

社群媒體分析

專案二

N104220008 王家得

N104020026 林冠宏

N104220004 莊佳宜

N104220027 楊蕙瑄



動機和分析目的

因為社群媒體的蓬勃發展，而造就了網紅當道的現象，
網紅的影響力甚至有超越了專家的趨勢，甚至開起了線上課程，
此次研究延續期中專案報告爬取的原始資料，
來自PPT的八卦版與房屋版的討論，
針對發文和留言分別進行主題分類與社群網路分析。



主題：PTT中對於線上課程討論之主題與網路分析

原始資料

選擇看板：八卦版與房屋版

搜尋關鍵字：線上課程

日期區間：2022/10/01-2023/04/17

資料集描述：



欄位：

system_id	artUrl	artTitle	artDate	artPoster	artCatagory	artContent	artComment	e_ip	insertedDate	dataSource
文章ID	網址	文章標題	發布日期	發布帳號	看板	文章內容	文章留言	來源	留言日期	PTT

資料分析過程

01

定義主題

分別使用
LDA主題模型與
GuidedLDA
建立主題模型

02

分析主題

計算主題
篇數與情緒
並視覺化呈現

03

分析留言

針對主題
計算留言數量與情緒
並視覺化呈現

04

社會網路圖

繪製發文者、留言者
與主題關係
的社會網路圖



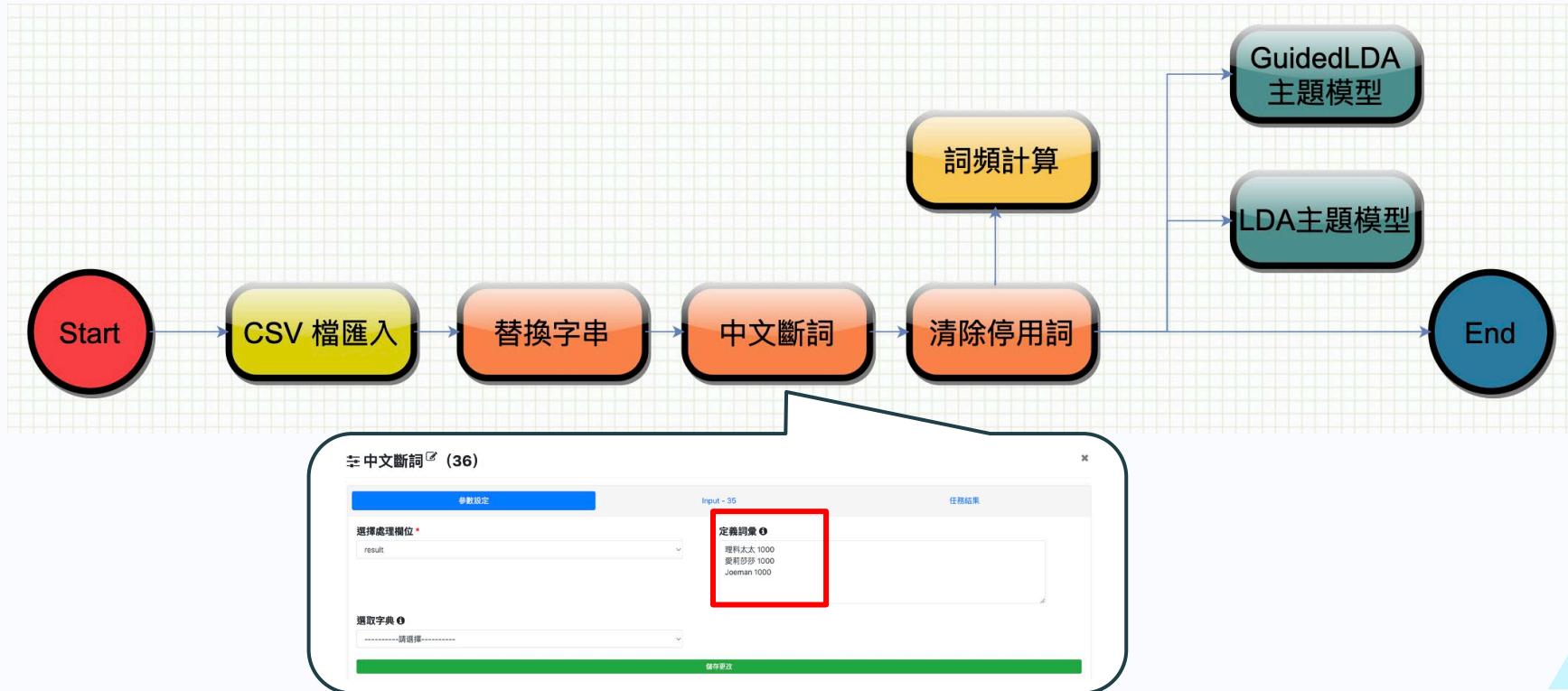
01

定義主題

分別使用LDA主題模型與GuidedLDA建立主題模型

實作步驟-LDA

匯入原始資料後，進行基本預處理，中文斷詞中定義出「理科太太」、「愛莉莎莎」、「Joeman」三個特殊名稱



LDA主題模型參數設定

LDA主題模型 (115)

x

參數設定

Input - 37

任務結果

目標欄位 *

result

主題數 *

3

詞彙頻率下限 ⓘ

5

alpha

預設為主題數/50

chucksize ⓘ

預設為2000

是否輸出字典

是

迭代次數

1000

主題保留關鍵字數量

10

詞彙頻率上限 ⓘ

0.8

Beta

預設為0.1

update_every ⓘ

1

儲存更改

LDA Vis – Topic 1

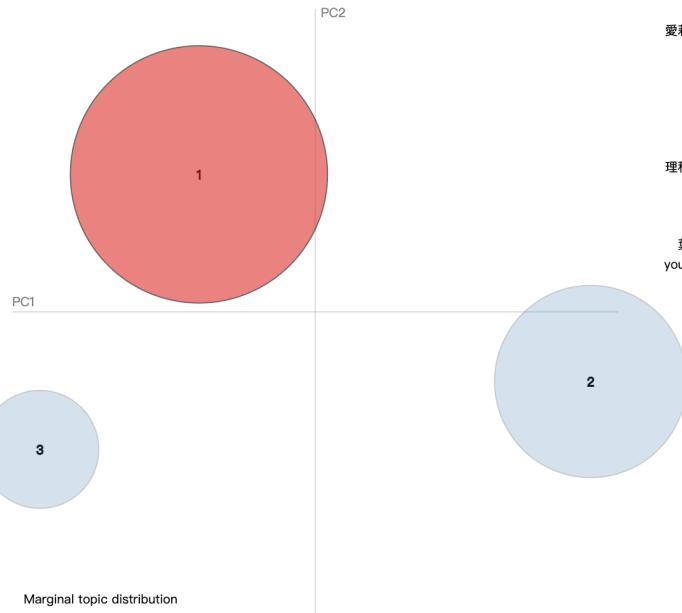
Selected Topic: 0

Slide to adjust relevance metric:(2)

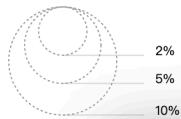
$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

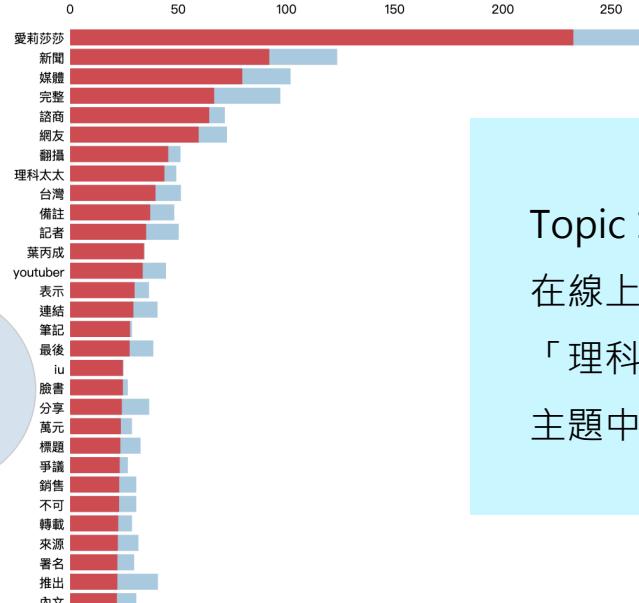
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 1 (56.6% of tokens)



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

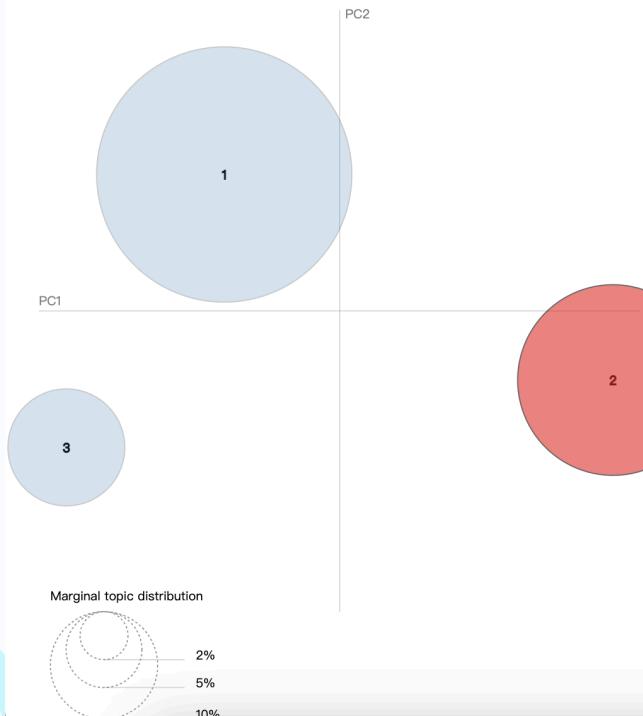
Topic 1-線上課程

在線上課程議題中的「愛莉莎莎」、
「理科太太」、「葉丙成」均在這
主題中被提到

LDA Vis – Topic 2

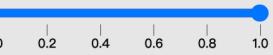
Selected Topic: 0

Intertopic Distance Map (via multidimensional scaling)

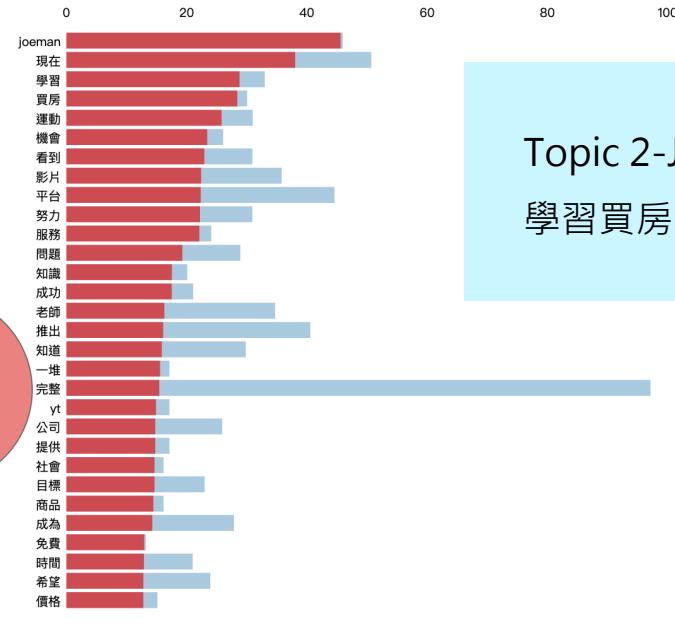


Slide to adjust relevance metric:(2)

$\lambda = 1$



Top-30 Most Relevant Terms for Topic 2 (31.6% of tokens)



Topic 2-Joeman買房課程
學習買房知識相關字眼

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

LDA Vis – Topic 3

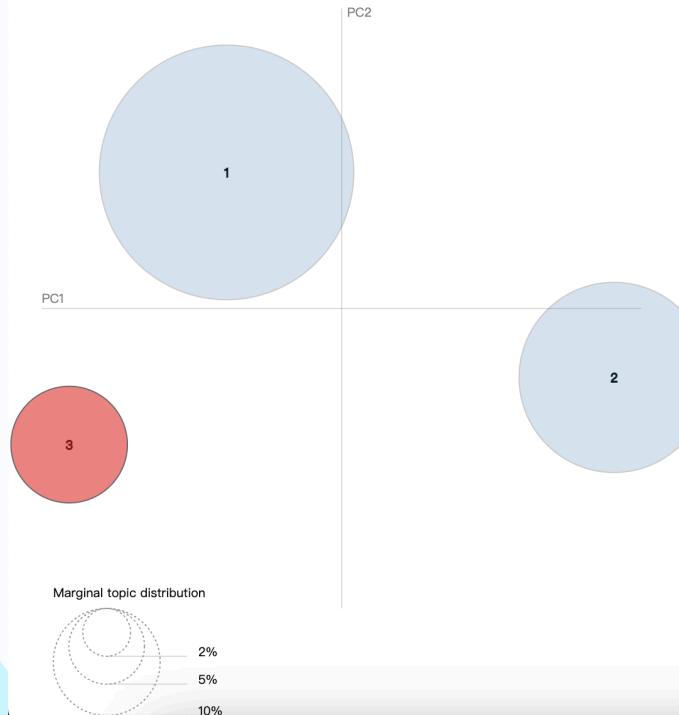
Selected Topic: 0

Slide to adjust relevance metric:(2)

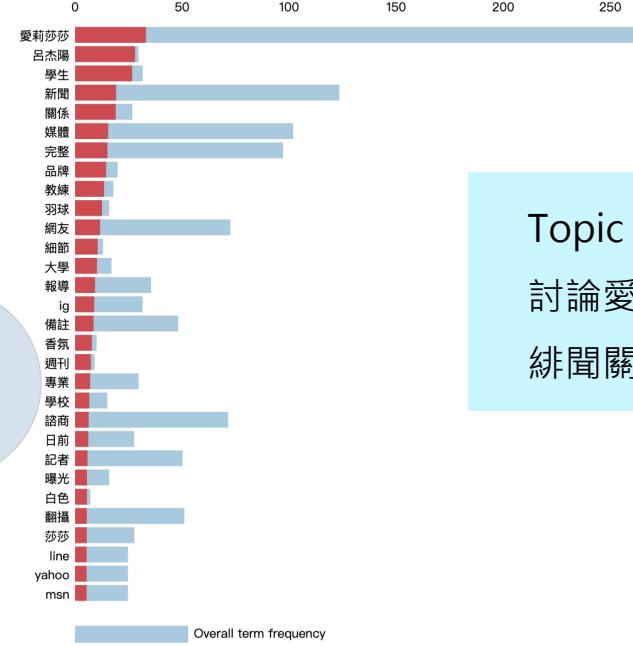
$\lambda = 1$

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 3 (11.9% of tokens)



Topic 3-愛莉莎莎戀情

討論愛莉莎莎與羽球教練的
緋聞關係

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

GuidedLDA模型參數設定

GuidedLDA 主題模型 (114)

參數設定

Input - 37

任務結果

目標欄位 *

result

主題數 *

5

詞彙頻率下限 ⓘ

4

alpha

預設為主題數/50

主題種子字 ⓘ

理科太太,筆記,心理
艾莉莎莎,呂杰陽,網紅
九妹,Joeman,買屋,買房,預售屋
葉丙成,教授,簡報,台大

迭代次數

1000

主題保留關鍵字數量

10

詞彙頻率上限 ⓘ

0.8

Beta

預設為0.1

是否輸出字典

是

儲存更改

GuidedLDA Vis – Topic 1

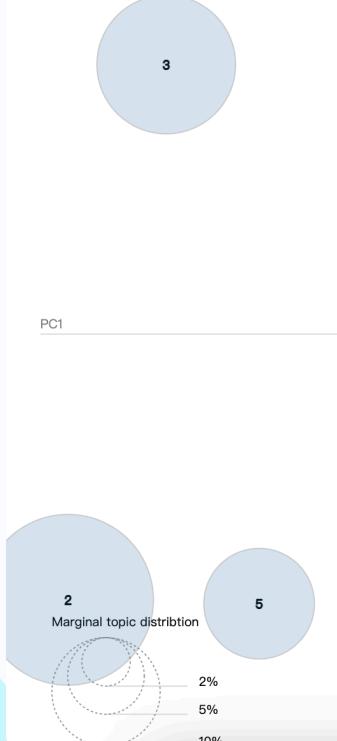
Selected Topic: 0

Slide to adjust relevance metric:(2)

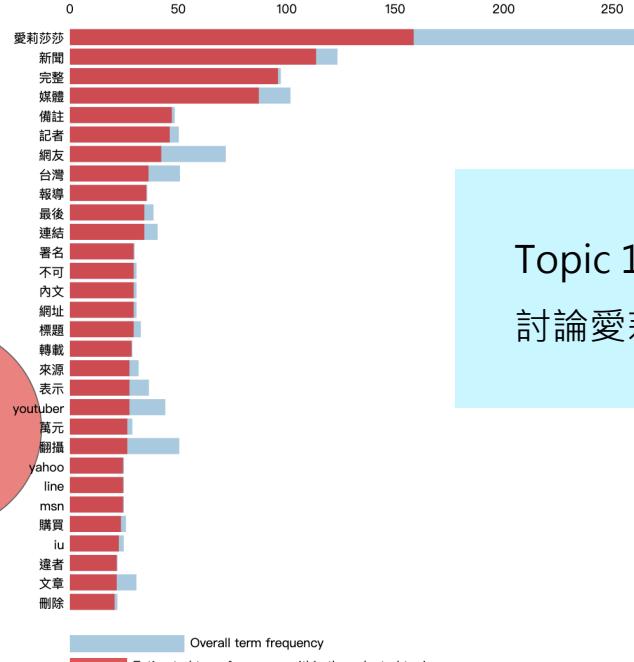
$\lambda = 1$

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 (35.7% of tokens)



Topic 1-愛莉莎莎的網紅課程

討論愛莉莎莎的新聞、媒體、報導

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)

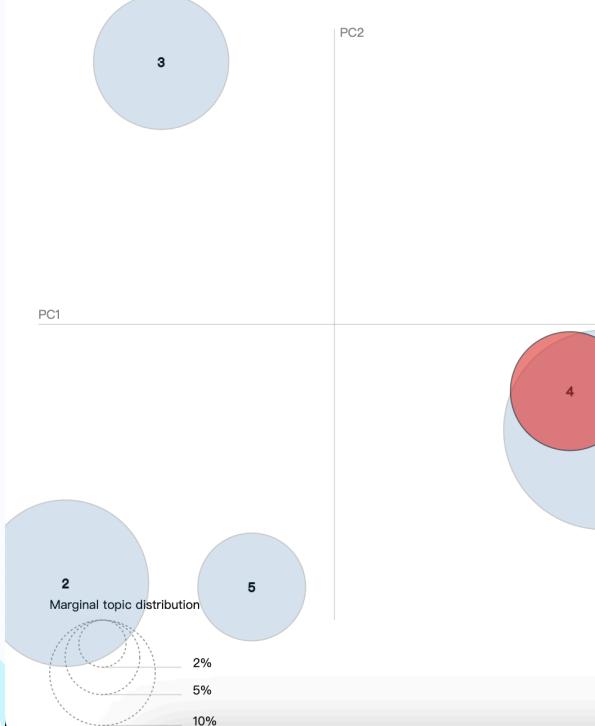
GuidedLDA Vis – Topic 4

(和Topic 1位置重疊性高)

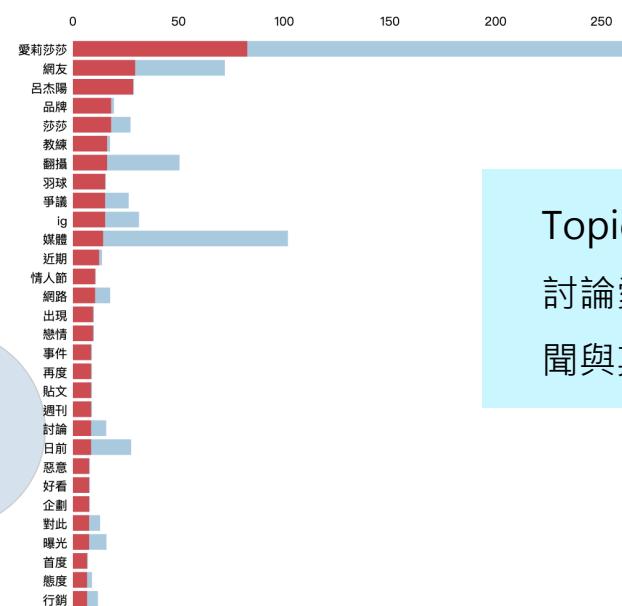
Selected Topic: 0

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

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 4 (12.7% 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)

Topic 4-愛莉莎莎與羽球教練緋聞
討論愛莉莎莎與羽球教練被拍的緋聞與其感情關係

GuidedLDA Vis – Topic 2

Selected Topic: 0

Intertopic Distance Map (via multidimensional scaling)

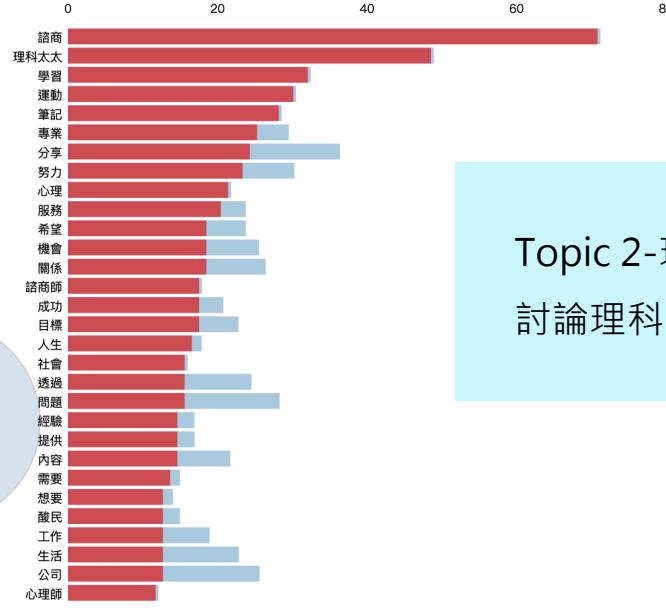


Slide to adjust relevance metric:(2)

$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

Top-30 Most Relevant Terms for Topic 2 (24.8% of tokens)



Topic 2-理科太太的諮詢筆記
討論理科太太的線上課程內容

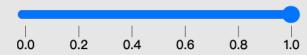
1. saliency(term w) = frequency(w) * $\sum_t p(t | w) * \log(p(t | w) / p(t))$ for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t) / p(w)$; see Sievert & Shirley (2014)

GuidedLDA Vis – Topic 3

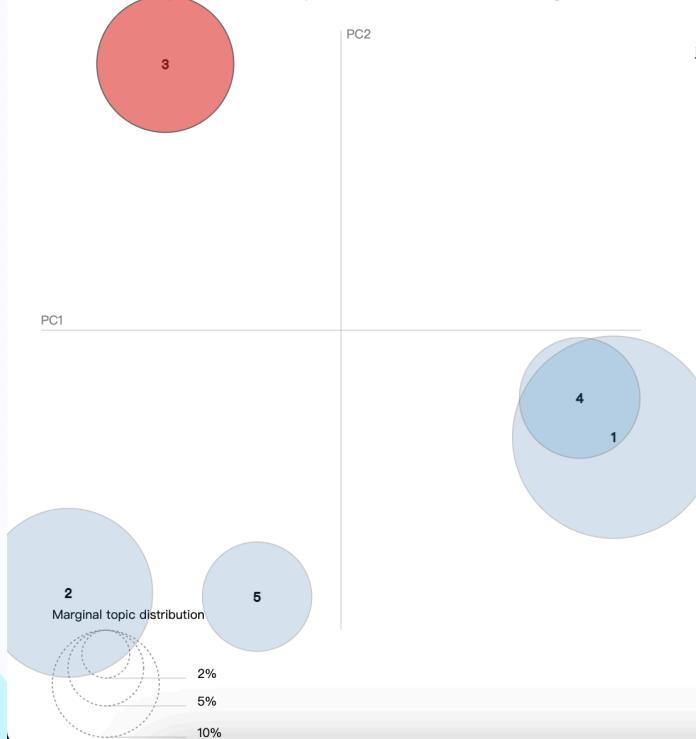
Selected Topic: 0

Slide to adjust relevance metric:(2)

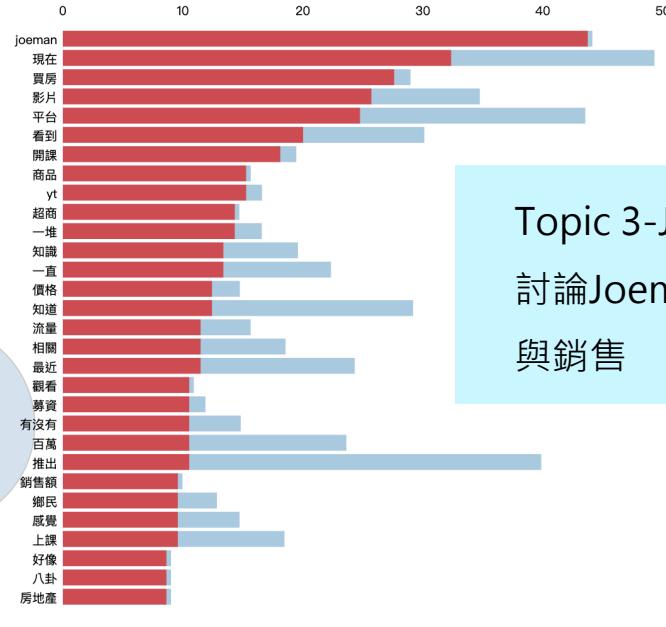
$\lambda = 1$



Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 3 (16.4% of tokens)



Topic 3-Joeman的買房課程
討論Joeman的線上課程進度
與銷售

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)

GuidedLDA Vis – Topic 5

Selected Topic: 0

Slide to adjust relevance metric:(2)

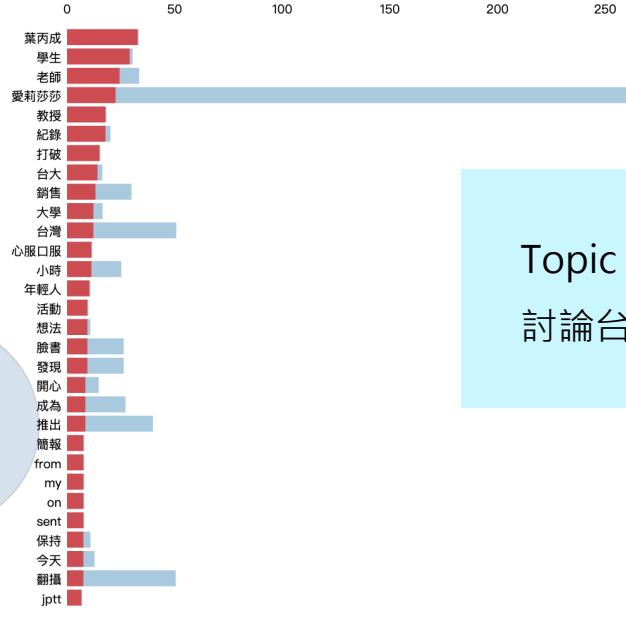
$\lambda = 1$



Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 5 (10.4% of tokens)



Topic 5-葉丙成的簡報課程

討論台大教授葉丙成的簡報課程

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)

五個主題定義

透過GuidedLDA主題模型定義出鮮明的五個主題，匯出其csv，供後續主題本身與其發文、回文者關係分析，包含篇數、情緒與視覺化分析

	org_system_id	理科諮商	莎莎男友	九妹買房	台大簡報	莎莎賺錢
0	1	0.330612	0.004082	0.004082	0.657143	0.004082
1	2	0.558621	0.006897	0.420690	0.006897	0.006897
2	3	0.003390	0.037288	0.952542	0.003390	0.003390
3	4	0.015385	0.169231	0.630769	0.015385	0.169231
4	5	0.476923	0.015385	0.476923	0.015385	0.015385
5	6	0.113514	0.167568	0.708108	0.005405	0.005405
6	7	0.471895	0.001307	0.275817	0.001307	0.249673
7	8	0.006897	0.213793	0.213793	0.351724	0.213793
8	9	0.474510	0.003922	0.356863	0.160784	0.003922
9	10	0.581176	0.000784	0.000784	0.000784	0.416471

Showing 1 to 10 of 109 entries

Previous 1 2 3 4 5 ... 11 Next



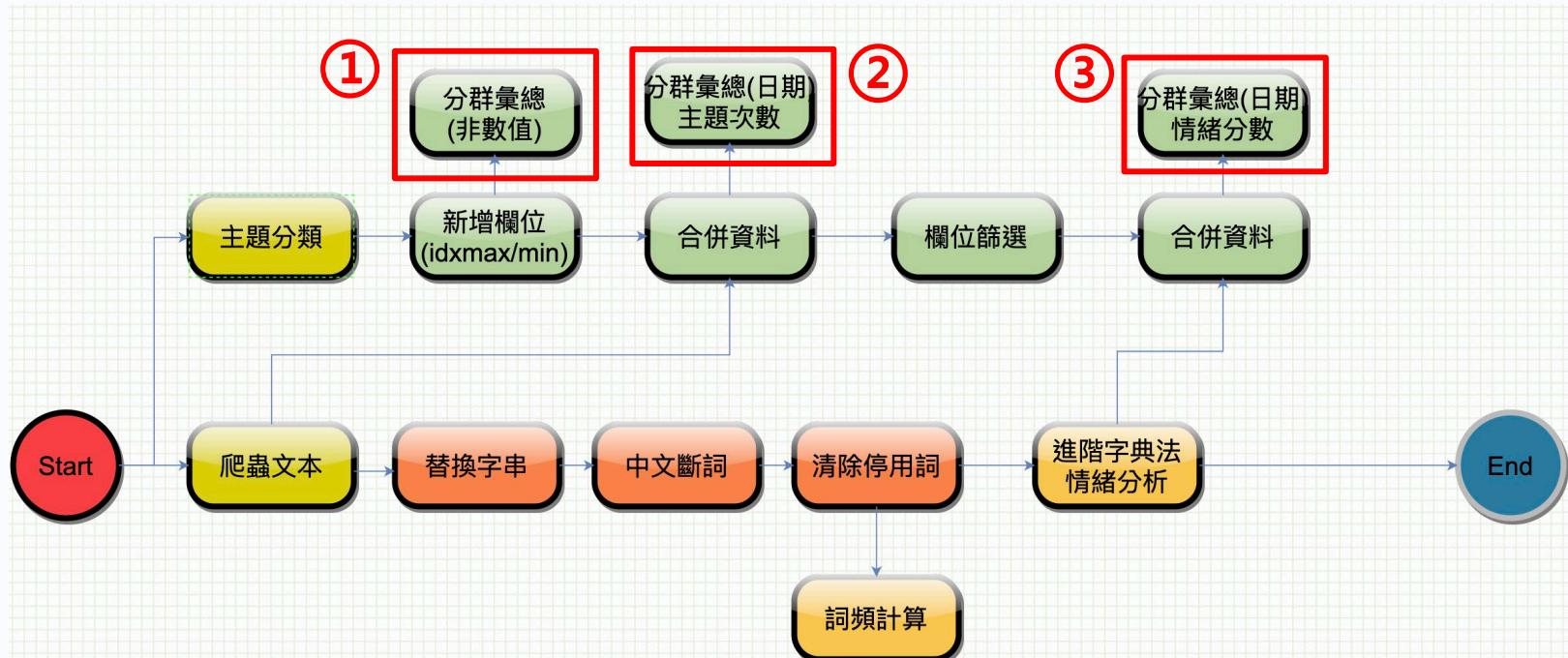
02

分析主題

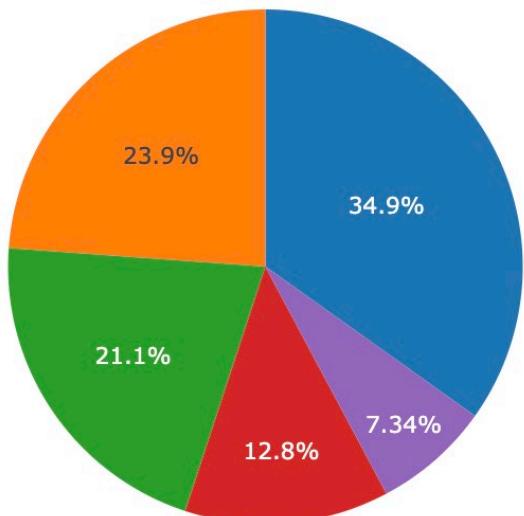
計算主題篇數與情緒並視覺化呈現

實作步驟

上方路徑分別計算五個主題彙總篇數，時間軸上發佈相關的主題次數與情緒，下方路徑為原始資料，供合併資料使用，以及計算字詞情緒。



① 五個主題分別篇數與分佈

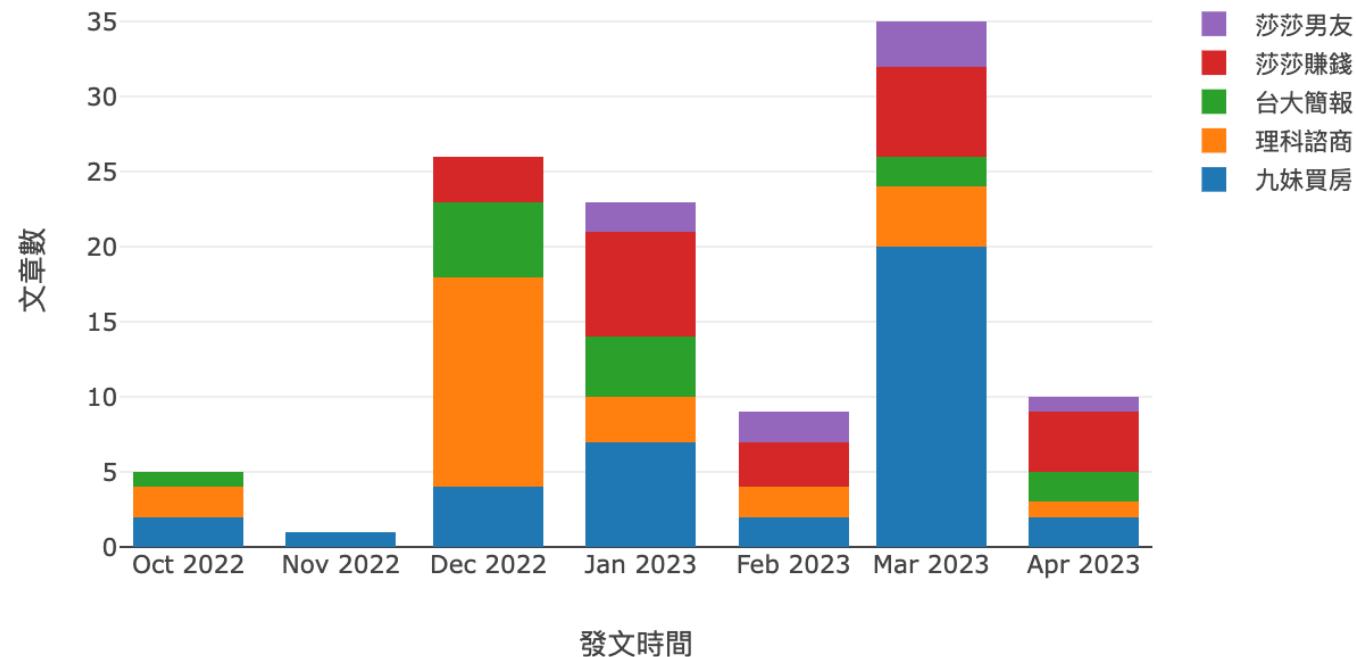


九妹買房
理科諮商
莎莎賺錢
台大簡報
莎莎男友

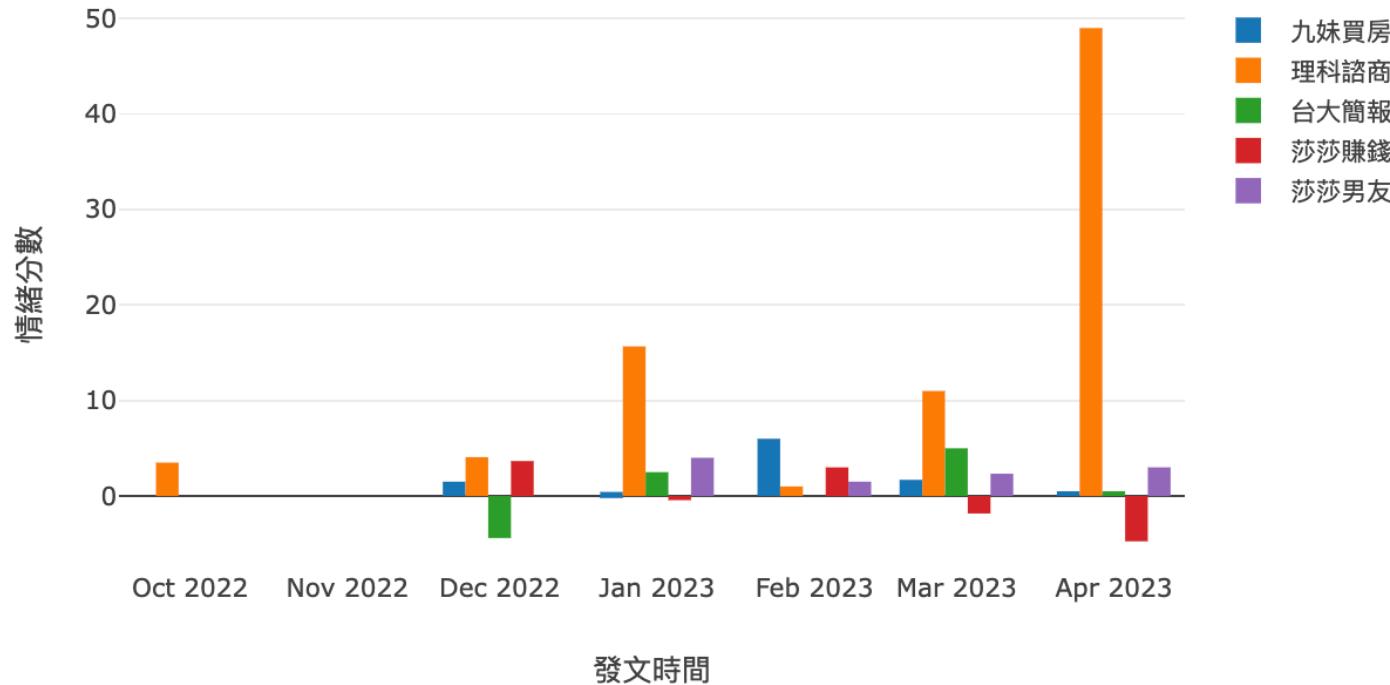
topic	system_id@count
九妹買房	38
台大簡報	14
理科諮商	26
莎莎男友	8
莎莎賺錢	23

Showing 1 to 5 of 5 entries

② 五個主題討論趨勢



③ 五個主題情緒走勢



發文原始資料的字詞情緒分析



77
最大正向情緒



0
最小正向情緒



0
最大負向情緒

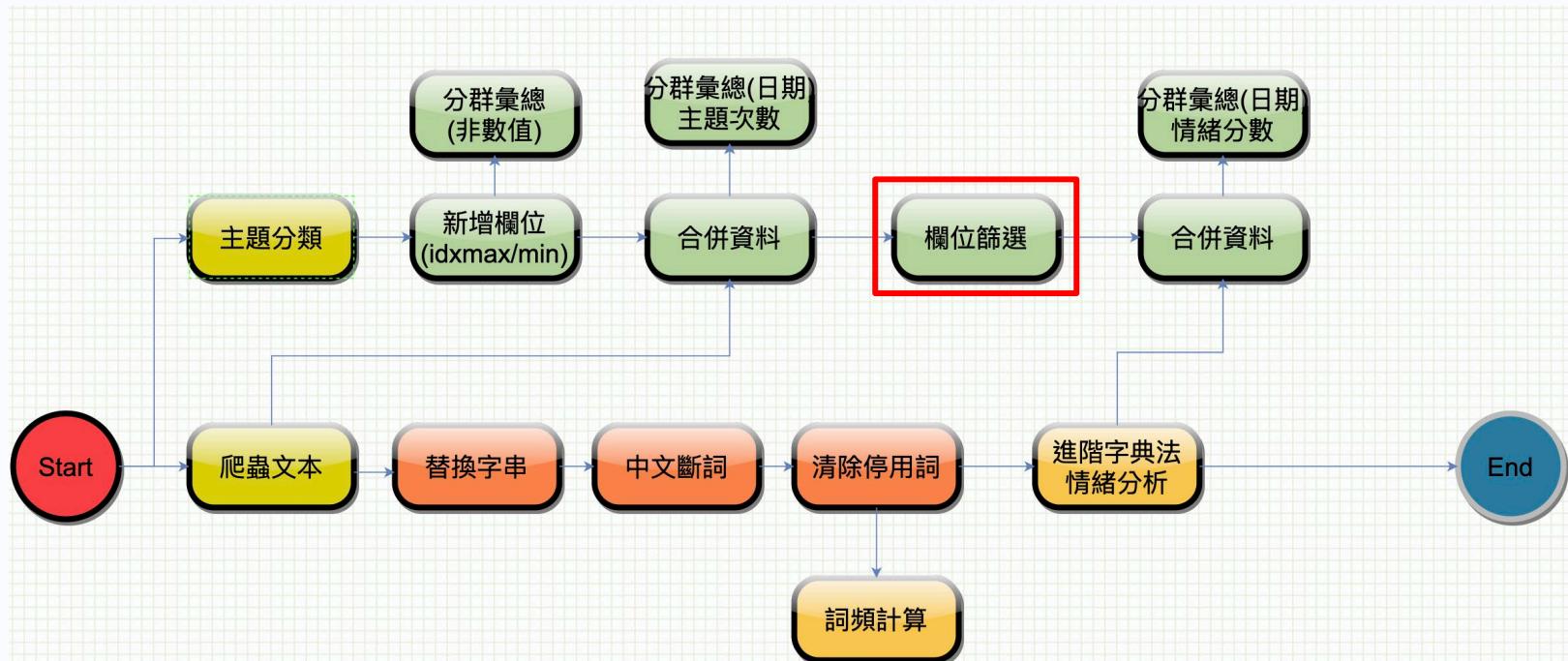


-28
最小負向情緒



實作步驟

匯出所需csv檔，作為繪製發文者社會網路圖的原始資料。



篩選繪圖所需資料

欄位篩選 (140)

參數設定 Input - 128 任務結果

任務結果

Show 10 entries Search:

system_id	topic	artTitle	artDate	artPoster	artCategory
1	台大簡報	Re:[新聞]高虹安爭議燒!領資策會「全新出差」遭疑	2022-10-07 11:55:40	killeryuan	Gossiping
2	理科諮商	Re:[新聞]快訊／兵役延長「4個月要補當兵？」 邱	2022-10-12 13:27:23	MEKURUMEKU	Gossiping
3	九妹買房	[問卦]結婚有什麼好處？	2022-10-13 10:31:36	askaleroux	Gossiping
4	九妹買房	[問卦]半年線上課程要價5萬欸...	2022-10-14 03:04:46	berkeley5566	Gossiping
5	理科諮商	[問卦]線上課程選哪個	2022-10-21 12:27:03	tg9456	Gossiping
6	九妹買房	[問卦]有沒有youtuber很好做的八卦	2022-11-20 00:01:52	ryusai	Gossiping
7	理科諮商	[新聞]廣告「免費英聽課30天」女控：退課「扣近	2022-12-02 12:05:13	pkuguy	Gossiping
8	台大簡報	[問卦]狂賀！愛莉莎莎開線上課程ㄌ！！	2022-12-02 20:06:51	blueskymaple	Gossiping
9	理科諮商	[問卦]理科大大在賣諮商筆記（線上課程）？！	2022-12-10 13:36:05	seand8088503	Gossiping
10	理科諮商	[新聞]心理諮商100hrs！理科太太「諮商筆記」賺	2022-12-10 21:55:03	ipipwrong	Gossiping

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

全螢幕瀏覽 點我下載完整CSV資料 點我下載完整Rdata 點我下載完整json資料



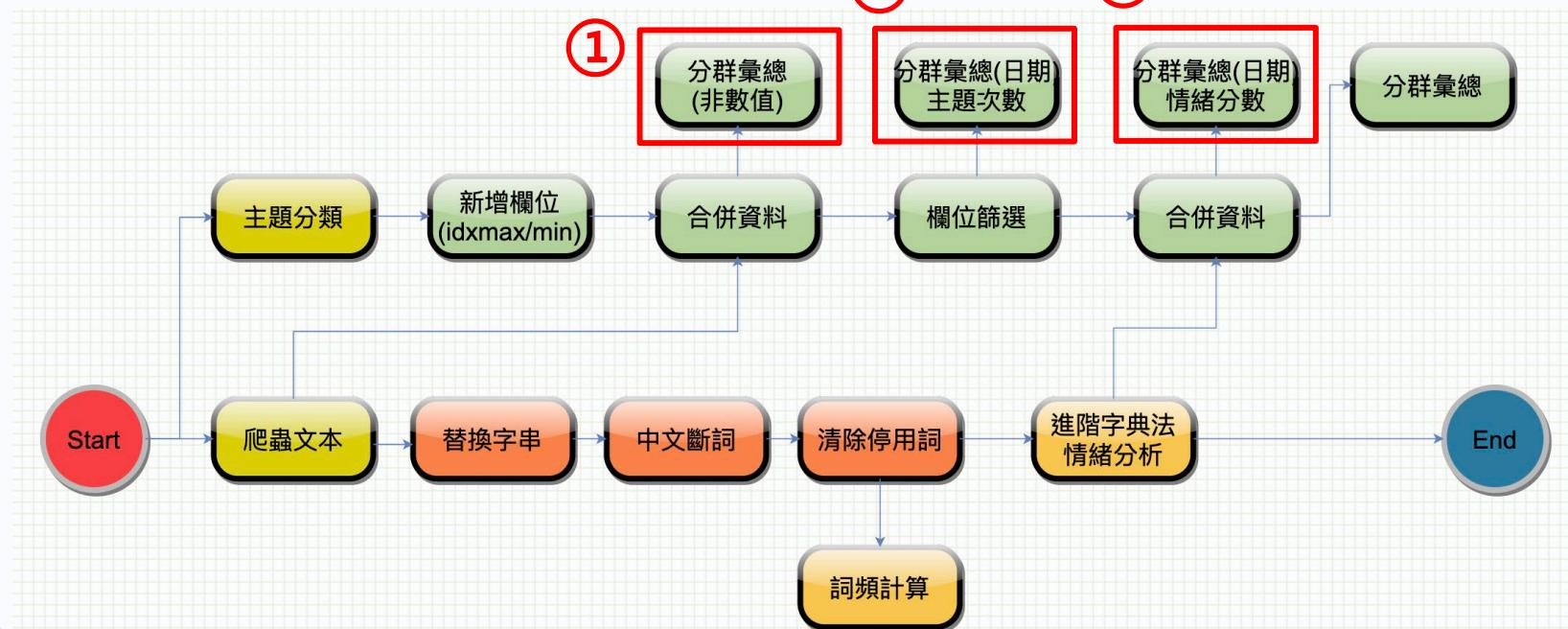
03

定義留言

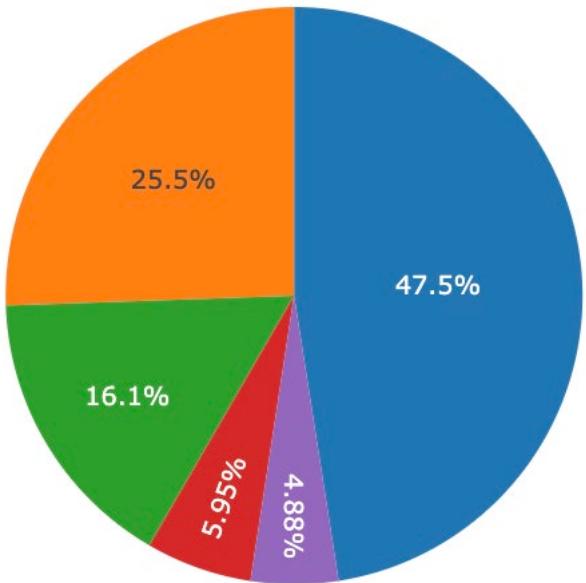
針對主題，計算留言數量與情緒並視覺化呈現

實作步驟

上方路徑分別針對五個主題的留言彙總篇數，時間軸上發佈相關的主題次數與情緒，下方路徑為原始資料，供合併資料使用，以及計算字詞情緒。



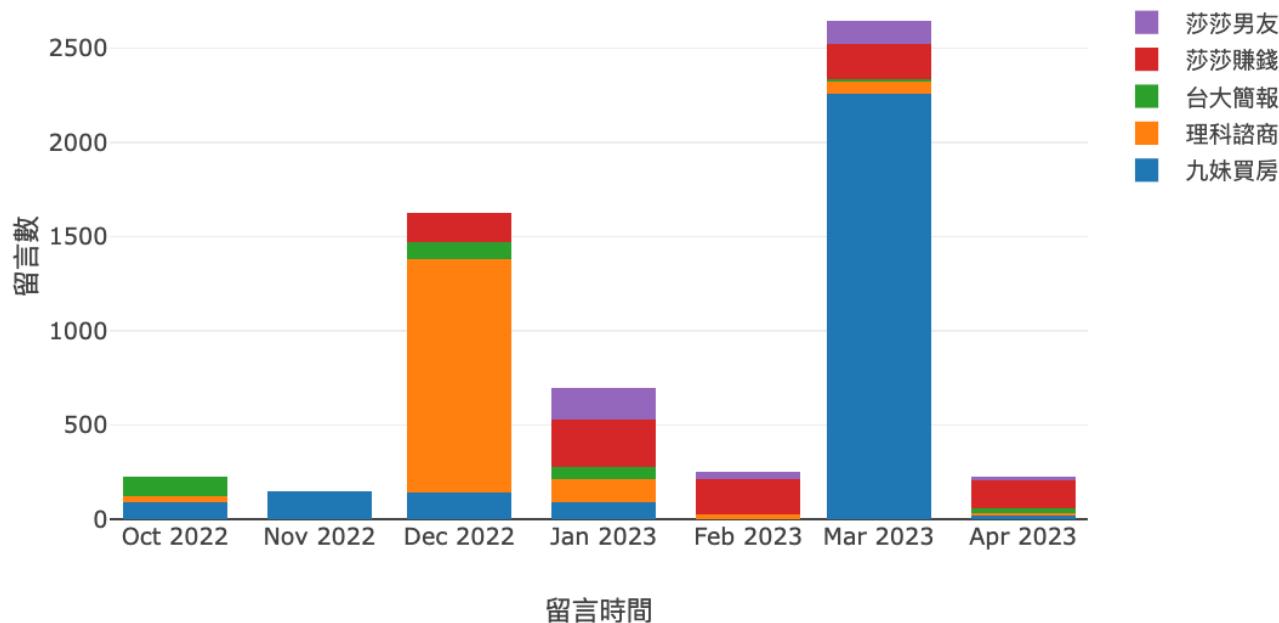
① 五個主題分別留言篇數與分佈



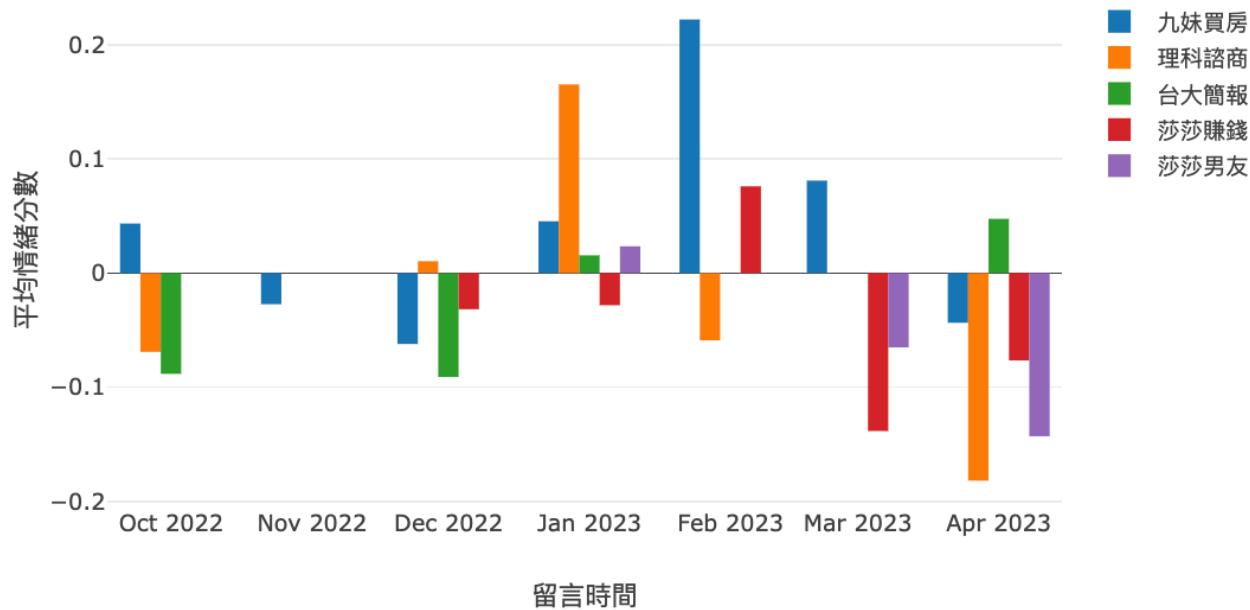
九妹買房
理科諮詢
莎莎賺錢
台大簡報
莎莎男友

topic	system_id_x@count
九妹買房	2765
台大簡報	284
理科諮詢	1486
莎莎男友	346
莎莎賺錢	936

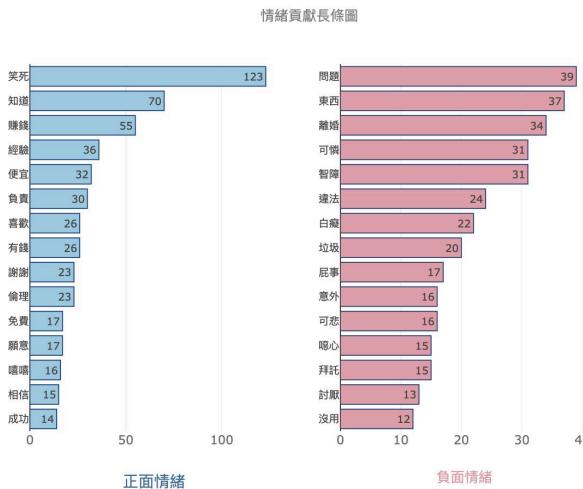
② 五個主題討論趨勢



③ 五個主題情緒走勢



留言原始資料的字詞情緒分析



4

最大正向情緒



0

最小正向情緒



0

最大負向情緒



-3

最小負向情緒





04

社會網路圖

繪製發文者、留言者與主題關係的社會網路圖

實作步驟-發文者

匯入紀錄發文者與主題的csv檔，彙總發文者於各主題的發文數量，作為社會網路圖的資料。

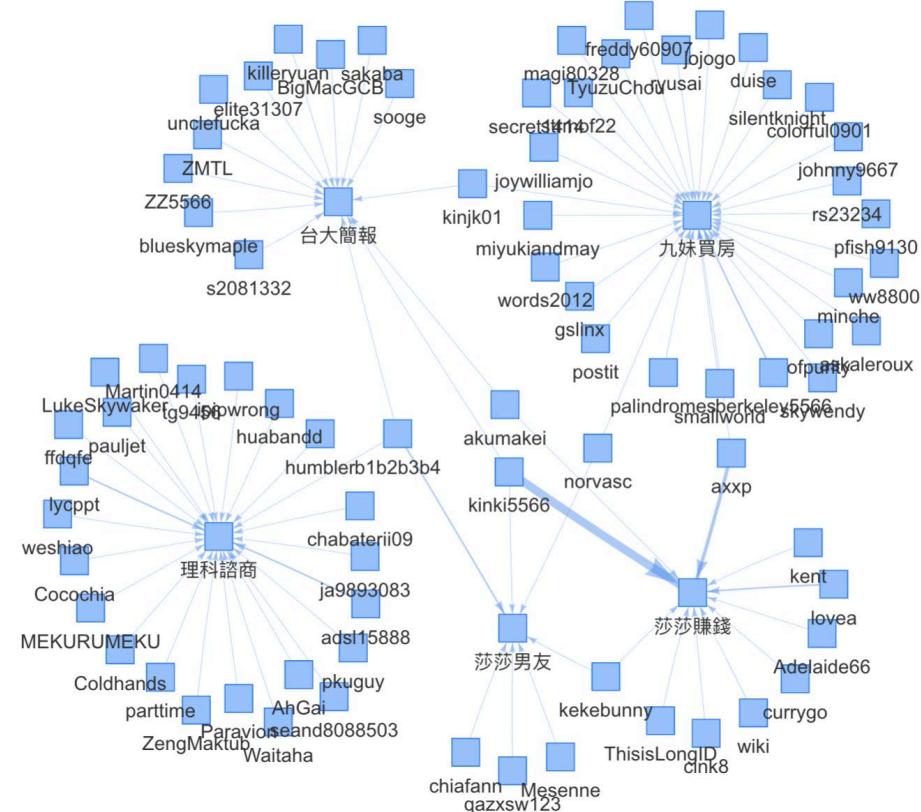


topic	artPoster	system_id@count
九妹買房	AzTiana	1
九妹買房	ForTheHorde	1
九妹買房	Goodinfo	1
九妹買房	Happybuddha	1
九妹買房	LEDG	1
九妹買房	Louismax	1
九妹買房	LukeSkywaker	1
九妹買房	Scorpio	1
九妹買房	TyuzuChou	1
九妹買房	askaleroux	1

Showing 1 to 10 of 93 entries

Previous 1 2 3 4 5 ... 10 Next

發文者與五個主題的社會網路關係



實作步驟-留言者

匯入紀錄留言者與主題的csv檔，彙總留言者於各主題的發文數量，作為社會網路圖的資料。



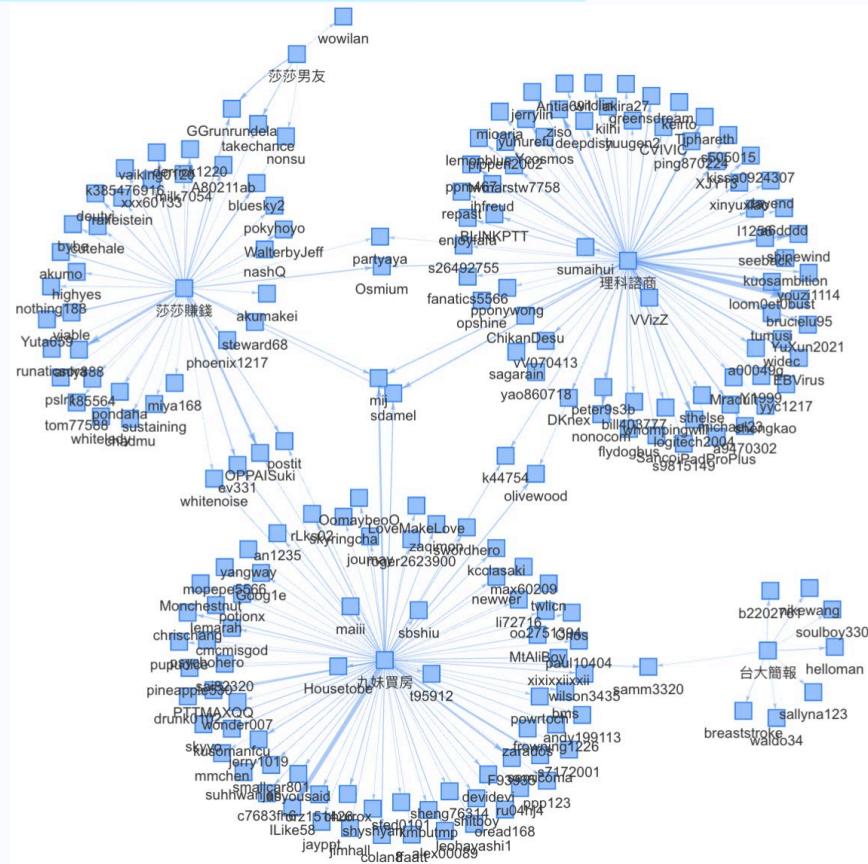
A screenshot of a data preview interface showing a table of comments. The columns are labeled: cmtPoster, topic, and system_id_x@count. The data shows various posters and their counts across different topics.

cmtPoster	topic	system_id_x@count
A0908	九妹買房	2
A0908	台大簡報	2
A555822000	九妹買房	1
A80211ab	九妹買房	3
A80211ab	理科諮商	2
A80211ab	莎莎賺錢	6
Agamemnonk	九妹買房	1
AGATELINK	九妹買房	1
AISC	九妹買房	1
AKATSKI	九妹買房	1

Showing 1 to 10 of 100 entries

Previous 1 2 3 4 5 ... 10 Next

留言者與五個主題的社會網路關係



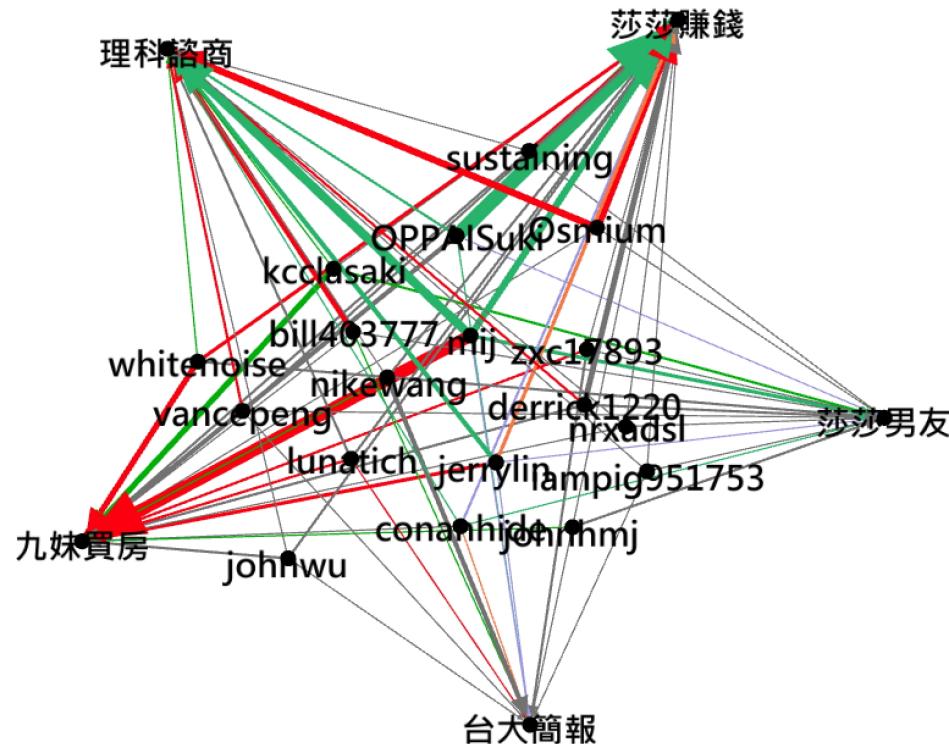
以Gephi繪製社會網路圖

匯出紀錄留言者對主題的回覆數量和情緒。



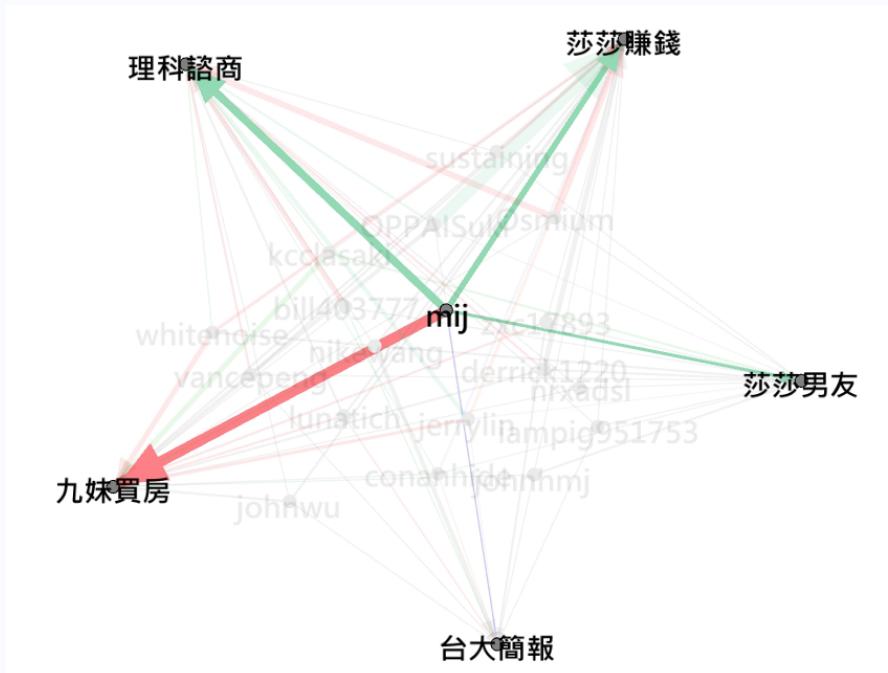
以Gephi繪製社會網路圖

綠色為正向情緒、紅色為負向情緒、灰色為中性情緒。線條粗細代表回應數量。



社會網路圖-特殊發現

以留言者mij為例，他對這五個主題均有回覆，給予「理科太太」、「愛莉莎莎」正面評價，對「葉丙成」為中性看法，唯獨對「Joeman」有較大量的負面評價，推測為Joeman的黑粉。



結語

透過網路評論的分析，可以讓我們使用分析工具了解評論者的真實想法。

有時候針對單一報導或是主題去接受關於此議題資訊時，極有可能會造成一些資訊的誤差及偏頗，

透過大量評論的統計，計算其整體情緒分數，會比單一觀點論述來得更為客觀。

情緒趨勢分析可以了解到這些網紅對網路言論的影響和觀感，

再透過文字雲分析了解每個主題所討論的關鍵字詞和詞頻，了解網路聲量對於這些話題的主要討論內容。

從分析結果可以看出網路聲量中對這些網紅大部分都是正面的情緒居多，

由此可證網紅對於網路的影響力還是具有一定的影響力及正面觀感，

本小組對此現象提供的建議如下：當你成為有一定知名度的網紅時，

請記得開線上課程及多多進行商業合作及聯名，

因為網紅對於大眾還是會有一定程度的吸引力。