

# 基于结构和适应度的社区发现

高启航, 景丽萍, 于剑, 林友芳

(北京交通大学交通数据分析与挖掘北京市重点实验室, 北京 100044)

**摘要:**复杂社会网络无处不在,对复杂社会网络进行社区发现越来越被人们重视.基于局部结构的社区发现可以在不用了解全局的情况下对某些节点进行划分;社会网络的社区适应度特性可以找出不同适应度下的社区结构.基于局部结构以及社区适应度的网络属性,提出一种新的社区发现算法.通过实验比较,算法能较好、较快的发现社区结构,在人工网络以及真实社会网络均取得较之已有方法更好的效果.

**关键词:**社区发现;结构社区;社区适应度;复杂社会网络;局部社区

**中图分类号:**TP311      **文献标识码:**A      doi:10.3969/j.issn.0253-2778.2014.07.004

**引用格式:** Gao Qihang, Jing Liping, Yu Jian, et al. Community detection based on structure and fitness[J]. Journal of University of Science and Technology of China, 2014,44(7):563-569.  
高启航, 景丽萍, 于剑, 等. 基于结构和适应度的社区发现[J]. 中国科学技术大学学报, 2014,44(7): 563-569.

## Community detection based on structure and fitness

GAO Qihang, JING Liping, YU Jian, LIN Youfang

(Beijing Key Lab of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing 100044, China)

**Abstract:** Many systems can be described as complex social networks, and increasing attention has been paid to the detection of social communities out of complex social networks. Structured-based community detection can be achieved locally without knowledge of the overall situation. The community fitness characteristics of social networks can help to identify community structures at different fitnesses. A new algorithm based on structure and fitness was proposed to test large generated networks and real networks. Experiments had shown its better efficiency and higher accuracy.

**Key words:** community detection; community structure; community fitness; large networks; local community

## 0 引言

社会网络是一种由节点和连接这些节点的边组成的复杂结构.这些节点表示对象,连接节点的边表

示对象之间的关系.这些关系可以是人与人之间的关系,或者生物之间、互联网页面之间的关系等.社区结构是社会网络的一种重要的拓扑属性,通过社区结构可以了解网络的个体组成以及网络的行为;

收稿日期:2014-03-21;修回日期:2014-06-15

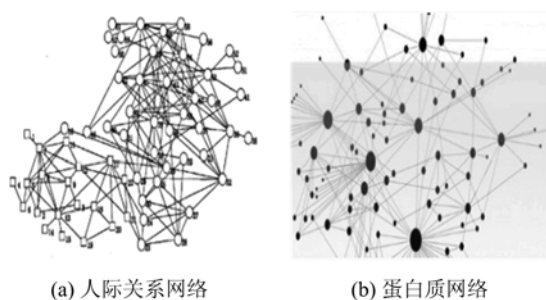
基金项目:中央高校科研业务费专项基金(2014JBM029)资助.

作者简介:高启航,男,1989年生,硕士生.研究方向:数据挖掘,机器学习. E-mail: gao7hang@163.com

通讯作者:景丽萍,博士/副教授. E-mail: lpjing@bjtu.edu.cn

通过分析社区结构,可以发现隐藏在网络表象下的社区实质规律.同时,根据网络拓扑结构特征,也可以掌握网络节点和行为的趋势.大体上来讲,复杂网络要研究的是各种看上去互不相干,但事实上密切关联的形形色色网络之间的共同属性与处理他们的普适性方法<sup>[1]</sup>.

人们已经提出了许多描述网络拓扑的统计特征,包括度节点、聚簇系数、介数、平均路径长度、同配系数等.研究发现了网络上的一些重要性质,比如度分布的幂律特征、网络上的小世界现象、社区结构的存在等<sup>[2]</sup>.目前研究者已经提出了大量的社区结构发现方法<sup>[3-7]</sup>. Girvan 等<sup>[8]</sup>提出了一种发现社区的 GN 算法,该算法已经成为社区结构分析的标准算法,但是该算法在社区数目未知的情况下不知道在哪一步终止.针对网络整体结构的社区发现算法,计算的结果也是网络所有社区划分结果<sup>[9-10]</sup>,这些算法的计算复杂度比较高,对一个  $m$  条边,  $n$  个节点的网络,GN 算法的计算复杂度是  $O(m^2 n)$ <sup>[8]</sup>,文献<sup>[11]</sup>的计算复杂度是  $O(n(m+n))$ ,文献<sup>[12]</sup>的计算复杂度是  $O(n^3)$ .



(a) 人际关系网络

(b) 蛋白质网络

图 1 各种社会网络

Fig. 1 Different social networks

当社会网络在规模很大,甚至更新很频繁的情况下,全局社区发现很难实现;而且若只需要发现社区中指定点(或区域)的社区情况,这种社区发现很耗时,所以局部社区发现技术应运而生.局部社区发现算法可以在局部范围内较快地发现指定点(或区域)所在的局部社区,因此其不需要全局计算所需的高昂时空开销<sup>[13-14]</sup>.其中,Xu 等<sup>[13]</sup>提出的 SCAN 算法侧重于从节点与节点之间的关系出发,将相似性较高的节点放在一起.这种算法可以较快发现具有明显结构特征的社区,缺点是发现的社区规模往往较小,不能充分反映真实社区特征. Martelot 等<sup>[14]</sup>提出的 MSCD 算法从核心节点出发,先产生初始社区,再利用社区适应度,不断将社区外节点加进和移

出.这种算法能解决 SCAN 算法的社区规模问题,但初始节点的选择往往是随意的,不同的初始节点会产生不同的划分结果;同时,社区扩大和缩小的过程都需要进行适应度判断,复杂度极高,不适合节点数和社团数都很多的网络.

本文针对局部社区的思想,提出一种新的社区发现算法,既可以有效地发现真实的社区情况,又可以保证时间复杂度较低.

## 1 已有研究工作

2007 年,Xu 等<sup>[13]</sup>提出了社区网络的结构特征.节点间的结构特征是基于节点的特殊结构,从某个节点出发,分析其邻居节点之间的连接形式,得到一种特殊结构,如图 2 所示.

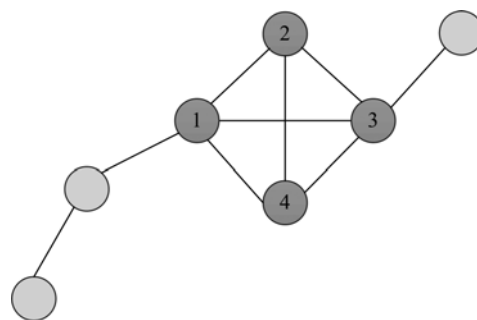


图 2 节点之间的结构特征

Fig. 2 Structural feature between nodes

图 2 中,节点 1 与节点 2 的共同节点是节点 3 和节点 4,其共同邻居节点数(2 个)与其总的邻居节点数(7 个)大于设置阈值,因此将节点 1 和节点 2 放入同一个社区.同理,节点 2 和节点 3 放入同一个社区;节点 3 和节点 4 放入同一个社区……最终,节点 1、2、3、4 都属于一个社区.通过这种划分方式,可以较快得到社区.通过该方法得到社区,能够比较快速地将很有可能在一个社区的节点划在一起,但是所得社区不会很大,因此不能较好地体现社区特征.

2009 年,Lancichinetti 等<sup>[15]</sup>提出一种基于社区适应度(fitness)的社区评价方法,并利用该评价方法提出一种局部社区发现算法.2013 年,Martelot 等<sup>[14]</sup>也通过适应度的方法提出了基于多尺度的 MSCD 算法.一个社区  $c$  的社区适应度如图 3 所示.

图 3 中,节点 5 未加入社区之前,其社区适应度  $f_c$  值为  $6/(6+4)=0.6$ ;节点 5 加入社区之后, $f_c$  值为  $8/(8+2)=0.8$ , $f_c$  值增加,社区更加健壮,所以应该将节点 5 加入社区.

文献<sup>[15]</sup>和 MSCD 算法能解决 SCAN 算法的

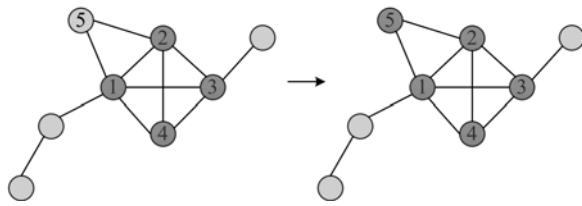


图 3 社区适应度演示图

Fig. 3 Demo of community fitness

社区规模问题,但初始节点的选择往往是随意的,不同的初始节点会产生不同的划分结果;同时算法复杂度极高,不适合节点数和社团数都很多的网络。

社会网络具有多尺度的特征,多尺度可以基于不同空间分辨率进行的数据分析<sup>[16]</sup>,通过网络的多尺度特征可以调节社区的适应度,从而找到不同适应度下的社区划分。

## 2 基于结构和适应度的社区发现

复杂网络中,某些节点之间存在结构特征,比如  $a$  节点与  $b$  节点有多个共同邻居节点,那  $a$  和  $b$  很有可能属于同一个社区.这种基于节点之间的结构特征来进行社区划分的时间复杂度较低,故本文引入社区适应度来划分不同尺度下的社区结构。

设定一个尺度参数,我们基于这个尺度,选择合适的结构参数、适应度参数.前者用来判断核心节点的个数以及社区大小;后者用来判断需要将社区扩大的规模。

确定核心节点之前,我们先确定两个定义:

**定义 2.1**  $\epsilon$  节点. 满足

$$N_{\epsilon}(v) = \left\{ w \in \Gamma(v) \mid \frac{|\Gamma(v) \cap \Gamma(w)|}{\sqrt{|\Gamma(v)| |\Gamma(w)|}} \geq \epsilon \right\} \quad (1)$$

的节点,称  $N_{\epsilon}(v)$  为  $v$  的  $\epsilon$  节点. 其中  $\Gamma(v)$  表示  $v$  的邻居节点集。

**定义 2.2** 核心节点. 当一个节点的  $\epsilon$  节点个数大于  $\mu$  时,此节点为核心节点. 即

$$\text{Core}_{\epsilon, \mu}(v) \Leftrightarrow |N_{\epsilon}(v)| \geq \mu \quad (2)$$

由式(2)知,核心节点的数目由参数  $\epsilon$  和  $\mu$  确定,所以选择合适的参数至关重要. 尺度既可以决定  $\epsilon$  节点的个数,也可以在  $\epsilon$  节点确定的前提下,确定核心节点的个数. 一般而言,核心节点过多,会使得网络中大多数节点都是核心节点,直接影响算法复杂度;核心节点过少,会使得由节点间结构划分出的社区个数太少,影响最终的划分结果。

核心节点确定后,此核心节点即产生一个社区. 将核心节点的  $\epsilon$  节点分别加入社区,如果此  $\epsilon$  节点也是核心节点,那么再将该核心节点的  $\epsilon$  节点分别加入此社区,依次类推,直到无法加入节点为止. 由此可见,每一个或多个核心节点就可以产生一个社区,称为初始社区。

通过核心节点,得到了一个初始社区. 我们利用社区的适应度模型,逐一判断社区外的节点是否应当加入社区内. 计算节点加入社区前后的适应度大小,如果后者大于前者,则将此节点加入,否则不加入. 逐步判断和加入,直到社区外的节点都无法加入社区内为止. 适应度模型如下:

$$f_c = \frac{K_{in}^c}{(K_{in}^c + K_{out}^c)^\alpha} \quad (3)$$

式中,  $K_{in}^c$  为其社区  $c$  的内度之和,  $K_{out}^c$  为社区  $c$  外度之和,  $\alpha$  为适应度参数,可以表现不同尺度社区的大小.  $f_c$  越大,此社区就越健壮. 适应度参数  $\alpha$  的选择非常重要,参数较小的话会导致整个网络是一个大的局部社区。

通过控制适应度的方法,可以将适合加入社区的节点加进去,更深层次地挖掘社区成员. 此时,社区扩大,社区之间的重叠程度也增加. 由于节点加入社区的顺序不同,社区内也有一些不属于社区的节点被加了进去. 所以,我们要删除不属于社区的节点以及将重叠度较高的社区融合. 删除社区内节点只需要判断社区内有外度的节点,同样用到公式(3),如果删除社区内节点可以增加此社区的适应度,则将其删除;否则,不删除. 如果两个社区的融合度大于社区重叠度阈值,则将其融合。

我们将连接多个社区且不属于任何社区的节点定义为枢纽节点. 将枢纽节点加入与其连接的社区中,可以获得重叠社区。

结合以上方法,基于结构和适应度的社区发现算法(CDSF 算法)流程如下:

**Step 1** 设置适应度参数;初始化核心节点.

**Step 2** 通过核心节点,根据节点结构属性,确定社区结构,得到初始社区.

**Step 3** 发现初始社区后,找枢纽节点;将枢纽节点加入与之相连的社区中,得到重叠社区.

**Step 4** 利用公式(3)将社区外节点有序加入社区;加入社区的顺序由节点的 factor 值大小排列,这样可以大大减少移出节点次数。

$$\text{factor}_v = \frac{2d_{in}}{(d_{in} + d_{out})^\alpha} \quad (4)$$

$d_{in}$ 指的是该节点指向社区内的度数,  $d_{out}$ 指的是该节点指向社区外的度数.

**Step 5** 将社区内节点按公式(3)移出社区;移出社区时,只判断社区内有出度的节点即可.

$$d_{out} \geq 1 \quad (5)$$

**Step 6** 将社区之间重叠度大于  $m$  的社区融合.

$$\max \left\{ \frac{|C_1 \cap C_2|}{C_2}, \frac{|C_1 \cap C_2|}{C_1} \right\} \geq m \quad (6)$$

式中,  $C_i$  指的是社区容量,  $m$  为可变值.

算法基于结构和适应度的基本思想,先从简单的节点之间的结构特征,将相似结构的节点划在一起,形成初始社区;再引入社区适应度的概念,可以将社区适应度增大的社区外节点加入进来.另外,枢纽节点的加入、社区内节点的移出以及社区之间的融合,均会在一定程度上优化社区划分结果.

在有  $m$  条边和  $n$  个节点的网络中, CDSF 算法先找到有相似节点结构的初始社区,需要将整个网络的所有边遍历一遍.假设节点  $i$  的度为  $\text{degree}(v_i)$ , 那么时间复杂度为  $O(\text{degree}(v_1) + \dots + \text{degree}(v_n))$ . 所有节点的度之和为  $2m$ , 所以, 时间复杂度为  $O(m)$ . 然后, CDSF 算法再对初始社区进行扩张, 将能使社区适应度增加的节点加入社区内. 先计算每个社区的适应度, 再对社区外的节点计算其加入社区之后的社区适应度, 对每个与社区相接的社区外节点, 均计算两遍时间复杂度. 这些社区外节点的个数决定了耗时大小, 由于事先形成了初始社区, 需要计算社区适应度的社区外节点不是很多, 一般而言, 其数目与初始社区的社区内节点数目相当, 所以, 其时间复杂度为  $2 * O(0.5 * n)$ , 即  $O(n)$ . 将社区内节点移出到社区外, 其计算次数仅是社区内出度大于 0 的社区内节点, 时间复杂度小于  $O(n)$ . 社区融合过程中的时间复杂度计算, 只取决于已经得到的社区个数. 最坏情况下, 需要遍历整个网络的一半节点, 时间复杂度小于  $O(0.5 * n)$ . 综上, 基于结构和适应度的社区发现算法时间复杂度成线性, 为  $O(m+n)$ .

### 3 实验结果与分析

为了证明上述算法的可靠性, 本文选用了人工生成网络和真实网络两组数据进行试验. 人工生成网络根据 Lancichinetti 等<sup>[17]</sup> 在 2009 年提出的算法. 本文选取两个真实网络数据集, 分别是美国橄榄球数据集和美国政治书籍数据集.

实验结果是将本算法与 MSCD 算法<sup>[14]</sup>、SCAN 算法<sup>[13]</sup>作比较, 本实验中所有算法均用单线程. 实验设备为 3G 内存, CPU 为 AMD 双核, 1.90 GHz, 操作系统是 Windows 7. 关于算法的性能评价指标, 用的是 NMI 值<sup>[18]</sup>和 ARI 值<sup>[19]</sup>. NMI 和 ARI 值越大, 表示社区划分结果越好.

#### 3.1 人工网络实验结果

节点数目和平均节点度是社会网络形成的两个重要指标, 直接反映不同社会网络的属性, 在本实验中有重要意义. 本文设置的人工生成网络的实验分为两组: 第一组保持平均节点度不变, 将网络节点数目逐步变化, 观察三种算法的划分结果; 第二组保持节点数目以及社区数目不变, 逐步增加平均节点度, 观察三种算法的划分结果.

参数的设置决定社区分辨率. 为达到相同尺度, 三种算法的参数设置力求一致. 其中,  $\epsilon$  的取值为  $[0, 1]$ ,  $\mu$  的取值为正整数. 经实验验证、人工干预和不断调整, 以下参数设置可以达到基本一致的尺度, 得到最好效果. 其中, CDSF 算法的参数设置为  $\epsilon = 0.5, \mu = 5$ ; SCAN 算法的参数设置为  $\epsilon = 0.5, \mu = 5$ ; MSCD 算法的参数设置为  $\rho = 0.7$ .

##### (I) 节点数变化

人工网络情况下的节点变化实验参数设置如表 1 所示.

表 1 人工网络参数设置

Tab. 1 Parameters in artificial networks

参数	值
节点数目	5 000~60 000
边数目	15 000~180 000
实际社团数目	200~2 400
平均节点度	约 5.6 左右

图 4 展示了三种算法分别在不同节点数目的社会网络中, 其社区划分结果在 NMI 值、运行时间、ARI 值、社团个数与真实社团个数比例四个方面的结果比较. 从实验结果可以看出, CDSF 算法在外部评价结果 NMI 上优于 MSCD 算法和 SCAN 算法; 在时间上, SCAN 算法和 CDSF 算法都能比 MSCD 算法在更短时间找到社区, 随着节点数的增加, MSCD 算法时间复杂度增加较快, 而 CDSF 算法和 SCAN 算法均有线性的时间复杂度; ARI 值的比较, CDSF 算法同样也优于 SCAN 算法和 MSCD 算法; 社团数目上, SCAN 算法和 CDSF 算法与实际社团数目相差很近, MSCD 算法之所以与实际相差甚

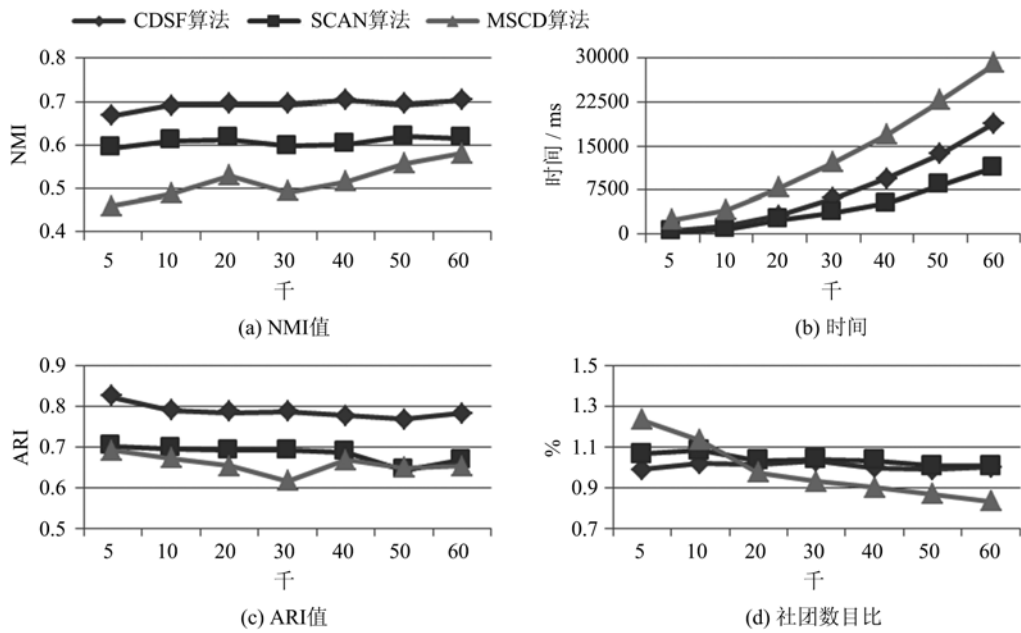


图 4 随节点数变化的社区发现结果展示

Fig. 4 Results of community detection with changes of node number

远,是因为它没有考虑节点的结构因素,单从某个节点出发,不断扩张,更易产生大社区;小社区与大社区重叠,更易融合,从而划分结果社团个数偏少。

CDSF 算法在精确度方面,随着节点的变化,与 SCAN 算法的趋势一致.原因是 CDSF 算法与 SCAN 算法在一开始选取了类似的种子节点. SCAN 算法通过找节点之间的相同结构,对社区不断进行扩张,直到获得最后结果;而 CDSF 算法在对节点结构进行扩张后,还有选择地将社团外的节点加入社区,从而将社区外的节点与社区内结构相似的节点加入进来.所以, CDSF 算法可以得到更好的结果,因此 NMI 值和 ARI 值也比 SCAN 算法高一些。

(II) 平均节点度变化

人工网络情况下的平均节点变化的实验参数如表 2 所示。

表 2 人工网络参数设置

Tab. 2 Parameters in artificial networks

参数	值
节点数目	50 000
边数目	88 000~200 000
实际社团数目	2 000
平均节点度	约 3.5~7.5

图 5 展示了三种算法分别在不同平均节点度的社会网络中,其社区划分结果在 NMI 值、运行时间、ARI 值、社团个数与真实社团个数比例四个方面的

结果比较. 在外部评价指标 NMI 上来看,平均节点度越大,CDSF 算法和 SCAN 算法的 NMI 值稍有增加;MSCD 算法有波动. 从时间上来看,平均节点度越大,三种算法耗时越大,但 SCAN 算法和 CDSF 算法增幅较小, MSCD 算法增幅较大. 在评价指标 ARI 值方面,三种算法均随着平均节点度的增加,ARI 值稍有减少,但 CDSF 算法一直高于 SCAN 算法和 MSCD 算法. 在相同尺度上,SCAN 算法和 CDSF 算法可以较精准地找到合适数目的社团,尤其在平均节点度的情况下,其更容易找到合适数目的社团. MSCD 算法在平均节点度较少的情况下,会找到更多的社团. 这与社团的扩张方式和社区融合过程有直接关系,当平均节点度较小时,社团不易扩张,更易形成小社区,且社区之间重叠度较小,不易融合;反之亦然。

CDSF 算法在平均节点度上各指标几乎趋于稳定,说明此算法的鲁棒性较好,适用于不同类型的网络. 随着平均节点度的增大,其 NMI 值一直保持在较高值,且耗时较少,ARI 值也显示,其能在不同密度的网络中找到合适的社团簇;与实际社团数目比较,CDSF 算法几乎能找到所有的社区结构。

3.2 真实网络实验结果

本文选取两个真实网络数据集,第一个数据集选取了美国橄榄球高校代表队的比赛情况,共含 115 支球队(点),616 场比赛(边). 第二个数据集选取亚

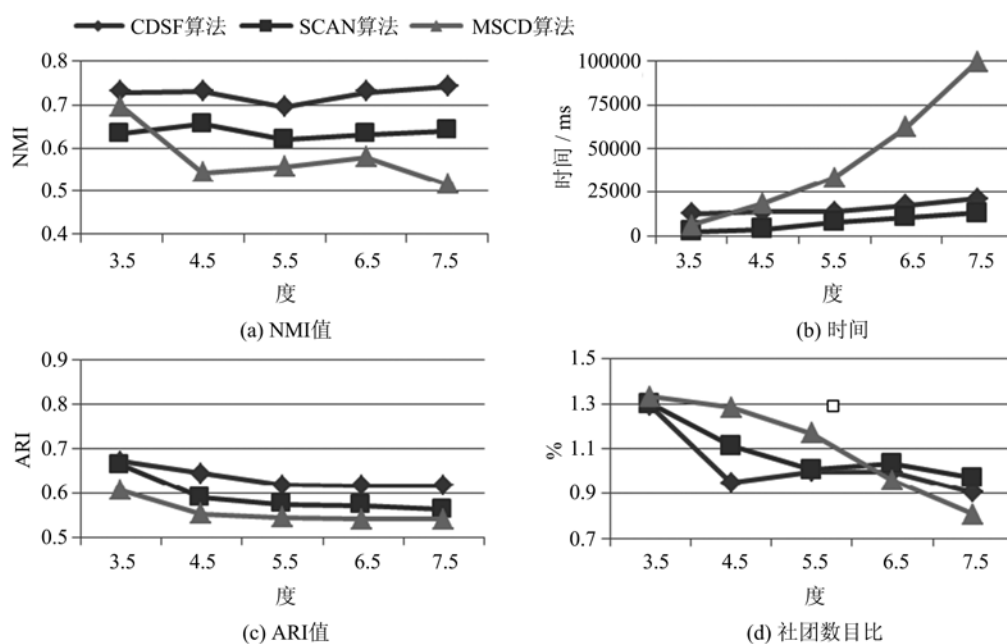


图 5 随平均节点度变化的社区发现结果展示

Fig. 5 Results of community detection with changes of average node degree

马逊关于美国政治书籍的销售情况,该数据集含有 105 个节点,441 条边。

#### (I) 美国橄榄球网络

美国橄榄球网络中,每个结点代表了参加美国 2000 年橄榄球赛季的高校代表队,连接两个结点之间的边则表示相应的两支球队之间至少曾有过一场比赛。美国橄榄球网络数据集包含 115 个节点,616 条边。这 115 支球队分别属于 12 个工会,社区划分的目的是找到每个工会包含的球队。

经实验验证、人工干预和不断调整,以下参数设置可以达到基本一致的尺度,获得最好效果。CDSF 算法的参数设置为  $\epsilon=0.5, \mu=5$ ; SCAN 算法的参数也设置为  $\epsilon=0.5, \mu=5$ ; MSCD 算法的参数设置为  $\rho=0.7$ 。

三个算法运行结果如表 3 所示。

表 3 美国橄榄球网络的社区划分结果

Tab. 3 Results of community detection in American football networks

算法	NMI	时间	ARI	社团个数
CDSF	0.966 636	47	0.970 067	12
SCAN	0.885 775	40	0.848 584	12
MSCD	0.892 965	2 513	0.670 796	10

由实验结果可以看出,三种算法几乎都能正确划分出 12 个左右的社团(MSCD 算法划分出 10 个)。从结果准确性上看,CDSF 算法最佳,其 NMI

值最高,达 0.97,与真实划分结果最接近;且耗时较小,ARI 值较高,可以取得理想的效果。

#### (II) 美国政治书籍网络

该数据集是关于美国政治书籍的分类。该数据集有 105 个节点,代表 105 本书籍,可以分为“民主”“保守”“中立”三类;441 条边表示买这本书的读者会同时买另一本书。该数据集的数据来源于亚马逊网站。

经实验验证、人工干预和不断调整,以下参数设置可以达到基本一致的尺度,获得最好效果。其中,CDSF 算法的参数设置为  $\epsilon=0.35, \mu=2$ ; SCAN 算法的参数设置为  $\epsilon=0.35, \mu=2$ ; MSCD 算法的参数设置为  $\rho=0.7$ 。

三个算法运行结果如表 4 所示。

表 4 美国政治书籍网络的社区划分结果

Tab. 4 Results of community detection in American political books networks

算法	NMI	时间	ARI	社团个数
CDSF	0.639 525	37	0.654 419	3
SCAN	0.531 241	32	0.639 913	3
MSCD	0.496 336	2 199	0.556 187	4

由实验结果看出,CDSF 算法和 SCAN 算法均可以正确划分出 3 个社区,分别代表“民主”“保守”“中立”三个社团;而 MSCD 算法划分了四个社区。就 ARI 值和 NMI 值而言,CDSF 算法的值均大于 MSCD 算法和 SCAN 算法,说明 CDSF 算法与实际

情况更加符合. 在运行时间上, CDSF 算法和 SCAN 算法相差不大, 均远远小于 MSCD 算法的运行时间.

人工网络数据集和真实数据集的实验均表明, CDSF 算法可以较好地划分出社区, 与实际划分结果相似度很高; 此外, 其运行时间也在可以接受的范围内, 比 SCAN 算法稍微耗时一点, 但在精确度方面远远优于 SCAN 算法.

## 4 结论

本文基于节点结构和节点簇结构, 提出一种新的社区划分算法, 这种算法可以较好、较快地找到社区. 人工生成网络和真实网络的实验表明, 该算法能够在时间与效果上达到很好的平衡, 能够在合理的时间范围内有效实现社区结构的优化.

现如今的社会网络是大数据的网络, 如何在最快时间准确找到社区, 一直是人们关注的重点. 如何在不同的社会网络找到更加准确的节点结构和节点簇结构特征, 是未来的研究方向.

### 参考文献(References)

- [1] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006.
- [2] Newman M E J. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167-256.
- [3] Li Y D, Liu J, Liu C L. A comparative analysis of evolutionary and memetic algorithms for community detection from signed social networks [J]. Soft Computing, 2014, 18(2): 329-348.
- [4] Bu Z, Zhang C C, Xia Z Y, et al. A fast parallel modularity optimization algorithm (FPMQA) for community detection in online social network [J]. Knowledge-Based Systems, 2013, 50: 246-259.
- [5] Shi C, Cai Y N, Fu D, et al. A link clustering based overlapping community detection algorithm[J]. Data & Knowledge Engineering, 2013, 87: 394-404.
- [6] Li K, Pang Y. A unified community detection algorithm in complex network[J]. Neurocomputing, 2014, 130(1): 36-43.
- [7] Bennetta L, Liub S S, Papageorgioub L G, et al. A mathematical programming approach to community structure detection in complex networks [C]// Proceedings of the 22nd European Symposium on Computer Aided Process Engineering. London, UK: Elsevier, 2012: 1 387-1 391.
- [8] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences, 2002, 99(12): 7 821-7 826.
- [9] Fortunato S. Community detection in graphs [J]. Physics Reports, 2010, 486(3): 75-174.
- [10] Wu F, Huberman B A. Finding communities in linear time: A physics approach[J]. The European Physical Journal B, 2004, 38(2): 331-338.
- [11] Newman M E J. Fast algorithm for detecting community structure in networks[J]. Physical review E, 2004, 69(6): 066133(1-5).
- [12] Zhou H J, Lipowsky R. Network brownian motion: A new method to measure vertex-vertex proximity and to identify communities and subcommunities [C]// Proceedings of the International Conference on Computational Science. Springer, 2004: 1 062-1 069.
- [13] Xu X W, Yuruk N, Feng Z D, et al. SCAN: A structural clustering algorithm for networks [C]// Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM Press, 2007: 824-833.
- [14] Martelot E L, Hankin C. Fast multi-scale community detection based on local criteria within a multi-threaded algorithm [EB/OL]. <http://arxiv.org/abs/1301.0955>.
- [15] Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure in complex networks[J]. New Journal of Physics, 2009, 11(3): 033015(1-18).
- [16] Serrano M Á, Boguñá M, Vespignani A. Extracting the multiscale backbone of complex weighted networks [J]. Proceedings of the National Academy of Sciences, 2009, 106(16): 6 483-6 488.
- [17] Lancichinetti A, Fortunato S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities [J]. Physical Review E, 2009, 80(1): 016118(1-9).
- [18] Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure in complex networks[J]. New Journal of Physics, 2009, 11(3): 033015(1-20).
- [19] Hubert L, Arabie P. Comparing partitions[J]. Journal of Classification, 1985, 2(1): 193-218.