

社群網站上之政治活動分析：以台灣 2016 年總統選舉為例*

學生：蒲郁文†

指導教授：孫春在 教授‡

摘要 — 本研究嘗試使用關鍵字擷取（**keyword extraction**）、情感分析（**sentiment analysis**）、社會網絡分析（**social network analysis**）等方法，分析 2016 年台灣總統選舉競選期間臉書各大政治性粉絲專頁之公開資料，探討使用者於社群網站上之政治活動的模式。本研究發現此次選舉有 26.63% 的社群媒體貼文是攻擊型競選，與台灣、中國或中國國民黨有關的議題是人們最關注的焦點，且特定族群會傾向於分享特定來源的資訊。同時，本研究也提供了後續研究者一套有效地自動化分析線上政治活動的方法。

關鍵字 — 社群媒體、自動化內容分析、社會網絡分析、數位行動主義、政治參與

一、緒論

社群網站的興起不僅改變了我們的社交生活，更改變了許多國家政治與權力的運作機制。在國內外皆有愈來愈多政治團體與公民團體透過網路來散佈其理念；許多具影響力的社會運動也是透過網路來連結潛在支持者或動員大眾；網路工具的靈活運用甚至還成為候選人能否贏得選舉的關鍵。

這種利用國際網路來號召群衆、表達訴求的作法被稱為數位行動主義（**digital activism**）。發生在伊朗的綠色革命（**Green Movement**）、北非及中東的阿拉伯之春（**Arab Spring**）、美國的佔領華爾街運動（**Occupy Wall Street**），皆被人們普遍認為與國際網路，特別是社群媒體，有密切的關連。國際網路、社群媒體對社會與政治的影響也成為許多學者關注的新焦點。

同樣的趨勢也發生在台灣。從洪仲丘事件、太陽花學運到反黑箱課綱運動，國際網路均對這些社會運動的促成扮演了相當關鍵的角色。這也讓許多社會學家、政治學家、傳播學家與計算機科學家開始對台灣的數位行動主義熱潮產生興趣。這不但是一個相當新穎的、跨領域的研究主題，這方面的研究對當今台灣社會也是相當重要的。

2016 年總統選舉競選期間，各政黨、候選人、利益團體與意見領袖皆大量使用臉書粉絲專頁（**Facebook Pages**）來與人民互動並宣揚其政治理念。多數民衆也相信這些網路宣傳確實有效影響了選舉的得票數。然而這些社群網站上的政治活動是否存在著某些機制？我們能

夠利用電腦來分析這些現象嗎？還有許許多多的問題有待我們回答。本研究的目的便是嘗試使用資訊科技，以創新的方法分析這些使用者在臉書粉絲專頁上的公開貼文，探討使用者於社群網站上之政治活動的模式。由於在台灣推特（Twitter）較不盛行，人們使用的社群網站以臉書（Facebook）為主，因此本研究的重心也將放在臉書上。

對數位行動主義的批判

儘管人們普遍同意數位行動主義能有效促進民主政治，這個趨勢仍然伴隨了一些潛在的隱憂。常見對數位行動主義的批評包含：(1) 數位落差（**digital divide**）：由於經濟、教育等差異，有些族群缺少機會接觸資訊科技，或者沒有能力使用資訊科技。(2) 民粹主義（**populism**）：網路本質上便具有民粹主義、安那其（**anarchy**）、去中心化（**decentralization**）等特性，網路輿論的品質可能偏低。(3) 懶人行動主義（**slacktivism**）：並沒有參與什麼實際的行動，僅是在網路上分享自己關心的社會議題，便自以為自己已經對這個議題做出了貢獻。(4) 鍵擊行動主義（**clicktivism**）：類似懶人行動主義，若一味地追求點擊率、做網路行銷，可能反而會忽略了實際的行動。(5) 網路巴爾幹化（**cyberbalkanization**）：由於商業壟斷、政治審查，或是個人不願意接觸與自己立場相左的訊息等因素，使國際網路逐漸分裂、解體。

資訊科技與人文社會跨領域研究的挑戰與契機

近年來，愈來愈多學者開始嘗試將計算機科學應用於人文社會科學領域，建立了數位典藏（**digital archives**）、數位人文（**digital humanities**）、計算社會科學（**computational social science**）、計算社會學（**computational sociology**）、社會模擬（**social simulation**）等新興研究領域。然而，這樣的跨領域研究是否真的有價值一直是備受爭議的。以下舉出一些此類研究常受到的批評，以及我從事此類研究的立場。

常見對數位人文的批評包含：(1) 只是用一些空洞的戲法來賣弄他們的研究能力，而沒有真正做出什麼新的分析。(2) 容易忽略種族、階級、性別、文化、意識形態等面向。(3) 在這個一切都已經被商品化了的現代

*學士論文。本研究未接收任何個人及團體贊助。

†國立交通大學資訊工程學系，ywpu@cs.nctu.edu.tw。

‡國立交通大學資訊工程學系，ctsun@cs.nctu.edu.tw。

社會中，人文學科日漸式微。但數位人文卻頂著科技的光環，而獲得了相當高的經費與聲望。(4) 研究者的背景不夠多元，其研究可能帶有某些偏見。

常見對社會模擬的批評包含：(1) 將現實社會過度地簡化為一般化的法則。(2) 模擬的成果仰賴預先建立好的模型，難以發現新的法則。(3) 預先建立的模型容易帶有偏見。(4) 模擬的成果並不能反映出現實社會的情況。

我的回應是，我認為傳統社會科學研究法仍是不可或缺的：(1) 並非所有問題都適合使用電腦來分析或模擬，有些現象無法被量化或歸納出一般化的法則。(2) 人文社會科學做的不只是資料分析、統計而已，還包含人文關懷、詮釋、思辨、批判等。(3) 因此田野工作、民族誌、訪談等研究法仍有其彌足珍貴的價值。

然而資訊科技仍然能對人文學科做出貢獻：(1) 數位人文提供了一個全新的視野來探究傳統上屬於人文社會科學領域的問題。(2) 資訊科技能幫助人文學者更有效率地獲得知識。(3) 資料探勘（data mining）、自然語言處理（natural language processing）等技術使一些過去被認為窒礙難行的研究成為可能。

二、文獻探討

過去已有多位學者考察了社群網站在競選活動中的使用，例如 Larsson and Moe (2012) 研究了 2010 年瑞典選舉競選期間推特的使用。^[1] 該研究的作法是，蒐集競選期間含有特定 hashtag 的推文，進行社會網絡分析（social network analysis）。

該研究做了以下分析：(1) 不同時間推文數量的變化。(2) 各種推文類型（一般、提及、轉推）所佔的比例。(3) 列出最活躍的使用者及其推文數量，並調查其身份背景。(4) 活躍用戶的「提及」網絡圖（透過資料視覺化）。(5) 呈上，將使用者分為發送者、接受者、兩者皆是，並調查其身份背景。(6) 活躍用戶的「轉推」網絡圖（透過資料視覺化）。(7) 呈上，將使用者分為轉推者、被轉推者、兩者皆是，並調查其身份背景。

該研究發現：(1) 推文數量的變化反應了線下競選活動的情勢。(2) 大部份的推文都是由少部份的活躍用戶所發出。(3) 大部份的活躍用戶都是政治菁英，然而，平民的聲音仍具有一定的影響力。(4) 大部份的推文都是一般推文，然而，許多活躍用戶仍會在推特上與人討論。

然而該研究也具有以下限制：(1) 只有極少數的民眾有使用推特。(2) 只蒐集含有特定 hashtag 的推文。(3) 應分析不同情境下的使用情形，加以比較。另外，該研

究也讓我想到一個值得進一步探究的問題，那就是，如何區別真人用戶與機器用戶？

另外 Zhang et al. (2010) 則研究了社群網站對公民的政治態度與行為的影響。^[2] 該研究參考了以下既有理論：(1) 社會資本（social capital）、大眾媒體（mass media）、人際討論（interpersonal discussion）、社群網站（social networking sites）都會影響公民參與（civic participation）和政治參與（political participation）。(2) 影響個人的政治態度與行為的因子有：政治興趣（political interest）、政治效能（political efficacy）、政治信任（political trust）、政黨認同（party identification）。

該研究的作法是採用電話調查，應變數（dependent variable）為政治參與、公民參與、對政府的信任，自變數（independent variable）為社群網站之依賴、政治之人際討論，控制變數（control variable）為政治效能、意識形態、政治興趣、年齡、性別、種族、教育程度。最後，進行階層式迴歸分析（hierarchical regression analysis）。

該研究發現：(1) 社群網站的使用與公民參與的提升呈現顯著的正相關，與政治參與、對政府的信任則無明顯關連。(2) 討論政治議題有助於促進公民參與和政治參與，無助於提升對政府的信任。(3) 教育程度愈高，愈有可能參與公民活動和政治活動，但對政府的信任卻愈低。(4) 然而，政治效能感較高者，其政治信任感也較高。

然而該研究也具有以下限制：(1) 樣本可能不夠具代表性，有超過四分之三的受訪者都完全不依賴社群網站。(2) 除了公民參與／政治參與的二分法外，是否還有更適當的分類法？另外，該研究也讓我想到兩個值得進一步探究的問題：(1) 調查結果顯示社群網站的使用無助於促進政治參與，這項發現在東方國家也成立嗎？(2) 調查結果顯示年長者較積極參與政治活動，當今台灣社會也是如此嗎？

最後 Hosch-Dayican et al. (2016) 則研究了 2012 年德國議會選舉競選期間公民如何在推特上遊說他人。^[3] 其研究方法為：(1) 資料蒐集（data collection）：先選定數個 hashtag，再透過滾雪球抽樣（snowball sampling），找出其他相關的 hashtag。蒐集選舉期間含有這些 hashtag 的推文。(2) 資料清理（data cleaning）：接著，需刪去誤判的推文。先刪去語言與地區皆不符的推文，再從剩下的推文中篩選出直接與選舉有關的推文。(3) 操作化（operationalization）：承上，先將推文分為「競選活動」與「非競選活動」兩種類型，再將屬於「競選活動」的推文分為「說服

型競選（persuasive campaigning）」與「攻擊型競選（negative campaigning）」兩種類型。(4)自動化內容分析（automated content analysis）：以上篩選與分類的過程仰賴質化內容分析（qualitative content analysis）。採用監督式學習（supervised learning），先由研究者做一小部份的樣本，再用它來訓練機器。具體作法包含克氏 α 係數（Krippendorff's alpha）、單純貝氏分類器（naive Bayes classifier）、一元語法（unigram）與二元語法（bigram）、詞幹提取（stemming）、十等分交叉驗證（10-fold cross-validation）等等。

該研究做了以下分析：(1)各種推文類型（非競選活動、說服型競選、攻擊型競選）與各種用戶類型（政黨／政治人物、媒體／記者、平民）的組合的推文所佔的比例。(2)從屬於「競選活動」的推文中隨機抽樣，進行更深入的內容分析。不論是「說服型競選」還是「攻擊型競選」，其推文皆可分為直接宣傳（明確表態要／不要投給誰）與間接宣傳（僅表達支持／反對意見）兩種。

該研究發現：(1)平民相當投入線上政治活動。(2)政黨／政治人物最常採取說服型競選，媒體／記者及平民最常發佈非競選活動的推文。(3)相較於其他用戶類型，平民採取攻擊型競選的比例最高。儘管在所有的用戶類型中，用戶採取攻擊型競選的比例都是最低的。(4)說服型競選相當程度地採用「轉推」的形式。直接宣傳的部份，常出現含有投票動機或投票行為的貼文，尤其是在選舉前夕和選舉當天。間接宣傳的部份，常出現對新聞報導、採訪、辯論的回應。(5)攻擊型競選則多為間接宣傳，採用「轉推」的形式，引用政黨／政治人物的推文。比起政治菁英，平民的推文更包含了對採訪與辯論的回應，以及對政黨／政治人物的不滿與憤怒。(6)平民被歸類為攻擊型競選的推文大部份都是在宣洩情緒與意見，而不是以進行攻擊型競選為目的。

該文作者建議後續研究者，進一步分析「轉推」網絡，並探討線上競選的後果。另外，該研究也讓我想到了幾個值得進一步探究的問題：(1)這份研究只有區別政治菁英與平民，忽略了使用者間的其他差異（例如社會經濟背景）。(2)採用監督式學習能降低解讀反諷文章的錯誤率到什麼程度？(3)台灣的使用者是否也具有如此高的素養，僅將近 20% 的網路貼文被歸類為攻擊型競選？

三、研究問題與方法

本研究分析的資料來自 2016 年台灣總統選舉競選期間臉書各大政治性粉絲專頁之公開資料。本研究進行的分析包含以下四個部份：

- 各政黨/候選人/意見領袖分別發佈了多少篇貼文？每日的貼文數量隨著時間有什麼樣的改變？
- 各政黨/候選人/意見領袖最關心的議題是什麼？人們關注的焦點隨著時間有什麼樣的改變？本次選舉人們最關心什麼？
- 哪些政黨/候選人/意見領袖最常使用攻擊型競選？什麼時候人們最常使用攻擊型競選？本次選舉有多少比例的貼文是攻擊型競選？
- 各發文者與其分享的資訊的來源有什麼樣的關係？特定族群是否會傾向於分享特定來源的資訊？

本研究主要採用量化研究（quantitative research），著重於資料分析。在研究方法上，本研究結合了多種現代計算機科學理論與技術：

- 資料探勘（data mining）
- 機器學習（machine learning）
- 自然語言處理（natural language processing）
- 資料視覺化（data visualization）

本研究具體的研究步驟如下所述。

資料蒐集（data collection）

使用 Facebook Graph API 蒐集競選期間（投票日前三個月至投票日止）台灣各大政治性粉絲專頁之公開資料。本研究蒐集的資料以貼文為主，同時亦包含一些附屬的詮釋資料（metadata），例如發文時間、分享的連結等。為保護個人隱私，本研究只蒐集粉絲專頁上的公開資料，不蒐集任何個別使用者的資料。

採用何種方法決定欲蒐集的粉絲專頁清單，對研究結果有極大的影響。為確保本研究的中立性，我參考 Socialbakers [4] 及 Likeboy [5] 提供之社群媒體統計數據，輔以我個人的自身經驗，整理出一份台灣熱門政治性粉絲專頁的清單。得到粉絲專頁清單後，我還做了進一步的篩選。我查詢這些粉絲專頁最新的按讚數，將它們依按讚數由高到低排序。擁有超過一萬個讚的粉絲專頁共計有 135 個，這些粉絲專頁便是我最終選定的蒐集對象。

完成後，本研究共取得了 18,967 筆臉書貼文，來自 135 個台灣熱門的政治性粉絲專頁。

斷詞（word segmentation）

下個步驟是進行文本斷詞，也就是將一篇文章拆解成一個個的單詞。例如「維根斯坦是二十世紀最有影響力的哲學家之一。」這句話可以被拆解成「維根斯坦」、「是」、「二十世紀」、「最」、「有」、「影響力」、「的」、「哲學家」、「之一」、「。」等單詞。斷詞是自然語言處理中最基本的步驟，文本必需先經過斷詞處理，才得以進一步分析。

本研究採用中央研究院資訊科學研究所中文詞知識庫小組所研發之中文斷詞系統 [6]，搭配我自行開發之中文斷詞系統 Python 客戶端程式 [7]，為蒐集到的貼文進行斷詞。本研究之所以選擇中央研究院的中文斷詞系統，是因為此系統正確率高（其研發團隊亦稱此系統之正確率約為 95% ~ 96%），且相較於其他非台灣開發的中文斷詞系統，此系統對台灣本地的文本有較好的表現。若要採用其他斷詞系統，可能得用台灣本地的語料重新訓練，才能達到同樣的表現。另外，良好的未知詞辨識能力也是此系統的一大優點。

由於本研究分析的文本是社群網站上的貼文，其形式較不嚴謹，常有錯別字、文法錯誤、使用空格或換行代替逗號或句號、夾雜表情符號或特殊字元等情形，故我在每則貼文開始斷詞前均會先對貼文進行簡單的預處理（pre-processing）。斷詞全部完成後，我更做了完整性測試（integrity test），檢查斷詞前與斷詞後的貼文是否一致。測試結果顯示斷詞系統相當穩定，僅極少數（不到 3%）的貼文含有錯誤（大多是出現亂碼）。

為了讓後續的分析更方便，我使用 peewee 物件關係對映（object-relational mapping）套件 [8]，將斷詞後的文本與詮釋資料儲存在資料庫中。往後若需查詢文本，只要透過 peewee 即可輕鬆完成，分析的結果也可透過 peewee 存入資料庫。

總體分佈（*overall distribution*）

先對研究資料做最初步的分析。使用 pandas 資料分析函式庫 [9] 和 Matplotlib、Plotly 等繪圖函式庫 [10, 11] 繪製貼文數量對時間的關係圖以及累計貼文數量對時間的關係圖。另外，統計各粉絲專頁分別有多少筆貼文、佔全部貼文的百分比。

關鍵字擷取（*keyword extraction*）

接著，本研究試圖分析貼文的關鍵字，藉以找出競選期間人們關注的焦點。由於每筆貼文篇幅不一，且本研究關心的是一個範圍內的貼文的整體特性，故本研究不採用「分析單一貼文的關鍵字再將結果整合」的作法，而是「先整合多筆貼文再分析其關鍵字」，如此可得到較好的效果。

本研究進行了三種不同層面的關鍵字分析。第一種分析方式是依作者劃分，合併相同粉絲專頁的貼文，分析各粉絲專頁發佈的貼文的關鍵字。第二種分析方式是依時間劃分，合併相同時間區間的貼文，分析各時間區間發佈的貼文的關鍵字。第三種分析方式是合併所有貼文，將之視為一個整體，分析整個競選期間所有人發佈的貼文的關鍵字。

透過以上三項分析，我們可以回答：(1) 各政黨／候選人／意見領袖最關心的議題是什麼？(2) 人們關注的焦點隨著時間有什麼樣的改變？(3) 本次選舉人們最關心什麼？

在技術上，本研究使用了 SnowNLP 函式庫 [12] 實作的 TextRank 演算法 [13]。SnowNLP 是一套針對中文的文本處理函式庫，惟其訓練資料僅包含簡體中文文本。為符合本研究之需求，本研究對該函式庫做了一些調整。本研究替換掉它的斷詞部份，只採用它的 TextRank 模組。另外，本研究所使用的停用字（stop word）清單，則是以它的停用字清單為基礎加以改良而成的。

依上述方法完成分析後，我發現某些持續被廣泛爭論的議題會掩蓋掉其他短暫出現的議題，使我們不容易找出不同時間區間人們關注的焦點的變化。為此，我嘗試了另一種分析各時間點之熱門議題的方法 [14]。首先，尋找出現次數有明顯高峰的單詞：找出所有出現超過一百次的單詞，統計這些單詞每日的出現率，然後以「平均值 + 2 × 標準差」為門檻，找出連續三日出現率高於門檻的單詞。接著，找出在相近時間點出現高峰的單詞，組成詞組。再從詞組反推事件。最後將事件中的每個單詞的出現率正規化，並繪圖。

情感分析（*sentiment analysis*）

一個範圍內的貼文，可能都是在談論相近的主題，但每篇貼文所傳達的情感卻有很大的差異，彼此毫無關連。因此，與關鍵字擷取不同，情感分析不宜「先整合多筆貼文再分析其情感」，而應該採用「分析單一貼文的情感再將結果整合」的作法。

本研究將貼文分為兩類：說服型競選（persuasive campaigning）貼文，即正面（positive）貼文，和攻擊型競選（negative campaigning）貼文，即負面（negative）貼文。本研究對攻擊型競選的定義是：貼文內容是以批評反對陣營為主、貼文重點是在批評反對陣營。本研究對說服型競選的定義是：非屬攻擊型競選的貼文。也就是說，中立的貼文在本研究裡也會被歸類到說服型競選貼文／正面貼文。

本研究先分析個別貼文的情感，再統計各個時間區間內各個粉絲專頁一共發佈了多少筆貼文、其中攻擊型競選貼文有多少筆、佔全部的百分比。基於這些數據，我們可以回答：(1) 哪些政黨／候選人／意見領袖最常使用攻擊型競選？(2) 什麼時候人們最常使用攻擊型競選？(3) 本次選舉有多少比例的貼文是攻擊型競選？

在技術上，本研究使用了我自行實作的單純貝氏分類器（naive Bayes classifier）[15]。單純貝氏分類器雖然原理簡單，但用在情感分析上卻有相當優異的表

現。單純貝氏分類器又有多種不同的模型，如多項單純貝氏（multinomial naive Bayes）、二元化多項單純貝氏（binarized multinomial naive Bayes）、伯努利單純貝氏（Bernoulli naive Bayes）、高斯單純貝氏（Gaussian naive Bayes）等，其中又以二元化多項單純貝氏較適合用於情感分析 [16]，故本研究決定實作二元化多項單純貝氏。

二元化多項單純貝氏是從多項單純貝氏發展出來的。先去除文章中重覆的單詞，再輸入多項單純貝氏模型，即為二元化多項單純貝氏。「二元化」的精神是，只管一個單詞在一篇文章中有沒有出現，不管該單詞在文中究竟出現了多少遍。另外，關於此模型的實作，在一些細節上我也有特別處理。譬如，我使用對數以避免浮點下溢（floating point underflow），我使用拉普拉斯平滑法（Laplace smoothing）以避免計算結果為零。

我從所有貼文中隨機選取 948 筆（約 5%）作為訓練／測試資料，這些訓練／測試資料由我手動標記情感類型。這些選中的貼文呈均勻分佈（uniform distribution），以確保訓練／測試資料的代表性。在這裡我也用了 Matplotlib 來視覺化訓練／測試資料各情感類型的分佈。

我使用十等分交叉驗證（10-fold cross-validation）來評估模型的表現（我並未使用此方法來決定模型）。將訓練／測試資料分成十等分，其中一分作為測試資料，其餘九分作為訓練資料。重覆十次，每次選不同分當測試資料，得到十組表現度量。取其平均，即為此模型最終的表現度量。最後再用所有已標記的資料來訓練。

社會互動（social interaction）

最後一項分析屬於社會網絡分析（social network analysis）的範疇。在臉書上，我們可以分享他人的貼文，或是將網址夾帶於貼文中，分享網站連結。這裡我想探討的便是這種社群網站上的「分享」網絡。

對每篇貼文，皆去探查它的作者是誰、它分享了來自哪裡的貼文或連結，將這些「被分享者-分享者」對映整理成一份清單。接著，再將此清單輸入 Gephi 網絡視覺化軟體 [17]，即可繪製各發文者與其分享的貼文／連結的來源之網絡圖。此網絡圖是一個有向圖（directed graph），藉由此圖，我們將能夠回答：特定族群是否會傾向於分享特定來源的資訊？

這步驟有幾個技術上的細節需要留意，首先是資料處理的部份。本研究不僅在貼文裡搜尋連結，也檢查臉書貼文的「story」和「link」屬性，尋找所有該貼文分享的貼文和連結。每個連結皆需檢查它是否為短網址（經 goo.gl、bit.ly、tinyurl.com、t.co、ppt.cc 等平台縮短過的

網址），如為短網址，必需找出它的原始網址，這樣做出的分析才有意義。所有網址皆只保留其主機名稱（host）部份，捨去其路徑（path）。另外，對於臉書的連結，必需找出它導向的粉絲專頁的名稱。若該臉書連結並非導向粉絲專頁，則捨棄該連結。最後，排除「自己分享自己的貼文」的情況。

在網絡圖繪製的部份，我採用了 Force Atlas 佈局演算法（layout algorithm），此演算法最能清楚呈現本次探討的網絡。接著，我也進行了模組性（modularity）分析來偵測社群（community）。為了進一步了解使用者在網絡裡的行為模式，我還使用了 HITS 網絡連結分析。最後，為了讓網絡圖更簡潔易讀，我只保留分支度（degree）不小於 3 的節點（node）。

四、研究結果與討論

總體分佈（overall distribution）

圖 1 是貼文數量對時間的關係圖，橫軸為時間，以日為單位，縱軸為貼文數量，以篇為單位。圖 2 是累計貼文數量對時間的關係圖，橫軸為時間，以日為單位，縱軸為累計貼文數量，以篇為單位。從這兩張圖我們可以看出貼文數量呈現週期性變化，每逢假日貼文數量就會顯著降低。隨著投票日的接近，各政黨／候選人／意見領袖也愈來愈躍躍地在社群網站上發佈貼文。選舉前幾天的貼文數量更達到其餘時間的兩倍以上。

接下來，表 1 列出了各政黨／候選人／意見領袖的貼文數量，並依貼文數量遞減排序。發佈最多貼文的粉絲專頁是「柯建銘」，其貼文佔了全部貼文的 2.56%。此次選舉的三位總統候選人也都有相當高的貼文數。此外，有幾個粉絲專頁在競選期間沒有發佈任何貼文（也可能是有發佈貼文但事後又被刪除，或因故無法透過 Facebook Graph API 存取），因此並未出現在表上。

關鍵字擷取（keyword extraction）

表 2 至表 4 是關鍵字擷取的結果，#N 表示排名第 N 名的關鍵字。表 2 是依作者劃分，表 3 是依時間劃分，表 4 則是合併了所有貼文。本項分析的成果相當出色，從表 2 我們可以看出各政黨／候選人／意見領袖重視的議題。例如，「王丹網站 Wang Dan's Page」關注「中國」方面的議題，「中國國民黨 KMT」關注「朱立倫」、「經濟」、「兩岸」等事務，「周芷萱 斬父權獨立劍 老娘就是反骨！」關注「女性」、「性別」等議題，「王金平」關注「國會」、「改革」等事務，「綠黨」關注「勞工」、「環境」等議題，而「馬躍·比吼 Mayaw Biho」則關注「原住民」的文化和「土地」等等。

表 1: 各政黨／候選人／意見領袖的貼文數量
(限於篇幅，僅節錄部份資料)

粉絲專頁名稱	貼文數量	百分比
柯建銘	485	2.56%
Taiwan Fugue	361	1.90%
台灣新力量/羅致政粉絲團	337	1.78%
宋楚瑜找朋友	334	1.76%
反馬英九聯盟	330	1.74%
林昶佐 Freddy Lim	320	1.69%
基進黨（基進側翼）	319	1.68%
BillyPan 潘建志醫師	306	1.61%
宋楚瑜	301	1.59%
綠黨	297	1.57%
民主進步黨	293	1.54%
鄭永金後援會	282	1.49%
我是台灣人 (I am Taiwanese)	281	1.48%
江啓臣服務讚	262	1.38%
鄭正鈞	259	1.37%
打馬悍將粉絲團	255	1.34%
中國國民黨 KMT	254	1.34%
徐欣瑩	250	1.32%
馬躍·比吼 Mayaw Biho	250	1.32%
蔡英文 Tsai Ing-wen	249	1.31%
大安推范雲	246	1.30%
公民1985行動聯盟	246	1.30%
洪慈庸	242	1.28%
馬英九下台下台粉絲團	241	1.27%
挺馬英九聯盟	231	1.22%
呂欣潔 Jennifer Lu	230	1.21%
王如玄	220	1.16%
黃韻涵粉絲頁	219	1.15%
社會民主黨	219	1.15%
朱立倫	214	1.13%
吳秉叡	199	1.05%
段宜康	198	1.04%
鄭弘儀	197	1.04%
陳其邁 Chen Chi-Mai	195	1.03%
公民廟口-立委在做天在看	194	1.02%
時代力量 New Power Party	194	1.02%
沃草 Watchout	193	1.02%

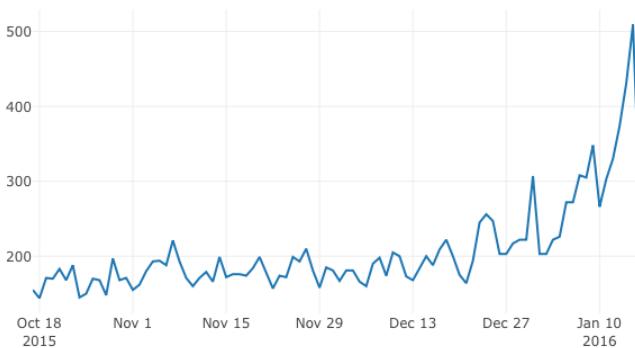


圖 1: 貼文數量對時間的關係圖

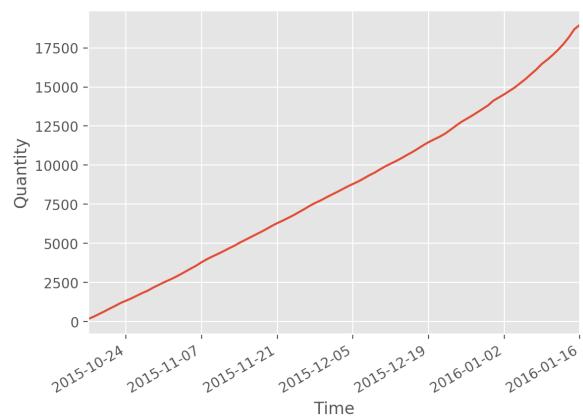


圖 2: 累計貼文數量對時間的關係圖

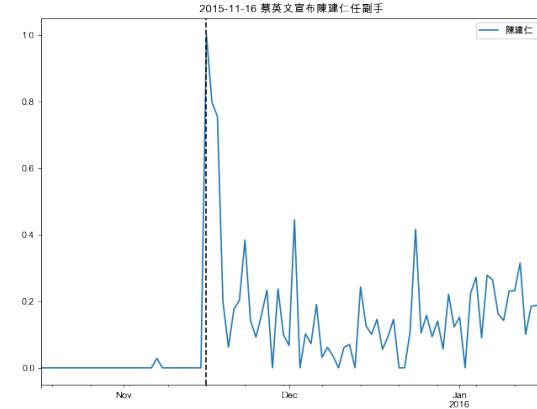
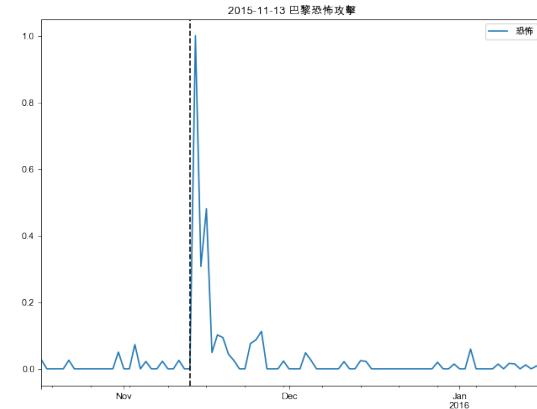
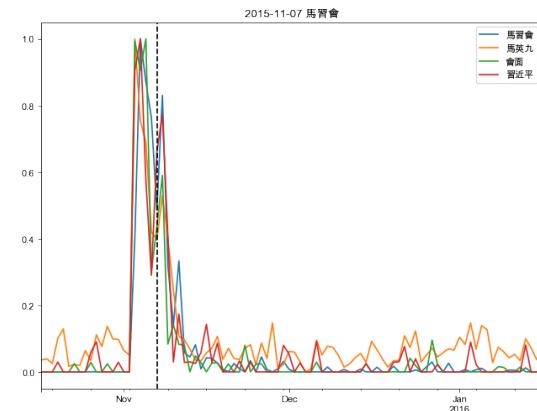


圖 3: 各時間點的熱門議題

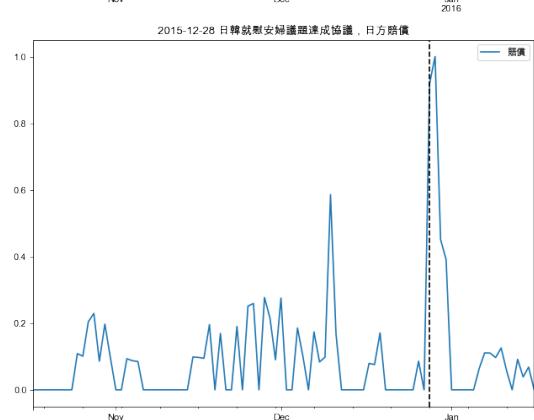
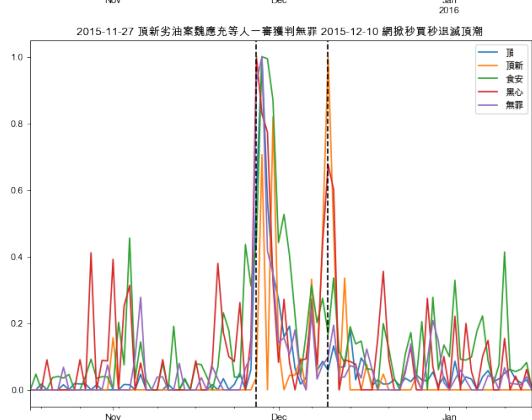
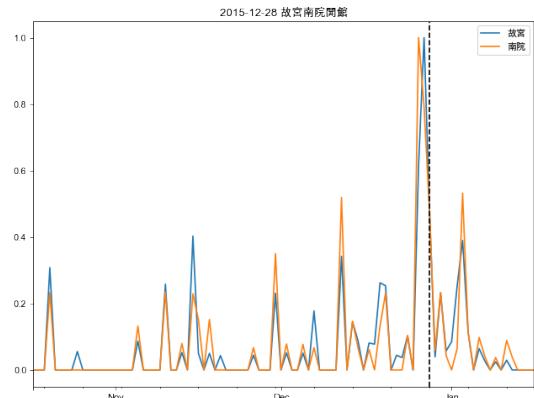
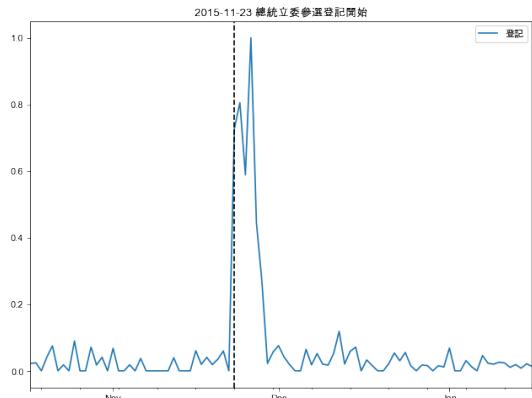
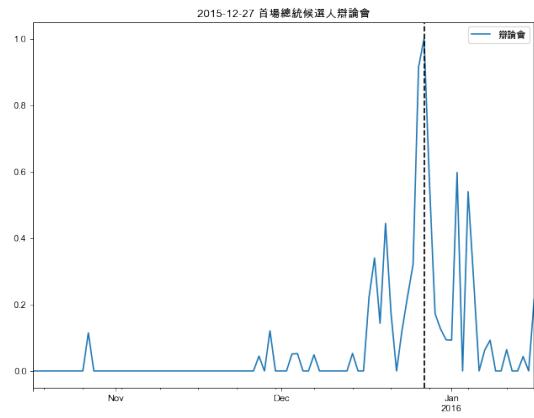
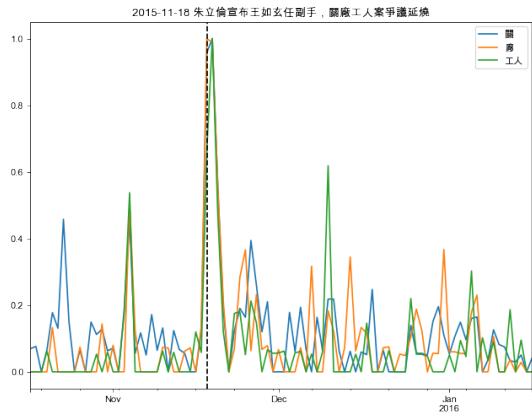
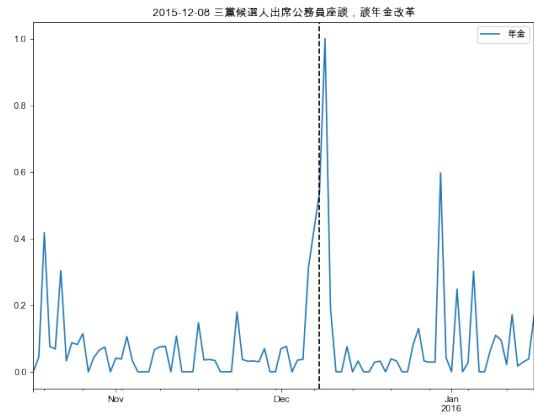
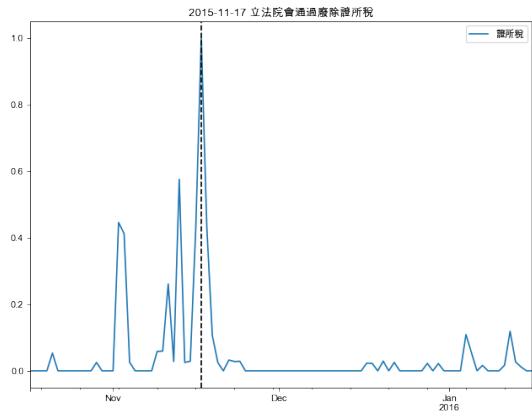


圖 3: 各時間點的熱門議題

圖 3: 各時間點的熱門議題

表 2: 關鍵字擷取，依作者劃分（限於篇幅，僅節錄部份資料）

粉絲專頁名稱	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
朱立倫	台灣	新	孩子	支持	推動	政策	國民黨	北市	經濟	中
大安推范雲	范雲	台灣	政治	中	感謝	朋友	支持	新	立委	大安
王丹网站 Wa...	中國	中	大學	朋友	運動	台灣	件	國家	臺灣	老師
國民黨青年團	青年	台灣	活動	力量	總統	朱立倫	主席	青年團	報名	時間
呂欣潔 Jennif...	政治	欣潔	台灣	中	呂欣潔	朋友	希望	支持	新	照顧
管碧玲 (kuan...	管媽	市長	台灣	高雄	朋友	旗津	中	總部	市民	第一
段宜康	朱立倫	國民黨	王如玄	台灣	新	公司	買	委員	兩	候選人
鄉民實業坊	黑子	種	書	中	現場	文章	真的	幫	Chenglap	11
我是台灣人 (...	台灣	綠黨	聯盟	國民黨	政黨	中	政治	立委	支持	新
林智堅	新竹市	新竹	市民	服務	中	城市	客家	朋友	台灣	座
BillyPan 潘...	醫師	台灣	潘建志	潘	潘醫師	中	支持	朋友	號	感謝
蔡錦隆加油讚	蔡錦隆	錦隆	台灣	總部	競選	台中	鄉親	今日	立委	舉辦
馬英九	臺灣	中	政府	兩	阿嬤	我國	兩岸	中華民國	國際	孩子
醫勞盟	醫師	醫療	健保	醫院	元	病人	醫勞盟	人員	月	健保署
陳其邁 Chen...	國民黨	元	中	台灣	食品	總統	民進黨	朱立倫	抹黑	約
中國國民黨...	台灣	朱立倫	國民黨	未來	經濟	兩岸	政策	立委	主席	國家
徐欣瑩	台灣	欣瑩	徐欣瑩	民國黨	民衆	人民	中	政治	主席	鄉親
全國廢核行...	能源	核	台電	核電	核電廠	政府	核災	台灣	中	核四
柯文哲	台北	政府	市民	市府	中	台北市	臺北市	工作	台灣	臺北
白色正義聯盟	台灣	支持	白盟	民進黨	兩岸	國民黨	小編	馬	中	總統
台灣勞工陣...	勞工	台灣	勞動	工作	政府	工會	中國	勞陣	元	經濟
謝長廷	國民黨	台灣	候選人	民進黨	總統	立委	英文	蔡	兩	曾
柯建銘	柯建銘	新竹	國會	新竹市	台灣	國民黨	民進黨	柯	改革	總召
法操FOLLOW	檢察官	大法客	證據	法官	起訴	新	案	頂	司法	Q
王定宇	台南	定宇	永康	國民黨	點	仁德	中	來到	台灣	參加
鄭永金後援會	新竹	鄭永金	台灣	新竹縣	鄉親	蔡英文	支持	客家	活動	感謝
黃韻涵粉絲頁	韻涵	活動	岡山	高雄	希望	中	支持	感謝	鄉親	一同
小聖蚊的治...	國民黨	台灣	朱立倫	話	馬英九	支持	新聞	民進黨	姐	食安
民主進步黨	全文	台灣	蔡英文	號	民進黨	產業	新	立委	國民黨	改革
周芷萱 斬...	台灣	性	中	種	社會	女性	性別	女人	工作	老娘
賴清德	台南	台灣	朋友	文化	市民	中	新	市府	臺南市	感謝
Taiwan Fugue	國民黨	朱立倫	台灣	馬英九	立委	新	中國	總統	王如玄	中
林佳龍	台中	市府	台灣	台中市	中	希望	元	市民	未來	活動
公民廟口-立...	台灣	立委	真的	種	中國	想	國民黨	國會	政府	中
吳秉叡	新莊	台灣	秉	鄉親	大秉	新	小英	立委	秉叡	中
張承中	政府	兩	想	民調	朋友	11	整合	企業	公益	希望
幸福好粽 鍾...	易仲	鄉親	美濃	邱議瑩	立委	高雄市	高雄	台灣	水	鍾易仲
鄭弘儀	寶島	台灣	大哥	世界	節目	鄭	中國	聯播網	馬英九	管理員
趙正宇	正宇	桃園	活動	趙正宇	八德	改變	支持	八德區	力量	新
我是中壢人	中壢	桃園	仲介	台灣	學生	點	學校	中	時間	教學
王金平	國會	改革	台灣	立委	國民黨	委員	立法院	支持	民進黨	人民

表 3: 關鍵字擷取，依時間劃分

時間區間	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
2016-01-10 ~ 2016-01-17	台灣	國民黨	中	立委	新	支持	政治	中國	候選人	選舉
2016-01-03 ~ 2016-01-10	台灣	中	國民黨	號	政府	新	立委	支持	兩	候選人
2015-12-27 ~ 2016-01-03	台灣	中	政府	新	國民黨	政治	立委	兩	社會	支持
2015-12-20 ~ 2015-12-27	台灣	中	國民黨	新	政府	立委	號	國會	候選人	支持
2015-12-13 ~ 2015-12-20	台灣	中	新	國民黨	政府	立委	中國	政治	國會	支持
2015-12-06 ~ 2015-12-13	台灣	中	國民黨	政府	新	社會	活動	立委	政治	支持
2015-11-29 ~ 2015-12-06	台灣	中	新	立委	政治	政府	國民黨	中國	支持	種
2015-11-22 ~ 2015-11-29	台灣	中	新	立委	國民黨	政府	中國	政治	活動	兩
2015-11-15 ~ 2015-11-22	台灣	中	政府	新	政治	立委	活動	社會	兩	國家
2015-11-08 ~ 2015-11-15	台灣	中	中國	政府	政治	新	活動	立委	支持	國民黨
2015-11-01 ~ 2015-11-08	台灣	中	中國	政府	總統	馬英九	馬習會	人民	國民黨	兩岸
2015-10-25 ~ 2015-11-01	台灣	中	國民黨	新	政府	政治	支持	立委	活動	兩
2015-10-17 ~ 2015-10-25	台灣	中	新	國民黨	政治	活動	政府	朱立倫	文化	朋友

表 4: 關鍵字擷取，合併所有貼文

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
全部貼文	台灣	中	國民黨	新	政府	兩	中國	政治	活動	立委

表 5: 貼文情感分析（限於篇幅，僅節錄部份資料）

粉絲專頁名稱	01/10	01/03	12/27	12/20	12/13	12/06	11/29	11/22	11/15	11/08	11/01	10/25	10/17	所有時間區間	
朱立倫	1/30	1/18	2/18	0/18	0/17	0/20	2/17	0/16	2/6	0/15	1/15	0/15	1/5	10/210	
(3.33%)	(5.56%)	(11.11%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(11.76%)	(0.00%)	(33.33%)	(0.00%)	(6.67%)	(0.00%)	(20.00%)	(4.76%)	
大安推范雲	2/35	0/27	3/20	0/17	0/23	0/21	2/18	1/18	0/12	1/12	2/13	2/10	1/13	14/239	
(5.71%)	(0.00%)	(15.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(11.11%)	(5.56%)	(0.00%)	(8.33%)	(15.38%)	(20.00%)	(7.69%)	(5.86%)	
王丹网站 Wang	3/14	6/11	6/16	2/12	2/13	7/12	0/9	1/7	0/5	0/9	0/4	5/13	5/12	37/137	
Dan's Page	(21.43%)	(54.55%)	(37.50%)	(16.67%)	(15.38%)	(58.33%)	(0.00%)	(14.29%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(38.46%)	(41.67%)	(27.01%)
國民黨青年團	1/4	0/4	0/3	0/2	1/2	0/0	0/1	1/11	0/6	1/8	0/7	0/3	0/0	4/51	
(25.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(50.00%)	(-)	(0.00%)	(9.09%)	(0.00%)	(12.50%)	(0.00%)	(0.00%)	(-)	(7.84%)	
呂欣潔 Jennifer	2/32	2/25	0/19	0/18	1/16	1/16	0/18	3/11	0/17	1/17	4/11	3/11	1/15	18/226	
Lu	(6.25%)	(8.00%)	(0.00%)	(0.00%)	(6.25%)	(6.25%)	(0.00%)	(27.27%)	(0.00%)	(5.88%)	(36.36%)	(27.27%)	(6.67%)	(7.96%)	
管碧玲	2/24	3/23	0/12	1/16	2/11	2/13	3/7	2/10	0/13	1/13	0/3	1/7	1/9	18/161	
(kuanbiling)	(8.33%)	(13.04%)	(0.00%)	(6.25%)	(18.18%)	(15.38%)	(42.86%)	(20.00%)	(0.00%)	(7.69%)	(0.00%)	(14.29%)	(11.11%)	(11.18%)	
段宜康	12/35	10/19	13/21	9/11	11/14	9/14	7/9	10/12	8/11	8/11	7/13	11/14	9/14	124/198	
(34.29%)	(52.63%)	(61.90%)	(81.82%)	(78.57%)	(64.29%)	(77.78%)	(83.33%)	(72.73%)	(72.73%)	(53.85%)	(78.57%)	(64.29%)	(60.00%)	(62.63%)	
鄉民實業坊	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	5/9	3/11	6/12	3/7	6/10	23/50
(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(0.00%)	(55.56%)	(27.27%)	(50.00%)	(42.86%)	(46.00%)	
我是台灣人 (I am Taiwanese)	8/47	14/46	6/28	7/22	1/22	7/18	3/13	7/15	3/8	8/13	9/17	5/12	1/16	89/277	
林智堅	0/13	0/9	0/7	0/8	0/8	0/8	0/7	0/10	0/9	0/10	0/7	0/12	0/11	0/119	
(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	(0.00%)	
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
宋楚瑜	7/57	4/51	6/27	3/18	2/12	3/15	0/10	1/17	0/16	2/18	2/18	1/13	1/22	32/294	
(12.28%)	(7.84%)	(22.22%)	(16.67%)	(20.00%)	(0.00%)	(5.88%)	(0.00%)	(11.11%)	(11.11%)	(7.69%)	(4.55%)	(4.55%)	(10.88%)		
特急件小周的 人渣文本	19/25	13/20	9/12	8/11	4/5	6/7	5/9	6/8	3/10	9/10	4/6	9/13	4/6	99/142	
所有粉絲專頁	509/2355	446/1868	392/1507	305/1384	340/1303	384/1255	318/1179	373/1240	337/1204	308/1187	468/1244	311/1127	336/1275	4827/18128	
(21.61%)	(23.88%)	(26.01%)	(22.04%)	(30.60%)	(26.97%)	(30.08%)	(27.99%)	(25.95%)	(27.62%)	(27.60%)	(27.62%)	(26.35%)	(26.63%)		

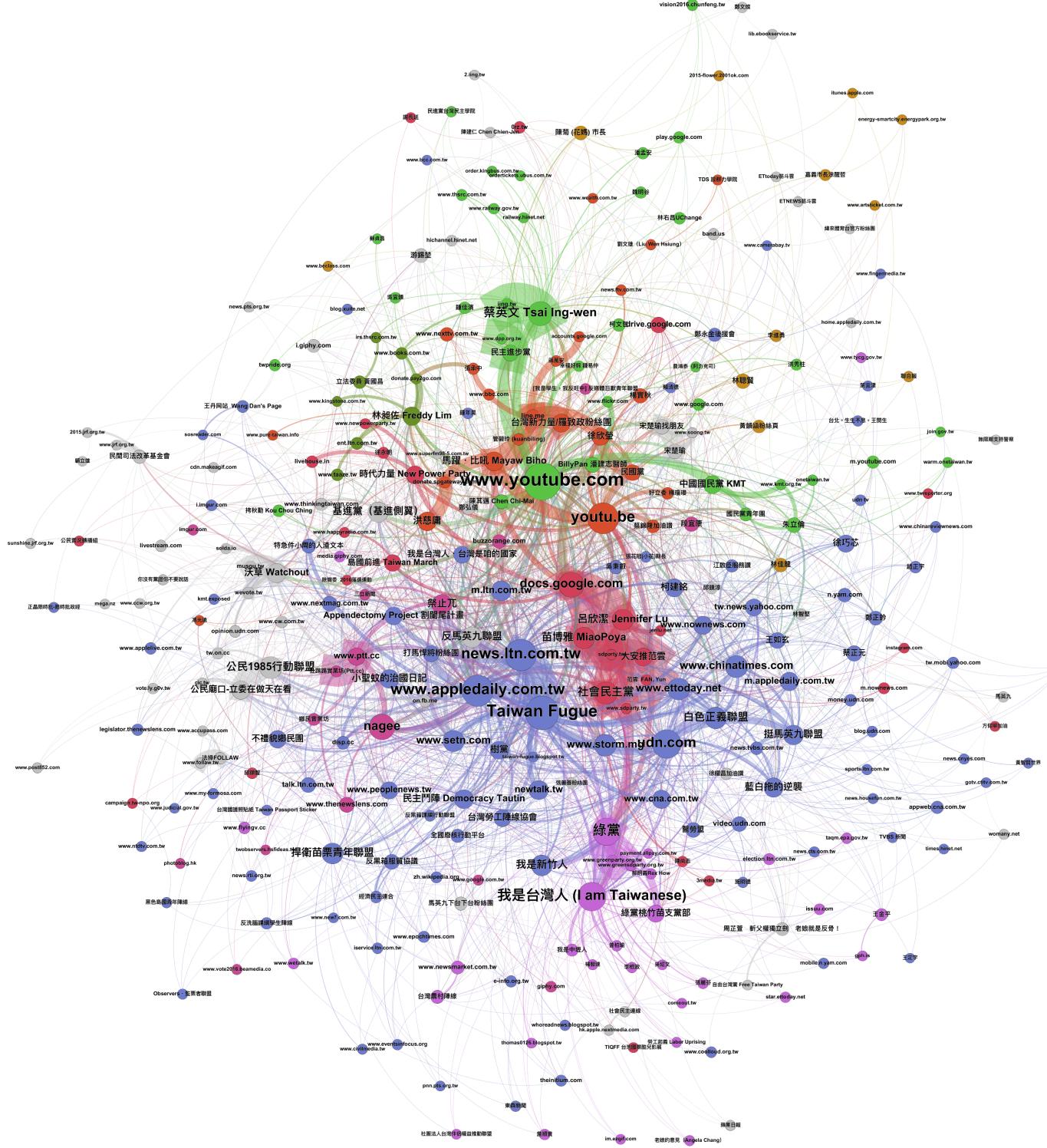


圖 4: 貼文／連結的「分享」網絡圖

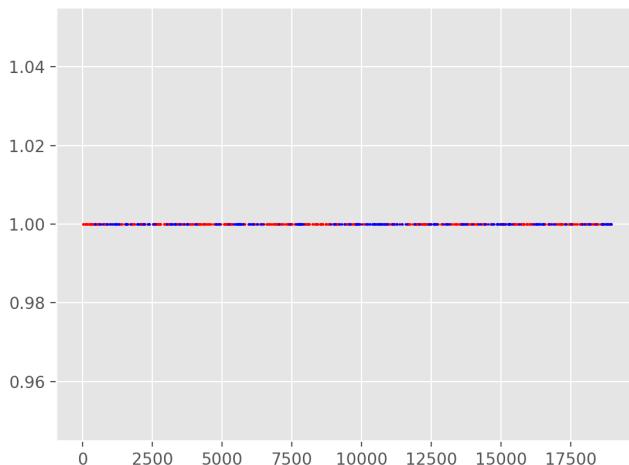


圖 5：訓練／測試貼文的分佈圖

表 3 找出了每個時間區間最熱門的關鍵字分別是哪些，本研究以一週為一個時間區間。透過此表，我們可以發現人們關注的焦點皆圍繞在「台灣」、「中國」、

「國民黨」上，唯有在 11 月 1 日至 11 月 8 日間，由於舉行兩岸領導人會面，「馬英九」、「馬習會」也躍升為社群網站上的討論焦點。

第二種分析方法的結果見圖 3，使用此方法能找出更多在不同時間發生的事件。單詞的出現率的高峰不一定是在事件當天，這是因為對於突發事件，可能要過一小段時間才會為衆人所知；對於預先計劃好的事件，可能在事前就已廣泛地被討論。

表 4 告訴我們本次選舉人們最關心跟「台灣」和「中國」有關的議題，而「國民黨」也是人們討論的焦點。值得一提的是，此次選舉的另一大黨，民主進步黨，並未出現在前十大關鍵字之中。我想這是因為選前的執政黨是中國國民黨，因此人們不但會討論他們發表的政見，更會躍躍地評論他們過去的政績。

情感分析 (*sentiment analysis*)

本項分析共有 948 筆訓練／測試貼文，經研究者手動標記後，有 651 筆被標記為說服型競選，佔 68.67%，有 297 筆被標記為攻擊型競選，佔 31.33%。圖 5 顯示訓練／測試貼文在所有貼文中確實是呈均匀分佈的。圖中的每個點皆代表一筆訓練／測試貼文，點的水平位置代表它在所有貼文中的位置。紅點代表說服型競選，藍點代表攻擊型競選。

由於攻擊型競選貼文的比例較低，而本項分析的目標便是要找出攻擊型競選貼文，故將攻擊型競選定義為陽性。經測試，本情感分類模型的準確度 (accuracy) 為 86.82%，查準率 (precision) 為 85.25%，查全率 (recall) 為 70.29%，F 度量 (F-measure) 為 76.75%。

表 5 是情感分析的結果，A/B (C%) 表示該粉絲專頁在該時間區間一共有 B 筆貼文，其中有 A 筆是攻擊型競選，佔 C%。本項分析不計入無內文之貼文。本研究發現，此次選舉只有 26.63% 的社群媒體貼文是攻擊型競選，其餘貼文皆是說服型競選。

三位總統候選人攻擊型競選貼文的比例更大幅低於此平均值，許多政治人物也都不常使用攻擊型競選。然而，具鮮明政治立場，或以抨擊反對陣營為目的的意見領袖，則較常使用攻擊型競選。「Taiwan Fugue」、「反馬英九聯盟」、「不禮貌鄉民團」、「藍白拖的逆襲」等粉絲專頁攻擊型競選貼文的比例皆超過八成。

對此，有一種可能的情形是，政治人物在其官方粉絲專頁上保持正面競選的形象，但在暗中卻支持其他非官方的粉絲專頁攻擊、批評其反對者。透過這種有意無意的「分工」，政治人物可以在表面上營造清新高尚的形象，同時也實現攻擊型競選的目標。這種競選策略並不罕見，不過此次選舉究竟是否有這樣的情形發生，有待進一步探討。

使用攻擊型競選的比例並未隨著時間而有太大的變化，唯有在 11 月 1 日至 11 月 8 日間，或許是因為舉行兩岸領導人會面的緣故，攻擊型競選貼文的比例創下高峰，比平時高出了一成左右。另外，使用攻擊型競選的比例則在投票日前一週達到最低點，比平時低了 5% 左右。

社會互動 (*social interaction*)

本部份探討的是社群網站上的「分享」網絡，分析使用者分享的貼文／連結的來源，用 Gephi 畫出社會網絡圖。本研究從全部貼文中，一共整理出 12,428 筆分享紀錄。結果見圖 4，圖中的一個圓點表示一個粉絲專頁或網站，一條邊從某個圓點以逆時針方向連到另一個圓點，表示某個粉絲專頁分享了另一個粉絲專頁／網站。圓點的大小表示有多少條邊連到這個圓點（即分支度為多少），邊的粗細表示分享次數。分群的結果以不同顏色表示。

從圖 4 我們可以發現一些規則，例如，「nagee」較常分享蘋果日報 (www.appledaily.com.tw) 及批踢踢 (www.ptt.cc) 的消息；「Taiwan Fugue」較常分享自由時報 (news.ltn.com.tw) 及蘋果日報 (www.appledaily.com.tw) 的消息；「白色正義聯盟」較常分享聯合新聞網 (udn.com) 的消息；「我是台灣人 (I am Taiwanese)」較常分享「綠黨」的消息等等。

另外，此圖也呈現出各政黨／候選人／意見領袖之間的關係。例如「苗博雅 MiaoPoya」、「呂欣潔 Jennifer Lu」和「大安推范雲」皆屬於同一個政黨，因此他們在圖中也屬於同一個社群。同樣地，「白色

表 6: 權威粉絲專頁排名

Label	Authority
Taiwan Fugue	0.245
我是台灣人 (...	0.194
白色正義聯盟	0.173
nagee	0.171
綠黨	0.166
挺馬英九聯盟	0.164
呂欣潔 Jennif...	0.161
民主鬥陣 De...	0.158
藍白拖的逆襲	0.155
社會民主黨	0.154
反馬英九聯盟	0.152

表 7: 匯集網站排名

Label	Hub
www.youtube.com	0.324
news.ltn.com.tw	0.289
www.appledaily...	0.288
udn.com	0.277
youtu.be	0.263
docs.google.com	0.200
www.setn.com	0.196
www.storm.mg	0.194
www.ettoday.net	0.188
www.chinatimes...	0.188
m.ltn.com.tw	0.163

「正義聯盟」、「挺馬英九聯盟」和「藍白拖的逆襲」彼此有相近的政治立場，因此他們在圖中的位置也比較相近。

表 6 和表 7 是 HITS 的分析結果。從 HITS 的分析結果，或從圖 4 中圓點的大小，我們還可以看出哪些消息來源最常被人們引用。在此次選舉中最常被人們引用的消息來源是 YouTube (www.youtube.com 或youtu.be)、自由時報 (news.ltn.com.tw)、蘋果日報 (www.appledaily.com.tw) 和聯合新聞網 (udn.com) 等等。

從這裡我們或許也能看出各消息來源在使用者心中的可信度 (credibility)。一個消息來源會被一個使用者引用，通常就代表該使用者信任該消息來源。因此，一個消息來源若引用頻率較高，或許也隱含著該消息來源在人們心中擁有較高的可信度。

五、結論

本研究發現此次選舉有 26.63% 的社群媒體貼文是攻擊型競選，與台灣、中國或中國國民黨有關的議題是人們最關注的焦點，且特定族群會傾向於分享特定來源的資訊。同時，本研究也提供了後續研究者一套有效地自動化分析線上政治活動的方法。

本研究做的只是初步的、嘗試性的分析，後續研究者還可以再做更深入的探究。譬如散播不實消息、操縱輿論「帶風向」等行為會對整個網路社群造成什麼樣的影響？網路社群是否容易被誇張聳動卻未經查證的消息煽動？此外，我認為以下幾點也值得進一步探討：

(1) 比較使用者於網路上對各政黨／候選人／意見領袖表達支持的人數與實際的選舉結果，考察社群網站上的輿論反應出社會大眾意見的程度。(2) 社會大眾意見極化之現象是否存在？該如何理解？我的推論是，有愈來愈多人使用社群媒體來接收政治新聞，然而台灣許多發佈政治新聞的粉絲專頁都有很鮮明的意識形態，報導

內容也相當偏頗。但是社群媒體的檢索演算法會將我們最可能感興趣的文章排序在最前面，因此抱持特定立場者會關注特定的粉絲專頁。機器用戶於粉絲專頁大量留言，或擁護自己的立場，或對反對者批評謾罵，強化對立並製造出擁有人數優勢的假象。這使得使用者對自己的立場更加堅定，缺少機會傾聽反對者的想法，人們意見更加分歧，難以達成共識。(3) 線上政治參與可能為台灣政治帶來哪些正面／負面影響？

要回答這些問題，可能還需搭配其他計算機科學以外的研究方法。例如設計問卷以調查使用者使用社群網站之習慣、張貼政治評論之動機等。

誌謝

首先，我要感謝我的指導教授孫春在教授，對我的研究給予了許多寶貴的建議。另外，我要感謝中央研究院資訊科學研究所中文詞知識庫小組，提供中文斷詞線上服務。有您們的協助，這份研究才得以順利完成。

參考文獻

- [1] A. O. Larsson and H. Moe, "Studying political microblogging: Twitter users in the 2010 swedish election campaign," *New Media & Society*, vol. 14, no. 5, pp. 729–747, 2012.
- [2] W. Zhang, T. J. Johnson, T. Seltzer, and S. L. Bidard, "The revolution will be networked," *Social Science Computer Review*, vol. 28, no. 1, pp. 75–92, 2010.
- [3] B. Hosch-Dayican, C. Amrit, K. Aarts, and A. Dassen, "How do online citizens persuade fellow voters? Using twitter during the 2012 dutch parliamentary election campaign," *Social Science Computer Review*, vol. 34, no. 2, pp. 135–152, 2016.
- [4] Socialbakers, "Definitive portal for social media statistics globally." <https://www.socialbakers.com/statistics/>. Accessed: 2016-09-28.
- [5] Likeboy FB Ranking, "台灣粉絲團總排名名單." <http://likeboy.tw>. Accessed: 2016-09-28.
- [6] 中央研究院資訊科學研究所中文詞知識庫小組, "中文斷詞系統." <http://ckipsvr.iis.sinica.edu.tw>. Accessed: 2016-10-08.
- [7] Y.-W. Pwu, "ckip-client: A Python client for CKIP Chinese Word Segmentation System." <https://pypi.python.org/pypi/ckip-client>. Accessed: 2016-10-08.
- [8] C. Leifer, "peewee: A small, expressive ORM." <http://docs.peewee-orm.com>. Accessed: 2016-10-08.
- [9] W. McKinney, "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in Science Conference* (S. van der Walt and J. Millman, eds.), pp. 51–56, 2010.
- [10] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing In Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
- [11] Plotly, "Plotly: Plotly for Python." <https://plot.ly/d3-js-for-python-and-pandas-charts/>. Accessed: 2017-11-19.
- [12] R. Wang (isnowfy), "SnowNLP: Simplified chinese text processing." <https://github.com/isnowfy/snownlp>. Accessed: 2017-02-16.
- [13] R. Mihalcea and P. Tarau, "TextRank: Bringing order into texts," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (D. Lin and D. Wu, eds.), (Barcelona, Spain), pp. 404–411, Association for Computational Linguistics, July 2004.

- [14] 謝明恩, 林立秦, 張逸群, and 蔡明娟, “P T T 軍情室,” *2016 Data Mining Final Project Slides*, no. 24, 2016.
- [15] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [16] D. Jurafsky and C. D. Manning, “Natural language processing: Lecture slides from the Stanford Coursera course.” <http://web.stanford.edu/~jurafsky/NLPCourseraSlides.html>. Accessed: 2017-02-18.
- [17] M. Bastian, S. Heymann, and M. Jacomy, “Gephi: An open source software for exploring and manipulating networks,” in *International AAAI Conference on Weblogs and Social Media*, 2009.