

資料科學與社會研究 期末報告

經研二 R05323042 蘇敬倫
經研二 R05323044 陳論陞
經研二 R05323055 劉得顯
經濟四 B03303070 李雅雯
農經三 B04204025 鄭曉今

1 研究主題

- 八卦板是否有特定族群的存在？

1.1 問題敘述與初步想法

本研究的宗旨是爲了理解社群網絡的使用者，在社群平台上是否會有特殊的習性，而這些具有相似習性的使用者是否會因互動頻繁而形成社群網絡的特定族群？如果會，應該如何界定這些特定族群？而這些特定族群又有怎麼樣的特性呢？

由於對於這些問題並沒有一個特定的假設或想法，本研究將透過資料探索的方式，先利用演算法來定位每個社群網絡的使用者，再分析每位使用者的相對關係，以找出社群網絡中潛在的特定族群。若是能夠順利將使用者分群，透過觀察每個族群參與討論（推文）的文章分類，將能定義族群的興趣類別，而透過更細微的統計分析，能夠將不同類別的族群所帶有的使用者特性做簡單的描述。

爲了避免研究範圍過於分散，必須有一個較明確的研究標的。本研究選擇以 PTT 八卦版當作研究的目標，因爲八卦版具有幾個不錯的優點：(1) 八卦版的發文數與參與人數（版友）皆有一定的規模，不會有難以取得資料或樣本點過少的問題。(2) 八卦版並不具有特定的主題，任何的主題都可以在該版上討論，因此較容易吸引不同種類的族群發文或參與討論。(3) PTT 的資料爲公開資料，但 ID 具有匿名性，使用者較容易跟隨自己真實的喜好在該版上活動。雖然八卦版還是有一些缺點，像是只有會使用 bbs 的人才會出現在該版，所以並不具有整體社會的代表性，但由於多數的社群媒體都有類似的問題，基於八卦版的優點，我們仍選擇該版爲我們的研究標的。

1.2 問題的重要性

由於網際網路的發達，人與人之間的溝通管道、情感聯繫，到政治角力、商業行爲都出現了巨大的變革。從近來數次臺灣的選舉、川普當選到英國脫歐等重要事件，社群網絡的影響可說是舉足輕重，有許多人嘗試透過分析社群網絡的風向、聲量及族群間關聯關係來預測事件的結果，或是透過分析得知社群網絡的結構，以確定影響事件的關鍵樞紐，並從中施力。除此之外，許多的公司也嘗試對網路使用者的行爲進行分析，適當地將網路使用者分群，從中找出公司的目標客群，以達到提升公司業績與利潤的目的。

從上述例子便可得知，研究網路使用者的行為，並依使用者的特性分群，在現今社會中已經非常普及且重要。再加上未來網路普及度、參與度，還有透過網路能進行的活動種類，只會愈來愈多，所以這方面的研究也將更加的重要。因此，本組選擇將對網路使用者進行行為研究與分群當作這次期末研究的主題。

2 研究方法

2.1 資料變數

本研究所使用的變數包含 PTT 八卦版中每篇文章的：

- 標題
- 發文者 ID
- 推文者 ID
- 文章網址

本研究最終使用了 2017/12/06 2018/01/01 共 3000 頁網頁範圍的文章，其中有效文章總數為 59572 篇，活躍人數共計 93670 人。¹

2.2 資料分析步驟大綱

1. 利用爬蟲取得上述資料。
2. 將資料整理成特定的串列形式放入 word2vec 套件中進行轉換，獲得每個使用者 ID 的向量。
3. 將代表使用者 ID 的向量放入 Kmeans 套件中，進行分群，獲得八卦版使用者的分群。
4. 觀察各分群使用者熱烈討論的文章，找出各分群可能的興趣類別為何。
5. 利用集中度、參與比例等特性來分析各分群可能的特徵為何。

2.3 資料分析的方法

2.3.1 資料前處理

為了定位各個八卦版使用者，我們將 PTT 八卦版 2017/12/06 2018/01/01 期間的所有文章的 html 檔案透過 python 的 requests 套件爬下來，接著拆解文章標題、發文者 ID 和推文者 ID，並創造一個大的 python 的字典變數，將每篇文章網址設為鍵 (key)，並將另一個小的字典變數設為該鍵的值 (value)，這個較小的字典包含的元素為發文者 ID、文章標題及包含所有該文推文者 ID 的串列變數。為了使得之後能夠將發文者 ID 與推文者 ID 之後能放入 Word2vec 的套件中分析，將發文者 ID 與推文者 ID 以下列方式組成一個新的串列變數：

[..., 推文者 9ID, ..., 推文者 1ID, 發文者 ID, 推文者 2ID, ..., 推文者 10ID, ...]

完成以上串列的前處理之後，將該串列放入 word2vec 套件進行文字轉換成向量的工作。

¹活躍人數是指在這段期間有發文或推文至少 1 次的人員總數。

2.3.2 利用 Word2vec 套件將文字轉換成向量

Word2vec 為一群用來將單詞轉換成向量的模型²，這些模型為雙層的神經網路，用來建構詞與詞之間的關係，本模型使用 Skip-gram 來建立神經詞的嵌入，基本構想就是用輸入去預測兩旁的單詞，跟本次想取得的結果相當，Skip-gram 把一個詞從詞窗剔除。在 skip-grams 下給定 n 詞圍繞著詞 w ，Word2vec 預測一個句子中其中一個缺漏的詞 c ，即以機率 $p(c|w)$ 來表示。相反地，CBOW 給定詞窗中的文本，預測當前的詞 $p(w|c)$ 。

當模型訓練完成後，Word2vec 模型可以將每一個單詞都投射到多維空間，這邊的維度可以自己決定，我們選定 25 維，當單詞都投射到多維空間以後，我們可以計算兩點的距離得知單詞與單詞之間的關係以及關係是否緊密。

Word2vec 通常依賴 Skip-gram 和 CBOW 兩種方式來建立嵌入。Word2vec 為 Tomas Mikolov 在 Google 帶領的研究團隊創造，於 2013 年發表論文：Distributed Representations of Words and Phrases and their Compositionality。

在這個部分，Word2vec 將每一個發文者及推文者的 ID 轉換為具有 25 個維度的向量，本文給予它一個新的名字：ID2vec，如此一來，每一個發發文者及推文者的 ID 就如同一個有 25 個維度特性的個體，我們可以利用 K-平均演算法 (Kmeans) 將所有的 ID 進行分群。值得注意的是，為了確保每個使用者的資料有足夠的出現次數，來判定該使用者的行為，只將出現次數大於 5 次的使用者 ID 放入 Word2vec，最後剩下 52056 筆使用者 ID。

2.3.3 利用 Kmeans 套件將 ID 進行分群

Kmeans 則是關於分群聚類的演算法³，能將高維空間中的樣本分成 K 群，並找出 K 個中心點，然後計算每一個樣本點跟哪個中心點距離最近，並歸類在那群，之後再重新計算一次組內的中心點，藉此不斷的去更新其中心點的位置，達到差異最小的分群模型。

這個部分，我們參考平方誤差和 (Sum of the Squared Errors, SSE) 隨組數的變化，發現 SSE 在 10 組的時候有一個較大的下降，所以決定使用 Kmeans 演算法將所有的樣本分為 10 組，再從中找出具有特色的組別進行研究與分析。

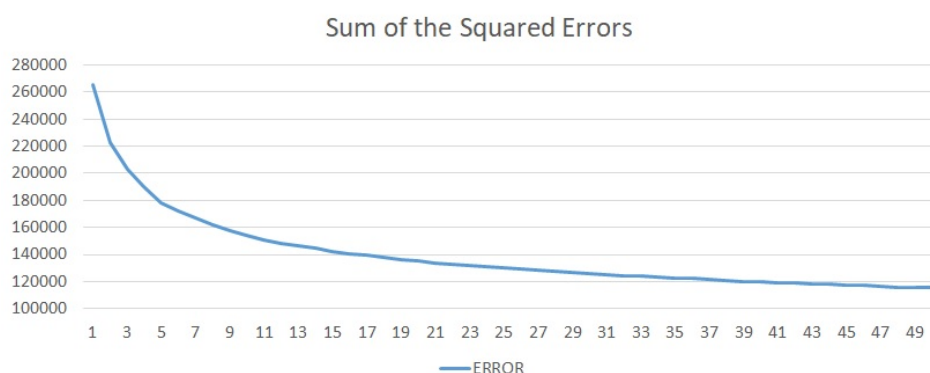


圖 1: 平方誤差和 (Sum of the Squared Errors, SSE) 隨組數的變化

²Word2vec 的資料參考維基百科：<https://zh.wikipedia.org/zh-tw/Word2vec>

³Kmeans 的資料參考維基百科：<https://zh.wikipedia.org/wiki/K-平均算法>

2.3.4 判斷各分群的興趣類別

這個部分的分析，是透過觀察各分群成員熱烈討論的文章，研究各個分群的成員對於哪類議題的文章較感興趣。這裡原本是要使用 Jieba 套件對文章標題進行斷詞，再統計出現最多的語詞來決定各分群的興趣類別，不過由於技術上遇到一些問題，這次就先以人工判斷，由本小組成員依投票取得各分群興趣類別的共識。

2.3.5 各分群成員的參與集中度的計算與呈現

各分群成員的參與集中度計算方式如下：

1. 先計算每一篇文章中的推文者 ID 有多少是屬於該分群的。
2. 將計算出來的 ID 數除以該分群的總人數，即可得知該分群有多少比例的人參與該文的討論。
3. 取出比例高於 0.1 的文章，繪出 X 軸為參與比例，Y 軸為該比例區間累積文章數的直方圖。

$$\text{參與比例} = \frac{\text{該篇文章屬該分群的推文 ID 數}}{\text{該分群成員數}} \quad (1)$$

透過觀察每一個分群的參與集中度，可以理解一個分群的成員在討論的參與上，是經常集中於某些文章中有熱烈的參與，還是該分群的成員很少聚集於同一篇文章中討論，而只是鬆散地參與討論。透過結合參與集中度與分群的興趣類別，或許能夠推論出該分群成員的特色與行為模式。

2.3.6 各分群成員的參與比例散布圖

繪製參與比例散布圖的方式如下：

1. 先對各個組別兩兩比較，挑出兩組成員都有參與討論的文章。
2. 將兩組成員在這些文章中的參與比例分別設為 X 軸與 Y 軸。
3. 將資料點描繪於圖上。

透過觀察圖上散布點與 45 度線的關係，可以理解欲比較的兩個分群之間的關係，若散布點皆離 45 度線很遠，呈現 L 形，代表兩個分群的成員參與的文章很不一樣，兩個分群的成員的興趣類別差異很大。若散布點皆距離 45 度線很近，則代表兩個分群的成員在相同類型的文章參與比例雷同，兩個分群的成員的興趣類別可能是很相近的。

3 分析結果

3.1 分群的結果

下表是我們將 ID 分群的結果：

表 1: 分群結果

群數	人數	興趣類別
第 0 群	27854	吃瓜群眾
第 1 群	774	勞基法修法
第 2 群	1746	性愛趴
第 3 群	1806	即時新聞
第 4 群	805	未定義
第 5 群	720	意識形態
第 6 群	10612	未定義
第 7 群	696	未定義
第 8 群	2953	未定義
第 9 群	4090	未定義

^a. 共有 52056 筆使用者 ID。

^b. 僅挑選較有特色的分群進行後續分析。

在 10 個分群中，透過觀察各分群成員參與討論的文章，找出了其中 5 個比較具有特色或代表性的分群，分別為：(1) 吃瓜群眾 (2) 勞基法修法關心者 (3) 性愛趴關心者 (4) 即時新聞關心者 (5) 意識形態討論者。以下將針對各群體可能會有的特性及行為模式加以說明。

3.1.1 吃瓜群眾

表 2: 吃瓜群眾推文數最高的 10 篇文章

文章標題	回應次數
Re: [新聞] 911、海嘯全命中已故靈媒預言：明年中	450
[問卦] 再見 2017 這個垃圾年	320
Re: [爆卦] 安定力量/「罷昌」案件秀下限懶人包!	310
[問卦] 雲的符號到底要怎打?	270
Re: [新聞] 檢調阻王炳忠開門 律師隔門大喊：要	261
Re: [問卦] 通識課重修開課啦!	234
[問卦] 完整圖 5 路車最後一班	215
Fw: [公告] LLsolo 法務站長辭職聲明 +2017 年行政報告	209
[問卦] 兄弟們上車囉	204
[問卦] 大家快去參加反過勞惡勞基法遊行	182

^a. 這個分群有 27854 筆使用者 ID。

^b. 相同 Repo 名稱的文章視為同一篇。

所謂的吃瓜群眾是指對事情的本質不了解，對討論、推文及各種意見抱持「圍觀」態度的一類人群，現今的網路使用者，可能有一大部分的比例都屬於這類型。

從吃瓜群眾的數量、參與的文章與推文的內容可以得知，這群成員並沒有一個比較關心的議題，通常是有跟風的心態，一但什麼話題很火，或是什麼東西被推至風口浪尖，他們就會想要參一脚。

3.1.2 勞基法修法關心者

表 3: 勞基法修法關心者推文數最高的 10 篇文章

文章標題	回應次數
Re: [問卦] 台灣勞工運動是不是已經沒救了?	884
Re: [新聞] 基本工資 3 萬中小企業：會倒一半	697
Re: [新聞] 燙手泳池何處去？台南、高雄也說 NO	522
Re: [新聞] 賴香伶參加勞團遊行柯文哲：至少我沒反對	506
Re: [新聞] 走到抽筋癱瘓西門町萬人反勞基「就這樣	483
[新聞] 柯 P 跳票法輪功 101 大樓前練功仍遭騷擾	465
Re: [問卦] 台菜二世祖的鄉民宴是不是過譽了？	378
Re: [問卦] 反核是不是真的辯不贏擁核啊？	313
Re: [新聞] 賴清德：勞力需求增加起薪定提高	298
Re: [新聞] 藍委詢問「蔣介石 3 大貢獻」 鄭麗君冷回	253

^a. 這個分群有 774 筆使用者 ID。

^b. 相同 Repo 名稱的文章視為同一篇，故單一文章的筆數可能超過 774。

勞動部於 2017 年 10 月 31 日預告「勞基法修正草案」，包括「休息日加班工資核實計算」「延長工時上限」「放寬輪班間隔」「鬆綁七休一」「特休假遞延」等五大方向之後，引起各界熱烈討論，也造成許多勞團嚴正抗議，對於勞基法修法的原因，是否有獨厚資方的討論也持續延燒。這股有關勞基法的討論，以及各處抗議活動的報導，似乎也在 PTT 八卦版上面拓展開來。本研究也發現似乎有一群 PTT 使用者持續地在八卦版討論相關議題。

3.1.3 意識形態討論者

表 4: 意識形態討論者推文數最高的 10 篇文章

文章標題	回應次數
Re: [新聞] 藍委詢問「蔣介石 3 大貢獻」 鄭麗君冷回	1065
Re: [新聞] 基本工資 3 萬中小企業：會倒一半	812
Re: [問卦] 為何中國皇帝無法在國內建立天皇等級權威？	668
Re: [新聞] 被酸「長得漂亮別說台語」暴怒謝忻霸氣	641
Re: [爆卦] 王炳忠侯漢廷開記者會抗議	635
Re: [問卦] 日治時代台灣是最美好的時代嗎？	518
Re: [問卦] 蔣經國有這麼讓人痛恨嗎 ???	514
Re: [新聞] 彭淮南護匯率政策左批張忠謀右打蔡明忠	465
[問卦] 日治時代台灣是最美好的時代嗎？	434
Re: [新聞] 台灣人為何對南京大屠殺冷感？教授點出	416

^a. 這個分群有 720 筆使用者 ID。

^b. 相同 Repo 名稱的文章視為同一篇，故單一文章的筆數可能超過 720。

本研究也發現有一個族群的成員對於意識形態相關的文章有特別關注的情況。該族群討論議題多圍繞「台灣」「中國」等台灣最核心兩岸政治議題，涉及歷史、軍事、意識形態及兩岸商業往來等討論。如「日治時代」「共軍武力」「觀光業」等。打開相關貼文會發現，該族群探討政治話題的回復中，一開始先由新聞切入，各抒己見，不乏認真想探討的人，但卻有許多無實質內容的吐槽，各方雞同鴨講後，風向漸漸跑偏。而由於 2017 年 12 月王炳忠事件燃燒，因此「王炳忠」成為此段截取資料中，關心政治與意識形態的這一族群的熱門討論話題。在回復數量最多的幾則王炳忠相關帖文中，此一族群民衆大多抒發對王「背叛國家」，違背主流意識形態之嫌厭，不乏謾罵；亦有人將其與白色恐怖對比，藉此討論藍綠兩黨執政。

3.1.4 性愛趴關心者

表 5: 性愛趴關心者推文數最高的 10 篇文章

文章標題	回應次數
Re: [問卦] 雜交趴裡面為什麼有一個肥宅？	425
Re: [新聞] 性愛趴當事人懇求刪掉外流照：我才大一	385
Re: [問卦] 免費給圖會破壞行情嗎？	378
[問卦] 末班車	374
[問卦] 收到一個好多內容的壓縮檔	351
[問卦] 謝謝大大	337
[問卦] 剛剛夢到一些東西	306
Re: [問卦] 真的有影片檔嗎？	298
Re: [問卦] 收到啦免費隨機 10 位	294
[問卦] 免費給圖會破壞行情嗎？	287

^{a.} 這個分群有 1746 筆使用者 ID。

^{b.} 相同 Repo 名稱的文章視為同一篇。

本研究也發現有一個分群是和 2017 年底爆發的性愛趴事件有關的，從該群組最多互動的篇章往下看，可以直接從標題看出這群人大概的目的，他們為何會群聚？從留言更可以得到答案，在互動高的幾篇文章裡面，大部分的留言都屬於同一個方向（私、求圖片、500P 求站內等等），目的是為了要求取性愛趴的圖片以及影片大飽眼福；其次，在「性愛趴當事人懇求刪掉外流照：我才大一」一篇文章中，也可以找到這群人為何聚集的目的，其留言以及討論方向著重在事情發生的始末以及主觀的探討。

表 6: 即時新聞關心者推文數最高的 10 篇文章

文章標題	回應次數
Re: [問卦] 台菜二世祖的鄉民宴是不是過譽了?	688
Re: [新聞] 拒檢舉魔人! 2 天 5 千人連署「勸導取代開	615
Re: [新聞] 基本工資 3 萬中小企業: 會倒一半	533
Re: [新聞] 墾丁觀光 8 年來最慘, 業者點出: 陸客減少	426
Re: [問卦] 只有我覺得世新女生處理方式太”過”了嗎	396
[新聞] 砍學妹男大生苦戀 5 年 LINE 對話紀錄讓人	352
[問卦] 請問爭勞權是害自己嗎?	351
[新聞] 開槍擊斃狂匪遭起訴警: 國家給我槍當然	325
Re: [問卦] 台灣勞工運動是不是已經沒救了?	308
[新聞] 役政署長強闖成功嶺! 駕駛嗆哨兵: 你們是	283

^a. 這個分群有 1806 筆使用者 ID。

^b. 相同 Repo 名稱的文章視為同一篇。

本研究也發現一個很有趣的分群，這個分群的成員大多數對於即時新聞很有興趣，會花時間回應新聞內容。在這些推文中，有人從政黨切入評論、有人戰南北、有人單單用一個詞描述心情作罷，有時候感受一致，有時候關注點分散，此族群所留下的評論大多不是為了與人互動或辯論，也較少嘩眾取寵，而是單純地出發己見，這種感覺就像是網路新聞文末供人選擇的心情選項，「將看完新聞的心情或第一個想法簡單描述出來」便是這些人一致的作風，他們並不期望和其他樓層的留言有更多互動。另外，這些貼文留言時間從早晨到晚上皆有，可以推測這一族群在我們所選擇的連續時間裡經常地造訪本版；因此，我會認為這一族群的劃分，「成員間是否具有相似的政治傾向或偏好」這點並不能肯定，但成員極有可能都具有以批踢踢為新聞消息來源的習慣。

3.2 參與集中度的分析結果

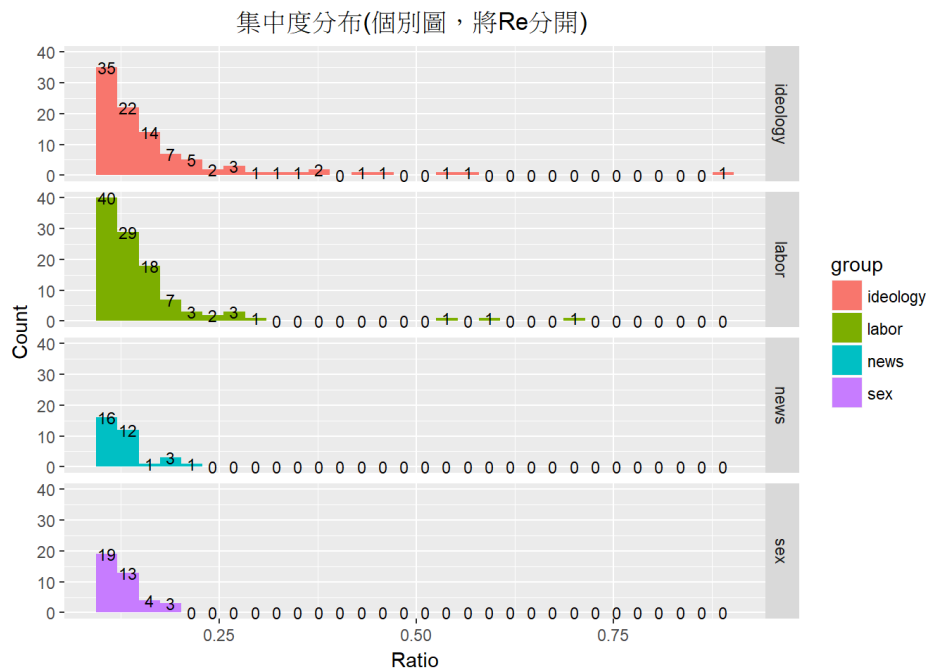


圖 2: 參與集中度分圖 (將 Repo 分開)

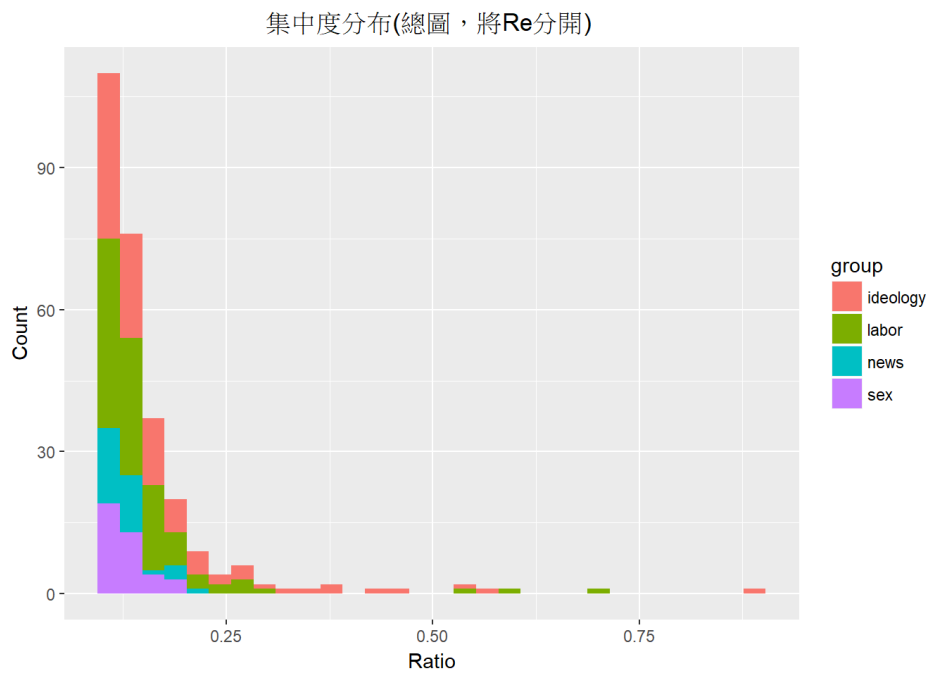


圖 3: 參與集中度疊圖 (將 Repo 分開)

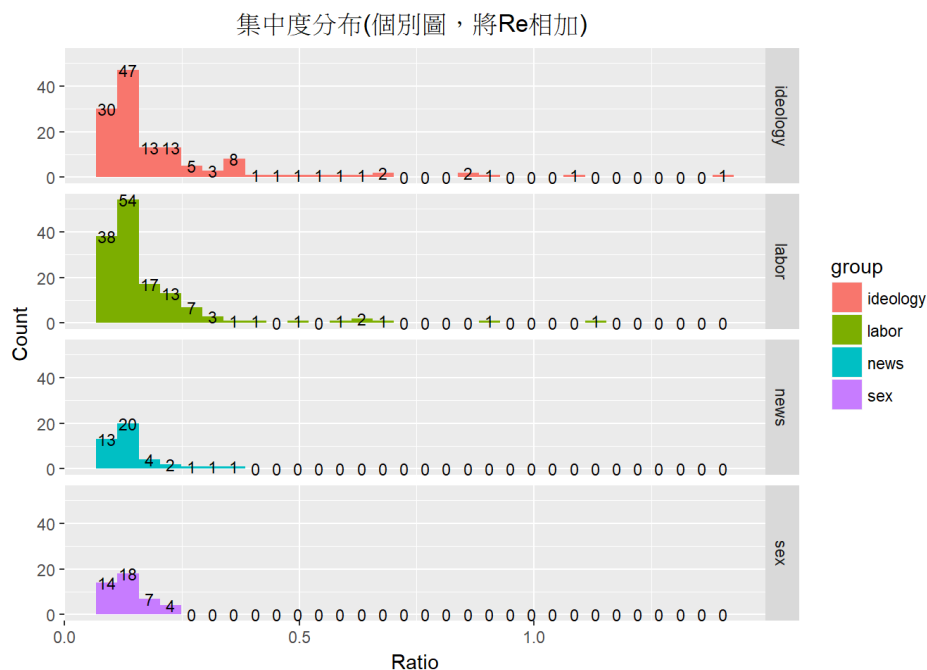


圖 4: 參與集中度分圖 (將 Repo 分開)

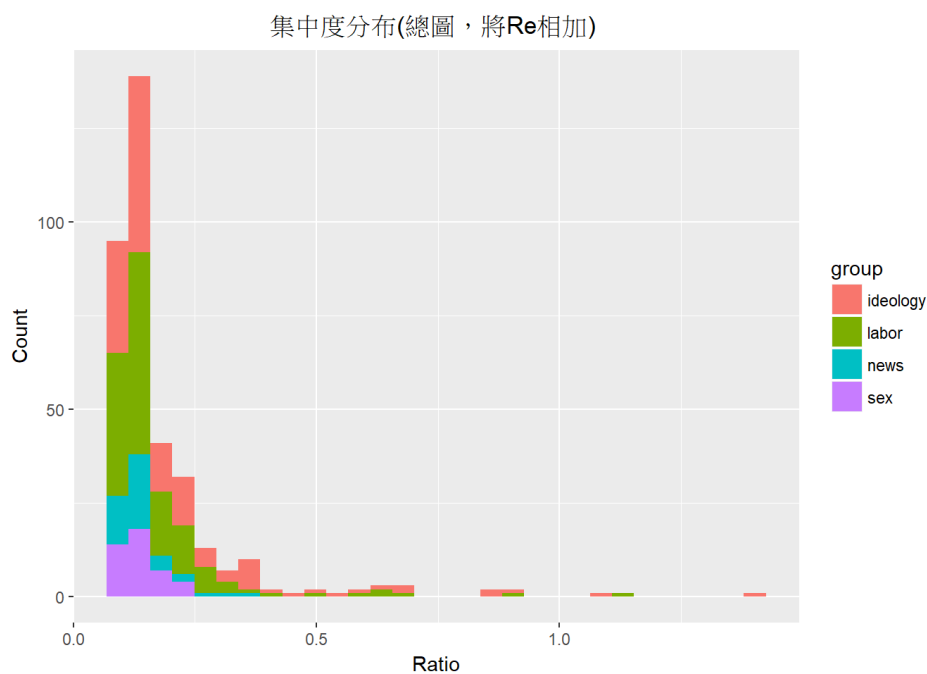


圖 5: 參與集中度疊圖 (將 Repo 分開)

不論是否將 repo 的文章分開計算，都可以發現在關心勞基法組別及意識形態組別相較於性愛趴及即時新聞組別，ID 都有較高的集中度。首先，若將 repo 相加，關心勞基法組別及意識形態組別分別有 141 及 132 篇文章該組別有超過 0.1 的比例有推文，而性愛趴及即時新聞組別分別只有 43 及 42 篇文章該組別有超過 0.1 的比例推文。若是將 Re 分開，四個組別，依順序則分別有 106、98、39 及 33 篇文章超過 0.1 比例的門檻。在集中度的分布上，關心勞基法組別及意識形態組別分別至少都有 3 篇 (含) 以上的文章有超過 50% 的組別成員參與，而性愛趴及

即時新聞組別則是沒有任何一篇文章有超過 25% 的組別成員參與。就算考慮到後面兩個組別的人數約為前面兩個組別的兩倍左右，還是可以觀察得出來前面兩個組別的集中度較高。

除了觀察參與程度大於 0.1 的文章，我們也式 (2) 的方式分別計算了各個組別的平均參與比例，平均參與比例代表的意義是該組別的成員在至少有一個成員加入討論的文章，普遍會有多少的參與比例。以勞基法關心者組別為例，該組別的平均參與比例為 0.008234668，成員總人數為 774 人，所以每篇文章平均有 $0.008234668 \times 774 = 6.37$ 個該組成員參與。從計算的結果也可以發現勞基法關心者與意識形態討論者的平均參與比例明顯地較另兩個組別為高。

$$\text{平均參與比例} = \frac{\text{該組別成員曾參與討論的文章參與比例加總}}{\text{該組別成員曾參與討論的文章總數}} \quad (2)$$

表 7: 勞基法修法關心者推文數最高的 10 篇文章

組別	平均參與比例
勞基法關心者	0.008234668
意識形態討論者	0.007373829
性愛趴關心者	0.004312062
即時新聞關心者	0.00451015

我們認為這樣的結果可能代表關心勞基法組別及意識形態組別的成員雖然較少，但結構是較為緊密的，這些成員可能是長期並持續的關心這類議題，對文章的討論是較熱烈且議題集中度高。而性愛趴及即時新聞組別的成員較前兩個組別相對來得多，但是結構鬆散，且集中度不高。對性愛趴組別而言許多的的文章是用來「求圖」，在同一篇文章中熱烈討論的比例反而少，而對即時新聞組別而言，因為即時新聞的面向過於寬廣，也較難集中於某些文章中引起熱烈討論。

3.3 參與比例比較散布圖的分析結果

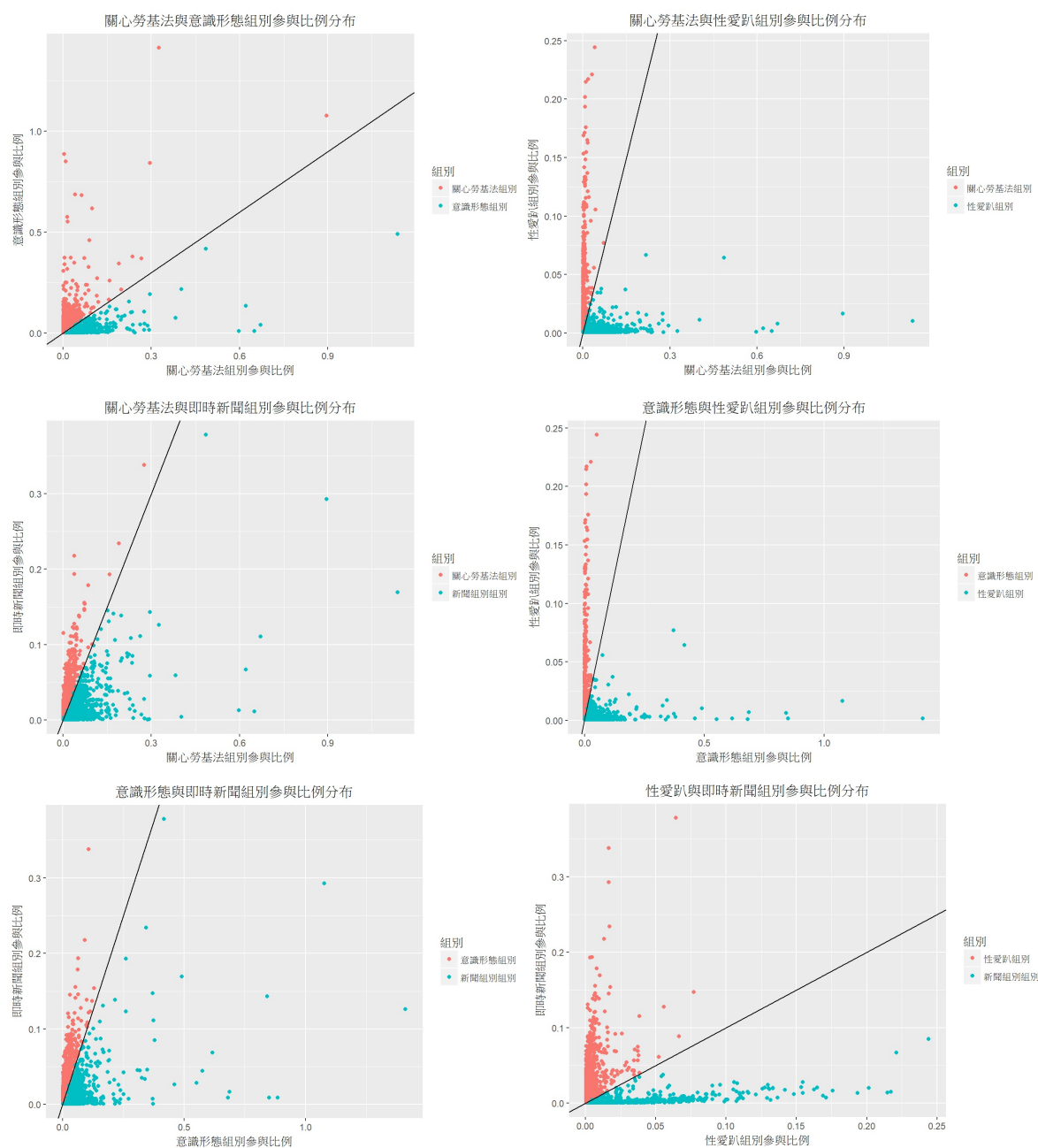


圖 6: 參與比例比較散布圖

首先，只要是將關心勞基法或意識形態組別對性愛趴或即時新聞組別繪圖，45 度線與水平線的實際夾角都會較大，這是因為前兩個組別成員在文章中的參與比例較後兩個組別成員的參與比例來得高的緣故。若是關心勞基法組別對意識形態組別繪圖或性愛趴組別對即時新聞組別繪圖則不會有這個問題，因為它們之間的參與程度相當。這某種程度上是呼應了集中度分析的結果。

再者，仔細觀察散布點與 45 度線的關係，可以發現，只要有性愛趴組別出現的圖，散布點都呈現 L 形，距離 45 度線非常的遠，這代表性愛趴組別成員與其他組別的成員，熱烈參與討論的文章有明顯的不同。總結來說，性愛趴組別的成員，在 PTT 八卦版這個社群媒體，與其他組別的成員，有一段較明顯的「距離」。

4 模型的可信度

爲了確認模型的可信度，我們在完成 Kmeans 分群步驟之後，曾以人工的方式確認我們分群的結果是否與現實符合，也就是我們預測相似度高的兩個 ID，是否真的存在緊密的關係？以 ID obov 爲例，表 8 是行爲模式與他相似的 ID，本研究挑選相似度最高的 ID Nchuxxxxxx，以 Google 搜尋的方式，確認兩者之間真的有非常緊密的關係 (如圖 7)。因此，本組的模型具有一定的可信度。

表 8: 即時新聞關心者推文數最高的 10 篇文章

使用者 ID	相似度
Nchuxxxxxx	0.9487780928611755
Bbmelonxxxxx	0.9077775478363037
Haruhitxxxxx	0.9046445488929749
Lunxxxxx	0.8945542573928833
Bigdxxxxx	0.887511134147644
Kasumxxxxx	0.8867853879928589
Mxxxxx	0.8813595175743103
Sakaxxxxxx	0.88044273853302
Zhiwxxxxx	0.8782533407211304
Sf8xxxxx	0.8747169971466064

^a. 爲了保護當事人，將 ID 末 5 碼以 xxxxx 代替。



圖 7: 以 Google 搜尋確認 obov 與 Nchuxxxxxx 的關係

5 結論

爲了了解網絡媒體使用者之間的關係，並找出可能的特定族群，本組爬取 PTT 八卦版的資料進行分析，結果發現了八卦版的確有特定族群存在。以此次的分析，至少就能找到 5 群具有特性的使用者，分別是：(1) 吃瓜群眾 (2) 勞基法修法關心者 (3) 性愛趴關心者 (4) 即時新聞關心者 (5) 意識形態討論者。再透過參與比例集中度的分析，可將第 (2) 至第 (5) 群再依其結構特性歸類，分別爲：

- 結構緊密，長期且持續關注某類議題的使用者，爲長久性的族群
 - 勞基法修法關心者
 - 意識形態討論者
- 結構鬆散，對突發性的議題有興趣但不會持續關注，通常是暫時性的族群
 - 性愛趴關心者
 - 即時新聞關心者

而透過參與比例比較散布圖的分析結果則可得知，在這 4 個分群中，性愛趴關心者所參與討論的文章和另外 3 群成員參與的文章最爲不同，性愛趴關心者是與另外 3 群距離較遠的一個分群。至於現實中究竟是哪個群體爲「鄉民」的大多數，則需要做進一步的研究才可得知。