我要上热门！

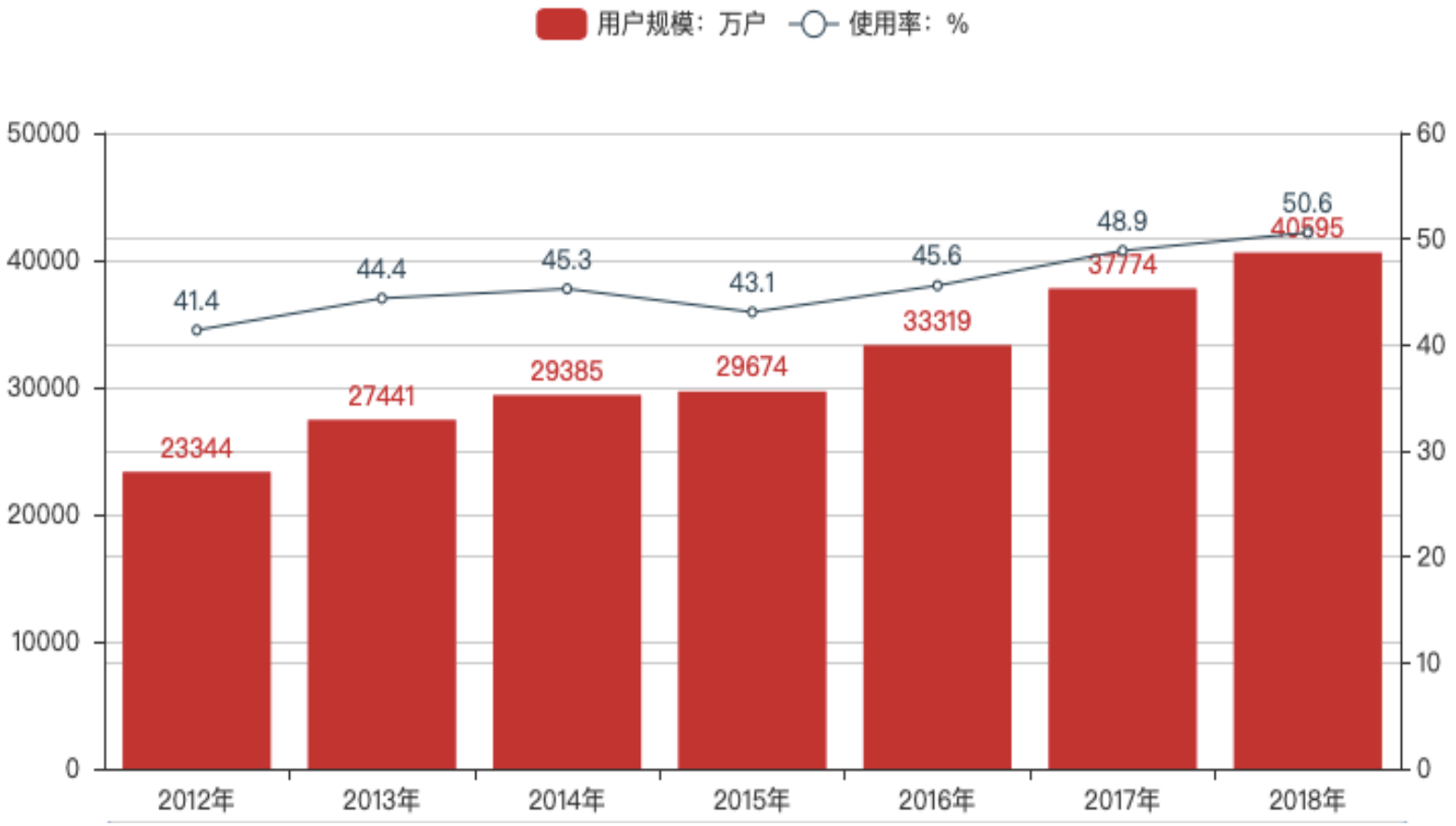
--起点中文网热门小说分析

01

背景

01 背景

2012-2018年中国网络文学用户规模及使用率情况



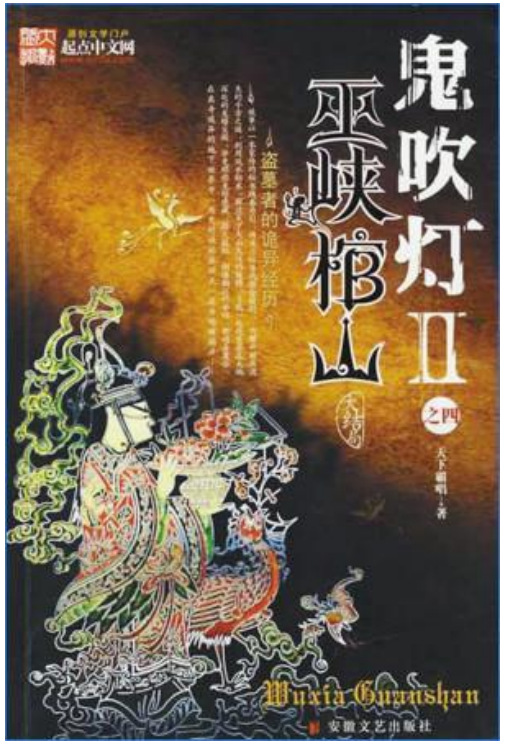
数据来源：中国产业研究院

2018上半年用户规模已达到**4.06亿**人。

预计2020年，中国网络文学作品总数达到**2240万**，作家数达到850万。

01 背景

- 网络作家：唐家三少、南派三叔、唐七公子、匪我思存。
- 由网络小说改编的影视作品：《花千骨》、《鬼吹灯》、《三生三世十里桃花》。



01 研究问题

起点中文网上有哪些人气作家?

热门小说具有哪些特征?

什么类型的小说更热门?

如何按套路写热门小说?

02

描述分析

02 数据说明

从起点中文网上爬取得到22912条小说数据，保留**22285**条数据。

数据预处理： a）删除数据存在缺失或错误的记录

b）将带有单位的数值转化为数字形式

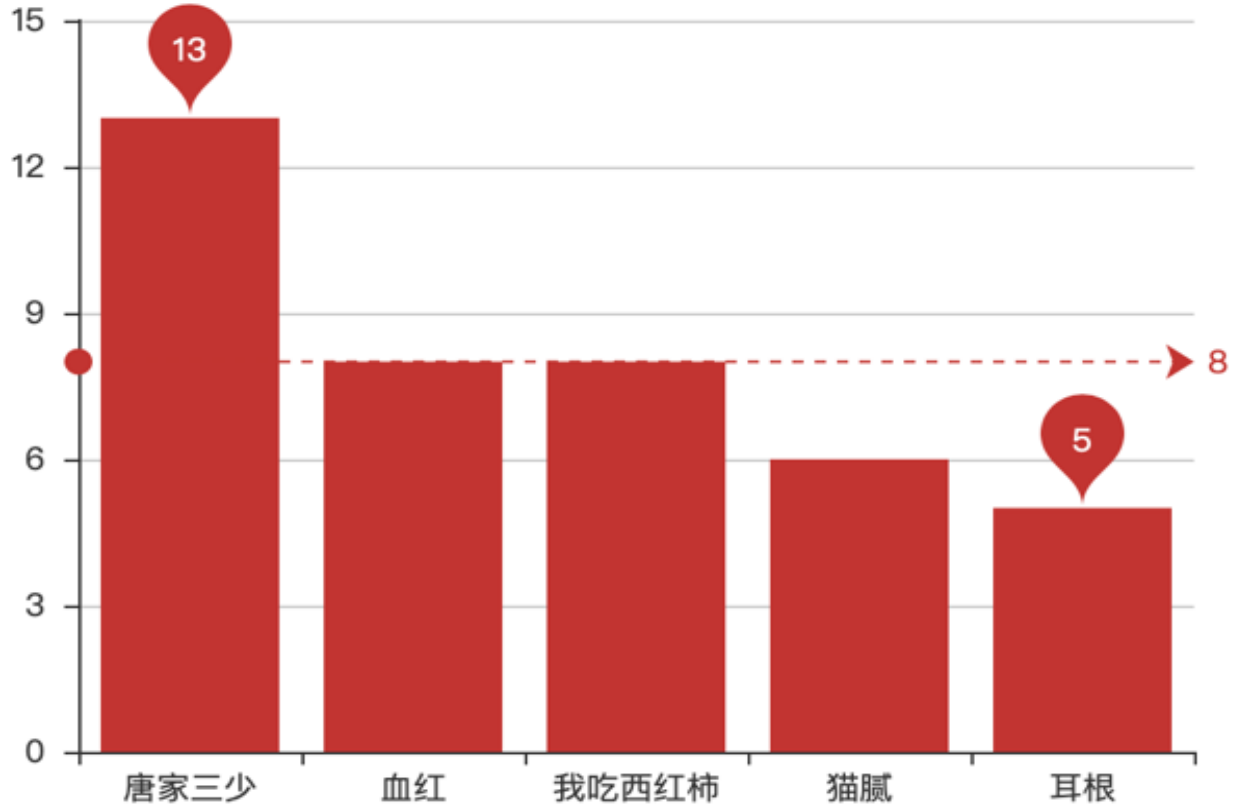
变量类型： **4**个定量变量，**10**个定性变量

小说基本信息： 标签，作者，书名，分类1，分类2，宣传语，作品链接，连载状态，字数，小说ID，简介

小说人气信息： 收藏数，点击数，推荐数

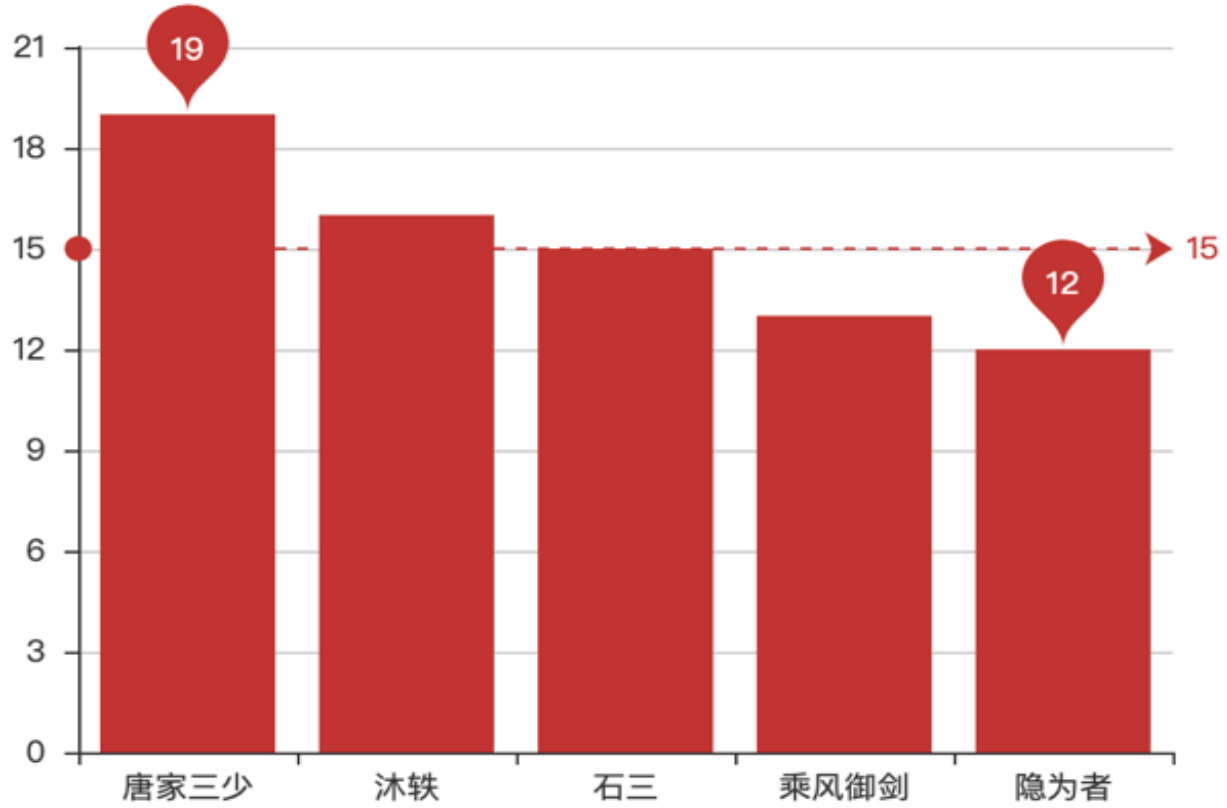
02 最受欢迎的作者

- 统计每个作者写的书有多少本内能够进入起点排行榜前300；
- 进入前300的书最多的五个作者分别是:唐家三少、血红、我吃西红柿、猫腻和耳根；
- 其中，**唐家三少**进入前300的作品有13部，远超第二名。



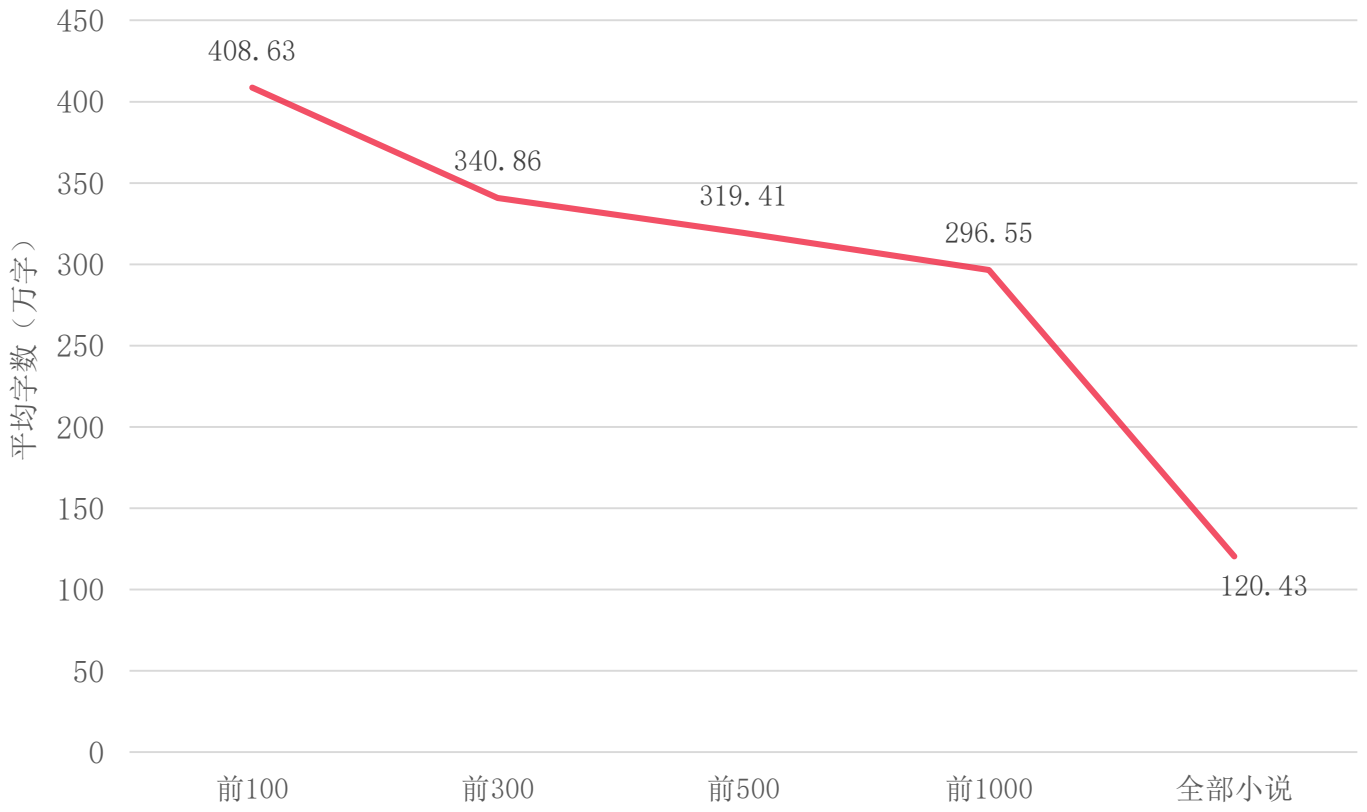
02 最高产的的作者

- 统计每个作者的全部作品数目；
- 虽然总共的书数量有2万多本，但是一个作者的作品数量也不多，不会超过20本；
- **唐家三少**凭借19本的数量也称为高产第一名。

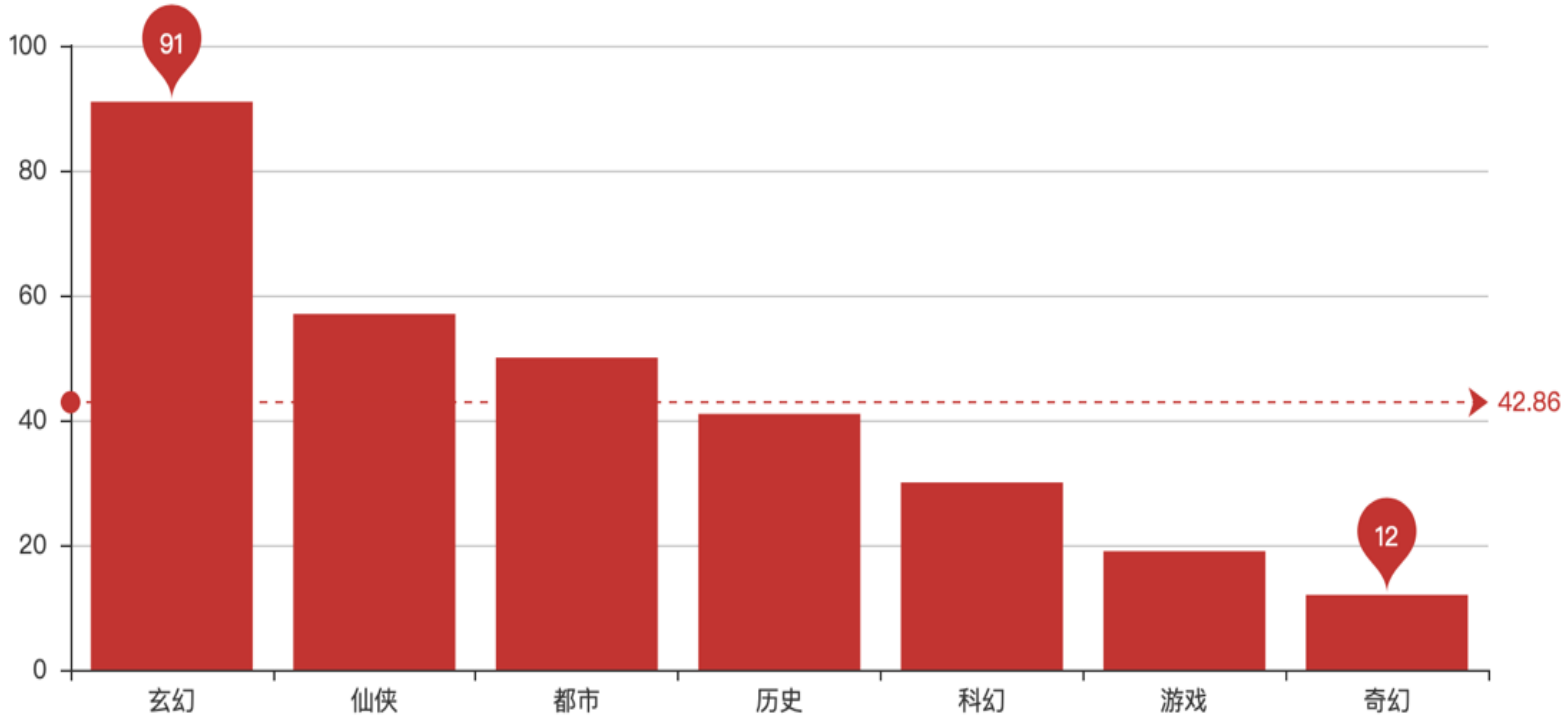


02 小说字数

- **排行越高的小说的平均字数越多；**
- 受欢迎的小说会受到读者支持和打赏，作者更有动力更新；
- 流行的小说本身质量更好，故事更加丰富和完整；
- 同时，热门作者们的作品长度也在不断增长。



02 小说类型

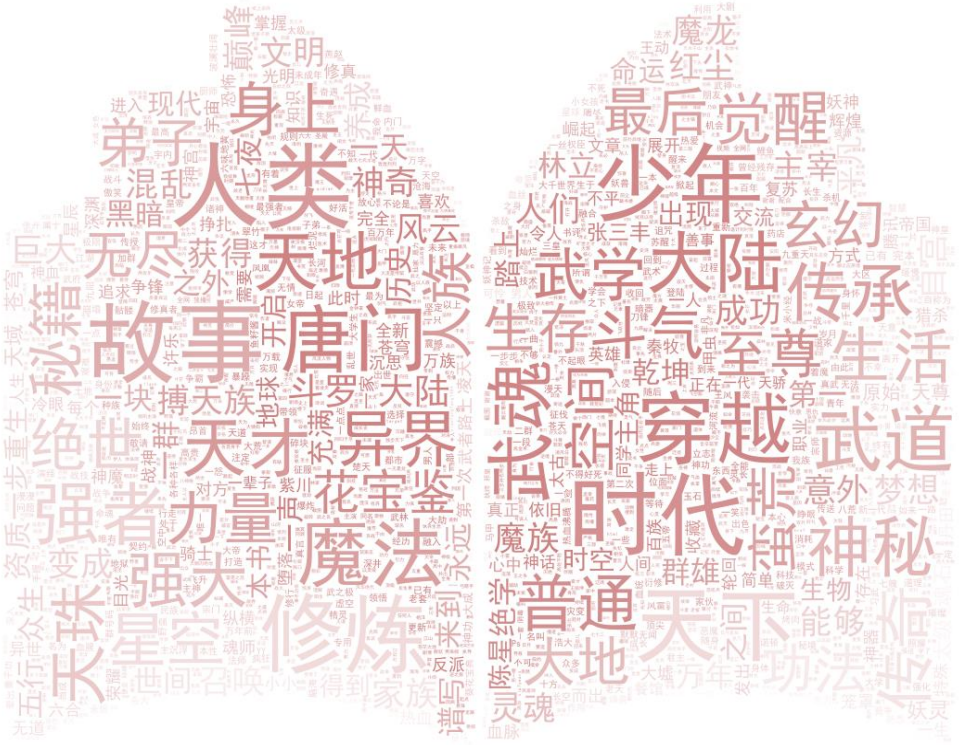


- 统计起点网排行榜前300的小说类别；
- 其中，**玄幻**小说最受欢迎，前300本小说有91本是玄幻小说。

02 小说类型与关键词

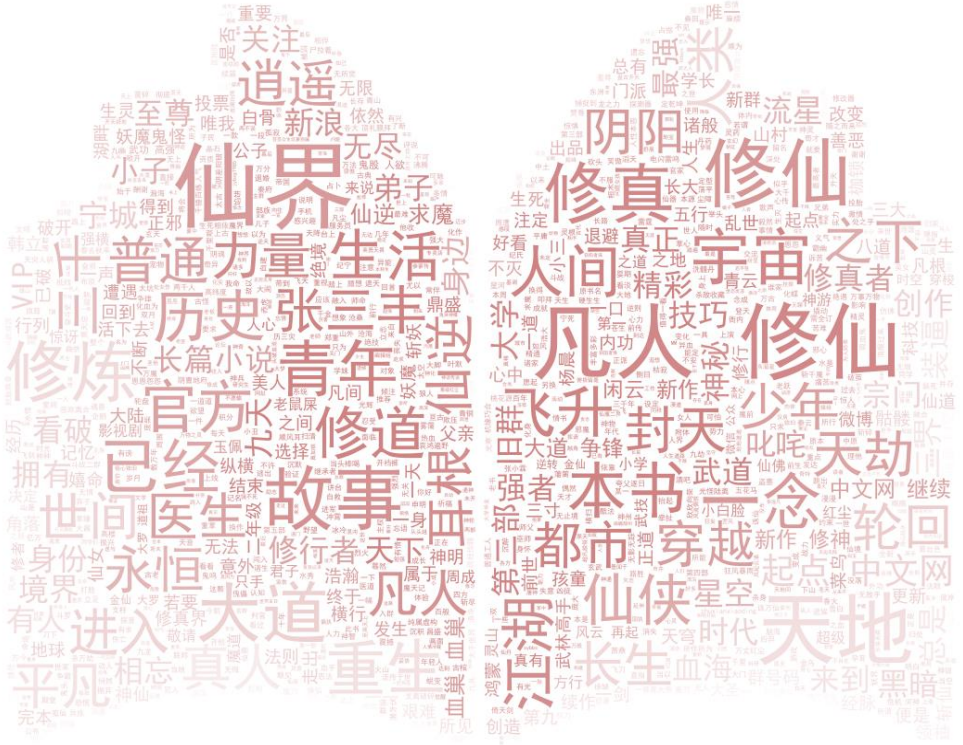
玄幻：

“人类”、“少年”、“武道”



仙侠：

“仙界”、“修真”、“修仙”



02 小说类型与关键词

都市：

“故事”、“人生”、“传奇”

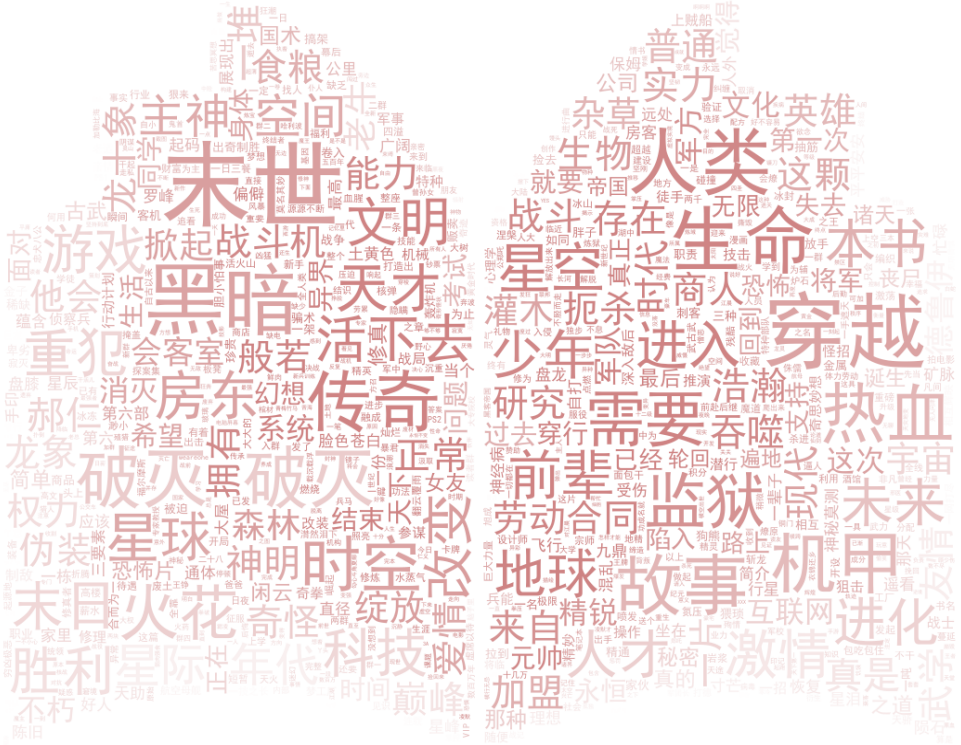
历史：

“穿越”、“大汉”、“皇帝”

02 小说类型与关键词

科幻：

“机甲”、“末世”、“星球”



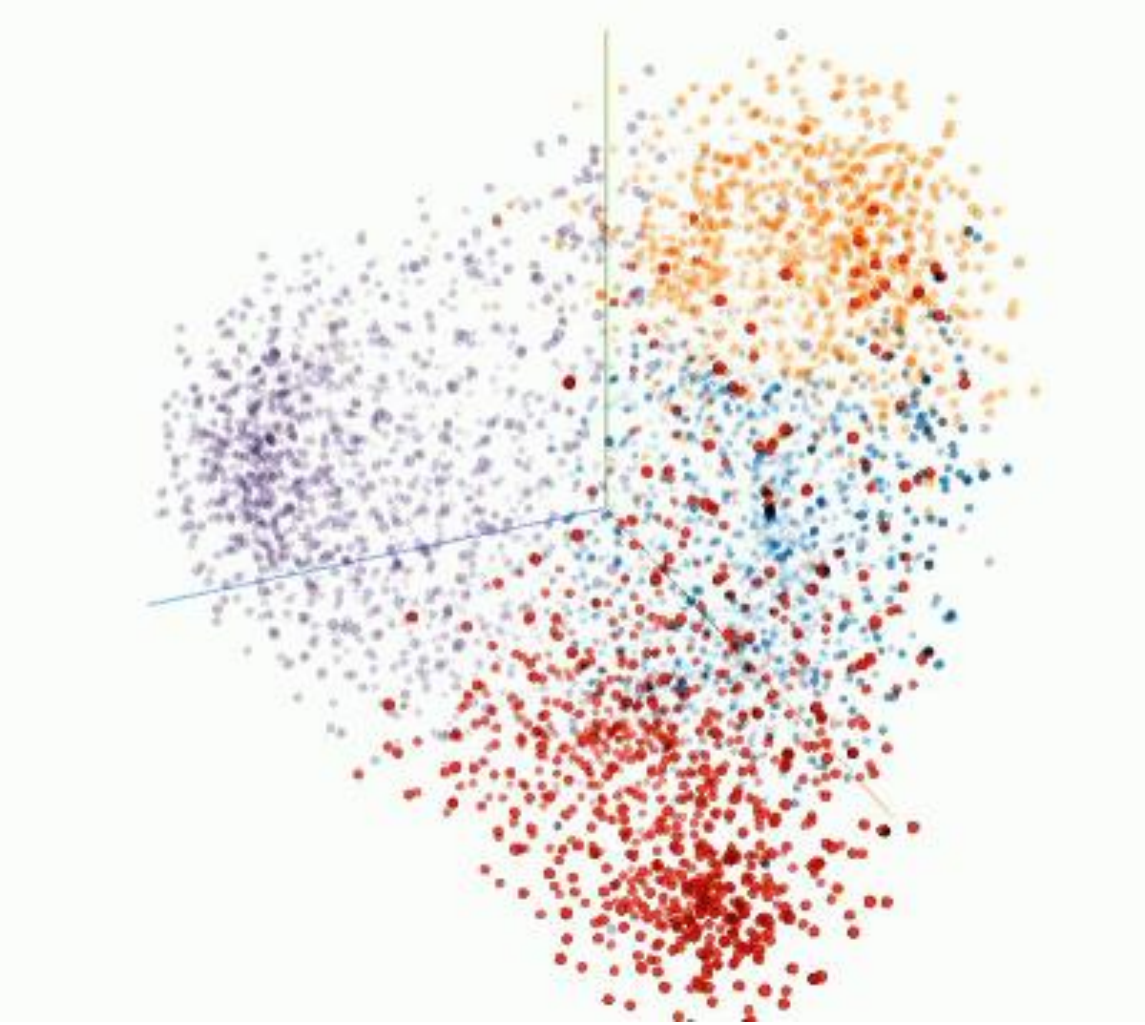
游戏：

“游戏”、“网游”、“法师”

03

模型分析

03 小说分类的降维



生成文档词向量

基于文学作品预训练好的 Chinese Word Vectors 模型，将书名和摘要转化为300维的向量。

词向量降维

将高维词向量进行t-sne降维。

K-means聚类

采用K-means对降维后的向量进行聚类，当聚类数为4的时候聚类效果最佳。

03 聚类结果

1

玄幻修仙类

代表作者：耳根、天蚕土豆、辰东
典型词：“天地”、“宇宙”、“世界”
代表作品：《一念永恒》，《大主宰》，《完美世界》

2

二次元异世界穿越类

代表作者：圣骑士的传说、天运老猫、姐姐的新娘
典型词：“中二”、“穿越”、“异世界”
代表作品：《修真聊天群》，《重生之最强剑神》《文化入侵异世界》

3

玛丽苏穿越类

代表作者：禿笔居士、鱼人二代、王梓钧
典型词：“玛丽苏”、“傻白甜”、“霸道总裁”、“穿越”
代表作品：《大唐仙医》，《校花的贴身高手》，《民国之文豪崛起》

4

神魔奇幻类

代表作者：辰东、我吃西红柿
典型词：“魔法”
代表作品：《圣墟》，《盘龙》

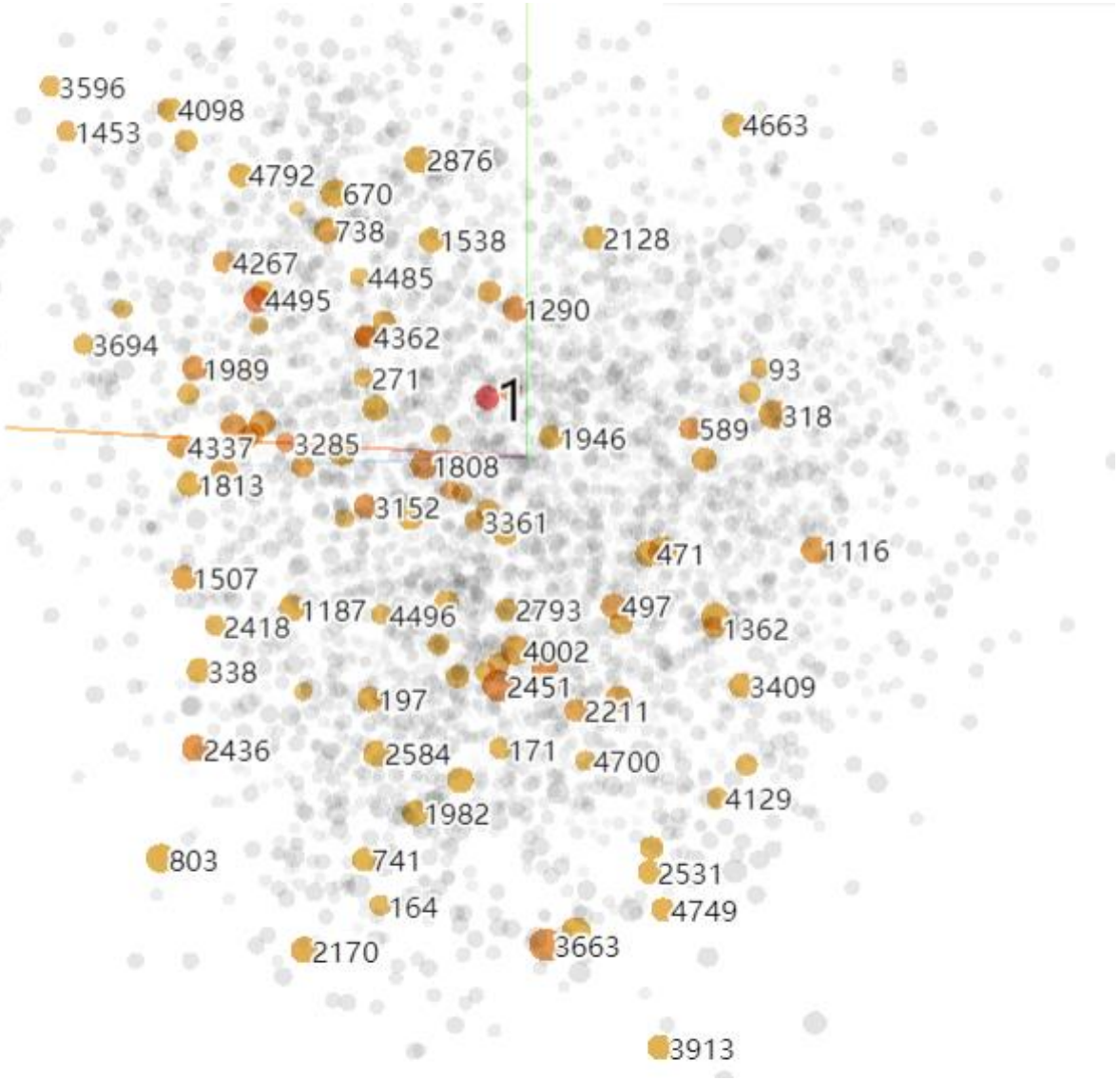
03 同类相似性分析

优秀作品在相似作品中表现突出

对所有小说按照收藏量排序，**序号越小收藏量越多**

序号为1的小说是作者辰东的**《圣墟》**，**收藏量468万。**有颜色的点为与《圣墟》最相似的300本小说。

这300本小说的收藏量远远低于《圣墟》，说明《圣墟》**在其同类别的小说中的创作水平最突出**



03 小说简介潜在主题挖掘——LDA模型

LDA（Latent Dirichlet Allocation）隐含狄利克雷分配模型，是一种文档主题生成模型，也称为一个三层贝叶斯概率模型，包含词、主题和文档三层结构。可以用来识别大规模文档集（document collection）或语料库（corpus）中潜藏的主题信息。

根据之前所聚类模型形成的4大类别，我们分别选取每类中搜藏数最多的前100本小说的简介，进行主题挖掘。

03 小说简介潜在主题挖掘结果

玄幻修仙类

主题1： "传奇" + "江湖" + "仙人" + "世间" + "至尊" + "大道" + "身怀" + "人心"
主题2： "大道" + "轮回" + "凡人" + "长生" + "强者" + "天地" + "一剑" + "神仙"
主题3： "苍穹" + "重生" + "都市" + "大陆" + "星空" + "人间" + "纵横" + "踏上"

二次元异世界穿越类

主题1： "游戏" + "系统" + "时代" + "人生" + "重生" + "异界" + "主角" + "熟悉"
主题2： "系统" + "人生" + "重生" + "历史" + "希望" + "改变" + "传奇" + "穿越"
主题3： "电影" + "游戏" + "文娱" + "冠军" + "中国" + "神话" + "历史" + "高手"

玛丽苏穿越类

主题1： "系统" + "重生" + "回到" + "穿越" + "高手" + "师父" + "功法" + "享受"
主题2： "穿越" + "系统" + "游戏" + "老婆" + "医生" + "发生" + "弟子" + "家伙"
主题3： "穿越" + "喜欢" + "人生" + "兄弟" + "皇帝" + "系统" + "现实" + "游戏"

神魔奇幻类

主题1： "人类" + "时代" + "地球" + "魔法" + "回到" + "进化" + "力量" + "系统"
主题2： "游戏" + "骷髅" + "装备" + "强者" + "科技" + "末世" + "时代" + "奇迹"
主题3： "时代" + "地球" + "人生" + "人类" + "修炼" + "命运" + "改变" + "空间"

03 小说简介主题分析—玄幻修仙类示例

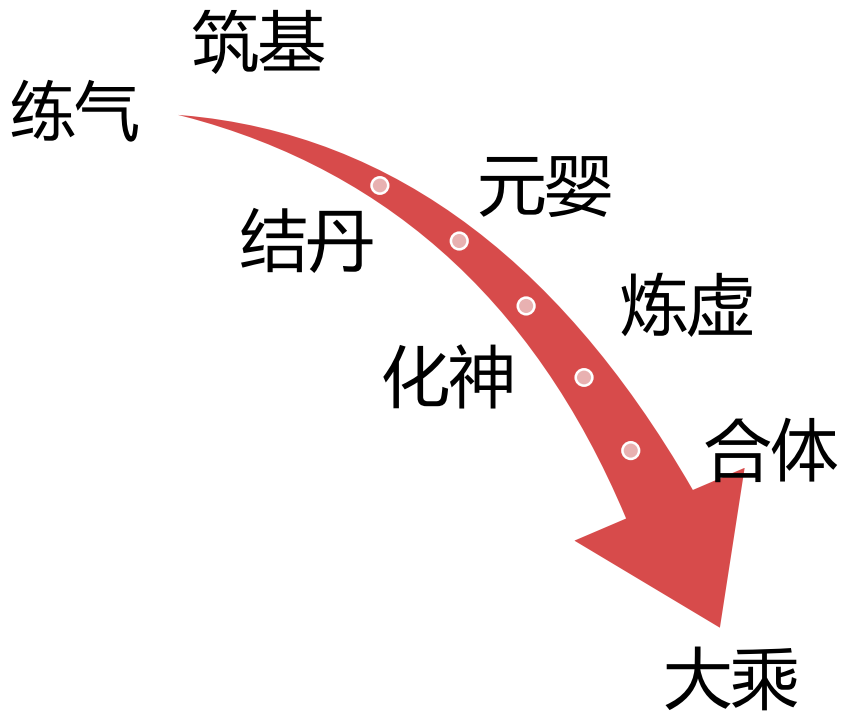
主题1是典型的**以武入道流**，从习武闯荡江湖开始，当小说人气不断累加，问鼎世俗后，引入修仙的地图，以武道作为大道，从而成为世间传奇；

主题2**是凡人修仙流**，从凡人一步步一个个小境界提升从而证道长生；

主题3是**都市修仙流**，在一个大家都不陌生的环境下（都市），接触到修仙，从而纵横人间。

03 升级类小说模式与情节分析

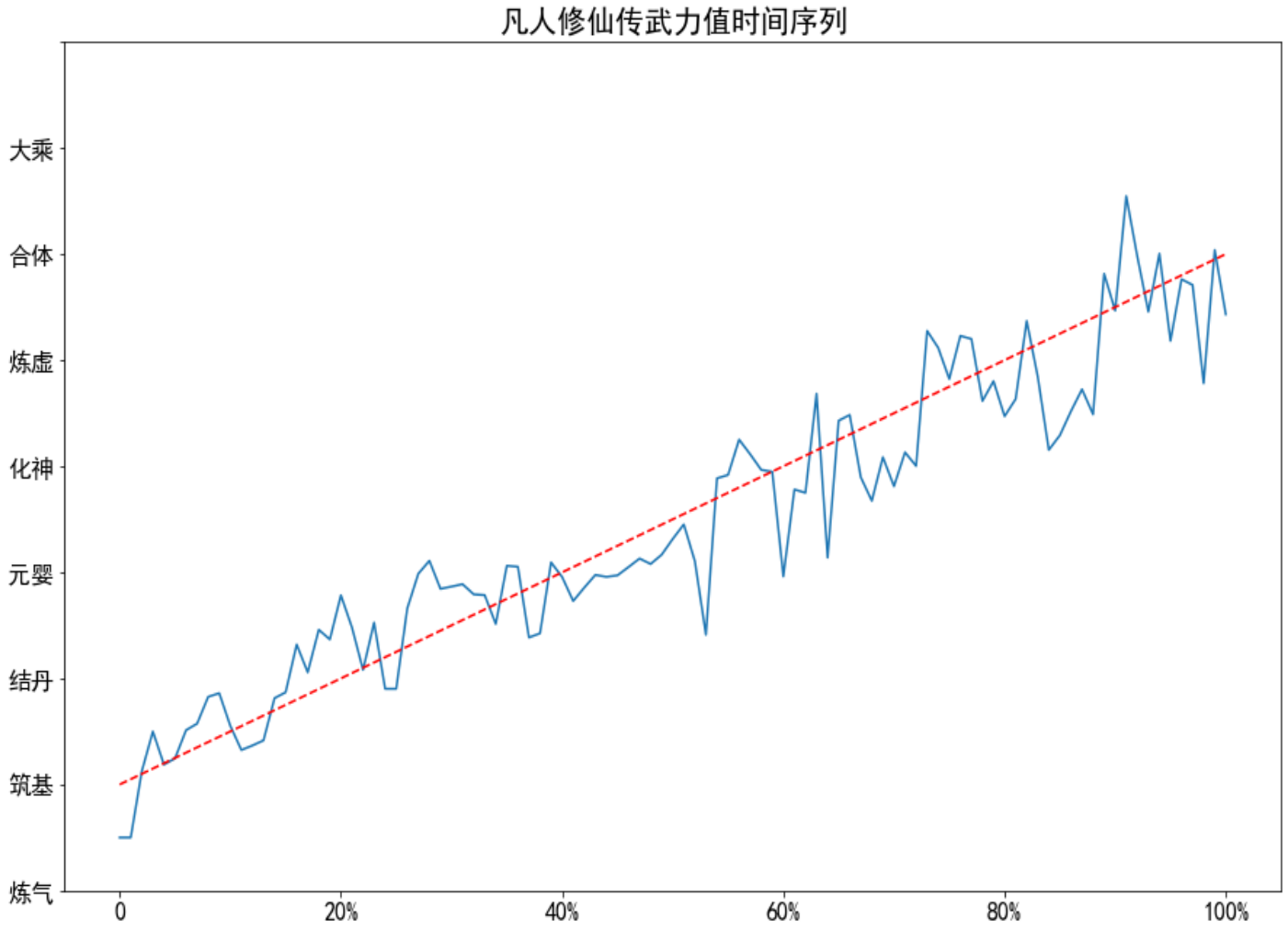
纵观4个分类，12大主题，我们发现每个分类中的小说至少有一种升级模式，即主角从一个普通人逐渐成长为小说世界的巅峰人物，我们称之为“升级”类小说。



03 升级类小说模式与情节分析 —《凡人修仙传》

开创了“升级模式”先河

提取小说每章中出场人物的**境界**，加权平均，将小说章节作为时间轴，去除量纲并进行平滑后形成时间序列。



03 升级类小说模式与情节分析 —《斗破苍穹》

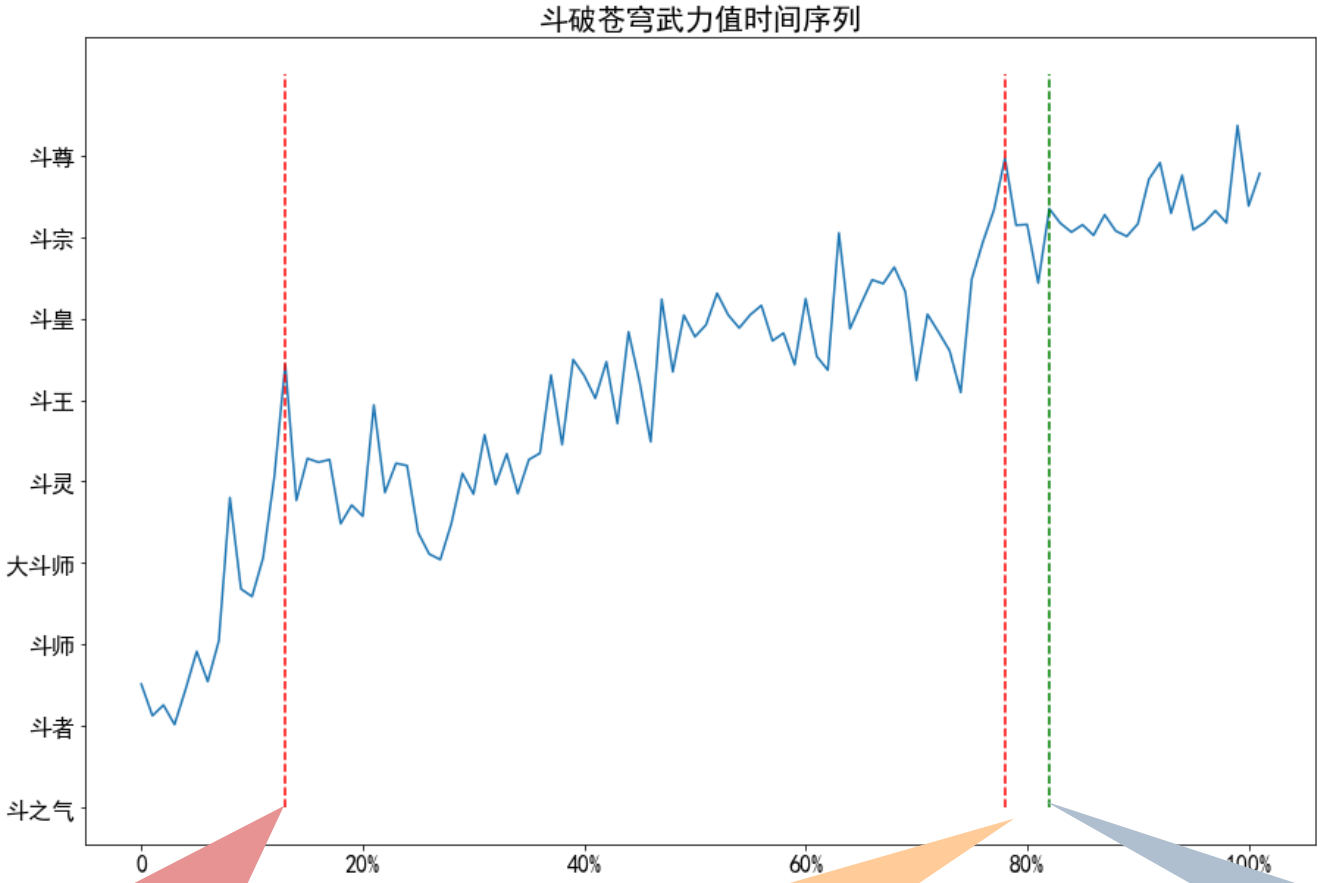
斗破苍穹简介

讲述了天才少年萧炎在创造了家族空前绝后的修炼纪录后突然成了废人，种种打击接踵而至。就在他即将绝望的时候，一缕灵魂从他手上的戒指里浮现，一扇全新的大门在面前开启，经过艰苦修炼最终成就辉煌的故事……

斗破苍穹等级设定

斗之气，斗者，斗师，大斗师，斗灵，斗王，斗皇，斗宗，斗尊，斗圣，斗帝

03 升级类小说模式与情节分析 —《斗破苍穹》



第一个突变：
主角萧炎被“斗皇”
强者追杀

第二个突变：
主角萧炎与众“斗尊”
强者抢夺“斗圣”骸骨

第二个大突变后的
平静：复活药老

04

结论

04 结论

1. 小说可以降维为四大类
2. 优秀作品在相似作品中水平突出
3. 热门小说的简介主题存在固定模式，可以模仿
4. 升级类小说依然有提升空间，可以让情节跌宕起伏

谢谢大家