◎博士论坛◎

半指导的核聚类检测网络社团方法

付立东[1,2]
FU Li-dong[1,2]

1.西安科技大学 计算机学院,西安 710054
2.西安电子科技大学 计算机科学工程学院,西安 710071
1.School of Computer,Xi'an University of Science and Technology,Xi'an 710054,China
2.School of Computer Science & Engineering,Xidian University,Xi'an 710071,China
E-mail:fulidong2005@163.com

FU Li-dong.Kernel approach for detecting communities in complex networks based on semi-supervised clustering.Computer Engineering and Applications,2010,46(5):1-3.

Abstract: In recent years,the problem of community structure detection has attracted more and more attention and many approaches have been proposed.In this context,Li et al recently propose modularity density objective function for community detecting called the *D* function.Empirically,higher values of the *D* function have been shown to correlate well with good community structures.However,optimization of the function is a NP-hard problem.In this paper,how to optimize the *D* function can be formulated as a semi-supervised approach problem.The equivalence of the semi-supervised and the kernel *k*-means based on modularity density are also proved and a new semi-supervised kernel clustering approach is proposed.The approach is illustrated and compared with direct kernel approach based on modularity density by using a classic computer generated networks.Experimental results show the significance of the proposed approach,particularly,in the cases when community structure is obscure.
Key words: complex networks;community structure;modularity density;kernel approach;semi-supervised clustering

摘 要:近年来,复杂网络中的社团发现越来越受到研究人员的关注并且许多方法被提了出来。在这种背景下,最近李等人提出了一种用来评估社团质量的函数,称之为模块密度函数(即 *D* 值)。该函数显示了较高的 *D* 值对应于较好的社团结构,然而,优化该函数是一个 NP 难问题。通过模块密度函数 *D* 的半指导聚类优化,论证了模块密度函数的半指导聚类与核 *k* 方法的等价性并提出了一种新的半指导核聚类检测复杂网络社团方法。在一个经典的计算机产生的随机网络中检验了该算法,并与基于模块密度的直接核方法做了比较。特别地,当网络中社团结构变得模糊时,实验结果显示这种新的算法在发现复杂网络社团上是有效的。
关键词:复杂网络;社团结构;模块密度;核方法;半指导聚类
DOI:10.3778/j.issn.1002-8331.2010.05.001 **文章编号**:1002-8331(2010)05-0001-03 **文献标识码**:A **中图分类号**:TP399

1 引言

复杂网络中,基于元素间相互作用模式发现社团结构越来越受到研究人员的关注[1-2]。许多领域中,网络扮演着重要的角色,例如英特网,社会网络,以及生物网络。所有复杂网络可以模型化成图 $G=(V,E)$ 的形式,这里 V 表示一组顶点,并且 E 表示成顶点之间相互作用的一组边。在网络中,一个社团结构被定义成内部连接紧密,而外部连接相对松散的顶点子集。这些社团结构暗含了复杂网络的功能,并用来帮助研究人员理解复杂网络的增长机制。结果,复杂网络中社团刻画与检测成了当今最杰出的问题之一[3-4]。

复杂网络中的社团鉴定与检测能帮助进一步理解和探索网络。近年来,许多方法被提出来用以检测复杂网络中的社团结构,例如基于边介数方法[5],非负矩阵分解方法[6],模糊聚类方法[7],其中有许多算法是基于由 Newman 提出的用来评估复杂网络中社团结构的模块函数 Q[5]的优化上得到的。然而,最近的研究表明[8],这个模块函数不能检测出小于一种内在尺寸的社团结构。这主要是因为一个网络的划分高敏感于该网络边的总数目。

为了克服先前这种社团评估函数,最近,李等人提出了一种新的社团评估函数[9],称之为模块密度(所谓 *D* 值)。这个函数 *D* 值在评估网络的划分,即社团结构时,不仅考虑了社团边的多少,而且考虑了社团中节点的多少。通过一种过渡操作核矩阵,李等人同时证明了模块密度函数和目标核 *k* 方法的等价

基金项目:国家自然科学基金重点项目(the Key National Natural Science Foundation of China under Grant No.60933009);国家自然科学基金(the National Natural Science Foundation of China under Grant No.60970065);教育部高校博士点基金资助项目(the Ph.D Programs Foundation of Ministry of Education of China under Grant No.200807010013)。
作者简介:付立东(1973-),男,博士,讲师,主要研究领域为复杂网络社团检测,中心度量。
收稿日期:2009-11-02 **修回日期**:2009-12-17

性。然而,核 k 方法敏感于网络的初始划分,可看作是该方法的一种缺陷。利用潜藏的马尔科夫随机场半指导模型 HRMF 优化了模块密度并证明了基于模块密度的半指导方法与核 k 方法的等价性,基于这种等价性,使用半指导化了模块密度来初始化核 k 方法,从而提升社团的检测结果。

2 核 k-means 方法

有必要首先介绍一下核 k-means[10]方法,因为将看到这种核 k-means 方法将等价于基于模块密度的半指导方法。给定一组 $\boldsymbol{v}_i \in R^n$ 维的数据向量$\{\boldsymbol{v}_i\}_{i=1}^n$,核 k-means 目的是发现 p 个不连接的划分$\{\boldsymbol{V}_c\}_{c=1}^p$,以便下列的目标函数最小化:

$$H(\{\boldsymbol{V}_c\}_{c=1}^p)=\sum_{c=1}^{p}\sum_{\boldsymbol{v}_i\in\boldsymbol{V}_c} w_i \parallel \phi(\boldsymbol{v}_i)-m_c \parallel^2 \tag{1}$$

这里 $m_c=\left(\sum_{\boldsymbol{v}_i\in\boldsymbol{V}_c}\phi(\boldsymbol{v}_i)\right)\Big/|\boldsymbol{V}_c|$,$\phi$ 是一种映射$\{\boldsymbol{v}_c\}_{i=1}^n$中的向量到一个高维空间的函数,而不需要知道 ϕ 的确切表达。如果扩展核 k-means 函数中的距离 $\parallel \phi(\boldsymbol{v}_i)-m_c \parallel^2$,可得到:

$$\parallel \phi(\boldsymbol{v}_i)-m_c \parallel^2=\phi(\boldsymbol{v}_i)\cdot\phi(\boldsymbol{v}_i)-\frac{2\sum_{\boldsymbol{v}_j\in\boldsymbol{V}_c}\phi(\boldsymbol{v}_i)\cdot\phi(\boldsymbol{v}_j)}{|\boldsymbol{V}_c|}+\frac{\sum_{\boldsymbol{v}_j,\boldsymbol{v}_l\in\boldsymbol{V}_c}\phi(\boldsymbol{v}_j)\phi(\boldsymbol{v}_l)}{|\boldsymbol{V}_c|^2} \tag{2}$$

注意到公式(2)包含的数据点仅是以内积的形式出现。因此,能使用的给定的核矩阵 $\boldsymbol{K}$ 去计算映射空间中节点之间的距离,这里 $\boldsymbol{K}_{ij}=\phi(\boldsymbol{v}_i)\cdot\phi(\boldsymbol{v}_j)$。假设输入量是内积形式的核矩阵 $\boldsymbol{K}$;应用这个核矩阵 $\boldsymbol{K}$,距离计算公式(2)$d(\boldsymbol{v}_i,m_c)=\parallel \phi(\boldsymbol{v}_i)-m_c \parallel^2$ 可以写成:

$$\boldsymbol{K}_{ii}-\frac{2\sum_{\boldsymbol{v}_j\in\boldsymbol{V}_c}\boldsymbol{K}_{ij}}{|\boldsymbol{V}_c|}+\frac{\sum_{\boldsymbol{v}_j,\boldsymbol{v}_l\in\boldsymbol{V}_c}\boldsymbol{K}_{jl}}{|\boldsymbol{V}_c|^2} \tag{3}$$

结果一旦获得一个合适的核矩阵 $\boldsymbol{K}$,可以设计一种相似于 k-means 算法以单调递减这个核 k-means 目标函数。基本的核 k-means 算法如参考文献[10]所示。特别地,核 k-means 可以表达成矩阵迹的形式[10]

$$H(\{\boldsymbol{V}_c\}_{c=1}^k)=\text{trace}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})-\text{trace}(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{Y}) \tag{4}$$

其中 $\boldsymbol{\Phi}$ 是所有 $\boldsymbol{v}_i$ 向量组成的矩阵。对任何给出的正半定核矩阵 $\boldsymbol{K}$,Dhillon 等人指出 $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}$ 能够等于这个核矩阵,并且 $\boldsymbol{Y}$ 是一个正交的 $N\times k$ 数据向量粘沾矩阵,即 $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}=\boldsymbol{I}_k$,如果 $\boldsymbol{v}_i$ 属于子集,那么它的元素 $Y_{ic}=\dfrac{(w(v_i))^{1/2}}{(\sum_{\boldsymbol{v}_i\in\boldsymbol{V}_c}w(v_i))^{1/2}}$,否则 $Y_{ic}=0$。一旦 $\boldsymbol{K}$ 确定,注意到 $\text{trace}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})$是一个常数,那么核 k-means 的最小化等价于 $\text{trace}(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{Y})$最大化,即

$$\min H\propto\max\text{trace}(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{Y}) \tag{5}$$

3 半指导聚类 HMRF 模型

当聚类时,通常有一些关于聚类数据的背景知识,利用这些背景知识可以帮助提升聚类结果。假设这些背景知识来自于分段的必然关联或非关联的约束形式。这些约束对复杂网络也是自然的,因为约束关系可以通过网络中的边被明确地捕获。

介绍一种新近提出的半指导聚类目标函数模型,这种模型可以优化社团发现与评估的模块密度函数。Basu 等人基于潜藏的马尔科夫随机场为半指导聚类提出了一种 HMRF 模型框架[11]。选择平方欧氏距离作为聚类偏离策略并且使用推广的波兹潜力作为约束偏离惩罚,这个半指导聚类目标可以表达成:

$$J(\{\boldsymbol{V}_c\}_{c=1}^k)=\frac{1}{2}\sum_{c=1}^{k}\sum_{\boldsymbol{x}_i\in p^*}\parallel \boldsymbol{x}_i-\mu_c \parallel^2+\sum_{\substack{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in M\\ \text{s.t.}l_i\neq l_j}}w_{ij}+\sum_{\substack{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in C\\ \text{s.t.}l_i=l_j}}w_{ij} \tag{6}$$

这里 M 是一组必然关联约束,C 是一组不关联约束,w_{ij} 是对顶点向量 $\boldsymbol{x}_i$ 与 $\boldsymbol{x}_j$ 之间偏离约束的惩罚代价,l_i 指示 $\boldsymbol{x}_i$ 的聚类标签。这个目标函数的第一项是标准的 k-means 函数,第二项是对偏离必然关联约束的一种惩罚,并且第三项是对偏离非关联约束的一种惩罚。

为方便起见,改变一下公式(6)惩罚函数:如果两个顶点在不同的聚类中,对必然关联偏离,替换添加一个惩罚项,而是从目标中减去对应的惩罚。如果对所有的必然约束的权重和是一个常数,那么这就等价于在最初目标函数附上一个额外约束。因此最小化 $J(\{p_c\}_{c=1}^k)$等价于最小化:

$$\frac{1}{2}\sum_{c=1}^{k}\sum_{\boldsymbol{x}_i\in p^*}\parallel \boldsymbol{x}_i-\mu_c \parallel^2-\sum_{\substack{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in M\\ \text{s.t.}l_i=l_j}}w_{ij}+\sum_{\substack{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in C\\ \text{s.t.}l_i=l_j}}w_{ij} \tag{7}$$

4 模块密度与社团结构

让 $G=(V,E,A)$表示一个无向网络,其中 V 是一个包含着 n 个顶点的集合,E 是一个包含 m 条边的集合,$\boldsymbol{A}$ 是一个 $n\times n$ 维的对称邻接矩阵。A_{ij} 代表着顶点 i 和顶点 j 之间的边的权重。想要划分网络 G 的顶点到 k 个非重叠社团结构,那么模块密度函数被定义成:

$$D(\{\boldsymbol{V}_c\}_{c=1}^k)=\sum_{c=1}^{k}\frac{L(\boldsymbol{V}_c,\boldsymbol{V}_c)-L(\boldsymbol{V}_c,\overline{\boldsymbol{V}}_c)}{|\boldsymbol{V}_c|} \tag{8}$$

这里 $L(\boldsymbol{V}_c,\boldsymbol{V}_c)$等于 $\sum_{i,j\in V_c}A_{ij}$,度量了社团内部边的权重和,$\overline{\boldsymbol{V}}_c$ 是 $\boldsymbol{V}_c$ 的补集。进一步,让 $\boldsymbol{V}_c$ 的度等于社团 $\boldsymbol{V}_c$ 内的顶点到所有其他顶点的边数,即 $degree(\boldsymbol{V}_c)=L(\boldsymbol{V}_c,\boldsymbol{V})$,有 $L(\boldsymbol{V}_c,\overline{\boldsymbol{V}}_c)=degree(\boldsymbol{V}_c)-L(\boldsymbol{V}_c,\boldsymbol{V}_c)$。

结果,$L(\boldsymbol{V}_c,\overline{\boldsymbol{V}}_c)$度量了不同社团之间边的权重。$|\boldsymbol{V}_c|$是社团 $\boldsymbol{V}_c$ 内顶点的数目,度量了一个社团的大小。

因此,模块密度 D 值等于所有社团平均内部度减去所有社团平均外部度。较大 D 的值意味着较强壮的社团结构。如果一个划分给出了所有社团平均内部度不大于平均外部度,那么模块密度 D 值小于等于零。

5 半指导的核聚类方法

使用半指导方法来优化模块密度,并论证了基于模块密度的半指导聚类与核 k-means 方法之间在数学计算上具有等价性。基于此,提出了半指导的核聚类算法。

5.1 模块密度的半指导聚类优化

正如上面提到的,HMRF 目标函数使用三项函数来表达:一项是要聚类的目标函数,一项是强制必然关联约束,一项是

强制非关联约束。使用相同的策略,现在考虑半指导图聚类的模块密度目标函数。用公式(8)替换公式(7)中的第一项,那么,模块密度的半指导聚类可以被表达成:

$$J(\{\boldsymbol{V}_c\}_{c=1}^{k})=\sum_{c=1}^{k}\frac{L(\boldsymbol{V}_c,\boldsymbol{V}_c)-L(\boldsymbol{V}_c,\overline{\boldsymbol{V}}_c)}{|\boldsymbol{V}_c|}+\sum_{\substack{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in M\\ \text{s.t.}l_i\neq l_j}}\frac{w_{ij}}{|\boldsymbol{V}_{l_i}|}-\sum_{\substack{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in C\\ \text{s.t.}l_i=l_j}}\frac{w_{ij}}{|\boldsymbol{V}_{l_j}|} \tag{9}$$

在公式(9)中,注意到第二和第三项有相反的符号,与 HMRF 模型的表达有所不同,这是因为想要最大化公式(9)。重写公式(9),有:

$$J(\{\boldsymbol{V}_c\}_{c=1}^{k})=\sum_{c=1}^{k}\frac{L(\boldsymbol{V}_c,\boldsymbol{V}_c)-L(\boldsymbol{V}_c,\overline{\boldsymbol{V}}_c)}{|\boldsymbol{V}_c|}+\sum_{c=1}^{k}\sum_{\substack{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in \boldsymbol{V}_c\\ (\boldsymbol{x}_i,\boldsymbol{x}_j)\in M}}\frac{w_{ij}}{|\boldsymbol{V}_c|}-\sum_{c=1}^{k}\sum_{\substack{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in \boldsymbol{V}_c\\ (\boldsymbol{x}_i,\boldsymbol{x}_j)\in C}}\frac{w_{ij}}{|\boldsymbol{V}_c|} \tag{10}$$

定义一个对角度矩阵 $\boldsymbol{D}$,有 $\boldsymbol{D}_{ii}=\sum_{j=1}^{n}\boldsymbol{A}_{ij}$,并且 $\boldsymbol{W}$ 是约束矩阵,以便对一个非关联 W_{ij} 等于 $-w_{ij}$,对一个必然关联 W_{ij} 等于 w_{ij},其他为 0。进一步,介绍一个聚类 c 的指示向量 $\boldsymbol{x}_c$,有 $\boldsymbol{x}_c(i)=1$,如果聚类 c 包含顶点 i。令 $\tilde{\boldsymbol{x}}_c=\boldsymbol{x}_c/(\boldsymbol{x}_c^{\mathrm{T}}\boldsymbol{x}_c)^{1/2}$,那么模块密度函数的半指导聚类可被写成:

$$J(\{\boldsymbol{V}_c\}_{c=1}^{k})=\sum_{c=1}^{k}\frac{\boldsymbol{x}_c^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_c-\boldsymbol{x}_c^{\mathrm{T}}(\boldsymbol{D}-\boldsymbol{A})\boldsymbol{x}_c}{\boldsymbol{x}_c^{\mathrm{T}}\boldsymbol{x}_c}+\sum_{c=1}^{k}\frac{\boldsymbol{x}_c^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}_c}{\boldsymbol{x}_c^{\mathrm{T}}\boldsymbol{x}_c}+\sum_{c=1}^{k}\frac{\boldsymbol{x}_c^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}_c}{\boldsymbol{x}_c^{\mathrm{T}}\boldsymbol{x}_c}=$$

$$\sum_{c=1}^{k}\tilde{\boldsymbol{x}}_c^{\mathrm{T}}(2\boldsymbol{A}-\boldsymbol{D}+2\boldsymbol{W})\tilde{\boldsymbol{x}}_c \tag{11}$$

定义一个对应的 $n\times k$ 顶点粘质矩阵 $\boldsymbol{Z}$,并且 $\boldsymbol{Z}=\tilde{\boldsymbol{X}}$,这里 $\tilde{\boldsymbol{X}}$ 的第 c 列等于 $\tilde{\boldsymbol{x}}_c^{\mathrm{T}}$,且有 $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}=\boldsymbol{I}_k$。那么能进一步简化等式(7)中模块密度 D 值的半指导聚类为:

$$J(\{\boldsymbol{V}_c\}_{c=1}^{k})=\mathrm{trace}(\boldsymbol{Z}^{\mathrm{T}}(2\boldsymbol{A}-\boldsymbol{D}+2\boldsymbol{W})\boldsymbol{Z}) \tag{12}$$

令 $\boldsymbol{K}=2\boldsymbol{A}-\boldsymbol{D}+2\boldsymbol{W}$,在公式(12)中,因为 $\boldsymbol{Z}$ 相似于公式中的 $\boldsymbol{Y}$(都是正交的指示向量)。因此,这个半指导聚类的最大化等价于 $\max_{\boldsymbol{Z}}\mathrm{trace}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{Z})$。结果模块密度半指导聚类同样能被表达成 $\mathrm{trace}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{Z})$迹的最大化。对核矩阵来说,一个要求就是要保证核 $\boldsymbol{K}$ 是正半定定义,以确保核算法收敛,可以添加 $\sigma\boldsymbol{I}$ 到核矩阵中去。其中 σ 是任意实数,$\boldsymbol{I}$ 是一个单位矩阵。但不会影响聚类结果,因为有

$$\mathrm{trace}(\boldsymbol{Z}^{\mathrm{T}}(\sigma\boldsymbol{I}+\boldsymbol{K})\boldsymbol{Z})=\sigma\cdot\mathrm{trace}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})+\mathrm{trace}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{Z})=$$

$$\sigma k+\mathrm{trace}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{Z}) \tag{13}$$

因此,基于模块密度的半指导聚类在 $\boldsymbol{K}=\sigma\boldsymbol{I}+2\boldsymbol{A}-\boldsymbol{D}+2\boldsymbol{W}$ 的情况下与核 k-means 在数学计算上是等价的。

5.2　半指导的核聚类算法

由于核 k-means 对网络的初始划分比较敏感,基于模块密度的半指导聚类与核 k-means 方法的等价性,可以使用半指导方法来产生初始的聚类,在半指导方法中,可以使用网络的约束条件及深度优先搜索算法来产生初始聚类,然后使用核 k-means 方法对这些初始的聚类再一次聚类。基于以上分析,可以设计出一种模块密度的半指导核聚类算法以检测复杂网络的社团结构。这个算法可以描述为:

输入 $\boldsymbol{A}$:网络的邻接矩阵,k:社团数目,$\boldsymbol{W}$:约束矩阵,$\boldsymbol{K}$:核矩阵,$t_{\max}$:迭代的最大次数

输出 $\{\boldsymbol{V}_c\}_{c=1}^{k}$:最终的社团划分

1.生成一个顶点权重向量 $\boldsymbol{\alpha}$。

2.形成矩阵 $\boldsymbol{K}=2\boldsymbol{A}-\boldsymbol{D}+2\boldsymbol{W}$。

3.通过添加 $\sigma\cdot\boldsymbol{I}$ 到 $\boldsymbol{K}$ 中,以保证 $\boldsymbol{K}$ 是正半定定义。

4.计算 $\boldsymbol{K}=\sigma\cdot\boldsymbol{I}+2\boldsymbol{A}+2\boldsymbol{W}$。

5.应用矩阵 $\boldsymbol{W}$ 中的约束条件及深度优先搜索算法产生初始聚类 $\{\boldsymbol{V}_c^{(0)}\}_{c=1}^{k}$。

6.返回 $\{\boldsymbol{V}_c\}_{c=1}^{k}=Kernel\text{-}k\mathrm{means}(\boldsymbol{K},k,t_{\max},\boldsymbol{\alpha},\{\boldsymbol{V}_c^{(0)}\}_{c=1}^{k})$如参考文献[10]中所示。

6　算法检验

在一个众所周知的预定义计算机产生的网络社团结构中[5],检验了提出的基于模块密度的半指导算法并与随机初始化的核 k 算法做了比对。这个人工网络包含 128 个顶点,这些顶点被划分成相同大小的 4 个社团结构,每个社团结构包含 32 个顶点。边使用了两个概率被随机放置,以使得每一个顶点的平均度等于 16,即 $z_{in}+z_{out}=16$,这里 z_{in} 是每一个顶点连接的平均落在相同社团中的边数目,z_{out} 是每一个顶点连接的平均落在不同社团间的边数目。图 1 展示了该随机网络中 128 个顶点在 $z_{in}=12$,$z_{out}=4$,时被划分成的 4 个社团结构时的大概情况。这个随机网络被广泛应用于检验社团发现算法[4-5]。

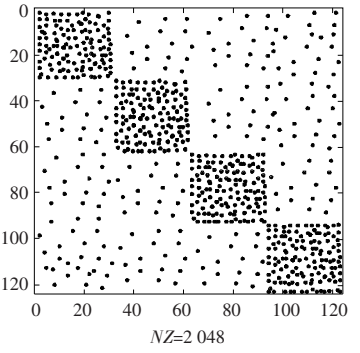

图 1　计算机产生的网络中 128 个顶点在 $z_{in}=12$,$z_{out}=4$ 时被划分成的 4 个社团结构时的大概情况

对这个实验的计算结果概括在图 2 中。图 2 展示了通过基于模块密度 D 的半指导核算法及直接的核 k-means 算法随着 z_{out} 的变化能正确检测出顶点的百分比。从图 2 可看出,当 $z_{out}\leqslant 5$ 时,两种方法都能正确检测出多于 96%的顶点数。然而,当 $5<z_{out}\leqslant 8$ 时,暗示着这 4 个社团的顶点越来越难以检测,基于

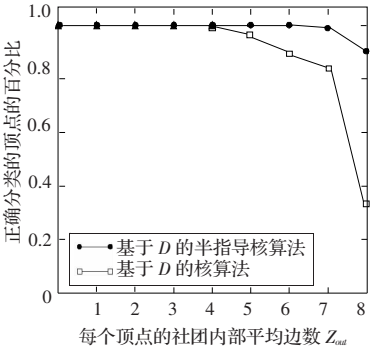

图 2　两种方法在预知社团结构的计算机产生图上对顶点的分类结果。刻画了随着 z_{out} 的变化,两种方法能正确分类出顶点的百分比

(下转 16 页)