政治科學論叢 / 第 71 期 / 民國 106 年 3 月 / 頁 113-166

網路統獨的聲量研究：大數據的分析*

劉嘉薇**

摘　要

統獨議題是臺灣最重要的政治分歧，本研究透過 7,000 多個網路頻道大數據，分析網路上的統獨輿論。因爲統獨屬於較敏感的政治性議題，網路的匿名性也有助於探知敏感議題的意見表達。網路民意的大數據分析在國內政治學研究仍屬少見，本文也用了一些篇幅探討它的信效度。

本研究發現在統獨聲量的分布中，偏向獨立者高於統一。以「對象」而言，「中國」及「臺灣」被提及次數最多。在聲量來源方面，兩岸議題討論聲量中，高達 48% 集中在社群網站，其中單一網站聲量最多者爲 PTT。再進一步考慮到統獨議題的聲量情緒分析，平均每出現 6 篇正面情緒聲量時，同時會出現 10 篇負面情緒聲量。以上網路大數據的分析不同於民意調查的結果，網路大數據偏向獨立的聲量較高，維持現狀是民調主要的結果，但在網路的聲量很少，可能與網路具有匿名特質有關，讓網民亦在網上表達意見，值得重視。

---

* DOI:10.6166/TJPS.71(113-166)。本論文曾發表於「兩岸四地民意調查：現況與展望」學術研討會，國立政治大學選舉研究中心主辦，2015 年 11 月 07 日，臺北：政治大學。作者感謝評論人林聰吉教授的意見，以及兩位期刊匿名審查人和編輯委員會的建議。在論文寫作過程中也謝謝羅清俊教授與作者討論大數據分析的想法。當然，本論文若有疏漏，文責由作者自負。

** 國立臺北大學公共行政暨政策學系副教授，E-mail: jwliu@mail.ntpu.edu.tw。

收稿日期：105 年 6 月 20 日；通過日期：105 年 12 月 5 日

臺灣人學學術期刊資料庫

本研究分析網路大數據，而這與以民意調查進行統獨研究時，研究者想方設法希望挖掘「維持現狀者」骨子裡想法的做法異曲同工。不同於民意調查之處在於，民意調查以個體爲分析單位，網路大數據分析以聲量爲分析單位，兩者獲致不同結果，兩者都是民意，但當政府和政黨疾呼重視網民意見時，網路大數據呈現異於民調的結果，民調試圖理解「全民」的意見，但實際抽樣時可能缺乏年輕人的意見，網路大數據有助於理解「網民」的意見，兩者皆需做爲當局執政依據。尤其當網路論壇或社群媒體民意四起時，更彰顯網路大數據分析的價值。

關鍵詞：統一、獨立、統獨、大數據、網路

臺灣大學學術期刊資料庫

壹、前言

統獨議題牽動著臺灣政治和選舉的發展（王甫昌，2008；吳乃德，1992；1993；1997；2005；吳玉山，1999；2001；盛杏湲，2002；陳文俊，2003；游盈隆，1996；游清鑫等，2013；劉嘉薇，2016；Chu & Lin, 2001；Dittmer, 2005；Hsieh & Niou, 1996；Keng et al., 2006；Lin et al., 1996；Rigger, 2001；Wachman, 1994）。中國大陸因素對臺灣社會內部造成分歧，民眾分別主張統一、獨立或是維持現狀，但從 1990 年代以來，我們可以從民意調查資料觀察到，「支持統一者」在整體趨勢中漸少，「支持獨立者」與「維持現狀者」在整體中漸增（國立政治大學選舉研究中心，2017）。然民意調查多由電訪取得樣本，觸及的受訪者多爲訪問期間能在家中接聽電話者，對於晚歸或出外求學的年輕人較難以觸及。雖民調可以加權或插補（imputation）等方式彌補一些未能觸及民眾的意見。然年輕人卻是在網路上經常表達意見的族群（科技部臺灣傳播調查資料庫 1 期 3 次）（科技部，2017），因此對於網路上的統獨意見表達，有必要以其他研究方式瞭解。

綜上，統獨立場容許各種形式的務實抉擇，包括「先維持現狀，以後走向統一」、「先維持現狀，以後走向獨立」、「維持現狀，看情形再決定獨立或統一」或者「永遠維持現狀，不談統一或獨立」之類。因此，在民意調查當中，我們經常得到如此的選項，我們從民調資料瞭解到某些民眾隱身在維持現狀中，然而，在網路此一匿名的世界中，是否將可能使民眾表達出對於兩岸統一或獨立此一議題更眞切的想法？民眾統獨立場在網路大數據與民意調查中有何不同？因此以下我們將接續介紹爲何要以網路大數據探尋統獨聲量，以及大數據研究方法的特色。

大數據分析無意取代民意調查，民意調查的強項在於可以瞭解每位受訪者的基本人口背景，雖然我們「無法完全」取得大數據分析中每則聲量背後「發言者」的背景，但這也並非大數據的主要目的，研究大數據的目的在於瞭解「聲量」，每一個「個人」也可能有數則「聲量」，這是大數據研究的特色，重點在於「聲量」研究，而非「個體」研究。民調的分析單位是人，而網路大數據的分析單位是聲量，一個用戶有多筆聲量的情況自

然是網路大數據分析的特色，因爲我們非研究每一位公民的投票抉擇，自是不需要以個人爲分析單位，反之，網路聲量代表意見的方向和強度，也適合作爲公共政策的參考，試想一位不對公共政策表態的公民，在網路上沒有任何聲量的產出，自然對公共政策的決策影響較小，反之，政府念茲在茲的「婉君」，不就是他們在網路上發表的聲量也具有一定的數量和強度嗎？因此本研究除了關注網路聲量，也關注意見領袖的意見。

綜上，歸納本文研究目的：

1.民意調查爲抽樣研究，若有某些年齡層不足之處，便進行加權。然網路大數據分析的使用者多爲年輕人，可以補足民意調查年輕人不足之處（科技部臺灣傳播調查資料庫 1 期 3 次）（科技部，2017），這也是以大數據分析網路統獨的好處和目的。

2.在匿名的網路世界中，統獨立場是否更能被表達，亦即在匿名的狀況下，維持現狀者會減少嗎？

3. 不同於民意調查一個個體對統獨立場僅能表達一次意見，網路中一個個體可能發出多次聲量，進而「強化」某些意見的「量」，量大便可能值得重視，分析大數據的優點便在於能找出量有多大以及聲量的立場，這也是本文試圖探究的面向，甚至是可能的貢獻。

針對本文三項研究目的，以下第貳點首先說明在民調資料中民眾統獨立場的分類與分布，並加入以民調進行統獨研究的問題（針對研究目的 1）。第參點討論在「匿名」的狀況下，統獨此一「敏感」議題中的「維持現狀」者會減少嗎？（針對研究目的 2），以及以大數據研究方法運用於統獨研究的優點和信效度的討論（針對研究目的 3）。

貳、民眾統獨立場的分類與分布

在歷年的統獨研究中，儘快統一和儘快獨立這兩種極端意見較少，但儘快獨立都超過儘快統一。維持現狀此一趨中的意見在歷年都較多，而偏向獨立在 2003 年首次超過偏向統一。永遠維持現狀則幾乎是一路攀升（國立政治大學選舉研究中心，2017）。

臺灣本土學術期刊資料庫

統獨測量經反覆嘗試調整後，終於發展出在政治學界目前最常見的「統獨六分」測量，分別爲「儘快宣布獨立」、「維持現狀，以後走向獨立」、「維持現狀，看情形再決定統一或獨立」、「永遠維持現狀」、「維持現狀，以後走向統一」，及「儘快統一」。使用此分類的作者相當多，如 Liu & Ho (1999)、陳義彥與陳陸輝（2003）、陳陸輝與周應龍（2004）、張佑宗（2006）、陳陸輝、耿曙與王德育（陳陸輝等，2009a）、張傳賢與黃紀（2011）、楊婉瑩與李冠成（2011）、林瓊珠（2012）、游清鑫、林長志與林啓耀（2013）、劉嘉薇（2016）、臺灣選舉與民主化調查（TEDS）（科技部人文司，2017）。

然而，亦有學者以條件統獨探知民眾眞實的統獨態度，他們發現民眾在兩岸條件相當或中國大陸不攻打臺灣時，民眾願意選擇統一或獨立（陳陸輝等，2009a；2009b；耿曙等，2009；劉嘉薇等，2009）。再者，蕭怡靖與游清鑫（2012）亦感知到維持現狀者佔多數對調查結果的影響，嘗試在「保持單一面向」又能「有效拆解維持現狀」的前提下，利用正反面探詢的方式，重新建構民眾的統獨立場，更能有效拆解「維持現狀」者的統獨偏好。俞振華與林啓耀（2013）亦認爲，由於民眾統獨立場背後的價值衝突或矛盾立場程度不一，造成價值衝突或矛盾立場程度高的人對統獨議題表態時陷入立場矛盾的兩難困境（變異數大）。

一般而言，統獨議場的分布相當趨中，不過有些選民在某些條件下會偏向統一或獨立，例如兩岸各方面條件接近時或中國大陸不會攻打臺灣時，其中具有條件式統一的獨派選民在 2012 年總統選舉時也有部分投給馬英九，但在區域立委選舉中卻不見得投給國民黨候選人（林繼文，2015）。這的確顯示具有隱藏在維持現狀中，但實際上具有統獨立場的民眾，當眞實的統獨立場浮現時，便可能影響具有兩岸論述的總統選舉。

因爲條件統獨的設計可以減少維持現狀者的比例，說明了統獨六分的測量的確讓一些民眾隱藏眞實的統獨立場，因此討論匿名的網路統獨聲量有其研究價值。網路大數據爲何如此重要？Haire 與 Mayer-Schönberger (2014) 以及 Nickerson 與 Rogers (2014) 就指出，儘管巨量資料的出現對於經濟層面的衝擊較大，但其對於我們社會的影響力量也不容小覷，因其將有可能改變許多重要的政策決定。因此巨量資料並不僅是商業分析工具，也是進行社會研究的分析途徑之一。網路此一匿名的空間，可能讓民眾更願意表

臺灣大學圖書館期刊資料庫

示意見，更容易營造有威望的權力化身，得到個人賦權（吳筱玫，2003；陳萬達，2007；Peng, 2014）。

當然，以上述民意調查進行的統獨研究，已獲得一定豐碩的成果，然在本研究研究期間（2013～2014 年，容後再說明爲何選擇此一段時間）以及更早的 2012 年電訪民調資料中，加權前年輕人的比例少了許多，以「臺灣選舉與民主化調查」爲例，2012～2014 年歷次電訪的年輕人（20～29 歲）加權前的比例(A)、加權後比例(B)以及上述兩者相減的比例(A-B)如「表 1：歷次電訪 20～29 歲加權前不足比例」，從第一次到第十次電訪「20-29 歲加權前不足的比例」依序爲 -5.8、-4.0、-5.5、-3.2、-6.1、-5.4、-6.2、-3.8、-5.0、-6.0，說明了民意調查中年輕人比例偏低、不足，需靠加權調整。造成年輕人不足的原因與游清鑫（2016）提出的原因有關「年紀較輕的受訪者晚上不在家的比例較高」，而這正是電訪進行的時間。這些現象也說明了民意調查的問題，以及以網路大數據分析統獨民意的需要。

表 1　歷次電訪 20～29 歲加權前不足比例

電　訪　名　稱	執 行 日 期	20～29 歲 加權前比例 (A)	20～29 歲 加權後比例 (B)	20～29 歲 加權前不足比例 (A-B)
第一次總統滿意度電訪	2012/09/22-26	12.4	18.2	-5.8
第二次總統滿意度電訪	2012/12/19-23	14.4	18.4	-4.0
第三次總統滿意度電訪	2013/03/20-24	12.9	18.4	-5.5
第四次總統滿意度電訪	2013/06/26-30	15.1	18.3	-3.2
第五次總統滿意度電訪	2013/09/26-30	11.8	17.9	-6.1
第六次總統滿意度電訪	2013/12/13-12/17	12.2	17.6	-5.4
第七次總統滿意度電訪	2014/03/14-18	12.3	18.5	-6.2
第八次總統滿意度電訪	2014/06/6-6/10	13.8	17.6	-3.8
第九次總統滿意度電訪	2014/09/25-30	12.3	17.3	-5.0
第十次總統滿意度電訪	2014/12/11-15	11.2	17.2	-6.0

資料來源：作者整理自臺灣選舉與民主化調查（科技部人文司，2017）

臺灣大學學術期刊資料庫

參、以大數據研究方法運用於統獨研究

前一節所提的民意調查和本節所提的大數據分析都屬於資料科學（data science），只是前者強調民意調查技術，其資料是結構性的；後者強調資訊科學（computer science）於分析中的運用，其資料可以是結構性的，例如政府釋出的資料，每一筆數據排列得整整齊齊；但大數據也可能是非結構性的，非結構性的資料包括文字、圖片、影音。在方法論上。再者，民意調查多數以抽樣理論研究「樣本」，大數據則是直接研究「母體」，然此母體爲「可得的」母體，「可得的」意謂著某些數據雖然存在，但若無法取得，仍然無法進行大數據分析（劉嘉薇，2015）。一旦擁有大量或足夠的資料，便可以看出資料的分布趨勢，建立大量數據運作的模式，對比出資料的關係（Bertolucci, 2013）。透過這些比對，研究者能夠預測出未來資料發展的趨勢，進而做出更好的決策（廖洲棚等，2013；陳敦源等，2015；Jin et al., 2015）。

一、為何要以網路大數據適探尋統獨聲量？

網路上的統獨意見除了涵蓋一般民意調查 20 歲以上成年人的意見，更包括 20 歲以下族群的意見。因此我們要說明：爲何要以網路大數據探尋統獨聲量？包括網路匿名與統獨爲敏感議題兩大原因。

(一) 網路匿名

正因爲過去的研究已經顯示，統獨立場總隱藏在「維持現狀」的黑盒子中，網路的特性可能更適合探詢隱藏在維持現狀的統獨立場。Derlega 和 Chaikin (1977) 指出所謂的匿名性是一種他人無法明確得知這個人是誰的狀態，因此，自我隱私的界限就不會被明確地限制在一定範圍內，且在此種狀態下就容易發生自我揭露的情形。Wallace (1999) 指出，如果你無法得知我的眞實身分，我就不需要害怕你會對我進行報復行動，所以我就能暢所欲言，而這就是匿名性所帶來的好處。在發表較敏感性資訊時，往往傾向匿名，而每個人也都有權利這麼做。換言之，在網路上可以隱藏自己的眞實身分，所以使用者較願意表達自我的意見。Turkle (1995) 認爲，網路的

臺灣大學圖書館期刊資訊

匿名性使得使用者可以在網路互動上扮演任何一個你所想要表達的自我認同，甚至，有時網絡上的自我認同會和現實生活中的認同相互融合或是產生矛盾。Wallace (1999) 也認爲，網路的匿名性有助於民眾意見表達，因爲能無懼地說出自己所要表達的意見，網路匿名性保護了說話者的身分。匿名發言對政治論述帶來一定程度的貢獻。

(二) 敏感議題

至於敏感性問題在調查研究中有何問題，又可以用什麼方法克服？以下學者進行了許多努力，最主要是利用實驗設計找出受訪者對於敏感性問題（主要是對於民主實行、政權以及政府施政的看法）的眞實回答，同時發現年輕人在敏感性問題中容易吐露眞實想法，因此應該竭盡所能探詢年輕人的意見。Lei 與 Lu (2016) 以中國年輕人（大學生）和一般民眾爲樣本進行實驗設計，樣本以隨機方式置入實驗組和對照組，以面對面（face to face）方式進行調查研究，並設計敏感問題置入問卷中，敏感問題在本研究中也就是政治正確（politically correct）的題目，包括中國民主品質的問題以及對中國政治系統的信任。研究發現，在調查研究中，一般民眾對於這些題目較偏向政治正確的回答，包括對民主品質有較高的評價，以及對政治系統較信任。對年輕人來說，這些問題反而比較不敏感，他們較願意吐露眞實的評價，也就是對民主品質較低的評價，對政治系統有較低的信任。以上的研究發現，都是在控制其他變數情況下的結果，說明了敏感題目設計對不同世代的影響。本文研究認爲，正因爲年輕人較易吐露眞實的想法，然年輕人在電訪中的比例容易偏低（如第貳點後半的討論），因此分析網路大數據較能納入年輕人的想法。

與上述類似的研究成爲最近熱門的主題，Pan、Wu 與 Chang (2017) 訪問中國十個重要城市民眾來臺的感受，詢問中國民眾的題目屬於政治敏感的題目，研究發現大多數民眾對政府的態度與中國對兩岸關係官方宣傳的方向幾乎一致，而且對於臺灣民眾支持與中國統一有不正確的認知。Wu、Chang 與 Pan (2017) 亦發現，許多民意調查顯示，在中國共產黨的威權規則之下，民眾對中國政治體制具有高度的支持，對此類敏感題目的回答，多數民眾表示對中國政治體制的支持。然而，當分析 2011 年的「亞洲民主

臺灣政治學會出版

動態調查」（Asian Barometer Survey）資料時，發現民眾階級和對中國自由民主的偏好之間具有倒 U 型的關係，亦即仍能在中產階級身上，看到他們對民主的支持。

關於匿名或敏感性議題，最近更有許多討論。Rosenfeld、Imai 與 Shapiro (2016) 認為，當研究敏感問題，包括腐敗或偏見時，研究者越愈來愈依賴透過間接詢問的技術減輕直接調查的無反應（item nonresponse）。再者，Blair、Fair、Malhotra 與 Shapiro (2013) 研究巴基斯坦貧窮與支持「好戰政治」（militant politics）的關係時亦有類似發現。同樣地，Lyall、Blair 與 Imai (2013) 研究戰爭受害者對戰爭的態度，他們在阿富汗普什圖省塔利班叛亂中心的 204 個村莊進行的調查實驗，使用認可實驗間接地引出這個敏感問題的眞實答案。

類似的研究應用於 Blair、Imai 與 Lyall (2014) 的成果，認可實驗在社會科學中愈來愈受歡迎，可作爲敏感問題的間接測量技術。再者，Blair、Imai 與 Zhou (2015) 以類似方法，研究對奈及利亞（Nigeria）好戰組織的看法，藉著審查研究的標準化設計，開發應用於分析的各種多變量迴歸分析技術，達到敏感問題回答的最佳測試。

綜上，整體而言，在民意調查詢問政治敏感的題目，受訪者的回答無法反映出眞實情況，增加網路匿名性能讓網民更暢所欲言表達。以上已實質討論既有文獻發現的內容，說明網路聲量研究可以達到過去民意調查此一研究方法不足之處。

具體來說，本研究透過網路大數據（big data）分析，企圖瞭解整體「網路」統獨的趨勢。網路大數據分析的重要性在於，串聯人文與科技，將人文的議題以科技的方式研究，也將科技的演算納入人文社會脈絡的意義（Manovich, 2011）。網路大數據分析的重要性便在於補充民意調查長期發展的問題，藉此得知網路族群的意見，當然，民意調查以個體爲單位，網路大數據分析以聲量爲單位，兩者分析單位不同，然大數據分析方法的引進，能讓我們在「網路政治」的研究領域中，探析「婉君」怎麼想？對於民主政治民意表達的理解往前邁進。以民意調查得知支持獨立者與維持現狀者在整體中較多，網路世界也是如此嗎？因此我們以下將介紹大數據研究方法。

臺灣大學圖書館期刊資料庫

二、大數據研究方法

關於大數據研究方法，其定義和研究上的信效度，會是此類研究關注的問題。嚴格來說，大數據到目前爲止，還沒有一個統一的定義，大多數的說法是，「超過典型資料庫工具的硬體環境和軟體工具所能獲取、存儲、管理和分析能力者」即被視爲大數據或海量資料。更簡單地來說，「大數據」一詞指的是無法以傳統流程或工具所處理、分析的資料（胡世忠，2013）。「大數據」又稱作是「巨量（鉅）資料」或「海量資料」，是晚近幾年相當流行的名詞，因爲資訊科技的進步所衍發出來的新興概念，不僅改變了我們的生活方式，也改變了我們的思維模式。Haire 與 Mayer-Schönberger (2014) 就指出，儘管巨量資料的出現對於經濟層面的衝擊較大，但其對於我們社會的影響力量也不容小覷，因其將有可能改變許多重要的政策決定。是故，巨量資料並不僅是商業分析的工具，其也是未來進行社會研究的分析途徑之一。

關於大數據的定義，目前普遍的看法都認爲，其是源自於 3V 的概念。大數據的定義乃是以 3V 爲核心，即容量（volume）、速度（velocity）和多樣性（variety），與傳統資料相較，其資料來源是多元的，且大多是非結構性資料，也因更新速度極快，所以資料量龐大（Jordan & Lin, 2014；李欣宜，2015）。Haire 與 Mayer-Schönberger (2014) 也指出，大多的研究機構或其他企業組織多以高容量、高速度和高多樣性來描述大數據的特性，同時期還包含了高準確性以及高可靠性的意涵在裡頭。

關於大數據信效度的問題，以下包括大數據的大小和範圍，以及我們應該以因果關係或相關性看待大數據。再者，大數據處理的流程也與研究的信效度有關。Borgman（2015）進行了討論，包括大數據的「大小」（size）重不重要，無疑地，在研究範圍內蒐集所有的數據將可達到大數據分析的效果，亦即沒有抽樣的問題。而網路聲量的研究將可以取得比民意調查更多的觀察值。至於網路大數據蒐集「時間點」的問題，Borgman（2015）認爲所有人類在網路上的活動都可以視爲大數據的一部分，端視研究時間的界定，藉由研究者的分析，可將這些大數據的「來源」（sources）轉換爲「資源」（resources）。

即使有「大量的」數據，就一定具有測量效度嗎？Lazer、Kennedy、

臺灣大學學術期刊資訊庫

King 與 Vespignani (2014) 指出，數據量並不意味著可以忽略測量的基本問題，並構造數據之間的有效性、可靠性以及依賴性。核心挑戰是，已經受到廣泛關注的大數據不是那些具有可信度儀器的輸出結果。

那麼，大數據分析信效度核心的問題是什麼？Ekbia、Mattioli、Kouper、Arave、Ghazinejad、Bowman、Suri、Tsou、Weingart 與 Sugimoto (2015) 指出，目前對於大數據分析信效度的核心辯論集中在這是因果關係或相關性的分析。大數據研究並沒有消除以上科學史中主要和長期的困境，相反地，它已經重新定義和放大了它們。其中重要的是相關性問題（什麼被算入，什麼不被算入），有效性（validity）（如何有意義），這種研究也促使研究者超越傳統的量化和質性研究的二分法，大數據分析可能是量化的，也可能是質性的分析。

而大數據處理的流程也與信效度息息相關，Driscoll 與 Walker (2014) 認爲，鑑於社會科學方法的信度和效度取決於其透明度，大型社會數據（big social data）幾乎在封閉的商業組織中普遍產生。然而，沒有對基礎數據如何蒐集、存儲、清理和分析的清晰描述，讀者無法評估其效度，以上這些對於大數據的看法和分析過程的嚴謹度，都與信效度息息相關。

肆、研究設計

根據上述大數據的特性，本研究以 2013 年到 2014 年網路聲量爲研究範圍，之所以選擇 2013 年到 2014 年的網路聲量爲範圍，是因爲 2014 年是太陽花學運發生的那一年，此議題與兩岸經貿談判有關，因此本研究特別關心此一年兩岸統獨聲量。至於 2013 年，是太陽花學運的前一年，網路輿情有無相關情緒的醞釀，同樣值得關注。本研究分析依序包括兩岸議題聲量分析、統獨立場比較分析以及關注面向維度分析三大面向，委託意藍資訊股份有限公司，於 2015 年 6 月 1 日到 6 月 30 日，採用 OpView（社群口碑資料庫）進行分析，從大量網路輿論中擷取與關注與兩岸統獨相關內容，並加以儲存、歸類、分析；藉由網友產生的大量文本資訊挖掘其統獨立場意向。[1]

[1] 公司出具證明表示：意藍資訊股份有限公司（以下簡稱本公司），受劉嘉薇（以下簡稱委託人）委託，於 2015 年 6 月 1 日到 6 月 30 日，提供 OpView（社群口碑資料庫）

臺灣人文與學術期刊資訊網

OpView 包含全臺灣最熱門的社群網站、討論區、部落格等 7,000 個網站頻道，已涵蓋重要和大型頻道。之所以不選擇單一網站或論壇作爲分析範圍，是因爲選擇之後便不能稱爲大數據分析，而僅爲個案分析，即使 7,000 多個網站使用者背景相當多元，然這較接近網路世界的面貌，如果僅取其中幾個網站作爲分析範圍，將失去大數據的意義。其中社群網站的網站數有 6 個，頻道數（討論版數）有 3,978 個，較具有代表性的網站有 LineQ、Facebook 公眾訊息、[2] Facebook 關鍵意見領袖、FB 粉絲團、PTT 八卦版及其他版面、YouTube 關鍵字搜尋、Plurk 搜尋、Yahoo！知識＋……等。討論區的網站數有 86 個，頻道數（討論版數）有 2,115 個，較具有代表性的網站有 Mobile01、卡提諾王國、伊莉討論區、捷克論壇、SOGO 論壇、Fashionguide……等。部落格的網站數有 26 個，頻道數（討論版數）有 931 個，較具有代表性的網站有痞客邦 PIXNET、udn 部落格、PCHome 部落格、中時部落、Blog Spot、樂多日誌 blog……等。

在正式進入分析前，本研究首先定義大數據分析相關專有名詞的定義，這些名詞是進行大數據常用的分析模式（Haire, Mayer-Schönberger, 2014；Jordan & Lin, 2014）。包括聲量、頻道、情緒、正面情緒聲量、負面情緒聲量、P(positive)/N(negative) 比、維度、立場。網路大數據研究的分析單位不同於民意調查，民意調查的分析單位是個人，網路大數據的分析是聲量，包括聲量本身的分布、維度以及聲量的立場。以上分析的理論意義在於「聲量本身的分布」類似於以民調進行統獨研究的樣本基本描述；「維度」則類似以民調進行統獨研究的不同面向，比如文化上的對兩岸關係的看法、政治上對兩岸關係的看法，而「聲量的立場」類似於「統獨立場」的研究。「聲量的立場」與「維度」的相關也納入分析，類似過去以民調進行的統獨研究中，民眾在文化和政治上的統獨立場可能不同。以上這些分析面向皆能在過去的文獻中找到對應的部分。至於頻道、情緒、正面情緒聲量、負面

平台作爲「網路統獨：大數據的分析」分析之用。此文研究問題、理論建構、分析方向由委託人決定，資料亦由委託人自行詮釋，因此本公司不需列爲共同作者。

2 公眾訊息主要來源是素人或一般的網友，當發文的隱私設定選擇的是公開，則隱私程度最低，所有的好友或非好友都會看到這篇發文，這種情形下的發文，就會是所謂的公眾訊息。

臺灣人學社會科學資料庫

情緒聲量、P/N 比因爲與主要理論無直接關，因此移至附錄 5。

聲量意指文章篇數，主文與回文皆算在內。頻道爲一個網站的子討論板，如：PTT 的 Gossiping 板。情緒意指在一篇文章中，含有情緒詞數量多寡的量化分數指標，分數範圍爲 0～100；每篇文章皆含有正面及負面情緒兩種分數指標，因一篇文章中可能同時提及正面與負面情緒詞。正面聲量意指正面情緒分數超過門檻值（Positive>=30）的文章，內容提及正面用詞較多。負面聲量意指負面情緒分數超過門檻值（Negative>=30）的文章，內容提及負面用詞較多。

情緒分數的計算方式，爲針對文章內容用詞，透過正負情緒語料庫，並輔以字詞長度、文章長度、字詞頻率、各字詞的情緒權重，以文章爲單位進行正向與負向情緒分數的計算，該分數是經過加權計算及正規化之後的數值，尺度落在 0～100 之間。至於情緒判定的準確率，所使用的分類資料包含人工標記以及網友在特定網站發言時即須提供的情緒標記，舉例而言，PTT 的 Happy 及 Hate 版以及 Yahoo 新聞底下網友留言的心情標籤等；本研究使用此類情緒標籤，作爲衡量情緒分類模型準確度的判定依據。在反覆測試不同的門檻數值之後，最後可歸納出門檻值爲 30 時，模型準確率可達八成五以上的較佳表現（意藍資訊，2016）。

進一步而言，P/N 比可衡量特定議題或多篇文章所呈現情緒的平均數值，比值越高表示網友討論內容越正面，反之越負面。再者，維度意指統獨議題所包含的多種特性或面向，可與其他分析角度做交叉比對。立場意指網友使用文字所展現出來，對兩岸將走向獨立、統一或維持現狀的態度，本分析僅考慮此三種態度，由於維持現狀者較少，分析時將視資料數量省略維持現狀者的分析。

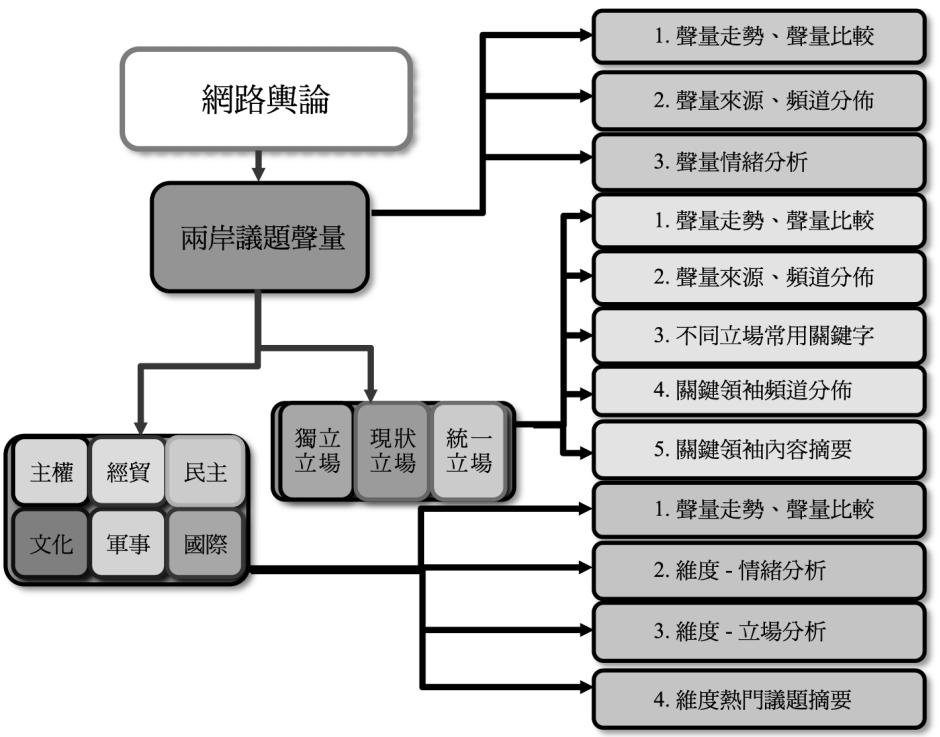
至於 OpView 的資料來源是否完整以及不同來源頻道資料是否適合加總在一起。依據本研究所使用的 OpView 資料來源，其資料蒐集機制遵循以下蒐集機制與維護方式，使用自動化程式，針對網路上公開、無設定隱私權限的網站與社群，進行結構與非結構資料的蒐集。在資料收錄之餘，OpView 亦有一定的資料維護方式。資料更新頻率從每小時至每日更新 3～4 次以上不等（意藍資訊，2016）。

在分析流程架構圖（圖 1）中，本研究依據上述 OpView 的範圍儲存網路

臺灣大學學術期刊資料庫

輿論中的兩岸議題聲量，首先將對兩岸議題聲量進行鳥瞰，包括整體聲量走勢和聲量比較的分析、聲量的來源以及頻道的分布，再者是聲量情緒的分析（上述亦即圖 1 右方第一區塊）。再者本研究將針對兩岸議題聲量區分成統一、獨立和維持現狀的聲量。區分出這三大立場後，本研究同樣先進行聲量走勢和聲量比較的分析、聲量的來源以及頻道的分布，並進一步分析不同立場常用的關鍵字。更甚者，在這三大立場中，關鍵領袖的頻道分布爲何？關鍵領袖發言的內容摘要又爲何？（上述亦即圖 1 右方第二區塊）除了統獨聲量的立場，網友討論統獨的實質內容又是那些？亦即討論的維度分布狀況。因此我們也先討論統獨議題聲量各維度的聲量走勢、聲量比較。除了維度（主權、經貿、民主、文化、軍事、國際，後續將說明爲何是這六個維度）的分布，我們也好奇維度與情緒關聯的分析、維度與立場關聯的分析，更甚者是維度中有那些熱門議題（上述亦即圖 1 右方第三區塊）。

圖 1　分析流程架構圖



資料來源：本研究自製

首先，什麼樣的聲量可以稱作兩岸議題聲量？本研究有四步驟蒐集網路聲量。第一步驟爲「定義相關聲量」，第二步驟爲「設定關鍵字距離」，以下將兩步驟合併說明較爲清楚。爲提高資料品質，限定網路輿論需同時提及「對象詞」（中國、大陸、內地、兩岸），及「關係詞」（獨立、統一、維持現狀、臺獨）者，才納入分析範圍內。爲避免無關文章納入，需限定【對象詞】及【關係詞】詞組之間的距離，本分析設定上限爲 10 字，[3] 範例如下：

【無關】:「……中國當局對於近期的克里米亞獨立一事……」
（【對象詞】及【關係詞】詞組之間的距離超過 10 字）

【有關】:「……許多人認爲兩岸已經不太可能統一……」
（【對象詞】及【關係詞】詞組之間的距離低於 10 字）

透過上述設定，已可獲得統獨議題聲量基本內容。本研究第三步驟「擴充關鍵字」，使用中文斷詞處理方式，分析出網友習慣用語。接著再將新詞組納入分析範圍，重新搜尋兩岸統獨議題聲量。新詞組包括：臺灣、臺澎、中華民國、中共、支那、祖國、北京、ROC、PRC。

最後，第四步驟還需要「排除雜訊」，亦即排除無關內容，主要以統一企業相關詞組爲主，文章中出現以下詞組者，不納入分析範圍：「統一獅、統一企業、統一阪急、統一布丁、獨家、免費」等（統獨聲量判斷最終關鍵字如附錄 1）。本分析是使用關鍵字設定做爲資料篩選的門檻，但因爲人類語言的使用情境及前後脈絡都很複雜，無法只靠關鍵字，就能在大量文字資料當中，精確的找出完全符合的資料。因此必須依賴人工方式，對資料進行校對，找出雖符合篩選條件但卻是雜訊的文章。在本次分析中，我們在每個月隨機抽取 100 篇文章，人工判斷是否與統獨議題有關。若爲無關者，則條列出可排除雜訊的關鍵字。

雜訊類型有二大類，一是廣告文，那就會設定廣告文常用用語，本次分析的設定爲：獨家、尋參、細商、考資、品規、尋結、品網、貸、搶購、細介、品訊、買網址、特惠商、免費、成人、色情、線上學英文。二是其

[3] 設定關鍵字距離爲 10 字是同時考量資料豐富度及精準度之後，所完成的設定。這部分的設定，會隨著議題或領域的不同，而有不同的字詞距離設定，可以參考「詞語共現距離」（co-occurrence window size）概念的相關文獻，例如：Indurkhya 與 Damerau (2010)。

臺灣人文學術期刊資料庫

他議題，但用語恰好相同者，在本次分析中，最常出現的是與統一企業相關的食安討論，所以本次分析的設定爲：統一獅、統一集團、統一企業、統一公司、統一超、統一阪急、統一食品、統一麵、統一泡麵、統一布丁、統一超。最後分析結果如下：

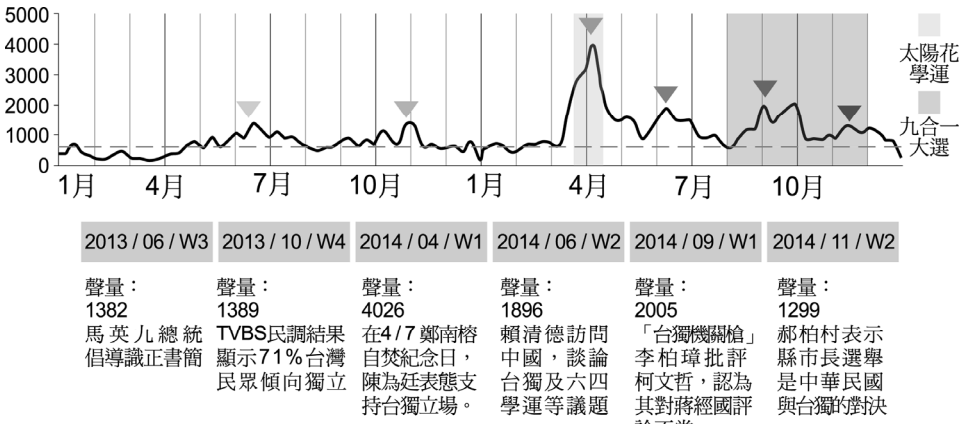
伍、研究結果

依據分析架構，本研究主要分析結果包括：整體統獨聲量、統獨聲量所指對象分析、統獨聲量來源與頻道分析、統獨聲量立場分析、統獨聲量關鍵字分析，以及統獨聲量維度分析，分述如下，次要分析結果則置於附錄 5。

一、整體統獨聲量

在整體聲量走勢方面，從 2013 年到 2014 年爲止，網友針對兩岸統獨議題的討論聲量，總計有 101,912 篇（不包括新聞）。每週平均聲量 961 篇，其中 W1 代表第一周，其他依此類推。其中有六大高峰超出平均（圖 2 線條所示高度），2013 年有 2 個，2014 年有 4 個。各聲量高峰的主要來源皆是轉錄新聞消息，引起大量網友議論或發表意見。2014 年太陽花學運時期爲聲量最大高峰，並持續發酵至當年 5 月中旬；其次爲九合一大選前「臺獨機關槍」李柏璋批評柯文哲，認爲其對蔣經國評論不當。

圖 2　兩岸統獨議題討論聲量



資料來源：本研究自製

臺灣大

二、統獨聲量所指對象分析

再以「對象」而言，表 2 整理了中國大陸相關字詞分布狀況，表 3 整理了臺灣相關字詞分布狀況，表 2 和表 3 百分比計算方式皆爲次數除以總篇數 101,912（非次數除以總次數），因此百分比加總大於 1，「中國」及「臺灣」被提及次數最多，逾 8 萬（表 2）及 17 萬次（表 3）。整體而言，每篇提到中國大陸相關字詞的平均次數爲 1.6 次（表 2），每篇提到臺灣相關字詞的平均次數爲 2.3 次（表 3）。而「關係」則是描述兩岸應有的關係，其中提及「獨立」（表 4）爲 43,263 次，提及「統一」爲 35,063 次，提及「獨立」較「統一」多了 8,000 多次，每篇提到兩岸相關字詞的平均次數爲 1.1 次（表 4）。

表 2　中國大陸相關字詞分布狀況

	次　數	比　例
中國	81,126	0.8
大陸	43,525	0.4
中共	20,718	0.2
北京	7,755	0.1
祖國	2,308	0.0
支那	1,810	0.0
內地	830	0.0
相關字詞總提及次數	158,082	1.6 次／篇

資料來源：本研究整理

表 3　臺灣相關字詞分布狀況

	次　數	比　例
臺灣	176,621	1.7
臺獨	31,601	0.3
中華民國	21,932	0.2
相關字詞總提及次數	230,154	2.3 次／篇

說明：台灣及臺灣合併計算。
資料來源：本研究整理

臺灣學術期刊資料庫

表 4　兩岸相關字詞分布狀況

	次　數	比　例
獨立	43,263	0.4
維持現狀	2,429	0.0
統一	35,063	0.3
兩岸	30,883	0.3
相關字詞總提及次數	111,638	1.1 次／篇

資料來源：本研究整理

三、統獨聲量立場分析

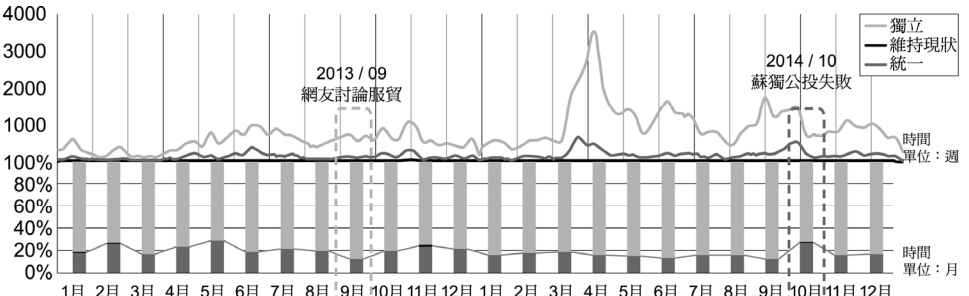
除了統獨議題情緒聲量的分布，統獨的「立場」又是如何？主要立場判定方式是「短句判斷法」，網友對兩岸議題發表意見時，也會表態出個人統獨立場；當意見長度較短時（本分析設定爲 50 字內），其情緒通常也會對應對統獨的好惡。因此可依據提及「獨立、統一」和正、負面情緒出現的四種（2×2）情形進行立場的判斷。輔以「習慣用語判斷法」，若網友討論兩岸議題時帶有特定立場，通常會出現特定慣用用語，尤其以蔑稱相對立場者用詞居多，網友習慣用語列表如附錄 2。例如獨立的習慣用語包括：蛆蛆、支那、舔共、阿共、共諜等，統一的習慣用語包括：吱吱、皇民、民粹、學匪、臺巴等。同時採用「主回文對立判斷法」，網友針對具有特定立場的主文章，時常會留言或回覆來加入討論，且內容會帶有正面或負面情緒，此時可透過該情緒來判斷回覆網友的統獨立場。再加上「機器學習判斷法」，由於前三個判斷方法（短句判斷法、習慣用語判斷法和主回文對立判斷法），僅能處理 40% 左右聲量資料，剩餘 60% 聲量資料，需透過資料探勘技術，使用 Naïve Bayes 分類演算法進行自動化立場判斷，此處理方式準確率可達到 70% 左右。上述四種判斷方法，都會爲每筆聲量判斷統獨立場，經驗證過程發現，簡單投票決（判斷爲獨立者，多於判斷爲統一者，即爲獨立，反之亦然）的準確率可達到 80% 以上。

以標題「國民黨慘敗竟幫了北京一大忙：統一大局已定」爲例，其內容摘要爲「很好笑！夠腦殘了！如果這是支那人寫的，就不意外。」關鍵

字爲「支那」，以四種方法判定此聲量的統獨立場，第一種方法爲「短句判斷法」，關鍵字爲統一，情緒正負面皆有（正面情緒＞＝30，負面情緒＞＝30），因此將立場判斷爲「維持現狀」；第二種方法爲「習慣用語判斷法」，因爲關鍵字爲「支那」，便判斷立場爲「獨立」；第三種方法爲「主回文對立法」，此聲量主文關鍵字爲「統一」，但回文情緒爲負面，亦即反統一，因此將立場判斷爲「獨立」；最後一種是「機器學習判斷法」，亦即透過以上三種方法的，讓機器學習，前述三種方法便是給機器「訓練教材」的概念，之後再給機器其他聲量，機器便可過之前的學習，自主判斷聲量的統獨立場。判斷結果爲「獨立」者有爲 3 種方法，維持現狀票者爲 1 種，統一者無，根據上述結果，以多數判斷結果爲結果—「獨立」，以上僅爲一例，其他例子以此類推。

透過前述方法，可定義出各篇聲量的統獨立場，而支持獨立的聲量明顯多於支持統一者。從圖 3 可以觀察，在這兩年，網路聲量支持獨立的比例穩定維持在八成左右，2013 年 9 月聲量雖較少，但該月有出現服貿相關的新聞，有部分網友因此而持有終將會統一的看法，引起諸多網友批評，使得支持獨立的聲量佔比，達到 88% 的最高峰。此外，2014 年 10 月，可能由於蘇格蘭獨立公投失敗，在網路上也出現較多認清現實的討論，因此使網友支持統一的聲量上升至 27% 的最高點。

圖 3　網友統獨立場聲量走勢及各月統獨立場比例變化



	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
獨立	82.2	73.4	84.0	77.1	71.7	82.2	78.7	81.3	88.1	80.6	75.8	78.9	84.3	82.7	82.2	84.9	86.0	87.1	84.5	84.5	88.0	72.8	84.9	83.9
維持現狀	0.3	1.0	0.0	0.0	0.3	0.0	0.2	0.0	0.0	0.0	1.4	0.7	0.0	0.0	0.1	0.1	0.0	0.1	0.1	0.0	0.1	0.9	0.2	0.0
統一	17.6	25.7	16.0	22.9	28.0	17.8	21.1	18.7	11.9	19.3	22.8	20.4	15.7	17.3	17.7	15.0	14.0	12.9	15.4	15.5	11.9	26.2	14.9	16.1

單位：%

資料來源：本研究自製

臺灣大學

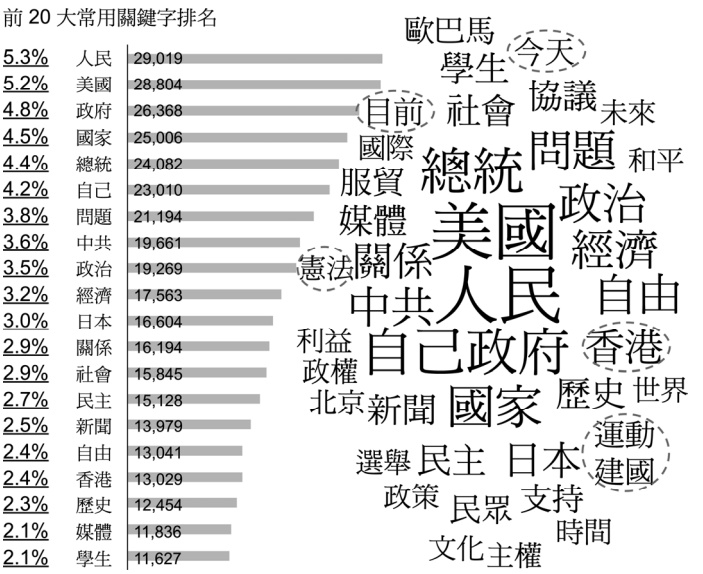
四、統獨聲量關鍵字分析

那麼，在這些統獨議題聲量的討論中，熱門討論關鍵字爲何？我們從圖 4、圖 5 發現，不論立場爲何，統獨雙方在論述時，最常提及的字詞包括以下幾項，一一解讀之。這些詮釋的意義在於了解網路上在討論統獨時涉及那些關鍵字，抓取了關鍵字即能掌握統獨討論的重點，至於這些關鍵字出現在統一聲量和獨立聲量則是不同的比例。

- 人民（獨立：29,019，5.3%；統一：7,002，6.1%）：較多的聲量在聲稱「人民」時偏向獨立，然而在統一聲量中使用「人民」爲關鍵字者較獨立聲量比例多。
- 政府（獨立：26,368，4.8%；統一：5,700，5.0%）：較多的聲量在聲稱「政府」時偏向獨立，然而在統一聲量中使用「政府」爲關鍵字者較獨立聲量比例多。
- 國家（獨立：25,006，4.5%；統一：6,944，6.1%）：較多的聲量在聲稱「國家」時偏向獨立，然而在統一聲量中使用「國家」爲關鍵字者較獨立聲量比例多。
- 自己（獨立：23,010，4.2%；統一：5,146，4.5%）：較多的聲量在聲稱「自己」時偏向獨立，然而在統一聲量中使用「自己」爲關鍵字者較獨立聲量比例多。
- 美國（獨立：28,804，5.2%；統一：4,574，4.0%）：較多的聲量在聲稱「美國」時偏向獨立，且在統一聲量中使用「美國」爲關鍵字者較獨立聲量少，可見使用美國一詞時，很可能是指向獨立聲量。
- 日本（獨立：16,604，3.0%；統一：4,481，3.9%）：較多的聲量在聲稱「日本」時偏向獨立，然而在統一聲量中使用「日本」爲關鍵字者較獨立聲量比例多，與上述使用「美國」爲關鍵字的情況不同。

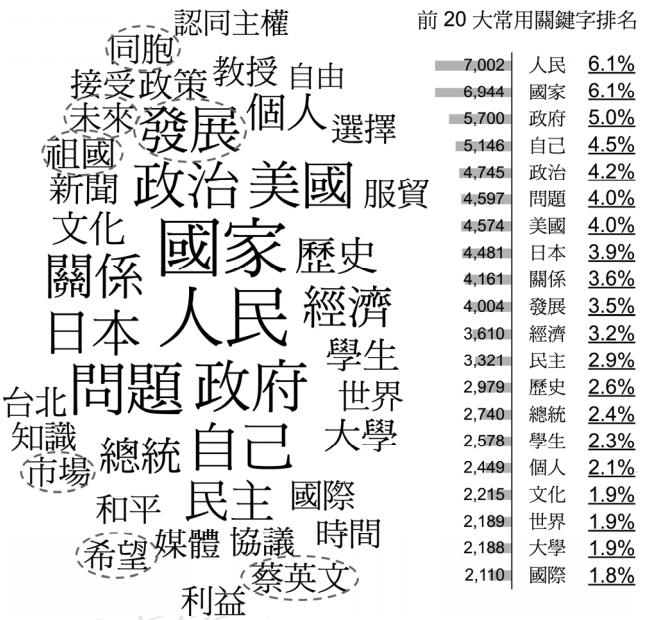
臺灣大學學術期刊資料庫

圖 4　獨立立場常用字



資料來源：本研究自製

圖 5　統一立場常用字



資料來源：本研究自製

臺灣大學

五、統獨聲量維度分析

在維度方面，本研究分析的維度包括主權、經貿、民主、文化、軍事和國際。之所以分析這些維度，乃因前述分析也發現，網路媒體討論的議題有與新聞事件相同之處，本研究曾進行報紙統獨議題的內容分析（劉嘉薇，2014-2016），因此雖本研究在分析前未能瞭解網路統獨聲量的維度，但仍能透過報紙內容分析，發現這六大維度是新聞報導兩岸新聞最主要的維度（如附錄 3），既然網路討論議題具有隨著新聞報導走的趨勢，新聞報導的維度將有助於瞭解網路統獨聲量可能具有的維度。根據新聞報導維度爲基礎，本研究再加入網友慣用用語，形成網路維度關鍵字，如附錄 4，其中主權維度聲量有 43,173 篇，佔 42.4%。經貿維度聲量有 12,883 篇，佔 12.6%。民主維度聲量有 23,660 篇，佔 23.2%。文化維度聲量有 24,626 篇，佔 24.2%。軍事維度聲量有 20,582 篇，佔 20.2%。國際維度數量有 4,507 篇，佔 4.4%。

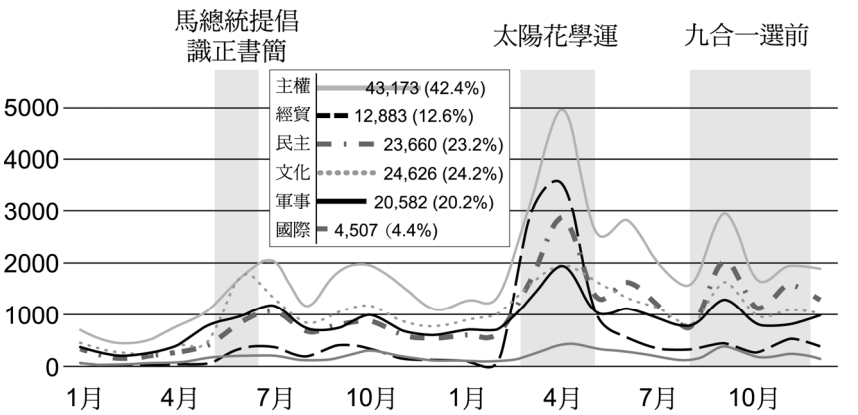
整體而言，根據圖 6，這兩年期間當中，網友在討論統獨議題時，主權始終是最關注的維度。除國際議題外，其餘五維度在學運及九合一選前期間，皆有明顯漲幅，經貿議題的討論量在學運時暴增，而後回復至原先的較低水準。民主、文化及軍事等三維度聲量相近，且受事件刺激的漲跌程度類似。民主議題自 2014 年後的漲幅相對較大；除學運期間，國際與經貿維度的聲量表現相近。

而在維度情緒分析方面，根據圖 7、圖 8，網友討論兩岸議題時，整體有出現正負面情緒的情緒顯著比爲 68%（圖 7），[4] 而 P/N 比則爲 0.61（圖 8）。網友討論這些維度時，比起整體討論，更容易帶有情緒，但情緒稍稍正面一些（六個維度的顯著情緒比都高達 80% 左右，且 P/N 比都超過 0.65）。然網友在討論主權時，有「不鳴則已，一鳴不爽」的情況，討論時相對冷靜不帶情緒，但若有情緒，也是負面居多（在六維度、兩年的時間中，「主權」維度的「顯著比」和「P/N 比」都最低）。此外，網友在 14 年討論經貿議題時，比 13 年冷靜且比較樂觀（「經貿」維度顯著比下降，P/N 比上升）。

4 情緒顯著比＝1－（正負面情緒皆不顯著聲量／所有聲量）。

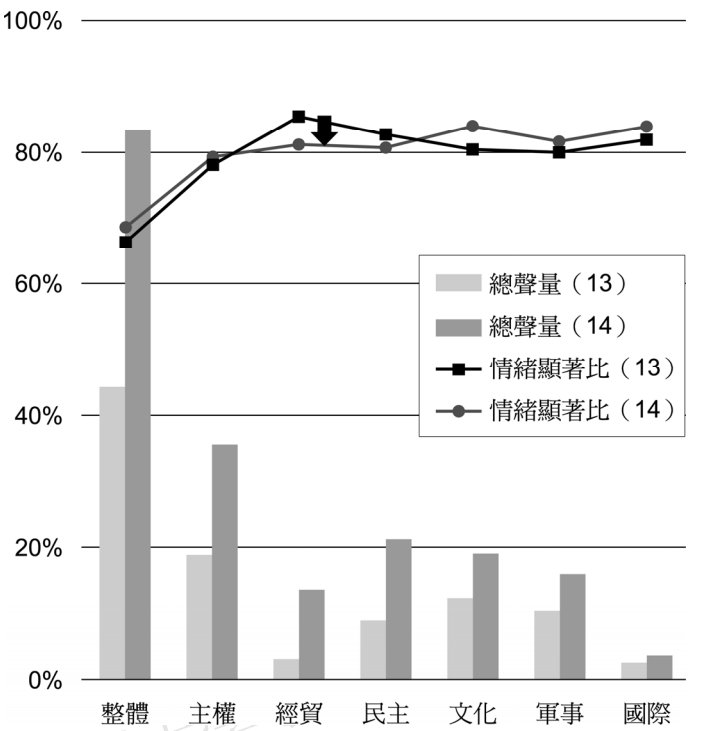
臺灣大學學術期刊資料庫

圖 6　統獨聲量維度趨勢圖



資料來源：本研究自製

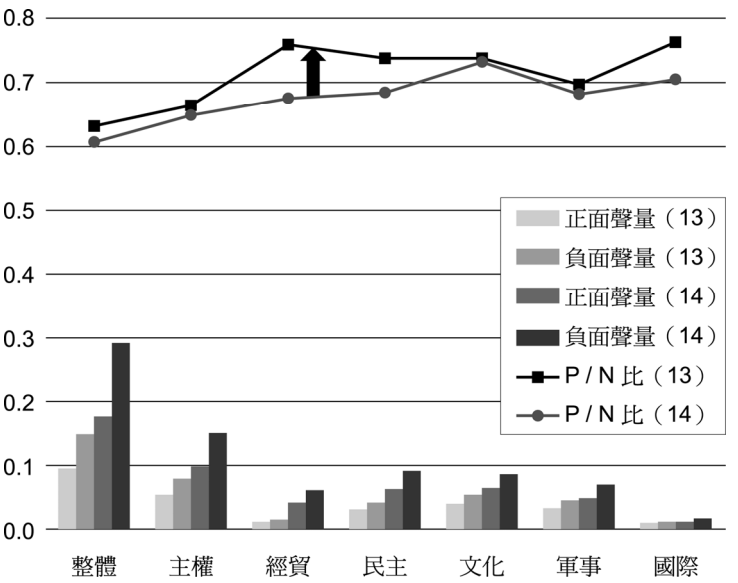
圖 7　2013-2014 年之聲量及情緒顯著比



資料來源：本研究自製

臺灣大學

圖 8　2013-2014 年正負面情緒及 P/N 比



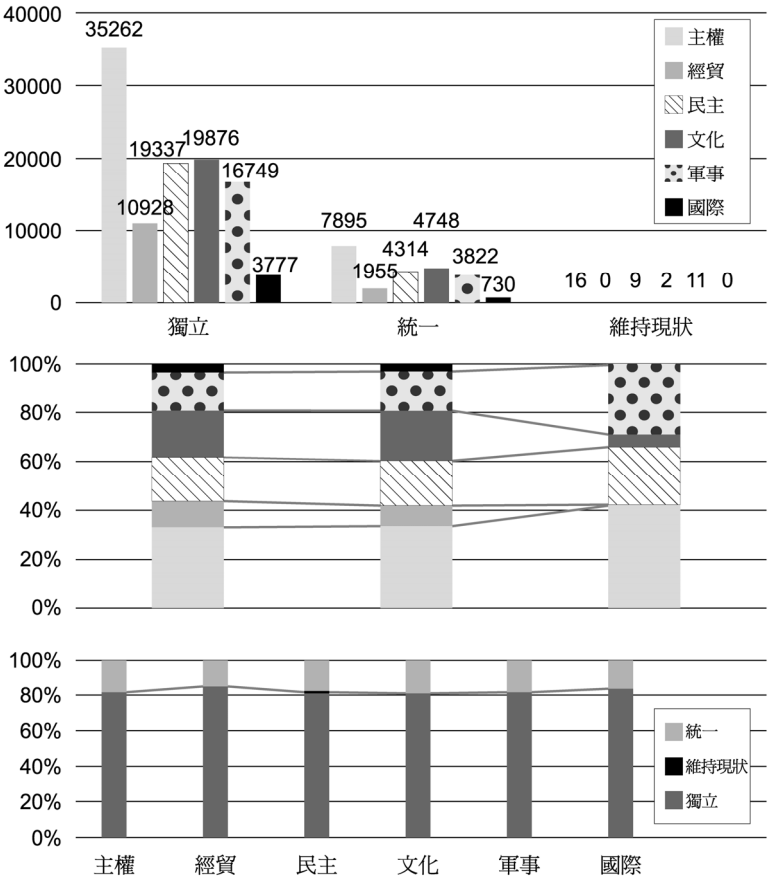
資料來源：本研究自製

若以統獨立場在各維度的聲量來看，由圖 9 比例分布可知，不論網友統獨立場爲何，其在各維度討論的比例，幾乎沒有差異（圖 9 的上圖和中間圖）。相反地，網友討論各個維度時，統獨立場的比例都與整體立場比例接近（圖 9 的下圖），而維持現狀立場的討論雖少，仍會關注主權等，但卻完全沒有經貿及國際。統一立場者相對考慮較多維度，其提及維度關鍵字平均爲 1.3 個（統一立場共有 18,018 篇，六維度在其中共出現 23,464 次，每篇出現 1.30 次維度關鍵字），略多於獨立者的 1.26 個（獨立立場共有 83,721 篇，六維度在其中共出現 105,929 次，每篇出現 1.27 次維度關鍵字）。需要特別說明的是，各篇文章提及維度可能超過一個，加總各維度數量後，即代表提及所有維度數。

在中文語言中，脈絡是由大量的中文字以非結構化資料組成，而非結構化的資料，因此具有脈絡意義，包括網路聲量的頻道來源差異、網路社群本身的立場與屬性等皆有其脈絡意義，因此除了量化分析，質性分析也可輔助非結構化資料統獨立場的判讀。

臺灣大學圖書館期刊資料庫

圖 9　立場－維度聲量分布



資料來源：本研究自製

以下我們將討論各維度的討論狀況，以質性資料的方式呈現。在討論主權議題的網路口碑中，提及兩岸統獨立場的內容多數在爭論：不同的稱呼方式與界定方式，是否隱含了網友對於臺灣主權在當前 / 未來是否從屬於他者的認知或期望。在主權維度上，我們挑選回文則數較多的標題檢視，以「臺灣國」搜尋到「TVBS 民調 71% 希望臺灣獨立」這則標題中網友 aids790202 認爲，「倡議成立臺灣國的臺獨人士，身爲中華民國的國民卻不承認中華民國，是恥辱」，回文共有 637 則。以「族群」搜尋到「臺獨機關槍大爆發」這則標題中網友 mmt0827 認爲，「眞夜自承，他的國家認同是

臺灣，不是中國。臺灣歷史是經歷過荷蘭人、西班牙人、漢人、滿人、日本人、支那敗犬這些人統治的歷史，不是中國史」，回文共有 479 則。以「臺灣人」搜尋到「馬英九總統居然打算推動臺灣寫簡體字」這則標題中網友螞蟻吳認爲，「自己生是臺灣人、死是臺灣鬼，因此只寫繁體字。在這個討論串裡，寫繁體字或簡體字，分別被指涉身分爲臺灣人或中國人」，回文共有 471 則。以「中國人」搜尋到「臺獨的悲哀每下愈況日趨下流」這則標題中網友 829102402 自稱，「自己是臺灣人，也是中國人。在此討論串中，網友自稱爲中國人，隱含了網友對於兩岸未來將趨向統一的認知或期望」，回文共有 449 則。以「一邊一國」搜尋到「從香港看臺灣顏銘緯：獨立才是唯一活路」這則標題中網友 georgetan5168 質疑，「臺獨人士所使用的獨立、一邊一國、臺灣國等詞彙，背後的意義不同，暗示臺獨人士態度丕變」，回文共有 341 則。

在討論經貿議題的網路口碑中，提及兩岸統獨立場的內容多數在爭論：反對兩岸服務貿易協議，背後是否隱含了希望臺灣獨立的政治態度。在經貿維度上，我們挑選回文則數較多的標題檢視，以「服貿」搜尋到「TVBS 民調 71% 希望臺灣獨立」這則標題中網友 900108 指出，「臺灣與大陸之間的聯系無法切斷，服貿協議也不能退回，加上還有 200 多萬名臺商在大陸，因此臺灣獨立不可行」，回文共有 637 則。以「TPP」搜尋到「從香港看臺灣顏銘緯:獨立才是唯一活路」這則標題中網友 swag88 認爲，「TPP（跨太平洋戰略經濟夥伴關係協議）是比「糖衣毒藥」更毒藥的協議，呼籲臺灣要小心美國。口碑顯示，網友在意美國對兩岸政治往來的看法」，回文共有 348 則。以「服貿」搜尋到「自稱無畏統媒、馬政府抹黑陳爲廷表態：我主張臺獨」這則標題中網友 GBA12963 質疑，「網友太陽花學運參與者反對服貿協議，是民進黨所帶領的，且反對服貿協議的目的是要增加臺獨勢力」，回文共有 340 則。以「服貿」搜尋到「我叫林飛帆，我主張臺灣獨立」這則標題中網友林語宸質疑，「反對服務貿易協議的人，背後帶有希望臺灣獨立的政治立場」，回文共有 316 則。以「服貿」搜尋到「美敲打學運再搞臺獨就棄臺」這則標題中網友 redmoon 認爲，「臺灣爲了在夾縫中求生存，應該澄清反服貿不等於支持臺獨。網友認爲在中、美兩國環伺下，臺灣沒有辦法臺獨」，回文共有 315 則。

臺灣大學圖書館期刊合作版權

在討論民主議題的網路口碑中，提及兩岸統獨立場的內容多數在爭論：採取民主方式如公投、修憲、政黨政治，能否讓臺灣走向獨立時，避免受到其他國家的武力威脅。在民主維度上，我們挑選回文則數較多的標題檢視，以「公投」搜尋到「TVBS 民調 71% 希望臺灣獨立」這則標題中網友 esutiya781119 認爲，「民主方式如公投、修憲都無法讓臺灣獨立，因爲臺灣沒有足夠的軍事力量，無法抵抗中國大陸的威脅」，回文共有 637 則。以「公投」搜尋到「馬英九總統居然打算推動臺灣寫簡體字」這則標題中網友 f121394745 認爲，「若執政當局要更改臺灣官方使用文字，應由全民公投決定」，回文共有 471 則。以「人權」搜尋到「臺獨的悲哀每下愈況日趨下流」這則標題中網友 ssngn 認爲，「臺灣是主權獨立的國家，ssngn 還諷刺大陸網友，『回去管管中國人的人權吧』」，回文共有 449 則。以「政黨」搜尋到「從香港看臺灣顏銘緯：獨立才是唯一活路」這則標題中網友色之水呼籲，「支持民主政黨蔡英文主席，就能夠建立獨立且強大的臺灣人民共和國」，回文共有 341 則。以「政黨」搜尋到「自稱無畏統媒、馬政府抹黑陳爲廷表態：我主張臺獨」這則標題中網友 oscarfym 批評，「參與反對兩岸服務貿易協議的太陽花運動的學生，背後其實是以加入民進黨爲目的」，回文共有 336 則。

在討論文化議題的網路口碑中，提及兩岸統獨立場的內容透露出：兩岸文化中差異的部分，如風俗民情、簡體／繁體字，被獨立立場的網友作爲兩岸無法統一的理由。在文化維度上，我們挑選回文則數較多的標題檢視，以「風俗民情」搜尋到「TVBS 民調 71% 希望臺灣獨立」這則標題中網友 kilua7679 認爲，「兩岸風俗民情差異太大，是兩岸無法統一的原因之一」，回文共有 637 則。以「殘體字」搜尋到「馬英九總統居然打算推動臺灣寫簡體字」這則標題中網友 ken chan 批評，「簡體字是殘體字，呼籲臺灣的同胞絕不能接受殘體字」，回文共有 471 則。以「二二八」搜尋到「從香港看臺灣顏銘緯：獨立才是唯一活路」這則標題中網友 nono2000 認爲，「二二八沒有發生，受害者多數都是小偷犯罪之人。是政治故意炒作才會出現這個歷史錯誤」，回文共有 341 則。以「陸生」搜尋到「蔡英文答陸生問：不認『臺獨』兩岸無法合作」這則標題中網友 a854677988 表示，「支持蔡英文和陸生表明立場，希望兩岸合作的前提是立場對等、臺灣人要有

臺灣大學圖書館期刊資料庫

尊嚴」，回文共有 279 則。

在討論軍事議題的網路口碑中，提及兩岸統獨立場的內容透露出：網友懼怕中國軍事壓力而傾向維持現狀，或是選擇兩岸統一。這是一種受到脅迫所產生的認同妥協。在軍事維度上，我們挑選回文則數較多的標題檢視，以「飛彈」搜尋到「TVBS 民調 71% 希望臺灣獨立」這則標題中網友 waipeiwang 認爲，「因爲中國大陸以飛彈武力威脅臺灣，所以才想要獨立」，回文共有 637 則。以「戰爭」搜尋到「臺獨的悲哀每下愈況日趨下流」這則標題中網友 cyt20120115 預期，「目中無人的中國二十年後若不進行非經濟上（如政治）的開放，一定會對臺灣或日本發動戰爭」，回文共有 449 則。以「潛艦」搜尋到「從香港看臺灣顏銘緯：獨立才是唯一活路」這則標題中網友 a0987758120 認爲，「臺灣尚未建立優勢軍力之前，若政府宣布臺灣獨立，臺灣會被中國擊敗」，回文共有 348 則。以「戰爭」搜尋到「自稱無畏統媒、馬政府抹黑陳爲廷表態：我主張臺獨」這則標題中網友 876 批評，「臺獨人士應該自行前去大陸打仗，避免牽連到不支持臺灣獨立的平民老百姓」，回文共有 336 則。以「武力犯臺」搜尋到「美敲打學運再搞臺獨就棄臺」這則標題中網友斐月認爲，「中國假如以武力犯臺，美國不可能坐視不管。但爲了減少開戰所造成的損失，希望兩岸維持現狀」，回文共有 317 則。

在討論國際議題的網路口碑中，提及兩岸統獨立場的內容多數在爭論：臺灣的政治立場一旦作出抉擇，是否會影響臺、美關係和臺灣與邦交國之間的關係。在國際維度上，我們挑選回文則數較多的標題檢視，以「聯合國」搜尋到「TVBS 民調 71% 希望臺灣獨立」這則標題中網友 2uup 指出，「臺灣獨立需要總統舉行獨立公投、改國號爲臺灣，再用臺灣名義加入聯合國。但網友質疑這些臺獨階段很難施行」，回文共有 637 則。以「辦事處」搜尋到「馬英九總統居然打算推動臺灣寫簡體字」這則標題中網友 Jessie Lai 認爲，「兩岸互設辦事處既浪費公帑，又有侵犯臺灣主權的危險性」，回文共有 471 則。以「聯合國」搜尋到「從香港看臺灣顏銘緯：獨立才是唯一活路」這則標題中網友 twpointtw 認爲，「只要臺灣不加入聯合國、不變更國號國旗、不廢憲，中國願意讓臺灣維持現狀」，回文共有 341 則。以「邦交」搜尋到「自稱無畏統媒、馬政府抹黑陳爲廷表態：我主張臺獨」這則標題中網友 justeat 認爲，「臺灣已經是實質上的國家，有自己的邦交

國，因此不必再臺獨了。網友擔憂眞的宣布臺獨，會引發戰爭」，回文共有 336 則。以「邦交」搜尋到「美敲打學運再搞臺獨就棄臺」這則標題中網友 cavity 認爲，「因爲臺灣的邦交國少，所以臺灣要更加依賴美國。」，回文共有 317 則。

最後，至於頻道、情緒、正面情緒聲量、負面情緒聲量、P/N 比因爲與主要理論無直接關，因此移至附錄 5「統獨聲量的其他分析」。

陸、結論與討論

兩岸關係的經驗研究受到學界的關注，除了嘗試以不同的測量瞭解、分析民眾的統獨立場，從 1990 年代，學界便有系統地蒐集民眾統獨立場的趨勢。當民意調查捕捉到民眾統獨的變遷時，網路上的統獨同時間也在持續發酵。網路上統獨立場的研究透過大數據的分析方法，可以不經抽樣，直接研究母體，爲網路統獨研究打開一扇窗。然此「母體」不是以每一位民眾爲分析單位，而是以每一「聲量」爲分析單位，當一位民眾在網路上發表多於一則關於統獨的言論時，其意見是被加權的，亦即當民眾發表多則聲量時，其統獨立場將被進行更大的代表，這是網路意見的特色，因爲我們看到的是聲量，不是個人。本研究曾經試圖以網路個別 ID 爲分析單位進行研究，然研究結果與以聲量爲分析單位進行研究類似，因此，不另外呈現相關結果。

回到研究問題，以民意調查得知支持獨立者與維持現狀者在整體中較多，網路世界也是如此嗎？從統計學的角度出發，對大數據的使用仍然會致力瞭解大數據資料與現實世界之間的差異。本研究發現，在統獨聲量的分布中，偏向獨立者高於統一，其中提及獨立爲 43,263 次，提及統一爲 35,063 次，提及「獨立」較「統一」多了 8,000 多次。在 2013-2014 這兩年中，支持獨立聲量的比例穩定維持在 8 成左右。然而，在統獨立場的民意調查中，具有一定比例的民眾選擇「維持現狀再決定」和「永遠維持現狀」，但在網路統獨聲量中，維持現狀的比例甚低，多數的聲量皆能判定其偏統或偏獨，且偏獨居多。

這項在網路上突出的「獨立」立場，與過去民意調查結果大相逕庭，

臺灣大學學術期刊資訊網

在政府、各政黨都強調「網軍」意見時，他們說的不是民意調查結果，是網路上呈現的政治意見聲量，統獨便是其中之一。亦即當政府或政黨關注民調和網路大數據的意見時，將會得到不同的結果，民調結果理論上有助於理解「全民」的意見，而網路大數據有助於理解「網民」的意見，至於潛水或不表態的網民，這也算是網路輿情嗎？本研究認爲不是，儘管這些潛水者可能在民意調查中表態，但他們因爲眾多原因不在網路中表態，意味著當媒體大幅報導政府或政黨應該重視網民意見時，這些潛水者的意見未被報導，自然也就不是記者筆下「婉君」的意見，這也是本研究主要要推廣的論點－網路大數據分析的目的不同於民意調查分析，也會得到不同結果，但兩者可以相輔相成。當然，不論是民意調查或大數據分析，也唯有將這些意見轉化成實際投票的政治參與，才是對政府、政黨或政治人物的徹底評價。至於潛水的那些沉默者是誰？那些人選擇沉默？將列爲本研究未來的研究目標。

本研究「理論上」將統獨聲量區分爲：獨立、統一、維持現狀三種，「維持現狀」本身應當與「統」或「獨」被同等看待，然這是植基於過去民意調查對統獨的分類，然在「實際上」分析維持現狀者卻較少。本研究發現網路上的統獨聲量獨多於統，雖與民調資料呈現的趨勢類似，但是網路聲量卻幾乎沒有維持現狀一類，亦即本研究可能的貢獻或洞見在於看到統獨網路聲量的分布的極化，而非中立化。這非常有可能如同文獻檢閱（王嵩音，2007；吳筱玫，2003；吳齊殷，1998；陳萬達，2007；Bargh et al., 2002；Derlega & Chaikin, 1977；Peng, 2104；Turkle, 1995；Wallace, 1999）提出與網路的匿名性有關，在匿名的世紀中，隱身在中間立場者，便可能因爲網路的「保護」而「挺身表態」。雖網路非完全匿名，但多數的網路意見表達仍爲匿名，一些非匿名的意見表達不會影響本文的結論。再者，雖然網路上可能也有大陸人士進行留言，影響臺灣網路民意的分析，然從分析結果中僅有少數的統一聲量，這項因素的干擾應該很小。

網路的匿名性已經增加民眾討論的尺度，並減少表達的隱性限制。加上統獨爲較敏感的議題，過去的研究經常發現敏感問題透過不同的實驗設置，可以讓民眾說出較爲眞實的想法（Blair et al., 2014；Lyall et al., 2013；Blair et al., 2015）。本研究分析網路大數據，網路上的政治討論正可減少在

臺灣社會與政治研究資料庫

眞實世界表達政治意見的壓力，更能展現網民的想法，而這與以民意調查進行統獨研究時，研究者想方設法希望挖掘「維持現狀者」骨子裡想法的做法異曲同工。

在「網路政治」的研究領域中，「婉君」怎麼想？再以「對象」而言，「中國」及「臺灣」被提及次數最多。在聲量來源方面，兩岸議題討論聲量中，高達 48% 集中在社群網站。再進一步考慮到統獨議題的聲量情緒分析，兩岸議題總 P/N 比爲 0.61，即平均每出現 6 篇正面情緒聲量時，同時會出現 10 篇負面情緒聲量。

雖然大數據分析的強項不在於解釋因素間的因果關係（causal relation），然我們卻透過各種分析看到一些「相關」（correlation），因此我們也無需認爲大數據的結果僅能描述，而不能解釋。例如：媒體選擇報導的統獨議題與網路討論的統獨議題便有相關，發展趨勢雷同，圖 2 顯示，發生重大統獨議題時，網路討論聲量也較大；圖 3 也顯示，服貿議題熱議時，獨立聲量較多，以上兩點都說明了時間與聲量有關。再者，在這兩年的統獨聲量中，聲量越多時，P/N 比值愈小，負面情緒通常越強烈（圖 7、圖 8），我們也藉此看到聲量與 P/N 比的關聯。而在太陽花學運時，主權維度的聲量較多（圖 6），這也是時間（事件）與討論維度的關聯。

再者，圖 7 到圖 8 皆爲談論相關性，包括統獨維度和情緒顯著比的相關、統獨維度和 P/N 比的相關性。網友在討論主權時，討論時相對冷靜不帶情緒，但若有情緒，也是負面居多。此外，網友在 14 年討論經貿議題時，比 13 年冷靜且比較樂觀。最後，統獨立場中維持現狀立場的討論雖少，仍會關注主權等，但卻完全沒有經貿及國際。統一立場者相對考慮較多維度（圖 9），以上爲統獨立場和維度的相關性討論。

本研究從不同方式研究網路統獨，大數據分析的強項在於廣泛解釋現象，也因爲研究大數據，所以此類資料便無代表性檢定的問題，大數據便是母體。然而，大數據的母體與眞實的母體並不一致，因爲研究者使用了關鍵字過濾聲量，亦即本研究在推論時需要更謹慎。

對於解釋現象背後的原因則非大數據分析主要關心的焦點，但本研究也試圖探尋上述所謂的相關，在大數據中找到相關的型態。雖本研究樣本不完全等於母體，因爲有些網站不開放，然 7,000 個網站頻道，已涵蓋重

臺灣本土學術期刊資料庫

要和大型頻道，亦即在恪遵研究倫理、不以其他技術侵入非開放網站的情況下，本研究已做最大努力。

對臺灣民眾而言，統獨立場總隱藏在「維持現狀」的選項之中。網路是匿名的空間，可能讓民眾更願意表示意見，可以不需要暴露自己太多背景，更能暢所欲言、自我揭露，表達自身對事物的根本看法，民眾在網路上更加「無懼」。從網路統獨聲量的研究，我們探詢到更多偏統一或偏獨立的聲量，亦即民眾在網路上是會「探出頭來」，比較不會「躲在牆角」，面對面討論敏感議題的尷尬減少了，也許民眾在統獨立場上有「外柔內剛」的特質，民調中（外表）表達的是維持現狀，網路上（內心）表達的偏獨，無論民意調查千方百計想要探詢民眾的統獨態度，維持現狀總是主流意見。

然而，網路上的統獨聲量卻非如此，如此反差反映了研究來源的不同，以及方法的差異。研究來源的不同是指民調研究對象爲民眾，網路大數據的研究對象爲聲量，前者試圖推論全國民眾統獨立場，後者要觸及的便是網路統獨輿情，以兩種方法進行的研究相輔相成。本文希望在民調之外，挖掘「網路統獨」，也的確得到非常不同的研究結果，亦即當民調聲稱民眾的統獨如何如何，其實它還必須考慮那些不在家接受電話訪問的民眾，或許這些不接電話的民眾會上網討論，這才顯得用網路大數據研究統獨立場的必要和特色。

最後，網路聲量和研究倫理的問題也需要受到討論。本研究所涵蓋的網路社群多樣，這些網路社群的用戶各有不同的人口背景。本研究所提供的資料僅爲針對網路上的公開資料進行截取與分析，並未蒐集網路社群用戶的個人資料；基於個人資料保護法的規定，也未針對實名制網路社群的個人資料進行分析。以上爲研究倫理的遵守，應不至於列爲研究限制。儘管我們知道不同網路社群用戶的人口組成各不相同、不同網路社群間用戶、虛擬／實質影響力、代表性的比較基準並不一致、也不等值；但爲了掌握統一、獨立立場在網路上的佔比差異，且網路社群之間的影響力比較並不是本研究的研究範圍，因此本研究使用「聲量」此一量化單位做爲不同網路社群中統獨立場的比較單位，是基於研究倫理和實際研究時的妥協做法。

再者，作者未來的研究將可以朝幾個方向繼續深入探討：臺灣年輕人

臺灣大學學術期刊資料庫

「天然獨」或「自然獨」的內涵是什麼？當年輕人提到人民、政府、國家、自己、中國等關鍵字時，他們彼此之間是如何討論的？跟上個世代有何不同？或是嘗試建立因果關係的解釋。作者將不會僅停留在描述現象階段，希望未來擴大對學術的貢獻。然所有的研究皆從描述開始，本文從各個面向描述網路統獨聲量，是一個起始，即使只是描述，但我們也已經研究出網路統獨聲量豐富的面向。

至於網路聲量是否可以完全辨別出它們的統／獨的立場，本文雖以四大步驟（頁 127-128 第一步驟爲「定義相關聲量」，第二步驟爲「設定關鍵字距離」，第三步驟「擴充關鍵字」，使用中文斷詞處理方式，分析出網友習慣用語。接著再將新詞組納入分析範圍，重新搜尋兩岸統獨議題聲量。最後，第四步驟還需要「排除雜訊」，亦即排除無關內容）分析網路統獨聲量，然是否能百分之百掌握網路統獨聲量的「立場」，也許仍有落差存在，畢竟立場與聲量還是有所不同，聲量主要分析的是它們的「量」，有可能無法完全辨別網路聲量統／獨的立場，也值得未來繼續精進。

參考書目

一、中文部分

王甫昌，2008，〈族群政治議題在臺灣民主轉型中的角色〉，《臺灣民主季刊》，5(2): 89-140。

王嵩音，2007，〈網路使用之態度、動機與影響〉，《資訊社會研究》，12: 57-85。

吳乃德，1992，〈國家認同和政黨支持：臺灣政黨競爭的社會基礎〉，《中央研究院民族學研究所集刊》，74: 33-60。

吳乃德，1993，〈省籍意識、政治支持與國家認同〉，張茂桂（編），《族群關係與國家認同》，臺北：業強，頁 27-51。

吳乃德，1997，〈國家認同和民主鞏固：衝突、共生與解決〉，游盈隆（編），《民主鞏固或崩潰：臺灣二十一世紀的挑戰》，臺北：月旦，頁 15-30。

吳乃德，2005，〈麵包與愛情：初探臺灣民眾民族認同的變動〉，《臺灣政治學刊》，9(2): 5-39。

吳玉山，1999，〈臺灣的大陸政策：結構與理性〉，包宗和、吳玉山（編），《爭辯中的兩岸關係理論》，臺北：五南，頁 153-210。

臺灣政治學學術期刊資料庫

吳玉山，2001，〈兩岸關係中的中國意識與臺灣意識〉，《中國事務》，4: 71-89。
吳清山，2014，〈教育名詞－大數據〉，《教育資料與研究》，115: 290-280。
吳筱玫，2003，《網路傳播概論》，臺北：智勝。
吳齊殷，1998，〈虛擬社區 vs. 眞實生活〉，《科學月刊》，29(8): 668-674。
李欣宜，2015，〈一次搞懂大數據（下）〉，數位時代網頁，http://www.bnext.com.tw/article/view/id/35809，2015/4/2。
林瓊珠，2012，〈穩定與變動：臺灣民眾的「臺灣人／中國人」認同與統獨立場之分析〉，《選舉研究》，19(1): 97-127。
林繼文，2015，〈論述如何框限選擇？條件式統獨偏好對 2012 年臺灣總統選舉的影響〉，《政治科學論叢》，63: 55-90。
俞振華、林啓耀，2013，〈解析臺灣民眾統獨偏好：一個兩難又不確定的選擇〉，《臺灣政治學刊》，17(2): 165-230。
科技部，2017，〈臺灣傳播調查資料庫 1 期 3 次〉，科技部臺灣傳播調查資料庫網頁，http://www.crctaiwan.nctu.edu.tw/，2015/09/30。
科技部人文司，2017，〈臺灣選舉與民主化調查〉，TEDS 網頁，http://www.tedsnet.org/main.php，2017/01/20。
胡世忠，2013，《雲端時代的殺手級應用：海量資料分析》，臺北：天下。
耿　曙、劉嘉薇、陳陸輝，2009，〈打破維持現狀的迷思：臺灣民眾統獨抉擇中理念與務實的兩難〉，《臺灣政治學刊》，13(2): 3-56。
國立政治大學選舉研究中心，2017，〈臺灣民眾統獨立場趨勢分布〉，選舉研究中心網頁，http://esc.nccu.edu.tw/news/news.php?Sn=167，2016/09/28。
張佑宗，2006，〈選舉事件與選民的投票抉擇：以臺灣 2004 年總統選舉爲分析對象〉，《東吳政治學報》，22: 121-160。
張傳賢、黃　紀，2011，〈政黨競爭與臺灣族群認同與國家認同間的聯結〉，《臺灣政治學刊》，15(1): 3-71。
盛杏湲，2002，〈統獨議題與臺灣選民的投票行爲：一九九〇年代的分析〉，《選舉研究》，9(1): 41-80。
陳文俊，2003，〈藍與綠：臺灣選民的政治意識型態初探〉，《選舉研究》，10(1): 41-80。
陳陸輝、周應龍，2004，〈臺灣民眾統獨立場的持續與變遷〉，《東亞研究》，35(2): 143-186。
陳陸輝、耿　曙、王德育（陳陸輝等），2009a，〈兩岸關係與 2008 年臺灣總統大選：經濟交流、武力威脅與選民投票取向〉，《選舉研究》，16(2): 1-22。
陳陸輝、耿　曙、涂萍蘭、黃冠博（陳陸輝等），2009b，〈理性自利或感性認

臺灣政治學術期刊資料庫

同？影響臺灣民眾兩岸經貿立場因素的分析〉，《東吳政治學報》，27(2): 87-125。

陳敦源、蕭乃沂、廖洲棚，2015，〈邁向循證政府決策的關鍵變革：公部門巨量資料分析的理論與實務〉，《國土及公共治理》，3(3): 33-44。

陳萬達，2007，《網路新聞學》，新北：威仕曼。

陳義彥、陳陸輝，2003，〈模稜兩可的態度還是不確定的未來：臺灣民眾統獨觀的解析〉，《中國大陸研究》，46(5): 1-20。

游盈隆，1996，《民意與臺灣政治變遷：1990 年代臺灣民意與選舉政治的解析》，臺北：月旦。

游清鑫，2016，〈民意調查的類型〉，陳陸輝（編），《民意調查研究》，臺北：五南，頁 59-94。

游清鑫、林長志、林啓耀，2013，〈臺灣民眾統獨立場的問卷設計與測量：以 TEDS 爲例〉，黃紀（編），《臺灣選舉與民主化調查（TEDS）方法論之回顧與前瞻》，臺北：五南，頁 55-88。

意藍資訊，2016，《意藍資訊網路口碑分析學術研究文獻探討》，未出版。

楊婉瑩、李冠成，2011，〈一個屋簷下的性別權力關係對國家認同的影響（1996-2008）〉，《選舉研究》，18(1): 94-136。

廖洲棚、陳敦源、蕭乃沂、廖興中，2013，《運用巨量資料實踐良善治理：網路民意導入政府決策分析之可行性研究》（行政院研究發展考核委員會委託研究報告，RDEC-MIS-102-003），臺北：行政院。

劉嘉薇，2014-2016，《臺灣民眾媒體使用與統獨立場的因果交融：十二年來的實證分析》（行政院科技部專題研究計畫－優秀年輕學者研究計畫，編號：MOST 103-2628-H-305-001-MY2），臺北：行政院。

劉嘉薇，2015，〈政治學研究途徑與方法〉，陳義彥（編），《政治學》（增訂第六版），臺北：五南，頁 99-126。

劉嘉薇，2016，《臺灣民眾的媒體選擇與統獨立場》，臺北：五南。

劉嘉薇、耿　曙、陳陸輝，2009，〈務實也是一種選擇：臺灣民眾統獨立場的測量與商榷〉，《臺灣民主季刊》，6(4): 141-68。

蕭怡靖、游清鑫，2012，〈檢測臺灣民眾六分類統獨立場：一個測量改進的提出〉，《臺灣政治學刊》，16(2): 65-116。

二、英文部分

Bargh, John A., Katelyn Y. McKenna & Grainne M. Fitzsimons. 2002. “Can You See the Real Me? Activation and Expression of the ‘True Self’ on the

臺灣期刊學術期刊資料庫

Internet.” *Journal of Social Issues* 58(1): 33-48.

Bertolucci J. 2013. “Big Data Analytics: Descriptive vs. Predictive vs. Prescriptive.” in http://www.informationweek.com/big-data/big-data-analytics/big-data-analytics-descriptive-vs-predictive-vs-prescriptive/d/d-id/1113279. Latest update 11 December 2016.

Blair, Graeme, C. Christine Fair, Neil Malhotra & Jacob N. Shapiro. 2013. “Poverty and Support for Militant Politics: Evidence from Pakistan.” *American Journal of Political Science* 57(1): 30-48.

Blair, Graeme, Kosuke Imai & Jason Lyall. 2014. “Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan.” *American Journal of Political Science* 58(4): 1043-1063.

Blair, Graeme, Kosuke Imai & Yang-Yang Zhou. 2015. “Design and Analysis of the Randomized Response Technique.” *Journal of the American Statistical Association* 110(511): 1304-1319.

Borgman, Christine L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, Massachusetts: The MIT Press.

Chu, Y. H. & J. W. Lin. 2001. “Political Development in 20th-century Taiwan: State-building, Regime Transformation and the Construction of National Identity.” *China Quarterly* 165: 102-129.

Derlega, Valerian J. & Alan L. Chaikin. 1977. “Privacy and Self-Disclosure in Social Relationships.” *Journal of Social Issues* 33(3): 102-115.

Dittmer, L. 2005. “Taiwan’s Aim-inhibited Quest for Identity and the China Factor.” *Journal of Asian & African Studies* 40(1/2): 71-90.

Driscoll, Kevin & Shawn Walker. 2014. “Working Within a Black Box: Transparency in the Collection and Production of Big Twitter Data.” *International Journal of Communication* 8: 1745-1764.

Ekbia, Hamid, Michael Mattioli, Inna Kouper, G. Arave, Ali Ghazinejad, Timothy Bowman, Venkata Ratandeep Suri, Andrew Tsou, Scott Weingart & Cassidy R. Sugimoto. 2015. “Big Data, Bigger Dilemmas: A Critical Review.” *Journal of the Association for Information Science and Technology* 66(8): 1523-1545.

Haire, A. J. & V. Mayer-Schönberger. 2014. “Big Data: Opportunity or Threat?” *InterMedia* 42(3): 16-21.

Hsieh, J. F. S. & E. M. S. Niou. 1996. “Salient Issues in Taiwan’s Electoral Politics.” *Electoral Studies* 15: 219-235.

Indurkhya, Nitin & Fred J. Damerau. eds. 2010. *Handbook of Natural Language Processing*. Boca Raton, Fla.: Chapman & Hall/CRC Press.

Jin, X, B. W. Wah, X. Cheng & Y. Wang 2015. “Significance and Challenges of Big Data Research.” *Big Data Research* 2: 59-64.

Jordan, J. M. & D. K. J. Lin. 2014. “Statics for Big Data: Are Statisticians Ready for Big Data?” *Journal of Chinese Statistical Association* 52: 133-149.

臺灣大學傳播學院資料庫

Keng, S., L. H. Chen & K. B. Huang. 2006. “Sense, Sensitivity, and Sophistication in Shaping the Future of Cross-strait Relations.” *Issues & Studies* 42(4): 23-66.

Lazer, David, Ryan Kennedy, Gary King & Alessandro Vespignani. 2014. “The Parable of Google Flu: Traps in Big Data Analysis.” *Science* 343: 1203-1205.

Lei, Xuchuan & Jie Lu. 2016. “Revisiting Political Wariness in China’s Public Opinion Surveys: Experimental Evidence on Responses to Politically Sensitive Questions.” *Journal of Contemporary China* 26(104): 213-232.

Lin, T. M., Y. H. Chu & M. Hinich. 1996. “Conflict Displacement and Regime Transition in Taiwan: A Spatial Analysis.” *World Politics* 48(4): 453-481.

Liu, I. C. & S. Y. Ho. 1999. “The Taiwanese/Chinese Identity of the Taiwan People.” *Issues and Studies* 35(3): 1-34.

Lyall, Jason, Graeme Blair & Kosuke Imai. 2013. “Explaining Support for Combatants During Wartime: A Survey Experiment in Afghanistan.” *American Political Science Review* 107(4): 679-705.

Manovich. 2011. “Trending: The Promises and the Challenges of Big Data.” In *Debates in the Digital Humanities*, ed. M. K. Gold. Minneapolis: University Of Minnesota Press, 460-475.

Nickerson, David W. & Todd Rogers. 2014. “Political Campaigns and Big Data.” *Journal of Economic Perspectives* 28(2): 51-73.

Pan, Hsin-Hsin, Wen-Chin Wu & Yu-Tzung Chang. 2017. “How Chinese Citizens Perceive Cross-Strait Relations: Survey Results from Ten Major Cities in China (*Journal of Contemporary China*).” in http://dx.doi.org/10.1080/10670564.2017.1274835. Latest update 10 January 2017.

Peng, Kun. 2014. *Anonymous Communication Networks: Protecting Privacy on the Web*. Boca Raton: CRC Press.

Rigger, S. 2001. “Maintaining the Status Quo: What It Means, and Why the Taiwanese Prefer It.” *Cambridge Review of International Affairs* 14(2): 103-114.

Rosenfeld, Bryn, Kosuke Imai & Jacob N. Shapiro. 2016. “An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions.” *American Journal of Political Science* 60(3): 783-802.

Turkle, Sherry. 1995. *Life on the Screen: Identity in the Age of the Internet*. New York: Simon and Schuster.

Wachman, A. M. 1994. *Taiwan: National Identity and Democratization*. New York, NY: M. E. Sharpe.

Wallace, Jonathan D. 1999. “Nameless in Cyberspace Anonymity on the Internet.” *Cato Institute Briefing Papers* 54: 1-8.

Wu, Wen-Chin, Yu-tzung Chang & Hsin-Hsin Pan. 2017. “Does China’s Middle Class Prefer (Liberal) Democracy?” *Democratization* 24(2): 347-366.

臺灣大學圖書館期刊資料庫

附錄 1

統獨聲量判定最終關鍵字

關　　鍵　　字
統一**兩岸｜統一**中國｜統一**大陸｜統一**內地｜獨立**兩岸｜獨立**中國｜獨立**大陸｜獨立**內地｜維持現狀**兩岸｜維持現狀**中國｜維持現狀**大陸｜維持現狀**內地｜兩岸**統一｜中國**統一｜大陸**統一｜內地**統一｜兩岸**獨立｜中國**獨立｜大陸**獨立｜內地**獨立｜兩岸**維持現狀｜中國**維持現狀｜大陸**維持現狀｜內地**維持現狀｜臺獨｜統一**臺灣｜統一**臺灣｜統一**中華民國｜獨立**臺灣｜獨立**臺灣｜獨立**中華民國｜維持現狀**臺灣｜維持現狀**臺灣｜維持現狀**中華民國｜臺灣**統一｜臺灣**統一｜中華民國**統一｜臺灣**獨立｜臺灣**獨立｜中華民國**獨立｜臺灣**維持現狀｜臺灣**維持現狀｜中華民國**維持現狀｜統一**臺澎｜統一**臺澎｜統一**ROC｜統一**PRC｜獨立**臺澎｜獨立**臺澎｜獨立**ROC｜獨立**PRC｜維持現狀**臺澎｜維持現狀**臺澎｜維持現狀**ROC｜維持現狀**PRC｜臺澎**統一｜臺澎**統一｜ROC**統一｜PRC**統一｜臺澎**獨立｜臺澎**獨立｜ROC**獨立｜PRC**獨立｜臺澎**維持現狀｜臺澎**維持現狀｜ROC**維持現狀｜PRC**維持現狀｜統一**中共｜統一**支那｜統一**北京｜統一**祖國｜獨立**中共｜獨立**支那｜獨立**北京｜獨立**祖國｜維持現狀**中共｜維持現狀**支那｜維持現狀**北京｜維持現狀**祖國｜中共**統一｜支那**統一｜北京**統一｜祖國**統一｜中共**獨立｜支那**獨立｜北京**獨立｜祖國**獨立｜中共**維持現狀｜支那**維持現狀｜北京**維持現狀｜祖國**維持現狀！統一獅！統一集團！統一企業！統一公司！統一超！統一阪急！統一食品！統一麵！統一泡麵！統一布丁！統一超！獨家！尋參！細商！考資！品規！尋結！品網！貸！搶購！細介！品訊！買網址！特惠商！免費！成人！色情！線上學英文

資料來源：本研究整理

註：**代表內含字元，**不超過 10 字。｜代表關鍵字之間的間隔，無特殊意義。！代表否定，亦即排除該關鍵字。

臺灣大學學術期刊資料庫

附錄 2

網友習慣用語列表

下列習慣用語的取得方式有二：其一爲人工閱讀 PTT、Mobile01、卡提諾論壇等熱門網站文章，紀錄各立場網友常用詞語，並刪減雙方互有重疊者。其二爲對原始資料抽樣並人工標記 300 篇文章的統獨立場後，分別對二立場的文章，使用中文斷詞技術處理，以抽取出各自立場的特定用語整併以上二個方式分析結果，得到如下習慣用語。

立　　場	習　　慣　　用　　語
獨立（37 個詞）	黨產、傷害中國、人民感情、感情脆弱、中時不意外、聯合重工、獨裁、玻璃心、心碎、支那、中共、善液、舔共、史明、丁貴、捧 LP、賤畜、國民黨不倒、帶風向、共諜、紅色、侵略、恐嚇、螞蝗、馬皇、小刀、阿共、中國黨、八國聯軍、貴國、共匪、馬冏、蛆蛆、狗黨、一邊一國、瞄準、文攻武嚇
統一（16 個詞）	祖國、內地、美國走狗、扁貪、空心菜、冥進、皇民、吱吱、學匪、民粹、臺巴、灣灣、中華民族、同文同種、一家人、和平統一

資料來源：本研究整理

臺灣大學學術期刊資料庫

附錄 3

新聞報導兩岸新聞主要維度

總類目	次 類 目	說　　明	關　　鍵　　字
主　權	意識形態	討論統一、獨立或維持現狀等與兩岸主權有關之主題。	九二共識、統一、獨立、統獨、國統綱領、兩岸和平協議、終極統一、急統、舊金山和約、一中各表、反分裂國家法、兩國論、一中憲法、臺獨議題、維持現狀、臺灣國
	國　　籍	與國籍相關議題都歸在此類。	外國籍、國籍、國籍問題
	族　　群	談論與族群相關議題都歸在此類。	族群、種族、臺灣人、中國人、血統論
	司　　法	與司法議題相關都歸於此類。	兩岸司法互助機制、司法、打擊犯罪、加強查緝
經　貿	經　　濟	論及經濟、貿易、通商事項都歸類在此。	ECFA、兩岸經濟合作架構協議、兩岸投保協議、FTA、自由貿易區、TPP、兩岸共同市場、臺商企業（臺商）
	中資企業	與中資企業相關議題都歸在此類。	中資企業、中資財團、陸資、陸資企業、陸資財團
民　主	民主價值	強調人權、自由、平等、民主的思潮都屬於民主價值主題。	民主價值、民主化、人權、民主化、六四、自由、選舉、公投
	政治思想	談論與政治思想相關議題都歸在此類	馬克思、共產主義
	政　　黨	與政黨相關議題都歸在此類。	政黨、政黨輪替

臺灣大學學術期刊資料庫

總類目	次類目	說明	關鍵字
文化	文字、風俗	關於語言、文字的認知以及兩岸文化交流活動，不管是藝術乃至於文學都歸類於此。	簡體字、繁體字、清明祭祖、故宮國寶、中國文化、通婚
	宗教	論及各種宗教議題、宗教交流活動都歸類於此。	法輪功、藏傳佛教、達賴喇嘛
	體育	談論各種體育活動，如奧運、亞運，或是兩岸交流賽事都歸於此類。	奧運、亞運、世界警察消防運動會、奧運聖火、京奧
	教育	談論各種教育議題都歸於此類。	陸生來臺、兩岸學術研討會、三限六不、高中公民教育
	旅遊	談論旅遊相關議題都歸於此類。	觀光、自由行、大陸觀光客、陸客、臺胞證、兩岸直航
	歷史	談論與歷史事件相關議題都歸在此類。	歷史、辛亥百年、共同記憶、二二八事件、臺灣史
	網路	談論與網路相關議題都歸在此類。	網路、網軍、網站、網際網路
軍事	軍事	談論國防、匪諜或是軍購等歸類在國防主題。	軍事、陸軍司令部、共諜、將軍、軍購、參謀本部、武力犯臺、解放軍
國際	外交	論及國際交流與外交政策的議題都屬於外交主題。	外交部、外交政策、活路外交、外交休兵、外館、外交辦事處
	國際局勢	以國際事件爲主題，並論及兩岸或是臺灣發展都屬於這項類別。	WHA、WTO、APEC、國家競爭力、加入聯合國（入聯、返聯）
	衛生	談論與衛生安全相關議題都歸在此類。	瘦肉精、狂牛症、肺結核、食品安全、禁藥、毒物、農委會、環保署、海關、標準檢驗局、動物殘留檢驗標準、黑心食品

資料來源：本研究整理

臺灣大學學術期刊論文集

附錄 4

網路維度關鍵字

項次	維度	關　　鍵　　字
1	主權	九二共識｜國統綱領｜和平協議｜終極統一｜終極獨立｜急統｜急獨｜舊金山和約｜中日和約｜波茲坦｜開羅｜一中各表｜一個中國｜反分裂國家法｜兩國論｜一中憲法｜臺灣國｜國籍｜族群｜種族｜民族｜臺灣人｜中國人｜血統｜後代｜司法｜查緝｜犯罪｜判刑｜公安｜法官｜法院
2	經貿	ECFA｜FTA｜TPP｜經濟合作｜投保協議｜自由貿易區｜自貿區｜共同市場｜臺商｜中資｜陸資｜服貿｜服務貿易｜伏冒
3	民主	民主價值｜民主化｜人權｜六四｜民主自由｜自由民主｜選舉｜公投｜馬克思｜馬列｜共產主義｜共產思想｜民主集中｜三民主義｜政黨
4	文化	簡體字｜繁體字｜正體字｜殘體字｜簡體中文｜繁體中文｜正體中文｜殘體中文｜書簡識繁｜清明｜祭祖｜故宮國寶｜中國文化｜中華文化｜通婚｜習俗｜法輪功｜佛教｜達賴｜喇嘛｜基督教｜回教｜伊斯蘭教｜奧運｜亞運｜運動會｜京奧｜陸生｜學術交流｜三限六不｜課綱｜教育｜陸客｜中客｜觀光｜自由行｜臺胞證｜兩岸直航｜歷史｜辛亥｜記憶｜二二八｜中國史｜臺灣史｜抗戰｜內戰｜軍閥｜遷臺｜淪陷｜建政｜新中國｜網路｜互聯網｜網軍｜婉君｜五毛｜翻牆｜因特網｜網站｜封鎖｜長城
5	軍事	國防｜軍購｜軍事｜陸軍｜海軍｜空軍｜三軍｜司令｜統帥｜參謀｜共諜｜將軍｜將領｜上將｜中將｜少將｜上校｜中校｜少校｜上尉｜中尉｜少尉｜武力｜戰爭｜解放｜後備｜現役｜戰車｜坦克｜陸戰｜步兵｜砲兵｜工兵｜砲灰｜登陸｜搶灘｜巷戰｜斬首｜前線｜艦隊｜潛艇｜潛水艇｜潛艦｜飛彈｜戰機｜戰鬥機｜直升機｜東風｜航母｜雄二｜天弓｜半徑｜愛國者｜中科院｜雄風｜核彈｜導彈｜預警｜武器｜九鵬
6	國際	外交部｜外交政策｜活路外交｜外交休兵｜邦交｜外館｜使館｜辦事處｜WHA｜WHO｜APEC｜國家競爭力｜入聯｜返聯｜瘦肉精｜狂牛症｜肺結核｜食品安全｜食安｜禁藥｜毒物｜農委會｜環保署｜海關｜標準檢驗局｜標檢局｜動物殘留｜殘留檢驗｜檢驗標準｜黑心食品｜黑心食物

資料來源：本研究整理

臺灣大學圖書館期刊資料庫

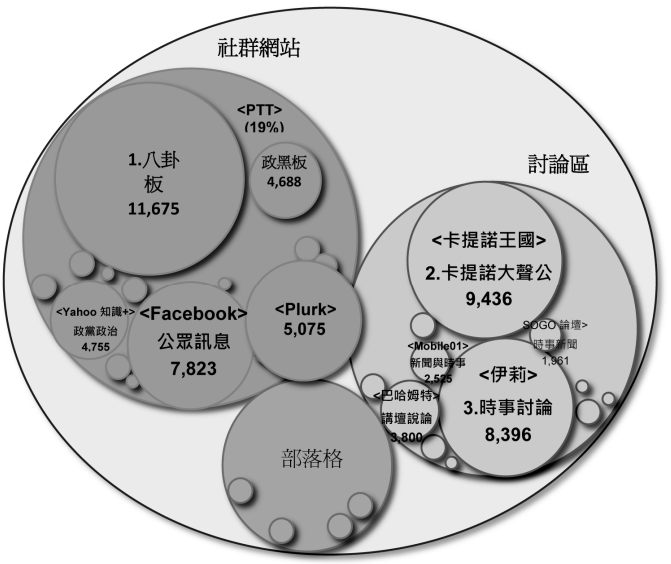
附錄 5

統獨聲量的其他分析

一、統獨聲量來源與頻道分析

聲量來源及頻道分布狀況如何（如附錄圖 1），兩岸議題討論聲量中，高達 48%集中在社群網站，其次討論區達 41%，部落格僅佔 10%。其中，單一網站聲量最多者爲 PTT，聲量佔比 19%，其次爲卡提諾王國及伊莉論壇。單一頻道聲量最多者爲 PTT 八卦板（11,675 聲量），接著爲卡提諾大聲公（9,436 聲量）及伊莉時事討論（8,396 聲量）。此外，我們還發現，「臺灣」相關詞最常被提及，而社群網站文章內含字數較多，即內容較偏向長篇大論。

附錄圖 1　主要聲量來源及頻道佔比示意圖



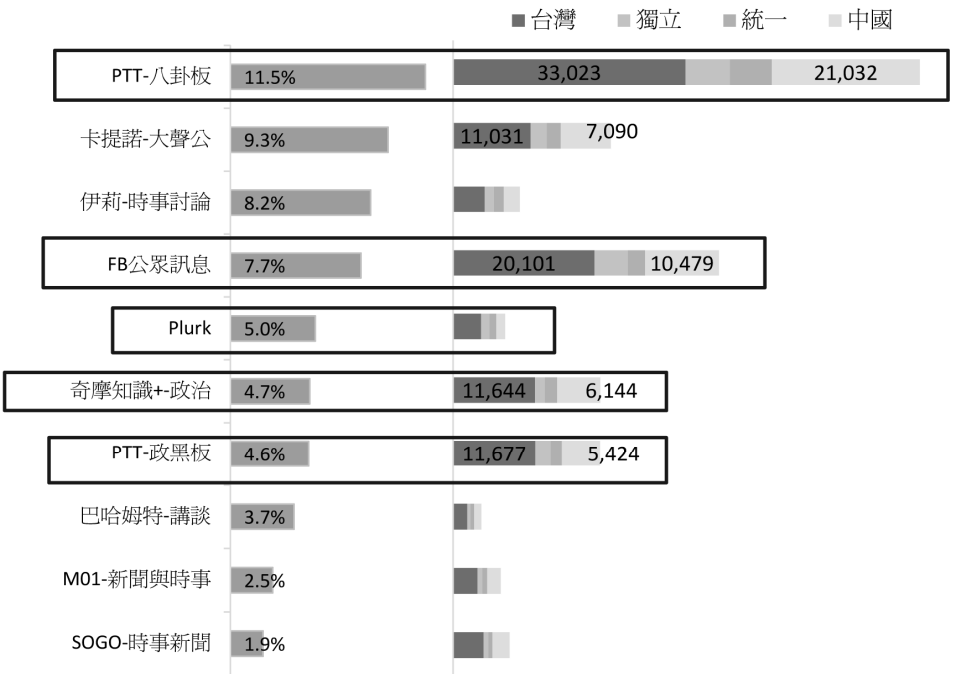
資料來源：本研究自製

既然多數的聲量都來自社群網站，社群網站中的重要頻道爲何？其討論本研究設定的關鍵字情況爲何？在附錄圖 2 中，以線條框住的部分爲社

臺灣本土學術期刊資料

群網站，其它未被框住的部份爲討論區或部落格。在社群網站頻道聲量的討論中，首先，PTT 八卦版佔前十大頻道聲量的 11.5%，其中討論「臺灣」者最多（臺灣及臺獨等字詞合併成「臺灣」項目，中國及大陸等字詞合併成「中國」項目），有 33,023 聲量，討論「中國」者次之，有 21,032 聲量。再者，Facebook 公眾訊息佔前十大頻道聲量的 7.7%，其中討論「臺灣」者最多，有 20,101 聲量，討論「中國」者次之，有 10,479 聲量。緊接著，討論聲量依序是 Plurk 佔前十大頻道聲量的 5.0%、奇摩知識＋的政治佔 4.7%，PTT 政黑板佔 4.6%，以上三者的討論多爲「臺灣」提及的次數最多。

附錄圖 2　前十大頻道討論聲量及重要字詞提及次數



資料來源：本研究自製

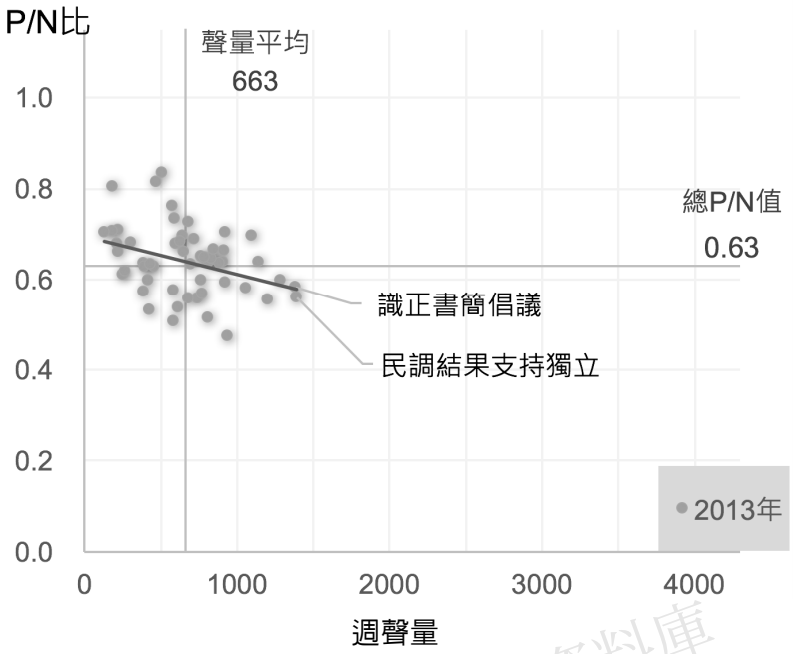
二、統獨聲量情緒分析

再進一步考慮到統獨議題的聲量情緒分析，情緒指的是正面或負面的評價，以下 P 所指爲正面（positive），N 所指爲負面（negative），P/N 比爲

臺灣大學學術期刊資料庫

正面評價和負面評價的比值。整體而言，兩岸議題總 P/N 比爲 0.61，即平均每出現 6 篇正面情緒聲量時，同時會出現 10 篇負面情緒聲量。在附錄圖 3、附錄圖 4 中，每一資料點即代表一週，X 軸即代表該週總聲量大小，Y 軸即代表該週 P/N 比高低。不論圖橫軸的周聲量多少，在圖的縱軸上，所有點 P/N 比的分布都在 1 以下，亦即不論 2013 年或 2014 年，各週的負面聲量都多於正面聲量（P/N 值皆小於 1）。若另外觀察兩年分布趨勢的話，則聲量越多時，P/N 比值愈小，負面情緒通常越強烈。若比較兩年之間差異時，則 2014 年比 2013 年的聲量更多，周聲量點的分布更往右移。此外，負面情緒也更強烈，周聲量點的分布更往下移。

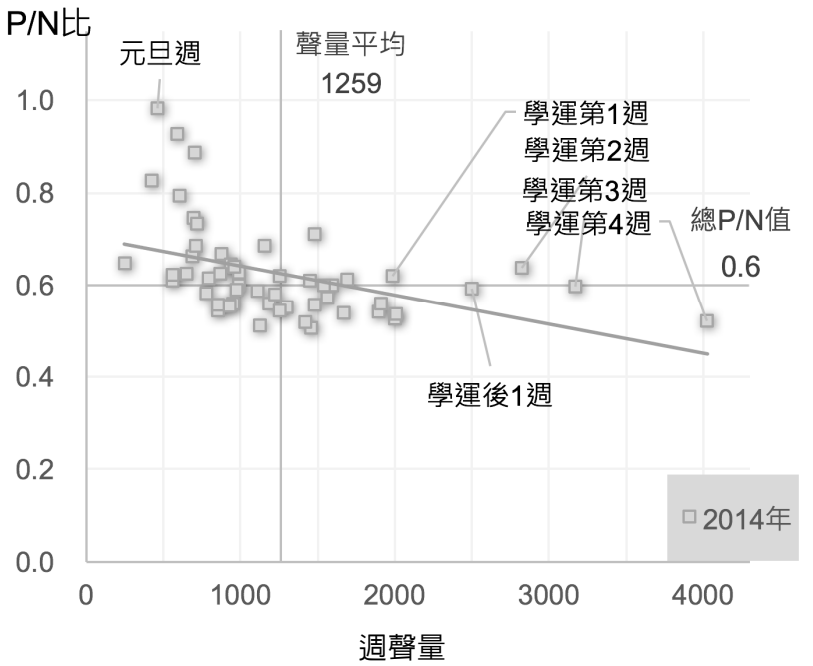
附錄圖 3　2013 年每週總聲量及 P/N 比分布圖
（每一點代表一周）



資料來源：本研究自製

臺灣大學學術期刊資料庫

附錄圖 4　2014 年每週總聲量及 P/N 比分布圖
（每一點代表一周）

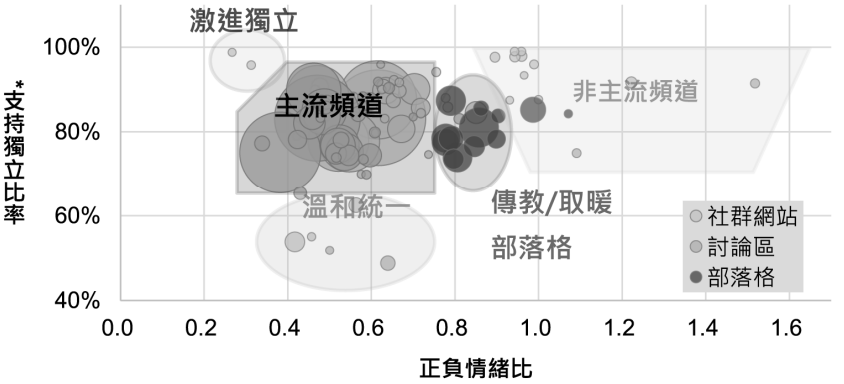


資料來源：本研究自製

而統獨立場的主流頻道主要來源是社群網站及討論區，從附錄圖 5 得知，主流頻道（60 個，例如：捷克論壇、PCDVD、蔡英文粉絲專頁等）情緒表現在 0.3～0.7 之間，支持獨立比例〔獨立立場聲量／（獨立立場聲量＋統一立場聲量）〕則在 70%～90% 左右，另一大族群本研究將之命名爲「傳教／取暖」（18 個，例如：UDN 部落格、奇摩部落格等）性質的部落格，其情緒明顯較主流頻道正面。此外，另一種類型分別落在立場維度的兩個極端處，本研究命名爲「激進獨立及溫和統一」類型，然而聲量不多。另有統獨討論的非主流頻道，以藝人 FB 粉專爲主，其情緒較正面，但相關聲量較少。之所以以支持獨立比例爲縱軸，原因在於獨立聲量較統一聲量多，較需要關注。

臺灣大學學術期刊資料庫

附錄圖 5　聲量前百大頻道、統獨立場及情緒分布



資料來源：本研究自製

三、統獨聲量關鍵領袖分析

而在這巨量的網路聲量中，有無所謂的關鍵領袖？關鍵領袖是指出沒頻道在獨立及統一立場分別的聲量中，文章篇數最多的前十名頻道，且在獨立及統一立場的聲量中，文章篇數分別最多的前三名作者 ID，當然這些 ID 背後也不排除是同樣的人，即使是同樣的人，我們都會納入分析，因爲大數據分析更在乎的是聲量。在附錄表 1 的第二欄中，獨立立場關鍵領袖出沒頻道最多者爲 PTT 的八卦版（Gossiping），佔獨立聲量的 11.64%；Facebook 次之，佔獨立聲量的 8.04%；再者爲「卡提諾王國」下的「卡提諾大聲公」，佔獨立聲量的 7.46%。附錄圖 6 的區塊大小便是討論量的多寡，標示 1 的區塊便是 PTT 的八卦版，其它依序爲 Facebook、「卡提諾王國」的「卡提諾大聲公」、「伊莉討論區」的「時事討論」、Plurk、Ptt 的政黑版（HatePolitics）、Yahoo！奇摩知識＋下社會人文的政黨政治等等。且附錄圖 6 區塊之間的連結緊密，代表網友會同時在不同的頻道討論獨立立場。在附錄表 1，第一欄爲「全體聲量排序」，第二欄爲「獨立聲量排名」，在第二欄中若畫上向上的箭號者，代表箭號對應的頻道獨立聲量排序高於全體聲量排序，其中 Facebook、PTT 政黑版有獨立聲量排序高於全體聲量排序的情況。之所以以獨立聲量排序，原因在於獨立聲量較統一聲量多，較需要關注。

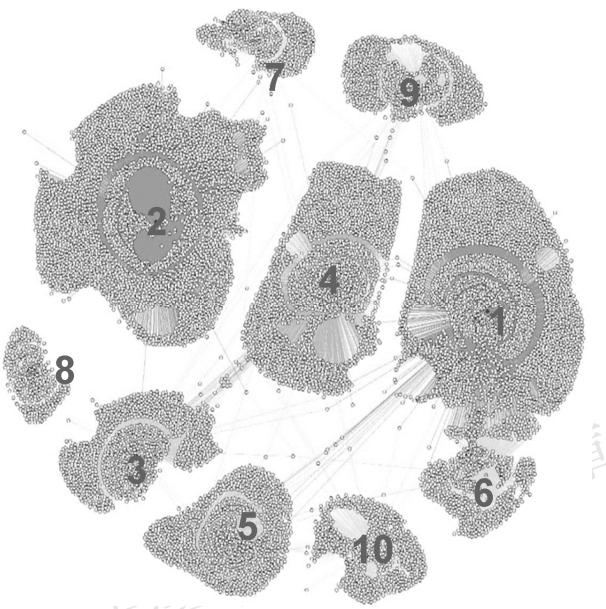
臺灣大學圖書館期刊資料庫

附錄表 1　獨立立場關鍵領袖出沒頻道

全體聲量排序	獨立聲量排序	頻道	文章數量	佔獨立聲量之比例
1	1	Ptt＞Gossiping	9,741	11.64%
4	2 ⬆	Facebook	6,735	8.04%
2	3	卡提諾王國＞卡提諾大聲公	6,248	7.46%
3	4	伊莉討論區＞時事討論（時事新聞）	6,240	7.45%
5	5	Plurk	4,461	5.33%
7	6 ⬆	Ptt＞HatePolitics	3,643	4.35%
6	7	Yahoo！奇摩知識＋＞社會人文－政黨政治	3,402	4.06%
8	8	巴哈姆特討論區＞講談說論	3,400	4.06%
9	9	mobile01＞新聞與時事（生活娛樂）	2,142	2.56%
11	10	痞客邦 PIXNET	1,589	1.90%

資料來源：本研究整理

附錄圖 6　獨立立場關鍵領袖出沒頻道示意圖
（每一點代表一個作者 ID）



資料來源：本研究自製

臺灣大學

另一方面，在附錄表 2 的第二欄中，統一立場關鍵領袖出沒頻道最多者爲「卡提諾王國」的「卡提諾大聲公」，佔獨立聲量的 14.42%；「伊莉討論區」的「時事討論」次之，佔獨立聲量的 11.49%；再者爲 PTT 的八卦版（Gossiping），佔獨立聲量的 10.73%。附錄圖 7 的區塊大小便是討論量的多寡，標示 1 的區塊便是「卡提諾王國」的「卡提諾大聲公」，其它依序爲「伊莉討論區」的「時事討論」、PTT 的八卦版、PTT 的政黑版、「Yahoo! 奇摩知識＋」下「社會人文」的「政黨政治」、「Facebook 公眾訊息」下的「FbPublicMsg」、「Plurk 搜尋」下的「Plurk 搜尋」等等。且區塊之間的連結相對附錄圖 6 較不緊密，代表網友較不會同時在不同的頻道討論統一立場。在附錄表 2 中，第一欄爲「全體聲量排序」，第二欄爲「獨立聲量排名」，在第二欄中若畫上向上的箭號者，代表箭號對應的頻道統一聲量排序高於全體聲量排序，其中「卡提諾王國」中的「卡提諾大聲公」、「伊莉討論區」的「時事討論」、PTT 的政黑版、「Yahoo! 奇摩知識+」下「社會人文」的「政黨政治」、「SOGO 論壇」下的「時事新聞討論區」、「捷克論壇」下的「兩岸政治論壇」有統一聲量排序高於全體聲量排序的情況。

本研究挑選了關鍵領袖的討論內容，獨立聲量排行最高的是 gustav350，發文篇數有 320 篇，臺灣提及率是 33%[5]，臺獨提及率是 11%，中國提及率是 81%，大陸提及率是 11%，中國大陸提及率爲 2%（提及率加總超過 100% 的原因在於每篇聲量會提及各種關鍵字）。gustav350 的討論範疇包括「如何能臺獨」、「促調臺獨黨綱」。獨立立場鮮明，期待民進黨以激進的手段推動臺灣獨立。在「臺灣」關鍵詞的發文範例有「車輪黨 1945 接收臺灣，卻是來了另一位總督」，在「臺獨」關鍵詞的發文範例爲「共產黨有資格操弄民族主義、高喊反臺獨嗎」，在「中國」關鍵詞的發文範例爲「高級外省人假借三民主義，統一中國」，在「大陸」關鍵詞的發文範例爲「大陸不民主，等大陸民主了再來和臺灣政治溝通」。需要特別說明的是，兩岸稱呼用語會隨著立場而改變，獨立立場者較常使用「臺灣」而非「臺獨」、且較常

[5] 網友 gustav 發文篇數 320 篇，臺灣提及率 33%，表示在 320 篇他發表的留言或文章中，有 105 篇（320×0.33）提到「臺灣」二字。因爲一篇文章會同時出現很多字，所以這些提及率加起來會超過 100%。

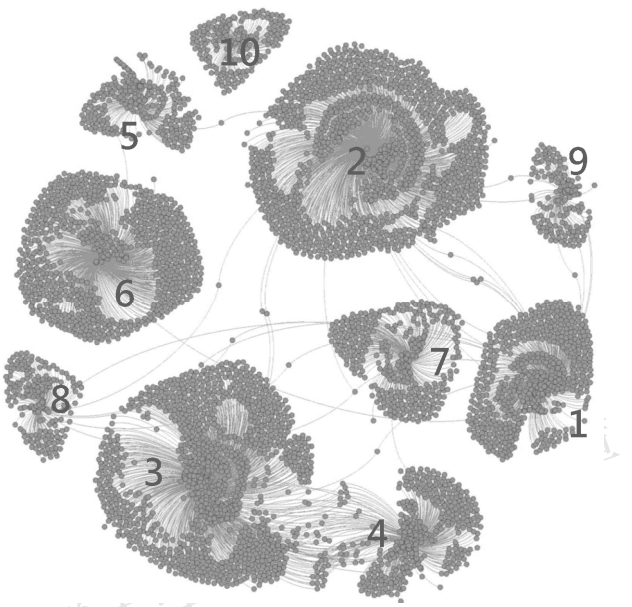
臺灣政治學術期刊資料庫

附錄表 2　統一立場關鍵領袖出沒頻道

全體聲量排序	統一聲量排序	頻道	文章數量	佔統一聲量之比例
2	1⬆	卡提諾王國＞卡提諾大聲公	2,599	14.42%
3	2⬆	伊莉討論區＞時事討論（時事新聞）	2,070	11.49%
1	3	Ptt＞Gossiping	1,934	10.73%
7	4⬆	Ptt＞HatePolitics	1,044	5.79%
6	5⬆	Yahoo！奇摩知識＋＞社會人文－政黨政治	965	5.36%
4	6	Facebook 公眾訊息＞FbPublicMsg	955	5.30%
5	7	Plurk 搜尋＞Plurk 搜尋	597	3.31%
10	8⬆	SOGO 論壇＞時事新聞討論區	446	2.48%
12	9⬆	捷克論壇＞兩岸政治論壇	442	2.45%
8	10	巴哈姆特討論區＞講談說論（電玩人專區）	386	2.14%

資料來源：本研究整理

附錄圖 7　統一立場關鍵領袖出沒頻道示意圖
（每一點代表一個作者 ID）



資料來源：本研究自製

臺灣大學

使用「中國」。[6]

獨立聲量排行次高的是段王爺，發文篇數有 258 篇，臺灣提及率是 70%，臺獨提及率是 15%，中國提及率是 76%，大陸提及率是 15%，中國大陸提及率爲 0%。段王爺的討論範疇包括「大陸人爲什麼無法接受臺灣獨立」、「臺灣獨立所需面對的問題該如何解決」。獨立立場鮮明，時常在 Yahoo! 奇摩知識＋爲臺獨辯護。在「臺灣」關鍵詞的發文範例有「如果沒有太陽花學運，臺灣與中國統一已鳴槍起跑了」，在「臺獨」關鍵詞的發文範例爲「支持臺獨是臺灣兩千三百萬人的信念」，在「中國」關鍵詞的發文範例爲「寧願與菲律賓統一也不願跟中國統一」，在「大陸」關鍵詞的發文範例爲「臺灣人何必去反攻大陸解救中國同胞」。

獨立聲量排行第三高的是 frogha1，發文篇數有 223 篇，臺灣提及率是 43%，臺獨提及率是 79%，中國提及率是 28%，大陸提及率是 28%，中國大陸提及率爲 3%。frogha1 的討論範疇包括「臺灣到底要怎麼獨立？」、「kmt 不倒，臺灣不會好」。獨立立場鮮明，是活躍於 PTT Gossiping 八卦版的臺獨支持者。在「臺灣」關鍵詞的發文範例有「如果臺灣被中國統一，血是必然要流的」，在「臺獨」關鍵詞的發文範例爲「臺獨其實才是臺灣的主流民意」，在「中國」關鍵詞的發文範例爲「以後全都是統媒，中國洗腦工程就大功告成」，在「大陸」關鍵詞的發文範例爲「蘇貞昌認爲中華民國包含臺灣地區跟大陸地區」。

統一聲量排行最高的是 qd11061，發文篇數有 390 篇，臺灣提及率是 33%，臺獨提及率是 11%，中國提及率是 81%，大陸提及率是 11%，中國大陸提及率爲 4%。qd11061 的討論範疇包括「如何能臺獨」、「促調臺獨黨綱」。獨立立場鮮明，期待民進黨以激進的手段推動臺灣獨立。在「臺灣」關鍵詞的發文範例有「對臺獨有利的必然不利臺灣」，在「臺獨」關鍵詞的發文範例爲「臺獨高喊阿扁無罪，這不是無的放矢嗎？和平統一的方式最好我怕打仗，民進黨到綠島去搞獨立吧。愛和平的選擇統一愛打仗的選擇臺獨。」，在「中國」關鍵詞的發文範例爲「中國人用的字就是中文」，在

[6] 林毓芝，2003，《《聯合報》、《中國時報》、《自由時報》統獨偏向與立論觀點之比較－以中共 1993 年與 2000 年兩次對臺白皮書爲例》，臺北：政治大學新聞學系碩士論文。

臺灣人與學術期刊資料庫

「大陸」關鍵詞的發文範例爲「開放大陸大學學歷，必定又會引爆臺獨新一輪抗議」。需要特別說明的是，稱呼方式隱含了認知或期望，統一立場網友較常使用「大陸」指稱中國大陸，且提及「臺獨」時多爲批評。[7]

統一聲量排行次高的是 winnie001，發文篇數有 390 篇，臺灣提及率是 47%，臺獨提及率是 72%，中國提及率是 25%，大陸提及率是 34%，中國大陸提及率爲 3%。winnie001 的討論範疇包括「臺灣未獨立臺灣人就是中國人」、「臺獨扭曲了 228」。統一立場鮮明，認爲臺灣人當然是中國人。在「臺灣」關鍵詞的發文範例有「其實，大陸沒有飛彈瞄準臺灣」，在「臺獨」關鍵詞的發文範例爲「統一的好處很多但有一樣不好那就是『臺獨當美日奴僕的希望落空了』」，在「中國」關鍵詞的發文範例爲「二岸統一後，可改名爲中國」，在「大陸」關鍵詞的發文範例爲「某些臺獨看到大陸強大就氣憤難平」。

統一聲量排行第三高的是 829102402，發文篇數有 369 篇，臺灣提及率是 27%，臺獨提及率是 63%，中國提及率是 19%，大陸提及率是 23%，中國大陸提及率爲 3%。829102402 的討論範疇包括「臺獨的悲哀每下愈況日趨下流」、「臺灣者，中國之土地也」。統一立場鮮明，認爲臺獨是全世界最大的詐騙集團。在「臺灣」關鍵詞的發文範例有「統一也是臺灣的選項，既然如此，掛五星旗也是應該的，生命比什麼都重要」，在「臺獨」關鍵詞的發文範例爲「何必在意臺獨呢！！！重點還是把經濟搞好吧」，在「中國」關鍵詞的發文範例爲「簡體字和繁體字都是中國字」，在「大陸」關鍵詞的發文範例爲「以大陸現在的實力，打日本輕而易舉的」。

臺灣人學學術期刊資料庫

7 林平，2009，〈從居住空間看臺灣人對中國大陸當地的認同〉，《臺灣政治學刊》，13(2): 57-111。

Public Opinion of Unification-Independence on the Web:

A Big-Data Analysis*

*Jia-wei Liu***

Abstract

The Unification-independence issue is an major political cleavage in Taiwan. This study analyzed the public opinion of unification-independence on the internet through big data. We found that in the distribution of unification-independence sound volume, the number of the independence sound volume was higher than that of unification. In the aspect of the target, “China” and “Taiwan” were mentioned the most. In the source of the sound volume, social media was the main source, and PTT is the main channel. Furthermore, the sentiments of sound volume were that there were 6 positive sound volume and 10 negative sound volume at the same time. The above results were not different from the results of the surveys, and it deserves to be taken seriously by the administration. The value of big data analysis is obvious, especially when online forum or social media rise.

Keywords: Unification, Independence, Unification-independence, Big Data, Internet

臺灣人文學術期刊資料庫

* DOI:10.6166/TJPS.71(113-166)

** Associate Professor, Department of Public Administration and Policy, National Taipei University. E-mail: jwliu@mail.ntpu.edu.tw

臺灣大學學術期刊資料庫