COMP396
Meeting 5

Summary

1. Recall
2. Relevant Papers
3. Supporting Material
4. JSD
5. Code
6. What I don't understand yet?

Recall

- We had an overview of DIM objective functions
- Discussed points to focus on (i.e infograph and JSD)

Relevant Papers

-LEARNING DEEP REPRESENTATIONS BY MUTUAL INFORMATION ESTIMATION AND MAXIMIZATION->https://arxiv.org/pdf/1808.06670.pdf

-DEEP GRAPH INFOMAX: https://arxiv.org/pdf/1809.10341.pdf

-INFOGRAPH: UNSUPERVISED AND SEMI-SUPERVISED GRAPH-LEVEL REPRESENTATION LEARNING VIA MUTUAL INFORMATION MAXIMIZATION->https://arxiv.org/pdf/1908.01000.pdf

Supporting Material

-Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results-> https://arxiv.org/pdf/1703.01780.pdf

FASTGCN: FAST LEARNING WITH GRAPH CONVOLUTIONAL NETWORKS VIA IMPORTANCE SAMPLING-> https://arxiv.org/pdf/1801.10247.pdf

Mutual Information Neural Estimation->https://arxiv.org/pdf/1801.04062.pdf

f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization
https://arxiv.org/pdf/1606.00709.pdf

HOW POWERFUL ARE GRAPH NEURAL NETWORKS?

JSD

The mutual information estimator modeled by discriminator Tψ

$$I_{\phi,\psi}(h^i_\phi(G); H_\phi(G)) := \mathbb{E}_{\mathbb{P}}[-\mathrm{sp}(-T_{\phi,\psi}(\vec{h}^i_\phi(x), H_\phi(x)))] - \mathbb{E}_{\mathbb{P}\times\tilde{\mathbb{P}}}[\mathrm{sp}(T_{\phi,\psi}(\vec{h}^i_\phi(x'), H_\phi(x)))] \quad (5)$$

Unsupervised learning

h is the summarized patch representation centered at node u and Hφ(G) is the global representation after applying READOUT.

$$\hat{\phi}, \hat{\psi} = \arg\max_{\phi,\psi} \sum_{G\in\mathbf{G}} \frac{1}{|G|} \sum_{u\in G} I_{\phi,\psi}(\vec{h}^u_\phi; H_\phi(G)). \quad (4)$$

Here we use GIN:

Semi-supervised learning

$$L_{\text{total}} = \sum_{i=1}^{|\mathbb{G}^L|} L_{\text{supervised}}(y_\phi(G_i), o_i) + \sum_{j=1}^{|\mathbb{G}^L|+|\mathbb{G}^U|} L_{\text{unsupervised}}(h_\varphi(G_j); H_\varphi(G_j)) \quad (7)$$

$$- \lambda \sum_{j=1}^{|\mathbb{G}^L|+|\mathbb{G}^U|} \frac{1}{|G_j|} \sum_{k=1}^{K} I(H_\phi^k(G_j); H_\varphi^k(G_j)). \quad (8)$$

Code (setup)

https://colab.research.google.com/drive/1XC2H9SzhT7myepPNJO8WX0hoDBRuhFcV

NB: parameters are not correctly set as in the actual experiment

Materials & Resources

InfoGraph Code: https://github.com/fanyun-sun/InfoGraph

Deep Graph Infomax (Original): https://github.com/PetarV-/DGI

Deep Graph Infomax: https://github.com/rusty1s/pytorch_geometric/blob/master/torch_geometric/nn/models/deep_graph_infomax.py

What I don't understand yet

- relationship between mean teacher and how mean teacher is used here (function) is ambiguous

To do:

Focus on global global (smiles + graph) + local global (node, readout) unsupervised DIM