层次聚类社区发现算法的研究*

龚尚福，陈婉璐，贾澎涛
（西安科技大学 计算机科学与技术学院，西安 710054）

摘 要： 概述了社区发现算法的研究现状；介绍了因分析对象的不同而产生的四类社区发现方法：矩阵谱分析方法、层次聚类方法、基于边图思想的方法和基于极大团思想的方法。对其中性能最优的层次聚类方法进行了详细的综述，并对其典型算法进行了分析比较。最后，提出了社区发现算法可能的研究方向，为今后的研究提供参考。

关键词： 社区发现；复杂网络；矩阵谱分析；层次聚类；边图思想；极大团方法
中图分类号： TP391 **文献标志码：** A **文章编号：** 1001-3695(2013)11-3216-05
doi:10.3969/j.issn.1001-3695.2013.11.003

Survey on algorithms of community detection

GONG Shang-fu, CHEN Wan-lu, JIA Peng-tao
(*College of Computer Science & Technology, Xi'an University of Science & Technology, Xi'an* 710054, *China*)

Abstract: This paper briefly introduced research situation of community detection. Secondly, it classified community detection methods into broad categories as spectral algorithms, hierarchical clustering, link clustering and clique percolation, then gave a brief introduction about them. Since hierarchical clustering had better performance, it especially presented and compared some typical methods of it. Finally, it pointed out the future research directions that would be hopefully beneficial to the researchers from related fields.

Key words: community detection; complex network; spectral algorithms; hierarchical clustering; link clustering; clique percolation

在复杂网络中，社区是根据某种或某几种性质对网络进行的划分，由节点和边组成，也被称为群组(group)、聚类(cluster)或模块(module)。在研究复杂网络拓扑结构的过程中，人们逐渐发现网络中存在社区。最初，因为无法找到网络内部拓扑结构的规律，研究人员便假设网络是随机分布的，并在此基础上进行研究；但是在后来的研究过程中发现，绝大多数真实复杂网络的拓扑结构既不完全随机也不完全规则，而是具有局部聚类与局部随机特征的、极为复杂的链接关系[1,2]。后来，人们发现了“小世界效应”(small-world effect)[2]、“无标度”(scale-free)[3]特性以及网络中的聚簇特征。这些成果充分证明了复杂网络中存在社区结构，并且社区内部节点间联系相对紧密或性质相似，社区间节点联系相对稀疏或性质相异[4~6]。

根据社区结构的特点，可以通过一些方法挖掘出复杂网络中的社区结构，这个过程称为社区发现；能够实现社区发现的算法称为社区发现算法。通过社区发现算法挖掘复杂网络中潜在的社区结构，对于进一步研究网络的拓扑结构和层次结构起着至关重要的作用。因为复杂网络中包含海量的节点和边，并且结构复杂，一般的图分析方法很难对其进行系统的研究。但是通过社区发现，可以得到网络的层次结构关系，并根据需要在不同层次上进行网络分析，以发现其中蕴藏的知识。另外，社区发现具有很高的应用价值。比如，可以聚类互联网上位置相近的用户，从而为其提供更加个性化的服务[7]；可以构建买家和商品的网络，从中发现具有相似兴趣的顾客，从而建立高效的商品推荐系统[8]；可以对大型网络进行聚类，从而改善数据的存储结构，使其易于查询[9]。目前，社区发现已被广泛应用于涉及网络分析的各个领域，如社会网络[5]、蛋白质作用关系网络[10~12]、Web 网络[13,14]、新陈代谢网络[15,16]等。

因此，在复杂网络中进行社区发现具有非常重要的意义。《Nature》《Science》《PNAS》《Physical Review E》等世界顶级的学术期刊上出现了大量与社区发现相关的论文，越来越多的国际会议与工作组如 SIGKDD、ICDM、WWW、SIGIR、ECML PKDD、CIKM 等都将社区发现作为研究探讨的问题。自此，社区发现成为了研究的焦点。本文对社区发现领域进行了研究，分析了现有的社区发现算法，并提出了目前该领域已解决和亟待解决的问题，希望能为日后的研究工作有所助益。

1 社区发现算法

对于复杂网络中的社区发现，目前已经提出了许多思想和算法。根据它们分析对象的不同，分为以下四类：a) 矩阵谱分析方法，通过对网络的邻接矩阵进行谱分析来实现社区发现；b) 层次聚类方法，从网络中节点出发，采用某种策略衡量节点间的相似性，并以此聚类划分社区；c) 基于边图思想的方法，从网络中的边出发进行社区发现；d) 基于极大团思想的方法，

收稿日期：2013-02-19；修回日期：2013-04-19 基金项目：陕西省自然科学基金资助项目(2012JQ8035)

作者简介：龚尚福(1954-)，男，宁夏平罗人，教授，主要研究方向为网络计算与信息安全(gongsf@xust.edu.cn)；陈婉璐(1989-)，女，硕士研究生，主要研究方向为网络计算与信息安全；贾澎涛(1977-)，女，河南新郑人，副教授，博士，主要研究方向为人工智能及应用、数据挖掘、矿山安全及可视化.

通过分析网络中的完全子图来实现社区发现。

1.1　基于矩阵谱分析的方法

基于矩阵谱分析的方法[17~21]是最早用于社区发现的方法,它通过对 Laplace 矩阵或模块度矩阵进行谱分析来发现网络中的社区。具体来说,因为对任意实对称矩阵而言,它的非退化特征值所对应的特征向量总是正交的,所以,非零特征值对应的特征向量中总是包含正、负两种元素。通过矩阵谱分析方法进行社区发现正是基于这种思想提出来的。将复杂网络映射到 Laplace 矩阵中,计算矩阵的特征值和特征向量,找到次小特征值所对应的特征向量,按照特征向量中元素的正负划分社区。将正元素对应节点划分到一个社区,负元素对应节点划分到另一个社区,这样就可以实现最简单的网络社区划分。考虑到一次划分只能将网络分为规模相近的两个社区,若想发现网络中真实的社区结构,可以迭代进行上述步骤,从而在网络中发现相对真实的社区结构。

1.2　基于层次聚类思想的方法

基于层次聚类思想的方法脱离了社区发现对矩阵分析的依赖,通过计算节点的相似性进行迭代聚类。聚类的最终结果会形成一棵层次聚类树,通过模块度函数对树进行切割,可以找到最优的社区划分。模块度函数[6]是由 Newman 与 Girvan 提出的,定义为

$$Q = \sum_r (e_{rr} - \alpha_r^2)$$

其中: e_{rr} 表示社区 r 内部边数,α_r 表示社区 r 内部边数与外部边数之和。若将社区看成是一个子图,则与该子图对应的随机模型中的边数量应小于该子图中实际存在的边数量,即社区结构越好,对应的模块度函数值越大。

1.3　基于边图思想的方法

基于边图的思想是在研究重叠社区时被引入到社区发现中的。不同于层次聚类方法,基于边图思想的方法通过计算邻接边之间的相似性来实现社区发现,当邻接边属于不同社区时,它们的交点一定是两个社区的重叠部分。

Ahn 等人[22]于 2009 年首次提出基于边图的重叠社区发现方法,计算边对的相似度,对其进行降序迭代合并,直到所有的边都在同一个社区中。黄发良等人[23]受到边图思想的启发,将其与粒子群优化的思想相结合,提出基于边图的粒子群优化方法,将原图转换为边图,对边图进行非重叠划分,之后运用粒子群优化方法找到最优划分。该方法的时间复杂度为 $O(t \times m^2/r)$,其中,t 表示迭代次数,m 表示边数,r 表示一次迭代产生的社区数。

1.4　基于极大团思想的方法

层次聚类方法和基于边图思想的方法进行社区发现时,关注的焦点都是网络中的单个节点或边,而基于极大团思想的方法则是从网络中的完全子图出发,通过分析完全子图划分社区。在极大团思想中,社区被定义为一系列相邻的 k 完全图的集合,k 完全图即为具有 k 个节点的完全图。如果两个 k 完全图之间共享 $k-1$ 个节点,则称这两个 k 完全图是相邻的。

最早的基于极大团思想的方法是由 Palla 等人[16]提出的极大团过滤算法(clique percolation method,CPM) 。CPM 的基本思想是: 社区是由 k 完全图在其多个邻接 k 完全图上滚动而来的。Kumpula 等人[24]对 CPM 进行了优化,提出了时序性极大团过滤算法(sequential CPM,SCP) 。该方法不仅加速了极大团过滤算法,而且成功地将极大团思想应用到了加权网络中。在此基础上,Shen 等人[25,26]将极大团思想与层次聚类方法相结合,提出了 EAGLE 算法。该算法将网络中大于某一阈值的 k 完全图凝聚为谱系树,通过 cover 评价社区质量,选取切割点以得到最佳社区结构。

1.5　几类方法的比较

上述几类方法从不同的角度实现了社区发现,其典型算法比较如表 1 所示。

表 1　社区发现算法比较

方法	时间复杂度	寻优能力	优点	缺点
矩阵谱分析方法	一般: $O((m+n)\times n) \sim O(n^3)$; 对稀疏图: $O(n^2)$	社区结构清晰时,准确度有波动; 社区结构模糊时,准确度较好	准确度较高; 社区结构简单时效果较好	需要先验知识; 不识别重叠社区; 无法自动识别社区数目
层次聚类方法	一般: $O((m+n)\times n) \sim O(m^2 n)$; 对稀疏图: $O(n^2)$; 优化后可达: $O(md\log n)$	不同的聚类方法得到的社区划分结果准确度差别较大,其中准确度最高的是 GN	准确度较高; 能揭示网络的层次结构; 优化后可发现大规模网络中的重叠社区	对于不好的划分,不能回溯地进行修正
基于边图思想的方法	一般: $O((m+n)\times n)$; 对稀疏图: $O(n^2)$	能较为正确地发现重叠社区; 对稀疏图和稠密图同样有效	能揭示网络的层次结构; 能发现重叠社区	应用较少; 不具备迭代停止的条件
基于极大团思想的方法	依赖于网络的拓扑结构(参考: 对含 12700 条边的网络,耗时 7200 s; GN: 6 374 s; 快速 GN: 0.3 s; 谱方法: 5.6 s)	准确度很高,特别是能准确地发现重叠社区	能预测网络的动态变化; 能找到网络中的重要社区	不能揭示网络层次结构; 寻找极大团复杂度较高; 发现过多的重叠

矩阵谱分析方法在网络结构简单时可以取得较好的结果,但是其不可避免地会受到一般谱方法的限制。比如,需要将社区数目作为算法迭代的停止条件,而这在实际的网络分析中是无法事先获得的。另一方面,将网络映射到 Laplace 矩阵需要很高的时间复杂度和空间复杂度,所以基于矩阵谱分析的方法难以在复杂网络分析中得到广泛的应用。

基于边图思想的方法为社区发现方法提供了一种新的思路,它巧妙地解决了网络中社区的重叠问题。但是基于边图思想的方法还没有被广泛接受,因此还未得到较为深入的研究。

基于极大团思想的方法可以较好地发现网络中的重叠社区结构。但是运用极大团思想的方法一方面需要很高的时间复杂度来发现网络中的极大团,另一方面会发现社区结构中存在大面积重叠现象,而这很可能并不符合网络中的实际情况。

与以上三种方法相比,层次聚类方法虽然也具有一些局限性,但是因为层次聚类方法产生的聚类树可以非常直观地反映网络的层次结构,对于进一步研究网络的拓扑结构起着至关重要的作用,所以层次聚类方法得到了普遍的关注。

2 层次聚类方法

2.1 分裂式方法

分裂式方法[5,13,27~29]的基本原理是首先将整个网络看成一个社区，然后以某种策略计算节点对的相似性，将相似性较低的节点对划分到不同的社区中。通过这样的反复迭代操作，将网络划分为若干个子图，即为社区。

分裂式方法的典型代表是 GN 算法[5]。GN 算法通过迭代删除边介数最大的边来实现社区发现，但是精确计算边介数的代价非常高，如果网络中的边数为 m，节点数为 n，则边介数计算的时间复杂度为 $O(mn)$，整体算法的时间复杂度高达 $O(m^2n)$。GN 算法过程如下所示：

输入：网络 $G(V,E)$；聚类方法。
输出：$G(V,E)$ 的层次聚类树图。
聚类过程：
将 $G(V,E)$ 初始化为一个社区；
计算网络中所有边的边介数：
$B(e) = \sum_{i\in V, j\in V, i\neq j} \text{shortestPath}_{i,j}(e)$；
/* $\text{shortestPath}_{i,j}(e)$ 表示点 i、j 之间的最短路径是否经过 e，经过 e 则为 1，否则为 0 */
迭代聚类划分社区；
while 网络中还存在边
　删除具有最大边介数的边 e；
　重新计算与 e 相关联的所有的边介数；
end while

GN 算法是社区发现算法的开创性算法，它为社区发现算法的发展奠定了良好的基础，但是其复杂度限制了它的应用和普及。通过对 GN 算法进行改进，可以发现社区间的重叠现象。Gregory[30] 在 GN 算法的基础上，提出 CONGA 算法，沿用 GN 算法的边介数思想，并引入节点对的 split betweenness。将节点对 (v_1,v_2) 的 split betweenness 定义为通过 (v_1,v_2)、经过其邻居节点的最短路径数目。比较边介数和节点对的 split betweenness，若边介数大，则删除边；若节点对的 split betweenness 大，则将该点删除并分别复制到两个子图中。Pinney 等人[31]提出，在比较边介数和节点对的 split betweenness 时，不单纯地比较大小，而是计算两者的比率。若两者的比率在 $[\alpha,1/\alpha]$ 之间，则删除边；否则对节点进行带复制的删除操作。Gregory 后来将这种思想进一步发展，提出了 PEACOCK 方法[32]。PEACOCK 的基本思想是对具有最大 split betweenness 的点 v 进行复制，即可得到 (v,v_1)，其中 v_1 是 v 的一个复制；之后在 v 与 v_1 中间添加一条边，实现图的扩充。重复进行上述操作，直到图中不存在具有较大 split betweenness 的点。将原图用 PEACOCK 方法进行转换，得到更大的图。对该图进行一般的社区划分，再将得到的社区结构映射到原图上，实现重叠社区的发现。

分裂式方法在应用的过程中因为侧重于发现社区间的边界，往往具有很高的时间复杂度，而且对于稀疏网络，可能会发现大量的孤立节点不属于任何社区，因此并不能取得很好的结果。但是通过调整原图，将现有算法应用于重叠社区发现的思想十分值得借鉴。

2.2 凝聚式方法

凝聚式方法与分裂式方法类似，但是方向相反。凝聚式方法的基本思想是首先将网络中的每个节点都视为一个独立社区，迭代计算社区之间的相似性，将相似度高的社区进行合并。相似性策略根据是否对网络中的全部节点进行判断，可分为：

a) 基于全局相似性的方法[33~36]。它是指在计算节点之间的相似性时，要对网络中全部节点或边进行判断，找出全局最优的合并策略。

通过计算全局相似性进行凝聚式聚类的典型算法是 Newman[33] 提出的快速 GN 算法。快速 GN 算法的思想是迭代合并使模块度增量最大的两个社区。每次迭代的时间复杂度为 $O(mn)$，其中 m 为边数，n 为节点数，整体算法的时间复杂度为 $O((m+n)n)$，在稀疏网络中为 $O(n^2)$。Clauset 等人[34]提出 CNM 算法，将快速 GN 算法的数据结构改为堆结构，使得时间复杂度减少到 $O(md\log n)$，其中 d 为社区聚类树的深度。快速 GN 算法过程如下所示：

输入：网络 $G(V,E)$；聚类方法。
输出：$G(V,E)$ 的层次聚类树图。
聚类过程：
将 $G(V,E)$ 中每一个节点初始化为一个社区；
迭代聚类划分社区
while G 中社区数目大于 1
　计算社区 i、j 之间的边 e_{ij}；
　计算 $a_i = \sum_j e_{ij}$；
　计算任意两个社区间的模块度增量 $\Delta Q = e_{ij} + e_{ji} - 2a_ia_j = 2(e_{ij} - a_ia_j)$；
　合并使模块度增量最大的两个社区；
　将合并结果保存到层次聚类树中；
end while

b) 基于局部相似性的方法[37~41]。它是指在计算节点的相似性时，只考虑该点的邻居节点，以发现局部最优的合并策略。

一种典型的算法是标记传播算法(label propagation algorithm，LPA)[38,39]。标记传播算法在初始时刻以唯一的标签标志每个节点，在迭代的过程中，每一个节点都以邻居节点中最流行的标签替换自己原有的标签，最后使连接紧密的节点具有统一的标签从而形成社区。标记传播算法过程如下所示：

输入：网络 $G(V,E)$；聚类方法。
输出：$G(V,E)$ 的社区结构。
聚类过程：
初始化网络中所有节点的标签 $C_x(0) = x$；
/* $C_x(t) = a$ 表示节点 x 在第 t 次迭代后的标签为 a */
初始化 $t = 1$；
迭代聚类划分社区
while 存在需要更新标签的节点
　将节点随机排序放到 X 中；
　对所有 X 中的节点，更新其标签值 $C_x(t) = f(C_{x_1}(t),\cdots,C_{x_m}(t),C_{x_{m+1}}(t-1),\cdots,C_{x_n}(t-1))$；
　/* $C_{x_i}(t)$ 表示节点 x 的邻接点 i 在第 t 次迭代时的标签值，f 函数用来返回在邻居节点中出现频率最高的标签值 */
　$t = t + 1$；
end while

另一种是基于适应度的算法[40,41]。因为社区内连接数大于社区间连接数，所以基于适应度的算法以此作为社区发现的切入点，任选网络中的一点作为起始点，对其邻居节点进行迭代聚类。Lancichinetti 等人[41]在此基础上将适应度函数调整为 $f_g = k_{\text{in}}^g/(k_{\text{in}}^g + k_{\text{out}}^g)^\alpha$ 以发现网络中的重叠社区。其中 k_{in}^g、k_{out}^g 分别表示模块 g 中所有节点的入度和和出度和，变量 α 用来控制社区的大小。算法过程如下所示：

输入：网络 $G(V,E)$；阈值 α（α 越大，得到的每个社区的规模越小）；聚类方法。
输出：$G(V,E)$ 的社区结构。
聚类过程：
while 存在不属于任何社区的节点
　任选不属于任何社区的节点 A；
　将 A 作为社区 g 的起点，$k_{\text{in}}^g = 0$；
　迭代划分社区

```
    do
        获得 g 的邻居节点集合 B;
        对 ∀v∈B 计算 f_g^v = f_{g+{v}} − f_{g−{v}};
        /* f_{g+{v}} 和 f_{g−{v}} 分别为 v 加入 g 和不加入 g 的适应度* /
        获得 max{f_g^B},将其加入 g,形成 g';
        对 ∀v'∈g',删除 g'中所有使 f_{g'}^{v'}<0 的点;
    while 对 ∀v∈B 有 f_g^v<0
end while
```

c) 基于中心节点的聚类方法。它结合了以上两种方法的思想，先遍历网络中的节点，以某种策略确定中心节点，将相似性较小的中心节点划分到不同的社区中，再对其邻居节点进行判断，确定社区划分。通过将网络中度最大的点作为中心节点，陈端兵等人[43]提出根据中心节点及其邻居节点确定初始社区，之后对各个小社区进行合并。基于中心节点聚类算法过程如下所示：

```
输入: 网络 G(V,E);阈值 C_L(C_L=0.45)。
输出: G(V,E) 的社区结构。
聚类过程:
初始化 p^in,用来存储抽取的小社区;
迭代聚类划分社区
while G 中存在还未被访问过的节点
    由度最大的节点其邻接点形成社区 c;
    依次计算 c 中节点 v 对社区 c 的连接度 C(v,c) = (Σ_{u∈c} w_uv)/k_v;
    /* k_v 为 v 的度数,u 与 v 之间有连接时,w_uv=1,否则 w_uv=0* /
    将连接度小于 C_L 的节点删除;
    获得 c 的邻接点;
    对每个邻接点 v',计算 C(v',c);
    将连接度大于等于 C_L 的节点加入 c 中;
    标记 c 中的节点,保存 c 到 p^in;
end while
```

计算 p^{in} 中任意一对社区间连接度 $C(c1,c2) = \sum_{u\in c1,v\in c2} w_{uv} / \min\left(\sum_{u,v\in c1} w_{uv}, \sum_{u,v\in c2} w_{uv}\right)$；

```
    if C(c1,c2) > C_L && 模块度增量 >0
        合并 c1、c2;
    end if
```

此外，骆挺等人[43]根据邻居节点是否与中心节点形成完全子图来决定邻居节点的归属社区；而万雪飞、马兴福等人[44,45]则是根据适应度函数来判断其邻居节点的归属社区。王莉军、淦文燕等人[46,47]将社区发现方法与物理学中电势场的思想相结合，引入数据场的概念，将社区建模为电势场，将势函数值最大的点定义为中心节点，以此划分社区。

基于全局相似性的方法和基于中心节点的方法因为需要计算全部节点的相似性，所以时间复杂度较高。而且如果网络随时间发生动态变化，这两种方法无法得到比较客观的社区划分结果。基于局部相似性的凝聚式算法因为具有很强的局部特性，所以很难发现网络中的全局最优解，但在大规模复杂网络中节点数目比较多，计算全局相似性十分困难，此时，基于局部相似性的凝聚式算法不失为一种比较好的策略。

2.3　层次聚类方法比较

在层次聚类方法中，凝聚式算法的时间复杂度普遍低于分裂式算法，但是因为专注于一些相似性较高的节点，导致在迭代的最后一些外围的相似度较低的节点得不到很好的划分。但总体上来看，凝聚式算法比分裂式方法要更加符合社区的形成过程，也因此具有更多的应用前景。层次聚类方法的典型算法比较如表 2 所示。

表 2　层次聚类方法的典型算法比较

算法	时间复杂度	寻优能力	优点	缺点
GN	$O(m^2n)$	准确度很高，但是不存在社区时仍识别出社区结构	适宜小规模简单网络；关注社区边界，为社区结构的分析提供一种新的角度	计算边介数复杂度高；在极端情况下易对社区结构产生过高估计；无法应用于复杂网络
快速 GN	一般：$O((m+n)\times n)$；对稀疏图：$O(n^2)$；优化后：$O(md\log n)$	准确度随社区结构模糊程度的增加而平稳下降，大部分时候低于 GN	适用于大型加权网络；获得聚类树图，便于分析大规模网络的社区结构	无法适应网络动态变化；易受到模块度自身局限性的影响
LPA	每次迭代 $O(m)$，大部分网络五次迭代即可	准确度随网络社区结构模糊程度的增加而迅速下降	时间复杂度符合大规模网络分析的要求；完全利用拓扑结构，无须人为设置变量	算法准确度低；无法用于发现网络中的重叠社区
基于度的局部方法	$O(n^2\log n)$	准确度较好	能同时发现网络中的层次结构和重叠结构；能较好地适应网络的动态变化	无法得到最优划分；随机点的选取可能会导致社区的重复发现
基于中心节点聚类	$O(n^2)$	准确度较好	发现的社区质量高；在大规模网络中同样有效	只能用于无向无权网络；包含的网络层次结构信息较少；易发现过多的小社区

3　结束语

本文在阅读整理了大量文献后，对当前的社区发现算法进行了梳理，对现有的社区发现算法进行了分析比较。通过对现有研究成果的分析可以发现，尽管目前的研究已经取得了卓有成效的进展，但是关于社区发现算法的研究仍然存在一些问题，这些存在的问题为后续研究指明了方向。

a) 模块度的优化问题亟待解决。研究发现[48,49]，模块度 Q 并不能如实地反映出真实的网络社区结构。比如，某些明显不好的社区结构却能得到相对较高的模块度值，得到的结果较为粗糙等。因此，继续优化模块度或找到一个可以替代模块度的社区结构衡量标准，将是层次聚类方法未来的一个研究方向。

b) 关于重叠社区的研究。对社区重叠现象的研究仍然有一些问题还未涉及到。比如，在所有的社区发现算法中，都未考虑到模糊重叠[50]的情况，即一个节点归属于不同社区的程度有所差异。这种想法更符合真实世界的情况，但是如何在这种情况下进行社区发现还有待研究。

c) 关于链接社团的研究。与传统方法将社区看成是节点的集合不同，链接社团[51]将社区看成是由若干条边组成的集合，通过对边进行聚类，得到边的聚类树图。目前使用的用来衡量社区划分质量的函数 D 并不能取得很好的结果[52]，因此如何判断社区结构的优劣是链接社团后续研究中亟待解决的问题。

d) 尽管社区发现的算法层出不穷，但是却鲜有探讨如何对网络中的社区结构进行分析的。比如，网络中的社区结构和

社区的重叠现象是如何形成的，以及不同层次的社区结构对网络分析的意义何在。虽然大量的社区发现算法被提出来，但是总体而言还是相当基础和简单的，还未达到通过社区发现获取知识的程度。

参考文献:

[1] ROSVALL M. Information horizons in a complex world [D]. Umea: Umea University, 2006.

[2] WATTS D J, STROGATZ S H. Collective dynamics of 'small world' networks [J]. Nature, 1998, 393(6684): 440-442.

[3] BARABASI A L, BONABEAU E. Scale-free networks [J]. Scientific American, 2003, 288(5): 60-69.

[4] FORTUNATO S. Community detection in graphs [J]. Physics Reports, 2010, 486: 75-174.

[5] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks [J]. Proceedings of the National Academy of Sciences, 2002, 99(12): 7821-7826.

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

[7] KRISHNAMURTHY B, WANG Jia. On network-aware clustering of Web clients [C]//Proc of Conference on Applications, Technologies, Architectures and Protocols for Computer Communication. New York: ACM Press, 2000: 97-110.

[8] REDDY P K, KITSUREGAWA M, SREEKANTH P, *et al*. A graph based approach to extract a neighborhood customer community for collaborative filtering [C]//Proc of the 2nd International Workshop on Databases in Networked Information Systems. London: Springer-Verlag, 2002: 188-200.

[9] WU A Y, GARLAND M, HAN Jia-wei. Mining scale-free networks using geodesic clustering [C]//Proc of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2004: 719-724.

[10] RIVES A W, GALITSKI T. Modular organization of cellular networks [J]. Proceedings of the National Academy of Sciences, 2003, 100(3): 1128-1133.

[11] SPIRIN V, MIRNY L A. Protein complexes and functional modules in molecular networks [J]. Proceedings of the National Academy of Sciences, 2003, 100(21): 12123-12128.

[12] CHEN Jing-chun, YUAN Bo. Detecting functional modules in the yeast protein-protein interaction network [J]. Bioinformatics, 2006, 22(18): 2283-2290.

[13] FLAKE G W, LAWRENCE S, GILES C L, *et al*. Self-organization and identification of Web communities [J]. Computer, 2002, 35(3): 66-70.

[14] DOURISBOURE Y, GERACI F, PELLEGRINI M. Extraction and classification of dense communities in the Web [C]//Proc of the 16th International Conference on World Wide Web. New York: ACM Press, 2007: 461-470.

[15] GUIMERA R, AMARAL L A N. Functional cartography of complex metabolic networks [J]. Nature, 2005, 433(2): 895-900.

[16] PALLA G, DERENYI I, FARKAS I, *et al*. Uncovering the overlapping community structure of complex networks in nature and society [J]. Nature, 2005, 435(6): 814-818.

[17] FIEDLER M. Algebraic connectivity of graphs [J]. Czechoslovak Mathematical Journal, 1973, 23(2): 298-305.

[18] DONETTI L, MUNOZ M A. Detecting network communities: a new systematic and efficient algorithm [J]. Journal of Statistical Mechanics: Theory and Experiment, 2004, 2004(10): 10012.

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

[20] ZHANG Shi-hua, WANG Rui-sheng, ZHANG Xiang-sun. 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.

[21] 黄发良，肖南峰. 用于网络重叠社区发现的粗糙谱聚类算法 [J]. 小型微型计算机系统，2012，33(2)：263-266.

[22] AHN Y Y, BAGROW J P, LEHMANN S. Communities and hierarchical organization of links in complex networks [EB/OL]. (2010-09-04) [2012-12-03]. http://arxiv.org/abs/0903.3178.

[23] 黄发良，肖南峰. 基于线图与 PSO 的网络重叠社区发现 [J]. 自动化学报，2011，37(9)：1140-1144.

[24] KUMPULA J M, KIVELA M, KASKI K, *et al*. Sequential algorithm for fast clique percolation [J]. Physical Review E, 2008, 78(2): 026109.

[25] SHEN Hua-wei, CHENG Xue-qi, CAI Kai, *et al*. Detect overlapping and hierarchical community structure in networks [J]. Physica A: Statistical Mechanics and Its Applications, 2009, 388(8): 1706-1712.

[26] SHEN Hua-wei, CHENG Xue-qi, GUO Jia-feng. Quantifying and identifying the overlapping community structure in networks [J]. Journal of Statistical Mechanics: Theory and Experiment, 2009, 2009(7): 07042.

[27] TYLER J R, WILKINSON D M, HUBERMAN B A. E-mail as spectroscopy: automated discovery of community structure within organizations [C]//Proc of the 1st International Conference on Communities and Technologies. Netherlands: Kluwer Academic Publishers, 2003: 81-96.

[28] RADICCHI F, CASTELLANO C, CECCONI F, *et al*. Defining and identifying communities in networks [J]. Proceedings of the National Academy of Sciences, 2004, 101(9): 2658-2663.

[29] 陈东明，徐晓伟. 一种基于广度优先搜索的社区发现算法 [J]. 东北大学学报，2010，31(3)：346-349.

[30] GREGORY S. An algorithm to find overlapping community structure in networks [C]//Lecture Notes in Computer Science, vol 4702. Berlin: Springer, 2007: 91-102.

[31] PINNEY J W, WESTHEAD D R. Betweenness-based decomposition methods for social and biological networks [C]//Interdisciplinary Statistics and Bioinformatics. Leeds: Leeds University Press, 2007: 87-90.

[32] GREGORY S. Finding overlapping communities using disjoint community detection algorithms [C]//Proc of International Workshop on Complex Networks. Berlin: Springer, 2009: 47-61.

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

[34] CLAUSET A, NEWMAN M E J, MOORE C. Finding community structure in very large networks [J]. Physical Review E, 2004, 70(6): 066111.

[35] WU Fang, HUBERMAN B A. Finding communities in linear time: a physics approach [J]. The European Physical Journal B: Condensed Matter and Complex Systems, 2004, 38(2): 331-338.

[36] 王刚，钟国祥. 基于信息熵的社区发现算法研究 [J]. 计算机科学，2011，38(2)：238-240.

[37] 刘旭，易东云. 基于局部相似性的复杂网络社区发现方法 [J]. 自动化学报，2011，37(12)：1520-1529.

[38] 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.

[39] LEUNG I X Y, HUI Pan, LIO P, *et al*. Towards real-time community detection in large networks [J]. Physical Review E, 2009, 79(6): 066107.

[40] BAGROW J P, BOLLT E M. Local method for detecting communities [J]. Physical Review E, 2005, 72(4): 046108.

(下转第 3227 页)

[47] LI Nan, CHEN Guan-ling. Analysis of a location-based social network [C] //Proc of International Conference on Computational Science and Engineering. Washington DC: IEEE Computer Society, 2009: 263-270.

[48] SCELLATO S, NOULAS A, LAMBIOTTE R, *et al*. Socio-spatial properties of online location-based social networks [C] //Proc of the 5th International AAAI Conference. 2011.

[49] YING J J, LEE W, YE Mao, *et al*. User association analysis of locales on location based social networks [C] //Proc of the 3rd ACM SIGSPATIAL International Workshop on Location Based Social Networks. New York: ACM Press, 2011: 69-76.

[50] LEE R, SUMIYA K. Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection [C] //Proc of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks. New York: ACM Press, 2010: 1-10.

[51] De LONGUEVILLE B, SMITH R S, LURASCHI G. "OMG, from here, I can see the flames!": a use case of mining location based social networks to acquire spatio-temporal data on forest fires [C] //Proc of ACM SIGSPATIAL International Workshop on Location Based Social Networks. New York: ACM Press, 2009: 73-80.

[52] GUY M, EARLE P, OSTRUM C, *et al*. Integration and dissemination of citizen reported and seismically derived earthquake information via social network technologies [C] //Advances in Intelligent Data Analysis. Berlin: Springer-Verlag, 2010: 42-53.

[53] WAJAMIYA S, LEE R, SUMIYA K. Crowd-based urban characterization: extracting crowd behavioral patterns in urban areas from Twitter [C] //Proc of the 3rd ACM SIGSPATIAL International Workshop on Location Based Social Networks. New York: ACM Press, 2011: 77-84.

[54] FERRARI L, ROSI A, MAMEI M, *et al*. Extracting urban patterns from location-based social networks [C] //Proc of the 3rd ACM SIGSPATIAL International Workshop on Location Based Social Networks. New York: ACM Press, 2011: 9-16.

[55] SCHLIEDER C, YANENKO O. Spatio-temporal proximity and social distance: a confirmation framework for social reporting [C] //Proc of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks. New York: ACM Press, 2010: 60-67.

[56] POZDNOUKHOV A, KAISER C. Space-time dynamics of topics in streaming text [C] //Proc of the 3rd ACM SIGSPATIAL International Workshop on Location Based Social Networks. New York: ACM Press, 2011: 1-8.

[57] VICENTE C R, FRENI D, BETTINI C, *et al*. Location-related privacy in geo-social networks [J]. IEEE Internet Computing, 2011, 15(3): 20-27.

[58] LI Nan, CHEN Guan-ling. Sharing location in online social networks [J]. Network, 2010, 24(5): 20-25.

[59] KOFOD-PETERSEN A, GRANSAETHER P A, KROGSTIE J. An empirical investigation of attitude towards location-aware social network service [J]. International Journal of Mobile Communications, 2010, 8(1): 53-70.

[60] TSAI J Y, KELLEY P G, CRANOR L F, *et al*. Location-sharing technologies: privacy risks and controls [J]. Journal of Law and Policy for the Information Society, 2010, 6(2): 119-151.

[61] TOCH E, SADEH N M, HONG J. Generating default privacy policies for online social networks [C] //Extended Abstracts on Human Factors in Computing System. New York: ACM Press, 2010: 4243-4248.

[62] BRUSH A J B, KRUMM J, SCOTT J. Exploring end user preferences for location obfuscation, location-based services, and the value of location [C] //Proc of the 12th International Conference on Ubiquitous and Computing. New York: ACM Press, 2010: 95-104.

[63] MANO M, ISHIKAWA Y. Anonymizing user location and profile information for privacy-aware mobile services [C] //Proc of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks. New York: ACM Press, 2010: 68-75.

[64] PUTTASWAMY K P N, ZHAO B Y. Preserving privacy in location-based mobile social applications [C] //Proc of the 11th Workshop on Mobile Computing Systems & Applications. New York: ACM Press, 2010: 1-6.

[65] MASCETTI S, FRENI D, BETTINI C, *et al*. Privacy in geo-social networks: proximity notification with untrusted service providers and curious buddies [J]. The VLDB Journal, 2011, 20(4): 541-566.

[66] KOSTAKOS V, VENKATANATHAN J, REYNOLDS B, *et al*. Who's your best friend: targeted privacy attacks in location-sharing social networks [C] //Proc of the 13th International Conference on Ubiquitous and Computing. New York: ACM Press, 2011: 177-186.

[67] FRENI D, RUIZ VICENTE C, MASCETTI S, *et al*. Preserving location and absence privacy in geo-social networks [C] //Proc of the 19th International Conference on Information and Knowledge Management. New York: ACM Press, 2010: 309-318.

[68] FUSCO S J, MICHAEL K, MICHAEL M G, *et al*. Exploring the social implications of location based social networking: an inquiry into the perceived positive and negative impacts of using LBSN between friends [C] //Proc of the 9th International Conference on Mobile Business and Global Mobility Roundtable. Washington DC: IEEE Computer Society, 2010: 230-237.

[69] DOYTSHER Y, GALON B, KANZA Y. Querying geo-social data by bridging spatial networks and social networks [C] //Proc of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks ACM SIGSPATIAL. New York: ACM Press, 2010: 39-46.

[70] DOYTSHER Y, GALON B, KANZA Y. Storing routes in socio-spatial networks and supporting social-based route recommendation [C] //Proc of the 3rd ACM SIGSPATIAL International Workshop on Location Based Social Networks. New York: ACM Press, 2011: 49-56.

[71] CHOW C, BAO Jie, MOKBEL M F. Towards location-based social networking services [C] //Proc of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks. New York: ACM Press, 2010: 31-38.

[72] 刘经南,郭迟,彭瑞卿. 移动互联网时代的位置服务 [J]. 中国计算机学会通讯, 2011, 7(12): 40-49.

(上接第 3220 页)

[41] 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.

[42] 陈端兵,尚明生,李霞. 重叠社区发现的两段策略 [J]. 计算机科学, 2013, 40(1): 225-228.

[43] 骆挺,钟才明,陈辉. 基于完全子图的社区发现算法 [J]. 计算机工程, 2011, 37(18): 41-43.

[44] 万雪飞,陈端兵,傅彦. 一种重叠社区发现的启发式算法 [J]. 计算机工程与应用, 2010, 46(3): 36-41.

[45] 马兴福,王红. 一种新的重叠社区发现算法 [J]. 计算机应用研究, 2012, 29(3): 844-846.

[46] 王莉军,杨炳儒,谢永红. 一种基于数据场的社区发现算法 [J]. 计算机应用研究, 2011, 28(11): 4142-4145.

[47] 淦文燕,赫南,李德毅,等. 一种基于拓扑势的网络社区发现算法 [J]. 软件学报, 2009, 20(8): 2241-2254.

[48] GUIMERA R, SALES-PARDO M, AMARAL L A N. Modularity from fluctuations in random graphs and complex networks [J]. Physical Review E, 2004, 70(2): 025101.

[49] FORTUNATO S, BARTHELEMY M. Resolution limit in community detection [J]. Proceedings of the National Academy of Sciences, 2007, 104(1): 36-41.

[50] GREGORY S. Fuzzy overlapping communities in networks [J]. Journal of Statistical Mechanics: Theory and Experiment, 2011, 2011(2): 02017.

[51] AHN Y Y, BAGROW J P, LEHMANN S. Link communities reveal multi-scale complexity in networks [J]. Nature, 2010, 466(8): 761-764.

[52] LESKOVEC J, LANG K J, MAHONEY M. Empirical comparison of algorithms for network community detection [C] //Proc of the 19th International Conference on World Wide Web. New York: ACM Press, 2010: 631-640.