

doi:10.11835/j.issn.1000-582X.2018.11.009

基于信息传递和峰值聚类的自适应社区发现算法

赵建军¹,汪清¹,由磊¹,洪文兴²

(1.天津大学 电气自动化与信息工程学院,天津 30072;2.厦门大学 自动化系,厦门 361005)

摘要:信息传递是网络具有的基本特征,基于此提出了一种基于信息传递和峰值聚类的自适应社区发现算法。首先,定义了节点与邻居之间的信任度函数,每个节点基于信任度独立的向网络中扩散信息量。扩散结束后,节点总信息量即为峰值聚类中的密度;网络中节点之间的距离通过所含节点信息量的倒数替代。然后,提出一种自动选取核心节点方法并为核心节点分配不同社区,把剩余节点分配到与它距离最短的核心节点所在社区,完成社区划分。本算法的优点在于无需额外参数并且能够发现社区内部结构。实验结果表明本算法发现的社区结构更加接近网络真实社区结构。

关键词:社交网络;社区发现;信息传递;峰值聚类

中图分类号:TP393

文献标志码:A

文章编号:1000-582X(2018)11-076-08

An adaptive community detection method based on information transfer and density peaks

ZHAO Jianjun¹, WANG Qing¹, YOU Lei¹, HONG Wenxing²

(1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, P.R.China;
2. Automation Department, Xiamen University, Xiamen 361005, Fujian, P.R.China)

Abstract: Information transfer is the basic feature of the network. Accordingly, an adaptive community detection method based on information transfer and density peaks is proposed. Firstly, the trust degree function between nodes and neighbors is defined, and each node independently spreads amount of information to the network based on the trust degree. After the diffusion, the total information amount of the node is the density of the density peaks. The distance between the nodes in the network is replaced by the reciprocal of the information amount of the destination node. Then, a method is proposed that can automatically select core nodes which are divided into different communities, and the remaining nodes are allocated to the community of the closest core node. The algorithm has the advantage that no additional parameters are needed and the internal structure of the community can be found. The experimental results show that.

Keywords: social network; community detection; information transfer; density peaks

收稿日期:2017-04-20

基金项目:国家自然科学基金资助项目(2015AA01A706)。

Supported by the National Natural Science Foundation of China(2015AA01A706).

作者简介:赵建军(1992—),男,天津大学硕士研究生,主要从事社交网络分析研究,(Tel)13370338283;(E-mail)349746280@qq.com。

洪文兴(联系人),男,厦门大学副教授,硕士生导师,主要从事信息网络方向研究,(E-mail)hwxx@xmu.edu.cn。

社交网络中存在关系不均匀的现象,形成了社交网络中的社区结构^[1-3]。网络中的社区结构有助于理解网络的拓扑结构,揭示复杂系统内部的规律,能够为信息推荐和信息传播控制提供有力支撑。目前学者们提出了很多社区发现算法,如谱平分法,模块度优化算法,标签传播算法^[4](LPA, label propagation algorithm),基于信息编码的算法^[5](Infomap),基于随机游走的算法^[6](Walktrap)。但是上述算法并不能够发现社区内部结构。

Rodriguez 等^[7]于 2014 年提出密度峰值聚类算法,能够得到稳定的聚类结果和聚类内部结构。该算法的核心思想在聚类中心的描述上,作者认为聚类中心同时具有 2 个特点:本身密度大,即被密度不超过它的邻居包围;与其它密度更大的数据点之间的“距离”相对较大。由于社区发现本质上是对网络节点进行聚类,因此该算法能够应用于社区发现中,并且在发现社区的基础上,确定每个社区内部的核心节点。但是在确定数据点的密度时需要给定截断距离,截断距离的选取会直接影响聚类效果。

黄岚等^[8]通过相似度定义网络中各节点之间的距离,将密度峰值聚类算法应用于社区发现中,提出一种基于点距离和密度峰值的社区发现算法(VDDPC, vertex distance and density peaks dustering);Wang^[9]等引入了箱线图选取核心节点,将密度峰值聚类算法应用于重叠社区发现算法中;Wang^[10]等基于密度峰值聚类算法,提出一种局部扩展的社区发现算法(LCCD, local expanding algorithm)。上述 3 种算法中截断距离是人为给定的,该参数的选取对社区发现的结果影响很大。Chen^[11]等基于密度峰值聚类提出一种线性复杂度的社区发现算法,但是并未提出一种自动的核心节点获取方法。

为了避免参数选取不当对算法性能的影响,笔者提出一种基于信息传递和峰值聚类的自适应社区发现算法(AID, adaptive community detection method based on information transfer and density peaks)。首先,引入信息量来度量节点密度与节点间距离。节点信息量通过节点间的信息传递过程获取。定义信息传递过程基于节点间的信任度,信任度越高的目的节点获取的信息量越大。用信息量来代替峰值聚类中密度,可避免截断距离的参数选取,且节点传播信息是独立的,易实现并行化。然后,提出一种自动选取核心节点的方法,从而完成整个网络的社区划分。

本算法主要优势:1)无需参数选取;2)能够发现社区内部结构;3)易实现并行化。

1 密度峰值聚类算法思想

Rodriguez 等人提出的密度峰值聚类算法^[7]认为每一个聚类都有一个聚类中心。聚类中心的特点包括:1)核心节点本身密度大,即被密度不超过它的邻居包围;2)与其它密度更大的数据点之间的“距离”相对较大。

设数据集中总的的数据点数为 N ,对于数据集中的每一个数据点 i ,用局部密度 ρ_i 和距离 δ_i 标记。

其中局部密度 ρ_i 定义为

$$\rho_i = \sum_j X(d_{ij} - d_c), \quad (1)$$

式中:当时 $d_{ij} < d_c$ 时, $X(d_{ij} - d_c) = 1$;否则 $X(d_{ij} - d_c) = 0$, d_{ij} 是数据点 i 和数据点 j 之间的距离, d_c 是截断距离。数据点 i 的密度等于距该点距离小于 d_c 的数据点个数。因此 d_c 是否选取适当,对聚类结果影响很大。

设 $\{q_i\}_{i=1}^N$ 表示 $\{\rho_i\}_{i=1}^N$ 的一个降序排列的下标序,即 $\rho_{q_1} \geq \rho_{q_2} \geq \dots \geq \rho_{q_N}$,则距离 δ_i 定义为

$$\delta_{q_i} = \begin{cases} \min_{q_j, j < i} \{d_{q_i q_j}\}, & i \geq 2; \\ \max_{j \geq 2} \{\delta_{q_j}\}, & i = 1; \end{cases} \quad (2)$$

由公式(2)可知,数据点 i 的距离 δ_i 表示比其密度大的数据点距离的最小值。

对于数据集中的每一个数据点 i ,计算得 (ρ_i, δ_i) ,以 ρ 为横轴,为 δ 纵轴画出该数据集的散点图,即决策图。选取同时具有较大的 ρ 值和 δ 值作为核心节点。对于剩余数据点,将其划分到距离其最近的核心节点所在的类簇中,完成整个数据集的聚类。

2 基于信息传递和峰值聚类的自适应社区发现算法

研究提出了一种基于信息传递和峰值聚类的自适应社区发现算法(AID)。该算法包含 4 个部分:1)信

息传递,2)距离矩阵计算,3)核心节点获取,4)社区划分。

2.1 信息传递

社交网络可表示为图 $G=(V,E)$,其中 $V=\{v_1,v_2,\dots,v_n\}$ 表示网络中节点集, $E=\{e_1,e_2,\dots,e_m\}$ 表示网络中边集, n 表示网络中节点的个数, m 表示网络中边的条数。

定义网络信息量矩阵 $S_{n \times n}$,其中 S_{ii} 表示节点 i 的初始信息量, S_{ij} 表示源节点 i 传递到节点 j 的信息量。

节点间信息传递遵循如下假设:

信息传递过程中,节点对于信任度不同的邻居节点所传信息量不同,信任度越大,所传递的信息量越大。

基于此定义了源节点 i 与邻居节点 $C(i,j)=|N_i \cap N_j|$ 的信任度,由公式(3)至公式(6)获得信任度函数。

$$C(i,j)=|N_i \cap N_j|, \quad (3)$$

$$\alpha(i,j)=\frac{|C(i,j)|+1}{|N_i|}, \quad (4)$$

$$\beta(i,j)=\begin{cases} \frac{2 \times |E(C(i,j))|}{(|C(i,j)|)(|C(i,j)|-1)}, \\ 0, \end{cases} \quad (5)$$

$$T(i,j)=\alpha(i,j) \times (\beta(i,j)+1), \quad (6)$$

式中: $|C(i,j)|$ 表示节点 i 与节点 j 的公共邻居个数, N_i 表示与节点 i 相邻的邻居节点, N_j 表示与节点 j 相邻的邻居节点, $|E(C(i,j))|$ 表示节点 i 与节点 j 公共邻居间的连边总数。公式(4)中,分母为源节点的邻居个数,使得 $T(i,j) \neq T(j,i)$,信息量的传递不对称,度数大的节点获得的信息量大。公式(5)表示 2 节点公共邻居间连边总数与其最大可能的连边总数的比值。由于 $\beta(i,j)$ 是建立在 $\alpha(i,j)$ 的基础上,因此总的信任度定义为公式(6)。

信息传递步骤如下:

1)初始化所有节点的信息量为 1,即信息量矩阵 $S_{n \times n}$ 为单位矩阵。

2)遍历网络中的节点,将每个节点依次作为源节点,对其初始信息量 1 进行扩散传递,此时并不考虑其余节点的信息量。

3)源节点在扩散传递信息量时,采用广度优先算法。以 i 为源节点,其传递到邻居节点 j 的信息量 $S_{ij}=1 \times T(i,j)$,邻居节点 j 要把源节点 i 传递的信息量 S_{ij} 继续扩散到 j 的邻居节点 k , k 节点获得信息量 $S_{ik}=S_{ij} \times T(j,k)=1 \times T(i,j) \times T(j,k)$,直到网络中所有节点都含有节点 i 的信息量时,节点 i 的信息传递完成。

信息传递结束后, $S_{ii}=1$ 表示节点 i 的原始信息量, $S_{ij}(j=1,2,\dots,n;j \neq i)$ 表示源节点 i 传递到网络上其它节点的信息量。 $M_j=\sum_{i=1}^n S_{ij}$ 表示节点 j 的信息总量。利用节点总信息量代替密度峰值聚类中各个节点的密度,从而避开了截断距离 d_c 的选取。

2.2 距离矩阵计算

本算法中,每一个节点需要局部密度 ρ 和距离 δ 标记。因此需要计算不同节点之间的距离,即距离矩阵。可通过信息量矩阵 $S_{n \times n}$ 获得距离矩阵 $D_{n \times n}$,具体步骤如下

1)将 $D_{n \times n}$ 对角线元素置为 0,即节点到自身的距离为 0;

2)令 $D_{ij}=S_{ij}(i \neq j)$;

3)依次求出 $D_{n \times n}$ 每行的最大值 $M_i(i=1,2,\dots,n)$;

4) $D_{ij}=\frac{M_i}{D_{ij}}(i \neq j)$,即得到归一化距离矩阵。

距离矩阵 $D_{n \times n}$ 中, D_{ij} 表示节点 i 到节点 j 的距离。节点 i 的距离 δ_i 表示节点 i 到比它密度大的节点的最短距离,当节点 i 的密度最大时,令 $\delta_i=\max\{\delta_i(j=1,2,\dots,n;j \neq i)\}$ 。这样网络中的每一个节点都可以

用密度 ρ 和距离 δ 去标记,即 (ρ, δ) 。

2.3 核心节点获取

对于任意节点 i ,用 (ρ_i, δ_i) 进行标记,并定义了混合参数^[7] $\gamma_i = \rho_i \times \delta_i, i = 1, 2, \dots, n$ 。 γ_i 的值越大,越有可能是核心节点。

为了使得整个算法达到自适应的目的,提出一种自动选取核心节点方法:

首先把网络中的节点分为 3 类,即非核心节点集,核心节点集和未确定节点集。分类依据为:非核心节点集为 $\delta = 1$ 的节点构成的集合;核心节点集为 $\delta \geq E(\delta_i) + \sigma(\delta_i), i = 1, 2, \dots, n$ 的节点构成的集合,其中 $E(\delta_i)$ 和 $\sigma(\delta_i)$ 为节点距离 δ 的均值和方差;剩余节点构成的集合即为未确定节点集。

然后进一步确定未确定节点集中的节点是否属于核心节点,采用如下算法

- 1) 计算核心节点集中混合参数 γ 的最小值 γ_{\min}^c 和非核心节点集中混合参数 γ 的最大值 γ_{\max}^{nc} 和最小值 γ_{\max}^{nc} ;
- 2) 把未确定节点集中 $\gamma \leq \gamma_{\max}^{nc}$ 的节点划分到非核心节点集中,被划分的节点总数记为 N ;
- 3) 把未确定节点集中的 $\gamma \geq \gamma_{\min}^c$ 节点划分到核心节点集中;
- 4) 对于未确定节点集中剩余的节点,依据 γ 值的大小进行升序排序。按序遍历节点,如果公式(7)成立则该节点属于非核心节点集,并令 $\gamma_{\max}^{nc} = \gamma_i$;如果公式(7)不成立,则计算终止,把本节点和剩余节点划分到核心节点集中。

$$\gamma_i \leq \gamma_{\max}^{nc} + \frac{\gamma_{\max}^{nc} - \gamma_{\min}^c}{N} \times \frac{\lambda_{\min}^c - \gamma_i}{\gamma_i - \gamma_{\max}^{nc}} \quad (7)$$

通过上述方法,确定了网络的核心节点,即核心节点集中的元素。

2.4 社区划分

首先,为每一个核心节点分配唯一的且不同于其他核心节点的社区号。

对于非核心节点 i ,距离矩阵 $D_{n \times n}$ 中 D_{ji} 表示节点 j 到节点 i 的距离,所有核心节点距节点 i 的距离可通过距离矩阵 D 确定。把节点 i 划分到与它距离最小的核心节点所在的社区内。依次遍历所有非核心节点,即可完成整个网络的社区发现。

完成社区发现后,可通过节点总的信息量,确定节点在社区内的重要程度,即社区内部结构。

2.5 时间复杂度分析

本算法由 4 个步骤组成,第一步信息传递中,单个节点信息传递复杂度为 $O(n-1)$,因为网络中节点都要作为源节点进行信息扩散传递,因此总的复杂度为 $O(n^2-n)$,其中 n 为网络的节点数;第二步中,获得距离矩阵每行最大值的时间复杂度为 $O(n^2)$,对距离矩阵归一化处理的复杂度为 $O(n^2)$,总的复杂度为 $O(2n^2)$;第三步核心节点选取中只需对网络中的每个节点遍历即可,因此复杂度为 $O(n)$;第四步社区划分中,需依次遍历普通节点,其复杂度为 $O(n-n_c)$, n_c 为核心节点的个数。综上,总的算法复杂度为 $O(n^2)$ 。由于各个节点的信息传递是相互独立的,因此本算法易实现并行化。

3 实验与分析

在真实网络数据集和人工网络数据集上与经典社区发现算法 LPA^[4]、Infomap^[5]、Fastgreedy^[12]、Walktrap^[6] 和基于峰值聚类的 VDDPC^[8]、LCCD^[10] 社区算法进行比较,实验结果验证了本算法的可行性和有效性。

3.1 真实网络数据集

表 1 中列出了 4 个真实数据集都具有已知的社区结构,包括 Karate 数据集^[13]、Dolphins 数据集^[14]、Football 数据集^[15] 和 Polbooks 数据集。

Karate 网络数据反映了某大学空手道俱乐部成员之间的关系,节点之间的连边表示成员之间的好友关系,由于校长与俱乐部主管的矛盾,该网络分裂成为 2 个社区。

Dolphins 网络数据中每个节点代表了一只宽吻海豚,节点之间的连边表示 2 只海豚之间接触频繁。经过长期观察,该网络被分为 2 个社区。

Football 网络数据集中每一个节点代表一支球队,节点之间的连边表示球队之间存在一场或多场比赛。由于球队属于不同的州,每个州内的比赛场次多,因此该网络被分为 12 个社区。

Polbooks 网络数据集中每一个节点代表一本在亚马逊销售的政治类书籍,节点之间的连边表示 2 本书被同时购买的频率高。由于政治类书籍代表的政治立场不同,所以该网络被划分为 3 个社区,分别代表“自由”、“保守”和“中立”。

表 1 真实网络数据
Table 1 Real-world networks

数据集	节点数	边数
Karate	34	78
Dolphins	62	158
Football	115	612
Polbooks	105	441

3.2 人工网络数据集

Andrea Lancichinetti 等人提出了著名的 LFR^[16]人工基准网络构造模型。该模型的优势在于可以通过参数控制整个网络和各个社区的规模、节点度分布等拓扑性质,使其更加符合真实社交网络的拓扑结构。

该模型中 N 为网络节点数, $\langle k \rangle$ 为平均节点度数, k_{\max} 为最大节点度数, C_{\min} 为网络最小社区所含的节点数, C_{\max} 为网络最大社区所含的节点数, t_1 和 t_2 分别为度分布和社区大小分布的幂指数, μ 为模糊参数,表示网络社区的模糊程度, μ 值越大,网络的社区结构越难发现,其取值范围为 $(0, 1)$ 。为了验证本算法的有效性,生成 3 组网络,分别为 LFR1($N=1\ 000$), LFR2($N=5\ 000$), LFR3($N=10\ 000$),网络其它参数设置如下:平均度数 $\langle k \rangle=20$,最大度数 $k_{\max}=50$, $t_1=2$, $t_2=1$,社区大小 $([C_{\min}, C_{\max}])$ 取值为 $[10, 50]$ 。

3.3 评价指标

模块度 Q ^[17] 被广泛的应用于社区划分结果的评价中。其定义如公式(8)所示。

$$Q = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{k_i k_j}{2m}] \delta_{c_i, c_j}, \quad (8)$$

Q 表示网络中社区内部节点之间的连边所占的比例和另一个随机网络中社区内部节点连边所占比例的期望相减得到的差值,其取值范围为 0 到 1。数值越接近 1 表示社区划分的结果越好。

标准化互信息 NMI(normalized mutual information)^[18] 是由 Danon 等人提出的一种衡量检测社区与真实社区接近度的指标,其定义如公式(9)所示

$$\text{NMI}(R | P) = 1 - [H(R | P) + H(P | R)]/2. \quad (9)$$

NMI 的取值范围为 0 到 1,其值越高表明所划分的社区结构与已知的社区结构越接近。当其取值为 1 时,所划分的社区结构与已知社区结构相同。

3.4 实验结果与分析

将本算法与 Infomap、LPA、Fastgreed、Walktrap、VDDPC、LCCD 等算法进行对比。VDDPC 和 LCCD 算法需要确定截断距离,依照作者研究选取最优的参数。LPA 为随机算法,表 2 中的实验结果是独立运行 50 次的平均值。

3.4.1 真实网络数据集结果分析

图 1 为本算法在 4 个真实网络上的决策图,表 2 列出了不同算法在真实数据集上所划分社区的评价结果。

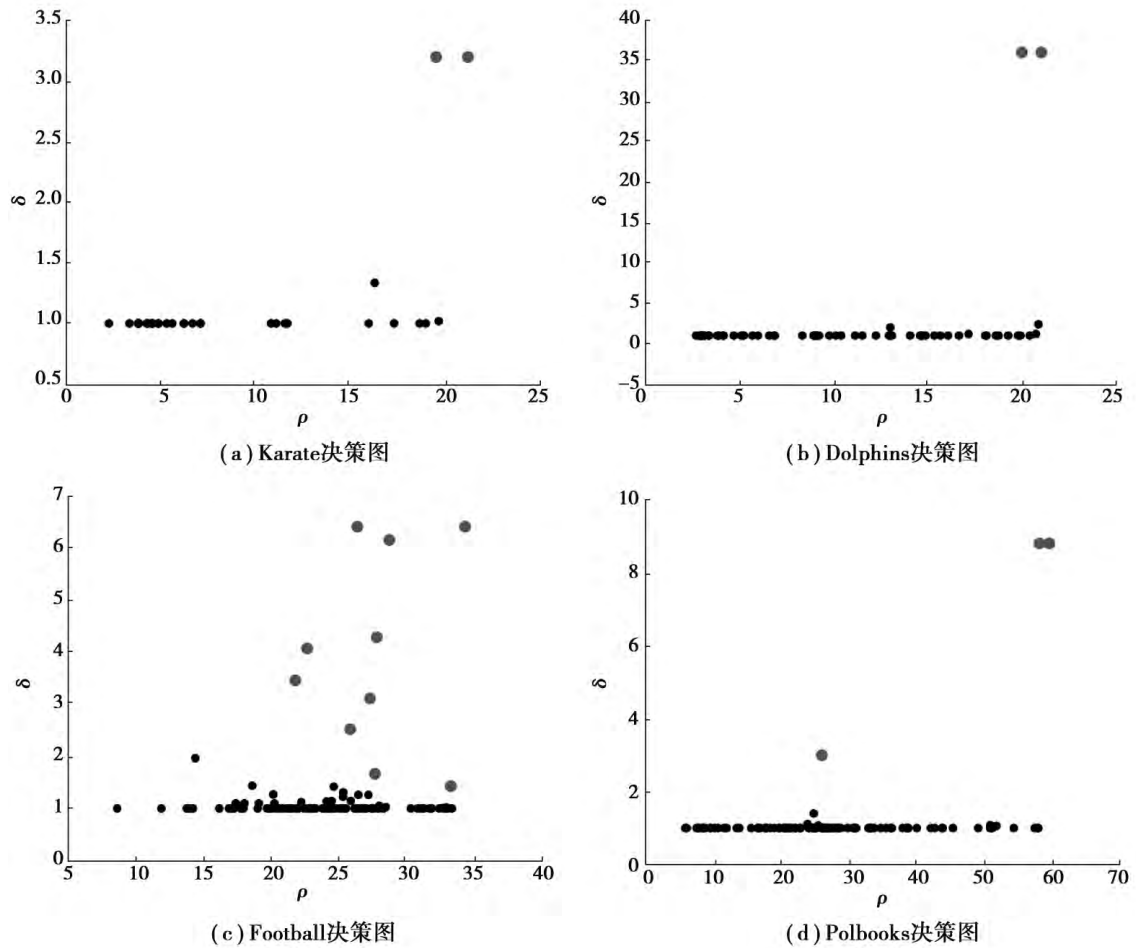


图 1 真实网络决策图

Fig.1 Decision graph of real-world networks

表 2 在已知社区结构的真实网络下的算法分析

Table 2 Algorithm analysis of real-world networks with community structure

网络	AID		Fastgreed		Infomap		Walktrap		LPA		VDDPC		LCCD	
	Q	NMI	Q	NMI	Q	NMI	Q	NMI	Q	NMI	Q	NMI	Q	NMI
Karate	0.371	1.000	0.381	0.692	0.402	0.699	0.353	0.504	0.416	0.602	0.371	1.000	0.420	0.699
Dolphins	0.385	0.814	0.482	0.604	0.519	0.481	0.487	0.418	0.427	0.745	0.315	0.697	0.526	0.589
Football	0.583	0.895	0.550	0.698	0.601	0.924	0.603	0.887	0.593	0.860	0.603	0.887	0.607	0.884
Polbooks	0.502	0.563	0.502	0.531	0.523	0.493	0.507	0.543	0.504	0.544	0.476	0.513	0.526	0.533

图 1 所示决策图中,红色节点为核心节点,本算法利用节点总信息量代替节点密度,使得核心节点与非核心节点产生明显的区别,提高网络社区划分的准确性。

表 2 所示,在 Karate、Dolphins 和 Polbooks 3 个真实网络数据集上,相较其它 6 个算法本算法所划分的社区结构具有最高的 NMI 值,即与网络已有的社区划分最接近。而在 Football 数据集上,本算法所划分社区的 NMI 值为 0.895,只比 Infomap 算法低,并且 2 者相差不大,本算法所划分的社区结构依然有较高的水平。而本算法在 Q 值上,并未优于其它算法,这主要与 Q 与 NMI 的定义有关。模块度 Q 是从理论上定义社区,其仅仅考虑社区内外节点之间连边数量的比值。而 NMI 评估的是算法所划分社区与网络真实社区的相似度。以 Karate 网络为例,本算法所划分的社区与网络原有社区相同(NMI=1),相应的 Q 值仅为 0.371。

对于已有社区结构的网络来说,越接近真实值,算法越有意义, NMI 比 Q 具有更大的指导意义。因此与其它算法相比,本算法所划分的社区结构更加准确。

在真实网络数据集上,与前 4 种经典社区发现算法相比本算法不仅在准确度上优于上述 4 种算法,同时本算法还能发现社区内部结构。与后 2 种基于峰值聚类的社区发现算法相比,本算法无需确定截断距离等额外参数,同时 NMI 值也优于上述 2 种算法。

3.4.2 人工网络数据集结果分析

图 2,图 3 和图 4 分别是 7 种算法在 LFR1,LFR2,LFR 人工数据集上的实验结果,其横轴为模糊参数 μ ,纵轴为 NMI。

由 3 图可知,随着 μ 的增加,7 种算法的 NMI 值都呈现下降趋势。图 2 中,当 $\mu < 0.5$ 时,除 Fastgreed 算法,其它算法都能准确的发现网络社区,这是因为 Fastgreed 算法是针对模块度优化的。当 $\mu > 0.5$,各算法的 NMI 值出现明显下降,LPA 与 Infomap 算法分别在 $\mu \geq 0.6, \mu > 0.7$ 时,NMI 变为 0,即整个网络被划分为一个社区。其它 4 种算法,在 $\mu > 0.7$,仍能够发现社区,但明显本算法(AID)与 Walktrap 算法的 NMI 值要高于其它 2 种算法。在 $\mu \geq 0.8$ 时,本算法的 NMI 值大于其他 6 种算法。

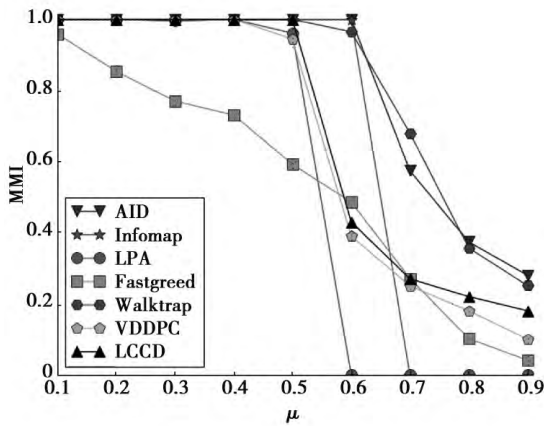


图 2 各算法在 LFR1 上的 NMI 值

Fig.2 NMI comparison of various algorithms on LFR1

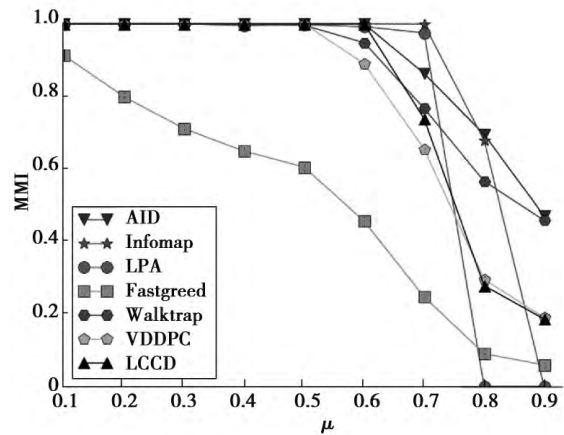


图 3 各算法在 LFR2 上的 NMI 值

Fig.3 NMI comparison of various algorithms on LFR2

对 3 图作横向对比,随着网络节点数增加时,各算法的 NMI 值的下降趋势减缓。在图 3,图 4 中,当 $\mu = 0.7$,本算法的 NMI 值低于 Infomap 和 LPA 算法,但是本算法的 NMI 随着 μ 的增加,下降缓慢,当 $\mu = 0.9$ 时,NMI 值是 7 种算法中最高。

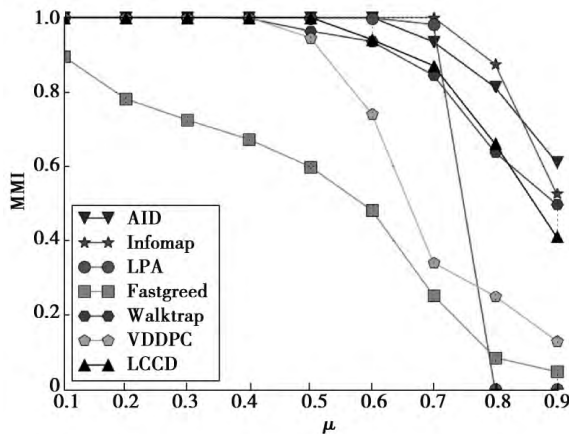


图 4 各算法在 LFR3 上的 NMI 值

Fig.4 NMI comparison of various algorithms on LFR3

综上,本算法在人工数据集上,能够稳定的发现网络的社区结构。当社区结构明显时,能够准确的发现网络社区;当社区结构模糊时,相对于其它算法,依然有较高的社区发现能力。

4 结 论

笔者提出一种基于信息传递与峰值聚类的自适应的社区发现算法。利用节点间基于信任度的信息传递确定峰值聚类中的节点密度和节点距离,其能够很好的区分核心节点与非核心节点;然后通过核心节点完成社区划分。

本算法在整个社区划分过程中,无需额外参数,能够自适应的应用于各种网络中。同时,当节点作为源节点进行信息传递时,与其他节点的信息量无关,因此本算法易实现并行化。实验结果验证了本算法的可行性和有效性。

参考文献:

- [1] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences, 2002, 99(12): 7821-7826.
- [2] Fortunato S. Community detection in graphs[J]. Physics Reports, 2009, 486(3-5):75-174.
- [3] Newman M E J. Communities, modules and large-scale structure in networks[J]. Nature Physics, 2012, 8(8):25-31.
- [4] Raghavan U N, Albert R, Kumara S. Near linear time algorithm to detect community structures in large - scale networks[J]. Physical Review E, 2007, 76(3): 36-40.
- [5] Rosvall M, Bergstrom C T. Maps of random walks on complex networks reveal community structure[J]. Proceedings of the National Academy of Sciences, 2008, 105(4): 1118-1123.
- [6] Pons P, Latapy M. Computing communities in large networks using random walks[C]// Computer and Information Sciences-ISCIS 2005. Berlin: Springer Heidelberg, 2005: 284-293.
- [7] Rodriguez A, Laio A. Clustering by fast search and find of density peaks[J]. Science, 2014, 344(6191): 1492-1496.
- [8] 黄岚,李玉,王贵参,等. 基于点距离和密度峰值聚类的社区发现方法[J]. 吉林大学学报工学版, 2016, 46(6):2042-2051. HUANG Lan, LI Yu, WANG Guishen, et al. Community detection method baased on vertex distance and clustering of density peaks[J]. Journal of Jilin University, 2016, 46(6):2042-2051(in Chinese).
- [9] Huang L, Wang G, et al. A link density clustering algorithm based on automatically selecting density peaks for overlapping community detection[J]. International Journal of Modern Physics B, 2016, 30(24): 1650167.
- [10] Wang X, Liu G, Li J, et al. Locating structural centers: a density-based clustering method for community detection[J]. PLoS ONE, 2017, 12(1): e0169355.
- [11] Chen Y. Finding communities by their centers[J]. Scientific Reports, 2016,6:24017.
- [12] Clauset A, Newman M E, Moore C. Finding community structure in very large networks[J]. Physical review E, 2004, 70(2): 066111.
- [13] Zachary W W, An information flow model for conflict and fission in small groups[J]. Journal of Anthropological Research, 1997, 33(4):452-473.
- [14] Lusseau D. The emergent properties of a dolphin social network[J]. Proceedings Biological Sciences, 2003 (270): S186-S188.
- [15] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy Sciences of USA , 2002, 99(12):7821-7826.
- [16] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms[J]. Physical Review E, 2008, 78(4): 46-52.
- [17] Newman M E, Girvan M. Finding and evaluating community structure in networks[J]. Physical review E, 2004, 69(2): 026113.
- [18] Danon L, Diaz-Guilera A, Duch J, et al. Comparing community structure identification[J]. Journal of Statistical Mechanics, 2005, 2005(09):9-18.

(编辑 侯 湘)