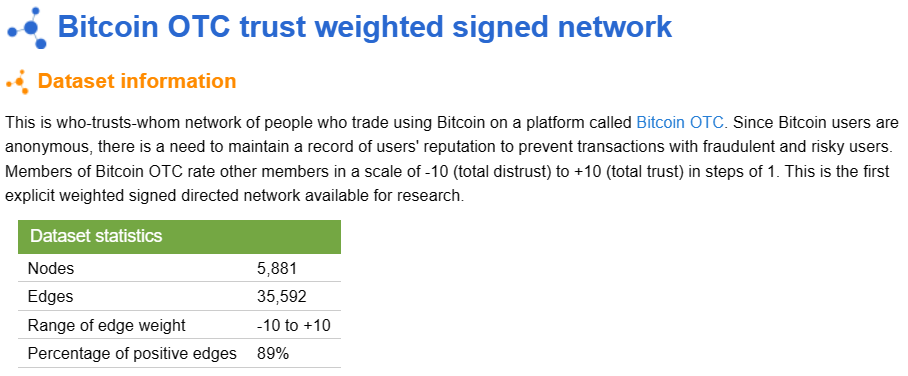
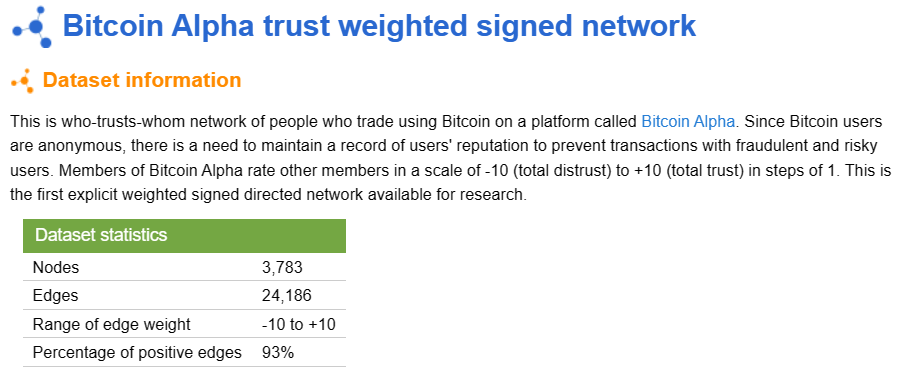
选择的数据集：SNAP(Stanford Large Network Dataset Collection):

<http://snap.stanford.edu/data/index.html>

[SNAP: Signed network datasets: Bitcoin OTC web of trust network (stanford.edu)](https://snap.stanford.edu/data/soc-sign-bitcoin-otc.html)



[SNAP: Signed network datasets: Bitcoin Alpha web of trust network (stanford.edu)](https://snap.stanford.edu/data/soc-sign-bitcoin-alpha.html)



两个均为图数据，带权，有向。实际意义是比特币交易记录，所以频繁模式为图数据中出现的频繁子图，即现实中具有较高信用的比特币交易社区，或者信用较低用户群的不良交易记录（对应原数据中边权值为负值的交易记录）

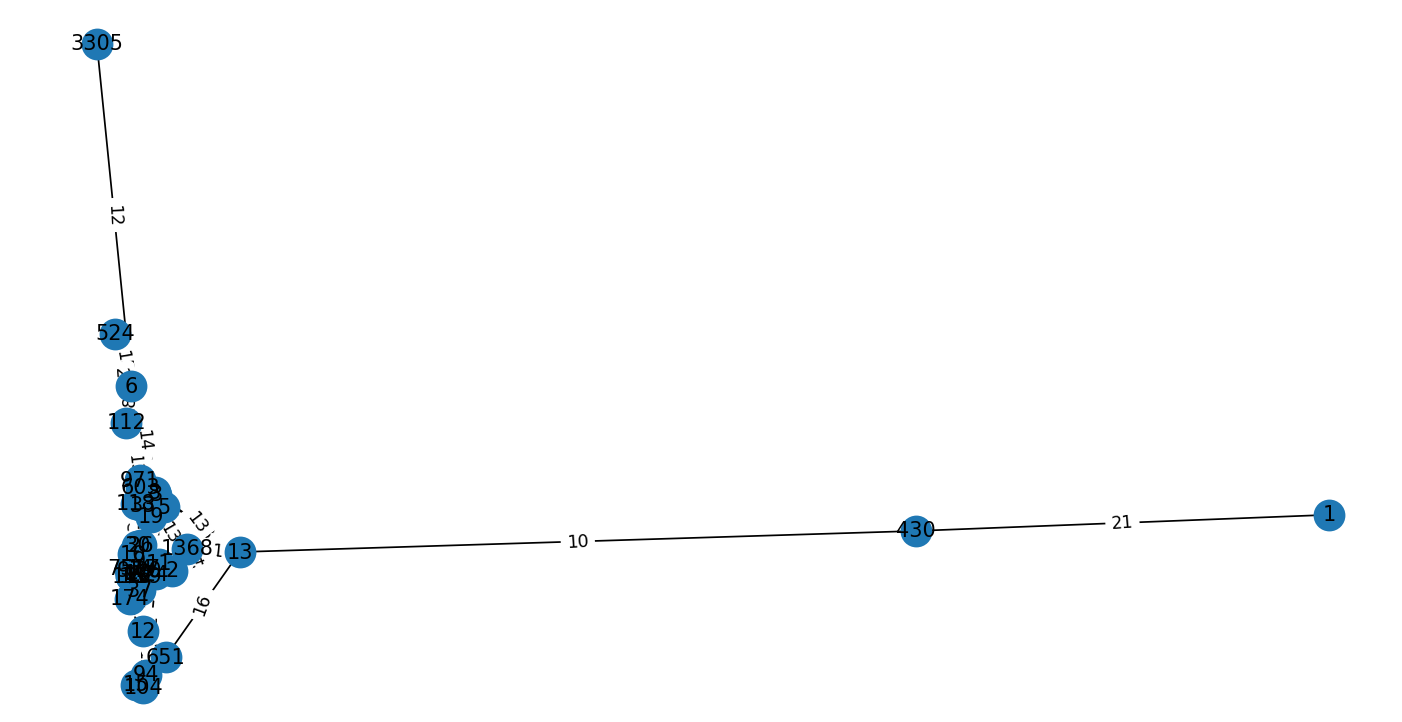
gaspan算法：首先使用DFS对图中节点生成编码。然后计算子图的支持度，对于每一个生成的子图模式，gspan计算其在整个图数据库中出现的频率，即支持度。算法还加入了子图扩展和剪枝优化，来提高频繁子图的挖掘效率

挖掘结果：

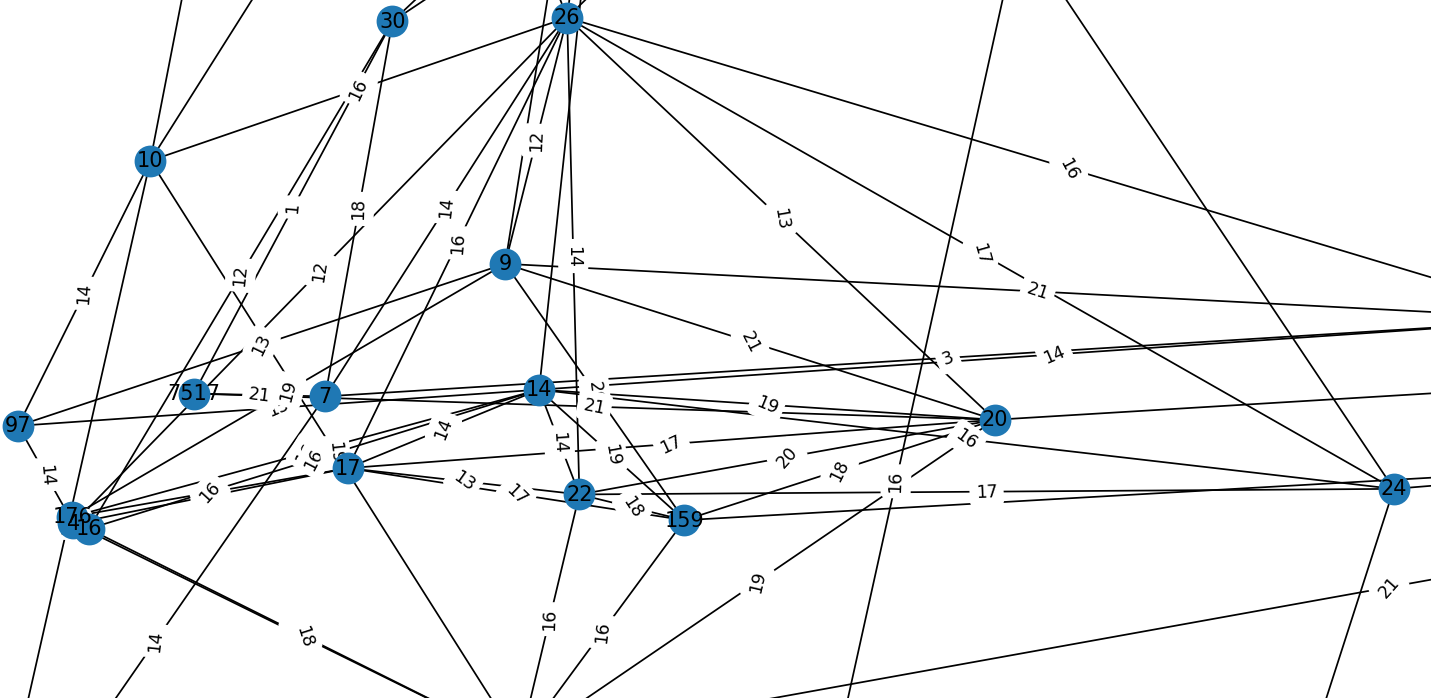
由于时间关系，未能计算出所有结果，部分结果可在data/result\_\*.data中查询。虽然只有部分数据，但通过对原数据的分析和其实际意义的理解，同样可以发现一些规律，较小id的用户，其注册时间较早，交易记录也比较可靠，更容易出现在频繁子图中（可通过观察发现其边的权值通常比较大），而较大id的用户注册时间晚，一些不可靠的交易会发生在这一用户群体中。

可视化分析：

这里选择某一频繁子图进行分析。



从图中可以看出3305和430这两个用户交易记录较少，且边的权值不是很高，说明其在用户群体中信用度还不算高，而430与1之间有权重为21（最高值）的边，这可能说明，430和1号用户之间可能彼此相识，或者这次交易记录两者早已计划好。



上图为中心密集区域，可以看出id较小的节点间边的权值都比较高，验证了刚刚的分析。图中7517和30节点的边权值为1（最小值），可能是两个用户之一遭到了对方的诈骗。