Bayesian Deep Learning for Graphs

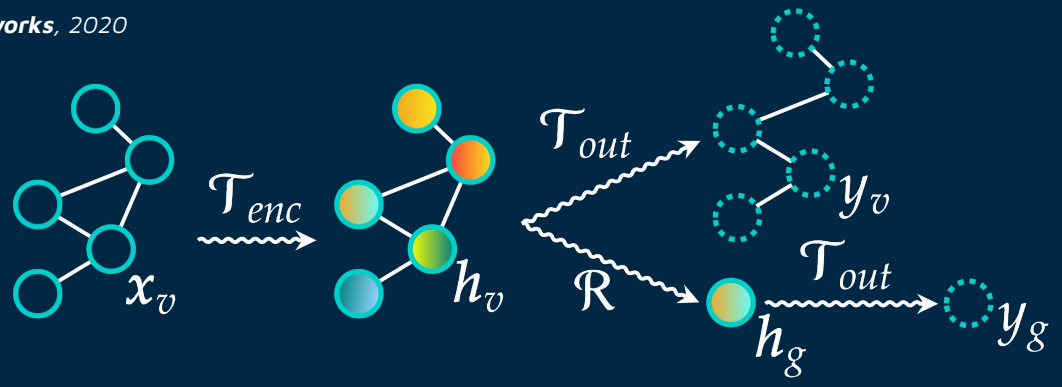
Candidate: Federico Errica
Supervisors: Davide Bacciu, Alessio Micheli

Why ML on Graphs?

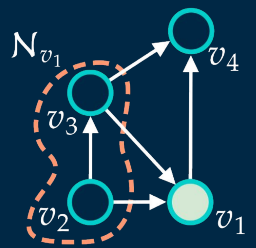
Table of Contents

The Bigger Picture

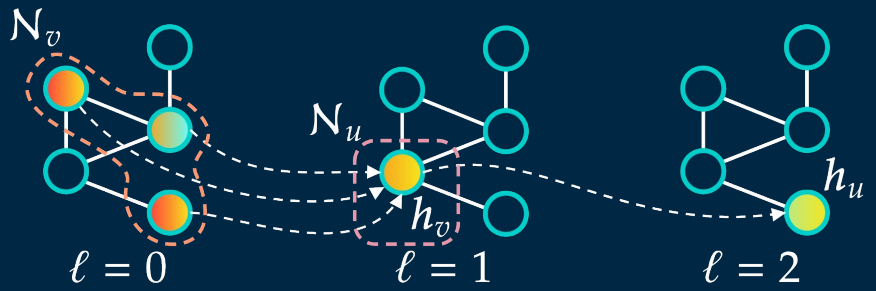
*Published in **Neural Networks**, 2020*



(1) local



(2) iterative



Graph Convolutional Layer

*Published in **Neural Networks**, 2020*

$$\underbrace{\boldsymbol{h}_v^{\ell+1}}_{v\text{'s state}} = \phi^{\ell+1}\Big(\boldsymbol{h}_v^{\ell},\ \underbrace{\Psi}_{\text{perm. invariant function}}(\overbrace{\{\psi^{\ell+1}(\boldsymbol{h}_u^{\ell}) \mid u \in \mathcal{N}_v\}}^{\text{set of } v\text{'s neighboring states}})\Big)$$

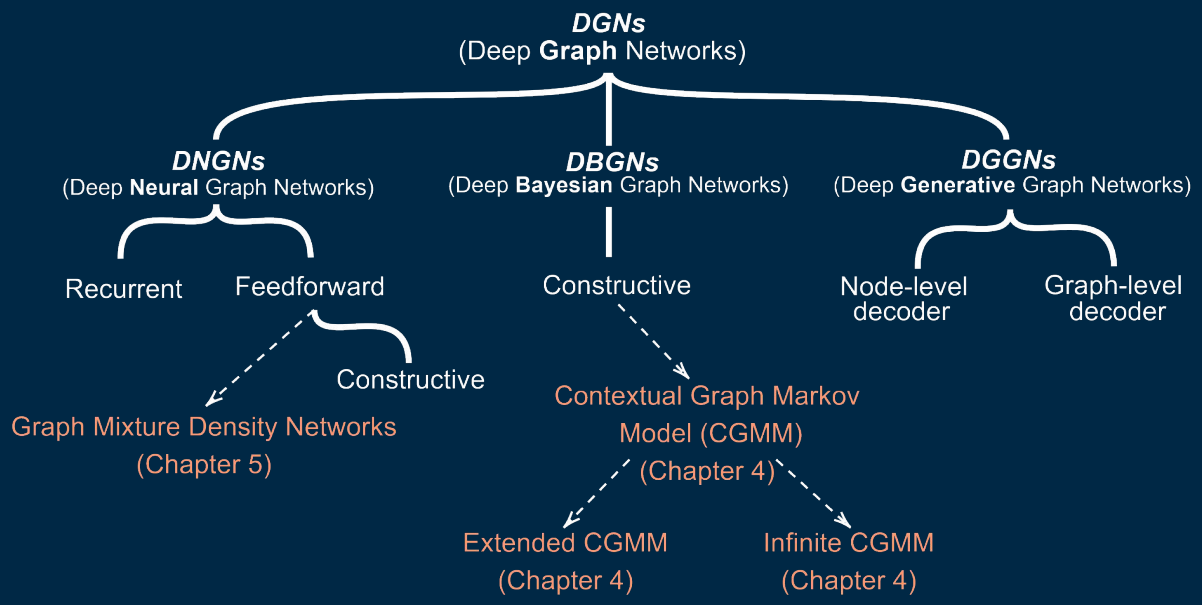
Variants/extensions:

- Edge-aware convolution
- Attention over neighbors
- Laplacian-normalized

Model	Neighborhood Aggregation
NN4G	$\sigma\Big(\mathbf{w}^{\ell+1^T}\mathbf{x}_v + \sum_{i=0}^{\ell}\sum_{c_k\in\mathcal{C}}\sum_{u\in\mathcal{N}_v^{c_k}} w_{c_k}^i * \mathbf{h}_u^i\Big)$
GNN	$\sum_{u\in\mathcal{N}_v} MLP^{\ell+1}\Big(\mathbf{x}_u, \mathbf{x}_v, \mathbf{a}_{uv}, \mathbf{h}_u^{\ell}\Big)$
GraphESN	$\sigma\Big(\mathbf{W}^{\ell+1}\mathbf{x}_u + \hat{\mathbf{W}}^{\ell+1}[\mathbf{h}_{u_1}^{\ell}, \ldots, \mathbf{h}_{u_{\mathcal{N}_v}}^{\ell}]\Big)$
GCN	$\sigma\Big(\mathbf{W}^{\ell+1}\sum_{u\in\mathcal{N}(v)} L_{vu}\mathbf{h}_u^{\ell}\Big)$
GAT	$\sigma\Big(\sum_{u\in\mathcal{N}_v} \alpha_{uv}^{\ell+1} * \mathbf{W}^{\ell+1}\mathbf{h}_u\Big)$
ECC	$\sigma\Big(\frac{1}{\lvert\mathcal{N}_v\rvert}\sum_{u\in\mathcal{N}_v} MLP^{\ell+1}(\mathbf{a}_{uv})^T\mathbf{h}_u^{\ell}\Big)$
R-GCN	$\sigma\Big(\sum_{c_k\in\mathcal{C}}\sum_{u\in\mathcal{N}_v^{c_k}} \frac{1}{\lvert\mathcal{N}_v^{c_k}\rvert}\mathbf{W}_{c_k}^{\ell+1}\mathbf{h}_u^{\ell} + \mathbf{W}^{\ell+1}\mathbf{h}_v^{\ell}\Big)$
GraphSAGE	$\sigma\Big(\mathbf{W}^{\ell+1}(\frac{1}{\lvert\mathcal{N}_v\rvert}[\mathbf{h}_v^{\ell}, \sum_{u\in\mathcal{N}_v}\mathbf{h}_u^{\ell}])\Big)$
CGMM	$\sum_{i=0}^{\ell} w^i * \Big(\sum_{c_k\in\mathcal{C}} w_{c_k}^i * (\frac{1}{\lvert\mathcal{N}_v^{c_k}\rvert}\sum_{u\in\mathcal{N}_v^{c_k}}\mathbf{h}_u^i)\Big)$
GIN	$MLP^{\ell+1}\Big((1+\epsilon^{\ell+1})\mathbf{h}_v^{\ell} + \sum_{u\in\mathcal{N}_v}\mathbf{h}_u^{\ell}\Big)$

A Taxonomy

*Published in **Neural Networks**, 2020*

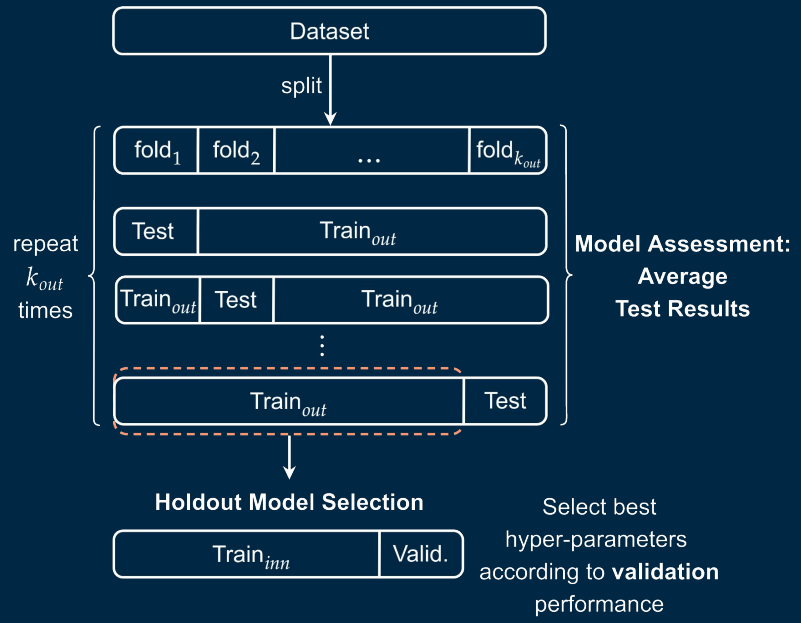


Analyzing Scholarship Issues

*Published in **ICLR**, 2020*

	DGCNN	DiffPool	ECC	GIN
Data preprocessing code	Y	Y	-	Y
Model selection code	N	N	-	N
Model evaluation code	Y	Y	-	Y
Data splits provided	Y	N	N	Y
Label Stratification	Y	N	-	Y
Report accuracy on test	Y	A	A	N
Report standard deviations	Y	N	N	Y

Y → yes
N → no
A → ambiguous
- → not available



A STANDARD & UNIFORM EVALUATION SCHEME

Re-evaluation Results

*Published in **ICLR**, 2020*

Results on chemical datasets

	D&D	NCI1	PROTEINS	ENZYMES
Baseline	**78.4** ± 4.5	69.8 ± 2.2	**75.8** ± 3.7	**65.2** ± 6.4
DGCNN	76.6 ± 4.3	76.4 ± 1.7	72.9 ± 3.5	38.9 ± 5.7
DiffPool	75.0 ± 3.5	76.9 ± 1.9	73.7 ± 3.5	59.5 ± 5.6
ECC	72.6 ± 4.1	76.2 ± 1.4	72.3 ± 3.4	29.5 ± 8.2
GIN	75.3 ± 2.9	**80.0** ± 1.4	73.3 ± 4.0	59.6 ± 4.5
GraphSAGE	72.9 ± 2.0	76.0 ± 1.8	73.0 ± 4.5	58.2 ± 6.0

Results on social datasets

		IMDB-B	IMDB-M	REDDIT-B	REDDIT-5K	COLLAB
No Features	Baseline	50.7 ± 2.4	36.1 ± 3.0	72.1 ± 7.8	35.1 ± 1.4	55.0 ± 1.9
	DGCNN	53.3 ± 5.0	38.6 ± 2.2	77.1 ± 2.9	35.7 ± 1.8	57.4 ± 1.9
	DiffPool	68.3 ± 6.1	45.1 ± 3.2	76.6 ± 2.4	34.6 ± 2.0	67.7 ± 1.9
	ECC	67.8 ± 4.8	44.8 ± 3.1	OOR	OOR	OOR
	GIN	66.8 ± 3.9	42.2 ± 4.6	**87.0** ± 4.4	**53.8** ± 5.9	**75.9** ± 1.9
	GraphSAGE	**69.9** ± 4.6	**47.2** ± 3.6	86.1 ± 2.0	49.9 ± 1.7	71.6 ± 1.5
With Degree	Baseline	70.8 ± 5.0	**49.1** ± 3.5	82.2 ± 3.0	52.2 ± 1.5	70.2 ± 1.5
	DGCNN	69.2 ± 3.0	45.6 ± 3.4	87.8 ± 2.5	49.2 ± 1.2	71.2 ± 1.9
	DiffPool	68.4 ± 3.3	45.6 ± 3.4	89.1 ± 1.6	53.8 ± 1.4	68.9 ± 2.0
	ECC	67.7 ± 2.8	43.5 ± 3.1	OOR	OOR	OOR
	GIN	**71.2** ± 3.9	48.5 ± 3.3	**89.9** ± 1.9	**56.1** ± 1.7	**75.6** ± 2.3
	GraphSAGE	68.8 ± 4.5	47.6 ± 3.5	84.3 ± 1.9	50.0 ± 1.3	73.9 ± 1.7

We report the median number of selected layers per model

	IMDB-B		IMDB-M		REDDIT-B		REDDIT-M		COLLAB	
	1	DEG	1	DEG	1	DEG	1	DEG	1	DEG
DGCNN	3	3	3.5	3	4	3	3	2	4	2
DiffPool	1	2	2	1	2	2	2	1	2	1.5
ECC	1	2	1	1	-	-	-	-	-	-
GIN	3	2	4	2	4	4	4	3	4	4
GraphSAGE	4	3	5	4	3	4	3	5	3	5

Metric: ACCURACY

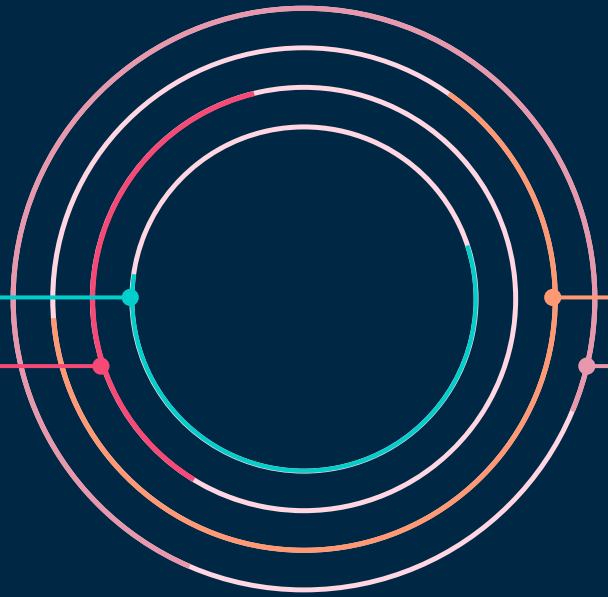
Deep Bayesian Graph Nets

Contextual Graph Markov Model

*Published in **ICML**, 2018 & **JMLR**, 2020*

Exploit vast amounts of raw unlabelled data

UNSUPERVISED EMBEDDINGS



Does not suffer from vanishing/exploding gradient

DEEP

PROBABILISTIC

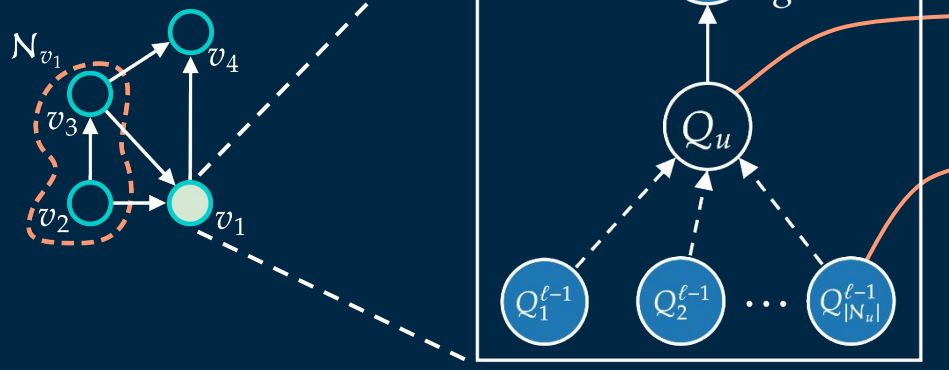
Relies on simple conditional Bayesian networks

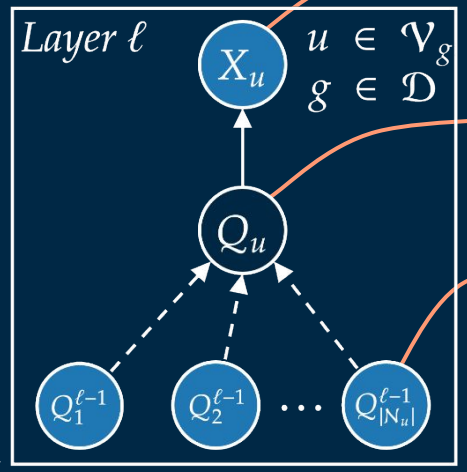
EFFICIENT

Linear (Space/Time) in the number of edges

Contextual Graph Markov Model

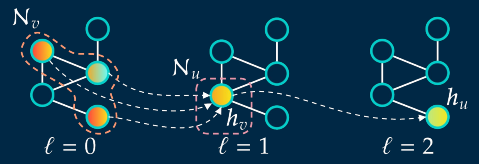
*Published in **ICML**, 2018 & **JMLR**, 2020*





NODE FEATURES

CATEGORICAL DISTRIBUTION (C STATES)

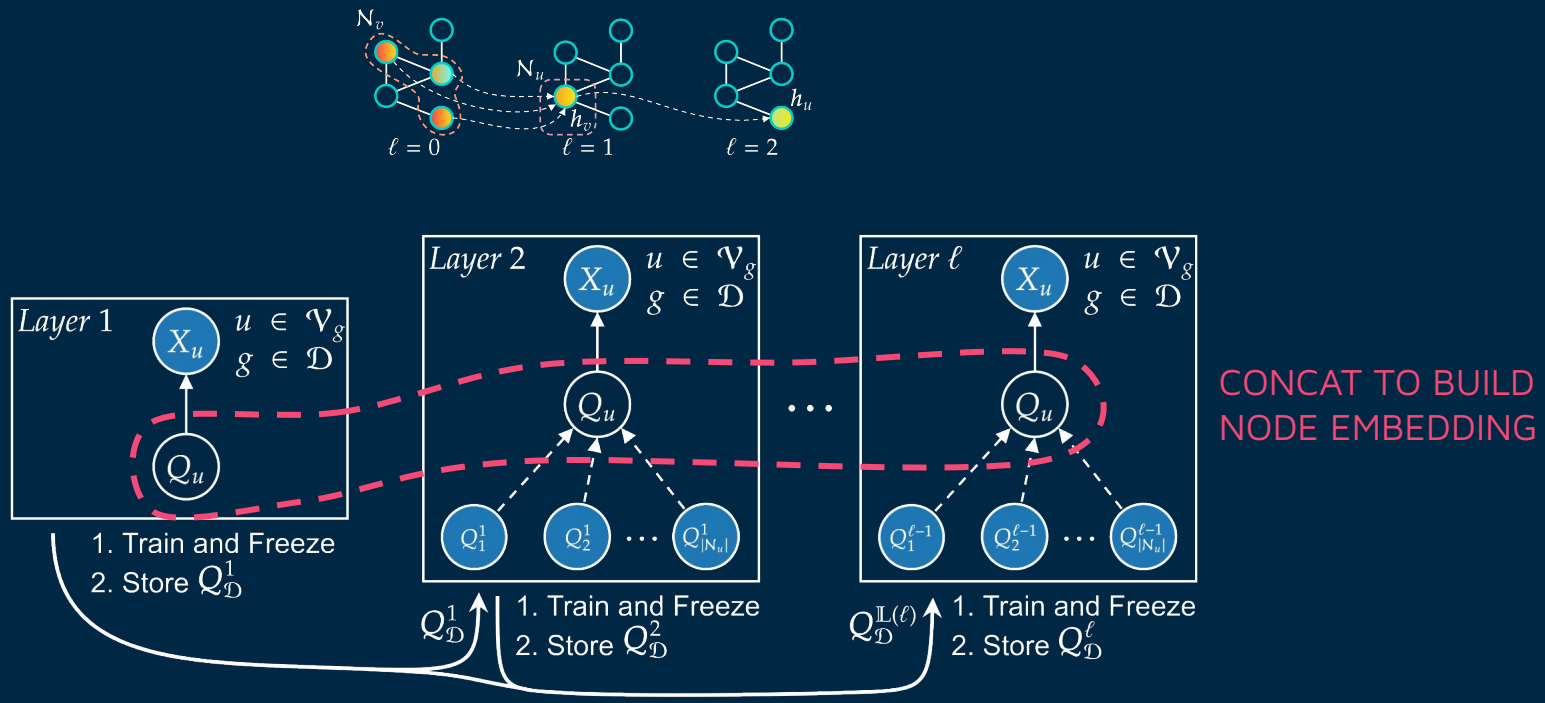


TRAINING ONE LAYER AT A TIME!

INTRACTABLE
P(Q|set of neighbors)

How-to CGMM

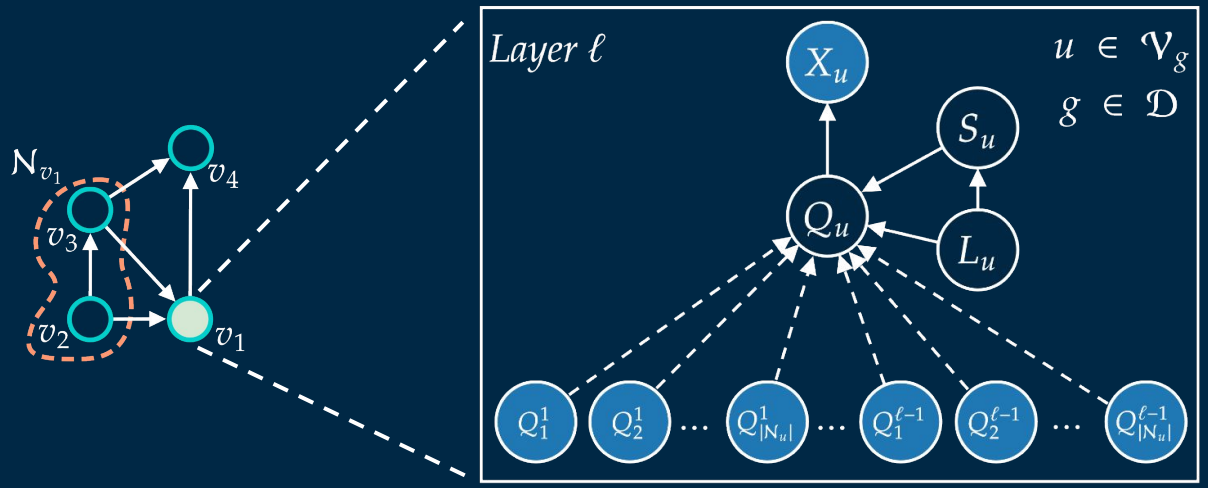
*Published in **ICML**, 2018 & **JMLR**, 2020*



High-level description of CGMM's incremental construction

Handling Discrete Edge Features

*Published in **ICML**, 2018 & **JMLR**, 2020*



INTRODUCE **SWITCHING PARENT** VARIABLES

GROUP FROZEN NEIGHBORS BY **EDGE TYPE** AND **LAYER**

ADDRESS **INTRACTABILITY** + ENHANCED **FLEXIBILITY**

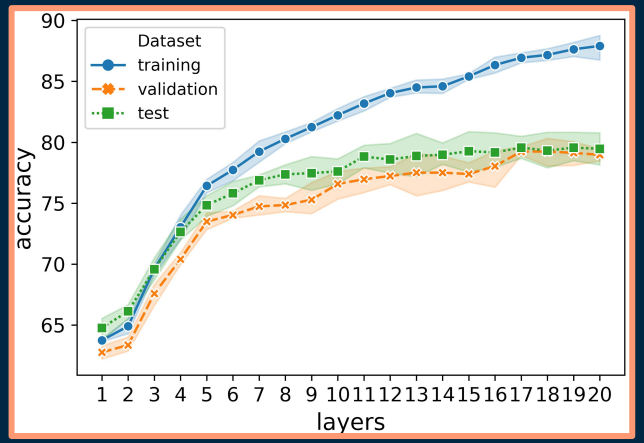
Graph Classification Results

*Published in **ICML**, 2018 & **JMLR**, 2020*

	D&D	NCI1	PROTEINS
BASELINE	**78.4** $\pm$ 4.5	69.8 $\pm$ 2.2	**75.8** $\pm$ 3.7
DGCNN	76.6 $\pm$ 4.3	76.4 $\pm$ 1.7	72.9 $\pm$ 3.5
DIFFPOOL	75.0 $\pm$ 3.5	76.9 $\pm$ 1.9	73.7 $\pm$ 3.5
ECC	72.6 $\pm$ 4.1	76.2 $\pm$ 1.4	72.3 $\pm$ 3.4
GIN	75.3 $\pm$ 2.9	**80.0** $\pm$ 1.4	73.3 $\pm$ 4.0
GRAPHSAGE	72.9 $\pm$ 2.0	76.0 $\pm$ 1.8	73.0 $\pm$ 4.5
CGMM	74.9 $\pm$ 3.4	76.2 $\pm$ 2.0	74.0 $\pm$ 3.9

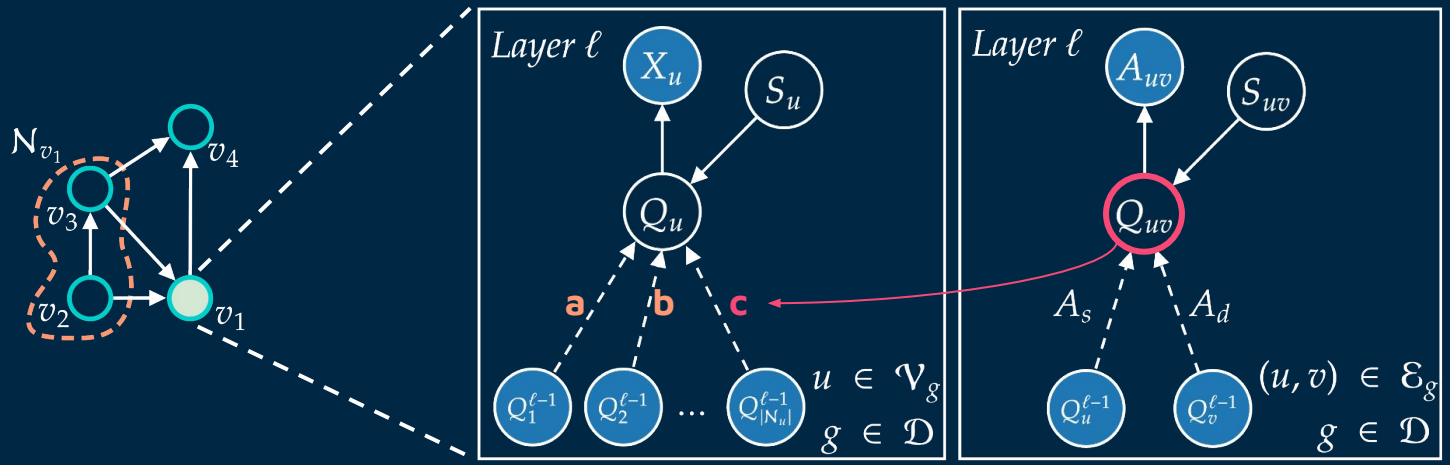
	IMDB-B	IMDB-M	REDDIT-B	REDDIT-5K	COLLAB
BASELINE	70.8 $\pm$ 5.0	**49.1** $\pm$ 3.5	82.2 $\pm$ 3.0	52.2 $\pm$ 1.5	70.2 $\pm$ 1.5
DGCNN	69.2 $\pm$ 3.0	45.6 $\pm$ 3.4	87.8 $\pm$ 2.5	49.2 $\pm$ 1.2	71.2 $\pm$ 1.9
DIFFPOOL	68.4 $\pm$ 3.3	45.6 $\pm$ 3.4	89.1 $\pm$ 1.6	53.8 $\pm$ 1.4	68.9 $\pm$ 2.0
ECC	67.7 $\pm$ 2.8	43.5 $\pm$ 3.1	-	-	-
GIN	71.2 $\pm$ 3.9	48.5 $\pm$ 3.3	**89.9** $\pm$ 1.9	**56.1** $\pm$ 1.7	75.6 $\pm$ 2.3
GRAPHSAGE	68.8 $\pm$ 4.5	47.6 $\pm$ 3.5	84.3 $\pm$ 1.9	50.0 $\pm$ 1.3	73.9 $\pm$ 1.7
CGMM	**72.7** $\pm$ 3.6	47.5 $\pm$ 3.9	88.1 $\pm$ 1.9	52.4 $\pm$ 2.2	**77.32** $\pm$ 2.2

NCI1



Extension – Continuous Edge Features

*Published in **IJCNN**, 2021*

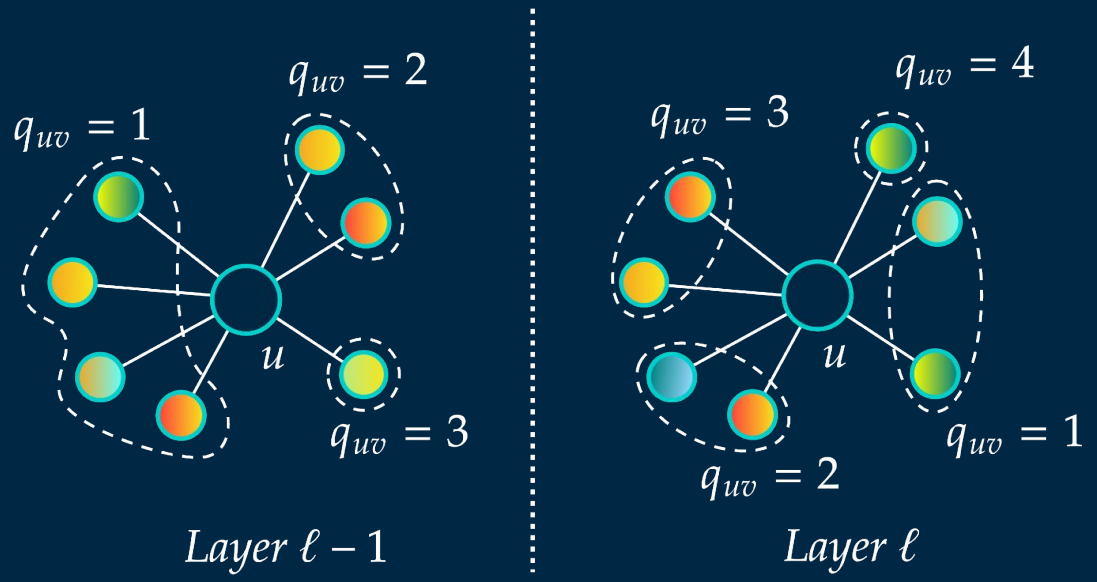


NODE COMPONENT
(ORIGINAL CGMM LAYER)

NEW EDGE COMPONENT

Dynamic Neighborhood Aggregation

*Published in **IJCNN**, 2021*



EVEN WHEN EDGE FEATURES ARE NOT AVAILABLE!

Graph Class./Regr. Results

*Published in **IJCNN**, 2021*

GRAPH CLASSIFICATION

	D&D	NCI1	PROTEINS
BASELINE	**78.4** $\pm$ 4.5	69.8 $\pm$ 2.2	**75.8** $\pm$ 3.7
DGCNN	76.6 $\pm$ 4.3	76.4 $\pm$ 1.7	72.9 $\pm$ 3.5
DIFFPOOL	75.0 $\pm$ 3.5	76.9 $\pm$ 1.9	73.7 $\pm$ 3.5
ECC	72.6 $\pm$ 4.1	76.2 $\pm$ 1.4	72.3 $\pm$ 3.4
GIN	75.3 $\pm$ 2.9	**80.0** $\pm$ 1.4	73.3 $\pm$ 4.0
GRAPHSAGE	72.9 $\pm$ 2.0	76.0 $\pm$ 1.8	73.0 $\pm$ 4.5
CGMM	74.9 $\pm$ 3.4	76.2 $\pm$ 2.0	74.0 $\pm$ 3.9
E-CGMM	73.9 $\pm$ 4.1	78.5 $\pm$ 1.7	73.3 $\pm$ 4.1

GRAPH REGRESSION on QM7b

	MAE	Relative Improvement
CGMM-no edge attributes	1.52 $\pm$ 0.05	19%
CGMM-discretized edges	1.49 $\pm$ 0.07	17%
E-CGMM	**1.23** $\pm$ 0.06	-

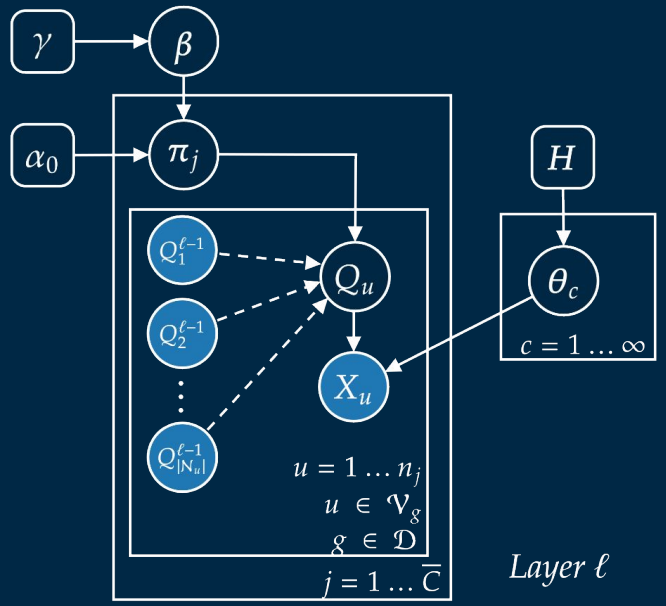
	IMDB-B	IMDB-M	REDDIT-B	REDDIT-5K	COLLAB
BASELINE	70.8 $\pm$ 5.0	**49.1** $\pm$ 3.5	82.2 $\pm$ 3.0	52.2 $\pm$ 1.5	70.2 $\pm$ 1.5
DGCNN	69.2 $\pm$ 3.0	45.6 $\pm$ 3.4	87.8 $\pm$ 2.5	49.2 $\pm$ 1.2	71.2 $\pm$ 1.9
DIFFPOOL	68.4 $\pm$ 3.3	45.6 $\pm$ 3.4	89.1 $\pm$ 1.6	53.8 $\pm$ 1.4	68.9 $\pm$ 2.0
ECC	67.7 $\pm$ 2.8	43.5 $\pm$ 3.1	-	-	-
GIN	71.2 $\pm$ 3.9	48.5 $\pm$ 3.3	**89.9** $\pm$ 1.9	**56.1** $\pm$ 1.7	75.6 $\pm$ 2.3
GRAPHSAGE	68.8 $\pm$ 4.5	47.6 $\pm$ 3.5	84.3 $\pm$ 1.9	50.0 $\pm$ 1.3	73.9 $\pm$ 1.7
CGMM	**72.7** $\pm$ 3.6	47.5 $\pm$ 3.9	88.1 $\pm$ 1.9	52.4 $\pm$ 2.2	77.32 $\pm$ 2.2
E-CGMM	70.7 $\pm$ 3.8	48.3 $\pm$ 4.1	89.5 $\pm$ 1.3	53.7 $\pm$ 1.0	**77.45** $\pm$ 2.3

“Can we extend CGMM to automatize the choice of its hyper-parameters?”

The Infinite CGMM

Under Review

- A Hierarchical Dirichlet Process
  - Automatic choice of # states for Q_u
  - Gibbs sampling inference
  - Can estimate all hyper-parameters with suitable hyper-priors
- Choice of observations' groups determined by neighbors' states
- Batch version for larger datasets



Graph Classification Results

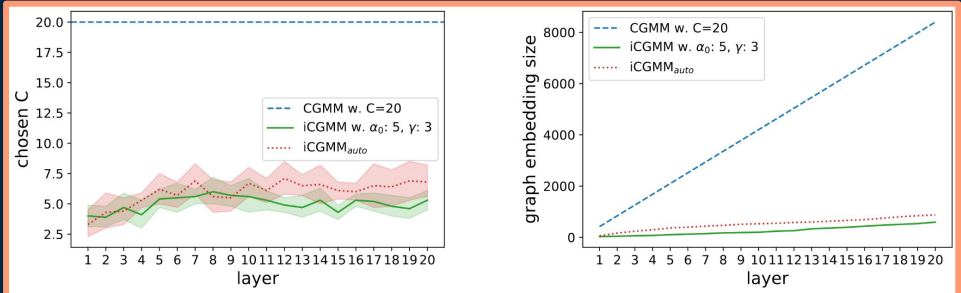
Under Review

GRAPH CLASSIFICATION

	D&D	NCI1	PROTEINS
BASELINE	**78.4** $\pm$ 4.5	69.8 $\pm$ 2.2	**75.8** $\pm$ 3.7
DGCNN	76.6 $\pm$ 4.3	76.4 $\pm$ 1.7	72.9 $\pm$ 3.5
DIFFPOOL	75.0 $\pm$ 3.5	76.9 $\pm$ 1.9	73.7 $\pm$ 3.5
ECC	72.6 $\pm$ 4.1	76.2 $\pm$ 1.4	72.3 $\pm$ 3.4
GIN	75.3 $\pm$ 2.9	**80.0** $\pm$ 1.4	73.3 $\pm$ 4.0
GRAPHSAGE	72.9 $\pm$ 2.0	76.0 $\pm$ 1.8	73.0 $\pm$ 4.5
CGMM	74.9 $\pm$ 3.4	76.2 $\pm$ 2.0	74.0 $\pm$ 3.9
E-CGMM	73.9 $\pm$ 4.1	78.5 $\pm$ 1.7	73.3 $\pm$ 4.1
ICGMM	75.6 $\pm$ 4.3	76.5 $\pm$ 1.8	72.7 $\pm$ 3.4
ICGMM$_f$	75.0 $\pm$ 5.6	76.7 $\pm$ 1.7	73.3 $\pm$ 2.9
ICGMM$_{auto}$	76.3 $\pm$ 5.6	77.6 $\pm$ 1.5	73.1 $\pm$ 3.9
ICGMM$_{fauto}$	75.1 $\pm$ 3.8	76.4 $\pm$ 1.4	73.2 $\pm$ 3.9

	IMDB-B	IMDB-M	REDDIT-B	REDDIT-5K	COLLAB
BASELINE	70.8 $\pm$ 5.0	**49.1** $\pm$ 3.5	82.2 $\pm$ 3.0	52.2 $\pm$ 1.5	70.2 $\pm$ 1.5
DGCNN	69.2 $\pm$ 3.0	45.6 $\pm$ 3.4	87.8 $\pm$ 2.5	49.2 $\pm$ 1.2	71.2 $\pm$ 1.9
DIFFPOOL	68.4 $\pm$ 3.3	45.6 $\pm$ 3.4	89.1 $\pm$ 1.6	53.8 $\pm$ 1.4	68.9 $\pm$ 2.0
ECC	67.7 $\pm$ 2.8	43.5 $\pm$ 3.1	-	-	-
GIN	71.2 $\pm$ 3.9	48.5 $\pm$ 3.3	89.9 $\pm$ 1.9	**56.1** $\pm$ 1.7	75.6 $\pm$ 2.3
GRAPHSAGE	68.8 $\pm$ 4.5	47.6 $\pm$ 3.5	84.3 $\pm$ 1.9	50.0 $\pm$ 1.3	73.9 $\pm$ 1.7
CGMM	72.7 $\pm$ 3.6	47.5 $\pm$ 3.9	88.1 $\pm$ 1.9	52.4 $\pm$ 2.2	77.32 $\pm$ 2.2
E-CGMM	70.7 $\pm$ 3.8	48.3 $\pm$ 4.1	89.5 $\pm$ 1.3	53.7 $\pm$ 1.0	77.45 $\pm$ 2.3
ICGMM$_f$	**73.0** $\pm$ 4.3	48.6 $\pm$ 3.4	91.3 $\pm$ 1.8	55.5 $\pm$ 1.9	78.6 $\pm$ 2.8
ICGMM$_{fauto}$	71.8 $\pm$ 4.4	49.0 $\pm$ 3.8	**91.6** $\pm$ 2.1	55.6 $\pm$ 1.7	**78.9** $\pm$ 1.7

- =/> performances
- Less waste of states



(a) # states chosen at each layer (b) Cumulative graph embedding size on NCI1

Graph Mixture Density Nets

Why GMDNs ?

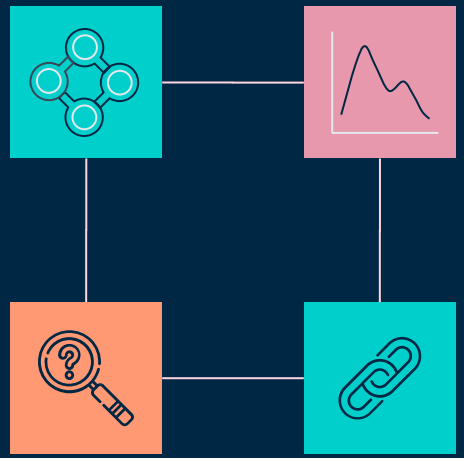
*Published in **ICML**, 2021*

Structure

Like Deep Graph Networks

Multimodality

Like Mixture Density Networks



Uncertainty

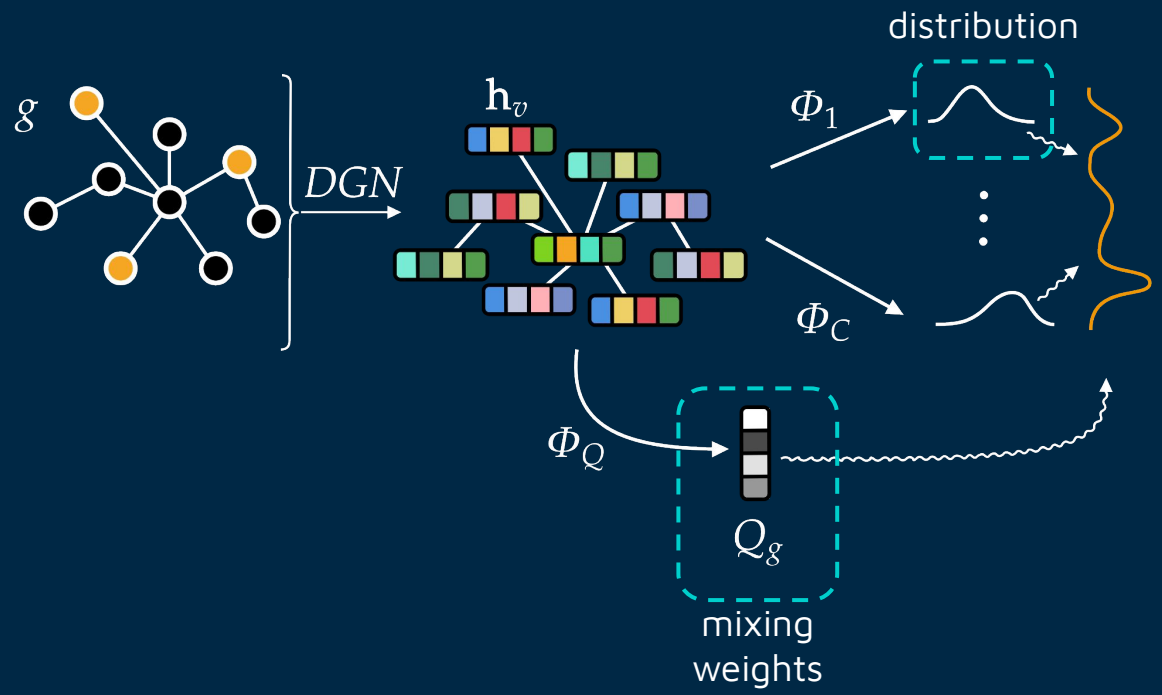
1) Data Representation
2) Encoder Expressiveness

End-to-end

Get the best of neural and probabilistic worlds

GMDN in a Nutshell

*Published in **ICML**, 2021*

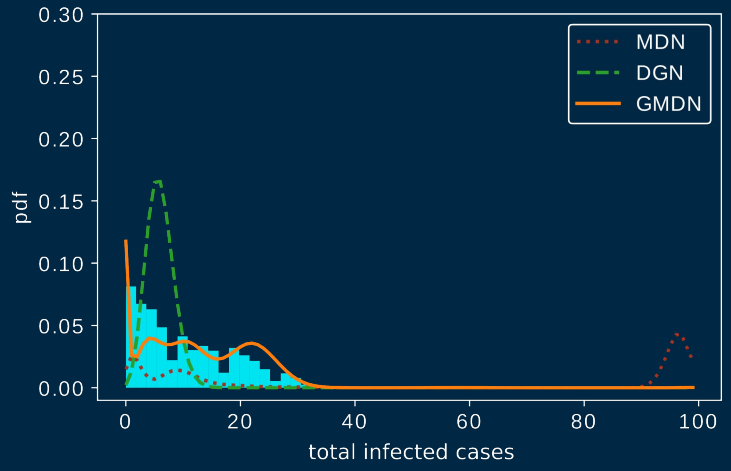


Results

*Published in **ICML**, 2021*

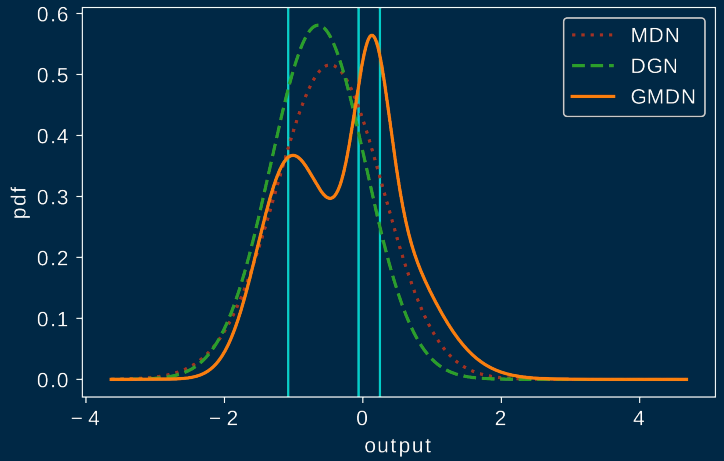
Synthetic Epidemic Simulations

	BA-100	ER-100	Structure	Multimodal
RAND	-4.60	-4.60	✗	✗
HIST	-1.16	-2.32	✗	✓
MDN	-1.17(.05)	-2.54(.07)	✗	✓
DGN	-0.90(.35)	-1.96(.16)	✓	✗
GMDN	**-0.67**(.02)	**-1.56**(.04)	✓	✓



Chemical Tasks

	alchemy_full		ZINC_full	
	$\log \mathcal{L}$	MAE	$\log \mathcal{L}$	MAE
RAND	-27.12	-	-4.20	-
HIST	-21.91	-	-1.28	-
MDN	-1.36(.90)	0.62(.01)	-1.14(.01)	0.67(.00)
DGN	-7.19(1.3)	0.62(.01)	-0.90(.10)	0.49(.03)
GMDN	**-0.57**(1.4)	**0.61**(.02)	**-0.75**(.10)	**0.49**(.04)

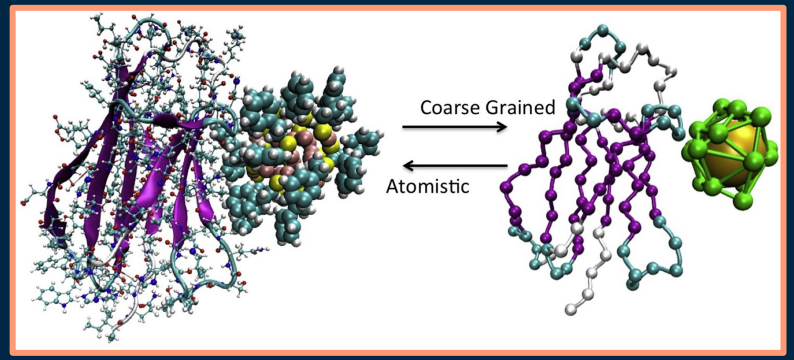


Applications

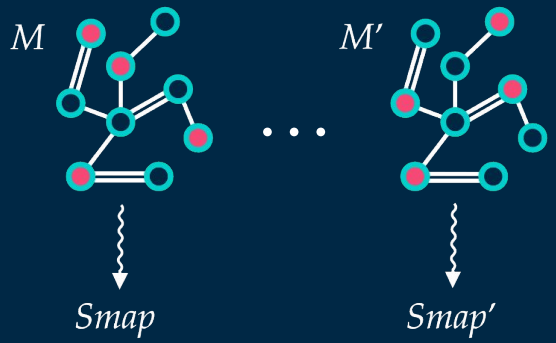
Accelerate Molecular Dynamics

*Published in **Frontiers in Molecular Biosciences**, 2021 (with Uni Trento)*

Coarse Graining



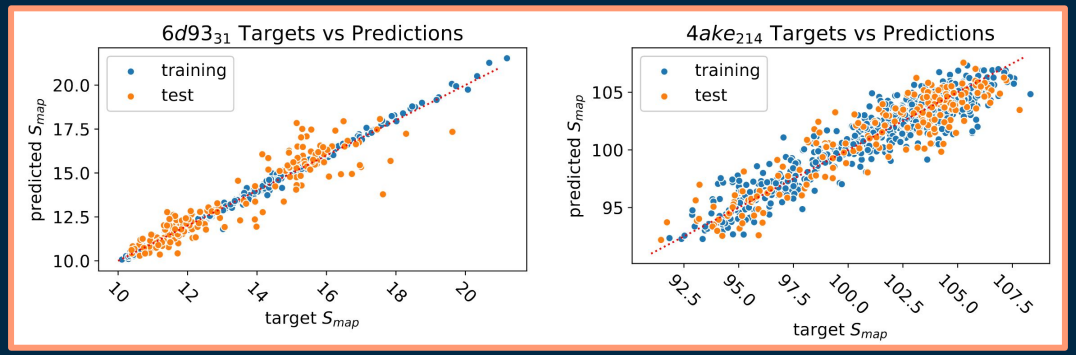
Simulations



find the optimal mapping according to known criterion → huge cost!

Results

*Published in **Frontiers in Molecular Biosciences**, 2021 (with Uni Trento)*



Model / Protein	TR MAE	TR R^2	VL MAE	VL R^2	TE MAE	TE R^2
Baseline / *6d93*	0.55	0.86	0.63	0.83	0.65	0.82
DGN / *6d93*	0.13	0.99	0.33	0.95	0.33	**0.96**
Baseline / *4ake*	1.78	0.70	1.75	0.65	1.86	0.69
DGN / *4ake*	0.91	0.92	1.2	0.85	1.35	**0.84**

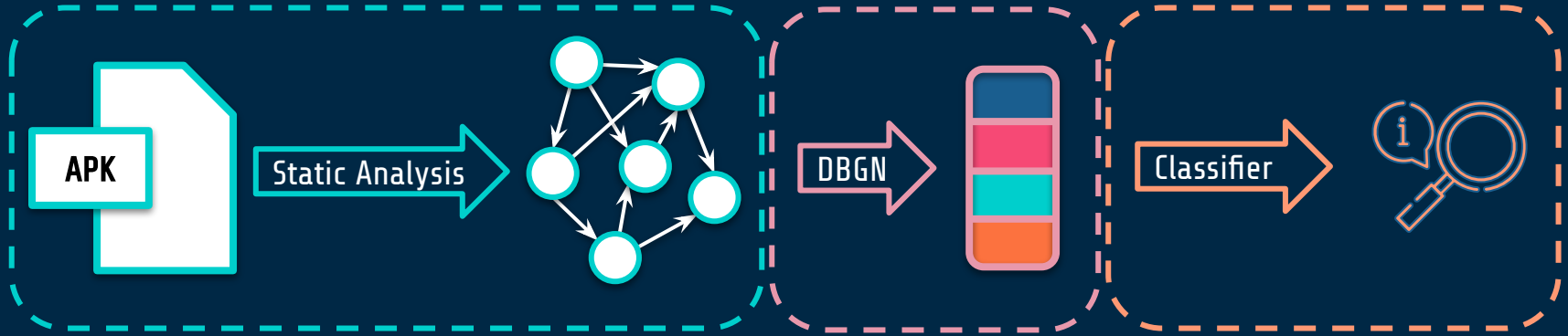
Accurate

Protein	Single measure	Inference GPU (CPU)	Improvement GPU (CPU)
6d93	$\simeq$ 2.1 mins	$\simeq$ 0.9(98.7) ms	$\simeq$ 140000 $\times$ (1276$\times$)
4ake	$\simeq$ 8.0 mins	$\simeq$ 4.8(1103.2) ms	$\simeq$ 100000 $\times$ (435$\times$)

Fast Prediction

Robust Malware Classification

*Published in **ESANN**, 2021 (with IIT CNR)*



1. Create a **Call Graph** dataset
2. **Do not consider** node features, use node out-degrees
3. Build unsupervised graph embeddings
4. **Robust** structural malware classification

	TR Loss	TR Acc.	VL Loss	VL Acc.	TE Loss	TE Acc.
Baseline	$1.2_{\pm 0.05}$	$55.6_{\pm 0.5}$	$1.1_{\pm 0.01}$	$60.6_{\pm 0.9}$	$1.1_{\pm 0.03}$	$56.7_{\pm 0.5}$
CGMM	$0.01_{\pm 0.01}$	$99.8_{\pm 0.4}$	$0.16_{\pm 0.01}$	$97.9_{\pm 0.2}$	$0.13_{\pm 0.01}$	$96.4_{\pm 0.6}$
E-CGMM	$0.03_{\pm 0.02}$	$99.4_{\pm 0.6}$	$0.59_{\pm 0.003}$	$98.4_{\pm 0.3}$	$0.19_{\pm 0.02}$	$\mathbf{97.3}_{\pm 0.4}$
iCGMM$_f$	$0.05_{\pm 0.01}$	$98.7_{\pm 0.5}$	$0.27_{\pm 0.04}$	$94.8_{\pm 0.5}$	$0.35_{\pm 0.03}$	$93.6_{\pm 0.6}$
iCGMM$_{f_{auto}}$	$0.07_{\pm 0.03}$	$97.93_{\pm 0.9}$	$0.25_{\pm 0.02}$	$95,8_{\pm 0.5}$	$0.42_{\pm 0.1}$	$92.7_{\pm 0.5}$

Conclusions

- **Introductory** contributions
  - Mathematical systematization of DL4G literature
  - Addressing empirical reproducibility issues
- **Methodological** contributions
  - Basic architecture (CGMM)
  - Extensions (E-CGMM, iCGMM)
  - Modeling multimodality with graphs (GMDN)
- **Practical** contributions
  - Molecular dynamics
  - Malware classification

Future Directions

- **Uncertainty** Modeling
  - Epistemic / Aleatoric via sampling from learned distributions
  - Useful in physics engine simulators
- Handling **missing features**
- **Temporal** graphs
- **Interpretability** analyses
  - Variations in likelihood / multimodal distribution w.r.t. changes in the graph

THANK YOU!

Davide Bacciu

Alessio Micheli

Daniele Castellana

Marco Podda

Internal Committee: Roberto Grossi (UniPi), Luca Oneto (UniGe)
International Reviewers: Mark Coates (McGill), Shirui Pan (Monash)

List of Publications

- ★ Bacciu D., **Errica F.**, Micheli A., *"Contextual Graph Markov Model: A Deep and Generative Approach to Graph Processing"*, ***ICML*** *2018*
- ★ **Errica F.**, Podda M., Bacciu D., Micheli A., *"A Fair Comparison of Graph Neural Networks for Graph Classification"*, ***ICLR*** *2020*
- ★ Bacciu D., **Errica F.**, Micheli A., Podda M., *"A Gentle Introduction to Deep Learning for Graphs"*, ***Neural Networks*** *2020*
- ★ Bacciu D., **Errica F.**, Micheli A., *"Probabilistic Learning on Graphs via Contextual Architecture"*, ***JMLR****, 2020*
- ★ **Errica F.**, Bacciu D., Micheli A., *"Theoretically Expressive and Edge-aware Graph Learning"*, ***ESANN*** *2020*
- ★ **Errica F.**, Giulini M., Bacciu D., Menichetti R., Micheli A., Potestio R., *"A deep graph network-enhanced sampling approach to efficiently explore the space of reduced representations of proteins"*, ***Frontiers in Molecular Biosciences*** *2021*
- ★ Atzeni D., **Errica F.**, Bacciu D., Micheli A., *"Modeling Edge Features with Deep Bayesian Graph Networks"*, ***IJCNN*** *2021*
- ★ **Errica F.**, Edizel B., Denoyer L., Petroni F., Plachouras V., Silvestri F., Riedel S., *"Context Matching for Low-Resource Classification"*, ***IJCNN*** *2021*
- ★ **Errica F.**, Bacciu D., Micheli A., *"Graph Mixture Density Networks"*, ***ICML*** *2021*
- ★ **Errica F.**, Iadarola G., Martinelli F., Mercaldo F., Micheli A., *"Robust Malware Classification via Deep Graph Networks on Call Graph Topologies"*, ***ESANN*** *2021*
- ★ Carta A., Cossu A., **Errica F.**, Bacciu D.,*"Catastrophic Forgetting in Deep Graph Networks: an Introductory Benchmark for Graph Classification", Workshop on Graph Learning Benchmarks,* ***WWW*** *2021* - ***spotlight***

Questions?

THANKS

federico.errica@neclab.eu
diningphil.github.io

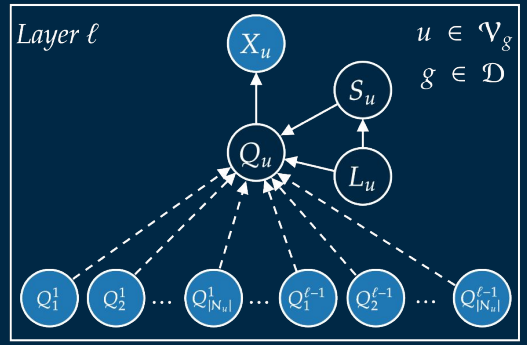
CGMM Formulation

To summarize, the likelihood of a graph under the extended formulation of CGMM can be therefore written as

$$\mathcal{L}(\boldsymbol{\theta} \mid g) = \prod_{u \in \mathcal{V}_g} \sum_{i=1}^{C} P_{\boldsymbol{\theta}}(X_u = x_u | Q_u = i) P(Q_u = i \mid \boldsymbol{Q}_{\mathcal{N}_u}^{\mathbb{L}(\ell)}) \approx$$

$$\approx \prod_{u \in \mathcal{V}_g} \sum_{i=1}^{C} \underbrace{P_{\boldsymbol{\theta}}(X_u = x_u | Q_u = i)}_{emission} \sum_{\ell' \in \mathbb{L}(\ell)} \underbrace{P_{\boldsymbol{\theta}}(L_u = \ell')}_{SP\ layer} \sum_{a=1}^{|\mathcal{A}_g|} \underbrace{P_{\boldsymbol{\theta}}^{\ell'}(S_u = a)}_{SP\ edge} \times$$

$$\times \frac{1}{|\mathcal{N}_u^a|} \sum_{j}^{C} \underbrace{P_{\boldsymbol{\theta}}^{\ell',a}(Q_u = i \mid Q_*^{\ell',a} = j)}_{SP\text{-}aware\ transition} \sum_{v \in \mathcal{N}_u^a} q_v^{\ell'}(j). \tag{4.4}$$



Complexity: Linear in # of edges

iCGMM Formulation

Overall, the generative process of a single iCGMM layer can be formalized as follows:

$$
\begin{aligned}
\boldsymbol{\beta} \mid \gamma &\sim \mathrm{Stick}(\gamma) & j_u \mid \boldsymbol{Q}^{\ell-1}_{\mathcal{N}_u} &= \psi(\boldsymbol{Q}^{\ell-1}_{\mathcal{N}_u}) \\
\boldsymbol{\pi}_j \mid \boldsymbol{\beta}, \alpha_0 &\sim \mathrm{DP}(\alpha_0, \boldsymbol{\beta}) & q_u \mid j_u, (\boldsymbol{\pi}_j)_{j=1}^{\bar{C}} &\sim \boldsymbol{\pi}_{j_u} \\
\boldsymbol{\theta} \mid \boldsymbol{H} &\sim \boldsymbol{H} & x_u \mid q_u, (\boldsymbol{\theta})_{c=1}^{\infty} &\sim F(\boldsymbol{\theta}_{q_u}),
\end{aligned}
\tag{4.7}
$$