Generating Counterfactual Hard Negative Samples for Graph Contrastive Learning

Haoran Yang[1], Hongxu Chen[1], Sixiao Zhang[1], Xiangguo Sun[2], Qian Li[3], Xiangyu Zhao[4], Guandong Xu[1]

[1]University of Technology Sydney,

[2]Chinese University of Hong Kong,

[3]Curtin University,

[4]City University of Hong Kong

1 Introduction

1.1 Background about graph contrastive learning (GCL)

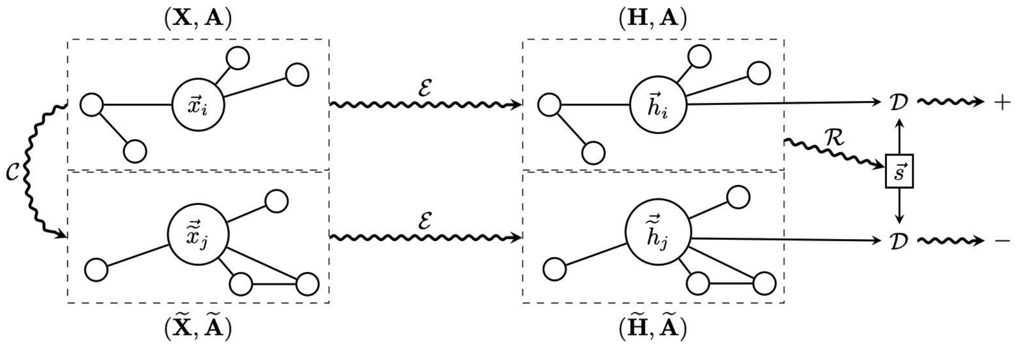

Fig. Overview of Deep Graph Infomax (DGI)[1]

Data augmentation	Type	Underlying Prior
Node dropping	Nodes, edges	Vertex missing does not alter semantics.
Edge perturbation	Edges	Semantic robustness against connectivity variations.
Attribute masking	Nodes	Semantic robustness against losing partial attributes per node.
Subgraph	Nodes, edges	Local structure can hint the full semantics.

Tab. Summary of graph augmentation methods.[2]

➢Graph perturbation:

- ***Node dropping*** drops nodes and related edges to augment the graph, the underlying prior is that vertex missing does not alter semantics.
- ***Edge perturbation*** drops or adds some edges in the graph, the underlying prior is that semantic robustness against connectivity variations.
- ***Attribute masking*** discards partial attributes, the underlying prior is that semantic robustness against losing partial attributes per node.

[1] Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, R Devon Hjelm, "Deep Graph Infomax", ICLR 2019
[2] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, Yang Shen, "Graph Contrastive Learning with Augmentations", NeurIPS 2020

1 Introduction

1.2 Limitations and chanllenges in current research works

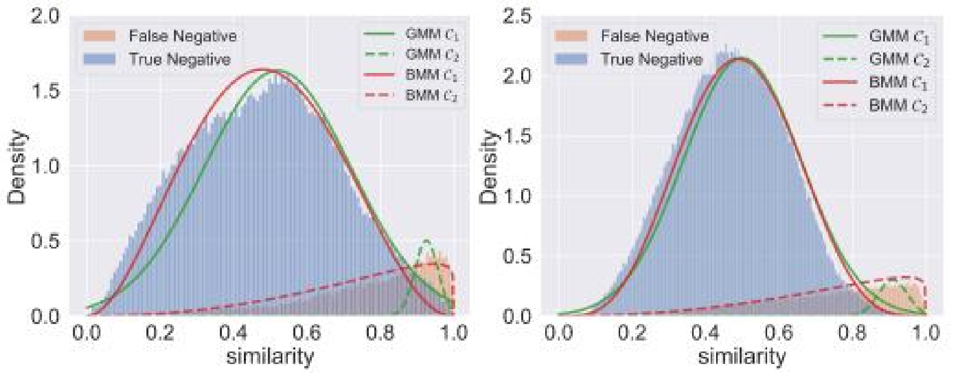

Fig. False negative samplesare more likely occurs when similarity is high.[1]

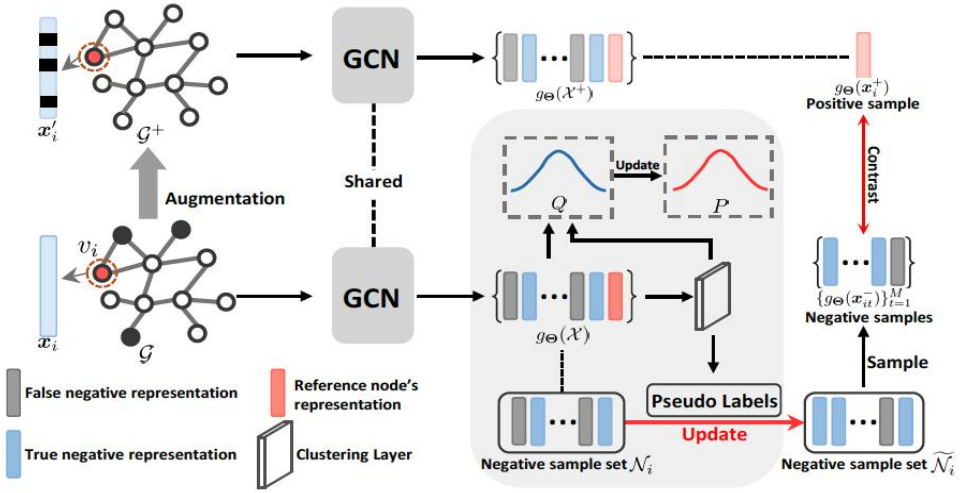

Fig. An attempt to reduce the impact brought by false negative samples.[2]

- Random factors also affect GCL performance:
  - Graph augmentations introduce noise into the data.
- False negative samples affect GCL performance:
  - False negative samples are more likely occur when the negative samples are very similar to the target.
  - Some current research works try to relieve the impact brought by false negative samples.
- Complex protocols increase complexity:
  - Current post-processing methods reduce the probabilities by sophisticated desgins after contrasting sample generation.

[1] Jun Xia, Lirong Wu, Jintao Chen, Ge Wang, Stan Z. Li, Qiang Liu, Shu Wu, Liang Wang, "Debiased Graph Contrastive Learning", arXiv 2021
[2] Han Zhao, Xu Yang, Zhenru Wang, Erkun Yang, Cheng Deng, "Graph Debiased Contrastive Learning with Joint Representation Clustering", IJCAI 2021

2 Methodology

2.1 Counterfactual mechanism

Algorithm 1: Counterfactual generation heuristic

sample a random instance as the initial x'
optimise $L(x, x', y', \lambda)$ with initial x'
while $|\hat{f}(x') - y'| > \varepsilon$ **do**
    increase λ by step-size α
    optimise $L(x, x', y', \lambda)$ 1 with new x'
return x'

x: interested instance
x′ : counterfactual
y′: desired outcome
λ : balances two terms
ϵ : tolerance for the distance

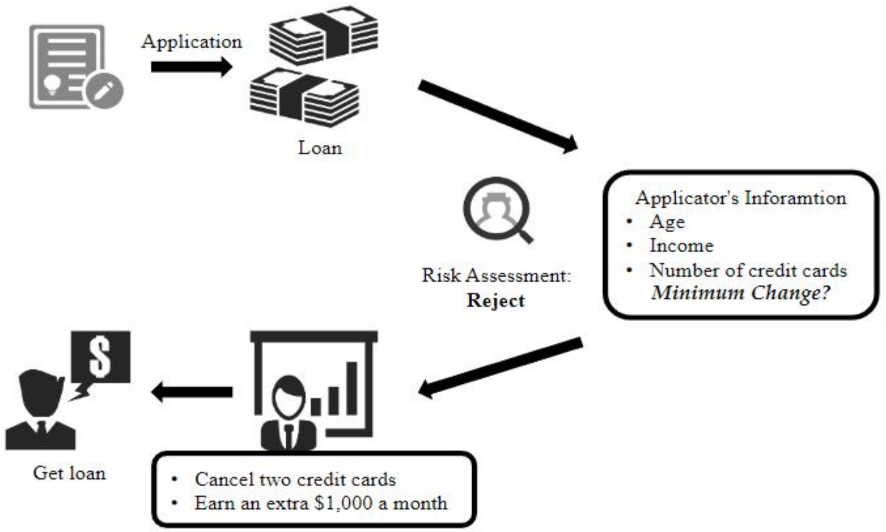

Fig. A toy example for understanding the rationale of the counterfactual mechansim.

- True hard negative samples:
  - High-quality negative samples help contrastive learning model capture critical information.
  - However, most sampled hard negative instances are false.

- Counterfactual mechansim:
  - Minimum changes to ensure the ‘hard’.
  - Different outcomes to ensure the ‘negative’.

2 Methodology

2.2 The overview of the proposed method

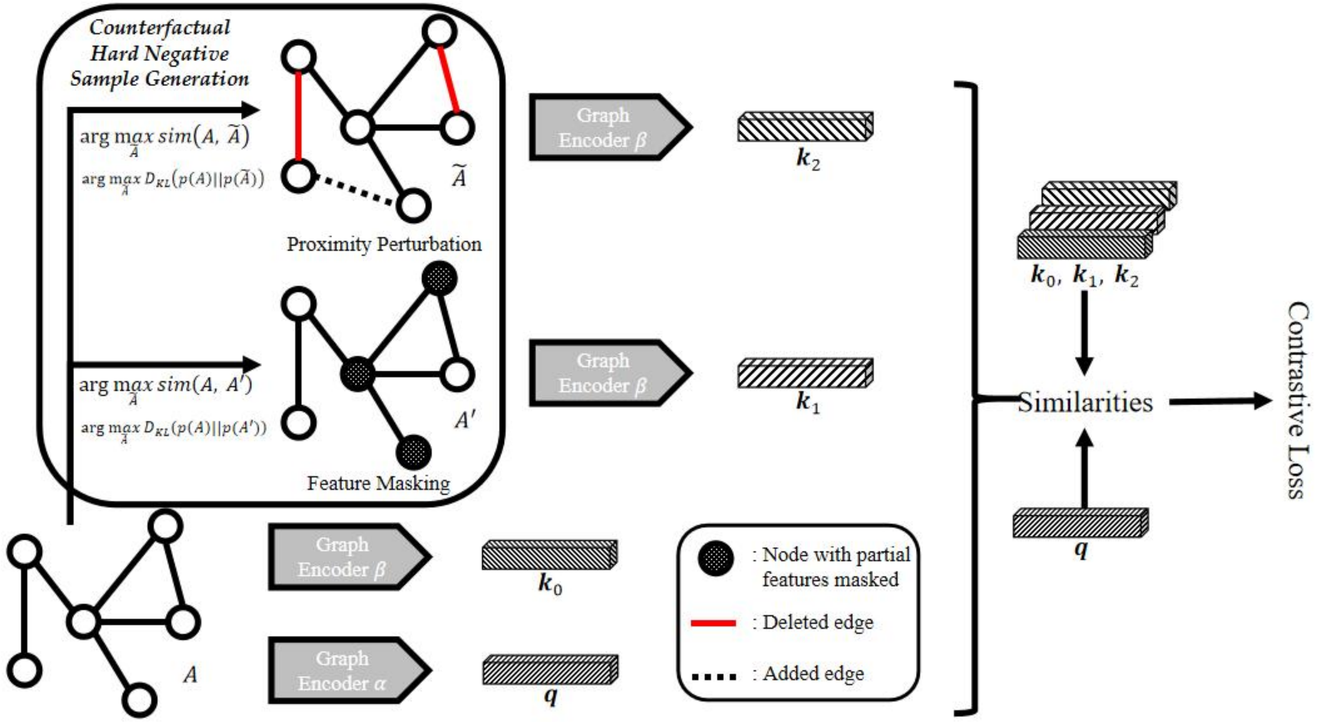

Fig. The overview of CGC. We first conduct counterfactual hard negative sample generation to acquire a proximity-perturbed and feature-masked sample. Then, the target and the two generated hard negative samples will be fed into the graph contrastive learning module to learn graph embeddings

- Adaptive graph augmentations:
  - Proximity perturbation - deleting and adding edges in the graph.
  - Feature masking - masking a portion of values of the graph feature matrix.
- Counterfactual palys the role:
  - to make the predictor to give a different result for the augmented graph.
  - to ensure the perturbations on the graph is minimum.
- Graph contrastive learning:
  - The augmented graphs will be coupled with the target graph.

3 Experiments

➢ Dataset Statistics

Dataset	Num. of Graphs	Avg. Num. of Nodes	Avg. Num. of Edges	Node Attr. (Dim.)	Classes
ENZYMES	600	32.63	62.14	18	6
PROTEINS_full	1,113	39.06	72.82	29	2
Synthie	400	95.00	172.93	15	4
FRANKENSTEIN	4,337	16.90	17.88	780	2

➢ Comparison experiments

Method \ Dataset	PROTEINS_full		FRANKENSTEIN		Synthie		ENZYMES	
	F1-Micro	F1-Macro	F1-Micro	F1-Macro	F1-Micro	F1-Macro	F1-Micro	F1-Macro
RandomWalk	-	-	57.97(std 2.15)	57.45(std 1.92)	18.50(std 4.06)	16.86(std 3.58)	-	-
ShortestPath	70.88(std 4.91)	69.88(std 4.99)	62.39(std 1.95)	59.81(std 2.02)	50.75(std 9.69)	47.32(std 9.86)	27.83(std 6.37)	27.18(std 5.93)
GL	69.89(std 3.25)	68.57(std 3.43)	61.26(std 2.85)	53.94(std 2.46)	52.50(std 10.49)	50.24(std 10.37)	31.67(std 7.03)	30.02(std 7.13)
WL	72.32(std 3.11)	71.36(std 3.41)	-	-	-	-	37.83(std 4.95)	36.42(std 5.78)
sub2vec	70.17(std 2.06)	66.26(std 0.44)	54.97(std 1.80)	46.83(std 4.00)	29.75(std 4.67)	22.07(std 3.75)	19.67(std 3.64)	13.34(std 4.33)
graph2vec	68.65(std 3.45)	64.16(std 5.00)	61.70(std 3.04)	59.68(std 0.22)	54.25(std 0.62)	35.17(std 0.26)	25.67(std 4.84)	22.41(std 5.04)
InfoGraph	71.61(std 4.67)	70.48(std 5.06)	63.57(std 2.12)	62.95(std 2.20)	54.5(std 8.05)	54.17(std 7.87)	38.33(std 7.03)	37.07(std 6.89)
MVGRL	72.06(std 3.29)	69.53(std 3.61)	61.89(std 1.40)	59.65(std 1.50)	62.00(std 9.07)	61.59(std 9.52)	40.50(std 7.85)	38.7(std 9.12)
GraphCL	73.05(std 3.29)	**71.04(std 3.35)**	62.62(std 2.49)	61.89(std 2.57)	57.50(std 9.08)	55.87(std 8.87)	33.67(std 4.58)	33.46(std 4.96)
GCA	71.71(std 4.40)	69.59(std 4.44)	63.20(std 1.70)	62.17(std 1.57)	52.25(std 5.18)	43.27(std 9.85)	34.00(std 5.01)	33.62(std 5.01)
CGC	**73.48(std 4.90)**	70.03(std 5.75)	**64.93(std 1.98)**	**63.25(std 2.04)**	**63.75(std 6.91)**	**63.23(std 6.71)**	**47.50(std 6.25)**	**46.99(std 6.30)**

Thanks for your listening!