

软件体系结构的复杂网络社团特性研究

秦怀斌, 梁 斌, 邵明文

(石河子大学 信息科学与技术学院, 新疆 石河子 832003)

摘 要: 作为系统设计的高层抽象和系统实现的蓝图, SA 体现了系统的基本思想和系统开发的早期决策. 对所开发系统的 SA 结构的复杂网络社团特性研究, 是了解整个系统的网络结构和功能特征的重要途径. 介绍了复杂网络及复杂网络社团理论和方法, 提出借助复杂网络社团思想进行 SA 复杂网络社团特性度量. 将 SA 结构转化为复杂网络模型, 并通过 SA 的模块度、SA 的层次性及 SA 的重叠性等来进行 SA 结构的复杂网络社团特性研究. 最后, 通过实例对所提出的研究思路进行验证. 实践表明, 此方法能较好地对所开发系统的 SA 结构的复杂网络社团特性进行评估及度量, 为系统后续开发等工作提供依据.

关键词: 软件体系结构; 复杂网络; 社团特性

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2012)04-0006-04

Research on Complex Networks Community Characteristics for Software Architecture

QIN Huai-bin, LIANG-Bin, SHAO Ming-wen

(College of Information Science and Technology, Shihezi University, Shihezi 832003, China)

Abstract: This paper introduces the theory and method of complex networks and complex networks community, puts forward the complex networks community characteristics metrics method of software architecture based on complex networks community. Transforms SA structure into a complex network model, and through Modularity, hierarchical structure, overlapping communities to metric complex networks community characteristics of SA. Finally, verifies the research though by example. Practice shows that this method can better metric and assessment for SA's complex networks community characteristics of a system, and provides the basis for further development.

Key words: software architecture; complex Networks; community characteristics

1 引言

作为软件系统的高层抽象, 软件体系结构 (Software Architecture, SA) 体现了系统的基本思想和系统开发的早期决策, SA 在明确了系统各个组成部分的同时, 也限定了各个组成部分的交互方式.

复杂网络作为复杂性科学研究的一部分, 它打破了线性、均衡、简单还原系统的传统模式, 致力于系统非线性和非均衡方面的研究. 系统的复杂性主要体现在结构复杂性、节点复杂性及各种复杂性因

素的相互影响等三方面^[1-2]. 借助于复杂网络, 不仅研究构成系统各个组成部分的复杂性, 更重要的是研究系统整体的复杂性.

近几年, 许多研究者对一些真实系统从复杂网络角度进行刻画, 发现这些系统存在一个共同特性: 复杂网络社团特性. 复杂网络社团特性是指, 网络中的众多节点按照某种方式进行分组, 组内节点间的连接比较稠密, 组间节点的连接比较稀疏^[1-2]. 对系统复杂网络社团特性的研究, 是了解整个系统的网络结构和功能特征的重要途径.

鉴于此, 本文将软件体系结构和复杂网络社团

收稿日期: 2011-07-14; 修回日期: 2011-09-18

基金项目: 国家科技支撑计划项目 (2007BAH12B07); 国家自然科学基金 (60963006); 石河子大学科学技术研究发展计划 (ZRKX2010YB18)

特性进行融合,通过复杂网络社团特性理论与方法进行软件体系结构相关度量研究,以此从较高的抽象层理解系统的复杂网络社团特性,为后续系统开发提供指导。

2 复杂网络社团及软件体系结构社团

2.1 复杂网络

复杂网络借助于图论和统计物理方法来捕获并描述系统的整体行为. 复杂网络抽象为一个图 $G=(V,E)$, 其中, $V=\{V_1, V_2, \dots, V_n\}$ 为节点集合, $E=\{E_1, E_2, \dots, E_n\}$ 为边集合, V 中的元素称为节点, E 中的元素称为边. $N=|V|$ 为节点总数, $L=|E|$ 为边总数. E 中的每条边 E_i 与 V 中的一对节点 (u,v) 对应, 如果 E 中任意节点对 (u,v) 和 (v,u) 对应同一条边, 则该网络称为无向网络, 否则为有向网络; 如果 E 中所有的边的长度均为 1, 即 $|E_i|=1$, 则称网络为无权网络, 否则为加权网络。

国内外目前对无向无权复杂网络研究较多, 并主要围绕节点的度 (degree) 及度分布 (degree distribution)、平均路径长度 (average path length)、簇系数 (clustering coefficient)、度相关性 (degree correlation)、介数 (betweenness) 等指标进行研究. 其中, 节点的度分布用函数 $P(k)$ 描述, 表示一个随机选定的节点恰好有 k 条边的概率, 即度为 k 的概率 $P(k)$ 随节点度 k 的变化规律; 簇系数用于衡量复杂网络集团化程度, 即复杂网络中节点之间的连通程度, 或网络节点的局部传递程度; 平均路径长度指复杂网络中所有节点对之间的平均最短距离. 大量研究表明, 真实复杂网络具有较大的簇系数和较小的平均路径长度。

2.2 复杂网络社团特性

复杂网络中的社团结构, 使用较多的定义是: 网络中的节点可以分成组 (group), 组内节点连接稠密而组间节点连接稀疏. 如图 1 所示, 图中的网络包含三个社团, 分别对应于图中三个虚线圆圈包围的部分, 在这三个社团内部, 节点之间的联系非常紧密, 而社团之间的联系比较稀疏. 同时, 研究者还提出了强社团和弱社团的概念: 如果子图 V 中任何一个顶点与 V 内部顶点连接的度大于其与 V 外部顶点连接的度, 则为强社团; 如果子图 V 中所有顶点与 V 内部顶点的度之和大于 V 中所有顶点与 V 外部顶点连接的度之和, 则为弱社团. 与此同时, 为了表述更加严格意义上的社团, 研究者提出了另一个社团定义——LS 集合, 一个 LS 集合是一个由顶点构成

的集合, 它的任何真子集与该集合内部的连边都比与该集和外部的连边多^[2-3]。

目前, 在复杂网络社团方面研究较多的是社团的发现及划分算法, 除此之外, 对社团的模块度 (modularity)、层次性 (hierarchical structure)、重叠性 (overlapping communities)、计算复杂度、动态演化等研究也较多。

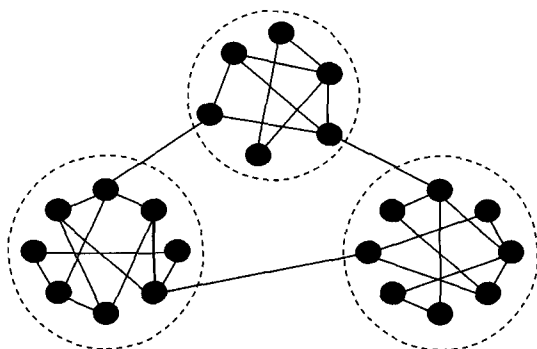


图1 复杂网络社团结构示意图

2.3 软件体系结构社团

根据 Garlan & Shaw 的定义: $SA=\{\text{构件}, \text{连接件}, \text{约束}\}$. 定义认为系统是有构件、连接件和约束构成. 其中, 构件 (component) 是一组代码或一个独立的程序, 构件相对独立, 仅通过接口与外部相互作用; 连接件 (connector) 是构件的粘合剂, 把不同的构件连接起来; 约束 (constrain) 是构件与连接件连接时的规则, 指明构件连接的势态和条件^[4-5]。

最近几年, 部分研究者借助复杂网络对大量系统的类图、模块结构图等进行研究, 发现软件系统的 SA 结构及内部通信过程既不是随机和无序的, 也不是规则的, 大多具有“小世界” (Small-World) 和“无标度” (Scale-Free) 等复杂网络的特征, 即具有较大的簇系数和较短的平均路径, 以及节点的连接呈指数分布或幂律分布. 另一个重要发现, 就是软件系统的 SA 具有复杂网络社团特征: 将 SA 划分为若干个群组 (group), 每个群组内部的节点之间的连接相对非常紧密, 各个群组之间的连接却相对比较稀疏^[6-7]。

本文作者及团队成员通过逆向工程对部分系统的 SA 结构进行研究, 将构件映射为复杂网络的节点, 连接件映射为复杂网络的边, 发现: 早期自由设计的软件规模较小, 没有遵循软件开发思想, 受程序设计者的个人影响较大, 软件呈现随机网络特征, 社团特征不明显; 而在开发中遵循了软件工程的模块化、层次化等开发设计思想的系统, 控制结构趋于严格, 具有规则网络的部分特性, 并具有复杂网络

的许多特征(如“小世界”、“无标度”等),并且社团特征明显.这一点正好与软件工程强调的“高内聚、低耦合”、“模块化”等思想吻合.同时,团队成员对部分系统的 SA 按照功能(函数)调用进行群组划分,发现其复杂网络社团特征尤为明显.

3 软件体系结构的社团特性

通过以上分析,为了简化研究,本课题假定构件和连接件的连接为无向连接,并且连接件不存在权值问题.所以,在建立 SA 的复杂网络时,将 SA 转化为无向无权复杂网络模型,其中,将 SA 中的构件映射为节点,连接件映射为边.

3.1 SA 的模块度

SA 的模块度用于衡量 SA 结构的网络划分质量,标记为 Q .一般情况下,一个软件系统的 SA 结构的网络划分 Q 值越大,则划分出的网络社团结构越好,越符合“社团内部连接紧密,社团间连接稀疏”的思想.

定义 1 SA 的模块度

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j) \quad (1)$$

式中, k_i 和 k_j 是节点 i 和 j 的度值; C_i, C_j 是节点 i 和 j 所属社团; m 是网络总边数.当 $C_i = C_j$ 时, $\delta(C_i, C_j) = 1$, 否则为 0. Q 值在 0 和 1 之间.研究及统计表明,一般以 $Q = 0.3$ 作为网络具有明显社团结构的下界.

按照软件工程思想,将 SA 中的构件、类等按照包的语义划分为多个包,包内的构件或类之间的联系紧密,而包之间的关系较松散(即用较少的、简单的连接件进行连接).其中,软件工程中的包的概念与复杂网络的社团概念一致.

3.2 SA 的层次性

复杂网络社团结构的层次性,指网络中的节点可能具有不同层次结构,大的社团内部可能包含较小规模的社团,而这些小社团还可能包含其他更小规模的社团,以此类推^[8].也可以做出网络的倒树图,在最底层,每一个节点看作一个社团,节点不断向上聚合从而形成较大规模的社团,最顶层则表示将整个网络看作一个大的社团.每一个较底层次的社团都被完全包含在高一层的社团中.不存在孤立的社团.通过选择在不同位置断开可以得到不同的网络社团结构.

这一特性正好与软件工程的“模块化”、“抽象与逐步求精”等思想相吻合.在分析及设计系统的 SA

结构过程中,为了减小因为问题的规模和复杂性不断增大而带来的困难,通过包机制,将系统划分为多个模块,包内节点联系紧密,包间关系松散;通过在不同抽象级别考虑和处理问题,用逐步求精的方法不断地分解,逐步确立过程细节,直至某一功能用程序语言描述的算法实现为止,具有不同详细程度的各种描述,组成了系统的层次结构.这与复杂网络社团的层次性特性一致.

在复杂网络社团研究中,研究者们提出了一些层次性社团结构查找算法,比如基于相似性矩阵,提出了聚类算法和分裂算法,在此基础上,提出了自上向下的分裂算法等.

3.3 SA 的重叠性

SA 的重叠性,是指在 SA 结构的复杂网络中存在一些“骑墙节点”,这些节点同时被多个社团所包含,属于这些社团的交叉部分.对应于软件工程中,在软件开发过程中,为了减少冗余,提高模块的复用率,将多个模块的公共部分提取出来,设计成一个公共模块,供其它模块调用,那么,这个公共模块就是这些模块(社团)的“骑墙节点”.

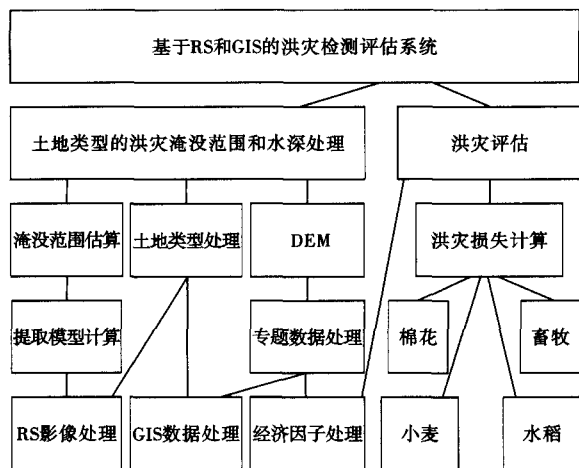
目前在复杂网络社团研究中,已经设计出了一些寻找重叠性社团的算法,如 Palla 于 2005 年提出了派系过滤算法 CPM(clique percolation method),在此基础上,研究者们提出了加权网络中的派系过滤算法 CPMw、有向网络中的派系过滤算法 CPMd、基于 GN 算法的改进算法等.

4 实例分析

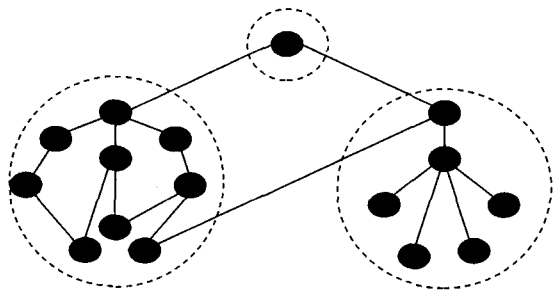
随着计算机技术、网络技术及 3S 技术的发展,新疆生产建设兵团(以下简称“兵团”)在国家科技部的大力支持下,结合兵团屯垦戍边职责及发展实际,开发了兵团区域空间信息服务系统,分别服务于兵团、各职能部门、师(局)、农牧团场、连队,实现各个部门对信息的及时了解及部门之间的信息共享.该系统包括农业、水利、土地、生态环境、公共安全等五个子系统.由于本系统非常庞大,作为实例,本文只选择水利子系统基于遥感(Remote Sensing, RS)和地理信息系统(Geographical Information System, GIS)的洪灾检测评估系统进行分析.基于 RS 和 GIS 的洪灾检测评估系统主要实现对洪灾区调查的基础上(洪水发生前的土地利用数据、行政界限、社会、经济、人口等数据),建立洪灾遥感监测快速识别模型,实现对不同土地利用类型的淹没情况的实时监测,为汛情的动态监测和评估系统提供

迅速直观、可靠的现状和变化信息。

通过对子系统分析,得出该子系统的顶层 SA 结构如图 2 所示。



在图2中,每个矩形框视为一个构件,构件的连线视为构件间的调用关系。将图2的SA结构图映射为复杂网络模型图,如图3所示。映射过程中,SA中的构件映射为复杂网络的节点,连接件映射为复杂网络的边。为了表述直观,将其按照复杂网络的社团结构进行划分。并将其划分为具有3个社团的网络模型。(注:根据不同的社团发现及划分算法,划分出的社团结构图可能有一些差距)



在图3所示的SA结构映射的复杂网络社团结构图中,利用式(1),计算得出该社团的模块度值约为0.5,介于具有明显网络社团结构特征值0.3~1之间。通过分析图2和图3可知,该SA结构具有明显的层次性,并且该层次结构还可以继续细化,直到得出具体可实现的最小构件为止。根据SA的重叠性分析,图3所示的网络结构中,3个社团共有3条边互相连通,这3条边所连接的4个节点(分别对应于“基于RS和GIS的洪灾检测评估系统”构件、“土地类型的洪灾淹没范围和水深处理”构件、“经济因

子处理”构件和“洪灾评估”构件等4个构件),就是“骑墙节点”,网络的重叠性特性明显。

5 结束语

本文基于复杂网络及复杂网络社团理论和方法,将SA结构转化为复杂网络模型,并通过SA的模块度、SA的层次性及SA的重叠性等进行SA结构的复杂网络社团特性研究。最后,通过兵团区域空间信息服务系统的水利子系统内的基于RS和GIS的洪灾检测评估系统为例进行分析,验证了其SA结构的复杂网络社团特性。实践表明,此方法能较好地对所开发系统的SA结构的复杂网络社团特性进行验证及度量,并能了解及分析所开发系统的网络结构及功能特征,为以后的系统设计、实现、集成、测试、维护等阶段提供依据。

参考文献:

- [1] 汪小帆,刘亚冰. 复杂网络中的社团结构算法综述[J]. 电子科技大学学报,2009,38(5):537-543.
- [2] 李晓佳,张鹏,狄增如. 复杂网络中的社团结构[J]. 复杂系统与复杂性科学,2008,5(3):19-42.
- [3] 刘晋霞,曾建潮,薛耀文. 复杂网络强社团结构探测[J]. 小型微型计算机系统,2011,32(4):686-690.
- [4] 赵佳华,伦立军. 软件体系结构层的耦合性度量研究[J]. 微电子学与计算机,2009,26(5):103-105.
- [5] 焦锋,王丽平,侯建民. 一种软件体系结构复杂性的评价方法[J]. 计算机应用研究,2008,25(8):2377-2379.
- [6] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of community hierarchies in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment,2008(10):10008.
- [7] Lancichinetti A, Fortunato S, Kertesz J. Detecting the overlapping and hierarchical community structure of complex networks[J]. New J. Phys. 2009(11):033015.
- [8] 刘鑫炎,黄美发,刘夫云,等. 基于复杂网络的产品族结构树的研究与实现[J]. 微电子学与计算机,2009,26(11):177-180.

作者简介:

秦怀斌 男,(1980—),硕士,副教授. 研究方向为软件工程、复杂网络。

梁斌 男,(1959—),硕士,教授. 研究方向为计算机软件与理论。

邵明文 男,(1974—),博士后,教授. 研究方向为事软件工程、粗糙集理论与应用。