一种基于节点影响力的信息传播概率算法

张 永,和 凯
ZHANG Yong, HE Kai

兰州理工大学 计算机与通信学院,兰州 730050
College of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

ZHANG Yong, HE Kai. Algorithm for information propagation probability based on influence of origin. Computer Engineering and Applications, 2018, 54(10):126-134.

Abstract: In the information propagation research of social network, it is the most common way to set up a propagation probability and simulate the information propagation by using various models. However, the appointed propagation probability has a great influence on the propagation. Based on the idea of finding important nodes in complex networks, this paper presents a method to calculate the probability of propagation. Experiment analyzes different results caused by the fixed propagation probability and the probability with influence of origin, and shows the calculated probability more satisfies the truth by validating the algorithm of node's influence.
Key words: social network; information propagation probability; influence

摘 要:在社交网络上的信息传播的研究中,设定一个传播概率,运用各种传播模型来模拟信息传播过程,是最常见的一种方式,然而人为设定的传播概率对传播过程有很大影响。根据复杂网络的相关研究,计算信息源节点的影响力,并以此为基础提出了一种计算信息传播概率的方法。实验对比了人为设定的传播概率与考虑了信息源节点影响力的传播概率对传播结果造成的差异,并通过证明影响力算法的有效性,说明了计算后的传播概率更加合理。
关键词:社交网络;信息传播概率;影响力
文献标志码:A **中图分类号:**TP393 doi:10.3778/j.issn.1002-8331.1612-0414

1 引言

社交网络服务SNS(Social Network Service)近年来发展迅速,其凭借网络的强大连通力将人们的社交范围从现实的人际关系扩展到虚拟的网络中来。通过即时聊天工具、微博、博客、网络社区等网络应用将人们的社交范围逐步扩大,最终形成一个人与人关联的巨大的复杂网络。Facebook是目前世界上最大的在线社交网络,目前已拥有超过22亿的总用户,并据Facebook预测到2030年用户总数将会达50亿人。社交网络不但具有互联网络的物理特性,还包含了人际关系的社交特性,是一个典型的复杂网络,其规模及影响范围正在不断扩大。

针对社交网络上的信息传播问题已有一定的研究成果,如谣言传播问题[1-3]、社交网络中的信息转发预测[4-6]、信息传播的模型研究[7-11]等,其中用户影响力问题[12-20]一直是一个热点。由于社交网络的复杂特性,影响信息传播的因素非常多,要提取所有对信息传播有影响的特征不现实,过多的特征会使得模型复杂度过高。在上述研究问题中,无论是谣言传播与转发预测,还是构造信息传播模型,信息源节点的权威性这一特征会对信息的传播结果有重要的影响。因此量化信息源节点的权威性,对精确描述信息传播的过程有重要意义。其中文献[13]利用网络拓扑寻找重要节点,文献[19]则研究了在有社区结构的网络中如何寻找重要节点。

本文主要解决的问题是如何通过衡量信息源节点的影响力来确定一条信息的初始传播概率。本文的研究重点是在实际传播开始之前给出一个明确的传播概率,取代以往研究中根据经验设定的固定值,而不考虑在传播过程中,由于舆论导向,人际关系的相互影响等因素而引起的动态传播状态变化。本文参考了基于随

作者简介:张永(1963—),男,教授,研究领域为智能信息处理,图像处理,E-mail:yzhang@lut.edu;和凯(1991—),男,硕士研究生,研究领域为智能信息处理,社会计算。
收稿日期:2016-12-26 **修回日期:**2017-03-20 **文章编号:**1002-8331(2018)10-0126-09
CNKI**网络出版:**2017-09-14, http://kns.cnki.net/kcms/detail/11.2127.TP.20170914.1519.004.html

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

机游走的知识图谱问题的解决方案[21-22],提出一种基于节点影响力的初始传播概率计算方法。实验以SIR传染病模型[23]为信息传播基础模型,首先证明节点影响力对传播结果有重要影响,其次证明了基于影响力的算法的有效性,最后对比了基于节点影响力的信息传播概率与固定传播概率在传播过程中的差异。信息传播概率可以为预测信息传播规模、分析信息传播特点、挖掘舆论导向等问题提供一定的依据。

2　基于节点影响力的信息传播概率

2.1　节点影响力问题

社交网络上影响力的研究是社交计算的重要内容,找出有影响力的节点在社会舆论导向、商业营销、谣言识别以及专家发现等问题上都有重要意义。目前对于如何确定一个社交网络用户的影响力有很多的研究,其方法大致可以归结为两类:基于网络拓扑的方法与基于用户行为的方法。其中基于网络拓扑的影响力算法相比基于用户行为的算法更加简单且复杂度低,常用的有基于节点度(Degree Centrality)、基于最短路径的介数中心度(Betweenness Centrality)、紧密中心度(Closeness Centrality)、基于随机游走的特征向量中心度等算法。

确定节点的影响力问题,类似于PageRank算法对网页排名的问题,需要对每一个节点确定影响力。针对大规模社交网络而言,传统的节点影响力度量指标效果均不理想。比如用度中心度来衡量节点影响力的效果很差,而介数中心度与紧密中心度虽然效果较好,但是时间复杂度高达 $O(n^3)$,性能无法接受。

节点 u_i 的度中心度以 $C_k(u_i)$ 表示:

$$C_k(u_i)=degree(u_i) \tag{1}$$

节点 u_i 的介数中心度 C_B 用以衡量网络中包含节点 u_i 的任意两个节点间的最短路径的条数,占所有最短路径条数的比例大小。它可以较好地描述节点 u_i 在网络中的中心性,即对其他节点的影响力大小,以 $C_B(u_i)$ 表示:

$$C_B(u_i)=\sum_{s\neq u_i\neq t\in v}\frac{\sigma_{st}(u_i)}{\sigma_{st}} \tag{2}$$

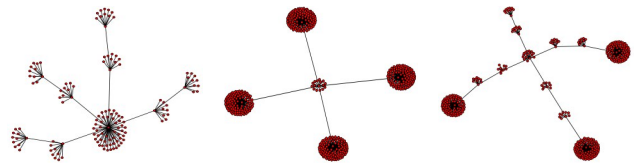
其中,σ_{st} 是节点 s 与节点 t 间所有最短路径的条数,而 $\sigma_{st}(v)$ 则是包含节点 u_i 的 s 与 t 间最短路径的条数。

节点 u_i 的紧密中心度 C_C 用以衡量节点 u_i 到网络中其他节点的距离之和,即如果节点 u_i 发出一条信息,需要多久能传播到所有能够到达的节点。

$$C_C(u_i)=\frac{1}{\sum_{t\in U/u_i}d_G(u_i,t)} \tag{3}$$

其中,$d_G=(u_i,t)$ 是节点 v 到节点 t 的距离。

但是,在社交网络中,有一种如图1所示的常见现象。在图1(a)中,中心节点虽然本身具有很大影响力,但它的邻居节点却都是小影响力节点。而图1(b)中,中心节点虽然本身影响力小,但它有4个影响力大的邻居。这种情况会使得图1(b)的中心节点比图1(a)的中心节点拥有更大的影响力,而图1(b)中心节点的度却比图1(a)中心节点的度小。进一步的,本文发现如果有图1(c)这样的结构,在中心节点本身度很小的情况下,在它的附近有几个权威节点,信息可能需要经过一次传播,或经过较少的两次或三次的传播后,能到达这几个重要的节点。在这种情况下,如图1(c)这样的中心节点仍会拥有较大的影响力。图2为以类似图1(a)、图1(b)、图1(c)的中心节点为信息源节点,在SIR模型上的传播结果,横轴 t 为传播轮次,纵轴为最大感染率。在图2中可以看到度数最大的中心节点影响力却最小,这表明简单的节点影响力度量单位不能很好反映出潜在重要节点所带来的影响力变化。



(a)直接影响力较大　(b)实际影响力较大　(c)潜在影响力较大

图1　三种不同情况下的信息源节点

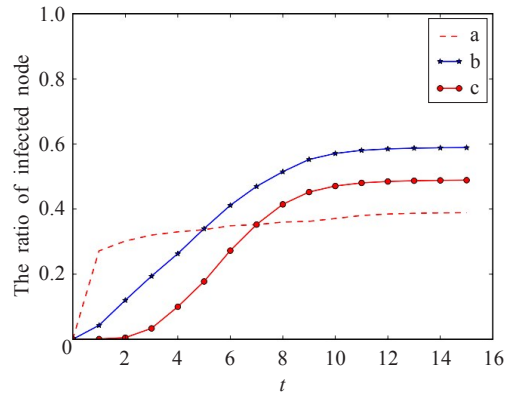


图2　三种不同情况下的信息源节点影响力比较

文献[7]中的方法考虑到了图1(b)中的这种情况,如果一个节点的邻居节点或邻居节点的邻居节点是影响力很大的节点,即沿着网络拓扑向外传播2层时,若这个信息源节点周围有重要节点与之相连,那么这个节点因此影响力也会比较大。但是考虑到图1(c)的情况,也许在一个节点向外传播3层或4层就会有许多重要节点,那么上述的各种方法便不能反映出这些重要节点对信息源节点影响力的作用。

2.2　基于节点影响力的传播概率算法

本文针对社交网络上的信息传播特点,用基于随机游走的节点影响力算法来确定信息传播概率。其主要思想为:对于一个社交网络 $G=(V,E)$,其中 V 为所有顶点的集合,E 为所有边集合,设置 N 个从信息源节点 v 出发随机游走器,游走长度为 L 的路径,计算每条路

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

径权重,最后将沿着信息源节点出发游走到的路径权重求和,得到信息源节点的影响力大小,归一化后设为信息传播概率。详细算法描述如下。

从所求节点 v 出发,设置随机游走器数量为 N ,游走路径长度为 L 。则随机游走器 N_i 的一次长度为 L 的随机游走所带来的权重为:

$$h_{N_i}(v)=\sum_{e'\in L_{N_i}} C(v') \tag{4}$$

其中,本文 N 的取值根据计算量要求可灵活设置, L_{N_i} 为随机游走器 N_i 在长度 L 的游走路径上的所有节点的集合。 $C_{v'}$ 为节点 v' 的权重值,计算公式如下:

$$C(v')=\sum_{u\in\Gamma(v')} Q(u) \tag{5}$$

其中, $\Gamma(v')$ 为节点 v' 的最近邻居节点集合, $Q(u)$ 的计算公式为:

$$Q(u)=\sum_{z\in\Gamma(u)} degree(z) \tag{6}$$

其中, $\Gamma(u)$ 为节点 u 的最近邻居节点集合, $degree$ 为节点 z 的度。然后将所有随机游走器带来的权值求平均,即得最终节点的影响力评分,表示为:

$$score(v)=\frac{\sum_{N_i\in N_c} h_{N_i}(v)}{N} \tag{7}$$

其中 N_c 为设置的 $degree(e)\times 3$ 个随机游走器的集合。最后利用 $Score$ 的最大与最小值的差将 $Score$ 值归一化为传播概率。

$$P(v)=\frac{score(v)-\min\limits_{v_0\in V}(score(v_0))}{\max\limits_{v_0\in V}(score(v_0))=\min\limits_{v_0\in V}(score(v_0))} \tag{8}$$

以图3为例解释上述影响力算法。节点1为信息源节点,假设有2个随机游走器,游走到了2条路径,分别是 $L_{N_1}=(1,3,6), L_{N_2}=(1,12,13)$ 。节点1的影响力评分 $Score(1)=h_{N_1}(1)+h_{N_2}(1)=78+43=121$ 。其中有 $h_{N_1}(1)=C(1)+C(3)+C(6)$, $C(3)=Q(1)+Q(6)$, $Q(3)=degree(1)+degree(6)$,其余同理可得。

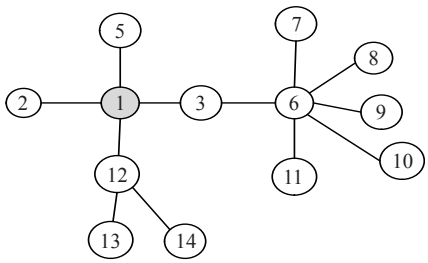


图3　随机游走示例图

3　实验仿真

3.1　数据描述

本文实验中使用四种社交网络:MSN Space(MS)、NetsScience(NS,一种科学家发表论文的合作关系网络)、Twitter(TW)、新浪微博(XL)。其中实验所用MS的数据取自 http://www.cs.bris.ac.uk/~steve/networks/peacockpaper;NS数据取自 http://www.personal.umich.edu/~mejn/netdata;TW数据取自 http://snap.stanford.edu/data/;XL数据取自 http://www.nlpir.org/?action-viewnews-itemid-299。假设这四种网络均为无向无权图,以 $G=(V,E)$ 表示,其中 V 表示网络节点的集合, E 表示链接 V 的边的集合。详细网络拓扑数据如表1所示,具体网络度分布如图4所示。

表1　网络参数

网络	节点数	边数	平均度	度相关系数	聚类系数	最大度
MS	3 982	6 803	3.416 9	−0.133 0	0.283 8	189
NS	4 381	4 608	2.103 6	−0.227 7	0.051 2	121
TW	5 835	80 381	27.551 3	0.046 1	0.504 4	364
XL	5 287	6 110	2.311 3	−0.398 9	0.027 9	2 000

首先可以看到图4所示的四种社交网络均很好地服从了幂率分布,符合典型社交网络的分布特点。其中Twitter的聚类系数相对较大,为同配性网络,MSN Space、NetsScience、Twitter与新浪微博为异配性网络。需要注意的是Twitter与新浪微博中存在权威节点的现象相对更加明显,可以看到大量节点分布在 10^3 量级附近,而MSN Space与NetsScience中权威节点则更多分布在 10^2 量级附近。在图4(d)中可以看到,在度2 000的位置有一个明显的尾部上升趋势,这是因为新浪微博本身限制最大关注2 000人,在节点度数达到2 000后,便不能再增长。

3.2　传播模型及传播规则

SIR传染病模型可将社交网络中信息的传播过程描述如下:某几个节点首先发出一条信息,成为初始信息源节点,这些处于传播信息状态的节点为感染状态(Infected,I),其余暂时未接触到这条信息的节点为易感染状态(Susceptible,S)。这些处于感染状态节点的直接邻居节点会接收到信息,并以一定的概率传播这条信息,由易感染状态变为感染状态,这个概率就是上述提到信息的初始传播概率。而处于感染状态节点不会一直处于感染状态,它们会在一定时间后,结束传播过程,转变为免疫状态(Recovered,R),不再传播该信息。SIR模型可用下列微分方程组描述:

$$\begin{cases}\dfrac{\mathrm{d}S(t)}{\mathrm{d}t}=-P_{\mathrm{SI}}(t)I(t)\\ \dfrac{\mathrm{d}I(t)}{\mathrm{d}t}=P_{\mathrm{SI}}S(t)I(t)-P_{\mathrm{IR}}I(t)\\ \dfrac{\mathrm{d}R(t)}{\mathrm{d}t}=P_{\mathrm{IR}}I(t)\end{cases} \tag{9}$$

实际上使用简单的SIR模型不能完全描述社交网络中各种复杂的节点状态,且模型在计算节点状态的改变概率时并没有考虑到社交网络中节点相互影响的重

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

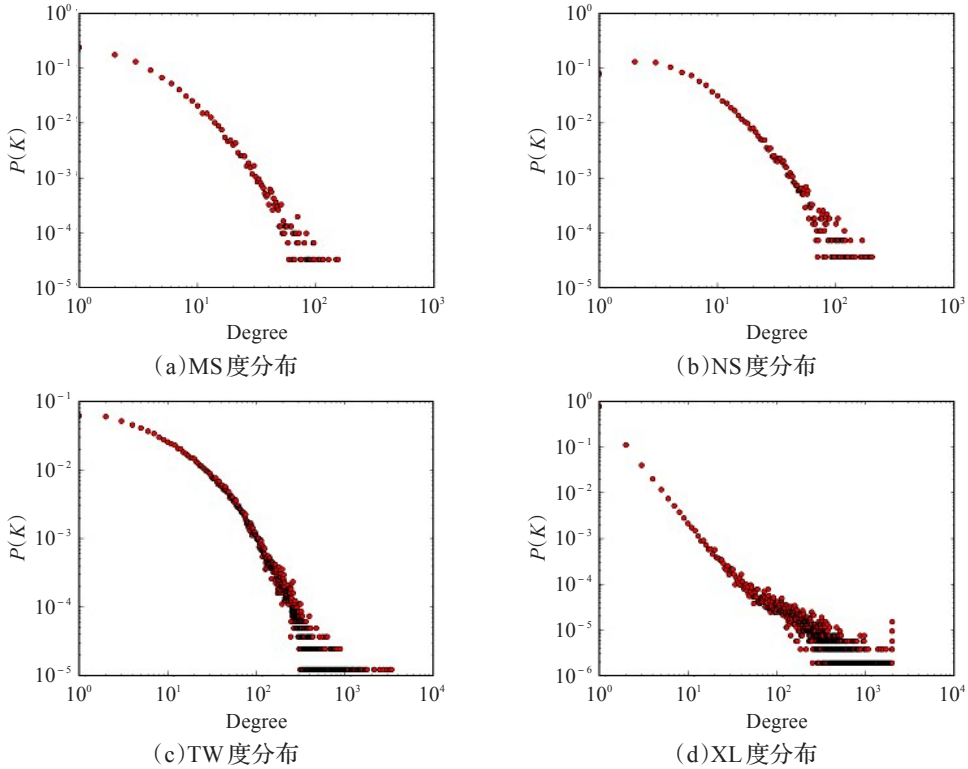


图4　各网络的度分布

表2　各网络中影响力前五的节点

排名		MS/值	NS/值	TW/值	XL/值
CK	1	cjun50/189	AKIMITSU,J/121	48485771/364	40343/2 000
	2	wll20508200/78	AEPPLI,G/103	40981798/334	10413/2 000
	3	Kevincancer/55	AMARAL,LAN/60	11348282/227	11086/271
	4	evafu1/50	ALA-NISILA,T/41	7861312/204	10469/208
	5	Sukyroom/46	ALLOUL,H/36	7860742/188	11088/114
CB	1	coolrock007/0.307 7	AFFLECK,I/0.120 80	3359851/0.096 70	40343/0.621 8
	2	Xhzd/0.233 7	AKIMITSU,J/0.111 90	783214/0.076 40	40783/0.269 2
	3	cjun50/0.127 9	AEPPLI,G/0.076 50	10228272/0.068 29	10145/0.130 0
	4	liuxiao1114/0.111 8	ALCARAZ,FC/0.0666	17093617/0.068 10	10413/0.117 8
	5	Yqqyeah/0.1065	XAVIER,JC/.066 49	41716273/0.067 50	11022/0.115 0
CC	1	coolrock007/0.261 9	AKIMITSU,J/0.107 20	3359851/0.322 10	40343/0.446 0
	2	Xhzd/0.258 9	AFFLECK,I/0.105 20	48485771/0.307 80	1618051664/0.444 4
	3	cjun50/0.239 8	POILBLANC,D/0.101 40	15234657/0.302 10	1197161814/0.425 8
	4	Maggiexia/0.234 1	TAKAGI,H/0.101 20	115485051/0.301 10	1282005885/0.401 5
	5	cat-gg/0.230 0	ITO,T/0.100 70	3829151/0.299 70	1182391231/0.397 6
Score	1	cjun50/49 084	AKIMITSU,J/19 459	48485771/3 665 656	40343/4 297 245
	2	Chanxiner/22 173	AEPPLI,G/14 817	3359851/2 555 351	10413/1 665 921
	3	sophiachenxm/18 648	TAKAGI,H/11 602	358775055/2 441 255	1224801831/16 159 216
	4	sahala4 620/17 627	ALTSHULER,BL/11 544	463952369/2 391 163	1400672683/15 137 640
	5	wll20508200/16 483	LEE,KK/11 454	100581193/1 344 021	1395412207/14 365 732

要特性。针对此问题,文献[2],文献[9]与文献[11]的研究内容主要是在模型信息传播过程时,对简单病毒传播模型的改进。目前主要的改进手段为添加新类型的节点以描述社交网络中节点的传播状态,或设定一些符合特定社交网络中信息的规则。本文为了保持除初始传播概率这一影响传播的因素外,其他影响因素不变,故使用了传统的SIR模型,以方便对照实验结果。

3.3　实验模拟

实验首先通过对比已有的节点权威性度量单位:度中心度、介数中心度、紧密中心度来确定本文影响力评分Score计算方法的有效性。最后对比了使用基于影响力算法的传播概率,与使用固定传播概率的最终传播结果的差异。

表2为根据上述指标得到影响力最大的前5个节

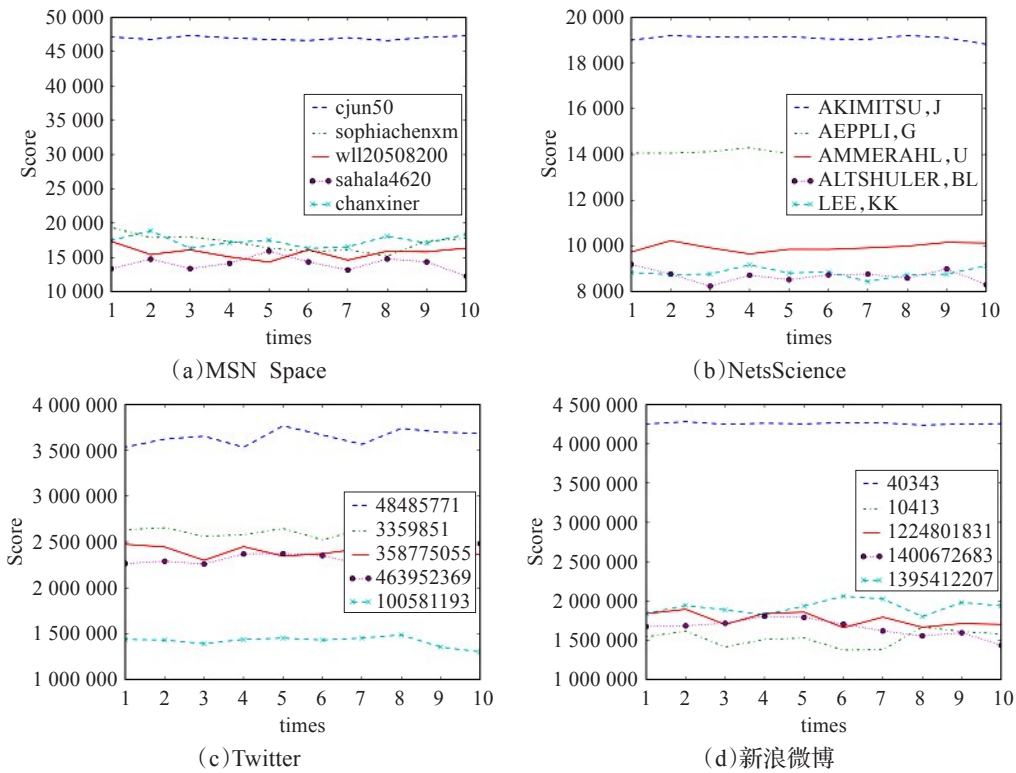
(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

点。其中 C_K 表示度中心度的计算结果，C_B 表示介数中心度的计算结果，C_C 表示紧密中心度的计算结果，*Score* 为本文方法的计算结果，游走长度 L 取值为5，随机游走器数量取值为 $degree(e)\times 100$。如表2所示，总体来看本文提出的Score值与度中心度的衡量结果相似度更大，和介数中心度与紧密中心度的衡量结果相似度较小。介数中心度与紧密中心度的衡量结果较为相似，因为这两个指标都是基于最短路径的，而本文提出的方法采用的是随机游走与度中心度结合的一种方法，因而结果更接近于度中心度。现今社交网络的规模越来越大，许多社交网络上有数亿个节点，即使截取部分网络，如同介数中心度与紧密中心度这样时间复杂度高达 $O(n^3)$ 的算法仍然是不可接受的。本文Score算法好处在于通过控制随机游走器的数量与游走的路径长度，可以很好地控制算法的时间消耗，在计算效率与精确度之间的关系上具有良好的灵活性。

因为 *Score* 的计算方法是基于随机游走的，而随机游走本身带有一定的不确定性，所以为确保实验结果的稳定性，本文在随机游走器的数量分别设为 $degree(e)\times 10$、$degree(e)\times 50$、$degree(e)\times 100$ 的情况下对每个网络上的Score评分最大的节点进行了10次评分。实验结果表明，在 $degree(e)\times 100$ 的情况下，随机游走的结果能够达到比较稳定的状态，特别是针对权威节点的评分更加稳定。如图5所示，横轴 t 为实验轮次，纵轴为 *Score* 评分。随着随机游走器数量的增加，基于随机游走的方法的实验结果是越来越稳定的，所以 *Score* 算法应具有足够的健壮性。

针对上述结果，本文根据以往研究，采取固定传播概率 $P_{SI}=0.2$ 进行轮次 $t=15$ 的传播仿真。最终感染比率 $I(t)$ 分别与表2中的四种指标比较其关联性，绘制如图6与图7所示的关系图。好的影响力指标应当在影响力增大的同时，使最终感染人数增大。可以在图6与图7中看到，在四种网络中紧密中心度的表现相对较好，特别是在MSN Space与NetsScience中，紧密中心度的结果与最终感染人数比率有很强的正相关性，在Twitter与新浪微博中的相关性相对较小。这可能是因为相比MSN Space与NetsScience，Twitter与新浪微博这类社交网络的社交性更强，网络更加复杂。人们在这种网络上传播信息会受到更多种因素的影响，如是否是热点话题，是否含有图片与超链接等因素都会影响最终的传播结果，所以信息源节点对权威性这一因素的作用可能会被稀释。在MSN Space与NetsScience网络中，*Score* 值的效果虽不如紧密中心度，但明显优于度中心度与介数中心度。而在Twitter网络中紧密中心度表现出一定的关联性，其余指标表现结果均不理想。在新浪微博中，*Score* 值略优于其他三种指标。

为了验证 *Score* 值的有效性，图8为在MSN Space、NetsScience、Twitter与新浪微博中以各影响力指标排名第一的节点作为信息源节点，同样以固定传播概率



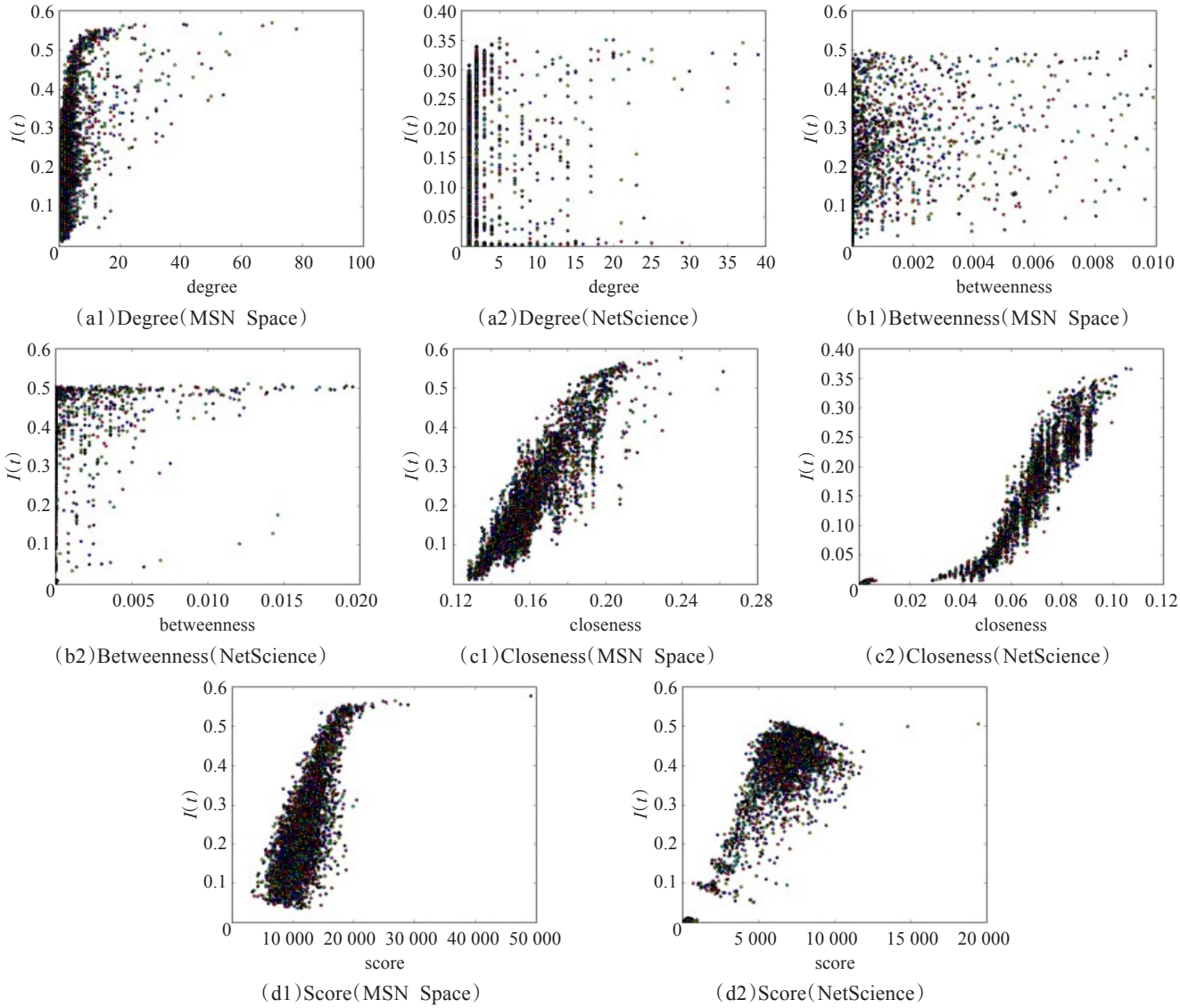
(a)MSN Space　(b)NetsScience

(c)Twitter　(d)新浪微博

注：横坐标为时间轮次

图5　Score评分的稳定性示意图

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net



(a1)Degree(MSN Space)　(a2)Degree(NetScience)　(b1)Betweenness(MSN Space)

(b2)Betweenness(NetScience)　(c1)Closeness(MSN Space)　(c2)Closeness(NetScience)

(d1)Score(MSN Space)　(d2)Score(NetScience)

图6　MSN Space与NetsScience上最终感染比率与各影响力指标间的关系

$P_{SI}=0.2$在SIR模型中传播10次,得到的最终感染率变化的曲线图。可以看到在MSN Space中,四种影响力指标均排名第一的节点cjun50在最终感染比率上明显高于其他节点,这首先可以说明节点权重对信息传播结果有明显的影响,高影响力节点传播的信息会感染更多节点。其次,可以看到cjun50节点的最终感染率是明显高于其他节点的,由表2所示数据可知,*Score*值将此节点排名第一,且cjun50的*Score*值是明显高于其他节点的,而其他评价值却相差不大,这显示出*Score*值一定的优越性。由*Score*值识别出的一个重要节点chanxiner,与紧密中心度、介数中心识别为第二重要节点的xhzd其最终感染率相近,但其他指标却未识别出此节点。在NetsScience网络中,由度中心度、紧密中心度与Score识别为排名第一的AKIMITSU,J,它的最终感染率的确较高,而介数中心度识别出的排名第一的节点AFFLECK,I的最终感染率也比较高,这说明这四种指标NetsScience网络中均有良好的表现。其中*Score*值将AEPPLI,G节点排在第二位,而度中心度与介数中心度也识别出了这个重要节点,且*Score*排名前五个重要节点中,有三个节点由其他评价指标确定为排名前五的重要节点,说明这些节点的评价指标有一定程度的相似性。在Twitter网络中,首先可以看到由各个评价指标选出的影响力排名前五的节点差异很大,与图8数据所示结论一致,各个评价指标的表现均不稳定,差异较大。感染率最高的节点7861312由度中心度与介数中心度识别出,但得分不高。在新浪微博中,感染率较高的10413节点,除了紧密中心度为将其排名前五外,其他三种指标均识别出了这个重要节点。总的来说,*Score*评分与其他影响力指标选择出的重要节点,对信息传播的结果均有一定影响,证明*Score*值在评价节点影响力方面的有效性。且各指标在不同的网络上,性能好坏有差异。

实验最后,在上述四种网络上,任选一个节点为信息源节点,分别采用由本文提出的基于影响力的信息传播概率$P(v)$,与固定概率$Fixed=0.2$进行传播模拟。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

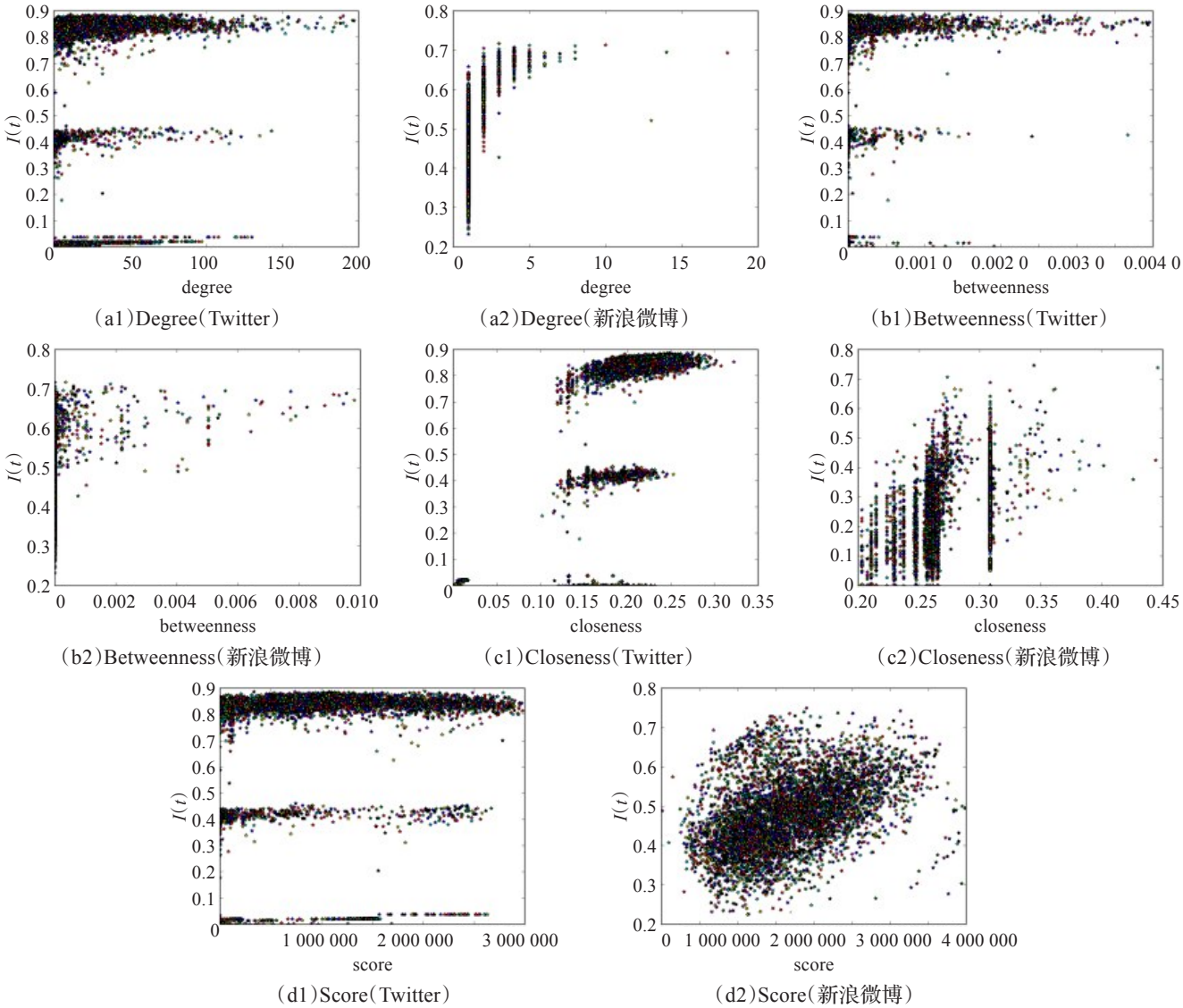


图7　Twitter与新浪微博上最终感染比率与各影响力指标间的关系

表3　对比固定传播概率与基于影响力的传播概率

网络名:节点名	度中心度	介数中心度	紧密中心度	Score/(Max/Min)	Fixed/$P(v)$
MS:chanxiner	42	0.021	0.223	22 173/(49 084/580)	0.20/0.45
NS:AEPPLI,G	103	0.076	0.101	14 817/(19 459/7)	0.20/0.76
TW:3359851	355	0.097	0.322	2 555 351/(3 665 656/4)	0.20/0.68
XL:10413	370	0.118	0.349	1 665 921/(4 297 245/11 385)	0.20/0.39

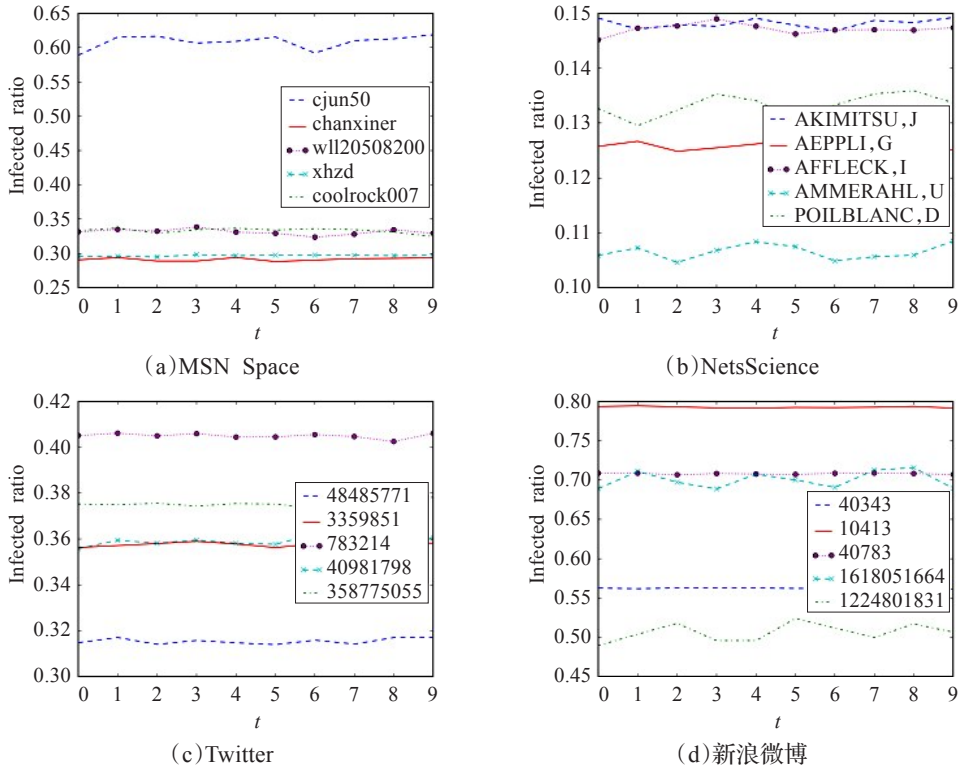
具体选取情况见表3。实验结果如图9所示,横轴t为传播轮次,纵轴为感染节点比例。可以明显看到,不同的初始传播概率对信息的传播过程影响极大。对于不同的初始传播概率,曲线斜率代表的信息传播的速度不一致,曲线顶点代表的最大信息传播范围不一致,达到信息传播的最大范围时间也不一致,同时感染节点比例再次归零,即信息消亡的时间也不一致,尤其在NetScience网络上差异极大。这些结果证明了在以往的研究中,所有传播过程均指定固定概率,忽略不同信息源节点的差异,是极为不准确的。根据节点的影响力,给信息源节点不同的传播概率,相比人为设定固定的传播概率更为合理。

4　结束语

在以往的研究中,研究者的研究重点往往在于信息的传播过程中,而忽略在传播开始之前的环境影响。本文提出了一种基于节点影响力的信息传播概率算法,用以确定不同信息源节点的初始传播概率。实验通过在SIR模型上模拟信息传播过程,通过验证影响力算法的有效性,证明计算后的传播概率更加合理。影响力算法不仅可以用于计算信息传播概率,同样在谣言传播、专家发现、传染病控制等方面均有重要价值。

本文重点是在模拟传播开始之前,根据节点本身属性确定一个合适的初始传播概率,代替以往研究中人为

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

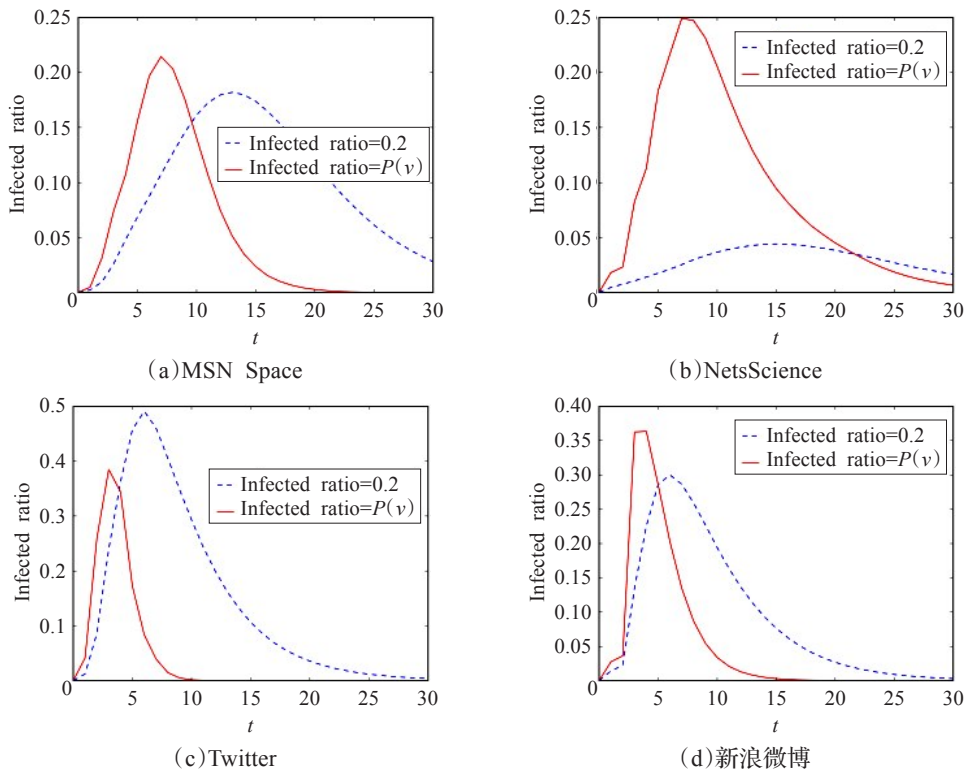


(a)MSN Space　(b)NetsScience

(c)Twitter　(d)新浪微博

注:横坐标 t 为传播轮次

图8　各权威节点的最终感染比率



(a)MSN Space　(b)NetsScience

(c)Twitter　(d)新浪微博

注:横坐标 t 为传播轮次

图9　固定传播概率对比基于节点影响力的传播概率

设定的固定概率。应当明确的是,在实际信息传播过程中,传播概率会受到很多因素的影响,如会受到周围权威节点,或当前舆论导向的影响,这些在传播过程中动态的影响因素将会是日后的研究方向。

参考文献:

[1] Zhao L J,Wang J J,Chen Y C,et al.SIHR rumor spreading model in social networks[J].Physica A,2012,391(7):2444-2453.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

[2] 顾亦然,夏玲玲.在线社交网络中谣言的传播与抑制[J].物理学报,2012,61(23).
[3] 王辉,韩江洪,邓林,等.基于移动社交网络的谣言传播动力学研究[J].物理学报,2013,62(11).
[4] 曹玖新,吴江林,石伟,等.新浪微博网信息传播分析与预测[J].计算机学报,2014(4).
[5] Hong L,Dan O,Davison B D.Predicting popular messages in Twitter[C]//Proceedings of the 20th International Conference Companion on World Wide Web,2011.
[6] Dickens L,Molloy I,Lobo J.Learning stochastic models of information flow[C]//IEEE 28th International Conference on Data Engineering(ICDE),2012.
[7] 张彦超,刘云,张海峰,等.基于在线社交网络的信息传播模型[J].物理学报,2011,60(5).
[8] Yang J,Leskovec J.Modeling information diffusion in implicit networks[C]//Proceedings of the 2010 IEEE International Conference on Data Mining,Sydney,Australia,2010:599-608.
[9] 王金龙,刘方爱,朱振方.一种基于用户相对权重的在线社交网络信息传播模型[J].物理学报,2015,64(5).
[10] Lü L,Chen D B,Zhou T.Small world yields the most effective information spreading[J].New Journal of Physics,2011,12.
[11] 唐朝生.在线社交网络信息传播建模及转发预测研究[D].河北秦皇岛:燕山大学,2014.
[12] Kimura M,Saito K,Nakano R,et al.Extracting influential nodes on a social network for information diffusion[J].Data Min Knowl Disc,2010,20:70-97.
[13] Chen D,Lü L,Shang M S,et al.Identifying influential nodes in complex networks[J].Fuel & Energy Abstracts,2012,391(4):1777-1787.
[14] Kitsak M,Gallos L K,Havlin S,et al.Identification of influential spreaders in complex networks[J].Nature Physics,2010,6(11):888-893.
[15] Batagelj V,Zaversnik M.An $O(m)$ algorithm for cores decomposition of networks[J].Advances in Data Analysis and Classification,2011,5(2):129-145.
[16] Hou B,Yao Y,Liao D.Identifying all-around nodes for spreading dynamics in complex networks[J].Physics A:Statistical Mechanics and Its Applications,2012,391(15):4012-4017.
[17] Hu Q,Gao Y,Ma P,et al.A new approach to identify influential spreaders in complex networks[J].Acta Physica Sinica,2013,62(14):99-104.
[18] Lü L,Zhang Y C,Chi H Y,et al.Leaders in social networks,the delicious case[J].Plos One,2011,6(6).
[19] Zhang X,Zhu J,Wang Q,et al.Identifying influential nodes in complex networks with community structure[J].Knowledge-Based Systems,2013,42(2):74-84.
[20] Wei D,Deng X,Zhang X,et al.Identifying influential nodes in weighted networks based on evidence theory[J].Physica A Statistical Mechanics & Its Applications,2013,392(10):2564-2575.
[21] Lao N,Cohen W W.Relational retrieval using a combination of path-constrained random walks[J].Machine Learning,2010,81(1):53-67.
[22] Lao N,Cohen W W.Fast query execution for retrieval models based on path-constrained random walks[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2010:881-888.
[23] Moreno Y,Nekovee M,Pacheco A F.Dynamics of rumor spreading in complex networks[J].Physical Review E,2004,69(6).

(上接72页)

[11] Ren C,Wang D,Urgaonkar B,et al.Carbon-aware energy capacity planning for datacenters[C]//2012 IEEE 20th International Symposium on Modeling,Analysis & Simulation of Computer and Telecommunication Systems(MASCOTS),2012:391-400.
[12] Feller E,Morin C,Leprince D.State of the art of power saving in clusters and results from the EDF case study[D].Institut National de Recherche en Informatique et en Automatique(INRIA),2010.
[13] 罗亮,吴文峻,张飞.面向云计算数据中心的能耗建模方法[J].软件学报,2014(7):1371-1387.
[14] Kontorinis V,Zhang L E,Aksanli B,et al.Managing distributed UPS energy for effective power capping in data centers[C]//2012 39th Annual International Symposium on Computer Architecture(ISCA),2012:488-499.
[15] 孙鉴,李战怀,张晓,等.基于统计量的存储系统磁盘功耗建模方法研究[J].计算机研究与发展,2016,53(7):1517-1531.
[16] Pinheiro E,Bianchini R.Energy conservation techniques for disk array-based servers[C]//ACM International Conference on Supercomputing,2014:369-379.
[17] 邢文训,谢金星.现代优化计算方法[M].2版.北京:清华大学出版社,2005:113-208.
[18] Srikantaiah S,Kansal A,Zhao F.Energy aware consolidation for cloud computing[C]//Proceedings of the 2008 Conference on Power Aware Computing and Systems,2008:1-5.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net