

# Research on Social Network User Impact Ranking Method Based on Motif

Xianfei Qu, Zhaoxia Tian, Xu Chen

School of Information Science and Technology, Dalian Maritime University, Dalian Liaoning  
Email: xianfeiv@126.com

Received: May 15<sup>th</sup>, 2020; accepted: May 28<sup>th</sup>, 2020; published: Jun. 4<sup>th</sup>, 2020

---

## Abstract

LeaderRank has been widely used in ranking algorithms to measure the influence or importance of users in social networks, but traditional LeaderRank only uses edge-based relationships and ignores high-order relationships captured by a subgraph composed of a small number of nodes. In this paper, we propose a Motif-based LeaderRank (MLR) algorithm, which incorporates the higher-order relationships of Motif into the LeaderRank algorithm to improve the ranking effect of social network user influence. We perform experiments on the Twitter dataset, and the results not only show the feasibility of the MLR algorithm, but also significantly improve the accuracy of the ranking of user influence in social networks. In addition to comparing with the baseline algorithm, the parameters of the MLR algorithm are also analyzed, which shows that the MLR algorithm is better than LeaderRank when ranking user influence in social networks.

## Keywords

LeaderRank Algorithm, Social Network, Higher-Order Relationship, Ranking Method

---

# 基于Motif的社交网络用户影响力排序方法研究

曲贤菲, 田朝霞, 陈旭

大连海事大学信息科学技术学院, 辽宁 大连  
Email: xianfeiv@126.com

收稿日期: 2020年5月15日; 录用日期: 2020年5月28日; 发布日期: 2020年6月4日

---

## 摘要

LeaderRank已被广泛地用于衡量社交网络中用户的影响力或重要性的排序算法, 但传统的LeaderRank

仅利用基于边的关系,而忽略由少量节点组成的子图捕获的高阶关系。在本文中,我们提出一种基于Motif的LeaderRank (MLR)算法,将Motif的高阶关系合并到LeaderRank算法中,提高社交网络用户影响力的排序效果。我们在Twitter数据集上进行实验,结果不仅表明MLR算法的可行性,并且还显著提高了在社交网络用户影响力排序的准确性。除了与基线算法比较之外,还对MLR算法的参数进行分析,表明在社交网络中对用户影响力排序时,MLR算法比LeaderRank要更好。

## 关键词

LeaderRank算法, 社交网络, 高阶关系, 排序方法

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来,网络在人们生活中的地位越加凸显,在线社交网络已成为大多数人日常交流和互动的平台。社交网络中的用户影响力排序是意见领袖挖掘[1]、影响力分析[2] [3]、社会信任度[4]的普遍问题。PageRank [5]算法被视为在社交网络中对用户影响力排序的通用算法,可将其用来衡量意见领袖挖掘以及影响力分析和社会信任度分析的一种措施。但该算法在应对悬挂节点(出度为0的节点)、不连通子图或出现环形有向图时会出现排序结果不唯一的现象,并且该算法用到的阻尼系数 $d$ 对每个用户都是统一的,这点用在不同的用户身上存在着缺陷。Lv等[6]针对相关的情况,提出了一种LeaderRank算法,其是在网络中添加一个背景节点,并且将该背景节点与网络中的其他节点双向连接,得到一个 $G(N+1, M+2N)$ 的新的强连通网络,如图1所示,对网络的噪音有很好的容忍性,相比PageRank算法收敛更快、准确性更高。

但是LeaderRank在社交网络分析中有一些限制,在网络中,可能存在一些重要的结构,这些结构可能会对节点的影响力或可信度产生影响。如图1所示,在LeaderRank算法中边 $e_{12}$ 、边 $e_{13}$ 和边 $e_{14}$ 的权重设为相同的值,但是我们可以看到,相比节点4,节点1更信任节点2和节点3,是因为节点2和节点3存在相互关系,节点1、节点2和节点3构成三角形结构,因此将边 $e_{12}$ 、边 $e_{13}$ 和边 $e_{14}$ 的权重设为相同的值是不合理的。因此在社交网络中考虑涉及多个节点的高阶结构是非常重要的。

在本文中,我们提出了一种基于Motif的LeaderRank (Motif-based LeaderRank, MLR)算法,我们捕获网络中Motif的高阶关系,并将该高阶关系合并到LeaderRank算法中,用于社交网络用户影响力排序。文献[7]提到,由于在社交网络方面,许多工作着重于分析三重闭合作为社交网络的重要结构特征以及分析三元配置作为各种社交网络理论的基础,因此,在本文中,我们主要考虑社交网络的三角关系,在本文中我们主要考虑基于社交网络三角关系的3-节点简单Motif。图2显示了一些典型的基于社交网络三角关系的3-节点简单Motif,其中 $M_6$ 表示图1中节点1、节点2和节点3的3-节点简单Motif。我们在美国社交网络及微博客服务数据集Twitter上进行实验,来提取有影响力用户,我们使用基线算法度中心性方法IND,原始LeaderRank算法和加权LeaderRank算法,结果表明我们提出的MLR算法明显优于基线算法度中心性方法IND,原始LeaderRank算法和加权LeaderRank算法。在这项工作中,我们采用线性关系来对基于社交网络三角关系的3-节点简单Motif的高阶关系和基于边的直接关系进行组合,通过对参数的调节,表明了基于社交网络三角关系的3-节点简单Motif的高阶关系和基于边的直接关系在计算节点的权威性方面是相互补充的。

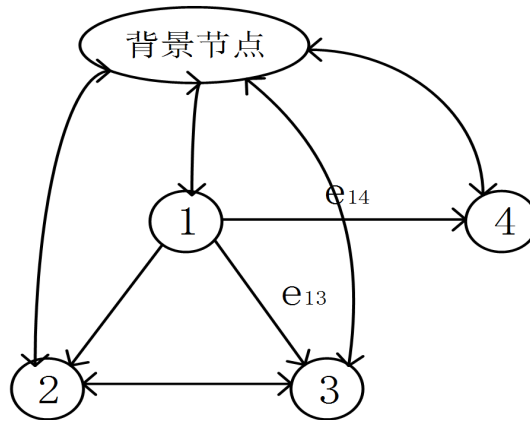


Figure 1. A schematic diagram of the LeaderRank algorithm. The algorithm introduces a background node in the original network

图 1. LeaderRank 算法示意图。该算法在原网络中引入一个背景节点

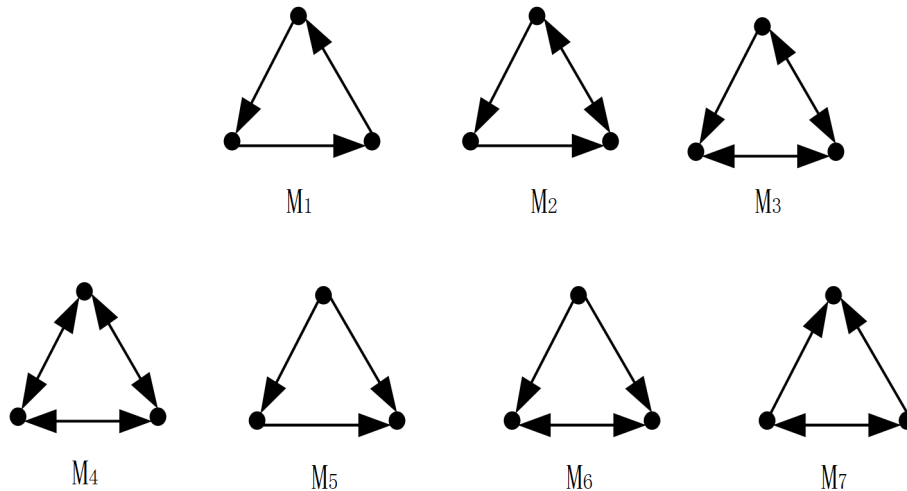


Figure 2. 3-Node simple Motif based on social network triangle relationship in directed unweighted graph

图 2. 有向无权图中的基于社交网络三角关系的 3-节点简单 Motif

## 2. 相关研究

在本节中，我们介绍在复杂网络上使用 LeaderRank 和 Motif 进行计算的相关工作。

### 2.1. LeaderRank 算法

首次引入 LeaderRank 来对网络上的网页进行排序是 Lv 等[6]，该算法主要是参考 PageRank 算法的基本思想，并在 PageRank 算法的基础上来进行不断的改进，并且在  $N$  个节点里面加上背景节点这样就获取了  $N + 1$  的网络  $G(N + 1, M + 2N)$ 。除了进行网页排名之外，LeaderRank 还用于许多其他领域，例如引文网络分析和意见领袖挖掘。在引文网络方面，邓启平等[8]充分考虑到了引用、合作这样不同的选项，分析哪些人更有影响力，并且按照合作数量还有被引的情况进行分析，并对作者开展相应的排序活动。在意见领袖方面，徐郡明等[9]则是充分想到使用者彼此有比较大的想法区别，都会按照自身的感情进行分析问题，就以此进行研究，并对 LeaderRank 算法进行了完善，添上了情感这一项，经过相应的完

善, 该算法具有了比较强的抗干扰水平, 并且更加精准。此外, Zhang 等[10]则充分分析了多种不一样的算法, 并且据此开展研究, 将节点聚类, 获取其中边的具体权值。Bakshy 等[11]在具体的验证过程中对其中的一些 Twitter 数值进行比较, 了解用户的影响力的用途。顾亦然等[12]则充分依照到 LeaderRank 算法, 通过相似度分析彼此的用途, 这样就能让具体的效果更好, 并且通过相应的模型, 让算法精准度更好。与我们的工作相比, 所有先前的研究仅是在 LeaderRank 算法中仅仅考虑了边关系, 而忽略了多个节点之间的高阶关系。

## 2.2. 复杂网络中 Motif 的识别问题

网络是一个由多个节点组成的集合, 节点之间有一定的连接, 而不同于规则网络和随机网络的一些网络被称为复杂网络。复杂网络具有许多统计特征, 其中最重要的是小世界效应[13][14]和无标度特性[15][16]。随着复杂网络研究的进一步深入, 网络中 Motif 的识别问题越来越受到关注, Milo 等[17]人首次对在复杂网络中识别网络 Motif 的问题进行系统的阐述, 激发人们对网络 Motif 识别技术研究的极大兴趣, 在社交网络[18][19], 学术网络[20], 生物学[21]等方面应用广泛。此外, 之前的大部分工作都集中在如何对子图计数, 在文献[22]中, 提出一种快速高效的并行算法对 3-节点 Motif 和 4-节点 Motif 进行计算。在文献[23]中, 提出一种采样算法对 4-节点 Motif 进行计算。在文献[24]中提出了一种向左向右摇摆的蹒跚随机游走算法, 该算法通过使用随机协议对步行走访的节点附近的子图进行采样来实现其计算效率。Motif 也被应用在图聚类或社区检测[25]方面。在文献[20]中, Wang 等人提出通过衡量节点在网络中的参与程度来衡量其重要性。与这些先前的研究相比, 我们结合了 Motif 来探索基于社交网络三角关系的 3-节点简单 Motif, 然后将该高阶关系与基于边的直接关系进行结合通过 LeaderRank 来对社交网络用户影响力进行排序。

## 3. 基于 Motif 的 LeaderRank 算法

### 3.1. 问题描述

社交网络中的社区由网络中一定数量的节点组成, 令社交网络  $G = (V', E', W')$ , 其中  $V' = \{v'_i | i = 1, 2, \dots, f\}$  是节点集合,  $E' = \{e'_{ij} | i, j = 1, \dots, f\}$  是边的集合,  $e'_{ij}$  表示从  $v'_i$  到  $v'_j$  的一条边。  $W'$  是邻接矩阵,  $W'_{ij}$  代表  $e'_{ij}$  的权重。  $G$  是由  $\{G_1, G_2, \dots, G_r\}$  多个社区构成, 如图 3 所示, 利用社区发现算法划分社区, 对某一社区进行社交网络用户影响力排序。令社区  $G_1 = (V, E, W)$ , 其中  $V = \{v_i | i = 1, 2, \dots, n, n+1\}$  是节点集合,  $E = \{e_{ij} | i, j = 1, \dots, n, n+1\}$  是边的集合,  $e_{ij}$  表示从  $v_i$  到  $v_j$  的一条边。  $W$  是邻接矩阵,  $W_{ij}$  代表  $e_{ij}$  的权重。除背景节点外的有向无权图, 如果  $e_{ij}$  存在则  $W_{ij} = 1$  否则  $W_{ij} = 0$ 。然后归一化邻接矩阵得到转移概率矩阵  $P$ , 其中  $P_{i,j} = W_{i,j} / \sum_i W_{i,j}$ 。图上的 LeaderRank 可以定义如下。

$$x = P^T \cdot x \quad (1)$$

其中  $x \in RN + 1$ , 并且  $x_i$  表示当迭代达到稳定时的 MLR 值。达到稳定之后, 将背景节点的 MLR 值平分给  $N$  个节点, 最终的 MLR 值计算为

$$\text{MLR}_i = x_i + \frac{x_{n+1}}{N} \quad (2)$$

公式(2)中,  $i = 1, 2, \dots, n$ ,  $N$  表示网络中除背景节点以外的所有节点的个数。

在一个社区  $G_1$ , 我们可以将该定义概括如下。如果存在从节点  $v_i$  到  $v_j$  的一条边  $e_{ij}$ , 则我们用  $W_{ij}$  表示  $v_i$  对  $v_j$  的认可强度或  $v_j$  对  $v_i$  施加的影响强度。这个权重可以基于交互频率, 用户帖子的内容相似度等来计算, 但是, 以上所有权重仍然是两个用户之间的一阶关系的计算。如图 1 所示, 对于用户 1,  $W_{13}$  应

该大于  $W_{14}$ , 因为节点 1 和节点 3 参与了可信度评估中的三角关系。因此, 理想的是, 权重还可以包含关于这种高阶关系的信息。

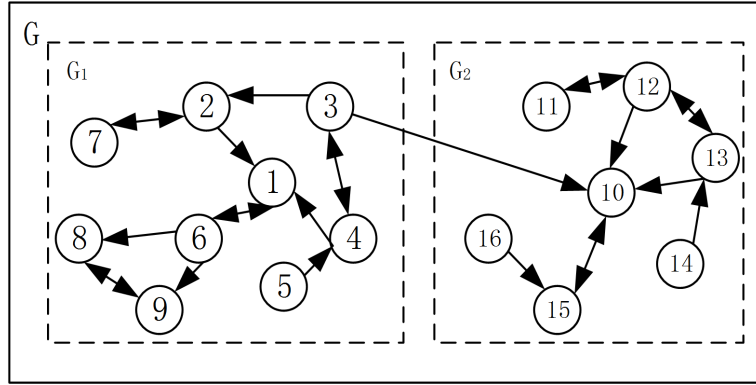


Figure 3. Network without attributes  
图 3. 不包含属性的网络

### 3.2. Motif 定义

**定义 1: 网络 Motif.** 一个 Motif 由元组  $(B, A)$  表示, 其中  $A \subset \{1, 2, \dots, k\}$ ,  $B$  表示  $k \times k$  二进制矩阵。在文献[25]中给出锚节点代表我们感兴趣的节点。通常情况下, 锚节点是所有  $k$  个节点, 在这种情况下, 我们称它为简单锚定, 否则, 我们称它为锚定 Motif。在本文中, 我们关注 3-节点简单 Motif, 图 4 给出了一个 3-节点简单 Motif 示例, 说明 Motif 的定义。

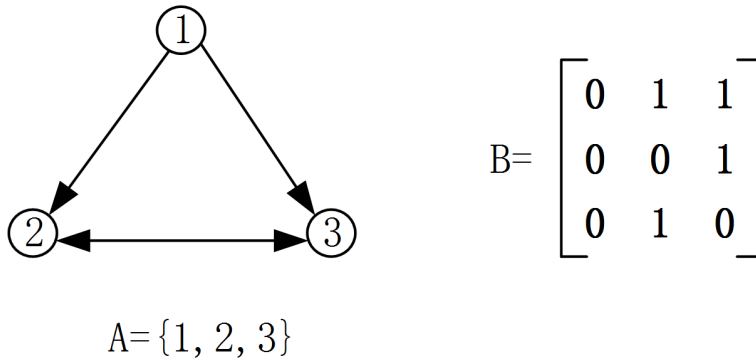


Figure 4. An example for a simple 3-nodes Motif  
图 4. 一个 3-节点简单 Motif 示例

**定义 2: Motif 集合.** 带有邻接矩阵  $W$  的有向无权图的 Motif 集合用  $F(B, A)$  表示, 定义为

$$F(B, A) = \{(\text{set}(v), \text{set}(\delta A(v))) \mid v \in V^k, v_1, \dots, v_k, \text{distinct}, W_v = B\} \quad (3)$$

其中  $\delta A$  是一个选择函数, 为由  $A$  索引的  $k$  个元组的子集, 而  $\text{set}(\cdot)$  是一个将有序元组转换为无序元组的运算符  $\text{set}((v_1, v_2, \dots, v_k)) = \{v_1, v_2, \dots, v_k\}$ 。  $v$  是表示  $k$  个节点的有序向量,  $W_v$  是有序向量  $v$  的  $k$  个节点在子图上引起的  $k \times k$  个邻接矩阵。

我们用  $F(B, A)$  表示图 5 的示例如下。

$$F(B, A) = \{(\{1, 2, 3\}, \{1, 2\}), (\{1, 2, 5\}, \{1, 2\})\} \quad (4)$$

### 3.3. 基于 Motif 的邻接矩阵

当给定一个 Motif 集合  $F$ , 我们使用  $F$  中两个节点的共现来捕获 3-节点简单 Motif 的高阶关系, 基于 Motif 的邻接矩阵或共现矩阵的定义如下:

$$(W_M)_{ij} = \sum_{(v, \delta A(v)) \in F} I(\{i, j\} \subset \delta A(v)) \quad (5)$$

其中  $i \neq j$ ,  $I(s)$  是真值指标函数, 即如果语句  $s$  为真, 则  $I(s) = 1$ , 否则为 0。请注意, 对于 3-节点简单 Motif, 要求  $i$  和  $j$  是  $set(v)$  的成员。

基于 Motif 的邻接矩阵表示在给定 Motif 上的两个节点出现的频率。 $(W_M)_{ij}$  越大, 表示在 Motif 上  $i$  和  $j$  之间的关系越明显。如图 5 所示, 该图是由图二中的  $M_1$  和  $M_6$  的 3-节点简单 Motif 构成, 当我们以  $M_6$  的 3-节点简单 Motif 来计算邻接矩阵时, 构成  $M_6$  的 3-节点简单 Motif 的边权值  $W_{M_{ij}} = 1$ , 其他  $W_{M_{ij}} = 0$ 。注意, 边  $e_{12}$  既属于节点 1、节点 2 和节点 5 构成的  $M_6$  的 3-节点简单 Motif 又属于节点 1、节点 2 和节点 3 构成的  $M_6$  的 3-节点简单 Motif, 因此边  $e_{12}$  的权值为  $W_{M_{12}} = 2$ 。

令  $W$  为  $G$  的邻接矩阵,  $U$  和  $B$  分别为  $G$  的单向链接和双向链接的邻接矩阵。这里, 我们集中讨论  $W$  中的元素为 1 或 0 的非加权图。例如, 在图 5 中,  $e_{23}$  是双向边, 而  $e_{13}$  是单向边。那么我们有  $B = W \Theta W^T$  和  $U = W - B$ , 其中  $\Theta$  表示逐项乘积。注意,  $B$  是一个二进制矩阵, 表示有向图中两个节点之间存在双向边。

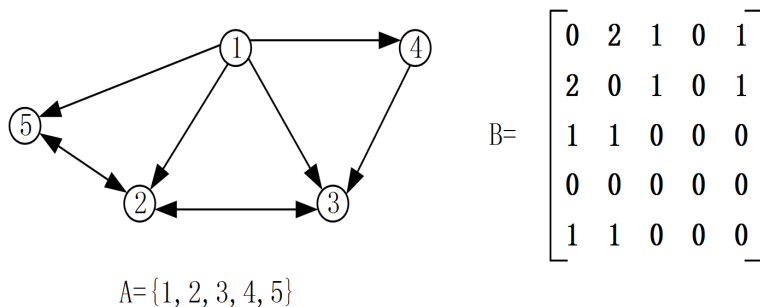


Figure 5. An example for computing  $M_6$  based adjacency matrix  
图 5. 基于  $M_6$  的邻接矩阵示例

我们使用一个示例来说明  $M_6$  的计算过程。如图 4 所示, 三个用户  $v_i, v_k, v_j$  有六种不同的情况  $\{(3,1,2), (2,1,3), (1,2,3), (3,2,1), (1,3,2), (2,3,1)\}$ ,  $e_{23}$  是双向边, 而  $e_{12}$  和  $e_{13}$  是单向边。为了计算  $M_6$  参与下的  $v_i$  和  $v_j$  的频率, 我们需要将在六种情况下的所有频率相加, 在  $(1,2,3)$  中,  $v_i$  为节点 1,  $v_j$  为节点 3, 频率由  $(U \cdot B) \Theta U$  所得, 其中  $U \cdot B$  计算的是不包括边  $e_{13}$  的路径 1-2-3,  $\Theta U$  表示用边  $e_{13}$  补充图。因此  $M_6$  邻接矩阵的计算为

$$W_{M_6} = (U \cdot B) \Theta U + (B \cdot U^T) \Theta U^T + (U^T \cdot U) \Theta B \quad (6)$$

### 3.4. 基于 $M_6$ 邻接矩阵的案例

#### 1) 社区发现算法

这里使用文献[26]中所提到的基于随机游走方法来计算节点属性相似性、节点影响力, 评估节点对社区重要程度, 使节点向社区核心靠拢来划分社区。

第一步: 子空间构造。发现子属性空间, 构建属性增强图。

将属性转化为网络中的虚拟节点来构造属性增强图, 在网络  $G$  中, 节点与其对应的虚拟节点间建立一条双向边, 如图 6 所示:

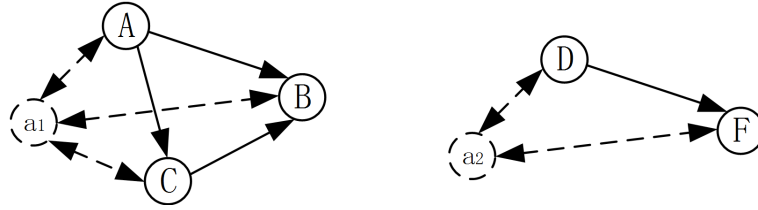


Figure 6. Virtual node  
图 6. 虚拟节点

构造属性信息矩阵  $Y = (y_{ij})_{f \times f}$  来描述虚拟节点对原节点关系的影响，构造增量邻接矩阵  $W^{Attr} = (w_{ij}^{Attr})_{f \times f}$  来描述加入虚拟节点后新的网络拓扑结构。通过属性影响度  $Affect(A_i)$  来反映将属性信息转化为拓扑结构信息后，原节点间的联系受到多大的影响，计算公式为

$$Affect(A_i) = \frac{\sum_i \sum_j (W^{Attr^2})_{ij} - \sum_i \sum_j (W^{r2})_{ij}}{R^{\beta A}} \quad (7)$$

其中  $A_i$  为属性； $\beta A$  为矩阵缩放因子，作用为更加明显的区分属性的结构影响度。

第二步：社区划分。计算转移概率，计算核心系数，确定聚类方向，创建初始社区及边缘修剪。

构造节点影响力  $strength$  来描述任意节点的邻接节点对该节点产生的影响， $strength_{ij}$  的计算公式为

$$strength_{ij} = s_{ij}^{out} + s_{ij}^{in} + s_{ij}^{Attr} \quad (8)$$

其中  $s_{ij}^{out}$  为节点  $v_i$  指向节点  $v_j$  的出链产生的影响力， $s_{ij}^{in}$  为节点  $v_j$  指向节点  $v_i$  的入链产生的影响力， $s_{ij}^{Attr}$  为属性关系产生的影响力。

$$s_{ij}^{out} = \begin{cases} 0, w'_{ij} = 0 \\ 1 + e^{-\beta_{out} \sum_h w_{ih}}, w'_{ij} = 1 \end{cases} \quad (9)$$

$$s_{ij}^{in} = \begin{cases} 0, w'_{ji} = 0 \\ e/2^{1-\beta_{in} \sum_h w_{hi}}, w'_{ji} = 1 \end{cases} \quad (10)$$

$$s_{ij}^{Attr} = \begin{cases} 0, y_{ij} = 0 \\ \frac{1-1.1^{1-\beta_{Attr} \sum_h w_{ih}}}{2/(1+\log D')}, y_{ij} = 1 \end{cases} \quad (11)$$

其中  $\beta_{out}$ 、 $\beta_{in}$  和  $\beta_{Attr}$  分别表示出链系数、入链系数和属性系数，用于控制函数的收敛速度， $D'$  为子属性空间包含的属性数目。

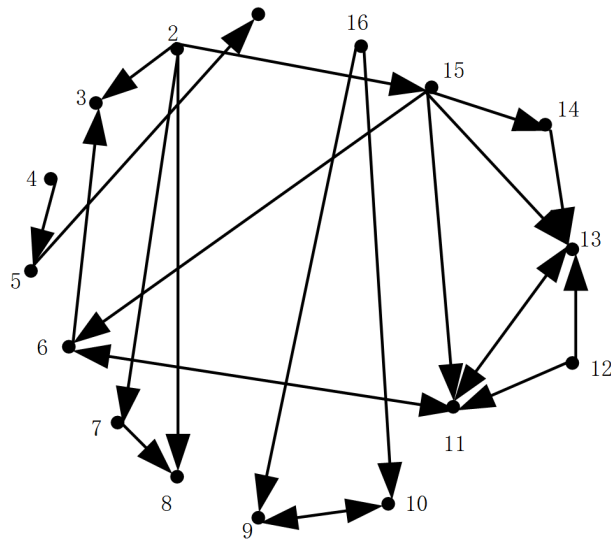
利用影响力矩阵，得到转移概率矩阵  $P'$ ， $P' = (p'_{ij})_{f \times f}$ 。

$$P'_{ij} = \frac{s_{ij}}{\sum_k s_{ik}} \quad (12)$$

引入节点核心系数  $Cr = (cr_1, cr_2, \dots, cr_f)^T$ 。节点的转移概率和节点的核心系数共同作用于节点的聚类方向  $Dr = (dr_1, dr_2, \dots, dr_f)^T$ 。取其中系数最大的作为聚类方向，得到初始社区，进行边缘修剪工作，输出社区序列。

2) 基于 Motif 的邻接矩阵

我们构建社区  $G_1(16,17)$  的网络拓扑结构，如图 7 所示。



**Figure 7.** Community  $G_1(16,17)$  subgraph structure diagram  
**图 7.** 社区  $G_1(16,17)$  子图结构示意图

我们对图 7 识别 3-节点简单 Motif  $M_6$ ，首先我们构建一阶矩阵  $W$ ，如图 8 所示。

$$W = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

**Figure 8.** Community  $G_1(16,17)$  first-order matrix  $W$   
**图 8.** 社区  $G_1(16,17)$  一阶矩阵  $W$

之后依据算法 1 得出基于  $M_6$  的邻接矩阵。

算法 1：基于  $M_6$  的邻接矩阵算法

输入：一阶矩阵  $W$

输出：基于  $M_6$  的邻接矩阵  $W_{M_6}$

a)  $W^T \leftarrow W$  矩阵的转置矩阵

b)  $B \leftarrow W \odot W^T$  单项链接矩阵

c)  $U \leftarrow W - B$  双向链接矩阵

d)  $U' \leftarrow U$  矩阵中的每个元素取绝对值

e)  $U'^T \leftarrow U'$  矩阵的转置矩阵



- f) result1 ← U' · B
- g) result1' ← result1 ⊕ U'
- h) result2 ← U' · U'<sup>T</sup>
- i) result2' ← result2 ⊕ U'<sup>T</sup>
- j) result3 ← U'<sup>T</sup> · U'
- k) result3' ← result3 ⊕ B
- l) W<sub>M<sub>6</sub></sub> ← result1' + result2' + result3'
- m) return W<sub>M<sub>6</sub></sub>

由算法 1 可得如图 9 所示的基于 W<sub>M<sub>6</sub></sub> 的邻接矩阵。

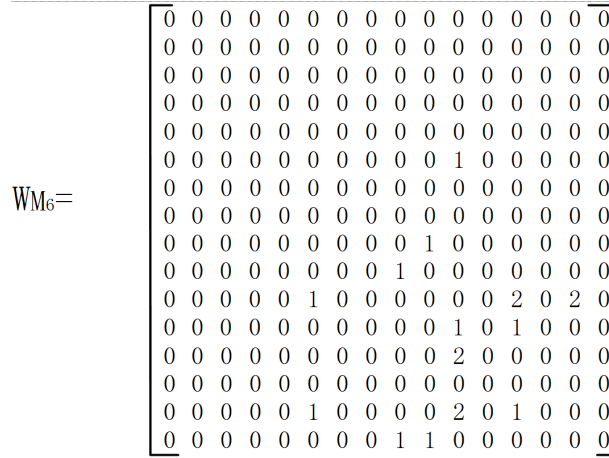


Figure 9. Community G<sub>1</sub> (16,17) based on the M<sub>6</sub> adjacency matrix W<sub>M<sub>6</sub></sub>

图 9. 社区 G<sub>1</sub> (16,17) 基于 M<sub>6</sub> 邻接矩阵 W<sub>M<sub>6</sub></sub>

由图 7 的案例，基于边的关系所得的邻接矩阵为 W，仅考虑两个节点 i 和 j，若存在 e<sub>ij</sub> 即节点 i 指向节点 j，则 W<sub>ij</sub> = 1，否则 W<sub>ij</sub> = 0。但由图 7 可观察到节点 11 与节点 13 所连成的边 e<sub>1113</sub> 的权重要明显大于其他边的权重，节点 11 与节点 15 所形成的 e<sub>1115</sub> 边权重也明显大于其他边的权重，主要原因为节点 11 与节点 13 所连成的边 e<sub>1113</sub> 既属于由节点 11、节点 13 和节点 15 所围成的三角关系，也属于由节点 11、节点 12 和节点 13 所围成的三角关系，因此，我们引入基于 Motif 的高阶关系来计算基于 Motif 的邻接矩阵，来对节点 11 与节点 13 所连成的边 e<sub>1113</sub> 的权重和节点 11 与节点 15 所形成的 e<sub>1115</sub> 边权重重新赋值，提高社交网络用户影响力排序的准确性。

### 3.5. 线性合并后的 LeaderRank

在计算出基于 Motif 的邻接矩阵之后，我们将它们合并到用于对用户影响力排序的排序模型中。由于基于 Motif 的邻接矩阵中的非零元素将不超过原来基于边的邻接矩阵中的非零元素，因此我们将基于社交网络三角关系的 3-节点简单 Motif 的高阶关系与基于边的关系结合，替代原来的基于边的关系。这样，基于 Motif 的边可以视为对使用 LeaderRank 进行计算的补充。我们使用线性组合来融合基于边和基于 Motif 的邻接矩阵。具体来说，对于给定的 Motif M<sub>k</sub>，我们生成如下的新矩阵。

$$H_{M_k} = \alpha \cdot W + (1 - \alpha) \cdot W_{M_k} \tag{13}$$

其中  $\alpha \in [0,1]$  平衡了原来的基于边的关系和由 Motif 提供的 3-节点简单 Motif 的高阶关系，

$$(P_{M_k})_{ij} = (H_{M_k})_{ij} / \sum_i (H_{M_k})_{ij} \tag{14}$$

之后我们可以利用公式(14)计算转移概率矩阵  $P_{M_k}$  并将其代入公式(1)中的转移概率矩阵  $P$ ，计算迭代稳定时的 MLR 值，迭代稳定后利用公式(2)得到节点的最终 MLR 值。

算法 2: 基于 Motif 的 LeaderRank (MLR)算法。

输入: motif 集合  $F_{mk}$ , 邻接矩阵  $W$ 。

输出: MLR<sub>i</sub> 的值。

```

a) for( $v, \delta A(v)$ )  $\in$  FMK do
b)   for  $i \in v$  do
c)     for  $j \in v$  do,  $i \neq j$ , do
d)       if  $w_{ij} > 0$  then
e)         ( $WMK$ ) $_{ij}++$ ;
f)       end if
g)     end for
h)   end for
i) end for
j) return  $Wmk$ 
k) 使用线性组合来融合基于边和基于 Motif 的邻接矩阵, 利用公式(13)得到新矩阵 HMK
l) 利用公式 (14) 计算转移概率矩阵 PMK
m)  $t \leftarrow 0, \nabla \leftarrow 1$ 
n) while( $\nabla > 10^{-5}$ )
o)   代入公式(1)
p)    $\nabla = x(t+1) - x(t)$ 
q)    $t \leftarrow t+1$ 
r) end while
s) 代入公式(2)
s) return MLRi

```

## 4. 实验

### 4.1. 实验数据集

我们选用美国社交网络及微博客服务数据集 Twitter, 依据关注关系, 选取 2000 个节点, 57,544 条组成, 在构建社交网络时, 当用户  $v_i$  信任用户  $v_j$  时, 将添加边  $e_{ij}$ , 网络数据是从 <http://law.di.unimi.it/webdata/twitter-2010/>上面下载的。

### 4.2. 实验环境

我们的实验环境为在 Windows10 操作系统上进行操作, 使用 python 2.7.16, numpy 1.11.2, Networkx 1.11 的开发环境进行实验。

### 4.3. 评价指标

为了评估所提出 MLR 算法的可行性与准确性, 我们比较了按不同算法排序的前  $K$  个用户的质量。具体来说, 我们提取具有最大 LeaderRank 值的  $K$  个用户, 然后计算归一化折损累积增益(NDCG), 这个指标通常用来衡量和评价搜索结果算法。

NDCG 计算公式表示为

$$NDCG @ N = \frac{DCG @ N}{IDCG @ N} \quad (15)$$

其中

$$DCG @ N = \sum_{i=1}^N \frac{2^{rel_i} - 1}{\log_2(i+1)} \quad (16)$$

$$\text{IDCG}@N = \sum_{i=1}^{|REL|} \frac{2^{rel_i} - 1}{\log_2(i+1)} \quad (17)$$

其中  $rel_i$  代表给定查询的文档相关性得分,  $|REL|$  表示文档按照相关性从大到小排序, 取前  $N$  个文档组成的集合,  $\log_2(i+1)$  表示如果将较高的相关项放在较低的位置, 则使用  $\log_2(i+1)$  来惩罚该算法,  $\text{IDCG}@N$  指的是理想情况下, 根据相关性得分对结果进行排序。

**基线算法。**在对社交网络进行社区划分时, 节点的转移概率和节点的核心系数共同作用于节点聚类方向, 主要是考虑节点的度中心性, 因此选用基线算法度中心方法 IND, 它根据度中心性来选择有影响力的节点, BLR 方法是原始的 LeaderRank 算法。WLR 方法是在加权网络中运行 LeaderRank, 其中从  $i$  到  $j$  的边的权重设置为  $i$  引用  $j$  或  $i$  信任  $j$  的频率。

对于 MLR, 我们分别处理  $M_1 \sim M_7$  这七种情况下的 3-节点简单 Motif, 然后分别显示每个 3-节点简单 Motif 的结果。为了进行准确性比较, 我们展示了在不同  $\alpha$  值中可以获得的每个 Motif 的最佳结果, 并展示参数  $\alpha$  如何影响 MLR 的准确性。

#### 4.4. 实验结果分析

##### 1) 可行性与准确性分析

对 Twitter 数据集分别使用基线算法 IND、BLR、WLR 和我们的 MLR 算法计算 top10、top50 和 top200 的 NDCG 值, 如表 1 所示。

**Table 1.** NDCG for top10, top50, top200 users from Twitter

**表 1.** Twitter 数据集 top10、top50、top200 的 NDCG 值

	Twitter		
top	10	50	200
IND	0.9637	0.9515	0.9334
BLR	0.9677	0.9434	0.9256
WLR	0.9677	0.9434	0.9256
MLR	0.9875	0.9649	0.9479

总体结果如表 1 所示。我们显示了使用不同算法得出的 top10, top50, top200 排序结果的 NDCG 值。从表 1 中, 我们可以看到, 通过我们提出的 MLR 算法, 具有不同  $K$  值的三个数据集它的 NDCG 值都显示出很高的准确性, 展现了将 3-节点简单 Motif 的高阶关系合并到 LeaderRank 算法中的可行性。

由表 1 可以得出 3 个观察结果。第一, 在 Twitter 数据集上, 我们可以看到,  $K$  越大, NDCG 值越小, 这意味着更多的用户获得理想排序列表的难度更大。第二, 我们可以看到在 Twitter 上的 BLR 和 WLR 在  $K = 10$  时 NDCG 大于 0.96,  $K = 50$  时 NDCG 大于 0.94,  $K = 200$  时, NDCG 大于 0.92, 而我们提出的 MLR 可以将 NDCG 在  $K = 10$  时从 0.9637 提升到 0.9875, 在  $K = 50$  时从 0.9434 提升到 0.9649, 在  $K = 200$  时 0.9256 提升到 0.9479。第三, 我们发现 BLR 和 WLR 的 NDCG 值相似, 这意味着使用频率作为权重并不会带来太多额外的好处。

从表可知, 将基于社交网络三角关系的 3-节点简单 Motif 的高阶关系与基于边的直接关系线性组合到 LeaderRank 可以提高社交网络用户影响力排序的准确性。

##### 2) $\alpha$ 值分析

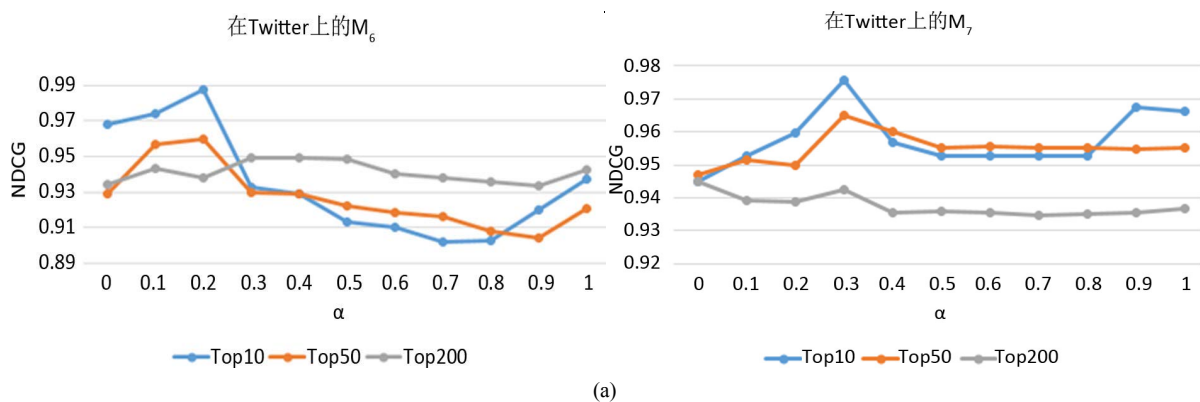
**Table 2.** NDCG values of  $M_1 \sim M_7$  in Twitter dataset top10, top50, top200  
**表 2.**  $M_1 \sim M_7$  在 Twitter 数据集 top10、top50、top200 的 NDCG 值

Twitter			
TopK	10	TopK	10
$M_1$	0.9742	0.9565	0.9360
$M_2$	0.9767	0.9484	0.9371
$M_3$	0.9826	0.9595	<b>0.9479</b>
$M_4$	0.9772	0.9475	0.9259
$M_5$	0.9781	0.9627	0.9452
$M_6$	<b>0.9875</b>	0.9599	0.9378
$M_7$	0.9758	<b>0.9649</b>	0.9424

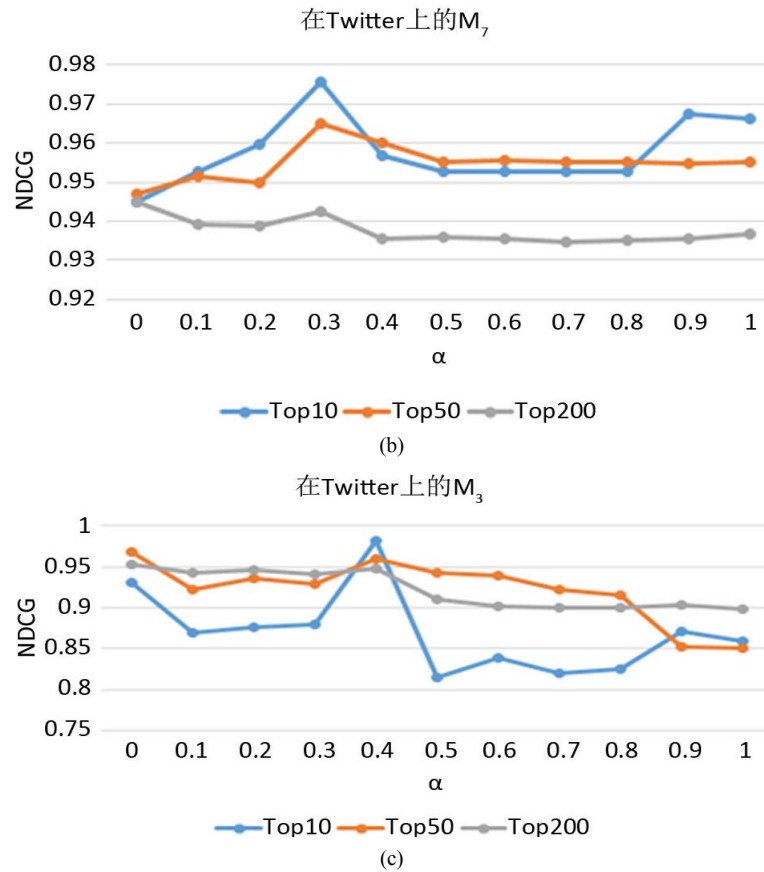
在公式(8)中，我们用参数  $\alpha$  来控制基于边的直接关系和基于社交网络三角关系的 3-节点简单 Motif 的高阶关系之间的平衡。当  $\alpha = 0$  时，它仅使用高阶关系进行计算。当  $\alpha = 1$  时，这意味着我们仅将基于边的直接关系进行计算。在本节中，我们将说明此参数如何影响 NDCG 的准确性。

为简单起见，我们分别从表 2 中取  $K = 10$ ， $K = 50$  和  $K = 200$  NDCG 值最大时所指向的 3-节点简单 Motif，即  $K = 10$  时使用  $M_6$  进行实验， $K = 50$  时使用  $M_7$  进行实验和  $K = 200$  时使用  $M_3$  进行实验。我们可以看到，在大多数情况下，趋势是一致的，并且在  $[0, 1]$  中的某个值上可以获得最佳准确度。这意味着通过融合基于边的直接关系和基于社交网络三角关系的 3-节点简单 Motif 的高阶关系，NDCG 准确性会更高。Twitter 上  $M_7$  的  $K = 200$  排序结果和  $M_3$  上的  $K = 50$  和  $K = 200$  排序结果的最佳准确性是在  $\alpha = 0$  时实现的，即仅使用基于社交网络三角关系的 3-节点简单 Motif 的高阶关系。这再次表明，基于社交网络三角关系的 3-节点简单 Motif 的高阶关系可以为社交网络用户影响力排序提供有用的信息。

由图 10 所示的  $M_6$ 、 $M_7$  和  $M_3$  在  $K = 50$  和  $K = 200$  排序结果曲线趋势相似，而 top10 则更加多样化，这可能是因为在社交网络中的前十名用户非常突出，并且行为可能与其他用户不同。在与 BLR 和 WLR 进行对比，可以计算得到 Twitter 上  $M_6$  在 top10 方面效果提高了 2%，在 top50 方面效果提高了 1.7%，top200 方面效果提高了 1.3%；Twitter 上  $M_7$  在 top10 方面效果提高了 0.8%，在 top50 方面效果提高了 2.2%，top200 方面效果提高了 1.8%；Twitter 上  $M_3$  在 top10 方面效果提高了 1.5%，在 top50 方面效果提高了 1.7%，top200 方面效果提高了 2.4%。由此表明，将基于社交网络三角关系的 3-节点简单 Motif 的高阶关系与基于边的直接关系线性组合到 LeaderRank 可以提高社交网络用户影响力排序的准确性。



(a)



**Figure 10.** Parameter analysis of  $\alpha$  on the three datasets. (a) Running alpha parameter analysis of  $M_6$  on Twitter; (b) Running alpha parameter analysis of  $M_7$  on Twitter; (c) Running alpha parameter analysis of  $M_3$  on Twitter

**图 10.** 在 Twitter 数据集上  $\alpha$  参数分析。(a) 在 Twitter 上运行  $M_6$  的  $\alpha$  参数分析; (b) 在 Twitter 上运行  $M_7$  的  $\alpha$  参数分析; (c) 在 Twitter 上运行  $M_3$  的  $\alpha$  参数分析

### 3) 案例分析

为了显示 MLR 算法的准确性,我们从社区序列里提取另一个社区,提取前 10 个用户并计算其 NDCG 值的平均值。结果如表 3 所示,在基于  $M_1 \sim M_7$  的网络 Motif 上运行 MLR 算法,所得到的前 10 名用户的 NDCG 值的平均值。

**Table 3.** The average of the NDCG values of the top 10 users

**表 3.** 前 10 名用户 NDCG 值的平均值

Twitter	
	Average
BLR	0.9412
$M_1$	0.9632
$M_2$	0.9558
$M_3$	0.9768
$M_4$	0.9582
$M_5$	0.9647
$M_6$	0.9783
$M_7$	0.9604

从表 3 中可以看出, 在 Twitter 上, MLR 的 NDCG 值的平均得分大于 BLR 的 NDCG 值的平均得分, 这意味着 MLR 可以将最相关的结果排在前面, 可以选择更有影响力的作者或可信赖的用户。这展现了基于社交网络三角关系的 3-节点简单 Motif 的高阶关系合并到 LeaderRank 算法中的准确性。因此, 它表明 MLR 可以更合理对有影响的用户进行排序, 再次显示了在社交网络用户影响力排序时将基于社交网络三角关系的 3-节点简单 Motif 的高阶关系合并到 LeaderRank 算法中的优势。

## 5. 结束语

在本文中, 我们提出 MLR 算法, 将基于社交网络三角关系的 3-节点简单 Motif 的高阶关系与基于边的关系相结合代入到 LeaderRank 中进行计算。使用基于 Motif 的邻接矩阵来重新权衡社交网络中边的链接。然后, 将基于社交网络三角关系的 3-节点简单 Motif 的高阶关系和基于边的关系组合在一起执行 LeaderRank。我们在 Twitter 数据集上进行实验, 通过归一化折损累积增益(NDCG), 表明基于社交网络三角关系的 3-节点简单 Motif 的高阶关系和基于边的关系组合在一起的 LeaderRank 算法不仅在社交网络用户影响力排序上的可行性, 而且可以显著提高社交网络用户影响力排序的准确性。

## 参考文献

- [1] Song, X., Chi, Y., Hino, K., et al. (2007) Identifying Opinion Leaders in the Blogosphere. *Proceedings of the 16th ACM Conference on Information and Knowledge Management, CIKM 2007*, Lisbon, 6-10 November 2007, 971-974. <https://doi.org/10.1145/1321440.1321588>
- [2] Tang, J., Sun, J., Wang, C., et al. (2009) Social Influence Analysis in Large-Scale Networks. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, 28 June-1 July 2009, 807-816. <https://doi.org/10.1145/1557019.1557108>
- [3] Xiang, B., Liu, Q., Chen, E., et al. (2013) PageRank with Priors: An Influence Propagation Perspective. *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, AAAI Press, Palo Alto, 2740-2746.
- [4] Wang, Y., Wang, X., Tang, J., et al. (2015) Modeling Status Theory in Trust Prediction. *AAAI*, 1875-1881.
- [5] Page, L., Brin, S., Motwani, R. and Winograd, T. (1999) The Pagerank Citation Ranking: Bringing Order to the Web. *Proceedings of the 7th International World Wide Web Conference*, Brisbane, 1998, 161-172.
- [6] Lü, L.Y., Zhang, Y.-C., Yeung, C.H. and Zhou, T. (2011) Leaders in Social Networks, the Delicious Case. *PLoS ONE*, 6, e21202. <https://doi.org/10.1371/journal.pone.0021202>
- [7] Jiang, X., Sun, X. and Zhuge, H. (2013) Graph-Based Algorithms for Ranking Researchers: Not All Swans Are White! *Scientometrics*, 96, 743-759. <https://doi.org/10.1007/s11192-012-0943-y>
- [8] 邓启平, 王小梅. 利用 LeaderRank 识别有影响力的作者[J]. 现代图书情报技术, 2015, 31(9): 60-67.
- [9] 徐郡明, 朱福喜, 刘世超, 等. 改进 LeaderRank 算法的意见领袖挖掘[J]. 计算机工程与应用, 2015(1): 110-114.
- [10] Zhang, Z.H., Jiang, G.P., Song, Y.R., et al. (2017) An Improved Weighted LeaderRank Algorithm for Identifying Influential Spreaders in Complex Networks. An Improved Weighted LeaderRank Algorithm for Identifying Influential Spreaders in Complex Networks. *IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, Guangzhou, 21-24 July 2017. <https://doi.org/10.1109/CSE-EUC.2017.146>
- [11] Bakshy, E., Hofman, J.M., Mason, W.A., et al. (2015) Everyone's an Influencer: Quantifying Influence on Twitter. *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*, Hong Kong, 9-12 February 2011.
- [12] 顾亦然, 朱梓嫣. 基于 LeaderRank 和节点相似度的复杂网络重要节点排序算法[J]. 电子科技大学学报, 2017, 46(2): 441-448.
- [13] Watts, D.J. and Strogatz, S.H. (1998) Collective Dynamics of Small World Networks. *Nature*, 393, 440-442. <https://doi.org/10.1038/30918>
- [14] Sen, P., Dasgupta, S., Chatterjee, A., et al. (2003) Small-World Properties of the Indian Railway Network. *Physical Review E*, 67, 036106. <https://doi.org/10.1103/PhysRevE.67.036106>
- [15] Barabási, A.-L. and Albert, R. (1999) Emergence of Scaling in Random Networks. *Science*, 286, 509-512. <https://doi.org/10.1126/science.286.5439.509>

- 
- [16] Barabási, A.-L., Albert, R. and Jeong, H. (2000) Scale-Free Characteristics of Random Networks: The Topology of the World-Wide Web. *Physica A*, **281**, 69-77. [https://doi.org/10.1016/S0378-4371\(00\)00018-2](https://doi.org/10.1016/S0378-4371(00)00018-2)
- [17] Milo, R., Shen-Orr, S., Itzkovitz, S., *et al.* (2002) Network Motifs: Simple Building Blocks of Complex Networks. *Science*, **298**, 824-827. <https://doi.org/10.1126/science.298.5594.824>
- [18] Ugander, J., Backstrom, L. and Kleinberg, J.M. (2013) Subgraph Frequencies: Mapping the Empirical and Extremal Geography of Large Graph Collections. *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro, May 2013, 1307-1318. <https://doi.org/10.1145/2488388.2488502>
- [19] Rotabi, R., Kamath, K., Kleinberg, J., *et al.* (2017) Detecting Strong Ties Using Network Motifs. *Proceedings of the 26th International Conference on World Wide Web*, Perth, 3-7 April 2017, 983-992. <https://doi.org/10.1145/3041021.3055139>
- [20] Wang, P., Lü, J.H. and Yu, X. (2014) Identification of Important Nodes in Directed Biological Networks: A Network Motif Approach. *PLoS ONE*, **9**, e106132. <https://doi.org/10.1371/journal.pone.0106132>
- [21] Nataša, P. (2007) Biological Network Comparison Using Graphlet Degree Distribution. *Bioinformatics*, **23**, e177-e183.
- [22] Ahmed, N.K., Neville, J., Rossi, R.A., *et al.* (2015) Efficient Graphlet Counting for Large Networks. 2015 IEEE International Conference on Data Mining (ICDM). 2015 *IEEE International Conference on Data Mining*, Atlantic City, 14-17 November 2015. <https://doi.org/10.1109/ICDM.2015.141>
- [23] Jha, M., Seshadhri, C. and Pinar, A. (2014) Path Sampling: A Fast and Provable Method for Estimating 4-Vertex Subgraph Counts. *Proceedings of the 24th International Conference on World Wide Web*, Florence, 18-22 May 2015, 495-505. <https://doi.org/10.1145/2736277.2741101>
- [24] Han, G. and Sethu, H. (2016) Waddling Random Walk: Fast and Accurate Mining of Motif Statistics in Large Graphs. 2016 *IEEE 16th International Conference on Data Mining (ICDM)*, Barcelona, 12-15 December 2016. <https://doi.org/10.1109/ICDM.2016.0029>
- [25] Benson, A.R., Gleich, D.F. and Leskovec, J. (2016) Higher-Order Organization of Complex Networks. *Science*, **353**, 163-166. <https://doi.org/10.1126/science.aad9029>
- [26] 许平华, 胡文斌, 邱振宇, 等. 节点不对称转移概率的网络社区发现算法[J]. 软件学报, 2019, 30(12): 3829-3845.