# 新疆棉分析过程

# 高频词分析

由于热点词汇的出现频率会随着热点事件的发生而快速增长，通过统计热点词汇出现的频率的变化情况可以看出热点事件的主题，下面是统计3月-6月新疆棉高频词出现的最高次数



由这里不能看出，聚焦点还是以新疆，中国，人民，西方，耐克这些单词为主，为了更好的看出差异，我们采用词云图的方式，这样更为直观的看出主要的还是中国和西方之间的相关词出现频率较为密切

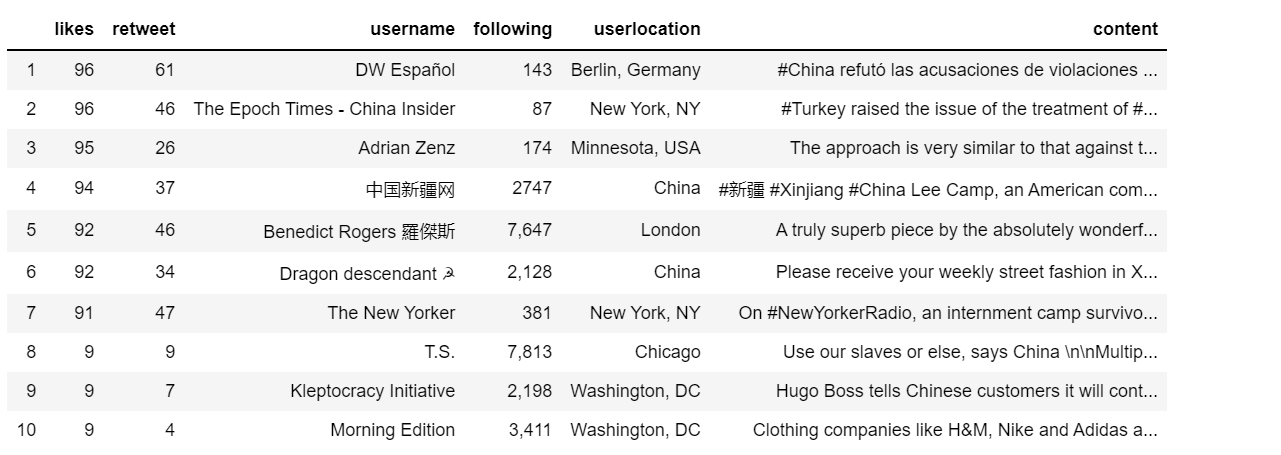


## 聚类

为了进一步探讨词与词之间的关系，我们采用了聚类的分析，但却因为它们的出现，反而淹没了在线网民发表的其他意见，这些意见会反映在其他的关键词里，然后，对总数据集词语共现做了聚类分析，得出总数据集词语共现聚类分析图，在这张图中也不能发现有效的信息，更不能从中得出关于在线网民如何讨论该排放门事件所产生的观点意见，也阻碍了对数据的深入分析

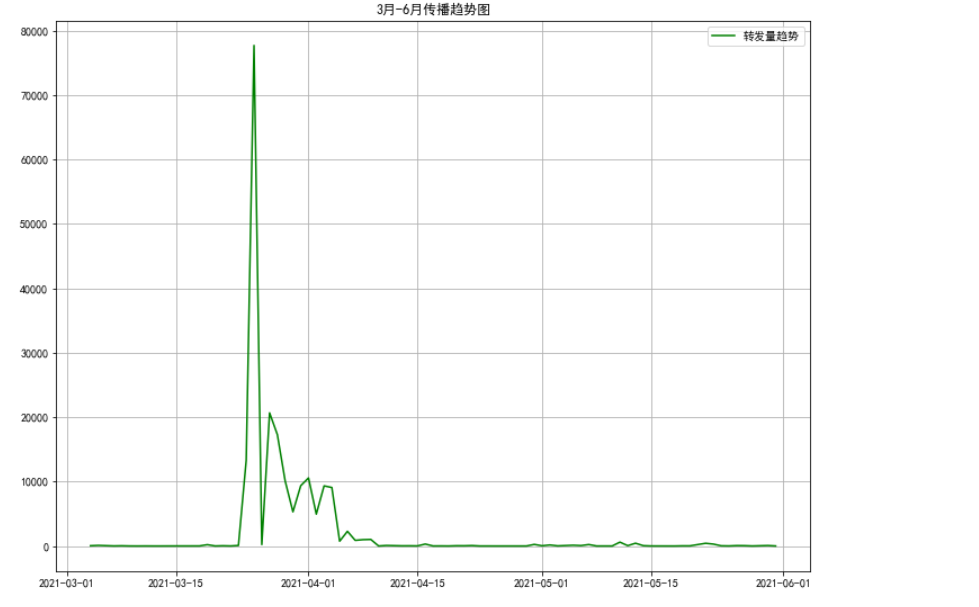


## 查看转发量点赞量前十的网民的观点



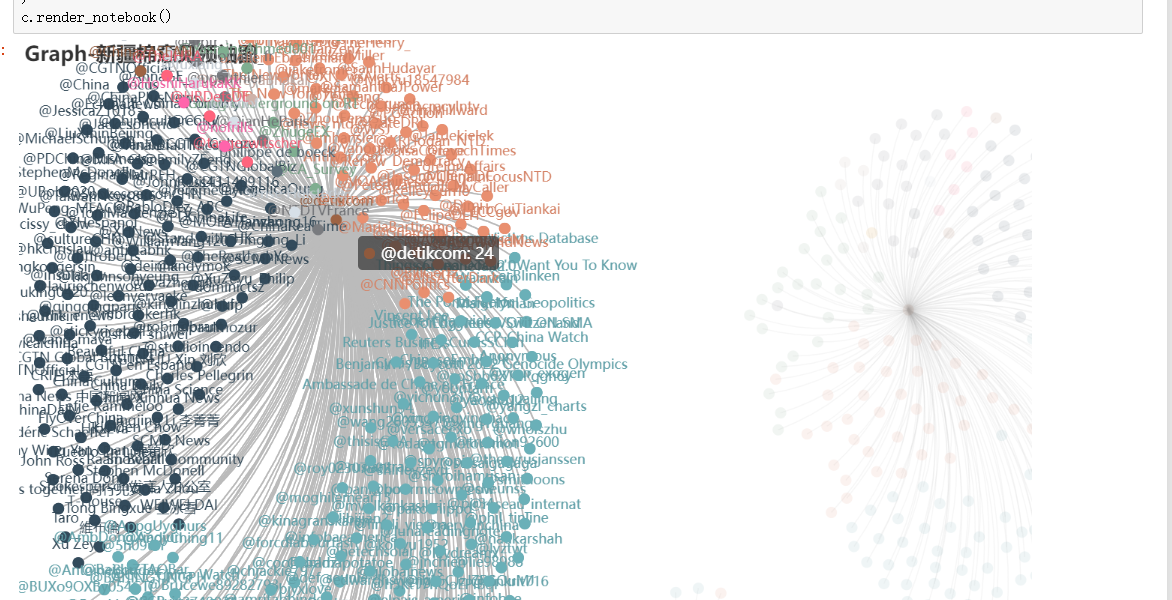
在这个前十的网民中，其大部分来自西方国家，中国的只有两名，还有其中一名为官媒，表明其实大多数西方国家的网民还是高度重视这件事情

## 传播量



传播量在3月底到4月这有一个明显的爆发点，我们可以大胆推测这段时间正是该事件的高潮，人们对此有热烈的讨论，在意见领袖的催动下引起大量微博转发和评论的行为，同时意见领袖所提及新疆棉的问题成为公众意见和态度的雏形

## 意见领袖



我们从意见领袖图中，找出每一块的意见领袖代表