**“那年杏花微雨，你说你是……？”**

**——从人物角色分析看《甄嬛传》的出圈原因**

**第6组：曹毓文 陈思祺 石婧丹 宋文轩**

****

**目录**

[一、背景介绍 1](#_Toc2900)

[二、研究思路 1](#_Toc25027)

[三、数据来源 2](#_Toc11456)

[四、《甄嬛传》的出圈原因分析 3](#_Toc21017)

[（一）评论词云图分析 3](#_Toc15246)

[（二）评论LDA主题模型 3](#_Toc28002)

[（三）评论星级分析 4](#_Toc23869)

[五、数解各宫娘娘——人物分析 6](#_Toc19930)

[（一）剧中人物形象分析 6](#_Toc2794)

[1. 人物形象词云图分析 6](#_Toc30895)

[2. 人物出场情况分析 7](#_Toc2927)

[3. 人物亲密度矩阵分析 9](#_Toc7733)

[（二）观众对人物评价分析 1](#_Toc10049)0

[1. 人物评论数量分析 1](#_Toc8369)0

[2. 人物评论情感分析 10](#_Toc11411)

[六、研究结论 1](#_Toc16008)2

**一、背景介绍**

《甄嬛传》是我国的一部现象级大型电视剧，开播至今11年一直传奇般地维持着非常高的讨论度，也被不少观众反复观看。在豆瓣上，超过45万人对《甄嬛传》打出了稳定的9.3分的高分；而在网络上，《甄嬛传》也能奇妙地融合进各种新鲜的时事热点，为人们所津津乐道，如近期剧中“宝鹃，我的嗓子”、“剪秋，本宫的头好痛”等台词也被网友用来形容“阳过”的症状，再度冲上热搜；同时，有爱好者们甚至开发了“甄学”，对剧情和任务进行了逐帧解读，乐此不疲。

在当下流量当道而越来越多的国产剧却乏善可陈之时，《甄嬛传》却能经久不衰，其持续热度背后的原因也常常被讨论与分析。本文希望能从统计的角度，结合剧情和观众评价等相关数据，对该剧的火热现象进行剖析，并聚焦剧中塑造的精彩人物群像，从人物形象的角度出发进行深度分析。由此，希望能将《甄嬛传》作为国产剧的优秀范例，对其成功总结经验，由此对我国影视剧行业的发展有一定的指导意义。

**二、研究思路**

我们将通过对这部电视剧剧本、评论播放量等指标分析《甄嬛传》在十年多的时间内持续出圈的原因，以问题作为导向，得出具体的结论。具体研究思路如图1。

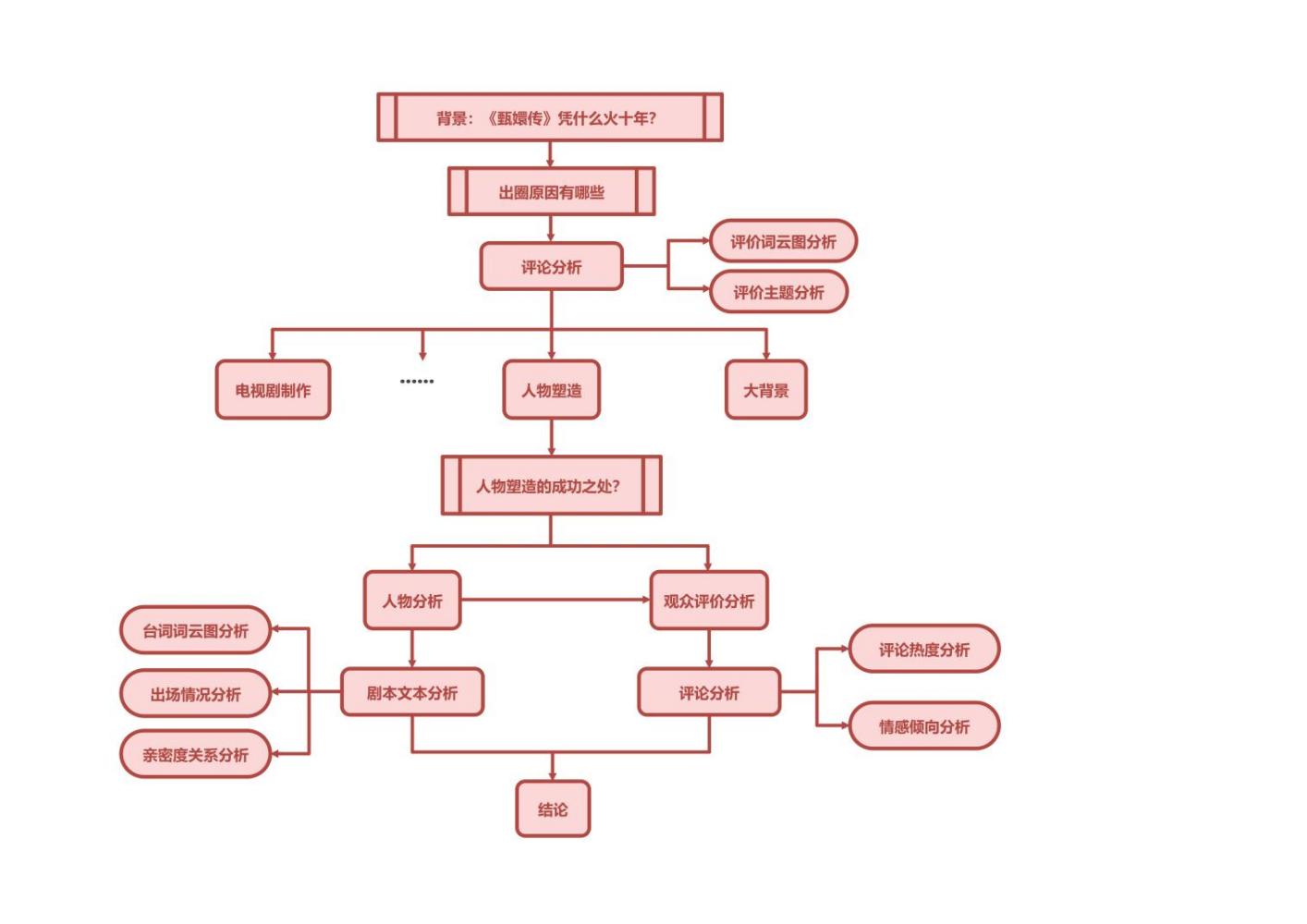


图1 研究思路

首先，探究观众对《甄嬛传》的总体态度。通过对豆瓣上《甄嬛传》的影评和短评进行综合分析，基于丰富的评论文本资料，构建了词云图；在此基础上使用LDA主题模型和基于经验的主题数设定，对各个主题的含义加以归纳，提炼出3个评论的主题；对评论星级进行分析并以关键词为定位原评论，了解观众对《甄嬛传》的总体评价态度和为什么给予好评或差评。

其次，对人物形象进行分析，探究《甄嬛传》的人物塑造。通过对剧本进行文本分析，绘制人物形象词云图、人物亲密度矩阵，分析人物出场频次情况，从而对人物的性格、说话作风、故事经历，与其他人物之间的互动关系形成较为全面的汇总。为了解观众对人物塑造的看法，我们再次对评论进行处理分析，以分割后精确到人物角色的短句为分析材料，比较主要人物出场次数，对评论进行情感分析，深层次地挖掘人物塑造给观众带来的体验。

最后，通过以上数据分析手段，得出《甄嬛传》在播出十年后依然长盛不衰、不断出圈的原因。

**三、数据来源**

我们的数据来源有三个：《甄嬛传》电视剧剧本，豆瓣平台《甄嬛传》评论和Youtube平台《甄嬛传》的集数及播放量。

数据的获取常常在数据分析过程中起到关键作用，高质量的数据才能得到有效的结论。为真实反映电视剧的情节和人物特色，我们选择《甄嬛传》电视剧剧本这一第一手资料来分析剧中人物形象；在观众反馈部分，考虑到豆瓣平台在国内影评领域具有较高地位，评论数量多，最具代表性，而Youtube作为全球最受欢迎的流媒体播放平台，其受众范围广，我们选择这两个平台作为观众反馈的数据来源。为了保证数据的完整性和可靠性，我们爬取了这两个平台关于《甄嬛传》的全部相关数据。下面的表格是对数据类型的具体说明。

表1 数据来源

|  |  |  |  |
| --- | --- | --- | --- |
| 来源 | 变量名 | 说明 | 数据示例/取值范围 |
| 《甄嬛传》电视剧剧本 | 幕数 | 剧本中显示的幕数 | 1-1265（若出现续幕则幕数加1） |
| 剧本内容 | 剧本中背景、人物讲话、动作或旁白 | “百官（叩首）：谢主隆恩！” |
| 豆瓣平台《甄嬛传》评论 | 内容 | 评论具体内容 | “不晓得别人怎么看这部戏,不过对我而言,这真是近年难得一见的好剧……” |
| 星级 | 评论人对《甄嬛传》做出的星级评价 | 1-5 |
| Youtube平台《甄嬛传》 | 集数 | 《甄嬛传》电视剧的集数 | 1-76 |
| 观看量 | 电视剧每一集在YouTube上的播放量 | 1-9999999 |

**四、《甄嬛传》的出圈原因分析**

**（一）评论词云图分析**

我们从豆瓣网站中的“后宫甄嬛传”讨论部分爬取了601条短评和1220条长剧评，做了分词和停用词处理，提取出高频词绘制词云图，如图2所示。

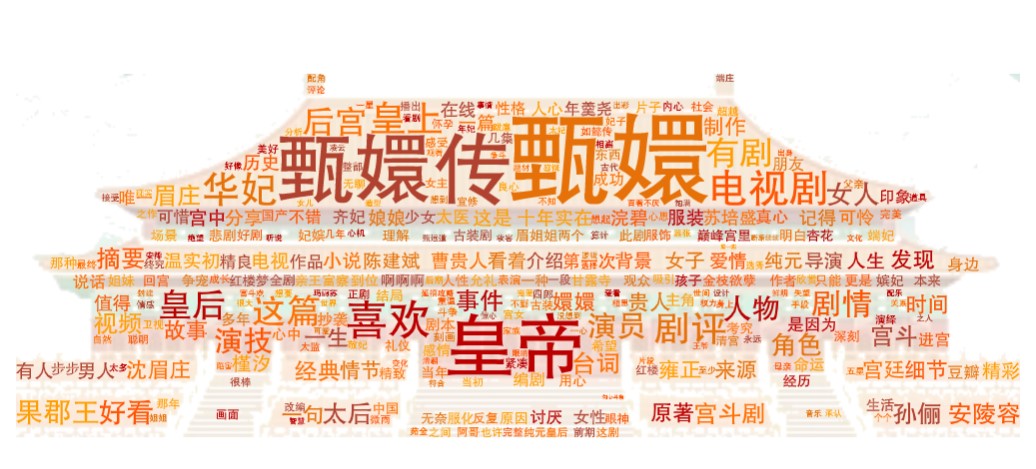


图2 评论词云图

从词云图可以看出，“甄嬛”、“皇帝”、“华妃”、“皇后”、“沈眉庄”等主要人物赫然在列，并由此上升到对演员的讨论，如“孙俪”、“陈建斌”两位饰演甄嬛和皇帝的演员频繁出现，而“女子”、“爱情”、“剧情”、“宫斗剧”等词多次提及也说明了《甄嬛传》的宫斗题材与剧情走向为人所津津乐道，而“演技”、“制作”、“服装”等与电视剧制作的关键词也被多次提及，可见电视剧制作也是观众的关注点。

**（二）评论LDA主题模型**

为了进一步提炼出观众评论的几类主题，我们对爬取的评论建立了LDA（Latent Dirichlet Allocation）主题模型。主题模型是近年来文本挖掘中较为流行的建模方法，常用于从文本中提取隐含主题，从而对文本内容进行总结和概括。这里，我们提取出3个主题，并根据每个主题中出现概率较高的词对各个主题的含义加以归纳，分别是：电视剧制作、大背景、角色演绎。

表2 评论LDA主题

|  |  |
| --- | --- |
| 主题 | 高频词 |
| 电视剧制作 | 喜欢、演员、好看、小说、视频、台词、不错、故事 |
| 大背景 | 皇帝、皇上、皇后、后宫、女人、宫廷、雍正 |
| 角色演绎 | 甄嬛、华妃、角色、孙俪、果郡王 |

**（三）评论星级分析**

在601条短评中，对甄嬛传的评价有褒有贬，我们绘制了评价星级的柱状图，如图3所示，其中4星和5星好评占81.1%，而1星和2星的差评仅占7.6%。

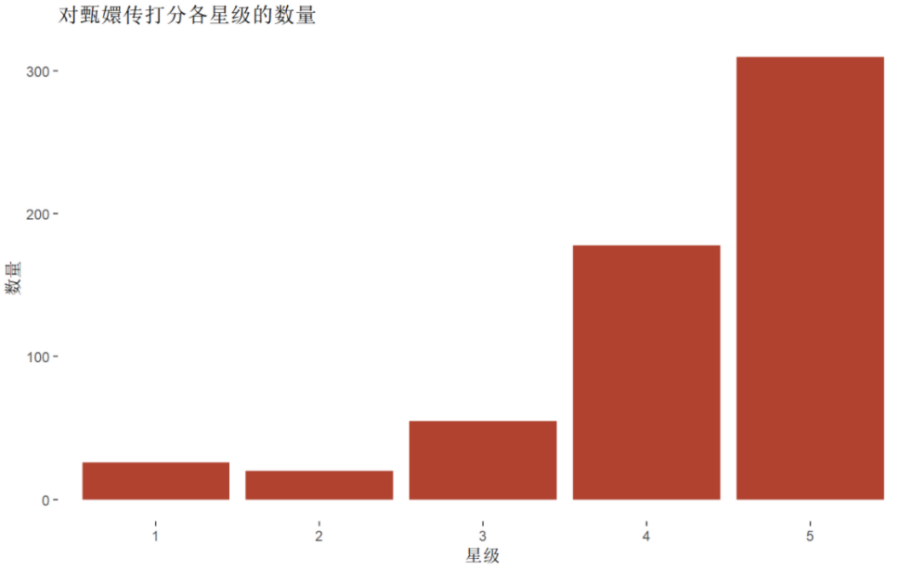


图3 评论星级柱状图

为继续观察甄嬛传的好评与差评，我们将601条短评中星级为1和2的作为差评数据，将星级为4和5的作为好评数据，分别提取10个高频词，绘制出柱状图如图4和图5所示。

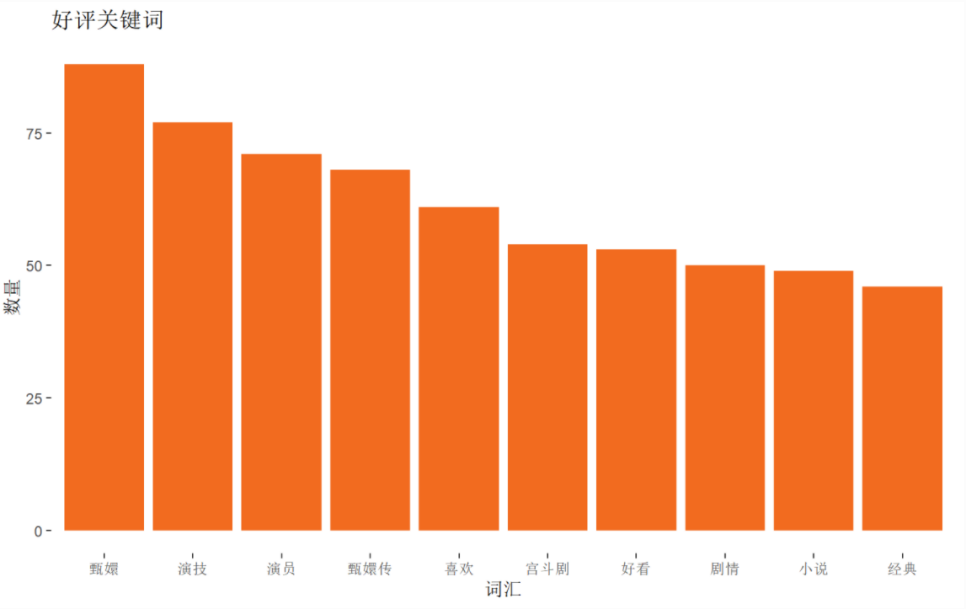


图4 好评关键词柱状图

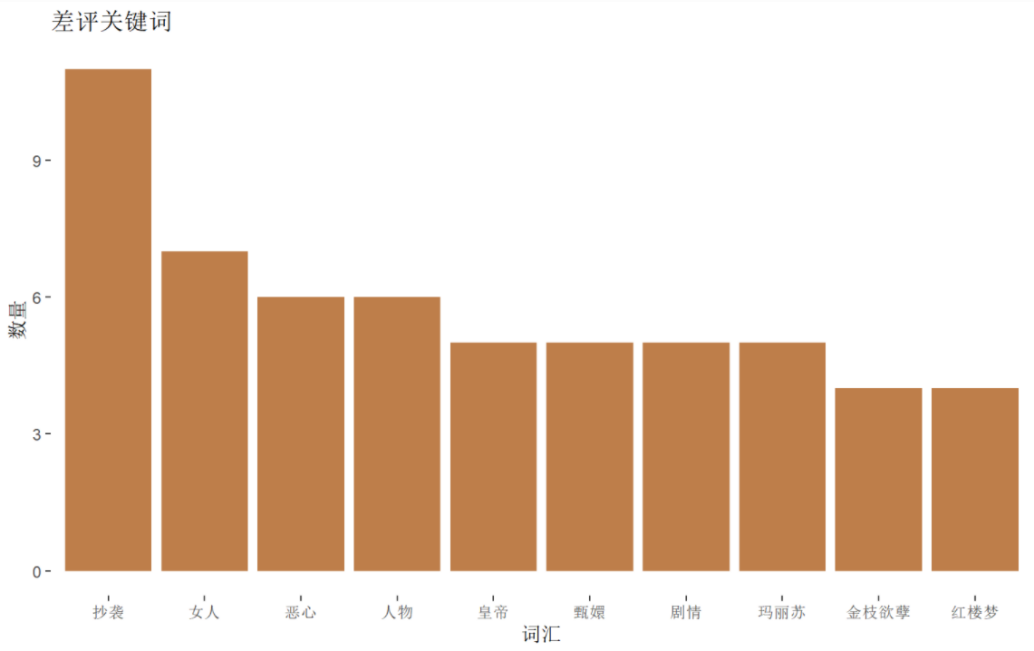


图5 差评关键词柱状图

在好评关键词中，“好看”、“喜欢”这类正向评价词汇较多，“演技”、“演员”、“剧情”等频繁出现体现了观众对于《甄嬛传》在电视剧制作方面有较高的评价。

*“近年来看过最好看的剧，服化道，演员演技，剧情等各方面都是一流的，说是经典也不为过。”*

*“真心喜欢这剧，制作精良，场景华丽，演员表演到位。”*

*“看过最好的一部宫斗剧！剧中人物饱满，个性鲜明，剧情紧凑。”*

在差评关键词中，“抄袭”、“玛丽苏”等词的提及体现出观众对于《甄嬛传》原小说的态度。通过关键词定位原评论，我们发现一些差评中在提及了原小说的抄袭事件的同时也肯定了甄嬛传电视剧制作方面的精良。

“*人设套路，计谋套路，抛却演技和精美的服化道，内里依然是玛丽苏式意淫。哪一天宫斗剧里的女人幡然醒悟联手做掉皇帝，而不是为了一个油腻薄情的男子抛弃自我断送一生，才是真的反封建。”*

*“剧其他的都不错——班底、服饰、演技外加改编，但架不住原著是个傻逼抄袭小说。给一星望周知甄嬛传抄袭。”*

*“甄嬛传，播再多遍，也只是个玛丽苏剧。到处是对红楼梦拙劣的模仿和盗用。”*

**五、数解各宫娘娘——人物分析**

**（一）剧中人物形象分析**

**1. 人物形象词云图分析**

首先，我们在剧本中分别提取八个女性角色的台词，去除名字后进行分词并统计词频，使用TF-IDF算法计算得分后均乘以10万作为新“词频”，并用WordArt绘制词云图。借助词云图，通过其中的人物可以进行关系和情感倾向的分析；通过其中的口头禅可以看出人物性格和说话作风；通过其中的事物、特征可以大致看出人物的故事经历。以华妃和宁嫔的词云图为例，如图6和图7所示。

图6 华妃词云图 图7 宁嫔词云图

以下，我们将人物主要分为正反两大阵营进行简要解读。

**（1）正面：“嬛嬛战队”**

嬛嬛战队中包括甄嬛、沈眉庄和敬妃三个人物。词云图分别如图8、图9和图10所示。

图8 甄嬛词云图 图9 沈眉庄词云图 图10 敬妃词云图

在嬛嬛战队中的三位正面人物我们可以看到，大部分关键词都是他们所关心和在意的人和事，这也符合他们积极正面的人物形象。以敬妃为例，额娘一词可以看出她最终母女情，妹妹、真乖、月儿等等都和她温婉不与人相争的性子相符。

**（2）反面：“安琪拉打胎大队”**

“安琪拉打胎大队”包括皇后、祺嫔和安陵容三个人物。词云图分别如图11、图12和图13所示。

图11 皇后词云图 图12 祺嫔词云图 图13 安陵容词云图

该阵营中主要是反面角色，观察词云图时，映入眼帘的关键词基本都与算计有关，这通过一胎、厉鬼、贱人、麝香等词就暴露出来。比如皇后身上出现的华妃、莞贵人、莞嫔等词都可以看出她的心中装的都是眼中的对手。

**2. 人物出场情况分析**

**（1）关键人物出场幕数密度**

从时间维度看，我们选取了四个出场时间比较有代表性的角色，分别是甄嬛、华妃、宁嫔和祺嫔，绘制了关键人物出场幕数密度图，如图14所示。

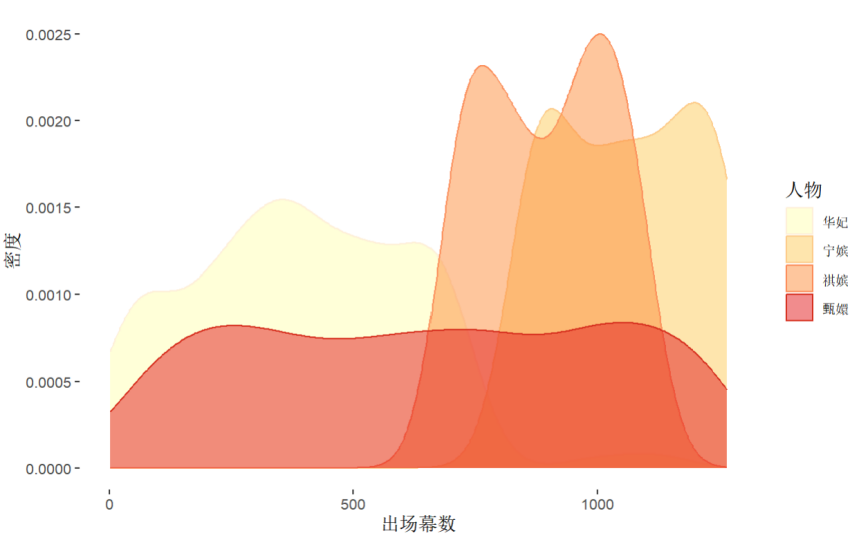


图14 关键人物出场幕数密度图

从图中可以看出，人物的高光时刻分散在剧中不同部分。其中甄嬛作为剧中女一号，她的出现贯穿整部剧；华妃主要在剧中的前半部分出现，而宁嫔和祺嫔主要在剧的后半部分出现。这也帮助推动了不同阶段剧，充实了剧情的完整性。

同时，我们对从YouTube上爬取的《甄嬛传》剧中各集的播放量绘制了折线图，如图15所示。

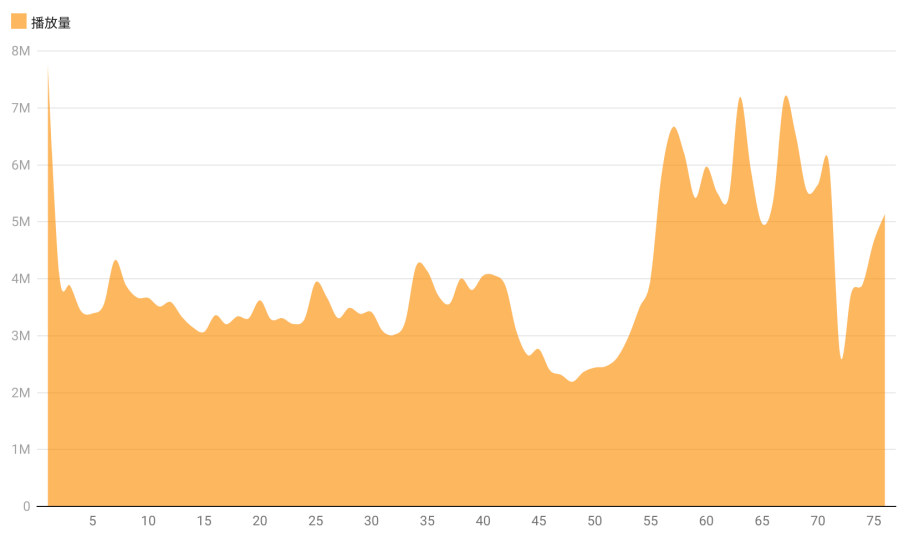


图15 全剧各集播放量折线图

结合全剧各级的播放量看，除了50集左右是主角甄嬛被废除位份出宫修行的低谷剧情播放量较低，长达76集的剧能维持相对稳定和蓬勃的观看量也和不同人物高光时刻有关。在剧的后半部分播放量明显高于前半部分，可以推测丰富的人物形象与更多的主要角色的出场起到了对剧情的推动的作用，进而吸引观众。

**（2）主要人物出场幕数**

我们统计了主要人物的出场幕数，来观察哪宫娘娘的戏最多。主要人物出场幕数柱状图如图16所示。

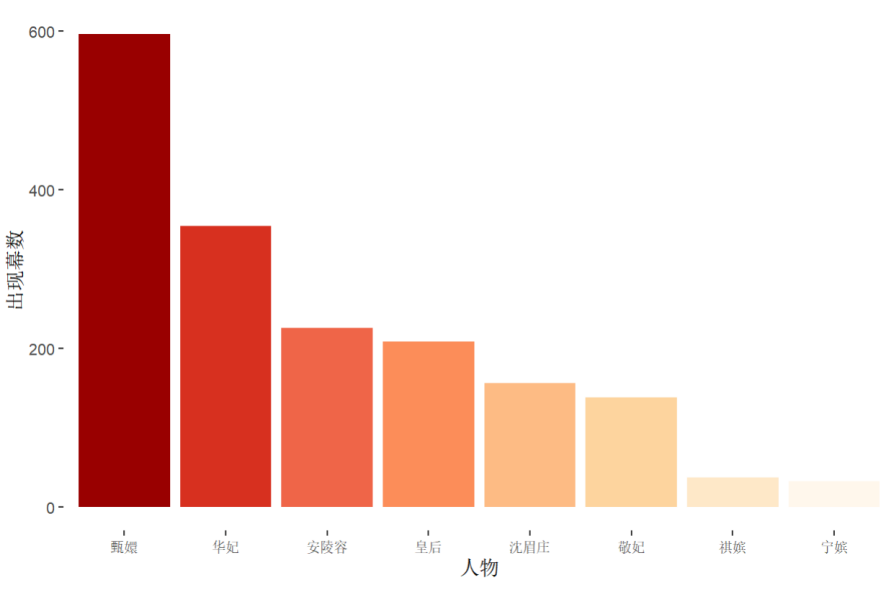


图16 主要人物出场幕数柱状图

可以看出，人物塑造篇幅上有详有略，重点刻画了甄嬛、华妃、安陵容等更多面、心路历程更丰富的人物。同时我们也产生了一些疑问：戏多的娘娘就是最受圣恩的吗？戏多的娘娘就是最受观众喜爱的吗？我们在接下来两个部分对这两个问题进行一一解答。

**3. 人物亲密度矩阵分析**

我们通过统计两个角色在同一幕中出场的幕数，绘制了主要人物亲密度矩阵，如图17所示。

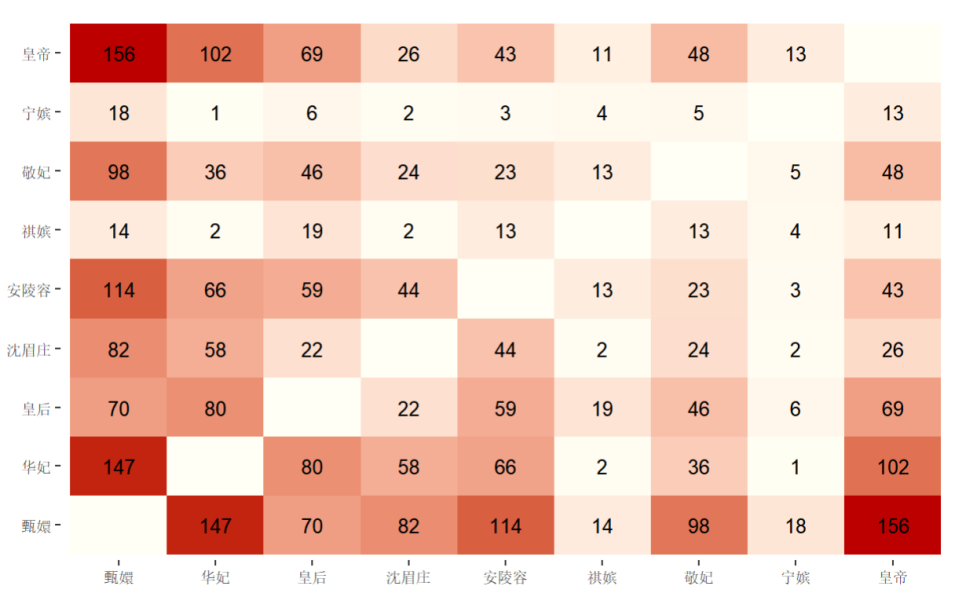


图17 主要人物亲密度矩阵

可以看出，大部分角色与甄嬛的接触最为频繁，这也是主角光环的必然结果，与出场情况相对应。再来看看，作为全剧最大boss的皇后和大女主的甄嬛，两人对手戏最多的女人居然不是彼此，而都是只上线了四十集的华妃，这也许说明华妃娘娘最难对付，这些有趣的现象都值得我们探究。再看看皇上的情况，值得注意的是皇上与敬妃的对手戏多于宠妃安陵容，这也许可以说明受宠的妃子并不一定是相处最多的妃子。

**（二）观众对人物评价分析**

**1. 人物评论数量分析**

我们将601条短评和1220条长剧评按照句号、空格、逗号分割，用主要女性人物的名字作为筛选词进行筛选，先绘制了提及这些人物频率的柱状图如图18所示。

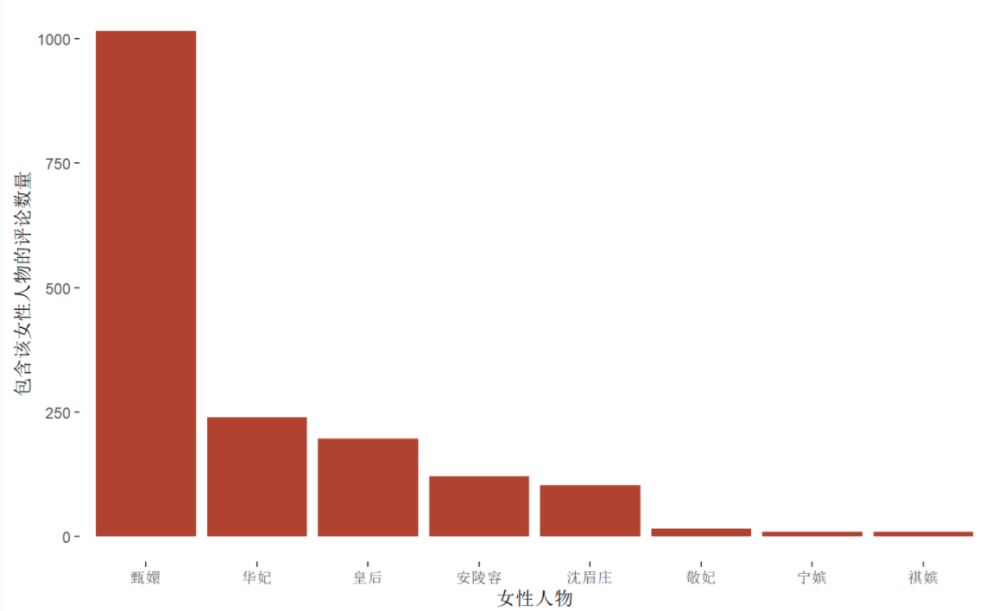


图18 包含女性人物的评论数量柱状图

由于不少评论中用“甄嬛”指代“甄嬛传”，所以提及“甄嬛”的数量远高于其他人物；除了甄嬛外的女性人物，提及频次由高到低分别是华妃、皇后、安陵容、沈眉庄、敬妃、宁嫔、祺嫔。将结合之前的出场幕数可以看出，评论数量与出场次数基本正相关，而其中比较有热度的是皇后，这可能与她上线时间长、前后差别大相关。

**2. 人物评论情感分析**

为了解观众对几个主要女性人物的态度，我们先以逗号、句号、问号、空格为分隔符把每条豆瓣评论分成几个小句子，尽可能让一小句中仅出现一位女性角色，进而以台湾NTUSD正向负向情感词汇语料库为基础，并添加了一些自定义情感词汇，对评价各女性人物的评论（其中，“甄嬛”评论去除了包含“甄嬛传”的评论，“皇后”的评论去除了包含“纯元皇后”的评论）做情感分析，得到如下结论。

表3 对评论中女性人物的情感分析

|  |  |
| --- | --- |
| 女性人物 | 正向情感占比 |
| 沈眉庄 | 80.0% |
| 甄嬛 | 49.9% |
| 华妃 | 40.0% |
| 安陵容 | 39.6% |
| 皇后 | 21.4% |

通过在R中查看各变量（对某一人物的评论）的具体内容，我们认为在评论中提及“甄嬛”的大多是对电视剧的评价，所以在情感分析上不能简单地认为49.9%是对该人物的评价。而提及沈眉庄、华妃、安陵容、皇后的评论大多是对该人物的评价，所以我们可以关注这四个人物的情感分析。

**（1）沈眉庄**

*“喜欢眉姐姐的端庄”*

*“三个人中最爱的还是沈眉庄”*

**（2）华妃**

*“华妃人如其名，华丽而高贵，其美有如夏花般绚烂”*

*“蒋欣的华妃真的可以封神了”*

*“"在看的时候看到开始年妃的狠毒真想按下快进"”*

**（3）安陵容**

*“安小鸟真的很可怕”*

*“多疑敏感、自卑内向的安陵容”*

*“我很欣赏安陵容”*

**（4）皇后**

*“皇后的狠毒和善于伪装”*

*“剧中腹黑阴狠的皇后”*

事实上，正向情感占比的情况与我们最初的设想是相符的，沈眉庄作为剧中甄嬛为数不多的朋友之一，她的温婉端庄深入人心，对她的正向情感占比也位于榜首。而从华妃和安陵容的正向情感评论占比与具体评论中可以看到一个人物的塑造是多面的，对两者评论的正面情感在40%附近表明这样传统意义上的“反面角色”也有她们的亮点和可爱之处；而对于正面情感只有21.4%的皇后，观众则多用狠毒、腹黑去评价，比较符合电视剧中塑造的性格。

**六、研究结论**

通过以上的分析，我们从人物分析的角度总结了对《甄嬛传》出圈原因的三点解读。

一，剧中塑造的丰富人物“群像”是《甄嬛传》经久不衰的重要原因。从对剧的总体评论分析可以看出剧中丰富的人物群像讨论度较高，观众针对许多不同的人物展开了热烈的讨论。本剧并不是主角“一枝独秀”的剧，而是创造了许许多多精彩的人物形象，这也为观众创造了很大的讨论空间。

二，立体多面的人物形象增加了人物的可解读性。从人物形象分析可以发现，本剧的人物都展现了丰富和立体的性格特征，如“华妃”、“安陵容”一类的反派其实也具有她们值得同情和欣赏的特点。剧中的人物不能简单以正反派划分，而更应该理解为一个个生动而真实的人。正是这种人物的多面性使得观众可以从多种角度去理解和解读每个人物，也被人物的魅力所吸引。

三，不同人物的高光时刻贯穿全局使剧情饱满紧凑。在剧情编排上，在主线，即主角甄嬛的成长之路外，不同人物的出场情况和故事的叙述都有详有略，各个人物的高光时刻遍布全剧，如前期对安陵容的“黑化”过程、华妃的由盛转衰详细地刻画，而后期则侧重对祺贵人、敬妃等人物的刻画，这样的安排使得整部剧从头到尾都具有很强的可看性，剧情持续呈现精彩而紧凑的状态。