On contextual architectures for probabilistic learning on graphs

CIML Reading Workshop, Jan 23, 2020
Speaker: Federico Errica

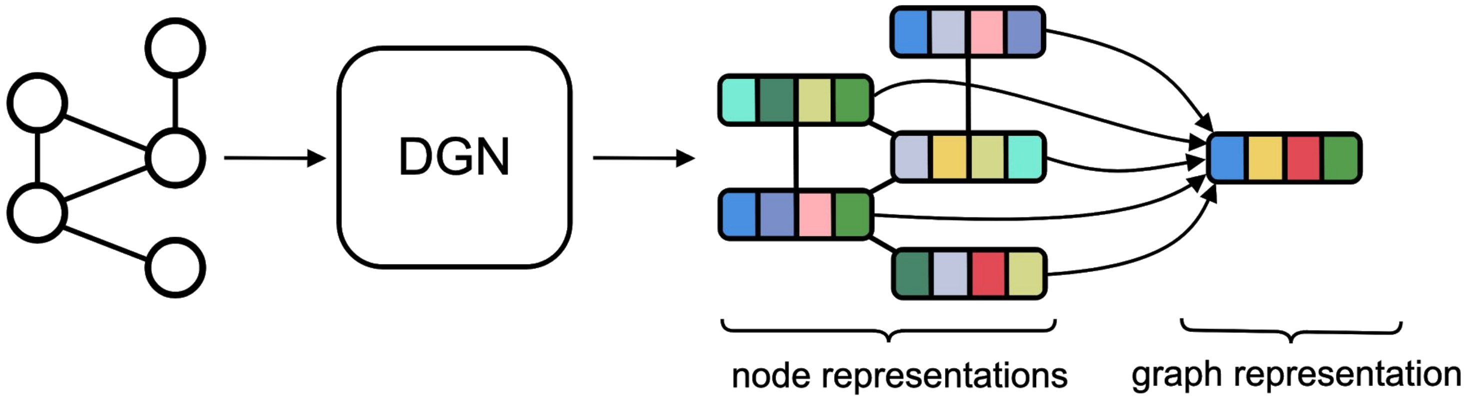
Outline

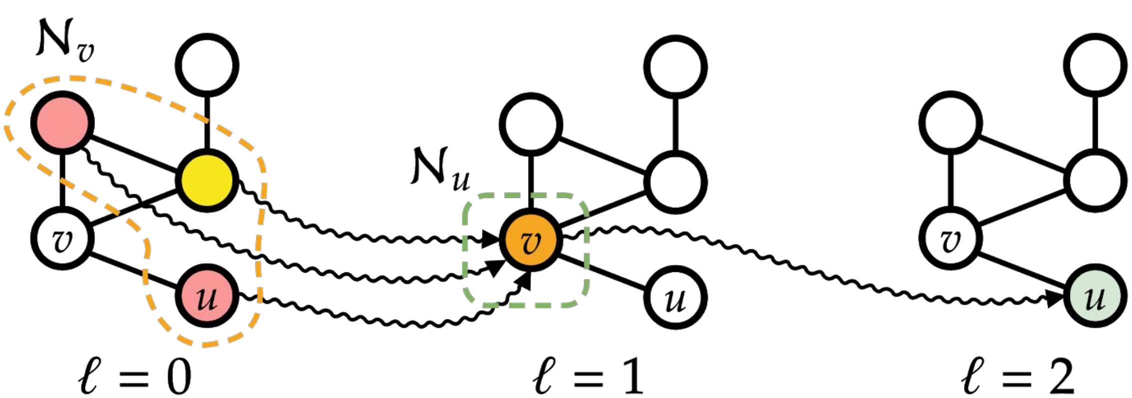
- Historical Background
- Representation learning on graphs
- Research proposal
- How
- Q&A

A bit of history

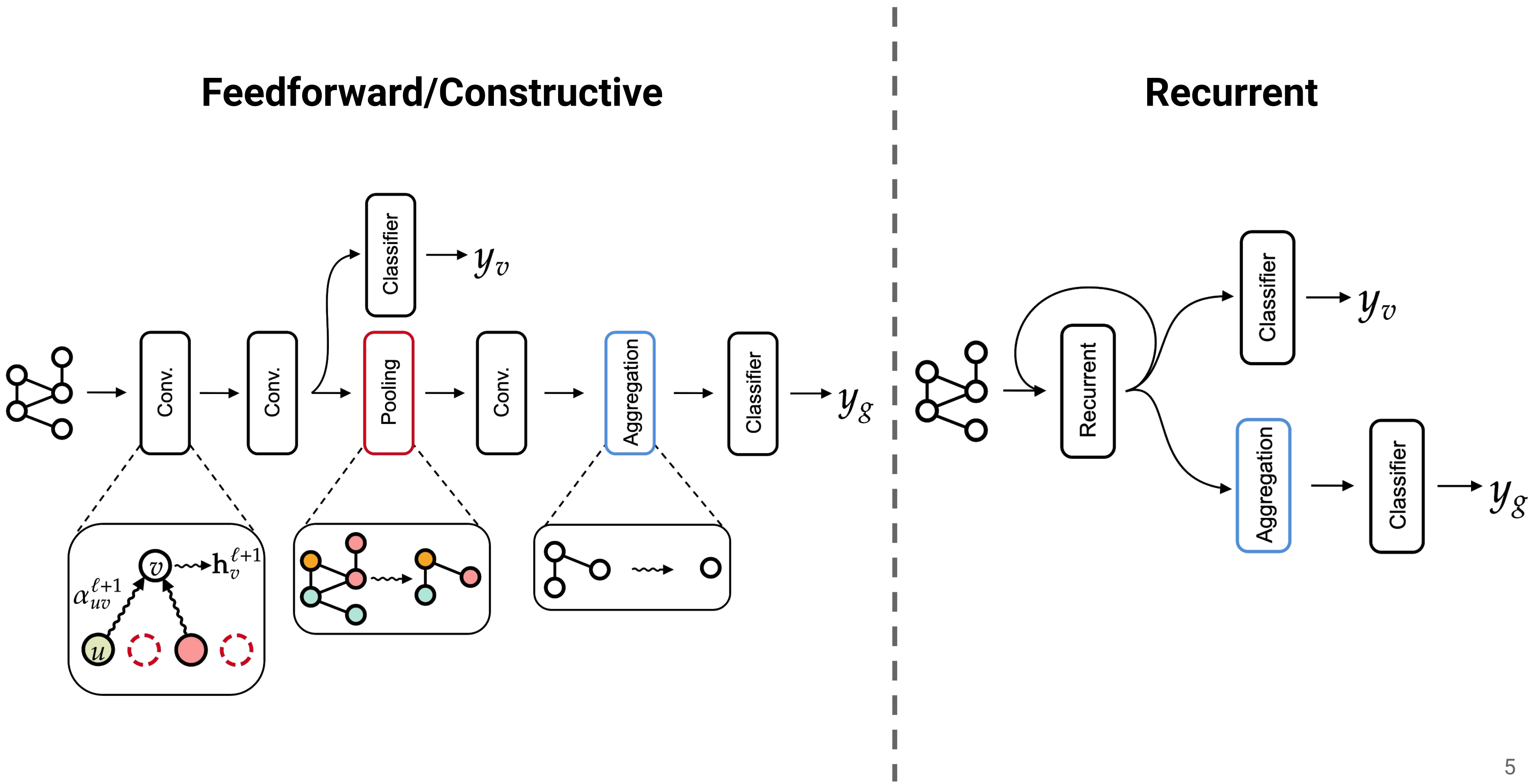
- Supervised neural networks for the classification of structures
  Sperduti A. and Starita A., IEEE Transactions on Neural Networks, 1997.
- A general framework for adaptive processing of data structures
  Frasconi P., Gori M., and Sperduti A., IEEE transactions on Neural Networks, 1998
- Application of Recursive Cascade-Correlation Networks to Chemical Structures
  Bianucci A. M., Micheli A., Sperduti A., and Starita A., Applied Intelligence, 2000.
- *Neural network for graphs: A contextual constructive approach*
  Micheli A., IEEE Transaction on Neural Networks, 2009
- *The graph neural network model*
  Scarselli F., Gori M., Tsoi A.C. , Hagenbuchner M., Monfardini G. IEEE Transactions on Neural Networks, 2009
- Graph echo state networks
  Gallicchio C. and Micheli A. The 2010 International Joint Conference on Neural Networks (IJCNN), 2010.

Representation learning on graphs





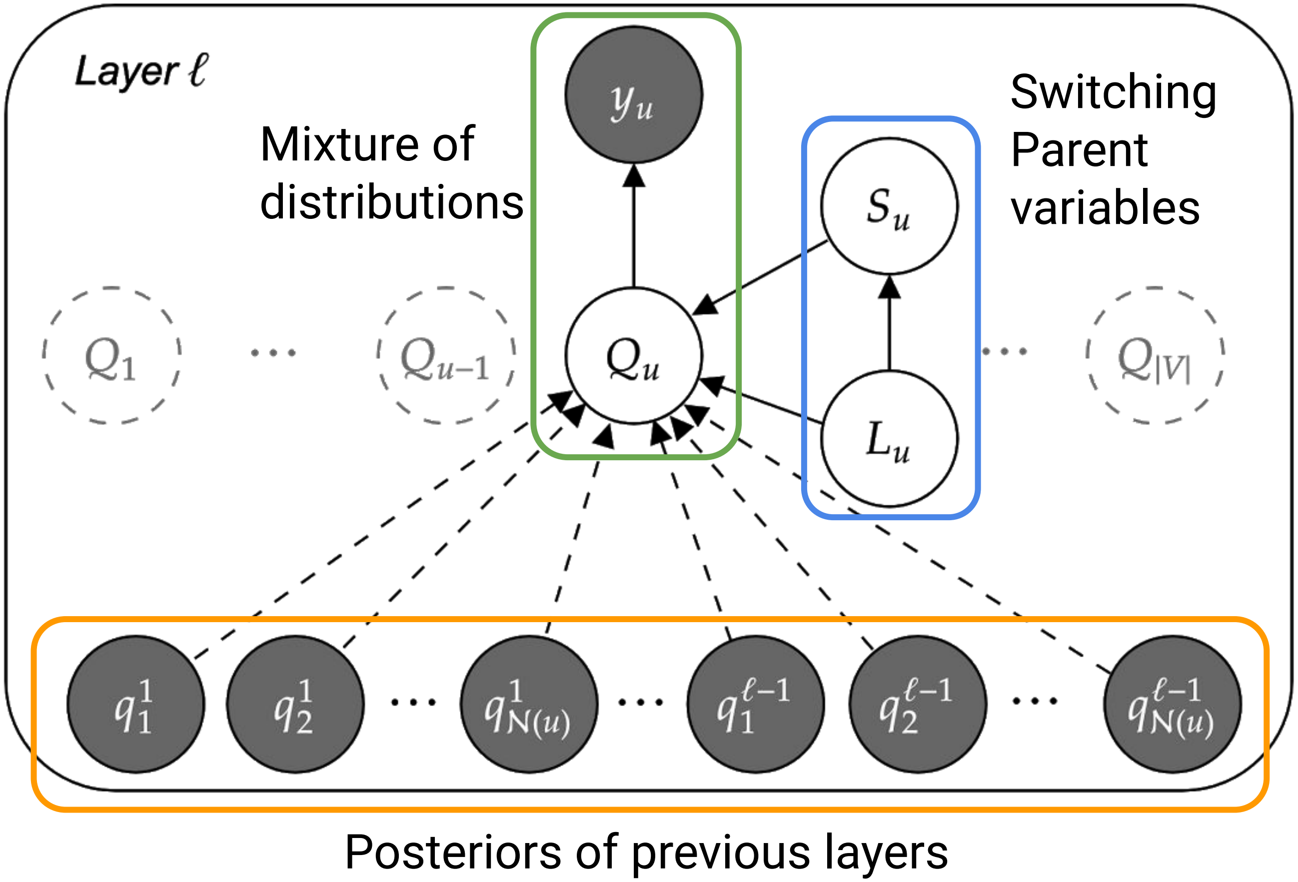
Examples of architectures[1]



My Research Topic

- A **probabilistic** framework for “Deep Graph Networks”
- From neural to probabilistic:
  - Neighborhood Aggregation
  - Handle edge features
  - Incremental construction
  - Sampling neighbors
  - Pooling (?)

The How[3]


Layer ℓ
Mixture of distributions
Switching Parent variables
y_u
S_u
Q_1
Q_{u-1}
Q_u
L_u
$Q_{|V|}$
q_1^1
q_2^1
$q_{N(u)}^1$
$q_1^{\ell-1}$
$q_2^{\ell-1}$
$q_{N(u)}^{\ell-1}$
Posteriors of previous layers

Ongoing Works

- Deal with **arbitrary** edge values
  - Model the edge distribution
  - Cluster edges into discrete types

- Unsupervised criterion to stop the construction
  - No “best” criterion
  - The idea is to reach convergence somehow
  - How to enforce it?

- *Future Works*
  - Sampling techniques to reduce time complexity
  - Neural aggregation functions (Generalized EM)

More References

1. *A Gentle Introduction to Deep Learning for Graphs* (under review)
   Bacciu D., Errica F., Micheli A., Podda M.
   https://arxiv.org/abs/1912.12693

2. *Probabilistic Learning on Graphs via Contextual Architectures* (under review)
   Bacciu D., Errica F., Micheli A.

3. Contextual Graph Markov Model: A Deep and Generative Approach to Graph Processing
   Bacciu D., Errica F., Micheli A., ICML 2018
   http://proceedings.mlr.press/v80/bacciu18a/

4. A Fair Comparison of Graph Neural Networks for Graph Classification
   Errica F., Podda M., Bacciu D., Micheli A., ICLR 2020
   https://openreview.net/pdf?id=HygDF6NFPB

5. *Semi-supervised classification with graph convolutional networks*
   Kipf T., Welling M., ICLR 2017

Thank you!

Q&A

Reading Workshop: your own research topic

- Introduce the group to your research topic
- Try to make it interesting and pleasant!
- Foster **collaborations**, exchange and promote **ideas**
- Have fun (possibly 😉)

Feel free to interrupt, start a discussion, and correct me

DL on graphs: motivations

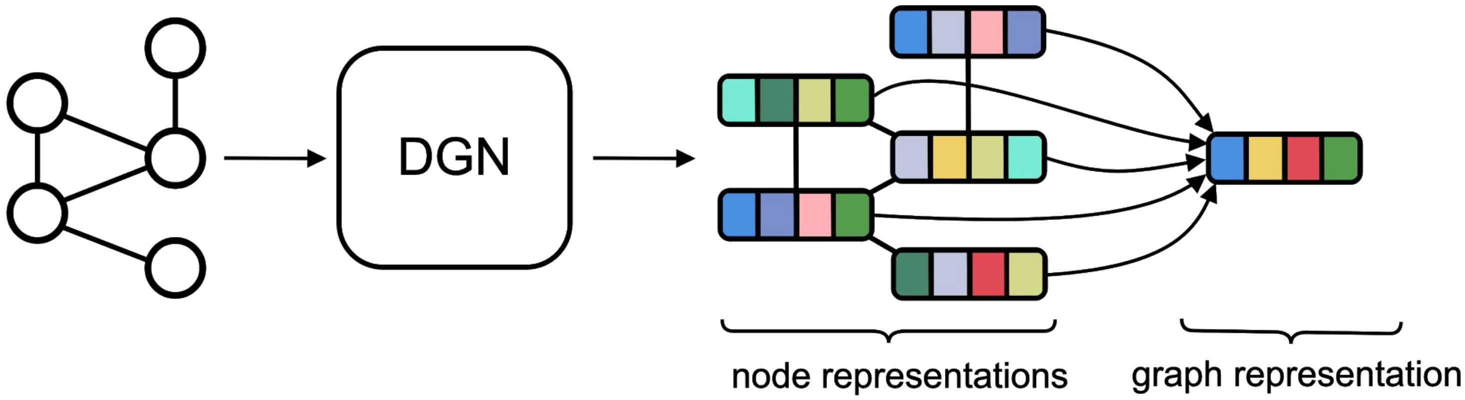
- Adaptive processing of general structures

- Unknown size & topology
  - → No causal dependencies that impose an ordering
  - Cycles complicate things

- Automatic feature extraction
  - vs CNNs for flat data

- We still like black boxes ;)

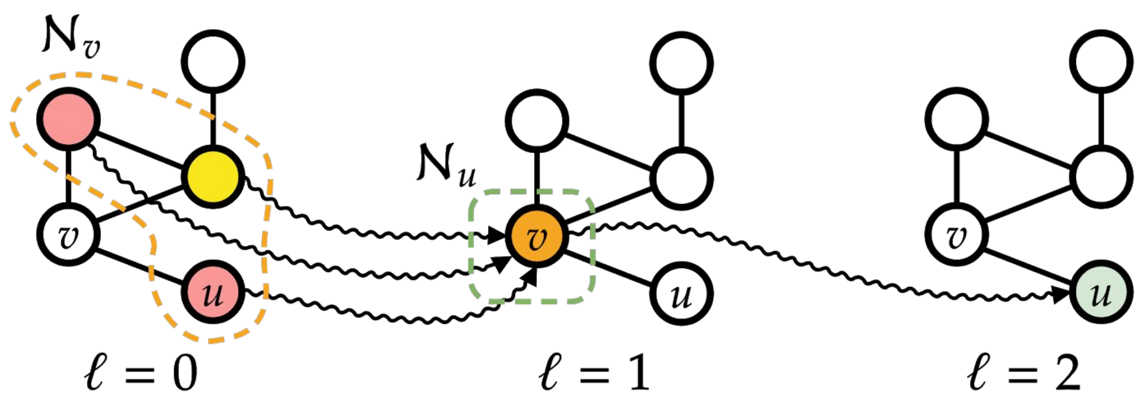
Representation learning on graphs



- Deep Graph Networks (DGNs)[1]:
  - Neural
  - Probabilistic
  - Generative

- “Graph Neural Networks” → ambiguous and restrictive

DL on graphs: context spreading



- **Local** and **iterative** processing of graphs

- Many mechanisms to spread context[1]:
  - Feedforward (e.g. Graph Convolutional Network[4])
  - Constructive (e.g. Neural Network for Graphs[5])
  - Recurrent (e.g. Graph Neural Network[6])

Contextual Graph Markov Model

- A **probabilistic** formalization of a DGN
- It truly is probablistic
  - All parameters are probability distributions (for now..)
- Deep?
  - A Stack of Bayesian Networks
  - **Incremental construction**
- Unsupervised
  - Maximum Likelihood Estimation via EM
  - Closed-form learning equations

Contextual Graph Markov Model (cont.)

- Learn via Expectation-Maximization

Generation

“Attention”

$$\mathcal{L}(\theta|\mathcal{G}) \approx \prod_{\substack{g\in\mathcal{G}\\ u\in\mathcal{V}_g}} \Big(\sum_{i=1}^{C} \boxed{P(y_u|Q_u = i)} \sum_{\ell'\in\mathbb{L}(\ell)} \boxed{P(L_u = \ell')}$$

$$\sum_{a=1}^{A} \boxed{P^{\ell'}(S_u = a)} P^{\ell',a}(Q_u = i|\mathbf{q}^{\ell',a}_{\mathcal{N}(u)}) \Big)$$

$$P^{\ell',a}(Q_u = i|\mathbf{q}^{\ell',a}_{\mathcal{N}(u)}) \approx \underbrace{\frac{1}{|\mathcal{N}^{\ell',a}(u)|} \sum_{j}^{C} P^{\ell',a}(Q = i|q = j) \sum_{v\in\mathcal{N}^{a}(u)} q^{\ell'}_{v}(j)}_{\textbf{Neighborhood Aggregation (mean)}}$$