电子学

文章推荐系统中图神经网络方法的偏差放大比较分析

Nikzad Chizari、Niloufar Shoeibi 和María N. Moreno-García *

计算机科学与自动化系，理学院，萨拉曼卡大学，Plaza de los Caídos sn，37008 萨拉曼卡，西班牙* 通讯作者：mmg@usal.es

摘要： 推荐系统（RSs）用于向用户提供个性化的物品推荐，帮助他们克服信息过载问题。目前，基于深度学习的推荐方法正逐渐超越矩阵分解等传统方法，这是因为它们能够表示用户与物品之间的复杂关系，并整合额外信息。这些数据具有图结构，而图神经网络（GNNs）具备更强的结构学习能力，因此图神经网络已成功应用于推荐系统。然而，在使用这些算法时，偏差放大问题仍有待研究。偏差会导致不公平决策，由于社会失望情绪和环境损害，这可能对公司的声誉和财务状况产生负面影响。本文旨在通过文献综述，以及分析不同基于图神经网络的算法相较于最新进展方法在应对偏差方面的表现，全面研究这一问题。我们还将探讨适当的解决方案，以尽可能小地影响模型性能为代价解决这一问题。

关键词： 推荐系统；图神经网络（GNN）；偏差放大；平均流行度；基尼指数；敏感特征

check for updates

引用： Chizari, N.; Shoeibi, N.; Moreno-García, M.N. 推荐系统中图神经网络方法的偏差放大比较分析。《电子学》**2022年**, *11*, 3301. https://doi.org/10.3390/electronics11203301

学术编辑：João Carneiro、Patrícia Alves 和 Goreti Marreiros

收稿日期：2022年8月31日 接受日期：2022年10月10日 发表日期：2022年10月13日

出版商说明： MDPI 对已发表地图中的司法管辖权主张及机构隶属关系保持中立。

版权： © 2022年 作者所有。被许可方为瑞士巴塞尔的 MDPI。本文为开放获取文章，依据知识共享署名（CC BY）许可条款和条件发布（https://creativecommons.org/licenses/by/4.0/）。

1. 引言

如今，许多用户受益于在线购买各种产品和服务的便利。当人们面对过多选择时，问题便随之而来，这可能导致信息过载，使决策过程变得困难。为了解决这一问题，人们使用推荐系统（RS）向用户提供个性化物品推荐 [1–5]。

鉴于推荐系统在众多应用领域中得到广泛应用，确保推荐结果公平至关重要。偏差是推荐系统中最重要的问题之一，会影响推荐系统的有效性 [1,2]。有偏的推荐结果可能导致资源和机会分配不平衡且不公平 [6]。这些有偏的决策可能给个人和公司造成严重的经济、社会和声誉损害。此外，偏差还可能导致歧视，这违反了法规 [7–9]。偏差还可能造成严重的法律、技术和安全损害 [10]。近年来，上述问题促使人们开始研究推荐系统中的偏差问题，探讨这一主题的文章数量也有所增加 [2]。推荐系统中的偏差所引发的这些问题，是本文的主要研究动机。

近年来，深度学习方法也取得了显著进展，并开始应用于推荐系统。深度学习方法能够构建多层、非线性的层间互连网络结构，旨在自动从数据中提取多层次特征的表示 [11]。这些方法可以显著提升推荐系统的性能，并解决其面临的问题。在这些方法中，图神经网络（GNN）算法已被证明在许多需要处理图数据的学习任务中十分有效，因为图数据包含丰富的元素间关系信息。图神经网络

可以通过消息传递沿图边传播信息，从而捕捉图中的依赖关系 [12]。

为了解决传统推荐系统的局限性，图神经网络算法已成功应用于推荐系统。近年来，图神经网络在多个应用领域的应用迅速普及。这主要归功于它们从图形信息表示中学习的能力，以及深度学习方法所具备优势的重要性。

在推荐系统中，用户–物品交互可以建模为图。此外，还可以利用额外数据改进推荐，包括社交信息或上下文信息。在推荐系统领域，基于神经网络的方法，尤其是采用深度学习算法的方法，已被提出作为协同过滤方法的替代方案，这得益于它们学习用户与物品之间复杂关系的能力。然而，这些方法只能处理欧几里得空间中的数据，因为它们并非为处理高阶信息结构而设计的 [6,12]。近期的图神经网络技术能够克服这些缺点，它们将深度学习算法扩展到了非欧几里得空间 [13]。

在图神经网络中，会执行一个迭代过程，将信息在节点之间传播，并从邻居节点聚合信息。目前，已有多种基于图神经网络的推荐模型方案，既有通用模型，也有专注于序列或会话推荐的模型 [14]。不同的图神经网络变体可用于完成这些任务。其中，图卷积网络（GCN）、图注意力网络（GAT）和图循环网络（GRN）在包括推荐系统在内的众多领域中，性能均优于其他机器学习方法 [12], 。尽管基于图神经网络的模型被公认为具有诸多优势，但要使这些方法胜过传统方法，仍有一些挑战亟待解决。其中之一是如何应对对推荐系统产生负面影响的多种偏差。一些研究表明，图结构的使用可能会放大公平性问题和其他社会偏差 [15]。

根据 Baeza-Yates（2016）和 Chen 等人（2020）[2,16]的研究，推荐系统中的偏见通常可分为四大类：数据偏差、算法偏差（模型偏差）、结果与响应偏差，以及会加剧这一问题的反馈回路偏差。这些偏差都可以进一步划分为多个类别。偏差还会在不同方面引发严重问题，包括经济、法律、社会、安全和技术层面 [17], ，以及内容、隐私、自主性、不透明性、公平性和社会层面等伦理方面 [18,19]。此外，效用、覆盖率、多样性、新颖性、意外发现性、可见性和曝光等目标也会受到偏差影响 [20]。偏差还可能通过用户评论体现出来，例如序列偏差、观点偏差和文本偏差 [21,22]。不同平台上的偏差类型各异，包括功能偏差、规范偏差、外部偏差以及非个人账户 [23]。偏差对多个研究领域和应用领域的重大影响，凸显了深入分析偏差并提出有效解决方法的重要性。推荐系统领域对偏差的研究已有数年。然而，在图神经网络等新算法背景下，对偏差的研究仍十分有限。

本文旨在分析总体上困扰基于图神经网络的推荐系统的不同类型偏差。我们还深入分析图神经网络算法应对不同类型偏差时的表现（鉴于目前针对这些算法应对某些类型偏差的研究较为有限），并将其与其他广泛使用的算法进行比较。我们探究了这样一个问题：除了在推荐结果中实现较高的准确率之外，是否还可以相信该算法不存在偏差，也不会作出不公平决策。我们还考察了哪些偏差在图神经网络算法中更为突出，并研究应对这些偏差的方法。

在以下章节中，我们介绍机器学习（ML）、图神经网络算法和推荐系统中的偏差问题的最新进展，最后重点讨论基于图神经网络的推荐系统中的偏差放大问题。在本节中，我们将考察这一特定研究领域中最重要的相关工作。随后，我们将详细说明实验研究。比较传统方法与基于图神经网络的推荐系统方法以及两个真实数据集的比较研究结果将在第

4节中呈现。最后，我们将讨论该领域中最重要的问题以及未来的工作。

2. 最新进展

本节将从机器学习、图神经网络算法、推荐系统以及基于图神经网络的推荐系统等不同方面，综述偏差问题。文献中已有大量研究从不同视角探讨偏差。不过，本节也将分析有关偏差问题的一般性重要研究，重点关注推荐系统，尤其是基于图神经网络的推荐系统。

2.1. 机器学习中的偏差

为了更好地理解不同方面的偏差问题，以及该问题如何影响推荐系统的结果，本文在此给出不同层面偏差的定义，例如偏差的统计学含义以及机器学习（ML）中的偏差。

根据[24]，，统计偏差是指由测量或抽样过程产生的系统性误差。其中，有必要区分由随机机会造成的误差和由偏差造成的误差。

数据是机器学习算法的输入。当该输入存在偏差时，这种偏差可能会被传递到算法生成的模型中，导致不公平决策并降低质量 [25,26]。这些问题可能对公司产生严重的财务、社会和声誉影响[8]。

基于机器学习的模型中的算法偏差源于异常数据集、薄弱模型、糟糕的算法设计或历史上的人类偏差[27]。在训练阶段，欠拟合问题也可能导致算法偏差，这可能是训练数据的局限性与模型容量问题共同造成的[28]。影响这一机制的因素包括不可约误差（贝叶斯误差）、正则化机制、类别不平衡以及代表性不足的类别[28]。

在CRISP-DM的不同阶段会出现各种类型的常见偏差。CRISP-DM是由IBM于2015年提出的、广为人知的数据挖掘标准流程，将数据处理过程分为六个不同阶段，包括业务理解、数据理解、数据准备、建模、评估和开发。所述偏差类型包括社会偏差、测量偏差、表征偏差、标签偏差、算法偏差、评估偏差、部署偏差和反馈偏差[8]。当数据大规模地将社会中的偏差传递给模型时，就会产生社会偏差。测量偏差源于业务理解阶段中的人为错误，尤其是在处理年龄和性别等敏感特征时。这类偏差也可能发生在数据准备阶段。此外，当样本或数据分布不能代表真实的底层分布时，表征偏差会在数据收集和抽样过程中出现。标签偏差可能出现在数据准备阶段，即为预测任务选择标签时。由于含糊性以及文化或个体差异，为数据集选择最佳标签可能非常困难。算法偏差也可能在建模阶段因模型参数或模型误分类等技术问题而产生[29]。

同样重要的是，根据训练数据统计，了解模型是否会放大数据中已有的偏差 [29]。评估偏差发生在评估阶段，可能是由于训练数据和测试数据的人群差异所致。最后，在复杂的社会技术环境中，模型实施后可能会出现部署偏差。

机器学习中的偏差也可能导致不公平结果。机器学习公平性可分为十类：统计平等、均等化赔率、机会均等、差别影响、差别性不当对待、待遇平等、广义熵指数、个体公平（以前称为通过认知实现公平）、通过无知实现公平，以及反事实公平 [30]。

[29] 对偏差放大进行了系统且受控的研究。为此，研究采用了一个高度受控的简单图像分类问题。结果表明，多个因素与偏差放大相关，包括模型的准确率、模型容量、模型过度自信程度以及训练数据的规模。此外，结果还表明，偏差放大在训练期间可能有所不同；而分类任务识别群体成员身份的难度也会影响偏差放大。

2.2. 图神经网络中的偏差

图神经网络及其变体在多种图学习任务中都展现出优异的性能。然而，由于当前数据集规模不断扩大，它们面临着显著的计算挑战。图卷积的多层结构意味着需要自顶向下递归地进行邻居聚合，这可能导致邻域大小随层数增加而不断扩大。如果图是无标度且稠密的，那么即使只有几层，计算嵌入也需要涉及图中的很大一部分，这对于大规模图而言是不可行的 [31,32]。

其他研究表明，与异质节点相比，图神经网络在同质节点上的表现更好。为考察图是否具有同质性，研究者定义了同质性比率。同质性比率较高的图被视为同质图，而比率较低的图则被视为非同质图 [33]。

尽管图神经网络通常能取得更高的准确率，但现有的大多数图神经网络并未考虑公平性问题，因此可能会歧视具有特定敏感特征值的人口统计子群体，例如年龄、性别和种族。已实现的图神经网络所做出的决策可能会受到这类歧视的显著影响 [32,34,35]。此外，许多机器学习系统都是利用人工生成数据训练的；因此，我们显然有必要理解并缓解图神经网络方法中针对人口统计群体的偏差 [36]。

图神经网络算法中的有偏差结果可能源于多种原因，其中最重要的是网络结构存在偏差。尽管找出网络结构中导致偏差的部分至关重要，但人们认为，这种偏差可能源于图神经网络核心操作中的消息传递机制。要理解网络结构中的偏差，面临若干挑战，包括公平性概念差距、可用性差距和忠实性差距。公平性概念差距关注的是如何在实例级别衡量偏差。可用性差距指出，找出计算图中对预测的公平程度影响最大的边也至关重要。最终识别出的边不一定是对公平性贡献最大的边。忠实性差距则意味着需要确保所收集的偏差解释能够反映基于所选模型得出的真实推理结果 [34]。此外，偏差还可能导致训练数据与测试数据之间出现分布偏移，尤其是在标签方面 [37]。

2.3. 推荐系统中的偏见

不同推荐系统提供的推荐质量会因用户的特征及其敏感信息（包括年龄、性别、种族和人格）而有所不同。这种行为与欧盟委员会（EC）的法规相冲突：“要求对人工智能系统进行事前测试、风险管理和人工监督，以最大限度降低在教育与培训、就业、重要服务、执法和司法等关键领域作出错误或有偏差的人工智能辅助决策的风险” [7]。根据这项法规，人工智能系统应遵循欧盟基本权利，例如不受歧视的权利、尊重个人的私人生活以及个人数据保护权 [7]。此外，推荐系统中的有偏差结果可能导致用户不满 [38]。

鉴于推荐系统中的偏差问题是导致不公平决策和歧视的最重要因素之一，而且这一问题显然与上述法规相悖。[2] 中提出的研究指出，偏差可分为三种潜在类型

，可首先对其进行识别：输入数据偏差、计算偏差（可能源于算法，也可能被引入团队决策中）以及结果偏差。这是对 Baeza-Yates (2016) [16], 先前提出的偏差分类的扩展；该分类以循环形式将循环行为细分为七种不同类型的偏差。数据偏差属于观察性偏差而非实验性偏差，当训练数据的分布与理想的测试数据分布不同时便会发生，包括：选择偏差、曝光偏差、从众偏差和位置偏差。算法偏差也可能出现在建模过程的不同阶段，包括训练、评估和特征工程。流行度偏差、不公平性和归纳偏差都可能是这类偏差造成的结果。推荐系统中的流行度偏差源于长尾现象。当少数热门物品在系统中获得最多交互时，就会出现这一常见问题。这可能导致模型忽视不热门物品，并为更热门的物品赋予更高分数 [2]。上述偏差共同作用，可能形成一个循环图：有偏数据从一个阶段流向下一个阶段，同时引入更多新的偏差 [2,39]。这种偏差的循环行为增加了识别需要采取行动之处的复杂性。曝光偏差是由于用户只能接触到物品的特定部分而产生的，这种情况由隐式反馈造成；它也可能由流行度偏差引起，因为系统会推荐最热门的物品 [2,40]。换言之，偏差可能限制用户选择并污染用户反馈，进而加剧曝光偏差 [41]。在上述偏差类型中，流行度偏差被认为是推荐系统领域最重要的一种偏差 [42]。

Ashokan 和 Hass（2021年）提出的另一种偏差分类法[30] 将偏差分为三个主要类别，并进一步细分为：数据生成偏差，包括历史偏差、表征偏差、测量偏差、总体偏差、抽样偏差和辛普森悖论。历史偏差指由社会技术问题造成的既有偏差。表征偏差可能在抽样阶段产生。测量偏差发生在选择、使用和测量特定特征的过程中。总体偏差发生在数据集分布与现实世界总体不同时。抽样偏差也可能源于创建随机子组时出现的错误。辛普森悖论指总体子组在聚合阶段的行为差异所导致的偏差[43]。

模型构建和评估偏差包括评估偏差、聚合偏差、流行度偏差、算法偏差、遗漏变量偏差、人口统计偏差和时间偏差。评估偏差可能出现在模型评估阶段。对总体如何影响模型结果作出错误假设，可能导致聚合偏差。热门物品获得更多交互，因而会产生流行度偏差 [42,44]。所用算法内部的技术问题可能导致算法偏差。遗漏变量偏差源于模型未选取一个或多个关键变量。对不同用户人口统计群体（例如年龄和性别）的差异采取不同处理，可能导致人口统计偏差[30,45]。随着时间推移，行为和总体差异会造成时间偏差[30,46]。

部署和用户交互偏差包括行为偏差、内容生产偏差、链接偏差、呈现偏差、社会偏差、涌现偏差、观察者偏差、交互偏差和排名偏差。数据集中用户行为的差异可能导致行为偏差。用户生成内容在结构、词汇、语义和句法变体方面的差异会造成内容生产偏差。当用户活动所体现的网络属性无法真实反映用户行为时，就会产生链接偏差。呈现信息的过程中会出现呈现偏差。当偏好被刻意赋予某些群体并影响其判断时，也会产生社会偏差 [30,47]。现实用户的行为与数据集中用户的行为存在差异，因此会产生涌现偏差。如果研究人员的预期无意间被带入研究数据中，就可能出现观察者偏差。用户与系统交互的方式存在差异，可能造成交互偏差。最后，排名靠前的结果获得了更多曝光时，就会产生排名偏差 [30]。

鉴于偏差会影响模型的决策，因此有必要考虑各种类型的偏差；然而，近期发表的研究主要致力于解决曝光偏差、流行度偏差、不公平性和偏差循环效应。以往方法面临的最重要挑战之一，是模型性能与偏差缓解之间的权衡，而人们认为这种权衡取决于所选场景[2]。公平性的定义也可能因领域而异，但这一问题近来已受到广泛关注 [48]。

在推荐领域，有研究提出了一种模型，用于模拟社交网络中偏差反馈循环的多个轮次，以分析这种反馈循环的长期影响。该模型采用了不同的控制参数，包括网络中的同质性程度、群体的相对规模、从多种新链接推荐器中进行选择，以及在三种不同的随机用户行为模型之间进行选择；这些参数决定了用户是否会接受每条推荐。这项实验研究的结果表明，同质性程度较高的少数群体可以从所有链接推荐器中获得过多的曝光优势。另一方面，如果该群体具有异质性，其曝光就会不足。此外，少数群体的同质性程度会影响差异化曝光速度，而少数群体的相对规模会放大这种影响。无论少数群体还是多数群体，都会因其同质性程度不同而经历“富者愈富”效应。

在 [1], 中，作者致力于对话式推荐系统（CRS），并从曝光率、成功率和对话效用等多个角度，系统地研究了最新进展中的 CRS 所存在的流行度偏差问题。本文提出了一套专为 CRS 设计的流行度偏差指标。[49] 中的研究也聚焦于推荐系统中的流行度偏差和长尾问题。此外，该论文还提出了用于衡量物品长尾现象的实用指标。为进一步探讨上述问题，参考文献 [30] 继续深入分析，衡量了基于评分的推荐系统中的算法偏差与公平性。这项研究考虑了多种类型的偏差和公平性。此外，研究通过分析两个领域提出了公平性指标。

在包括 [50,51], 在内的一些研究中，性别这一敏感属性被纳入考量。[50] 提出了无偏性别推荐（UGRec），旨在平衡男性和女性用户的性能。为捕捉用户偏好，研究设计了一个信息聚合组件，用于从用户-物品图中学习用户和物品的表征。为改进表征，研究提出了一种聚合用户高阶邻居的多跳机制。此外，研究还采用了结合对抗学习的端到端训练框架，以避免影响准确率。该框架能够去除性别特定特征，同时保留共有特征。 [51]对音乐推荐系统中的性别偏差和歧视进行了探索性分析。这项研究的主要目标是探究哪种协同过滤方法会增强或减轻艺术家性别偏差。为此，研究采用偏好比率（PR）和偏差差异度（BD）指标来衡量结果。结果表明，在真实世界的 LastFM 数据集中，协同过滤推荐系统可能会放大性别偏差问题。

[52] 提出的另一项研究也关注两个图书评分数据集（亚马逊和 Book-Crossing）中的推荐系统性别偏差。该研究针对系统的准确率，提出了一种模型无关的偏差缓解方法。研究采用了两种基于 K 近邻的推荐系统方法。结果显示，偏差显著降低，而模型准确率几乎未受影响。

2.4. 基于图神经网络的推荐系统中的偏差

基于图神经网络的推荐系统网络中，能够强化既有偏差的特定敏感属性，引发了人们对如何衡量监督方法公平性的关注。为此使用的指标旨在使受保护群体中被判定为正向的敏感属性值比例与未受保护群体相同 [14,53]。

用户–项目交互行为并未明确包含用户的任何敏感信息，但由于用户与其属性之间存在很强的相关性，直接应用现代用户和物品表示学习可能导致用户敏感信息泄露 [14]。此外，考虑到推荐系统基于图的特性，用户并非彼此独立；他们与其他行为相似的用户存在隐式关联，这可能给先前的模型及协同过滤推荐的基础带来严重问题 [14]。此外，当前图神经网络算法受到数据中社会偏见的影响，从而限制了模型的泛化能力 [15]。在图结构中，具有相似敏感属性的节点往往彼此相连；由于具有相似敏感信息的节点与具有其他敏感特征的节点之间存在表示差异，这一特性可能导致决策中出现严重偏差 [15]。一些方法还考虑了在线社交网络中使用的图嵌入方法。图嵌入方法是数据挖掘的优秀工具之一，它将每位用户映射为一个低维向量，其中包含网络的结构信息。这些信息可能包括用户的邻域、流行度等 [53]。根据先前的研究，在线社交网络存在歧视，偏袒多数群体，这违反了反歧视法 [53]。

[54] 中介绍的研究重点是校准基于会话的推荐中的长尾问题，这类推荐可分为基于循环神经网络的模型和基于图神经网络的模型。该研究使用不同指标评估模型（例如，平均倒数排名和召回率），并通过覆盖率和长尾覆盖率等指标衡量流行度偏差。此外，研究还提出了一种校准模块，利用会话表示预测推荐列表中长尾物品的比例。研究还采用了包含两个阶段的课程训练策略，以提高校准模块预测的准确率。

[55], 中对推荐系统的基于图的协同过滤方法进行了调查。该研究考虑了当前使用的基于图的协同过滤方法在准确率和新颖性方面的双重性能。结果表明，大多数基于图的协同过滤模型中的对称邻域聚合会加剧推荐系统中的流行度偏差。此外，随着图传播深度的增加，这种偏差加剧现象也会进一步扩大。

鉴于关于基于图神经网络的推荐系统中偏差与公平性问题的研究十分有限，该领域的大多数研究都聚焦于推荐系统中的敏感信息。 [14] 中的研究致力于消除表示学习中的敏感信息，从而实现公平表征学习，促进公平推荐。为解决这一问题，研究者提出了一种用于公平性感知表示学习的模型无关图视角。该模型以任意推荐模型生成的用户和物品嵌入作为输入，并定义了一个敏感特征集。此外，该模型还充当过滤器，在不损害推荐准确率的前提下，模糊该集合中定义的任何敏感信息。在这种结构中，每个用户都可以被视为一个自我中心图结构，从而使过滤器能够在基于图的对抗训练过程中发挥作用。判别器旨在预测相关属性，而过滤器的训练则着眼于移除可能从以用户为中心的图结构中泄露的任何敏感信息。研究者在两个真实世界数据集上评估了该模型，结果显示其性能优异。

[15] 旨在克服敏感信息有限时基于图神经网络的推荐系统公平性面临的两大挑战：第一，如何通过弥补敏感属性不足来消除歧视；第二，如何确认推荐系统中的图神经网络分类器是否公平。为解决上述问题，研究提出了一种名为 FairGNN 的新方法，用于公平的节点分类。FairGNN 使用图神经网络中的敏感属性估计器，在存在对手的环境中，基于不同数据集预测带有噪声的敏感属性，从而实现公平分类。在真实世界数据集上进行的实验结果表明，所提出的模型在公平性和分类性能方面均表现有效。

另一项研究 [53] 聚焦于量化并解决图嵌入方法中的公平性问题，并将 node2vec 方法用于基于图神经网络的推荐系统。为解决这一问题，本文提出了一种研究 node2vec 算法公平性的新方法。此外，研究扩展了统计均等方法（该方法利用用户对的敏感属性衡量群体公平性），并提出了表征平等这一新理念，用于计算友谊推荐系统中的公平性。随后，研究将 node2vec 方法应用于真实世界的在线社交网络数据集，以发现不公平的图嵌入所导致的推荐偏差。最后，研究在 node2vec 的基础上提出了一种名为 Fairwalk、具备公平性感知能力的新型图嵌入算法。

文献 [56] 解决了基于图神经网络的推荐系统中的曝光偏差问题。本文提出了一种采用逆倾向方法的邻居聚合方法。该方法通过获取图中每次交互对应的用户—物品倾向得分，平衡每个目标节点存在偏差的局部结构；随后，将经过拉普拉斯归一化的逆倾向得分用作邻居聚合过程中的边权重。这使嵌入能够突出较不受欢迎的邻居。结果表明，该去偏方法效果显著，因此提升了模型的性能。

鉴于推荐系统中提到的这些挑战，偏差放大是最需要考虑的重要议题之一。此外，基于图神经网络的结构可以提高模型的准确率，但也可能使偏差问题更加严重。这一问题显然与上述准则相悖，需要进一步研究。另一方面，基于图神经网络的推荐系统中的性别等敏感属性十分重要，因此也需要考虑对这些属性的公平性。

3. 实验研究

本研究旨在探讨基于图神经网络的方法应对影响推荐系统的不同类型偏差时的表现。本实验重点关注流行度偏差和性别偏差。我们利用实验结果判断：这种推荐方法在大多数情况下是否能提升性能，同时加剧偏差。为此，我们在两个真实世界数据集（MovieLens 和 LastFM）上，为推荐系统实现了三类不同的方法：协同过滤（CF）、矩阵分解（MF）和基于图神经网络的方法。每类方法均采用不同的具体模型：深度矩阵分解（DMF）、基于物品的K近邻（ItemKNN）、神经协同过滤（NeuMF）、基于交互邻域的神经协同过滤模型（NNCF）、神经图协同过滤（NGCF）、用于推荐的轻量级图卷积网络（LightGCN），以及自监督图学习（SGL）。我们根据衡量性能和偏差的不同指标对结果进行了比较。

3.1. 基准数据集

本研究使用两个真实世界数据集来实现推荐系统。这两个数据集中的物品都存在前文提到的长尾效应，即数据中存在流行度偏差，因此非常适合用于偏差研究。下文将介绍这两个数据集。

3.1.1. MovieLens 100K [57]

MovieLens 是推荐系统领域中广泛使用的数据集之一，尤其常用于偏差研究。该数据集从非商业性的基于网络的电影推荐系统 MovieLens 网站逐步收集，并经过随机抽样。数据集包含用户对电影的评分，采用的星级量表范围为 [1, 5]。在该网站提供的不同规模的数据集中，本项目使用了 ml-100 K 数据集，其中包含 10 万条评分记录。此外，MovieLens 由三个不同的文件组成，分别包含用户信息、物品信息，以及用户对物品给出的评分。敏感特征位于

用户数据集中，根据 capAI 的指导，[58], “性别”和“年龄”被认定为敏感特征。该数据集还存在物品流行度偏差，也就是说，与其他电影相比，热门电影获得了更多评分。上述原因使该数据集非常适合用于研究偏差问题。MovieLens 数据集的信息见表1。

表1. MovieLens 数据集信息。

特征	描述	数据类型	计数	Mean	Std
Age	用户年龄	int	100 K	32.96	11.56
评分	用户对电影的评分	浮点数	10 万	3.52	1.12
用户 ID	用户 ID	int	10 万	-	-
电影 ID	电影 ID	int	100 K	-	-
性别	用户的性别	字符串	100 K	-	-
职业	用户职业	字符串	10 万	-	-
电影标题	已评分电影的标题	字符串	100 K	-	-

本研究提供了探索性数据分析（EDA），以便更好地了解这两个数据集。长尾效应见图 1。

图 1 显示，MovieLens 数据集中物品评分的分布高度集中于热门物品。这意味着该数据集存在物品流行度偏差。该数据集包含1682个物品，其中87.3%属于长尾部分。

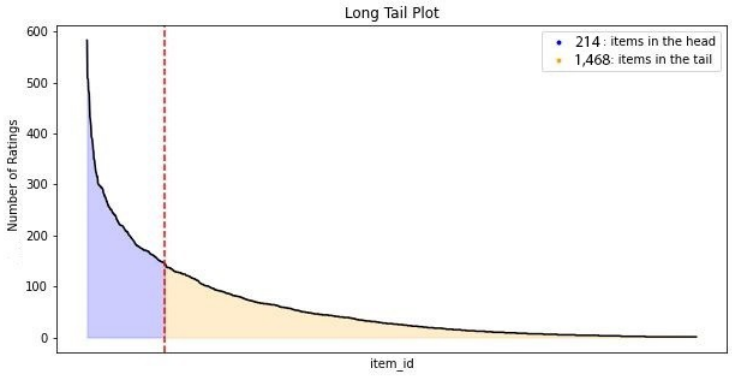


图1. MovieLens长尾图。

图2 表示各评分区间的评分数量。评分区间依据用户对物品的 1 至 5 分评分划分。该图展示了平均评分在 1 至 5 分之间的分布。

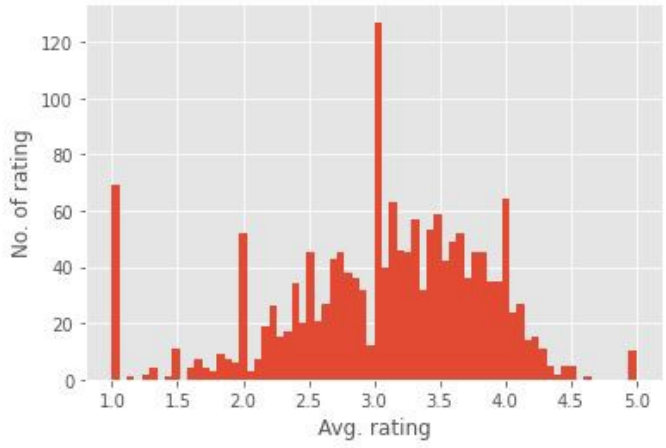


图 2. MovieLens平均评分图。

图 3 展示了年龄和性别的分布情况，表明用户大多是年轻人，且男性评分的物品数量明显多于女性。

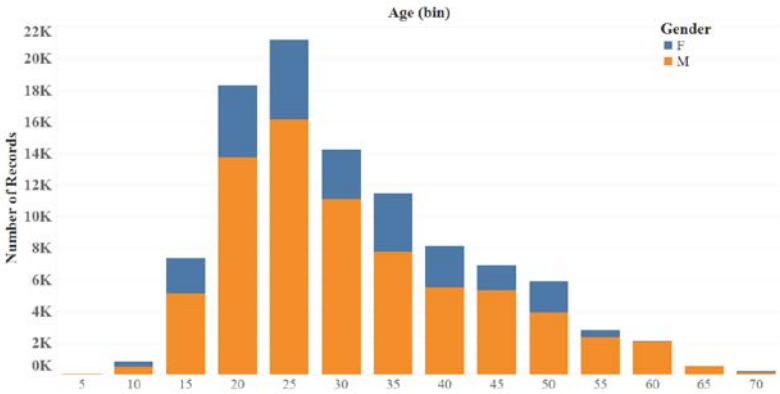


图 3. MovieLens 年龄/性别分布。

图4 显示男性用户的数量远高于女性用户。

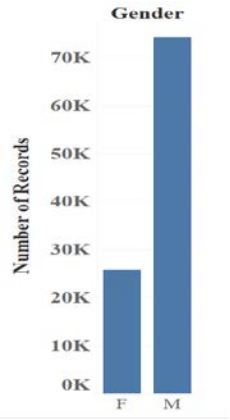


图 4. MovieLens 性别分布。

图 5 展示了根据用户职业划分的评分项目数量。可以看出，数据集中用户的主要职业是学生。

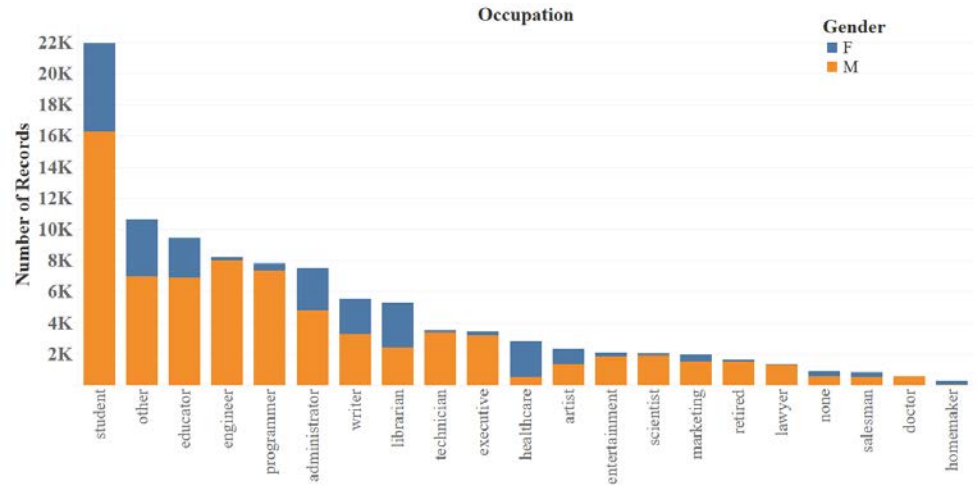


图 5. MovieLens 职业分布。

3.1.2. LastFM [59]

LastFM 数据集也广泛应用于推荐系统，尤其是在处理流行度偏差问题时。该数据集是规模最大的 数据集之一，包含来自世界各地的用户和艺术家信息。该数据集展示了每位用户收听每位艺术家的次数。LastFM 数据集包含用户和艺术家的特征，以及他们之间的交互。根据 capAI 指南 [58]，，性别可被视为该数据集中的敏感特征。对该数据集的分析显示，热门物品的评分显著高于其他物品。LastFM 数据集的信息和详情见表 2。

表2. LastFM 数据集信息。

特征	描述	数据类型	计数	Mean	Std
权重	每位艺人的聆听次数	浮点数	100 K	745.24	3751.32
用户 ID	用户 ID	int	100 K	-	-
项目 ID	艺术家 ID	int	100 K	-	-
性别	用户年龄	字符串	100 k	-	-
国家	用户所在国家	字符串	100 K	-	-
Name	艺术家名称	字符串	100 K	-	-

在下一节中，我们将对 LastFM 数据集进行探索性数据分析。

图 6 展示了 LastFM 数据集中艺术家评分的分布情况。图中显示，由于评分集中在少数物品上，因此存在显著的流行度偏差。具体而言，数据集中的 41,269 个物品中，有 90.3% 属于长尾。

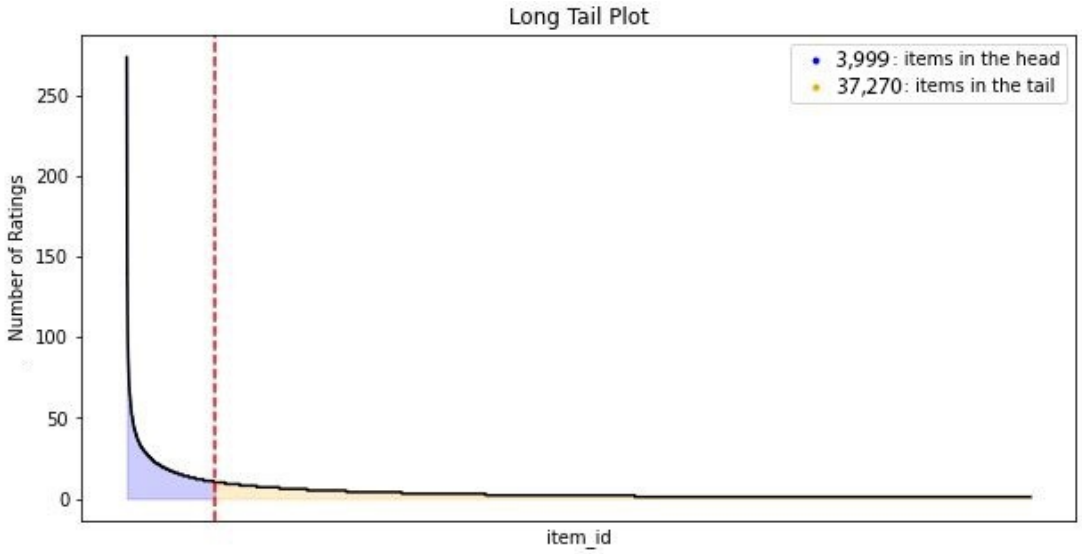


图 6. LastFM 长尾分布。

图 7 展示了 LastFM 数据集的性别分布。与 MovieLens 数据集一样，男性用户的数量远高于女性用户。

图 8 展示了按记录数量排名的最受欢迎的十位艺术家。

图 9 展示了记录在世界地图上的分布情况。可以看出 n 大多数记录都来自美国。

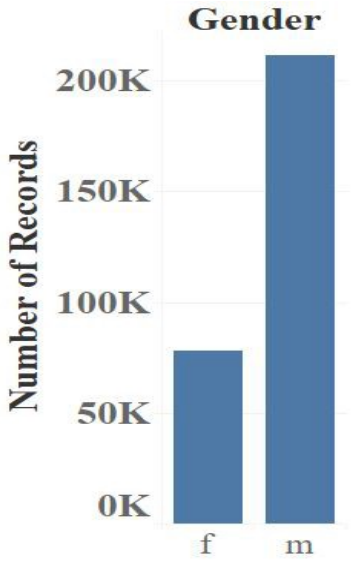


图7。 LastFM性别分布。

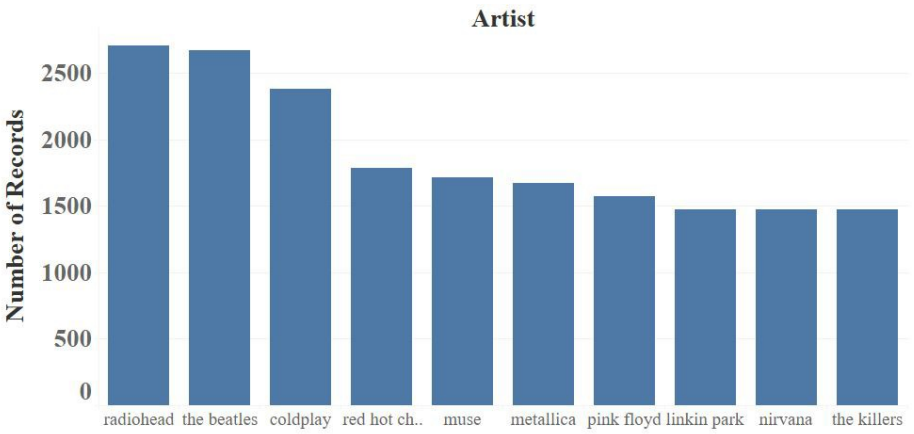


图8。 LastFM排名前十的艺术家。

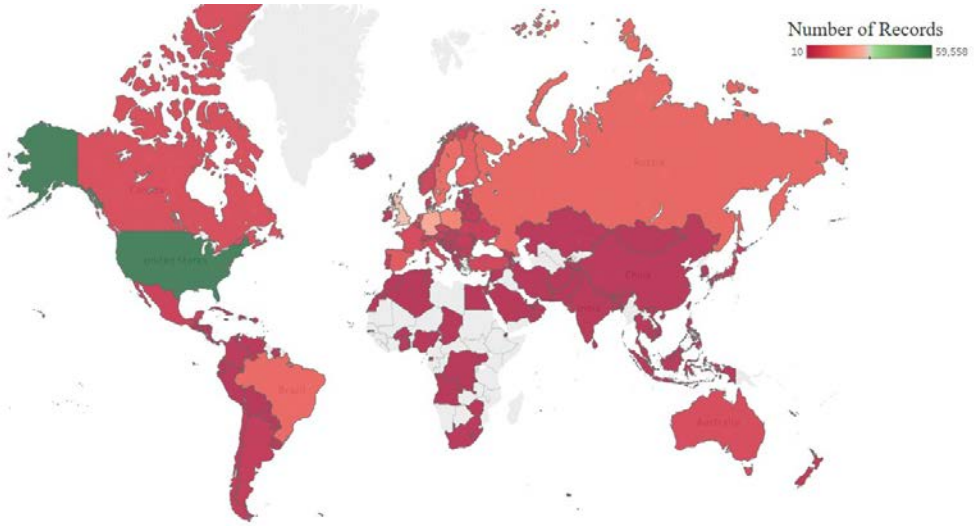


图9。 LastFM地图上的唱片分布。

3.2. 推荐方法

在本研究中，我们为推荐系统采用了三种推荐方法：协同过滤（CF）、矩阵分解（MF）和基于图神经网络的推荐方法。我们还针对每种方法实现了不同的模型，以收集更广泛的结果，从而得出更可靠的结论。实现不同的模型有助于我们进一步分析偏差问题；因此，应比较不同模型，以详细考察偏差放大情况。以下介绍这些方法及选用的模型：

1. 协同过滤：

协同（或社交）过滤技术基于用户偏好，并利用用户对物品的评分信息来计算用户或物品之间的相似度。如果两位用户以相似的方式为物品评分，那么他们很可能也会以类似的方式为新物品评分。因此，目标用户是这样一类用户：与其品味相同的用户对其推荐的物品给出了较高评分。另一种协同过滤策略是推荐与用户已消费或给予正面评价的物品相似的物品；这种物品间的相似度是根据用户给出的评分计算得出的。在协同过滤方法中，会收集用户对物品的评分，以构建用户—物品评分矩阵。该矩阵用于寻找用户或物品之间的相似度。协同过滤方法可以解决基于内容的方法中的一些问题：基于内容的方法也会推荐与用户已消费或评价较高的物品相似的物品，但这种相似度是根据物品特征计算的。基于内容的方法的一个缺点是可能无法获取物品特征，或难以获得这些特征；而协同过滤可以避免这一问题，因为它仅利用其他用户的反馈进行推荐。此外，这些技术的质量更高，因为它们基于用户评价过的物品，而不依赖质量可能较低的内容。与基于内容的系统不同，协同过滤方法可以推荐内容多样、与用户先前消费过的物品并不相似的物品，只要其他用户已经表现出对这些不同物品的兴趣即可。协同过滤技术采用多种方法，包括：

- 基于用户：这类系统利用目标用户的邻居（即评分行为相似的用户）对某物品给出的评分，评估目标用户对该物品的偏好 [4]。
- 基于物品：这类方法根据用户对相似物品给出的评分，预测该用户对某物品的评分。在这类方法中，如果系统中的多个用户对两个物品给出的评分相似，那么这两个物品就被视为相似物品。这与基于内容的方法不同，后者根据物品的特征或属性来确定物品之间的相似性。对于常见的商业推荐系统，这种方法更为适用，因为目录中的用户数量远高于物品数量。通常，基于物品的方法更可靠、计算时间更短，而且无需频繁更新 [4]。

图 10 展示了基于用户的方法与基于物品的方法之间的区别。所采用的协同过滤方法如下：

- 基于物品的K近邻：该方法是一种基于物品的方法，根据用户对物品的评分计算物品之间的相似度。该方法的主要依据是，顾客更倾向于购买与自己以往购买的物品相匹配的物品。用户-物品矩阵中的历史购买信息有助于识别相似物品集合，并据此生成 Top-K 推荐。从整体上看，该算法包含两个主要部分。第一部分构建模型，以捕捉各种物品之间的关系。第二部分将上述计算得到的模型应用于为用户获取 Top-K 推荐。与其他类似的协同过滤方法相比，该方法也表现出较高的性能 [60–62]。

• 基于交互邻域的神经协同过滤模型（NNCF）：该模型利用深度学习对用户与物品之间复杂的交互进行建模。该新型模型还利用邻域信息补充用户–项目交互数据，从而提升模型性能。NNCF 模型能够克服传统算法的不足，例如简单的线性分解可能无法充分支持用户与物品之间的复杂交互。该方法还能够提供质量较高的用户/项目嵌入 [63–65]。

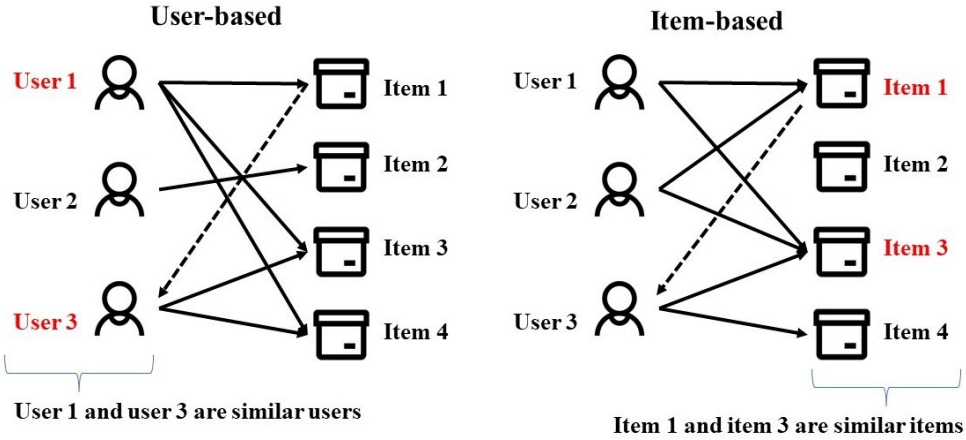


图10. 基于用户的方法与基于物品的方法之间的差异。

2. 矩阵分解：

矩阵分解包含一组基于模型的技术，这些技术将评分矩阵转换为两个分别表示用户和物品的潜在因子矩阵，以解决评分矩阵的稀疏性问题。这是一种低维因子模型，其假设用户和物品潜在因子之间的内积会影响用户对某一物品的偏好 [66]。目前，矩阵分解已成为实现推荐系统最流行的方法之一 [67]。

所使用的矩阵分解方法如下：

• 深度矩阵分解（DMF）：该方法采用神经网络架构，利用显式评分和表示非偏好的隐式反馈构建用户-物品矩阵。随后，将该矩阵作为输入，以便为深层结构学习架构学习一个共同的低维空间。该方法还采用一种基于二元交叉熵的新型损失函数，同时考虑显式评分和隐式反馈，以增强优化效果。通过应用隐式反馈，DMF 相较于传统模型能够提供更好的 Top-K推荐，并通过利用显式历史评分学习隐藏结构来重建用户评分。该方法还支持双通道结构，可结合来自用户和物品的侧信息。一些文章还指出，在准确率和训练效率方面，该方法优于新型推荐算法 [68–70]。

• 神经协同过滤（NeuMF）：鉴于协同过滤模型中最重要的因素是用户特征与物品特征之间的交互，这些方法中的内积可以由神经网络架构取代。基于神经网络的协同过滤（NCF）是一种表达并推广矩阵分解的框架，可通过使用非线性核加以增强。为此，可以使用多层感知机来学习用户–项目交互函数 [71]。深度神经网络良好的建模能力和非线性是其性能出色的主要原因。此外，

NeuMF 中通用的 NCF 使我们有机会结合使用各种模型 [67]。

3. 基于图神经网络：

近年来，增长最快且潜力巨大的技术之一是图学习（GL）[14]。这种方法涉及应用于图结构数据的机器学习。利用这些优势来学习关系数据，研究人员提出了基于图学习的推荐系统（GLRS）[6]。事实上，大多数对象之间存在显式或隐式的关联。这些关系可以通过图来表示。在推荐系统中，对象可被视为用户、物品、属性和上下文，这一特征尤为明显。

这些对象彼此紧密关联，并通过不同的关系相互影响。运用图技术可以显著提升推荐系统的质量。图学习具有学习复杂关系的强大能力，并且在从各种图中获取蕴含的知识方面潜力巨大[72]。

推荐系统中包含不同类型的实体，包括用户、物品和属性，它们彼此之间存在不同类型的关系，因此可以用不同类型的图来表示。众所周知，推荐模型使用的三类主要对象是用户、物品和用户–项目交互，不过也可能使用有关用户和/或物品的其他信息。据此，推荐系统使用的数据可分为两大类：用户–项目交互数据（用户对物品的点击、购买或评分）和侧信息数据（用户和项目属性）。此外，根据交互是序列式还是一般性交互，交互数据还可分为两类。[72]每一类还可进一步划分为多个子类，如表3所示。

表3。 推荐系统及其表示图中的数据表示总结 [72]。

数据类别	数据子类别	表示图
一般交互	显式交互、隐式交互	加权二部图，无权二部图
序列交互	单类型交互 多类型交互	有向同质图，有向异质图
辅助信息	属性信息，社交信息、外部知识	异质图，同质图树，或异质图

用户-物品矩阵中的每个输入都包含用户与物品之间发生的交互类型信息。交互数据可根据类型分为显式交互和隐式交互。显式交互是指要求用户对物品发表意见（例如，用户对物品的评分）。隐式交互则是从用户的行为（例如，点击、浏览）中推断出来的 [72,73]。

所使用的图神经网络方法如下：

• LightGCN：该模型是图卷积网络（GCN）的一个简化版本，保留了 GCN 中对推荐任务最重要的组成部分。LightGCN 在用户-物品交互图上传播用户和物品嵌入。随后，该模型将所有层学习到的嵌入加权求和，作为最终嵌入 [74]。LightGCN 中的对称归一化与标准 GCN 相同，用于控制图卷积操作导致的嵌入规模增长。与传统方法相比，该方法也展现出优异的性能 [75,76]。

• 神经图协同过滤（NGCF）：本研究还选择了这种模型作为一种方法。该模型将图结构引入用户–物品交互中。该方法通过在用户–物品图结构上生成嵌入，从中获益，进而在

用户–物品图中形成高阶连通性。协同信号以显式方式注入嵌入过程 [72]。此外，该方法使用多个嵌入传播层，并将各层输出拼接起来，以生成推荐任务的最终预测。NGCF 在模型优化方面也表现出色 [77]。

- 自监督图学习（SGL）：该方法是对 GCN 模型的改进版本，在准确率和鲁棒性方面表现更佳。这些模型在处理含噪交互时也表现更好。该方法在经典的有监督推荐任务基础上加以改进，并辅以自监督任务，通过自我判别强化节点表示学习。这种结构会为节点生成不同视图，并最大化同一节点不同视图之间的一致性，使其高于该节点与其他节点之间的一致性。此外，还设计了三种算子来生成上述视图——节点丢弃、边丢弃和随机游走，它们从不同方面改变图结构。SGL 方法在推荐系统任务中也展现出优异性能，因此适用于本实验 [78–80]。

3.3. 评估指标

为了从推荐可靠性以及对偏差的敏感性两个角度评估采用前述方法实现的模型，我们同时使用了性能指标和用于衡量偏差放大的指标。考虑有一组待排序的n 个物品，并给定物品集合I 和用户集合U，$R\hat{}(u)$用于表示模型生成的物品排序列表，而$R(u)$表示用户u 曾与之交互的物品真实集合。对于 Top-K 推荐，只需考虑排名靠前的物品。因此，在 Top-K 评估场景中，我们将推荐列表截断为长度为K的列表。

为了更好地理解这些指标，下面在表 4中提供了一个符号表。

表 4. 符号表。

符号	定义
U	用户集合
V	物品集合
u	一位用户
v	一件物品
$R(u)$	用户 u 交互过的物品真实集合
$R\hat{}(u)$	模型生成的物品排序列表
K	推荐列表的长度
$M(x)$	输入为x 、输出为y的推荐系统算法机制
θ	生成x的分布
Θ	生成每个x实例的一组θ分布

表 5 展示了用于评估推荐可靠性的主要指标。与当前大多数推荐系统一样，评估基于Top-K物品推荐列表进行，其中K表示列表的大小。

表6 展示了本研究用于评估模型对推荐系统中最相关偏差类型的敏感性的指标。在本研究中，我们主要关注与推荐物品的流行度和多样性相关的偏差放大。为此，我们使用了三种不同的指标（平均流行度、基尼指数和物品覆盖率），如表6所示。此外，我们还采用差异公平性（DF）指标评估性别偏差，该指标也在表6中进行了说明。该指标旨在评估推荐算法对受保护属性的所有取值是否表现一致，或是否会对某些取值在输出中产生一定程度的偏差。在本研究中，我们考察了受保护属性“性别”的“男性”和“女性”两种取值。这意味着，预测的

物品得分在该属性的两种取值下应当相近。DF 特别关注交叉性[81,82]，，其中包括分别保障每一种受保护属性的公平性，以及对少数群体的行为；后者与前文提到的反歧视法相关。

表5. 性能评估指标。

指标名称	描述
MRR	计算由某个系统检索到的第一个相关项目的倒数排名算法。将排名$_u^*$视为第一个相关算法为用户 u 找到的条目所处的排名位置。 平均倒数排名@K = $\frac{1}{\|U\|}\sum_{u\in U}\frac{1}{\text{排名}_u^*}$
NDCG	是一种衡量排名质量的指标，排名位置按对数方式折扣，以此计算。它通过为命中位置赋予更高的排名靠前的分数转化为命中。$\delta(0)$ 是一个指示函数。 归一化折损累计增益@K = $\frac{1}{\|U\|}\sum_{u\in U}\frac{1}{\min(\|R(u)\|,K)}\sum_{i=1}^{K}\frac{\delta(i\in R(u))}{\log_2(i+1)}$ $\frac{1}{2^1+1}$ $\log(i)$
精确率	是用于计算相关项占所有推荐物品比例的指标。其最终值是每位用户的指标值的平均值 $\|\hat{R}(u)\|$ 表示$\hat{R}(u)$的物品数量。 $Precision@K = \frac{1}{\|U\|}\sum_{u\in U}\frac{\|\hat{R}(u)\cap R(u)\|}{\|R(u)\|}$
召回率	是用于计算相关项占所有相关项比例的指标。$\|R(u)\|$ 表示$R(u)$的物品数量。 $Recall@K = \frac{1}{\|U\|}\sum_{u\in U}\frac{\|\hat{R}(u)\cap R(u)\|}{\|R(u)\|}$
命中率（命中）	此指标也称为截断命中率。它用于计算包含在长度为 K 的排序项列表中的“命中”数量。如果至少有一个物品属于真实集合，我们就称其为一次命中。$\delta(0)$ 是指示函数。$\delta(b) = 1$ 如果b 为真，则为 1；否则为 0。$\varnothing$ 表示空集。 $HR@K = \frac{1}{\|U\|}\sum_{u\in U}\delta(\hat{R}(u)\cap R(u) \neq \varnothing)$

表6。 偏差测量指标。

指标名称	描述
平均流行度	计算推荐物品的平均流行度。在在下方公式中，$\varphi(i)$ 是物品 i 的交互次数在训练数据 [83] 中。 $$\text{平均流行度@}K = |U| \sum_{u \in U} \frac{1}{\frac{\sum_{i \in R_u} \varphi(i)}{|R_u|}}$$
基尼指数 [41]	表示推荐项的多样性。它用于衡量分布的不平等程度。在下列公式中，$P(i)$ 表示物品 i 出现的次数在推荐列表中，该列表按非递增序索引 [84]。 $$GiniIndex@K = \frac{\sum_{i=1}^{|I|}(2i - |I| - 1)P(i)}{|I| \sum_{i=1}^{|I|} P(i)}$$
物品覆盖率	计算推荐物品相对于所有物品的覆盖率 [85] $$ItemCoverage@K = \frac{|\cup_{u \in U} \hat{R}(u)|}{|I|}$$
差异公平性（DF），针对敏感属性性别 [81,86]	确保所有受保护群体都受到无偏差的对待。该指标也表示差异的隐私解释。机制 $M(x)$ 相对于 (A, θ) 满足 e-差分公平性，对所有满足条件的 $\theta \in \Theta$ 均成立，其中 $x \sim \theta$ 且 $y \in Range(M)$。对于所有 $(s_i, s_j) \in A \times A$，其中 $P(s_i) > 0$，$P(s_j) > 0$。s_i、$s_j \in A$ 是由所有受保护属性值构成的元组（此处为男性和女性）。 $$e^{-e} \leq \frac{P_{M,\theta}(M(x) = y | s_i, \theta)}{P_{M,\theta}(M(x) = y | s_j, \theta)} \leq e^{e}$$

4. 结果

本节展示上一节所述推荐方法在 MovieLens 和 LastFM 数据集上得到的结果。这些结果是通过应用上述指标得出的。研究旨在评估不同的推荐方法，确定哪些方法在性能与对偏差的敏感性之间实现了最佳平衡，因为前者的提升通常会导致后者恶化。结果还将用于判断“基于图神经网络的方法会产生更可靠、但偏差也更大的模型”这一假设是否成立。若该假设得到证实，则说明有必要进一步研究如何应对这些方法中的偏差放大问题。

在表7 和8中，可以看到所提及模型在 Movie- Lens 和 LastFM 数据集上的结果。此外，还提供了不同的指标，以便更好地了解各模型的性能及其偏差放大情况。

表 7 展示了 LastFM 数据集上的结果。

表 7. MovieLens 数据集上的结果。

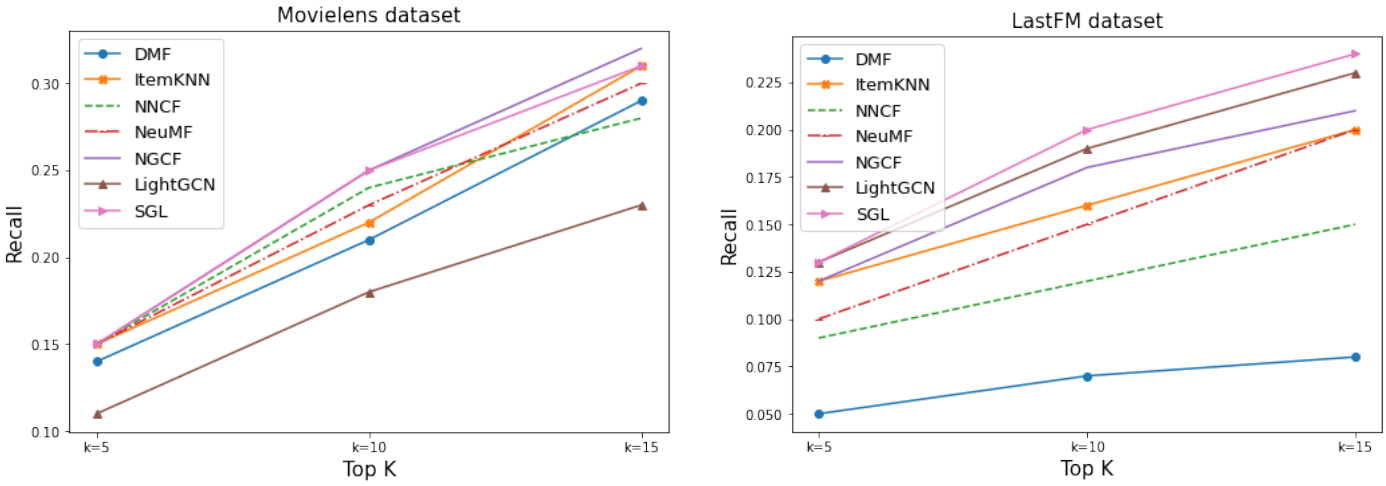
方法	方法	前K项	召回率	精确率	MRR	NDCG	命中	物品覆盖率	基尼指数	平均流行度
MF	DMF	K = 5	0.14	0.22	0.43	0.26	0.62	0.18	0.94	256.29
MF	DMF	K = 10	0.21	0.17	0.42	0.25	0.73	0.20	0.93	252.25
MF	DMF	K = 15	0.29	0.16	0.45	0.28	0.83	0.28	0.90	219.49
MF	NeuMF	K = 5	0.15	0.23	0.45	0.27	0.65	0.25	0.91	228.52
MF	NeuMF	K = 10	0.23	0.18	0.46	0.27	0.78	0.36	0.89	212.41
MF	NeuMF	K = 15	0.30	0.16	0.46	0.28	0.83	0.40	0.86	196.89
CF	基于物品的K近邻	K = 5	0.15	0.23	0.44	0.28	0.63	0.19	0.93	231.96
CF	基于物品的K近邻	K = 10	0.22	0.18	0.46	0.27	0.75	0.24	0.93	249.74
CF	基于物品的K近邻	K = 15	0.31	0.16	0.46	0.29	0.84	0.29	0.89	208.12
CF	NNCF	K = 5	0.15	0.24	0.47	0.29	0.64	0.17	0.95	284.47
CF	NNCF	K = 10	0.24	0.19	0.46	0.22	0.78	0.25	0.91	217.70
CF	NNCF	K = 15	0.28	0.15	0.47	0.27	0.81	0.30	0.91	231.28
GNN	NGCF	K = 5	0.15	0.24	0.48	0.29	0.66	0.15	0.95	277.85
GNN	NGCF	K = 10	0.25	0.20	0.49	0.30	0.77	0.25	0.93	255.49
GNN	NGCF	K = 15	0.32	0.17	0.49	0.31	0.86	0.32	0.89	219.13
GNN	LightGCN	K = 5	0.11	0.17	0.36	0.21	0.55	0.05	0.98	245.13
GNN	LightGCN	K = 10	0.18	0.14	0.37	0.21	0.67	0.07	0.97	312.47
GNN	LightGCN	K = 15	0.23	0.12	0.38	0.21	0.76	0.10	0.96	292.8
GNN	SGL	K = 5	0.15	0.25	0.47	0.29	0.66	0.24	0.91	229.24
GNN	SGL	K = 10	0.25	0.20	0.49	0.29	0.80	0.31	0.89	209.39
GNN	SGL	K = 15	0.31	0.17	0.49	0.30	0.85	0.34	0.88	200.63

表 8 展示了 LastFM 数据集上的结果。

表 8. LastFM 数据集上的结果。

方法	方法	前K项	召回率	精确率	MRR	NDCG	命中	物品覆盖率	基尼指数	平均流行度
MF	DMF	K = 5	0.05	0.05	0.11	0.05	0.20	0.01	0.99	377.81
MF	DMF	K = 10	0.07	0.03	0.12	0.06	0.25	0.02	0.99	341.85
MF	DMF	K = 15	0.08	0.02	0.12	0.07	0.30	0.02	0.99	309.64
MF	NeuMF	K = 5	0.10	0.10	0.25	0.12	0.40	0.05	0.98	167.49
MF	NeuMF	K = 10	0.15	0.07	0.27	0.14	0.52	0.06	0.98	157.12
MF	NeuMF	K = 15	0.20	0.06	0.27	0.16	0.60	0.09	0.98	140.17
CF	基于物品的K近邻	K = 5	0.12	0.11	0.29	0.14	0.41	0.12	0.96	152.64
CF	基于物品的K近邻	K = 10	0.16	0.08	0.30	0.16	0.50	0.23	0.93	131.54
CF	基于物品的K近邻	K = 15	0.20	0.06	0.30	0.18	0.57	0.31	0.91	118.00
CF	NNCF	K = 5	0.09	0.07	0.16	0.09	0.31	0.04	0.98	195.14
CF	NNCF	K = 10	0.12	0.06	0.17	0.10	0.38	0.05	0.98	185.23
CF	NNCF	K = 15	0.15	0.05	0.19	0.12	0.49	0.06	0.99	177.06
GNN	NGCF	K = 5	0.12	0.11	0.29	0.14	0.44	0.03	0.99	202.55
GNN	NGCF	K = 10	0.18	0.09	0.32	0.17	0.59	0.06	0.98	155.32
GNN	NGCF	K = 15	0.21	0.07	0.31	0.18	0.64	0.08	0.98	160.38
GNN	LightGCN	K = 5	0.13	0.14	0.31	0.15	0.47	0.05	0.98	174.23
GNN	LightGCN	K = 10	0.19	0.09	0.33	0.18	0.59	0.09	0.98	148.15
GNN	LightGCN	K = 15	0.23	0.07	0.34	0.20	0.66	0.12	0.97	132.52
GNN	SGL	K = 5	0.13	0.13	0.33	0.15	0.48	0.06	0.98	142.71
GNN	SGL	K = 10	0.20	0.10	0.35	0.19	0.62	0.10	0.97	114.49
GNN	SGL	K = 15	0.24	0.07	0.35	0.21	0.69	0.14	0.96	103.11

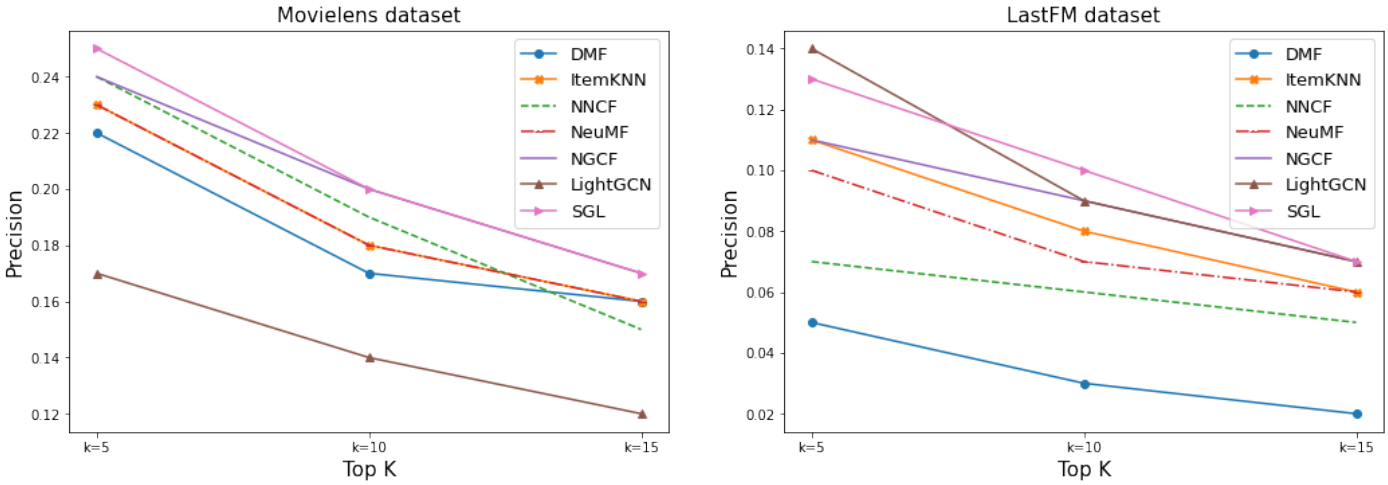
图 11 展示了在两个数据集上，所实现模型的三种不同长度 Top-K 列表对应的召回率。根据这些结果，三种基于图神经网络的方法中，有两种（NGCF 和 SGL）在两个数据集上的召回率指标表现更好，而第三种 LightGCN 在两个数据集上的表现则有所不同。



(a) (b)

图 11。 Movielens 和 LastFM 的召回率结果。**(a)** MovieLens 召回率。**(b)** LastFM 召回率。

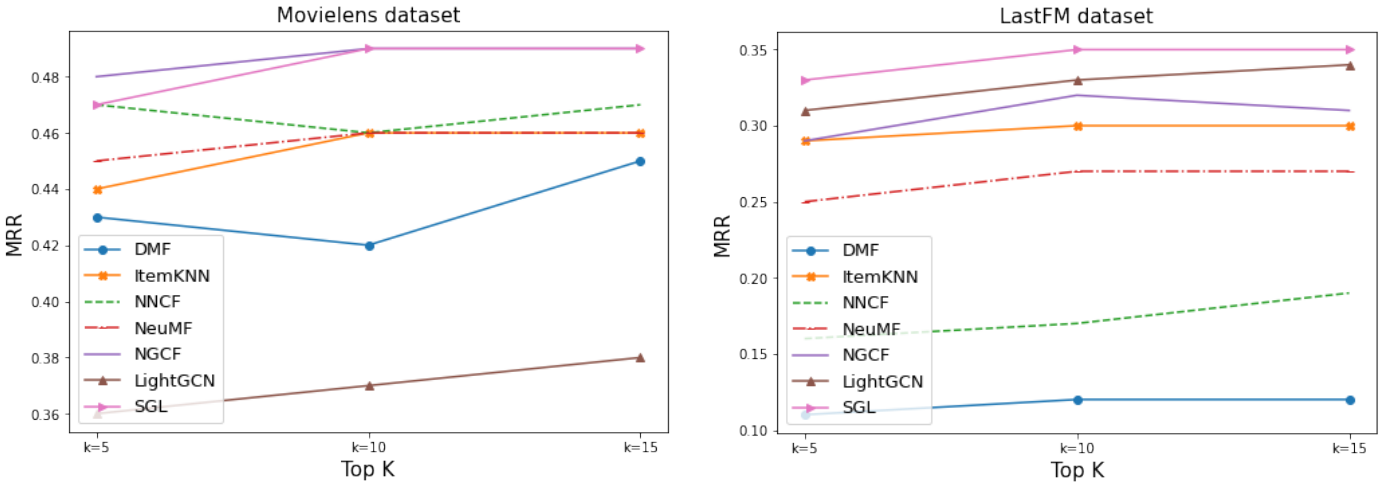
关于图中所示模型的精确率 12，结果相似。从该图可以看出，NGCF 和 SGL 在两个数据集上都提供了更高的精确率，但对于 LastFM 数据集，LightGCN 也是一个不错的选择。



(a) **(b)**

图12。 Movielens 和 LastFM 的精确率结果。**(a)** MovieLens 精确率。**(b)** LastFM 精确率。

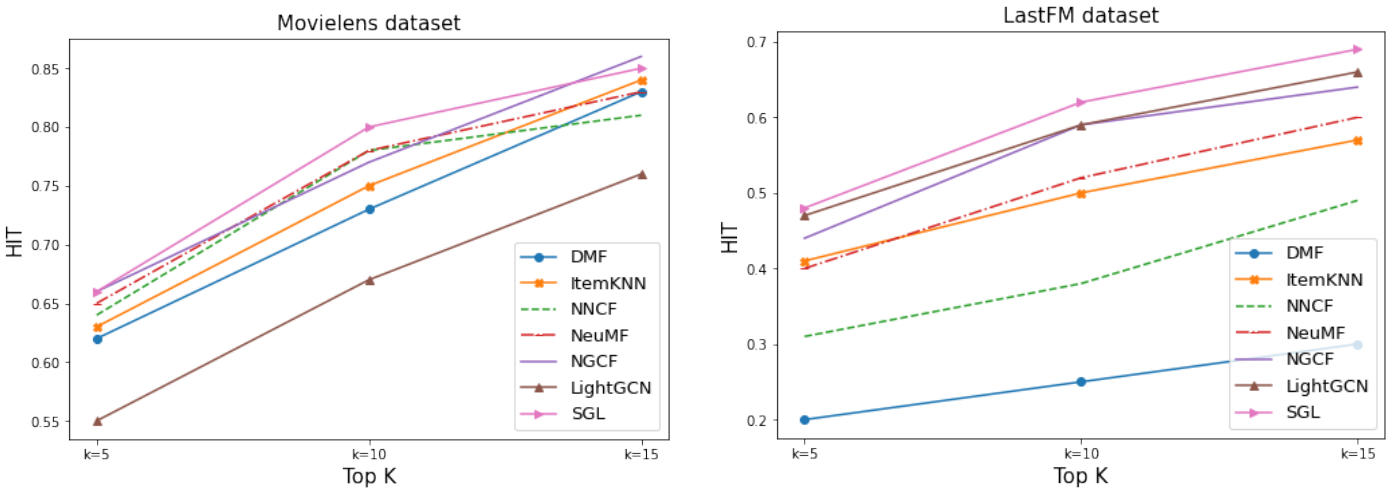
图 13 展示了平均倒数排名指标。结果表明，与之前的指标一样，NGCF 和 SGL 模型的平均倒数排名较高。对于 LastFM 数据集，LightGCN 的表现也较为合理。



(a) (b)

图13。Movielens 和 LastFM 的平均倒数排名结果。（**a**）MovieLens 平均倒数排名。（**b**）LastFM 平均倒数排名。

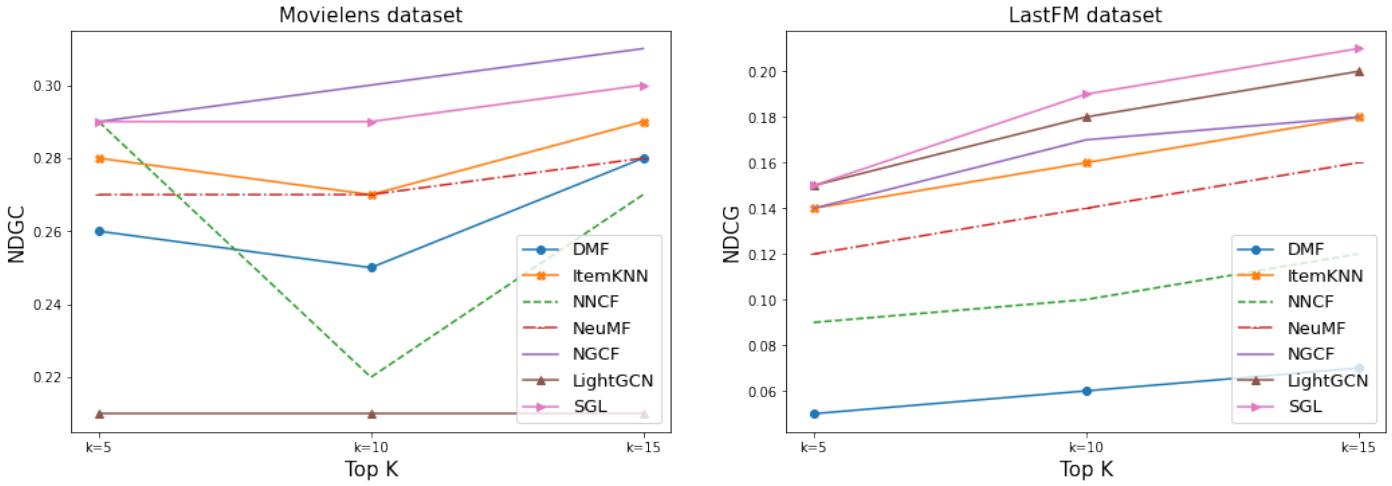
图 14 展示了命中度量。从这一指标来看，除了 LightGCN 表现相当差之外，所实现的大多数模型在 MovieLens 上的结果相近，而 SGL 和 NGCF 取得了最高值。另一方面，该指标表明，对于 LastFM 数据集，SGL、LightGCN 和 NGCF 这三种基于图神经网络的方法取得了较高的命中值。



(a) (b)

图14。Movielens 和 LastFM 的命中结果。（**a**）MovieLens 命中。（**b**）LastFM 命中。

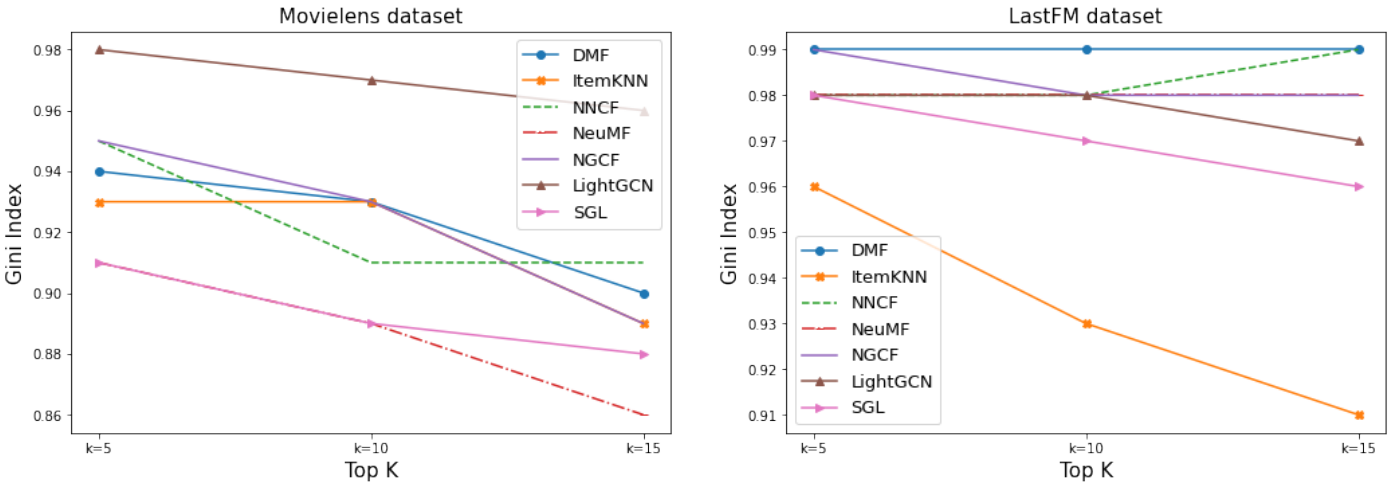
图 15 展示了所实现模型的归一化折损累计增益。我们可以从图中观察到，与其他方法相比，两种基于图神经网络的方法 NGCF 和 SGL 在 MovieLens 上取得了更好的结果。在 LastFM 数据集上，三种基于图神经网络的算法的归一化折损累计增益值均为最佳，尽管 LightGCN 的表现优于 NGCF。



(a) (b)

图15。MovieLens 和 LastFM 的归一化折损累积增益结果。（a）MovieLens 归一化折损累计增益。（b）LastFM 归一化折损累计增益。

在介绍了与性能指标相对应的结果后，我们转而讨论偏差指标。图 16 展示了给定模型的基尼指数。这些结果表明，SGL 等准确性更高的模型具有更低的基尼指数，这意味着更低的多样性和更严重的偏差放大。然而，与前述指标相比，测试模型在该指标上的结果差异并不显著，尤其是在 LastFM 数据集的情况下。

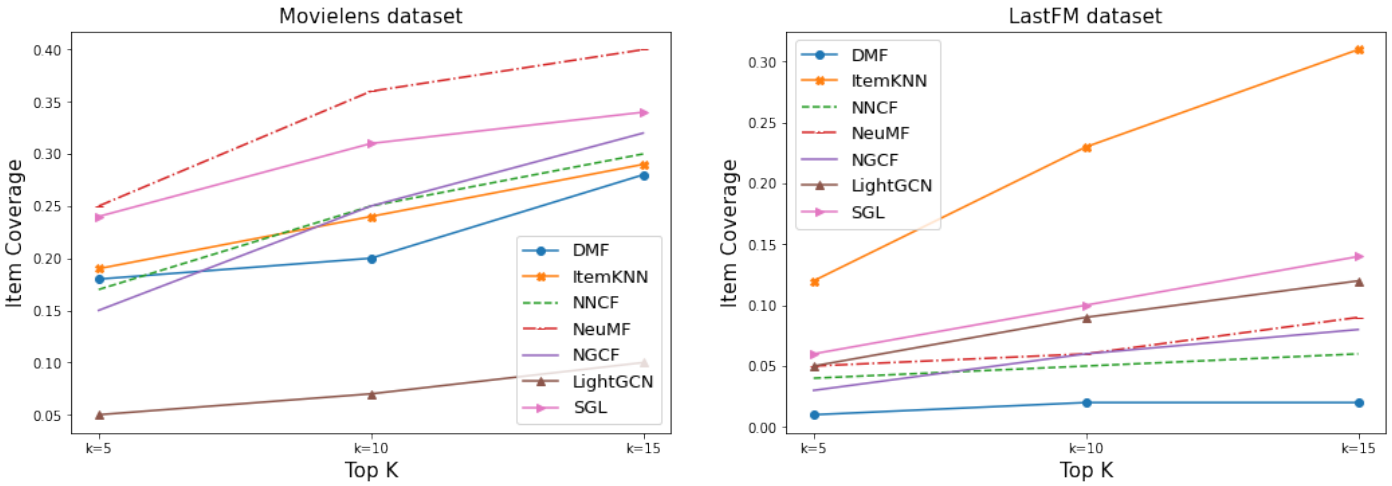


(a) (b)

图16。MovieLens 和 LastFM 的基尼指数结果。（a）MovieLens 基尼指数。（b）LastFM 基尼指数。

图17 展示了物品覆盖率。覆盖率低意味着系统对用户可能喜欢但未推荐的某些物品存在歧视。总体而言，MovieLens 的覆盖率优于 LastFM，这与后者数据集中的物品数量更多相一致：LastFM 数据集中有 41,269 件物品，而 MovieLens 中有 1682 件。可以看出，在 MovieLens 数据集中，NeuMF 的物品覆盖率最高，LightGCN 最低。相比之下，在 LastFM 数据集中，基于物品的K近邻的物品覆盖率最高。我们还可以看到，基于图神经网络的 SGL 方法在两个数据集的覆盖率排名中均位居第二，而其他方法的表现则因数据集而异。在两个数据集中，NGCF 的覆盖率与其他非图神经网络方法相近；而在

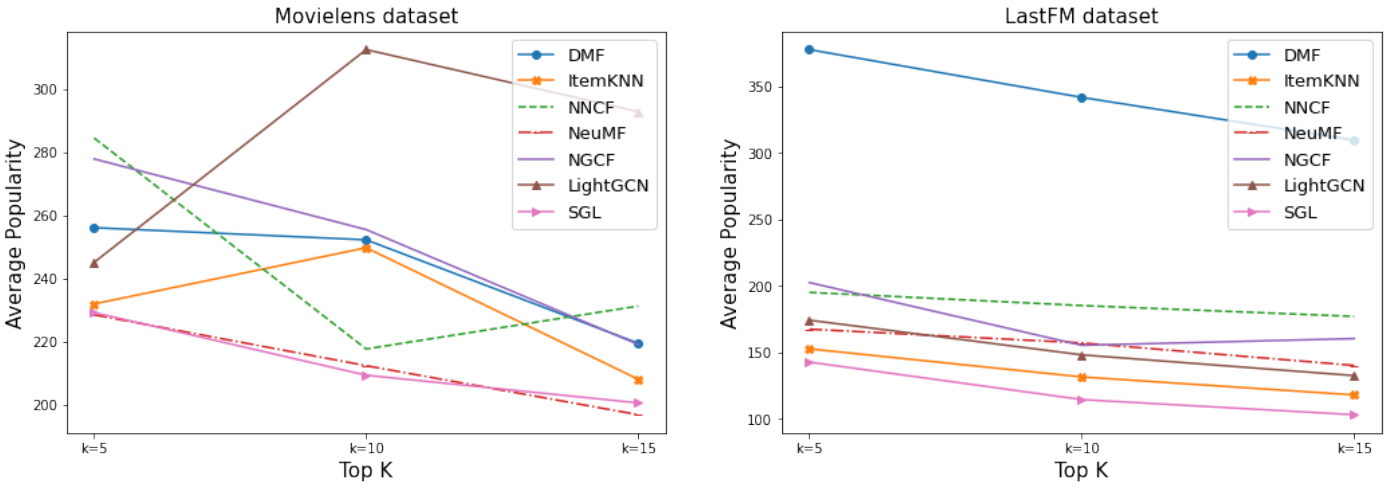
MovieLens 数据集中，对于 k = 10 和 K = 15，其表现优于大多数此类方法。这表明，基于图神经网络的方法在覆盖率方面表现尚可。



(a) (b)

图17. Movielens 和 LastFM 的物品覆盖率结果。（**a**）MovieLens 物品覆盖率。（b）LastFM 物品覆盖率。

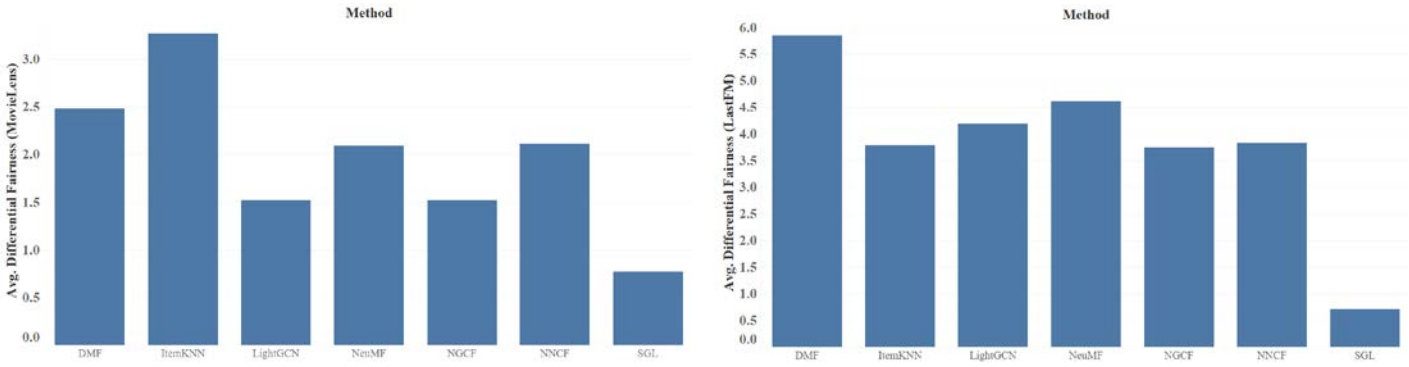
图 18 展示了两个数据集上物品的平均流行度。要尽量减少这种偏差，需要生成包含低流行度物品的推荐，因此该指标的值越低越理想。从这些结果中可以看出，各方法在 MovieLens 数据集上的表现并不一致：它们的数值差异很大，而且在不同 Top-K 列表长度下也有较大波动。结果表明，SGL 算法在两个数据集上的流行度值最低，而 NeuMF 在 MovieLens 数据集上的数值与之非常接近。在 MovieLens 数据集上，LightGCN 的性能最差；在 LastFM 上，表现最差的是 NNCF。NGCF 在两个数据集上的表现也逊于大多数经典方法。因此，这证实了基于图神经网络的方法会放大流行度偏差，但 SGL 除外；就流行度及许多其他指标而言，SGL 均优于其他方法。



(a) (b)

图18. Movielens 和 LastFM 的平均流行度结果。（**a**）MovieLens 平均流行度。（b）LastFM 平均流行度。

图 19 展示了敏感属性性别的差分公平性结果。需要注意的是，该指标的值越低，性别偏差越小。该指标适用于评分预测，因为它使用用户—物品对的评分预测结果；但不适用于 Top-K 推荐列表。因此，图中没有展示不同 k 值的结果，而是给出了一个单一数值，即测试集中所有样本的评分预测结果的均值。图中显示，在 LastFM 数据集上，DMF 存在显著的性别偏差；在 MovieLens 数据集上，基于物品的K近邻也存在显著的性别偏差。另一个值得注意的现象是，在 MovieLens 数据集上，基于图神经网络的方法表现优于其他方法；而在 LastFM 上，各方法的数值则较为接近。在这两种情况下，SGL 方法都优于其他方法，与前文讨论的结果一致。因此，不仅基于图神经网络的方法会受到性别偏差的影响，可靠性较低的其他方法也会如此。与所用的其他方法相比，甚至有些模型（如 SGL）表现更好。



(a) (b)

图19。MovieLens 和 LastFM 上敏感属性性别的差分公平性结果。(a) MovieLens 的平均性别差分公平性。（**b**）LastFM 的平均性别差分公平性。

综合分析结果可知，基于图神经网络的方法提供了更可靠的推荐，因为它们的性能指标值最高。然而，应用的偏差指标显示，使用这些方法时，某些类型的偏差会被放大，尤其是与多样性和覆盖率相关的偏差。在这一方法类别中，SGL 算法表现最佳，因为它可靠性良好，且受到偏差的影响不及该组中的另外两种算法。SGL 引入的用户-物品图自监督学习旨在解决基于图卷积网络的推荐系统所面临的一些问题（稀疏性、偏斜分布和噪声），本研究结果也表明，这种做法有助于提升算法应对不同类型偏差的表现。此外，该算法在两个数据集和所有列表长度下的表现最为一致。与之相反，LightGCN 对某些偏差高度敏感，其结果在不同数据集之间差异很大。

5. 结论与未来工作

偏差是推荐系统中最重要的问题之一，而且由于其会给公司和个人带来高昂的代价，这一问题值得深入研究。在本研究中，我们对基于图神经网络的推荐系统中的偏差放大进行了实证研究。我们研究了通过推荐 Top-K 物品列表向用户提供产品或服务的公司目前最为关注的偏差。此外，我们还通过计算敏感属性“性别”的差分公平性，对性别偏差进行了研究。除了列表应尽可能贴合用户的喜好之外，列表最受重视的要求还包括多样性、尽可能涵盖更多物品，以及

包含不受欢迎的物品，以拓宽并丰富面向用户的产品选择，并帮助他们发现原本可能找不到的产品。

为实现研究目标，我们选取了两个受到偏差影响的真实世界数据集作为研究对象。所选数据集呈幂律分布，也称为长尾现象，这表明物品存在流行度偏差。此外，这些数据集还包含敏感特征——性别，因此适用于本研究。

我们实现了来自三种方法（协同过滤、矩阵分解和图神经网络）的多种不同模型，以比较基于图神经网络的方法与其他类型算法的表现。为评估结果，我们采用不同指标来衡量模型在前K项推荐中的可靠性和偏差放大情况。用于评估各模型在本研究所分析的偏差类型上的指标包括基尼指数、物品覆盖率和平均流行度。

结果表明，基于图神经网络的方法在精确率、召回率、归一化折损累计增益及其他可靠性指标方面通常表现更佳，但根据计算所得的偏差指标，这些方法也更容易放大偏差。在所有方法中，LightGCN 的表现受数据集和偏差类型影响最大，而 SGL 最稳定，对偏差最不敏感，尤其是流行度偏差和覆盖率偏差。这凸显了进一步研究的必要性，以便在准确率与偏差放大之间实现更好的权衡。此外，性别偏差的结果显示出不同的偏差放大模式。在大多数情况下，图神经网络方法在两个数据集上的差分公平性方面表现更佳。这表明，图神经网络方法对不同类型偏差的放大效应可能有所不同。

未来的工作中，我们将致力于研究基于图神经网络的推荐算法中偏差放大的成因，并提出对推荐可靠性影响最小的缓解方案。此外，我们计划通过采用更多数据集和更多指标来扩展本研究，从而分析性别以外的其他敏感特征。此外，还可以利用非二元性别属性研究性别偏差。

作者贡献： 概念构思，N.C.、N.S. 和 M.N.M.-G.；方法，N.C. 和 M.N.M.-G.；软件，N.C. 和 N.S.；验证，N.C. 和 N.S.；正式分析，N.C.；调查研究，N.C. 和 N.S.；资源，M.N.M.-G.；数据整理，N.C.；论文初稿撰写，N.C.；论文审阅与编辑，N.C.、N.S. 和 M.N.M-G；可视化，N.C.；监督指导，M.N.M.-G.；项目管理，M.N.M.-G.；经费获取，M.N.M.-G.。所有作者均已阅读并同意发表的稿件版本。

资助： 本研究未获得外部资助 ng.

知情同意声明： 不适用。

数据可用性声明： 本研究使用了公开数据集。详情见第 3.1节。

致谢： 本节可致谢作者贡献或资助部分未涵盖的任何支持。这可能包括行政和技术支持，或实物捐赠（例如，实验中使用的材料）。

Conf利益冲突： 作者声明不存在利益冲 突。

参考文献

1. Lin, S.; Wang, J.; Zhu, Z.; Caverlee, J. 量化并缓解对话式推荐系统中的流行度偏差。 *arXiv* **2022年**, arXiv:2208.03298.2. C hen, J.; Dong, H.; Wang, X.; Feng, F.; Wang, M.; He, X. 推荐系统中的偏差与去偏：综述及未来方向。 *arXiv* **2020**, arXiv:2010.03240.3. Zhang, S.; Yao, L.; Sun, A.; Tay, Y. 基于深度学习的推荐系统：综述与新视角。 *ACM*计算调查（*CSUR*） **2019**, *52*, 1–38. [CrossRef]4. Ricci, F.; Rokach, L.; Shapira, B. 《推荐系统手册》；施普林格：美国纽约州纽约，2015。

5. Alam, M.; Iana, A.; Grote, A.; Ludwig, K.; Müller, P.; Paulheim, H. 利用情感与立场检测分析新闻推荐系统偏差。arXiv 2022年，arXiv:2203.05824. 6. Gao, C.; Wang, X.; He, X.; Li, Y. 用于推荐系统的图神经网络。载于第十五届 ACM 国际网络搜索与数据挖掘会议论文集，美国亚利桑那州坦佩，2022年2月21–25日；第1623–1625页。7. Di Noia, T.; Tintarev, N.; Fatourou, P.; Schedl, M. 欧洲人工智能法规下的推荐系统。Commun. ACM

2022, – 。

65, 69 73 [ ]
CrossRef 8. Fahse, T.; Huber, V.; Giffen, B.v. 管理机器学习项目中的偏差。载于国际经济信息学会议论文集，德国纽伦堡，2022年2月21–23日；施普林格：德国柏林/海德堡，2021年；第94–109页。9. Kordzadeh, N.; Ghasemaghaei, M. 算法偏差：综述、综合分析与未来研究方向。Eur. J. Inf. Syst.

2022, – 。

31, 388 409 [ ]
CrossRef 10. Boratto, L.; Fenu, G.; Marras, M. 在协同推荐的流行度去偏中连接用户和物品视角。Inf. Process. Manag.

2021, *58*, 102387 [ ] . CrossRef 。Mu, R. 基于深度学习的推荐系统综述。《IEEE开放获取》
11

2018, *6*, 69009 69022 [ ] – 。CrossRef 。
12
Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. 图神经网络：方法与应用综述。《人工智能开放》

2020, *1*, 57 81 [ ] – 。CrossRef 。Bronstein, M.M.; Bruna, J.; LeCun, Y.; Szlam, A.;
13
Vandergheynst, P. 几何深度学习：超越欧几里得数据。IEEE Signal Process. Mag.

2017, *34*, 18 42 [ ] – 。CrossRef 。Wu, S.; Sun, F.; Zhang, W.; Xie, X.; Cui, B. 推荐系统中的图神经网
14
络：综述。ACM计算调查（CSUR）2020年。[CrossRef]15. Dai, E.; Wang, S. 向歧视说不：利用有限的敏感属性信息学习公平图神经网络。载于第14届 ACM 国际网络搜索与数据挖掘会议论文集，线上，2021年3月8–12日；第680–688页。16. Baeza-Yates, R. 网络中的数据偏差与算法偏差。载于第8届 ACM 网络科学会议论文集，德国汉诺威，2016年5月22–25日；第1页。17. Boratto, L.; Marras, M. 网络偏差感知推荐进展。载于第14届 ACM 国际网络搜索与数据挖掘会议论文集，线上会议，2021年3月8–12日；第1147–1149页。
18 **2020**, *35*, 957 967 [ ] .
19 **2013**, *15*, 209 227 [

Milano, S.; Taddeo, M.; Floridi, L. 推荐系统及其伦理挑战。AI Soc. – 。CrossRef 。Bozdag, E. 算法过滤与个性化中的偏差。Ethics Inf. Technol. – 。CrossRef

] . Ciampaglia, G.L.;
20
Nematzadeh, A.; Menczer, F.; Flammini, A. 算法流行度偏差如何阻碍或促进质量。Sci. Rep. **2018**, *8*, 15951 [ ] .
21
CrossRef 。Eryarsoy, E.; Piramuthu, S. 在线客户评论中序列偏差的实验评估。Inf. Manag. 2014, 51, 964–971. [CrossRef]22. Vall, A.; Quadrana, M.; Schedl, M.; Widmer, G. 下一首歌曲推荐中的顺序、上下文与流行度偏差。Int. J. Multimed. Inf. Retr.

2019, *8*, 101 113 [ ] – 。CrossRef 。Olteanu, A.; Castillo, C.; Diaz, F.; Kıcıman, E. 社交数据：偏差、方法论陷
23
阱与伦理边界。Front. Big Data 2019, 2, 13. [CrossRef] [PubMed]24. Bruce, P.; Bruce, A.; Gedeck, P. 《数据科学家的实用统计学》，第2版；O’Reilly Media, Inc.：美国加利福尼亚州塞瓦斯托波尔，2020年；第50–51页。25. Gu, J.; Oelke, D. 理解机器学习中的偏差。arXiv 2019, arXiv:1909.01866. 26. Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A. 机器学习中的偏差与公平性综述。ACM计算调查（CSUR）2021年，54, 1–35. [CrossRef]27. Akter, S.; Dwivedi, Y.K.; Sajib, S.; Biswas, K.; Bandara, R.J.; Michael, K. 基于机器学习的营销模型中的算法偏差。J. Bus. Res. 2022, 144, 201–216. [CrossRef]28. Blanzeisky, W.; Cunningham, P. 影响机器学习偏差的算法因素。载于欧洲机器学习与数据库知识发现联合会议论文集，西班牙毕尔巴鄂，2021年9月13–17日；施普林格：德国柏林/海德堡，2021年；第559–574页。29. Hall, M.; van der Maaten, L.; Gustafson, L.; Adcock, A. 偏差放大的系统性研究。arXiv 2022, arXiv:2201.11706. 30. Ashokan, A.; Haas, C. 评分预测的公平性指标与偏差缓解策略。Inf. Process. Manag. 2021, 58, 102646. [CrossRef]31. Zhang, Q.; Wipf, D.; Gan, Q.; Song, L. 具有近似最优遗憾值的有偏图神经网络采样器。Adv. Neural Inf. Process. Syst. 2021, 34, 8833–8844. 32. Dong, Y.; Liu, N.; Jalaian, B.; Li, J. Edits：图神经网络数据偏差的建模与缓解。载于2022年ACM 网络会议论文集，线上会议，法国里昂，2022年4月25–29日；第1259–1269页。33. Liu, Y.; Ao, X.; Feng, F.; He, Q. UD-GNN：半同配图上的不确定性感知去偏训练。载于第28届 ACM SIGKDD知识发现与数据挖掘会议论文集，美国华盛顿哥伦比亚特区，2022年8月14–18日；第1131–1140页。34. Dong, Y.; Wang, S.; Wang, Y.; Derr, T.; Li, J. 图神经网络偏差的结构性解释。载于第28届 ACM SIGKDD知识发现与数据挖掘会议论文集，美国华盛顿哥伦比亚特区，2022年8月14–18日；第316–326页。

35. Xu, B.; Shen, H.; Sun, B.; An, R.; Cao, Q.; Cheng, X. 面向消费贷款欺诈检测：采用角色约束条件随机场的图神经网络。载于AAAI人工智能大会论文集，线上会议，2021年10月11日至15日；第35卷，第4537–4545页。36. Zeng, Z.; Islam, R.; Keya, K.N.; Foulds, J.; Song, Y.; Pan, S. 异构信息网络的公平表征学习。载于AAAI国际网络日志与社交媒体大会论文集，线上会议，2021年6月7日至10日；第15卷。37. Chen, Z.; Xiao, T.; Kuang, K. BA-GNN：偏差感知图神经网络的学习。载于2022年IEEE第38届国际数据工程大会（ICDE）论文集，马来西亚吉隆坡，2022年5月9日至12日；第3012–3024页。38. Gao, C.; Lei, W.; Chen, J.; Wang, S.; He, X.; Li, S.; Li, B.; Zhang, Y.; Jiang, P. CIRS：通过反事实交互式推荐系统打破过滤气泡。arXiv 2022，arXiv:2204.01266。39. Fabbri, F.; Croci, M.L.; Bonchi, F.; Castillo, C. 人物推荐系统中的曝光不平等：长期影响。载于AAAI国际网络与社交媒体大会论文集，美国佐治亚州亚特兰大，2022年6月6日至9日；第16卷，第194–204页。40. Mansoury, M.; Abdollahpouri, H.; Pechenizkiy, M.; Mobasher, B.; Burke, R. 一种缓解推荐系统中多侧曝光偏差的基于图的方法。《ACM信息系统汇刊》 **2021**, *40*, 1 31 [ ] – 。CrossRef 。Sun, W.; 41
Khenissi, S.; Nasraoui, O.; Shafto, P. 在协同过滤中消除人类—推荐系统反馈循环的偏差。载于2019年万维网大会论文集，美国加利福尼亚州旧金山，2019年5月13日至17日；第645–651页。42. Ahanger, A.B.; Aalam, S.W.; Bhat, M.R.; Assad, A. 推荐系统中的流行度偏差——综述。载于国际计算机工程新兴技术大会论文集，印度斋浦尔，2022年2月4日至5日；施普林格：德国柏林/海德堡，2022年；第431–444页。43. Wu, P.; Li, H.; Deng, Y.; Hu, W.; Dai, Q.; Dong, Z.; Sun, J.; Zhang, R.; Zhou, X.H. 推荐系统中因果学习的机遇：基础、估计、预测与挑战。载于国际人工智能联合会议论文集，奥地利维也纳，2022年7月23日至29日。44. Kowald, D.; Lacic, E. 基于协同过滤的多媒体推荐系统中的流行度偏差。arXiv 2022，arXiv:2203.00376。45. Neophytou, N.; Mitra, B.; Stinson, C. 重新审视推荐系统评估中的流行度偏差和人口统计偏差及其有效性。载于欧洲信息检索大会论文集，挪威斯塔万格，2022年4月10日至14日；施普林格：德国柏林/海德堡，2022年；第641–654页。46. Rahmani, H.A.; Naghiaei, M.; Tourani, A.; Deldjoo, Y. 探究时间偏差对兴趣点推荐的影响。arXiv 2022，arXiv:2207.11609。47. Ekstrand, M.D.; Das, A.; Burke, R.; Diaz, F. 推荐系统中的公平性。载于《推荐系统手册》；施普林格：德国柏林/海德堡，2022年；第679–707页。48. Abdollahpouri, H.; Mansoury, M.; Burke, R.; Mobasher, B. 流行度偏差、校准与推荐公平性之间的联系。载于第十四届ACM推荐系统大会论文集，线上会议，巴西，2020年9月22日至26日；第726–731页。49. Abdollahpouri, H.; Burke, R.; Mobasher, B. 通过个性化重排序管理推荐系统中的流行度偏差。载于第三十二届国际FLAIRS会议论文集，美国佛罗里达州萨拉索塔，2019年5月19日至22日。50. Liu, H.; Wang, Y.; Lin, H.; Xu, B.; Zhao, N. 通过对抗学习缓解公平推荐系统中的敏感数据暴露。《神经计算与应用》 **2022**, 1 15 [ ] – 。 51
CrossRef 。Shakespeare, D.; Porcaro, L.; Gómez, E.; Castillo, C. 探究音乐推荐中的艺术家性别偏差。arXiv 2020，arXiv:2009.01715。52. Saxena, S.; Jain, S. 探究并缓解采用显式反馈的推荐系统中的性别偏差。arXiv 2021，arXiv:2112.02530。53. Rahman, T.; Surma, B.; Backes, M.; Zhang, Y. Fairwalk：迈向公平的图嵌入。载于国际人工智能联合会议论文集，中国澳门，2019年8月10日至16日。54. Chen, J.; Wu, W.; Shi, L.; Zheng, W.; He, L. 基于校准的长尾会话推荐。《应用智能》2022，1–18。[CrossRef]55. Zhao, M.; Wu, L.; Liang, Y.; Chen, L.; Zhang, J.; Deng, Q.; Wang, K.; Shen, X.; Lv, T.; Wu, R. 探究基于图的协同过滤的准确性—新颖性表现。arXiv 2022，arXiv:2204.12326。56. Kim, M.; Oh, J.; Do, J.; Lee, S. 消除推荐系统中图神经网络邻居聚合的偏差。arXiv 2022，arXiv:2208.08847。57. Movielens。2021年。在线地址：https://grouplens.org/datasets/movielens/（访问日期：2022年6月31日）。58. Floridi, L.; Holweg, M.; Taddeo, M.; Amaya Silva, J.; Mökander, J.; Wen, Y. capAI——一项用于开展合格评定的流程，使人工智能系统符合欧盟《人工智能法案》。2022年。在线地址：https://papers.ssrn.com/sol3 /papers.cfm?abstract_id=4064091（访问日期：2022年6月31日）。59. 百万歌曲数据集。在线地址：http://millionsongdataset.com/lastfm/（访问日期：2022年6月31日）。60. Al-Ghamdi, M.; Elazhary, H.; Mojahed, A. 推荐系统的协同过滤评估。《国际先进计算机科学与应用期刊》 **2021**, *12*, 559 565 [ ] – 。CrossRef 。Airen, S.; Agrawal, J. 使用K近邻变体的电影推荐系统。《国家科学院快报》 61
2022，45，75–82。[CrossRef]

62. Deshpande, M.; Karypis, G. 基于物品的Top-N推荐算法。《ACM信息系统汇刊》2004，22，143–177。[CrossRef]63. Sang, L.; Xu, M.; Qian, S.; Wu, X. 利用残差循环网络增强知识图谱的神经协同过滤。《神经计算》 **2021**, *454*, 417 429 [ ] – 。CrossRef 。Girsang, A.S.; Wibowo, A.; Jason; Roslynlia. 用于音乐推荐系统的64
神经协同方法。《IOP会议系列：材料科学与工程》 **2021**, *1071*, 012021 [ ] 。CrossRef 。Bai, T.; Wen, J.R.; Zhang,
65
J.; Zhao, W.X. 一种基于交互的邻域神经协同过滤模型。载于2017年ACM信息与知识管理会议论文集，新加坡，2017年11月6日至10日；第1979–1982页。66. Himabindu, T.V.; Padmanabhan, V.; Pujari, A.K. 基于保形矩阵分解的推荐系统。《信息科学》 **2018**, – 。
467, 685 707 [ ]
CrossRef 67. Kuang, H.; Xia, W.; Ma, X.; Liu, X. 用于跨域推荐的深度矩阵分解。载于2021年IEEE第五届先进信息技术、电子与自动化控制会议（IAEAC）论文集，中国重庆，2021年3月12日至14日；第5卷，第2171–2175页。68. Xue, H.J.; Dai, X.; Zhang, J.; Huang, S.; Chen, J. 面向推荐系统的深度矩阵分解模型。载于IJCAI论文集，澳大利亚墨尔本，2017年8月19日至25日；第17卷，第3203–3209页。69. Yi, B.; Shen, X.; Liu, H.; Zhang, Z.; Zhang, W.; Liu, S.; Xiong, N. 采用隐式反馈嵌入的深度矩阵分解推荐系统。《IEEE工业信息学汇刊》 **2019**, *15*, 4591 4601 [ ] – 。CrossRef 。Liang, G.; Sun, C.;
70
Zhou, J.; Luo, F.; Wen, J.; Li, X. 面向多接入边缘计算网络中推荐系统的通用矩阵分解框架。《移动网络与应用》 **2022**, 1 13 [ ] – 。CrossRef 。He, X.; Liao, L.; Zhang, H.;
71
Nie, L.; Hu, X.; Chua, T.S. 神经协同过滤。载于第26届万维网国际会议论文集，澳大利亚珀斯，2017年4月3日至7日；第173–182页。72. Wang, S.; Hu, L.; Wang, Y.; He, X.; Sheng, Q.Z.; Orgun, M.A.; Cao, L.; Ricci, F.; Yu, P.S. 基于图学习的推荐系统：综述。arXiv 2021，arXiv:2105.06339。73. Schafer, J.B.; Frankowski, D.; Herlocker, J.; Sen, S. 协同过滤推荐系统。载于《自适应网络》；施普林格：德国柏林/海德堡，2007年；第291–324页。74. He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; Wang, M. LightGCN：简化并增强用于推荐的图卷积网络。载于第43届国际ACM SIGIR信息检索研究与发展会议论文集，中国西安，2020年7月25日至30日；第639–648页。75. Broman, N. 股票灵感推荐系统比较。硕士论文，林雪平大学，瑞典林雪平，2021年。76. Ding, S.; Feng, F.; He, X.; Liao, Y.; Shi, J.; Zhang, Y. 用于推荐系统再训练的因果增量图卷积。《IEEE神经网络与学习系统汇刊》 **2022** [ ] [ ] 。CrossRef PubMed 。Sun, W.; Chang, K.; Zhang, L.;
77
Meng, K. INGCF：一种基于NGCF的改进推荐算法。载于并行处理算法与体系结构国际会议论文集，中国厦门，2021年12月3日至5日；施普林格：德国柏林/海德堡，2021年；第116–129页。78. Wu, J.; Wang, X.; Feng, F.; He, X.; Chen, L.; Lian, J.; Xie, X. 用于推荐的自监督图学习。载于第44届国际ACM SIGIR信息检索研究与发展会议论文集，线上会议，2021年7月11日至15日；第726–735页。79. Yang, C. 面向推荐的监督式对比学习。arXiv 2022，arXiv:2201.03144。80. Tang, H.; Zhao, G.; Wu, Y.; Qian, X. 用于Top-k推荐的基于多样本的对比损失。《IEEE多媒体汇刊》2021。[CrossRef]81. Foulds, J.R.; Islam, R.; Keya, K.N.; Pan, S. 差分公平性。载于NeurIPS 2019具有保证的机器学习研讨会，加拿大温哥华；马里兰大学巴尔的摩县分校教师文集：美国马里兰州巴尔的摩，2019年。82. Foulds, J.R.; Islam, R.; Keya, K.N.; Pan, S. 公平性的交叉性定义。载于2020年IEEE第36届国际数据工程大会（ICDE）论文集，美国得克萨斯州达拉斯，2020年4月20日至24日；第1918–1921页。83. Naghiaei, M.; Rahmani, H.A.; Dehghan, M. 图书推荐中流行度偏差的不公平性。arXiv 2022，arXiv:2202.13446。84. Lazovich, T.; Belli, L.; Gonzales, A.; Bower, A.; Tantipongpipat, U.; Lum, K.; Huszar, F.; Chowdhury, R. 使用分布不平等度量指标衡量内容推荐算法产生的差异化结果。arXiv 2022，arXiv:2202.01615。85. Wang, X.; Wang, W.H. 为深度推荐系统提供物品侧个体公平性。载于2022年ACM公平性、问责制与透明度会议论文集，韩国首尔，2022年6月21日至24日；第117–127页。86. Islam, R.; Keya, K.N.; Zeng, Z.; Pan, S.; Foulds, J. 通过神经公平协同过滤实现职业推荐去偏。载于网络会议论文集，斯洛文尼亚卢布尔雅那，2021年4月19日至23日；第3779–3790页。