摘要

近年来，随着互联网经济的发展和在线社交平台的日益普及，用户数量和商业平台所提供的产品数量都呈现了指数级的增长。推荐系统在商业中得到了越来越普遍的应用。商业平台通过推荐系统帮助用户发现机遇自身兴趣的个性化物品；部分互联网商业平台,如豆瓣和Netflix，依赖推荐系统的准确度以最大化用户的粘性和满意度。由于优异的性能，协同过滤非常广泛地被运用在计算推荐中, 并在此基础上发展出了众多社会化推荐系统。但是，由于评分系统和社交网络的开放性，以协同过滤为代表的推荐非常容易受到恶意攻击。在一种被称为托攻击(Shilling Attack)的攻击模式下，推荐系统会出现偏差，从而产生巨大的服务风险。

针对社交网络的服务风险，本文利用多维度的方法对潜在的风险展开定量和定性的评估。本文从社交网络分析的角度出发，进行了原始数据集的风险评估；在社交网络分析的基础上，本文对托攻击实现了完整的模拟和仿真，并在社会化推荐系统上进行风险评估；此外，本文还从托攻击检测效率的角度调研了推荐系统的服务风险。本文的主要贡献包括：

1. 通过社交网络分析对社交网络数据集进行预处理和可视化；
2. 使用了三种经典的托攻击算法对FilmTrust数据集进行了模拟攻击，并在社交推荐系统上进行验证
3. 设计托攻击检测算法，从均方根误差(RMSE)以及系统检测托攻击能力等方面评估推荐系统的服务风险

关键词：社交网络，协同过滤，社会化推荐系统，托攻击，风险量化

Abstract

With the development of the network economy and the growing popularity of social platforms these years, the number of users and products provided by business platforms has grown exponentially. Recommender systems are increasingly gaining popularity. Business platforms utilize recommender systems to help users identify personalized products based on their own interests; Some platforms including Facebook and Netflix, rely on the accuracy of recommender systems to maximize user stickiness and satisfaction. Collaborative filtering is a widely used method for computing recommendations due to its good performance. However, because of the open nature of rating systems and social networks, the social recommender systems are susceptible to malicious attacks. Under shilling attack, the recommendation service could be biased and lead to enormous service risks.

Aiming at service risks of social networks, this article uses multi-dimensional approaches to carry out quantitative and qualitative assessments of potential risks. By deploying Social Network Analysis (SNA), this paper conducts risk evaluation of the original data set; based on SNA, this paper implements complete simulations of shilling attack methods and analyzes attack performance under various social recommender systems. In addition, this article also investigates the service risk of the recommendation system from the perspective of the detention of shilling attacks. The main contributions of this article include:

1. Preprocessed and visualized social network dataset using social network analysis;
2. Three classic shilling attack algorithms were used to attack the FilmTrust data set, and their performance was assessed under social recommender systems;
3. Evaluated the service risk of the social recommendation system in terms of RMSE and other metrics to check systems' ability to detect shilling attacks

Keywords: Social Network, Collaborative Filtering, Social Recommender Systems, Shilling Attack, Risk Management

第一章 绪论

1.1 研究背景和意义

互联网的数字信息和用户数量的爆炸性增长造成了信息过载这一潜在挑战。信息检索系统在一定程度上解决了这个问题，但是针对用户的个性化信息供给仍然缺位。为了解决这个问题，推荐系统应运而生。推荐系统是一种信息过滤系统，它通过根据用户的喜好，兴趣或观察到的针对项目的行为，从大量动态生成的信息中过滤出重要信息片段，从而解决了信息过载的问题的[1]。推荐系统可以根据用户的档案预测是否某个特定的用户会对特定的物品产生兴趣。推荐系统对于服务提供商以及用户都有很大的到价值[2]。推荐系统可以降低寻找和选择物品的成本，并能极大地改善决策的过程和质量。商业平台通过推荐系统帮助用户发现机遇自身兴趣的个性化物品；部分平台, 如豆瓣和Netflix，依赖推荐系统的准确度以最大化用户的粘性和满意度。推荐系统的成功应用是提高商业平台盈利水平的有力手段[2]。因此，一个能够为用户提供准确并值得依赖的推荐的推荐系统具有很大的商业价值。

通常情况下，推荐系统以协同过滤为基础，该技术可根据两个用户的评级概况计算相似度，并推荐由相似用户高度评价的项目[4]。协同过滤被凭借它优良的效率被广泛地运用在以Amazon为代表的电子商务系统中。但是，协同过滤技术也有明显的缺点——它需要足够的历史评价来计算用户之间的相似度。大多数用户只会对可达上百万的物品中的极少数进行评分，共同评价的缺失可能会降低推荐系统的准确度[5]。随着在线社交平台的愈发流行，引入了社交网络因素的推荐算法也应运而生。这些方法在推荐中合并了评分信息和社交关系信息，并假设用户的倾向会在一定程度上会被其他与该用户建立社交关系的用户所影响。先前的研究已经证明结合了社交信息的推荐系统可以做出更加准确的推荐，并且针对冷启动现象特别有效[6]。

社会化推荐系统可以被看作传统的基于评分的推荐系统与社交网络服务的结合。它们都极大地依赖于用户概貌(user profile)。但是，由于评分系统和社交网络的开放性，以协同过滤为代表的推荐非常容易受到恶意攻击。在推荐系统中，恶意攻击者可以通过注入虚假的评分信息和用户关系信息，以实现攻击推荐系统的目的。这种攻击被称作托攻击(shilling attack)[7]。在面向社交网络的托攻击下，推荐系统服务会出现偏差，面临着极大的服务风险。潜在的托攻击影响了商业平台的安全性，使得用户对推荐结果出现不信任，并对用户和提供商双方的利益造成不可估量的损失。综上所述，针对该具体情境进行服务风险评估，并为商业平台服务商提供评估结果和解决方案，具有巨大的商业价值。

针对推荐系统的托攻击和相应的检测技术在近年引发了很多关注，但是很少有相关研究涉及到社交网络分析技术；此外，先前的研究大多基于传统的推荐系统，面向社会化推荐系统的托攻击仍然是一个相对比较陌生的领域。因此，对此领域展开研究具有较大的学术价值与强烈的现实意义。

1.2 国内外相关工作

1.2.1 社交网络分析相关研究

社交网络作为一种重要的web服务有着广泛的应用，如协同工作、协同服务质量评级、资源共享和发现新朋友[10]。社交网络的概念不仅停留在理论方面，而且进入了实践领域。只要可以描述和分析用户的关系，就可以找到并定义社交网络的应用和在线服务。网络理论是关于结点之间关系表示的研究[11]。社交网络是由一组连接参与者的特定联系形成的网络。社会网络理论表明，参与者在关系网中的位置会影响他们对资源，朋友和信息的访问[12]。

近年来，越来越多的研究将社交网络分析理论引入了商业决策与商业推荐中。先前已经有许多关于社交网络分析的应用。例如，DeMeo 等[13]开发了一个基于社交网络分析的架构来推荐相似的用户和资源；Zhen et 等人[14]应用了社交网络的概念开发了可以进行端到端知识分享的推荐系统。

商业平台通过研究如何利用社交关系来改善客户的购买决策，从而增加销量[15]。与其他用户有紧密的社交关系的用户被认为能够对他人施加更大的影响力[16]。在先前的研究中，Carchiolo 等人[17]发现社交网络中朋友和朋友的朋友之间的关系在涉及到信任与可靠信息时有关键的作用。Albert和Barabasi [18]指出，社交网络是一种复杂的网络，以社交实体作为节点，以链接显示这种关系。 社交网络的本质更多地侧重于构成网络的组件之间的关系，而不是其自身的结构。评估社会关系的亲密性，可以得出社

会网络中社会节点的等级或分数，以表示影响力或信任的强度[19]。在实际应用中，Wang和Chiu [20]结合社会亲密性和社会声誉来发现可信赖的在线拍卖卖家。 在针对目标群体的广告投放的研究中，Kempe 等人[21]的研究表明，具有较高亲密关系的人们传播的信息将对网络中的其他节点产生更大的影响。

1.2.2 基于传统推荐系统的托攻击及检测研究现状

托攻击中的攻击者借助不同的攻击手段，生成攻击概貌，伪装为系统中的真实用户向推荐系统注入不实信息，以达到提高或降低目标项目被系统推荐的概率，进而扰乱系统的推荐结果。“托攻击”的概念最早由Riedl和Lam提出，他们引入最初的两种攻击模式，随机攻击和均值攻击[7]，并在基于用户的协同过滤与基于项目的协同过滤中加以实现。后来的研究者提出了新的攻击模型。Mobasher等人提出了流行攻击和分段攻击[Mobasher B, Burke R, Williams C, et al. Analysis and detection of segment-focused attacks against collaborative recommendation[C]//International Workshop on Knowledge Discovery on the Web. New York, USA, 2005: 96-118.]；O’Mahony, Smyth等研究者以及Feng等研究者分别提出了爱憎攻击[O’Mahony M P, Smyth B. Collaborative web search: a robustness analysis[J]. Artificial Intelligence Review, 2007, 28(1): 69-86.]和探测攻击[Feng Q,Liu L, Dai Y. Vulnerabilities and countermeasures in context-aware social rating services [J]. ACM Transactions on Internet Technology (TOIT), 2012, 11(3): 11.]。对于国内外研究人员来说，研究的重点在于检测出托攻击概貌，并将虚假的概貌从推荐系统中识别出来以保证系统的正常运作。托攻击的检测方法根据是否需要先验知识和训练样本进行区分，分为无监督方法、半监督方法和有监督方法。

有监督的检测方法依赖大量的训练数据进行训练，提取出托攻击攻击者的概貌特征实现托攻击的检测。Chirita[白杨14]等人在研究中使用了RDMA和WDMA的特征去区分系统中的用户是否为托攻击用户；Williams[白杨15]等人将用户特征进行了分型，定义了6种与攻击类型无关的通用特征与7种描述已知攻击类型的专用特征，并在此基础上利用K-近邻学习、决策树和SVM等有监督学习方法对用户概貌进行分类。该方法对特定攻击类型有出类拔萃的效果，但在其他攻击下容易造成误判；Zhou等人[白杨18]等针对AOP攻击提出了一种攻击检测方法，通过IF-IDF方法提取AOP攻击生成的AOP攻击下的托攻击概貌，并使用了支持向量机进行分类与检测；Zhang等人[白杨19]从用户行为学的角度出发，针对异常用户打分模式的差异，通过引入互信息知识，构建区别真实用户和异常用户的的分类特征，并运用在决策树分类器下。该方法在较大的填充规模下有较高的准确率。

由于样本标签获得较困难，无监督学习方法在检测托攻击中具有较大优势。Bhaumik[白杨21]等研究人员通过K-means聚类将系统中的用户分为两簇，并判定用户数较少的一簇为托攻击攻击者。这种方法具有高适用性，适用于所有的攻击类型，但误判率高；Mehta等人[白杨22]发现攻击概貌具有高相关性，并在此基础上基于PCA（主成分分析），利用Z-score与协方差矩阵计算用户概貌的主成分系数得分，进而过滤出可疑的用户概貌。

1.2.3 基于社会化推荐系统的托攻击及检测研究现状

与针对传统推荐系统的托攻击相比，面向社会化推荐系统的托攻击还会与真实用户建立社交关系以达到虚拟的信息传播效果。Junliang Yu[25]等人在实验中设计了多重指标，验证了结合社交信息的混合托攻击模式在社会化推荐系统上的可行性和有效性；在托攻击检测方面，Yishu Xu[Detecting shilling attacks in social recommender systems based on time series analysis and trust features]等人提出了一种利用时间序列和信任特征的社会化托攻击检测方法；程[白杨27]等人基于微博用户的时间戳、关系图等社交属性特征，开发了一种使用机器学习算法区分微博水军的托攻击检测方法。

1.3 论文内容及结构

本文主要研究协同过滤推荐系统在托攻击下所承受的服务风险和攻击下推荐系统的稳健型。通过社交网络分析模型将社交关系信息结合入协同过滤推荐系统。主要的工作包括：

1. 通过社交网络分析对社交网络数据集进行预处理和可视化；
2. 使用了三种经典的托攻击算法对FilmTrust数据集进行了模拟攻击，并在社交推荐系统上进行验证
3. 设计托攻击检测算法，从均方根误差(RMSE)以及系统检测托攻击能力等方面评估推荐系统的服务风险

本文可分为五个章节，具体的组织结构如下：
第一章绪论。主要介绍了社交网络服务风险评估研究工作的背景和意义，简要论述了国内外相关工作的进展，介绍了本文的内容和组织结构。
第二章预备知识。主要介绍了本文涉及到的相关概念和定义，包括社交网络分析、推荐系统和托攻击的概念。
第三章模型与算法。总结了随机攻击(Random Attack)、普通攻击(Random Attack)和流行攻击(Bandwagon Attack)三种经典的托攻击模型，介绍了社交偏移的算法和与用户-项目评分矩阵结合的方法；引入了服务风险评价指标。
第四章实验过程与结果。对FilmTrust数据集进行了模拟攻击，从均方根误差(RMSE)以及系统检测托攻击能力等方面评估社交推荐系统的服务风险。
第五章总结与展望。首先对本文的工作进行总结，然后分析攻击和评估算法中存在的一些问题，对以后的研究工作做出展望。

第二章 预备知识

2.1 社交网络分析

社交网络分析(Social Network Analysis,SNA)是网络科学在社交网络中的应用。其中，社会现象是由重叠的二元组数据作为观察单位来表示和研究的(Brandes et al.,2013c)。因此，作为一种直接、方便的数学表示，图(Graph)是社交网络分析的基础[8]。社交网络分析在微观角度研究个体的行为，在宏观的角度研究关系(网络结构)的模式，以及上述两者的交互[9]。社交网络为个人选择提供并限制了机会，而与此同时，个人发起，构建，维持和打破关系，并由此确定了网络的整体结构。哪些网络结构或位置创造强大的机会，或者约束，取决于所研究关系的工具性价值。社会资本是由社会关系创造的机会结构。在社交网络分析中，已经开发出许多措施来表征和比较网络结构和网络位置。根据决定机会结构差异的因素，分析可以侧重于中心性差异、对紧密联系集群的调查、在网络中结构性等效的位置等。

2.1.1 中心度

在图分析中，中心度[22]是识别图中重要节点的非常重要的概念。 它用于测量图中各个节点的重要性（或“中心”，如节点在图中的“中心”程度。取决于如何定义“重要性”，每个节点都可能很重要。中心度具有不同的衡量标准，每种标准或度量都从不同的角度定义节点的重要性，并进一步提供有关图及其节点的相关分析信息。

点度中心度

对于图中的有权和有向网络，一个节点的入度被定义为直接连接该节点的边数，出度则正好相反。最具影响力的节点是具有最大入度的节点。d_i指的是节点v_i的度(邻接节点的个数)。

$$C_d(v_i) = d_i^{in} \tag{1}$$

接近中心度

接近中心性度量节点到所有其他节点的平均最短距离。它反映了网络结构的整体连接性。$d(i,j)$定义了两个节点i与j之间的最短距离，n为节点数。

$$C_c(i) = \frac{n-1}{\sigma_j^n - 1 d(i,j)} \tag{2}$$

中介中心度

中介中心性指的是一个节点担任其他两个节点之间最短路的桥梁的次数。一个节点充当“中介”的次数越高，它的中介中心度就越大。公式中，p_{jk}表示两个节点j与k之间最短路径的条数，$p_{jk}(i)$表示节点j与k之间包含节点i的最短路径数。

$$C_b(i) = \sigma_{j<k} \frac{P_{jk}(i)}{P_{jk}} \tag{3}$$

2.2 推荐系统

推荐系统帮助用户从网络中的大量数据中过滤出有用的信息。推荐系统同时帮助商业平台对目标用户售卖商品。简而言之，推荐系统是通过各种各样的的统计和机器学习模型去预测用户给陌生项目的评分实现的，而后在预测的评分或倾向性的基础上特定的商品将被推荐给用户。因此推荐系统成为了一种可以帮助终端用户决定购买什么、帮助组织将产品呈现给合适的用户的利器。

2.2.1 协同过滤

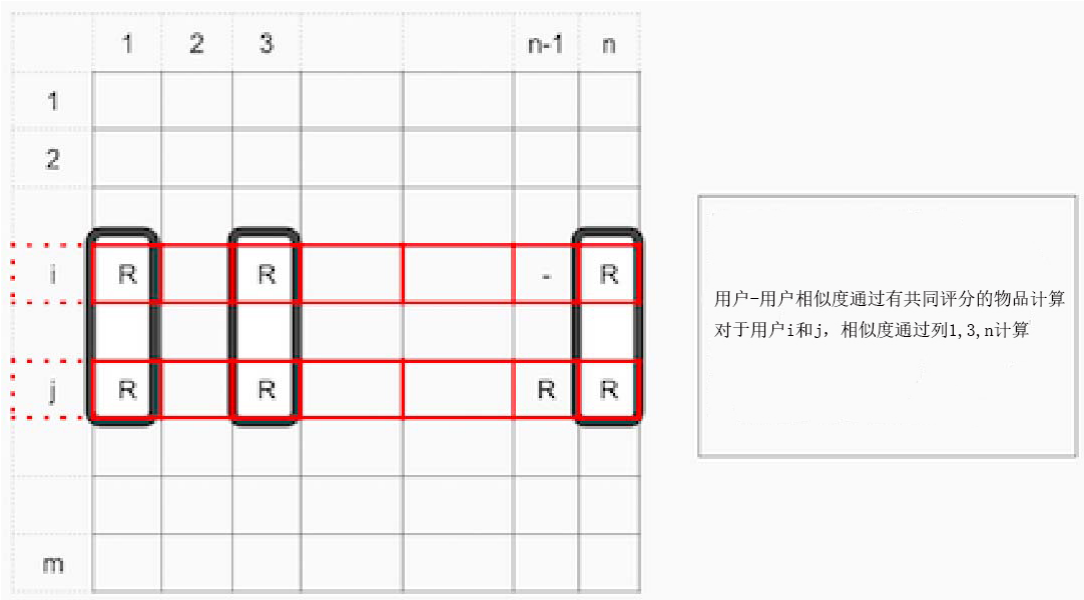
协同过滤技术[23]通过计算用户之间或项目的相似度去预测一个项目的评分。协同过滤有两种形式：基于用户的协同过滤；基于项目的协同过滤：

基于用户的协同过滤

在基于用户的协同过滤中，通过用户-项目评分矩阵计算两个用户的相似度。“用户1”与“用户2”的相似度通过余弦距离进行计算，其中考虑n个被两个用户都评分过的项目。当计算出相似度后，我们可以根据计算出的相似度来预测“用户1”对已由“用户2”评分的项目的评分。通过对数据库中所有的用户都进行相同的操作，可以得到用户k对特定物品m的评分为：

$$x_{m,k} = \overline{x_k} + \frac{\sum_{u_a} sim(u_k, u_a)(x_{a,m} - \overline{x_a})}{\sum_{u_a} |sim(u_k, u_a)|} \tag{4}$$

其中$\overline{x_k}$代表用户k的评分打分，用于归一化。它定义了一个特定的用户如何对打分项目1-5的范围中，有的用户会对不喜欢的项目打3分，而在个别情况下有的用户会打1分。因此计算$\overline{x_k}$对预测每个不同用户的评分非常有必要。第二部分的分母同样用于归一化，将预测得到的评分控制在1-5。

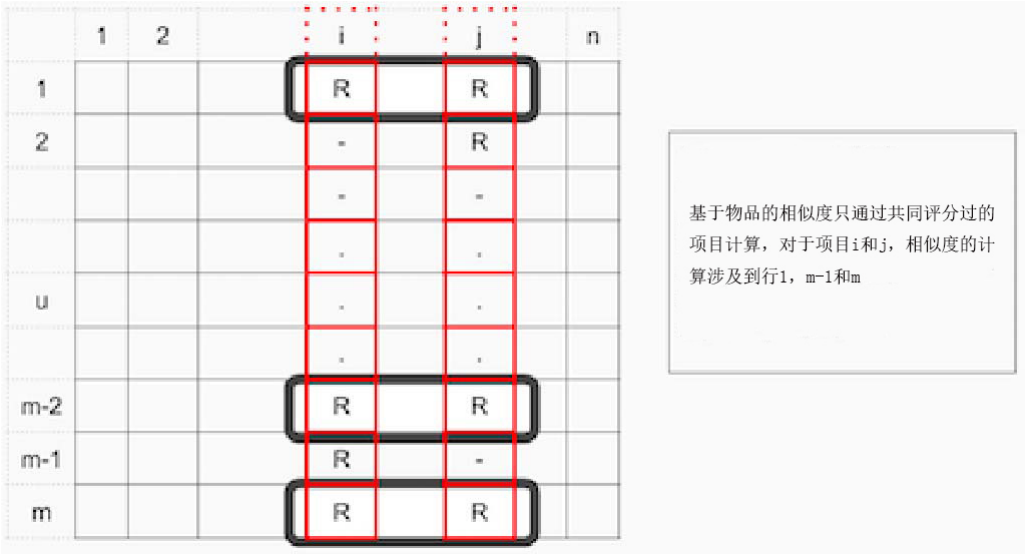


基于项目的协同过滤

基于项目的协同过滤计算两个项目的相似度。两个物品，“项目1”和“项目2”之间的相似度通过对n个均评价过这两个项目的用户计算余弦相似度实现。如果一个用户喜欢“项目1”，且没有见过“项目2”，并且“项目1”和“项目2”有较高的相似度，那么“项目2”就会被推荐给“项目1”。用户k对特定项目m的评分为：

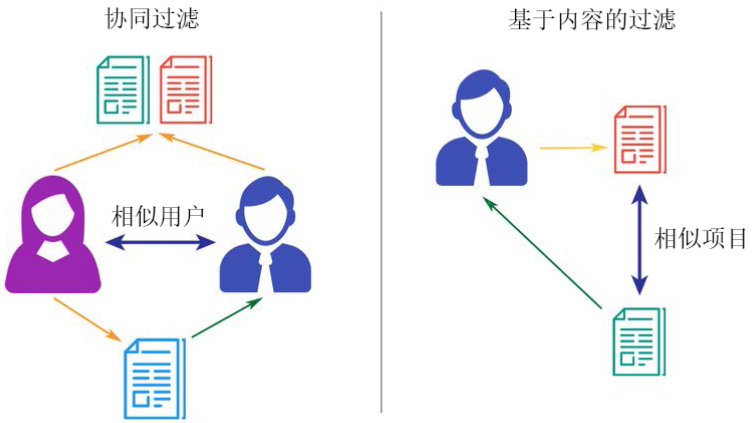
$$x_{m,k} = \frac{\sum_{i_b} sim(i_m, i_b)(x_{k,b})}{\sum_{i_b} |sim(i_m, i_b)|} \tag{5}$$

在这个公式中，计算项目m与数据库中的每个项目b的相似度，然后根据用户k对项目b的评分，计算项目m的评分。分母用作归一化因子，将预测的评分控制在1-5的范围内。



2.2.2 基于内容的过滤

基于内容的过滤模型通过浏览用户以前喜欢的项目，根据每个用户的“口味”推荐用户。在该方法中，每个用户会根据先前评价过的项目类型创建一个单独的profile。基于内容的推荐系统通常会要求用户填写一个调查问卷来了解用户的兴趣，从而消除冷启动问题。在协同过滤推荐系统中，当新用户进入系统时，由于该特定用户的评分矩阵内缺少必要的资料，因此无法计算相似度，从而导致推荐准确度降低。但是，基于内容的推荐系统可能面临过于推荐专一的问题。在实践中，协同过滤比基于内容的过滤有更好的性能。[23]



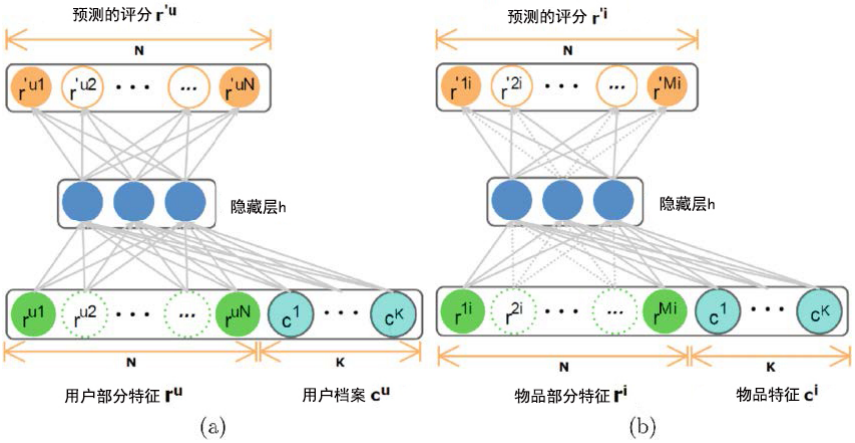
2.2.3 基于模型的过滤

基于模型的过滤利用用户评分矩阵去寻找特定用户有可能感兴趣的项目的潜在的或隐藏的特征。该方法同样计算用户相似度以避免推荐过度专一的问题。为了实现这个方法，基于模型的过滤使用了矩阵分解技术，如SVD(奇异值分解)。SVD的公式如下：

$$X = USV^T \tag{6}$$

对于一个给定的$m \times m$矩阵X, U是一个$m \times r$的正交矩阵；V是一个$m \times r$的对角矩阵，且对角线都是非负实数；V^T是一个$r \times n$的正交矩阵。

S对角线上的元素被称为X的奇异值。矩阵X可以被分解为U,S和V。矩阵U的行为用户，列为隐藏特征向量；矩阵V的行为隐藏特征向量，列为项目。这些隐藏特征向量实际上是特征向量，X中的奇异值是对应特征向量的特征值。模型的预测通过计算U,S和V^T的乘积实现。



2.3 社会化推荐

由于协同过滤被广泛地运用在推荐系统中，大多数现存的社会化推荐系统都基于协同过滤技术。社会化推荐系统有两个输入，如评分信息和社交关系信息。现有的大多数社交推荐系统都将协同过滤模型作为基本模型来构建系统，并根据社交网络分析的结果，提出捕获社交信息的方法[29]。因此，大多数社交推荐系统可以被表述为：

$$\text{社交推荐协同过滤模型} = \text{基础协同过滤模型} + \text{社交信息模型} \tag{7}$$

社交推荐协同过滤模型中的基本系统过滤模型创建了一种对社交推荐系统进行分类的方法。 在基于协同过滤的推荐系统分类之后，可以根据其基本的协同过滤模型将社交推荐系统分为两大类：基于记忆(memory-based)的社交推荐系统和基于模型(model-based)的社交推荐系统。

在本文中，使用基于模型的社会化推荐系统去评估托攻击造成的影响。与传统的基于矩阵分解的推荐系统不同，社会化推荐系统需要获取用户的社交信息，并作为算法的输入。基于模型的社交推荐系统的统一最优目标函数被定义为：

$$min_{U,V,\Omega} \parallel (R - U^T V) \parallel_F^2 + \alpha Social(T, S, \Omega) + \lambda(\parallel U \parallel_F^2 + \parallel V \parallel_F^2 + \parallel \Omega \parallel_F^2) \tag{8}$$

其中$R \in \mathbb{R}^{m \times n}$是用户-物品评分矩阵，$U \in \mathbb{R}^{k \times m}$是用户隐含特征矩阵，$V \in \mathbb{R}^{k \times n}$是项目隐含特征矩阵，$\lambda$为惩罚因子。$Social(T, S, \Omega)$被用来引入社交网络中的社交信息。$T \in \mathbb{R}^{m \times m}$为原始用户-用户关系矩阵，$S \in \mathbb{R}^{m \times m}$为归一化的或转置的关系矩阵，$\Omega$为从社交信息中学习到的参数的集合，$\alpha$被用来控制社交性的影响。根据对$Social(T, S, \Omega)$的不同定义，基于模型的社交推荐方法可以被划分为三类：协分解（co-factorization）方法、集成（ensemble）方法与正则化（regularization）方法[J. Tang, X. Hu, H. Liu, Social recommendation: a review, Soc. Netw. Anal. Min. 3 (4) (2013) 1113–1133.]

2.3.1 协分解方法

协同分解方法属于基于模型的社交推荐系统。该组系统的基本假设是，第i个用户u_i应当在评分空间(评分信息)和社交空间(社交信息)中共享相同的用户偏好向量u_i。该组中的社交推荐系统通过共享相同的用户偏好隐因子，在用户-项目矩阵和用户-用户矩阵中执行协分解。如果$Z \in \mathbb{R}^{k \times m}$是特定要素隐含特征矩阵，那么$R = U^T V$且$T = U^T Z$。SoRec[Hybrid18]是协分解方法的代表方法，该方法的优化问题可以被定义为：

$$min_{U,V,\Omega} \parallel (R - U^T V) \parallel_F^2 + \alpha \parallel (S - U^T Z) \parallel_F^2 + \lambda(\parallel U \parallel_F^2 + \parallel V \parallel_F^2 + \parallel \Omega \parallel_F^2) \tag{9}$$

其中$Social(T, S, \Omega)$被定义为$\|(S - U^T Z)\|_F^2$

2.3.2 集成方法

集成方法的基本假设为用户的倾向由用户和朋友的品味决定。因此，一个给定用户的缺失评分最终被预测为当前用户和他朋友的评分的线性组合。集成法方法的代表算法是RSTE[Hybrid19]。在RSTE中，评分$R_{i,j}$的线性表达式被定义为：

$$R_{i,j} = U_i^T V_j + \beta \sum_{U_k \in N_i} S_{i,k} U_K^T V_j \tag{10}$$

其中N_i是用户i的朋友的集合，$S_{i,k}$用于归一化用户i朋友的总评分数，β在用于平衡用户自身的属性和用户朋友品味的信息。该方法的优化问题可以被定义为：

$$min_{U,V,\Omega} \parallel (R - U^T V) - \beta S U^T V \parallel_F^2 + + \lambda(\parallel U \parallel_F^2 + \parallel V \parallel_F^2) \tag{11}$$

其中$Social(T, S, \Omega)$被定义为$\|R - \beta S U^T V \|_F^2 - 2Tr((R - U^T V)\beta S U^T V)$

2.3.3 正则化方法

正则化方法的基本思想一个用户的倾向应当与他的朋友相似。因此在训练过程中，正则化方法着重于强制使给定用户的潜在特征接近其朋友的潜在特征。SocialMF[Hybrid20]是一种典型的正则化方法。在这种方法中，$Social(T, S, \Omega)$被定义为$\sum_{i=1}^{n}(U_i - \sum_{U_k \in N_i} S_{i,k} U_k^T)^2$，用户$i$的被迫使向其朋友的平均倾向靠拢。使用这个部分融合社交信息后，SocialMF旨在优化下面这个问题：

$$min_{U,V,\Omega} \parallel (R - U^T V) \parallel_F^2 + \alpha \sum_{i=1}^{n}(U_i - \sum_{U_k \in N_i} S_{i,k} U_k^T)^2 + \lambda(\parallel U \parallel_F^2 + \parallel V \parallel_F^2 + \parallel \Omega \parallel_F^2) \tag{12}$$

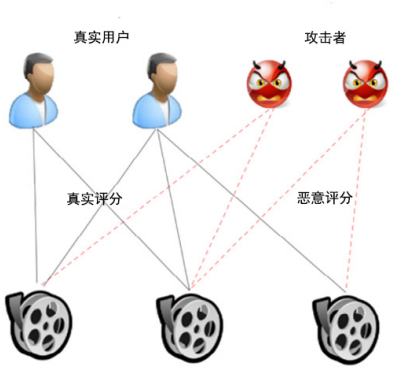
在本文的实验中，使用了这三个社交化推荐算法

第三章 模型与算法

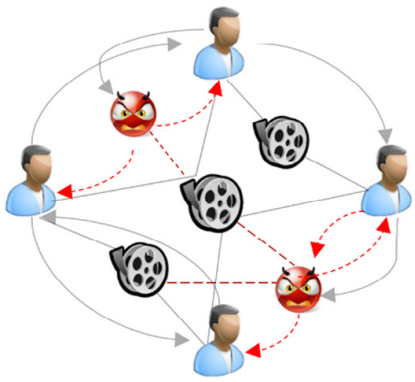
3.1 托攻击

协同过滤推荐系统和社交网络系统存在重大的缺陷。这些缺陷大多源于这两个系统对外开放并依赖用户概貌的特性。如果推荐系统在计算推荐时应用了用户之间的相似度，则可能存在一些虚假的的用户概貌文件(User profile)，并试图使推荐系统出现偏差。这些由攻击者创建的虚假用户概貌被称为托攻击概貌(Shilling profile)。 根据攻击者的目标, 托攻击可以被划分为以提高目标项目推荐率为目标的的推攻击(Push attack)和打击目标项目的核攻击(Nuke attack)。通俗而言，托攻击可以理解为网络中的水军攻击。以国内最大的电影评分网站豆瓣为例，由于豆瓣的评分在国内具有不可替代的参考性，许多质量较差的电影发行方会雇佣大量的水军去给电影打虚假的高分，从而增加电影被推荐的概率与提高电影的商业吸引力。

推荐系统中用户概貌的结构特性决定了它会削弱推荐系统抵抗攻击的能力，并且使攻击者概貌难以检测。这些概貌使推荐出现偏差，推荐系统的潜在用户可能对系统的推荐出现强烈的不满，并致使商业平台出现经济损失。下图呈现了托攻击的结构：

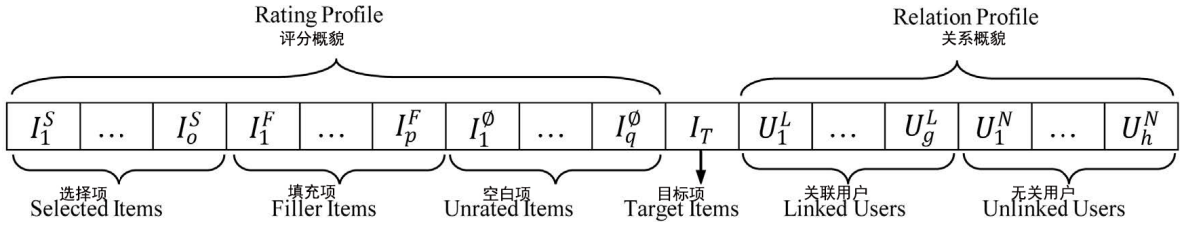


基于恶意评分的攻击


基于虚假用户关系和恶意评分的攻击

3.1.1 攻击概貌

社会化托攻击的攻击概貌[24,25]的总体结构如下图所示：



针对社会化推荐系统的攻击需要由攻击者注入一系列的攻击概貌。攻击概貌由评分概貌和关系概貌两部分组成，对应了用户评分数据库和用户关系数据库。对于攻击者而言，评分概貌用于伪装身份以及对目标项目施加积极或消极的影响；关系概貌用于帮助攻击者获得更多的关注者以提高在社交网络中的影响力。尽管托攻击具有不同的攻击策略，攻击者概貌通常情况下可以被划分为六个部分：选择项I_S(Selected items)，填充项I_F(Filler items)，空白项(Unrated items)$I_\emptyset$，目标项(Target items)I_T，关系用户(Linked users)U_L和无关用户(Unlinked users)U_N。前面的四部分组成了评分概貌，后面的两部分构成了关系概貌。对这六部分详细的解释如下：

- 选择项是攻击者利用以成为真实用户的近邻(neighborhood)的项目的集合。选择项根据攻击策略的不同被给予不同的评分，在一些攻击模式中不使用
- 填充项是使攻击者伪装成真实用户的项目的集合。填充项根据攻击策略的不同被给予不同的评分
- 空白项是攻击者没有评分的项目的集合
- 目标项是攻击者要施加积极/消极影响的项目的集合。目标项的评分通常情况下根据攻击目标的不同被给与最高分或最低分（如在1-5的评分系统中，推攻击评5分，核攻击评1分）这些项目的均方根误差(RMSE)会被影响（在攻击前后对这些项目的预测评分会产生非常大的变化）[26]
- 关联用户是与攻击概貌有信任关系的真实用户概貌。这些真实的用户概貌可能信任攻击者，或被攻击者信任，或双向信任
- 无关用户是与攻击者没有信任关系的真实用户概貌

3.1.2 攻击模式

为了清除地区分攻击模式，在本文中根据不同攻击模式使用到的信息，将攻击模式分为了三大类。攻击模式如下所示：

- 评分攻击：通过注入歪曲的评分概貌使推荐系统生成有利于攻击者的推荐
- 关系攻击：依赖建立广泛的关系连接，扩大在社交网络中的影响力，以在庞大的用户群中获利
- 混合攻击：融合了评分攻击和关系攻击，以提高在推荐系统中破坏性

值得注意的是在这三种攻击模式中，关系攻击并不是一个典型的针对推荐系统的攻击，但它可以作为增强破坏性的辅助攻击。

通常情况下，评分攻击具有如下几种攻击模型：

随机攻击

在随机攻击中，攻击概貌组随机从评分数据库中选取填充项，并给与这些项目整个数据库的平均分。填充项评分的公式如下所示：

$$r(I_F) \sim (\mu, \sigma^2) \tag{13}$$

其中μ和σ为数据库中所有项目的均值和标准差。目标项目根据攻击目标的不同分别给与最高分或最低分。随机攻击的执行花销较低，但效果欠佳

均值攻击

在均值攻击中，攻击概貌组随机从评分数据库中选取填充项，并给与这些项目当项目的平均分。填充项评分的公式如下所示：

$$r(I_F) \sim (\mu_{item}, \sigma^2_{item}) \tag{14}$$

其中μ_{item}和σ_{item}为当前项目的均值和标准差。目标项目的评分与随机项目使用相同的策略。

流行攻击

在流行攻击中，填充项随机选择，可以按照随机攻击和均值攻击的评分策略赋值。目标项目的评分也使用相同的策略。流行攻击的主要区别在于使用了一个额外的选择项列表，这些列表中的项目需要被众多用户评分过，且评分是积极的。因此攻击者会对这些项目打高分以成为大批量用户的近邻。就根据用户之间的相似度计算的推荐而言，这是一种非常强大的攻击。流行攻击的攻击概貌更接近真实用户，所以攻击效果较好。

分段攻击

分段攻击是一种与众不同的攻击方式，攻击者需要得到关于项目的额外知识，如特定项目的商品分类。攻击者选择与目标项目种类一致的项目作为选择项，且$r(I_S) = r_{max}$。填充项除填充项和目标项中随机选择，并且$r(I_F) = r_{min}$

随机关联攻击

在随机关联攻击中，U_L为随机选择的用户。这种策略下随机选择关联的用户，因为不需要需要被关联的用户的进一步的信息，所以在实现上比较简单。

目标关联攻击

在目标关联攻击中，U_L为针对特定的攻击计划选择的用户。这种策略下需要需要被关联的用户的信息，例如被关注者的数量，关注者的数量，评分分布等。了解这些信息可以帮助攻击者缩小链接范围，然后轻松捕获将追随他们的用户。

充分利用有关项目和链接的信息通常会使攻击者更多地获利，并减少被检测到的机会[27]。 然而，收集和利用相关的信息是有成本的。攻击者必须在成本和有效性之间进行权衡。此外，为攻击模型指定适当的参数可以使攻击者保持隐蔽性并难以被发现。 如下是有关参数的一些说明：

- 填充规模：填充项占系统中所有项目的百分比
- 连接规模：关注的人占系统中所有用户的百分比
- 选择规模：选择项占系统中所有项目的百分比
- 攻击规模：攻击概貌占系统中所有概貌的百分比

在社会化推荐系统中，通常将以上的攻击模式相结合，得到如下表所示的攻击模式：

3.2 托攻击检测算法

在托攻击下，推荐系统往往会给出错误的建议，并降低系统现有用户的信任度。因此，通过适当的检测算法来发现试图执行托攻击的攻击概貌是非常有意义的。在本文中，使用了一个无监督学习的方法去实现托攻击检测。作为一种通用的算法，无监督学习的托攻击检测方法可以运用到任何推荐系统中。基于托攻击概貌与其他概貌有高度相似地协方差，PCA算法是检测托攻击概貌的理想算法[28]。

```
输入：
R,用户的评分矩阵(n个用户,m个项目)
c,相关阈值(correlation threshold)
p,概貌阈值(profile threshold)
输出：托攻击用户的列表
Corr_matrix(n×n)：用户的相关矩阵
for i:1->n(Corr_matrix中的每个用户) do
    count=0
    for j:1->n(第i个用户的每个相关值)
        if Corr_matrix(i,j)>=c then
            count=count+1;
        end if
    if count>p(概貌阈值)
```

```
        将用户i添加入托攻击用户的列表
    end if
return Shilling_list;
```

由于托攻击概貌与其他用户有非常高的相关性，相关阈值的值必须被设置得很高。设置偏低的相关阈值有可能会导致错误地把真实用户检测为托攻击概貌，从而使检测算法失真。概貌阈值是能够影响推荐准确性并使推荐系统出现偏差的用户数，由一些启发式方法找到[28]。

在上述算法中，PCA检测法被运用到检测托攻击概貌中。初始的相关性由评分矩阵中的用户计算。相关性矩阵中存放每两个用户之间的相关值。在PCA中，与其他概貌有高度相关性的概貌被认为是托攻击。如果一个概貌的相关性超过了与其他概貌的相关性阈值，并且计数超过了概貌阈值，那么该概貌将被判定为托攻击概貌。

由于托攻击成组进行，托攻击概貌表现出高相关性。因此，在这些攻击概貌下，这些概貌对目标项和填充项都有相似的评分。如果一个攻击者试图对填充项打不同的分，攻击的开销会极大地提高。因此，如果概貌阈值设置正确，则托攻击概貌具有高相关性的假设将更可能给出准确结果。

3.3 评价指标

本文将从三个角度对社交网络进行风险评估。

第一部分从社交网络分析的角度出发，通过网络节点分布、连接性、节点的影响力等指标与可视化结果，结合部分社会学与传播学现象分析FilmTrust社交网络潜在的服务风险；

第二部分使用了三种经典的托攻击算法对FilmTrust数据集进行了模拟攻击，并在SoRec和SBPR算法上验证。使用到的评价指标如下：

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{15}$$

$$NDCG = \sum_{i \in T_u} \frac{2^{r_i} - 1}{log(pi + 1)} \tag{16}$$

第三部分评估了在托攻击风险下使用检测算法能力。风险的量化采用了类似机器学习中的F1公式：

$$F - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{17}$$

其中准确率定义为检测到的托攻击概貌占所有判定为托攻击概貌的占比；
召回率定义为检测到的托攻击概貌占系统中所有托攻击概貌的占比

第四章 实验过程与结果

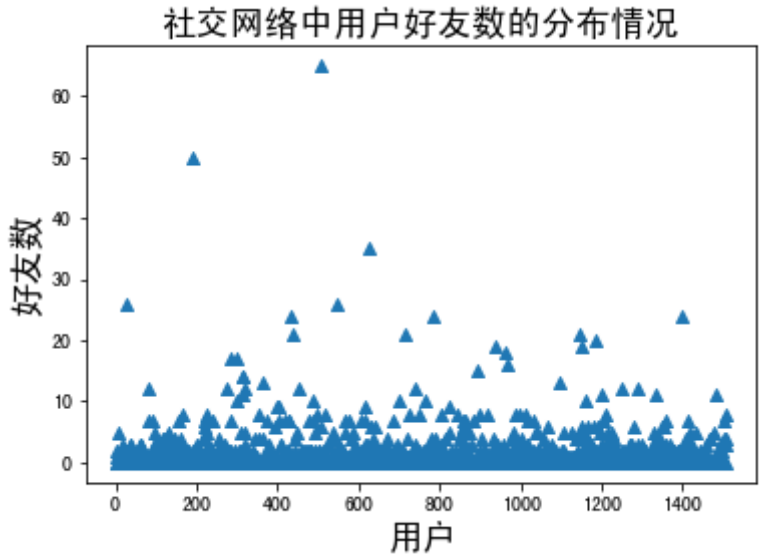
4.1 社交网络分析结果

FilmTrust为2011年从网站FilmTrust 完整抓取下来的数据集。本数据集由两部分组成：ratings.txt 和 trust.txt。其中rating数据集包含了编号1-1508的用户对电影的打分数据，trust数据集包含了编号1-1642的用户之间的社交关系。由于trust数据集的用户范围超过了rating数据集的用户范围，为了简化，在读入数据时将trust数据集中的用户范围限定与rating数据集相同。筛选之后数据集的内容如下表所示：

用户数	**1508**
项目数	2071
评分记录条数	35497
社交关系条数	1853

本文通过NetworkX[]和分析该社交网络组成情况。NetworkX是一个Python包，用于创建，操作和研究复杂网络的结构，动态和功能。在NetworkX中，社交网络的结构通过邻接表的形式储存。将trust数据集的信息转换为邻接表后，即可用NetworkX计算该社交网络的连接性、中心度等属性。

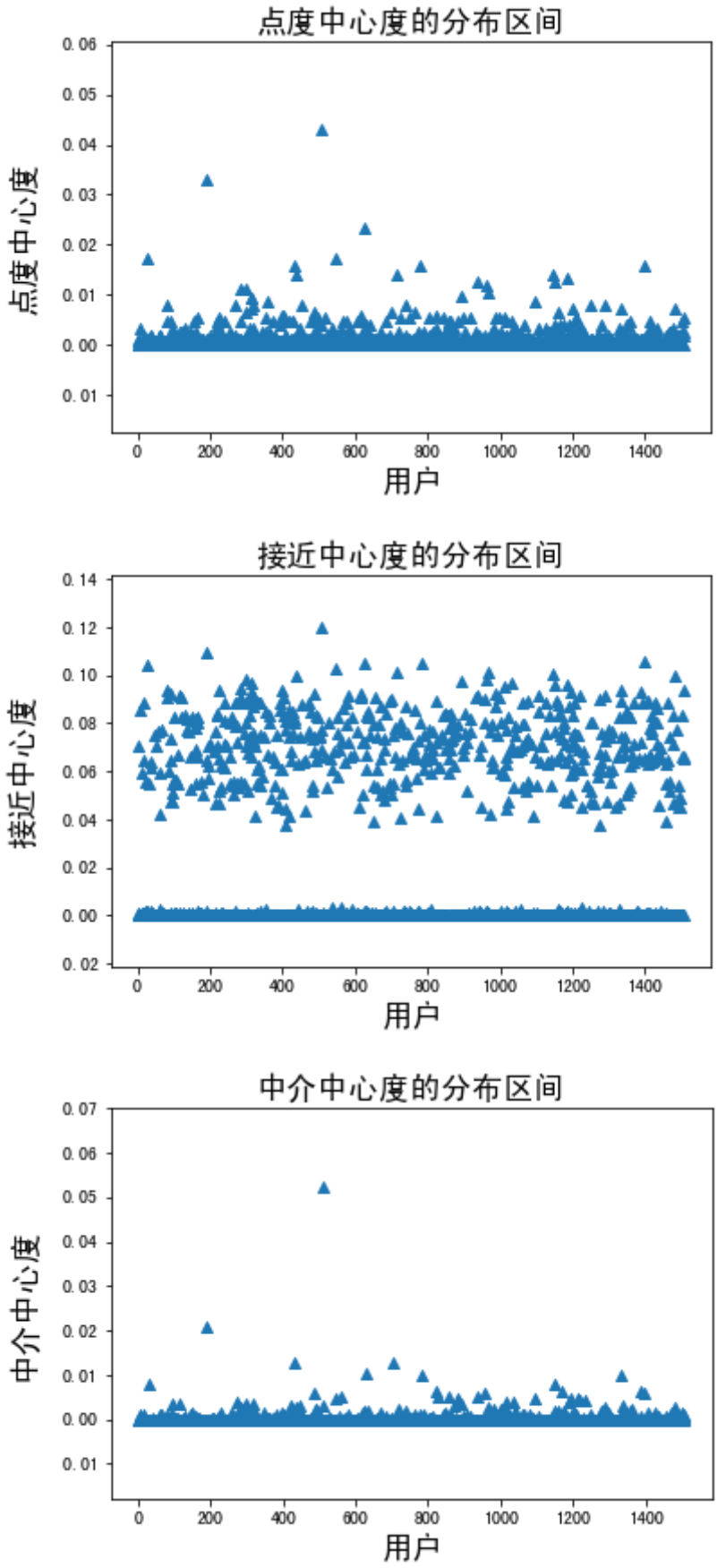
在连接性方面，1508个已知用户中，522个用户至少有一名好友，而986个用户在社交网络中处于孤立状态；分布情况如下图所示：



经过计算，社交网络非孤立节点的平均度为： 2.995 ，标准差为： 4.568，此数据将会用于下一部分针对社交化推荐系统的托攻击；

经过Gephi[Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.]的Yifan Hu Propositional布局可视化(不包括孤立节点)后，可以得到当前数据集社交网络的结构图：

中心度的度量如下图所示：



从社交网络分析的角度出发，FilmTrust社交网络的服务风险体现在如下几个方面：

- 孤立用户相关：超过一半的用户在社交网络中处于孤立状态，在进行社会化推荐时，由于缺少好友评分的倾向性，这些孤立用户会面临冷启动的困境；事实上，当前FilmTrust数据集所体现出的社会关系是符合社交网络现状的——例如豆瓣和YouTube，大多数用户属于网络上的“隐形人”，并不

会主动去与陌生人建立社交关系。商业平台需要针对这一现象发掘解决方案，如引导用户积极建立社交关系等；

- 脱网用户相关：从Gephi的可视化结果不难发现，在建立过关系的用户中，仍有部分用户只是与少数几个其他用户建立了关系，这些小型的关系游离在主干网络之外。虽然早期的社交推荐系统主要依靠直接的朋友关系，但在近年的研究中，间接关系愈发成为考量的焦点。因此，推荐系统对这些用户的推荐可能会出现偏差，缺乏准确性；
- 中心度分布相关：作为衡量网络中节点影响力的三个重要指标，点度中心度、接近中性度和中介中心度也可以作为评估风险的指标。接近中心度的结果说明该社交网络非孤立节点具有较均匀的分布，有利于维持社交网络的健壮性；而点度中心度和接近中心度的分布则较为极端——只有极少数的节点具有广泛的社交关系/能够为关系的建立提供桥梁的服务；因此提高网络的健壮性可以从这两个角度出发去改善

4.2 针对不同推荐算法的风险评估

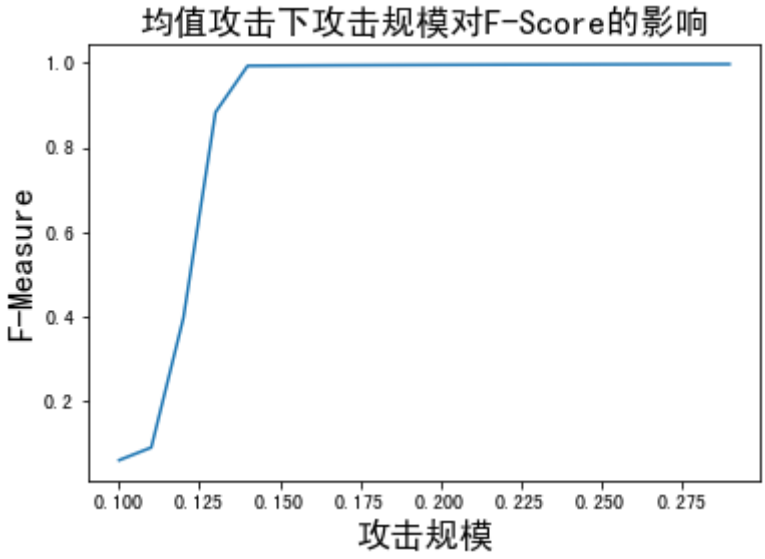
为了实现这一部分的目标，需要实现托攻击概貌的模拟。由于现实世界中几乎没有数据集给出了攻击者的标签，因此在本文中攻击者采用随机生成的方式。对于每次攻击，在给定攻击规模后，计算出当前需要伪造的托攻击概貌的个数。混合攻击由评分攻击和关系攻击组成。

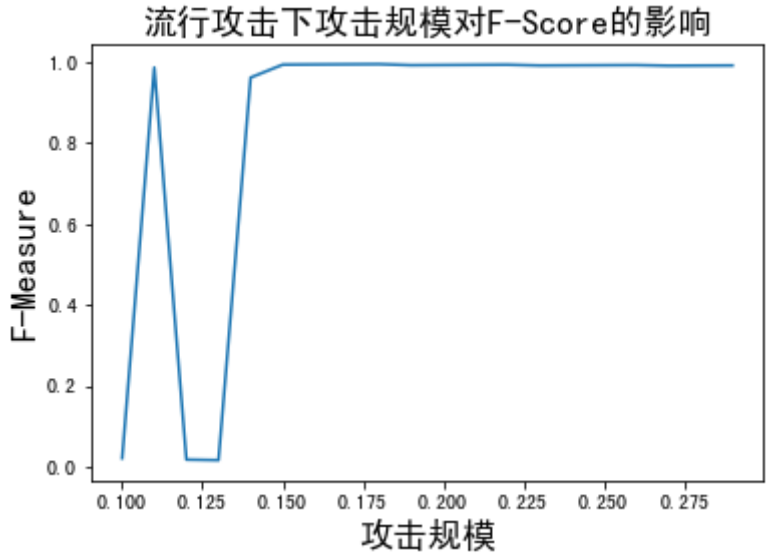
评分攻击：针对不同的攻击策略，在用户-项目评分矩阵中分别给所有托攻击概貌的填充项和目标项评分，得到篡改后的评分文件；

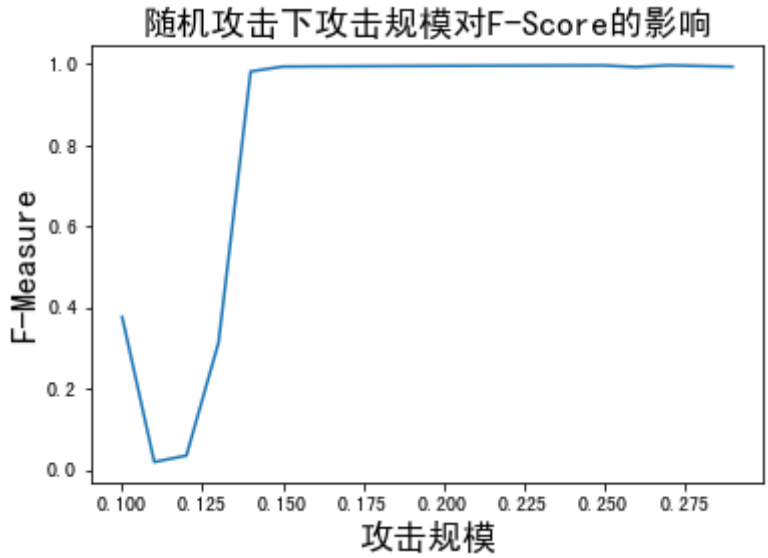
关系攻击：本文的关系攻击部分使用的是随机关联攻击。假设与每个托攻击概貌相关联的用户数符合均值为μ，标准差为σ的标准分布，其中μ与σ为数据集社交网络中非孤立用户度的均值和标准差。每个托攻击概貌通过标准分布随机函数得到关联的用户数并与之建立联系，得到篡改后的关系文件；

4.3 托攻击检测效率的风险评估

在第三部分中，根据不同的攻击策略(随机攻击、均值攻击、流行攻击)和攻击规模(5%-25%)生成全新的托攻击概貌数据集。在参数设置上，相关阈值设置为0.95，概貌阈值设置为10%。由之前提到的F-score指标对检测算法性能进行评估。







从实验结果可以推断出，当系统中的托攻击概貌数量大于或等于概貌阈值是，托攻击检测算法将会呈现出非常高的准确度；反之，如果托攻击概貌小于概貌阈值，推荐系统不会受到很大的影响。

第五章 总结与未来展望

致谢

随着毕业论文的最终定稿，我的本科生涯也即将划上句号。在下半年的时候，我也将背上行囊前往一片陌生的大陆，继续的我的学术生涯。此刻我并没有太多复杂的心情，只想对这一路走来遇到的人和事表示由衷的感谢。

首先，我要向我的指导老师肖阳导师表示最诚挚的谢意。在完成毕业设计的过程中，肖老师极尽他的耐心与关切，指导我们从无到有完成了大学四年最后一部宏大的诗篇。没有肖老师悉心的指导和启发性的建议，这篇论文远远不可能达到迄今的程度。

其次，我要感谢在我留学之路上给与我教诲、帮助和陪伴的人。感谢我的德语老师尚天欣、谈书婷和Jonas，在他们幽默生动的讲授下，枯燥机械的语法也变得妙趣横生；我也非常感激张俊伟、杨超、刘志宏和肖阳老师在推荐信上给我提供的帮助，没有他们，我的申请之路会充满困难；此外，回首过去的两年刷绩点、准备雅思考试和德累斯顿工业大学面试的历程，我也不得不感谢我的朋友胡同煦、王嘉辉和张迅齐的陪伴，是他们陪我度过了C楼和图书馆的日日夜夜。

我非常庆幸选择了西安电子科技大学，在这里的四年给与了我人生中最宝贵的经验，也留下了最快乐的回忆。师长们严谨的学术态度与谦逊的处事作风为我树立了正确的人生观；丰富的教学资源让我享受到了高水平的发展平台；西安得天独厚的地理位置也让我充分发展了自己的兴趣爱好，见识了祖国大地不同的风土人情。

我还想感谢这次新冠疫情，虽然经历了隔离、网课和史上最难的申请季，但这场全球性的灾难让我在人文、社会和地缘政治上脱离了过去肤浅的认知，使我能够从更宏观和中立的角度去思考我的位置、责任和使命，并坚定我选择的道路。感谢B站UP主峰哥亡命天涯和Reddit社区China_irl，在我每一次情绪低谷时给与我莫大的安慰和快乐，并在精神上让我成为一个更加羽翼丰满、有格局的人。

本人学识有限，论文中的不足还请老师批评指正，我会在今后的生活和学习中不断完善。

最后，再次感谢各位老师。我将铭记我校“厚德、求真、砺学、笃行”的校训，书写美好未来。坚守初心，回归本心，保持信心，守护真心。

引用 Citations

[1] Recommendation systems: Principles, methods and evaluation
[2] Pu P, Chen L, Hu R. A user-centric evaluation framework for recommender systems. In: Proceedings of the fifth ACM conference on Recommender Systems (RecSys'11), ACM, New York, NY, USA; 2011. p. 57–164.
[4] X. Su, T.M. Khoshgoftaar, A survey of collaborative filtering techniques, Adv. Artif. Intell. 2009 (2009) 4.
[5] J.B. Schafer, D. Frankowski, J. Herlocker, S. Sen, Collaborative filtering recommender systems, in: The Adaptive Web, Springer, 2007, pp. 291–324.
[6] H. Gao, J. Tang, H. Liu, gscorr: modeling geo-social correlations for new check-ins on location-based social networks, in: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, ACM, 2012, pp. 1582–1586.
[7] S.K. Lam, J. Riedl, Shilling recommender systems for fun and profit, in: Proceedings of the 13th International Conference on World Wide Web, ACM, 2004, pp. 393–402.
[8] Ulrik Brandes, in International Encyclopedia of the Social & Behavioral Sciences (Second Edition), 2015
[9] F.N. Stokman, in International Encyclopedia of the Social & Behavioral Sciences, 2001
[10] J. Domingo-Ferrer, A. Viejo, F. Sebe, U. Gonzalez-Nicolas, Privacy homomorphisms for social networks with private relationships, Computer Networks 52 (2008) 3007–3016.
[11] C. Kiss, M. Bichler, Identification of influencers — measuring influence in customer networks, Decision Support Systems 46 (1) (2008) 233–253.
[12] P. Van Baalen, J. Bloemhof-Ruwaard, E. van Heck, Knowledge sharing in an emerging network of practice, European Management Journal 23 (2005) 300–314.
[13] P. DeMeo, A. Nocera, G. Terracina, D. Ursino, Recommendation of similar users, resources and social networks in a social internetworking scenario, Information Sciences 181 (7) (2011) 1285–1305.
[14] L. Zhen, Z. Jiang, H. Song, Distributed recommender for peer-to-peer knowledge sharing, Information Sciences 180 (18) (2010) 3546–3561.
[15] Y.A. Kim, J. Srivastava, Impact of social influence in e-commerce decision making, Proceedings of the ninth international conference on Electronic commerce, ACM, New York, NY, USA, 2007, pp. 293–302.
[16] K.O. Lee, N. Shi, M.K. Cheung, H. Lim, C.L. Sia, Consumer's decision to shop online: the moderating role of positive informational social influence, Information Management 48 (6) (2011) 185–191.
[17] V. Carchiolo, A. Longheu, M. Malgeri, Reliable peers and useful resources: searching for the best personalised learning path in a trust- and recommendation-aware envi- ronment, Information Sciences 180 (10) (2010) 1893–1907.

[18] R. Albert, A. Barabasi, Statistical mechanics of complex networks, Reviews of Modern Physics 74 (47) (2002) 47–97.
[19] J. Srivastava, N. Pathak, S. Mane, M.A. Ahmad, Data mining for social network analysis, IEEE International Conference on Data Mining, Hong Kong, 2006, pp. 18–22.
[20] J.C. Wang, C.C. Chiu, Recommending trusted online auction sellers using social network analysis, Expert Systems with Applications 34 (3) (2008) 1666–1679.
[21] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence through a social network, Proceedings of ACM SIGKDD'03, 2003, pp. 137–146.
[22] Analysis D D and Raya V 2013 Social Network Analysis, Methods and Measurements Calculations pp 2-5
[23] Xiaoyuan Su and Taghi M. Khoshgoftaar, 'A Survey of Collaborative Filtering Techniques' Hindawi Publishing Corporation Advances in Artificial Intelligence , Volume 2009, Article ID 421425, 19 pages, August 2009.
[24] Yishu Xu, Fuzhi Zhang, Detecting Shilling Attacks in Social Recommender Systems Based on Time Series Analysis and Trust Features
[25] Junliang Yu, Min Gao, Wenge Rong, Wentao Li, Qingyu Xiong, Junhao Wen, Hybrid Attack on Model-based Social Recommender Systems
[26] Youquan Wang, Lu Zhang , Haicheng Tao , Zhiang Wu, Jie Cao,'A Comparative Study of Shilling Attack Detectors for Recommender Systems,' IEEE 2015 12th International Conference on Service Systems and Service Management (ICSSSM), pp. 1-6, June 2015.
[27] Y. Wang, L. Zhang, H. Tao, Z. Wu, J. Cao, A comparative study of shilling attack detectors for recommender systems, in: 2015 12th International Conference on Service Systems and Service Management (ICSSSM), IEEE, 2015, pp. 1–6.
[28] ADVANCING RECOMMENDER SYSTEMS BY MITIGATING SHILLING ATTACKS
[29] Social Recommendation on review