



---

# 社群媒體分析第三次讀書會

題目：公眾人物八卦新聞

指導老師：黃三益教授

第三組 鍾承玫 N074220001  
羅育琦 N094220002  
楊雯琿 N094220006  
周昆穎 N094220007  
陳冠汶 N094220012  
李明章 N094220023  
鄭雅丹 N094220027



## 一、動機和分析目的

探討民眾對於藝人的私生活負面新聞，有助於給世人謹惕，因此探討民眾對於此事件的情緒分析做法具有一定的參考價值。

## 二、資料集的描述

因本組欲探討民眾對於近期知名藝人負面新聞的議題討論及情緒分析，從工作流程平台既有資料來源爬取 PTT 在 Gossiping(八卦)討論板的相關議題，考量資料量過多，僅抓取在 12/20~12/21 區間有關王力宏及其妻李靚蕾的資料。

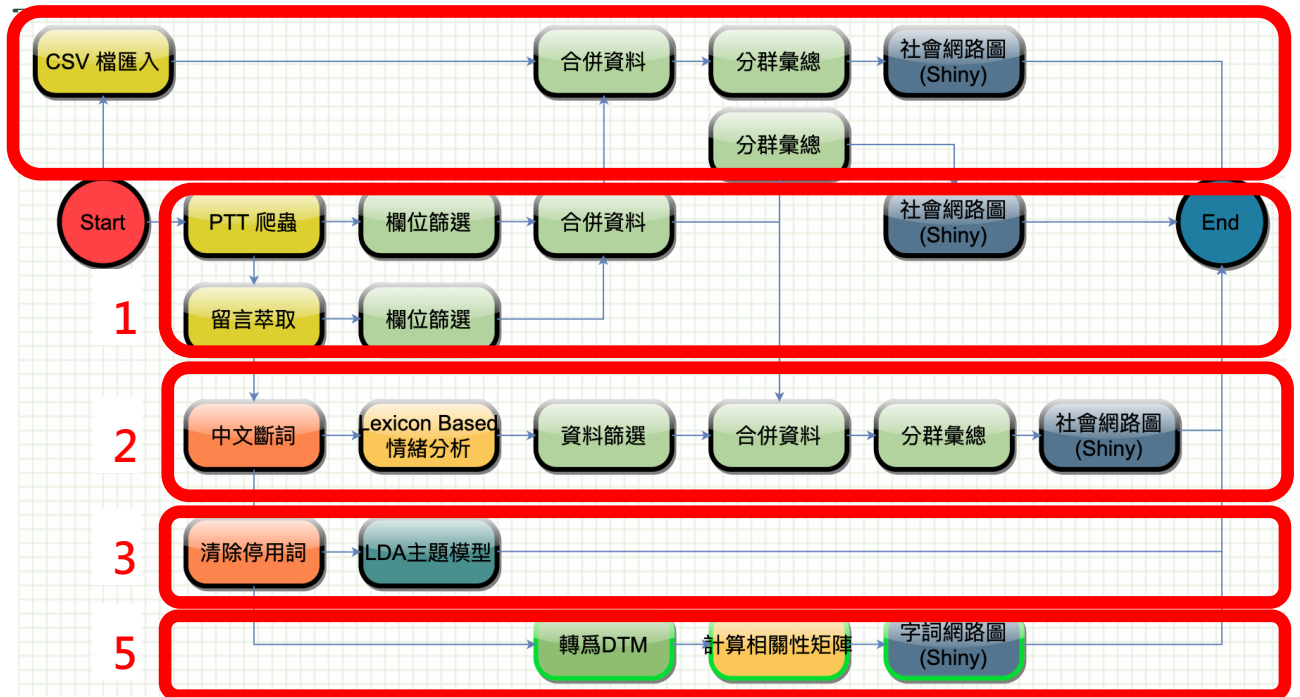
## 三、工作流程分析

本次作業流程包含了 5 項的分析作業，

1. 首先是從 PTT 文本探討民眾對於此事件的討論程度，並做成社會網路圖。
2. 接下來探討民眾對於此事件的情緒影響分析，同樣做成社會網路圖。
3. 進行主題模型分析並定義其主題名稱。
4. 有了主題模型的資料之後，可以知道這一篇文章主要是探討哪一個主題，並將其資料先行分類後，拿來分析回應者都回應了哪些文章上，也以社會網路圖來做呈現。
5. 最後計算相關性矩陣，進行字詞網路圖的分析。



4



爬蟲結果總計取得 695 筆的文章筆數。

## PTT 爬蟲 (4)

參數設定

### 統計資訊

10

欄位數



695

資料筆數



留言萃取流程則取得 42,112 筆留言資料，平均每一篇文章有 61.39 的留言數。



## 留言萃取 (28)

參數設定

Input - 4

### 統計資訊

42112

總留言數量

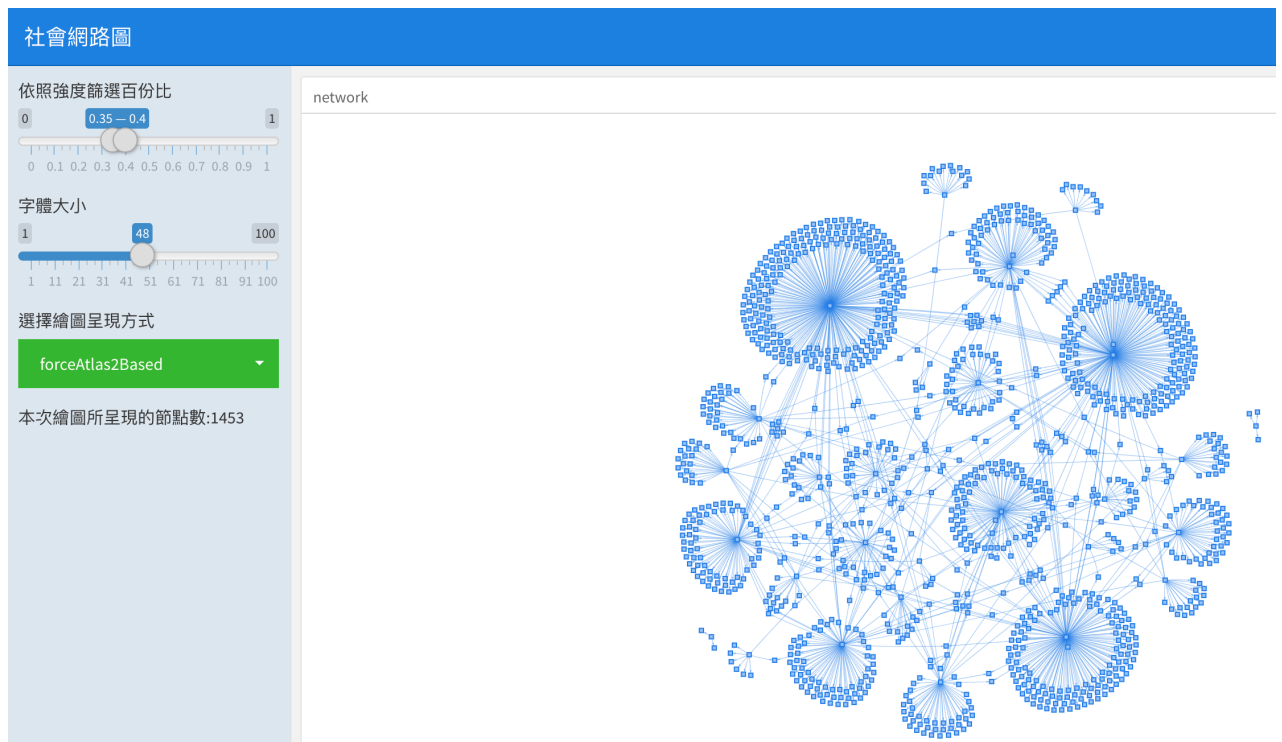


61.39

平均單篇留言數



將 PTT 文本數和留言萃取流程，保留 PO 文者和留言者的欄位之後進行合併，並做分群匯總，最後做成社會網路圖，繪出各個留言者在一篇文章留言的數量。



本組也探討了民眾對於此事件的情緒分析影響，首先在中文斷詞的流程中處理留言的內容，並將此事件的關鍵人物設定較高的權重。



## 中文斷詞 (22)

參數設定	Input - 28
<b>選擇處理欄位 *</b> cmtContent	<b>定義詞彙 ⓘ</b> 王力宏 1000 李靚蕾 1000

在 Lexicon Based 情緒分析自定義正/負面詞彙，正面詞彙有支持、接受、喜歡、有錢等。負面詞彙則將離婚、支持、渣男、噁心等詞彙納入。

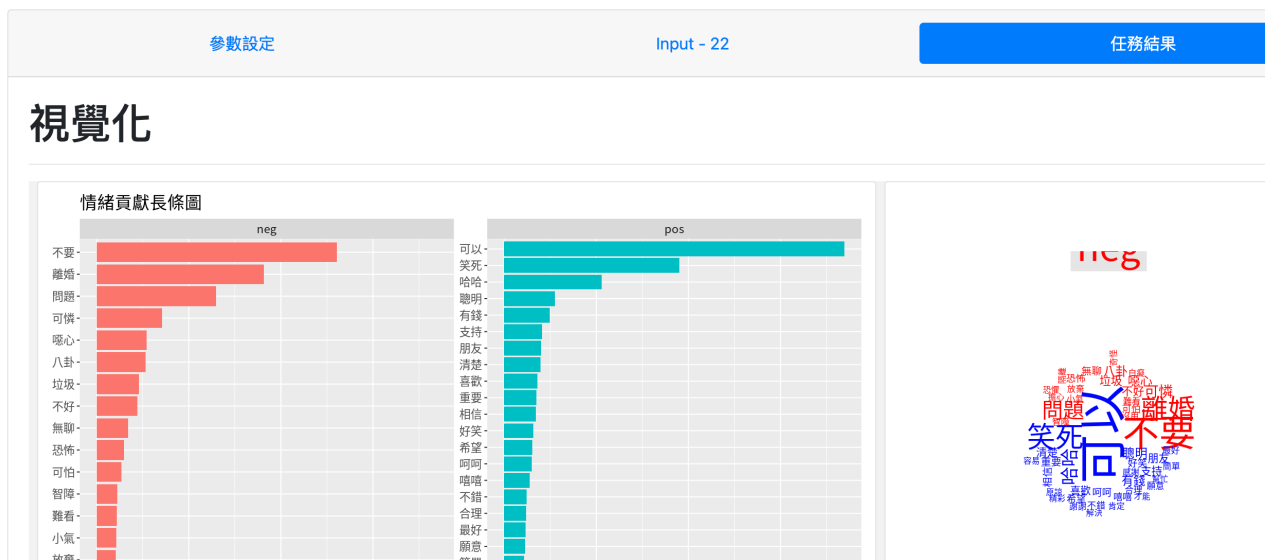
## Lexicon Based 情緒分析 (51)

參數設定	Input - 22
<b>語言 *</b> Chinese	<b>載入預設情緒字典 *</b> 是
<b>定義正面詞彙 ⓘ</b> 支持 接受 喜歡 有錢	<b>定義負面詞彙 ⓘ</b> 離婚 八卦 渣男 噁心

將情緒分析視覺化可以看到，離婚、問題、和渣男皆排在負面情緒的前幾名。



## Lexicon Based 情緒分析 (51)



接下來把情緒分析結果中情緒值小於 0 的負面情緒篩選出來。

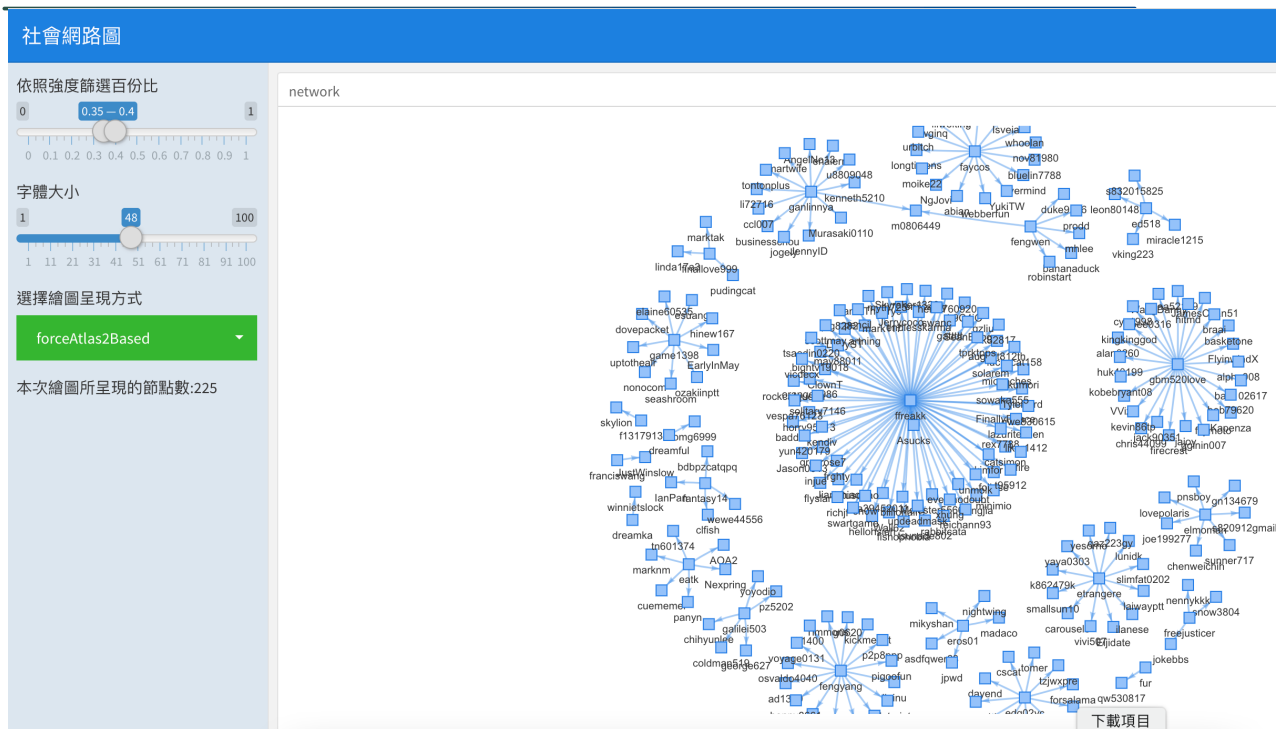
## 資料篩選 (53)

參數設定

條件式 \* ⓘ

`'$sentiment_value<0'`

和 PTT 文本進行合併後，將 PO 文者和留言者分群做成社會網路圖，除了取得 PO 文者和留言之間的網路關係之外，也計算出留言者在這一篇文章的留下負面情緒的留言數。



本組也進行主題模型分析，將其文本分為 5 個主題，詞彙頻率上下限分別設定 0.5、10，其餘均為系統預設值。並定義其主題名稱。

### 📄 LDA主題模型 (68)

參數設定	Input - 65	任務結果
<b>目標欄位 *</b> result	<b>迭代次數</b> 50	
<b>主題數 *</b> 5	<b>主題保留關鍵字數量</b> 20	
<b>詞彙頻率下限 ⓘ</b> 10	<b>詞彙頻率上限 ⓘ</b> 0.5	
<b>alpha</b> 預設為主題數/50	<b>Beta</b> 預設為0.1	
<b>chucksize ⓘ</b> 預設為2000	<b>update_every ⓘ</b> 1	

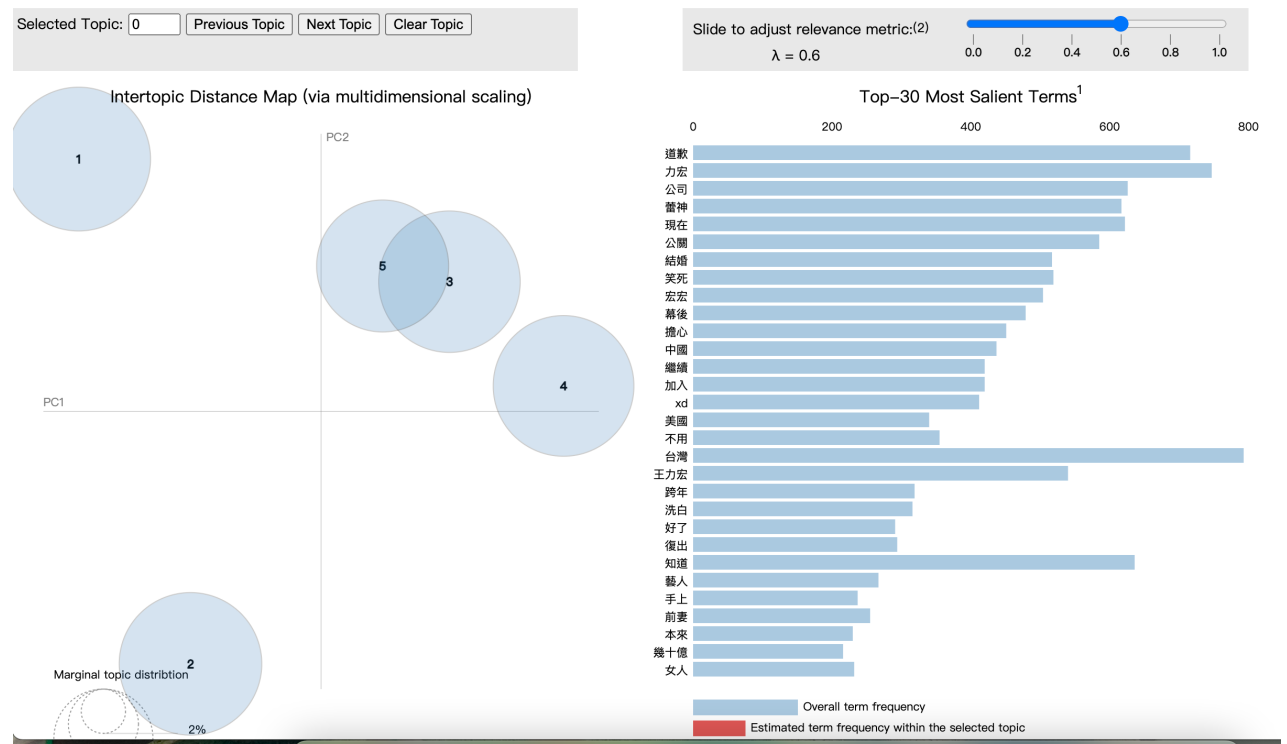
並定義其主題名稱。

- 1.當事人經紀公司的後續動作。
- 2.學音樂的小孩不會變壞？
- 3.渣男的復出(花蓮？)



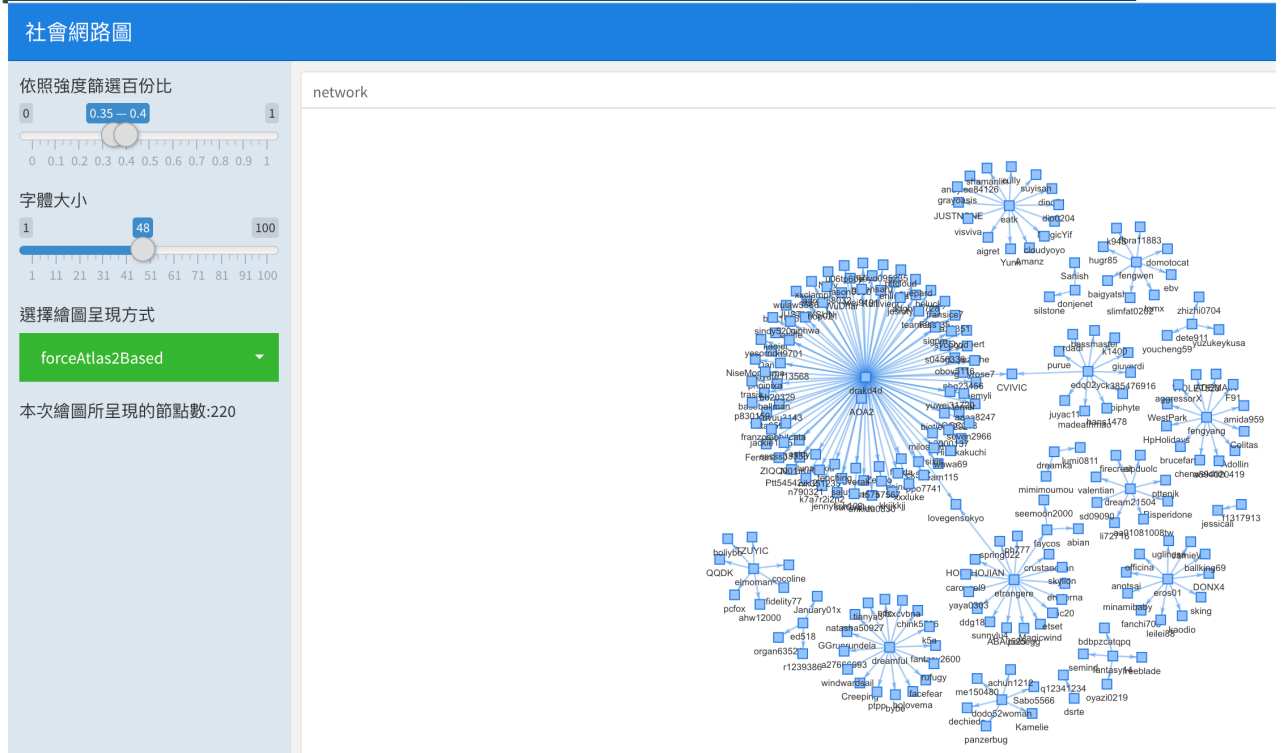
4.道歉聲明在社群媒體的反應。

5.此事件之後當事人要損失多少財物。



如同前面提到，有了主題模型的資料之後，可以知道這一篇文章主要是探討哪一個主題。所以將其資料先行分類，和 PTT 文本合併後，拿來分析回應者都回應了哪些文章上，也以社會網路圖來做呈現。





## 三 字詞網路圖 (Shiny) (87)

