arXiv:2305.00089v1 [cs.DL] 28 Apr 2023

学术文章参考文献列表长度模型

FATEMEH GHAFFARI 和 MARK C. WILSON

Abstract. 我们提出并分析了一种简单的文章产出与引文行为概率模型，该模型明确假设，某篇文章的可引用性不会随时间推移而下降。模型对文章参考文献列表中条目的数量和年代作出预测。后一个主题此前已有研究，但仅限于数据分析；据我们所知，尚未有人提出相关模型。随后，我们对多个学科的参考文献列表长度进行了大规模分析。结果表明，我们无法否定这个简单模型，而且它与汇总后的参考文献列表数据拟合得相当好。过去几十年来，出版物总数与参考文献列表平均长度之间的关系高度近似线性。尽管我们的模型显然过于简化，但很可能有助于进一步构建学术文献模型。最后，我们将这项工作与有关学术出版物“老化”或“过时”的大量文献联系起来，并指出该研究领域的重要性如今已不再明晰，而现有文献中有许多内容既令人困惑，也容易使人困惑。

1. Introduction

科学发展（我们将其视为“已发表的学术研究成果”的同义词，包括人文学科和社会科学等领域）可以通过已发表文章的数量、作者数量、被引文章数量等指标来衡量。这些指标都已得到广泛研究，并且每一个都是需要完整统计的“全球性”指标。更“局部”的指标是文章中参考文献列表的长度（即参考文献条目的数量）。通过对相对少量的文章进行抽样，可以较好地估计其分布，或仅估计平均值或中位数等统计量。另一个便捷且显而易见的特点是，已发表文章的参考文献列表会随时间保持不变。

参考文献列表长度的总体分布显然与引文的总体分布相关。如果我们考虑这样一个有向图：每篇文章都是一个节点，当且仅当文章 X 在其参考文献列表中包含文章 Y 时，就存在一条从 X 指向 Y 的边，那么总出度等于总入度：换言之，所有文章的参考文献数量之和等于所有文章获得的引文总数。然而，某篇文章获得的引文数量会随时间增加，这使得参考文献列表在许多计算中更为方便。例如，参考文献列表长度的相关细节已被用于引文分析中的标准化，以及其他主题 [4]。这种对同步（考察参考文献列表）与历时（考察引文）文献计量分析的区分，自 20 世纪 70 年代初 Line & Sandison 以来，已被许多作者提出。术语跨时同步也被用于描述同步研究的时间序列。

参考文献列表长度相较于收到的引文数量或发表的文章数量，受到的研究要少得多。我们推荐 [7] ，以回顾较早期的文献；这些文献大多汇集了关于不同研究领域或特定期刊中参考文献列表的小样本统计数据。从早期数据中很容易看出，几十年前的论文所列参考文献往往少于如今撰写的论文。更精细的分析表明，参考文献数量

Date：2023年5月2日。*Key words and phrases.* 文献计量学、Scientometrics、科学文献增长。

一直在增加，不同研究领域的参考文献列表长度在平均值或中位数方面差异显著，不同文章类型之间也存在很大差异（例如，综述文章所列参考文献通常多于原创研究文章）。近期汇编了更详尽的数据 [6, 2]。

1.1. **我们的贡献**。 尽管已有多位研究者对已发表文章数量的增长进行了建模，但据我们所知，尚无人对参考文献列表的长度进行建模——我们所知的所有先前研究都只是基于数据进行描述。我们提出了一个简单的文章产出连续时间模型，用以关联各项数量，并探究其行为。该模型明确排除了任何“过时”过程，即较早发表的文章比新近发表的文章更不容易被引用。我们使用了一个大型数据集，其中收录了2006—2016年发表的文章，并表明尽管该模型显然过于简化，但对我们的研究目的而言，其拟合效果出奇地好。具体而言，我们发现，在所研究的时期内，参考文献列表平均长度的增长可以很好地用线性函数建模，文章产出数量的增长亦是如此；而文章总数与参考文献列表平均长度之间则呈现出明显的线性关系。我们还发现，该模型对参考文献列表长度分布的拟合效果优于预期（尽管并不理想）；这对模型而言是更为严格的检验。总体而言，该模型与数据拟合良好，这为其在其他情境中的应用提供了支持。我们还对模型进行了推广，这可能有助于开展更精细的建模。

我们将上述研究与有关老化和过时的文献联系起来，并回顾该领域的一些混淆之处。

2. A BASIC MODEL

此前的作者（例如 [7, 6]）提出了一些可能会增加参考文献列表长度的关键因素：

- 随着时间推移，研究人员可引用的作品自然会越来越多；
- 现代作者使用电子关键词搜索，相较于纸质期刊时代，这让他们更容易找到相关参考文献；
- 由于在线出版，页数限制有所放宽，因此有更多篇幅可以引用其他作品；
- 引用实践的变化（例如：如今同行评审更加严格，审稿人经常建议引用某些论文；礼节性引文现在可能也更为普遍）。

最主要的可能制衡因素是科学文章的“过时”，许多作者都对此进行过研究。文献的快速增长往往伴随着理解的提升，这意味着较早的参考文献会被较新的文献完全涵盖，不再被引用（当然，牛顿的原始著作如今获得的引文很少）；此外，作者可能重视新颖性，而且用于文献综述的精力有限。因此，较早的论文可能会变得不那么“值得引用”，这里的意思仅仅是，与较新的论文相比，它们对引用者的吸引力较小。不过，与此相对的是，参考文献往往通过“滚雪球抽样”获得：研究者阅读一篇文章，然后追溯其中的参考文献，如此等等。这种倾向想必会使较早的论文占据优势。

因此，我们很难判断可引用性的时间演变应呈现何种规律。不过，上述考量引出了两个明确的研究假设。假设 A 认为，文章的可引用性不随时间变化。换言之，对较早发表的文章既没有明显的偏好，也没有明显的排斥。假设B 认为，文章的可引用性会随时间下降，因此人们会偏好较新的文章。在有关学术文章老化／过时的文献中，许多作者都采用了假设B（认为可引用性在出版后不久先有所上升，随后便会下降）。

模型。 考虑以下文章产出模型。假设文章连续出版，并令 $P(t)$ 表示截至时间 t 已出版文章的总数，因此 $P'(t)$ 表示时间 t 的瞬时出版率。假设每篇新文章引用每篇此前已出版文章的概率为 $q(a)$，且该概率仅取决于文章的年龄 a。换言之，时间 t 出版的文章 X 对时间 $s < t$ 出版的每篇文章 Y 所作的引用数是一个伯努利随机变量$c_{t-s,\ XY}$；对于给定的值 $t - s$，这些变量独立同分布。

明确采用假设 A 的最简单特例，是 q 与 a相互独立的情形。换句话说，文章 X 在时间 t 发表时，对每篇在时间 $< t$ 发表的文章 Y 的引文数量是伯努利随机变量 c_{XY}；这些变量独立同分布，且与 t独立。下文将其称为“均匀模型”。

模型中的量 q 即使在最简单的情形下保持恒定，也预计会因研究领域而异。下文的数据表明，不同学科之间的参考文献数量差异显著（这是由不同领域引文行为的文化差异造成的）。例如，据我们的经验，数学文章通常写得相当简洁，所引文献大多是文章直接建立在其基础之上的研究。相比之下，经济学文章以引言冗长、引文繁多而闻名——事实上，审稿人经常会拒绝那些没有讨论相关性较弱的文献的稿件。

模型行为。 我们的第一个预测是，在时间 t 发表的每篇文章的参考文献列表的期望长度 $L(t)$为

$$L(t) = \int_{-\infty}^{t} q(t-s)P'(s)\,ds.$$

其直观含义是， $P'(s)\ ds$ 是小时间间隔 $[s, s+ds]$内发表的论文的大致数量，这些论文的年龄约为 $t - s$，因此每篇论文被引用的概率为$q(t - s)$。

在均匀模型下，我们有以下特殊情形

$$L(t) = qP(t).$$

求导后，我们看到 在这种情况下， L 的增长率与 P的增长率成正比 .

通常，从某个 $t_0 > -\infty$ 开始计时会很方便（例如，因为我们只有时间 t_0之后的数据）。在这种情况下，我们只需改写即可得到

$$L(t) = L(t_0) + \int_{t_0}^{t} q(t-s)P'(s)\,ds.$$

该模型的一个重要特点是，由于我们假设被引用的文章是均匀随机选取的，因此，如果我们考虑只统计在时间 t_0 或之后发表的文章所获引文这一限制，分析结果基本不变。例如，将 $L^*(t)$ 定义为在时间 t_0 或之后发表的参考文献数量，并将 $P^*(t) = P(t) - P(t_0)$ 定义为在时间 t_0 或之后发表的文章数量，我们有

$$L^*(t) = \int_{t_0}^{t} q(t-s)(P^*)'(s)\,ds.$$

3. 实验设置

我们使用了 Academic Analytics 公司提供的数据集 https://aarcresearch.com/。该数据集相当全面，涵盖了北美 390 所授予博士学位的大学中教职人员在 2006—2016年期间的所有出版物。有关所用流程的更多详情，请参见

附录 A。我们指出，对于本文的大部分分析而言，这样全面的数据集并非必需，因为随机抽样已经足够。

我们剔除了数据缺失的条目，以及参考文献少于 5 条的条目（以排除致编辑的信）。该数据集是从给定时间段内发表的全部文章中抽取的样本；就参考文献列表长度而言，我们认为它很可能不存在偏差。该数据集中，2006—2016年化学、数学、经济学和肿瘤学领域发表的文章分别为（147149、53776、30160、16957）篇，并包含每篇文章的页数和参考文献数量。

我们还需要估算 $P(t)$。为此，我们在 Scopus 中进行了手动查询。当然，我们不知道 Scopus 使用的研究领域分类是否与所提供数据集中的分类相同。这也是我们将研究范围限定在上述四个领域的原因之一——原始数据集还包含历史学和地理学等其他研究领域，而我们预计这些领域的分类可能存在一些不一致之处。

所有统计分析均使用 Python 及 numpy、pandas 和 matplotlib 等标准软件包完成。

4. Results

在所研究的每个研究领域中，我们发现，2006–2016年间，每年发表的文章平均数量以及每篇文章的参考文献数量均稳步增加。普通线性回归对两者的拟合效果都很好（见表 1），这表明参考文献列表长度随时间呈线性增长。我们还发现，不同研究领域的平均参考文献列表长度存在显著且一致的差异，如图 1（右图）所示。

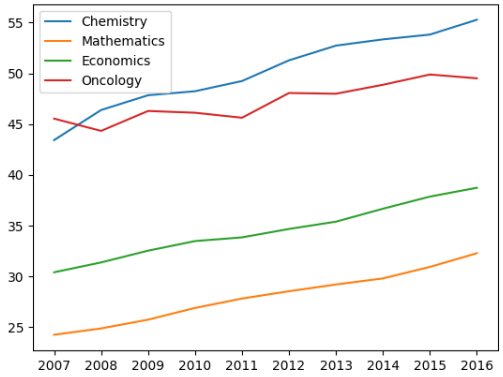
领域	R^2 对于 $L(t)$	R^2 对于 $P(t)$	R^2 针对 L与 P的比较	R^2 针对 $P\ (t)\ - P\ (t-1)$
化学	0.405948	0.997372	0.963810	0.964448
数学	0.605758	0.997737	0.993011	0.947762
经济学	0.521143	0.968208	0.978862	0.908734
肿瘤学	0.269766	0.993056	0.888816	0.969280

Table 1. 线性模型对 $L(t)$和 $P(t)$的拟合，2006—2016年（如图1 和2所示）。

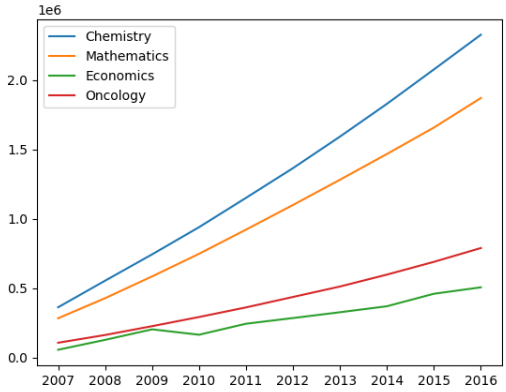
我们将平均参考文献列表长度 $L(t)$与我们根据 Scopus 对 $P(t)$的估计值进行了比较。我们的模型在均匀情形下预测两者呈仿射关系（其中存在一个非零常数项，源自截至 2006年发表的所有文章），其形式为 $L(t) = q\,(P(t) - P(t_0)) + d$。拟合效果极佳（见图2和表1）。我们预计，不同的研究文化会使模型中的 q 取值有所不同；例如，如果经济学与数学在一定程度上属于不相交领域，我们预计经济学模型中的 q 取值会大于数学模型。我们通过上述最小二乘拟合得到的斜率值确实存在显著差异，经济学的斜率远大于数学。

我们对上述结果进行了若干稳健性检验。附录展示了将平均值替换为中位数只会带来相对较小的变化。我们还使用另一个数据集进行了检验。详细研究 [2] 的作者慷慨提供的数据清楚地表明，大约从 1980年起， P 和 L呈现大致线性增长，并且 P 可以用关于 L (见图3)的线性函数很好地拟合。

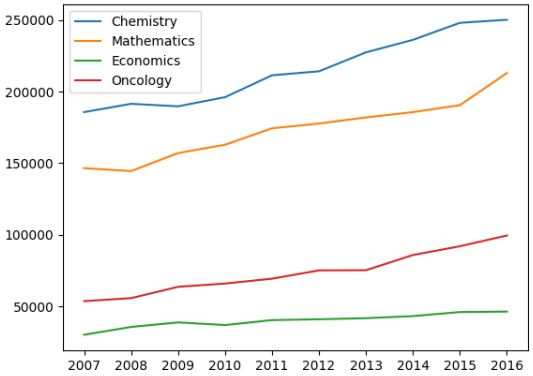
4.1. **我们的模型与指数增长阶段。** 我们的均匀模型预测，在文章数量呈指数增长的时期，参考文献列表长度也将呈指数增长。作者提供的 [2] 数据对此明确提出了反证：1900–2004 年各汇总子领域中参考文献列表的增长、 P的增长，以及 L 与 P 之间的关系见图4。



(A) $L(t)$ 与 t



(B) $P(t)$ 与 t



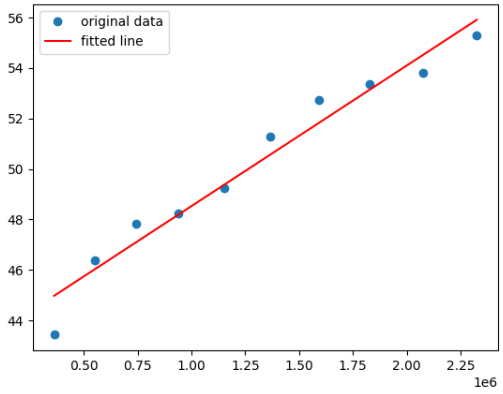
(C) $P(t) - P(t-1)$ 对比 t

FIGURE 1. t、 $P(t)$ 与 $L(t)$ 之间的关系。

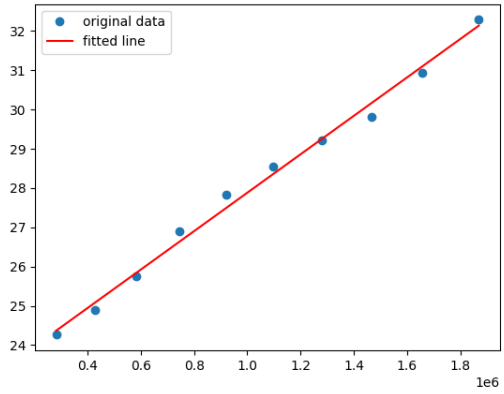
领域	R^2
医学	0.982209
自然科学	0.994795
社会科学	0.990064
艺术与人文学科	0.925461

TABLE 2. 线性模型对 $L(t)$ 与 $P(t)$ 的拟合，数据集来自 [2], 1985–2004 .

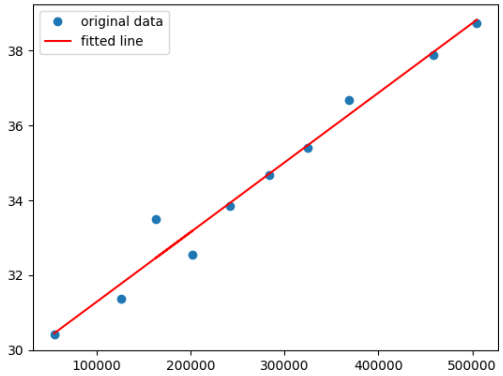
事实上，正如下列计算所示，假设 $P(t) = Ce^{kt}$，在我们的更一般模型中，任何函数 q 都无法在文献呈指数增长时使参考文献列表呈次指数增长：



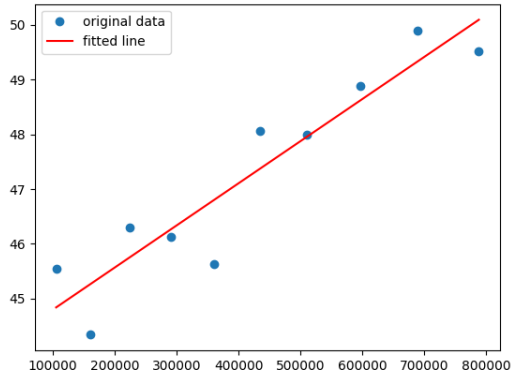
(A) 化学 – 斜率： 6×10^{-6}



(B) 数学 – 斜率： 5×10^{-6}



(C) 经济学 – 斜率： 19×10^{-6}



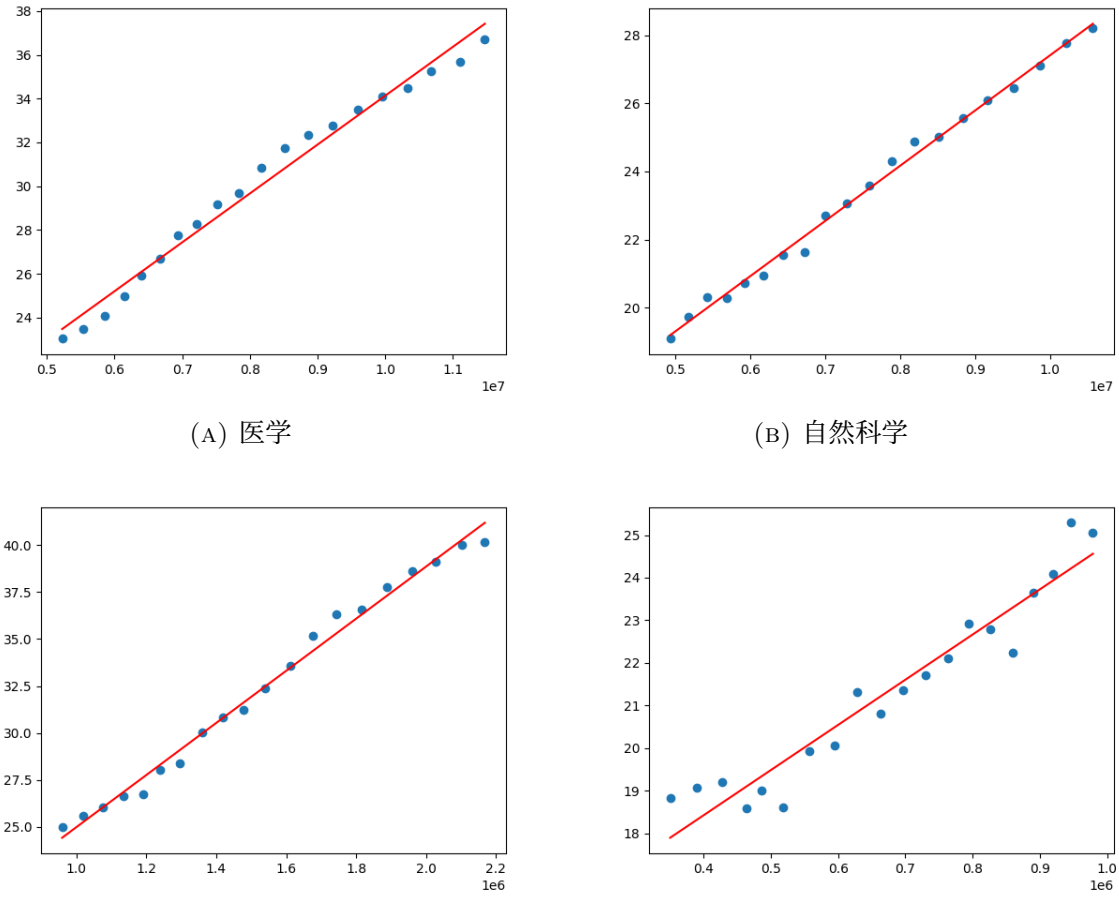
(D) 肿瘤学 – 斜率： 8×10^{-6}

FIGURE 2. $P(t)$ 与 $L(t)$ 之间的线性关系。

$$\begin{aligned} L(t) &= Ck \int_{-\infty}^{t} q(t-s)e^{ks}\, ds \\ &= Ck \int_{0}^{\infty} q(a)e^{k(t-a)}\, da \\ &= Cke^{kt} \int_{0}^{\infty} q(a)e^{-ka}\, da. \end{aligned}$$

最后一行中的积分是一个非负常数。

这些结果并不令人意外，因为在这种情况下，该模型缺乏依据。我们不能合理地假设，在科学的指数增长阶段（20世纪80年代之前），作者能够始终找到所有相关参考文献。例如，那段时期还没有出现广泛用于检索出版物的电子检索。更合理的假设是，作者在自己读过的文章的参考文献列表中增添了少量参考文献。这会使 L增长相当缓慢，或许略快于线性增长。



(A) 医学 (B) 自然科学

(C) 社会科学 (D) 艺术与人文学科

FIGURE 3. .....之间的线性关系 $P(t)$ 和 $L(t)$ 来自.....的数据集 [2], 1985–2004。

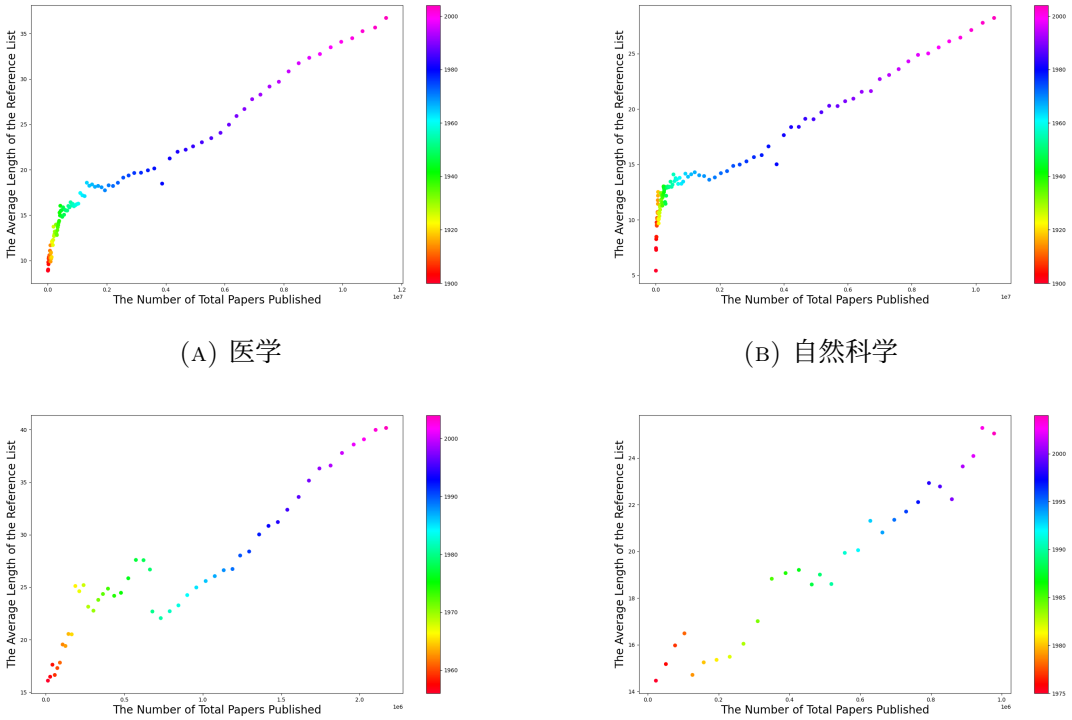
5. CONNECTION TO THE LITERATURE ON OBSOLESCENCE AND AGING

5.1. 过时与老化。 文献计量学中一个看似截然不同的主题是科学文献的老化/过时。我们审阅了至少30篇文献计量学领域有关这些主题的文章，并对该领域文献提出了几点批判性意见。

首先，我们来厘清术语。有些作者将“老化”用作“过时”的同义词。过时是特定文章的一种属性（即随着时间推移，其被引用的吸引力逐渐降低）。文章会随着时间推移而变旧，但仅仅因为一篇文章发表时间更久，就说它变得过时，并不十分准确。事实上，这也是因为在此期间发表了更多相关性更强、更值得引用的文章。

有些作者用“老化”一词来指参考文献列表中条目的年龄（平均值或中位数）随时间增加。这种用法很容易造成混淆，因为参考文献列表中条目的年龄增加，是由于文章“老化”（即过时意义上的老化）速度降低了！

从现在起，我们只讨论文章的过时和参考文献列表的近因偏差。我们关于参考文献列表长度的研究与这些主题的联系如下。



(A) 医学 (B) 自然科学

(C) 社会科学 (D) 艺术与人文学科

FIGURE 4. 之间的关系 $P(t)$ 与 $L(t)$ 来自数据集 [2], 1900–200 4.

（1）科学文献的增长会导致参考文献列表出现近因偏差：数量更多的新条目会挤掉较早的条目。即使参考文献列表的长度与文章产出的增长保持同步，这一点仍然成立。即便不假设存在任何老化效应（文章的可引用性随时间推移而下降），参考文献列表也会出现近因偏差。（2）不过，任何与年龄相关的老化效应都会加剧参考文献列表的近因偏差。（3）许多作者将过时（即假设B）视为既定事实，并利用参考文献列表中的近因偏差来估算过时程度。（4）参考文献列表中的近因偏差由文献增长和过时共同导致。

第一点成立，因为正如下面第5.2节所讨论的，这种现象在我们的均匀模型中也能观察到。在这个模型中，在时间 t 被引用且年龄至少为 a 的论文所占比例，简单来说就是 $P(t-a)/P(t)$，并且会随着 a 增大而下降。这种显而易见的效应此前已有论述。例如，在 [2]中，我们发现：“我们认为，被引用的科学信息年龄增长的主要原因，是对战后已出版材料数量惊人增长以及当前科学发展增长放缓所作出的一种相当机械化的反应。”

第二点很清楚。如果 q 是 a的递减函数，至少在初始的短暂时期之后如此，那么我们预期这个关于 $L(t)$的表达式会小于 q 保持不变时的值。

现在我们来讨论上文列出的第三点。Line & Sandison [3] 指出，过时是一种必须加以验证的假设，而且有证据表明它确实会发生。Larivière、Archambault 和 Gingras [2] 则将过时视为常识（按照这类假设的惯例，先假定可引用性会有所提升）。他们对发表于1900–2004年间的文章参考文献列表中的条目年代进行了异时研究

，发现：“事实上，与普遍看法相反，如今科学文献并没有更快过时；实际观察到的情况恰恰相反。自20世纪70年代中期以来，科学出版物的有效寿命一直在稳步延长。”只要其他条件相同，过时速度加快就应导致参考文献列表中条目的平均年代降低；就这一点而言，上述结论是正确的。然而，这项研究并未证明过时确实存在——正如上文所述，参考文献列表中的近因偏差无需借助过时也能解释。

现在我们讨论上面列出的第四点。Line & Sandison [3] 大约在50年前对这一领域进行了综述，我们的大多数观察结果当时已经被他们或更早的作者记录下来。然而，文章的过时与参考文献列表中的近因偏差之间的区别，以及过时和文献增长都会对参考文献列表的近因偏差产生作用这一事实，后来却被几位作者忽视了。例如，Gupta [1, 第 336]页明确地将以下内容作为采用同步方法研究过时现象的依据：“如果出版物的使用率随其年代增长而下降，则可推定存在出版物过时现象。随着时间推移，使用文献的减少（老化或衰减）可以通过图书馆使用研究，或通过研究出版物中引文的年代，或被引文章的年代来确定。”

对过时与近因偏差的某些混淆，可能是由条件概率中的一个错误造成的：我们想知道 $P(cited \mid old)$ 是否很小，但我们实际观察到的却是 $P(old \mid cited)$。贝叶斯公式表明

$$P(cited \mid old)P(old) = P(old \mid cited)P(cited).$$

由于 $P(old)$ 可能远小于 $P(cited)$，我们无法得出所期望的结论。

在 [2] 中，我们发现：“我们认为，被引用的科学信息平均年龄增加的主要原因，是对战后已出版材料数量惊人增长以及当前科学发展增速放缓的一种相当机械的反应。”正如我们的均匀模型所示，增长本身无法解释参考文献列表中的近因偏差。我们的模型推论之一是，文章产出放缓会导致参考文献列表中条目的平均年龄上升。然而，如果过时确实是一个影响因素，那么这一特性就未必成立。

作为一项更为普遍的批评，我们认为，如今关于过时的文献在实践中的相关性相当有限，其中一些研究从一开始似乎就存在误解。正如 Line & Sandison [3], 所解释的，我们应当区分文献的过时与知识的过时。在印刷品馆藏时代，书架空间有限，因此图书馆员必须判断哪些连续出版物今后不太可能被读者查阅。然而，在电子时代，人们可以轻松获取那些不太受欢迎的长尾文献，文献过时便不再那么令人担忧。知识过时是否是一个重大问题，甚至它是否曾经是个问题，都尚不明确。

总之，我们发现关于老化/过时的文献亟待改进，并希望我们在此所作的贡献能推动未来开展更多严谨的研究。

5.2. **均匀模型与年龄分布**。在我们的通用模型下，考虑在时间 t 发表的文章的参考文献列表。此类列表中所有条目的预期总年龄为

$$A(t) := \int_{-\infty}^{t} (t-s)q(t-s)P'(s)\,ds.$$

请注意，$A(t)/L(t)$ 给出了在时间 t 发表的随机选取文章的参考文献列表中一个条目的平均年龄。

现在我们将范围限定在均匀模型中，其中对所有 $a \geq 0$，$q(a)$ 都是常数。由于缺乏关于 P 的历史数据，要全面分析这一问题会很困难，因此我们将范围限定为

由在 t_0 或之后发表的文章对在 t_0 或之后发表的文章所作的引文，从而得到公式

$$A^*(t) = q \int_{t_0}^{t} P^*(s)\, ds.$$

通过分部积分，我们可以将这个量改写为

$$\begin{aligned} A^*(t) &= q \int_{t_0}^{t} (t-s) P'(s)\, ds \\ &= qt \int_{t_0}^{t} P'(s)\, ds - q \int_{t_0}^{t} sP'(s)\, ds \\ &= qt\,(P(t) - P(t_0)) - q\,[sP(s)]_{t_0}^{t} + q \int_{t_0}^{t} P(s)\, ds \\ &= q \int_{t_0}^{t} P(s) - q(t-t_0)P(t_0)\, ds \\ &= q \int_{t_0}^{t} (P^*(s) - P(t_0))\, ds \\ &= q \int_{t_0}^{t} P^*(s)\, ds. \end{aligned}$$

我们采用“平均场近似”方法，通过以下公式估算条目的预期平均年龄：

$$\frac{A^*(t)}{L^*(t)} = \frac{\int_{t_0}^{t} P^*(s)}{P^*(t)}. \tag{1}$$

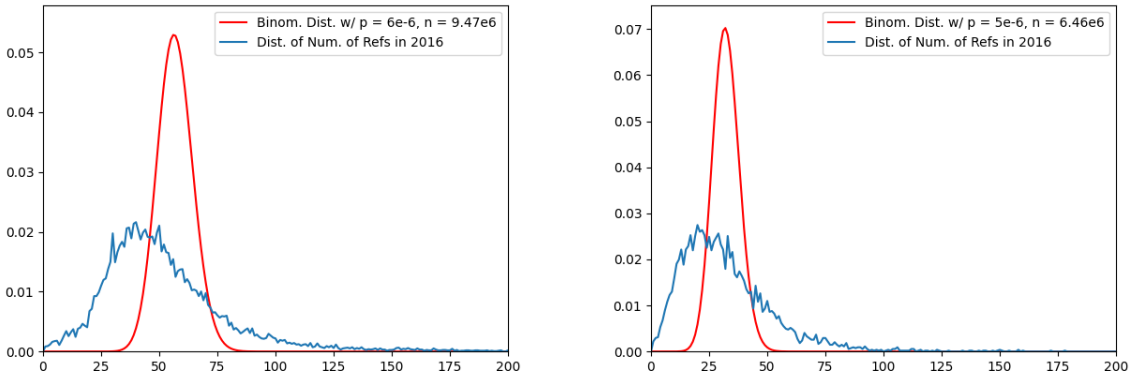
通过求解 $P^*(t-a) = P^*(t)/2$ 来得到中位数，其中 a。请注意，预期平均年龄至多为 $t-t_0$，且等于由点 $(t_0, P(t_0))$、$(t, P(t_0))$、$(t, P(t))$ 和 $(t_0, P(t))$ 构成的矩形中位于曲线 $y = P(t)$ 下方部分所占比例的倒数。例如，如果 P^* 随着 t 呈二次增长（对应于单位时间内产出的文章数量线性增加），那么这些近似值给出的平均年龄为 $(t-t_0)/3$，中位数为 $(t-t_0)(1-2^{-1/3}) \approx 0.206(t-t_0)$。我们在数据集上检验了公式 (1)——表3展示了预测结果。为了获取参考文献列表中条目的年龄数据，我们从每个领域随机抽取了30篇2016年出版的文章，并从 CrossRef 提取其数据。这需要手动查询：由于无法通过 CrossRef 的 API 获取出版物参考文献的详细信息，因此必须为每个条目手动提取完整的参考文献列表。随后，我们使用 Python 程序分析参考文献列表，并计算被引用条目的样本均值和中位数。仅考虑了2006年之后出版的参考文献。结果也列于表3。考虑到样本量，拟合效果良好。

领域	平均年龄（预测）	平均年龄（抽样）	中位年龄（预测）	中位年龄（抽样）
化学	5.36	4.74	5	5
数学	5.40	5.19	5	5
经济学	4.62	5.37	5	5
肿瘤学	5.65	4.79	5	5

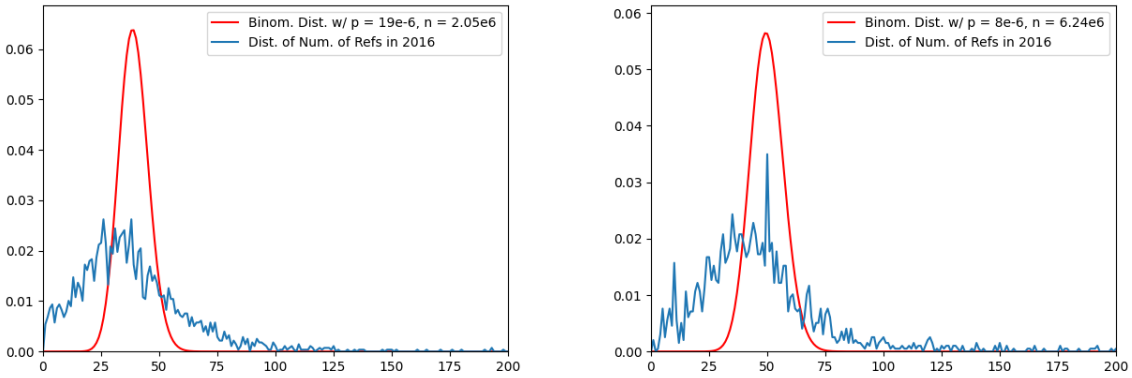
TABLE 3. 在文章发表于2016年且参考文献发表于2006年或之后的条件下，参考文献年代的预测值与样本统计量。

不能指望我们的平均场方法能够准确预测在时间 t发表的随机选取文章的参考文献列表长度分布。

模型预测，这应当遵循二项分布 $Bin\left(P\left(t\right)\ c\right)$。然而，我们确实在图5中考虑了对该模型的这一更严格检验。结果并不特别理想，这在意料之中，因此还应对参考文献列表长度的分布进行更为细致的建模。



(A) 化学　　(B) 数学



(C) 经济学　　(D) 肿瘤学

FIGURE 5. 201 年发表的文献平均参考文献列表长度的二项拟合　　6.

6. DISCUSSION AND CONCLUSION

我们的结果表明，所采用的均匀引文模型足以解释参考文献列表长度的时间演化，并将其与已发表文章数量的增长联系起来，至少在电子检索时代是如此。我们认为，该模型适用于此类统计研究，也可能用于其他领域。我们更一般的模型或许能够给出更详细的预测，但如何估算老化函数 q是一个棘手的问题。研究一个过去20年间才基本兴起、且发展迅速的现代领域，将会非常有趣，也能检验我们的均匀模型是否仍然适用。

Price [5] 早已讨论过研究领域在引文模式上的差异。数学等领域乍看之下似乎更适合均匀模型，在这种模型中，研究成果一旦具备可引用性，

其可引用性下降的速度可能不会像变化更快的领域那样迅速。我们希望能激励其他研究者对引文行为进行足够细致的研究，从而改进本文提出的模型，并更清楚地区分增长与过时。

上述实证和建模结果引出了有关政策的问题。每篇文章的参考文献数量不断增长（也称为“引文膨胀”），如果持续足够长时间，可能会令人担忧。此前已有作者指出这一点 [6, 7]。在单位时间阅读成本不变的情况下，学者核查参考文献将越来越困难。例如，我们的线性模型预测，平均每十年至少增加5条参考文献，而且这一数字很可能还会继续上升。这表明，我们需要更高效地运用机器阅读来处理文章，而这又要求原始研究文献能够开放获取。另一个更为激进的做法是由期刊限制参考文献的数量。这类似于资助或晋升申请者将重点放在最重要的出版物上。当然，这不会减少作者的工作量，但会帮助读者，也让人们更容易比较研究人员随时间变化的引文产出。或者，可以明确说明引用理由，并将文献引用标准化。例如，所有礼节性引文（作者提及他人的研究，以使审稿人相信作者熟悉相关知识，但本文并未直接以这些研究为基础）都可以列在一个标准位置，供读者将其归为此类；如果时间紧张，也可以忽略这些引文。

Funding and/or Conflicts of interests/Competing interests

作者声明不存在利益冲突。本研究未获得资助。

References

[1] Brij Gupta。《理论群体遗传学中文献的增长与过时》。载于：*Scientometrics* 42.3（1998年），第335–347页。[2] Vincent Larivi`ere、 ´Eric Archambault和Yves Gingras。《1900–2004年科学文献老化的长期模式》。载于：国际科学计量学与信息计量学会（*ISSI*）第*11*届会议论文集，马德里，西班牙。2007年，第449–456页。[3] Maurice B. Line和Alexander Sandison。《文献工作进展：“过时”与文献使用随时间的变化》。载于：*Journal of Documentation* 30.3（1974年），第283–350页。[4] Werner Marx和Lutz Bornmann。《视角转变：从被引参考文献的角度看文献计量学——关于文献计量学中被引参考文献评价方法的文献综述》。载于：*Scientometrics* 109.2（2016年），第1397–1415页。[5] Derek J. De Solla Price。《科学论文网络：书目参考文献的模式揭示了科学研究前沿的性质》。载于：*Science* 149.3683（1965年），第510–515页。[6] I˜naki Ucar等。《1972–2013年期间工程学期刊论文中参考文献数量的增长》。载于：*Scientometrics* 98.3（2014年）。出版商：Springer，第1855–1864页。[7] Moshe Yitzhaki和David Ben-Tamar。《生物化学及其他领域中的参考文献数量：Journal of Biological Chemistry 1910–1985年间的案例研究》。载于：*Scientometrics* 21.1（1991年）。出版商：Akad´emiai Kiad´o，与Springer Science+ Business Media BV联合出版……，第3–22页。

AppendixA. Datasets

以下数据集描述由 Academic Analytics 提供 .

我们从 Aca- demic Analytics 商业数据库（AAD；http://www.academicanalytics.com/）中筛选出学者在 2007年至 2019年 期间合著或独著的期刊文章的数字对象标识符（DOI）。AAD 由一份年度更新的教职人员名册组成，收录了受雇于

390 所授予博士学位的美国大学的教职人员。AAD 将教职人员与其合著或独著的、具有 CrossRef-DOI 的每篇期刊文章关联起来，采用人工和半自动化的消歧方法。AAD 还包含每位教职人员的学术院系隶属关系；这些院系又依据 NCES CIP 代码（https://nces.ed.gov/ipeds/cipcode/）由人工划分为不同学科。我们从 AAD 中提取了院系被归类到以下学科的学者所合著或独著的每篇期刊文章的 DOI：化学、经济学、英语语言文学、地理学、历史学、数学和肿瘤学。我们将每个 DOI 作为搜索词输入 CrossRef API（https://www.crossref.org/），以获取参考文献列表长度以及首页和末页的页码，据此计算每篇文章的总页数。

Appendix B. Robustness checks

B.1. **中位数与平均值**。我们使用中位数重复了正文中的分析，结果与之前相似，如图 6 和 7 以及表 4)。

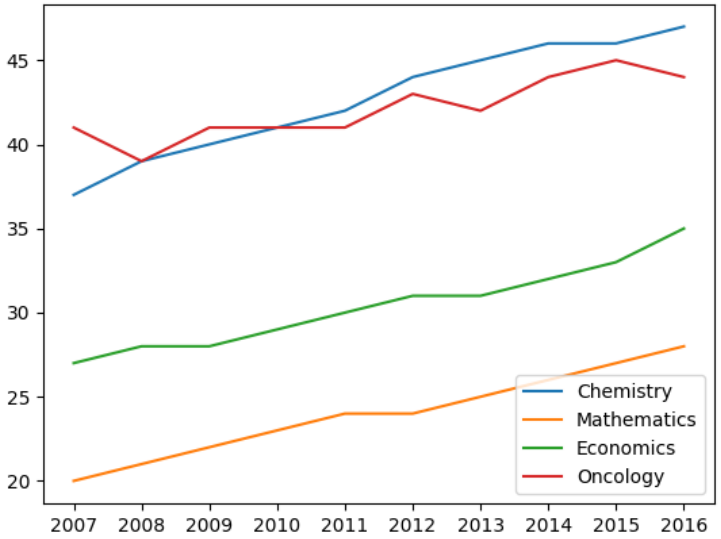
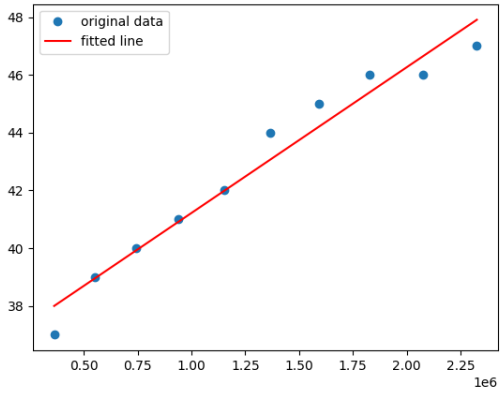


Figure 6. 按年份划分的参考文献列表长度中位数

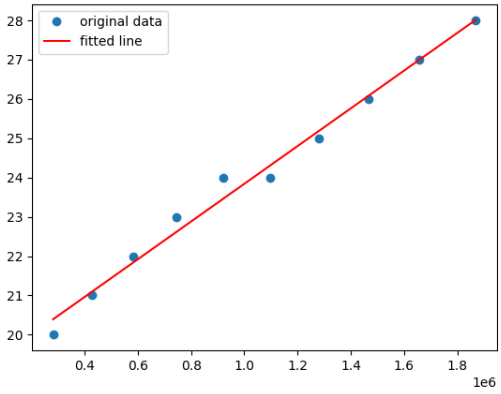
领域	R^2
化学	0.962511
数学	0.988723
经济学	0.980056
肿瘤学	0.805837

Table 4. L（t）与 P（t）的线性模型拟合 - 中位数

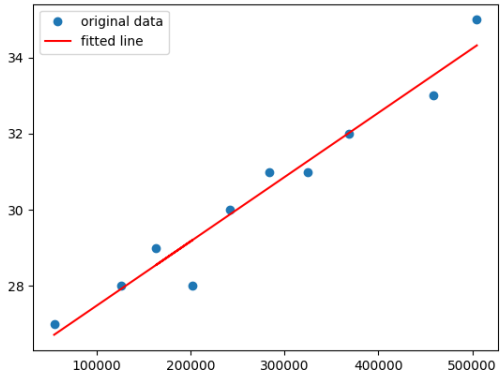
Manning College of Information and Computer Sciences, University of Massachusetts Amherst
Department of Mathematics and Statistics, University of Massachusetts Amherst



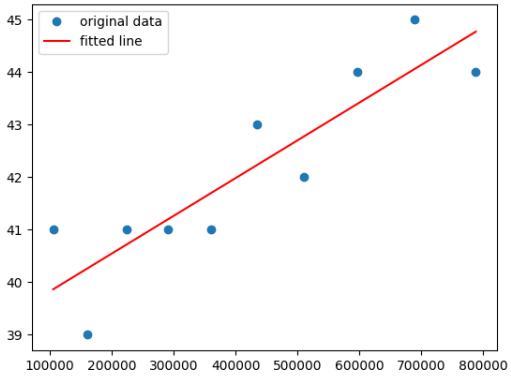
(A) 化学 – 斜率： 5×10^{-6}



(B) 数学 – 斜率： 5×10^{-6}



(C) 经济学 – 斜率： 17×10^{-6}



(D) 肿瘤学 – 斜率： 7×10^{-6}

FIGURE 7. $P(t)$ 与 $L(t)$ 之间的线性关系（中位数）。