# 社群媒體分析

**讀書會報告**

**主題：郭台銘總統之路輿情分析PART2**

**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 視覺化圖表分析

1. LDA主題分佈
2. 討論度
3. 情緒趨勢

第3節 LDA主題模型

1. 初選
2. 總統大選
3. 策略動向
4. 外交政策
5. 兩岸關係
6. 疫苗採購
7. 政策觀點
8. 七大分類前10關鍵字與分類任務結果

第4節 社會網路關係圖

1. 發文與回文者之社會網路關係圖
2. LDA分類之社會網路關係圖
3. Gephi社會網路關係圖by帳號

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：2023/3/1~2023/5/12

搜尋關鍵字：郭董、郭台銘、果凍

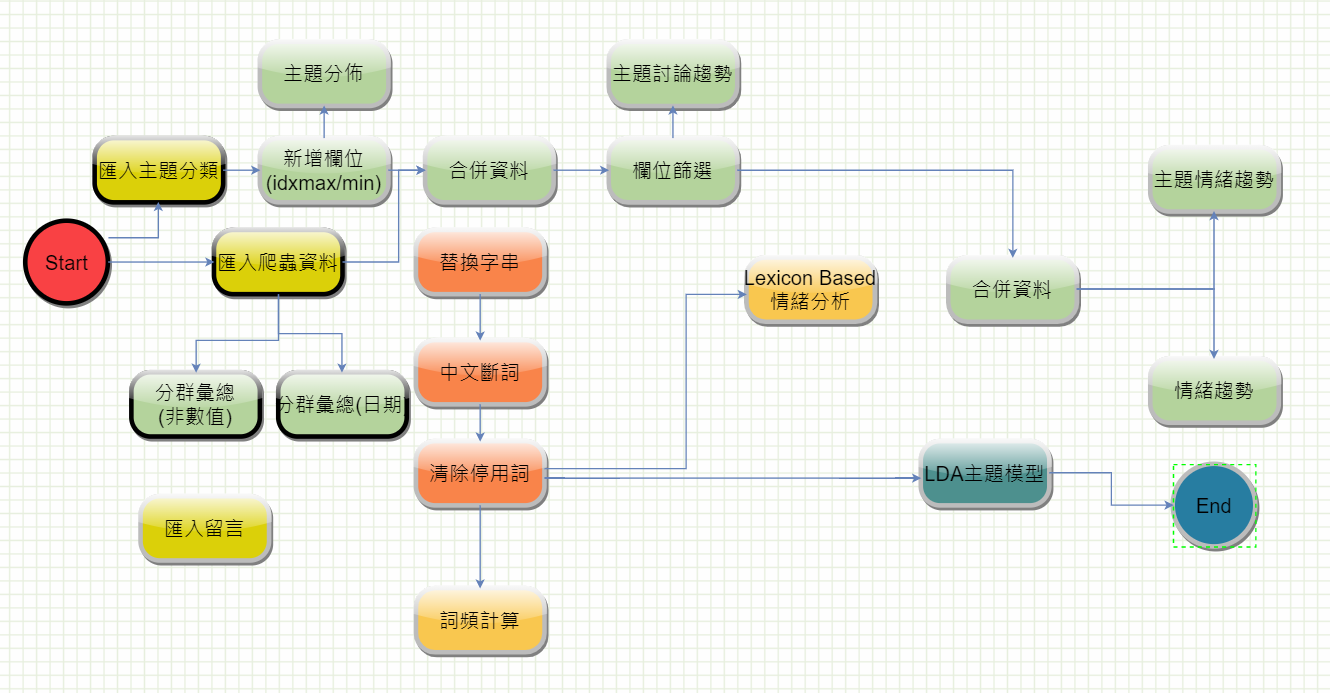
排除關鍵字：國棟

PTT爬蟲任務結果：資料筆4,846筆

# 第3章 分析說明

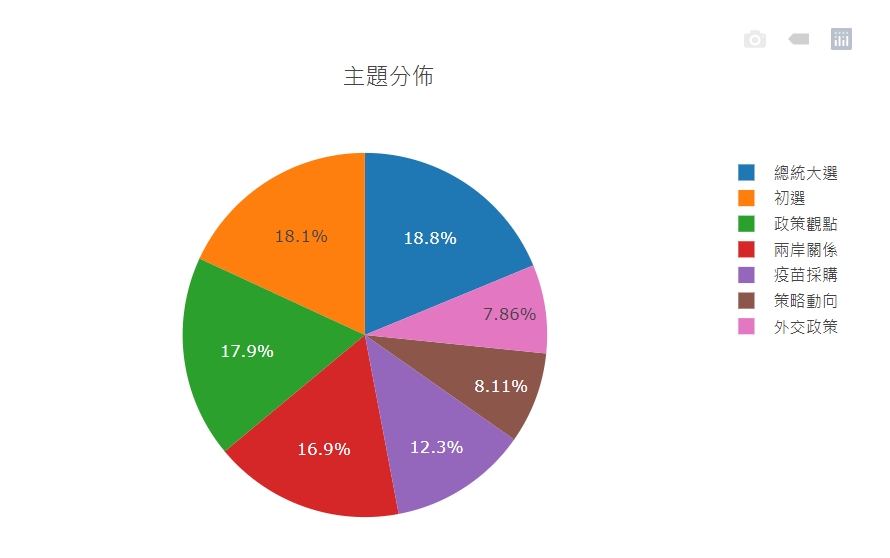
## **第1節 TARFlow流程**

TARFlow工作流程檔名：GoodTiming\_Part2\_topic、goodtiming\_article\_topic\_noun

* 爬蟲
* 斷詞
* 文章情緒分析
* LDA主題分類
* 合併主題與情緒分數
* 製作社會網路圖

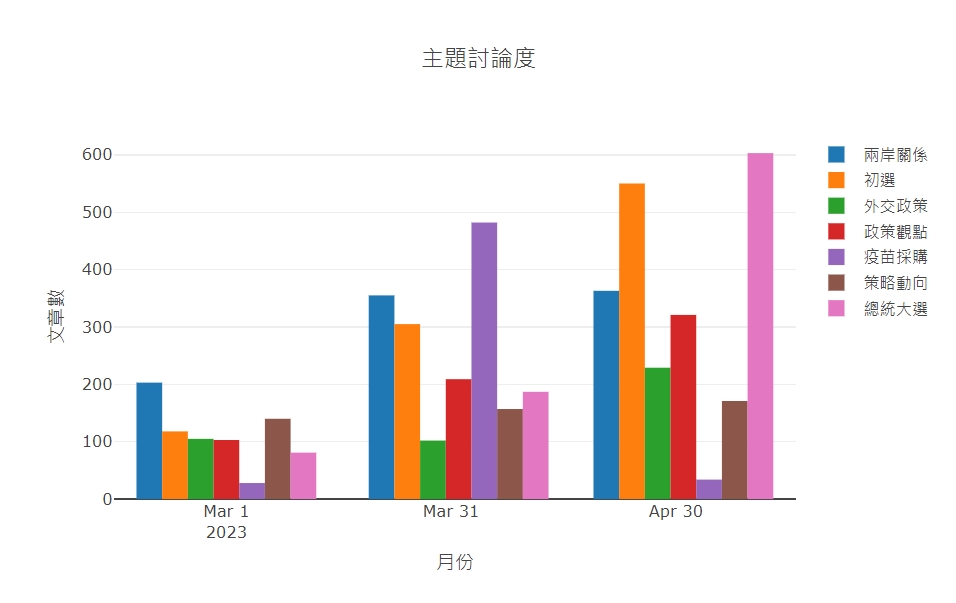
## **第2節 視覺化圖表分析**

### LDA主題分佈

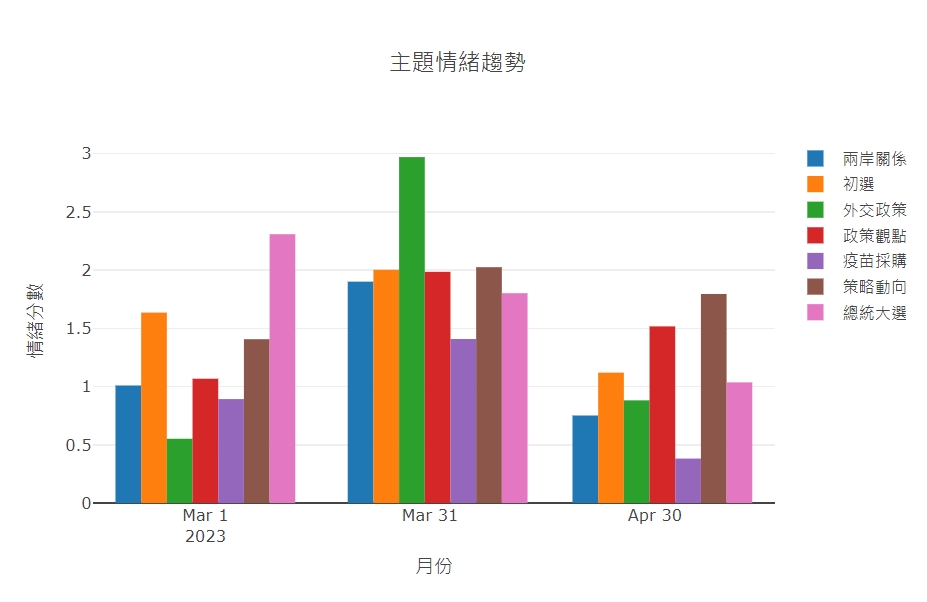
根據主題分布提供的數據，我們可以觀察到以下七個主題文章數高到低的趨勢，這比例僅能提供一個文章數量分布的趨勢，但實際的趨勢和重要性可能需要更全面的分析和評估，由主題的分佈比例可看出聲量最高的前三名分別是總統大選、初選、政策觀點，可見文章熱度還是圍繞著選舉氛圍前進，反而是跟國家外交政策以及策略動向相關等比較生硬的議題乏人問津。

### 第2項 討論度

而關於主題討論度的部分，可以觀察出隨著時間的推移，越接近選舉初選以及總統大選的文章數日漸增加，**疫苗採購**的部分則是隨著媒體的討論度只有在四月時達到高峰，**兩岸關係議**題根據推估也會在越接近選舉時越趨向熱門話題。



### 第3項 情緒趨勢

根據下圖可以明顯看出**外交政策**的情緒分數在3月底時達到了最高峰，這跟前兩組圖片給人所呈現的趨勢有所落差，可解釋為影響文章情緒分數的重點不在熱度以及文章數量多寡，政策本身的議題正當性可能是造結果的主因，而疫苗採購的部分則是已經過了該議題的高原期，相信未來**「疫苗」**這個議題在影響整體的選舉的主流趨勢力量上將變得有限。

### **第3節 LDA主題模型 -**因七項分類混淆度最低，且各項無重疊，故本組認為7項為最佳分類。

### 第1項 初選

### 一張含有 文字, 螢幕擷取畫面, 圖表, 圓形 的圖片 自動產生的描述

### 第2項 總統大選

**一張含有 文字, 螢幕擷取畫面, 圖表, 圓形 的圖片

自動產生的描述**

### 第3項 策略動向

一張含有 文字, 螢幕擷取畫面, 圖表, 圓形 的圖片

自動產生的描述

### 第4項 外交政策

一張含有 文字, 螢幕擷取畫面, 圓形, 圖表 的圖片

自動產生的描述

### 第5項 兩岸關係

一張含有 文字, 螢幕擷取畫面, 圖表, 圓形 的圖片

自動產生的描述

### 第6項 疫苗採購

一張含有 文字, 螢幕擷取畫面, 圓形, 圖表 的圖片

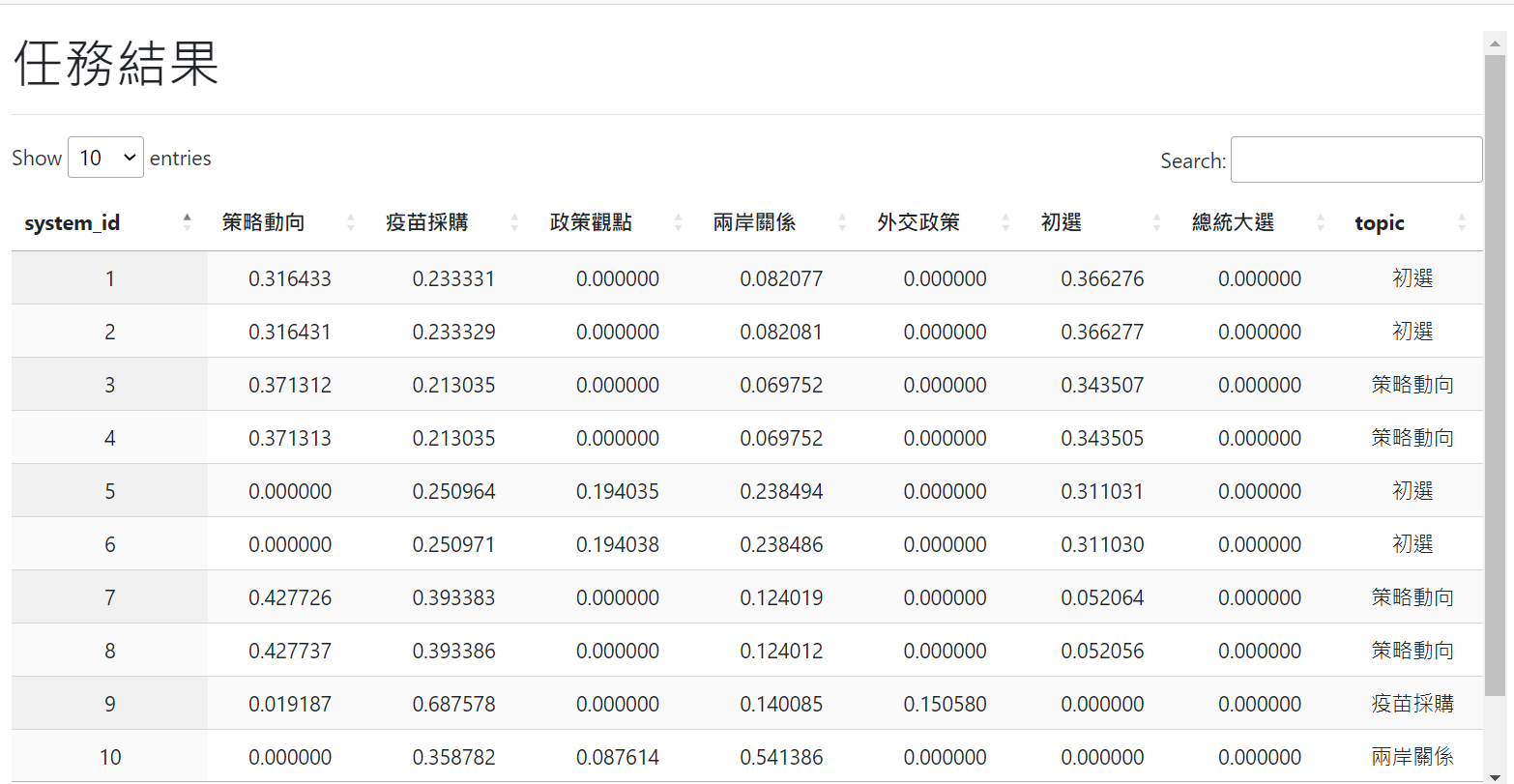
自動產生的描述

### 第7項 政策觀點

一張含有 文字, 螢幕擷取畫面, 軟體, 圖表 的圖片

自動產生的描述

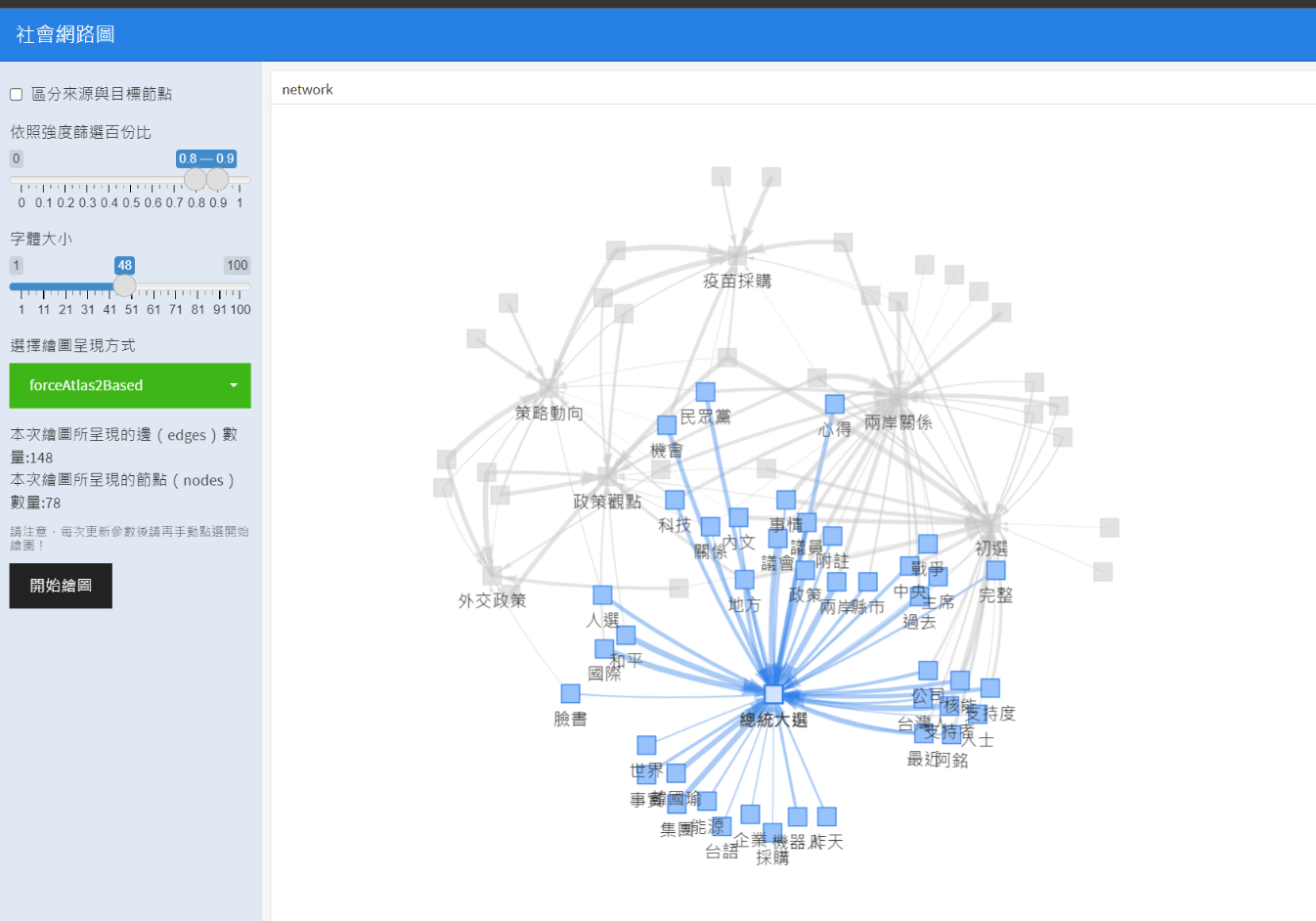
### 第8項 七大分類前10關鍵字與分類任務結果



## **第4節 社會網路關係圖**

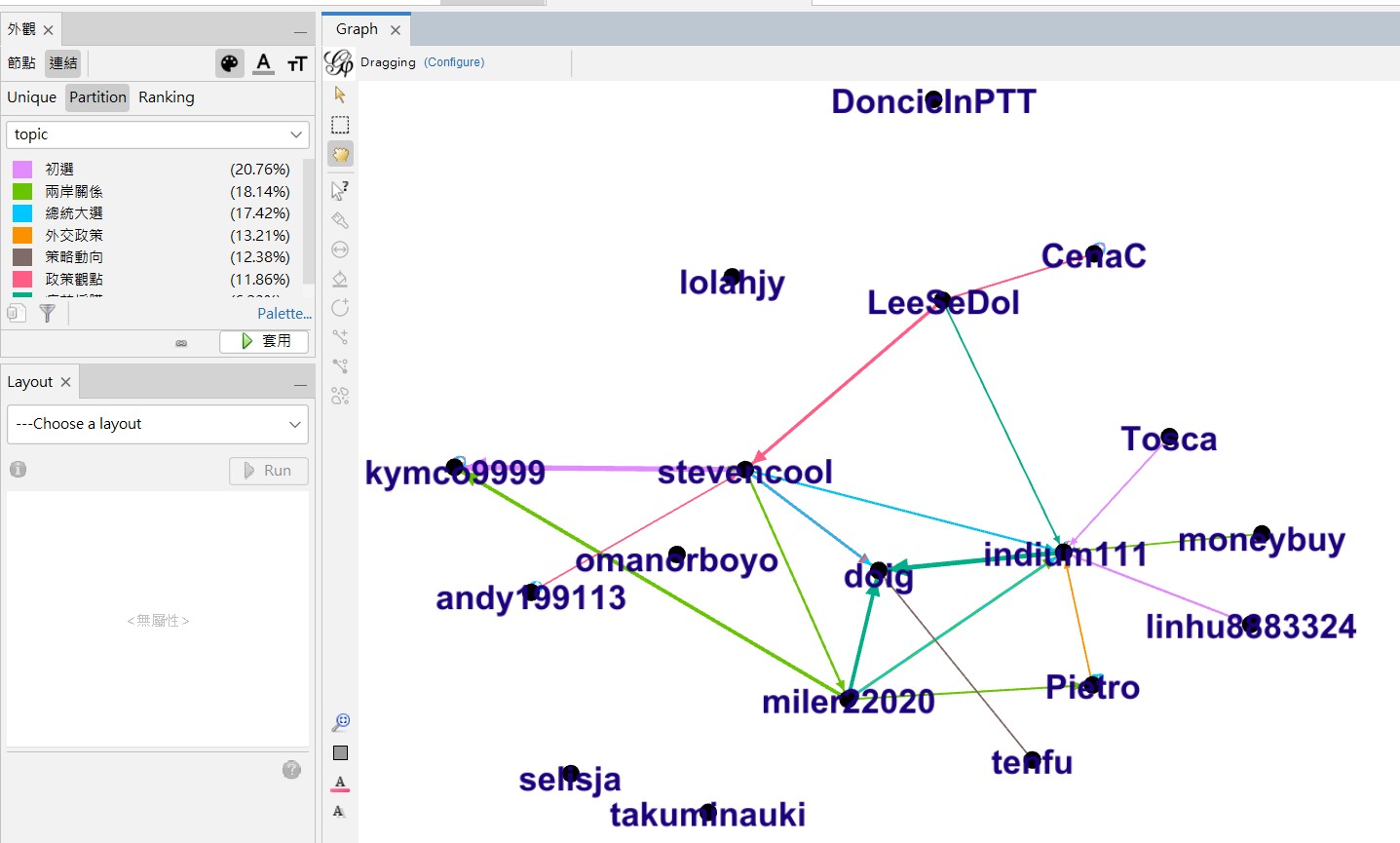
### 第1項 發文與回文者之社會網路關係圖

### LDA分類之社會網路關係圖

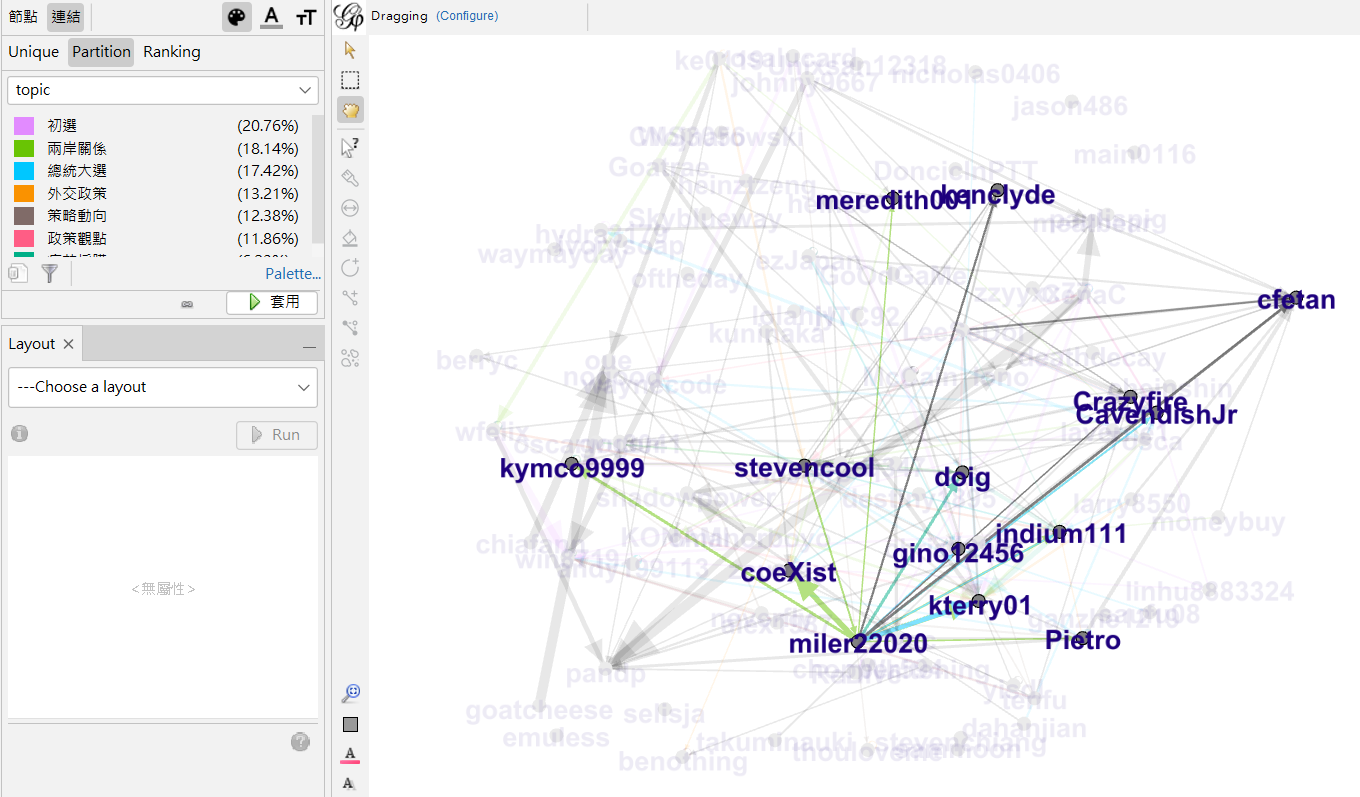


### 第3項 Gephi社會網路關係圖by帳號

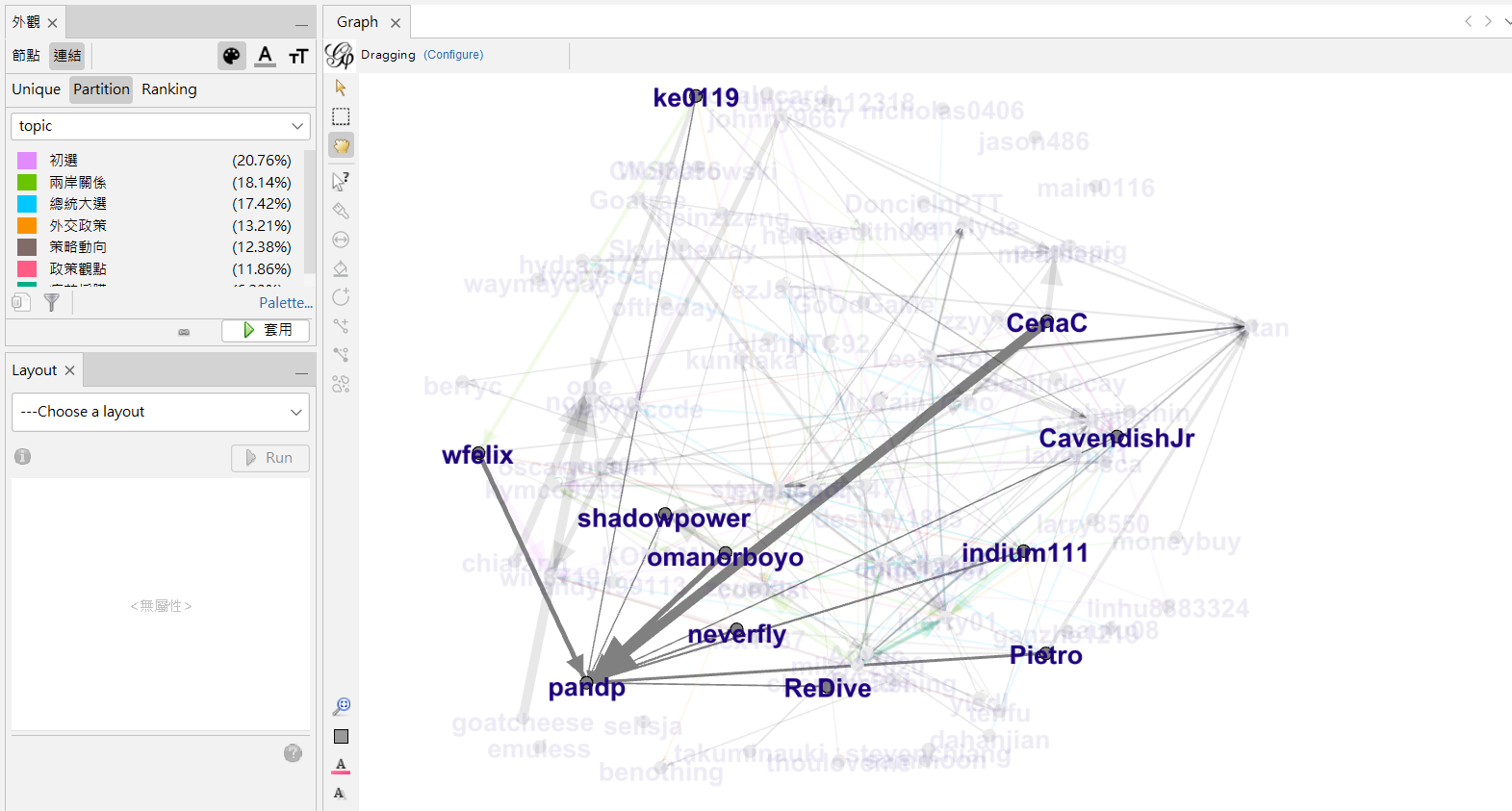
圖一：Inditon111帳號共有7條線(文章類別)指向，發文討論面向較廣。連結不同群體，且是一個重要的節點，具有BetweennessCentrality特質。



圖二：miler22020熱愛討論兩岸關係，且主要對coeXist文章有較多的回覆。



圖三：pandp熱愛發策略動向文章，CenaC較常回復pandp文章。



# 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設定七個主題，透過LDA快速識別並掌握包括策略動向、疫苗採購、政策觀點、兩岸關係、外交政策、初選、總統大選等七大主題分支，分類七群為混淆度最低，且各群無重疊，因此本組認為分七群為最佳數字。

本次讀書會另一個重點為社會網路關係圖，本組嘗試將帳號與LDA分類放入社會網路關係圖，並嘗試使用Gephi進行視覺化分析。透過強度篩選百分比調整至最佳化，即能發現部分帳號具有Betweenness Centrality特質，這些帳號在社群網路裡具有連結不同主題特質，且扮演重要節點角色；另發現多數帳號多屬對單一主題有興趣。 透過分析和研究社會網路關係圖，我們可以揭示人與人之間的互動模式、社交群體的形成、資訊流動的路徑等，從而瞭解社交系統的運作和影響。