

程序员分析唐诗朋友圈

白居易最“大V” 李白的确“负心”

全唐诗作数量“无名氏”排第四

2017年2月26日，“前进四”在自己的微信公众号“前进日志”上贴出文章《当我们在读唐诗时，我们在读什么》。

他自称“航天工程师”，日常工作围着数据分析、写程序转，是标准的“程序员”。用“前进四”的话来描述：“这是一个非常严谨的行业，容不得半点差错。”用数据分析遇到的各类问题，是他的职业习惯。

“前进四”说，用科技手

段来分析文艺作品在学术界并不新鲜，其之前也读过相关论文，如浙江大学徐永明《中国古典文学研究的几种可视化途径——以汤显祖研究为例》。只不过这些文章多流传在学术圈中。

为分析唐朝诗歌，他先从网上找来一份全唐诗，其中包括2609位作者，收录42974首诗。通过Python（一种计算机程序设计语言）这种工具，他首先分析唐诗中出现的字、词，找找有哪些好玩的点。

第一个问题，唐朝诗人谁的作品数量最多？程序统计的结果前三名是：白居易（2643首）、杜甫（1151首）、李白（897首）。而位居第四的是“无名氏”（843首）。

“前进四”说，这个统计只局限在全唐诗里，提醒读者其中可能存在误差：“比如全唐诗中白居易的诗最多，这是因为白居易生前自己编纂过文集，李杜则没有，所以白的文章流传到后世的就比李杜的多一些。”

唐诗中出现最多的字是“不”字

紧接着，他又用计算机统计全唐诗中常见的字、词，例如出现最多的字、季节、颜色、植物、动物等词汇分别是哪些。“前进四”说，这些检索十分简单，只需要几行代码就可以实现；而检索的关键词是他本着对于诗歌常见意象的阅读经验想出来的。

借助THULAC分词工具（一款由清华大学开发的中文词法分析工

具），他得出结论：

唐诗中出现最多的字是“不”字；四季出现的频率中，“春”和“秋”呈现出压倒性的场面；颜色中，诗人用得最多的是“白”色；植物中，“松”“竹”最受诗人喜爱；动物中，“龙”“马”出场次数高于其他种类。对此“前进四”开玩笑——“难道唐朝也讲龙马精神”？

根据“前进四”的统计，唐诗

中最常出现的地名是“江南”和“长安”，他在文中如是说——“毕竟一个地方环境好，一个地方是首都，想必这两个地方的房价一定也很贵。”

那么，什么具体场景能引发诗人们的兴致呢？门前、海上、江边、楼上，“这就跟我们现在到景区门前要合影留念，到海边、河边、高楼上都要拍照发朋友圈是一个意思吧”。

最热烈“朋友”：陆龟蒙和皮日休

2017年3月12日，“前进四”又贴出另一篇分析成果《计算机告诉你，唐朝诗人之间的关系到底是什么样的》，10万+的阅读量让他感到惊讶：“最初只是为了好玩，没想到有这么多人喜欢。”

这篇文章，他把关注焦点放在诗人身上，试图理清诗人的“朋友圈”。如何解决呢？借助全唐诗，他把“关系”简化为“引用关系”，即诗的标题和正文中只要提到过对方，那么两者之间的引用关

系加1。一首诗如果提到多次对方，算一次引用。

首先，让我们看一下著名的李白与杜甫。

全唐诗中，杜甫写了12首与李白有关的诗，李白则只有3首与杜甫有关的诗。虽然“前进四”也调侃“李白这种朋友确实差劲了一点”，但对于一段时间网上流传的李杜二人的段子，他觉得“看看笑笑就好”。

“前进四”认为，杜甫写李白

的诗多，一方面因为李白是长辈，比杜甫大了十来岁，成名时间也早得多，更多的是因为杜甫和李白的性格不同：李白飘逸，杜甫深情。

那么，唐代诗人间的唱和互动以哪两人最为频繁呢？排名第一的是晚唐时期的陆龟蒙和皮日休，堪称“绝对的好基友”。

两人互相提到对方的次数均达百次以上，常年酬唱的成果被集成唱和诗集《松陵集》，文学史上也留下“皮陆”的美名。

白居易是“大V中的大V”

另外，从排名前三十的引用关系来看，白居易还与李逢吉、崔玄亮、李建、李绅等名字相连次数很多。

“前进四”说，白居易绝对是唐朝诗人朋友圈中的明星，是“大V中的大V”。

两人间关系有了排名，“前进四”利用ECharts将前一百名引用关系图示化，得到诗人们的社交网络

圈。依据引用度的强弱，关系图显示出了唐朝诗人的两个大型朋友圈：盛唐杜甫-李白朋友圈、中唐白居易朋友圈。

对于白居易这个“社交核心”，“前进四”解释说，社交网络图的核心人物，就是与其他文人发生联系比较多，并且在文坛上有一定名望的人。白居易是其中典型，白与同时代的其他文人都有联系，并且

是当时的诗坛领袖。

除此之外，从初唐、盛唐、中唐、晚唐分阶段看，初唐诗人中关系最好的是宋之问和沈佺期，两位正是宫廷诗人的代表，确定了近体诗的格律，史称“沈宋”；晚唐诗人的社交网络比较散乱，没有明显的核心。其中最重要的就是李商隐和杜牧。

来源：南方都市报



“大V中的大V”白居易（资料图片）

1200多年前，唐朝诗人圈的“小透明”杜甫听说了一个消息：自己的偶像、诗人圈的大V李白，被流放到夜郎国！消息让杜甫辗转难眠，提笔写下“三夜频梦君，情亲见君意”。写诗表达友情或绝交，是唐朝诗人圈的惯例了。

1200多年后，网友“前进四先生”想弄清诗人的关系。唐朝诗人最喜欢用什么词？谁是全唐诗中最大V？唐朝各时期，诗坛产生了怎样的“朋友圈”？他没有被4万多首全唐诗吓倒：“用计算机来做，再多也无所谓”。