

# 复杂网络集团特征研究综述<sup>\*</sup>

张光卫<sup>1</sup> 康建初<sup>1</sup> 夏传良<sup>2</sup> 李鹤松<sup>1</sup>

(北京航空航天大学软件开发环境国家重点实验室 北京 100083)<sup>1</sup>

(中国科学院数学与系统科学研究院计算机科学研究室 北京 100080)<sup>2</sup>

**摘要** 自然界和人类社会的许多系统可以用复杂网络进行建模,复杂网络已成为多个学科的研究热点。分析复杂网络的一个关键问题是如何理解其全局组织,网络的健壮性和稳定性在很大程度上取决于其集团结构特征。本文简要介绍了复杂网络的基本概念并详细总结了近年复杂网络集团特性的研究进展,重点分析了社区发现算法的最新研究成果,最后提出这一领域几个有待解决的问题和可能的发展方向。

**关键词** 复杂网络,社区,幂率分布,小世界,无尺度

## Discovering Communities in Complex Networks

ZHANG Guang-Wei<sup>1</sup> KANG Jian-Chu<sup>1</sup> XIA Chuan-Liang<sup>2</sup> LI He-Song<sup>1</sup>

(State key Lab. of Software Development Environment, Beijing University of Aeronautics and Astronautics, Beijing 100083)<sup>1</sup>

(Department of Computer Science, Academy of Mathematics and System Sciences, The Chinese Academy of Sciences, Beijing 100080)<sup>2</sup>

**Abstract** Complex networks have been studied extensively due to their relevance to many real systems as diverse as the Internet, biological and social networks. A key question is how to interpret the global organization of such networks as the coexistences of their structural subunits(communities). There has been considerable recent interest in algorithms for finding communities in networks groups of vertices within which connections are dense but between which connections are sparser. The present article reviews some basic concepts, important progress, and significant results in the current studies of various complex networks, with emphasis on the concepts and algorithms for discovering communities.

**Keywords** Complex network, Community, Power-law, Small world, Scale free

## 1 引言

“结构决定功能”是系统科学的基本观点<sup>[1]</sup>。如果将复杂系统内部的各个元素抽象为结点,元素之间关系视为连接,那么就构成了一个具有复杂连接关系的网络,通常称为复杂网络,该网络反映了系统的拓扑结构。自1999年Barabási和Albert发现复杂网络拓扑结构的无尺度性质以来<sup>[2]</sup>,众多物理学家和数学家的研究表明,从各种现实网络(食物链网络<sup>[3,4]</sup>、新陈代谢网<sup>[5~8]</sup>、神经细胞相互连接形成的神经网络<sup>[9]</sup>、WWW网<sup>[10,11]</sup>、Internet<sup>[12]</sup>、科学家合作网<sup>[13,14]</sup>、社会关系网<sup>[15,16]</sup>等)中抽象出来的复杂网络有3个基本统计特征:一是小世界(Small World)<sup>[9]</sup>特征,即网络中结点对之间的平均距离较短,这里距离的定义为连接两点的最短路径所经过的边的数目。大量的实验研究表明,真实网络几乎都具有小世界效应。二是无尺度(Scale Free)<sup>[17]</sup>特征,即网络结点的度服从幂率(Power-law)分布,结点的度是指与该结点关联的边的数目。三是集聚(Clustering)特征,集聚特征是由结点的集聚系数(Clustering Coefficient)<sup>[9,17,38]</sup>衡量的。2002年,Newman提出了复杂网络的社区结构(Community Structure)<sup>[19]</sup>概念,社区(Community)就是网络中结点的集合。社区中的结点之间具有紧密的连接,而社区之间则为松散的连接,这样的集合也被称为簇、内聚组、模块(clusters, cohesive groups, modules)<sup>[16,19~21]</sup>,图1为一个社区的示意图。虽然社区没有

一个被广泛接受的、统一的、量化的定义,但它确实是复杂系统层次和模块结构的标志<sup>[8,22]</sup>。

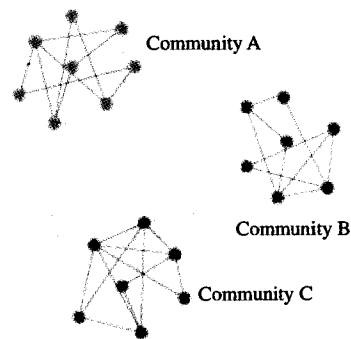


图1 社区结构示意图

图中有3个社区,社区内部结点的联系紧密而社区之间的联系松散。

目前人们除了研究复杂网络的统计特征和动力学特征外,越来越重视社区结构的研究<sup>[19,23,26,39,44,45,47~52]</sup>,力图揭示看上去错综复杂的网络是怎样由相对独立又互相交错的社区组成的。这对于理解网络的结构和功能特性有深刻的理论和现实意义;相对独立的基本模块可能是功能相关的蛋白质组<sup>[22,23]</sup>,也可能是相互协作的工业社区<sup>[24]</sup>或相互关系密切的人群组<sup>[26]</sup>,根据结点所处的社区便能够推断出其性质;另外,网络社区结构对网络的抗毁性、健壮性和稳定性<sup>[54]</sup>,对传染

<sup>\*</sup>国家自然科学基金(60496323,60375016)、现代设计大型应用软件的共性基础973计划(2004CB719401)项目。张光卫 博士生,研究方向:不确定性人工智能、复杂网络、数据挖掘。

病传播和防控,对如何在网络化数据上进行知识发现和数据挖掘,以及对复杂网络的化简等均有重要的意义<sup>[18,25]</sup>。

本文着重对社区发现的策略和方法进行总结和分析,第2部分阐述了传统图论和社会科学领域在网络结构研究中的成果,对传统网络的分割算法进行了总结和分析。第3部分总结了近几年复杂网络社区结构研究中的新进展,着重讨论了研究思想和发现算法。最后给出本领域有待解决的几个问题和发展方向。

## 2 传统的网络分割算法

如何把网络分割为相对独立的部分,在传统图论和社会科学中已有多年研究历史,我们把这些研究归类于传统的方法。传统网络分割策略和算法对研究复杂网络中社区发现有一定借鉴意义。

### 2.1 图论方法

图论中一个典型的问题是如何把网络中的  $n$  个结点划分到  $g$  个组中,使得各个组中结点的数目大致相同,并且在各个组之间的边的数目尽量少,称为“图分割问题”(Graph Partitioning Problem)<sup>[27]</sup>。图分割方法大多采用 2 选 1 的策略,首先把整个图分解为最优的两个子图,然后对两个子图也进行同样的处理,反复进行,直到得到足够数目的子图。算法大体分为两类:

(1)基于谱分解(Spectral bisection)的方法<sup>[28,29]</sup>。通过分析网络拉普拉斯算子(Laplacian)<sup>[30]</sup>的特征向量完成分割。当网络能够恰好被分解成两个社区时,算法的执行效果非常好,而且一般情况下谱分解算法运算相当快。由于大多数情况下拉普拉斯算子是个稀疏矩阵,特征向量的计算可以用 Lanczos 算法<sup>[33]</sup>在较短的时间内完成。算法的缺点为每一次分割必须把网络分解成两个部分,通过重复调用算法来完成多个社区的分割,由于不能判断整个网络被分解为多少个社区合适,算法很难达到满意的分割效果。

(2)Kernighan-Liu 算法(简称 KL 算法)<sup>[31]</sup>。通过基于贪婪优化的启发式过程把网络分解为两个社区,优化函数  $Q$  为落在两个社区内部的边数与落在两个社区之间的边数的差。算法的执行过程:①初始化,人为指定社区的大小(结点数)以及两个社区的最初配置(可按照指定的大小把结点随机划分到两个社区中)。②考察所有分别取自两个社区的结点对,计算优化函数的变化量  $\Delta Q$ ,选出使  $\Delta Q$  最大的结点对,然后交换两个结点。凡是被交换过的结点在后面的过程中不再被交换,重复执行,直到有一个组中的所有结点均被交换过。③考察步骤②的所有交换,找出其中使  $Q$  最大的那次,该时刻结点在两个社区中的分布便为网络的最佳分割。算法的最坏时间复杂度为  $O(n^2)$ 。启发式算法的缺点为没有利用网络的全局结构进行分析,而且必须事先指定社区的大小,如果指定的值和实际的情况不一致就会出现错误的划分,这使得 KL 算法几乎难以在现实的网络中应用。

谱分解与 KL 算法都是二分法,难以把网络分解为合理个数的社区。

### 2.2 社会学方法

社会学家有若干解释和分析社会网络的理论方法,最典型的是基于同类相近原则的分层聚类(Hierarchical Clustering)<sup>[32]</sup>,根据任意两个结点对  $(i, j)$  之间相似程度  $x_{ij}$  把社会网络自然划分成组。相似程度有多种不同的定义,常用的是基于结点无关路径条数的定义<sup>[35]</sup>,认为结点间的相似程度由

结点无关路径的条数决定。结点无关路径是指两条路径仅仅起点和终点相同,而经过的其他结点均不同。结点  $i, j$  之间的结点无关路径的条数等于割断它们之间的联系需要移走的顶点的最小数目。另外,相似程度还有类欧氏距离的定义<sup>[34]</sup>、基于 Pearson 相关系数的定义<sup>[15]</sup>等。算法分为两种:聚集(Agglomerative)和分裂(Divisive),取决于算法通过添加还是移去边来完成<sup>[35]</sup>。

一种聚集算法的执行过程为:①初始化起始状态为  $N$  个孤立结点。②计算网络中每一对结点的相似程度。③根据相似程度从强到弱连接相应结点对。④算法可在任何时刻结束,边的添加过程用一个树状图(Dendrogram)来表示。如图2所示,树叶表示网络中的结点,树枝把结点或一组结点连接在一起,形成的社区具有嵌套的层次结构。参考文<sup>[36]</sup>给出了一个典型的社区分割实例。

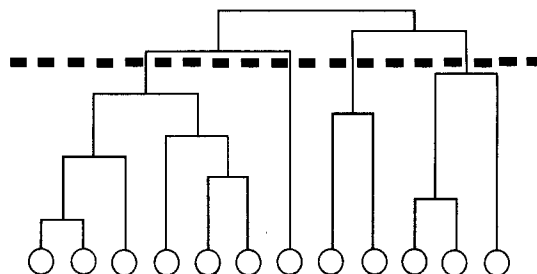


图2 表示分层聚类结果的层次树

底层的圆形结点为网络中的结点。从底向上,结点之间通过树枝连接成越来越大的社区,最终形成一个大的社区。从不同的层次可以把网络截成不同规模的社区,在图中虚线部分截取能够把网络分解为4个社区。

分层聚类方法的缺点首先是算法的适应性问题,算法对于那些社区结构已知的网络有较好的结果,比如在电影演员合作网络<sup>[9,37]</sup>(演员被看成结点,如果两个演员在同一部电影中充当角色,那么认为他们之间存在连接)中能够清楚地得到演员合演电影的个数<sup>[38]</sup>。对于社区结构未知的网络,算法准确度较差<sup>[39]</sup>。另外,算法虽然能对核心结点进行很好的分类,但是对于外围结点的分类却经常出错<sup>[39]</sup>。如图3所示,核心结点(深色)在网络中有较强的连接,它们能够在算法执行的早期被连接在一起。浅色的为外围结点,这些点与社区往往只有一条边相连,图中结点1属于社区A是比较合适的。

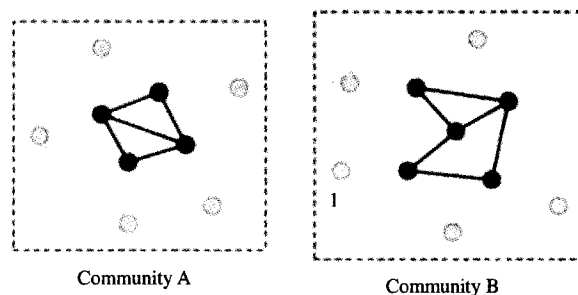


图3 分层聚类方法对核心结点能进行很好的分类,对于外围结点的分类却经常出错,如图结点1未能正确地分配社区

## 3 复杂网络中社区发现算法研究

针对复杂网络的特点,Palla 提出了社区发现算法应该满

足的5个条件<sup>[40]</sup>:(1)算法不能太严格;(2)应当基于连接的密度;(3)应当有局部性;(4)不能产生截断点和截断边;(5)允许社区的交错。这说明在复杂网络中社区的发现存在一定的不确定性和模糊性,在大量的结点面前,没有必要进行过分严格的划分。以下分别讨论近几年提出的重要社区发现算法。

### 3.1 GN 算法

Girvan 和 Newman 于 2001 年提出一个基于边介数的社区发现算法(GN 算法)<sup>[19]</sup>。算法采用分割法,依据边不属于社区的程度逐步把不属于任何社区的边(即社区之间相连接的边)删除,直到把所有的边都删除掉。

作者用边介数衡量边在社区间连通中的重要程度和不属于任何社区的程度。边介数是指所有最短路径通过该边的次数。按照社区的定义,社区内部结点之间联系紧密,而社区之间通过少量的边连接在一起,社区之间的边比社区内部的边有更大的边介数,通过逐步移去这些介数较高的边就能够把它们连接的社区分割开来。算法如下:

- ①计算网络中所有边的边介数;
- ②移去介数最高的边;
- ③重新计算所有受影响边的介数;
- ④从 2 重复执行,直到所有的边都被移走。

类似社会学分层聚类方法,GN 算法的执行过程也是一个把网络结点映射到树状图上的过程。算法最坏情况下每移走一条边就需要重新计算所有边介数。如果采用 Newman 提出的计算边介数的快速算法<sup>[41]</sup>,对于  $n$  个结点和  $m$  条边的网络,求解出所有边的介数的时间复杂度为  $O(mn)$ ,那么整个算法的复杂度为  $O(m^2n)$ 。GN 算法是社区探测领域的一个重要里程碑,它从网络的全局结构出发避免了传统算法的若干缺点<sup>[39,42]</sup>,成为目前进行网络社区分析的标准算法,得到了广泛的应用<sup>[43~45]</sup>。

GN 算法也有明显缺点。首先,算法的时间复杂度高,因为需要多次计算边介数,而几乎每一次计算都要分析整个网络。其次,通过树状图把网络分解到结点,强迫任何一个结点必须属于一个社区,而不考虑是否真正有意义。

针对算法的时间复杂性高的问题,Newman 2004 年提出了一个基于模块化的快速算法<sup>[21]</sup>。首先定义一个  $k \times k$  对称矩阵  $e$ ,它的每个元素  $e_{ij}$  表示网络中连接社区  $i$  和  $j$  的边所占的比例<sup>[46]</sup>,矩阵的迹  $Tr \square e = \sum_i e_{ii}$  表示落在社区内的边占总边数的比例。矩阵中行(或列)的和  $a_i = \sum_j e_{ij}$  表示与社区  $i$  中的点相连接的边所占的比例。另外,不考虑社区的存在,假设边在结点之间是随机连接的,那么  $e_{ij} = a_i a_j$ 。因而社区模块化程度可以由  $Q = \sum_i (e_{ii} - a_i^2)$  来度量, $Q$  表示落在社区内的边所占比例与随机连接时的期望值之间的差,可以用来衡量网络模块化的程度。在对网络结点进行划分的时候,根据  $Q$  的值决定一个结点是否应当属于一个社区。如果把该点放入某个社区,导致社区内的边数小于随机连接时的期望值,那么  $Q=0, Q>0.3$ (经验值)意味着该划分形成的社区结构是显著的。文<sup>[39]</sup>给出了几个实际的例子。 $Q$  的优化非常耗时,把  $n$  个结点划分到  $g$  个组中(所有组均不能为空)的分法的总数为  $\sum_{g=1}^n S_n^{(g)}$ ,其中  $S_n^{(g)}$  为第二类 Stirling 数,Newman 使用贪婪策略进行了优化。算法执行时的开始状态为  $n$  个孤立结点形成的  $n$  个社区,重复对社区进行合并,并使得每一次合并  $Q$  的值增长最快(或下降最慢),每次合并的结果可以用树状图来表示。在  $Q$  最大的层次截断树状图将得到最佳的社区划分。

### 3.2 改进的 GN 算法

Radicchi 等人认为 GN 算法由于缺少量化的社区定义,造成划分的合理性难以保证<sup>[47]</sup>,并于 2003 年提出了一种量化定义社区的方法,在此基础上对 GN 算法进行了改进。用  $G$  表示整个网络,网络结点的邻接矩阵为  $A_{ij}$ ,结点  $i$  的度  $k_i = \sum_j A_{ij}$ ,考虑子图  $V \subset G (i \in V), k_i^n(V) = \sum_{j \in V} A_{ij}$  表示  $V$  内部与结点  $i$  相连的边数,  $k_i^m(V) = \sum_{j \notin V} A_{ij}$  表示  $V$  外部与结点  $i$  相连的边数,于是结点  $i$  的度按照贡献被划分成内部连接数和外部连接数两类  $k_i(V) = k_i^n(V) + k_i^m(V)$ 。以此为基本 Radicchi 给出量化社区的强定义和弱定义。

**定义 1** 子图  $V$  为满足强定义的社区,当且仅当  $k_i^n(V) > k_i^m(V), \forall i \in V$ 。

满足强定义的社区中每个结点与社区内部结点的连接数大于与外部结点的连接数。

**定义 2** 子图  $V$  为满足弱定义的社区,当且仅当  $\sum_{i \in V} k_i^n(V) > \sum_{i \in V} k_i^m(V)$ 。

满足弱定义的社区中所有点与内部结点的连接数的和大于与社区外部结点的连接数的和,显然满足强定义的社区一定满足社区的弱定义。

基于对社区的量化定义,Radicchi 提出 GN 算法的改进版本:

- ①选择一种社区的量化定义方式(强定义或弱定义);
- ②计算所有边的介数;
- ③把介数最大的边去掉;
- ④如果步骤③不能使网络分解为至少两个子网,那么转步骤②执行;
- ⑤测试所有子网是否满足步骤①选择的社区定义方式。如果至少有两个子网满足定义,那么在树状图上画出相应的部分;
- ⑥返回步骤②,对所有子图重复执行,直到网络中没有边存在。

Radicchi 提出的算法在一定程度上改善了 GN 算法,得到了较好的实验效果。改进后算法的时间复杂度为  $O(aM + bM^2)$ ,其中  $M$  为边的数目,比 GN 算法有较好的改善。

### 3.3 关于社区的交错和嵌套

GN 算法以及诸多改进版本<sup>[21,44,47]</sup>在网络由相互独立的社区组成时非常有效。但现实中的网络往往是由互相交错和嵌套的社区组成的,如图 4 所示。比如,在科学家合作网中一个物理学家也可能是一个数学家,他将同时处于两个社区中<sup>[40]</sup>;在蛋白质交互网络中,一种蛋白质可能被划分到多个功能组中。这些情况是 GN 算法所没有考虑的。

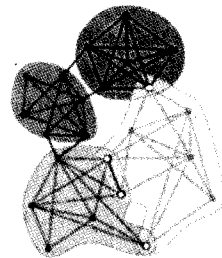


图 4 社区的交错现象

图中的空心结点均为社区的交错结点,它们同时属于多个社区。

Palla 等人 2005 年在“Nature”杂志上发表文章探讨了这个问题<sup>[40]</sup>,认为社区的交错和嵌套是复杂网络的一个重要结构特征。为了方便研究这些特征,作者给出了基于  $k$ -

clique<sup>[48]</sup>的社区定义。 $k$ -clique 是一个全连通的子图, 可以看成构成网络的基本元素, 它们是重复出现的重要网络连接模式, 这些小的有清晰定义子图对理解网络的结构有重要的意义<sup>[49,50]</sup>。 $k$ -clique 有如下特点:

①  $k$ -clique 邻接。如果两个  $k$ -clique 共享  $k-1$  个结点, 那么认为它们是邻接的。

②  $k$ -clique 链。指一系列邻接的  $k$ -clique 组成的集合。

③  $k$ -clique 连接。如果两个  $k$ -clique 出现在一个  $k$ -clique 链中, 那么认为它们是连接的。

④  $k$ -clique 渗透簇。指由多个  $k$ -clique 链组成的最大的集团。

如图 5 所示。

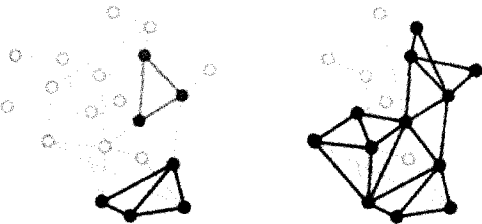


图 5  $k=3$  的  $k$ -clique 示意图

左边深色结点为两个小的  $k$ -clique 簇, 右边深色结点构成一个较大的  $k$ -clique 簇。

Palla 给出了基于  $k$ -clique 的社区定义<sup>[40]</sup>: 社区可以是多个  $k$ -clique 的连接, 它们之间可以通过邻接的  $k$ -clique 互相到达。这个定义的好处是能够表示一个社区内部连接的紧密性, 说明了社区内部的成员是可达的。这类类似于搭积木, 每个  $k$ -clique 为一个积木, 整个社区是由多个积木搭建而成的。

基于  $k$ -clique 定义的社区发现算法能够分析出社区之间的交错情况, 首先求出网络中所有的  $k$ -clique, 得到网络的 clique 交错矩阵<sup>[51]</sup>, 然后通过分析该矩阵, 得到划分的社区。算法的时间代价为指数形式。交错社区发现算法对真实数据的分类较为科学和合理。GN 算法及其改进算法均使用层次结构的树状图表示划分的结果, 无法表示社区的重叠现象, Palla 给出的算法通过分析 clique 交错矩阵直接得到社区的划分, 从而避免了层次树图的产生。

**总结** 本文介绍了复杂网络的集团结构特征, 重点对社区发现算法进行了讨论和分析。指出了图论方法、社会学方法和 GN 算法的优缺点。社区发现是一个 NP 完全问题。图论中已经有若干启发式算法可以解决这个问题, 其缺点在于不利于从全局结构分析和理解网络, 难以判断结果的最优性, 即把网络分解为多少个社区最合理的问题。社会学领域层次聚类方法的适应性存在问题, 虽然对核心结点能够进行很好的分类, 但对外围结点的分类却经常出错, 也不适合结构未知的复杂网络。GN 算法从网络全局结构进行分析, 通过边介数进行社区划分, 缺点是强迫任何一个结点必须属于一个社区, 而没有考虑是否真正有意义。GN 算法的多个改进版本部分地降低了算法的时间复杂度和提高了准确率。传统的和以 GN 算法为代表的社区分类算法均没有考虑社区的交错和嵌套特征。Palla 等 2005 年在“Nature”上发表文章, 讨论和分析了复杂网络中社区嵌套和交错的问题, 指出了这个领域一个新的方向。

复杂网络的集团结构特征研究是一个新的研究领域, 存在诸多值得研究的问题: ①在社区的探测和发现算法方面, 如

何进一步降低算法的时间复杂度, 使得算法能够处理具有大量结点的复杂网络? 如何综合考虑局部和全局特征, 提高社区划分的合理性? 如何衡量社区探测中的不确定性和模糊性? ②复杂网络的演化过程中社区特征的进化规律如何, 全局化和社区化的关系如何? ③复杂网络的集团化特征对病毒在网上的传播有什么影响? 对网络的抗毁性、健壮性和稳定性的影响如何? ④能否通过把社区抽象成点进行复杂网络的简化? ⑤怎样在网络化数据(侧重数据之间的联系而不是数据的各种属性)上进行知识发现和数据挖掘? PNAS 2003 年举办了关于“科学地图”(Mapping knowledge domains)<sup>[52]</sup>研讨会, 具体就如何充分利用 PNAS 1997 到 2002 年间的 93,000 多页科技论文进行了讨论, 认为研究和理解科学协作网络<sup>[13,41]</sup>的社区结构在海量电子文献的分析、分类以及展示中有重要的意义。

值得指出的是, 复杂网络是近几年非常热的一个新兴学科交叉研究领域, 研究人员绝大多数来自物理、数学、生物等领域。计算机科学的研究人员也应该深入到复杂网络的前沿阵地, 做出应有的贡献。

## 参考文献

- 1 许国志, 等. 系统科学[M]. 上海: 上海科技教育出版社, 2000
- 2 Barabasi A L, Albert R. Emergence of scaling in random networks. *Science*, 1999, 286, 509~512
- 3 Cohen J E, Briand F, Newman C M. *Community Food Webs: Data and Theory*. Springer, Berlin, 1990
- 4 Williams R J, Martinez N D. Simple rules yield complex food webs. *Nature*, 2000, 404, 180~183
- 5 Dorogovtsev S N, Mendes J F F. *Evolution of Networks: From Biological Nets to the Internet and WWW*. New York: Oxford University Press, 2003
- 6 Hartwell L H, Hopfield. From molecular to modular cell biology. *Nature*, 1999, 402, C47~C52
- 7 Bhalla U S. Emergent properties of networks of biological signaling pathways. *Science*, 1999, 283, 381~387
- 8 Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic networks. *Nature*, 2000, 407, 651~654
- 9 Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks. *Nature*, 1998, 393, 440~442
- 10 Albert R, Jeong H, Barabasi A L. Diameter of the world-wide web. *Nature*, 1999, 401, 130~131
- 11 Broder A, Kumar R, Maghoul F. Graph structure in the web. *Computer Networks*, 2000, 33, 309~320
- 12 Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the internet topology. *Comp. Comm Rev*, 1999, 29, 251~262
- 13 Newman. The structure of scientific collaboration networks. *Proc Natl Acad Sci USA*, 2001, 98, 404~409
- 14 Redner S. How popular is your paper? An empirical study of the citation distribution. *Eur J Phys*, 1998, B4, 131~134
- 15 Wasserman S, Faust K. *Social Network Analysis*. Cambridge: Cambridge University Press, 1994
- 16 Scott J. *Social Network Analysis: A Handbook*, 2nd edition. London: Sage Publications, 2000
- 17 Stogatz S H. Exploring complex networks. *Nature*, 2001, 410, 268~276
- 18 Complex networks and networked data mining. *Lect Notes Artif Int*, 2005, 3584, 10~12
- 19 Girvan M, Newman M E J. Community structure in social and biological networks. *Proc Natl Acad Sci USA*, 2002, 99, 7821~7826
- 20 Everitt B S. *Cluster Analysis*. 3rd edn. London: Edward Arnold, 1993
- 21 Newman M E J. Fast algorithm for detecting community structure in networks. *Phys Rev*, 2004, E 69, 066133
- 22 Ravasz E, Somera A L, Mongru D A, et al. archical organization of modularity in metabolic networks. *Science*, 2002, 297, 1551~1555
- 23 Spirin V, Mirny L A. Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci USA*, 2003, 100, 12123~12128
- 24 Onnela J P, Chakraborti A, Kaski K, et al. Dynamics of market correlations: Taxonomy and portfolio analysis. *Phys Rev*, 2003, E 68, 056110
- 25 Classifying class and finding community in UML metamodel network. *Lect Notes Artif Int*, 2005, 3584, 690~695

(下转第 28 页)

**结论** 本文提出了一种新的固定速率分层组播拥塞控制算法 FLMCC。组播会话中的每层按照固定速率发送数据包。各接收端采用窗口机制,利用 TCP 友好而平滑的 GAIMD 算法完成窗口调整,并根据当前窗口值计算期望速率;根据期望速率订购数目不等的层,从而获得不同的吞吐量。为测量 RTT,采用了一种精确测量和粗略测量相结合的

策略;精确测量时,为避免大量反馈产生反馈内爆,采用了随机定时器策略。仿真表明,算法具有良好的 TCP 友好性、响应性和协议内公平性,链路利用率高。算法可为异构的组播成员有效提供多速率服务,增强了组播拥塞控制的可扩展性,并且实现简单。

表 1 多个 FLMCC 组播流的公平性指数及链路利用率

参数	组播流数目									
	2	4	6	8	10	12	14	16	18	20
公平性指数	0.9972	0.9986	0.9970	0.9984	0.9976	0.9982	0.9973	0.9991	0.9971	0.9986
链路利用率	0.9818	0.9857	0.9833	0.9923	0.9953	0.9915	0.9938	0.9974	0.9960	0.9973

**参 考 文 献**

- 1 Nonnenmacher J, Biersack E W. Scalable feedback for large groups. *IEEE/ACM Trans on Networking*, 1999,7(3): 375 ~ 386
- 2 Bhattacharyya S, Towsley D, Kurose J. The Loss Path Multiplicity Problem in Multicast Congestion Control. In: Doshi B, ed. *Proc. of IEEE INFOCOM*. New York, USA, 1999
- 3 Kwon G, Byers J W. Smooth Multirate Multicast Congestion Control. In: *Proc. of IEEE INFOCOM 2003*, San Francisco, USA, 2003
- 4 Byers J, Horn G, Luby M, et al. FLID-DL: Congestion Control for Layered Multicast. *IEEE JSAC Special Issue on Network Support for Multicast Communication*, 2002, 20(8): 1558~1570
- 5 McCanne S, Jacobson V, Vetterli M. Receiver-Driven Layered

- Multicast. In: *Proc. of ACM SIGCOMM '96*, 1996
- 6 Vicisano L, Rizzo L, Crowcroft J. TCP-like Congestion Control for Layered Multicast Data Transfer. In: *Proc. of IEEE INFOCOM '98*, 1998
- 7 Legout A, Biersack E W. PLM: Fast convergence for cumulative layered multicast transmission schemes. In: *Proc. of ACM Sigmetrics*, 2000
- 8 Yang R Y, Simon S L. General AIMD Congestion Control. In: *Proc. of the 8th International Conference on Network Protocols*. Osaka, Japan, 2000
- 9 Widmer J, Handley M. Extending Equation-based Congestion Control to Multicast Applications. In: Floyd S, ed. *Proc of ACM SIGCOMM*. San Diego, USA. 2001
- 10 Chiu D-M, Jain R. Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks. *Computer Networks and ISDN Systems*, 1989, 17(1): 1~14

(上接第 4 页)

- 26 Watts D J, Dodds P S, Newman M E J. Identity and search in social networks. *Science*, 2002, 296: 1302~1305
- 27 Karypis G, Kumar V. Multilevel Algorithms for Multi-Constraint Graph Partitioning. In: *Proceedings of the 1998 ACM/IEEE Conference on Supercomputing(CDROM)*, San Jose, CA, November, 1998, 7~13
- 28 Fiedler M. Algebraic connectivity of graphs. *Czech Math J*, 1973, 23: 298~305
- 29 Pothan A, Simon H, Liou K P. Partitioning sparse matrices with eigenvectors of graphs. *SIAM J Matrix Anal Appl*, 1990, 11: 430~452
- 30 Mohar B. The Laplacian spectrum of graphs. *Graph Theory, Combinatorics, and Applications*, 1991, 2: 871~898
- 31 Kernighan W, Lin S. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal*, 1970, 49: 291~307
- 32 Scott J. *Social Network Analysis: A Handbook*. 2nd edition. Sage, London, 2000
- 33 Golub G H, Van Loan C F. *Matrix computations*. Baltimore, MD; Johns Hopkins University Press, 1989
- 34 Burt R S. Position in networks. *Social Forces*, 1976, 55: 93~122
- 35 Scott J. *Social Network Analysis: A handbook*. 2nd edition. London; Sage Publication, 2000
- 36 Breiger R L, Boorman S A, Arabie P. An algorithm for clustering relations data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology*, 1975, 12: 328~383
- 37 Amaral L A N, Scala A, Barthélemy M, et al. Classes of small-world networks. *Proc Natl Acad Sci USA*, 2000, 97: 11149~11152
- 38 Marchiori M, Latora V. Harmony in the small-world. *Physica A*, 2000, 285: 539~546

- 39 Newman M E J, Girvan M. Finding and evaluating community structure in networks. *Phys Rev*, 2004, E 69: 026113
- 40 Pallal G, Derenyi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 2005, 435: 814~817
- 41 Newman M E J. Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. *Phys Rev*, 2001, E 64:016132
- 42 Girvan M, Newman M E J. Community structure in social and biological networks. *Proc Natl Acad Sci USA*, 2002, 99: 7821~7826
- 43 Holme P. Subnetwork hierarchies of biochemical pathways *Bioinformatics*, 2003, 19: 532~538
- 44 Wilkinson DM, Huberman BA. A Method for Finding Communities of Related Genes. *Proc Natl Acad Sci USA*, 2004, 101 (Suppl 1):5241~5248
- 45 Gleiser P, Danon L. Community structure in jazz. 2003, 0307434
- 46 Newman M E J. Mixing patterns in networks. *Phys Rev*, 2003, E 67:026~126
- 47 Radicchi F, Castellano C, Cecconi F, et al. Defining and identifying communities in networks. *Proc Natl Acad Sci USA*, 2004, 101: 2658~2663
- 48 Derenyi I, Palla G. Clique Percolation in Random Networks. *Phys Rev Lett PRL*, 2005, 94: 160~202
- 49 Milo R, et al. Network Motifs: Simple Building Blocks of Complex Network. *Science*, 2002, 298:824
- 50 Shen-Orr S, Milo R, Mangan S, et al. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet*, 2002, 31: 64
- 51 Everett M G, Borgatti S P. Analyzing clique overlap. *Connections*, 1998, 21: 49~61
- 52 Shiffrin R M, Börner K. Mapping knowledge domains. *Proc Natl Acad Sci USA*, 2004, 101: 5183~5185