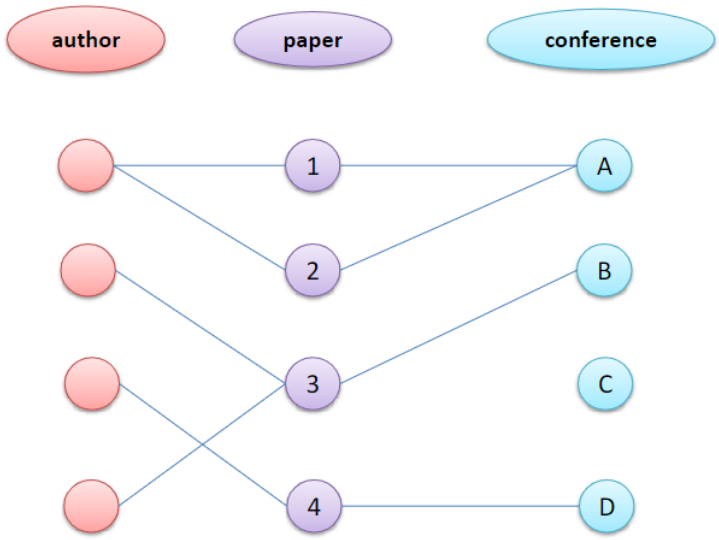
1. 训练使用的图中不能包含 paper 和 conference 的边（**上次会议提到**）

原始图如下，假定训练边集合{1-A，2-A}，测试边集合{3-B}，验证边集合{4-D}。

Original graph



为了用来训练的 Train Graph，首先，会在 original graph 中删除验证边和测试边。如果此时再删去训练集中的 paper-conference 边，conference 节点会变成完全孤立的节点，这使得它们好像不能在经过 HGT 的 embedding 后获得和原图其他节点有关的信息。注意，original graph 中只包含 paper-published on-conference 的单向边。如果将 original graph 中 paper 和 conference 的边设置为双向边，上述节点孤立的问题可以解决，但是是否会有信息泄露的问题。

2. 关于负采样（**上次会议提到**）

实验中需要通过采样得到不存在的边，以保证样本种类均衡。上次会议提到，负采样得到的边可能具有不同的分布，但是绝大多数源码都是随机的对不存在的边进行采样，分布不一致确实是一个问题。

3. 关于 embedding

实验中使用 HGT 模型来编码节点特征。在训练、验证、测试阶段，需要分别进行一次节点特征编码，还是只在训练时进行一次节点特征编码，即

a) h = model.forward(…)
   train_pred(h)
   val_pred(h)
   test_pred(h)

b) h = model.forward(…)　　train_pred(h)
   h1 = model.forward(…)　　val_pred(h1)
   h2 = model.forward(…)　　val_pred(h2)