

复杂网络社团的投影聚类划分

李伟, 杨晓峰, 张重阳, 汤可宗, 杨静宇
(南京理工大学 计算机系, 江苏 南京 210094)

摘要: 社团结构划分对研究复杂网络有重要作用, 由于该问题的复杂性, 复杂网络中的社团划分问题成为近期的一个研究热点. 从经典数据分析的角度研究了复杂网络的社团结构, 首先依据网络的拓扑信息, 将网络节点投影成高维空间的点, 使得一个网络对应到高维空间中的一个点分布; 接着使用主分量分析方法 PCA 对高维点分布降维, 保留点群分布的主要结构信息; 再通过 K-means 聚类结果来推断网络的社团结构. 基于 2-mode 数据和 1-mode 网络数据实验表明, 该方法可以快速、可靠地找出网络的社团. 将经典数据分析的聚类方法应用到网络分析中, 验证了该思路的有效性, 为网络社团分析提供一个新视角.

关键词: 复杂网络; 社团划分; 聚类; 主分量分析

中图分类号: TP311; TP393; N94 **文献标识码:** A **文章编号:** 1673-4785(2011)01-0057-06

A clustering method for community detection on complex networks

LI Wei, YANG Xiaofeng, ZHANG Chongyang, TANG Kezong, YANG Jingyu
(Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: Community detection is important for understanding complex networks. Because of its high complexity, community detection in complex networks has recently attracted significant interest from research groups. In this work, a data analysis perspective was proposed for community detection on complex networks. First, based on the network topology, the nodes of the studied network were projected as data points in a high-dimensional space, and the network was associated with a data cloud. Second, principal component analysis (PCA) was used to reduce the high-dimensional data cloud into a low-dimensional one, which kept the main structural information. Third, K-means algorithms were used to find clusters of the data points in the reduced data cloud, which inferred the communities of the studied network. Experiments on datasets DGG (2-mode) and Zachary (1-mode) indicated that the proposed method can reveal network communities effectively. The proposed method provided a novel perspective of the community detection of complex networks.

Keywords: complex networks; community detection; cluster; PCA

近 10 年来, 伴随着互联网的普及, 计算技术的发展, 人们共享和处理大量数据的能力得到很大提高, 这使得需要大量现实数据支撑的复杂系统的实证研究成为可能. 研究者采用整体研究模式, 以探索现实系统的宏观性质为目标, 在多个学科领域取得了重要进展^[1-3]. 特别地, 通过忽略原系统中各个体自身细节, 将组成系统的个体抽象为网络节点, 即无论是细胞还是社会中的成员, 一律看作是无属性的节点, 再将它们之间的相互关系抽象成网络的边, 这

样原来多样的复杂系统就可以从一个通用的网络视角来研究, 称之为“复杂网络”研究^[1-6].

复杂网络方法被广泛地应用到各个研究领域, 比如社会学中人际交互网络、合作网络、商业网的研究, 信息技术领域的文献索引网、互联网、万维网研究, 生物学中的蛋白质作用网络、神经网络、捕食网等的研究^[4-6]. 通过这样一个独特的研究视角, 大量实证研究证实了复杂网络模型在自然界和人类社会中的普遍性和有效性, 而且来自于不同领域的复杂系统惊奇地呈现出一些相同的性质, 其中最具有代表性的成果有 Watts 小世界效应^[7]和 Barabasi 无标度特征^[8].

收稿日期: 2010-05-24.

基金项目: 国家自然科学基金资助项目 (60632050, 60873151).

通信作者: 杨静宇. E-mail: yangjy@mail.njust.edu.cn.

近来,复杂网络的另一个共同属性“社团结构”也引起了普遍的关注.人们发现在许多社会、生物网络中都存在着社团结构^[9-11],即整个网络由多个社团构成,这些社团的内部节点连接紧密,外部节点连接稀疏.研究表明这些社团常常与系统的功能性质有着很强的对应关系,如在人际交互网中,社团对应着某些社群,这些社群的内部成员具有相似的职业、政治倾向等社会属性^[9-11].同样,社团结构也反应在其他的社会网、信息网以及生态网中.社团结构的分析有助于研究者进一步探索网络的内部结构,对认识原系统的属性有重要意义,对现实的社会分析、系统优化以及商业决策起着指导性作用.

由于网络中社团结构分析的重要意义,近年来,社团结构划分算法研究受到广泛的关注^[9-11].然而,由于社团划分问题自身的复杂性,现有的方法往往只在某个领域或某些条件下表现较优.因而,网络中的社团划分问题仍然是研究人员面前的一个挑战.下面,首先介绍经典的社团划分方法;然后引入本文方法:将网络社团划分看作一个数据挖掘问题,首先将依据网络节点的影响力,将网络投影成高维空间的一个数据分布,再使用主分量降维,最后通过低维空间中的聚类结果来考察原网络的社团结构;文章最后讨论所提方法的优缺点及其改进.

1 网络社团划分经典算法

经典社团划分算法的思想很多来源于社会学中的层次聚类(hierarchical clustering)和计算机科学中的图分割(graph partition)^[12].这些算法大致可以分成:凝聚方法(agglomerative method)、分裂方法(divisive method)、搜索方法和其他方法.其中凝聚方法和分裂方法来源于社会学中寻找社团结构的层次聚类方法.Kernighan-Lin 算法^[13]和谱平分算法^[14]则是图分割方法的代表.

1.1 层次聚类算法

凝聚方法和分裂方法是依据节点间的相似性,通过向网络中逐渐添加边或是从网络中移除边,把网络自然地划分为各个社团.具体地,凝聚方法的思想是将初始的网络看成一个节点数为 n 而边数为 0 的空网络,首先计算出两两节点之间的相似性,然后依次向相似性最高的节点对之间添加边,当该过程停止时,这个网络的组成就认为是其原网络的社团划分,见图 1 从下往上.

对应地,分裂算法是直接原网络着手,首先计算出两两节点之间的相似性,然后删除相似性最低的节点对之间的边,重复这个过程,网络就逐渐被细

分成各个小部分,见图 1 自上而下.

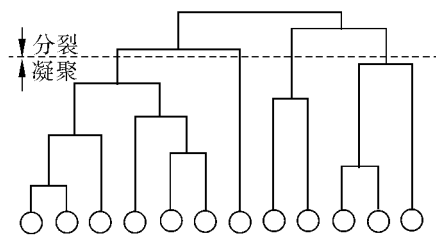


图1 层次聚类的凝聚、分裂方法

Fig.1 Agglomerative method and divisive method for hierarchical clustering

1.2 图分割方法

Kernighan 和 Lin 在 1970 年提出了针对图分割问题的 Kernighan-Lin 算法^[13].该方法首先将网络随机地分成 2 个社团,然后通过重复交换来自这 2 个社团的节点对,并在社团内部边数减去社团之间边数达到最大时停止.

谱平分法则是依据无向图 G 的 Laplace 矩阵(若 G 有 n 个节点,则其 Laplace 矩阵为 $n \times n$ 维对称矩阵 L .其对角元素等于点的节点度,若节点 i 与节点 j 连接,则 L_{ij} 值为 -1 ,否则为 0)的第二小特征值所对应的特征向量的元素的正、负符号将网络节点分成 2 类^[14].

考虑上述层次聚类 and 图分割算法,可以发现大多数算法具有 1 个共同的不足之处,即在不知网络确切的社团数目时,很难确定算法何时终止.针对该问题,Girvan 和 Newman 在 2004 年提出了网络社团化评价函数 Q ^[10].设网络被划分成 n 个社团,则 Q 值计算如下:

$$Q = \sum (e_{ii} - a_i^2) = T_{re} - \|e^2\|.$$

式中: e_{ij} 表示社团 i 与社团 j 顶点之间的边占网络所有边的比例; $a_i = \sum e_{ij}$ 表示与社团 i 中节点相连的边占网络所有边的比例; $T_{re} = \sum e_{ii}$ 则表示连接 n 个社团内部节点的边占网络所有边的比例; $\|e^2\|$ 为矩阵 e^2 的模,即 e^2 中元素的加总.物理意义上, Q 函数定义了社团内实际连接数目与随机连接情况下社团内连接数目之差,可以定量地刻画某种方法划分结果的社团化程度.在具体算法迭代过程中,可以对每一次社团划分结果计算 Q 值,当某次划分的 Q 值达到峰值时,则可认为此时社团划分最优.基于该思想,一批基于传统的层次划分和图分割算法,并通过最优化目标函数 Q 来实现复杂网络的社团结构划分的方法涌现出来^[11].

然而,由于现实网络的规模巨大、结构复杂,使得网络社团划分问题通常需要搜索非常广阔的解空间.在一般情况下,找到这类分割问题的精确解是一

个 NP-hard 问题^[15]. 许多实际算法存在着需要预设参数,选择社团尺度,或者每次只能二分网络等限制^[16],因而,网络中的社团划分问题仍然是一个挑战. 本文从一个新的角度考虑网络划分问题:首先将网络投影成高维空间的一个数据分布,接着使用主分量方法抽取主要分布结构,最后通过 K-means 算法聚类低维空间中的数据,进而反推原网络的社团结构.

2 网络社团划分的投影聚类算法

近来,数据分析方法被有效地应用在网络分析领域^[17-19],文献[19]首先将网络投影成高维空间的点群分布,然后使用数据分析的主分量分析(PCA)来抽取点群的主结构,并依据点群结构特征来反推原网络的结构属性. 结果表明,该方法可以重现复杂网络研究领域中 small-world 网、scale-free 网、Internet 的层次结构等经典结果. 在目前网络分析方法的研究尚未成熟时,通过将网络问题转化成一个传统的数据分析问题,并有效地利用数据分析领域中经典方法来解决网络问题,具有广泛的实际意义.

本文沿用上述思路,首先将网络节点投影成高维空间点,接着使用 PCA 对生成的点分布降维,最后采用 K-means 方法聚类低维空间点,再根据聚类结果来划分网络社团,方法原理见图 2.

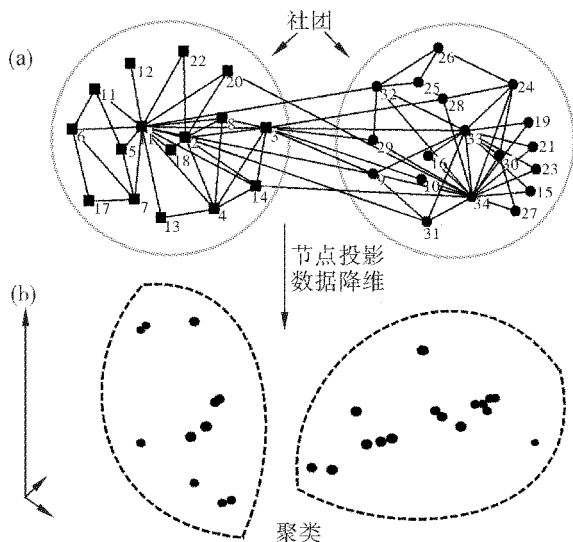


图2 网络社团结构即聚类划分算法

Fig. 2 Network community structure and illustration of clustering algorithm

2.1 节点投影

首先度量网络节点对网络的影响力,并依据节点的影响力度量将其投影到度量空间,具体步骤如下.

1) 随机选择 p 个网络节点作为度量基准点,计算当前度量节点 i 到 p 个基准点的图论距离 $d_{ij}, j =$

$1, 2, \dots, p$;

2) 当距离越大时,节点间的影响越小,计算距离的倒数 $1/d_{ij}$ 作为节点间影响力度量;

3) 节点 i 对 p 个基准点的影响力向量为 $f_i = (1/d_{i1}, \dots, 1/d_{ip})$, 当基准点选取的具有充分代表性时, f_i 可以表示节点 i 的网络影响力,向量 f_i 将节点 i 投影到高维度量空间.

2.2 PCA 降维

在得到所有 n 个网络节点的影响力向量后,构建 $p \times n$ 影响矩阵 $F = [f_1 f_2 \dots f_n]$. 再使用主分量分析 PCA 对矩阵 F 作降维,抽取主要结构信息.

1) 首先在 p 维测量空间 R^p 中计算 F 的标准化矩阵 X , X 的行均值为 0, 方差为 1.

2) 计算 X 的协方差矩阵 $C = XX^T$, 对 C 作奇异值分解:

$$C = XX^T = (U\Sigma V^T)(V\Sigma U^T) = UAU^T.$$

3) 选择 U 的前 q 列记作 U_q , 计算 $R = U_q X$, R 即为 X 的 q 维投影.

接着用经典的 K-means 算法^[20] 对网络低维投影点 R 作聚类分析,将同一类的点所对应的原网络顶点划分到一个社团.

3 实验

3.1 数据集

我们使用了 2 个数据集作为实验数据,分别是南方女士数据集 DGG 和空手道俱乐部数据集 Zachary.

南方女士数据集是美国 5 位民族学者于 20 世纪 30 年代在研究一个南部小镇阶层时收集的,根据几位收集者的姓名首字母简称为 DGG^[21]. DGG 数据记录了 9 个月期间,18 位女士(P1 ~ P18)参加 14 件非正式的社会活动(E1 ~ E14)的情况,是一个反应行动者与事件之间关系的 2-mode 数据集. DGG 数据集的网络表示见图 3,矩阵表示见表 1,其中 'X' 表示参加了该活动.

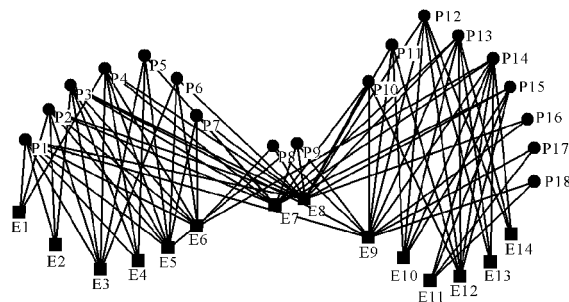


图3 南方女士数据集 DGG 的网络结构

Fig. 3 Illustration of DGG network

表1 南方女士数据集 DGG
Table 1 Participants-events of DGG

参与者	参与事件													
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
P1	X	X	X	X	X	X	—	X	X	—	—	—	—	—
P2	X	X	X	—	X	X	X	X	—	—	—	—	—	—
P3	—	X	X	X	X	X	X	X	X	—	—	—	—	—
P4	X	—	X	X	X	X	X	X	—	—	—	—	—	—
P5	—	—	X	X	X	—	X	—	—	—	—	—	—	—
P6	—	—	X	—	X	X	—	X	—	—	—	—	—	—
P7	—	—	—	—	X	X	X	X	—	—	—	—	—	—
P8	—	—	—	—	—	X	—	X	X	—	—	—	—	—
P9	—	—	—	—	X	—	X	X	X	—	—	—	—	—
P10	—	—	—	—	—	—	X	X	X	—	—	X	—	—
P11	—	—	—	—	—	—	—	X	X	X	—	X	—	—
P12	—	—	—	—	—	—	—	X	X	X	—	X	X	X
P13	—	—	—	—	—	—	X	X	X	X	—	X	X	X
P14	—	—	—	—	—	X	X	—	X	X	X	X	X	X
P15	—	—	—	—	—	—	X	X	—	X	X	X	—	—
P16	—	—	—	—	—	—	—	X	X	—	—	—	—	—
P17	—	—	—	—	—	—	—	—	X	—	X	—	—	—
P18	—	—	—	—	—	—	—	—	X	—	X	—	—	—

空手道俱乐部数据集是 Zachary 在 20 世纪 70 年代初,用了 2 年来观察美国某大学的空手道俱乐部成员间的人际关系,并依据这些成员平时的交往,建立的一个 1-mode 网络,它反应了成员之间的社交状况^[22],见图 2(a).空手道俱乐部网有 34 个点,代表俱乐部成员;78 条边,代表成员间的人际关系.在 Zachary 调查过程中,由于针对是否提高收费这一问题,俱乐部管理者(1 号顶点)与俱乐部教师(33 号顶点)之间产生了分歧并引发激烈的争论,最终导致俱乐部网络分裂成 2 个部分:其中方形顶点代表支持俱乐部管理者的成员,圆形顶点代表支持俱乐部教师的成员.

3.2 实验结果及讨论

DGG 和 Zachary 数据是具有代表性的社团结构,常常被使用作为示例来测试算法的效果,本文也在这 2 个数据集上作算法测试:

- 1) 计算各个网络节点的网络影响力,再根据节点的影响力度量,将节点投影成高维空间点;
- 2) 使用主分量分析 PCA 对生成的高维点降维,在具体实验中,保留前 2 维结构信息;
- 3) 计算点点间的距离(余弦),作 K-means 聚类,根据聚类结果来决定网络的社团结构.

3.2.1 DGG 数据实验结果

在 Freeman 的综述性文章^[23]中,列出了自 1940 年以来,超过 20 种方法在数据集 DGG 上作社团分

析的结果,见表 2.表 2 中数字 1(2)表示参与者被划分到第 1(2)社团中;数字 1/2 表示算法判断某个参与者分到第 1 或第 2 个社团均可;NA 表示算法对某个参与者的社团归属未能做出确切判断.本文结果在表 2 最后一行 L&Y10 给出.

考察表 2,讨论以下 3 点.

1) 绝大多数算法将参与者 P1 ~ P7 划到一个社团,本文方法得到相同结果.

2) 在表 2 中,一个值得注意的地方是,部分算法认为 P16、P17 和 P18 的社团不可确定,理由是这 3 个参与者参加活动少,因而仅仅由可知的信息量不能得到一个确定的判断.本文方法是一个强算法,对每个参与者都会给出一个硬性划分.而且我们认为根据 P16、P17、P18 已有的活动信息,将她们划入到社团二是合理的.

3) DGG 数据社团划分的难点在于判断参与者 P8 的归属,各种方法给出了不同的结论.表 1 中,P8 记录的社会活动有 E6、E8 和 E9,其中 E8 是大多数人都会参与的大众活动,因而由活动 E8 不能推断参与者 P8 的偏好,首先将 E8 排除不考虑.活动 E6 和 E9 则是具有代表性的社团活动,其特征为绝大多数参与者是同一个社团内成员,很少有另一个社团成员参加.表 2 中部分算法认为 P8 具有更强的社团一倾向,解释如下:比较 E6 和 E9,参与 E6 仅有一个社团外部成员 P14,而参与 E9 的有 3 个非社团成

员 P1、P3 和 P9,因而,E6 是比 E9 更典型的社团活动,则 P8 的“社团一”性质更强. 本文认为 P8 应归入社团二,依据如下:比较 E6 和 E9,E9 的社团代表性高过 E6;因为 E9 是社团二中最具代表性的活动,社团二中几乎人人参加,而社团一的最具代表性活

动是 E5,而非 E6,P8 参加了社团二最重要的活动 E9,未参与社团一最重要活动 E5,因而 P8 的“社团二”性质更强. 当然,关于 P8 的划分问题,关系到社会学定量等复杂问题,在此仅作简要讨论,进一步结论仍需进一步的研究.

表 2 DGG 社团划分比较

Table 2 The comparison of community detection on DGG

算法	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18
DGG41	1	1	1	1	1	1	1	1	1/2	2	2	2	2	2	2	2	2	2
HOM50	1	1	1	1	1	1	1	1/2	1	NA	NA	2	2	2	2	NA	2	2
P&C72	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
BGR74	1	1	1	1	1	1	1	NA	1	2	2	2	2	2/3	2/3	NA	2/3	2/3
BBA75	1	1	1	1	1	1	1	2	1	2	2	2	2	2	2	2	2	2
BCH78	1	1	1	1	1	1	NA	NA	NA	2	2	2	2	2	2	NA	NA	NA
DOR79	1	1	1	1	1	1	1	NA	1	2	2	2	2	2	2	NA	NA	NA
BCH91	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
FRE92	1	1	1	1	1	1	1	NA	1	2	2	2	2	2	2	2	NA	NA
E&B93	1	1	1	1	1	1	1	NA	1	2	2	2	2	2	2	NA	NA	NA
FR193	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
FR293	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2
FW193	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	1/2	2	2
FW293	1	1	1	1	1	1	1	NA	1	2	2	2	2	2	2	NA	2	2
BE197	1	1	1	1	1	1	1	NA	1	2	2	2	2	2	2	NA	NA	NA
BE297	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
BE397	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
S&F99	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	NA	2	2
ROB00	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
OSB00	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2
NEW01	1	1	1	1	1	1	1	2	1	2	2	2	2	2	2	2	2	2
L&Y10	1	1	1	1	1	1	1	2	1	2	2	2	2	2	2	2	2	2

3.2.2 Zachary 数据实验结果

Zachary 空手道俱乐部的实际社团结构见图 2(a),2 个社团分别以方形和圆形顶点标出. 本文方法得到 2 个社团,其中社团一中顶点为 1、2、3、4、5、6、7、8、11、12、13、14、17、18、20、22,社团二为 9、10、15、16、19、21、23、24、25、26、27、28、29、30、31、32、33、34,可以看出,它能够实际社团准确地划分出来^[10-12].

值得指出的是,DGG 和 Zachary 均是数据集,实验中取所有网络顶点作为度量基准点;另外,K-means 方法作聚类时需要类别数作为预设参数,本文依据社团 Q 函数峰值设定类别数均为 2.

4 结束语

复杂网络的社团结构往往反应了系统的功能与性质,社团结构的分析有助于研究者进一步探索网络的关键结构,进而对认识原系统的性质有着指导性意义.

本文针对复杂网络中的社团划分问题,提出将网络的社团划分问题转化为数据聚类问题,给网络中社团划分问题提供了一个新的解决视角. 通过将经典的 PCA、K-means 等数据分析方法引入复杂网络分析领域,为社团划分提供新的理论方法,并在实验中验证了经典方法在复杂网络社团划分问题中的有效性. 本文方法不仅可以用于 1-mode 网络的社团划分,而且可以直接处理 2-mode 网络,并且能得到优良的划分结果,不需要像传统方法那样,首先需要将 2-mode 网络转化成 1-mode 网络,再进行社团划分,这使得所提方法在处理网络社团划分问题时具有更高的通用性.

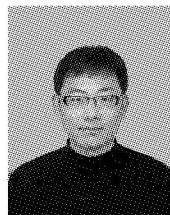
目前,本文方法还只是应用在 2 个示例数据集上,而更多的现实网络往往要复杂的多,比如边是有关值和方向的、网络规模又十分巨大等. 如何将方法推广应用到更多的现实网络是下一步的工作:一是将算法泛化到有权有向网络;二是在大规模数据上测试算法性能,进一步改进算法.

本文提出的方法具有发展成通用方法的潜力,通过将更多的经典数据方法引入到网络分析领域,可以为网络分析提供更多的理论工具,同时也为发现和解释更多的网络特性提供了可能.我们希望通过与数据分析领域的科研人员以及网络科学工作者的进一步讨论,将本文方法进一步完善和发展.

参考文献:

- [1] JASNY B R, ZAHN L M, MARSHALL E. Special online collection: complex systems and networks[EB/OL]. [2010-05-20]. <http://www.sciencemag.org/complexity/>.
- [2] DOROGOVISEV S N, GOLTSEV A V, MENDES J F F. Critical phenomena in complex networks[J]. Reviews of Modern Physics, 2008, 80(4): 1275-1335.
- [3] 汪秉宏, 周涛, 何大韧. 统计物理与复杂系统研究最近发展趋势分析[J]. 中国基础科学, 2005, 7(3): 37-43.
WANG Binghong, ZHOU Tao, HE Daren. The trend of recent research on statistical physics and complex systems[J]. China Basic Science, 2005, 7(3): 3743.
- [4] ALBERT R, BARABASI A L. Statistical mechanics of complex networks[J]. Reviews of Modern Physics, 2002, 74(1): 4797.
- [5] NEWMAN M E J. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167-256.
- [6] BOCCALETTI S, LATORA V, MORENO Y, et al. Complex networks: structure and dynamics[J]. Physics Reports, 2006, 424(4/5): 175-308.
- [7] WATIS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. Nature, 1998, 393(6638): 440-442.
- [8] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
- [9] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2001, 99(12): 7821-7826.
- [10] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2): 026113.
- [11] FORTUNATO S. Community detection in graphs[J]. Physics Reports, 2010, 486(3/4/5): 75-174.
- [12] 解伯, 汪小帆. 复杂网络中的社团结构分析算法研究综述[J]. 复杂系统与复杂性科学, 2005, 2(3): 1-12.
XIE Zhou, WANG Xiaofan. An overview of algorithms for analyzing community structure in complex networks[J]. Complex Systems and Complexity Science, 2005, 2(3): 1-12.
- [13] KERNIGHAN B W, LIN S. A efficient heuristic procedure for partitioning graphs[J]. Bell System Technical Journal, 1970, 49(2): 291-307.
- [14] FIEDLER M. Algebraic connectivity of graphs[J]. Czech Math Journal, 1973, 23(98): 298-305.
- [15] BRANDES U, DELING D, GAERTLER M, et al. Maximizing modularity is hard[EB/OL]. (2006-08-30) [2010-05-20]. <http://arxiv.org/abs/physics/0608255>.
- [16] FORTUNATO S, BARTHELEMY M. Resolution limit in community detection[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(1): 3641.
- [17] COSTA LD F, RODRIGUES F A, TRAVIESO G, et al. Characterization of complex networks: a survey of measurements[J]. Advances in Physics, 2007, 56(1): 167-242.
- [18] LI Wei, YANG Jingyu. Comparing networks from a data analysis perspective[J]. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 2009, 5: 1907-1916.
- [19] LI Wei, YANG Jingyu, HADDEN W C. Analyzing complex networks from a data analysis viewpoint[J]. Europhysics Letters, 2009, 88(6): 68007.
- [20] DUDA R O, HART P E, STORK D G. Pattern classification[M]. New York, USA: John Wiley & Sons, Inc., 2001: 114-121.
- [21] DAVIS A, GARDNER B B, GARDNER M R. Deep south[M]. Chicago: The University of Chicago Press, 1941: 147.
- [22] ZACHARY W W. An information flow model for conflict and fission in small groups[J]. Journal of Anthropological Research, 1977, 33: 452-473.
- [23] FREEMAN L. Dynamic social network modeling and analysis[M]. Washington, DC, USA: The National Academic Press, 2003: 39-97.

作者简介:



李伟,男,1978年生,博士.主要研究方向为复杂网络、模式识别、机器学习.



杨晓峰,男,1982年生,博士,主要研究方向为网络安全、人工智能.



杨静宇,男,1941年生,教授,博士生导师,教育部图像信息处理与智能控制重点实验室学术委员会委员,国际信息处理联合会(IFP)观察员.主要研究方向为模式识别、智能机器人、智能系统.曾获奖14项,其中国家级2项,省部级12项.发表学术论文300余篇,出版论(译)著7部.