
摘要

近年来，复杂网络的社团研究相对比较成熟，可针对以 Internet 拓扑结构为基础的社团特征研究至目前为止还是相对不足，并且没有很好的从 Internet 特有的结构特征入手。系统的社团结构与其功能息息相关，对于 Internet 社团结构特征的深刻理解，可以帮助人们发现互联网新的功能单元、理解其复杂的网络特征结构、化简网络结构、寻找防治互联网病毒传播新策略、网络控制等。

针对互联网特有的结构特征，如 Internet 网络边缘树状结构特征、Internet 的链状特征(海底光缆、网络专线、BGP 外部路由线路)、网络中局部区域高核数节点相互聚类特征、中心节点特征等，本文在互联网社团研究的过程中，通过对比传统的复杂网络的社团发现算法，总体上运用了分而划之、合而聚之的思想，分别设计了层次折叠收缩算法、链状探测算法、高核节点聚类特征探测算法、中心节点算法，通过这些算法，大致可以找出 Internet 基本的较小的功能组织结构，比如，层次折叠收缩算法可以探测到 Internet 网络边缘树状结构特征组织，链状探测算法可以探测到 Internet 网络专线、BGP 外部路由线路等链状的结构特征组织，高核节点聚类探测算法可以探测到网络中局部区域高核数节点相互聚类特征组织，中心节点算法可以探测到 Internet 中心节点特征组织。通过这些探测算法，网络中的大部分节点都会被归类到相应的组织结构中去，可是还有一些少部分节点可能不在上述的几种特征组织内，上面的几种有针对性的探测算法就无法把它归类到相应的组织结构内，所以就这些没有被探测到的每一个节点归为单独的组织结构（便于后续处理）。经过前面几步的处理，一个完整网络拓扑的大大小的组织结构就呈现出来，下一步就需要把这些大大小小的网络组织结构合并成为一个个具有高内聚低耦合特征的社团，这就是合而聚之。可是要使合并后的网络的社团模块度达到最优化是一个 NP 难题，在时间复杂度和空间复杂度上都比较大，对于具有众多较小组织结构的网络的社团模块度寻优处理，本文设计的小社团合并算法采用贪婪算法来寻找局部最优的模块度，最终一个具有典型社团结构划分特征的网络拓扑浮出水面。本文设计的社团发现算法由一组多粒度社团发现算法混合而成，称之为 105 算法。目前在 IPV6 网络的社团发现方面：就社团划分质量的主流衡量标准模块度而言，105 算法微弱于 fast unfolding of communities in large networks(简称 FUOCILN)算法

，在模块度评价方面，FUOCILN 是目前国际上最优秀的算法这一，这也从侧面印证了 105 算法在模块度评价标准上的优秀性；就多粒度特征而言，105 算法发现的社团内部包含多种不同粒度的互联网基本拓扑结构，比如链状、高核聚类状、局部树状、中心节点状等，而 FUOCILN 算法发现的社团内部只有节点信息，105 算法的社团内部信息容量要远大于 FUOCILN。

对于互联网的社团可视化问题，传统的一些可视化算法仅仅只能做到整个网络拓扑级别的可视化，社团结构的可视化特点并不明晰，针对这种问题，本文在参考传统节点布局方式的基础上全新构建节点布局，这种布局分为两个层次：整个网络的社团之间位置布局、每个社团的内部节点位置布局。其中，第一个层次又分两个小层次：其一，根据物理类比法，模拟物理系统环境，全自动布局社团节点；其二，在第一步完成之后，采用动态交互布局模式，如果第一步布局有不合理的地方，可手动调节社团节点位置。由于大规模数据的复杂性以及考虑了突出社团特征等问题，这必将牺牲画布资源。为此，本可视化算法开辟一个新的画布资源，采用射线布局算法显示某个特定社团内部节点的拓扑结构。鉴于本可视化算法的特点，称之为 visualCommunity 算法（简称 VC）。VC 算法能够合理清楚的可视化互联网的社团结构特征。

关键词：Internet 拓扑结构；社团发现；多粒度；社团可视化

Community Structure Detecting of Multiple Granularity and Visualization Based on Internet Network Topology Abstract

The community studies of complex network characteristics of the Internet topology so far is relatively lacked, and not starts from the Internet's unique structural features. Community structure of the system is closely related to the functions. a deep understanding of the characteristics of the Internet community structure can feature (submarine cable、network、BGP external route), the clustering feature of the the course of the study of the Internet community compares to the traditional complex network of community detecting algorithms, use the dividing and getting together methods and design the level folding contraction method、the chain-detection method、node clustering feature of the method、the central node method. These methods can be used to roughly find the Internet smaller functional organizational structure, for example, the level folding contraction method can detect the edge of the tree structure feature organization of the Internet network, chain detection method can detect the structural method organization of the Internet submarine cable, network, chain characteristics organization like external route of BGP, method can detect organization with the some areas , the Center node algorithm can detect the feature organization of the Internet center node. Through these detection method, most of the network nodes will be classified to the appropriate organizational structure, but some small part of the nodes may be not in above several features organization. The above targeted detection method unable to be classified to the appropriate organizational structure, so they are classified as a separate organizational structure (to facilitate in the follow-up treatment). After processing on the front steps, a whole organizational structure of the entire network topology is revealed, the next step is to merge these different sizes of the network organizational structure to become one community with and low coupling characteristics, which is together. But to make the merged network community modularity to achieve optimal is an NP- process of community module for many smaller organizational structure of the network community, small communities merge method designed by this article uses a greedy algorithm to find the local optima of modularity, and eventually a network topology with the characteristics of the typical community structure division surfaced. In this paper, the community detect algorithm

source not found.

is made up of a group of multi-granularity community detect method, known as 105 algorithm. This algorithm in the IPV6 network community is superior to fast unfolding of communities in large networks (referred FUOCILN) algorithm, which is known as beyond many traditional societies detecting algorithm. Nowadays, this algorithm is further optimized and promising.

For visualization of the Internet community, some of the traditional visualization algorithms only can do the whole level visualization of network topology, visualization feature of the community structure is not clear. For this problem, this article builds a new graph layout on the foundation of the traditional graph layout; this layout is divided into two big steps: the layout of the entire network community, the layout of the internal nodes in each community. The first step consists of two levels: firstly, according to the physical analogy, it simulates the physical system environment, and automatically layouts community node; secondly, after the first step, it use dynamic interactive layout mode. If the first step manually adjust the location of the Associations node. The complexity of large-scale data and the problem of outstanding association characteristic must sacrifice the canvas resources of the internal layout. For this, the visualization algorithm uses the ray layout algorithm to display the topology of the internal nodes of a particular community and open a new canvas resource, In view of the characteristics of the visualization algorithms, it is called visual Community. Visual Community can reasonably and clearly visualize the characteristics of the Internet community structure.

Keywords: the Internet topology structure;community detecting ;multi-granularity community;visual Community;

目 录

独创性声明.....	I
摘 要.....	II
Abstract.....	III
第 1 章 绪 论.....	1
1.1 课题研究的背景与意义.....	1
1.1.1 互联网的社团结构研究现状.....	1
1.1.2 互联网可视化研究现状.....	2
1.1.3 本文研究的意义.....	3
1.2 本文组织结构.....	4
第 2 章 互联网网络数据来源及处理.....	6
2.1 互联网网络数据来源.....	6
2.2 原始网络数据预处理.....	8
2.2.1 数据格式的处理.....	8
2.2.2 原始网络拓扑.....	9
2.2.3 最大子图的获取.....	10
2.3 相关特征量.....	12
2.4 本章小结.....	13
第 3 章 互联网拓扑基本特征分析与多粒度社团发现算法设计.....	14
3.1 互联网拓扑基本特征分析.....	14
3.2 多粒度社团发现算法设计的总体思想.....	17
3.3 分而划之.....	19
3.3.1 层次折叠收缩算法.....	20
3.3.2 链状探测算法.....	25
3.3.3 高核数节点聚合探测算法.....	27
3.234 中心节点组织探测算法.....	33
3.3.5 各种网络探测算法处理后的网络效果.....	36
3.4 合而聚之.....	37
3.4.1 模块度.....	37
3.4.2 小社团合并算法.....	38
3.5 社团划分比较.....	41

3.6 对于全球 IPV6 拓扑的划分比较.....	42
3.7 105 算法在其它网络上的测试.....	44
3.7.1 Karate 俱乐部.....	44
3.7.2 东北大学嵌入式实验室聊天网络—网络的共产主义.....	46
3.7.3 东大嵌入式实验室日常人际交流网络—地理位置隔离的社团结构.....	47
3.8 新的衡量社团划分质量方法的探讨.....	49
3.8.1 新的社团发现思路.....	49
3.8.2 新的求解最短路径的算法.....	50
3.9 本章小结.....	50
第 4 章 一种新的面向社团的网络拓扑可视化算法.....	52
4.1 可视化算法综述.....	52
4.2 面向社团结构特征的可视化.....	54
4.2.1 所用开发工具.....	54
4.2.2 算法思想.....	55
4.2.3 算法设计与实现.....	55
4.3 本章小结.....	60
第 5 章 总结与展望.....	61
5.1 本文所做的工作.....	61
5.2 不足及展望.....	62
参考文献.....	63
致 谢.....	67
攻读硕士期间发表的论文.....	69

第1章 绪论

1.1 课题研究的背景与意义

1.1.1 互联网的社团结构研究现状

近年来,随着信息系统(如、电信网、广播电视网、移动网络等)的迅猛发展,网络拓扑数据的规模快速增大,以致人们不能通过传统技术和方式来管理和运作这些复杂网络。人们通过对生物网络、社会关系网、Web网等的研究,发现这些网络都具有某些共同特点,包括:整体相对稀疏,局部比较密集;顶点度值服从幂率分布,也被称为无标度特性^[1];整体分布具有高聚集度、低平均最短路径(平均最短路径为 $O(\log\log N)$)的小世界特性^[2]。具有以上性质(无标度性、小世界性)的网络被称为复杂网络。互联网具有典型的复杂网络结构,本文后续的研究手段都是以复杂网络理论为基础。

在现实世界中,存在大量复杂系统,可把这些复杂系统看作各种网络来研究。一个典型的网络是由若干节点 $v_i (v_i \in V)$ 以及连接任意两个在实际拓扑中有连接关系的节点 v_j 和 $v_k (v_j, v_k \in V)$ 的边 $e_{jk} (e_{jk} \in E)$ 组成。网络的研究最早可追溯到18世纪,数学家欧拉在对“Konigsberg七桥问题”的研究时提出的一个数学分支——图论,图论在较长的时间内一直未能有突破性进展。上世纪60年代,Erdős和Rényi建立了随机图理论^[3](Random Graph Theory),开创了复杂网络理论系统性研究。自此以后的很长时间内随机图理论一直是复杂网络结构研究的基本理论^[4]。随机图理论有明显的缺点,它不能很好的描述很多实际网络,因为大部分实际网络并不是完全随机的。例如Internet上的两个站点之间是否有超文本链接,Internet中的两个AS域之间是否有直接联系,两个作者之间是否有合作,这些都不能由随机选择来决定。

从20世纪末开始,复杂网络理论的研究已经拓展到了从物理学到社会学等众多学科中,其中Watts和Strogatz在Nature杂志上发表的文献[5]探讨了复杂网络中存在“小世界”现象,Barabási和Albert在Science杂志上发表的文献[6]探讨了随机网络中“无标度”现象,这两篇文献被认为是复杂网络研究新纪元的标志。目前对于复杂网络的研究主要包括网络结构发现、网络拓扑建模、网络特性分析及控制(稳定性、数据流通等)等几个方面。在利用复杂网络理论对实际网络拓扑结构进行研究时,用来刻画其宏观拓扑结构的网络特征量主要有度分布,平均路径长度、聚类系数、介数等等。随着对网络性质的物理意义及数学特性的研究分析,人们发现包括Internet在内的许多网络有一个共同的性质,即社团特性或者称为群聚现象。也就是说网络是由若干个“群(group)”或

“模块(module)”，在群内部节点的连接非常紧密，而相对的在各个群组之间的连接则较为稀疏。如图 1.1 所示为一个具有明显社团结构的网络，虚线圆中包含的部分就代表了网络的三个子团。

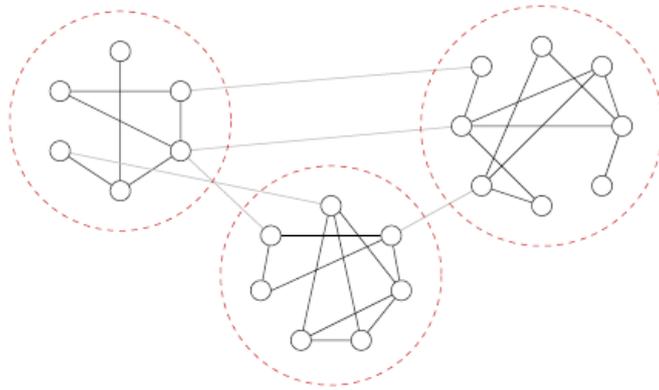


图 1.1 社团结构划分示意图

Fig.1.1 the illustration of community structure

一般而言，社团包含模块、类、群、组等各种含义。例如由大量网站社团组成，同一社团的各站点涉及的都是一些共同话题（见文献[7~9]）。在电信网络或生物网络中，也可以根据不同的性质将各节点化为不同的社团（见文献[10~12]）。提出网络中的社团结构，对于分析网络结构和特性至关重要。社团结构发现在很多领域都有广泛应用，比如生物学、物理学、计算机图形学、社会学（见文献[13~14]）等。

社团特性作为复杂网络的一个普遍现象已经在很多实际网络中得到了证实：文献[15]根据爵士乐音乐家是否曾在一个乐队演奏作为二者有联系的边，构造了一个由 jazz 音乐家组成的网络，并发现在此网络中存在着明显的社团现象；文献[16]揭示了 Internet AS 级拓扑中的社团结构与 AS 地理分布之间的关系。目前对于复杂网络中的社团特性的研究主要集中在社团划分算法、社团结构研究及基于社团结构的建模：例如文献[17]提出了一种基于信息编码为标准的社团划分算法，扩展了之前通常是基于模块度^[18,19]或是特征向量^[20]的社团划分算法；文献[21]则将社团划分算法又扩展到了加权网络中。文献[22]发现了在很多实际网络中的社团结构存在着叠加性和层次性；文献[23]提出了一种基于社团的网络演化模型。

目前对于复杂网络社团结构的研究在国际上已经相对成熟，可是针对以互联网拓扑特征为基础的社团结构研究还相对不足，传统社团发现算法并不能很好的反应互联网特有的拓扑特征。本文将根据这些不足展开对互联网特有的社团结构特征的研究。

1.1.2 互联网可视化研究现状

互联网的结构非常复杂，如果仅用数据表格或文字的形式来表示网络，理解起来非常困难，导致网络所包含的信息无从体现。将复杂网络方便、直观地表示出来的最好方法是将其进行可视化。科学计算可视化的思想是上个世纪八十年代美国科学基金会（ICV）提出的。当时在科学计算中产生了大量数据，人们很难清楚知道这些数据所表示的含义以及数据之间的关系，于是提出了将它们以图形化的方式显示出来的可视化思想。复杂网络研究的兴起进一步促进了网络可视化技术的发展，同时对可视化技术提出了更高的要求。复杂网络可视化一方面可以通过精确的展示帮助人们认识网络的内部结构，另一方面可以帮助挖掘隐藏在网络内部的有价值的信息。作为信息可视化的一个重要分支，复杂网络可视化的研究从上世纪90年代中期开始，在 Graph Drawing, InfoVis(IEEE Symposium on Information Visualization), IV(International Conference on Information Visualization)等重要国际会议中都成为一个越来越受关注的议题，引起了各国学者的高度重视。目前研究网络拓扑可视化的机构和组织有很多：以采集 Internet 数据为主要研究对象的 CAIDA，推出了 Walrus, Otter, Skitter, Plankton, MantaRay 等；SCAN.ISI.USC^[24]利用可视化工具 IGN2PCI 提供可视化接口；Bell 实验室^[25]的可视化工作则充分借鉴了力引导布局思想，对力学公式做了局部修改，并且在布局较大规模的网络拓扑时采用了树型布局思想；此外，还有 NLANR^[26]开发的 CiChild, DELIS 开发的 BGPlay, 伯克利大学的 gnuTellaVison, Graphviz, Pejek, Gephi 等。这些可视化工具中包含着各式各样的图布局算法。目前常见的布局算法有这几种：随机布局、粒子布局、放射布局、力引导布局、GEM 布局、环布局等。其中弹力布局算法运用物理类比法来模拟物理系统，ForceAtlas2 算法是弹力布局中的经典之作。这些算法针对侧重的某些特点各有不同的节点布局方式，可是能突出社团结构特征布局的可视化算法则相对较少。本文将根据这些不足展开对突出社团结构特征的可视化研究。

1.1.3 本文研究的意义

复杂网络中的社团结构发现具有重要意义，比如发现复杂系统中的功能单元（相同主题的站点组成的社团、生物网络中相同功能的细胞单元组成的社团、蛋白质相互作用网络的社团、人际社会网络中的社团），帮助人们观察和理解网络特征结构，简化复杂的网络以便更加观察和理解网络拓扑，指导人们更好的研究网络动力学行为（病毒传播的防护、网络牵引控制等）。

目前对于复杂网络社团结构的研究在国际上已经相对成熟,可是针对以互联网拓扑特征为基础的社团结构研究还相对不足,传统社团发现算法并不能很好的反映互联网特有的拓扑特征。本文提取 CAIDA 上的互联网拓扑数据,建立互联网拓扑,并利用经典可视化算法 forceAtlas2^[27]对互联网拓扑结构可视化,发现了互联网特有的一些拓扑特征,如边缘局部特征(类树状结构)、链状结构特征、局部区域高核数节点相互聚类特征、中心节点特征等,这些特殊的结构特征并不是偶然形成的,而是有着很强的规律性的,比如类树状结构就是一个真实的局域网络的结构,链状结构特征则说明这是一条网络专线或者是 BGP 的外部路由线路等,局部区域高核数节点相互聚类特征则是由互联网的非均质性和混沌性所引起(后面会详细解释),而中心节点特征则说明网络中存在一些重要的节点。如果能设计一种针对这些特有的互联网的拓扑特征探测算法,那么将对互联网的社团研究提供帮助。

如今,许多科学研究者都开始应用复杂网络理论和数据挖掘方法去分析各种现象,解决实际的问题。虽然各个学科学者的研究方向都不尽相同,但研究的思路有着一定的共性,研究的方法有着相似的过程:问题开始是对所研究领域的数据进行网络建模,之后对建立的网络模型应用复杂网络和传统数据挖掘的各种算法进行统计分析,计算网络静态几何特征量,研究网络随时间的演化过程,挖掘网络所包含的信息与知识,最后对网络结构和研究结果进行可视化展示,得出结论。在整个研究过程中,仅系统建模的步骤只需要数学理论和相关的领域知识,而其它步骤如:对网络进行算法分析和对网络结构进行可视化展示等,都需要计算机软件的辅助来完成。然而,在不同学科的研究过程中,很多学科专家对计算机工具的开发并不熟悉,从而无法集中精力来解决其领域问题,影响了他们的研究进程和效果。因此复杂网络的可视化对于处理大规模关系数据的科学家们来说至关重要。互联网是典型的复杂网络,互联网的可视化研究也有十几年的历史,互联网的可视化可以帮助人们更加直观的观察网络,发现潜在的网络特征和功能,同时下一代互联网的设计也需要对网络的结构和功能有深刻了解。对互联网拓扑结构中的社团特征的可视化研究在很多方面都有着重要意义,比如,简化整个网络拓扑,帮助人们理解网络的整个拓扑架构,社团还可以映射具体的地理位置,使各地的互联网的发展普及情况一目了然。

1.2 本文组织结构

本文共分为五章,各章的主要内容如下:

第一章为绪论,介绍了互联网的社团结构研究以及可视化研究的现状,论述了互联网的社团结构研究以及可视化研究的重要意义。

第二章介绍了互联网拓扑数据的来源，以及数据的预处理过程，并且还简单介绍了后续论文用到了一些特征量的含义。

第三章介绍了互联网拓扑基本特征分析与多粒度社团发现算法设计

，在本章中，着重分析了互联网拓扑中基本的拓扑特征结构，并且提出了探测每种拓扑结构的算法，最后对算法做了验证。

第四章介绍了传统的一些可视化布局算法，并提出了新的面向社团的网络拓扑可视化算法 `visualCommunity`。`visualCommunity` 在可视化网络拓扑时分为三步：第一步，模拟物理系统环境，全自动布局社团节点；第二步，采用动态交互模式，手动调整社团节点位置；第三步，采用射线布局模式，可视化社团节点内部的拓扑结构。

第五章是本文的总结与展望。

第 2 章 互联网网络数据来源及处理

互联网是一个分层设计的网络，Internet 中的计算机在物理层面上基于物理连接关系可以构成一种拓扑，而在应用层面上还可以构成应用层拓扑，比如在 P2P 应用中，peer^[28,29]之间构成的拓扑。目前所讨论较多的 Internet 拓扑有三种：IP 级拓扑^[30,31]、路由级拓扑和自治域级拓扑。

定义 2.1 IP 级拓扑：traceroute 机制测量得到的选路路径是 IP 地址级的路径，除最后一跳之外，其它地址各对应一个路由器接口地址。直接从 IP 级路径生成的拓扑图称为 IP 级拓扑^[31]图，其中一个节点代表一个 IP 地址。

定义 2.2 路由级拓扑：指互联网中的路由器基于互联关系而构成的一种拓扑。图中的节点指代的是一个路由器，而边则表示两个路由器之间存在着直接的连接关系。而由于实际物理线路比较难于测量，当前的 Internet 路由器级拓扑数据的获得通常基于 traceroute 类的工具。

定义 2.3 自治域级拓扑：表示 Internet 中各个独立的自治域之间的互联关系，这里的一个节点表示 Internet 中的一个自治域（Autonomous System），而图中的边则代表两个自治域之间通过 BGP^[32, 33]边界网关协议存在着互联关系，两个自治域之间有一条边。自治域级拓扑又称 AS 级拓扑。

本文研究社团发现选用的是 IP 级拓扑。

2.1 互联网网络数据来源

目前针对互联网探测数据有多种来源，本文使用的是 CAIDA（The Cooperative Association for Internet Data Analysis，一个对全球范围 Internet 结构及数据进行研究的国际合作机构）提供的探测数据。CAIDA 是互联网拓扑分析研究领域中的一个具有代表性和影响力的大型科研项目。CAIDA 的探测架构是 Ark（Archipelago）架构，这是一种分布式测量方式，通过元组空间实现各个探测节点之间的通信，可实现探测源点之间的协作测量。目前探测节点主要分布在北美洲、欧洲众多国家的研究院所、高校、军事机构中，而亚洲、大洋洲、南美洲、非洲分布较少。东北大学经过该组织授权建立 CAIDA 中国第一节点(Neu Node, she-cn)，成为该组织在中国地区的首个合作者。图 2.1 为 Ark 探测源点在全球的分布。

found.



图 2.1 Ark 探测源点在全球的分布

Fig.2.1 The distribution of Ark monitors in the world

Ark 项目迄今的活动节点共 54 个，其中亚洲 5 个，亚洲的 5 个测量点如表 2.1 所示。

表 2.1 Ark 位于亚洲的监测点

Table2.1 Ark monitors in Asia

Name	City	Country	Continent	Organization	Organization classification
cjj-kr	Daejeon	KR	Asia	KREONet1	research network
mnl-ph	Quezon City	PH	Asia	ASTI	research
nrt-jp	Tokyo	JP	Asia	APAN	research network
she-cn	Shenyang	CN	Asia	Northeastern University	university
tpe-tw	Hsinchu	TW	Asia	TWAREN	research network

CAIDA 的测量源点分布在世界几大洲内，从不同的视角对互联网拓扑结构进行全面的测量与分析。其测量范围的广泛程度，就目前而言在主动测量项目中较有优势，此外 CAIDA 测量源点拥有完全自主的控制权与所属权，可以不间断地持续测量而不受影响，其测量结果的数据量十分可观，有效性也能得到保障。作为 Ark 项目的成员，Neu 节点不仅自身采集互联网拓扑数据，还可以以合作者的身份，获得 CAIDA 遍布全球众多测量源点所提供的海量的、更新及时的数据。

为了清晰简单的分析互联网拓扑结构，本文选取的是 2011 年 9 月 29 号 CAIDA 全球所有站点探测到的合并的数据。

2.2 原始网络数据预处理

2.2.1 数据格式的处理

1. CAIDA 原始数据格式

CAIDA 探测的传统的数据格式不能直接在可视化工具上加以分析，需要转换格式。CAIDA 利用 traceroute 原始的数据格式如下：

```
traceroute from 203.181.248.51 to 83.148.147.42
 1 203.181.248.60 0.257 ms
 2 203.181.249.97 0.383 ms
 3 203.181.102.129 0.393 ms
 4 118.155.197.129 0.627 ms

traceroute from 203.181.248.51 to 46.182.233.99
 1 203.181.248.60 0.254 ms
 2 203.181.249.97 0.356 ms
 3 *
 4 *
 5 217.141.106.153 292.489 ms
 6 46.182.232.30 289.670 ms
```

2. traceroute 的工作原理

如果要正确的解析 CAIDA 原始数据格式，必须要深刻理解 traceroute 的工作原理。traceroute 程序的设计是利用 ICMP 及 IP 是指图 $G(V, E)$ 中所有节点的度的分布情况，可用分布函数 $p(k)$ 来描述。 $p(k)$ 表示的是一个随机节点的度恰好是 k 的概率，
$$p(k) = \frac{P(k)}{V(G)}$$
，其中 $P(k)$ 为度为 k 的节点数目， $V(G)$ 为节点总数。

2.4 本章小结

本章主要介绍了后续研究用到的互联网拓扑数据来源以及原始数据的预处理过程，最后还介绍了下文涉及到的一些基本特征量的概念。在本章的基础上，接下来一章将进行基于互联网拓扑特征的多粒度社团发现研究。

第 3 章 互联网拓扑基本特征分析与多粒度社团发现算法设计

本章主要针对互联网特有的基本拓扑特征展开分析，并设计了一组探测算法（总称 105 算法）可以分别探测这些基本的拓扑结构，之后评价了本文提出的社团发现算法；为了考验算法的普适性，在其它网络上做了验证；最后探讨了社团划分的新思路。

3.1 互联网拓扑基本特征分析

互联网作为一种典型的复杂网络，它的拓扑特征必定有复杂网络的特征。本文将用复杂网络的理论分析互联网的拓扑特征结构。对于互联网拓扑特征进行分析，有效的手段就是对其可视化，在可视化的基础上通过观察统计比较等方式来发现互联网的基本特征结构。在用可视化的手段分析前，先根据互联网的网络协议做一个实际互联网结构的猜想。

首先，互联网的结构应该是非均质的，有许许多多的局域网组成；其次，在这些局域网之间，应该有外部路由器连接；再次，每个分离的大陆之间应该有至少一条链状线路；再次，在局域网内部，应该还可以包含更小的局域网；再次，互联网作为一种典型的复杂网络，整个网络拓扑的节点的重要性程度应该满足二八定律，互联网中应该会有一些比较重要的中心节点，一般来说度值越大，节点的重要性也越大，最后，应该有许多互联网的服务节点，这些节点不是内网节点，并且相关服务的节点连接可能会很密切。整体上，根据复杂网络理论，互联网的结构应该具有分形特征。

为了验证上述的猜想，本文借助 Gephi^[34] 对互联网拓扑可视化。为了节省篇幅，只截取相关的局部图形加以说明验证。

1. 互联网中的局域边缘结构

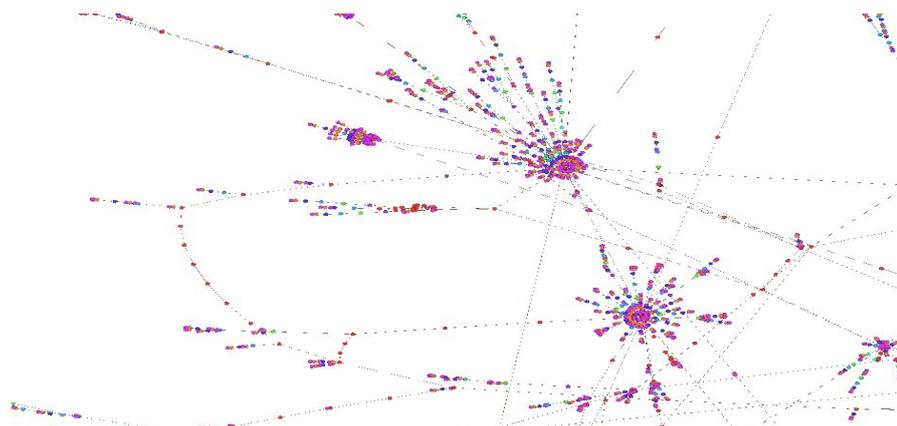


图 3.1 真实互联网局部结构
Fig .3.1 local structure of real Internet

纵观整个网络拓扑，其中包含了大大小小的许多类似上图的局部聚集结构，最小的结构只包含几个节点，最大的则包含数百个节点，在这些高度聚集的局部结构之间稀疏的连接一些中间节点，节点度分布很不均匀，有很多度值较小的节点，但也有度值较大的 Hub 节点，这些 Hub 节点实际上可能是连接外网的路由节点，从图 3.1 可看出，这些 Hub 节点都有一条或几条链路与其它局部区域相连接，节点度分布直观上出现了具有无尺度特征的长尾现象，这样的特征是典型的互联网局域网结构特征。

2. 互联网中的链路特征

在图 3.1 中，可以发现有很多链状的节点，这些节点连接不同的局域网，它们的度值基本上是 2，这些度值为 2 的节点完全是中转节点，这些节点具有路由来自其它网络的数据包功能。这些链状节点大都是 BGP 外部路由器，当然也有可能是一个大的局域网内部的路由链路。

3. 互联网高核节点聚类特征

为了便于分析，去掉了互联网中核数为 0 的节点，即去掉了互联网中的局部区域结构，图 3.2 便是 1 核网络集合视图。从图中看出，网络中的一些高核节点出现聚类现象，这种聚类节点应该是一个完成某种功能的功能模块，比如一些大型的服务器中心，至少这些聚类节点一定位于一个网络非常发达的地区。关于这种高核聚类特征的详细探讨会在章节 3.3.3 中给出。

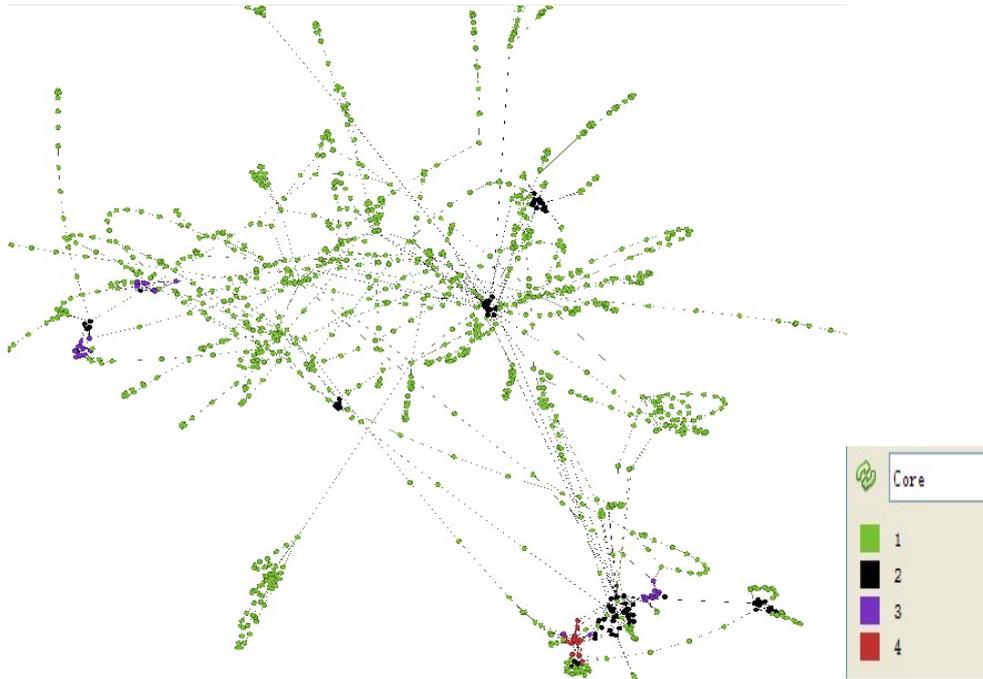


图 3.2 高核节点聚类特征

Fig.3.2 Cluster feature of ;

}

Integer 父容器size = 父节点集合.size();

Integer k = 0;

while(k 小于 父容器size)

{

 parentNode = 父节点集合.获取元素(k);

 childNode = 叶子节点集合.获取元素(k);

 k++;

 parentNode.获取邻居节点集合().remove(childNode);

 if (parentNode.层次到子节点映射.包含键值(fold_level))

 parentNode.层次到子节点映射.获取键值
 (fold_level).add(childNode);

 else

 {

 容器 叶子节点集合1 = new 容器();

 叶子节点集合1.add(childNode);

 parentNode.层次到子节点映射.添加键值对(fold_level

```
    ,叶子节点集1);  
        }  
    }  
}  
fold_level = fold_level+1 ;  
End program
```

4. 算法验证

经过层次折叠收缩算法的处理，网络中所有核数为 0 的边缘树状节点全部收缩到它们的根节点上，图 3.7 便是 0 核节点收缩后的图。在图 3.7 中，核数为 0 的节点以及全部被去掉，图中所有节点的度值都大于等于 2，这就说明本算法执行成功。

但是本算法的最终目的是发现局域网络树状结构组织，那么该如何还原这树状结构组织呢？在反复删除度值为 1 的叶子节点的过程中，本算法都在孜孜不倦的把这些删除的节点信息一一保存到与之相邻的父节点的数据结构中，即所有被删除的节点信息最终都保存在尚未被删除的核数大于零的那些根节点中，只要找到根节点，就可以还原此根节点下面的所有的子孙节点以及这些子孙节点的连接情况，那么该如何从图 3.7 中寻找那些根节点呢？在本算法中，每个节点都有一个 `childrenSum` 属性，在图 3.7 的 1 核网络拓扑中，只有属性 `childrenSum` 的值大于 0 的那些节点才是根节点。所以只需遍历图 3.7 的 1 核网络拓扑，找到这样的根节点，再根据根节点中的数据结构保存的子孙节点信息根据一定的递归算法即可复原整个网络拓扑结构，具体的递归算法不再累述，本文直接给出局部的树状结构图，在图 3.8 中，可以明显的看到，同属于一个根节点下的所有的核数为 0 的节点的颜色是一样的，这说明本文的复原算法也是有效的。

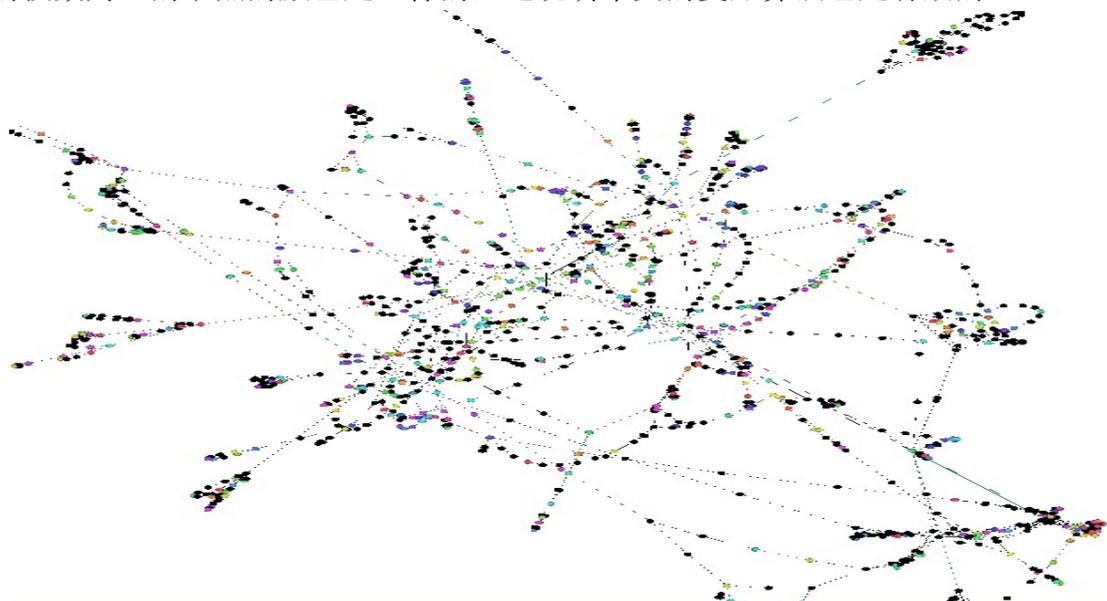


图 3.7 层次折叠之后的全局图

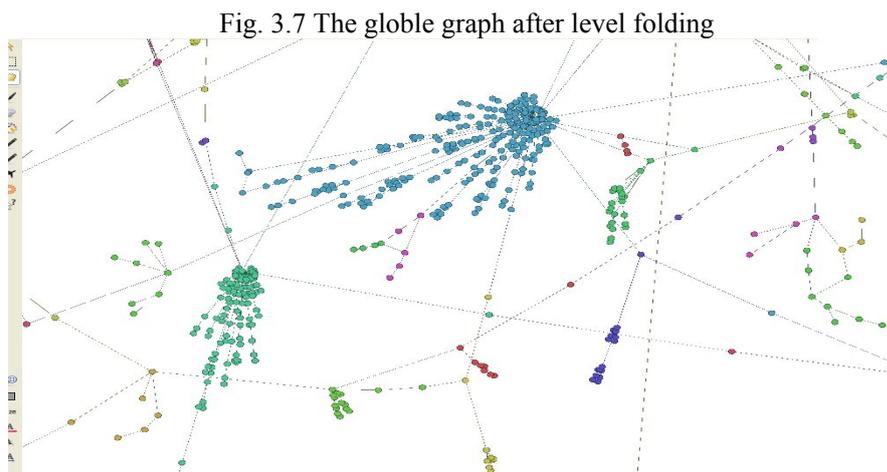


图 3.8 层次折叠算法局部效果图
Fig.3.8 The local result graph of level folding algorithm

为简化网络，便于观察，反复折叠收缩叶子节点到它的父节点中，最终得到一个 1 核网络拓扑，此拓扑中，节点度值大于等于 2。下面所有的研究将在这个简化的 1 核网络（图 3.7）上进行。

3.3.2 链状探测算法

1. 算法提出

在图 3.7 中，黑色节点表示尚未被层次折叠算法探测到的节点，即不属于局部区域树状结构，其它颜色节点表示局部区域树状结构特征的根节点。从图 3.7 中可看出，在剥去整个网络拓扑最外层核数为 0 的节点后，即剥去整个局域网结构后，整个网络拓扑变的相对简单，并且一个很明显的结构便是链状结构，图中有很多由尚未被探测算法探测到的黑色节点组成的链状结构。如果能把这些链状结构探测出来作为一个基本的网络组织结构，效果将会不错。如何探测这种结构呢？探测到的链状结构必须是一个只包含黑色节点的链状结构，如果包含了其它颜色节点，即使在此图中看起来是一个链状结构，可是在把最外层核数为 0 的节点还原到整个网络拓扑后，这些包含着其它颜色的链状结构将不再是一个真实的链状结构。下面将探讨一下探测此种结构的具体的算法设计。

2. 算法设计

- 1) 初始化 $community = 0$;
- 2) 依次遍历当前 1 核网络中的每一个节点 N 。如果遍历完毕则直接跳转到第 7 步。
- 3) 如果节点 N 是黑色节点，度值不大于 2，返回第 2 步 ， 否则执行下一步。
- 4) 依次遍历 N 的邻居节点集合 $N[i]$ ， $community++$ 。遍历完毕时返回第 2 步。
- 5) 当 N 的邻居节点 $N[i]$ 的度值不等于 2 或者此邻居节点是非黑色节点时，返回第 4 步 ； 否则 $N[i]munityNum = community$ 。执行下一步。

6), 返回第 5 步。

7) 算法结束。

3. 算法流程图

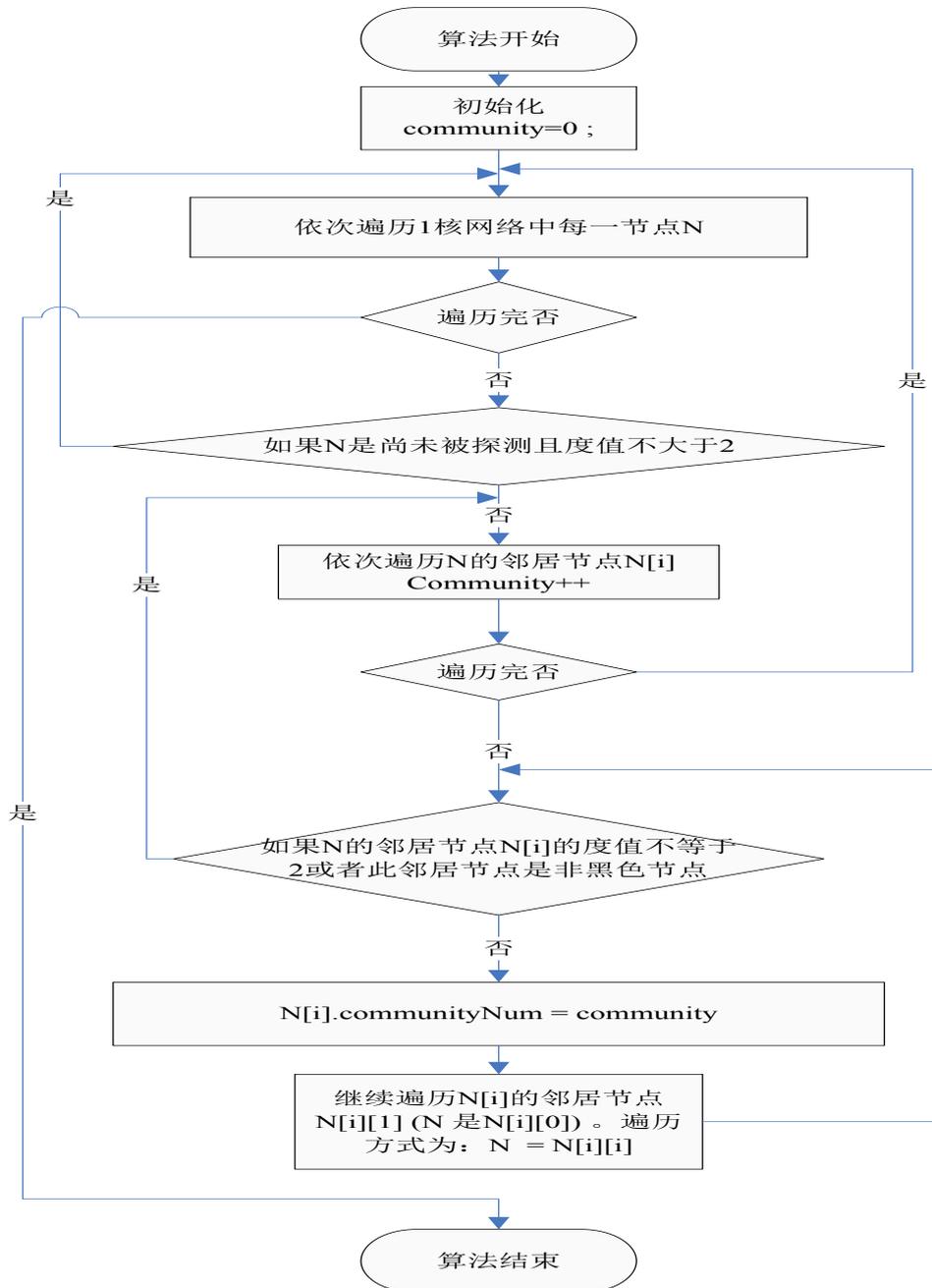


图 3.9 链状探测算法流程图

Fig. 3.9 The flow chart of Chain detecting

4. 算法验证

在图 3.10 的简化的 1 核网络拓扑中, 很明显的可以看到, 那些度值为 2 的节点中是颜色一致的, 说明本文的链状探测算法成功的探测到了网络中的链状结构组织。

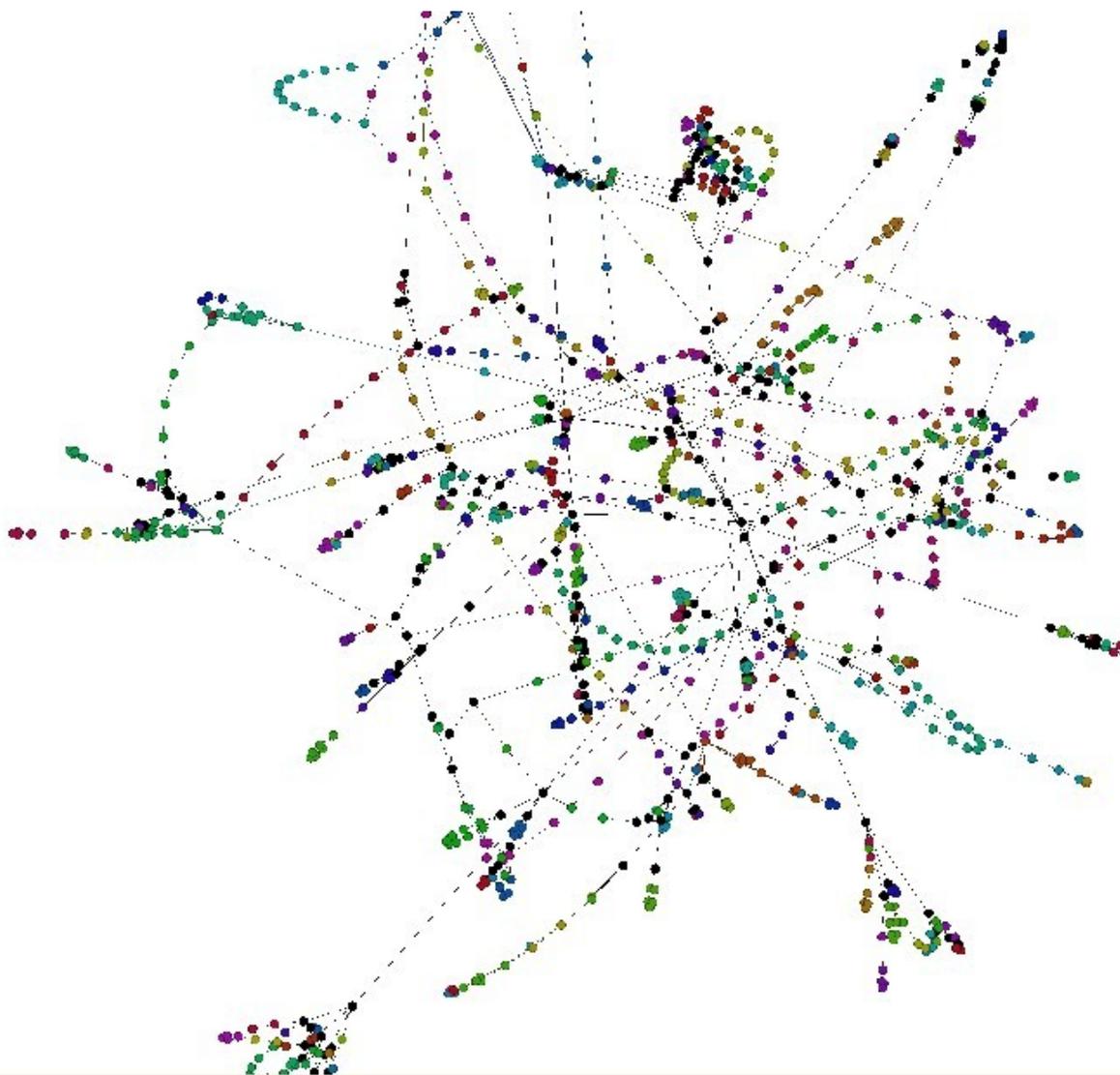


图 3.10 链状探测算法效果图

Fig. 3.10 The result graph of Chain detecting

3.3.3 高核数节点聚合探测算法

1. 算法提出

通过之前两个算法，可以探测到两种基本的互联网拓扑特征结构。下面介绍第三种拓扑结构的探测方法。

在观察本文所用的可视化的网络拓扑结构时，为了能更清晰的看到网络中的每个节点大致在网络中所处的位置，本文设计一个网络拓扑的 K 核解析算法，通过此算法，可以得到每个节点的核数，即在网络中的层次。图 3.11 是本文研究所用网络的基于 K 核解析算法的节点核数分布图。

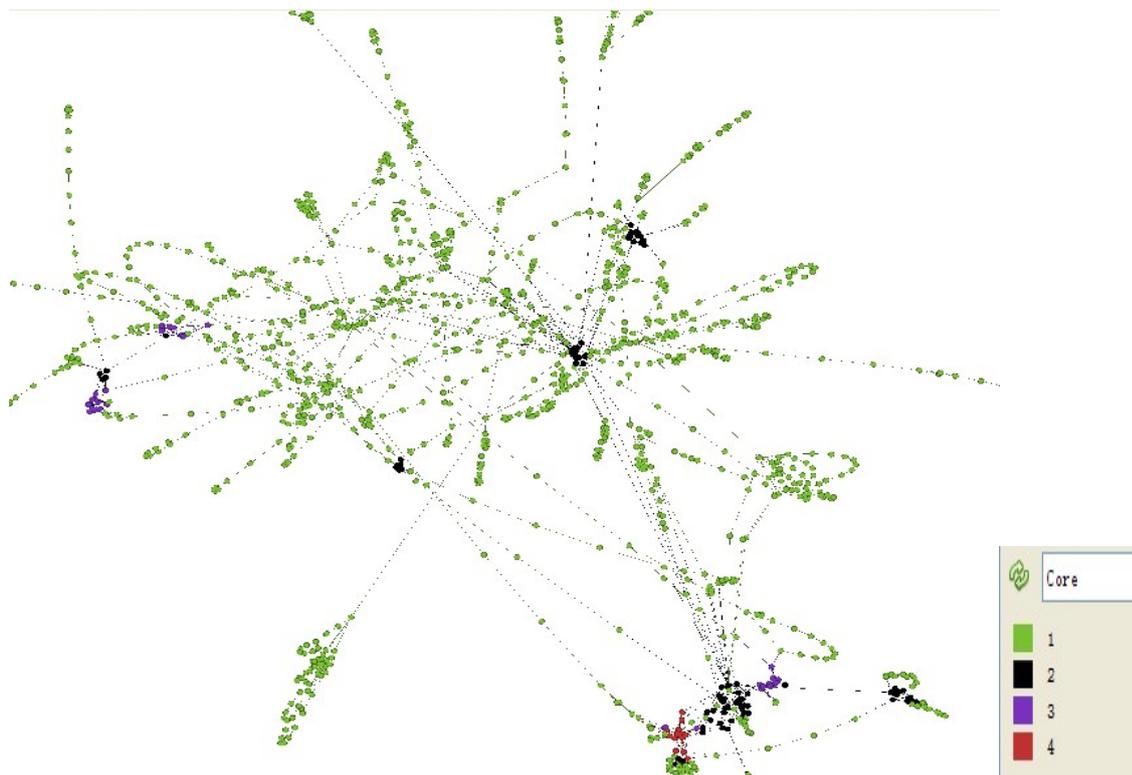


图 3.11 K 核解析
Fig.3.11 K-Core analysis

a) 网络拓扑中的节点层次分布

从上图（删除了 0 核节点）可以清晰的看到：简化的 1-核的网络拓扑结构的最高核数是 4；大部分节点分布在一个较低的核数为 1 的层次上；在一些局部区域，出现了高核节点的聚类现象，他们之间连接较为紧密，而它们与周围的低核节点连接先对稀疏，这是典型的高内聚低耦合现象，并且这些高核节点的核数要比周围的节点的核数大，以这些局部区域的最高核节点为中心向四周扩散时节点的核数越来越低。当一片局部区域越大时，一般来说，中心区域的节点核数越大，这种情况使在这片区域中更容易出现较小的类似于整体特征结构的分形特征结构。但也有一些区域的高核节点并没有出现低耦合现象。是什么导致了互联网的这种特征呢？

b) 原因分析

本文的理解：由于互联网的节点分布非均质特征、聚类特征、以及在不同规模尺度上的分型特征，导致互联网出现了某些特定核数的节点的聚集现象，这些节点的核数一般要比它们周围的节点的核数要高，即这些节点是这片区域中核数最高的节点，反应到物理层面上，可理解为，这些节点是这片区域的中心节点，所以可以把具有这类特征的节点，理解为一个特定的网络组织。其实在现实的社会网络中，也很容易去举出具有这种特征的例子。

例子: 在社会网络中, 有 4 个公司, 华为 (10 万员工)、58 同城 (5 千员工)、东大新业 (300 员工)、东科电力下属的研发中心 (30 员工)。本文故意拿出这四个具有不同规模尺度上的公司, 以便说明问题。也许这 4 个公司在业务上没有往来, 即没有直接连接的边, 但放眼整个社会网络, 他们肯定可以通过社会的一些其他团体或公司间接到达; 这 4 个公司均具有内部人员关系高度聚类, 内部与外部关系相对生疏的现象; 这 4 个公司在不同规模尺度上出现了类似的分形特征, 在结构上, 大致具备“领导核心、中层领导、低级领队、普通成员”这种类型的层次特征, 只是在较小的团体内部, 它的这种组织结构相对简单, 不全具备复杂团体的那种明显的层次特征结构的复杂的层次性。

比如华为员工的级别可以划分到 1-18 级, 而规模尺度较小的东科电力的员工只有 4 个级别: 普通员工、工程师, 副高、教高。在自然界中也会出现这种现象: 作为高级灵长类的人, 它的生命系统具有非常的组织结构, 大致有血管系统、皮肤系统、神经系统、听觉系统、视觉系统、消化系统、呼吸系统、泌尿系统等; 而进化相对低级的蚯蚓就不具备听觉系统、视觉系统、泌尿系统等。人与蚯蚓的这种生命系统都是整个大自然生物的生命之“道”的不同但又相似的分形系统。即使是在一个公司内部, 也会出现高核节点聚类的现象, 比如在华为内部, 作为华为最高行政的董事管理层可能每个人都相互熟悉, 可是对于处在华为最外层的任意的两个普通员工, 他们之间相识的概率估计会很小 (10 万人呢), 这就是典型的内层耦合性强外层相对分散的结构特征, 在华为的介于最内层和最外层之间也会分布许许多多的类似于整体结构特征的分形小群落。

c) 大道无形, 生育天地

中国有句俗语叫“麻雀虽小, 五脏俱全”, 人有五脏, 麻雀亦有五脏, 只是形体不同而已, 功能却何其相似! 于子云: “混沌自有相似处, 分形藏身于森木; 待到回归奇点时, 万事万物终化无”。天地统一于“道”, 生物和非生物都在某种神秘力量的驱使下延续着不同但又相似的演化, 外部的分形特征, 如草、农作物、树木, 内部的分形特征, 如白桦树、洋槐树、橡树, 再内部的分形特征, 如叶子、小枝条、较大的枝条、大枝条、整棵树, 再比如, 江流、河流、小溪, 再比如, 大海、湖泊、池塘、小水坑等等, 看似无序繁杂往复的大自然又是多么有序多么相似啊, 探索发现有这样一句话: Nature is a miraculous place where order exists within disorder.

d) 此类特征的网络组织结构的探测

那么如何寻找这些具有此类特征的网络组织结构呢? 首先需要有一个 K-核解析算法, 计算网络中每个节点的核数; 其次需要一个探测具有聚类特征的高核组织结构的算法。

2. 算法设计

➤ K-核解析算法

算法思想

- 1) 对于一个网络拓扑，初始化全局变量 $core=0$ 。
- 2) 依次遍历网络中每一个节点 N
- 3) 如果 N 度值小于或等于 $(core+1)$ (N 是叶子节点), 则 $N.core=core$ ，并且从此拓扑结构中删除节点 N ， N 的核数信息必须在其他数据结构下保留；
- 4) 如果网络中不存在节点度值小于等于 $(core+1)$ 的情况，则 $core++$ 。
- 5) 如果网络中所有节点都被删去，则算法结束；否则返回第 2 步。

以上只是本算法的简单思想介绍，限于篇幅，具体算法流程图和实现将不再给出，直接给出本算法的验证。

算法验证

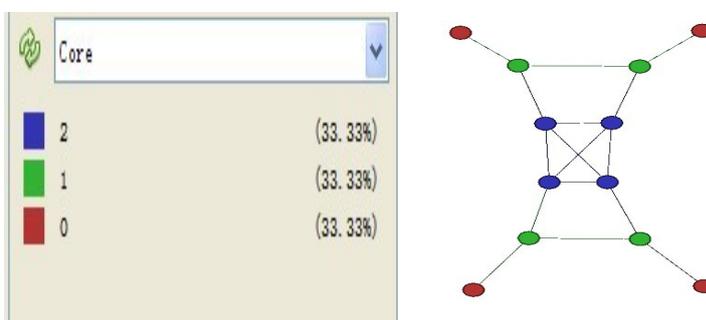


图 3.12 K 核解析验证图

Fig.3.12 verification graph of K-Core analysis

从 K-核解析算法验证图来看，本算法是正确的，能准确解析网络中每个节点的核数。

➤ 聚类特征的高核组织结构探测算法

算法思想

以 K-核解析算法为基础，解析出每个节点的核数。高核聚类特征应该至少从核数为 2 的节点开始的，因为 0 核是网络的最外层，现在需要探测的是局部网络最内层的聚类特征。在探测此类结构时，需要遍历 1 核网络中的每一个节点 N ，记录此节点 N 的核数 $core$ ，然后在顺着节点 N 寻找所有核数为 $core$ 且能通过核数为 $core$ 的节点与 N 联通的节点，计算这些节点集合的总度值 TD ，然后在计算这些节点与其它核节点连接的总边数 TE ，当 TE/TD 的比值小于一定值 V 时，可认为这些节点集合具有高核聚类特征，值 V 可根据具体情况设定。

算法设计

- 1) 初始化 1 核集合每一个节点 $N.visited=false$,全局变量 $community=0$;
- 2) 遍历 1 核集合网络拓扑中每一个节点 N ，遍历完毕时，跳转第 7 步。如果 $N.visited=true$ 或者 $N.core \leq 1$ 则重复执行这一步，否则执行下一步。
- 3) 设置 $N.visited=true$ ， $community++$ ， $communityNum=community$ 。

- 4) 利用广度优先遍历策略, 依次遍历 N 的邻居节点 $N[i]$, 如果 $N[i].core == N.core$, 则 $N[i].community = community$, $ajacentNodeVector.add(N[i])$ 。
 - 5) 依次遍历 $ajacentNodeVector$ 中的每一个元素 $E[i]$, 对于每一个元素 $E[i]$, 设置 $E[i].visited = true$, 依次遍历 $E[i]$ 的每一个邻居节点 $J[i]$, 如果 $J[i].core == N.core$, 则 $J[i].community = community$, $nextAjacentNodeVector.add(J[i])$ 。
 - 6) 如果 $nextAjacentNodeVector.size() == 0$, 返回第二步; 否则, 返回第 5 步。
3. 下面给算法流程图。

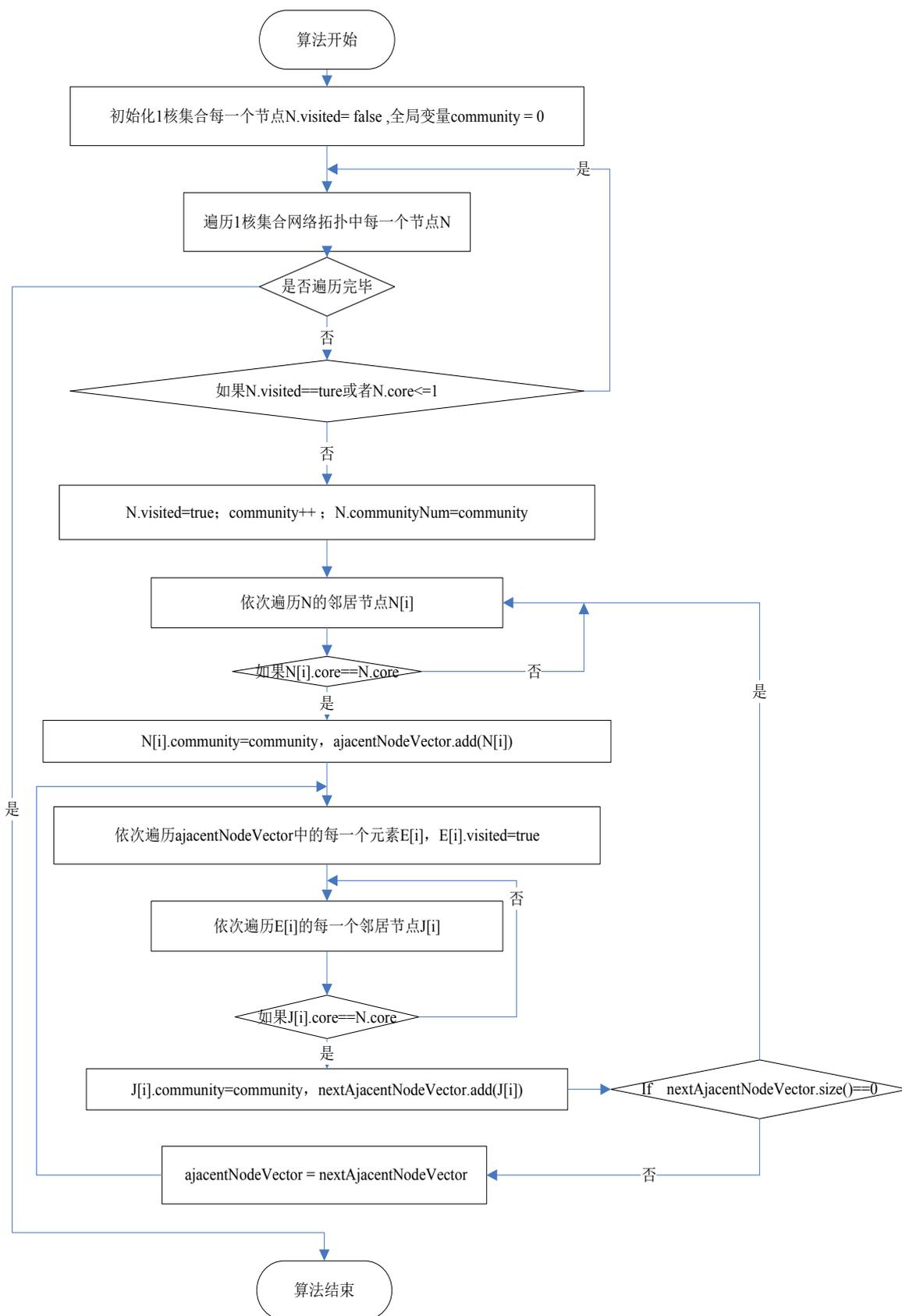


图 3.13 高核聚类探测算法流程图

Fig.3.13 The flow chart of 和 Girvan 提出一种衡量网络社区结构的指标 Q

值，它是基于 association mixing^[35]来定义的，模块度公式如下。

$$Q = \sum_i (e_{ii} - a_i^2) = Tr e - \|e^2\|$$

e_{ij} 表示网络中子团 i 和子团 j 之间连接的边数在网络所有边中所占的比例。

$a_i = \sum_j e_{ij}$ 表示与子团 C_i 相连的边在所有的边中所占的比例。

Q 值的上限为 1，值越接近于 1，说明网络社团结构就越明显。

3.4.2 小社团合并算法

1) 算法思想

在把小粒度的网络组织合并成较大粒度的网络模块时，该按照什么标准呢？如果仅仅是把耦合性较大的网络组织合并在一起，那么耦合性的强弱又该如何界定呢？

Newman 和 Girvan 提出一种衡量网络社区结构的指标：模块度。尽管把模块度当做一种衡量社团结构的标准还存在一定争论，但目前国际上研究社团发现的人基本上都在采用这个指标，模块度已经是事实上的标准，就像 TCPIP 协议栈一样。本文将采用模块度作为衡量社团结构的一种标准。

所以在合并小粒度网络组织时，需要模块度的优化。寻找模块度的全局优化解在数据量很大的情况下是一个 NP 难题，在时间复杂度上和空间复杂度上都比较大。在借鉴传统的优化算法的基础上，本文将采用基于局部区域优化的贪婪算法思想，通过每一次的局部最佳选择以此来作为全局最优的近似解。

一般情况下，要选出最优量度标准并不是一件容易的事，但对某问题能选择出最优量度标准后，用贪婪算法求解则特别有效。最优解可以通过一系列局部最优的选择即贪婪选择来达到，根据当前状态做出在当前看来是最好的选择，即局部最优解选择，然后再去解这个选择后产生的相应的子问题。每做一次贪婪选择就将所求问题简化为一个规模更小的子问题，最终可得到问题的一个整体最优解。

2) 算法设计

模块度优化的局部优化问题虽然大致思路比较简单，但在具体的本问题解决过程中是异常麻烦，为了快速给出思路，在此只做简单大概的陈述，详细的算法实现会在后章节给出。

假设网络中现存 n 个网络组织， C_i 表示第 i 个网络组织， i 从 0 开始。

1. 依次遍历网络中的第 i 个网络组织 C_i ，对于网络组织 C_i 的第 j 个邻居网络组织 C_j ，计算出 C_i 与 C_j 合并后的模块度 M_{ij} 。

2. 找出合并后最大的模块度

$$M_{i-\max} = \max(M_{ij} | j \geq 0 \ \&\& \ j \leq C_i \text{的邻居网络组织总数}) = M_{ik} ;$$

if $M_{i-\max} > 0$, then 合并 C_i 与 C_k 为一个新的网络组织。

以模块度局部最优化为方法，在合并完整个网络中耦合性较高的网络组织后，网络的社团分布如图 3.20。在某一局部区域，相同颜色的节点为一社团，直观上，网络的社团结构比较明显，在与 3.2.5 节中的没有经过模块度优化的网络结构图相比，下图的社团分布则没有出现零碎的网络组织现象。由此，可认为此合并算法是有效的。

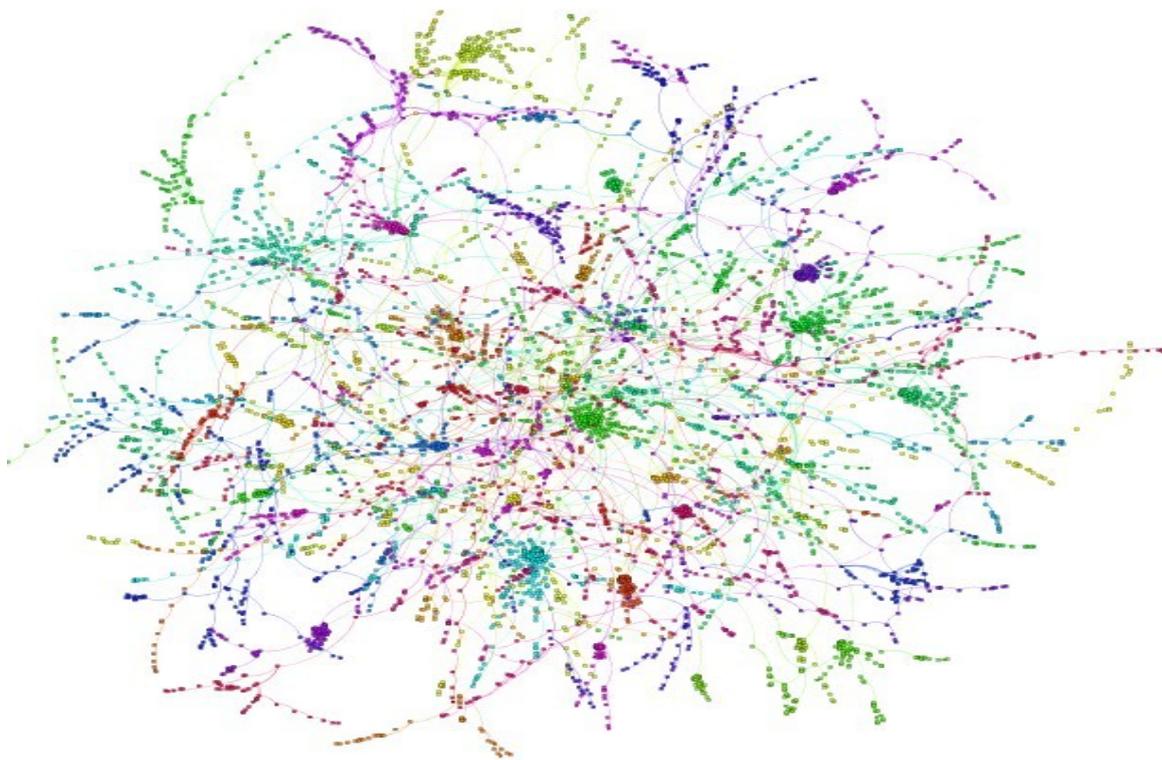


图 3.20 最终的社团划分分布图
Fig.3.20 The final graph of community detecting

4) 化简网络

如果能把网络中的社团表示为抽象节点，社团中的节点规模表示为抽象节点的大小，社团内部的边的稠密程度用抽象节点的颜色深浅表示，抽象节点的关联用带有权重的边表示，这样，网络的简化如图 3.21。根据这幅图，网络中的社团分布情况以及社团内部情况可以很清晰的展示出来。

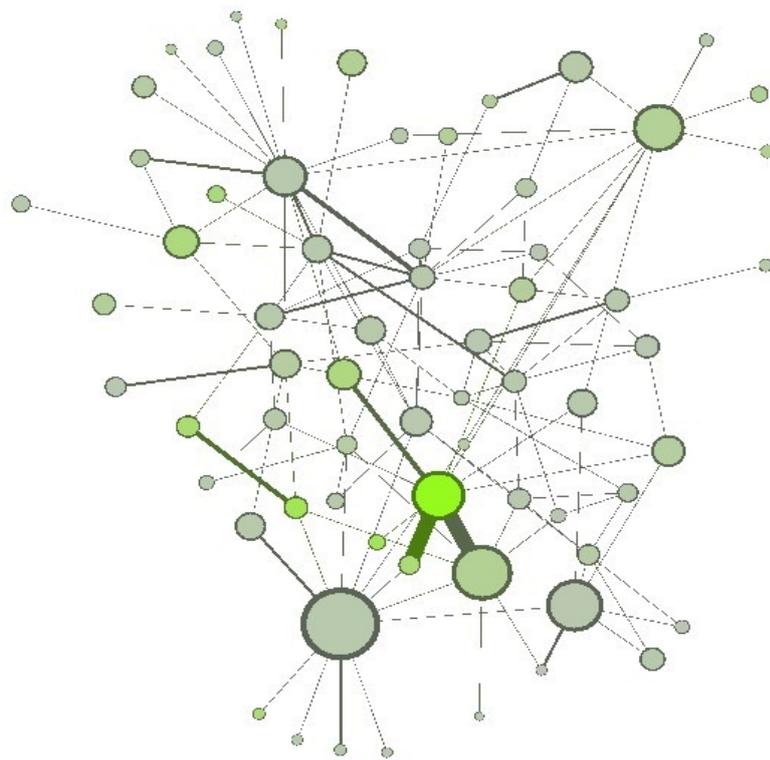


图 3.21 社团节点抽象图
Fig.3.21 the abstract graph of community node

3.5 社团划分比较

本社团发现算法是针对互联网基本拓扑特征的一组算法，不是单一的算法，总体上，称之为 105 算法。对于本文选取的网络拓扑 studiedNet（节点数 6782，边数 7542），分别用 105 算法和 fast unfolding of communities in large networks^[36](简称 FUOCILN)算法处理此网络，下面是社团划分的结果。对于这两种算法，由于每一次处理网络，模块度的值都是有变化，所以本文选取的是三次处理的最高的模块度值。

表 3.1 社团发现结果对比
Table3.1 result comparison of community detecting

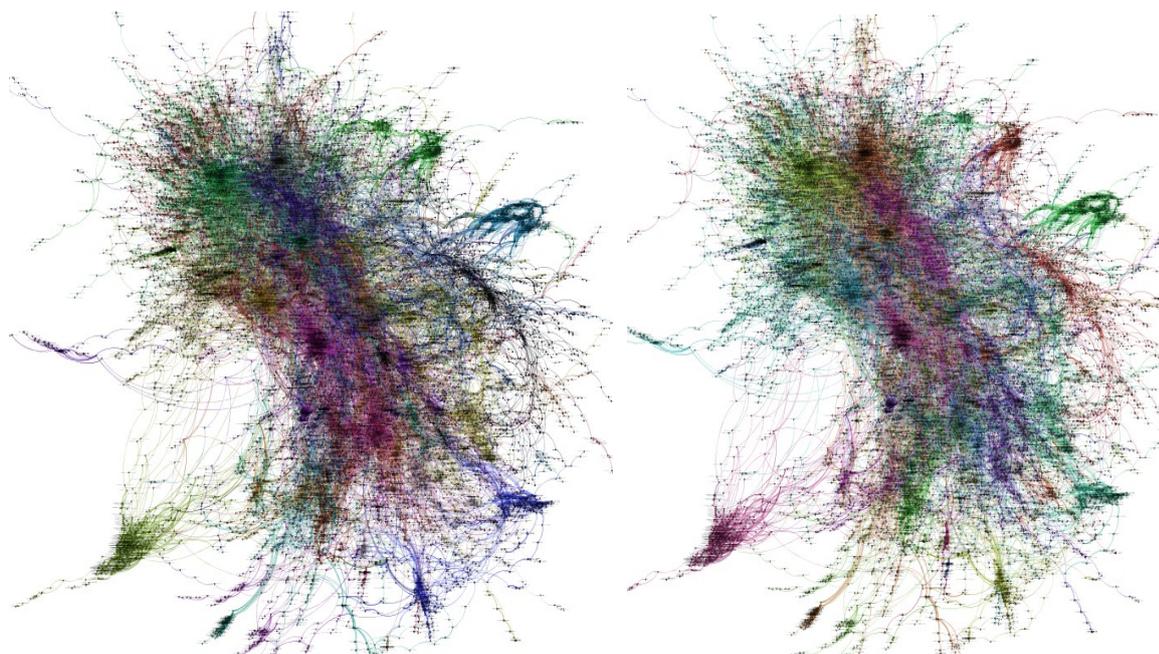
社团发现算法	105	FUOCILN
模块度	0.957	0.956
社团数目	65	78

在上表的划分结果中，令人惊喜的是，105 算法发现的社团结构在模块度上与传统的经典的 FUOCILN 算法不相上下，那么在社团的划分结果上有什么区别吗？

两种算法都是基于模块度优化的，但是具有的社团发现方式完全不一样。105 算法的大致思路是首先探测网络的基本特征组织，然后在用模块度优化的思想合并这些网络组织。而 FUOCILN 直接把每一个节点当做一个初始社团，然后直接用模块度优化的思想划分社团。所以，105 算法发现的社团结果是一个多粒度的，即在一个社团内部，有可能会有局域边缘的树状结构、链状结构、高核聚类结构、中心节点结构这四种结构中的一个或几个，而且 105 算法可以很轻易的来解析这种社团内部基本的互联网的特征结构。FUOCILN 则是一种单纯追求模块度的算法，根本不具备发现社团内部的互联网的基本特征组织结构的能力。105 算法具有这种发现社团结果多粒度的原因在于 105 算法是一组算法，它所实现的复杂的功能远非 FUOCILN 算法可比。

3.6 对于全球 IPV6 拓扑的划分比较

上一节对 105 算法与 FUOCILN 算法在社团发现结果上做了比较：在以模块度为衡量标准的基础上，由于采用了模块度的贪婪优化方式，105 算法的社团划分结果与 FUOCILN 不相上下；但在更细致的互联网拓扑结构上，由于采用了一组探测算法，105 算法能够探测到社团内部的一些典型的多粒度的互联网特有的拓扑特征，而 FUOCILN 则做不到这样的划分。由于全球主干 IPV6 路由节点大约有 2 万~2.3 万个之间，而之前用于研究的 studiedNet 网络只是全球总的 IPV6 的一部分，为了更加科学的评价本文提出的社团发现算法，本文采集了 CAIDA 全球 26 个探测节点在 2011 年 9 月这一整月探测的网络拓扑数据集合，然后处理合并成一个完整的没有重复节点和重复边的 IPV6 网络拓扑，称之为 wholeNet（节点数 20207，边数 47981），此时，可以把 wholeNet 看作是一个全球的 IPV6 拓扑。下面比较一下 105 算法和 FUOCILN 算法在 wholeNet 上的划分比较。



(a) 105 社团发现图

(b) FUOCILN 社团发现图

图 3.22 两种算法划分结果的视觉比较

Fig.3.22 visual comparison of two algorithm detecting

图 3.22 (a)、3.22 (b) 分别是 105 算法和 FUOCILN 社团发现在 forceAtlas2 上的可视化效果图，根据上图，在一个区域内相同颜色的被划分为同一社团，在不同区域内相同颜色的为不同社团。由于 forceAtlas2 没有区别两个相邻社团颜色是否一致，导致一个相同颜色区域内有可能包含不止一个社团，但整体上，forceAtlas2 可视化算法还是能反应社团划分的效果。直观上看，两种社团发现效果不相上下，在很多区域，有着很强的相似性。FUOCILN 算法是一个非常经典的社团发现算法，这也直接印证了本文提出的 105 算法的优秀程度。但为了更加科学的比较本文提出的 105 算法与 FUOCILN 算法，本节将从两个方面加以探讨：模块度、多粒度。

1. 模块度

模块度是评价社团发现算法的一个主流标准，很多传统的社团发现算法也都是用模块度来评价算法的优秀性。所以，本节也用模块度来衡量一下 105 算法。下面是两种社团发现算法对同一网络 wholeNet 的社团发现结果。

表 3.2 社团发现结果对比
Table3.2 result comparison of community detecting

社团发现算法	105	FUOCILN
模块度	0.838	0.845
社团数目	45	54

对于全球的 IPV6 网络拓扑的社团发现，很遗憾，本文提出的 105 算法以微弱劣势败于 FUOCILN，这也在意料之中，原因在接下来给出。

2. 社团内部的多粒度性

由于 105 算法是一组专门针对 IPv6 互联网拓扑的基本拓扑特征设计的，在模块度局部优化的过程中，是以 IPV6 互联网基本拓扑结构为粒度单位的；FUOCILN 算法采用的同样是模块度局部优化算法，可是 FUOCILN 算法是以每个基本节点为粒度单位的，FUOCILN 算法的基本粒度要比 105 算法细化很多；可以基本断定，在采用模块度局部优化的算法中，FUOCILN 算法在模块度方面是很难被超越的。天之道，损有余而补不足。正由于这个原因，105 算法才有自己的其它的优秀特点：105 算法发现的社团内部还可以继续划为多种互联网基本拓扑结构，比如链状、高核聚类状、局部树状、中心节点状等。这些基本的结构是 FUOCILN 算法发现不了的。105 算法的数据结构中保留了这些结构属性，只是由于 forceAtlas2 可视化工具的局限性，这些特征并不能直观的展现出来。这正所谓：“东边日出西边雨，道是无晴却有晴”。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

<https://d.book118.com/675110133104011131>