doi:10.3969/j.issn.1000-436x.2013.z1.025

基于公开信息的社交网络隐私泄露

吕少卿[1]，张玉清[1,2]，倪平[2]
（1. 西安电子科技大学 综合业务网理论及关键技术国家重点实验室信息安全研究中心，陕西 西安 710071；
2. 中国科学院大学 国家计算机网络入侵防范中心，北京 100190）

摘 要：针对社交网络公开信息提出了一种隐私信息推测算法，通过对用户的好友关系网络进行社区发现，利用社区内一部分好友公开信息推测其他好友隐私信息。实验表明，该算法只需利用少量公开信息就能以较高的准确率推测出其他用户大量的隐私信息。

关键词：社交网络；隐私泄露；社区发现
中图分类号：TP393.08 **文献标识码**：B **文章编号**：1000-436X(2013)Z1-0190-07

Privacy leakage in online social networks based on public information

LV Shao-qing[1], ZHANG Yu-qing[1,2], NI Ping[2]
(1. Information Security Research Center of State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China;
2. National Computer Network Intrusion Protection Center, University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract: An algorithm was proposed to predict privacy information in social network with public information. This algorithm detected communities in the friend relationship network and inferred privacy information using public information of some friends. The results show that users' privacy information can be estimated with high accuracy when given only a little public information.

Key words: online social network; privacy leakage; community detection

1 引言

随着社交网络的飞速发展，越来越多的用户通过社交网络来记录自己的生活、与好友沟通交流。为了更好地交流以及得到感兴趣的信息，用户在社交网络中通过教育信息以及兴趣等条件添加其他用户为好友，同时这些用户由于具有相同的信息或爱好，它们之间也可能具有好友关系。在社交网络中由一个用户的好友以及这些好友之间的关系所组成的网络结构称之为该用户的好友关系网络。在好友关系网络中，根据联系的紧密程度，这些用户能够分为不同的群，其中群内用户的好友关系紧密，不同群之间用户的关系稀疏，那么这些群就称为这个好友关系网络中的社区[1]。

如图 1 所示，图中每个圆环代表一个用户，这些用户都与用户 u 具有好友关系，圆环内的字符表示该用户的标识，圆环间的连线表示这 2 个用户存在好友关系。这些用户以及用户间的好友关系所组成的网络称为用户 u 的好友关系网络。根据这些用户之间的好友关系，利用社区发现算法能够将它们分为 A、B、C、D 4 个社区，分别由图中椭圆内用户组成。A 社区由用户 $A1$、$A2$ 组成，B 社区由用户 $B1$、$B2$、$B3$ 组成，C 社区由用户 $C1$~$C5$ 组成，用户 $D1$~$D4$ 组成了 D 社区。同一社区内用户的关系紧密，不同社区之间用户的关系比较稀疏。

为了更好地标识自己，大量用户在社交网络中填写了个人真实信息。社交网络对用户信息提供了

收稿日期：2013-06-26
基金项目：国家自然科学基金资助项目（61272481）；北京市自然科学基金资助项目（4122089）
Foundation Items: The National Natural Science Foundation of China(61272481); The Natural Science Foundation of Beijing (4122089)

相应的设置以保护个人隐私。用户通过隐私设置来控制哪些信息能够被哪些用户访问。但大量的研究发现，根据用户已公开的信息，利用网络分析以及数据挖掘的方法能够推测出用户未公开的信息，从而造成用户的隐私泄露[2~6]。

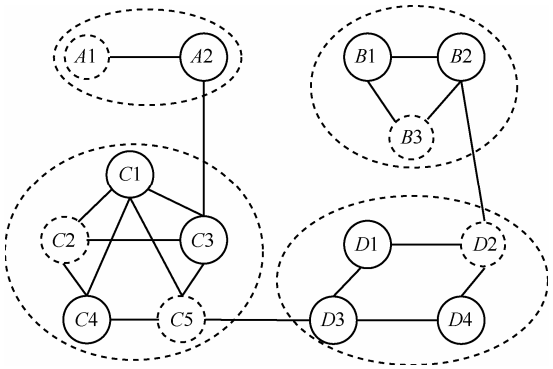

图 1　社区示意

本文以人人网为例，提出了一种基于社交网络中用户公开信息来推测隐私信息的算法。通过社区发现将用户好友关系网络分为不同的社区，根据同一社区的用户具有共同信息这一特点，利用社区内一部分用户公开的教育信息推测社区内其他用户未公开的教育信息。

本文的主要贡献如下。

1) 首次利用社交网络中单个用户好友关系网络的社区结构来推测用户隐私。

2) 分析了人人网中 40 位用户的好友关系网络以及社区结构。

3) 利用所分析的社区结构和社区内用户公开的教育信息推测社区内其他用户未公开的教育信息。

2　相关工作

大量的工作涉及到利用社交网络中公开信息来推测用户个人隐私信息，这些工作利用了社交网络中用户的个人信息、好友关系或这两者的结合。

Zheleva 与 Getoor[2]利用用户的好友关系以及群关系来推测用户的隐私信息，并且比较了利用不同的攻击模型所推测结果的正确率，结果显示利用群关系所得到的结果准确率更高，但是在实验的过程中作者直接假设用户 50%的好友的信息是公开的，而本文的实验数据完全来自真实的社交网络。

在文献[3]中作者利用修改的朴素贝叶斯方法，通过对用户个人信息以及好友关系的分析推测用户的其他信息(政治观点)，并且比较了只用个人信息、只用关系信息以及结合个人信息和关系信息这 3 种情况下的正确率。

Dey 与 Tang 等人[4]利用用户不同程度的公开信息采用不同的方法来推测用户的年龄信息。1）利用用户公开的上学时间来确定；2）利用用户公开的好友列表中好友的年龄来确定；3）没有公开上学时间也没有公开好友列表的用户通过逆向好友查找来确定用户的好友，然后再执行第 2）步。作者最后提出了一种迭代算法，不只利用用户的好友信息，而且包括好友的好友以及 3 层好友关系来推测用户的年龄。

虽然这些工作利用了用户的公开信息以及好友关系来推测隐私信息，但它们都是将用户的好友关系网络作为一个整体去推测用户自身的信息，而本文的工作是利用用户好友关系网络的社区结构，通过社区内一部分用户的公开信息来推测其他用户的隐私信息，相比之下，本文所推测的隐私信息更多。

在文献[5]中作者只利用了用户的兴趣信息来推测用户的隐私信息（性别、关系、国籍、年龄）。作者先用 LDA(latent dirichlet allocation)算法将用户填写的音乐爱好进行聚类，然后通过同一类中一些用户的公开信息来推测其他用户的信息。作者认为具有相同爱好的用户应该有相似的个人信息。但作者只利用了用户的兴趣信息，而且所推测的隐私信息是性别、关系、国籍、年龄等，而本文所推测的信息为教育信息，更具有隐私性。

与本文的工作相类似，Mislove[6]等人的工作同样利用用户公开信息和社区发现来推测用户的隐私信息，但只针对整个数据集进行社区发现，而本文的工作是对单个用户的好友关系网络进行社区发现，所需的初始公开信息更少、算法更灵活。

3　基于公开信息的社交网络隐私泄露

3.1　社区结构

网络中社区结构是指由节点组成的群，群内节点之间的连接比较紧密，而群之间节点的连接相对比较稀疏。其具体定义为[7]：用 G 来表示网络所对应的图，i 为其中一个节点，则 i 的度为

$$k_i = \sum_j A_{ij} \tag{1}$$

若节点 i 与节点 j 存在连接，则 A_{ij}=1，否则 A_{ij}=0。

节点 i 所在的图为 S，此时 S 的节点集合与连接的集合均为 G 中节点集合与连接集合的子集。即 S 是 G 的子图。这种情况下，节点 i 的度可分为 2 部分。

$$k_i(S)=k_i^{\text{in}}(S)+k_i^{\text{out}}(S) \tag{2}$$

其中，

$$k_i^{\text{in}}(S)=\sum_{j\in S}A_{ij} \tag{3}$$

即节点 i 与 S 内部节点的连接数。

$$k_i^{\text{out}}(S)=\sum_{j\notin S}A_{ij} \tag{4}$$

即节点 i 与 S 外部节点的连接数。

若

$$k_i^{\text{in}}(S)>k_i^{\text{out}}(S),\forall i\in S \tag{5}$$

则子图 S 为原网络的一个强社区结构。

若

$$\sum_{i\in S}k_i^{\text{in}}(S)>\sum_{i\in S}k_i^{\text{out}}(S) \tag{6}$$

则子图 S 为原网络的一个弱社区结构。即子图 S 内所有节点在 S 内的度之和比在 S 外的度之和大。

参数 modularity 是 Newman 专门为社区结构提出的度量[1]，是网络中社区结构与随机网络中社区结构之间的差别，其函数表达[8]为

$$Q=\frac{1}{2m}\sum_{ij}(A_{ij}-\frac{k_ik_j}{2m})\delta(C_i,C_j) \tag{7}$$

其中，k_i 和 k_j 分别为节点 i 和 j 的度，C_i 是节点 i 所属的社区；m 是网络中总连接数。若节点 i 与节点 j 存在连接，则 A_{ij}=1，否则 A_{ij}=0。当 C_i =C_j 时，$\delta(C_i,C_j)=1$，否则为 0。modularity 范围为−1~1，0 表示该网络中没有比随机网络更多的社区结构，modularity 的 Q 值越大表示社区结构越强。

有大量的算法可用来在网络结构中发现社区，传统的如计算机科学中的图形分割(graph partitioning)以及社会学中的层次聚类(hierarchical clustering)[9~13]等，也有基于 modularity 的算法，具体可参见文献[14]。本文采用 Blondel D 等人提出的以 modularity 为度量的社区发现算法[15]。该算法首先将网络内每个节点作为一个社区，然后每次将节点移动到使 modularity 达到最大的社区中，直到 modularity 不再增加或者只剩下一个节点。

根据文献[6,16]可知，社交网络中具有相同信息的用户更容易成为好友，它们之间具有较强的联系，因此会组成社区。反之在社交网络中根据用户之间联系的紧密程度可以分为不同的社区[17]，社区内用户具有共同的信息，如教育经历、兴趣爱好等。

3.2 隐私推测

在社交网络中大量用户的个人信息是公开的[18]，因此能够通过统计一个社区内信息公开用户的个人信息出现的频率来推测出该社区内用户所共有的信息，即其他信息未公开用户的隐私信息。

对于社交网络中用户 u，由其好友组成的好友关系网络 $G_u=(V_u,E_u)$，$V_u=N_u$ 即与用户 u 邻接节点的集合，$E_u=\{i,j:i,j\in V_u\}$ 即用户 u 的好友之间关系的集合。利用社区发现算法可以将 G_u 分为社区 C_i, i=1,2,…,N，N 为总的社区数。对社区 C_i 内公开信息的用户 Pub_i 计算

$$T_i=\frac{|Pub_i|}{|C_i|} \tag{8}$$

其中，$|C_i|$为社区 C_i 内总用户数，$|Pub_i|$为社区 C_i 内信息公开的用户数。

为了确保所推测信息的有效性，T_i 必须满足：

$$T_i>\theta \tag{9}$$

其中，θ 为实验过程中所设定的阈值。

对于社区 C_i 内不同的公开信息 A_j, j=1,2,…,M，M 为总的公开信息数，分别计算

$$P_j^i=\frac{|A_j|}{|Pub_i|} \tag{10}$$

其中，$|A_j|$为社区 C_i 内具有公开信息 A_j 的用户数。

则社区 C_i 内用户所共有的信息 I_i 是使 P_j^i 达到最大时的 A_j，即

$$\begin{cases}P_k^i=\max\limits_{1\leqslant j\leqslant M}P_j^i\\ I_i=A_k\end{cases} \tag{11}$$

同时为了确保所推测信息的有效性，P_k^i 必须满足

$$P_k^i>\varepsilon \tag{12}$$

其中，ε 也为实验过程中所设定的阈值。

如图 1 所示，实线圆环表示用户的信息是公开的，可以被任何其他用户访问，虚线圆环表示用户的信息是被保护的，只能被该用户的好友或者用户自身访问。社区 C 中用户 $C1$、$C3$、$C4$ 的个人信息

是公开的，因此可以通过统计这些用户个人信息中不同信息所出现的频率，在满足阈值条件下，频率最高的信息就是此社区内用户所共有的信息，从而推测出用户 *C*2、*C*5 具有某种信息。由于用户 *C*2、*C*5 这些信息是未公开的，因此造成了这些用户的隐私泄露。

4　针对人人网的基于公开信息的隐私泄露实验

人人网是当前国内用户数最多的实名制社交网络平台，在 2011 年已经达到 1.6 亿的注册用户[19]，并且主要用户为在校大学生，用户的好友主要是大学同学、高中同学、家人以及同事[16]，因此能够利用本文的算法来推测比较敏感的教育信息。相比其他文献之前的工作，如推测用户的性别、年龄、国籍等，本文所推测的教育信息对用户具有更大的威胁性，恶意用户可以利用所推测的教育信息构造更真实的虚假账号来通过用户的好友请求，甚至可直接对用户进行社会工程攻击。

4.1　数据获取

根据第 3 节的说明，为了执行本文所提出的算法，需要获得某个用户的好友列表以及这些好友之间的好友关系和这些用户公开的教育信息。虽然人人网对用户的个人信息设置了相应的隐私保护策略，但是通过利用这些策略所存在的一些缺陷，作者能够获得所需的数据。

4.1.1　获取好友列表

在社交网络中，用户可以通过与其他用户建立好友关系来保证自己的信息只能被某些用户访问到，从而防止隐私信息泄露。一个用户所有的好友集合就是该用户的好友列表。在人人网中，用户无法对自身的好友列表进行隐私设置，同时用户无法直接访问到其他用户的好友列表，只有与用户具有好友关系或者存在共同好友的账号才能够访问到该用户的好友列表。如图 2 所示，用户 *U* 与用户 *F*1、*F*2、*F*3 具有好友关系，用户 *V* 通过与用户 *U* 建立好友关系，就能够访问到用户 *F*1、*F*2、*F*3 的好友列表。

根据这个策略，通过添加好友数较多的用户为好友即可获取大量用户的好友列表。在实验中，测试账号添加一位好友数超过 6 000 的用户为好友，然后在其好友列表中随机选取 40 位用户作为样本，通过爬虫工具直接访问“http://friend.renren.com/GetFriendList.do?id=用户 id”即可访问该用户的好友列表。这 40 位用户共获取到 14 186 位好友，其好友分布如图 3 所示。其中，好友数最少的为 59 位，最多为 1 223 位，而且分布比较均匀，因此具有一定的代表性。

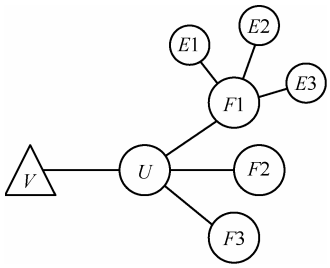

图 2　好友关系示意

4.1.2　获取好友之间关系

根据第 3 节的方法，为了在用户好友关系网络中找出社区，需要确定用户好友之间的关系。如图 2 所示，虽然用户 *V* 通过与用户 *U* 建立好友关系获取了用户 *F*1、*F*2、*F*3 的好友列表，却无法得到这些好友之间的关系，即无法确定 *F*1 的好友 *E*1、*E*2、*E*3 之间是否存在好友关系。

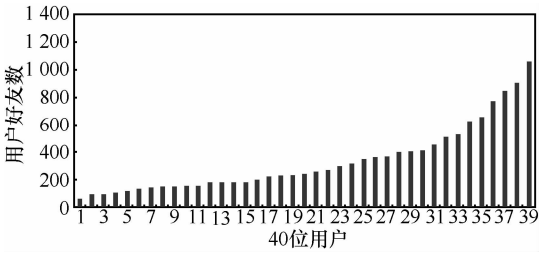

图 3　用户好友数分布

通过分析人人网的隐私策略，人人网开放平台所提供的应用程序接口中有判断 2 组用户是否互为好友关系的接口 Friends.areFriends[20]，而且调用时不需要额外的权限。为了使用此接口，作者开发了人人网应用 FriendTest1，在应用中调用 Friends.areFriends 就能够判断 2 组用户是否为好友。由于人人网开放平台在一个应用时内对应用程序接口的调用有配额限制，因此作者注册了 5 个测试账号，将查询请求分配到不同的账号及时间段。通过对 4.1.1 节中所获得的好友列表的判断，共获得 425 857 组好友关系。

4.1.3　获取用户信息

为了保护用户的隐私，社交网络为用户的个人信息提供了不同等级的隐私保护策略。以人人

网为例，隐私保护分为 4 个等级，分别为：“只有我自己可见”、“只有我的好友可见”、“好友及同城、同公司、同学校的人可见”、“所有人可见”，其中默认的隐私保护等级是“好友及同城、同公司、同学校的人可见”。为了更好地方便用户交流，人人网根据用户的个人信息将用户加入到某一个网络中。如果用户填写了具体的大学信息，那么用户就加入了这个大学的网络，与其他同样填写了该大学的用户处于同一网络。有大量用户使用社交网络默认的隐私设置[18,19]，这些用户的个人信息对同一个网络内的其他用户是公开的。为了能够访问到使用默认隐私设置的用户的信息，就需要与该用户处在同一网络，即“同城、同公司、同学校”。通过分析，利用人人网开放平台的应用程序接口 user. getProfileInfo[22]就能够获取任意用户所在的网络。

同时用户通过修改自己的个人信息就能够加入到对应的网络。因此首先利用应用程序接口 user.getProfileInfo 获得用户所在网络，然后将所有用户按照网络进行分类。通过脚本修改测试账号的网络，然后访问同一网络内的所有用户，如果用户是默认的隐私设置，测试账号就能够访问到该用户的个人信息。针对 4.1.1 节所获得的好友列表内的用户共获得了 6 446 位用户的个人教育信息，其余 7 740 位用户由于隐私设置为“只有我自己可见”或“只有我的好友可见”而不能够直接获取。

4.2 社区发现

本文使用 R 语言开发环境[23]，利用 igraph[24]分组对 4.1 节所获得的 40 个好友关系网络进行社区发现。为了排除人人网中虚假账号的影响，实验中只选取好友关系网络中度大于 3 的用户，即与同一好友关系网络中其他用户的联系大于 3 的用户。经过处理得到 13 297 位用户，好友关系 423 995 组。其中，信息公开的用户数为 6 239 位。对 40 个好友关系网络分别执行社区发现算法 multilevel.community，共得到 199 个社区。每个好友关系网络的社区数分布如图 4 所示。其中，12 个好友关系网络存在 4 个社区，12 个好友关系网络存在 5 个社区，11 个好友关系网络存在 6 个社区。每个好友关系网络的 *modularity* 值分布如图 5 所示，其中，36 个好友关系网络的 *modularity* 值大于 0.3，表明这些网络中存在很强的社区结构。

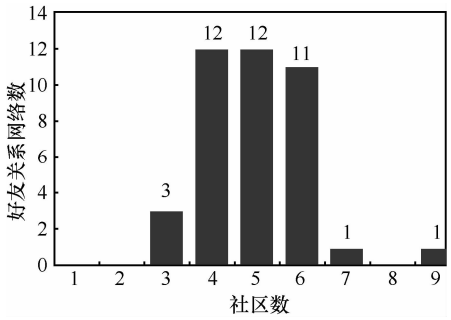

图 4 好友关系网络社区分布

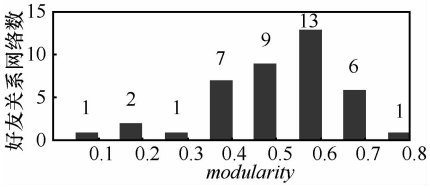

图 5 *modularity* 值分布

4.3 信息推测

通过对人人网用户特征的分析，本文主要针对用户的教育信息，即大学、高中、初中、小学这 4 个属性。由于人人网用户填写的初中和小学信息较少，所以设定阈值 θ=0.3，ε=0.5，即社区内信息公开的用户要占社区总用户的 30%以上，社区内出现频次最高的属性值占信息公开用户的 50%以上，这样才能保证所推测的信息是正确的。

在发现的 199 个社区中，其中，155 个能够满足所设定的阈值，判定为由相应的教育信息所构成的社区。表 1 列出了所推测的社区属性及结果。其中，由大学和高中属性所构成的社区最多，分别为 63 个和 74 个，这与人人网中用户的好友主要为大学和高中同学相符。这 155 个社区内的用户数为 10 147，其中，信息公开的用户数为 4 739，通过推测可得到 2 544 位用户的大学信息、2 464 位用户的高中信息、391 位用户的初中信息以及 9 位用户的小学信息。利用人人网公共主页 XSS 漏洞，通过获取这些用户的个人信息并与实验所推测的结果比较，其中，2 031 位用户的大学信息、1 949 位用户的高中信息、196 位用户的初中信息以及 6 位用户的小学信息推测正确。由于有些用户的信息填写不完整，无法判断推测结果是否正确，因此正确率的计算公式为

$$\text{正确率} = \frac{\text{正确数}}{\text{信息未公开用户数} - \text{缺失数}} \tag{13}$$

其中，大学和高中信息的正确率较高，分别为 91.57%

表 1　　好友关系网络社区分布及推测结果

					推测结果			
社区类	社区数	用户数	信息公开用户数	信息未公开用户数	正确数	错误数	缺失数	正确率
大学	63	4 543	1 999	2 544	2 031	187	326	91.57%
高中	74	4 805	2 341	2 464	1 949	152	363	92.77%
初中	17	783	392	391	196	60	135	76.56%
小学	1	16	7	9	6	0	3	100.00%
总计	155	10 147	4 739	5 408	4 182	399	827	91.29%

和 92.77%，总的推测正确率为 91.29%。

5　防御方法

根据本文所使用的方法，分别对用户以及社交网络提出相应的建议，以保护用户的个人隐私。

对于用户，可以通过提高自己的隐私设置等级来防止个人隐私信息被恶意获取并使用，如果人人网中大多数用户的隐私设置为“只有我自己可见”或“只有我的好友可见”，那么本文中提到的方法就会因样本数据太少而无法推测其他用户的隐私信息。

对于社交网络，以本文中提到的人人网为例，可以通过以下 2 个方面提高对用户隐私信息的保护。1）完善隐私保护策略，本文实验所获取的数据利用了人人网在隐私设置方面的缺陷，如好友列表无法隐藏、开放平台的接口以及用户修改个人信息无次数限制等。只要相应的缺陷不存在，就会对数据的获取带来很大的开销，使本文提到的方法在有效的时间内无法得到所需的数据。2）提高用户隐私保护意识，社交网络需要明确显示用户所选择的隐私设置效果，同时提醒用户可能的隐私泄露，并且在用户开始使用社交网络时引导用户进行相应的隐私设置，避免由于默认的隐私设置而造成用户的隐私泄露。

6　对比与讨论

正如本文第 2 节所提到的，之前有大量的工作关注于利用社交网络中用户的公开信息来推测用户未公开的隐私信息。其中一些工作[3,4]在推测用户隐私时是将单个用户的好友关系作为一个整体，通过整体的特征去推测该用户的信息。有些工作[6,18,25~27]是将整个社交网络作为一个整体，通过分析整个社交网络的特征来推测用户隐私信息。与这些工作相比，本文的工作是利用单个用户好友关系网络中的社区结构，根据同一社区内用户具有某种共同信息，通过社区内用户公开的信息来推测其他用户未公开的信息。由于本文只是利用了单个用户的好友以及这些好友之间的关系，所以可以只针对一个用户执行信息推测，而不需要获取大量的初始数据。同时作者是根据社区内用户公开的信息来推测其他用户未公开的信息，与其他工作相比，本文的工作所推测的信息更多。

由于本文主要关注教育信息，所以在实验中如果社区是由其他原因（如共同兴趣等）形成，就无法推测该社区内用户的共同信息，所以在针对人人网的实验中忽略了不是由教育信息构成的 44 个社区。同时为了防止虚假账号的影响，在进行社区发现前只选择了度大于 3 的用户。由于社区是好友关系网络自然形成的结构，因此本文所推测的信息只能确定是教育信息内的某一属性，如大学或高中等，而不能指定是哪一个属性。根据社区发现算法，网络中的用户只存在于一个社区中，本文只考虑了一个社区具有一种主要的共同信息，因此只能推测出用户的一条隐私信息。

7　结束语

社交网络作为一种新的社交平台对用户的交流沟通起着重要的作用，实名制社交网络由于其实名制以及用户数量巨大的特点，包含着大量的个人隐私信息。虽然用户通过隐私设置使其隐私信息不被公开，但是能够通过其他方式泄露这些信息。本文针对人人网所做的实验发现，只需要获取用户的好友关系网络以及其中一些用户公开的信息，就能够以较高的正确率推测出其他用户未公开的隐私信息。因此，社交网络需要完善自身的隐私保护策略，同时提高用户的隐私保护意识。

参考文献：

[1] NEWMAN M. Modularity and community structure in network[J].

The National Academy of Sciences, 2006, 103(23):8857-8582.

[2] ZHELEVA E, GETOOR E. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles[A]. Proceedings of the 18th International Conference on World Wide Web (WWW'09)[C]. Madrid, Spain, 2009. 531-540.

[3] LINDAMOOD J, HEATHERLY R, KANTARCIOGLU M. Inferring private information using social network data[A]. Proceedings of the 18th International Conference on World Wide Web (WWW'09)[C]. Madrid, Spain, 2009. 1145-1146.

[4] DEY R, TANG C, ROSS K. Estinmating age privacy leakage in online social networks[A]. Proceedings of the 31th Conference on Computer Communications IEEE(INFOCOM 2010)[C]. Orlando, Florida, USA, 2012. 2836-2840.

[5] CHAABANE A, ACS G, ALI KAAFAR M. You are what you like! Information leakage through users' interests[A]. Proceedings of the 19th Annual Network and Distributed System Security Symposium (NDSS'12)[C]. San Diego, California, USA, 2012.312-325.

[6] MISLOVE A, VISWANATH B, GUMMADIL K P. You are who you know: inferring user profiles in online social networks[A]. Proceedings of the Third ACM International Conference on Web Search and Data Mining(WSDM'10)[C]. New York, NY, USA, 2010. 251-260.

[7] RADICCHI F, CASTELLANO C. Defining and identifying communityes in networks[J]. Proc Natl Acad Sci USA, 2004, 101:2658-2663.

[8] PARK J, NEWMAN M. The origin of degree correlations in the Internet and other networks[J]. Phys Rev E, 2003, 68(2):26112-26119.

[9] NEWMAN M E J. Detecting community structure in networks[J]. Eur Phys J B, 2004, 38(2):321.

[10] SALES-PARDO M R, GUIMERA A, MOREIRA A, *et al*. Extracting the hierarchical organization of complex systems[J]. Proc Natl Acad Sci USA, 2007, 104:15224-15229.

[11] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R. Fast unfolding of community hierarchies in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 10:10008.

[12] BOCCALETTI S, LATORA V, MORENO Y. Comp lex networks: structure and dynamics[J]. Phys Rep, 2006, 424:175-308.

[13] SCOTT J. Social Network Analysis: A Handbook[M]. London: Sage Publications, 2002.

[14] FORTUNATO S. Community detection in graphs[EB/OL]. http://arxiv.org/abs/0906.0612v1, 2009.

[15] BLONDEL V D, GUILLAUME J, LAMBIOTTE R. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 10(10):108-120.

[16] TANG S, YUAN J, MAO X. Relationship classification in large scale online social networks and its impact on information propagation[A]. Proceedings of the 32th Conference on Computer Communications IEEE(INFOCOM 2011)[C]. Shanghai, China, 2011. 2291-2299.

[17] FANG L, KIM H, LEFEVRE K. A privacy recommendation wizard for users of social networking sites[A]. Proceedings of the 17th ACM Conference on Computer and Communications Security(CCS'10)[C]. Chicago, USA, 2010. 630-632.

[18] LIU Y, PGUMMADI K, KRISHNAMURTHY B. Analyzing facebook privacy settings: user expectations reality[A]. Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference (IMC'11)[C]. Berlin, Germany, 2011. 61-70.

[19] Renren[EB/OL]. http://en.wikipedia.org/wiki/Renren, 2013.

[20] Renren open platform wiki-Friends.areFriends[EB/OL]. http://wiki.dev. renren.com/ wiki/Friends.areFriends, 2013.

[21] DEY R, JELVEH Z, ROSS K. Facebook users have become much more private: a large-scale study[A]. Proceedings of the International Workshop on Security and Social Networking[C]. San Diego, California, USA, 2012. 346-352.

[22] Renren open platform wiki-Users.getProfileInfo[EB/OL]. http://wiki.dev.renren.com/wiki/Users.getProfileInfo, 2013.

[23] The R prohect for statistical computing[EB/OL]. http://www.r-project.org/, 2013.

[24] Igraph: Network analysis and visualization[EB/OL]. http://cran.r-project.org/web/packages/igraph/index.html, 2013.

[25] TANG C, ROSS K, SAXENA N. What's in a name: a study of names, gender inference, and gender behavior in facebook[A]. Proceedings of the 16th International Conference on Database Systems for Advanced Applications (DASFAA'11)[C]. Hong Kong, China, 2011. 344- 356.

[26] YANG S, LONG B, SMOLA A. Like like alike-joint friendship and interest propagation in social networks[A]. Proceedings of the 20th International Conference on World Wide Web(WWW'11)[C]. Hyderabad, India,2011. 537-546.

[27] GONG N, TALWALKAR A. Jointly predicting links and inferring attributes using a social-attribute network(SAN)[A]. The 6th SNA_KDD Workshop Proceedings[C]. Beijing, China, 2012. 76-87.

作者简介：

吕少卿（1987-），男，山西忻州人，西安电子科技大学博士生，主要研究方向为社交网络安全。

张玉清（1966-），男，陕西宝鸡人，博士，中国科学院大学教授、博士生导师，主要研究方向为网络与信息系统安全。

倪平（1988-），女，山东枣庄人，中国科学院大学硕士生，主要研究方向为 Web 安全、社交网络隐私保护。