

基于复杂网络的软件开发方法研究

何明东,熊建斌,李振坤

(广东工业大学 计算机学院,广东 广州 510006)

摘要:近年来,发展中的复杂网络理论越来越多地应用到各个学科中,如生物学、社会学、城市交通、计算机网络等。复杂网络理论在软件工程中的应用也得到很多研究者的青睐。然而之前大部分研究都是采用从代码出发的“反向”研究途径。文中探讨了复杂网络理论在软件工程中的应用,根据两者的结合点将复杂网络理论的最新研究成果应用到软件开发过程中,提出了一种基于功能点和复杂网络理论的软件模块化模型。最后在广东省高中阶段教育招生统一服务平台的开发过程中证明了该模型的实用性和有效性。

关键词:复杂网络;软件工程;模块化

中图分类号:TP311.52

文献标识码:A

文章编号:1673-629X(2011)06-0059-04

Complex Network-Based Software Development Method Research

HE Ming-dong, XIONG Jian-bin, LI Zhen-kun

(College of Computer, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: In recent years, developing increasingly complex network theory applied to various disciplines such as biology, sociology, urban transport, computer networks. Complex network theory in software engineering has also been a lot of researchers of all ages. However, most studies are based on the previous departure from the code of “reverse” research approach. Discuss the complex network theory in software engineering, the latest research results of complex network theory point applied to software development process according to combination of both, a modular software model based on function points and complex network theory is presented. It is proved the model is practical and effective in Guangdong Province in the final stages of education in the enrollment process of a unified service platform for the development.

Key words: complex network; software engineering; modular

0 引言

随着计算机硬件和信息技术的快速发展,网络和软件的融合,使软件系统呈现出从两个方面的转变:一方面软件运行平台从集中、封闭的单机环境向开发、动态和多变的网络环境转变^[1];另一方面,软件开发也从面向产品向面向网络服务转变,网络资源成为生产软件系统考虑的主要因素。以网络为基础的软件系统随着集成规模越来越庞大,用户量和数据交互都会呈现喷井式的增长,成为异常复杂的软件系统。软件设计和开发者对于不断膨胀的软件系统的理解和控制越来越困难,网络化软件在这种环境应运而生。网络化软件系统^[1],是指人工构建的,用于对基于计算机网络

的信息系统进行控制和管理,面向应用领域,向大众用户提供各种服务的大规模软件系统。自从20世纪60年代,随机图论一直是研究复杂网络结构的基本理论,绝大多数实际的复杂网络结构都不是完全随机的。Watts和Strogatz在1998年发现了复杂网络的小世界特征^[2],随后一年,Barabasi和Albert提出复杂网络的无标度特征^[3]。对这两种特征的发现极大地促进了复杂网络理论的发展,不断有人提出各种的复杂网络模型来模拟和解析现实世界的复杂网络,比如演化网络等。复杂网络的研究也不再局限于数学领域,在社会学、物理学、生物学、计算机等众多学科也掀起了研究复杂网络的热潮。如:复杂网络中的社团结构、复杂网络中的传播动力学、计算机病毒在Internet上的传播、蛋白质之间的相互作用、经济网络等。美国卡内基·梅隆大学(Carnegie Mellon University)软件工程研究所在2004年就开始了对复杂巨型系统软件专项调研,以解决在开发复杂系统软件所遇到的软件工程问题。随后在2006年开展网络化软件的研究。在国内,何克清等人将复杂网络理论和研究方法应用到软件工程的

收稿日期:2010-11-09;修回日期:2011-02-17

基金项目:广东省自然科学基金资助项目(9151009001000021);广东省教育部产学研合作专项资金资助项目(2009B090300341)

作者简介:何明东(1985-),男,广东兴宁人,硕士研究生,研究方向为复杂网络和分布式系统;李振坤,教授,硕士生导师,研究方向为软件项目管理、计算机网络与分布式系统。

设计之中,提出网络化软件的概念^[3]。复杂网络理论和方法引入软件工程,为软件工程的发展注入了新鲜的血液。文中提出了一种基于复杂网络的模块化方法,并将复杂网络应用于计算机软件开发中。

1 复杂网络与软件工程

1.1 复杂网络

Watts 和 Strogatz 提出大量真实网络都具有小世界效应^[4](Small-World Effect),即具有大的聚类系数和小的平均路径长度。继小世界特性被揭示之后,Barabasi 和 Albert 惊奇地发现万维网中的节点度服从幂律分布,由于幂律分布具有标度无关性,所有把节点度服从幂律分布的网络叫做无标度(Scale-Free)网络^[2]。通过大量实证研究发现各种大型网络随着规模增大都会自我演化为这种无标度网络。无标度特性刻画出了复杂网络的节点度的不均匀性,即少数节点拥有大量的边,而大部分节点只有少量边。随后 Barabasi 和 Albert 进一步提出 BA 模型。BA 模型两个最主要的特性是增长特性和优先连接特性。增长特性指网络的规模随着时间在不断地扩大;优先连接特性指新增加的节点更倾向于与网络中那些具有较高连接度的“大”节点相连。一个新节点与一个已经存在的节点 i 的相连接概率为 $p(k_i) = k_i / \sum k_j$,其中 k 为节点 i 的度数。优先连接特性显示,BA 模型中,众多节点优先连接到度大的节点上,呈现一种“富者愈富”的现象^[5]。

这种度聚集现象使整个网络分化成由若干的“集体”组成,集体内节点间连接率远远大于集体间的连接率,研究人员把网络中的这种“集体”称为社区结构。这与软件工程中要求的模块间的内聚性与耦合性非常相似。

1.2 软件工程

软件工程要求软件要有强的内聚性和弱的耦合性。

内聚性是模块内各部分之间联系紧密程度的度量,一个模块内,内聚性强,标志模块的独立性强;内聚性弱,标志模块的独立性差。在一个理想的软件系统中,每个模块只有单一的功能,只做需求的一件事情,但在实现中一个模块往往执行若干方法结合在一起的任务,这些方法组合方式的紧密程度不同就构成了不同的内聚性。在进行模块设计时,尽量争取使模块内聚性强。

耦合性是模块之间联系紧密程度的度量。耦合性强,标志互联的强,模块独立性差;耦合性弱,标志互连的弱,模块独立性强。耦合强弱取决于划分模块造成模块间接口的复杂程度。软件工程的模块性的特性与复杂网络的特性-社团结构不谋而合。

2 复杂网络在软件工程中的应用

如果将软件系统的每一个功能点或者类当做一个节点,一个功能点或类与另外一个功能点或类之间有共享的数据、继承、聚合等关系就将这两个节点加上一条边。这样软件结构就会转换成一个网络结构。在一个模块内部会有大量的连接,在模块之间则只会有少量的连接。

大量研究表明软件系统形成的网络具有明显的复杂网络特征。Potanin, wheeldon^[6-8]等人以对象为点,对象之间的关系为边分析了 C++, Java, Smalltalk, Linux 等大型软件系统,所形成的网络具有很明显的复杂网络特征。图 1 是分析 Linux V0.12 内核所生成的拓扑网络图。

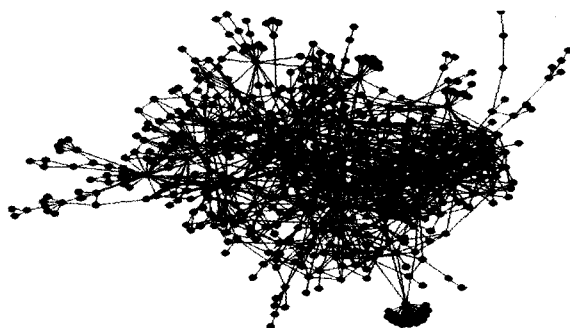


图 1 Linux V0.12 内核所生成的拓扑网络图

很明显,整个网络是由若干个“集体”构成的,每个集体内部的节点之间的连接非常紧密,但是各个群之间的连接却相对来说比较稀疏。说明了软件系统网络里存在着社区结构,这也验证了软件系统网络的复杂网络特性。

然而之前的绝大部分研究都是采取后验的方式进行的。即是采用从代码出发的“反向”研究途径,没有真正涉及软件系统开发设计所遵循的组织结构^[8]。主要采取的策略是以软件系统中的类库、类、接口、编译单元等作为网络中的节点,它们之间的关联、依赖、聚合、组合等关系作为边,建立软件网络模型,应用于软件度量^[9-11]、测试^[12]等。这些研究主要根据软件网络的拓扑结构,使用各种网络模型和研究方法得出软件系统元素间复杂的关系以达于了解软件系统整体性质和演化规律的目的。这些关系为复杂软件系统的设计和开发提供了有价值的参考和依据。然而,复杂网络在软件工程中的应用不能仅仅在发现软件网络的结构特性和解释为什么会产生这种特性上,而更应该在发现开发复杂软件系统新的软件方法学等方面展开深入研究^[4]。

3 复杂网络与软件模块化

在需求分析阶段,开发者不能控制有多少的方法、

类或者其他粒度单位。可以知道的是功能点,比如人员管理、财务报表等等。也可以分析出数据库的基本表结构,大概包括有哪些表。下面给出一个针对功能点和数据点建立的软件复杂网络模型。

3.1 功能点-数据点模型(Function-Data, FD)

定义1 G 是无向图,表示软件网络。包括节点集 V 和边集 E , $G = \langle V_1, V_2, E \rangle$ 。

定义2 软件网络的节点 V_1 表示功能点, V_2 表示数据点。

定义3 边集 E 中的边对应 V_1 与 V_2 的关系。

关系1 V_1 使用 V_2 中的数据。

关系2 V_1 产生 V_2 中的数据。

3.2 FD模型的模块性度量

Girvan 和 Newman 在 2002 年提出 GN 算法。GN 算法是一种分裂方法。它的思想是不断地从网络中移除介数(经过每条边的最短路径数目)最大的边。GN 算法为区别社团的内部边和社团之间的边提供了一种有效的度量标准^[5]。

GN 算法的基本流程如下:

- 1) 计算网络中每条边的介数。
- 2) 不断移除网络中介数最大的边,直到每个节点为孤立的节点为止。

Girvan 和 Newman 还提出了计算衡量社区模块性 (Modularity, M) 的计算公式^[13,14]。

$$M = \sum_{i=1}^r \left[\frac{l_i}{L} - \left(\frac{d_i}{2L} \right)^2 \right] \quad (1)$$

其中 $L = \sum |E|$ 为 G 的总边数, $\delta(v, \omega) =$

$$\begin{cases} 1, & \text{if } (v, \omega) \in E \\ 0, & \text{otherwise} \end{cases}, l_i = \sum_{v \in C_i, \omega \in C_i} \delta(v, \omega) \text{ 为社区 } C_i \text{ 内各顶点的边数总和, } d_i = \sum_{v \in C_i, \omega \in V} \delta(v, \omega) \text{ 为社区 } C_i \text{ 所有节点的总度数。}$$

社区内的边数越多,社区间的边越少, M 值就越大,社区结构特征就越明显。反之,社区结构特性就越不明显。在软件网络中, M 值大,说明内聚性强,耦合性弱; M 值小,说明内聚性弱,耦合性强。

3.3 耦合性定义

耦合性是模块独立性的度量之一,可以由度分布系数来定义。模块 C 的度分布系数定义为:

$$P_c = 1 - \left(\frac{k_c}{\sum k} \right)^2 \quad (2)$$

其中 k_c 为模块 C 内节点的边数, $\sum k$ 为模块 C 内节点的总度数。模块内节点之间的连接越密切,则 $\left(\frac{k_c}{\sum k} \right)$ 越大, P_c 越小,表明模块间耦合性越小;反之,模块间耦合性越大。

3.4 FD模型实际应用

自 2009 年起广东省高中阶段学校(含普通高中和中等职业学校、技工学校,下同)招生(以下简称中招)工作实行统一的高中阶段教育招生服务平台(以下简称省中招服务平台)、统一初中毕业生数据、统一在省中招服务平台编制生源计划、市(县、区)教育局招生部门组织实施、各类学校分类分层次录取、地市及时报送录取信息的招生录取工作实施办法,努力实现全省高中阶段学校招生工作统筹管理、数据集中、实时监控的目标。

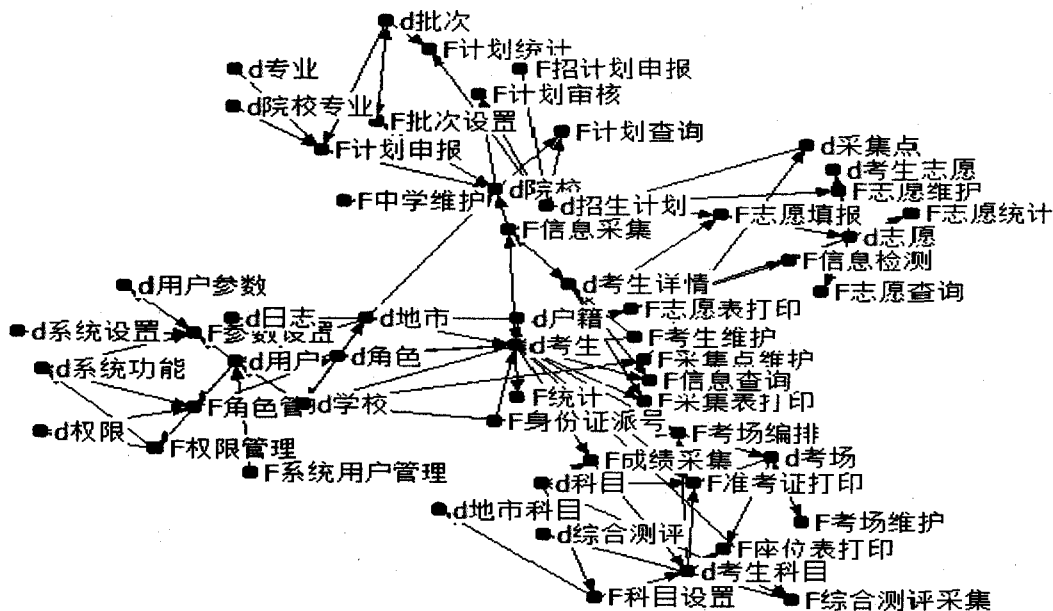


图2 广东省中招平台功能网络图

根据省招生办招生工作的需求,中招服务平台主要功能表现在如下几个方面:考生数据采集、院校计划采集、数据维护、数据统计、系统管理、报表打印、录取等。数据集中主要表现在考生信息、院校计划、考生志愿、考务数据、系统参数等。

图 2 是根据 FD 模型抽象出的广东省中招统一服务平台的主要功能的网络图。

由 GN 算法计算得到的中招服务平台网络图分成五个模块,平均 M 值为 0.583。模块的度分布系数为 0.132。功能网络图模块性特征非常明显。

广东省中招服务平台面向考生、院校、各地市招办三种类型的用户,于是该平台系统分成三个应用。考生端包含考生信息采集、考生志愿信息采集两个模块;院校端包含院校计划申报模块;地市端包含考务和数据操作、系统功能设置等模块。

图 3 为设计后的中招服务平台总体结构图。

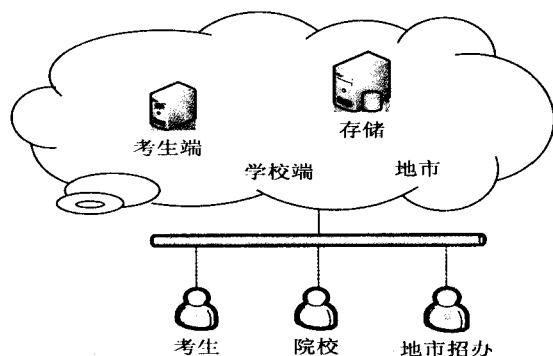


图 3 广东省中招服务平台总体结构图

4 结束语

文中探讨了复杂网络理论在计算机软件工程中的应用。针对之前复杂网络理论在软件设计方法方面研究的缺乏,在前人研究的基础上提出了一种基于功能点和数据点的软件模块化模型。该模型可以为软件开

发过程提供参考,也在实际项目应用中得到了良好的效果。

参考文献:

- [1] 左开中. 复杂网络与网络化软件系统[J]. 计算机应用与软件, 2009, 26(5): 94-95.
- [2] Barabasi A L, Albert R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
- [3] 吕金虎, 王红春, 何克清. 复杂动力网络及其在软件工程中的应用[J]. 计算机研究与发展, 2008, 45(12): 2052-2059.
- [4] Watts D J, Strogatz S H. Collective dynamics of 'small world' networks[J]. Nature, 1998, 393(6684): 440-442.
- [5] 汪小凡, 李翔, 陈荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006.
- [6] Potanin A, Noble J, Frean M, et al. Scale-free Geometry in object-oriented Programs[J]. Comm. ACM, 2005, 48: 99-103.
- [7] Wheeldon R, Counsell S. Power Law Distributions in Class Relationships[C]//Proc. Third IEEE Int'l Workshop Source Code Analysis and Manipulation. [s. l.]: [s. n.], 2003.
- [8] Wen L, Dromey R G, Kirk D. Software Engineering and Scale-free Networks[J]. IEEE Transactions on Systems, 2009, 39(4): 845-854.
- [9] 李冰, 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(B12): 2372-2375.
- [10] Gao Song, Li Chunping. Complex Network Model for Software System and Complexity Measurement[C]//2009 World Congress on Computer Science and Information Engineering. [s. l.]: [s. n.], 2009.
- [11] 王树森, 顾庆, 陈焘, 等. 基于复杂网络的大型软件系统度量[J]. 计算机科学, 2009, 36(2): 287-290.
- [12] 李丽萍, 缪准扣, 钱忠胜. 基于复杂网络面向对象集成测试的研究[J]. 计算机科学, 2008, 35(12): 254-257.
- [13] 郭雷, 许晓鸣. 复杂网络[M]. 上海: 上海科技教育出版社, 2006.
- [14] 姚园园, 张志鸿. 从复杂网络角度对即时消息网络进行拓扑建模[J]. 微计算机信息, 2006, 22(8): 207-211.

(上接第 58 页)

- [4] 包磊, 姚放吾. 基于远程设备的汇编源码调试器的设计与实现[J]. 计算机技术与发展, 2009, 19(11): 155-158.
- [5] 周洪波. 嵌入式系统软件开发环境中调试器的设计[J]. 微计算机信息, 2006, 22(5-2): 60-61.
- [6] 赵全忠. 嵌入式系统及其调试手段的研究[J]. 沈阳工业学院学报, 1999(12): 75-79.
- [7] 仇玉章. 32 位微型计算机原理与接口技术[M]. 北京: 清华大学出版社, 2000.
- [8] Wiley Rosenberg. How Debuggers Work: Algorithms, Data Structures, and Architecture[M]. [s. l.]: Wiley Computer

Publishing, 1996: 1-256.

- [9] Azzerini B, Lopriore L. Program debugging environments: design and utilization[M]. New York: Ellis Horwood, 1992.
- [10] 李红卫. 嵌入式远程调试工具的研究与实现[J]. 微计算机信息, 2009, 25(1-2): 87-89.
- [11] 杨宏业. 清华嵌入式软件系统的解决方案[J]. 计算机测量与控制, 2002, 10(8): 539-540.
- [12] Agrawal H. Towards automatic debugging of computer programs[M]. [s. l.]: Purdu University, 1991.