

社交网络中的社团隐私研究综述

蒋忠元, 陈贤宇, 马建峰

(西安电子科技大学网络与信息安全学院, 陕西 西安 710000)

摘 要: 社团是社交网络的重要特征, 社团检测技术的发展给网络用户带来隐私泄露的危险。如何保护敏感的社团信息不被泄露, 保障用户与社团安全已经成为网络安全领域的研究热点。近几年, 社团保护技术取得了初步进展, 但针对社交网络中的社团隐私或社团安全研究进展综述较少, 不利于该研究方向的长远发展。因此, 主要针对社团结构隐私方面的研究进展进行综述, 对最新的相关研究成果进行系统地归纳、总结、比较, 并展望了未来的研究方向。

关键词: 社交网络; 社团检测; 社团隐私; 社团安全; 社团隐藏

中图分类号: TP309

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2021021

Survey of community privacy in social network

JIANG Zhongyuan, CHEN Xianyu, MA Jianfeng

School of Cyber Engineering, Xidian University, Xi'an 710000, China

Abstract: Community is an important feature of social network and the development of community detection technology brings the danger of privacy disclosure to network users. How to protect sensitive community information from being leaked and ensure the security of users and communities has become a research hotspot in the field of network security. In recent years, community privacy protection technology has made initial progress, but there is a few survey on the research of community privacy or community security in social networks, and it may limit the potential and long-term development of this research topic. The research on the privacy of community structure was mainly reviewed and the related works on community security were classified, summarized, compared. The hot issues of community security in the future were proposed.

Keywords: social network, community detection, community privacy, community safety, community hiding

收稿日期: 2020-04-11; 修回日期: 2020-08-02

通信作者: 陈贤宇, xychen_3@stu.xidian.edu.cn

基金项目: 国家自然科学基金(61502375); 陕西省自然科学基金基础研究计划(2020JM-203); 中央高效基本科研业务费项目(10223196150)

Foundation Items: The National Natural Science Foundation of China(61502375), The Natural Science Basis Research Plan in Shaanxi Province of China (2020JM-203), The Fundamental Research Funds for the Central Universities (10223196150)

论文引用格式: 蒋忠元, 陈贤宇, 马建峰. 社交网络中的社团隐私研究综述[J]. 网络与信息安全学报, 2021, 7(2): 10-21.

JIANG Z Y, CHEN X Y, MA J F. Survey of community privacy in social network[J]. Chinese Journal of Network and Information Security, 2021, 7(2): 10-21.

1 引言

随着科学技术的发展,人们日常生活的方方面面越来越离不开形态各异的网络,文献[1]对复杂网络概念进行了表述。复杂网络是由大量的节点以及节点之间错综复杂的连接关系共同构成的网络结构,它是人们日常生活中各种复杂系统的抽象表达,如社会网络^[2-3]、生物网络^[4]、电力网络^[5]、金融网络^[6]等。网络科学由于其高度的跨学科性质以及与社会生活相关联的高度密切性质吸引着一大批专家学者们的广泛关注与研究,他们提出了许多有意义的研究方向并获得了一系列可观的研究成果,如链路预测^[7-8]、节点分类^[9]、网络重构^[10]、网络舆情控制与阻断^[11-12]、生物学学科中对未知生物功能的预测^[13-14]、计算机学科中的社团检测算法^[15-17]等问题。

近年来,社交网络的社区(community)也被称为社团、圈子、模块等,作为社交网络科学的一个重要组成部分而被广大研究员们进行深入研究。研究表明,社交网络中的一个社区代表了该网络中联系最密切的相似节点之间的网络拓扑,不同的社区之间代表了不同群体之间的差异性或联系的稀疏性^[16-20],而针对社区结构发掘的社团检测算法为社团发挥其重要价值起到了关键作用。例如,兴趣组推荐主要依赖于社团的用户往往有较相似的兴趣爱好以及紧密的联系。其次,社团的划分也可以被运用在关联信息、关联好友推荐系统等。然而社团本身富含用户与群体的隐私信息,敏感社团信息泄露可被攻击者用来从事不良活动。例如,基于社团的谣言扩散,因为相同社团的成员对某种谣言信息的敏感性近似,研究表明社团会加快谣言传播、加大传播范围^[12]。综上,社团机制在社交网络中是一把双刃剑,既有利于人们的社交活动,又会给用户带来安全隐患。

针对社交网络的隐私保护技术已较成熟^[21],而针对社团隐私保护的研究尚处于发展阶段。社团隐私与社交网络隐私有较大的区别,主要体现在对象与范围不同。社团隐私面向具体的社团,而社交网络的隐私则包括整体网络、部分网络、单个用户安全等。本文涉及的社团隐私主要聚焦在社团结构的隐私问题。本文主要从以下几方面

进行介绍。

1) 社团隐私研究基础。本文首先介绍社交网络中社团隐私问题的相关概念;其次,由于社团是研究社团隐私的基础,因而本文将回顾现有的社团检测相关工作^[15-17];最后将引入社团隐私保护的研究动机和目标等。

2) 社团安全或社团隐私。社团检测研究导致的社团暴露可能会造成严重的社团与用户隐私泄露,引发社团安全危机。为此人们逐渐开始研究社团安全隐私的保护工作,本文将详细介绍现有社团保护技术的核心思想和方法路线。

3) 总结与未来工作展望。本文针对目前社团隐私技术的研究,总结研究中的不足之处,对未来可能出现的热门研究点进行了展望,为未来社团隐私保护技术的研究提供基础。

2 社交网络中社团隐私研究基础

2.1 社交网络中的社团隐私问题

在科技发展迅速的时代,个体与个体间的信息沟通会产生具有不同含义的社交网络,如邮件社交网络、同事社交网络、朋友亲戚社交网络等。近十几年来,社团检测技术作为研究社交网络特性的重要入口一直占据着研究人员的研究视线,他们致力于探索如何更加高效快捷地挖掘网络社团划分结构。然而,随着社团检测研究技术的发展,人们逐渐受到该技术的滥用应用导致用户个人隐私泄露等安全问题的困扰。与此同时,用户对于个人隐私保护的意识逐步提高,用户更加关注社交网络的社团隐私保护问题,而不再无止尽地在社交网络中挖掘各种潜在的社区信息,因为社交网络的社区检测研究所追求的对网络社团结构进行深度剖析的想法势必会带来安全隐患。因此,如何在社交网络中通过隐藏社团来规避社交网络的社团检测工具就显得异常重要。

社团检测技术的发展和用户对社团隐私保护的要求形成技术冲撞,而且这种技术冲撞是不平衡不对等的,现有的研究视线应该关注社团隐私保护技术的研究,努力实现上述两种技术的平衡发展。现有的社团隐私保护技术的研究主要分为两种思路:① 基于目标社团(target-community)的社团隐私保护研究^[22-26]:该研究认为网络中某

个目标社团结构（或者有隐私保护需求的群体）是敏感的，需要通过社团隐藏技术打破该目标社团在网络中的社团划分结构，如对社交网络结构进行适当扰动（删除真实链路、增加虚假链路、删除真实节点或增加干扰节点等）来降低现有社团检测算法检测出目标社团的概率，以此逃避社团检测技术对该社团结构的挖掘。② 基于全社团（all-communities）的社团隐私保护研究^[27-28]：该研究默认社交网络中的全部社团结构都是敏感的，需要通过一种或多种社团隐私保护技术对所有的社团进行隐藏操作，以此达到全社团隐私保护的目标。针对目标社团隐私保护的技术更加贴近实际需求，实施代价相对较小，并且能较好地保持原始网络的可用性。而全社团保护技术使网络的可用性损失较大，实施代价略高。

2.2 社交网络中社团检测的相关研究

社团是社交网络中社团隐私保护的重要基础结构，而社团检测是发掘网络中社团结构的主要方法^[29-30]。从是否具有重叠社团的角度将现有的社团检测方法分为非重叠社团检测算法（non-overlapping community detection algorithm）与重叠社团检测算法（overlapping community detection algorithms）。非重叠社区检测算法旨在挖掘无节点相交的社团划分结构；而重叠社团检测算法需要考虑某些节点可以同时存在于不同社团内部的特性对检测性能的影响，以此来发掘存在相交节点集的社团划分。

初期的非重叠社团检测算法主要采用层次聚类思想，包括自底向上层次聚类（聚合形式）和自顶向下层次聚类（分裂形式）。Girvan 等^[16]提出了基于边介数（edge betweenness）的社区检测算法，边介数主要包括最短路径中的边介数、随机行走边介数等。Newman 等^[17]提出了基于模块度优化（modularity optimization）的方法，该方法通过寻找模块度最大的网络划分来发现社区划分结构。此外，Newman^[31]从矩阵谱分析的角度提出了基于模块度矩阵谱分析的社区发现算法。为了改进文献[16]中 GN 算法的性能，Newman^[32]在 2004 年又结合检测算法提出了新的快速社团发现算法，该算法是一个聚合类检测算法，其假

设每个节点作为一个单独的社团，每次合并的两个社团操作都能使整个网络的模块度 Q 产生最大的增量，直至模块度无法增加。模块度的定义如式(1)所示。

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(C_i, C_j)$$

$$P_{ij} = \frac{1}{2m} d_i d_j \quad (1)$$

其中， A 为邻接矩阵，当节点 i 和节点 j 属于同一个社团时， $\delta=1$ ，否则， $\delta=0$ ； d_i 为节点 i 的度数， m 为网络中的总边数。文献[33]为了提高检测算法的性能水平，基于快速 GN 算法改进了相应的数据结构。Li 等^[34]提出了一种同时考虑社团内部边数以及社团外部边数的模块密度的度量函数；Rosvall^[35-36]提出了一种基于互信息（mutual information）的目标函数来发现社区划分；Blondel^[37]提出了一种基于贪婪层次聚类 and 局部模块度最优化思想的 BGLL 算法；除上述算法以外，研究人员还提出了许多其他的非重叠社团检测算法，如基于标签传播检测算法^[19]、遗传算法^[38]、随机游走算法^[39-40]等。

大规模网络结构通常极其复杂，并非单纯地由非重叠社区组合而成，而是由一个个重叠的社团构成的^[41-42]。重叠社区的社团检测算法主要分为以下 4 类。① 基于边划分的检测算法：Ahn 等^[43]提出了一种重叠社区检测算法——ABL（Ahn, Bagrow, Lehmann）算法，该算法将挖掘紧密相连的点社团转变为挖掘网络中紧密相连的边社团；Kim 等^[44]扩展了 InfoMap 算法到边图上；Ye 等^[45]提出了一种多解析度的边社团发现算法 MLP。② 基于优化的重叠社区检测算法：Lancichinetti 等^[46]提出了一种从随机选取的种子节点开始构造社区结构，终止条件为适应度函数获得局部最大化值；OSLOM 算法通过最优化一种社区随机波动的统计适应度函数来发现重叠社区^[47]；基于优化的重叠社区检测算法还包括 OCA 算法^[48]、GCE 算法^[49]、COCOD 算法^[50]。③ 基于模糊划分的重叠社区发现算法：Zhang 等^[51]提出了一种基于 LDA 和 K-means 聚类算法的检测算法；Psorakis^[52]对网络进行非负矩阵分解后进行特征抽取并使用降维算法来挖掘社团结构；Nepusz^[53]使用模拟退火

算法来解决重叠社区检测问题；Newman 等^[54]提出了使用概率图模型的社区划分方法；Airoldi 等^[55]从统计学角度提出了随机分块模型；④ 基于节点特性的社区检测算法：Abrahao 等^[56]提出了使用中介系数来划分社区的 CONGO 算法；Jiang 等^[57]通过使用相似度比较的方式提出了一种对抗女巫攻击的用于大规模网络的社团检测方法等。有关社团检测的详细内容参考文献[58-59]。

2.3 社团隐私保护研究的动机

社团划分的主要动机是检测网络结构中存在的社团信息，其有助于种群发现、蛋白质组功能分析、兴趣组挖掘等。社团保护的本质是社团检测的逆工作，其研究动机主要有两方面。

社团隐私保护（社团欺骗、社团隐藏）技术对于想要在社交网络中隐藏的用户（作为一个群体）是有利的。例如，网络警察需要以便衣的身份在社交网络中进行犯罪取证，需要有点保障。第一，便衣警察需要与普通的社交用户进行正常的社交活动（即保留每位用户参与社交活动的权利），以免引起非法用户的察觉而销毁罪证或逃离。第二，犯罪团伙的警觉性比普通人高，即信息、行为模式等仅在团伙内部共享。同时犯罪团伙需要不断吸纳新成员来壮大其不法组织并获取利润。此时，便衣警察可以趁机打入犯罪团伙的社团，即卧底形式。与现实生活中的卧底形式相似，所有便衣警察及其所在的社团需要采取一些措施，包括与目标社团成员请求建立链接关系，同时疏散警员之间的链接关系，从而取得犯罪团伙多数成员的链接建立请求。犯罪团伙认为某便衣警察符合其社团成员标准，即符合社团划分的结果，则接纳该警察为团伙成员，共享组织内部信息以及活动。因此，在这种情况下，警员社团是敏感的，是需要被保护的对象。综上所述，积极的社团隐私保护技术有助于降低网络犯罪、净化网络环境、保障用户以及整个社会的安全。

与此同时，社团保护技术可能被恶意利用。例如，犯罪组织想在不被察觉的情况下采用与上述类似的社团渗透等手段渗透到各个目标社团，收集社团内的用户个人资料以及共享的私密信息

等，从而锁定目标受害者进行敲诈、勒索、虚假信息传播等犯罪活动。恶意的社团欺骗等方法对社会的危害极大，会降低用户对社交网络的信任度，引发信任危机，甚至给社会带来巨大的负面影响。

因此，社团隐私保护技术是一把双刃剑。已有科研人员在社团防护方面做了相关的研究工作。接下来将详细对其中的工作进行介绍。

社团检测与社团保护的基础是真实的社团数据，表 1 展示了一些被广泛使用的真实社交网络数据集。

表 1 相关工作所使用的社交网络数据集
Table 1 Social network dataset used in related work

社交网络	节点数目	边数目	描述
kar	34	78	Zachary Karate's Club
dol	62	159	Dolphins association
mad	62	159	Madrid Train Bombing
lesm	77	154	Co-appearances of characters in Les Misérables
polb	105	441	Books about US politics
words	112	425	Adjectives/nouns in the novel David Copperfield
nets	1 589	5 486	Net Theory co-authors
erdos	6 100	9 939	Erdos collaborations
faceb	6 387	2.17×10^5	Facebook
pow	4 941	6 594	USA Power Grid
astr	1.79×10^4	1.97×10^5	Astrophysics co-authors
4sq	6.39×10^5	3.2×10^6	Foursquare
dblp	5.405×10^5	1.52×10^7	DBLP
amz	5.48×10^5	9.25×10^5	Amazon metadata
you	1×10^6	3×10^6	Youtube dataset
ork	3×10^6	1.172×10^8	Orkut social network

3 社交网络中社团隐私保护方法

随着研究人员对社团安全的关注度越来越高，他们的研究成果不断涌现。本文将现有的社团隐私保护方法归结为两类主要研究思路：基于目标社团（target community）和基于全社团的社团隐私保护，本节将分别以这两类研究思路为切入点对各种算法的核心思想、基本技术、研究成果以及评价指标等进行综述性研究，并对各自存在的问题和挑战进行分析与说

明。

3.1 基于目标社团的社团隐私保护

基于目标社团的隐私保护技术的研究方法主要通过删除或增加相应数量的合适链路边来扰动目标社团在网络中的内部结构以及目标社团与外部成员的联系程度,使目标社团尽可能地隐藏于网络中。据此,现有的研究有几种不同的方法:基于社团安全性优化和模块度的社团隐藏算法^[22]、基于新的社团安全性优化的社团隐藏算法^[23]、基于模块度启发式的社团隐藏算法^[24]、基于节点集模块度的社团隐藏算法^[25]以及基于模块度和遗传算法的社团隐藏算法^[26]。

(1) 基于社团安全性优化和模块度的社团隐藏算法

文献[22]提出了社区欺骗 (community deception) 的概念,并定义了一个衡量社团欺骗程度的指标,即欺骗分数 (deception score)。任意给定一个社交网络 $G=(V,E)$ 和社团检测算法 A_d ,该检测算法可获得一个社团划分集合 $\bar{C}=\{C_1, C_2, \dots, C_k\}$,待隐藏的目标社团定义为 $C \subseteq V$ 。当社团 C 被 A_d 完全发掘时,说明该隐藏技术获得最差的结果,即 $C \in \bar{C}$;另外,当 $C \notin \bar{C}$ 时, C 中的成员可以以多种不同的方式被隐藏在 $\bar{C}$ 中。在引入社团欺骗概念之前,文献[22]首先指明将 C 中的成员以最好的方式隐藏在 $\bar{C}$ 中的 3 个必要条件:信息流动性保护、目标社区成员的扩散性程度以及成员分布的社区规模程度。据此,该文献定义了 A_d 对应于某个社区 C_i 的召回率 (R, recall) 和精确度 (P, precision),分别如式(2)和式(3)所示,其中 $\zeta_C^{C_i}$ 表示网络在社团检测算法 A_d 的检测结果下,目标社团 C 的成员在社团 C_i 中的个数。

$$R(C_i, C) = \frac{\zeta_C^{C_i}}{|C|}, \forall C_i \in \bar{C} \quad (2)$$

$$P(C_i, C) = \frac{\zeta_C^{C_i}}{|C_i|}, \forall C_i \cap \bar{C} \neq \emptyset \quad (3)$$

随后,文献[22]定义了目标社团欺骗分数 $H(C, \bar{C})$,如式(4)所示, $|S(C)|$ 表示社团 C 中的成员在子图中连通组件的数目。

$$H(C, \bar{C}) = \left(1 - \frac{|S(C)| - 1}{|C| - 1}\right) \cdot \left(\frac{1}{2} (1 - \max_{C_i \in \bar{C}} \{R(C_i, C)\})\right) + \frac{1}{2} \left(1 - \frac{\sum_{C_i \cap C \neq \emptyset} P(C_i, C)}{|C_i \cap C \neq \emptyset|}\right) \quad (4)$$

文献[22]通过重新排列社团 C 中成员的 β 条边来提高 C 的隐藏得分 $H(C, \bar{C})$,为了获取最合适的 β 条边,定义了节点 u 相对于社团 C 的安全性 (node safeness) 的概念,如式(5)所示。

$$\sigma(u, C) = \frac{1}{2} \frac{|V_C^u| - |E(u, C)|}{|C| - 1} + \frac{1}{2} \frac{|\tilde{E}(u, C)|}{\deg(u)} \quad (5)$$

其中, $V_C^u \subseteq C$ 表示节点 u 与社团 C 中节点存在联通路径的节点集, $\deg(u)$ 代表节点 u 的度数, $E(u, C)$ 和 $\tilde{E}(u, C)$ 分别代表节点 u 与社团 C 内部其他节点之间存在的边的集合和节点 u 在目标社团 C 与外部社团节点之间存在的边的集合。

紧接着,该文献定义了社团安全性,如式(6)所示。

$$\sigma(C) = \sum_{u \in C} \sigma(u, C) / |C| \quad (6)$$

根据式(6),提出社区欺骗方法 D_s ,即当安全增益 $\xi_C = \sigma(C') - \sigma(C)$ 以贪心策略获得最大增量时,此时的边更新操作能够使目标在社团获得最佳安全性,其中 C' 是执行一次或多次边更新后的目标社团。

文献[22]提出了基于模块度的对比方法 D_m 和 D_r 。 D_m 方法使用模块度的损失度量来指导重新连边操作,模块度的损失定义为 $ML = M_G(\bar{C}) - M_{G'}(\bar{C})$,其中, $M_G(\bar{C})$ 表示网络的模块度,如式(7)所示。

$$M_G(\bar{C}) = \frac{\eta}{m} - \frac{\delta}{4m^2} \quad (7)$$

其中, $\eta = \sum_{C_i \in \bar{C}} |E(C_i)|$, $\delta = \sum_{C_i \in \bar{C}} \deg(C_i)^2$, $\deg(C_i) = \sum_{v \in C_i} \deg(v)$ 。 D_r 方法随机选择各种类型的节点间连边进行删除或添加操作,实验结果表明, D_r 方法很难达到隐藏目标社团的效果,该方法的欺骗分数 H 约等于 0。

针对该文献研究的不足,未来的研究途径有:首先,考虑设计更加健全或通用的计算社团安全

性的公式；其次，实验仅考虑在社团内部删除边操作以及在社团之间增加新边操作，社团内部加边和社团之间删除边操作不具有现实意义，从而导致在实验中人为排除了这两种操作的可行性，然而在约束条件满足之下，后两种方案也可能产生正面影响。

(2) 基于新的社团安全性优化的社团隐藏算法

为了消除上述方案中的约束条件对社团隐藏的影响，文献[23]提出了优化后的社团隐藏方案 H_s ，经过优化后的方案明确证明了在社团内部删除边操作和在社团之间增加边操作的可行性以及在社团内部增加边操作和在社团之间删除边操作的不可行性。文献[23]首先提出社团的离散度的概念，定义为 $\rho(C) = \sum_u \sum_v \text{sp}(u, v)$ ，其中 $u, v \in C$ ，且 $u \neq v$ ， $\text{sp}(u, v)$ 表示节点 u 和 v 之间的最短路径长度。该文献对其进行归一化表达以此来评估目标社团内部的安全性，当一个目标社团的节点个数为 n 时，其社团内部安全性定义如式(8)所示。

$$\psi(C) = \frac{\rho(C) - 2(n-1)^2}{\sum_{k=1}^n \left(\sum_{i=1}^{n-k} i + \sum_{j=0}^{k-1} j \right) - 2(n-1)^2} \quad (8)$$

该文献沿用了文献[22]中对于目标社团与社团外部之间安全性的定义，如式(9)所示。

$$\phi(C) = \sum_{u \in C} \frac{|\tilde{E}(u, C)|}{\deg(u)} \quad (9)$$

综合目标社团内部安全性和目标社团与社团外部安全性的评估方案，提出了均衡性的目标社团安全性的定义，如式(10)所示。

$$\sigma'(C) = \frac{1}{2}\psi(C) + \frac{1}{2}\phi(C) \quad (10)$$

图1演示了 H_s 的流程，相同颜色的节点集代

表同一个社团，该原始网络包含4个社团，如图1(a)所示，其中红色节点所在的社团为目标社团，算法经过启发式的删边操作和加边操作后得到新的社交网络 G' ，经过二次社团检测后，重新获得网络中的社团结构，由图1(d)可知，经过社团隐藏算法保护的目标社团成员被成功隐藏到其他社团中。

文献[23]在文献[22]的基础上设计了新的计算社团安全的公式，排除了文献[22]中证明部分的不利之处，然而该文献的研究关注点仍然比较单一，如该方法只针对非重叠社团检测技术且仅对静态网络结构进行研究等。

(3) 基于模块度启发式的社团隐藏算法

文献[24]提出了一个针对社团隐私保护的基于模块度的启发式算法——DICE (disconnect internally, connect externally) 算法，该算法给定一个操作预算 b ，首先随机删除 d 个目标社团 C 的内部链路，然后随机添加 $b-d$ 个 C 的成员与其他社团成员之间的链路来进行社区的隐藏操作。为了评估 DICE 算法在一个网络社团划分结果 $CS = \{C_1, C_2, \dots, C_k\}$ 下，隐藏社团 C 的效果，文献[24]提出了两个基于不同角度进行隐藏评价的指标 (u' 和 u'')，分别如式(11)和式(12)所示。式(11)表明了 C 的成员在 CS 中的分布情况，而式(12)展示了 C 在网络中的隐藏效果。综合 u' 和 u'' ，文献[24]提出了一个均衡的社团隐藏评价指标 u ，如式(13)所示。

$$u'(C, CS) = \frac{|\{C_i \in CS : C_i \cap C \neq \emptyset\}| - 1}{\max(|CS| - 1, 1) \max_{C_i \in CS} (|C_i \cap C|)} \quad (11)$$

$$u''(C, CS) = \sum_{C_i \in CS} \frac{|C_i \setminus C|}{\max(n - |C|, 1)} \quad (12)$$

$$u(C, CS) = \alpha u'(C, CS) + (1 - \alpha) u''(C, CS) \quad (13)$$

其中 $\alpha \in [0, 1]$ 。 $u(C, CS) = 0$ 表明 C 完全暴露于检

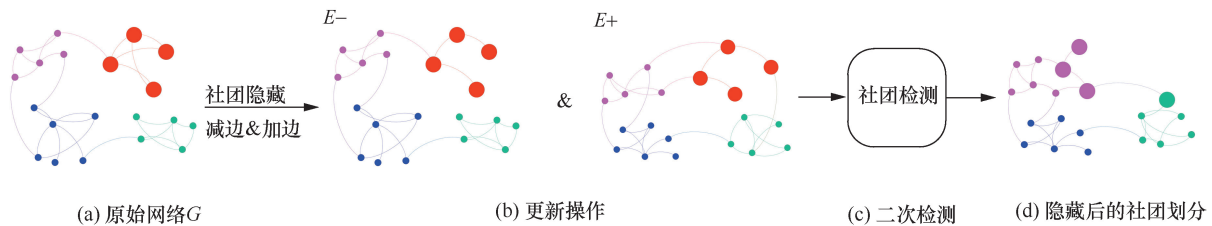


图1 社团隐藏算法 H_s 的流程
Figure 1 Schematic diagram of community hiding algorithm H_s

测算法之下, 此时隐藏效果最差; 相反地, $u(C, CS)=1$ 表明了 C 的所有成员均匀地被隐藏在其他社团中。

综上, 文献[24]的研究突破了小型网络变化对实验的影响, 并且对中心度测量和社区检测算法的灵敏度分析进行了扩展; 另外, 文献[24]的研究发现有助于绘制出在社交网络中保护隐私的界限, 同时揭示了在安全应用程序中使用通用的社会网络分析工具的含义。然而该文献仍未指明如何从链路预测算法的角度隐藏个体或社团间的关系或者如何通过特征向量中心来逃避检测等。

(4) 基于节点集模块度的社团隐藏算法

文献[25]提出了通过隐藏一个目标社区来抵抗社区检测算法的思想, 首先定义了诱导边(induced edge)和诱导图(induced graph)的概念。给定一个网络 $G=(V, E)$, 假设 $S \subseteq V$ 是图 G 的一个节点子集, 则节点集 S 中的节点所包含的边组成了诱导边, 记为 $L(S, E)$, 由 S 和 $L(S, E)$ 构建的子图被定义为诱导图 $I(G, S)$ 。由此定义了节点集的模块度 (S, V) -modularity, 如式(14)所示。

$$Q(S, V) = \frac{a(S) - r(S)}{\sum_{i \in S} d_i^G} \quad (14)$$

其中, d_i^G 是图 G 中节点 i 的度数, $a(S)$ 是节点集 S 中所有节点在诱导图中的度数总和, $r(S)$ 是节点集 S 在随机图 (G_{random}) 中诱导图中的度数总和, G_{random} 是与图 G 具有相同节点数的随机网络, 且每条边出现的概率为 $d_i d_j / 2|V|$ 。

根据上述定义, 文献[25]提出了目标社区隐藏模型, 即 (α, β) -hidden community, 该模型必须满足以下 3 个条件: ① 目标社团的节点集 S 必须是节点集 V 的通用子集; ② 诱导图 $I(G, S)$ 的边数至少为 $\alpha \cdot |V|$; ③ 诱导图 $I(G, S)$ 的模块度大小至少为 β , 即 $Q(S, V) \geq \beta$ 。该文献通过增加新的边来使隐藏模型满足上述 3 个条件。该文献引入丢失率 (miss-ratio) 的概念来评价社区隐藏的效果, 具体的定义如式(15)所示。

$$A((\alpha, \beta), \Omega) = \frac{V - \text{FN} \cdot S}{|V - S|} \quad (15)$$

其中, Ω 代表某种社团检测策略, FN 表示目标社团节点被错误的环划分至节点集 $V - S$ 中的比例。

文献[25]提出的关于社团隐藏的研究时间较早, 实验中只关注于向具有高中心性值(如度中心性)的节点添加链路, 关注点单一; 此外, 该文献仅对邮件网络进行了实验, 应在大规模网络上进行扩展。

(5) 基于模块度和遗传算法的社团隐藏算法

文献[26]引入了社区检测攻击的问题, 并通过重新连接少量边制定有效的策略来攻击社区检测算法。全文提出了 4 种社团欺骗方法: 随机攻击(RA)策略、社区检测攻击(CDA)、基于节点度的攻击(DBA)、基于遗传(GA)算法的模块度(GA-based Q-attack)策略。该文献将模块度(modularity)和标准化互信息(normalized mutual information)作为评价指标, 实验结果表明, 在考虑大多数的情况下, GA-based Q-attack 的性能优于 RA、CDA 和 DBA。

基于遗传算法的模块度策略将社团的隐藏技术类比于遗传过程, 需经过基因初始化、优势选择、交叉匹配和基因突变 4 个过程, 其中优势选择过程中的适应度函数受模块度的启发, 该适应度函数的定义如式(16)所示, 其中 Q 是社交网络的模块度。

$$f = e^{-Q} \quad (16)$$

虽然 GA-based Q-attack 方法有一定优势, 但该方法仅以模块度作为单优化目标, 然而基于多目标优化的方法也值得深入研究。此外, 通过简化遗传算法中适应度计算和选择时间复杂度较低的优化算法来设计效率更高的攻击策略也是未来的研究重点之一。

表 2 归纳总结了上述基于目标社团隐私保护的相关工作。由表 2 可知, 文献[25]仅使用了添加新的链路操作, 未考虑删除链路操作对社团隐私保护的影响; 文献[24-26]仅针对模块度优化检测技术进行社团隐私的保护, 然而, 目前还存在其他大量的检测算法未被加以考虑, 如随机游走算法^[2]或标签传播算法^[19]等。此外, 文献[25]在进行社团隐私保护前需要获得网络中每一个节点的中心性值, 因而相比于其他只需要目标社团节点所包含的边的方法而言, 文献[25]所需的网络信息更多。文献[22-23]提出了一种有针对性的被称为社区安全的欺骗目标函数, 并通过对目标函数

表2 基于目标社团中隐私保护的相关工作对比
Table 2 Comparisons of related work based on privacy protection on target community

文献	操作	技术	所需网络知识	不足之处
文献[22]	增边、删边	安全性、模块度	目标社团节点的边	检测单一、反欺骗问题
文献[23]	增边、删边	新的安全性	目标社团节点的边	检测单一、真实应用
文献[24]	增边、删边	模块度	目标社团节点的边	研究知识不足、有局限性
文献[25]	增边	模块度	节点中心性	网络规模小、关注点单一
文献[26]	增边、删边	模块度、遗传算法	目标社团节点的边	检测单一、单目标优化

的优化来进行社团隐藏操作。其中，文献[23]改进了文献[22]的社团安全性的定义，该优化方案的突出贡献是排除了文献[22]中有关在社团内部增加边和在社团之间删除边的操作对实验结果的影响，即假设欺骗算法不会新增目标社团内部的链接和不会删除目标社团与外部成员之间已存在的链路。由表2可知，现有的基于目标社团隐私保护研究工作的不足之处主要有：① 研究的关注点或关注目标较为单一；② 社交网络类型不足，只关注非重叠社团的网络；③ 反欺骗等真实应用问题还未深入研究等。

3.2 基于全社团的社团隐私保护

基于全社团的社团隐私保护技术主要探索有关社团的匿名化问题，现有的方法主要有基于子图 K -同构和 K^+ 同构的全社团隐藏算法^[27]以及基于社交网络分级隐私保护的全社团隐藏算法^[28]。

(1) 基于子图 K -同构和 K^+ 同构的全社团隐藏算法

荣欢^[27]对社交网络中社团 K -匿名进行了研究。文献[27]明确指出对社团隐藏问题的研究，可以着手于对社团隐私保护中有关子图 K -匿名的研究，而 K -同构算法是子图 K -匿名技术中最常用的方法，该方法分为初始化阶段、修改阶段以及结束阶段3个阶段。该算法在初始化阶段把原始网络划分为 K 个没有公共节点的不同社团（子图），记为 $\bar{C} = (C_1, C_2, \dots, C_k)$ ，且 $C_i \cap C_j = \emptyset (i \neq j)$ ；在修改阶段通过添加或者删除某些节点或者边，使在算法执行结束时这 K 个社团之间互相同构。该算法的流程如图2所示。

K -同构算法在对社团进行隐私保护的同时需要关注对原社交网络结构的修改程度。文献[27]

中引入图的编辑距离这一评价指标来描述匿名化处理后边的变化程度。假设社交网络 $G = \langle V, E \rangle$ ，经过 K -同构处理后的网络记为 $GP = \langle VP, VE \rangle$ ，定义经过匿名化处理后的编辑距离如式(17)所示。

$$GED(G, GP) = |\{E \cup EP\} - \{E \cap EP\}| \quad (17)$$

其中， $\{E \cup EP\}$ 表示 G 与 GP 所有的边的集合， $\{E \cap EP\}$ 表示 G 与 GP 公有的边的集合。

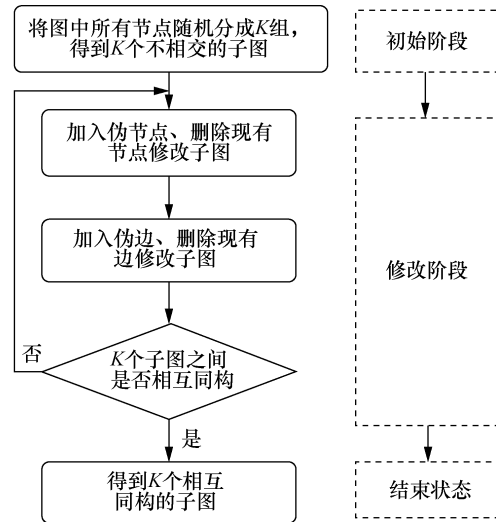


图2 K -同构算法的流程
Figure 2 The process of the K -isomorphic algorithm

K -同构算法所需要的匿名成本如式(18)所示。

$$\text{cost}(G, GP) = GED(G, GP) + \|VP\| - |V| \quad (18)$$

其中， $|VP|$ 和 $|V|$ 分别表示网络 G 和 GP 的节点个数。

文献[27]结合 K -同构算法，针对现有算法在社团检测划分存在的不足，提出节点角色类型改进策略，并以“指针-计数”机制为原型提出基于结构相关的社团检测方法通用处理模型 BSCHEF，使社团划分更加合理清晰。此外，文献[27]还提出基于最大共有子图的添加结果集的

验证环节的图相似性检测 (MPD_V) 算法, 该算法可以在层次性链式上进行搜索, 有效降低了计算的复杂度。

文献[27]综合运用 BSCHEF 和 MPD_V 算法, 对 K -同构算法的初始阶段进行了改进, 提出了新的社团匿名方法, 记为 K^+ 同构算法 (如图3所示)。该算法将社交网络划分为 m 个节点集合, 且每个节点集合至少包含 K 个具有相似性的社团, 然后对每一个节点集合中的所有社团进行添加/删除节点和边的操作, 使对于每一个节点集合中的任意一个社团都至少有 $K-1$ 个社团与它同构。实验结果表明, 当 K 取得合理的范围时, K^+ 同构算法所需的匿名化成本远低于 K -同构算法。但是, K^+ 同构算法不宜选择较大的参数 K , 否则会造成社团隐藏的匿名成本过大, 破坏社交网络的可用价值。

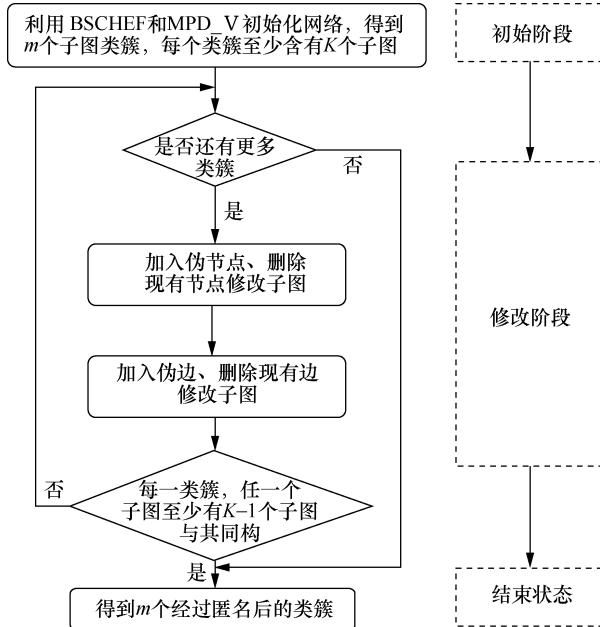


图3 K^+ 同构算法的流程

Figure 3 K^+ isomorphic algorithm model process

文献[27]还提出两个可行性研究方向: ① 研究在多角色节点和大量数据背景下的并行化方案; ② 研究 BSCHEF 和 MPD_V 算法在隐藏效果上抵抗攻击的能力以及能否对算法的性能进行提升等。

(2) 基于社交网络分级隐私保护的全社团隐藏算法

陆越^[28]针对社交网络真实数据公布环节中

的数据可用性与节点隐私信息保密性间的平衡性需求, 提出了融合社团划分的社交网络分级隐私保护算法, 该算法解决了社交网络的不同结构特征中节点均一化的隐私保护问题。文献[28]首先利用节点权重影响力排序提出了基于节点影响力因素的社团划分方法, 将节点集隐私保护等级分为 H (high)、M (middle) 和 L (low) 3 种等级。针对 H 等级成员集合的保护提出了基于敏感信息 (节点的度、节点标签、边的敏感分布等) 分布的 (k, l, p_1, p_2) 算法; 针对 M 等级的匿名化技术, 该文献引入 k -degree- l -diversity 匿名化方法来实现对节点的度以及敏感属性多样性的保护; 针对 L 等级的匿名化技术, 使用 k -degree 匿名化方法来保护节点的度信息。图4给出了上述的分级隐私保护技术示意。

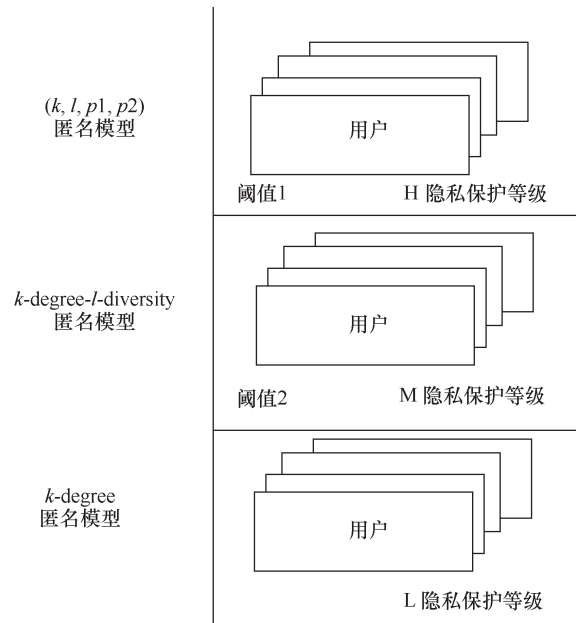


图4 分级隐私保护技术示意

Figure 4 Schematic diagram of Hierarchical privacy protection technology

此外, 文献[28]引入了文献[60]提出的匿名化算法评价标准评估算法的效用: 信息丢失率 (COSTA) 和平均路径长度 (APL), 具体的计算方法分别如式(19)和式(20)所示。

$$\text{COSTA} = \frac{\sum_{i=1}^N |D'_G(i) - D_G(i)|}{\sum_{i=1}^N |D_G(i)|} \quad (19)$$

$$APL = \frac{2}{(N-1)N} \sum_{i \neq j} d_{ij} \quad (20)$$

其中, $D'_G(i)$ 表示隐藏操作后网络图节点 i 的度, $D_G(i)$ 表示原始网络中节点 i 的度, d_{ij} 表示节点 i 和 j 之间的最短距离。

分级匿名模型在社团匿名化过程中的信息丢失率虽然有所降低, 但由图 4 可知, 该文献主要依据单一的影响力等级划分的计算结果, 隐私等级的划分条件较为薄弱, 算法的可扩展性较低, 未来需要结合数据的具体特征、隐私需求等因素细化等级, 划分方法。该文献的研究结果是基于攻击者的背景知识为节点的度, 尚未考虑背景知识为子图等情况, 还需进行深入研究。

表 3 详细归纳比较了上述基于全社团隐私保护思想的相关工作。文献[27]以增边或删边操作来达到 K 个社团相互同构的目的, 以此获得社团结构被扰动后的社交网络, 从而保护社团隐私; 文献[28]提出了一种通过划分社交网络社团隐私保护等级进行针对性社团保护的分级隐私保护方法, 该方案的不足之处在于划分隐私等级时的划分依据单一, 应该考虑加入更多的划分因素等。

表 3 基于全社团中隐私保护的相关工作对比
Table 3 Comparisons of related work based on privacy protection on the all communities

文献	操作	技术	不足之处
文献[27]	增边或删边	K -同构、 K^+ 同构	节点角色不足、算法优化
文献[28]	匿名化	分级隐私保护	等级划分依据单一、薄弱

4 结束语

由于社交网络社团隐私保护关系到用户信息泄露的隐患, 科研人员已经着手于在社交网络中保护社团内部信息免受社交网络社团分析工具的挖掘攻击。本文以社交网络的社团为切入点, 通过社区检测算法的弊端来引出研究内容: 社交网络社团隐私保护。本文针对社团结构安全问题综合整理了现有的关于社交网络社团隐私保护的文献, 对文献中所提出的核心思想、核心方法、评价指标、对照实验等进行了较详细的综合性论述。目前这一领域仍有广阔的研究前景, 其潜在的研

究方向有以下几个方面。

1) 跨社团隐私保护技术。真实的社交网络中, 对单个目标社团的隐私保护往往会使保护力度不够, 可能存在多社团或跨社团保护的需求。现有方法要么针对单目标社团防护, 要么针对所有社团进行防护, 缺乏多目标或跨目标社团防护技术。多目标社团防护技术不单单是多个单社团防护技术的组合, 因为在多社团或跨社团中, 扰动一条链路可能导致复杂多样的结果, 如有助于 A 社团隐藏, 但会加剧 B 社团的暴露, 使目标函数特性更加复杂, 增加计算过程的复杂度。因此, 高效的多社团或跨社团防护技术的研究将是重点与难点。

2) 社团推断 (community inference)。社团隐私保护技术本质上是社团检测的逆过程, 社团保护的目的是混淆社团结构, 达到欺骗社团检测的目的。那么, 挖掘隐藏的社团就成为新的问题, 具有重要的实际应用需求。因此, 社团推断的研究迫在眉睫, 其与社团检测不同, 社团推断需要补足缺失信息, 识别虚假链路, 进而采用先进的社团划分方法挖掘隐藏的社团。

3) 动态社交网络的社团隐私保护。当前的社团研究尤其社团隐私保护都聚焦在静态网络结构上, 但实际的社交网络都是动态演进的。演进过程中不断有新的链路出现, 旧的链路消失, 社团检测的结果在不断演进, 给社团隐藏的效果带来不确定性。因此, 如何设计新的社团保护机制或优化现有机制使其能够自适应动态社交网络, 为目标社团提供动态的防护有待解决。进一步地, 社团推断是否可以自适应动态网络进行精准推断仍然是未解之谜。

4) 多层网络或异质信息网络的社团隐私问题。现实生活中的不同网络之间具有互相依赖关系, 或呈现多层网络形态, 如社交网络的支撑网络是因特网。更进一步, 社交链接的定义或表征含义多样, 本质上是异质信息网络 (heterogenous information network), 那么针对复杂多样的网络形态进行社团隐藏需要更加复杂有效的解决办法, 有待广大科研工作者的努力奉献。

5) 基于博弈论的社团安全对抗研究。社团发现 (包括社团检测与社团推断) 与社团防护之间的对抗属于博弈论的范畴。因此, 基于博弈论的社团

安全对抗研究将成为一种必需的解决方案。

6) 社团隐私防护在实际系统中的应用问题。当前研究集中在已公开的社交网络数据集上进行, 离真正在实际社交网络应用还有距离。如何系统地评估社团隐私保护技术的影响、代价、用户与运营商的接受程度等, 并付诸应用是社团研究的最终目的。

参考文献:

- [1] Wikipedia. Complex network[EB].
- [2] XUAN Q, ZHANG Z Y, FU C, et al. social Synchrony on complex networks[J]. IEEE Transactions on Cybernetics, 2018, 48(5): 1420-1431.
- [3] XUAN Q, FANG H, FU C, et al. Temporal motifs reveal collaboration patterns in online task-oriented networks[J]. Physical Review E, 2015, 91(5): 052813.
- [4] GARCIA J O, ASHOURVAN A, MULDOON S, et al. Applications of community detection techniques to brain graphs: algorithmic considerations and implications for neural function[J]. Proceedings of the IEEE Institute of Electrical and Electronics Engineers, 2018, 106(5):1-22.
- [5] CHEN Z, WU J, XIA Y, et al. Robustness of interdependent power grids and communication networks: a complex network perspective[J]. IEEE Transactions on Circuits & Systems II Express Briefs, 2018, 65(1): 115-119.
- [6] SCHIAVO S, REYES J, FAGIOLO G. International trade and financial integration: a weighted network analysis[J]. Quantitative Finance, 2010, 10(4): 389-399.
- [7] FU C, ZHAO M, FAN L, et al. Link weight prediction using supervised learning methods and its application to yelp layered network[J]. IEEE Transactions on Knowledge & Data Engineering, 2018, 30(8): 1507-1518.
- [8] CLAUSET A, MOORE C, NEWMAN M E. Hierarchical structure and the prediction of missing links in networks[J]. Nature, 2008, 453(7191): 98-101.
- [9] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[C]//ICLR. 2016.
- [10] ZHANG H F, XU F, BAO Z K, et al. Reconstructing of networks with binary-state dynamics via generalized statistical inference[J]. Circuits & Systems I Regular Papers IEEE Transactions, 2019, 66(4): 1608-1619.
- [11] 王永刚, 蔡飞志, LUA E K, 等. 一种社交网络虚假信息传播控制方法[J]. 计算机研究与发展, 2012, 49(S2): 131-137.
WANG Y G, CAI F Z, LUA E K, et al. A diffusion control method of fake information in social networks[J]. Journal of Computer Research and Development, 2012, 49(S2): 131-137.
- [12] ZHENG J G, PAN L. Least cost rumor community blocking optimization in social networks[C]//2018 Third International Conference on Security of Smart Cities, Industrial Control System and Communications (SSIC). 2018: 1-5.
- [13] SPIRIN V, MIRNY L A. Protein complexes and functional modules in molecular networks[J]. Proceedings of the National Academy of Sciences, 2003, 100(21): 12123-12128.
- [14] GOH K I, CUSICK M E, VALLE D, et al. The human disease network[J]. Proceedings of the National Academy of Sciences, 2007, 104(21): 8685-8690.
- [15] NEWMAN M E. The Structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167-256.
- [16] GIRVAN M, NEWMAN M E. 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.
- [17] NEWMAN M E, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2): 026113.
- [18] PALLA G, DERENYI I, FARKAS I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. Nature, 2005, 435(23): 814-818.
- [19] 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): 036106.
- [20] RHEINGOLD H. The virtual community: homesteading on the electric frontier[M]. New York: Harper Perennial, 1993.
- [21] 姚瑞欣, 李晖, 曹进. 社交网络中的隐私保护研究综述[J]. 网络与信息安全学报, 2016, 2(4): 33-43.
YAO R X, LI H, CAO J. Overview of privacy preserving in social network[J]. Chinese Journal of Network and Information Security. 2016, 2(4): 33-43.
- [22] FIONDA V, PIRRO G. Community deception or: how to stop fearing community detection algorithms[J]. IEEE Transactions on Knowledge & Data Engineering, 2018, 30(4): 660-673.
- [23] CHEN X Y, JIANG Z Y, LI H, et al. Community hiding by link rewiring in social networks[J]. IEEE Transactions on Big Data, 2020.
- [24] WANIEK M, MICHALAK T, RAHWAN T, et al. Hiding individuals and communities in a social network[J]. Nature Human Behaviour, 2018, 2: 139-147.
- [25] NAGARAJA S. The impact of unlinkability on adversarial community detection: effects and countermeasures[C]//Privacy Enhancing Technologies, International Symposium. 2010: 253-272.
- [26] CHEN J Y, CHEN L H, CHEN Y X, et al. GA-based q-attack on community detection[J]. IEEE Transactions on Computational Social Systems. 2019, 6(3): 491-503.
- [27] 荣欢. 社会网络中社团 K-匿名研究[D]. 南京: 南京信息工程大学, 2016.
RONG H. The study of community K-anonymization in social network[D]. Nanjing: Nanjing University of Information Science & Technology, 2016.
- [28] 陆越. 基于社团划分的社交网络分级隐私保护算法研究[D]. 上海: 上海交通大学, 2018.
LU Y. Hierarchical privacy protection algorithm for social network based on community division[D]. Shanghai: Shanghai Jiao Tong University, 2018.
- [29] FORTUNATO S. Community detection in graphs[J]. Physics Reports, 2010, 486(3-5): 75-174.
- [30] 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.
YANG B, LIU D Y, JIN D, et al. Complex network clustering algorithms[J]. Journal of Software, 2009, 20(1): 54-66.
- [31] NEWMAN M E. Modularity and community structure in networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(23): 8577-8582.
- [32] NEWMAN M E. Fast algorithm for detecting community structure in networks[J]. Physical Review E, 2003, 69: 066133.
- [33] CLAUSET A, NEWMAN M E, MOORE C. Finding community structure in very large networks[J]. Physical Review E, 2004, 70: 066111.
- [34] LI Z P, ZHANG S H, WANG R S, et al. Quantitative function for community detection[J]. Physical Review E, Statistical Nonlinear & Soft

- Matter Physics, 2008, 77: 036109.
- [35] ROSVALL M, BERGSTROM C T. An information-theoretic framework for resolving community structure in complex networks[J]. Proceedings of the National Academy of Sciences, 2007, 104(18): 7327-7331.
- [36] 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.
- [37] 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(10): 10008.
- [38] SHI C, YAN Z, WANG Y, et al. A genetic algorithm for detection communities in larger-scale complex networks[J]. Advances in Complex Systems, 2010, 13(1): 3-17.
- [39] DONGEN S M. Graph clustering by flow simulation[D]/University of Utrecht, The Netherlands, (CWI). 2000: 238-247.
- [40] 尹欣红. 基于连接模式的相关分析及带偏置的随机游走的社团检测方法研究[D]. 兰州: 兰州大学, 2019.
- YIN X H. Research on correlation analysis based on link pattern and community detection method with random walk with bias[D]. Lanzhou: Lanzhou University, 2019.
- [41] YUAN Y L, SOH D, YANG H, et al. Learning overlapping community-based networks[J]. IEEE Transactions on Signal and Information Processing over Networks, 2019, 5(4): 684-697.
- [42] YANG J, LESKOVEC J. Overlapping community detection at scale: a nonnegative matrix factorization approach[C]/Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. 2013: 587-596.
- [43] AHN Y Y, HAN S, KWAK H, et al. Analysis of topological characteristics of huge online social networking services[C]/International Conference on World Wide Web. ACM, 2007: 835-844.
- [44] KIM Y, JEONG H. Map equation for link communities[J]. Physical Review E, 2011, 84(2): 026110.
- [45] YE Q, WU B, ZHAO Z, et al. Detecting link communities in massive networks[C]/International Conference on Advances in Social Networks Analysis & Mining. IEEE Computer Society, 2011: 71-78.
- [46] LANCICHINETTI A, FORTUNATO S, JANOS K. Detecting the overlapping and hierarchical community structure in complex networks[J]. New Journal of Physics, 2009, 11(3): 033015.
- [47] ANDREA L, FILIPPO R, RAMASCO J J, et al. Finding statistically significant communities in networks[J]. Plos One, 2011, 6(4): 18961.
- [48] PADROL S A, PERARNAU L G, PFEIFLE J, et al. Overlapping community search for social networks[C]/IEEE International Conference on Data Engineering. 2010: 992-995.
- [49] LEE C, REID F, MCDAID A, et al. Detecting highly overlapping community structure by greedy clique expansion[J]. arXiv: 1002.1827, 2010.
- [50] DU N, WANG B, WU B. Detecting highly overlapping community structure by greedy clique expansion[C]/Proceedings of the 17th ACM Conference on Information and Knowledge Management. 2008: 1371-1372.
- [51] ZHANG H, QIU B, GILES C L, et al. An LDA-based community structure discovery approach for large-scale social networks[C]/IEEE International Conference on Intelligence & Security Informatics. 2007: 200-207.
- [52] PSORAKIS I, ROBERTS S, EBDEN M, et al. Overlapping community detection using Bayesian non-negative matrix factorization[J]. Physical Review E, 2011, 83(6): 066114.
- [53] NEPUSZ T, PETROCZI A, NEGYESSY L, et al. Fuzzy communities and the concept of bridgeness in complex networks[J]. Physical Review E, 2008, 77(1): 016107.
- [54] NEWMAN M E, LEICHT E A. Mixture models and exploratory analysis in networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(23): 9564-9569.
- [55] AIROLDI E M, BLEI D M, FIENBERG S E, et al. Mixed membership stochastic blockmodels[J]. Journal of Machine Learning Research, 2008, 9(3): 33-40.
- [56] ABRAHAO B, SOUNDARAJAN S, HOPCROFT J, et al. A separability framework for analyzing community structure[J]. ACM Transactions on Knowledge Discovery from Data, 2014, 8(1): 101-129.
- [57] JIANG Z Y, LI J, MA J F. Similarity-based and sybil attack defended community detection for social networks[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2020.
- [58] 李欢, 莫欣岳. 复杂网络重叠社团检测算法研究综述[J]. 传感器与微系统, 2017, 36(1): 1-4.
- LI H, MO X Y. Research review on detection algorithm of overlapping community in complex network[J]. Transducer and Microsystem Technologies, 2017, 36(1): 1-4.
- [59] 于乐. 社交网络中社团发现及网络演化分析[D]. 北京: 北京邮电大学, 2014.
- YU L. Research on community detection and evolution analysis in social network[D]. Beijing: Beijing University of Posts and Telecommunications, 2014.
- [60] LU K, TERZI E. Towards identity anonymization on graphs[C]/Proceedings of the 2008 ACM SIGMOD, International Conference on Management of Data ACM. 2008: 93-106.

[作者简介]



蒋忠元 (1988—), 男, 陕西榆林人, 博士, 西安电子科技大学副教授, 主要研究方向为大数据安全、复杂网络安全研究、复杂网络传输动力学、网络功能虚拟化、城市计算。



陈贤宇 (1996—), 男, 安徽六安人, 西安电子科技大学硕士生, 主要研究方向为社交网络隐私与保护、社团检测。



马建峰 (1963—), 男, 陕西西安人, 博士, 西安电子科技大学教授、博士生导师, 主要研究方向为应用密码学、无线网络安全、数据安全、移动智能系统安全。