

微博舆论：“清华学姐”事件的发酵

第二小组

组长：陈婧涵

组员：刘凯瑞、郁博文、王佳鹏、周吴越、江澜泉韵

分工：

1. 陈婧涵：负责微博数据的爬取与采集、工作分配安排。
2. 刘凯瑞：负责对爬取结果进行分词并统计词频。
3. 郁博文：负责词云制作。
4. 王佳鹏：负责表格处理和绘制曲线图。
5. 周吴越：负责分析，Word 的汇总与编写。
6. 江澜泉韵：负责制作结课展示的 PPT。

一. 事件回顾：

17 日在清华食堂，事件发生，女生在朋友圈和群里发了男生的个人信息和误解攻击，当晚删除朋友圈，并说明未有明确结论，18 日查监控发现是误会，女生在朋友圈、院系群和树洞等进行澄清，线下联系到男生并道歉。之后，女生在传播过之前的信息的各个平台发布道歉文章，澄清事实并取得了男生的谅解。

事件网络发酵阶段：

1. 知乎出现匿名帖，冒充男生的名义，号称想自杀写遗书，觉得人生被毁了。然后在微博和营销号上广泛传播。舆情主要集中在不当的挂人至其社会性死亡、性骚扰事件中男性与女性。
2. 营销号开始传播事件女生的个人信息和照片，网上开始出现外貌羞辱和反社会性死亡（反而使学姐社会性死亡）
3. 另一种声音开始出现，种种证据发现这件事情的传播后面有资本在消费舆论，有人拿着学姐的私人信息找各个营销号曝光。清华内部人员开始澄清此事件已经在校内解决。
4. 余波：JM 帝国（侮辱女性的漫画）开始卷入，出现舆论认为，这种侮辱女性的漫画的作者不被举报和曝光，对比被人肉的清华学姐，是微博舆论环境恶臭的体现。又出现舆论认为，可以不举报这个漫画，因为也可以创作侮辱男性的漫画。此后，舆论场转向了小说创作和出版自由。

研究价值：

事情影响和传播非常广泛，在微博热搜第一挂了一整天，“清华”“性骚扰”“社会性死亡”其中涉及的诸多关键词都是微博特别容易上升和发酵的主题。令人值得注意的是，这场事件原本只发生于当事人朋友圈和清华树洞，在当事人互相道歉之后再次被挂到微博上引起传播，属于一种间接的传播，两位当事人并无此意。

而舆论的内容也朝着不同主题和矛盾进行着，难以界定的性骚扰和长期缺乏的举报方式、社会性死亡、借助舆论被舆论反噬、对清华学姐的外貌羞辱……

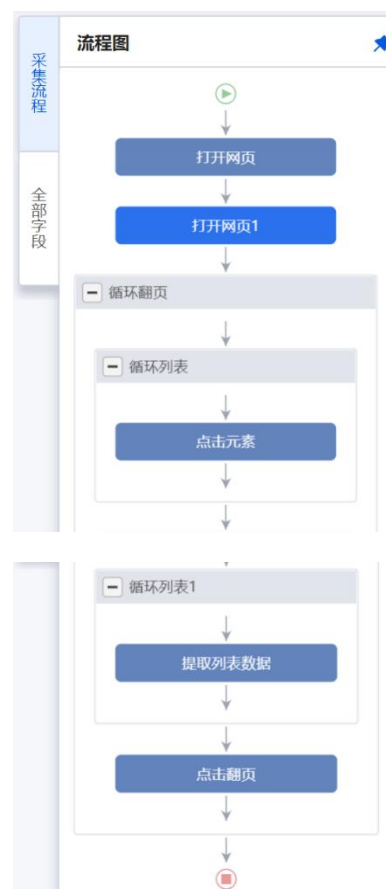
让人非常惊奇的是，这场事件的余波在十天之后仍然没有平息，甚至朝着各种意想不到的地方发展，我们的词云中出现了许多匪夷所思的词语：“法西斯”“美院”“文艺创作”“JM”“女房东”……

综上所述，这场事件传播广，影响大，时间长，舆论风向变化多端，讨论的主题非常丰富，汇集了舆论场上几大主要矛盾，具有代表性和研究价值。

二．微博数据采集

采用软件：八爪鱼

采集流程图：如下



具体过程：

利用微博高级搜索，搜索 11 月 19 日至 11 月 27 日的有关“清华学姐”的热门微博数和总微博数（由于微博搜索的翻阅页数最多 50 页，所以总微博数的数据不全面），再循环采集每条微博的 ID、内容、发布时间、转发、评论、点赞数，导出制成 excel 表格。

步骤一、打开网页

步骤二、微博登录

步骤三、在【高级搜索】中选择搜索条件

步骤四、点击【展开全文】，展开微博全文

步骤五、使用自动识别，自动识别列表页和翻页

步骤六、优化字段

步骤七、设置翻页次数

步骤八、启动采集

过程截图：

Sina Visitor System新

微博搜索

提取到的数据

#	ID	内容	时间	转发	评论	点赞
1	梦醒惊起梦中人	#清华学姐#怎么说呢...	11月20日 23:47			
2	迁/sss	//@文明大兴清华者...	11月20日 23:47			
3	lita爱吃抹茶芝士蛋糕	女性真的好惨//@晚...	11月20日 23:47			
4	修二轮车了的	//@笑着七千万剩男...	11月20日 23:47			
5	桃色山峦	#清华学姐#爱情#...	11月20日 23:47			
6	非山到海	哈哈哈哈哈了分享...	11月20日 23:46			
7	高草草小公举	在车上不小心碰到别...	11月20日 23:46			
8	淮南郭白衣	我上个厕所，你让我...	11月20日 23:46			

< 1 ... 7 8 9 >

已采集88条 已用时: 8分钟25秒 平均速度: 10条/分钟

停止采集

Sina Visitor System新

微博搜索

提取到的数据

#	ID	内容	时间	转发	评论	点赞
1	是你永远得不到的人...	事情原本在18号晚...	11月20日 23:11		1	
2	抓药超人小李	#清华美院老师回应...	11月20日 23:11			22
3	今天也是一颗开心的...	//@Aqzfp//@蓝蓝的...	11月20日 23:10			
4	京原二少	我参与了@中国新闻...	11月20日 23:10			
5	金小毛今天健身了吗	#清华学姐# 赶紧吧...	11月20日 23:10			2
6	一剑之距	她身体面，那就让她...	11月20日 23:10	3	4	34

< 1 ... 29 30 31 >

已采集306条 已用时: 20分钟43秒 平均速度: 15条/分钟

停止采集



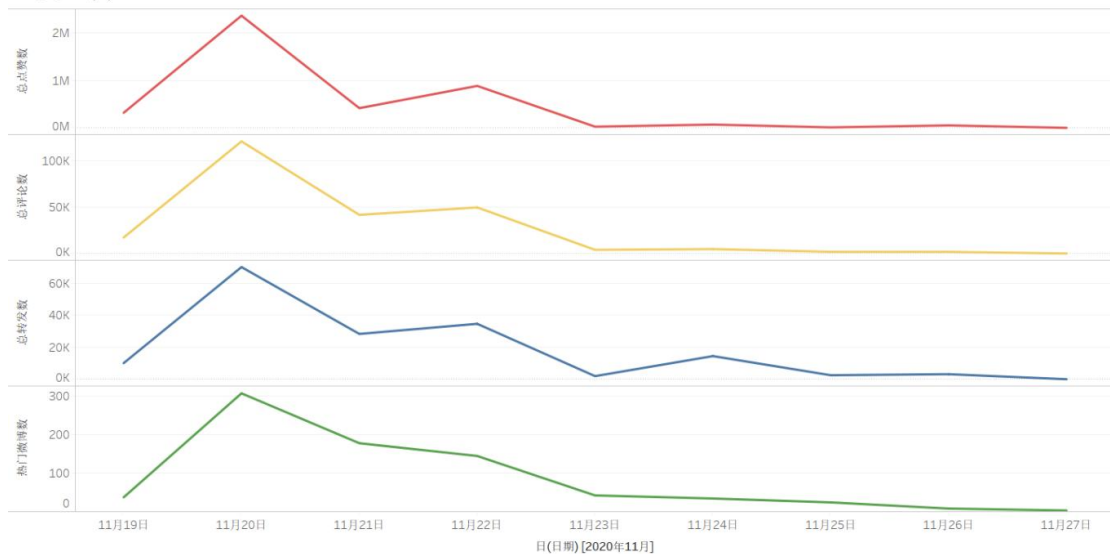
三、微博数量统计可视化

1. 对微博数据采集导出的 Excel 表格进行一部分优化处理,包括将单元格格式设置为数值格式,增加“总计”工作表等。在总计工作表对数据进行分组求和,便于接下来的可视化。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	日期	热门微博数	总转发数	总评论数	总点赞数																
2	11月19日	37	10124	17375	317798																
3	11月20日	306	70169	121319	2363358																
4	11月21日	177	28380	41914	415859																
5	11月22日	144	34708	49934	884580																
6	11月23日	42	2004	4111	25918																
7	11月24日	34	14536	4907	70648																
8	11月25日	24	2570	1951	10767																
9	11月26日	8	3219	1955	50964																
10	11月27日	3	83	116	1271																
11																					
12																					
13																					
14																					
15																					
16																					
17																					
18																					
19																					
20																					
21																					
22																					
23																					
24																					
25																					
26																					
27																					
28																					
29																					

2. 通过老师上课所教授的 Tableau 帮助绘制可视化图表。以日期为横轴,热门微博数,总转发数,总评论数,总点赞数分别为纵轴绘图,并使用不同颜色区分。绘制成功后将数据导出为图片格式。

工作表 1 (2)



日期天的总点赞数总和, 总评论数总和, 总转发数总和与热门微博数总和的趋势。

四、分析词频

分词这里我们使用的是 python 的 jieba 库, jieba 库是一个优秀的中文分词第三方库 (需另外安装), 调用其中的函数可以将一句话或一篇文章分成一个个词语, 以便后续的词频分析。词频分析阶段, 我们在 python 中写了一段代码来对刚刚分好的词进行统计并输出。在这里, 为了更好的处理数据, 反应数据的真实情况, 我们人工在 jieba 库里添加了新词“社会性死亡”, 并将“清华大学”与“清华”, “女孩子”与“女生”等同义词组合并成一个词, 同时我们还排除了“链接”, “知乎”, “网页”等干扰词汇。我们统计了 2020 年 11 月 19 日到 2020 年 11 月 27 号的数据, 得出每一天的词频和总体这九天的词频。然后我们绘制了相应的词云以更直观地显示出各个词出现的频次, 下面是代码的展示和统计结果的展示。

1. 方法:

```

1 import jieba
2 txt = open("20201127-.txt", "rb").read()
3 jieba.add_word("社会性死亡", frequency=1)
4 excludes = ["链接", "知乎", "网页", "这个", "知乎", "23", "11", "20", "22", "19", "21", "27", "事件", "##"]
5 words = jieba.lcut(txt)
6 counts = {}
7 for word in words:
8     if len(word) == 1:
9         continue
10    elif word == "女孩子":
11        word = "女生"
12    elif word == "别人":
13        word = "某位"
14    elif word == "je":
15        word = "je"
16    elif word == "20":
17        word = "20"
18    elif word == "清华大学":
19        word = "清华"
20    else:
21        counts[word] = counts.get(word, 0) + 1
22 for word in excludes:
23     del counts[word]
24 items = list(counts.items())
25 items.sort(key=lambda x: x[1], reverse=True)
26 for i in range(22):
27     word, count = items[i]
28     print("<{0}>{1}>".format(word, count))
  
```

词云制作：

采用工具：凡科快图

具体过程：利用小组其他组员做出的词频统计，将关键词按天分别做成词云。根据词频决定关键词的数量、大小、位置、颜色。词频较高的关键词会显示的较为显眼。

具体步骤：

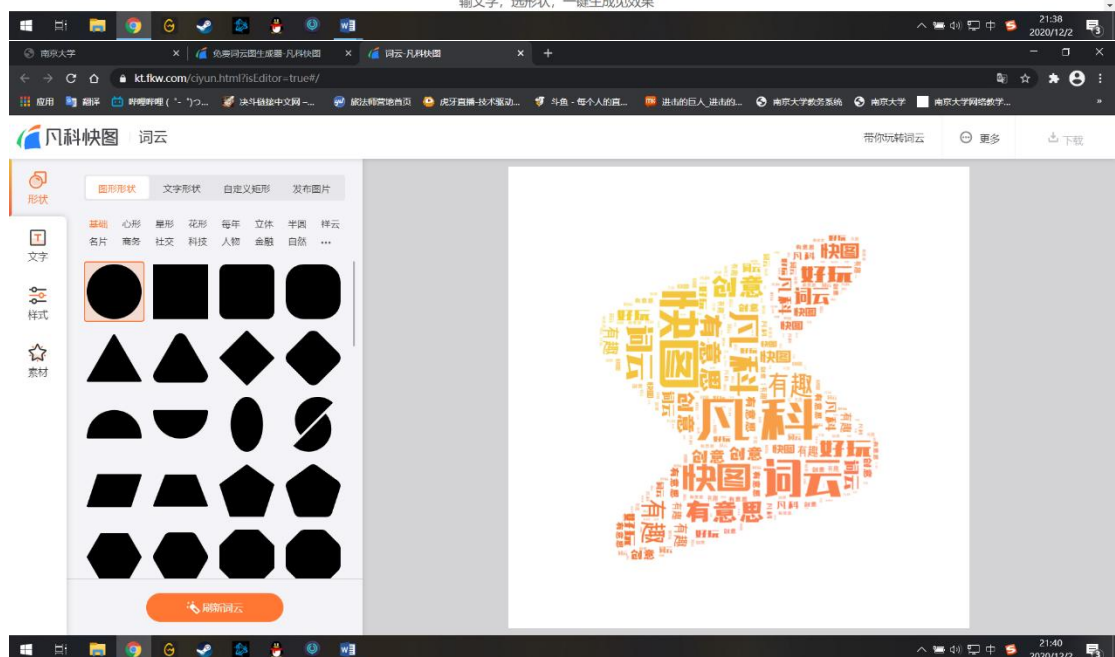
- 一、打开凡科快图
- 二、将关键词分别录入
- 三、根据词频，调整关键词的出现频率、颜色、大小、位置
- 四、选择合适的词云形状。
- 五、点击刷新词云并保存。
- 六、过

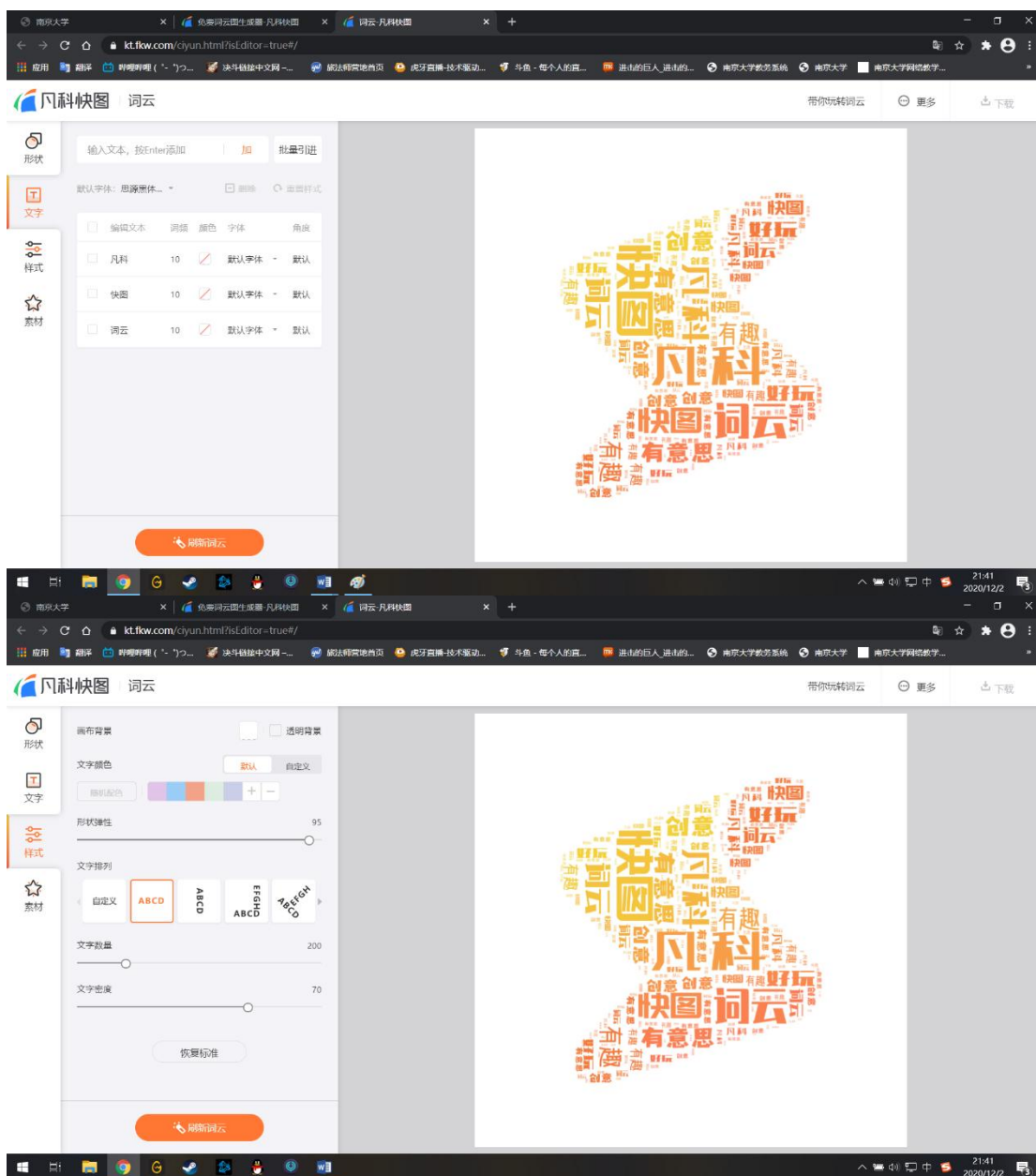
程 截 图



简单四步，生成词云图

输文字，选形状，一键生成效果





2. 相关评论总词频

整体的（经过简单的筛选，去掉了“那些”，“他们”等没有实际含义的词）



学姐	4452
清华	4402
自己	1078
没有	1027
女性	842
不是	810
性骚扰	791
学弟	688
JM	616
道歉	544
女生	528
如果	466
不要	423
男性	422
女人	408
看待	372
真的	370
以后	331
保护	320
时候	319
社会	308
社会性死亡	304
希望	294

误会	276
事情	270
中国	269
监控	262
法西斯	261
正义	249
信息	248
我们	242
性侵犯	238
问题	237
这种	234
性别	233
女权	232
威胁	230
诬陷	227
造谣	227
朋友	227
网暴	218
别人	207
人肉	207
租客	207
回应	206
女房东	206
应该	205
太原市	200
男生	198
美院	195
发声	193
当事人	192
舆论	179
老师	179
曝光	176
骚扰	174
支持	169
感谢	169
仇女	169
朋友圈	167
转发	160
大家	158
一起	155
认为	155
本质	155

3. 每日词频
2020.11.19:



学姐	359
清华	238
学弟	156
道歉	78
性骚扰	75
信息	65
监控	65
公开	54
自己	52
一个	51
没有	48
误会	48
身份	40
行为	40
证实	39
是否	38
女性	37
社交	37
媒体	36
欠妥	36

2020.11.20:



清华	656
学姐	651
女性	202
学弟	156
以后	139
自己	129
道歉	127
不是	95
还有	94
那么	83
女生	80
性骚扰	79
没有	78
希望	77
好事	74
误会	72
就是	71
如果	69
这样	68
事情	67

2020.11.21:



清华	389
学姐	385
学弟	108
一个	80
信息	78
自己	73
老师	72
全文	66
如何	65
收起	64
美院	61
女生	59
回应	58
道歉	57
表示	53
性骚扰	53
女性	52
监控	52
和解	50
潇湘晨报	49

2020.11.22:



清华	1042
学姐	917
自己	317
性骚扰	282
没有	232
性侵犯	231
你们	200
女性	191
不要	174
朋友	166
已经	154
诬陷	146
以后	135
男性	133
出来	127
社会	126
他们	124
支持	124
当事人	122
不是	122

2020.11.23:



学姐	286
清华	282
法西斯	247
没有	126
评论	115
自己	115
不是	112
一个	111
绝对	97
自由	95
可不是	94
不至于	93
嘴脸	93
人类	93
非法出版	93
文艺创作	92
存在	92
客观	92
何况	91
作品	91

2020.11.24:



学姐	518
清华	510
24	470
JM	313
没有	212
不是	190
什么	152
知道	149
威胁	137
如果	126
那么	111
所以	105
造谣	102
豆瓣	97
本质	96
鸣不平	95
恐怖主义	93
没人会	92
仇女	92
那生	91

2020.11.25:

2020.11.26:

学姐	556
清华	552
没有	205
可以	196
性骚扰	165
那些	151
女性	138
社会性死亡	111
一个	109
他们	108
不是	107
就是	101
特权	99
时候	93
这样	93
正义	90
还有	89
环境	86
不会	85
女生	85



现在，有证据表明，此次清华学姐事件，极有可能是场第三方有意为之的营销狂欢、流量狂欢、猎巫狂欢，学姐和学弟都被利用了。

图一，时间线没错的话，17日18日这两天，此事已经完结，曝光是pyq，是小范围的，澄清连发五篇，pyq、树洞各个平台都做了澄清，还有学姐、学弟、辅导员三方共同在场进行道歉，不说完全覆盖到，但肯定覆盖到pyq所能传播的绝大范围了吧。

图二，证实18日学姐已经做了道歉。

图三，证实学弟已经原谅，知乎那篇非学弟所发，学弟更是被“伸张正义”之人骚扰。

图四，关键的来了，19日有人直接拿着学姐信息私信某博主要求曝光，不过被拒绝了。

那么在双方都想息事宁人，双方都未参与的情况下，有第三方在知乎装作学弟说自己被网爆被社会性死亡，还是第三方拿着学姐信息私信要求曝光。

知道学姐那么多详细信息的人，肯定认识学姐，既然认识，学姐连发五篇道歉和澄清文章，那么这人不可能不知道，既然知道，还故意隐瞒结果，到处私信大v说学弟被网暴了。

可以猜测，这次学姐事件，是第三方有组织

五、特点：

● 过度阐发延伸

从清华学姐上升到对清华、美院、艺术生等类似地图炮的攻击，从性骚扰延伸到其他性骚扰事件，舆论的过度阐发和延伸造成了真相无人在意，人们只是在乎这个由头去说他们想说的事情。

● 不就事论事，过度抽象

这个事件中有非常多细节可以延伸到舆论场中几个主要矛盾，在这场舆论风波中最显著的是女性和男性的矛盾。舆论场上，他们讨论的就不再是事件的真相和细节，而是搬出过往实锤的性骚扰和这次性骚扰诬告的传播率对比，来回应“男孩子要保护自己”这种舆论。

● 有力量在刻意制造对立（断章取义，捏造自白书）

几个营销号放出资姐私人信息，知乎上匿名的自白书，这些为了扩大流量的言论背后可能有资本在作怪，挑起更多的矛盾来达到舆论的进一步发酵和挑起公众的情绪。毕竟营销号需要一大笔钱。但是其目的尚不明确。（猜测可能后续会有新的舆论事件爆发，而清华学姐事件作为舆论铺垫。）

● 把人肉，社会死亡的譬喻在其他事件上

JM 帝国、法西斯、创作自由，这其实已经是另外一个事件了，但是如果提到人肉，由于清华学姐事件人肉影响力之大，舆论会联系此事件。

六、结论：

从清华学姐事件舆论发酵中得到的启示：

- 就事论事，不要上升，从个人到群体，从个例到全部，例如清华、美院和艺术生的无脑延伸
- 首先追求事件的真相而非妄加评判是非对错
- 谨防舆论中的资本，资本制造对立制造舆论数据，而个人要思考谨防成为资本的排头兵
- 谨防别有用心的人将事件扭曲，将人与人之间的矛盾转化为性别间的矛盾，扩大事态，营造异性对立，转嫁矛盾。
- 让子弹飞一会

.....

从清华学姐事件中得到的启示：

- （举报性骚扰靠挂人）当事人认为权利被侵害时应正常发声而不是首先借助舆论的力量。维权者维权并非炫耀权利，也不是一味地试图报复自以为的施害者。网络是一把双刃剑，贸然将一件事炒热可能火焰会反噬自己，是不可取的行为。诚然，社会应鼓励人们受到性骚扰时敢于发声，敢于举证，但不是不举证就急于恶化事态。有关部门也应完善性骚扰举报制度。
- （关于社会性死亡）很多微博上的社会性死亡都是秩序和规则缺失时的无奈之举，是不可取的行为。令一个人社会性的死亡成本可以忽略不计，然而威力无穷。这并不是是一种维权的方式，反而只会引起一部分无事可做的网民的狂欢，一些暗中营造对立的别有用心者的窃喜。
- （关于网民）作为这件事情的局外人，网民们可以永远站在道德制高点上对他人评价，指摘。而清华学姐这件事因其复杂性和戏剧性很快就被他们作为情绪的宣泄口，其中不乏别有用心和暗怀怨恨的人；不乏怀有阶级歧视，性别歧视，学历歧视的冲动分子；不乏坐山观虎斗，只看好戏，不讲原则的人。他们将网络环境变成一滩浑水，用以发泄生活中的种种阴暗面。我们在生活中要避免偏听则暗，静待花开。
- 总结：学会筛选网络信息，避免成为被人利用的利刃。

.....