关于论文《Modularity in complex multilayer networks with multiple aspects: a static perspective》的学习报告

多层网络的表示模型：对于复杂的网络系统，多层网络的所有层可以进行分组，意味着这些层中的某些层是从不同的aspect进行观测与表示的。因此多层网络可以表示成aspect-aspect的形式。然而，每增加一个aspect，就会带来数量不一定的layers，从而使得表示的矩阵的大小很难预先确定。因此出现了第二种多层网络的表示方式——aspect-layer的表示方式，这种表示方式实际上是把具有层次性的多个aspect合并成同一个aspect，而这些原来的aspect就降级成layer。

多层网络进行社区探测的衡量标准：目前，使用得比较多的衡量标准是模块性。

$$Q = \frac{1}{4e}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2e})h_i h_j,$$

模块性衡量的是一个社区的划分效果情况，Q值越大表示社区的划分效果越好。从公式上解释，A_{ij}是原来存储网络信息的邻接矩阵，而k_i是i节点的度，e是边的总数，因此2e就是整个网络图中的总度数，而$\frac{k_i k_j}{2e}$表示的是当网络中的边是任意添加的时候节点ij之间期望出现的边数。因此，这个公式衡量的是当前图划分的实际情况与理想的划分情况之间的差别，当实际划分情况越接近理想的划分情况时，表明社区的划分效果越好。Mucha将这一概念延展到了多层网络中，通过比较随机漫步模型在时间t跟静态情况下停留在同一社区的可能性，从而测量社区的稳定性，这是一个动态过程，涉及了随机过程，意味着相同的节点在任意层的网络之间都能传递信息。而对于时序网络，只有相邻层（时间片）之间是存在相互关联的，并不是任意层都存在联系，因此节点在任意层中间传递是不合适的，因此Mucha的方法有缺陷。这一缺陷后来在时序网络的拓展方面提出了解决方法。（这里对论文中提出的动态方面以及解决措施的理解不知是否正确？）

一个通用的衡量社区划分的质量函数：

$$\begin{aligned}\mathcal{H}(g) = &-\sum_{i\neq j} a_{ij} \underbrace{A_{ij}\delta(g_i,g_j)}_{\text{Internal existing edges}} + \sum_{i\neq j} b_{ij} \underbrace{(1-A_{ij})\delta(g_i,g_j)}_{\text{Internal non-existing edges}} \\ &+ \sum_{i\neq j} c_{ij} \underbrace{A_{ij}\left[1-\delta(g_i,g_j)\right]}_{\text{External existing edges}} - \sum_{i\neq j} d_{ij} \underbrace{(1-A_{ij})\left[1-\delta(g_i,g_j)\right]}_{\text{External non-existing edges}},\end{aligned} \tag{1}$$

从公式上进行初步解释这个质量函数的意义：考虑社区划分的一个准则：假如网络中的某两个点之间存在一条边，表明这两个节点存在一定的关系，那么这两个节点属于同一个社区的可能性就远大于不存在边连接的两个节点。这个质量函数由四部分组成，分别对应着奖励被划分到同一个社区的存在边的两个节点；惩罚被划分到同一个社区的不存在边的两个节点；惩罚被划分到不同社区的存在边的两个节点；奖励被划分到不同社区的不存在边的两个节点。

以上公式是针对单层网络的情况。在论文中，作者为了将该公式应用到多层网络中，引入了不同层之间的耦合。引入了层间耦合的质量函数在考虑了原来同层网络的节点之

间的社区划分情况以后，也考虑了同一节点在不同层网络里面是否被划分到同一个社区中，这一点扩展的现实意义是该节点在不同层的网络中是否扮演了相同的角色。对最终的多层网络的质量函数自由选择参数，得到以下汉密尔顿形式函数：

$$\mathcal{H}_M(g) = -\sum_{ijsv}\left(A_{ijs}^{\{v\}} - \gamma_s^{\{v\}} p_{ijs}^{\{v\}}\right)\left[2\delta\left(g_{is}^{\{v\}}, g_{js}^{\{v\}}\right) - 1\right]$$
$$- \sum_{isrvw} e_{isr}^{\{vw\}}\left[2\delta\left(g_{is}^{\{v\}}, g_{ir}^{\{w\}}\right) - 1\right]\left(2C_{isr}^{\{vw\}} - 1\right). \tag{3}$$

通过一系列的分析，将公式分成与社区划分相关的部分，以及仅跟网络架构相关的部分。最终推导出最优化哈密尔顿函数就相当于优化网络的modularity：

$$Q_M(g) = \sum_{isjrvw}\left[\left(A_{ijs}^{\{v\}} - \gamma_s^{\{v\}} p_{ijs}^{\{v\}}\right)\delta_{sr}\delta_{vw} + \tilde{C}_{isr}^{\{vw\}}\delta_{ij}\right]\delta\left(g_{is}^{\{v\}}, g_{jr}^{\{w\}}\right). \tag{6}$$

这一目标函数与《Community Structure in Time-Dependent, Multiscale, and Multiplex Networks》中的目标函数下相似，但是在该论文中没有讨论不同网络之间的耦合强度。作者认为，不同层网络之间的耦合强度应该根据网络的任务而定。在论文中，针对两类典型网络，作者提出了不同的耦合强度测量方法。

1. 针对具有多个非均衡分布层的网络，在这种情况下，层之间的耦合强度应该通过网络层的相似程度衡量，因此得到耦合强度的幅度公式为：

$$e_{isr}^{\{vw\}} = \omega \cdot \frac{M_{sr}^{\{vw\}}}{\max_{s,r,v,w} M_{sr}^{\{vw\}}},$$

2. 针对时序网络，不同的层代表网络演化过程中在不同时间点网络情况。在这种情况下并不是所有层之间都存在联系的，尤其是针对不同的情况，相邻时间片的网络的同一个节点之间也可能不存在关系。因此考虑以上提出的情况，作者在论文中定义不同层之间某个点的耦合强度的幅度为：当这两层网络不在相邻的时间点，或即使在两个相邻的时间点，假如同一个节点在这两个时间点之间不存在持续联系，则耦合强度的幅度为0。

mSpec优化方法：对于单层网络，有人提出了一种谱方法用来求解单层网络的模块性最优化问题。受这种方法的启发，作者在论文中提出了针对多层网络的模块性最优化问题提出了mSpec方法。这种方法首先把多层网络表示成超毗邻矩阵的表示形式（Supra-adjacency representation），所谓Supra-adjacency representation就是把网络的多层表示整合成一个单层网络，具体方法就是：对于一个2层的网络，这个网络是通过2个N*N的图以矩阵的形式表示，mSpec把这两个图扩展成一个2*2的块矩阵，其中对角线的块表示原来两个层中的连接程度；而非对角线的块表示不同层之间的耦合程度。采用的这种表示方法实际上就是把同一个在不同层中出现的点变成两个单独的点，而这些属于x点的copy的点在跨层的连通中只会与x点存在耦合关系。通过这样一种表示方法，就能应用原始的Spectral方法对单层网络的模块性优化问题进行求解。

当希望将原始的多层网络划分为多个社区时，采用的是基于模块性增益的子社区划分算法。具体操作是对于一个社区C，对应的模块性为M1，假设将这个社区划分为两个子社区B,C，对应的模块性为M2和M3，则划分后的模块性增益就变成了M2+M3-M1。我们的目标是找到使原始网络的模块性最大的划分方法，因此对于一个社区，问题就变成使这个社区被划分后得到的模块性更大，即使得该社区划分的模块性增益最大。而这个

问题和使模块性最大的本质是一样的，因此可以通过Spectral方法进行求解。