
第二次讀書會報告

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

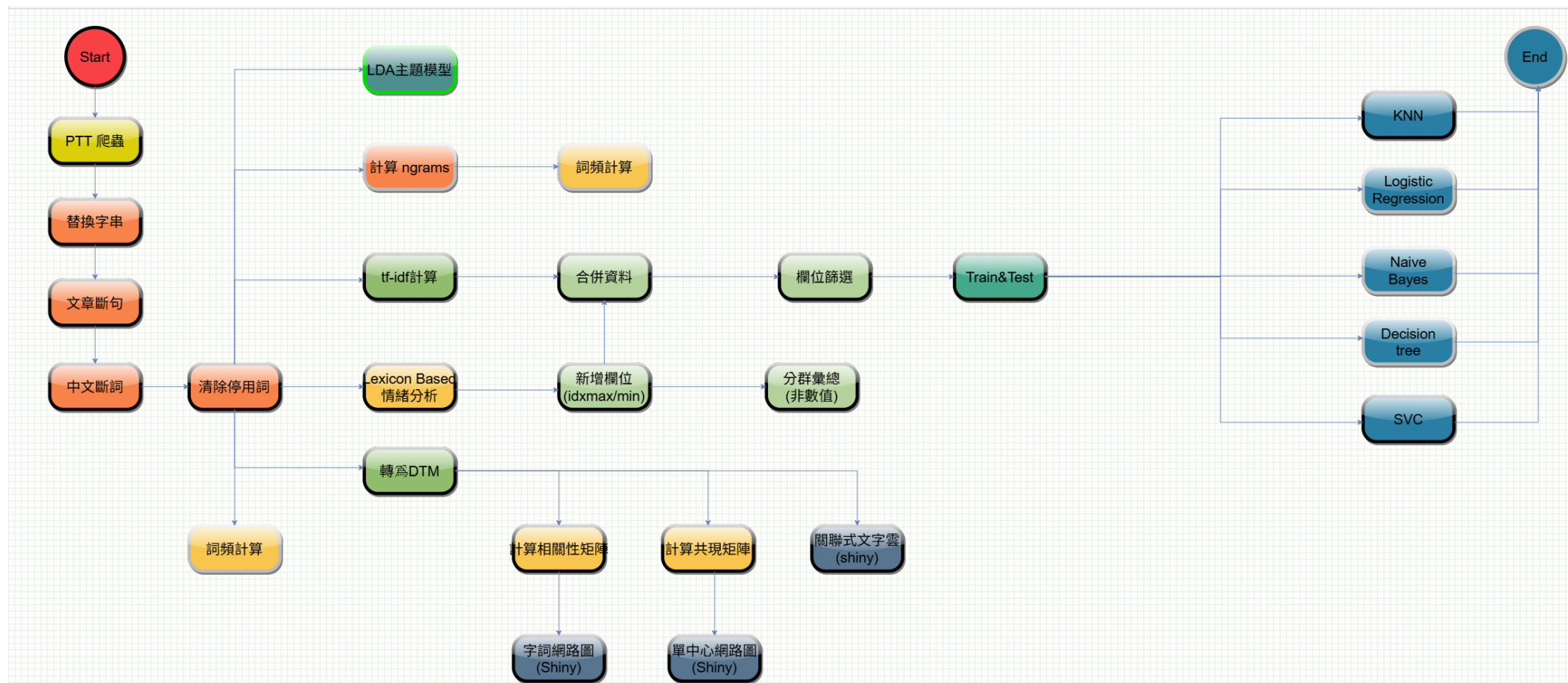
N124320026 施泳呈

專案動機

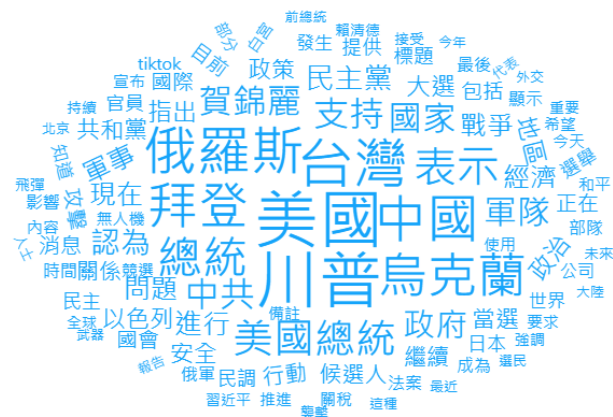
現在的人越來越習慣在網路上討論政治，PTT 這類匿名論壇更是許多台灣網友發表意見的重要地方。特別是在美國總統大選期間，PTT「美國總統大選」政黑板（HatePolitics）、八卦版（Gossiping）上會有大量的討論，這些留言可能反映網友對候選人的看法，甚至影響其他人的意見。

但因為這些討論內容多是文字，意見也非常多元，很難直接用肉眼判斷整體的情緒趨勢及關連性。因此，本研究希望分析 PTT 上的留言情緒及文字關係，看看大家對不同候選人的態度是偏向支持、反對，還是中立。進一步來說，我們也想了解這些情緒變化及字詞的相關性。這樣的研究不但能幫助我們更了解網路輿論的變化，甚至對更了解網路輿論主要的內容有甚麼關聯，甚至對未來的政治情緒分析或選舉預測也可能有所幫助。

TARFLOW 流程說明

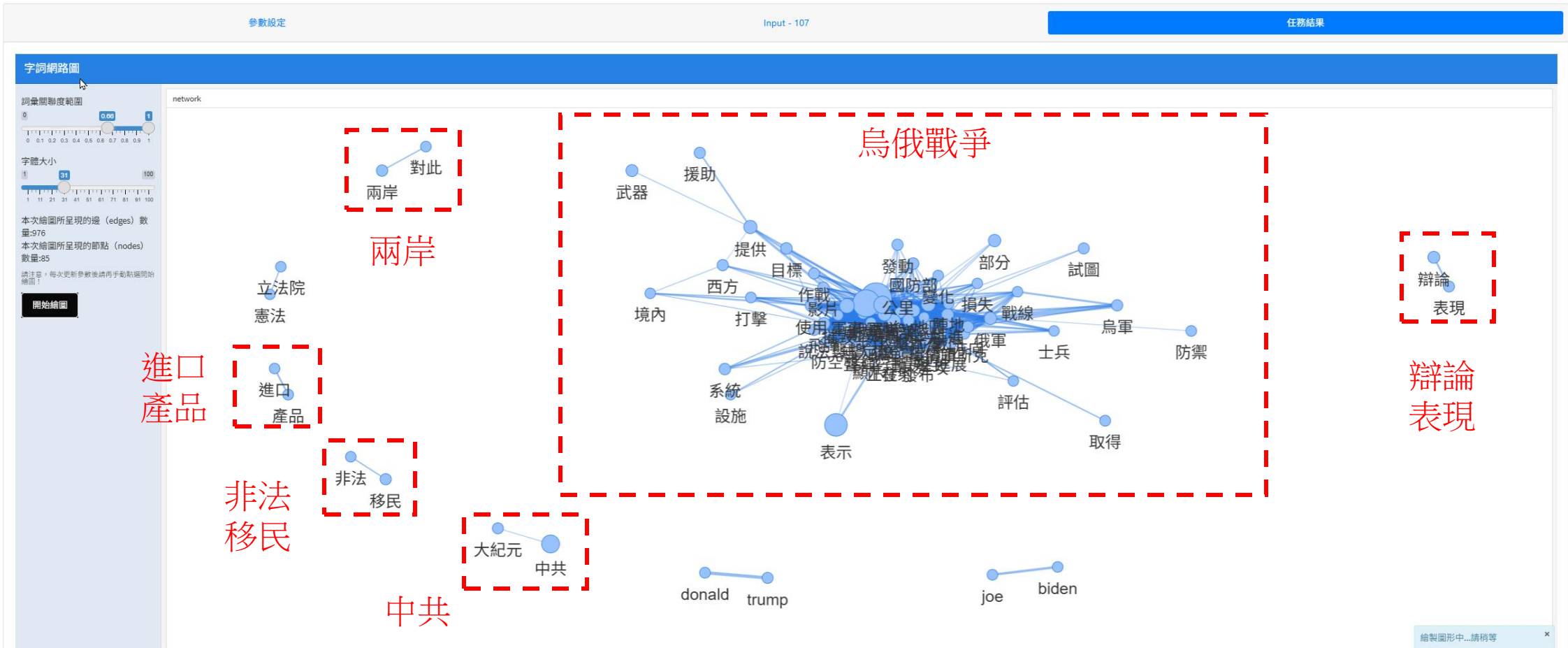


N-GRAM(N=2) VS 原始文字雲



字詞網路圖

字詞網路圖 (Shiny) (116)



關聯式文字雲



SVC

SVC

(92)

參數設定

Input - 84

任務結果

統計資訊

1.177

訓練時間

0.71

測試資料macro-F1

0.73

測試資料加權精確率

1

懲罰係數

0.667

推論時間

0.716

測試資料加權F1

0.723

測試資料micro召回率

rbf

核函數

0.723

測試資料準確度

0.723

測試資料micro精確率

0.709

測試資料macro召回率

3

維度

0.723

測試資料micro-F1

0.733

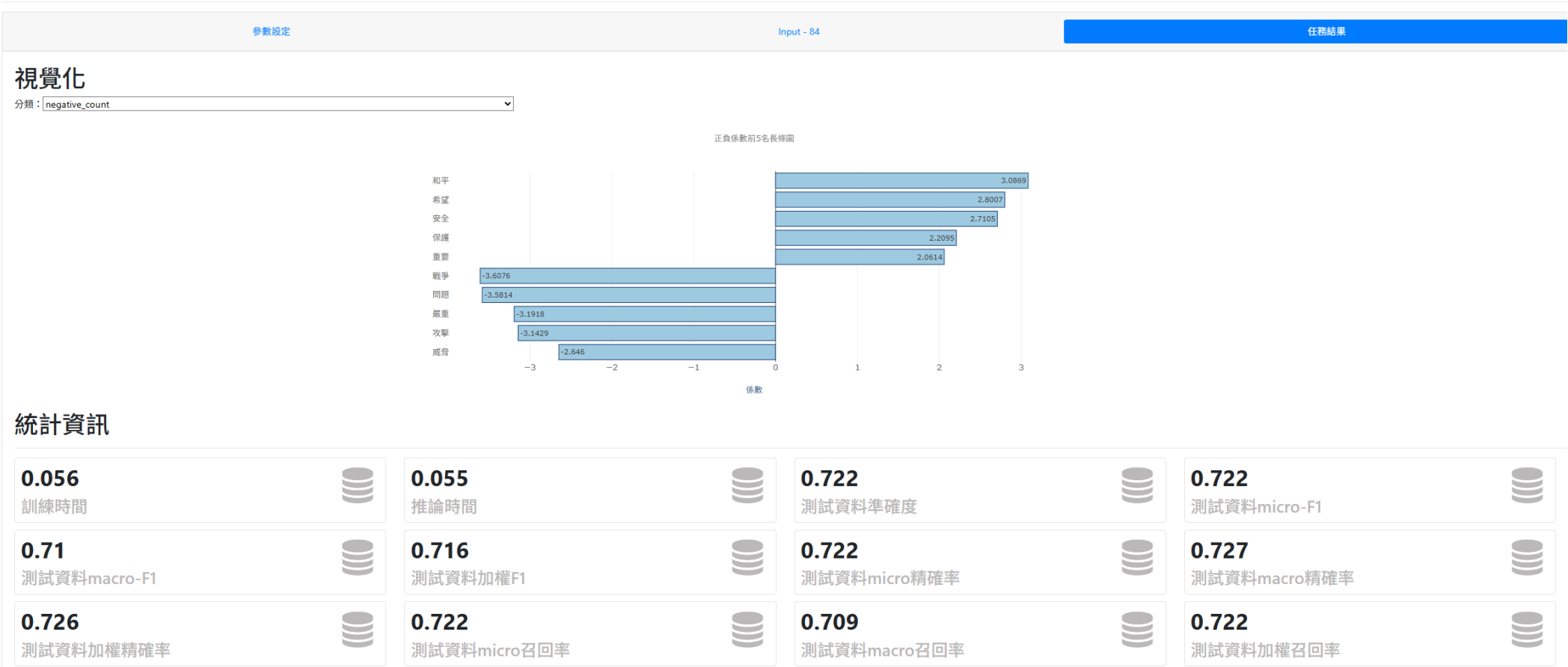
測試資料macro精確率

0.723

測試資料加權召回率

LOGISTIC REGRESSION

Logistic Regression (91)



主題模型

LDA Vis

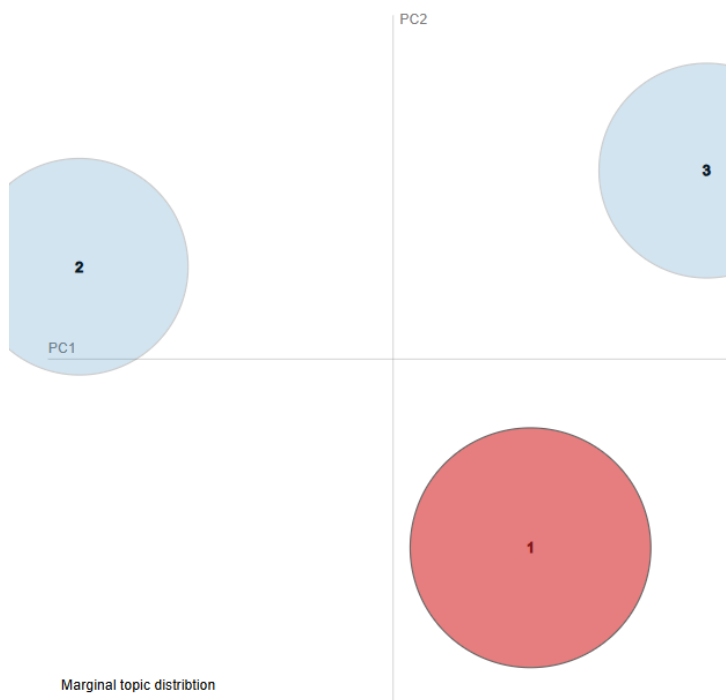
Selected Topic:

Slide to adjust relevance metric:⁽²⁾

$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

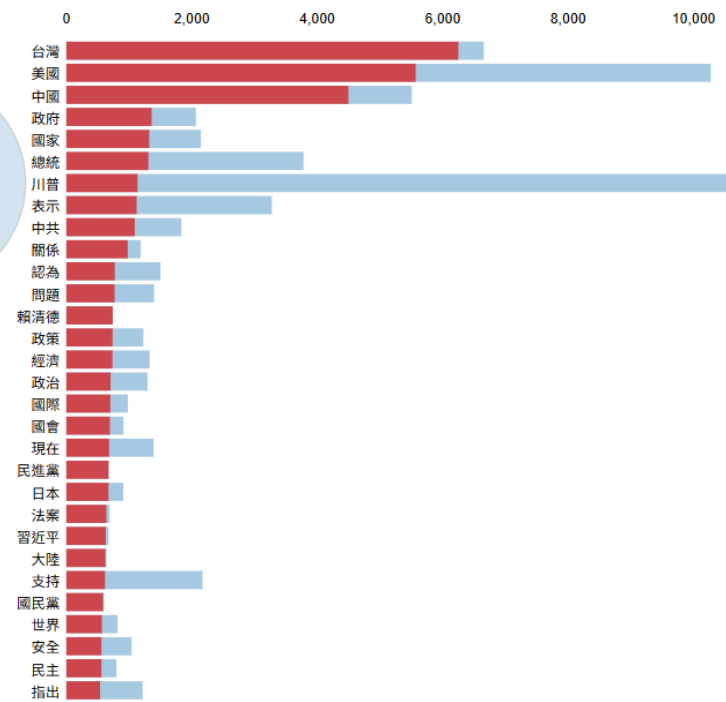
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 1 (38.2% of tokens)



Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))]; see Chuang et. al (2012)

2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

主題模型

LDA Vis

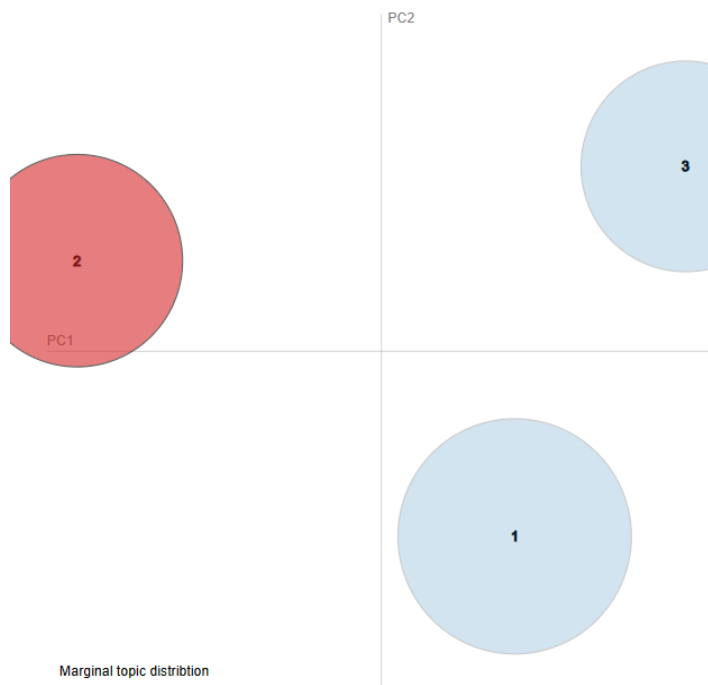
Selected Topic:

Slide to adjust relevance metric:⁽²⁾

$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

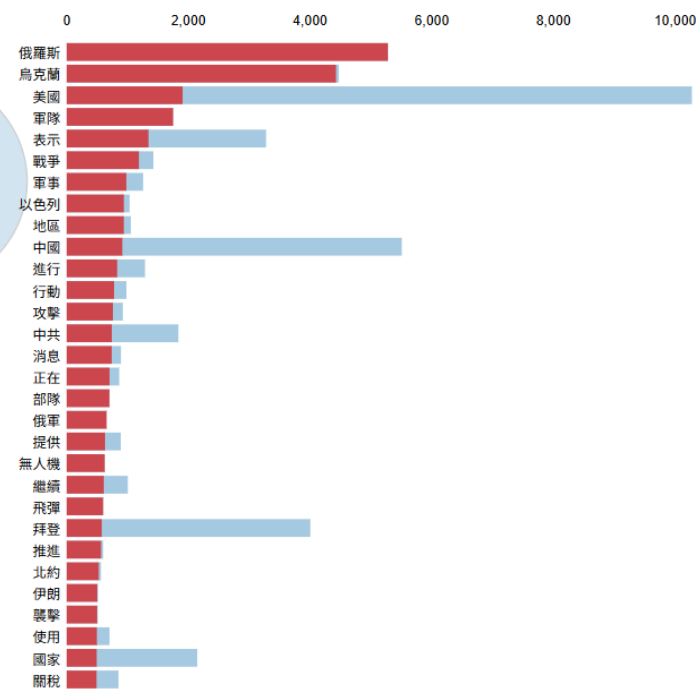
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 2 (31.2% of tokens)



Overall term frequency

Estimated term frequency within the selected topic

1. $\text{saliency}(\text{term } w) = \text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w)/p(t))]$ for topics t ; see Chuang et. al (2012)

2. $\text{relevance}(\text{term } w | \text{topic } t) = \lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

主題模型

LDA Vis

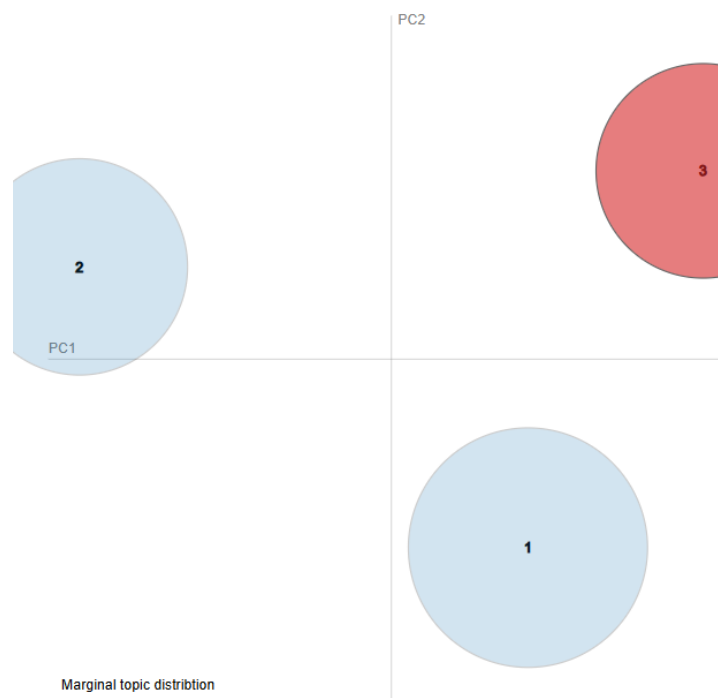
Selected Topic:

Slide to adjust relevance metric:⁽²⁾

$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

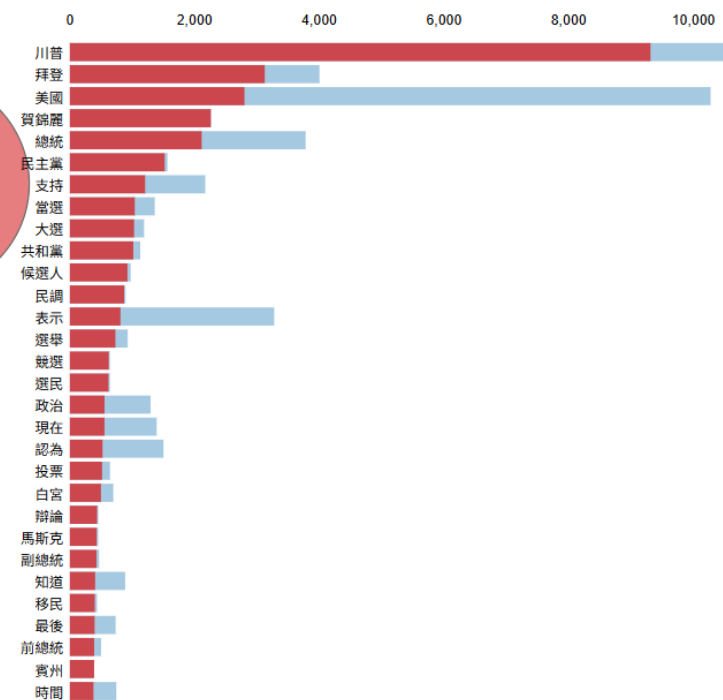
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 3 (30.6% of tokens)



Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)

2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

結語

- 本研究以 PTT 上的留言為分析對象，觀察台灣網友對美國總統大選的情緒反應。
- 顯示台灣網友對國際政治具高度關注，且能透過網路論壇即時參與討論與表達觀點。
- 發現特定事件（如俄烏戰爭、中共議題、非法移民等）會顯著影響討論熱度與情緒波動。
- 藉由文字探勘與情緒分類模型（如 SVC、Logistic Regression），掌握網友對候選人的正負面情緒傾向。
- 主題模型結果分為三個主題美國選舉、烏俄戰爭、兩岸關係
- 藉由分析網路輿情，有助於未來進行政治預測、媒體監控與公共政策回應。