

数据库复杂网络构造算法及特征分析

李春芳^{*①③} 刘连忠^② 刘振国^③

^①(北京航空航天大学自动化科学与电气工程学院 北京 100191)

^②(北京航空航天大学计算机学院 北京 100191)

^③(河北体育学院网络中心 石家庄 050041)

摘要: 数据库复杂网络是对管理信息系统软件复杂性的一种度量, 提供了一种精简的可视化业务逻辑描述, 为软件增量设计提供了自说明文档。为抽取数据库复杂网络, 提出了两种算法: (1) 基于数据表主外键关联的网络构造算法; (2) 基于隐性语义关联的构造算法及扩展算法。通过对 9 个软件数据库的分析, 研究了该类网络的统计特性, 发现其无标度特性入度分布比出度分布更不均匀, 入度大的节点是系统的骨干节点。为精准构造直接辅助软件工程的数据库复杂网络, 提出了针对语义隐性关联的一组命名规范。

关键词: 管理信息系统; 复杂网络; 软件网络; 数据库网络

中图分类号: TP311

文献标识码: A

文章编号: 1009-5896(2012)11-2700-07

DOI: 10.3724/SP.J.1146.2012.00491

Algorithms and Features Analysis of Database Complex Networks

Li Chun-fang^{*①③} Liu Lian-zhong^② Liu Zhen-guo^③

^①(School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China)

^②(School of Computer Science and Engineering, Beihang University, Beijing 100191, China)

^③(Network Center, Hebei Institute of Physical Education, Shijiazhuang 050041, China)

Abstract: DataBase Complex Networks (DBCN) is a kind of metric for management information systems, which provides a simplified and visualized description of business logic and a self-introduced documentary. To extract DBCNs, two algorithms are proposed: (1) based on the primary and foreign key associations of database tables algorithm; (2) based on the hidden semanteme associations algorithm and its extension. Through the analysis on 9 software databases, the statistical features of DBCN are investigated and found that in-degree distribution is more disperse than that of out-degree, and tables with greater in-degree are the backbone nodes. In order to accurately construct DBCN to facilitate the software engineering, a group of naming criteria is proposed for hidden semanteme associations.

Key words: Management Information System (MIS); Complex networks; Software networks; Database networks

1 引言

数据库是各类管理信息系统 (Management Information System, MIS) 的数据核心。随着软件规模扩大, 数据库的规模和关联复杂性相应增加。软件网络是软件复杂系统的骨架, 而数据库复杂网络是业务系统内部关联性的骨架。在顶层设计中, 主要是将业务内部的关联性映射为数据表之间的关联性, 尽管业务系统的功能不可能全部在数据库设计中实现, 但是显而易见, 优化的数据库关联设计,

能尽可能多地覆盖业务逻辑, 减少系统实现的复杂性和代码量。由于在数据表的基本字段确定后, 非主键和非外键字段的添加和删除, 不会影响其他的关联性, 因此通过主键和外键建立数据表之间的复杂网络, 作为对 MIS 系统宏观层面的描述是合理的。

从文献看, 复杂网络的解释为, 从实际复杂系统抽取的网络结构, 具有明显的“无标度”和“小世界”特性, 其网络特征介于随机网络和规则网络之间, 即不是完全随机, 也不完全规则。软件是一类人工复杂系统, 复杂网络理论引入软件工程来描述和度量软件的复杂性, 形成了软件网络^[1]。软件网络主要探索了软件包级^[2]、类级^[3-6]、方法级^[4-6]的网络构造和统计特征, 以及软件的动态运行环境网络^[7], 从复杂网络角度探索了软件的宏观结构和软件

2012-04-26 收到, 2012-09-03 改回

国家 863 计划项目(2005AA113040)和河北省高校自然科学基金(Z2010279)资助课题

*通信作者: 李春芳 LCF1974@126.com

度量^[8,9]以及软件网络的应用^[10-12]，然而对软件网络的分支数据库复杂网络的研究较少。何克清等认为软件网络没有被准确定义，泛指从软件系统中抽取的网络结构，该结构具有复杂网络的特征^[1]。以数据库中表作为节点，将外键到主键形成的参照依赖作为有向边，MIS 系统的数据库可以建模成一个有向图，本文发现该结构也具有“小世界”和“无标度”特征，可以认为，数据库复杂网络是软件网络的一个分支，笼统指从数据库中抽取的网络结构，研究数据层关联形成的网络统计特性。

早在 1983 年，文献 [13] 利用超图 (hyper networks) 超网络来描述数据库，其节点是一个数据模式的所有数据表的字段集合，一个数据表定义为一个超边 (hyper edge)，它能连接多于两个节点。由于超图的可视化信息不便于直观理解，远不如复杂网络描述数据库内部关联更直接。

尽管一个 MIS 系统的数据库网络可能仅有几十到上千的节点，但是其关联复杂性足以让软件开发人员难以应对，需要 CASE (Computer Aided Software Engineering) 工具辅助描述其复杂性。CASE 工具，如 Rational Rose, PowerDesigner, Microsoft Visio, Visual Paradigm 等均采用了数据库关系图描述数据表间的关联，但是其抽象粒度不够，开发人员无法看到数据库网络更宏观的视图，难以驾驭大型系统的数据库设计。数据库关系图没有从复杂网络的角度统计其网络结构特性，本文从更抽象的复杂网络的角度研究 MIS 系统数据模式的关联，以更加“骨感”的形式来描述和度量数据层的复杂性，使软件开发人员能掌控数据关联的概貌。

2 形式化表示

数据库完整性指数据的正确性和相容性，完整性包括域、实体、参照和用户定义完整性，其完整性设计就是约束设计。在大型 MIS 系统设计中，参照完整性是核心，它集中体现了业务逻辑，也是软件设计复杂性的根源之一，本文将数据表间的参照完整性建模为复杂网络，以更简洁的方式描述其复杂关联性。完整性约束可以通过数据库管理系统 (DataBase Management System, DBMS) 或应用程序实现，通过 DBMS 实现的参照完整性即主外键关联，而由应用软件实现的参照完整性本文定义为语义隐性关联，简称语义关联，它隐含表示了数据表主外键之间的参照关系。

大量的开发经验表明，外键关联和语义关联各有优势，表 1 列出了两种关联的比较，在本文分析的 9 个数据模式中，4 个采用了外键，5 个采用了隐

表 1 外键关联和语义关联比较

比较项	外键关联	隐性语义关联
跨表查询	性能差	速度快
管理	集中	分散
故障发生	能够保证数据的一致性	可能导致数据不一致
代码开发难度	降低软件设计的复杂性，外键羁绊代码实现	自主维护参照完整性，增加软件设计复杂性
支持 CASE	支持	部分支持

性语义关联。外键关联采用数据库集中管理数据关系，在海量数据查询情况下跨表查询性能比语义关联差。由于 DBMS 能自动维护数据的完整性，因此在故障发生时，能够保证数据的一致性，而语义关联可能会引入不一致数据。在开发阶段，由于系统的不完备性，外键会羁绊代码实现，而语义关联需要增加代码来维护参照引用数据的一致性。CASE 工具支持外键关联和部分语义关联的图结构分析。

大部分的 DBMS 系统和逆向工程工具都采用了数据库关系图，其节点是数据表结构信息，边是外键到主键的关联依赖，因此在数据库复杂网络构造上用数据表作为节点，外键关系作为边是合理的。在已有的数据库关系图中，节点信息复杂，然而对复杂性起关键作用的是主键和外键，其他非主键和非外键字段的添加和删除对关联复杂性没有影响，因此进一步化简节点是可能的。

数据库复杂网络 (DataBase Complex Networks, DBCN) $G = \{V, E\}$ 是从数据库模式中抽取的有向网络模型， $V = \{v_1, v_2, \dots, v_n\}$ 是节点集，代表数据库模式中的数据表， $E = \{(v_i, v_j) | i, j = 1, 2, \dots, n\}$ 是有向边集合，代表一个数据表外键 (或隐含外键) 到其它数据表主键的依赖。

基于复杂网络的统计参数^[1]，定义如下特征参数：

- (1) 节点数 N ：数据表个数；
- (2) 边数 E ；
- (3) 平均最小路径长度 L ；
- (4) 平均聚类系数 C ；
- (5) 平均出度/入度 D_i/D_o ；

(6) 节点对可达率 R ：定义为有向网络上任意不同的两个节点之间存在的可达路径总数除以 $N \times (N-1)$ ， R 是对内部关联紧密程度的一种度量，网络内聚度越高， R 越大，如果系统内部存在级联删除和级联更新，在数据库内部传递的可能性越大；

(7) 网络连接率 S ：定义为最大连通图上的节点数除以总节点数。

在 DBCN 上, 节点的入边是其它数据表的外键(或隐含外键)字段到该数据表主键的引用, 节点的出边是该数据表的外键(或隐性外键)到其它数据表主键的引用。数据库主键和字段的关系是一对多的关系, 主键可能是一个字段, 或者多个字段, 即联合主键, 绝大多数是一个字段, 在联合主键情况下每个独立字段都是外键, 即到其他表主键的引用。因此入边都是对节点主键的引用, 而出边是对其他节点主键的引用。

由于一个数据表中可能有多于一个字段引用其它表的同一个主键字段, 因此 DBCN 是一个可能包含重边的有向图, 在以下的构造算法中去掉了重边, 只考虑简单有向图构造的网络。

3 构造算法

3.1 基于主外键的构造算法

本节研究了基于复杂网络的数据库关联分析技术, 实现对 Oracle, MySQL, SQLServer 数据库中表的元数据的关联分析。针对基于数据库主外键的关联和无外键的语义隐性关联提出了两种网络构造算法。

算法 1 基于主外键的构造算法(FK-DBCN)

(1) 查询系统表读取一个数据模式下的所有数据表名称, 作为网络节点信息;

(2) 查询系统约束表, 读取数据模式的所有外键约束, 作为从一个数据表节点(外键)到参照表节点(主键)的有向边信息。

用算法 1 构造了非开源软件学校管理系统 Yundl 的数据库网络, 其 DBMS 为 SQLServer, 它采用了外键实现参照完整性约束, 显示度排名前三位的节点是 tbStudentBasis, tbTeacherBasis 和 tbYearBasis, 反映该系统中所有的关系中最重要的是学生、教师在学年中的各种关系, 这种直接宏观的业务逻辑表示对于变更的开发团队、软件的二次开发以及更新维护非常重要, 可以缩短增量开发时间。

3.2 基于语义关联的构造算法

由于 DBCN 内部关联的复杂性, 很多的软件开发实践表明, 业务的主、外键关联往往通过程序内部映射实现, 而不采用 DBMS 本身的完整性约束, 从而避开某些关联控制的难题。实际上, 无论在软件的设计还是实现上, 关联引起的复杂性无法回避。开源软件网络与主机监控系统 Zabbix 的数据库关联采用了 MySQL 数据库, 内部无外键约束。从对 Zabbix 的分析中发现, 尽管没有主外键, 但是从规范的数据表字段命名规则可以推理语义隐性关联, 进而推理出关联关系, 而 CASE 工具 PowerDesigner

也采用这种语义推理。例如: 在 Zabbix 数据库中有两个数据表 History 和 Items, 其中 History 有 3 个非主键的字段 itemid, clock 和 value, 在 Items 数据表中仅有一个主键 itemid, 那么可以推出表 History 到表 Items 有一个隐性外键关联。根据数据表的字段命名规则推理出外键约束, 进而构造 DBCN 的有向边, 这种推理一般近似正确, 主要取决于隐性主外键的命名规则。根据语义隐性关联设计了算法 2。

算法 2 基于语义隐性关联的 DBCN 构造算法(HFK-DBCN)

(1) 查询一个数据模式的所有数据表名称, 作为网络节点信息, 节点数为 n ;

(2) 查询每个数据表节点 $Node_i(i=1,2,\dots,n)$ 的非主键字段个数 fc 和名称 FFields[1,2,...,fc];

(3) 查询每个数据表节点 $Node_j(j=1,2,\dots,n, \text{且 } j \neq i)$ 所有主键字段个数 pc 和名称 PFields[1,2,...,pc];

(4) 如果 FFields[1,2,...,fc] 的某个字段等于 PFields[1,2,...,pc] 中某个字段, 则添加一条从节点 i 到节点 j 的边。

针对语义关联下, 通过算法 2 仍不能构造 DBCN 的情况, 提出了 HFK-DBCN+算法。以开源课程管理系统 Moodle 为例, 其中所有表的主键均采用 id 命名, 通过 HFK-DBCN 检索不到任何边信息, 但是在大量的引用表中采用了“数据表名称”+“id”方式引用主键表中的字段, 例如在 mdl_user 表中主键为 id, 而在引用表 mdl_course_display 等多个数据表中均使用 userid 隐性关联。

算法 2+ 扩展的语义隐性关联的数据库复杂网络构造算法(HFK-DBCN+)

(1) 查询一个数据模式的所有数据表名称, 作为网络节点信息, 节点数为 n ;

(2) 查询每个数据表节点 $Node_i(i=1,2,\dots,n)$ 的非主键字段个数 fc 和名称 FFields[1,2,...,fc];

(3) 查询每个数据表节点 $Node_j(j=1,2,\dots,n, \text{且 } j \neq i)$ 所有主键字段个数 pc 和名称 PFields[1,2,...,pc], 增加其扩展的可能主键字段名 PFields[pc+1,...,pc+e];

(4) 如果 FFields[1,2,...,fc] 的某个字段等于 PFields[1,...,pc+e] 中某个字段, 则添加从节点 i 到节点 j 的一条边。

在 Moodle 中, 所有数据表名以前缀“mdl_”开始, 去掉前缀后, 人工分析后增加了 4 项可能匹配的主键字段名, 4 条规则如下: 取出表示数据表的名字, 例如表 mdl_course 中原仅有一个主键“id”, 增加了“course”; 取出表示数据表的名字,

直接加“id”，例如表 mdl_user 中原仅有一个主键“id”，增加了“userid”；取出表示数据表的名字，去掉“s”后加“id”，例如表 mdl_external_services 中原仅有一个主键“id”，增加了“externalserviceid”；取出表示数据表的名字，去掉“ies”直接加“yid”，例如表 mdl_glossary_entries 中原仅有一个主键“id”，增加了“glossaryentryid”。

需要指出的是，由于字段名命名的随意性，通过算法 2 和算法 2+构造的 DBCN 是近似正确的，而在 CASE 工具中不支持 HFK-DBCN+类型的数据模式图结构分析。

3.3 语义隐性关联下确定性 DBCN 构造的数据表字段命名规范

为准确构造语义隐性关联的 DBCN，在数据库设计中，需要满足以下条件：

(1)主键使用表名 TableName，或者表名的子串，并拼接“id”，描述为 subString(TableName).concat(“id”)。

(2)所有对主键引用的外键字段，采用与所引用的主键同名字段标识。

基于以上规范，不仅基于算法 2 能构造准确的 DBCN，也支持使用 CASE 工具的语义推理。

4 实验与分析

4.1 实验数据

本节实现了一个数据库复杂网络分析工具 (DBCN Tool)，其系统结构如图 1 所示，通过 JDBC 连接各种数据库管理系统，提供用户名、密码和连接数据源信息，构造一个以 Pajek(Pajek 是一个成熟的社会网络分析工具)格式存储的 DBCN，网络结构可以通过 Pajek 软件分析，或在 DBCN Tool 中进行可视化分析，其可视化部分采用了网络分析的 Java API 开发包 JUNG。

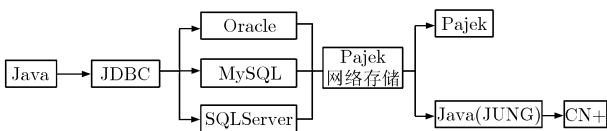


图 1 数据库复杂网络分析工具的系统结构

在数据库复杂网络中，由于数据表作为节点，网络分析的粒度较大，一般节点数在几十到几百、甚至上千个数量级，可视化的意义远大于参数分析。与 PowerDesigner 提供的图结构关系表示不同，DBCN 弱化了节点信息，将数据表视为一个节点，不考虑其内部的字段数和名称，凸显了节点间的参照引用关系。

实验抽取了 9 个中大型 MIS 系统的 DBCN，其数据基本信息如表 2，其中 2 个非开源系统，7 个开源系统；1 个 Oracle，2 个 SQLServer，6 个 MySQL 数据模式。

表 2 实验数据基本信息

软件名称	类别	应用	数据库	边	表数
Uam	非开源	统一审计与监控	Oracle	外键	69
Zabbix	开源	监控系统	MySQL	语义	88
Yumdl	非开源	新课改学校管理	SQLServer	外键	206
Moodle	开源	课程管理系统	MySQL	语义	287
WebErp	开源	通用 ERP	MySQL	外键	125
CeoErp	开源	通用 ERP	SQLServer	外键	97
EcShop	开源	电子商务	MySQL	语义	88
OpenCms	开源	通用 ERP	MySQL	语义	37
NseerErp	开源	通用 ERP	MySQL	语义	443

4.2 网络参数

9 个 MIS 系统的 DBCN 统计参数如表 3 所示，其节点和边的数量级接近，与随机网络相比可达节点之间的平均长度 L 较小， $1 < L < 3$ ；聚类系数 C 较大；平均入度和出度 D_i/D_o 较小， $0 < D_i < 3$ ；可达节点比例 R 较小，它反映级联更新和级联删除的传播能力；网络连通性参数 S 较大，反映数据表孤立节点数较少，大部分数据表与其他数据表有参照引用关联。

作为一个反例，笔者一直犹豫将对开源系统 NseerErp 的分析写进来，该系统拥有 443 个数据表，它采用隐性语义关联，由于数据表和字段命名的随意性，通过算法 2 和算法 2+都无法获得其真实的 DBCN。鉴于 NseerErp 在所讨论的 9 个系统中规模最大，DBCN 在大型系统的作用非常明显，讨论该系统的数据库和字段命名格式非常必要。由于 NseerErp 所有数据表的第 1 个字段名为主键且名称均为“id”，其可能真正的主键没有被定义为主键，如 crm_order 表中主键为系统自增量“id”，其后跟一个非主键“order_id”，在 crm_order_details, crm_discussion 等多个表中也存在“order_id”，因此有理由认为这些“order_id”，参照引用表 crm_order 中的“order_id”字段。由于 crm_order 的 order_id 在数据库中没有被定义为主键，尝试根据表名提取关键字，构造可能的主键，即 crm_id, order_id，而由 crm_order_details 构造可能的主

表3 DBCN 统计参数

数据	N	E	L	C	D_i/D_o	R	S	抽取算法
Uam	69	81	1.82	0.006	1.17	0.047	1.00	算法 1
Zabbix	88	156	1.54	0.020	1.77	0.037	0.92	算法 2
Yundl	206	294	2.10	0.021	1.43	0.034	0.90	算法 1
Moodle	287	243	1.12	0.031	0.85	0.003	0.61	算法 2+
WebErp	125	142	2.16	0.035	1.032	0.027	0.69	算法 1
CeoErp	97	112	1.79	0.121	1.15	0.029	0.82	算法 1
EcShop	88	239	1.87	0.087	2.71	0.080	0.84	算法 2
OpenCms	37	100	1.58	0.130	2.70	0.136	0.89	算法 2
NseerErp	443	2114	2.07	0.645	1.96	0.011	0.52	算法 2+

键也包含 `crm_id`, `order_id` 和 `details_id`, 无法确定哪个表中的 `order_id` 是主键, 进一步以最短数据表名确定其中一个为主键, 按照以上规则进一步筛选, 其 DBCN 参数见表 3, 可以断定该网络具有不确定性, 近似真实地反映系统的数据表关系, 由于字段命名的不规范, 可能存在遗漏和虚假的边。如果系统采用 3.3 节的命名规范, 满足外键字段和相关联的主键字段同名, 则容易构造确定性的 DBCN, 为软件的理解和增量设计提供直观的关系视图。

4.3 距离分布和度分布

图 2 所示为 6 种 DBCN 的节点距离的概率分布, 从确定性的主外键 DBCN 实例 CeoErp, Yundl 和 WebErp 可以看出, 密度最大的距离都为 2, 即存在传递一次的关联关系最多, 最长的传递距离分别是 4, 5 和 5, 它表示操作一个数据表在级联更新和级联删除时影响其他数据表的深度。

图 3 分析了 Yundl, EcShop, WebErp 和 CeoErp 中每个节点的入度, 对入度按照降序排序, 发现入度分布离散, 出度分布比较均匀, 例如在 Yundl 中节点 `tbStudentBase` 入度达到 33, EcShop 的节点 `ecs_group_goods` 入度为 24, WebErp 的节点 `stockmaster` 入度为 18, CeoErp 中节点 `tCustomer` 的入度为 13。

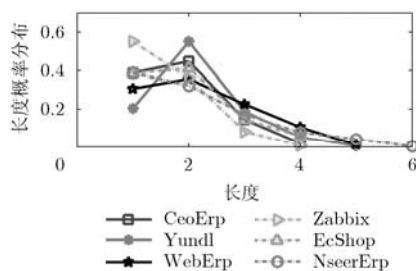


图2 6种 DBCN 的节点距离分布

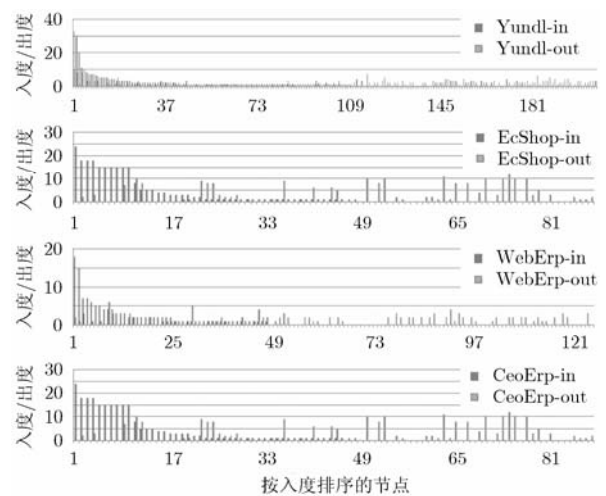


图3 4种 DBCN 的各节点入出度

图 4(a)所示为 4 种 DBCN 的入度和出度的概率分布散点图, 图 4(b)为在双对数坐标下近似拟合的直线, 发现 DBCN 网络具有比较明显的无标度特征。以 Yundl 为例, 数据表中 47% 的节点入度为 0, 有 43% 的节点出度为 0, 网络结构不均匀, 少数节点有较大的度, 大量节点具有较小的度。

4.4 核网络特性

一个图的 k -核指反复去掉度小于或等于 k 的节点后, 所剩余的子图。一个节点存在于 k -核网络, 而在 $(k+1)$ -核中被移除, 那么节点的核数为 k , 核数表明节点在核中的深度, 也表明在网络传递关系上, 节点的复杂程度。Yundl-DBCN 是基于主外键的 DBCN, 其网络结构准确, 它是一个 3-核网络。其中 0-核即网络全集, 节点数较多, 不易区分重要节点, 难以观察到网络核心层。在 3-核网络上, 迭代删除了度为 0, 度为 1 和度为 2 的节点, 节点数由初始的 206 个减少到 35 个, 从网络上能直观获得其骨干和深层的关联信息。软件工程过程使用 DBCN

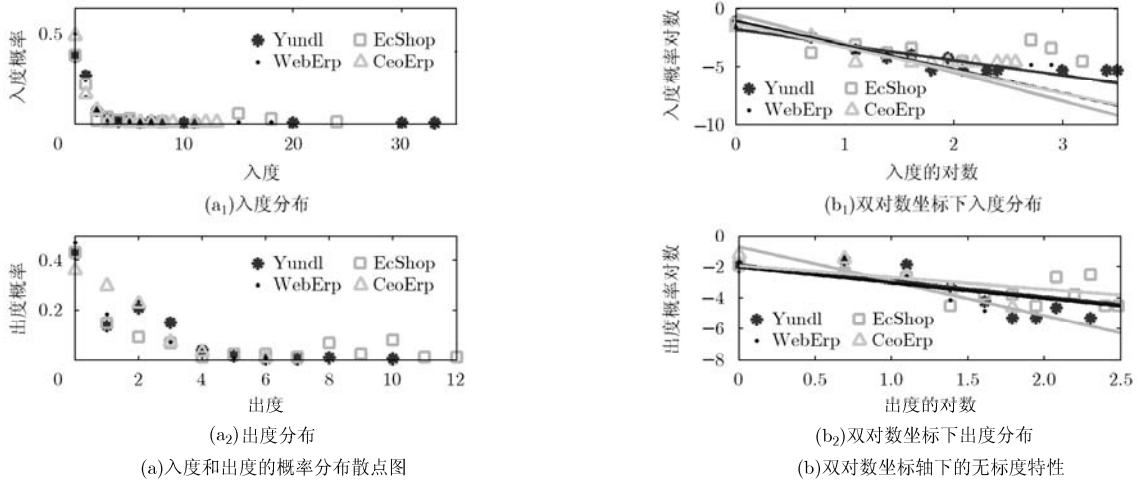


图 4 4 种 DBCN 的节点度分布概率

的核网络可视化数据层的关联，能降低分析设计的难度。

在逐次构造的核网络上，其参数，即以节点数表示的网络规模，网络聚类系数，平均长度和网络密度的变化随 k -核网络变化分别如图 5(a)-5(d)所示，随着 k -核的增大，网络节点数下降，在最大核网络上，从确定性的主外键 DBCN 的实例 CeoErp, Yundl 和 WebErp 看，网络规模下降到 0.2 倍以下，DBCN 的核在 3 或 4 之间，而带有不确定性的隐性语义关联的 DBCN 实例 Zabbix, EcShop 和 NseerErp，从图上可以推测，构造算法可能引入了实际不存在的关联，网络的统计特性异常。聚类系数随 k -核的增大变化不确定。网络平均长度随着 k -核的增大在减少。网络密度表示网络节点间实际存在的边数和可以存在的边数比，其随 k -核的增大而增大，表明高核网络上节点之间边多，关联紧密，反映出用高核网络分析网络的骨干和深层关联是合理的。

5 结束语

本文提出了数据库复杂网络的概念，其工程意义在于为管理信息系统提供了一种规模度量，为软件的数据库设计提供一种新的辅助工具，研究了常用数据库管理系统的主外键关联和语义隐性关联形成的复杂网络构造算法，并分析了网络的统计特性。在大型的 MIS 系统中，或者企业内部的多个 MIS 系统关联形成的数据中心，DBCN 的工程意义更加明显，可以化简数据层的复杂性。数据库复杂网络表示了数据库关联图的统计特性，其节点(数据表)在网络上的重要性是不同的，少量的节点之间蕴含了大多数的业务逻辑关系。此外，数据库复杂网络视图也为软件开发文档的数据库描述提供了一种新的形式，是一种自说明的文档，使得在软件版本变更、增量设计以及开发人员变动情况下，尤其在开源软件的二次开发中，能快速理解数据库内部的关

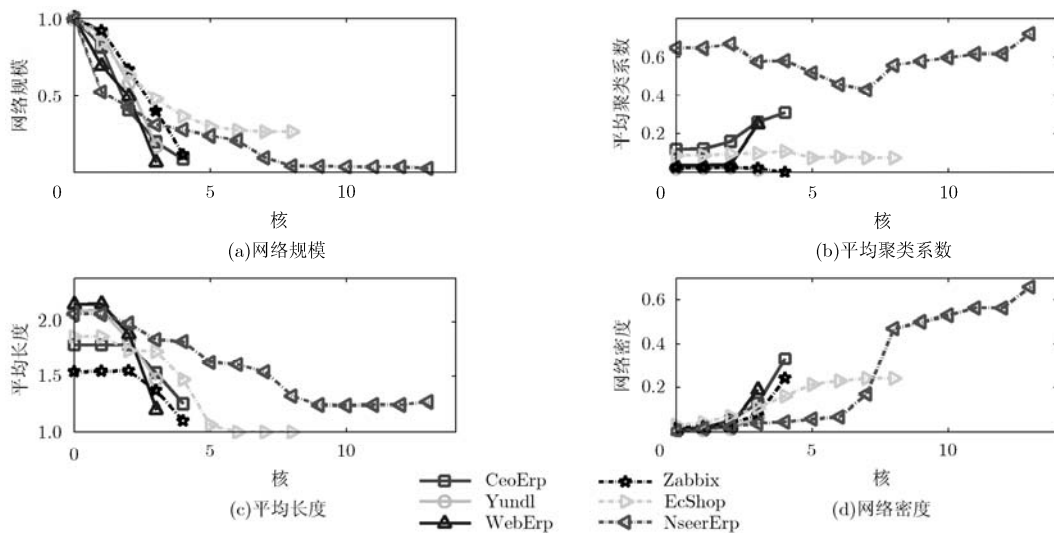


图 5 不同核深度下 6 种 DBCN 网络参数比较

联, 从而理解业务逻辑关系, 缩短开发时间。鉴于构造基于隐性语义关联的数据库复杂网络具有不确定性, 提出了按照外键字段名与参照引用主键同名的原则来规范命名数据字段, 能支持构造准确的网络结构, 辅助于软件开发人员对系统逻辑的理解。

参 考 文 献

- [1] 何克清, 马于涛, 刘婧, 等. 软件网络[M]. 北京: 科学出版社, 2008: 5-37.
He Ke-qing, Ma Yu-tao, Liu Jing, *et al.* Software Networks[M]. Beijing: Science Press, 2008: 5-37.
 - [2] LaBelle N and Wallingford E. Inter-package dependency networks in open-source software[OL]. <http://arxiv.org/abs/cs/0411096>, 2011, 2.
 - [3] Myers C R. Software systems as complex networks: structure, function, and evolvability of software collaboration graphs[J]. *Physical Review E*, 2003, 68(4): 046116.
 - [4] Pan Wei-feng, Li Bing, and Ma Yu-tao. Multi-granularity evolution analysis of software using complex network theory[J]. *Journal of Systems Science and Complexity*, 2011, 24(6): 1068-1082.
 - [5] 韩言妮, 李德毅, 陈桂生. 软件多粒度拓扑特性分析及其应用[J]. 计算机学报, 2009, 32(9): 1711-1721.
Han Yan-ni, Li De-yi, and Chen Gui-sheng. Analysis on the topological properties of software network at different levels of granularity and its application[J]. *Chinese Journal of Computers*, 2009, 32(9): 1711-1721.
 - [6] 王红春. 网络化软件多粒度动态特性分析[D]. [博士学位论文], 武汉大学, 2010.
Wang Hong-chun. Multi-granularity dynamic characteristic analysis of networked software[D]. [Ph.D. dissertation], Wuhan University, 2010.
 - [7] Cai K Y and Yin B B. Software execution processes as an evolving complex network[J]. *Information Sciences*, 2009, 179(12): 1903-1928.
 - [8] Li Huan and Li Bing. A pair of coupling metrics for software networks[J]. *Journal of Systems Science and Complexity*, 2011, 24(1): 51-60.
 - [9] Ma Y T, He K Q, Li B, *et al.* A hybrid set of complexity metrics for large-scale object-oriented software systems[J]. *Journal of Computer Science and Technology*, 2010, 25(6): 1184-1201.
 - [10] Ma Yu-tao, Liu Yu-chao, Zhang Hai-su, *et al.* A hybrid method for prioritizing software requirements in terms of use cases[J]. *Journal of Convergence Information Technology*, 2012, 7(5): 17-27.
 - [11] Li C F, Liu L Z, and Li X Y. Software networks of java class and application in fault localization[C]. Proceedings of ISDEA, Sanya, China, January 7-8, 2012: 1117-1120.
 - [12] 马于涛, 何克清, 李兵, 等. 网络化软件的复杂网络特性实证[J]. 软件学报, 2011, 22(3): 381-407.
Ma Yu-tao, He Ke-qing, Li Bing, *et al.* Empirical study on the characteristics of complex networks in networked software[J]. *Journal of Software*, 2011, 22(3): 381-407.
 - [13] Fagin R. Degree of acyclicity for hypergraphs and relational database schemes[J]. *Journal of the Association for Computing Machinery*, 1983, 33(3): 514-550.
- 李春芳: 女, 1974年生, 副教授, 博士生, 研究方向为故障定位、软件网络。
刘连忠: 男, 1958年生, 研究员, 研究方向为信息系统与安全、电子政务。
刘振国: 男, 1982年生, 讲师, 硕士, 研究方向为信息检索、网络管理。