

一. 摘要

近年來社群媒體越來越興盛，無形之中改變了大眾接收資訊的習觀。多數公眾人物都會創建社群媒體帳號，宣傳自身想法和立場。近年來 Instagram 普獲年輕族群歡迎。公眾人物也紛紛在 Instagram 經營帳號。

除了政治人物，關心政治的每個人都有其政治傾向偏好。雖然現已有由諸如 Facebook、Twitter 或 PTT 上政治傾向的研究，但卻沒有 Instagram 上公眾人物政治傾向的研究，也沒有個不同政黨傾向在 Instagram 上展現的風格特色的比較。因此，本計畫將根據 Instagram 上公眾人物所形成的社群網絡，研究 Instagram 上公眾人物的政治光譜探勘分析技術。並進一步析各個群體間的重要人物、緊密程度、風格文化差異。

二. 研究動機與研究問題

隨著網際網路發展成熟，社群媒體也跟著隨之興起，從早期的 BBS 論壇、部落格到後來的 Facebook、Twitter、Instagram。其中因為能即時發布相片、輕易編輯照片和操作便利，Instagram 在年輕族群中成長極快，其在全球的活躍使用者數量甚至超過十億，而 18~34 歲用戶便占了 65%。

Instagram 簡稱 IG，讓使用者可以分享其貼文（文字、照片）並讓使用者的粉絲可以看到，而使用者 A 也可以追蹤（Following）其他使用者 B，而成為 B 的粉絲（follower），進而能在第一時間看到 B 的貼文。

隨著 Instagram 的熱門，不少的公眾人物（如藝人、網紅、政治人物），創辦 IG 帳號，透過自身魅力和特有風格，吸引更多年輕人的注意，並將自身的想法傳達出去。

一般而言，政治人物往往有其鮮明的政治立場。非政治領域的公眾人物，例如藝人雞排妹、閃靈樂團公認支持綠營的立場。近年來，眾所皆知方文山、歐陽娜娜往中國發展，毫不掩飾其政治立場。但多數公眾人物，為了避免爭議或市場考量，往往隱晦其政治傾向，避免踩到敏感的政治紅線。例如五月天多次無奈地捲入政治表態事件。又如近年崛起的博恩夜夜秀，其政治傾向也往往是大眾八卦的議題。儘管如此，關心政治的每個人都有其政治傾向與偏好。即使是政治人物，其政治立場也並非絕對的藍綠白，而是有親疏遠近的政治光譜。例如綠營的高嘉瑜立委就與民眾黨有較親近的關係。

本研究因此將針對公眾人物在 IG 上的追蹤所形成的社群網絡和追蹤中種人物的粉絲，探勘公眾人物的政治光譜。此光譜不僅能顯現非政治人物的政治立場，還能表示政治人物立場的親疏遠近，如大眾所說的深藍、淺藍、深綠、淺綠。本計畫也將進一步探勘分析不同立場光譜在 IG 上的權威人物、緊密連結程度及文章相片所展現的風格。我們將開發爬蟲程式擷取 IG 公眾人物資料，並結合社群網絡技術、矩陣降維演算法、中心性分析、風格探勘技術。

三. 文獻回顧與探討

Carlos H.G. Ferreira 等學者有做 Instagram 分群的相關研究 [7]，他們以文章內容和留言作為分析目標，進而探勘政治人物及名人立場。其中他們還對留言、文章內容做特質、情感分析。

近年來，Facebook 的研究透過大量使用者對政治人物的粉絲專頁按讚行為，結合 Singular Value Decomposition，找出政治傾向的光譜 [5, 6]。系上學長則是以使用者的推噓文及留言行為，在批踢踢實業坊（PTT）做政治立場判斷，並結合 Bipartite Graph，找出可疑帳號 [8]。

現有針對 Instagram 的研究，多為商管學院針對廣告、購買行為做出探討，或是對於圖像色彩的相關研究，至於政治立場光譜則沒有相關研究。

因此，我們將針對追蹤關係與共同粉絲特性，結合 SVM、Community Detection 技術，繪製出政治立場光譜，並進一步分析光譜中群體的重要人物與風格文化。

四. 研究方法及步驟

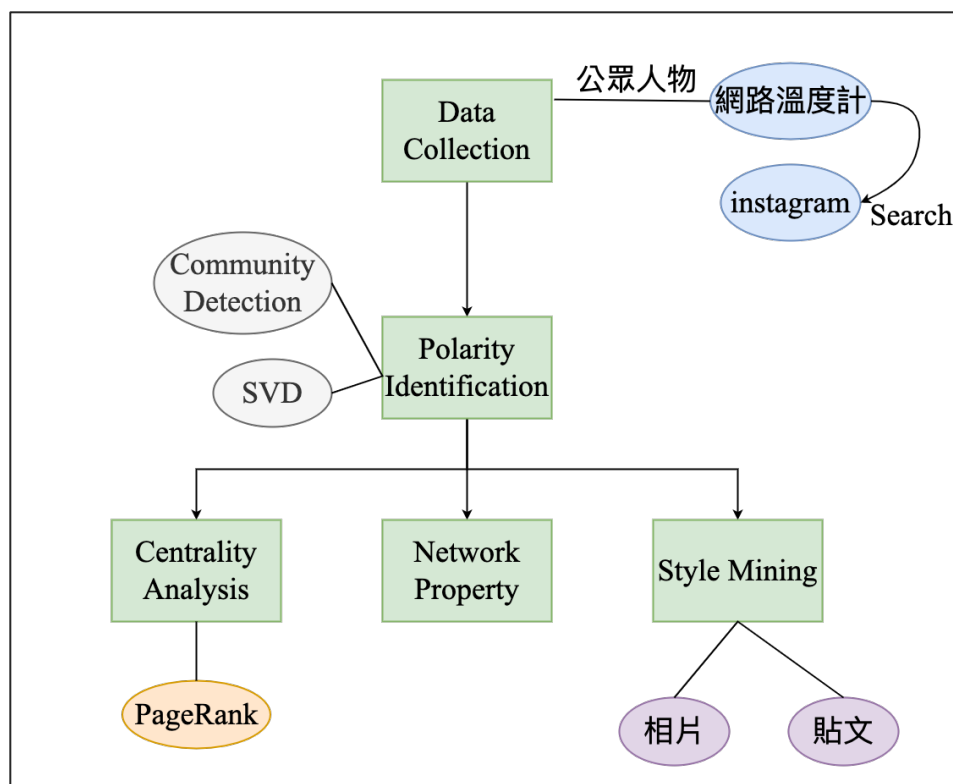


圖 1:系統架構圖

1. 研究架構

我們的研究架構如圖 1。首先需要公眾人物名單，其中包含政治人物、演員、歌手，作為主要研究對象，再來至 Instagram 上將其相關資料依照個人 ID 分類建儲存。接著，對這些公眾人物的政治立場光譜進行運算及分群，其中包含兩種方法：Community Detection on Following Graph 和 SVD on Follower Matrix。之後，系統會對分群完後的各個 Community 進行 Centrality Analysis，找出在 IG 上重要程度較高

的人物、分析 Community 內的關係緊密程度、探勘 Community 發文風格文化，分析不同政治立場的 Community 其風格差異。

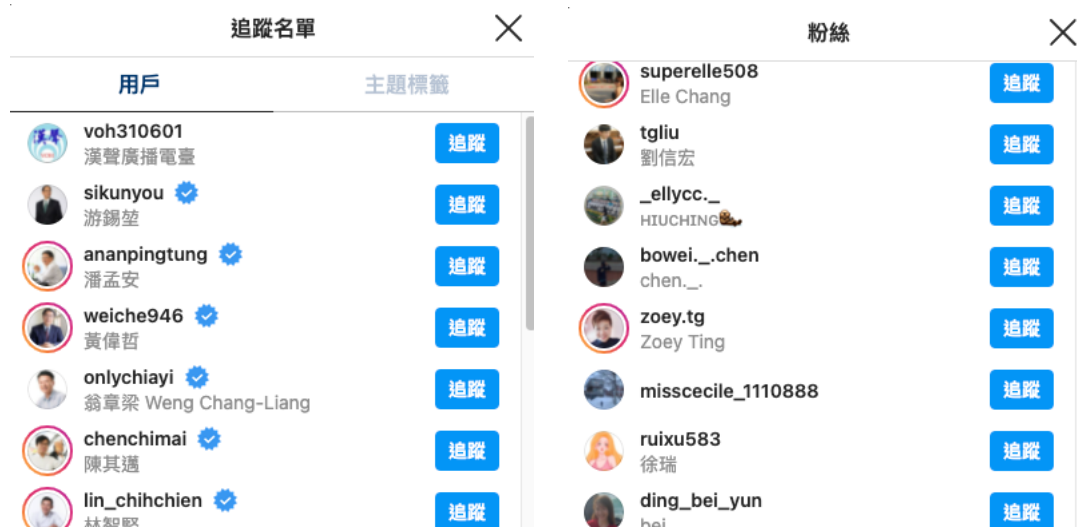


圖 3：蔡英文的追蹤名單（左）及粉絲（右）¹



圖 3：柯文哲貼文照片（左）和文章（右）²

2. Data Collection

透過 Crawler 爬蟲程式，可以在網路上查找大量資料。而 Python 這個程式語言有相當多便利的套件可以使用，如 Selenium、Requests、BeautifulSoup4 庫。我將用 Python 寫爬蟲程式來抓以藍綠白為主的重要官職的人物（如：蔡英文、朱立倫、柯文哲）和目前有名的公眾人物（歌手、Youtuber）。而資料來源主要來自「[網路溫度計](#)」，這個網站會分析目前時事熱門的人事物，因此我將在此網站爬取政治人物、政壇名嘴、歌手、Youtuber 四個類別共 450 人，作為此次研究目標。

在獲得公眾人物名單後，我將到 IG 查找其個人的帳號，其中該帳號超過一定粉絲量或有 IG 官方認證才會被採用，並利用爬蟲程式將其追蹤名單、粉絲（圖 2）、

¹ https://www.instagram.com/tsai_ingwen/

² <https://www.instagram.com/p/CZa61byB0W7/>

貼文內容與照片（圖 3）分別抓下來。

3. Polarity Identification

我們計畫研究的立場判斷包括兩種方法。一種是透過公眾人物的追蹤情形。另一種是根據公眾人物的粉絲。

3.1 Community Detection on Following Graph

為了探討公眾人物立場光譜的分析，本實驗第一個研究的方向是公眾人物彼此間的追蹤關係。根據一人物 a 的追蹤名單，若是他追蹤了一些 P 政黨的政治人物，那我們便可以推斷此人的政治立場是靠近 P 政黨的。以圖 5 為例，由左至右分別是蔡英文、柯文哲、蔣萬安的部分追蹤名單，其中不難發現蔡英文和蔣萬安都是追蹤各自陣營的人，而柯文哲大部分都是追蹤公家機關。因此我們將透過下列兩步驟來分析公眾人物彼此間的追蹤關係，並分群其政治立場。

(1) Following Graph

在資料收集完，我們有了公眾人物的追蹤關係後，可以將公眾人物帳號視為一個節點（Node），追蹤視為一個有方向的邊（Edge），使追蹤名單的關係轉換為一張有方向性的圖（Directed Graph）（圖 5）。

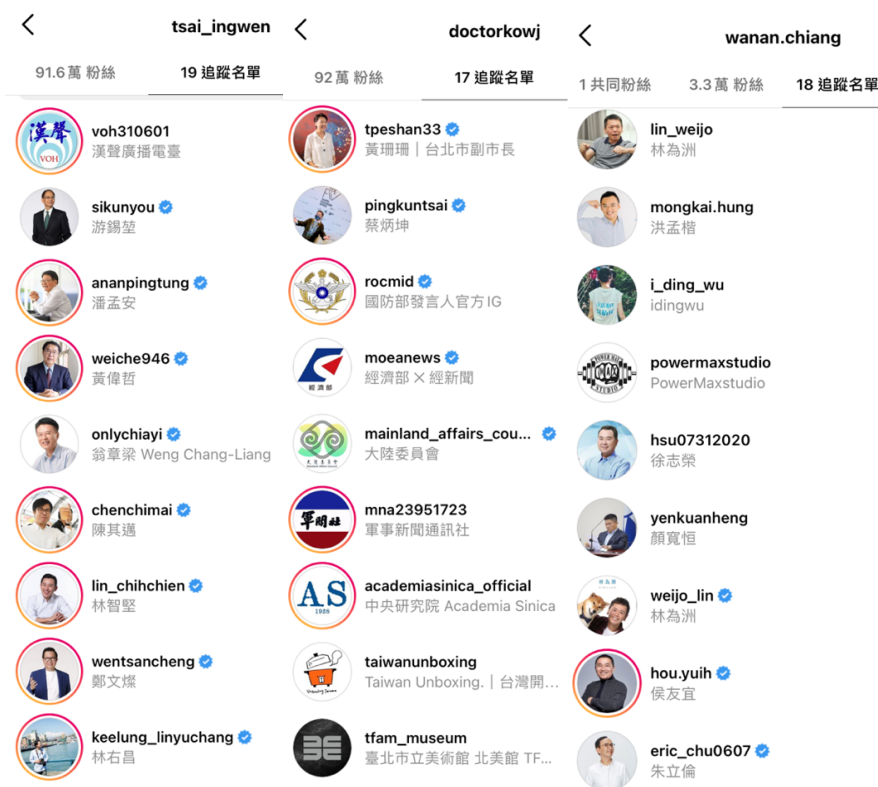


圖 4

(2) Community Detection

Community Detection 的定義為：給定一個 Graph，希望把所有的 Nodes 做分群，使得同一群內的 Nodes 互動較密切，不同群之間的 Nodes 的互動較不密切。Community Detection 包含多個演算法，其中包含 Speaker-listener Label Propagation

Algorithm (SLPA)、Hop Attenuation & Node Preference (HANP)、Louvain Algorithm。考慮到我的追蹤關係圖是 Directed Graph，因此我將使用 Louvain Algorithm。

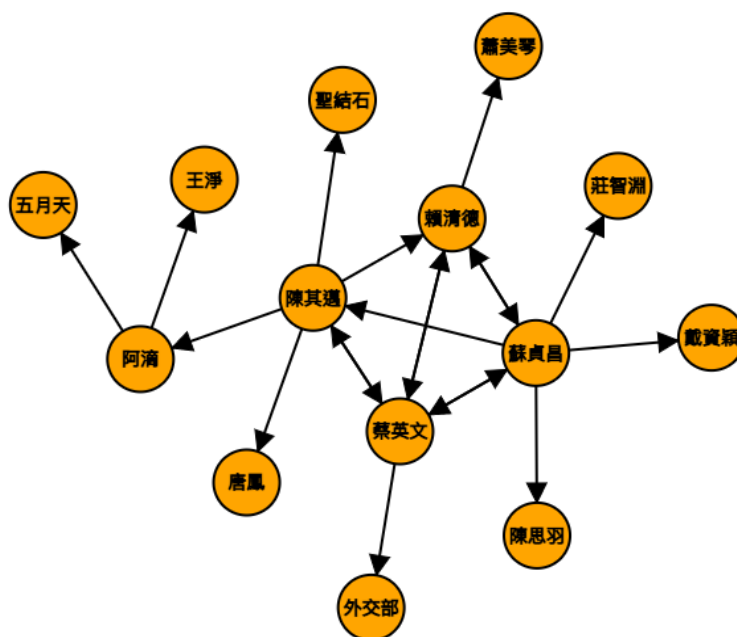


圖 5：Following Graph 示意圖

Louvain Algorithm 是一個針對優化 modularity 最大值，不斷進行迭代的演算法，其中針對 Directed Graph 的 modularity 的定義如圖 4。而迭代方式主要分成下列兩個步驟：

$$Q_d = \frac{1}{m} \sum_{ij} \left[A_{ij} - \frac{k_i^{in} k_j^{out}}{m} \right] \delta(c_i, c_j),$$

i. Optimize Local Modularity

將所有的節點視為一個 community，並針對每一個節點 i，計算出與其相連節點 j 之間的 modularity gain (ΔQ) (圖 5)，進而將 i 與提供最高 ΔQ 的節點 j 視為同一個 community。若 ΔQ 為負值，則 i 不做移動。

接著遍歷所有的節點，對其找到一個最合適的 community，並合併進去，直到所有的節點都無法再移動。而合併順序通常採用貪婪演算法 (Greedy Algorithm)。

$$\Delta Q = \left[\frac{\Sigma_{in} + 2k_{i,in}}{2m} - \left(\frac{\Sigma_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\Sigma_{in}}{2m} - \left(\frac{\Sigma_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right]$$

ii. Restructure Community as New Node

將第一步驟得到的各個 community 視為一個新的節點，並重新建構子圖，接著再重複第一步驟，重複執行。兩個新節點之間權值為原兩社群之間各邊的

權值總和。概念圖如圖 6。

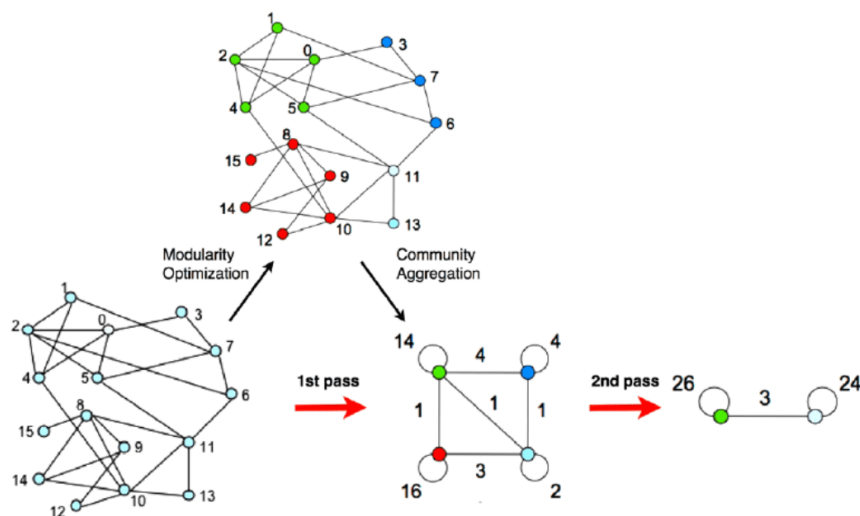


圖 6 [2]

此外，Louvain Algorithm 是假設一個人的政治立場為獨立，不會重疊 (overlapping) 情況下，所使用的演算法。而考慮到會有 overlapping 的狀況下，我將另外再參考 BigCLAM 演算法 (Cluster Affiliation Model for Big Networks)，其透過非負矩陣分解 (Non-negative Matrix Factorization) 來解決 overlapping 的問題。

3.2 SVD on Follower Proximity Matrix

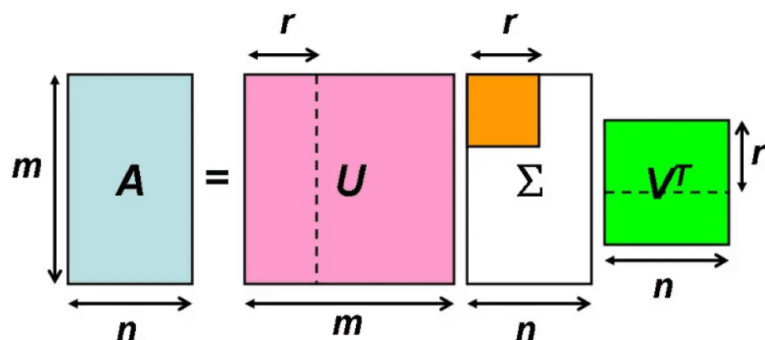
上述 2.1 方法是利用公眾人物彼此間的追蹤關係，做出立場的分群。接著我將考慮公眾人物彼此間的共同粉絲對政治立場傾向的因素。在 IG 中，假設有一人物 a，則 a 的粉絲代表一定程度上喜好 a 的立場、風格，才會選擇追蹤 a。那麼若一政治人物 b，與 a 擁有越多的共同粉絲，則我們可以猜想 b 的政治立場與 a 越相近。因此我將用公眾人物間的共同粉絲數，作為第二種測量公眾人物的政治立場光譜的方法。而做法我拆成下列兩個步驟：

(1) Follower Proximity Matrix

根據先前爬蟲抓下來的公眾人物及其粉絲，我們可以計算出公眾人物彼此之間共同粉絲數。接著，建一個 Proximity Matrix M ，其中 M 的元素 m_{ij} 代表 i 人物與 j 人物彼此間的共同粉絲數。由於假設 a 人物的粉絲數本身就很多，與其他人的共同粉絲數自然就會較多，考量到這點，我會對 M 矩陣中的每個元素 m_{ij} 除以該元素對應到的列人物的粉絲數 ($\text{diag}(m_i)$)，得到新的矩陣 A ，其中 A 元素 $a_{ij} = m_{ij} / \text{diag}(m_i)$ 。

(2) SVD (Singular Value Decomposition)

SVD 被譽為矩陣分解的「瑞士刀」和「勞斯萊斯」，前者說明其用途非常廣泛，後者則表達它是值得珍藏的精品。在此，我將使用 SVD 降維的特性，萃取出前兩個最重要維度，作為公眾人物政治立場的二維座標。SVD 示意圖 (其中 r 在本實驗中為 2)：



3

將上述的 A 矩陣做 SVD 分解可以得到： $A = U \cdot \Sigma \cdot V$

其中 U 是 Left Singular Matrix， $U(i,1)$, $U(i,2)$ 便是就是人物 i 的二維座標，也是人物 i 的政治立場二維光譜座標。

有了所有人物的座標後，便可以套入分群演算法，例如非監督式學習的 k-means 演算法，求得政治立場群組。

4. Centrality Analysis

在計算完立場分群後，我們想更進一步分析在這些 community 經營較成功的主要人物是否符合大眾的一般認知，例如民進黨的核心人物應為蔡英文、賴清德、蘇貞昌，而國民黨的核心人物為朱立倫、江啟臣等。

在 IG 上，一個使用者 a 的重要程度將大部分來自於其粉絲數的多寡，因為這會直接反應其文章的曝光程度，進而決定其影響力。而針對其粉絲也要考量到如果該粉絲追蹤越多人，那麼 a 對這名粉絲的影響力也就會越小。

PageRank 演算法會考慮到上述考量點，因此我將使用 PageRank 作為這次研究的使用。PageRank 目的是測量元素在集合中的相對重要性，而我將用來計算並預測在政治立場分析完的 community 中重要人物或經營較成功的公眾人物。PageRank 會給節點 i 一個權重，稱為 i 的 PageRank，並表示為 $PG(i)$ 。

根據 PageRank 定義並套用在本實驗上，假設有 a, b, c, d 四個節點，其中 b, c, d 皆有追蹤 a，而 $PR(a)$ 的算法如下：（其中 $L(j)$ 為 j 的總共追蹤人數）

$$PR(a) = \frac{PR(b)}{L(b)} + \frac{PR(c)}{L(c)} + \frac{PR(d)}{L(d)}$$

因此， $PR(a)$ 的算法可以寫成：

$$PR(a) = \sum_{u \in F_a} \frac{PR(u)}{L(u)}$$

即人物 a 的 PageRank 值會由集合 F_a （包含所有追蹤 a 的人物的集合）中的每個 u 人物其 PageRank 值除以 u 的總追蹤數的加總而成。

再進一步思考，假設一個公眾人物的帳號是剛申請而沒有粉絲，但這不代表說

此人重要程度為 0，還是有一定的機率是重要的，所以根據 PageRank 論文，給一個 Damping Factor (d)，作為其機率。(論文建議 0.15 [4]) 因此最終的 PageRank 公式可以寫成：

$$PR(a) = \frac{d}{N} + (1 - d) * \sum_{u \in F_a} \frac{PR(u)}{L(u)}$$

在集合中，所有的 PageRank 值可以用一個特殊相鄰矩陣的特徵向量表示。其特徵向量 R 為下圖，其中 p_1, p_2, \dots, p_N 為目標人物。

$$\mathbf{R} = \begin{bmatrix} PR(p_1) \\ PR(p_2) \\ \vdots \\ PR(p_N) \end{bmatrix}$$

同時，R 也是下列方程式的解，其中相鄰函數 $\ell(p_i, p_j)$ 代表“ p_j 是否追蹤 p_i ”與“ p_j 的總追蹤數”的比值。

$$\mathbf{R} = \begin{bmatrix} (1-d)/N \\ (1-d)/N \\ \vdots \\ (1-d)/N \end{bmatrix} + d \begin{bmatrix} \ell(p_1, p_1) & \ell(p_1, p_2) & \dots & \ell(p_1, p_N) \\ \ell(p_2, p_1) & \ddots & & \\ \vdots & & \ell(p_i, p_j) & \\ \ell(p_N, p_1) & & & \ell(p_N, p_N) \end{bmatrix} \mathbf{R} \quad 4$$

接著再經過幾次的迭代運算後，便可以求出 PageRank 值的特徵向量，也就是我們的目標。

5. Network Property

在網路理論中，Small World Network 是一類特殊的複雜網路，在此網路中，部分的節點並不相連，但經過少數幾步後，便可以到達。本實驗將此概念套用至上述立場判斷完的 Following Graph，藉由運算在一群體間網路的直徑與平均路徑長度，進而分析此社群的關係緊密程度。其中直徑 D、平均路徑長度 L 計算如下，而 $(d_{i,j})$ 為節點 i 和 j 的最短路徑。

$$D = \text{Max}_{i,j} (d_{i,j})$$

$$L = 1/(n(n-1)/2) \sum_{i>j} d_{ij}$$

6. Style Mining

接著我們想更進一步探討同一社群其發文風格及特色，其中本實驗將針對同一個社群內的所有公眾人物的最新 20 篇貼文的文章與照片進行風格分析。

- (1) 文章：中央研究院-中文斷詞系統 (<http://ckipsvr.iis.sinica.edu.tw/>) 提供線上使用，使得將公眾人物的文章做斷詞變得容易。本實驗將以關鍵詞和情緒詞作探

⁴ <https://en.m.wikipedia.org/wiki/PageRank>

勘。

- 關鍵詞：首先以社群為單位，針對同一社群的文章使用 TF-IDF 演算法 (Term Frequency-Inverse Document Frequency)，找出每個文章中前 10 名重要程度的詞（每個社群預估將會有超過 2000 個），再使用 Python 中 Gensim 套件的 Word2Vec 模型，便可以得到詞向量 (Word Vector)。最後使用分群演算法，分析出最多且密集的詞，進而推論出每個社群常使用的關鍵詞的風格。
- 情緒詞：中文維度型情感辭典 ([CVAW](#))，是一個針對中文詞做二維向量分佈的詞典，其中 X 軸代表正面及負面的情感，而 Y 軸代表平淡及興奮的情緒。因此，透過 TF-IDF 演算法得出的詞，我們還可以帶入 CVAW，找出分佈密度較高的情緒詞。

(2) 照片：顏色時常能表達出發文者的情緒，例如紅色表現活力、熱情正面情緒，綠色給人有寧靜的感覺。因此對於發文者而言，顏色的選擇、調配，往往是需要經過反覆琢磨的。而一張照片中的人物往往隱藏了相當多的風格資訊，例如為了吸引年輕人注意，偏綠陣營的照片出現的人物年齡，可能就相對低，而偏藍派陣營可能會想吸引婦女們的注意，所以出現人物就多為女性，亦或者某陣營為了表達該團隊十分和諧，所以常常發出多人的合照。因此本研究將針對顏色和人物特徵做社群的風格分析。

- 顏色：首先我們將原先使用 RGB 表示的照片轉換成 HSV 表示，HSV 是以色相 (H)、飽和度 (S)、明度 (V) 來表達顏色，會比 RGB 表示更加直觀。接著再以 Color Histogram 來分析照片 (圖 7)，將頻率較高的顏色記錄下來，最後統計同一社群裡，常出現的顏色分佈。
- 人物特徵：針對人臉偵測及屬性，Microsoft 提供了一個強大的 API ([臉部偵測和屬性](#))，我們將透過此 API，偵測出每張照片中人物的年紀、性別以及平均每張相片出現的人臉個數，接著經過統計運算，進而分析出不同社群間針對照片人物的風格差異。

7. Performance Evaluation

最後的評估我們將以人為方式做最終確認，其中包含重要人物、community 緊密關係、風格文化。以風格文化為例，藍營貼文常出現中華民國、兩岸一家親等詞，相片風格則常出現國旗；而綠營貼文則是常出現台灣、民主等，相片風格通常出現多面積的綠色。

五. 預期結果

建立公眾人物在 Instagram 上的政治立場光譜，並分析出各個群體內部的
重要人物、發文風格文化以及群體關係的緊密程度。

六. 參考文獻

- [1] E. A. Leicht, M. E. J. Newman: Community Structure in Directed Networks, Phys. Rev. Lett. 100, 118703 (2008)

- [2] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre: Fast unfolding of communities in large networks, Journal of Statistical Mechanics Theory and Experiment (2008)
- [3] Jaewon Yang, Jure Leskovec: Overlapping Community Detection at Scale: A Nonnegative Matrix Factorization Approach, Proceedings of the sixth ACM international conference on Web search and data mining (2013)
- [4] Sergey Brin, Lawrence Page: The anatomy of a large-scale hypertextual Web search engine, Computer Networks and ISDN Systems, Vol. 30, Issue 1-7, 1998.
- [5] Robert bond, Solomon Messing: Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook, American Political Science Review , Vol. 109, Issue 1 (2015)
- [6] Ming-Hung Wang, Alex Chuan-Hsien Chang, Kuan-Ta Chen, Chin-Laung Lei: Estimating Ideological Scores of Facebook Pages: An Empirical Study in Taiwan, The Computer Journal, Vol. 60, Issue 11 (2017)
- [7] Carlos H.G. Ferreira, Fabricio Murai, Ana P.C. Silva, Jussara M. Almeida, Martino Trevisan, Luca Vassio, Marco Mellia, Idilio Drago: On the dynamics of political discussions on Instagram: A network perspective, Online Social Networks and Media Vol. 25 (2021)
- [8] 吳宗樺, 黃懷萱, 沈錕坤: 社群媒體上使用者政治傾向之探勘分析: 以 PTT 八卦板為例, 計算傳播研討會 (2021)

七. 需要教授指導內容

此研究涉及多方領域的知識，需要向教授請教相關問題及文獻研讀與技術方面的指導。尤其是在社群網絡這方面的演算法和分析應用，需要教授在實作途中適時給予引導修正。