计 算 机 研 究 与 发 展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202330533
61(7): 1850−1862, 2024

基于随机块模型的社区隐藏统一框架

刘 栋[1,2,3] 刘 侠[1] 贾若雪[1] 张文生[4]

[1](河南师范大学计算机与信息工程学院 河南新乡 453007)

[2](教育人工智能与个性化学习河南省重点实验室(河南师范大学) 河南新乡 453007)

[3](教学资源与教育质量评估大数据河南省工程实验室(河南师范大学) 河南新乡 453007)

[4](中国科学院自动化研究所 北京 100190)

(liudong@htu.edu.cn)

A Unified Framework for Community Hiding Based on Stochastic Block Model

Liu Dong[1,2,3], Liu Xia[1], Jia Ruoxue[1], and Zhang Wensheng[4]

[1](*College of Computer and Information Engineering*, *Henan Normal University*, *Xinxiang*, *Henan* 453007)

[2](*Henan Provincial Key Laboratory of Artificial Intelligence and Personalized Learning in Education* (*Henan Normal University*), *Xinxiang*, *Henan* 453007)

[3](*Big Data for Teaching Resources and Education Quality Evaluation Henan Engineering Laboratory* (*Henan Normal University*), *Xinxiang*, *Henan* 453007)

[4](*Institute of Automation*, *Chinese Academy of Sciences*, *Beijing* 100190)

Abstract As one of the important tools for complex network analysis, community detection can be used to help gain insight into the community structure of the network and the potential relationship between nodes. However, it also brings privacy leakage problems. As a concomitant problem of community detection, community hiding aims to destroy the community structure of the network with minimal edge disturbance cost, and it has received more and more attention from scholars in recent years. However, the existing community hiding methods ignore the network generation mechanism and lack a unified framework for hiding at different scales. Therefore, we propose a community hiding algorithm based on a stochastic block model (HC-SBM), which constructs a unified community hidden framework from the perspective of network generation mechanism, and launches three-scale attacks against community detection algorithm, namely, micro (individual), mesoscopic (community), and macro (network). The principle of this method is to illustrate the generation mechanism of the network based on the stochastic block model, especially the rules and patterns of the formation and division of the network community, mining critical links and link collections in the process of network generation.Finally, the network community structure is destroyed at the minimum cost of perturbation. Through extensive experiments on real networks and comparisons with several mainstream baseline algorithms, the proposed HC-SBM algorithm is shown to be superior in terms of community hiding effect.

Key words community hiding; community detection; stochastic block model; generation mechanism; social network analysis

收稿日期: 2023-06-21; **修回日期**: 2023-10-09

基金项目: 国家自然科学基金项目(62072160)

This work was supported by the National Natural Science Foundation of China (62072160).

通信作者: 张文生(wensheng.zhang@ia.ac.cn)

摘　要　社区检测是复杂网络分析的重要工具之一，可帮助深入了解网络的社区结构和节点间潜在的关系，但同时也带来了隐私泄露问题. 社区隐藏作为社区检测的伴生问题，旨在以最小的边扰动代价破坏网络的社区结构，近年来受到越来越多学者的关注. 但现有的社区隐藏方法忽略了网络的生成机制且缺少针对不同尺度隐藏的统一框架，因此提出了一种基于随机块模型的社区隐藏（community hiding-stochastic block model, HC-SBM）算法，该算法从网络生成机制角度构建了社区隐藏的统一框架，即实现微观（个体）、介观（社区）、宏观（网络）3 个尺度上的社区检测算法攻击. 其基本思想是基于随机块模型刻画网络的生成机制，特别是网络社区形成和分裂的规律和模式，挖掘生成过程中的关键性链接以及链接集合，最终通过最小代价扰动策略破坏网络社区结构. 通过在真实网络上的大量实验，并与 4 种先进的基准算法进行比较，表明了提出的 HC-SBM 算法在社区隐藏效果更优.

关键词　社区隐藏；社区检测；随机块模型；生成机制；社会网络分析

中图法分类号　TP391

复杂网络是一种由海量节点和边组成的复杂关系系统，可用来建模现实世界中各类系统主体之间错综复杂的联系. 社区检测是复杂网络的重要分析工具之一，旨在将网络划分为内部链接紧密、外部链接稀疏的各个社区，以帮助人们更好地了解网络的社区结构以及节点间的潜在关系[1]. 近年来，社区检测已被广泛应用于许多现实场景中，如社交网络的用户关系分析[2]、科学合作网络的研究热点挖掘[3]等，为企业和组织带来了巨大的经济和社会利益. 但同时，真实网络的社区结构以及潜在关系也被过度呈现，隐私泄露问题被越来越多的用户关注[4].

因此，如何对抗主流社区检测算法从而达到用户隐私保护目的，逐渐成为被关注的研究方向，即社区隐藏. 社区隐藏本质上是对网络结构中特定边进行扰动（增边或删边），旨在以最小的扰动代价破坏网络社区结构，降低主流社区检测算法精度，减少用户隐私泄露风险. 自文献 [5] 提出社区隐藏概念后，诸多学者尝试通过模块度、互信息、安全指数等社区隐藏度量指标，以启发式算法、遗传算法等为求解工具，寻求社区隐藏的最优方案.

目前，社区隐藏算法主要分为微观社区隐藏（个体）、介观社区隐藏（社区）和宏观社区隐藏（网络）3 类. 其中微观社区隐藏主要是通过一系列攻击策略使得特定节点的社区隶属发生变化，其代表性算法包括 Waniek 等人[6] 提出的启发式算法 ROAM（remove one add many）；介观社区隐藏主要是通过一系列攻击策略使得目标社区结构发生变化，其代表性算法包括 Fionda 等人[7] 提出的基于安全性社区隐藏（deception via safeness, Ds）算法和基于模块度社区隐藏（deception via modularity, Dm）算法以及 Waniek 等人[6] 提出的启发式算法 DICE（disconnect internally connect externally）；宏观社区隐藏主要是通过一系列攻击策略使得网络整体的社区结构发生变化，其代表性算法包括 Chen 等人[8] 和 Liu 等人[9] 提出的基于遗传算法的 Q-Attack 和 GNA（genetic algorithm based NMI attack）算法. 上述算法[6-9] 分别在不同尺度的社区隐藏上取得了较好的效果，但目前还缺少能够统一解决微观、介观和宏观社区隐藏的方法或框架，另外大多算法基于启发式或遗传算法，未考虑网络生成机制对社区结构隐藏的影响，无法更近一步解释攻击策略的合理性.

针对上述问题，本文面向微观、介观和宏观社区隐藏，引入节点重连概率矩阵（详细介绍见 3.2 节），为不同尺度下社区隐藏问题提供了统一解决思路. 同时，从网络生成机制角度出发，提出了基于随机块模型的社区隐藏（community hiding-stochastic block model, HC-SBM）算法. 该算法利用随机块模型（stochastic block model, SBM）[10] 可解释性、表达能力强等优点，刻画网络生成机制，特别是网络社区形成和分裂的规律和模式，从中挖掘维持网络社区结构最关键链接及链接集合，以最小代价扰动，最大限度地攻击网络社区结构，最终实现微观、介观和宏观社区隐藏. 此外，该算法与主流社区隐藏算法进行实验对比，从多个角度验证所提算法的性能优势.

本文的主要贡献包括 3 个方面：

1）提出了面向不同尺度社区隐藏问题的统一解决思路. 通过引入重连概率矩阵，实现了微观、介观和宏观 3 个不同尺度社区隐藏.

2）提出了基于随机块模型的社区隐藏算法 HC-SBM. 该算法从网络生成机制角度考虑，利用随机块模型刻画网络社区形成和分裂的规律和模式，挖掘出维持网络社区结构最关键性链接及链接集合，以最小代价扰动策略破坏网络社区结构.

3) 广泛的实验和评估. 在 4 个真实网络数据集上，将所提算法与主流社区隐藏算法进行实验对比，从多角度验证所提算法的性能优势.

1 相关工作

本节从社区检测和社区隐藏 2 个研究方向的进展阐述本文相关工作.

社区检测主要基于网络的拓扑结构，采用模块划分的方式将网络划分为组内链接紧密、组外链接稀疏的社区；基于是否考虑重叠节点的角度将社区检测分为非重叠社区检测和重叠社区检测 2 类[11]. 在非重叠社区检测中，每个节点只属于 1 个社区. 自文献 [12−13] 提出用模块化来衡量社区结构质量以来，诸多学者在此基础上做了优化，如多级聚类(multi-level clustering)算法[14]、贪婪优化(greedy optimization)算法[15]、光谱优化(spectral optimization)算法[16]等，这些算法将社区检测转化为一个优化问题，并试图通过最大化适应度函数求解最优社区划分. 此外，Yang 等人[17]从信息传播的角度出发，提出了标签传播算法(label propagation algorithm, LPA)，该算法根据每个节点的邻居节点来更新其标签. 同时，Fan 等人[18]则从机器学习的角度出发，提出了一种基于谐波函数的半监督社区检测降噪方法. 在重叠社区检测中，一些节点可以归属多个社区，社区之间存在交集. 目前也有一些相对成熟的重叠社区检测方法，如 Palla 等人[19]研究通过发现联通的派系来发现重叠社区的算法 CPM(clique percolation method)；文献 [20] 提出了一种基于节点表示的两步遗传算法来检测重叠社区. 这些方法[14−20]均为社区检测提供了新思路和创新性解决方案. 然而，随着社区检测结果越来越精确，个人隐私信息泄露的风险也日益突出.

社区隐藏是社区检测的逆问题. 其本质是使用某种策略对网络的边进行扰动以攻击网络结构，降低主流社区检测算法精度. 目前社区隐藏相关研究主要分为微观社区隐藏、介观社区隐藏和宏观社区隐藏 3 类. 针对微观社区隐藏，Waniek 等人[6]提出了启发式算法 ROAM，该算法通过移除目标节点与其邻居的边和添加对应节点与目标节点其他邻居的边实现微观社区隐藏. Li 等人[21]提出基于图学习的社区检测模型对抗攻击算法 CD-Attack，该算法结合局部相似性和全局相似性设计边的篡改策略，利用黑箱攻击实现微观社区隐藏. 针对介观社区隐藏，Fionda 等人[7]引入了社区欺骗的概念，提出基于安全度的社区隐藏算法 Ds 和基于模块度的社区隐藏算法 Dm，并给出欺骗得分 H 作为 Ds 和 Dm 的评价指标. 在此基础上，Chen 等人[22]提出一种新的改进算法 Hs(hiding via safeness)，通过设计增益函数的具体实例，将社区隐藏问题转化为优化问题，实现将敏感目标社区的成员隐藏到网络的其他不敏感社区中. 此外，Waniek 等人[6]采用删除社区内部边、增加社区之间边的方法，提出了启发式算法 DICE，由于 DICE 随机地选择重连链接，无法区分添加或删除的最佳边，性能相对较低. 针对宏观社区隐藏，Chen 等人[8]和 Liu 等人[9]都从遗传算法角度出发，分别利用模块度 Q 和归一化互信息 NMI 设计适应度函数以实现宏观社区隐藏，并取得了较好的社区结构隐藏效果. Liu 等人[23]从信息理论视角研究社区结构隐藏，提出基于社区的结构熵来表达社区结构所揭示的信息量，并设计残差熵最小化(residual error minimization, REM)算法实现全局网络隐藏. 此外，近年来也有学者对重叠社区隐藏进行研究，其中 Liu 等人[24]提出基于节点重要度的重叠社区隐藏算法，利用节点重要度来删除或添加相应的社会关系以将节点移除重叠社区. Yang 等人[25]提出基于多级邻域信息的重叠社区隐藏算法.

上述社区隐藏方法研究，大多针对单一尺度社区隐藏问题进行优化，普适性差，缺少能够统一解决微观、介观和宏观社区隐藏的方法或框架. 同时，大多数方法是以启发式或遗传算法为基础隐藏策略，通常依赖于某些先验知识设计特定的适应度函数，对网络生成机制刻画不足，存在代价高、鲁棒性差等问题，且无法更近一步解释攻击策略的合理性. 因此，本文尝试从网络生成机制的角度出发，构建了一种基于随机块模型的社区隐藏算法 HC-SBM，该算法利用随机块模型具有可解释性、表达能力强以及灵活等优点，刻画网络社区形成和分裂的规律或模式，引入节点重连概率矩阵，为实现微观、介观和宏观社区隐藏提供了统一解决思路. 该算法与主流社区隐藏方法之间的区别如表 1 所示.

2 符号表示与社区隐藏问题描述

给定 $G=(V,E)$ 表示一个无向、无权网络，其中 V 表示网络中节点的集合，E 表示网络中边的集合. 社区发现的目标是利用社区检测算法 $\boldsymbol{A}_{\mathrm{D}}$ 将网络 G 划分为 k 个子图 $G_i=(V_i,E_i)$，每个子图表示 $C=\{C_1,C_2,\cdots,C_k\}$ 中一个社区，其中对于任意的 $i,j\in\{1,2,\cdots,k\}$，C_i

Table 1 Comparison Between HC-SBM and the Mainstream Community Hiding Algorithms

表 1 HC-SBM 与主流社区隐藏算法的对比

算法	操作	隐藏尺度	是否考虑生成机制
ROAM	加边和减边	微观	否
DICE	加边或减边	介观	否
SBA[7]	加边或减边	介观	否
RAT[5]	加边和减边	宏观	否
HC-SBM（本文）	加边和减边	微观、介观和宏观	是

$\cap C_j=\varnothing$. 此外，在预算β下添加$(E+)$或删除$(E-)$$G$中节点对的链接，得到扰动网络$G'=(V,E')$. 本文采用的具体符号总结如表 2 所示.

Table 2 Symbolic Representation

表 2 符号表示

符号	描述
G/G'	原网络/扰动网络
G_i	网络G的一个子图
C/C'	原网络社区划分/扰动网络社区划分
C_i/C_i'	任意一个原网络社区/任意一个扰动网络社区
A_{D}	社区检测算法
β	扰动预算
$E+/E-$	添加链接/删除链接
C_{t}	目标社区
v_{t}	目标节点
$\boldsymbol{P}$	重连概率矩阵

问题 1. 微观社区隐藏. 给定目标节点v_{t}，假设v_{t}在原始网络中所属社区为C_i，通过以最小的扰动代价对原始网络的边进行扰动后v_{t}所属的社区为C_j，且$C_i\neq C_j$，可以通过求解优化问题实现微观社区隐藏：

$$\arg\min\left\{\varphi\left(C_i,E,E+,E-,E',C_j\right)\right\},\tag{1}$$

其中φ是衡量改变节点社区所需预算的函数. $(E+)$和$(E-)$表示链接的添加和删除，为求得最小预算下改变目标节点v_{t}的社区. 本文主要围绕目标节点v_{t}进行链接扰动：

$$\begin{aligned}&E+\subseteq\{(s,v_{\mathrm{t}}):s\in V,(s,v_{\mathrm{t}})\notin E\},\\&E-\subseteq\{(s,v_{\mathrm{t}}):s\in V,(s,v_{\mathrm{t}})\in E\}.\end{aligned}\tag{2}$$

问题 2. 介观社区隐藏. 从社区划分集合 C 中选择一个目标社区$C_{\mathrm{t}}\in C$. 通过给定的预算β添加$(E+)$或删除$(E-)$$C_{\mathrm{t}}$中节点的链接（目的是破坏$C_{\mathrm{t}}$的社区结构），得到扰动网络$G'=(V,E')$. 然后用同样的社区检测方法$A_{\mathrm{D}}$对扰动网络进行社区检测，得到社区划分$C'=\{C_1',C_2',\cdots,C_r'\}$，可以通过求解最优化问题实现介观社区隐藏：

$$\arg\max\{\sigma(A_{\mathrm{D}}(G,E,C_{\mathrm{t}}),A_{\mathrm{D}}(E',E+,E-,G',C_{\mathrm{t}},\beta))\},\tag{3}$$

其中σ是隐藏指标度量函数，用于衡量社区结构C和C'之间的差异，通过最大化这种差异来达到社区隐藏的目的. $E'=E\cup(E+)\cup(E-)$是链接扰动后边的集合，为最好隐藏C_{t}，本文主要考虑在目标社区内删除边、目标社区间加边：

$$\begin{aligned}&E+\subseteq\{(s,r):s\in C_{\mathrm{t}}\oplus r\in C_{\mathrm{t}},(s,r)\notin E\},\\&E-\subseteq\{(s,r):s\in C_{\mathrm{t}}\wedge r\in C_{\mathrm{t}},(s,r)\in E\},\\&(E+)+(E-)\leqslant\beta.\end{aligned}\tag{4}$$

问题 3. 宏观社区隐藏. 给定网络G，在预算为β情况下对其链接进行扰动，得到扰动网络$\tilde{G}'$. 对$\tilde{G}'$用同样的社区检测算法得到社区$\tilde{C}=\left\{\tilde{C}_1,\tilde{C}_2,\cdots,\tilde{C}_s\right\}$，利用宏观社区隐藏度量指标函数$\tau$比较$C$和$\tilde{C}$两个集合的差异. 可以通过优化问题实现全局社区隐藏：

$$\arg\min\left\{\tau(A_{\mathrm{D}}(G,E,C),A_{\mathrm{D}}(E',E+,E-,\tilde{G}',\tilde{C},\beta))\right\}.\tag{5}$$

一般情况下，社区的隐藏效果通常会随着预算β的增加而增加，本文工作目标是在尽可能小的预算β下，使得社区达到最大隐藏效果. 因此，本文全局考虑节点链接的扰动：

$$\begin{aligned}&E+\subseteq\{(s,r):s\in V\wedge r\in V,(s,r)\notin E\},\\&E-\subseteq\{(s,r):s\in V\wedge r\in V,(s,r)\in E\},\\&(E+)+(E-)\leqslant\beta.\end{aligned}\tag{6}$$

其中式(1)中的函数φ、式(3)中的函数σ、式(5)中的函数τ用于衡量社区隐藏效果（详细内容见 4.2.2 节度量指标部分），本文力求最小化φ和τ、最大化σ.

从上述 3 个问题的描述可表明，社区隐藏可以转换为组合优化问题. 常用的社区隐藏算法往往采用启发式算法或遗传算法寻求最优解，但存在代价高、鲁棒性差等问题. 因此，本文提出了 HC-SBM，该算法从网络生成机制的角度考虑，运用随机块模型刻画网络的生成和分裂的规律或模式，得出节点重连概率矩阵，挖掘出破坏社区结构的关键链接及链接的集合，为不同尺度社区隐藏问题提供统一的解决思路，以最小的链接扰动来最大化破坏网络社区结构.

3 基于随机块模型的社区隐藏算法（HC-SBM）

3.1 随机块模型（SBM）

随机块模型是一种经典的生成模型，结合了随机模型与块模型，能够捕获以社区分布为隐变量的概率生成过程，通过最大化节点社区成员的似然函数，可以重新构建社区. 此外，随机块模型具有可解

释性，能够刻画给定网络G的生成过程[26]. 主要分为2步：

1)参数估计. 该过程需要与真实网络拟合，通过估计参数得出最优随机块模型. 需要估计的参数主要有节点与块的隶属度矩阵$\boldsymbol{z}\in\mathbb{R}^{n\times k}$以及块与块的链接概率矩阵$\boldsymbol{\pi}\in\mathbb{R}^{k\times k}$，$\boldsymbol{z}$和$\boldsymbol{\pi}$中的元素都是来自[0,1]的概率值，分别表示节点属于某个社区的概率以及块间节点和块内节点间是否有链接的概率.

2)生成网络. 任意节点v_i和v_j之间是否生成链接取决于节点所属的块、两者所属的块以及块与块之间的链接概率. 节点v_i和v_j分别属于块k和l，则2个节点之间链接a_{ij}的生成过程服从伯努利分布：

$$a_{ij}|c_{ik}c_{jl}=1\sim Ber(\pi_{kl}), \tag{7}$$

其中c_{ik}表示节点v_i属于社区k的概率，$Ber(\pi_{kl})$表示服从以概率为π_{kl}的伯努利分布.

基于上述过程，随机块模型完整的似然函数是：

$$P(\boldsymbol{A},C|\boldsymbol{z},\boldsymbol{\pi})=\prod_{i=1}^{n}\prod_{r=1}^{k}z_k^{c_{ir}}\times\prod_{i\neq j}^{n}\prod_{r,l}^{k}\left(\pi_{rl}^{a_{ij}}(1-\pi_{rl})^{(1-a_{ij})}\right)^{c_{ir}c_{jl}}. \tag{8}$$

本文采用期望最大化(expectation maximization, EM)算法求解似然函数，估计出最优模型参数.

3.2 节点重连概率

由于标准随机块模型假设节点的连接是随机的，连接的概率由节点的社区隶属度和块链接概率矩阵决定. 为了探究社区不同节点链接的重要性，找到破坏网络社区结构关键链接，统一解决微观、介观和宏观社区隐藏问题，本文引入了重连概率矩阵，该矩阵假设具有高重连概率的节点链接在维持网络社区的结构方面更重要. 形式上，对于任意节点v_i和v_j，利用随机块模型刻画节点v_i与其邻居节点v_j在网络中的链接生成过程，多次迭代生成，挖掘出节点v_i所在社区形成和分裂的规律，得出维持节点v_i所在社区的关键性链接及链接集合(算法1中行⑤~⑩). 节点v_i与其他节点的重连概率计算公式为：

$$P_{ij}=\frac{\sum_{t=1}^{T}f\left(v_j,N_t(v_i)\right)}{T}, \tag{9}$$

其中T为迭代生成网络的次数，$N(v_i)$表示节点v_i的邻居集合，$f\left(v_j,N_t(v_i)\right)$为判断节点$v_j$是否出现在$N(v_i)$集合中的函数. 重连概率高的节点间的链接是维持网络社区结构的关键链接，选择重连概率高的节点链接进行扰动(加边或减边)，以最小代价扰动破坏网络社区结构. 实现伪代码如算法1所示.

算法1. 节点重连概率算法.

输入：网络$G=(V,E)$，迭代次数T；

输出：节点重连概率P_{ij}.

① 初始化 Num=0;

② 将随机块模型与网络 G 拟合得到最优模型参数$\boldsymbol{z}$和$\boldsymbol{\pi}$;

③ $v_i,v_j\leftarrow RandomChooseNode(V)$; /* 选择扰动节点对*/

④ $\boldsymbol{z}'(v_i)\leftarrow ChangetNodeMembers(\boldsymbol{z}(v_i))$;

⑤ while $T>0$ do /* 迭代生成网络，挖掘节点v_i社区形成和分裂规律*/

⑥ $G_T\leftarrow GenerationNetwork(\boldsymbol{z}',\boldsymbol{\pi})$; /* 根据模型参数$\boldsymbol{z}'$和$\boldsymbol{\pi}$迭代生成网络*/

⑦ $N_T(v_i)\leftarrow G_T$; /* 根据每次迭代生成的网络得出节点v_i的邻居节点集合*/

⑧ if $v_j\in N_T(v_i)$/*判断节点v_j是否在节点v_i的邻居节点集合中*/

⑨ Num++;

⑩ end if

⑪ T--;

⑫ end while

⑬ $P_{ij}=Num/T$.

3.3 HC-SBM 模型结构

基于随机块模型的社区隐藏模型结构大致分为3个部分，模型框架如图1所示. 1)对原网络G使用社区测算法A_{D}进行社区检测，得到社区划分集合. 2)利用重连概率矩阵$\boldsymbol{P}$，依次选择概率矩阵中最大值对应的索引，得到破坏网络社区结构的关键链接及链接的集合，如图1中计算得出节点v_3的概率矩阵中最大值依次对应的索引为(v_5,v_2,v_0,v_1)，则破坏节点v_3社区结构的扰动链接为$SetEdge=\{(v_3,v_5),(v_3,v_2),(v_3,v_0),(v_3,v_1)\}$，然后依次对原网络进行边扰动，得到扰动网络$G'$. 3)使用同样的社区检测算法$A_{\mathrm{D}}$对$G'$进行社区检测，得到社区划分$C'$，比较$C$与$C'$，得出度量隐藏效果. HC-SBM算法构建了社区隐藏的统一解决思路，根据社区隐藏尺度ω的取值，实现了微观、介观和宏观3个不同尺度的社区隐藏以及3个尺度之间的变换. 实现伪代码如算法2所示.

算法2. HC-SBM 算法.

输入：网络$G=(V,E)$，预算β，社区隐藏尺度ω；

输出：扰动网络G'.

① if $\omega\in\{v_{\mathrm{t}},C_{\mathrm{t}},G\}$ /* 选择社区隐藏尺度*/

② $\tilde{V}_\omega\leftarrow getSetSpecificNode(\omega)$; /* 社区隐藏尺度对应节点的集合*/

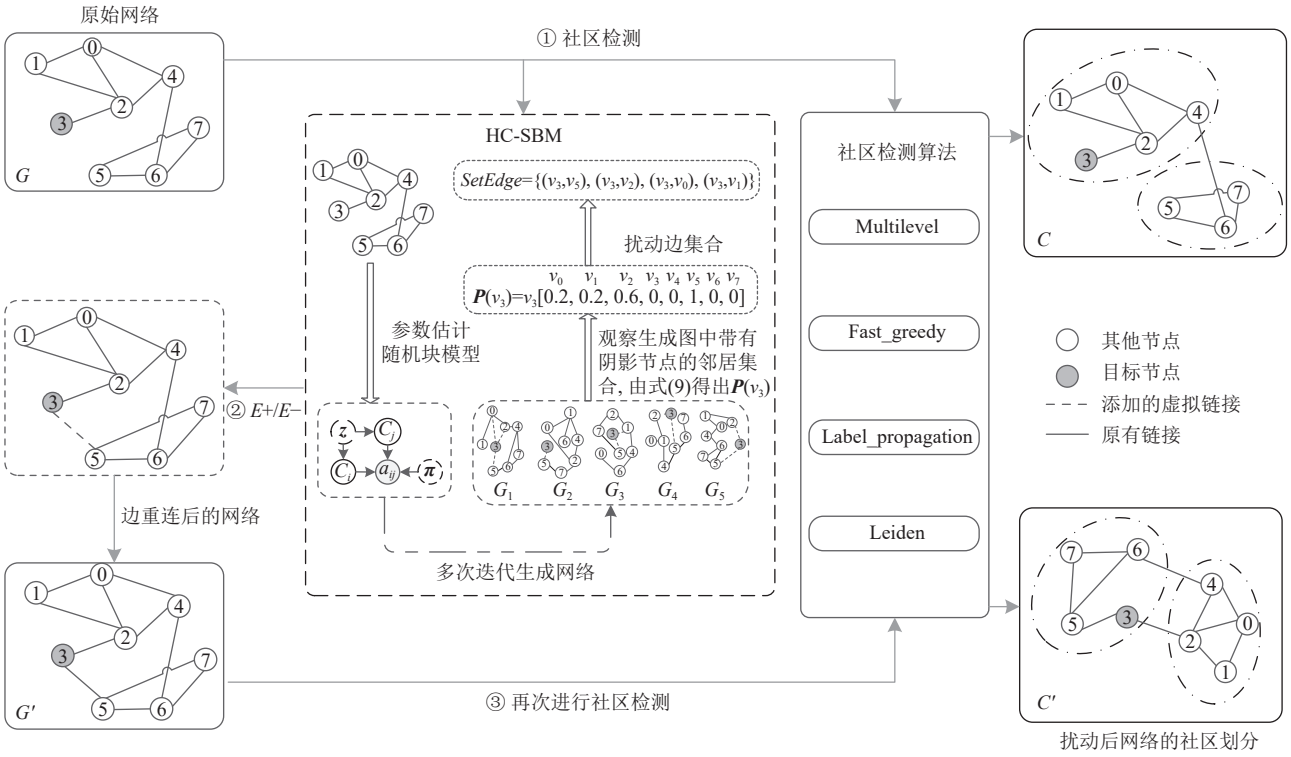


Fig. 1 Illustration of HC-SBM

图 1 HC-SBM 示意图

③ end if

④ while β>0 do /* 迭代选择扰动链接*/

⑤ $v_i \leftarrow chooseNode(\tilde{V}_\omega)$;

⑥ 由算法 1 得到节点v_i的重连概率矩阵$\boldsymbol{P}(v_i)$;

⑦ $v_j \leftarrow getMaxIndex(\boldsymbol{P}(v_i))$; /* 根据节点重连概率矩阵选择最大值对应的索引*/

⑧ if v_i和v_j 在同一个社区

⑨ 从网络 G 中删除节点对 (v_i,v_j) 间的链接;

⑩ β−−;

⑪ end if

⑫ if v_i和v_j不在同一个社区

⑬ 将节点对(v_i,v_j) 间的链接加入到网络 G 中;

⑭ β−−;

⑮ end if

⑯ end while

⑰ $G' \leftarrow update(G)$. /* 更新网络*/

1)微观社区隐藏. 由算法 2 可知, 给定网络G和目标节点v_t, 得到特定节点集合$\tilde{V}_{v_t}$(算法 2 行②). 首先由算法 1 得出目标节点v_t的重连概率矩阵$\boldsymbol{P}(v_t)$, 然后依次选择$\boldsymbol{P}(v_t)$中最大值对应的节点v_j, 得到维持目标节点v_t所属社区的关键性链接(v_i, v_j)(算法 2 中行③④). 若v_t与v_j不在同一个社区且两者间不存在链接, 则在两者间添加链接; 若v_t与v_j在同一个社区且两者间存在链接, 则删除两者的链接(算法 2 中行⑥~⑫). 多次迭代扰动关键性链接, 直到目标节点v_t的社区发生变化.

2)介观社区隐藏. 由算法 2 可知, 给定网络G和目标社区C_t, 得到特定节点的集合$\tilde{V}_{C_t}$(算法 2 行②). 首先, 由算法 1 可得出集合$\tilde{V}_{C_t}$中任意节点的重连概率矩阵$\boldsymbol{P}(v_k)$, 然后依次选择$\boldsymbol{P}(v_k)$中最大值对应的节点v_l, 得到维持目标社区C_t结构的关键性链接(v_k, v_l)(算法 2 中行③④). 若v_k与v_l不在同一个社区且两者间不存在链接, 则在两者间添加链接; 若v_k与v_l在同一个社区且两者间存在链接, 则删除两者的链接(算法 2 中行⑥~⑫). 在预算β下, 迭代扰动关键性链接, 使得目标社区中的节点尽可能分散到其他社区中, 最大化欺骗得分H和隐藏性能μ的值.

3)宏观社区隐藏. 由算法 2 可知, 给定网络G, 得到特定节点集合$\tilde{V}_G$. 首先由算法 1 得出集合$\tilde{V}_G$中任意节点重连概率矩阵$\boldsymbol{P}(v_p)$, 然后依次选择$\boldsymbol{P}(v_p)$中最大值对应的节点v_s, 得到维持网络社区结构关键性的链接(v_p, v_s)(算法 2 中行③④). 若v_p与v_s不在同一个社区且两者间不存在链接, 则在两者间添加链接; 若v_p与v_s在同一个社区且两者间存在链接, 则删除两者的链接(算法 2 中行⑥~⑫). 在预算β下, 迭代扰动关键性链接, 使得每次扰动最大化破坏网络的社区结

构, 最小化网络的度量指标 *NMI*, *ARI*, *Jaccard*.

根据基于随机块模型的社区隐藏算法的伪代码, 现对其时间复杂度进行分析. 由算法 1 和算法 2 可知, 微观社区隐藏的时间复杂度为$O\left(T\times n\log n\right)$, 介观社区隐藏的时间复杂度为$O\left(\beta\times T\times n\log n\right)$, 宏观社区隐藏时间复杂度为$O\left(\beta\times T\times n\log n\right)$. 其中, n为节点数, $T\times n\log n$为迭代生成网络得到的节点重连概率矩阵的时间复杂度, β为扰动预算.

4 实验分析

4.1 实验数据集及社区检测方法

本文在 4 个真实网络数据集[27−29]上进行实验,包括 Blogs, Power, Hep, Astro. 表 3 给出了数据集所对应的节点数、边数以及在不同社区检测方法下的社区数等详细信息.

Table 3 Experimental Datasets

表 3 实验数据集

数据集	节点数	边数	各社区检测算法下的社区数				数据集描述
			Multilevel	Fast_greedy	Label_propagation	Leiden	
Blogs	1 222	16 714	11	11	3	11	超链接博客网络
Power	4 941	6 594	40	41	481	490	美国电网
Hep	8 361	15 751	1 376	1 411	1 858	1 375	高能理论科学家合作网络
Astro	16 706	121 251	1 072	1 168	1 550	1 078	天体物理学合作作者网络

本文选取了 4 种主流的社区检测算法, 包括 Multilevel, Fast_greedy, Label_propagation, Leiden.

1)Multilevel[14]. 该算法将社区发现问题视为一个多级模块度优化问题. 首先从孤立的节点开始, 以最大化模块度收益为目的, 迭代将节点移动到社区, 直至无法进一步改进为止. 然后通过将社区合并成一个单一的超级节点重新建立新的网络. 重复这 2 个步骤, 直到算法稳定下来.

2)Fast_greedy[15]. 该算法基于贪心模块度最大化策略, 将每个节点看作是一个社区, 每次迭代选择产生最大模块度的 2 个社团合并, 直至整个网络融合成一个社区.

3)Label_propagation[30]. 该算法为每个节点分配一个标签, 随着信息的传播, 更新标签为节点的邻居中出现最频繁的标签, 直至没有节点更改其标签.

4)Leiden[31]. 该算法是 Louvain[32] 算法的一个改进版, 通过局部移动节点改善划分结果, 根据改善后的划分情况凝聚网络.

4.2 基准算法及度量指标

本文分别从微观、介观和宏观 3 个不同尺度的社区隐藏来选取各自的基准算法及度量指标.

4.2.1 基准算法

在微观社区隐藏方面, 本文提出启发式随机攻击算法 Dr(add and delete random)作为基准算法, 即在目标节点和其他社区节点之间随机增加和删除链路.

在介观社区隐藏方面, 主要基于安全度攻击的社区隐藏算法 SBA (safeness based attacks)和 DICE 作为基准算法. 其中 SBA 是以最大化安全增益为基础, 通过移除社区内部的边和添加社区间的边来削弱社区结构, 欺骗社区检测算法. DICE 是一种用于伪装社区的启发式算法, 通过删除社区的内部链接、增加社区间的链接, 使目标社区中的节点分散到其他社区.

在宏观社区隐藏方面, 主要用随机攻击(random attacks, RAT)算法和 Pro-SBM[33] 算法作为基准算法. RAT 主要是通过随机断开原有网络中的一条链路, 并随机连接网络中无链接的节点对. Pro-SBM 算法先通过随机、公平地分配扰动资源, 然后通过似然最小化的方法选择适当的扰动边缘.

4.2.2 度量指标

在微观社区隐藏方面, 主要统计了从原社区中移出目标节点所需扰动的边数, 力求扰动最少的链接将目标节点移出原社区.

在介观社区隐藏方面, 主要使用 Fionda 和 Pirro 提出的欺骗得分H和 Waniek[6] 提出的度量性能μ来评估 HC-SBM 攻击的效果.

给定目标社区C_{t}和社区划分C', 欺骗得分H计算为

$$H(C_{\mathrm{t}},C')=\left(1-\frac{|S(C_{\mathrm{t}})|-1}{|C_{\mathrm{t}}|-1}\right)\times\left(\frac{1}{2}\left(1-\max_{C_i\in C'}\{R(C_i,C_{\mathrm{t}})\}\right)+\frac{1}{2}\left(1-\frac{\sum_{C_i\cap C_{\mathrm{t}}\neq\varnothing}P(C_i,C_{\mathrm{t}})}{|C_i\cap C_{\mathrm{t}}\neq\varnothing|}\right)\right),\tag{10}$$

其中$|S(C_{\mathrm{t}})|$为C_{t}的成员诱导的子图中连通分量的个

数. $R(C_i,C_t)$和$P(C_i,C_t)$分别是目标社区C_t的召回率和精确度. 欺骗得分H的取值范围为0～1. $H=1$表示C_t中的节点分散到了除目标社区外其他的各个社区中，并且以最好的方式隐藏在网络中；$H=0$表示C_t中任意节点的社区都没发生改变.

隐藏程度μ是衡量目标社区中C_t的节点在社区结构C'中隐藏程度的指标，从社区中的节点分散程度和隐藏程度2个方面来度量. 隐藏程度μ计算为

$$\mu(C_t,C')=\alpha\frac{|C_i\in C':C_t\cap C_i\neq\varnothing|-1}{(k-1)\times\max\limits_{C_i\in C'}(C_t\cap C_i)}+(1-\alpha)\sum_{C_i\in C'}\frac{|C_i/C_t|}{\max(n-|C_t|,1)}. \tag{11}$$

其中k表示网络中社区数，且$k>1$；n表示网络中的节点数；α表示参数，其取值对于实验结果评估没有影响，本文实验中α=0.5. 此外，$\mu\in[0,1]$，其数值越大表示目标社区隐藏效果越好.

在宏观社区隐藏方面，主要使用*NMI*[34]，*ARI*，*Jaccard*等指标来度量HC-SBM对于整个网络扰动的效果. 给定网络G的原社区划分$C=\{C_1,C_2,\cdots,C_n\}$，攻击后的社区划分为$C'=\{C_1',C_2',\cdots,C_n'\}$.

$$NMI=\frac{2I(C,C')}{H(C)+H(C')}, \tag{12}$$

$$ARI=\frac{RI-E(RI)}{\max(RI)-E(RI)}, \tag{13}$$

$$Jaccard(C,C')=\frac{|C\cap C'|}{|C|+|C'|-|C\cap C'|}, \tag{14}$$

其中$H(C)=-\sum\limits_{i=1}^{n}\frac{|C_i|}{|N|}\mathrm{lb}\frac{|C_i|}{|N|}$是原网络划分$C$的熵，$I(C,C')=H(C)-H(C|C')$是社区划分$C$和$C'$之间的互信息，$H(C|C')=-\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{k}\frac{|C_i\cap C_j'|}{|N|}\mathrm{lb}\frac{|C_i\cap C_j'|}{|C_j'|}$，$RI=\frac{i+j}{\binom{n}{2}}$为兰德指数，其中$i$表示在$C$中被划分为同一类，而在$C'$中也被划分为同一簇的节点数；$j$为在$C$中被划分为不同类别，在$C'$中被划分为不同簇的节点数. $E(RI)=\frac{\sum\limits_{i}\binom{r_i}{2}\sum\limits_{j}\binom{c_j}{2}}{\binom{n}{2}}$为兰德指数期望值，其中$r_i$表示列联表第$i$行的和，$c_j$表示列联表第$j$列的和.

与社区检测结构相反，在社区隐藏问题中*NMI*，*ARI*，*Jaccard*的值越接近于零目标对象隐藏效果越好.

4.3 实验结果与分析

为了评估基于随机块模型的社区隐藏算法HC-SBM的有效性，本文主要从HC-SBM的攻击是否能过躲避各个主流社区检测算法，以及HC-SBM是否比主流社区隐藏算法效果更佳2个方面进行分析，通过广泛实验得出3点结论.

1)在宏观社区隐藏方面. HC-SBM对4种社区检测算法的攻击效果如表4所示，其中实验预算分别选取数据集网络边数的1%，3%，5%，实验结果为迭代10次取平均值. 总的来说，对所有的攻击策略，*NMI*，*ARI*，*Jaccard*的值都随着预算的增加而减小，即改变的边越多，社区隐藏效果越显著. 以HC-SBM对Fast_greedy社区检测算法为例，Blogs数据集在边扰动预算为边数5%的情况下*NMI*的值能达到0.421，而Astro数据集在边扰动预算为边数5%的情况下*ARI*的值平均能达到0.261，*Jaccard*的值平均能达到0.208. 此外，HC-SBM还与基线算法Pro-SBM和RAT进行了对比. 由表4可知，超过90%的情况下，HC-SBM的*NMI*，*ARI*，*Jaccard*值都要小于Pro-SBM和RAT，即HC-SBM社区隐藏效果比Pro-SBM和RAT更好.

总之，针对宏观社区隐藏，HC-SBM能很好地躲避4个主流社区检测算法，并且隐藏性能优于现有主流社区隐藏算法. 主要是由于HC-SBM是从网络的生成机制入手，挖掘出破坏网络社区结构最关键链接及链接集合，每次扰动对社区结构影响最大的边，从而有较好的社区隐藏效果.

2)在介观社区隐藏方面. HC-SBM算法对4种社区检测算法攻击效果以及同介观社区隐藏基准算法的实验对比结果如表5所示. 其中实验预算的选择依赖于选择的目标社区的大小(本文选取目标社区内边数的10%)，H和μ的值越高表示目标社区隐藏越好. 由表5可知，HC-SBM对4种社区检测算法都有较好的攻击效果. 以Blogs为例，在对Multilevel社区检测算法的攻击下，欺骗得分H平均(本文实验是迭代10次取平均值)能达到0.515，隐藏性能μ平均能达到0.491. 同时将HC-SBM算法与基线算法DICE和SBA的性能进行了比较，在超过95%的情况下，HC-SBM的隐藏效果要比SBA和DICE的效果好. 这种优势主要是由于HC-SBM扰动的是网络社区结构最关键链接集合，对社区结构影响最大，从而使得对社区检测算法有较好的攻击效果.

以Blogs为例，图2给出了在不同预算下HC-SBM，DICE，SBA的欺骗得分H和隐藏性能μ的对比. 大部分情况下，HC-SBM的欺骗得分H和隐藏性能μ要优于SBA和DICE，即HC-SBM的折线高于DICE和SBA. 尤其是在预算较小的情况下，HC-SBM对于社区检测算法的攻击都要比DICE和SBA的效果好. 因为HC-SBM每次扰动的是对社区结构影响最大的链接，从

Table 4 Performance Comparison Between HC-SBM and Macro Community Hiding Baseline Algorithms

表 4 HC-SBM 与宏观社区隐藏基线算法性能对比

数据集	度量 指标	对比算法	Multilevel			Fast_greedy			Label_propagation			Leiden		
			1%	3%	5%	1%	3%	5%	1%	3%	5%	1%	3%	5%
Blogs	*NMI*	HC-SBM	**0.724**	**0.594**	**0.438**	**0.814**	**0.567**	**0.421**	**0.806**	0.671	0.567	0.836	**0.568**	**0.473**
		Pro-SBM	0.841	0.621	0.480	0.822	0.595	0.481	0.827	**0.596**	**0.534**	**0.792**	0.639	0.489
		RAT	0.897	0.737	0.727	0.839	0.772	0.764	0.894	0.842	0.802	0.862	0.698	0.680
	ARI	HC-SBM	**0.841**	**0.734**	**0.456**	**0.897**	**0.716**	**0.571**	**0.882**	0.768	0.681	0.901	**0.556**	**0.488**
		Pro-SBM	0.912	0.765	0.615	0.907	0.743	0.632	0.895	**0.704**	**0.642**	**0.892**	0.785	0.662
		RAT	0.950	0.861	0.859	0.922	0.880	0.873	0.946	0.909	0.887	0.932	0.838	0.824
	Jaccard	HC-SBM	**0.841**	**0.734**	**0.491**	**0.893**	**0.720**	**0.594**	**0.889**	0.793	0.726	0.895	**0.558**	**0.508**
		Pro-SBM	0.907	0.760	0.620	0.903	0.741	0.642	0.900	**0.744**	**0.699**	**0.889**	0.780	0.662
		RAT	0.946	0.854	0.852	0.917	0.875	0.868	0.948	0.913	0.893	0.927	0.831	0.824
Power	*NMI*	HC-SBM	**0.841**	**0.804**	**0.769**	**0.883**	**0.800**	**0.754**	**0.893**	**0.884**	**0.873**	**0.887**	**0.824**	**0.786**
		Pro-SBM	0.886	0.814	0.771	0.883	0.838	0.798	0.896	0.887	0.876	0.894	0.845	0.816
		RAT	0.852	0.819	0.801	0.890	0.839	0.811	0.901	0.889	0.889	0.900	0.866	0.857
	ARI	HC-SBM	**0.646**	**0.615**	**0.575**	**0.748**	**0.594**	**0.536**	**0.493**	**0.471**	**0.435**	0.781	**0.659**	**0.596**
		Pro-SBM	0.731	0.638	0.579	0.762	0.705	0.638	0.523	0.491	0.469	**0.770**	0.691	0.647
		RAT	0.693	0.632	0.592	0.770	0.700	0.649	0.540	0.506	0.503	0.801	0.732	0.713
	Jaccard	HC-SBM	**0.488**	**0.456**	**0.415**	**0.608**	**0.436**	**0.380**	**0.329**	**0.309**	**0.280**	0.649	**0.502**	**0.435**
		Pro-SBM	0.586	0.480	0.419	0.625	0.556	0.480	0.356	0.326	0.308	**0.635**	0.538	0.489
		RAT	0.541	0.473	0.432	0.636	0.550	0.493	0.371	0.340	0.337	0.676	0.587	0.563
Hep	*NMI*	HC-SBM	**0.854**	**0.819**	**0.794**	**0.923**	**0.839**	**0.801**	**0.920**	**0.933**	**0.892**	**0.878**	**0.824**	**0.802**
		Pro-SBM	0.875	0.861	0.814	0.929	0.846	0.815	0.947	0.942	0.929	0.891	0.861	0.837
		RAT	0.896	0.868	0.842	0.944	0.863	0.838	0.945	0.942	0.938	0.888	0.877	0.857
	ARI	HC-SBM	**0.514**	**0.493**	**0.422**	**0.820**	**0.617**	0.582	**0.621**	**0.601**	0.569	**0.561**	**0.535**	**0.501**
		Pro-SBM	0.520	0.562	0.439	0.824	0.618	0.580	0.644	0.676	**0.422**	0.596	0.552	0.537
		RAT	0.598	0.567	0.472	0.848	0.627	**0.559**	0.616	0.597	0.633	0.564	0.601	0.512
	Jaccard	HC-SBM	**0.351**	**0.346**	**0.277**	**0.710**	0.591	**0.403**	**0.469**	**0.428**	0.377	**0.429**	**0.380**	**0.339**
		Pro-SBM	0.358	0.397	0.289	0.719	**0.562**	0.424	0.476	0.512	**0.269**	0.432	0.388	0.374
		RAT	0.433	0.402	0.316	0.744	0.617	0.427	0.497	0.437	0.466	0.499	0.436	0.351
Astro	*NMI*	HC-SBM	**0.671**	**0.643**	**0.587**	0.631	**0.520**	**0.479**	**0.874**	**0.780**	**0.654**	**0.710**	**0.676**	**0.615**
		Pro-SBM	0.710	0.652	0.603	**0.626**	0.561	0.503	0.890	0.811	0.688	0.725	0.688	0.668
		RAT	0.688	0.647	0.657	0.726	0.606	0.600	0.910	0.851	0.847	0.742	0.692	0.696
	ARI	HC-SBM	**0.362**	**0.340**	**0.319**	0.353	**0.306**	**0.261**	**0.800**	**0.669**	**0.489**	**0.411**	**0.429**	**0.403**
		Pro-SBM	0.396	0.373	0.321	**0.344**	0.318	0.273	0.801	0.683	0.503	0.420	0.445	0.430
		RAT	0.348	0.363	0.372	0.577	0.364	0.334	0.848	0.735	0.738	0.472	0.451	0.444
	Jaccard	HC-SBM	**0.255**	**0.239**	**0.202**	0.303	**0.250**	**0.208**	**0.731**	**0.643**	**0.493**	**0.315**	**0.270**	**0.265**
		Pro-SBM	0.262	0.244	0.206	**0.270**	0.254	0.211	0.749	0.644	0.512	0.381	0.302	0.290
		RAT	0.277	0.247	0.246	0.460	0.284	0.263	0.802	0.689	0.690	0.394	0.374	0.300

注：1%，3%，5% 为扰动预算占网络总链接数的百分比；加粗部分为最小值，值越小宏观社区隐藏效果越好.

而以最小的预算得出最好的攻击效果. 同时也可以看到有时 SBA 的隐藏得分要优于 HC-SBM, 这可能是因为 SBA 倾向于移除社区边缘, 当大多数预算进行边缘移除时, SBA 的效果会高于 HC-SBM. 整体来看, HC-SBM 的隐藏性能比 DICE 和 SBA 更佳.

3)在微观社区隐藏方面. 本文实验主要统计了成功将目标节点移出原社区所需扰动的边数, 力求用最少的预算达到目标节点移出原社区效果, 实验结果如表 6 所示.

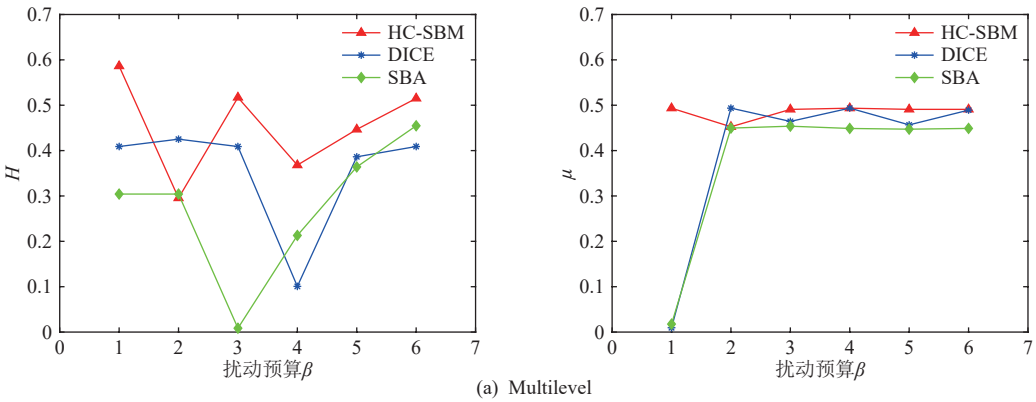
由表 6 可知, HC-SBM 在 4 种社区检测算法中都能将目标节点从原社区移出到其他社区中, 并且大部分情况下 HC-SBM 社区隐藏算法用的预算要小于 Dr 算法. 这种优势主要是由于 HC-SBM 能够挖掘出破坏网络社区结构最关键性的链接及链接的集合,

Table 5 Performance Comparison Between HC-SBM and Mesoscopic Community Hiding Baseline Algorithms

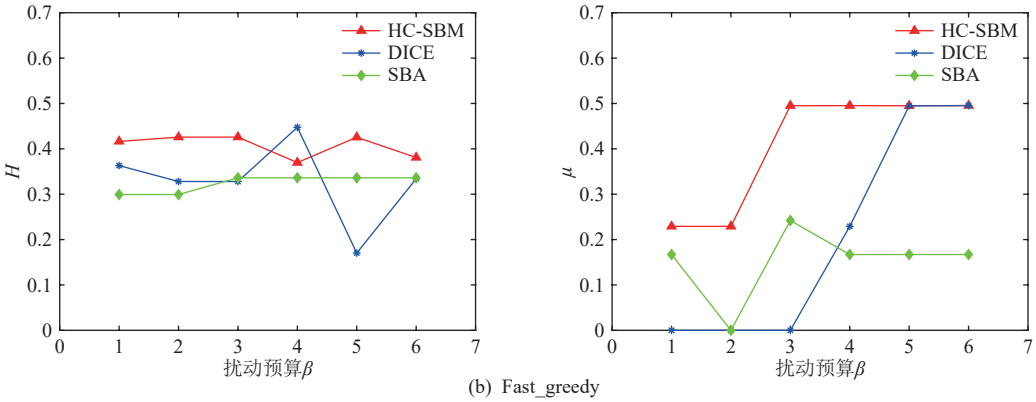
表 5 HC-SBM 与介观社区隐藏基线算法性能对比

数据集	对比算法	Multilevel		Fast_greedy		Label_propagation		Leiden	
		H	μ	H	μ	H	μ	H	μ
Blogs	HC-SBM	**0.515**	**0.491**	**0.381**	**0.497**	**0.500**	**0.497**	**0.305**	**0.487**
	DICE	0.395	0.489	0.334	0.495	0.350	0.492	0.272	0.486
	SBA	0.345	0.449	0.336	0.167	0.449	0.495	0.281	0.484
Power	HC-SBM	**0.485**	**0.776**	**0.445**	**0.581**	0.332	0.640	**0.454**	**0.563**
	DICE	0.264	0.183	0.354	0.180	**0.402**	0.460	0.430	0.553
	SBA	0.480	0.684	0.425	0.569	0.318	**0.770**	0.439	0.562
Hep	HC-SBM	**0.357**	**0.426**	**0.385**	**0.149**	**0.443**	**0.140**	**0.421**	**0.536**
	DICE	0.326	0.235	0.350	0.087	0.189	0.056	0.340	0.465
	SBA	0.102	0.143	0.362	0.053	0.312	0.076	0.400	0.378
Astro	HC-SBM	**0.402**	**0.431**	0.440	**0.647**	**0.491**	**0.341**	**0.413**	**0.754**
	DICE	0.340	0.401	**0.455**	0.524	0.375	0.253	0.327	0.731
	SBA	0.272	0.327	0.330	0.221	0.372	0.258	0.240	0.154

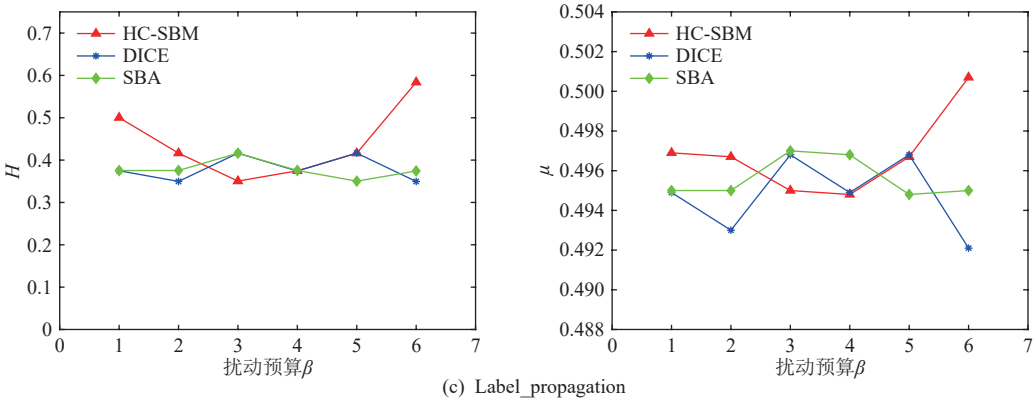
注：加粗部分为最大值，值越大介观社区隐藏效果越好.



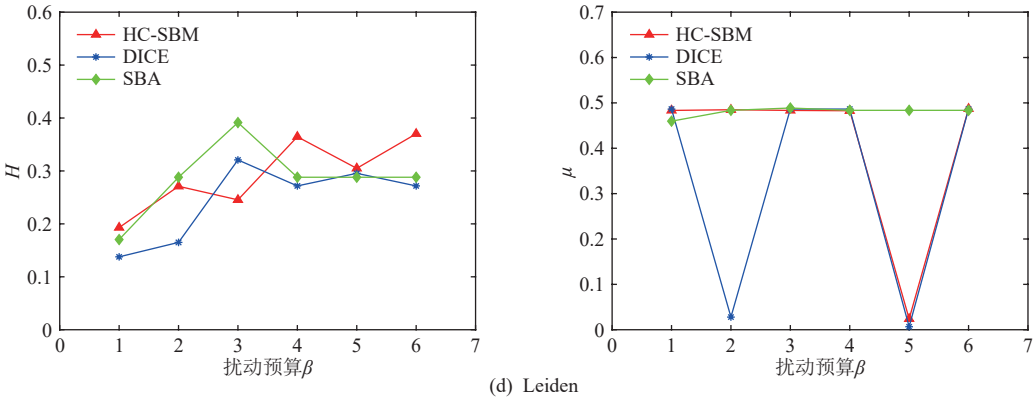
(a) Multilevel



(b) Fast_greedy



(c) Label_propagation



(d) Leiden

Fig. 2 Comparison of the metric H, μ between HC-SBM and the comparative algorithms under Blogs

图 2 Blogs 下 HC-SBM 与对比算法度量指标 H, μ 比较

Table 6 Performance Comparison Between HC-SBM and Micro-Community Hiding Baseline Algorithms

表 6 HC-SBM 与微观社区隐藏基线算法性能对比

数据集	对比算法	Multilevel	Fast_greedy	Label_propagation	Leiden
Blogs	HC-SBM	**1**	**3**	**2**	**1**
	Dr	3	6	6	1
Power	HC-SBM	**2**	**1**	**1**	2
	Dr	3	2	3	2
Hep	HC-SBM	**1**	**2**	**2**	**1**
	Dr	3	3	3	2
Astro	HC-SBM	**3**	**4**	**3**	3
	Dr	4	5	4	3

注：加粗部分为最小值，值越小微观社区隐藏效果越好.

每次扰动的是能够改变目标节点社区影响最大的边，从而以最小的预算策略达到最好的隐藏效果.

5 结 论

本文提出了一种基于随机块模型的社区隐藏算法 HC-SBM，该算法利用随机块模型刻画网络的生成机制，特别是网络社区形成和分裂的规律和模式，挖掘生成过程中关键性链接以及链接集合，通过引入节点重连概率矩阵，以最小代价扰动策略破坏网络社区结构，构建了微观、介观和宏观社区隐藏统一框架. 同时，在 4 个真实网络上进行了大量实验，并与 4 种主流社区隐藏算法进行多角度对比实验，结果表明，算法 HC-SBM 的社区隐藏效果更佳.

未来可行的研究方向包括从网络生成机制角度出发，刻画属性与链接在维持社区结构的规律和模式，实现属性网络的社区隐藏；其次，揭示时间维度在网络社区结构形成中的机制，探索动态网络社区隐藏方法和机理.

作者贡献声明：刘栋提出算法思路和实验方案，以及指导论文修改；刘侠负责完成实验并撰写论文；贾若雪协助完成实验并给出指导意见；张文生提出指导意见并修改论文.

参 考 文 献

[1] Albert R, Barabási A L. Statistical mechanics of complex networks[J]. Reviews of Modern Physics, 2002, 74(1): 47−97

[2] Newman M E. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167−256

[3] Zhou Yang, Cheng Hong, Yu Xu. Graph clustering based on structural/attribute similarities[J]. ACM Transactions on Accessible Computing, 2009, 2(1): 718−729

[4] Oukemeni S, Rifà-Pous H, Puig J M. Privacy analysis on microblogging online social networks: A survey[J]. ACM Computing Surveys, 2019, 52(3): 1−36

[5] Nagaraja S. The impact of unlinkability on adversarial community detection: Effects and countermeasures[C]// Proc of the 10th Int Symp on Privacy Enhancing Technologies. Berlin: Springer, 2010: 253−272

[6] Waniek M, Michalak T P, Rahwan T, et al. Hiding individuals and communities in a social network[J]. Nature Human Behaviour, 2018, 2(2): 139−147

[7] Fionda V, Pirrò G. Community deception or: How to stop fearing community detection algorithms[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(4): 660−673

[8] Chen Jinyin, Chen Lihong, Chen Yixian, et al. GA-based Q-attack on community detection[J]. IEEE Transactions on Computational Social Systems, 2019, 6(3): 491−503

[9] Liu Dong, Chang Zhengchao, Yang Guoliang, et al. Hiding ourselves from community detection through genetic algorithms[J]. Information Sciences, 2022, 614: 123−137

[10] Holland P W, Laskey K B, Leinhardt S. Stochastic block models: First steps[J]. Social Networks, 1983, 5(2): 109−137

[11] Tao Haicheng, Bu Zhan, Cao Jie. Community hidden framework based on multi-objective reinforcement learning[J]. SCIENTIA SINICA Informationis, 2021, 51(7): 1131−1145 (in Chinese)
(陶海成, 卜湛, 曹杰. 基于多目标强化学习的社区隐藏框架[J]. 中国科学: 信息科学, 2021, 51(7): 1131−1145)

[12] Newman M E. Modularity and community structure in networks[J]. Proceedings of the National Academy of Sciences, 2006, 103(23): 8577−8582

[13] Girvan M, Newman M E. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences, 2002, 99(12): 7821−7826

[14] Yang Liang, Cao Xiaochun, He Dongxiao, et al. Modularity based community detection with deep learning[C]//Proc of the 25th Int Joint Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2016: 2252−2258

[15] Bisma S K, Muaz A N. Network community detection: A review and visual survey[J]. arXiv preprint, arXiv: 1708.00977, 2017

[16] Barnes E R. An algorithm for partitioning the nodes of a graph[J]. SIAM Journal on Algebraic Discrete Methods, 1982, 3(4): 541−550

[17] Yang Gui, Zheng Wenping, Che Chenhao, et al. Graph-based label propagation algorithm for community detection[J]. International Journal of Machine Learning and Cybernetics, 2020, 11(6): 1319−1329

[18] Fan Lilin, Song Kaiyuan, Liu Dong. A noise reduction method for semi-supervised community detection based on harmonic function[J/OL]. International Journal of Modern Physics B, 2018, 32(14)[2023-09-11].https://www.x-mol.com/paper/130828782641948 6720?adv

[19] Palla G, Derenyi I, Farkas I, et al. Uncovering the overlapping community structures of complex networks in nature and society[J]. Nature, 2005, 435(7043): 814−818

[20] Jaswant M, Devi V S. Overlapping community detection in social network using disjoint community detection[C]//Proc of the 8th Symp Series on Computational Intelligence. Piscataway, NJ: IEEE, 2015: 764−771

[21] Li Jia, Zhang Honglei, Han Zhichao, et al. Adversarial attack on community detection by hiding individuals[C] //Proc of the 29th Int World Wide Web Conf. New York: ACM, 2020: 917−927

[22] Chen Xianyu, Jiang Zhongyuan, Li Hui, et al. Community hiding by link perturbation in social networks[J]. IEEE Transactions on Computational Social Systems, 2021, 8(3): 704−715

[23] Liu Yiwei, Liu Jiamou, Zhang Zijian, et al. Rem: From structural entropy to community structure deception[C] // Proc of the 33rd Conf on Neural Information Processing System. Cambridge, MA: MIT, 2019: 12938−12948

[24] Liu Dong, Yang Guoliang, Wang Yanwei, et al. How to protect ourselves from overlapping community detection in social networks[J]. IEEE Transactions on Big Data, 2022, 8(4): 894−904

[25] Yang Guoliang, Wang Yanwei, Chang Zhengchao, et al. Overlapping community hiding method based on multi-level neighborhood information[J]. Symmetry, 2022, 14(11): 23−28

[26] Liu Xueyan. Research on stochastic block models for complex network analysis [D]. Changchun: Jilin University, 2021(in Chinese)
(刘学艳. 面向复杂网络分析的随机块模型研究[D]. 长春: 吉林大学, 2021)

[27] Lada A, Natalie S. The political blogosphere and the 2004 U. S. election: Divided they blog[C]//Proc of the 11th ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2005: 36−43

[28] Haiko L W, Duncan J. Collective dynamics of small world networks[J]. Nature, 1998, 393(6684): 440−442

[29] Newman M E. The structure of scientific collaboration networks[J]. Proceedings of the National Academy of Sciences, 2001, 98(2): 404−409

[30] Razieh H, Alireza R. AntLP: Ant-based label propagation algorithm for community detection in social networks[J]. CAAI Transactions on Intelligence Technology, 2020, 5(1): 34−41

[31] Traag A V, Waltman L, Eck V J. From Louvain to Leiden: Guaranteeing well-connected communities[J]. arXiv preprint, arXiv: 1810.08473, 2019

[32] Vincenet D B, Jean-Loup G, Renaud L, et al. Fast unfolding of communities in large networks[J/OL], Journal of Statistical Mechanics: Theory and Experiment, 2008, 2008(10)[2023-09-11].https://www.x-mol.com/paper/1415733752522543104?adv

[33] Liu Xuecheng, Fu Luoyi, Wang Xinbing, et al. ProHiCo: A probabilistic framework to hide communities in large networks[C] // Proc of the 40th IEEE Conf on Computer Communications. Piscataway, NJ: IEEE, 2021: 1−10

[34] Danon L, Diaz-Guilera A, Duch J, et al. Comparing community structure identification[J]. Journal of Statistical Mechanics: Theory and Experiment, 2005, 2005(9): 219−228



Liu Dong, born in 1976. PhD, professor. His main research interests complex network analysis and educational data mining.

刘　栋, 1976 年生. 博士，教授. 主要研究方向为复杂网络分析和教育数据挖掘.



Liu Xia, born in 1996. Master candidate. Her main research interest includes complex network analysis.

刘　侠, 1996 年生. 硕士研究生. 主要研究方向为复杂网络分析.



Jia Ruoxue, born in 1998. Master candidate. Her main research interest includes complex network analysis.

贾若雪, 1998 年生. 硕士研究生. 主要研究方向为复杂网络分析.



Zhang Wensheng, born in 1965. PhD, professor. His main research interests include artificial intelligence, statistical machine learning, big data pattern mining, medical data analysis and reasoning, and computer vision.

张文生, 1965 年生. 博士，教授. 主要研究方向为人工智能、统计机器学习、大数据模式挖掘、医学数据分析与推理、计算机视觉.