

doi:10.11918/j.issn.0367-6234.2015.08.006

配水管网管段改造排序的 PageRank 算法

李飞宇¹,石振锋²,吴晨光¹,于美婷²,袁一星¹

(1. 哈尔滨工业大学 市政环境工程学院,150090 哈尔滨;2. 哈尔滨工业大学 数学系,150001 哈尔滨)

摘 要: 针对城市配水管网管段改造比选排序问题,提出了基于 PageRank 的改进的 MPR-Pipe 算法,实现了对管网节点和管段多种水力属性的 *PR* 值求解.利用经济流量和管段单位水力坡降 *PR* 值,定义了管段改造的重要性度量,并以此作为改造比选排序的依据.理论计算表明该算法求解效率较高,工程应用案例表明该算法提出的管段改造比选排序方案有效、可行.

关键词: 配水管网;管段改造;PageRank;单位水力坡降;管段重要性

中图分类号: TU821.3 文献标志码: A 文章编号: 0367-6234(2015)08-0025-05

PageRank-based selection sort of pipe renewal in water distribution system

LI Feiyu¹, SHI Zhenfeng², WU Chenguang¹, YU Meiting², YUAN Yixing¹

(1.School of Municipal and Environmental Engineering, Harbin Institute of Technology, 150090 Harbin, China;

2.Department of Mathematics, Harbin Institute of Technology, 150001 Harbin, China)

Abstract: A novel modified Page Rank algorithm for pipe renewal (MPR-Pipe) is presented to realize selection sort and provide Page Rank values of static and dynamic attributions of junctions and pipes in water distribution systems. The measurement of pipe renewal significance, well defined on the grounds of economic flow and unit hydraulic gradient, can be taken as the criterion for the selection sort of pipe renewal. Theoretical calculations confirm the high efficiency of the algorithm, and engineering application cases indicate the effectiveness and feasibility of the renewal plan based on MPR-Pipe algorithm.

Keywords: water distribution system; pipe renewal; PageRank; unit hydraulic gradient; pipe significance

结合我国配水管网的实际情况,管网急需改造的原因主要包括:管网设备陈旧与老化;管网运行负荷较重;管网布局欠合理;水源新建及改造;更高的水质安全需求;节能降耗.因此,在对供水管网进行改造过程中应用管网水力模型,分析供水管网的工况,合理地供水管网进行改造,特别是通过优化管网改造的顺序可以节约投资,最大限度地降低能量消耗,而且可以提高整个供水系统的安全性能,对于企业和社会都具有重要意义.为此,国内外开展了管网改造方面的研究,主要包括:管段老化模型研究^[1]、管网运行现状模型研究^[2-3]、管网运行成本模型研究^[4]及管网改造决策模型研究^[5-6],其中管网改造决策模型的建立与求解是最主要的研究方向,

其结果可以直接指导管网改造问题^[7].管段老化模型、管网运行现状分析、管段运行成本经济型分析,一般在构建和求解管网改造决策模型中起辅助作用^[8].同时,国内外还建立了多种优化改造模型用于管网的改造,如基于多目标优化的管网改造决策模型^[6]、基于遗传算法的管网优化模型^[9].这些模型建模的思想、形式、复杂程度差异性较大,各有侧重.

赋予静态属性和动态的状态属性后的供水管网在应用过程中可抽象为网络.管网改造过程中管段的排序问题即可转化为网络中对连接边的排序问题.Google 网页搜索过程中,通过网页间的链接来评价网页的重要性,这就是 Google 网页搜索引擎中的核心算法——PageRank 算法^[10].随着对 PageRank 算法的广泛研究,其不仅在自身网页排序中应用效果显著,在很多领域都得到了应用,如图书管推荐技术、网络安全评估、Web 数据挖掘等.借鉴 PageRank 算法在其他行业中的应用,利用管网改造管段排序问题和网页排序问题的相似性,将 PageRank 算法引

收稿日期: 2014-06-20.

基金项目: 国家自然科学基金(51178141);水体污染控制与治理科技重大专项(2012ZX07408-002-004-002).

作者简介: 李飞宇(1981—),男,博士研究生;
袁一星(1957—),男,教授,博士生导师.

通信作者: 吴晨光, wu.cg@126.com.

入城市配水管网改造的应用中目前未见报道.本文首先介绍了 PageRank 算法,并评述了其优缺点.结合城市配水管网的特殊性,提出了适于管网改造排序应用的改进的 PageRank 算法,并定义了管段的重要性度量,进而给出管段改造排序的计算方法.最后,根据管段排序的结果并通过实例验证,表明该算法描述管段重要性度量的定义是合理的、可行的,可为城市配水管网提供可行的科学改造方案.

1 PageRank 算法简介及评述

1.1 PageRank 算法简介

PageRank 算法是 Google 网页搜索引擎中的核心算法,由 Google 的创始人 Sergey Brin 和 Lawrence Page^[10]于 1998 年提出并得到了成功应用.其基本思想主要来自传统文献计量学中对文献引用的研究和分析^[11],即一篇文献被其他文献引用的越多,说明文献的质量就越高.假设网页 A 链接网页 B,则可看作是网页 A 对网页 B 进行投票,根据投票数来判断网页的重要性(PageRank 值,简称 PR 值).PageRank 技术根据整个 Web 的链接结构来计算各个网页的重要性值,构造有向网页图 $G = \langle V, E \rangle$,其中顶点 V 为所有的网页集合,边 E 为网页中所有的链接集合,网页 A 存在一条指向网页 B 的链接即表示为网页有向图的边 E .PageRank 算法的基本定义为

$$PR(P_i) = (1 - d) + d \sum_{P_j \in B(P_i)} PR(P_j) \frac{1}{C(P_j)}, \quad i = 1, 2, \dots, n. \quad (1)$$

式中: $\{P_i\}_{i=1}^n$ 表示网页; $PR(P_j)$ 表示网页 P_j 的重要性值; $B(P_i)$ 表示链接到网页 P_i 的总链接网页的集合; $C(P_j)$ 表示 P_j 向外指数的总连接数; d 为阻尼因子,一般取为 $(0, 1)$.在解上述方程中 d 越小,方程收敛的速度越快,但 d 也不能太小,否则该算法中网页链接排序的思想就被弱化了,一般 d 选取 0.85 ^[12].

1.2 PageRank 算法的不足

PageRank 算法在 Web 页面重要性应用中存在两个主要的不足,即 PR 值沉淀现象和“黑洞效应”.在多个网页彼此链接后形成的有向图中,当网页链接形成的有向图之外的网页链接到该有向图内部的网页时,就会出现有向图内部网页传递不出去 PR 值而导致传递进来的 PR 值一直滞留在此有向图内部,即 PR 值沉淀现象.图 1 解释了该现象.在图 1 中,网页 A、B、C、D 内部相互链接,无向外链接,网页 E、F 分别链接网页 B 和 D,在应用 PageRank 算法进行迭代计算中,网页 E、F 的 PR 值不能在网页 A、B、C、D 形成的组内贡献出去而出现沉淀的现象.为此,Sergrey Brin 和 Law Page 对 PageRank 算法进

行改进^[13],引入衰退因子 $D(P_i)$,对应某一网页的初始 PR 值,改进算法即

$$PR(P_i) = (1 - d) + d \sum_{P_j \in B(P_i)} PR(P_j) \frac{1}{C(P_j)} + dD(P_i), \quad (2)$$

其中 $\|PR(P_i)\| = 1$.

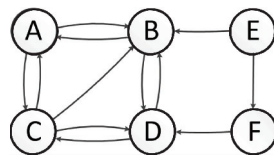


图 1 PR 值沉淀现象的 Web 有向图

在求解式(2)描述的方程组时,其收敛性问题是关键.对于连通的有向 Web 图 $G = \langle V, E \rangle$,设其中某个顶点的入度为 0,由于其没有对外贡献 PR 值,PR 值会逐渐减少,经过有限次迭代之后,该连通图内所有顶点的 PR 值将全部为 0,这就是 PageRank 算法存在的“黑洞效应”,即入度为 0 的顶点就像黑洞一样,将整个有向图内的 PR 值全部慢慢“吸收”.然而,针对“黑洞效应”,Google 并没有公布其处理办法,有些研究者认为可以将 Web 有向图内所有的网页均作为入度为 0 网页的链接,这样就能有效避免“黑洞效应”.

针对 PageRank 算法存在的上述两个不足,在管段排序应用中,应该根据给水管网的实际特点,对 PageRank 算法进行适当改进.

2 面向管网改造管段排序的改进的 PageRank 算法——MPR-Pipe

2.1 改进的 PageRank 算法——MPR-Pipe

为将 PageRank 算法更好地应用在给水管网的管段改造排序问题,对 PageRank 算法作如下两点改进:

1) 在网页模型中,网页的属性值只体现了链接数,各个链接只是数值上平分此网页的 PR 值.给水管网中,每个管段具有更多实际意义的属性值,如流量、流速、单位水头损失等,不同的管段属性值对管段的最终改造排序结果有重要的影响,因此,在管网模型中应以管段属性信息值的比例来分配该管网节点的 PR 值.本文选用单位水力坡降.

2) 根据 PageRank 算法的介绍,由式(1)可知,PageRank 的名次和该网页被链接的数目基本一致,无论该网页链接多少网页均几乎不会影响 PR 值.相反有多少网页链接该网页从根本上决定网页 PR 值的大小.本文提出对管段进行排序的改进的 PageRank 算法由管网节点的正向链接和反向链接共同决定,这种改进一方面结合了管网模型的实际

情况, 另一方面也有效地避免了在网页模型中 PageRank 算法存在的沉淀现象和“黑洞效应”。

记 $I(J_i, J_j)$ 为管网节点 J_i 到节点 J_j 的单位水头损失, $I(:, J_i)$ 为所有流入节点 J_i , 并与节点 J_i 链接的节点构成的 J_i 单位水头损失总和, $I(J_i, :)$ 为流出节点 J_i , 与节点 J_i 链接的节点构成的 J_i 单位水头损失总和。在节点总数为 N_j 的给水管网中, 改进的 PageRank 算法 (modified PageRank for pipe, MPR-Pipe) 可以概括为如下线性方程:

$$PR(J_i) = (1 - d) + d[c_1 \cdot W(v, J_i)^{\text{in}} + c_2 W(v, J_i)^{\text{out}}] \quad (3)$$

式中: $i = 1, 2, \dots, N_j$;

$$W(v, J_i)^{\text{in}} = \sum_{J_j \in \text{In}(J_i)} PR(J_j) \frac{I(J_j, J_i)}{I(J_j, :)};$$

$$W(v, J_i)^{\text{out}} = \sum_{J_k \in \text{Out}(J_i)} PR(J_k) \frac{I(J_i, J_k)}{I(:, J_k)};$$

$\text{In}(J_i)$ 为流入管网节点 J_i 的管网节点集合; $\text{Out}(J_i)$ 为从管网节点 J_i 流出的管网节点集合; c_1 和 c_2 分别为流入与流出权重系数 (介于 0~1 的常数), 且 $c_1 + c_2 = 1$; d 为阻尼因子, 用于保证线性方程 (3) 的收敛性。

2.2 管段重要性度量的定义

通过方程组 (3) 求得供水管网中每个节点在流量、绝对水头和自由水头等属性上的 PR 值。然而, 管网改造过程中都是基于管段进行的, 因此, 需要利用每个节点的 PR 值定义管段重要性的度量。考虑到管段的单位水力坡降和流量是刻画管段输水能力和经济指标计算时的重要指标, 本文将结合管段的单位水力坡降和实际计算流量与经济流量之间的关系来定义管段的重要性。设连接管段 P_i 的两个节点分别为 $J_i^{(1)}$ 和 $J_i^{(2)}$, P_i 的实际计算流量为 $P_{\text{Flow}}^{(i)}$, 其经济流量为 $P_{\text{BestFlow}}^{(i)}$ 。则定义管段改造的重要性程度为

$$PR(P_i) = [e_1 \times PR(J_i^{(1)}) + e_2 \times PR(J_i^{(2)})] \frac{P_{\text{Flow}}^{(i)}}{P_{\text{BestFlow}}^{(i)}} \quad (4)$$

其中 $e_1 + e_2 = 1$ 。本文取 $e_1 = e_2 = 0.5$, 利用供水管网中管段的单位水力坡降属性来定义管段重要性度量, 并称为管段的 PR 值。由式 (4) 可以看出, 当管道实际流量越大于经济流量时, 管段的 PR 值就越大, 据此排序在前, 表明管段越急需改造, 因此, 式 (4) 具有明显的物理意义。该 PR 值标明了所有管段改造的顺序 (不考虑管段改造的经济成本等因素)。当实际流量远小于经济流量时, 从物理意义上讲也急需改造。在实际工程应用中, 对这种管道的改造一般采用关阀改变供水管网逻辑拓扑结构, 或调整阀门开度改变这部分管道的水力工况状态等方案来实现管网的改造, 不在本文的讨论范围之内。

2.3 MPR-Pipe 算法的数值求解

给定管网有向图 $G = \langle V, E \rangle$, 定义管网的邻接矩阵 P , 若管段 i 链接管段 j , 且水流方向为从管段 i 到管段 j , 则 $P_{i,j} = I(J_i, J_j)$, 否则 $P_{i,j} = 0$ 。即

$$P_{i,j} = \begin{cases} I(J_i, J_j), & i \text{ 流向 } j \\ 0, & \text{其他} \end{cases} \quad (5)$$

注意到 $\frac{I(J_j, J_i)}{I(J_j, :)} (\text{s.t. } J_j \in I(J_i))$ 即为对管网管段邻接

矩阵 P 的行向量归一化, 记作 $P_1, \frac{I(J_i, J_k)}{I(:, J_k)} (\text{s.t. } J_k \in O(J_i))$ 为管网管段邻接矩阵 P 的列向量归一化, 记作 P_2 , 于是可将式 (3) 简记为

$$x = Ax, A = (1 - d)ve^T + d(c_1P_1 + c_2P_2^T) \quad (6)$$

其中 $e = (1, 1, \dots, 1)^T$ 与 $v = \frac{e}{n}$ 为列向量, d 为阻尼因子。整理 $x = Ax$ 可得

$$[I - d(c_1P_1 + c_2P_2^T)]x = mv \quad (7)$$

其中 $m = m(x) = (I - d)e^Tx = (I - d)\|x\|_1$ 为常数, 于是可将式 (6) 简化成线性方程组。记 $B = I - d(c_1P_1 + c_2P_2^T)$, $m = (I - d)v$, 则式 (6) 的求解过程转化为对线性方程组 (8) 的求解, 即

$$Bx = m \quad (8)$$

注意到管网中的节点数量通常较大, 基于高斯消元法进行求解效率较低。为此, 本文采用 Jacobi 迭代法, 建立迭代公式

$$x^{k+1} = Jx^k + b, \quad (9)$$

则 MPR-Pipe 算法转换后的等效线性方程组 $Bx = m$ 的 Jacobi 迭代矩阵为

$$\begin{cases} J = d(c_1P_1 + c_2P_2^T), \\ b = (1 - d)v. \end{cases} \quad (10)$$

取迭代过程中节点的 PR 初始值为 $\frac{1}{N_j}$ 。注意到迭代矩阵 J 的谱半径小于 1, 因此, 该迭代过程是收敛的。

综上, MPR-Pipe 算法可描述为

算法 1 面向管段改造排序的改进的 PageRank 算法

Input 管网水力计算结果数据文件; 阻尼因子 d ; 流入和流出权重系数 c_1, c_2 ; 迭代终止精度 ε ; 最大迭代次数 MAX_Iter ;

Output 城市供水管网管段和节点的 PR 值;

Step 1 读取供水管网水力计算结果数据文件;

Step 2 建立管段邻接关系矩阵 A_E ;

Step 3 由流入和流出权重系数 c_1, c_2 , 建立与邻接矩阵 A_E 相关的水力工况参数系数矩阵 C_E ;

Step 4 初始化每个节点的 PR 值为 $\frac{1}{N_j}$;

for $i = 1$ to MAX_Iter do

利用线性方程组的 Jacobi 数值求解
 算法求解公式(6);
 计算 $\delta^{k+1} = |x^{k+1} - x^k|$;
 if $\delta < \varepsilon$ goto Step 5;

end

Step 5 根据节点的 PR 值,计算管段的 PR 值;

Step 6 保存管段和节点的 PR 值信息为数据文件。

3 结果及分析

3.1 数据结构设计

设计如下数据结构实现 MPR-Pipe 算法:

■节点的数据结构描述如下:

```
struct NODE { // 节点结构
    long innNo; // 内部编号
    double dFlow_PR; // 流量 PR 值
    double dSpeed_PR; // 流速 PR 值
    double dFall_PR; // 压降 PR 值
};
```

■管段的数据结构描述如下:

```
struct PIPE { // 管段结构
    long innNo; // 内部编号
    long from_innNo; // 流出节点内部编号
    long to_innNo; // 流入节点内部编号
    double dFlow_PR; // 流量 PR 值
    double dSpeed_PR; // 流速 PR 值
    double dFall_PR; // 压降 PR 值
    double dFall_1000_PR; // 单位水损 PR 值
};
```

■管网拓扑结构描述如下:

```
typedef struct ADJACENT_ELEMENT {
    int no;
    PIPE * pipe;
} A_E;
typedef struct COEFFICIENT_ELEMENT {
    int col;
    double value;
} C_E;
typedef struct WaterPipeNet {
    vector<PIPE> pipe_list; // 管段数组.
    vector<NODE> node_list; // 节点数组.
    vector< vector<A_E> > row_type; // 邻接矩阵
    vector< vector<A_E> > col_type; // 邻接矩阵
    vector< vector<C_E> > cem; // 系数矩阵
} GRAPH;
```

3.2 计算实例及分析

本文基于 Embarcadero RAD XE 5 平台,利用C++

编程实现了 MPR-Pipe 算法.以北方某市部分管网为例进行了仿真计算和分析,其管网拓扑如图 2 所示。

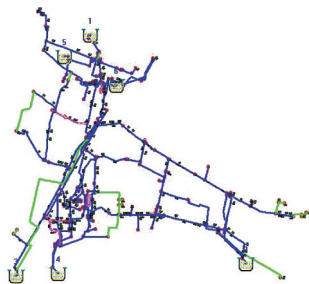


图 2 北方某市部分管网拓扑结构

在该管网拓扑结构中,由 291 条管段、259 个节点及 6 个水厂构成.在水力计算过程中,共计算了 24 个时段,水力计算的边界约束不做列出.利用本文的 MPR-Pipe 算法针对该管网 24 个时段水力计算结果中的管段流量、流速、水头损失和单位水头损失等动态水力属性分别进行求解,花费的总计算时间为 0.058 s.图 3 给出了上述管网在 23 时、9 时、14 时和 18 时的管段改造重要性排序示意图。

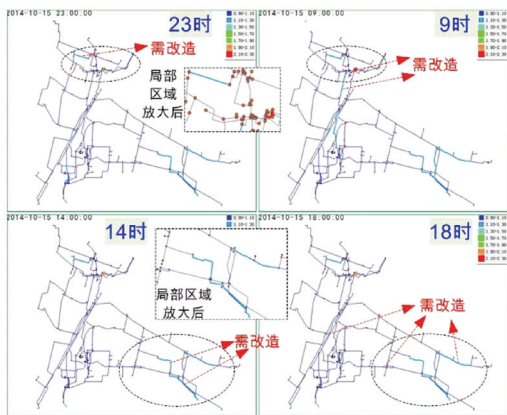


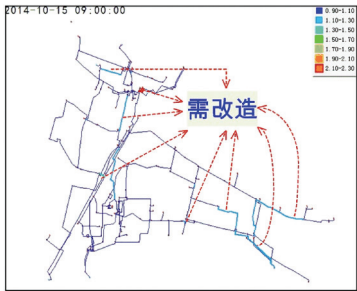
图 3 北方某市部分管网管段改造排序结果

图 3 中 4 个子图中共标识了 8 条管段,其在 24 个时段中,有超过 18 个时段均具有较大的 PR 值.也就是说,从管段改造应用的意义上看,这 8 条管段为急需改造的管段,也是上述管网中的较“薄弱”管段.对于具有相同管材的管段,其阻力系数 C 值会随着敷设年代的增加而逐渐变小,在相同的用水形式下,其单位管段水力坡降也会变大.根据管段重要性的定义,对于单位水力坡降较大的管段,如果计算得到的流量小于经济流量的管段,则其重要性将按照指数级(小于 1)平缓;对于实际计算流量大于经济流量的管段,其重要性将按照指数级(大于 1)提升,从而加速这部分管段的排序,成为急需改造的管段。

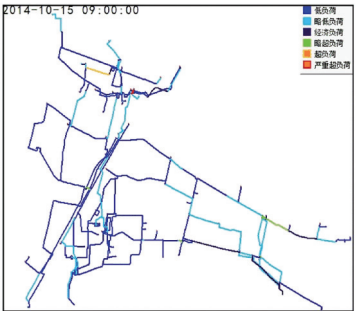
3.3 MPR-Pipe 算法与传统分析方法的比较

为了验证本文算法的合理性和有效性,以高日高时为水力计算的基本工况,与传统的管段运行负荷计算方法^[14]得到的结果进行了比较,结果如图 4 所示.其中管段运行负荷中的经济指标等因素与本

文中计算经济流量时设定的参数相同.可以看出,本文得到的 8 条急需改造的管段,其运行负荷普遍高于其他管段.在这一点上本文定义的重要性较好地包括了传统的管段运行负荷分析的范畴.然而,在运行负荷分析方法中,有较多的管段处于略低负荷、经济负荷及略超负荷等,尚不能较好地给出管段改造的排序结果.本文提出的管段重要性度量可较好地对运行负荷偏高的管段实现数值上较大范围的放大,而运行负荷较低的管段则实现了数值上较大范围的缩小,从而可更有效地实现管段重要性两个不同层级上的区分,也更有利于快速实现急需改造管段的选择.因此,从传统的管段改造角度看,本文算法得到的排序结果要优于通过管段运行负荷分析得到的改造的管段集合.



(a) 本文算法的排序结果



(b) 管段运行负荷分布

图 4 高日高时(上午 9 时)工况下本文算法排序结果及管段运行负荷的计算比较

综上,本文提出的改进的 PageRank 算法快速地对管段改造方案的比选排序,该排序结果综合反映了管段的单位水头损失和经济流量的情况.对于敷设年代较久远的管段,因其单位水力坡降较大而可能成为排序中急需改造的管段.本文算法得到的排序结果已成功应用于该城市这部分供水管网的实际改造.由于管径的优化选择不在本文的讨论范围之内,这里仅给出了需要改造管段的优先顺序,而未给出这些管段改造后的管径选择方案.

4 结 论

1) 基于 PageRank 算法,提出了面向城市配水管网改造管段排序应用的改进的 PageRank 算

法——MPR-Pipe 算法.

2) 基于 MPR-Pipe 算法,定义了管段改造排序的重要性度量——管段改造的 PR 值.

3) 基于管段改造的 PR 值得到管段改造排序方案,实现了对北方某城市部分管网的实际改造应用,表明本文的方法具有实用的工程意义和理论意义.

参考文献

[1] WATSON T, MASON A. A hierarchical Bayesian model for the maintenance of water pipe networks [C]//Proceedings of the 7th International Conference on Hydroinformatics. France: Nice, 2006: 2991-2998.

[2] GYERGYEK L, PERSEN S. Simulation and optimal control of large water distribution systems [J]. Mathematics and Computers in Simulation, 1982, 96(2): 214-216

[3] 舒诗湖, 何文杰, 赵明, 等. 供水管网漏失检测技术现状与进展[J]. 给水排水, 2008, 34(6): 114-116.

[4] SAVIC D A, WALTERS G A. Genetic algorithms for least-cost design of water distribution networks [J]. Journal of Water Resources Planning and Management, 1997, 123(2): 67-77.

[5] MALM A, SVENSSON G, BÄCKMAN H, et al. Prediction of water and wastewater networks rehabilitation based current age and material distribution [J]. Water Science & Technology: Water Supply, 2013, 13(2): 23-28.

[6] ROSHANI E, FILION Y R. Multi-objective rehabilitation planning of water distribution systems under climate change mitigation scenarios [J]. Bridges, 2014, 10: 9780784412312.321.

[7] SARGAONKAR A, KAMBLE S, RAO R. Model study for rehabilitation planning of water supply network [J]. Computers, Environment and Urban Systems, 2013, 39: 172-181.

[8] BEALE D J, MARLOW D R, COOK S. Estimating the cost and carbon impact of a long term water main rehabilitation strategy [J]. Water Resources Management, 2013, 27 (11): 3899-3910.

[9] KANTA L, ZECHMAN E, BRUMBELOW K. Multiobjective evolutionary computation approach for redesigning water distribution systems to provide fire flows [J]. Journal of Water Resources Planning and Management, 2011, 138 (2): 144-152.

[10] BRIN S, PAGE L. The anatomy of a large-scale hypertextual web search engine [J]. Computer Networks and ISDN Systems, 1998, 30(1): 107-117.

[11] GIBSON D. Device discovery and power management in embedded systems [C]//Proceedings of the Linux Symposium 2003. Ottawa: [s.n.], 2003: 213-224.

[12] AUSTIN D. How Google finds your needle in the web's haystack [J]. American Mathematical Society Feature Column, 2006, 10: 12.

[13] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: bringing order to the web [J]. Stanford Infolab, 1999, 9(1): 1-14.

[14] 赵洪滨. 给水管网系统理论与分析 [M]. 北京: 中国建筑工业出版社, 2003: 312-313. (编辑 刘 彤)