

# YouTube Comments Analysis

## Introduction and Motivations

現今網路世代崛起，自媒體成為了新興娛樂產業的主流，YouTuber 網紅也如雨後春筍般大量出現，但也因此充斥各種褒貶不一的言論。除了影片本身傳遞的資訊外，網友們在留言區的討論有時甚至比影片本身更精采，出現兩組甚至多組極端相反的評語也是常見的情況，因此若並非對事件本身有足夠的了解，容易因為留言讓人變得難以判斷是非對錯。我們希望能夠透過分析 YouTube 影片下的繁體中文留言，找出留言中的隱藏資訊及大眾普遍的想法。我們分析討論內容及動機包含：

1. 留言能夠反應網紅目前的風評  
我們將會比對 PTT 及 YouTube 留言中，針對同一名 YouTube 網紅的評論進行正負面分析，討論該網紅目前在大眾的印象中偏向為何。
2. 留言很難分辨真假粉  
單憑一則留言難以分辨留言者的真實心態，因此將會分析留言所包含的正負面內容進行分析，也會比較同個人在同一個頻道中不同影片的留言區出現的次數與頻率進行統計與分析。

## Youtuber 網紅選擇

我們根據網紅分析網站 influencer 中目前臺灣地區 YouTuber 網紅粉絲數排名，選擇了符合排名靠前、非大型公司或電視台、非發布 MV 為主的歌手個人頻道、粉絲留言以中文為主等條件的六個頻道進行分析與討論：

JOEMAN、木曜 4 超玩、白癡公主、阿滴英文、這群人 TGOP、蔡阿嘎，其中多組擁有多個頻道，我們暫不討論副頻道中的影片及留言。

## Tool and Implement

Language：Python

### YouTube crawler：

使用 YouTube Data API v3 抓取頻道中最新發布影片中的所有留言。

### PTT crawler：

我們選擇了在 PTT 上較為熱門的四個版的內容：

KoreaDrama 韓劇版、Gossiping 八卦版、Movie 電影版、C\_Chat 希洽版。

每一個版爬二十篇推數超過 50 的文章，計算正負留言占所有留言之比例。

$$\text{Ratio} = \frac{\text{Comments type count}}{\text{Total comments count}}$$

### Toolkit：senti\_c

Sentiment analysis toolkit for traditional Chinese

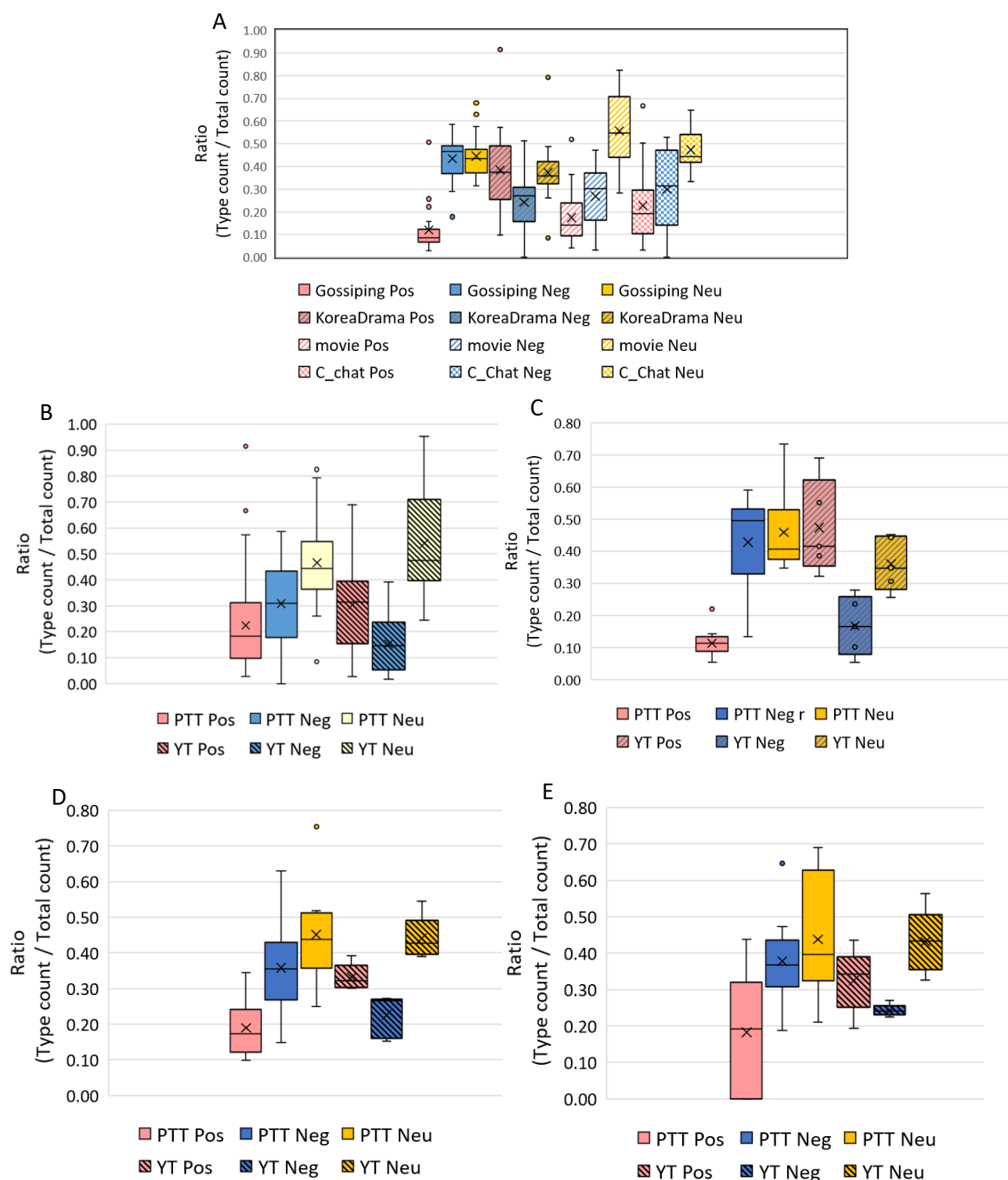
## Results

### PTT 和 YouTube 的留言偏好比較

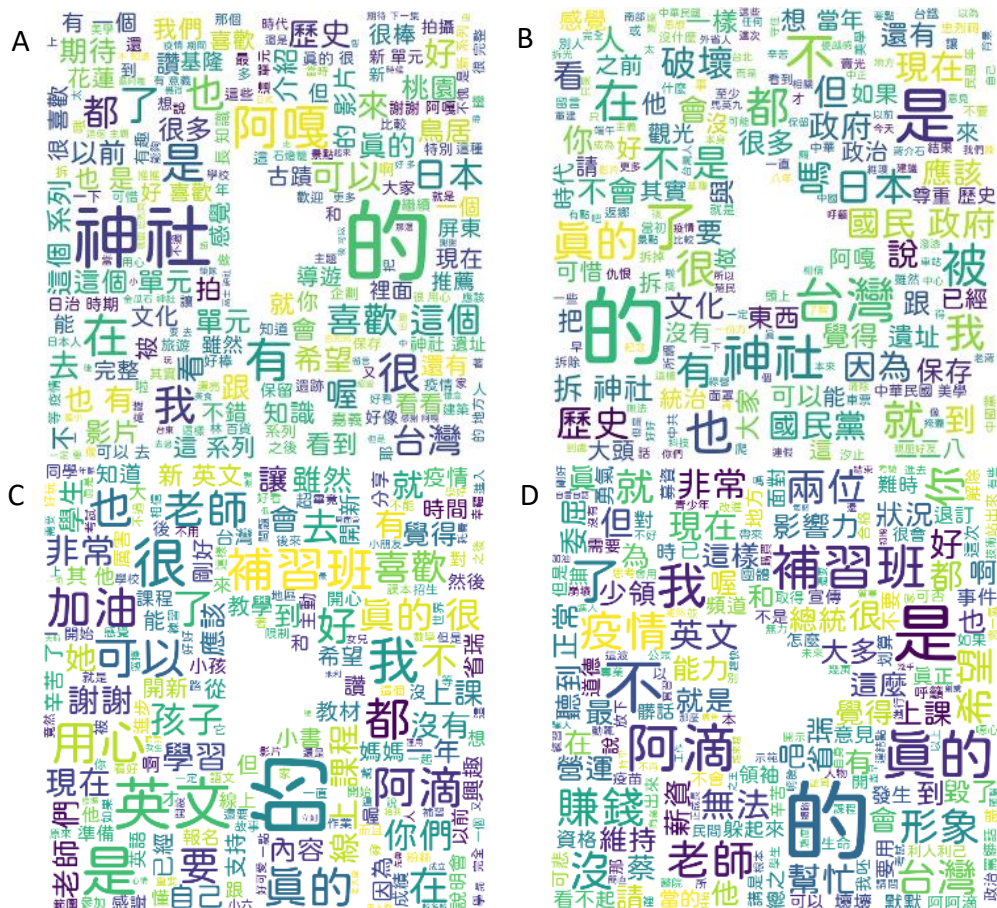
圖一 A 為各版整體留言分布，其中八卦版的正面評論最少而負面評論最多，電影版中性評論最多是所有分類中最為突出的，韓劇版則是正面評論相對其餘三版最多，透過分析不同版的留言情況，也可以大致獲得不同版上的網民的特性。

分析 PTT 各版的特色後，我們也觀察了 PTT 和 YouTube 上的留言差異，如圖一 B 所示，根據圖可以發現，YouTube 上的中性、正面評論比例皆多於 PTT，因此可以暫時定義 YouTube 的留言環境相對友善，出現在影片底下的留言也以真正支持、想法相近的粉絲所留下的為主。

之後我們針對同一位 YouTuber 進行兩個平台的比較，我們挑選出特別進行圖表分析的三組 Youtuber 在 PTT 和 YouTube 兩個平台中所獲得的正負面評論，分別為蔡阿嘎、Joeman 和這群人 TGOP。在這三組我們可以看到 Youtube 和 PTT 上對於同一位創作者的留言偏好皆呈現相反的情況，這和分析 PTT 和 YouTube 整體留言分布時得到的結論類似，在其他 YouTuber 身上也可以得到類似的結果。



圖一、PTT 和 YouTube 的留言偏好整體比較。Pos (Positive comments ratio) 為正面留言占所有留言的比例，Neg (Negative comments ratio) 為負面留言占所有留言的比例，Neu (Neutral comments ratio) 為中性留言占所有留言的比例。YT 代表 YouTube。圖 A 為 PTT 各版整體留言偏好傾向分布。圖 B 為 PTT 整體蒐集到的留言和 YouTube 整體蒐集到的留言做偏好傾向的比較。圖 C 為蔡阿嘎在 YouTube 幾部影片下留言和在 PTT 相關文章的留言做偏好傾向比較。圖 D 為 Joeman 在 YouTube 幾部影片下留言和在 PTT 相關文章的留言做偏好傾向比較。圖 E 為這群人 TGOP 在 YouTube 幾部影片下留言和在 PTT 相關文章的留言做偏好傾向比較。



圖二、字頻分析 (Word Frequency Analysis)。圖 A、B 為蔡阿嘎【尋找台灣神社#1：基隆新北】的字頻分析。圖 A 為正面留言，圖 B 為負面留言。圖 C、D 為阿滴英文【開補習班沒賺錢！有人信嗎？】的字頻分析。圖 C 為正面留言，圖 D 為負面留言。

### 針對 YouTube 留言分析

分析影片正負面留言，找出大部分正負面留言所包含的字詞，並找尋正負面留言和 YouTuber 相對應關係。這部份分析將會分成兩階段：

#### 1. 抓取高頻字詞

統計詞頻分別選出正面、負面、中性的詞。

#### 2. 加入 POS tagging 增加精確度

不僅在統計詞頻時做斷詞，同時在斷詞時做 POS tagging 進行詞性的標記藉以提高精確度，相同的詞也可能因為不同的詞性標記而分開計數，最後做去相依性得到最後結果。

### 留言字頻分析 (Word Frequency Analysis)

第一階段分析將以單支影片底下的留言去做統計：

### 蔡阿嘎【尋找台灣神社#1：基隆新北】

圖二 A 為此影片留言中正面評論的高頻詞，而圖二 B 為負面評論的高頻詞。在正面評論中，多是稱讚為知識型影片、歷史地理相關等內容，而負面評論中則是包含了國民黨、政府破壞古蹟等內容。

### 阿滴英文【開補習班沒賺錢！有人信嗎？】

圖二 C 為此影片留言中正面評論的高頻詞，而圖二 D 為負面評論的高頻詞。正面評論中的內容多是感謝阿滴為了孩子英文教學的用心及鼓勵，而負面評論中則是認為阿滴開補習班主要是為了賺錢。

抓取高頻詞得到的結果發現，雖然能夠大略的獲知正反面評論的主張，但圖片中可見重複的字詞及語助詞，像是第一支影片

中正負面皆出現了「臺灣」、「歷史」等字詞，而第二支影片則是都有「補習班」、「阿滴」等字詞，並且四張圖片中

皆可見明顯的「的」，因此加入了第二階段的分析。

A

Positive Sentiment correlation Word	POS	Frequency
喜歡	VK	83
~	FW	45
愛	VL	39
好看	VH	37
哈	D	36
讚	VC	36
超	Dfa	35
好喝	VH	34
實體書	Na	34
好吃	VH	33

B

Negative Sentiment correlation Word	POS	Frequency
只	Da	30
疫情	Na	29
到	P	28
沒	VJ	27
直接	VH	27
再	D	26
用	P	24
戴	VC	23
被	P	22
一樣	VH	22

C

Positive Sentiment correlation Word	POS	Frequency
感謝	VK	309
感動	VHC	107
友好	VH	85
台	Nc	84
加油	VB	83
日	Nc	79
影片	Na	76
非常	Dfa	54
帥	VH	53
喜歡	VK	53

D

Negative Sentiment correlation Word	POS	Frequency
違規	VA	141
檢舉	VC	97
違停	VA	83
違停	VC	76
自己	Nh	64
但	Cbb	63
停	VHC	61
沒	VJ	58
不要	D	58
交通	Na	53

E

Positive Sentiment correlation Word	POS	Frequency
好笑	VH	271
可愛	VH	152
集	Nf	128
喜歡	VK	113
超	Dfa	107
演	VC	92
愛	VL	72
最	Dfa	66
好看	VH	65
❤	FW	65

F

Negative Sentiment correlation Word	POS	Frequency
、	PAUSECATEGORY	120
自己	Nh	77
一直	D	63
天	Nf	62
這樣	VH	60
沒	VJ	60
只	Da	59
越	D	53
新人	Na	53
陷阱題	Na	49

圖三、第二階段的留言分析。圖 A、B 為 Joeman，圖 C、D 為木曜 4 超玩，圖 E、F 為這群人 TGOP。

## 第二階段的留言分析

1. 使用中性高頻詞過濾掉正負面高頻詞中的語助詞
2. 使用負面高頻詞過濾掉正面高頻詞中和影片相關聯的主題詞

第二階段的分析則是以該 YouTuber 最近五支影片的留言進行統計：

這階段分析的三組為 Joeman、木曜 4 超玩、這群人 TGOP。圖三所示，從三組結果中可見，正面高頻詞可抓取出較明顯的

正面詞彙，而負面相對少，且由於 Scope 非常小，斷詞後經常重複出現一樣的單一字元因而其排名十分靠前，使獲得的結果較難以直接進行正負面內容分析。

我們主要使用 Term Frequency 因而濾出的中性結果還是相對少，造成無法有效的過濾助詞及連結詞等，之後若是有繼續進一步研究此主題，使用 Term Frequency-inverse document frequency 應該能夠得到更好的結果。

表一、粉絲留言分析。左半部份為臺灣區各大 YouTuber 訂閱數排名，而右半部份則是粉絲留言數排名。

排名	YouTuber	訂閱數 (萬)	排名	YouTuber		留言數
1	這群人 TGOP	341	1	阿神	↑	250
2	阿滴英文	267	2	黃氏兄弟	↑	87
3	阿神	261	3	這群人 TGOP	↓	69
4	Pan Piano	249	4	Pan Piano		51
5	蔡阿嘎	248	5	千千進食中	↑	51
6	Joeman	209	6	蔡阿嘎	↓	29
7	木曜 4 超玩	209	7	白癡公主	↑	27
8	黃氏兄弟	171	8	阿滴英文	↓	15
9	千千進食中	164	9	木曜 4 超玩	↓	9
10	白癡公主	154	10	Joeman	↓	5

A

小快樂 很可悲阿 阿滴喔 不是台灣可以幫忙 不是很可悲阿 阿滴喔 不是台灣可以幫忙 不是很會宣傳 很可悲阿 阿滴喔 不是台灣可以幫忙 不是很0 3

B

Simply Peterlogical Holoosexual Second 119 Lol 0 0

圖四、部分特殊情況留言。A 為阿滴英文頻道連續三部影片皆留言之粉絲。B 為白癡公主的頻道中其中一名連續三部影片皆留言之粉絲。

## 粉絲留言分析

除了分析留言正負面內容外，透過觀察粉絲留言情形也可以發現目前大眾對於該 YouTuber 的態度，如表一所示。我們定義連續三部影片都留言者為該 YouTuber 的粉絲並進行排序，表一左半部份為臺灣區各

大 YouTuber 訂閱數排名，而右半部份則是粉絲留言數排名，其中阿滴英文和 Joeman 的訂閱數都很高但忠實粉絲的留言數卻很少，可以判斷該頻道可能正在走下坡或容易吸引新粉絲卻無法長期留住觀眾的現況。



此外在分析時還發現，有兩種特殊情況，如圖四所示。圖四 A 留言者為小快樂的例子，他在阿滴英文頻道連續三部影片皆留言但全是負評，因此被定義為「黑粉」。

圖四 B 則是在白癡公主的頻道中其中一名連續三部影片皆留言的粉絲，雖頻繁的進行留言但留言內容皆是不確定其想表達的意思的內容，因此被定義為「假粉」。

## Member List and Contributes

109065532 孫協廷

- Paper research
- Comment analysis

109062507 郭源哲

- Paper research
- Comment analysis

109065711 朱彥如

- PTT crawler
- Comment analysis

106062122 丁憶涵

- YouTube Crawler
- 資料整理

## Reference

1. <https://tw.noxinfluencer.com/youtube-channel-rank/top-100-tw-all-youtuber-sorted-by-subs-weekly>
2. <https://pypi.org/project/senti-c/0.1.0/>
3. YouTube
4. PTT