数据集方面：

整理完了丹麦提供的COFs数据集，存在一个问题，单个COFs太大了，多达800个原子。数据的预处理最少需要大概200个小时，还只是将COFs的cif文件转换为图数据。

训练方面：

为了测一下训练所需要花费的时间，用了其中100个COFs做了一个缩小版数据集，跑一轮epoch大概需要1个小时。如果将26万个COFs都用来跑的话，一轮epoch需要将近100天。这个速度实在不能接受。

解决的方法：

1. 用多进程处理数据，服务器满载的情况下，大概200个小时。
2. 用MOFs数据来做一下实验，检验一下自监督模型的效果，然后再用到COFs上。MOFs大概有两万个比较好的数据，而且每个MOFs都比较小。可以用来做为自监督的训练数据。还有一万多个数据是包含带隙的，可以用做后面预测器的训练。

任务安排：

1. COFs整理数据：整理了丹麦那个26个COFs数据，还有北化工的40万个COFs，以及之前六万个COFs的数据还没有整理好。实验合成的COFs已经整理好了。
2. 自监督在MOFs数据集上的实验（由于COFs数据不够，先在MOFs做实验）
   1. 自监督的代码已经调通了，可以开始跑任务了。
   2. MOFs 数据整理，已经整理好了2万个MOFs，还需要整理一下虚拟合成的MOFs数据集（事情不紧急）
   3. 跑两个实验，一个是加了自监督的预测器，另一个是没有添加自监督的预测器。
3. 拓扑神经网络
   1. 利用跑MOFs实验的空余时间学习拓扑神经网络，跑他的demo（这块的任务目前还有具体的安排）