



西安电子科技大学
XIDIAN UNIVERSITY

复杂网络与群体智能

吴建设

Email: jshwu@mail.xidian.edu.cn

西安电子科技大学人工智能学院

课件下载地址:

<https://web.xidian.edu.cn/jshwu/teach.html>



- (1) 复杂网络社区检测简介
- (2) 社区检测方法



➤ 社区检测:

在具有**社区结构**的网络中，节点呈社区分布，同一社区内部节点连边密集，不同社区之间节点连边稀疏。

社区检测，就是将网络中的社区找出来，也称社区发现、社区挖掘。

➤ 社区检测算法的分类:

- (1) 层次聚类算法：包括凝聚算法和分裂算法
- (2) 基于模块度优化的社区检测算法
- (3) 基于网络动力学的社区检测算法
- (4) 谱聚类法
- (5) 基于学习的方法：图神经网络



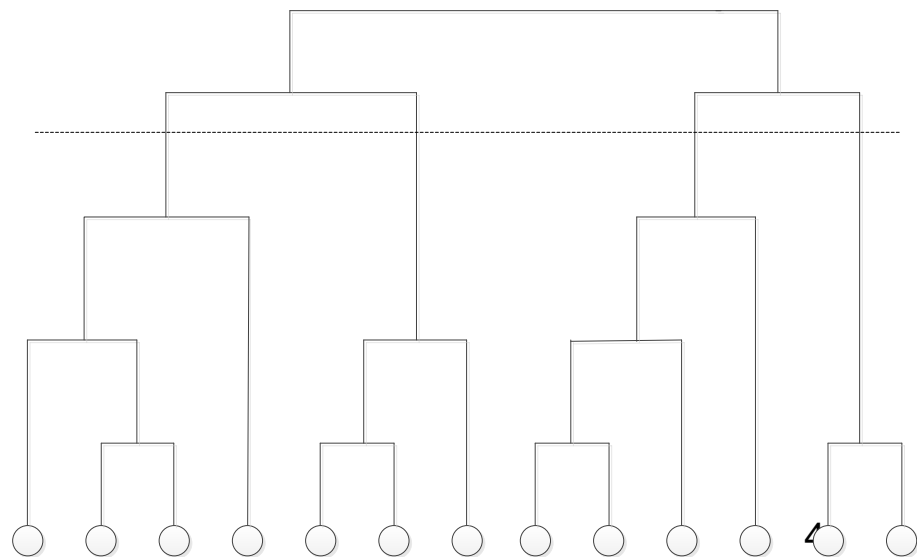
➤ 社区检测算法的分类:

(1) 层次聚类算法: 包括凝聚算法和分裂算法

Girvan和Newman于2002年提出了GN算法是典型的分裂算法。该算法提出了边介数的概念, 用于衡量网络中某条边的重要程度。边介数定义为经过该边的最短路径数目。GN算法通过不断删除边介数最大的连边来划分社区。

缺点:

1. 不容易确定从哪里停止删边; 即不容易确定最终的社区划分;
2. 由于删除边介数最大的连边后需要重新计算网络中所有节点的边介数, 且GN算法不断删除连边, 直到所有网络节点都各自成为一个社区时才能得到树状图, 因此该算法的时间复杂度较高, 为 $O(N^3)$ 。



层次聚类算法中的树状图



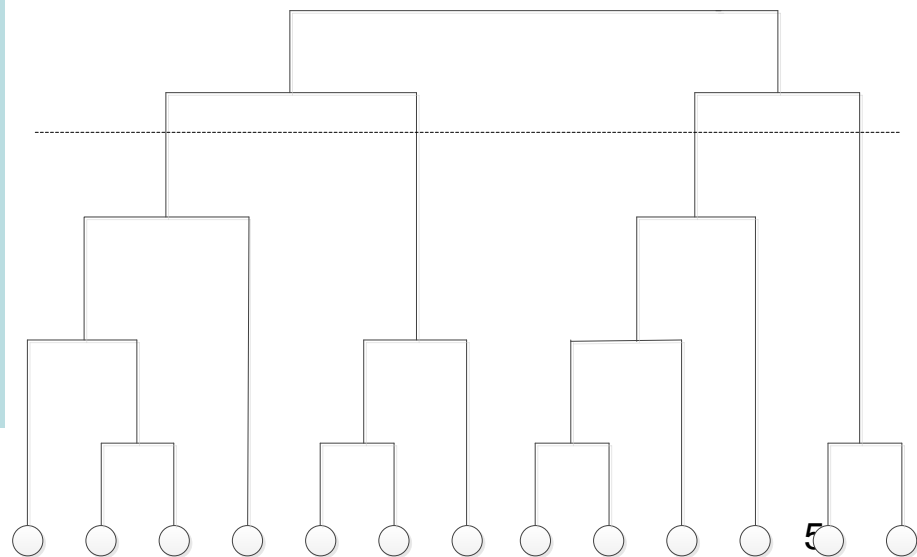
➤ 社区检测算法的分类:

(1) 层次聚类算法: 包括凝聚算法和分裂算法

Girvan和Newman于2002年提出了GN算法是典型的分裂算法。该算法提出了边介数的概念, 用于衡量网络中某条边的重要程度。边介数定义为经过该边的最短路径数目。GN算法通过不断删除边介数最大的连边来划分社区。

改进:

Fortunato等提出了信息中心度指标, 以及Zhou等定义的网络中连边的相异性参数。用这些评价指标取代边介数进行社区检测, 大大缩短了GN算法的运行时间



层次聚类算法中的树状图



➤ 社区检测算法的分类:

(2) 基于模块度优化的社区检测算法

Newman于2004年提出了纽曼快速算法FA。FA算法是基于凝聚的算法，属于模块度优化算法的一种。FA算法不断合并模块度函数值增加最多的两个社区。该社区合并的过程可表示为一个树状图，树状图中使得模块度函数 Q 最大的层次被用来划分社区。FA算法的时间复杂度为 $O(MN)$ 。

模块度函数 Q 还可以用公式表示为:

$$Q = \sum_{i=1}^C \left[\frac{L_i}{M} - \left(\frac{d_i}{2M} \right)^2 \right], \quad (4-9)$$

L_i : 社区内的连边总数, M : 网络中的连边总数。 i 表示社区而非节点, d_i 表示社区 i 内所有节点的度的总和。公式的第一项 L_i/M 表示社区内连边数与网络总边数的比值; 第二项 $d_i/2M$ 表示随机网络中, 社区内连边概率的期望值。



➤ 社区检测算法的分类:

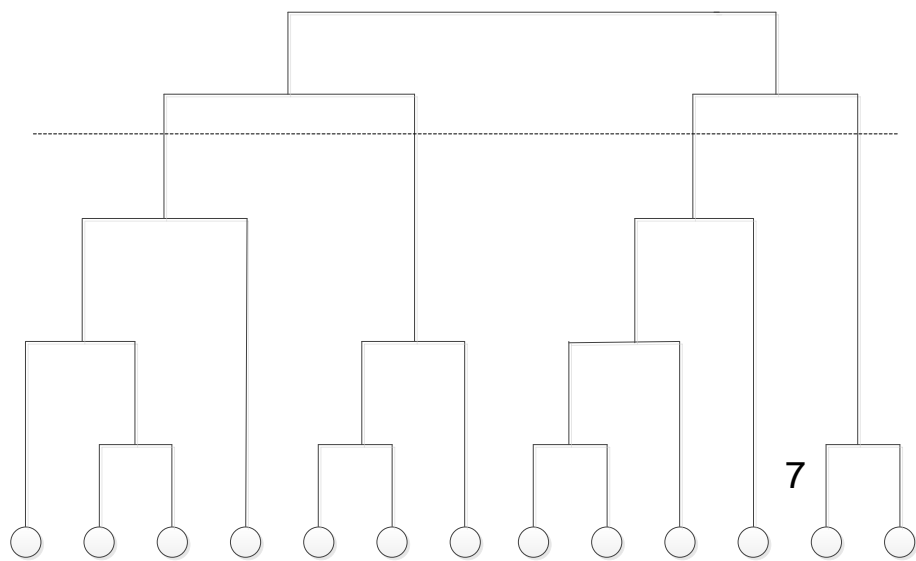
(2) 基于模块度优化的社区检测算法

FA算法的主要步骤为:

(1) 将网络中的每个节点看作一个社区，节点之间的连边被看作社区间的连边，网络包含多少个节点，就存在多少个社区；

(2) 计算某一社区与相连所有社区合并时模块度函数的增加值，进而将模块度函数数值增加最大的社区与该社区合并；

(3) 不断重复步骤(2)，直到整个网络中的所有节点都被合并为一个社区为止。该社区合并的过程可表示为一个树状图，树状图中使得Q函数最大的层次被用来划分社区。





➤ 社区检测算法的分类:

(2) 基于模块度优化的社区检测算法

- 在FA方法基础上进行了改进, Blondel等提出了BGLL算法, 简化了社区合并的过程, 大大地降低了FA算法的时间复杂度。BGLL算法的主要步骤为:
- (1)将网络中的每个节点看作一个社区, 节点之间的连边被看作社区间的连边, 网络包含多少个节点, 就存在多少个社区; 然后计算某一社区与相连所有社区合并时模块度函数的增加值, 进而将模块度函数值增量最大的社区与该社区合并, 不断重复上一步骤, 直到模块度函数值不再增加。
- (2)将上一步中的每一个社区都看作新网络的一个节点。其中, 两个新节点之间的连边权重等于两个原社区之间连边的权重之和, 而社区内部连边的权重看作新节点的自环权重。这样就得到了收缩后的新网络, 并可继续用上一步描述的方式对新网络进行处理。
- (3)重复以上两步, **直到模块度不再增加为止**。此时得到的社区划分即为BGLL算法的检测结果。BGLL算法的时间复杂度为 $O(M\log(N))$ 。



➤ 社区检测算法的分类:

(2) 基于模块度优化的社区检测算法

把模块度作为优化目标的各种算法，遗传算法、粒子群算法、等等



(2) 基于模块度优化的社区检测算法

$$Q = \sum_{i=1}^C \left[\frac{L_i}{M} - \left(\frac{d_i}{2M} \right)^2 \right], \quad (4-9)$$

缺点：根据公式(4-9)对模块度函数的定义，只要网络的图 G 中的一个子图满足不等式

$$\frac{L_i}{M} - \left(\frac{d_i}{2M} \right)^2 > 0, \quad (4-10)$$

该子图就被视作一个社区。若社区 i 的内部连边数目 L_i 满足不等式 $L_i < \sqrt{M/2}$ ，不论社区内的节点与其他社区的连边稀疏与否，该社区都难以通过最大化模块度函数检测出来，这种模块度函数的局限性称为**分辨率限制**。因此，在处理存在不同规模社区的现实世界网络时， Q 函数的效果不理想。为了克服模块度函数的局限性，学者们提出了各种改进指标，如模块度密度，但是这个问题仍然没有解决。

- 作业:

- 1 简单叙述社区检测中的**FA**算法和了**BGLL**算法的主要步骤, 说明它们异同之处。

- **大作业**(可作为本课程大作业) :

观察研究一种群体智能现象， 蚂蚁、蜂群、鸟群、鱼群，
总结其产生群体智能的机理，编程实现验证，分析其可能的
用途，整理成报告。

(大作业做完的同学直接发我邮箱，不用等最后收齐后发我)



西安电子科技大学
XIDIAN UNIVERSITY

谢谢

