模块度最大化

1、综述:

本篇文章首先介绍了社团结构,模块度的概念.

随后介绍了单层网络和多层网络模块度最大化的算法.之后使用了一个数据集测试了多层网络模块度最大化的方法.并将单层网络方法用于多层网络的每一层得到结果.随后进行了效果对比.

最后介绍了模块度最大化近年来的进展.

2、社团结构:

2.1定义：

如果一个网络中的节点能够分组,组内节点联系紧密,就称该网络具有社团结构.

2.2特点：

网络中的社团内部连接紧密，常对应某种特点或功能，而社团外部的连接则相对稀疏，即“内紧外松”.

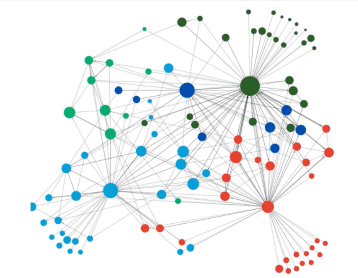
2.3可视化：

图1.1.3中相同颜色的点属于同一个社团

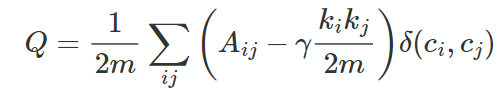
图1.1.3

3、模块度:

3.1用途:

衡量网络中节点分组质量,即判断社团划分的效果.越大越好.

3.2计算公式:

1

M:图中边数 A:邻接矩阵 ki:点的度数 r:解参数(可变,通常为1) :(1:i,j属于同一社团,否则为0)

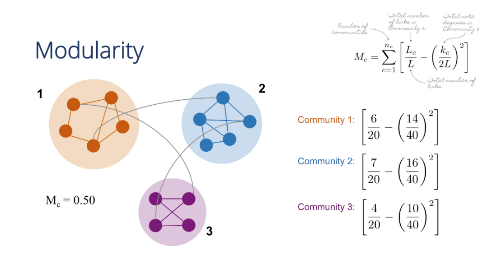
3.3示例:

图3.3.1

4、单层网络模块度最大化:

4.1用途:

从所有可能的分组找到使模块度最大的分组,完成社团检测.

4.2 NP-Hard

当Modularity Max问题指的是求使得模块度最大的社团划分方案的时候，它是一个优化问题，即一个NP-hard问题，既不是P问题，也不是NP问题.

根据[1]，3-PARTITION问题可以被多项式时间归约为求一个构建图的使得模块度最大的社团划分方案，而3-PARTITION是一个NPC问题，故而Modularity Max问题是一个至少比NPC问题要难的问题，故是NP-hard问题.其中，图的构建方式如下：

对于一个3-PARTITION问题，对于K，构建K个完全子图，对于每个Ai，构建一个元素点，使得该点与每个子图中的Ai个节点相连，在[1]中证明了使得模块度最大的社团划分必然有k个社团，每个社团一个完全子图，每个社团3个元素节点，这就形成了一个3-PARTITION问题的解.

图示

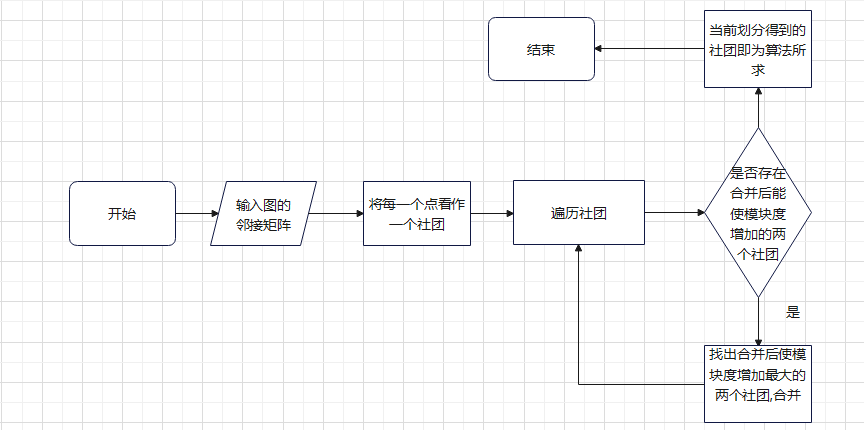
描述已自动生成

图4.2.1

4.3常见算法:

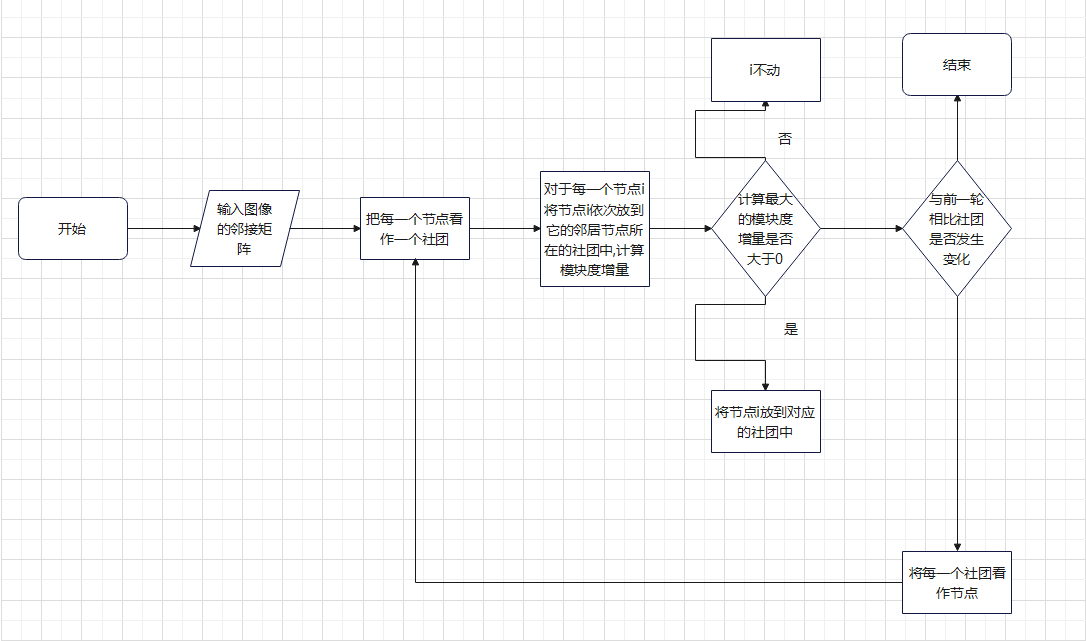
4.3.1贪心算法一[2]:

算法流程



4.3.2贪心算法二(Louvain算法)[3]:

算法流程

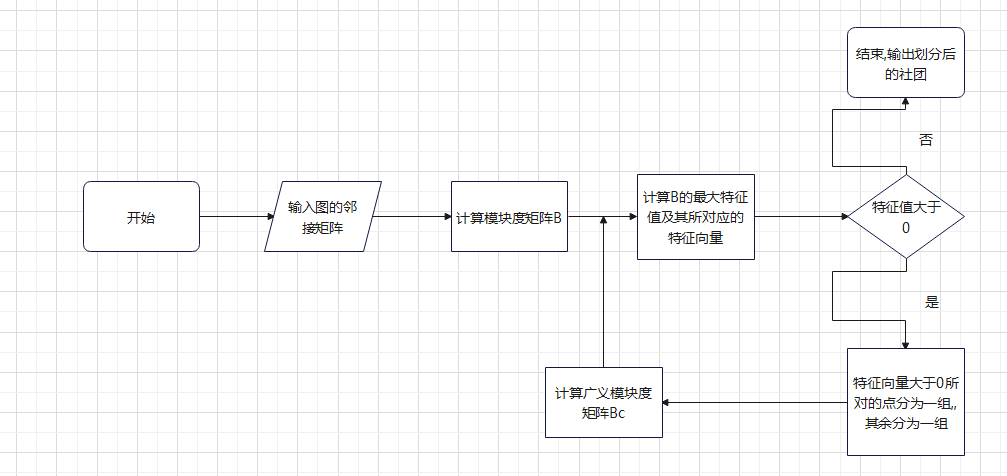


4.2.3、谱算法:[4]

广义上来说，任何在算法中用到SVD/特征值分解的，都叫谱算法.

算法使用了拉格朗日乘数法

算法流程



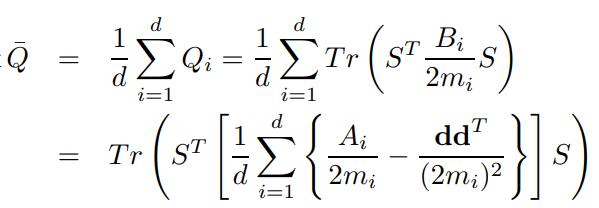
5、多层网络模块度最大化

5.1、研究目的:

群体之间的关系无法用一个网络来表示,比如用餐,社交软件,工作,娱乐,共同作者关系.发现在多种关系下的社团结构.

5.2研究方法:

5.2.1PMM算法[5]

5.2.1.1定义

为模块度

1 如果点i属于社团j

Sij(n\*k)=

0 其它

此方法是单层网络的一个推广,目的是寻找一个划分,使得所有层网络模块度之和的平均值最小.

5.2.1.2困难:

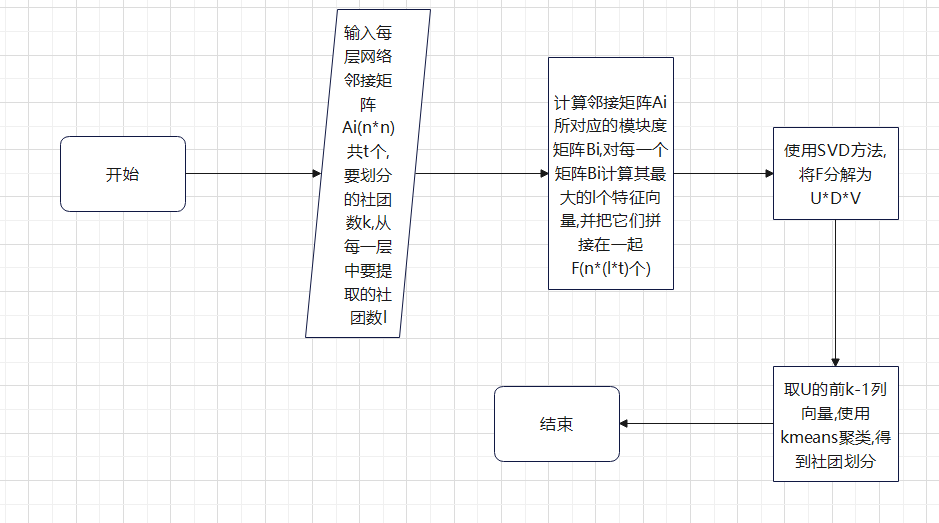
S的离散性使得模块度最大化对单层网络来说是一个NP-hard问题.但假使S连续,最优的S可以通过计算模块度B的最大的k个特征向量得到.

5.2.1.3灵感来源:

1、模块度矩阵特征向量可以看作网络的重要结构特征.

2、SVD方法中可以用最大的k个的奇异值和对应的左右奇异向量来近似描述矩阵.

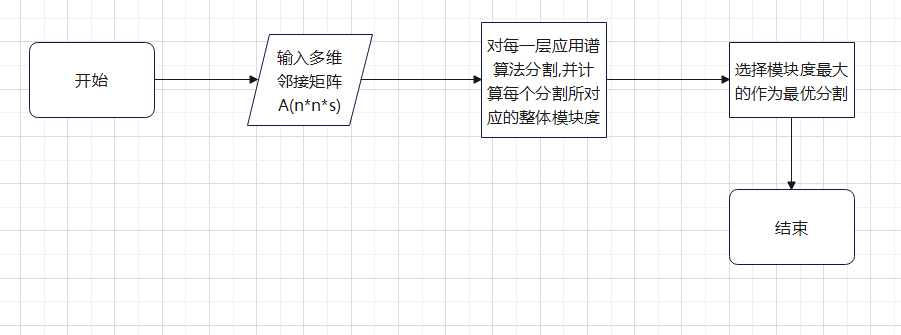
5.2.1.4算法流程:



5.2.2谱方法二

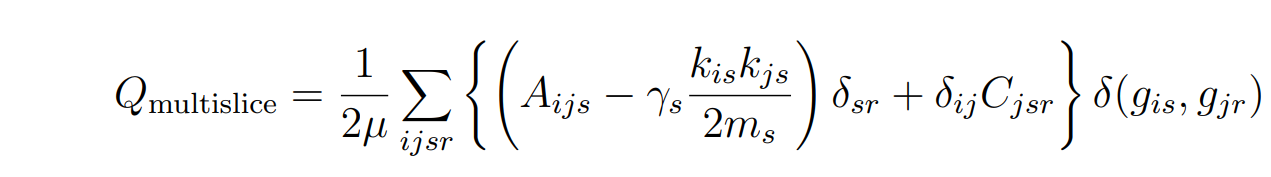
5.2.2.1算法流程:

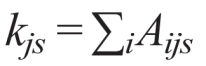
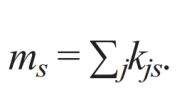
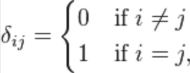
对每一层应用谱方法[6]



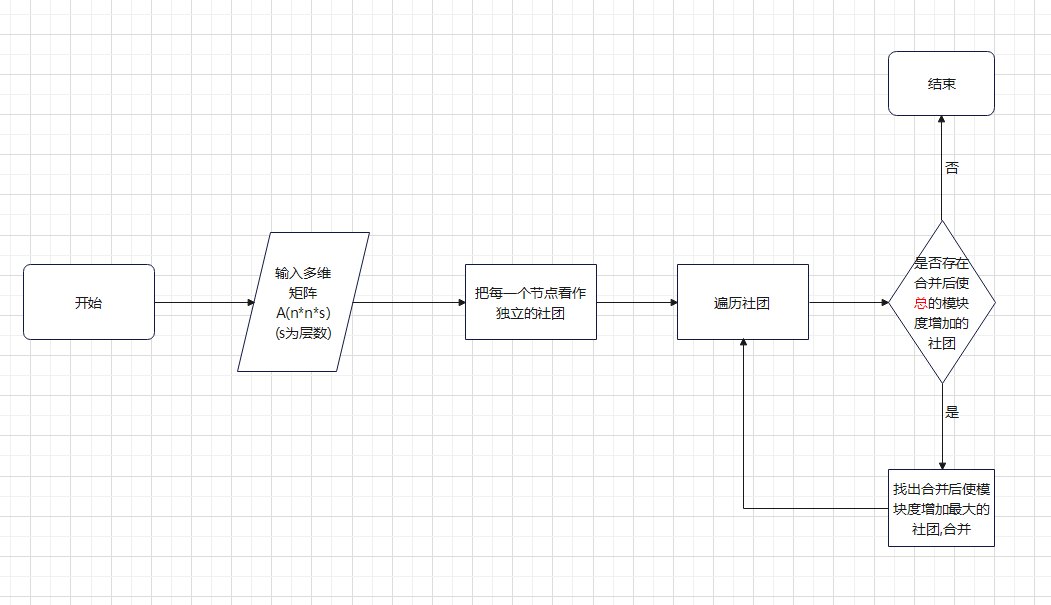
5.2.3贪婪算法(对单层网络贪婪算法的推广):

5.2.3.1模块度[7]:



u 为常数,A代表多层邻接矩阵, ,C层间耦合矩阵,取值为(Cjrs=1如果第s层的节点j与第r层节点相连否则为0) gis为第s层节点i所在的社团

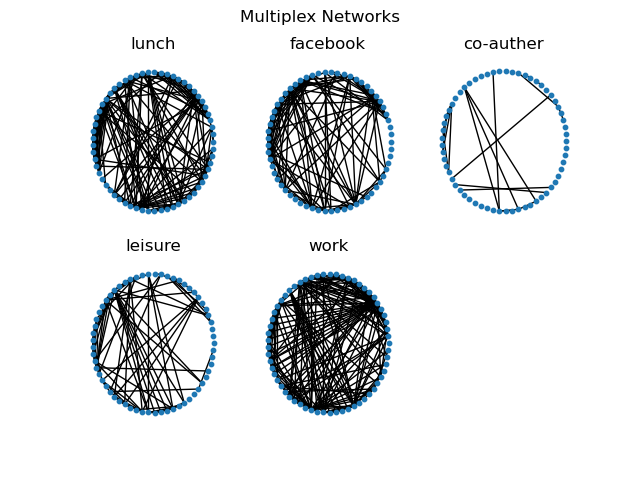
5.2.3.2:算法流程:



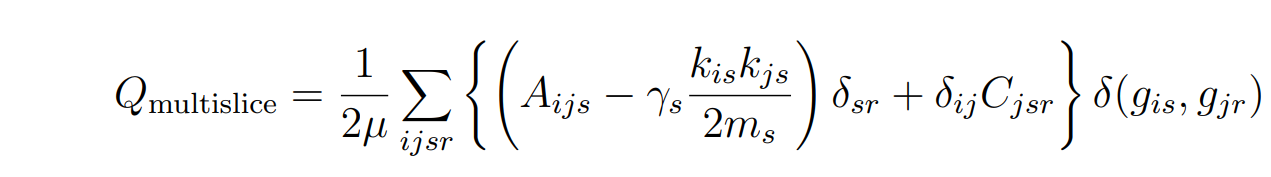
6、使用数据集对代码进行测试,并对比结果:

6.1数据集[8]介绍:

五种线上线下关系对应五层网络:facebook,Leisure,Work,Co-authorship,Lunch.



6.2衡量标准:



其中rs设置为1

6.3结果:

1,使用贪婪算法二得到的模块度: 0.3036

2使用广义louvain算法[9]得到模块度: 0.2975

3使用PMM算法得到模块度: 0.2859

4对每一层使用谱算法得到模块度: 0.2697

(分组见附件)

对每一层使用谱算法得到的效果最差,原因分析:虽然对五层分别使用了谱算法并从中选取了最好的划分,然而此划分并没有充分使用所有层的数据,其余方法都使用了.

6.4改进:

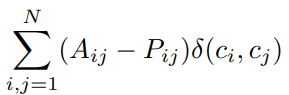
显然,不同层对于社团划分的重要性不同.比如共同co-authorship这一关系,明显强于facebook关系.后续考虑改进模块度公式对每一层分配合理权重重新计算模块度进行划分

7、近年进展:

7.1 Bazzi M这篇论文[10]讨论了不同的空网络和层间边在多层网络模块度最大化的影响

7.1.1空网络(null network):

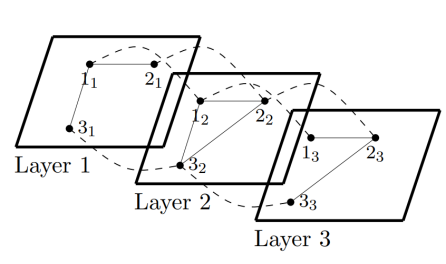
基于原网络随机生成的网络,并非真实存在.我们可以使用它的邻接矩阵来和我们划分得到的社团相比较.



Pij 即为空网络的邻接矩阵

7.1.2层间边:

层与层之间的边(图中虚线)

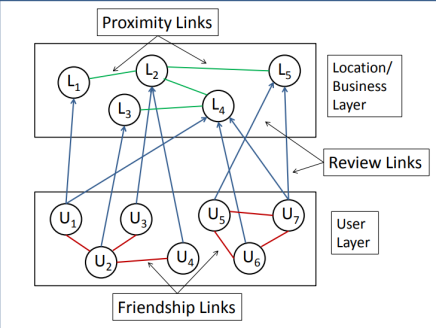


.

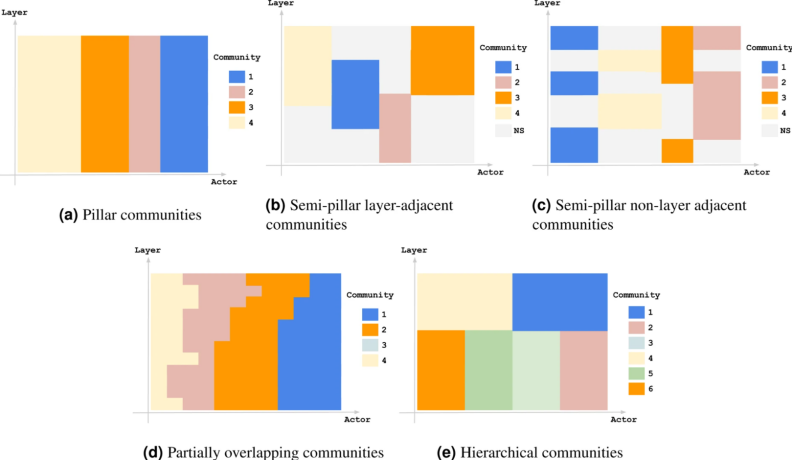
7.2 Asep等人[11]把多层网络模块度最大化看作是一个多目标优化求解

7.3 Pramanik S等人[12]论文中提出的方法是无参数的,多尺度并且可以检测出只包含一种类型节点的社团和包含多种类型节点的社团

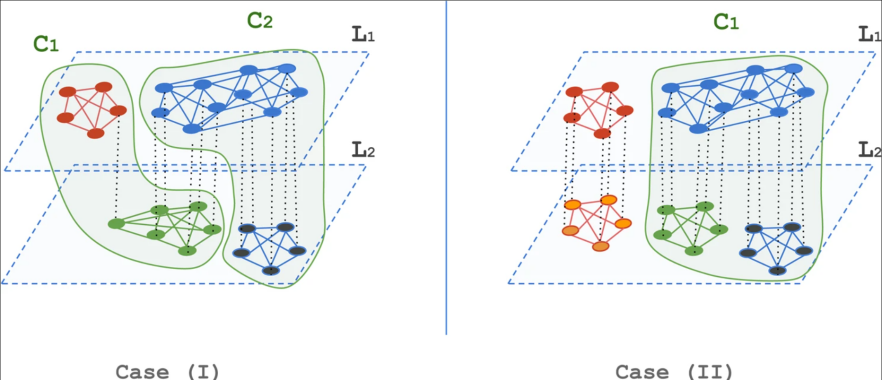
7.3.1不同类型节点:

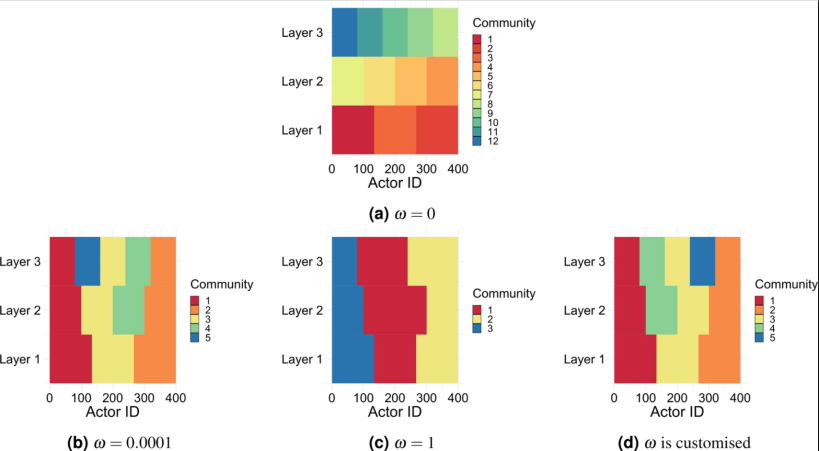


7.4 Obaida Hanteer等人[13]向模块化函数引入了一个新参数耦合强度ω，并提出关于基于模块化最大化的方法的局限性的讨论.

新提出的参数ω假设不同层中相同actor的节点耦合边的权重等于ω.通过该扩展，模块化最大化方法不仅可以最大化社区内层内边缘并最小化跨社区层内边缘，还可以最大化耦合边缘权重的总和.  
  
图7.4.1  
（不同模型的多层群落图示）

从实验中得出结论，论文提出的多层群落中只有少数几个简单模型可以通过模块化最大化方法恢复，而更复杂的模型在任何ω的调谐下都无法准确恢复.

  
图7.4.2  
（通过模块度最大化来恢复两种类型的社区图示）

  
图7.4.3  
（通过模块度最大化来恢复多层社区图示.模块化最大化的准确性似乎会随着模型复杂度提升而而下降.）

8、参考文献

[1]U. Brandes et al., "On Modularity Clustering," in IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 2, pp. 172-188, Feb. 2008, doi: 10.1109/TKDE.2007.190689.

[2][networkx.algorithms.community.quality.modularity — NetworkX 2.6.2 documentation](https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.community.quality.modularity.html?highlight=modularity" \l "networkx.algorithms.community.quality.modularity)

[3] [Louvain method - Wikipedia](https://en.wikipedia.org/wiki/Louvain_method)

[4]M.E.J.Newman[美]郭世泽,陈哲译 网络科学引论.北京:电子工业出版社

[5] Tang L, Wang X, Liu H (2009) Uncoverning groups via heterogeneous interaction analysis. In: 2009 Ninth IEEE international conference on data mining. IEEE, pp 503–512

[6] <https://zhiyzuo.github.io/python-modularity-maximization/>

[7] Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP (2010) Community structure in time-dependent, multiscale, and multiplex networks. Science 328(5980):876–878

[8]“Combinatorial Analysis of Multiple Networks” Matteo Magnani, Barbora Micenkova, and Luca Rossi arXiv:1303.4986 (2013)

[9]Lucas G. S. Jeub, Marya Bazzi, Inderjit S. Jutla, and Peter J. Mucha,      
 \*"A generalized Louvain method for community detection implemented  
    in MATLAB,"\* https://github.com/GenLouvain/GenLouvain (2011-2019).

[10] Bazzi M, Porter MA, Williams S, McDonald M, Fenn DJ, Howison SD (2016) Community detection in temporal multilayer networks, with an application to correlation networks. Multiscale Model Simul 14(1):1–41(2016)

[11] Modularities Maximization in Multiplex Network Analysis Using Many-Objective Optimization Asep Maulana Valerio Gametto Diego Garlaschelli Iryna Yevesyeva Michael Emmerich

[12] Pramanik S, Tackx R, Navelkar A, Guillaume JL, Mitra B (2017) Discovering community structure in multilayer networks. In: 2017 IEEE international conference on data science and advanced analytics (DSAA). IEEE, pp 611–620(2017)

[13] Obaida Hanteer and Matteo Magnani. Unspoken Assumptions in Multilayer Modularity maximization. Scientific Reports, 10(1):11053, 2020.