DOI:10.7544/issn1000-1239.20200779
59(4):796-812,2022

基于抖音共同联系人的群体用户关系分析

乐洪舟[1,2] 何水龙[1] 王 敬[1]
[1](信阳师范学院计算机与信息技术学院 河南信阳 464000)
[2](河南省教育大数据分析与应用重点实验室(信阳师范学院) 河南信阳 464000)
(yuehz@xynu.edu.cn)

Analysis of Group Users' Relationship Based on TikTok Mutual Contacts

Yue Hongzhou[1,2], He Shuilong[1], and Wang Jing[1]
[1](*School of Computer and Information Technology, Xinyang Normal University, Xinyang, Henan* 464000)
[2](*Henan Key Laboratory of Analysis and Applications of Education Big Data (Xinyang Normal University), Xinyang, Henan* 464000)

Abstract Many popular social apps have the function of showing mutual relationship between users. However, the exposure of mutual relationship may lead to the occurrence of user privacy security problems. Taking China's most famous short video software TikTok as the research object, a privacy disclosure security vulnerability in the mutual contacts function of TikTok is analyzed. A method of vulnerability exploiting and attacking for group users is proposed. The attack effect is that even if some users are not allowed to find themselves through their mobile phone numbers by some settings, an attacker can still use the known mobile phone numbers of group users and the internal connections among group users to get these users' TikTok accounts. After getting as many TikTok accounts of the group users as possible, attackers can collect the following, contacts, video likes and comments information among group users, and use this information to calculate users' relationship, which can provide some assistance for launching further effective attacks. Two indexes—intimacy and group-activeness—are proposed to describe users' relationship, and the calculation method of these two indexes is given. Through the experiment of three real groups in society, the effectiveness of user relationship calculation is verified. In the end, the security threats to users are analyzed and the security prevention suggestions are given.

Key words TikTok; mutual relationship; privacy disclosure; security vulnerability; user relationship

摘 要 很多流行的社交App都有展示用户之间的共同关系的功能,然而,共同关系的暴露也可能导致用户隐私安全问题的发生.以中国最知名的短视频软件抖音为研究对象,分析了其共同联系人功能存在的用户隐私泄露的安全漏洞.提出了一种针对群体用户的漏洞利用和攻击方式,该攻击方式可以达到的效果是,即使群体中某些用户设置了不允许通过手机号找到自己,攻击者仍然可以利用已知的群体用户的手机号码和群体用户之间的内在联系获得这些用户的抖音账号.攻击者在获得群体中尽可能多的用户的抖音账号后,可以对这些用户相互之间的关注信息、通信录信息、视频点赞和评论信息进行收集,

收稿日期:2020-09-23;**修回日期**:2021-04-21
基金项目:国家自然科学基金项目(31900710);河南省自然科学基金项目(212300410236)
This work was supported by the National Natural Science Foundation of China (31900710) and the Natural Science Foundation of Henan Province(212300410236).

并利用这些信息计算群体用户之间的关系,为发起进一步的有效攻击提供一定的辅助.提出了描述用户关系的2个指标——亲密度和群体活跃度,并给出了这2个指标的计算方法.通过对现实社会中3个真实群体的实验,验证了用户关系计算的有效性,分析了对用户所造成的安全威胁,并给出了安全防范建议.

关键词 抖音;共同关系;隐私泄露;安全漏洞;用户关系

中图法分类号 TP309

随着互联网的飞速发展,网络社交成为一种新的时尚,一些流行的社交App已经成为人们生活中不可缺少的一部分,为人们提供了信息交流和娱乐消遣的重要平台.当前很多流行的社交App都有查找共同关系的功能,2个用户可以查看彼此之间有共同关系的人,包括共同联系人、共同好友等.表1列出了6个具有共同关系功能的社交App.从表1中可以看出,除QQ外,其他App的用户查看共同关系都需要满足一定的前提,例如:微信用户在双方互为好友的情况下才能查看共同好友,查看的位置是微信朋友圈,朋友圈中只有共同的好友的评论和点赞才能被双方看到.除QQ和微信以外的其他App只需要满足一定的隐私设置即可查看共同关系.这些共同关系主要包括共同关系的数量和共同关系的用户名单2方面的内容,不同的App显示共同关系的内容不一样,除QQ仅能显示共同好友的数量以外,其他几个App既能显示共同关系的数量,也能显示共同关系的用户名单.我们也发现,这些App为用户提供查看共同关系的功能,主要目的是为了向用户推荐潜在的好友(如QQ、抖音、Facebook、LinkedIn),然而,也有App是为了帮助用户回避现实中熟悉的人,避免遇到熟人所造成的尴尬,例如探探这种陌生人交友软件.

查找共同关系的功能在为社交媒体用户提供便利的同时,也带来了很多用户隐私安全问题.例如:一些研究者发现,已知社交网络中用户节点间共同关系的数量,可以实施友谊攻击,挖掘和分析用户的身份数据,对用户身份进行识别[1-2].甚至有研究者提出,即使某些社交媒体App运用了一些匿名化方法来防御共同关系功能造成的隐私泄露问题,攻击者也能采取一定的攻击手法,绕过匿名化技术的安全限制,暴露匿名用户的身份,推断目标用户的社交关系[3-4].一些社交App为用户提供了一些安全设置,使用户能够自主控制与共同关系有关的隐私信息,但一些研究者也提出一些攻击方法,能够使用户的隐私设置失效[5].其他的与共同关系有关的安全问题还包括邻域攻击问题[6-8]、共谋攻击问题[9]等.

现有的关于共同关系的安全问题的相关研究基本上没有将现实中真实的社会群体作为研究对象,研究他们在社交媒体中所暴露的各成员之间亲密关系的隐私泄露问题.本文将以中国最知名的短视频软件抖音为例,研究抖音App的共同联系人功能泄露现实中真实群体用户的账号信息和群体用户之间的亲密关系的安全问题.从表1可以看出,抖音的共同联系人查看功能可以针对任意用户,无需双方互为好友,这为攻击者进行信息收集提供了便利条件.我们发现了抖音共同联系人功能存在的一个安全漏洞,在目标用户已经进行了个人信息屏蔽设置(关闭"允许将我推荐给好友"选项)的情况下,攻击者仍然可以利用共同联系人功能发现目标用户的账号.

Table 1 Apps That Can Show Mutual Relationship

表1 显示共同关系的App

应用	显示位置	能查看的内容	查看前提	谁可以查看
微信	朋友圈评论	共同好友数量和名单	加对方好友	微信好友
QQ	好友推荐	共同好友数量	无	任何人
抖音	用户推荐	共同联系人数量和名单	使用了"查看通信录好友"功能	任何人
探探	匹配到某一用户的结果页面	共同联系人数量和名单	开启"显示共同手机联系人"选项	任何人
Facebook	用户主页	共同好友数量和名单	"谁可以看你的好友列表"选项设置为"仅自己"以外的选项	根据设置而定
LinkedIn	用户主页	共同关系的数量和名单	"谁可以看你的关系"选项设置为"我的关系"	关系圈的人

我们结合这个安全漏洞,在已知抖音群体用户的电话号码信息后,能够获得群体中尽可能多的用户的抖音账号信息,并通过群体用户相互之间的关注信息、通信录信息、视频点赞和评论信息来推算群体用户之间的亲密关系信息(亲密度和群体活跃度),这些亲密关系信息能够为攻击者实施进一步有效攻击提供一定的辅助.我们通过对现实中一个真实群体的实验,证明了通过抖音共同联系人功能发掘群体用户抖音账号的有效性,并验证了亲密关系信息的计算结果的合理性.此外,我们也进行了安全威胁分析,并提出了安全防范建议.

本文主要创新和贡献包括 3 个方面:

1) 发现了中国最知名的短视频软件抖音的共同联系人功能存在的安全漏洞,可以使用户的相关个人隐私设置失效,使攻击者在已知群体用户的电话号码信息后,利用群体用户之间的通信录关系,获得群体中尽可能多的用户的抖音账号信息.

2) 提出了结合抖音共同联系人安全漏洞,并通过群体用户相互之间的关注信息、通信录信息、视频点赞和评论信息挖掘群体用户之间的亲密度和用户群体活跃度的方法.该方法除了考虑到用户之间关注关系、通信录关系和点赞关系的连接数量,而且也把用户之间的评论的情感考虑在内,利用基于称谓词词典和情感词典的方法计算评论的情感值,从而推算用户之间的亲密度.

3) 提出攻击者针对群体用户的亲密度和群体活跃度的计算结果能够为攻击者发起针对群体用户的进一步有效攻击提供一定的辅助,并分析了对群体用户可能造成的安全威胁,给出了相应的安全防范建议.

1 抖音相关背景

抖音是目前中国最知名、用户量最大的短视频软件,用户量超 5 亿[10].用户可以通过抖音制作和发布短视频对外展示自己,也可以观看其他用户发布的视频.抖音从 2016 年 9 月上线以来就迅速积累了大量的人气,为各种年龄段用户提供了娱乐和社交的平台,赢得了广大用户的欢迎.抖音鼓励用户查看通信录好友,在启动 App 的时候甚至会弹窗提醒用户查看通信录好友,而一旦用户使用了这个功能,则抖音 App 就会获取用户手机通信录信息,由此也导致了一些隐私泄露的安全问题和法律纠纷[11].

图 1 展示了某一抖音用户的个人主页,访问者可以点击位置①查看该用户关注的用户列表,点击位置②查看该用户发布的视频列表,点击位置③可以查看该用户点赞了哪些视频.

Fig. 1 TikTok user personal homepage

图 1 抖音用户个人主页示例

然而,抖音(以 2020 年发布的 11.6.0 版本为例)中有表 2 所示的设置选项,可以使用户有权决定是否公开某些信息.如果某用户对这些选项不进行设置,那么这 4 个选项都会是默认设置,即其他用户可以看到该用户的关注用户列表、粉丝列表、点赞视频列表,并且可以通过“查询通信录好友”的功能找到他.

如果用户进行了设置,例如,当把“谁可以看我对作品的点赞”选项设置为“仅自己和作者”时,仅视频作者和用户自己可以看到用户对某视频的点赞行为.当把“谁可以看我的关注、粉丝列表”选项设置为“仅自己”时,只有用户自己可以看他的关注和粉丝列表.当把“允许将我推荐给好友”选项设置为关闭时,其他用户将不能通过“查找通信录好友”功能找到他.当“私密账号”选项设置为“开”时,其他用户将不能看该用户的关注用户列表、粉丝列表、点赞视频列表等信息.

经过研究发现,大多数用户不会专门对这些选项进行设置,这便给攻击者收集用户的相关信息提供了便利.

Table 2 Options of TikTok User Personal Setting

表 2 抖音用户个人设置选项

设置	默认选项	其他选项	备注
私密账号	关	开	账号设为私密时,只有你批准的用户才能关注你,并能看到你的内容和喜欢
允许将我推荐给好友	开	关	关闭后,你不会在“好友推荐”中被推荐给可能认识的人,其他用户也不能通过他上传的通信录找到你
谁可以看我对作品的点赞	所有人	仅自己和作者	设置为“仅自己和作者”时,其他用户将无法从你的个人主页看到你赞过的作品,且不会在推荐中看到你的点赞信息
谁可以看我的关注、粉丝列表	所有人	仅自己	设置为“仅自己”时,其他用户将无法查看你的关注、粉丝列表

2 共同联系人泄露安全漏洞

抖音提供了“查找通信录好友”的功能,当某用户向服务器发送“查找通信录好友”请求后,服务器会返回该用户的通信录好友对应的抖音用户列表,然而,如第 1 节中所提到的那样,若用户设置“允许将我推荐给好友”选项为关闭时,其他用户将不能通过“查找通信录好友”功能找到他.

然而,抖音中有个共同联系人功能能够让设置了“允许将我推荐给好友”选项为关闭的用户账号(抖音 ID)也通过电话号码被查询出来.抖音中的共同联系人功能是为了向用户推荐存在共同联系人关系的用户,增加用户之间的关联.然而,如果用户 a 和用户 b 存在共同联系人,那么抖音在向 a,b 进行相互推荐的时候,不但可以向 a,b 展示共同联系人的数量,而且可以展示共同联系人的姓名和抖音 ID,如图 2 所示,而不管共同联系人是否设置“允许将我推荐给好友”选项为关闭.除了可以获取抖音的推荐用户与攻击者的共同好友信息以外,抖音的共同联系人功能的相关接口甚至允许攻击者查询任一用户与攻击者的通信录的共同联系人,这是一个明显的安全漏洞.

共同联系人功能存在的这个安全漏洞给攻击者提供了挖掘群体用户隐私的有利条件.由于一个社会群体中的用户一般相互间会保留对方的联系方式,因此如果一个社会群体(如学生班级群体、单位职工群体)的电话号码信息被攻击者窃取,那么攻击者将可能获得群体用户中尽可能多的抖音 ID,建立群体用户电话号码和抖音 ID 的对应关系.方法是:攻击者首先把群体用户的手机号码存储在通信录,并通过“查询通信录好友”功能查询通信录好友的抖音 ID,如果只能查找到部分通信录好友的抖音 ID,那么攻击者可以通过“共同联系人”的功能接口向服务器发送请求,获取这些抖音 ID 对应的每一位用户与攻击者的共同联系人的 ID,新获得的抖音 ID 还可以继续被拿来查询共同联系人,如此迭代往复,就可以把群体里面尽可能多的用户对应的抖音 ID 获取到.

图 3 通过实例展示了群体用户抖音 ID 迭代查询的过程,图 3 中假设群体用户编号为 1~10,他们的电话号码都被攻击者存放到通信录,攻击者通过查看通信录好友功能只能查到编号为 1~5 的抖音用户.然而,由于用户 2 的通信录中存储了用户 7,8,则攻击者通过查询与用户 2 的共同好友就可以查到用户 7,8 的抖音 ID.在得到了新的用户 7,8 的 ID 后,攻击者继续查询与用户 7,8 的共同好友关系.由于

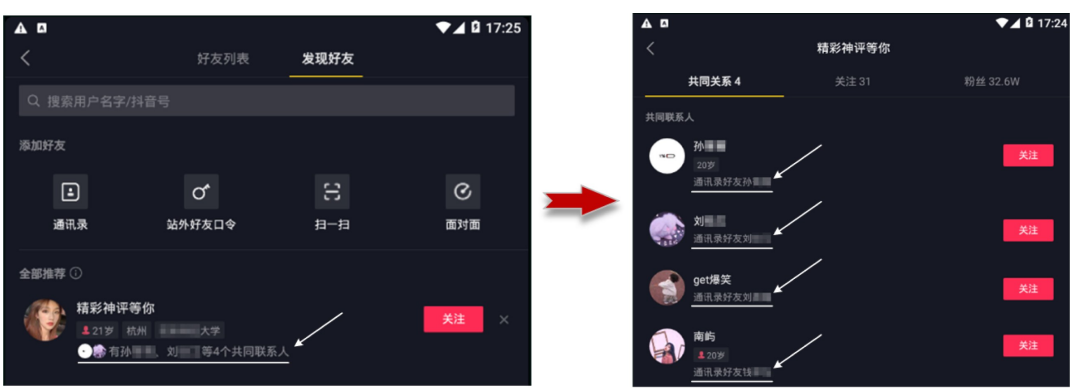


Fig. 2 The mutual contacts function of TikTok

图 2 抖音共同联系人功能

用户 8 的通信录中存储了用户 6,9,则攻击者通过查询与用户 8 的共同好友就可以查到用户 6,9 的抖音 ID.因此最终攻击者可以获取到群体中用户 1～9 的抖音 ID,仅仅只有用户 10 是查不到抖音 ID 的.

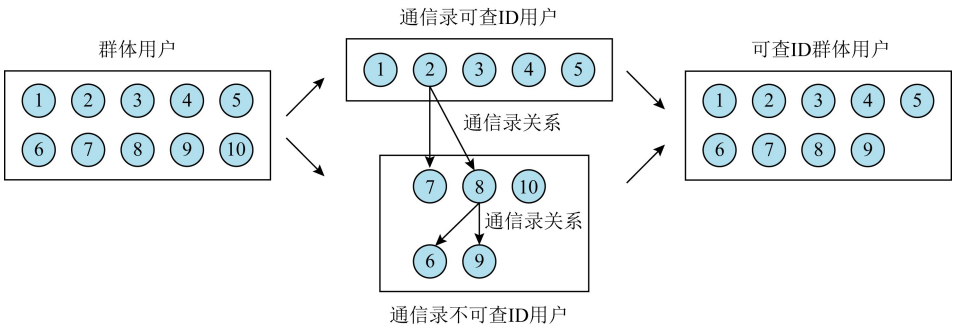


Fig. 3 Iterative query of group users' TikTok IDs based on mutual contacts function

图 3 基于共同联系人功能的群体用户抖音账号迭代查询

3 攻击者视角的用户关系分析

群体用户既然是一个社会群体,那么他们之间肯定有密切的关系,如果一个攻击者既能获得群体用户的抖音账号信息,又能利用抖音账号信息推算出这个群体中的用户之间的亲密关系,知道哪些用户在群体中较为活跃、人缘好等信息,那么对他实施进一步攻击会带来很大的辅助作用.例如,攻击者伪装为某用户身边的朋友对该用户进行诈骗,或者重点对群体活跃性较高的用户进行攻击,再借助于他们影响周围的人等.

在第 1 节中已经介绍了攻击者可以从抖音用户主页获得用户的关注用户列表、粉丝列表、点赞视频列表等信息.在第 2 节中也讨论了攻击者可以利用查找通信录好友和共同好友的功能获得群体中尽可能多的用户的抖音 ID.攻击者可以通过分析这些信息来挖掘用户之间的关系.本文在这里讨论 2 种用户关系分析:亲密度分析和群体活跃度分析.

3.1 亲密度分析

亲密度分析是讨论群体中任意 2 个用户之间的亲密关系.可以这样假设:如果一个用户 a 关注了另一个用户 b,或在 a 的通信录中保存了用户 b 的联系电话,或者 a 对 b 发布的视频点赞或评论过,那么二者之间可能存在亲密关系.而如果攻击者能够对群体用户相互之间的关注、通信录、点赞和评论关系进行合理分析,则可能大致推算出群体用户相互之间的亲密关系.

本文首先根据用户之间的关注、通信录、点赞、评论 4 种关系计算 4 种亲密分量,每种亲密分量都被分为单向亲密分量和双向亲密分量,然后将这 4 种亲密分量加权求和,计算用户之间的亲密度.

3.1.1 关注关系亲密分量计算

关注关系单向亲密分量:用户 a 对用户 b 的关注关系单向亲密分量表示为 QF_a^b,如果用户 a 关注了 b,则 QF_a^b 的值为 1.如果 a 没有关注用户 b,则 QF_a^b 的值为 0.

关注关系双向亲密分量:用户 a 与用户 b 的关注关系双向亲密分量表示为 QF_{ab},QF_{ab} 的值等于双方单向亲密分量的加和,即 $QF_{ab}=QF_a^b+QF_b^a$.

3.1.2 通信录关系亲密分量计算

通信录关系单向亲密分量:用户 a 对用户 b 的通信录关系单向亲密分量表示为 QC_a^b,如果 b 在 a 的通信录中,则 QC_a^b 的值为 1.如果 b 不在 a 的通信录中,则 QC_a^b 的值为 0.

通信录关系双向亲密分量:用户 a 与用户 b 的通信录关系双向亲密分量表示为 QC_{ab},QC_{ab} 的值等于双方单向亲密分量的加和,即 $QC_{ab}=QC_a^b+QC_b^a$.

3.1.3 点赞关系亲密分量计算

点赞关系单向亲密分量:用户 a 对用户 b 的点赞关系单向亲密分量表示为 QL_a^b,如果 a 对 b 的所有视频的点赞数为 m_a^b,a 给所有群体用户的视频点赞总数为 n_a,b 的所有视频得到群体中的用户点赞总数为 n^b,求 QL_a^b 的值:

$$QL_a^b=\frac{1}{2}\left(\frac{m_a^b}{n_a}+\frac{m_a^b}{n^b}\right),\ n_a>0,\ n^b>0. \tag{1}$$

如果 a 没有给群体中的任何用户点赞($n_a=0$),则 QL_a^b 的值为 0.如果 b 没有得到群体中任何用户的点赞($n^b=0$),则群体中任意用户 x 对 b 的点赞关系亲密分量 QL_x^b 的值为 0.

点赞关系双向亲密分量:用户 a 与 b 的点赞关系双向亲密分量表示为 QL_{ab},QL_{ab} 的值等于双方单向亲密分量的加和,即 $QL_{ab}=QL_a^b+QL_b^a$.

3.1.4 评论关系亲密分量计算

本文通过计算评论内容的情感值来计算用户之间评论关系的亲密分量,评论内容的情感值包括称谓的情感值和语义的情感值 2 个方面.这样设计的原因是基于 2 点:1)如果用户 a 对用户 b 的评论中有比较亲密的称谓,则说明 a 与 b 的亲密度较高;2)如果 a 对 b 的正面的评论多,负面的评论少,则也能反映 a 对 b 的亲密程度高.

本文首先构建称谓语词典和情感词典,然后利用这 2 个词典对评论的情感值进行计算:

1) 称谓语词典构建

本文以吉常宏编著的《汉语称谓大词典》[12] 为基准,结合刘静敏[13]、吴超[14]、徐爽[15] 关于汉语称谓语的研究成果,整理出一个称谓词词典.根据称谓词的亲密程度,将称谓词分为 6 个等级,亲密度最高级别 R1 为具有亲属关系的称谓,不同的亲密级别所赋予的亲密分值不同,表 3 展示了称谓语词典中各亲密级别的部分称谓词信息.

Table 3 Examples of Appellation Words Dictionary

表 3 称谓词词典示例

级别	称谓	分值	备注
R1	爸,妈,哥哥,妹妹,老公,老婆	3	具有亲属关系的称谓
R2	亲爱的,宝贝,宝宝,乖乖,honey,baby	2.5	恋人、亲密朋友之间的称谓
R3	小淘气,小可爱,小笨蛋,小傻瓜,小猪,小猫,小妮	2	
R4	小 X,X 儿,X,阿 X,老 X	1.5	X 为姓名中的一个字
R5	X 哥,X 姐,X 弟,X 妹,兄弟	1.2	
R6	老弟,哥们,老铁,美女,帅哥	0.5	其他亲密称谓

2) 情感词典构建

本文通过基于情感词典的方法来计算评论的情感值.通过对台湾大学中文情感极性词典(NTUSD)[16]、知网 HowNet 词典[17]、大连理工大学中文情感词汇本体库(DUTIR)[16]、清华大学李军中文褒贬义词典[18] 和 BosonNLP 情感词典[19] 进行整合和筛选,构建了包含情感词、程度副词和否定词的情感词典,所构建的情感词典的部分信息如表 4 所示.此外,本文还采用了结巴词典[20] 对句子进行分词.

Table 4 Examples of Sentiment Words Dictionary

表 4 情感词词典示例

词性	类别	词数	示例	分/权值
情感词	正面情感词	6 094	高兴,喜欢,漂亮,好看,灿烂	1
	负面情感词	11 445	伤心,愤怒,讨厌,惭愧,暗	−1
程度副词	over	31	超,过度,过分,过甚,忒	2
	most	65	充分,极其,极为,十分,非常	1.75
	very	42	大为,多么,分外,很,实在	1.5
	more	37	更,较为,那么,愈发,真	1.25
	ish	29	略微,稍微,挺,有点儿,有些	0.75
	insufficiently	14	半点,不大,不甚,不怎么,微	0.5
否定词		58	不,没,无,非,勿	−1

情感词词典中包含 6 094 个正面情感词和 11 445 个负面情感词,将每个正面情感词的分值设为 1,负面情感词的分值设为 −1.程度副词用于对情感词进行修饰,增强或减弱情感.本文将程度副词划分为 6 个类别:over,most,very,more,ish,insufficiently,权值分别为 2,1.75,1.5,1.25,0.75,0.5.同时选取了 58 个否定词,每个否定词的权值设为 −1,否定词作用在情感词前面,可能导致情感发生变化(奇数个否定词),也可能不改变情感(偶数个否定词).

3) 评论的情感值计算

对于某一条评论句 S,求其情感值的步骤:

首先,以标点符号和情感词为基准,将评论 S 分为分句 $S_1,S_2,\cdots,S_n$,每个分句最多只包含一个情感词.

然后,对于每一个分句 S_i,求分句的情感值.具体步骤有 5 步.

① 用结巴词典对分句进行分词.

② 从分词中寻找称谓词,并通过称谓词词典获取称谓词的情感值 q_1,若没有称谓词,则 q_1 为 0.

③ 查找分句的情感词,并根据情感词典判断其为正面还是负面的情感,获取情感词的情感分值 q_2,q_2 的值可能为 1 或 −1(正面或负面的情感).

④ 寻找修饰情感词的程度副词,记录程度副词的数量 m,并根据情感词典确定各程度副词的权值 $\omega_1,\omega_2,\cdots,\omega_m$.

⑤ 向情感词前方寻找否定词,统计否定词出现的次数,记否定词的个数为 n.

于是,对于由多个程度副词修饰的情感词,采取

求程度副词的几何平均数的方法确定程度副词的权值,分句 S_i 的情感值 PS_i 为

$$PS_i = q_1 + (-1)^n \times \left(\prod_{j=1}^{m} \omega_j\right)^{\frac{1}{m}} \times q_2. \quad (2)$$

最后,求整条评论句 S 的情感值 PS,一条评论句的情感值等于各个分句的情感值之和:

$$PS = \sum_{i=1}^{n} PS_i. \quad (3)$$

举例说明:如果某条评论的内容为"玲儿今天穿的真漂亮,视频拍得很好看,就是光线有点儿暗.",其中,"玲"为姓名中的一个字,则这条评论的情感值计算过程为

首先,将评论分为 3 个分句:"玲儿今天穿的真漂亮"(S_1),"视频拍得很好看"(S_2),"就是光线有点儿暗"(S_3).

然后,计算各分句的情感值,S_1 中包含正面形容词"漂亮"(分值 1)和程度副词"真"(权重 1.25),并且还包含有亲密称谓"玲儿"(分值 1.5);S_2 中包含正面形容词"好看"(权重 1)和程度副词"很"(权重 1.5);S_3 中包含负面形容词"暗"(权重 −1)和程度副词"有点儿"(权重 0.75).

于是,各分句的情感值为

$$PS_1 = 1.5 + 1 \times 1.25 = 2.75;$$
$$PS_2 = 1 \times 1.5 = 1.5;$$
$$PS_3 = -1 \times 0.75 = -0.75.$$

最后,得出评论的情感值 $PS = \sum_{i=1}^{3} PS_i = 3.5$.

4) 评论关系的亲密分量计算

得到每条评论的情感值后,采取如下方法计算评论关系的亲密分量:

① 评论关系单向亲密分量.用户 a 对用户 b 的评论关系单向亲密分量表示为 QT_a^b,QT_a^b 的值为 a 对 b 的所有评论的情感值中的最大值.

② 评论关系双向亲密分量.用户 a 与用户 b 的评论关系双向亲密分量表示为 QT_{ab},QT_{ab} 的值等于双方单向亲密分量的加和,即 $QT_{ab} = QT_a^b + QT_b^a$.

3.1.5 亲密度计算

计算完 4 种亲密度分量后,将它们进行综合,求出每 2 个用户之间的亲密度.本文分别设置用户之间的关注、通信录、点赞和评论关系的亲密分量的系数为 $\alpha,\beta,\gamma,\delta$,于是,用户 a 对用户 b 的单向亲密度 Q_a^b 和双向亲密度 Q_{ab} 分别计算为:

$$Q_a^b = \alpha \times QF_a^b + \beta \times QC_a^b + \gamma \times QL_a^b + \delta \times QT_a^b, \quad (4)$$

$$Q_{ab} = \alpha \times QF_{ab} + \beta \times QC_{ab} + \gamma \times QL_{ab} + \delta \times QT_{ab}. \quad (5)$$

3.2 群体活跃度分析

攻击者在获得了群体各用户之间的亲密度后,也可以通过分析这些亲密度信息大致分析出某一用户在群体中是否活跃,在实施攻击时可以重点针对一些群体活跃性较高的用户.例如:重点对这一部分人灌输一些有害的信息(欺骗性质的商业宣传、不健康行为、恶意软件的安装与使用等),再通过他们影响周围的人,借助于这些群体活跃用户在群体中的影响力,可能取得更好的有害信息传播效果.

本文将群体活跃度分为被动活跃度和主动活跃度 2 种情况来讨论.

1) 被动活跃度分析

某一用户的被动活跃度是指群体中其他用户对该用户的单向亲密度的汇总,反映了群体中其他用户对该用户的亲密程度(或者说是受欢迎程度).例如,如果某个用户很少被群体中的其他人放在通信录,或者即使发布了很多抖音视频,也很少有人关注他,很少有人给他点赞和评论,则可以近似认为该用户的被动活跃度低,受欢迎程度低.相反,如果一个用户被群体中的很多人放在通信录,或者即使发布了很少的抖音视频,但却有很多人关注他,很多人给他点赞和评论,则可以近似认为该用户的被动活跃度高,受欢迎程度高.

对于用户 a 来说,如果群体中所有关注 a 的用户为 $\{x_1, x_2, \cdots, x_i\}$,群体中把 a 放在通信录的用户为 $\{y_1, y_2, \cdots, y_j\}$,群体中对 a 发布的视频点赞过的用户为 $\{z_1, z_2, \cdots, z_m\}$,群体中对 a 发布的视频评论过的用户为 $\{u_1, u_2, \cdots, u_n\}$,$a$ 发布的视频数为 v_a,则 a 的被动活跃度 H^a 可以计算为

$$H^a = \beta \times \sum_{k=1}^{j} QC_{y_k}^a + \left(\alpha \times \sum_{k=1}^{i} QF_{x_k}^a + \gamma \times \sum_{k=1}^{m} QL_{z_k}^a + \delta \times \sum_{k=1}^{n} QT_{u_k}^a\right) \times \frac{1}{\ln(v_a + 1)}. \quad (6)$$

计算被动活跃度的时候,需要考虑到一个用户发布的视频数(也可以认为是抖音的使用程度)对群体活跃度的影响,有的用户虽然视频数很多,但其他用户对其亲密度却很低,那他很可能在群体中不活跃,而有的用户虽然视频数很少,但其他用户对其亲密度却很高,那么他可能在群体中很活跃.因此,我们将用户关注、点赞和评论关系的亲密度之和除以系数 ln(视频数+2),加 2 是为了避免视频数为 0 的情况.

2) 主动活跃度分析

某一用户的主动活跃度是指该用户对群体中的

其他用户的单向亲密度的汇总,反映了用户对群体中其他用户的亲密程度(或者说是热情程度).如果某个用户的通信录中很少有群体中的其他人,或者很少在抖音上关注其他人,很少给其他人的视频发表评论,则我们可以近似认为该用户的主动活跃度低,对他人的热情程度低.相反,如果一个用户把群体中的很多人放在通信录,或者在抖音上关注了很多人,给很多人的视频发表了评论,则我们可以近似认为该用户的主动活跃度高,对他人的热情程度高.

对于用户 a 来说,如果 a 关注的该群体中的所有用户为$\{x_1,x_2,\cdots,x_i\}$,a 的通信录中包含的该群体中的用户为$\{y_1,y_2,\cdots,y_j\}$,该群体中被 a 点赞过的用户为$\{z_1,z_2,\cdots,z_m\}$,该群体中被 a 评论过的用户为$\{u_1,u_2,\cdots,u_n\}$,则 a 的主动活跃度 H_a 为

$$H_a=\alpha\times\sum_{k=1}^{i}QF^{a}_{x_k}+\beta\times\sum_{k=1}^{j}QC^{a}_{y_k}+\gamma\times\sum_{k=1}^{m}QL^{a}_{z_k}+\delta\times\sum_{k=1}^{n}QT^{a}_{u_k}. \tag{7}$$

无论是被动活跃度还是主动活跃度,都可以反映一个用户在群体中是否活跃、人缘好不好,群体活跃度分析的结果可以为攻击者实施进一步有效攻击提供辅助.

4 数据采集方法

攻击者要想实现第 3 节中介绍的亲密度分析和群体活跃度分析,必须对群体用户的关注、通信录、点赞和评论信息进行收集.在第 1 节中已经介绍过,攻击者可以通过收集用户主页上的信息获知一个用户关注了哪些其他用户,点赞过哪些视频,该用户发布的视频被哪些人评论过,可以通过查找通信录好友和共同联系人的功能获得尽可能多的群体用户的抖音 ID.攻击者可以将这些收集到的群体用户的信息,汇总为 5 张数据表:群体用户身份信息数据表、群体用户关注信息数据表、群体用户通信录信息数据表、群体用户点赞信息数据表和群体用户评论信息数据表.这 5 张表的建立过程如图 4 所示(图 4 中的加粗菱形即表示为数据表):

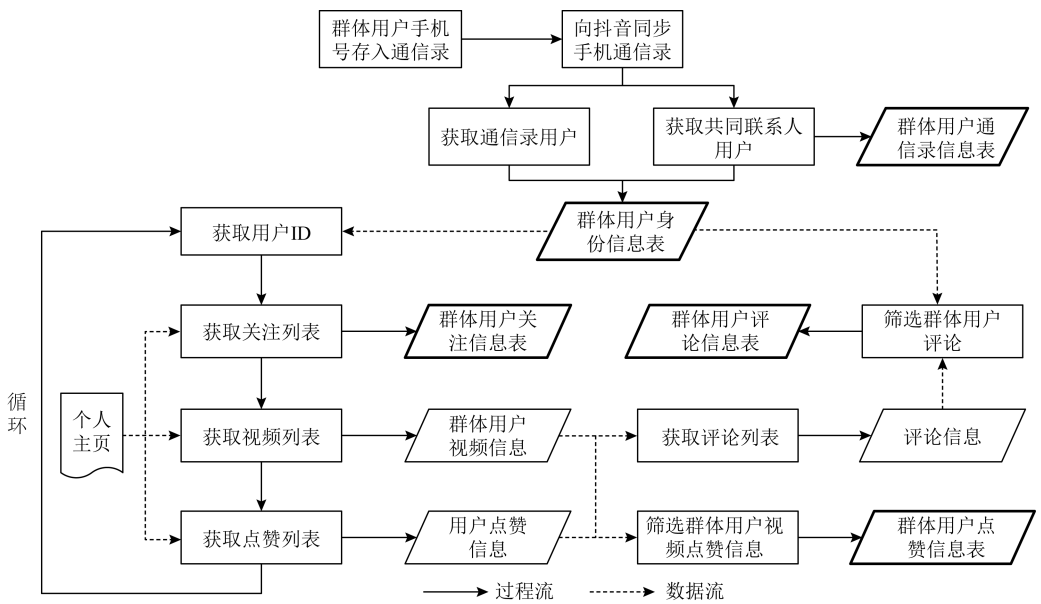


Fig. 4 The collecting process of TikTok user data

图 4 抖音用户数据采集过程

攻击者首先将已经获得的群体用户电话号码存入手机的通信录,通过“查找通信录好友”功能向抖音后台同步手机通信录,或者通过自动化的方式、以欺骗的手段向抖音服务器上传通信录,或让抖音服务器获知该用户的通信录发生了更新.通过查找通信录好友获得所有可以发现抖音 ID 的通信录好友用户集合(以 A 表示),并通过共同联系人接口自动发送数据包,逐一查找攻击者与这些用户的共同好友列表,并通过第 2 节中介绍的迭代查询的方法获得这个群体中的用户的尽可能多的抖音 ID 集合(以 B 表示).在这个过程中,攻击者可以获知每一位用户的通信录上有群体中的哪些人,由此建立“群体用户通信录信息数据表”,并获得群体中尽可能多的用户身份信息集合,由此建立“群体用户身份信息数据表”.

然后,通过自动化程序发送请求,对 B 中的任意一位用户 u,查询其关注列表(被关注者 ID 集合),根据被关注者 ID 集合,筛选出属于本群体的被关注者,建立"群体用户关注信息数据表".查询 u 发布的视频列表(视频号、视频标题),获取其视频列表中的每个视频的评论信息(评论者 ID、评论内容等),根据评论者 ID 集合筛选出属于本群体的用户的评论,建立"群体用户评论信息数据表".查询 u 点赞过的视频列表(视频发布者 ID、视频号、视频标题),根据发布者 ID 或视频号,筛选出属于本群体用户的视频,建立"群体用户点赞信息数据表".

经过分析发现,抖音 App 采用了基于客户端加密的反爬虫策略,所有的请求都需要携带 X-Gorgon 参数,X-Gorgon 是对 Url, Postdata, Cookies, SesseionID 进行多次 MD5 加密后组成的字符串,而 X-Gorgon 的加密计算在客户端就可以完成,不需要服务器端的参与.Android 版抖音 App 的 X-Gorgon 的加密计算在 libcms.so 库中完成,为了消除破解 libcms.so 的麻烦,本文采取直接利用抖音 App 中的 libcms.so 库的方法.方法是先从抖音 App 中取出 libcms.so,然后开发一个 App,编写 Java 代码通过 JNI 机制调用 libcms.so 中的 X-Gorgon 的加密计算函数.然后将 App 放在 Android 设备中安装并运行,这个 App 通过开放端口的办法,接收外来请求,并为请求者计算并返回 X-Gorgon.

发送请求的方法如图 5 所示.我们让 PC 端电脑和 Android 设备处于同一个局域网 WiFi 中,PC 端程序负责构造发送给抖音服务器的请求,包括 Url, Postdata, Cookies, SesseionID 等信息,并向 Android 设备中的服务 App 发送计算 X-Gorgon 请求(请求中携带上一步中构造的 Url, Postdata, Cookies, SesseionID),服务 App 收到计算 X-Gorgon 请求后,计算并返回 X-Gorgon,PC 端收到 X-Gorgon 后,重新构造发送给抖音服务器的 Web 请求数据,并向抖音服务器发送请求.抖音服务器收到请求后,验证 X-Gorgon 参数合法,便会把该请求当做合法请求对待.

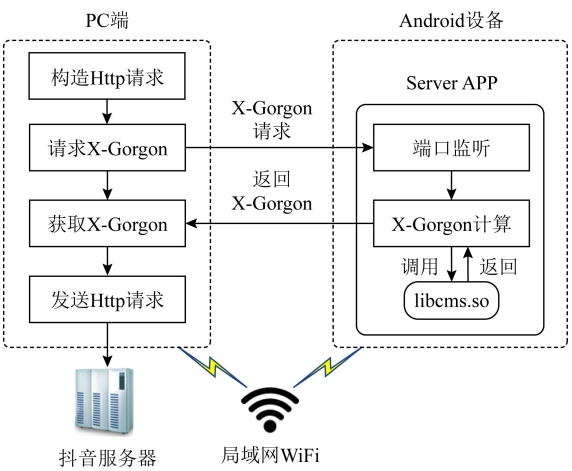


Fig. 5 Construction method of TikTok legitimate request data

图 5 抖音合法请求数据构造方法

5 实验与分析

为了验证基于抖音共同联系人的群体用户关系分析的准确性,我们选择了 3 个真实的社会群体进行了实验,本节将分实验 1 和实验 2 对这 3 个社会群体分别进行实验和分析.

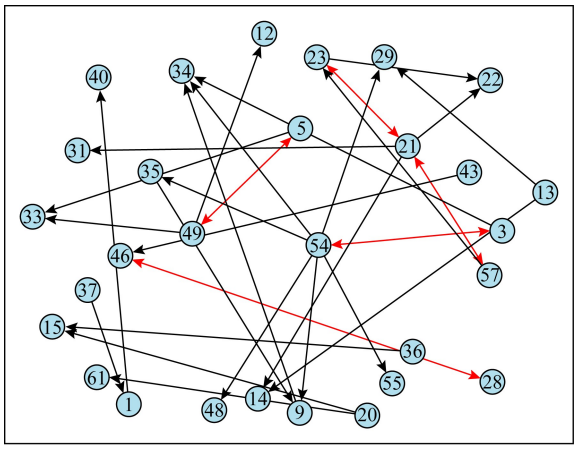
5.1 实验 1

实验 1 选择了某一个学校的某一个班级的 72 位学生做实验,做实验之前我们已经告知学生实验的细节,并征得了学生的同意.我们将学生的手机号码添加到手机通信录,并通过"查看通信录好友"功能同步到抖音服务端.实验中有 61 人的手机号和抖音 ID 的对应关系被找到,其中有 35 人通过手机联系人找到,26 人通过共同联系人找到,说明这 26 人设置了"允许将我推荐给好友"选项为关闭.能获得抖音 ID 的 61 人中有 6 人的关注列表不能浏览,说明这 6 人设置了"谁可以看我对作品的点赞"选项为"仅自己和作者".有 5 人的点赞视频列表不能浏览,说明这 5 人设置了"谁可以看我对作品的点赞"为"仅自己和作者".这些被屏蔽的信息会导致实验对群体用户的关系分析不全面的问题,但所涉及数据不多,对总体影响并不大.

72 人中有 11 人不能通过手机号码获得抖音 ID.通过与学生的交流,我们了解到其中有 8 人没有将自己的手机通信录同步到抖音 App,还有 3 人没有用过抖音,即没有抖音账号.这些缺失的群体用户抖音信息也会导致对群体用户的关系分析不全面的问题,因此,本实验只针对群体中能找到抖音 ID 的 61 人之间的相互关系进行分析,不把此 11 人考虑在内.

图 6 所示为 61 个用户相互间的关注关系,图 6 中的单向箭头从关注者指向被关注者,双向箭头表示双方相互关注.图 7 所示为 61 个用户相互间的通信录关系,图 7 中的单向箭头从某一用户指向其手机通信录中的用户,双向箭头表示双方都在对方的通信录上.图 8 所示为 61 个用户相互间的点赞关系,

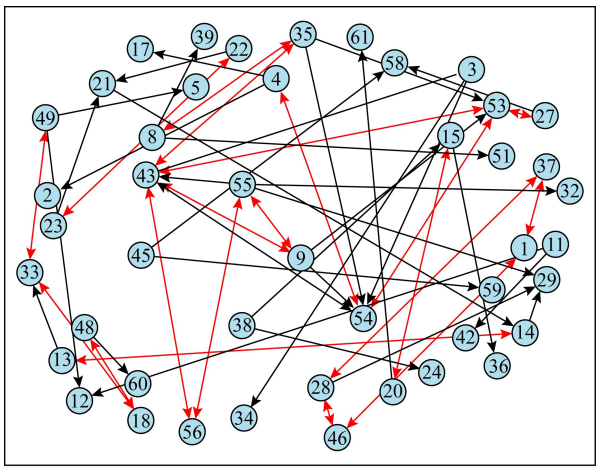
图 8 中的单向箭头从点赞者指向被点赞者,箭头上的数字代表点赞者对被点赞者发布的视频点赞了多少次.双向箭头表示双方相互进行了点赞,箭头上的数字代表双方分别点赞了多少次,例如:用户 54 和 55 双方相互进行了点赞,箭头上的数量 4:3代表用户 54 向用户 55 点赞了 4 次,用户 55 向用户 54 点赞了 3 次.图 9 所示为 61 个用户相互间的评论的亲密分量的计算结果信息,图 9 中的单向箭头从评论者指向被评论者,箭头上的数字代表评论者对被评论者发布的视频的评论的亲密分量.双向箭头表示双方相互进行了评论,箭头上的数字代表双方评论的亲密分量分别是多少,例如:用户 21 和 22 双方相互进行了评论,箭头上的数量 2.25:3代表用户 21 对用户 22 的评论亲密分量为 2.25,用户 22 对用户 21 的评论亲密分量为 3.



圆圈中的数字表示用户编号.

Fig. 6 The following relationship between group users

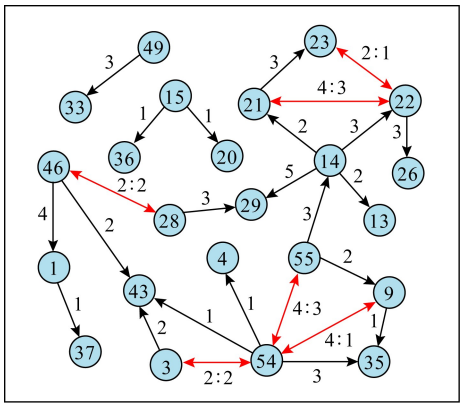
图 6 群体用户相互之间的关注关系



圆圈中的数字表示用户编号.

Fig. 7 The contacts relationship between group users

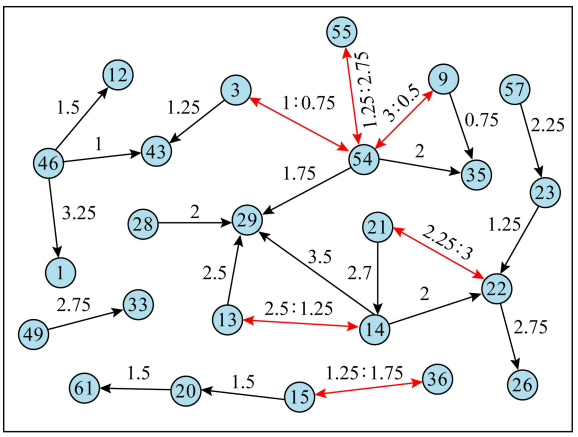
图 7 群体用户相互之间的通信录关系



圆圈中的数字表示用户编号.

Fig. 8 The video like relationship between group users

图 8 群体用户相互之间的点赞关系



圆圈中的数字表示用户编号.

Fig. 9 The comment intimacy components between group users

图 9 群体用户相互之间的评论亲密分量

我们按照式(5)对群体用户之间的双向亲密度进行了计算,为了将用户之间的亲密度限制在 1 以内,我们设置 α, β, γ, δ 系数的值为保证每个类别的最大双向亲密度之和为 1.在确定系数的取值之前,我们对 300 个用户进行了问卷调查,让参与调查的人选出 3 种其认为的最能体现亲密度的亲密分量,关注、通信录、点赞和评论所得的票数如表 5 所示,化简比例 173∶186∶262∶279 为 2∶2.15∶3.03∶3.23,近似为 2∶2∶3∶3的关系.因此我们设置 α, β, γ, δ 系数的值分别为保证这 4 种类别的最大双向亲密度为 0.2, 0.2, 0.3, 0.3,以平衡 4 种亲密分量关系.这 4 个数值也代表了 4 种亲密分量在用户关系分析上所占的比重.例如:对于关注类别的亲密度,若双方互为关注,则亲密分量为 2,为最大值,那么我们将 α 设置为 0.2/2=0.1,此时,关注类别的最大亲密度为 0.2.表 6 所示为各系数的取值说明.

Table 5 Survey Results on the Importance of Intimacy

表 5 亲密分量重要性调查结果

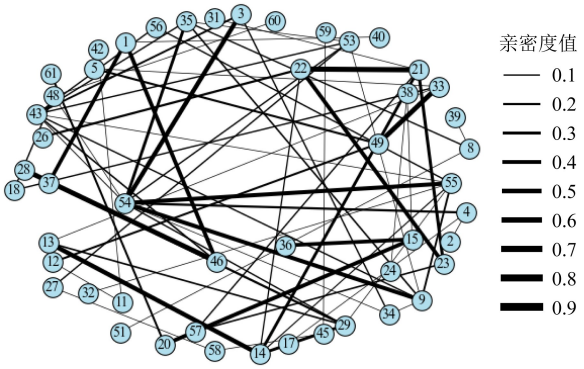
关注	通信录	点赞	评论
173	186	262	279

Table 6 Value of Intimacy Component Coefficient

表 6 亲密分量系数取值

类别	亲密分量最大值	所占比重	系数设置
关注	2	0.2	$\alpha=0.1$
通信录	2	0.2	$\beta=0.1$
点赞	1.33	0.3	$\gamma=0.23$
评论	5.25	0.3	$\delta=0.057$

图 10 所示为亲密度的计算结果,图 10 中线条(边)的粗细代表亲密程度,线条越粗代表双方关系越亲密.图 10 右边展示了亲密度值 0.1～0.9 的图例,群体中用户之间亲密度最大值为 0.73,最小值为 0.1.



圆圈中的数字表示用户编号.

Fig. 10 Calculation results of two-way intimacy of group users

图 10 群体用户双向亲密度计算结果

为了验证亲密度计算结果的准确性,我们对班级的学生进行了问卷调查,问卷的问题之一是“你和你们班关系最好的 3 个人是谁?”,要求每位学生填写与自己关系最好的 3 个同班同学.由于这 61 人相互间的亲密度分布在 0.1～0.8 之间,因此实验对图 10 中不同亲密度值对应的边进行单独统计,分析不同亲密度值对应的边与问卷调查结果匹配的程度,方法如下:

1) 将每位学生的回答结果形成 3 条边,分别是从学生自己指向他(她)填写的 3 位同学,例如:用户 28 的问卷调查结果为{28,[29,46,66]},则可以形成 3 条边(28,29), (28,46), (28,66),本文称这些边为实际边,若学生的回答结果中包含 11 个未获得抖音 ID 的用户,则不参与匹配.

2) 将图 10 中的边形成双向关系,例如:边(1,37)形成 2 条边(1,37), (37,1),并将这些边按照亲密度值进行归类,本文称这些边为测试边.

3) 将实际边与测试边进行匹配,看不同亲密度值范围对应的测试边分别有多少条被匹配,并计算匹配比例.例如:亲密度为 0.7～0.8 之间的 4 条测试边为(21, 22),(22, 21),(28, 46),(46, 28),编号为 21,22,28,46 的用户回答问卷的结果为:{21,[23,43,57]},{22,[21,26,54]},{28,[29,46,66]},{46,[1,28,33]}.则亲密度为 0.7～0.8 之间的 4 条测试边中有 3 条边匹配,只有(21, 22)没有匹配,则匹配率为 75%.

测试结果如表 7 所示,从表 7 中可以看出,总体上亲密度值越高的边被匹配的比例越高,说明实验对群体用户的亲密度计算结果是有效的.

Table 7 Validation of Intimacy Calculation Results

表 7 亲密度计算结果验证

亲密度	边数	匹配数	匹配比例/%
$0.1\leqslant Q<0.2$	54	10	18.52
$0.2\leqslant Q<0.3$	44	16	36.36
$0.3\leqslant Q<0.4$	8	3	37.50
$0.4\leqslant Q<0.5$	8	5	62.50
$0.5\leqslant Q<0.6$	14	7	50.00
$0.6\leqslant Q<0.7$	6	4	66.67
$0.7\leqslant Q<0.8$	4	3	75.00

我们按照式(6)(7)对群体用户的群体活跃度进行了计算,结果如图 11 所示,图 11 展示了各用户的主动活跃度和被动活跃度.为了验证群体活跃度计

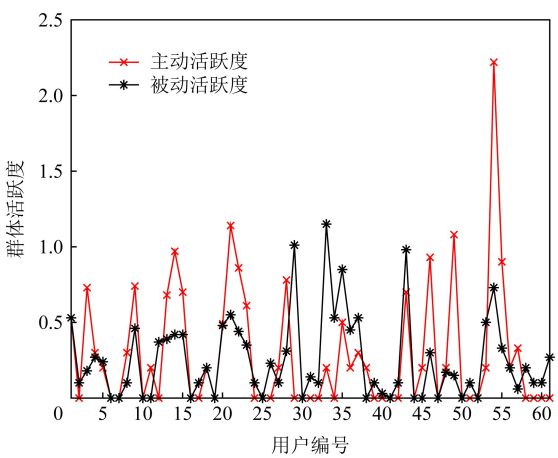


Fig. 11 Calculation results of group users' group-activeness

图 11 各群体用户群体活跃度计算结果

算结果的准确性,我们对学生进行的问卷调查的问题之二是“你们班最合群的3个人是谁?”(“合群”可以近似认为是对1个人的群体活跃性的一种表述,易于被调查人员理解),我们对各位学生的得票数进行了统计,其中得票数最高的10人如表8所示,表8中列出了这些人的得票数、视频数、主动活跃度和被动活跃度,得票数排名前10用户都在前述61人之中.

Table 8 Information about the Most Group-Active 10 Users

表8 群体中最活跃的10位用户的相关信息

用户编号	得票数	视频数	主动活跃度	被动活跃度
33	27	0	0.20	1.15
28	25	7	0.78	0.31
21	18	3	1.14	0.55
54	16	59	2.22	0.73
22	15	10	0.86	0.44
14	15	16	0.97	0.42
35	13	0	0.50	0.85
43	11	0	0.70	0.98
15	11	2	0.70	0.42
46	9	40	0.93	0.30

结合表8和图11可以看出,这些用户的主动活跃度和被动活跃度都较高,说明实验对用户的群体活跃度的计算结果是有效的.编号为33,35,43的用户虽然没有在抖音上发视频,但仍然具有很高的群体活跃度和得票数,而编号为46和54的用户虽然发布抖音视频很多,但群体活跃度和得票数并没有排在最前面,说明用户的人缘好坏、在群体中是否活跃与发布抖音视频多少没有直接关系.

5.2 实验2

在实验2中,我们选择2个社会群体进行了实验,群体1为某公司的员工群体,总人数68人;群体2为某兴趣爱好团体,总人数92人.在征得用户同意的情况下,我们采取与实验1相同的方式收集了群体用户的电话号码,并添加到手机通信录,通过“查看通信录好友”功能同步到抖音服务端.

群体1有54人的手机号和抖音ID的对应关系被找到,群体2有76人的手机号和抖音ID的对应关系被找到.我们采取和实验1相同的方法对用户之间的亲密度进行了计算.由于实验1中亲密度小于0.2的2个用户之间的亲密关系预测成功率较低,因此实验2主要对0.2以上的亲密度进行分析,对亲密度大于0.2的2个用户设置问卷调查.群体1中亲密度在0.2以上的边数为76条,群体2中亲密度在0.2以上的边数为112条.

我们制作了网页版的调查问卷,并通过微信群消息的形式邀请用户参与问卷调查.问卷的问题设置与实验1有所不同,主要包括2个问题:1)在本团体(单位)中你与以下几个人关系比较亲密;2)在本团体(单位)中以下几个人与其他人关系处理得最好.问题1主要用来分析亲密度计算结果的有效性,在问题1中我们会给出与被调查用户亲密度在0.2以上的用户选项供用户选择,用户对选项进行正确与否的判断.问题2主要用来分析群体活跃度计算结果的有效性,在问题2中我们会给出群体活跃度最高(主动活跃度和被动活跃度的和最高)的10位用户供用户选择,用户同样对选项进行正确与否的判断.

群体1中实际有42人参与了问卷调查,所涉及的亲密度在0.2以上的边数为50条,形成问题1的选项数为50项.群体2中实际有63人参与了问卷调查,所涉及的亲密度在0.2以上的边数为86条,形成问题1的选项数为86项.我们对问卷调查中问题1的结果进行了整理,结果如表9所示.从表9可以看出,总体上亲密度值越高的边(选项)被用户认可的可能性越高,说明实验对群体用户的亲密度计算结果是有效的.

Table 9 Results of Two Groups' Intimacy Questionnaire

表9 2个群体的亲密度问卷调查结果

亲密度	群体1			群体2		
	边数	匹配数	匹配比例/%	边数	匹配数	匹配比例/%
$0.2\leqslant Q<0.3$	24	9	37.50	34	12	35.29
$0.3\leqslant Q<0.4$	10	4	40.00	18	8	44.44
$0.4\leqslant Q<0.5$	6	4	66.67	14	8	57.14
$0.5\leqslant Q<0.6$	4	3	75.00	8	5	62.50
$0.6\leqslant Q<0.7$	4	2	50.00	6	5	83.33
$0.7\leqslant Q<0.8$	2	2	100.00	6	4	66.67

我们也对问题2的结果进行了统计,群体1中参与调查的42人共对420个选项(每个用户对10个活跃度最高的用户进行正确性判断)进行选择,其中281个选项被用户判断为正确,正确率为66.90%.群体2中参与调查的63人共对630个选项进行选择,其中379个选项被用户判断为正确,正确率为60.16%.群体活跃度的判断准确率均超过60%,说明实验对群体活跃度的计算结果是有效的.群体1

的正确率高于群体 2,我们推测原因是群体 1(单位)的成员相互之间的联系紧密程度高于群体 2(兴趣爱好团体).

6 安全威胁和防范建议

通过第 5 节中的实验可以了解到,抖音泄露共同联系人的安全漏洞为攻击者挖掘群体用户的关系提供了可能,为攻击者对群体用户实施进一步的网络攻击提供了有力的辅助,给用户造成了一定的安全威胁.本文将可能造成的安全威胁概括为 3 个方面:

1) 抖音群体用户的抖音账号被暴露.攻击者在获得群体用户的电话号码信息后,能够获得群体中尽可能多的用户的抖音账号信息,使用户对自己的隐私设置失效(关闭"允许将我推荐给好友"选项).

2) 攻击者伪装为某一用户的亲密朋友的身份对该用户进行诈骗.攻击者通过计算用户之间的亲密度,能够大致掌握某一用户在群体中的社交关系,找到与其较为亲密的朋友.攻击者通过某些攻击手段,伪装为该用户的亲密朋友,对该用户进行诈骗,可能导致攻击更容易取得成功.例如,攻击者可以通过植入恶意软件的方法,以亲密朋友的 QQ 或微信账号的身份向该用户发送信息进行诈骗.

3) 攻击者对一些群体活跃性较高、人缘较好的用户进行重点攻击.通过群体活跃度分析,攻击者能够了解到群体中哪些用户群体活跃度较高,群体活跃度较高的用户可能对周围的用户更具有影响力.因此攻击者针对这些用户实施攻击,并借助于他们将攻击扩散到群体中的其他用户,可能更易使攻击效果扩大化.

为了防御这种安全问题,我们分别从厂商和普通用户的角度提出了安全防范建议.

1) 从抖音厂商的角度.抖音 App 开发者不应当把共同联系人的账号信息向用户展示出来(如图 2 所示),而是只向用户展示共同联系人的数量即可.如果必须要展示的话,也应当禁止展示那些把"允许将我推荐给好友"选项设置为关闭的用户.此外,由于抖音 App 基于客户端加密的反爬虫策略已经被实验证明是无效的,因此,抖音 App 开发者应当采取有效的服务器端防爬虫措施来防止攻击者大量收集用户数据的行为,而不应当仅仅依靠客户端参数加密来防爬虫.

2) 从用户的角度.我们认为用户应当谨慎地使用"查找通信录好友"的功能,尽量不要将自己的通信录信息传给抖音服务器,不仅给自己带来潜在的危害,也可能损害通信录好友的利益.此外,用户在从事各种网络活动时,也应当保护好自己的手机号码信息,防止被恶意攻击者所利用.

7 相关工作

与本文研究社交 App 群体用户的用户关系类似,一些研究者在社交网络的用户关系强度、用户亲密度等研究方面做过一些相关工作.

社交网络用户之间的关系强弱可以通过用户关系强度来度量,目前关于社交网络用户关系强度的研究主要建立在 3 种模型基础上:图模型、相似度模型和概率模型.在图模型方面,Zhuang 等人[21]提出一个基于核的学习算法来推断社交媒体 Flickr 用户的社交关系强度,利用核函数成对学习法来估算社会关系强度.Zhao 等人[22]提出了一个综合考虑用户个人信息、交互活动和活动域来衡量用户之间关系强度的通用框架.Zhong 等人[23]提出了一个基于用户特征和交互信息的潜在特征来推算用户社会关系强度的模型,将矩阵分解与多核函数相结合.Guo 等人[24]将多视图下的用户关系学习和关系强度建模耦合到一个过程框架中来发现潜在的用户社会关系强度.针对因隐私设置或是数据丢失等问题导致的节点之间的关系不可见的问题,一些研究者运用边缘权重预测的方法来分析社交网络中节点之间的关系强度.Aicher 等人[25]提出了一种加权随机块模型,将随机块模型推广到具有任意指数族分布的边权的网络,推断链接和边缘权重的存在.Zhu 等人[26]提出了一种新的基于局部网络结构的权值预测方法,形成每个节点的邻居集,能够预测在部分链接权重信息丢失的情况下的链路权值.Sá 等人[27]基于有监督的机器学习进行节点链接预测,使用权值来改善有监督的链接预测的相关性.在相似度模型方面,Singla 等人[28]提出搜索主题的相关性高的用户之间存在较强的用户关系.Yu 等人[29]通过对微博用户兴趣的研究发现,微博用户之间的关系强度与兴趣相似性正相关.随着基于位置服务的兴起,一些研究者提出通过分析用户轨迹相似度来挖掘移动社交网络中的用户关系强度的方法[30-31].在概率模型方面,Pham 等人[32]根据用户在时间和空间上的共现关系,构建熵模型来计算用户关系强度.He 等人[33]提出多元逐步线性回归模型,利用朴素贝叶斯分类器计算用户关系强度.还有一些研究者提出隐变量

模型,通过用户之间的交互行为隐变量计算用户关系强度[34-35].Xiong 等人[36]提出了一个概率图模型来衡量社会网络中不同用户之间的关系强度,该模型考虑了用户特征之间的相似性、用户名的共现性和不同活动域中的交互活动.

在社交网络用户亲密度的研究方面,一些研究者提出了利用社交网络用户亲密度进行社交社区发现的方法.刘瑶等人[37]提出了有向加权社交网络中节点之间的亲密度定义,在此基础上设计了一种基于节点亲密度和度的社区结构检测方法,为无向无权、有向无权、无向加权、有向加权等社交网络的社区划分提供了统一的解决方法.Zheng 等人[38]针对移动社交网络中高效的数据转发所面临的网络连通性的不确定性的难题,提出了一种基于地理亲密度的移动社交网络(mobile social networks, MSNs)数据转发方案,通过一种新的度量地理亲密度的方法,对社交节点的地理信息和节点之间的友谊进行量化.Wang 等人[39]提出了一种基于节点间非对称亲密度的分层社区检测算法.在有向加权网络中,利用节点聚类算法有效地检测出社群结构,生成一组最优的社群,并采用了一个新的非对称参数来度量节点之间的密切关系.Kong 等人[40]提出了一种基于用户亲密度的社区发现算法,利用社交网络局部拓扑信息构造一个亲密度矩阵来度量节点间的亲密度,通过计算节点的重要性,保证节点按特定的顺序更新,并在标签传播过程中评估标签对更新节点的影响.Chen 等人[41]提出了一种基于亲密度进化聚类的动态社区结构检测算法,利用时间加权相似度矩阵,计算出社群演化过程中的时间变化,并利用微分方程来学习亲密度演化行为,在交互过程中根据迭代模型更新节点间的亲密度.

以上相关研究工作在计算用户之间的关系强度或亲密度时都没有考虑到用户之间的文本交流内容的情感值,没有如本文类似的工作,即采用社交媒体评论的文本情感分析来计算用户之间的关系强度和亲密度,因此,本文的研究工作是对已有研究工作的一种创新.

本文提出攻击者利用群体用户之间的关系能够实施更有效的攻击,与之类似,一些研究者也提出过关于攻击者利用社交网络用户之间的关系进行攻击的安全问题和相应的安全防御方法.

Wondracek 等人[42]提出利用社交网络上的用户所属组的成员身份信息,可以唯一地标识一位匿名用户,实施反匿名攻击.Tai 等人[1]提出利用社交网络中 2 个相连接的用户节点的度数进行友谊攻击,对已发布的社交网络数据集中的相关受害者进行识别.Sun 等人[2]在 Tai 等人[1]的工作的基础上进行了扩展,提出了一种基于 2 个用户的共同联系人数量的攻击的新方式,并提出了一种 k 次非负矩阵分解(k-degree non-negative matrix factorization, k-NMF)匿名性方法保护用户共同好友数据.Macwan 等人[3]提出一种基于共同朋友序列的匿名化方法来防御共同好友攻击,保证了在共同好友序列中至少有 k 个元素具有相同的值,增加一条边的共同好友数量能够降低其他边的共同好友匿名化要求.Jin 等人[5]发现,在社交网站 LinkedIn 和 Google Plus 中,即使用户通过隐私设置限制了其他用户访问共同好友列表,攻击者也可以采取某些方法发起隐私窃取攻击,识别目标用户的朋友和远程邻居.他们也提出了一种通过用户好友列表进行友谊识别和推断的友谊识别推断(friendship identification and inference, FII)攻击[4],可以利用用户的偏好冲突导致的策略不一致来识别和推断目标用户的很多好友.Liu 等人[9]提出了一种利用朋友搜索引擎挖掘社交网络用户的好友关系的共谋攻击方式,通过一个精心设计的由多个恶意请求者协同发起的查询序列能够使受害者用户的友谊隐私设置失效.此外,一些研究者提出了利用社交网络用户节点之间的关系实施邻域攻击的方法[6-8],在攻击者对目标受害者在社交网络上的邻居节点和邻居之间的关系有一定了解的基础上,邻域攻击可以使攻击者突破传统的匿名技术来识别受害者的身份.为了防御邻域攻击,研究者们提出了一些新的匿名方法,包括 K-匿名(K-anonymity)方法[6]、基于邻域邻接矩阵(ONAM)的匿名方法[7]、动态 l-分集(dynamic-l-diversity)的匿名方法[8]等.

通过对这些研究工作的分析,可以发现这些研究工作所提出的攻击者利用社交网络用户之间的关系进行攻击的方式都没有把实际的社会群体作为攻击对象,而本文所提出的攻击手法是在攻击者已知的社会群体的电话号码等信息的基础上,通过社交媒体来挖掘群体中用户之间的关系,从而为攻击者实施有效的攻击提供辅助,是对已有研究工作的一种创新.

8 结束语

本文提出了抖音共同联系人功能存在的泄露群体用户的抖音账号的安全问题,攻击者还可以利用

收集到的群体用户相互之间的关注信息、通信录信息、视频点赞和评论信息,计算群体用户之间的亲密度和用户的群体活跃度信息,这些信息能够为攻击者实施对群体用户的进一步有效攻击提供一定的辅助.本文给出了具体的计算用户亲密度和群体活跃度的方法,并通过实验验证了所计算的群体用户的亲密度和群体活跃度信息的有效性,同时提出了这种安全问题对群体用户可能造成的安全威胁,给出了相应的安全防范建议.

本文的研究工作也存在一些局限性,例如,本文只是针对抖音 App 进行了研究,实际上这项研究还可以扩展到更多的 App 上,综合运用更多的 App 来分析群体用户的隐私泄露问题,不过由于我们在其他社交 App 上难以找到类似的安全漏洞,并且多个 App 之间通过什么形式进行关联依然是一个待解决的难题,这将是我们未来的研究工作需要解决的问题.此外,由于本文的实验对象是现实中真实的用户群体,考虑到群体用户的个人隐私问题,因此,实验规模和验证力度有限,未来我们将寻找到既能有效验证所提出的理论,又不危害测试人员个人隐私安全的有效实验验证方法.

作者贡献声明:乐洪舟提出研究思路,负责实验方案设计、实验数据分析和论文撰写;何水龙负责实验方案的具体实施、程序设计和测试;王敬负责实验数据采集、算法设计和论文校对.

参 考 文 献

[1] Tai C H, Yu P S, Yang D N, et al. Privacy-preserving social network publication against friendship attacks [C] //Proc of the 17th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2011: 1262-1270

[2] Sun Chongjing, Yu P S, Kong Xiangnan, et al. Privacy preserving social network publication against mutual friend attacks [J]. Transactions on Data Privacy, 2014, 7(2): 71-97

[3] Macwan K R, Patel S J. Mutual friend attack prevention in social network data publishing [C] //Proc of the 2017 Int Conf on Security, Privacy, and Applied Cryptography Engineering. Berlin: Springer, 2017: 210-225

[4] Jin Lei, Takabi H, Long Xuelian, et al. Exploiting users' inconsistent preferences in online social networks to discover private friendship links [C] //Proc of the 13th Workshop on Privacy in the Electronic Society. New York: ACM, 2014: 59-68

[5] Jin Lei, Joshi J B D, Anwar M. Mutual-friend based attacks in social network systems [J]. Computers & Security, 2013, 37: 15-30

[6] Liu Chuangang, Liu I H, Yao Wunsheng, et al. *K*-anonymity against neighborhood attacks in weighted social networks [J]. Security & Communication Networks, 2016, 8(18): 3864-3882

[7] Diwakar A K, Singh N K, Tomar D S. End user privacy preservation in social networks against neighborhood attack [C/OL] //Proc of the 2017 ISEA Asia Security and Privacy. Piscataway, NJ: IEEE, 2017 [2020-09-03]. https://ieeexplore.ieee.org/document/7976991

[8] Hu Xiaoyi, Wang Li'e, Tang Jiaqi, et al. Anonymizing approach to resist label-neighborhood attacks in dynamic releases of social networks[C/OL] // Proc of the 19th IEEE Int Conf on e-Health Networking, Applications and Services. Piscataway, NJ: IEEE, 2017 [2020-09-03]. https://ieeexplore.ieee.org/document/8210763

[9] Liu Yuhong, Li Na. An advanced collusion attack against user friendship privacy in OSNs [C] //Proc of the 40th IEEE Annual Computer Software and Applications Conf. Piscataway, NJ: IEEE, 2016: 465-470

[10] Quest Mobile. Spring 2020 China mobile Internet quarterly report [EB/OL] [2020-09-03]. https://www.questmobile.com.cn/research/report-new/90

[11] Sohu Tech. TikTok, WeRead was sentenced to infringe on user personal information [EB/OL] [2020-09-03]. https://www.sohu.com/a/410574111_115565

[12] Ji Changhong. Chinese Appellation Dictionary [M]. Shijiazhuang: Hebei Educational Press, 2001 (in Chinese)
(吉常宏. 汉语称谓大词典[M]. 石家庄: 河北教育出版社, 2001)

[13] Liu Jingmin. Intimate appellation and Chinese aesthetic taste [J]. Shandong Social Sciences, 1998 (3): 3-5 (in Chinese)
(刘静敏. 亲昵称谓与中国人的审美趣味[J]. 山东社会科学, 1998 (3): 3-5)

[14] Wu Chao. Research on the evolution of modern Chinese widely-used social addressing term [D]. Hohhot: Inner Mongolia University, 2015 (in Chinese)
(吴超. 现代汉语通用社会称谓语的嬗变研究[D]. 呼和浩特: 内蒙古大学, 2015)

[15] Xu Shuang. Cognitive function perspective of semantic generalizations of Chinese intimate address terms [D]. Beijing: China University of Petroleum, 2018 (in Chinese)
(徐爽. 汉语亲昵称谓语语义泛化的认知功能视角[D]. 北京: 中国石油大学, 2018)

[16] Wang Ke, Xia Rui. A survey on automatical construction methods of sentiment lexicons [J]. Acta Automatica Sinica, 2016, 42(4): 495-511 (in Chinese)

(王科,夏睿.情感词典自动构建方法综述[J].自动化学报,2016,42(4):495-511)

[17] Mei Lijun, Zhou Qiang, Zang Lu, et al. Merge information in HowNet and TongYiCi CiLin [J]. Journal of Chinese Information Processing, 2005, 19(1): 64-71 (in Chinese)
(梅立军,周强,臧路,等.知网与同义词词林的信息融合研究[J].中文信息学报,2005,19(1):64-71)

[18] Miao Yuqing, Gao Han, Liu Tonglai, et al. Sentiment analysis based on grid clustering [J]. Journal of University of Science and Technology of China, 2016, 46(10): 874-882 (in Chinese)
(缪裕青,高韩,刘同来,等.基于网格聚类的情感分析研究[J].中国科学技术大学学报,2016,46(10):874-882)

[19] Min Kerui, Ma Chenggang, Zhao Tianmei, et al. BosonNLP: An ensemble approach for word segmentation and POS tagging [C] //Proc of the 2015 Natural Language Processing and Chinese Computing. Berlin: Springer, 2015: 520-526

[20] Peng Haiyun, Cambria E, Hussain A. A review of sentiment analysis research in Chinese language [J]. Cognitive Computation, 2017, 9(4): 423-435

[21] Zhuang Jinfeng, Mei Tao, Hoi S C H, et al. Modeling social strength in social media community via kernel-based learning [C] //Proc of the 19th ACM Int Conf on Multimedia. New York: ACM, 2011: 113-122

[22] Zhao Xiaojian, Yuan Jin, Li Guangda, et al. Relationship strength estimation for online social networks with the study on Facebook [J]. Neurocomputing, 2012, 95: 89-97

[23] Zhong Youliang, Du Lan, Yang Jian. Learning social relationship strength via matrix co-factorization with multiple kernels [C] //Proc of the 2013 Int Conf on Web Information Systems Engineering. Berlin: Springer, 2013: 15-28

[24] Guo Dongyan, Xu Jingsong, Zhang Jian, et al. User relationship strength modeling for friend recommendation on Instagram [J]. Neurocomputing, 2017, 239: 9-18

[25] Aicher C, Jacobs A Z, Clauset A. Learning latent block structure in weighted networks [J]. Journal of Complex Networks, 2015, 3(2): 221-248

[26] Zhu Boyao, Xia Yongxiang, Zhang Xuejun. Weight prediction in complex networks based on neighbor set [J]. Scientific Reports, 2016, 6(1):38080

[27] Sá H R D, Prudencio R B C. Supervised link prediction in weighted networks [C] //Proc of the 2011 Int Joint Conf on Neural Networks. Piscataway, NJ: IEEE, 2011: 2281-2288

[28] Singla P, Richardson M. Yes, there is a correlation—From social networks to personal behavior on the web [C] //Proc of the 17th Int World Wide Web Conf. New York: ACM, 2008: 1627-1628

[29] Yu Yan, Mo Lingfei. Investigating correlation between strength of social relationship and interest similarity [C] // Proc of the 4th Int Conf on Computational Social Networks. Berlin: Springer, 2015: 172-181

[30] Yuan Yihong, Raubal M. Measuring similarity of mobile phone user trajectories—A spatio-temporal edit distance method [J]. International Journal of Geographical Information Science, 2014, 28(3): 496-520

[31] Li Quannan, Zheng Yu, Xie Xing, et al. Mining user similarity based on location history [C/OL] //Proc of the 16th ACM SIGSPATIAL Int Conf on Advances in Geographic Information Systems. New York: ACM, 2008: 1-10 [2020-09-03]. https://dblp.uni-trier.de/rec/conf/gis/LiZXCLM08.html

[32] Pham H, Shahabi C, Liu Yan. Inferring social strength from spatiotemporal data [J]. ACM Transactions on Database Systems, 2016, 41(1): No.7

[33] He Yaxi, Zhang Chunhong, Ji Yang. Principle features for tie strength estimation in micro-blog social network [C] // Proc of the 12th IEEE Int Conf on Computer and Information Technology. Piscataway, NJ: IEEE, 2012: 359-367

[34] Sheng Dakui, Sun Tao, Wang Sheng, et al. Measuring strength of ties in social network [C] //Proc of the 15th Asia-Pacific Web Conf. Berlin: Springer, 2013: 292-300

[35] Xiang Rongjing, Neville J, Rogati M. Modeling relationship strength in online social networks [C] //Proc of the 19th Int World Wide Web Conf. New York: ACM, 2010: 981-990

[36] Xiong Liyan, Lei Yin, Huang Weichun, et al. An estimation model for social relationship strength based on users' profiles, co-occurrence and interaction activities [J]. Neurocomputing, 2016, 214: 927-934

[37] Liu Yao, Kang Xiaohui, Gao Hong, et al. A community detecting method based on the node intimacy and degree in social network [J]. Journal of Computer Research and Development, 2015, 52(10): 2363-2372 (in Chinese)
(刘瑶,康晓慧,高红,等.基于节点亲密度和度的社会网络社团发现方法[J].计算机研究与发展,2015,52(10):2363-2372)

[38] Zheng Yi, Zhang Dafang, Xie Kun. A geography-intimacy-based algorithm for data forwarding in mobile social networks [J]. Chinese Journal of Electronics, 2016, 25(5): 936-942

[39] Wang Xingyuan, Qin Xiaomeng. Detecting communities by asymmetric intimacy in directed-weighted network [J]. International Journal of Modern Physics C, 2017, 28(1): 1750006

[40] Kong Hanzhang, Kang Qinma, Liu Chao, et al. An improved label propagation algorithm based on node intimacy for community detection in networks [J]. International Journal of Modern Physics B, 2018, 32(25): 1850279

[41] Chen Jianrui, Liu Danwei, Hao Fei, et al. Community detection in dynamic signed network: An intimacy evolutionary clustering algorithm [J]. Journal of Ambient Intelligence and Humanized Computing, 2020, 11(2): 891-900

[42] Wondracek G, Holz T, Kirda E, et al. A practical attack to de-anonymize social network users [C] //Proc of the 2010 IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2010: 223-238

Yue Hongzhou, born in 1987. PhD, lecturer. His main research interests include the mobile app security, Android system security, and social network security.

乐洪舟,1987 年生.博士,讲师.主要研究方向为移动应用安全、Android 系统安全和社交网络安全.

He Shuilong, born in 1984. Master candidate. His main research interests include mobile networks security and cryptographic application.

何水龙,1984 年生.硕士研究生.主要研究方向为移动网络安全和密码学应用.

Wang Jing, born in 1989. PhD, lecturer. His main research interests include social network security, artificial intelligence, and machine learning.

王 敬,1989 年生.博士,讲师.主要研究方向为社交网络安全、人工智能和机器学习.