具有社区结构的无标度网络生成算法

郑文萍[1,2,3] 曲 瑞[1] 穆俊芳[1]

(山西大学计算机与信息技术学院 太原 030006)[1]

(山西大学计算智能与中文信息处理教育部重点实验室 太原 030006)[2]

(大数据挖掘与智能技术山西省协同创新技术中心(山西大学) 太原 030006)[3]

摘 要 近年来,生成图模型在复杂网络研究中的作用越来越重要。图的生成过程对于研究疾病的蔓延和信息的传播具有重大意义,同时图模型的生成也有助于更深入地研究复杂网络的特性。为了能够生成既符合真实网络特征又具有结构多样性的复杂网络,提出了一种具有社区结构的可调节聚集系数和模块性的无标度网络生成算法——TCMSN(Scale Free Network with Tunable Clustering Coefficient and Modularity)。通过调节混合参数可以调节生成网络的模块性,通过调节社区内连边的概率和混合参数可以对网络聚集系数进行调节。TCMSN采用了合理的连边策略,在不破坏网络结构多样性的情况下,能尽可能维持网络的无标度特性。人工构造数据和真实网络数据的对比实验结果表明,TCMSN算法能够生成可调节聚集系数和模块性的无标度网络模型,且能够生成最接近真实网络社区结构特征的网络模型。

关键词 网络生成模型,BA无标度网络,聚集系数,社区结构

中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.02.013

Generation Algorithm for Scale-free Networks with Community Structure

ZHENG Wen-ping[1,2,3] QU Rui[1] MU Jun-fang[1]

(School of Computer & Information Technology, Shanxi University, Taiyuan 030006, China)[1]

(Key Laboratory of Computation Intelligence & Chinese Information Processing, Shanxi University, Ministry of Education, Taiyuan 030006, China)[2]

(Collaborative Innovation Center of Shanxi Province for Big Data Mining and Intelligence Technology (Shanxi University), Taiyuan 030006, China)[3]

Abstract Generating complex network models can help researchers to understand network behaviors and simulate the transmission processes of disease epidemics and information diffusion. It is also important to generate complex networks meeting the characteristics of real networks and having structural diversity. A network generation algorithm TCMSN (Scale-free Network with Tunable Clustering Coefficient and Modularity) was proposed to generate scale-free complex networks with tunable clustering coefficient and modularity. TCMSN can adjust modularity by changing the mixing parameter and adjust clustering coefficient by changing the global preferential attachment probability and mixing parameter of the network. It adopts a reasonable strategy about adding edges in networks to maintain the scale-free characteristics, as much as possible without destroying network diversity. Experimental results on artificial data sets and real networks show that the proposed TCMSN algorithm can not only generate scale-free network model with tunable clustering coefficient and modularity, but also generate network model closed to the community structure of the real networks.

Keywords Network generation models, BA scale free network, Clustering coefficient, Community structure

1 引言

自然界中存在大量的复杂网络,如社会网络、蛋白质互作用网络、新陈代谢网络等[1],深入研究和探索这些复杂网络的内在结构、性质和行为是一项亟需研究的课题。建立合理的复杂网络模型,有助于理解和研究复杂系统的结构和动力学行为,并为面向复杂网络的相关算法[2-5]提供数据基础。不同领域的复杂网络(如万维网、电子邮件网络等信息领域网络;

到稿日期:2017-04-01 返修日期:2017-07-11 本文受国家自然科学基金(61572005),山西省回国留学人员科研基金(2017-014)资助。

郑文萍(1979—),女,博士,副教授,CCF会员,主要研究方向为图论算法、生物信息学等,E-mail:wpzheng@sxu.edu.cn(通信作者);曲 瑞(1993—),女,硕士生,CCF会员,主要研究方向为复杂网络建模;穆俊芳(1991—),女,博士,CCF会员,主要研究方向为复杂网络建模。

移动通信网、电力网、道路交通网等技术领域网络；人际关系网络、恐怖犯罪网络等社会系统领域网络；基因调控网络、蛋白质互作用网络等生物技术领域网络）在网络规模、度分布、混合特性和社区结构等方面表现各异。通常可以用一个图 $G=(V,E)$ 来描述复杂网络，节点集 V 代表网络研究对象个体的集合，边集 E 代表对象个体之间的相互作用或关系的集合。对复杂网络内在社区结构和产生机理的研究可以归结为对相应图模型的研究，利用图论知识可以有效地建立符合实际的网络模型。

Watts 和 Strogatz[6] 于 1998 年给出了“WS 小世界模型”的构建方法，该方法基于规则图以固定的概率重新连接网络中的每条边，形成了节点度服从泊松分布的网络模型。为了体现真实复杂网络的无标度特性，Barabási 和 Albert[7] 于 2001 年将复杂网络的无标度特性归结为增长机制和优先连接机制（如万维网、金融网络和美国飞机航班网络等），提出了节点度服从幂律分布的“BA 无标度模型”。然而，BA 无标度模型不能反映现实网络中节点连接的变化情况，只能生成幂指数近似为 3 的无标度网络。考虑到实际网络中的小世界特性、聚集性以及社区结构等特征，研究者也提出了一些 BA 无标度网络的改进模型[8-13]。

聚集性是复杂网络的重要特征之一，刻画了复杂网络的传递性，即网络中若两个节点与同一节点关联，则这两个节点在很大程度上也存在关联。通常用聚集系数[14] 来表示复杂网络中节点邻域的聚集性，通过网络中节点邻域内三角形的平均比例来量化网络的聚集性。区别于聚集性，社区结构[15-16]（也称为集簇结构）表示网络中节点的抱团性，即将复杂网络中的节点分组，使组内节点连边相对紧密而组间节点连边相对稀疏。社会系统领域的人际关系网络、生物技术领域的蛋白质互作用网络等都呈现了明显的社区结构，通常用模块性[17] 来衡量复杂网络是否具有明显的社区结构。

Newman 等人[17] 于 2002 年提出了 GN-Benchmark 算法，构造了具有明显社区结构的正则网络模型，将 128 个节点平均分成 4 个社区，社区中的节点度数均为 16。GN-Benchmark 算法构造的网络聚集系数和模块性不可调，网络结构比较单一，缺乏多样性。Fortunato 等人[18] 于 2008 年提出了 LFR-Benchmark 算法，构造了节点度序列和社区规模序列都服从幂律分布且有明显社区结构的复杂网络模型。该算法随机选择满足条件的社区添加节点，并根据“配置模型”[19] 添加节点间的连边以形成网络，但 LFR-Benchmark 算法无法生成聚集系数可调的网络。Tmara 等人[20-21] 假设一个网络是由若干个 ER 子图构成的，并在此基础上提出了服从幂律度分布的 Block Two-Level Erdös-Rényi（BTER）模型，该模型可以生成可调聚集系数的具有社区结构的复杂网络。

BTER 算法将一些度基本相同的节点构成的子图作为社区，社区内部构成了 ER 图。BTER 并不能生成模块性可以调整的复杂网络。本文对具有社区结构的无标度复杂网络进行建模，给出了一种具有聚集系数和模块性可调整的无标度网络生成算法（Scale Free Network with Tunable Clustering Coefficient and Modularity，TCMSN）。该算法首先生成符合幂律分布的参考度序列和参考社区规模序列；参考 BTER 过程构造社区内部的边；通过混合参数控制社区内外边的比例，进而对网络的模块性进行调整，保证了所构造网络的多样性。TCMSN 算法也可以构造与真实网络具有相同度序列的复杂网络，且能确保网络聚集系数和模块性也与真实网络接近。

2　图的相关知识

自然界存在大量可以表示成网络（图）$G=(V,E)$ 的复杂系统，其中 $V=\{v_1,v_2,\cdots,v_N\}$ 是网络 G 的节点集，表示复杂系统中研究对象的集合，令 $N=|V|$；E 是网络 G 的边集，表示研究对象之间的关系，令 $L=|E|$。除非特别声明，本文仅考虑无向简单图。通常用邻接矩阵 $A_{N\times N}$ 表示图 G，其中：

$$a_{ij}=\begin{cases}1, & \text{若}(v_i,v_j)\in E\\ 0, & \text{若}(v_i,v_j)\notin E\end{cases}$$

令 $N_G(v_i)=\{u|(u,v_i)\in E\}$ 表示图 G 中节点 v_i 的邻域，则节点 v 的度数 $d(v_i)=|N_G(v_i)|$，简记为 d_i。图 G 中所有的节点度按降序（或升序）排列可得到图 G 的度序列，记作 $(d_1,d_2,\cdots,d_N)$。图 G 的平均度记作 $d_{avg}=\frac{1}{N}\sum_i^N d_i$；图 G 的节点子集 S 导出的子图 $G[S]$ 是由节点集 S 和 G 中连接 S 中节点的边所构成的子图，即 $V(G[S])=S$ 且 $E(G[S])=\{(u,v)|(u,v)\in E,u\in S,v\in S\}$。在不引起混淆的情况下，将 $G[S]$ 简记为 S。

聚集系数又称为传递性，用来衡量一个网络的集团化程度。对于一个无向简单图 $G=(V,E)$，节点 v 的聚集系数 $CC(v)$ 可以用式(1)表示：

$$CC(v)=\begin{cases}\dfrac{2|E(G[N_v])|}{d_v(d_v-1)}, & \text{若 } d_v>1\\ 0, & \text{否则}\end{cases}\tag{1}$$

聚集系数刻画节点 v 邻域的导出子图的边密度，聚集系数越大，则 v 邻域的导出子图越接近完全图。聚集系数也可以看作图 G 中节点 v 邻域中包含 v 的完全图 K_3（也可称为三角形）的数目。

网络 G 的聚集系数由式(2)定义：

$$CC=\frac{1}{N}\sum_{i=1}^{N}CC(v_i)\tag{2}$$

通常，不同领域的复杂网络往往具有不同的聚集系数。表 1 给出了部分复杂网络实例的基本统计指标，可以看出，真实网络中聚集系数的变化范围比较大。因此，生成具有可调节聚集系数的复杂网络是网络生成模型的一项基本功能。可以通过调节局部节点的聚集系数来调节整个网络的聚集系数。若希望得到更大聚集系数的网络，则应适当增加网络中节点的三角形数目，反之亦然。

表 1 部分复杂网络实例的基本统计指标[22-30]

Table 1 Typical statistical indicators of some complex network instances

网络		类型	节点数	平均路径长度	聚集系数	度指数
社交网络	电影演员合作网络	无向图	449913	3.48	0.78	2.3
	电子邮件网络	有向图	59912	4.95	0.16	1.5/2.0
信息网络	万维网	有向图	269504	11.3	0.29	2.1/2.4
	词同现网络	无向图	460902	—	0.44	2.7
技术网络	电子电路网	无向图	24097	11.1	0.03	3
	对等网络	无向图	880	4.28	0.01	2.1
生物网络	代谢网络	无向图	765	2.56	0.67	2.1
	海洋食物网	有向图	135	2.05	0.23	—

社区结构(集簇结构)表示网络中节点的抱团性。通常用 2004 年 Newman 提出的模块度来衡量社区结构,如式(3)所示:

$$m=\frac{1}{2L}\sum_{v_i,v_j\in V}\left[a_{ij}-\frac{d_i d_j}{2L}\right]\sigma(\tau_i,\tau_j) \tag{3}$$

其中,$\sigma(\tau_i,\tau_j)$表示节点 i 和节点 j 是否在同一社区,即:

$$\sigma(\tau_i,\tau_j)=\begin{cases}1, & \tau_i=\tau_j\\ 0, & \tau_i\neq\tau_j\end{cases}$$

由式(3)可知,对于网络 G 的一个社区划分而言,如果社区内部的连边越稠密于社区之间的连边,则模块性越高。通常一个复杂网络的模块性 m 的取值范围为 0.3～0.7,m 值越大,网络的社区结构越明显。通过调整社区内部连边和社区间连边的比例,可以对网络的社区结构进行调节。

为了对所构造网络的社区结构进行度量,本文采用 Louvain 算法[31]对所研究的网络进行社区划分,以计算其模块性。

3 复杂网络生成算法 TCMSN

具有聚集系数和模块性可调整的无标度网络生成算法(Scale Free Network with Tunable Clustering Coefficient and Modularity,TCMSN)以用户给定的节点数、社区数和混合参数等作为输入参数,构造具有可调聚集系数和模块性的无标度网络。算法的主要过程为:首先生成服从幂律分布的参考度序列和参考社区规模序列,以保证网络的“无标度”特性;然后通过混合参数控制社区内外边的比例,以生成内部参考度序列和外部参考度序列,进而调节模块性;再以社区内部参考度序列为基础,将社区节点分块,块内节点以较高的概率进行连边,其余边按照一定的规则进行块间连边,并形成一个初始社区;最后以社区外部参考度序列为基础构造社区间连边。由此,便可生成具有可调聚集系数和模块性的无标度网络。

3.1 社区节点分配过程 Assignment

算法以网络节点数 N、参考度序列的幂指数 r_c、参考社区规模序列幂指数 r_d、混合参数 μ 等作为基本输入参数。首先,根据混合参数和参考度序列生成社区内部参考度序列和社区外部参考度序列;然后,对每个节点优先选择合适规模的社区进行分配。

由于一个节点的度数只可能是非负整数,而无标度网络的参考度序列中存在着大量度数很小的节点,取整操作通常会产生大量的社区内部度过小的节点,造成社区内部连边的比例远低于预期比例,因此采用式(4)修正由于度数取整操作而引起的小度节点社区内部边偏少的问题。以网络平均度 d_{avg} 为参考,若节点 v 的度数 $d_v<2\times d_{avg}$,则适当提高其内部边所占比例,并令其内部度为$\left\lceil d_v\times(1-\frac{2}{3}\mu)\right\rceil$。为了保证整体的社区内部边比例符合混合参数的预设,调整度数偏大的内部度为$\left\lceil d\times\frac{sum\times(1-\mu)-sum1\times(1-\frac{2}{3}\mu)}{sum-sum1}\right\rceil$,其中 sum 表示 N 个节点的度之和,$sum1$ 表示度小于 2 倍平均度的节点的度之和。

具体的节点分配如算法 1 所示。

算法 1 社区节点分配算法 Assignment

输入:节点数 N,最大度 d_{max},最小度 d_{min},参考度序列的幂指数 r_d,社区数 k,最小社区规模 D_{min},最大社区规模 D_{max},参考社区规模序列的幂指数 r_c,混合参数 μ

输出:社区节点集合 $V\{C_1,C_2,\cdots,C_k\}$

Step 1 根据 $p(d)=\frac{r_d-1}{d_{min}}(\frac{d}{d_{min}})^{-r_d}$ 生成符合幂指数 r_d 的 N 个节点的非递减的参考度序列$\{d_1,d_2,\cdots,d_N\}$。根据公式:

$$d_{in}=\begin{cases}\left\lceil d\times(1-\frac{2}{3}\mu)\right\rceil, & d<2\times d_{avg}\\ \left\lceil d\times\frac{sum\times(1-\mu)-sum1\times(1-\frac{2}{3}\mu)}{sum-sum1}\right\rceil, & d\geqslant 2\times d_{avg}\end{cases} \tag{4}$$

计算内部度序列为$\{d_{in}^1,d_{in}^2,\cdots,d_{in}^N\}$,并根据 $d_{out}=d-d_{in}$ 计算外部序列为$\{d_{out}^1,d_{out}^2,\cdots,d_{out}^N\}$。

Step 2 根据公式 $p(D)=\frac{r_c-1}{D_{min}}(\frac{D}{D_{min}})^{-r_c}$ 生成 k 个社区规模非递减的参考序列$\{D_1,D_2,\cdots,D_k\}$,且 $D_1<D_2<\cdots<D_k$,对应的社区为$\{C_1,C_2,\cdots,C_k\}$。

Step 3 按照社区规模从小到大为每个社区 $C_i(1\leqslant i\leqslant k)$分配节点。若当前社区规模为 D_i,则从尚未分配社区的节点中随机选择 D_i 个节点度小于当前社区规模的节点,并将其分配到社区 C_i 中。

Step 4 得到社区节点集合 $V\{C_1,C_2,\cdots,C_k\}$,结束。

为了生成具有明显社区结构的复杂网络,往往希望社区内部的连边相对紧密,社区间的连边相对稀疏。在算法 1 中,混合参数 μ 是指所生成网络中社区间边数占总边数的比例,μ 越小,社区内部边所占比例越大,社区结构越明显,其模块度值越大。因此,可以通过控制混合参数 μ 来调节网络的模块性。此外,混合参数 μ 的变化也会对网络的聚集系数产生影响,当 μ 增大时,社区内部连边减少,聚集系数会相应减小;当 μ 减小时,社区内部连边增多,聚集系数会相应增大。

3.2 社区内部边的构造过程 IntraLink(C_i)

以社区内部参考度序列为基础进行社区内部的连接。为了得到可调节聚集系数的网络生成算法,首先将社区内部的节点进行分块,分别进行块内连边和块间连边;然后通过调节块内连边的概率实现聚集系数的调整。

式(5)给出了块内连边的概率公式,增大参数 a 可以提高社区内部连边的概率,进而较大幅度地增大网络的聚集系数;反之,降低参数 a 的取值也可以较大幅度地减小网络的聚集系数。由于无标度网络中存在着大量度数偏小的节点,对小

度节点连边的概率进行调节，可以使得整个网络的聚集系数得到小幅度调整，降低参数 b 的取值可以小幅度增大网络聚集系数，而增大参数 b 的取值则可以小幅度减小网络聚集系数。通常参数 a 的取值范围是 $0<a\leqslant 1$；参数 b 的取值应保证连接概率 p 非负，本文中一般取 $0.05\leqslant b\leqslant 1.5$。

具体构造过程如算法 2 所示。

算法 2　社区 C_i 内部边的构造 IntraLink(C_i)

输入：社区 C_i 的内部参考度序列 $\{d_{in}^1, d_{in}^2, \cdots, d_{in}^{D_i}\}$，参数 a 和 b

输出：边集 $E(C_i)$

Step 1　将社区 C_i 内的节点按参考内部度序列非递减排序，即 $d_{in}^1\leqslant d_{in}^2\leqslant\cdots\leqslant d_{in}^{D_i}$，则社区 C_i 内部的最大度 $d_{in}^{max}=d_{in}^{D_i}$。

Step 2　从度为 2 的节点开始，将排好序的社区内节点划分为 x 块，除最后一个块外，每个块有 $d_{b\ min}^j+1(1\leqslant j<x)$ 个节点，其中 $d_{b\ min}^j$ 为第 j 个块中首节点的内部度，即第 j 个块的最小内部度。

Step 3　构造第 $j(1\leqslant j<x)$ 块的内部连边。选择块内任意两点，按概率

$$p=a\left[1-b\left(\frac{\log(d_{b\ min}^j+1)}{\log(d_{in}^{max}+1)}\right)^3\right] \tag{5}$$

进行连边。

Step 4　计算社区 C_i 中每个节点 v 的剩余内部度 $d'_{in}(v)$。

Step 4.1　计算当前社区 C_i 的平均内部剩余度 $avgd_{in}$。

Step 4.2　随机选择 C_i 中的社区内部度 $d'_{in}(v)\geqslant 2\times avgd_{in}$ 的节点 v，如果不存在这样的节点，则随机选择 $d'_{in}(v)>0$ 的节点 v。

Step 4.3　随机选择 C_i 中与 v 相异的 $d'_{in}(w)>0$ 且 w 与 v 间尚未连边的节点 w，添加边 wv，即 $E(G)=E(G)\cup\{(w,v)\}$。更新内部剩余度 $d'_{in}(w)=d'_{in}(w)-1$，$d'_{in}(v)=d'_{in}(v)-1$。

Step 4.4　若社区 C_i 中所有节点的内部剩余度 $d'_{in}(v)=0$ 或非零内部剩余度节点之间已无法添加新边，则转 Step 5；否则转 Step 4.1。

Step 5　对于剩余内部度 $d'_{in}(v)>0$ 的节点 v，更新其外部参考度 $d_{out}(v)=d_{out}(v)+d'_{in}(v)$。

Step 6　结束。

算法 2 的 Step 4 根据社区内部剩余度对块间添加边，由于内部剩余度序列基本符合幂律分布，因此存在大量度数偏小的节点。如果随机选择两个节点进行连接，可能使小度节点之间的连接过早饱和，而序列中的大度节点在社区内部仅有很小的连接可能。为了不破坏初始的参考度序列，会将一部分内部连接转换成外部连接，使得网络的混合参数发生较大变化。

为了避免上述情况，Step 4 中优先选择内部剩余度偏大的节点进行连接，在不破坏初始参考度序列的情况下，最大可能地维持网络的混合参数。

3.3　社区间连边的构造过程 InterLink

在算法 2 构造了每个社区 $C_i(1\leqslant i\leqslant k)$ 的内部连接之后，以网络节点已更新的外部参考度序列 $\{d_{out}^1, d_{out}^2, \cdots, d_{out}^N\}$ 为输入，添加社区间连接。具体过程如算法 3 所示。

算法 3　社区间连边的构造 InterLink

输入：图 G 的外部参考度序列 $\{d_{out}^1, d_{out}^2, \cdots, d_{out}^N\}$，$\bigcup_{1\leqslant i\leqslant k} E(C_i)$

输出：E(G)

Step 1　计算当前网络 G 的平均外部剩余度 $avgd_{out}$。

Step 2　随机选择 G 中外部度 $d'_{out}(v)\geqslant 2\times avgd_{out}$ 的节点 v，若不存在这样的节点，则随机选择 $d'_{out}(v)>0$ 的节点 v。

Step 3　随机选择 G 中与 v 相异的 $d'_{out}(w)>0$ 且 w 与 v 间尚未连边的节点 w，添加边 wv，即 $E(G)=E(G)\cup\{(w,v)\}$。更新外部剩余度 $d'_{out}(w)=d'_{out}(w)-1$，$d'_{out}(v)=d'_{out}(v)-1$。

Step 4　若 G 中所有节点的外部剩余度 $d'_{out}(v)=0$ 或非零外部剩余度节点之间已无法添加新边，则结束；否则转 Step 1。

在算法 3 的步骤 4 中，当图 G 中所有节点的外部剩余度为 0 时，会得到一个符合初始参考度序列的没有重边和自环的简单图。然而，在大多数情况下，算法结束时会存在一些非零外部剩余度节点，为了确保最终网络是简单图，这些剩余度将被舍弃。

由于幂律分布的统计特性，舍弃少量的剩余度不会破坏网络度序列的幂律分布特性。但是，如果最终舍弃的外部度偏多，则会引起度序列分布特征的改变。因此，在构造过程中应该保证最小可能地舍弃外部度。鉴于此，算法 3 优先选择了外部度偏大的节点进行连接，尽可能优先满足大度节点的连接。为了使所构造的网络具有结构多样性，在度数偏大的节点中随机选择一个作为新连接的一个端点，另外一个端点则在其他可用节点中随机选择。这样，既可以最小可能地改变参考度序列，又可以维持网络结构的多样性。

3.4　TCMSN 生成算法

综合算法 1—算法 3，下面给出具有社区结构的可调节聚集系数和模块性的复杂网络生成算法的基本框架，如算法 4 所示。

算法 4　TCMSN 生成算法

输入：节点数 N，最大度 d_{max}，最小度 d_{min}，参考度序列的幂指数 r_d，社区数 k，最小社区规模 D_{min}，最大社区规模 D_{max}，参考社区规模序列的幂指数 r_c，混合参数 μ，参数 a 和 b

输出：复杂网络 G=(V,E)

Step 1　执行社区节点分配算法 Assignment，生成幂律分布的参考度序列和参考社区规模序列，并将节点分配至社区，从而得到社区节点集合 $V\{C_1, C_2, \cdots, C_k\}$。

Step 2　对每个社区 $C_i(1\leqslant i\leqslant k)$ 执行社区内部边构造算法 IntraLink(C_i)，以构造社区内部连边。

Step 3　执行社区间连边的构造算法 InterLink(算法 3)，最终生成基本符合参考度序列和混合参数的简单图 G。

4　实验分析

本文所提算法 TCMSN 能够通过式(5)给出的社区块内连接概率来调节网络的聚集系数，增大(或减小)参数 a，可较大幅度地提高(或降低)网络的聚集系数；增大(或减小)参数 b，可较小幅度地降低(或提高)网络的聚集系数。通过调整混合参数 μ 可以调节网络的模块性，参数 μ 越小，社区间连边所占比例越小，社区结构越明显；反之亦然。同样，混合参数 μ 增大(或减小)，网络的聚集系数会降低(或提高)。

4.1　TCMSN 算法聚集系数和模块性调节性实验

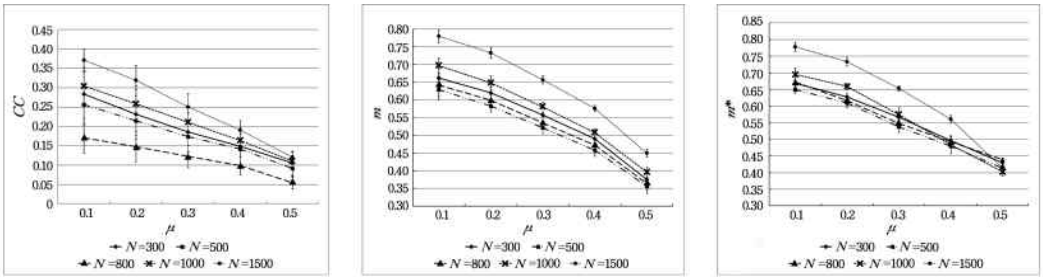
为了验证本算法对聚集系数和模块性的调节性，首先在给定参数(见表 2)下分别生成节点数为 300，500，800，1000 和 1500 的 5 个符合幂律分布的参考度序列和参考社区规模序列。分别通过改变参数 a，b，μ 的取值，生成相应的复杂网络模型，实验结果如图 1—图 3 所示。

表 2 数据参数

Table 2 Parameters of data

N	$d_{\max}$	$d_{\min}$	r_d	K	$D_{\max}$	$D_{\min}$	r_c
300	50	3	3	4	150	40	1.5
500	60	3	3	4	250	40	1.5
800	60	3	3	4	400	40	1.5
1000	80	3	3	6	400	40	1.5
1500	80	3	3	10	400	60	1.5

图 1 给出了在 a 和 b 固定的情况下，利用本文算法 TCMSN 所生成的复杂网络的聚集系数(见图 1(a))和模块性(见图 1(b))随参数 μ 变化的情况。为了验证初始社区划分结果的合理性，采用经典社区划分算法 Louvain 对所研究的网络进行社区划分，并计算相应的模块性(见图 1(c))。可以看出，网络的聚集系数和模块性都随着 μ 的增大而减小。

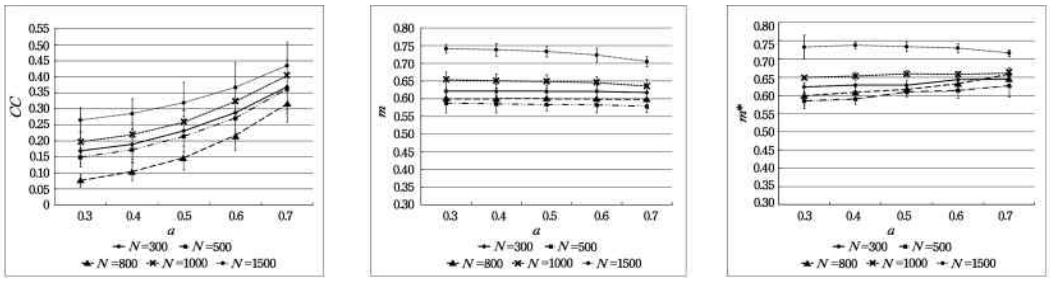


(a)CC (b)m (c)m^*

图 1 $a=0.5$，$b=0.5$ 时不同参数 μ 的聚集系数和模块性

Fig. 1 Clustering coefficient and modularity for different μ while $a=0.5$ and $b=0.5$

图 2 给出了在 μ 和 b 固定的情况下，利用本文算法 TCMSN 所生成的复杂网络的聚集系数(见图 2(a))和模块性(见图 2(b))随参数 a 变化的情况。可以看出，随着参数 a 的增大，网络的聚集系数也有较大幅度的提高(图 2(a))。当混合参数 μ 固定不变时，网络模块性基本保持稳定(图 2(b))。图 2(c)给出了 Louvain 算法社区划分后的模块性计算结果。

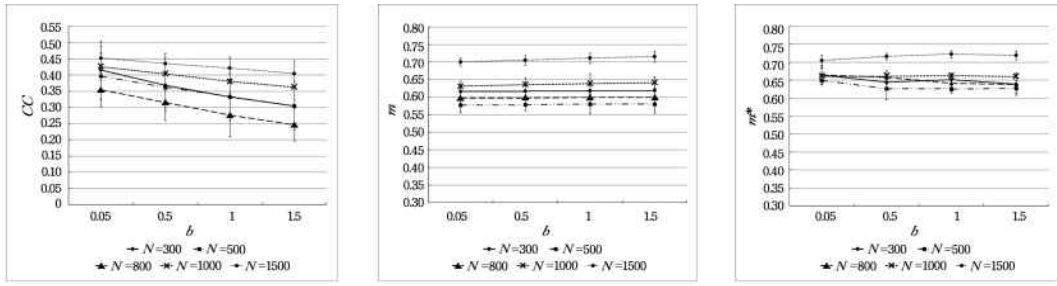


(a)CC (b)m (c)m^*

图 2 $\mu=0.2$，$b=0.5$ 时不同参数 a 的聚集系数和模块性

Fig. 2 Clustering coefficient and modularity for different a while $\mu=0.2$ and $b=0.5$

图 3 给出了 μ 和 a 固定的情况下，利用本文算法 TCMSN 所生成的复杂网络的聚集系数(见图 3(a))和模块性(见图 3(b))随参数 b 变化的情况。可以看出，随着参数 b 的增大，网络的聚集系数有小幅度降低(图 3(a))。此时，网络模块性基本保持稳定(图 3(b))。图 3(c)给出了 Louvain 算法社区划分后的模块性计算结果。



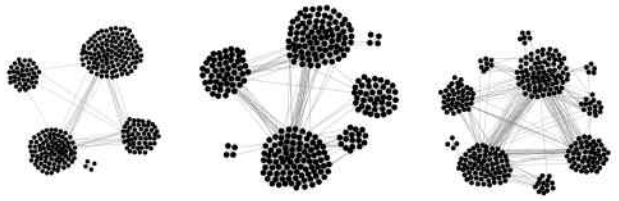
(a)CC (b)m (c)m^*

图 3 $a=0.7$，$\mu=0.2$ 时不同参数 b 的聚集系数和模块性

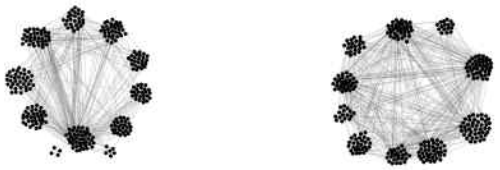
Fig. 3 Clustering coefficient and modularity for different b while $a=0.7$ and $\mu=0.2$

图 4 进一步比较了 TCMSN 算法与经典算法 Louvain 在混合参数 μ 变化时的社区划分结果。TCMSN 算法的初始参数设定如下：社区数为 $k=4$，网络节点数 $N=300$，$a=0.5$，$b=0.5$。当混合参数 $\mu=0.1$ 时，二者的社区划分结果基本相同，模块度值也比较接近；随着混合参数 μ 的逐渐增大，TCMSN 算法的社区结构逐渐模糊，Louvain 算法也很难划分出

与 TCMSN 算法相似的社区结构。表 3 给出了本文算法与 Louvain 算法的具体比较结果，其中模块度值 m 是按照 TCMSN 算法的初始社区划分计算所得的；模块度值 m^* 由 Louvain 算法的社区划分计算所得。可以看出，二者的模块性随着混合参数的增长逐步下降，社区结构变得逐渐模糊。



(a)$\mu=0.1, m^*=0.668$　(b)$\mu=0.2, m^*=0.627$　(c)$\mu=0.3, m^*=0.566$



(d)$\mu=0.4, m^*=0.496$　(e)$\mu=0.5, m^*=0.429$

图 4　$N=300, a=0.5, b=0.5$ 时不同参数 μ 下 Louvain 算法的运行结果示例

Fig. 4　Results of Louvain algorithms for different μ while $N=300$, $a=0.5$ and $b=0.5$

表 3　$N=300, a=0.5, b=0.5$ 时不同参数 μ 下 TCMSN 与 Louvain 算法的模块性比较结果

Table 3　Comparison of TCMSN and Louvain in modularity for different μ while $N=300$, $a=0.5$ and $b=0.5$

N	L	a	b	μ	CC	m	m^*
				0.1	0.283±0.062	0.662±0.027	0.668±0.011
				0.2	0.231±0.064	0.620±0.016	0.627±0.014
300	806	0.5	0.5	0.3	0.185±0.049	0.557±0.018	0.566±0.013
				0.4	0.148±0.049	0.490±0.020	0.496±0.017
				0.5	0.105±0.031	0.375±0.021	0.429±0.012

4.2　仿真实验

本文选择了 Dolphins，polBooks，ca-AstroPH，cit-HepPH 和 ca-CondMat 5 个数据集进行对比实验，相关参数的设置如表 4 所列。根据真实网络数据提取相应的度序列，因此没有对参考度序列的幂指数进行设置。

表 4　TCMSN 的参数设置

Table 4　Parameter setting of TCMSN

name	N	k	$D_{\min}$	$D_{\max}$	r_c	μ	a	b
Dolphins	62	4	7	25	1.5	0.25	0.5	0.1
polBooks	105	4	10	41	1.5	0.25	0.85	0.1
ca-AstroPH	18772	32	40	1700	1.5	0.3	0.995	0.01
cit-HepPH	34546	20	40	4700	1.5	0.25	0.740	0.1
ca-CondMat	23133	46	40	1500	1.5	0.3	0.985	0.05

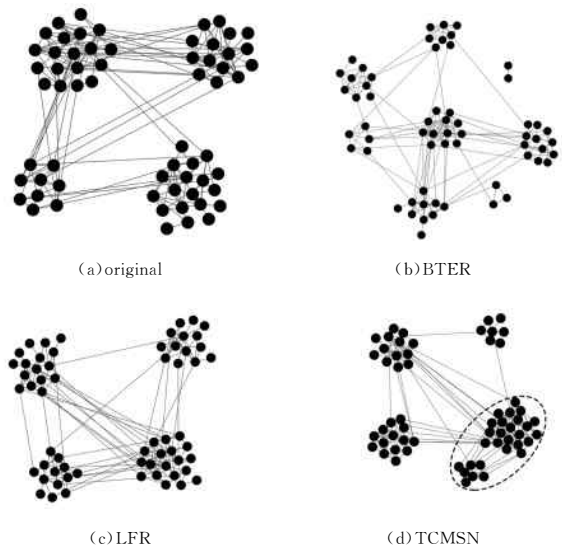
选择具有社区结构的复杂网络生成算法 BTER[20-21] 和 LFR[18] 作为比较对象，表 5 给出了本文算法 TCMSN 与 BTER 和 LFR 在表 4 所列参数设置下所生成网络的基本指标。可以看出，TCMSN 算法产生的网络节点数、边数、最大度等都与真实网络基本保持一致，说明本文算法在社区内部块间连边和社区之间连边时，优先选择度数偏大的节点作为新边端点，最大程度地维持了初始度序列，不破坏初始度序列的分布特性。通过合理调整参数 a，b 和 μ，TCMSN 算法所构造网络的聚集系数和模块性也可与真实网络基本保持一致。

表 5　真实数据集上 BTER，LFR，TCMSN 的比较

Table 5　Comparison of BTER, LFR and TCMSN on real data set

name	model	N	L	$d_{\max}$	CC	m^*
Dolphins	original	62	159	12	0.259	0.526
	BTER	61	165	15	0.361	0.581
	LFR	62	159	12	0.287	0.535
	TCMSN	**62**	**159**	**12**	**0.271**	**0.522**
polBooks	original	105	441	25	0.488	0.519
	BTER	125	564	24	0.530	0.625
	LFR	105	**441**	25	0.309	**0.523**
	TCMSN	**105**	438	**25**	**0.497**	0.532
ca-AstroPH	original	18772	198050	504	0.631	0.621
	BTER	18679	205594	454	0.645	0.750
	LFR	18772	197608	504	0.165	**0.632**
	TCMSN	**18772**	**198050**	**504**	**0.620**	0.644
cit-HepPH	original	34546	420877	846	0.285	0.707
	BTER	34496	469327	841	0.293	0.626
	LFR	34546	420875	846	0.098	0.695
	TCMSN	**34546**	**420877**	**846**	**0.284**	**0.698**
ca-CondMat	original	23133	93439	279	0.633	0.719
	BTER	23133	94689	265	0.642	0.726
	LFR	23133	93431	279	0.075	0.726
	TCMSN	**23133**	**93439**	**279**	**0.633**	**0.713**

图 5 给出了本文算法 TCMSN 与 BTER 算法和 LFR 算法在 Dophins 数据集上的生成模型的比较结果。



(a)original　(b)BTER

(c)LFR　(d)TCMSN

图 5　Louvain 算法的社区划分(Dolphins 数据集)

Fig. 5　Community division under Louvain algorithm(Dolphins data set)

如图 5(a)所示，原始网络构成了 4 个社区。采用 Louvain 算法分别对 BTER 生成网络、LFR 生成网络和 TCMSN 生成网络进行社区划分，结果如图 5(b)－图 5(d)所示。尽管 Louvain 算法对 TCMSN 网络进行划分后得到了 5 个社区，但图 5(d)中虚线框内的两个社区在 TCMSN 网络生成过程中

被实际标识为同一社区。因此，TCMSN 生成网络在社区结构上与真实网络更接近。

图 6 给出了本文算法 TCMSN 与算法 BTER 和算法 LFR 在 polBooks 数据集上的生成模型的比较结果。同样可以看出，相较于 BTER，本文算法 TCMSN 给出的社区结构更接近于实际网络的模型。

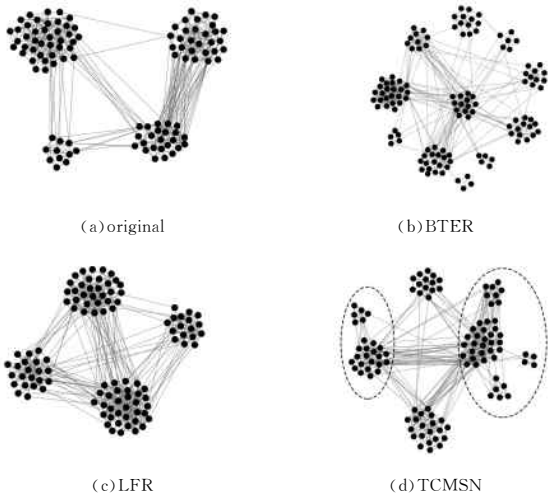


图 6 Louvain 算法的社区划分(polBooks 数据集)

Fig. 6 Community division under Louvain algorithm(polBooks data set)

结束语 本文给出了一种具有社区结构的可调节聚集系数和模块性的无标度网络生成算法——TCMSN，该算法以用户给定的节点数、社区数、混合参数、参考度序列幂指数和参考社区序列幂指数等作为输入参数，通过社区节点分配、社区内边的构造和社区间边的构造 3 个过程，保证了所生成的网络最大程度地维持了网络的无标度特性，同时确保了生成网络的结构多样性。基于人工构造数据和真实网络数据的对比实验结果表明，TCMSN 算法能够生成可调节聚集系数和模块性的无标度网络模型，并且能够生成最接近真实网络社区结构特征的网络模型。

合理构造符合复杂网络特性的生成图模型，不仅为研究复杂系统的功能提供了模型基础，也可为相关算法提供数据基础。随着复杂网络的不断产生与丰富，如何构造符合不同领域(如社交网络、生物网络、信息网络、技术网络)特征的复杂网络是一个新的挑战；另外，重叠社区结构是复杂网络的一个重要特征，如何对具有重叠社区结构的复杂网络建模是一个亟待研究的课题。总之，为了更好地研究复杂系统的功能和性质，仍然需要改进或者发展更好的生成图模型。

参 考 文 献

[1] YANG B, LIU D Y, JIN D, et al. Complex network clustering algorithms[J]. Journal of Software, 2009, 20(1): 54-66. (in Chinese)
杨博，刘大有，金弟，等. 复杂网络聚类方法[J]. 软件学报，2009，20(1)：54-66.

[2] WANG J, LIANG J Y, ZHENG W P. A graph clustering method for detecting protein complexes [J]. Journal of Computer Research and Development, 2015, 52(8): 1784-1793. (in Chinese)
王杰，梁吉业，郑文萍. 一种面向蛋白质复合体检测的图聚类方法[J]. 计算机研究与发展，2015，52(8)：1784-1793.

[3] XING Y K, MA S P. A clustering algorithm based on markov chain models [J]. Journal of Computer Research and Development, 2003, 40(2): 129-135. (in Chinese)
邢永康，马少平. 一种基于 Markov 链模型的动态聚类方法[J]. 计算机研究与发展，2003，40(2)：129-135.

[4] ZHOU S G, ZHOU A Y, CAO J, et al. A fast density-based clustering algorithm [J]. Journal of Computer Research and Development, 2000, 37(11): 1287-1292. (in Chinese)
周水庚，周傲英，曹晶，等. 一种基于密度的快速聚类算法[J]. 计算机研究与发展，2000，37(11)：1287-1292.

[5] 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.

[6] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks [J]. Nature, 1998, 393(6684): 440-442.

[7] BARABÁSI A, ALBERT R. Emergence of scaling in random networks [J]. Science, 1999, 286(5439): 509-512.

[8] HOLME P, KIM B J. Growing scale-free networks with tunable clustering [J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2002, 65(2): 26017.

[9] DOROGOVTSEV S N, MENDES J F F. Evolution of networks [J]. Advances in Physics, 2001, 51(4): 1079-1187.

[10] LIU J G, DANG Y Z, WANG Z T. Multistage random growing small-world networks with power-law degree distribution[J]. Chinese Physics Letters, 2006, 23(3): 746-749.

[11] ZHANG L H. Simulation and application researchers on complex network modeling [D]. Dalian: Dalian University of Technology, 2013. (in Chinese)
张兰华. 复杂网络建模的仿真与应用研究[D]. 大连：大连理工大学，2013.

[12] CUI A X. Research on modeling and spreading dynamics of complex networks[D]. Chengdu: University of Electronic Science and Technology of China, 2014. (in Chinese)
崔爱香. 复杂网络建模及其传播动力学研究[D]. 成都：电子科技大学，2014.

[13] LIU S J. Construction of complex network models and analysis of their properties [D]. Chengdu: Southwest Jiaotong University, 2015. (in Chinese)
刘胜久. 复杂网络模型构建及特性分析[D]. 成都：西南交通大学，2015.

[14] FRONCZAK A, FRONCZAK P, HOŁYST J A. Mean-field theory for clustering coefficients in Barabási-Albert networks [J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2003, 68(4): 046126.

[15] GIRVAN M，NEWMAN M E J. Community structure in social and biological networks [J]. Proceedings of the National Academy of Sciences of the United States of America，2002，99(12)：7821-7826.

[16] NEWMAN M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics，2004，69(6 Pt 2)：066133.

[17] NEWMAN M E J，GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E Statistical Nonlinear &Soft Matter Physics，2004，69(2)：026113.

[18] LANICHINETTI A，FORTUNATO S，RADICCHI F. Benchmark graphs for testing community detection algorithms [J]. Physical Review E Statistical Nonlinear & Soft Matter Physics，2008，78(4)：046110.

[19] MOLLOY M，REED B. A critical point for random graphs with a given degree sequence [J]. Random Structures & Algorithms，1995，6(2-3)：161-180.

[20] SESHADHRI C，KOLDA T G，PINAR A. Community structure and scale-free collections of Erdös-Rényi graphs[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics，2012，85(5)：056109.

[21] KOLDA T G，PINAR A，PLANTENGA T，et al. A scalable generative graph model with community structure[J]. Siam Journal on Scientific Computing，2014，36(5)，C424-C452.

[22] AMARAL L A，SCALA A，BARTHELEMY M，et al. Classes of small-world networks [J]. Proceedings of the National Academy of Sciences of the United States of America，2000，97(21)：11149-11152.

[23] ECKMANN J P，MOSES E. Curvature of co-links uncovers hidden thematic layers in the World Wide Web [J]. Proceedings of the National Academy of Sciences of the United States of America，2002，99(9)：5825-5829.

[24] BARABÁSI A L，ALBERT R，JEONG H. Scale-free characteristics of random networks：the topology of the world-wide Web [J]. Physica A Statistical Mechanics & Its Applications，2000，281(1-4)：69-77.

[25] ALBERT R，JEONG H. Diameter of the World Wide Web [J]. Nature，1999，401(6)：130-131.

[26] DOROGOVTSEV S N，MENDES J F. Language as an evolving word web[J]. Proceedings Biological Sciences，2001，268(1485)：2603-2606.

[27] RIPEANU M，FOSTER I，IAMNITCHI A. Mapping the gnutella network：properties of Large-Scale Peer-to-Peer systems and implications for system design [J]. IEEE Internet Computing，2002，6(1)：50-57.

[28] JEONG H，TOMBOR B，ALERT R，et al. The large-scale organization of metabolic networks [J]. Nature，2000，407(6804)：651-654.

[29] HUXHAM M，RAFFAELLI D. Do parasites reduce the chances of triangulation in a real food web [J]. Oikos，1996，76(2)：284-300.

[30] CANCHO R F I，JANSSEN C，SOLÉ R V. Topology of technology graphs：small world patterns in electronic circuits [J]. Physical Review E，2001，64(4)：046119.

[31] BLONDEL V D，GUILLAUME J L，LAMBIOTTE R，et al. Fast unfolding of communities in large networks [J]. Journal of Statistical Mechanics Theory & Experiment，2008，2008(10)：155-168.

(上接第 56 页)

[14] SOXT E. Ethereum[M]. Wiesbaden：Springer Fachmedien Wiesbaden，2017.

[15] CHANDRAN N，GROTH J，SAHAI A. Ring signatures of sublinear size without random oracles[C] // International Colloquium on Automata，Languages，and Programming. Springer，2007：423-434.

[16] ZHANG Y Q，WANG X F，LIU X F，et al. Survey on Cloud Computing Security[J]. Journal of Software，2010，27(6)：1328-1348.(in Chinese)
张玉清，王晓菲，刘雪峰，等. 云计算环境安全综述[J]. 软件学报，2010，27(6)：1328-1348.

[17] CASTRO M，LISKOV B. Practical Byzantine Fault Tolerance and Proactive Recovery[J]. ACM Transactions on Computer Systems，2002，20(4)：398-461.

[18] YUAN Y，WANG F Y. Blockchain：The State of the Art and Future Trends[J]. Acta Automatica Sinica，2016，42(4)：481-494.(in Chinese)
袁勇，王飞跃. 区块链技术发展现状与展望[J]. 自动化学报，2016，42(4)：481-494.

[19] CHEN H，WEI S M，ZHU C J，et al. Security Certificateless Aggregate Signature Scheme[J]. Journal of Software，2015，26(5)：1173-1180.(in Chinese)
陈虎，魏仕民，朱昌杰，等. 安全的无证书聚合签名方案[J]. 软件学报，2015，26(5)：1173-1180.

[20] LU H J，YU X Y，XIE Q. Provably Secure Certificateless Aggregate Signature with Constant Length[J]. Journal of Shanghai Jiaotong University，2012，46(2)：259-263.(in Chinese)
陆海军，于秀源，谢琪. 可证安全的常数长度无证书聚合签名方案[J]. 上海交通大学学报，2012，46(2)：259-263.