



2017第八届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2017

# 可视化技术研究

## ——传媒大数据可视化

主讲人：王鑫



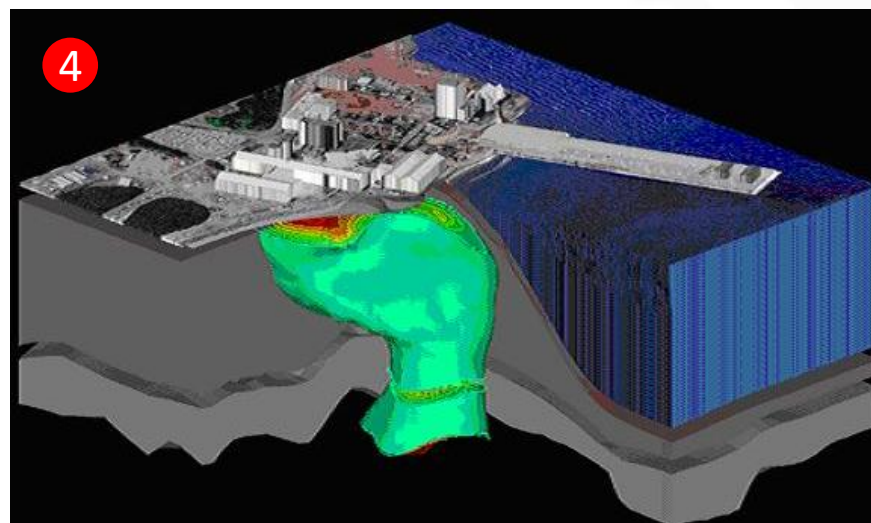
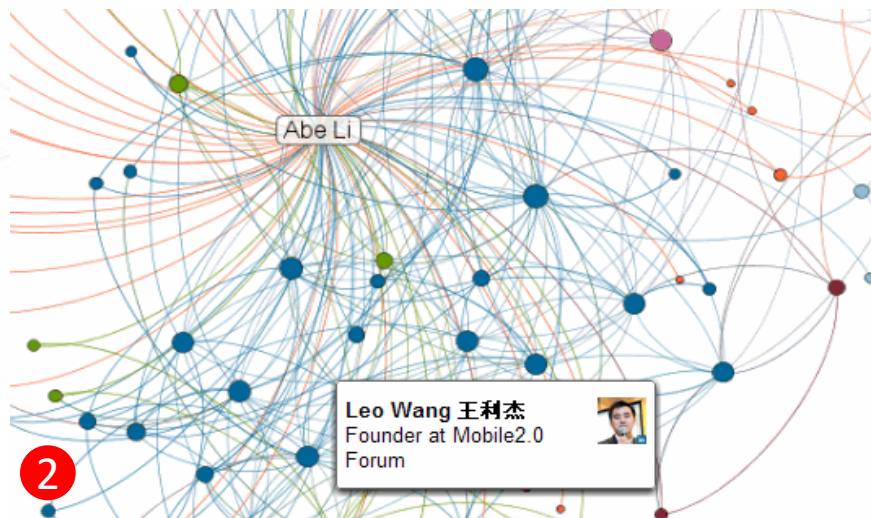
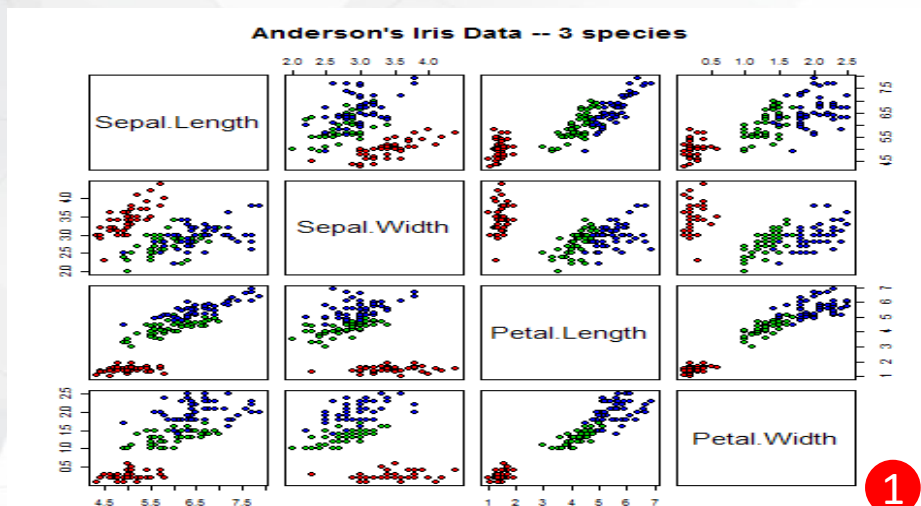
数据可视化不仅是一门包含各种算法的技术，还是一个具有方法论的学科。数据可视化不仅是在信息丛林中找到方向的最好方法，还能帮助人们并发现全新的视角。

——大卫·麦克德里斯《数据可视化之美》



现代的数据可视化技术，综合运用**计算机图形学**、**图形处理**、**人机交互**等技术，将采集或模拟的数据变换为可识别的**图形符号**、**图形**、**视频**或**动画**，并以此呈现对用户有价值的信息。

# 可视化技术分类

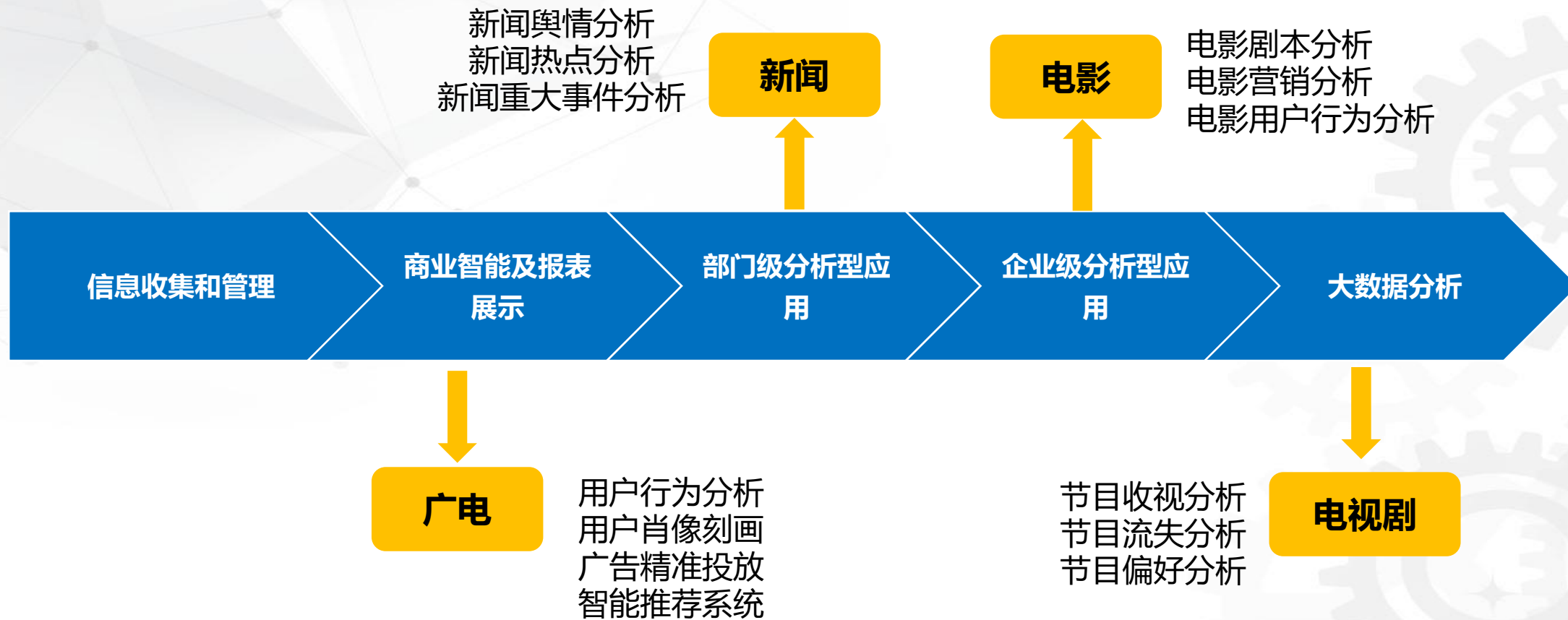


可视化主流方向：

- ①数据可视化
- ②社交可视化
- ③文本可视化
- ④地理可视化



# 可视化技术行业应用





日记卡

数据采集方式

4岁以上人员，人工填写，  
每周回收

记录时间单位

15分钟

数据提供速度

15工作日，人工收集

数据准确性

一张卡记录一人一周的收视，  
基于回忆

测量仪

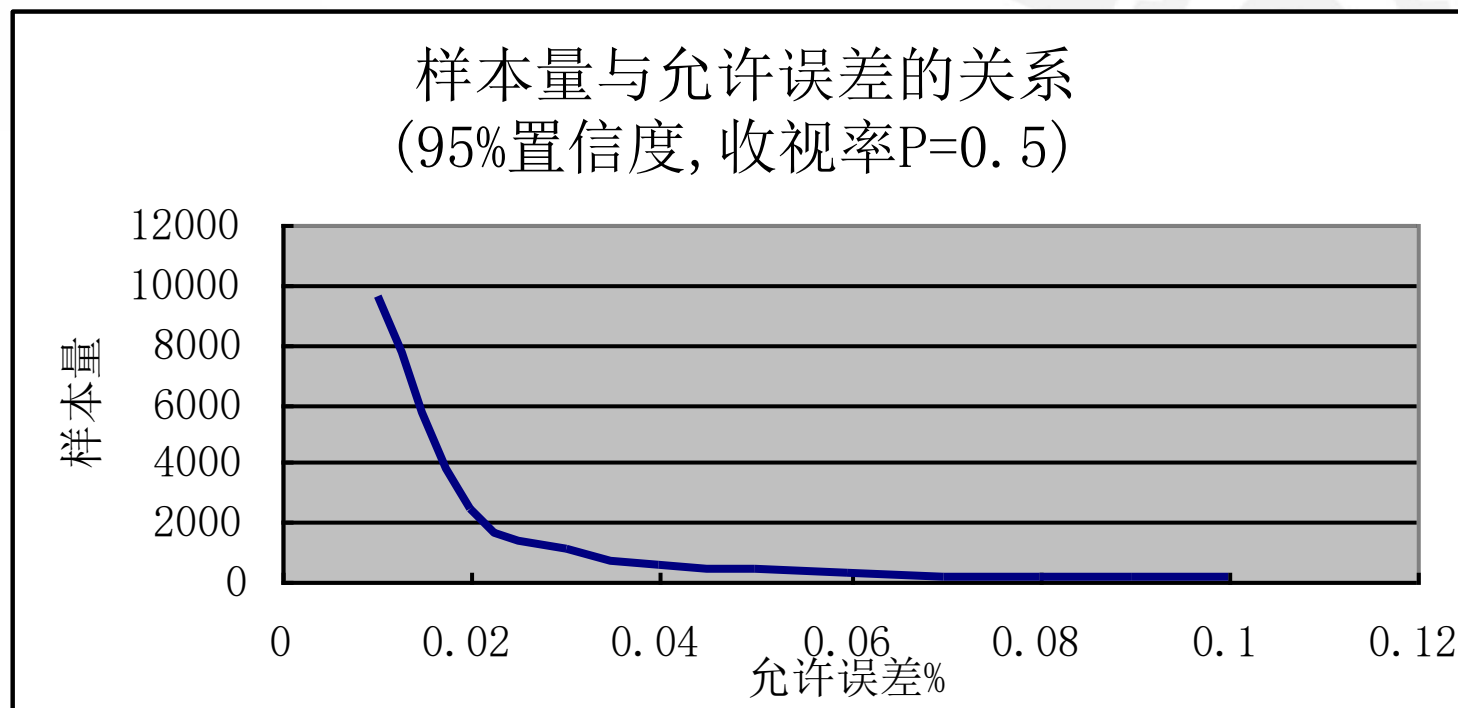
遥控器特殊操作，仪器  
调查

1秒钟

24小时，凌晨固定时间  
回传遥控器家庭成员键配合度低  
不能实时

- 样本量与误差的关系

允许误差	样本量
1%	9604
2%	2401
3%	1067
4%	600
5%	384
6%	267
7%	196
8%	150
9%	119
10%	96

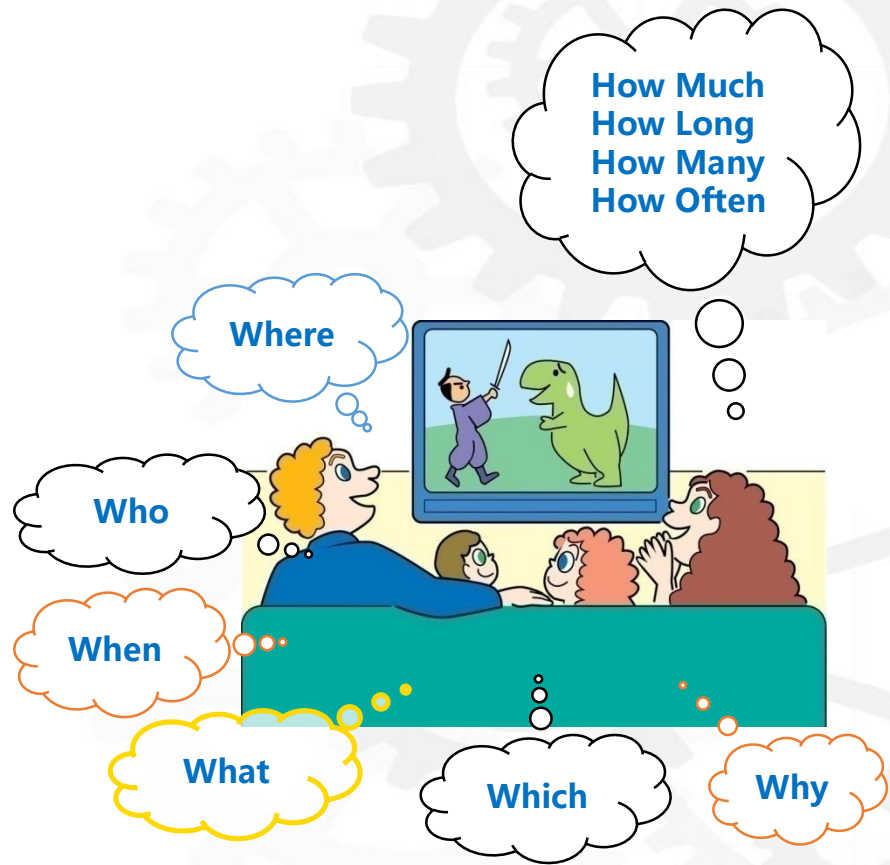


## 传统收视率调查方法 – 问题

- 抽样误差
  - 央视-索福瑞 (CSM) : 全国650个城市, 总样本户<5万户, 平均每城市不足百户, 只有北、上、广等少数城市的样本户达到500个。
- 样本户污染
  - 样本“污染”难以避免。样本户相对固定, 隐蔽性差;
  - 且日记法、测量仪要求受众参与性强, 可以被“收买”。
- 代表性
  - 样本更换及跟踪难度大, 要求人员固定, 只能是常驻特定居民。
- 数据单一
  - 受制于采集手段, 往往只有直播数据, 缺乏常见的时移、回看
- 只支持传统指标
- 时效性差; 遇到特殊情况, 需要人工修改数据

## 广电大数据 – 采集

- 用户行为数据
  - 谁在什么时间看了什么频道、节目、页面
- 用户特征信息
  - 用户是什么人（地区、年龄、性别、职业、学历、收入）
- 媒体资源信息
  - 什么渠道、在什么时间播了什么类型的节目
- 用户消费信息
  - 谁购买了什么服务
- 服务端业务信息
  - 谁在什么时间使用了什么服务
- 终端采集
  - 覆盖三网、多屏；全网数据采集





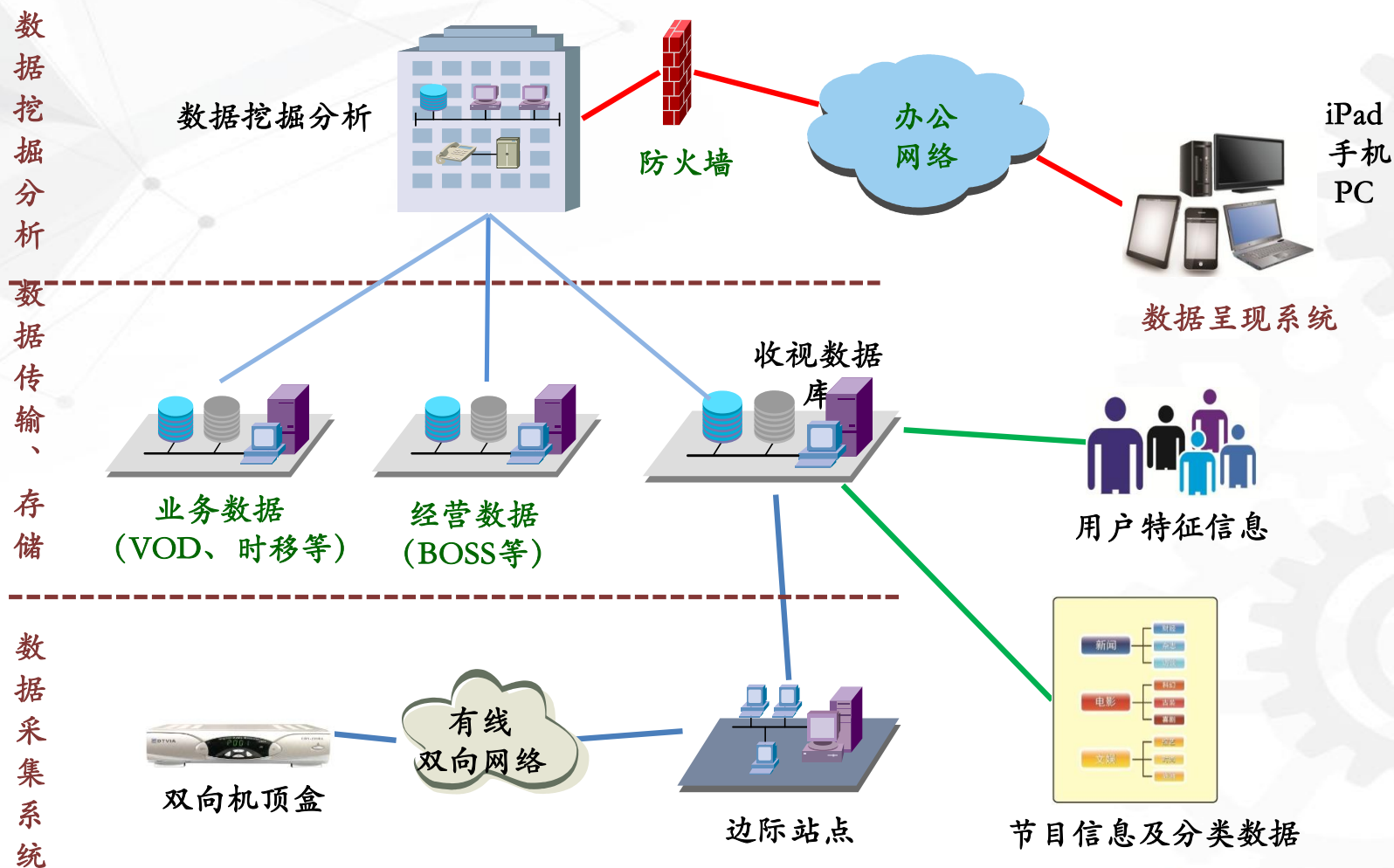
## 广电大数据 – 分析

- 直播
  - 收视时长、收视率、到达率、接触度、市场份额、观众忠诚度等。
  - 创新指标：节目相对吸引力、收视率分布等。
- 点播
  - VOD业务使用及VOD节目指标。
  - 各时段在线户数、在线率；到达户数、到达率、点播户数、点播率、收看时长、页面点击等。
  - 创新的竞争力指标：时间转化率、点击转化率。
  - 按栏目、按供应商分别分析。
- 时移、回看
  - 业务各时段在线户数、在线率、页面点击率等。
  - 各频道及节目的收视时长、收视率、到达户数、到达率、市场份额等。
- 广告业务
  - 按各广告位、广告包进行指标分析。
  - 各广告位曝光频次、n+曝光率、n+到达户数、有效曝光率、有效到达户数等。
- 其它增值业务
  - 业务各时段在线户数、在线率、页面点击率等。

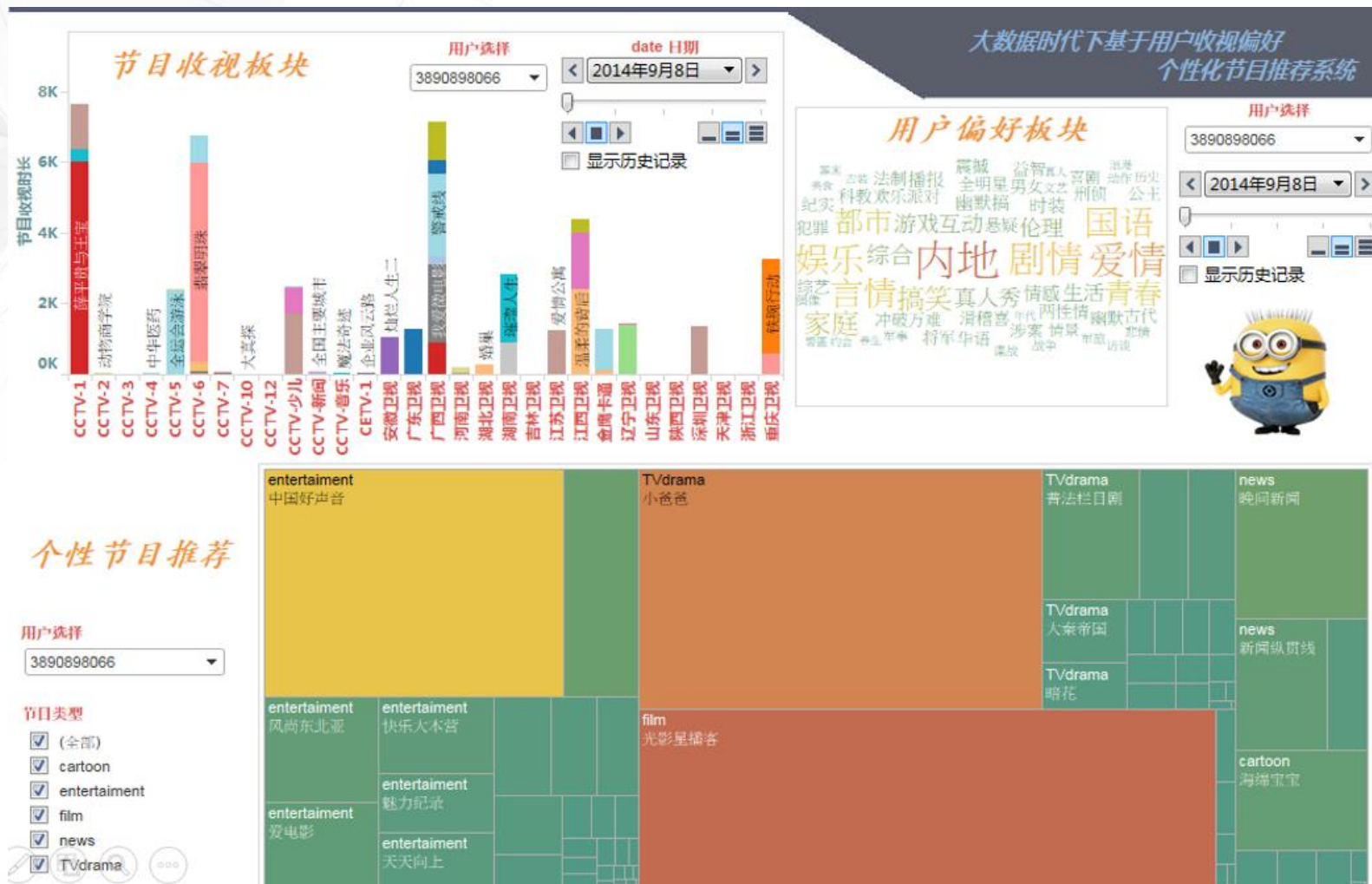
## 广电大数据 – 特点

- 数据准确、分析计算误差小；公正、抗污染、不易造假
  - 全网海量用户收视数据分析，全方位、无死角，尤其对弱势频道、非黄金时段节目的数据分析更准确，使其数据有意义，更能反映实际情况；
  - 全网双向用户可达千万以上，用户污染影响微乎其微；
- 指标更有价值
  - 例如忠诚度、竞争力指标等，对低收视率的频道和节目，能提供更多参考依据；
- 能提供舆情分析
  - 根据全网用户的收视行为，结合节目播出信息，可以分析舆情；
  - 涉及舆情及国家信息安全，不建议外资公司参与；
- 能了解每一个用户的偏好，提供个性化服务
- 技术难度大
  - 采用传统手段采集海量用户收视数据，成本太高；
  - 需掌握大数据处理系统的架构、算法、专业工具等核心技术；

## 广播电视大数据系统技术架构

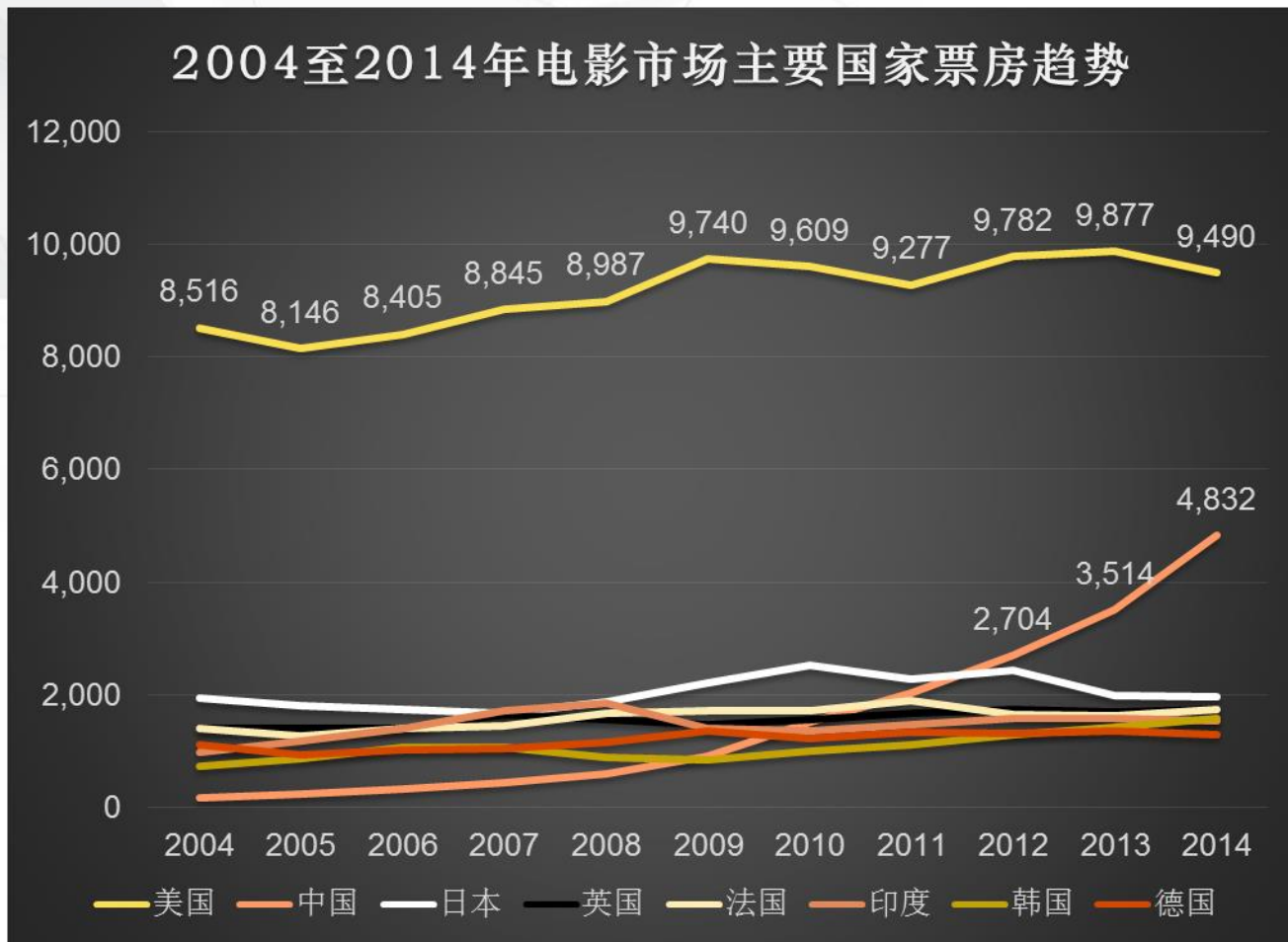


## 广播电视大数据系统技术应用

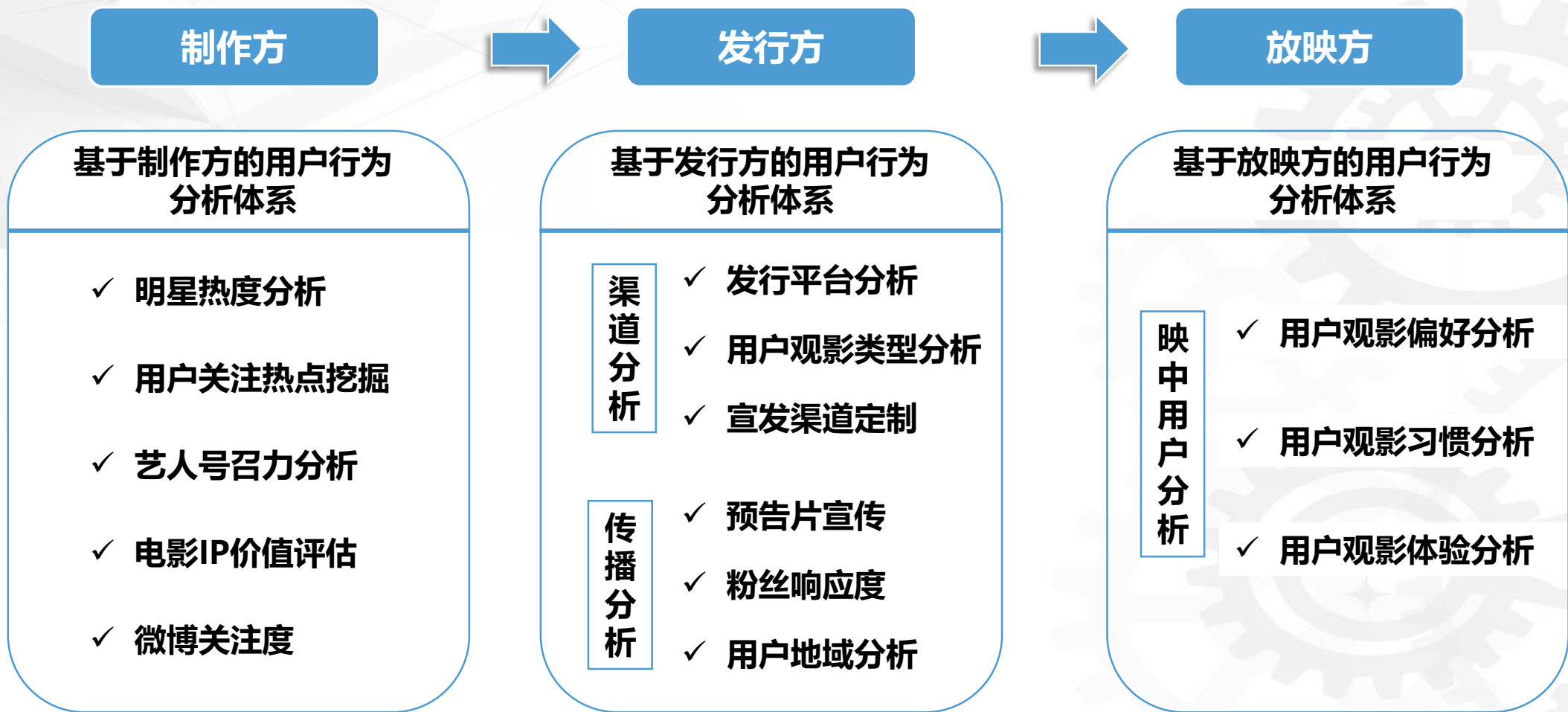




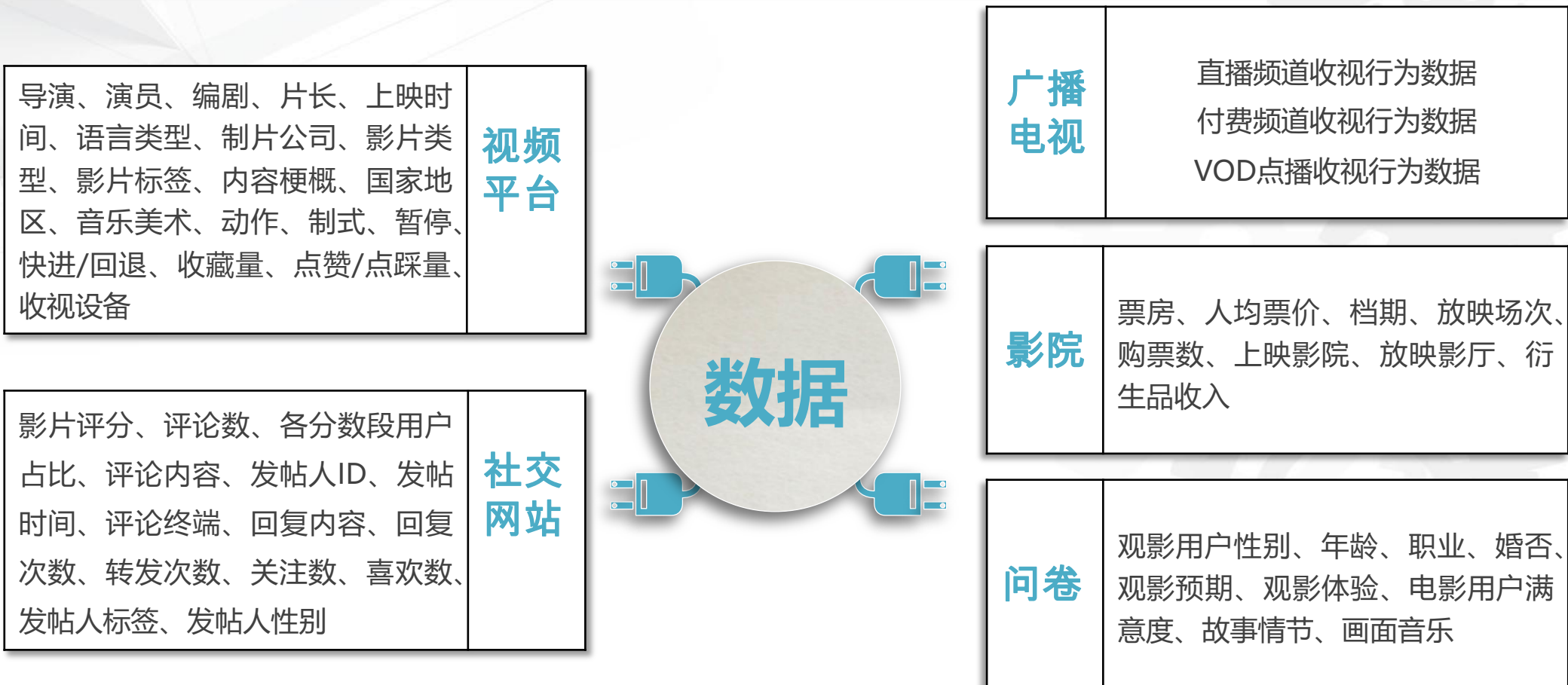
## 中外电影产业发展现状



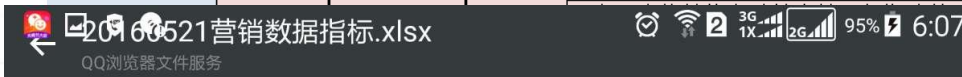
- ◆ 截止2015年底，**全年票房破440亿**，同比增速48.7%。
- ◆ 目前，**中国电影年产700部左右**，预估2020年每年电影产量将超过1200部。
- ◆ 据保守估计，全国电影行业**企业超8000家**，**从业人员超120万人**。
- ◆ **影院6300多家**，**屏幕超31000块**。
- ◆ **观影人次12.6亿**。
- ◆ 资本市场电影**投资储备金超千亿**。



## 中影指数数据支撑



大权项	大权比例	小权项	小权比例	选 项
结构	0.15	结构元素	0.2	a. 基本结构元素完整, 各元素比例安排合理
				b. 基本结构元素完整, 各元素比例安排不当
				c. 基本结构元素不完整, 各元素比例安排不当
				d. 基本结构元素不完整, 各元素比例杂乱无章
		故事核	0.5	a. 故事核具有一定的创新度, 异于市场同类
				b. 故事核普遍市场常见, 但完整不散乱。
		线索	0.3	c. 故事核有新奇点, 但散乱。
				d. 故事核散乱, 无新奇点。
				a. 故事主线清晰, 副线与主线相互作用。
				b. 故事主线清晰, 副线作用不明显, 对情节无影响。



20160521 Sheet1

	K	L	M	N	O	P	Q	R	S	T	U
1											
2											
3	060个视频	113张图片	3条新闻	536	631	89	2	6	41	12913042	532836
4	29个视频	256张图片	93条新闻	454	953	151	26	3	33	10325461	399593
5	616个视频	461张图片	55条新闻	162	431	128	0	1	29	3970816	178214
6	625个视频	181张图片	114条新闻	160	675	79	0	4	22	4137949	133513
7	321个视频	233张图片	30条新闻	155	612	29	0	9	89	2450678	70066
8	144个视频	933张图片	225条新闻	269	1009	283	28	4	57	6165094	286312
9	29个视频	267张图片	169条新闻	283	465	92	1	14	96	7434362	215316
10	50个视频	161张图片	2条新闻	262	2488	139	12	15	75	2869419	285348
11	26个视频	220张图片	88条新闻	431	620	38	2	11	93	4581209	140291
12	213个视频	145张图片	30条新闻					2	40	1300134	30108
13	16个视频	140张图片	17条新闻					3	60	826744	12200
14	216个视频	146张图片	42条新闻					6	91	2859971	85626
15	3个视频	10张图片	1条新闻					1	5	28143	341
16	0个视频	5张图片	0条新闻								
17	21个视频	51张图片	0条新闻								
18											
19	18个视频	84张图片	0条新闻					4	29	5276	18
20	9个视频	84张图片	14条新闻					3	36	35904	18
21	2个视频	12张图片	0条新闻					0	3	31804	16

电影营销分析

电影名称	电影英文名	电影拼音	更多片名	原著	导演	主演	IMDB评分	IMDB人数	时光网评分	时光网人数	豆瓣评分	豆瓣人数	电影类型	电影票房	剧情介绍	色彩
小时代4:灵魂尽头	Tiny Times 4.0	xiaoshidaisilingjintou	小时代4:灵魂尽头	小时代	郭敬明	杨幂 郭采洁 陈学冬 郭碧婷 谢依霖	3.1	144	3.3	23566	4.7	65513	爱情, 青春	48748万	《小时代4:灵魂尽头》作为影片《小时代》系列的终	彩色
评论影片	评论人	评论时间	支持人数	评论内容			score									
湄公河行动	Riverose	2016-10-01	439	彭于晏和张涵予拍合照的时候推荐			3.5									
湄公河行动	大島	2016-10-06	336	糯米脾气太差了, 吃不到早餐推荐			3.5									
湄公河行动	frozenmoon	2016-09-30	292	一部类型混乱的电影。唯一较差			3.5									
湄公河行动	PixelGirl	2016-10-04	352	就是恩求以后留条活狗 力荐			3.5									
湄公河行动	m89	2016-09-28	277	林超贤又拍了一部《逆战》, 较差			3.5									
湄公河行动	indicent time	2016-09-30	364	《毒战》完成警匪片审查破推荐			3.5									
湄公河行动	王大根	2016-10-03	246	3.5, 主旋律能拍成这样真的推荐			3.5									
湄公河行动	what'swilliam	2016-09-30	249	3.6星 1. 张涵予爷们力爆表 还行			3.5									
湄公河行动	月亮上的猫	2016-09-30	252	怪不得主旋律电影烂, 人物很差			3.5									
湄公河行动	隐	2016-09-25	262	林超贤这样能拍出硬货的导演推荐			3.5									
湄公河行动	车瑞茜	2016-10-04	262	当你坐在电脑面前安静上着推荐			3.5									
湄公河行动	小岩井	2016-10-01	181	娃娃兵让人揪心, 天杀的毒贩推荐			3.5									
湄公河行动	刘小黛	2016-09-29	157	彭于晏现场很会耍花枪。《毒战》推荐			3.5									
湄公河行动	小瞞皮	2016-09-30	227	很好的边境犯罪片。可圈可点推荐			3.5									
湄公河行动	奥蒂安斯	2016-10-02	144	向中国警犬致敬! 推荐			3.5									
湄公河行动	河狸	2016-09-30	187	国庆档4部大片里最实至名归推荐			3.5									
湄公河行动	锐利修蕊	2016-09-30	144	对彭于晏和赵健全程好感爆棚 还行			3.5									
湄公河行动	小心和小欣	2016-09-26	96	呵呵, 八一制片厂新导演林超贤较差			3.5									
湄公河行动	陆支羽	2016-09-26	149	3.5。1. 非常林超贤的金三角推荐			3.5									
湄公河行动	半导	2016-10-01	92	角色性格不是黑就是白, 注水较差			3.5									
湄公河行动	xyan	2016-10-03	130	看完仔细一想, 敌军的所有推荐			3.5									
湄公河行动	liverbird	2016-10-01	137	全程吓着看完脖子疼...而淋推荐			3.5									

电影用户行为分析





日票房 2016年10月9日 ▾

前一天  
今日实时 **1473.0** 万  
后一天

北京时间 09:57:48 ⓘ

票房排名

指标自定义 &gt;

影片	实时票房 (万元)	票房占比	排片占比	上座率
湄公河行动 上映10天 6.31亿	<b>738.72</b>	50.1%	31.6%	3.4%
从你的全世界 路过 上映11天 6.34亿	<b>338.78</b>	23.0%	26.0%	2.4%
爵迹 上映10天 3.52亿	<b>211.23</b>	14.3%	16.0%	2.6%



猫眼评分

想看

9.3(305816人评分)

57407

类型: 动作,犯罪,剧情

时长: 124分钟

制式: 2D/中国巨幕

上映日期: 2016-09-30大陆上映

总票房:63104万元

首周票房:16529万元

点映票房:2万元



猫眼评分

想看

9.3(305816人评分)

57407

类型: 动作,犯罪,剧情

时长: 124分钟

制式: 2D/中国巨幕

上映日期: 2016-09-30大陆上映

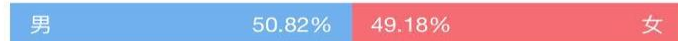
总票房:63104万元

首周票房:16529万元

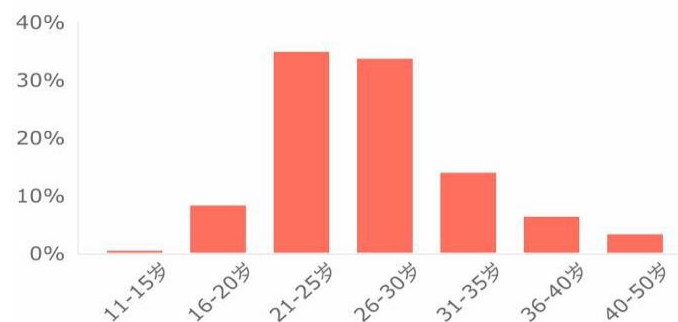
点映票房:2万元

票房 日排片 城市 影投 受众分析 想看指数

性别分布



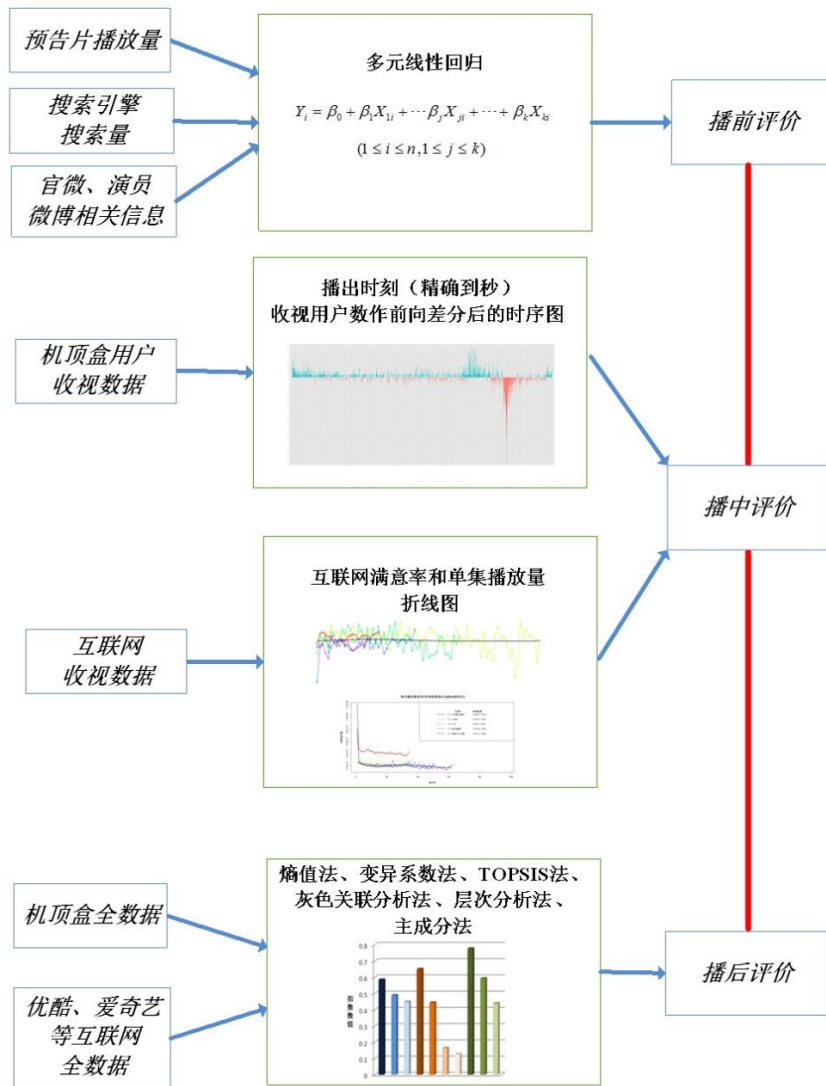
年龄分布



票房 日排片 城市 影投 受众分析 想看指数

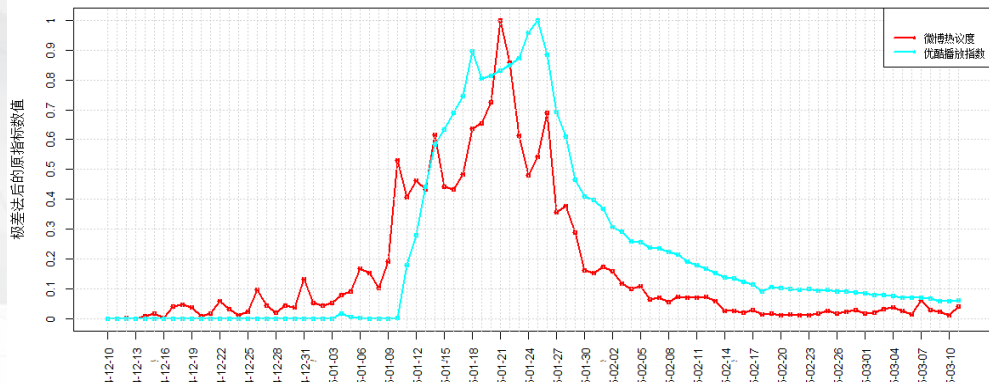
想看趋势



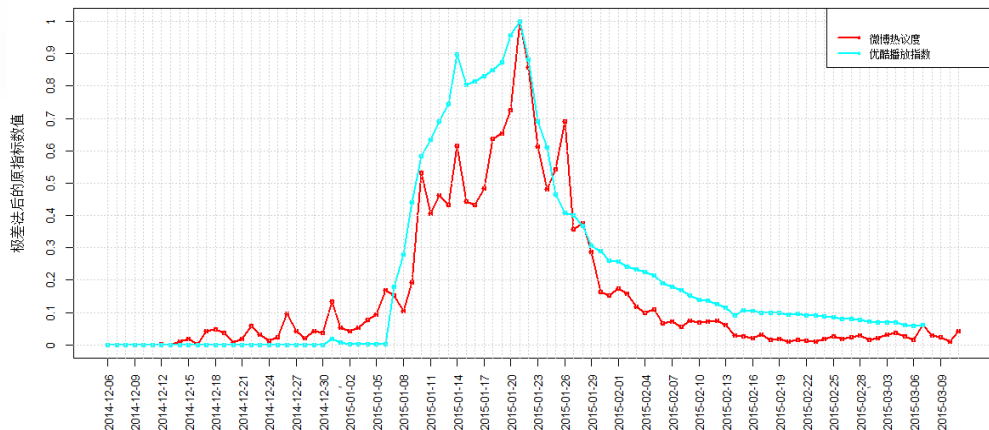


# 跨平台电视剧 播前播中播后评价 模型

《何以笙箫默》微博热议度&amp;优酷播放指数——时间序列对比(视频网站指数不平移)

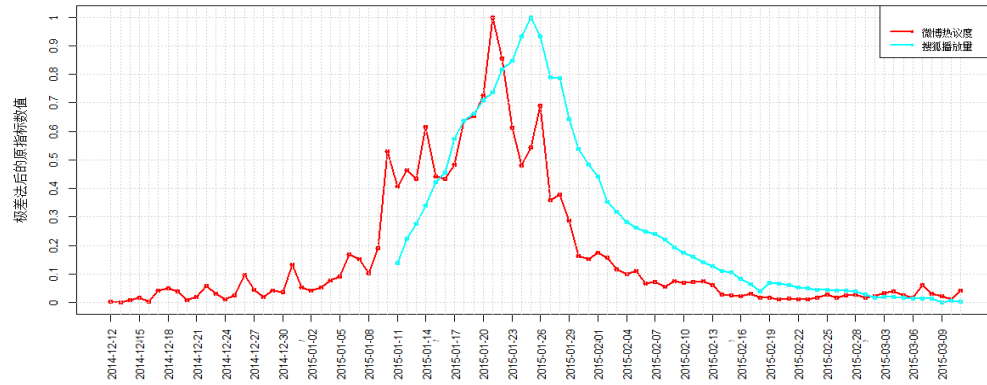


《何以笙箫默》微博热议度&amp;优酷播放指数——时间序列对比(视频网站指数向左平移最高点之间差值)

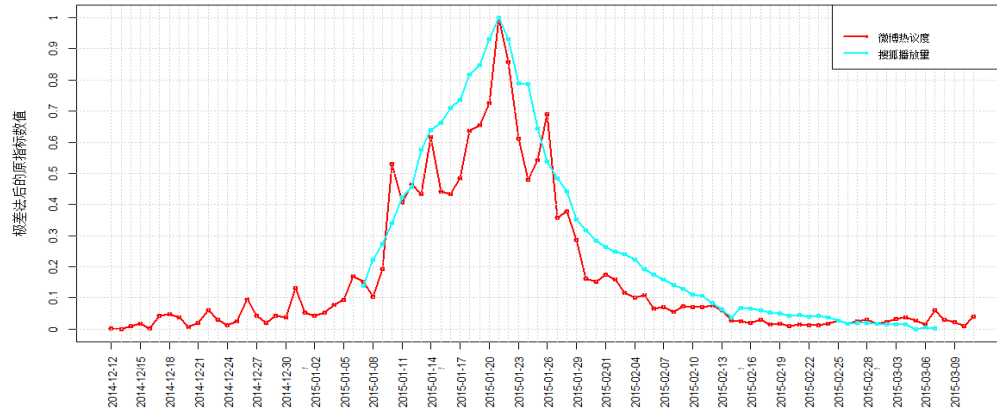


播前评价——微博热议度与优酷播放指数的关系

《何以笙箫默》微博热议度&amp;搜狐播放量——时间序列对比(视频网站指数不平移)

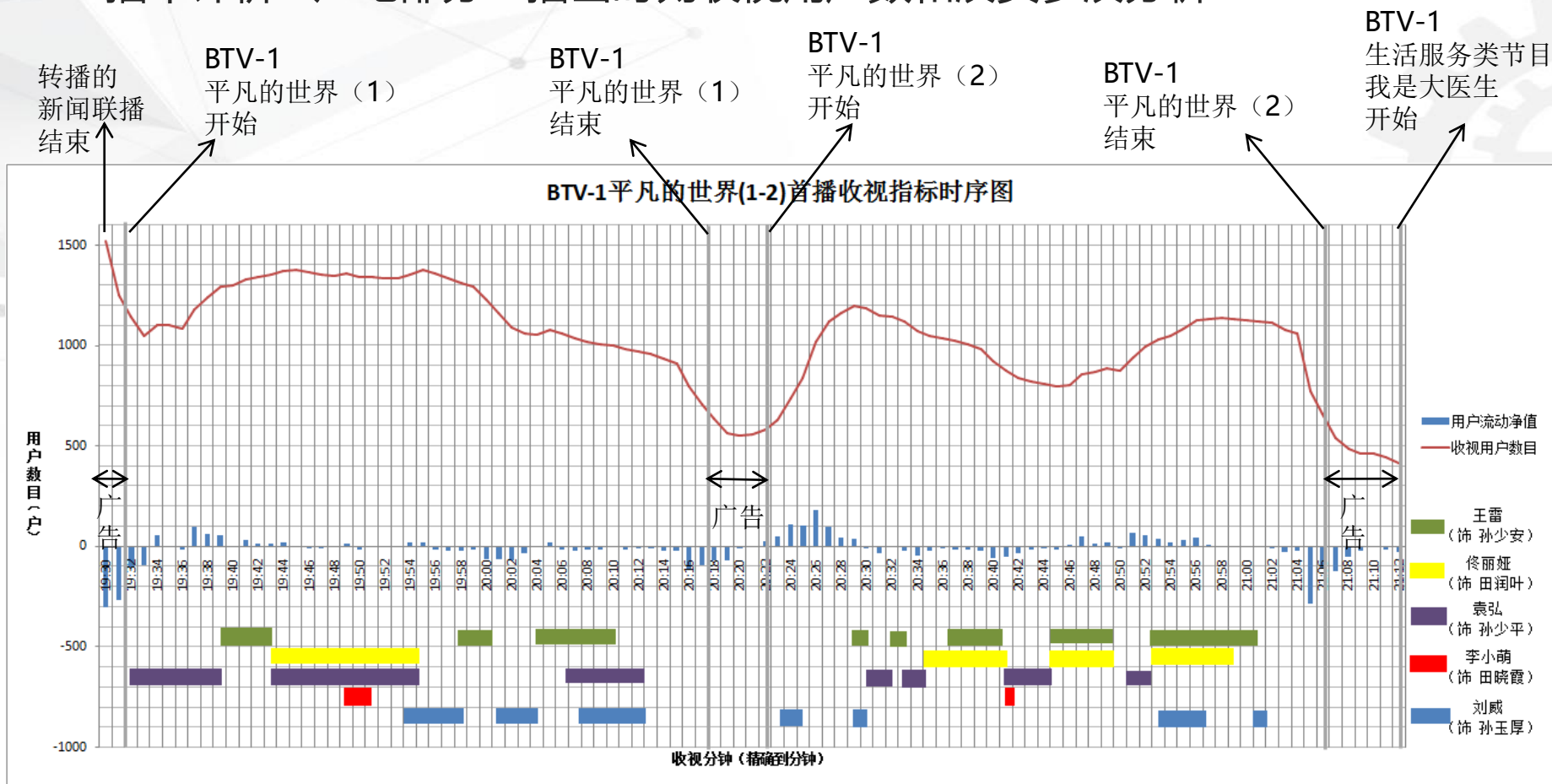


《何以笙箫默》微博热议度&amp;搜狐播放量——时间序列对比(视频网站指数向左平移最高点之间差值)



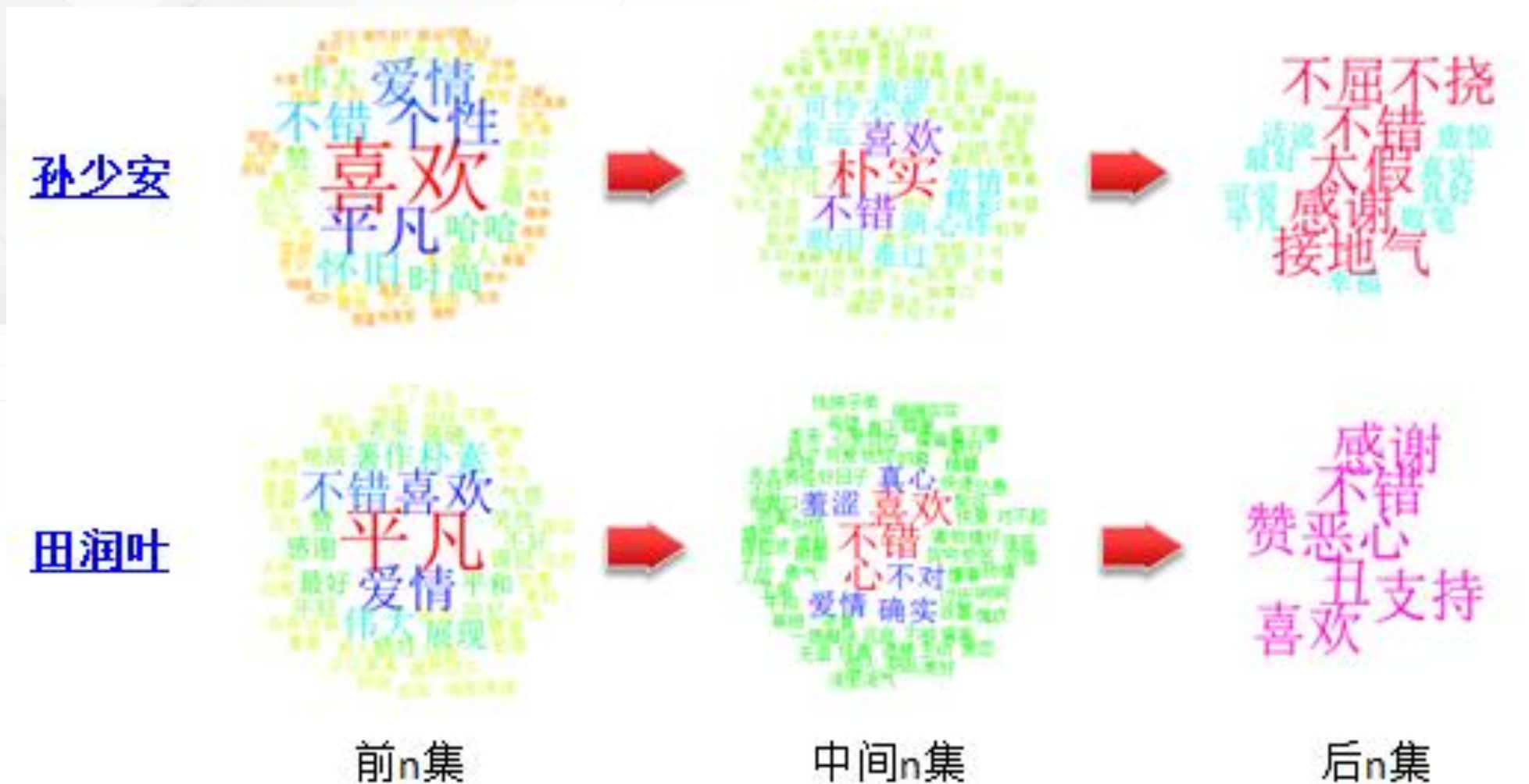
播前评价——微博热议度与搜狐播放量的关系

## 播中评价—广电部分：播出时刻收视用户数和演员参演分析



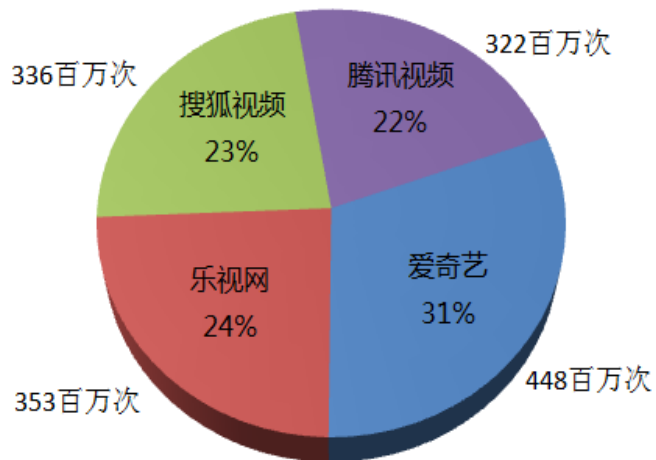
2015-02-26 BTV-1平凡的世界 (1-2) 首播收视用户数及演员参演情况时序图



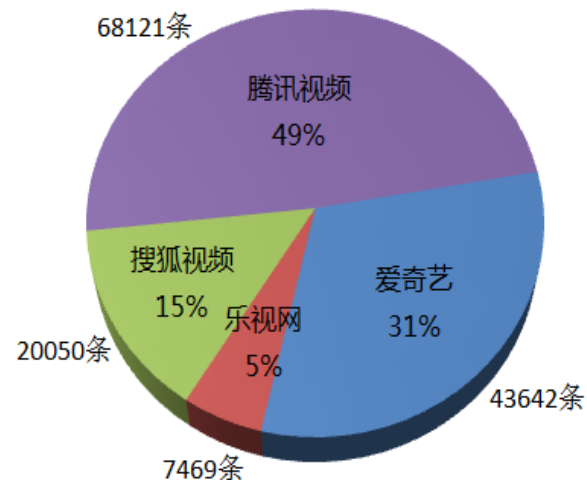


电视剧《平凡的世界》主要角色的词云随剧集的变化趋势图

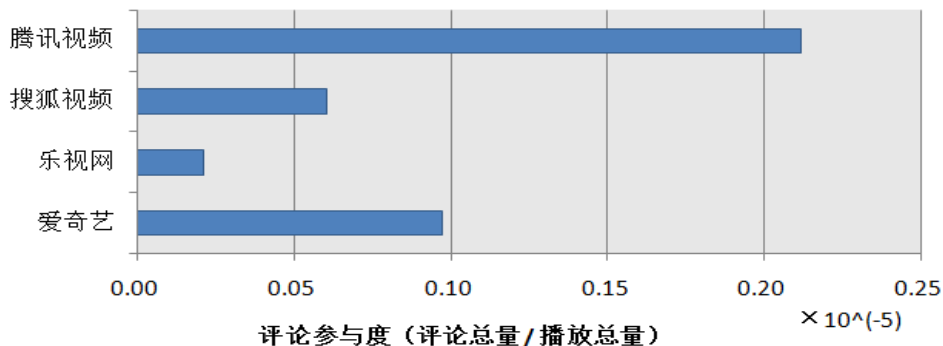
电视剧《平凡的世界》在不同视频网站的播放总量比较



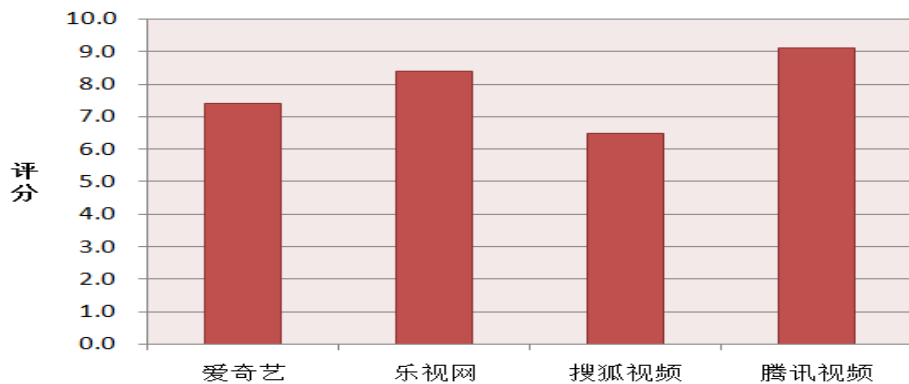
电视剧《平凡的世界》在不同视频网站的评论总量比较



电视剧《平凡的世界》在不同视频网站的评论参与度比较



电视剧《平凡的世界》在不同视频网站的评分比较



电视剧《平凡的世界》在不同视频网站的播放情况比较



2015年 5月2日，黑龙江庆安县火车站候车大厅发生枪击事件，一名曾多次上访的男子徐纯合被工作人员拦截在进站口， 争执中涉嫌暴力袭警，被执勤民警开枪击毙，引发民警用枪是否合理等一系列争议。

随后，有媒体通过调取监控录像，试图还原当时现场情况。但完整视频直到事发十多天后才公开，这期间不免引发公众的种种想象和猜测，让枪击事件更加疑云重重。视频公开后，人们对视频的真实性又产生怀疑，而警察是否应该开枪等话题热度适中持续。

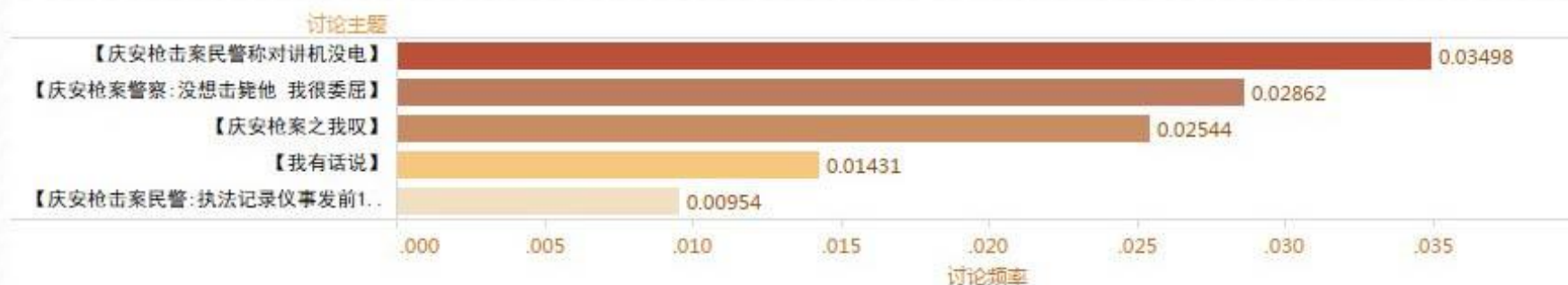
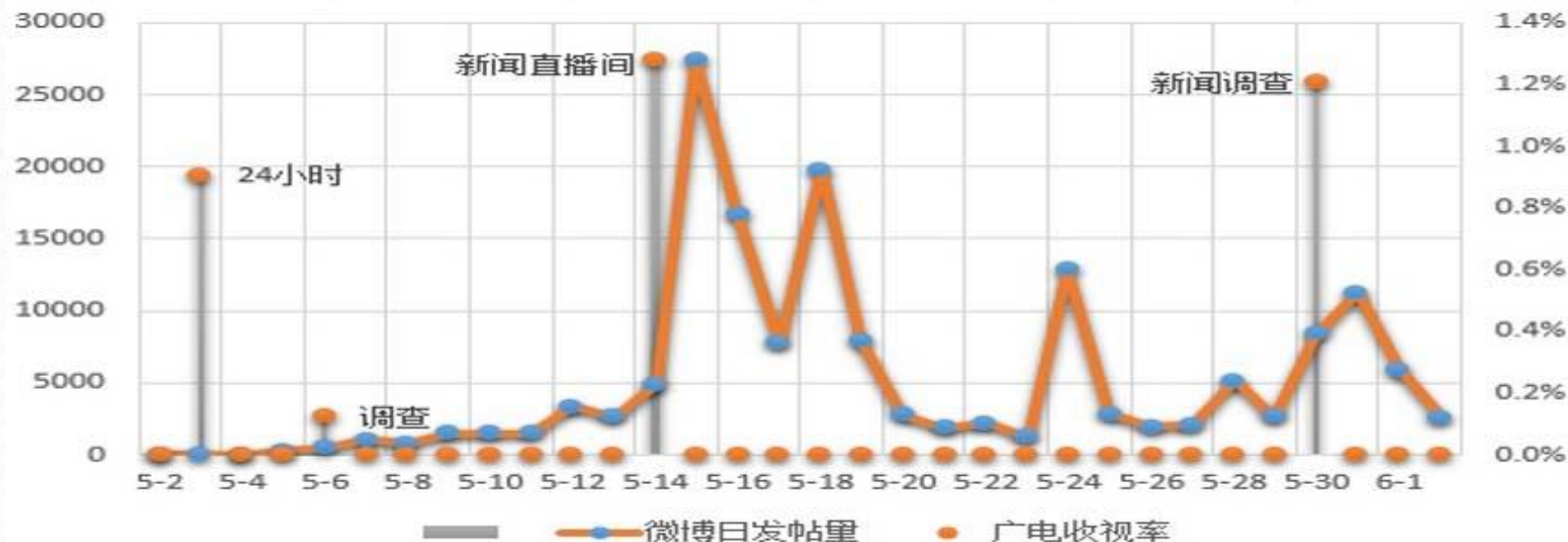






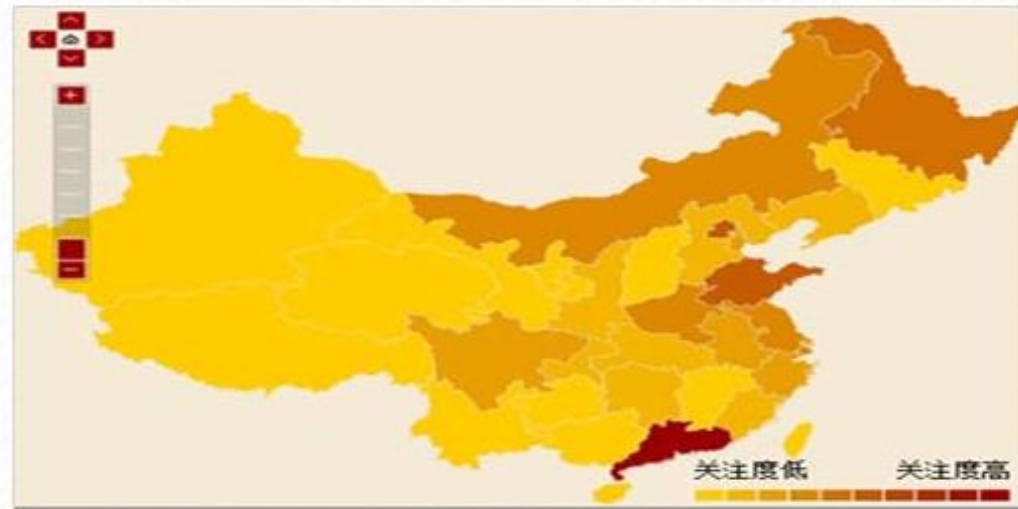
事发后一个月的时间里，庆安事件在微博上持续热议。虽然事情起源于2日，但5月14日之前，微博舆情始终处于缓慢发酵状态，直到15日，微博转发量出现最高爆发点，后面又出现三次高点。通过调查，我们发现这四个时间点与主流媒体报道相关。主流媒体在突发事件的话语构建中，仍占有主导地位。

14日前，之所以网民揣测、夸大事实的言论较多，因为还原真相的现场监控视频一直没有公布。14日，央视《新闻直播间》公布了监控录像。18日，央视《新闻周刊》中，白岩松说：“庆安枪击案不该就此终结。”24日，词频的焦点集中在“白岩松的评论”和“视频是否造假”。30日，央视《新闻调查》专访了开枪民警，出现第四个微博舆情高潮。





# 新闻大数据



群众B



群众C



DTCC

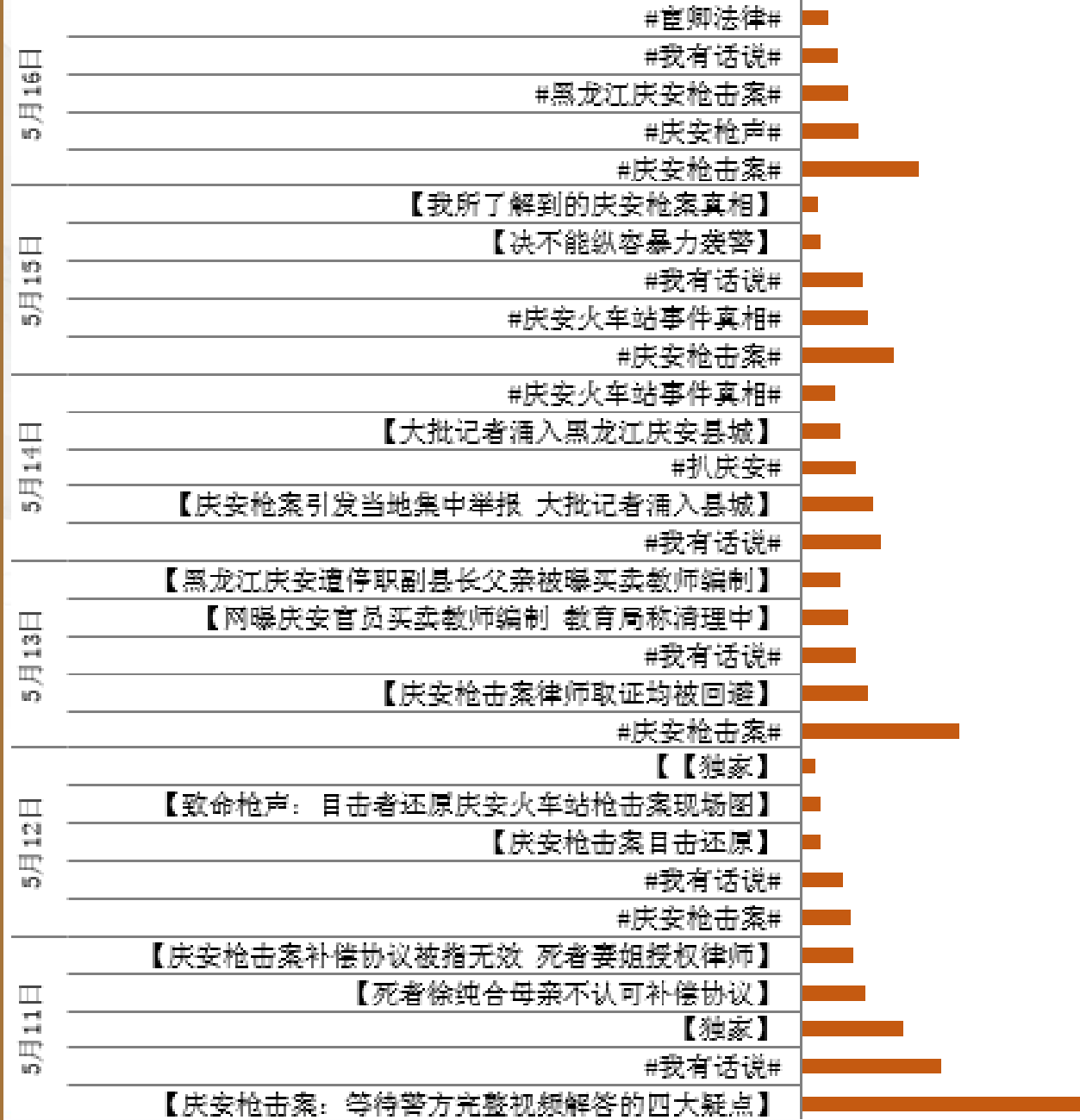
# 2017年第八届中国数据库技术大会

## DATABASE TECHNOLOGY CONFERENCE CHINA 2017

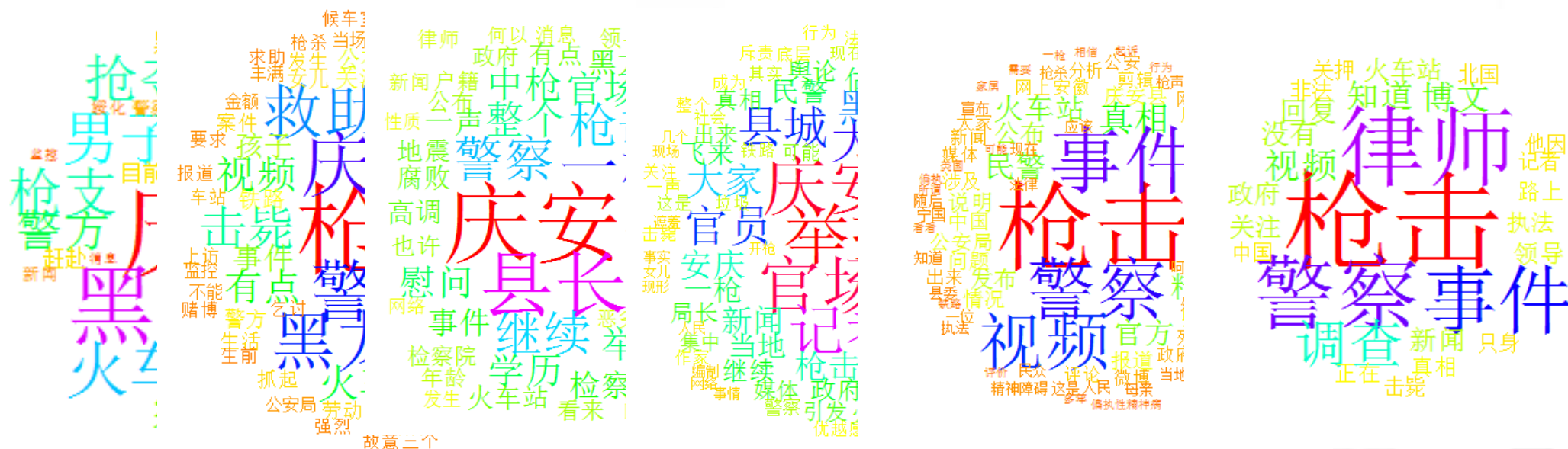
SequeMedia  
 盛拓传媒

# 大数据案例剖析

## 新闻大数据



# 新闻大数据



5月  
2日

5月  
5日

5月  
12日

5月  
14日

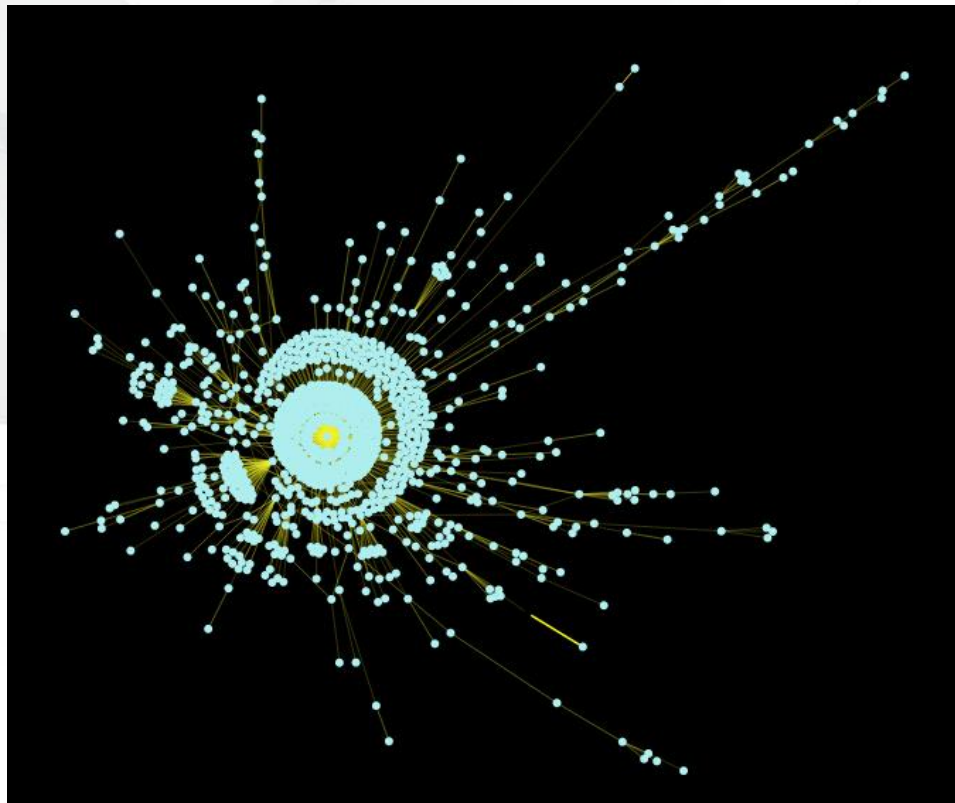
5月  
16日

5月  
30日



# 大数据案例剖析

## 新闻大数据



头条新闻  
5月30日



新浪视频  
5月30日







北京鑫青年 (● ^ o ^ ●) 

北京 西城



扫一扫上面的二维码图案，加我微信

# THANKS