

面向结构洞的复杂网络关键节点排序

韩忠明 吴杨 谭旭升 段大高 杨伟杰

Ranking key nodes in complex networks by considering structural holes

Han Zhong-Ming Wu Yang Tan Xu-Sheng Duan Da-Gao Yang Wei-Jie

引用信息 Citation: *Acta Physica Sinica*, 64, 058902 (2015) DOI: 10.7498/aps.64.058902

在线阅读 View online: <http://dx.doi.org/10.7498/aps.64.058902>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2015/V64/I5>

---

您可能感兴趣的其他文章

Articles you may be interested in

非均齐超网络中标度律的涌现-----富者愈富导致幂律分布吗?

Emergence of scaling in non-uniform hypernetworks-----does 搵 he rich get richer□ lead to a power-law distribution?

物理学报.2014, 63(20): 208901 <http://dx.doi.org/10.7498/aps.63.208901>

基于复杂网络理论的微博用户关系网络演化模型研究

An evolution model of microblog user relationship networks based on complex network theory

物理学报.2014, 63(20): 208902 <http://dx.doi.org/10.7498/aps.63.208902>

复杂网络中带有应急恢复机理的级联动力学分析

Analysis of cascading dynamics in complex networks with an emergency recovery mechanism

物理学报.2014, 63(15): 158901 <http://dx.doi.org/10.7498/aps.63.158901>

动态银行网络系统中系统性风险定量计算方法研究

Calculation of system risk in a dynamical bank network system

物理学报.2014, 63(3): 038902 <http://dx.doi.org/10.7498/aps.63.038902>

一种复杂网络路由策略的普适优化算法

A pervasive optimized algorithm for complex network routing strategy

物理学报.2014, 63(2): 028901 <http://dx.doi.org/10.7498/aps.63.028901>

# 面向结构洞的复杂网络关键节点排序\*

韩忠明<sup>†</sup> 吴杨 谭旭升 段大高 杨伟杰

(北京工商大学计算机与信息工程学院, 北京 100048)

(2014年7月13日收到; 2014年9月30日收到修改稿)

复杂网络中的结构洞节点对于信息传播具有重要作用, 现有关键节点排序方法多数没有兼顾结构洞节点和其他类型的关键节点进行排序. 本文根据结构洞理论与关键节点排序相关研究选取了网络约束系数、介数中心性、等级度、效率、网络规模、PageRank 值以及聚类系数 7 个度量指标, 将基于 ListNet 的排序学习方法引入到复杂网络的关键节点排序问题中, 融合 7 个度量指标, 构建了一个能够综合评价面向结构洞节点的关键节点排序方法. 采用模拟网络 and 实际复杂网络进行了大量实验, 人工标准试验结果表明本文排序方法能够综合考虑结构洞节点和核心节点, 关键节点排序与人工排序结果具有较高的一致性. SIR 传播模型评估实验结果表明由本文选择 TOP-K 节点发起的传播能够在较短的传播时间内达到最大的传播范围.

**关键词:** 结构洞, 关键节点, 排序学习, 复杂网络

**PACS:** 89.75.Hc

**DOI:** 10.7498/aps.64.058902

## 1 引言

复杂网络的研究一直备受关注, 为了解释不同网络所具有的不同特征, 研究焦点从网络的社团结构<sup>[1-3]</sup>到网络中的关键节点发现<sup>[4]</sup>, 关键节点对复杂网络上的信息传播、防止传染病的传播和扩散等现实情境起到重大作用. 网络中关键节点排序问题不仅与网络拓扑结构特征有关还应考虑到其功能特征因素. Burt 的结构洞理论<sup>[5]</sup>认为, 在社会结构中占据结构洞位置的企业会获得更多的竞争优势, 从而使企业获得累加收益, 包括信息利益和控制利益. 在信息社会中, 处于结构洞位置的成员能够获取更关键的信息, 从而影响甚至控制社会关系与信息的传播. 如图 1 所示, 该图为互联网上某个热点话题的传播图, 图中红色节点为处于结构洞的节点, 社交媒体上一些热点话题被拥有粉丝较多的蓝色节点发起, 而处于边缘地带的结构洞节点却带动了将近 50% 的话题传播量, 如果没有这些处于结构洞的节点发挥作用, 话题将不会有广泛的传

播范围, 这说明话题在传播过程中, 核心节点固然很重要, 但处于结构洞的节点也起到重大作用. 所以, 关键节点排序问题关注的焦点不能仅局限于网络中的核心节点, 也不可忽略处于结构洞位置的节点. 现有关键节点排序研究很少考虑结构洞节点, 如何在复杂网络的关键节点排序问题中综合考虑结构洞节点和其他类型的关键节点是一个值得深入研究的问题.

近年来, 复杂网络中的关键节点的排序问题已经成为研究热点. 定量的衡量网络中节点的重要程度通常会用到一些中心性度量指标, 如节点的度<sup>[6]</sup>、接近中心性<sup>[7]</sup>、流介数中心性<sup>[8]</sup>, Katz 中心性<sup>[9]</sup>等, 但是这些指标对网络的拓扑结构依赖性很强. 而 Kitsak 等<sup>[10]</sup>从新的视角提出利用 K-核分解来研究网络中的关键节点, 该方法认为关键节点的重要性与其所处网络的位置有关, 将外层的节点层层剥去, 处于内层的节点即为网络中的关键节点. 另外, 在搜索引擎领域, 排序问题的研究已经广为人知, 如著名的 Google 网页排名算法 Pagerank<sup>[11]</sup>, Lü 等提出的 LeaderRank 算法<sup>[12]</sup>以

\* 国家自然科学基金(批准号: 61170112)、中央财政支持地方高校发展专项资金人才培养和创新团队建设项目(批准号: 19005323132)、教育部人文社会科学研究基金项目(批准号: 13YJC860006)资助的课题.

<sup>†</sup> 通信作者. E-mail: hanzm@th.btbu.edu.cn

及 Hits 算法<sup>[13]</sup>等,但这些算法对于网络中处于结构洞位置的节点的识别能力有限.除此之外还有基于节点删除和收缩的排序方法<sup>[14-16]</sup>、基于互信息的节点重要性排序方法<sup>[17]</sup>等等,这些研究分别从不同角度对网络中的关键节点进行排序.

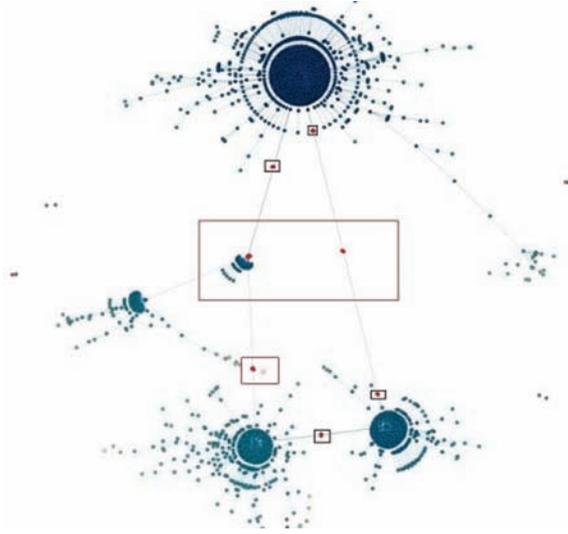


图1 (网刊彩色) 热点话题传播图

由于上述这些关键节点排序方法各有利弊,于是学者们寻找综合性的方法对网络中的关键节点进行评价.于会等<sup>[18]</sup>提出了一种多属性决策的排序方法,对网络中单个节点的度中心性、介数中心性等多个指标作为决策评价方案的属性进行综合计算,从而对网络中关键节点进行排序;Hou等<sup>[19]</sup>考虑度、介数、K-核三个不同指标对节点重要性的影响,采用欧拉距离公式计算三种不同指标的综合作用得到节点的排序结果;Comin等<sup>[20]</sup>综合考虑介数与度的关系,定义了一个关键节点排序的指标;任卓明等<sup>[21]</sup>提出了一种基于度与聚类系数的网络节点重要性度量方法对大规模网络的节点重要性进行有效分析;周漩等针对节点删除法、节点收缩法和介数法的不足综合考虑了节点效率、节点度值和相邻接点重要度贡献,提出了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<sup>[22]</sup>.通过实验证明,这些综合利用度量指标的方法取得的结果均优于单个指标获得的关键节点排序结果,但是这些排序方法也没有融入具有结构洞特征的因素,因而无法考虑处于结构洞位置的关键节点.

对于复杂网络中面向结构洞的关键节点排序问题的研究目前不多,一方面是由于处于结构洞位置的节点拥有独特特征,这些特征将结构洞节点与

复杂网络中的核心节点区分开,因此即使处于结构洞的关键节点在网络中发挥重要作用却依然容易被忽略;另一方面,在复杂网络中关于结构洞的研究还不深入,如何在复杂网络中科学的测量结构洞节点还值得探索. Burt提出的几个定量描述结构洞的指标<sup>[23]</sup>,如网络约束系数(Constraint, CT)、网络有效规模(EffectiveSize, ES)、效率(Efficient, EF)、等级度(Hierarchy, HI)等;Freeman提出用介数中心性(Betweenness Centrality, BT)测量方法描述结构洞<sup>[24]</sup>,这些指标能够定量的刻画结构洞的一些局部特征.根据相关结构洞指标,本文提出一种能够融合结构洞特征与网络节点其他重要性特征的排序方法,从而更加全面的对网络中的关键节点进行综合评价.

排序学习方法<sup>[25]</sup>是当前文本检索领域的研究热点,它将机器学习方法引入到信息检索的文档相关性排序问题中,充分考虑各种排序方法对最终排序结果的影响,通过训练学习排序参数,将各种排序方法视为特征,对文档的相关性做综合的评估.本文将该方法引入到复杂网络中面向结构洞节点的排序问题上,对不同的度量指标进行排序学习,最后获得参数.

排序学习方法能够有效结合多个特征的关键在于该方法能够学习多维特征的参数估计问题.而排序学习方法根据训练样本的不同可以分为三类:基于单个样本的Pointwise算法、基于样本对的Pairwise算法和基于样本队列的Listwise算法.Pointwise算法将每个训练数据集作为学习样本,将分类问题转化为单个文档的分类和回归问题;Pairwise算法将训练数据集中有不同相关性标注的一对数据作为样本,根据这些样本对将排序问题转化为二值分类问题,如应用比较广泛的Ranking SVM<sup>[26]</sup>算法;Listwise算法将训练数据的相关性序列作为样本,该样本序列与标注序列的距离作为损失函数进行学习,这三类方法中Listwise算法的实际效果最好,因此这种方法是排序学习领域当前的研究热点,而其中应用效果较好的算法是ListNet算法<sup>[27]</sup>.所以本文将ListNet排序学习方法引入复杂网络多指标关键节点排序问题中.

综上所述,本文将基于ListNet排序学习方法对网络中单个节点的CT, EF, ES, HT, BC, Pagerank值(PageRank, PR)、聚类系数(Clustering Coefficient, CC)这7个度量指标进行融合,综合评价复杂网络中面向结构洞的节点的重要性,从而对网

络中的关键节点进行排序.

## 2 基于 ListNet 排序学习的面向结构洞的关键节点排序方法

将 CT, ES, EF, HI, BT, PR, CC 7 个度量指标作为学习特征, 基于 ListNet 排序学习方法根据

训练数据中的指标值学习得到参数, 参数再经过交叉验证得到最终的参数值, 从而构建面向结构洞的关键节点排序学习方法. 实际应用过程中, 我们首先构造适合的学习样本网络, 然后学习参数, 对于测试排序网络, 则采用相同的参数, 直接排序, 得到关键节点排序结果, 基本流程如图 2 所示. 由于本文方法只需要学习一组参数, 对于待排序的网络没有任何的预处理和学习, 具有很高的实用性.

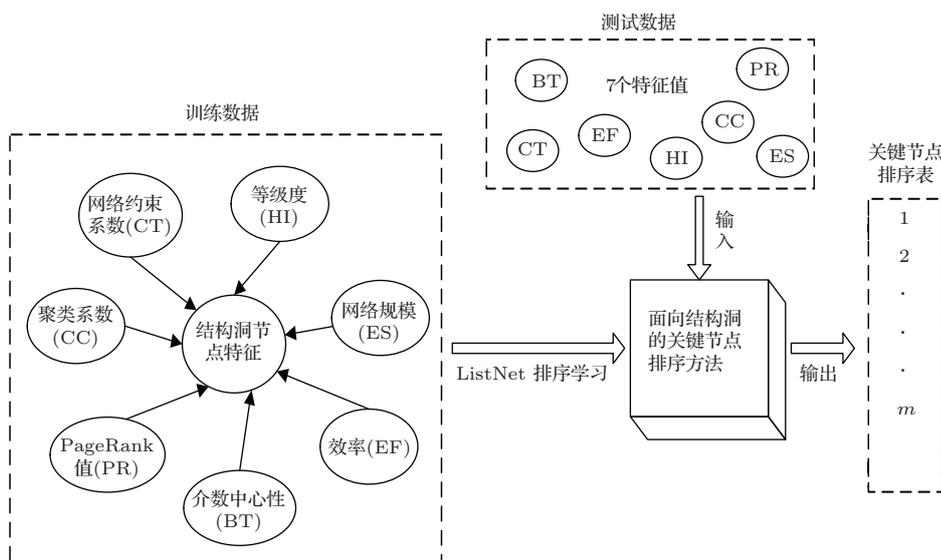


图 2 面向结构洞的关键节点排序方法的基本流程图

设一个复杂网络中有  $m$  个待排序的节点, 表示为  $M = \{1, 2, \dots, m\}$ , 排序标准分别为上述 7 个度量指标. 我们采用线性神经网络模型融合这 7 个度量指标, 用评分函数作为节点排序的准则, 评分函数用  $h_{\theta}(x) = \langle \theta, x \rangle$  表示, 其中向量  $\theta$  为需要学习的参数, 向量  $x = \{x_1, x_2, \dots, x_7\}$  表示一个节点的 7 个指标值, 则  $\theta = \{\theta_1, \theta_2, \dots, \theta_7\}$  对应为该节点各个指标值的模型参数, 用  $X = \{x_1, x_2, \dots, x_m\}$ , 表示网络中所有节点的指标值的集合, 则参数集合  $\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}$ , 而  $\Theta' = \{\theta'_1, \theta'_2, \dots, \theta'_m\}$  则表示学习后所得的参数集合, 线性神经网络模型学习参数的过程如图 3 所示.

上述神经网络模型学习的参数来自损失函数, ListNet 排序学习引入 Luce [28] 模型度量网络中节点排序的概率, 根据排序概率构造损失函数. Luce 模型是 Luce 提出的一个高度抽象化的理论化模型, 奠定 ListNet 排序学习的理论基础. 这种模型能够将排序对象的任意一种排序表示成一个排序概率, 假设排序对象用  $N = \{1, 2, \dots, n\}$  表示, 排序对象通过某种排序方式得到的分数集合用  $S = \{s_1, s_2,$

$\dots, s_n\}$  表示,  $\phi(\cdot)$  为严格递增的正函数, Luce 模型将排序概率定义为

$$P_s = \prod_{j=1}^q \frac{\phi(s(j))}{\sum_{t=k}^n \phi(s(t))}. \quad (1)$$

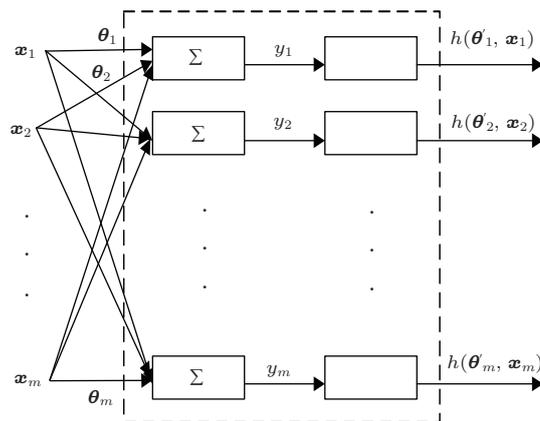


图 3 线性神经网络模型

依据 Luce 模型, 对于结构洞节点的排序问题, 排序对象为网络中的节点用  $N = \{1, 2, \dots, n\}$  表

示, 排列方式为上述的7种度量指标, 各节点的指标值用  $X = \{x_1, x_2, \dots, x_n\}$  表示, 根据指标值得到的评分函数用  $h(x_i)$  表示, 那么根据 Luce 模型, 评分函数的排列概率可以表示为

$$P_{h(x)} = \prod_{j=1}^q \frac{\exp(h(x_j))}{\sum_{t=k}^n \exp(h(x_t))}, \quad (2)$$

其中  $n$  表示网络中节点的总数,  $j$  和  $t$  均表示节点的序列数,  $q$  表示选用节点序列数. 结合7个指标得到网络中每个节点的排列概率, 采用交叉熵构造出损失函数, 表示为

$$L(y, h_\theta(x)) = - \sum_{i=1}^n P_y \times \log(P_{h_\theta(x)}), \quad (3)$$

其中  $y$  为网络中面向结构洞关键节点排序的相关性判断条件,  $n$  为训练数据中网络的数量.

通过梯度下降算法学习模型中的参数, 计算损失函数对  $\theta$  的梯度, 结果为

$$\Delta\theta = \frac{\partial L(y, h_\theta(x))}{\partial \theta} = - \sum_{i=1}^n \frac{\partial P_{h_\theta(x)}}{\partial \theta} \frac{P_y}{P_{h_\theta(x)}}. \quad (4)$$

$L(y, h_\theta(x))$  的优势在于该函数是连续并且是凸函数, 其误差损失在0—1之间, 但是计算量又成为  $L(y, h_\theta(x))$  的致命缺点, 而前  $k(k \ll q)$  项排列

概率<sup>[25]</sup>就可以优化该函数. 虽然会牺牲一定的准确度, 但是有效的节省了运算时间. 将排列概率中的  $q$  值设为1, 则可以得到

$$\Delta\theta = - \sum_{i=1}^n \frac{\exp(y_1)}{\sum_{j=1}^m \exp(y_j)} \times \left\{ x_1 - \frac{\sum_{j=1}^m [\exp(h_\theta(x_j)) \times x_j]}{\sum_{j=1}^m \exp(h_\theta(x_j))} \right\}. \quad (5)$$

采用 Luce 模型和交叉熵作为线性神经网络模型的损失函数, 我们可以得到基于 ListNet 的网络关键节点排序算法, 算法如表1所示.

算法输入训练网络每个节点的7个指标值  $x$  以及人工标注打分  $y$ 、迭代次数  $T$  和学习步长  $\eta$ ; 输出学习到的神经网络模型的参数  $\theta$ . 基本过程是首先随机给定参数的初始值, 利用评分函数得出网络每个节点的得分(算法中步骤3—5), 再代入到线性神经网络模型中, 利用梯度下降法计算参数的梯度  $\Delta\theta$ (算法中步骤6), 然后更新参数  $\theta' = \theta - \eta \cdot \Delta\theta$ (算法中步骤7), 利用迭代次数  $T$  控制循环, 直到所得模型的参数值达到稳定状态.

表1 基于 ListNet 的网络关键节点排序算法

Input:	Training Data $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$ , Number of Iterations $T$ , Learning Rate $\eta$
Output:	$\theta$
1:	for $t = 1$ to $T$ do:
2:	for $i = 1$ to $m$ do:
3:	Input $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$ and compute $h_\theta(x_i)$ for $x_t$ with current $\theta$
4:	Compute $P_{h(x)}$ with $h_\theta(x_i)$ in the network
5:	Get $P_{h(x)}$ into neural network and compute $L(y, h_\theta(x))$ with $y_i$
6:	Compute $\Delta\theta$ using Gradient Descent Algorithm
7:	$\theta' = \theta - \eta \cdot \Delta\theta$ ; /*Update $\theta$ */
8:	end for
9:	end for

### 3 实验分析

#### 3.1 实验设计

我们人工构建了21个模拟网络, 选择了4个代表性真实复杂网络作为实验网络进行实验, 其中20个模拟网络作为训练网络进行参数学习, 剩下的模拟网络和4个真实网络作为测试网络, 然后根据测

试网络得到的结果进行评价和分析. 我们人工排序了部分网络中的关键节点, 然后进行了具有人工标注结果下的准确排序实验. 对于难以人工排序的较大网络, 则利用 SIR 模型评价关键节点的传播能力.

##### 3.1.1 人工标注实验网络

实验选取了20个模拟网络作为训练网络, 这些模拟网络参考真实网络中的一些结构进行构造, 每个网络中有30到40个节点; 将一个模拟网络

和 Coauthor 网络中抽取的一个子集网络作为测试网络, 这两个测试网络的规模分别为 37 个节点和 50 个节点, 由于这些网络规模较小, 本文对这些网络中的节点基于结构洞理论进行人工标注和排序打分, 并将它们作为本文方法中的相关性判断条件. 为了评估实验效果, 人工标注的网络采用  $NDCG@n$  指标作为评价标准, 该指标用来评价不同排序方法得到排序结果中前  $n$  个节点和人工排序关键节点的准确性.

### 3.1.2 真实网络

为了验证实现效果的真实性, 本文选取了节点较多 3 个真实网络作为测试网络, 每个网络的统计特性如表 2 所示.

表 2 实验网络

网络	节点数量	边数量
USA Air Lines	332	2461
Coauthor	462	975
Power Grid	4940	6594

USA Air Line 网络中的节点代表着机场, 边代表航空线路, 在这个网络拓扑图中研究面向结构洞的关键节点, 可以找到对交通的通畅性影响较大的机场, 将这些机场作为重点维护对象, 以避免交通的不便.

Coauthor 网络是 C-DBLP 科研合作网中提取的一个连通子集, 这个网络中的节点表示的是科研人员, 边代表的是科研人员之间的合作关系, 用该网络对面向结构洞的关键节点进行研究可以挖掘出在课题组的合作中起到重要联系的科研人员.

Power Grid 网络是美国西部电力网网络, 1996 年美国西部电网连续两次发生大面积停电事故导致了巨额的经济损失, 该事故引起了人们对网络中关键节点的重视, 所以该网络对于本文的研究十分有意义.

由于在这些网络中节点数量较大, 无法通过人工标注进行评价, 因此本文采用 SIR 模型来评价真实网络中各种方法所得排序结果的优劣. 将不同排序方法得到的有序关键节点作为 SIR 模型的初始感染节点, 取相同数目的 TOP-K 节点设为初始感染节点, 用 SIR 模型在这些网络上进行实验, 若排序后的结果使得网络传播得更广更快, 则说明该排序效果更佳.

## 3.2 人工标注实验网络实验结果分析

对于测试网络中模拟网络和 Coauthor 子集网络的实验结果, 本文分别采用  $NDCG@2$ ,  $NDCG@3$ ,  $NDCG@5$  和  $NDCG@8$  评价方法进行评价. 图 4 和图 5 分别是模拟网络和 Coauthor 子集网络的  $NDCG$  对比图, 分别给出本文方法与其他 7 种指标对网络中节点进行排序的  $NDCG@n$  柱状对比, 表 3 则是模拟网络和 Coauthor 子集网络中本文方法与介数中心性计算出的 TOP-8 节点编号. 从图 4 和图 5 中可以看出, 本文方法排序结果在大多数情况下是优于其他几个度量指标. 图 5 中本文方法的排序结果前 3 个节点与介数中心性的结果一致, 但从表 3 中可以看出两种方法所得前 3 个节点并不完全相同, 这是因为  $NDCG$  计算过程中节点 34 和节点 14 的在两种方法中排序得分加和正好相等; 而图 5 中本文方法  $NDCG@5$  和  $NDCG@8$  结果略优于介数中心性的排序结果, 与表 2 中节点排序结果对应. 所以, 整体分析本文方法对面向结构洞关键节点排序具有较高有效性.

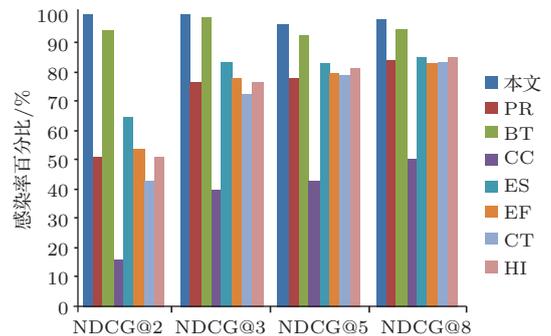


图 4 (网刊彩色) 模拟网络  $NDCG@n$  柱状对比图

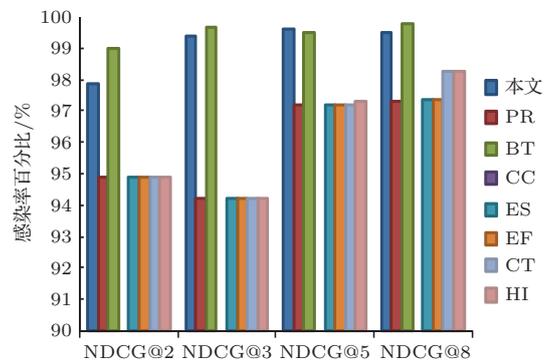


图 5 (网刊彩色) Coauthor 子集网络  $NDCG@n$  柱状对比图

图 6、图 7 分别给出了模拟网络和 Coauthor 子集网络的拓扑结构图, 根据基于模块度的 Force At-

las 算法所得的社团用不同的颜色表示, 图中较大的节点为本文方法所得排序结果分别选取的前5个和前8个关键节点, 与表3中本文方法排序结果节点编号对应, 结合图6和图7可以看出, 本文排序最前的节点很好地覆盖了网络中的关键节点, 如图6中连接三个社团的节点49以及图7中处于网络中心位置的节点20等等. 而在这些关键节点中, 图6中节点13, 29和图7中节点1, 20, 37处于社团的中心位置, 图7中节点7, 24, 38处于社团边缘位置, 图6节点49和图7中节点14, 34处于结构洞位置, 说明本文方法不仅挖掘出处于核心位置的关键节点, 同时也能识别出处于结构洞位置的关键节点, 证实本文方法充分融合了各个指标的特点, 从而能够更加全面的描述网络中的关键节点. 在表3的模拟网络中节点49在本文方法下排名第一, 而在介数中性指标计算结果中排在第3位, 这从另一个角度说明本文方法能够更准确的识别出面向结构洞的关键节点.

从表3可以看出, 在 Coauthor 子集网络中本文方法没有将具有结构洞特征的节点14优先识别出来, 因而在图5中展示的本文方法的效果与介数中心性效果差别不大, 然而在模拟网络中本文方法却可以优先识别出处于结构洞位置的节点49从而达到挖掘面向结构洞的关键节点的目的. 而根据图6和图7可以观察到, 在模拟网络中节点49连接着三个拓扑结构不同的社团, 该节点为典型的处于结构洞位置的节点; 而在 Coauthor 子集网络中

处于结构洞位置的节点特征不是很明显, 如节点14和节点34, 同时这些节点连接的社团都均是星型结构, 网络中社团的拓扑结构比较单一, 而本文方法的特点是将网络拓扑结构与节点功能性特征相结合来识别网络中的关键节点, 所以在 Coauthor 子集网络中不能显示出本文方法的特殊性.

表3 TOP-8 节点编号表

排名	模拟网络			Coauthor 子集网络		
	人工标注	本文	BT	人工标注	本文	BT
1	49	49	13	14	20	20
2	13	13	29	34	34	14
3	29	29	49	20	14	34
4	19	26	19	24	24	24
5	26	31	26	37	37	7
6	40	32	31	7	7	37
7	31	33	33	1	1	1
8	32	19	32	38	38	6

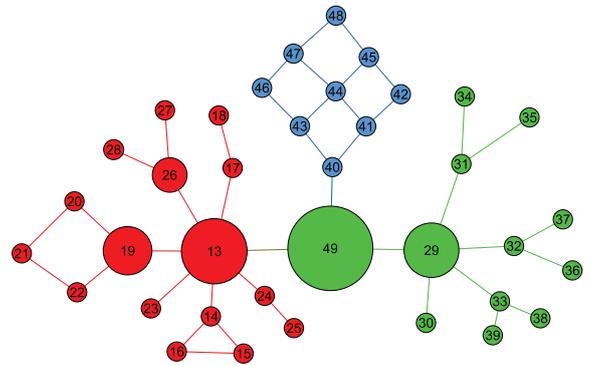


图6 模拟网络 top-5 的节点

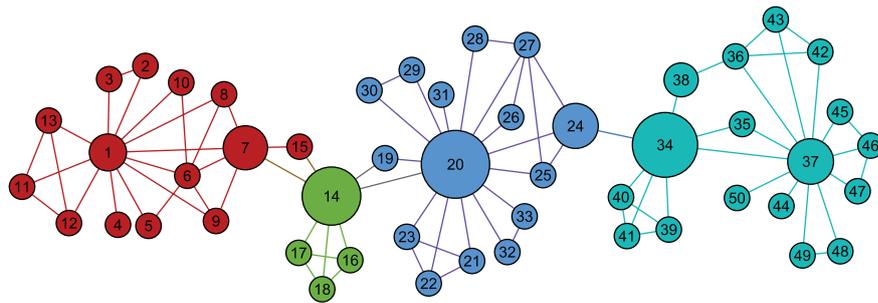


图7 Coauthor 子集网络 top-8 的节点

### 3.3 非人工标注实验网络实验结果分析

在本文方法以及7个度量指标得出的排序结果中, 分别取美国航空网 (USA Air Lines) 和科研合作网 (Coauthor) 的前10个节点、美国西部电力网 (Power Grid) 中的前30个节点, 将各个度量指标以及本文方法所得的这些节点作对比, 选择不同指标中不重叠的节点作为初始感染源进行 SIR 传

播分析, SIR 模型采用相同的感染和恢复系数进行实验. 选择不同的初始感染节点, SIR 模型所能达到的最大感染比例与所需时间不同. 感染比例越高、所需时间越短, 说明初始节点在传播过程中的作用越大. 不同网络在不同方法选择节点作为初始感染节点时, SIR 模型得到最大感染节点数的感染率和对应的感染时间如表4所示.

表4 SIR模型结果表

方法	AirLine(332)		Coauthor(462)		PowerGrid(4940)	
	感染率	时间	感染率	时间	感染率	时间
本文	1	5	0.88	5	0.62	11
CT	1	5	0.86	5	0.57	10
本文	1	4	0.75	8	0.51	15
BT	1	5	0.65	6	0.37	15
本文	1	5	0.91	5	0.62	11
PR	0.93	5	0.90	6	0.57	8
本文	1	5	0.91	5	0.62	11
ES	0.91	5	0.89	6	0.55	10
本文	1	5	0.91	5	0.62	11
EF	0.91	5	0.89	6	0.55	10
本文	1	5	0.82	7	0.62	11
HI	0.92	5	0.73	7	0.51	10
本文	1	5	0.87	5	0.62	11
CC	0.93	6	0.87	7	0.51	10

从表4可以看出,在美国航空网(USA Air Lines)网络中,本文方法识别出的关键节点感染了网络中所有节点,而对比的度量指标中只有CT和BT识别出的节点有类似效果,但是这两个度量指标对应的感染时间均大于本文,说明本文方法得出的排名靠前的节点与各度量指标相比,能在更短时间内感染相同数量的节点.在科研合作网(Coauthor)和美国西部电力网(Power Grid)这两个实验网络中,本文方法所得排序结果感染的节点比例均比其他度量指标所得排序结果感染的节点比例大,在Power Grid网络中,由于采用本文方法得到的关键节点作为感染源感染节点的数目更多,所以才使得所需传播时间略高于其他方法;另一方面,由于本文方法所识别的关键节点是面向结构洞的,一些关键节点处于网络的边缘,所以感染时间比一些直接从中心位置的节点发起感染所需的时间较长.与此同时,从实验数据分析,本文方法与其他方法相比感染时间差异并不显著,并且面向

结构洞的关键节点的优势在于能够感染到更多的节点.

为了更加深入分析不同方法得到的关键节点排序结果的差异,我们对Power Grid网络上取出不同方法排序得到的TOP-30节点,按照排序顺序依次逐步增加作为初始感染源,采用相同参数的SIR模型试验不同方法排序结果作为初始感染节点所能感染的最大比例与所需时间,以此来评价不同方法排序结果的性能差异,结果如图8和图9所示.图8为感染范围对比图,图9为对应的感染时间的对比图,其中图8和图9的横坐标均表示感染源节点依照排序次序逐步加入初始感染节点的数量,图8的纵坐标表示感染节点占总节点数目的百分比,图9中的纵坐标表示感染时间.为了便于观察,图8和图9均分为(a), (b)两张图,图8和图9中的(a)的将本文方法(RK)与BT, CT以及PR所得结果做对比,而(b)则是将RK与CC, ES, EF以及HI做对比.从图8中可以看出,本文方法感染节点数目的百分比从总体趋势上多于其他方法,对比图8(a), (b)两图, RK的排序效果较优于CT, CC, ES, EF以及HI.而与BT, PR相比,在初始节点增加时更为接近,究其原因是因为RK与BT, PR这两个度量指标找到的某些关键节点部分相同,从而使得感染节点的能力相同.但是结合图8和图9可以观察到,感染源为前7个节点的时候,虽然RK与BT, PR所感染节点数目的百分比接近,但所需感染时间最短,说明与其他方法相比RK将感染力更强的节点排在了前面.当初始感染节点数量逐渐增加,感染的时间越来越短,当感染节点多余20个后,各方法的感染时间均趋于稳定,而在这种情况下,本文方法排序结果所得感染源能够感染到整个网络60%的节点,进一步说明了本文方法能够在真实网络上选择出影响传播的关键节点并能够较好排序.

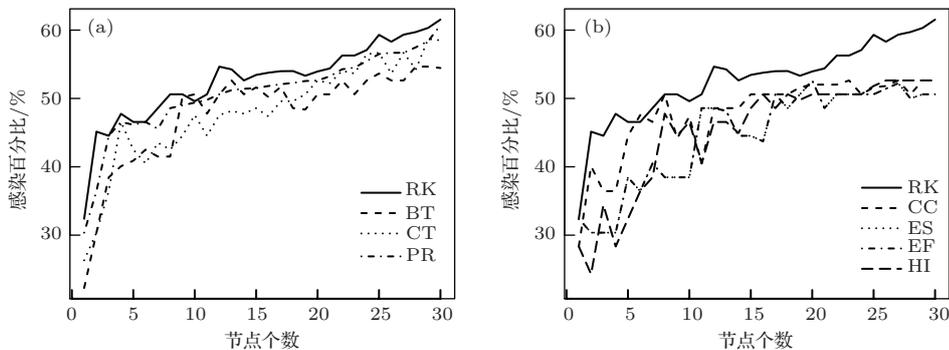


图8 感染范围对比图

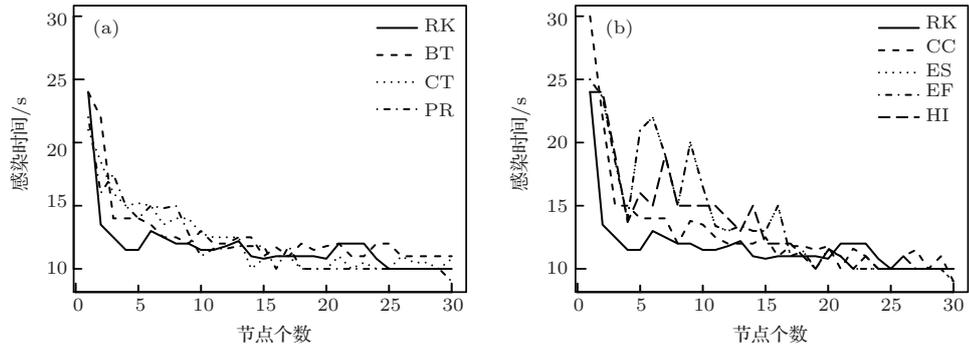


图9 感染时间对比图

## 4 结 论

本文基于ListNet的排序学习方法融合CT, ES, EF, HI, BT, PR, CC 7个度量指标, 构造出面向结构洞的关键节点排序方法, 通过美国航空网(USA Air Lines)、科研合作网(Coauthor)和美国西部电力网(Power Grid)三种实际网络进行实验验证, 利用SIR模型评价本文所构造方法的有效性. 最后, 本文主要得出结论有:

1) 根据具有人工排序的实验结果分析, 综合7个度量指标评价复杂网络中的关键节点比单一指标评价效果更好, 验证了相关研究阐述的集成性排序方法优于单个指标排序的结论;

2) 对于大型复杂网络, 本文方法排序得到的TOP-K节点能够在较短的时间内达到最大的传播范围, 在网络拓扑结构具有多样性的情况下效果更佳, 说明本文方法选择出的节点具有较高的传播能力, 对于现实复杂网络中的实用性比较强.

由于介数中心性计算复杂度高, 下一步工作中, 我们将优化本文方法, 探索利用简单指标实现有效排序, 将本文方法运用在更大型的复杂网络关键节点排序上.

## 参考文献

- [1] Shen Y 2013 *Chin. Phys. B* **22** 058903
- [2] Liang Z W, Li J P, Yang F, Athina P 2014 *Chin. Phys. B* **23** 098902
- [3] Shen Y 2011 *Chin. Phys. B* **20** 040511
- [4] Lü L Y, Lu J A, Zhang Z K, Yan X Y, Wu Y, Shi D H, Zhou H P, Fang J Q, Zhou T 2010 *Complex System and Complex Science* **7** 173 (in Chinese) [吕琳媛, 陆君安, 张子柯, 闫小勇, 吴晔, 史定华, 周海平, 方锦清, 周涛 2010 复杂系统与复杂性科学 **7** 173]
- [5] Burt R S 1992 *Networks and Organizations: Structure, Form, and Action* **65** 57
- [6] Bonacich P 1972 *J. Math. Sociol* **2** 113
- [7] Freeman L C 1979 *Soc. Netw.* **1** 215
- [8] Freeman L C, Borgatti S P, White D R 1991 *Soc. Netw.* **13** 141
- [9] Katz L 1953 *Psychometrika* **18** 39
- [10] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nature Physics* **11** 22
- [11] Arasu A, Cho J, Garcia-Molina H 2001 *ACM Transactions on Internet Technology* **1** 2
- [12] Lü L, Zhang Y C, Yeung C H 2011 *PLoS One* **6** 21202
- [13] Kleinberg J M 1999 *JACM* **46** 604
- [14] Li P X, Ren Y Q, Xi Y M 2004 *System Engineering* **22** 13 (in Chinese) [李鹏翔, 任玉晴, 席酉民 2004 系统工程 **22** 13]
- [15] Chen Y, Hu A Q, Hu X 2004 *Journal of China Institute on Communications* **25** 129 (in Chinese) [陈勇, 胡爱群, 胡啸 2004 通信学报 **25** 129]
- [16] Tan Y J, Wu J, Deng H Z 2006 *System Engineering-Theory and Practice* **788** 79 (in Chinese) [谭跃进, 吴俊, 邓宏钟 2006 系统工程理论与实践 **788** 79]
- [17] Zhang Y, Liu Y H, Xu K H, Luo Z R 2011 *Computer Science* **38** 88 (in Chinese) [张翼, 刘玉华, 许凯华, 骆珍荣 2011 计算机科学 **38** 88]
- [18] Yu H, Liu Z, Li Y J 2013 *Acta Phys. Sin.* **62** 020204 (in Chinese) [于会, 刘尊, 李勇军 2013 物理学报 **62** 020204]
- [19] Hou B N, Yao Y P, Liao D S 2012 *Physica A* **391** 4021
- [20] Comin C H, Costa L D 2011 *Phys. Rev. E* **84** 056105
- [21] Ren Z M, Shao F, Liu J G 2013 *Acta Phys. Sin.* **62** 128901 (in Chinese) [任卓明, 邵凤, 刘建国 2013 物理学报 **62** 128901]
- [22] Zhou X, Zhang F M, Li K W, Hui X B, Wu H S 2012 *Acta Phys. Sin.* **61** 050201 (in Chinese) [周漩, 张凤鸣, 李克武, 惠晓滨, 吴虎胜 2012 物理学报 **61** 050201]
- [23] Burt R S 2004 *American Journal of Sociology* **110** 349
- [24] Freeman L C 1977 *Sociometry* **40** 35
- [25] Liu T Y 2009 *Foundations and Trends in Information Retrieval* **3** 225
- [26] Joachims T 2002 *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Edmonton, Canada, July 23–26, 2002 p133
- [27] Cao Z, Qin T, Liu T Y, Tsai M F, Hang L 2007 *Proceedings of the 24th Annual international conference on Machine learning*. Corvallis, June 20–24, 2007 p129
- [28] Luce R D 1959 *Individual Choice Behavior: A Theoretical Analysis* (New York: Wiley) pp147–177

# Ranking key nodes in complex networks by considering structural holes\*

Han Zhong-Ming<sup>†</sup> Wu Yang Tan Xu-Sheng Duan Da-Gao Yang Wei-Jie

(Beijing Technology and Business University, Beijing 100048, China)

( Received 13 July 2014; revised manuscript received 30 September 2014 )

## Abstract

Structural hole nodes in complex networks play important roles in the network information diffusion. Unfortunately, most of the existing methods of ranking key nodes do not integrate structural hole nodes and other key nodes. According to the relevant research on structural hole theory as well as the key node ranking methods, network constraint coefficient, betweenness centrality, hierarchy, efficiency, network size, PageRank and clustering coefficient, 7 metrics are selected to rank the key nodes. Based on the 7 metrics, a ranking learning method based on ListNet is introduced to solve ranking key nodes by multi metrics. Comprehensive experiments are conducted based on different artificial networks and real complex networks. Experimental results with manual annotation show that the ranking method can comprehensively consider the structural hole nodes and other nodes with different important features. The ranking results on different networks are highly consistent with the manual ranking results. The spreading experiment results using signed to interference ratio propagation model show that SIR model can reach a maximum propagating ratio in a shorter propagating time initiated by TOP-K key nodes selected by our method than TOP-K key nodes selected by other methods.

**Keywords:** structural hole, key node, learning ranking, complex network

**PACS:** 89.75.Hc

**DOI:** [10.7498/aps.64.058902](https://doi.org/10.7498/aps.64.058902)

---

\* Project supported by the National Natural Science Foundation of China (Grant No.61170112), the central finance special fund to support the development of local colleges and Universities, China (Grant No. 19005323132), and the of Humanities and Social Science Research Fund Ministry of Education, China(Grant Nos 13YJC860006).

<sup>†</sup> Corresponding author. E-mail: [hanzm@th.btbu.edu.cn](mailto:hanzm@th.btbu.edu.cn)