

基于网络结构的节点中心性排序优化算法^{*}

张 凯, 马英红

(山东师范大学 管理科学与工程学院, 济南 250014)

摘 要: 针对社交网络中节点中心性排序算法存在的不足, 从网络结构的角度提出了一种准确有效的节点中心性排序算法(CentraRank)。运用新浪微博数据和随机数据的模拟实验证明了算法的可行性和有效性, 并根据佩龙—佛罗贝尼乌斯定理证明算法的收敛性。实验结果表明, 该算法能够克服其他中心性算法的缺陷, 在精度和收敛度方面均有所提升。最后在该算法的基础上提出了一种基于网络结构的边中心性排序优化算法(Edge-Rank), 并验证了算法的正确性。

关键词: 中心性; 排序算法; 网络结构; 精度

中图分类号: TP393; TP301.6

文献标志码: A

文章编号: 1001-3695(2016)09-2596-05

doi: 10.3969/j.issn.1001-3695.2016.09.007

Centrality ranking algorithm based on network structure

Zhang Kai, Ma Yinghong

(School of Management Science & Engineering, Shandong Normal University, Jinan 250014, China)

Abstract: To solved the problem that nodes centrality ranking algorithm of networks worked badly in quality, this paper developed an efficient algorithm based on network structure. This paper applied the random data to confirm the feasibility and validity and used the Perron-Frobenius Theorem to prove the convergence of the algorithm. Simulation experiments show that the algorithm is more effective than the compared ones in the precision and convergence. On the bases of nodes centrality ranking algorithm, this paper proposed the edges centrality ranking algorithm and then verified the validity.

Key words: centrality; ranking algorithm; network structure; precision

0 引言

信息技术的快速增长带动了生活的网络化, 人们的生活中也存在各种各样的网络: 在线社交网络、科研网络以及交通网络等。社会网络是人们通过各种关系建立起来的联系, 并通过成员之间的交互作用形成的一种网络化结构。社会网络分析法就是对于社会网络的关系结构或者属性进行分析, 行动者可以是人、社区或者群体等, 他们之间的关系能够反映出一定的现象或者数据。诸多数据可以表示为二元图 $G=(V, E)$, 其中的 V 表示研究的成员集合, E 是成员之间的关系的集合。社会网络的一个突出特点是存在少数的关键节点, 这类节点对研究网络的功能和保持网络的稳定性具有重要的作用。例如网络在遭受外界蓄意攻击时, 关键节点遭到攻击就会导致整个网络瘫痪。研究网络节点重要(中心)性排序是亟待解决的问题, 这对处理信息流、预防传染性的传播以及网络上的其他重要的行为具有非常重要的意义。王林等人^[1]通过社会网络、交通网络等几个实例指出了中心化的意义。孙睿^[2]介绍了国内外节点重要性排序的研究现状, 指出节点重要性排序主要涉及基于网络结构和基于节点属性的重要性排序两个方面。

基于网络结构的节点重要性排序衡量指标较多, 主要研究有四个方面: a) 利用网络局部属性的指标进行节点中心性排序, 常见的是度排序算法^[3,4], 在节点中心性衡量方面占据较

大优势; b) 利用网络全局属性的指标进行节点中心性排序, 如特征向量^[5]、邻接度^[6]以及紧密性^[7]等, 这些算法往往复杂度较高, 但是考虑的网络信息也更加全面; c) 利用网络位置的节点中心性排序算法, 最具代表性的就是 K-核分解^[8,9]算法; d) 利用随机游走^[10]进行节点中心性排序, 这种方法在基于节点属性中比较常见, 也可以利用网络结构判断节点的中心性。李星等人^[11]借鉴随机游走的思想, 提出了一种中心性的快速算法, 并证明了算法在复杂度方面有一定的改进。

许多基于节点属性衡量指标都是基于 PageRank 算法和 HITS 算法, 并结合相关的网络属性进行排序。Cha 等人^[12]从粉丝数量、转发次数以及引用次数三个方面衡量节点用户的重要性, Romero 等人^[13]运用 HITS 算法结合粉丝的消极相关性提出新的中心性衡量指标。研究社会网络的重要任务之一就是对于节点的中心性进行排序, 不同网络的结构存在差异性, 所以如何有效地评估节点在不同网络结构中的中心性是研究社会网络关键问题之一。目前, 基于信息流的边排序算法并不少见。Fortunato 等人^[14]用移除网络中的边而导致的网络效率改变值来衡量边的中心性; Newman 等人^[15]用边上经过的最短路径的条数定义边中心性; Meo 等人^[16]提出 K-path 指标, 该指标基于信息传播能力计算网络中边的重要性。

PageRank 算法最初是用来衡量 Google 网页重要性的算法。PageRank 算法已应用到很多领域, 有学者将算法理论和

收稿日期: 2015-05-08; 修回日期: 2015-06-23 基金项目: 国家自然科学基金资助项目(71471106); 高等学校博士学科点专项科研基金资助项目(20133704110003)

作者简介: 张凯(1989-), 女, 山东德州人, 硕士研究生, 主要研究方向为社交网络(970055982@qq.com); 马英红(1971-), 女, 教授, 主要研究方向为复杂网络。

数学的知识结合解决 PageRank 算法的收敛速度问题^[17];有将传统 PageRank 算法应用到网页排序方面。例如戚华春等人^[18]提出了 PageRank-Time 算法,解决了原算法偏重旧网页问题;李凯等人^[19]将网页被点击次数作为集体个性化向量应用于 PageRank 算法中,提高了搜索引擎查询结果的精度。也有学者将 PageRank 算法应用到社会网络的领域,用于网络中节点的中心性排名。Lyu 等人^[20]提出 LeaderRank 算法对网络的节点进行排序,结果表明 LeaderRank 算法比 PageRank 算法排序更精准;陈少钦等人^[21]提出了一种实时中心性算法(MU-Rank);刘建国等人^[22]指出在复杂网络中节点排序方面,PageRank 算法考虑网络的全局拓扑特性,但是也忽略了一些实际因素。PageRank 算法在社会网络领域都是结合节点属性指标进行中心性排序,基于网络结构并与节点度中心性、中介中心性排序算法结合的不多见。本文借鉴经典的 PageRank (以下简称 PR) 算法的节点中心性排序的思想,加入紧密性和中介性指标,提出了 CentraRank 中心性算法模型,并证明了 CentraRank 的收敛性。通过搜集的新浪微博数据证实了算法的可行性和有效性。用度排序指标作为原始的中心性节点,加快算法的收敛速度;并在此基础上提出基于节点的边中心性排序算法(EdgeRank),并验证了算法的正确性。

1 算法的相关准备

本文研究的网络主要是有向无权网络,复杂网络可以表示成一个二元图 $G(V, E)$, 其中的 V 表示所有节点或顶点的集合, E 是图中边的集合, 节点之间的关系可以用邻接矩阵表示。

在网络中成员的重要性往往用节点的中心性进行衡量。中心性衡量指标可以分为四大类^[23]: a) 节点的度中心性, 描述一个节点是如何被连接的; b) 紧密性表示一个节点抵达其他节点的难易程度; c) 中介性表示一个节点在网络中与其他节点连接的重要性; d) 邻居节点度的大小反映节点邻居的重要程度。

节点的度中心性。用度测量的中心性是网络中最直接的地位指标, 一个节点的度中心性可以表示为

$$d_i(g)/(n-1) \quad (1)$$

式(1)描述节点的度中心性, 其取值范围为 $(0, 1)$, 度中心性越接近 1 表示节点越重要。但是度测量的中心性指标只能衡量网络的局部信息。紧密性指标考察给定的节点和任意其他节点的紧密程度, 该指标表示为

$$m_i = (n-1) / \sum_{k \neq j} P(k, j) \quad (2)$$

尽管紧密性能够衡量一个节点和其他中心性节点的接近程度, 但该算法仅考虑节点局部接近性, 没有远程连接和网络的整体结构。中介性指标是测量节点在网络中的位置优势, 中介性指标表示为

$$Ce_i = \sum_{k \neq j: i \in \{k, j\}} \frac{P_i(k, j) / P(k, j)}{(n-1)(n-2)} \quad (3)$$

节点的中介性越大越能够反映它在网络传播中的重要位置, 中介性针对节点的局部位置的关键性进行衡量有比较显著的效果。邻居度表示的是在网络中节点随机选取它的一个邻居度分布的情况, 这个指标认为一个节点的重要性是由它的邻居的重要性来决定的。邻居节点度分布可以表示为

$$P(d) = \frac{P(d)}{\langle d \rangle}, \langle d \rangle = \sum_d P(d) d \quad (4)$$

PageRank 算法也是衡量节点中心性的传统算法之一, PageRank 的数学模型如下:

$$PR(A) = (1-d) + d \left(\sum_{i=1}^n \frac{PR(T_i)}{C(T_i)} \right) \quad (5)$$

其中: $PR(A)$ 为网页 A 的 PageRank 值; $PR(T_i)$ 为链接到网页 T_i 的 PageRank 值; $C(T_i)$ 为网页 T_i 的出度数量; d 为阻尼系数 $0 < d < 1$, 通常 d 被设定为 0.85。从式(5)中可以看出算法考虑了度和出(入)度对于节点结构的影响。一个网页的入度越多表示该网页越重要。

2 节点中心性算法的优化

本文从网络的结构出发, 提出了衡量节点中心性的新算法, 并验证了算法的正确性和有效性。

2.1 算法优化思想

2.1.1 算法模型的优化

本文用 7 个点的无向网络(图 1)^[24]说明目前算法存在的问题以及改进的原因。表 1 给出了图 1 中节点的中心性比较。

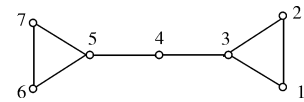


图 1 7 个点的无向网络

表 1 图 1 中的节点中心性的比较

中心性指标	节点 1、2、6 和 7	节点 3 和 5	节点 4
度	0.33	0.5	0.33
紧密性	0.4	0.55	0.6
中介性	0	0.53	0.6
PageRank	0.33	0.5	0.33

由图 1 和表 1 中可以得出以下结论:

a) 考虑度中心性。从表 1 中的结果可以看出节点 4 的中心性与节点 1、2、6 和 7 的相等, 但从图 1 中看出节点 4 的中心性要远远大于节点 1、2、6 和 7 的中心性。如果去掉节点 4, 网络的结构会发生变化, 对于信息传输产生很大的影响, 所以度中心性衡量该网络时会出现偏差。

b) 考虑 PageRank 算法的中心性。从式(5)中看出 PageRank 算法仅考虑了节点的度和邻居的特点, 所以和度中心性产生了相同的结果, PageRank 算法在衡量该网络时也会导致偏差。

c) 紧密性和中介性的值衡量节点在抵达其他节点的难易程度和重要程度。从表 1 中看出在 PageRank 算法中缺乏这方面的指标, 并且从表 2 中可以看出紧密性和中介性指标与 PageRank 算法的相关性不大, 所以将这两个指标和 PageRank 算法线性结合能够达到互补的效果。选取均值是因为在式(2)和(3)中可以看出两个指标均与最短路径有关系, 并且两个指标代表不同的意义, 如果单独考虑三个指标需要设置两个调整参数, 增加模型的复杂度。

刘建国等人^[22]在节点重要性排序进展中也指出四大中心性指标各有利弊: 度指标简单直观, 但是只反映了节点的局部特征; 邻居度能够考虑不同邻居的影响; 中介性考虑了节点的负载能力; 紧密性考虑节点的局部接近性, 依赖于网络的拓扑结构。他们在最后的结论中还指出中心性指标均是在不同的角度评价节点。付立东等人^[25]指出利用全局和局部结合的角度, 设计出一种基于结构的中心性度量方法成为必要。

所以本文提出新的算法模型, 既保留 PageRank 算法的优点, 又加入紧密性和中介性指标进行调整, 简单地线性相加能够将不同指标的优缺点组合, 得到的新算法命名为 CentraRank (以下简称 CR 算法)。由于不同的指标适用于不同的网络, μ

能够根据实际网络左右调整,使本文的模型能更符合网络特点。算法的模型如下:

$$CR_i = \mu \sum_{j=1}^n w_{ji} CR_j + (1 - \mu) C_i \quad (6)$$

式中各参数的含义如下:

a) CR_i 表示节点 i 的中心性值,是衡量节点中心性的最终指标。

b) W 是邻接矩阵。有向网络中的邻接矩阵是仅包含 0 和 1 的矩阵,且

$$w_{ji} = \begin{cases} 1 & (v_j, v_i) \text{ 是 } G \text{ 中的边} \\ 0 & \text{否则} \end{cases} \quad (7)$$

c) μ 是调整系数,是用来衡量原始 PR 算法与紧密性和中介性指标占据的权重大小。 μ 取值会决定参数占据的比例。 μ 较大,说明 PR 算法占据的权重大。本文确定调整系数采取的实证法,通过具体的案例获取最佳的调整系数。

d) C_i 是节点 i 的中介性和紧密性指标的均值。

$$C_i = (m_i + Ce_i) / 2 \quad (8)$$

其中: m_i 是节点 i 的紧密性值,计算方法如式(2)所示; Ce_i 是节点 i 的中介性值,计算方法如式(3)所示。

2.1.2 初始中心值的选取

虽然初始值选取并不影响最终的结果,但是初始中心值选取对于算法至关重要,能在一定程度上减少算法的运行次数,提升算法的性能。由于本文算法延续了 PR 算法思想,所以在选取初始值时以各种指标和 PR 算法的皮尔逊相关性为依据,如表 2 所示。

表 2 各指标的皮尔逊相关性

	中介性	入度	出度	总度	紧密性
PR	0.140	0.622	0.146	0.564	0.120

在表 2 中可以看出,入度指标和 PR 算法的最终结果的相关性最大。在算法运行过程中能够加速算法的收敛、降低算法的运行次数,所以本文选取网络的入度指标作为初始中心值。

数据量是影响算法复杂度的一个非常重要的因素。原算法为降低运行次数采取了 0~1 的随机数作为初始中心值。本文为进一步地提升算法的性能,将标准化的入度指标作为初始的中心值,并在结果中证明了本算法的优势。

2.2 算法设计

算法的具体步骤如下:

- 输入邻接矩阵 P_i ,并将矩阵进行标准化,得到矩阵 P ;
- 根据式(2)和(3)计算每个节点的 m_i 和 Ce_i 值,并根据式(8)得出每个节点的 C_i 值;
- 选取式(1)中标准化的网络度值作为初始的中心性值 CR' ,并确定 μ 的值;
- 对每一个节点 i 根据式(6)依次计算每个节点的 CR_i 值;
- 判断结果的收敛性,若 $\max(\text{abs}(CR' - CR)) > 0.0001$,输出 CR ;否则重复步骤 d)。

2.3 算法收敛性证明

算法主要取决于式(6),因此,只需要证明式(6)的收敛性即可。而式(6)的收敛性等价于 $CR_i = \mu \sum_{j=1}^n w_{ji} CR_j$ 的收敛性。

证明 对每一个节点 i ,有 $CR_i = \mu \sum_{j=1}^n w_{ji} CR_j$ 。因此,对网络中所有节点, $CR = \mu \cdot W \cdot CR$ 。对于给定 $n \times n$ 矩阵 W , CR 即为对应 μ 的特征向量。式(6)的收敛性等同于求特征向量的唯一性。根据佩龙—佛罗贝尼乌斯定理^[26],如果非负的随机

矩阵 W ,使得其列的每个项相加等于 1,存在一个唯一的非负特征值的单位特征向量(特征值为 1 的特征向量)。若 $\mu = 1$,标准化的矩阵满足佩龙—佛罗贝尼乌斯定理,则一定存在唯一的非负的 CR ,则算法一定是收敛的。若 $\mu \neq 1$,则令 $s = CR/\mu$,原式转换成 $s = Ws$,同样特征值为 1。因此,特征向量是唯一的,故算法收敛。

经过以上的证明过程可知,式(6)是收敛的。

3 节点中心性算法实验结果与分析

新浪微博中的“关注”是一种单向、无须对方确认的关系,只要用户喜欢就可以关注对方,通过新浪微博的关注关系构建的网络更能体现用户的主动需求。网络数据来自于社交网站新浪微博,利用 253 个单个用户,采集用户之间的关注关系,如节点 i 关注节点 j ,那么 $i \rightarrow j$ 构成有向连接。数据的时间点是 2014 年 12 月。

新浪微博关注关系网络由 Pajek 软件实现可视化,如图 2 所示。网络中节点代表新浪微博的用户,如节点 1 关注节点 2,那么 $1 \rightarrow 2$ 构成有向连接。网络共有 253 个节点 510 条边。从图中看出节点之间的互动比较少,这与新浪微博的性质相符。新浪微博是一种弱关系构建的网络,网络中关注倾向于在陌生人之间展开,呈现一种比较稀疏的状况。数据是以自我为中心构建的网络,所以网络中不存在孤立的节点。网络中大部分节点的总度比较小,仅少数节点的总度较大,这和大部分的实际网络基本相符。新浪微博网络是比较常见的一类有向网络,并且通过关注关系构建更能体现用户的主观需求,用于算法的结果分析更准确。

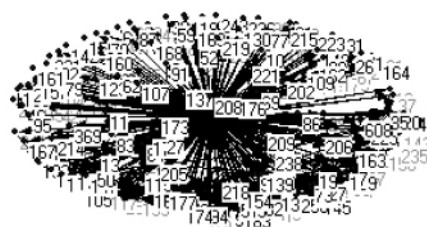


图 2 新浪微博关注关系网络

3.1 μ 值的确定

无论对于 PR 算法还是 CR 算法来说, μ 值的确定都是一个难点。本文采用实际数据拟合,用误差来区分不同取值的效果,误差值越小,效果就越好。由以上的证明可知,式(6)中的 μ 是求解 CR 的关键。为了求解 μ 的值,利用图 2 中的新浪微博数据进行实证计算。

μ 值表示核心算法模型中的调整系数,为了获取最佳值,本文采取误差大小评定 μ 值。本文误差的公式为 $(x_i - x_{i-1})^2$,其中 x_{i-1} 和 x_i 分别代表在排序结果中相邻两点的 CR 值,由于 CR 算法的值非唯一性,本节在多次运算过程中取总误差大小均值评定 μ 值。分别取 $\mu = \{0.1, 0.2, \dots, 0.9\}$ 在 MATLAB 中模拟运行 100 次,获取 9 个对应的 μ 值进行比较。在图 3 中发现误差能够较好地区分 μ 值,当 $\mu = 0.8$ 和 $\mu = 0.9$ 时,误差的值最小,且误差是逐渐减小的,这与 PR 算法中的 0.85 非常接近。图 2 新浪微博案例中本文也选取 $\mu = 0.85$ 作为调整系数。

3.2 中心性排序结果对比

对于图 2 的新浪微博关注关系网络数据,本节依据 CR 算法选取了前八名的用户,并获取这些用户在节点总度、节点入度、中介性以及 PR 等代表性的排序结果进行对比,分析这些

用户的变化情况,结果如表3所示。

表3 各类中心性排序算法的排序结果对比

编号	用户名	总度	入度	中介性	PR	CR
237	张泉灵	2	1	6	1	1
240	赵晓	4	3	2	2	2
163	数据挖掘	5	4	4	4	3
166	宋丹丹	3	2	43	3	4
186	王子多	8	10	1	12	5
134	倪正东	9	9	3	8	6
51	爱国者冯军	7	5	8	6	7
248	周晓鹏	6	7	7	7	8

从表3中可以看出,PR算法的排序结果和入度相关性较大,但是没有考虑中介性等指标,将本文CR算法和PR算法的排序结果进行对比也可以看出本文算法的准确性。表中编号163和166的CR和PR的排序结果正好相反,用户“数据挖掘与数据分析”在CR算法中排名第3,但是在PR算法中排名第4,用户“数据挖掘与数据分析”比用户“宋丹丹”在CR算法中排名有所提前,是因为考虑表中的中介性指标,用户“数据挖掘与数据分析”的中介性排名远远大于用户“宋丹丹”,导致排名有所提前,表中编号51和248也出现了同样的情况。表中编号134和186的CR算法排名较PR算法排名都有上升,同样可以看出是中介性指标起的作用。

3.3 网络鲁棒性结果对比

本文还通过鲁棒性考察节点排名的变动更加符合实际。鲁棒性是进行系统分析的关键性指标,并且对于网络的稳定性具有重要的意义。本文中用最大连通子图来说明CR算法比PR算法的实际意义更加明确,使用图2新浪微博关注关系网络数据,在MATLAB 7.0软件和Windows XP系统环境下仿真,结果如图4所示。图中横坐标1,5,...,10分别代表在PR和CR算法中移除排名前1、前5以及前10名的节点;纵坐标代表最大连通子图中节点的个数。

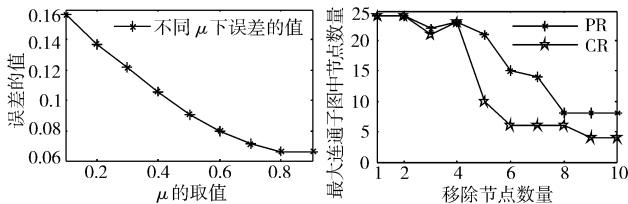


图3 不同 μ 值下的误差大小

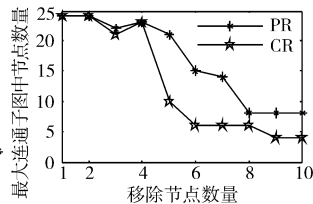


图4 恶意攻击下最大连通子图中节点数量的变化

在图4中可以看出,PR算法和CR算法在1、2和4这三个点出现重合,是因为PR算法和CR算法排在前两名的用户都是“张泉灵”和“赵晓”,并且两算法前四名的用户也完全相同;但是在后续的过程中,CR算法面对恶意攻击最大连通子图中节点数量变化得更快,这说明CR算法的排序结果对于网络的鲁棒性更重要,可以得出CR算法的结果更符合现实意义的结论。并且横坐标达到9之后,最大连通子图的数量趋于稳定,仿真过程终止。

3.4 算法精度分析

算法精度是衡量算法性能的重要指标之一,本文对比了五种不同算法的误差,用误差大小体现算法精度。误差的公式也采用 $(x_i - x_{i-1})^2$,其中 x_{i-1} 和 x_i 分别代表在排序结果中相邻两点的CR值,运用图2新浪微博关注关系网络数据,最终求得的结果如表4、图5所示。表4中1、8、16...分别代表前1、8、16项...的累加值,因为72项之后误差都为0,所以考察72项之前的项。

表4 各类算法误差比较

比较项	1	8	16	24	32	40	48	56	64	72
CR	0.0097	0.0445	0.0542	0.0565	0.0593	0.0601	0.0604	0.0609	0.0616	0.0618
PR	0.0171	0.0503	0.0566	0.0598	0.0622	0.0628	0.0632	0.0638	0.0646	0.0656
入度	0.026	0.0562	0.0632	0.0667	0.0684	0.0702	0.0702	0.0702	0.0702	0.0702
中介数	0.0424	0.1356	0.1539	0.1645	0.1668	0.1680	0.1681	0.1681	0.1681	0.1681

从表4和图5中能够得到以下结论:

a) 本文的算法精度要远远大于其他四种中心性排序算法(误差远远小于其他算法),在误差趋于稳定的过程中,算法的优势还是比较明显的,并且CR算法在节点数量非常大的情况下,最大平均误差为0.0618,远远小于PR的0.0656,所以能够接受该算法。

b) 在节点个数达到10之后,误差指标变化幅度较小,达到70左右基本保持不变,这是因为在新浪微博中的用户关注关系是一种有向弱关系,所以用户之间的双向信息交互行为为相对少,也就是说处于活跃的用户数量较少。在实际的社会网络中,处于中心位置的用户数量很少,大部分的节点都处在无交互的状态。实验结论与现实情况吻合。

3.5 时间复杂性分析

本节总共选取7个案例,前6个案例是随机生成的有向网络,第7个案例是图2的新浪微博关注关系网络,算法在MATLAB 7.0软件和Windows XP系统环境下平均运行100次,得到的结果如图6所示。图6中 n 表示矩阵的规模 $n = \{4, 8, 16, 32, 64, 128, 254\}$ 分别表示的是该点应用的案例是 $4 \times 4, 8 \times 8, \dots$ 的邻接矩阵构造的网络,依此类推。

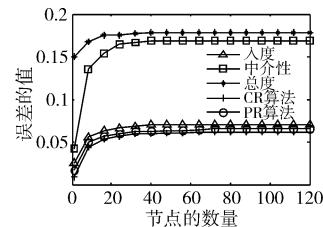


图5 各类中心性排序算法的精度对比

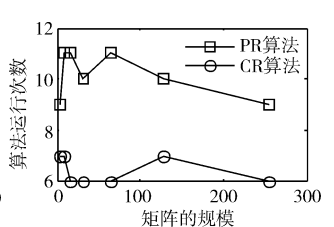


图6 PR和CR算法在不同矩阵规模下的运行次数对比

从图6中能够得到以下结论:

a) 矩阵规模和运行的次数不相关。伴随着矩阵规模的增大,运行次数呈现先降低后升高的现象,表明两者不存在一定的相关关系。

b) 在同等网络规模下,本文算法的运行次数少,远远快于原来的PR算法。例如在 $n = 32$ 的随机网络中,算法运行100次之后得到的PR平均运行结果为10,得到的本文算法的平均运行次数是6次,结果显示了本文算法的可行性。并且在图中还能够看出伴随网络规模的增加,算法的运行次数并不增加。

4 基于节点的边中心性排序算法

本文从网络结构的角度,在节点中心性算法的基础上将网络中的边转换成网络中的节点,提出一种基于节点的边中心性排序EdgeRank算法(以下简称ER算法)。

4.1 算法思想

ER算法的构造如下:

a) 将网络中的边转换成点,然后运用点中心性排序算法解决边排序问题。转换方法如图7所示。其中(a)代表的是有向网络转换,图7中上半部分的边1、2和3转换成下半部分的点1、2和3,转换前后均是有向网络;(b)代表的是无向网络

转换 图中上半部分的边 1、2 和 3 转换成下半部分的点 1、2 和 3, 转换前后均是无向网络。

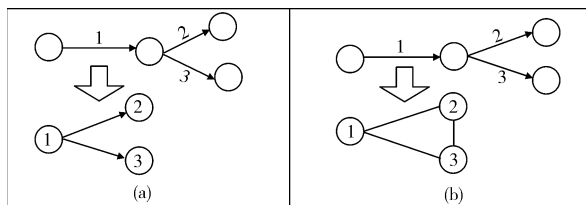


图7 有向和无向网络转换模型

为了说明网络转换前后的等价性, 本文运用网络的度指标和中介性指标来说明。在图7中得出有向网络中编号为1的边出度为2, 编号2和3的边的位置平行, 并且入度都为1, 通过编号为1的边才能够到达编号为2和3的边, 转换后的图形也是如此。无向网络中三条边是全连通的, 转换后的网络是完全图。可以看出这种转换方法不仅能够保证从边到点的等价替换, 还能够保证各边在网络中的位置不变, 转换之后的模型还能够简化网络, 容易计算边的各种属性值。

b) 用 CentraRank 算法对转换后的网络排序, 得到网络边中心性排序结果。

4.2 算法准确性分析

4.2.1 无向网络算法准确性分析

迈克尔罢工网络是进行边中心性算法的经典案例^[27], 网络中包含24个点和37条边以及3个社团。网络描述的是1973年发生的一场罢工活动, 最初指定的 Sam-Wendle 不能够及时终止罢工活动, 迈克尔指出 Bob-Norm 这条边最利于终止罢工, 这条边位于三个社团的重叠部分, 在网络中发挥着最重要的作用。Simko 运用 NetworGame 软件计算出迈克尔罢工网络中居于最中心的边是 Bob-Norm, 他们能够100%终止罢工, 并且指出最不利于罢工终止的边是 Sam-Wendle, 他们终止罢工的可能性仅有8%。无向网络边排序结果如图8所示。

从图8(其中横坐标表示转换之后的边序号)中可以看出, 17号边中心性排序结果为1, 正好对应边 Bob-Norm, 说明17号边在网络中发挥着领导者的作用, 在这条边的领导下能够迅速地终止罢工。37号边的中心性排序结果为37, 正好对应边 Sam-Wendle, 算法得出的排序结果不仅与 NetworGame 软件^[27]一致, 还完全符合实际情况, 都能够说明算法的正确性。

4.2.2 有向网络算法准确性分析

有向网络选取的是图2中新浪微博网络节点排序算法中前8名的用户构建的网络, 网络中包含8个节点和13条有向边, 进行网络转换后得到的排序结果如图9所示。其中横坐标表示转换之后的边序号。

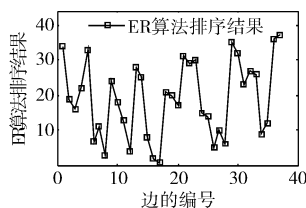


图8 无向网络边排序结果

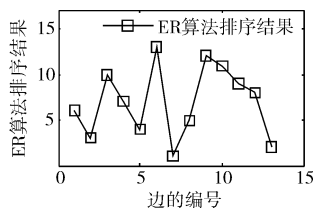


图9 有向网络 ER 算法排序结果

从图9可以看出, 序号为7的边位于第1名, 序号为6的边居于第13名, 单凭排序结果无法确定算法结果的优劣。为了更好地说明算法的性能, 采取鲁棒性对于排序结果进行测试, 说明边的中心性排序结果的实际意义。为了对边的重要性

和节点的重要性进行比较, 本文定义新指标进行对比分析。

网络的破坏率 = 总节点个数 -

$$\frac{\text{最大连通子图中点个数}}{\text{总节点个数}} \quad (9)$$

从式(9)可以看出网络的破坏率与边或者节点的中心性成正比, 破坏率越大, 说明该边或者节点在网络中的位置越重要。图10为有向网络中攻击节点和边的破坏率对比。其中, 横坐标表示边和节点排序的结果序号, 每次删除对应的一个点。实验是在 MATLAB 7.0 软件, Windows XP 系统环境下仿真的。从图10可以看出, 随着边排序编号的增加, 网络的破坏率递减, 说明了边排序算法的准确性。攻击节点和边网络的破坏率出现很大的不同, 攻击网络的边对于网络的破坏率明显增大, 在实际生活中也是如此, 切断一条交通要道往往使两地成为孤立的个体。攻击边导致网络的最大破坏率能够达到0.76%, 能够导致高达80%的网络瘫痪, 这对于实际生活中的网络能够产生较大影响; 而攻击网络的节点最高的破坏率仅有50%, 这与边的破坏率相比还差距比较大。

图11是攻击边的破坏率。其中横坐标表示边排序的结果序号总和, 如2就代表排在前2名的用户。ER 算法排序代表攻击在 ER 算法中居于前六名的边序号; random 表示在随机生成的数据抽取三组同规模的数据与 CR 算法对比。

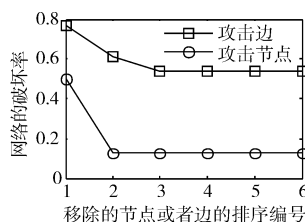


图10 有向网络中攻击节点和边的破坏率对比

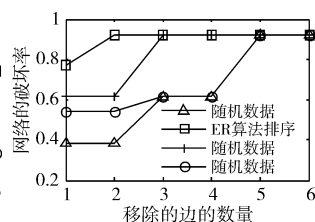


图11 有向网络中攻击边的破坏率

在图11中, 为了降低随机性的概率, 在破坏率前10名的随机数据中选取三组数据。ER 算法产生的网络破坏率远远大于随机数组, 由于网络中仅包含13条边, 所以在删除4条边之后网络的破坏率基本重合, 说明 ER 算法排序的准确性。

5 结束语

本文提出了节点中心性排序 CR 算法, 证明算法的收敛性, 并进行仿真实验验证算法的误差和收敛次数远远小于 PR 算法; 在此基础上提出了基于节点的边中心性排序 ER 算法, 证明了算法的正确性。本文还存在很多的不足: a) 仅仅考虑了网络结构指标, 忽视了节点的属性指标, 以后进一步将结构指标和属性指标进行融合, 定义更加符合实际的中心性排序算法; b) 将边转换成节点后, 网络中邻接矩阵规模增大, 复杂度提升。

参考文献:

- [1] 王林, 张婧婧. 复杂网络的中心化[J]. 复杂系统与复杂性科学, 2006, 3(1): 13-20.
- [2] 孙睿, 罗万伯. 网络舆论中节点重要性评估方法综述[J]. 计算机应用研究, 2012, 29(10): 3606-3608, 3628.
- [3] 任卓明, 邵凤, 刘建国, 等. 基于度与集聚系数的网络节点重要性度量方法研究[J]. 物理学报, 2013, 62(12): 12-18.
- [4] Freeman L C. A set of measures of centrality based on between-ess[J]. Sociometry, 1977, 40(1): 35-41.
- [5] Bonacich P. Factoring and weighting approaches to status scores and clique identification[J]. Journal of Mathematical Sociology, 1972, 2(1): 113-120.

(下转第2605页)

的平衡,仍有待考量。

5 结束语

本文提出了一种求解多目标调度问题的改进混沌烟花算法(CPAFWA),设计了一种双元锦标赛和动态淘汰制相结合的帕累托非劣解集构造方法,该策略在求取帕累托非劣解集上的时间复杂度明显优于传统的双元锦标赛法。为了避免算法陷入局部最优,提高求解质量,根据逻辑自映射产生混沌序列,对20%的种群进行混沌搜索。通过求解本文所构建的三目标作业车间调度模型,证明所提算法的有效性。

在Job-Shop标准问题上的仿真实验结果表明,混沌搜索过程可以有效提高作业车间调度问题的求解质量。然而,结果同时表明,该搜索过程会降低算法的整体时间效率,导致算法的实际运行时间较长。因此,如何在不影响算法求解质量的情况下提高混沌搜索效率仍值得进一步研究。

参考文献:

- [1] Baker K. Introduction to sequencing and scheduling [M]. New York: Wiley, 1974: 1-15.
- [2] 雷德明, 吴智铭. 基于粒子群优化的多目标作业车间调度[J]. 上海交通大学学报, 2007, 41(11): 1796-1800.
- [3] 覃朝勇, 刘向, 郑建国. 求解多目标 Job-Shop 生产调度问题的量子进化算法[J]. 计算机应用研究, 2010, 27(3): 849-852.
- [4] 王伟玲, 李俊芳, 王晶. 求解多目标作业车间调度问题的双种群遗传算法[J]. 计算机集成制造系统, 2011, 17(4): 808-815.
- [5] 任晓莉. LSO 改进 CGA 解决多目标作业车间调度问题[J]. 计算机应用与软件, 2015, 32(3): 60-64.
- [6] 肖晓伟, 肖迪, 林锦国. 等. 多目标优化问题的研究概述[J]. 计算机应用研究, 2011, 28(3): 805-827.
- [7] Tan Ying, Zhu Yuanchun. Fireworks algorithm for optimization [C]//Advances in Swarm Intelligence. Berlin: Springer, 2010: 355-364.
- [8] Ding Ke, Zheng Shaoqiu, Tan Ying. A GPU-based parallel fireworks algorithm for optimization [C]//Proc of the 15th Annual Conference on Genetic and Evolutionary Computation Conference. Yew York: ACM Press, 2013: 9-16.
- [9] 王培崇, 高文超, 钱旭. 等. 应用精英反向学习的混合烟花爆炸优化算法[J]. 计算机应用, 2014, 34(10): 2886-2890.
- [10] Zheng Yujun, Song Qin, Chen S Y. Multi objective fireworks optimization for variable-rate fertilization in oil crop production [J]. Applied Soft Computing, 2013, 13(11): 4253-4263.
- [11] 谭营, 郑少秋. 烟花算法研究进展[J]. 智能系统学报, 2014, 9(5): 515-528.
- [12] Coello C A, Lechuga M S. MOPSO: a proposal for multiple objective particle swarm optimization [M]. Washington D C: IEEE Press, 2002: 1051-1056.
- [13] Sha D Y, Lin H H. A multi-objective PSO for Job-Shop scheduling problems [J]. Expert Systems with Applications, 2009, 37(2): 1-6.
- [14] El-Gohary A, Al-Ruzaiza A S. Chaos and adaptive control in two prey-one predator system with nonlinear feedback [J]. Chaos, Solitons and Fractals, 2007, 34(2): 443-453.
- [15] 刘长平, 叶春明. 基于逻辑自映射的变尺度混沌粒子群优化算法[J]. 计算机应用研究, 2011, 28(8): 2825-2827.
- [16] Mehrabian A R, Lucas C. A novel numerical optimization algorithm inspired from weed colonization [J]. Ecological Informatics, 2006, 1(4): 355-366.
- [17] Beasley J E. OR-Library: distributing test problems by electronic mail [J]. Journal of the Operational Research Society, 1990, 41(11): 1069-1072.
- [18] Ponnambalam S G, Ramkumar V, Jawahar N. A multi-objective genetic algorithm for Job-Shop scheduling [J]. Production Planning and Control, 2001, 12(8): 764-774.
- [19] Zitzler E, Thiele L. Multi-objective evolutionary algorithms: a comparative case study and the strength Pareto approach [J]. IEEE Trans on Evolutionary Computation, 1999, 3(4): 257-271.
- [16] De Meoa P, Ferrarab E, Fiumara G, et al. A novel measure of edge centrality in social networks [J]. Knowledge-Based Systems, 2012, 30(6): 136-150.
- [17] Soon I Y, Koh S N. Speech enhancement using 2D Fourier transform [J]. IEEE Trans on Speech and Audio Processing, 2003, 11(6): 717-724.
- [18] 戚华春, 黄德才. 具有时间反馈的 PageRank 改进算法[J]. 浙江工业大学学报, 2005, 33(3): 272-275.
- [19] 李凯, 赫枫龄, 左万利. PageRank-Pro: 一种改进的网页排序算法[J]. 吉林大学学报: 理学版, 2003, 41(2): 175-179.
- [20] Lyu Linyuan, Zhang Yicheng, Yeung C H. Leaders in social networks, the delicious case [J]. PLoS ONE, 2011, 6(6): 1-9.
- [21] 陈少钦, 范磊, 李建华. MURank: 社交网络用户实时影响力算法[J]. 通讯技术, 2013(3): 88-92.
- [22] 刘建国, 任卓明, 郭强. 等. 复杂网络中节点重要性排序的研究进展[J]. 物理学报, 2013, 62(17): 178901.
- [23] Borgatti S P. Centrality and network flow [J]. Social Networks, 2005, 27(1): 55-71.
- [24] Jackson M O. Social and economic networks [M]. [S. l.]: Princeton University Press, 2008.
- [25] 付立东, 高琳. 模块密度谱分的复杂网络社团发现方法[J]. 西安电子科技大学学报, 2010, 37(5): 916-921.
- [26] Meyer C D. Matrix analysis and applied linear algebra [M]. [S. l.]: SIAM, 2007.
- [27] Simko G I, Csrmely P. Nodes having a major influence to break co-operation define a novel centrality measure: game centrality [J]. PLoS ONE, 2013, 8(6): 6-12.

(上接第2600页)

- [6] Bonacich P. Power and centrality: a family of measures [J]. American Journal of Sociology, 1987, 92(5): 1170-1182.
- [7] Jackson M O, Wolinsky A. A strategic model of social and economic networks [J]. Journal of Economic Theory, 1996, 71(1): 44-74.
- [8] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks [J]. Nature Physics, 2010(6): 888-893.
- [9] Carmi S, Havlin S, Kirkpatrick S. A model of Internet topology using k-shell decomposition [J]. Proceedings of the National Academy of Sciences of USA, 2007, 104(27): 11150-11154.
- [10] Li Ping, Zhang Jie, Xu Xiaoke, et al. Dynamical influence of nodes revisited: a Markov chain analysis of epidemic process on networks [J]. Chinese Physics Letters, 2012, 29(4): 048903.
- [11] 李星, 钟志农, 李洋. 一种随机游走中心性的快速算法[J]. 计算机应用研究, 2013, 30(8): 2337-2340.
- [12] Cha M, Haddadi H, Benevenuto F, et al. Measuring user influence in twitter: the million follower fallacy [C]//Proc of the 4th International Conference on Weblogs and Social Media. 2010: 97-105.
- [13] Romero D M, Galuba W, Asur S, et al. Influence and passivity in social media [C]//Machine Learning and Knowledge Discovery in Databases. Berlin: Springer-Verlag, 2011: 113-114.
- [14] Fortunato S, Latora V, Marchiori M. A method to find community structures based on information centrality [J]. Physical Review E, 2004, 70(5): 056104.
- [15] Newman M, Girvan M. Finding and evaluating community structure in networks [J]. Physical Review E, 2003, 69(2): 26113.