doi:10.11959/j.issn.1000-436x.2017028

基于多核心标签传播的复杂网络重叠社区识别方法

邓琨[1]，李文平[1]，余法红[1]，张健沛[2]
（1.嘉兴学院数理与信息工程学院，浙江 嘉兴 314001；2. 哈尔滨工程大学计算机科学与技术学院，黑龙江 哈尔滨 150001）

摘 要：针对传统基于标签传播的重叠社区识别方法存在较强的随机性，以及需要预设相关阈值来辅助完成社区识别等缺陷，提出基于多核心标签传播的重叠社区识别方法(OMKLP)。在分析节点度以及节点与邻居节点的局部覆盖密度后提出核心节点评价模型，并在此基础上给出局部核心节点识别方法；基于局部核心节点，提出新的面向重叠社区的异步标签传播策略，该策略能够快速地识别出社区内部节点与边界节点，以获得重叠社区结构；提出重叠节点分析方法，进一步提高识别重叠节点准确度。OMKLP 算法无需掌握任何先验知识，仅在掌握网络基本信息(点、边)基础上，便能够准确识别出重叠社区结构，从而有效解决了传统标签传播算法所存在的缺陷。在基准网络和真实网络上进行测试，并与多个经典算法进行对比分析，实验结果验证了所提算法的有效性和可行性。
关键词：复杂网络；社区识别；标签传播；重叠节点
中图分类号：TP391 **文献标识码**：A

Overlapping community detection in complex networks based on multi kernel label propagation

DENG Kun[1], LI Wen-ping[1], YU Fa-hong[1], ZHANG Jian-pei[2]
(1.College of Mathematics Physics and Information Engineering, Jiaxing University, Jiaxing 314001,China;
2.College of Computer Science and Technology, Harbin Engineering University, Harbin 150001，China)

Abstract: In view of the strong randomness and pre-setting the related threshold of traditional overlapping community detection method based on label propagation, overlapping community detection in complex networks based on multi kernel label propagation (OMKLP) was proposed. Evaluation model of kernel nodes was proposed after analyzing the node's degree and local covering density of nodes and their neighbor nodes. And on this basis, the detection method of local kernel nodes was also presented. Based on local kernel nodes, a new asynchronous label propagation strategy oriented to overlapping community was proposed, which can rapidly distinguish inner nodes and outer nodes of communities so as to obtain overlapping community structure. The analysis method of overlapping nodes was proposed to increase the accuracy of detecting overlapping nodes. Without any prior knowledge, only on the basis of the basic network information (nodes and links), the algorithm can detect the structure of overlapping communities accurately. Therefore, it effectively solved the defect of the traditional label propagation algorithm. The algorithm was tested over benchmark networks and real-world networks and also compared with some classic algorithms. The experiment results verify the validity and feasibility of OMKLP.
Key words: complex networks, community detection, label propagation, overlapping nodes

收稿日期：2016-06-16；**修回日期**：2016-12-11
通信作者：邓琨，dengkun@hrbeu.edu.cn
基金项目：国家自然科学基金资助项目（No.61672179, No.61370083, No.61402126）；教育部人文社会科学研究青年基金资助项目（No.15YJCZH088）；浙江省自然科学基金资助项目（No.LY16F020027）；浙江省教育厅科研基金资助项目（No.Y201636127, No.Y201533771）

Foundation Items: The National Natural Science Foundation of China (No.61672179, No.61370083, No.61402126), The Humanity and Social Science Youth Foundation of Ministry of Education of China (No.15YJCZH088), Zhejiang Provincial Natural Science Foundation of China (No.LY16F020027), Zhejiang Provincial Education Department Research Foundation of China (No.Y201636127, No.Y201533771)

1 引言

现实世界中的很多复杂系统可由复杂网络的形式表示出来，如人际关系网络、蛋白质作用网络和科学家合作网络等。通过对这些复杂网络的分析，发现其中存在小世界特性[1]、无标度特性[2]和社区结构特性[3]等复杂网络的基本统计特性。本文所研究的“社区结构特性”是指复杂网络中普遍存在着“社区内部连接紧密、社区之间连接松散”的特点。复杂网络社区识别旨在揭示出网络中真实存在的社区结构，其对于复杂网络的拓扑结构分析、功能分析和行为预测具有重要的理论意义和实际意义[3]，它不仅吸引了大量的学者进行研究，而且被广泛应用于蛋白质功能预测[4,5]、舆情分析与控制[6]和搜索引擎[7]等众多领域中，已成为当今研究的热点问题。

近年来，经典的社区识别算法层出不穷，如基于分裂的 GN 算法[8]、基于模块度优化的 FN 算法[9]、BGLL 算法[10]和基于标签传播的 LPA 算法[11]等，虽然这些算法比较优秀，但其仅能将复杂网络分解为若干个互不相连的社区，即网络中的每个节点仅属于唯一社区。但在真实网络中，各社区之间并不完全独立，可能是相互重叠的，即一些节点可同时属于多个社区，如在社会网络中，人们可以根据不同的分类（家庭、朋友、职业、爱好等）同时属于多个不同社区。由此，复杂网络中重叠社区识别通常更具实际意义。

至今，许多重叠社区识别算法已被提出。基于派系渗透的 CPM 算法[12]认为社区内部的边有较大可能形成大的完全子图，而社区之间的边几乎不可能形成较大的完全子图，因此，CPM 算法通过派系渗透的方式来识别网络中的社区结构，类似算法包括文献[13,14]。基于边聚类的 LC 算法[15]利用边通常存在单一角色且属于唯一社区的特性完成社区识别，该算法可直观理解为当边所属的社区确定后，相应的重叠节点将自然归属多个社区，完成社区识别任务，类似算法包括文献[16,17]。基于局部优化的 LFM 算法[18]分别从不同的种子节点出发，并对适应度函数进行优化，不断扩展不同种子所在的局部社区，最终得到全局社区结构，类似算法包括文献[19]。Cao 等[20]采用非负矩阵生成的模型来完成社区识别，在社区数量已知的情况下，采用归一化的对称非负矩阵对网络进行分解，在社区数量未知的情况下，应用贝叶斯对称非负矩阵来自动确定社区数量，对网络完成分解，其类似算法还包括文献[21,22]。

近年来，标签传播算法凭借简单、高效等优势，已得到普遍关注，并被应用于重叠社区识别领域，如 COPRA[23]、SLPA[24]等算法。虽然标签传播算法在重叠社区识别方面依然有效，但也存在着识别精度不稳定，以及需要预先设置阈值参数来辅助确定重叠节点等缺陷。

本文针对现有标签传播方法存在的不足，提出基于多核心标签传播的重叠社区识别(OMKLP, overlapping community detection in complex networks based on multi kernel label propagation)方法，该方法首先在考虑节点度以及节点与邻居节点的局部覆盖密度的基础上给出核心节点评价模型，并通过局部搜索找到局部核心节点，同时，将核心节点的邻居节点标签赋值为与核心节点相同的标签，以确保在标签传播的过程中，在局部范围内有更多的相同标签发出，使算法快速收敛；然后，通过分析每个节点的所有邻居节点归属系数最大值的标签，更新该节点的标签存储空间，完成标签传播操作，获得重叠社区结构；最后，采用无需预先设置参数阈值的重叠节点分析方法进一步确认重叠节点。该算法无需掌握任何先验知识，仅在掌握网络基本信息（点、边）基础上，完成重叠社区识别。其具体创新点主要包括以下 3 点。

1) 定义核心节点评价模型判断各节点在网络中的角色，并使用局部搜索策略找到核心节点。

2) 给出新的适用于重叠社区识别的异步标签更新策略以完成社区识别。

3) 提出无需预先设置参数阈值的重叠节点分析方法帮助整理确认重叠节点。

2 问题分析

最近，Steve 提出的 COPRA 算法将标签传播方法引入重叠社区识别领域，该方法首先为每个节点分配社区标签及相应的归属系数存入标签存储空间；然后根据每个节点的邻居节点标签更新该节点标签存储空间中所属社区标签及归属系数，其中，节点所属社区标签与对应的归属系数并不唯一，如果节点的某个所属社区标签对应的归属系数小于 $\frac{1}{v}$（v 是算法参数），那么删除该社区标签，如果节

点的所属社区标签对应的归属系数都小于$\frac{1}{v}$，那么将随机选择保留一个社区标签，删除其余的社区标签；最终存储空间中标签相同的节点为一个社区，重叠节点则拥有多个标签。虽然 COPRA 能够有效完成社区识别任务，但是其依然存在一些缺陷。在图 1 中可以明显看出，节点 a、b、c 应为一个社区，节点 d、e、f 应为一个社区，但在算法执行过程中，在节点 a 的标签存储空间中，所属社区标签及归属系数可表示为$(b,\frac{1}{3})$、$(c,\frac{1}{3})$、$(d,\frac{1}{3})$，若此时设置参数 $v=2$，因为 a 的所属社区标签对应的归属系数都小于$\frac{1}{2}$，因此，节点 a 的标签需要随机地在标签 b、c、d 中产生，从而 a 的社区标签有可能被更新为 d 标签，导致 a 和 d 在一个社区中。显然，这种较强的随机性将严重影响社区识别质量。

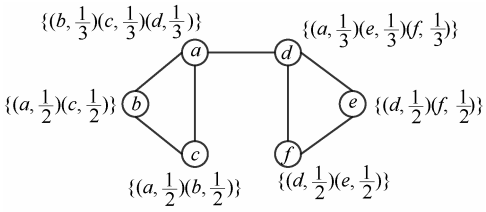


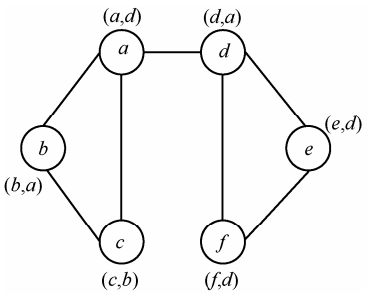
图 1　COPRA 算法识别社区示意

此后，Xie 等又提出了 SLPA 算法，该算法首先为每个节点初始化标签及标签存储空间；然后在迭代过程中，每个节点向邻居节点发出标签，同时从邻居节点随机接收一个标签存入标签存储空间；最终在每个节点的标签存储空间中，相同标签所占比例大于参数 r 的标签将保存下来，此时，若某节点的标签存储空间中标签数量大于 1，其对应的节点将被确定为重叠节点。图 2 所示为 SLPA 算法识别社区的过程，图 2(a)~图 2(d)分别为 SLPA 算法进行 4 次标签传播后可能出现的情况。可以看出，在图 2(d)中，若 r 值选择为 0.4，那么在标签存储空间中，标签所占比例小于 0.4 的标签将被删除，从而每个节点的标签存储空间中，仅剩下标签 d，因此，使整个网络形成一个社区，从而影响社区识别质量。

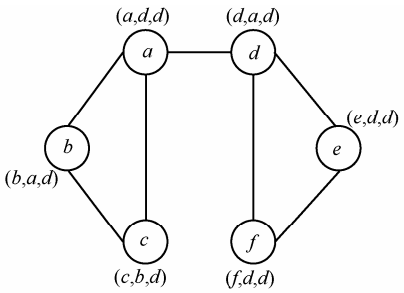
经过分析可知，现有基于标签传播的重叠社区识别算法，出现上述情况的原因有 2 点：1) 算法具有较强的随机性，影响社区识别质量；2) 需要预先设置参数，以帮助算法确定更新规则及确定重叠节点。而在纷繁复杂的网络中，掌握这些先验知识是极其困难的，在很多未知网络中也是不现实的。鉴于此，本文提出基于多核心标签传播的重叠社区识别方法。

3　OMKLP 算法

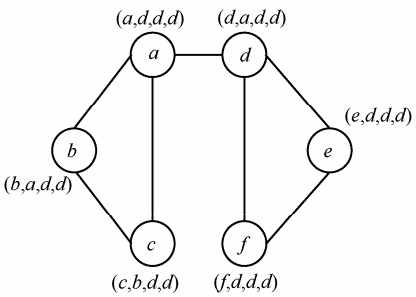
在分析了现有基于标签传播重叠社区识别算法所存在缺陷的基础上，本节给出基于多核心标签传播的重叠社区识别方法 OMKLP，该方法首先找到局部核心节点，并将核心节点的邻居节点标签赋



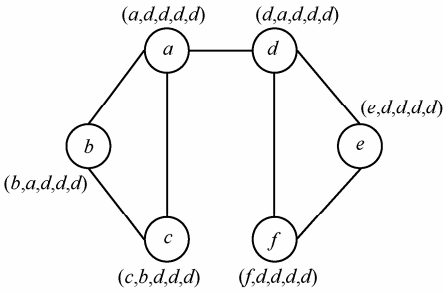
(a) 第1次标签传播后可能出现的结果



(b) 第2次标签传播后可能出现的结果



(c) 第3次标签传播后可能出现的结果



(d) 第4次标签传播后可能出现的结果

图 2　SLPA 算法识别社区示意

值为与核心节点相同的标签，确保在标签传播的过程中，在局部范围内有更多的相同标签发出，从而降低随机成分，达到快速收敛的目的；然后，采用新的适用于重叠社区识别的异步标签更新策略，识别出重叠社区；最后，通过重叠节点分析方法以准确地确定重叠节点，提高重叠社区的识别精度。其中，局部核心节点识别方法(KNDA, kernel nodes detection algorithm)、面向重叠社区的异步标签传播策略(ALPOC, asynchronous label propagation for overlapping community)以及重叠节点分析(AON, analysis of overlapping nodes)方法为 OMKLP 的核心，下面将对其进行详细阐述与分析。

3.1 局部核心节点识别方法

针对现有基于标签传播的重叠社区识别算法具有较强的随机性这一缺陷，OMKLP 算法首先搜索局部核心节点，并利用局部核心节点在局部范围内所具有的影响力将局部核心节点的邻居节点快速吸引，使局部范围内其他节点在收敛过程中更具目的性，从而降低算法的随机性。在此，给出局部核心节点识别算法 KNDA。

在通常情况下，局部核心节点被认为是在一个局部区域内拥有高节点度的节点[25]。但是，研究表明[26]，以节点最大度来判断某节点是否为局部核心节点的正确率仅为 67%，其原因在于节点度高的节点有可能是网络中的 hub 节点，而 hub 节点通常有较高的概率为边界节点。如图 3 所示，虽然节点 a 的度为 6，但是它处于 2 个社区之间，属于边界节点，而节点 b，虽然节点度为 5，但是与其连接的邻居节点之间连接紧密，从而是社区 A 的核心节点。因此，以传统节点最大度的方式来判断某节点是否为核心节点显然并不适用。

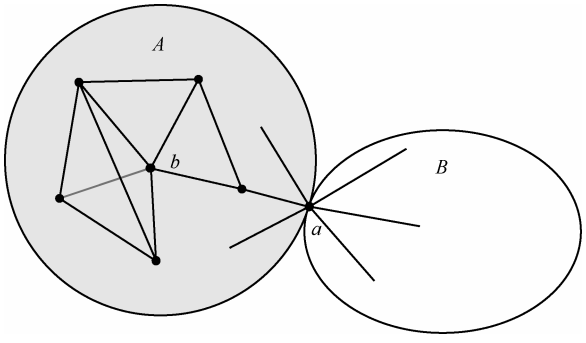


图 3 局部核心节点示意

综上，对于局部核心节点更直观的理解应该是节点度较高，并且与其连接的邻居节点之间也要有较紧密的连接，因此，从节点最大度以及节点与其邻居节点连接的紧密度（局部覆盖密度）两方面判断节点在网络中的角色显然更加合理。现将局部覆盖密度与评价节点重要性的局部核心节点评价值定义如下。

定义 1 （局部覆盖密度）若节点 v 是网络 G 中的一个节点，则节点 v 周围的局部覆盖密度可表示为

$$ED(v)=\frac{\sum_{t_1,t_2\in adj(v)\bigcup\{v\}}|e_{t_1,t_2}|}{|adj(v)\bigcup\{v\}|} \tag{1}$$

其中，$adj(v)$是 v 的邻居节点集合，如果节点 t_1 和 t_2 存在连接边，则$|e_{t_1,t_2}|$=1；否则，$|e_{t_1,t_2}|$=0。

由定义 1 可知，在一个节点和其邻居节点组成的集合中，若它们之间的关联边越多，则它们之间的连接越紧密，其局部覆盖密度值越高。在对社区结构的直观认识中可知，社区边界节点的局部覆盖密度不会很高，而社区中心节点往往具有较高的局部覆盖密度。

定义 2 （局部核心节点评价值）若节点 v 是网络 G 中的一个节点，k_v 是节点 v 的度，$ED(v)$ 是局部覆盖密度，则节点 v 的局部核心节点评价值可表示为

$$CV(v)=k_vED(v) \tag{2}$$

可以看出，局部核心节点评价值是从节点度与局部覆盖密度两方面来考虑节点在网络中的角色。以图 3 为例来说明局部核心节点评价值的合理性，若仅使用节点最大度来判断图 3 中节点 a 和 b 哪一个为局部核心节点时，有 $k_a=6$，$k_b=5$，因此，节点 a 将作为局部核心节点，而使用本文提出的局部核心节点评价值进行判断时，有 $CV(b)\approx 8.3$，$CV(a)\approx 5.14$，则节点 b 将作为局部核心节点。综上可知，使用局部核心节点评价值来判断局部核心节点可以有效避免使用节点最大度来判断局部核心节点所存在的局限。

基于局部核心节点评价值指标，局部核心节点识别算法 KNDA 被给出，具体过程如算法 1 所示。

算法 1 局部核心节点识别算法

输入 复杂网络 $G=(V,E)$

输出 局部核心节点集合

// $G=(V,E)$为复杂网络，V 为 G 的节点集合，E 为 G 的边集合

begin

1) while 存在未被搜索过的节点

2)　　RN←在未被搜索的节点中随机选取一个节点

3)　　CV←计算 RN 节点的 CV 值

4)　　maxCV←−1

5) while CV>maxCV

6)　　maxCV←CV

7)　　E_{RN}←RN 的邻居节点集合

8) while　$n_i \in E_{RN}$

9)　　CV_i←计算节点 n_i 的 CV 值

10)　　if　$CV_i > CV$

11)　　$CV \leftarrow CV_i$

12)　　$RN \leftarrow n_i$

13)　　end if

14)　　将节点 n_i 标记为已搜索

15)　end while

16)　end while

17)　$CNS \leftarrow CNS \cup RN$

18) end while

end

由算法 1 可知，KNDA 首先比较节点与其邻居节点的局部核心节点评价值，并沿着使该值增长的方向展开搜索，在一个区域(KNDA 算法在搜索任意一个局部核心节点的过程中，所遍历到的节点集合)中找到局部核心节点评价值最大的节点后，此节点被确定为局部核心节点，然后，在另一个区域继续搜索局部核心节点，直到网络中所有节点均被遍历过为止，最终，经过 KNDA 算法处理后，网络中局部核心节点评价值 CV 较大的节点将被搜索到。

3.2　面向重叠社区的异步标签传播策略

为提高 OMKLP 算法的标签传播能力，本文提出面向重叠社区的异步标签传播策略，在介绍该策略之前，先给出如下定义。

定义 3　（社区标签）设节点 v 归属社区 C，社区 C 的标识为 $L(C)$，若 $L(v)=L(C)$，则称 $L(v)$ 为节点 v 的社区标签。

定义 4　（归属系数）设节点 v 同时属于多个社区（C_1，C_2，⋯，C_r），r 为 v 同时归属的社区数量，则 v 属于社区 C_i (i=1,⋯,r)的归属系数 q 可定义为

$$q = \frac{w}{d_v} \tag{3}$$

其中，d_v 为节点 v 的度，w 表示节点 v 与社区 C_i 的连接边数。

定义 5　（社区内部节点）设 v 为社区 C 内一个节点($v \in C$)，$adj(v)$ 为 v 的邻居节点集合，若 $\forall t \in adj(v)$，如果 $t \in C$，则 v 为社区 C 的内部节点。

定义 6　（社区边界节点）设 m 为社区 C 内一个节点($m \in C$)，$adj(m)$ 为 m 的邻居节点集合，若 $\exists t \in adj(m)$，并且 $t \notin C$，则 m 为社区 C 的边界节点。

在给出上述定义后，现将面向重叠社区的异步标签传播策略 ALPOC 进行如下描述。

1) 为每个节点 v 初始化标签存储空间 S_v，初始标签存储空间可表示为：$S_v = \{(L(v), b(v))\}$，其中，$L(v)$ 为节点 v 的标签，初值为 v，$b(v)$ 为节点 v 属于标签为 $L(v)$ 社区的归属系数。

2) 网络中的节点 v 接收其每个邻居节点的最大归属系数所对应的社区标签，组成标签集合，并按式(4)更新节点 v 的标签存储空间 S_v。

$$S_v = \left\{ \left(x_i, \frac{count(x_i)}{|L_v|} \right) \right\}, x_i \in L_v \tag{4}$$

其中，L_v 为具有重复元素的集合，x_i 为 L_v 中不重复的元素，i=1,2,⋯,N，N 为不同元素的数量，$count(x_i)$ 表示 L_v 中 x_i 元素的个数。

3) 采用异步更新方法，重复执行步骤 2)，直到网络中任何节点 v 的存储空间 S_v 中所有社区标签都不发生变化，该策略终止执行。

4) 最终，社区标签相同的节点为一个社区，若标签存储空间中包含多个社区标签，则该节点为重叠节点。

在以上标签传播方式中，每个节点采用异步更新方式接收其所有邻居节点的标签，接收每个邻居节点的标签为该节点最大归属系数对应的社区标签，将接收的全部社区标签按式(4)进行计算，根据计算结果更新节点标签存储空间中的社区标签及归属系数。需要说明的是，虽然同步更新策略相较于异步更新策略在稳定性上略有优势，但在迭代次数上异步更新策略要少于同步更新策略[11]，同时，在面对二部网络(在一个无向网络中，若顶点集可分割为 2 个互不相交的子集，且网络中每条边所关联 2 个顶点分别属于 2 个不同顶点集)及星型结构网络时，异步更新产生的振荡效应要低于同步更新策略[27]。综上，从运行效率及普适性的角度来说，异步更新

策略往往更有优势，因此，ALPOC 策略采用异步更新方式完成标签传播任务。

通过分析可知，ALPOC 策略在核心节点影响力的驱动下，社区内部节点仅受到社区内节点标签的影响，更容易取得相同社区标签，因此，针对社区内部节点，该策略类似于传统的 LPA 算法，能够快速收敛，而社区的边界节点由于受到来自不同社区标签的影响会形成重叠节点。可以看出，ALPOC 策略可以将社区内部节点及边界节点快速区分，从而找到重叠社区结构。需要指出，若将所有的边界节点都作为重叠节点，显然并不合理，但在 OMKLP 算法中，并不鼓励使用设置阈值的方式去消除某些归属系数小的标签，其目的是在复杂的网络环境中，并不盲目决定某些边界节点是否为重叠节点，以保证社区识别结果更为准确。鉴于此，本文提出重叠节点分析方法。

3.3 重叠节点分析

经过分析可知，现有基于标签传播的重叠社区识别算法通常要预先设置相关参数以帮助算法确定重叠节点，而在纷繁复杂的网络中，准确掌握先验知识，精确设置参数几乎是无法完成的。鉴于此，本文提出一种在无需掌握任何先验知识的情况下，能够准确识别重叠节点的重叠节点分析方法。在给出重叠节点分析算法之前，进行如下定义。

定义 7 （边归属值）若 u 、 v 为网络 G 中的 2 个节点， B_{u-C} 、 B_{v-C} 分别为节点 u 、 v 归属于社区 C 的归属系数，则边 $e_{u,v}$ 归属于社区 C 的归属值可表示为

$$B_{e_{u,v}} = \sqrt{B_{u-C}B_{v-C}} \tag{5}$$

定义 8 （社区局部度量）设 $ind(C)$ 为社区 C 的边界节点与 C 内节点的连接边数， $outd(C)$ 为社区 C 的边界节点与 C 的外部节点连接的边数，此时，针对社区 C 的局部度量 D 可表示为

$$D = \frac{ind(C)}{outd(C)} \tag{6}$$

由定义 8 可知，社区局部度量是边界节点与社区内部的连接数和边界节点与社区外部连接数的比例，其中，若边界节点与社区内部连接越紧密，则社区局部度量越高，反之，社区局部度量越低，因此，这也反映出边界节点在局部环境下，对归属某一社区的倾向性。在此基础上，给出社区局部增量定义。

定义 9 （社区局部增量）若节点 a 为一个重叠节点，D 为 a 不属于社区 C 时的社区局部度量，D_o 为 a 属于社区 C 时的社区局部度量，此时，社区局部增量 ΔD 可表示为

$$\Delta D = D_a - D \tag{7}$$

显然，若 $\Delta D > 0$ ，则说明将节点 a 加入社区 C ，会使社区 C 的内部紧密度增加，若 $\Delta D < 0$ ，则说明将节点 a 加入社区 C ，会使社区的紧密度降低。因此，通过社区局部增量可以判断出节点在加入某社区时，其是否对社区紧密度的增加存在贡献。

基于社区局部度量与社区局部增量的定义，本文提出重叠节点分析方法，具体算法描述如算法 2 所示。

算法 2 重叠节点分析方法

输入 O //执行 ALPOC 策略后给出的重叠节点集合

输出 S //执行 AON 算法输出的重叠节点集合

begin

1) for each $i \in O$

2) $E_i \leftarrow$ 与节点 i 关联的边集

3) for each $e_{ij} \in E_i$

4) $L(e_{ij}) \leftarrow e_{ij}$ 的最高边归属值对应的社区标签

5) $L(E_i) \leftarrow L(E_i) \bigcup L(e_{ij})$

6) end for

7) $S_i \leftarrow$ 从节点 i 的标签存储空间 S_i 中删除未包含在 $L(E_i)$ 中的标签

8) if S_i 中拥有多个标签

9) $T \leftarrow$ 计算节点 i 与 S_i 中标签对应社区的社区局部增量 ΔD 值

10) $S_i \leftarrow$ 从 T 中删除 ΔD 为负值时所对应的社区标签，若 T 中的 ΔD 全为负值，则保留最大 ΔD 值所对应的社区标签

11) end if

12)end for

end

由算法 2 可知，AON 算法首先计算与重叠节点相连接每条边的边归属值，并将每条边归属于边归属值最大的社区，同时，对其赋予社区标签，然后从重叠节点的标签存储空间中删除与该节点的连接边中未出现的社区标签，随后，对剩余的重叠节点进行社区局部度量增量分析，以判断重叠节点是

否对增加所属社区连接紧密度存在贡献，若社区局部度量增量值为负，则从标签存储空间中删除该社区标签。在最终的社区结构中，标签相同的节点为一个社区，标签存储空间中存在多个标签的节点为重叠节点。

综上，AON 方法首先从网络中点与边的原始角度分析某重叠节点是否属于某社区，以便能够自然地去除某些重叠节点，避免在计算社区局部增量值时，出现振动现象。然后，从社区连接紧密度角度分析重叠节点对某社区的贡献以准确找到重叠节点。

3.4　OMKLP 算法描述

经过上述分析，OMKLP 算法的具体过程如下。

1) 为网络中每个节点初始化标签。

2) 通过 KNDA 算法搜索局部核心节点。

3) 局部核心节点的邻居节点标签更新为局部核心节点的标签。

4) 采用 ALPOC 标签传播策略更新标签。

5) 运用 AON 算法分析重叠节点，完成社区识别。

OMKLP 算法首先为每个节点初始化唯一标签，并通过 KNDA 算法找到局部核心节点，并将局部核心节点的邻居节点标签更新为局部核心节点的标签。其目的在于，利用局部核心节点在局部所具有的影响力将其邻居节点快速吸引，以形成最初的社区，以此在局部范围内发出更多的相同标签信号，确保算法快速收敛，降低算法的随机成分。经过上述操作后，OMKLP 算法采用 ALPOC 标签传播策略进行标签更新，该策略在已经形成的以局部核心节点为中心的小区域中进行标签传播，由于局部范围内已经存在发出较多相同标签的信号源，因此，社区内部节点较容易达成共识，而社区边界节点由于受到不同标签的影响，可能成为重叠节点。为使识别的重叠节点更为准确，OMKLP 算法运用 AON 算法分析整理重叠节点，该算法首先采用点与边的网络原始要素分析某重叠节点是否属于某社区，去除容易辨别的重叠节点，避免在计算社区局部增量值时出现干扰，然后，该方法使用社区局部增量值的方式分析现有的每个重叠节点对所属社区的贡献，以准确识别重叠节点。经过以上操作，最终完成了社区识别任务。

3.5　算法时间复杂度分析

设网络 G 中包括 n 个节点，k 为节点的平均度，c 为网络 G 中局部核心节点数，t 为 ALPOC 标签更新策略的迭代次数，o 为在标签传播过程中所识别的重叠节点数。以下为 OMKLP 算法的时间复杂度分析。

显然，OMKLP 算法为每个节点初始化标签的时间复杂度为 $O(n)$；由于在 KNDA 算法中，计算每个节点 $ED(V)$ 值所需的最坏时间复杂度为 $O\left(\frac{k(k-1)}{2}\right)$，每个节点与其邻居节点 $ED(V)$ 值比较的时间复杂度为 $O(k)$，从而计算 n 个节点的 $ED(v)$ 值及进行比较的时间复杂度为 $O\left(\frac{k(k-1)}{2}+k\right)n$，因此，KNDA 算法的最坏时间复杂度也不会超过 $O(k^2n)$；将局部核心节点的邻居节点标签更新为与核心节点相同的社区标签，其时间复杂度为 $O(kc)$；在标签更新策略中，由于需要 t 次迭代，同时，每个节点需要接收邻居节点社区标签，因此，标签传播的时间复杂度为 $O(tkn)$。因为在 AON 算法中，需要对 o 个重叠节点进行分析，首先分析每个重叠节点的连接边，时间复杂度为 $O(ko)$，然后分析每个重叠节点的社区局部增量值，其时间复杂度不会超过 $O(ko)$。最终，OMKLP 算法的时间复杂度为 $O(n+k^2n+kc+ktn+2ko)$。考虑到 c、o 远小于网络中的节点数 n，由此，OMKLP 算法的时间复杂度可表示为 $O(jk^2n)$，其中，j 为常数。

在此，选取 LFM[18]、LC[15]、COPRA[23]和 SLPA[24]等经典重叠社区识别算法与OMKLP算法进行时间复杂度对比分析，以展示各算法的运行效率。从表 1 中可知，除 LFM 算法的时间复杂度为 $O(n^2)$外，其他算法的时间复杂度都接近于线性。表 1 中 n、k、t 与本节符号一致，s 表示网络中的社区数量，m 为网络中的边数，$k_{\max}$ 为网络中最大节点度。

表 1　　各算法的时间复杂度对比

序号	算法	时间复杂度
1	LFM	$O(n^2)$
2	COPRA	$O\left(sm\log\left(\frac{sm}{n}\right)\right)$
3	LC	$O(nk_{\max}^2)$
4	SLPA	$O(tm)$
5	OMKLP	$O(jnk^2)$

4　实验

为了测试 OMKLP 算法的性能，算法在基准网

络数据集和真实网络数据集上进行测试，并与经典算法 CFINDER[12]、LFM、LC、COPRA 和 SLPA 进行对比分析，以验证其有效性与可行性。

4.1 评价指标

为评价各算法的性能，本文采用 3 个经典的评价指标，分别从社区识别精度、重叠节点识别精度、社区连接紧密度 3 个方面来衡量各算法的优劣。

1) 社区识别精度的评价指标为扩展统一化互信息(NMI, normalized mutual information)[18]评价指标。*NMI* 的取值范围为 0~1，如果 *NMI* 取值是 1，则识别的社区结构与真实社区结构完全一致；如果 *NMI* 取值是 0，则识别的社区结构与真实社区结构是截然不同的。也可以说算法识别社区结构的准确率越高，则 *NMI* 值越大；否则 *NMI* 值越小。

2) 重叠节点识别精度的评价指标为 *F-score*[28]。*F-score* 的取值范围为 0~1，即算法识别重叠节点的准确率越高，则 *F-score* 取值越大。

3) 社区连接紧密度的评价指标为扩展模块度函数（EQ，extend Q）[13]。扩展模块度代表的含义是 *EQ* 值越高，其识别的社区结构越紧密，否则，社区结构越松散。

4.2 基准网络数据集

由于 LFR benchmark 基准网络[29]的节点度与社区规模之间存在幂率分布统计特性，与真实复杂网络的分布状况极其相似，因此，本文使用该数据集作为本文所提算法与其他对比算法的测试数据集。

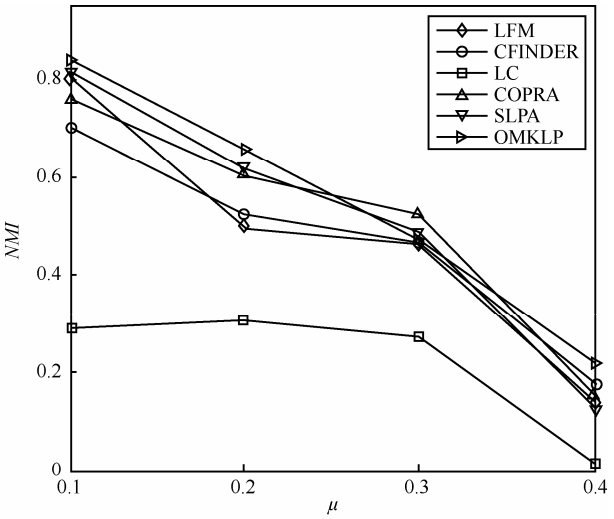
LFR benchmark 基准网络包括以下参数：N 表示网络中节点总数；k 表示节点平均度；$k_{\max}$ 代表节点最大度；$C_{\max}$ 与 $C_{\min}$ 分别为最大社区节点数及最小社区节点数；μ 为混合比例数，若 μ 值越小，则社区结构越清晰，否则社区结构越模糊；O_n 表示重叠节点数；O_m 代表重叠节点最多可同时归属的社区数。在此，基准网络的共享参数为 N=200，k =10，$C_{\min}$=20，$C_{\max}$=50，$k_{\max}$=30，其他参数设置如表 2 所示。

表 2 LFR 基准网络参数设置

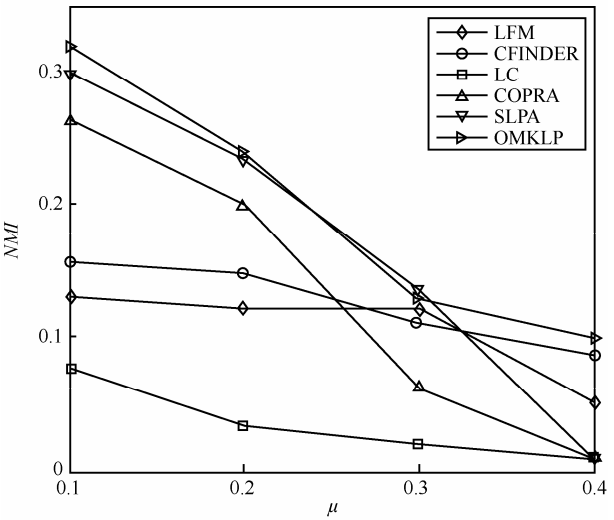
网络	O_n	O_m	μ
R1	20	2	0.1~0.4
R2	100	2	0.1~0.4
R3	20	2~6	0.1
R4	20	2~6	0.3

4.2.1 算法的识别精度分析

图 4 中给出了分别在低重叠度网络(R1)以及在高重叠度网络(R2)中，各算法在 *NMI* 方面的比较结果，在与 CFINDER 和 LFM 算法比较过程中可知，虽然 μ 值的逐渐增加对 2 种算法的识别精度影响较小，但整体来看，2 种算法的识别精度要低于 OMKLP 算法。在与 LC 算法的比较中可以看出，LC 算法的识别精度都远低于其他算法，其原因是由于 LC 算法在识别社区的过程中产生了过多的重叠节点，因此，影响社区识别质量。在与 SLPA、COPRA 算法的比较中可知，虽然 SLPA、COPRA 算法在初始时识别的社区精度较高，但随着社区结构逐渐模糊，其算法的识别精度下降过快，因此说明 SLPA 和 COPRA 算法并不稳定，而 OMKLP 算法在社区识别的过程中虽然也受到了 μ 值增加的



(a) 针对R1网络的识别结果

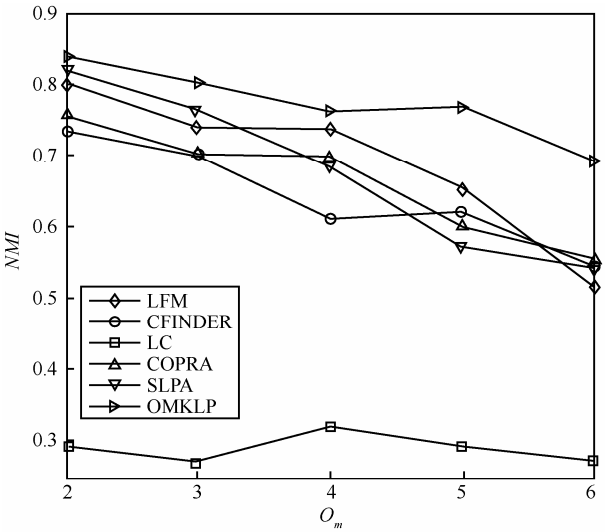


(b) 针对R2网络的识别结果

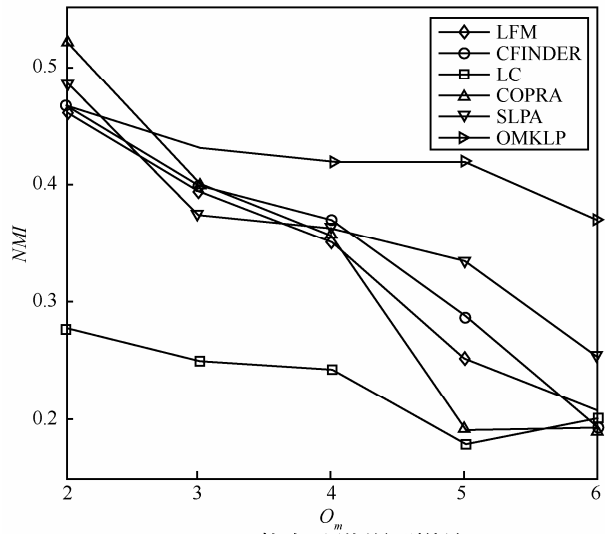
图 4 μ =01~0.4 时，各算法识别精度比较结果

影响，但其下降速度要慢于 SLPA 和 COPRA 算法。综上可知，在社区识别精度方面，随着社区结构越来越模糊，对 SLPA 与 COPRA 算法的影响最大，随后是对 OMKLP 算法有较大影响，对 CFINDER、LFM 和 LC 算法影响最小，但从整体识别精度来看，OMKLP 算法优于其他对比算法。

图 5 展示了分别在社区结构较清晰的网络(R3)与社区结构较模糊的网络(R4)中，各算法的对比结果。从图 5 中能够看出，在与 LC 算法的比较过程中，其算法的识别精度远低于 OMKLP 算法。在与 CFINDER、LFM、COPRA 和 SLPA 算法的比较过程中可知，虽然各对比算法初始时社区识别精度较高，但其都受到了重叠节点同时归属社区数 O_m 增加的影响，使其他对比算法的识别精度降低较快，



(a) 针对R3网络的识别结果



(b) 针对R4网络的识别结果

图 5　节点所属社区数 O_m=2~6 时，各算法识别精度比较结果

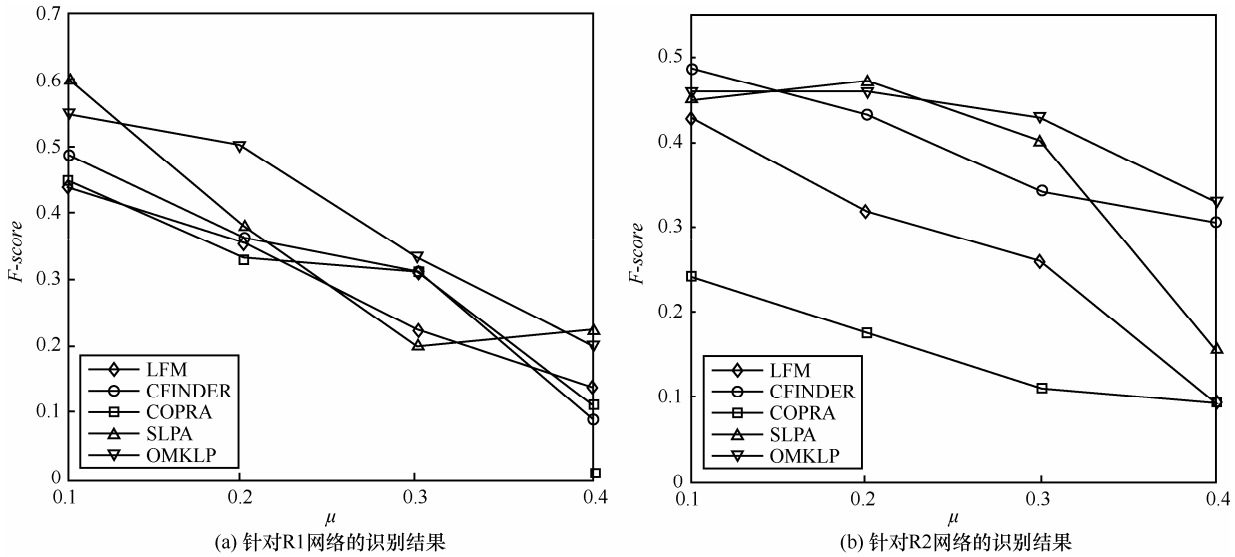
而 OMKLP 算法，随着 O_m 的增加对其识别精度影响不大。因此，在社区识别精度方面，重叠节点同时归属社区数对 OMKLP 算法不构成太大影响。

综上，在基准网络数据集中，随着 O_m (重叠节点最多可同时归属的社区数)的增加，对 OMKLP 算法的识别精度影响不大，而随着 μ 值(混合比例数)的增加，OMKLP 算法在识别精度方面出现了较大振幅。需要指出的是，由于标签传播类算法在更新节点标签时，是在分析其邻居节点标签的基础上完成的，因此，此类算法在识别精度方面对社区内部节点间的连接紧密程度较为敏感，从而随着 μ 值的增加，在社区结构逐渐变得模糊时，SLPA、COPRA、OMKLP 算法均受到较大影响，从实验结果可知，OMKLP 算法在识别稳定性上相较于 SLPA 和 COPRA 等传统算法，也得到了较大提高。虽然随着 μ 值增加 OMKLP 在识别准确率上出现一些振动，但在整体识别精度上，OMKLP 算法相较于其他对比算法也展示出较好的识别精度。

4.2.2　重叠节点识别精度分析

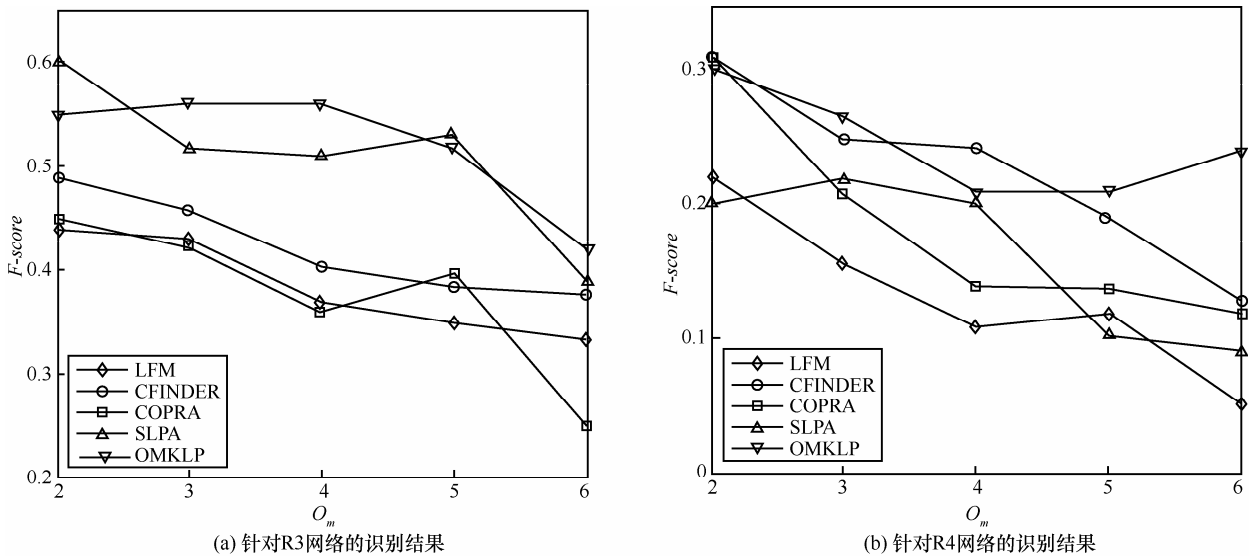
在对各算法进行了社区识别精度测试的基础上，本节将对它们的重叠节点识别能力进行分析。图 6 给出了在低重叠度网络(R1)以及在高重叠度网络(R2)中，各算法在 *F-score* 方面的比较结果，由于 LC 算法在识别社区的过程中，将所有节点全部作为了重叠节点，因此，无法评价其识别重叠节点的精度，故本节仅将 OMKLP 算法与其他 4 种算法进行对比。由图 6 可知，虽然 CFINDER、COPRA 与 LFM 算法识别精度较为稳定，但其识别精度普遍低于 OMKLP 算法。在与 SLPA 算法的比较中可知，尽管在社区结构较为清晰的网络中，SLPA 识别重叠节点的精度较高，但随着社区结构逐渐模糊，其识别精度下降过快，因此，在多数情况下，SLPA 的识别精度低于 OMKLP 算法。

图 7 中列出了在社区结构较清晰的网络(R3)与社区结构较模糊的网络(R4)中，各算法在 *F-score* 方面的比较结果。由图 7(a)可知，CFINDER、LFM 和 COPRA 算法在社区结构较清晰的网络中，识别重叠节点的稳定性较好，但识别精度低于 OMKLP 算法。虽然在 O_m=2 时，SLPA 算法识别重叠节点精度较高，但在 O_m 不断变化时，该算法的识别精度波动较大，在多数情况下低于 OMKLP 算法。由图 7(b)可知，由于受到模糊社区以及重叠节点同时



(a) 针对R1网络的识别结果　　(b) 针对R2网络的识别结果

图 6　μ =01~0.4 时，各算法重叠节点识别精度比较结果



(a) 针对R3网络的识别结果　　(b) 针对R4网络的识别结果

图 7　节点所属社区数 O_m=2~6 时，各算法重叠节点识别精度比较结果

归属社区数量不断增加的影响，各算法识别重叠节点的精度波动较大，由图 7 可以看出，各算法在社区结构较为模糊的网络中，对重叠节点的识别准确率都有较大影响。但也不难看出，OMKLP 算法相对于其他算法，依然有较高的 *F-score* 值。

4.3　真实网络数据集

由于基准网络与真实网络的拓扑特性稍有不同，本节通过真实网络数据集进一步测试 OMKLP 算法的性能与识别社区的合理性。在此，通过 *EQ* 指标来衡量各算法识别社区的质量。表 3 列出了用于测试各算法的 9 种真实网络数据集，其中包含几十个节点的小规模网络以及几万个节点的大规模网络。

表 3　　真实网络数据集

网络	节点	边	平均度	描述
Karate	34	78	4.59	空手道俱乐部网络[30]
Dolphins	62	159	5.13	海豚社会网络[31]
Lesmis	77	254	6.6	悲惨世界关系网络[32]
Polbooks	105	441	8.4	美国政治之书网络[33]
Email	1 133	5 451	9.62	电子邮件交往网络[34]
Polblogs	1 490	19 022	25.53	美国大选博客网络[35]
Netscience	1 588	2 742	3.45	作者合作网络[36]
PGP	10 680	24 316	4.55	信任网络[37]
Internet	22 963	48 436	4.22	互联网快照网络[35]

表 4 给出了 OMKLP 算法及其他对比算法针对

表 4　　**各算法的 *EQ* 值比较结果**

网络	OMKLP	CFINDER	LFM	COPRA	SLPA	LC
Karate	0.367 9	0.107 2	0.214 6	0.323 9	0.347 2	0.122 0
Dolphins	0.519 1	0.288 5	0.237 4	0.420 6	0.387 9	0.151 4
Lesmis	0.433 8	0.185 5	0.481 2	0.477 9	0.320 9	0.248 4
Polbooks	0. 484 2	0.430 4	0.347 6	0.458 6	0.456 8	0.219 8
Email	0.307 9	0.264 1	0.182 2	0.352 3	0.183 7	0.026 1
Polblogs	0.196 3	—	0.192 2	—	0.007	0.015
Netscience	0.910 9	0.590 5	0.259 9	0.718 6	0.712 6	0.719 2
PGP	0.700 8	—	—	0.433 5	0.692 8	0.160 1
Internet	0.185 3	—	0.195 8	0.034 5	0.116 1	0.023 4

9 个真实网络得到的 *EQ* 值的结果(“—”表示算法识别社区失败或所得 *EQ* 值小于 0.001)。从中可知，OMKLP 算法在 Karate、Dolphins、Polbooks、Polblogs、PGP、Netscience 这 6 个网络中所取得的 *EQ* 值为第一；在 Email、Internet 这 2 个网络中取得的 *EQ* 值为第二，其在 Email 网络中所取得的 *EQ* 值低于 COPRA 算法，在 Internet 网络中取得的 *EQ* 值低于 LFM 算法；在 Lesmis 网络中取得的 *EQ* 值为第三，仅低于 LFM 和 COPRA 算法。综上可知，OMKLP 算法在真实网络中依然具有较强的社区识别能力，其适用的网络更加广泛。

下面给出 OMKLP 针对 Karate、Dolphins、Lesmis 和 Polbooks 这 4 个网络所得的社区识别结果，以分析算法执行的合理性。图 8 展示了面对 Karate 网络执行 OMKLP 算法所得到的社区结构。Karate 是美国空手道俱乐部网络，该网络是针对一个美国大学空手道俱乐部进行 2 年观察分析而构建的。俱乐部中有 34 个成员作为节点，成员之间的友谊关系作为连接 2 个节点的边。由于出现分歧，该俱乐部最终分为以校长(34 节点)和教练(1 节点)为中心的 2 个新俱乐部。从图中可以看出，OMKLP 算法将 Karate 网络分成了 3 个社区，其整体上将节点 1 为中心的社区分成了 2 个更小、更紧密的社区，而节点 3 由于和两边的社区都有较为紧密的连接，因此，确定为重叠节点。在对 Karate 网络的一些讨论中[30]可知，成员 3 分别在 2 个俱乐部都有较多朋友，该成员具体属于哪一俱乐部是较难决定的。因此，将其作为重叠节点也相对合理，同时，由于 OMKLP 算法能够发现更小的社区结构，所以有助于从更多角度研究社区结构。

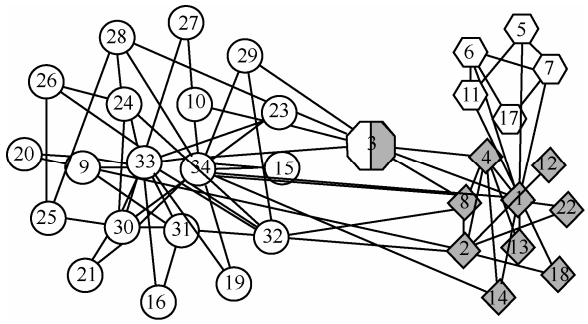


图 8　OMKLP 算法面对 Karate 所得到的社区结构

图 9 给出了 OMKLP 算法面对 Dolphins 网络所得到的社区结构。Dolphins 网络是对居住在神奇湾的 62 只海豚进行观察而建立的。其中，每个海豚代表一个顶点，若 2 只海豚经常联系，则它们之间就有一条边连接。从图中可知，OMKLP 将 Dolphins 右边的一个大的社区划分成了 3 个更小的社区(以正方形、菱形和圆表示的社区)，由于节点 31 与节点 37 分别与 2 个社区都有较紧密的联系，因此，将它们确定为重叠节点。

图 10 给出了 OMKLP 算法面对 Lesmis 网络所得到的社区结构。Lesmis 网络是雨果的小说《悲惨世界》中基于 77 个不同人物在相同章节中是否共同出现而构成的。77 个节点表示小说中的人物，边表示 2 个人物在相同章节中共同出现过，共 254 条边。可以看出，OMKLP 算法将 Lesmis 网络划分成了 5 个社区，这与小说情节的社会划分非常相似，由此说明 OMKLP 算法针对 Lesmis 网络的划分结果非常合理，而节点 24、32、40、50、52、56 由于在不同章节中出现，而且与不同章节中的人物有紧密的联系，因此，确定为重叠节点。

图 11 给出了 OMKLP 算法面对 Poolbooks 网络

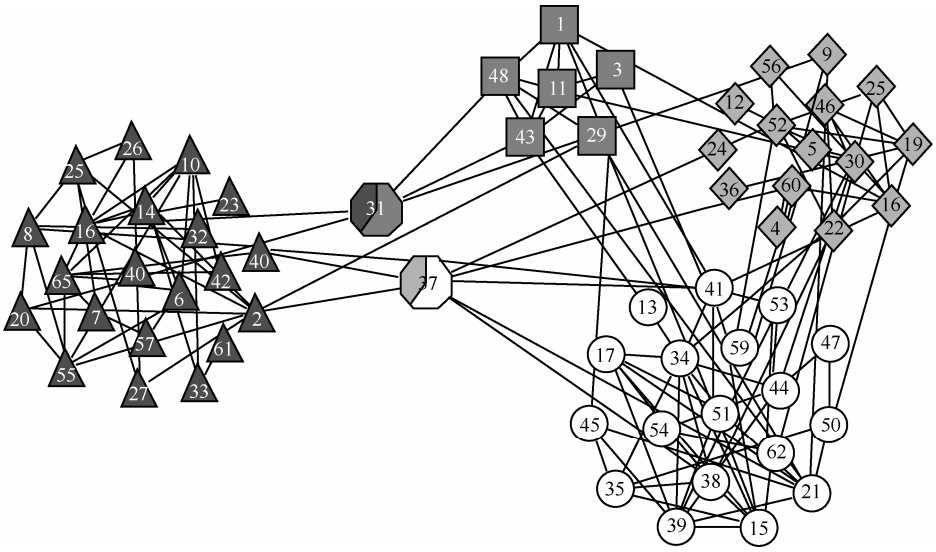


图 9 OMKLP 算法面对 Dolphins 所得到的社区结构

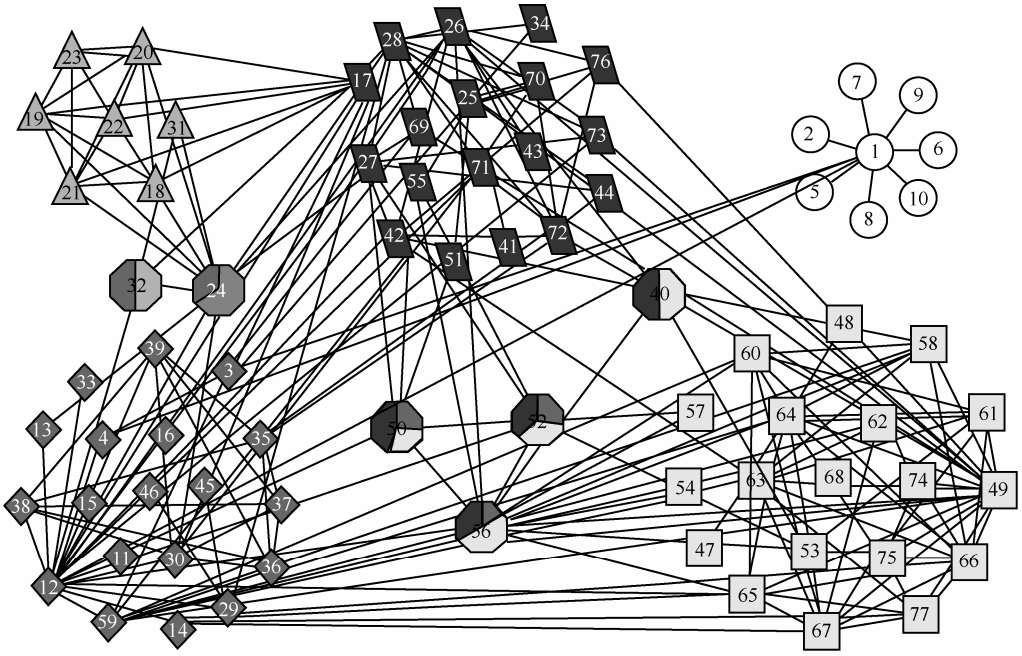


图 10 OMKLP 算法面对 Lesmis 网络所得的社区结构

所得到的社区结构。Polbooks 是美国政治之书网络，该网络的节点代表在亚马逊网站上销售的美国政治书籍，共 105 本书，边代表同一顾客共同购买的书籍，共 441 条边。这些书籍的政治观点分别是进步派、中立派和保守派。从中可知，OMKLP 算法将 Polbooks 网络近乎完美地划分了 3 个社区，由于节点 8、15、50、86 这几本书籍的政治观点不是非常清晰，因此将其作为重叠节点。

5 结束语

本文针对现有基于标签传播的重叠社区识别方法所存在随机性较强，导致社区识别结果不稳定，以及需要预先设置阈值参数来辅助完成重叠社区识别任务等缺陷，提出了基于多核心标签传播的重叠社区识别方法。该方法首先在分析节点度与节点的局部覆盖密度的基础上，给出局部核心节点识别方法，进而将找到的核心节点的邻居节点标签赋值为与核心节点相同的标签，以确保在标签传播的过程中，局部范围内有更多的相同标签发出，从而降低随机成分；然后，采用新的适用于重叠社区识别的异步标签更新策略，识别出重叠社区；最后，通过重叠节点分析方法以准确地确定重叠节点，提

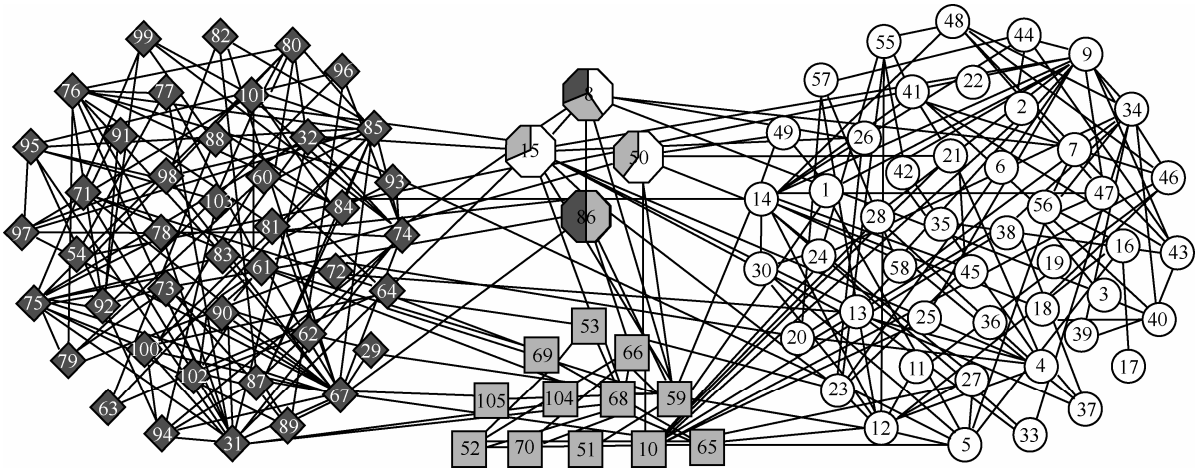


图 11　OMKLP 算法面对 Polbooks 网络所得到的社区结构

高重叠社区的识别精度。需要说明，OMKLP 算法无需预先设置任何参数，仅在掌握点与边等原始网络信息的基础上即可完成重叠社区识别，因此，该算法更具普适性。

在基准网络数据集中，通过对 *NMI* 评价指标的分析可知，μ 值的增加，对 OMKLP 算法有较大影响，但其整体的社区识别精度依然高于其他对比算法，需要指出，虽然 OMKLP 算法受 μ 值影响较大，但其较之 COPRA 与 SLPA 等标签传播算法更为稳定。在重叠节点归属社区数 O_m 不断增加时，OMKLP 相对于其他对比算法的识别精度更为稳定。此外，通过对 *F-score* 评价指标的分析可知，OMKLP 算法在多数情况下均优于其他对比算法，这也验证了 OMKLP 算法较为稳定，而且识别重叠节点能力较强。在对 9 个真实网络进行实验分析的过程中，OMKLP 算法仅在 Lesmis 网络、Email 网络和 Internet 网络没有取得 *EQ* 最大值，在其余 6 个网络中均取得与对比算法的最优结果。通过分析社区结构划分的合理性方面可知，OMKLP 算法在面对不同网络时均能较合理地分解社区，并且所分解的社区结构往往更小。综上，OMKLP 算法虽然会受到社区结构模糊的影响，使算法识别精度(*NMI*)的振幅较大，但在包括 *NMI*、*F-score* 和 *EQ* 的整体评价中，OMKLP 算法在与其他算法比较中均展现出良好的性能，因此，验证了 OMKLP 算法是有效且可行的。

未来的研究工作将集中于：在对节点进行标签传播的基础上，将边元素加入其中，由于边同时连接 2 个节点，相较于节点的标签更新，边标签的更新更为稳定，以便进一步提高算法的稳定性，解决算法在社区结构模糊的网络中识别精度振幅较大的问题。

参考文献：

[1] WATTS D J, STROGATZ S H. Collective dynamics of “small-world” networks[J].Nature, 1998, 393(84):440- 442.

[2] BARABÁSI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439):509-512.

[3] 金弟，刘大有，杨博，等. 基于局部探测的快速复杂网络聚类算法[J]. 电子学报, 2011, 39(11): 2540-2546.
JIN D, LIU D Y, YANG B, et al. Fast complex network clustering algorithm using local detention[J]. Acta Electronica Sinica, 2011, 39(11): 2540-2546.

[4] WANG Z, ZHANG J. In search of the biological significance of modular structures in protein networks[J]. Plos Computational Biology, 2007, 3(6): 1011-1021.

[5] FARUTIN V, ROBISON K, LIGHTCAP E, et al. Edge-count probabilities for the identification of local protein communities and their organization[J]. Proteins Structure Function and Bioinformatics, 2006, 62(3): 800-818.

[6] QIAN C, CAO J D, LU J Q, et al. Adaptive bridge control strategy for opinion evolution on social networks[J]. Chaos：An Interdisciplinary Journal of Nonlinear Science, 2011, 21(2): 025116.

[7] SIDIROPOULOS A, PALLIS G, KATSAROS D, et al. Prefetching in content distribution networks via web communities identification and outsourcing[J].World Wide Web, 2008, 11(1): 39-70.

[8] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2): 026113.

[9] NEWMAN M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E, 2004, 69(6): 066133.

[10] 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: 10008.

[11] RAGHAVAN U N, ALBERT R, KUMARA S. Near linear time algorithm to detect community structures in large-scale networks[J]. Physical Review E, 2007, 76(3): 036106.

[12] PALLA G, DERÉNYI I, FARKAS I, et al. Uncovering the overlapping

community structure of complex networks in nature and society[J]. Nature, 2005, 435(7043): 814-818.

[13] SHEN H W, CHENG X Q, GUO J F. Quantifying and identifying the overlapping community structure in networks[J].Journal of Statistical Mechanics-Theory and Experiment, 2009, 53(7): 07042.

[14] ZHANG Z W, WANG Z Y. Mining overlapping and hierarchical communities in complex networks[J]. Physica A: Statistical Mechanics and its Applications, 2015, 421: 25-33.

[15] AHN Y Y, BAGROW J P, LEHMANN S. Link communities reveal multiscale complexity in networks[J]. Nature, 2010, 466(7307): 761-764.

[16] MENG F, ZHANG F, ZHU M, et al. Incremental density-based link clustering algorithm for community detection in dynamic networks[J]. Mathematical Problems in Engineering, 2016, 2016(6): 1-11.

[17] KIM P, KIM S. Detecting overlapping and hierarchical communities in complex network using interaction-based edge clustering[J].Physica A: Statistical Mechanics and its Applications, 2015, 417: 46-56.

[18] LANCICHINETTI A, FORTUNATO S, KERTESZ J. Detecting the overlapping and hierarchical community structure in complex networks[J]. New Journal of Physics, 2009, 11(3): 033015.

[19] WANG M, YANG S, WU L. Improved community mining method based on LFM and EAGLE[J]. Computer Science and Information Systems, 2016, 13(2): 515-530.

[20] CAO X, WANG X, JIN D, et al. The (un)supervised detection of overlapping communities as well as hubs and outliers via (bayesian) NMF[C]//International Conference on World Wide Web Companion. 2014: 233-234.

[21] 常振超, 陈鸿昶, 黄瑞阳, 等. 基于非负矩阵分解的半监督动态社团检测[J]. 通信学报, 2016, 37(2): 132-142.
CHANG Z C, CHEN H C, HUANG R Y, et al. Semi-supervised dynamic community detection bused on non-negative matrix factorization[J]. Journal on Communications, 2016, 37(2): 132-142.

[22] HE D, WANG H, JIN D, et al. A model framework for the enhancement of community detection in complex networks[J]. Physica A Statistical Mechanics & Its Applications, 2016, 461: 602-612.

[23] GREGORY S. Finding overlapping communities in networks by label propagation[J].New Journal of Physics, 2010, 12(10): 103018.

[24] XIE J R, SZYMANSKI B K , LIU X .Slpa: uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process[C]//IEEE ICDM Workshop on DMCCI. IEEE, Vancouver, Canada, 2011: 344-349.

[25] CHEN Q, WU T T. A method for local community detection by finding maximal-degree nodes[C]//9th International Conference on Machine Learning and Cybernetics. 2010: 8-13.

[26] ZHANG T, WU B. A method for local community detection by finding core nodes[C]//International Conference on Advances in Social Networks Analysis and Mining, IEEE. Istanbul, Turkey, 2012: 1171-1176.

[27] LEUNG I X, HUI P, LIÒ P, et al. Towards real-time community detection in large networks[J]. Physical Review E, 2009, 79(2): 853-857.

[28] XIE J, KELLEY S, SZYMANSKI B K. Overlapping community detection in networks: the state of the art and comparative study[J]. Acm Computing Surveys, 2011, 45(4): 115-123.

[29] LANCICHINETTI A, FORTUNATO S, RADICCHI F. Benchmark graphs for testing community detection algorithms[J]. Physical Review E, 2008, 78(4): 046110.

[30] ZACHARY W W. An information flow model for conflict and fission in small groups[J]. Journal of Anthropological Research, 1977, 33(4): 452-473.

[31] LUSSEAU D. The emergent properties of a dolphin social network[J].Proceedings of the Royal Society B: Biological Sciences, 2003, 270(S2): 186-188.

[32] KNUTH D E. The Stanford graphbase: a platform for combinatorial computing[EB/OL]. http://www-cs-faculty.stanford.edu/~uno/sgb.html.

[33] NEWMAN M E J. Modularity and community structure in networks[J]. Proceedings of the National Academy of Science, 2006, 103(23): 8577-8582.

[34] GUIMERA R, DANON L, DIAZ-GUILERA A, et al. Self-similar community structure in a network of human interactions[J]. Physical Review E, 2003, 68(6): 065103.

[35] NEWMAN M E J. Network data from mark Newman's home page[EB/OL]. http://www-personal.umich.edu/~mejn/netdata/.

[36] NEWMAN M E J. Finding community structure in networks using the eigenvectors of matrices[J].Physical review E, 2006, 74(3): 036104.

[37] BOGUÑÁ M, PASTOR-SATORRAS R, DÍAZ-GUILERA A, et al. Models of social networks based on social distance attachment[J]. Physical Review E, 2004, 70(5): 056122.

作者简介：

邓琨（1980-），男，黑龙江哈尔滨人，博士，嘉兴学院讲师，主要研究方向为数据挖掘、复杂网络结构分析等。

李文平（1979-），男，贵州大方人，博士，嘉兴学院讲师，主要研究方向为数据挖掘、隐私保护等。

余法红（1977-），男，湖北监利人，嘉兴学院讲师，主要研究方向为复杂网络、智能计算、大数据分析等。

张健沛（1956-），男，黑龙江哈尔滨人，博士，哈尔滨工程大学教授、博士生导师，主要研究方向为数据库理论与应用、数据挖掘、复杂网络等。