

# “远读”——电子时代的中国文学基本图景

## 1. 背景

“远读”（distant reading）是对“近读”（close reading）的戏仿。德勒兹与加塔利提出了“树与块茎”的概念，这个理论认为启蒙以来的印刷时代的本文形态是“树”，意义是树的主根，作者是意义的中心；而电子时代的文本形态是“块茎”，文本是互相关联的，并非锚定在一个作者中。弗兰克·莫莱蒂又提出了“树与波浪”的文学图景，认为全球化以前的文学图景是“树”，是民族文化的独立生长；全球化后的文学图景是“波浪”，大量的文本在世界范围内互相影响、互相塑造。莫莱蒂认为，细读只能让我们以严谨的态度对待极少量的文本，而要想了解整体的世界文学体系，就需要我们放弃对单一文本的关注，而转向发现众多文本之间的联系，即“文本间性问题”<sup>1</sup>，勾勒对世界文学进行宏观分析的图景。在世界文学的研究中，他认为有必要牺牲单个文本的细节赏析，而聚焦比文本小得多或大得多的单位，使用定量分析方法从中发现规律，从而获取一种“鸟瞰式的宏观视野”<sup>2 3</sup>。

因此，远读不是文本精读、逐字逐句地阅读，而是运用数字人文、计算批评的方法，勾勒出文学世界的基本图景。对于中国文学来说，我们不乏纤毫毕现的文本分析（如清代的文字、训诂，民国的“整理国故”等），我们缺乏的是把握中国文学世界的视野和能力。因此，在这里，人文学者以躬身其中的操作，通过解释工具和“通过工具思考”，发展出了更全面的文化批评。这种文化批评的着力点，并不止于目的和对象的政治关切，更在于它能够遵从并运用数字人文自身的形式语言和操作规则，来完成关于工具的反思。<sup>4</sup>

中国文学 bot 是微博拥有 93.6 万粉丝的博主，有 1164.4 万的转评赞，是微博众多文学 bot 里数据最好的一位。它日常会发布粉丝投稿的中国文学作品中的节选文本，一般在 10-250 字之间，有持续的热度。通过分析它对投稿文本的选择和微博用户对不同文本的态度反应，对了解电子时代的中国文学基本图景有一定帮助。

<sup>1</sup> 高树博：“弗兰克·莫莱蒂对‘细读’的批判”，《学术论坛》4（2015）：99-104。

<sup>2</sup> Jockers, Matthew. Macroanalysis: Digital Methods and Literary History. Urbana-Champaign: University of Illinois Press, 2013.

<sup>3</sup> 向帆,何依朗.“远读”的原意:基于《远读》的引文和原文的观察[J].图书馆论坛,2018,38(11):44-48+43.

<sup>4</sup> 赵薇.作为计算批评的数字人文[J].中国文学批评,2022(02):157-166+192.



## 2.数据处理

首先用 python 程序爬取“中国文学 bot”2023 年 11 月 20 日到 2024 年 1 月 20 日两月间所发布的 254 条博文，再将所得数据清洗，分为内容、时间、点赞数、转发数、评论数、文本长度、作者六个维度，最后制作饼图、符号条形图、蝴蝶图、双向柱图、词云以及相关图进行可视化呈现，帮助读者理解数据。

遗憾的是作者本意想爬取近一年的数据量，但没有绕过微博的反爬虫机制，后续索性取最近两月的数据进行更加细致的分析。

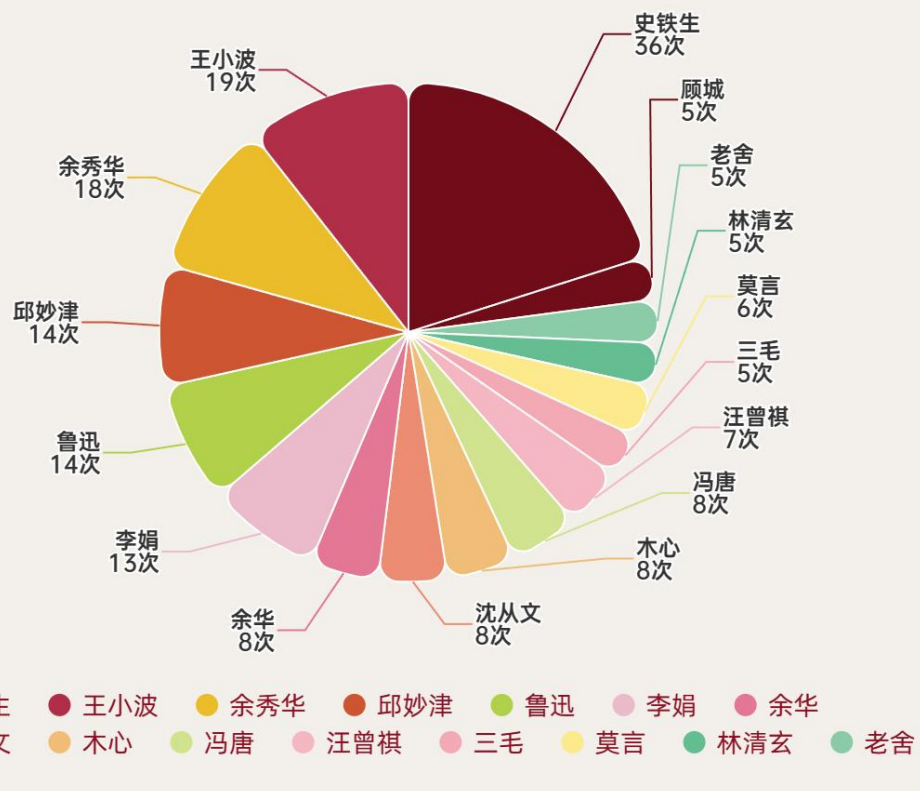
## 3.可视化分析

### 3.1 文本作者分析

毫无疑问被引用次数是与 bot 运营者和受众的品味相关的。

254 条博文，由 60 位作者的作品节选组成，其中史铁生的作品被选取最多，达 36 次。饼图仅统计被引用达 5 次以上的作者，有 44 位被引用次数未达 5 次的作者并未计入饼图。其中被引用 1 次的有 27 位，被引用 2 次的有 9 位，被引用 3 次的有 2 位，分别为巴金、林语堂，被引用 4 次的有 6 位，分别是陈春成、海子、季羡林、钱钟书、杨绛、周国平。

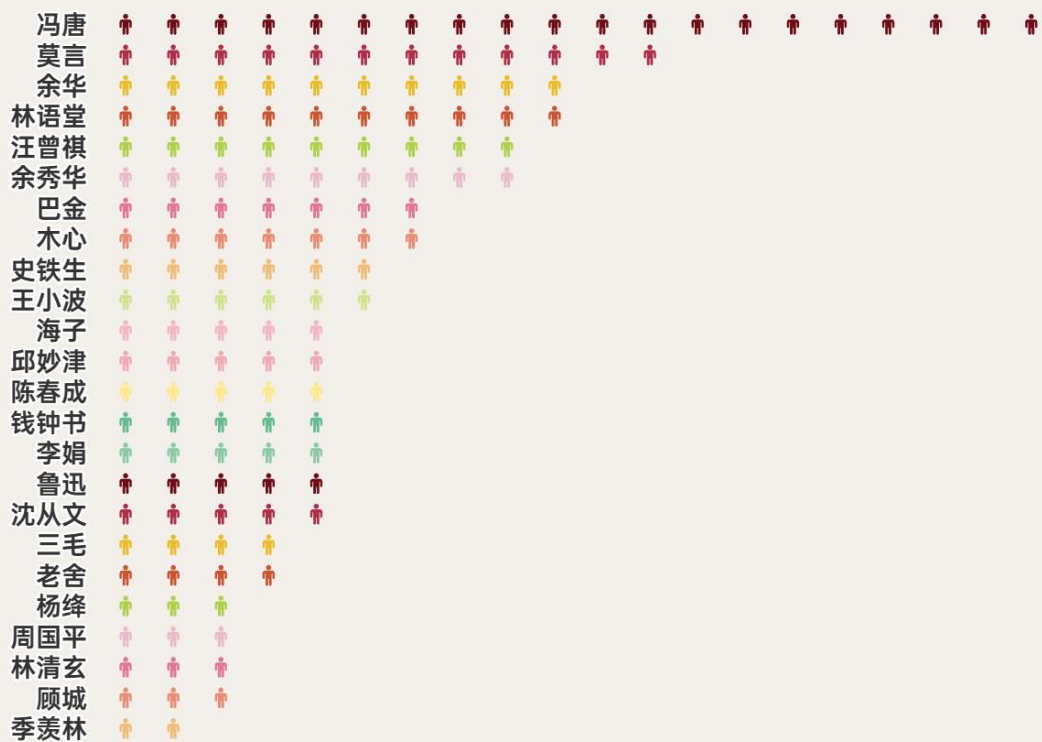
## 11.20-1.20中国文学bot引用各作者次数



为了更清楚地看出微博用户对中国文学作者对应文本的喜爱程度，用该作者所有被引用的文本获得的点赞、转发和评论数量之和，除以作者被引用次数，来计算受用户喜爱程度。

最后用如下符号条形图来表示。

## 11.20-1.20中国文学bot文本作者受用户喜爱度

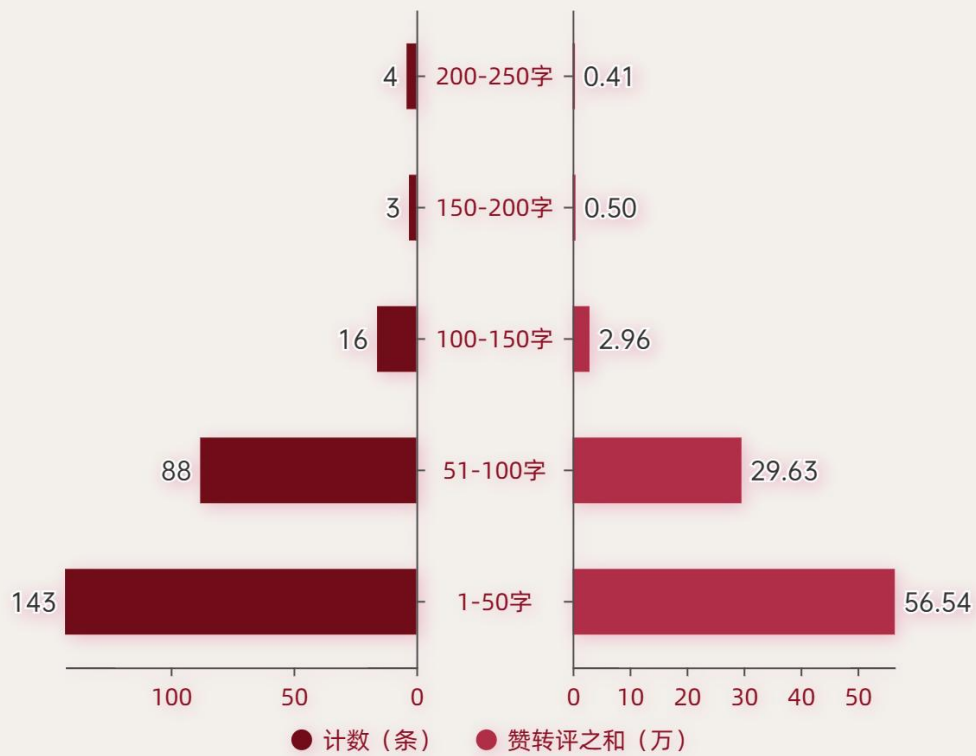


用户喜爱度计算方式：作者的文本所获转评赞之和/被引用次数

### 3.2 文本篇幅分析

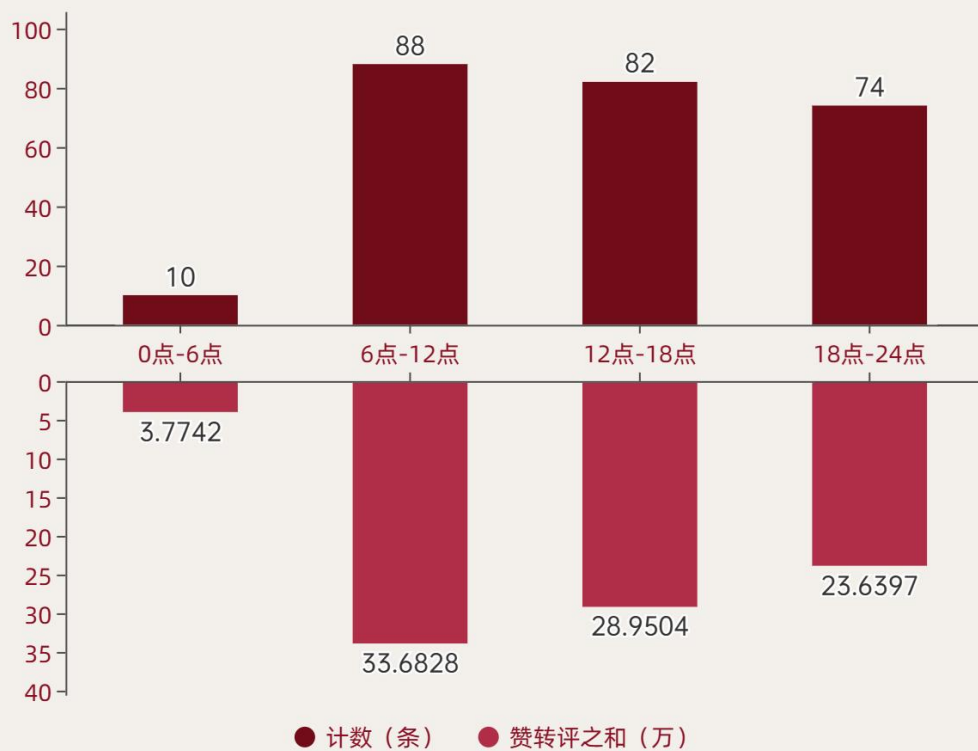
社交平台传播对文本的需求是重内容、轻形式、篇幅短，以“严肃文学”，“经典”标榜的中国文学 bot 也不例外地去迎合这样的趋势，它近两月所发布内容中 100 字以内的篇幅占 90.94%。

不同文本长度的被引用次数和转评赞之和



### 3.3 文本发布时间段分析

一天中不同时间段的文本计数和所获转评赞之和





### 3.4 文本词频分析

在社交媒体中，具有情绪导向，感染力强的文字更有传播力度。从词云中可以看出，爱仍是最能引起读者共鸣的一个字。爱、生活、爱情、希望仍有打动人心力量。

其中所用停用词为中文停用词表 stopwords.txt。



### 3.5 相关度分析

相关系数定量地刻画了 X 和 Y 的相关程度，即相关系数越大，相关程度越大；对应相关系数为 0 时，相关程度最低。

相关系数 $|\rho(X, Y)|$ 的划分：

0.8-1.0 极强相关

0.6-0.8 强相关

0.4-0.6 中等程度相关

0.2-0.4 弱相关

0.0-0.2 极弱相关或无相关

所取五个维度：点赞、转发、评论、赞转评合计、文本长度

所得相关系数矩阵如下：

[	[	1.	0.88519401	0.8538666	0.99943605	-0.14811555]
[	0.88519401	1.	0.65057541	0.8995089	-0.1204377	]
[	0.8538666	0.65057541	1.	0.85117625	-0.05672995]	
[	0.99943605	0.8995089	0.85117625	1.	-0.14559761]	
[	-0.14811555	-0.1204377	-0.05672995	-0.14559761	1.	]

分析可得，点赞和转发、评论都是强烈正相关，互相具有相关性，而评论和

转发的相关度只有 0.65，说明评论的人不一定转发，反之亦然，而点赞、转发、评论和文本长度都是极弱相关或者无相关。

