



## 期末分組\_14

主題：吉伊卡哇

組員：N124320008 邱懷英

N124320009 吳亭緯

N124320017 郭良益 (組長)

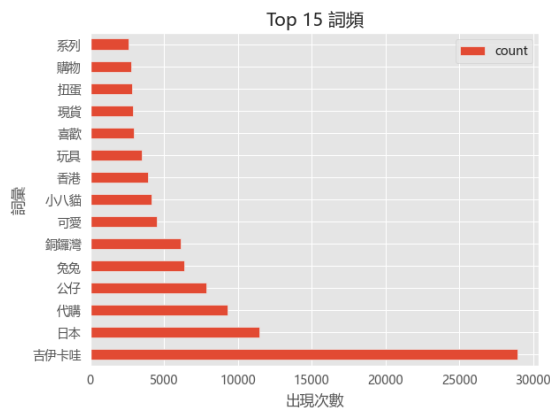
專案連結：[https://github.com/liangyiK/SMA\\_FR](https://github.com/liangyiK/SMA_FR)

### 1. 前言

近年來，吉伊卡哇（ちいかわ，Chiikawa）以其可愛與社會寫實的劇情風格，在社群平台上迅速走紅，引起廣泛關注與討論。本專案資料來自 PTT、YT、IG 與吉伊卡哇相關的文本為研究對象，結合多項自然語言處理與社群分析方法，探討其討論熱潮的語言特徵、情緒波動、主題趨勢與社群結構。

本研究主要分析方法包括：重要字排名、關鍵字提取與文字雲視覺化、TF-IDF 權重分析、情緒分析、時間點事件對照、Bigram 語詞搭配、文章聚類、分類器比較、主題模型（LDA 與 BERTopic）以及社群網路圖建構。透過這些方法，我們希望能全面理解吉伊卡哇在社群中的討論動態與受歡迎的內在機制。

### 2. 重要字詞(排名、文字雲)



每日正負向情緒詞量頻率

詞量頻率

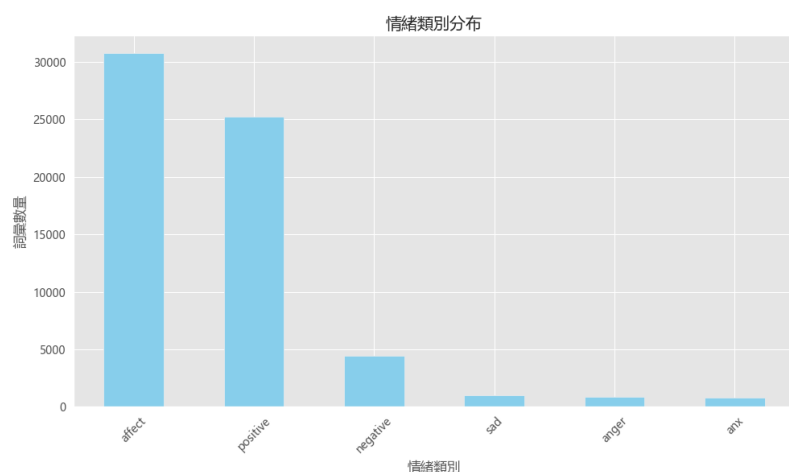
正向情緒詞

負向情緒詞

日期

2024-01 2024-03 2024-05 2024-07 2024-09 2024-11 2025-01 2025-03

- ## ● 情緒字文字雲

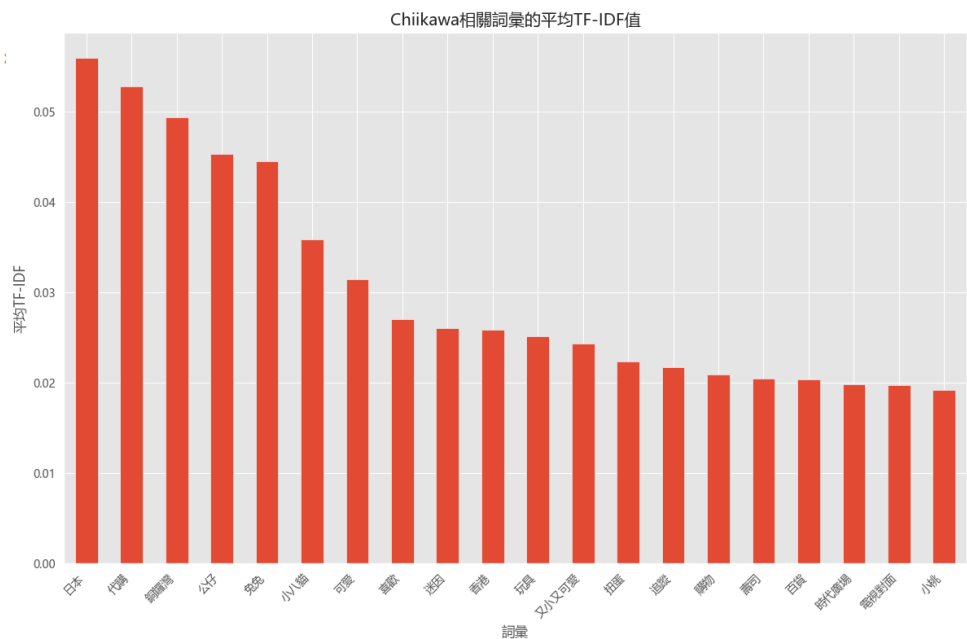


4. TF-IDF 分析

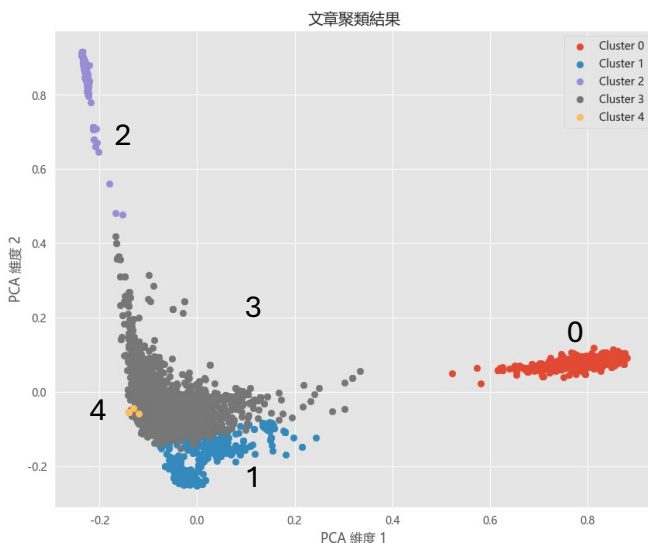
前20個最重要詞彙（基於平均TF-IDF）：

日本	0.055930
代購	0.052809
銅鑼灣	0.049382
公仔	0.045378
兔兔	0.044553
小八貓	0.035842
可愛	0.031493
喜歡	0.027042
迷因	0.026059
香港	0.025921
玩具	0.025161
又小又可愛	0.024369
扭蛋	0.022351
追蹤	0.021777
購物	0.020968
壽司	0.020509
百貨	0.020356
時代廣場	0.019813
電視對面	0.019712
小桃	0.019204

dtype: float64



5. 文章聚類結果



各聚類關鍵詞：

Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
0 銅鑼灣	代購	迷因	日本	又小又可愛
1 時代廣場	旺角	追蹤	兔兔	華山
2 電視對面	八字	喜歡	小八貓	閃店
3 百貨	日本	翻譯	可愛	傢伙
4 公仔	公仔	幽默	壽司	網上

6. Bigram

使用Bigram模型，'吉伊卡哇'後面最可能的詞：

吉伊卡哇：0.2285

代購：0.0880

公仔：0.0556

小八貓：0.0439

兔兔：0.0244

使用Bigram模型生成的文本（種子詞：吉伊卡哇）：

吉伊卡哇 絕區 吉伊卡哇 吉伊卡哇 壽司 系列 超級 可愛 突擊 壽司 吉伊卡哇 吉伊卡哇 小八貓 兔兔 師傅 師父

Bigram模型在測試集上的困惑度：15.53

## 7. 小結：主要著重討論角色及購物資訊，討論熱度也隨著熱門事件而有變化

1156, 'PTT': 861, 'YT': 462) 時間範圍: 2024-01-01 至 2025-03-16

主要高頻詞彙:

- 吉伊卡哇: 28964次
- 日本: 11436次
- 代購: 9272次
- 公仔: 7858次
- 兔兔: 6368次
- 銅鑼灣: 6101次
- 可愛: 4493次
- 小八貓: 4146次
- 香港: 3896次
- 玩具: 3481次

本聚類結果:(後續於LDA、BERT會看到不同的呈現方式)

- 聚類 0 (1163篇): 銅鑼灣, 時代廣場, 電視對面, 百貨, 公仔
- 聚類 1 (931篇): 代購, 旺角, 八字, 日本, 公仔
- 聚類 2 (510篇): 迷因, 追蹤, 喜歡, 翻譯, 幽默
- 聚類 3 (9623篇): 日本, 兔兔, 小八貓, 可愛, 壽司
- 聚類 4 (252篇): 又小又可愛, 華山, 閃店, 傢伙, 網上 ... 情緒分析結果:
- 整體正向情緒詞比例: 85.04%
- 情緒分數平均值: 0.6453
- 情緒分數中位數: 0.9442

## 8. 分類器選擇(1.隨機森林 2.SVM)

模型建議

**首選隨機森林**：綜合指標最佳，適用於大多數場景。 **次選 SVM**：若計算資源充足且需極致性能。 **淘汰多項式貝氏**：除非特徵明顯符合多項分佈假設。

5-fold 交叉驗證結果：

邏輯斯迴歸 → 準確率: 0.729，精確率: 0.797，召回率: 0.796

多項式貝氏 → 準確率: 0.686，精確率: 0.593，召回率: 0.531

決策樹 → 準確率: 0.616，精確率: 0.730，召回率: 0.699

隨機森林 → 準確率: 0.802，精確率: 0.837，召回率: 0.814

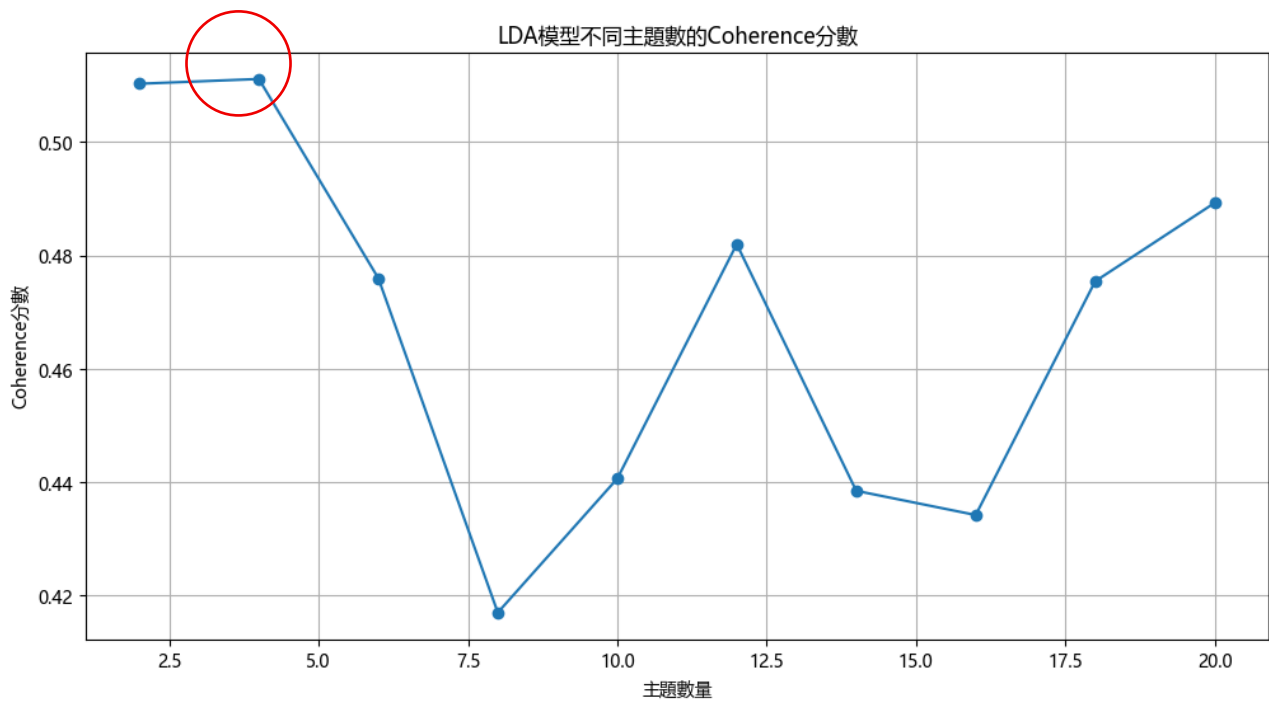
SVM → 準確率: 0.760，精確率: 0.809，召回率: 0.804

## 9. LDA 主題模型

前處理做好之後，進行自動化尋找最佳 LDA 主題數量，並進行主題模型評估與可視化。主要分為四個階段：若使用 Tarflow 可能需要先視覺化後人工辨識適合的主題數，這邊採用自動化判斷，主要分為四個步驟

- 主題數量範圍設定，測試 2~20 的偶數列模型表現，也能避免跑太久
- 模型訓練與評估，用 Coherence(越高越好)、Perplexity(越低越好)評估
- 可視化分析，這邊的可視化指的是說模型在哪個主題數表現的最好，而非先產出 vis 圖
- 最佳模型選擇，根據結果自動化選擇最佳主題，大致上主題可以這樣解讀

0: '商品代購', 1: '角色討論', 2: '線下活動', 3: '迷因創作'



最佳主題數量: 4, Coherence 分數: 0.5111

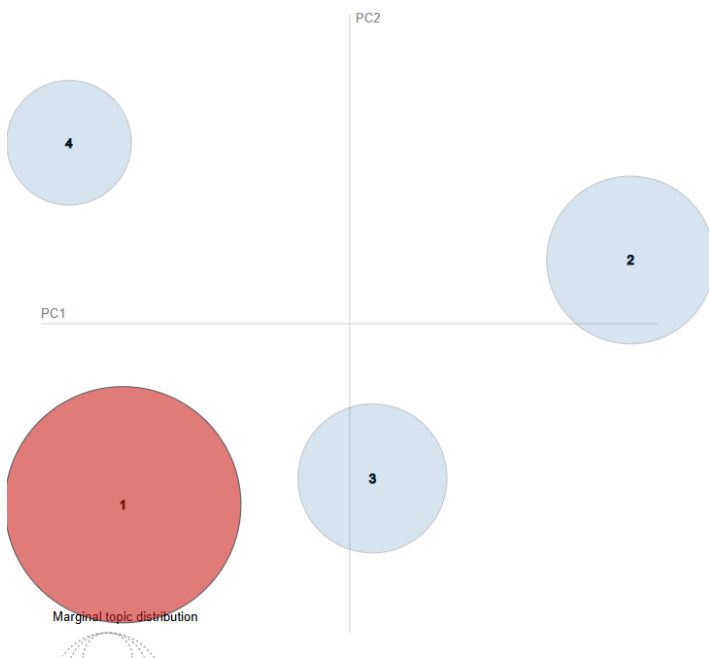
最佳 LDA 模型的主題:

主題 #0: 0.014\*"可愛" + 0.010\*"日本" + 0.008\*"美食" + 0.007\*"兔兔" + 0.007\*"小八貓" + 0.005\*"角色" + 0.005\*"活動" + 0.005\*"聯名" + 0.005\*"限定" + 0.005\*"真的"

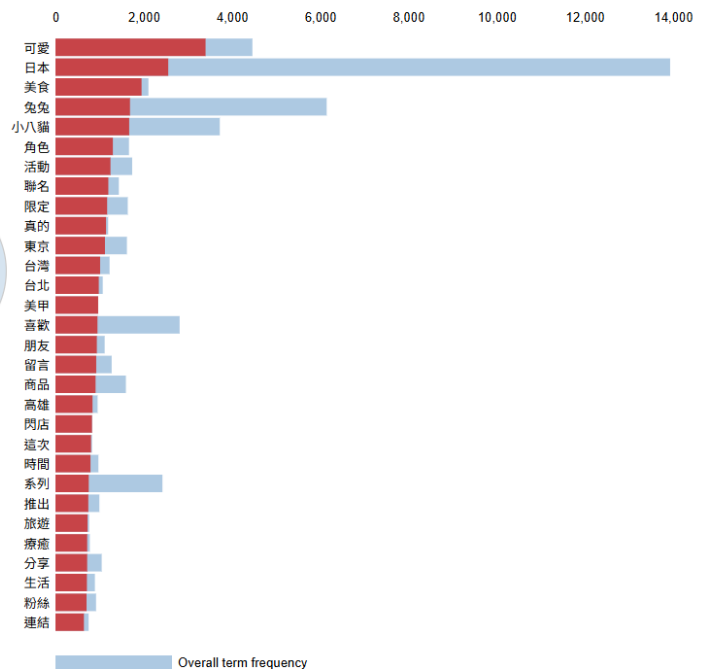
Selected Topic:

Slide to adjust relevance metric: (2)   $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 1 (45.9% of tokens)



主題 #1: 0.057\***"銅鑼灣"** + 0.052\***"日本"** + 0.045\***"代購"** + 0.041\***"公仔"** + 0.037\***"玩具"** + 0.022\***"對面"** + 0.019\***"精品"** + 0.019\***"香港"** + 0.018\***"旺角"** + 0.017\***"購物"**

Selected Topic:

Previous Topic

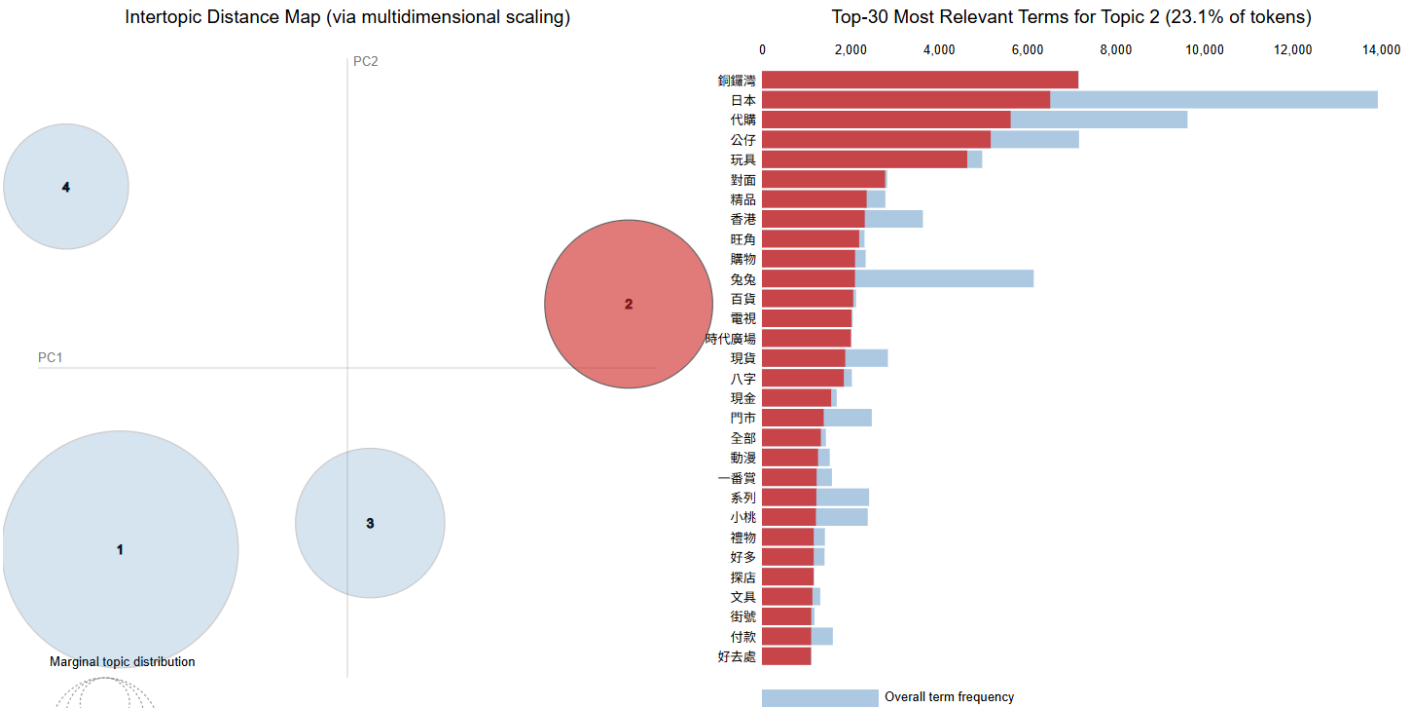
Next Topic

Clear Topic

Slide to adjust relevance metric:<sup>(2)</sup>

$\lambda = 1$

0.00.20.40.60.81.0



主題 #2: 0.046\***"日本"** + 0.040\***"代購"** + 0.022\***"兔兔"** + 0.015\***"小八貓"** + 0.015\***"公仔"** + 0.009\***"下單"** + 0.009\***"網站"** + 0.008\***"現貨"** + 0.008\***"貨品"** + 0.007\***"小桃"**

Selected Topic:

Previous Topic

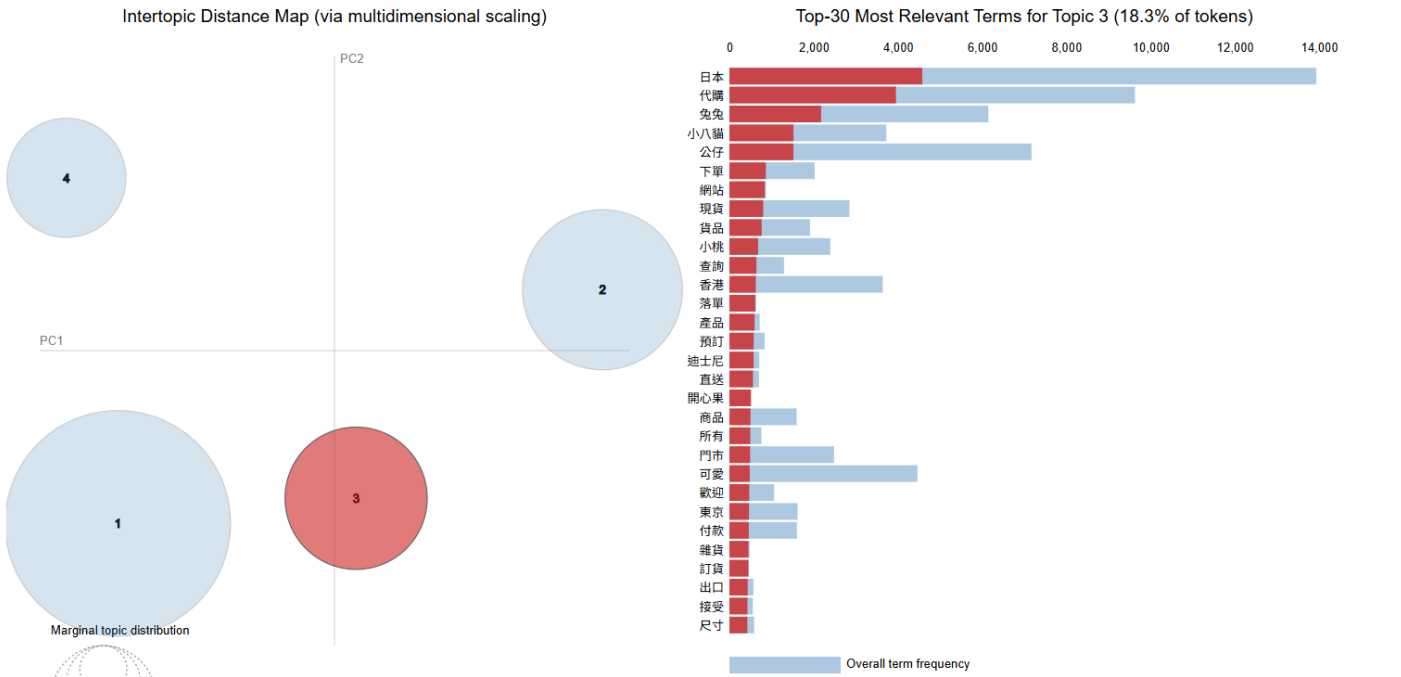
Next Topic

Clear Topic

Slide to adjust relevance metric:<sup>(2)</sup>

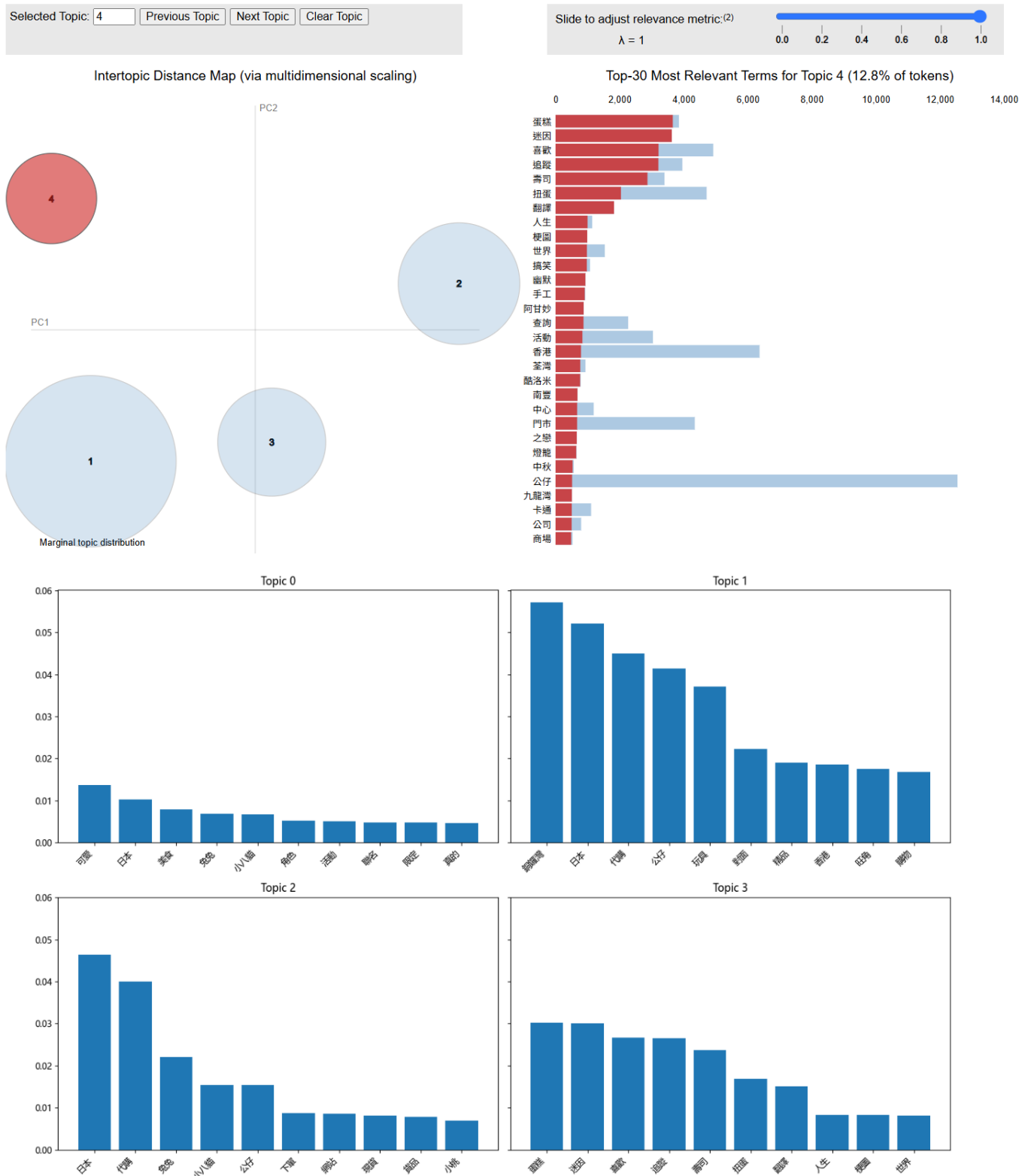
$\lambda = 1$

0.00.20.40.60.81.0





主題 #3:  $0.030 \times \text{"蛋糕"} + 0.030 \times \text{"迷因"} + 0.027 \times \text{"喜歡"} + 0.027 \times \text{"追蹤"} + 0.024 \times \text{"壽司"} + 0.017 \times \text{"扭蛋"} + 0.015 \times \text{"翻譯"} + 0.008 \times \text{"人生"} + 0.008 \times \text{"梗圖"} + 0.008 \times \text{"世界"}$



### 'IG', 'YT', 'PTT'三平台的 LDA :

由於前一段的 LDA 模型是跑所有平台的分析，可能各個平台會有不同的可能性，故我們分別在處理了，結果如下

- IG：可愛形象、迷因、美食、代購、精品分享
- YT：頻道經營、商品開箱、角色介紹、國際話題
- PTT：活動/新聞、商品交易、角色討論

**IG 平台最佳主題數量: 3, Coherence 分數: 0.6663**

IG 平台最佳 LDA 模型的主題:

主題 #0: 0.012\*"可愛" + 0.011\*"喜歡" + 0.010\*"蛋糕" + 0.009\*"追蹤" + 0.008\*"迷因" + 0.007\*"小八貓" + 0.007\*"美食" + 0.006\*"壽司" + 0.006\*"兔兔" + 0.006\*"活動"

主題 #1: 0.044\*"日本" + 0.037\*"代購" + 0.016\*"旺角" + 0.014\*"門市" + 0.013\*"八字" + 0.012\*"下單" + 0.011\*"兔兔" + 0.009\*"現貨" + 0.009\*"付款" + 0.009\*"東京"

主題 #2: 0.052\*"銅鑼灣" + 0.047\*"日本" + 0.044\*"公仔" + 0.035\*"玩具" + 0.033\*"代購" + 0.021\*"兔兔" + 0.020\*"香港" + 0.020\*"精品" + 0.017\*"購物" + 0.016\*"對面"

**YT 平台最佳主題數量: 8, Coherence 分數: 0.4720**

YT 平台最佳 LDA 模型的主題:

主題 #0: 0.032\*"新聞" + 0.016\*"頻道" + 0.014\*"訂閱" + 0.012\*"悠遊" + 0.012\*"一番賞" + 0.012\*"黃牛" + 0.010\*"直播" + 0.009\*"粉絲" + 0.008\*"主播" + 0.007\*"扭蛋"

主題 #1: 0.032\*"扭蛋" + 0.028\*"壽司" + 0.018\*"日本" + 0.016\*"影片" + 0.012\*"介紹" + 0.010\*"角色" + 0.007\*"訂閱" + 0.006\*"可愛" + 0.006\*"分析" + 0.006\*"粉絲"

主題 #2: 0.023\*"娃娃" + 0.017\*"新聞" + 0.015\*"影片" + 0.011\*"訂閱" + 0.010\*"可愛" + 0.009\*"閃店" + 0.009\*"會員" + 0.009\*"遊戲" + 0.008\*"歡迎" + 0.008\*"零食"

主題 #3: 0.023\*"著作" + 0.015\*"劉萱" + 0.012\*"漫畫" + 0.011\*"画像" + 0.011\*"侵害" + 0.011\*"合作" + 0.010\*"英雄" + 0.010\*"瑪麗" + 0.009\*"動畫" + 0.009\*"精華"

主題 #4: 0.012\*"美国" + 0.011\*"表示" + 0.009\*"娃娃" + 0.009\*"頻道" + 0.008\*"訂閱" + 0.008\*"迷因" + 0.008\*"报道" + 0.008\*"中国" + 0.007\*"日本" + 0.006\*"去年"

主題 #5: 0.032\*"波子" + 0.028\*"鳴話" + 0.018\*"開箱" + 0.016\*"影片" + 0.015\*"玩具" + 0.014\*"貼圖" + 0.012\*"娃娃" + 0.012\*"訂閱" + 0.011\*"公仔" + 0.008\*"系列"

主題 #6: 0.026\*"東京" + 0.021\*"日本" + 0.018\*"壽司" + 0.015\*"聯名" + 0.010\*"新聞" + 0.010\*"可愛" + 0.010\*"扭蛋" + 0.009\*"起司" + 0.007\*"推出" + 0.007\*"蛋姬"

主題 #7: 0.013\*"直播" + 0.013\*"尊重" + 0.012\*"聊天室" + 0.010\*"會員" + 0.009\*"齋戒" + 0.008\*"穆斯林" + 0.008\*"探店" + 0.008\*"学生" + 0.007\*"台灣" + 0.007\*"提及"

**PTT 平台最佳主題數量: 3, Coherence 分數: 0.5973**

PTT 平台最佳 LDA 模型的主題:

主題 #0: 0.029\*"吉伊卡哇" + 0.020\*"高雄" + 0.012\*"新聞" + 0.010\*"完整" + 0.010\*"活動" + 0.009\*"燈會" + 0.008\*"記者" + 0.008\*"媒體" + 0.008\*"提燈" + 0.007\*"冬日"

主題 #1: 0.026\*"商品" + 0.015\*"代買" + 0.014\*"代購" + 0.011\*"吉伊卡哇" + 0.009\*"運費" + 0.009\*"價格" + 0.008\*"面交" + 0.008\*"購入" + 0.006\*"購買" + 0.006\*"匯率"

主題 #2: 0.057\*"吉伊卡哇" + 0.016\*"小八貓" + 0.015\*"兔兔" + 0.010\*"可愛" + 0.008\*"看到" + 0.007\*"日本" + 0.006\*"高雄" + 0.006\*"真的" + 0.006\*"知道" + 0.005\*"動畫"

## Guided LDA 模型分析

三個種子詞是設定三主角的名字跟暱稱 (註:使用縮寫是因為之前的替代詞已經建立，所以才只建立少的種子詞)

['吉伊卡哇', '小可愛'],

['小八', 'ハチワレ'],

['烏薩奇', '兔兔', 'うさぎ']

產出三個主題字詞如下

Topic 0 銅鑼灣、香港、購物、百貨 香港代購/購物主題 - 主要討論香港地區的購物資訊

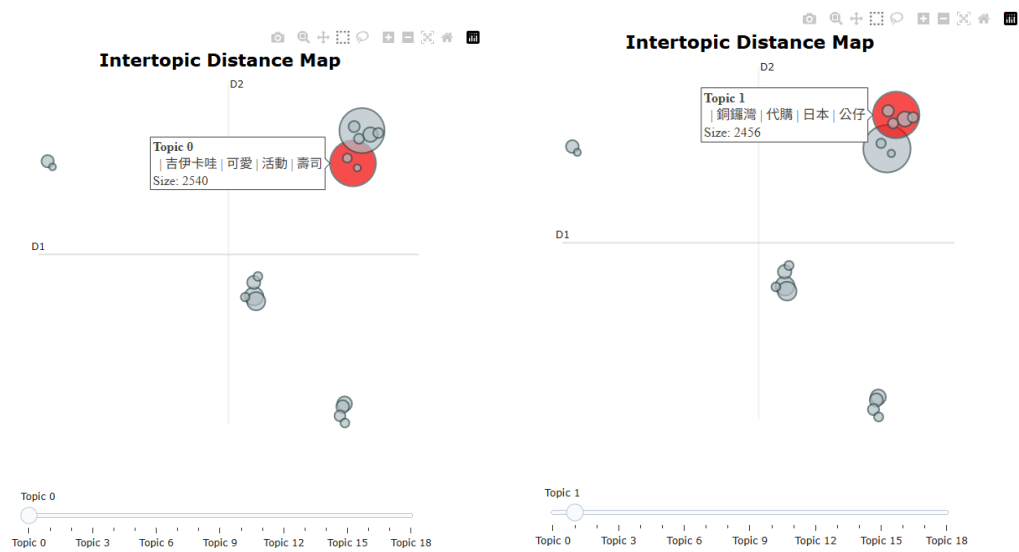
Topic 1 可愛、蛋糕、美食、聯名 美食/周邊商品主題 - 討論角色相關的美食或聯名商品

Topic 2 代購、門市、下單、付款 代購服務主題 - 討論代購流程與服務

結果顯示可以看到**購買相關主題**還是大大的影響了主題分析，原始的資料集有許多討論皆是集中於購買身上



10.BERT Topic 主題模型說明



主題	關鍵字	GPT 推論摘要
0	吉伊卡哇, 可愛, 活動, 壽司, 日本, 聯名, 台, 藏, 高雄	以吉伊卡哇為主題的高雄活動・結合日本文化與壽司聯名・吸引民眾參與
1	銅鑼灣, 代購, 日本, 公仔, 玩具, 小可愛, 吉伊卡哇, 精品, 對面	圍繞銅鑼灣的日本商品代購與流行文化
2	吉伊卡哇, 可愛, 吉伊, 知道, 烏薩奇, 小八, 真的, 題, 喜歡	描繪角色之間的情感連結與可愛互動
3	迷因, 追蹤, 喜歡, 翻譯, 阿甘妙, 梗圖, 梗, 幽默, 搞笑, 人生	迷因文化表達生活情感・透過幽默產生共鳴
4	係, 嘅, 網站, 唔, 貨品, 佢, 睇, 咗, 日本	查找日本貨品時的失望與困惑情緒
5	又小又可愛, 吉伊卡哇, 著小編, 可愛, 早操, 跳, 幾點, 福井, 太太	分享角色可愛日常活動・吸引粉絲喜愛
6	露天, 特價, 吉伊卡哇, 克店, 北約, 華景, 天地, 旗艦店, 日空版	露天商場特價促銷與旗艦店活動
7	吉伊卡哇, 烏薩奇, 小八, 栗, 栗子饅頭, 卡哇, 吉依, 小八貓, 兔哥	角色與文化飲食結合・呈現童趣與視覺創意
8	小八貓, 吉伊卡哇, 代購, 烏薩奇, 公仔, 小八, 小可愛, 吉伊, 嚕	角色公仔與收藏文化・反映年輕人對萌文化熱愛
9	手工, 紮, 作, 燈籠, 包, 九龍灣, 中秋, 號, 早	九龍灣中秋手工活動體驗傳統文化
10	绝区, 吉伊卡哇, 布, 抓, 哥布林, 服装, 金刚, 与, 吉伊, 狼	奇幻冒險主題角色與裝備・融合戰鬥情節
11	係, 唔, 咗, 咁, 冇, 嘅, 抽, 尋日, 啲	日常生活口語對話・傳達隨性與感受
12	開心果, 日本, 代購, 現貨, 小本經營, 留港, 外均, 免問, 註明	香港小本代購經營・主打日本商品與現貨銷售
13	美甲, 蠟染, 手繪, 預約, 星期六, 光療, 凝膠, 指甲, 星期日	美甲服務分享與預約體驗
14	吉伊卡哇, 吉伊, 口哇, 吉其, 蕉, 猜, 推坑, 興奮, 吉, 開心	表達對可愛事物的喜愛與社群互動
15	探店, 盒, 玩, 扭蛋, 吉伊卡哇, 公仔, 伯, 公道, 集, 開箱	探索扭蛋與收藏公仔的開箱樂趣
16	撮, 写, 錄, 活記, 影友会, 繫, 好人, 攝影, 又小又可愛, 生活	攝影愛好者拍攝可愛事物並分享交流
17	商品, 代買, 面交, 購入, 請, 運費, 價格, 匯率, 費用	商品代買與交易過程中的費用與方式
18	烏薩奇, 代購, 兔哥, 發聲, 小可愛, 日本, 兔, 寶寶, 公仔	兔哥與烏薩奇角色代購與收藏經驗

11.主題模型小結：

模型類型	主題結構	主題摘要	優點	限制
整體 LDA	4 主題	商品代購、角色討論、線下活動、迷因創作	主題分明、詞彙直觀	缺乏平台語境與情感層次
Guided LDA	3 主題(三個主角為種子)	香港代購、美食聯名、代購服務流程	精準聚焦預設角色/情境	主題受限於種子詞，無法發掘未知主題
分平台 LDA	IG:3, YT:8 PTT:3	IG 偏可愛/精品 YT 偏開箱/國際話題 PTT 偏活動/交易	能比較平台差異	模型建置與解釋較複雜
BERTopic	18 主題	涵蓋活動、情感互動、迷因、開箱、交易等細緻面向	語意連貫、主題豐富	命名需人工解釋、視覺不明確

12. 社群網路圖

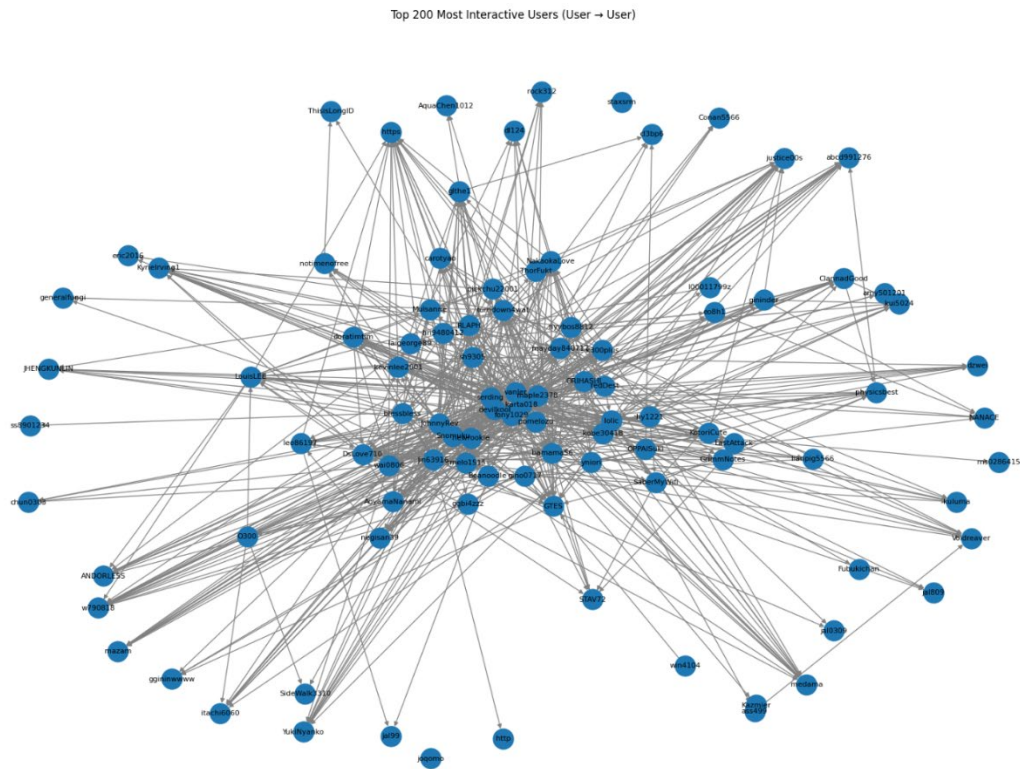
使用者社群指標計算 ( Social Network Metrics )

針對先前篩選出的互動子圖 `G_top` ( 前100名活躍使用者 )，計算多種常見的社群網路分析指標：

- Degree / In-Degree / Out-Degree：互動總數、收到與主動的互動數。
- Betweenness Centrality：衡量一個節點是否為資訊流通的橋樑。
- Closeness Centrality：距離其他節點的平均距離，代表資訊接觸速度。
- Eigenvector Centrality：連到「重要節點」的節點更具影響力。
- PageRank：衡量使用者穩定的影響力 ( 類似 Google 排名演算法 )。

這些指標可用來找出：

- 最有影響力的使用者
- 社群中的資訊中介者或領導者
- 活躍或邊緣使用者的分布狀況



## ◆ 結論

本專案透過多面向的文本探勘與語意分析方法，深入解析吉伊卡哇在社群平台上的討論內容與情緒脈絡，綜合研究成果可歸納以下幾點關鍵發現：

### 1. 熱門字詞與語意焦點

文字雲與 TF-IDF 分析顯示，吉伊卡哇相關的高頻詞多聚焦於角色名稱（如小八、兔兔）、周邊產品（如布丁、公仔）與熱門活動（如點盤、聯名壽司），反映出使用者高度關注「角色互動」與「商品收集」的雙主軸。

### 2. 情緒分析結合事件節點

搭配討論熱點時間軸分析，我們發現特定事件如藏壽司活動、聲優變聲爭議與冬季遊樂園等，明顯牽動情緒波動，且呈現正負交錯的社群反應，顯示粉絲在熱愛角色的同時也具備批判與參與意識。

### 3. 多元主題模型揭示語境層次

本研究採用四種主題建模技術進行分析：整體 LDA、分平台 LDA、BERTopic 與 Guided LDA。

- **LDA**：辨識出清晰的主題如「商品代購」、「活動參與」、「迷因創作」等
- **Guided LDA**：藉由指定角色種子詞，有效聚焦在角色導購、美食聯名與購物體驗等精準主題上。
- **分平台 LDA**：則呈現平台間語境差異，IG 偏好可愛分享，PTT 集中於新聞與商品交易實體活動，YT 則傾向影片與頻道內容以及開箱討論
- **BERTopic**：能細緻地刻畫語意主題與口語語氣，擴展了角色情感、開箱經驗、迷因文化等層面；

### 4. 分類器比較與文章聚類分析

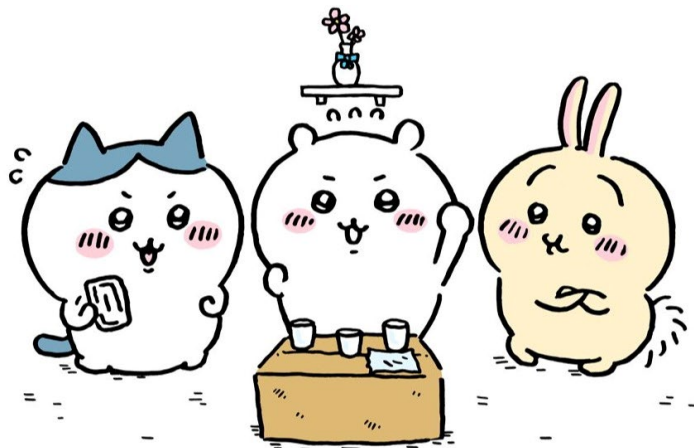
多模型比較中以**隨機森林分類器**表現最佳（準確率達 0.802），未來可應用於社群情緒自動化分析；而 KMeans 聚類結果亦顯示討論主題分布集中、結構清晰，有助後續語意分類任務應用。

### 5. 社群網絡結構與影響節點觀察

社群網路圖揭示出部分高度連結節點，顯示吉伊卡哇社群中存在潛在的意見領袖與擴散中心，雖以單向發文為主，但仍可見到具傳播力的互動聚落。

### 6. 總結：

吉伊卡哇的社群熱潮背後，不僅體現出角色魅力與品牌經營策略，更是一場跨足情緒投射、商品收藏與迷因文化的集體參與現象。本研究透過本學期所學到的文字探勘技術，成功找出時下熱門的討論焦點與語意結構，也為日後動漫角色行銷、社群輿情監測與內容推薦系統，提供具體可行的分析框架與資料基礎。



THANK YOU