

文章编号: 1001 - 2486(2011)01 - 0047 - 06

复杂网络社团发现算法研究新进展*

骆志刚, 丁凡, 蒋晓舟, 石金龙

(国防科技大学 计算机学院, 湖南 长沙 410073)

摘要: 社团结构是复杂网络普遍存在的拓扑特性之一, 发现复杂网络中的社团结构是复杂网络研究的基础性问题。针对非重叠社团发现和重叠社团发现两类问题, 全面综述了当前复杂网络社团发现算法研究的新进展, 分析了每类社团发现算法的特点, 指出该领域值得进一步探索的研究方向。

关键词: 复杂网络; 社团发现算法; 重叠社团

中图分类号: TP391 **文献标识码:** A

New Progress on Community Detection in Complex Networks

LUO Zhi-gang, DING Fan, JIANG Xiao-zhou, SHI Jin-long

(College of Computer, National Univ. of Defence Technology, Changsha 410073, China)

Abstract: Community structure is one of the common topological characteristics of complex networks. Community detection has become a fundamental problem in the research field of complex networks. The new progress of current algorithms for community detection was reviewed. The characteristics of these algorithms were discussed. Finally, future direction of this active area was proposed.

Key words: complex networks; community detection; overlapping community

复杂网络一般指节点众多、连接关系复杂的网络。由于其灵活普适的描述能力, 能够广泛应用于各科学领域对复杂系统进行建模、分析, 近年来吸引了越来越多的人对其进行研究。随着研究的深入, 人们发现许多实际网络均具有社团结构, 即整个网络由若干个社团组成, 社团之间的连接相对稀疏、社团内部的连接相对稠密。社团发现则是利用图拓扑结构中所蕴藏的信息从复杂网络中解析出其模块化的社团结构, 该问题的深入研究有助于以一种分而治之的方式研究整个网络的模块、功能及其演化, 更准确地理解复杂系统的组织原则、拓扑结构与动力学特性, 具有十分重要的意义。

自2002年 Girvan 和 Newman 基于边介数提出 GN 算法以来^[1], 国际上掀起一股社团发现的研究热潮, 来自生物、物理、计算机等各学科领域的研究者带来了许多新颖的思想和算法, 并广泛应用于各个学科领域的具体问题中。本文在归纳总结的基础上, 从非重叠社团发现和重叠社团发现两个方面综述当前社团发现算法的新进展, 并展望该领域未来的一些研究方向。

1 非重叠社团发现算法

非重叠社团发现是指识别出的社团之间互不重叠, 每个节点有且仅属于一个社团。社团发现早期的研究工作大部分都围绕非重叠社团发现展开, 相关综述可参见文献[2-3]。近年来, 基于对社团结构的不同理解, 研究者们在对节点集划分时采用的标准和策略不同, 衍生出许多风格迥异的新算法, 典型算法有模块度优化算法、谱分析法、信息论方法、标号传播方法等。

1.1 基于模块度优化的社团发现算法

基于模块度优化的社团发现算法是目前研究最多的一类算法, 其思想是将社团发现问题定义为优化问题, 然后搜索目标值最优的社团结构。由 Newman 等^[4]首先提出的模块度 Q 值是目前使用最广泛的优化目标, 该指标通过比较真实网络中各社团的边密度和随机网络中对应子图的边密度之间的差异来度量社团结构的显著性。模块度优化算法根据社团发现时的计算顺序大致可分为三类。

第一类算法采用聚合思想, 自底向上进行, 典

* 收稿日期: 2009 - 11 - 02

基金项目: 国家自然科学基金资助项目(60673018)

作者简介: 骆志刚(1962—), 男, 研究员, 博士生导师。

型代表算法有 Newman 快速算法^[4]、CNM 算法^[5]和 MSG-MV 算法^[6]等。Newman 快速算法将每个节点看作是一个社团,每次迭代选择产生最大 Q 值的两个社团合并,直至整个网络融合成一个社团。整个过程可表示成一个树状图,从中选择 Q 值最大的层次划分得到最终的社团结构。该算法的总体时间复杂度为 $O(m(m+n))$ 。在 Newman 快速算法的基础上,CNM 算法采用堆数据结构来计算和更新网络的模块度,大大提高了计算速度;MSG-MV 算法则引入多步扩展,迭代过程中每次可合并多对社团以避免过早地收缩到少数较大的社团。

第二类算法主要采用分裂的思想,自顶向下进行。例如,Newman 最早提出的 GN 算法就属于这类算法^[1],算法通过依次删去网络中边介数(即网络中经过每条边的最短路径数)最大的边,直至每个节点单独退化为社团,然后从整个删边过程中选取对应最大 Q 值时的结果。该算法复杂度较高,为 $O(n^3)$ 。随后,Newman 等人通过定义模块度矩阵,将模块度用矩阵的特征向量表示,提出一种用于划分网络社团结构的谱方法^[7]。该算法通过求解模块度矩阵的最大正特征值以及对应的特征向量,依据特征向量中元素的符号将网络不断递归二分,直至子网络再细分已不能增大 Q 值。整个算法的平均时间复杂度为,较 GN 算法在计算速度和准确度上均有较大提高。

第三类算法则是直接寻优法,如 Duch 等提出的 EO 算法^[8]以及 Agarwal 等提出的整数规划方法^[9]。EO 算法的思想是将每个节点对模块度 Q 值的贡献大小定义为局部变量,然后在随机初始划分的基础上,通过贪婪策略调整局部变量(具有最小贡献度的变量)来提高全局目标函数 Q 值。整数规划方法则通过求解对应的松弛线性规划问题能够给出最大模块度的一个上界,这是以前方法所不具备的。此外,还有一些基于遗传算法、蚁群算法等智能算法的社团发现算法也可归为此类。

近年来越来越多的研究发现:模块度优化方法无法发现小于一定粒度的社团^[10-11]。在实际网络中,尤其是大规模网络中,社团的大小不一,该问题尤为突出。为此,研究者们提出一些局部调整策略。如 Ruan 等^[12]结合谱平分法和局部搜索方法提出的 HQcut 算法,在分裂网络前增加统计测试来判断是否需进一步细分。此外,部分研究者提出新的模块度来避免 Q 值存在的粒度问题。如李珍萍等^[13]提出的模块度 D 值,在衡量社

团内外连接度的差异时,引入了社团大小作为分母进行平均,从理论和数值试验上证明了作为模块度 D 值要优于 Q 值。

总的来说,模块度优化算法是目前应用最为广泛的一类算法,但是在具体分析中,很难确定一种合理的优化目标,使得分析结果难以反应真实的社团结构,尤其是分析大规模复杂网络时,搜索空间非常大,使得许多模块度近似优化算法的结果变得更不可靠。

1.2 基于谱分析的社团发现算法

谱分析法建立在谱图理论基础,其主要思想是根据特定图矩阵的特征向量导出对象的特征,利用导出特征来推断对象之间的结构关系。通常选用的特定图矩阵有拉普拉斯矩阵和随机矩阵两类。图的拉普拉斯矩阵定义为 $L = D - W$,其中 D 为以每个节点的度为对角元的对角矩阵, W 为图的邻接矩阵;随机矩阵则是根据邻接矩阵导出的概率转移矩阵 $P = D^{-1}W$ 。这两类矩阵有一个共同性质,同一社团节点对应的特征分量近似相等,这成为目前谱分析方法实现社团发现的理论基础。

基于谱分析的社团发现算法普遍做法是将节点对应的矩阵特征分量看作空间坐标,将网络节点映射到多维特征向量空间中,运用传统的聚类方法将节点聚成社团。例如,Donetti 等^[14]基于节点之间的距离度量,在不同维度的特征空间中建立聚类树图,从中选择全局模块度最大的划分作为社团发现结果。Capocci 等^[15]则基于同一社团的节点对应的随机矩阵特征分量强相关这一性质提出计算特征向量的 Pearson 相关系数来度量节点之间的相似度。

应用谱分析法不可避免地要计算矩阵特征值,计算开销大,但由于能够通过特征谱将节点映射至欧拉空间,并能够直接应用传统向量聚类的众多研究成果,灵活性较大。

1.3 基于信息论的社团发现算法

从信息论的角度出发,Rosvall 等^[11]把网络的模块化描述看作对网络拓扑结构的一种有损压缩,从而将社团发现问题转换为信息论中的一个基础问题:寻找拓扑结构的有效压缩方式。如图 1 所示,原拓扑结构 X 通过编码器产生模块描述 Y ,解码器对 Y 进行解码,推测出原结构 Z ,那么何种模块描述 Y 是最优的?以信息论的观点来看,互信息 $I(X, Y)$ 最大时,即最能反应原始结构 X 的 Y 是最优的。在该框架下,互信息 $I(X, Y)$

最大等价于求条件信息 $H(X|Y)$ 最小, Rosvall 等给出了条件信息的量化表示并运用模拟退火优化算法进行求解, 可实现上千个节点的网络社团发现。测试表明, 对于社团大小及边密度不一的社团发现问题, 该发现算法要明显优于基于模块度优化的社团发现算法。后来, Rosvall 等进一步以描述图中信息的扩散过程为目标^[16], 将问题转换为寻找描述网络上随机游走的有效编码方式, 使该方法更适合于捕捉社团内部节点之间的长程相关性, 已有文献测试表明该方法是目前非重叠社团发现算法里准确度最高的一类方法^[17]。

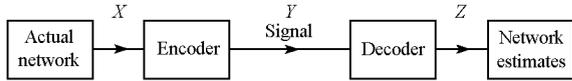


图 1 从信息论的角度看社团发现^[11]
Fig. 1 Community detection from the view of information theory^[11]

1.4 基于标号传播的社团发现算法

很多情况下, 复杂网络的边表示的是个体之间的信息传播, 其传播的结果通常是使得社团内部节点之间共享相同的信息。Raghavan 等^[18]基于该思想提出一种快速的标号传播算法(简称为 LPA 算法)。如图 2 所示, LPA 算法首先为每个节点指派唯一标号, 在每一步迭代中, 每个节点将自身标号更新为其邻节点出现次数最多的标号, 如果存在多个相同的最多标号, 则随机选择一个作为更新值, 若干次迭代后密集相连的节点会收敛于同一标号, 最终, 具有相同标号的节点归为一个社团。

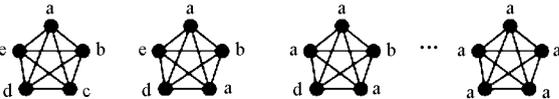


图 2 标号传播过程示意^[18]
Fig. 2 The process of label propagation^[18]

LPA 算法最大的优点在于不需要任何参数输入, 比如社团的数目、大小等, 而且算法具有线性的时间复杂度为 $O(m)$, m 为网络的边数, 收敛速度非常快, 可适用于规模较大的复杂网络。随后, Leung 等对标号更新规则进行了改进^[19]。在 LPA 迭代过程中引入衰减因子, 每次迭代由衰减因子、边权重、偏好特征决定更新的标号, 同时增加标号更新的选择性, 比如共享最大标号的邻居少于某个比例时才进行更新, 降低了计算开销。

2 重叠社团发现算法

前面所介绍的非重叠社团发现方法把每个节

点严格地划分到某个社团中, 而真实世界中这种硬划分并不能真正反应节点和社团的实际关系, 例如蛋白质相互作用网络中由于蛋白质功能的多样性, 单个蛋白质在不同的时空条件下参与不同的功能模块中。同样的现象普遍存在于各种真实网络之中, 如社会网络中的人属于多个集体、网络中的网页属于多个主题等。因此, 重叠社团发现更符合真实世界的社团组织规律, 成为近几年社团发现研究的新热点, 涌现出许多新颖算法。

2.1 基于团渗透改进的重叠社团发现算法

由 Palla 等^[20]提出的团渗透算法是首个能够发现重叠社团的算法, 文献[2]中已对其进行详细介绍。该类算法认为社团是由一系列相互可达的 k -团(即大小为 k 的完全子图)组成的, 即 k -社团。算法通过合并相邻的 k -团来实现社团发现, 而那些处于多个 k -社团中的节点即是社团的“重叠”部分。

Kumpula 等在前人工作基础上进一步提出一种快速团渗透算法(SCP 算法)^[21]。该算法分两阶段进行, 第一阶段将网络的边按顺序(如加权网络按权值大小顺序)插入到网络中, 并同时检测出现的 k -团; 第二阶段将检测的 k -团, 根据是否与已有 k -社团相邻, 并入 k -社团或形成新的 k -社团。由于边插入的顺序性, 在第二阶段检测时 SCP 算法只需依次对 k -团进行局部判断; 而且 SCP 算法能够在一遍运行中检测不同权重阈值下的 k -社团, 较大地提高了团渗透算法的计算速度。

基于团渗透思想的算法需要以团为基本单元来发现重叠, 这对于很多真实网络尤其是稀疏网络而言, 限制条件过于严格, 只能发现少量的重叠社团^[22]。

2.2 基于模糊聚类重叠社团发现算法

另一观点认为可将重叠社团发现归于传统模糊聚类问题加以解决, 以计算节点到社团的模糊隶属度来揭示节点的社团关系。这类算法通常从构建节点距离出发, 再结合传统模糊聚类求解隶属度矩阵。

张世华等首先应用这一思想^[23], 他们结合谱分析方法将网络中的节点近似映射到欧拉空间中的数据点, 进而利用 FCM 算法对空间中的数据点进行聚类, 从而得到节点与社团之间的隶属度矩阵。由于模糊聚类算法 FCM 本身要求预先知道社团数, 该算法在模块度 Q 的基础上引入新的模块度指标模糊模块度 \tilde{Q} , 选取使得 \tilde{Q} 值最大的

模糊聚类结果作为最终的社团划分结果。

上述方法在判断社团数上需要预先给定或花费大量计算以确定合理的社团数目。我们在文献[24]中提出基于通信-时间核构建距离矩阵,输入到模糊相似性传播聚类来实现重叠社团发现,在考虑节点长程相关性的同时,可以自适应地确定社团数。值得一提的是,此类算法的关键在于所构建的距离矩阵,采用何种节点距离更符合实际情况在具体应用中是一个值得探索的问题。

2.3 基于非负矩阵分解的重叠社团发现算法

非负矩阵分解方法是一种聚类和降维技术,已经在多个领域取得广泛应用。它将特征矩阵 V 近似分解为两个非负矩阵 $W_{n \times k}$ 和 $H_{k \times n}$,使得两者的乘积尽可能等于原有特征矩阵。将非负矩阵分解应用于社团发现,其思想是特征矩阵 V 中的每个行向量 v 可以近似地看作 H 的列向量的线性组合: $v = wH$, w 其中是 W 的行向量,从而可以用相对于 V 更少的基向量 H 来表示整个数据空间,而 w 则表示各基向量对于产生数据点 v 所起到的权重。因此,可以利用 W 将各节点聚成 k 个类,而 W 各分量值表示每个节点到每个社团的绝对隶属度。从图3所示的 w 分量曲线可以看出,节点6和节点11分别与两个社团呈现出一定的隶属关系,说明节点6和节点11是潜在的社团重叠区域。算法关键在于如何选取特征矩阵以表示复杂网络中所蕴含的拓扑信息。

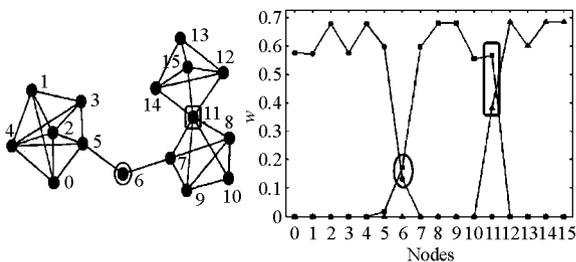


图3 一个包含两个重叠节点的简单网络及其对应的 w 分量

Fig.3 A simple network with two overlapping node and corresponding w components

张世华等首次将 NMF 应用于重叠社团发现^[25],他们以扩散核为基础,将正规化后的核矩阵作为特征矩阵 V ,同时采用一种梯度下降的迭代方法来求解 W 和 H 。该算法最大的优点在于能够给出节点与社团之间所属关系的绝对概率,这是上述算法所不具备的。但该算法依然需要预先给定 k 参数,不可避免在确定 k 值时带来较高的计算开销,仅适用较小规模的网络。Zarei 等则引入新的节点-节点关联矩阵作为特征矩阵^[26],

降低算法对初始条件的敏感性,扩大了适用范围。

2.4 基于种子扩展思想的重叠社团发现算法

此类算法的基本思想是以具有某种特征的子网络为种子,通过合并、扩展等操作向邻接节点扩展,直至获得评价函数最大的社团。该类算法近两年来得到了迅速发展。

Lancichinetti 等^[27]提出以若干个节点为种子,通过扩展形成对整个网络的覆盖,即 LMF 算法。该算法定义了两个适应度函数:社团的适应度和节点对社团的适应度。其种子扩展的基本步骤是:以单个节点 v 为初始社团 g ,考虑所有与其相邻的节点 A ,将对 g 的适应度最大的邻节点加入到当前社团形成 g' ,重新计算 g' 中各节点的适应度,将适应度为负的节点剔除,重复上述过程直至扩展后的社团其邻节点对它的适应度均为负。通过一定的种子选取策略(每次选取尚未指派到社团的节点),可以由扩展得到的若干个局部社团生成整个网络的覆盖,保证网络中每个节点至少被一个局部社团所覆盖。

此外,尚明生、陈端兵等也提出类似的种子扩展思想^[28-29],主要区别在于种子选择策略、扩展评价函数上的不同。如文献[28]提出节点强度的概念,即节点所有邻接边的权重之和,依次以最大强度的单个节点为种子;文献[29]则是以最大团为种子,通过合并和扩展两步逐步形成最终的社团,而两者均以考虑节点隶属度的重叠模块度 Q_c 为扩展评价函数。

2.5 基于混合概率模型的重叠社团发现

前述很多算法均是先给出社团结构的定义或量化特征,然后再设计相应的算法寻找社团。换句话说,这些方法均需对社团结构作出假设。针对此问题,Newman 等建立了社团结构的混合概率模型,以概率方法对复杂网络的社团结构进行探索,以求得期望最大的社团结构,从而避开社团定义的问题^[30]。他们认为社团是由具有相似连接的节点组成的,因此定义两个模型参数 θ_r 和 π_r 用来刻画社团 r 的特征,其中 θ_r 表示社团 r 中某节点对节点 i 的连接概率, π_r 表示社团 r 的节点数占整个网络节点数的比例。通过这两个参数,可以将网络的邻接矩阵 A 看作是观察数据,并引入缺失数据 $g = \{g_i\}$,用来表示节点 i 所属社团,由此将问题转换为典型的带缺失数据的最大似然估计问题,其似然度计算如下:

$$\begin{aligned} L &= \ln P(A, g | \pi, \theta) \\ &= \ln P(A | g, \pi, \theta) P(g | \pi, \theta) \end{aligned}$$

$$= \sum_i \left[\ln \pi_{g_i} + \sum_j A_{ij} \ln \theta_{g_i, j} \right]$$

通过简单变形后可直接运用 EM 算法进行求解。由于该算法未对社团结构作过多假设,具有较强的灵活性,能够发现不同结构类型的社团,如具有二分结构的社团,这是之前方法所不具备的。同时所求得参数 θ_{ri} 中暗含了节点 i 到社团 r 的隶属程度,因此,通过该算法能够识别出重叠社团,并得到隶属程度大小。然而,该方法基于 EM 算法来估计未知参数,收敛速度较慢,计算复杂度较高,一定程度上制约了算法的应用规模。

2.6 基于边聚类的重叠社团发现

以往社团发现算法的研究均以节点为对象,考虑如何通过划分、聚类、优化等技术将节点归为重叠或不重叠的社团。最近,Evans 等^[31]和 Ahn 等^[32]分别发表了以边为研究对象来划分社团。虽然节点属于多重社团,但边通常只对应某一特定类型的交互(真实网络中的某种性质或功能)。因此,以边为对象使得划分的结果更能真实地反映节点在复杂网络中的角色或功能。文献[31-32]虽然都是对边进行聚类来提示重叠社团,但在处理方式上有所不同。

文献[31]将原网络转换为加权线图,即原网络中的边映射为线图上的节点,线图上的节点存在边当且仅当原网络中所对应的边存在共享节点,而边权值等于 $1/(k-1)$,其中 k 为共享节点的度。通过网络的转换,最大的好处是可以直接应用非重叠社团发现算法来提示网络的重叠社团。

文献[32]则着重解决社团的重叠性与层次性之间的冲突,提出边社团的概念,即将社团看作是紧密相连的边集合。作者在定义边间相似度的基础上,采用层次聚类建立聚类树状图。由于每个节点拥有一条到多条边,每条边存在于一个边社团中,因此对该树状图不同层次的切割得到的社团结构对应着不同层次的重叠社团。

3 总结与展望

从异常复杂的网络解构出其中的社团结构,已成为当今复杂系统研究领域一项具有挑战性的研究课题。虽然该课题近些年受到广泛关注,涌现出一批新颖的算法,但目前相关研究仍未形成统一的框架和度量标准,尤其是重叠社团发现算法的研究,尚存许多问题亟待解决。我们认为,复杂网络社团发现的进一步研究可从以下几个方面展开。

(1)建立统一的度量标准。由于复杂网络的类型众多,连接规律各有不同,很难以社团结构的某种统一的模块度(如 Q 值)来刻画社团发现算法的优劣。一种更为科学的方式是建立一套包含多种复杂网络的统一标准测试集,以评判算法在不同类型网络中的优劣,明确算法的适用范围。

(2)适用于大规模复杂网络的社团发现算法。复杂网络的规模越来越大,对算法的计算复杂度提出了更高要求。虽然在不考虑重叠社团的情况下,已出现一些接近于线性时间复杂度的算法,但这些算法通常采用较为激进的贪婪策略,网络规模变大且非稀疏时,其结果变得不可靠。在重叠社团发现算法中,很多算法需要通过多次计算来获得最佳的社团数,计算开销过大。因此,考虑复杂网络社团密度不均的特点,从局部社团出发研究网络的社团结构是未来的重要研究方向之一。此外,设计适合于大规模网络分析的高效并行算法也是未来重要的研究方向之一。

(3)重叠社团与层次社团的结合。一般认为,社团之间共享部分边缘节点从而产生重叠社团,然而重叠社团结构远比想象的要复杂。实际上,除了重叠性,层次性也是社团结构的另一大特性。例如,第 i 层中的中心节点,可能在第 j 层中就变成了边缘节点。可见,重叠性与层次性两者联系十分紧密,有必要将两者融和在一起来解构复杂网络。在目前的众多方法中,唯有边社团给出了社团重叠性和层次性普遍并存的合理解释,未来以边为对象来研究网络社团结构将是一个值得深入研究的方向。

参考文献:

- [1] Girvan M, Newman M E J. Community Structure in Social and Biological Networks[J]. P Natl Acad Sci USA, 2002, 99(12): 7821 - 7826.
- [2] 解岱,汪小帆.复杂网络中的社团结构分析算法研究综述[J].复杂系统与复杂性科学,2005,2(3):12.
- [3] 李晓佳,张鹏,狄增如,等.复杂网络中的社团结构[J].复杂系统与复杂性科学,2008,5(3):24.
- [4] Newman M E J. Fast Algorithm for Detecting Community Structure in Networks[J]. Phys Rev E, 2004, 69(6): 066133.
- [5] Clauset A. Finding Local Community Structure in Networks[J]. Phys Rev E, 2005, 72(2): 026132.
- [6] Schuetz P, Cafilisch A. Multistep Greedy Algorithm Identifies Community Structure in Real-world and Computer-generated Networks[J]. Phys Rev E, 2008, 78(2): 026112.
- [7] Newman M E J. Modularity and Community Structure in Networks[J]. Proc Natl Acad Sci USA, 2006, 103(23): 8577 - 8582.
- [8] Duch J, Arenas A. Community Detection in Complex Networks Using Extremal Optimization[J]. Phys Rev E, 2005, 72(2):

- 027104.
- [9] Agarwal G, Kempe D. Modularity-maximizing Graph Communities Via Mathematical Programming[J]. *The European Physical Journal B-Condensed Matter and Complex Systems*, 2008, 66(3): 409 – 418.
- [10] Fortunato S, Barthelemy M. Resolution Limit in Community Detection[J]. *P Natl Acad Sci USA*, 2007, 104(1): 36 – 41.
- [11] Rosvall M, Bergstrom C T. An Information-theoretic Framework for Resolving Community Structure in Complex Networks[J]. *P Natl Acad Sci USA*, 2007, 104(18): 7327 – 7331.
- [12] Ruan J H, Zhang W X. Identifying Network Communities with a High Resolution[J]. *Phys Rev E*, 2008, 77(1): 016104.
- [13] Li Z, Zhang S, Wang R S, et al. Quantitative Function for Community Detection [J]. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2008, 77(3): 036109.
- [14] Donetti L, Munoz M. Detecting Network Communities: A New Systematic and Efficient Algorithm [J]. *Journal of Statistical Mechanics*, 2004, P10012.
- [15] Capocci A, Servidio V D P, Caldarelli G, et al. Detecting Communities in Large Networks [J]. *Physica A: Statistical Mechanics and Its Applications*, 2005, 352(2 – 4): 669 – 676.
- [16] Rosvall M, Bergstrom C T. Maps of Random Walks on Complex Networks Reveal Community Structure [J]. *Proc Natl Acad Sci USA*, 2008, 105(4): 1118 – 1123.
- [17] Lancichinetti A, Fortunato S. Community Detection Algorithms: A Comparative Analysis[J]. *Physical Review E*, 2009, 80(5): 056117.
- [18] Raghavan U N, Albert R, Kumara S. Near Linear Time Algorithm to Detect Community Structures in Large-scale Networks[J]. *Phys Rev E*, 2007, 76(3): 036106.
- [19] Leung I, Hui P, Lio P, et al. Towards Real-time Community Detection in Large Networks[J]. *Physical Review E*, 2009, 79(6): 66107.
- [20] Palla G, Derenyi I, Farkas I, et al. Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society [J]. *Nature*, 2005, 435(7043): 814 – 818.
- [21] Kumpula J M, Kivela M, Kaski K, et al. Sequential Algorithm for Fast Clique Percolation [J]. *Physical Review E*, 2008, 78(2): 026109.
- [22] Zhang S H, Ning X M, Zhang X S. Identification of Functional Modules in a PPI Network By Clique Percolation Clustering[J]. *Comput Biol Chem*, 2006, 30(6): 445 – 451.
- [23] Zhang S, Wang R, Zhang X. Identification of Overlapping Community Structure in Complex Networks Using Fuzzy C-means Clustering [J]. *Physica A: Statistical Mechanics and its Applications*, 2007, 374(1): 483 – 490.
- [24] Ding F, Luo Z, Shi J, et al. Overlapping Community Detection By Kernel-based Fuzzy Affinity Propagation [C]//*Proceedings of the 2nd International Workshop on Intelligent Systems and Applications(ISA'2010)*, Wuhan, China, 2010.
- [25] Zhang S, Wang R S, Zhang X S. Uncovering Fuzzy Community Structure in Complex Networks[J]. *Phys Rev E*, 2007, 76(4): 046103.
- [26] Zarei M, Izadi D, Samani K A. Detecting Overlapping Community Structure of Networks Based on Vertex-vertex Correlations [J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2009, P11013.
- [27] Lancichinetti A, Fortunato S, Kertesz J. Detecting the Overlapping and Hierarchical Community Structure in Complex Networks [J]. *New Journal of Physics*, 2009, 11:033015.
- [28] Chen D, Shang M, Lv Z, et al. Detecting Overlapping Communities of Weighted Networks Via a Local Algorithm [J]. *Physica A: Statistical Mechanics and its Applications*, 2010, 389(19): 4177 – 4187.
- [29] Shang M S, et al. Detecting Overlapping Communities Based on Community Cores in Complex Networks [J]. *Chinese Physics Letters*, 2010, 27(5): 058901.
- [30] Newman M E, Leicht E A. Mixture Models and Exploratory Analysis in Networks[J]. *Proc Natl Acad Sci USA*, 2007, 104(23): 9564 – 9569.
- [31] Evans T, Lambiotte R. Line Graphs, Link Partitions, and Overlapping Communities[J]. *Physical Review E*, 2009, 80(1): 16105.
- [32] Ahn Y Y, Bagrow J P, Lehmann S. Link Communities Reveal Multiscale Complexity in Networks[J]. *Nature*, 2010, 466: 761 – 764.