

# 電機宅男的社科叢林奇幻之旅

王銘宏

[tonymhwang@gmail.com](mailto:tonymhwang@gmail.com)

世新大學經濟系 資料科學概論



你哪位？

# 個人簡介

- 王銘宏 Ming-Hung Wang
- 1987年7月生於嘉義市
- 學歷
  - 國立台灣大學電機工程研究所博士 ( 2017 )
  - 國立清華大學通訊工程研究所碩士 ( 2010 )
  - 國立清華大學資訊工程學士 ( 2008 )



你不是學電機的嗎？跟政治有啥關係？

# 為什麼我踏入這個坑？

- 318學運

- 鄉民
- 會寫程式的鄉民
- 會寫程式又喜歡資料的鄉民

- 然後呢？

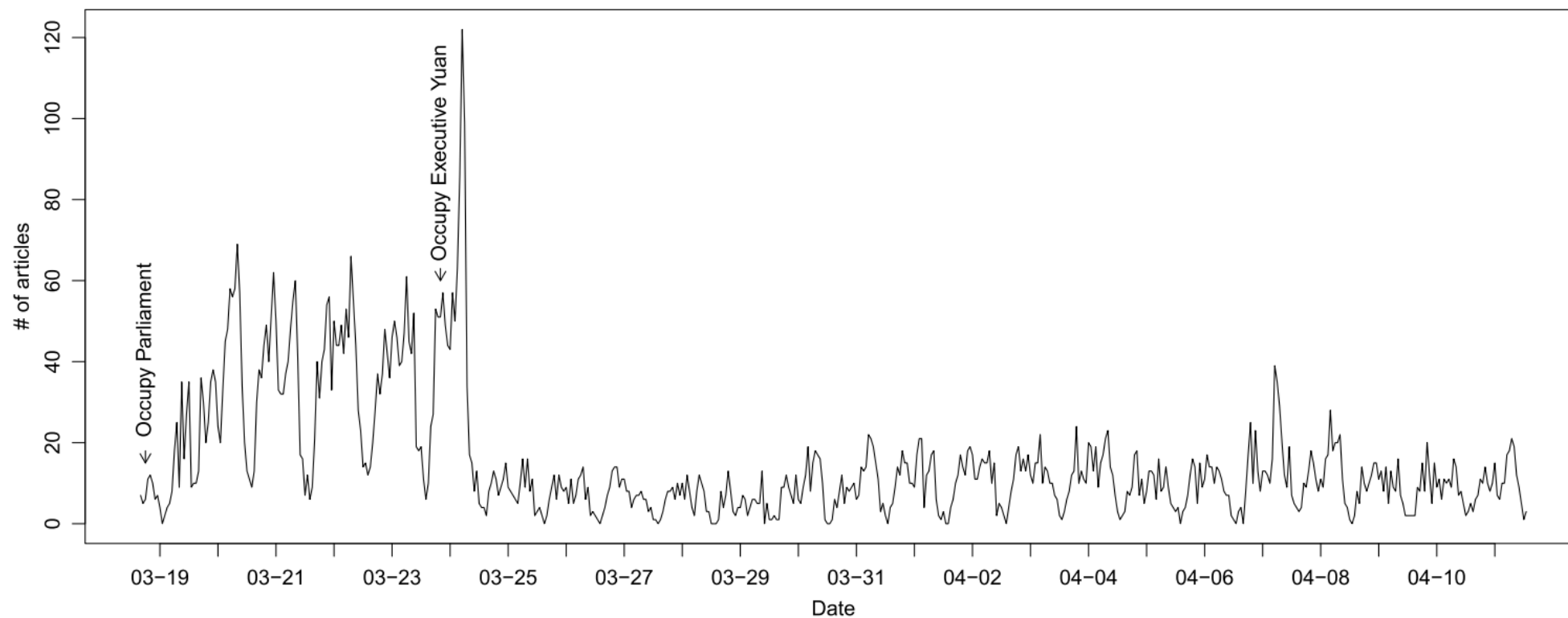
- PTT經歷 2004 年起
- 5000人同時上線
- 陪我成長的平台
- 研究看看

# PTT 也能研究？

- 學運時

- 鄉民瘋狂發文

- 鄉民瘋狂推文





我只想看重要文章，不想只是消勾勾

## 重要文章

## ● 看有沒有推爆

## 一面倒推

## 一面倒噓

## ● 看回文數量

## 一狂戰

—PTT 網紅

【板主: Bignana/RS5566/lovesooman/Kay731】 八卦板					系列《Gossiping》
[←]離開 [→]閱讀 [Ctrl-P]發表文章 [d]刪除 [z]精華區 [i]看板資訊/設定 [h]說明					人氣:16696
編號	日期	作者	文章	標題	
1066	+	10/25	MaInDog5566	<input type="checkbox"/> [問卦]	宮水三葉算母豬嗎?
1067	+	310/25	kazekaze	R: [問卦]	這次母豬事件誰最傷? 苗妞? 渣本? 蘇
1068	+	510/25	icepet0015	<input type="checkbox"/> [問卦]	有女權主義母豬化的八卦嗎?
1069	+	8610/25	pgimjjkk	R: [問卦]	這次母豬事件誰最傷? 苗妞? 渣本? 蘇美?
1070	+	4110/25	trustlove	R: [問卦]	這次母豬事件誰最傷? 苗妞? 渣本? 蘇美?
1071	+	310/25	KSHLO	R: [問卦]	這次母豬事件誰最傷? 苗妞? 渣本? 蘇美?
1072	+	610/25	bigloser5566	R: [問卦]	這次母豬事件誰最傷? 苗妞? 渣本? 蘇美?
1073	+	10/25	GasaiMayu	<input type="checkbox"/> [問卦]	母豬是否受折磨的八卦?
1074	+	210/25	tffc	R: [問卦]	這次母豬事件誰最傷? 苗妞? 渣本? 蘇美?
1075	+	410/25	ThorFukt	<input type="checkbox"/> [問卦]	革命機的女主角是母豬嗎?
1076	+	510/25	monocyte123	<input type="checkbox"/> [問卦]	母豬飼料多少錢
1077	+	310/26	andy2011	<input type="checkbox"/> [問卦]	兩津巡查長夠資格當母豬教主嗎?
1078	+	510/26	thundelet	<input type="checkbox"/> [問卦]	第一篇母豬文是蘇美人發的嗎?
1079	+	910/26	LKK549777	<input type="checkbox"/> [問卦]	如果一個交換, 渣男變渣女, 母豬變公豬?
1080	+	710/26	td123	<input type="checkbox"/> [問卦]	☺ 卜英奎☺ 母豬☺
1081	+	210/26	kilig	R: [問卦]	如果一個交換, 渣男變渣女, 母豬變公豬?
1082	+	410/26	jason0330	<input type="checkbox"/> [問卦]	帥到分手是在靠北母豬的歌嗎?
1083	+	210/26	ptt987654321	R: [問卦]	有沒有母豬教戰不倒的八卦?
1084	+	110/26	phnx	轉 [爆卦]	母豬教的起源: 社會階層的壓迫與箝制
1085	+	2110/26	there801021	<input type="checkbox"/> [問卦]	有沒有母豬戰到約砲去的八卦?

文章選讀 (y)回應(X)推文(^X)轉錄 (=□<>)相關主題(/?a)找標題/作者 (b)進板畫面



# 那我有沒有辦法可以早一點知道重要訊息？

- 學運期間
  - 衝突爆發
  - 內部訊息
- 選舉期間
  - 吸引目光
  - 重要議題



Photo credit: UDN.com

所以，我想在發文後一下子就知道文章  
會不會推(噓)爆？



此外，我也想在誰發文後一下子就會被  
推(噓)爆？

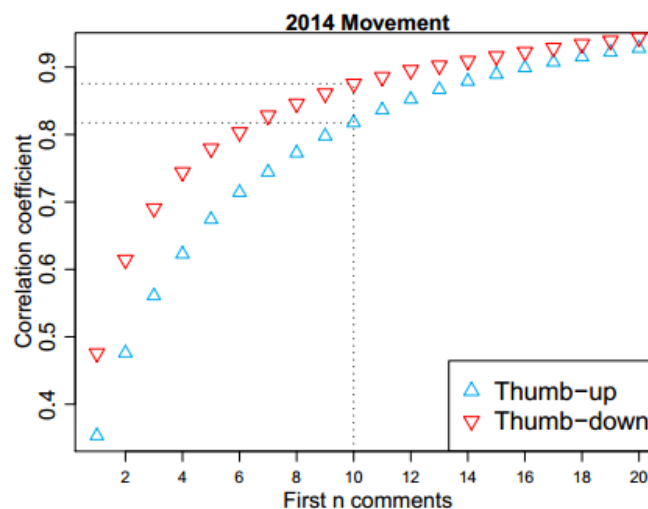
# 運用回歸模型預測推文數量

- 運用時序變化來做為自變數
  - 每隔5分鐘紀錄一次
- 文章屬性變數
  - 文字數
  - 發文時間
  - 發文在平日與否
- 回歸模型

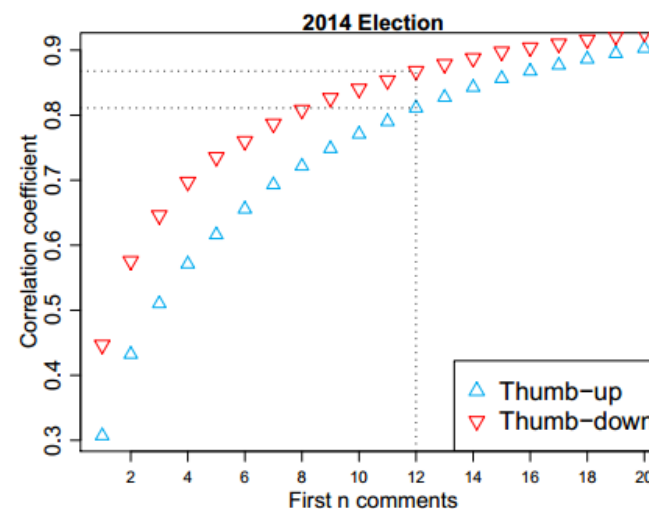
$$T_p = \alpha_0 + \sum_{i=1}^n \alpha_i D_{i,p} + \beta_1 Text_p + \beta_2 is.workhour_p + \beta_3 is.workday_p. \quad (1)$$

# 運用回歸模型預測推文分數

- 觀察前幾推結果推估文章是否會爆
  - 發現前10-12推噓比例跟整體推噓比例接近
  - 關聯性



(a) The Sunflower movement



(b) The 2014 election

# 運用回歸模型預測推文分數

- 早期觀察結果

- 前  $n$  個評論推噓比例

- 文章屬性變數

- 評論頻率

- 文章長度

- 關鍵字出現次數

- 回歸模型

$$Article.score.t_p = \text{sgn}(Article.score_p) * \log_{10}(|Article.score_p + 1|)$$

$$\text{sgn}(x) = \begin{cases} 1, & \text{if } a \geq 1, \\ -1, & \text{otherwise.} \end{cases}$$

$$\begin{aligned} Article.score.t_p = & \alpha_0 TU_{p,n} + \alpha_1 TD_{p,n} + \alpha_2 Time_{p,n}^* + \\ & \alpha_3 TU_{p,n} : TD_{p,n} + \alpha_4 TU_{p,n} : Time_{p,n}^* + \\ & \alpha_5 TD_{p,n} : Time_{p,n}^* + \beta_0 Text_p + \\ & \beta_1 Match_p. \end{aligned}$$

# 辨識誰是熱門(引戰)用戶？

- 發文評論數

- 引起熱烈討論

$$Consensus_p = \frac{Article.score_p}{R_p}$$

- 評論比例

- 一面倒推

- 一面倒噓

$$Author.score_v = \frac{\sum_{p \in v} Consensus_p}{Post_v}$$

- Consensus

- 共識程度

$$Influential.score_v = \frac{\sum_{p \in v} (Consensus_p * \log_{10}(R_p))}{Post_v}$$

- Author score

- 平均文章共識程度

- Influential score

- 共識程度加權文章評論數



# 案例分析



# 關鍵字 (以318事件為例)

Issue	Cross-Strait Service Trade Agreement
Location	Legislative Yuan, Executive Yuan
Abbreviation	318 movement, Sunflower Movement
Action	Occupy Parliament, Occupy Executive Yuan
Opposition	Chen Wei-Ting, Lin Fei-Fan, Dr. Huang Guo-Chang
Government	President Ma Ying-jeou, Premier Jiang Yi-huah, Parliament Chairman Wang Jin-pyng

# 蒐集的資料集

	2014 Movement	2014 Election
Start	2014/03/18 06:00:00	2014/09/30 06:00:00
End	2014/04/11 06:00:00	2014/11/30 06:00:00
# Articles	8,431	7,211
# Replies	609,232	623,340
Words/post	574.3	612.5
Ratings/post	72.3	86.4
# Users	49,682	36,452
# Authors	4,518	2,483
# Raters	48,832	35,983

# 過濾相關文章

---

**Algorithm 1** Filtering topic-related articles.

---

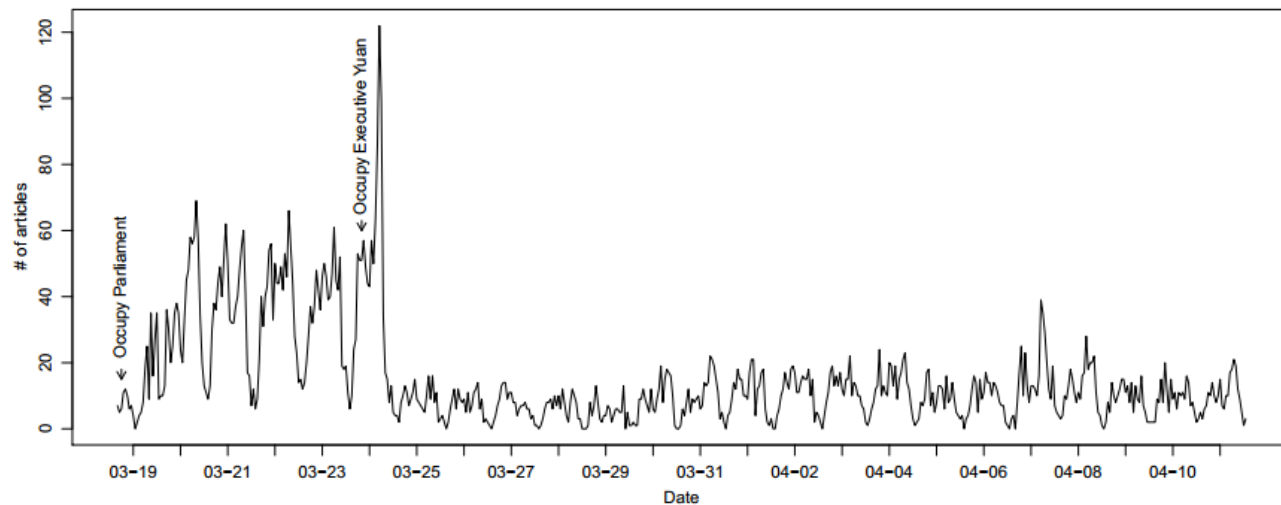
**Input:** Set of Chinese articles  $\mathcal{A}$ , Keyword lexicon  $\mathcal{L}$

**Output:** Set of articles related to the topic  $\mathcal{A}_{election}$

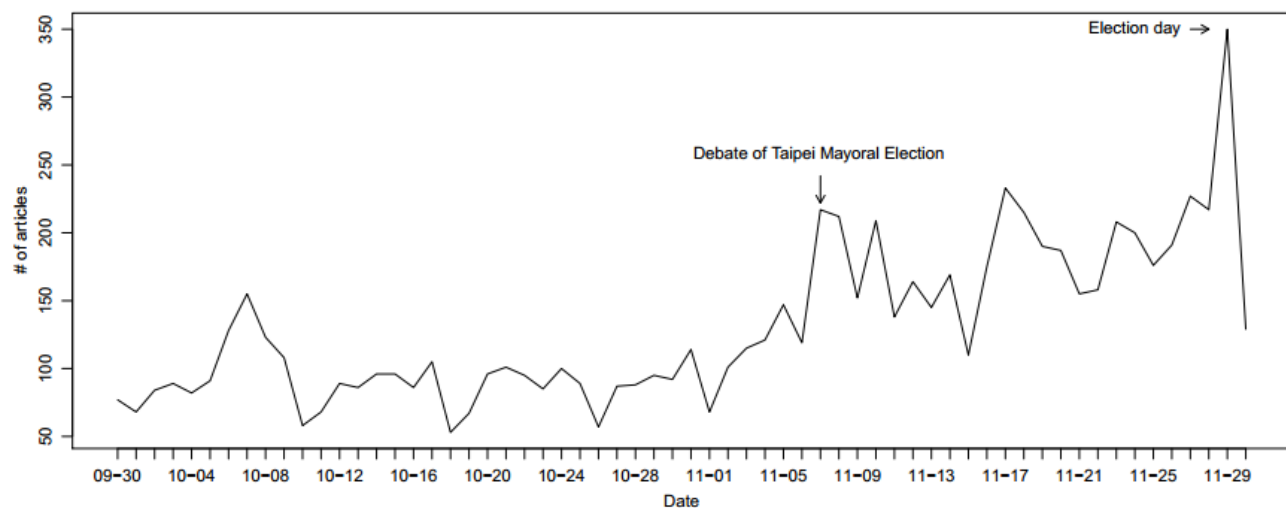
```
1: initial  $\mathcal{A}_{topic} \leftarrow \emptyset$ 
2: for every article  $A \in \mathcal{A}$  do
3:   words  $\leftarrow$  Chinese.segment( $A.text$ )
4:   for every word  $w \in words$  do
5:     if  $w \in \mathcal{L}$  and  $length(A.reply) \geq 20$  then
6:       Add  $A$  into  $\mathcal{A}_{topic}$ ;
7:       break;
8: return  $\mathcal{A}_{topic}$ ;
```

---

# 文章數量時序變化



(a) The Sunflower movement (hourly)



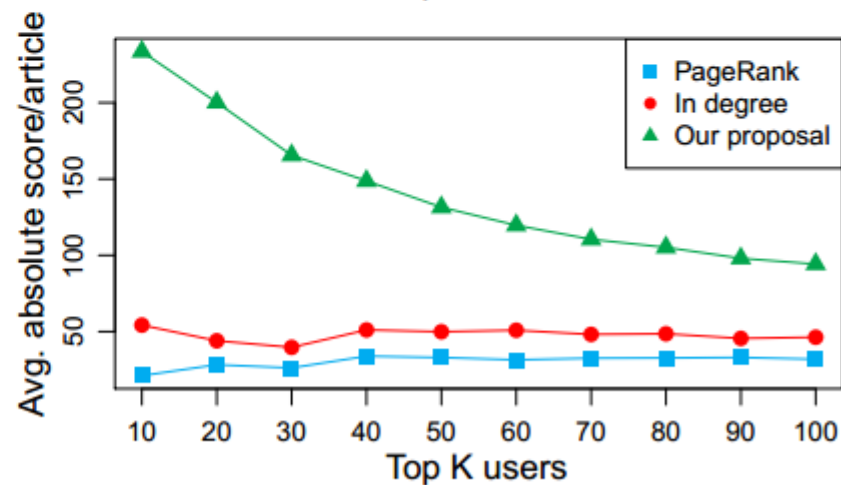
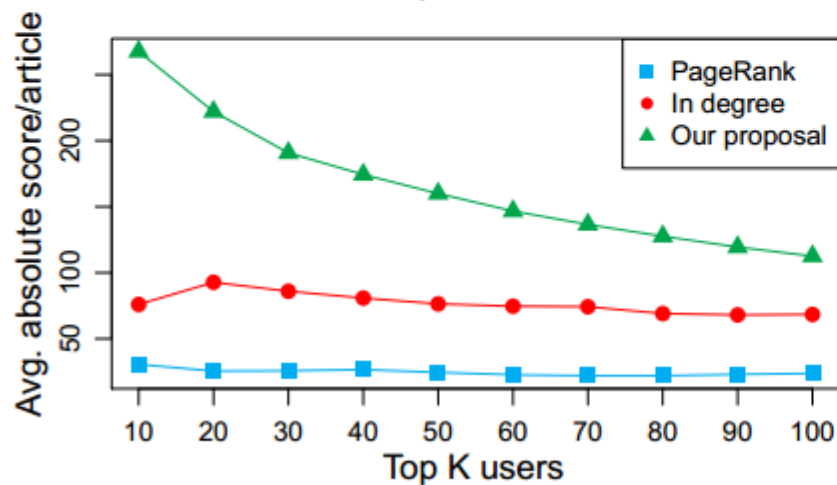
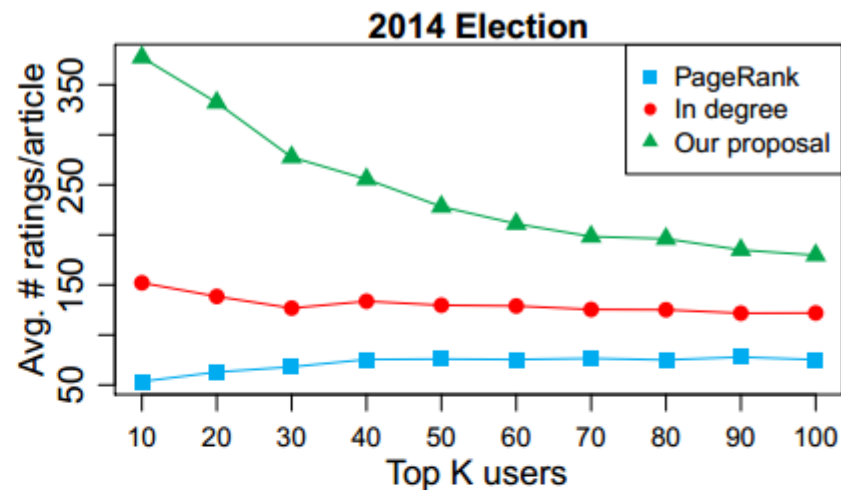
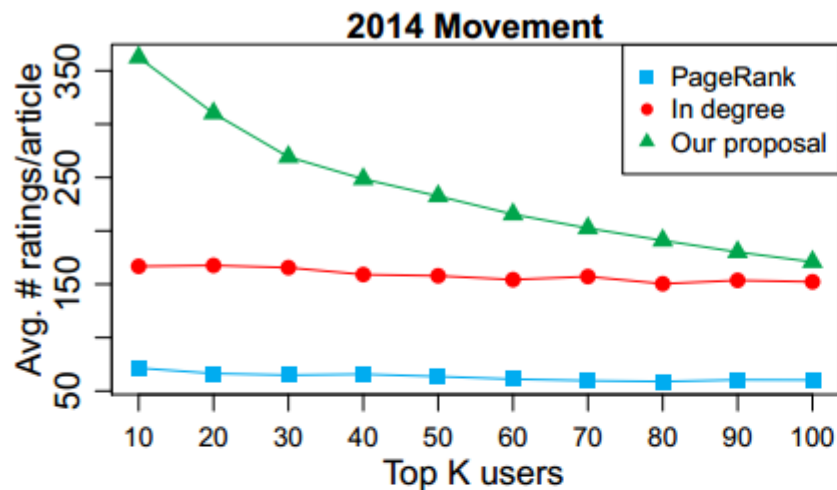
(b) The 2014 local election (daily)

# 不同的先期觀察數量與預測準確度

2014 Sunflower Movement				2014 Local Election			
$n$	Adj. $R^2$	RMSE	RMSE <sub>cv</sub>	$n$	Adj. $R^2$	RMSE	RMSE <sub>cv</sub>
10	0.740	0.624	0.627	12	0.718	0.686	0.687
11	0.756	0.603	0.606	13	0.735	0.666	0.667
12	0.773	0.583	0.586	14	0.751	0.645	0.646
13	0.786	0.565	0.568	15	0.767	0.624	0.625
14	0.796	0.552	0.555	16	0.778	0.609	0.610
15	0.808	0.536	0.538	17	0.787	0.596	0.597

\*RMSE<sub>cv</sub> denotes the average RMSE of 10-fold cross validation

# 意見領袖辨識效果



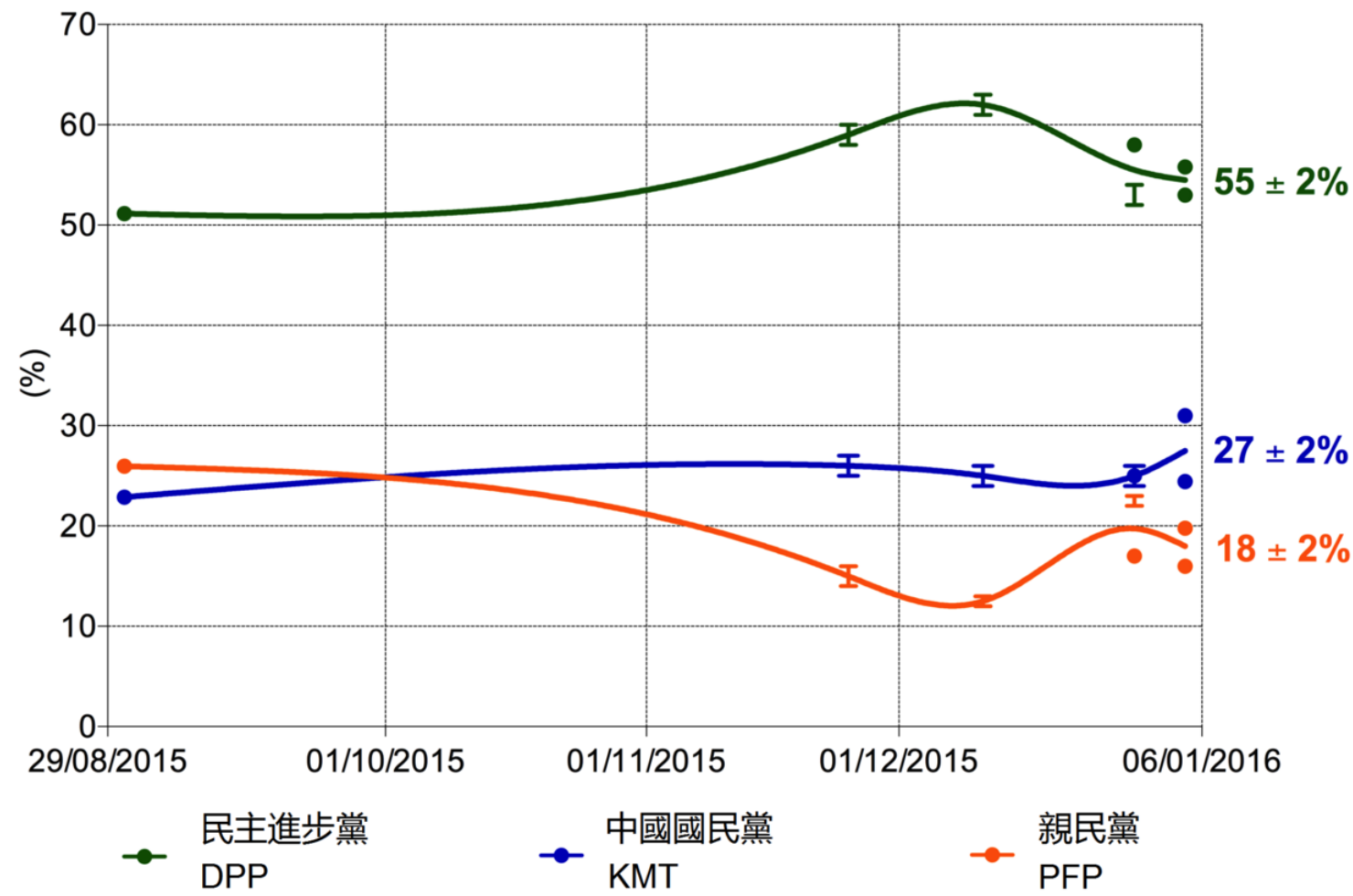
(a) The Sunflower movement

(b) The 2014 election




# 運用社群媒體進行選舉預測

# 2016總統選舉民調





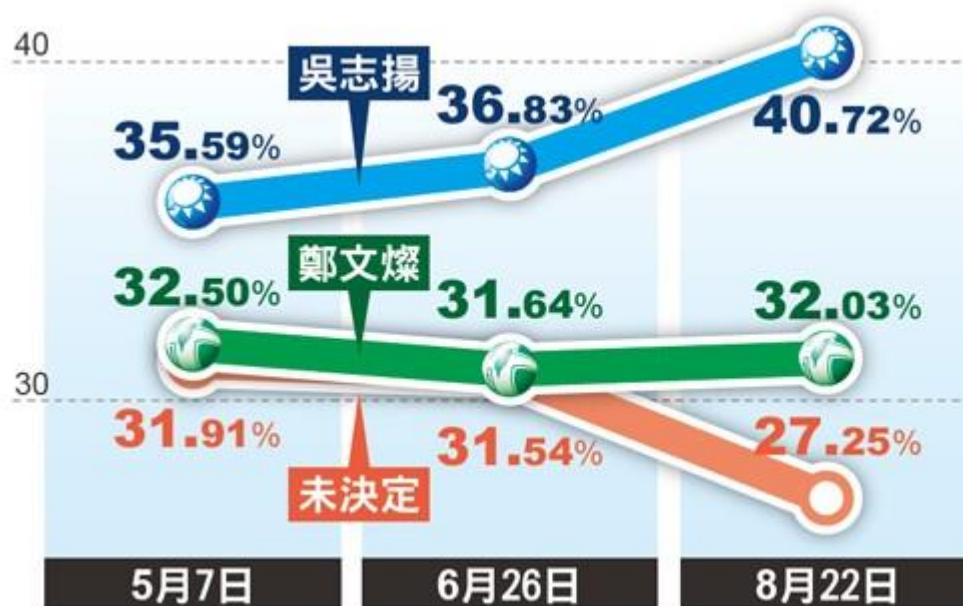
### 2016年中華民國總統選舉結果

號次	候選人	政黨	得票數	得票率	當選標記
1	朱立倫、王如玄	 中國國民黨	3,813,365票	31.04%	
2	蔡英文、陳建仁	 民主進步黨	6,894,744票	56.12%	
3	宋楚瑜、徐欣瑩	 親民黨	1,576,861票	12.84%	
選舉日期		2016年1月16日	選舉人數	18,782,991 <sup>[153]</sup> 人	
投票率		66.27%	投票人數	有效：12,284,970 <sup>[154]</sup> 人 無效：163,332 <sup>[154]</sup> 人	



準嗎？

# 2014地方選舉民調 (桃園)







資料來源：本報民調中心

桃園市長選舉民調

本報最近三次

# 2014年桃園市市長選舉結果

[隱藏]

號次	候選人	政黨	得票數	得票率	當選標記
1	鄭文燦	 民主進步黨	492,414	51.00%	
2	吳志揚	 中國國民黨	463,133	47.97%	
3	許睿智	 無黨籍	9,943	1.03%	
選舉人數			1,568,713		
投票數			984,099		
有效票			965,490		
無效票			18,609		
投票率			62.73%		

# 民意調查方法

## ●電話民調

### –最廣泛運用

- 有定義母群體，可對於電話抽樣調查結果進行完善的推估。
- 有非常嚴格的理論基礎。

### –but

- 電話訪問的涵蓋率出問題（抽樣母體的代表性問題）。
- 民眾不表態。

# 民意調查方法

## ●網路民調

—快速且便宜

—But

- 無法得知母群體與抽樣的關係
- 可能會有重複表態

—常出現誇張的結果



# 民意調查方法

- 社群媒體熱度

- 兩大方法

- 討論文章數量 A. Tumasjan et al. (2010)

- 計算候選人出現文章數量

- 候選人有多夯？

- 文章情緒分數 B. O' Connor et al. (2010)

- 計算文章情緒分數 (正負面)

- 候選人評價正反？

# 網路民調方法缺點 (戰點)

## ● 抽樣誤差

### — 人口抽樣誤差

- 年齡數位落差
- 居住地數位落差

### — 角色落差

- 活躍者
- 觀察者
- 沉默螺旋



# 網路民調方法缺點 (戰點)

- 上述兩種方法致命缺點

- 計算候選人文章數

- 高討論就代表高得票？
    - 文章灌水、洗文章都會導致預測失敗

- 計算候選人文字情緒分數

- 大家都說好就是好？
    - 沉默的多數怎麼辦？
    - 只有作者看法



怎麼解決？

# 運用更多群眾的智慧

- 群眾評分機制

- 鄉民的眼睛是雪亮的
- 你可以洗文章，我也可以噓爆你
- 群眾可以提供一個文章與事實查核的依據
- 評分門檻比發表低，更多人的意見可以被觀察到

- 採用各方法的長處

- 文章數 (候選人熱門程度)
- 情緒分數 (候選人相關的形容文字情緒)
- 公共接受度 (群眾的評分)

# 文章過濾與斷詞演算法

---

## Algorithm 2 Filtering candidate-related articles

---

### Input:

Set of Chinese articles  $\mathcal{A}$

An election candidate  $c$

### Output:

Set of articles related to the candidate  $c$  as  $\mathcal{A}_c$

```
1: initial  $\mathcal{A}_c \leftarrow \emptyset$ 
2: for every article  $A \in \mathcal{A}$  do
3:   words  $\leftarrow \text{jieba.segment}(A.\text{text})$ ;
4:   for every word  $w \in \text{words}$  do
5:     if  $w == c$  then
6:       Add  $A$  into  $\mathcal{A}_c$ ;
7:       break;
8: return  $\mathcal{A}_c$ ;
```

---

---

## Algorithm 3 Calculating sentiment scores of a candidate

---

### Input:

Election candidate  $c$ , Set of candidate-related articles  $\mathcal{A}_c$

Positive  $\mathcal{L}_{pos}$  and negative lexicon  $\mathcal{L}_{neg}$  from NTUSD

### Output:

Positive sentiment score of candidate  $c$   $Sent.pos_c$

Negative sentiment score of candidate  $c$   $Sent.neg_c$

```
1: initial  $Senti.pos_c \leftarrow 0$ ;  $Senti.neg_c \leftarrow 0$ ;
2: initial  $pos \leftarrow 0$ ;  $neg \leftarrow 0$ ;
3: for every article  $A \in \mathcal{A}_c$  do
4:   words  $\leftarrow \text{jieba.segment}(A.\text{text})$ ;
5:   for every word  $w \in \text{words}$  do
6:     if  $w \in \mathcal{L}_{pos}$  then
7:        $pos \leftarrow pos + 1$ ;
8:     if  $w \in \mathcal{L}_{neg}$  then
9:        $neg \leftarrow neg + 1$ ;
10:   $Sent.pos_c \leftarrow Sent.pos_c + pos$ ;
11:   $Sent.neg_c \leftarrow Sent.neg_c + neg$ ;
12: return  $Sent.pos_c$ ;
13: return  $Sent.neg_c$ ;
```

---

# 案例分析：2014 年地方選舉

- 針對六都選舉
- 選前三個月的文章
  - PTT八卦板
  - 含有六都候選人的文章

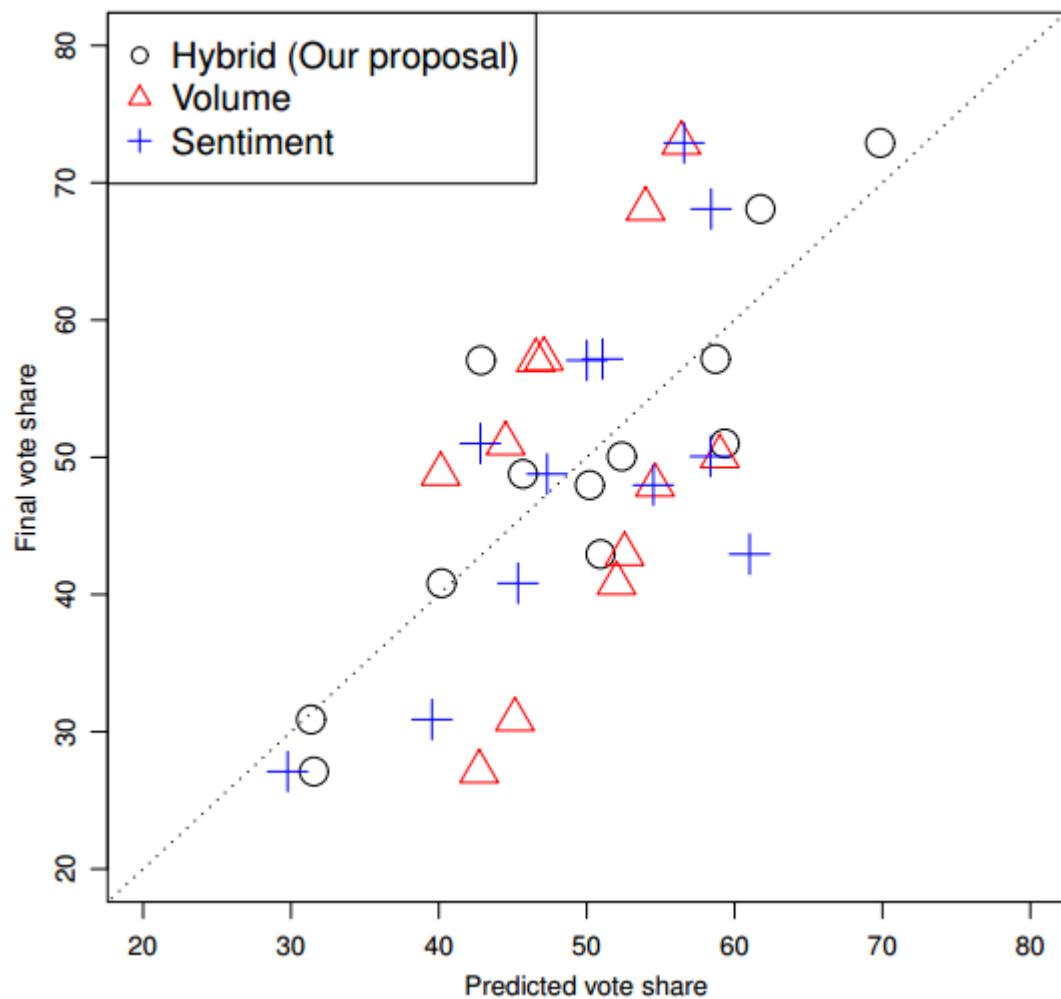
Table 4.1: A summary of our dataset

Start crawling	2014-08-29 06:00:00
End crawling	2014-11-29 06:00:00
Total articles	26,735
Total ratings	1,203,941
Avg. # ratings / article	45
Total users identified	47,658
Authors	5,843
Raters	46,625

# 回歸分析

- 自變數
  - 文章數
  - 文字情緒分數
  - 公共接受度
  - 變數間的 interaction
- 依變數
  - 兩大參選人得票率
- 比較先前方法

# 預測得票與實際得票



$$VS_p = (Volume_p + Sentiment_p + Public.Acceptance_p)^2$$

# 預測贏家與得票率絕對誤差

Table 4.2: Predicted vote shares of three methodologies (volume-based, sentiment-based, and the hybrid method proposed by us) and the absolute prediction errors.

City	Cha./Inc.	Volume	Sentiment	Hybrid	Vote Share	Winner Predicted
Taipei	Challenger	0.475	0.530	0.594	0.583	S, H
	Incumbent	0.525	0.470	0.406	0.417	
Taichung	Challenger	0.470	0.450	0.457	0.571	
	Incumbent	0.530	0.550	0.543	0.429	
Kaohsiung	Challenger	0.455	0.404	0.337	0.312	V, S, H
	Incumbent	0.545	0.596	0.663	0.688	
New Taipei	Challenger	0.405	0.448	0.466	0.494	V, S, H
	Incumbent	0.595	0.552	0.534	0.506	
Taoyuan	Challenger	0.449	0.440	0.542	0.515	H
	Incumbent	0.551	0.560	0.458	0.485	
Tainan	Challenger	0.431	0.345	0.311	0.271	V, S, H
	Incumbent	0.569	0.655	0.689	0.729	
MAE		0.111	0.076	0.040	-	



# 臉書粉絲頁政治傾向分析



我懷疑PTT是綠的  
我懷疑臉書是綠的

- 
- 
-

# 所以...



## GreenifyFacebook

提供者: Cindy Wang (CindyLinz)

★★★★★ (32)

[趣味休閒](#)

1,385 位使用者

+ 加到 CHROME

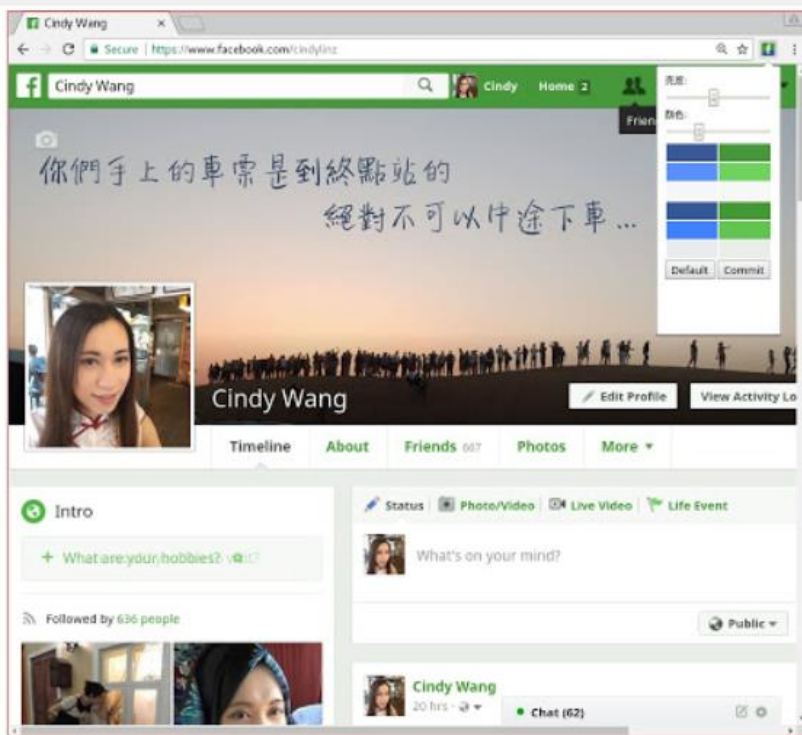


總覽

評論

支援

相關項目



與你的裝置相容

### Change the blue Facebook into a green Facebook

綠化 Facebook 的畫面

Source code:  
<https://github.com/CindyLinz/Chrome-GreenifyFacebook>

CHANGES:

v1.9 - 可以調整為其他色調, 效能略為改善

v1.8 - 點文章的「時間」的時候會進入顯示該篇文章的畫面, 而剛載入的時候會出現藍色框框若

! 檢舉濫用情形

### 其他資訊

版本: 1.9

更新日期: 2017年8月5日

大小: 9.34KiB

語言: 中文 (繁體)

# 發現文獻

- Bond, R., & Messing, S. (2015). Quantifying social media' s political space: Estimating ideology from publicly revealed preferences on Facebook. *American Political Science Review*, 109 (1), 62-78.
  - 使用臉書共同按讚行為評估政治人物政治光譜位置
  - 使用記名投票結果來驗證

# 我們能做什麼

- 做一樣的事情，場域在台灣？
  - 有，等等會講。
- 除了人，是不是有哪些粉絲頁是藍綠聚集地？
  - 哪些粉絲頁是網軍？
- 除了政治人物，常有人或組織自稱中壢李姓選民。
  - 是真中壢？
  - 還是假中立？

# 粉絲頁政治傾向推估方法

- 蒐集政治人物共同按讚人數
  - 台灣 113 立委
  - 建立  $113 \times 113$  的 matrix
  - 再針對粉絲頁人數作 normalization
- 作奇異值分解 (Singular Value Decomposition)
- 取前幾個左奇異向量當作推估結果
  - 取第一？
  - 取第二？
  - 哪個比較好？

# 粉絲頁政治傾向推估方法

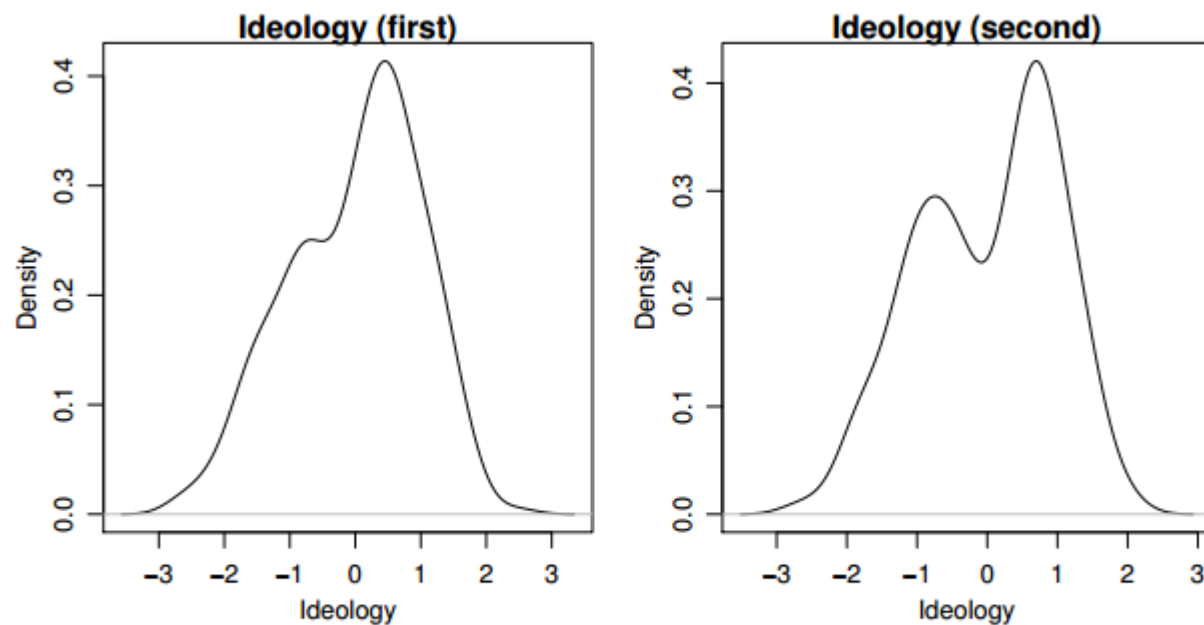


Figure 5.2: The density of ideology scores derived from the first and second left singular values of the SVD results.

# 使用者政治傾向推估方法

- 方法一：

- 使用推估粉絲頁的方法
- 建構 70 million \* 70 million 的使用者共同按讚粉絲頁數量矩陣
- 不可行 且 慢...

- 方法二：

- 使用剛剛算出來的113個粉絲頁政治傾向分數
- 看使用者按藍的多還是按綠的多
  - 取平均
  - 但沒有按政治人物粉絲頁的使用者我們就沒辦法評估



# 普遍粉絲頁政治傾向推估方法

- 使用推估的使用者政治傾向分數
- 看使用者又去按了哪些粉絲頁
  - 依照到該粉絲頁讚使用者取平均政治傾向分數
  - 泛藍/綠支持者常按哪些粉絲頁
    - 評估粉絲頁政治性
    - 評估粉絲頁政治傾向

# 資料集

Item	Value
Crawl period	2015/03 – 2015/09
# of like records	5,395,078,061
# of posts	28,702,869
# of pages	212,022
# of users	70,089,483

資料來源: QSearch

# 台灣粉絲頁類型統計

Category	Percentage
Community	10.27%
Artist	3.95%
Musician/Band	3.63%
Public Figure	3.31%
Clothing	3.28%
Local Business	3.17%
Fictional Character	2.58%
Product/Service	2.55%
Personal Blog	2.44%
Restaurant/Cafe	2.27%

# 立委位置圖

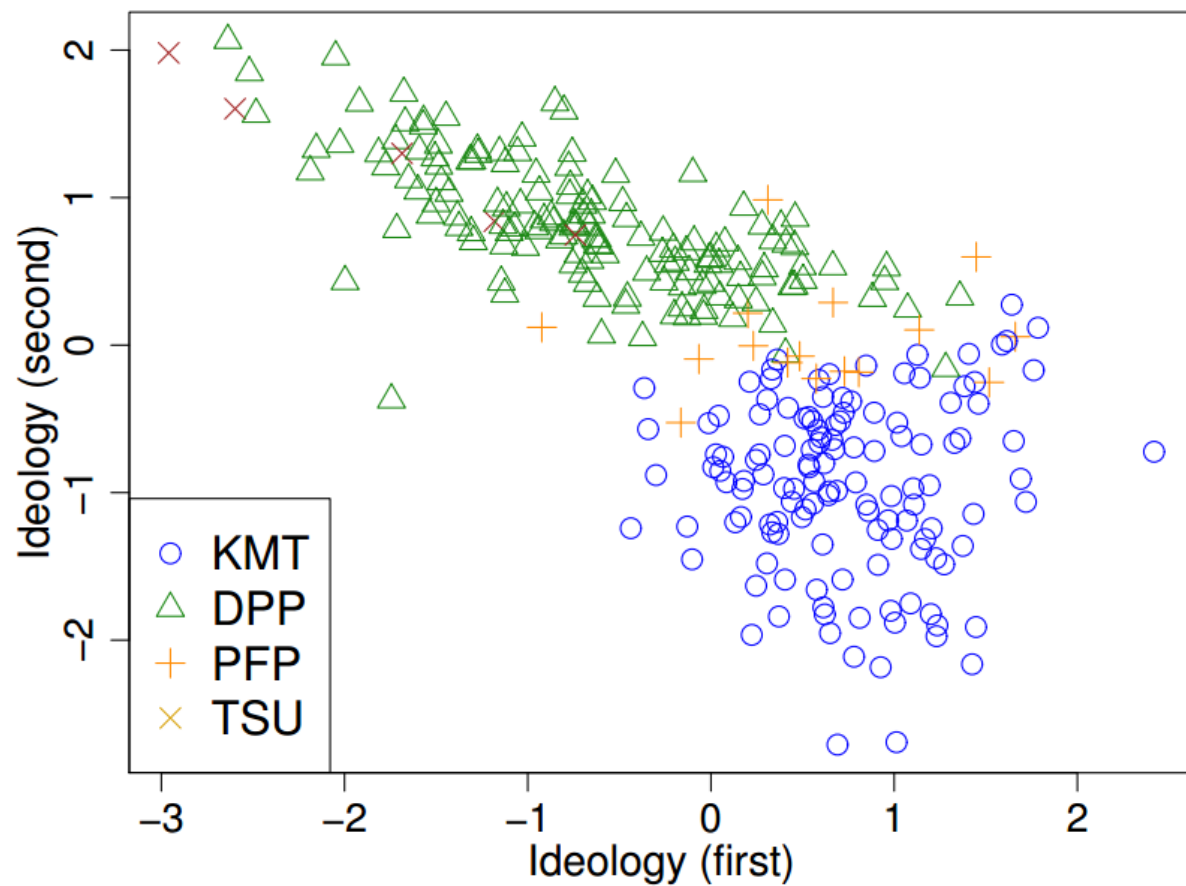


Figure 5.1: Ideology scores of all political pages from the first and second left singular values of the SVD results.

# 第一及第二左奇異向量分布

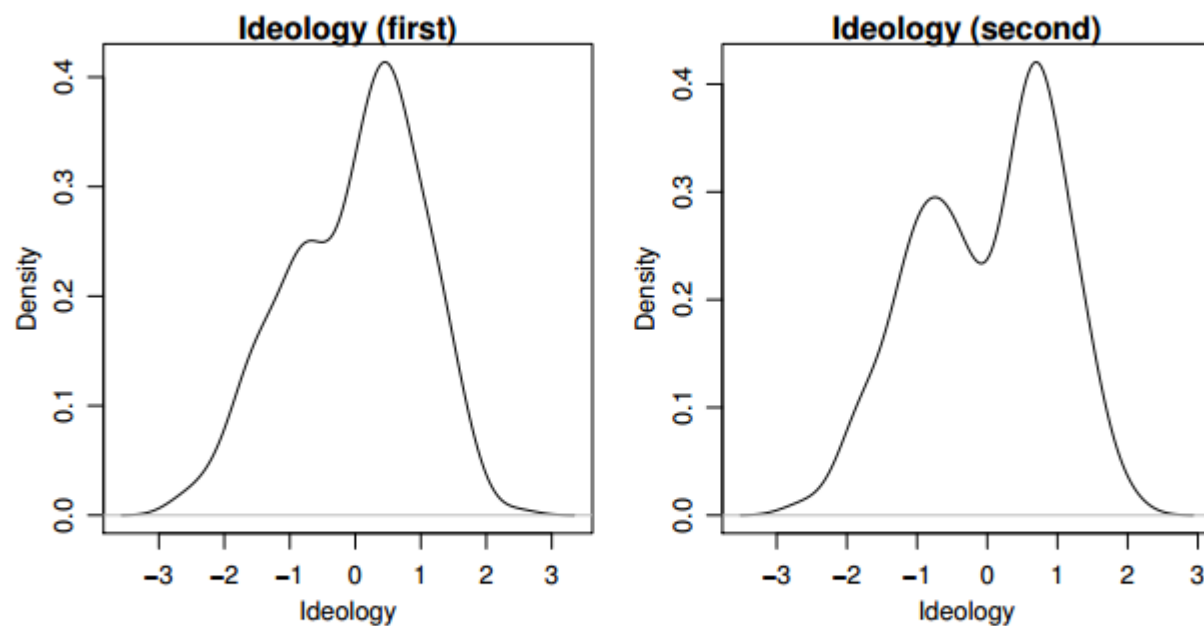


Figure 5.2: The density of ideology scores derived from the first and second left singular values of the SVD results.



哪些粉絲頁最藍跟最綠？

# 政治性高且極藍/綠

Table 5.2: Top Polarized and High Politicalness Pages

Label	Page name	Ideology	Politicalness
B1	Xiong Xiong Xiang Dao Bu Dui Jin (熊熊想到不對勁)	-1.46	0.87
B2	Fei Die Wu Can - Dong Zhi-Sen Shi Jian (飛碟午餐-董智森時間)	-1.46	0.81
B3	Ting Hung Hsiu-Chu Can Xuan Zhong Guo Guo Min Dang Dang Zhu Xi (挺洪秀柱參選中國國民黨黨主席)	-1.43	0.80
B4	Yu Zhou Hua Yuan (宇宙花園)	-1.41	0.83
B5	Lan Jun Jiao Fu Tang Xiang-Long (藍軍教父唐湘龍)	-1.41	0.83
G5	ZM	0.84	0.89
G4	Tai Wan Du Li Jian Guo Da Qi Dui (台灣獨立建國大旗隊)	0.85	0.93
G3	Du Tai Xin She (獨臺新社 Taiwan Independence Reformation Association)	0.85	0.93
G2	Professor Tsay Ting-Kuei - Tai Wan Jian Chi De Li Liang (蔡丁貴教授-台灣堅持的力量)	0.85	0.90
G1	Hsu Chun-Hsin (許忠信)	0.88	0.92

- Ideology: 政治光譜位置
- Politicalness: 政治性
- 極端藍綠
- 高政治性

# 不同類別粉絲頁藍綠與政治性分布

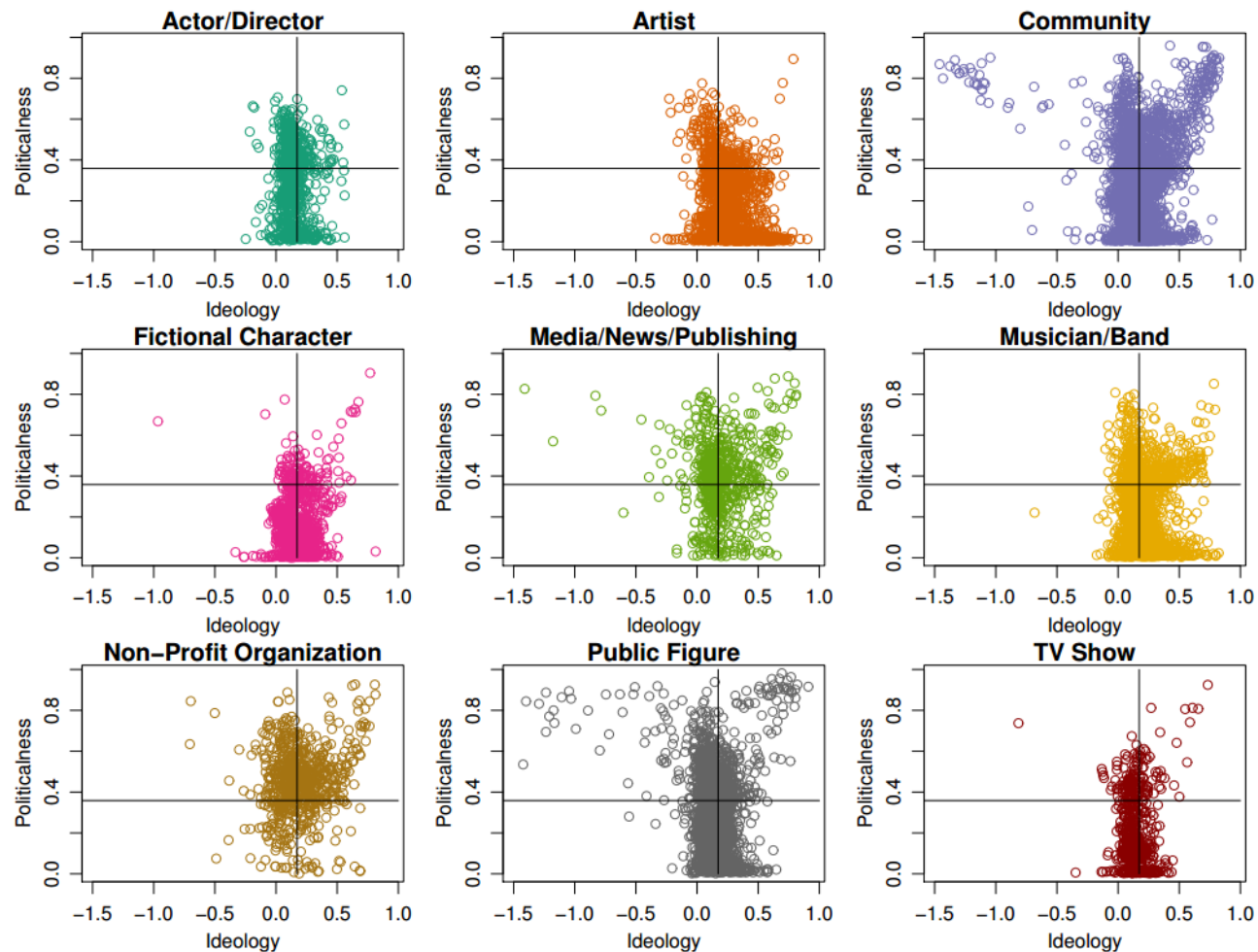


Figure 5.4: The ideology of Facebook pages grouped by different categories.



# 藍綠粉絲頁- Community 類別

Table 5.3: Top Polarized Pages in the Community Category

Label	Page name	Ideology	Politicalness
CB1	Xiong Xiong Xiang Dao Bu Dui Jin (熊熊想到不對勁)	-1.46	0.87
CB2	Ting Hung Hsiu-Chu Can Xuan Zhong Guo Guo Min Dang Dang Zhu Xi (挺洪秀柱參選中國國民黨黨主席)	-1.43	0.80
CB3	Zheng Yi Wu Di, Li Ting Ma Ying-Jeou (正義無敵・力挺馬英九)	-1.37	0.86
CB4	President Ma Ying-Jeou Hou Yuan Hui (馬英九總統後援會)	-1.33	0.89
CB5	Zhong Hua Min Guo - Ma Ying-Jeou Wang Lu Hou Yuan Tuan (中華民國-馬英九網路後援團)	-1.33	0.85
CG5	Wu Xian Qi Zhi Chi Zhan Shen Kuo-Chang Lao Shi (無限期支持戰神國昌老師)	0.81	0.89
CG4	Tai Peng Guo Ji Fa Fa Li Jian Guo Lian Xian (臺澎國際法法理建國連線)	0.81	0.80
CG3	Tai Da Fa Lu Xue Sheng Ting 318 (台大法律學生挺 318)	0.82	0.83
CG2	Han Wei Tai Wan Wen Shi Qing Nian Zu He (捍衛台灣文史青年組合)	0.82	0.85
CG1	Qian Ze Fang Yang-Ning Ji Qi Zhang Guan Wei Xian Bao Duo Gong Tou Meng Ji Hui Quan Li (譴責方仰寧及其長官違憲剝奪公投盟集會權利)	0.83	0.90

# 藍綠粉絲頁- Musician/Band 類別

Table 5.4: Top Polarized Pages in the Musician/Band Category

Label	Page name	Ideology	Politicalness
MB1	Taipei Wind Orchestra & Symphonic Band	-0.10	0.53
MB2	臺北市立國樂團 (Taipei Chinese Orchestra)	-0.08	0.65
MB3	周華健 (Emil Wakin Chau)	-0.08	0.42
MB4	金佩珊 (Jin Pei-Shan)	-0.07	0.60
MB5	台北香頌室內樂團 (Taipei Chansontpe Chamber Orchestra)	-0.06	0.55
MG5	2HRs	0.74	0.52
MG4	農村武裝青年 (Nong Cun Wu Zhuang Qing Nian)	0.74	0.73
MG3	拍謝少年 (Pai Xie Shao Nian)	0.75	0.60
MG2	淡水男孩-Tamsui BoyZ (Dan Shui Nan Hai)	0.78	0.85
MG1	濁水溪公社 LTK Commune (Zhuo Shui Xi Gong She)	0.79	0.73

# 藍綠粉絲頁- Publisher 類別

Table 5.5: Top Polarized Publishers

Label	Page name	Ideology	Politicalness
PB1	道聲出版社	0.05	0.37
PB2	大田出版	0.08	0.49
PB3	四也童書出版公司	0.08	0.70
PB4	小熊出版社 LittleBearBooks	0.13	0.53
PG4	唐山書店／唐山出版社	0.66	0.58
PG3	衛城出版	0.74	0.60
PG2	玉山社出版公司	0.78	0.75
PG1	前衛出版社	0.81	0.80

# 藍綠粉絲頁- 主要媒體

Table 5.6: The Ideology and Politicalness of Mass Media in Taiwan

Label	Page name	Ideology	Politicalness
N1	ChinaTimes Breaking News (中時電子報 【即時新聞】頻道)	-0.46	0.68
N2	CTV News (中視新聞)	-0.34	0.53
N3	DaAi World News (大愛全球新聞)	-0.11	0.38
N4	United Evening News (聯合晚報)	-0.07	0.57
N5	GOOD TV (GOOD TV 好消息電視台)	-0.01	0.38
N6	工商時報	0.01	0.45
N7	財經新報	0.19	0.59
N8	NOWnews 今日新聞	0.20	0.48
N9	Liberty Times (自由時報)	0.21	0.37
N10	The Epoch Times - Taiwan (大紀元時報 - 台灣)	0.23	0.58
N11	New Tang Dynasty Television (新唐人亞太 電視台)	0.30	0.78
N12	ERA News (年代新聞)	0.31	0.58
N13	SET News (三立新聞)	0.37	0.60

# 攻擊性/反諷性粉絲頁

- 反串
  - 無限期支持連勝文參選總統
- 攻擊
  - 柯文哲有病
- 反諷語意分析尤其困難
  - 好棒
  - 好棒棒

# 自動化偵測相關反諷/攻擊粉絲頁

- 看政治人物與該粉絲頁政治傾向落差
  - 支持的粉絲頁大多會落在附近 (追蹤者屬性大多類似)
- 落差非常多的
  - 無限期支持連勝文參選總統 vs. 連勝文
  - 胡志強的芭樂園 vs. 胡志強
  - 柯文哲有病 vs. 柯文哲

# 實際案例

Table 5.7: Largest Ideology Difference between Pages and The Mentioned Politicians

Lable	Page name	Mentioned	Diff.	Type
AS1	Lai Ching-Te You Zhong Bu Jin Yi Hui, Mei Zhong Bu Jin Shi Fu! (賴清德有種不進議會，沒種不進市府！)	Lai Ching-Te (賴清德)	-1.84	opposed
AS2	Qu Zhu Ker Chien-Ming Li Kai Guo Hui Lian Meng (驅逐柯建銘 離開國會聯盟)	Ker Chien-Ming (柯建銘)	-1.82	opposed
AS3	Da Jia Yi Qi Qu Chuai Chiu Yi-Ying De Men, Biao Da Si Fa Bu Gong De Fen Nu (大家一起去踹 邱議瑩的門，表達司法不公的憤怒)	Chiu Yi-Ying (邱議瑩)	-1.79	opposed
AS4	Wu Xian Qi Deng Dai Lai Ching-Te Shi Zhang Zun Shou Zheng Zhi Cheng Nuo! Zheng Zhi Yao Cheng Xin Bu Ke Shua Wu Lai! (無限期等待賴清德市長遵守政治承諾！政治要誠信不可耍無賴！)	Lai Ching-Te (賴清德)	-1.66	opposed
AS5	Tsai Ing-wen Chao Di Quan Gui, Min Jin Dang Tan Yu Shou Hui Mei Ri Lan Dang (蔡英文炒地權 貴、民進黨貪汙收賄媚日爛黨)	Tsai Ing-wen (蔡英文)	-1.65	opposed
AS6	Ko Wen-je You Bing (柯文哲有病)	Ko Wen-je (柯文哲)	-1.64	opposed
AS7	Wu Xan Qi Zhi Chi Lien Sheng-Wen Can Xuan Zong Tong (無限期支持連勝文參選總統)	Lien Sheng-Wen (連勝文)	1.63	sarcastic
AS8	Hu Chih-Chiang De Ba Le Yuan (胡志強的芭樂園)	Hu Chih-Chiang (胡志強)	1.62	sarcastic
AS9	Zhong Jie Guo Min Dang! Ma Ying-Jeou Xia Tai! Huan Gei Tai Wan Gong Ping Zheng Yi! (終結 國民黨！馬英九下台！還給台灣公平正義！)	Ma Ying-Jeou (馬英九)	1.61	opposed
AS10	Wo Ai Lien Sheng-Wen (我愛連勝文)	Lien Sheng-Wen (連勝文)	1.60	sarcastic

# 落差分布

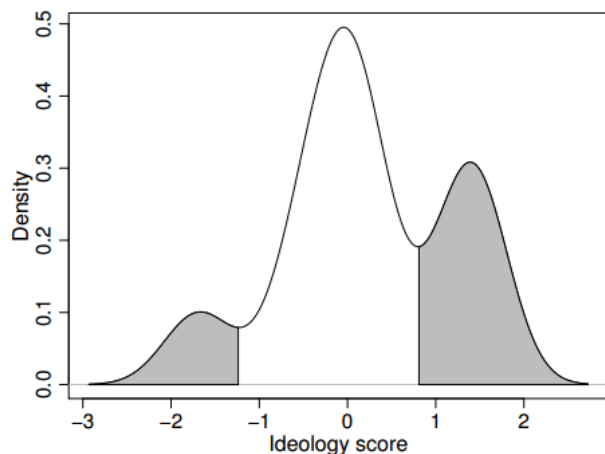


Figure 5.5: The distribution of the ideology score difference between pages and their affiliated politicians' pages. Pages in the grey area are identified as opposed/sarcastic pages.

- 把粉絲頁名稱有政治人物名字的挑出來
- 計算跟官方粉絲頁分數落差
- 密度圖
  - 找出兩個 turning points
  - 灰色部分推估為反諷/攻擊
- 人工驗證
  - 30/31
  - 96.97% 正確度



# 未來研究方向

- 其他的預測標的
  - 股票 (老題目，但做不好)
  - 其他選舉與地區
- 議題立場
  - 脫歐、反恐/親恐、同婚
- 應用到其他領域
  - 泛藍的人都到哪裡吃飯？
  - 泛綠的人都到哪裡運動？
  - 結合各種類型的社群媒體資料



感謝聆聽