

引用格式:乔虹,田玉玲,马建芬. 基于多属性融合策略的复杂网络社团划分算法[J]. 科学技术与工程, 2018, 18(32): 51—57
Qiao Hong, Tian Yuling, Ma Jianfen. Community detection algorithm based on multi-attribute fusion strategy[J]. Science Technology and Engineering, 2018, 18(32): 51—57

基于多属性融合策略的复杂网络社团划分算法

乔虹 田玉玲* 马建芬
(太原理工大学信息与计算机学院, 晋中 030600)

摘要 为解决目前社团划分算法依赖于单一度量方法,划分结果不够准确,无法适应不同类型的网络划分需求的问题,通过一种多属性融合策略方法研究社团划分。该算法首先基于节点连接,综合度量了多个影响社团划分的属性,并引入模块度概念确定各属性融合的权重,为划分增加了客观的参考信息,从而提高划分准确率;其次,针对传统划分方法迭代次数过多、效率低的问题,利用人工免疫网络内在的全局并行搜索能力实现对社团核心节点的快速寻优,并提出动态算子、免疫检测因子和反向学习机制对人工免疫网络的收敛速度和局部最优问题加以改进,从而提高寻优效率,缩短算法执行时间。最后,在三个经典真实数据集(Zachary、Dolphin、College Football)上进行实验,并将结果与经典算法对比。结果表明,该算法能适应不同的网络,且在较短的执行时间里实现更加精确的划分。可见,相比传统算法,本文算法具有更高的划分效率。

关键词 社团结构 社团划分 多属性 人工免疫网络
中图分类号 TP311; **文献标志码** A

社团结构具有同一社团的节点联系紧密,不同社团的节点连接稀疏的特点。近年来,复杂网络的社团结构特性被发掘,引起了学术界和社会的足够重视,因此针对其的研究与应用成为热点。基于层次的划分方法不依赖于社团个数,而是依据复杂网络所呈现出的结构进行划分,为社团划分的发展开辟了新的思路。

不同的层次算法采用不同的标准衡量节点的关系强度,根据关系强度对节点进行聚类,先得到小社团,再对小社团聚类,得到大社团。近年来,已经有很多方法被用来度量节点关系强度,如聚类系数^[1]、欧氏距离^[2]、链路预测^[3]、信息中心度^[4]、信息熵^[5]、传递信息量^[6]、公共邻居紧密性^[7]、节点动态连接度^[8]、结构相似度^[9]、传递增益系数^[10]、密度^[11]、增量相关^[12]、节点重要性^[13]等。但是,上述方法的度量标准往往是单一的,使得算法在特定的网络结构上存在缺陷,无法实现全面的度量,最终影响到划分效果,且由于传统算法在每次迭代时只添加一个节点,这使得重复计算操作过多,降低了划分效率。

近年来,多属性融合方法作为信息融合的关键

技术之一,解决了单一属性作为依据而引起的多解性问题,被应用于多个研究领域^[14,15]。本文采用多属性融合策略,通过考虑多个影响节点关系强度的属性,增加社团划分时可参考的信息,使算法适应多种不同的网络结构,实现更加全面的度量,从而提高社团划分的质量。同时,利用人工免疫网络实现网络的遍历寻优,确定社团核心节点,以核心节点为中心,实现其余节点的快速划分。为了加快寻优收敛速度,并有效避免搜索过程中的局部最优问题,引入动态算子以改进人工免疫网络中的克隆与变异操作,实现快速收敛,并结合反向学习机制,在保留局部最优结果的同时,快速跳出局部最优,扩大全局搜索范围。

1 人工免疫网络模型

人工免疫网络(artificial immune network model, aiNet)是一种模仿生物免疫系统的人工智能方法,在数据处理方面,具有学习、记忆以及通过竞争实现并行分布处理的能力。

生物免疫中,抗原是一种物质,它能刺激免疫系统并诱导其发生免疫反应。抗体则是在免疫系统受到抗原刺激后,实现与抗原特异性结合的免疫球蛋白。在aiNet免疫网络中,抗原和抗体都被表示为 L 维向量,对应于形态空间的一个特定位置,它们在形态空间里相互联系,每一对抗体和抗原之间的亲和度反映了它们的交互作用。在网络学习中,抗体通

2018年6月19日收到 国家自然科学基金(61472271)资助
第一作者简介:乔虹(1993—),女,汉族,山西定襄人,硕士研究生。E-mail:491926887@qq.com。
*通信作者简介:田玉玲(1963—),女,汉族,山西太原人,博士,教授。E-mail:tianyuling@tyut.edu.cn。

过竞争获取生存的权利,与抗原亲和度高的抗体,通过克隆选择原理进行克隆分裂,引导网络活化和细胞扩增;亲和度过低的抗体则被消除。另外,相似抗体之间的识别导致网络抑制,太过相似的抗体被消除,抗体不断进化,产生记忆细胞。aiNet 免疫网络最终会输出记忆细胞矩阵。

aiNet 免疫网络是一个边界加权图,由抗体节点组成,无须全部连接,以节点对集合为边界,每个连接对应分配的权值,当抗原入侵时,其结构如图 1 所示。

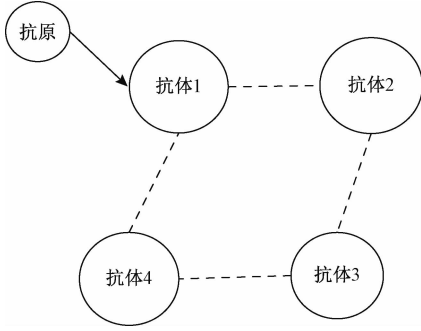


图 1 aiNet 免疫网络结构

Fig. 1 aiNet immune network structure

aiNet 免疫网络具有内在并行性和强大的全局搜索能力,但是,其寻优具有一定随机性,缺乏对寻优结果整体质量的监督和指导,且抗体的多样性使得网络收敛速度变慢,在搜索过程中容易陷入局部最优。因此,引入动态算子改进 aiNet 免疫网络的克隆和变异操作,以加快算法收敛速度,采用免疫检测因子度量每一代寻优结果的质量,并结合反向学习机制改善局部最优问题,使算法快速跳出局部最优,扩大全局搜索范围。

2 多属性融合策略社团划分算法

算法基于多属性融合策略和改进的 aiNet 免疫网络,多属性融合的主要思想为:基于一定数学运算基础,综合考虑多个属性的影响,并通过融合手段使这些属性相互补充,得出最优结果。

2.1 属性选择及亲和度计算

在对网络进行社团划分时,节点归属于哪个社团,往往与其邻居节点的归属有关。一个节点的归属既与和它有最多共享邻居的节点有关,也与共享邻居节点之间形成的连通子图有关,此外,当两节点不存在共享邻居节点时,其间路径长度将会对节点归属产生影响。因此,在度量网络中的节点关系强度时,综合考虑以下三个属性。

2.1.1 共享邻居

算法通过两节点间共享邻居数占其各自邻居节

点总数的比例来度量两节点间的关系强度,节点 v_i 和节点 v_j 的共享邻居 $CN(i,j)$ 由式(1)表示:

$$CN(i,j) = \frac{c[n(i),n(j)]}{n(i) + n(j)} \quad (1)$$

式(1)中: $n(i)$ 表示节点 v_i 的邻居节点数量; $n(j)$ 表示节点 v_j 的邻居节点数量; $c[n(i),n(j)]$ 表示节点 v_i 与节点 v_j 的公共邻居数量。

2.1.2 邻居相似度

算法采用邻居节点之间连边数量占其各自邻居节点数量乘积的比例实现基于邻居相似度的关系度量,节点 v_i 和节点 v_j 的邻居相似度 $ES(i,j)$ 由式(2)表示:

$$ES(i,j) = \frac{2e[n(i),n(j)]}{n(i)n(j)} \quad (2)$$

式(2)中: $e[n(i),n(j)]$ 表示节点 v_i 和节点 v_j 的邻居节点之间的连边数。

2.1.3 路径长度

算法定义节点间的路径长度为连接两节点的最短路径包含的边数,节点 v_i 和节点 v_j 的路径长度 $LS(i,j)$ 由式(3)表示:

$$LS(i,j) = \frac{1}{D(i,j) + 1} \quad (3)$$

式(3)中: $D(i,j)$ 表示节点 v_i 和节点 v_j 之间的最短路径所包含的边的数目。若节点 v_i 到 v_j 无通路,则认为 $D(i,j) = d$, d 为网络平均路径长度。

综上,为满足具有不同网络结构下的社团结构划分需求,基于多属性融合策略,节点 v_i 和节点 v_j 的亲密度 $I(i,j)$ 由式(4)定义:

$$I(i,j) = \frac{aCN(i,j) + bES(i,j) + cLS(i,j)}{a + b + c} \quad (4)$$

式(4)中: a 、 b 、 c 分别为 $CN(i,j)$ 、 $ES(i,j)$ 、 $LS(i,j)$ 的权值。

2.2 属性权值计算

需要注意的是,式(4)中,不同的 a 、 b 、 c 权值组合将影响相应的属性对亲和度的贡献,进而影响算法的划分效果。因此,在实现社团划分之前,应分析不同取值下,权值参数对划分效果的影响,确定最佳权值组合。

为了提高权值的客观性,避免人为决策的主观性和局限性,采用标准的模块度 Q 衡量在不同取值的权值参数下,依照各属性进行社团划分产生的划分效果。 Q 取值越大,表示该权值下划分效果越理想, Q 值计算方法由式(5)表示:

$$Q = \frac{1}{2N_b} \sum_{vw} \left[A_{vw} - \frac{k_v k_w}{2N_b} \right] \delta(v,w) \quad (5)$$

式(5)中: N_b 表示网络中的边总数; v 和 w 表示任意两节点,两点存在关系时,有 $A_{vw} = 1$,否则 $A_{vw} = 0$; k_v

表示节点 v 的度; $\delta(v, w)$ 用来表示节点 v 和 w 是否在同一个社区内, 若在同一社区, $\delta(v, w) = 1$, 否则 $\delta(v, w) = 0$ 。

2.3 改进的克隆与变异操作

aiNet 免疫网络采用克隆与变异操作保持抗体多样性, 其中, 克隆数目与亲和度成正比, 变异数目与亲和度成反比。但是, 传统的克隆和变异操作往往无法平衡种群多样性和算法的收敛速度。

为此, 引入动态算子, 在不同阶段分别强调收敛性和多样性。在亲和度影响克隆和变异的前提下, 使克隆与变异数目随着算法迭代次数动态变化, 以实现算法在搜索前期保证多样性, 后期保证收敛性, 实现收敛更快, 精度更高的寻优。

选用克隆数目 $cnum$ 与变异数目 $vnum$ 的计算公式分别见式(6)和式(7):

$$cnum(j) = [I(i, j) - V_i] \left(1 - 0.01 \frac{It_i}{It_n} \right) 100 \quad (6)$$

$$vnum(j) = \frac{0.01}{I(i, j) - V_i} \left(1 - 0.01 \frac{It_i}{It_n} \right) \quad (7)$$

式(6)和式(7)中: $I(i, j)$ 表示节点 v_i 和节点 v_j 的亲和度; V_i 表示亲和度阈值; It_i 表示第 i 次迭代; It_n 表示迭代总数 n 。同时, 为了避免冗余寻优, 实现更加精确的局部搜索, 本文算法采用邻域变异, 抗体节点将变异为其邻居节点。

2.4 免疫检测因子

引入免疫检测因子 op 来度量 aiNet 免疫网络中每一代记忆细胞库的质量, 其计算公式为

$$op(M_i, M_{i+1}) = \frac{1}{1 + \frac{\sum_{\alpha \in M_{i+1}} I(\alpha)}{\|M_{i+1}\|} - \frac{\sum_{\beta \in M_i} I(\beta)}{\|M_i\|}} \quad (8)$$

式(8)中: M_i 表示第 i 代记忆细胞集, $\|M_i\|$ 表示第 i 代记忆细胞集的规模大小; $I(i)$ 表示记忆细胞库中第 i 个记忆细胞对应的抗体受抗原刺激后, 计算得到的累计亲和度。

显然, 在一定迭代次数内, 当 op 值恒为 1 时, 表示当前记忆细胞集已经达到局部最优, 不再进化。

2.5 反向学习机制

反向学习机制是一种避免局部最优的有效方法, 其主要思想是在随机寻优的前提下, 反向解有近 50% 的概率优于当前最优解。

为了避免算法陷入局部最优, 引入反向学习策略, 当免疫检测因子值在一定迭代次数内的恒为 1 时, 认为算法陷入局部最优, 开始反向学习。规定反向学习的初始 R 个抗体 $Ab^* = \{Ab_1^*, Ab_2^*, \dots, Ab_r^*\}$ 与当前抗体 $Ab = \{Ab_1, Ab_2, \dots, Ab_r\}$ 的关系为

$$Ab_i^* = \psi_i + \xi_i - Ab_i, \quad Ab_i \in [\psi_i, \xi_i] \quad (9)$$

式(9)中: Ab_i 表示第 i 个抗体; $[\psi_i, \xi_i]$ 表示 Ab_i 的取值范围。

2.6 基于改进 aiNet 免疫网络的社团划分过程

算法的具体实现过程如下(假设有 N 个待划分节点)。

步骤 1 输入待划分节点信息及连接关系, 随机选择 R 个节点(研究表明, 一般情况下, N 个节点构成的网络, 其社团个数不超过 $\sqrt{N}$, 本算法设置 $R = \sqrt{N}$) 作为 aiNet 网络初始抗体。

步骤 2 将剩余 $N - R$ 个节点分别作为抗原提呈给网络进行自主学习, 按式(4)计算每一个抗原与网络中的抗体的亲和度, 把亲和度作为对应抗体受到的刺激值。

(1) 免疫克隆与变异, 对于第 i 个抗原 Ag_i , 与它亲和度高于阈值 V_i 的抗体, 按式(6)克隆一定数目, 产生克隆抗体集 Ab_i , 低于 V_i 的抗体按式(7)变异一定数目, 产生变异抗体集 Ab'_i 。

(2) 重新计算 Ab'_i 与抗原 Ag_i 的亲和度, 亲和度高于 V_i 的抗体及 Ab_i 中的抗体经过相互抑制, 进化为记忆细胞, 存入记忆细胞集 M_i 。若 M_i 中各记忆细胞受到抗原 Ag_i 的刺激值高于 V_i , 更新 V_i 为当前最小刺激值。

(3) 抗原提呈结束后, 所有记忆细胞集进入记忆细胞库 M , 统计群内记忆细胞对应抗体受到的累积刺激度, 从高往低排序, 保存前 R 个记忆细胞, 计算并记录当前免疫检测因子 op 。若达到反向学习条件, 则按式(9)初始化抗体, 构成新的网络; 否则, 记忆细胞库 M 内的记忆细胞分化成为下次迭代的初始抗体群。

步骤 3 对步骤 2 进行迭代, 若达到迭代次数, 结束, 得到最优记忆细胞库。

步骤 4 把最优记忆细胞库中的每个记忆细胞都作为社团核心节点, 计算其余节点与各个社团核心节点的亲和度, 并将它们归入与亲和度最高的核心节点所在社团, 若节点与两个及以上核心节点的亲和度相等, 则该节点为社团重叠节点。

3 实验结果与分析

3.1 实验数据集

为了验证所提划分方法的合理性与有效性, 采用以下三个具有不同网络结构的真实数据集进行实验。

3.1.1 Zachary 网络

通过观察一个美国大学空手道俱乐部成员间的

友谊关系而得出的社会网络,以俱乐部的成员为节点,成员之间存在的友谊关系为边,该网络包含 34 个节点和 78 条边。

3.1.2 Dolphin 网络

通过观察新西兰 Doubtful Sound 海峡 62 只海豚群体的交流情况而得出的社交网络,以海豚为节点,海豚间的频繁接触为边,该网络包含 62 个节点和 159 条边。

3.1.3 College Football 网络

通过美国大学生足球联赛而得出的社会网络。以参赛的足球队为节点,球队之间进行过比赛即形成边,该网络包含 115 个节点和 613 条边。

3.2 对比算法

主要将本文算法与以下三种经典算法进行对比。

3.2.1 快速 Newman 算法

该算法基于贪婪算法思想,通过计算各节点对之间的相似性,高相似性的节点不断凝聚进而形成社团结构,时间复杂度降低的同时划分精确度也有所降低。

3.2.2 GN 算法

该算法基于边介数,通过删除网络中的边进而找出社团结构,具有较高的划分精确度和时间复杂度。

3.2.3 LPA 算法

该算法基于凝聚思想,通过不断更新节点标签找出社团结构,具有较低的时间复杂度,稳定性较差。

3.3 评价指标

3.3.1 准确率

准确率是社团划分效果的衡量标准之一,其值一般为社团标号准确的节点占全部节点的比例。

假设网络 G 具有 N 个节点,算法发现的社团结构为 $C = \{C_1, C_2, \dots, C_k\}$,而网络实际社团结构为 $C^* = \{C_1^*, C_2^*, \dots, C_m^*\}$,则准确率 A 计算式为

$$A = \frac{\sum_{i=1}^N \text{sgn}(l_i, l_i^*)}{N}$$

(10)

式(10)中: l_i 代表第 i 个节点所属的社团标号; l_i^* 代表第 i 个节点的真实社团标号;当社团标号与实际标号一致时, $\text{sgn} = 1$; 否则, $\text{sgn} = 0$;

3.3.2 运行时间

算法的运行时间 T 能反映一个算法的运算速度和时间效率。为了保证数据的可靠性,统计一定次数内的算法的运行时间,并去掉最大值和最小值,求取平均运行时间, T 的计算式为

$$T = \frac{\sum_{j=1}^{\eta} T_j - \min - \max}{\eta - 2}$$

(11)

式(11)中: T_j 表示第 j 次算法的运行时间; $\min$ 表示统计次数内算法运行时间的最小值; $\max$ 则表示最大值; η 表示统计次数。

3.4 实验结果

3.4.1 对 Zachary 网络进行社团划分

(1) 选用参数

$N = 34, R = 6, V$ 的初值设为 0.05, 通过比较各属性权值的变化对模块度的影响(图 2), 得出对应的最佳权值组合为 ($a = 0.40, b = 0.80, c = 0.90$)。通过多次实验,发现迭代次数为 5 次以上时,算法执行结果稳定,因此设置迭代次数为 6。

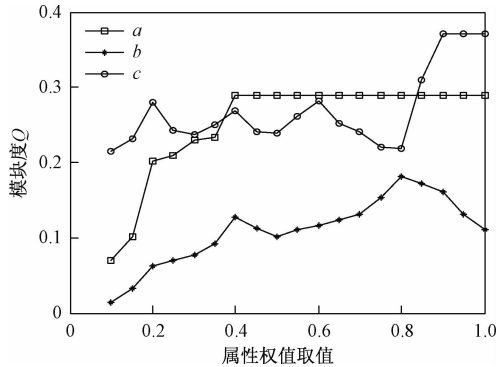


图 2 各属性不同权值下 Zachary 网络的模块度 Q
Fig. 2 The modularity Q of Zachary network under different weights of each attribute

(2) 算法执行结果

得到社团核心节点 1 和节点 34, 通过划分得到社团结构如表 1 所示。

表 1 Zachary 网络的社团划分结果
Table 1 Results of community detection of Zachary

社团编号	社团节点编号
社团 1	1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 17, 18, 20, 22
社团 2	9, 10, 15, 16, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34

(3) 与经典算法对比

本文算法得出的划分结果与 Zachary 网络的实际划分结果一致。而对比快速 Newman 算法和 GN 算法在划分 Zachary 网络时均错划 1 个节点, LPA 算法错划 7 个节点^[16], 本文算法的划分精确率要高于三个算法, 分别提升了 2.94%、2.94%、17.85%。

3.4.2 对 Dolphin 网络进行社团划分

(1) 选用参数

$N = 62, R = 8, V$ 的初值设为 0.05, 通过比较各属性权值的变化对模块度的影响(图 3), 得出对应

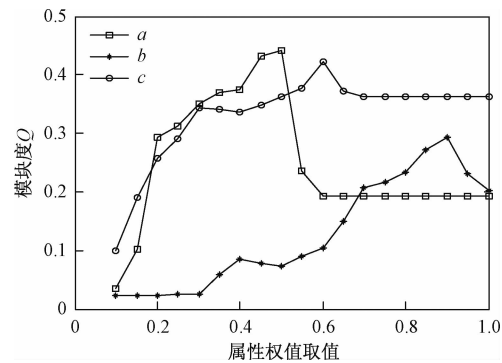


图3 各属性不同权值下 Dolphin 网络的模块度 Q
Fig. 3 The modularity Q of Dolphin network under different weights of each attribute

的最佳权值组合为 ($a = 0.50, b = 0.90, c = 0.60$)。通过多次实验,发现迭代次数为 10 次以上时,算法执行结果稳定,因此设置迭代次数为 11。

(2)算法执行结果

得到社团核心节点 15 和节点 58,通过划分得到社团结构如表 2 所示。

表 2 Dolphin 网络的社团划分结果	
Table 2 Results of community detection of Dolphin network	
社团编号	社团节点编号
社团 1	2,6,7,8,10,14,18,20,23,26,27,28,32,33,
	40,42,49,55,57,58,61
社团 2	1,3,4,5,9,11,12,13,15,16,17,19,21,22,24,
	25,29,30,31,34,35,36,37,38,39,41,43,44,
	45,46,47,48,50,51,52,53,54,56,59,60,62

(3)与经典算法对比

采用本文算法得出的划分结果与 Dolphin 网络的实际划分结果一致。而对比 GN 算法错划 1 个节点,快速 Newman 算法错划 3 个节点,LPA 错划 15 个节点^[16],本算法的划分精确率高于三个算法,准确率分别提升了 4.84%、1.61%、25.32%。

3.4.3 对 College Football 网络进行社团划分

(1)选用参数

$N = 115, R = 11, V$ 的初值设为 0.05,通过比较各属性在该网络的表现(图 4),得出该网络对应的最佳权值组合为($a = 0.5, b = 0.9, c = 0.2$)。通过多次实验,发现迭代次数为 20 次以上时,算法执行结果稳定,因此设置迭代次数为 21。

(2)本文算法执行结果

得到社团核心节点分别为 2,6,4,8,9,16,62,70,83,88,89,93,105,20,通过划分得到社团结构如表 3 所示,

(3)与经典算法对比

本文算法得出的划分结果与 College Football 网络的实际划分结果相比,划分错误的节点为 7 个。

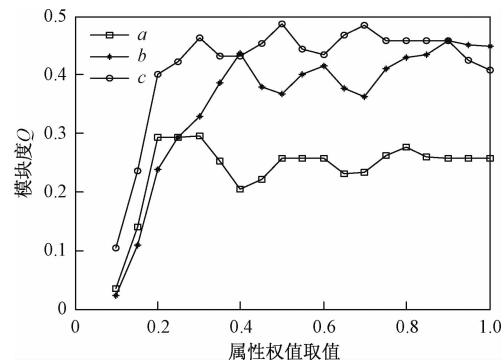


图4 各属性不同权值下 Football 网络的模块度 Q
Fig. 4 The modularity Q of Football network under different weights of each attribute

表 3 College Football 网络的社团划分结果	
Table 3 Results of community detection of College Football network	
社团编号	社团节点编号
社团 1	2,26,34,38,46,64,90,104,106,110
社团 2	6,4,11,41,53,73,75,82,85,98,99,103,108
社团 3	8,9,22,23,52,69,78,79,109,112
社团 4	16,3,7,14,33,40,48,61,65,101,107
社团 5	62,13,15,19,27,32,35,39,43,44,55,72,86,100
社团 6	70,12,25,29,51,91,
社团 7	83,81
社团 8	88, 21,28,57,60,63,66,71,77,96,97,114
社团 9	89, 47,50,54,59,68,74,84,111,115
社团 10	93,45,49,58,67,76,87,92,113

而对比快速 Newman 算法和 GN 算法划分错误的节点个数分别为 12 个和 10 个,LDA 错划 23 个^[16],发现本算法的划分精确率仍然高于三个算法,准确率分别提升了 4.34%、2.61%、13.91%。

通过上述三个真实数据集的实验,本算法与快速 Newman 算法、GN 算法、LPA 算法在划分准确率和执行时间方面的对比结果分别如图 5 和图 6 所示。

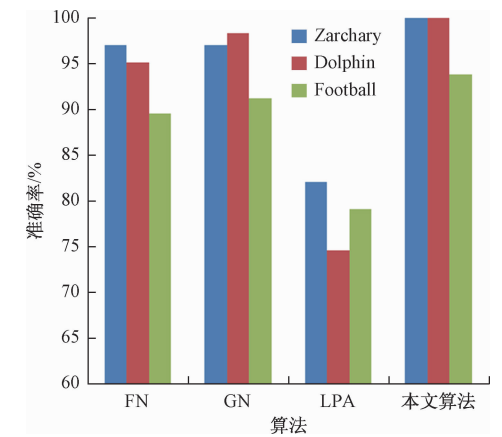


图5 不同算法对三个数据集的划分精确率对比
Fig. 5 Comparison of detection accuracy of different algorithms on three datasets

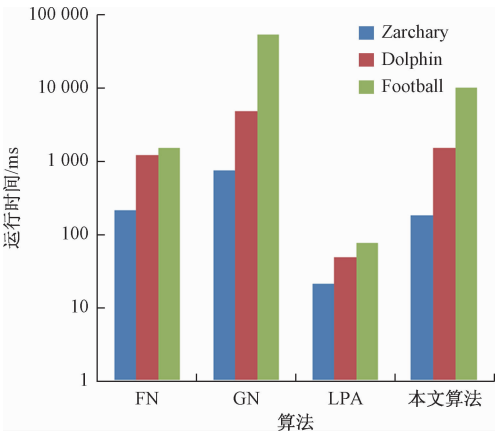


图 6 不同算法对三个数据集的划分执行时间对比
Fig.6 Comparison of detection runtime of different algorithms on three datasets

由实验结果可知,在上述不同的网络中,与经典的 GN 算法和快速 Newman 算法、LPA 算法,本文算法均能得到更加精确的划分结果,划分准确率得到了明显提高。

在运行时间方面,本文算法在节点数较少的 Zachary 网络和 Dolphin 网络上表现良好,能在较短的运行时间内得到更加精确的结果。随着网络节点数的增多,在 College Football 网络上,运行时间虽然低于 GN 算法,但是高于快速 Newman 算法和时间复杂度较低的 LPA 算法,虽然准确率提升了,但是运行时间仍然较长,因此,后续应继续加以改进。

4 结论

针对目前社团划分算法存在的问题,提出了一种新的社团划分算法,该算法基于多属性融合策略全面度量节点关系,并采用改进的 aiNet 免疫网络实现社团结构的快速划分。

通过 3 个经典真实网络的实验测试,结果表明,相对于传统算法,该方法不需要事先知道社团个数,能适应不同的网络,并取得更精确的划分结果,在一定程度上也减少了运行时间,具有一定的应用价值。后续研究将进一步考虑社团划分精度的影响因素,改进算法的准确性和性能,并在有向网络数据集上验证算法的准确性和运行性能,提升算法效果。

参 考 文 献

1 Radicchi F, Castellano C, Cecconi F, *et al.* Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2004; 101(9): 2658—2663

2 Pons P, Latapy M. Computing communities in large networks using random walks. *The Proceedings of 20th International Symposium on Computer and Information Sciences Conference*. Berlin: Springer-

Verlag, 2005: 284—293

3 Clauset A, Moore C, Newman M E J. Hierarchical structure and the prediction of missing links in networks. *Nature*, 2008; 453(7191): 98—101

4 赵凤霞, 谢福鼎. 基于 K-means 聚类算法的复杂网络社团发现新方法. *计算机应用研究*, 2009; 26(6): 2041—2043

Zhao Fengxia, Xie Fuding. Detecting community in complex networks using K-means cluster algorithm. *Application Research of Computers*, 2009; 26(6): 2041—2043

5 邓小龙, 王 柏, 吴 斌, 等. 基于信息熵的复杂网络社团划分建模和验证. *计算机研究与发展*, 2012; 49(4): 725—734

Deng Xiaolong, Wang Bo, Wu Bin, *et al.* Modularity modeling and evaluation in community detecting of complex network based on information entropy. *Journal of Computer Research and Development*, 2012; 49(4): 725—734

6 刘 瑶, 康晓慧, 高 红, 等. 基于节点亲密度和度的社会网络社团发现方法. *计算机研究与发展*, 2015; 52(10): 2363—2372

Liu Yao, Kang Xiaohui, Gao Hong, *et al.* A community detecting method based on the node intimacy and degree in social network. *Journal of Computer Research and Development*, 2015; 52(10): 2363—2372

7 Hu X, Guo J, Chen X, *et al.* Research of signed networks community detection based on the tightness of common neighbors. *The Proceedings of 6th International Conference on Digital Home Conference*, New York: Institute of Electrical and Electronics Engineers, 2016: 155—160

8 贾 珺, 胡晓峰, 贺筱媛. 基于节点动态连接度的网络社团划分算法. *复杂系统与复杂性科学*, 2016; 13(4): 56—61

Jia Jun, Hu Xiaofeng, He Xiaoyuan. Finding community structure in networks using node's dynamic connection degree. *Complex Systems and Complexity Science*, 2016; 13(4): 56—61

9 聂祥林, 张玉梅, 吴晓军, 等. 基于节点依赖度和相似社团融合的社团结构发现算法. *计算机工程与科学*, 2017; 39(7): 1273—1280

Nie Xianglin, Zhang Yumei, Wu Xiaojun, *et al.* A community detection algorithm based on node dependence and similar community fusion. *Computer Engineering and Science*, 2017; 39(7): 1273—1280

10 邓小龙, 翟佳羽, 尹栾玉. 基于矢量影响力聚类系数的高效有向网络社团划分算法. *电子与信息学报*, 2017; 39(9): 2071—2080

Deng Xiaolong, Zhai Jiayu, Yin Luanyu. Vector influence clustering coefficient based efficient directed community detection algorithm. *Journal of Electronics and Information Technology*, 2017; 39(9): 2071—2080

11 Vogiatzis D, Keros A D. A density based algorithm for community detection in hyper-networks. *The Proceedings of 12th International Workshop on Semantic and Social Media Adaptation and Personalization Conference*. New York: Institute of Electrical and Electronics Engineers, 2017: 59—64

12 Li G, Guo K, Chen Y Z, *et al.* A dynamic community detection algorithm based on parallel incremental related vertices. *The Proceedings of 2nd International Conference on Big Data Analysis*. New York: Institute of Electrical and Electronics Engineers, 2017: 779—783

13

Moayedikia A. Multi-objective community detection algorithm with node importance analysis in attributed networks. *Applied Soft Computing Journal*, 2018; 67: 434—451

14

付 凯,夏靖波,陈 珍. 一种改进的多属性系统结构分析模型. *科学技术与工程*, 2017; 17(4): 239—243
Fu Kai, Xia Jingbo, Chen Zhen. An improved multi-attribute system structure analysis model. *Science Technology and Engineering*, 2017; 17(4): 239—243

15

徐 浩,邢清华,王 伟. 基于投影的不完全信息直觉模糊多属性决策方法. *科学技术与工程*, 2017; 17(8): 101—105
Xu Hao, Xing Qinghua, Wang Wei. Projection based Intuitionistic Fuzzy Multi-attribute decision making method for incomplete Information. *Science Technology and Engineering*, 2017; 17(8): 101—105

16

吕伟明. 基于线性阈值模型与协同方法的社团检测算法研究. 兰州: 兰州大学, 2013
Lv Weiming. The research of community detection based on LT model and cooperative method. Lanzhou: Lanzhou University, 2013

Community Detection Algorithm Based on Multi-attribute Fusion Strategy

QIAO Hong, TIAN Yu-ling*, MA Jian-fen

(College of Computer Science and Technology, Taiyuan University of Technology, Jinzhong 030600, China)

[Abstract] Most community detection methods rely on a single method to measure the connection strength between nodes, and lead to not perfect effects, and they are unable to adapt to different types of Network, either. In order to solve above problems, a novel community detection algorithm based on multi-attribute fusion strategy was used to investigate community detection. Firstly, from the view of multi-attribute fusion strategy, the proposed algorithm comprehensively measured multiple factors which affecting community detection, and introduced the modularity to determine weights of attributes to get more objective reference for detection and achieve more accurate effects. Secondly, to decrease excessive iteration times and improve efficiency of traditional methods, the algorithm further used artificial immune network with powerful parallel computing ability to find the core nodes of the community rapidly, and improved convergence rate and solved the local optimal problem by using dynamic operators, immune detection operators and opposition-learning. The experimental results show that the proposed algorithm gets better results on three different datasets of Zachary, Dolphin and College Football compared with traditional methods. It is concluded that the proposed algorithm achieves more accurate division effects in shorter time and has greater efficiency.

[Key words] community structure community detection multi-attribute artificial immune network