

图匹配中idea3

针对噪声点和图结构的问题

- 具体解决图网络输入的问题

目前网络是通过将两个图前后输入同一个GNN中然后得到匹配，会导致网络无法得到两个图之间的相似信息。

PCA的工作提出了将目标域按照相似度变换到源域，进行拼接后输入一个GNN，该方法可能会造成过度平滑。

Related work

Outlier的匹配相关工作

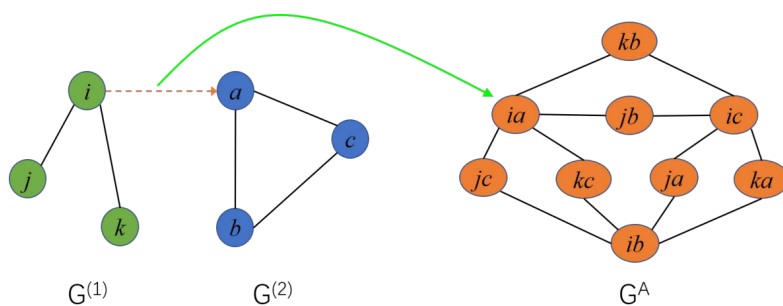
图匹配的outlier定义为源graph中不与目标graph匹配的节点。outlier无法从单个graph中确定，而是要通过两个graph的联合信息确定。

网络对outlier的处理上：网络需要对输入的两个graph同时进行处理，而不是依次单独处理。首个深度网络GMN[1]将深度网络引入图匹配问题中用于特征的embedding，然而没有将两个graph**联合输入网络**中，PCA-GM[2]将把目标域的特征按**相似度线性变换**后拼接入源域特征中，但这种直接的线性变换可能会导致影响节点特征的非线性关系，从而导致过度平滑。BBGM只通过将单个graph依次输入共享权重GNN中，无法让源图在forward中学习到目标图的信息。

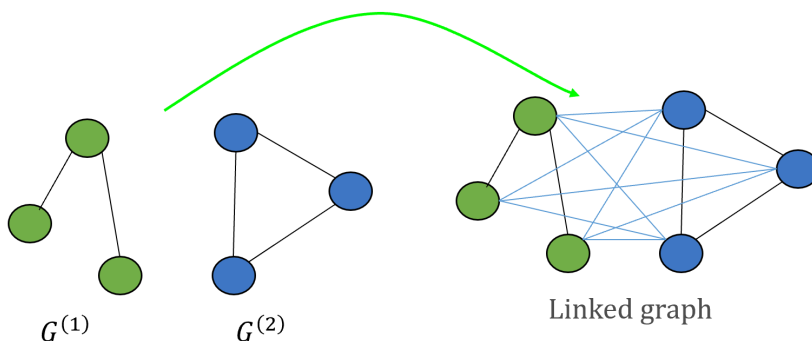
在后处理阶段对outlier的处理：图匹配问题要满足**一一匹配**的约束，即每个节点的匹配可能性求和为1。传统计算的Hangrain算法由于不能求梯度而无法作为网络训练的匹配求解算法。PCA-GM, DQGM[3]使用可求导的sinkhorn迭代求解匹配，但是由于sinkhorn会让所有节点满足**一一匹配**约束，因此同时也会让**outlier的与目标域的匹配可能性求和为1**，从而无法被剔除。

图匹配问题转化为分类问题

- 可以将图匹配转化为节点分类问题。Tao等人[4]直接将两个图结构构造成为associate graph，考虑每一个边之间的关系。计算复杂度为 $O((N_e * M_e)^2)$ 。然而

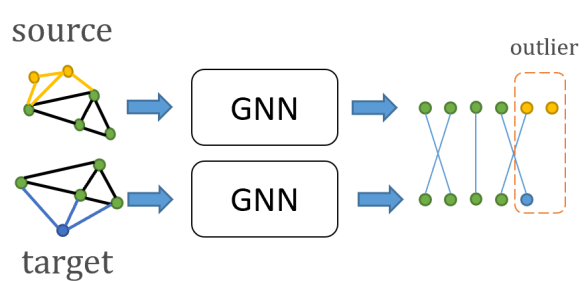


- 而我们直接可以将两个图直接连接起来，构造link graph，此时可以更好地分理出outlier，并且计算复杂度为 $O((N_v + M_v)^2)$ ，当节点的数量增加时，该方法效率会更加高。

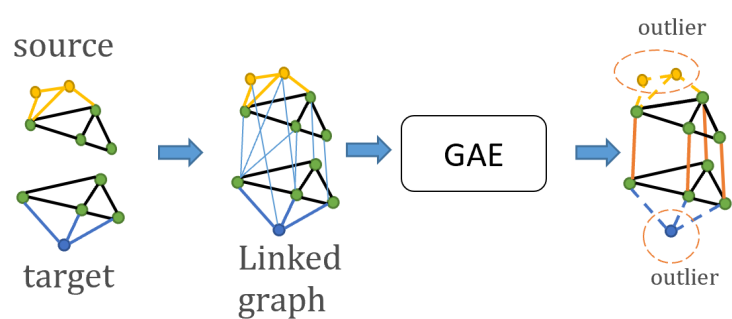


motivation

- 针对outlier的特性，我们应该考虑让网络能够同时处理两个图，而不是分别进行输入。下图为前人的工作，GNN之间虽然是共享权值，但是在forward过程中是无法得知目标图的信息，这导致outlier的区分度仅仅依靠单个图上的特征分布决定，而无法通过两个图的相互关系得到。我们的目的：
- 针对让网络剔除outlier的问题，将两个图同时输入网络，让网络能够学习到两个图的信息。



Method



- 如上图所示，将两个网络首先计算节点相似度，然后生成对应的连接边，构建一个包含两个图的连接图linked graph。将连接图输入到图生成网络中，输出的是包含了改进后的各个单图**独立边**结构和求解匹配的**匹配边**，并且对outlier的边进行抹除，后处理中通过检测孤立节点从而剔除outlier。
- 该方法将图匹配问题转化为图结构生成问题，不仅仅让网络求解匹配关系，同时还对图的结构进行重生成，让图的结构更加一致。

Loss

- 匹配边的分类Loss：将两个图之间的**匹配边** \hat{Y}_e 进行交叉熵。 $L = -Y_e \ln(\hat{Y}_e)$
- 图结构一致性Loss：用匹配关系作为监督，令两个图之间的结构更加一致。 $L = \|\hat{Y}_s - X\hat{Y}_tX^T\|_2$
- outlier Loss：outlier的应该是孤立点，减少边节点 \hat{Y}_e 的Loss。 $L = \sum_{Y_e \in Y_{outlier}} (1 - \hat{Y}_e)$

贡献点

- 将图匹配问题转化为图结构生成问题，并构建连接图求解问题。
- 对图匹配中的噪声点进行定义，并通过检测生成图中的孤立节点剔除噪声点。
- （希望）在带有噪声的图匹配问题中效果比已有方法要好（F1>60%）

未来可以提升的点

- 生成的匹配边是全连接图，感觉会也导致求解困难。要解决全连接图的求解复杂度：使用全连接边的图会增加求解空间的复杂度。因此需要先验策略减少生成的边缘。

计算梯度反传，当匹配矩阵的forward得到的值比较小的时候，我们可以认为他跟对应节点之间的关系较小，反传参数的值也比较小。可以直接删除对应的边，从而减少计算量。

目前的不足

- 未能进行显式的一一匹配约束，但是DGMC的工作表明不一定需要进行匹配的约束。

