

# 社群媒體分析 讀書會 2

## Group3

B104020008	林冠儒	B104020044	黎懿綺
B114020046	蔡佳倚	M134020017	李恩榮
M134020024	盧荻原	M134020039	顏淳如

# 目錄

1. Week\_6 文辭與文件分析
2. Week\_7 文件分類
3. Week\_9 主題模型

# 1. Week\_6 文辭與文件分析

# 1-1 文辭與文件分析\_資料集

- 資料來源：PTT八卦&電影版
- 關鍵字：進擊的巨人
- 時間範疇：2024/01/01~2025/04/11

# 1-2 文辭與文件分析\_Bigram & Trigram

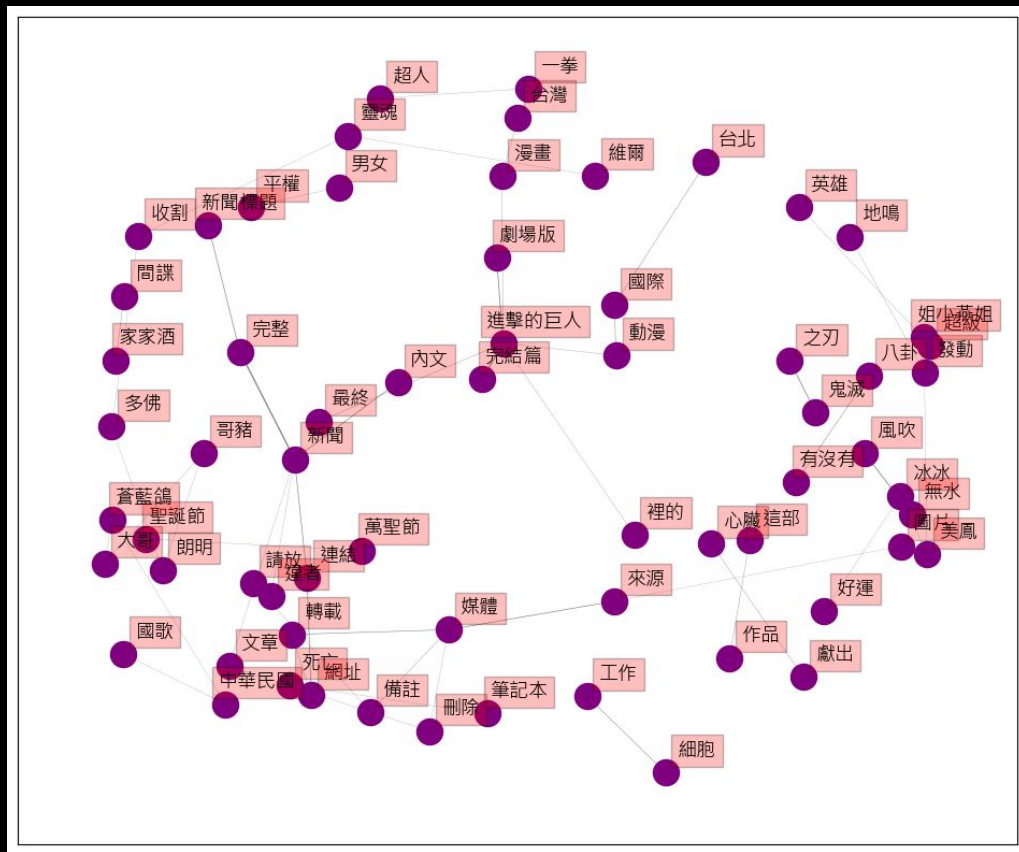
	word	count
0	進擊 巨人	806
1	完整 新聞	87
2	巨人 完結篇	79
3	風吹 無水	54
4	調查 兵團	51
5	劇場版 進擊	51
6	鬼滅 之刃	49
9	完整 新聞標題	44
10	媒體 來源	44
8	新聞 內文	44

▲ Bigram

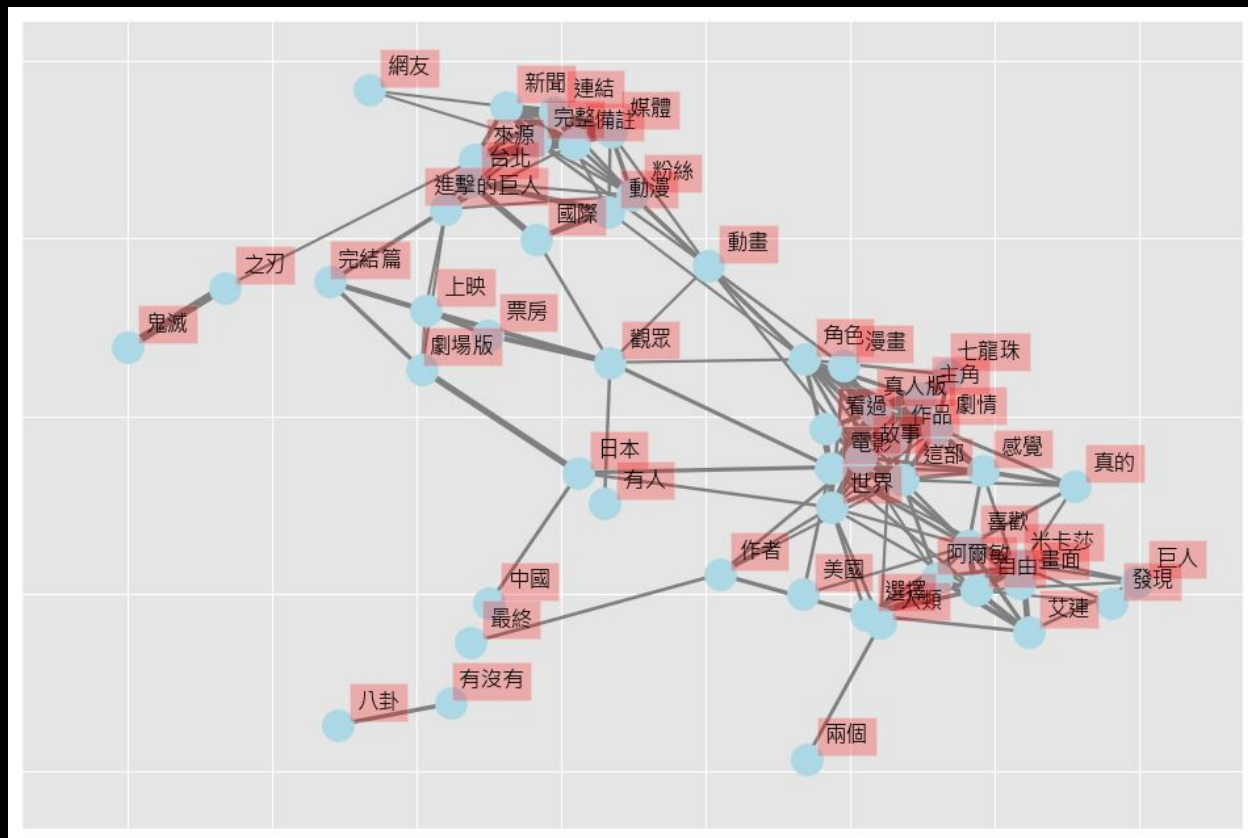
	word	count
0	進擊 巨人 完結篇	75
1	劇場版 進擊 巨人	51
2	完整 新聞 連結	44
3	新聞 連結 網址	43
4	完整 新聞 內文	43
5	台北 國際 動漫	26
6	網址 轉載 媒體	24
7	連結 網址 轉載	24
8	進擊 巨人 最終	22
9	轉載 媒體 備註	22

▲ Trigram

## 1-3 文辭與文件分析\_Bigram視覺化



# 1-4 文辭與文件分析\_詞頻共線圖 (f=60, thres=0.3)



# 1-5 文辭與文件分析\_Ngram預測模型

- 輸入字為 "艾蓮"

```
next word: 米卡莎, probability: 0.0421
next word: 葉卡, probability: 0.0421
next word: 崩潰, probability: 0.0421
next word: </s>, probability: 0.0316
next word: 繼承, probability: 0.0316
```

- 輸入字為 "巨人"

```
next word: 之力, probability: 0.0337
next word: </s>, probability: 0.0241
next word: 動畫, probability: 0.0193
next word: 世界, probability: 0.0193
next word: 巨人, probability: 0.0145
```



## 2. Week\_7 文件分類

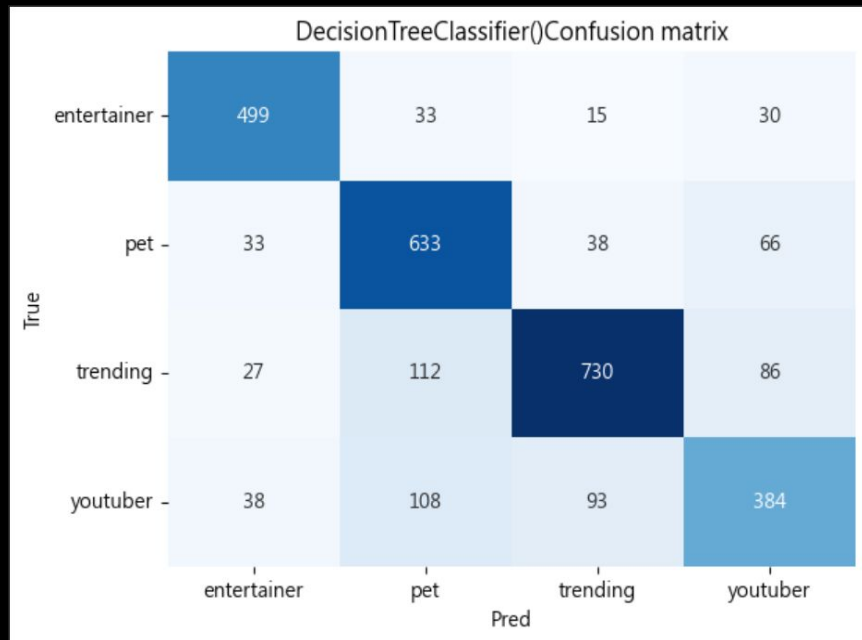
## 2-1 文件分類\_資料集

- 資料來源：Dcard
- 版別：追星、寵物、時事、Youtuber
- 時間範圍：2024/11/01 ~ 2024/11/30
- 資料筆數：4488

## 2-2 文件分類\_模型訓練

- 資料切割比例 : 7:3
- 訓練模型種類 : 邏輯回歸、決策樹、SVM、隨機森林
- 採cross-validation,  $k = 5$

## 2-3 文件分類\_決策樹

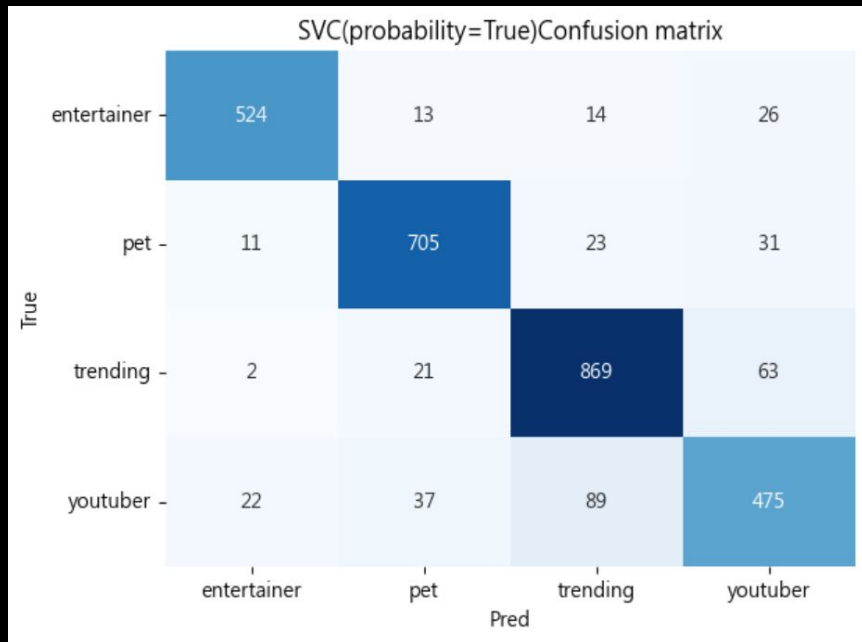


```
now training: clf_dtree
              precision    recall  f1-score   support

entertainer    0.84      0.86      0.85     577
    pet        0.71      0.82      0.76     770
    trending    0.83      0.76      0.80     955
    youtuber    0.68      0.62      0.65     623

   accuracy          0.77     2925
  macro avg    0.77      0.77      0.76     2925
weighted avg    0.77      0.77      0.77     2925
```

## 2-3 文件分類\_SVM

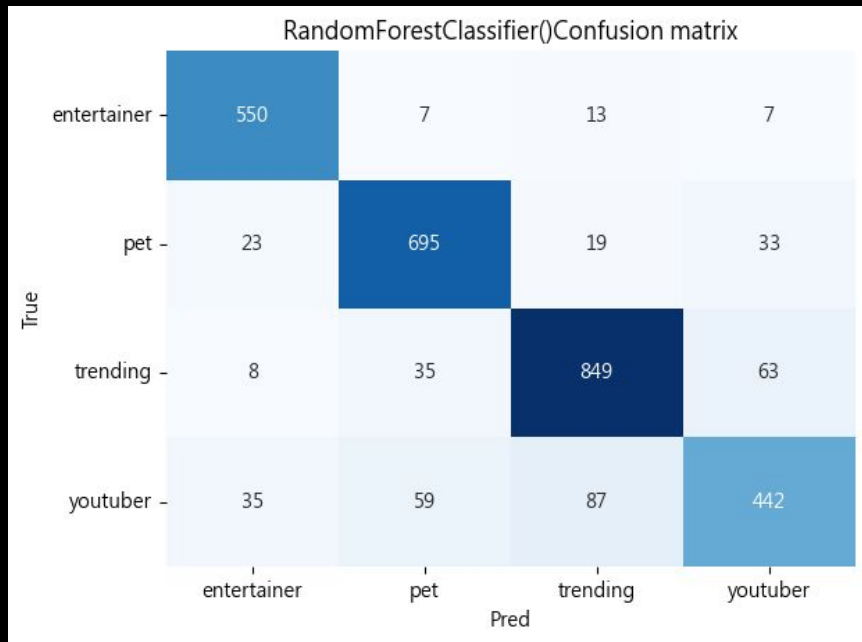


```
now training: clf_svm
precision    recall  f1-score   support

entertainer   0.94    0.91    0.92     577
pet           0.91    0.92    0.91     770
trending      0.87    0.91    0.89     955
youtuber      0.80    0.76    0.78     623

accuracy          0.88    2925
macro avg         0.88    0.87    0.88    2925
weighted avg      0.88    0.88    0.88    2925
```

## 2-3 文件分類\_隨機森林

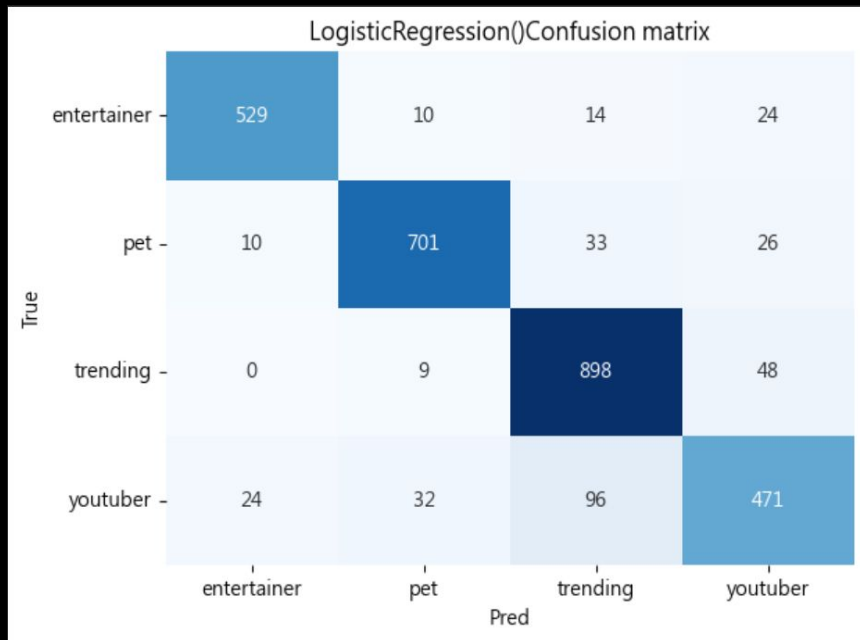


```
now training: clf_rf
precision    recall  f1-score   support

entertainer   0.89    0.95    0.92     577
pet           0.87    0.90    0.89     770
trending      0.88    0.89    0.88     955
youtuber      0.81    0.71    0.76     623

accuracy      0.86    0.86    0.86    2925
macro avg     0.86    0.86    0.86    2925
weighted avg  0.87    0.87    0.87    2925
```

## 2-3 文件分類\_Logistic Regression

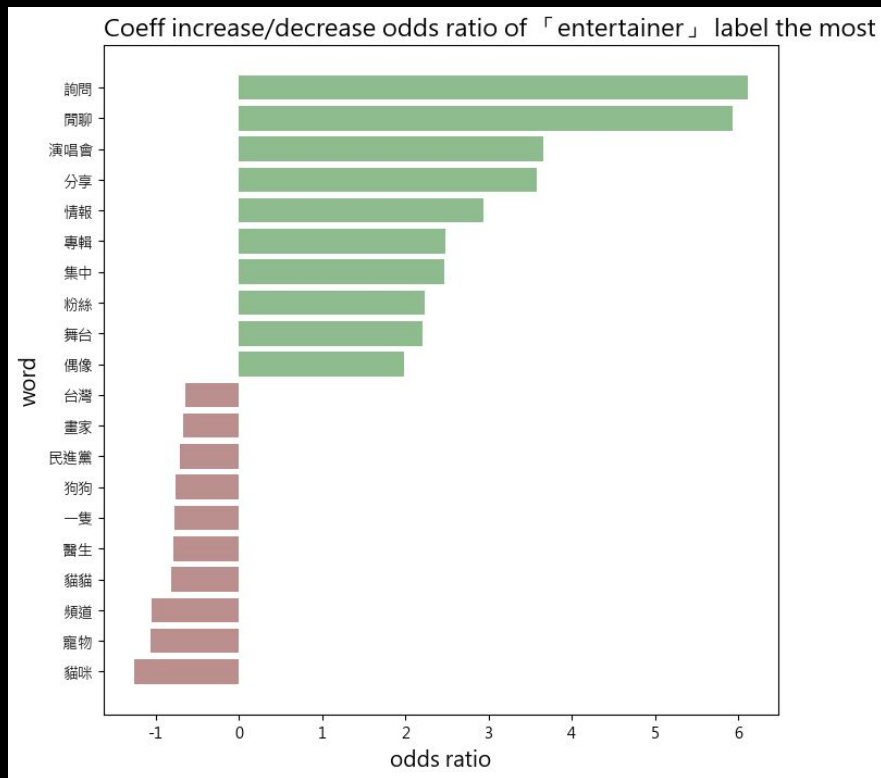


```
now training: clf_logistic
              precision    recall  f1-score   support

entertainer   0.94      0.92      0.93       577
    pet       0.93      0.91      0.92      770
trending      0.86      0.94      0.90     955
youtuber      0.83      0.76      0.79     623

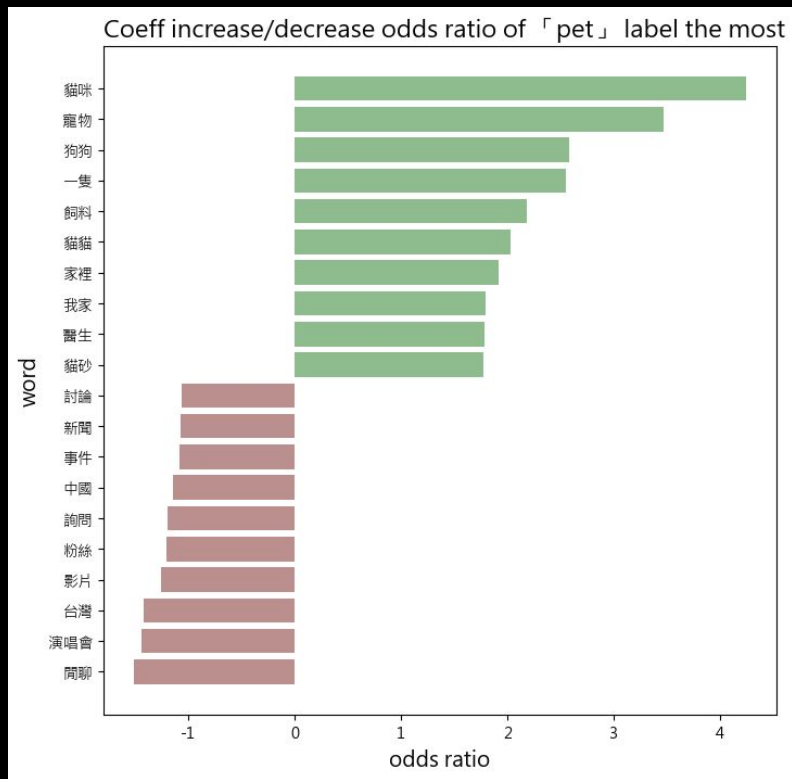
accuracy      0.89      0.88      0.89    2925
macro avg     0.89      0.88      0.88    2925
weighted avg  0.89      0.89      0.89    2925
```

## 2-4 文件分類\_名詞特徵(追星版)

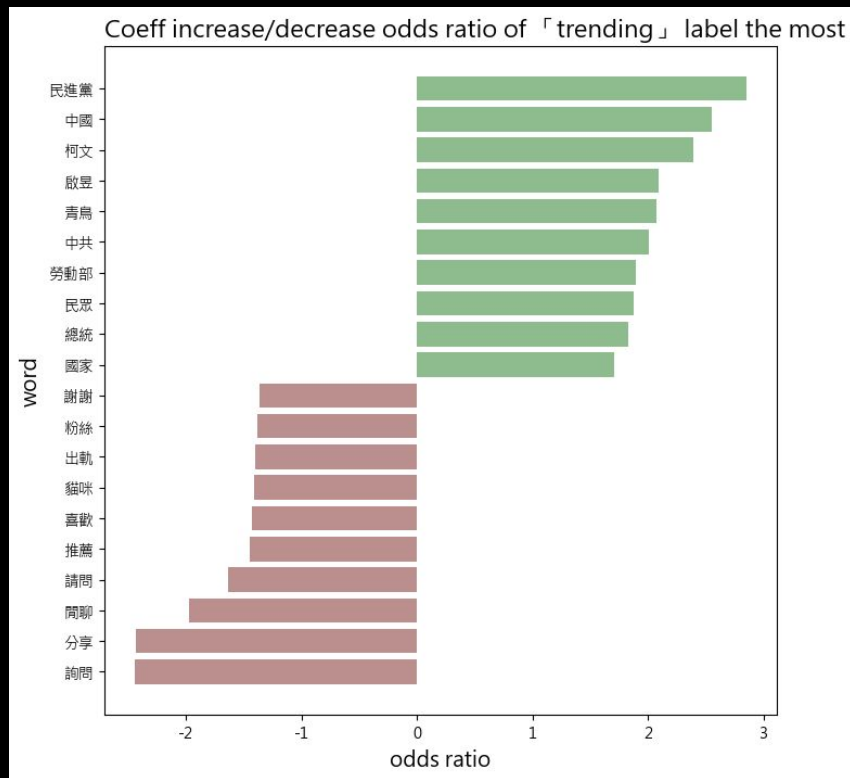




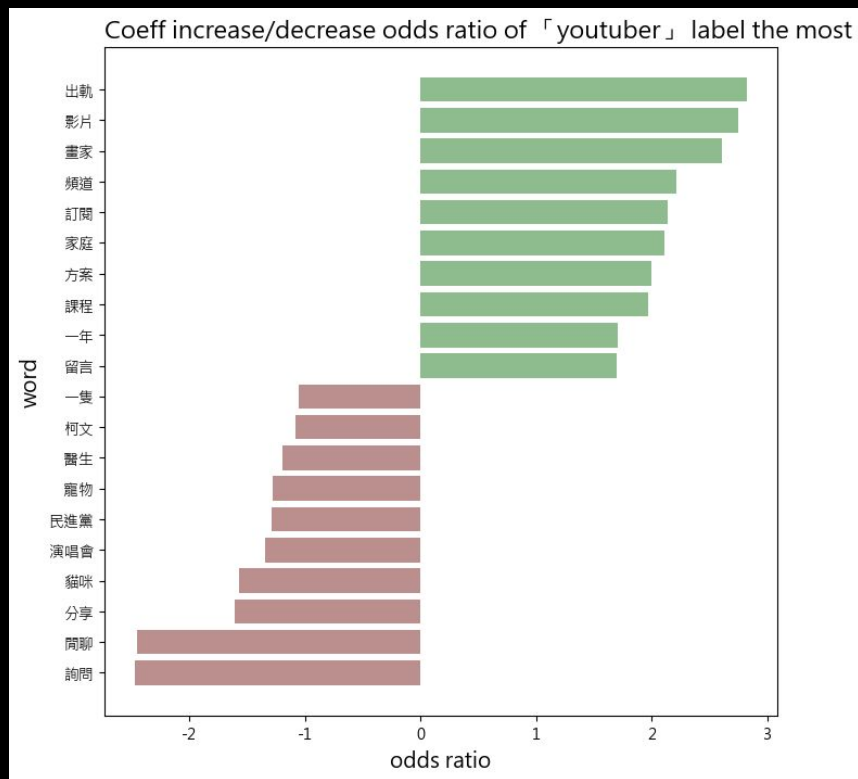
## 2-4 文件分類\_名詞特徵(寵物版)



## 2-4 文件分類\_名詞特徵(趨勢版)



## 2-4 文件分類\_名詞特徵 (youtuber版)



## 2-5 文件分類\_預測分類

- 資料來源：Dcard
- 版別：追星、寵物、時事、Youtuber
- 時間範圍：2025/03/01 ~ 2025/03/15
- 資料筆數：2598
- 採用模型：邏輯回歸模型

## 2-5 文件分類\_分類模型評估

	precision	recall	f1-score	support
entertainer	0.94	0.89	0.91	284
pet	0.92	0.89	0.91	376
trending	0.82	0.95	0.88	542
youtuber	0.82	0.66	0.73	265
accuracy			0.87	1467
macro avg	0.88	0.85	0.86	1467
weighted avg	0.87	0.87	0.87	1467

('YouTuber 必用 AI 查證 工具 避免 說錯 資訊 信錯 假業配 今天 分享 兩個 實用工具 YouTube 媒體 最怕 無非 第一 說錯 觀眾 ' 攻擊 公信力 受損 第二 信錯 業配 背上 法律 風險 兩個 痛點 分鐘 努力 付諸流水 賠錢 影響 事業 說錯 風險 YouTuber 熱點 潮流 ' 時事 搶快 往往 只會 簡單 Google 幾篇 文章 認為 消息 真的 拍片 發表意見 消息 推翻 影片 變成 錯誤 資訊 來源 觀眾 信任感 下降 ' 嚴重 觀眾 更會 舉報 同行 公開批評 信譽 谷底 難以 翻身 信錯 風險 YouTuber 接業配 機關 企業 合作 確定 對方 可靠 合作方 詐騙 ' 問題 頻道 拖下水 影響 法律責任 看到 案例 理財 YouTuber 投資 平台 業配 這家 公司 證實 龐氏 騙局 這位 YouTuber 觀眾 ' 痛罵 幫助 騙子 帶貨 最終 公開 道歉 警方 喝茶 麻煩事 一籬 YouTuber 知道 自保 需要 查證 問題 網上 資訊 真假 混雜 時間 比對 ' 來源 說法 確認 真偽 資訊 藏得 光靠 Google 搜尋 不到 真正 原始 來源 嚴重 搜尋 容易 SEO 操作 騙子 花錢 資訊 Google ' 第一頁 真的 找到 有用 資料 還要 逐篇 長文 閱讀 想想 特別 還要 花費 時間 精力 製作 影片 額外 時間 成本 有沒有 快速 方法 答案 善用 ' AI 今天 我要 介紹 兩個 工具 解決 問題 而設 兩個 工具 ChatGPT Perplexity ChatGPT 目前 ChatGPT 已經 原本 ' 要加 200 美元 Deep Research 功能 下放 Plus 用戶 使用 理解 需求 自動 閱讀 總結 最後 自動 生成 一份 完整 調查報告 ' 全程 30 分鐘 完成 Perplexity Perplexity AI 搜尋器 奠基者 App 專為 搜尋 研究 設計 提供 正確 來源 上線 類似 ' ChatGPT Deep research 功能 自動 蒐集 整理 重要 Perplexity 免費 能夠 使用 Deep research 高性價比 ' 選擇 使用 以下 實用 Prompt 搭配 ChatGPT Perplexity Deep research 使用 熱點 資訊 查證 確保 新聞 時事 ' 資訊 正確 確認 這則 新聞 真實性 新聞標題 內文 提供 來自 官方 來源 政府 學術 機構 知名 媒體 相關 資訊 交叉 比對 觀點 確認 說法 ' 是否 推翻 避免 跟風 錯誤 資訊 查證 這則 消息 是否 後續 澄清 修正 新聞標題 說法 確認 這則 資訊 是否 有效 是否 已有 更新 研究 官方 ' 聲明 推翻 說法 背景 調查 快速 查核 業配 品牌 合作方 信譽 分析 這家 公司 品牌 公司 名稱 官網 提供 品牌 註冊 資訊 過往 新聞 用戶 ' 評價 是否 詐騙 爭議 紀錄 合作 合約 風險 分析 避免 簽下 霸王 條款 分析 這份 合作 合約 找出 公平 條款 潛在 風險 合約 內容 標記 ' 吃虧 條款 提供 修改 建議 確保 合約 有利 總結 一下 YouTuber 具備 AI 查證 能力 保護 希望 說錯 觀眾 質疑 專業 希望 接到 ' 爛業配 影響 信譽 觀眾 帶貨 騙子 希望 辛苦 影片 最後 假消息 詐騙 公司 毀掉 現在 試試 AI 用來 查證 YouTube 內容 專業 可靠 ' 整理 一份 AI 查證 提供 Prompt 提問 技巧 查詢 方便 歡迎 索取 記得 AI 決定 節省 蒐集 閱讀 資料 時間 精力 整個 過程 變得 ' 簡單 快捷 作出 明智 決定')

# 3. Week\_9 主題模型

## 3-1 文件分類\_資料集

- 資料來源：Dcard
- 版別：追星、寵物、時事、Youtuber
- 時間範圍：2024/11/01 ~ 2024/11/30
- 資料筆數：4488

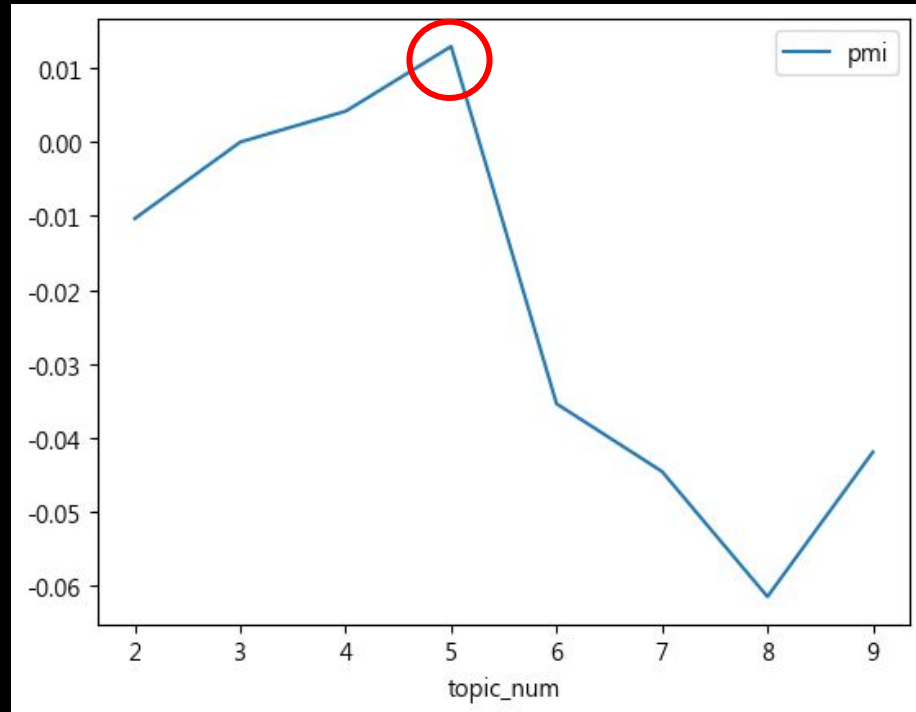
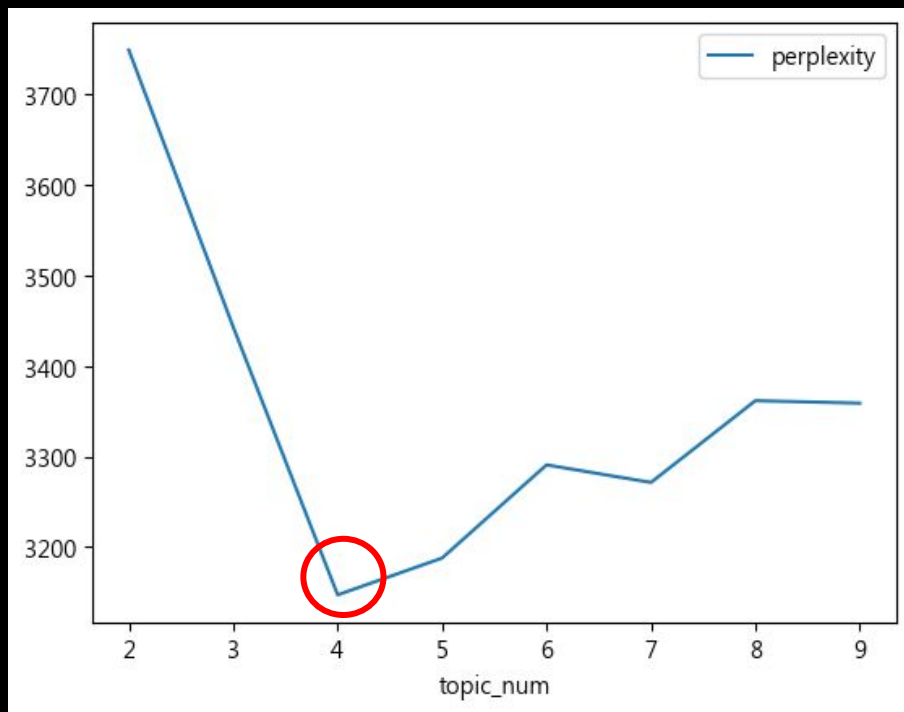


## 3-2 查看LDA主題代表字

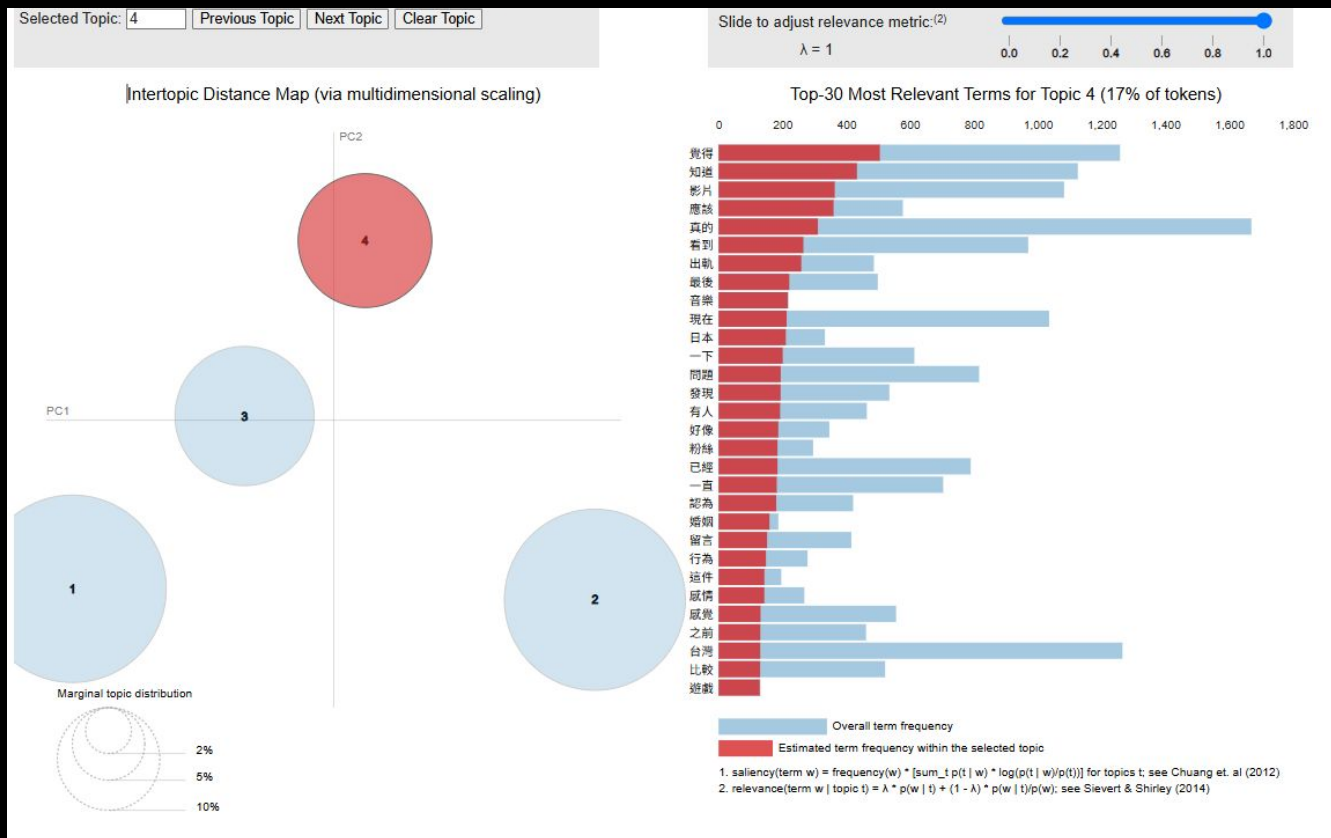
- 主題一:畫家事件
- 主題二:出軌相關
- 主題三:家庭相關
- 主題四:遊戲直播相關
- 主題五:youtuber相關
- ...
- 主題十:政治相關

類別「時事」的佔比  
為0.77

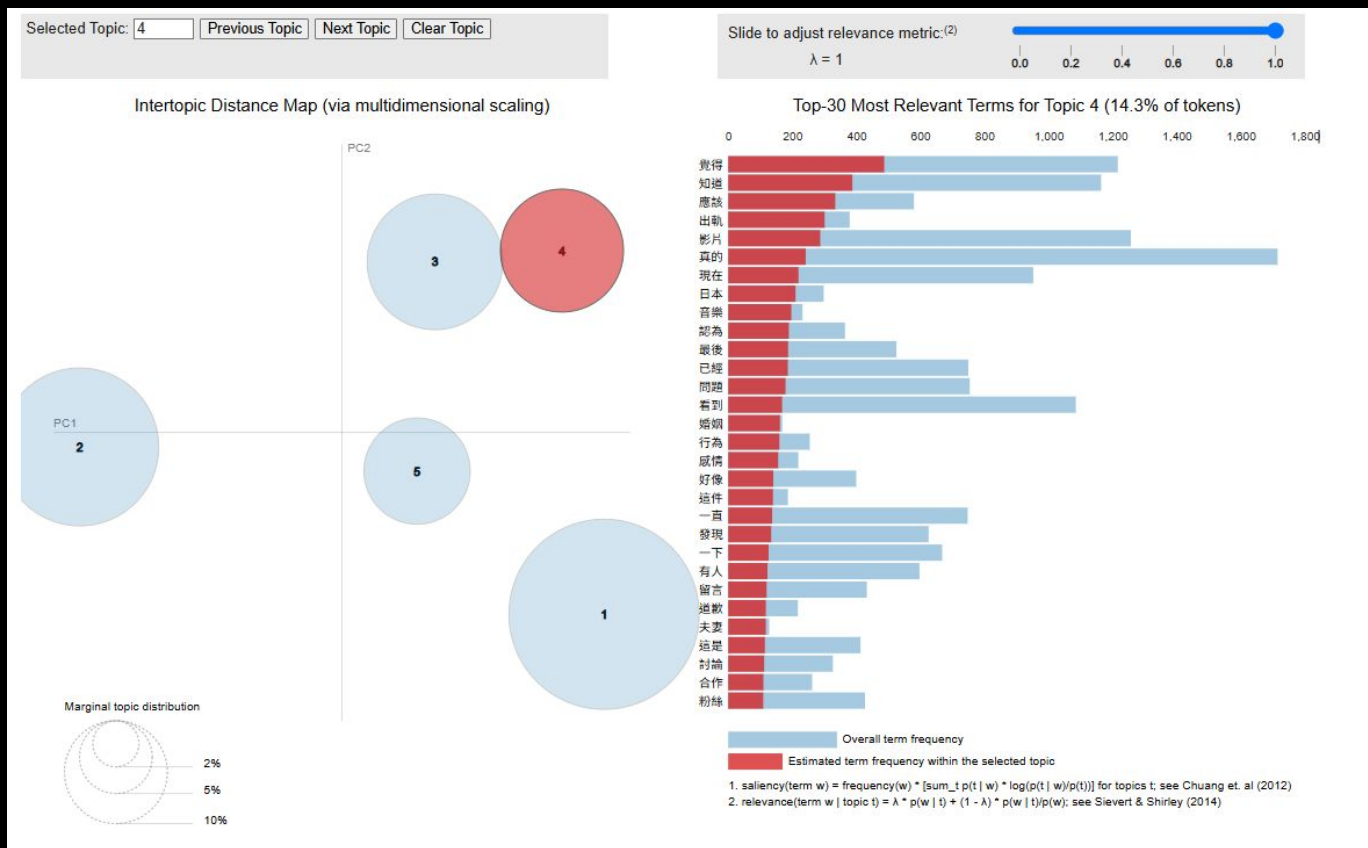
### 3-3 透過LDA模型指標 (perplexity、PMI) 找出最佳模型



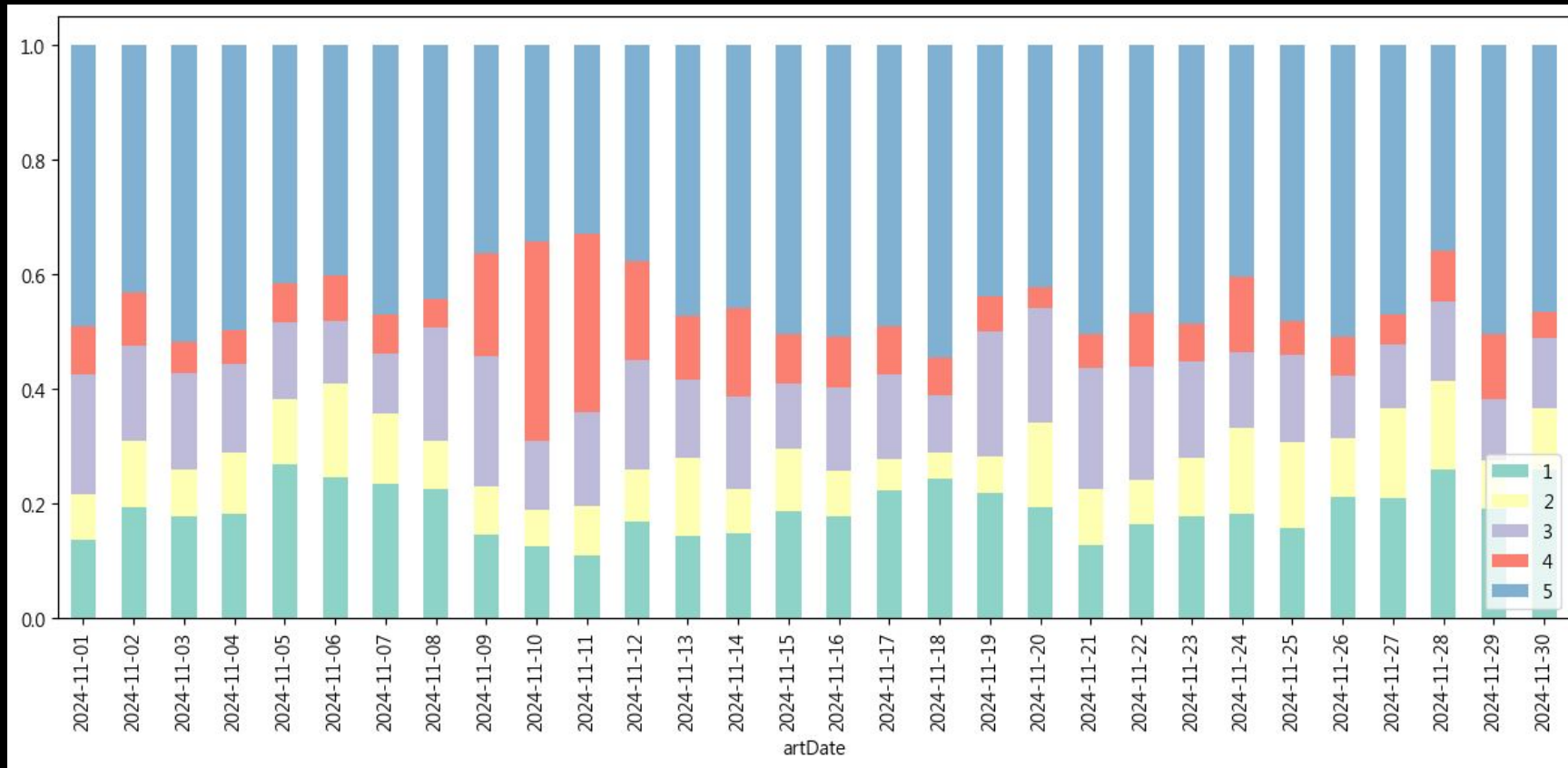
# 3-4 視覺化呈現 (主題數:4)



# 3-4 視覺化呈現 (主題數:5)



# 3-5 視覺化呈現文章趨勢變化



## 3-5 視覺化呈現文章趨勢變化

358 發一個陰陽怪氣的文章來嗆我你講說演唱會什麼嘴形沒對上真不知道自己家的也是對嘴而已嗎但為啥自己...  
392 之前覺得中粉在視訊簽售要求的掐脖子好噁原來我討厭的不是掐脖子而是土味音樂和沒有美感的舞蹈還有...  
398 想不到欸賴耀翔的票數斷層上升把帝翁擠出去了也算是皇天不負苦心人吧等待了這麼久應該是離出道最近...  
542 剛剛看到一個留言想反駁但是也不知道要反駁什麼原本影片是有出現福富月直拍底下留言就有留言到一個...  
2420 多慧再度登上新聞版面那看看多慧害怕的神情到底是哪支哥布林跟蹤偷拍啦還是女哥布林忌妒派人去的  
2478 看到一堆民進黨支持者或是其他人一直講敢跟中共打仗還笑別人投降主義但叫他敢露臉拍影片幹中共嗎就...  
2489 如題最近偶然在隔壁版看到這篇是否該全面禁止支語抖小紅書雖然老生常談不過因為沒有標準答案一直...  
3791 交往超過十年的前妻大爆料帶新人帶到床上去了大書之後也出來拍影片道歉了大書這一對沒有小王另一對...  
3809 別再出來硬凹丟女生的臉了  
3828 結論就是那位中國女畫家很雷這一切的源頭其實並不是一開始發的那篇文因為大家早就看出他們兩人的關...  
3844 女畫家反擊一剛開始發現老婆出軌時渾身發麻卻沒有對她發過任何一次脾氣還自責說可能我給不了想要的...  
3866 從昨天被小三爆料開始至今完全神隱超過小時的機會你經紀公司擬不出一篇稿來沒做可以快點否認有做可...  
3873 最終更文更補充小細節進來昨天看下來發現好多事情爆發才加入的吃瓜民眾不分青紅皂白的亂噴所以想要...  
3874 雖然一看到和的事情我也蠻驚訝但人的感情本來就會隨時間因為遇到不同的人事物而發生改變不管是感情...  
3876 四個人都非常糟糕呢許多其他帖文已經指出了的不當行為而加以分析我認為我有必要再發千篇一律的文章...  
3882 就結果論來說都各自再找了另一半當然現在的方向會有人希望找到他們在婚姻中是誰先出軌誰先對不起誰...  
3883 你說都很早睡或熬夜工作不知道自己昨晚湧進大量留言所以消失可以理解結果現在都隔天中午了日本甚至...  
3884 再次更新我對發文和二度發文的判讀三次私會沒有繼續聯絡了藉這個機會和他們個人解讀很隱晦地說他們...  
3886 他們兩個放在一起更像是夫妻有相同的氣質他們一起拍片應該蠻有趣的是說老觀眾應該知道其實很好色  
3888 如果只有一方的話早就吵翻了吧真是這樣和平離婚也好

**感謝聆聽**