内容概述：

1. COFs量化属性数据集
   1. 数据集简介（留个mark）
2. 多任务学习
   1. 结合homo、lumo和bandgap之间的关系进行多任务学习
3. 周期图结构
   1. 不同于分子的图结构，晶体具有周期性重复结构，所以需要更加一般化的数据结构

具体内容：

1. **COFs量化属性数据集**

**论文名称：**《In Search of Covalent Organic Framework Photocatalysts: A DFT-Based Screening Approach》

**论文内容：**通过DFT计算几个和光催化水分解相关的属性，作者将这些属性分为三类：energy-based descriptors；charge separation descriptor；charge-carrier mobility descriptor。首先在CURATED COFs数据集上基于PBE泛函进行了高通量计算，然后再利用PBE0泛函对结果进行修正。

**对我的作用：**论文提出了光催化水分解的计算方法，可以和吴优师兄在这个方法上讨论针对我们的问题设计计算方法以及目标属性的后处理脚本。第二点，作者提供了COFs的数据，其中包含，带隙、导带、价带等对于我们任务有用的目标属性，这些数据对于我们做机器学习预测器来说都是比较优质的数据。

1. **多任务学习**

**任务目标：**考虑到导带、价带和带隙的关系，导带-价带=带隙。所以在设计预测器的时候并不需要每个任务都训练一个预测器，相反，可以利用多任务学习的特征，这些任务可以共用一个骨架网络，这里暂时使用cgcnn，在cgcnn后接入多个预测器头，分别预测不同属性，可以设计loss为每个预测loss的平均，同时考虑到导带、价带和带隙的关系，可以设计一个正则项，即mse(导带-价带，带隙)，注入一定的化学先验。

**多任务学习结果：**目前只做了QMOF数据集上bandgap的训练，以及多任务学习的训练。目前发现多任务学习的收敛非常快，并且最终学习的效果也不算太差，但是实验并不算多，需要本周做更多的实验来说明多任务的效果。

1. **周期图结构**

相关论文：《CrystalNets.jl: Identification of Crystal Topologies》

内容：不同于分子的图表示，晶体是具有周期性，需要使用更加一般的数据结构，可以使用周期图来表示晶体，通过给边增加一个offset的信息来反映目标节点处于哪个晶格内。使用周期图来表示晶体，需要确定顶点和边分别表示什么，当晶体比较小的时候，可以将原子作为顶点，原子间的化学键作为边。但是当晶体比较大的时候，比如MOF、COF这些结构大小是纳米级别的，继续采用原子作为顶点，化学键作为边，将会是整体的图结构非常大，从而难以体现出整体结构的拓扑，所以对于MOF、COF这类纳米级晶体材料，一般将第二构建单元作为周期图的顶点，有时，甚至将第二构建单元作为周期图的边。

**对我的作用：**论文中提到了简化MOF、COF的结构来构建周期图，如果弄清楚作者的算法，可以在此基础上设计出MOF、COF的组合复形，并利用拓扑神经网络学习多孔材料的表示。

**本周安排：**

1. 理解周期图的构建方法，以及如何简化MOF、COF的周期图。
2. 多任务学习，再跑几组实验
   1. 三个目标单任务训练时的效果。
   2. 多任务训练时的效果。