

复杂网络中的社区发现——理论与应用

Community Finding in Complex Networks——Theory and Applications

王 林/WANG Lin^{1,2}, 戴冠中/DAI Guan-zhong¹

1.西北工业大学自动化学院,西安 710072; 2.西安理工大学自动化学院,西安 710048

1. School of Automation, Northwestern Polytechnical University, Xi'an 710072, China;

2. School of Automation, Xi'an University of Technology, Xi'an 710048, China

[摘要] 复杂网络是对于复杂系统的高度抽象,其中许多性质如小世界性质、无标度性质以及聚集性质等等已经得到了充分的研究。大量文献表明,复杂网络呈现出的社区结构(Community structure)特性,以及如何在大型网络中高效地发现社区(Community finding)问题是近年来复杂网络的研究热点。本文较为全面地综述了关于社区发现方面的概念、理论、算法及应用等,期望对于社区发现问题的进一步研究及若干基本问题的早日解决起到一定作用。

[关键词] 复杂网络,社区发现,边介数,模块度,层次聚类

[中图分类号] N94

[文献标识码] A

[文章编号] 1000-7857(2005)08-0062-05

Abstract : Complex networks are abstracted from complex systems, some properties of which such as small-world characteristics, scale-free characteristics and clustering are already thoroughly researched. As illustrated in lots of literatures, the existence of community structures in complex networks and how to find such communities effectively are focus of many recent researches in the branch of complex networks. This paper is intended as an overview of concepts, theories, algorithms and applications concerned with community finding.

Key Words: complex network, community finding, edge betweenness, modularity, hierarchical clustering

CLC Number : N94

Document Code: A

Article ID: 1000-7857(2005)08-0062-05

1 前言

复杂网络是复杂系统的抽象,网络中的节点是复杂系统中的个体,节点之间的边则是系统中个体之间按照某种规则而自然形成或人为构造的一种关系。现实世界中包含着各种类型的复杂网络,如社会网络^[1-4](朋友关系网络及合作网络等)、技术网络^[2,3,5-7](Internet、万维网以及电力网等)、生物网络^[2,8-10](神经网络、食物链网络以及新陈代谢网络等)。经过近几年的努力,复杂网络的研究取得了许多重要进展,发现了复杂网络的若干统计特征,其中包括小世界性质^[11,12](即网络中节点之间的平均距离很短,对数依赖于网络中的节点数)、无标度性质^[6,7,13-15](即网络中节点的度分布右偏斜,具备幂函数或指数函数的形式)以及聚集性(Clustering)或网络传递性^[2,16](Network Transitivity)。网络的聚集性是指同一个节点的两个相邻节点仍然是相邻节点的概率较大。

复杂网络的另外一个重要特征就是网络中所呈现出的社区结构。大量实证研究

表明,许多网络是异构的,即复杂网络不是一大批性质完全相同的节点随机地连接在一起的,而是许多类型的节点的组合。相同类型的节点之间存在较多的连接,而不同类型的节点之间的连接则相对较少(如图1所示)。我们把满足同一类型中的节点以及这些节点之间的边所构成的子图称为网络中的社区(Community)。

在大型复杂网络中自动搜寻或发现社区,具有重要的实用价值。如,社会网络中的社区代表根据兴趣或背景而形成的真实的社会团体;引文网络中的社区代表针对同一主题的相关论文;万维网中的社区就是讨论相关主题的若干网站^[9,10];而生物化学网络或者电子电路网络中的社区可以是某一类功能单元^[11,12]。发现这些网络中的社区有助于我们更加有效

地理解和开发这些网络。

与社区发现相关的理论包括图论^[17]以及模式识别^[18,19]等。复杂网络中社区发现的研究起源于社会学的研究工作^[20],Wu和Huberman^[21]以及Newman和Girvan^[22]的研究成果,使得复杂网络中的社区发现成为近几年复杂网络领域的一个研究热点并形成了复杂网络中一个重要的研究方向。

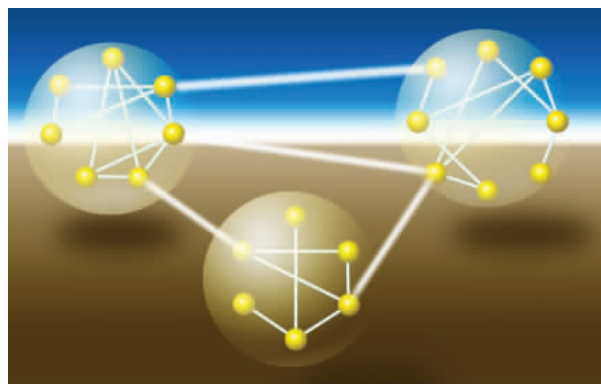


图1 具有社区结构的网络

收稿日期:2005-06-20

基金项目:国家“863”计划资助项目

作者简介:王 林,男,西安市金花南路5号西安理工大学自动化学院教授,西北工业大学在职博士生,主要研究方向为复杂系统及控制理论;E-mail: wanglin@xaut.edu.cn

戴冠中(通讯作者),男,俄罗斯宇航科学院外籍院士,西安市友谊西路127号西北工业大学自动化学院教授,主要研究方向为复杂系统、控制及网络信息安全;E-mail: daigz@nwpu.edu.cn

Newman 和 Girvan 把社区发现问题定义为将网络节点划分成若干组,使得组内节点之间的连接比较稠密,而不同组节点之间的连接则比较稀少。他们提出了基于边介数(Edge Betweenness)概念的分割方法,尽管该方法计算量很大,但由于其性能优越,从而成为社区发现研究的重要参考模型。其它一些方法涉及到很多概念,如电路理论^[21]、超顺磁聚类^[23]、网络的三角环(Triangular Loop)^[24]、网络的谱性质分析^[25]、网络的 Laplace 矩阵以及聚类技术^[26]等等。

尽管复杂网络的社区发现问题得到了大量的研究,但还存在一些尚未解决的基本问题,如社区概念虽然大量使用,但却缺少严格的数学定义;大多数社区发现算法虽然性能优越,但所需计算量却很大。这说明复杂网络中社区发现的研究还需要付出大量的努力。

本文按照历史发展的线索综述社区发现问题的若干研究成果,并作适当的性能比较和优缺点分析。文中所涉及的网络均假定是无向、无权值、无自环的简单图。

2 迭代二分法 (Iterative Bisection)

计算机科学中的一个典型问题,是将一个网络分解成若干节点数基本相等的子网,使得不同子网中的节点之间的连接数最少,称为图分割(Graph Partitioning)。图分割问题(GPP)可应用于并行计算机的处理器合理分配进程。实际应用中,大多数图分割方法均基于迭代二分法:首先将图按要求分割成 2 个子图,然后再分别对这 2 个子图进行分割,如此迭代下去直至获得所要求的子图个数。

我们介绍 2 个最具代表性的迭代二分法:一是基于图的 Laplace 矩阵的特征向量的谱二分法(Spectral bisection)^[27,28];另一个是 Kernighan-Lin 算法^[29],该算法使用贪婪(Greedy)算法对社区内以及社区间的边数进行优化,从而达到改进网络的初始分解的目的。

2.1 谱二分法

设 G 是一个具有 n 个节点的无向图, G 的 Laplace 矩阵 L 是一个 $n \times n$ 的对称矩阵,其对角线元素 L_{ii} 是节点 i 的度,表示节点 i 和节点 j 的连接情况,当 i 和 j 之间有边相连时, $L_{ij}=-1$,否则 $L_{ij}=0$ 。显然, L 每一行的和以及每一列的和均为 0,因而,向量 $1=(1, 1, L, 1)$ 是 L 的相应于特征值 0 的特征向量。

如果网络可以被分解成 g 个互不重叠、互不连接的子图(即图的分支) G_k ,则其 Laplace 矩阵 L 是一个块对角矩阵,每个对角块就是相应的分支子图的 Laplace 矩阵。容易知道,此时 L 存在 g 个与特征值 0 对应的特征向量 $v^{(k)}, k=1, 2, L, g$, 其中, $v^{(k)}$ 的对应于分支 G_k 的分量为 1,而其余分量为

0。

如果网络可以被分解成若干子图,但子图之间存在少量的连接,此时情形与上面有所区别。一般来说,此时 L 只存在 1 个特征值为 0 的特征向量,但存在 $g-1$ 个特征值接近于 0,相应的特征向量可以近似地看成上述 $v^{(k)}$ 的线性组合。因此,只要找到 Laplace 矩阵的几个接近于 0 的特征值,然后求对应的特征向量的线性组合,我们就可以近似地得到这些子图。

我们考虑一个特例,即将网络分割成 2 个子图。由于对称矩阵的任意 2 个特征值所对应的特征向量相互正交,故 Laplace 矩阵 L 的任意对应于非零特征值的特征向量均正交于向量 $1=(1, 1, L, 1)$,从而所有非零特征值的特征向量必须具有正分量和负分量。若 2 个子图间连接很少,则将存在一个特征向量,其特征值近似于 0;该特征向量的正分量对应一个子图,负分量对应另一个子图。因此,我们可以通过观察第二个最小的特征值所对应的特征向量,根据特征值元素的正负将一个网络分解成 2 个社区,这就是谱二分法。该第二个特征值 λ_2 称为图的代数连接度,它可度量图的分割是否合理,该值越小越合理。

作为谱二分法的一个应用,我们考虑社会学网络中的一个著名网络:Karate club 网络^[30]。该网络是 Zachary 对一所美国大学的 kKarate club 的成员中的朋友关系进行历时 2 年的观察而获得的。研究期间,由于一次关于时尚的争论,该网络自然地分割成两半。该网络的代数连接度 λ_2 为 0.469,与 0 有一定差异;但如图 2 所示,应用谱二分法所得结果与实际情况仍非常吻合,只有节点 3 判断失误。

谱二分法的计算速度较快。通常计算一个 $n \times n$ 矩阵的特征向量需要 $O(n^3)$ 次运算,但对于大多数复杂网络,Laplace 矩阵是稀疏矩阵,可以用 Lanczos 方法快速计算出最小的几个特征值所对应的特征向量,而计算第二特征向量所需运行时间大约为

$m/(\lambda_3-\lambda_2)$,其中 m 是图的边数。可以看出,如果图可以清晰地分解成 2 个社区,收敛速度较快。

谱二分法的主要缺点是只能将图分成 2 个子图,若要分成多个子图,则只能多次应用该方法。即使这样可行,我们预先也无法确定究竟将图分割成多少个子图才合适。

2.2 Kernighan-Lin 算法

Kernighan 和 Lin 提出了一个完全不同的图分割方法,该方法实际应用情况良好,最差运行时间为 $O(n^2)$,其中 n 是网络中的节点数。

首先定义一个效益函数 Q :位于子图内部的边数之和减去子图之间的边数。该算法要求必须预先指定 2 个子图的大小。算法分成以下 2 步。

1) 考察所有的节点对(节点对中的 2 个节点分别取自 2 个子图),交换节点对中 2 个节点的位置时,计算效益函数 Q 所产生的变化 ΔQ 。选择使所产生的变化 ΔQ 最大的那对节点进行实际交换,如此重复下去,并且要求已经交换过的节点不允许再交换。当一个子图中的所有节点均已交换过 1 次时,第一步完成。

2) 检查第一步中所进行的每一次交换,寻找使得效益函数 Q 最大的那一次交换,此即为所求得的图的分割。

Kernighan-Lin 算法的主要缺点是必须预先指定 2 个社区的大小,该算法应用于 Karate club 网络所得结果与实际情况一点不差,这一点显然优于谱二分法。但前提条件是必须指定 2 个子图的大小,否则会得到错误的结果。这使得 Kernighan-Lin 算法无法应用于大多数真实网络。即使 Kernighan-Lin 算法的这一缺点得以克服,作为图分割方法,其先天性不足仍然难以解决。

3 层次聚类法 (Hierarchical Clustering)

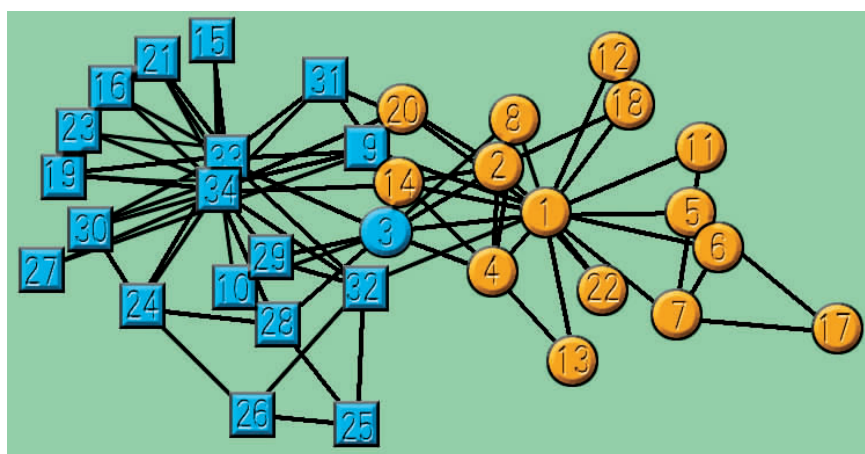


图 2 Karate club 网络的谱二分法分解

社会学家研究社会网络时所采用的社区发现的主要方法是层次聚类^[31],它是对每一对节点定义一个基于网络结构的相似度 x_{ij} ,然后从一个具有 n 个节点、但没有边的空网出发,按照相似度由大到小的顺序,1次在1对节点之间加上1条边。需要指出的是,所加上的边与网络中原来的结构无关,原来的网络结构只影响到节点对的相似度。

3.1 单连接法

从网络中提取社区结构的方法之一是单连接法(Single Linkage),该方法把在逐次加边过程中所形成的图的连通分支作为社区。若最近所加入边的相似度是 x ,那么显然就有相似度大于或等于 x 的2个节点一定属于同一个社区。当然同一个社区中可能存在相似度小于 x 的节点对。随着 x 越来越小,边越多,一些小的连通分支逐步合并成大的连通分支,结果连通分支数越来越少,开始具有 n 个社区,最后则形成1个大社区。每一步中的社区一定是包含在下一步的某一个社区中,整个算法进行的过程可以表示成1个树形结构(见图3)。

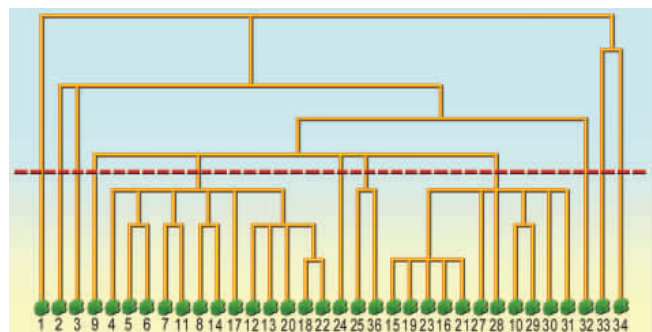


图3 算法的树形结构

图中,底端的树叶表示算法开始时的单个节点,而顶部的树根则表示所有节点都进入了1个社区。在树的不同高度所进行的水平切割,表示算法进行到相应步数后所得到的社区。与二分法相同,层次聚类法也无法确定应该将图分解成多少个社区。

3.2 完全连接法

另一个确定社区的方法是完全连接法(Complete Linkage)。单连接法将算法进行到某一步时所得到的分支作为社区,而完全连接法则是将算法进行到某一步时,网络中的极大节点集团(Maximal Clique)称为社区。所谓节点集团实际上是一个完全子图,所谓极大节点集团是指不存在包含该集团的更大的集团。

虽然完全连接法具有更好的特性,但基于下列原因很少被使用。①在图中寻找节点集团非常艰难,可供选择的算法是Bron-Kerbosch算法,其最差运行时间与节点数呈指数关系;②一般来说节点集团不唯一,一个节点可能会同时属于多个集团。

有多种方式可以定义节点间的相似度。社会学研究中倾向于基于结构等价性(Structural Equivalence)定义相似度。如果2个节点具有完全相同的相邻节点,则称这2个节点为结构等价。例如,在朋友关系网络中,如果两人拥有完全相同的朋友,则这两人就是结构等价的。

由于真实网络中的结构等价很少见,人们可以定义一个等价程度的度量指标。其中一个指标称为Euclid距离^[1,32],定义为:

$$x_{ij} = \sqrt{\sum_{k \neq i,j} (A_{ik} - A_{jk})^2} \quad (1)$$

式中 A_{ij} 是图的邻接矩阵的元素。对于结构等价的节点对, Euclid距离为0;而对于没有任何共同相邻节点的节点对, Euclid距离取得最大值。因此,根据该度量指标进行的层次聚类应该按照 x_{ij} 增加的顺序在网络中逐步加边。

另一个常用的相似性度量指标是邻接矩阵的行或列之间的相关性^[33]。首先给出邻接矩阵行向量的均值和方差如下:

$$\mu_i = \frac{1}{n} \sum_j A_{ij}, \quad \sigma_i^2 = \frac{1}{n} \sum_j (A_{ij} - \mu_i)^2 \quad (2)$$

i 行和 j 行的相关系数为:

$$x_{ij} = \frac{\frac{1}{n} \sum_k (A_{ik} - \mu_i)(A_{jk} - \mu_j)}{\sigma_i \sigma_j} \quad (3)$$

结构等价程度高的节点对必然对应较高的 x_{ij} 。

另外还有一个相似度定义为节点之间的边独立的路径数。所谓边独立是指两条路径不包含相同的边。显然,该相似度并非基于结构等价性。根据最大流-最小割定理(max-flow/min-cut),这种路径数等于网络中相应两节点间可以通过的最大流量。计算路径数所需时间为 $O(m)$,其中 m 是图中的边数。使用独立路径数作为相似度而形成的完全连接社区中,任意两节点之间至少存在 k 条边独立路径,其中 k 是最近加入的边的相似度。图论文献中称这种社区为 k -分支。已经有若干专门用于确定 k -分支的算法,其中最著名的是Hopcroft和Tarjan的2-分支和3-分支算法^[33,34]。网络的 k -分支常常用于社会网络的分析中。

作为层次聚类法的一个例子,我们将Euclid距离单连接聚类应用于Karate Club网络,并得到如图3所示的树形图。该方法获得了2个子图的绝大多数成员,但也有些成员不归属任何一个子图。这是由层次聚类方法的特点所决定的。该方法善于发现社区的一部分,发现那些根据所选择的相似度定义的高相似的节点对,而处于社区边界的节点常常因为相似度小而无法确定其归属。

层次聚类方法计算速度分析:相似度计算通常需要 $O(mn)$ 次运算,相似度排序所需时间为 $O(n^2 \log n)$ 。树形图的构造可以通过Fischer的union/find算法^[35,36]实现,所需时间基本上线性依赖于 n 。因此,对于1个稀疏图,整个算法的运行时间为 $O(n^2 \log n)$ 。

层次聚类方法无需事先指定子图的大小,但仍然无法确定网络最终应该分解成多少个社区,即便不是这样,许多节点无法确定归属的问题,也使得层次聚类方法在分析许多大型真实网络时无法获得满意的结果。

4 G-N 算法

层次聚类方法实际上是一种基于加边的方法,正因为如此,应用层次聚类算法将导致若干节点无法确定归属。下面介绍的一些算法是基于去边的方法。

对于一般网络发现算法来说,其基本要求是可发现网络的最自然的分割,无需事先指定要分解成多少个社区以及社区的大小,同时不存在如层次聚类算法那样的缺陷。Girvan和Newman^[37]所提出的G-N算法部分地达到上述要求,该方法的思路如下。

社区之间所存在的少数几个连接应该是社区间通信的瓶颈,是社区间通信时通信流量必经之路。如果我们考虑网络中某种形式的通信并且寻找到具有最高通信流量的边,该边就应该是连接不同社区的通道。若将这样的边去除就应该获得了网络的最自然的分解。

作为通信流量的度量,Girvan和Newman引入了边介数的概念,这是Freeman的点介数^[38](Vertex betweenness)概念在边上的推广。一条边的边介数定义为所有节点对之间的最短路径中经过该边的路径数。在一个具有 m 条边和 n 个节点的图中计算出每条边的介数所需时间为 $O(mn)$ ^[39,40]。

G-N算法包括计算网络中每条边的介数、去除边介数最大的那条边、重复进行直至网络中没有任何边存在。将该算法应用于Karate club网络,所得结果与谱二分法所得结果完全相同。然

而,对于一般的网络分析,该方法要比谱二分法优越得多。正如层次聚类法一样,G-N 算法可以将网络分裂成任意数量的社区,可以从算法的树状图看出网络社区结构形成的动态过程。

G-N 算法存在下列 2 点不足:① 该算法也无法预知网络最终应该分裂成多少社区;② 计算速度缓慢,最差运行时间为 $O(m^2n)$,其中 m, n 分别为网络中的边数和节点数。

Tyler 等^[41]在研究电子邮件网络时,对 G-N 算法进行了修改,大大提高了其计算速度,但精度有所损失。

G-N 算法计算边介数时,首先计算以给定节点作为起点的所有路径对于给定边的边介数的贡献,然后将所有节点的贡献相加,得到给定边的边介数,因此计算所有边的边介数所需计算时间为 $O(m^2n)$ 。Tyler 等认为无须计算所有节点的贡献,而只要计算部分节点的贡献即可。如果随机抽取样本节点,便可以得到对于边介数的 Monte Carlo 估计值。虽然 Monte Carlo 估计会产生统计误差,但 Tyler 等论证了只需较少的样本节点即可获得理想的计算结果,从而大大减少了计算时间。

5 Radicchi 算法

该算法与 G-N 算法相同,都是基于去边,但不是根据边介数选择要去除的边,而是引进了边聚集系数 (Edge Clustering Coefficient)^[24]的新指标。整个算法的运行时间为 $O(m^4/n^2)$ 。显然,对于稀疏图,其计算速度要比 G-N 算法快一个数量级。

Radicchi 等考虑网络中的三角环 (Triangular loop, 即边数为 3 的闭合路径)。若 1 个三角环包含一条连接不同社区的边,则该三角环中的另 2 条边中的某一条仍然连接这 2 个社区的可能性将很大。但是由于连接不同社区的边非常稀少,故包含 1 条给定的连接不同社区的边的三角环不可能很多。因此,将 1 条边的边聚集系数定义为包含该边的三角环所占比例:

$$C_{ij} = \frac{z_{ij} + 1}{\min(k_i - 1, k_j - 1)} \quad (4)$$

其中 k_i, k_j 分别表示节点 i 和 j 的度, z_{ij} 表示网络中实际包含该边的三角环的个数。上式中的分母表示包含该边的最大可能的三角环的个数。

Radicchi 算法每一步去除的是网络中边聚集系数最小的边,每次去除后,再重新计算每一条边的边聚集系数,如此进行下去,直至网络中不存在任何边。

Radicchi 算法的不足是该算法依赖于网络中的三角环,如果网络中三角环很少,那么该算法将失去意义。实证研究表明,社会网络中三角环的数量比较大,而在非社会网络中,三角环的数量则相对较少。这意味着 Radicchi 算法更加适合于社会网络。

6 W-H 算法

Wu 和 Huberman^[21]基于电阻网络的性质提出了 W-H 算法,其主要思路如下。

将网络中的每条边想象成电阻为单位值的导线,且在网络中任意选择的 2 个节点上加上单位值的电位差,通过求解 Kirchhoff 方程,可以得到每个节点处的电位值。将计算结果表示成图 4,其中横坐标标识的是每个节点的电位值,图 4 表示网络的电位谱图。Wu 和 Huberman 认为,如果网络可以分解成 2 个社区,并且所选择的 2 个节点分属不同社区,那么电位谱在连接 2 个社区的边的两端会产生一个较大的间隙。因此,我们首先确定电位谱的最大间隙处的某个电位值,然后根据每个节点处的电位是否大于该值而确定节点属于哪个社区。由于计算每个节点处的电位值需要求解 Laplace 矩阵的逆矩阵,所需计算量通常为 $O(n^3)$,显然速度太慢。为此,Wu 和 Huberman 采用了下列近似方法。

依次更新每个节点处的电位值为其相邻节点的电位值的平

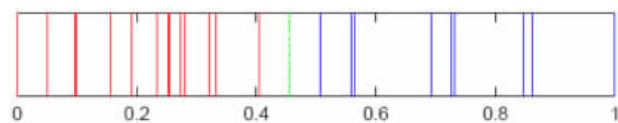


图 4 网络电位谱图

均值,如此进行多次,则将得到每个节点处的电位的近似值,并且运行次数与网络的大小无关。这说明计算出每个节点处的电位值的给定精度所需运行时间为 $O(m+n)$ 。

W-H 算法的一个重要特点是可以用来确定包含指定节点的社区,无须计算出所有的社区,该特点可以应用于一些特定场合,如 Web 搜索引擎。

W-H 算法的不足是如果预先不知道关于网络社区结构的部分信息,则很难应用该算法确定社区结构。

7 网络分解评价

前面所介绍的所有算法均无法确定网络究竟分解成为多少个社区比较合适,为此,Newman 和 Girvan^[22]定义了一个用以评价网络分解满意度的指标,称为模块度 (Modularity)。模块度的定义为:假定网络已经被分裂成 g 个社区,定义一个 $g \times g$ 的矩阵 e ,其元素 e_{ij} 表示原网络中连接社区 i 和社区 j 中的节点的边数所占比例。该矩阵的迹 Tre 表示网络中连接同一社区中节点的边所占比例,该矩阵行元素的和 $a_i = \sum_j e_{ij}$ 表示连接社区 i 中的节点的边所占的比例 (该数值只依赖于网络中节点的度值,不依赖于其它网络特征,如社区结构)。如果将网络看成一个节点度值给定前提下所生成的随机网络,那么连接社区 i 和社区 j 中的节点的边数所占比例应该为 $a_i a_j$,特别地,连接社区 i 中的节点的边所占比例应该为 a_i^2 。因此,若将网络看成一个节点度值给定的随机网络,则连接同一社区中节点的边所占比例为:

$$\sum_i a_i^2 = \sum_i \sum_{jk} e_{ij} e_{ik} = \|e^2\| \quad (5)$$

其中, $\|x\|$ 表示 x 的所有元素之和。于是我们可以将网络的模块度定义为:

$$Q = \text{Tre} - \|e^2\| \quad (6)$$

应该指出,模块度的定义独立于任何特定的社区发现算法,只是用来衡量网络分解是否合理。

类似于 Kernighan-Lin 算法,可以利用网络的模块度来设计出网络分解的贪婪算法^[42]。

8 结论

本文比较全面地综述了关于复杂网络中的社区结构的性质以及发现这样的社区结构的若干算法,并对这些算法的适用范围、计算复杂度以及准确性等方面进行了比较。期望通过我们的努力能够对我国在社区发现方面的研究工作起到一定的推动作用。

参考文献 (References)

- [1] S. Wasserman, K. Faust. Social Network Analysis[M]. Cambridge University Press, Cambridge, 1994
- [2] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393: 440~442
- [3] Amaral L A N, Scala A, Barth M, e_l_emy, and Stanley H E. Classes of small-world networks [M]. *Proc. Natl. Acad. Sci. USA* 97, 11 149~11 152 (2000)
- [4] Newman M E J. The structure of scientic collaboration networks [M]. *Proc. Natl. Acad. Sci. USA* 98, 404~409 (2001)
- [5] Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the internet topology [J]. *Computer Communications Re-*

- view, 1999,29: 251~262
- [6] Albert R, Jeong H, Barabasi A L. Diameter of the world-wide web [J]. *Nature*, 1999, 401:130~131
- [7] Broder A, Kumar R, Maghoul F, et al. Graph structure in the web [J]. *Computer Networks*, 2000,(33): 309~320
- [8] Williams R J, Martinez N D. Simple rules yield complex food webs[J]. *Nature*, 2000, 404: 180~183
- [9] Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic networks[J]. *Nature*, 2000,407: 651~654
- [10] Fell D A, Wagner A. The small world of metabolism [J]. *Nature (Biotechnology)*, 2000, (18): 1 121~1 122
- [11] Pool I, Kochen M. Contacts and inuence [J]. *Social Networks*, 1978, (1):1~48
- [12] Milgram S. The small world problem[J]. *Psychology Today*, 1967, (2): 60~67
- [13] Barabasi A L, Albert R. Emergence of scaling in random networks [J]. *Science*, 1999, 286: 509~512
- [14] Krapivsky P L, Redner S, Leyvraz F. Connectivity of growing random networks [J]. *Phys. Rev. Lett.*, 2000, (85): 4 629~4 632
- [15] Dorogovtsev S N, Mendes J F F, Samukhin A N. Structure of growing networks with preferential linking [J]. *Phys. Rev. Lett.*, 2000, (85):4 633~4 636
- [16] Newman M E J, Strogatz S H, Watts D J. Random graphs with arbitrary degree distributions and their applications [J]. *Phys. Rev. E*, 2001, 64: 026 118
- [17] West D B. Introduction to Graph Theory [M]. Prentice Hall, Upper Saddle River, 2001
- [18] Duda R O, Hart P E, Stork D G. Pattern Classification[M]. Wiley-Interscience, New York, 2001
- [19] L. da F. Costa, R. M. C. Jr.Shape Analysis and Classification: Theory and Practice [M]. CRC Press, Boca Raton, 2001
- [20] Scott J. Social Network Analysis: A Handbook[M]. Sage Publications, London, 2000
- [21] Wu F, Huberman B A. Finding communities in linear time: A physics approach[J]. *Euro. Phys. J B*, 2003,38: 331~338
- [22] Newman M E J, Girvan M. Finding and evaluating community structure in networks [J]. *Physical Review E*,2004,69: 026 113
- [23] Reichardt J, Bornholdt S. Detecting fuzzy community structures in complex networks with a Potts model. cond -mat/0402349, 2004
- [24] Radicchi F, Castellano C, Cecconi F, et al. Defining and identifying communities in networks. cond-mat/0309488. (2004)
- [25] Capocci A, Servedio V D P, Caldarelli G, et al. Detecting communities in large networks. cond-mat/0402499. (2004)
- [26] Donetti L, Munoz M A. Detecting Network Communities: a new systematic and efficient algorithm. cond -mat/0404652. (2004)
- [27] Fiedler M. Algebraic connectivity of graphs [J]. *Czech Math J*. 1973, 23: 298~305
- [28] Pothen A, Simon H, Liou K-P.Parti
- [29] Kernighan B W, Lin S. An efficient heuristic procedure for partitioning graphs [J]. *Bell System Technical Journal*,1970,49: 291~307
- [30] Zachary W W. An information flow model for conflict and fission in small groups[J]. *Journal of Anthropological Research*, 1977, 33: 452~473
- [31] Scott J. Social Network Analysis: A Handbook[M]. Sage, London, 2nd edition, 2000
- [32] Burt R S. Positions in networks[J]. *Social Forces*, 1976, 55: 93~122
- [33] Tarjan R E. Depth-first search and linear graph algorithms [J]. *SIAM J Comput*. 1972, (1): 146~160
- [34] Hopcroft J E, Tarjan R E. Dividing a graph into tri-connected components[J]. *SIAM J Comput*. 1973, (2): 135~158
- [35] Fischer M J.Efficiency of equivalence algorithms [M]. In Miller R E and Thatcher J W. (eds.), *Complexity of Computer Computations*. Plunum Press, New York,1972, 153~167
- [36] Tarjan R E. Efficiency of a good but not linear set union algorithm [J]. *J ACM*,1975, 22: 215~225
- [37] Girvan M, Newman M E J. Community structure in social and biological networks [J]. *Proc. Natl Acad*. 2002, 99: 7 821 ~ 7 826
- [38] Freeman L C. A set of measures of centrality based upon betweenness [J]. *Sociometry*, 1977, 40: 35~41
- [39] Newman M E J. Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality [J]. *Phys Rev E* 2001,64: 016 132
- [40] Brandes U. A faster algorithm for betweenness centrality[J]. *Journal of Mathematical Sociology*, 2001, 25: 163~177
- [41] Tyler J R, Wilkinson D M, Huberman B A. Email as spectroscopy: Automated discovery of community structure within organizations [C]. In Huysman M, Wenger E, Wulf V. (eds.), *Proceedings of the First International Conference on Communities and Technologies*, Kluwer, Dordrecht (2003)
- [42] Newman M E J. Fast algorithm for detecting community structure in networks[J]. *Phys Rev E*, 2004, 69: 066 133

(责任编辑 李慧政)

(制 图 严佳君)

·图片科技新闻·

我国多种形式纪念郑和下西洋 600 周年



图 1 郑和画像

2005 年 7 月 11 日,是我国伟大的航海家、探险家郑和下西洋 600 周年纪念日,郑和下西洋 600 周年纪念大会在北京人民大会堂隆重举行。600 年前,郑和率领由 208 艘船组成的庞大船队,完成了 7 次下西洋的创举,途经 30 多个国家和地区,成为世界航海史上时间最早、规定最大、技术先进、活动范围最广的洲际航海活动,传播了中华文明,促进了经济文化交流和经贸往来,扩大了中外友好关系,为世界航海事业和人类文明进步做出了巨大贡献。

为纪念郑和下西洋 600 周年,国家决定把每年的 7 月 11

日定为中国“航海日”。各地举行了各种形式的纪念活动。6 月 28 日,世界上航程最远的波音 777-200LR 在北京首都国际机场亮相,并被命名为“郑和号”。同一天,中国内地邮政和港、澳特别行政区邮政联合发行“纪念郑和下西洋 600 周年”同一主题邮票,并推出“郑和下西洋 600 周年”系列纪念邮品。7 月 11 日,“郑和下西洋铜像”在福建泉州落成。交通部组织开展了以“热爱祖国、睦邻友好、科学航海”为主题的首届“航海日”系列活动。全国有关省市和学术团体纷纷举行纪念大会和学术研讨会,纪念郑和下西洋 600 周年。



图 2 “郑和宝船”仿真模型



图 3 郑和下西洋海航图



图 4 被命名为“郑和号”的波音飞机



图 5 “纪念郑和下西洋 600 周年”纪念邮票