

# 基于加权内容-结构网络和 随机游走的社团划分算法

牛新征<sup>1</sup>, 牛嘉郡<sup>2</sup>, 苏大壮<sup>3</sup>, 余 堃<sup>4</sup>

(1. 电子科技大学计算机科学与工程学院, 四川成都 611731; 2. 电子科技大学计算机科学与工程学院, 四川成都 611731;  
3. 大众点评网, 上海 200050; 4. 电子科技大学信息与软件工程学院, 四川成都 611731)

**摘 要:** 针对传统模块优化社团划分算法仅能利用网络的结构信息, 而无法利用同样丰富的内容信息, 导致划分精度较低的问题, 提出一种结合内容属性并通过给连边加权来全面优化网络拓扑结构的社团划分算法 CCSRW (Classification with Content-Structure and Random Walk). 设计利用随机游走理论计算结构节点与内容节点间的相似性关系矩阵, 并将结构节点映射到内容属性空间上, 最终把社团划分问题转化为多维无监督聚类问题. 通过在真实数据集上进行的全面实验分析, 展示了相比于传统社团划分算法, 本文的算法能更准确的描述网络结构, 显著提高划分性能, 并有效解决小社团不敏感问题, 更适用于大规模复杂信息网络的社团划分.

**关键词:** 社团划分; 加权内容-结构网络; 随机游走; 模块优化

**中图分类号:** TP391      **文献标识码:** A      **文章编号:** 0372-2112 (2017)09-2135-08

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2017.09.012

## Community Detection Based on Weighted Content-Structural Network and Random Walks

NIU Xin-zheng<sup>1</sup>, NIU Jia-jun<sup>2</sup>, SU Da-zhuang<sup>3</sup>, SHE Kun<sup>4</sup>

(1. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China;  
2. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China;  
3. Dianping, Shanghai 200050, China;  
4. School of Software Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China)

**Abstract:** For the traditional module optimization community partition algorithms can only use the structure information of network, and cannot use the rich content information, leading to low precision problem. A community partition algorithm that is combined with the content attribute and empowers the edge to fully optimize the topology of the network, called CCSRW (Classification with Content-Structure and Random Walk) is proposed. We use random walk theory to calculate the similarity relationship matrix between structure nodes and content nodes, and map structure nodes onto the content attribute space, finally divide the community partition problems into multidimensional unsupervised clustering problems. Comprehensive experimental analysis on the real data sets shows that compared to the traditional community partition algorithms, this algorithm can describe the network structure more accurately, improve the classification performance significantly, and solve the problem that is not sensitive to small community effectively, and it is more suitable for the large-scale complex information network community partition.

**Key words:** community detection; weighted content-structure network; random walking; modularity optimization.

收稿日期: 2016-04-11; 修回日期: 2016-07-31; 责任编辑: 梅志强

基金项目: 国家科技支撑计划 (No. 2013BAH33F02); 国家自然科学基金 (No. 61300192); 中央高校基本科研业务费电子科技大学项目 (No. ZYGX2014J052); 2015 年省科技厅支持计划 (No. 2015GZ0102); 四川省自贡市公安局-基于智能视频分析的交通流量监控与事故预测系统的研究与实现; 四川省公安厅科研项目 (No. 2015SCYYCX06); 成都市科学技术局软科学研究项目 (No. 2015-RK00-00247-ZF)

## 1 引言

在关于网络拓扑结构的科学研究中,为进一步理解网络的构成和演变,社团划分问题得到越来越多的关注<sup>[1]</sup>. 社团划分是根据网络的内部功能属性,将其中具有相似特征的节点划分到同一社团,使得社团内部节点间联系紧密,外部节点间联系稀疏.

大部分社团划分算法仅根据结构是否紧凑划分社团<sup>[2]</sup>. 随着互联网的发展,人们积累了网络文档、博客等形式的大量信息,从而产生了很多与传统网络大不相同的,新兴的社交网络和信息网络,具有更丰富的内容信息. 在对具有内容属性的网络进行划分时,节点的内容相似,意味着它们所代表的个体具有相似的社会行为,即使不属于同一个密集区域也应被划分到同个社团. 因此不能只考虑单一的结构信息,而应该同时根据结构和内容双重属性进行社团划分,才能保证划分结果的全面性和有效性.

目前社团划分的研究可分为模块优化和非模块优化两大类. 非模块优化方法未能将社团划分问题从 NP 难题中简化出来,因此模块优化算法是当前最流行的社团划分算法<sup>[4]</sup>. Newman 和 Girvan 教授在文献<sup>[5]</sup>中利用社团结构的模块化系数作为判断标准,并认为社团内部边密度较高而社团之间边密度较低. Clauset<sup>[6]</sup>在 Newman 的研究基础上提出使用最大堆、模块度增量矩阵等数据结构存储数据,有效缩减了时间复杂度. 但其方法过分偏重大规模社团,导致在实际应用中容易陷入最坏的时间复杂度. Schuetz 在文献<sup>[7]</sup>中对 Clauset 的算法做了进一步改进,每次同时融合多个社团,有效改善了大规模社团的情况. 文献<sup>[8]</sup>的算法基于贪心策略,使初始社团的构建始于一个具有最高强度的节点,然后通过不断的添加节点来扩展社团. 但以上基于模块优化的方法并不能准确评价社团结构,尤其对小社团不敏感,即算法无法判断一个子网应该是一个大社团还是多个小社团组合而成. 并且忽视了连接社团之间的边所包含的信息,只根据网络节点间的连接关系判断紧密性,导致其在实际应用中效果并不理想. 从信息论的角度来看,以模块方法来表示的复杂结构所含有的信息熵并不足以准确判断网络结构.

由于模块优化算法有着以上缺陷,本文提出一种社团划分新思路,主要有以下四个方面的贡献:

(1) 采用改进的加权内容-结构网络模型有效融合内容和结构信息,能真实体现节点间关系强弱的同时,避免了建立复杂的统一模型.

(2) 基于加权内容-结构网络模型,改进传统的随机游走算法使其充分利用网络的边权信息,更精确地计算节点间的相似度,创新性地提出了以网络中的内

容节点表征结构节点的社团划分新思路.

(3) 利用网络中节点的相对关系来计算其绝对属性,通过多个属性的组合将其映射到多维坐标空间中,并用较成熟的聚类技术解决社团划分问题,克服了模块化算法低准确度的同时,避免了非模块化方法高复杂度的缺陷.

(4) 从多个角度不同层面构建实验,与同类型的多个算法进行对比,验证了本文社团划分的准确性和处理大规模数据的有效性. 同时设计实验获得令聚类效果最优的转移步数,并用主成分分析得到步数对聚类的重要影响.

## 2 算法原理

### 2.1 加权内容-结构网络模型

本文提出一种能够有效融合内容和结构信息的加权网络模型.

**定义 1 加权内容-结构网络模型:** 如图 1 所示,将网络划分成一个具有二分类的偶图,包含两类节点,一是结构节点,该部分节点由原始网络节点组成,且有边相连,设为  $N$ ; 二是内容节点,由所有节点的内容属性去重后得到,设为  $M$ . 结构节点之间的连边,结构节点和内容节点间的连边都赋有权值,用来区分节点间相互作用的强弱程度.

通过这样的部署,能真实反映现实特征的同时又摒弃了在社团划分过程中内容节点间可能存在的毫无意义的间接结构信息.

**定义 2 结构节点间连边的权重  $W_s$ :** 结构节点之间交互的频次.

**定义 3 结构节点与内容节点间连边的权重  $W_c$ :** 结构节点中出现该内容属性的频次.

我们将构建加权内容-结构网络分成三步:

(1) 若结构节点  $n_i$  和  $n_j$  在原始网络中相连,则在加权内容-结构网络中,两节点依然相连. 根据两节点间交互的频次,设置连边的权重. 结构节点之间的边集合用实线表示,权重为  $W_{sij}$ .

(2) 若结构节点  $n_i$  存在内容属性  $p$ ,则在内容-结构网络中,结构节点  $n_i$  与表示该内容属性的内容节点  $m_p$  相连. 根据内容属性  $p$  在节点  $n_i$  中出现的次数,设置内容节点  $m_p$  和结构节点  $n_i$  之间连边的权重. 连接结构节点和内容节点的边为虚线,权重为  $W_{cip}$ .

(3) 由于不同连边的权重差异较大,为了方便处理,需要对权重进行数据规范化处理. 本文采用线性比例变换法对数据进行标准化,设原值为  $W_{ij}$ ,标准化后的值为  $a_{ij}$ ,则:

$$a_{ij} = \frac{W_{ij}}{\max W_{ij}} (1 \leq i \leq n, 1 \leq j \leq m) \quad (1)$$

经过标准化处理后的权重均在 0 ~ 1 之间,最大值为 1,最小值为 0.

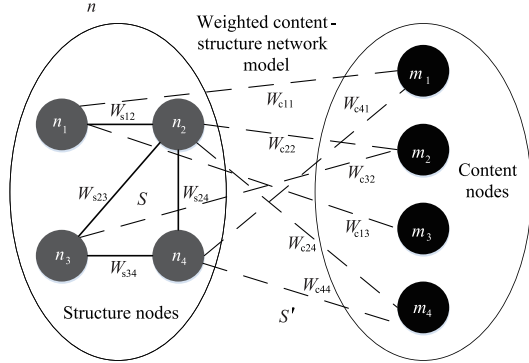


图1 加权内容-结构网络结构图

图 1 左边的结构节点集  $N = \{n_1, n_2, n_3, n_4\}$  和右边的内容节点集  $M = \{m_1, m_2, m_3, m_4\}$  共同构成了内容-结构网络的节点集合,由实线表示的边为结构节点之间的边  $S$ ,由虚线表示的边为连接内容节点和结构节点的边  $S'$ ,  $W_s$  为结构节点之间连边的权重,  $W_c$  为结构节点和内容节点之间连边的权重.

**定义 4 边权属性** 设内容-结构网络的关联矩阵为  $A$ ,  $A$  为带权无向矩阵,大小为  $n \times n$ ,设  $A_{ij} = a_{ij}$ ,则第  $i$  个节点和第  $j$  个节点相连,连边权重经标准化处理之后为  $a_{ij}$ .若  $A_{ij} = 0$ ,则两个节点未相连,由此可用新的加权内容-结构网络的对应矩阵来表示一个图的边权属性,如式(2)所示.

$$A_{ij} = \begin{cases} a_{ij}, & 1 < i, j < n, i \neq j \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

### 2.2 随机游走计算网络节点相似性

网络中的随机游走是指从某一节点出发,按照一定概率转移到邻居节点的过程.本文创新性地提出了以网络中的内容节点来表征结构节点的社团划分新思路:以结构节点为起点,内容节点为终点,通过加权内容-结构网络中的随机游走,分别计算结构节点与每个内容节点间的相似度.该相似度组合包含了每一个结构节点和内容节点之间的连接关系,并共同构成了结构节点在不同内容节点维度上的映射.最终实现以结构节点与各个内容节点的相似度组合,将结构节点映射到以内容节点为基准的多维空间中的点,并将社团划分问题转换成较为简单的多维空间的点聚类问题.

给定一个网络  $G = (V, E)$ ,其边权矩阵  $A = (a_{ij})_{n \times n}$ ,其中  $a_{ij}$  表示节点  $i$  和  $j$  边的权重,若  $i$  和  $j$  无连接则为 0.根据随机游走模型,可作出以下定义:

**定义 5 加权网络中的转移概率** 节点  $i$  向节点  $j$  的转移概率为  $P_{ij} = (a_{ij})/d(i)$ ,其中  $d(i)$  为包含节点  $i$  的边的权重之和,即

$$d(i) = \sum_{k=1}^n a_{ik} \quad (3)$$

通过迭代计算,可求得节点  $i$  经过  $t$  步到达  $j$  的转移概率  $P_{ij}^t$ .根据 Pons 的相关结论,最终求得节点  $i$  与  $j$  在整个网络层面上的相似度如式(4)所示:

$$r_{ij} = \sqrt{\frac{\sum_{k=1}^n (P_{ik}^t - P_{jk}^t)^2}{d(k)}} \quad (4)$$

对于网络  $G$ ,以及其权重矩阵  $A$ ,根据式(4)就能很容易计算出结构节点和内容节点间的相似度.

### 2.3 平衡节点结构及内容的社团划分

我们结合内容-结构网络模型和随机游走理论,通过计算结构和内容节点之间的相似性并将其设置为该节点在内容维度上的属性值,最终将结构节点映射为多维空间中的一个点,如图 2 所示.

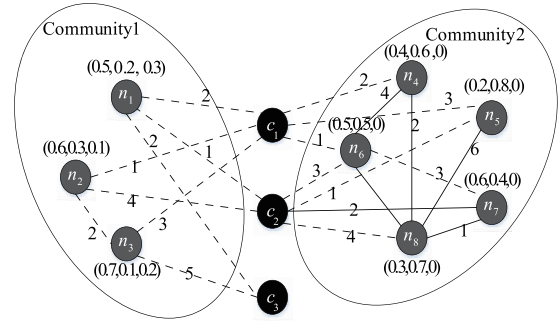


图2 结构节点在内容节点上的映射举例

图 2 中,原始网络被构建成内容-结构网络,包含结构节点集  $N = \{n_1, n_2, n_3, \dots, n_8\}$  和内容节点集  $C = \{c_1, c_2, c_3\}$ .图中的实线表示结构节点之间的边  $S$ ,虚线为连接内容节点和结构节点的边  $S'$ ,边上的数值表示边的权重.利用 2.2 节所述的随机游走模型,即可计算出结构节点与所有内容节点间的相关性,该过程的算法伪代码如算法 1 所示:

#### 算法 1 网络构建及计算相似性算法伪代码

```

输入:原始网络  $N$ ,内容信息  $C$ 
输出:内容-结构关系矩阵  $M$ 
//构建新网络模型
for( $n \leftarrow 0$  to  $N$ )
  for( $c \leftarrow 0$  to  $C$ )
    if( $n$ .hasProp( $c$ )) then  $n$ .link( $c$ )//连接结构和内容节点
     $N$ .add( $C$ ) //在网络  $N$  中添加属性节点
//计算结构节点( $0 \sim N$ )和属性节点( $N \sim N + C$ )之间的关系
Matrix  $M(N$ .num, $C$ .num) $\leftarrow$ new Matrix()
for( $i \leftarrow 0$  to  $N$ .num)
  for( $j \leftarrow 0$  to  $C$ .num)
    //计算相似度
     $M(i,j) \leftarrow$ calDistance( $N$ .getNode( $i$ ), $N$ .getNode( $j + N$ .num))
return  $M$ 
    
```

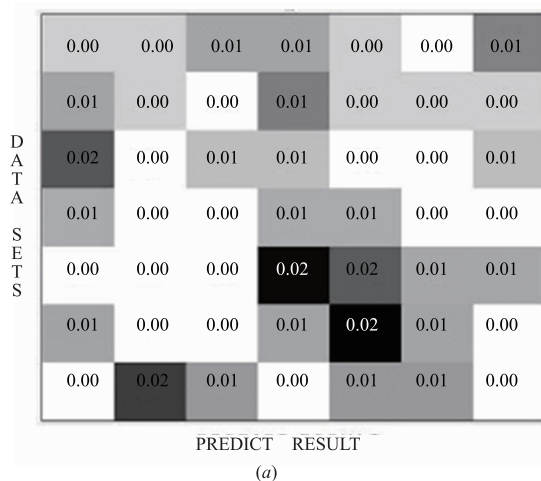
该算法返回的内容-结构关系矩阵  $M$  中存储的为结构节点在内容维度的映射信息,其规模为  $\text{num}(N) * \text{num}(C)$ . 得到了内容-结构矩阵之后,社团划分问题转化成多维无监督聚类问题. 现关于无监督聚类的算法很多,本文仅以经典的 k-means 测试.

### 3 实验

在本节中,我们将设计 4 个实验对本文提出的 CCSRW 算法进行分析评估. 分别是(1)与传统算法的对比实验,(2)与 DYCOS 算法的对比实验,(3)步数对聚类效果的影响分析实验,(4)主成分分析实验. 算法的准确性主要由使用真实数据的聚类准确度来度量. 对于算法的实现,本文选择开源网络分析工具 igraph 和 Orange 工具简化操作流程.

#### 3.1 CORA 实验数据

在评估实验中,本文使用 CORA<sup>[9]</sup> 真实数据集. 包括了多篇论文的引用情况,有 24519 个节点,94589 条边,79 个不同的标签. 每个节点代表一篇论文,每条边描述两篇论文的引用情况,标签则表示论文的内容分类信息. 为符合小世界网络<sup>[10]</sup> 特性,避免因抽取节点少而与原始网络相差大. 本文的所有实验都会抽取不同数量的节点进行测试,研究节点数目对社团划分结果会造成怎样的影响,并对比本算法与其他算法在网络规模增大时的不同表现.



#### 3.2 与传统方法的对比实验

首先,进行数据预处理之后,在完整网络中使用传统方法(随机游走方法和模块优化方法)对原始数据进行社团划分,在分类结果中按照社团的比例随机抽取一定数量的节点,形成新的网络. 其次再次进行两个并行的社团划分:(1)同样使用传统社团划分算法,(2)使用 CCSRW 算法进行社团划分. 最后根据两次结果计算混淆矩阵,与第一次社团划分结果对比,就能判断在信息缺失的情况下,所采用的算法是否能够依然保证结果的准确率.

##### 3.2.1 性能评价指标

混淆矩阵是一种用于监督学习的可视化工具,可直观的分析正确预测和错误预测的实例比. 混淆矩阵中每个方格内的比例表示实际数据集和聚类数据集的交叉比例,若任意一个方格的比例要么接近 1,要么接近 0,则说明该聚类算法的结果与实际数据集较为接近,效果更好.

NMI 指数是基于信息论的量化分类相似性指标,用来计算网络划分的互信息,可以衡量两个数据分布的吻合程度,数值越大表示聚类结果与真实情况越相近.

##### 3.2.2 分类模型的性能评估

图 3~5 分别是提取节点数目为 800,1600,2400 的分类对比,图(a)为第二次社团划分依然采用原始方法的混淆矩阵,图(b)为采用本文所提出的新方法得到的混淆矩阵.

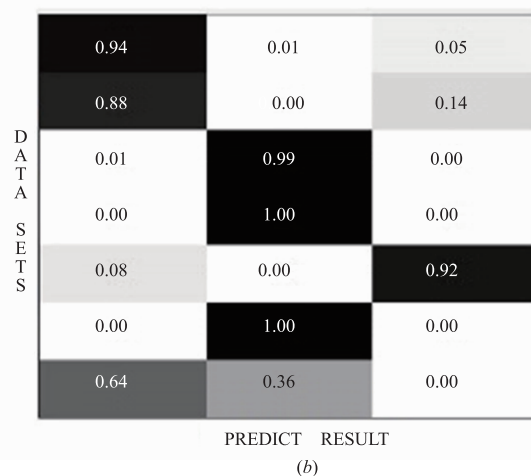


图3 对比随机游走的混淆矩阵 $n=800$

从如图 3~5 可以看出,利用随机游走模型进行社团发现的混淆矩阵较为分散且占比很小,基本不能表示原始数据集的分类信息. 相反利用内容-结构网络得出的混淆矩阵较为完整,虽然存在两个数据集在测试时会被合并为一个数据集的现象,但原始数据集的整体结构并没有被过度破坏,预测结果能够在一定程度上表达原始数据集的信息.

我们将 CCSRW 算法与模块优化算法 fast unfolding 进行对比,如图 6~8 所示. 模块优化同随机游走一样无法较为完整的体现出网络的社团结构,而 CCSRW 方法较高的保证原始数据的完整性.

图 9(a)将抽取节点数目在 1000~10000 范围内的混淆矩阵转化成了 NMI 曲线,并与传统随机游走和模块优化算法进行对比. 在抽取节点数目较多时(如

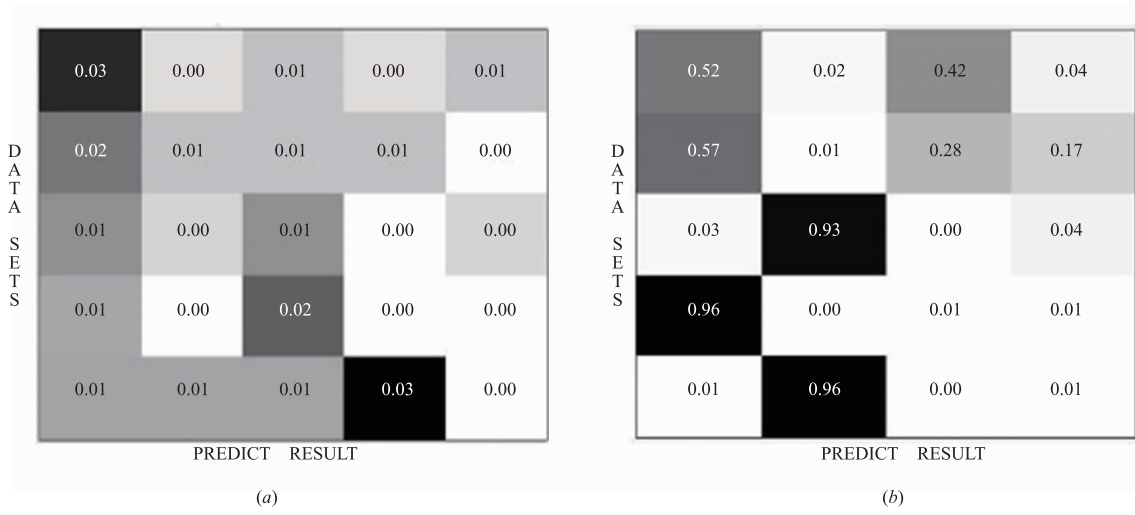


图4 对比随机游走的混淆矩阵 $n=1600$

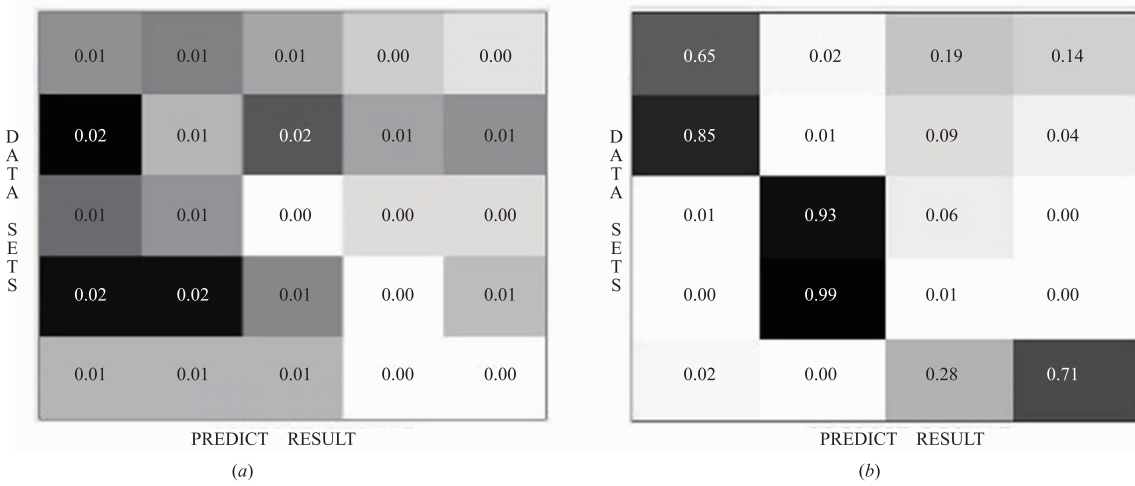


图5 对比随机游走的混淆矩阵 $n=2400$

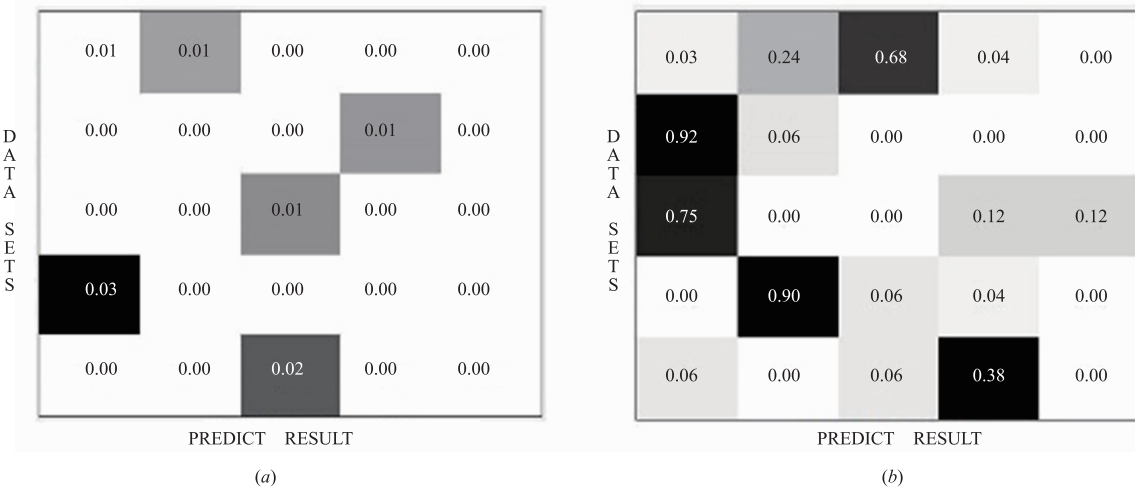


图6 对比模块优化的混淆矩阵 $n=800$

10000 个), NMI 指数较高,两种方法均达到了 0.9 以上,随着抽取数目减少(损失信息变多),仅依据结构进

行社团发现的随机游走理论 NMI 下降到了 0.3 左右, 很难表达原始数据的信息. 而本文的社团划分算法在

节点数目仅有 1000 时, 依然可以保证 0.7 的准确度.

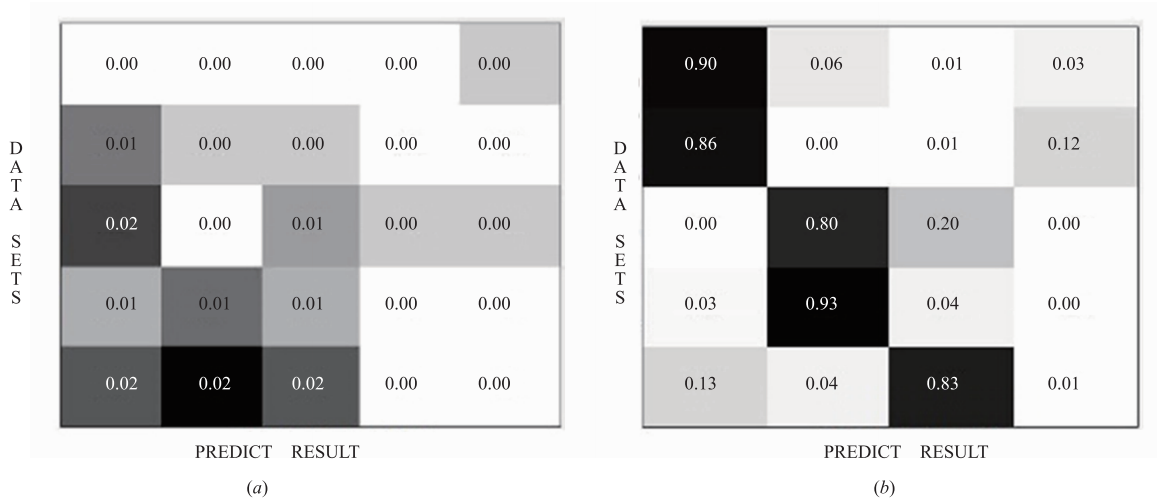


图7 对比模块优化的混淆矩阵n=1600

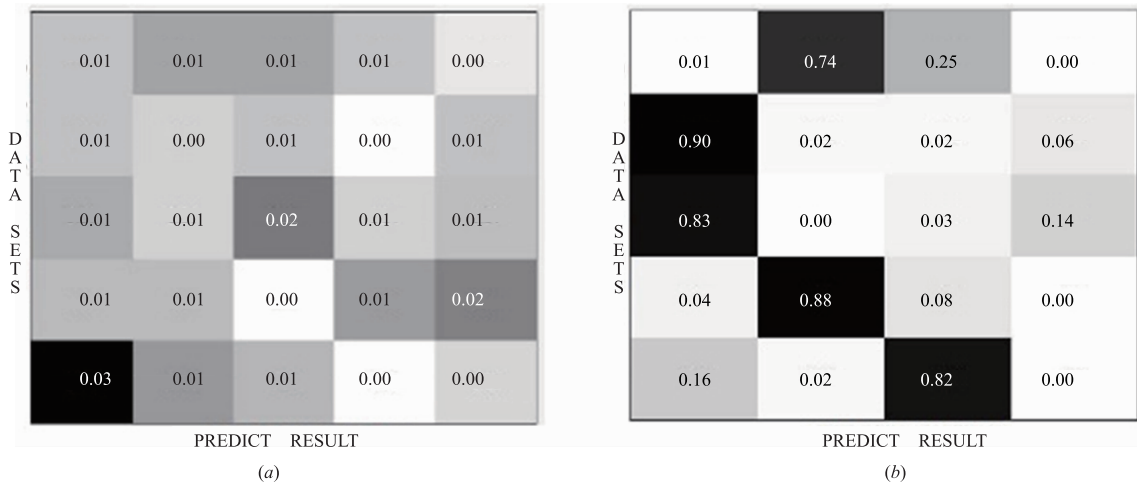


图8 对比模块优化的混淆矩阵n=2400

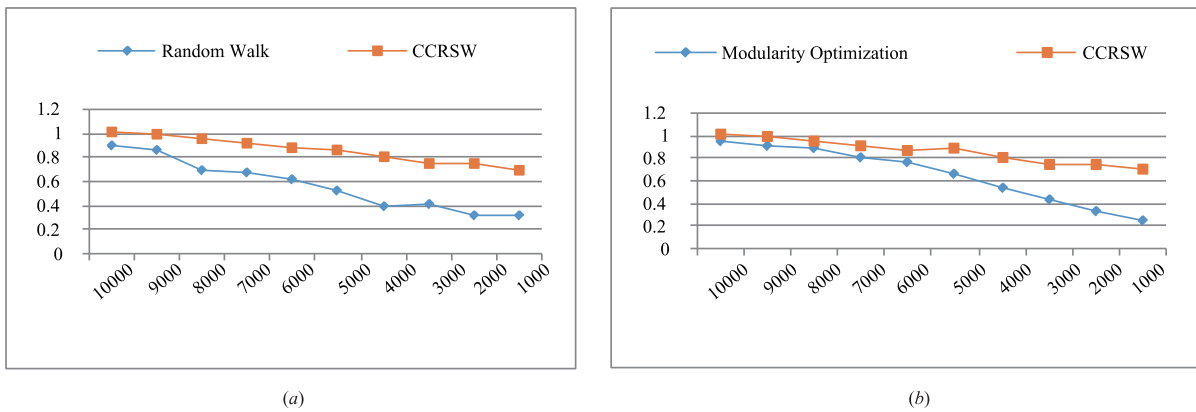


图9 NMI指数对比

根据图 9(b), CCRSW 和模块优化在抽取节点数目较多时, 能保证原始数据的完整性. 但是在抽取节点数

目降低到 1000 时, 内容-结构方法保持在了 0.7 左右, 而模块优化的 NMI 指数仅比 0.2 略高.

### 3.3 与 DYCOS 算法的对比实验

本算法与基于内容-结构网络模型的节点划分算法 DYCOS<sup>[3]</sup> (DYnamic Classification algorithm with Content and Structure) 进行分类准确度的对比, 选用 CORA 中七组不同时间段递增的数据进行测试. 对比结果如图 10 所示. 在每个时段 CCRSW 的准确率都比 DYCOS 高, 至少高出 10.13%. 而随着不同时段数据量的递增, DYCOS 的分类准确率有所波动, 而 CCRSW 的准确率较为稳定. 说明对于大规模数据, CCRSW 的分类性能优于 DYCOS.

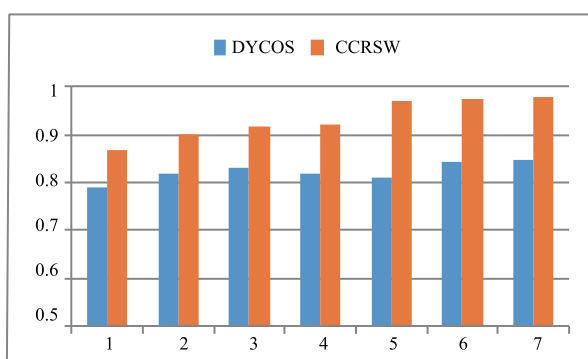


图10 CCRSW与DYCOS的分类准确率对比图

### 3.4 步数对聚类效果的影响分析

在本文选用的网络模型中, 我们利用了基于内容-结构网络的思想<sup>[3]</sup>, 原文为了避免处理大规模矩阵相乘, 仅使用了一次转移矩阵, 但实际上, 一个用户节点在某一内容节点上的维度可能会经过几个用户节点的中转, 因此, 我们对比了新算法 CCRSW 在不同步数下的聚类效果.

图 11(a) ~ (f) 为转移矩阵步数在 1 ~ 6 情况下不同颜色的线性投影, 图中逆时针 0 ~ 0.75 分别代表 75 个不同的列向量值. 从投影位置和聚类结果来看, 在步数为 1 或 2 的时候, 不同的用户节点倾向于部署在一条直线上. 这说明步数较小时, 节点的区分度较低, 在计算判定边界时, 简单的线性函数不能满足聚类要求, 聚类效果也不理想; 在步数为 5 或 6 时, 用户节点不再倾向于成一条直线, 而是大量节点聚集在一小片区域, 其余节点随机的散布在零点和区域周围. 在分类结果上, 通常是节点密集区域分成一到两个社团, 其余散列节点杂乱的分成几个小型社团, 主要特征是数据通常集中在一些边缘节点, 在聚类时, 边缘节点的线性投影被大跨度的散列在投影空间中, 而密集区域从整体角度上又区分度不够, 为了照应边缘节点, 在计算聚类个数评分时, 节点密集区域的结果相应的会做出一些“牺牲”; 而在步数为 3 或 4 时, 可以看到节点投影空间主体呈一个扇形, 相比步数较小时

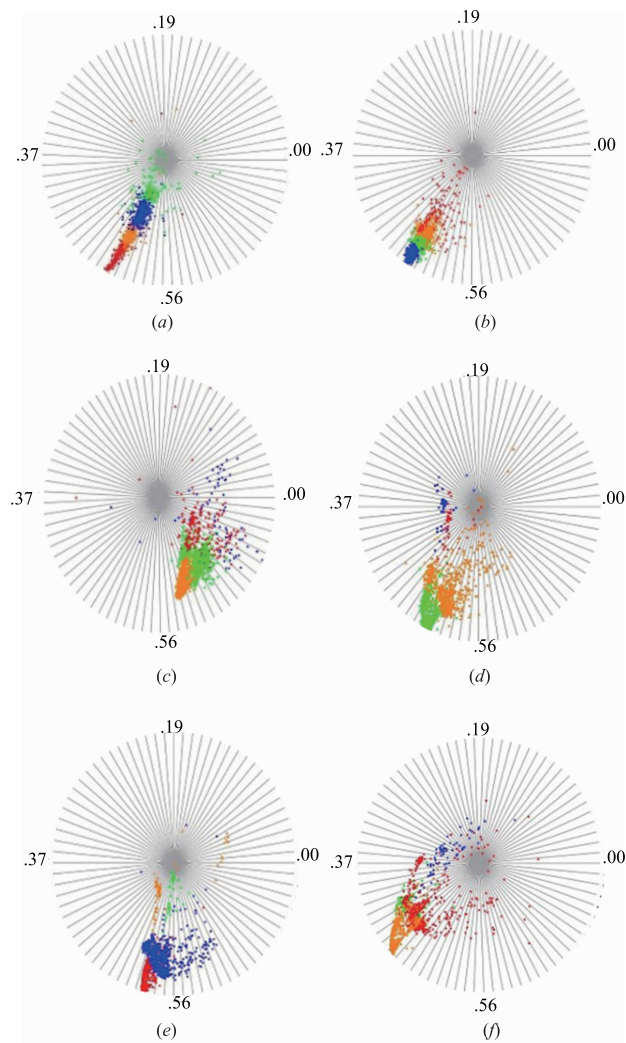


图11 不同步数对聚类结果的影响

的密集集中以及步数较大时的杂乱无章, 3 或 4 可以是步数的较好选择.

### 3.5 主成分分析 (PCA)

随着节点数目增加, CCRSW 有着较强的纠错性, 节点增加时, 出现错误划分节点的情况也越来越少, 能够在一定程度上保证社团的完整性. 但是, 与一般社团划分算法类似, CCRSW 也存在划分小社团的问题. 说明用于分类的列向量区分度还不够. 列向量通过简单处理和除重后, 主成分分析如图 12 所示:

从 PCA 来看, 在累计方差达到 98% 的时候, 占用维数仅有全部维数的三分之一左右, 这说明获取的大量数据并没有全部在聚类时发挥应有的作用, 此外, 单一元素方差的影响下降过快, 在一维时, 单元元素方差接近 80%, 但是第二维的方差仅有 5% 左右, 此后元素方差依次降低, 同样说明用于聚类的列向量各元素之间存在着同质性, 区分度不高.

参考实验 3.4 结论, 同时考虑社交网络中小世界模

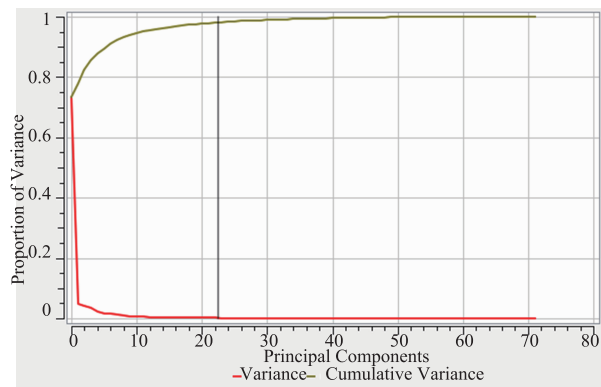


图12 CORE数据集PCA

型理论——网络中任意两个节点的间隔不会超过5个,即步数不会超过6。这也说明在该模型中,提高步数会让整个网络具有较高的连通性,这在一定程度上让聚类降低了维度的区分度。

#### 4 结论

针对基于模块系数优化的传统社团划分算法所具有的缺陷,并考虑到复杂网络内容信息的重要性,本文提出了一种新的社团划分算法 CCRSW,该算法通过在传统网络中加入内容节点,设置连边权重,并利用随机游走模型将原始网络节点映射到内容节点维度,将问题转化成在一个普遍的聚类问题。通过实验证明,虽然该算法依然存在无法判断小社团的问题,但是在信息缺失的情况下表现明显优于基于随机游走和模块优化的社团划分算法,在处理大规模数据时准确率明显高于考虑内容信息的分类算法。

#### 参考文献

- [1] 张健沛,邓琨,杨静,等. 基于边标签传播的复杂网络社区识别方法[J]. 电子学报,2015,43(6):1113-1118.  
Zhang J P, Deng K, Yang J, et al. Community detection in complex networks based on link label propagation[J]. Acta Electronica Sinica, 2015, 43(6): 1113-1118. (in Chinese)
- [2] Silva A, Meira J W, Zaki M J. Mining attribute-structure correlated patterns in large attributed graphs[J]. Proceedings of the VLDB Endowment, 2012, 5(5): 466-477.
- [3] Aggarwal C C, Li N. On node classification in dynamic content-based networks[A]. Eleventh Siam International Conference on Data Mining[C]. Arizona:SDM,2011. 355-366.
- [4] Jin J. Fast community detection by SCORE[J]. The Annals of Statistics, 2015, 43(1): 57-89.
- [5] Newman M E J. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167-256.
- [6] Clauset A, Newman M E J, Moore C. Finding community structure in very large networks[J]. Physical Review E, 2004, 70(6): 264-277.
- [7] Schuetz P, Cafilisch A. Efficient modularity optimization by multistep greedy algorithm and vertex mover refinement[J]. Physical Review E, 2008, 77(4): 357-361.
- [8] Chen D, Shang M, Lv Z, et al. Detecting overlapping communities of weighted networks via a local algorithm[J]. Physica A: Statistical Mechanics and Its Applications, 2010, 389(19): 4177-4187.
- [9] Andrew McCallum. Andrew McCallum Data[DB/OL]. <http://people.cs.umass.edu/~mccallum/data.html>, 2015-11-10.
- [10] 李兵,王浩,李增扬,等. 基于复杂网络的软件复杂性度量研究[J]. 电子学报,2006,34(12A):2371-2375.  
Li B, Wang H, Li Z Y, et al. Software complexity metrics based on complex networks[J]. Acta Electronica Sinica, 2006, 34(12A): 44-56. (in Chinese)

#### 作者简介



牛新征 男,1978 年生于贵州贵阳,电子科技大学计算机科学与工程学院副教授。研究方向为数据挖掘和网络计算。  
E-mail: xinzheniu@uestc.edu.cn

牛嘉郡 女,1994 年生于山西忻州,电子科技大学计算机科学与工程学院研究生在读,研究方向为数据库与数据挖掘。

E-mail: janetniu@126.com

苏大壮 男,1989 年生于辽宁阜新,大众点评网大数据平台工程师,研究方向为数据挖掘和网络计算。

E-mail: 905533181@qq.com

余堃 男,1968 年生于湖北武汉,电子科技大学信息与软件工程学院教授,主要研究领域为网络计算、人工智能。

E-mail: kunshe@126.com