
第三次讀書會報告

美國總統大選

第九組

TARFLOW

組員名單:

N124320001 沈世英

N124320005 趙千霈

N124320008 邱懷英

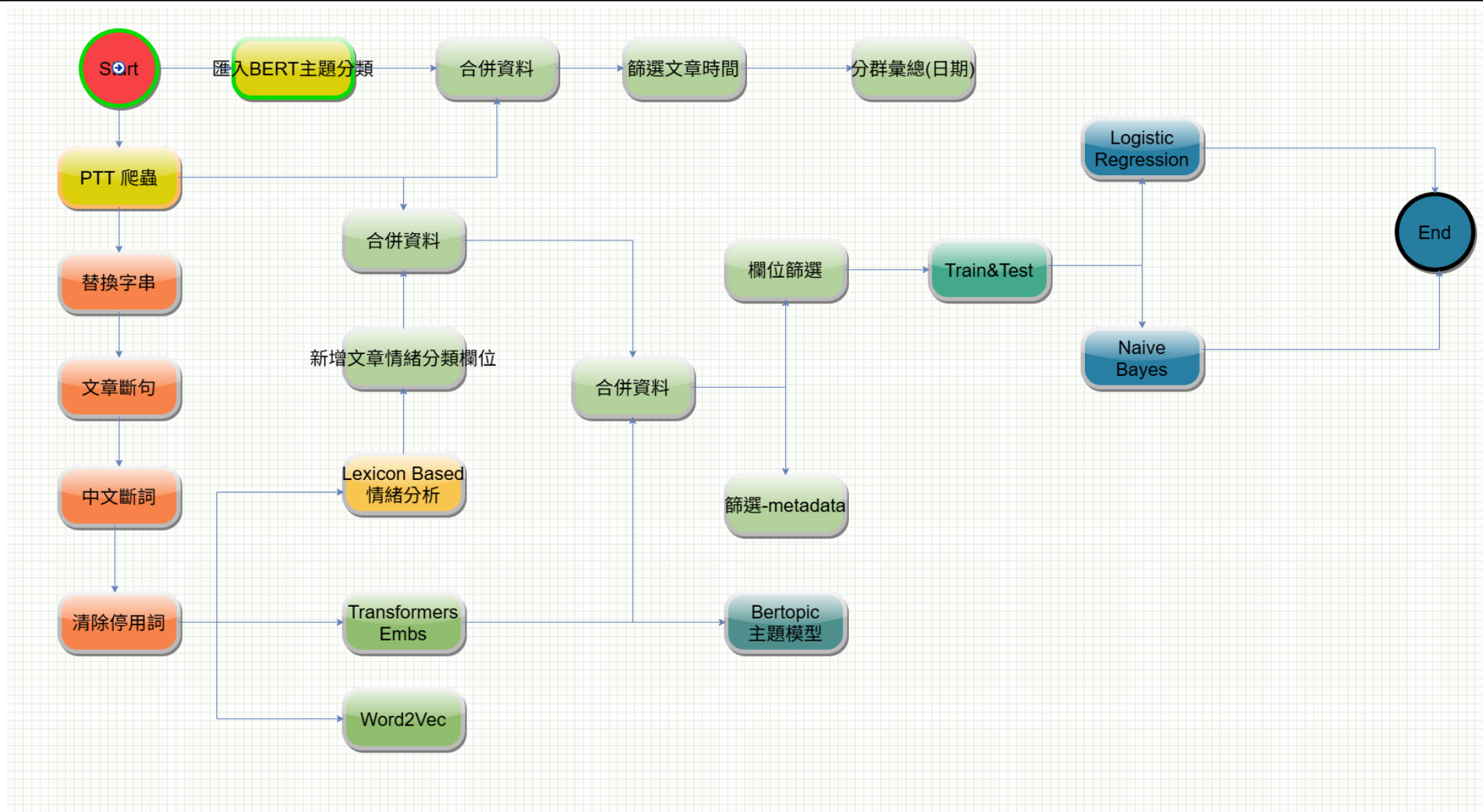
N124320014 陳柏豪

N124320018 邱翔

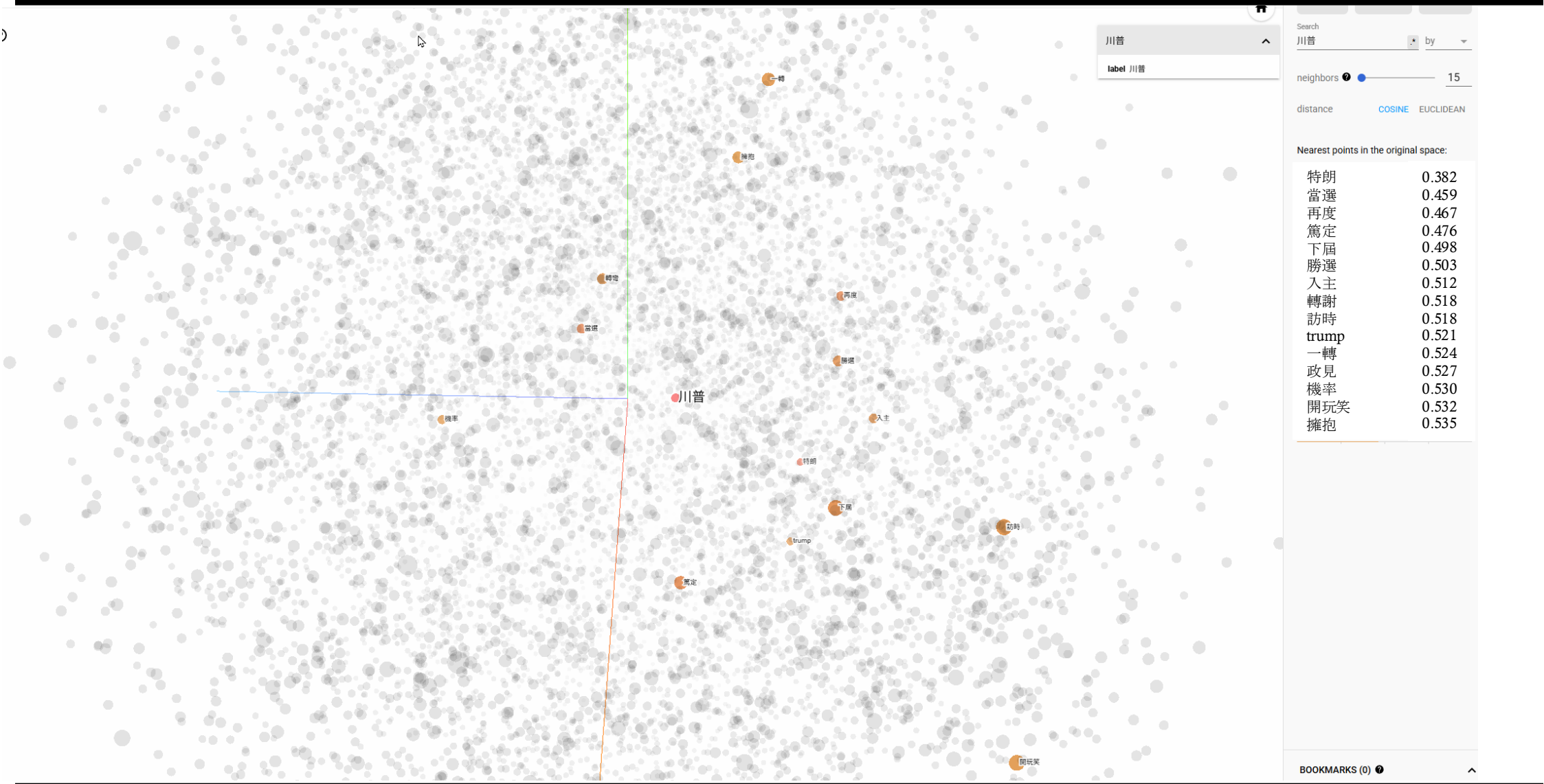
N124320021 永瑞華

N124320026 施泳呈

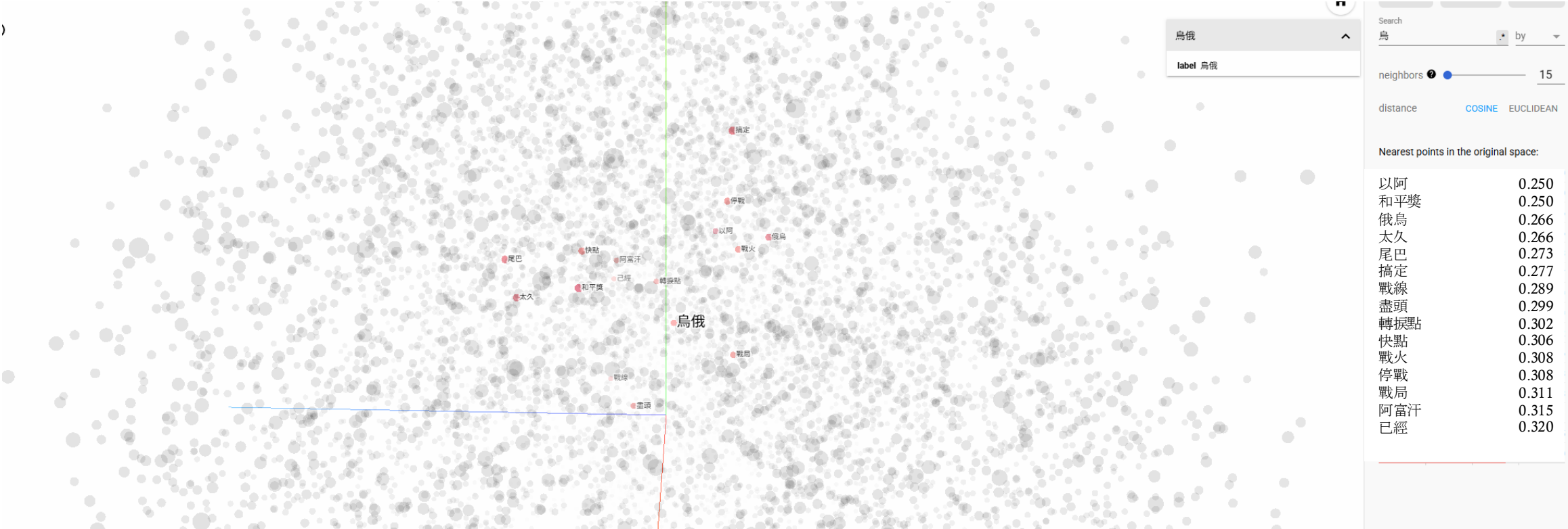
TARFLOW 流程說明



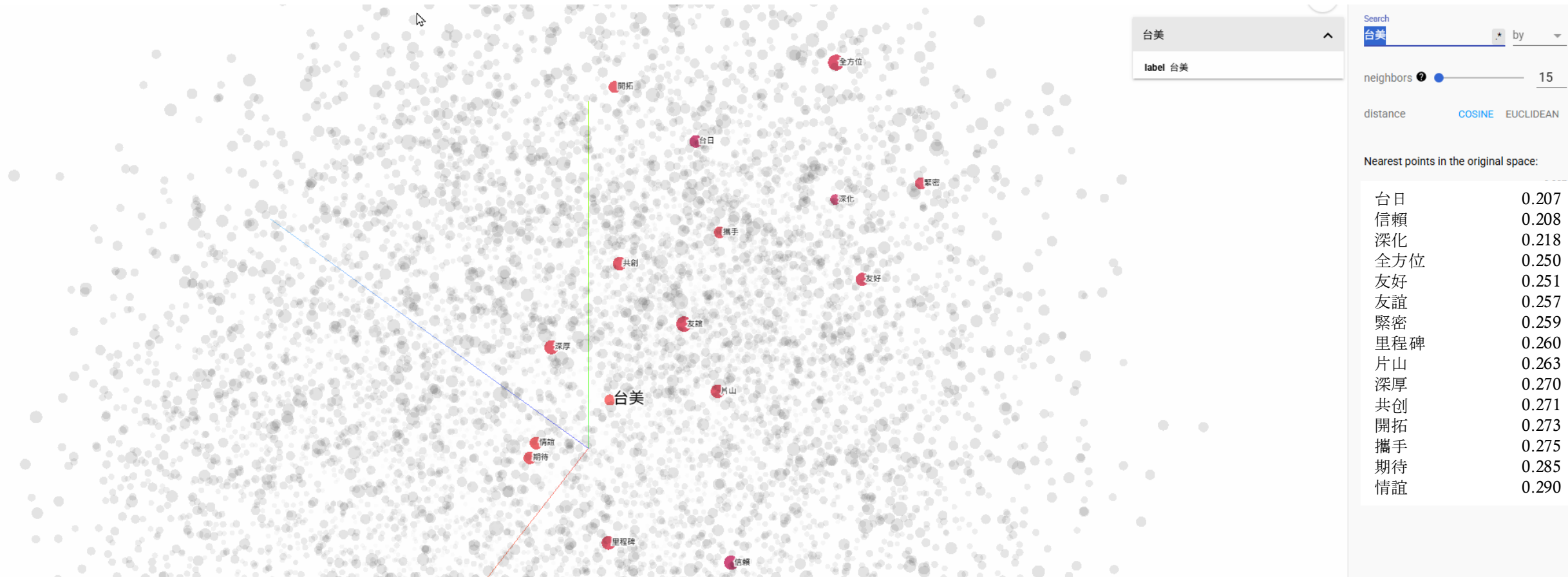
WORD2VEC-"川普"



WORD2VEC-"烏俄"



WORD2VEC-"台美"



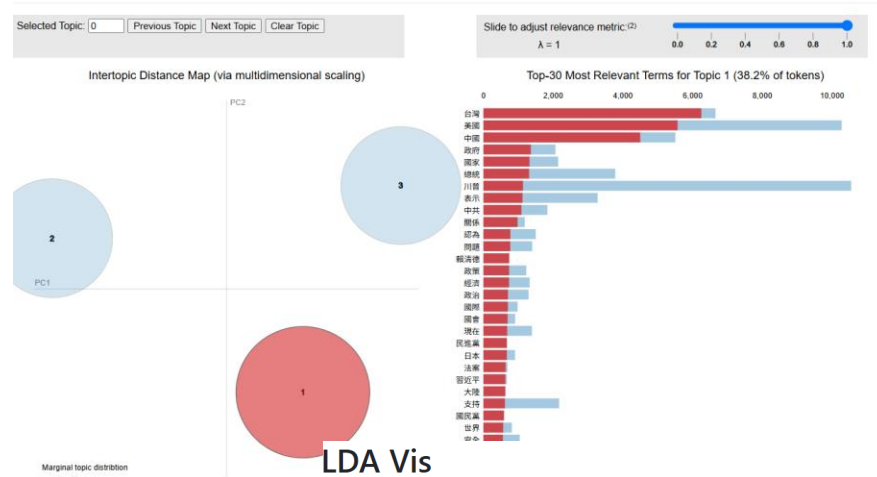
主題模型比較-LDA

烏俄戰爭

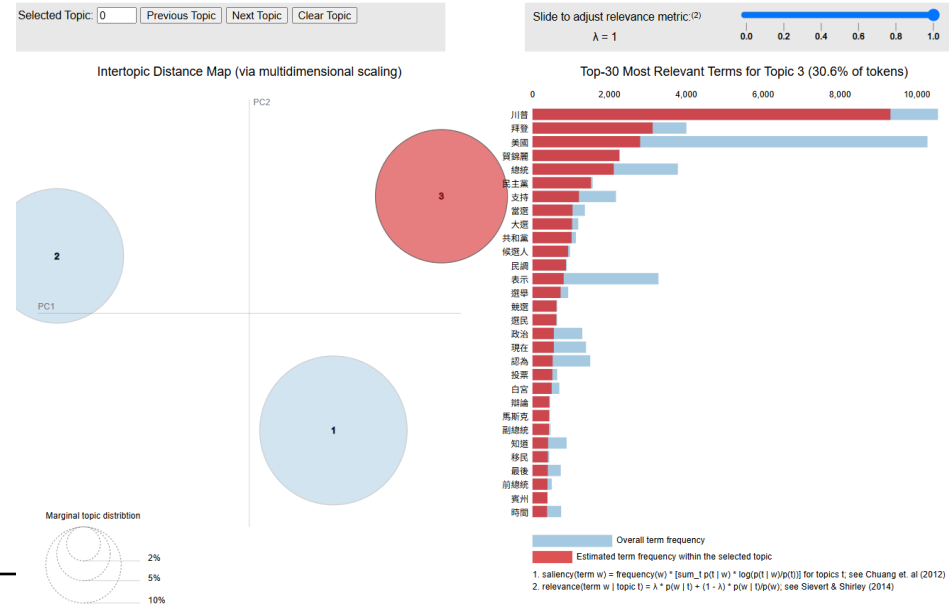
國家關係
(台美、兩岸關係)

美國總統

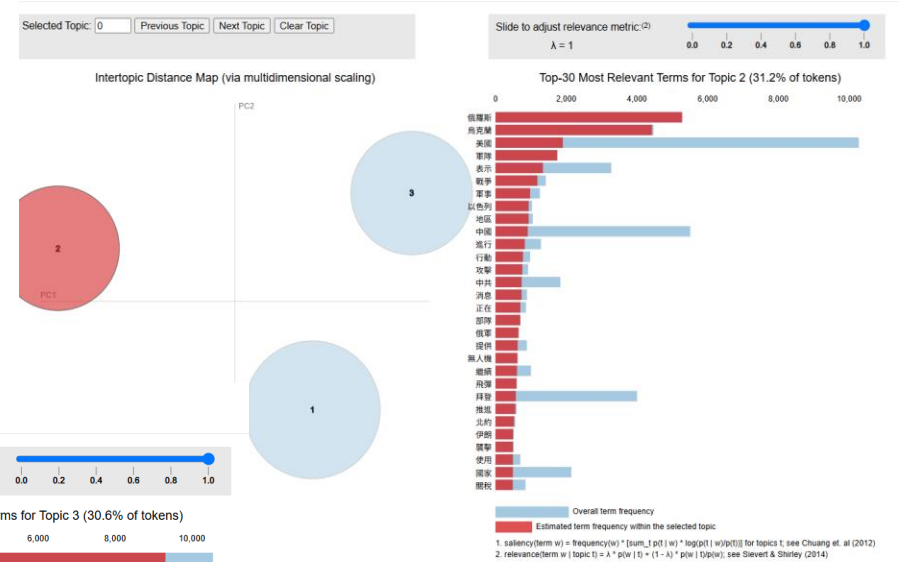
LDA Vis



LDA Vis



LDA Vis



主題模型比較-BERTOPIC(10)

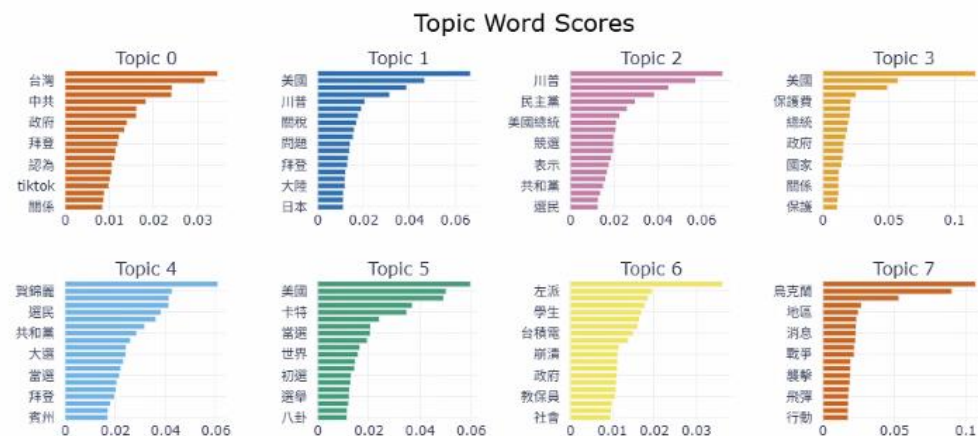
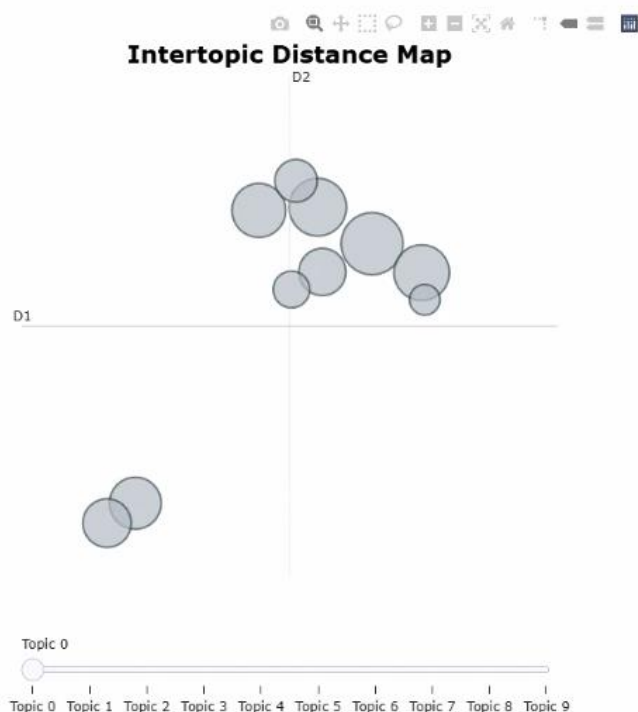
Bertopic 主題模型 (170)

參數設定

Input - 149

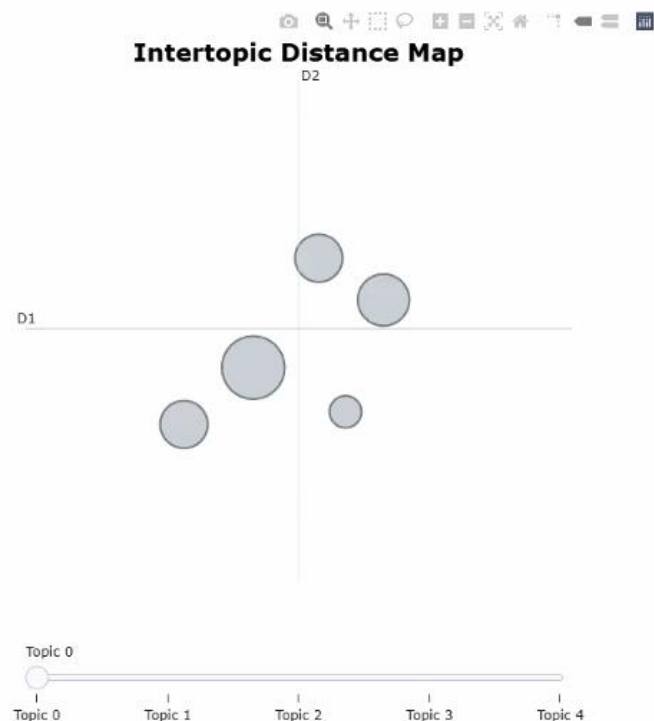
任務結果

BERTopic Vis



主題模型比較-BERTOPIC(5)

BERTopic Vis



情緒分類器比較

TF-IDF

Transformer Embedding

Naive Bayes (89)

參數設定

Input - 84

任務結果

統計資訊

0.035

訓練時間

0.652

測試資料macro-F1

0.67

測試資料加權精確率

Gaussian

分佈

0.048

推論時間

0.659

測試資料加權F1

0.668

測試資料micro召回率

0.668

測試資料準確度

0.668

測試資料micro精確率

0.654

測試資料macro召回率

0.668

測試資料micro-F1

0.671

測試資料macro精確率

0.668

測試資料加權召回率

參數欄位

Show10entries

0

Search:

Naive Bayes (89)

參數設定

Input - 84

任務結果

統計資訊

0.037

訓練時間

0.052

推論時間

0.629

測試資料準確度

0.629

測試資料micro-F1

0.619

測試資料macro-F1

0.625

測試資料加權F1

0.629

測試資料micro精確率

0.625

測試資料macro精確率

0.627

測試資料加權精確率

0.629

測試資料micro召回率

0.619

測試資料macro召回率

0.629

測試資料加權召回率

Gaussian

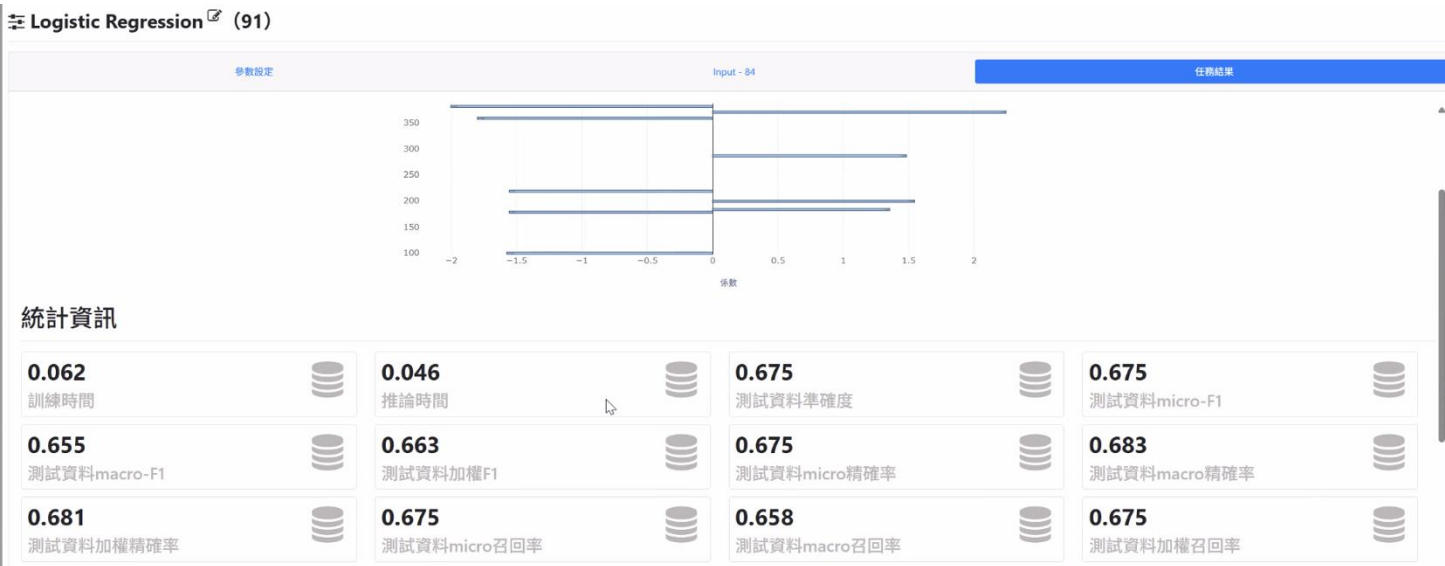
分佈

情緒分類器比較

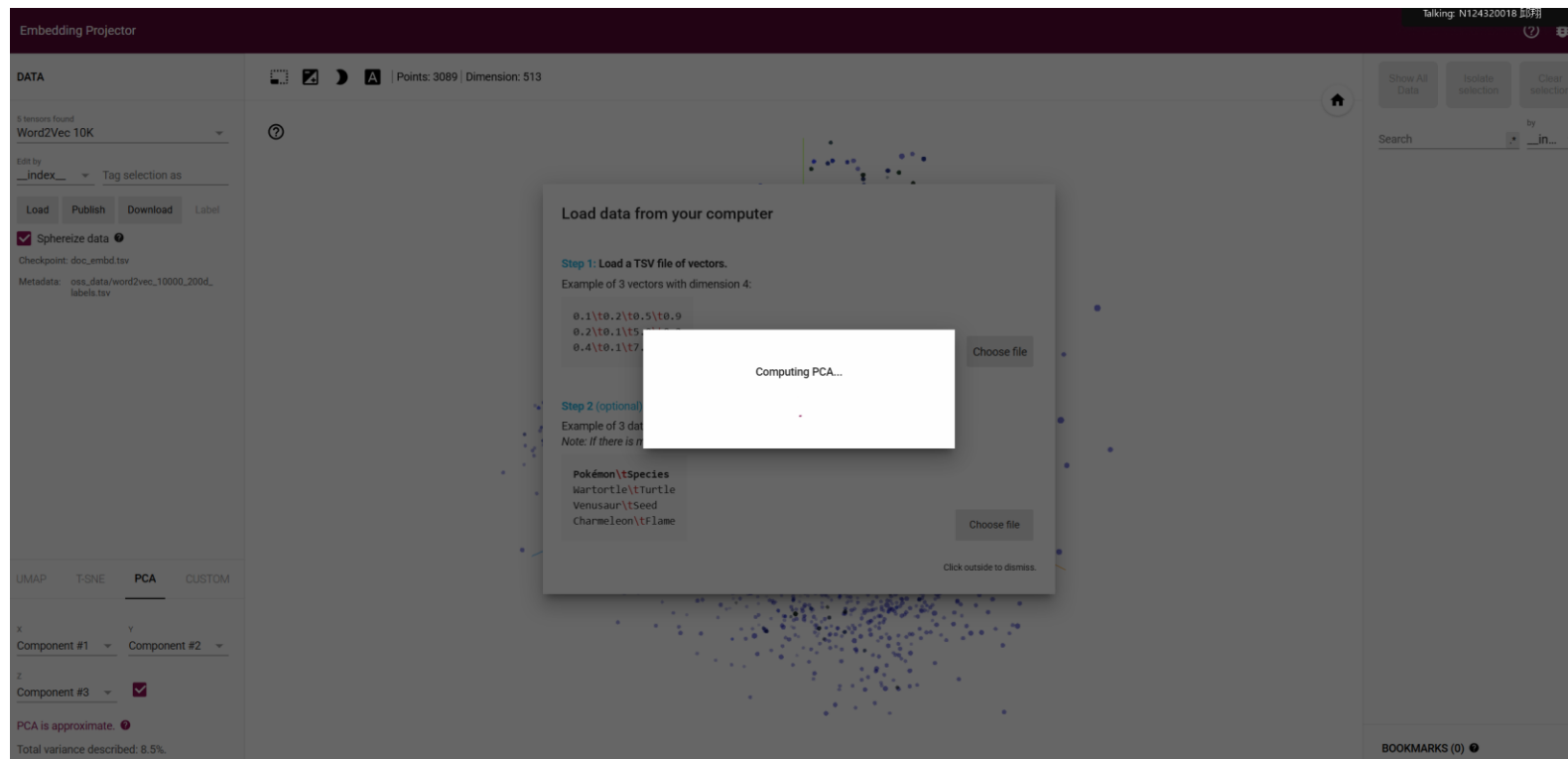
TF-IDF



Transformer Embedding

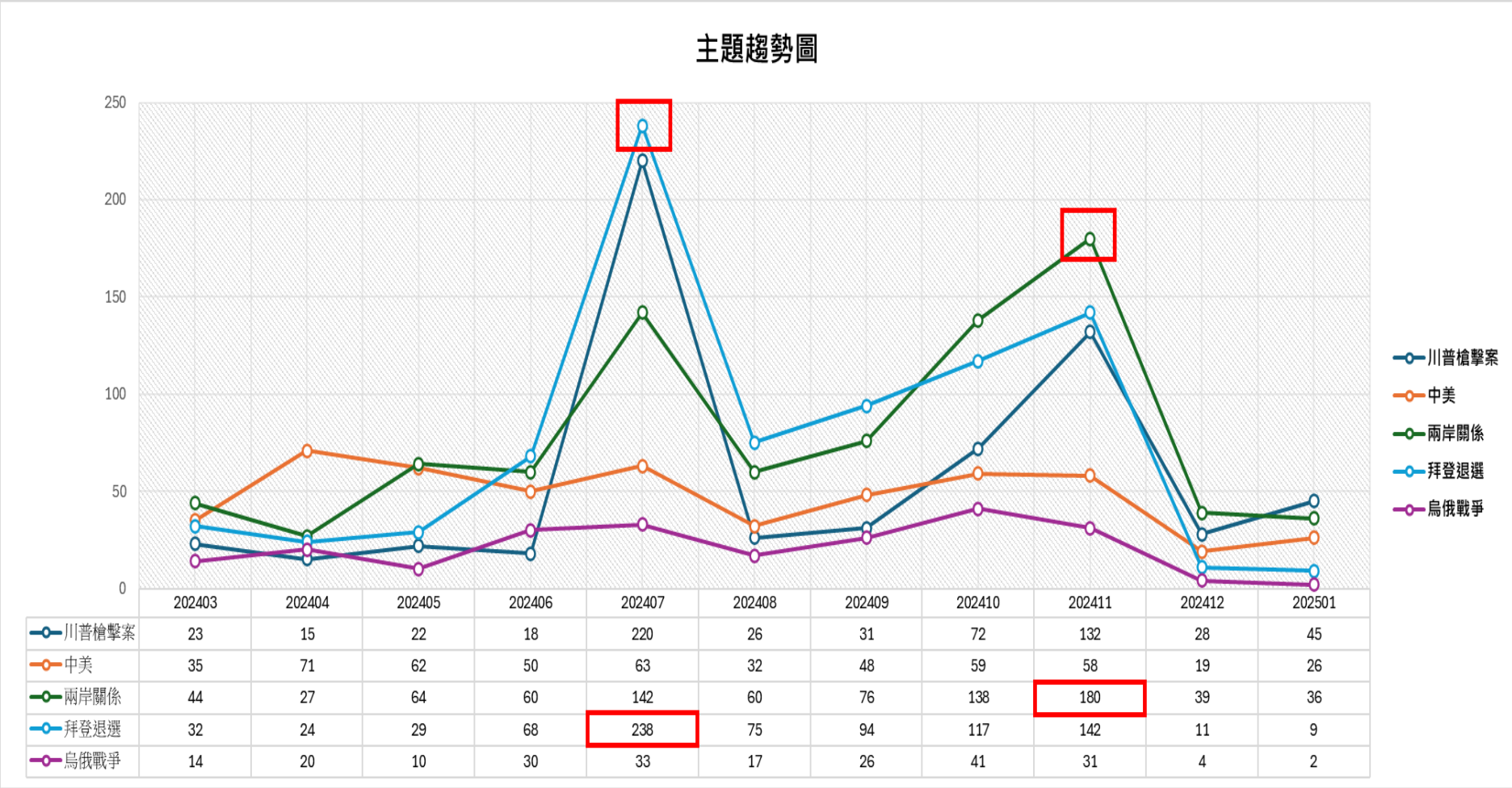


以情緒分類再做一次，可惜資料量太大無法QQ



主題趨勢圖(討論文章數)

日期	事件描述
2024/3/12	川普鎖定提名：獲共和黨總統候選人資格。
2024/6/27	首場辯論：拜登表現受批評，健康受質疑。
2024/7/13	川普遭槍擊未遂：賓州集會遇襲，右耳受傷。
2024/7/15	共和黨提名：川普獲提名，J-D·萬斯任副手。
2024/7/21	拜登退選：支持賀錦麗接棒競選。
2024/7/22	賀錦麗獲提名：民主黨正式提名為總統候選人。
2024/9/10	第二場辯論：川普與賀錦麗於費城交鋒。
2024/9/15	川普再遭刺殺未遂：佛州遇襲，嫌犯被捕。
2024/9/24	賀錦麗辦公室遭槍擊：亞利桑那州發生槍擊，無人傷亡。
2024/11/5	大選日：川普勝出，獲 312 張選舉人票。
2025/1/20	川普就職：華盛頓宣誓，展開第二任期。



結語

- BERTopic的主題分類較LDA更為明確，出現"槍擊案"與"拜登退選"
- 主題數太多仍會失焦，主題間的相似性高
- 使用tf-idf做情緒分類器的效果較Transformer Embedding佳，因情緒分類常常跟明確的字詞有關，而TF-IDF 是直接抓關鍵詞的專家。不需要理解上下文、語氣或多輪對話，反而更直接有效
- 主題趨勢圖與實際事件時間軸吻合