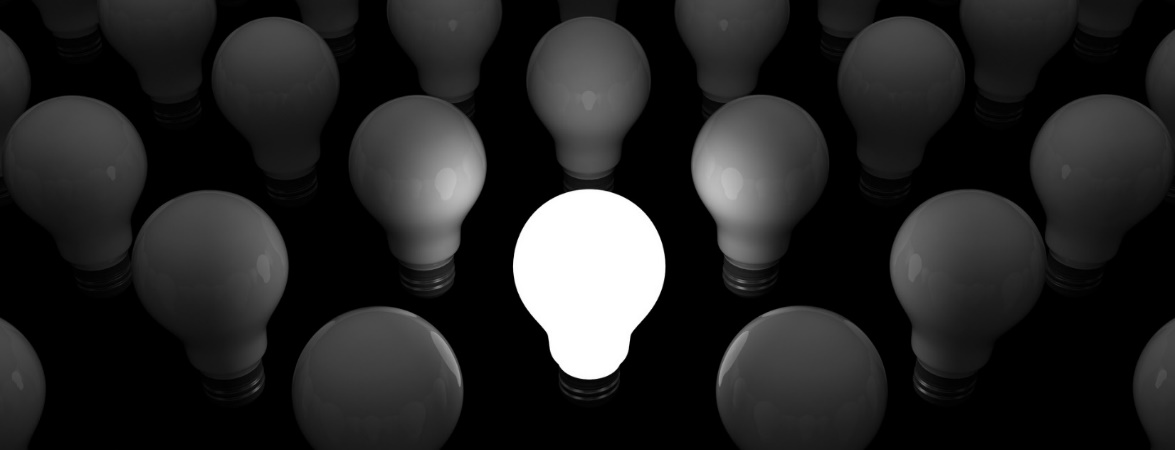
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **台灣推動電力市場改革- Taiwan Business TOPICS 工商雜誌** | | | | |
|  |  |  | |  |
| 社群媒體分析  **期末專案報告- PTT八卦版缺電議題分析** | | |
| **第12組**  **組員 林志勳 N094220014**  **洪嘉駿 N094220020**  **李明章 N094220023**  **陳雅詩 N094220026** | |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
|  |  |  | |  | |
|  | 動機和分析目的  **2021年夏天，台灣發生了大規模的停電現象，而且在同一週內還不只一次的大停電。甚至政府發布的官方消息還告訴民眾，未來缺電可能是一種常態，全國分成很多輪流停電組別，如果遇有緊急供電狀況時，將依該照組別進行輪流限電措施，帶給民眾生活上的諸多不便，也造成很多企業及工廠巨大的損失。但是年底舉行公投案，提到有關能源政策時，目前的執政黨政府卻又告訴民眾核四不用建、台灣不缺電，在野黨則是利用此一前後矛盾狀況對政府發起猛攻，執政黨也立刻反擊，提出各種替代能源政策等辯駁。因此，我們想了解在批踢踢八卦板，對於2021年有關「缺電」的主題，批踢踢鄉民們對於該議題的相關討論重點是甚麼，或是表達出來的關鍵字、文章情緒是呈現甚麼樣的狀況。** | |  | |

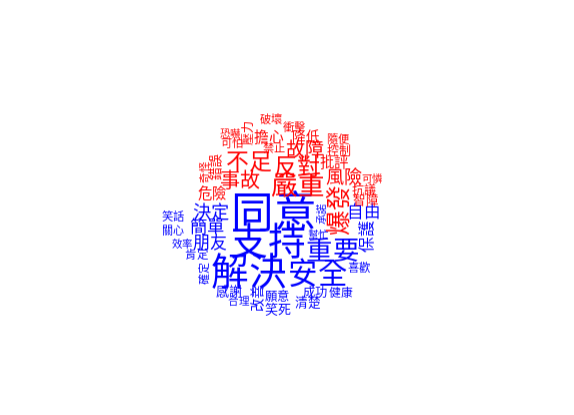
|  |  |
| --- | --- |
|  |  |
| 資料集描述資料來源：PTT八卦版 **因批踢踢的社群討論版裡面，沒有特別針對本議題的專屬討論版，因此我們選擇了長期都處於聲量與討論文章數最高的前幾名熱門版別PTT八卦版，作為我們主要的資料集分析文本來源，並使用文字分析平台的PTT爬蟲抓取原始資料集。** 資料時間區間： 2021/03/31 - 2021/12/22 **為觀察本議題在社群討論內的長期變化，我們選擇停電時期的前後幾個月以及在核電公投案成立並開始有社群討論熱度的時期，作為主要觀察及分析文本的時間區間。選擇爬取資料集的日期區間為2021/03/31 - 2021/12/22。** | |
|  | |  |
| 關鍵字：缺電資料筆數：3942筆 2021年在5月份的時候同一週內就發生了兩次大停電，本來我們猜想台灣缺電的議題，關注的人並不是很多，經過大停電事件之後，觀察這段時間的社群討論版上的文章數量，一下子從一、兩百篇的文章爆增為兩千多篇，因此我們覺得這個題目應該具有相當的可研究性。 | |  |

|  |  |
| --- | --- |
|  |  |
| 使用工作流程平台分析資料 | |

****

|  |  |
| --- | --- |
|  |  |
| 資料前處理替換字串 **去除換行符號、網址、html標籤、底線、多餘的空格** 中文斷詞 進行每次斷詞任務之後反覆觀察任務結果，並將部分沒有斷開的詞加入自定義詞彙及權重(例如：天然,氣>天然氣、備轉,容量>備轉容量、用愛,發電>用愛發電)，另外，有一些可能影響資料分析結果的斷詞，也依序加入權重自定義詞彙，有些重要字詞的前面有「不」的必須先進行斷詞後，才進行沒有「不」的斷詞，才能有比較準確的結果(例如：不,缺電>不缺電、不,合理>不合理、不,合法>不合法、不,嚴重>不嚴重)；或是專有名詞、人名的斷詞錯誤修改(例如：韓國,瑜>韓國瑜、小,粉紅>小粉紅、供,五,停,二>供五停二)   清除停用詞 進行每次清除停用詞任務之後重複觀察每次任務結果，並將可能干擾分析結果的停用詞加入自定義停止詞 | |

|  |  |
| --- | --- |
|  |  |
| 資料的分析過程Lexicon Based 情緒分析 **將無法明確確定意義的負面字詞例如「八卦」、「看看」、「垃圾」、「反正」等字先做停用字去除後做一次Lexicon Based情緒分析，得到的：**  **負面字詞有嚴重、爆發、反對、不足、事故、故障…等**  **正面字詞有同意、支持、解決、安全、重要、決定…等** | |

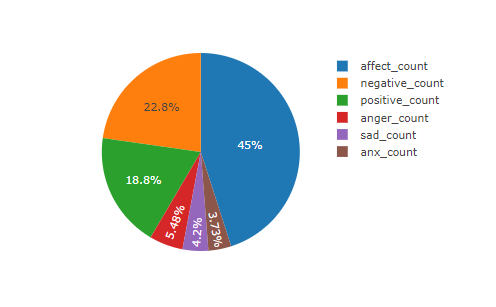


## LIWC 情緒分析

**觀察實際的正負面文章數量，負面文章數量略多於正面文章數量**

**但是並沒有太過懸殊的明顯差距**

**推測本分析主題在這段分析期間的正、負面聲量(情緒)沒有太顯著的差異**



|  |  |
| --- | --- |
|  |  |
| 資料的分析過程詞頻計算及文字雲 **經過詞頻計算及所產生的文字雲，以「缺水」、「缺電」兩個詞最多** | |

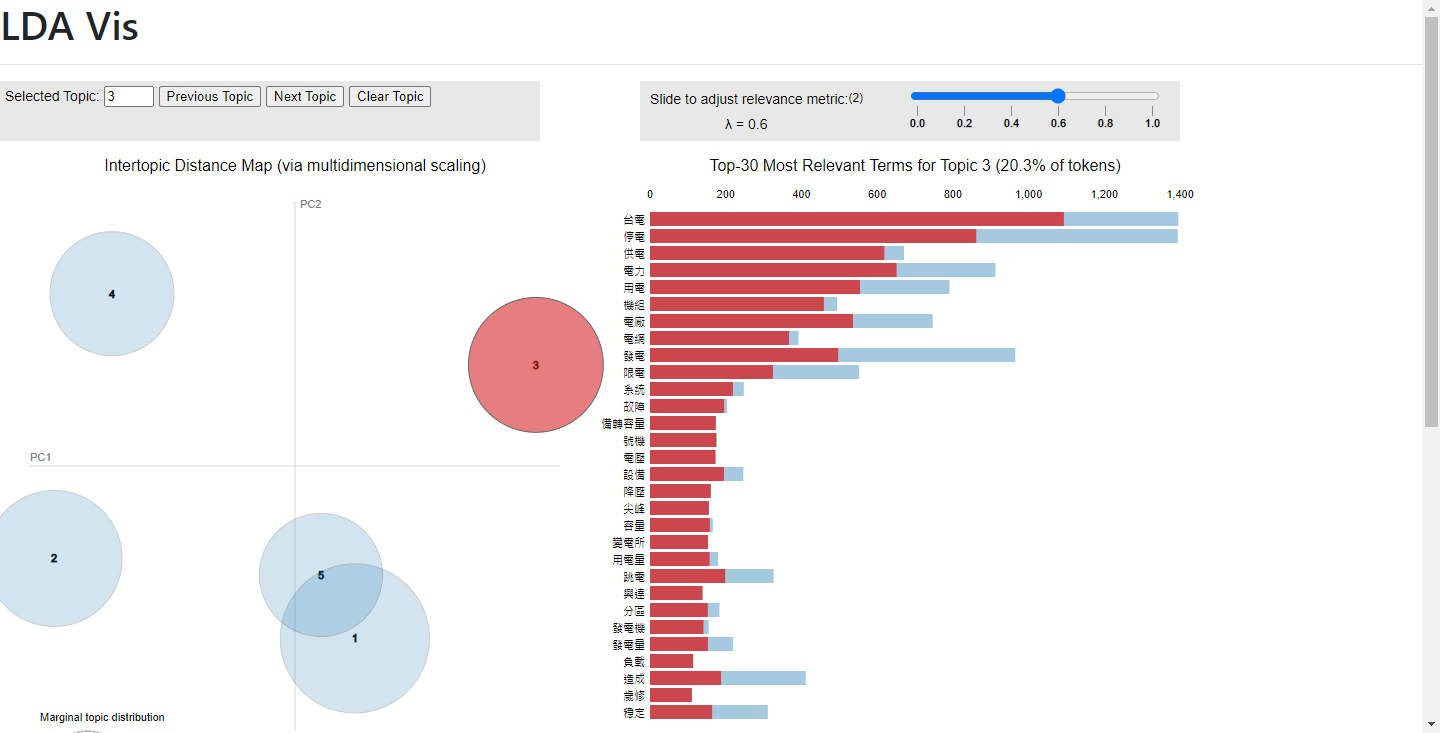


|  |  |
| --- | --- |
|  |  |
| 視覺化的分析結果與解釋LDA主題分析 **透過資料前處理後，再進行LDA主題模型分析出「缺電」關鍵字相關聯的主題，由於資料量的關係，我們選擇設定5個主題數來觀察其中的關聯性。** | |

|  |  |
| --- | --- |
| **結果觀察** | **主題1**的關鍵字在各文章出現最多的是**「核四」**，其次為公投、能源政策與政黨政治等相關議題，如天然氣、三接、藻礁、核電、反核等。  **主題5**的關鍵字在各文章出現最多的是**「中國」**，其次為經濟投資方面的相關詞彙。 |
| **主題分析** | **主題1**：分析出的字詞反映出文章議題呈現較為偏向國家政策與即將舉辦公投等相關議題有關，尤其民眾對於核四議題討論最為激烈，顯示與能源或電力有關的新聞事件發生時，媒體或網路輿論針對核四議題討論度總是最高。  **主題5**：反映出文章的議題為國際能源方面的探討，以及近期討論度較高的台積電在高雄設廠等議題。且經由缺電事件討論，延伸的議題不僅是攸關重要產業甚至國際企業的投資，更影響著國內未來的經濟發展，以及國際上的競爭優勢。  LDA圖中可看出主題一與主題五的討論相似度較高，皆較偏向為政治方面的議題。 |

|  |
| --- |
|  |
|  |
| 視覺化的分析結果與解釋 **LDA主題分析** | |

|  |  |
| --- | --- |
| **結果觀察** | **主題2**的關鍵字在各文章出現最多的是「**疫苗**」，其次與政府或政策較有關的字詞，並連帶提出其他政策缺失的詞彙，如缺水、缺疫苗。 |
| **主題分析** | **主題2**：分析出的字詞，反映出文章的議題為民眾較為負面對於政策不滿的討論，以及民眾對於目前執政者所執行的國家政策，提出相關的問題討論。 |



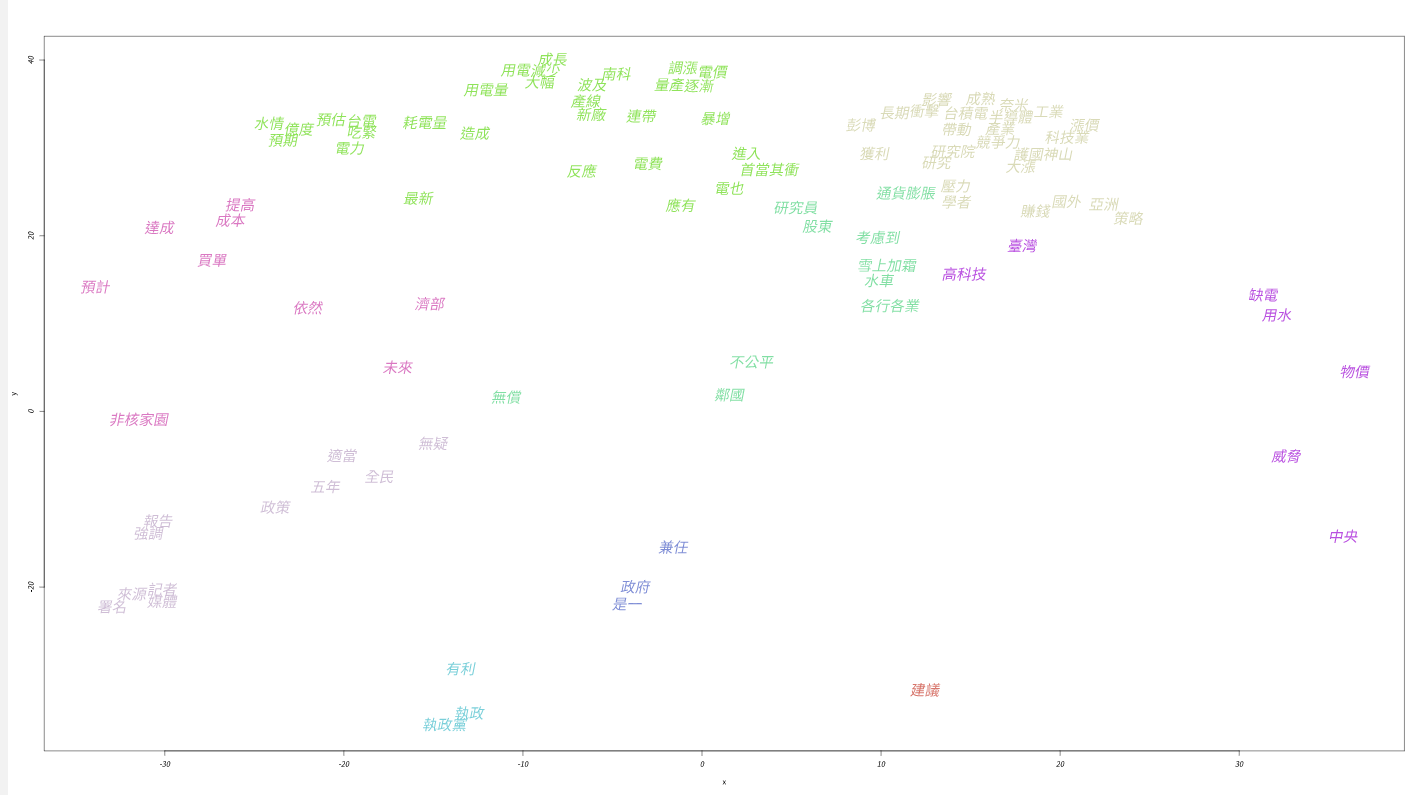
|  |  |
| --- | --- |
| **結果觀察** | **主題3**的關鍵字在各文章出現最多的是「**台電**」，其次為與電力或供電方面較為直接的詞彙，如發電、限電、故障、備轉容量等。 |
| **主題分析** | **主題3**：分析出的字詞，反映出文章的議題為台電公司相關**供電**方面的探討，以及用電負載導致停電與如何限電相關的問題討論。 |

|  |
| --- |
|  |
| 視覺化的分析結果與解釋 **LDA主題分析** | |

|  |  |
| --- | --- |
| **結果觀察** | **主題4**的關鍵字在各文章出現最多的是「政府」，其次為民眾探討其他議題所關聯到的詞彙，與缺電本身的關聯性不是很顯著。 |
| **主題分析** | **主題4**：分析出的字詞，反映出文章的議題經由缺電的關鍵字，關連到其他網路關注的話題，或是負面的言論以及謠言的討論，與缺電關鍵字相關的討論關聯性較低。 |

**由主題模型分析，可觀察到在資料集鎖定的這段時間區間內，與本專案議題相關的討論主題有出現與「缺電」相關討論字串(如主題3)，而這些字詞可以再進一步做為接下來要進行字詞網絡關係圖或是建立字典以匯入單中心網路圖之用。**

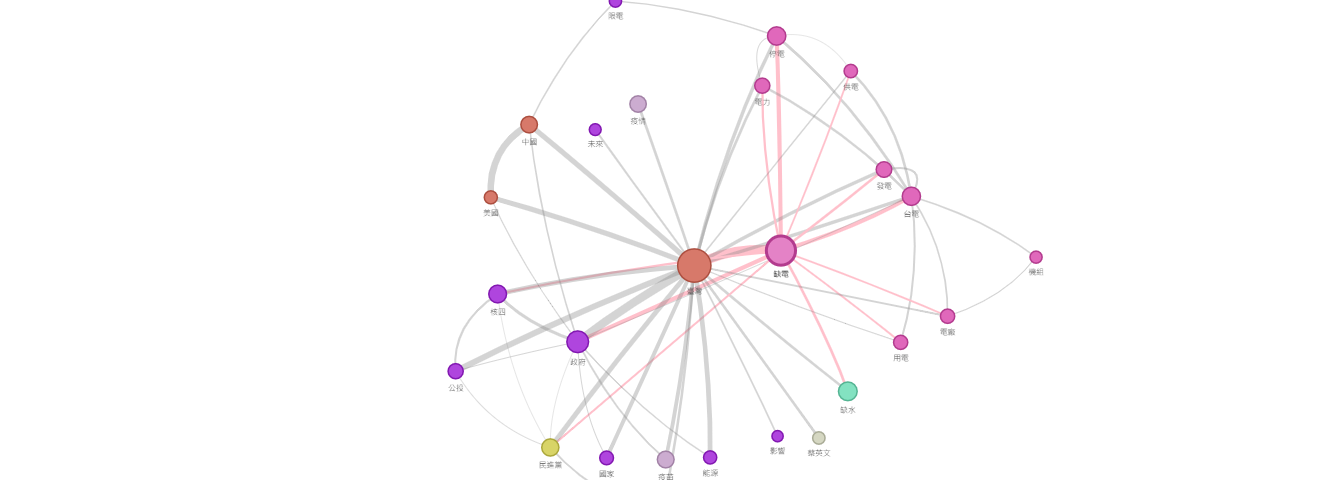
|  |
| --- |
|  |
| 視覺化的分析結果與解釋 **使用Word2Vec產生關聯式文字雲**  **使用Word2Vec模組搭配關聯式文字雲視覺化具產生文字雲圖，觀察各分群主題關鍵字的相關性程度。** | |



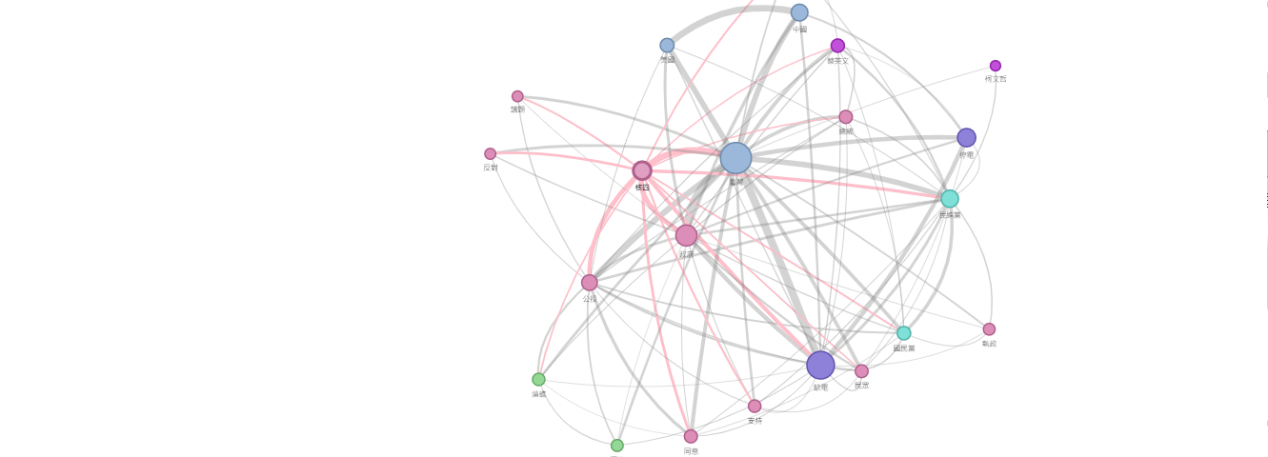
|  |
| --- |
|  |
| 視覺化的分析結果與解釋 **使用Word2Vec產生關聯式文字雲**  **使用Word2Vec模組搭配關聯式文字雲視覺化具產生文字雲圖，觀察各分群主題關鍵字的相關性程度。** | |
|  | |

|  |
| --- |
|  |
| 視覺化的分析結果與解釋 **使用關聯式文字雲將分群文字自定義為分類字典**  **觀察前面所產生的關聯式文字雲圖，各分群主題關鍵字分別有哪些，轉為DTM後將其共現矩陣csv檔案下載，將其分為幾個類別後，依據字詞自定義類型再重新存成csv-UTF8的字典檔，並匯入工作流程系統的單中心網路圖參數中。**  **建立自定義字典e-dictionary.csv** | |

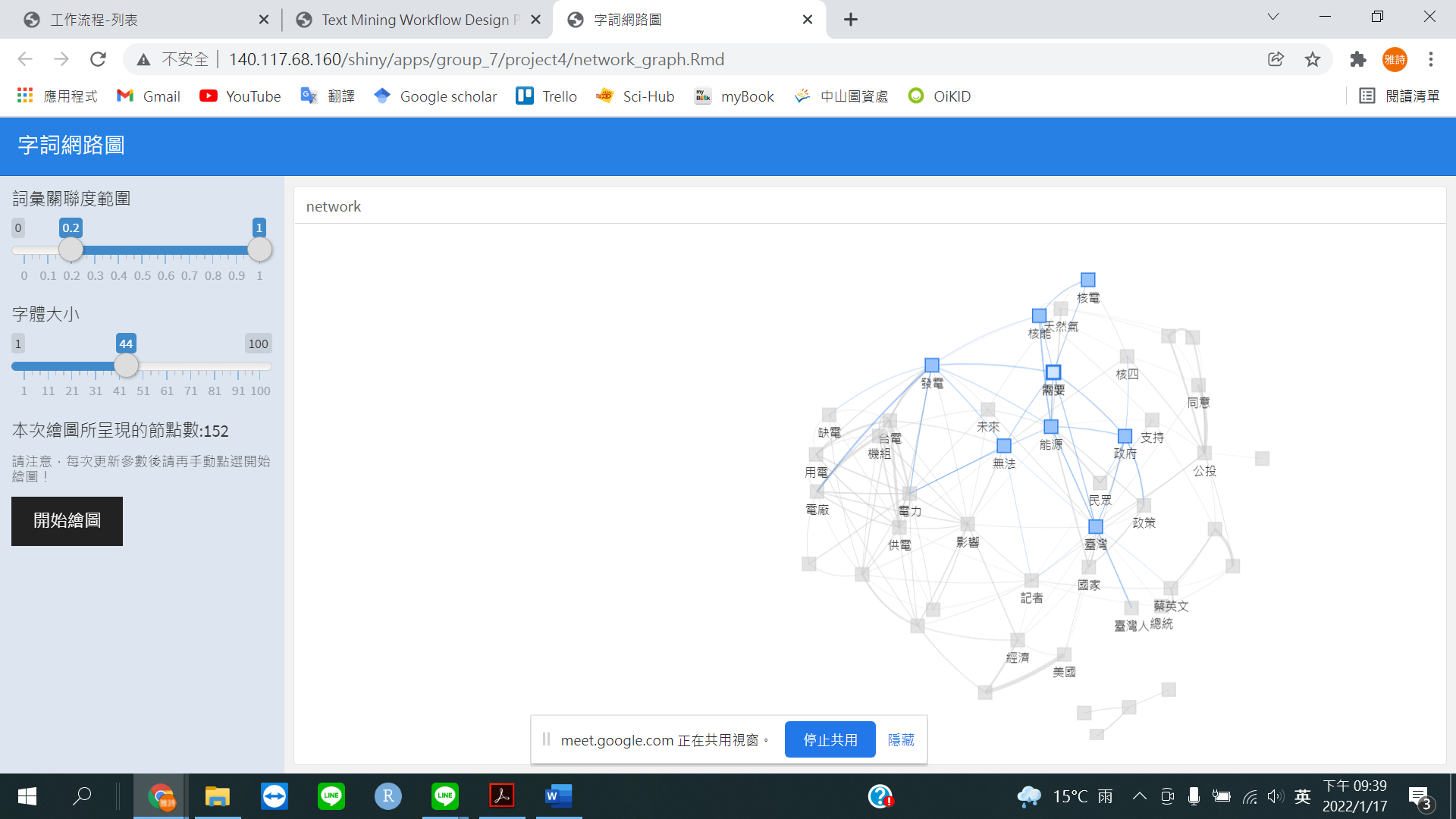
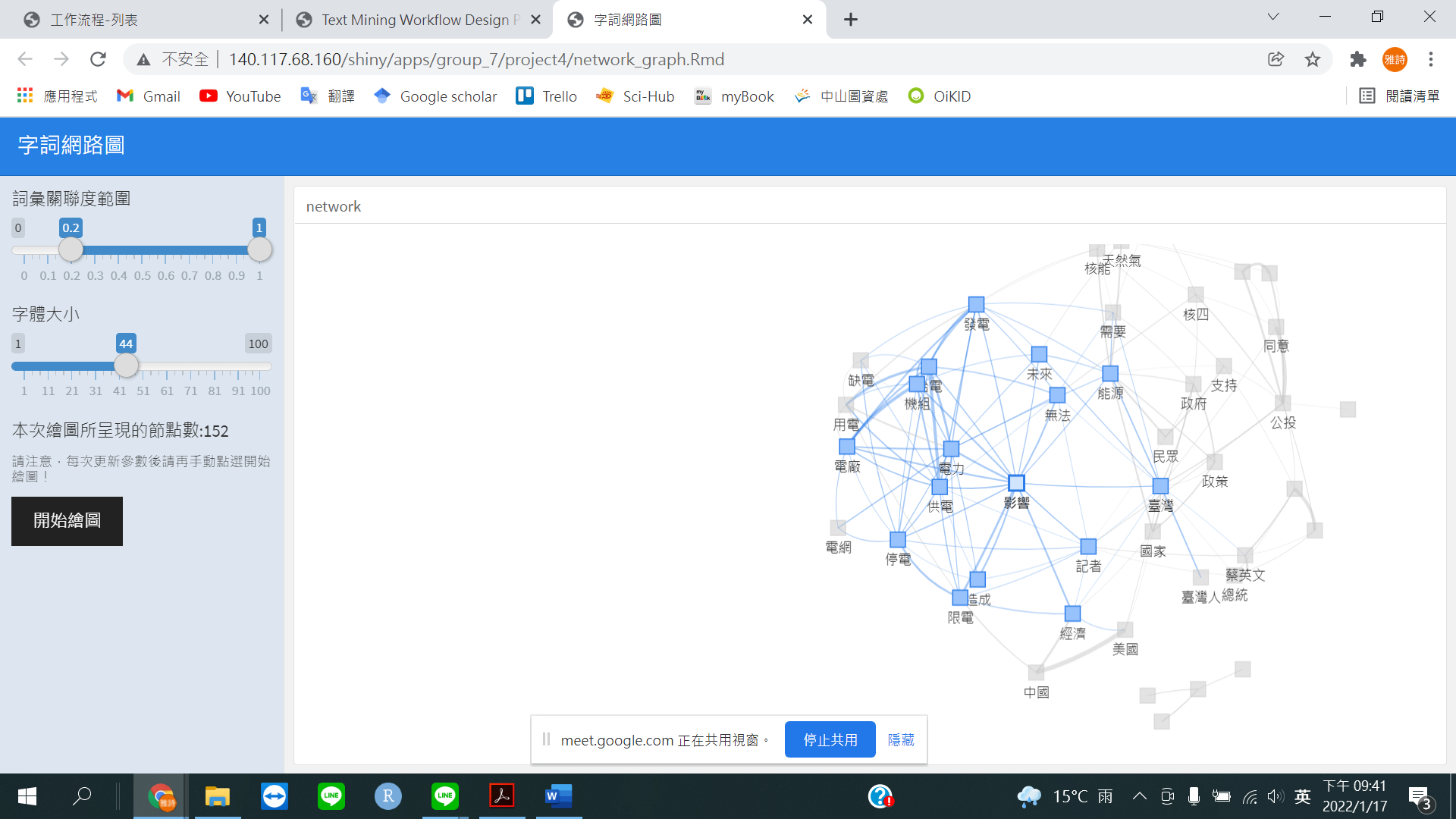
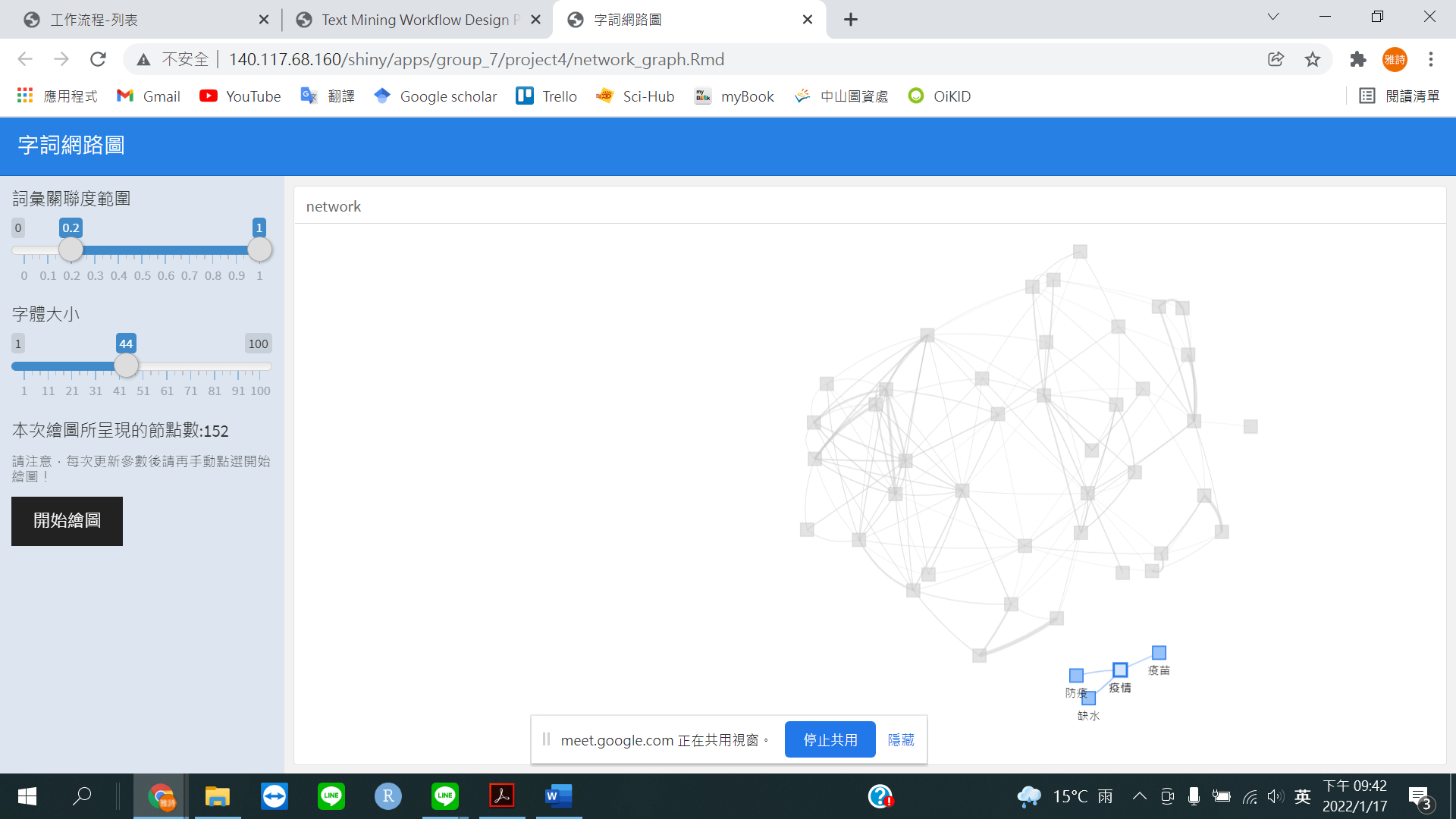
|  |
| --- |
|  |
| 視覺化的分析結果與解釋 **單中心網路圖**  **與「缺電」字詞相關的詞，關係強度最強的是「臺灣」、「台電」、「政府」、「停電」等** | |



|  |
| --- |
|  |
| 視覺化的分析結果與解釋 **單中心網路圖**  **與「核四」字詞相關的詞，關係強度最強的是「臺灣」、「公投」、「政府」等** | |



|  |
| --- |
|  |
| 視覺化的分析結果與解釋 **字詞網路關係圖** | |



另外從網路圖也可發現有一群字與主要的核心網路斷開，且位置距離較遠，是相較與本議題字詞關聯性較低的字詞

以「影響」這個詞為中心，發散出相關聯的字詞有電力、供電、發電、用電、機組等

以「需要」這個詞為中心，發散出相關聯的字詞有能源、核能、發電、核電等

|  |
| --- |
|  |
| 結論 **這次專案報告選擇的資料集來源是PTT八卦版，八卦版雖然在PTT的各討論版裡面一直都是位居前幾名討論熱度的熱門板別，但也因為八卦版沒有鎖定特定主題或對象，因此在蒐集討論本專案議題的文本資料時，在資料集時間區間的選擇，還有資料前處理的過程部分最耗時，也最費工，必須重複多次的觀察資料變化與輸出結果，才能完整的呈現出想要看的主要內容。LDA主題分析以及詞頻計算的工具幫助我們有效聚焦在重要的關鍵字詞上，也能藉此訂出比較相關的群集資料。**  **在缺電議題的社群討論串中，有許多是關於政府政策評論的相關討論內容，可以發現除了供需問題的面向外，社群使用者對於政府制定政策是否有科學客觀的依據，也是其關注的重點之一。** | |