# 社群媒體分析

**讀書會報告**

**主題：郭台銘總統之路輿情分析**

**第12組**

**黃足維 N104220007**

**王家得 N104220008**

**吳尚蓉 N104220010**

**呂伉弘 N104220011**

**陳韻元 N104220015**

**林宗憲 N104220020**

**陳聖宗 N104220029**



# 目錄

第1章 探討主題

第2章 資料範圍、來源、關鍵字

第3章 分析說明

第1節 TARFlow流程

第2節 LDA主題模型

1. LDA主題分佈
2. 郭朱配
3. 郭侯配
4. 公共議題
5. 政黑
6. 郭柯配
7. 機器人
8. LDA分類

第3節 圖表分析

1. 各版討論趨勢
2. 主題情緒趨勢
3. 主題情緒趨勢-郭O配
4. 討論主題各版分佈
5. 情緒趨勢
6. 主題討論趨勢

第4章 結論

# 探討主題

台灣企業家郭台銘的政治參與歷程。他在 1970 年加入中國國民黨，曾力挺馬英九連任總統。在 2014 年太陽花學運期間，他表示不願意擔任總統，並認為這是害人的位置。在 2019 年，他獲得中國國民黨頒發榮譽狀，並宣布參選國民黨總統初選，但最終輸給了韓國瑜。9 月，他退出中國國民黨，並呼籲支持者投票給親民黨和台灣民眾黨。2023 年 2 月，他表態參選 2024 年的總統大選。本組希望透過TARFlow工作流程平台，來了解郭台銘究竟會對2024年的總統大選帶來什麼樣的變數與影響。

# 第2章 資料範圍、來源、關鍵字

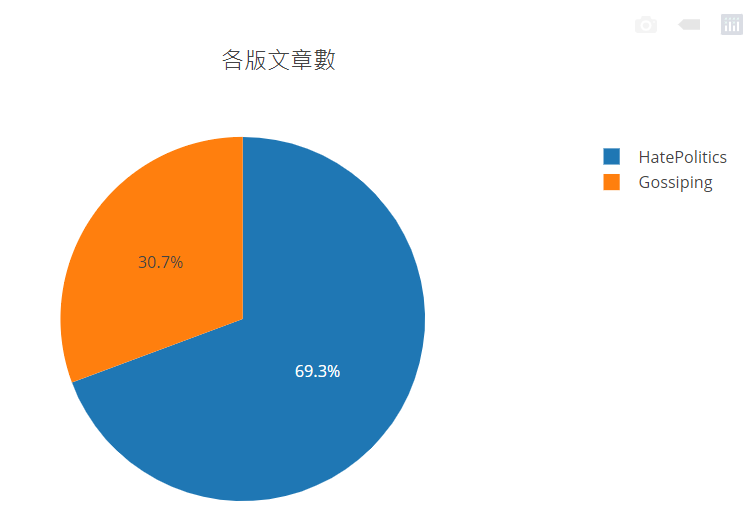
資料範圍、來源、關鍵字：

資料來源：PTT八卦版、政黑版。

資料範圍：2023/1/1~2023/5/5

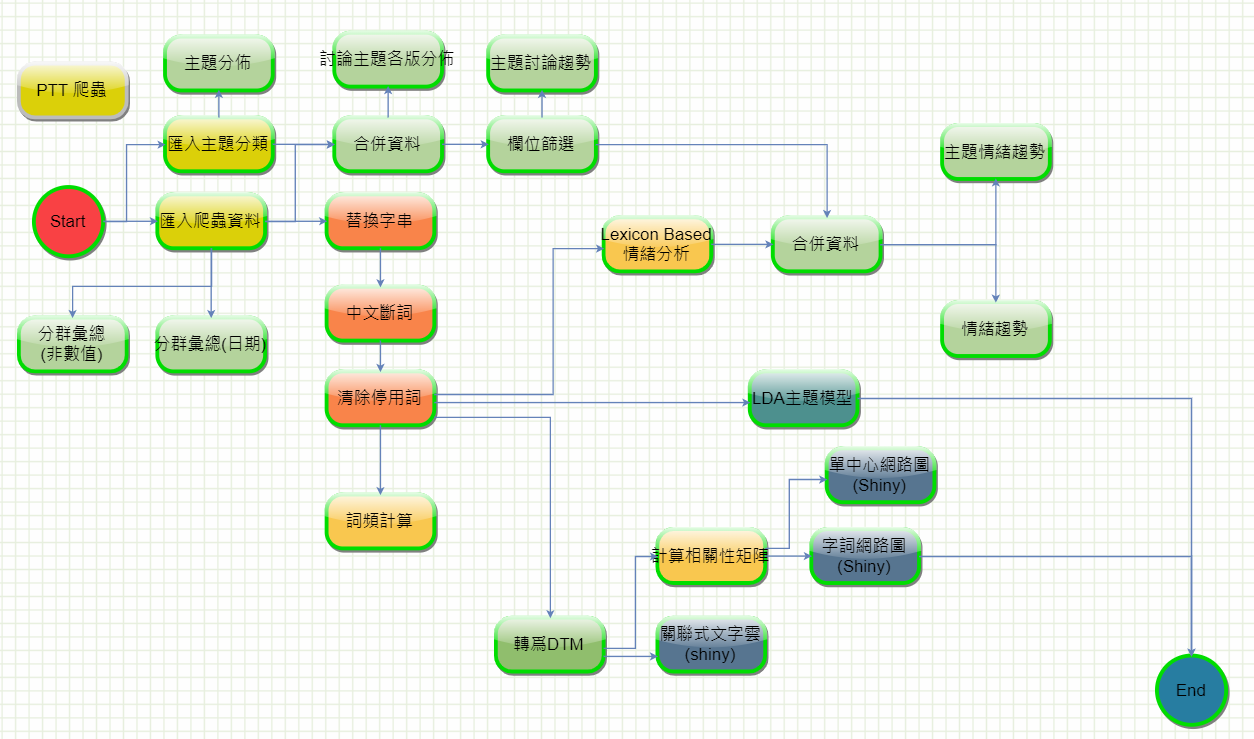
搜尋關鍵字：郭董、郭台銘、果凍

排除關鍵字：國棟

PTT爬蟲任務結果：資料筆4,823筆  
  


# 第3章 分析說明

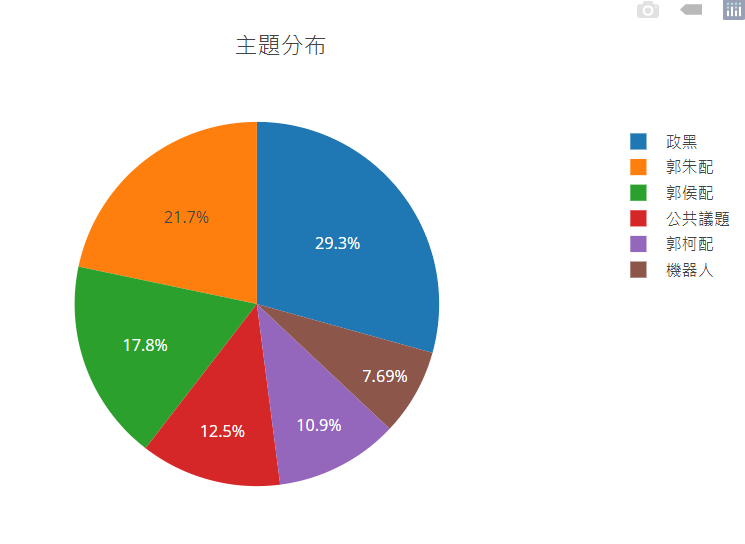
## **第1節 TARFlow流程**

TARFlow工作流程檔名：GoodTiming\_PTT\_Gossip

## **第2節 LDA主題模型**

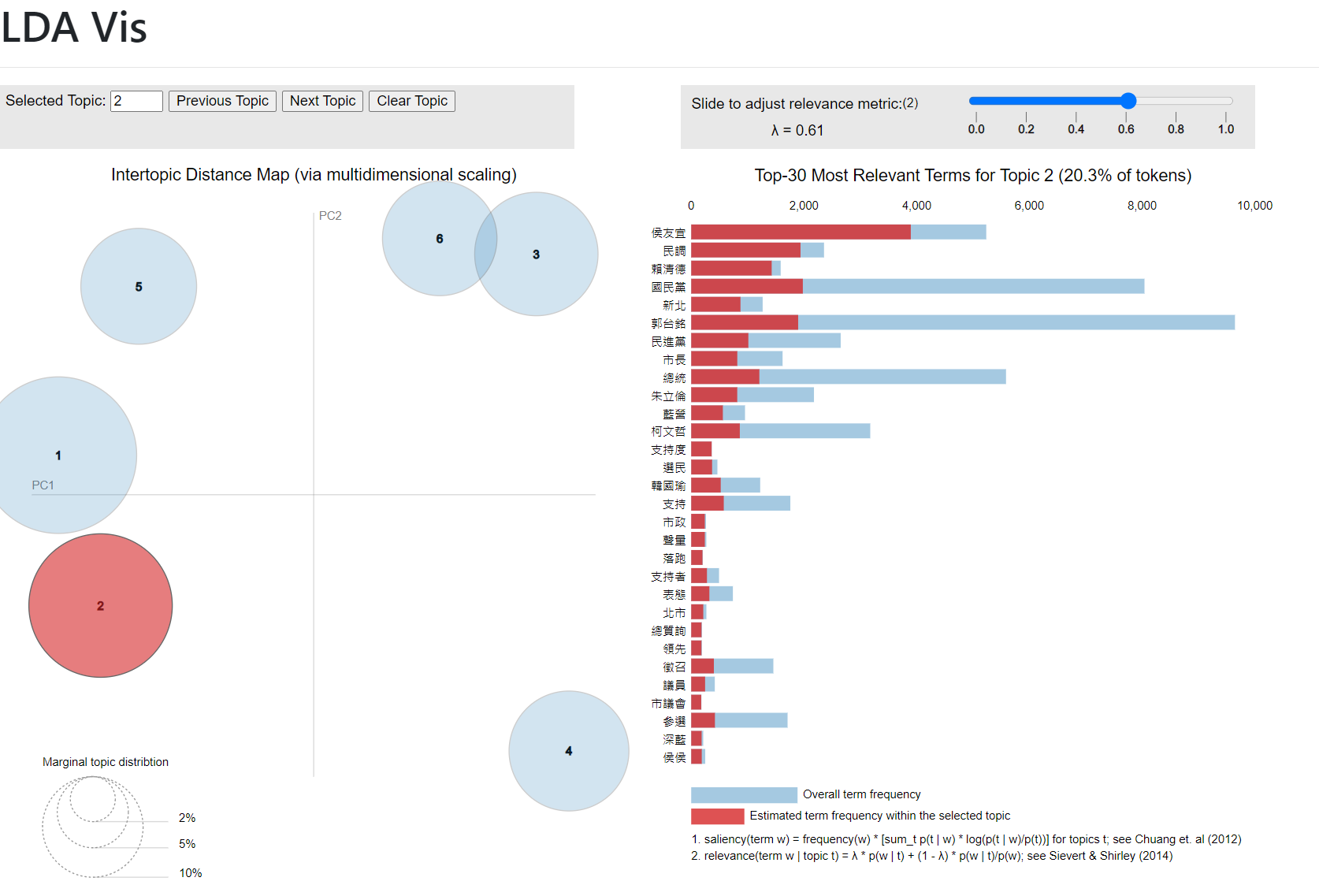
### LDA主題分佈

由主題分佈聲量可看出PTT酸民多，對於政治討論有許多負面評論。其次則是郭朱配聲量比例來到21.7%，第三名則是郭侯配，甚少討論郭韓配。

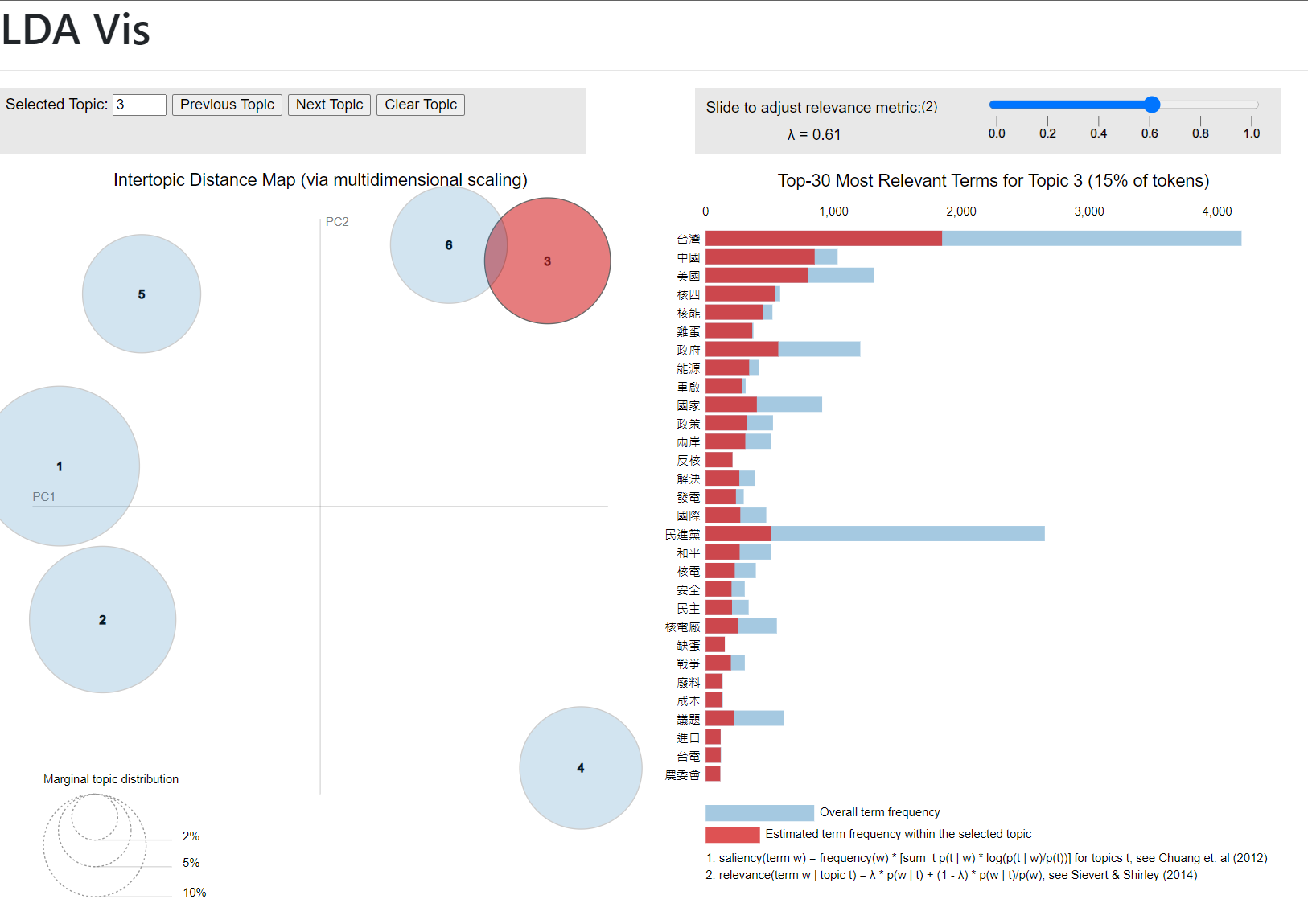


### 第二項 Topic1郭朱配

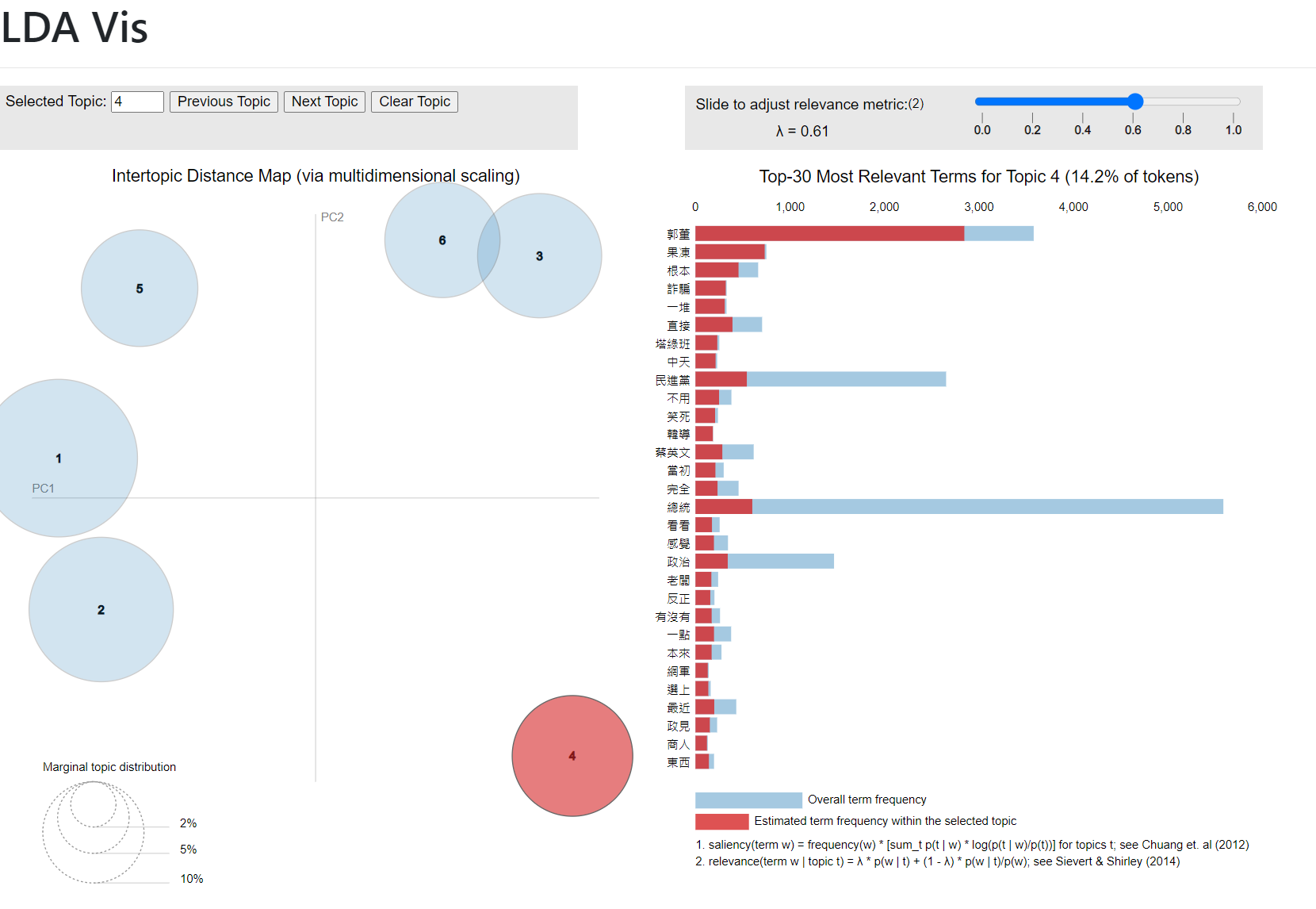
### 第三項 Topic2郭侯配



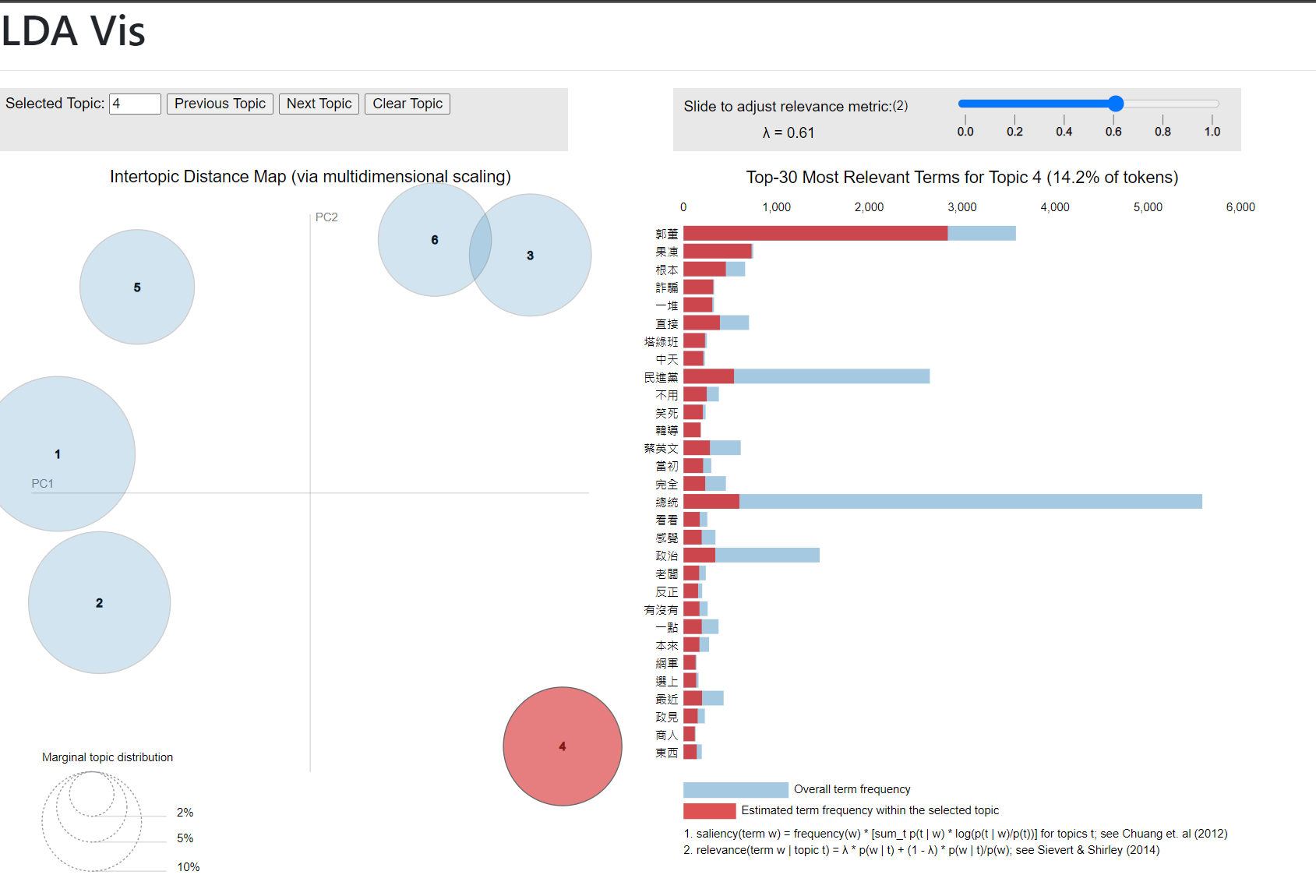
### 第四項 Topic3公共議題



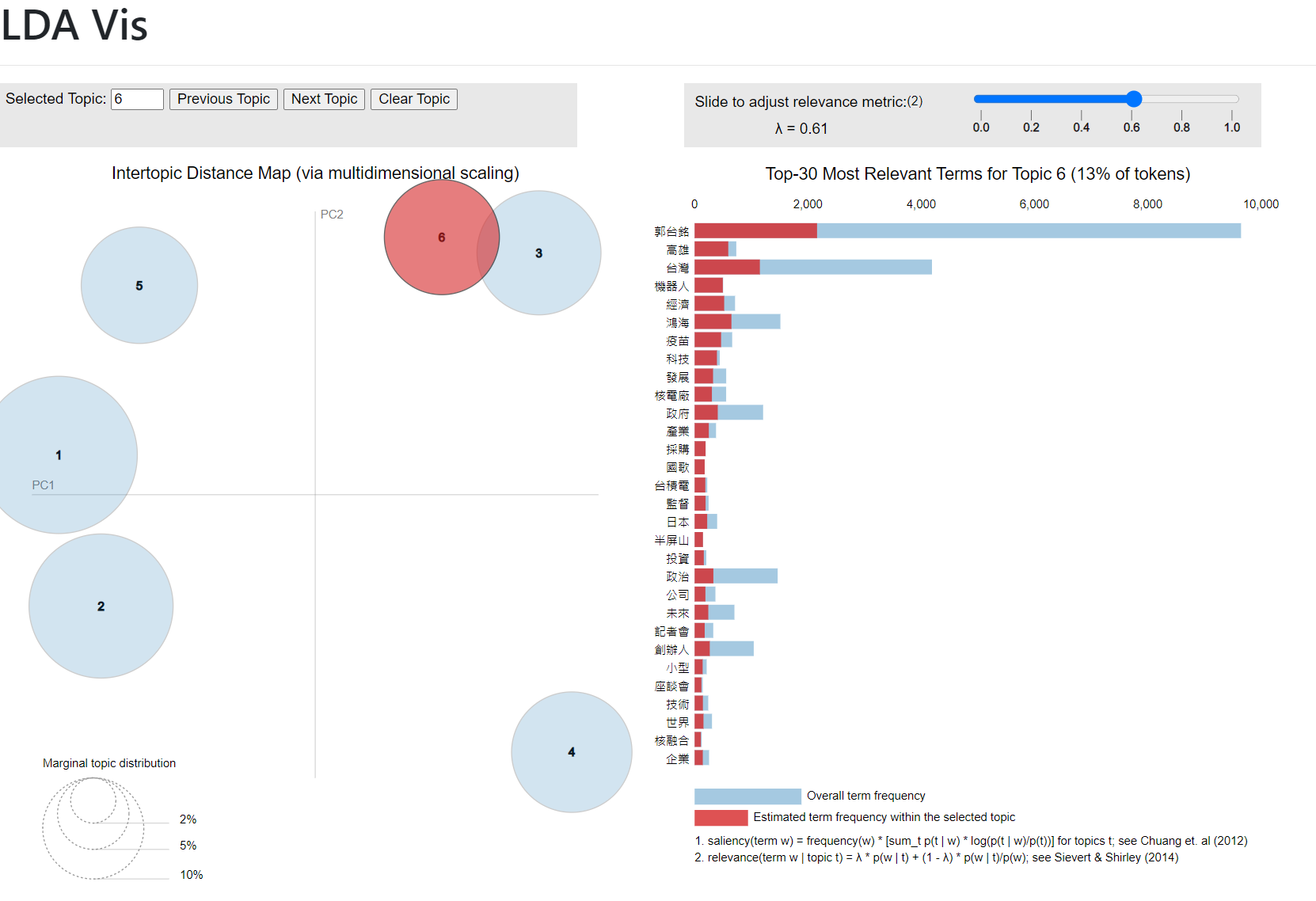
### 第五項 Topic4 政黑



### 第六項 Topic5郭柯配



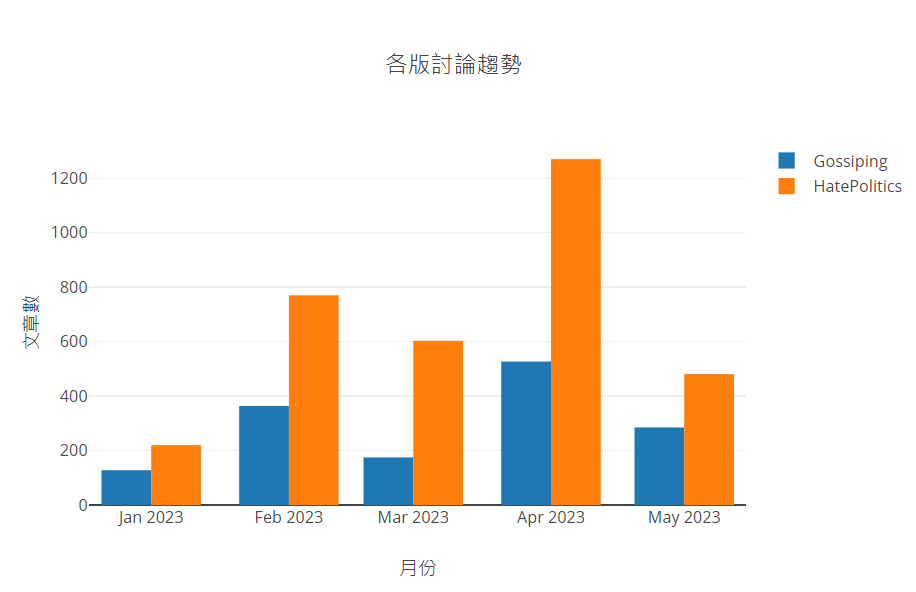
### 第七項 Topic6機器人



### 第八項 LDA分類

## 第3節 圖表分析

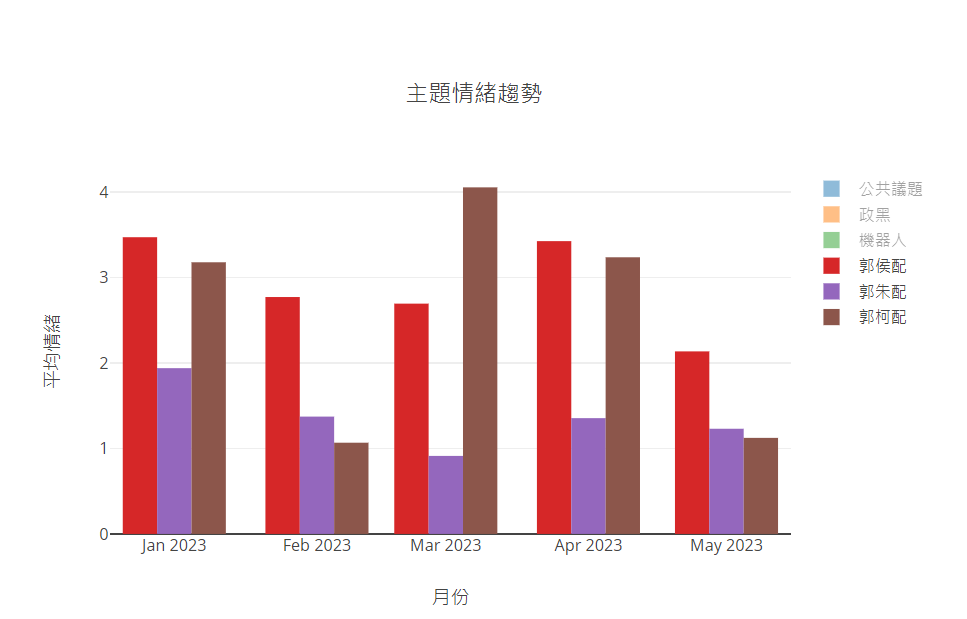
### 各版討論趨勢



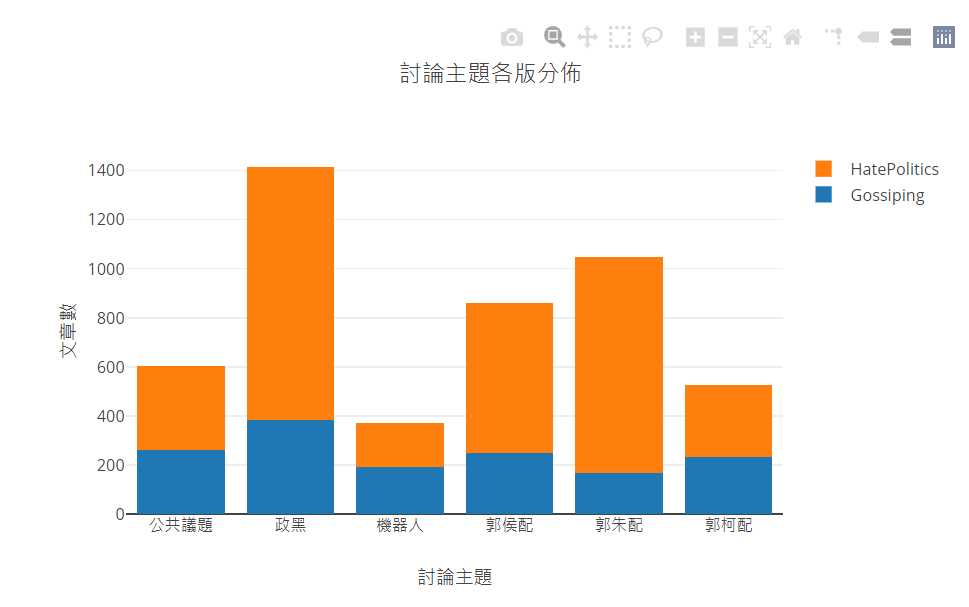
### 主題情緒趨勢

### 

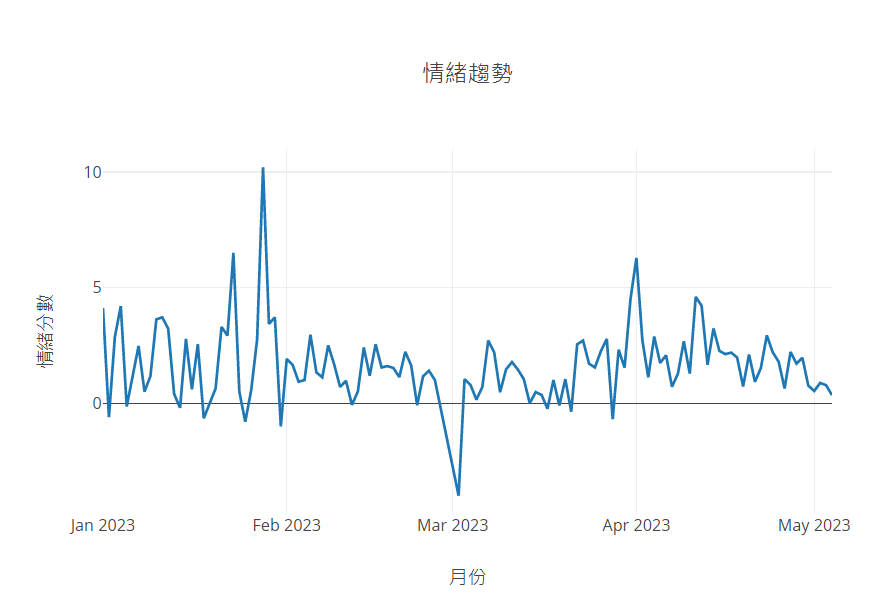
### 主題情緒趨勢-郭O配



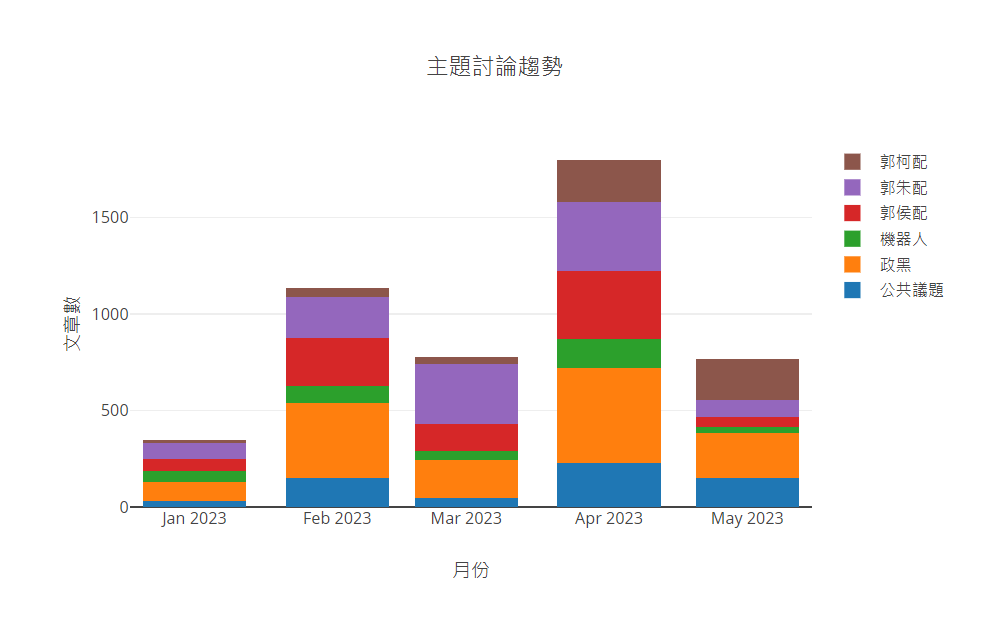
### 討論主題各版分佈



### 情緒趨勢



### 主題討論趨勢



# 第4章 結論

​ 透過LDA主題模型可以從文本中提取出潛在的主題，並且能夠自動分類並識別主題，本組透過LDA設定六個主題，透過LDA快速識別並掌握包括郭朱配、郭侯配、公共議題、政黑、郭柯配、機器人等六大主題分支。

且LDA能夠處理大規模文本數據，對於大規模文本分析非常有用，本組此次主題抓取資料範圍設定為2023/1/1~2023/5/5期間約4個多月，並爬取PTT八卦與政黑板，資料筆數共4,823筆，LDA亦能快速且有效處理

當然LDA尚有不足之處，例如需要預先處理文本數據，例如斷詞、去除停用字等，以便讓模型能夠更好的處理文本。此外在主題解釋性也可能不太精確，雖然後續採人工的方式來做解釋，但仍有可能解釋方向有誤，造成主題誤解，為避免此情形，仍需與主題分類文章抽樣比對，避免錯誤。最後，LDA模型需要調整參數，例如主題數量、λ值等等，這需要有一定的經驗與技巧，以避免主題重疊，分類不明確、相同字詞重複出現於各主題等。

綜上，LDA是一個非常有用的主題模型工具，使用上只要注意LDA工具的侷限與待解決問題，就能夠協助我們在這個大數據的時代，快速的分類與掌握輿情重點，特別是選舉這種風向與討論變化莫測的議題上，助益很大。