

基于 PageRank 的有向加权复杂网络节点重要性评估方法

张 琨 李配配 朱保平 胡满玉

(南京理工大学计算机科学与工程学院, 南京, 210094)

摘要: 现有复杂网络节点重要性评估研究主要集中在无向、无权复杂网络上, 未能全面客观反映真实复杂网络的情况。本文基于有向加权复杂网络模型, 借鉴 PageRank 排名算法, 并结合复杂网络节点重要性评估特点, 提出节点重要性评估的新指标——DWCN-NodeRank 和相应评估方法, 该指标既反映出节点局部连接的特性, 又从全局体现了有向加权复杂网络中整体链接关系对节点重要性的影响。采用真实的复杂网络数据集所进行的仿真实验结果表明, 该方法能快速、有效地评估有向加权复杂网络节点的重要性, 提高了复杂网络节点重要性评估的实用价值。

关键词: 复杂网络; 节点重要性; 评估方法; PageRank

中图分类号: TP393

文献标志码: A

文章编号: 1005-2615(2013)03-0429-06

Evaluation Method for Node Importance in Directed-Weighted Complex Networks Based on PageRank

Zhang Kun, Li Peipei, Zhu Baoping, Hu Manyu

(School of Computer Science and Engineering, Nanjing University of Science & Technology, Nanjing, 210094, China)

Abstract: The existing evaluation methods for node importance in complex network mainly focus on undirected and unweighted complex networks, and can not reflect objectively the reality of real world status. Based on directed-weighted complex network model, the PageRank ranking algorithm and combined with characteristics of evaluation for node importance of the directed-weighted complex network, a new evaluation index named DWCN-NodeRank for node importance and corresponding evaluation method are put forward. Besides the local link characteristics, the proposed index can reflect the influence of whole link relationship on node importance in directed-weighted complex networks from the global view. Simulation experiment runs on the data set of real complex networks, and the results show that the proposed method may quickly, effectively evaluate node importance in directed-weighted complex networks.

Key words: complex networks; node importance; evaluation method; PageRank

如何正确评价节点的重要性已成为复杂网络研究中的一项具有重要意义的课题。目前, 复杂网络节点重要性评估主要集中在不含权重的无权复杂网络上^[1-4], 即只给出节点之间的相互作用存在与否的定性描述, 并不描述节点间相互作用的强弱

程度。此外, 在现实生活中, 除了要考虑节点之间相互作用的强弱之外, 还要考虑这种相互作用的方向^[1], 如食物链网络、引文网络、Web 页面网络及 Internet 上的信息流网络等。在这些网络中, 节点间相互作用是有明确方向且可能不对等的。因此,

基金项目: 国家自然科学基金(61003210)资助项目; 江苏省自然科学基金(BK2010491, BK2011023)资助项目; 江苏省“六大人才高峰”基金(11-C-028)资助项目。

收稿日期: 2012-05-22; 修订日期: 2012-12-16

通信作者: 张琨, 女, 教授, 博士生导师, 1977 年出生, E-mail: zhangkun@njjust.edu.cn。

研究有向加权复杂网络中的重要节点评估,对于提高复杂网络鲁棒性和抗毁性、推动复杂网络发展具有重要意义和实用价值。

目前加权复杂网络节点重要性评估的研究刚刚起步,主要针对无向加权复杂网络,其重要性评估方法一般有两种:第一种以社会网络分析为基础,该方法使用复杂网络结构的某种统计特性,如节点介数、节点度、特征向量、信息、接近度、簇系数、累计提名等。例如,文献[5]通过定义节点效率和节点重要度评价矩阵,综合考虑了节点效率、节点度值和相邻节点的重要度贡献,提出一种利用重要度评价矩阵来确定复杂网络关键节点的方法,能够克服节点删除法和节点收缩法的不足。文献[6]在节点收缩法的基础上,引入节点连边的重要度,提出一种基于凝聚度的节点重要度评估方法,通过调节加权比例系数,还可以调节节点连边对节点重要度排序的影响。文献[7]考虑了加权网络的动态性和社区性,综合考虑节点增删、边权值变换情况下的动态距离矩阵更新和先分组后排序思想,给出了基于距离增量分组的动态加权网络节点重要度评估算法。总体来说,上述方法通过使用某一种指标来评估节点重要性,从而导致评估过于片面,评估结果欠准确。还有一些文献通过挖掘各种指标之间的内在联系,使用多种指标综合评估,结果相对要准确,但是这类方法计算复杂度高,且只能在保证网络整体性的前提下进行。第二种方法是以系统科学分析为基础,该方法基于系统的“核与核度”理论,认为重要节点一旦遭到破坏会造成整个系统瘫痪或造成重大损失。文献[8]定义虚拟的理想“核心节点”,将灰色关联度作为测度,评价网络中每个节点和理想“核心节点”的关联度,提出一种定量评估复杂网络节点重要度的算法。

为此,本文在有向加权复杂网络模型基础上,综合考虑复杂网络关联边的方向和权值大小,借鉴搜索引擎中 PageRank 排名算法思想,旨在提出有向加权复杂网络节点重要性评估的新指标及具体方法,从而能准确、有效地度量有向加权复杂网络节点的重要度。

1 有向加权复杂网络模型

有向加权复杂网络模型用 G 表示, $G = (V, E)$ 。 $V = \{v_1, v_2, \dots, v_n\}$ 为节点集合, $E = \{e_1, e_2, \dots, e_m\} \subseteq V \times V$, 为边集合。网络的节点数目为 $n = |V|$, 边数为 $m = |E|$ 。 $v_i \in V, (i = 1, 2, \dots, n)$, 表示网络中的一个节点, $(v_i, v_j) \in E$, 表示节点 v_i 到

节点 v_j 的一条有向边。 $w(v_i, v_j)$ 表示有向边 (v_i, v_j) 的权值(或连接强度); 类似的, $w(v_j, v_i)$ 表示有向边 (v_j, v_i) 的权值, 一般情况下 $w(v_i, v_j) \neq w(v_j, v_i)$ 。

本文引用基于文献[9]的有向加权网络模型^[10-11], 该模型在文献[9]模型基础上, 依据典型网络方向性特点, 结合方向性和网络演化特性, 定义节点强度包括入强度和出强度, 且引入概率参数 p 及 q , 在演化过程中根据节点强度进行择优选择。该模型可以通过调节 p 及 q 的值使得网络平均路径长度和聚类系数完全符合小世界和无标度特性, 实验表明节点出入度和出入强度分布均满足幂律分布。

依据该模型, 每个节点的强度分为入强度和出强度, $S_{in}(v_i)$ 表示节点 v_i 的入强度, $S_{in}(v_i) = \sum_{v_j \in V_{in}(v_i)} w(v_j, v_i)$, $S_{out}(v_i)$ 表示节点的出强度, $S_{out}(v_i) = \sum_{v_j \in V_{out}(v_i)} w(v_i, v_j)$ 。其中, $V_{in}(v_i)$ 表示指向节点 v_i 的所有节点集合; $V_{out}(v_i)$ 表示节点 v_i 所指向的所有节点集合。

2 基于 PageRank 的节点重要性评估方法

2.1 节点重要性评估指标——DWCN-NodeRank

PageRank 算法^[12] 是用于搜索引擎中网页排序的经典算法。该算法基于“从优质网页链接而来的网页必定还是优质网页”的回归关系, 其基本思想是: 当网页 A 有一个连接指向网页 B, 就认为 B 获得了 A 对它贡献的分值, 该值的多少取决于网络 A 本身的重要程度, 即网页 A 的重要性越大, 网页 B 获得的贡献值就越高。由于网络中网页连接的相互指向, 该分值的计算为一个迭代过程, 最终网页根据所得分值进行检索排序。PageRank 的数学公式如下

$$PR(x) = \frac{(1-\sigma)}{n} + \sigma \sum_{i=1}^n \frac{PR(Y_i)}{C_{out}(Y_i)} \quad (1)$$

式中: $PR(x)$ 为网页 x 的 PageRank 值; $PR(Y_i)$ 为链接到网页 x 的网页 Y_i 的 PageRank 值; $C_{out}(Y_i)$ 为网页 Y_i 的出链数量; σ 为阻尼系数, 表示在任意时刻, 用户到达某页面后并继续向后浏览的概率, 阻尼系数越大, 页面级别的收益越大; n 为网页总数。根据 PageRank 计算公式可知, x 的链接源越

多,即指向页面 x 的链接数目 n 越大,则 x 的重要性越高;而且网页 x 的链接源页面的级别越高,即 $PR(Y_i)$ 值越大,则网页 x 越重要, x 的 PageRank 值越大。

因此,本文借鉴 PageRank 算法将网页链接价值概念作为重要性排名因素的思想,并将其引入复杂网络节点重要性评估中,提出一个有向加权复杂网络节点重要性评估指标——DWCN-NodeRank。

在一个具有 n 个节点的有向加权复杂网络中,假定对于节点 v ,其连接分别来自节点 $v_1, v_2, \dots, v_l, w(v_i, v), 1 \leq i \leq l$ 分别为节点 v_i 到节点 v 的权值,则评估节点 v 的重要性指标 DWCN-NodeRank 值 $NR(v)$ 的计算公式如下所示

$$NR(v) = \frac{(1-\sigma)}{n} + \sigma \left(\frac{w(v_1, v)}{S_{out}(v_1)} NR(v_1) + \frac{w(v_2, v)}{S_{out}(v_2)} NR(v_2) + \dots + \frac{w(v_l, v)}{S_{out}(v_l)} NR(v_l) \right) = \frac{(1-\sigma)}{n} + \sigma \left(\frac{w(v_1, v)}{\sum_{j=1}^{m_1} w(v_1, z_j)} NR(v_1) + \frac{w(v_2, v)}{\sum_{j=1}^{m_2} w(v_2, z_j)} NR(v_2) + \dots + \frac{w(v_l, v)}{\sum_{j=1}^{m_l} w(v_l, z_j)} NR(v_l) \right) = \frac{(1-\sigma)}{n} + \sigma \sum_{i=1}^l \frac{w(v_i, v)}{\sum_{j=1}^{m_i} w(v_i, z_j)} NR(v_i) \quad (2)$$

式中: $\sigma(0 < \sigma < 1)$ 为阻尼系数; $NR(v_i)$ 为指向节点 v 的节点源 v_i 自身的 DWCN-NodeRank 值;

$\sum_{j=1}^{m_i} w(v_i, z_j)$ 为节点源 v_i 的出强度 $S_{out}(v_i)$, 其中 $V_{out}(v_i) = \{z_1, z_2, \dots, z_{m_i}\}$, 即节点源 v_i 与包括节点 v 在内的 z_1, z_2, \dots, z_{m_i} 等 m_i 个节点直接相连。节点 v 获得节点源 v_i 的权重可以用边 (v_i, v) 的权值与节点 v_i 的出强度(从节点 v_i 发出的所有连接强度的总和)的比值来表示

$$\frac{w(v_i, v)}{\sum_{j=1}^{m_i} w(v_i, z_j)} \quad (3)$$

通过该权重可对传统 PageRank 算法中平均分配的不合理性进行改进,通过对不同连接赋予不同的权值,提高重要节点的 DWCN-NodeRank 值,降低不重要节点的 DWCN-NodeRank 值。

公式(2)表明其符合 3 个条件:(1)节点连接数

条件。与节点 v 相连的节点个数越多,则说明节点 v 的重要度越高,则 $NR(v)$ 值越高,这与直观判断相符;(2)连接源节点级别条件。节点 v 的连接源节点 v_i 级别越高,即 $NR(v_i)$ 值越高,则 $NR(v)$ 值越高,说明节点 v 的重要度越高;(3)权重条件。节点 v 所占连接源的权重越高,则 $NR(v)$ 值越高,则说明节点 v 的重要性越高。

2.2 节点重要性评估方法

定义 1 邻接矩阵 M 。 M 用来描述具有 n 个节点的有向加权网络的连接关系,当两个节点之间有连接,则该矩阵元素表示两个节点之间的权值(或连接强度),否则该矩阵元素为 0

$$M = (m_{ij})_{n \times n} = \begin{cases} w(v_i, v_j) & (v_i, v_j) \in E \\ 0 & \text{其他} \end{cases} \quad (4)$$

定义 2 悬虚节点。假设节点 v_i 没有向外的连接,那么邻接矩阵 M 的第 i 行的值全为 0,则节点 v_i 称为悬虚节点。

定义 3 概率转移矩阵 Q 。将邻接矩阵 M 的每一行除以该行非零元素总和,进行归一化,得到概率转移矩阵 Q ,该矩阵元素记录一个节点到达另一个节点的概率

$$Q = (m_{ij} / \sum_{j=1}^n m_{ij})_{n \times n} \quad (5)$$

定义 4 处理过悬虚节点的概率转移矩阵 Q' 。为了便于节点重要度计算,定义矩阵 Q' 为处理过悬虚节点的概率转移矩阵,即用向量 $(1/n)e^T$ 来替换矩阵 Q 中元素全为 0 的行

$$Q' = Q + \eta \times (1/n)e^T = (m_{ij} / \sum_{j=1}^n m_{ij} + \eta \times 1/n)_{n \times n} \quad (6)$$

式中: η 表示标记节点是否为悬虚节点,等于 0 表示节点 v_i 不是悬虚节点,等于 1 表示节点 v_i 是悬虚节点。

定义 5 DWCN-NodeRank 矩阵 A 。矩阵 A 是根据 DWCN-NodeRank 值的计算公式(2)得到的

$$A = \sigma \times Q'^T + (1-\sigma) \times \frac{1}{n} \times E^T = \sigma \times Q'^T + (1-\sigma) \times \frac{1}{n} \times e \times e^T \quad (7)$$

式中: σ 为阻尼系数; Q'^T 为处理过悬虚节点的概率转移矩阵 Q' 的转置矩阵,使用转置矩阵考虑的是 DWCN-NodeRank 计算中主要关注节点的链入关系而不是链出。 $E^T = e \times e^T$ 为全 1 的行列式。

基于上述定义和 DWCN-NodeRank 评估指标,本文提出的基于 PageRank 的节点重要性评估方法如图 1 所示。

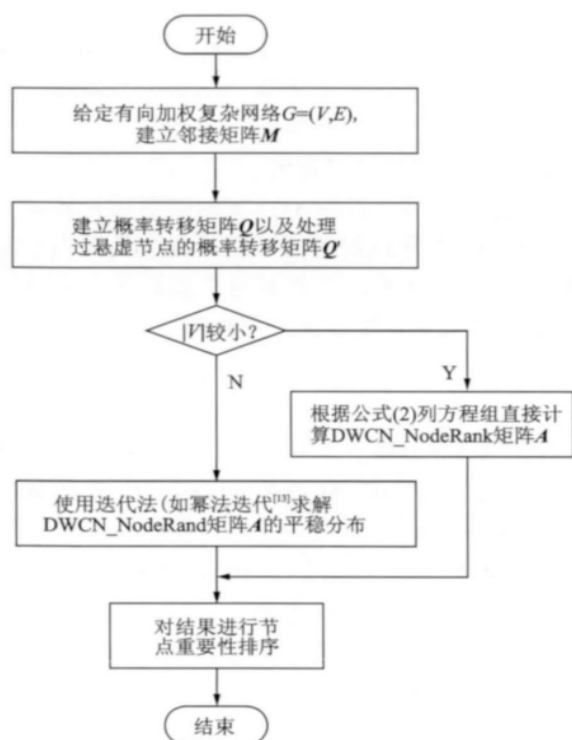


图1 节点重要性评估方法

2.3 复杂度分析

定理1 本文提出的节点重要性评估方法的空间复杂度为 $O(n^2)$, 时间复杂度为 $O(cn^2)$, 其中 c 为迭代次数。

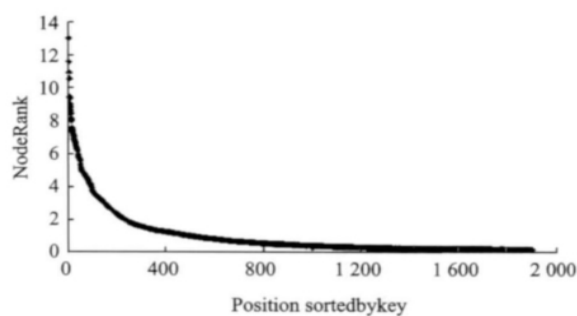
证明 本文使用一维数组存储相关矩阵, 邻接矩阵 M 及 DWCN-NodeRank 矩阵 A 一般均为稀疏矩阵, 因此, 评估方法的空间复杂度为 $O(n^2)$ 。对于时间复杂度而言, DWCN-NodeRank 值的计算是一个迭代过程, 其中矩阵与向量相乘所需的时间复杂度为 $O(n^2)$, 假设迭代次数为一个常数 c , 因此, 评估方法迭代计算的总体时间复杂度为 $O(cn^2)$, 最后使用排序算法对 DWCN-NodeRank 值进行排序, 排序算法通常的时间复杂度为 $O(n^2)$, 因此, 评估方法总的时间复杂度为 $O(cn^2)$, 其中 c 为迭代次数。证毕。

3 实验分析

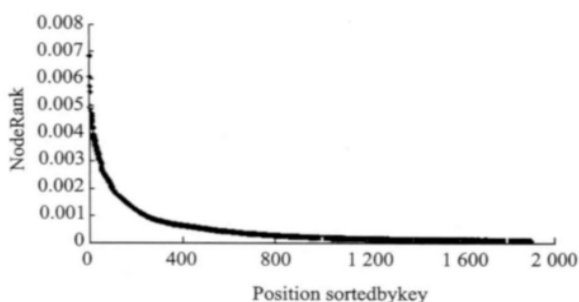
本节通过仿真实验来验证 DWCN-NodeRank 评估指标和方法的有效性和准确性。实验使用课

题组研制的复杂网络仿真平台, 复杂网络数据集来源于加州大学欧文分校学生在线交际网络 Facebook, like 真实数据集^[14], 其共有 1 899 个节点, 20 296 条有向边, 有向边表示用户之间信息传递的发送者到接受方的连接及信息传送方向, 以两个用户之间发出或接收的信息条数为权值, 权值之和为 59 835。

使用本文提出的节点重要性评估方法对 Facebook, like 数据集做重要性分析。一般来讲, 阻尼系数越大更真实, 而且越能更好地区分节点相对重要性, 突出重要节点的重要度, 同时弱化非关键节点的重要度。但同时, 阻尼系数越大, 迭代次数越多, 算法收敛速度越慢。因此, 综合考虑上述因素, 本文设置阻尼系数 $\sigma = 0.85$, 得到按 DWCN-NodeRank 值从大到小排序的相对重要度曲线图 (图 2)。横坐标表示按重要度高低排列的位置, 1 表示重要度最高位, 1 899 表示重要度最低位, 纵坐标为节点 DWCN-NodeRank 值。



(a) 节点初始向量均为1



(b) 节点初始向量为入强度

图2 Facebook, like 数据集按关键节点排序后的相对重要度

图 2(a) 设置节点初始向量均为 1, 图 2(b) 设置初始向量为节点入强度, 可以看出虽然重要度大小有区别, 但是重要度分布曲线是完全吻合的。

根据按重要度排序后的节点序号, 统计出对应

节点入度和入强度分布图,以及对应节点的源节点占排名前10%的节点个数分布图(图3),横坐标表示节点关键度高低排名。

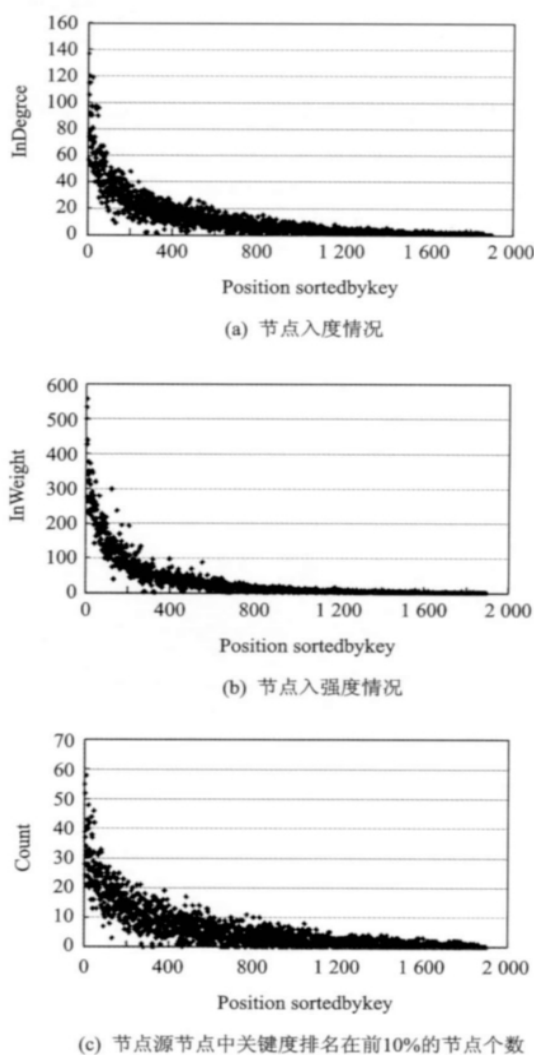


图3 节点入度、入强度和源节点关键度分布情况

从图3可以看出,虽然节点入度、入强度、源节点关键度与节点排名之间并不具有严格的递减函数关系,然而大体分布与重要度分布是相符的,这与本文 DWCN-NodeRank 算法提出的思想相吻合,从而验证了本文提出方法的有效性。

4 结束语

复杂网络节点重要性评估对于提高复杂网络的鲁棒性和抗毁性具有重要意义。目前研究主要集中在不含权重的无权复杂网络上,还无法反映具有方向和权值大小的现实复杂网络的实际情况。

因此,为了提高复杂网络重要性节点评估的实用性,本文基于有向加权复杂网络模型,借鉴了传统的 PageRank 排名算法,并将其引入有向加权复杂网络节点重要性评估中,提出节点重要性评估的新指标 DWCN-NodeRank 及其评估方法,该指标即体现了节点局部连接特性,又从全局角度体现了有向加权复杂网络中整体链接关系对节点重要性的影响。仿真实验结果表明该方法能快速、准确、有效地评估有向加权复杂网络节点的重要性。

参考文献:

- [1] Landherr A, Friedl B, Heidemann J. A critical review of centrality measures in social networks [J]. Business & Information Systems Engineering, 2010, 2(6): 371-385.
- [2] Kermarrec A, Merrer E L, Sericola B, et al. Second order centrality: Distributed assessment of nodes criticality in complex networks[J]. Computer Communications, 2011, 34(5): 619-628.
- [3] 谭跃进,吴俊,邓宏钟.复杂网络中节点重要度评估的节点收缩方法[J].系统工程理论与实践,2006, 26(11): 79-83.
Tan Yuejin, Wu Jun, Deng Hongzhong. Evaluation method for node importance based on node contraction in complex networks[J]. Systems Engineering-Theory & Practice, 2006,26(11): 79-83.
- [4] 吴俊,谭跃进,邓宏钟,等.考虑级联失效的复杂网络节点重要度评估[J].小型微型计算机系统,2007,4(4): 627-630.
Wu Jun, Tan Yuejin, Deng Hongzhong, et al. Evaluating node importance considering cascading failure in complex load-networks [J]. Journal of Chinese Computer Systems, 2007,4(4): 627-630.
- [5] 周璇,张凤鸣,李克武,等.利用重要度评价矩阵确定复杂网络关键节点[J].物理学报,2012,61(5): 1-7.
Zhou Xuan, Zhang Fengming, Li Kewu, et al. Finding vital node by node importance evaluation matrix in complex networks[J]. Acta Physica Sinica, 2012,61(5): 1-7.
- [6] 王甲生,吴晓平,廖巍,等.改进的加权复杂网络节点重要度评估方法[J].计算机工程,2012,38(10): 74-76.
Wang Jiasheng, Wu Xiaoping, Liao Wei, et al. Improved method of node importance evaluation in weighted complex networks[J]. Computer Engineering, 2012,38(10): 74-76.
- [7] 李玉华,贺人贵,钟开.动态加权网络中节点重要度

- 评估[J]. 计算机科学与探索, 2012, 6(2): 134-143.
- Li Yuhua, He Rengui, Zhong Kai, et al. Node importance evaluation in dynamic weighted network[J]. Journal of Frontiers of Computer Science and Technology, 2012, 6(2): 134-143.
- [8] 张益. 一种定量评估复杂网络节点重要度的算法[J]. 计算机工程, 2011, 37(20): 87-88.
- Zhang Yi. Quantitative evaluation algorithm for node importance of complex networks[J]. Computer Engineering, 2011, 37(20): 87-88.
- [9] Barrat A, Barthélemy M, Vespignani A. Modeling the evolution of weighted networks[J]. Physical Review E, 2004, 70(6): 066149-1-066149-12.
- [10] 王桂英, 周健, 谢颢. 基于 BBV 的有向加权网络模型[J]. 计算机工程, 2010, 36(12): 141-143.
- Wang Guiying, Zhou Jian, Xie Yang. C-T directed weighted network model based on BBV[J]. Computer Engineering, 2010, 36(12): 141-143.
- [11] 胡满玉. 基于链接关系的有向加权复杂网络关键点识别技术研究[D]. 南京: 南京理工大学计算机科学与工程学院, 2012.
- Hu Manyu. Research on evaluation of vital nodes in directed weighted complex networks based on link relation[D]. Nanjing, China: School of Computer Science and Engineering, Nanjing University of Science & Technology. 2012.
- [12] Page L, Brin S, Motwani R, et al. The PageRank citation ranking: bringing order to the web [EB/OL]. <http://ilpubs.stanford.edu:8090/422/>. 2012-05-13.
- [13] 李庆扬, 王能超, 易大义. 数值分析[M]. 北京: 清华大学出版社. 2003.
- Li Qingyang, Wang Nengchao, Yi Dayi. Numerical analysis [M]. Beijing: Tsinghua University Press. 2003.
- [14] Tore O, Pietro P, Kathleen M. C. Patterns and dynamics of users' behaviour Interaction: network analysis of an online community[J]. Journal of The American Society For Information Science And Technology, 2009, 60 (5): 911-932.