内容概述：

1. COFs的高阶网络构造
   1. 阅读了几篇论文
   2. 构造了部分COFs的RACs描述符（构造过程中的中间结果可以用于构造COFs的高阶网络）

**具体内容：**

1. **COFs的高阶网络构造**
   1. **阅读了几篇论文**

**《Understanding the diversity of the metal-organic framework ecosystem》**

**内容：提出了使用几何描述符和化学描述符来描述MOFs数据库的多样性**，文章的做法是使用zeo++计算MOFs的几何描述符，使用RACs计算MOFs的化学描述符。使用两种描述符组合来表示MOFs是因为MOFs的气体吸附属性不仅受到MOFs的孔道结构的影响，还会受到MOFs的组成元素（金属团簇，有机连接体）的影响。为了验证这个观点，作者做了几组实验，使用一个简单的机器学习模型KRR预测气体吸附属性；主要包含两种气体——CO2和CH4；两种属性——高压气体吸附和低压气体吸附；分别使用不同的描述符——只包含几何描述符，几何描述符加化学描述符。实验结果表明，只有几何描述符的时候，在CH4的高压吸附任务上效果较好，包含化学描述符，在所有任务上都有了一定提升，说明气体吸附任务上，包含整体的几何特征和化学组成成分的特征更加重要。

**对我的作用：**

**文章的实验结果和我在cgcnn上所做的实验一样**，结果会发现，cgcnn在低压吸附任务上的结果还可以，但是高压吸附任务上的误差会非常差。这与cgcnn提取特征的方式有关，cgcnn本质上是在做一个原子嵌入，嵌入得到的原子特征，直接通过一个readout，对所有原子进行一个求平均操作得到整个晶体的表示，这样做会损失掉整体的几何特征。这种问题在图嵌入问题中也很常见，当需要处理一张比较大的图时，往往会需要先做几次池化处理，然后在readout，从而能够更好地提取整张图的结构。

文章提出了**对MOF进行拆分的方法**，比较简单，就是所有金属原子，然后剔除，剩下的部分就是有机连接体了。对MOF的拆分是构造高阶网络的关键，因为目前对于MOF，甚至对于COF的深度学习技术都是基于原子层面的，并没有分析高阶的拓扑信息（也就是底层网络），而为了得到底层网络，第一步需要做的事情就是对MOF，COF进行拆分，将拆完后的原子团簇缩小为一个超级顶点，这些超级顶点之间的拓扑就是我们所说的底层网络。这样，我们可以对MOF，COF构建两张图，**第一张是原子图，第二张是底层拓扑图**，这样可以从原子图中提取化学组成成分特征，从第二张图提取整体几何特征。从而解决cgcnn不能处理大体系晶体的问题。

**《ReDD-COFFEE》**

**内容：**基本和上一篇类似，不过做的是COF，文章提出了如何对COF进行拆分，因为COF和MOF不同，不包含金属，所以不能通过识别金属原子的方式来拆分，所以文章先对所有可能的连接位点进行枚举，然后构造模式，再对COF的图进行模式搜索，找到连接位点的位置，去掉这些原子后，就完成了对COF的拆分，拆分后不同组成部分对应于底层网络中的顶点或者是边。

**对我的作用：**文章提出了如何对COF进行拆分，搞明白了他的拆分逻辑后，我可以利用这种方法对任意一个COF进行拆分，从而可以构造COF的高阶网络。

**总结：上述的两篇论文对我的作用在于都提出了如何对MOF或者COF的拆分方法，并且这两种方法有一定的互通性，完成拆分后，我可以对这些原子进行聚类，从而构造粗粒度的周期图（也就是底层网络），从而使用拓扑神经网络学习特征。**

* 1. **构造了部分COFs的RACs描述符**

最后，根据**《ReDD-COFFEE》**提供的代码，修改完bug后，先计算了一遍core-cof的RAC描述符，目的是为了将来对比RAC和使用图网络提取的特征的效果，因为从目前多数论文的实验效果来看，图网络会达到更好的效果。

下周任务：

1. **《ReDD-COFFEE》**提供计算RAC的代码，相当于提供了拆分COF的代码，下周需要研究他的代码，抽取出对我有用的拆分COF的代码，从而写出自己的构造COFs高阶网络的代码。
2. 做一个实验，使用cgcnn学习化学组成特征，在cat一个几何描述符，看看对于气体吸附任务会不会有性能提升。如果有性能提升的话，说明了cgcnn对于整体几何特征的捕捉较弱。