

# 社群媒體 第十組

## 第二次讀書會報告

組員

N124320007 藍筱琦 N124320011 林紀吟 N124320012 鄭義璋  
N124320017 郭良益 N124320019 吳邦齊 N124320020 陳軍弦  
N124320024 何牧 N124320029 黃靖紋

教授

黃三益老師

## 目錄

一、	前言.....	3
二、	分析議題說明.....	3
(一)	主題.....	3
(二)	資料收集.....	3
(三)	資料收集區間.....	3
三、	工作流程設計.....	3
(一)	使用平台.....	3
(二)	流程名稱.....	3
(三)	工作流程.....	3
(四)	流程概述.....	4
四、	流程演示.....	5
(一)	資料匯入與前處理.....	5
(二)	情緒分析.....	6
(三)	主題模型分析.....	8
	將主題數設定為 3，並透迭代 50 次方式，產出結果。.....	8
(四)	詞彙相關性與共現性關係.....	12
(五)	分類器演示.....	16
五、	結論.....	18

## 一、前言

近年來，隨著社群媒體的蓬勃發展與跨文化交流的日益頻繁，中國美食在台灣的普及度與影響力不斷攀升。從重慶麻辣火鍋、西安肉夾饅、廣西螺螄粉到上海小籠包，各式各樣來自中國不同省分的特色美食已悄然融入台灣人的日常飲食選擇中。這些美食不僅在實體店面如連鎖餐廳、夜市攤位蓬勃發展，更透過社群媒體平台的推波助瀾，成為網路熱議話題，掀起一波又一波的品嚐風潮。

過去，台灣民眾對中國食品存有安全疑慮，不少人對嘗試中國進口食品持保留態度。然而，隨著兩岸交流的深化與食品安全監管機制的完善，這種疑慮逐漸消滅。台灣商家看準市場潛力，紛紛引進或模仿中國各地特色美食，不僅豐富了台灣的飲食文化，也為市場注入新活力。本研究團隊中有成員是廣西台商，曾在疫情期間於台灣推出自製螺螄粉產品，成為台灣市場上首批正規銷售的螺螄粉之一。這段親身經歷，加上周遭友人對中國美食的持續推薦與討論，激發了我們對此議題的研究興趣。

## 二、分析議題說明

延續前次讀書會研究之台灣民眾對中國美食的態度、情感與關注重點，本次讀書會計劃建立文章分類與主題模型，分析社群媒體上的相關討論。

### (一)主題

台灣民眾對中國美食的文章分類與主題模型建立

### (二)資料收集

Instagram (關鍵字：「中華美食」、「爆紅美食」、「特色小吃」、「街邊小吃」、「川味」、「天津小吃」、「重慶小吃」、「上海小吃」等)

### (三)資料收集區間

2024 年 1 月~2024 年 12 月

## 三、工作流程設計

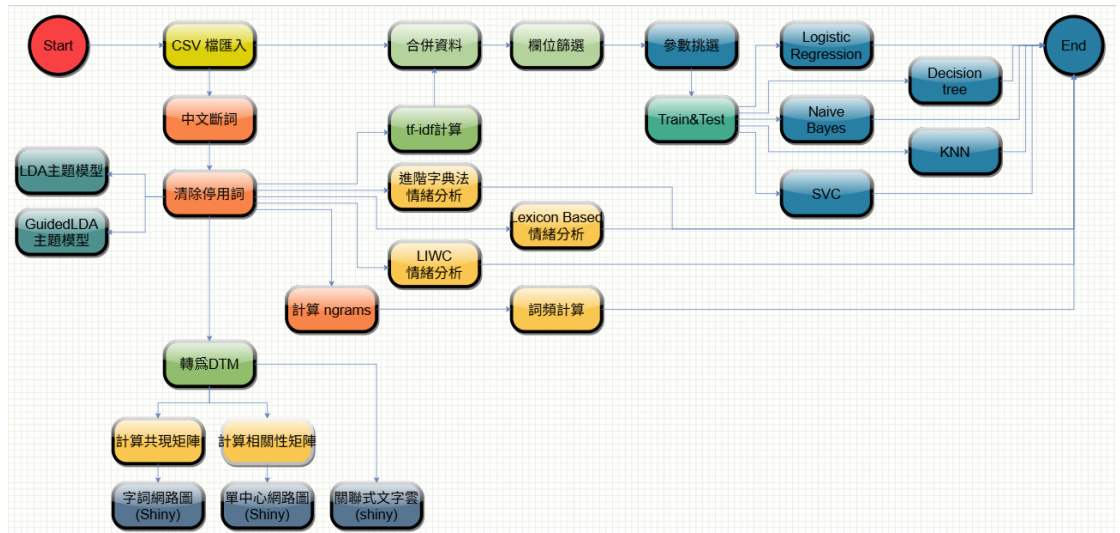
### (一)使用平台

TARflow 文字探勘工作流程設計平台

### (二)流程名稱

第十組讀書會報告\_主題模型

### (三)工作流程



#### (四) 流程概述

1. 爬取 Instagram 貼文，設定關鍵字「中華美食」、「爆紅美食」、「特色小吃」、「街邊小吃」、「川味」、「天津小吃」、「重慶小吃」、「上海小吃」。
2. 進行「中文斷詞」，將文章內容分解成字詞單位。
3. 進行「清除停用詞」，去除不必要或無意義之符號與字元。
4. 計算 n-grams、詞頻計算。
5. 進行情緒分析
  - (1) Lexicon Based 情緒分析。
  - (2) 進階字典法 情緒分析。
  - (3) LIWC 情緒分析
6. 主題模型分析
  - (1) LDA 主題模型
  - (2) Guided LDA
7. 詞彙相關性與共現性關係
  - (1) 進行「轉為 DTM」，將清理好的資料進行文本向量化。
  - (2) 繪製關聯式文字雲。
  - (3) 進行「計算共現矩陣」，分析哪些詞彙跟特定主題有關。
  - (4) 繪製字詞網路圖。
  - (5) 進行「計算相關性矩陣」，分析詞彙之間的關聯性。
  - (6) 繪製單中心網路圖。
8. 分類器演示
  - (1) 欄位調整與參數挑選。
  - (2) 進行訓練與分群。
  - (3) 各分類器的表現並比較。

#### 四、流程演示

### (一) 資料匯入與前處理

## 1. Instagram 爬蟲資料匯入

由於 Instagram 爬蟲一次只能搜索 1000 筆，將 Instagram 文章自 2024 年 1 分月份至 12 月份，分月抓取後再合併為一筆。

≡ CSV 檔匯入 (7)
X

參數設定

任務結果

## 統計資訊

**10**

欄位數

**3771**

資料筆數

## 任務結果

Show 10 entries

system_id	dataSource	artId	artDate	artContent	artPoster	artPosterUrl	commentCount	likeCount	type
1	IG	<a href="https://www.instagram.com/p/CzA5Xhm/">https://www.instagram.com/p/CzA5Xhm/</a>	2023-01-22 14:48:48	【🌸八色小籠、百福蓮藕】🍲兩天早茶都來囉！2023 1月新年如意如意，恭喜大家發財發福，👉 因為是愛吃的小籠包，滋味吸引，古典趣味的鍋湯，空間感廣，適合一家人聚餐！❤️ Promotion 已28%...	HFFoodie 🍷 Eat Hard, Play Hard 🍷	<a href="https://www.instagram.com/melody.eathard">https://www.instagram.com/melody.eathard</a>	38	3410	CAROUSEL_ALBUM
2	IG	<a href="https://www.instagram.com/reel/CZZaSHRusQ8/">https://www.instagram.com/reel/CZZaSHRusQ8/</a>	2023-01-22 09:09:29	♣️貴賓卡專屬私人料理🍴零失敗自飽到手～真寵物的店開張啦！1⃣️晚餐快速輕鬆完成超滿足下肚的兩道菜～2⃣️三杯雞好美味！【圖中標題左起順時】1v1n今日主角-油鹽香宅客- kimenfoods1v1n...	真寵1寵真真的讚 ❤️不專業幫人料理	<a href="https://www.instagram.com/_shuangshuang">https://www.instagram.com/_shuangshuang</a>	79		VIDEO
3	IG	<a href="https://www.instagram.com/reel/CZYwmsjHsUO/">https://www.instagram.com/reel/CZYwmsjHsUO/</a>	2023-01-22 12:28:25	辣到失火的四川麻辣料理 🔥🔥🔥結合傳統川菜特色的蜀鄉川味辣鍋🍲可以吃小到晚餐，川味辣魚和煮水餃🥟每一道都鮮甜！交朋接席🍻愛吃辣的人千萬別錯過這一家！1v1n... 川味辣魚500元...	肉欲 台中 台北 新竹 雲林 美食旅遊分享🍷	<a href="https://www.instagram.com/taberu.food">https://www.instagram.com/taberu.food</a>	18		VIDEO
4	IG	<a href="https://www.instagram.com/p/CY1GMHuAKKA/">https://www.instagram.com/p/CY1GMHuAKKA/</a>	2023-01-22 03:46:30	梅花梅花幾月開?今年正紀念第1月梅花盛開中! 1v1 滿風光采的大寒，卻也是梅花盛開的季節! 1v1 溫馨提醒，外出活動賞花請做好保暖喔! 1v1暖衣 #正紀念寒 萬通梅	台北市商業處 - 親睦有1人	<a href="https://www.instagram.com/sunsun_tpe">https://www.instagram.com/sunsun_tpe</a>	0		CAROUSEL_ALBUM

## 2. 中文斷詞

因搜尋目標為中國美食，尋找 30 個中國常見但臺灣較少見之中國美食加入權重 500

中文斷詞 (8)

參數設定

Input - 7

任務結果

選擇處理檔位 \*

artContent

定義詞彙

北京牌機 500  
陸軍島 500  
中國 500  
水寨牛肉 500  
糖醋肉 500

選取字彙

.....請選擇.....

儲存設定

### 3. 清除停用詞

- (1) Instagram 的文章常出現無意義之表情符號 (Emoji)。
- (2) 因主題為台灣民眾對中國美食的看法，對於同樣使用繁體字的香港用戶，新增香港口語常使用（但台灣不使用）的繁體中文字，例如：嘅、係、着、歎。

清除停用詞

(12)

參數設定

Input - 0

任務結果

語言 \*

Chinese

是否清除單字元

是

清除英文字母 \*

是

清除換行符號 \*

是

清除html tag \*

是

使用預設停止詞

是

是否轉為小寫英文

是

清除數字 \*

是

清除特殊標點符號 \*

是

自定義停止詞

+  
+  
+  
+

儲存設定

#### 4. 計算 n-grams、詞頻計算

## 統計資訊

ngram數量



## 任務結果

Show 10 entries

Search:

system\_id

result

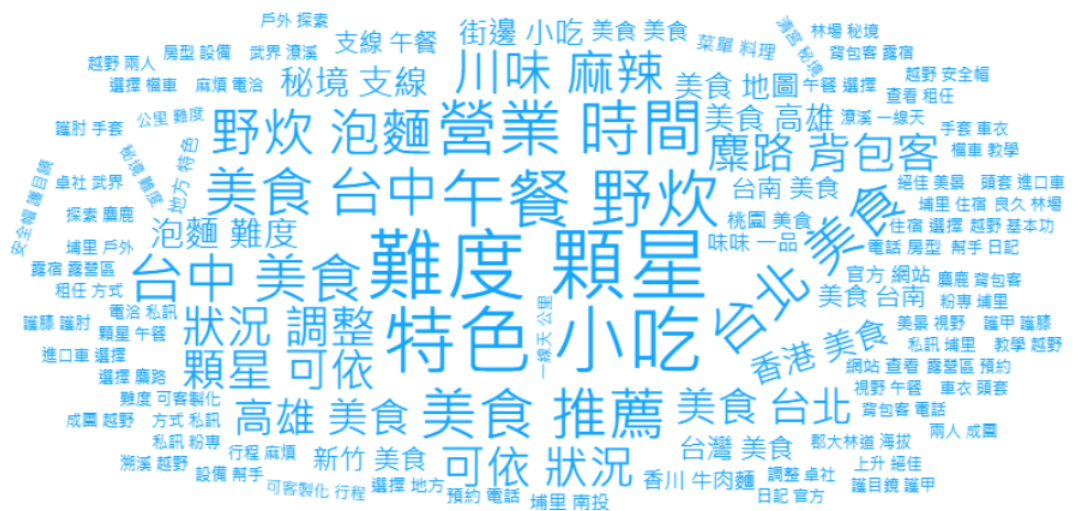
1

(八)色小龍蝦、百寶豆腐、鮑魚翅菜、天鰲燒雞、皇朝素齋、富貴合桃、世世如意、故土南珍、指南南燕、喜慶千禧、小籠包、蟹粉湯包、揚州炒飯、招牌菜三鮮、招牌菜臭豆腐、蔥油餅小籠包、小籠包水餃、麵點吸引：吸引古味、古典韻味，滋味爽滑、清透空間感、開胃、解渴、滋潤、適合一家人、大眾聚餐。餐前酒送一瓶——一送一換優惠，優待一小籠包、小籠包打打卡、打一打趣、一龍一鳳、原盅小籠包、小籠包全線、全線分店八折、八色小籠包、小籠包小籠包、小籠包包裹、包裹熱辣辣、熱辣辣湯汁；湯汁美味、味道醇正、甜酥黃松、甜酥黃松、甜酥黃松、蒜香芝士、芝麻綠瓜、綠瓜鹹肉、鹹肉小籠包小籠包皮薄不膩、爽口清香、湯底清爽、咬下去慢慢嚼、慢慢嚼、慢慢嚼、湯汁濃郁、滋味醇厚、雞肉飽滿、雞飽滿、雞飽滿、香油香、香油香、香油香、香味吸引：吸引新鮮、黑椒豬扒、蟹粉小籠包、小籠包子、牛肉鮮香、鮮香十足、十足飽滿、蟹肉蟹膏、蟹膏黏貼提提、提提鮮甜、招牌鴨腿、招牌出爐、出爐鴨腿、打開蓋子、蓋子新鮮、新鮮蒸起熟、蒸起熟燙燙、燙燙暖暖、暖暖夾死、夾死小籠包、小籠包湯汁、湯汁底部、底部滾動、滾動肉滿、肉滿湯滿、湧而出而、出而鮮香、鮮香美味、美味太上、太上

蘇絲、蘇美、美味高心、高心白淨。

2

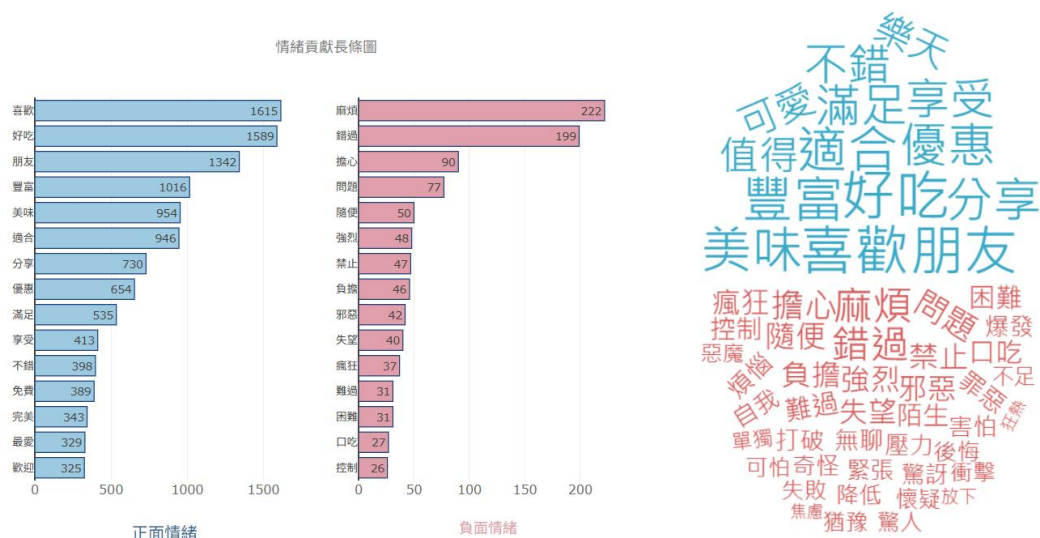
[**菜名**] 專車、專人、個人料理。料理失敗，收工白飯；白飯殺手，殺手夾餐。**賣點**：熱炒店、熟炒店開張、開張簡單、簡單快速、快捷輕鬆、輕鬆完成、完成下飯兩道、南撈三杯雞、三杯雙蔬、魚香茄片、茄汁豆腐等。今日主理：**生猛龍蝦**、金銀蛋卷、毛宅牛共有六種口味。**口味**：鹹甜、鹹苦、鹹酸、鹹辣、鹹香、鹹臭。三杯飯、飯蒸茄子、茄汁大鴨、大廚什錦湯、胡椒牛肉、川味魚香、魚香豆腐、私房鮑魚、私家秘製麻辣、麻辣宮保、宮保肉丁、煎餅新弄、新弄皮包、做出十一手一好易做、好買方便、便宜了大家，皆宜家常、緊密調料、調味提增本家、本來者怕、害怕調味、調味包會、包包太鹹、太鹹配茶、太鹹試吃、試吃過癮、過癮調味、調味套餐、醬色三口、口味三杯雞、三杯豆豉、雙菇材料、材料厚薄、腐竹蒜苗、蒜苗蒜苔、蒜苔荔枝、荔枝辣椒、辣椒雞蛋白、蛋九龍、九龍薯粉、鹽焗全雞、鹹醋三春、三春茄汁、茄汁豆腐、豆腐材料、材料蝦仁、蝦仁豆腐絲、妙炒芋蔥、芋蔥糯米、糯米芝麻、蒜末薑金、金線釀蟹、蟹茄汁好、好事、好康報喜、報喜即回、即日全順、全順好物、物齊滿堂、滿堂全順、全順宅宅在任、任搭三日廿四粥、粥廣濟、濟慈益金、益金密財...]



## 1. Lexicon Based 情緒分析

使用 Lexicon Based 情緒分析發現負面情緒最多的是「錯過」一詞。

## 視覺化



然美食文章中大多是「不能錯過」這類之正面情緒，因此改採可以移除情緒詞之進階字典法。

### 2. 進階字典法 情緒分析

使用 NTUSD 字典，根據結果新增正面詞彙「愛吃」、「上癮」，並將「錯過」、「限定」、「追蹤」、「老油條」、「酸味」…等詞加入移除情緒詞。

#### 進階字典法 情緒分析 (20)

參數設定

Input - 12

任務結果

選取字典 ❶

NTUSD

定義正面詞彙 ❶

愛吃  
上癮

是否使用否定詞 ❶

是

移除情緒詞 ❶

錯過  
限定  
追蹤  
老油條  
酸味

定義負面詞彙 ❶

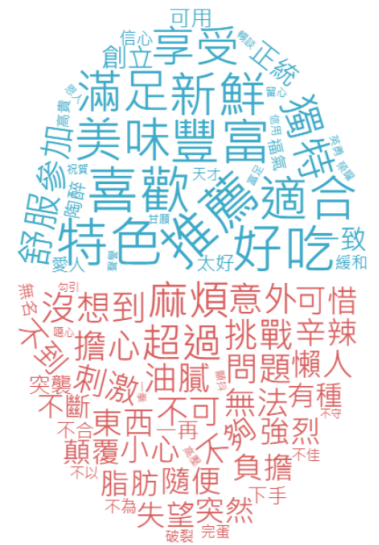
以換行符號區隔，e.g.  
難過  
生氣  
悲傷...

是否使用加強詞 ❶

是

儲存更改

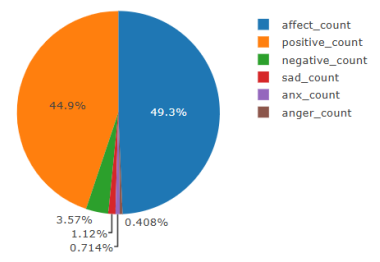
正面情緒	負面情緒
推薦 1259	麻煩 222
特色 1232	超過 197
喜歡 1008	意外 99
好吃 1002	沒想到 94
豐富 752	挑戰 91
適合 690	擔心 82
美味 647	油膩 75
新鮮 632	問題 70
滿足 463	刺激 70
享受 336	不可 65
過癮 313	無法 59
免費 308	可惜 57
自助 304	辛辣 54
歡迎 302	不夠 52
獨特 301	東西 50



使用 LIWC 情緒分析計算各類情緒詞在文本中所佔比例。

使用 LIWC 情緒分析計算各類情緒詞在文本中所佔比例。

正面情緒	負面情緒
喜歡	麻煩
好吃	錯過
朋友	擔心
豐富	問題
適合	強烈
美味	負擔
分享	隨便
滿足	失望
優惠	邪惡
享受	困難
不錯	禁止
免費	難過
歡迎	瘋狂
最愛	口吃
完美	陌生



## 1. LDA 主題模型

將主題數設定為 3，並透迭代 50 次方式，產出結果。

將主題數設定為 3，並透迭代 50 次方式，產出結果。



目標欄位 \*

result

主題數 \*

3

詞彙頻率下限 ⓘ

40

alpha

預設為主題數/50

chucksize ⓘ

預設為2000

是否輸出字典

是

迭代次數

50

主題保留關鍵字數量

20

詞彙頻率上限 ⓘ

0.5

Beta

預設為0.1

update\_every ⓘ

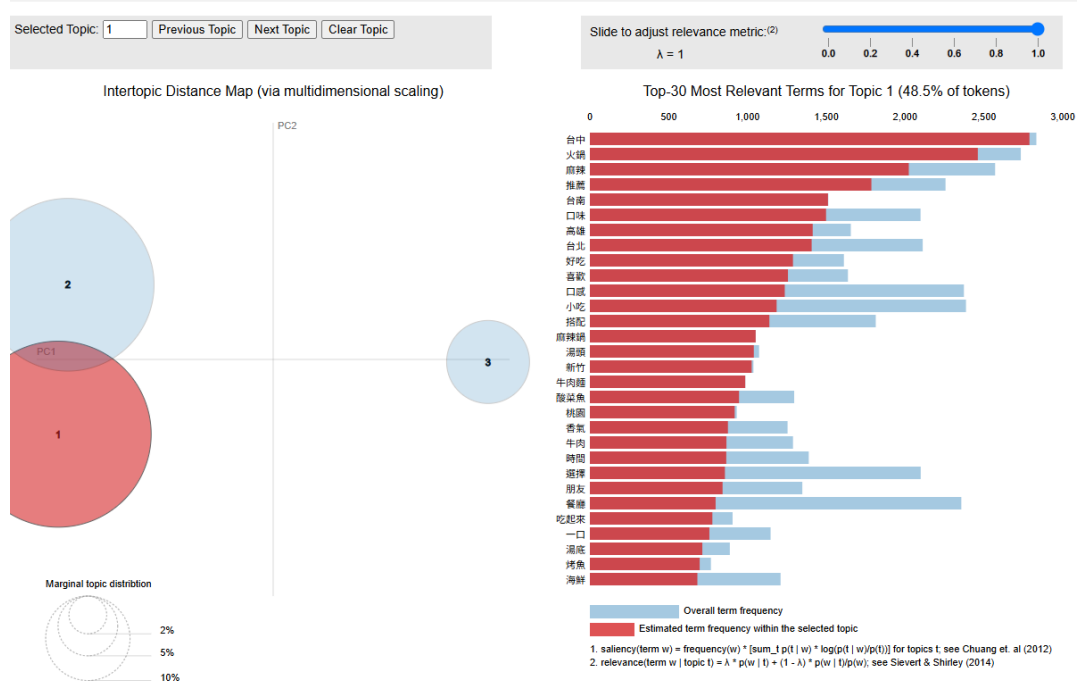
1

儲存更改

產出三大主題：

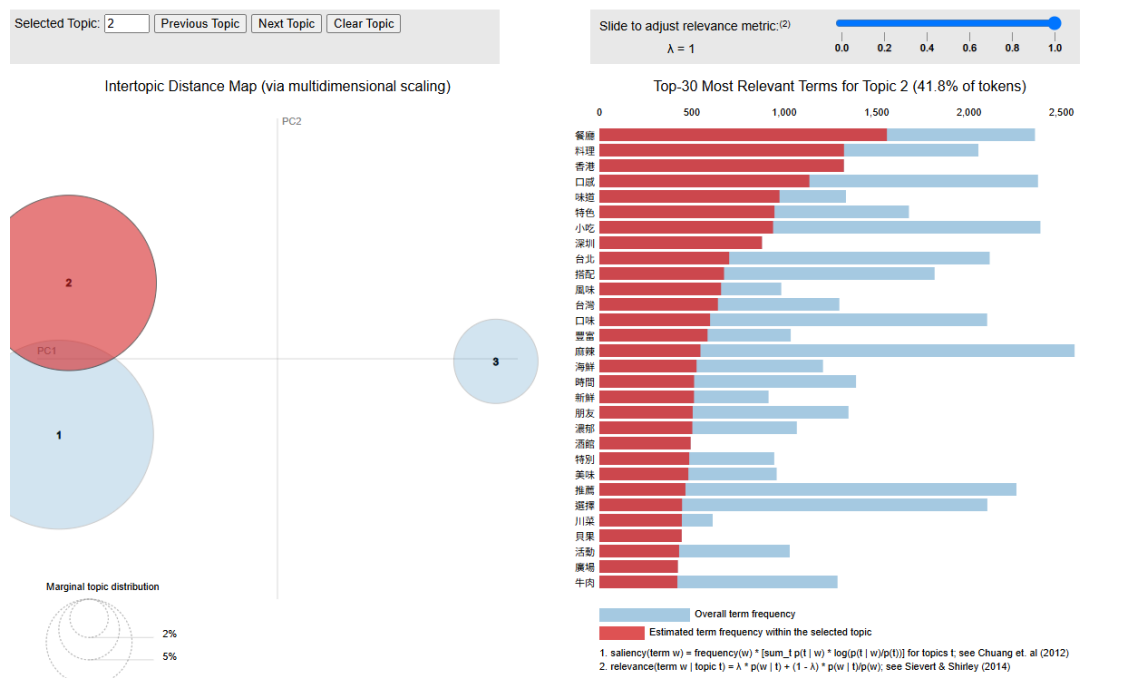
- (1) 第一類與小吃類別有關。關鍵字有「火鍋」、「麻辣鍋」、「牛肉麵」、「酸菜魚」等。

## LDA Vis



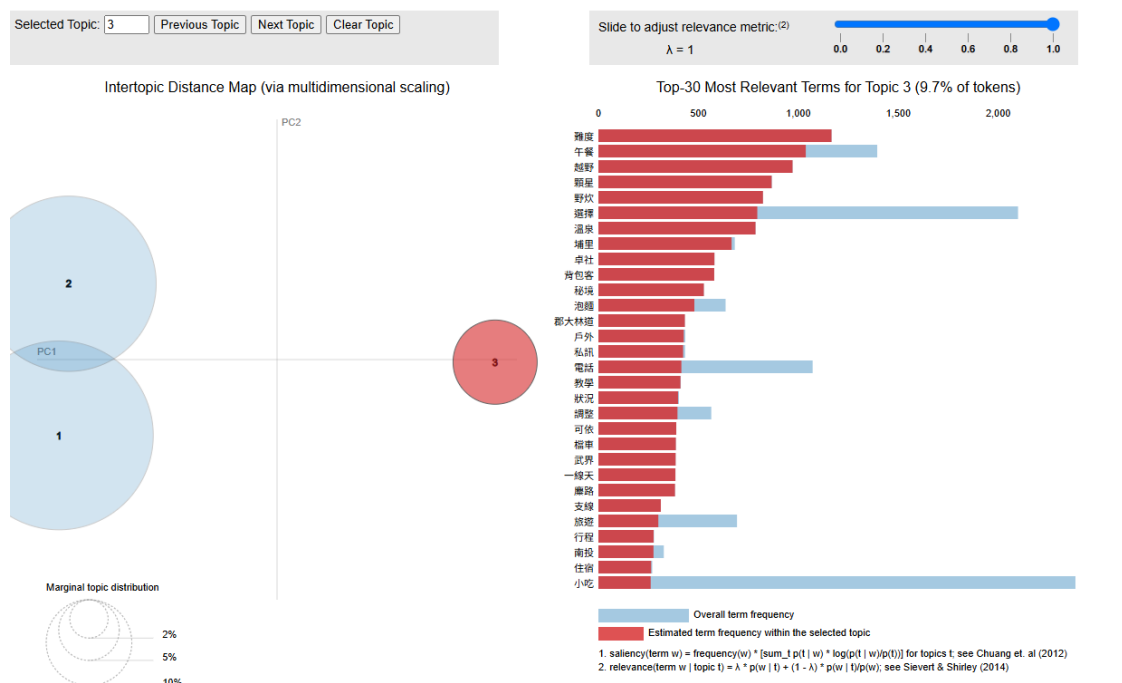
- (2) 第二類與描述餐點的口味或口感有關。關鍵字有「料理」、「口感」、「味道」、「特色」等。

## LDA Vis



(3) 第三群為戶外活動類。主要關鍵字有「越野」、「野炊」、「溫泉」、「背包客」等。

## LDA Vis



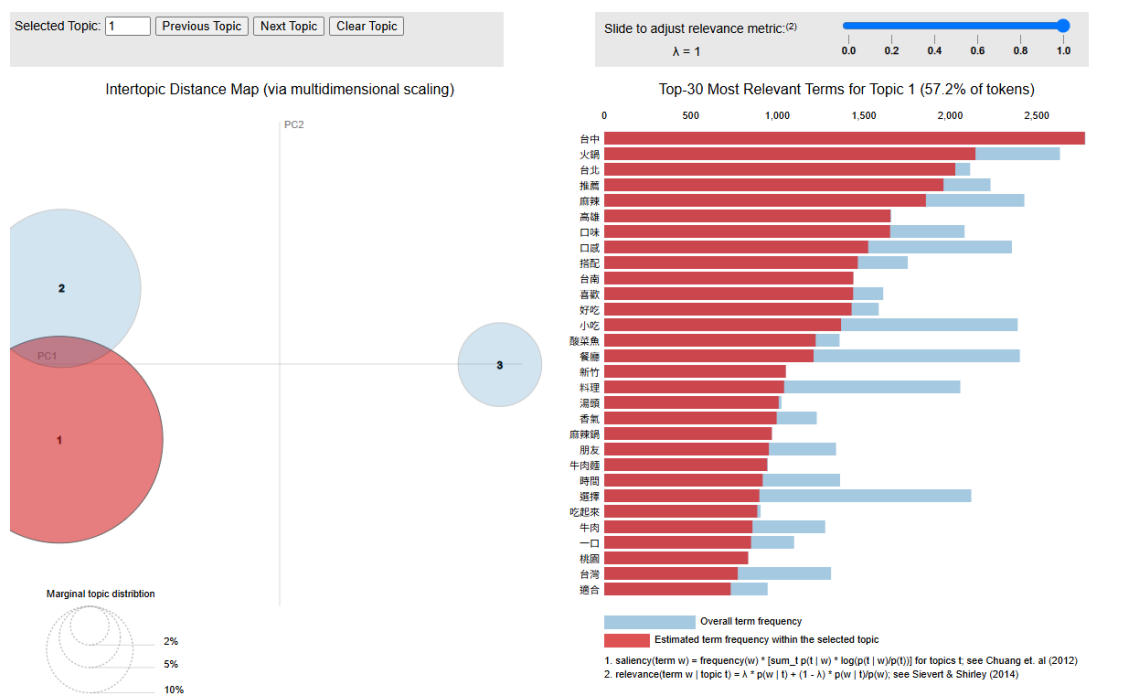
## 2. GuidedLDA 主題模型

設定主題數為 3，給主題種子字「火鍋、麻辣鍋、牛肉麵、酸菜魚」、「料理、口感、味道、特色」、「越野、野炊、溫泉、背包客」，並透

迭代 50 次方式，產出結果。

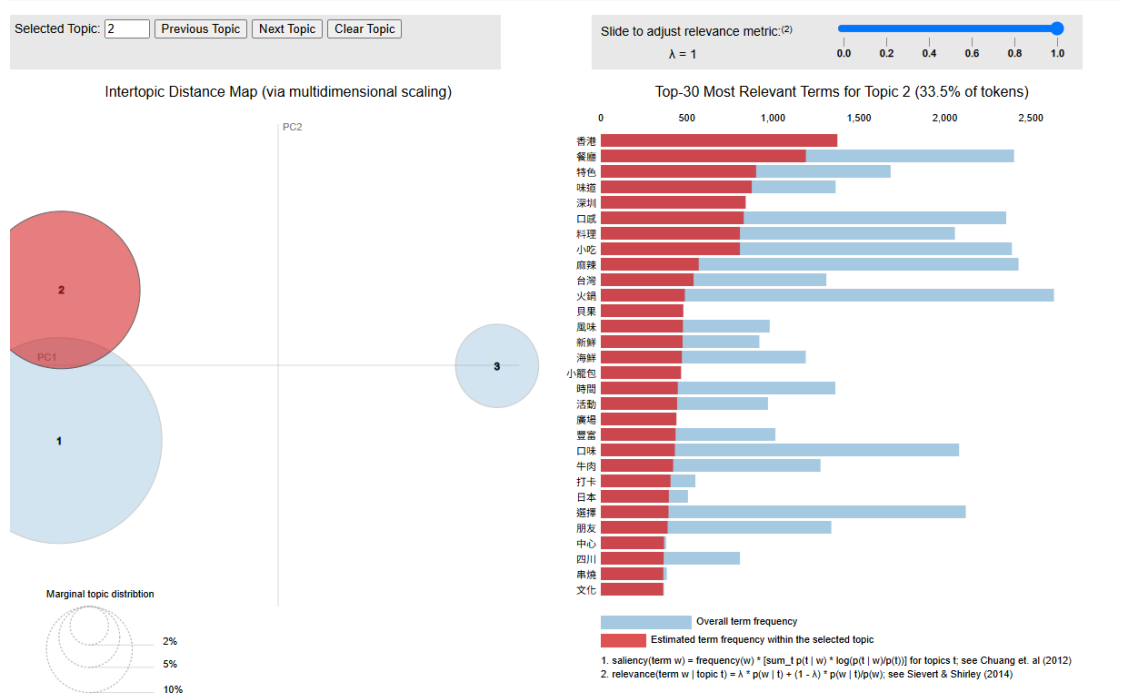
(1) 第一類「火鍋」、「麻辣」、「酸菜魚」等，與中國小吃類別相關。

## GuidedLDA Vis

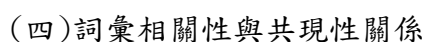


(2) 第二類關鍵字有「特色」、「味道」、「口感」等，與評論餐點的口味或口感有關。

## GuidedLDA Vis



## GuidedLDA Vis



篩選 200 個詞彙，將各詞彙轉成矩陣型態，以觀察詞彙之間的關係。

參數設定

Input - 12

任務結果

統計資訊

201  
字數

3771  
文章數

任務結果

Show 10 entries

Search

system_id	一口	一 塊 元	不 錯	中 心	車 打	乳 酪	享 受	人 氣	便 宜	傳 統	香 滑	充 滿	免 費	入 口	入 味	分 享	加 入	十 足	午 茶	南 洋	原 味	口 味	口 感	外 賣	可 依	台 中	台 市	台 光	台 市	台 灣	吃 到	吃 到	吃 起 來	味 道	咖 啡	品 牌	喜 歡	
1	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	2.0	1.0	0.0	0.0	1.0	0.0	0.0	2.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	2.0	0.0	0.0	0.0	1.0	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
5	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	2.0	0.0	2.0	0.0	0.0	0.0	6.0	1.0	0.0	0.0	2.0	1.0	0.0	1.0	2.0	0.0	0.0	3.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	
9	1.0	0.0	0.0	0.0	0.0	7.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	3.0	0.0	2.0	1.0	0.0	0.0	

全場熱度

點贊下載次數C/CV資料

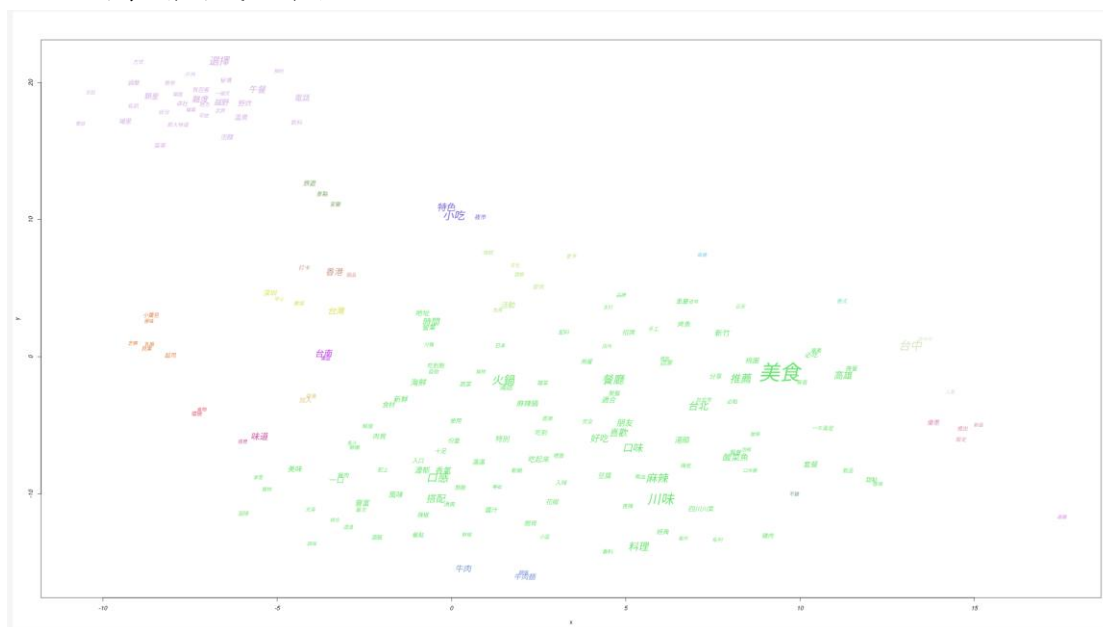
點贊下載次數like

點贊下載次數love資料

- (1) 對於食材與菜品類的詞彙，「串燒」一詞在多個文件中有較高頻率，表明燒烤類中國美食在台灣受到廣泛關注和討論。
- (2) 用餐體驗類詞彙中，「享受」、「人氣」等詞的出現，也說明了消費者對中國美食的正面情感。

- (3) 正面評價詞彙中，「多汁」、「好吃」、「新鮮」等正面評價詞在多個文本中出現，表明台灣民眾對許多中國美食持正面態度。
- (4) 「傳統」、「文化」、「正宗」等詞彙的出現頻率，反映了對中國美食文化背景的關注。

## 2. 繪製關聯式文字雲。



圖表說明

### (1) 整體結構與分佈特徵

關聯式文字雲圖以二維空間的方式呈現詞彙之間的關係，不同顏色的詞彙群組代表不同的語義聚類或主題：

- 綠色詞彙群組：位於圖表中央區域，形成較大且密集的詞彙網絡，可能代表討論中的核心主題和高頻詞彙。從圖中可見「美食」、「菜式」等詞彙屬於這一群組。
- 紫色詞彙群組：分佈在圖表左上方，形成相對獨立的詞彙集合，代表特定的菜系特色或地域性話題。
- 紅色詞彙群組：位於圖表左側，詞彙較為分散，代表較為獨特、小眾的討論點與菜式。
- 藍色詞彙群組：出現在圖表底部，詞彙數量較少但相互關聯性強，代表特定的味道描述或食材類型。

### (2) 詞彙關聯性分析

詞彙在空間中的距離反映了它們在文本中的關聯程度：

- 近距離詞彙：圖中可見的一些詞彙緊密聚集，表明這些詞彙在討論中經常一起出現，具有較強的語義關聯。例如，綠色群組中的詞彙共同描述常見的中國菜式（火鍋、湯底、麻辣鍋）或用餐體驗（口感、香氣、濃郁）。
- 遠距離詞彙：不同顏色群組之間的詞彙相距較遠，表明它

們代表不同的討論主題或面向，在文章中較少一起出現。

根據詞彙的分群和位置，可以識別出幾個主要的討論主題：

b. 紫色詞彙群組：代表美食的區域或地點（如「南投」、「埔里」、「卓社」等）的描述。

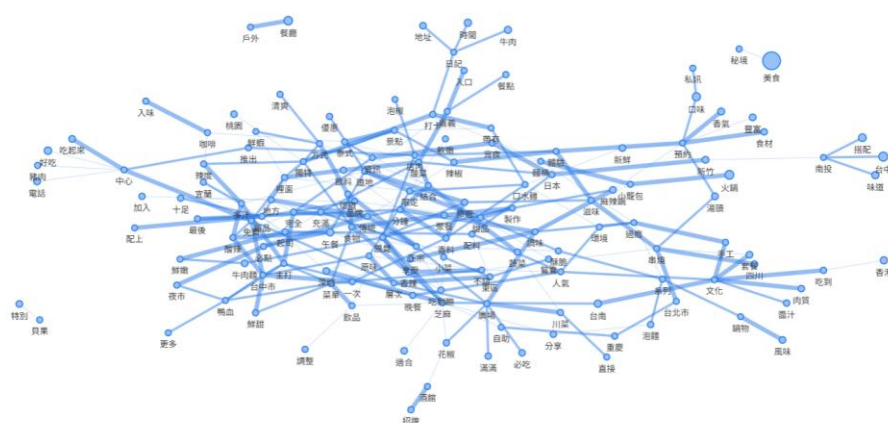
d. 藍色詞彙群組：與食材品質或特殊烹飪技法相關。

詞彙在圖中的大小反映了它們在語料庫中的重要性：

b. 小尺寸詞彙：散佈在各群組中的小詞彙，雖然出現頻率較低，但可能代表特定的細節描述或獨特評價角度。

熱圖

作為核心評價詞，「美食」與多種中國菜品詞彙有較強共現關係，表示它是描述中國美食的常用詞；而「夜市」則與特定食物的關聯，反映台灣民眾常在夜市環境中品嚐這些中國美食；從熱圖可見，「特色」與某些 X 軸詞彙形成較深色塊，表明大多時候，這些被視為具特色的中國菜品。

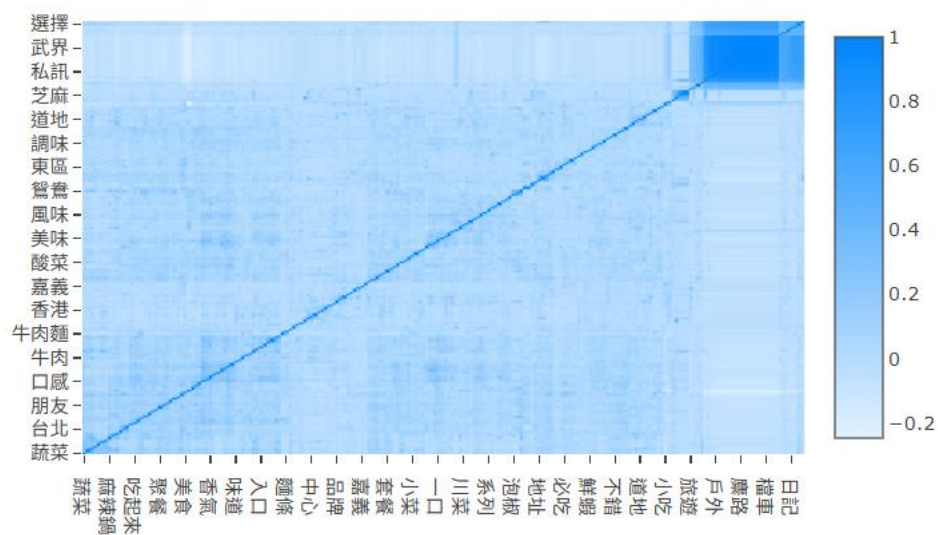


位於網路圖中心區域的詞彙（如「吃到飽」）與多個其他詞彙有連結，表示它們是討論中的核心概念；位於圖外圍的詞彙與中心詞彙的關聯相對較弱，或只與特定詞彙相關聯。

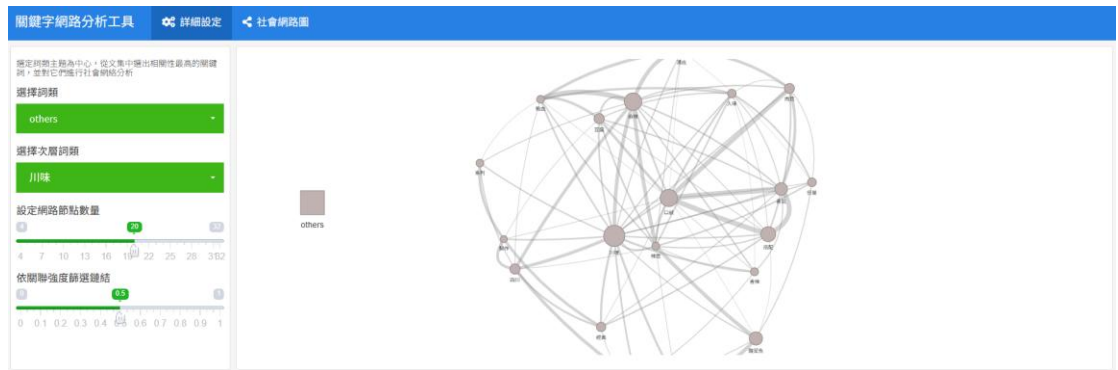
而圖中詞彙大致可分為「地理位置詞」、「菜系」、「評價」、「餐飲服務」等幾個主要類別。

5. 進行「計算相關性矩陣」，分析詞彙之間的關聯性。

熱圖



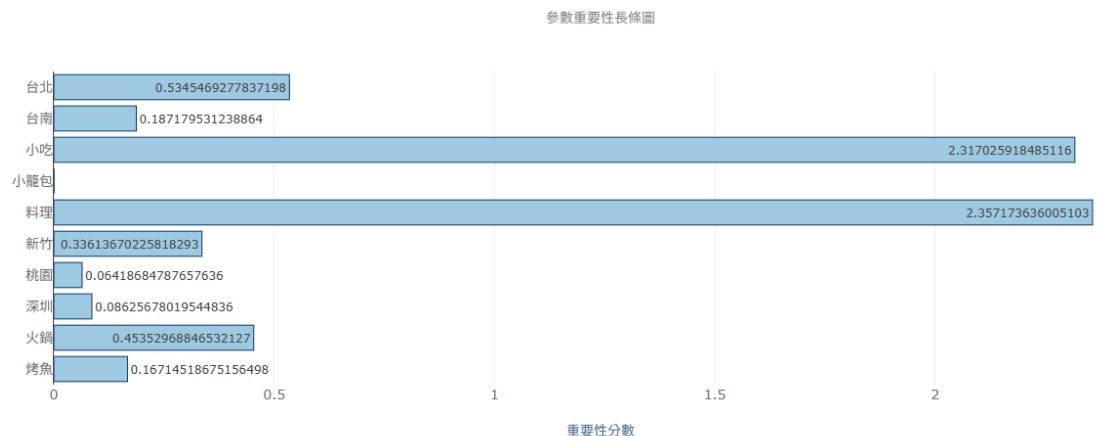
6. 繪製單中心網路圖。



## (五)分類器演示

### 1. 欄位調整

- (1) 清除停用詞後，計算 tf-idf 且最多篩選詞彙數量選擇預設的 200 後，與原始 csv 進行整合。
- (2) 篩選與訓練器有關的欄位後，透過參數挑選（參數量設定 17）評估哪些參數的重要性指標。



- (3) 進行 Train&Test 取 20%訓練資料、80%測試資料並隨機排序，亂數種子取 777。

- (4) 本次將五種分類器都跑過一次，一一描述各分類器表現：

#### a. Logistic Regression

Logistic Regression (129)

參數設定		Input - 127		任務結果	
統計資訊					
67.642		0.487		0.049	
訓練時間		推論時間		測試資料準確度	
0.016		0.037		0.049	
測試資料macro-F1		測試資料加權F1		測試資料micro精確率	
0.033		0.049		0.017	
測試資料加權精確率		測試資料micro召回率		測試資料macro召回率	
				0.015	
				0.049	
					測試資料macro精確率
					測試資料加權召回率

#### b. Decision tree



### Decision tree (130)

參數設定		Input - 127		任務結果	
統計資訊					
1.245 訓練時間	0.546 推論時間	0.078 測試資料準確度	0.078 測試資料micro-F1		
0.022 測試資料macro-F1	0.069 測試資料加權F1	0.078 測試資料micro精確率	0.022 測試資料macro精確率		
0.065 測試資料加權精確率	0.078 測試資料micro召回率	0.024 測試資料macro召回率	0.078 測試資料加權召回率		
112 樹深度	3251 葉節點數				

### c. Naive Bayes

### Naive Bayes (131)

參數設定

Input - 127

任務結果

統計資訊

0.975

訓練時間

1.326

推論時間

0.078

測試資料準確度

0.078

測試資料micro-F1

0.022

測試資料macro-F1

0.069

測試資料加權F1

0.078

測試資料micro精確率

0.021

測試資料macro精確率

0.064

測試資料加權精確率

0.078

測試資料micro召回率

0.024

測試資料macro召回率

0.078











測試資料加權召回率

Gaussian

分佈

### d. KNN

### KNN (132)

參數設定		Input - 127		任務結果	
統計資訊					
0.014 訓練時間		0.618 推論時間		0.036 測試資料準確度	
0.005 測試資料macro-F1		0.03 測試資料加權F1		0.036 測試資料micro精確率	
0.027 測試資料加權精確率		0.036 測試資料micro召回率		0.007 測試資料macro召回率	
5 鄰居數		euclidean 距離公式			

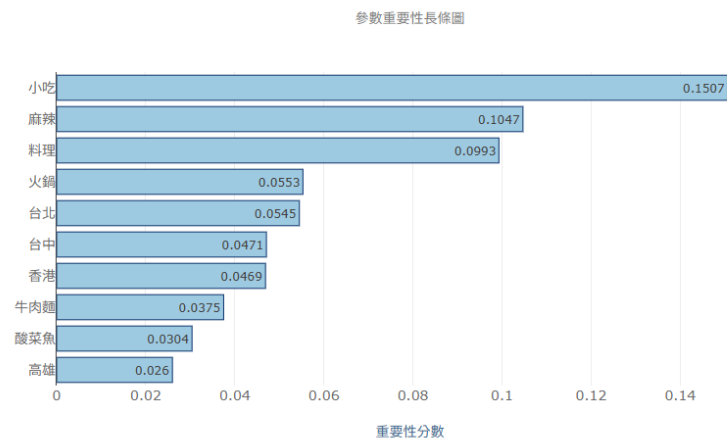
### e. SVC

### SVC (133)

參數設定		Input - 127		任務結果	
統計資訊					
11.136 訓練時間	71.795 推論時間	0.056 測試資料準確度	0.056 測試資料micro-F1		
0.017 測試資料macro-F1	0.041 測試資料加權F1	0.056 測試資料micro精確率	0.017 測試資料macro精確率		
0.036 測試資料加權精確率	0.056 測試資料micro召回率	0.019 測試資料macro召回率	0.056 測試資料加權召回率		
50 懲罰係數	rbf 核函數	3 維度			

(5) 經過五種分類器的結果，我們發現 Decision tree 跑出來的結果相較其他分類器佳。

其視覺化長條圖形與統計資訊如下



## 五、結論

我們發現「麻辣」、「火鍋」等高權重詞彙在文字雲圖中也占據重要位置，證實了這些詞彙在中國美食文化中的核心地位。城市分佈的重要性分數與文字雲圖中的地域詞彙群組亦相呼應，表現出中國美食文化的區域多樣性與地方特色。