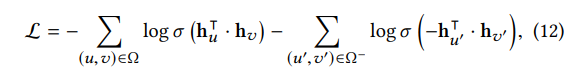
3.6 Train

~~通过上面提到的方法，我们最终得到了节点表示，它能够被用于很多下游任务。~~由于我们的数据是没有标签的，因此我们采用无监督的学习方法。对于无监督的学习，是没有任何标签的，我们优化模型的权重通过最小化以下的损失函数通过负采样【Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In NIPS. 3111–3119.】：

~~Through the methods mentioned above, we finally get the node representation, which can be used for different downstream tasks.~~ Because our data has not labels, we can train HO-GCN with unsupervised learning method. For unsupervised learning, without any node labels. We optimize the weight of the model by minimizing the following loss function through negative sampling



where σ(·) is the sigmoid function, Ω is the set of positive node pairs, Ω − is the set of negative node pairs(the complement of Ω)

4.4

具体的，我们将存在链接的节点对看作是正节点对，将不存在链接的边看作是负节点对。我们从原始数据中随机采样出节点组成负节点对，它的数量是与正节点对近似相等的，并按照相同的比例分为训练数据、验证数据和测试数据。GNNs模型是通过最小化公式XXX中的目标函数进行优化。

Specifically, we regard a pair of nodes that have a connection as a pair of positive nodes, and a pair of nodes that do not have a connection as a pair of negative nodes. We randomly sampled nodes from the original data to form negative node pairs, the number of which is approximately equal to the positive node pairs, and is divided into training data, verification data and test data in the same proportion. The HO-GNNs model is optimized by minimizing the objective function in formula XXX.

利用训练的模型我们会得到一对节点的嵌入embedding **h**u and the embedding **h**v, 计算节点之间存在链接的概率，公式如下：

Given embedding **h**v and the embedding **h**v generated by the trained model, Calculate the probability of links between nodes, the formula is as follows:【去掉激活函数】



对于正节点对，节点**h**u， **hv**嵌入是高阶网络中的节点，每个节点都会对应一个原始网络中的节点，即从生成的高阶网络到原始网络存在一个多对一的映射关系。因此，在高阶网络中存在多个节点对表达了原始网络中同一个节点对。在计算任意高级网络中节点对存在链接的概率时，如果这些节点对表达原始网络中同一个节点对，我们将其取平均值作为最终的链接概率。对于负节点对，从原始网络到高阶网络也存在一对多的关系，但是由于所有负节点对都是随机生成的，无法按照正节点对那样通过边的标签进行聚合，从而计算链接概率，因此我们直接将高阶网络中的负节点对的特征取平均，然后计算链接概率。我们通过area under the ROC curve (AUC) 和 f1-score值评估我们的模型用于链接预测的效果。

For a positive node pair, the node hu, hv embedding is a node in a higher-order network, and each node corresponds to a node in the original network, that is, there is a many-to-one mapping relationship from the generated higher-order network to the original network. When calculating the link probability of node pairs in any high-level network, if these node pairs represent the same node pair in the original network, we take the average value as the final link probability. For negative node pairs, there is also a one-to-many relationship from the original network to the higher-order network. However, since all negative node pairs are randomly generated, it is impossible to aggregate through the labels of the edges as the positive node pairs, so as to calculate the link probability. Therefore, we directly average the features of the negative node pairs in the high-order network, and then calculate the link probability. We use area under the ROC curve (AUC) and f1-score to evaluate the effectiveness of our model for link prediction.

结果被展示在表格XXX1和表格XXX2中，其中表格XXX1展示了在进行链接预测时的AUC值，表格XXX2展示了在进行链接预测时的f1-score值。我们在traces-100 、traces-1000 click-stream三个数据集上进行了对比试验，并且针对不同的数据集分别采用节点的度和度的one-hot形式作为节点的属性进行传播聚合。就像我们看到的，HO-GCN和HOA-GCN在两个标准上都具有很好的表现。GCN在每个数据集上的的表现很差，与GCN相比我们的HO-GCN和HOA-GCN表现有大幅度提升。这个结果支持了我们提出的高阶网络中存在的依赖对于高阶网络的表征是很重要的。

The results are displayed in Table XXX1 and Table XXX2, where Table XXX1 shows the AUC value during link prediction, and Table XXX2 shows the f1-score value during link prediction. We conducted comparative experiments on the three data sets of traces-100 and traces-1000 click-stream, and used the degree of the node and the one-hot form of the degree as the attributes of the node for propagation and aggregation for different data sets. As we have seen, HO-GCN and HOA-GCN have good performance on both standards. The performance of GCN on each data set is very poor. Compared with GCN, the performance of our HO-GCN and HOA-GCN has been greatly improved. This result supports that the dependence in our proposed high-order network is very important for the characterization of high-order networks.