

## 基于模糊聚类的多分辨率社区发现方法

汪晓锋 刘功申\* 李建华

(上海交通大学电子信息与电气工程学院 上海 200240)

**摘要:** 针对网络结构的复杂性和群体划分的不确定性, 该文提出一种基于模糊聚类的多分辨率社区结构发现方法。该方法用模糊方法来处理网络节点间的相似性, 以实现社区结构的模糊划分。基于节点间的局部交互信息, 考虑节点间的模糊关系和网络拓扑结构相似性传递, 实现网络社区的层次聚类。并通过调节模糊参数, 挖掘出不同分辨率下的社区结构。同时为了避免主观地确定社区数目, 引入一种新的模块度以度量社区划分结果。实验证明该方法能够有效且稳定地揭示潜在的社区结构。

**关键词:** 社交网络; 社区发现; 模糊聚类; 结构相似性

中图分类号: TP393; TP391

文献标识码: A

文章编号: 1009-5896(2017)09-2033-07

DOI: 10.11999/JEIT161116

## Multiresolution Community Detection Based on Fuzzy Clustering

WANG Xiaofeng LIU Gongshen LI Jianhua

(School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

**Abstract:** Focusing on the complexity of network structure and the indeterminacy of community partition, this paper puts forward a novel fuzzy clustering method for uncovering community structures. In contrast to previous studies, the proposed method disposes the similarity of connecting vertices with fuzzy relation. Based on local interactive information, it considers the fuzzy relation between vertices and the transitive similarity in network topology to divide vertices into communities. In addition, multiresolution communities can be detected by adjusting fuzzy parameter. In order to avoid subjectivity in the selection of cluster number, a new modularity is introduced to evaluate the effectiveness of the clustering analysis. It is proved by experiments that the method is efficient and stable to detect underlying communities.

**Key words:** Social network; Community structure; Fuzzy clustering; Structural similarity

### 1 引言

近年来关于复杂网络的研究受到国内外学者的广泛关注。复杂网络作为一个广泛的交叉学科, 其涉及计算机、物理学、信息科学、系统科学、网络科学等学科领域, 逐渐成为解决复杂问题的一个强有力工具, 并在众多领域有着广泛的应用, 如社交网络, 生物工程、经济金融、电力与交通、人类行为分析, 大数据分析等<sup>[1,2]</sup>。在大量的复杂网络研究中, 社区结构是一个重要的研究焦点。社区结构是复杂网络的一个重要结构特征, 分析社区结构对于理解网络的结构属性、交互模式以及群体行为至关

重要。网络社区或社团的概念很早被提出, 目前存在各种不同的定义<sup>[3-5]</sup>。一般说来, 对于给定的网络结构划分, 社区内部节点相对于社区之间有着更紧密的连接。在社交网络中, 由于社会交互性的增强, 基于不同兴趣、主题、职业、地域等特征形成的大量群体, 社区结构特征尤为明显。因此, 挖掘网络中紧密联系的社区结构对理解和分析网络结构属性、信息传播规律、人类社会组织结构等有重要的理论意义和广泛的应用价值。

社区结构发现旨在识别网络中带有某种自然属性的群体结构, 即根据网络的拓扑结构或节点属性, 将网络中互连的节点划分为若干模块, 使得各模块内部的联系相对稠密, 模块间的连接相对稀疏。尽管社区结构的概念便于理解, 但是由于复杂网络的多样性和复杂性, 提出一个有效揭示群体特征的社区发现算法是一个重大挑战。在复杂网络研究中, 已有许多社区发现算法被提出, 如基于图分割<sup>[6,7]</sup>、层次聚类<sup>[5,8]</sup>、派系过滤<sup>[9-11]</sup>、中心性度量<sup>[12,13]</sup>、谱

收稿日期: 2016-10-20; 改回日期: 2017-05-10; 网络出版: 2017-06-14

\*通信作者: 刘功申 lgshen@sjtu.edu.cn

基金项目: 国家 973 关键技术研究项目(2013CB329603), 国家自然科学基金(61472248, 61431008)

Foundation Items: The National 973 Key Basic Research Program of China (2013CB329603), The National Natural Science Foundation of China (61472248, 61431008)

聚类<sup>[14,15]</sup>、基于模块度优化<sup>[16-18]</sup>、连边聚类<sup>[19-21]</sup>等算法。尽管如此,但大部分方法依赖于先验知识的指导才能实现有效划分,同时没有提出一个统一的度量准则,因此存在一定局限性。对于社区结构划分,有两个关键问题需要解决:一是确定社区的数量,对于未知的网络结构,其社区划分和结构特征是预先未知的;另外就是各个社区成员的确定,避免社区结构的不合理划分。传统的社区划分方法如图分割将网络中的每一个节点硬性地划分到一个特定的社区,而忽略了其内在联系。在现实网络中,由于网络结构的复杂性和多样性,一个节点可能同时属于多个社区,或者一个节点以不同的参与度参与到不同的社区,即在社区发现过程中,存在不确定性或模糊性,社区结构的模糊划分更接近真实的网络结构。

针对上述社区划分数量选择的问题,早在2004年Newman等人<sup>[22]</sup>在提出了GN算法并提供了一种评价社区结构强度的模块度量,以此来确定网络社区结构的最佳划分数量。尽管GN算法在应用于大规模网络的社区结构发现时由于较高的算法复杂度而受到限制,但其模块度的概念在后来的研究中得到大量应用和进一步发展。针对社区结构的合理划分,存在大量研究与讨论,如基于节点相似性度量<sup>[23, 24]</sup>、中心度<sup>[13, 22]</sup>、聚类系数<sup>[25]</sup>以及重叠结构<sup>[21, 26]</sup>等。就社区结构本身来说,其内部节点之间是基于某种相似性或共同特征如社交网络中的不同兴趣、爱好、主题等关系联系在一起,这是人们研究社区划分的出发点之一。然而,大量现存的社区结构发现方法将个体与群体间相似性关系作为一种确定性或硬性的度量,即非此即彼,如此可能会导致社区的不合理划分。实际上,在真实网络结构如社交网络中个体间的相似关系是模糊的或不确定性的,而以一种确定性度量来划分可能忽略网络中其他重要信息。

因此,针对社区结构发现中的社区有效划分与划分数目两个基本问题,本文提出了一种模糊聚类技术与模块度相结合的多分辨率社区发现(FCD)方法,同时本文引入了新的相似性和模块度量。本文提出用模糊聚类的方法来揭示复杂网络中潜在的社区结构,是基于一种模糊关系模型而非以往的图模型来解决复杂网络中的社区划分问题。为实现社区结构的合理划分,本文算法先从根据节点的局部结构信息计算节点间的结构相似性,将节点间确定性关系转换为模糊关系,再将模糊关系通过传递闭包转换为模糊等价关系,然后将等价的模糊关系映射为模糊等价类,从而得到网络的层次性社区结构

划分。由于该方法得到的是不同规模、不同粒度的模糊社区,为了实现网络的最优划分结果,我们提出用一种新的模块度量来确定社区结构数目。

## 2 模糊聚类与社区发现

### 2.1 模糊聚类

模糊聚类以模糊集合理论为基础,用模糊数学的方法处理聚类问题。由于模糊聚类刻画了样本属于各个类别的不确定性程度,表达了个体类属的中介性,更能客观反映现实世界,从而成为聚类分析研究的主流。模糊聚类已经应用在诸多领域,如模式识别、图像处理、自动控制等<sup>[27]</sup>。在聚类分析中,当聚类边界特征不明显或不明确的对象时,模糊聚类机制显得更为合理。同时,模糊化的隶属度有助于发掘客观对象间复杂关系和潜在的集群<sup>[28]</sup>。

本文采用一种基于模糊关系的聚类分析方法,用模糊关系来描述待聚类模式间的相似性或隶属度,并建立一个模糊聚类模型来实现网络结构的模糊划分。在网络结构中,其模糊关系强调网络节点以不同的隶属程度同时属于多个社区类别。而非严格划分到某一特定社区。同时,模糊相似性关系可看作是确定性关系的一种泛化。接下来引入模糊聚类分析中的几个重要定义<sup>[28, 29]</sup>。

**定义 1** 对于非空集合  $U = \{u_1, u_2, \dots, u_m\}$  和  $V = \{v_1, v_2, \dots, v_n\}$ ,  $F(U \times V)$  为  $U \times V$  上的所有模糊关系集合。对于一个模糊关系  $R$ , 其中  $R \in F(U \times V)$ , 如果  $\forall (u, v) \in (U \times V)$ ,  $R(u, v)$  可表示为  $R$  中有序节点对  $(u, v)$  的隶属度值, 同时  $\exists a_{ij} = R(u_i, v_j)$  被称作模糊关系, 其中  $a_{ij} \in [0, 1]$ , 则  $R$  为  $V$  上的二元模糊关系。

**定义 2** 对于一个模糊关系  $R$ ,  $v \in V$ ,  $R$  为一个模糊等价关系, 如果满足以下条件:

- (1) 自反性:  $R(v_i, v_j) = 1$ ;
- (2) 对称性:  $R(v_i, v_j) = R(v_j, v_i)$ ;
- (3) 传递性:  $\exists \gamma \in (0, 1)$ , 若  $R(v_i, v_j) \geq \gamma$ , 且  $R(v_i, v_k) \geq \gamma$ , 则  $R(v_j, v_k) \geq \gamma$ 。

**定义 3** 对于一个模糊关系矩阵  $R = (a_{ij})_{m \times n}$ , 其传递闭包为包含  $R$  的最小传递关系  $t(R)$ , 其满足  $R \subseteq t(R)$ , 且  $t(R) = \bigcup_{k=1}^n R^k$ , 其中  $R \in F(U \times V)$ ,  $n = |V|$ 。

若  $R$  满足自反性和对称性, 则  $R$  为模糊相似关系矩阵, 且有  $R^2 \Rightarrow R^4 \Rightarrow \dots \Rightarrow R^{2^k}$ , 其中  $k \leq \lceil \log_2^n \rceil + 1$ 。  $t(R)$  为模糊等价关系, 且  $t(R) = e(R)$ , 其中  $e(R)$  表示包含  $R$  的最小模糊等价关系。模糊关系的传递闭包反应了模糊关系的一致收敛性。当  $R$  为一个给定的模糊等价关系, 对于节点集  $V$  中的节点  $v_i$ ,

其等价类为  $\mathbf{V}$  中所有节点的一个分类等价于：

$v_i : [v_i]_{\mathbf{R}} = \{v_j \in \mathbf{V} \mid \mathbf{R}(v_i, v_j)\}$ , (若  $v_j \in [v_i]_{\mathbf{R}}$ , 则  $[v_i]_{\mathbf{R}} = [v_j]_{\mathbf{R}}$ )。基于模糊关系的聚类方法实际上是一种动态聚类过程。在一个网络中, 对于一个给定的节点, 围绕其得到的模糊子集, 对应为按节点模糊关系的邻近程度构成的网络层次结构。通过选取模糊水平值, 可得到一个等价分类。

## 2.2 结构相似性度量

在一个网络结构中, 节点间的连接是一种二元的确定性关系, 可以用邻接矩阵来表示。为了将节点间这种确定性关系转化为模糊关系, 本文引入一种结构相似性度量。从社区的角度来说, 社区是基于某种共同属性或相似特征相互联系形成紧密的局部群体。所以社区可看作是各种关系紧密联系的不同群体, 社区成员之间彼此相似, 并且大量节点有着共同的邻居。因此, 不论任意两个节点间有无直接连边, 其相似性可以通过某种局部或全局的度量来衡量, 社区中的每一个节点都与其最相似的节点直接相连。本节提出一种基于节点共同邻居的相似性度量函数。

对于一个网络  $\mathbf{G}$  中任意两节点  $v$  和  $w$  之间的距离表示为  $d(v, w)$ , 在此表示两节点间最短路径长度。为简化计算并不失一般性, 各边的权重设为 1。

**定义 4** 对于任意正整数  $k$ , 节点  $v$  的  $k$  阶邻居为包含所有距节点  $v$  距离不大于  $k$  的节点集合, 也即  $N_k(v) = \{w \in \mathbf{V} \mid d(v, w) \leq k\}$ 。因此,  $N_1(v)$  为与节点  $v$  直接相连接的节点集合。对于  $k = 1$ , 称  $N_1(v)$  为节点  $v$  的直接邻居。对  $k \neq 1$ , 称  $N_k(v)$  为节点  $v$  的非直接邻居。

**定义 5** 结构相似性定义如下：

$$S(v, w) = \begin{cases} \frac{|N_k(v) \cap N_k(w)| + 1}{|N_k(v)| \cdot |N_k(w)|}, & v, w \in \mathbf{V}, v \neq w \\ 1, & v, w \in \mathbf{V}, v = w \end{cases} \quad (1)$$

其中,  $\mathbf{V}$  表示网络的节点集合,  $(v, w)$  为节点  $v$  和  $w$  之间的连边。通过标准化处理, 使得  $S(v, w) \in [0, 1]$ 。本文算法中仅采用结构相似性的一阶 ( $k = 1$ ) 的最简单形式。

基于共同邻居节点的相似度量可知, 两节点的共同邻居越多, 其相似度值越大。若两个节点具有相似的拓扑结构, 其在网络的功能和作用也相似。如社交网络中的节点相似表明节点间的共同的兴趣、地域、爱好等特征; 在 PPI 网络中, 相似性节点则表明相似的分子功能结构。因此, 网络拓扑结构中的相似性可以确定节点间的近似程度。

## 2.3 模块度量

模块度最初由文献[22]引入而被广泛采用的一

种评价社区划分质量的度量。一般地, 一个网络模块度定义为该网络的社区内部连边数量与相应的零模型的社区内部连边数量之差占整个网络连边的比例。对于一个特定网络, 不同的社区划分对应的模块度值一般也是不一样的。最大的模块度值表示最优的社区划分, 并对应合理的社区划分数目。因此, 模块度的各种变体相继被提出以适应特定的社区聚类算法<sup>[5]</sup>, 其中大部分方法转化为模块度优化的问题。基于本文中模糊聚类方法来说, 模块度的引入, 恰好弥补了模糊聚类技术要求预先确定社区划分数目的不足<sup>[30]</sup>。由于本文算法将节点间确定性的连接关系转化为模糊相似性, 常规定义的模块度量无法适用这种情况。因此, 针对模糊聚类的模糊特性, 本文提出了一种新的模块度, 以评价社区结构的模糊划分。

对于一个给定的具有  $K$  社区结构的网络划分  $P_K$ , 定义一个  $K \times K$  对称矩阵  $e$ , 其元素  $e_{ij}$  表示该网络划分中社区  $i$  中的节点和社区  $j$  中节点的连边集合。结合社区划分的模糊性, 本文引入一种基于相似性的模块度函数  $Q$  来度量社区结构。对给定的图  $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{S})$ , 其中  $\mathbf{V}$  为节点集合,  $\mathbf{E}$  为连边集合, 以相似度矩阵  $\mathbf{S} = (S(v, w))_{n \times n}$  作为连边的权重,  $n$  为网络节点规模。该模糊模块度函数定义为

$$Q(P_K) = \sum_{C=1}^K \left[ \frac{S(\mathbf{V}_C, \mathbf{V}_C)}{S(\mathbf{V}, \mathbf{V})} - \left( \frac{S(\mathbf{V}_C, \mathbf{V})}{S(\mathbf{V}, \mathbf{V})} \right)^2 \right] \quad (2)$$

其中,  $C \in [2, K]$ , 且  $S(\mathbf{V}', \mathbf{V}'') = \sum_{v \in \mathbf{V}', w \in \mathbf{V}''} S(v, w)$ 。

此模块度函数可作为 Newman 模块度函数的一种推广, 根据最大  $Q$  值对应最佳的社区数目  $K$ 。对于社区划分, 有两个极端情形: 当把整个网络划分为单一社区,  $Q$  取值为 0; 当网络中每一个节点视为一个社区, 模块度为负值。一个给定网络划分的最大模块度值满足  $0 \leq Q_m < 1$ 。

## 3 基于模糊聚类的多分辨率社区发现方法 (FCD)

上文描述了本文所提出的社区发现方法的理论基础和基本概念, 接下来将引入具体的社区发现方法。具体地, 本算法基于节点间的局部交互信息, 引入的结构相似性以度量节点间的模糊相似关系, 再通过模糊关系的传递性, 将交互的网络图映射到模糊等价类, 实现社区结构的模糊划分。并通过调节模糊参数, 挖掘出不同分辨率下的网络社区结构。同时为了避免主观地确定社区数目, 引入一种模糊模块度以度量社区划分结果。算法基本流程描述如下:

(1) 对于一个网络图  $G$ , 根据式(1)将其邻接矩阵  $A$  转换为模糊关系矩阵  $R$ ;

(2) 根据定义 3 计算模糊关系矩阵  $R$  的传递闭包, 即传递模糊关系  $t(R)$ ;

(3) 选择合适的阈值  $\varepsilon$  来从模糊关系矩阵中截取划分。设  $t(R) = (s(i, j))$ ,  $t(R)_\varepsilon$  为一个等价布尔关系, 则一个聚类划分为  $t(R)_\varepsilon = (\bar{S}(i, j))$ , 其中  $\bar{S}(i, j) = \begin{cases} 0, & S(i, j) < \varepsilon \\ 1, & S(i, j) \geq \varepsilon \end{cases}$ 。若  $\bar{S}(i, j) = 1$ ,  $v_i, v_j$  为同一等价类。等价类对应于与网络图中的节点社区;

(4) 调节参数  $\varepsilon$ , 按照步骤(3)中的聚类准则, 得到不同粒度的社区结构划分, 每个划分对应于不同的聚类数目;

(5) 针对得到的每一划分, 根据式(2)计算模糊模块度, 确定最佳的聚类数目。

本文算法基于节点间局部交互信息来度量结构相似性并揭示网络中的多分辨率社区结构, 因此具有相对较低的计算时间复杂度。假设给定网络节点规模为  $n$ , 网络连边数量为  $m$ , 可从如下几个方面对算法 FCD 的时间复杂度进行理论分析。首先, 在节点结构相似性计算过程中, 算法采用了向量化计算从而改进算法的计算效率, 其计算时间复杂度为  $O(n \lg n)$ , 其次, 相似性关系的模糊传递计算要求时间复杂度为  $O(n^2)$ , 最后, 多分辨率获取阶段需要将模糊阈值与稳定的节点相似性关系进行比较, 其接近线性计算时间。因此, 本文算法一般情况下计算时间复杂度为  $O(n^2)$ , 与实验中所用对比算法相比仍然存在一定优势。

## 4 实验与分析

为评价算法的有效性, 本文采用模块度( $Q$ )<sup>[22]</sup>和标准互信息(NMI)<sup>[31]</sup>这两种常用的衡量社区发现算法划分质量的评价指标。模块度作为一种常用的社区质量函数, 能有效刻画社区结构内部的紧密程度。标准互信息是一种基于信息论的度量, 用于衡量社区划分与已知的网络社区结构之间的差异性, 其值越大则表明社区划分越精确。

在实验中, 我们采用两类被广泛采用的网络数据集。一类是真实世界网络。真实世界网络数据是对现实世界中的关系网络的抽象, 实际上不存在明确的社区划分界限。真实世界网络包括空手道俱乐部网络、美国大学生足球联盟网络、海豚网络及 Amazon 政治书籍网络等<sup>[32]</sup>。另一类为人工合成网络。该类网络通过模拟现实网络中的社区结构特征而预设各种网络参数并用计算机程序生成的数据集, 因而这类网络存在已知的社区结构划分。目前

最常用的一种人工网络为 LFR 基准网络<sup>[33]</sup>。LFR 基准网络中的节点度和社区规模大小均服从幂律分布, 因此更接近现实网络结构特征。

作为对比, 本文引用了目前一些经典的社区发现算法, 如 CNM 算法<sup>[8]</sup>、Walktrap 算法<sup>[34]</sup>、Infomap 算法<sup>[35]</sup>和 BGLL 算法<sup>[17]</sup>, 从算法性能上进行了定量比较分析。此外, 本文通过实验对控制社区分辨率的模糊系数进行了讨论, 以及其最优模糊参数取值区间。

### 4.1 真实网络数据集

(1) Zachary 网络: 作为社交网络分析中一个经典的真实网络, Zachary 网络广泛用于测试社区检测算法的有效性和可靠性。该网络是由 Zachary 对该俱乐部成员间的社会交互进行了长达两年时间的调查和观察, 并构建了一个具有 34 个俱乐部成员 78 条边的交互网络。由于俱乐部管理者和教练间的意见分歧, 网络最终分裂成两个较小的社区结构。

在实验中, 本文方法能有效检测出网络结构中的分裂现象。同时通过调节聚类参数  $\varepsilon$ , 社团内部的紧密联系的更小粒度的社团被发现。当选取更大的参数值时, 该网络被划分为 4 个社团结构, 其与 Clauset 等人提出的 CNM 算法划分的结果一致。通过调节合适的参数值, 不同数量的社区结构被提取, 并对应不同的模块度值, 如图 1 所示。当被划分为 4 个社区结构时有最大的模块度, 而实际网络对应的 2 个社区的划分并非最优, 因此在模块度水平相当时, 应根据更有效的准则来适当选择。而在本实验中, 根据所提方法, 其社区划分结构是精确而有效的。

(2) Football 网络: Football 网络描述了美国大学生足球队在常规赛中的所有赛程, 其中的节点表示全部 115 支球队, 其中的连边表示 613 场比赛。这个网络被分为 10 个小组, 通常在小组内部球队之间的比赛比小组间的常规比赛更加频繁。由于这种特点, 该网络常用于社区发现实验中。

用所提方法对该网络进行聚类划分, 我们得到了有效的划分结果。通过对聚类参数的优化, 网络中的 10 个社区的划分结果被有效提取, 其与实际网络结构一致, 同时对应最大的模块度值。文献[4]中提出的 GN 算法对该网络进行社区划分, 仅准确发现其中 6 个社区, 且 CNM 算法仅能准确地检测到 4 个社区结构。通过调节聚类参数, 本算法能检测出更小分辨率的社区结构(多于 10 个社区结构)。由于网络本周具有的较强的社区结构特征, 大部分划分都能保持社区结构的完整性, 这一现象可以反映在针对不同社区划分的缓慢的模块度值变化上, 如图 2 所示。

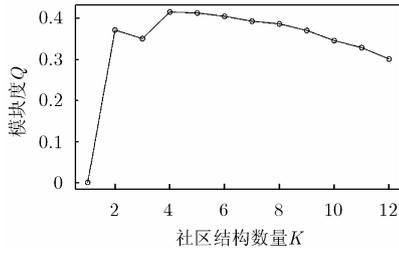


图 1 本文算法在Zachary网络中的模糊社区划分

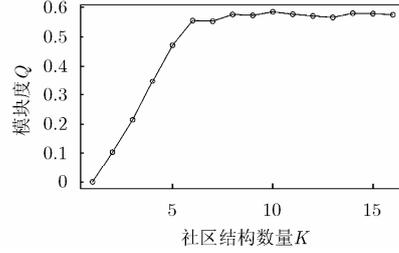


图 2 本文算法在Football网络中的模糊社区划分结果

(2)实验数据对比：除以上两个网络外，本文还对另外一些真实网络进行了实验，并定量对比了其他社区发现算法。本文实验所用真实网络基本信息如表 1 所示。表 2 为包括本文算法在内的 5 个算法在这些真实网络中的实验结果。

表 1 真实网络数据集统计信息

数据集	节点	边数	描述
Zachary	34	78	空手道俱乐部网络
Dolphins	62	159	海豚社会关系网络
Football	115	616	美国大学生足球网络
Polbooks	105	441	亚马逊美国政治书籍网络
Word	112	425	狄更斯小说词连接网络
Neural	297	2,148	Elegans 神经网络
Polblogs	1490	16718	美国政治博客网络

表 2 真实网络数据实验对比结果

数据集	CNM	Walktrap	Infomap	BGLL	FCD
Zachary	0.381	0.353	0.402	0.418	0.415
Dolphins	0.495	0.488	0.5247	0.518	0.526
Football	0.549	0.604	0.568	0.605	0.585
Polbooks	0.502	0.507	0.523	0.521	0.541
Word	0.295	0.216	0.009	0.281	0.284
Neural	0.372	0.353	0.358	0.387	0.341
Polblogs	0.427	0.425	0.423	0.427	0.453

从表 2 可以看出，本文算法在以上 7 个真实网络中均取得较好的结果，并在其中 4 个网络中得到了最大的模块度值。通过比较可知，CNM 算法仅在 Word 网络中取得了最大的模块度，而 BGLL 算法在其中的 3 个真实网络中表现最佳。总体而言，本文算法在有效性上相对其他算法有一定优势。

#### 4.2 人工合成网络

实验所用 LFR 基准网络中的节点度和社区规模均服从幂律分布，体现了真实网络中同质性特征。LFR 基准网络生成模型分别由网络节点规模、平均

度、最小社区规模、最大社区规模、幂律分布指数以及混合系数等参数来决定。对于一个基准网络来说，其混合系数越高，表示社区之间的边界越模糊，社区结构越难以区分。通过调节不同的混合系数，实验生成一组规模为 1000 的 LFR 网络，并与其他算法进行了比较，其结果如图 3 所示。LFR 网络模型参数： $N = 1000$ ， $k = 10$ ， $k_{max} = 50$ ， $C_{min} = 20$ ， $C_{max} = 50$ ， $t_1 = 2$ ， $t_2 = 1$ 。

通过对比可以发现，当混合系数  $\mu \leq 0.5$  时，本文所提出的 FCD 算法能获得最优的聚类结果，表明算法所得到社区划分与 LFR 网络中真实社区结构一致。同时，Infomap, Walktrap 和 BGLL 算法达到了相近的性能效果。然而，随着混合系数的增大 ( $\mu > 0.5$ )，各算法的精确性均逐步降低，其中 Infomap 算法下降最为明显，当混合系数大于 0.6 时，其未能识别出正确的社区结构。当混合系数在 0.5 至 0.6 之间时，BGLL 算法表现最佳，但随后其性能大幅降低。此外，Walktrap 算法在混合系数小于 0.6 时实现了与本文算法相近的性能，但当混合系数进一步增大后下降明显。总的来说，本文算法在较大的混合系数条件下，保持了较高的稳定性。

为了确定算法产生的多分辨率社区中最优社区划分，本文分析了模糊阈值的选取对算法的影响，如图 4 所示。图中曲线表示当混合参数从 0.1 到 0.8 增大时，算法在 LFR 网络中不同模糊系数取值条件下的 NMI 值变化。通过观察可以发现，不同混合系数条件下的 NMI 值呈现上升、下降及平稳的变化模式。然而 NMI 变化曲线收敛于不同的稳定值，这是由于 LFR 基准网络在不同参数控制下的网络结构差异性导致。此外，通过实验可以发现，对于 LFR 网络，当模糊阈值  $\epsilon \in [0.2, 0.4]$  时，算法能取得最优的聚类结果。最后，实验进一步证明了本文算法的稳定性。

#### 5 结束语

本文提出了一种基于模糊聚类的多分辨率社区结构发现方法。首先引入了一种结构相似性以度量节点间的模糊关系，将传统社区结构中的确定性划

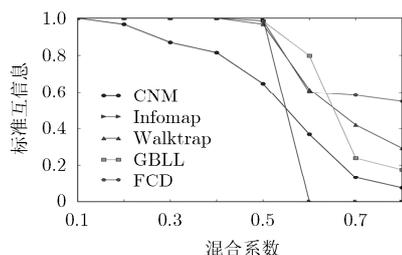


图3 LFR网络中各算法精确性比较

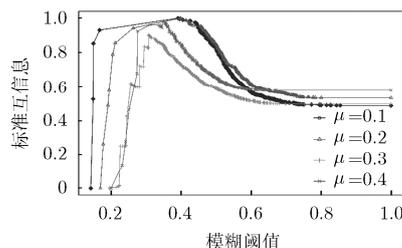


图4 LFR网络中模糊阈值对算法精确性的影响

分转化为模糊划分。通过调节模糊系数，本文所提的 FCD 算法实现了多分辨率的社区结构划分。同时，基于模糊特性，提出新的模块度函数以确定最优划分。通过对真实网络和人工合成网络的实验，并与其他社区发现算法进行了比较，证明了 FCD 算法能有效的揭示潜在的社区结构。算法所采用的结构相似性度量能有效刻画社区的局部特性，从而提高了算法的精确性和稳定性。

### 参考文献

- [1] 汪小帆, 李翔, 陈关荣. 网络科学导论[M]. 北京: 高等教育出版社, 2012: 1-27.  
WANG Xiaofan, LI Xiang, and CHEN Guanrong. Network Science: A Introduction[M]. Beijing: Higher Education Press, 2012: 1-27.
- [2] NEWMAN M E J. Complex systems: A survey[J]. *American Journal of Physics*, 2011, 79(8): 800-810. doi: 10.1119/1.3590372.
- [3] FORTUNAO S and DARKO H. Community detection in networks: A user guide[J]. *Physics Reports*, 2016, 659: 1-44. doi: 10.1016/j.physrep.2016.09.002.
- [4] ZHANG P, MOORE C, and NEWMAN M E J. Community detection in networks with unequal groups[J]. *Physical Review E*, 2016, 93(1): 012303. doi: 10.1103/PhysRevE.93.012303.
- [5] NEWMAN M E J. Communities, modules and large-scale structure in networks[J]. *Nature Physics*, 2012, 8(1): 25-31. doi:10.1038/nphys2162.
- [6] SCHAEFFER S E. Graph clustering[J]. *Computer Science Review*, 2007, 1(1): 27-64. doi: 10.1016/j.cosrev.2007.05.001.
- [7] MALLIAROS F D and VAZIRGIANNIS M. Clustering and community detection in directed networks: A survey[J]. *Physics Reports*, 2013, 533(4): 95-142. doi: 10.1016/j.physrep.2013.08.002.
- [8] CLAUSET A, NEWMAN M E J, and MOORE C. Finding community structure in very large networks[J]. *Physical Review E*, 2004, 70(6): 066111. doi: 10.1103/PhysRevE.70.066111.
- [9] LI Ming, DENG Youjin, and WANG Binghong. Clique percolation in random graphs[J]. *Physical Review E*, 2015, 92(4): 042116. doi: 10.1103/PhysRevE.92.042116.
- [10] LEE C, REID F, FCDAID A, et al. Detecting highly overlapping community structure by greedy clique expansion [C]. *Proceeding of 4th SNA-KDD Workshop on Social Network Mining and Analysis*, Washington DC, USA, 2010: 33-42.
- [11] AFSARMANESH N and MAGNANI M. Finding overlapping communities in multiplex networks[OL]. <https://arxiv.org/abs/1602.03746>. 2016.
- [12] YANG B and LIU J. Discovering global network communities based on local centralities[J]. *ACM Transactions on the Web*, 2008, 2(1): 1-32. doi: 10.1145/1326561.1326570.
- [13] NIKOLAEV A G, RAZIB R, and KUCHERIYA A. On efficient use of entropy centrality for social network analysis and community detection[J]. *Social Networks*, 2015, 40: 154-162. doi: 10.1016/j.socnet.2014.10.002.
- [14] NASCIMENTO M C V and CARVALHO A C. Spectral methods for graph clustering-A survey[J]. *European Journal of Operational Research*, 2011, 211(2): 221-231. doi: 10.1016/j.ejor.2010.08.012.
- [15] YANG J and LESKOVEC J. Defining and evaluating network communities based on ground-truth[J]. *Knowledge and Information Systems*, 2015, 42(1): 181-213. doi: 10.1007/s10115-013-0693-z.
- [16] XIANG Ju, HU Tao, ZHANG Yan, et al. Local modularity for community detection in complex networks[J]. *Physica A: Statistical Mechanics and Its Applications*, 2016, (443): 451-459. doi: 10.1016/j.physa.2015.09.093.
- [17] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, (10): P10008.
- [18] GOMEZ D, RODRIGUEZ J T, YANEZ J, et al. A new modularity measure for fuzzy community detection problems based on overlap and grouping functions[J]. *Approximate Reasoning*, 2016, 74: 88-107. doi: 10.1016/j.ijar.2016.03.003.

- [19] AHN Y Y, BAGROW J P, and LEHMANN S. Link communities reveal multiscale complexity in networks[J]. *Nature*, 2010, 466(7307): 761–764. doi: 10.1038/nature09182.
- [20] DING Z, ZHANG X, SUN D, *et al.* Overlapping community detection based on network decomposition[J]. *Scientific Reports*, 2016, 6: 24115. doi: 10.1038/srep24115.
- [21] HUANG Lan, WANG Guishen, WANG Yan, *et al.* A link density clustering algorithm based on automatically selecting density peaks for overlapping community detection[J]. *International Journal of Modern Physics B*, 2016, 30(24): 1650167. doi: 10.1142/S0217979216501678.
- [22] NEWMAN M E J and GIRVAN M. Finding and evaluating community structure in networks[J]. *Physical Review E*, 2004, 69(2): 026113.
- [23] PAN Y, LI D H, LIU J G, *et al.* Detecting community structure in complex networks via node similarity[J]. *Physica A: Statistical Mechanics and Its Applications*, 2010, 389(14): 2849–2857. doi: 10.1016/j.physa. 2010.03.006.
- [24] PAPAPOPOULOS F, KITSAK M, SERRANO M A, *et al.* Popularity versus similarity in growing networks[J]. *Nature*, 2012, 489(7417): 537–540. doi: 10.1038/nature11459.
- [25] RADICCHI F, CASTELLANO C, CECCONI F, *et al.* Defining and identifying communities in networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(9): 2658–2663. doi: 10.1073/pnas.0400054101.
- [26] XIE Jierui, KELLEY S, and SZYMANSKI B K. Overlapping community detection in networks: the state-of-the-art and comparative study[J]. *ACM Computing Surveys (CSUR)*, 2013, 45(4): Article No. 43. doi: 10.1145/2501654.2501657.
- [27] XU Rui and WUNSCH D. Survey of clustering algorithms[J]. *IEEE Transactions on Neural Networks*, 2005, 3(16): 645–678. doi: 10.1109/TNN.2005.845141.
- [28] 陈水利, 李敬功, 王向公. 模糊集理论及其应用[M]. 北京: 科学出版社, 2005: 1–448.  
CHEN Shuili, LI Jinggong, and WANG Xianggong. Fuzzy Set Theory and Its Application[M]. Beijing: Science Press, 2005: 1–448.
- [29] ZADEH L A. Toward a generalized theory of uncertainty (GTU)-an outline[J]. *Information Sciences*, 2005, 172(1): 1–40.
- [30] BARALDI A and BLONDA P. A survey of fuzzy clustering algorithms for pattern recognition I[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 1999, 29(6): 778–785. doi: 10.1109/3477.809032.
- [31] DANON L, DIAZ-GUILERA A, DUCH J, *et al.* Comparing community structure identification[J]. *Journal of Statistical Mechanics-Theory and Experiment*, 2005, (9): P09008.
- [32] NEWMAN M. Network data[OL]. <http://www-personal.umich.edu/~mejn/netdata>. 2016. 7.
- [33] LANCICHINETTI A, FORTUNATO S, and RADICCHI F. Benchmark graphs for testing community detection algorithms[J]. *Physical Review E*, 2008, 78(4): 046110. doi: 10.1103/PhysRevE. 78.046110.
- [34] PONS P and LATAPY M. Computing communities in large networks using random walks[C]. *Proceeding of 20th International Symposium on Computer and Information Sciences*, Turkey, 2005: 284–293. doi: 10.1007/11569596\_31.
- [35] ROSVALL M and BERGSTROM C T. Maps of random walks on complex networks reveal community structure[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(4): 1118–1123. doi: 10.1073/pnas.0706851105.
- 汪晓锋: 男, 1984年生, 博士生, 研究方向为复杂网络、社交网络分析、数据挖掘。
- 刘功申: 男, 1974年生, 博士, 副教授, 研究方向为自然语言处理、信息安全等。
- 李建华: 男, 1965年生, 博士, 教授, 研究方向为计算机通信网、信息安全等。