**社群媒體分析第九組第三次讀書會作業**

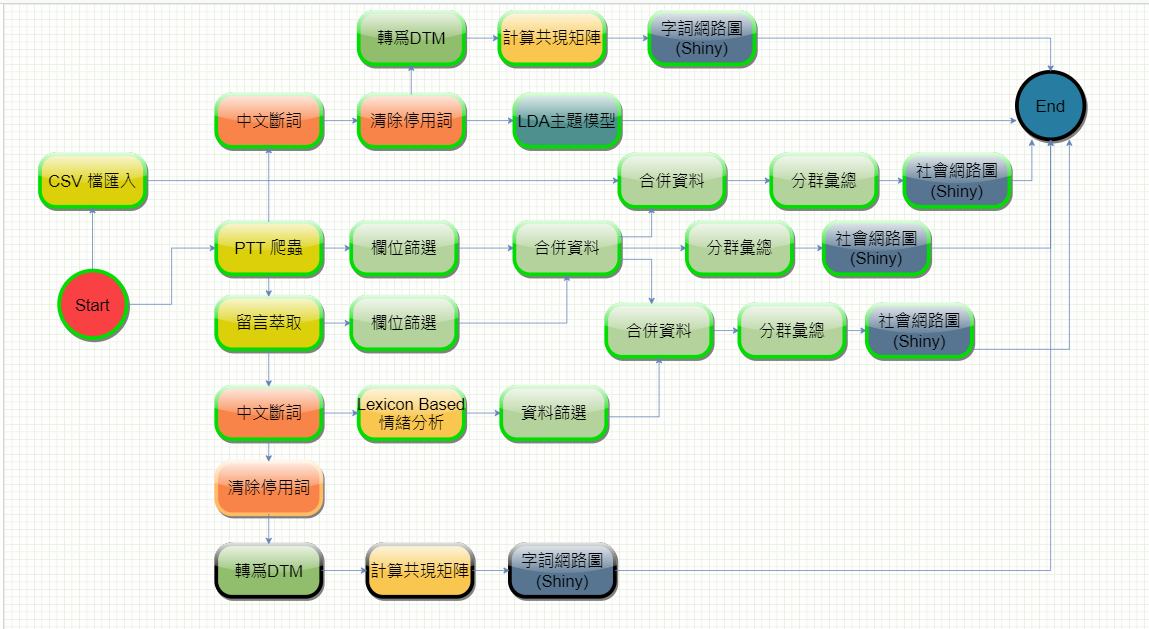
1. 分析熱門主題: 從宏蕾事件觀察社會輿論趨勢
2. 使用平台:中山大學工作流程平台
3. 所使用的工作流程名稱:Nice\_try
4. 建立時間: 2021/12/23 09:56:17

一張含有 桌 的圖片

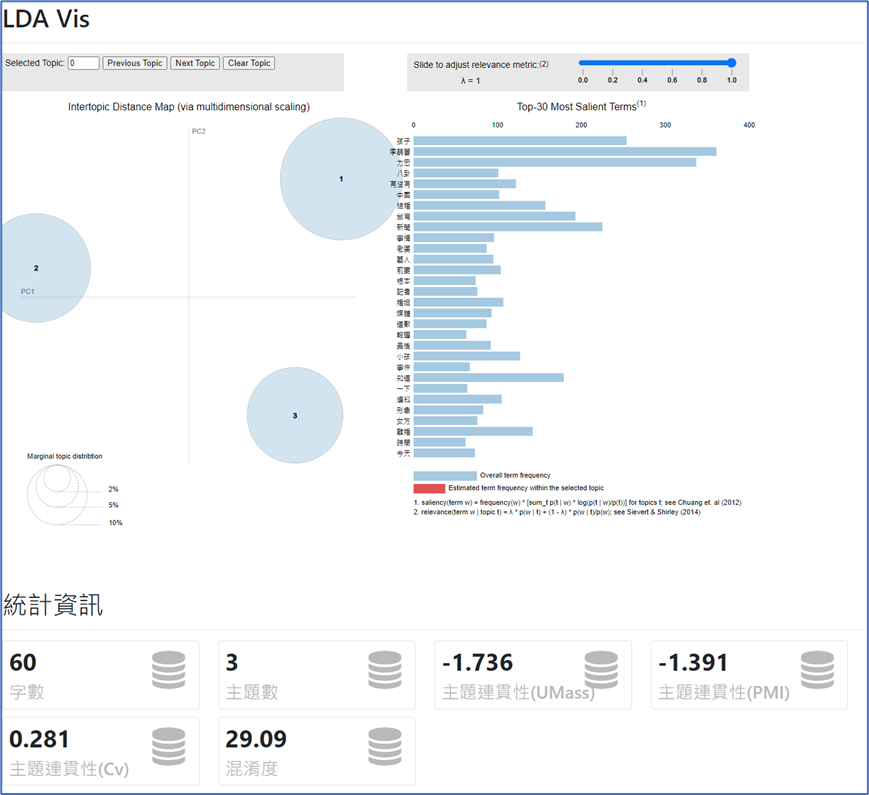
自動產生的描述

使用此工作流程

1. 工作流程圖:
2. 使用工作流程平台PTT元件爬取2021/12/19~2021/12/20、八卦版、關鍵字為: 王力宏、李靚蕾、蕾神之槌、宏荒之力等字詞，共計686篇文章，並使用LDA主題模型元件產生3個主題匯出csv檔，與686篇文章進行合併然後進行分群彙總產製社會網路圖。

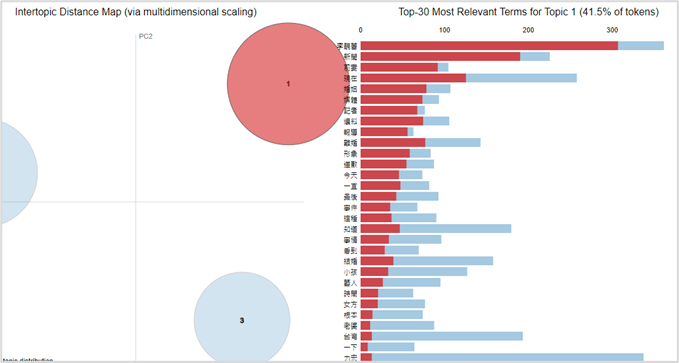


1. 主題模型的命名:

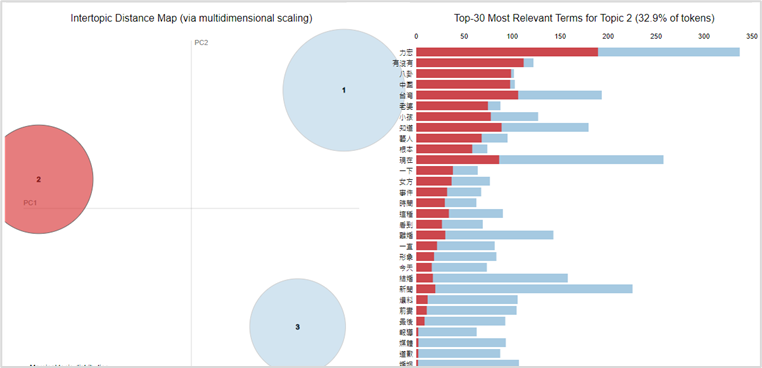


「家庭」通常是由男女雙方以及小孩所構成，在近期演藝圈所爆發的王力宏形象敗壞事件中，社會大眾亦是圍繞著這三者來做為茶餘飯後的討論內容，本小組依據LDA主題模型分為3種，分析結果說明如下：

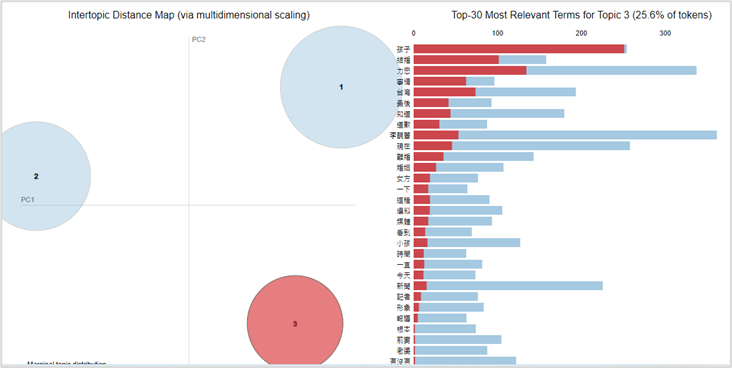
* 主題一：蕾神之槌／最多人討論的主題 (圈圈最大)
* 出現最高的前三名詞頻：李靚蕾、新聞、前妻
* 分析結果：最初王力宏與李靚蕾的離婚，社會輿論歸咎原因為婆媳問題，並直指李靜蕾強勢插手王力宏演藝事業，由於王力宏團隊刻意操作新聞風向，李靚蕾為了捍衛自身權利而選擇說出事實，並以前妻身份進行多項爆料，王力宏所建立的形象與前妻爆料內容明顯不符的情況下，讓新聞媒體爭相報導，大眾輿論亦逐漸轉變為支持女方。



* 主題二：宏荒之力／第二多人討論的主題
* 出現最高的前三名詞頻：力宏、有沒有、八卦
* 分析結果：隨著李靚蕾對外說明的文章內容，讓社會大眾驚訝地發現王力宏的實際作為與塑造的形象有相當大的落差，有愈來愈多的媒體、網友開始翻出王力宏之前的受訪紀錄或歌曲，為的就是從裡面找出王力宏有沒有其他的不良紀錄以及更多的八卦。



* 主題三：雙方的小孩／較少人討論的主題
* 出現最高的前三名詞頻：孩子、結婚、力宏
* 分析結果：由於王力宏澄清文裡面說明因為女方有小孩而被迫進行結婚，與李靚蕾對外說明的文章內容指出王家想要有很多小孩，所以結婚五年陸續生出三個孩子有明顯的落差。這點亦成為PTT八卦版的討論重點。



* LDA數值說明：

主題的連貫性(PMI)： -1.391

主題的連貫性(UMass)： -1.736

主題的連貫性(Cv) ：0.281

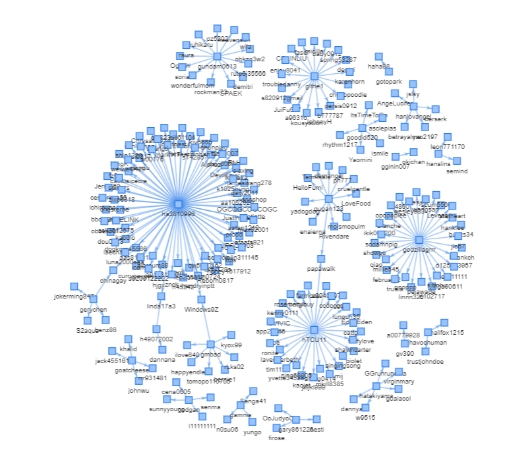
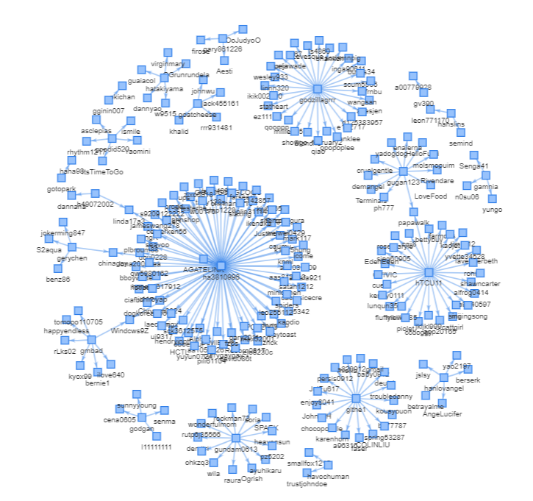
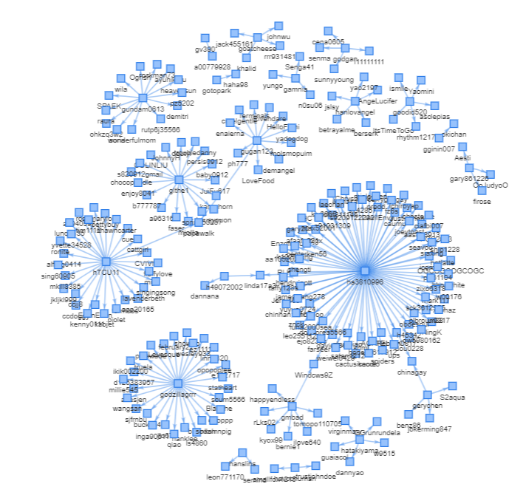
混淆度數值為： 29.09

1.由於主題的連貫性這三項數值偏低，是否代表主題與主題之間彼此相似?

2.混淆度29.09代表此為良好的模型?混淆度數值較低是否亦代表文章比較好理解?

1. 網路節點的定義與邊的權重:

* 網路節點分別是發文者(artPoster)和留言者(cmdPoster)，留言者為來源節點，發文者為目標節點
* 邊的權重則分別是依照三個主題(LDA元件產生)、發文數量及情緒值為正(大於0)的文章數量給定。

ha3810996

ha3810996

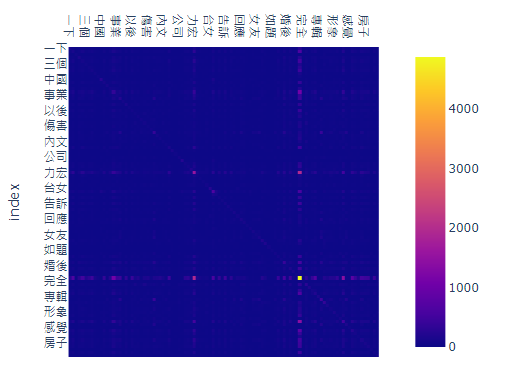
ha3810996

ha3810996的貼文最多人回應，觀其發文內容屬於語不驚人死不休的八卦文，適合終日群居言不及義的網民茶餘飯後閒聊使用

1. 視覺化圖表呈現的多樣性:

(一)字詞相關網路圖執行結果:

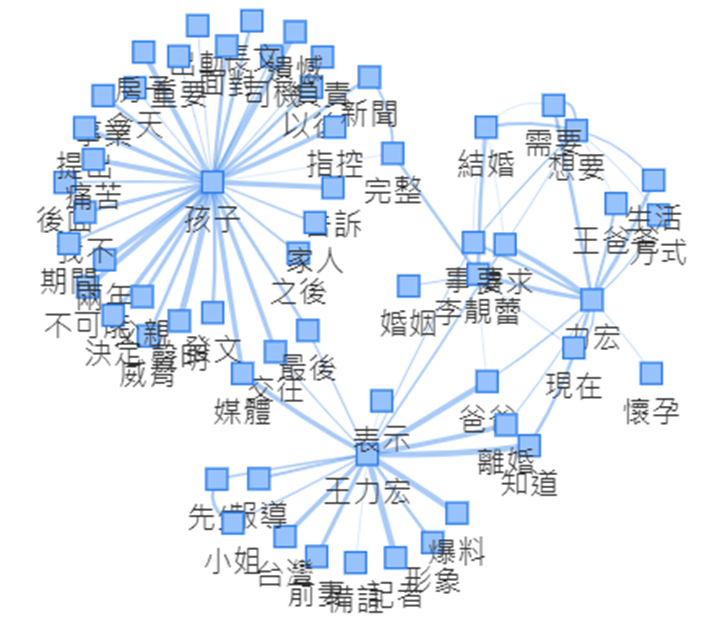
(Heat map )統計字數:200



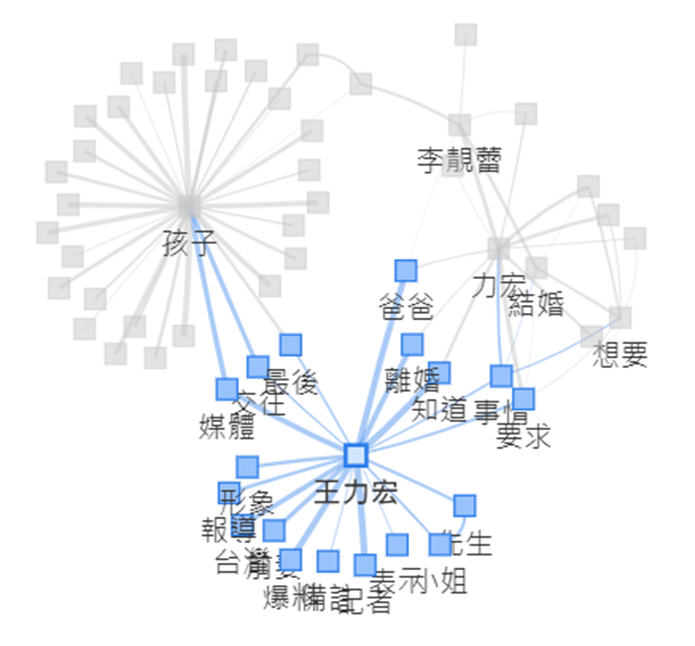
(二)全面性的字詞網路關聯圖

調整詞彙關聯度參數:詞跟詞的相關係數(關聯圖):人、事件、或詞跟詞之間的關聯度(由小而大)

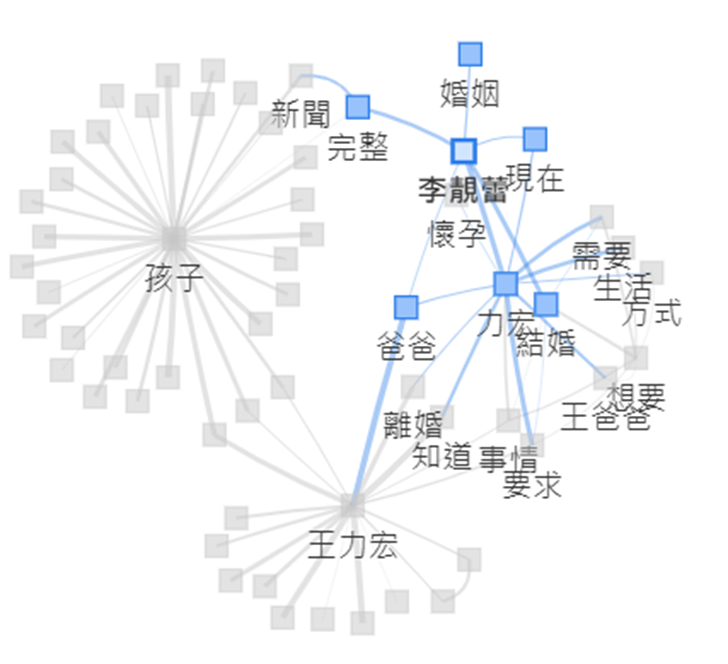
執行結果總覽:可看到三個相關字詞分布(王力宏、李靚蕾、孩子)



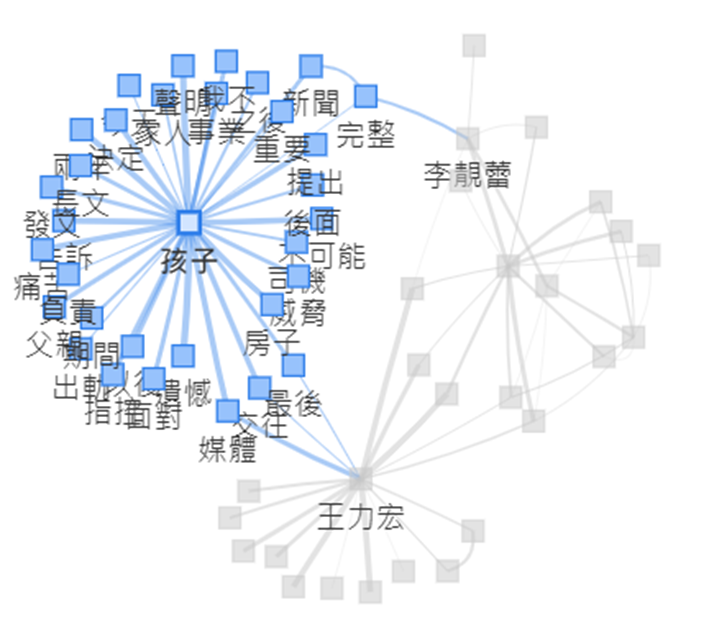
[王力宏]字詞關聯圖:形象、前妻、離婚、…等



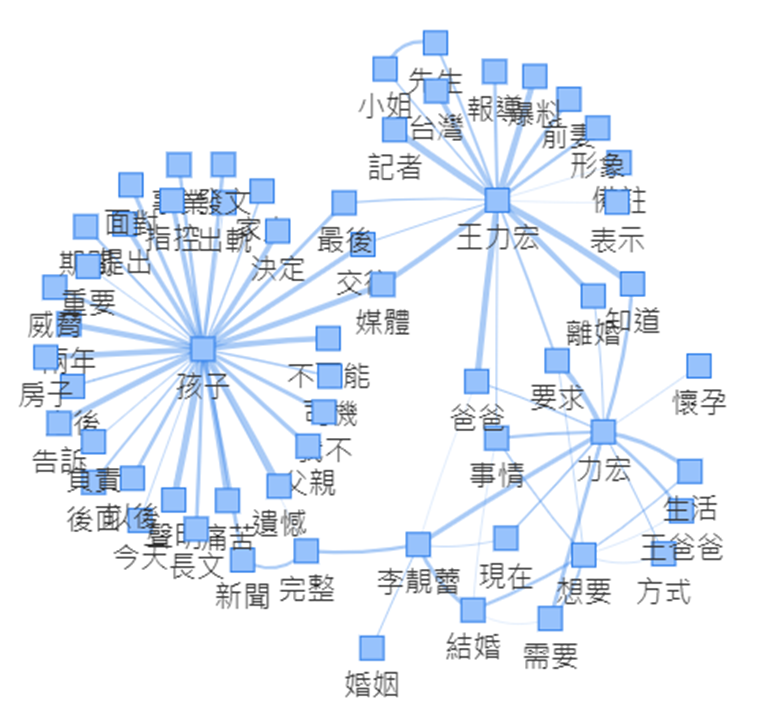
[李靚蕾]字詞關聯圖:婚姻、懷孕、完整、力宏、新聞…等



[孩子]字詞關聯圖:威脅、房子、家人、父親…等



留言萃取後之字詞網路圖:結果與上述接近

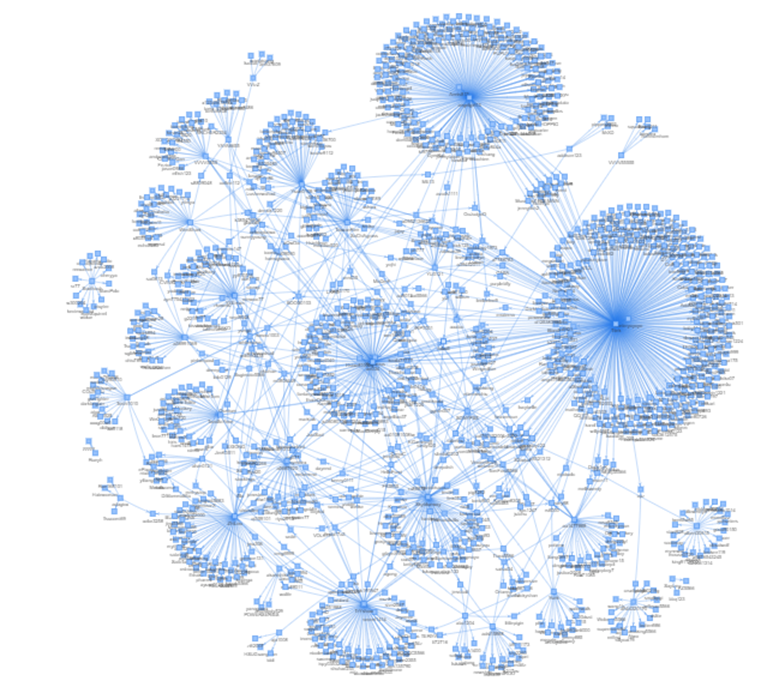
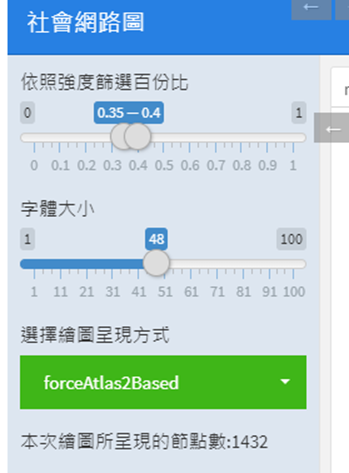


(三)社會網路圖:處理PO文跟回文(留言萃取)和情緒值為正(大於0)的文章數量的連結強度

本Project處理筆數:4502筆



首先進行參數調整

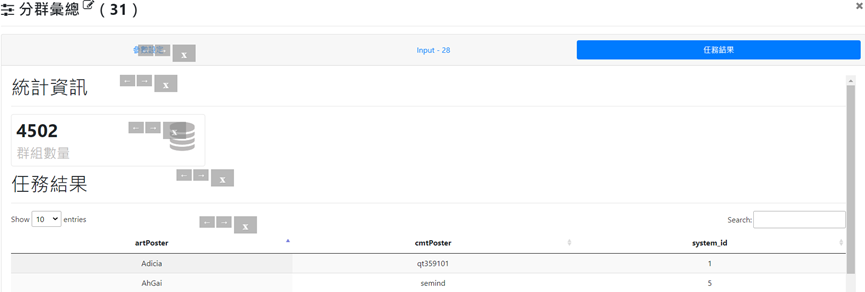


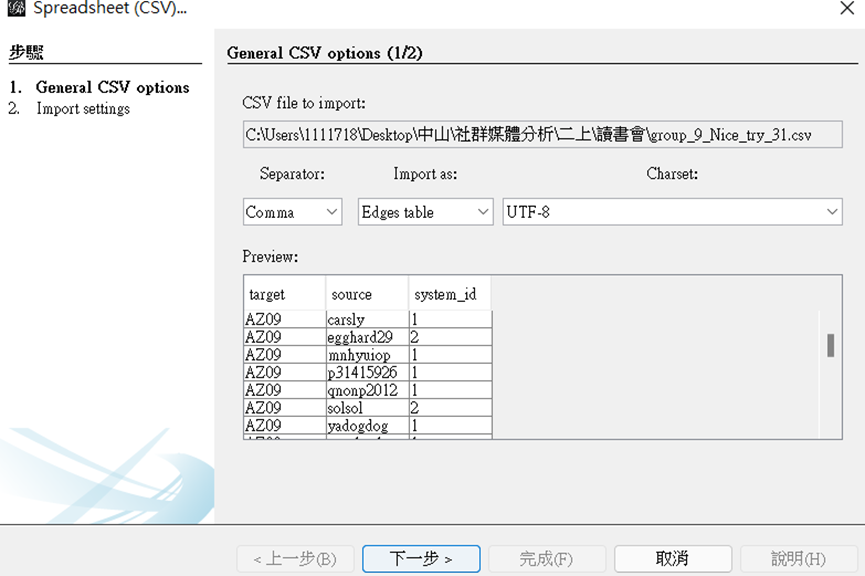
結果:

產出節點跟節點的結構圖，本處理資料筆數4502及1432個節點，呈現結果無法分析，本組組員欲調整參數，但因助教提過極為消耗資源無法調整，測試多次無法正常運作，故改採用Gephi工具執行後續分析。

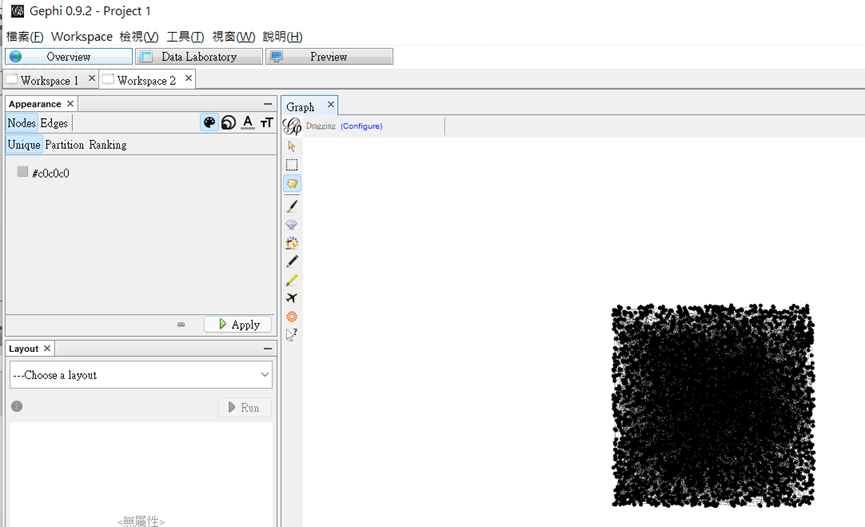
(四) 使用Gephi工具分析社會網路圖:

步驟1:並將以下工作平台資料匯入group\_9\_Nice\_try\_31.csv

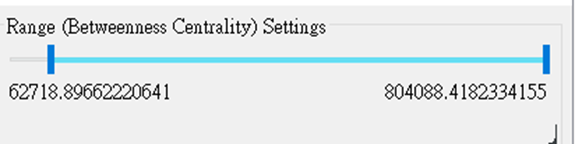


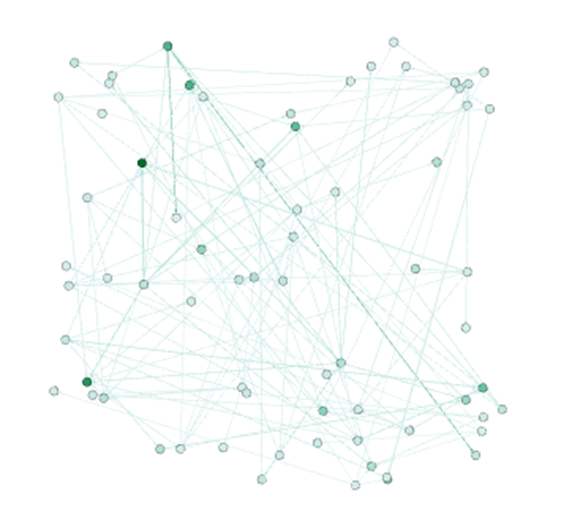


匯入及調整參數後結果:先期圖形結果仍無法清楚辨識內容



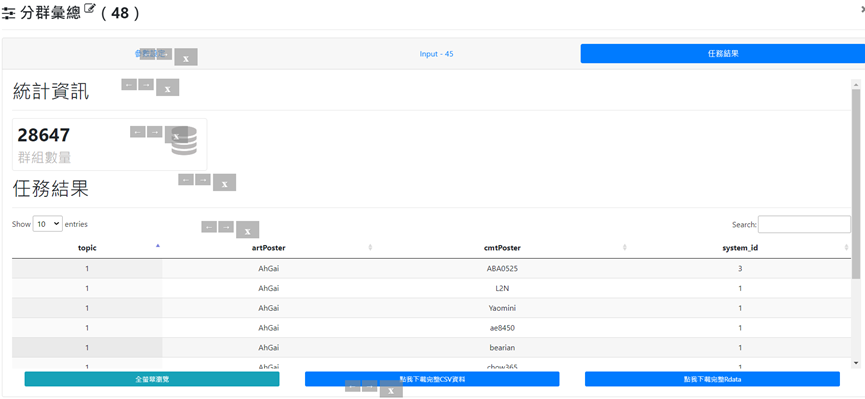
步驟2:調整node參數及Range





因資料結果為systemid故無法產出節點名稱，進行第二次資料處理及匯入作業

步驟3:下載工作平台Topic分群匯總group\_9\_Nice\_try\_31-topic.csv檔案並匯入Gephi工具，筆數28647

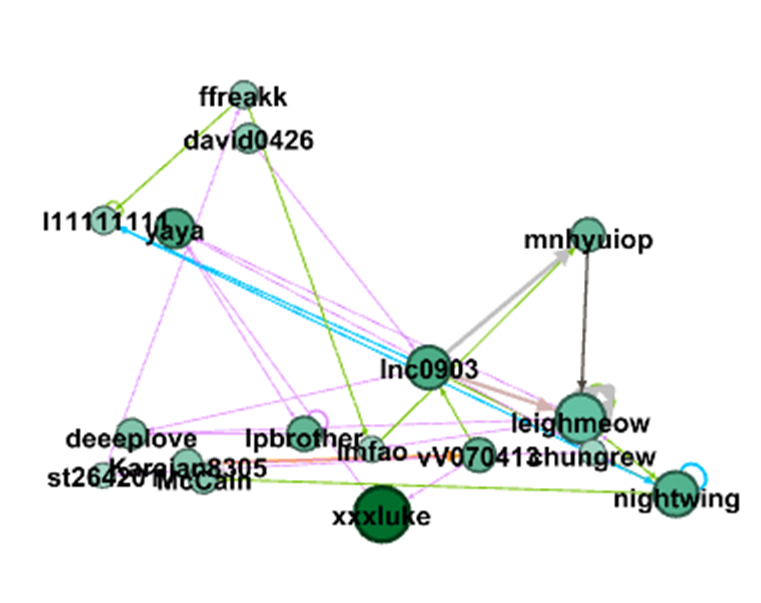


重複步驟匯入csv後調整Rang(Betweeness Centrality)參數及比例





執行結果:資料屬性已較能辨識之完整PO文及留言分布關係圖

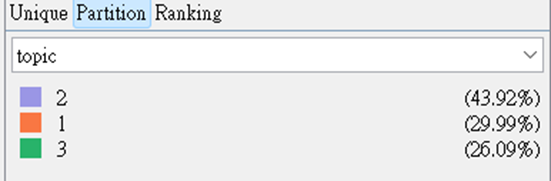


步驟4:重要節點的指標與該社會網路圖代表的意義:

說明:

步驟4.1:

以Topic分布百分比及發文及留言關連性



將關聯性及權重調整主要節點(degree centrality)分布如下，權重最高50如下

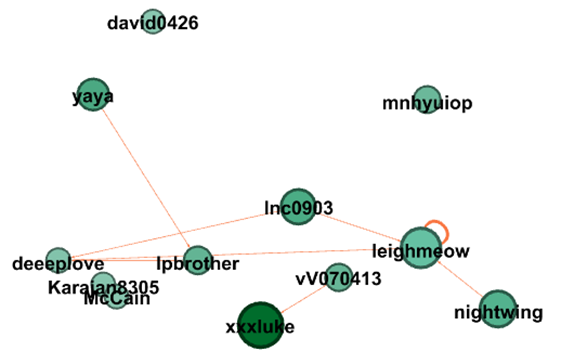
一張含有 文字 的圖片

自動產生的描述

社會網路分析: 參考LDA分析議題，以熱門PO文、留言者節點及方向性分析

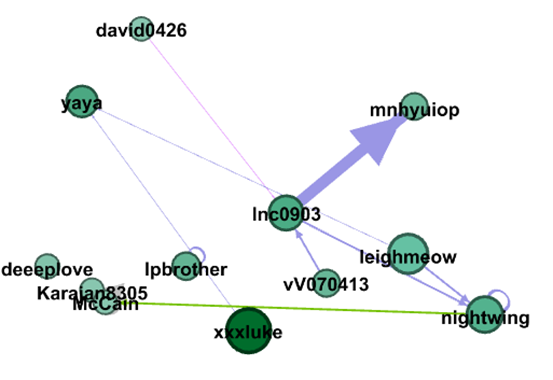
Topic-1(李靚蕾)之關係性(29.99%):最高為yaya、Inc0903、xxxluke…等

LDA議題:新聞、前妻、婚姻

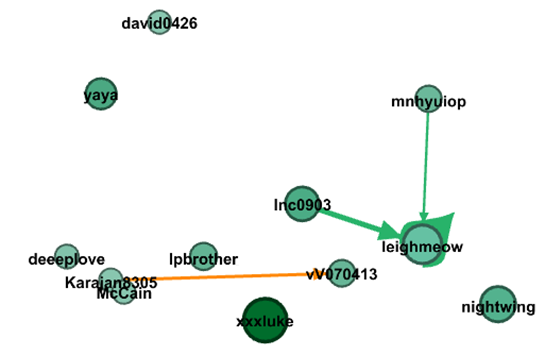


Topic-2(王力宏)之關係性(43.92%):mnyuiop、Inc0903、nightwing

LDA議題:力宏、八卦、有沒有



Topic-3(孩子)之關係性(26.09%):以leighmeow、vY070413最高，LDA議題:孩子、結婚、力宏



1. 結論
2. 我們發現對事件主題要有完整的了解，才能對LDA主題模型的分類做恰當的解釋。
3. LDA的混淆度(perplexity)越低越好，我們在調整參數後，混淆度從100以上調整到29.09，根據網路上找到的資料，介於60~20是不錯的模型，代表這樣的模型對於人類耳朵聽起來比較自然，也就是可以理解的程度比較好。
4. 在網路節點的定義和邊的權重部分，我們分別使用LDA主題、發文數量、情緒值大於0的文章數等三個資料來源，製作出社會網路圖，其中ha3810996在三個製作出來的社會網路圖中，都是最多人回應的，其主要原因為文章腥羶色內容吸引好奇八卦的網友熱烈回覆。
5. 字詞網路關聯圖做出來的結果與LDA模型的分類結果的關鍵字非常類似。主要的節點都有王力宏、王爸、李靚蕾、孩子。