

大数据安全与隐私保护技术初探

张 敏

中国科学院软件研究所

可信计算与信息保障实验室

mzhang@tca.iscas.ac.cn

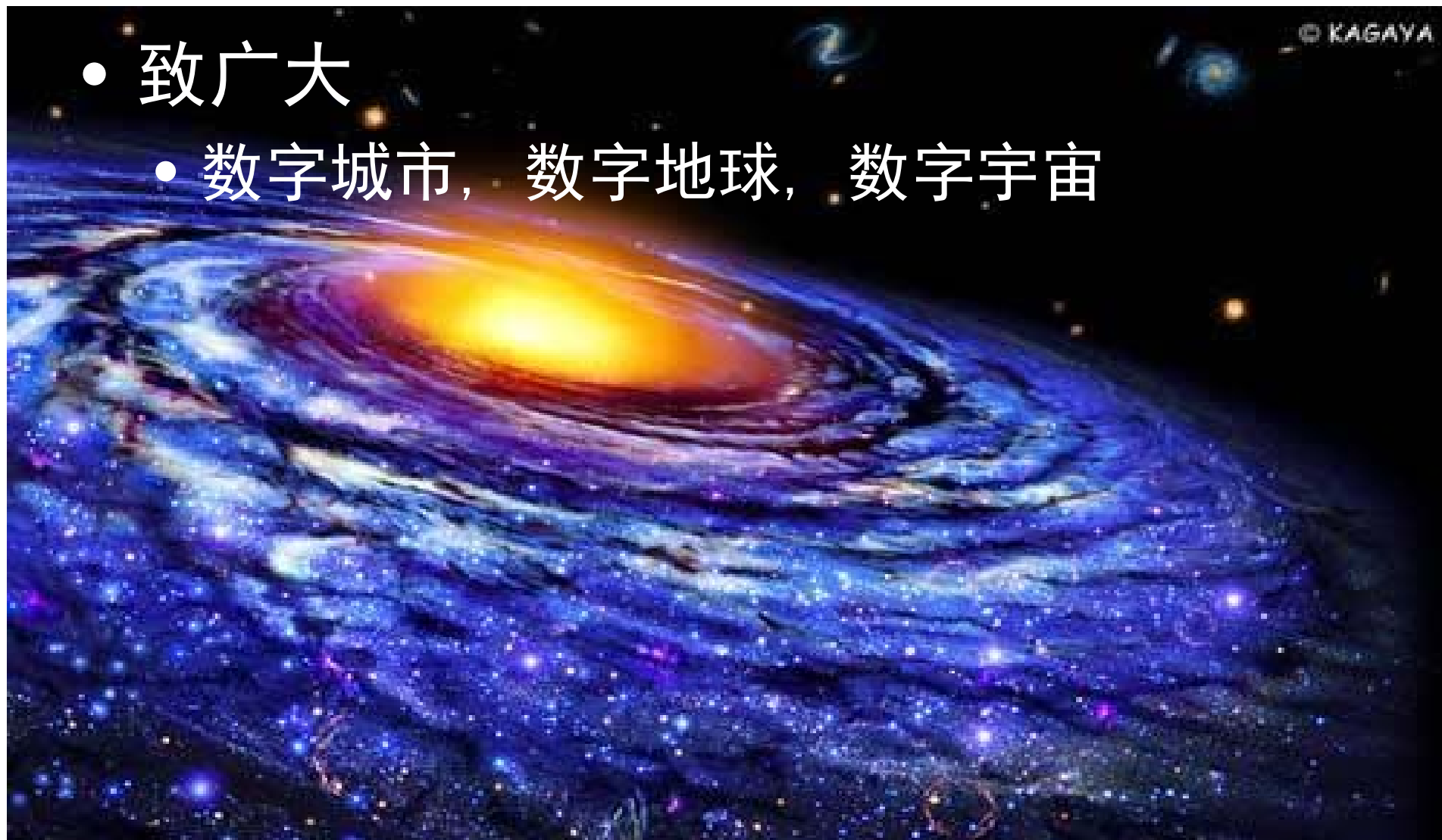


内容提要

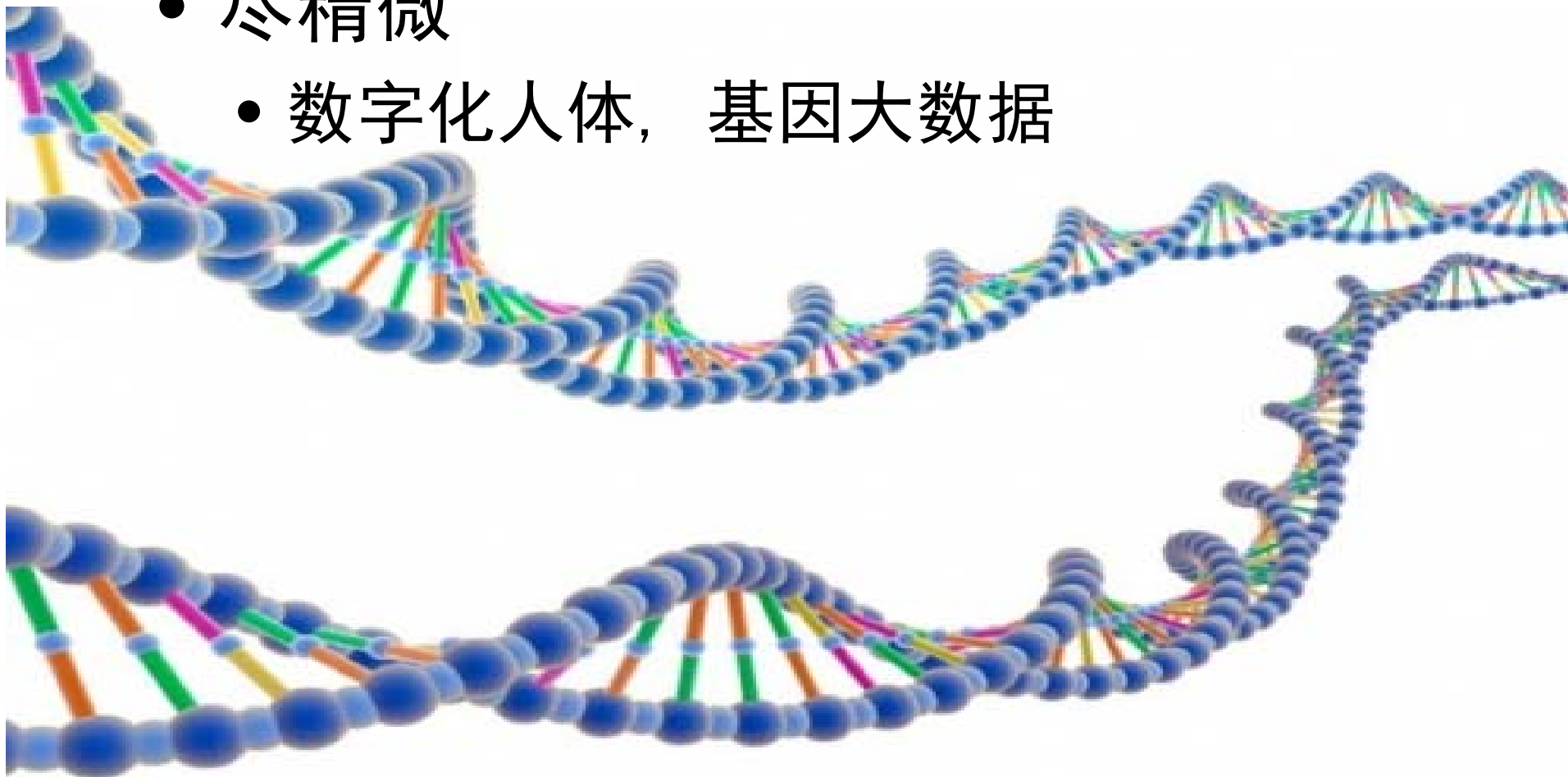
- 大数据时代
 - 一个例子
 - 隐私保护技术
 - 结束语
-

大数据时代

- 致广大
 - 数字城市，数字地球，数字宇宙



- 尽精微
 - 数字化人体，基因大数据



一个例子



四环堵死了！我联排迟到了



2010-3-6 13:46 来自彩信

光顾着看围脖留言忘记给老爸指路！都开到中关村了：（爸爸开始叨叨我说导航吧

@王培丹V：爸爸送我和小6去给<无人驾驶>配音的路上 原文转发(154) | 原文评论(310)



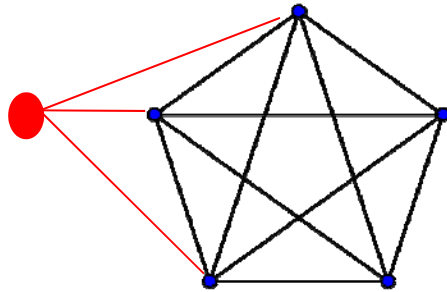
社交网络中的隐私信息



1. 身份匿名

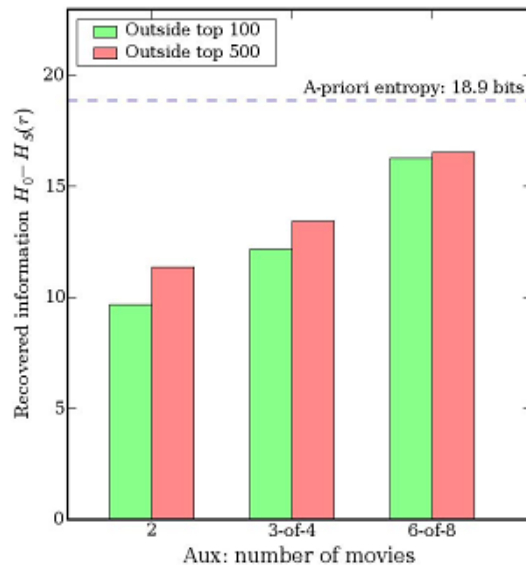


节点重识别攻击



Backstrom等人对攻击的攻击方式进行了划分，认为攻击者可通过主动或被动方式生成识别度高的社交结构，并与攻击目标连接，从而实现在匿名后的图中重新识别攻击目标的目的。

Backstrom, L. etc. 'Wherefore art thou r3579x? Anonymized social networks, hidden patterns, and structural steganography', World Wide Web 2007



Narayanan等人利用多个其他社交网络的信息作为背景知识，识别出攻击目标发布的匿名图中的某些特定节点作为种子节点，利用种子节点，进一步实现其邻居节点的识别。

Narayanan A, Shmatikov V. Robust De-anonymization of Large Sparse Datasets 2008

De-anonymizing social networks. Security and Privacy, Security and Privacy, 2009

去匿名化De-anonymity



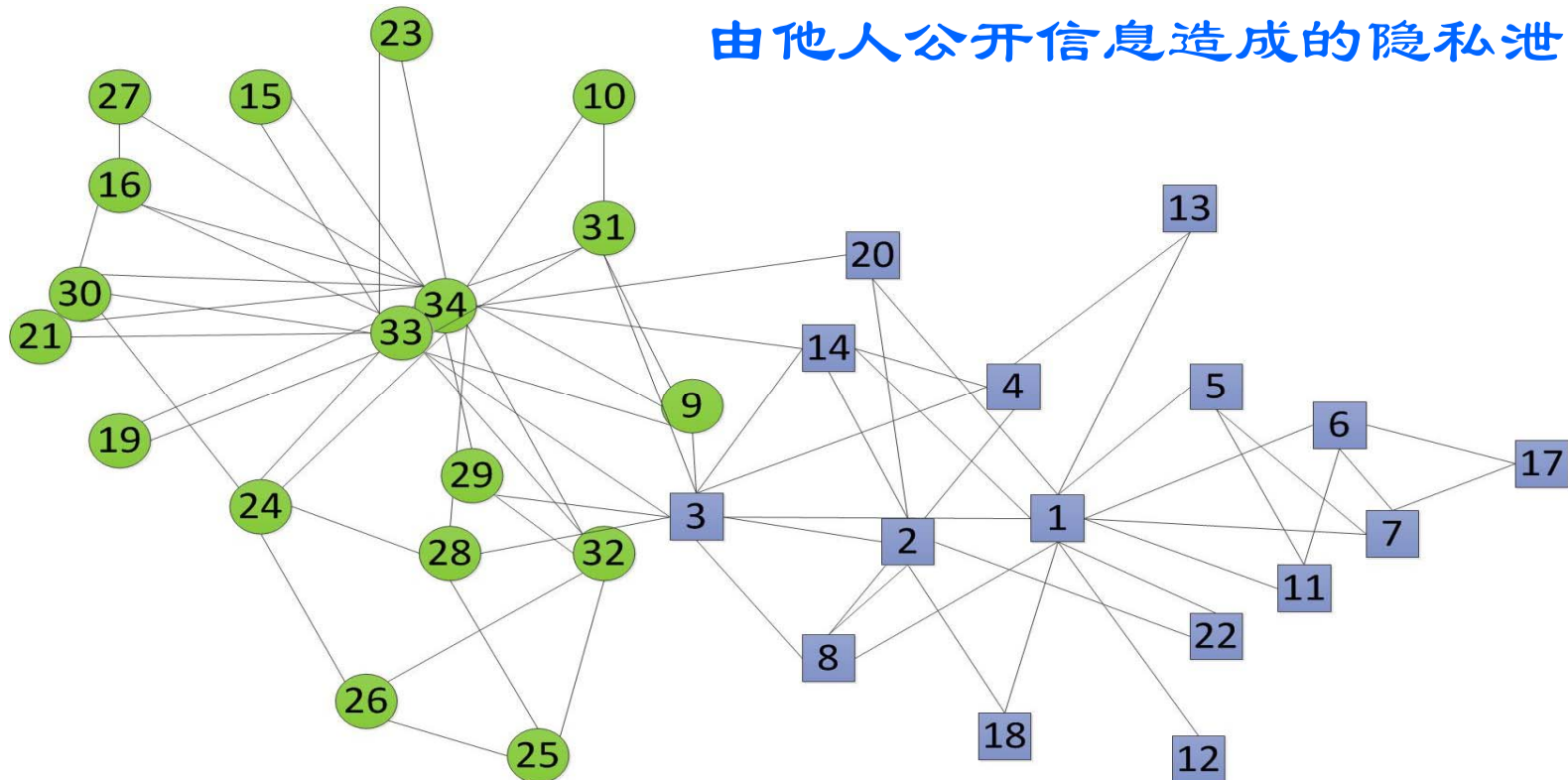
- 基于特定模式精确匹配
- 基于种子匹配
- 基于相似度的匹配

IMDB影评库

2. 属性匿名推测

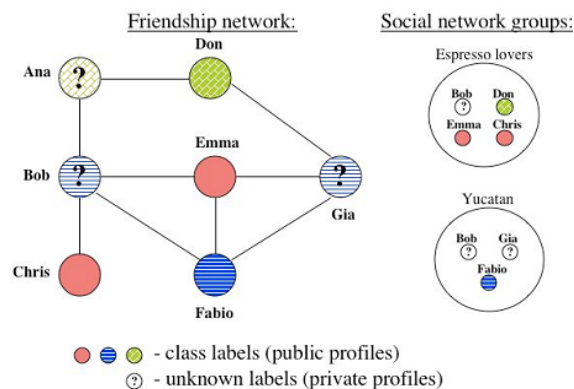
群组倾向性预测

由他人公开信息造成的隐私泄露



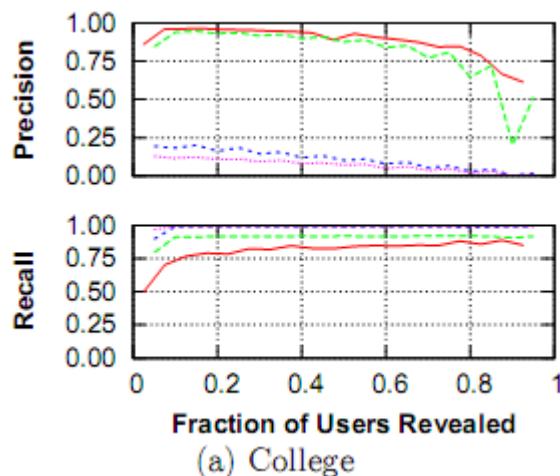
W. W. Zachary, An information flow model for conflict and fission in small groups, Journal of Anthropological Research

属性重识别攻击



Zheleva等人研究发现，参与同一小组的用户倾向于具有相似的属性。并可利用用户的群组标签对用户可能具有的属性进行预测。

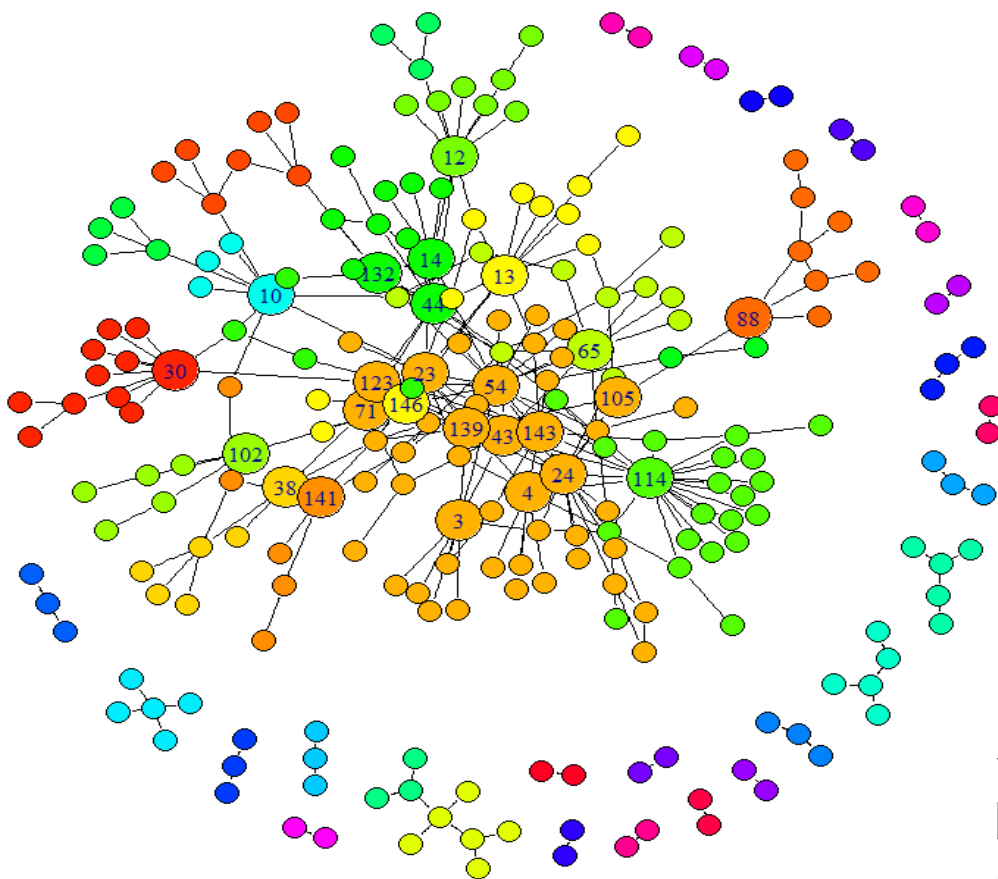
Zheleva E, Getoor L. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. WWW 2009



Mislove等人研究发现，用户可能与其好友具有类似的属性。可以通过好友的公开信息对用户未公开的信息进行推测。

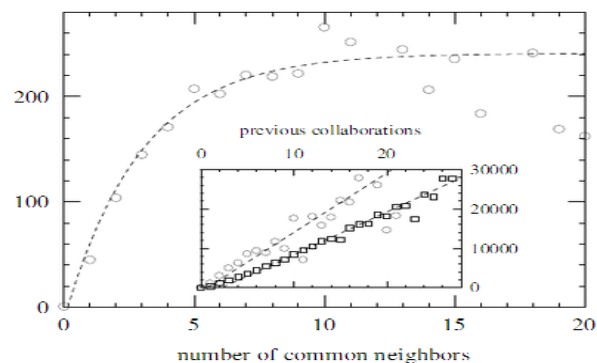
Mislove A, Viswanath B, Gummadi K P, et al. You are who you know: inferring user profiles in online social networks. In Proceedings of the third ACM international conference on Web search and data mining. ACM, 2010: 251-260

3. 关系匿名推测



边匿名猜测攻击： 社交网络中群组的存在，使得用户之间的匿名联系仍有可能推测出来。简单边匿名、随机边匿名方案匿名效果不理想，可用性差。

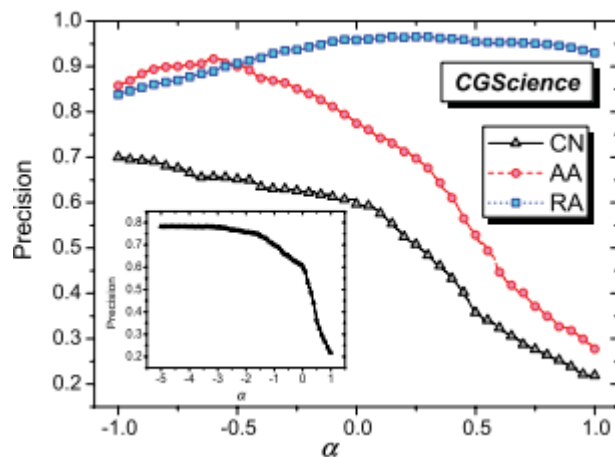
连接关系重识别攻击



Newman等人发现，两个用户间的共同朋友越多，两者间具有连接关系的可能性越大。提出根据共同朋友预测连接关系的模型

Adamic等人分析了节点间共同朋友的度数与节点间建立连接可行性之间的关系，Zhou等人建立了资源分配模型对节点间信息流动进行分析，并预测连接关系。


Zhou T.etc. Predicting missing links via local informatio]. The European Physical Journal B, 2009, 71(4): 623–630.



Zhou等人发现，在某些社交网络图中，与共同朋友间具有弱连接的两端，更容易形成朋友关系。据此提出了基于弱连接的朋友关系预测。

Lü L, Zhou T. Link prediction in weighted networks: The role of weak ties[J]. EPL (Europhysics Letters), 2010, 89(1): 18001.

4. 位置隐私

@杨承平 : 【深圳高二女生外出遇害 疑为玩微博定位惹祸上身!】宝安职业技术学校高二女生赖曾裕童@Mysshi 12日晚失踪, 13日尸体被发现。鉴定系他杀, 警方正全力侦破此案。她之前经常在微博晒自拍照和定位, 在此提醒广大网友, 在微博上保护自己的隐私, 避免过多定点自拍, 女性尤甚, 转发周知@袁裕来律师 @何兵

📌 收起 | 🗺️ 查看大图 | ↶ 向左转 | ↷ 向右转

复习📖📝 @ 港湾茶餐厅 <http://t.cn/zjBysET>

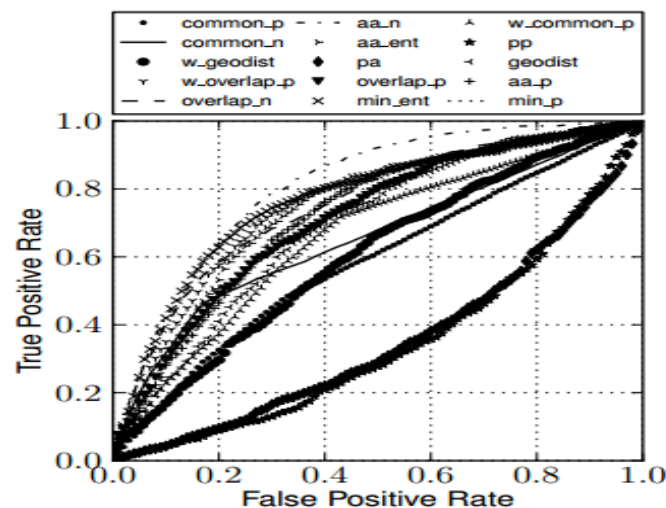


📍 广东省·深圳市·宝安区·电信路 - 显示地图

1月11日 15:33 来自Instagram | 举报



位置-社交关系攻击



(c) Place-social

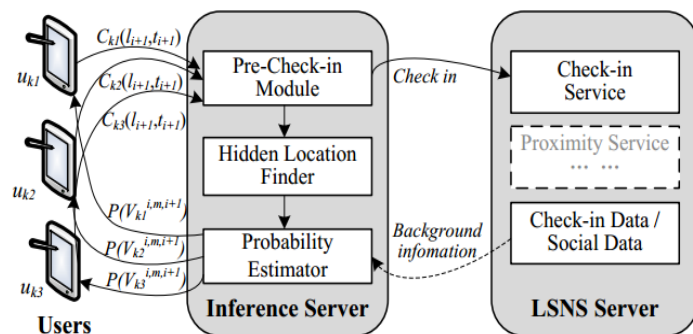


Fig. 4. System Architecture

Salvatore Scellato等人在社交网络中使用社交关系和用户的签到历史信息，对稀疏的预测空间进行压缩，通过机器学习的方法对用户的社交关系进行预测，并取得了良好的预测效果。

Exploiting place features in link prediction on location-based social networks 2011

作者Huo Zheng等人提出一种安全的社交网络签到系统框架，该系统以用户之间的社交关系、用户的签到历史轨迹信息以及地理位置信息为背景，对用户可能去过的位置进行预测，并将预测的位置反馈给用户，询问用户是否真实签到。

Feel Free to Check-in_Privacy alert against Hidden Location Inference Attacks in GeoSNs 2013

结束语

- 当前，用户使用社交网络时，难以避免个人隐私泄露威胁
 - 需要从国家与社会层面限定互联网企业对用户隐私信息的收集与使用，从根源上解决问题
 - 用户隐私使用三原则
 - 用户是隐私的所有者
 - 服务商承诺用户信息传输与存储安全
 - 双方公平交换
 - 从技术角度实现大数据隐私保护十分必要
 - 匿名保护
 - 访问控制
 - 等等
-

谢 谢!
