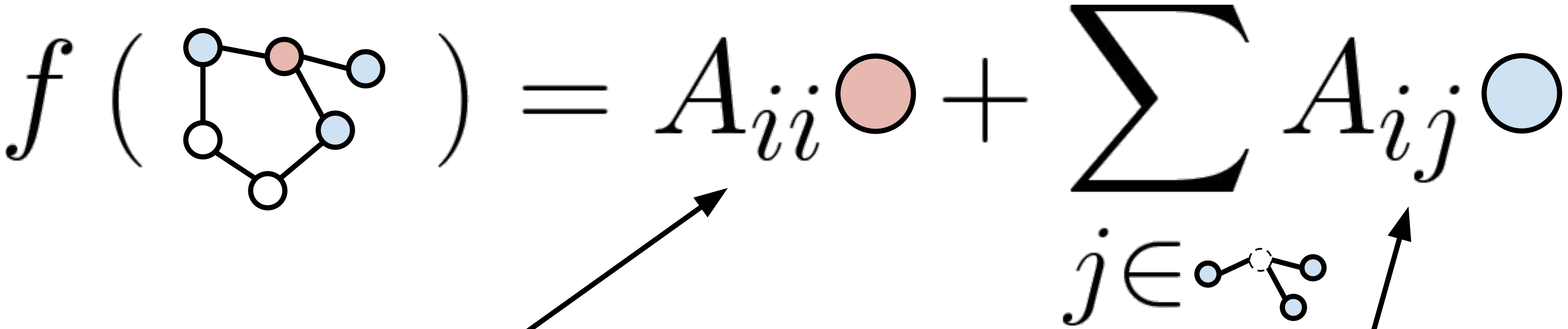
How much of the structure of an image do we find in graphs?

✔️ **Locality** (neighbourhood).

❌ Fixed number of neighbours.

❌ Neighbours have a definite ordering.

Graph convolutions

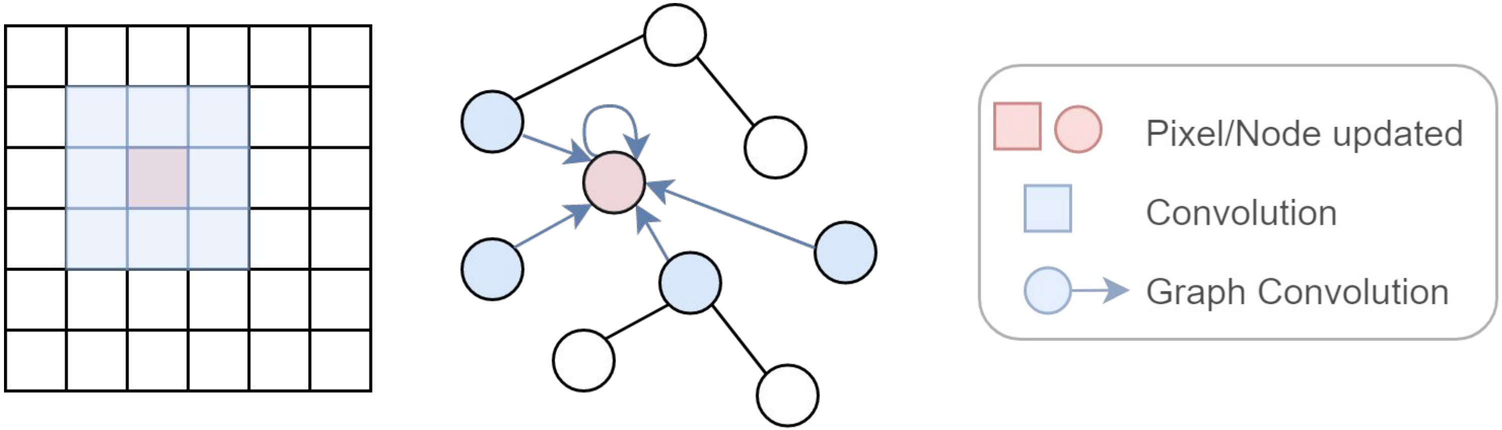


Local vertex operation

We use the adjacency matrix (or similar)

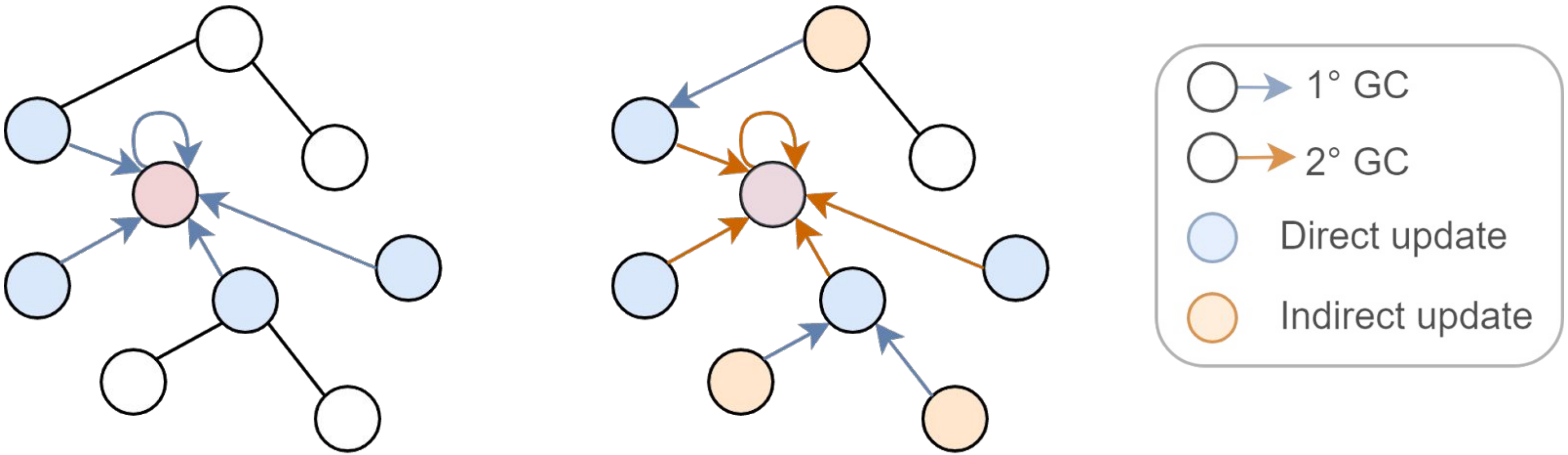
Note: the mixing coefficients are no more learnable!
Here, only local operations are trainable.

Graph convolution visualized



https://spindro.github.io/post/gnn/

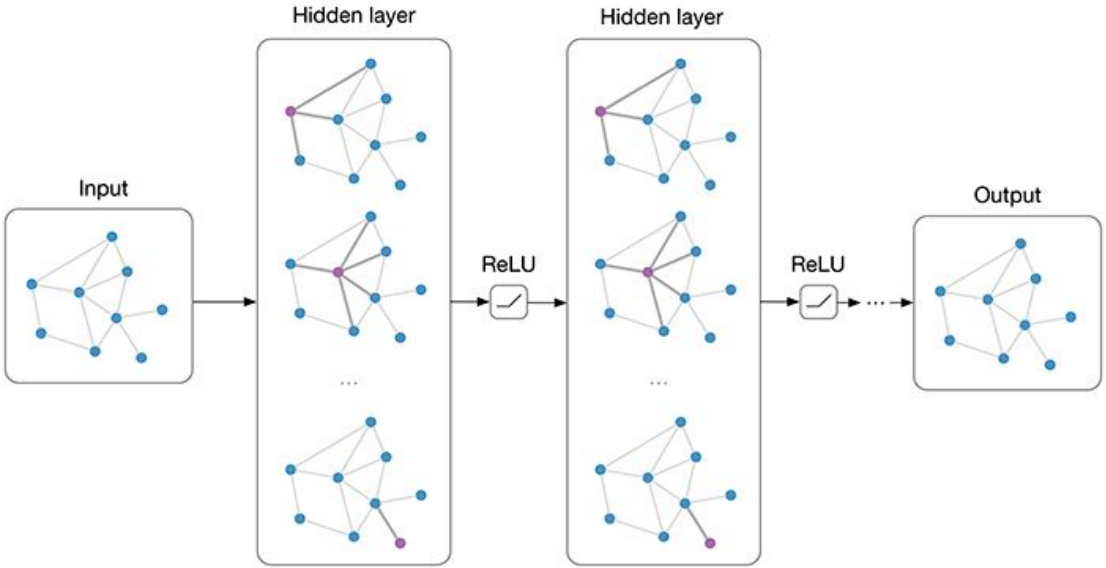
Stacking graph convolutional layers



Performing multiple updates increases the "receptive field" of each node.

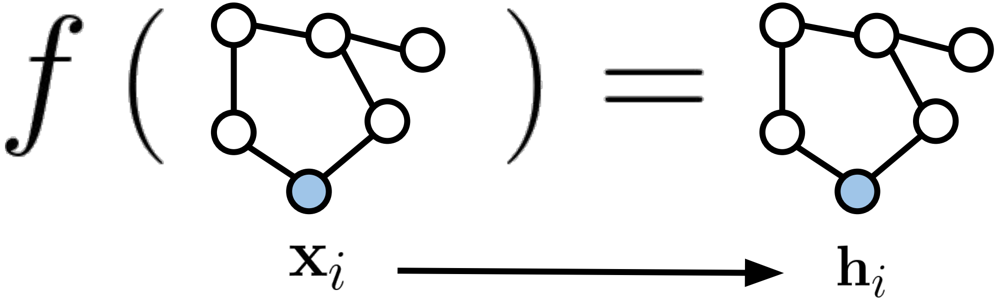
https://spindro.github.io/post/gnn/

Visualizing a graph convolutional network



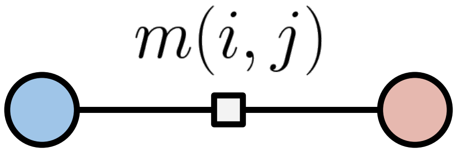
Kipf, T.N. and Welling, M., 2016. **Semi-supervised classification with graph convolutional networks**. arXiv preprint arXiv:1609.02907.

Tackling multiple tasks



1. Node classification: $$\mathrm{sotfmax}(\mathbf{h}_i)$$
2. Edge classification: $$\mathrm{sotfmax}(\mathbf{h}_i^T \mathbf{h}_j)$$
3. Graph classification: $$\mathrm{sotfmax}\left(\frac{1}{N}\sum_i \mathbf{h}_i\right)$$

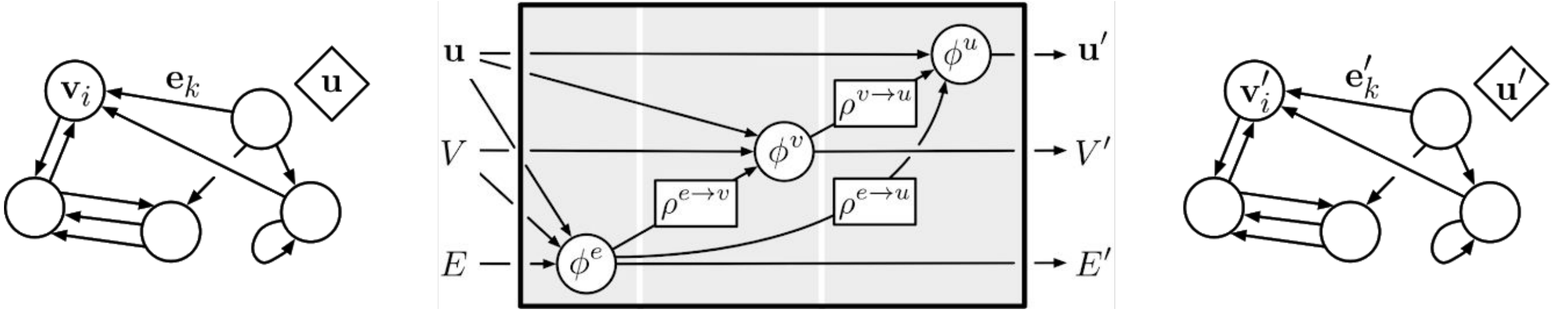
Message-passing neural networks



Instead of using directly the adjacency matrix, nodes can exchange **messages** with several mechanisms (e.g., attention models).

Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O. and Dahl, G.E., 2017, July. **Neural message passing for quantum chemistry**. In *International Conference on Machine Learning* (pp. 1263-1272). PMLR.

A more general setup



Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R. and Gulcehre, C., 2018. **Relational inductive biases, deep learning, and graph networks**. arXiv preprint arXiv:1806.01261.

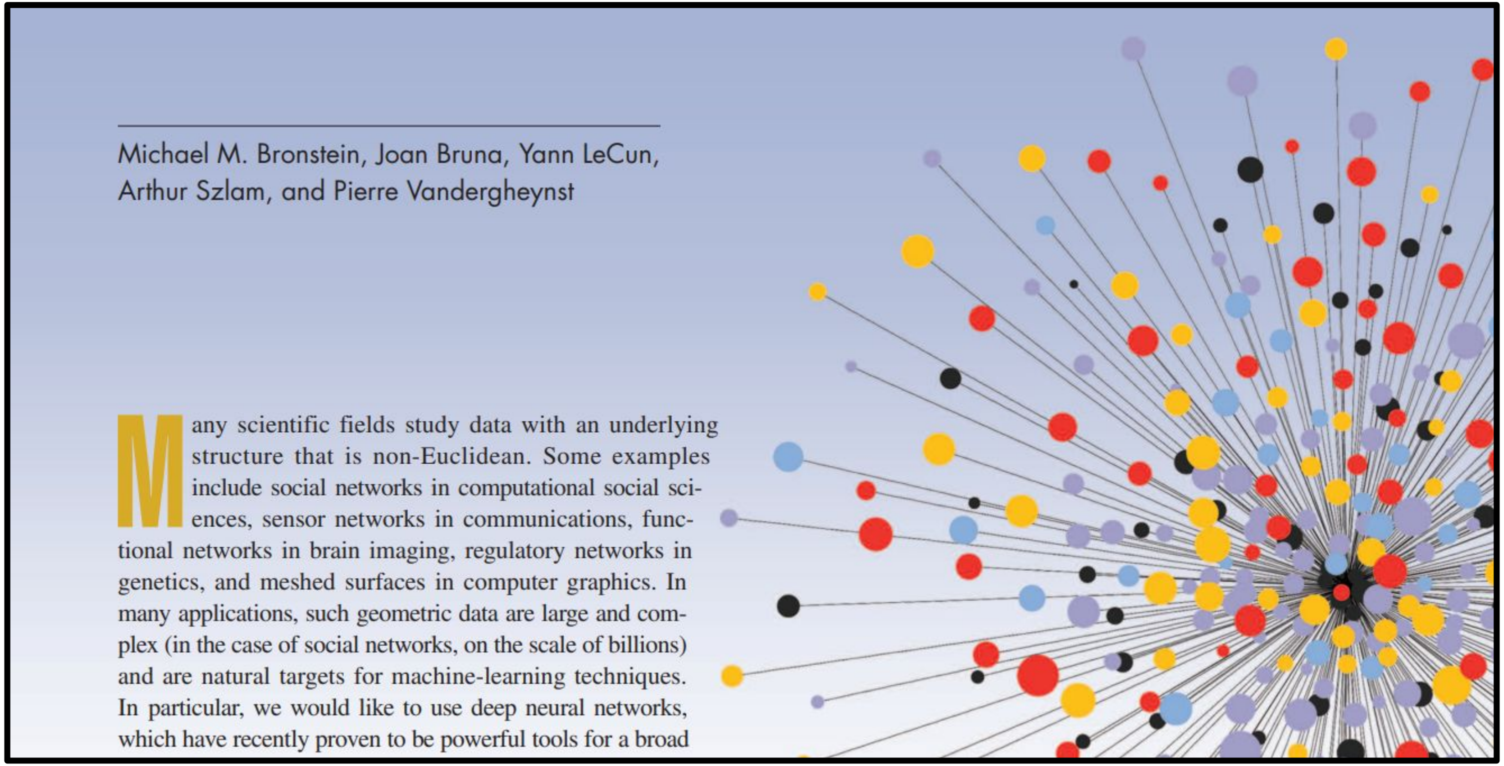
A zoo of techniques...

TABLE III: Summary of RecGNNs and ConvGNNs. Missing values ("-") in pooling and readout layers indicate that the method only experiments on node-level/edge-level tasks.

Approach	Category	Inputs	Pooling	Readout	Time Complexity
GNN* (2009) [15]	RecGNN	A, X, X^e	-	a dummy super node	$O(m)$
GraphESN (2010) [16]	RecGNN	A, X	-	mean	$O(m)$
GGNN (2015) [17]	RecGNN	A, X	-	attention sum	$O(m)$
SSE (2018) [18]	RecGNN	A, X	-	-	-
Spectral CNN (2014) [19]	Spectral-based ConvGNN	A, X	spectral clustering+max pooling	max	$O(n^3)$
Henaff et al. (2015) [20]	Spectral-based ConvGNN	A, X	spectral clustering+max pooling		$O(n^3)$
ChebNet (2016) [21]	Spectral-based ConvGNN	A, X	efficient pooling	sum	$O(m)$
GCN (2017) [22]	Spectral-based ConvGNN	A, X	-	-	$O(m)$
CayleyNet (2017) [23]	Spectral-based ConvGNN	A, X	mean/graclus pooling	-	$O(m)$
AGCN (2018) [40]	Spectral-based ConvGNN	A, X	max pooling	sum	$O(n^2)$
DualGCN (2018) [41]	Spectral-based ConvGNN	A, X	-	-	$O(m)$
NN4G (2009) [24]	Spatial-based ConvGNN	A, X	-	sum/mean	$O(m)$
DCNN (2016) [25]	Spatial-based ConvGNN	A, X	-	mean	$O(n^2)$

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C. and Yu, P.S., 2019. **A comprehensive survey on graph neural networks**. arXiv preprint arXiv:1901.00596.

Geometric deep learning


Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst

Many scientific fields study data with an underlying structure that is non-Euclidean. Some examples include social networks in computational social sciences, sensor networks in communications, functional networks in brain imaging, regulatory networks in genetics, and meshed surfaces in computer graphics. In many applications, such geometric data are large and complex (in the case of social networks, on the scale of billions) and are natural targets for machine-learning techniques. In particular, we would like to use deep neural networks, which have recently proven to be powerful tools for a broad

Bronstein, M.M. et al., 2017. **Geometric deep learning: going beyond Euclidean data**. IEEE Signal Processing Magazine, 34(4), pp.18-42.

GNNs before deep learning!

IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 20, NO. 1, JANUARY 2009

The Graph Neural Network Model

Franco Scarselli, Marco Gori, *Fellow, IEEE*, Ah Chung Tsoi, Markus Hagenbuchner, *Member, IEEE*, and Gabriele Monfardini

***Abstract*—Many underlying relationships among data in several areas of science and engineering, e.g., computer vision, molecular chemistry, molecular biology, pattern recognition, and data mining, can be represented in terms of graphs. In this paper, we propose a new neural network model, called graph neural network (GNN) model, that extends existing neural network methods for processing the data represented in graph domains. This GNN model, which can directly process most of the practically useful types of graphs, e.g., acyclic, cyclic, directed, and undirected, implements a function $\tau(\boldsymbol{G}, n) \in \mathbb{R}^m$ that maps a graph $\boldsymbol{G}$ and one of its nodes n into an m-dimensional Euclidean space. A supervised learning algorithm is derived to estimate the parameters of the proposed GNN model. The computational cost of the proposed algorithm is also considered. Some experimental results are shown to validate the proposed learning algorithm, and to demonstrate its generalization capabilities.**

ples a function τ that maps a graph G and one of its nodes n to a vector of reals[1]: $\tau(\boldsymbol{G}, n) \in \mathbb{R}^m$. Applications to a graphical domain can generally be divided into two broad classes, called *graph-focused* and *node-focused* applications, respectively, in this paper. In *graph-focused* applications, the function τ is independent of the node n and implements a classifier or a regressor on a graph structured data set. For example, a chemical compound can be modeled by a graph $\boldsymbol{G}$, the nodes of which stand for atoms (or chemical groups) and the edges of which represent chemical bonds [see Fig. 1(a)] linking together some of the atoms. The mapping $\tau(\boldsymbol{G})$ may be used to estimate the probability that the chemical compound causes a certain disease [13]. In Fig. 1(b), an image is represented by a region adjacency graph where nodes denote homogeneous regions of intensity of

Graph and learning

Some famous **applications**

Fake news detection on Twitter

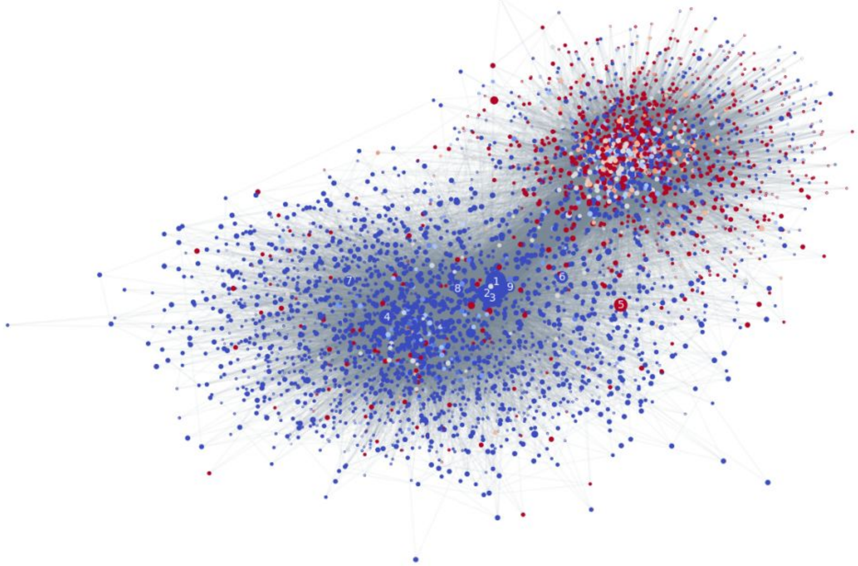
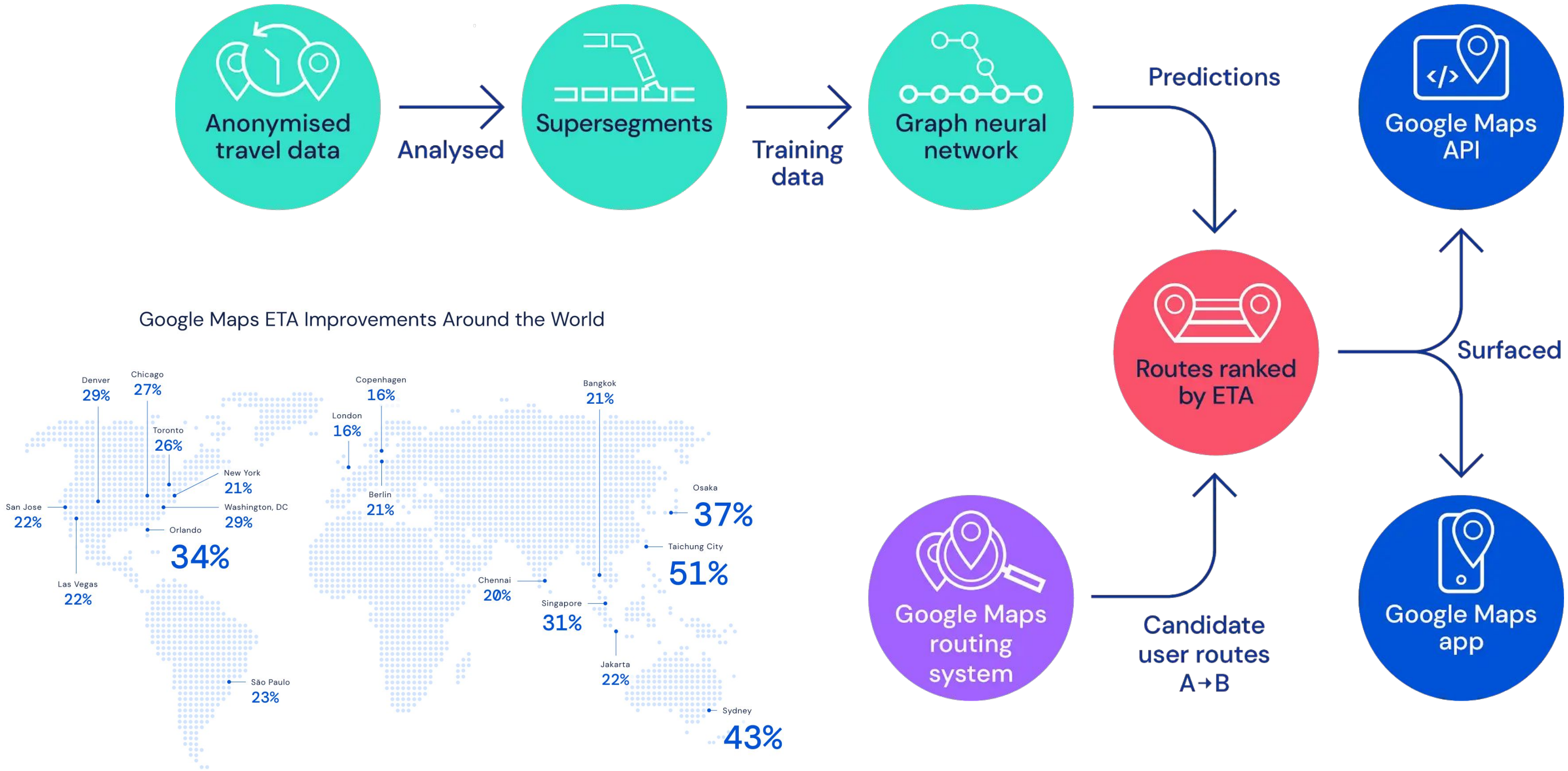


Figure 4: Subset of the Twitter network used in our study with estimated user credibility. Vertices represent users, gray edges the social connections. Vertex color and size encode the user credibility (blue = reliable, red = unreliable) and number of followers of each user, respectively. Numbers 1 to 9 represent the nine users with most followers.

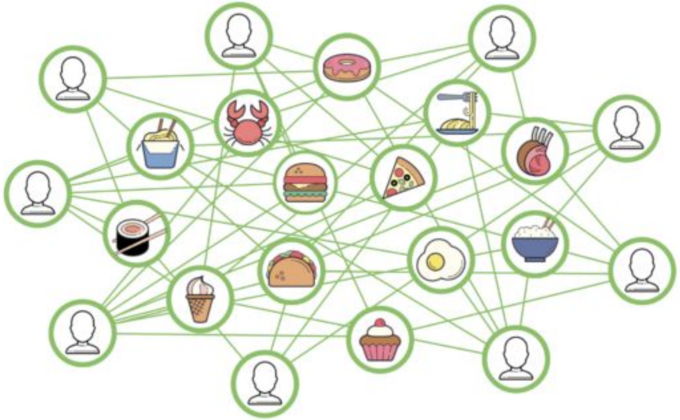
Monti, F., Frasca, F., Eynard, D., Mannion, D. and Bronstein, M.M., 2019. **[Fake News Detection on Social Media using Geometric Deep Learning]**. *arXiv preprint arXiv:1902.06673.*

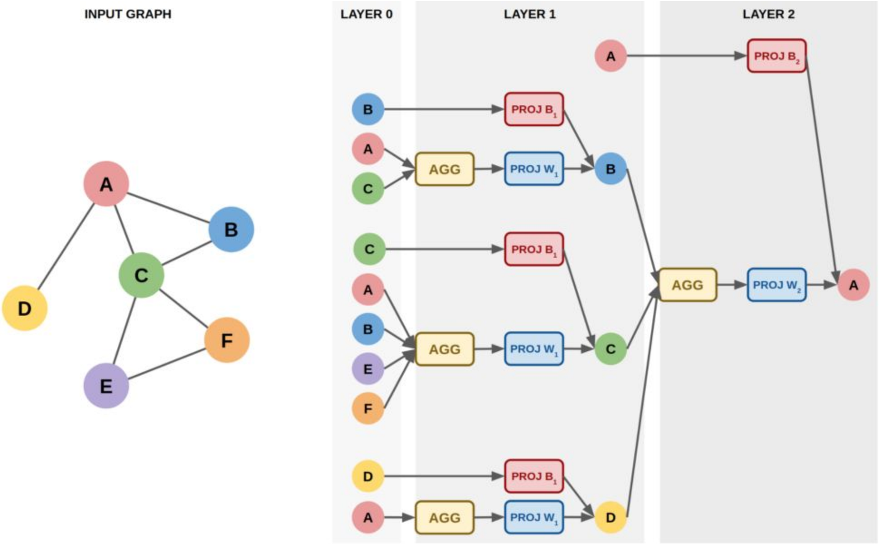
Traffic prediction on Google Maps



[Traffic prediction with advanced Graph Neural Networks]()

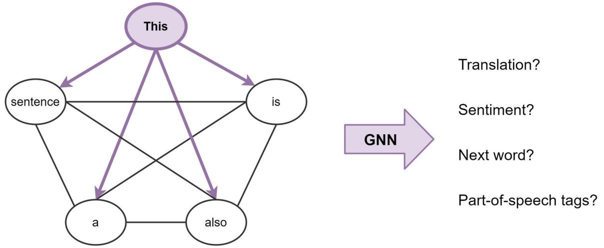
Recommending systems in Uber





[Food Discovery with Uber Eats: Using Graph Learning to Power Recommendations]

Graph networks in classic deep learning

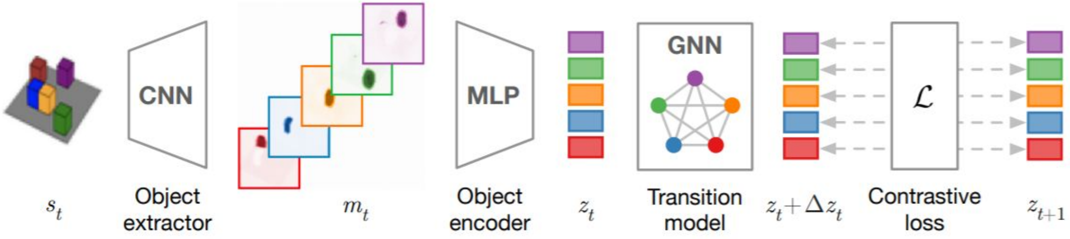


Transformers are basically GNNs on fully-connected graphs!

https://thegradient.pub/transformers-are-graph-neural-networks/

[[1911.12247] Contrastive Learning of Structured World Models]()

GNNs can be used to include **relational reasoning** in classical models!



Graph and learning

Open problems and **challenges**

Scaling up to huge graphs

Open Graph Benchmark

Benchmark datasets, data loaders and evaluators for graph machine learning

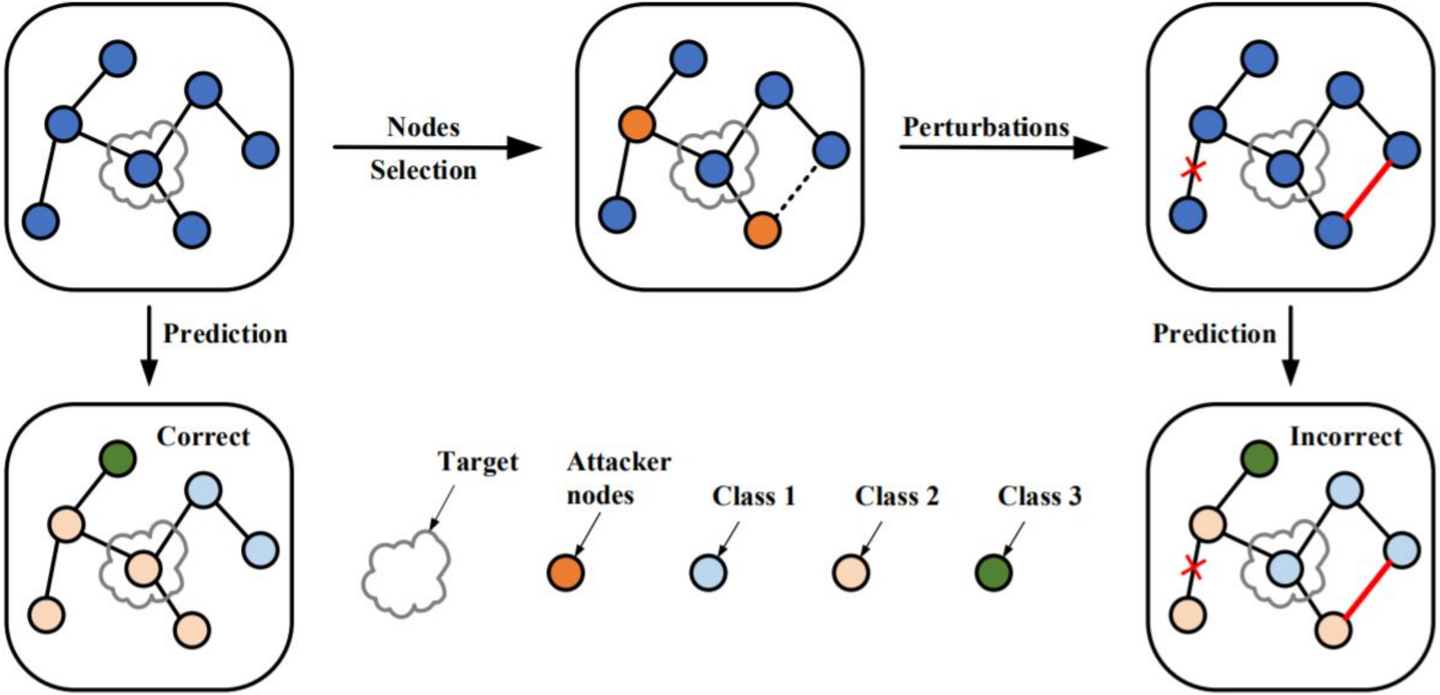
GET STARTED VIEW UPDATES

The Open Graph Benchmark (OGB) is a collection of realistic, large-scale, and diverse benchmark datasets for machine learning on graphs. OGB datasets are automatically downloaded, processed, and split using the OGB Data Loader. The model performance can be evaluated using the OGB Evaluator in a unified manner.

OGB is a community-driven initiative in active development. We expect the benchmark datasets to

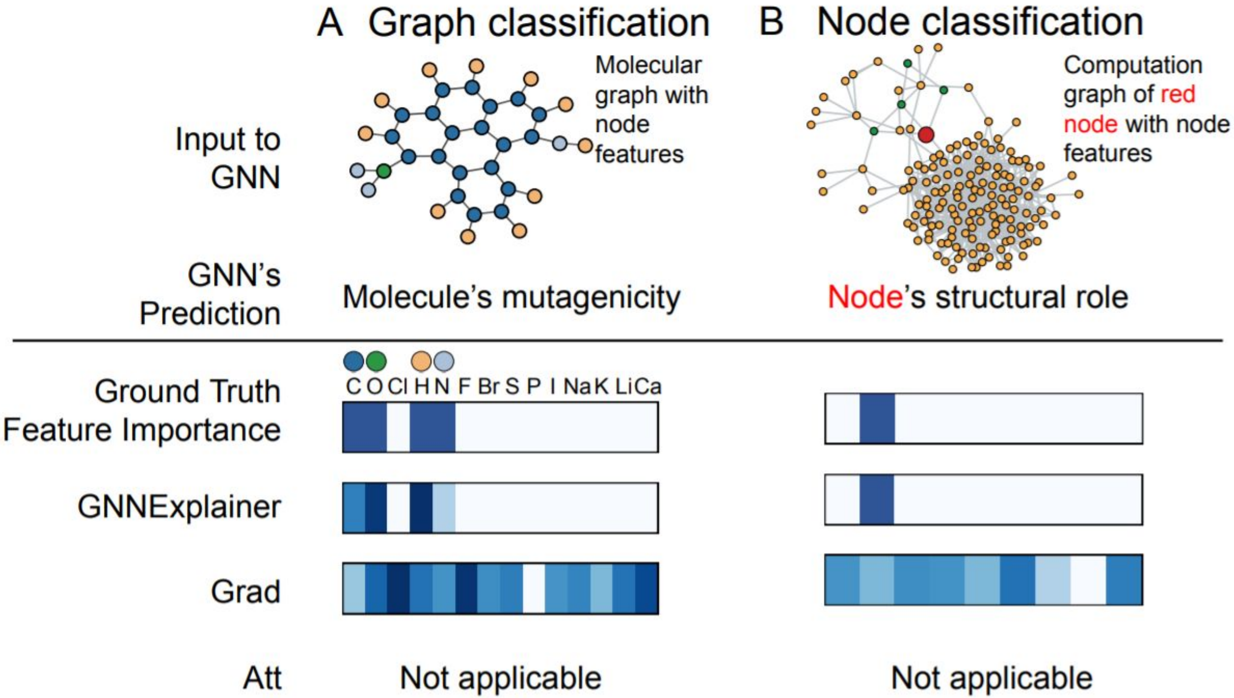
OPEN GRAPH BENCHMARK

Robustness



[Adversarial Attack on Large Scale Graph](https://github.com/safe-graph/graph-adversarial-learning-literature),
https://github.com/safe-graph/graph-adversarial-learning-literature

Interpretability



GNNExplainer (*Ying et al., 2019*)

Software?

DeepGraphLibrary

Amazon based, production-ready

PyTorch geometric

PyTorch, more research oriented

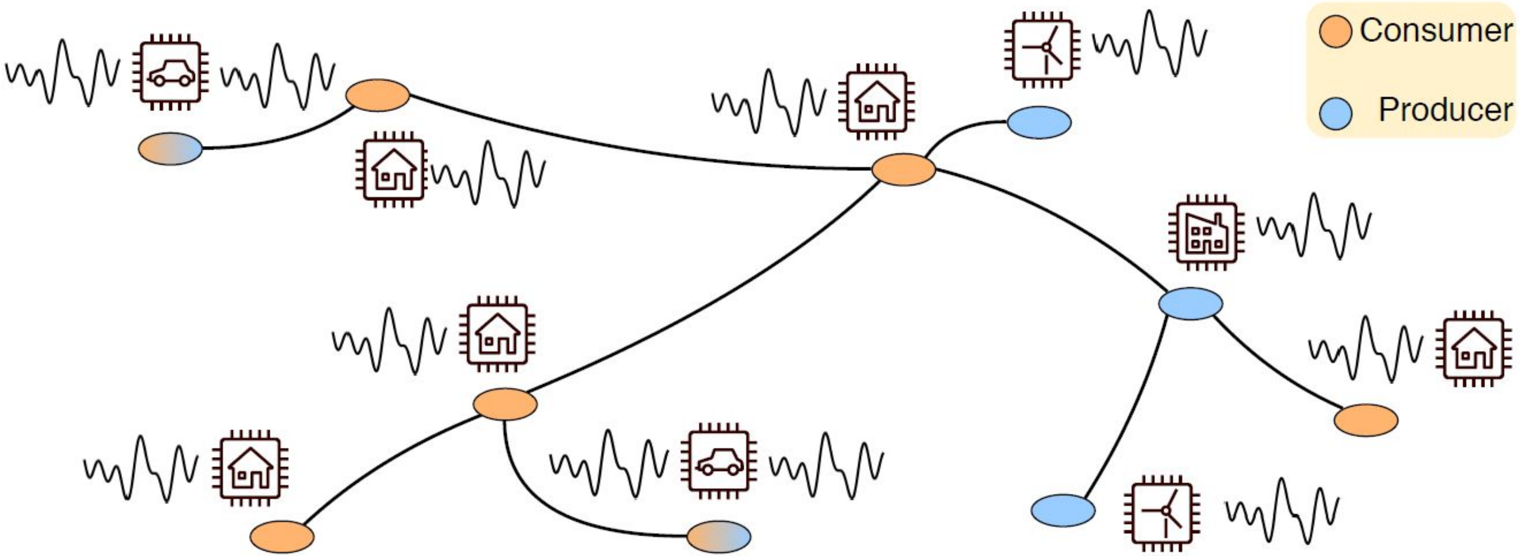
Spektral

Keras-like, TensorFlow 2.0

Distributed graphs

When graphs go **physical**

Distributed energy grids



What happens when our graph is also **physically distributed**?

An example of application

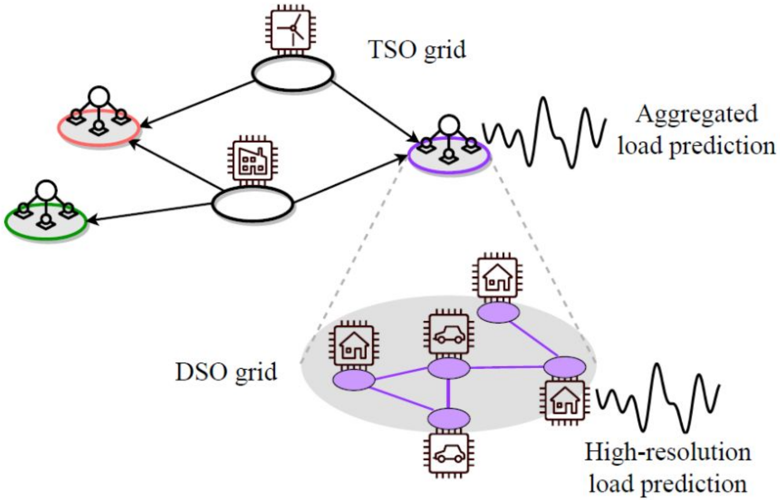
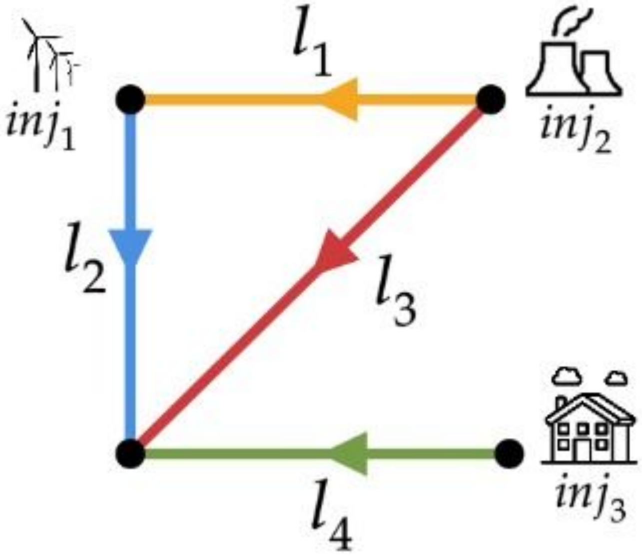
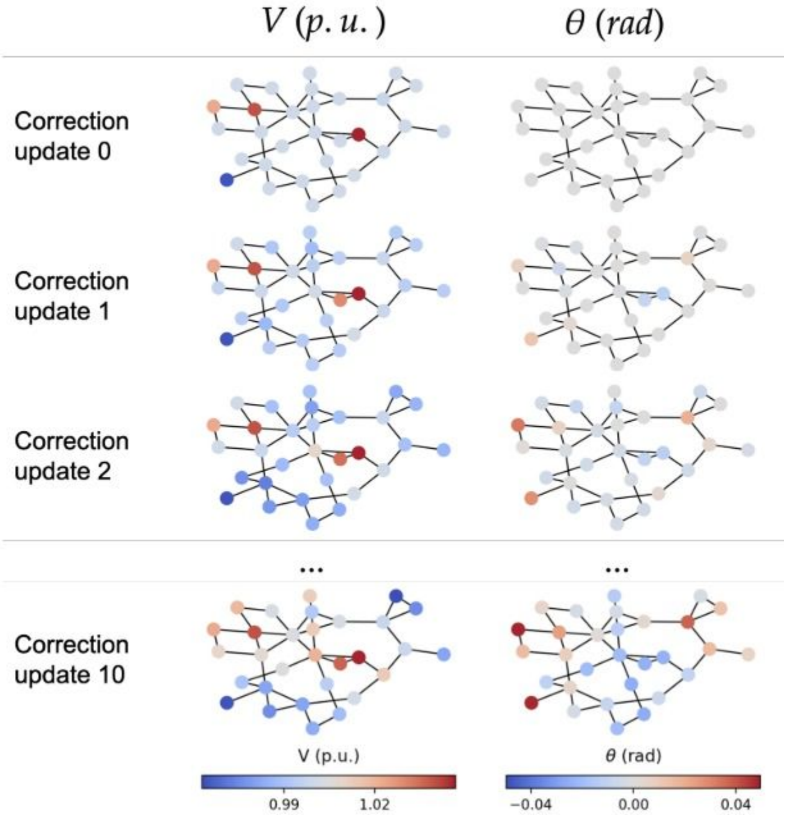


Figure 2: High-resolution load predictions are obtained on the single nodes (houses, electrical vehicles) in the low-voltage grid handled by the DSO. Through graph coarsening, the loads are aggregated at the level of resolution required to balance energy supply, provided by power plants, and the demand on the high-voltage TSO grid.

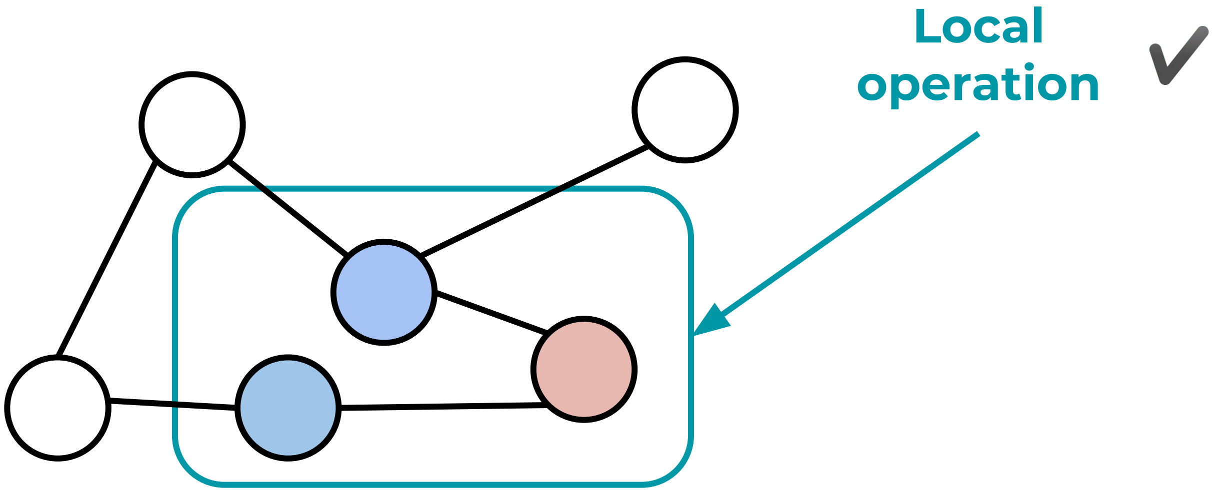
Another example



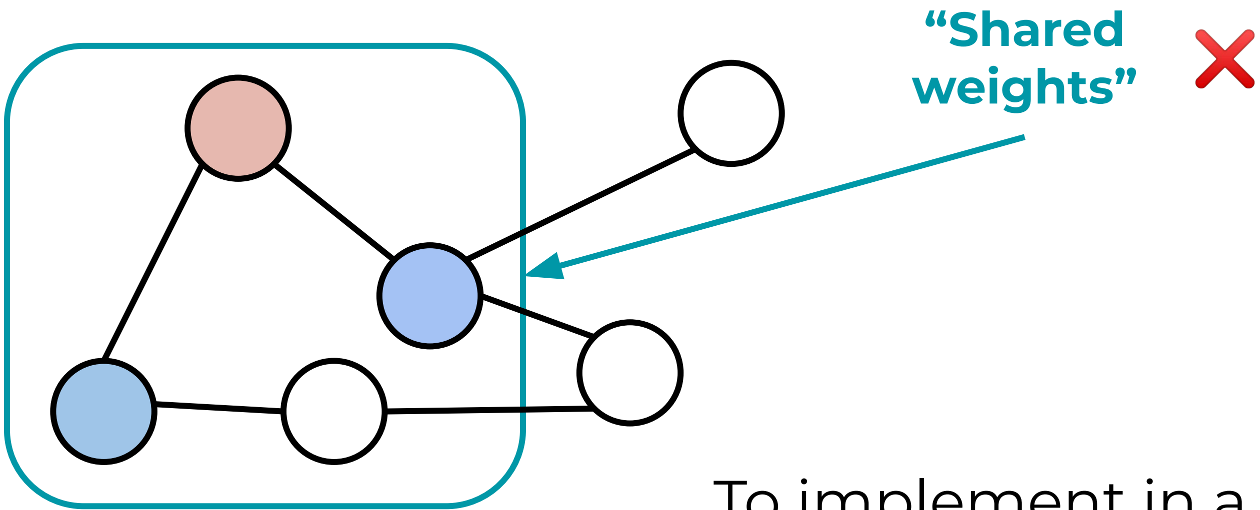


Donon, B., Donnot, B., Guyon, I. and Marot, A., 2019. **Graph neural solver for power systems**. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

Partially distributed graph networks



Partially distributed graph networks



To implement in a fully decentralized fashion, nodes require perfect coordination.

Fully distributed GCNs

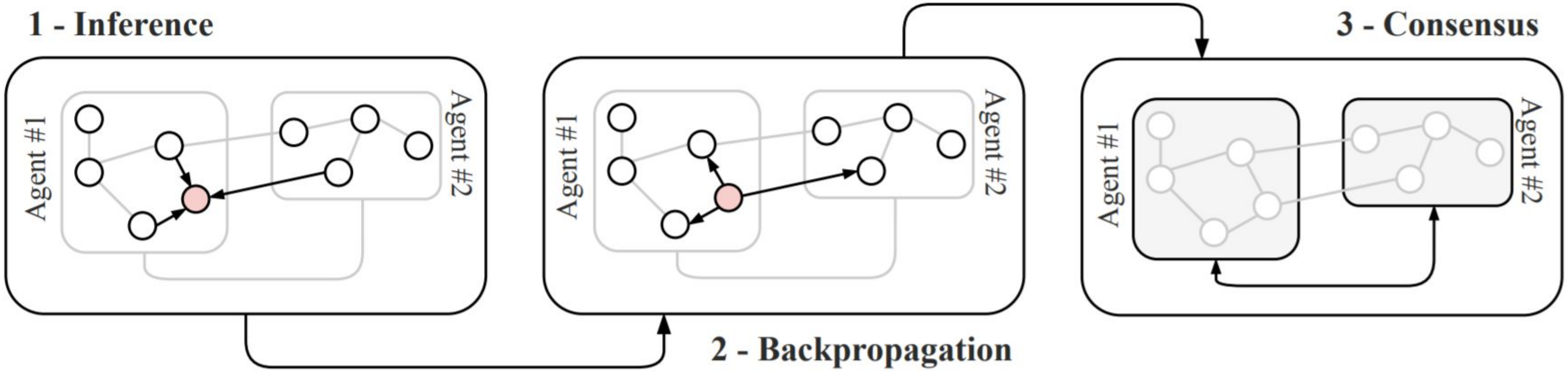


Fig. 1. Illustration of the proposed approach. In step 1, nodes communicate to perform inference. In step 2, a symmetric communication phase is executed to compute local gradients. In step 3, agents exchange local variables to asymptotically reach agreement. For steps 1-2, a representative active node is shown in red. Directed arrows show the flow of messages.

Scardapane, S., Spinelli, I. and Di Lorenzo, P., 2020. **Distributed Training of Graph Convolutional Networks**. *IEEE Transactions on Signal and Information Processing over Networks*.

It works!

TABLE II
RESULTS OF THREE DISTRIBUTED ALGORITHMS ON THE TRAFFIC PREDICTION BENCHMARKS FROM SECTION V-F

Dataset	Linear	GCN	Order-2 GCN
PeMSD4	0.231 ± 0.001	0.148 ± 0.002	0.126 ± 0.001
PeMSD8	0.266 ± 0.001	0.185 ± 0.002	0.147 ± 0.003

Nodes are traffic sensors in San Bernardino (California), with multiple base stations learning with no central coordination.

Scardapane, S., Spinelli, I. and Di Lorenzo, P., 2020. **Distributed Training of Graph Convolutional Networks**. *IEEE Transactions on Signal and Information Processing over Networks.*

Conclusions

Summarizing geometric deep learning

Key takeaways

1. Graph neural networks are powerful deep learning models that are surprisingly ubiquitous.

2. Scaling up and including temporal or heterogeneous information is challenging.

3. Huge research topic, still “under the radar” in industrial contexts.

Thanks for listening!

Simone Scardapane
Assistant Professor

Indro Spinelli
PhD Student

https://www.sscardapane.it/

https://spindro.github.io/

https://twitter.com/s_scardapane

https://twitter.com/IndroSpinelli

BTW, check out our association:
Italian Association for Machine Learning

And why not our podcast?
Smarter Podcast