基于零模型的社区检测基准网络构造及应用

任宏菲[1,2]，肖 婧[1*]，崔文阔[1]，许小可[1,2]

(1. 大连民族大学信息与通信工程学院 辽宁 大连 116600；2. 贵州省公共大数据重点实验室(贵州大学) 贵阳 550025)

【摘要】社区检测对于探索挖掘复杂网络的结构特性具有重要意义，社区检测算法性能对于检测结果具有重要影响。目前用于衡量社区检测算法性能的基准测试网络较为单一，主要包括人工合成网络和真实世界网络。由于真实世界网络中通常缺乏已知社区结构信息，人工合成网络成为衡量算法性能的主要途径，但普遍存在网络微观特性不可调且与真实世界网络差异较大、对检测算法区分度不高、无法更改局部网络结构等问题。为提升人工合成网络性能，该文提出基于零模型的基准测试网络构造方法，首先设计了能够保持中尺度特性的零模型，提升网络微观特性调整灵活度，使其更逼近真实世界网络结构特性；其次设计了能够调整社区结构强弱的零模型，提升网络社区检测的评价准确性；最后设计了能够调整局部拓扑结构的零模型，有效衡量局部社区结构特性变化对于整体网络结构及检测算法性能的重要性。实验结果表明，基于零模型的构造方法能够有效提升基准测试网络的多样性和灵活性，更加逼近真实世界网络特性，因此更能满足对于社区检测算法性能的评价需求，对于提升复杂网络社区检测性能具有重要意义。

关 键 词 社区检测； 复杂网络； 零模型； 基准网络

中图分类号 TP391 **文献标志码** A doi:10.3969/j.issn.1001-0548.2019.03.021

Construction and Applications of Benchmark Networks for Community Detection Based on Null Models

REN Hong-fei[1, 2], XIAO Jing[1*], CUI Wen-kuo[1], and XU Xiao-ke[1, 2]

(1. College of Information and Communication Engineering, Dalian Minzu University Dalian Liaoning 116600;
2. Guizhou Provincial Key Laboratory of Public Big Data, Guizhou University Guiyang 550025)

Abstract Community detection is of great significance for exploring the structural characteristics of complex networks while the performance of community detection algorithm makes important influence on the detection results. At present, the benchmark networks that are used to measure the performance of community detection algorithm mainly include artificial synthetic network and real-world network. Synthetic network has become the main method to measure the performance of the algorithm since the real-world network usually lacks information of known community structure. However, it is found that the microscopic characteristics of the network is unadjusted, which is different from the real-world network, the discrimination of the detection algorithm is not high, and it is inability to change the local network structure. In order to improve the performance of artificial synthetic network, a benchmark network construction algorithm on null-model is proposed. Firstly, an algorithm of null model that can maintain the mesoscale characteristics is built to improve the flexibility of network micro-feature adjustment and make it closer to the real-world network structural characteristics. Secondly, the null model of adjusting strengthen and weakness for community structure is designed for improving the evaluation accuracy of network community testing. Finally, a method based on null model is constructed so as to make some adjustments of the local topological structure for measuring the importance of the change with local community structure characteristics to the whole network structure and the performance on detection algorithm. Experimental results show that the algorithm in view of null model can effectively improve the diversity and flexibility of the benchmark network, thus making the network be more similar with the features of real-world network and meeting the demand for performance improvement of community detection algorithm.

Key words community detection; complex network; null model; synthetic network

收稿日期：2018 – 05 – 27；修回日期：2018 – 09 – 19

基金项目：国家自然科学基金(61773091, 61603073)；辽宁省自然科学基金(201602200)；辽宁省高等学校创新人才支持计划(LR2016070)；辽宁省重点研发计划指导计划项目(2018104016)

作者简介：任宏菲(1993 –)，女，主要从事复杂网络社区检测方面的研究.

通信作者：肖婧，Email: hrbeuxiaojing@aliyun.com

社区检测是指从复杂网络中挖掘出有实际意义的模块或层次结构的过程，能够更加深刻地认识和分析复杂系统中不同个体间的多层次联系[1]。社区检测提取出的网络结构特征，有助于理解和分析网络的拓扑特性、功能特性及动力学特性等，进而挖掘出网络中蕴含的深层次信息，并对网络行为进行预测[2]，因此具有重要的理论研究价值。此外，社区检测能够广泛应用于互联网基础设施优化[3]、在线销售产品推荐[3-4]、社交网络及生物网络分析等领域[5-6]，因此对其进行研究具有重要的现实意义。

针对社区检测算法的研究已成为近年来的研究热点，众多高效的社区检测算法被相继提出，典型代表包括派系过滤法[7]、链接聚类法[8]、局部扩展法[9-10]、模块度优化法[11-13]和标签传播法[14]等。然而，现有研究中能够用于衡量社区检测算法性能的基准测试网络较为缺乏，主要包括人工合成网络和真实世界网络两类。真实网络中一般仅有空手道俱乐部网络、海豚网络等少数具有较明确的社团信息，由于真实世界网络中通常缺乏已知的社区结构信息，无法准确评价社区检测的精确性，因此人工合成网络成为目前衡量算法性能的主要途径。社区检测研究中的人工合成网络主要包括GN网络[15]和LFR网络[16]，均存在微观特性不可调且与真实世界网络特性差异较大、对算法精确性和稳定性的评价能力较差且区分度不高、无法更改网络局部结构特性等问题，降低了对社区检测算法的评价能力，也在一定程度上制约了社区检测算法的研究。

为提升社区检测研究中基准测试网络的性能，本文提出了基于零模型的网络构造方法，并在此基础上设计了3种不同类型的人工合成基准测试网络。首先，通过设计能够保持中尺度特性(主要为社区结构特性)的零模型，以提升对于网络微观特性的调整灵活度，使其更加逼近真实世界网络的结构特性。其次，通过设计能够调整社区结构强弱的零模型，提升网络社区检测的评价准确性，即在网络社区结构模糊程度特性变化情况下，更加精确衡量社区检测算法的精确性和稳定性。最后，通过设计能够调整局部拓扑结构的零模型，有效衡量局部社区结构特性变化对于整体网络结构及检测算法性能的重要性，提升对于单个社区结构特性的分析能力。实验结果表明，基于零模型的网络构造方法，能够有效提升人工合成基准测试网络的多样性和灵活性，更加逼近真实世界网络特性，因此更能够满足对于社区检测算法性能的评价需求，对于促进复杂网络社区检测研究具有重要意义。

1　社区检测基准测试网络及局限性

现有社区检测研究中的基准测试网络类型较为单一，主要包括真实世界网络和人工合成网络两类。

1.1　真实世界测试网络

目前应用较为广泛的真实世界网络主要包括9种，具体如表1所示，其中Karate[17]、Dolphins[18]、Polbooks[19]和Football[19]是4种最典型的小规模测试网络，剩余5种为中等规模至大规模的测试网络。

表1　真实世界网络信息

网络	节点数量/个	连边数量/条
Karate	34	78
Dolphins	62	159
Polbooks	105	441
Football	115	613
C.elegans	453	2 040
Email	1 133	5 451
Erdös	6 927	11 850
PGP	10 680	24 316
Cond-Mat	27 519	116 181

由于真实世界网络中通常缺乏已知的社区结构信息，无法准确评价社区检测的精确性，因此通常采用模块度指标衡量检测所得社区划分质量。模块度函数[20]如下：

$$Q=\sum_{v=1}^{n_c}\left[\frac{l_v}{M}-\left(\frac{d_v}{2M}\right)\right] \tag{1}$$

式中，M是网络中所有连边数总和；n_c表示社区的数量；l_v是社区v内部所包含的边数；d_v是社区v中所有节点的度值之和。

模块度函数的基本思想是把社区划分后的网络与相应的1阶零模型进行比较。模块度值越高，说明社区结构越显著，社区检测质量越高；反之，模块度值越低，说明社区结构越模糊，社区检测质量越低。然而，越来越多的研究表明，模块度函数作为社区检测质量的评价标准也存在一定缺陷。首先，在随机网络研究中有时候也会得到较大的模块度值，所以模块度值高低不能完全代表社区划分质量。其次，在研究较大规模的网络时，发现模块度存在分辨率限制问题，即不易发现规模相对较小的社区。最后，模块度在数学理论上不够清晰，无法判断函数是单峰函数还是多峰函数[20-21]。

1.2　人工合成测试网络

相比真实世界网络，人工合成网络由于能够预

知真实的网络微观特性及社区划分，能够更有效衡量出检测所得社区划分的精确性。目前应用较广泛的人工合成网络主要包括GN网络和LFR网络两种。

GN网络的生成方式如下[15,22]：首先确定网络参数，包括节点数N，平均度K，社区数量C，节点与社区外部节点连边数目的期望值Z_{out}；然后根据以上参数将节点平均分成C个社区，保证每个社区的节点数和平均度相同；最后根据每个节点的社区归属信息及Z_{out}值大小随机构造连边，生成网络G。GN网络中每个社区中包含的节点数相同，网络的聚类特性和社区结构特性比较简单，一般与真实世界网络的拓扑结构特性相差较大。

LFR网络的生成方式如下[16,23]：首先确定网络参数，包括节点数N，平均度K，最大度k_{max}，混合参数μ，最大社区规模c_{max}，最小社区规模c_{min}，并根据参数生成度序列确定N个节点的度值；其次，在$[c_{min}, c_{max}]$范围内随机确定社区数目C，并将N个节点随机匹配到C个社区中；再次，根据配置模型算法随机选择N个节点中的任意节点对，构造各节点的社区内外部连边，保证网络的连通性；最后，根据网络连边信息及节点的社区归属信息生成网络G。相较于GN网络，LFR网络的节点度序列和社区规模序列服从幂律分布，更符合真实世界网络的拓扑结构特性。

由于人工合成网络中通常已知真实的社区结构，因此通常采用归一化互信息(normalized mutual information, NMI)[24-25]指标衡量检测所得社区划分与真实社区划分之间的逼近程度。NMI函数如下：

$$\mathrm{NMI}(\pi^a,\pi^b)=\frac{\sum_{h=1}^{k(a)}\sum_{l=1}^{k(b)} n_{hl}\log\frac{nn_{hl}}{n_h^a n_l^b}}{\sqrt{\left(\sum_{h}^{k(a)} n_h^a\log\frac{n_h^a}{n}\right)\left(\sum_{l}^{k(b)} n_l^b\log\frac{n_l^b}{n}\right)}} \tag{2}$$

式中，π^a和π^b表示社区划分方案a和b；n_h^a和n_h^b是属于π^a和π^b划分的第h个社区的节点数目；n_{hl}代表不同社区公共成员的数量；这些成员属于π^a划分的第h个社区，同时属于π^b划分的第l个社区。

归一化互信息NMI的基本想法是把检测所得社区划分与真实社区划分进行对比，度量二者之间的相似性。NMI值越大，社区划分越接近真实的社区结构，检测结果越精确[24]。

1.3 人工合成测试网络的局限性

在社区检测算法研究中，GN网络和LFR网络作为基准测试网络也暴露出诸多不足，主要包括以下3个方面：

1) 两种网络在平均度、度分布、匹配系数、平均聚类系数、社区数以及模块度等网络微观特性上和真实世界网络差异较大。此外，由于构造过程中只能按照固定优先级的顺序设定网络微观特性，如LFR网络构造过程中只能按照N、K、k_{max}、μ、c_{max}、c_{min}的顺序依次进行参数设置，因而忽略了优先级较低的网络微观特性，也使得网络微观特性的调整灵活度受到较大影响。

2) 两种网络均能够通过调整参数设置(GN调整Z_{out}，LFR调整μ)，构造社区结构逐渐模糊的网络集合，用以衡量社区检测算法的精确性和稳定性。然而，由于网络社区结构弱化过程中无法控制其他微观特性的变化，导致网络对于算法性能评估的准确性受到影响。

3) 两种网络构造过程中仅能对网络整体特性进行设置，无法调整网络中的局部拓扑结构，因而无法衡量局部社区结构特性变化对于整体网络结构及检测算法性能的重要性，对于单个社区结构特性的分析能力较弱。

根据以上分析可知，现有人工合成基准测试网络的灵活性和多样性较差，不能有效逼近真实世界网络特性，无法满足对于社区检测算法的性能评估需求，因此设计更高性能的基准测试网络具有重要研究价值。

2 基于中尺度零模型的网络构造

为生成更逼近真实网络特性的人工合成网络，本文提出一种新的基于中尺度特性零模型的网络构造方法。该方法能够保持中尺度特性(主要为社区结构特性)的零模型，生成具有不同网络微观特性的基准测试网络，通过自由调整网络的平均度、匹配系数、聚类系数等微观特性参数，有效提升测试网络的结构多样性。

2.1 基于随机重连的不同阶数零模型构造

用ER随机图或配置模型方法构造的零模型是从无到有生成新网络的过程，而用置乱算法构造的零模型则是将原始网络随机化的过程[26-29]。静态无权网络中常用的置乱算法是随机断边重连算法，断边重连的方法主要是在原始网络的基础上将网络中原有的连边随机的断开重连[30]。相比于配置算法，连边随机重连更简单、更容易操作，不需要理解和运用复杂的数学公式、不会产生自环和重边现象，能精确保持真实网络的一些物理属性[31]。在社区网络中构建零模型，能够帮助研究者更清晰地了解社区结构。参考文献[26]和[30]，随机断边重连的不同阶零模型构造过程如图1所示。

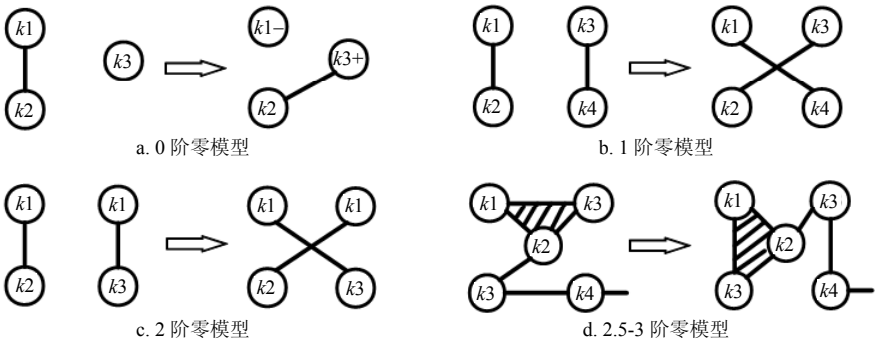


图1　0-3阶零模型的随机断边重连过程

随机断边重连的零模型是指在保证某种特性的前提下，对网络中的所有连边进行断开并且随机重连而产生的零模型。图中k表示度值，数字不同表示不同的度值。“−”表示度值减小，“+”表示度值增加。0阶零模型保证网络中所有节点平均度分布特性不变，如图1a所示；1阶零模型保证网络中节点度分布特性不变，如图1b所示；2阶零模型保证网络中联合度分布特性不变，如图1c所示；2.5阶零模型保证网络中联合度分布和断边前后度相关的聚类系数不变，如图1d所示；3阶零模型保证网络中联合边度分布特性不变，如图1d所示。由于0阶零模型是最简单且随机性较强的网络模型，所以实验研究过程中不对其进行测试。

2.2　保持中尺度特性的零模型构造

1) 网络初始化

确定网络节点数、度及最大度值、社区数、混合参数、零模型断边重连交换次数和最大尝试次数作为输入参数。根据网络参数，基于原始实证网络随机断边重连算法构造1-3阶零模型，定义计算社区内外部连边数的函数、计算网络模糊程度的系数μ的函数以及将社区划分列表转化为社区标号的函数。

2) 调用零模型构造社区内外部连边

在生成初始网络的基础上利用已构造的1-3阶零模型生成算法交换网络社区内外部连边，不同阶数零模型所代表的网络微观特性不同，所以可根据需要选择不同网络特性的零模型进而对社区内外部连边进行构造。

3) 根据混合参数μ控制内外部连边比例

在循环内部使μ值等于某一固定值，当计算零模型所构造网络内外部连边的比例值达到这一固定值时，停止循环输出最终网络结构，否则继续执行。表示网络模糊程度的混合参数μ是指所生成网络中社区间连边数占网络总连边数的比例，μ值越小表明社区内部连边所占比例越大，社区结构越明显，反之，网络模糊程度越强[30]。下面总结出保持中尺度特性的零模型构造算法的基本流程如下所示。

输入：节点数N，度值K，最大度kmax，连边数m，交换次数nswap，最大尝试次数maxtry，迭代次数steps。

输出：复杂网络$G=(V, E)$

1) 网络初始化：根据网络参数生成初始网络Gn、构造针对社区内部连边置乱和社区外部连边置乱的1-3阶零模型，社区外部连边置乱算法为inter_random_1k、inter_random_2k、inter_random_3k；社区外部连边置乱算法为inner_random_1k、inner_random_2k、inner_random_3k；定义计算网络连边的函数Edges、网络模糊程度的函数MU、社区划分列表转换社区归属信息函数network_community。

2) 调用1-3阶零模型构造社区内外部连边。

① 利用CNM算法对初始网络进行社区检测，得到社区划分列表community_list；

② 将社区划分列表community_list转化为字符串形式community_list_s；

③ 根据不同网络特性选择1-3阶零模型，依据参数nswap和maxtry，置乱初始网络Gn中的连边；

④ 当nswap=2m即执行2m次连边置乱时停止，生成网络GAS。

3) 通过模糊程度系数μ的大小控制内外部连边的比例。

① 再次引用CNM算法按照模块度最大的原则进行社区划分，结果根据network_community函数得到社区归属信息；

② 调用函数MU计算网络模糊程度系数μ，使μ等于某一固定值，若μ值小于或大于这一固定值，返回步骤2)的第③步，反之结束输出社区归属信息，得到网络G。

2.3　网络微观特性测试

为检测基于中尺度特性零模型网络构造方法的有效性，本文对生成网络的微观特性进行测试，评

价合成网络与真实世界网络微观特性的相似性。以空手道俱乐部网络为基准，测试采用保持中尺度特性的零模型构造算法生成100个测试网络，对所有网络微观特性进行统计分析，结果如表2所示。表中统计了所有网络在平均度、匹配系数、平均聚类系数、社区数和模块度这5种微观特性上的平均值和标准差。

观察表2中数据首先可知实证网络的结构统计特性如下：节点数为34，连边数为78，平均度为4.56，匹配系数为–0.48，平均聚类系数为0.57，社区数为2个，模块度为0.42。其次，GN网络的节点数和社区数目和原始网络相同，但它的连边数、平均度、匹配系数、平均聚类系数以及模块度的数值上都与原始网络相差较大，即GN网络只能保证社区数目特性不变。由于LFR网络与真实世界网络相似性比GN网络高，LFR网络在匹配系数、平均聚类系数和模块度在GN网络的基础上有所提高，但与原始网络依然相差较大，且在网络特性保持方面随机性较高。最后，保持中尺度特性的1-3阶零模型构造算法生成的网络能够与原始网络逐渐在节点数、连边数、平均度、社区数目以及平均聚类系数上基本保持一致。随着零模型阶数的升高，网络特性与真实网络越来越贴近，最终3阶零模型与真实网络特性完全相同。

表2 基于中尺度特性零模型构造网络的微观特性分析

真实网络	基准网络	节点数	边数	平均度	匹配系数	平均聚类系数	社区数	模块度
Karate 网络	原始网络	34	78	4.56	–0.48	0.57	2	0.42
	GN	34	89	5.24(±0.18)	–0.16(±0.07)	0.12(±0.03)	2(±0.00)	0.24(±0.01)
	LFR	34	73	4.31(±0.25)	0.27(±0.12)	0.62(±0.05)	6(±0.45)	0.68(±0.02)
	保持中尺度特性1阶零模型	34	78	4.56(±0.00)	–0.47(±0.01)	0.57(±0.03)	2(±0.00)	0.42(±0.00)
	保持中尺度特性2阶零模型	34	78	4.56(±0.00)	–0.48(±0.00)	0.57(±0.01)	2(±0.00)	0.42(±0.00)
	保持中尺度特性3阶零模型	34	78	4.56(±0.00)	–0.48(±0.00)	0.57(±0.00)	2(±0.00)	0.42(±0.00)

度分布特性能够很好地衡量网络中节点的连边情况，所以为进一步验证度分布特性在3种网络模型上与真实世界网络的相似程度，对度分布特性在3种网络模型上进行测试，分布曲线如图2所示。

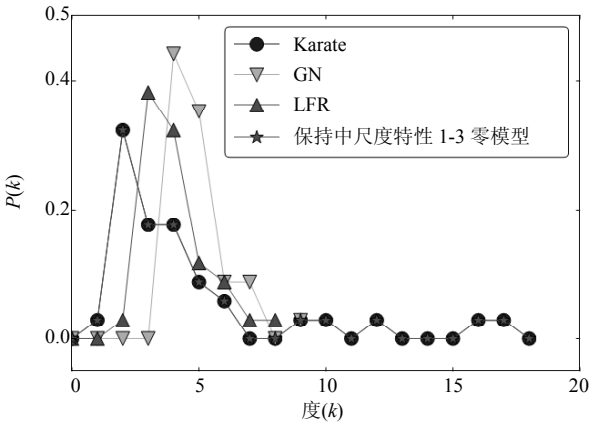


图2 保持中尺度特性零模型生成网络的度分布

从图2可以看出，GN网络度分布比较集中且与原始空手道俱乐部网络相差较大，LFR网络的度分布曲线虽然比GN网络分散，但与原始空手道俱乐部网络的分布曲线仍相差较大，而保持中尺度特性的1-3阶零模型对应的度分布曲线和空手道俱乐部网络完全相同。实验结果说明保持中尺度特性的零模型构造算法生成的网络模型能够保持度分布特性。

根据表2和图2的分析结果可得，保持中尺度特性的零模型构造算法在网络微观特性(平均度、度分布、聚类系数、模块度、同配系数)方面与真实世界网络保持一致。在参数调整方面，能够随时调整在生成网络过程中涉及到的所有参数以及网络特性，保证了网络社区结构的多样性。

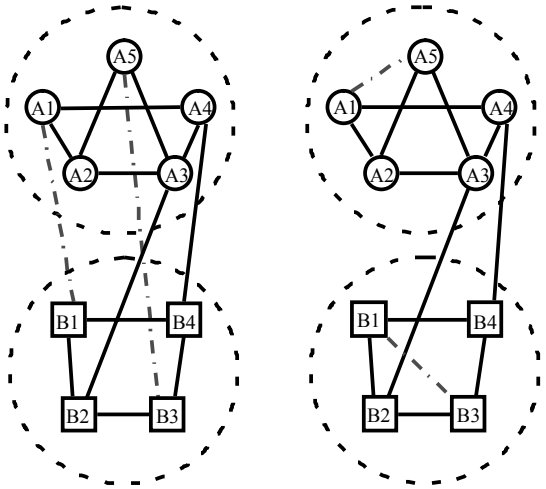
3 基于社区强弱变化零模型的网络构造

为提升基准测试网络对于社区检测算法精确性和稳定性的评估能力，本文设计了能够调整社区结构强弱的零模型，并利用其生成一系列基准测试网络。该网络能够在保证其他微观特性不变的情况下，使网络的单项特性，即社区结构模糊程度，按照检测需求进行调整。由于排除了其他微观特性因素的影响，该类网络能够更加精确地衡量社区检测算法对网络社区结构模糊程度的适应能力，提升网络对于社区检测算法精确性和稳定性的评价能力。

3.1 增强社区结构的零模型构造

社区结构一般会呈现出社区内部的节点之间连接稠密、属于不同社区的节点之间连接稀疏的特点。

如果要增强原始网络的社区结构，那就需要减少社区之间的连边，增加社区内部的连边，其过程如图3所示。首先将原始网络划分为多个社区，然后在保持社区结构不变的情况下，将两个社区之间的连边交换为社区内部连边。如图3a所示，采用断边重连1阶零模型的方式将社区A和社区B间的两条虚线连边A1-B1与A5-B3断开，然后分别将社区A中的两个节点A1与A5相连、将社区B中的两个节点B1与B3相连，最后生成的网络拓扑结构如图3b所示。通过反复使用此方式，可让实证网络的社区特性越来越强，但同时不破坏网络的度分布特性。在生成社区内部断边重连零模型时，2-3阶零模型同样适用，从而保留网络的高阶微观特性。



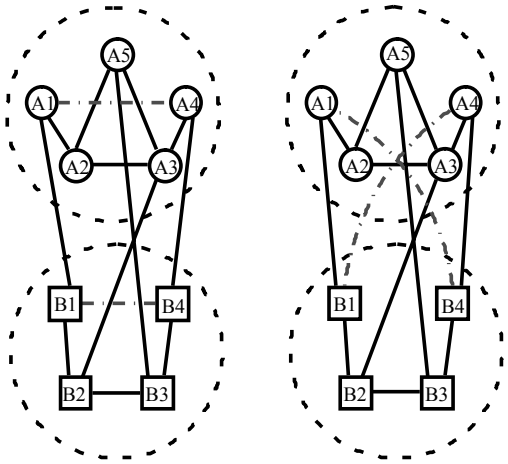
a. 原始网络　　b. 新生成网络

图3　增强社区结构的零模型构造

3.2　减弱社区结构的零模型构造

减弱社区结构零模型的原理与增强社区结构零模型构造方法相反，即通过增加社区间连边，减少社区内部连边，以达到减弱社区结构的目的，其构造过程如图4所示。首先将原始网络划分为多个社区，然后在保持其他连边结构不变的情况下，将社区内部的连边交换为两个社区之间的连边。采用断边重连1阶零模型的方式，将图4a中所示的社区A内部的虚线连边A1-A4和社区B内部的虚线连边B1-B4断开，然后将社区A和社区B间的节点A1与B4相连、A4与B1相连，重新连接后的社区结果如图4b所示。通过反复使用此方式，可让实证网络的社区特性越来越模糊，但同时不破坏网络的度分布特性。在生成社区内部断边重连零模型时，2-3阶零模型同样适用，从而保留网络的高阶微观特性。

通过构造增强或减弱社区结构的零模型可以在保持真实网络拓扑结构基本不变的情况下，增强或减弱社区结构特性，可有效识别出不同程度社区结构特性情况下，社区检测算法的稳定性和鲁棒性。



a. 原始网络　　b. 新生成网络

图4　减弱社区结构的零模型构造

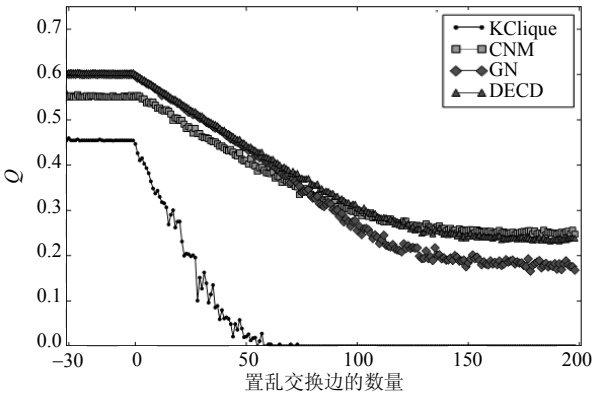
3.3　网络社区检测评价能力测试

为检测基于社区强弱变化零模型网络构造方法的有效性，本文在生成网络上进行社区检测测试，评价网络对于社区检测算法精确性和稳定性的评价能力。首先，通过增强或减弱社区结构零模型构造算法生成一个节点数为100，边数为518，度值为10和最大度为20的网络，往正向调节为减弱社区结构，往负向调节为增强社区结构。然后，采用基于模块度优化的贪心算法CNM[12]、基于分裂的层次聚类算法GN[15]、派系过滤算法KClique[7]这和基于模块度优化的差分进化算法DECD[11]4种典型的社区检测算法对上述网络分别进行社区划分，并计算出对应的评价指标函数Q和NMI。最后用可视化的方法对实验结果进行展示，结果如图5所示。

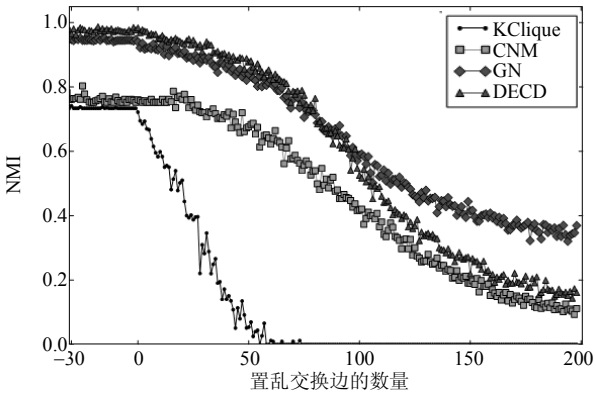
图中横坐标表示网络构造过程中零模型中置乱交换边的数量，取值范围为[–30,200]，其中横坐标值为0表示原始网络，小于0表示增强社区结构特性的网络，大于0表示减弱社区结构特性的网络。纵坐标表示在相应网络上所得社区划分结果的性能度量，图5a为模块度指标Q，而图5b为归一化互信息指标NMI。

从图5测试结果中可获得结论如下：1) 当网络社区结构较强时，继续增强社区结构特性对检测结果不会产生太大影响，但减弱社区结构特性能够准确衡量社区检测算法的鲁棒性。2) 随着社区结构特性的减弱，各算法对应的Q和NMI值逐渐下降，但变化程度和速度存在一定差异，由此体现出不同算法在社区结构单项特性上的检测能力。KClique算法曲

线波动较大，在增加社区之间连边达到60次时，Q值便降为0，说明KClique算法的性能较差。CNM算法的曲线比KClique算法平滑，但仍然存在一定的波动。GN算法虽然在网络模糊程度较强时NMI值对应的曲线表现最好，但Q值的下降速度却很快，所以稳定性和鲁棒性相对较差。4种算法中，DECD算法的稳定性和鲁棒性是相对较好的，它的Q值曲线和NMI曲线均相对平滑，在社区之间连边数达到200左右时趋于一个稳定的状态。3) 由于已知真实网络社区结构，因此NMI函数测量结果能够精确反映出所得社区划分的精确性和稳定性。然而，相同测试结果对应的Q值变化趋势与NMI变化趋势存在较大差异，说明在社区结构逐渐减弱的环境下，模块度指标Q对于算法检测质量的评价精确性受到影响。



a. 模块度指标 Q



b. 归一化互信息指标 NMI

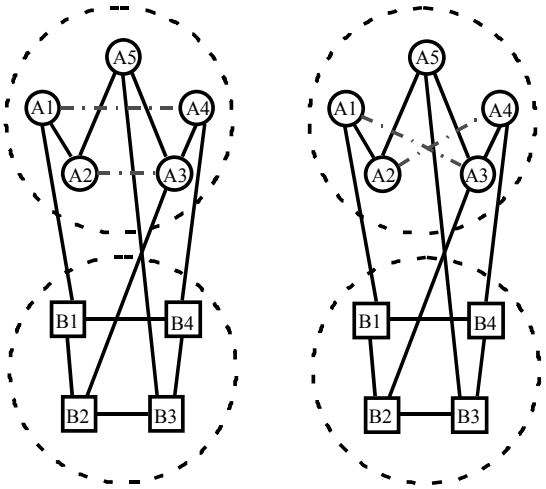
图5 社区检测算法鲁棒性的测试结果

4 基于社区局部变化零模型的网络构造

为有效衡量局部社区结构特性变化对于网络整体结构及检测算法性能的重要性，本文提出了一种基于社区内部局部断边重连的零模型构造算法，并在基础上构造了新的基准测试网络。该网络通过对单个社区内部连边进行置乱来控制整体的网络拓扑结构变化，进一步探究社区划分后哪一部分社区对于整体检测结果的影响较大，提升对于单个社区结构特性的分析能力。

4.1 社区内部局部断边重连的零模型构造

社区内部断边重连零模型只改变社区内部之间连边的拓扑结构，不改变社区间的连接关系，因此保持了原始网络的社区结构。社区内部断边重连零模型的构造过程如图6所示。首先采用某种社区检测算法将原始网络分为多个社区，然后在保持其他连边结构不变的情况下，每次只交换某一个社区内部连边。如首先选取图6a中社区A的两条虚线连边A1-A4和A2-A3，采用断边重连1阶零模型的方式先将其断开后将A1与A3相连、A2与A4相连，重连后的结果如图6b所示。在生成社区内部断边重连零模型时，2-3阶零模型同样适用。



a. 原始网络 b. 新生成网络

图 6 单个社区内部连边交换示意图

4.2 网络局部拓扑结构分析

以真实世界网络Dolphins网络为对象，首先使用社区检测算法对网络进行社区划分，然后使用基于单个社区内部局部断边重连的零模型构造算法对社区划分结果中单个社区内部连边进行置乱，最后分别对单个社区进行模块度计算来分析单个社区对整体模块度的影响，结果如图7所示。

Dolphins网络可被分为5个社区，从图7可以看出，在检测出的5个社区中，社区2的局部拓扑结构变化，对于整体网络结构及检测算法性能的影响最大，而社区3的影响最小。由此可以看出，Dolphins网络中社区2中包含的节点连边关系最为重要。此外，当保持社区结构不变，仅对社区内部结构特性进行调整时，也会对整体网络拓扑特性产生较大影响。如图中最右侧数据所示，当5个社区内部的拓扑

结构同时发生变化时，网络整体拓扑结构性能变化剧烈。基于上述分析可知，基于社区局部变化零模型的基准测试网络，一方面能够有效测量局部社区结构对于算法检测性能的影响，从而更细致地对算法性能进行评估；另一方面，能够在生成网络过程中通过局部调整，使生成的网络更加逼近真实世界网络拓扑结构，或者使生成的网络更加满足研究者设计的要求。

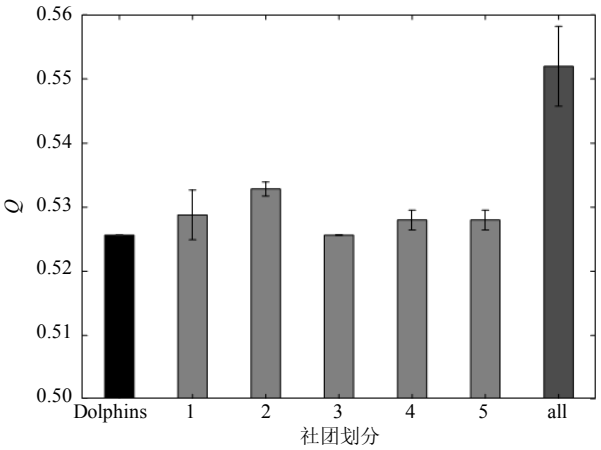


图7　单个社区结构划分结果对总体的影响

5　结 束 语

本文提出了基于3种零模型的新型社区检测基准网络构造算法。通过保持中尺度特性的零模型构造算法、增强或减弱社区结构的零模型构造算法和基于社区局部变化的零模型构造算法这3种网络生成算法分别解决了现有社区检测基准网络在网络特性与真实网络相差较大且网络特性不可调、对社区检测算法性能区分度不高以及无法改变网络局部拓扑结构等问题。根据网络微观特性测试、社区检测算法的鲁棒性测试以及网络局部拓扑结构分析的实验结果表明，3种算法生成的网络模型能够充分满足社区检测中作为测试基准网络的需求，保证了所生成的网络最大程度地维持真实世界网络的微观特性和社区结构特性，同时确保了生成的网络结构呈现多样性。

目前，基于零模型的社区检测基准网络构造算法仅在非重叠社区检测中有所应用，下一步将尝试扩展到模糊重叠社区检测的研究中。此外，算法中只展示了改变社区内部连边对整体的影响，改变社区间连边对整体影响还有进一步的研究空间。最后，本文基于零模型的社区检测基准网络构造算法只是针对现有社区检测基准网络GN和LFR存在的不足进行研究，对其他网络生成模型的劣势进行改进，以及如何进一步取长补短，与现有网络模型结合使用也是一个新的研究方向。

本文研究工作还得到大连市青年科技之星项目支持计划(2015R091)的资助，在此表示感谢。

参 考 文 献

[1] 肖婧, 张永建, 许小可. 复杂网络模糊重叠社区检测研究进展[J]. 复杂系统与复杂性科学, 2017, 14(3): 8-29.
XIAO Jing, ZHANG Yong-jian, XU Xiao-ke. Research progress of fuzzy overlapping community detection in complex networks[J]. Complex Systems & Complexity Science, 2017, 14(3): 8-29.

[2] 乔少杰, 韩楠, 张凯峰, 等. 复杂网络大数据中重叠社区检测算法[J]. 软件学报, 2017, 28(3): 631-647.
QIAO Shao-jie, HAN Nan, ZHANG Kai-feng, et al. Algorithm for detecting overlapping communities from complex network big data[J]. Journal of Software, 2017, 28(3): 631-647.

[3] YANG J, MCAULEY J, LESKOVEC J. Community detection in networks with node attributes[C]//IEEE International Conference on Data Mining. Piscataway: IEEE, 2014: 1151-1156.

[4] 汪小帆, 李翔, 陈关荣. 网络科学导论[M]. 北京: 高等教育出版社, 2012.
WANG Xiao-fan, LI Xiang, CHEN Guan-rong. Introduction to complex networks[M]. Beijing: Higher Education Press, 2012.

[5] 何东晓, 周翔, 王佐, 等. 复杂网络社区挖掘—基于聚类融合的遗传算法[J]. 自动化学报, 2010, 36(8): 1160-1170.
HE Dong-xiao, ZHUO Xu, WANG Zuo, et al. Community mining in complex networks — Clustering combination based genetic algorithm[J]. Acta Automatica Sinica, 2010, 36(8): 1160-1170.

[6] ZHANG X, CAO G. Transient community detection and its application to data forwarding in delay tolerant networks[J]. IEEE/ACM Transactions on Networking, 2017, 99: 1-15.

[7] GERGELY P, IMRE D, ILLÉS F, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. Nature, 2005, 435: 814-818.

[8] GIRVAN M, NEWMAN M E. Community structure in social and biological networks[J]. PNAS, 2002, 99(12): 7821-7826.

[9] LEE C, REID F, MCDAID A, et al. Detecting highly overlapping community structure by greedy clique expansion[R]. Dublin: University College Dublin, 2010.

[10] LANCICHINETTI A, FORTUNATO S, KERTÉSZ J. Detecting the overlapping and hierarchical community structure of complex networks[J]. New Journal of Physics, 2009, 11(3): 19-44.

[11] JIA G, CAI Z, MUSOLESI M, et al. Community detection in social and biological networks using differential evolution[C]//International Conference on Learning and Intelligent Optimization. Berlin: Springer, 2012, 7219:

71-85.
[12] CLAUSET A, NEWMAN M E, MOORE C. Finding community structure in very large networks[J]. Physical Review E, 2004, 70(2): 066111.
[13] XIAO Jing, ZHANG Yong-jian, XU Xiao-ke. Convergence improvement of differential evolution for community detection in complex networks[J]. Physica A, 2018, 503: 762-779.
[14] 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(2): 036106.
[15] GNVALDSSON T. Pattern discrimination using feedforward networks: A benchmark study of scaling behavior[J]. Neural Computation, 1993, 5(3): 483-491.
[16] LANCICHINETTI A, FORTUNATO S, RADICCHI F. Benchmark graphs for testing community detection algorithms[J]. Physical Review E, 2008, 78(4 Pt 2): 046110.
[17] BICKEL P J, CHEN A. A nonparametric view of network models and Newman-Girvan and other modularities[J]. PNAS, 2009, 106(50): 21068-21073.
[18] LUSSEAU D, NEWMAN M. E. Identifying the role that animals play in their social networks[J]. Proceedings of the Royal Society B: Biological Sciences, 2004, 271(Suppl 6): S477-S481.
[19] NEWMAN M. Modularity and community structure in networks[J]. PNAS, 2006, 103(23): 8577-8582.
[20] HE S, JIA G, ZHU Z, et al. Cooperative co-evolutionary module identification with application to cancer disease module discovery[J]. IEEE Transactions on Evolutionary Computation, 2016, 20(6): 874-891.
[21] SARZYNSKA M, LEICHT E A, CHOWELL G, et al. Null models for community detection in spatially embedded, temporal networks[J]. Journal of Complex Networks, 2018, 4(3): 363-406.
[22] XIE J, KELLEY S, SZYMANSKI B K. Overlapping community detection in networks: The state-of-the-art and comparative study[J]. ACM Computing Surveys, 2011, 45(4): 1-35.
[23] SUN Peng-gang, SUN Xi-ya. Complete graph model for community detection[J]. Physica A: Statistical Mechanics and Its Applications, 2017, 471: 88-97.
[24] DANON L, DÍAZGUILERA A, DUCH J, et al. Comparing community structure identification[J]. Journal of Statistical Mechanics, 2005(9): 09008.
[25] YANG J, LESKOVEC J. Defining and evaluating network communities based on ground-truth[C]//IEEE International Conference on Data Mining. Piscataway: IEEE, 2012, 42(1): 181-213.
[26] MAHADEVAN P, KRIOUKOV D, FALL K, et al. A basis for systematic analysis of network topologies[R]. San Diego: University of California San Diego, 2006.
[27] COLWELL R, WINKLER D. A null model for null models in biogeography[C]//Ecological Communities: Conceptual Issues and the Evidence Ecological Communities. Princeton: Princeton University Press, 1984: 344-359.
[28] 尚可可. 在线社交网络的零模型构造和行为预测研究[D]. 青岛: 青岛理工大学, 2013.
SHANG Ke-ke, The study of null model construction and user behavior prediction for online social networks[D]. Qingdao: Qingdao Technological University, 2013.
[29] MAHADEVAN P, KRIOUKOV D, FALL K, et al. Systematic topology analysis and generation using degree correlations[C]//ACM SIGCOMM. New York: ACM, 2006: 135-146.
[30] 尚可可, 许小可. 基于置乱算法的复杂网络零模型构造及其应用[J]. 电子科技大学学报, 2014, 43(1): 7-20.
SHANG Ke-ke, XU Xiao-ke. Construction and application for null models of complex networks based on randomized algorithms[J]. Journal of University of Electronic Science and Technology of China, 2014, 43(1): 7-20.
[31] WU J, ZHAO S, HE L, et al. Neural-Network-Based switching control for dc motors system with LFR[C]// International Symposium on Neural Networks. Berlin: Springer, 2007: 267-274.

编 辑 蒋 晓