

你的同溫層，暖嗎？

課程：R語言與資料科學

授課教師：謝舒凱

第七組 匿名@yunchiapig

許芸嘉 陳柄瑞

陳韋傑 陳信德

目錄

- ◆ 研究動機
- ◆ 研究方法
- ◆ 研究步驟
- ◆ 研究結果及討論

研究動機

研究動機

- 迴聲室效應 (Echo Chamber Effect)
- 封閉的環境中，相近意見的聲音不斷重複
- 俗稱：「同溫層」
- 大家的同溫層，究竟多厚呢？

PTT 批踢踢實驗坊



- 成立於1995/09/14
- 註冊人數 150~200百萬
- 同時上線人數 10 萬
- 25-45歲 上班族居多
- 政治、時事 熱度最高

Dcard



- 成立於2011/12/16
- 註冊人數 約100百萬
- 每日參訪人數 4.25 萬
- 18-25歲 大學生居多
- 感情、課業 討論為多

PTT 批踢踢實驗坊

八卦版(Gossiping)

女版(WomenTalk)

美妝版(MakeUp)

台大版(NTU)



Dcard



時事版(trending)

女孩版(girl)

美妝版(makeup)

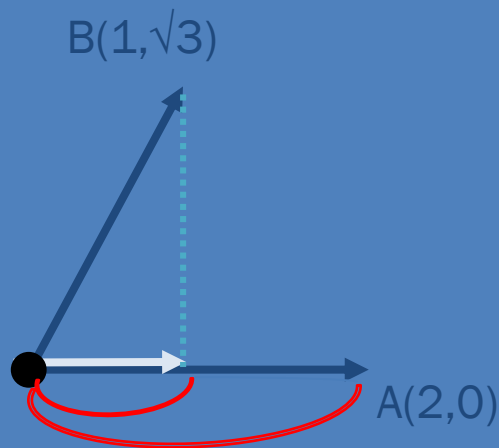
台大版(ntu)

研究方法

- ▣ 分析文章「內容」之一致性

- ▣ 一致性的指標 → Cosine Similarity 餘弦相似性

Cosine Similarity 餘弦相似性



$$\text{Cosine_similarity}(A,B) = 0.5$$

Cosine Similarity 餘弦相似性

	請	大家	IG	追蹤	@	yunchiapig
文章1	0.2	0.27	0.01	0	0.03	0.49	
文章2	0.1	0.3	0.05	0.3	0.15	0.1	
.....							

→ 文章1 (0.2 , 0.27 , 0.01 , 0 , 0.03, 0.49)

→ 文章2 (0.1 , 0.3 , 0.05 , 0.3 , 0.15 , 0.1)

Cosine Similarity 餘弦相似性

Terms



	請	大家	IG	追蹤	@	yunchiapig
文章1	0.2	0.27	0.01	0	0.03	0.49	
文章2	0.1	0.3	0.05	0.3	0.15	0.1	
.....							



Documents

Document-term matrix

(DTM)

研究假設

□ Dcard > PTT

□ 美妝 > 台大 > 女孩 > 八卦

研究步驟

研究步驟

1. API爬文

2. 文章內容斷詞

3. 建立DTM

4. 計算Cosine Similarity

1. API 爬文

```
```{r}
爬1000篇最新文章，將1000篇文章之id存入向量"idxs"
library(stringr)
library(httr)
idxs <- vector(mode = "character", length = 200)

req <- GET('https://www.dcard.tw/',
 path = c("_api", "forums", "trending", "posts?popular=false&limit=100"))
```

```
```{r}
# PTT 八卦版爬到之前Dcard所爬到186篇文最早的時間（即1/3 確保資料同為相同時段）
mine_ptt(ptt.miner = rookie_miner2,
  board = "Gossiping",
  min.date = "2020-01-03")
```
```

🐼 PTTmineR mining from ptt on your setting ...Item 2 has 0 rows but longest item has 1; filled with NAItem 3 has 0 rows but longest item has 1; filled with NAItem 4 has 0 rows but longest item has 1; filled with NAItem 5 has 0 rows but longest item has 1; filled with NA條件的長度 > 1，因此只能用其第一元素t  
longest item has 1; filled with NAItem 3 has 0 rows but longest item has 1; filled with NAItem 4 has 0 rows but longest item has 1; filled with NAItem 5 has 0 rows but longest item has 1; filled with NA

# 1. API 爬文

| 看板   | 八卦        |       | 女孩         |       | 美妝          |       | 台大       |       |
|------|-----------|-------|------------|-------|-------------|-------|----------|-------|
| 平台   | PTT       | Dcard | PTT        | Dcard | PTT         | Dcard | PTT      | Dcard |
| 時間   | 1/3 ~ 1/4 |       | 12/7 ~ 1/3 |       | 12/19 ~ 1/4 |       | 9/24~1/3 |       |
| 原始篇數 | 2201      | 186   | 3410       | 1000  | 211         | 1000  | 601      | 1000  |
| 抽樣篇數 | 186       | 186   | 500        | 500   | 211         | 211   | 500      | 500   |

## 2. 文章內容斷詞

斷詞

```
```{r}
# 對爬文結果斷詞
# ptt_tb[i]會出現第i篇文章的斷詞結果,文章被空白鍵斷開
library(jiebaR)
library(quanteda)

seg <- worker(symbol = T, bylines = F, user = "words.txt")

for (i in 1:length(dcard_tb$content)){
  dcard_tb$content[i] <-
  dcard_tb$content[i] %>% segment(seg) %>% paste(collapse = " ")
}
```
```

```
Loading required package: jiebaR
Package version: 1.5.2
Parallel computing: 2 of 4 threads used.
See https://quanteda.io for tutorials and examples.
```

```
Attaching package: 'quanteda'
```

```
The following object is masked from 'package:utils':
```

```
View
```



## 2. 文章內容斷詞

斷詞

### content

韓粉又爆盜用！「破億網紅蔡佩軒」影片整支被...

Youtuber「阿滴英文」日前號召 28 名包括千千、博...

此篇文章為轉貼文章，請更新至最新版本觀看完整...

[https://youtu.be/aZ5znkF\\_Umw](https://youtu.be/aZ5znkF_Umw) 阿滴英文提醒大家...

目前聽到的理由是因為討厭民進黨 滿好奇的想問問...

<https://i.imgur.com/qZr5iRK.jpg> 習近平也怕被民...

新聞來源：<https://news.ltn.com.tw/news/life/br...>

嗨，向大家分享這兩篇文 裡面談到很多我們平常...

這次是我第一次投票，但這總統素質令人堪憂，一個...

我真的不想看韓國瑜癸臉 4 年 不想看韓粉脫序行為...

我們前幾天有 po 過，但人數還差一點 QQ，希望大家...

<http://i.imgur.com/kVNLtM9.jpg> 空軍黑鷹直升機...

2020 大選迫在眉睫，國際政治金融專家汪浩認為，此...

我家是一個喜韓家庭 這幾天爸、媽一直問我要投給...

<http://i.imgur.com/lXoJTuf.jpg> 美國 CNN 報導，...

2020 大選倒數 6 天，1800 多位海內外學者連署公開支...

要要不是因為一塊「小英不修礦業法」的布條，幾...

新年才剛過，漢來海港餐廳也撐不住了！為了反映...

## 2. 文章內容斷詞

斷詞



建立語料庫

### content

韓粉又爆盜用！「破億網紅蔡佩軒」影片！

Youtuber「阿滴英文」日前號召 28 名包括千

此篇文章為轉貼文章，請更新至最新版本觀

[https://youtu.be/aZ5znkF\\_Umw](https://youtu.be/aZ5znkF_Umw) 阿滴英文

目前聽到的理由是因為討厭民進黨 滿好奇的

<https://i.imgur.com/qZr5iRK.jpg> 習近平也

新聞來源：<https://news.ltn.com.tw/news>

嗨，向大家分享這兩篇文 裡面談到很多我

這次是我第一次投票，但這總統素質令人堪

我真的不想看韓國瑜嘴臉 4 年 不想看韓粉脫

我們前幾天有 po 過，但人數還差一點 QQ，

<http://i.imgur.com/kVNLtM9.jpg> 空軍黑

2020 大選迫在眉睫，國際政治金融專家汪浩認

我家是一個喜韓家庭 這幾天爸、媽一直問我

<http://i.imgur.com/lXoJTuf.jpg> 美國 CNI

2020 大選倒數 6 天，1800 多位海內外學者連署

要要不是因為一塊「小英不修礦業法」的布

新年才剛過，漢來海港餐廳也撐不住了！為

韓國瑜

韓市長

蔡英文

蔡總統

柯文哲

柯P

蘇貞昌

吳敦義

張善政

黃國昌

賴清德

郭台銘

郭董

陳菊

馬英九

宋楚瑜

宋楚瑜

陳水扁

邱毅

謝龍介

李正皓

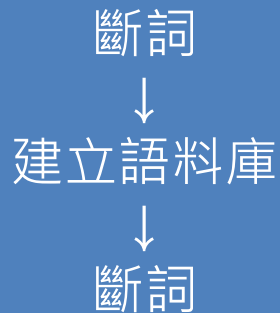
徐國勇

朱立倫

陳其邁

侯友宜

## 2. 文章內容斷詞



content

韓粉又爆盜用！「破億網紅蔡佩軒」影片！

Youtuber「阿滴英文」日前號召28名包括千千、博恩...

此篇文章為轉貼文章，請更新至最新版本觀看完整...

[https://youtu.be/aZ5znkF\\_Umw](https://youtu.be/aZ5znkF_Umw) 阿滴英文提醒大家...

目前聽到的理由是因為討厭民進黨 滿好奇的想問問...

<https://i.imgur.com/qZr5iRK.jpg> 習近平也怕被民...

新聞來源：<https://news.ltn.com.tw/news/life/br...>

嗨，向大家分享這兩篇文 裡面談到很多我們平常...

這次是我第一次投票，但這總統素質令人堪憂，一個...

我真的不想看韓國瑜嘴臉4年 不想看韓粉脫序行為...

我們前幾天有po過，但人數還差一點QQ，希望大家...

<http://i.imgur.com/kVNLtM9.jpg> 空軍黑鷹直升機...

2020大選迫在眉睫，國際政治金融專家汪浩認為，此...

我家是一個喜韓家庭 這幾天爸、媽一直問我要投給...

<http://i.imgur.com/lXoJTuf.jpg> 美國CNN報導，...

2020大選倒數6天，1800多位海內外學者連署公開支...

要要不是因為一塊「小英不修礦業法」的布條，幾乎...

新年才剛過，漢來海港餐廳也撐不住了！為了反映...

韓國瑜  
韓市長  
蔡英文  
蔡總統  
柯文哲  
柯P  
蘇貞昌  
吳敦義  
張善政  
黃國昌  
賴清德  
郭台銘  
郭董  
陳菊  
馬英九  
宋楚瑜  
宋楚瑜  
陳水扁  
邱毅  
謝龍介  
李正皓  
徐國勇  
朱立倫  
陳其邁

content

韓粉又爆盜用！「破億網紅蔡佩軒」影片 整支被搬...

Youtuber「阿滴英文」日前號召28名包括千千、博恩...

此篇文章為轉貼文章，請更新至最新版本觀看完整...

[https://youtu.be/aZ5znkF\\_Umw](https://youtu.be/aZ5znkF_Umw) 阿滴英文提醒大家...

目前聽到的理由是因為討厭民進黨 滿好奇的想問問...

<https://i.imgur.com/qZr5iRK.jpg> 習近平也怕被民...

新聞來源：<https://news.ltn.com.tw/news/life/br...>

嗨，向大家分享這兩篇文 裡面談到很多我們平常...

這次是我第一次投票，但這總統素質令人堪憂，一個...

我真的不想看韓國瑜嘴臉4年 不想看韓粉脫序行為...

我們前幾天有po過，但人數還差一點QQ，希望大家...

<http://i.imgur.com/kVNLtM9.jpg> 空軍黑鷹直升機...

2020大選迫在眉睫，國際政治金融專家汪浩認為，此...

我家是一個喜韓家庭 這幾天爸、媽一直問我要投給...

<http://i.imgur.com/lXoJTuf.jpg> 美國CNN報導，...

2020大選倒數6天，1800多位海內外學者連署公開支...

要要不是因為一塊「小英不修礦業法」的布條，幾乎...

新年才剛過，漢來海港餐廳也撐不住了！為了反映...

## 2. 文章內容斷詞

斷詞



建立語料庫



斷詞



修改語料庫



斷詞



.....

content

韓粉又爆盜用！「破億網紅蔡佩軒」影片！

Youtuber「阿滴英文」日前號召28名包括千千、博恩...

此篇文章為轉貼文章，請更新至最新版本觀看完整...

[https://youtu.be/aZ5znkF\\_Umw](https://youtu.be/aZ5znkF_Umw) 阿滴英文提醒大家...

目前聽到的理由是因為討厭民進黨 滿好奇的想問問...

<https://i.imgur.com/qZr5iRK.jpg> 習近平也怕被民...

新聞來源：<https://news.ltn.com.tw/news/life/br...>

嗨，向大家分享這兩篇文 裡面談到很多我們平常...

這次是我第一次投票，但這總統素質令人堪憂，一個...

我真的不想看韓國瑜膺臉4年 不想看韓粉脫序行為...

我們前幾天有po過，但人數還差一點QQ，希望大家...

<http://i.imgur.com/kVNLtM9.jpg> 空軍黑鷹直升機...

2020大選迫在眉睫，國際政治金融專家汪浩認為，此...

我家是一個喜韓家庭 這幾天爸、媽一直問我要投給...

<http://i.imgur.com/lXoJTuf.jpg> 美國CNN報導，...

2020大選倒數6天，1800多位海內外學者連署公開支...

要要不是因為一塊「小英不修礦業法」的布條，幾乎...

新年才剛過，漢來海港餐廳也撐不住了！為了反映...

韓國瑜  
韓市長  
蔡英文  
蔡總統  
柯文哲  
柯P  
蘇貞昌  
吳敦義  
張善政  
黃國昌  
賴清德  
郭台銘  
郭董  
陳菊  
馬英九  
宋楚瑜  
宋楚瑜  
陳水扁  
邱毅  
謝龍介  
李正皓  
徐國勇  
朱立倫  
陳其邁

content

韓粉又爆盜用！「破億網紅蔡佩軒」影片 整支被搬...

Youtuber「阿滴英文」日前號召28名包括千千、博恩...

此篇文章為轉貼文章，請更新至最新版本觀看完整...

[https://youtu.be/aZ5znkF\\_Umw](https://youtu.be/aZ5znkF_Umw) 阿滴英文提醒大家...

目前聽到的理由是因為討厭民進黨 滿好奇的想問問...

<https://i.imgur.com/qZr5iRK.jpg> 習近平也怕被民...

新聞來源：<https://news.ltn.com.tw/news/life/br...>

嗨，向大家分享這兩篇文 裡面談到很多我們平常...

這次是我第一次投票，但這總統素質令人堪憂，一個...

我真的不想看韓國瑜膺臉4年 不想看韓粉脫序行為...

我們前幾天有po過，但人數還差一點QQ，希望大家...

<http://i.imgur.com/kVNLtM9.jpg> 空軍黑鷹直升機...

2020大選迫在眉睫，國際政治金融專家汪浩認為，此...

我家是一個喜韓家庭 這幾天爸、媽一直問我要投給...

<http://i.imgur.com/lXoJTuf.jpg> 美國CNN報導，...

2020大選倒數6天，1800多位海內外學者連署公開支...

要要不是因為一塊「小英不修礦業法」的布條，幾乎...

新年才剛過，漢來海港餐廳也撐不住了！為了反映...

## 2. 文章內容斷詞

| 看板   | 八卦        |       | 女孩         |       | 美妝          |       | 台大       |       |
|------|-----------|-------|------------|-------|-------------|-------|----------|-------|
| 平台   | PTT       | Dcard | PTT        | Dcard | PTT         | Dcard | PTT      | Dcard |
| 時間   | 1/3 ~ 1/4 |       | 12/7 ~ 1/3 |       | 12/19 ~ 1/4 |       | 9/24~1/3 |       |
| 原始篇數 | 2201      | 186   | 3410       | 1000  | 211         | 1000  | 601      | 1000  |
| 抽樣篇數 | 186       | 186   | 500        | 500   | 211         | 211   | 500      | 500   |
| 斷詞個數 | 9862      | 9739  | 10423      | 13011 | 11629       | 20505 | 16157    | 7610  |

### 3. 建立DTM

```
```{r}
# 製作出dtm
ptt_dtm <- matrix(nrow = length(ptt_content), ncol = length(ptt_all))

for (i in 1:length(ptt_content)) {
  ptt_num <- vector("numeric",length(ptt_all))
  temp = ptt_toks[[i]]
  for (j in 1:length(ptt_all))
  {
    for(k in 1:length(temp))
    {
      if(temp[k] == ptt_all[j]){
        ptt_num[j] <- ptt_num[j] + 1
      }
    }
  }
  ptt_num <- ptt_num/sum(ptt_num)
  ptt_dtm[i,] <- ptt_num
}
```
```

### 3. 建立DTM

```
Console Terminal x Jobs x
~/
> length(dcard_all)
[1] 6432
> # 製作出dtm
> dcard_dtm <- matrix(nrow = length(dcard_content), ncol = length(dcard_all))
>
> for (i in 1:length(dcard_content)) {
+ dcard_num <- vector("numeric",length(dcard_all))
+ temp = dcard_toks[[i]]
+ for (j in 1:length(dcard_all))
+ {
+ for(k in 1:length(temp))
+ {
+ if(temp[k] == dcard_all[j]){
+ dcard_num[j] <- dcard_num[j] + 1
+ }
+ }
+ }
+ dcard_num <- dcard_num/sum(dcard_num)
+ dcard_dtm[i,] <- dcard_num
+ }
> dcard_dtm
 [,1] [,2] [,3] [,4] [,5] [,6] [,7]
 [,8] [,9] [,10] [,11] [,12] [,13] [,14]
 [,15] [,16] [,17] [,18] [,19] [,20] [,21]
 [,22] [,23] [,24] [,25] [,26] [,27]
 [,28] [,29] [,30] [,31] [,32] [,33] [,34]
 [,35] [,36] [,37] [,38] [,39] [,40] [,41]
 [,42] [,43] [,44] [,45] [,46] [,47] [,48]
 [,49] [,50] [,51] [,52] [,53] [,54] [,55]
 [,56] [,57] [,58] [,59] [,60] [,61]
 [,62] [,63] [,64] [,65] [,66] [,67]
```

### 3. 建立DTM

|       | T1 | T2 | T3 | T4 | T5 | ..... | T10000 ↑ |
|-------|----|----|----|----|----|-------|----------|
| D1    |    |    |    |    |    |       |          |
| D2    |    |    |    |    |    |       |          |
| D3    |    |    |    |    |    |       |          |
| ..... |    |    |    |    |    |       |          |
| D500  |    |    |    |    |    |       |          |



## 4. 計算Cosine Similarity

```
```{r}
# 任兩篇文章計算其Cosine Similarity
ptt_n <- length(ptt_content)
ptt_cosine_similarity <- vector("numeric", ptt_n*(ptt_n-1)/2 )
ptt_index <- 1

for(i in 1:(ptt_n-1)){
  for(j in (i+1):ptt_n){
    ptt_cosine_similarity[ptt_index] <- cos_sim(ptt_dtm[i,],ptt_dtm[j,])
    ptt_index <- ptt_index + 1
  }
}
```
```

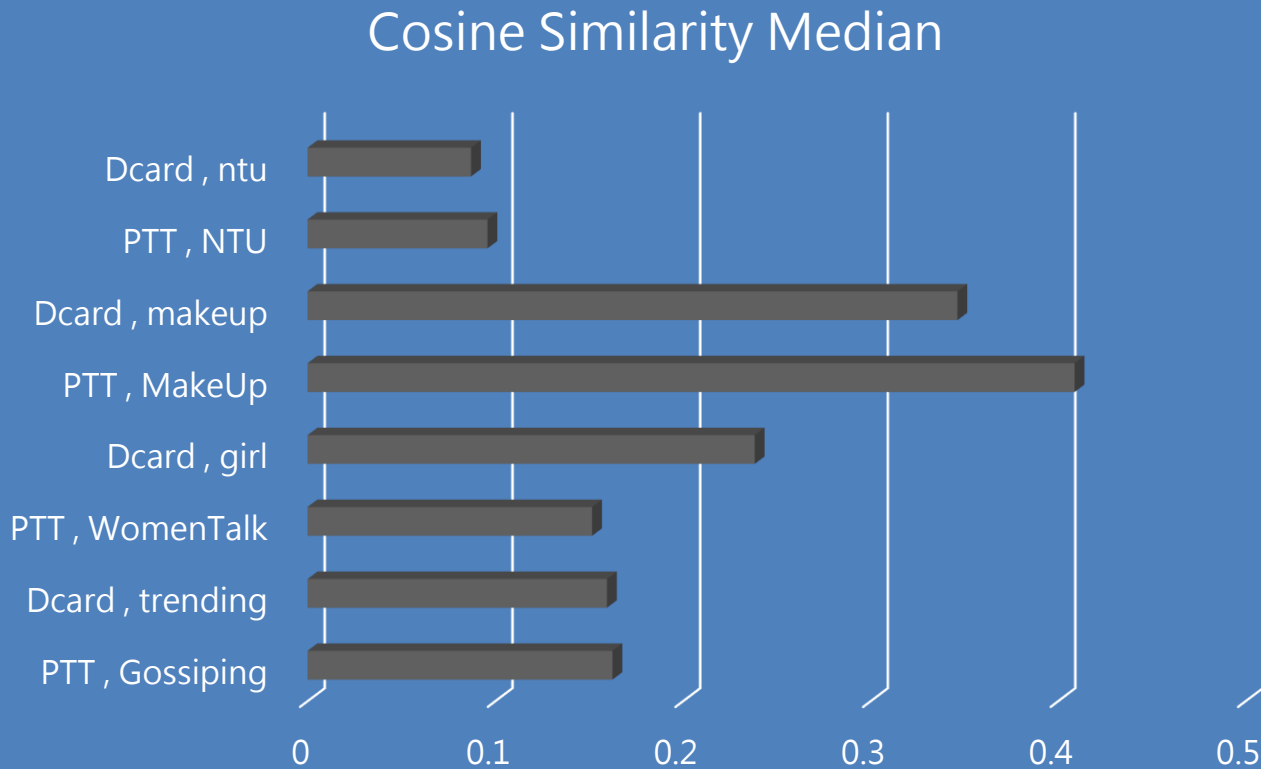
## 4. 計算Cosine Similarity

|       | D1 | D2   | D3   | D4   | ..... | D500 |
|-------|----|------|------|------|-------|------|
| D1    |    | 0.xx | 0.xx | 0.xx |       | 0.xx |
| D2    |    |      | 0.xx | 0.xx |       | 0.xx |
| D3    |    |      |      | 0.xx |       | 0.xx |
| ..... |    |      |      |      |       | 0.xx |
| D500  |    |      |      |      |       |      |

## 4. 計算Cosine Similarity

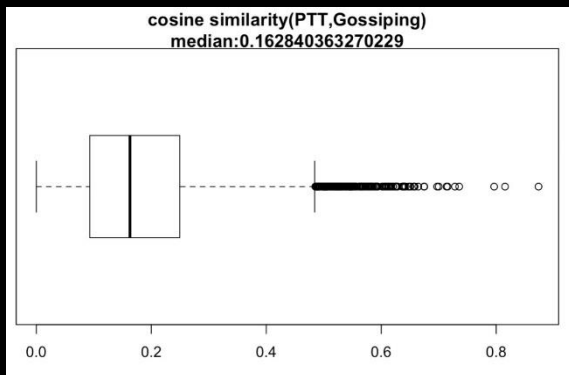
| 看板                       | 八卦        |        | 女孩         |        | 美妝          |        | 台大       |        |
|--------------------------|-----------|--------|------------|--------|-------------|--------|----------|--------|
| 平台                       | PTT       | Dcard  | PTT        | Dcard  | PTT         | Dcard  | PTT      | Dcard  |
| 時間                       | 1/3 ~ 1/4 |        | 12/7 ~ 1/3 |        | 12/19 ~ 1/4 |        | 9/24~1/3 |        |
| 原始篇數                     | 2201      | 186    | 3410       | 1000   | 211         | 1000   | 601      | 1000   |
| 抽樣篇數                     | 186       | 186    | 500        | 500    | 211         | 211    | 500      | 500    |
| 斷詞個數                     | 9862      | 9739   | 10423      | 13011  | 11629       | 20505  | 16157    | 7610   |
| Cosine Similarity<br>中位數 | 0.1628    | 0.1595 | 0.15157    | 0.2382 | 0.4088      | 0.3463 | 0.0959   | 0.0871 |

## 4. 計算Cosine Similarity



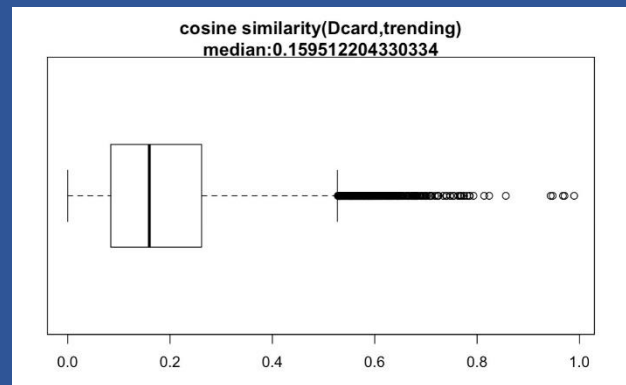
# 研究結果

# PTT 八卦版



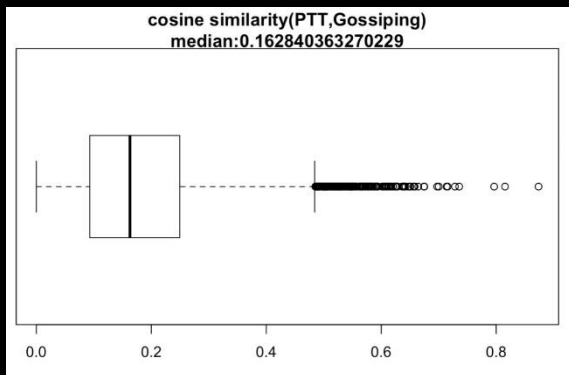
□ 中位數 : 0.1628

# Dcard時事版



● 中位數 : 0.1595

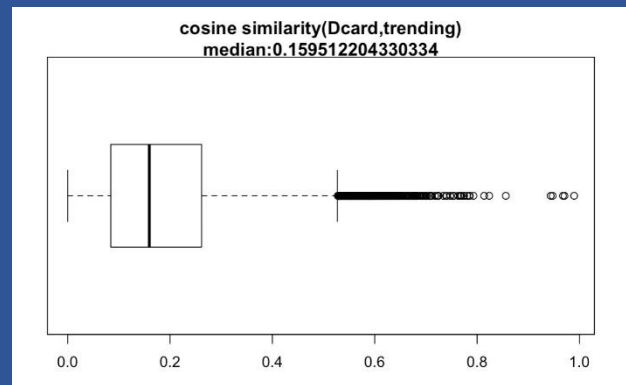
# PTT 八卦版



□ 平均數 : 0.1707

□ 標準差 : 0.1011

# Dcard時事版



● 平均數 : 0.1873

● 標準差 : 0.1377

# PTT 八卦版



# Dcard時事版



- ✓ 兩看板的Cosine Similarity 差不多
- ✓ 兩看板討論事物的內部一致性差不多
- ✓ 兩看板的同溫層差不多厚





# Dcard時事版

|       |      |           |       |       |      |      |         |       |         |       |
|-------|------|-----------|-------|-------|------|------|---------|-------|---------|-------|
| [1]   | "引述" | "a855341" | "崧"   | "樓下"  | "的"  | "腳印" | "之"     | "銘言"  | "媒體"    | "公視"  |
| [11]  | "新聞" | "黃"       | "千容"  | "王龍"  | "韜"  | "標題" | "南投"    | "縣長"  | "欲"     | "在"   |
| [21]  | "九九" | "峰建"      | "天空"  | "步道"  | "民團" | "應"  | "審慎"    | "評估"  | "內文"    | "縣長"  |
| [31]  | "林明" | "湊"       | "又"   | "有"   | "新"  | "點子" | "除了"    | "想"   | "在"     | "南投"  |
| [41]  | "九九" | "峰蓋"      | "王六"  | "王六"  | "王六" | "王六" | "王六"    | "王六"  | "王六"    | "王六"  |
| [51]  | "步道" | "上"       | [1]   | "媒體"  | "公視" | "新聞" | "黃"     | "千容"  | "王龍"    | "韜"   |
| [61]  | "已經" | "有"       | [11]  | "欲"   | "在"  | "九九" | "峰建"    | "天空"  | "步道"    | "民團"  |
| [71]  | "蓋"  | "天空"      | [21]  | "內文"  | "縣長" | "林明" | "湊"     | "又"   | "有"     | "新"   |
| [81]  | "步道" | "這是繼"     | [31]  | "在"   | "南投" | "九九" | "峰蓋"    | "天空"  | "步道"    | "而且"  |
| [91]  | "的"  | "新"       | [41]  | "搬到"  | "天空" | "步道" | "上"     | "不過"  | "南投"    | "天空"  |
| [101] | "峰"  | "森林"      | [51]  | "加總"  | "起來" | "已經" | "有"     | "座"   | "公民"    | "團體"  |
| [111] | "商店" | "和"       | [61]  | "地質"  | "不穩" | "蓋"  | "天空"    | "步道"  | "更要"    | "評估"  |
| [121] | "一步" | "解釋"      | [71]  | "蓋"   | "天空" | "步道" | "這是繼"   | "熱氣球" | "的"     | "點子"  |
| [131] | "處長" | "王源鍾說"    | [81]  | "林明"  | "湊"  | "的"  | "新"     | "想法"  | "想"     | "在"   |
| [141] | "的"  | "地質"      | [91]  | "公頃"  | "九九" | "峰"  | "森林"    | "園區"  | "內"     | "架設"  |
| [151] | "線"  | "的"       | [101] | "還要"  | "把"  | "商店" | "和"     | "藝品"  | "銷售"    | "都"   |
| [161] | "這個" | "步道"      | [111] | "處長"  | "則是" | "一步" | "解釋"    | "概念"  | "上"     | "比較"  |
| [171] | "的"  | "消費"      | [121] | "南投縣" | "觀光" | "處長" | "王源鍾說"  | "在"   | "九九"    | "峰"   |
| [181] | "廊道" | "https"   | [131] | "這麼"  | "特別" | "的"  | "地質"    | "這麼"  | "漂亮"    | "的"   |
| [191] | "有"  | "竹山"      | [141] | "南投縣" | "烏溪" | "線"  | "的"     | "風景"  | "所以"    | "我們"  |
| [201] | "等"  | "四座"      | [151] | "要"   | "把"  | "這個" | "步道"    | "把"   | "它"     | "高空"  |
| [211] | "吊橋" | "兩座"      | [161] | "把"   | "很多" | "的"  | "消費"    | "的"   | "概念"    | "把"   |
| [221] | "或是" | "熱潮"      | [171] | "高空"  | "的"  | "廊道" | "https" | "j"   | "imgur" | "com" |
|       |      |           | [181] | "南投"  | "已經" | "有"  | "竹山"    | "天梯"  | "清"     | "境"   |
|       |      |           | [191] | "寮"   | "龍鳳" | "等"  | "四座"    | "天空"  | "步道"    | "信義"  |
|       |      |           | [201] | "七彩"  | "等"  | "兩座" | "吊橋"    | "收益"  | "不錯"    | "但"   |
|       |      |           | [211] | "黯然"  | "歇業" | "或是" | "熱潮"    | "一過"  | "人潮"    | "不再"  |

# PTT 八卦版

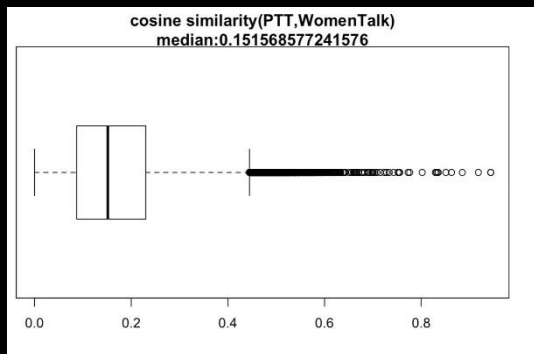
# Dcard時事版



|       |      |           |       |      |      |       |       |        |       |      |
|-------|------|-----------|-------|------|------|-------|-------|--------|-------|------|
| [1]   | "引述" | "a855341" | "崧"   | "樓下" | "的"  | "腳印"  | "之"   | "銘言"   | "媒體"  | "公視" |
| [11]  | "新聞" | "黃"       | "千容"  | "王龍" | "韜"  | "標題"  | "南投"  | "縣長"   | "欲"   | "在"  |
| [21]  | "九九" | "峰建"      | "天空"  | "步道" | "民團" | "應"   | "審慎"  | "評估"   | "內文"  | "縣長" |
| [31]  | "林明" | "湊"       | "又"   | "有"  | "新"  | "點子"  | "除了"  | "想"    | "在"   | "南投" |
| [41]  | "九九" | "峰蓋"      | "天空"  | "步道" | "而且" | "還要"  | "把"   | "商業活動" | "搬到"  | "天空" |
| [51]  | "步道" | "上"       | "不過"  | "南投" | "天空" | "步道"  | "及"   | "吊橋"   | "加總"  | "起來" |
| [61]  | "已經" | "有"       | "座"   | "公民" | "團體" | "指出"  | "九九"  | "峰"    | "地質"  | "不穩" |
| [71]  | "蓋"  | "天空"      | "步道"  | "更要" | "審慎" | "評估"  | "九九"  | "峰要"   | "蓋"   | "天空" |
| [81]  | "步道" | "這是繼"     | "熱氣球" | "的"  | "點子" | "後"   | "南投"  | "縣長"   | "林明"  | "湊"  |
| [91]  | "的"  | "新"       | "想法"  | "想"  | "在"  | "縣府"  | "管轄"  | "的"    | "公頃"  | "九九" |
| [101] | "峰"  | "森林"      | "園區"  | "內"  | "架設" | "空中"  | "步道"  | "而且"   | "還要"  | "把"  |
| [111] | "商店" | "和"       | "藝品"  | "銷售" | "都"  | "搬上去" | "南投縣" | "觀光"   | "處長"  | "則是" |
| [121] | "一步" | "解釋"      | "概念"  | "上"  | "比較" | "像是"  | "天空"  | "廊道"   | "南投縣" | "觀光" |
| [131] | "處長" | "王源鍾說"    |       |      |      |       |       |        |       |      |
| [141] | "的"  | "地質"      |       |      |      |       |       |        |       |      |

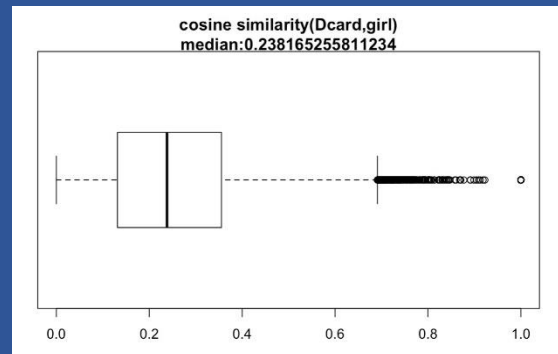
|       |       |      |      |        |       |      |      |       |       |        |
|-------|-------|------|------|--------|-------|------|------|-------|-------|--------|
| [1]   | "媒體"  | "公視" | "新聞" | "黃"    | "千容"  | "王龍" | "韜"  | "標題"  | "南投"  | "縣長"   |
| [11]  | "欲"   | "在"  | "九九" | "峰建"   | "天空"  | "步道" | "民團" | "應"   | "審慎"  | "評估"   |
| [21]  | "內文"  | "縣長" | "林明" | "湊"    | "又"   | "有"  | "新"  | "點子"  | "除了"  | "想"    |
| [31]  | "在"   | "南投" | "九九" | "峰蓋"   | "天空"  | "步道" | "而且" | "還要"  | "把"   | "商業活動" |
| [41]  | "搬到"  | "天空" | "步道" | "上"    | "不過"  | "空"  | "空"  | "步道"  | "及"   | "吊橋"   |
| [51]  | "加總"  | "起來" | "已經" | "有"    | "座"   | "公民" | "團體" | "指出"  | "九九"  | "峰"    |
| [61]  | "地質"  | "不穩" | "蓋"  | "天空"   | "步道"  | "更要" | "審慎" | "評估"  | "九九"  | "峰要"   |
| [71]  | "蓋"   | "天空" | "步道" | "這是繼"  | "熱氣球" | "的"  | "點子" | "後"   | "南投"  | "縣長"   |
| [81]  | "林明"  | "湊"  | "的"  | "新"    | "想法"  | "想"  | "在"  | "縣府"  | "管轄"  | "的"    |
| [91]  | "公頃"  | "九九" | "峰"  | "森林"   | "園區"  | "內"  | "空"  | "空中"  | "步道"  | "而且"   |
| [101] | "還要"  | "把"  | "商店" | "和"    | "藝品"  | "銷售" | "都"  | "搬上去" | "南投縣" | "觀光"   |
| [111] | "處長"  | "則是" | "一步" | "解釋"   | "概念"  | "上"  | "比較" | "像是"  | "天空"  | "廊道"   |
| [121] | "南投縣" | "觀光" | "處長" | "王源鍾說" | "在"   | "九九" | "峰"  | "你"   | "要"   | "欣賞"   |

# PTT 女孩版



□ 中位數 : 0.1516

# Dcard女孩版



● 中位數 : 0.2382

# PTT 女孩版



# Dcard女孩版



- ✓ Cosine Similarity : Dard > PTT 符合原假設
- ✓ Dcard女孩版討論事物的內部一致性較大
- ✓ Dcard女孩版的同溫層較厚

# PTT 女孩版



# Dcard女孩版



可能原因：

- ✓ Dcard女孩版多為大學女生之討論，使用者彼此的生活經驗較類似

# PTT 批踢踢實驗坊



- ❑ 成立於1995/09/14
- ❑ 註冊人數 150~200百萬
- ❑ 同時上線人數 10 萬
- ❑ 25-45歲 上班族居多
- ❑ 政治、時事 熱度最高

# Dcard



- 成立於2011/12/16
- 註冊人數 約100百萬
- 每日參訪人數 4.25 萬
- 18-25歲 大學生居多
- 感情、課業 討論為多

# PTT 女孩版



# Dcard女孩版



可能原因：

- ✓ Dcard女孩版多為大學女生之討論，使用者彼此的生活經驗較類似
- ✓ PTT女孩版的性質並不單純為女孩版.....

批踢踢實業坊 > 看板 WomenTalk

作者 sec5566 (sec)  
標題 [討論] 女孩們會專挑年紀比自己大的男生嗎？  
時間 Tue Jan 7 16:51:15 2020

# PTT 女孩版

多多少少會發現女生對於年紀是有要求的，  
普遍來說都希望男生年紀要比較大，  
次一級就是同年紀，  
只有小部分會考慮年紀小的，  
除非是帥小開鮮肉，  
為什麼女孩們大多還是都要年紀大的？  
那老屁孩呢？  
年紀小但臉老可以嗎？  
或反過來年紀大娃娃臉？

-----

Sent from JPTT on my Sony H4331.

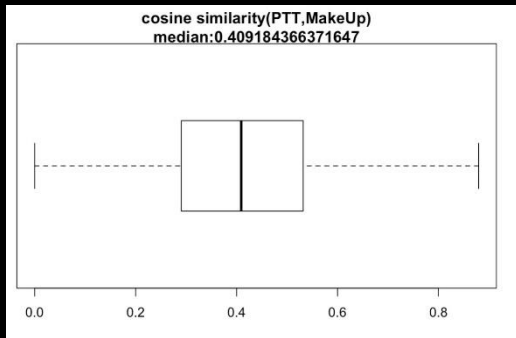
--

※ 發信站: 批踢踢實業坊([ptt.cc](http://ptt.cc)), 來自: 223.136.168.74 (臺灣)  
※ 文章網址: <https://www.ptt.cc/bbs/WomenTalk/M.1578387077.A.153.html>  
推 gaba000: 不是年紀大 是有錢懂嗎



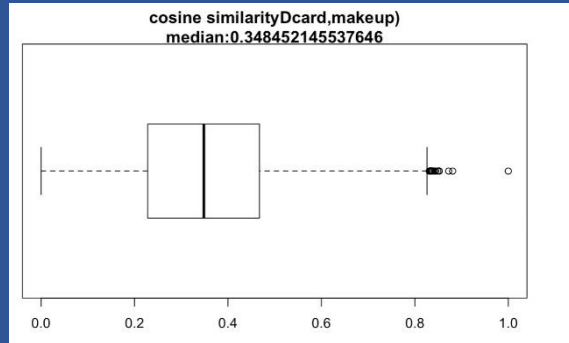


# PTT 美妝版



□ 中位數 : 0.4088

# Dcard美妝版



● 中位數 : 0.3463

# PTT 美妝版



# Dcard美妝版



- ✓ Cosine Similarity : PTT > Dcard , 不符合原假設
- ✓ PTT美妝版討論事物的內部一致性較大
- ✓ PTT女孩版的同溫層較厚

# PTT 美妝版



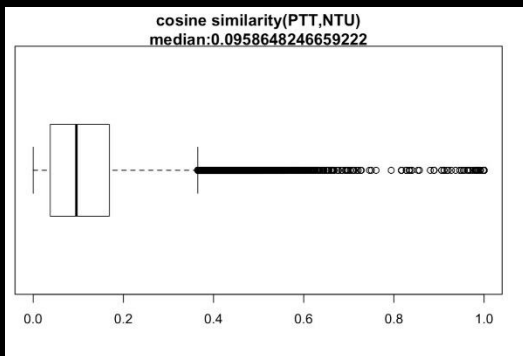
# Dcard女孩版



可能原因：

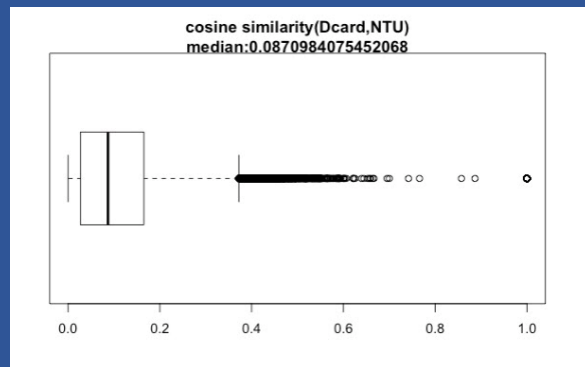
- ✓ Dcard美妝版的使用者運用大篇幅、細膩的文字  
像其他人推薦化妝品
- ✓ PTT美妝版的鄉民發問簡短，等待推文的回答

# PTT 台大版



□ 中位數 : 0.0959

# Dcard台大版



● 中位數 : 0.0871

# PTT 台大版

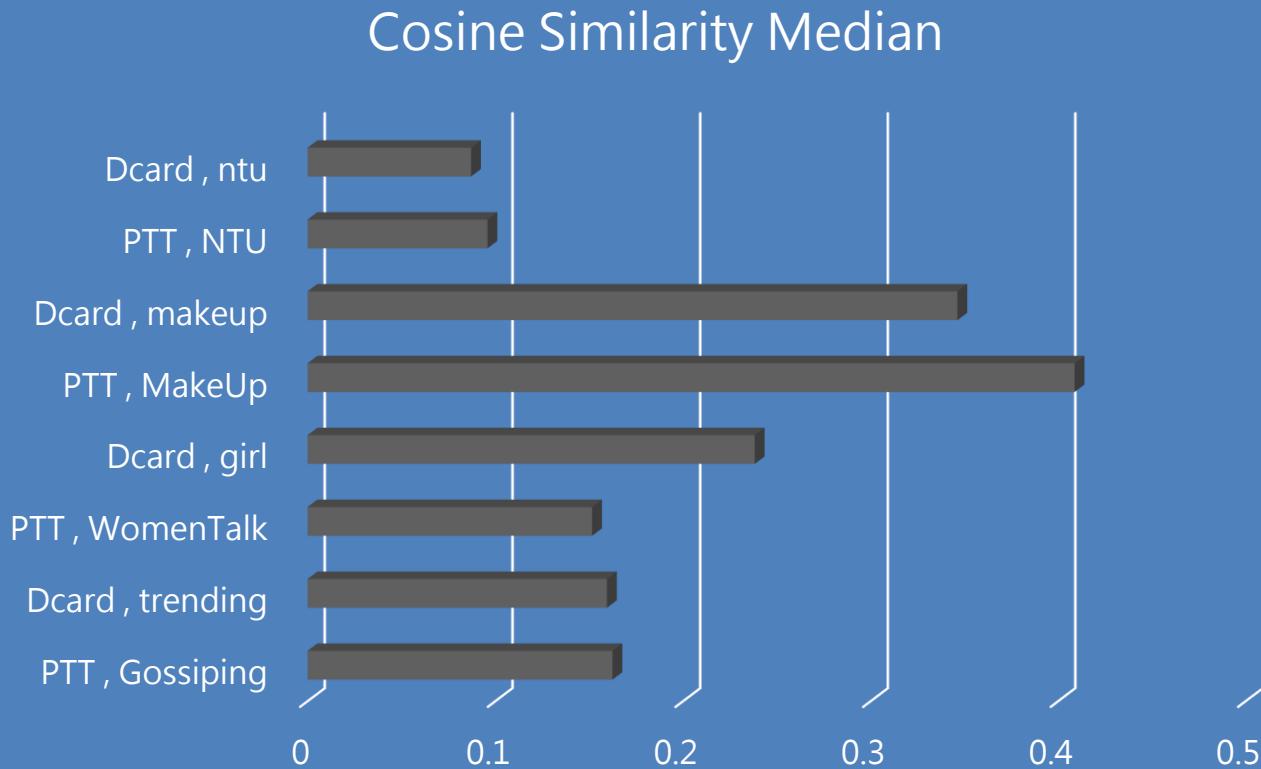


# Dcard台大版



- ✓ Cosine Similarity : PTT 與 Dcard 兩者差不多低
- ✓ 非常低....
- ✓ 不符合原假設

## 4. 計算Cosine Similarity



# PTT 台大版



# Dcard台大版



可能原因：

- ✓ 爬文時間截距長 ( 2019/09/24~2020/01/03 )
- ✓ 兩平台皆非台大學生的社群平台首選

# 研究限制



## 研究限制 + 未來改善方向

- ❑ 網址、英文、表情符號的使用
- ❑ PTT「引述」功能
- ❑ 錯字
- ❑ 功能詞（你我他、然後、總之...）
- ❑ 不符看板主題的文章
- ❑ 資料量過大、電腦運算能力不及
- ❑ Dcard API限制大量爬蟲

A solid blue geometric shape, resembling a large triangle or a parallelogram, occupies the left side of the frame. It is oriented diagonally, with its top-left corner at the top-left of the image and its bottom-right corner extending towards the bottom-right. The rest of the background is white.

**Thank you**