## 校长恩情还不完——基于LDA主题模型和词嵌入方法对神针中学公众号文本（2014-2025）的分析

### 摘要

自2017年朱华伟同志就任神针中学校长以来，校园各界日益感受到朱华伟不断散发的个人魅力与伟大恩情。理解校长的恩情形象建构历程，对充分把握朱华伟同志在“建设世界一流高中”历史进程中扮演的角色有重大意义。结合自然语言处理和计算社会科学技术的进展，本研究收集神针中学公众号文本（2014-2025），采用词频统计、LDA主题模型和词嵌入方法进行大数据文本挖掘，探索校长形象和神针形象的历史建构脉络。词频统计结果显示，朱华伟同志散发的伟大光辉远远胜过王占宝、赵立两位前任校长；神针中学校园形象从杰出公民教育向应试一流高中演变。LDA主题模型结果显示，“校长恩情”被算法识别为一个独立主题，自2020年起该主题占比水涨船高。词嵌入模型显示，……

本研究对于学习朱华伟教育思想、忠诚拥护朱华伟校长的英明领导有巨大现实意义。

### 一、引言：尊敬的朱校长，大家早上好！

「尊敬的朱校长，各位领导、亲爱的同事们」——《做有光的深中人！》，2025.7.12

「尊敬的朱校长、敬爱的各位老师、亲爱的学弟学妹们」——《当你问我，高三该如何度过》，2023.10.17

「尊敬的朱校长、各位老师、各位校友」——《七十余载弦歌不绝，校友情深携手共进——深圳中学校友会完成换届选举》，2019.3.27

与我们熟悉的国旗下演讲经典开场词「尊敬的老师、亲爱的同学们」，不知什么时候起，「尊敬的朱校长」早已悄然取代老师、同学和其他伟大领导，成为了神针中学校园演讲的固定发语词。有时，当校长出场时，我们还会看到若干排列组合的头衔，最外层用一个括号括起来：

「“凤翱翔于千仞兮，非梧不栖”。深圳中学校长朱华伟博士（二级教授，博士生导师，享受国务院政府特殊津贴专家，第50届国际数学奥林匹克中国国家队领队、主教练）携全体教职员工诚邀您的加入。」——《深圳中学2025年招聘启动！》2025.3.20

为什么要「用括号括起来」？笔者认为，这是校长在用胡塞尔现象学的语言，悬置自己的身份与头衔，用这种方式表达自己不慕名利、虚怀若谷的精神。在传统文本格式中，括号通常用于插入补充信息，而在现象学视角下，这种括号构成了一种哲学行为：悬置（epoche）。悬置意指对外在世界判断的暂时搁置，以回到意识本身的纯粹显现(Samuel Enoch Stumpf, 2019)。照此理解，我们或许可以说，朱华伟校长以括号悬置自身的社会身份与制度权威，用标点符号谦逊地告诫读者：“请暂时忘记这些浮名，我只是你们的朱校长”。

但是，尽管朱校长正在努力保持低调，作为学生的我们，恩泽于朱华伟校长的伟大光辉，实在无法将其悬置、视而不见。正如一篇荣获北京市嘉源(深圳)律师事务所律师函的文章指出：「当我们赞美学校时，总会说“感谢从天而降的绝世伟人祝华蔚！”当我们批评学校时，又会说“cnmzhw”。这说明，在学校机器的运转下，祝华蔚已经成了我们的神。」(暴论桌, 2025)由此可见，学生早已深刻意识到，我们应当以何种态度理解这位我们的天降恩人。

以上内容构成了本研究的核心问题意识。即，在神针中学沐浴着恩情的今天，我们应当如何理解这种恩情？需要指出，哪怕如朱华伟校长这样的伟人，其恩情也不是与生俱来的神谕，而是在社会建构中得以生成的。然而，这一恩情建构的具体历程目前尚不明朗。本文运用计算社会科学方法，试图理解校长的恩情建构历程，以及这一建构是如何与校园的形象变迁相关联。

### 二、理论回顾：何为恩情？

为此，我们回顾以下理论资源，以揭示校长恩情的象征性建构路径。

2.1 恩情作为合法性资源

布迪厄指出，象征资本是不同类型的资本被认知和认可为合法的形式。

通过媒体、仪式和符号的运用，恩情的建构过程也是学校机器日复一日的象征资本运作过程(皮埃尔·布迪厄, 2012)。当校长的权力关系被象征化、情感化，支配关系也便看起来像是一种道德义务，甚至是一种情感亲密(路易·阿尔都塞, 2019)。

在演讲中，从中层领导、家长校友到学生都下意识地以「尊敬的朱校长」作为演讲开头，

（福柯、布迪厄、阿尔都塞……）

### 三、数据、方法与实证

神针中学于2014年3月创建官方公众号，迄今已有12年，历经王占宝（2010-2016）、赵立（2016.2-2016.11）、朱华伟（2017.2至今）三任校长，发布深圳中学对外宣传的各种信息，是研究校长恩情建构与校园形象变迁的优质文本材料。因此，本研究选取神针中学公众号文章数据（2014-2025）开展实证性研究。

需要指出的是，公众号作为学校领导层对外宣传的窗口，展现的是神针中学领导层意图向考生、社会展现的官方叙事形象，不能代表神针中学的真正的校园文化，更不能代表老师、同学们在神针中学在地的生活体验。但是，我们仍旧可以通过公众号文本资料探究校长恩情建构这一议题——因为公众号作为象征资本运作的场域，事实上深度参与了校长形象生成和权力合法化的进程。这就像一家公司，有可能对内拖欠员工工资，对外欺骗市场，这些内容当然不会展现在公司的官方网站里；但是，通过公司官网三天两头放出的老板视察、领导讲话，投资者还是可以通过文本分析观察领导层试图向公司内外构建的形象及其演变。

我们收集了神针中学公众号自创建以来的所有推文，经过数据清洗、剔除较短或无意义的异常数据后，总共得到文章总数2503篇，平均长度2507字，最大长度12641字，总字符627万字，达到了LDA模型与词嵌入模型对数据集字符数的稳健性要求。下图展示了文章长度分布情况，观察到文章长度分布具有右偏长尾特征，经检验符合对数正态分布，验证了数据集的有效性。

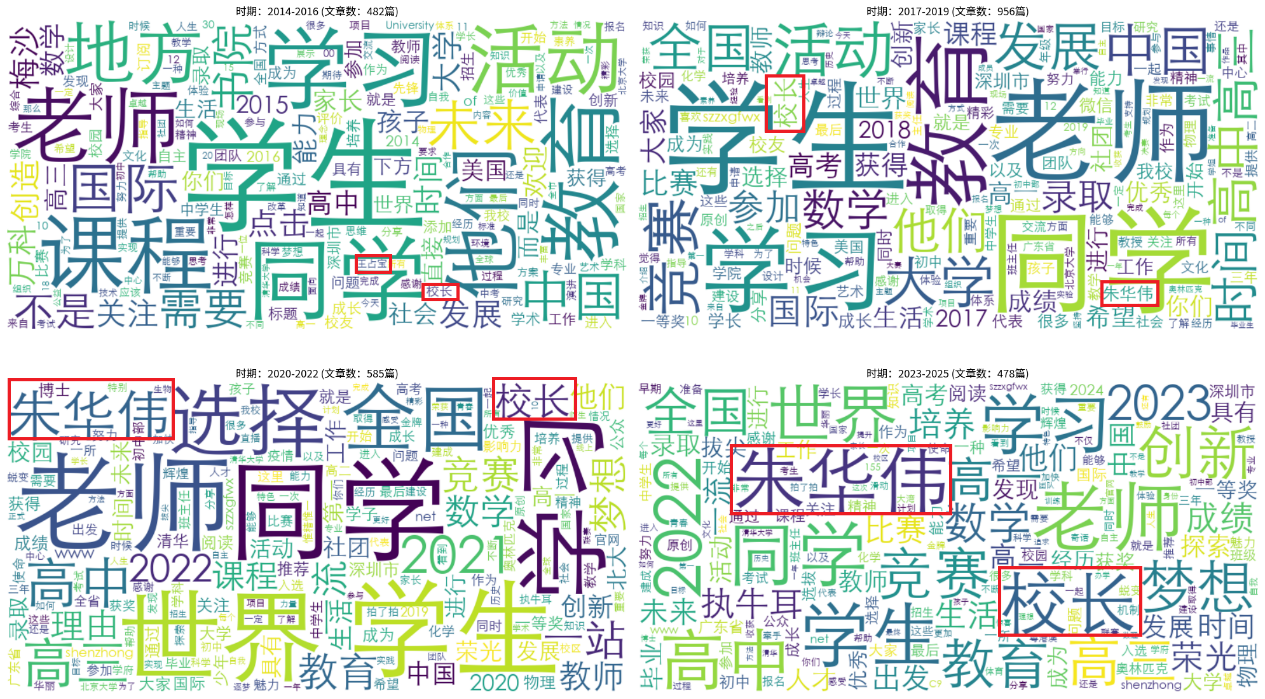
图表, 直方图

AI 生成的内容可能不正确。

我们首先运用词云图对公众号文本在不同时期的内容取向进行描述性观察。运用jieba库进行分词处理，并将无意义的虚词、单字词设置为停用词后，绘制总词云图如下所示：



为了进一步直观观察我们关心的词汇变化趋势，进一步将出现频率极高但无分析价值的词（深圳、中学、深中、我们等）设置为停用词，按照3年为单位分期，绘制分期词云图，如下所示：



注意到，在2014-2016年（王占宝-赵立时期），王占宝的姓名在词云图中占据面积极小，只出现在同学们的夹缝中。2017-2022年，“校长”开始逐渐膨胀。到2023-2025年期间，校长的大小已经膨胀至超越了老师、同学、学生等词汇，向社会公众宣示着，谁才是这座学校里真正的主人。

我们进一步精细观察部分词汇的提及次数变化曲线：

图表

AI 生成的内容可能不正确。

从词频曲线中可以更直观地观察到，如果说王占宝、赵立的曲线只是两座小山包，那么朱华伟就是一座高耸入云的珠穆朗玛峰。就任不到一个季度，词频曲线就已超越了王占宝、赵立两任校长的历史最高点。值得注意的是，在2020年以前（即华伟-晒布时代），「朱华伟」曲线和「校长」曲线有一定程度的分离，说明此时朱校长尚未充分完成自身的身份建构，对于宣传自己的姓名尚且较为羞涩，大部分时候以「校长」取代之。然而，在2020年后（即华伟-泥岗时代），两条曲线高度重合，意味着此时朱华伟已经充分进入了校长角色。

根据三位校长的词频和任期长度数据，我们计算并可视化了校长季度平均提及次数。王占宝和赵立，公众号的季度提及次数均为25次上下，而朱华伟则高达112.6次，大约是其前任的4.5倍：

日程表

AI 生成的内容可能不正确。

（说明：由于朱华伟先生非常爱护自己的肖像权，为了避免收到神针中学法务团队**北京市嘉源(深圳)律师事务所**的律师函，我们不得不以这种损害可视化艺术性的形式呈现，以表达我们对朱华伟先生肖像权的充分尊敬。当然，笔者依旧充分认可神中法律团队的业务能力，不知是否还会延续律师函治校的人类教育奇迹？。）

「校园民主」「公民教育」「学生权利」是王铮-王占宝时代的神针中学教育关键词。我们绘制了「民主」「公民」「权利」三词的词频曲线：

图表, 直方图

AI 生成的内容可能不正确。

观察到，三词均在王占宝-赵立年代有较为频繁的使用，而在朱华伟到任后销声匿迹。俗话说「不破不立」，在朱华伟校长逐步清理了前任校长遗留下来的神针中学历史垃圾后，我们好奇校长建立了一个什么样的新深中。我们计算了所有词在2017年，即朱华伟到任前后的词频差，分别对词频上升幅度最大、下降幅度最大的30个词进行可视化，结果如下：



上升幅度最大的词包括：朱华伟、校长、深中、中学、梦想、荣光、辉煌、拔尖、一流、清华、北大……

下降幅度最大的词包括：学生（top1）、校友、活动、书院、美国、国际……

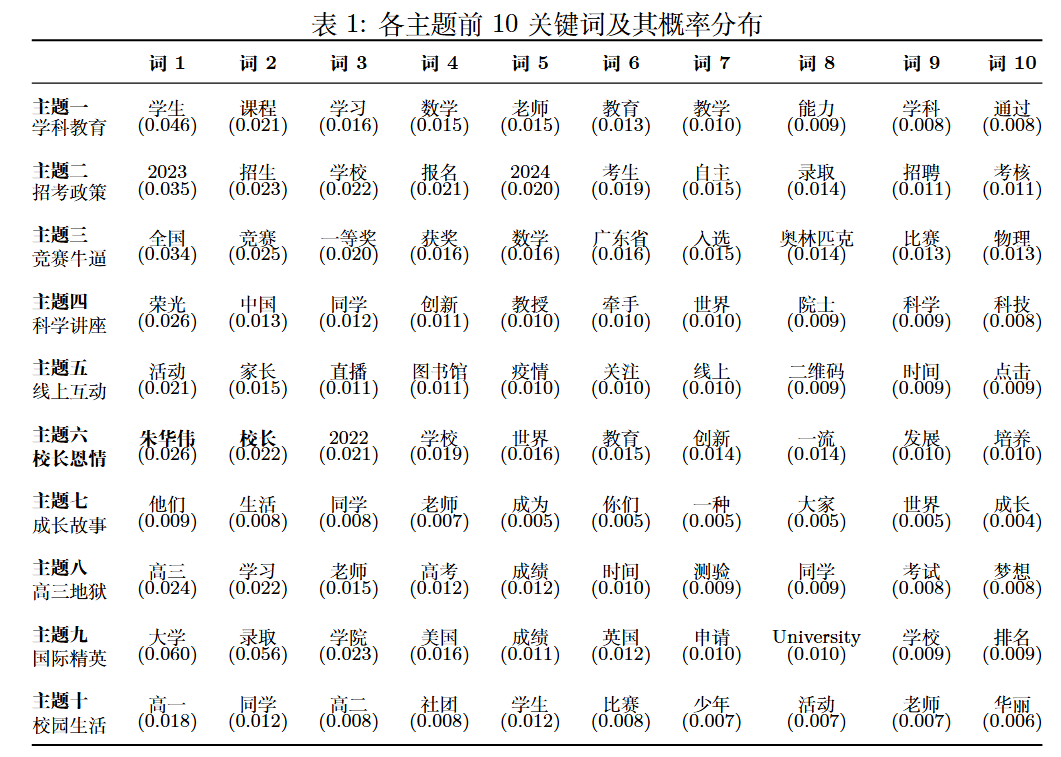
不难发现，词频的上升/下降与两个时代使用最频繁的校园-政治标语系统最直接相关。前朱华伟时代强调「杰出公民」，王占宝常用的表述是「未来不是**我们**要去的地方，而是**我们**要**创造**的**地方**」，而朱华伟常用的表述是「**梦想**与**荣光**」「**拔尖**创新人才」「中国特色**世界一流高中**」。从「杰出公民」到「**一流**高中」，从「丰富生命力」到「梦想与荣光」，关于梦想是谁的梦想、一流是哪一种一流，相信从这两张词频幅度的变化图中我们能够得到答案。

反复对比两张图里的词汇，笔者将两个时代的特征概括为三重转变：第一，「美国」「国际」的下降、「清华」「北大」「全国」的上升，这是从「世界公民」到「中国精英」的转变；第二，「学生」「书院」「活动」的下降，「清华」「北大」「拔尖」的上升，这是从「自由校园」到「一流应试」的转变；第三，「学生」「校友」「我们」的下降，「深中」「校长」「朱华伟」的上升，这是从「学生中心」到「校方-领导中心」的转变。事实证明，校长不仅善于破坏一个旧世界，更善于创造一个新世界。这三重转变，构成了朱华伟时代的神中底色。

词频分析能够初步揭示出不同时代的风气转变，有着强大的可解释性，然而这一较为原始的方法也有其局限性：它将每个词视为原子单位，忽略了词语的语境、多义性和词与词之间的联合使用模式。

为进一步揭示公众号文本中词汇背后的语义社群和主题分布，我们引入 LDA（Latent Dirichlet Allocation，潜在狄利克雷分配）主题模型。LDA 是一种无监督机器学习方法，假设每篇文章都是若干主题的混合体，每个主题又是若干词汇的概率分布。通过建模词与主题、主题与文档之间的共现关系，LDA 能够自动识别语料中反复出现的“主题语境”，从而超越词频所能触及的表层统计，进入更深层的话语结构。简言之，如果词频是一种“量”的统计，那么 LDA 更接近于“类”的划分。近年来，LDA方法已经成为社会科学、特别是传播学领域得到普及的基础性文本挖掘方法(Kushwaha R 等, 2025; Lazer et al., 2009; Macanovic A, 2022; Sivertsen, 2025)。

作为一种较早期的主题模型，LDA要求研究者手动选择主题数。通过综合困惑度（perplexity）和连贯性（coherence）两个指标的下降曲线，我们认为选择主题数为10能够最好地反应文本的语义结构特征。过滤单字词和停用词后训练主题模型，得到如下的主题-词汇概率分布：



LDA返回的是每个主题对应的高频词及其概率分布，每个主题的实际内容需要根据其关键词进行人工归纳。举例来说，观察主题六，我们惊喜地发现，「朱华伟」「校长」成为了该主题频率最高的词汇！结合「一流」等词，我们将主题六归纳为**「校长恩情」**。作为一种无监督机器学习算法，LDA的运行并没有任何人类输入的先验信息，即使如此，LDA也深刻领悟到了朱华伟同志满满溢出的恩情，连算法模型都为之动容。

根据同样的方法，我们把其他主题依次归纳为「学科教育」「招考政策」「竞赛牛逼」「科学讲座」「线上互动」「成长故事」「高三地狱」「国际精英」「校园生活」。

竞赛方向、国际方向、高一高二、高三，这四个在校园空间中被区隔的场域，在语义空间中同样存在着显著的区隔。「区隔」是形容朱华伟时代再恰切不过的词汇，场域和身份的区隔，决定着神中学生的日常生活的社会分层结构。依据LDA输出的结果，我们从中可以读出三重区隔：

图片包含 文本

AI 生成的内容可能不正确。第一重区隔，是高一高二与高三的区隔，参考主题模型中的「高三地狱」和「校园生活」。在20年以后，高一高二与高三被泥岗、晒布校区地理分割，也因为划开了泾渭分明的两个场域。高一高二的泥岗生活叙事呈现出一种鲜明的自由想象：社团、活动、比赛、逐梦，学生们似乎过着一种自定义的人生，至少在公众号文本中展现出来的叙事里如此。且不说泥岗提供的自定义空间如何被近年来的领导层蚕食的故事，学生们的生活选择本身即是文化资本分层的结果——考入名校、成为「神中学子」登上公众号栏目的精英们有权力把自己的故事嵌入右边的词云图，而更多的普通学生，或者说失败者，是事实上的沉默者。我们不得而知这群人是否真的在高一高二体验过何为「自由」。而当泥岗提供的自由幻光消失，学生们坐上地铁九号线转三号线来到晒布，很快发现自己坠入了高三、成绩、学习、考试等词语织就的大网。



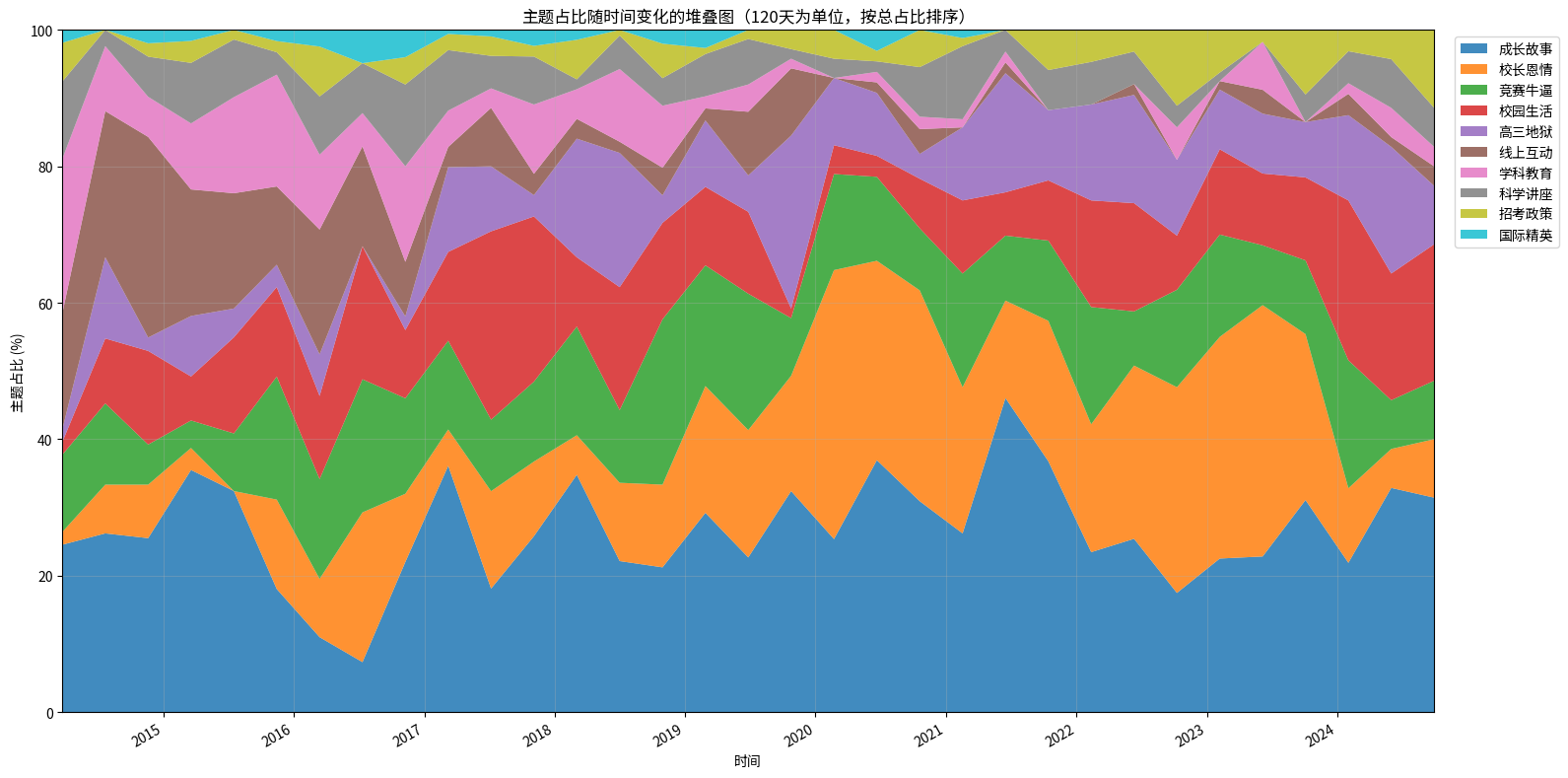
第二重区隔，是竞赛生与非竞赛生的区隔。下图展示了「竞赛牛逼」和「国际精英」两个主题的高频词词云。学校对竞赛成就、国际升学的宣传是高度精英主义的，在这两个主题中，我们看不到任何除了学生最终的耀眼成就以外的任何故事。对竞赛生尤甚：虽然笔者本人并未接触竞赛，但也对神中竞赛圈的一些故事有所耳闻。在这片土地上，享有最高层级占有资格的，比如教育界的神中，从来就是一个**采购商**，是**筛选和使用者**，而极少是一个**菜农**，去想着怎么费劲**培养和因材施教——**在竞赛体系尤甚(漫士呓语, 2023)。而这个过程中被丢弃的烂白菜，采购商自然不必施舍任何关怀。于是我们听说，学生因为没进国集而不被公众号「祝贺」「感谢」，成为了“等”字中的一员；我们听说，初中学了多年生竞的同学，连考试机会都不被允许。最后，采购商得到一条一条的一等奖推文，作为自己的功劳，在盛世的欢呼中装点着官方文件(孔飞力, 2012)。



第三重区隔，是国际方向与非国际方向的区隔。学校对国际升学的叙事同样高度精英化，夹带着University of xx的洋文。国际生的升学有着雄厚的家庭背景托底，他们因此在高中三年里走着不一样的教育路径、做着许多fancy的工作，这是更大的社会分层在教育场域中的符应（correspondence）(S·鲍尔斯 等, 1990)，本来并非学校的过错。然而，这一重区隔在2023届、2024届两届高三毕业典礼上得到了令人叹为观止的展现

……

为了展示这些主题在文档中的分布随时间的演化情况，我们用堆叠条形图进行可视化。主题分布模式的两个关键转折点发生在2017年与2020年，分别是校长就任、以及泥岗校区投入使用。以这两个时间点为分界，可以把2014-2025年分为三个历史阶段：「占宝-赵立时期」「华伟-晒布时期」与「华伟-泥岗时期」[[1]](#footnote-1)。下文将对主题分布变迁结果进行解释。



在2017年以前，「高三地狱」主题（浅紫色）占比极小。一方面，彼时尚未出现泥岗-晒布的空间区隔，高三生仍有机会深度参与自由的校园生活，我们看见，当时的学长学姐在高三仍旧深度参与社团活动者不在少数。另一方面，制度对高三学习的控制亦尚未强化到华伟时期的程度。可以想象，总的来说，当时的高三生的生活是比今天的晒布同学丰富自由的。而在2017年的拐点以后，该主题占比迅速攀升，暗示着高三与高一高二区隔的形成。在2020年后，该主题占比进一步上升，说明空间的区隔进一步强化了高三的非自由生活状态。

「校长恩情」主题的攀升同样发生在2017年，迅速占据大片面积，很快成为占比第二高（某些时间点甚至最高）的主题，验证了上文词频分析中关于

1. 感谢L学长对于历史分期的建议。 [↑](#footnote-ref-1)