

日期: /

## Finding Frequent Subgraphs

问题: 给定大图  $G$ , 给定尺寸  $k$ , 识别  $r$  个出现频率最高的子图

寻找某个图中最频繁出现的 size- $k$  子图通常面临两大难点:

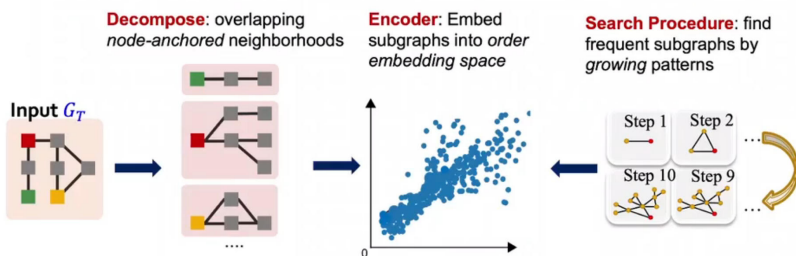
① 枚举: 需要枚举所有尺寸为  $k$  的子图

以渐增的方式构建子图

② 计数: 统计所有子图出现的频率

使用 GNN 预测频数而非显式计数

### [SPMiner]



给定查找  $G_Q$ , 只需将  $G_T$  随机分解为一系列  $k$ -hop 邻域  $G_{N_i}$ , 对其中嵌入  $Z_{N_i}$  满足  $Z_{N_i} \geq Z_Q$  的图计数。

日期:

/

Initial step: 从  $G_T$  中随机采样一个单节点  $u$  开始,  
令  $S = \{u\}$

Iteratively: 通过向模体中加入  $S$  中节点在  $G_T$  中的  
邻居节点对模体进行扩增  
在构建的过程中,贪婪地选择要加入的节点

[Total violation]

子图  $G$  的总违反度, 定义为不包含  $G$  的  
邻域数。即不满足  $2a \leq 2n_i$  的  $G_{N_i}$  数。  
最小化总违反度意味着最大化频率。