

强/弱连接对软件网络结构的影响*

田鹤¹, 赵海², 蔡巍²

(1. 辽宁科技学院 工程实践中心, 辽宁 本溪 117004; 2. 东北大学 计算机科学与工程学院, 沈阳 110004)

摘要: 抽取类级软件网络拓扑模型, 提出用紧密度来测度网络中的强连接, 用结构洞来刻画网络中的弱连接。由紧密度与度的相关性得出高度值节点周围一定存在强连接, 低度值节点间不一定用强连接相关联; 由中介度与结构洞的相关性得出网络中的弱连接与节点占据结构洞有关, 无洞和全洞结构都不存在弱连接; 利用 K 近邻指标分析强/弱连接的分布特征, 得出强连接具有网络中心性, 而弱连接的中心性并不明显。通过逐步删除一定百分比的强/弱连接节点和边来对比分析它们对软件网络结构的影响, 结果表明移去强连接可迅速破坏网络的连通性, 而移去弱连接在一定范围内对网络结构影响并不大, 但随着删除比例增多, 网络的连通性和信息流动受到严重破坏。所得结论对度量和控制软件结构复杂性提供了理论指导。

关键词: 软件网络; 强连接; 弱连接; 结构洞; 中心性; 连通性

中图分类号: TP311 **文献标志码:** A **文章编号:** 1001-3695(2018)05-1412-05

doi:10.3969/j.issn.1001-3695.2018.05.028

Impact on strong/weak ties on structural of software network

Tian He¹, Zhao Hai², Cai Wei²

(1. Engineering Practice Center, Liaoning Institute of Science & Technology, Benxi Liaoning 117004, China; 2. School of Computer Science & Engineering, Northeastern University, Shenyang 110004, China)

Abstract: This paper extracted the class-level software network topology model, and then proposed using closeness to measure the strong ties and structural holes to depict weak ties in network. There must be strong ties around the nodes in height-value degree by the correlation between closeness and degree, but the nodes in low-value were not necessarily associated with strong ties. Weak ties were related to occupying structural holes of nodes in network by the correlation between betweenness and structural holes, there were no weak ties in the no hole structure and full hole structure. It reached conclusion that strong ties had network centrality, whereas weak ties were not obvious by analyzing the distribution characteristics of strong ties and weak ties by K-neighbor indicator. Gradually it removed a percentage of nodes who had strong ties or weak ties to contrast and analyze their impact on the software network structure, which showed that removing strong ties would rapidly damage the connectivity of network, while removing weak ties would not have a big impact within a certain range, but as the deletion proportion increasing, the connectivity of network and the flow of information were damaged seriously. The conclusions above provide theoretical guidance for measuring and controlling the complexity of software structure.

Key words: software network; strong ties; weak ties; structural hole; centrality; connectivity

0 引言

随着软件技术的发展和需求的不断增加, 软件的规模不断增大, 其内部结构越来越复杂, 软件研发过程经常处于失控的状态, 传统经验型的软件评估方法遇到了瓶颈^[1]。如何认识、度量、管理和控制软件的复杂性, 在尽可能短的时间内高效率地开发出高质量的软件, 是当今软件领域亟待解决的问题之一。对于现代大型软件来说, 庞大的软件规模和复杂的内部交互关系使得软件结构向着模块化和层次化的方向发展^[2], 以降低软件开发和维护过程的复杂度, 这就为复杂网络和软件工程的结合提供了一个契机。将软件系统结构内的类模块视为网络中的节点, 类之间的作用关系视为连接边^[3], 所形成的软件网络拓扑特征从宏观上能够直观地反映软件内部结构特性。汪北阳等人^[4]从节点属性方面对软件结构展开分析; Subelj 等

人^[5]分析类级依赖软件网络中的社区结构, 检测结果可用于获得高度模块化的软件包。以往的研究主要集中于网络的静态结构和动态行为。而另一方面, 网络通常被看成信息传播的通道, 于是相互作用的节点之间便会产生连接边。在网络中节点间连接边的关系和连接方式影响着网络的布局 and 性能, 软件网络的连接情况与软件结构内功能模块的实现和调用相关联, 充分了解整个软件网络的连接性并合理搭建网络连接架构是软件开发和维护的前提。

节点和连接边是网络拓扑中两个主要元素, 由于节点的复杂性使网络中的连接多种多样, 基于连接和交互的节点产生了种种关系, 并且这些关系连成了一个复杂的关系网络^[6]。早在 1973 年, 美国社会学家 Granovetter^[7]提出了“强弱关系”理论, 指出人们在社会交往中存在两种性质的人际关系, 即强关系和弱关系, 并就此开始深入研究, 强调了弱关系(连接)在信

收稿日期: 2017-05-25; **修回日期:** 2017-07-10 **基金项目:** 国家自然科学基金资助项目(60973022); 国家科技支撑计划基金资助项目(2012BAH82F04); 辽宁省科学技术计划基金资助项目(2015401039)

作者简介: 田鹤(1985-), 女, 辽宁沈阳人, 助教, 硕士, 主要研究方向为软件网络、复杂网络、软件定义网络(tianher@sina.cn); 赵海(1959-), 男, 辽宁沈阳人, 教授, 博导, 主要研究方向为复杂网络、软件网络、普适计算、信息融合; 蔡巍(1983-), 男, 辽宁沈阳人, 博士, 主要研究方向为软件网络、复杂网络、计算机网络。

息流动中所起的作用。在此基础上,Bakshy 等人^[8]从网络拓扑角度进行了拓展和深化,指出在社会网络中有效的社会协同并不会出现在密集互锁的强关系中,相反,弱连接比强连接具有更大的影响力。近年来,研究人员尝试将“强弱关系”理论逐渐渗入到大脑网络^[9]、经济网络^[10]和互联网的关系网络^[11]等其他复杂网络,结果对其产生了不一样的声音。相关文献表明在大脑网络中强连接构成了分形网络结构,由弱连接组成的最短路径导致了小世界现象;在产业集群等经济领域的关系网络中往往是强/弱连接互补协作、共同作用;而在互联网的关系网络中又有不同的表现,这里强连接和弱连接没有明显的区分,甚至可以迅速相互转变。软件网络是一种典型的复杂网络,将强弱关系理论思想融入到软件网络是一次全新的尝试,本文从其网络拓扑中强/弱连接特性的角度展开研究,在类级软件网络模型基础上,提出用紧密度来测度网络中的强连接并用 K 近邻把软件网络中的节点分类,分析强/弱连接的拓扑特征和网络中心性;用结构洞来刻画网络中的弱连接并用中介度来测度节点占据结构洞的能力;分别逐步移除强/弱连接节点和它们的边,并统计波及到的强/弱连接数目来研究它们对软件网络连通性的影响,对比分析强/弱连接对软件网络结构的影响力,从而更深入地了解软件网络结构特性,指导软件的设计开发。

1 理论与建模

1.1 类级软件网络拓扑构建

复杂网络是通过将研究对象简化为由节点和边组成的网络形式来研究真实世界网络的共性的一种方法。软件结构实质上表现为一种内容互连的复杂网络拓扑的形态,节点代表软件的各组成元素,边代表组成元素间的相互作用关系。在面向对象的软件技术中,类是对具有相同属性和行为的一个或多个对象的描述,是最基本的组成单元,于是采用逆向工程方法,从面向对象的开源软件系统的源代码中抽取类层次网络^[12],以 $G=(V,E,f)$ 定义类级软件网络模型,将软件系统结构内所有的类模块抽象为节点集合 $V=\{v_1,v_2,\dots,v_n\}$,各类模块之间的相互作用关系抽象为节点间的连接边的集合 $E=\{e_1,e_2,\dots,e_n\}$, f 表示为一种关系,如泛化、关联、依赖、聚合等。本文着重研究软件网络中节点间的连接状态,忽略纷繁复杂的实现细节,相应地在构建类级网络模型时将其抽象为有向无权网络拓扑图。如图 1 所示是软件系统结构的局部区域内类模块抽象成网络拓扑图的过程,该子网区域由 KA 至 KH 的八个节点类构成,彼此之间存在八条连接边。

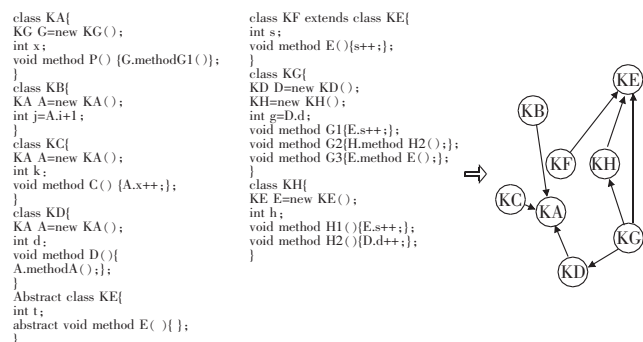


图 1 示例代码和相应的有向网络

1.2 理论基础与相关特征参数

社会网络的关系属性通常用连接强度来量化,一般来说,连接的强度可以是一定范围内的任意值,然而,为了简化概念,将社会网络中的所有关系归为强关系和弱关系两大类,引入到

网络拓扑中便形象化为强连接和弱连接。强连接是指那些关系亲密的个体,是一种紧密的人际关系,个体同质性较强,通常以小群体的形式出现^[13,14];弱连接则是一种不频繁的、疏远的弱关系^[15],是个体异质性较强的那部分连接。以此为出发点,纵观软件网络宏观拓扑结构,节点之间根据自身的属性有选择地连接,不同节点之间关联及交互作用程度是不一样的^[16],有些节点连接紧密,有些节点连接松散,使得整个网络的密度并不均匀,网络拓扑呈现出聚集性^[17]和层次性。通常,连接紧密的网络具有很高的连通性^[18],信息传递快,而连接稀疏的网络连通性很低,甚至存在很多孤立节点。由此,依据节点间不同的连接程度,把软件网络中的连接边分为强连接和弱连接。相关定义如下:

定义 1 强连接。由在网络中连接紧密的节点间产生,这些节点通常聚集在一定范围形成子网区域,彼此交互频繁。强连接维系下的网络区域的聚合程度较高。每一个子网区域都可视为软件结构中的类模块,具有相似的属性和结构信息。同时进一步说明了面向对象软件结构的模块化特征。

定义 2 弱连接。存在于不同的子网区域之间,是个体级别和群体级别之间的一种联系。它由单个节点产生,但是它们的存在不仅能影响该节点本身的状态和表现,也可以影响该节点所在的子网区域。

定义 3 紧密度。用于衡量网络中的节点通过网络到达其他节点的难易程度^[19],反映了一个节点对其他节点施加影响的能力。设网络具有 N 个节点,则节点 i 的紧密度为该节点到达其他所有节点的距离和的倒数。其计算公式为

$$C(i) = \left[\sum_{j=1}^N d_{ij} \right]^{-1} \quad (1)$$

其中: d_{ij} 是节点 i 到 j 的最短路径长度。紧密度值越大,说明与该节点相连的其他节点间的连接越紧密,该节点通过强连接与其他节点建立关系并频繁地相互作用。

弱连接通常联系着不同的群体,为不同群体间的信息交换提供了可能。强连接使节点间彼此携带的信息有重叠部分,网络中的节点通过弱连接能够跨过更多的子网区域传递更多的非冗余信息。在社会网络中一个重要概念就是结构洞理论^[20],所谓的结构洞是指非冗余联系人之间的缺口^[21]。由于结构洞的存在,两边的联系人可以获得累加的非重复的资源和信息。这样看来,弱连接与结构洞似乎在描述同样的现象,类似地,在软件网络中,结构洞看起来就是存在于两个没有紧密联系的节点集合之间的“空地”,跨越结构洞的节点能够获得异质性的信息^[22]。如图 2 所示,三条虚线为弱连接,去掉任一条都会产生结构洞。节点 v_6 连接着节点 v_5 和 v_7 ,如果 v_6 不存在,那么 v_5 和 v_7 就不能传递信息,也就是说 v_5 和 v_7 之间出现了结构洞,它们之间是通过弱连接交换信息,节点 v_6 不仅可以获得这两个子网区域的所有信息,还会整合这些信息,同时为两个子网区域之间的信息交流提供便捷,提高了传递效率。结构洞的观点就是强调要将弱连接存续下去。

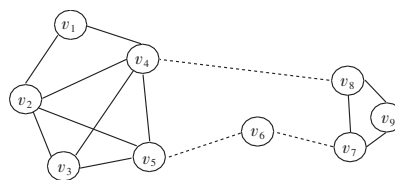


图 2 结构洞与弱连接

定义 4 约束度^[23]。它是结构洞的量化指标,用来衡量一个节点在网络中拥有结构洞的能力。将 Burt^[20]提出的约束度算法引入到软件网络中,定义其计算公式为

$$\text{constr}(i) = \sum_j \left(\frac{d_{ij} + d_{ji}}{\sum_k (d_{ik} + d_{ki})} + \sum_{k, k \neq i, k \neq j} p_{ik} p_{kj} \right)^2 \quad (2)$$

其中: p_{ij} 是 i 与节点 j 之间的最短路径长度同 i 与其所有邻接节点最短路径长度总和之比; d_{ij} 是指节点 i 到 j 的最短路径长度; k 是节点的邻接节点。节点的约束度越高, 形成结构洞的机会越小。若 i 为孤立节点, 则这个节点不存在结构洞, 则 $\text{constr}(i) = 1$ 。

定义5 中介度^[24]。它用来衡量网络中节点对资源控制的程度, 如果一个节点处于许多其他节点对之间的最短路径上, 则该节点具有较高的中介度, 处于重要地位, 同时拥有较多的结构洞。在软件网络拓扑 $G = (V, E, f)$ 中, 对任意节点 v , 定义中介度计算公式为

$$B(v) = \sum_{i \neq v \neq j \in V} \frac{\delta_{ij}(v)}{(N-1)^2} \quad (3)$$

其中: $\delta_{ij}(v)$ 为二值变量, 如果节点 i 到 j 的最短路径中经过节点 v , 那么 $\delta_{ij}(v) = 1$; 否则 $\delta_{ij}(v) = 0$; N 为网络节点总数。

定义6 最短路径长度。在软件网络拓扑 $G = (V, E, f)$ 中, 任意两个节点 i 与 j 之间最短路径 d_{ij} 定义为所有通路中连通这两个节点的最少边数。通常用来衡量节点间信息传递效率。

定义7 K 近邻。它是指网络中以某节点为起点最少经过 K 步路径到达的节点数目。按照不同的路径长度将到达节点分类, 经过相同路径长度到达的节点归于一类, 从自身类开始依次编号, 用 0 表示节点本身。

2 软件网络的连接性

结构洞形象地刻画了网络中的弱连接, 占据结构洞的节点间体现了一种弱关系, 通过弱连接相互传播信息。在软件网络拓扑结构中, 紧密依赖的节点通常呈现出集聚现象, 使具有相似特征或功能的节点聚集在一个子网区域, 这种局部连接分布特征可以结合度指标来分析。而另一方面, 如果一个子网区域的节点和另一个子网区域有连接, 那么这个节点应该靠近和它有连接的节点所在的子网, 其间经过的节点便很有可能占据着结构洞的位置, 中介度指标能够准确地度量节点占据结构洞的能力。以 Abiword 软件为例分析强/弱连接的特征, 如图 3 所示。

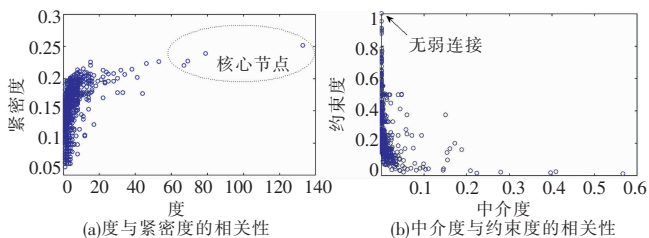


图3 软件网络中强/弱连接特性

从图 3(a) 可以看出, 紧密度大的节点并不一定有很高的度值, 而高度值的节点间一般都互相紧密依赖着。这样, 软件网络拓扑结构中的节点按照度值大小分为两类三种情况。第一类是度值很高的节点, 它们一定具有很强的连接性, 这些节点与周围节点通过强连接紧密聚集在一起, 支配着整个网络的信息传递, 实现软件的核心功能, 然而这类节点很少, 使软件结构复杂性不至于过高, 易于软件的维护; 第二类是度值很小的节点, 这类节点并不一定具有很强的连接性, 如果节点偏向网络拓扑的中心位置, 它们彼此间用强连接关联, 并且紧密连接着其他类模块以实现软件的主要功能; 如果节点处于网络拓扑的边缘, 那么它们不具有很强的连接性, 对应的类模块在软件结构中往往只关注自身的内部逻辑。因此, 高度值节点周围一定存在强连接, 低度值节点间并不一定用强连接相关联。

从图 3(b) 看出大部分节点的中介度都非常小, 数值集中在 0.1 以下, 说明通过这些节点的路径都很少, 节点占据的结构洞相对较少, 通过这些节点的弱连接很少, 极端的情况下随着中介度越来越小, 节点约束度趋近于 1, 这时节点周围不存在结构洞, 封闭的小网络无须弱连接。另外很明显地看到有几个节点的中介度非常大, 但约束度很小, 说明这些节点占据着较多的结构洞, 拥有较多的弱连接, 它们携带着很多关键信息, 通过弱连接传播到其他子网。一个极端特殊的情况是当约束度为 0 时, 出现全洞结构, 此时是一个典型的星形网络, Hub 节点的紧密度和度值最大, 通过强连接与周围的节点传递信息, 节点间并不存在弱连接。由此看来, 无论是无洞网络结构还是全洞网络结构都是不存在弱连接的。深入软件内部结构进一步对比分析软件网络拓扑特征参数值, 将 Abiword 软件网络中节点的各项特征参数测度值按降序排列。统计出四个测度值的关系有 10 种, 选取有典型代表的节点列于表中, 如表 1 所示。

表1 Abiword 软件网络中测度值的排名

节点序号	节点名称	节点度排名	强连接度排名	约束度排名	中介度排名
1605	UT_GenericVector	1	1	1 990	1
220	_wvExporter	262	8	1 756	8
711	ct_data_s	623	1 458	1 034	453
1117	IE_ExpSniffer	11	262	1 984	38
953	fp_Page	25	62	1 907	233
1994	XML_ParserStruct	27	1 568	1 935	729
1176	ie_imp_table	262	35	1 462	375
453	AP_Preview_Abi	150	20	1 804	186
1031	gzfilebuf	645	1 633	1 042	894
733	DTD	258	1 571	1 770	777

从表 1 可以看出, 大多数节点约束度排名都偏后, 这是因为网络中有一部分闭合子网和孤立节点。最有特殊性的节点是第 1605 号节点, 三个测度值都是最大的, 约束度却非常小, 说明 1605 号节点处于整个软件网络的中心, 其周围有很多相邻节点通过强连接紧密相连, 查找数据文件可知 UT_GenericVector 所在的子网区域共有 108 个相同类型不同功能的节点, 密集地分布在网络中心, 具有很高的中介度, 说明此节点占据着很多结构洞, 与其他子网区域中的节点有大量的弱连接, 是子网之间信息传递的关键节点, 如果去掉这个节点, 那么整个软件网络一定会粉碎。所以, 这样的节点有很大的影响力, 且构造比较复杂, 在设计软件结构时, 可以适当将其边缘化或者将子网中的节点类再分散。然而, 由节点度值的排名看到, 处于网络中心的节点虽然具有很强的连接性, 与其直接相连的节点不一定很多, 这与节点自身的逻辑属性和功能有关。

3 实证分析

选择四款具有代表性的面向对象的开源软件作为研究对象实例, 即用 C++ 开发的 Abiword、Mozilla 软件和 Java 开发的 Freemind、Eclipse 软件, 相应的软件网络中, Freemind 有 713 个节点, Abiword 有 2 002 个节点, Mozilla 有 8 356 个节点, Eclipse 有 12 583 个节点, 涵盖了小、中、大三个规模。通过计算特征参数值并分析它们的软件网络数据结构文件中节点和连接边的属性来深入了解软件系统结构的特征。

3.1 强/弱连接的中心性

K 近邻指标以节点间的路径为依据, 将软件网络中的节点分类, 所得结果既是节点间的距离, 也是节点类别。将四款软件网络中除去自身节点和“孤岛”节点(即孤立节点和不可达

节点),利用K近邻与其他参数的相关性可分析两种不同开发语言的软件网络中强/弱连接的分布特征,如图4所示。

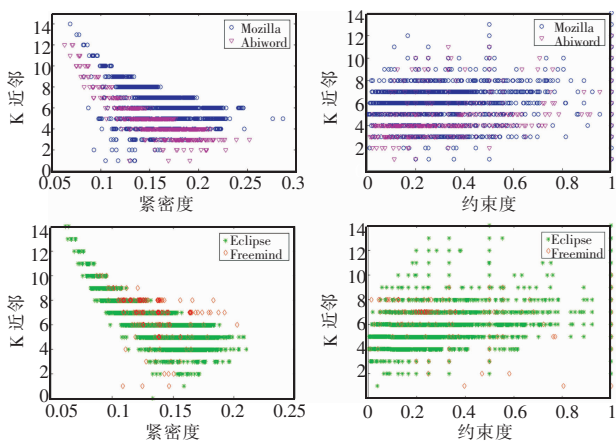


图4 软件网络中强/弱连接分布特性

从图4中可以直观地看到, Freemind、Abiword、Mozilla 和 Eclipse 软件网络的近邻节点类分别有 10、12、14 和 14 个,同时表示网络中从某一节点出发到达其他节点的最短路径长度的最大值,而无论选定哪个节点为出发点,它们的分布情况都很相似。进一步来看,四款软件网络的近邻节点类大多数都分布在路径长度的中间值附近,并且随着路径长度的不断增大,分布在每个近邻节点类的节点数目越来越少,说明软件网络拓扑具有中心化现象。由 K 近邻与紧密度分布图看到,聚集在网络偏中心区域的节点紧密度都比较大,说明这类节点间通过强连接相关联,它们彼此紧紧地依赖着,对应的软件结构是比较复杂的;由 K 近邻与约束度分布图看到,弱连接分布倾向于网络的中心,但相比较强连接而言分布相对散乱,在各类近邻节点间都会存在。特别地,虽然已经除去孤岛节点,还会存在 $\text{constr}(i) = 1$ 的情况,这些少量节点存在于封闭子网结构中,度值是非常小的,可能为 1 或 3,零星地分布在网络各个区域。而不论是强连接还是弱连接,对于近邻节点类别数值大的节点一般都非常少,到达这些节点的路径较长,它们之间并不简单地用强连接相关联,但也并不一定以弱连接方式相连,具体情况还要考虑节点的属性、功能和行为。

综合对比分析,在软件网络拓扑结构中,强连接具有一定的中心性,聚集分布在网络的中心,而弱连接并不具有明显的中心性,可以出现在网络中的各个地带。强连接与弱连接并非此消彼长的,集中在一个子网区域内的节点紧密相连,它们通过强连接传递信息共同完成模块的整体功能。网络中的弱连接随着结构洞的分布以及子网区域之间的作用关系而存在,对于软件结构内不同类型的模块之间,复杂的模块主要由相对简单的模块构成,而不是直接调用其他复杂的模块。弱连接为模块间的有效信息传递提供了捷径,在减小模块间的依赖性同时增加了模块间的交互,这种连接特性进一步验证了软件的“高内聚,低耦合”的设计原则。

3.2 强/弱连接的网络影响力

软件网络中的主体是不同类型的节点,节点的属性和特征决定了它们相互作用的方式,有效地实现软件的功能离不开各类模块间恰当的协调和调用,这种作用关系程度便体现在网络的强连接性或弱连接性上。单独分析节点本身的特性并不能全面地表达其在网络中施加影响的能力,每一次连接边的产生都意味着发生一次关系,而这种作用关系的强弱和波及范围便

体现在节点的强/弱连接在网络中的影响力。

大量的连接边搭建起来形成网状图,网络结构一个重要评价指标就是连通性,网络功能的实现依赖节点的连通性,若两个节点之间存在一条路径使其连接,信息则可以在两个节点间传播,甚至两个子网区域传播。然而,软件网络是动态变化的,在对软件维护和更新过程中需要对软件结构调整和完善,所以软件网络中节点或连接边随时都有可能更改或失效。节点的删改随之带来的是连接边的失效,从而破坏了网络的连通性,其破坏程度与节点间的连接程度是密切相关的,评估强连接与弱连接对网络的影响力就是当分别移去强连接和弱连接后对整个网络连通性造成的影响。对网络连通性的研究首先要获得网络拓扑的最大连通子图。以复杂网络中“富人俱乐部连通性”为指导思想,以高度值节点为基准,结合广度优先遍历方法寻找最大连通子图,同时统计四款软件网络中节点度值和 K 近邻的数据,去除度值为 0 的节点和不可达节点,抽取 Freemind、Abiword、Mozilla 和 Eclipse 软件网络的最大连通子图中节点数量分别为 664、1 567、6 555 和 10 375 个,连接边数分别为 1 345、2 774、13 251 和 22 117 条。依照强/弱连接程度的不同,蓄意删除一定比例的节点及其所拥有的强/弱连接边,在此基础上开始实验设计和分析,过程如下:

- a) 分别将强/弱连接按连接程度降序排列;
- b) 按比例分别移除网络中 1%、2%、3%、4%、5%、... 的节点,连同它们的强/弱连接一起移去,直至网络不能再分割;
- c) 每一次移除,都分别统计波及的强/弱连接的数目;
- d) 分析这样的移除过程对网络连通性的影响。

实验结果如图5所示。观察四个图中展现的强/弱连接的变化,当按比例删除强连接节点时,受影响的强连接数目总体趋势是减少的,尤其在最开始的 1% ~ 3% 区间曲线下降的斜率非常大,仅删除极小部分强连接节点就能波及大量的连接边,使网络连通性迅速遭到大面积破坏。这是由于网络中心的节点间存在着大量的强连接,移去网络中心的节点后,这些节点的强连接便随之移除,所以在对强连接节点逐渐移除的过程中,曲线最开始下滑的斜率非常陡峭,网络的中心性很快被削弱,对整个网络连通性的影响非常大。曲线骤然下降后会阶跃跳变,如图中 Abiword 软件中 6% ~ 9% 区间, Freemind 软件中 4% ~ 5% 区间, Mozilla 软件中 3% ~ 5% 区间, Eclipse 软件中 7% ~ 9% 区间尤为明显,产生这一现象的原因是这部分节点类的自身构造逻辑比较复杂,模块间的重用比较多,却又不是主要功能的类,这些节点及其所带的强连接在网络中的位置是偏离中心的,影响范围也仅限于它们自身所处的子网区域,所以强连接虽然具有中心性,但连边的多少还要取决于节点的类别和功能。随着删除节点数目的增加,波及范围相对减少,曲线下滑缓慢直至所剩无几,这是因为前期的攻击已经使网络遭到重创,所影响的强连接边数目的逐步减少主要是由于网络连通性迅速降低所致,整个网络快速被打散。

观察弱连接变化曲线,开始时删除一定百分比的节点后并没有大幅变化,曲线是缓慢增长的,但是随着移去节点比例的增加,波及到的弱连接数目突然增多,这是由于在最开始的连通子网中是没有结构洞的,移除的只是占据结构洞的节点,而随着移除节点百分比增多,便出现了结构洞,产生弱连接。随着结构洞的增多,波及的弱连接增多,而这样的现象是稍纵即逝的,当结构洞增多到一定程度甚至是出现全洞结构时,弱连

接减少,直至移尽,所以网络在一定破坏程度上,移去弱连接并不会对网络连通性造成很大的影响。随着移去弱连接节点数目的增多,影响的弱连接的数目也会增多,虽然存在连通网络,但是这些剩余的连通网络逐步形成一个个碎片,信息交流受到阻碍,达到一定程度也会破坏网络拓扑结构。

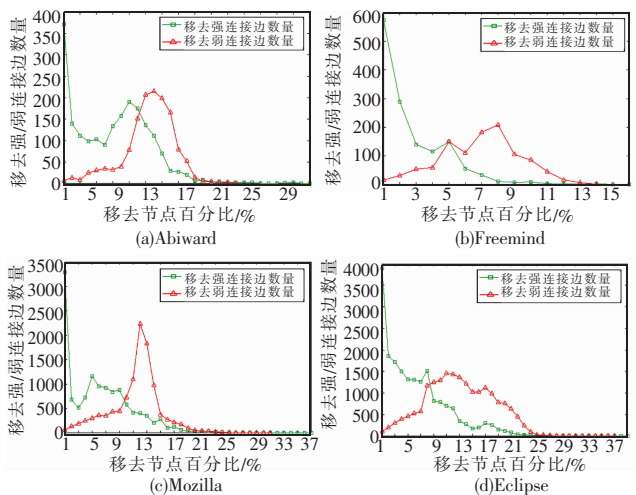


图5 强/弱连接对软件网络连通性影响的对比

无论是强连接还是弱连接,随着节点移除比例的增多,所波及的连边总体趋势都是下降的;尽管在此过程中都会出现突变现象,但只是暂时的。软件作为一种复杂的人工智能化系统,组织结构内单元模块的组成和实现直接影响软件的功能、性能和可靠性等指标。强连接与弱连接联系着节点(网络)的结构和活动,连接属性和方式的改变直接影响节点自身的变化及网络的变动。强连接主要集中在网络的中心区域,只需移除少量节点和边就可迅速破坏网络的连通性,对应着软件中大量的类模块结构被改变,使软件结构被破坏,在软件维护过程中要注意强连接的节点,轻易不要改动。而小比例节点和弱连接删改对网络的影响范围并不大,主要是由于弱连接承载着节点(网络)的信息流动,网络结构和功能大体可以维持,但随着节点和连边逐渐删去,结构洞增多,弱连接随之增多且也很快被移除,网络连通性严重受损,最后网络支离破碎。

4 结束语

研究软件网络中的强连接和弱连接有益于软件开发人员了解软件结构内各类模块间的作用关系。本文将强/弱连接理论引入到软件网络,从网络连接性的角度来深入了解软件内部结构,构建类级软件网络拓扑模型,在此基础上对网络拓扑特征量进行分析:利用紧密度来测度网络中的强连接;用结构洞来刻画网络中的弱连接;用K近邻指标分析强/弱连接的中心性。最后,逐步删除一定百分比的强连接与弱连接节点和边来对比分析它们对软件网络结构的影响。所得结论如下:

a) 强连接节点的度值不一定很大,但一定位于网络的中心;占据结构洞的节点通过弱连接与其他节点或子网传递信息。

b) 强连接具有网络中心性,其节点密集分布在网络的中心,节点间的紧密连接促使形成大量的子网区域;而弱连接相对来看中心性并不明显,伴随着网络中结构洞的出现及子网区域的信息传递而产生,在此路径的节点起到“桥梁”的作用,承担着重要责任。

c) 强连接能够迅速破坏网络的连通性,相比之下弱连接在一定范围内对网络结构影响是循序渐进的,但网络中信息传

播严重受到阻碍,长此下去软件结构依然会崩溃。

由此来看,在对软件结构更改时应适当调用弱连接节点类模块,避免对强连接节点类模块的大量改动,从而合理布局软件结构,高效实现软件功能,优化软件结构的设计。

参考文献:

- [1] Chidamber S R, Kemerer C F. A metrics suit for object-oriented design [J]. *IEEE Trans on Software Engineering*, 1994, 20(6): 476-49.
- [2] 李辉, 赵海, 徐久强, 等. 基于k-核的大规模软件宏观拓扑结构层次性研究 [J]. *电子学报*, 2010, 38(11): 2635-2643.
- [3] Li Hui, Zhao Hai, Cai Wei, et al. A modular attachment mechanism for software network evolution [J]. *Physica A*, 2013, 392(9): 2025-2037.
- [4] 汪北阳, 吕金虎. 复杂软件系统的软件网络节点影响分析 [J]. *软件学报*, 2013, 24(12): 2814-2829.
- [5] Subelj L, Bajec M. Community structure of complex software systems: analysis and applications [J]. *Physica A*, 2011, 390(16): 2968-2975.
- [6] Shirky C. *Cognitive surplus: creativity and generosity in a connected age* [M]. Beijing: China Renmin University Press, 2010.
- [7] Granovetter M S. The strength of weak ties [J]. *The American Journal of Sociology*, 1973, 78(6): 1360-1380.
- [8] Bakshy E, Rosenn I, Marlow C, et al. The role of social networks in information diffusion [C]//Proc of the 21st International Conference on World Wide Web. New York: ACM Press, 2012: 519-528.
- [9] Gallos L K, Makse H A, Sigman M. A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks [J]. *Proceedings of the National Academy of Sciences*, 2012, 109(8): 2825-2830.
- [10] 包丽华, 李南, 冯夏宗. 产业集群的强连接与弱连接现象 [J]. *经济视角*, 2007(5): 54-56.
- [11] Hsu M H, Chen H H. Efficient and effective prediction of social tags to enhance Web search [J]. *Journal of the American Society for Information Science and Technology*, 2011, 62(8): 1473-1487.
- [12] 王莹, 于海, 朱志良. 基于软件节点重要性的集成测试序列生成方法 [J]. *计算机研究与发展*, 2016, 53(3): 517-530.
- [13] Tao Haicheng, Wang Youquan, Wu Zhi'ang, et al. Discovering overlapping communities by clustering local link structures [J]. *Chinese Journal of Electronics*, 2017, 26(2): 430-434.
- [14] Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior [J]. *Proceedings of the National Academy of Sciences*, 2013, 110(15): 5802-5805.
- [15] 邵明星, 颜志军. 弱连接关系在SNS平台中影响作用的实证研究 [J]. *北京理工大学学报*, 2015, 17(2): 84-89.
- [16] Subelj L, Bajec M. Software systems through complex networks science: review, analysis and applications [C]//Proc of KDD Workshop on Software Mining. New York: ACM Press, 2012: 8-15.
- [17] 李博, 赵海, 王家亮, 等. 软件网络中的层次和群落 [J]. *东北大学学报: 自然科学版*, 2010, 31(6): 799-802.
- [18] Xu Guo'ai, Gao Yang, Qi Yana, et al. Contrastive analysis of software networks based on different coupling relationships [J]. *China Communications*, 2010, 7(4): 76-82.
- [19] Li Hui, Zhang Hui, Zhao Hai, et al. Research on structural holes and closeness of multi-granularity software networks [J]. *Journal of Software*, 2013, 8(2): 337-343.
- [20] Burt R S. *Structural holes: the social structure of competition* [M]. Boston: Harvard University Press, 1992: 53-58.
- [21] 苏晓萍, 宋玉蓉. 利用邻域“结构洞”寻找社会网络中最具影响力节点 [J]. *物理学报*, 2015, 64(2): 020101.
- [22] 刘倩倩. 加权网络的结构洞占据者获取方法的研究 [D]. 合肥: 安徽大学, 2015: 9-13.
- [23] Li Bo, Zhao Hai, Cai Wei, et al. A new metric method-improved structural holes researches on software networks [C]//Proc of International Conference on Graphic and Image Processing. 2013.
- [24] 李博, 赵海, 刘红, 等. 大规模软件网络类节点中介度分析 [J]. *东北大学学报: 自然科学版*, 2010, 31(7): 929-932.