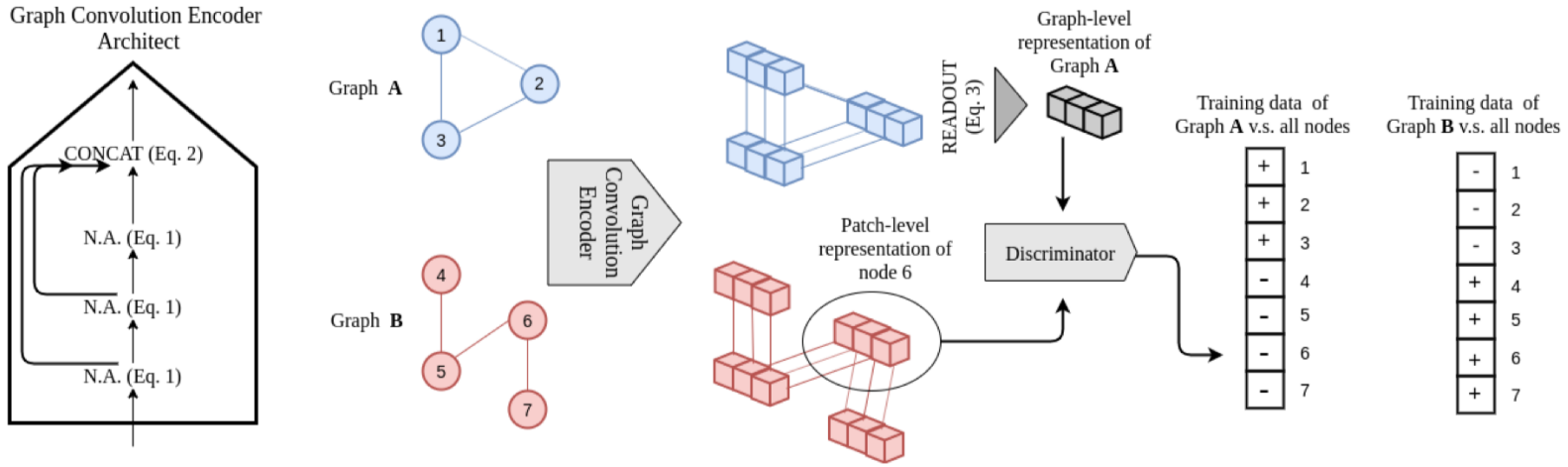
InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation Learning via Mutual Information Maximization

F. Sun, J. Hoffmann, V. Verma, J. Tang
ICLR 2020

Presenter: Minwoo Lee
2020.05.08

Methodology (Unsupervised)



Objective: Maximize Mutual Information of global representation and patch representation

- Patch representation: concatenation of node representations
- Global representation: A summary of all patch-representation for
- M.I. modeled with discriminator using Jensen-Shannon formulation

$$\hat{\phi}, \hat{\psi} = \arg\max_{\phi,\psi} \sum_{G \in \mathbf{G}} \frac{1}{|G|} \sum_{u \in G} I_{\phi,\psi}(\vec{h}^u_\phi; H_\phi(G)).$$

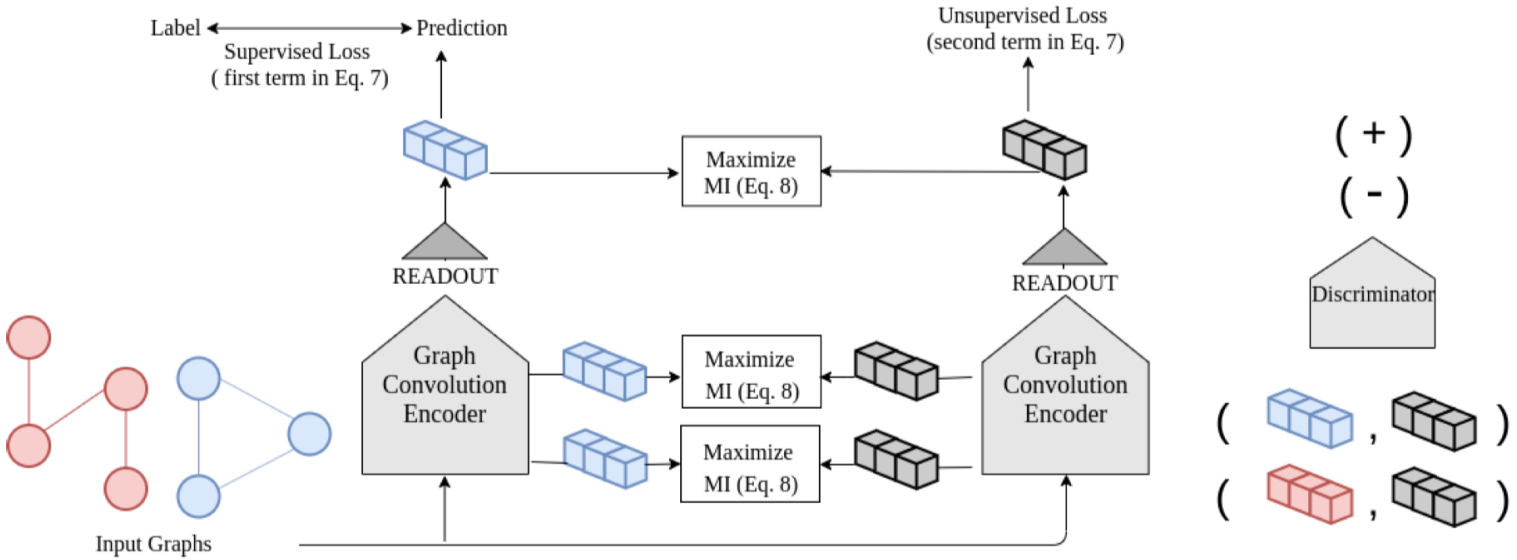
M.I.

$$I_{\phi,\psi}(h^i_\phi(G); H_\phi(G)) :=$$

$$\mathbb{E}_{\mathbb{P}}[-\mathrm{sp}(-T_{\phi,\psi}(\vec{h}^i_\phi(x), H_\phi(x)))] - \mathbb{E}_{\mathbb{P}\times\tilde{\mathbb{P}}}[\mathrm{sp}(T_{\phi,\psi}(\vec{h}^i_\phi(x'), H_\phi(x)))]$$

Discriminator

Methodology (Semi-supervised)



- Simple combination of both supervised & unsupervised terms could lead to negative transfer
- use student-teacher model & combine the models maximizing MI between each graph representation of each layers

$$L_{\text{total}} = \sum_{i=1}^{|\mathbb{G}^L|} L_{\text{supervised}}(y_\phi(G_i), o_i) + \sum_{j=1}^{|\mathbb{G}^L|+|\mathbb{G}^U|} L_{\text{unsupervised}}(h_\varphi(G_j); H_\varphi(G_j))$$

$$- \lambda \sum_{j=1}^{|\mathbb{G}^L|+|\mathbb{G}^U|} \frac{1}{|G_j|} \sum_{k=1}^{K} I(H_\phi^k(G_j); H_\varphi^k(G_j)).$$

student model

teacher model

Experiments

Unsupervised Learning

Dataset	MUTAG	PTC-MR	RDT-B	RDT-M5K	IMDB-B	IMDB-M
(No. Graphs)	188	344	2000	4999	1000	1500
(No. classes)	2	2	2	5	2	3
(Avg. Graph Size)	17.93	14.29	429.63	508.52	19.77	13.00
Graph Kernels						
RW	83.72 ± 1.50	57.85 ± 1.30	OMR	OMR	50.68 ± 0.26	34.65 ± 0.19
SP	85.22 ± 2.43	58.24 ± 2.44	64.11 ± 0.14	39.55 ± 0.22	55.60 ± 0.22	37.99 ± 0.30
GK	81.66 ± 2.11	57.26 ± 1.41	77.34 ± 0.18	41.01 ± 0.17	65.87 ± 0.98	43.89 ± 0.38
WL	80.72 ± 3.00	57.97 ± 0.49	68.82 ± 0.41	46.06 ± 0.21	72.30 ± 3.44	46.95 ± 0.46
DGK	87.44 ± 2.72	60.08 ± 2.55	78.04 ± 0.39	41.27 ± 0.18	66.96 ± 0.56	44.55 ± 0.52
MLG	87.94 ± 1.61	$\mathbf{63.26 \pm 1.48}$	> 1 Day	> 1 Day	66.55 ± 0.25	41.17 ± 0.03
Other Unsupervised Methods						
node2vec	72.63 ± 10.20	58.58 ± 8.00	-	-	-	-
sub2vec	61.05 ± 15.80	59.99 ± 6.38	71.48 ± 0.41	36.68 ± 0.42	55.26 ± 1.54	36.67 ± 0.83
graph2vec	83.15 ± 9.25	60.17 ± 6.86	75.78 ± 1.03	47.86 ± 0.26	71.1 ± 0.54	$\mathbf{50.44 \pm 0.87}$
InfoGraph	$\mathbf{89.01 \pm 1.13}$	61.65 ± 1.43	$\mathbf{82.50 \pm 1.42}$	$\mathbf{53.46 \pm 1.03}$	$\mathbf{73.03 \pm 0.87}$	49.69 ± 0.53

Semi-supervised Learning

Target	Mu (0)	Alpha (1)	HOMO (2)	LUMO (3)	Gap (4)	R2 (5)	ZPVE(6)	U0 (7)	U (8)	H (9)	G(10)	Cv (11)
MAE	0.3201	0.5792	0.0060	0.0062	0.0091	10.0469	0.0007	0.3204	0.2934	0.2722	0.2948	0.2368

Semi-Supervised	Error Ratio											
Mean-Teachers	1.09	1.00	**0.99**	1.00	**0.97**	0.52	0.77	1.16	0.93	0.79	0.86	0.86
InfoGraph	1.02	0.97	1.02	**0.99**	1.01	0.71	0.96	0.85	0.93	0.93	0.99	1.00
InfoGraph*	**0.99**	**0.94**	**0.99**	**0.99**	0.98	**0.49**	**0.52**	**0.44**	**0.58**	**0.57**	**0.54**	**0.83**