

Internet 宏观拓扑结构的可视化

张君 赵海 杨波

(东北大学信息科学与工程学院 沈阳 110819)

摘要 针对 Internet 宏观拓扑结构的复杂性,研究了 AS 级 Internet 宏观拓扑结构的可视化问题。首先利用常用的几种布点策略,对 Internet 宏观拓扑结构进行了可视化设计。结果表明,这些策略均不能展现 Internet 宏观拓扑结构的结构特性,因此并不适用于 Internet 的可视化。为此设计了一种新的方法,即中心层次布点策略。结果表明,该种策略能很好地体现 Internet 宏观拓扑结构的幂率性质和层次性质,对直观地观察 Internet 宏观拓扑结构提供了有力的手段。

关键词 复杂网络, Internet 拓扑, 可视化, 度, 核

中图分类号 TP301.5, TP393 **文献标识码** A

Visualization of Internet Macroscopic Topology Structure

ZHANG Jun ZHAO Hai YANG Bo

(College of Information Science & Engineering, Northeastern University, Shenyang 110819, China)

Abstract Aiming at the complexity of Internet macroscopic topology structure, we studied the visualization of Internet topology structure at AS level. We designed the algorithms on visualization of Internet topology using several popular strategies of distribution notes. The results showed that these strategies could not exhibit the structural features of Internet topology, and did not apply to the visualization of Internet. So we designed a new algorithm, which is the strategy of center-level distribution nodes. The results showed that the strategy well illustrated the power-law feature and hierarchical feature of Internet topology. It provides a powerful means for directly observing on Internet microscopic topology structure.

Keywords Complex network, Internet topology, Visualization, Degree, Core

网络作为系统的抽象,虽然每一个系统的网络都有其自身的特殊性质,有其紧密联系在一起的独特现象及自身的演化机制,但是由于都可以使用网络分析的方法,因此有其共性。研究网络的几何性质、网络的形成机制、网络演化的统计规律以及网络的结构稳定性,并把它的拓扑结构与具体系统结合起来是复杂网络研究的中心内容^[1-3]。Internet 作为一个典型的复杂网络实例,对其宏观拓扑结构的研究^[4-7],可以为其它类型复杂网络的相关研究提供一个有价值的原型基础,因此具有重要的指导作用。

目前,关于 Internet 宏观拓扑结构的研究主要分为如下 3 个方面:

- (1) 如何获得一份完整而准确的 Internet 拓扑数据;
- (2) 如何对 Internet 网络拓扑特征进行分析并建立模型;
- (3) 如何使用模型构造一幅类似于 Internet 的拓扑图。

这 3 个问题分别对应于 3 个研究方向,即拓扑结构测量、拓扑特征发现与建模以及拓扑可视化。

本文研究了 Internet 宏观拓扑结构的可视化问题。

1 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)提供的 AS 级互联网探测数据。CAIDA 是互联网拓扑分析研究领域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 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。Ark 分布范围极广,其探测范围的覆盖程度就目前而言在主动探测项目中占有优势,能够满足对互联网进行较为全面、准确的探测分析的需求;此外 CAIDA 探测源点有完全自主的所属权与控制权,可以不间断地持续测量而不受影响,其探测数据量十分可观,有效性可以得到很好的保障。作为 CAIDA 的合作者,位于东北大学的 Neu 节点不但可以自身采集数据,还可以获得由 CAIDA 提供的众多探测点的海量探测数据。

2 选点

在构造可视化拓扑图时,由于拓扑图中的点、线数量众多且分布位置不确定,因此在绘制过程中会出现点线重叠、主次不分的“麻团”现象。为了得到一幅比较清晰的拓扑图,在布点之前正确地选点是非常重要的。

本文受国家自然科学基金项目(60973022)资助。

张君(1967-),女,博士,讲师,主要研究方向为复杂网络, E-mail: zhangjun1@ise.neu.edu.cn; 赵海(1959-),男,教授,博士生导师,主要研究方向为复杂网络、普适计算、嵌入式系统及数据融合等。

为保持良好的视觉效果,在设计中可根据需要设置所显示节点的比例,依照这一比例按照节点在拓扑图中的重要程度进行选点。在本设计中通过节点的特征量来衡量节点在互联网中的重要程度,分别用节点度值大小和核所在层次深度两个特征来描述,选取节点时选择度值大的节点或核层次深的节点。

具体选点过程如下:

Input: 显示节点比例

Output: 所选节点和所选边

(1) 根据显示比例与节点数计算显示节点数量;

(2) 根据节点分布计算显示节点最小关键值 Keymin(度值或核值);

(3) 根据 Keymin 得到所显示节点的点集 P ;

(4) 根据点集 P 得到所选边集 V 。

3 布点

在 CAIDA 提供的 AS 级探测数据中并不包含节点的地理位置信息或相对位置信息,故此并不能按照物理位置进行互联网拓扑构建,这就需要根据数据的特征自行布点,构建可视化互联网拓扑。由于选用的自动布点策略不同,因此同样的数据所达到的显示效果完全不同,所体现的网络形态会在直观上产生巨大差异。

目前比较常用的布点策略有随机布点,环状布点以及 GEM 布点等。

3.1 随机布点

随机布点^[8]是布点中最简单的布点方式,布点后点比较均匀地分布在显示区域中,一般不会出现过于集中或空白的情况,如图 1 所示。其优点是速度快,算法简单,对布点量较大、规律不明显的图比较合适。由于一般随机布点图比较乱,相关度较高的点分散较远而无关点分布较近的情况可能发生,像社团规律较强的图极可能看不到规律,因此该种布点策略并不适合 Internet 拓扑。

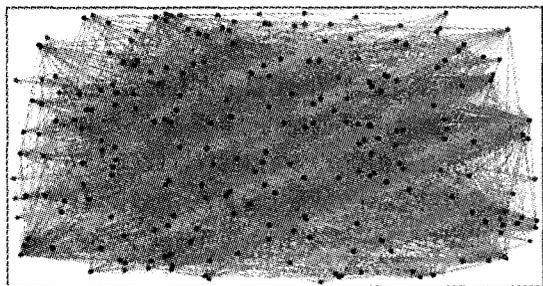


图 1 随机布点

3.2 环状布点

环状布点是一种比较常用的布点方法。该方法使点均匀分布在一个圆环上,环内是边。该方法保证了点不会和与之无关的边重叠,如图 2 所示。

环状布点特别适用于规则网络布点,比如全耦合网络或最近邻耦合网络。在一般布点中,由于环状布点的最大特点是点不会和与之无关的边相交,边的交际概率很大,故此环状布点最优状态是用于边较少、较集中的拓扑,因此不适用于 Internet 拓扑。

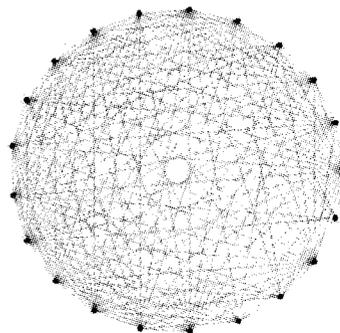


图 2 环状布点

3.3 GEM 布点

GEM 是 Graph Embedder 的简称,算法由 Arne Frick, Andreas Ludwig 和 Heiko Mehltau 在 A Fast Adaptive Layout Algorithm for Undirected Graphs 一文中提出,该算法思想来源于冷却系统^[9]。GEM 算法定义:每个节点都拥有一个初始温度,整个系统向温度降低的趋势运行,其局部温度取决于它的旧有温度和运动趋势。如果算法判定一个节点运动方向是朝着它的最终位置方向运动时,该点的局部温度降低,反之则局部温度升高。该系统的全局温度是所有节点的局部温度的平均。全局温度表示了整个拓扑结构的稳定程度,当全局温度低于设定的阈值时,算法认为拓扑结构达到稳定状态,算法结束。当拓扑中包含点数很多、节点之间关系较复杂时,要求结构稳定并非简单的事情,故此算法设定单点调整的最大次数,当算法调整点量超过这个次数限制时,即使系统并未完全稳定,算法依然停止。

GEM 布点的优势在于它可以将度值较高的节点集中在图中央,度较低的节点环绕在与之联系较为紧密的高度节点四周,整幅图呈现散射状。图 3 是典型 GEM 算法的布局效果。

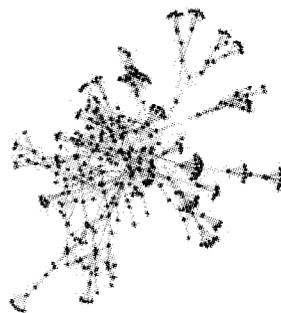


图 3 典型 GEM 布点

可见,GEM 最适合的是树状图形布点,对于联系极为紧密的互联网拓扑是否适用还有待进一步讨论。

综合以上几种布点方式,可见现有的布点策略并不能很好地体现互联网的拓扑结构,为此作者针对互联网的结构特性设计了一种新的布点方法,即中心层次布点。

3.4 中心层次布点

由于点是按照关键值(度大小或核层次)抽取的,因此考虑在图中将被选中的点中关键值较大的放在图中间,按照辐射状向四周扩展图形,关键值依次减小。最后使图形成圆形。

在实现过程中,考虑到互联网中节点关键值,特别是度值

在点显示比例较大的情况下会出现较大的差距,特别是大度值端会出现明显断层,如果单纯按照绝对值划分层次,则图形会出现中间只有少量点稀疏分布,并有明显空白区,外面大量点集中的情况。故在设计时,层次分布取关键值的自然对数,这样关键值同数量级的点在一个层次。每层内点随机分布。

层次布点节点坐标计算的流程为:

Input: 需要产生坐标的点集合

Output: 所有点的坐标

- (1) 计算显示区中心点(centX, centY);
- (2) 计算最大半径 maxR;
- (3) 计算每层宽度 width;
- (4) 用当前时间设置随机数种子;
- (5) while(点集未到末尾){
- (6) 根据该点 p 的关键值计算所处层次半径范围[r, R);
- (7) 生成半径范围内的固定值作为点 p 的半径 R_p ;
- (8) 随机生成点 p 的 X 坐标 x, 满足 $0 \leq |x| < R$;
- (9) 计算点 p 的 Y 坐标 y, 满足 $x^2 + y^2 = R_p^2$;
- (10) }
- (11) 结束。

该算法在计算节点坐标时只对节点集遍历一遍,在节点数为 n 时,时间复杂度为 $O(n)$ 。

图 4 描述了以节点度值作为关键值的中心层次布点效果,图 4(a)~图 4(d)为按照度值选点,比例分别为 0.2%、1%、5%和 50%的效果,显示数据选用的是 2009 年 1 月的 AS 级数据。从中可以发现,当显示节点数量较小时,中心层次效果并不明显;随着节点数量的上升,层次效果越来越好,越靠近外层,点越密集。至图 4(d)时无论从点的密集程度还是层次色彩,均可以明显看出层次变化。

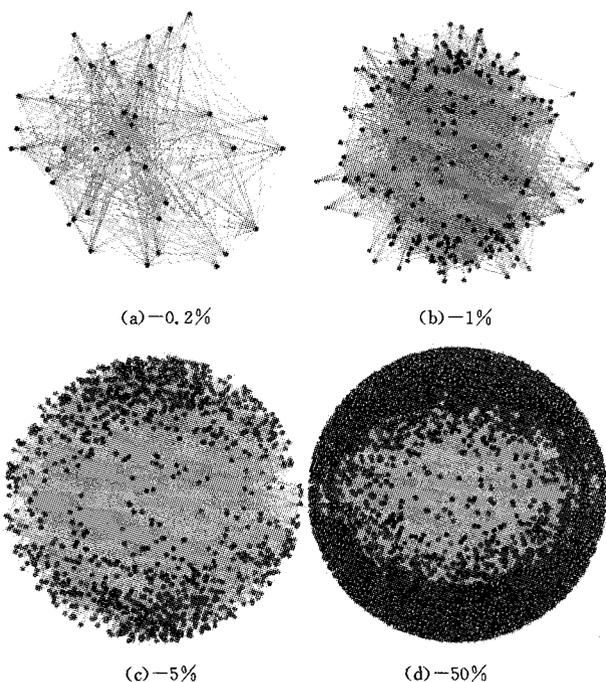


图 4 以度值为关键值的中心层次布点

通过实验发现,在点达到一定数量后,由于所增加点的度值集中在低度值范围,因此随着显示节点比例上升,只是外层

节点更加密集,对整体显示影响极小。图 5 是 100%节点显示效果,和 50%的几乎没有区别。

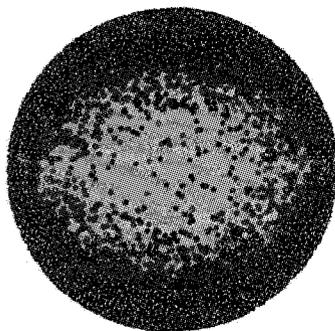


图 5 100%节点显示

以节点核值作为关键值选点并布点的算法与以节点度值作为关键值的算法相同,图 6 描述了显示 50%节点的效果。可见,节点核差距并没有度那样悬殊,由于层次设计的原因,图 6 的层次效果更加明显。

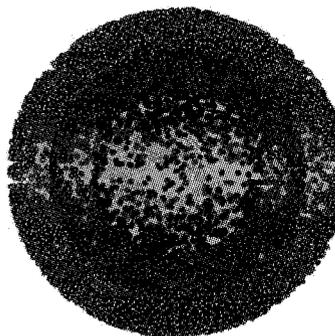


图 6 以核值为关键值的中心层次布点

结束语 从复杂网络的角度研究互联网的拓扑特征,已经成为当今互联网研究的一个重要分支,其研究成果对指导其它类型复杂网络的相关研究具有重要的意义。互联网的可视化是互联网拓扑研究的一个重要内容。

本文研究了 Internet 宏观拓扑结构的可视化问题。通过比较几种常用的布点策略,发现随机布点对布点量较大、规律不明显的图比较合适;环状布点适用于边较少、较集中的拓扑;GEM 适合于树状图形布点。Internet 拓扑满足幂率性质,是一种典型的复杂网络。因此在 Internet 的可视化方面,这几种布点策略均不适用,为此设计了一种新的布点策略,即中心层次布点,该种布点策略利用节点的度值或核值作为关键值进行选点,并且按照关键值的大小由内至外设置节点的位置,突出了 Internet 拓扑结构的幂率性质和层次性质,为大规模复杂网络的整体分析及其结构研究提供了一种方便的方法。

参考文献

- [1] Zhang Xiao-song, Chen Ting, Chen Rui-dong, et al. Complex network modeling with constant capacity restriction based on BA model[C]// Proceedings of the 2009 International Symposium on Computer Network and Multimedia Technology. 2009: 1-4
- [2] Tang Yang, Wang Zi-dong, Fang Jian-an. Pinning control of fractional-order weighted complex networks[J]. Chaos, 2009, 19

- [3] Cajueiro D O. Optimal navigation in complex networks [J]. *Physical Review E*,2009,79(4):1-7
- [4] Vilhar A,Novak R. Policy relationship annotations of predefined AS-level topologies [J]. *Computer Networks*, 2008, 52 (15): 2859-2871
- [5] Wang Jian,Liu Yan-heng,Zhang Cheng, et al. Model for router-level Internet topology based on attribute evolution[J]. *IEEE Communications Letters*,2009,13(6):447-449
- [6] Haddadi H,Uhlig S,Moore A, et al. Modeling Internet topology dynamics[J]. *Computer Communication Review*, 2008, 38 (2): 65-68
- [7] 张君,赵海,康敏. Skitter 与 Ark 探测架构下 AS 级 Internet 拓扑分析[J]. *计算机科学*,2010,37(11):38-40
- [8] Fisk C J, Caskey D L, West L E. ACCEL: Automated circuit card etching layout [J]. *Proceedings of the IEEE*,1967,55(11): 1971-1982
- [9] Frick A, Ludwig A, Mehldau H. A fast adaptive layout algorithm for undirected graphs [C]//*Proceedings of the DIMACS International Workshop on Graph Drawing*. 1994,10:388-403
-
- (上接第 229 页)
- [10] Yin Hao,Liu Xue-ning,Zhan Tong-yu, et al. Design and deployment of a hybrid CDN-P2P system for live video streaming; experiences with livesky[C]//*Proc. Of MM '09*. 2009:25-34
- [11] Armbrust M,Fox A,Griffith R, et al. A View of Cloud Computing[J]. *Communications of the ACM*,2010,53(5):862-876
- [12] Drezner Z, Hamacher H W. Facility location: applications and theory [M]. Springer,2004:132-141
- [13] Qiu LL,Padmanabhan V,Voelker G. On the placement of Web server replicas[C]//*Proc. INFOCOM'01*. 2001:1587-1596
- [14] Schobel A,Scholz D. The Big Cube Small Cube Solution Method for Multidimensional Facility Location Problems[J]. *Computers & Operations Research*,2010,37:115-122
- [15] Charikar M, Guha S. Improved Combinatorial Algorithms for the Facility Location and K-Median Problems[C]//*Proc. of the 40th Annual IEEE Conference on Foundations of Computer Science*. 1999
- [16] Bateni M,Hajiaghayi M. Assignment problem in content distribution networks; unsplitable hard-capacitated facility location [C]//*Proc. of SODA*. 2009:1-16
- [17] Abravaya S,Berend D. Multi-dimensional dynamic facility location and fast computation at query points[J]. *Information Processing Letters*,2009:386-390
- [18] BolorArabani A, Farahani R Z. Dynamic Facility Location: An Overview of Classifications, Models, Solutions and Applications [EB/OL]. <http://www.alirezabolori.com/wp-content/uploads/2010/12/Dynamic-Facility-Location-An-Overview-of-Classifications-Models-Solutions-and-Applications.pdf>,2009
- [19] Wendell P,Jiang J W,Freedman M J, et al. Donar: decentralized server selection for cloud services[C]//*Proc. of SIGCOMM'10*. 2010:231-242
- [20] Mohammad A F,Alexander L K,Amin V. A scalable, commodity data center network architecture[C]//*Proc. of SIGCOMM'08*. 2008:63-74
- [21] Yin Hao, Liu Xue-ning, Zhan Tong-yu, et al. LiveSky: Enhancing CDN with P2P[J]. *ACM Transactions on Multimedia Computing, Communications, and Applications*,2010,6(3):1-19
- [22] Salo J. Offloading content routing cost from routers[D]. Helsinki:University of Helsinki,2010
- [23] Ni J, Tsang D, Yeung I, et al. Hierarchical Content Routing in Large-Scale Multimedia Content Delivery Network[C]//*Proc. of ICC '03*. 2003:854-859
- [24] Sun Jing,Gao Sui-xiang,Yang Wen-guo, et al. Heuristic Replica Placement Algorithms in Content Distribution Networks[J]. *Journal of Networks*,2011,6(3):416-423
- [25] Borst S, Gupta V, Walid A. Distributed caching algorithms for content distribution networks[C]//*Proc. of INFOCOM'10*. 2010:1-9
- [26] Nygren E, Sitaraman R, Sun J. The Akamai Network: A platform for high-performance Internet applications [J]. *ACM SIGOPS Operating Systems Review*,2010,44(3):2-19
- [27] Repantis T,Cohen J,Smith S, et al. Scaling a Monitoring Infrastructure for the Akamai Network[J]. *ACM SIGOPS Operating Systems Review*,2010,44(3):20-26
- [28] Cohen J,Repantis T,McDermott S, et al. Keeping Track of 70,000+ Servers; The Akamai Query System[C]//*Proc. of USENIX LISA*. 2010:1-15
- [29] Buyya R,Ranjan R,Calheiros R N. Intercloud: Utilityoriented federation of cloud computing environments for scaling of application services[J]. *LNCS*,vol. 6081,2010:13-31
- [30] Sanjoy P,Yates R,Raychaudhuri D, et al. The Cache-And-Forward Network Architecture for Efficient Mobile content delivery services in the future internet[C]//*First ITU-T Kaleidoscope Academic Conference on Innovations in NGN; Future Network and Services*. 2008:367-374
- [31] Nielsen Online Report. Social networks & blogs now 4th most popular online activity [EB/OL]. <http://tinyurl.com/cfzjlt>, 2009
- [32] Pallis G,Zeinalipour-Yazti D,Dikaiakos M D. Online Social Networks;Status and Trends[J]. *New Directions in Web Data Management*,vol. 331,2011:213-234
- [32] Marcon M,Viswanath B,Cha M, et al. Sharing Social Content from Home: A Measurement-driven Feasibility Study [C]//*Proc. of NOSSDAV'11*. 2011:45-50
- [33] Scellato S, Mascolo C, Musolesi M, et al. Track Globally, Deliver Locally;Improving Content Delivery Networks by Tracking Geographic Social Cascades[C]//*Proc. of WWW'11*. 2011:457-466