Bringing Your Own View: Graph Contrastive Learning without Prefabricated Data Augmentations

Yuning You[1], Tianlong Chen[2], Zhangyang Wang[2], Yang Shen[1]
[1]Texas A&M University, [2]University of Texas at Austin

The 15th ACM International Conference on Web Search and Data Mining
wsdm Phoenix, AZ, USA 2022
February 21–25, 2022

Motivations

- Unlike images, graph data are inherently **heterogenous** (e.g. pandemics, product co-purchase relation, molecules).
- The SOTA GraphCL [1] handles the heterogenousity with ad-hoc choices of augmentations (**priors**) for every datasets.
- Rather than the **discrete** selection on prefabricated ones, can we directly learn and generate such priors, **continuously**?
- ***Question:*** What us the **space**, **principle** and **framework** for learnable GraphCL priors (GraphCL-LP)?

Method. Graph Generative Models as Learnable Priors

- We define the **space** of GraphCL priors as the set of stochastic mappings between graph manifolds $m: \mathcal{G} \rightarrow \mathcal{G}$.
- Naturally, we adopt the recent rising graph generative models (VGAE [2] here) for the prior space **parametrization**.
- Further, **principled** reward signals are introduced to convey messages from contrastive learning to generator training.
- Above components are assembled into the bi-level optimization **framework**.

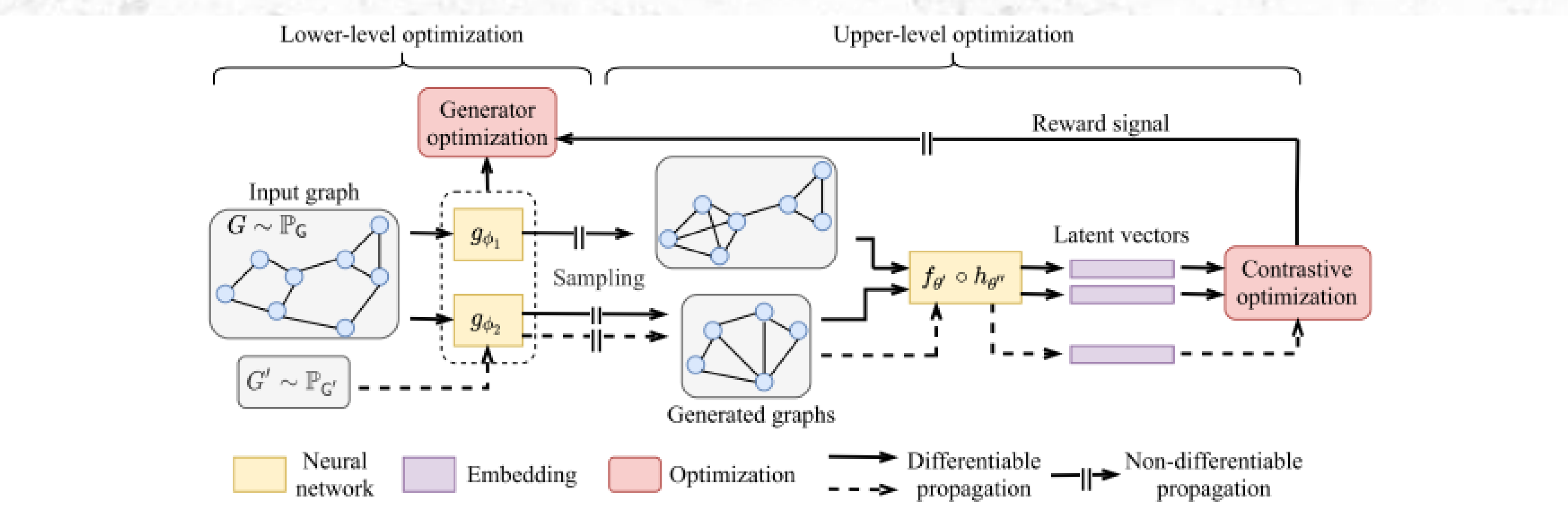

Figure 2: Pipeline of GraphCL with learned prior. Graph generative models g_{ϕ_1}, g_{ϕ_2} generate contrastive views for self-supervised contrasting, and then receive the reward for their parameter update.

Method. Principles for Learning Priors to Contrast

- Information minimization (**InfoMin**). Encouraging contrastive views to share less mutual information.
- Information bottleneck (**InfoBN**). Diminishing the information overlap between each contrastive view and its latent representation.

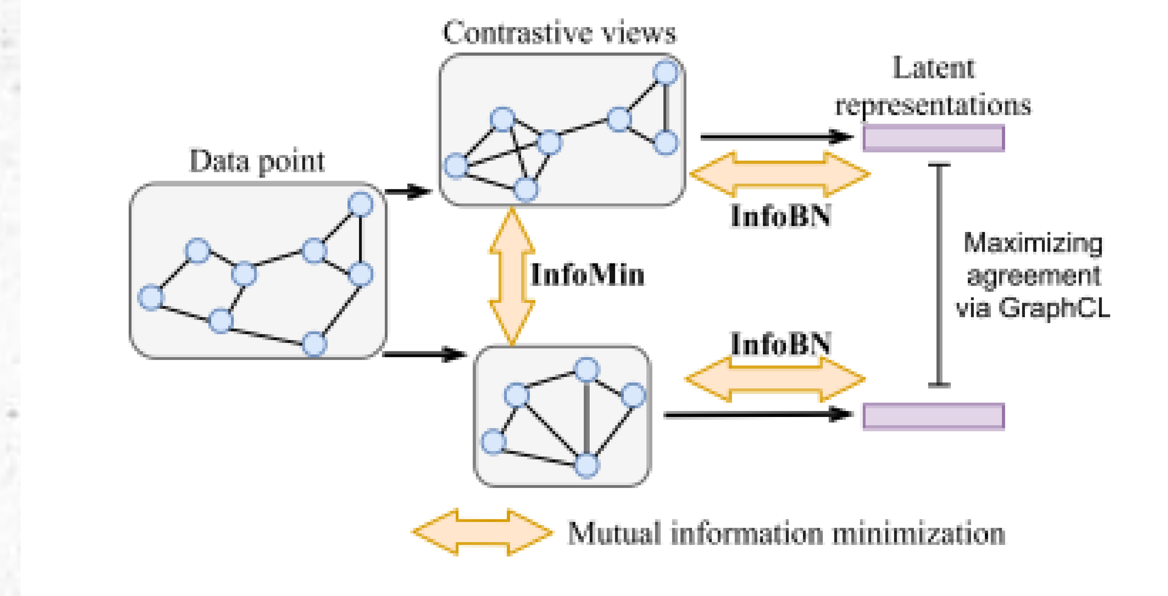

Figure 3: Schematic diagram of the InfoMin and InfoBN principles to guide the prior learning in GraphCL.

Experiments

- We numerically show the competitive performance of GraphCL-LP versus SOTA methods.

Table 2: Semi-supervised learning on small-scale benchmarks from TUDataset (the first four) and large-scale ones from OGB (the last two). Shown in red are the best three accuracies (%) for TUDataset and the best for ogbg-ppa and F1-score (%) for ogbg-code. The SOTA results compared here are as published under the same experimental setting (- indicates that results were not available in corresponding publications).

Methods	COLLAB	RDT-B	RDT-M5K	GITHUB	ogbg-ppa	ogbg-code
No pre-train	73.71±0.27	86.63±0.27	51.33±0.44	60.87±0.17	56.01±1.05	17.85±0.60
Augmentations	74.19±0.13	87.74±0.39	52.01±0.20	60.91±0.32	-	-
GAE	**75.09**±0.19	87.69±0.40	**53.58**±0.13	**63.89**±0.52	-	-
Infomax	73.76±0.29	**88.66**±0.95	**53.61**±0.31	**65.21**±0.88	-	-
ContextPred	73.69±0.37	84.76±0.52	51.23±0.84	62.35±0.73	-	-
GraphCL	74.23±0.21	**89.11**±0.19	52.55±0.45	**65.81**±0.79	57.77±1.25	22.45±0.17
LP-InfoMin	**74.66**±0.14	**88.03**±0.46	53.00±0.26	62.71±0.54	**59.10**±0.88	23.50±0.22
LP-InfoBN	74.61±0.28	87.64±0.33	53.05±0.14	62.64±0.37	55.48±0.97	23.31±0.22
LP-Info(Min&BN)	**74.84**±0.31	87.81±0.45	**53.32**±0.23	63.11±0.33	57.31±0.99	**23.61**±0.27

- Further analysis show: graph generation quality usually aligns with downstream performance; and also molecule-specific generator (GraphAF here) alone does not significantly benefit molecular datasets.

Table 5: Link prediction performance (AUROC and AUPRC, %) of VGAE generators on eight pre-training datasets. Better link prediction results are marked in red if accompanied with better downstream performances, as shown in Table 2, 3 and 4.

	Principles	COLLAB	RDT-B	RDT-M5K	GITHUB	ogbg-ppa	ogbg-code	Trans-Mol	Trans-PPI
AUROC (%)	InfoMin	**71.28**	**97.32**	99.08	78.68	**96.53**	92.39	64.54	71.20
	InfoBN	69.44	97.29	**99.31**	81.20	95.24	94.06	**83.55**	**71.32**
AUPRC (%)	InfoMin	**80.84**	**96.62**	98.67	78.49	**95.94**	90.08	64.14	69.34
	InfoBN	79.13	96.59	**98.97**	80.47	95.30	91.66	**82.51**	**70.65**

Table 6: Learned prior performance with different generators under the guidance of InfoMin, in the transfer learning setting on molecular datasets. Red numbers indicate the best performances (AUROC, %).

Methods	BBBP	Tox21	ToxCast	SIDER	ClinTox	MUV	HIV	BACE
VGAE	**71.47**±0.66	**74.60**±0.70	**63.13**±0.30	60.52±0.75	72.39±1.50	70.51±2.25	**76.43**±0.85	**78.86**±1.66
GraphAF	70.55±0.63	73.51±0.43	62.03±0.33	**61.32**±1.32	**77.47**±1.91	**72.25**±1.18	76.30±1.34	78.43±2.36

References

[1] Yuning You et al., "Graph Contrastive Learning with Augmentations", NeurIPS'20. [2] Thomas Kipf & Max Welling, "Variational Graph Auto-Encoders", NeurIPS'16 Workshop. [3] Chence Shi et al., "Graphaf: a flow-based autoregressive model for molecular graph generation ", ICLR'20.

Contact: {yuning.you,yshen}@tamu.edu, {tianlong.chen,atlaswang}@utexas.edu