

From @JoeBiden to @POTUS

Text Analysis of Joe Biden's @POTUS Account

組名：G06 有至誠之目，則有善術

組員：簡郁展、潘立恆、邱繼賢、陳暘、劉慧新

目錄

1. 專案簡介.....	2
1.1 探討問題.....	2
1.2 問題假設.....	2
1.3 預期結果.....	2
2. 研究方法.....	2
2.1 資料取得：.....	2
2.2 程式碼運作說明：.....	2
3. 結果.....	3
3.1 時間資料分析.....	3
3.1.1 RetweetCount 以及 FavoriteCount 的時間線分析.....	3
3.1.2 @POTUS 一天的發文習慣	3
3.2 文字分析.....	4
3.2.1 詞頻圖.....	4
3.2.2 文字來源傾向分析.....	5
3.3 情緒分析.....	5
3.3.1 @POTUS 情緒使用總覽	6
3.3.2 各情緒類別的字詞組成.....	6
3.3.3 與@JoeBiden 的情緒字詞比較.....	7
3.4 結論.....	8
4. 討論.....	8
5. 組員分工.....	9

1. 專案簡介

1.1 探討問題

拜登正式就任美國第 46 任總統後，Twitter 也將總統官方帳號@POTUS 轉交拜登團隊。多數美國媒體聲稱過去@JoeBiden 的經營主要面向美國大眾推廣政策理念，相反的川普帳號只跟自己的支持者對話。而時下未見有拜登@POTUS 的推特分析，我們有興趣的是拜登@POTUS 在公共關係經營上還是走傳統、規律路線嗎？拜登@POTUS 是僅僅將 twitter 當作一個宣傳政令跟政府講話的媒體平台，還是存在明顯的情緒宣傳？

1.2 問題假設

本專案預設即使去掉停詞之後，取得的關鍵字對於推文意涵的辨識度還是有權重差異的。有些特別具象化單詞會能夠指涉與特定政策意涵或與特定政策事件相關，例如 vaccine/covid/jobs/economy/plan 等等。而其他較模糊的政治相關用語(例如 president/senate/congress 等)、一般單詞則被視為不具有明確政策資訊內涵。

1.3 預期結果

因此，本專案簡單假設如果@POTUS 只是單純宣傳政令或講話的帳號，應該大量含有上述辨識度高的具象化單字以指涉特定議題或事件。或者，在喜歡數或轉推數較高的推文中，要可以觀察到上述情形。同時，本專案也會比較@JoeBiden 在情緒分析是否與@POTUS 有明顯不同。

2. 研究方法

2.1 資料取得：

資料取得與資料前處理：從 Twitter API 與 twitterR()的 userTimeline()函數載入推文擷取資料，並轉換成 data frame，時間跨度涵蓋拜登@POTUS 帳號截止至 2021 年 6 月 13 日共 760 筆，另有從 Kaggle 抓取@JoeBiden 現成資料。之後將取得的 data frame 整理出我們需要的變數。

2.2 程式碼運作說明：

- 從 biden_tweets_df 中，我們進行發推來源的提取與時區調整，整理出 id, source, text, created, retweetCount, favoriteCount, hour 七個變數。
- 資料的分析主要就是對這些變數進行關係操弄，其中涉及時間資料、詞彙提取、直接對圖文進行分析，在方法上都需要對字串進行處理以及偵測。因此主要使用到的 packages 包含：tidytext/stringr/tidyr/lubridate 等。

- 我們用 `unnest_tokens()` 斷詞後並去掉停詞，得到 `tweet_words` 的 data frame，主要用於文字分析與情緒分析。
- 本專案希望先呈現@POTUS 在時間資料上的表現，了解其整體發文動態。接著進入推文內容進行文字與情緒分析。最終嘗試回答我們在專案簡介中的提問。

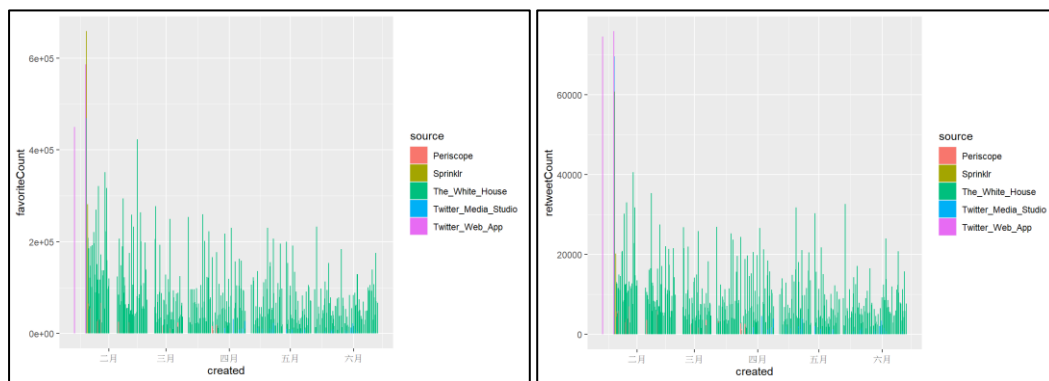
3. 結果

3.1 時間資料分析

3.1.1 RetweetCount 以及 FavoriteCount 的時間線分析

下面兩張圖可以明顯看出大部分推文都是從 The White House 發出來的，而 favoriteCount 的數量級大致上為 retweetCount 的十倍，兩者趨勢大體上相同。

- Twitter Web App 總共只發了四篇貼文，但其中的兩篇皆獲得了幾乎最高的 retweetCount 和 favoriteCount。
- Sprinklr 也只有 5 篇貼文，但 favoriteCount/retweetCount 前 10 名的推文中，就有 3 篇來自 Sprinklr。
- 但這些現象都只發生在一月拜登剛就職總統的時候。

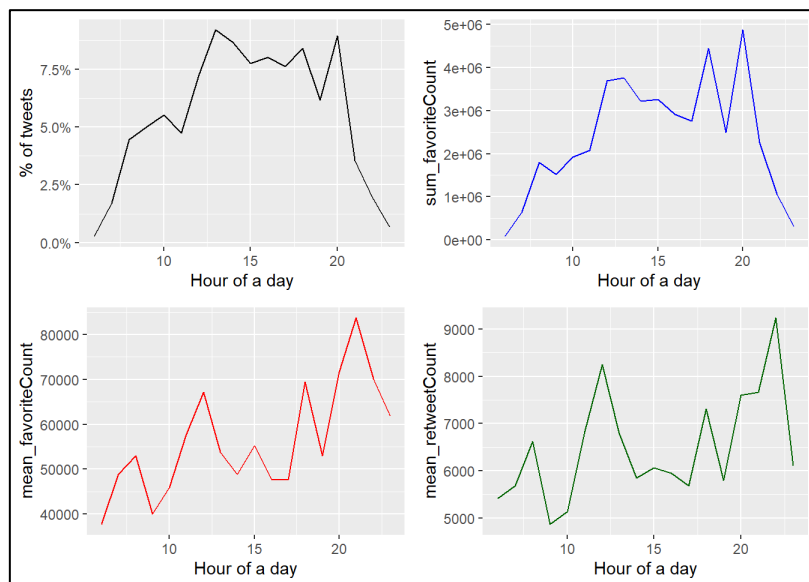


3.1.2 @POTUS 一天的發文習慣

將發文時間轉換成美東時間後，可以觀察到以下現象：

- @POTUS 發文在下午較頻繁，並在下午 1 點、晚上 8 點左右達到高峰，而且凌晨 1-5 點的時間段發文比例相當低，發文時間相當規律。
- 將每個時間段累積喜歡數加總後，發現與發文比例的趨勢相似，但發文數量多也可能造成喜歡數加總較多的現象。
- 若檢視每個時間段的喜歡數平均，在下午 12-1 點、下午 5-6 點、晚上 8-10 點都有明顯上升的趨勢，但與前兩張圖不同的是最高峰並非發生在晚上 8

點。除了晚上 8 點後發文比例比較少外，晚上 8 點後轉推數明顯較高也可能是提高觸及率與喜歡數的主要原因。此外，若內文包含重要議題或是受關注的社會事件，也可能獲得可觀的喜歡與轉推。

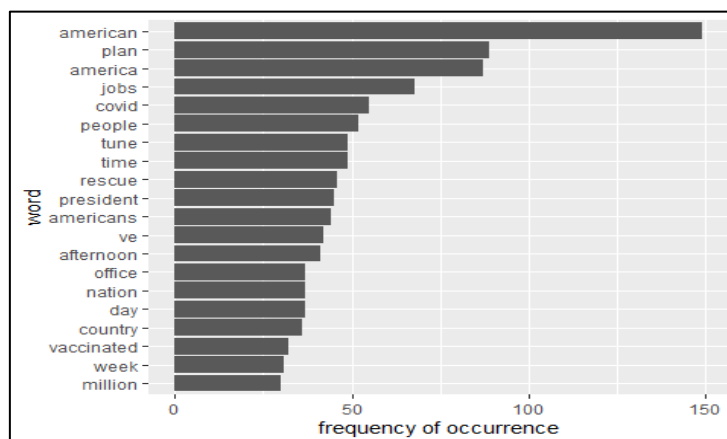


3.2 文字分析

在文字分析的環節，我們希望可以知道哪些單詞彙引起我們的興趣？包括 @POTUS 最常提到的詞彙，以及這些詞彙是否是 Problematic 的假設中提到在政策意涵上具有辨識度的字詞。我們希望進一步區分這些詞彙的來源傾向，藉此觀察推文管理情形。

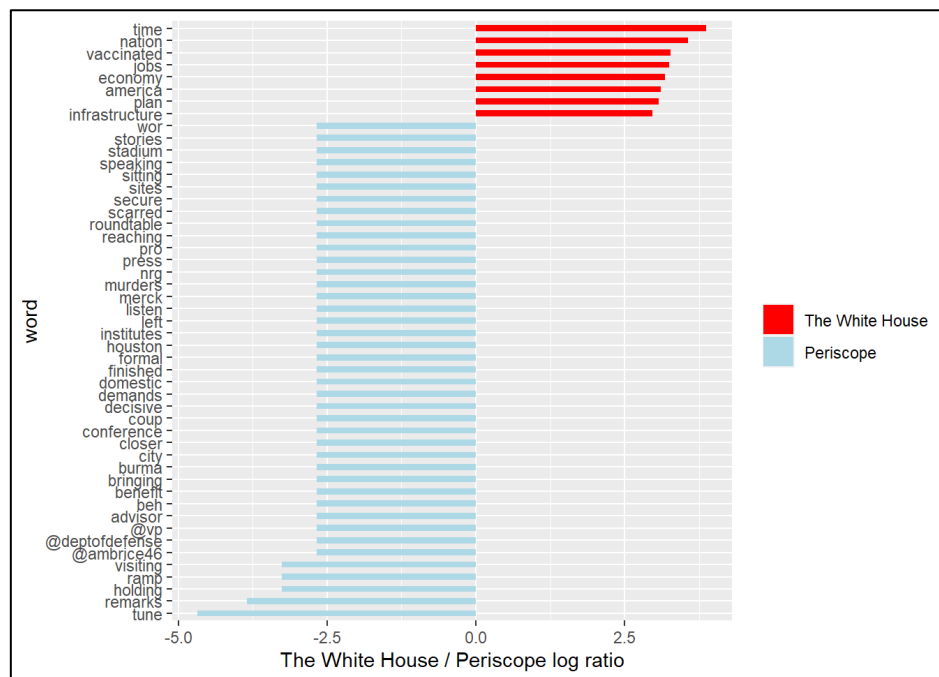
3.2.1 詞頻圖

可以發現@POTUS 整體而言政策辨識度高的詞彙包含 plan/jobs/covid/tune/rescue/vaccinated。推測與拜登的紓困法案《The American Jobs Plan》、《The American Rescue Plan Act of 2021》以及疫苗施打有關。



3.2.2 文字來源傾向分析

接著我們希望了解@POTUS 推文內容的來源傾向，我們更有興趣的東西是這些具有辨識度的字詞相對傾向被哪些發文來源使用，尤其是使用影音串流平台 (Periscope)以及一般發文管道 The White House 在文字使用傾向上區別。因此，可以透過計算勝算比(odds ratio)來知道相對於影音平台，The White House 更傾向具備那些文字特色，反之亦然。



可以發現@POTUS 對於推文內容具有系統性的管理：

- 從影音平台發布的推文多與舉辦會議或是記者會有關，包含 tune/remarks/holding/visiting；而且較 The White House 更傾向標註其他政府的 twitter 帳號，包含副總統(@vp)、國防部(@deptofdefense)、白宮國內政策委員會總監(@ambrice46)。
- The White House 則較傾向單純發布政策內容相關的字詞，例如 vaccinated/jobs/economy/plan/infrastructure，並且具有相較 Periscope 而言更高的文字辨識度(log ratio 的值比較有次序)，不會像 Periscope 一樣有好幾個單詞有一樣的勝算比。

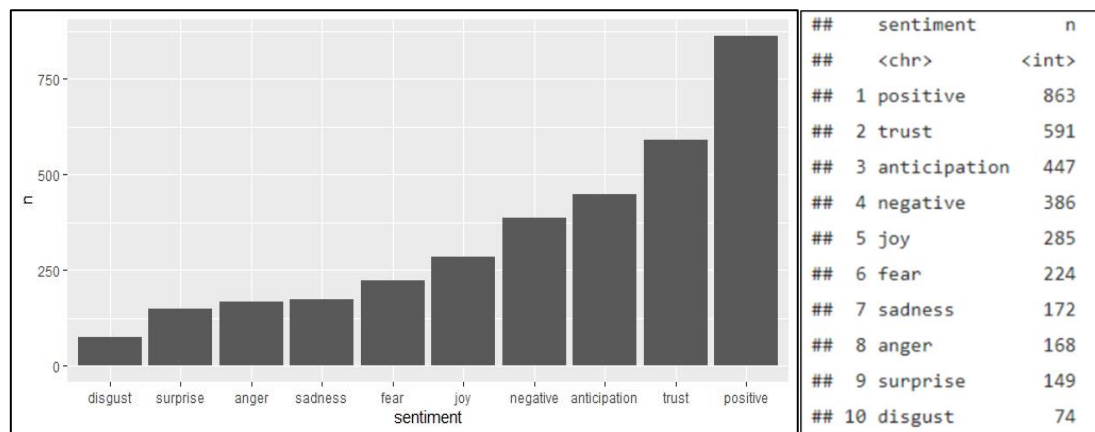
3.3 情緒分析

最後情緒分析的環節，我們想要知道@POTUS 的文字使用在情緒上表現，包含較常傳達哪些情緒、以及不同類別的情緒傳達分別由哪些字詞組成。對

此，我們採用 `textdata` package 中的 `get_sentiments()` 函數獲取 NRC Word-Emotion Association Lexicon，將 POTUS 使用的英文單詞分成十種情緒類別，比較@POTUS 跟@JoeBiden 的差異。

3.3.1 @POTUS 情緒使用總覽

@POTUS 整體較常使用正面的情緒字眼，包含 `positive/trust`，以及中性的 `anticipation`。

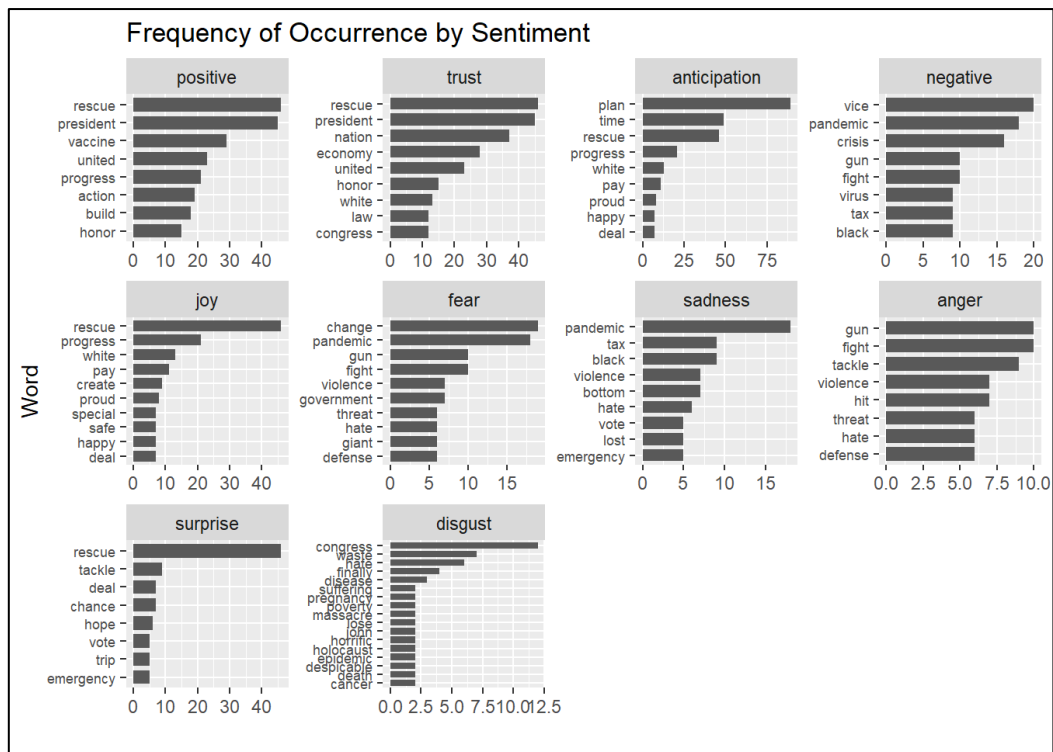


3.3.2 各情緒類別的字詞組成

看完情緒使用次數與發文來源的情緒傾向之後，我們想要知道到底是哪些字詞組成這些情緒的表達。綜合各情緒類別的字詞組成，可以發現出現頻率較高且具備政策意涵的詞彙有：

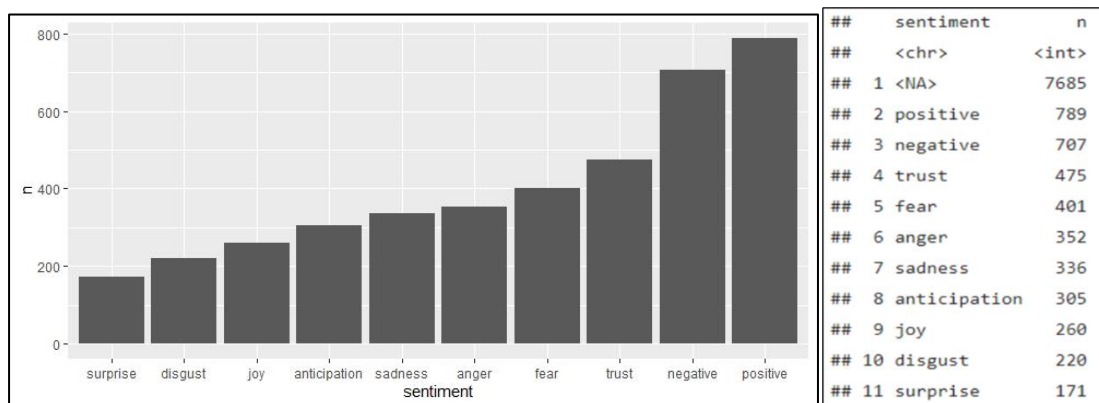
- `positive/trust`: `rescue/vaccine/economy`
- `anticipation`: `plan`
- `negative/fear/sadness/anger`: `pandemic/tax`

其他大部分的詞彙則無法分辨是否具備特別意義。



3.3.3 與@JoeBiden 的情緒字詞比較

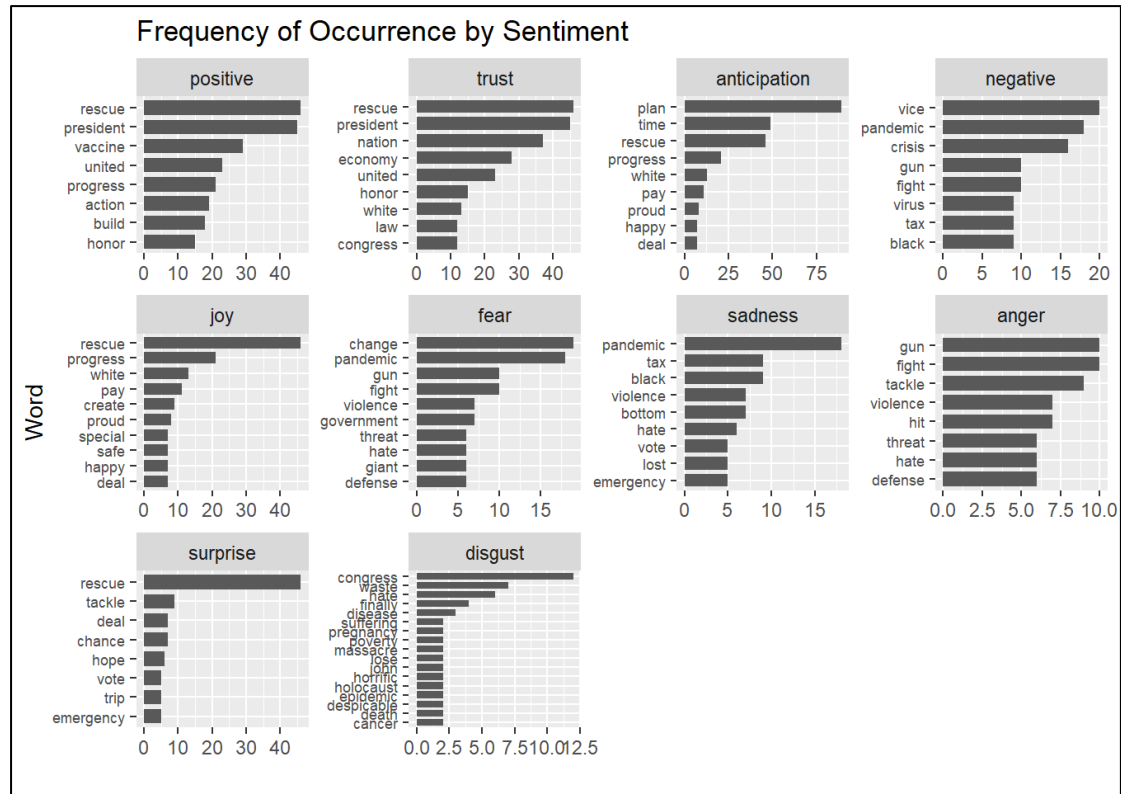
@JoeBiden 截至 2020 年 11 月使用 positive 與 negative 類別詞彙的傾向不相上下，而且兩者的使用量皆遠多過其他情緒類別。trust/fear/anger/sadness 等情緒類別則差距不大。這代表@JoeBiden 在情緒傳達上的複雜程度遠大於@POTUS。@POTUS 只傾向傳達正面情緒，但@JoeBiden 都會使用。



@JoeBiden 在各情緒類別的字詞組成上，較為顯著的有：

- positive: president
- negative/fear/anger: crisis/violence/fight/gun
- surprise: trump

- other/NA: vote/donald



3.4 結論

即使資料擷取時間有涵蓋總統選舉期間，@JoeBiden 除了選舉詞彙與人物詞彙之外(例如 president/vote/trump)還是多使用模糊字眼。即使我們加入不會被情緒詞表算進去的 NA 值去看，還是觀察不到政策辨識度比較高的單詞。可以推論相較@POTUS，@JoeBiden 更加的情緒化，而不傾向呈現具體政策意涵或社會事件，而@POTUS 單純就是一個政策導向的帳號。

4. 討論

本專案主旨在於釐清拜登總統帳號與個人帳號之間的不同。口頭報告結束後經過組內討論，我們認為即使是政策領域，分析資訊傳播者與接收者的行為模式相當重要。這涉及對於政策資訊擴散的研究，也就是政府與民眾的資訊落差是如何被填補的過程。其中，訊息接收者的解讀體驗會是相當重要的研究環節。就此，本專案定位算是對資訊傳播者行為研究的初見。

本專案的初步貢獻在於提升質化觀察特定網路社群的效率。在可能的延伸方向上，情緒傳播與其對應的政治、社會脈絡會如何影響訊息接收者的體驗，以及這些對於政策、新聞等資訊的體驗與解讀可以如何被設計，是長遠發展上我們有興趣的。

5. 組員分工

簡郁展 (投入程度 6): 題目發想、資料蒐集、@POTUS 帳號觀察、文字分析、@JoeBiden 情緒分析、統整製作 Rmarkdown、製作 PPT、口頭報告、書面報告
邱繼賢 (投入程度 4): Timeline 分析及繪圖、每日發文比例分析繪圖、發文照片/連結分析繪圖、各情緒用字繪圖、文獻閱讀
潘立恆 (投入程度 4): 喜歡數與轉推數貼文分析、Twitter 資料搜集、貼文情緒分析、文獻閱讀、Github site 建置
陳暘 (投入程度 3): 關鍵字搜尋、Tweet 資料閱讀、協助 Github site 建置
劉慧新 (投入程度 2): @JoeBiden 帳號觀察、轉推數跟喜歡數整理觀察