



## 注意力流网络中节点影响力的层级性研究

李 勇<sup>1</sup>,董思秀<sup>1</sup>,张 强<sup>1</sup>,程方颀<sup>2</sup>,王常青<sup>3</sup>

(1.西北师范大学 计算机科学与工程学院,兰州 730070; 2.北京航空航天大学 计算机学院,北京 100190;  
3.中国互联网络信息中心 互联网基础技术开放实验室,北京 100190)

**摘 要:** 复杂网络中节点影响力的层级性在网络结构与控制研究中至关重要。针对有向加权网络中节点影响力的层级性问题,基于海量在线用户行为数据,构建有向加权集体注意力流网络。通过定义节点的层级位置时间和位置约束指标,并结合节点的拓扑位置和时间序列,提出一种用于有向加权网络的节点影响力度量及排序算法。实验结果表明,该算法能有效区分网络层级结构,准确识别出最具影响力的节点,对于节点影响力评估与复杂网络可控性研究具有一定的借鉴意义和参考价值。

**关键词:** 注意力流网络;拓扑位置;时间序列;节点影响力;K-Shell算法

开放科学(资源服务)标志码(OSID):



**中文引用格式:** 李勇,董思秀,张强,等.注意力流网络中节点影响力的层级性研究[J].计算机工程,2021,47(8):109-115,123.  
**英文引用格式:** LI Y, DONG S X, ZHANG Q, et al. Research on the hierarchy of node influence in attention flow network[J]. Computer Engineering, 2021, 47(8): 109-115, 123.

## Research on the Hierarchy of Node Influence in Attention Flow Network

LI Yong<sup>1</sup>, DONG Sixiu<sup>1</sup>, ZHANG Qiang<sup>1</sup>, CHENG Fangqi<sup>2</sup>, WANG Changqing<sup>3</sup>

(1.College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China;  
2.School of Computer Science and Engineering, Beihang University, Beijing 100190, China;  
3.Domain Named System Laboratory, China Internet Network Information Center, Beijing 100190, China)

**[Abstract]** The hierarchy of node influence in complex networks is crucial for the research on network structure and controllability. As for the problem about the hierarchy of node influence in directed and weighted networks, a directed and weighted collective attention flow network is constructed based on massive data of online user behavior. By defining the Hierarchical Position Time (HPT) and position constraint of nodes, an algorithm for measuring and ranking the node influence in directed weighted networks is proposed. The algorithm also considers the topological positions and time series of nodes. Experimental results show that this algorithm can distinguish the hierarchical structure of network, identifying the most influential nodes accurately. It displays certain reference significance and reference value in evaluating the influence of nodes and the controllability of complex networks.

**[Key words]** attention flow network; topological position; time series; node influence; K-Shell algorithm

DOI: 10.19678/j.issn.1000-3428.0059141

### 0 概述

人类社会是一个类似混沌系统的复杂系统,几乎所有的人类社会现象和自然现象均可利用社交网络、生物网络等复杂网络模型进行描述。复杂网络内部结构错综复杂,存在少数关键核心节点,核心节点的微小扰动会引起网络的系统性涨落,甚至导致网络彻底崩溃。注意力流网络<sup>[1-3]</sup>是近年来兴起的一个新型的复

杂网络,由在线用户在不同信息源上连续的点击行为构成,其中,节点表示用户点击的信息源,边表示用户从一个信息源到下一个信息源的跳转,将人类的注意力看作抽象流动的物质。在注意力流网络研究领域,研究人员在前期研究中已发现多个演化普适模式,包括异速标度律、耗散律、引力定律和heaps律等<sup>[4-6]</sup>。LI等<sup>[2]</sup>基于在线集体注意力流,研究网站的影响力。WU等<sup>[3]</sup>分析了复杂网络上点击流的分散流结构。SHI等<sup>[7]</sup>提

**基金项目:** 国家自然科学基金“基于社会感知计算的公众环境感知与时空行为研究”(71764025);全国高等院校计算机基础教育教学研究项目(2020-AFCEC-355);甘肃省高等学校科研项目(2018A-001);西北师范大学青年教师科研能力提升计划项目(NWNU-LKQN-17-9)。

**作者简介:** 李 勇(1979—),男,副教授、博士,主研方向为大数据、社会计算;董思秀,硕士研究生;张 强,副教授、博士;程方颀,学士;王常青,高级工程师、博士。

**收稿日期:** 2020-08-03 **修回日期:** 2020-09-25 **E-mail:** dongsixiu@foxmail.com

出一种在不同网站之间分配和流动集体注意力的几何表示方法。GU等<sup>[8]</sup>提出一种基于流的几何嵌入及其数值近似改进算法,根据流距离定义的节点中心性对站点进行排名。

扩散模型用于度量节点的传播影响力<sup>[9-10]</sup>并可对节点重要性进行排序,主要包括阈值模型<sup>[11]</sup>、级联模型<sup>[12-13]</sup>、流行病模型<sup>[14-15]</sup>等模型。但是,在大型复杂网络中使用扩散模型度量节点的传播能力并对节点进行排序非常耗时。针对这一问题,近年来研究人员提出了多种节点排序的方法。这些方法主要分为结构化方法和超结构化方法,在结构化方法中节点的传播能力仅基于拓扑位置,在超结构化方法中除了节点拓扑位置外,还考虑了个体特征、用户兴趣等因素。由于对额外信息的需求较低,并且节点的传播能力仅基于网络结构确定,因此结构化方法受到了更多的关注。根据网络结构中使用的信息类型,结构化方法又分为局部、半局部、全局和混合方法。局部结构化方法仅依赖节点及其邻居来度量其影响力,如度中心性<sup>[16]</sup>和H-index中心性<sup>[17]</sup>。半局部结构化方法除了邻居的信息之外,还使用二阶邻居来度量节点的传播能力,如权重度中心性和扩展权重度中心性<sup>[18]</sup>。全局结构化方法需要遍历整个网络获取全局信息来度量节点的影响力,如紧密中心性<sup>[19]</sup>、中介中心性<sup>[20]</sup>、核数中心性<sup>[21]</sup>和K-Shell。混合结构化方法利用局部和全局信息来度量节点传播能力,如混合度分解<sup>[22]</sup>、邻域核数<sup>[23]</sup>、K-Shell迭代因子<sup>[24]</sup>和混合核心、度和熵<sup>[25]</sup>。除了上述结构化方法外,还有一些度量节点传播能力的博弈论模型。文献[26]考虑网络结构,使用合作博弈论提出多种节点中心性度量方法。文献[27]将节点传播能力的度量问题建模为非合作博弈问题,并根据该模型度量节点的传播能力。

网络中通过连接不同子网络的关键节点维持整个网络的凝聚力,对信息流动或控制十分关键。现有的关键节点识别方法多数仅关注单个或局部节点,很少从网络整体性、系统性上探讨节点影响力,而且这些方法多数是针对无向无权网络,关于有向加权网络节点影响力的层级性<sup>[28]</sup>研究较少。K-Shell<sup>[29]</sup>是网络中一种度量节点影响力的算法,但不能提供有关节点拓扑位置的充足信息。近年来,已有研究针对该问题从层级角度出发提出HKS (Hierarchical K-Shell)算法<sup>[30]</sup>。HKS算法在无向无权网络中能够准确且高效地度量节点的影响力并确定其拓扑位置,但在有向加权网络中HKS算法面临适用性问题。为解决上述问题,本文基于中国互联网络信息中心提供的海量在线用户行为大数据,构建集体注意力流网络,定义节点的层级位置时间和位置约束,同时考虑节点的拓扑位置和时间序列,提出一种用于有向加权网络的节点影响力度量及排序算法OHKS。

## 1 理论基础

### 1.1 研究框架

本文研究框架主要包括数据预处理、数据建模、OHKS算法及实验分析,如图1所示,具体过程如下:

1)数据预处理。通过分析在线用户行为日志数据,获取实验所需的点击流数据。

2)数据建模。对点击流数据进行建模,构造注意力流网络。

3)OHKS算法。在注意力流网络中,通过定义节点的层级位置时间(Hierarchical Position Time, HPT)和位置约束( $P$ )两项指标,同时考虑节点的拓扑位置和时间序列,使得每个节点都具有一个层级指标 $h$ ,然后计算其影响力。节点的HPT指标是从网络外围往核心分层计算,值越大,影响力越大,节点越靠近网络核心。节点的 $P$ 指标是从网络核心往外围分层计算,值越小,节点影响力越小,节点越接近网络外围。通过这两个指标对每个节点的位置和影响力进行层层约束,所得出的节点影响力不仅仅单独依赖于节点的度或节点的权重,而是更加综合有效。

4)实验分析。通过OHKS算法研究了注意力流网络中节点影响力的层级性,并将其实验结果和4种常规算法作对比实验分析。

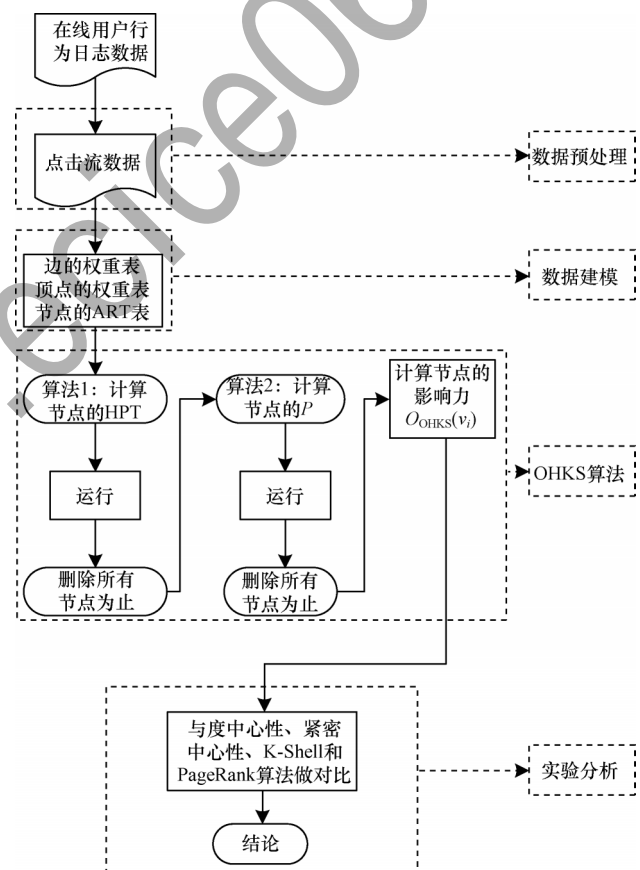


图1 研究框架

Fig.1 Research framework

## 1.2 注意力流网络

本文采用的实验数据是中国互联网络信息中心提供的在线用户行为日志数据,在保证用户个人隐私的前提下,详细记录了海量用户开关机时间、焦点窗口的窗口进程名和进程号、浏览器窗口的地址栏内容(已部分截断)、焦点窗口对应的程序版本号、程序所属公司名、用户人口属性等信息。用户每开关机一次就会建立一个相应的日志文件。每2秒就会扫描一次用户电脑显示屏最前端的焦点窗口,如果焦点窗口相比2s前已发生变化,则立即在日志中增加1条记录。为了方便分析,随机抽取200个用户1个月约800万条数据记录。

注意力流网络由一个加权有向图  $G=(V, E, T, W)$  表示,如图2所示,其中,  $V$  表示图的  $n+2$  个顶点集,  $source$  和  $sink$  是2个特殊节点,  $E$  表示图的边集,顶点权重  $T$  表示集体用户在一个站点上注意力停留的总时间,边的权重  $W$  表示注意力在各站点间转换的强度(不存在的边定义其权值为0)。

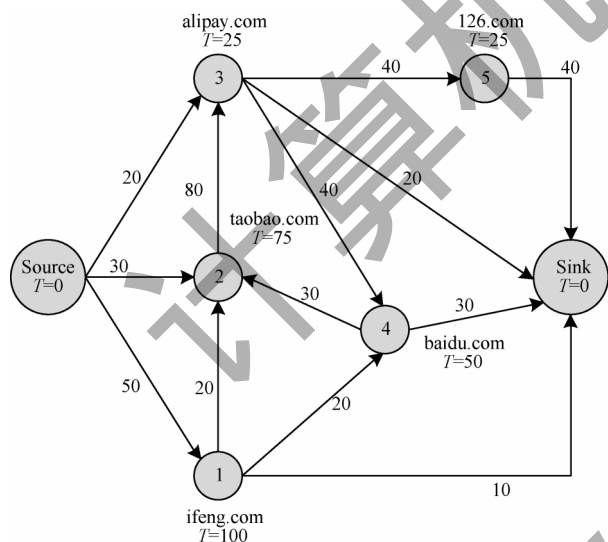


图2 注意力流网络

Fig.2 Attention flow network

## 1.3 HKS算法

K-Shell是网络中一种度量节点影响力的算法,但不能提供有关节点拓扑位置的充足信息。近年来,已有研究针对该问题从层级角度出发提出了HKS算法。ZAREIE等<sup>[30]</sup>指出图中有3类节点集可以影响节点  $v_i$  的传播能力:

- 1)在图核心的最短路径上访问  $v_i$  的节点集  $Pred_i$ 。
- 2)在图核心的最短路径上  $v_i$  访问的节点集  $Succ_i$ 。
- 3)在图核心的最短路径上  $v_i$  和  $v_j$  相互不访问的节点集  $Sibl_i$ , 其中  $v_j$  是  $v_i$  的邻域节点集。

HKS算法使用  $Pred_i$ 、 $Succ_i$ 、 $Sibl_i$  节点集指定节点

位置和影响力,利用  $b_i$  和  $f_i$  指标指定每个节点  $v_i$  的拓扑位置。 $b_i$  和  $f_i$  分别受  $Pred_i$ 、 $Sibl_i$ 、 $Succ_i$  节点集的影响,  $b$  和  $f$  分别决定了节点远离外围的程度和接近核心的程度。 $b$  实际上是节点  $v_i$  被删除的一个全局迭代计数器,计算  $b$  值的算法具体如下:1)设置  $Shell=1$  和  $b=1$  的初始值,从图中删除度等于  $Shell$  的节点,并为其分配  $b=1$ ,直到图中不再有度等于  $Shell$  的节点;2)  $b$  增加1,  $Shell$  增加1,度等于  $Shell$  的节点再次从图中被删除,并在给定计数器  $b$  全局值的情况下将它们设置为  $b_i$ ;3)不断重复该过程,直到删除图中所有节点。计算  $f$  值的算法具体如下:1)确定位于图核心的节点,规定它们的  $f$  值等于被分配的  $b$  值;2)从图的核心开始遍历,具有最高  $f_i$  值的每个节点  $v_i$  在每一步中用  $f_i-1$  值来改变未删除的邻居  $f$  值,然后删除节点  $v_i$ ;3)不断重复该过程,直到删除图中所有节点。

$H_{HKS}(v_i)$  表示节点  $v_i$  的传播影响力,计算公式如下:

$$H_{HKS}(v_i) = \sum_{v_j \in N(v_i)} S(v_j) \quad (1)$$

其中:  $v_j \in N(v_i)$  表示节点  $v_j$  是节点  $v_i$  的邻居节点;  $S(v_i)$  表示节点  $v_i$  的一阶邻域传播影响力之和。 $S(v_i)$  计算公式如下:

$$S(v_i) = \sum_{v_j \in N_i} d_j(b_j + f_j) \quad (2)$$

其中:  $v_j \in N_i$  表示节点  $v_j$  属于节点  $v_i$  的邻域;  $d_j$  表示节点  $v_j$  的度;  $b_j$  表示节点  $v_j$  的  $b$  值;  $f_j$  表示节点  $v_j$  的  $f$  值。

## 2 OHKS算法

传统HKS算法在无向无权网络中能够准确且高效地度量节点的影响力并确定其拓扑位置。然而,在有向加权网络中HKS算法面临适用性挑战。因此,本文利用OHKS算法来研究注意力流网络中节点影响力的层级性。在有向加权注意力流网络中,节点表示用户点击过的站点,用户在站点的停留时间表示节点的权重,边表示集体用户注意力从一个站点跳转到下一个站点,跳转次数表示节点的度,度表示边的权重(不区分节点的出度和入度)。OHKS算法采用净注意力流入来度量节点的影响力,用户浏览站点的先后顺序表示边的方向,根据入度计算边的权重(区分出度和入度),再结合顶点的权重得出点入中心性。

### 2.1 节点层级位置时间

节点的入度平均停留时间(Average Retention Time, ART)定义为:假设站点  $A$  的入度为  $x$ ,即在站点  $A$  产生停留时间的边数为  $x$ ,每条边在站点  $A$  产生的停留时间分别为  $a_1, a_2, \dots, a_x$ ,那么站点  $A$  的 ART 计算公式如下:



$$A_{ART, A} = \frac{\sum_{n=1}^x a_x}{x} \quad (3)$$

根据ART与节点拓扑位置,计算层级位置时间的算法具体如下:1)设置计数器count=0和层级指标h=1,从图中删除ART $\in$ [count, count+1)的节点,并使其HPT=h,不断重复该步骤直到图中不再有ART $\in$ [count, count+1)的节点;2)h增加1, count增加1,再次从图中删除ART $\in$ [count, count+1)的节点,根据给定的ART全局值得出HPT;3)不断重复该过程,直到删除图中所有的节点。节点的HPT越大,影响力越大,节点越靠近图的核心。

**算法1** 每个节点的HPT( $T_1, T_2$ )计算算法

输入 节点的邻接表 $T_1$ 、节点的ART表 $T_2$

输出 节点的HPT表 $T_3$

//建立空图 $G(V, E, ART)$ ,将 $T_1$ 中的节点和边添加到图

//G,将 $T_2$ 中的ART添加到相应节点

1. 设置 count = 0

2. 设置 h = 1

3. for each ( $V \neq \emptyset$ )

4. for each ( $v_i \in V$ )

5. 设置 flag = false

6. if (count < h and ART $_{v_i} \in$  [count, count+1))

7. HPT $_{v_i}$  = h

8. flag = true

9. end if

10. end for

11. if (flag = true)

12. for each ( $v_i \in V$ )

13. if (HPT $_{v_i}$  = h)

14. 从 V 中删除节点  $v_i$

15. end for

16. count++

17. else

18. h++

19. end if

20. end for

## 2.2 节点位置约束

节点的位置约束( $P$ )表示由节点HPT和一阶邻域同时约束。节点的 $P$ 越小,节点影响力越小,节点越接近网络外围。计算位置约束的算法具体如下:1)找到具有最大HPT的节点 $v_i$ ,定义 $Q=\max(HPT_{v_i})$ ;2)将节点 $v_i$ 的HPT赋给 $P$ ,其余节点的 $P$ 都赋值为0;3)从图的核心开始遍历,寻找 $P=Q$ 的节点 $v_i$ 未删除的邻居节点,并给它们赋值为 $P-1$ ,再删除节点 $v_i$ ;4) $Q$ 自减1,不断重复该过程,直到删除图中所有的节点。

若在核心节点附近存在非核心节点,非核心节点的一阶邻域会提高该节点的 $P$ 值。类似于通过高度值的邻接节点获得间接影响力的特征向量中心

性,如果一个节点的度很高,则说明该节点有较高的中心性;如果一个节点的度不是很高,它和一个有很高度值的节点邻接,则该节点的中心性也较高。

**算法2** 每个节点的 $P(T_1, T_3)$ 计算算法

输入 节点的邻接表 $T_1$ 、节点的HPT表 $T_3$

输出 节点的 $P$ 表 $T_4$

//建立空图 $G(V, E, HPT)$ ,将 $T_1$ 中的节点和边添加到

//图G,将 $T_3$ 中的HPT添加到相应节点

1. 设置  $Q = \max(HPT_{v_i})$

2. for each ( $v_i \in V$ )

3. if ( $\exists v_j \in N_i$  where  $HPT_{v_j} > HPT_{v_i}$ )

4.  $P_{v_i} = HPT_{v_i}$

5. else

6.  $P_{v_i} = 0$

7. end for

8. for each ( $V \neq \emptyset$ )

9. for each ( $v_i \in V$ )

10. if ( $P_{v_i} == Q$ )

11. for each ( $v_j \in N_i$ )

12. if ( $P_{v_i} - 1 > P_{v_j}$ )

13.  $P_{v_j} = P_{v_i} - 1$

14. end if

15. end for

16. end if

17. end for

18. 从 V 中删除节点  $v_i$

19.  $Q = Q - 1$

20. end for

## 2.3 节点影响力

$O_{OHKS}(v_i)$ 表示节点 $v_i$ 的影响力,计算公式如下:

$$O_{OHKS}(v_i) = \sum_{v_j \in N(v_i)} S(v_j) \quad (4)$$

$S(v_i)$ 计算公式如下:

$$S(v_i) = \sum_{v_j \in N_i} d_{v_j} (H_{HPT_{v_j}} + P_{v_j}) \quad (5)$$

其中: $H_{HPT_{v_j}}$ 表示节点 $v_j$ 的HPT值; $P_{v_j}$ 表示节点 $v_j$ 的 $P$ 值。

## 3 实验设置

### 3.1 数据建模

通过对在线用户行为日志数据进行提取,可获得用户的点击流数据如表1所示。在有向加权注意力流网络中,节点表示用户在浏览网页时所点击的站点,边表示用户注意力从该站点流出进入下一站点。在数据建模过程中,需要生成站点之间边的权重(dataew)站点的顶点的权重(datanw)、站点的入度平均停留时间,如表2~表4所示。在表4中, datanw1为表 datanw中相同站点的顶点的权重累加,

datanw2 为表 datanw1 中相同站点的顶点的权重累加, dataew1 为 dataew 中相同跳转的边的权重(不分出度和入度)累加, dataew2 为根据入度计算出边的权重(分出度和入度)累加。

表 1 点击流数据  
Table 1 Clickstream data

日期	站点	时间
2020-08-01	['2345.com']	12:07:03
2020-08-01	['236jj.com']	12:07:19
2020-08-01	['236jj.com']	12:07:23
2020-08-01	['236jj.com']	12:07:46
2020-08-01	['236jj.com']	12:07:50
2020-08-01	['youku.com']	12:07:54

表 2 边的权重  
Table 2 Edge weight

站点(出)	停留时间/s
['2345.com']	16
['236jj.com']	35
['youku.com']	15

表 3 顶点的权重  
Table 3 Vertex weight

站点(出)	站点(入)	跳转数
['2345.com']	['236jj.com']	1
['236jj.com']	['youku.com']	3
['youku.com']	['236jj.com']	2

表 4 站点的入度平均停留时间

Table 4 In-degree average retention time of site

站点(出)	站点(入)	datanw1	datanw2	dataew1	dataew2	ART/s
2345.com	236jj.com	16	16	1	1	16
236jj.com	youku.com	1 569	1 569	5	3	523
youku.com	236jj.com	73	707	0	5	141

### 3.2 节点影响力的层级性网络构建

通过分析在线用户行为点击流数据,构建包含 4 627 个节点、58 284 条边的注意力流网络,如图3所示。

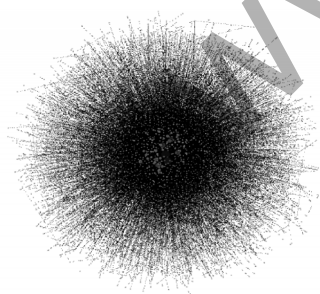


图 3 节点影响力的层级性网络

Fig.3 Hierarchical network of node influence

节点影响力的层级性网络由一个加权有向图  $G=(V, E, T, W)$  表示,主要以顶点的权重  $T$  和边的权重  $W$  为依据,得出的节点影响力不仅依赖节点的度或节点的权重,而且依赖顶点的权重和边的权重:

1)顶点的权重  $T$ 。评估节点影响力的 OHKS 值,由节点的层级位置时间和位置约束综合得出。OHKS 值越大,节点影响力越大,节点越靠近网络核心。对应于可视化过程中,节点半径越大,颜色越深。

2)边的权重  $W$ 。节点之间的跳转数,是一个累加的过程。连边越多,节点影响力越大,节点越靠近网络核心,即复杂网络中的“意见领袖”思想,它强调连结度高的个体在新的意见或信息传播中起重大作用。对应于可视化过程中,边越粗,颜色越深。

### 3.3 实验结果分析

为分析与验证 OHKS 算法得到的注意力流网络中节点影响力的层级性结果,将其与度中心性、紧密中心性、K-Shell、PageRank 算法进行对比,其中,度中心性属于局部结构化方法,紧密中心性属于全局结构化方法,K-Shell 是一种识别关键节点的经典全局结构化算法,PageRank<sup>[31-32]</sup> 是一种研究节点影响力的基本算法。

#### 3.3.1 对比算法分析

在 OHKS 算法中必须不断重复从网络中删除节点,算法 1 计算了每个节点的 HPT,时间复杂度为  $O(n)$ ,其中  $n$  是网络中的节点数,算法 2 计算了每个节点的  $P$ ,时间复杂度为  $O(n)$ ,节点  $v_i$  的一阶邻域影响力之和  $S(v_i)$  和影响力  $O_{OHKS}(v_i)$  的时间复杂度为  $O(n)$ 。因此,OHKS 算法的时间复杂度为  $O(n)$ 。

局部结构化方法的核心思想是具有大量邻居的高度节点更具影响力,并且具有结构简单和时间复杂度低等优点,但仅依赖节点及其邻居来度量影响力,忽略了网络的全局结构。度中心性衡量网络中一个节点和其他节点的关联程度,是最基本的中心性度量算法。对于一个有  $g$  个节点的无向图,节点  $i$  的中心度是  $i$  与其他  $g-1$  个节点的直接关联总数,计算公式如下:

$$C_D(n_i) = \sum_{j=1}^g x_{ij} \quad (6)$$

其中:  $C_D(n_i)$  表示节点  $i$  的中心度,将节点  $i$  在网络矩阵中对应的行或列所在的单元格值累加;  $\sum_{j=1}^g x_{ij}$  ( $i \neq j$ ) 表示节点  $i$  和  $g-1$  个节点  $j$  的直接关联数量;  $i \neq j$  表示排除  $i$  与自身的联系。

在全局结构化方法中需要遍历整个网络获取全局信息来度量节点的影响力,节点影响力由网络全局结构决定,因此它们具有更高的时间复杂度。紧密中心性为网络中节点在最短路径上的距离,表示节点  $v_i$  和

其他节点  $v_j$  的最短距离之和的倒数, 计算公式如下:

$$C_D(v_i) = \frac{n-1}{\sum_{j \neq i} g(v_i, v_j)} \quad (7)$$

其中:  $C_D(v_i)$  表示节点  $v_i$  的紧密中心度;  $g(v_i, v_j)$  表示  $v_i$  和  $v_j$  的最短路径距离。

K-Shell 是网络中一种度量节点影响力的算法, 但不能提供有关节点拓扑位置的充足信息, 算法具体过程如下: 1) 给每个节点分配一个  $k_s$  指标, 从图中删除度为 1 的节点, 直到不再有度为 1 的节点,  $k_s=1$  被分配给已删除的节点; 2) 从图中删除度为 2 的节点, 直到不再有度为 2 的节点,  $k_s=2$  被分配给已删除的节点; 3) 不断重复该过程, 直到从图中删除所有节点。

PageRank 由于遵循马尔科夫过程和随机游走设想, 需要反复迭代获取 PR 值, 其实验数据量要求大、实验设备性能要求高, 且运行时间长。PageRank 通过网页之间的链接结构来度量网页的重要性, 是类似于特征向量中心性的算法, 计算公式如下:

$$P_{PR_i} = \varphi \sum_j \frac{P_{PR_j}}{O_j} + (1-\varphi) \quad (8)$$

其中:  $\varphi \in (0, 1)$  是一个常数, 被称为阻尼系数, 表示任意时刻用户访问到某页面后继续访问下一个页面的概率;  $P_{PR_i}$  表示前一个节点  $j$  的 PageRank 值;  $O_j$  表示顶点  $j$  的出度。

### 3.3.2 节点影响力识别方式分析

基于 OHKS 算法得出的影响力前 15 名的站点如表 5 所示。根据度中心性、紧密中心性、K-Shell 和 PageRank 算法得出的影响力前 15 名的站点排名, 如表 6 所示, 其中 K-Shell 算法的站点排名不区分先后顺序。

表 5 基于 OHKS 算法的影响力前 15 名的站点排名

Table 5 Ranking of the top 15 influential sites based on the OHKS algorithm

排名	站点	OHKS 值
1	baidu.com	14 298
2	qq.com	5 854
3	sogou.com	4 082
4	sina.com	3 295
5	hao123.com	3 137
6	sohu.com	2 381
7	taobao.com	2 328
8	360.cn	2 120
9	weibo.com	2 094
10	youku.com	1 935
11	renren.com	1 859
12	163.com	1 792
13	xunlei.com	1 730
14	eastmoney.com	1 584
15	google.com	1 437

表 6 基于 4 种算法的影响力前 15 名的站点排名

Table 6 Ranking of the top 15 influential sites based on four algorithms

排名	度中心性算法	紧密中心性算法	K-Sell 算法	PageRank 算法
1	baidu.com	baidu.com	youku.com	baidu.com
2	qq.com	qq.com	soso.com	qq.com
3	sogou.com	sogou.com	2435.com	sogou.com
4	hao123.com	hao123.com	renren.com	hao123.com
5	taobao.com	taobao.com	sina.com	taobao.com
6	sina.com	sina.com	baidu.com	sina.com
7	360.cn	360.cn	sohu.com	360.cn
8	163.com	sohu.com	taobao.com	google.com
9	google.com	163.com	qq.com	163.com
10	sohu.com	google.com	163.com	sohu.com
11	weibo.com	weibo.com	google.co	weibo.com
12	soso.com	soso.com	hao123.com	soso.com
13	youku.com	youku.com	sogou.com	2345.com
14	2345.com	xunlei.com	weibo.com	youku.com
15	renren.com	2345.com	360.cn	renren.com

由表 6 可以看出, OHKS 算法与度中心性、紧密中心性、K-sell、PageRank 这 4 种常规算法得出的站点排名前 3 名站点一致, 其中, baidu.com 和 sogou.com 属于搜索引擎类站点, qq.com 属于信息类站点。OHKS 算法结合节点的全局拓扑位置和停留时间来识别最具影响力的站点, 当用户分别访问 baidu.com、qq.com 和 sogou.com 这 3 个站点时, 形成的多次跳转依然是在同站内访问, 即给站点带来了真正有效的停留时间。结合 baidu.com、qq.com 等网站在中国的受欢迎程度, 显然会获得高排名, 且该排名与中国的 Alexa 排名趋于一致。常规算法主要以跳转为核心来识别最具影响力的站点, 随着互联网的飞速发展和用户电子设备的快速进步, 访问速度越来越快, 使得用户可以轻松、快速地在搜索引擎类站点或信息类站点中实现多次跳转, 因此具有大量跳转数的网站排名较高。

结合表 5 和表 6 可以看出, 前 3 名以外的站点不尽相同, 因为 OHKS 算法从全局角度出发主要结合节点的拓扑位置和停留时间来识别最具影响力的站点。例如, sina.com 虽然属于门户类网站, 但是它在 OHKS 算法中的排名比在常规算法中靠前, 原因在于当用户访问 sina.com 时, 除了在 sina.com 中实现多次跳转以外, 它的微博、视频和游戏等专栏会使得用户长时间停留。视频类网站 youku.com 在 OHKS 算法中的排名也比常规算法靠前, 因为当用户访问 youku.com 时, 除了多次跳转外, 更多的是注意力的长时间的停留和聚焦。

### 3.3.3 算法适用性与性能分析

算法适用性与性能分析具体如下:

1) 适用性。OHKS 算法既适用于无向无权网络, 又适用于有向加权网络。度中心性、紧密中心



性、K-Shell 和 PageRank 算法仅适用于无向无权网络。

2)性能。OHKS 算法结合节点的全局拓扑位置,时间复杂度低、运行效率高。度中心性算法的时间复杂度低,但忽略了网络的全局结构。紧密中心性考虑了网络的全局结构,但时间复杂度高。K-Shell 算法不能提供有关节点拓扑位置的充足信息。PageRank 算法实验数据量要求大、实验设备性能要求高,且运行周期长。

3)节点影响力识别方式。OHKS 算法从全局角度出发,主要结合节点的拓扑位置和时间序列来识别有影响力的节点。度中心性、紧密中心性、K-Shell 和 PageRank 算法主要以跳转为核心来识别有影响力的节点。

#### 4 结束语

本文以在线用户行为点击流大数据为研究基础,生成点击流模型并构建注意力流网络,通过定义节点的层级位置时间和位置约束对 HKS 算法进行优化,提出一种用于有向加权网络节点影响力度量及排序的算法。实验结果表明,该算法适用于有向加权网络中的节点影响力分析,能对看似吸引了大量注意力的假象节点进行甄别,准确地识别出真正有影响力的节点,从而加深对网络层级结构的认识,有助于分析网络中心性、节点聚类、社区结构等特征。后续将进一步划分站点类别,并在不同类别的社区内部进行层级性或可控性算法研究,深入探索用户行为和互联网协同演化的关系。

#### 参考文献

- [1] 李勇,孟小峰,刘继,等. 基于小数据的在线用户兴趣长期演化研究[J]. 计算机研究与发展,2015,52(4):779-788.  
LI Y, MENG X F, LIU J, et al. Study of the long-range evolution of online human-interest based on small data[J]. Journal of Computer Research and Development, 2015, 52(4): 779-788. (in Chinese)
- [2] LI Y, ZHANG J, MENG X F, et al. Quantifying the influence of websites based on online collective attention flow[J]. Journal of Computer Science and Technology, 2015, 30(6): 1175-1187.
- [3] WU L F, ZHANG J. The decentralized flow structure of clickstreams on the Web[J]. European Physical Journal B, 2013, 86(6): 266-275.
- [4] LI Y, MENG X F, ZHANG Q, et al. Common patterns of online collective attention flow [J]. Science China Information Sciences, 2016, 60(5): 1-3.
- [5] WU F, HUBERMAN B A. Novelty and collective attention[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(45): 17599-17601.
- [6] LOU X D, LI Y, GU W W, et al. The atlas of Chinese world wide web ecosystem shaped by the collective attention flows[J]. PLoS One, 2016, 11(11): 1-13.
- [7] SHI P, HUANG X, WANG J, et al. A geometric representation of collective attention flows[J]. PLoS One, 2015, 10(9): 1-21.
- [8] GU W, GONG L, LOU X, et al. The hidden flow structure and metric space of network embedding algorithms based on random walks [J]. Scientific Reports, 2017, 7(1): 114-116.
- [9] WANG X, LAN Y, XIAO J. Anomalous structure and dynamics in news diffusion among heterogeneous individuals [J]. Nature Human Behaviour, 2019, 3(7): 709-718.
- [10] CENTOLA D. Influential networks [J]. Nature Human Behaviour, 2019, 3(7): 664-665.
- [11] BORODIN A, FILMUS Y, OREN J. Threshold models for competitive influence in social networks [C]//Proceedings of WINE'10. Berlin, Germany: Springer, 2010: 539-550.
- [12] CARNES T, NAGARAJAN C, WILD S M, et al. Maximizing influence in a competitive social network: a follower's perspective [C]//Proceedings of the 9th International Conference on Electronic Commerce. New York: ACM Press, 2007: 351-360.
- [13] GOLDENBERG J, LIBAI B. Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata [J]. Academy of Marketing Science Review, 2001, 9(3): 33-45.
- [14] BUSCARINO A, FORTUNA L, FRASCA M, et al. Disease spreading in populations of moving agents [EB/OL]. [2020-07-01]. <http://dx.doi.org/10.1209/0295-5075/82/38002>.
- [15] ZHOU J, CHUNG N N, CHEW L Y, et al. Epidemic spreading induced by diversity of agents' mobility [J]. Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 2012, 86(2): 98-115.
- [16] FREEMAN L C. Centrality in social networks conceptual clarification [J]. Social Networks, 1979, 1(3): 215-239.
- [17] LÜ L Y, ZHOU T, ZHANG Q M, et al. The H-index of a network node and its relation to degree and coreness [J]. Nature Communications, 2016, 7: 101-168.
- [18] LIU Y, TANG M, ZHOU T, et al. Identify influential spreaders in complex networks, the role of neighborhood [J]. Physica A: Statistical Mechanics and its Applications, 2016, 452: 289-298.
- [19] SABIDUSSI G. The centrality index of a graph [J]. Psychometrika, 1966, 31(4): 581-603.
- [20] FREEMAN L C. A set of measures of centrality based on betweenness [J]. Sociometry, 1977, 40: 35-41.
- [21] BORGATTI S P, EVERETT M G. Models of core/periphery structures [J]. Social Networks, 2000, 21(4): 375-395.
- [22] ZENG A, ZHANG C J. Ranking spreaders by decomposing complex networks [J]. Physics Letters A, 2013, 377(14): 1031-1035.
- [23] BAE J, KIM S. Identifying and ranking influential spreaders in complex networks by neighborhood coreness [J]. Physica A: Statistical Mechanics and Its Applications, 2014, 395(1): 549-559.

(上接第 115 页)

- [24] WANG Z, ZHAO Y, XI J, et al. Fast ranking influential nodes in complex networks using a k-shell iteration factor[J]. Physica A: Statistical Mechanics and its Applications, 2016, 461(1): 171-181.
- [25] SHEIKHAHMADI A, NEMATBAKHS M A, ZAREIE A. Identification of influential users by neighbors in online social networks[J]. Physica A: Statistical Mechanics and its Applications, 2017, 486(15): 517-534.
- [26] MOLINERO X, RIQUELME F, SERNA M. Cooperation through social influence[J]. European Journal of Operational Research, 2015, 242(3): 960-974.
- [27] IRFAN M T, ORTIZ L E. On influence, stable behavior, and the most influential individuals in networks: a game-theoretic approach[J]. Artificial Intelligence, 2014, 215: 79-119.
- [28] BASSOLAS A, BARBOSA-FILHO H, DICKINSON B, et al. Hierarchical organization of urban mobility and its connection with city livability[J]. Nature Communications, 2019, 10(1): 33-45.
- [29] KITSACK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6(11): 888-893.
- [30] ZAREIE A, SHEIKHAHMADI A. A hierarchical approach for influential node ranking in complex social networks[J]. Expert Systems with Application, 2018, 93: 200-211.
- [31] BIANCHINI M, GORI M, SCARSELLI F. Inside PageRank[J]. ACM Transactions on Internet Technology, 2005, 5(1): 92-128.
- [32] Wikipedia. PageRank algorithm[EB/OL]. [2020-07-01]. <http://wiki.swarma.net/index.php?title=PageRank算法&variant=zh-hant>.

编辑 陆燕菲