社群媒體分析

期末專案報告

主題：地方小吃

第 22 組

周嵩智 N104220009
吳尚蓉 N104220010
林素芬 N104220024
陳嬿雅 N104220026

目錄

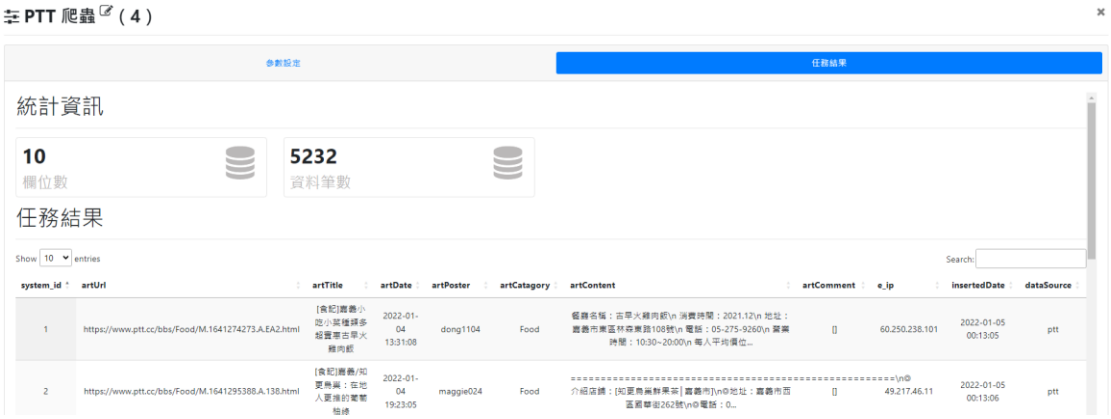
第1章　動機和分析目的

台灣作為一個美食文化豐富的地區，各地的小吃具有獨特的地理分布和口味特點。在台灣的 PTT 美食論壇上，許多網友分享和討論他們對各類小吃的喜好和見解。然而，過去研究尚未充分利用這些寶貴的數據資源，深入探討台灣小吃的地理分布和口味特點。因此，本研究以 "社群媒體分析" 課程專案報告為契機，利用 PTT 美食論壇的數據，全面分析台灣地區的小吃感興趣進而討論的主題為何，以期提供對台灣美食文化和產業的新視角和啟示。

第 2 章 資料集描述

資料來源：PTT 美食版/台中版/新竹版/台南版/高雄版/八卦版
資料範圍：2022/1/1~2023/5/31
搜尋關鍵字：小吃

PTT 爬蟲任務結果：資料 5,232 筆

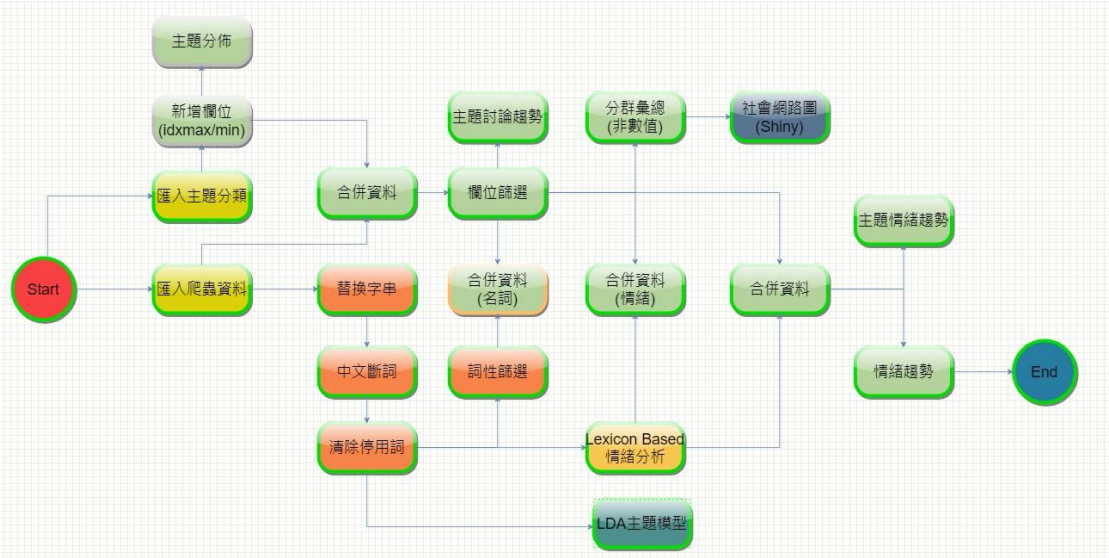


第 3 章 資料分析說明

第 1 節 TARFlow 流程

TARFlow 工作流程檔名：Taiwan_food

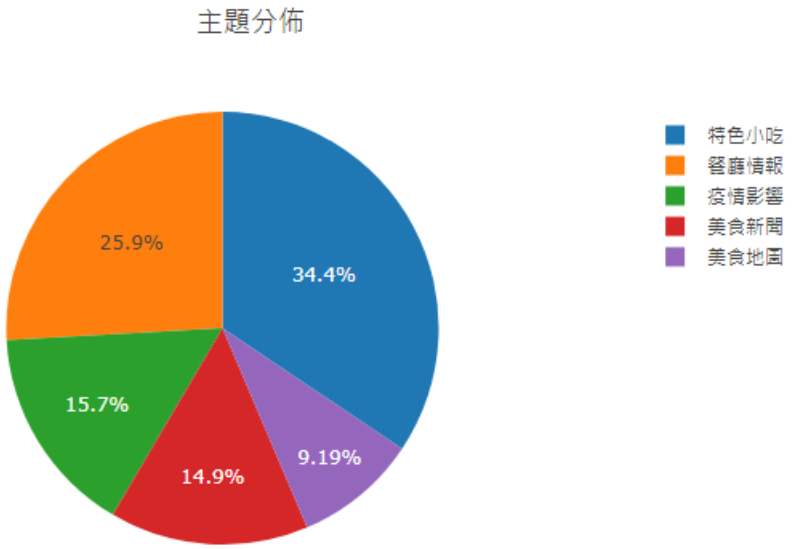
- 爬蟲
- 斷詞
- 文章情緒分析
- LDA 主題分類
- 合併主題與情緒分數
- 製作社會網路圖



第 2 節 視覺化圖表分析

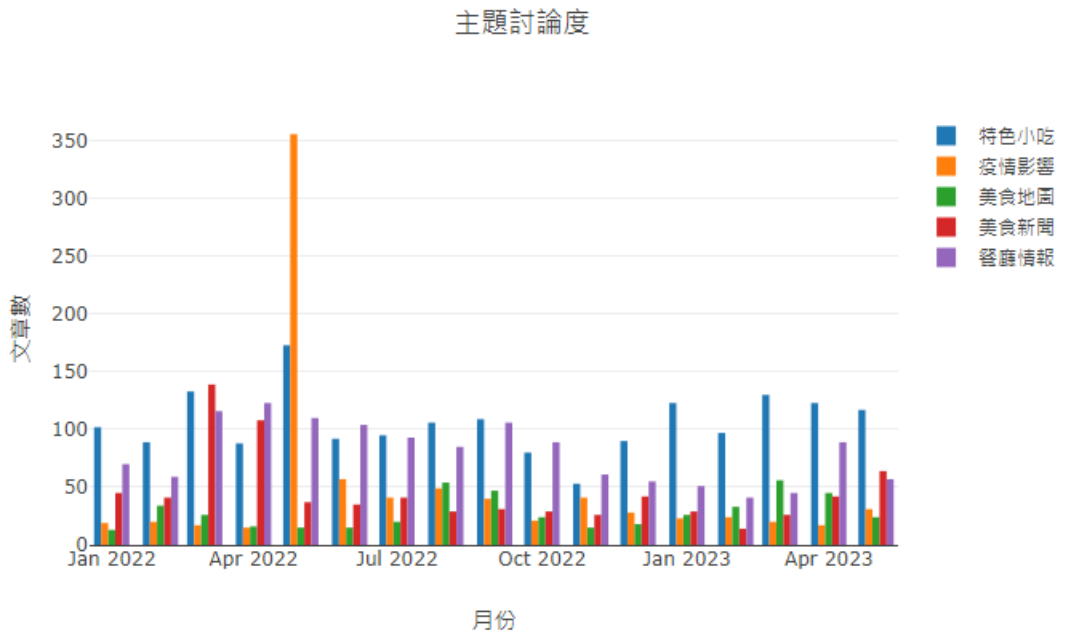
第1項 LDA 主題分佈

根據主題分布提供的數據，我們可以觀察到以下 5 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特色小吃、餐廳情報、疫情影響，可見文章熱度還是圍繞著特色小吃及影響美食的議題上。



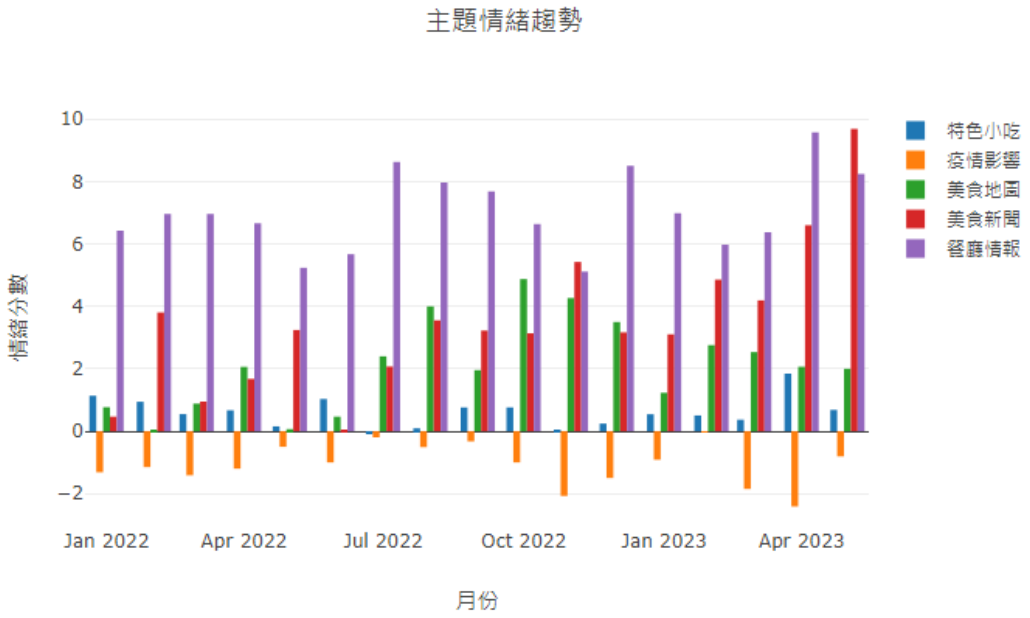
第 2 項 討論度

而關於主題討論度的部分，可以觀察出隨著時間的推移，特色小吃的討論熱度不減；在**疫情影響**的部分則是在 2022 月第 2 季因疫情影響而達到高峰。



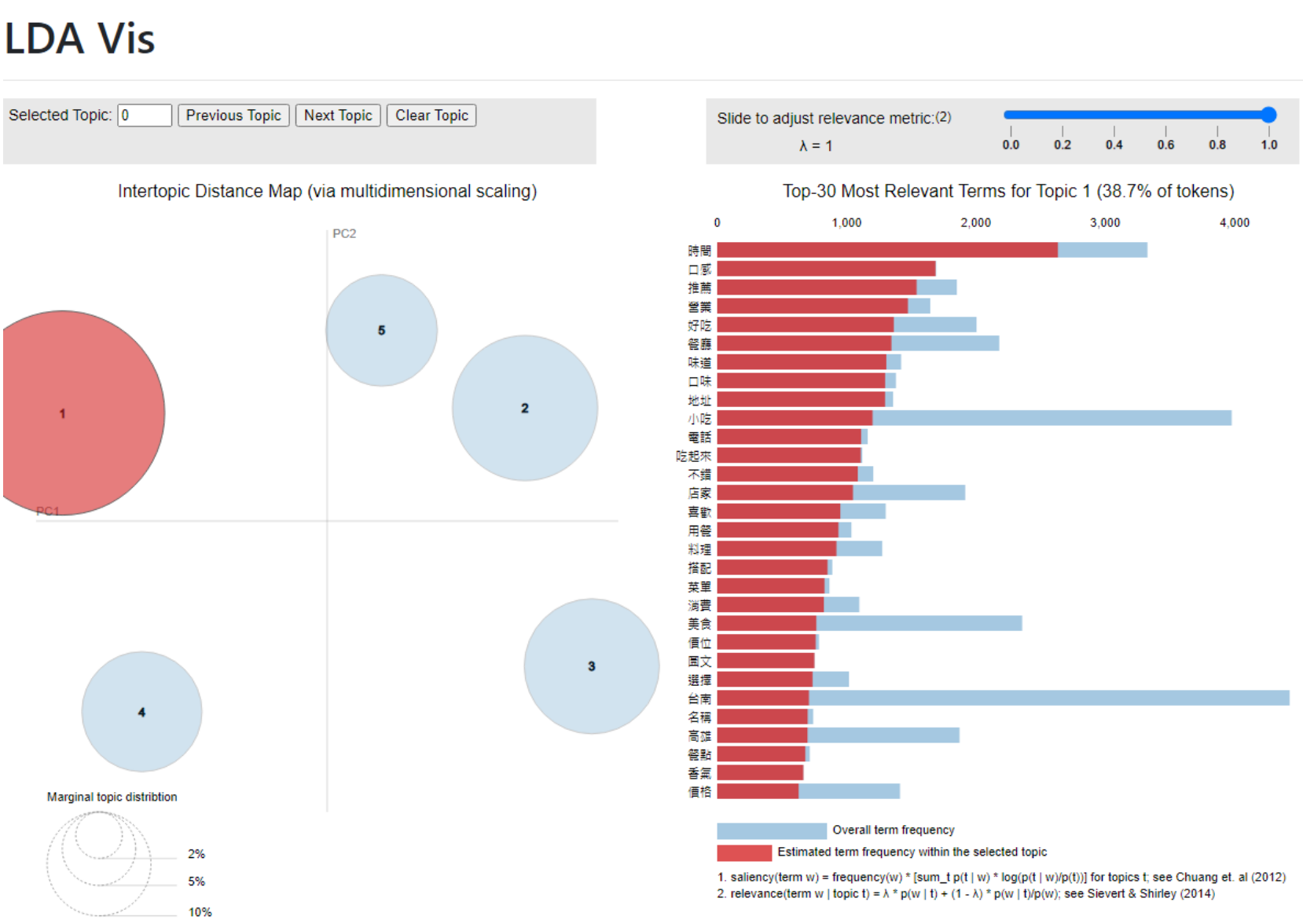
第 3 項 情緒趨勢

根據下圖可以明顯看出**疫情影響**的情緒分數對我們的主題，一直是負面情緒，可解釋為疫情不可外出，對民眾外食的影響很大，加上台灣夜市文化盛行，所以大家的情緒偏向負面；美食新聞的主題，則在 2023 慢慢解封後，情緒愈高昂。



第 3 節 LDA 主題模型 -因五項分類混淆度最低，且各項無重疊，故本組認為 5 項為最佳分類。

第 1 項 餐廳情報



第 2 項 美食新聞

LDA Vis

Selected Topic: 0 Previous Topic Next Topic Clear Topic

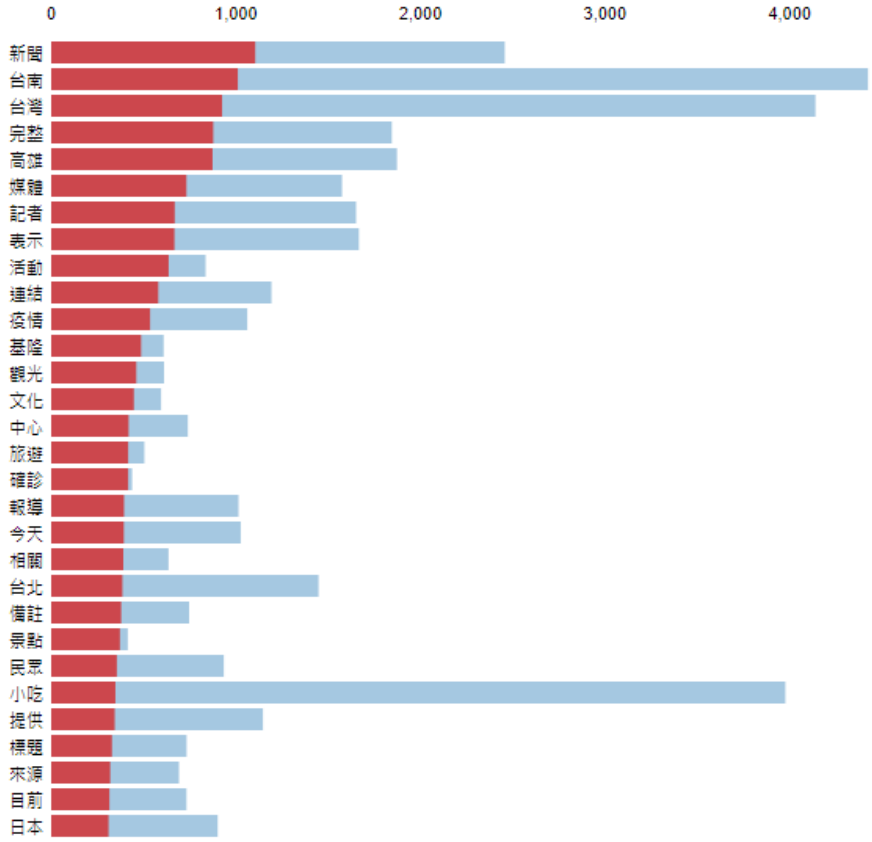
Slide to adjust relevance metric:(2)

λ = 1

0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)

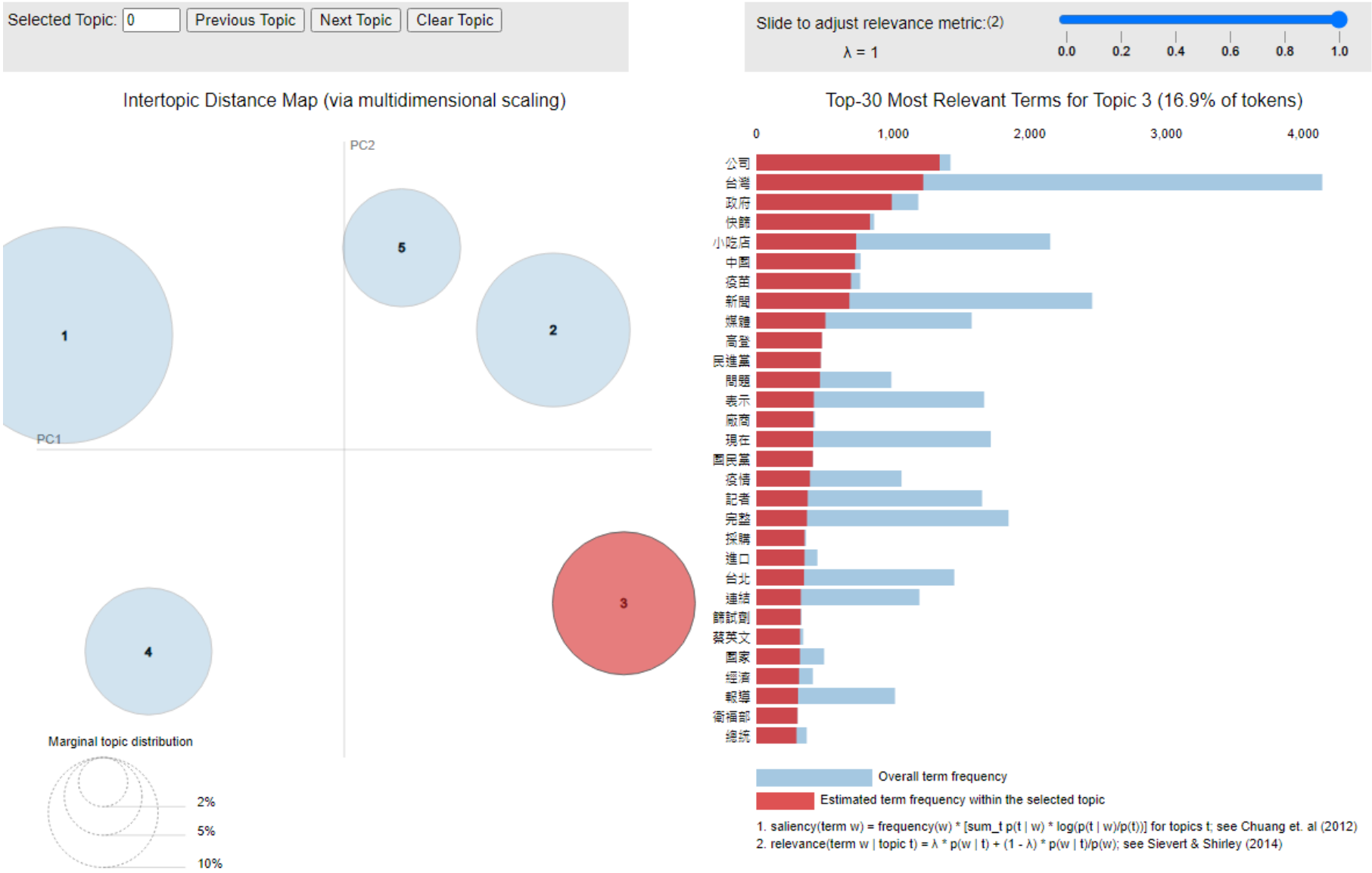
Top-30 Most Relevant Terms for Topic 2 (19.5% of tokens)



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) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

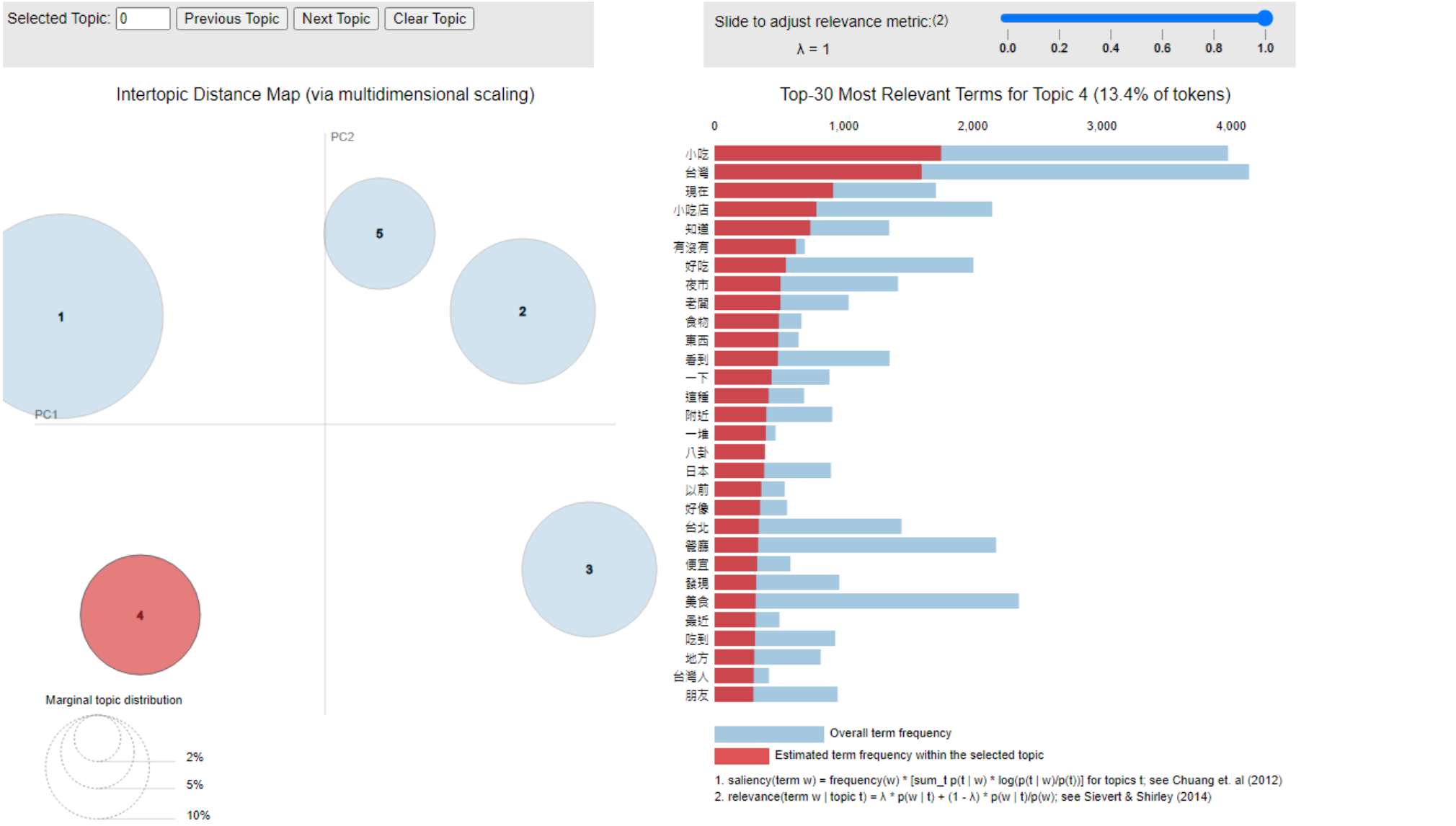
第 3 項 疫情影響

LDA Vis



第 4 項 特色小吃

LDA Vis



第 5 項 美食地圖

LDA Vis

Selected Topic: 0 Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:(2)

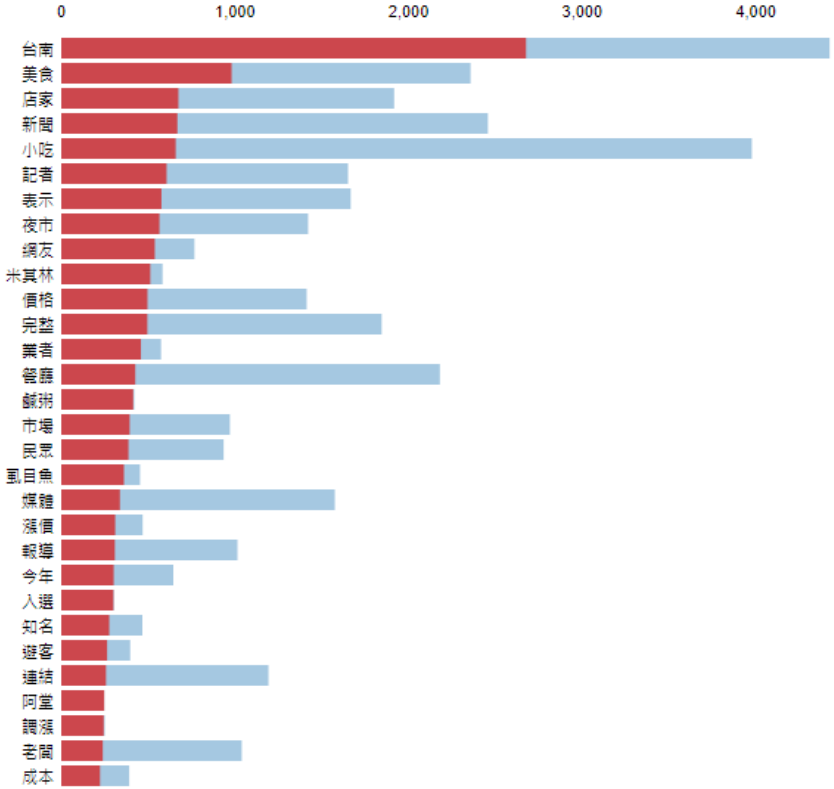
λ = 1

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 5 (11.5% of tokens)



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) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

第 8 項 五大分類前 10 關鍵字與分類任務結果

主題前10關鍵字										
探討議題	0	1	2	3	4	5	6	7	8	9
美食地圖	台南	美食	店家	新聞	小吃	記者	表示	夜市	網友	米其林
餐廳情報	時間	口感	推薦	營業	好吃	餐廳	味道	口味	地址	小吃
特色小吃	小吃	台灣	現在	小吃店	知道	有沒有	好吃	夜市	老闆	食物
疫情影響	公司	台灣	政府	快篩	小吃店	中國	疫苗	新聞	媒體	高登
美食新聞	新聞	台南	台灣	完整	高雄	媒體	記者	表示	活動	連結

任務結果

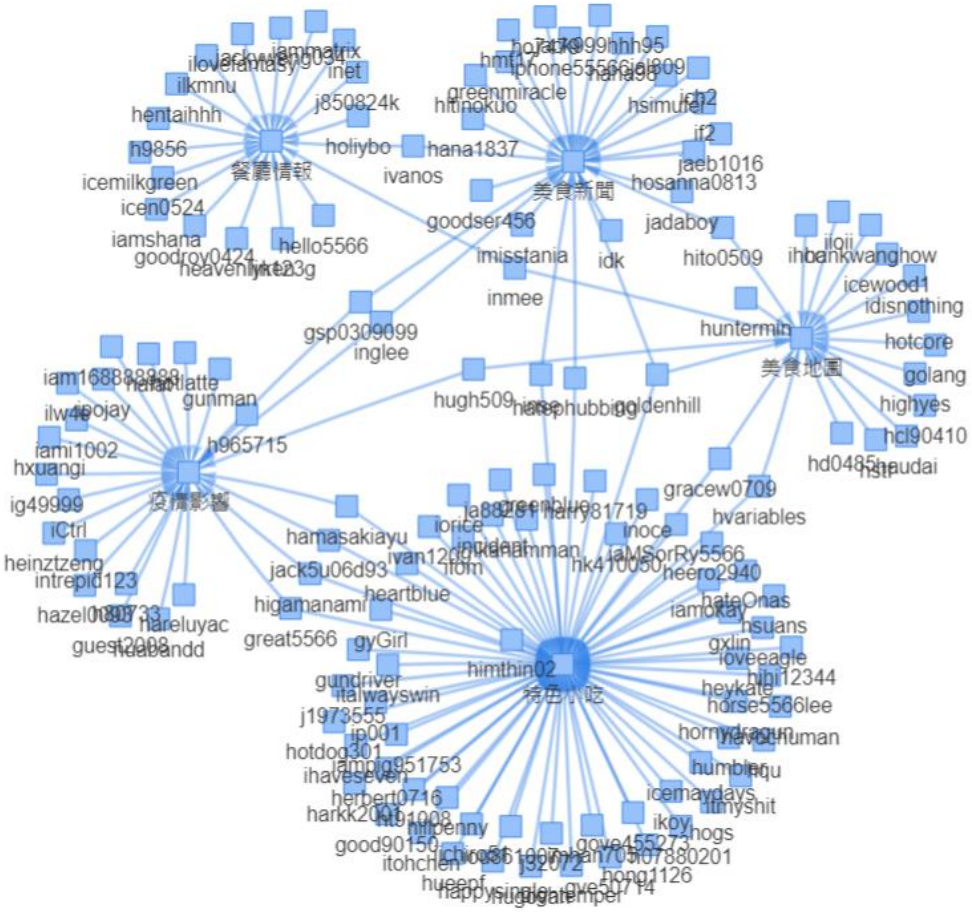
Show 10 entries Search:

system_id	美食地圖	餐廳情報	特色小吃	疫情影響	美食新聞	topic
1	0.000000	0.996837	0.000000	0.000000	0.000000	餐廳情報
2	0.071805	0.624540	0.295362	0.000000	0.000000	餐廳情報
3	0.000000	0.997286	0.000000	0.000000	0.000000	餐廳情報
4	0.000000	0.996812	0.000000	0.000000	0.000000	餐廳情報
5	0.000000	0.996404	0.000000	0.000000	0.000000	餐廳情報
6	0.000000	0.907582	0.000000	0.000000	0.091012	餐廳情報
7	0.000000	0.996651	0.000000	0.000000	0.000000	餐廳情報
8	0.000000	0.994157	0.000000	0.000000	0.000000	餐廳情報
9	0.000000	0.996886	0.000000	0.000000	0.000000	餐廳情報
10	0.000000	0.580712	0.136086	0.000000	0.282075	餐廳情報

Showing 1 to 10 of 100 entries Previous 1 2 3 4 5 ... 10 Next

第 3 章　視覺化的分析結果與解釋

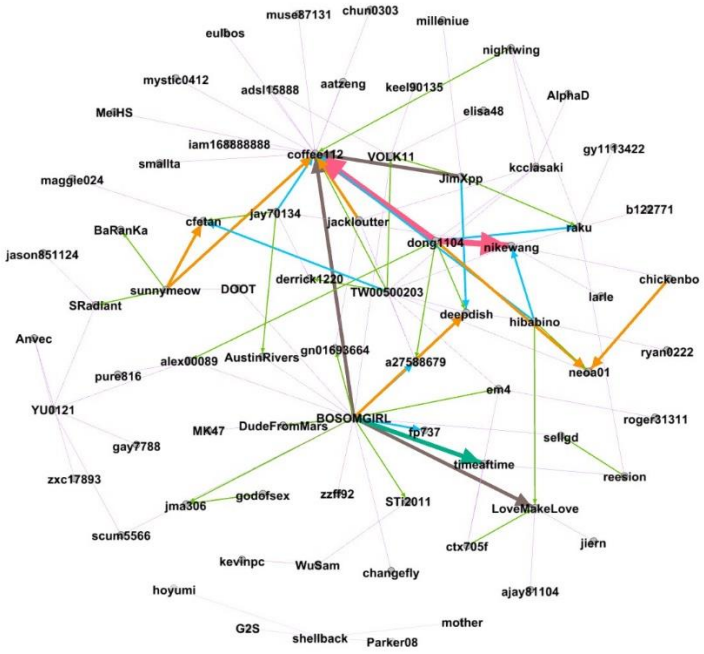
第 1 節 主題與發文者之社會網路關係圖



第 2 節 Gephi 社會網路關係圖 by 帳號

前處理：篩選前 48 大留言資料；畫圖時，排除發文者 vs 回文者,篇數小於 3 篇者。運用這些資料畫出留言者及回覆者社會網路關係圖。

透過社會網路圖，可以看出，coffee112 及 BOSOMGIRL 討論面向較廣，連結不同群體，具有 BetweennessCentrality 特質。



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 設定五個主題，透過 LDA 快速識別並掌握包括餐廳情報、美食新聞、疫情影響、特色小吃、美食地圖等五大主題分支，分類五群為混淆度最低，且各群無重疊，因此本組認為分五群為最佳數字。

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。