arXiv:2301.08177v1 [cs.SI] 19 Jan 2023

社交媒体影响者的个体公平性

Stefania Ionescu、Nicol`o Pagan 和 Anik´o Hann´ak

苏黎世大学社交计算组，瑞士苏黎世 8050，ionescu@ifi.uzh.ch，
pagan@ifi.uzh.ch，hannak@ifi.uzh.ch，

摘要。 如今，许多社交媒体平台都以内容创作者（CC）为中心。在这些平台上，关系形成过程取决于两个因素：（a）用户接触到内容创作者的曝光度（由推荐系统等决定），以及（b）用户的关注决策过程。近期研究强调了内容质量的重要性，指出在采用探索式推荐策略时，网络最终会趋于一种状态：内容创作者的质量越高，其关注者数量的期望值也越高。本文在先前工作的基础上，（a）不再局限于考察平均值，而是评估这一过程的公平性；（b）探究探索式推荐对于实现公平结果的重要性。通过分析方法，我们证明，非探索式推荐收敛速度很快，但通常会导致不公平的结果。此外，即使进行了探索，我们也只能保证质量最高（和最低）的内容创作者获得公平结果。

关键词：社会　　网络形成、个体公平性、马尔可　　夫链

1 引言

过去几十年间，社交媒体平台对我们生活的影响急剧增加，例如，它们塑造了我们接收的信息 [1] 以及我们形成的观点 [2]。在此期间，原本仅用于连接现实生活中朋友的平台，逐渐开始鼓励用户根据内容关注陌生人。如今，YouTube、Twitter、Instagram 和 TikTok 等平台都以用户生成内容（UGC）为核心，并利用推荐系统（RSs）促进内容探索。为适应这一变化，一些用户专注于创作甚至达到半专业水准的内容，以吸引越来越多的关注者，最终凭借受众实现收入。与劳动力市场类似 [3], ，我们自然会期待这些在线平台以公平的方式对待内容创作者（CCs），

本文为以下章节的预印本：Stefania Ionescu、Nicol`o Pagan 和 Anik´o Hann´ak，《社交媒体影响者的个体公平性》，发表于《复杂网络及其应用 XI》第一卷，由 Hocine Cher- ifi、Rosario Nunzio Mantegna、Luis M. Rocha、Chantal Cherifi 和 Salvatore Micci- che 编者，2023 年。经施普林格·自然瑞士股份公司许可，由 Springer 于 2023 年重印。最终认证版本可在线获取：https://doi.org/10.1007/978-3-031-21127-0 14。

即同等优秀的内容创作者应在可见性、受众规模乃至收入方面获得相似的回报。由于网络形成过程往往在很大程度上受到推荐系统的影响，因此有必要探讨这些系统是否为内容创作者带来了公平结果。

为了初步了解这些社交媒体平台的结构和特性， [4] 提出了一个简单的模型，其中：(a) 每位内容创作者都有内在且客观的质量；(b) 每一轮中，用户都会收到一条内容创作者推荐（该推荐可能从均匀分布中抽取，也可能通过优先连接（PA）过程产生）；(c) 如果被推荐的内容创作者质量高于用户当前所有被关注者，用户就会关注该内容创作者。通过模拟这一模型，Pagan 等人 [4] 表明，内容创作者的预期关注者数量（入度）符合齐夫定律 [5]。具体而言，按质量和按关注者数量得出的内容创作者预期排名相同。

虽然这些结果表明，这种网络形成过程会为内容创作者带来公平的结果，但我们发现了两个重要的局限。首先，分析仅限于两种探索式推荐过程（即，为了发现更好的选项而采取高风险行动的过程）。然而在实践中，并非所有推荐系统都是如此。推荐系统领域的近期文献阐述了我们为何以及如何通过找到探索与利用之间的适当平衡来促进多样性 [6–9]。在实践中，即使是热门项目，也可能无法同时被用户访问；也就是说，如果用户收到两个项目中的一个并关注它，就不会再收到另一个项目的推荐 [10]。这使现实世界中的推荐系统与优先连接和均匀随机（UR）推荐形成鲜明对比，因为在后两者中，每位用户都可能收到任何内容创作者的推荐。其次，作者只关注预期关注者数量在收敛时（即所有用户最终都收到最佳内容创作者的推荐之后）。然而，这种事前公平性并不意味着事后公平性（也就是说，即使内容创作者平均而言获得的关注者数量与其质量成正比，许多可能出现的实际结果仍然不公平） [11]。此外，正如作者也指出的那样 [4], ，收敛可能需要很长时间。这意味着，即使最终能够实现公平的结果，也可能无法在合理的时间内实现。

本文旨在通过连接网络科学、推荐系统和算法公平性领域，解决这两项挑战。更具体地说，本文（a）为内容创作者定义了事前和事后公平性指标；（b）使用极端优先连接扩展模型（这是一种受流行度推荐系统启发的非探索型推荐策略， [12], 该系统仅推荐关注者最多的内容创作者）；（c）使用马尔可夫链从理论上研究极端优先连接下的网络形成过程及其公平性；（d）比较极端优先连接、优先依附和均匀随机推荐下的网络形成过程及其公平性（或参考已有工作，或通过新的分析和数值结果进行比较）。

2 相关工作

网络。 多年来，复杂网络领域发展出各种简单却现实的机制，用于解释社交网络的形成

（例如，小世界网络模型 [13], 和优先连接模型 -PA- [14]）。在优先连接中，新生节点以与度成比例的概率连接到现有节点。这种富者愈富现象成功再现了这样的观点：受欢迎的用户拥有更高的可见性，进而获得更高的人气。另一方面，这种现象较少强调解释个体为何建立某些连接的社会经济微观基础。聚焦于这一替代方法，社会学中的一条研究路线（随机行动者导向模型 [15]）和经济学中的一条研究路线（战略性网络形成模型 [16]）采取功利主义视角：智能体建立连接以最大化某种收益，例如其网络中心性。Pagan 等人 [4]提出的基于质量的模型结合了这两种方法，使用均匀随机或基于优先连接的推荐系统，以及面向用户的功利主义关注决策函数。为加深我们对推荐系统与人类网络行为之间耦合关系的理解，我们加入一种非探索型推荐策略，并研究由此产生结果的公平性。

公平性。 研究者关注的不仅是过程的平均表现，也关注这些过程对个体造成影响时是否公平。这体现在大量关于制定公平性度量，以及根据应用领域选择最合适度量的方法论研究中 [17–19]。在这些公平性指标中，我们重点关注个体公平性，它评估资质相似的个体是否能获得质量相似的结果（参见 [20] ，了解其重要性以及它与其他公平性指标之间显而易见的不相容性）。一个重要现象是时机效应：根据这一效应，仅指定福利函数还不够，还必须明确应在何时对其进行衡量（事前或事后） [11]。基于这些研究，我们定义并考察内容创作者的事前和事后个体公平性。

推荐系统。 近年来，推荐系统领域主张在推荐系统评估中关注准确率以外的方面 [6] 。在设计层面，人们持续努力开发多样化推荐系统 [7–9] ，并确保任意两个项目都可以同时推荐给用户 [10]。这或许解释了为什么推荐系统领域采用的基于流行度的算法与优先依附不同 [12, 21]：它们不允许对备选项进行完整探索 [4, 14] 。我们分析极端优先连接的动机源于这些先前研究，不过我们的目标是理解网络形成过程及其公平性。

3 模型

如前两节所述，我们的工作基于 [4]提出的质量模型，下面将对其作简要介绍。与任何模型一样，该模型建立在一些简化假设之上；有关这些假设，请参阅原论文。请注意，为便于未来扩展，我们有时采用的符号与 [4] 有所不同。接下来的两个小节将介绍我们对该模型的改进：(a) 对网络形成过程进行基于推荐系统的表述（并定义新的推荐系统——极端优先连接），以及 (b) 公平性定义。

我们假设用户可分为 $n \geq 2$ 内容创作者（内容创作者）和m（非内容创作者）用户，其中 $m \gg n$。我们将内容创作者称为 CC_1，CC_2，...，CC_n，并假设他们按照现有的客观质量严格排序。也就是说，CC_1是质量最高的内容创作者，CC_2 是次高的内容创作者，依此类推。每个用户都可以关注任何内容创作者，因此形成一个二部、无权且有向的网络。我们用其邻接矩阵 $A \in \{0,1\}^{m \times n}$表示该网络，使得 (s.t.) $a_{u,\,i}$ 当用户 u 关注 CC_i时取值为 1，否则取值为 0。

网络形成是一个顺序动态过程，其中（a）网络最初为空；（b）在每个时间步，系统都会向每位用户推荐一位内容创作者，用户可以选择是否关注。形式上，令 A^t 表示时间t时网络的状态。根据假设（a），对于每位用户 $u \in m$ 和内容创作者 $i \in n$，$a^0_{u,\,i} = 0$ 。 [1] 下一小节将介绍如何生成推荐，以及这些推荐会产生什么影响。

3.1 推荐函数

我们将推荐视为把用户映射到项目的函数，即，时间 t 的推荐由 $R^t : m \to n$给出。 [2] 推荐系统是基于网络当前状态生成此类推荐函数的算法。本文考察以下三种算法，并按探索性从强到弱排列：

– 均匀随机（*UR*）。每位用户被推荐任意一位内容创作者的概率相同，即 $R_{\mathrm{UR}^t(u)} \sim \mathcal{U}(n)$。– 优先连接（*PA*）。每位内容创作者被推荐的概率与其当前关注者数量成正比。也就是说，如果$a_{\cdot,\,i} := \sum_{u \in m} a_{u,\,i}$ 是 CC_i的关注者数量，那么：

$$\mathbb{P}(R^t_{\mathrm{PA}}(u) = i) = \frac{a^t_{\cdot,i} + 1}{\sum_{j \in n}(a^t_{\cdot,j} + 1)}.$$

– *ExtremePA*。优先连接的一种极端变体，只推荐最热门的项目（之一）：

$$\mathbb{P}(R^t_{\mathrm{ext}}(u) = i) = \begin{cases} 0 & \text{, if } a^t_{\cdot,i} < \max_j a^t_{\cdot,j}, \\ 1/|\{i, \text{ s.t. } i \in \arg\max_j a^t_{\cdot,j}\}| & \text{, otherwise.} \end{cases}$$

每位用户 u 收到与其对应的推荐 $R^t(u)$ 后，当且仅当 $R^t(u)$ 比 u 的任何被关注者都更好（也可能是空真）时，才会关注 $R^t(u)$：

$$a^{t+1}_{u,i} = \begin{cases} 1 & \text{, if } R^t(u) = i \text{ and } i < j \ \forall j \ \text{ s.t. } a^t_{u,j} = 1 \\ a^t_{u,i} & \text{, otherwise.} \end{cases}$$

[1] 本文中，我们用 k表示不大于 k 的非零自然数集合，即 $k := \{1, 2, \ldots k\}$。 [2] 请注意，在完全一般的情况下，每位用户都会收到一组 *list* 推荐。不过，若允许每位用户只收到一条推荐，则我们的模型便可与 [4]进行比较。

3.2 感兴趣的指标

由于时间 $t+1$ 的网络只取决于时间 t 的网络，我们可以将这一网络形成过程视为马尔可夫链（MC）。更多细节将在第 4 节介绍。基于这一理解，我们首先想探究：(a) 这是否是一个吸收马尔可夫链（即是否存在某些网络——吸收态——最终会到达，并且无论经过多少时间步都不会改变）；以及如果是，(b) 吸收前的预期步数（即到达这种状态所需的步数）。

然而，正如引言中所述，我们也关注整个过程的公平性，包括事前和事后公平性。因此，我们定义：

– 内容创作者的事后个体公平性。我们称网络 A 是（个体）公平的，如果内容创作者的关注者数量排名与内容创作者的排名相同，即 $a_{\cdot,1} \geq a_{\cdot,2} \geq \cdots \geq a_{\cdot,n}$。我们也称 A 是（个体） CC_i-公平的，如果CC_i 是排名前 i 的内容创作者之一，即 $|\{j : a_{\cdot,j} > a_{\cdot,i}\}| < i$。特别是，这意味着当且仅当网络对所有 $i \in n$都是 CC_i-公平的时，该网络才是公平的。 [3]– 内容创作者的事前个体公平性。当这是一个吸收马尔可夫链时，我们也可以考察网络形成过程的事前公平性。更准确地说，如果吸收时内容创作者的关注者数量期望值递减，即如果$\mathbb{E}[a^{\infty}_{\cdot,1}] \geq \mathbb{E}[a^{\infty}_{\cdot,2}] \geq \cdots \geq \mathbb{E}[a^{\infty}_{\cdot,n}]$，我们就称该过程是事前（个体）公平的。类似地，如果 $|\{j : \mathbb{E}[a^{\infty}_{\cdot,i}] < \mathbb{E}[a^{\infty}_{\cdot,j}]\}| < i$，我们就称该过程是事前（个体） CC_i-公平的。

4 结果

在结果部分，我们将研究第 3.2 节中介绍的相关指标。因此，4.1 小节展示了如何将这一过程视为马尔可夫链，以及哪些状态是吸收状态。4.2 小节探讨达到吸收状态的预期时间。最后，在 4.3 小节中，我们基于前两节的内容研究公平性。每个小节都以两段开篇：（a）总结段，概述结果；（b）要点段，讨论这些结果的意义。每个小节其余部分则用于呈现结果本身。为节省篇幅，在后面几个证明中，我们省略了细节，仅提供证明的概要和直观解释。

4.1 一个吸收马尔可夫链

总结。 本小节首先证明我们的过程是一个吸收马尔可夫链（见定理 1）。在证明过程中，我们还定义了转移矩阵。图 1 展示了一个小规模案例中的这一过程，其中有两个内容创作者和两个（非内容创作者）

[3] 请注意，这是一个较弱版本的公平性定义。或者，我们可以说，如果一个网络 A 满足 $a_{\cdot,1} > a_{\cdot,2} > \cdots > a_{\cdot,n}$，那么它就是公平的。然而，正如我们稍后将在结果部分看到的那样，随着用户数量趋于无穷，达到相等的概率会趋于零。因此，在极限情况下，这两个定义是等价的。

用户。其余结果刻画了吸收态。首先，定理 2 表明，在探索式推荐过程（即优先依附或均匀随机）下，吸收态是每个用户都关注最佳内容创作者的状态。其次，定理 3 表明，在 ExtremePA 下，从**0** 状态可达的状态若满足以下条件，则为吸收状态：(a) 关注者数量最多的内容创作者只有一个；(b) 没有新用户想要关注该内容创作者。对于后一条件，我们使用引理 1。该引理指出，对于从 **0** 可达的任何状态，每个拥有最多关注者的 CC_i 都对应一个不关注任何比 CC_i 更好的内容创作者的用户。

要点。 本小节的结果构建了该过程的一种表示，有助于理解这一过程。重要的是，通过描述转移矩阵，我们可以看到三种推荐系统之间的差异：(a) ExtremePA 的吸收状态更多；(b) 在这些状态中， CC_1 是关注度最高的CC（即 CC_1-公平状态）的状态，是探索型推荐系统下的吸收状态；(c) 在探索程度方面，PA 介于 ExtremePA 和 UR 之间。所谓后者，是指与 UR 相比，PA 可能会在暂态中停留更长时间（这些暂态在公平性方面与仅在 ExtremePA 下为吸收状态的状态相似）。

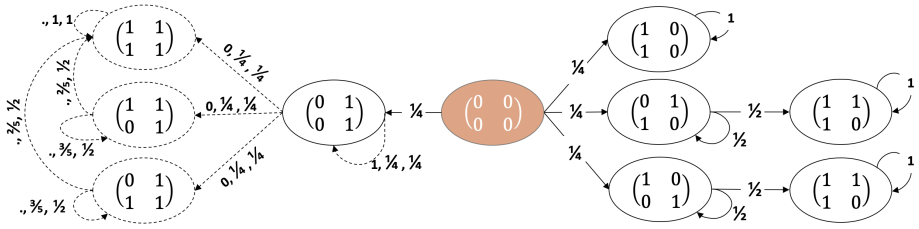


图 1。 $n = m = 2$的马尔可夫链表示示例。彩色节点为起始状态。我们用 (a) 实线表示三种推荐系统下相同的转移，用 (b) 虚线表示不同的转移（并分别标注 (1) ExtremePA、(2) PA、(3) UR 对应的概率）。当起始状态在 ExtremePA 下无法从**0** 到达时，则用圆点代替概率。

定理 1. $(A^t)_{t\geq 0}$ 是吸收马尔可夫链（对所有三种推荐系统均成立）。

证明。 首先，$(A^t)_{t\geq 0}$ 是一个马尔可夫链，因为 A^{t+1} 仅取决于前一步的网络配置。更具体地说，这是一个马尔可夫链，其中：(a) 状态空间为$\{0,1\}^{m\times n}$，(b) 初始分布为 λ，其中 $\lambda_A = 1$ 当且仅当 A 是零矩阵；(c) 转移矩阵由 $p_{B,\,C} := \mathbb{P}(A^{t+1} = C | A^t = B)$ 给出，其中：

– 从零矩阵出发，我们只能转移到这样一个矩阵：每个用户恰好关注一个内容创作者，即：$p_{\mathbf{0},C} = \begin{cases} 1/n^m, & \text{若 } c_{u\,\cdot} = 1 \ \forall\, u \in m_{\cdot}, \\ 0 & \text{否则；} \end{cases}$ – 从任何其他矩阵出发，我们只能转移到一个新矩阵，其中每个用户要么 (a) 关注与之前相同的内容创作者，要么 (b) 额外关注一个

内容创作者，且该创作者优于他们迄今关注过的最佳内容创作者；这类转移的概率取决于推荐过程。

此外，从转移矩阵的形式可以看出，这个过程是一个吸收马尔可夫链。为说明这一点，注意到 $p_{B,\ C}$ 非零，当且仅当 $C = B$，或者存在某个用户 u 多关注了一位内容创作者，即 $c_{u,\ .} > b_u$。因此，任意状态 B 要么是吸收状态，要么可以转移到元素总和严格更大的状态。由于 B 的元素总和有上界 $m \cdot n$，这样的转移序列必然是有限的。因此，在后一种情况下，经过一系列转移后最终会到达一个吸收状态。□

定理 2。 在优先依附和均匀随机推荐下，状态 B 是吸收状态，当且仅当所有用户都关注最佳内容创作者，即当且仅当 $b_{u,\ 1} = 1$ 对所有 $\overline{u} \in m$均成立。

证明。 (⇒) 假设存在某个用户 $u \in m$，使得 $b_{u,\ \overline{1}} = 0$。在优先依附和均匀随机推荐下，u 总有非零概率被推荐 CC_1。若发生这种情况，则 u 会关注 CC_1。因此，我们可以以非零概率转移到状态 $C \neq B$，其中$c_{u,\ 1} = 1$。所以，B 不是吸收状态。

(⇐) 反向推论显而易见。如果 $b_{u,1} = 1$，那么任何推荐都不会改变 u 的被关注者。这对所有用户 u都成立，因此 B 是吸收的。□

引理 1. 对于任何可从 ***0*** 到达的状态 B，以及任何满足 $b_{.,i} = \text{最大}_j\, b_{.,j}$ 的 $i \in n$，存在某个用户 $u \in m$ 满足 $\overline{b_{u,j}} = 0$，对所有 $j < i$ 均成立。

证明。 我们采用归纳论证。对于任何可从 **0** 直接到达的状态（即任何能在一个时间步内实现的网络），该断言显然成立，因为在这些状态中，每个用户恰好关注一个内容创作者。接下来，我们假设该断言对某个状态 B 成立。从该状态出发，要么保持在 B，要么转移到一个状态C，其中这些内容创作者的一个子集拥有更多关注者。由于该子集中的每个 CC_i 都增加了关注者人数，因此它们必定被推荐给某个用户 u，而该用户对它们的偏好高于其此前关注的任何内容创作者。因此，CC_i 是 C 中 u 所关注的最佳内容创作者，即 $i = \min\{j : c_{u,j} = 1\}$。□

定理 3. 在极端优先连接下，所有可从***0*** 到达的吸收态 B 都满足：***(a)*** 存在某个 $i \in n$，使得 $b_{.,i} > b_{.,j}$ 对所有 $j \in n - \{i\}$均成立；***(b)*** 对所有 $u \in m$，都存在某个 $j \leq i$，使得 $b_{u,j} = 1$。此外，每个满足 ***(a)*** 和 ***(b)*** 的状态都是吸收状态。

证明。 对于第一部分，设 B 是从 **0** 可到达的一个吸收状态。若条件 (a) 不成立，则至少有两个内容创作者（例如 CC_i 和 CC_j）拥有 $i < j$ 个关注者，且关注者数量最多。根据引理 1，存在某个尚未关注 CC_i 的用户 u。因此，如果向所有人推荐 CC_i，那么只有 CC_i 的关注者数量会至少增加一个（因为 u 会关注他们）。这种情况发生的概率非零，而一旦发生，就会到达一个新状态。因此，若条件 (a) 不成立，B 就不是吸收状态。接下来，若条件 (a) 成立但条件 (b) 不成立，那么在下一轮（将唯一一个关注者最多的内容创作者 CC_i 推荐给所有人时），新用户会关注 CC_i。因此，同样地，若条件 (a) 成立但条件 (b) 不成立，B 就不可能是吸收状态。

第二部分显而易见。若条件 (a) 成立，那么下一轮只能推荐关注者数量最多的那个唯一内容创作者。由于条件 (b) 成立，不会有新用户关注该内容创作者。因此，状态没有可能转移到新状态，也就是说，当前状态是吸收状态。□

4.2 吸收所需的预期时间

总结。定理 4 表明，随着用户数量趋于无穷，ExtremePA 下的预期吸收时间小于 2。证明这一结论需要两个前置结果：(a) 第一轮推荐后关注者数量出现并列的概率（引理 2），以及 (b) 第二轮中没有任何新用户关注关注者最多的用户 CC_i（如果 $i \neq n$，见引理 3）的概率，都会随着用户数量趋于无穷而趋于 0[4]。这些结果用于图 2 中的注释；图 2 概述了这一过程，并为定理的证明提供直观解释。

要点。这使 ExtremePA 与探索型推荐系统形成鲜明对比，因为先前研究表明，在 UR 和 PA 推荐情境下，收敛时间会随内容创作者数量呈对数级增长，随用户数量呈线性（或次线性）增长（见 [4] 中的图 2 和图 7b）。重要的是，这表明，对于 ExtremePA，仅分析吸收态中的公平性或许就足够了；但对于 PA 和 UR，情况可能并非如此：由于这些推荐系统在实践中可能导致较长的收敛时间，因此也应评估暂态中的公平性。对于 PA 而言尤其如此，因为如前所述，它可能比 UR 更长时间地停留在不公平状态中。

引理 2。对于任意 $i \neq j \in n$, $\mathbb{P}(a^1_{:,i} = a^1_{:,j}) \to 0$ 当 $m \to \infty$ 时。

证明。首先，令 Y_u 为以下取决于第一轮推荐的随机变量：

$$Y_u = \begin{cases} 0, \text{如果用户 } u \text{ 被推荐了某个 } CC_k \text{ 且满足 } k \neq i, j; \ 1, \text{如果用户 } u \text{ 被推荐了 } CC_i; \ -1, \text{如果用户 } u \text{ 被推荐了 } CC_j. \end{cases}$$

由于第一次推荐是均匀随机的，$(Y_u)_{u \in m}$ 是独立同分布的（其中 $\mathbb{P}(Y_u = 0) = (n-2)/n$，且当 $c = \pm 1$ 时 $\mathbb{P}(Y_u = c) = 1/n$）。因此，$\mathbb{E}[Y_u] = 0$ 且 $\mathrm{Var}(Y_u) = \mathbb{E}[Y_u^2] = 2/n$。根据中心极限定理可得，

$$\frac{\sum_{u \in m} Y_u - m \cdot 0}{2/n \cdot \sqrt{m}} \to^{\mathcal{D}} \mathcal{N}(0,1).$$

特别地，这意味着当 $m \to \infty$ $\mathbb{P}(a^1_{:,i} = a^1_{:,j}) = 0$ 时，极限为 □

引理 3. 对于任意 $i \in n$, $\mathbb{P}((a^1_{:,i} > a^1_{:,j} \ \forall j \in n) \curlywedge (\forall u \in m, \exists j \in i$ 使得 $a^1_{u,j} = 1)) \to 0$，当 $m \to \infty$ 时。

[4] 请注意，这两个初步结果相互独立，不依赖于推荐系统。

证明。 这是因为合取命题的后半部分对应于所有用户在第一轮都被推荐了排名前 i 的内容创作者的情形。因此，$\mathbb{P}((a^1_{:,\,i} > a^1_{:,\,j}\ \forall j \in n) \wedge (\forall u \in m,\ \exists j \in i\ a^1_{u,\,j} = 1)) \leq \mathbb{P}(\forall u \in m,\ \exists j \in i\ a^1_{u,\,j} = 1) = \binom{i}{n}^m \to 0$ 如 $m \to \infty$。□

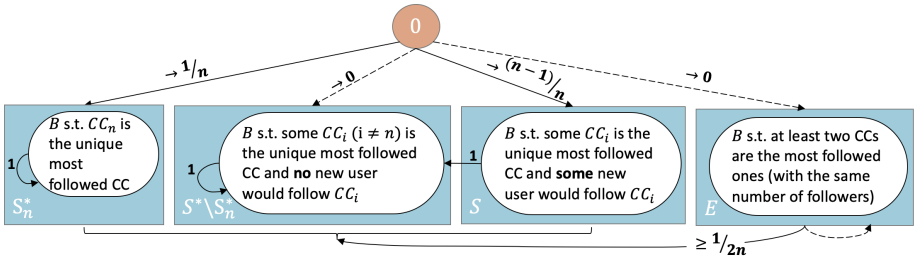


图2. ExtremePA 的马尔可夫链总结，如 $m \to \infty$ 所示。矩形表示状态集合。转移概率标注的是极限情况下的值。虚线表示在 $m \to \infty$ 时可忽略不计的转移。

定理 4. 在 *ExtremePA* 下，预期吸收时间趋于 $2-1/n$，当 m 趋于无穷时。

证明。 根据定理1和定理3，我们可以将状态划分为若干子集，如图2所示。令 μ_B 为从状态 B 到吸收的预期时间。于是 $\mu_{B:B\in S^*} = 0$（因为 S^* 中的所有状态都是吸收态），且 $\mu_{B:B\in S} = 1$（因为 S 中的所有状态都在一个时间步内转移到 S^* 中的某个状态）。因此，

$$\mu_{\mathbf{0}} = 1 \cdot p_{\mathbf{0},S^*_n} + 1 \cdot p_{\mathbf{0},S-S^*_n} + 2 \cdot p_{\mathbf{0},S} + \sum_{B\in E} (1+\mu_B) \cdot p_{\mathbf{0},B}$$

我们可以利用引理2和引理3求出各状态集合之间的转移概率，如 $m \to \infty$ 所示：(a) $p_{\mathbf{0},\,S^*_n} \to \frac{1}{n}$，(b) $p_{\mathbf{0},\,S^*-S^*_n} \to 0$，(c) $p_{\mathbf{0},\,S} \to \frac{n-1}{n}$，以及 (d) $p_{\mathbf{0},\,E} \to 0$。此外，由于 $\sum_{B\in E} p_{\mathbf{0},\,B} \cdot (1+\mu_B) \leq p_{\mathbf{0},\,E} \cdot (1+\mu_{B^*})$（其中 $B^* = \arg\max_{B\in E} \mu_B$），且 $\mu_{B^*} \leq c$（c 为常数）[5]，结论得证。□

4.3 内容创作者的公平性

总结。 基于先前的结果，我们证明探索型推荐系统在事前和事后都满足 CC_1-公平性（推论1），而 ExtremePA 仅满足事前公平性（定理5）。事实上，推论2表明，在 ExtremePA 下达到 CC_1-公平吸收状态的概率会随着用户数量趋于无穷而趋于 $1/n$；而推论3表明，达到

[5] 通过使用引理 1 证明 $p_{B^*,\,S\cup S^*} \geq 1/2n$，即可轻松证明这一点适用于 $c = 4n/(n-1)$（尽管可以得到更好的界）。

所有内容创作者都获得公平结果的概率会趋于 $1/n!$ 最后，图3展示了在三种推荐系统中的任意一种下，每位内容创作者获得 CC_i 公平结果的概率。

要点。 本分析传达了几个重要信息。首先，它表明探索是实现 CC_1公平结果的关键。其次，尽管 ExtremePA 具有事前公平性，但它很少具有事后公平性，这凸显了不能只关注预期关注者数量的重要性。第三，我们的数值分析表明，尽管探索总能带来 CC_1公平结果，但无法保证其他内容创作者的公平性。不过，分析结果显示，与 ExtremePA 相比，探索型推荐系统分配公平性的方式更加均衡。

推论1。 在均匀随机和优先依附推荐下，$(A^t)_t$ 同时具有事后和事前 CC_1公平性。

证明。 这是定理 2 的直接推论。由于 B 是吸收态当且仅当$b_{\cdot,1} = m$，所有吸收态都是 CC_1 公平的（即具有事后公平性），且$\mathbb{E}[b_{\cdot,1}] = m \geq \mathbb{E}[b_{\cdot,i}]$ （即具有事前 CC_1公平性）。□

推论2。 在 *ExtremePA* 下，结果满足事后 CC_1 公平的概率随着 $m \to \infty$ 的变化趋于 $1/n$ 。

证明。 如果 CC_1 是第一轮后关注者最多的唯一内容创作者，则最终结果 (a) 总是满足 CC_1 公平；而最终结果 (b) 只有在 CC_1 是第一轮后关注者数量最多的内容创作者之一时，才满足 CC_1 公平。因此，使用图2中的记号：$p_{\mathbf{0},\ S_1 \cup S_1^*} \leq \mathbb{P}$（$CC_1$ – 公平）$\leq p_{\mathbf{0},\ S_1 \cup S_1^*} + p_{\mathbf{0},\ E}$. 由于如前所示，$p_{\mathbf{0},\ S_1 \cup S_1^*} \to 1/n$ 且 $p_{\mathbf{0},\ E} \to 0$，故结论成立。□

推论3。 在 *ExtremePA* 下，结果满足事后公平的概率随着 $m \to \infty$ 的变化趋于 $1/n!$

证明。 根据引理 2，当 $m \to \infty$ 时，第一轮结束后关注者数量相同的概率趋于零。忽略平局后，只有满足 $a^1_{\cdot,1} > \cdots > a^1_{\cdot,n}$ 的结果才是公平结果。根据对称性，$(a^1_{\cdot,i})_i$ 的所有严格排序具有相同的概率。由于这样的排序共有 $n!$ 种，因此$\mathbb{P}(a^1_{\cdot,1} > \cdots > a^1_{\cdot,n}) \to 1/n!$，与 $m \to \infty$相同。结论由此得出。□

图3所示的行为与关于事后公平性的理论结果一致。具体而言，虽然 UR 和优先依附在收敛时保证 CC_1-公平性，但 ExtremePA 仅以较低的概率（约 1%）实现这一点。更准确地说，事后 CC_i-公平性的概率随质量指数 i线性增长，从 $1/n$增长至 1。相比之下，UR 和优先依附通常会为排名前半部分的 CC 提供更公平的结果，而为排名后半部分的 CC 提供较不公平的结果。在两者之中，探索性最强的推荐策略（UR）对大多数高质量 CC 更公平，而优先依附仅对少数排名靠前（和靠后）的高质量 CC 公平。

定理5。 在 *ExtremePA* 下，最终结果是事前公平的。

证明。 证明基于以下观察：当 CC_i 成为关注者最多的 CC时，最终所有此前未关注过更高质量内容创作者的内容创作者都会关注他们

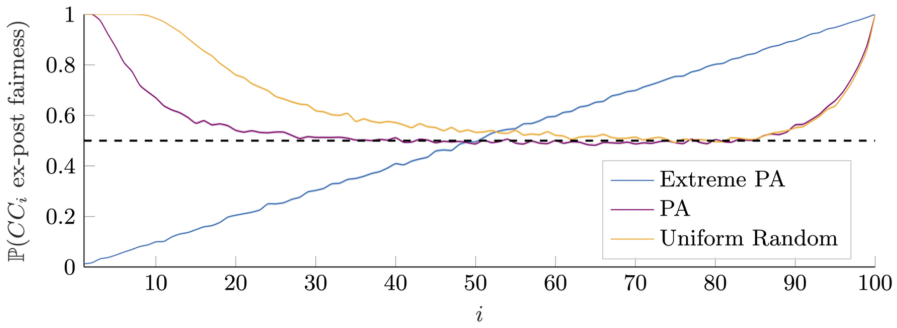


图3。 在三种推荐系统下，事后 CC_i-公平性的抽样概率：ExtremePA（蓝色）、优先依附（紫色）和均匀随机（橙色）。对于每种推荐系统，我们在 $n = 100$ 和 $m = 10000$ 条件下运行 10000 次模拟，直至收敛。虚线表示参考值 0.5。ExtremePA 下的公平性从质量最高的内容创作者到质量最低的内容创作者逐渐提高。优先依附和均匀随机对最佳内容创作者实现了更高的公平性。

。例如，如果 CC_2 在第一轮后拥有最多关注者，那么除被推荐 CC_1的用户外，所有用户都会在第 2 轮后关注CC_2。相反，如果 CC_1 拥有最多关注者，那么最终所有人都会转而关注 CC_1；而如果 CC_3 拥有最多关注者，那么除了原本关注 CC_1 或 CC_2 的用户外，所有人都会关注 CC_3。直观而言，这意味着 CC_2 的预期关注者数量多于 CC_3，但少于 CC_1。为简便起见，我们只对 $n = 2$形式化这一直觉：

$$\mathbb{E}[a^{\infty}_{\cdot,1}] = \sum_{k=0}^{m} \mathbb{P}(a^{1}_{\cdot,1} = k) \cdot \mathbb{E}[a^{\infty}_{\cdot,1} | a^{1}_{\cdot,1} = k] = \sum_{k=0}^{\left[\frac{m-1}{2}\right]} \frac{\binom{k}{m}}{n^m} \cdot k + \sum_{k=\left[\frac{m-1}{2}\right]+1}^{m} \frac{\binom{k}{m}}{n^m} \cdot m.$$

由于总和大于 $\mathbb{E}[a^{\infty}_{\cdot,2}] = \sum_{k=0}^{m} \binom{k}{m}_{n^m} \cdot k$，ExtremePA 具有事前公平性。 □

5 结论

在这项工作中，我们分析了推荐在基于用户生成内容的社交网络中对内容创作者个体公平性所发挥的作用。为此，我们（a）在先前模型的基础上扩展了推荐策略框架，（b）引入了一种非探索型推荐策略，并且（c）定义了事前和事后的公平性度量。我们的结果表明，网络形成过程是一个吸收马尔可夫链，其吸收状态、期望时间和公平性保证因推荐策略而异。具体而言，ExtremePA 下的期望吸收时间有上界 2，也就是说，远快于 UR 或优先依附下的期望吸收时间（后者随 m 线性或次线性增长）。此外，虽然所有受研究的推荐系统都能保证事前公平性，但事后公平性很少能够实现：在没有探索的情况下（即在

ExtremePA 下），实现事后公平性的概率为 $1/n!$；即使进行探索（即在 UR 或优先依附下），也只有对表现最佳（且最低）的内容创作者才能保证事后公平性。从本质上说，推荐系统中的探索，是以更快的吸收时间换取提高表现最佳内容创作者获得公平结果的概率。

参考文献

1. Eytan Bakshy、Itamar Rosenn、Cameron Marlow 和 Lada Adamic。社交网络在信息扩散中的作用。载于 *Proceedings of the 21st international conference on World Wide Web* 519 528, 2012 ，第 – 页。

2. Wendy Hall、Ramine Tinati 和 Will Jennings。从英国脱欧到特朗普：社交媒体在民主中的作用。 *Computer*，51(1):18–27，2018。 3. J Stacy Adams。探讨不公平现象。 *The journal of abnormal and social psychology* 67 5 422, 1963 ，( ): 。 4. Nicol`o Pagan、Wenjun Mei、Cheng Li 和 Florian D¨orfler。社交媒体影响者崛起的精英主义网络形成模型。 *Nature communications*，12(1 1( ): –12，2021。 5. George Kingsley Zipf。 *Human behavior and the principle of least effort: An introduction to human ecology* 2016 。Ravenio Books, 。 6. Sean M McNee、John Riedl 和 Joseph A Konstan。仅有准确性还不够：准确率指标如何损害推荐系统。载于 *CHI'06 extended abstracts on Human factors in computing systems* 1097 1101, 2006 ，第 – 页。

7. Matevˇz Kunaver 和 Tomaˇz Poˇzrl。推荐系统中的多样性——综述。 *Knowledge-based systems*，123:154–162，2017。 8. Natali Helberger、Kari Karppinen 和 Lucia D'Acunto。将曝光多样性作为推荐系统的一项设计原则。 *Information, Communication & Society*，21 2 191( ): –207，2018。 9. Pietro Gravino、Bernardo Monechi 和 Vittorio Loreto。迈向新颖性驱动的推荐系统。 *Comptes Rendus Physique*，20(4):371–379，2019。 10. Wenshuo Guo、Karl Krauth、Michael Jordan 和 Nikhil Garg。协同过滤推荐系统中的刻板印象问题。载于 *Equity and Access in Algorithms Mechanisms and Optimization* 1 10 2021 ， ， ，第 – 页。美国计算机协会，。 11. Roger B Myerson。功利主义、平等主义与社会选择问题中的时机效应。 *Econometrica: Journal of the Econometric Society*，第 883–897 页，1981。 12. Allison JB Chaney、Brandon M Stewart 和 Barbara E Engelhardt。推荐系统中的算法混杂如何增加同质性并降低效用。载于 *Proceedings of the 12th ACM Conference on Recommender Systems*，第 224–232 页，2018。 13. Duncan J Watts 和 Steven H Strogatz。" 小世界 " 网络的集体动力学。 *Nature*，393(6684):440，1998。 14. Albert-L´aszl´o Barab´asi 和 R´eka Albert。随机网络中标度的涌现。*Science*，286(5439):509–512，1999。 15. Tom AB Snijders。用于研究网络变化的随机行动者导向模型。*of mathematical sociology* 21 1 2 149 172, 1996 *Journal* ，(-): –。 16. Matthew O Jackson。 *Social and economic networks*。普林斯顿大学出版社，2010。 17. Sahil Verma 和 Julia Rubin。公平性定义解析。载于 *2018 ieee/acm international workshop on software fairness (fairware)* 1 7 2018 ，第 – 页。IEEE，。

18. Pratyush Garg、John Villasenor 和 Virginia Foggo。公平性指标：比较分析。载于 *2020 IEEE International Conference on Big Data (Big Data)*，第 3662–3666 页。IEEE, 2020。 19. Shira Mitchell、Eric Potash、Solon Barocas、Alexander D'Amour 和 Kristian Lum。算法公平性：选择、假设与定义。 *Annual Review of Statistics and Its Application* 8 141 163, 2021 ，：–。

20. Reuben Binns。论个体公平性与群体公平性之间表面上的冲突。载于 *Proceedings of the 2020 conference on fairness, accountability, and transparency*，第 514–524 页，2020。 21. Eli Lucherini、Matthew Sun、Amy Winecoff 和 Arvind Narayanan。T-recs：用于研究推荐系统社会影响的模拟工具。 *arXiv preprint arXiv:2107.08959* 2021 ，。