一种基于节点特征向量的复杂网络社团发现算法

陆亿红[1] 张振宁[1] 杨 雄[1,2]
(浙江工业大学计算机学院 杭州 310023)[1]
(常州工学院计算机信息工程学院通信工程系 常州 213002)[2]

摘 要 社团结构是复杂网络的一种很普遍且非常重要的拓扑特征,社团的发现有助于了解复杂网络的结构和功能。节点间相似度的评价指标对于社团发现的结果起着至关重要的作用,传统算法中使用的相似度指标存在着时间复杂度过高和不够精确的缺陷。为了弥补这两个缺陷,在信息传递理论的基础上将网络中的节点抽象成了多维数据集,结合传统聚类算法 K-means 提出了一种社团发现的新算法。基于 Zachary Karate Club 网络、Jazz Musician 网络和 Facebook 网络的实验结果表明,该算法是高效且准确的。

关键词 复杂网络,社团结构,信息传递理论,特征向量
中图法分类号 TP393 **文献标识码** A

Community Structure Detection Algorithm Based on Nodes' Eigenvectors

LU Yi-hong[1] ZHANG Zhen-ning[1] YANG Xiong[1,2]
(College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310023, China)[1]
(Department of Communication Engineering, School of Computer & Information Engineering, Changzhou Institute of Technology, Changzhou 213002, China)[2]

Abstract Community structure is one of the ubiquitous and significant topology characteristics of complex network. It can help us to learn the structure and functions of a complex network. The similarity index plays a vital role in community detection but it has the shortage of high time complexity and low accuracy. In order to improve the two shortages, nodes are abstracted from a complex network into a multi-dimension data set based on the theory of information transmission in the network. Combined with the traditional clustering algorithm K-means, a new community detection algorithm was proposed. The experimental results obtained from Zachary Karate Club network, Jazz Musician network and Facebook network show that the algorithm is effective and accurate.

Keywords Complex networks, Community structure, Theory of information transmission, Eigenvector

1 引言

最近几年,大型复杂系统在生物、物理、计算机科学、社会学、数学等诸多领域得到了越来越广泛的重视[1]。这些复杂系统大多可以用复杂网络或者图来表示,例如万维网、蛋白质网络、社交网络、食物网以及流行病传播网等[2]。在复杂网络中,节点表示系统中最基础的组成单元,边则表示它们之间的关系。社团结构是复杂网络中一种很普遍且非常重要的拓扑特征[3],社团的发现可以帮助我们了解复杂网络的功能、预测链路、研究拓扑结构等。目前尚没有一种公认的社团结构发现算法,但大家普遍认同的观点是同一社团内的节点联系紧密而不同社团间的联系相对较为稀疏[4]。

分级聚类是检测复杂网络中社团结构的传统方法,它基于各个节点之间的相似性或者联系强度把网络划分为若干个子网络即社团结构。根据向网络中添加边还是从网络中移除边,又可以分为凝聚方法和分裂方法两类。Girvan 和 Newman 就曾提出了一种非常经典的分裂类社团发现算法——GN 算法[5]。该算法基于边的介数去检测复杂网络中的社团结构,通过不断移除网络中介数高的边来构建分裂树,但是 GN 算法形成的分裂树缺乏分裂终止的条件。为了解决这个问题,Girvan 和 Newman 于 2004 年提出了模块度指标最为优化目标,用以判断是否进行进一步分列并评价社团检测的结果的质量。此后,研究者提出了 FN 算法、随机游走算法等众多以模块度来衡量社团划分质量的算法,模块度也成为了一种最常用的社团划分质量评价标准[6]。

对于这些分级聚类的算法,关键是如何准确地描述节点之间的相似度。根据描述的不同,大致可以分为局部相似度指标和全局性相似度指标两类。局部相似度指标主要有共同邻居(CN)相似度指标[7]、索尔顿(Salton)相似度指标、杰卡德(jaccard)相似度指标这类基于节点之间共同邻居特征的指标。局部相似度指标由于只考虑节点周围的一些拓扑情况而忽视了节点在整个复杂网络内部的作用,因此低估了节点对

本文受水利部公益性行业科研专项(201401044),常州市科技计划项目(CJ20159013)资助。
陆亿红(1968—),女,硕士,副教授,主要研究方向为软件理论、数据挖掘等,E-mail:lyh@zjut.edu.cn(通信作者);张振宁(1991—),男,硕士生,主要研究方向为复杂网络、数据挖掘;杨 雄(1980—),男,博士生,讲师,主要研究方向为复杂网络、数据挖掘。

于全局的作用。全局性相似度指标指的是基于全局信息的节点相似度指标，这类指标的计算复杂度比局部指标的高，但对节点之间相似性的描述也更为准确，例如局部路径(LP)相似度指标、卡茨(Katz)相似度指标等[8]。

现有的社团发现算法主要有两个缺陷：1)无法准确地刻画出社团的边界；2)算法具有很高的复杂度。为了在一定程度上弥补这些缺陷，结合信息传递理论[9-10]提出了一种社团发现新算法。该算法根据信息传递理论抽象出复杂网络中的节点的特征向量，再结合传统聚类算法 K-means 来发现复杂网络中的社团结构，利用模块度目标函数指标来调节 K-means 算法的参数 k，从而寻找到符合模块度函数的最佳社团结构。与一些传统的社团发现算法相比，该算法在将算法复杂度控制在 $O(n^2)$ 的同时尽可能准确地描述节点在网络中的地位，并在不降低社团结构合理性的情况下平衡了社团的规模。基于现实网络的实验结果表明，本文所提出的算法是高效且可行的。

2 相关算法和理论

现实中的绝大多数系统都可以被描述为一个无向无权的网络图 $G=(V,E)$，其中 V 为图中所有节点组成的集合，$|V|$ 为图中节点的数目；E 为图中所有边组成的集合，$|E|$ 表示图中边的数目。在假设网络中所有节点同时沿着它们的边向外传递信息且 v_i 与 v_j 均属于集合 V 的情况下，根据通讯的相关理论，节点 v_i 与 v_j 之间的信息传递效能 ε_{ij} 与节点 v_i 与 v_j 之间最短距离 l_{ij} 成反比，即为 $1/l_{ij}$[9]。根据这一理论，可以计算出节点到网络中所有节点的信息传递能效值，从而抽象出节点的特征向量用于描述该节点在复杂网络中的相对位置。同时，社团结构最重要的特征就是社团内的节点联系紧密，而社团之间的联系相对疏远。根据这一特征，节点 v_i 与 v_j 在复杂网络中的相对位置越接近，则 v_i 与 v_j 同属一个社团的概率就越高；反之，则属于不同社团的概率越高。

2.1 算法原理

矩阵 A 为节点之间的距离矩阵，矩阵 A 的元素描述如下：

$$a_{ij}=\begin{cases} l_{ij}, & i\neq j \\ 0, & i=j \end{cases} \tag{1}$$

其中，v_i 与 v_j 均属于集合 V，l_{ij} 表示 v_i 与 v_j 之间的最短距离。

矩阵 B 为节点之间信息的传递效能矩阵，节点之间的传输效率为最短路径的倒数。又由于传递效率值不大于 1，因此认为节点自身的信息传递效率为 1。显而易见，若节点之间的距离越远，则节点之间信息的传递效能越低；当节点之间不存在路径时，节点之间信息的传递效能则为 0。

$$b_{ij}=\begin{cases} \dfrac{1}{a_{ij}}, & i\neq j \\ 1, & i=j \end{cases} \tag{2}$$

矩阵 B 可以看作由 n 个 n 维向量组成，$n=|V|$，即 $B=(B_1,B_2,\cdots,B_n)$，$n=|V|$。其中任意向量 $B_i=(b_{i1},b_{i2},\cdots,b_{in})(i=1,2,\cdots,n)$ 反映的是节点 v_i 到该网络中其他任意节点的传输效率值，同时也可描述节点 v_i 在该网络中的相对位置。将 B_i 和 B_j 视作节点 v_i 和 v_j 在该网络中的特征向量，若特征向量相似性越大，那么节点 v_i 和 v_j 在该网络中的相对位置越接近，即节点 v_i 与 v_j 属于同一个社团的概率也就越高。显而易见，在节点 v_i 与 v_j 拥有的邻居节点完全相同的情况下，节点 v_i 与 v_j 极有可能属于一个社团。在如图 1 的网络中，可以很容易地获得节点 v_2、节点 v_5 与节点 v_7 的坐标：

$$B_2=(\frac{1}{2}\ 1\ \frac{1}{2}\ \frac{1}{2}\ 1\ \frac{1}{2}\ \frac{1}{3}\ \frac{1}{3}\ \frac{1}{3})$$

$$B_5=(1\ 1\ 1\ 1\ 1\ 1\ \frac{1}{2}\ \frac{1}{2}\ \frac{1}{2})$$

$$B_7=(\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}\ \frac{1}{2}\ 1\ \frac{1}{2}\ 1\ \frac{1}{2})$$

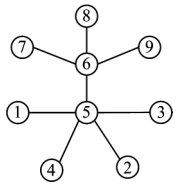

图 1 人工网络拓扑

2.2 节点之间的相似度

采用余弦距离表示节点之间的距离，其定义如下：

$$d_{ij}=\frac{(B_i,B_j)}{\|B_i\|\|B_j\|} \tag{3}$$

其中，向量 B_i 和 B_j 分别为节点 v_i 与 v_j 在网络 G 中的坐标，(B_i,B_j) 为向量的内积，$\|B_i\|\|B_j\|$ 为向量 B_i 和 B_j 的模长乘积。很明显，用 d_{ij} 来描述节点之间的相似度，其值越大，则表明向量之间的夹角越小，反映到图 G 中则是节点 v_i 与 v_j 越相似，它们在图 G 中传递信息的综合效率越相似，从而它们属于同一个社团的概率就越大。相反，d_{ij} 的值越小，则它们在同一社团内的概率就越小。

例如，图 1 中的节点 v_2、节点 v_5 与节点 v_7 之间的相似度分别为 $d_{25}=0.949$，$d_{27}=0.766$，$d_{57}=0.826$。其中节点 v_2 与节点 v_7 之间一个共同的邻居节点都没有，它们之间的相似度无法使用杰卡德(jaccard)指标、共同邻居(CN)指标等一些局部信息指标来描述，但是可以使用本文提出的指标来描述，并且由于本文的方法尽可能多地保留了原有网络的拓扑信息，因此可以很明显区分出节点 v_2、节点 v_5 与节点 v_7 的亲疏关系。

2.3 模块度

模块度是一种常用的社团划分质量评价标准，通过将社团内部的实际连边与其相应的零模型进行比对来反映社团划分的质量[5,12]。一个社团所对应的零模型就是与该社团具有某些相同拓扑性质而在其他方面完全随机的网络。由于度分布在网络中是最普遍的拓扑性质，因此目前在计算某种社团划分的模块度时往往将该种划分与和其具有相同度分布的零模型进行比较。模块度 Q 可表示为：

$$Q=\frac{1}{2M}\sum_{ij}(a_{ij}-p_{ij})\delta(C_i,C_j) \tag{4}$$

其中，$M=|E|$；a_{ij} 表示节点 v_i 与 v_j 之间是否存在连边：如果存在则 $a_{ij}=1$，如果不存在则 $a_{ij}=0$；C_i 与 C_j 分别表示节点 v_i 与 v_j 所属的社团，当 C_i 与 C_j 表示同一社团时，$\delta(C_i,C_j)$ 取 1，否则取 0；p_{ij} 表示在相应零模型中节点 v_i 与 v_j 之间存在连边的期望值，由于该零模型仅与网络 G 具有相同度分布，因此 $p_{ij}=k_i*k_j/2M$，其中 k_i 和 k_j 分别表示节点 v_i 与 v_j 的度数。

对给定的网络 G 进行最优划分，对应的模块度为 Q_G，则 Q_G 的理论取值范围为[0，1)。研究表明，如果 G 为真实的网络，那么 Q_G 的取值范围一般为 0.2～0.7，并认为模块度值超过 0.3 的网络具有较明显的社团结构[15]。

3　算法描述及分析

3.1　算法描述

K-means 算法通过不断地迭代计算出聚类中心，其聚类思想主要是使簇内的相似度尽量可能大，而簇间的相似度尽量小。将 K-means 算法的这一思想引入到社团发现中，即使得发现的社团内的节点具有较强的聚集性，而社团间的聚集性相对较弱。

算法的计算步骤如下：

输入：网络 G 的邻接矩阵

输出：网络 G 的社团结构及其模块度 Q 的值

Step1　使用广度优先算法获得节点之间的最短路径，得到网络 G 的最短路径矩阵；

Step2　通过式(2)计算出网络 G 的信息传递能效矩阵，得到每个节点的特征向量；

Step3　初始化社团个数 k=2，模块度 Q=0；

Step4　将节点特征向量所组成的多维数据集和 k 作为 k-means 算法的参数输入，以式(3)作为 k-means 相似函数，通过计算得到网络 G 的社团结构；

Step5　使用模块度指标作为评价函数来评价 Step4 得到的社团结构，从而得到模块度 Q_g，再计算得到 $\Delta Q = Q_g - Q$。如果 $\Delta Q>0$，设置 $Q=Q_g$，k=k+1 并跳转到 Step4；如果 $\Delta Q>0$，则输出网络 G 的社团结构和模块度 Q。

3.2　算法分析

本文算法使用了余弦距离来描述节点之间的相似度，即先抽象出节点的特征向量，再计算向量之间的余弦距离来衡量节点对的相似度，此方法可以在控制算法时间复杂度的同时尽可能少地丢弃节点的全局信息。如第 4 节中图 5 所示，图中节点 v_4 与多个社团的外围节点相连，一些现有的算法将节点 v_4 划分为了单独的社团，这显然是不合理的。使用本文算法可以明显提高算法的分辨率，使得节点 v_4 也能合理地被划归到某一社团，从而平衡社团的规模。尽管张学仁等人[11]提出社团中的核心节点往往有着高于该社团内其他节点的度，但是由于网络中存在的社团并不平衡，例如在同一网络中某些社团内节点的度普遍会高于另外社团内节点的度，因此以度、核度等指标选取的社团核心节点往往并不十分准确。本文算法在社团凝聚的过程中，社团的中心是虚拟的，这避免了一些算法对社团中心节点选取敏感的缺陷。

影响本文算法效率的主要因素为节点相似度特征向量的求取与社团的凝聚。通过采用广度优先的方法求取图中各个节点对之间的最短距离，再计算出节点相似度特征向量，其时间复杂度为 $O(n^2)$。社团凝聚采用 K-means 算法，因此其算法的时间复杂度为 $O(knt)$，其中 k 为社团个数，n 为节点数目，t 为迭代次数(为常数)。本文算法的整体时间复杂度为 $O(n^2)$。

4　实验结果与分析

为了验证算法的有效性，选取了 3 个比较知名的真实社会网络进行实验，分别是美国一所大学的空手道俱乐部成员之间相互的社会关系网络(Zachary Karate Club)[14]、P. Gleiser 爵士音乐家网络(Jazz Musician)[13]和 Facebook 社交网络，这 3 个网络的基本信息如表 1 所列。通过对这 3 个具有明显社团结构的网络进行实验并与一些经典的算法进行比较，来说明本文算法在社团发现中的优势。实验平台是 CPU 为 Intel Core i3-M350 内存 4GB 的 Win7 系统，实验编程环境为 R x64 3.1。

表 1　实验网络拓扑

网络名称	节点数	边数	最大度数	聚类系数
空手道网络	34	78	17	0.07
爵士音乐家网络	198	2742	100	0.14
Facebook 网络	1899	20295	339	0.05

4.1　空手道网络仿真分析

现实中绝大多数的网络都存在社团结构，本文在此处引用了一个经典网络空手道网络作为分析的对象。空手道网络是 Zachary 在 20 世纪 70 年代耗时两年多，通过观察美国一所大学的空手道俱乐部成员之间的相互关系而构建的一个关系网。在这期间，校长与空手道教练就是否应该收取会费而产生了分歧，最终导致该校的空手道俱乐部分裂为两个分别以校长和教练为中心的子社团。

使用本文所提出的算法对空手道网络进行社团划分，对实验过程中的模块度 Q 以及变量 k 进行记录，如图 2 所示。当 k 取 2 时算法将空手道网络分为了两个社团，此时模块度 Q 的取值为 0.371，具体社团结构如图 3 所示。图 3 中的社团结构与 Zachary 实际记录的社团结构一致，即最终空手道俱乐部分裂成了分别由节点 v_1 所代表的以教练为中心的子社团和节点 v_{33} 所代表的以校长为中心的子社团，但此时模块度 Q 的值并非为最佳。

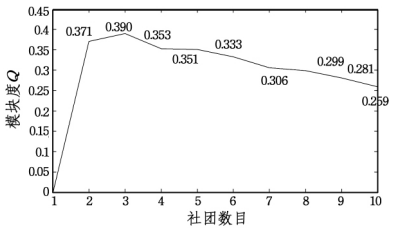

图 2　空手道网络算法划分过程

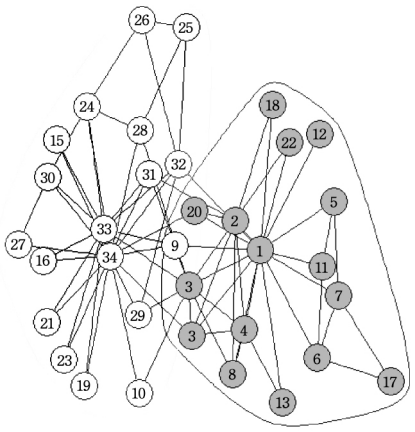

图 3　空手道网络的真实社团结构

由图 2 中可以看到模块度 Q 在 k 取 4 时才出现了下降，故算法在此时停止搜索，并输出 k 为 3 时的社团结构及模块

度 Q 的值作为最佳社团结构和最大模块度。此时模块度 Q 的值为 0.390，形成了如图 4 所示的 3 个社团。图 4 表明，节点 v_{25}、节点 v_{26} 和节点 v_{32} 形成了一个完全联通的子图，说明 3 人之间有着很密切的关系；节点 v_{28} 和节点 v_{29} 均只与两个较大社团的外层节点相连，因此这 5 个节点独立成为一个社团反而提高了模块度 Q 的值。同时，本文使用其他的社团发现算法对空手道网络进行了实验，将各个算法所得到的社团结构与空手道网络的实际社团结构做了比较，对各个算法结果的准确性做了比较，比较结果如表 2 所列。

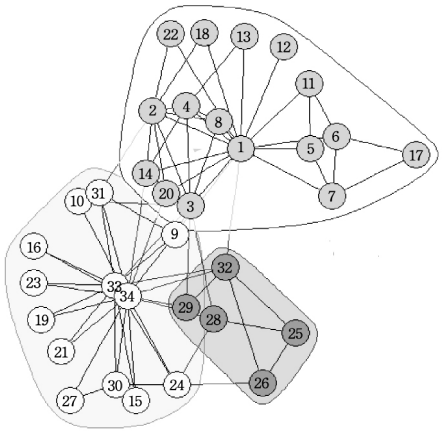

图 4 本文算法所得到的社团结构

表 2 各算法对于空手道网络结果的比较

算法名称	划分社团的数目	社团规模	准确率/%
本文算法	3	5,13,16	85.3
GN 算法	5	1,5,6,10,12	64.7
随机游走算法	5	4,5,7,9,9	52.9
标签扩散算法	3	5,10,19	82.4
spin-glass	4	5,6,11,12	67.6

4.2 爵士音乐家网络仿真分析

爵士音乐家网络是一个典型的社会网络，它描述的是爵士音乐家之间相互的联系。从表 1 的数据中可以看出，爵士网络比空手道网络更加复杂，无论是节点的数目还是网络的密度。为了检验本文算法的效果，分别将本文算法与其他一些经典的社团发现算法做比较，其中包括 GN 算法、随机游走算法、标签扩散算法和 spin-glass 算法，实验结果如表 3 所列。

表 3 爵士音乐家网络各算法结果的比较

算法名称	划分社团的数目	模块度 Q	社团规模
本文算法	3	0.440	60,61,77
GN 算法	39	0.405	49,45,59,4,4,3,2,剩余 32 个社团大小均为 1
随机游走算法	11	0.438	52,63,70,1,1,1,1,1,1,3,4
标签扩散算法	2	0.282	139,59
spin-glass	6	0.430	63,68,61,4,1,1

从表 3 可以发现，与特征向量法、GN 算法和标签扩散算法相比，本文算法发现的社团结构明显更符合模块度标准。尽管随机游走与 spin-glass 算法在模块度 Q 值方面与本文算法接近，但是它们的社团规模差异巨大，在划分社团结构的时候产生了许多大小为 1 的社团，这明显违背了我们对社团的认知。

进一步对这些小社团进行研究发现，这些节点往往处在社团的边界位置。例如爵士网络中的节点 v_{197}，尽管它的度达到了 v_{36}，但是与之相连的节点大多数为社团的外围节点，因此它被独立为了一个社团。类似这类节点起到了桥接各个社团的作用，其在网络中的位置如图 5 中的节点 v_4，它与节点 v_1，v_2，v_3 组成的社团 C_1、节点 v_5，v_6，v_7，v_8 组成的社团 C_2 以及社团 C_3，C_4，C_5 均存在联系。显然，节点 4 极有可能更加倾向某个社团，但是随机游走算法和 spin-glass 算法不能有效地识别这类边界节点的归属。而本的算法考虑了节点在网络中的全局效应，因此提高了识别的精度，平衡了社团的规模，找出的社团结构也更加符合模块度标准。

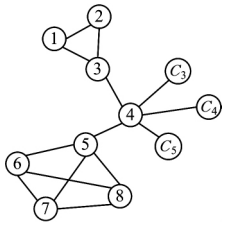

图 5 人工网路拓扑

4.3 Facebook 网络仿真分析

Facebook 网络是一个真实的社交网络，它描述的是 Facebook 用户的相互关注情况。从表 1 的数据中可以看出，Facebook 网络比爵士网络及空手道网络更加复杂。为了检验本文算法的效果，分别将本文算法与其他一些经典的社团发现算法做了比较，其中包括 GN 算法、随机游走算法、标签扩散算法和 spin-glass 算法，实验结果如表 4 所列。

表 4 Facebook 网络各算法结果的比较

算法名称	划分社团的数目	模块度 Q
本文算法	10	0.148
GN 算法	无法在有效时间内计算出结果(1 个小时)	
随机游走算法	452	0.0973
标签扩散算法	7	0.00158
spin-glass	由于本文使用的网络中存在孤立点，因此无法计算	

由于 GN 算法的时间复杂度过高，导致无法在有效时间内计算出结果。spin-glass 算法对网络拓扑有一定的敏感性，需要不存在孤立点的网络拓扑，因此在 Facebook 网络上也不适用。标签扩散算法和随机游走算法的时间复杂度很低，但是其获得的结果并不令人满意。随机游走算法得到了 452 个社区，导致产生了许多只有几个节点的社团，且模块度 Q 只有 0.0973。尽管标签扩散算法很好地控制了社团的规模，但是其聚类的精度过低，导致其模块度 Q 的值是几种算法中最低的，仅有 0.00158。而本文的算法有效地平衡了时间复杂度和精度，在 $O(n^2)$ 的时间复杂度内，有效地保证了社团发现的精度。

结束语 社团发现的实质是按照某一相似度指标对网络中的节点进行聚类。本文在信息传递理论的基础上对节点进行了二次抽象，将网络中的节点抽象成了多维数据集，再利用基于传统聚类算法的思想处理得到的多维数据集，最终得到了理想的社团结构。这一将网络的拓扑结构抽象为多维数据集的思想可以将许多成熟的处理多维数据集的算法引入到社团发现领域中，同时也可以处理带权的网络，这将是我们下一步的研究重点。

参 考 文 献

[1] WATTS D J. A twenty-first century science[J]. Nature, 2007,

445(7127):489-489.

[2] REN G,WANG X. Epidemic spreading in time-varying community networks[J]. Chaos: An Interdisciplinary Journal of Nonlinear Science,2014,24(2):023116.

[3] 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.

[4] MEHRA A. The Development of Social Network Analysis: A Study in the Sociology of Science[J]. Social Networks,2005,27(4):377-384.

[5] NEWMAN M E,GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics,2004,69(2 Pt 2):026113-026113.

[6] NEWMAN M E J. Detecting community structure in networks [J]. The European Physical Journal B-Condensed Matter and Complex Systems,2004,38(2):321-330.

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

[8] NEWMAN M E J. Clustering and preferential attachment in growing networks[J]. Physical Review E,2001,64(2):025102.

[9] ZHOU T,LÜ L,ZHANG Y C. Predicting missing links via local information[J]. The European Physical Journal B-Condensed Matter and Complex Systems,2009,71(4):623-630.

[10] FORTUNATO S,LATORA V,MARCHIORI M. Method to find community structures based on information centrality[J]. Physical review E,2004,70(5):056104.

[11] **刘海峰，刘守生，张学仁. 聚类模式下一种优化的** K-means **文本特征选择**[J]. **计算机科学**,2011,38(1):195-197.

[12] ZHANG T,WU B. A method for local community detection by finding core nodes[C]// Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012). IEEE Computer Society,2012:1171-1176.

[13] **邓智龙，淦文燕. 复杂网络中的社团结构发现方法**[J]. **计算机科学**,2012,39(z6):103-108.

[14] GLEISER P M,DANON L. Community structure in jazz[J]. Advances in complex systems,2003,6(4):565-573.

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

[16] **张聪，沈惠璋，李峰，等. 复杂网络中社团结构发现的多分辨率密度模块度**[J]. **物理学报**,2012,61(14):148902-148902.

(**上接第** 398 **页**)

结束语　本文利用差分隐私保护技术解决了社交网络项目推荐过程中推荐边权重泄露用户隐私问题。提出了一种集成拉普拉斯机制与指数机制的推荐方法 DPSR，**该方法解决了边权中异常点影响推荐结果的问题。在与同类方法对比中，**DPSR **具有较好的推荐精度。**

参 考 文 献

[1] SWEENEY L. k-anonymity: A model for protecting privacy[J]. International Journal of Uncertainty,Fuzziness and Knowledge-Based Systems,2002,10(5):557-570.

[2] DWORK C. Differential Privacy[C]// Proc of the 33rd Int Colloquium on Automata, Languages and Programming (ICALP 2009). 2006:1-12.

[3] DWORK C. Differential Privacy: A Survey of Results[C]// Proc of the 5th Int Conf on Theory and Applications of Models of Computation (TAMC 2008). 2008:1-19.

[4] DWORK C,MCSHERRY F,NISSIM K,et al. Calibrating noise to sensitivity in private data analysis[C]// Proc. of the 3rd Theory of Cryptography Conf (TCC 2006). 2006:265-284.

[5] MCSHERRY F,MIRONOV I. Differentially Private Recommender Systems: Building Privacy into the Netflix Prize Contenders[C]// Proc. of the 15th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining (KDD 2009). 2009:627-636.

[6] RIBONI D,BETTINI C. Differentially-private release of check-in data for venue recommendation[C]// Proc of IEEE Int Conf on Pervasive Computing and Communications (PerCom 2014). 2014:190-198.

[7] GUERRAOUI R,KERMARREC A,PATRA R,et al. D2P: Distance-Based Differential Privacy in Recommenders[J]. PVLDB,2015,8(8):862-873.

[8] MACHANAVAJJHALA A,KOROLOVA A,SARMA A. Personalized social recommendations: accurate or private[J]. PVLDB,2011,4(7):440-450.

[9] JORGENSEN Z,YU T. A Privacy-Preserving Framework for Personalized,Social Recommendations[C]// Proc of the Int Conf on Extending Database Technology (EDBT 2014). 2014:571-582.

[10] MCSHERRY F,TALWAR K. Mechanism Design via Differential Privacy[C]// Proc of the 48th Annual IEEE Symp on Foundations of Computer Science (FOCS 2007). 2007:94-103.

[11] Fortunato S. Community detection in graphs[J]. Phys Reps,2010,486(3):75-174.

[12] DINH N T,LI X,THAI M T. Network Clustering via Maximizing Modularity: Approximation Algorithms and Theoretical Limits[C]// 2015 IEEE Int Conf on Data Mining (ICDM 2015). 2015:101-110.

[13] MCSHERRY F. Privacy Integrated Queries: An Extensible Platform for Privacy-Preserving Data Analysis[C]// Proc of the ACM SIGMOD Int Conf on Management of Data (SIGMOD 2009). 2009:19-30.

[14] JARVELIN K,KEKALAINEN J. Cumulated gain-based evaluation of ir techniques[J]. TOIS,2002,20(4):422-446.

[15] LIBEN-NOWELL D,KLEINBERG J. The Link-Prediction Problem for Social Networks[J]. JASIST,2007,58(7):1019-1031.