

社群媒體分析 期末報告

從PTT論壇分析寵物版的情緒分析
及主題探討

指導老師：黃三益 教授

第十四組

N104020006 曾怡霽

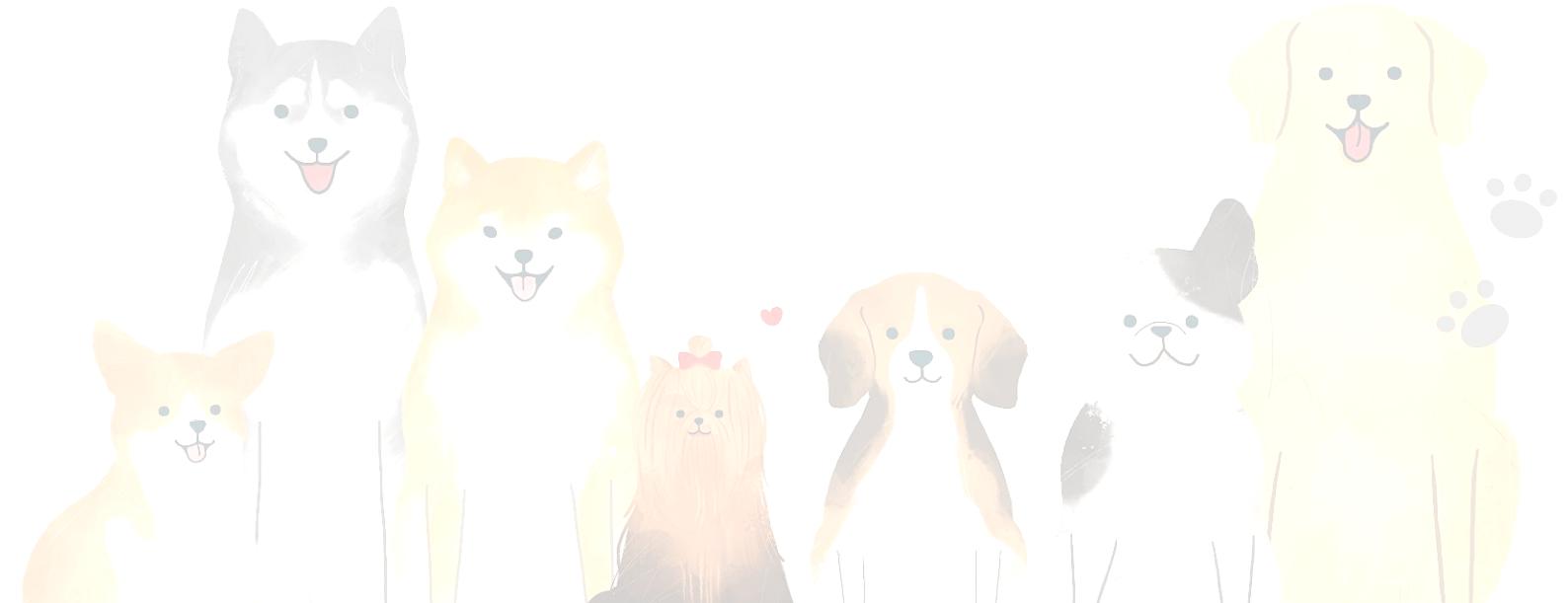
N104020012 蔡京叡

N104020013 吳映儒

N104020030 林欣民

目 錄

- 一、動機和分析目的**
- 二、資料集的描述**
- 三、資料的分析過程**
- 四、視覺化的分析結果與解釋**
- 五、結論**



一、動機與分析目的

1. 分析動機：

在繁忙的時代中，下班或下課後有個寵物陪在身邊是多麼好的一件事，但是如何照顧寵物、要怎麼選適合自己的寵物，這就是一門學問了，因此想藉由社會大眾的建議來了解大家養寵物的心得，哪一類型寵物的討論度最高，透過文字探勘與抓取PTT文章的方式，取得各項寵物版的內容，並將以合併，找尋最熱烈討論的議題。

2. 分析目的：

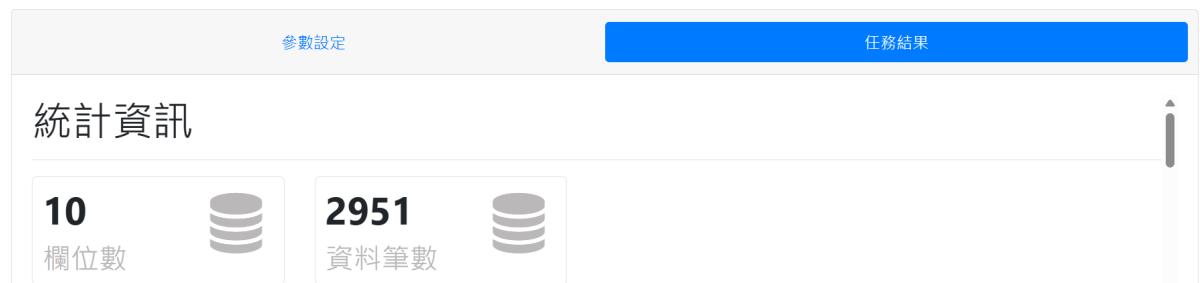
從PTT中抓取寵物相關板，進行資料搜集與分析，觀察民眾對寵物的看法，在工作流程平台中將會使用Lexicon Based情緒分析，了解民眾討論寵物時的正負面情緒以及使用LDA主題模型將主題劃分，了解將在不同主題的寵物版中，透過主題模型能夠區分成額外幾種主題，透過字詞網路圖得知關鍵詞彙之間的連接，並繪製社會網路圖了解寵物相關版的主要發言者與留言者之間的連結。

二、資料集描述

1.文章爬取：

使用工作流程平台爬取PTT論壇中2023/03/01 ~ 2023/05/31 之間的文章，範圍涵蓋PTT各個寵物專版，包含了「水族」、「鳥」、「貓」、「狗」、「特殊寵物」、「天竺鼠」、「倉鼠」、「寄居蟹」、「兔」、「兩棲爬蟲」等10個專版，總計共2951筆資料。

PTT 爬蟲 (4)



2.留言萃取：

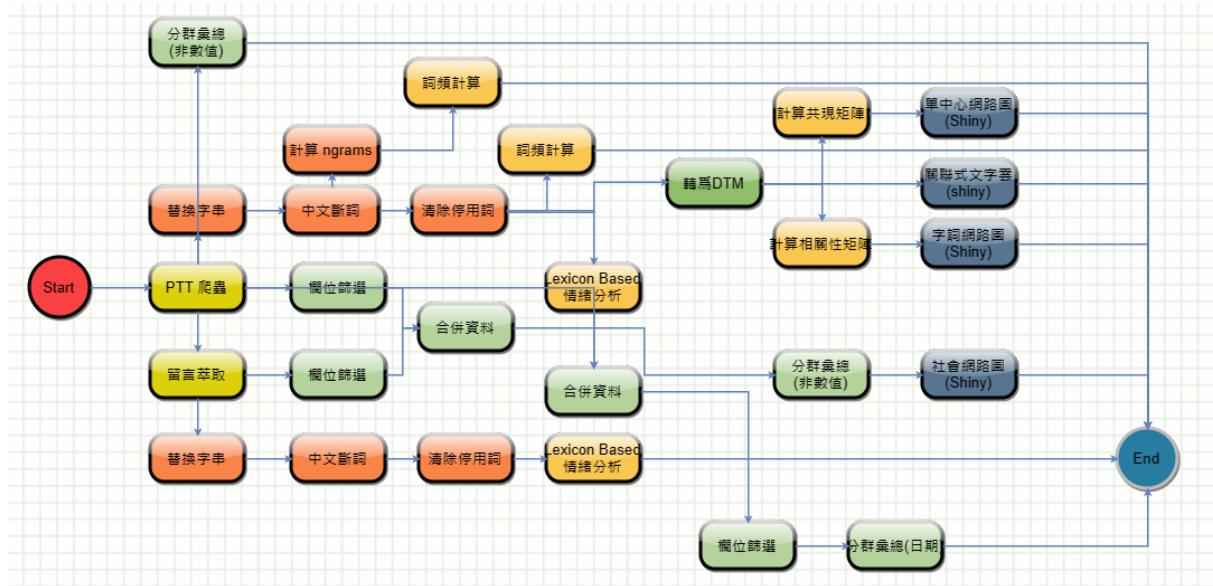
將PTT上各個寵物專版所爬取到的2951篇文章進行留言萃取，總計35465筆留言，平均單篇留言數21.48筆。

留言萃取 (7)



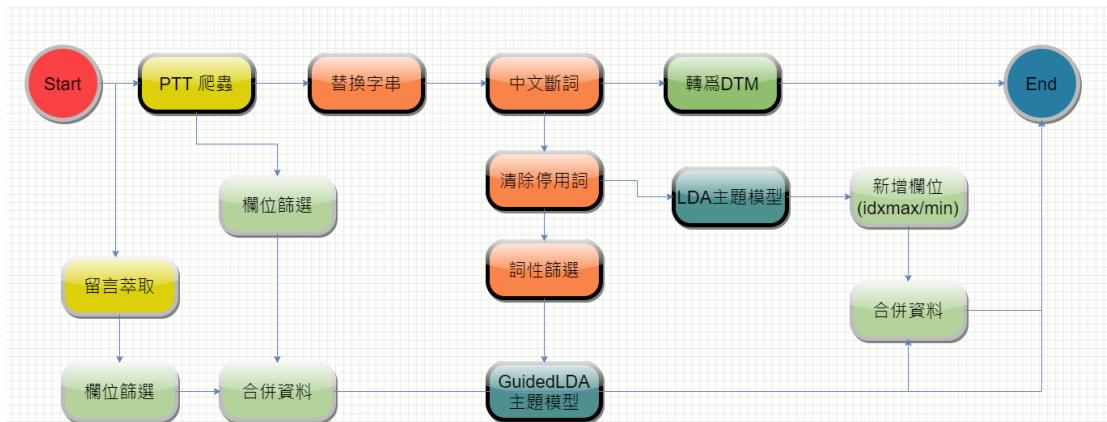
三、資料分析過程

1.情緒分析部分使用工作流程平台TarFlow工作流程:寵物版



爬取PTT論壇文章後，先去除換行符號、網址、html標籤、底線、多餘的空格，再進行中文斷詞、清除可能干擾分析結果的停用字，並運用「計算ngrams」以及「詞頻計算」功能，反覆觀察，進行斷詞及停用字處理。將文章轉換成DTM，繪製出單中心網路圖、關聯式文字雲、字詞網路圖，亦透過文章、留言的發文者及留言者，繪製社會網路圖。再透過Lexicon情緒分析取得每一篇文章及留言的情緒值，進行後續的相關分析。

2. 主題模型部分使用工作流程平台TarFlow工作流程:LDA



先將自行設定的主題csv檔案匯入至檔案管理作為字典，設定的主題有：「水族」、「鳥」、「貓」、「狗」、「特殊寵物」、「天竺鼠」、「倉鼠」、「寄居蟹」、「兔」、「兩棲爬蟲」。

工作流程平台爬取 PTT「水族」、「鳥」、「貓」、「狗」、「特殊寵物」、「天竺鼠」、「倉鼠」、「寄居蟹」、「兔」、「兩棲爬蟲」共10個專版後共2951筆資料，執行資料清理步驟：「替換字串」、「中文斷詞」、「清除停用字」後，執行 LDA 主題模型與

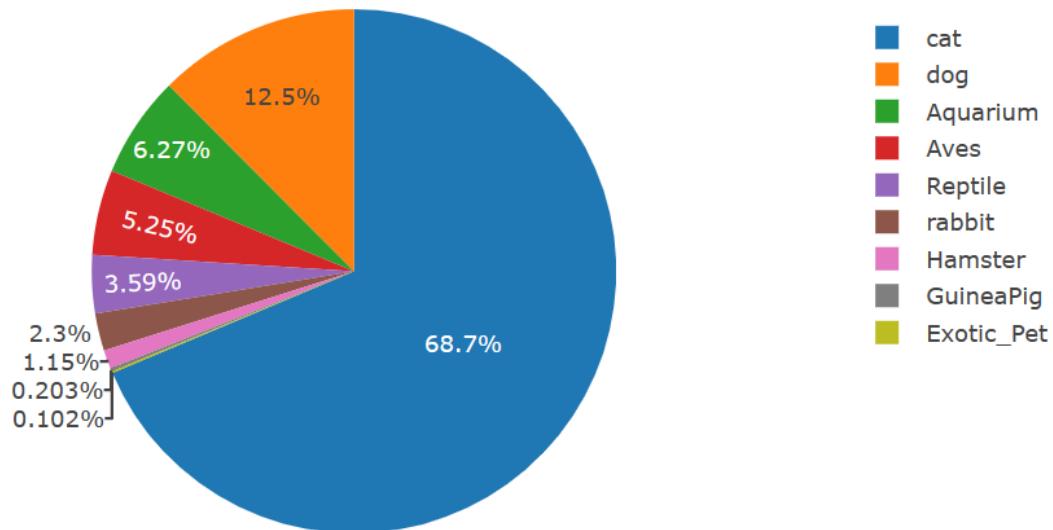
GuidedLDA。Gephi資料來源為上述PTT爬蟲資料使用留言萃取篩選欄位與PTT爬蟲資料合併後再與LDA主題模型的主題資料合併所匯出的CSV檔案。



四、視覺化的分析結果與解釋

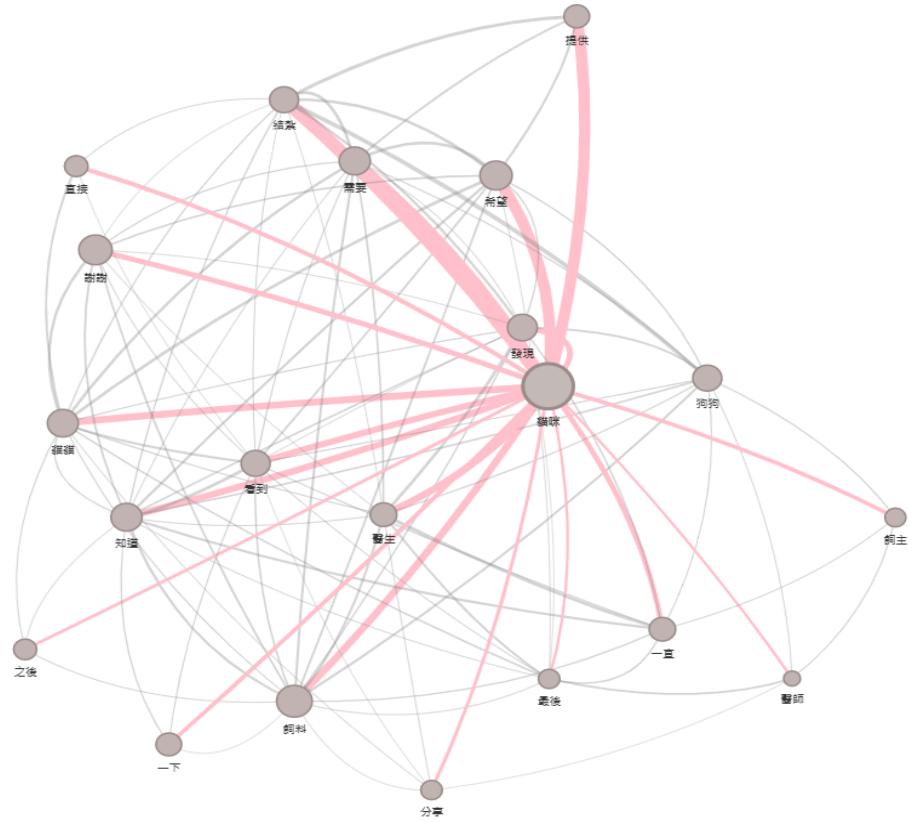
1.情緒分析

PTT寵物版各版討論次數比率圖

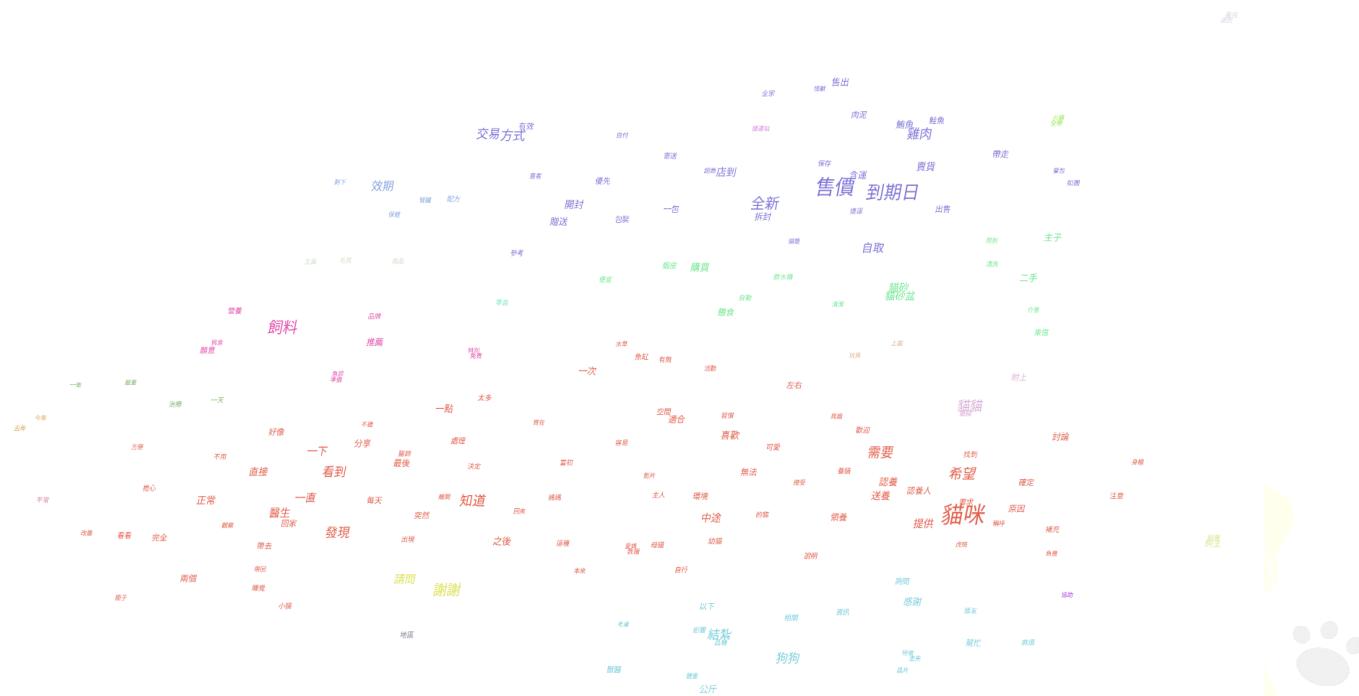


從PTT寵物版的比率圖可以看出，討論貓的文章為最多也是最主要的，其次為狗、水族、鳥、兩棲爬蟲、兔、倉鼠、天竺鼠、特殊寵物等。

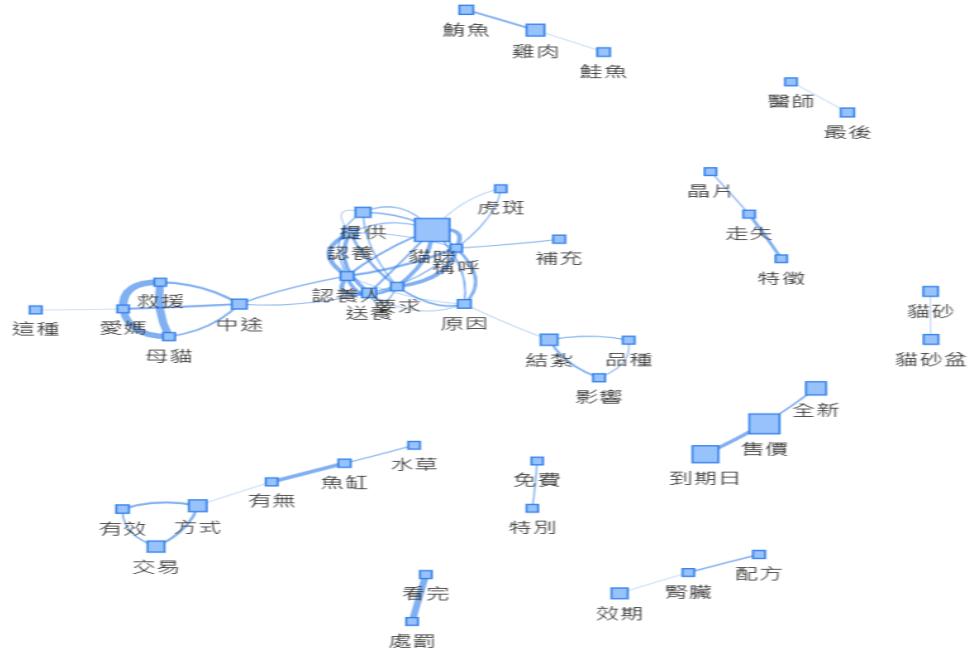




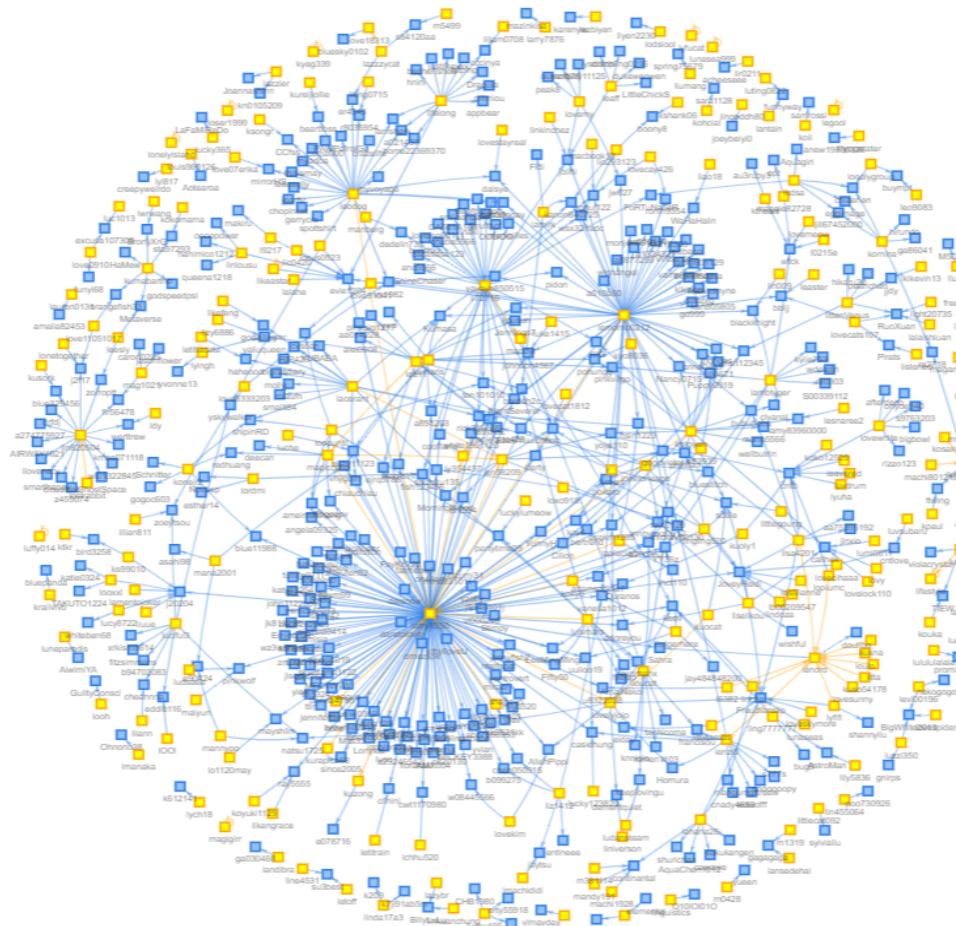
所有寵物版文章的單中心網路圖，以貓咪為主要的討論中心，衍生出相關議題。



所有寵物版文章的關聯式文字雲，主要可以分成貓咪、物品、飼料、醫療等有關聯的區域。



所有寵物版文章的字詞網路圖，也可以看出以貓咪為主發展出的字詞為最主要的。



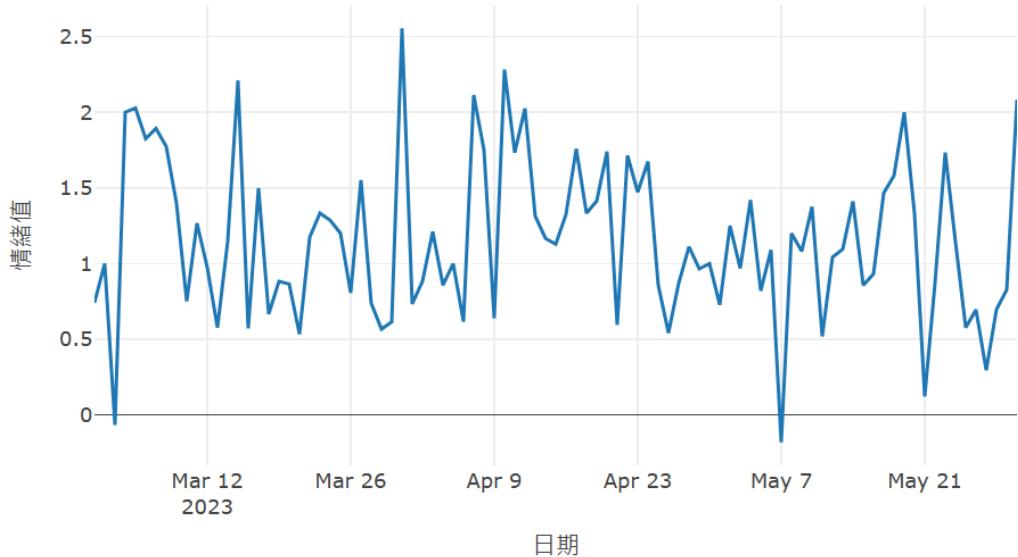
所有寵物版文章和留言，依據發文者及留言主所繪製出的社會網路圖，可以看出主要發文者為larry850、leomomo0212、leodog、lin820504等人。

視覺化



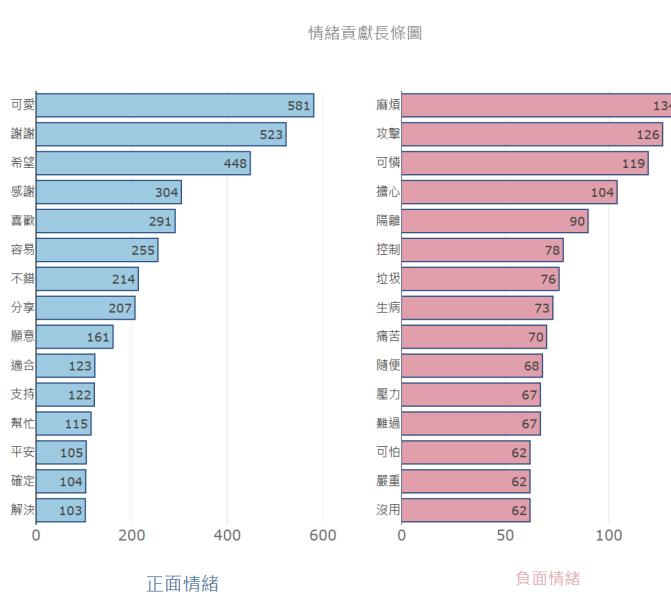
所有寵物版文章Lexicon Based的情緒分析，正面詞彙有謝謝、希望、喜歡、感謝、分享等，負面詞彙有擔心、麻煩、處罰、嚴重、負擔等，而且也觀察到正面情緒字數量遠大於負面情緒字數量。

總體情緒平均值分析圖



所有寵物版文章的情緒平均圖，可以看出民眾在討論寵物相關的議題，是比較偏向正面的。

視覺化

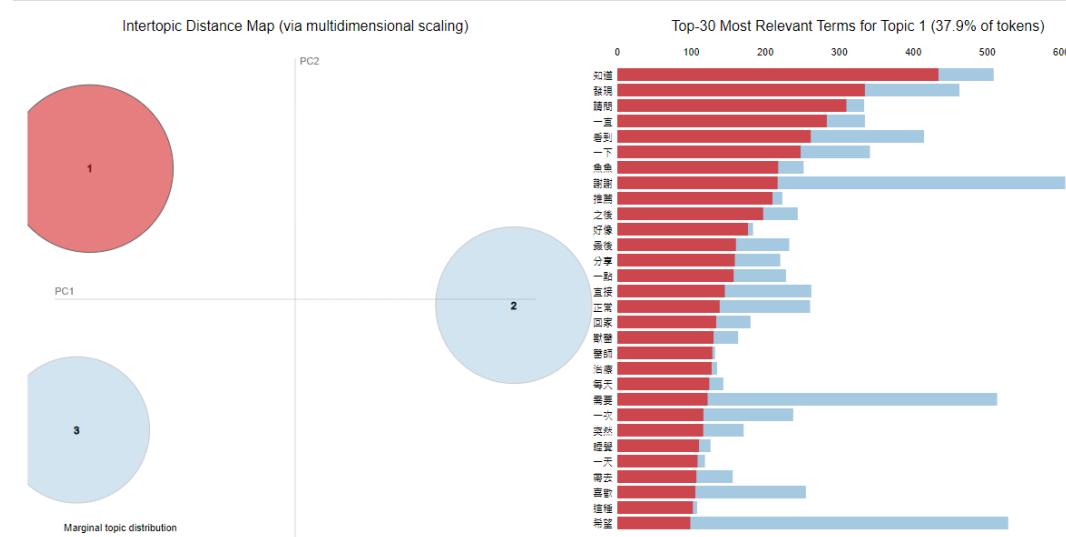


另外所有寵物版文章的留言Lexicon Based的情緒分析，正面詞彙有可愛、謝謝、希望、喜歡、容易等，負面詞彙有麻煩、攻擊、可憐、擔心、隔離等，留言正面情緒字數量也遠大於負面情緒字數量。

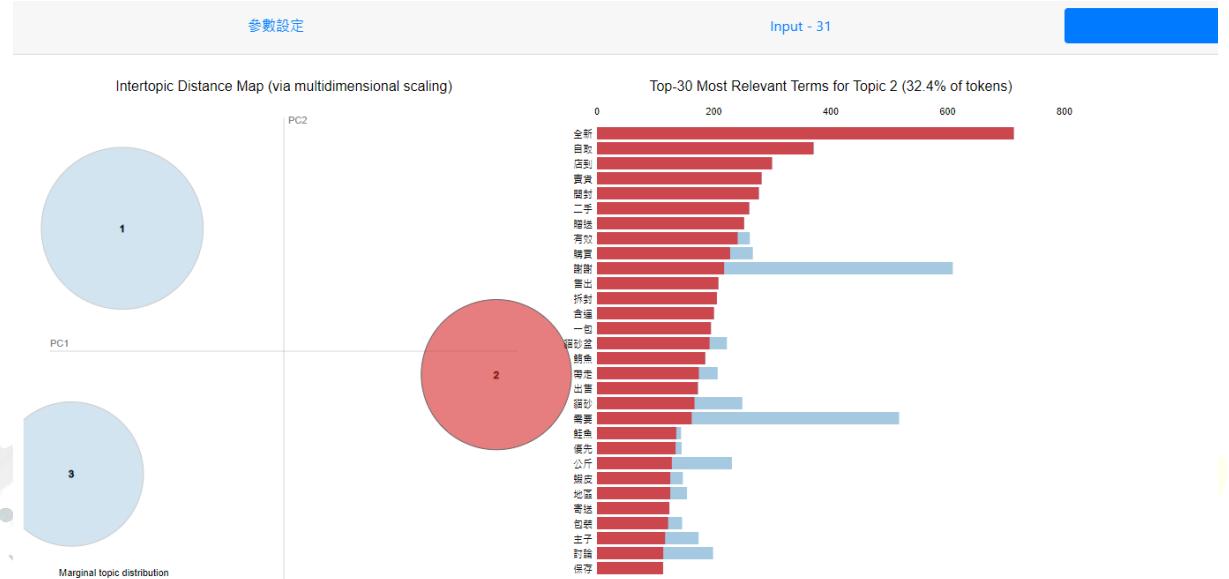


2.LDA

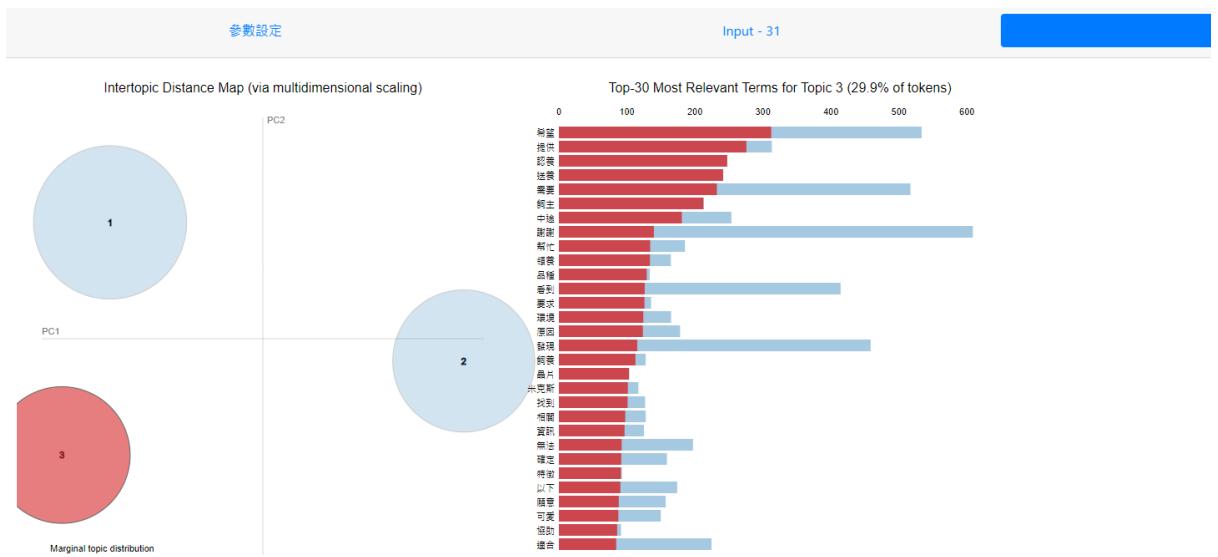
LDA 主題模型參數設定：詞彙頻率下限設定為20；詞彙頻率上限為0.3；主題數設定為3個來觀察其中的關聯性。



LDA主題1可以看到辭彙有獸醫、醫師、治療、希望、回家、請問等，可以推測該主題可能被歸類在寵物醫療相關的分類。



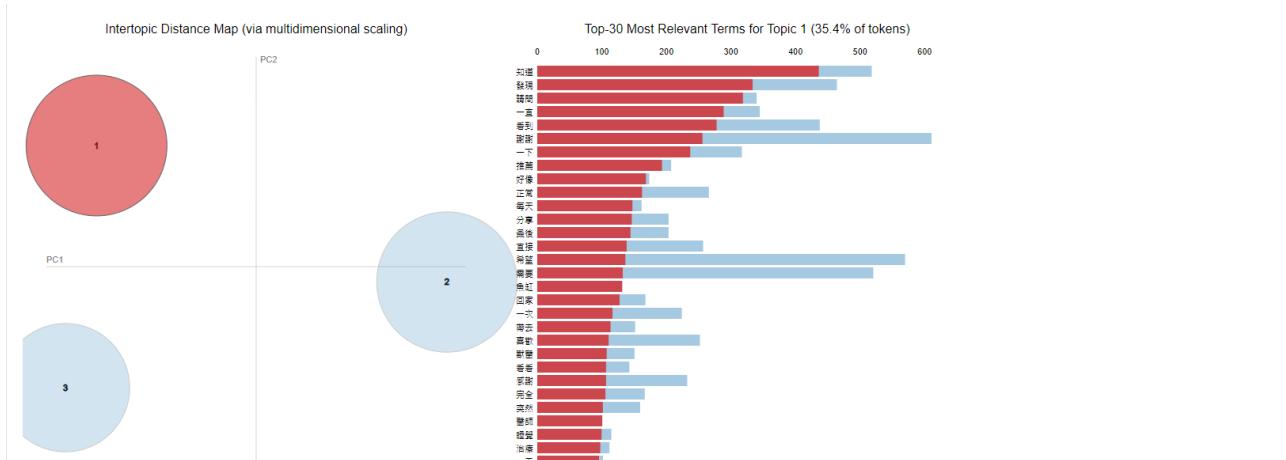
LDA主題2可以看到辭彙有自取、全新、二手、店到、貓砂、售出、蝦皮等，可推測為是寵物用品買賣相關的主題。



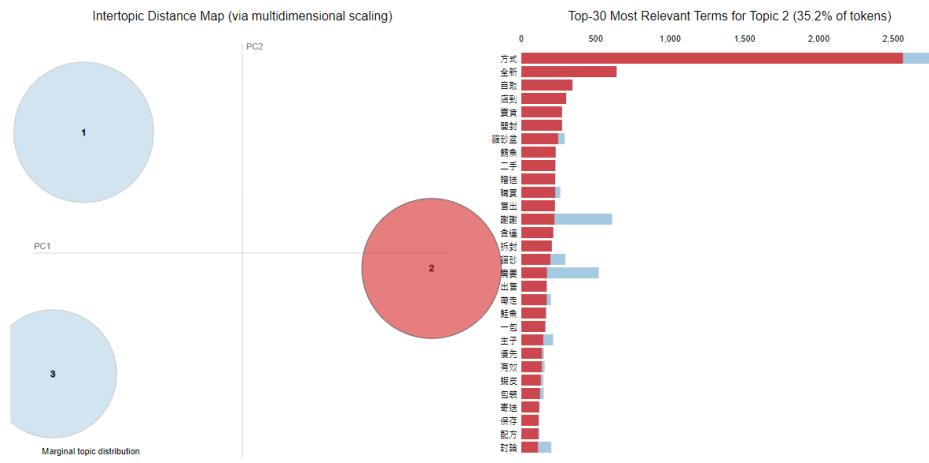
LDA主題3可以看到辭彙有希望、認養、送養、需要、飼主、飼養、協助、願意等，可推測為是寵物領養、寵物送養等相關的主題。

3. GuidedLDA

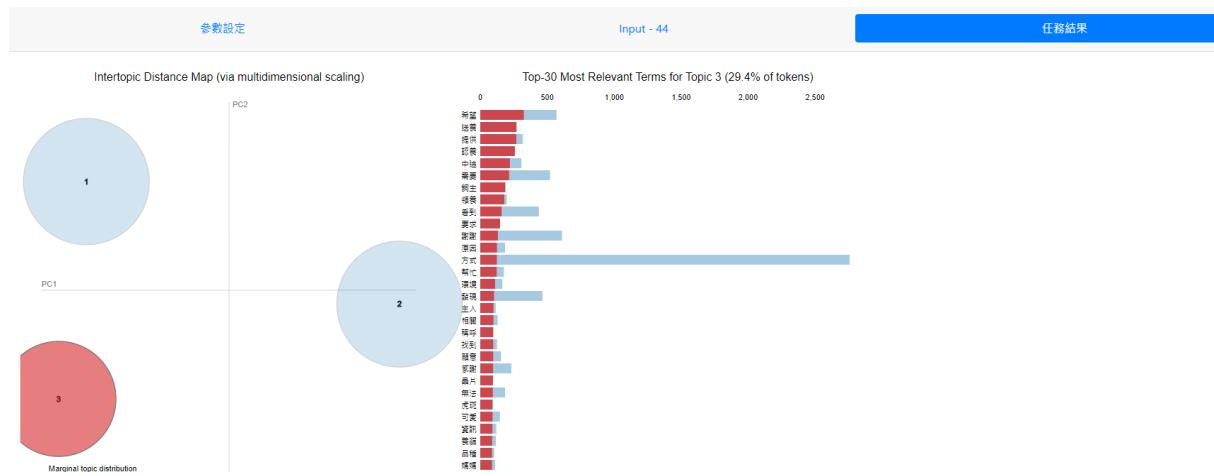
GuidedLDA 將種子字區分為PTT文章數較多的狗與貓兩個主題。詞彙頻率下限設定為15; 詞彙頻率上限為0.7; 主題數設定為3個來觀察其中的關聯性。



GuidedLDA主題1可以看到辭彙有獸醫、醫師、治療、睡覺、發現、請問、知道等，可以推測該主題可能被歸類在寵物醫療相關與寵物醫療諮詢的分類。



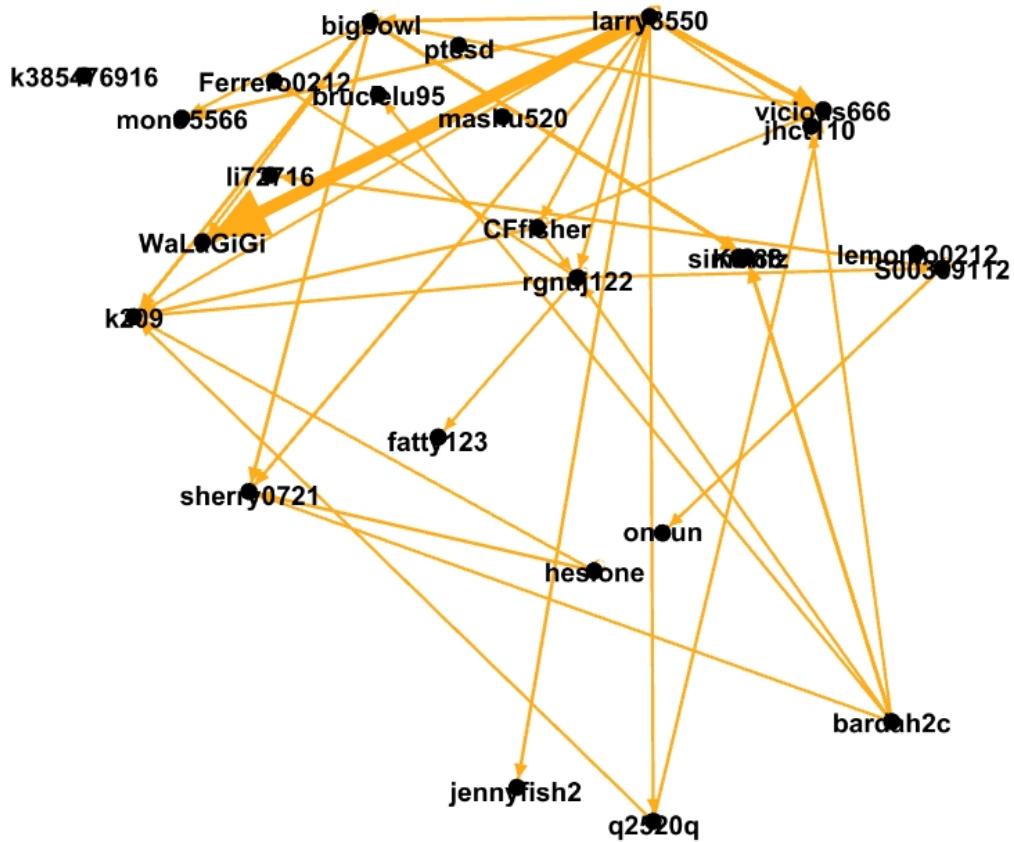
GuidedLDA主題2可以看到辭彙有自取、全新、二手、店到、貓砂、贈送、蝦皮等，可推測為是寵物用品買賣相關的主題。



LDA主題3可以看到辭彙有希望、認養、送養、提供、飼主、領養、找到、主人等，可推測為是寵物領養、寵物送養等相關的主題。

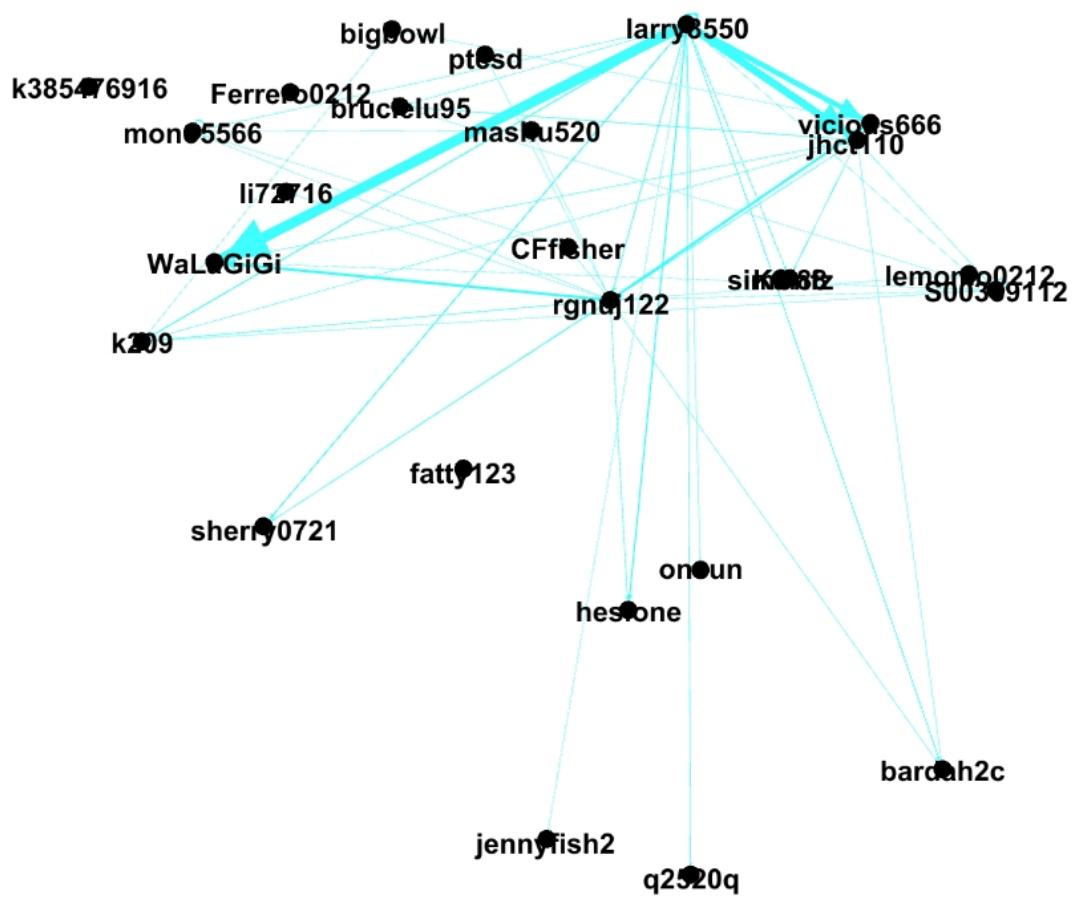


4. Gephi



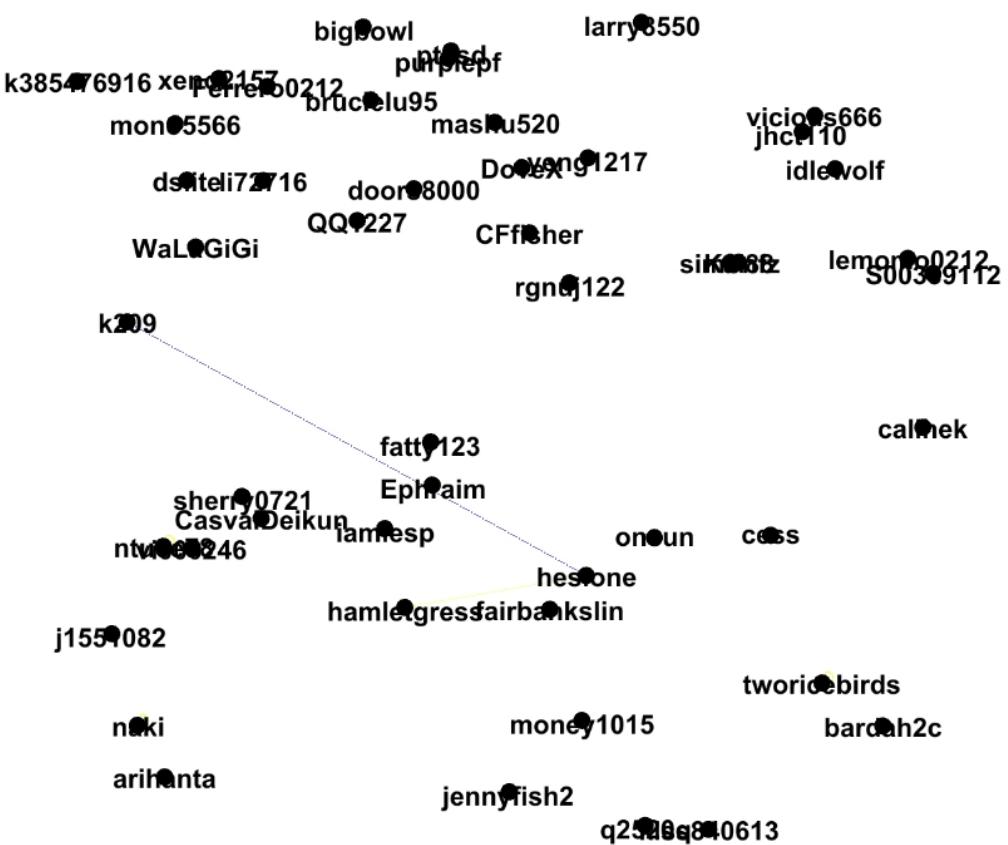
這是關於主題1寵物醫療相關的主題，可以發現larry6550與bardah2c經常回覆相關的主題文章，其中larry6550最常回覆WaLGiGi的文章，可推測larry6550與bardah2c對於寵物醫療有一定程度的了解與認知。



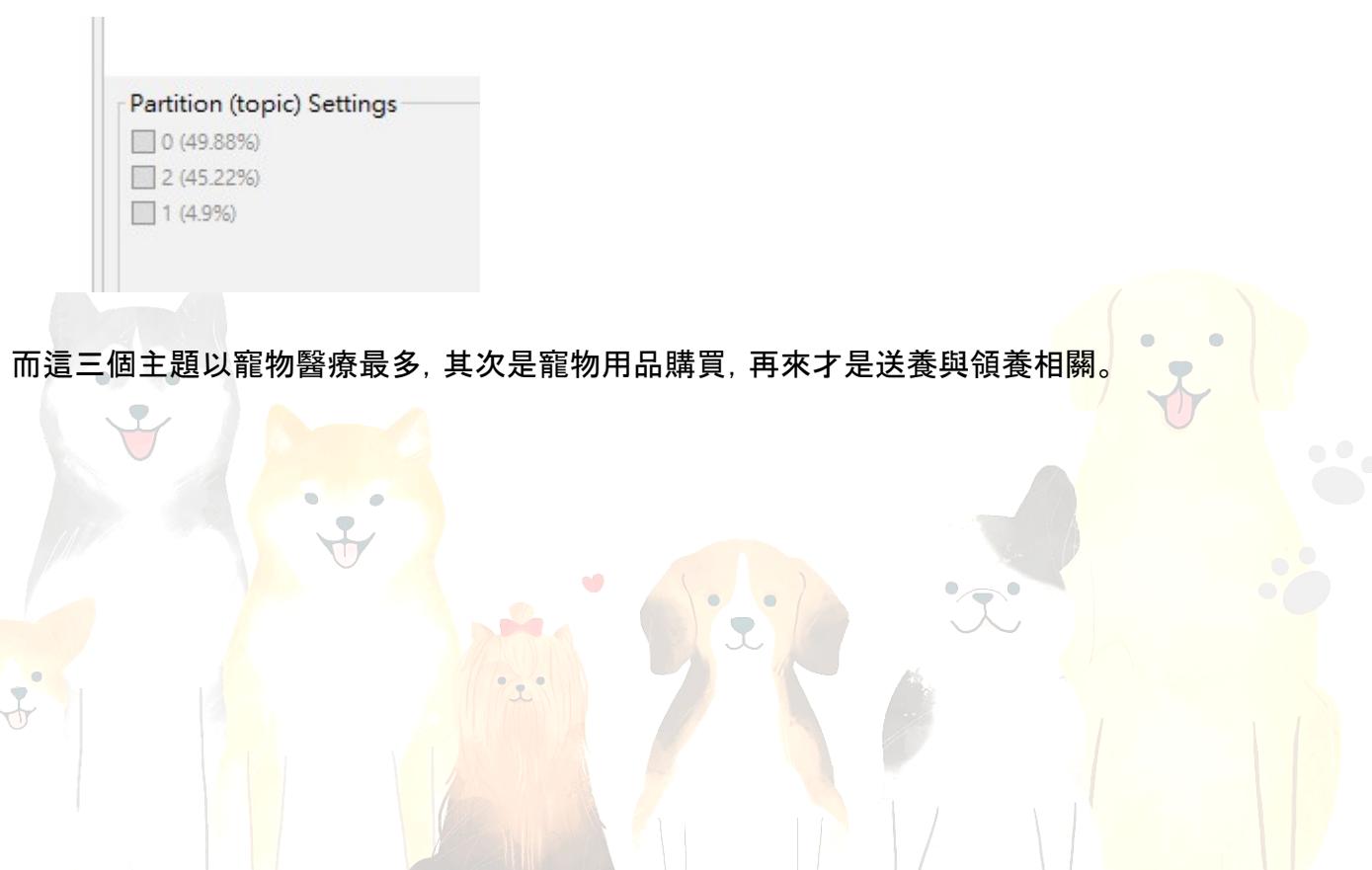


這是關於主題2寵物用品買賣的主題，可以發現larry6550與rgnuj122經常回覆相關的主題文章，其中最常回覆WaLGiGi的文章，可推測larry6550與rgnuj122也對於寵物用品買賣有一定程度的了解。





這是關於主題3寵物領養與認養相關的主題，可以發現hesone經常回覆相關的主題文章，其中最常回覆k209的文章，可推現hesone也對於寵物領養與認養有一定程度的了解。



五、結論

(一)情緒分析

從PTT的寵物版來看，民眾討論度最多的是在貓版，推論應當是許多人對貓有特別的喜爱，而且貓通常被視為可愛、有趣和有特殊性格的寵物，吸引了許多人在討論版上分享和討論有關貓的內容。另外貓在網路文化中也佔有重要的地位，許多網路迷因和內容都與貓有關，促使更多人在討論版上討論貓。

在進行情緒分析時，發現不論在文章或留言上，正面情緒詞彙數量皆大於負面情緒詞彙，推論應該是因為寵物具有正面影響，特別是狗和貓，通常會帶給人們正面的感受，可以提供陪伴、降低壓力、增強社交互動等，而且許多寵物討論版的版主或發文者希望創建一個積極、支持和友好的討論版，因此會傾向於分享正面的經驗。另外也有許多討論可圍繞著寵物的可愛行為、成功的訓練策略或是有用的飼養資訊，這些討論也是可以引發更多的正面評論。

但是，這並不代表著寵物討論版上的所有討論都是正面的，有些負面的詞彙可能反映出飼養寵物的挑戰和困難，包括寵物的健康問題、行為問題或飼養成本等，而這些討論提供民眾可以尋求幫助、分享經驗和學習如何更好地照顧他們的寵物。

(二)LDA主題分類分析：

1.LDA

LDA主題一：寵物醫療諮詢與醫療相關的主題。

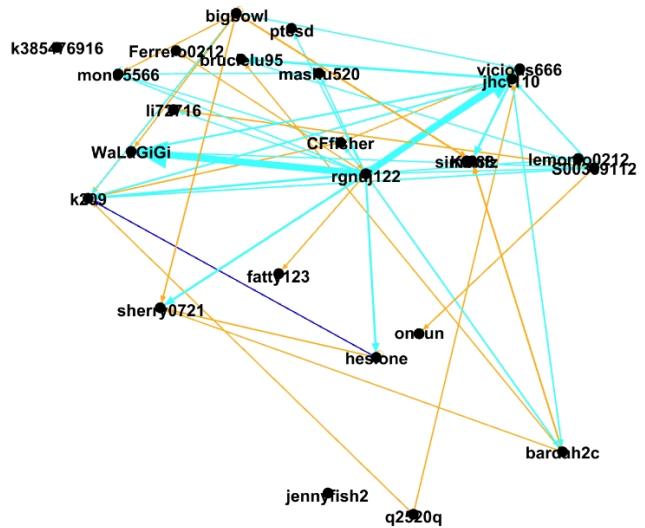
LDA主題二：寵物用品買賣相關的主題。

LDA主題三：寵物領養與送養相關的主題。

原本以為主題模型會將狗與貓及其他拆成三種主題，但結果顯示主題為醫療相關、用品買賣與領養相關等主題，故可能可以猜測無論何種寵物，大家都較關注這三種主題，並關健字詞會也都很相近



2.Gephi:



Gephi三種主題的留言圖，其中最主要的意見領袖為:rgnaj122，最與大家有串聯，而再來是bigbowl與bardah2c，可推測這三位PTT帳號可能有養寵物且可能對寵物的知識有一定程度的了解。

報告影片連結:<https://www.youtube.com/watch?v=C0MaO5f2uGc>

