Weibo Comments Analysis of Idol Public Opinion Reversal

--Take KUN as an Example

Group：Spider

LIU Zhaoxing 19456182

SHI Yi 19451512

HE Qing 19450141

WANG Yixue 19434871

Hong Kong Baptist University

COMM7350 AI and Digital Media Workshop

Professor: Dr Vincent WANG

May 15, 2020

Weibo Comments Analysis of Idol Public Opinion Reversal

--Take KUN as an Example

Background

Data and a huge fan base is the key to measure whether a star is popular or not. Massive amounts of data have created a potential commercial market for idols and stars. However, a large amount of data is also a double-edged sword, the more attention, the greater the pressure of public opinion of celebrities. If a celebrity is careless, he or she will be pushed into the abyss by the public opinion on the Internet platform.

CAI Xukun (Kun) is a controversial idol in Chinese major media and social media platforms in recent year. Since his debut as the center position in the reality show Idol Producer in 2018, Kun has experienced a process of attracting attention from tens of thousands of people on the Internet, being ridiculed by the whole Internet and becoming a good reputation. In the past two years, Kun has been the trending topic on Weibo several times.

Recently, with the broadcasting of reality show Young with you, Kun has got a high reputation in Chinese social media. A large number of people has changed the old image of him. Whether the data can prove that he really achieved the reversal of public opinion? How is the driving force behind the idols' online buzz? The goal of our project is to analyse the critical Weibo of these core communicators, including the texts of forwarding and comments. The significance lies in the discussion of when the star in the opinion reversal, the fan group brings how much power.

Method

In this report, we would use python to do the data analysis. Professional Weibo social network analysis tools are used to confirm the core communicator, and then python is used to capture comments of the core communicator's Weibo and analyse sentiments and comments of those comments. We also scraped some comments under Weibo of the fans leader of Kun to identify the characteristics of fans' ID to divide fans and non-fans from comments we got.

Data Collection

First, we conducted a preliminary analysis with the help of Weibo's official professional data analysis platform "微热点". It produced a report based on about 575,000 blogs about keywords "蔡徐坤+青春有你 2|青你 2+圈粉|转粉+导师"，from March to April. From this report, we get to know who played an important role in this event. Those bloggers who most frequently published or reposted related microblogs are the core communicators. Many hot messages were posted by them and caused a lot of comments and reposts.

We have selected the core communicators refers to the "activists" with the greatest influence (i.e. the largest number of followers) among the Weibo users who participated in the spread of the event, using python to scrap the comment data in every hot microblog which posted by the core communicators. The data frame mainly includes user name, timestamp, like count, and comment content. We divide these core

communicators into two categories. The first category is official media. These communicators are usually the accounts of official platforms, who have a certain right of speech and will guide the direction of the topic in an event. The second category is key opinion leaders. Compared with official institutions, they do not have much authority, but they can also influence the public opinion trend of an event. We have scarped 66110 comments from the core communicators' Weibo which the content is related to Cai Xukun. About 46480 comments is from official institutions such as "新浪综艺", "网易娱乐", "微博综艺", "芒果娱乐" and "新浪娱乐", and about 19630 comments is from key opinion leaders such as "鹅组情报员", "圈内教父", "荷兰豆养殖专家", "FashionWeek" and "熊猫 x 兔籽". The related Weibo content is showed as the following figure 1 and figure 2.

fig 1. Media accounts' Weibo related to Cai Xukun

fig2. Internet celebrities' Weibo related to Cai Xukun

Measures

Word Cloud can indicate the most frequently used words under demonstration discourse. The more the words were used, the bigger the words were shown in the picture. In content analysis, we would use word cloud and topic modelling to analyse the Weibo's comments.

We would use histogram to show the sentiment analysis result. the horizontal axis is the emotional score. The score is between 0-1, 0 represent the emotion is most negative and 1 represent the emotion is most positive. The vertical axis is the frequency of sentiments. We weighted the thumb up number of each comment, thumb up means that the user recognized the content of this comment, every comment gets one extra

score for every two likes, and we considered the score as the extra frequency in sentiment analysis. We use snowNLP package to measure sentimental score of each comment.

User Grouping

To make our findings more informative, we decided to divide the comments into different groups. We divided them into two groups by users' ID, comments from fans and comments from non-fans.

Since users are allowed to change their nickname for only one time a year, a user's ID really can lead to his/her identification and characteristics. So we scraped more than 15464 users' comments under Weibo from Kun's fan leader account "蔡徐坤全球粉丝后援会" to extract fans' features.

We found that fans are very willing to add part of Kun's name or nick name into their ID, like "蔡", "坤", "葵",, or some Homophones of those words, like "菜", "困", "奎", or some other features of Kun like birthday or constellation. We used these features to build a list and, if any part of a user's ID matches the features in the list, we consider this user as a fan of Kun, remaining users defined as non-fans.

At first we added quite a lot of phrases containing those words above to the feature list, and we double checked the frequency of words of remaining ID after filtering to find more obvious fan features and add them to the list to try more precise grouping. By using this step by step extracting, we finally found that the features of fans ID can

be briefly concluded into to some single words instead of many compound phrases. The detailed grouping dictionary was attached in appendix.

Result

Content Analysis

fig3. Word Cloud for Content

After grouping users in two dimensions (fan/non-fan, media/internet celebrity), we combined and integrated all the crawled content, and after deleting inessential words such as “蔡徐坤”, “啊”, “哈”, the text content is presented with four word clouds. As can be seen from the word cloud, generally speaking, contents in these four word clouds are similar, and the words “可爱”, “温柔”, and “喜欢” are more common. It is hard to distinguish the speakers of those four word clouds without titles, which means that the discourse of users in these four groups is generally consistent.

We also tried to present only part of comments with certain number of likes, including comments with no like, comments with at least two likes (to avoid thumb up by themselves) , and we found that the main discourse didn't have much differences.

In order to see the different characteristics of different groups more accurately, we conducted word cloud analysis of the content under each selected Weibo, a total of 10 Weibo, five from media, five from internet celebrities, and we divided the comments into fan comments and non-fan comments. Once again, the word cloud is presented according to different classifications. We found that the content of the word cloud is more related to the topic of the Weibo text, and the comments of fans and non-fan are not much different.

Subsequently, we further performed topic modelling on the text content. After deleting the common words , stop words and punctuation marks, we utilized Latent Dirichlet Allocation (LDA) to do that. We set the number of topics to five as we had 5 selected Weibo comments for each group, and we also divide the content into fan group and non-fan group. Topics are mainly about admiration or encouragement toward Kun. Four groups' results are shown below.

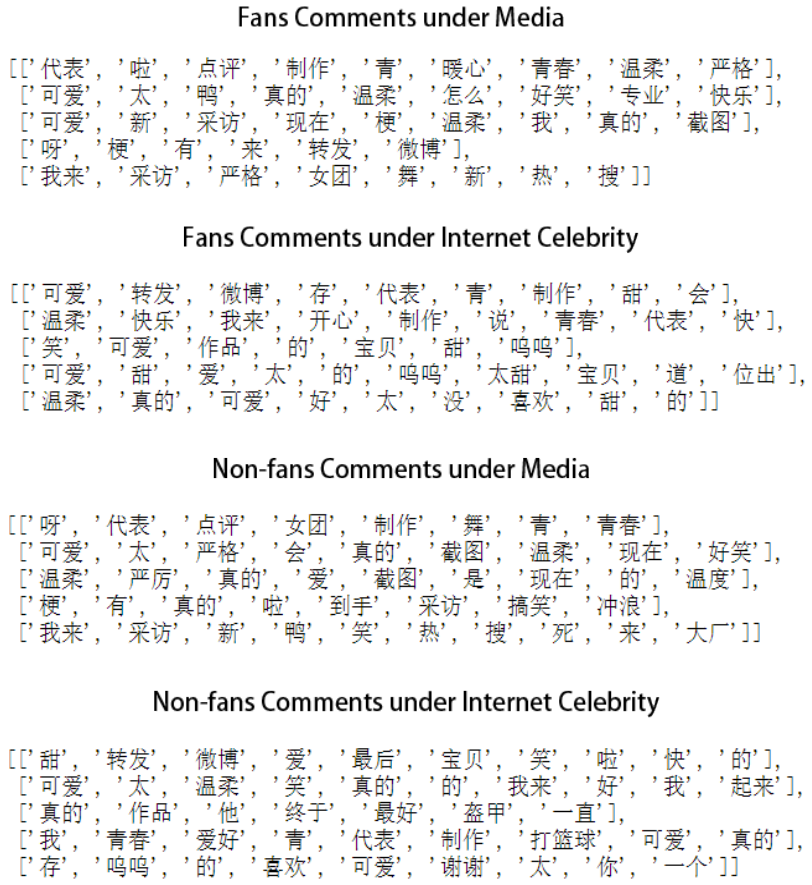


fig4. Topic modelling of comments

The number of likes of comments is also considered as an important metric to figure out topic model. Likes can be regard as agreements towards comments, so we gave every comment one extra score for every two likes, and we considered the score as the extra frequency in sentiment analysis. After weighting, some comments got strengthened, but the main discourse doesn't change too much. Both fans and non-fans are commending Kun. The topic model of comments after weighting is shown below.

Fans comments under Media

[['的', '舞台', '作品', '声音', '别着急', '热爱', '回应', '期待', '男孩', '坚持'],
['温柔', '回应', '啦', '很', '说', '采访', '新', '我'],
['期待', '今晚', '好', '会', '你', '都', '好帅', '最好', '人'],
['太', '代表', '喜欢', '专访', '我来', '严格', '爱', '来', '又'],
['可爱', '真的', '青春', '梗', '一个', '一起', '真是', '宝藏', '制作']]

Fans Comments under Internet Celebrity

[['可爱', '期待', '温柔', '女团', '好', '超级', '是', '不见不散', '暖心'],
['太', '梗', '现在', '温柔', '不错', '链接', '网页', '是'],
['真的', '的', '你', '啦', '会', '他', '网络', '觉得', '一个'],
['文字版', '的', '来', '来看', '热爱', '我来', '坚持', '看看', '专访'],
['喜欢', '人', '青春', '采访', '说', '新', '代表', '不', '也']]

Non-fans Comments under Media

[['偶练', '帅哥', '宝藏', '补', '赶紧', '男孩子', '网页', '链接'],
['制作', '优秀', '我家', '发现', '感觉', '白菜', '有种', '你', '代表', '爱'],
['作品', '他', '最好', '盔甲', '舞台', '超级', '真的', '铠甲', '的'],
['看', '动作', '杀', '营销', '好感', '号发', '直升', '开枪', '前两天'],
['温柔', '是', '就是', '真的', '开玩笑', '重生', '男孩子', '超级']]

Non-fans Comments under Internet Celebrity

[['没', '作品', '最好', '盔甲', '脸', '他', '微博', '后悔', '追', '开始'],
['一直', '温柔', '永远', '是', '作品', '自己', '说话', '成就', '突破'],
['的', '偶练', '觉得', '好', '人', '黑', '今天', '不是', '舞台', '我'],
['真的', '他', '真', '好看', '觉得', '巨', '青', '有点', '看到'],
['时候', '之前', '喜欢', '感觉', '很', '看到', '全网', '造型', '现在', '今晚']]

fig5. Topic modelling of comments after weighted

Sentiment Analysis

Official Accounts' Sentiment Analysis

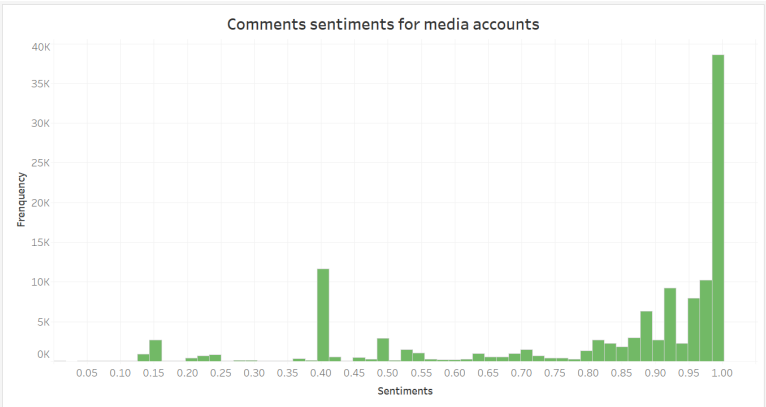


Fig6. Comments sentiments for media accounts

The figure shows the frequency of media accounts' comments with different sentiment scores, and it presents a very obvious tilt to a high positive attitude. After weighted with likes, the frequency of all comments is 121,963, and the comments with a high sentiment score (greater than 0.95) are about 60,000, which is close to 1/2 of the total.

```
Fan
1 吃菜菜不吃菜菜-:"不是自嘲 是释然是算了是大方回应被恶意剪辑鬼畜掩盖了的兴趣爱好本身  温柔地说出“ 篮球是我的爱好” 热爱真的无罪~"16527.0
---
Fan
2 282号发光虫:"多一些了解和靠近这个只是坚持自己的热爱的大男孩吧"10549.0
---
3 抱月亮睡:"“喜欢唱 跳 rap 篮球“小蔡一直说的都是“喜欢篮球“恶意鬼畜剪辑可以使一个人面目全非，让有的人还不认识他却已经开始讨厌他。男生喜欢打球是再正常不过的，恭喜小蔡终于能笑着说出来"9425.0
---
4 GG-Pink:"严厉有温度，温柔有原则！超级暖心的蔡PD啊！ "7426.0
---
Fan
5 八月狮子喵:"哈哈哈哈哈他真的好有梗好可爱哦"6759.0
---
Fan
6 SweetheartKUN_徐徐:"好看我好喜欢，“蓝”朋友，好帅，晚上不见不散蔡代表"6230.0
---
Fan
7 WiiiiKUN:"青春是漫长人生里最缤纷的色彩,热爱便是青春最美好的注解。同青春制作人代表@蔡徐坤 一起，为梦想增添一束光亮，用热爱记录最好的青春。"5321.0
---
8 墨鱼破壳仔:"哇，蔡徐坤的专访呀，快来看呀，蔡pd超有料呀"5241.0
---
Fan
9 KUN_柠萌:"蔡pd坚持你热爱的，我们一直在"5180.0
---
10 Ridiculous_Kun:"蔡徐坤太可爱啦"5020.0
---
```

fig7. Comments with score lager than 0.97 in Microblogs

We checked the content of comments with score lager than 0.97 in the order of like count. As shown in the picture, the comment most liked by people is from "吃菜菜不吃菜菜", who is identified as a fan of CAI by our grouping. This comment directly responds to Cai Xukun's biggest negative point before, we think which might make a great contribution to the reversal of public opinion about him. In addition, 6 of the 10 high sentiment comments with the highest number of likes (all more than 5000 likes) came from fans.

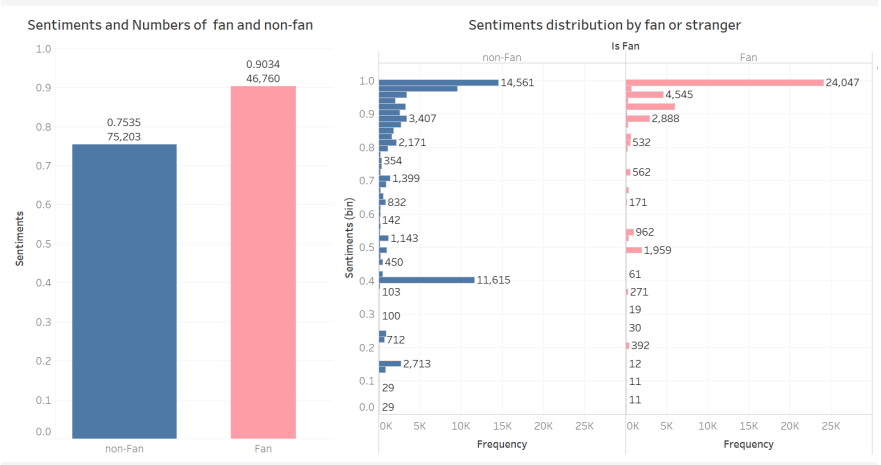


fig8. Dashboard for Sentiment Analysis

We also looked at the sentiment distribution of comments from fans and non-fans separately. As shown in the dashboard, the comments from non-fans take up 60% of the total (75,203), from fans about 40% (46,760). The comments from the two groups both are positive, but fans' sentiment score is extremely high, the average is 0.9034. However, the figure for non-fans is only 0.7535. We can see, from the sentiments distribution graph, over half of the fans' comments have extreme high sentiment score (close to 1.0), while the sentiment of non-fans' comments is more decentralized.

```
1 唐三是个小暖心:"其实小蔡本身就没什么实质性黑点吧 所以一时的玩梗也不会一直持续 现在看就是挺好的一个小朋友"23024.0
---
2 单单crush:"所以还是要了解才能下定论"5369.0
---
3 赞比星途璀璨:"网络上的你重拳出击 现实中的你不堪一击"1645.0
---
4 隨遇邇an:"来啦来啦，徐坤太有梗了高冷呀"1630.0
---
5 肉包里面全是菜:"又可以更了解他了，喜欢这个专访。  网页链接"761.0
---
6 肉包里面全是菜:"又可以更了解他了，喜欢这个专访。"761.0
---
Fan
7 刘爱坤Alan:"太可爱了吧啊啊啊啊啊啊我在现场我差点昏倒"742.0
---
8 kunwannaget炫:"你去便利店不都拎着三大袋回来的么蔡徐坤"583.0
---
Fan
9 蔡小葵的卷儿:"做好准备的人一定会得到一颗糖。"490.0
---
Fan
10 八月的小狮子冲呀:"所以不要跟风黑任何一个人 了解之后再评价吧"108.0
---
```

fig9. Comments with around 0.4 points in Micro blog

What is more, it is noticeable that there is a relatively high frequency (11,615) in the 0.4 sentiment score. We found most comments with around 0.4 points are indeed from non-fans, and their opinions present a trend of reversal. For example, the comment with 23024 likes said, “Now he looks like a good kid.” Another comment with a large number of likes also said, “So we still have to understand before making a conclusion.”

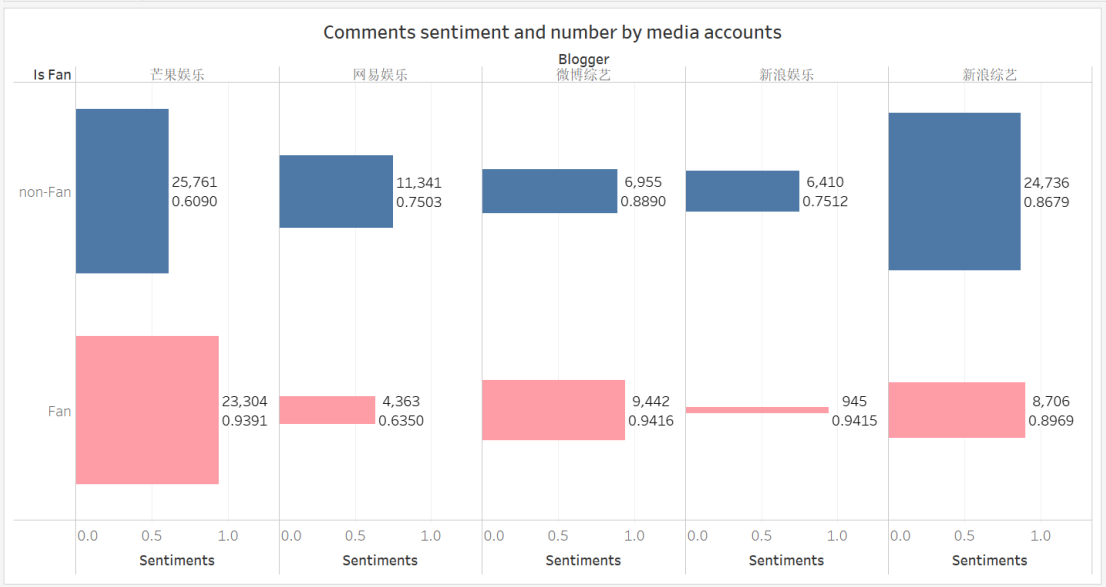


fig.10 Comment sentiment and number by media account

From the graph for comment sentiment and number by media account, the hot microblogs posted by “芒果娱乐” and “新浪综艺” have the most frequency of comments. But comments from “芒果娱乐” has a larger difference of sentiment between fans and non-fans, separately 0.9391 and 0.6090. It might because it posted the blog earliest after the Young with you getting online on March, 12, and the content was about Cai’s self-mocking to his controversial thing before, leading to intense discussion from people having different opinions.

fig11. Content from “芒果娱乐”

Individual Accounts’ Sentiment Analysis

We selected the five most influential non-official “activists” among the Weibo users who participated in the spread of the event to analyse the emotional tendency of the comments. However, in the preliminary analysis, we found that there was a user named “鹅组情报员” who posted a Weibo about Cai Xukun related to Yu Shuxin, another popular contestant in the reality show Young with You. Most of the comments below are about Yu Shuxin, which are all negative tendencies like affectation and disgust. In order to eliminate bias, we deleted this user's Weibo and comment content in the sentiment analysis, and analysed the sentiment of the other four Weibo comments. They are “圈内教父”, “荷兰豆养殖专家”, “FashionWeek” and “熊猫 x 兔籽”.

As shown in fig12, we could see that users have obvious emotional differences in the key opinion leader's comments. According to the micro blog about Cai Xukun posted by KOL, we can see that the frequency of positive emotions is relatively high, and the frequency of nearly 5000 emotions is 1.0, which is a very positive emotion.

Sentiment score greater than or equal to 0.5 accounted for most of the comment frequency, indicating that netizens' evaluation of Cai Xukun in reality show Young With You tends to be positive.

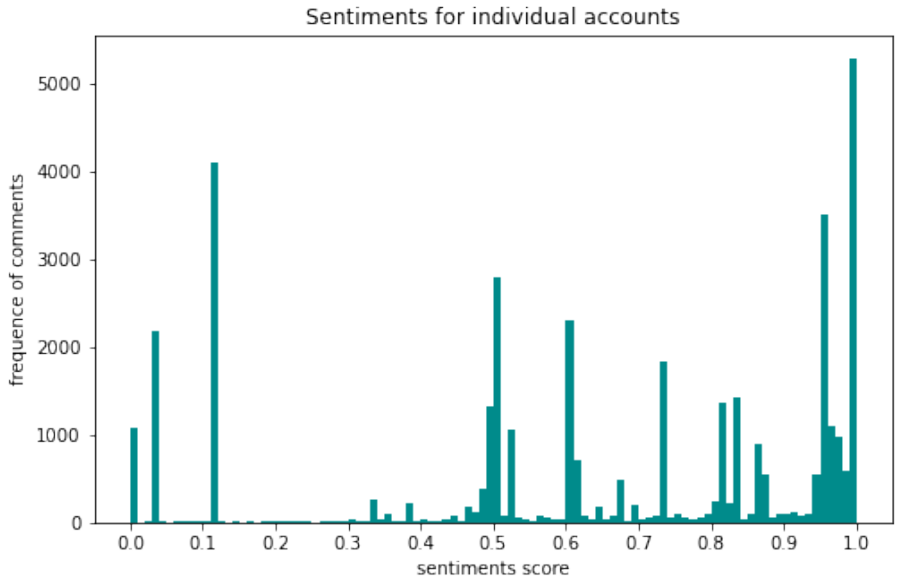
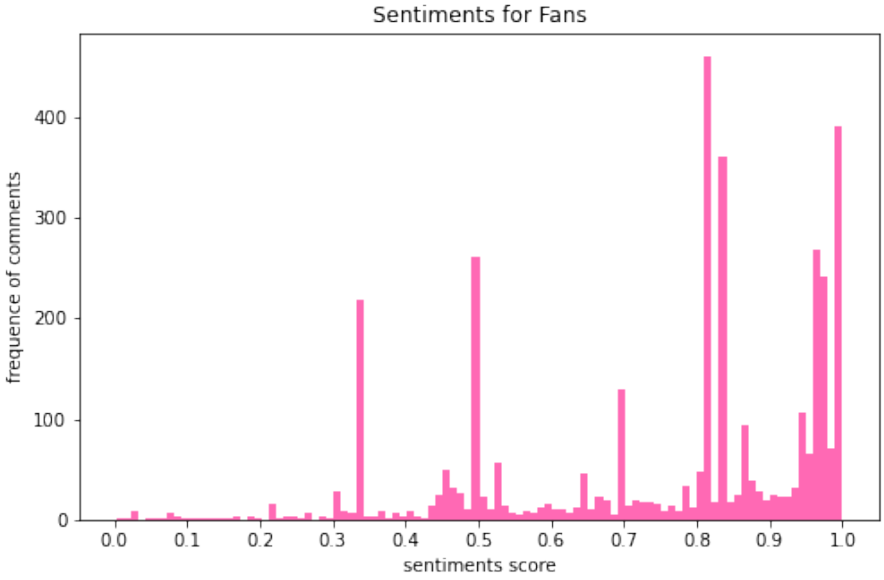


fig12. Sentiments for individual accounts

Compared with fig13 and fig14, we could see the difference between fan's sentiment and non-fan's sentiment. It is obvious that Cai's loyal fans have been positive about emotional tendencies on micro blog which include Cai. Most of the fans had an emotional score between 0.8 and 1.0. People who are not Cai Xukun's fans' sentiment have a severe polarization. The sentiment score equals one have almost 5000 frequency of comments. And the negative score (0 and 0.1) have 7000 frequency of comments. However, by comparing the vertical axis, it can be found that the number of fans and non-fans is different. The highest frequency of loyal fans is only 500, which indicates that most of the users who comment on these micro blog related to CAI Xukun are ordinary people. Loyal fans are also leading the way in the comments.



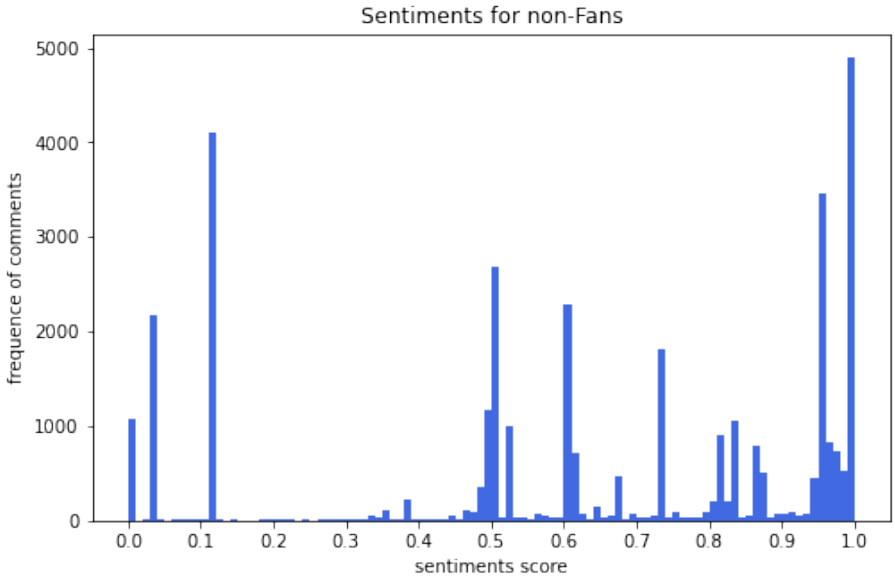


fig13 & fig14 Sentiments compared by fans and non-fans

As shown in fig15, we could see the sentiment score of Cai Xukun's microblog posted by the "圈内教父" is very high, even the average sentiment score of non-fans is higher than that of loyal fans. Meanwhile, the average sentiment score of Weibo account which called "荷兰豆养殖专家" is the lowest, but also more than 0.5. According to the content and comments on this micro blog shows in fig16, most of the comments with an average score of around 0.5 are expressing a changing attitude, from

the negative emotions in the previous negative news to a more positive one. Most of the netizens shows a positive mood in discussion of Cai Xukun as the PD in Young with you. There seems whether it is Cai Xukun's fans or not, there is no big difference in emotional judgment.

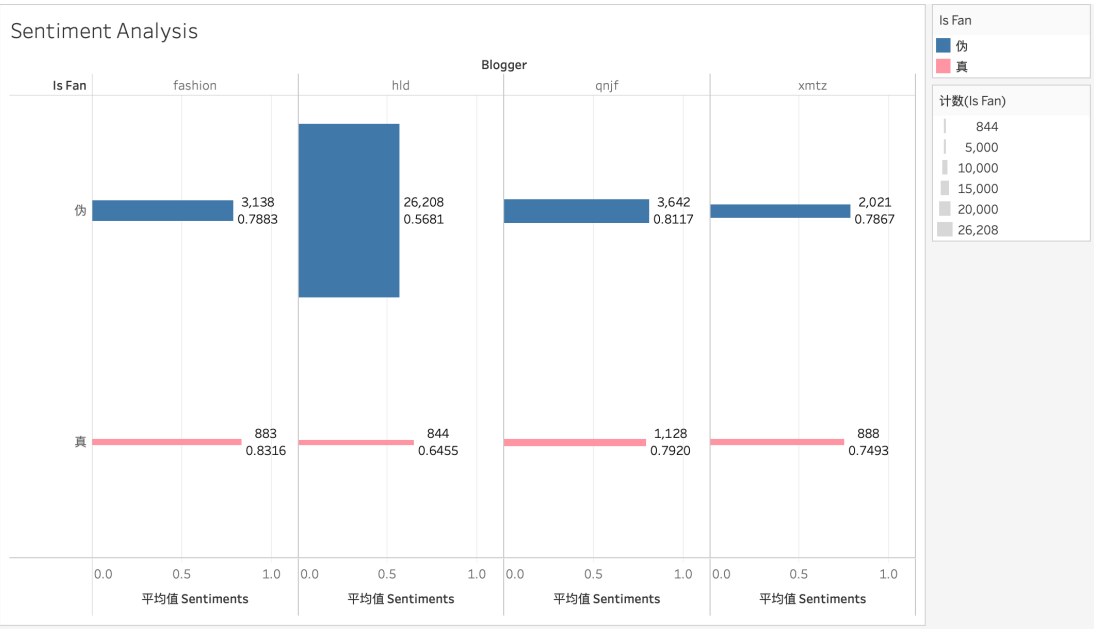


fig15. Overview for sentiment analysis on individual account

1 荷兰豆养殖专家:"我没看过偶，我承认我后悔了"8187
---
2 青春有你2-懒羊羊:"我好后悔当年没看偶练，让我失去了一个老公"5342
---
3 ulxxgnim:"我对他没啥好改观的，因为本来他就没做错什么，cxk一直都是这样的，无脑跟风黑的人需要改改自己的脑袋，而不是改对他的观。"4349
---
4 iamatudou:"说真的，青2让我看到了偶练巨C真的不是白叫的，有点后悔给他投路人票投少了"1871
---
5 帅了个仔:"他可真好看 我真肤浅"1182
---
6 春色梅子酒:"用时间来证明一切"653

Fan
7 快乐星球825号:"我看了之后对他的好感直升 而且前两天看营销号发的开枪的舞蹈动作真的有杀到我"419
---
8 小沈同学今天努力了吗:"最开始知道他是星动亚洲，妈耶！！！还有婴儿肥的脸听说偶练有他我才开始追的偶练"413
---
9 夏日赏味限定-:"之前和姐妹一起看偶练就觉得他站在舞台上好吸睛哦，一百多个人跳舞就只能看到他（不是因为他站c位的原因）所以后来全网黑他的时候我也黑不起来，毕竟能力在那儿啊！不过uls1，他是不是微调了？我感觉看偶练的时候他脸还有点不对称，但今晚看他好帅哦！（粉丝别来喷我，谢谢"268
---
10 是福猪陌陌吖:"你的作品是舞台！不要着急回应所有的声音！[簪]"248
---

fig16. Comments under "鹅组情报员"

Conclusion and Limitation

Conclusion

We can see from the content analysis that the main discourse no matter from fans or non-fans are all about admiration and commendation toward Kun. They all use words like “可爱”, “温柔”, “喜欢” to express their admiration. For comments with low sentiment score (lower than 0.35), we found that the main discourse is “高冷”or “早生贵子”, which seems not to be that negative.

With topic modelling, we had extracted keywords under different topics, and we found that the topics are still mainly about admiration and commendations towards Kun, with no obvious difference between fans and non-fans. Another finding is that keywords under different topics really have some relationships with the content of the original microblogs, which means that theme of comments under a microblog has kind of correlation with the theme of that original microblog content.

After analysing the sentiment of the comments of four group of users, we find that the overall comments showed a clear positive trend, and those with a particularly high emotional score (> 0.98) mainly came from fans. The average emotional scores of fans and non-fans are positive, but the fans are particularly high, 0.2 higher than non-fans.

The more neutral comments mainly come from non-fans, and the content presents a trend of reversing opinions. After Young With You broadcasted , the earliest comments on Weibo with content about Cai Xukun’s self-deprecating, published by “芒果娱乐”, showed the biggest difference between fans and non-fans.

Limitation

In judging whether the user is KUN's fan or not, we just use the specific words in their ID to divided them into fans and non-fans, there are still some fans without obvious ID seen as the non-fans. Therefore, the non-fans group are not just netizens who are not fans of KUN.

Through the content analysis and sentiment analysis, we both got the positive results about the four-group comments to KUN, which means his public opinion reversion is successful, compared to his previous bad reputation in Weibo, which is the past fact. However, we can analyse more comments at those periods to show the dynamic change of attitudes towards KUN: from negative to positive.

Reference

Nip, J. Y. M., & Fu, K. (2016). Networked framing between source posts and their reposts: An analysis of public opinion on China's microblogs. *Information, Communication & Society*, *19*(8), 1127–1149. https://doi.org/10.1080/1369118X.2015.1104372

Yuan, E. J., Feng, M., & Danowski, J. A. (2013). "Privacy" in Semantic Networks on Chinese Social Media: The Case of Sina Weibo: Privacy in Semantic Networks on Sina Weibo. *Journal of Communication*, *63*(6), 1011–1031. https://doi.org/10.1111/jcom.12058

Zou, X., Yang, J., & Zhang, J. (2018). Microblog sentiment analysis using social and topic context. *PLOS ONE*, *13*(2), e0191163. https://doi.org/10.1371/journal.pone.0191163

Author contributions: He Qing and Wang Yixue crawled data; Liu Zhaoxing and Shi Yi grouped users; Liu Zhaoxing and Wang Yixue did content analysis; He Qing and Shi Yi did sentiment analysis; all the members wrote the report.