

社群媒體分析 PROJECT 2

指導老師：黃三益教授

組長：N114320028 李晉安

組員：N114020004 蔡志強

N114320004 邱昱榕

N114320009 林育伶

Youtube網址

<https://youtu.be/7BZxX2UDLpA>



目錄

1. 動機和分析目的
2. 資料集的描述
3. 資料的分析過程
4. 視覺化的分析結果與解釋
5. 結論



動機與分析目的

動機

- 台灣人是個愛吃美食的民族，對於美食的追求可謂是孜孜不倦。在選擇美食時，台灣人會考量到多種因素，台灣人的美食選擇也會受到個人喜好、文化背景、地域特色等因素的影響。例如，有些人偏好中式料理，有些人則偏好西式料理；有些人喜歡吃辣，有些人則不喜歡吃辣；北部人的口味偏清淡，南部人的口味偏重鹹等。



分析目的

- 統計美食文章中出現的口味、食材、價格、環境、服務等相關詞彙的頻率，從而了解台灣人對這些因素的重視程度。此外，還可以對美食文章的評論進行分析，了解台灣人對不同美食的評價和偏好。



資料集的描述

PTT 爬蟲 (4)

參數設定

任務結果

選擇看板 *

dog(狗)

EAseries(歐美影集)

eat_disorder(飲食疾患)

Exotic_Pet(特殊寵物)

e_shopping(網購)

Finance(金融業)

Food(美食)

搜尋關鍵字 0

以換行區隔，e.g.
國立中山大學
西子灣
--

排除關鍵字 0

以換行區隔，e.g.
壽山動物園
猴子
--

搜尋起始日期

2023/06/01

搜尋結束日期

2024/06/01

儲存更改

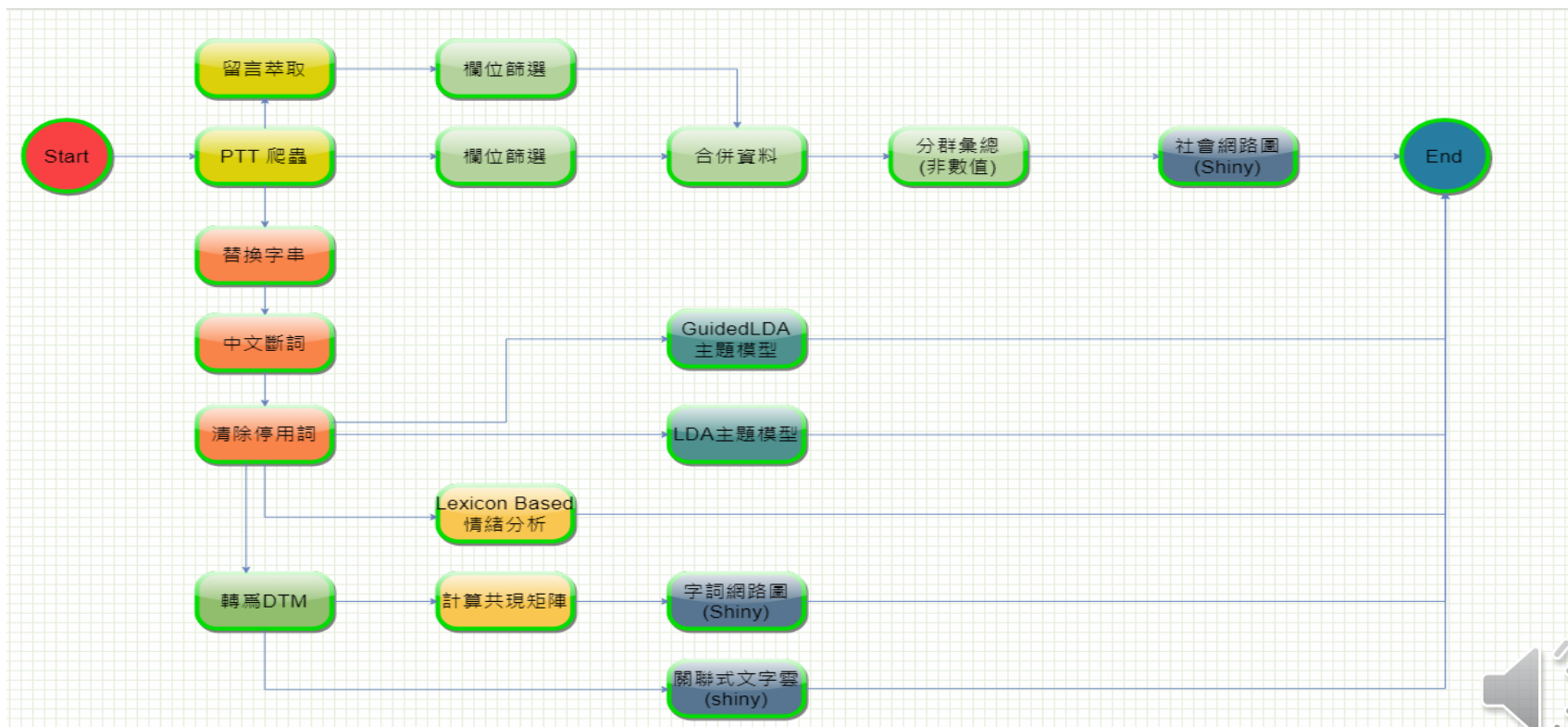
資料來源：PTT爬蟲的美食板

資料區間：2023/06/01-2024/06/01

運用這些資料與留言萃取繪製社會網路圖、情緒分析、LDA、字詞網路圖、關聯式文字雲。

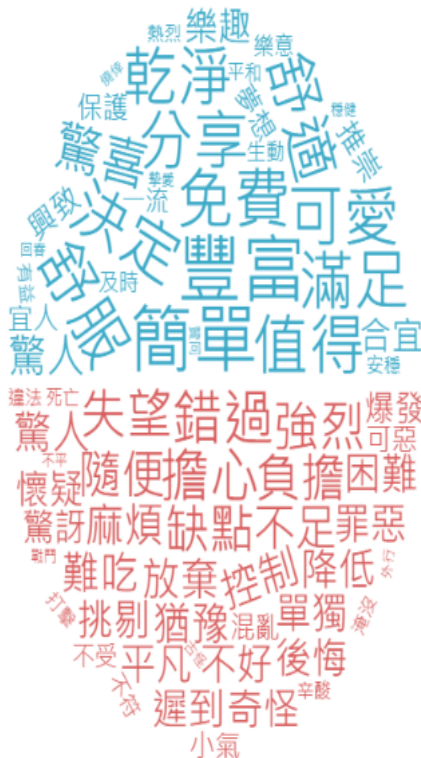
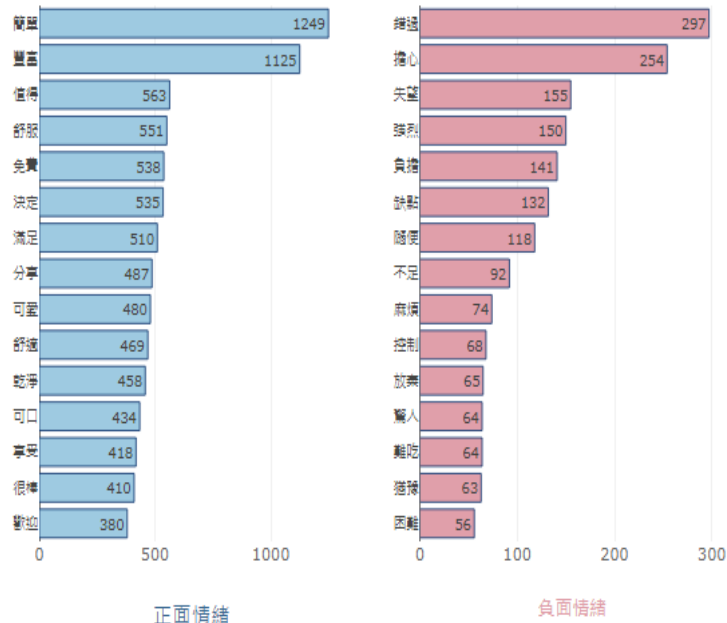


資料的分析過程-PTT美食板



視覺化的分析結果與解釋-情緒分析

情緒貢獻長條圖



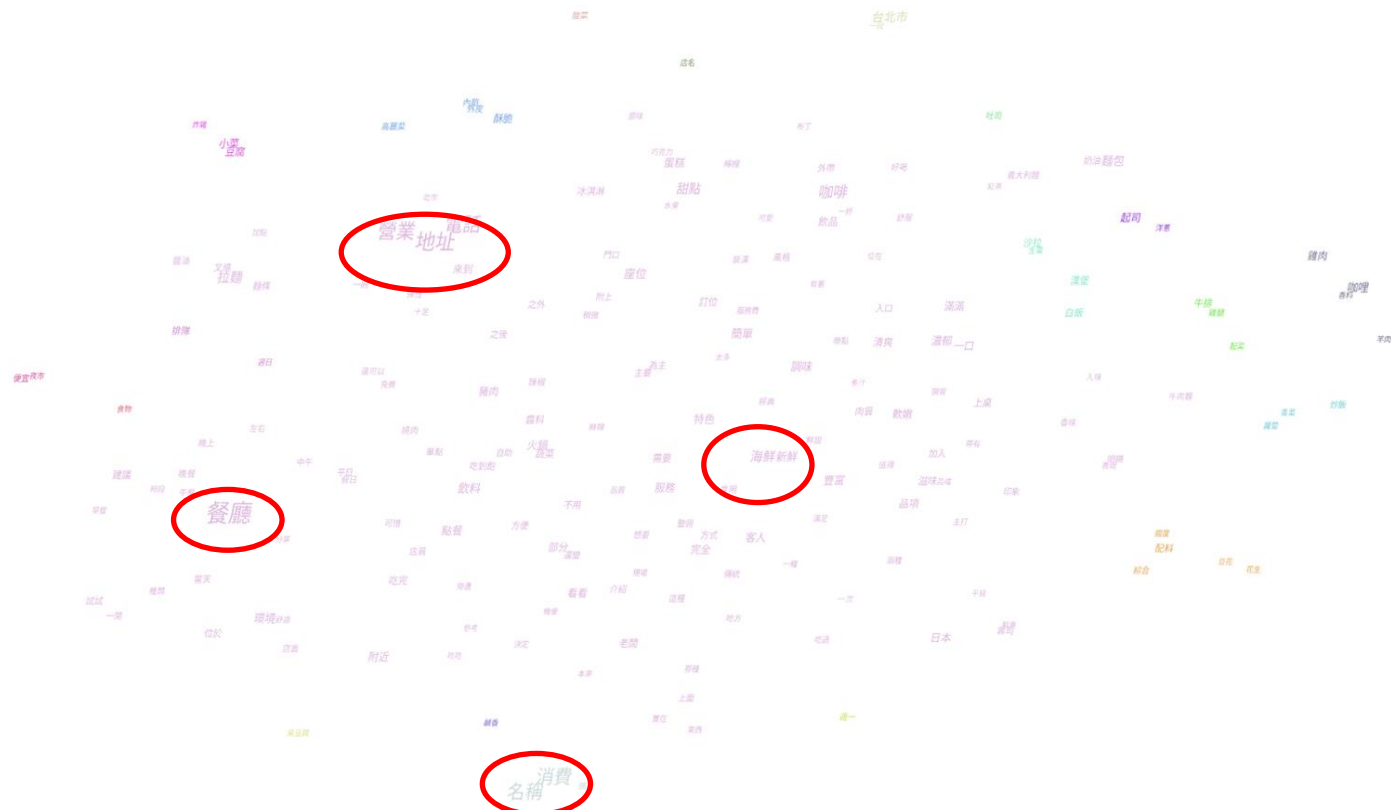
正面字詞：主要以簡單、豐富大幅領先其他字詞，還有顧客在意的舒適、乾淨、可口等相關字。

負面字詞：因為某些餐廳較為隱密饕客們害怕錯過，因此第一名為，其他還有像是擔心、失望、負擔、缺點、隨便等。

由於是美食板民以食為天，所以正面情緒字詞多於負面情緒字詞。

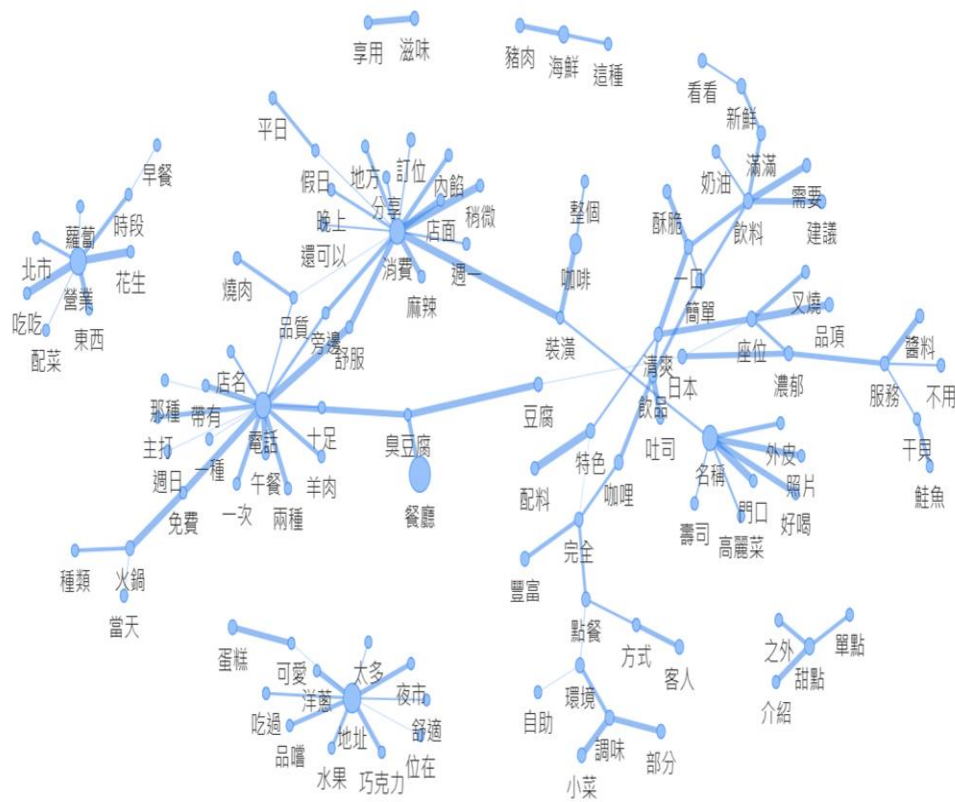


視覺化的分析結果與解釋-關聯式文字雲



透過關聯式文字雲可以發現，大家的討論還是聚焦在幾個區塊，包括餐廳的地址、營業相關、海鮮是否新鮮、消費的名稱等。





透過字詞網路圖分析台灣人對美食的偏好，發現討論的主要焦點集中在以下幾個區塊：**餐廳的地址、營業相關資訊、豬肉和海鮮及消費分享**。這些結果顯示消費者在選擇和評價餐廳時，特別重視餐廳位置、服務品質、食品新鮮度和整體消費體驗，為餐廳經營者提供了有價值的市場洞察。



(2) 這些詞語之間的**關聯複雜**，網路圖容易變得過於雜亂。

(3) **干擾字詞多**(例如：無關的聊天內容、感嘆詞等)，也讓圖變的雜亂



視覺化的分析結果 與解釋-社會網路圖

節點和連結定義：

節點 (Nodes)：每個節點代表一個使用者，包括發文者和回文者。

連結 (Edges)：每條連結表示發文者與回文者之間的互動，比如回覆、評論或引用。

從社會網路圖觀察發文者與回文著的關聯，可以看出經常回覆者包含幾個AlphaD、larle、KUNI0202、lulalachan、ggyyworld，這些經常回復他人的人，通常也會彼此回覆。透過以上分析方法，可以清晰地觀察發文者與回文者之間的關聯，並理解社會網路圖中不同使用者的角色和影響力。

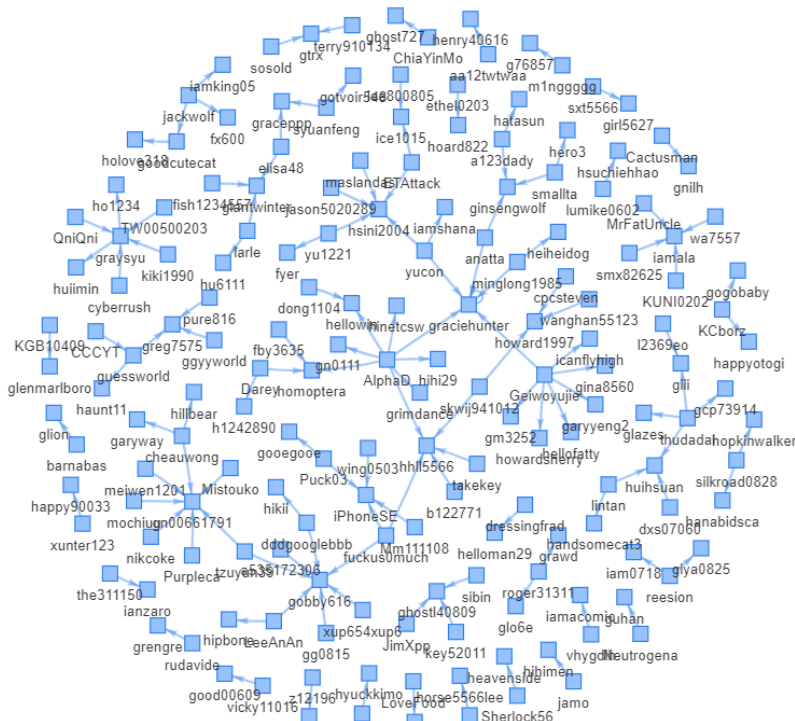


圖1、社會網路圖

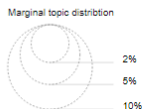
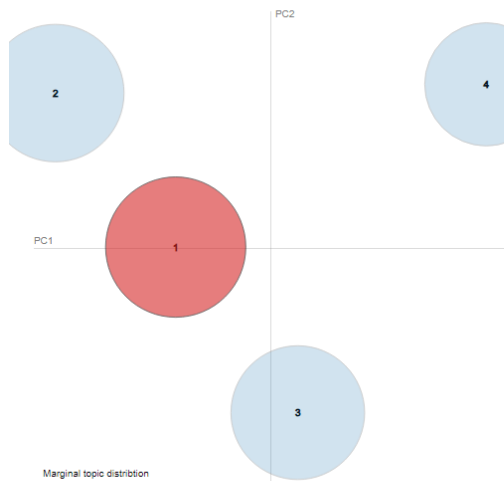
cmtPoster	artPoster	system_id@count
AlphaD	AlphaD	48
larle	larle	46
KUNI0202	KUNI0202	40
lulalachan	lulalachan	26
ggyyworld	ggyyworld	22
osakaleo	osakaleo	22
Shadouie	larle	20



視覺化的分析結果與解釋

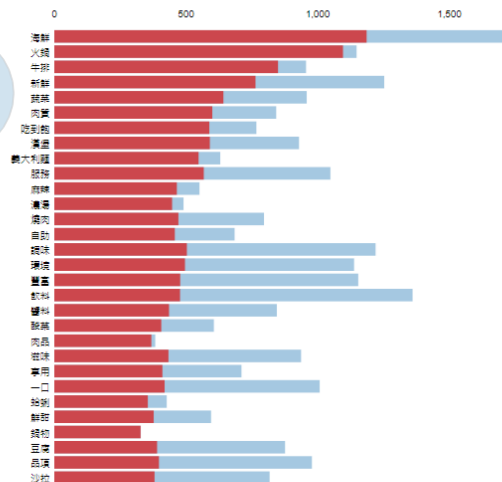
Selected Topic: Previous Topic Next Topic Clear Topic

Intertopic Distance Map (via multidimensional scaling)



Slide to adjust relevance metric:⁽²⁾ $\lambda = 0.81$

Top-30 Most Relevant Terms for Topic 1 (27.5% 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)

LDA

主題分類 1

比較偏向於**海鮮火鍋**的相關議題

有海鮮、火鍋、新鮮、吃蔬菜、肉質、吃到飽等關鍵字

由此可見現在的人都較喜歡吃清淡一點的海鮮火鍋並且吃到飽，也比較注重食材是否新鮮。



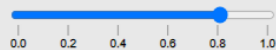
視覺化的分析結果與解釋

LDA

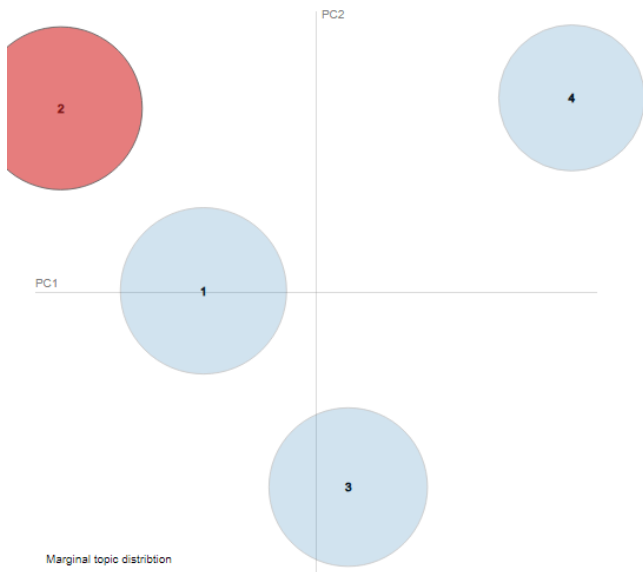
Selected Topic:

Slide to adjust relevance metric:⁽²⁾

$\lambda = 0.81$



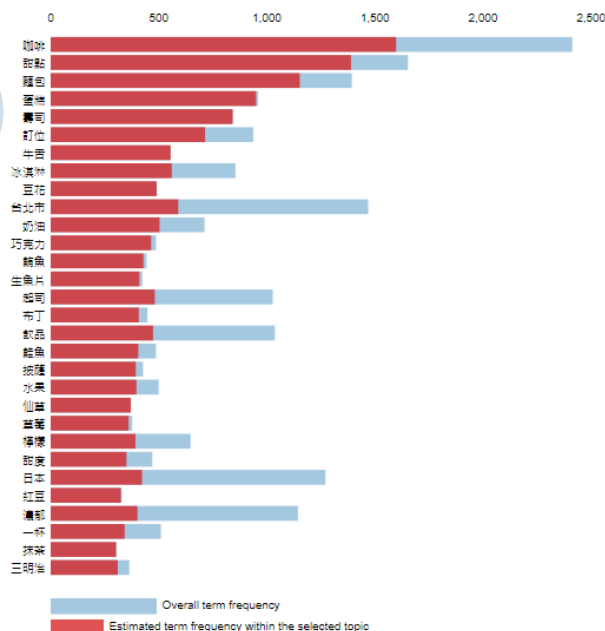
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 2 (26.4% 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) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Stevart & Shirley (2014)

主題分類2

比較偏向於下午茶相關的議題

可以看到關鍵字有咖啡、甜點、麵包、蛋糕等等

現在的人有下午茶的習慣，那也偏向喜歡和咖啡配甜點的習慣。



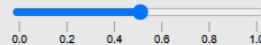
視覺化的分析結果與解釋

LDA

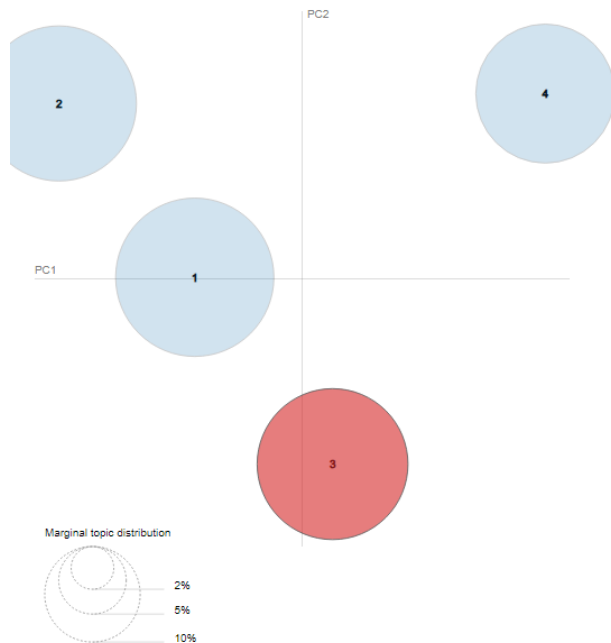
Selected Topic: 3 [Previous Topic](#) [Next Topic](#) [Clear Topic](#)

Slide to adjust relevance metric: (2)

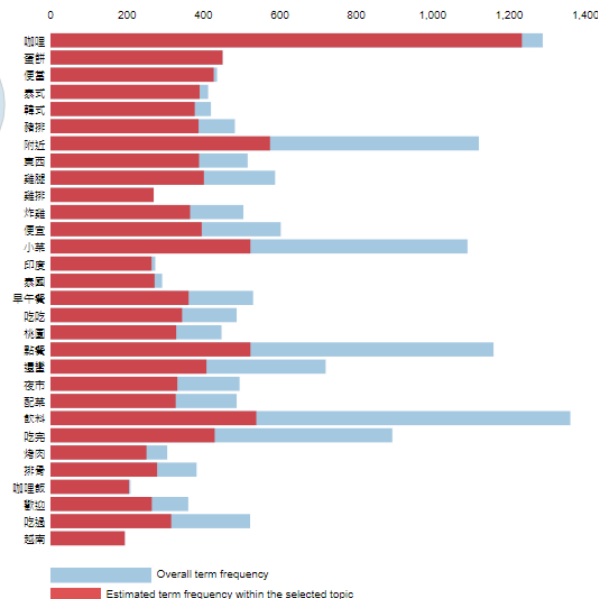
$\lambda = 0.51$



Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 3 (25% 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) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t) / p(w)$; see Sievert & Shirley (2014)

主題分類 3

比較偏向於早午餐的相關議題

有咖哩、蛋餅、便當等關鍵字

現代人較注重生活品質，故除工作日以外，都是早午餐一併用餐。



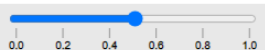
視覺化的分析結果與解釋

LDA

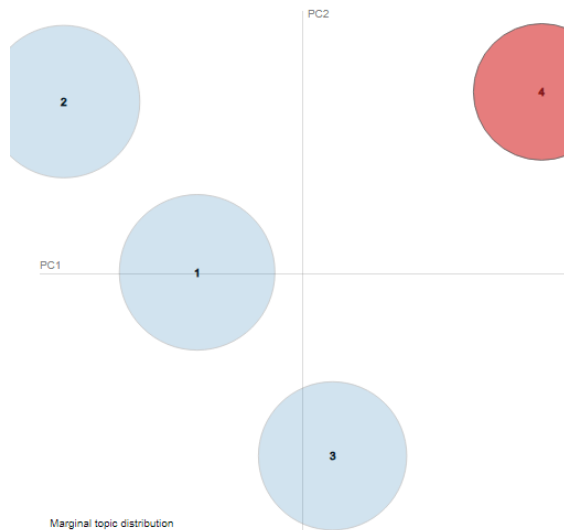
Selected Topic: 4 Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:(2)

$\lambda = 0.51$



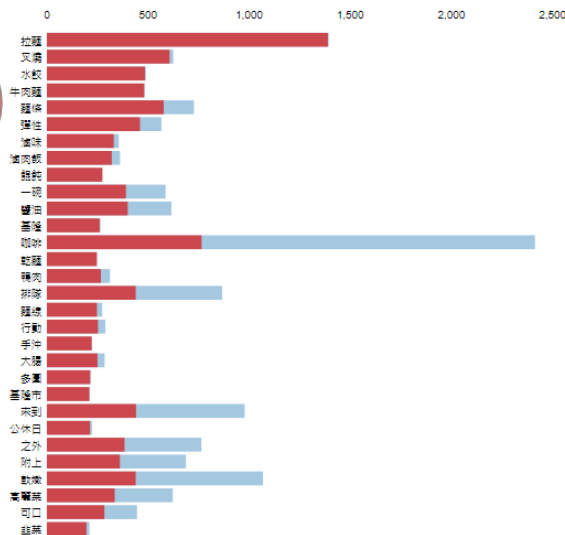
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 4 (21.2% 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))] 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)

主題分類 4

比較偏向於麵食方面

有拉麵、叉燒、水餃、
牛肉麵等關鍵字

也呼應了，現代人較喜
歡吃麵食類的食物。



視覺化的分析結果與解釋

GuidedLDA 主題模型 (65)

參數設定

Input - 7

任務結果

目標欄位 *

result

主題數 *

4

詞彙頻率下限 ⓘ

40

alpha

預設為主題數/50

主題種子字 ⓘ

台式,台灣,雞腿便當,排骨飯,珍珠奶茶,夜市
美式,美國,漢堡,薯條,可樂,熱狗,牛排
日式,日本,壽司,生魚片,拉麵,清酒
韓式,韓國,泡菜,年糕,韓式炸雞,部隊鍋

迭代次數

1000

主題保留關鍵字數量

20

詞彙頻率上限 ⓘ

0.7

Beta

預設為0.1

是否輸出字典

是

儲存更改

利用Guided LDA來檢視國人對
吃食分別關心的重點為何

透過種子字的設置，將四位可
能的候選人，設定成不同類別：

- 台式
- 美式
- 日式
- 韓式



視覺化的分析結果與解釋

Guided LDA

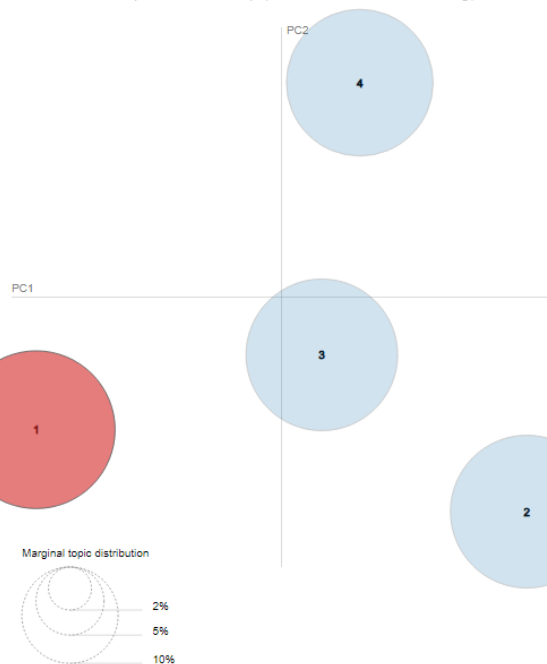
Selected Topic: 0 Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾

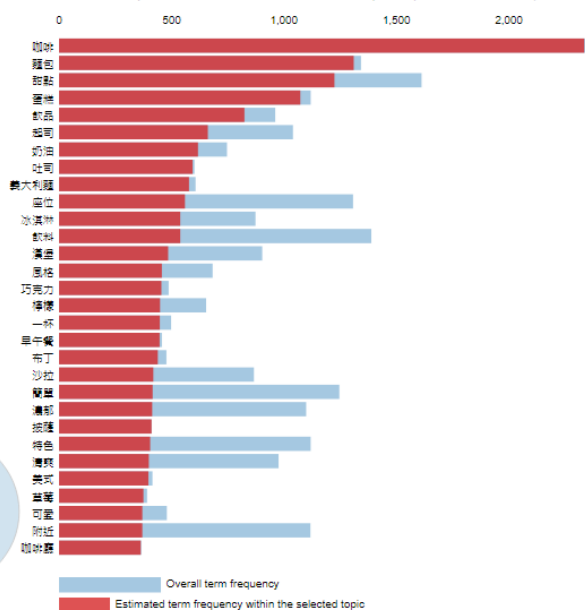
$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 1 (26.8% 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) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

美式，因為美國人較喜歡咖啡，麵包，甜點類的食物。

故關鍵字，都是甜食較多。



祖魯化的分析結果與解釋

Guided LDA

Selected Topic:

Slide to adjust relevance metric: λ

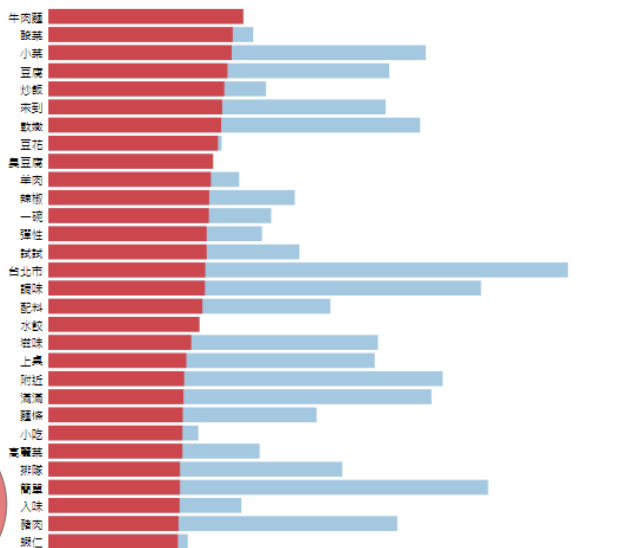
$\lambda = 1$

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 (25.2% of tokens)

0 200 400 600 800 1,000 1,200 1,400 1,600



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)

PC1

PC2



Marginal topic distribution



台式，較多澱粉類的
關鍵字。



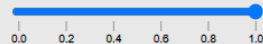
視覺化的分析結果與解釋

Guided LDA

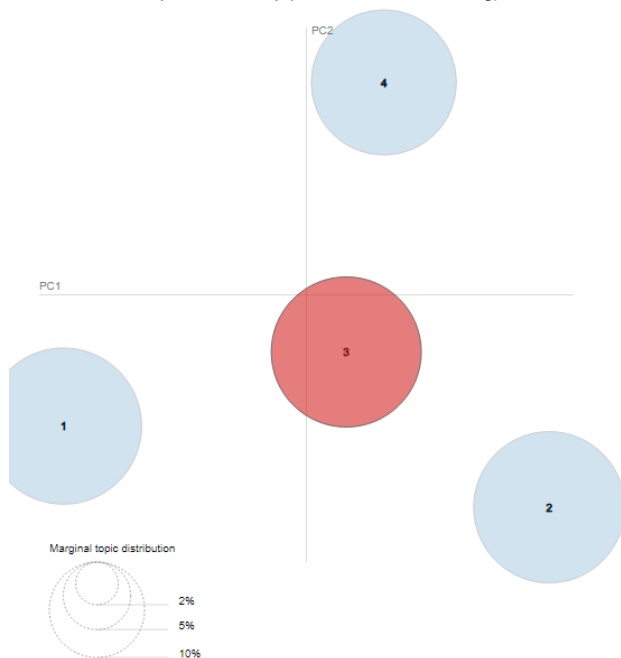
Selected Topic:

Slide to adjust relevance metric:⁽²⁾

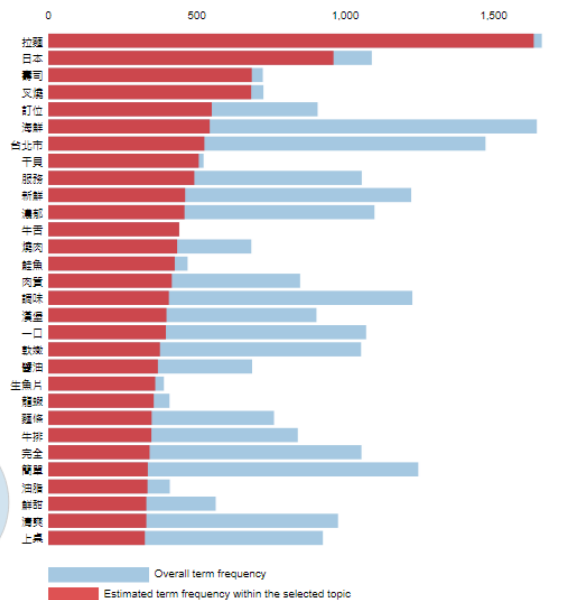
$\lambda = 1$



Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 3 (24.8% of tokens)



1. saliency(term w) = frequency(w) * [sum_t 1 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)

日式
關鍵字較多拉麵、壽司、
叉燒等



視覺化的分析結果與解釋

Guided LDA

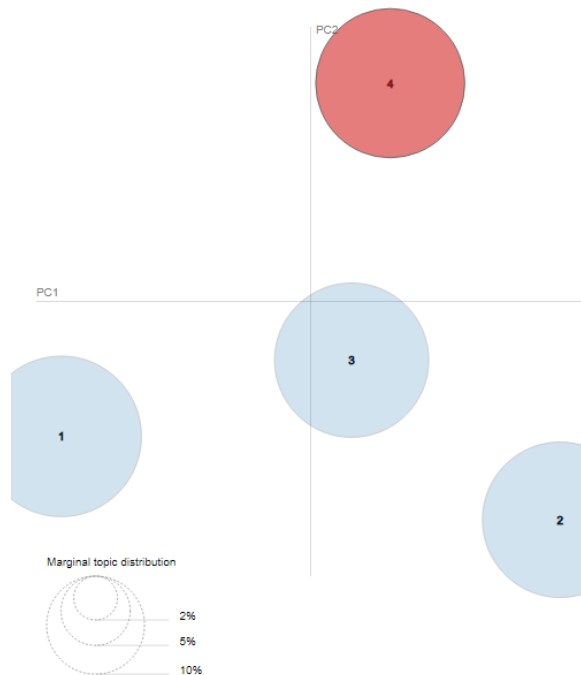
Selected Topic:

Slide to adjust relevance metric:⁽²⁾

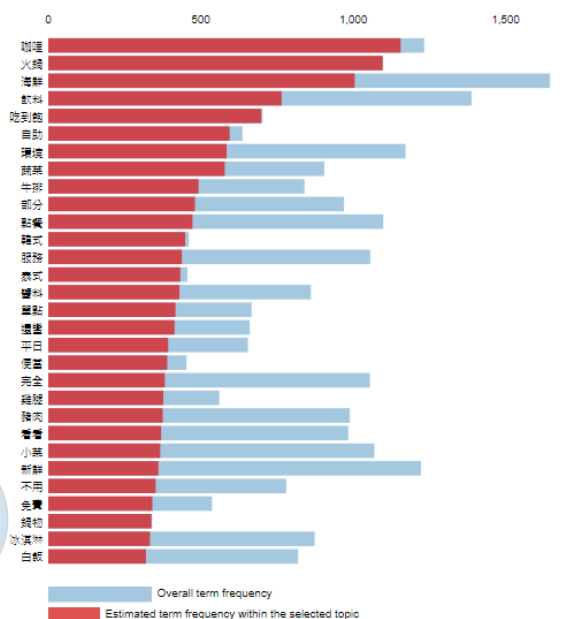
$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 4 (23.2% 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) = $\lambda \cdot p(w | t) + (1 - \lambda) \cdot p(w | t)/p(w)$; see Sievert & Shirley (2014)

韓式則是較多鍋類的相關的詞彙。



結論

結合TarFlow分析、字詞網路圖和社會網路圖，我們得出了以下結論：

1. 飲食偏好：

1. 消費者對於餐廳的食材新鮮度、餐廳的地理位置和營業時間非常關心。
2. 不同餐飲類型（如海鮮火鍋、下午茶、早午餐、麵食）有不同的關鍵字，反映出消費者的具體偏好和需求。

2. 互動模式：

1. 社會網路圖揭示了消費者之間的互動模式，中心節點代表影響力較大的使用者。
2. 關鍵意見領袖在消費者的決策過程中起著重要作用，他們的推薦和評論可能會顯著影響其他消費者的選擇。

3. 營銷策略：

1. 餐飲業者可以根據這些分析結果制定精準的營銷策略，針對不同的消費群體提供個性化的服務和推廣活動。
2. 強調食材新鮮度、提供詳細的餐廳資訊和提升服務質量，可以有效吸引和留住顧客。

- 這些綜合分析提供了對消費者行為的深刻理解，為餐飲業者在競爭激烈的市場中提供了有價值的指標。



Thanks!!

