数据计算与分析报告

我设计的数据指标体系在以实现为前提下进行了简化，共有两个二级指标：互动力和生产力，生产力指标以月均发布视频数作为唯一衡量标准，互动力指标下设总浏览数（views）、总点赞数（likes）和总评论数（comment-count），通过(总浏览数\*0.01+总点赞数+总评论数）计算，结果作为互动力指数，互动力指数的十万分之一与生产力指数之和作为评价频道的总指标，数字越大代表该频道在视频网站上影响力更大。我对总浏览量进行了加权，因为相对于点赞量和评论量，浏览量的数值更大，对最终互动指标结果影响大，但点赞行为和评价行为比浏览行为更能体现观看者对视频的积极反馈，因此需要对浏览量进行加权，减小其对最终结果的影响。另外，在计算总指标时对互动力指标进行了加权，以均衡生产力指标和互动力指标对总指标的影响。

从研究结果看，ibight、TheEllenShow、ESPN、The Tonight Show Starring Jimmy Fallon、Jimmy Kimmel Live等十五个频道排在总指标的前十五名，其中第一位ibight，第九位Dude Perfect，第十三位Marvel Entertainment和第十五位Logan Paul vlogs频道的影响力主要来自于互动力，ibight更是以超高的互动力指标获得最高的总指标值，位列第一。其余11位上榜频道的影响力主要来自生产力指标。

观察这些频道内的单个视频的浏览数、点赞数和评价数，我发现在这些频道发布的视频中，往往有几个视频的互动力指标额外得高。那么每个频道都是如此吗？在描述用户生产内容时，有UGC和PGC两种概念，区别主要在于提供的内容是否具有高质量、专业化的特点，偏向于PGC的平台在用户和粉丝量之间的关系存在极化现象，具有高粉丝量的用户较少且与其他用户之间在影响力方面差距极大。如果能够知道频道内每个视频的用户信息，我们也可以分析每个频道的用户生产内容特点，从而判断其影响力的来源和发展前景等等。但是由于数据有限，我只能先研究频道内单个视频的互动力指标分布，而不能具体到用户。因为研究结构需要有足够大的个体基数，因此我的研究对象集中在了所有在2017年11月到2018年5月期间发布视频数超过五十的频道。最后通过折线图的对比，我发现ESPN频道单个视频结构更加类似于UGC平台的用户和粉丝量结构，每个视频影响力较平均，质量稳定。

我想要研究的另一个问题是争议性和互动力之间的关系，是否争议性越大的视频会带来更多的讨论，因此有更高的互动力？还是有更多互动力的视频因为观看和讨论基数大就会产生更多的争议？经过对争议指数（|1-点赞量\点踩量|）和互动力指标的相关性分析，相关系数始终小于0.2，属于低度相关。所以从最简单的相关性分析来看争议性和互动力之间呈现低度的正相关关系。