

分子属性以及蛋白质功能预测可以看作是 graph 属性的预测，所以一个好的图表示是必须的。但是目前在这个领域存在两个挑战：

1. 昂贵的实验，使得标签数据稀缺；
2. 将模型推广到训练数据以外的数据集上。

这篇文章的主要工作是探索预训练，解决上述两个问题。特别地，

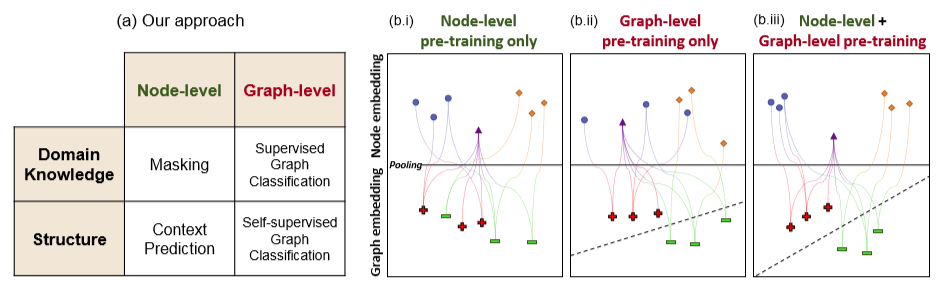
使用 GNNs 进行图表示学习，希望获得节点表示：

1) 捕获节点网络邻域结构的相似性；

2) 可以进行组合来提供准确的graph-level 表示；

3) 捕获 domain-knowledge。

要满足这个目标，我们提出了基于 node-level和 graph-level 一系列预训练 GNNs 的方法，使用的数据来自一些相关的监督任务，包括标签数据和没有标签的数据。



(a) Categorization of the pre-training methods for GNNs. Crucially, our methods, i.e., Context Prediction, Masking, and graph-level supervised pre-training, cover both node-level and graph-levelpre-training.

(b) Node and graph embeddings obtained by different pre-training strategies.

(b.i) When only node-level pre-training is used, nodes of different shapes (semantically different nodes) can be well separated, however, the node embeddings are not composable, and thus resulting graph embeddings (denoted by their classes, + and−) that are created by pooling embeddings of individual nodes are not separable.

(b.ii) With graph-level pre-training only, graph embeddings are well separated, however the embeddings of individual nodes do not necessarily capture their domain-speciﬁc semantics.

(b.iii) High-quality node embeddings are such that nodes of different types are well separated, while at the same time, the embedding space is also composable. This allows for accurate and robust representations of entire graphs, which allows robust transfer of pre-trained models to a variety of downstream tasks.

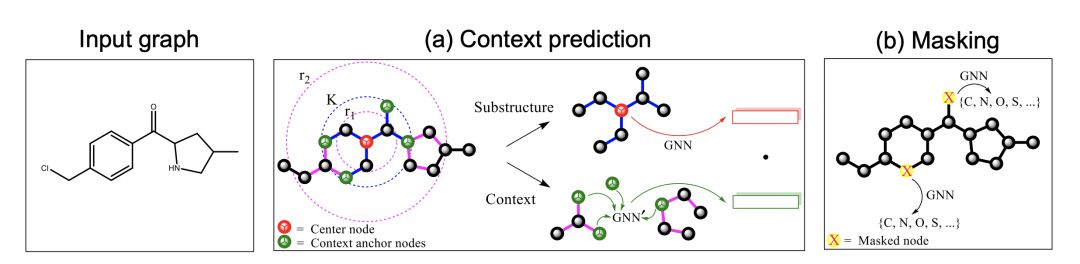


Figure 2: Illustration of Context Prediction and Masking for pre-training GNNs.

(a) In Context Prediction, the substructure is deﬁned as a K-hop subgraph around a selected center node, where K is the number of GNN layers and is set to 2 in the ﬁgure. The context is deﬁned as the surrounding subgraph that is between r1- and r2-hop from the center node, where we use r1 = 1 and r2 = 4 in the ﬁgure.

(b) In Masking, the input node/edge attributes are randomly masked, and the GNN is asked to predict them.

三个关键点

Context Prediction, which is a novel self-supervised node-level pre-training method that applies the distributional hypothesis [44, 31] to the graph domain. In particular, we use node embeddings to predict surrounding graph structure, so nodes that have similar surrounding graph structure will be mapped into similar representations.

Masking. To learn node embeddings that capture domain knowledge, we propose a novel selfsupervised node-level pre-training method called Masking. In Masking, we randomly mask input node/edge attributes and let GNNs predict the masked attributes from the surrounding structure. For example, in the chemistry application, we can use node embeddings to predict atom types of masked atoms, as illustrated in Figure 2 (b). This forces the model to capture chemistry domain knowledge, such as valency and the electronic or steric properties of functional groups [30].

Graph-level Prediction. To learn composable node embeddings that are useful for downstream tasks, we can either perform (1) supervised graph-level pre-training on domain-speciﬁc auxiliary tasks, or(2)self-supervisedpre-trainingtopredictstructuralpropertiesofthegraphs. Here, todirectly encode domain knowledge into graph embeddings, we take the ﬁrst approach and combine our novel Context Prediction and Masking methods with graph-level supervised pre-training. This ensures that individual node embeddings are easily composed to obtain domain-speciﬁc representations of an entire graph, as illustrated in Figure 1 (b.iii).

文章在分子属性和蛋白质功能两个应用中做了广泛测试。作者观察到，与未进行预训练的模型相比，graph-level 的监督预训练会导致边际性能提高，有时候甚至使性能下降。另一方面，同时结合了 node-level 和 graph-level 的预训练提高了在模型的泛化能力，同时比没有经过预训练的模型效果更好。

**论文链接：https://arxiv.org/pdf/1905.12265.pdf**