

大规模软件与 AS 级 Internet 宏观拓扑结构分析比较

李 辉 赵 海 艾 均 李 博

(东北大学信息科学与工程学院 沈阳 110004)

摘 要 大规模软件与 AS 级 Internet 宏观拓扑结构的研究,对两者结构的进一步理解和应用具有重要意义。根据大规模软件结构与 AS 级 Internet 宏观拓扑结构体现出来的复杂网络特征,分别将其抽取为网络拓扑,然后采用复杂网络的度量分析方法,从连通性、度分布特征、小世界特征和层次性 4 个方面进行了分析比较,得到了大规模软件与 AS 级 Internet 在宏观拓扑结构上的若干异同之处,并讨论了其产生原因。

关键词 复杂网络,连通性,无尺度,小世界,核数

中图法分类号 TP393 **文献标识码** A

Analysis and Comparison of the Macro-topology between Large-scale Software and AS-level Internet

LI Hui ZHAO Hai AI Jun LI Bo

(College of Information Science and Engineering, Northeastern University, Shenyang 110004, China)

Abstract The research of the macro-topology between large-scale software and AS-level Internet has important significances in the further comprehension and application for their structures. According to the complex network characteristic which is reflected by the macro-topology between large-scale software and AS-level Internet, their structures were converted to network topology, analyzed and compared from connectivity, degree distribution characteristic, small-world characteristic and hierarchy, using the methods of complex network metric and analysis, then some macro-topology similarities and differences between large-scale software and AS-level Internet were discovered. Additionally, the reasons of them were discussed.

Keywords Complex networks, Connectivity, Free-scale, Small-world, Coreness

在自然界和人类社会中,很多复杂系统的结构如互联网、人际网络、电力网络、人体神经网络等都体现出了一种网络化的组织结构。复杂网络以图论为理论研究基础,对复杂系统网络结构的几何性质、形成机制、结构稳定性等方面进行了大量的研究^[1],发现在复杂网络的复杂性背后蕴含着一些普遍的规律性,可以使用一些方法定量地对其加以表示和分析。其规律呈现的形式十分直观,成为了不同学科领域共同关心的热点。

1999 年以来,人们逐渐发现 Internet 具有复杂网络拓扑结构的特点^[2]。东北大学嵌入式技术实验室从 2001 年开始进行 Internet 拓扑结构的复杂性研究,2005 年 2 月成为 Internet 结构及数据研究的国际合作机构 CAIDA (Cooperative Association for Internet Data Analysis) 的一员,取得了许多重要研究成果。

近年来研究人员发现,大规模软件的结构也呈现出与互联网、人际网络等复杂系统相似的网络化结构特征,开始被作为一种复杂网络进行研究。2003 年 Myers 的一篇文章比较全面地阐述了软件系统结构与复杂网络的关系与研究前景^[3]。一些相关研究也从不同角度分析了复杂网络特征量在

软件系统中的体现^[4-5]。

目前,人们对大规模软件和 Internet 的结构已经做了大量有重要意义的研究^[6-9],但是对两者的直接对比分析很少。本文针对相同规模的软件与 AS 级 Internet 结构,采用复杂网络分析方法,从几个方面分别做了分析和比较,得出了两者在宏观拓扑结构上的若干异同点。

1 大规模软件结构与 AS 级 Internet 的网络化表示

一个具体的网络拓扑可以抽象为一个由点集和边集组成的图。本文将大规模软件结构与 AS 级 Internet 结构抽象为图。具体来说,将软件结构中的面向对象程序的类、接口、结构体、联合体等抽象数据类型看作图中的节点,将节点间的继承、关联、泛化、依赖等关系看作图中的边;将软件结构中的面向过程程序中的结构体等类型看作图中的节点,将它们之间的关联关系看作图中的边。在 Internet 拓扑结构中,将各个 AS (Autonomous Systems) 看作图中的节点,将它们之间通过边界网关的互联关系看作图中的边。

本文采用 CAIDA 组织提供的 2008 年 11 月的数据作为 AS 级 Internet 结构,选取 Eclipse 3. 0. 1 版本、MySQL 6. 0. 6

到稿日期:2009-01-20 返修日期:2009-04-17 本文受高等学校科技创新工程重大项目培育资金项目(708026)资助。

李 辉(1983-),男,博士生,主要研究方向为软件宏观拓扑结构分析,E-mail:lih2002@126.com;赵 海(1959-),男,博士生导师,主要研究方向为复杂网络、传感器网络、软件工程等;艾 均(1981-),男,博士生,主要研究方向为软件宏观拓扑结构分析;李 博(1980-),男,博士生,主要研究方向为软件工程。

版本和 Firefox 2.0.0.11 版本 3 个与 AS 级 Internet 节点规模相近的软件系统,其节点数与边数的规模比较如表 1 所列。并且,上述 3 个软件都是知名的大规模开源软件系统,分别由 Java,C,C++ 程序设计语言编写,软件应用类别分别是程序设计开发平台、数据库、浏览器,以便选取的软件系统具有相对普遍性。

表 1 选取软件与 AS 级 Internet 规模对比

样本	N	M
Eclipse 3.0.1	12583	23220
MySQL 6.0.6	3793	5368
Firefox 2.0.0.11	10595	18739
AS 级 Internet	9704	27492

2 软件结构与 AS 级 Internet 连通性比较

2.1 节点平均度分析

在网络拓扑图中,度是用来研究复杂网络拓扑结构的基本参数,它的定义如下:

定义 1(度) 一个节点 i 的度是指与其连接的其他节点的数目,记作 k_i 。

在网络拓扑中,节点度值的大小表示这个节点在网络中的重要程度。直观上看,一个节点的度越大,就意味着这个节点在某种意义上对整个网络的影响越大。

定义 2(节点平均度) 网络中所有节点的度的平均值称为网络的节点平均度,记为 $\langle k \rangle$ 。

节点平均度反映了网络拓扑的平均连通性能,这个值越大,表明网络拓扑的平均连通性能越好。从表 2 中数据可以看出,AS 级 Internet 平均连接度要明显大于软件结构平均连接度;同时,比较度值前 1% 和前 0.5% 的节点间平均度值,AS 级 Internet 也明显大于软件结构,表明 Internet 的连通性要好于软件结构。

表 2 节点平均度度量数据

样本	$\langle k \rangle$	Top 1%	Top 0.5%
Eclipse 3.0.1	3.69	91.08	121.26
MySQL 6.0.6	3.54	47.91	61.25
Firefox 2.0.0.11	2.83	85.60	145.24
AS 级 Internet	6.42	266.67	439.94

出现这种现象的原因,主要是 Internet 与软件结构对连通性的要求不同。相比之下,Internet 由于需要数据传输的可靠和高效,节点之间往往具有相当数量的冗余连接,以丰富路由选择。当网络拥塞或者遭到破坏时,在局部范围内不用中心节点的参与就可以有多条路由选择。而软件结构中的每个节点对应的模块职责是相对固定的,如果一个节点失效,是不能用其它节点替代的,也就不需要像 Internet 结构那么高的冗余度。因此,Internet 节点相比软件结构的节点就具有相对高的度值;相应地,Internet 也就具有更高的平均连接度。

2.2 富人俱乐部特性分析

定义 3(富人俱乐部连通性) 图中有少量的节点具有大量的边,这些节点被称为“富人节点”。这些富人节点倾向于彼此之间相互连接,构成“富人俱乐部”。富人俱乐部连通性 $\Phi(r/N)$ 用来刻画富人俱乐部现象,它表示的是网络中前 r 个度最大的节点之间实际存在的边数 L 与这 r 个节点之间可能存在的总边数 $r(r-1)/2$ 的比值。

富人俱乐部连通性主要是由高度值节点间的连接倾向形

成的^[7]。表 3 统计了 AS 级 Internet 与上述 3 个软件系统结构的前 1% 和前 0.5% “富人节点”之间的富人俱乐部连通性,从表 3 的统计数据可以看出,AS 级 Internet 结构的富人俱乐部连通性明显高于软件结构。

表 3 富人俱乐部特性度量数据

样本	Top 1%	Top 0.5%
Eclipse 3.0.1	0.0246	0.0387
MySQL 6.0.6	0.0847	0.1062
Firefox 2.0.0.11	0.0402	0.0492
AS 级 Internet	0.2491	0.4823

对 AS 级 Internet 结构来说,前 0.5% “富人节点”之间的富人俱乐部连通性又明显强于前 1% “富人节点”;而像 MySQL 6.0.6 版本和 Firefox 2.0.0.11 版本,前 0.5% 的“富人节点”之间的富人俱乐部连通性相对于前 1% 的“富人节点”并无明显优势。在软件结构中,“富人节点”多是一些如 String 等基础模块,它们之间相当一部分甚至是毫无关联的,因此联系并不如 Internet 的“富人节点”之间紧密。

对一个具体的网络结构来说,当网络遭受破坏而分裂成不同子块时,连通子图的连接冗余度是网络局部鲁棒性的一个有效度量,它从一个侧面反映了路由选择的弹性。冗余连接可以丰富路由的选择,在局部范围内不用中心节点的参与就可以有多条路由选择。Internet 的高连通特性使它具有较强的冗余度,软件拓扑结构核心节点的失效则很容易引发软件系统运行故障,所以 AS 级 Internet 的鲁棒性要强于软件系统。

3 软件结构与 AS 级 Internet 结构节点度分布特征比较

3.1 度分布分析

定义 4(度分布) 度分布指的是网络中节点的度的分布情况,用分布函数 $P(k)$ 来描述。 $P(k)$ 是指随机选定一个节点 i ,恰好使得 $k_i = k$ 的概率。

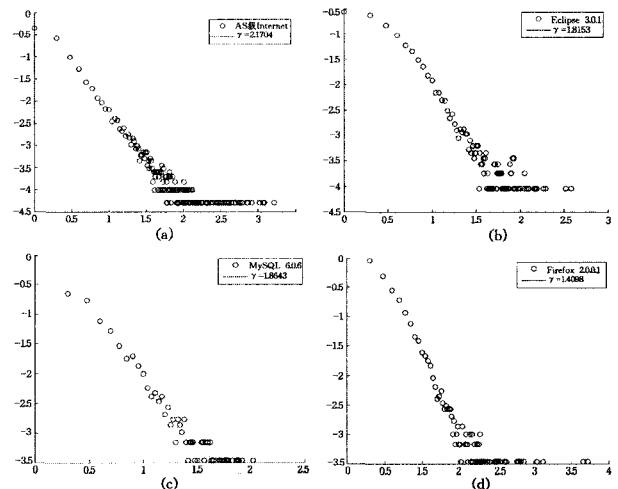


图 1 度分布图

度分布是用来研究复杂网络拓扑结构的基本参数之一。在度分布图中,度分布的幂指数 γ 描述了无尺度网络度分布的均衡程度^[10]。幂指数 γ 越大,度分布的区间越集中。特别地,当幂指数 γ 为 ∞ 时,所有节点的度都在同一个数值上,即为一个度均匀分布的网络。这说明幂指数 γ 越大,网络度分布越“均匀”。相反,幂指数 γ 越小,度分布的区间越发散。特

别是当幂指数 γ 为 0, 即所有的度都有相同概率的分布时, 最大度与最小度极为悬殊。这说明幂指数 γ 越小, 网络度分布越“不均匀”。

根据图 1 中 AS 级 Internet 度分布图和软件结构网络拓扑度分布图可以看出, 两者均呈现出幂率分布特性, 即无尺度分布特性。但是, 从表 4 度分布系数对比可以看出, 两者的无尺度分布还存在明显区别, 软件结构网络拓扑度分布系数 γ 在 1.4 到 2 之间, 而 AS 级 Internet 的度分布系数 γ 一般在 2 到 3 之间, 即 AS 级互联网的幂指数明显大于软件结构网络幂指数, 说明软件结构网络拓扑的度分布更不均匀, 节点间更不平等。

表 4 度分布系数度量数据

样本	γ
Eclipse 3.0.1	1.82
MySQL 6.0.6	1.87
Firefox 2.0.0.11	1.41
AS 级 Internet	2.17

与软件结构的无尺度特征相比, AS 级 Internet 结构比较弱化, 度值较大的点相对于度值较小的点数目减少较慢。软件系统中的度分布系数为 1.8 左右。如果将大度值作为一种重要节点的特征, 软件结构中对整个系统影响更大的节点相对 Internet 要更多一些。这些大度值的节点模块需要与更多的其他模块进行协作和数据交互, 也承担了更多的工作责任。这些节点的行为通常都比较复杂, 同时在这样的模块上出现错误的影响也极大, 这体现了软件的复杂性和脆弱性。

3.2 出入度分布分析

有向网络中一个节点的度分为出度和入度。节点的出度是指从该节点指向其他节点的边的数目, 节点的入度是指从其他节点指向该节点的边的数目。

相比软件结构, AS 级 Internet 更可以看作是无向图结构, 因为 Internet 的数据交换需要每个节点既要发送数据, 又要接收数据。如果按有向图分析, 一般来说每个节点有一个连接某个节点的出度, 必然有一个对应的入度。正是由于这种节点的出入度对称性, AS 级 Internet 可以用无向图的度来简化表示。与 AS 级 Internet 不同的是, 软件结构入度与出度不具备这样的对称性, 图 2 是 Eclipse3.0.1 和 Firefox2.0.0.11 的出入度分布图。

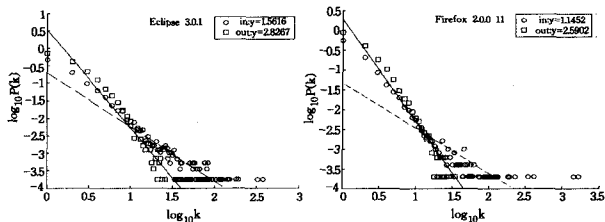


图 2 出入度分布图

从图 2 可以看出, 在软件结构的出入度分布中, 具有以下两个现象:

(1) 最大入度要明显大于最大出度。节点具有较大的入度, 说明节点具有较高的被复用程度, 这符合软件工程中优先复用已有模块的思想。而如果一个节点的出度较大, 则一方面说明这个节点较为重要, 实现功能较多, 另一方面这个节点的内部结构也较为复杂, 在使用中出现错误的概率也较大。所以, 当一个软件模块内部结构过于复杂时, 一般采取的方法是将其分解为几个小的模块来协作完成原先模块的功能, 这

样会降低出现错误的概率。

(2) 出度分布系数要明显大于入度分布系数。相比入度分布, 出于降低模块内部出现错误概率的考虑, 大出度节点出现的概率较小, 一般很少出现达到最大入度值的节点。因此, 度值较大的点相对于度值较小的点数目减少较快, 相应地出度分布系数普遍明显大于入度分布系数。

因此, 对大规模软件系统来说, 最大限度地复用现有模块, 并且合理分解内部结构复杂的模块, 可以降低软件系统出现故障的概率。

4 软件结构与 AS 级 Internet 结构小世界特性比较

定义 5(聚集系数) 节点的聚集系数 C 为节点所有相邻节点之间连接的边的数目占可能的最大边数目的比例, 网络拓扑的聚集系数则为所有节点聚集系数的平均值, 表示网络集团化的程度。

定义 6(平均路径长度) 网络中两个节点之间的距离定义为连接这两个节点的最短路径上的边数, 网络的平均路径长度 d 定义为任意两个节点之间距离的平均值。

聚集系数反映了网络拓扑中节点之间的内聚程度以及层次化的趋势。对于节点规模为 N 的网络, 聚集系数都远远大于 $O(N^{-1})$, 说明网络具有高聚集特性^[11]。从表 5 度量数据来看, 无论是 AS 级 Internet 还是软件结构, 其聚集系数都远大于节点数的倒数, 因此, 软件结构与 AS 级 Internet 结构的节点同样具有高聚集特性。同时, 软件结构与 AS 级 Internet 结构的平均路径长度分别相对于各自网络规模来说又很小, 表明软件结构与 AS 级 Internet 结构都具有小世界网络特性^[12]。两者所呈现的聚集性质, 主要是低度节点倾向与高度节点连接的异配性质所导致的, 低度节点倾向与高度节点相连, 而高度节点之间的连接又比较紧密, 所以网络整体上表现为节点拥有紧密相连的邻居。

表 5 聚类系数与平均路径长度度量数据

样本	C	d
Eclipse 3.0.1	0.0279	5.67
MySQL 6.0.6	0.0620	7.23
Firefox 2.0.0.11	0.0167	6.08
AS 级 Internet	0.4031	3.47

但是, 对比表 5 中软件结构和 AS 级 Internet 结构的聚集系数和平均路径长度的数据, 又可以看出两者具有明显的不同, Internet 比软件结构具有更高的聚集系数和更短的平均路径长度, 小世界特征更加明显。

出现上述现象的原因, 从根本上说是 Internet 与软件结构的设计思想和目的不同。对互联网来说, 其数据传输要求尽可能高的效率, 反映到 Internet 结构拓扑图中就是任意两点之间不但要求可达, 而且要求这两点之间的跳数尽可能少, 需要有一定的冗余连接来保证互联网的服务质量^[13], 节点之间的联系更加紧密。而软件结构本身在设计上并没有像互联网那样的要求和意图, 对软件结构网络拓扑中的一个节点来说, 它只需要与它存在协作关系的几个节点紧密联系即可, 不关心与它不存在协作关系的节点, 这从客观上增大了软件拓扑结构中节点之间的平均连接距离。

另一方面, 文献^[14]指出, Internet 平均路径长度随时间呈现出逐渐下降的趋势。而软件同一时间序列下多个版本的度量数据显示, 软件结构的平均路径长度却呈现出逐渐增大的趋势, 演化方向与 Internet 截然相反。

这种现象同样也是 Internet 与软件结构的设计思想和目

的不同造成的。在 AS 级 Internet 结构演化进程中,高度值节点之间的联系会更加紧密,某些较低度值节点为了自身数据传输效率的提高会增加与其它节点的连接边,也会在客观上提高网络的连通性。软件系统的功能扩展使得不断有新节点模块加入,这些新节点按照偏好依附原则,只与个别与其存在协调工作关系的节点直接联系,不存在 Internet 网络拓节点间相互连通的要求,客观上会增加软件拓扑结构中任意两点之间的连接距离。

AS 级 Internet 的高聚集特性也使它的结构相对软件结构具有较强的冗余度,再次表明 AS 级 Internet 的鲁棒性要强于软件系统。

对大规模软件系统来说,重视软件复用的思想,充分挖掘软件模块的内部“潜力”,复用现有源代码中的常用模块和优秀模块,既可以经济地提高软件开发的效率,又可以提高软件结构的聚集性以增强鲁棒性。

5 软件结构与 AS 级 Internet 结构层次性特征比较

5.1 核数分布及其与度数的相关性分析

定义 7(k -核) 图的 k -核是指在图中反复去掉度值小于等于 k 的节点及与其连接的边所余下的子图。 k 值越大,意味着该核在图中位置更为核心。

定义 8(节点核数/图核数) 节点核数表示包含该节点的最核心的核,即指若节点核数为 k ,则该节点存在于 k -核中,而不存在于 $(k+1)$ -核中。图中节点核数的最大值为该图的核数,记作 $k\text{-core}_{\max}$ 。

节点核数标识节点在拓扑中的深度,其分布也是认识复杂网络拓扑的一个重要度量^[7]。通过对节点核数以及拓扑图各级核数的分析和计算,可以量化考察网络的层次性^[15]。图 3 分别是 AS 级 Internet 与 MySQL 6.0.6 版本的核数分布图。从图中可以看到,AS 级 Internet 与软件结构的核数分布与度分布相似,也呈现幂指数分布,并且大部分节点的核数都很小,只有少数的节点具有较大的核数。

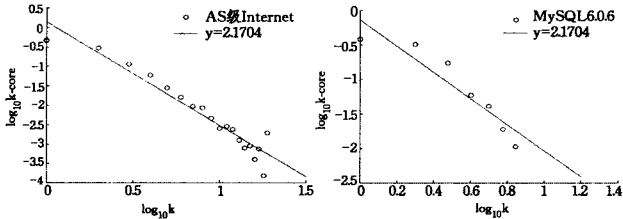


图 3 AS 级 Internet 与 MySQL 6.0.6 版本核数分布图

图 4 描述了 AS 级 Internet 与 Eclipse 3.0.1 版本、MySQL 6.0.6 版本和 Firefox 2.0.0.11 版本的核数度数相关性。从图 4 可以看到,在 Internet 与软件结构中,核数数值的变化比度值的变化要小。在 AS 级 Internet 低度值节点的核数与其度值之间有较强的正相关性,但是在节点度值较高的区域,核数不再显著增加。高度值节点主要集中于最高核内,但有少部分散落于其它层,说明并非全部“富”节点都具有较高的层次,也说明以核数相关性来分析网络层次较之传统的节点度分层方法更为有效。

5.2 核数数值及层次化分析

表 6 为 AS 级 Internet 与 3 个软件的核数度量数据。从表 6 可以看出,AS 级 Internet 拓扑结构的核数约为 21~23,明显高于软件结构的核数。AS 级 Internet 拓扑结构体现出一种自发的多层次的性质,这与互联网自治系统的等级划分

是密切关联的。

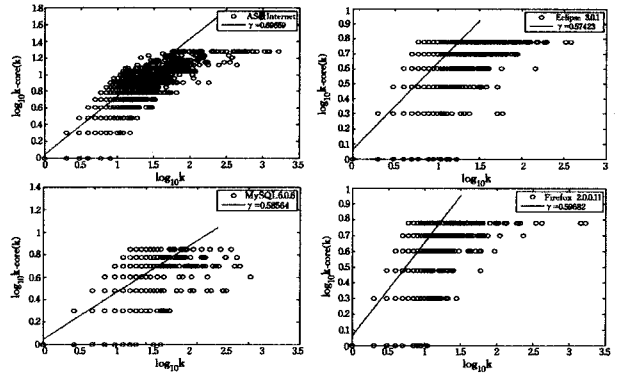


图 4 核数度数相关性图

对互联网来说,为网络提供服务的是因特网服务提供者 (ISP),一个 ISP 在管理上可以分成一个或几个 AS。由于只有极少数 ISP 申请多个 AS 号在 BGP 协议中作为该 ISP 部分网络的标识,因而可以认为 ISP 和 AS 只是在不同语境下对同一实体的不同称呼。这就为我们在划分 AS 时提供了便利,同时也是互联网具有层次性的最直接和最根本的原因。严格的互联网自治系统等级模型应该是这样的:一小部分全球传输处在互联网最顶级,下面各级分别是国家级 ISP、地区级 ISP 及本地 ISP,下一级 ISP 只与上一级 ISP 相连,每一级 ISP 都是上一级 ISP 的客户。

软件结构的核数比 AS 级 Internet 的核数要小得多,普遍在 4~7 之间,结构更为“扁平”。也就是说,当前单一软件系统的层次数目是相对有限的。尽管核数的分布特征与 Internet 相比不明显,但是节点的核数仍然表现出了软件结构具有层次特征,还可以作为度量节点影响范围大小的重要特征量。核数越大的节点在软件系统中的作用也就越重要,核数大的节点可以在不断的重构演化中逐渐发展为软件系统的核心框架。

表 6 核数度量数据

样本	k-core _{max}
Eclipse 3.0.1	6
MySQL 6.0.6	7
Firefox 2.0.0.11	6
AS 级 Internet	19

结束语 本文将大规模软件结构与 AS 级 Internet 结构看作复杂网络结构,用复杂网络的分析手段,分别从连通性、度分布特征、小世界特征、层次性 4 个方面对两者进行了分析比较,得到两者以下若干异同之处:

(1) AS 级 Internet 结构的平均节点度和富人俱乐部特性高于大规模软件结构,因此 Internet 结构的连通性要优于软件结构。

(2) 大规模软件结构与 AS 级 Internet 结构的度分布都呈现出无尺度分布特征;但是软件结构的无尺度分布要弱于 AS 级 Internet,软件结构的出入度分布呈现不对称性,AS 级 Internet 结构鲁棒性更强。

(3) AS 级 Internet 结构与大规模软件结构都具有较高的聚集系数和较短的平均路径长度,呈现小世界特征;AS 级 Internet 结构相比大规模软件结构,小世界特征更明显,并且具有相反的演化方向。

(4) AS 级 Internet 与大规模软件结构的核数分布都呈现无尺度分布特征,低度值节点与核数有更强的相关性;Internet 相比软件结构具有更高的核数,层次性更明显。

本文还详细讨论了两者结构异同之处产生的原因,并且认为,对比 AS 级 Internet 结构,大规模软件的结构有进一步完善的空间。本文所做的工作对 Internet 和软件结构的进一步理解和研究具有一定的帮助。

参考文献

[1] 汪小帆,李翔,陈关荣. 复杂网络理论及其应用[M]. 北京:清华大学出版社,2006:1-14
 [2] Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the Internet topology[J]. ACM SIGCOMM Computer Communication Review, 1999, 29(4): 251-262
 [3] Myers C R. Software systems as complex networks; Structure, function, and evolvability of software collaboration graphs[J]. Physics Review E, 2003, 68: 1-15
 [4] Ma Yu-tao, He Ke-qing, Du Dehui. A qualitative method for measuring the structural complexity of software systems based on complex networks[A] // 12th Asia-Pacific Software Engineering Conference[C]. Taiwan, 2005: 257-263
 [5] Lee Young, Yang Jeong, Chang K H. Metrics and evolution in open source software[A] // 7th International Conference on Quality Software[C]. 2007: 191-197

[6] Albert R, Barabási A L. Statistical mechanics of complex networks[J]. Rev Mod Phys, 2002, 74 (1): 47-97
 [7] Newman M E J. The structure and function of complex networks[J]. SIAM Rev, 2003, 45 (2): 167-256
 [8] 李兵,王浩,李增扬,等. 基于复杂网络的软件复杂度度量研究[J]. 电子学报, 2006, 34(12A): 2371-2374
 [9] 郑晓龙,曾大军,李慧倩,等. 开源软件的复杂网络分析及建模[J]. 复杂系统与复杂性科学, 2007, 4(3): 1-9
 [10] Zhou Shi, Mondragon R J. The rich-club phenomenon in the Internet topology[J]. IEEE Communication Letters, 2004, 8(3): 180-182
 [11] 张国强,张国清. Internet 网络的关联性研究[J]. 软件学报, 2006, 17(3): 490-497
 [12] Watts D J, Strogatz S H. Collective dynamics of 'smallworld' networks[J]. Nature, 1998, 393(4): 440-442
 [13] 赵金晶,朱培栋,牛薇,等. Internet 平均路径长度的定量研究[J]. 计算机工程, 2008, 34(2): 157-162
 [14] 张昕,赵海,王莉菲,等. AS 级 Internet 拓扑分析[J]. 通信学报, 2008, 29(7): 50-61
 [15] Zhang Guoqing, Zhang Guoqiang, Yang Qingfeng, et al. Evolution of the Internet and its cores[J]. New Journal of Physics, 2008, 10(12): 123027

(上接第 141 页)

为 1000×700 具有 53 个相互交叉的路段, 23 个三叉或四叉路口的交通网中模拟了 10k~50k 个移动对象的运动。初始时, 车辆是在路段上随机分布的, 车辆可以随时进入和离开路网, 原始数据每一个时间步长更新一次。实验主要从索引建立时的更新性能以及针对移动对象的数量和查询窗口的尺寸不同对查询效果的影响方面将 AUC 索引与 TPR 树进行了比较。

图 7 是从建立索引时随着移动对象数量的增加对索引更新所需磁盘访问次数的影响上将 TPR 树和 AUC 索引进行的比较。由于二者采用了不同的轨迹预测方法, TPR 树采用的是线性预测, AUC 采用的是 CA 的基于模拟的预测方法, 而且本文在环形交叉口中采用了圆弧拟合对象的轨迹, 更加接近对象的真实运动轨迹, 虽然计算代价有所增加, 但从图 7 可看出, AUC 索引更新的代价要远远小于 TPR 树。图 8 是查询过程中, 从范围查询窗口的规模对查询性能的影响上将 TPR 树和 AUC 索引进行的比较。由于 AUC 索引中考虑了交叉口的情况, 因此当查询窗口中包含了路段的交点时, 就要对该交点所对应的交叉口进行查询。这样, 随着查询窗口的增大, AUC 索引所需的磁盘 I/Os 数就会增加, 从图中可看出, 随着窗口的增大, AUC 索引查询所需的磁盘 I/Os 大约保持在 TPR 树的 2/3 左右。

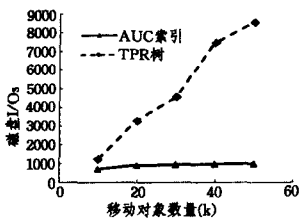


图 7 TPR 树和 AUC 索引建立时磁盘 I/Os 比较

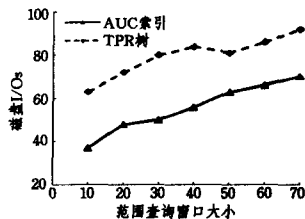


图 8 TPR 树和 AUC 索引查询效率与范围查询窗口的关系比较

图 9 和图 10 是在查询过程中从移动对象的数量与磁盘访问次数的关系方面对 AUC 索引与 TPR 树和 TB 树分别进行的比较。可见, 随着移动对象数量的增加, AUC 索引所需的磁盘 I/Os 都要远小于 TPR 树和 TB 树。

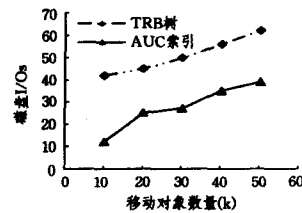


图 9 TPR 树与 AUC 索引查询性能与移动对象数量的关系比较

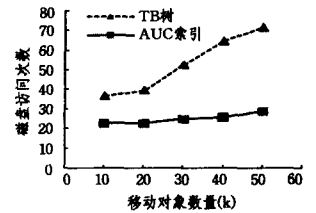


图 10 TB 树与 AUC 索引查询效率与对象数量的关系比较

结束语 本文根据城市交通网和交通系统中移动对象的运动特点, 采用了一种带有环形交叉口的元胞自动机模型模拟对象的运动, 并在此基础上提出了基于城市道路网络的 AUC 索引结构。本结构可以支持对路网中的移动对象的过去、当前和将来的轨迹的查询。通过实验表明, AUC 索引的查询和更新性能都要优于 TPR 树和 TB 树。

参考文献

[1] Chen Ji-dong, Meng Xiao-feng. Indexing Future Trajectories of Moving Objects in a Constrained Network[J]. Journal of Computer Science and Technology, 2007, 22(2): 245-251
 [2] 陈继东,胡志智,孟小锋,等. 一种基于城市交通网络的移动对象全时态索引[J]. 计算机研究与发展, 2007, 44(6): 1008-1014
 [3] Nagel K, Scheckenberg M. A cellular Automaton Model for Freeway Traffic[J]. J. Phys. I, France, 1992(2): 2221-2229
 [4] Biham, et al. Self-organization and a Dynamical Transition in Flow Models[J]. Pys Rev A, 1992, 46: 6124-6127
 [5] Schadschneider A, Chowdhury D, et al. A New Cellular Automaton Model for City Traffic[C] // Germany: Traffic and Granular Flow '99, 2000
 [6] Dupuis A, Chopard B. Parallel traffic simulation on Geneva using cellular automata[J]. Parallel and Distributed Computing Practices(PDCP), 1998, 1(3): 79-92
 [7] Pei Soo-chang, Horng Ji-hwei. Fitting Digital Curve Using Circular Arcs[J]. Pattern Recognition, 1995, 28(1): 107-116
 [8] 涂嘉文,徐守时,谭勇. 基于最小均方误差的圆弧分段曲线拟合方法[J]. 计算机应用, 2001, 21