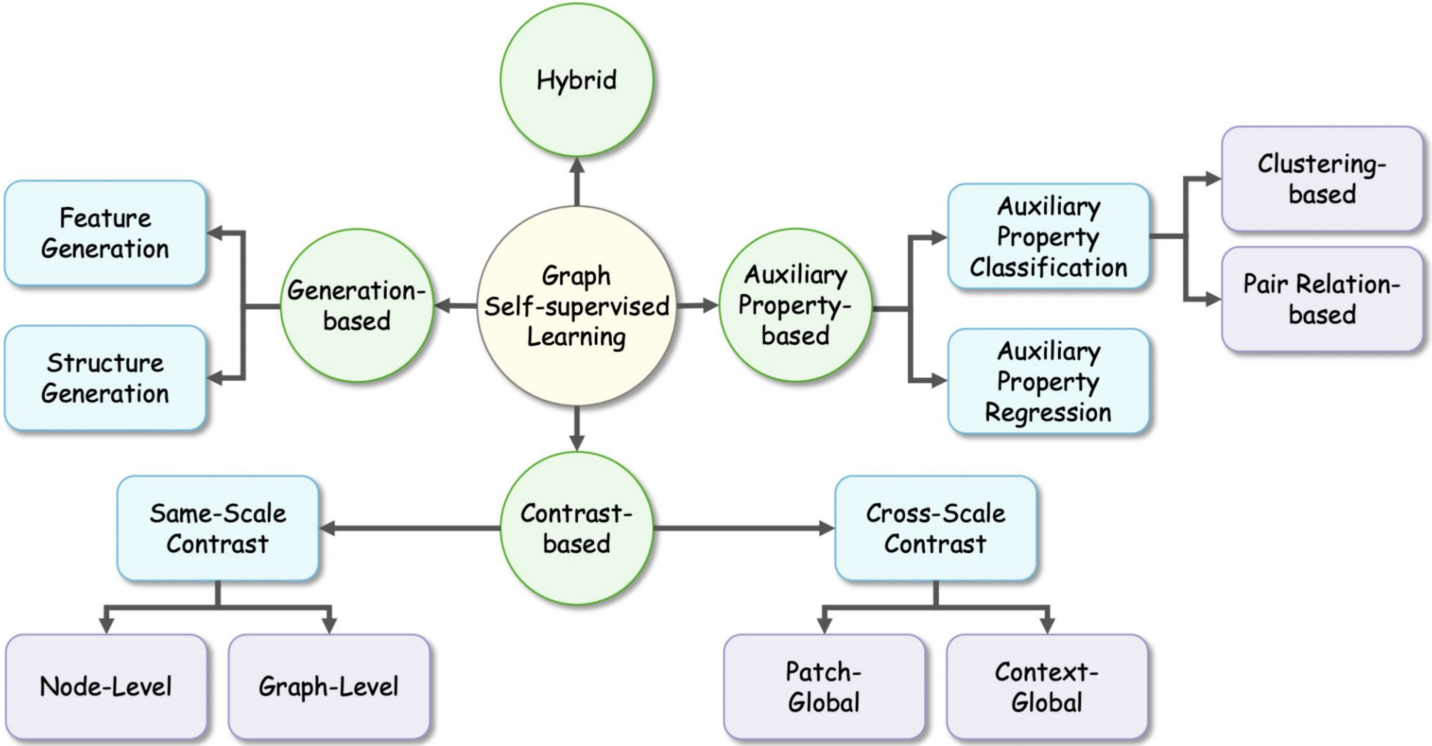
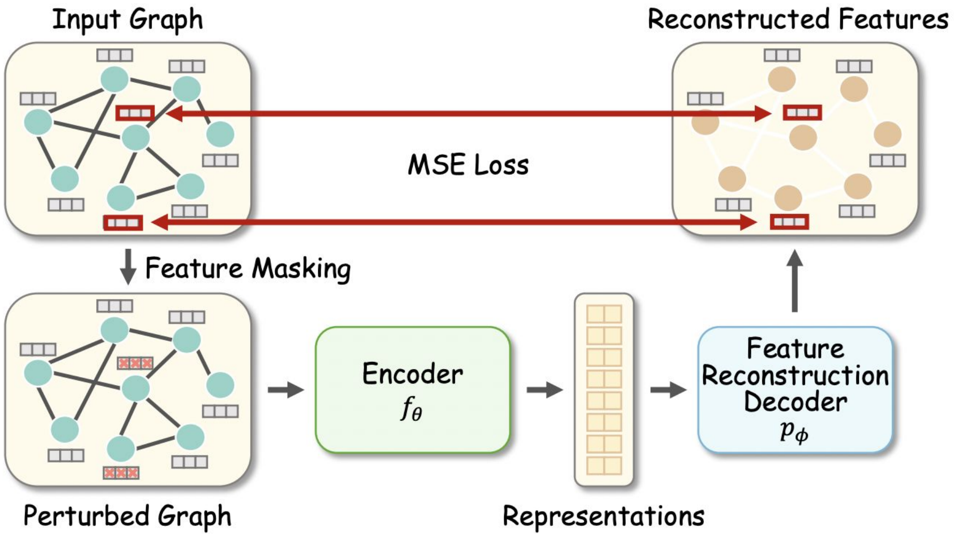
Self-supervised learning for graphs

Ildus Sadrtdinov, 14.03.23

SSL graph tasks taxonomy

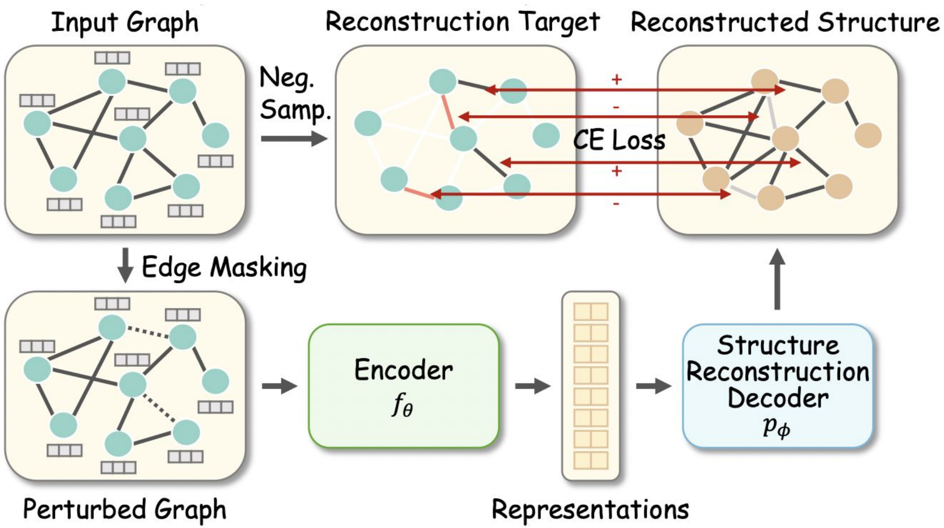


Generation-based: Graph Completion



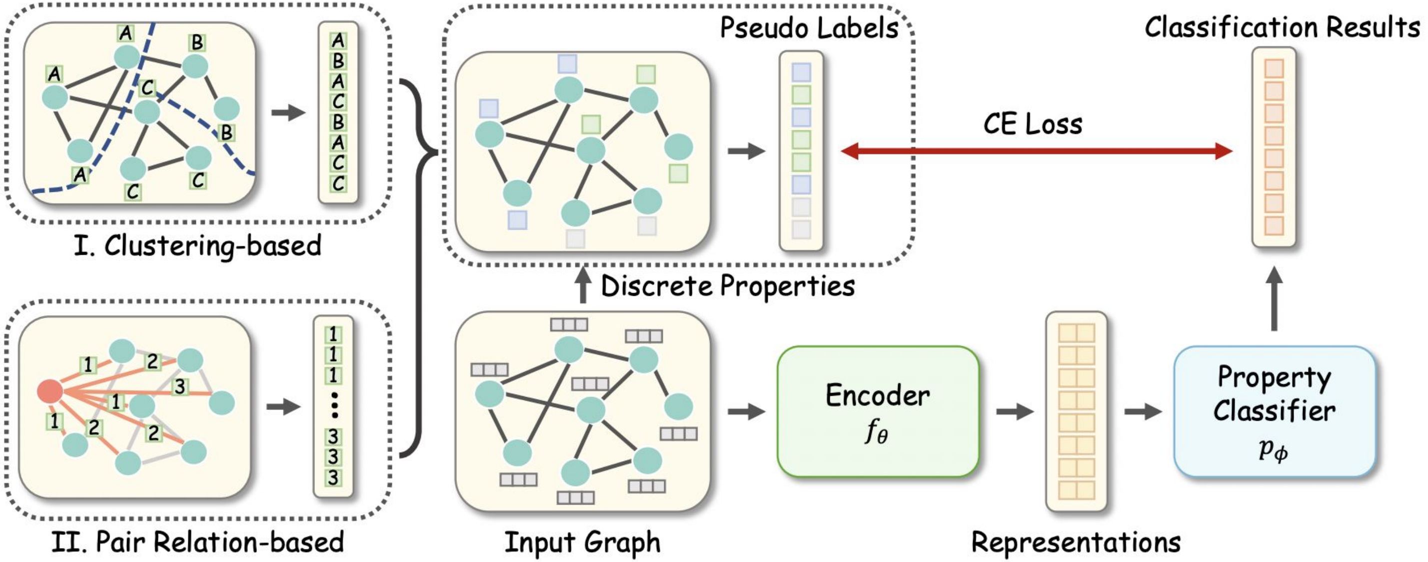
(a) Graph Completion is a representative approach of feature generation-based graph SSL. The features of certain nodes are masked and then fed into the model, and the learning target is to reconstruct the masked features. An MSE loss is used to recover the features.

Generation-based: Link Reconstruction



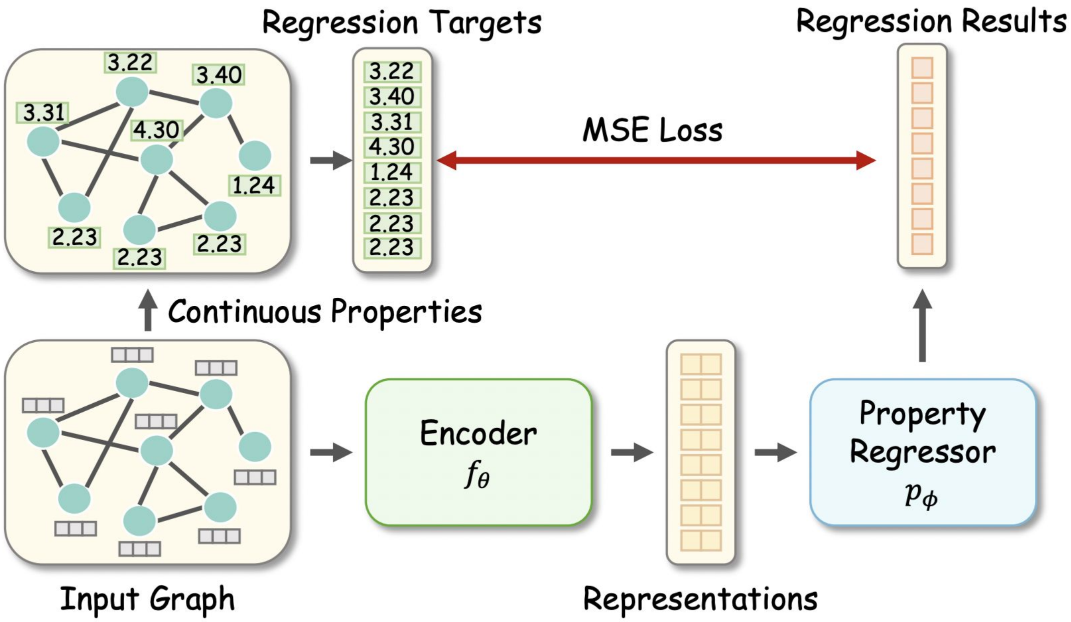
(b) The objective of Denoising Link Reconstruction is to rebuild the masked edges. A binary cross-entropy (BCE) loss is employed to train the model where existing edges are the positive samples and the unrelated node pairs are the negative samples. A negative sampling (Neg. Samp.) strategy is used to balance the classes.

Auxiliary property classification



(a) Auxiliary property classification methods extract discrete properties as pseudo labels, and the pretext decoder is used to predict the classification results. A CE loss function is used to train the models. Two types of properties (clustering-based and pair relation-based) can be used to define the pseudo labels.

Auxiliary property regression



(b) Auxiliary property regression methods aim to predict the continuous auxiliary properties with the decoder, where models are trained by the MSE loss.

Graph augmentations for CL

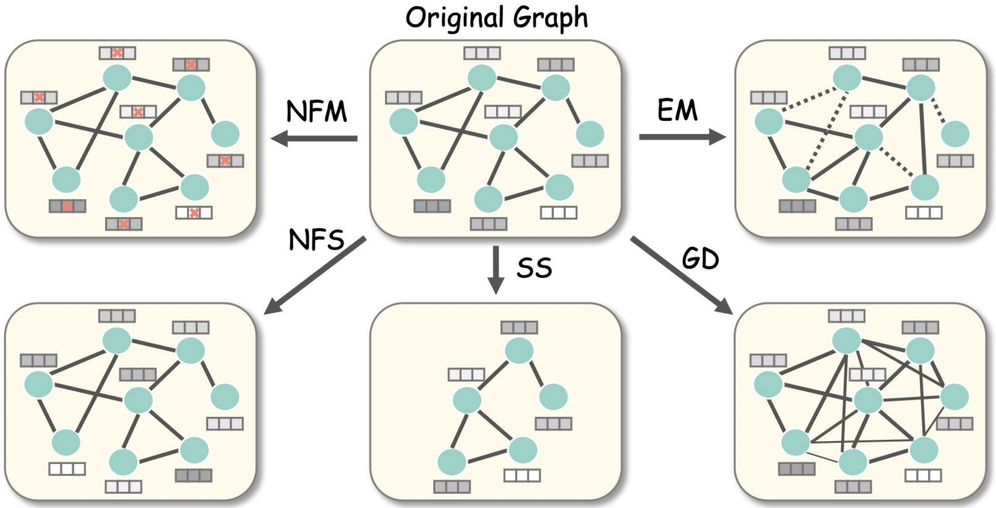


Fig. 7: Brief examples of five types of common graph augmentations, including Node Feature Masking (NFM), Node Feature Shuffle (NFS), Edge Modification (EM), Graph Diffusion (GD), and Subgraph Sampling (SS).

$$t_i(\mathbf{A}) = \sum_{k=0}^{\infty} \Theta_k \mathbf{T}^k$$

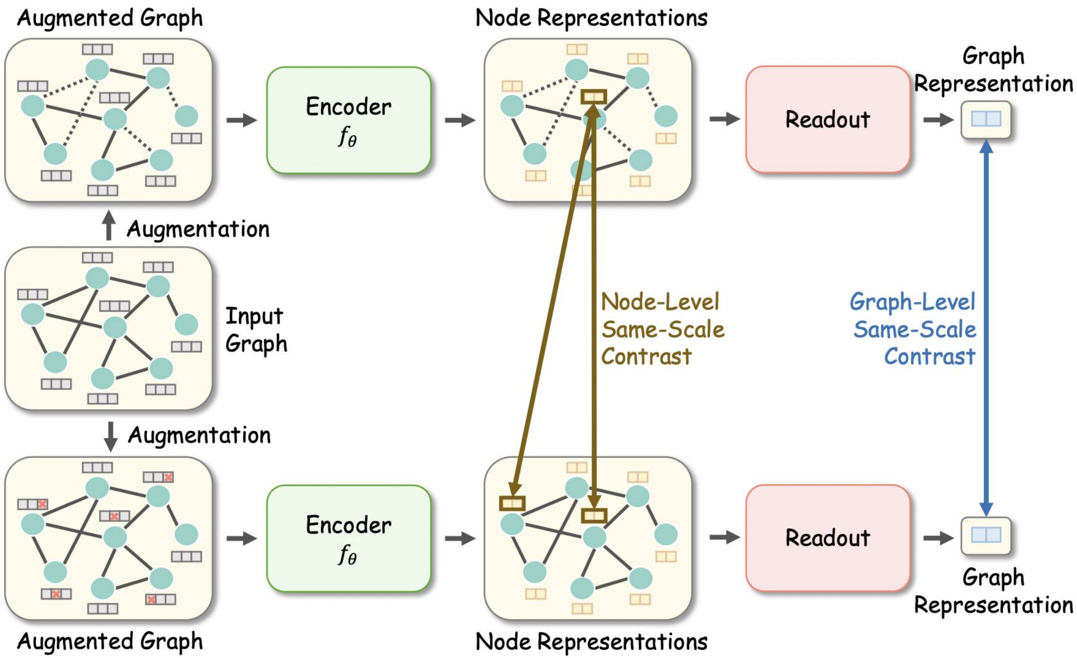
$$\Theta_k = \frac{e^{-\iota} t^k}{k!} \qquad \mathbf{T} = \mathbf{A}\mathbf{D}^{-1}$$

$$t_i(\mathbf{A}) = \exp\left(\iota \mathbf{A}\mathbf{D}^{-1} - \iota\right)$$

$$\Theta_k = \beta(1-\beta)^k \qquad \mathbf{T} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$$

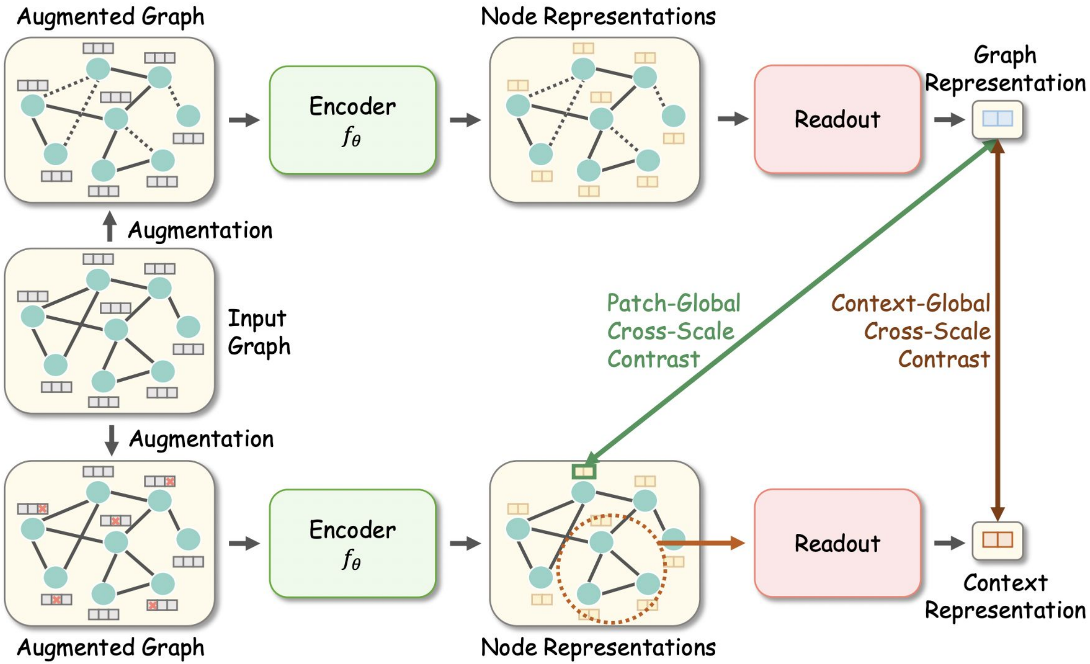
$$t_i(\mathbf{A}) = \beta\left(\mathbf{I} - (1-\beta)\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\right)$$

Same-scale contrastive learning



(a) In same-scale contrast, different views are generated with various graph augmentations based on the input graph at first. Then, the contrast can be placed on node or graph scales.

Cross-scale contrastive learning



(b) In cross-scale contrast, the augmented graph views are generated firstly. Then, the cross-scale contrast aims to discriminate patch- or context- level embeddings with global representations.