



Research and demonstration on the rule of information aging

信息老化规律研究与实证

汇报人：贺国芳

PPT制作：郝丹瑜

资料搜集：赵红玉、张靖英

时间：2023.10.16



目录



- 01 信息老化的度量指标
- 02 信息老化的数学模型
- 03 信息老化规律的应用
- 04 CNKI关于老化的指标
- 05 网络信息老化实证

PARTONE

01 信息老化的度量指标



度量指标 <<<<<

半衰期

指某学科(领域)尚在利用的全部文献中较新的一半是在多长一段时间内发表的。

01

中值引文年龄

也称“最大引文年限”。
引文峰值年龄即指引文数量达到最大值时对应的时间。

02

普赖斯指数

表示在某一时间点上，被引用的文献相对于在该时间点之前发表的文献的平均被引用次数的比值。
 $Pr = \frac{\text{出版年限不超过5年的被引文献数量}}{\text{被引文献总数量}} \times 100\%$

03

剩余有益性

某年份某期刊被用户所利用的文献数被称为期刊有益性，经过若干年，期刊还保留的有益性为剩余有益性。

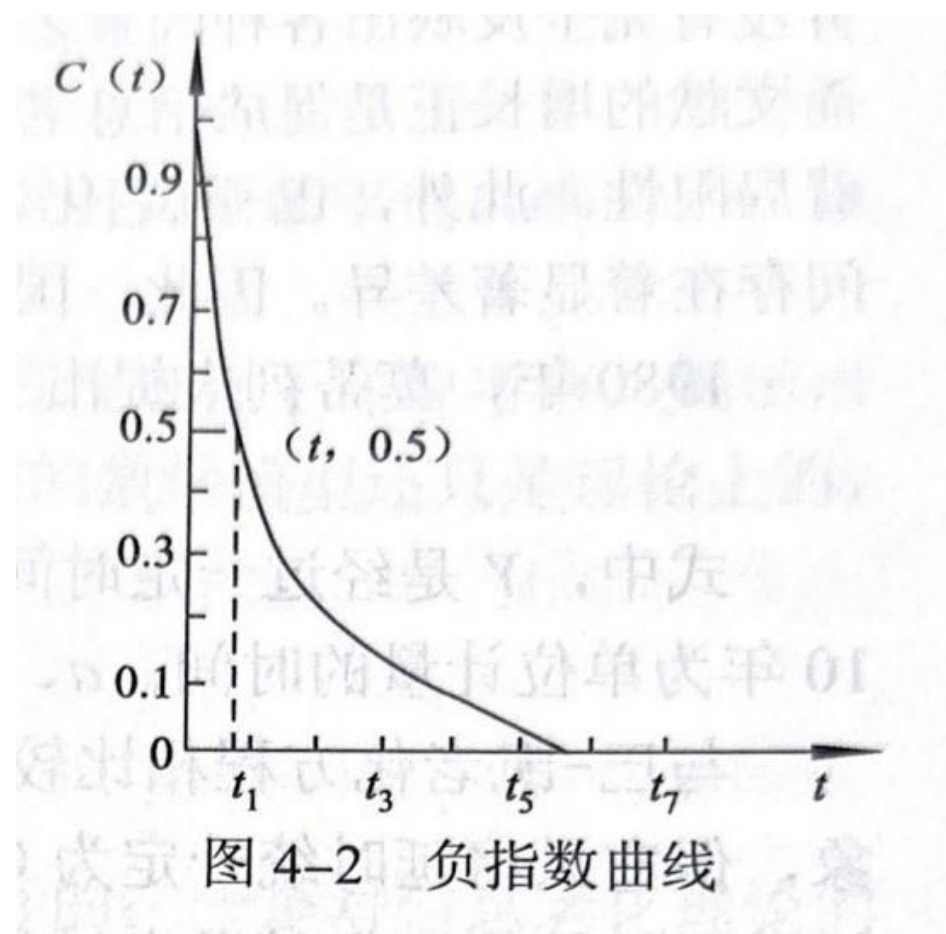
04

PARTTWO

02信息老化的数学模型



负指数模型 <<<<<



公式：

$$C(t) = Ke^{-at}$$

$C(t)$ 是引用文献中 t 年以前发表的相对数量

T 是文献的出版年龄

K 是常数，与学科性质有关

a 是文献老化率，与学科性质有关，且与半衰期有关

其基本观点是：把文献被利用的程度与文献“年龄”的关系是为一种负指数函数关系。负指数模型描述了文献老化规律，反映了**文献利用率的衰减现象**，基本符合实际观察结果。但是**未能将影响文献老化的因素与文献老化的关系直接反映出来**。

巴-凯老化方程 <<<<

巴尔顿和凯普勒选择了9个学科领域的期刊文献，进行了引文数据的统计分析，提出了**文献被引率与时间**的关系式：

$$Y = 1 - \left(\frac{a}{e^x} + \frac{b}{e^{2x}} \right)$$

Y是一定时间内的文献被引率

X是时间，以10年为单位

a,b是系数，满足 $a + b = 1$

根据半衰期定义，令 $y = 1/2$ ，这时x即为半衰期 X_H ， $X_H = 10(a + \sqrt{a^2 - 2a + 2})$

巴尔顿和凯普勒指出，一个学科的期刊文献是由多方面的内容组成的，很多相关学科都有可能使用。这样，半衰曲线也必然会受不同的内容的文献所支配。巴-凯老化方程只考虑到“老化”而**没有考虑到文献的增长**，同时文献出版年龄**以10年为单位也存在着局限性**。

PARTTHREE

03信息老化规律的应用



信息老化规律应用 <<<<<

有关信息老化的研究主要是围绕两个方面来进行的：一是对信息老化理论的研究，进而探索文献信息传播的动态规律；二是对信息拉票话的应用研究，以便更好地指导书刊选购、馆藏优化、排架流通等活动，更快地提高文献利用率和服务效率。

在科学学和科技史研究中的应用：

说明学科发展的速度，揭示科学发展的规律，反映出人类是如何继承和发展学科知识的。

在文献情报管理中的应用：

- 1.指导剔除老化的文献，优化馆藏、提高文献服务效率；
 - 2.为指定合理的文献工作原则提供依据；
 - 3.用于评价文献。
-

PARTFOUR

04CNKI数据库中关于 老化的指标



半衰期 <<<<<

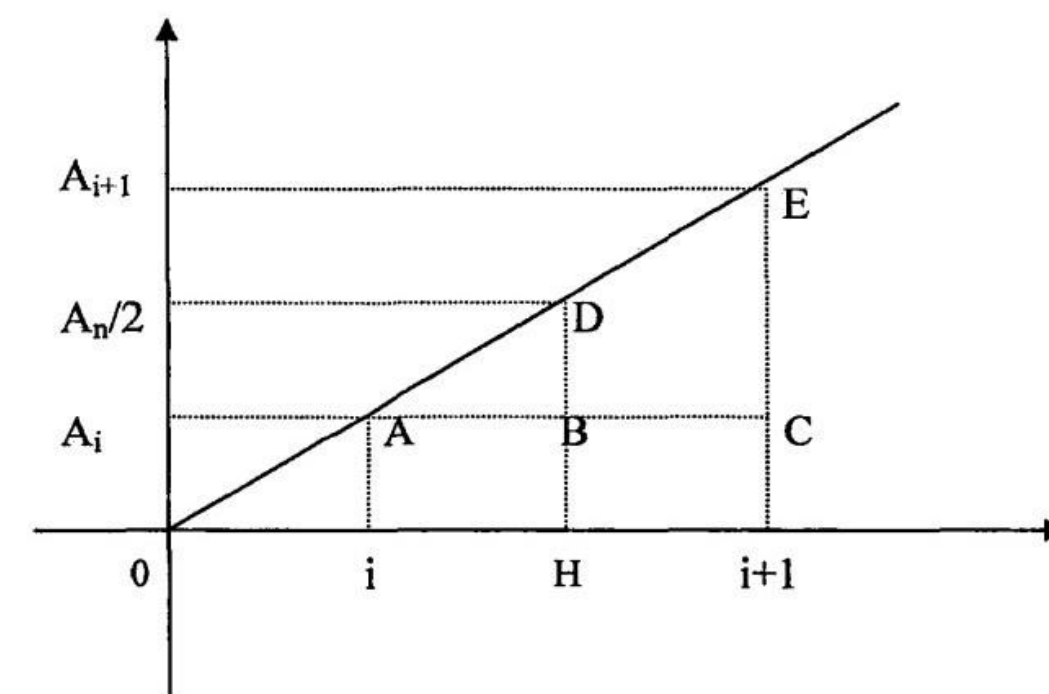
某学科领域现时尚在利用的全部信息资源中的一半是在多长一段时间内生产的。

计算方法：

Burton-Kebler方程：从**回溯**的角度出发，统计所选学科期刊当年发表文章的全部参考文献的出版时间，以**十年为一组**，将**积累百分比达到0.5时的时间**定义为半衰期，并给出引文累积百分比曲线数学模型 $y = 1 - (\frac{a}{e^x} + \frac{b}{e^{2x}})$ 。

作图法：基于引文频次累积百分比曲线，以累积百分比为纵坐标，时间为横坐标，直接在图中找出纵轴为时对应的时间。

中值法：与Burton做法类似，也是从引文的累积百分比着手。示意图如下



被引半衰期与引用半衰期 <<<<<

被引半衰期：指某一期刊论文在某年被引用的全部次数中，较新的一半被引论文发表的时间跨度。

以《Nature》2000年被引数据为例，半衰期为 $5 + (\frac{306184}{2} - 136096)/(156227 - 136096) \approx 5.84443^2$

表 3.1 《Nature》2000 年被引半衰期计算相关数据

引文年龄	引文年代	引文数	引文累积数	累积百分比 (%)
0	2000	5771	5771	1.88
1	1999	21013	26784	8.75
2	1998	30512	57296	18.71
3	1997	30066	87362	28.53
4	1996	24671	112033	36.59
5	1995	24063	136096	44.45
6	1994	20131	156227	51.02
7	1993	18258	174485	56.99
8	1992	15786	190271	62.14
9	1991	12378	202649	66.19
.....	103535	306184	100

表 3.2 《Nature》2000 年引证半衰期计算相关数据

引文年龄	引文年代	引文数	引文累积数	累积百分比 (%)
0	2000	1271	1271	4.75
1	1999	4174	5445	20.34
2	1998	3952	9397	35.10
3	1997	2982	12379	46.24
4	1996	2367	14746	55.08
5	1995	1810	16556	61.84
6	1994	1522	18078	67.53
7	1993	1228	19306	72.11
8	1992	943	20249	75.63
9	1991	901	21150	79.00
10	1990	709	21859	81.65
.....	1842-1989	4913	26772	100

引用半衰期：指该期刊引用的全部参考文献中，较新一半是在多长一段时间内发表的。

以《Nature》2000年引证数据为例,半衰期为 $3 + (\frac{26772}{2} - 12379)/(14746 - 12379) \approx 3.4254$

普赖斯指数 <<<<<

根据Price定义，普赖斯指数为当前被使用的文献中，近5年文献占总数的百分比。而在后人的研究中，具体以多少年为划分标准实则可以根据研究需要自行确定。如果近几年发表的文献的引文量占到了相当大的比重，则说明早期的文献被引量较小，从而推断出其价值的衰减。

表 4.8 《Nature》被引普赖斯指数表

年代	2000	2001	2002	2003	2004	2005
普赖斯指数	0.4445	0.4397	0.4457	0.4303	0.4156	0.4025
年代	2006	2007	2008	2009	2010	2011
普赖斯指数	0.3760	0.3563	0.3403	0.3311	0.3342	0.3204

《Nature》的普赖斯指数值总体上不超过50%，且呈下降趋势，数值范围在0.4457到0.3204之间。但《Nature》创刊于1869年，而在每一年，之前发表的全部文献的被引机会理论上是相等的，也就是说，自创刊起至5年以前这相当长一段时间内发表的大量文献所占的累积被引比重仅有60%到70%，可见早期文献的使用价值实际上相对较低。

其他指标 <<<<<

当年指标

定义：也叫快指标。指在指定年度里（2001年）引文中引及当年的引文数与总引文数之比。它揭示了研究者利用当年成果的敏锐程度。

计算方法：当年指标 = 目标年份内被引用的文献数 / 目标年份内所有文献的引用次数总和

剩余有益性

定义：某年份某期刊被用户所利用的文献数被称为期刊有益性，经过若干年，期刊还保留的有益性为剩余有益性。



引文峰值年龄

定义：指引文数量达到最大值时对应的时间。

计算方法：某刊在统计年得到的全部被引中，被引文献发表时间出现最多的那一年与统计年之间的时间差;引证分析中，引文峰值年龄为某刊在统计年的全部参考文献中，参考文献发表时间出现最多的那一年与统计年之间的时间差。

中值引文年龄

定义：指被引用的文献中较新一半的发表时间间隔。例如，某个领域中被引用的文献总数为N，其中N/2的被引用文献是在最近的n年中发表的，那么，这个领域的文献中值引文年龄为n年。

PARTFIVE

05网络信息老化现象实证



Twitter中的学术信息老化研究 <<<<<

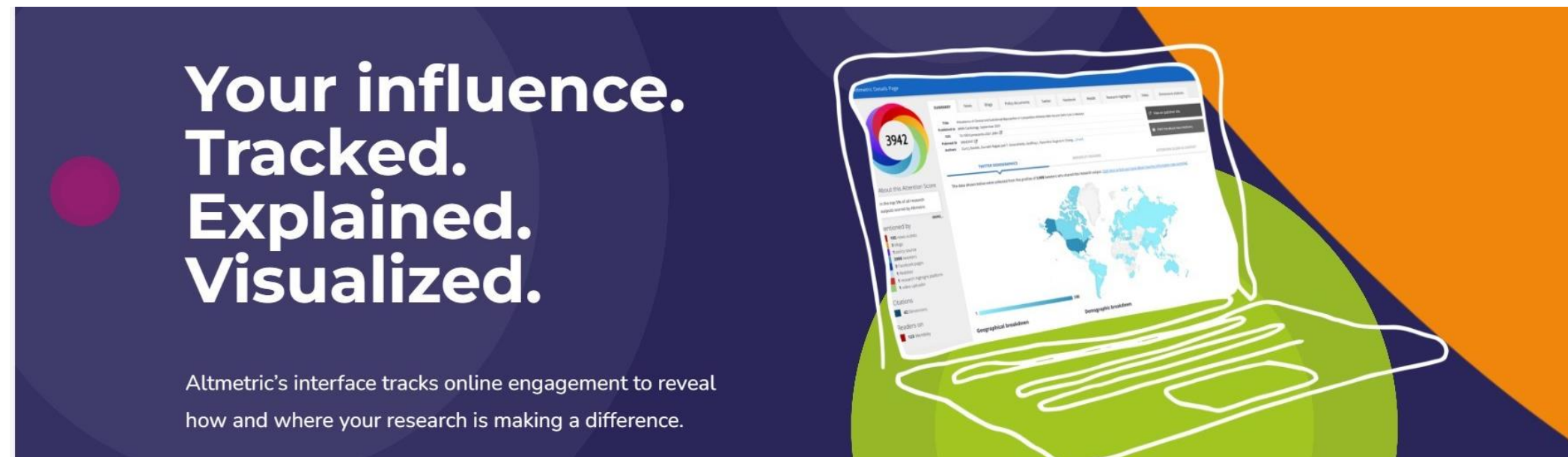
研究意义

在Web2.0时代，科学文献被搬上新媒体，学者通过社交网络分享和讨论科学知识。**网络的“零距离”与实时更新**则加剧了网络学术信息的老化。因此，开展基于社交网络的学术信息老化研究具有重要意义：理论上可以见微知著，**揭示网络学术信息的某些特征和规律，丰富网络信息计量学的研究内容**；实践上则为**预测科技热点、优化社交网络用户的信息需求**提供了科学依据。



Twitter中的学术信息老化研究 <<<<<

研究方法



社交媒体拥有Web2.0动态交互特性，科学家之间及科学家与公众之间可以直接进行交流，大大提高了科学的传播速度，并由此在学术圈得到普及。**Altmetrics**是Web2.0环境中的科学计量学研究，是建立在社交网络工具与开放存取分别在科学交流活动与科学成果出版平台中广泛应用的基础上而产生的，**为社交媒体信息老化研究提供了便利**。本文选取Altmetric.com平台上**2014-2017年每年得分最高的前100名科学文献**在Twitter平台上的转发记录为样本，针对不同学科对数据进行定量化分析，从**最大利用时限、半衰期、曲线拟合**等多角度透视学术信息在**社交网络中的老化特征和规律**。

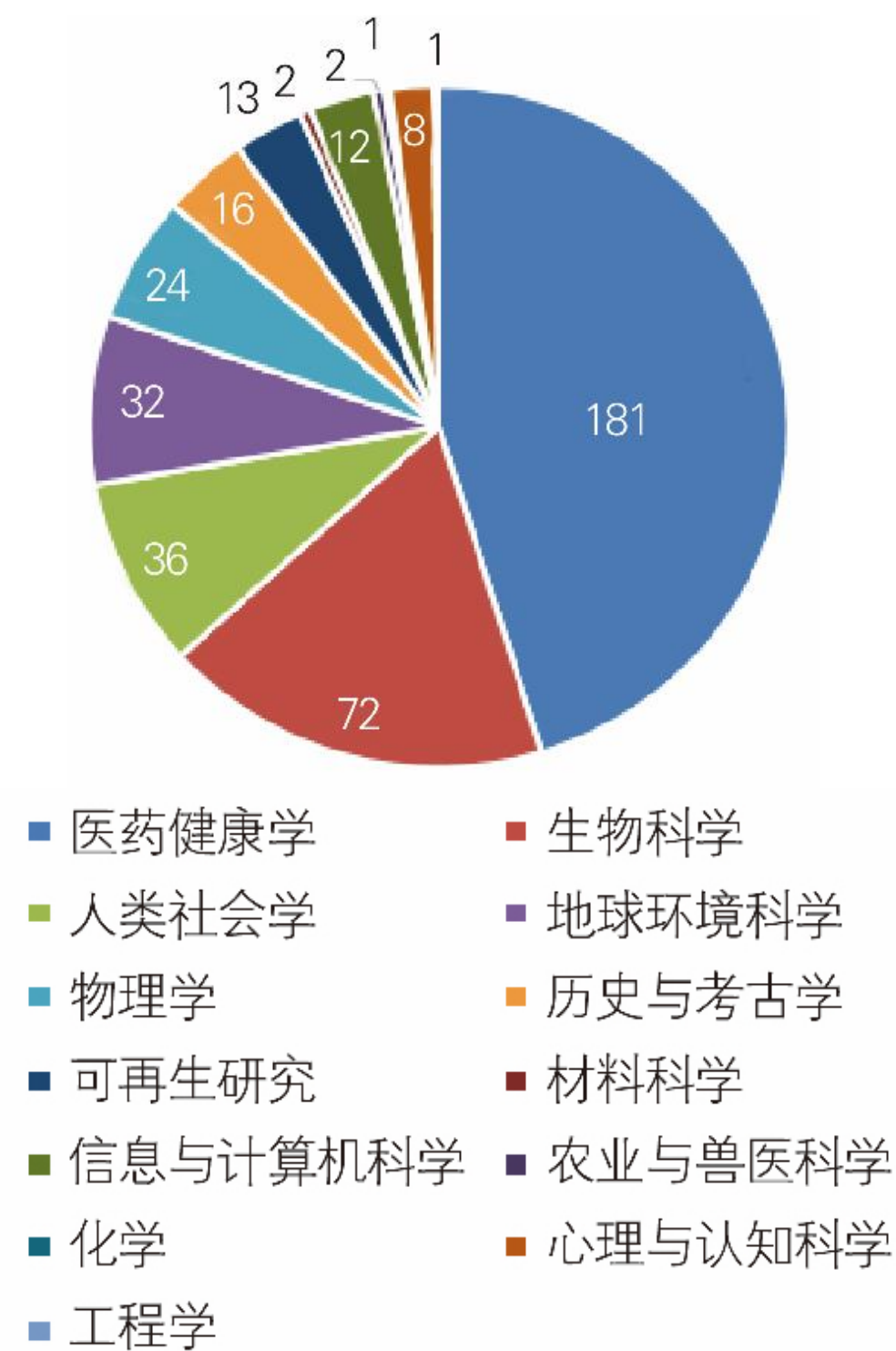
Twitter中的学术信息老化研究

<<<<<

数据来源

数据来源：数据来自Altmetric.com，Twitter是**学者使用最多，涵盖文献量最大**的社交媒体，以**2014-2017**年共4年每年**得分最高的top100科学文献**在Twitter上的转发记录为研究样本。

数据获取：先将top100文献按学科分类，共有13个学科，选择**文献量较大的4门学科**作为研究对象。随后利用Python语言爬取文献在Twitter上的转发日期，统计文献自公布于Twitter首日至数据采集日期间每日的转发次数。同一学科文献的**每日转发次数取平均值**，并以此数据**定量化分析不同学科的老化特征**，最后使用SPSS统计分析软件对**老化曲线拟合**，从而提出适用于社会网络的学术信息老化模型。



Twitter中的学术信息老化研究 <<<<<

老化特征分析

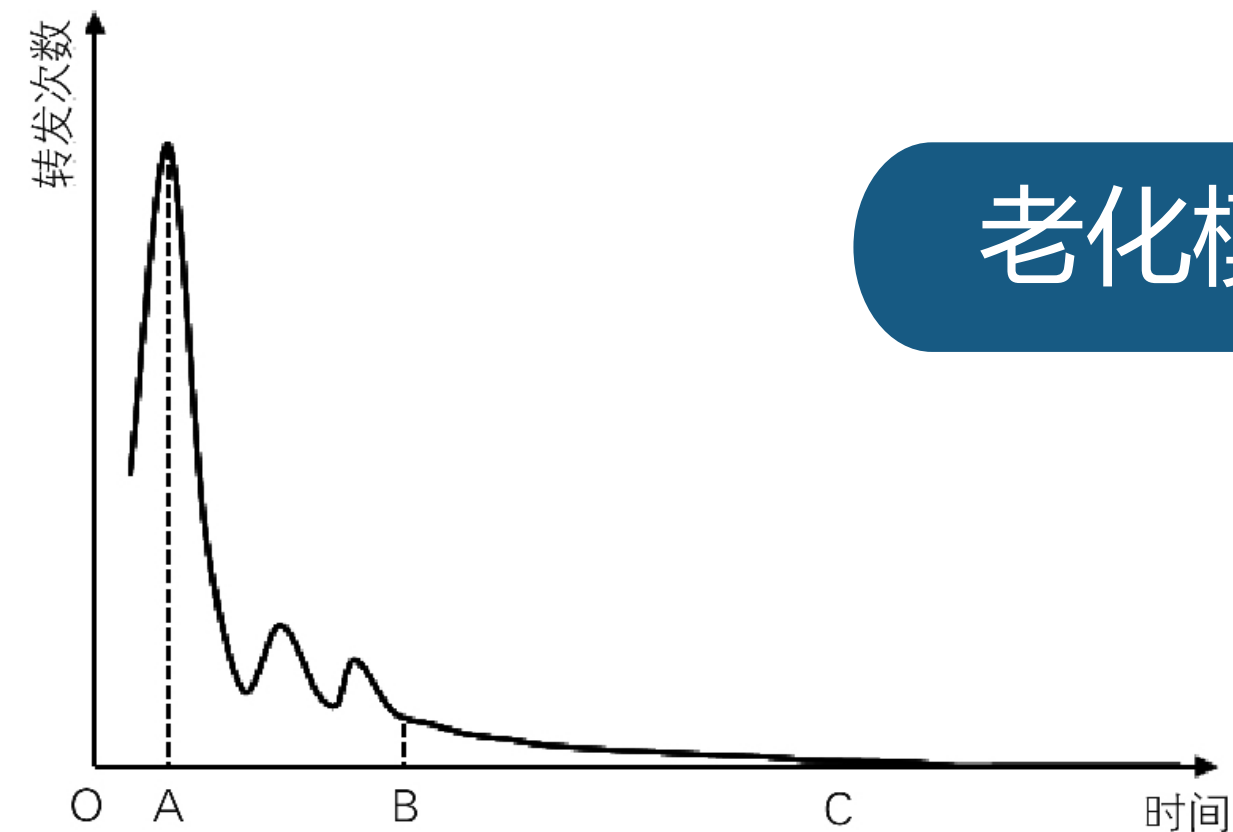
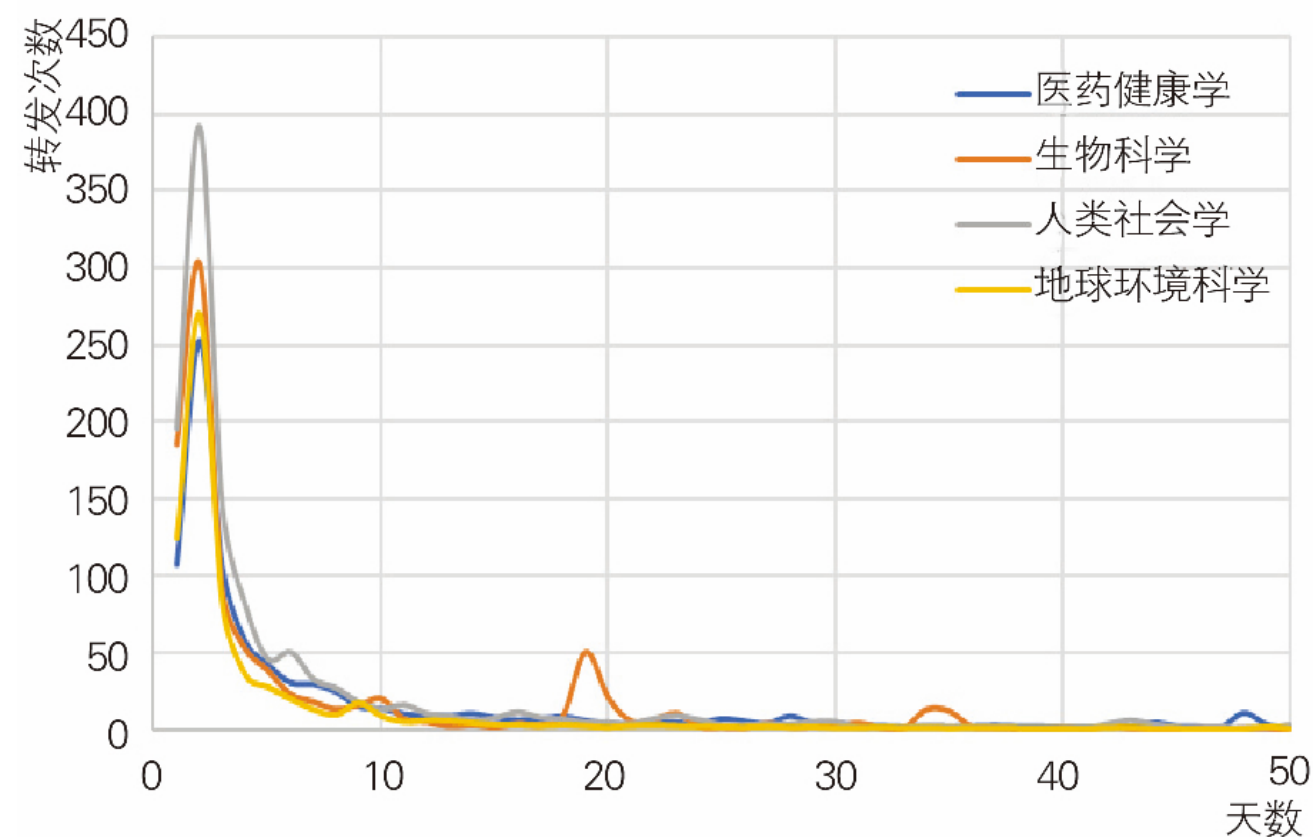
表 2 各学科的老化特征

学科	最大利用时限(天)	半衰期(天)
医药健康学	2	5
生物科学	2	3
人类社会学	2	3
地球与环境科学	2	2

不同学科的最大利用时限都是第2天，这说明基于社交网络传播的学术信息**通常在第2天的浏览量最大**，达到最大峰值。而**半衰期则依具体学科而定**，一般而言，应用技术学科的文献老化速度较快，基础理论学科的文献老化速度较慢；学科处于发展时期，老化速度较快，学科进入相对稳定期，老化速度相应变慢。

而社交网络的学术信息老化因素还需考虑到**大众群体对不同学科的兴趣度**，与生活息息相关的学科，热度会更持久(如医药健康学)，而对大众来说相对枯燥难懂的学科，则兴趣不大，相应老化得快些(如地球与环境科学)。

Twitter中的学术信息老化研究 <<<<<



以时间为横坐标，当日转发次数为纵坐标，绘制各学科的老化曲线图。根据曲线的变化过程，将老化曲线划分为**迅速递增期**、**震荡老化期**和**缓慢衰老期**3个阶段，分别对应图中的O~A段、A~B段、B~C段，其中O点表示社交网络学术信息的产生点，A点表示学术信息的利用价值达到最大值，B点表示学术信息价值经过震荡变化后的逐渐老化点，C点表示社交网络学术信息的消亡点。



Thank You for Your Watching

感谢您的观看

时间：2023.10.16

