



誰識KOL?

2020台灣總統大選在 Facebook 上的社群網絡分析

游騰林

2020-09-06

開始前來點小調查

專案 PM



- 你的產品有哪些客戶?
- 怎麼找出重要的客戶?
- 他們怎麼看你的產品?

分析師



- 社群網絡分析是什麼?
- 要怎麼分析文本資料?
- 如何視覺化分析成果?



游騰林 數據分析師



東華大學 社會學研究所



國泰世華 數據部



TLYu0419@gmail.com



<https://github.com/TLYu0419>



國泰世華
國泰人壽
國泰產險
國泰證券
國泰創投
國泰投信

By collaborating with Cathay's subsidiaries and consolidating the group resources, DDT provides innovative digital products and services driven by data with brilliant customer experiences, aiming to become the top in Asia Pacific Fintech ecology.

時間安排

介紹

SNA

NLP

總結

介紹

SNA

NLP

總結

- 專案目標
- 使用套件
- 資料結構

- 如何評估重要性?
- 找出 KOL
- 視覺化

- NLP 基礎
- LDA 分析
- 視覺化

- SNA + NLP
- 更多應用



「專案介紹 Introduction」

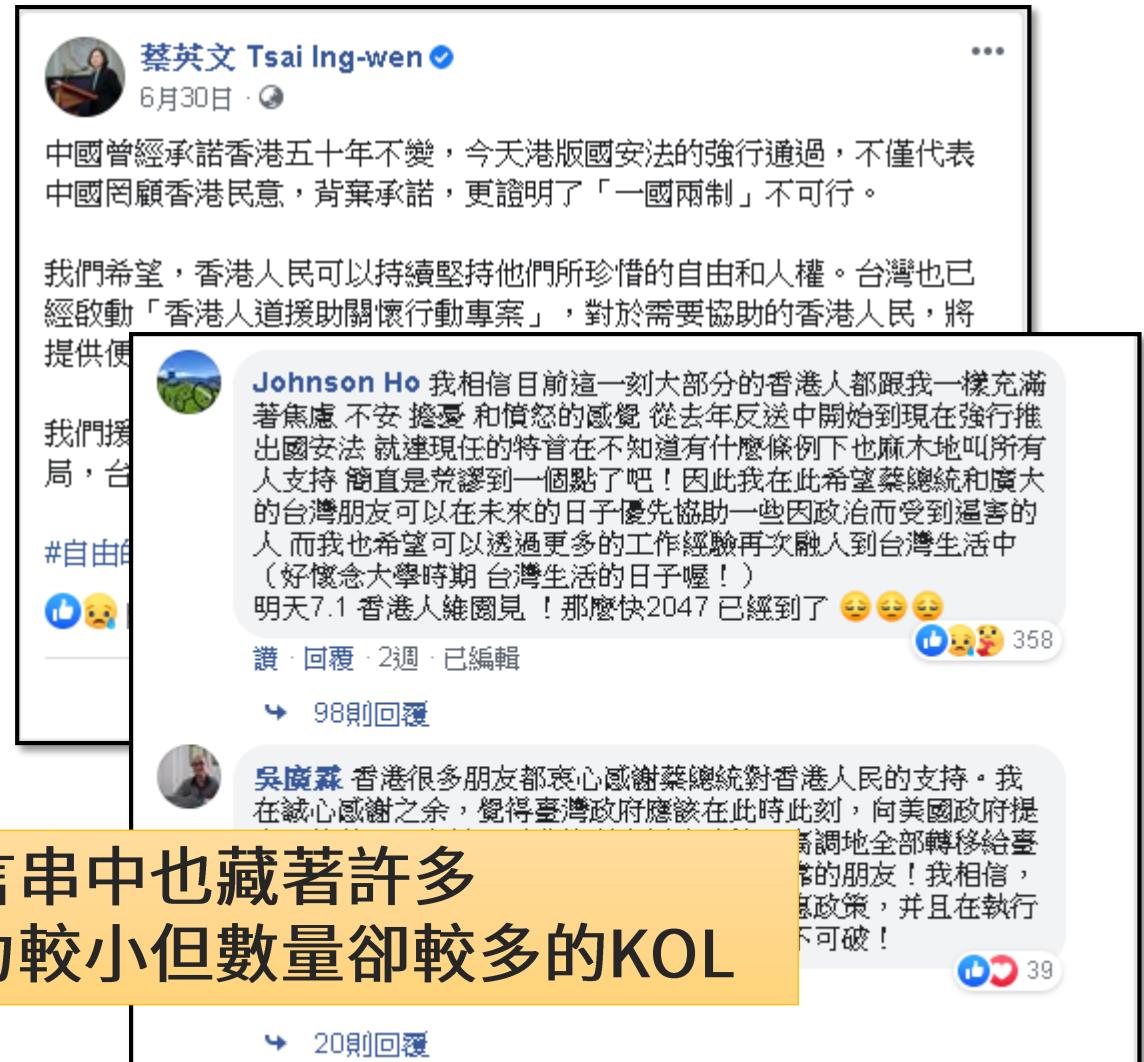
誰是你的 KOL？

介紹

SNA

NLP

總結



在留言串中也藏著許多
影響力較小但數量卻較多的KOL

專案目標

找出 KOL

透過 SNA 找出位在網絡核心的 KOL

認識 KOL

透過 NLP 認識 KOL 討論的熱門話題

為什麼需要 SNA？

介紹

SNA

NLP

總結

INSIDE 5G AI 新創 評論 焦點 活動 Jobs 好工作 繁 / 簡 訂閱

趨勢
抓到網軍！Twitter 永久刪除來自中、俄、土耳其的 3.2 萬個帳號
2020/06/12 · Heemie · Twitter、網軍、言論、仇恨言論
除了刪掉中國核心網軍的大量帳號，Twitter 也發現有額
軍帳號所發表內容的正當性。

中時電子報 真道理 真愛台灣
新冠肺炎 即時 政治 言論 生活 娛樂 財經
首頁 / 中國時報

網軍帶風向價碼 網站全
04:10 2020/06/26 | 中國時報 | 吳家豪
拍賣網站驚見網軍出售帶風向的服務，且價格全都露。圖為國民黨去年率眾到行政院陳情，要求消滅網軍東廠。(本報資料照)

蝦皮優選
蝦皮超便宜

facebook

粉絲專頁100(讚+追蹤) 40元
貼文100讚 40元
台灣真人50讚 50元
台灣真人留言讚 2.8元
直播 LIVE 80元

全館最便宜
全店免運

facebook 按讚 / 直播 / 真人 / 留言讚 / 粉絲專頁讚 / 直...

\$10 - \$50 Free

月銷量 1 萬
台南市佳里區

為什麼需要 SNA ?

介紹

SNA

NLP

總結



Python 套件

介紹

SNA

NLP

總結



Ref:

<https://tlyu0419.github.io/2020/03/17/Crawl-Facebook-Pages>

誰識KOL? 2020台灣總統大選在 Facebook 上的社群網絡分析

Mail: TLYu0419@gmail.com

Python 套件

介紹

SNA

NLP

總結



Facebook粉絲專頁爬蟲筆記

發表於 2020-03-17 | 更新於: 2020-04-08 | 閱讀次數: 2760

先前曾發表過 [網路爬蟲_Facebook粉絲團貼文與留言](#) 的文章，可以透過 Selenium 來抓取 Facebook 粉絲專頁的貼文與留言，但由於 Selenium 的效率實在太慢了，因此這篇文章是記錄我嘗試用 request 抓取資料的筆記，並且於文末附上台灣的2020總統大選 3位候選人從 2019 年 1 月至 2020 年 2 月間的貼文與留言資料~



完整的 Python 程式碼如下，有興趣的人請參考
<https://github.com/TLYu0419/FindAndMeetKOLs>



Ref:

<https://tlyu0419.github.io/2020/03/17/Crawl-Facebook-Pages>

誰識KOL? 2020台灣總統大選在 Facebook 上的社群網絡分析

Mail: TLYu0419@gmail.com

Facebook 上有哪些資料？

介紹

SNA

NLP

總結



蔡英文 Tsai Ing-wen 6月30日

中國曾經承諾香港五十年不變，今天港版國安法的強行通過，不僅代表中國罔顧香港民意，背棄承諾，更證明了「一國兩制」不可行。

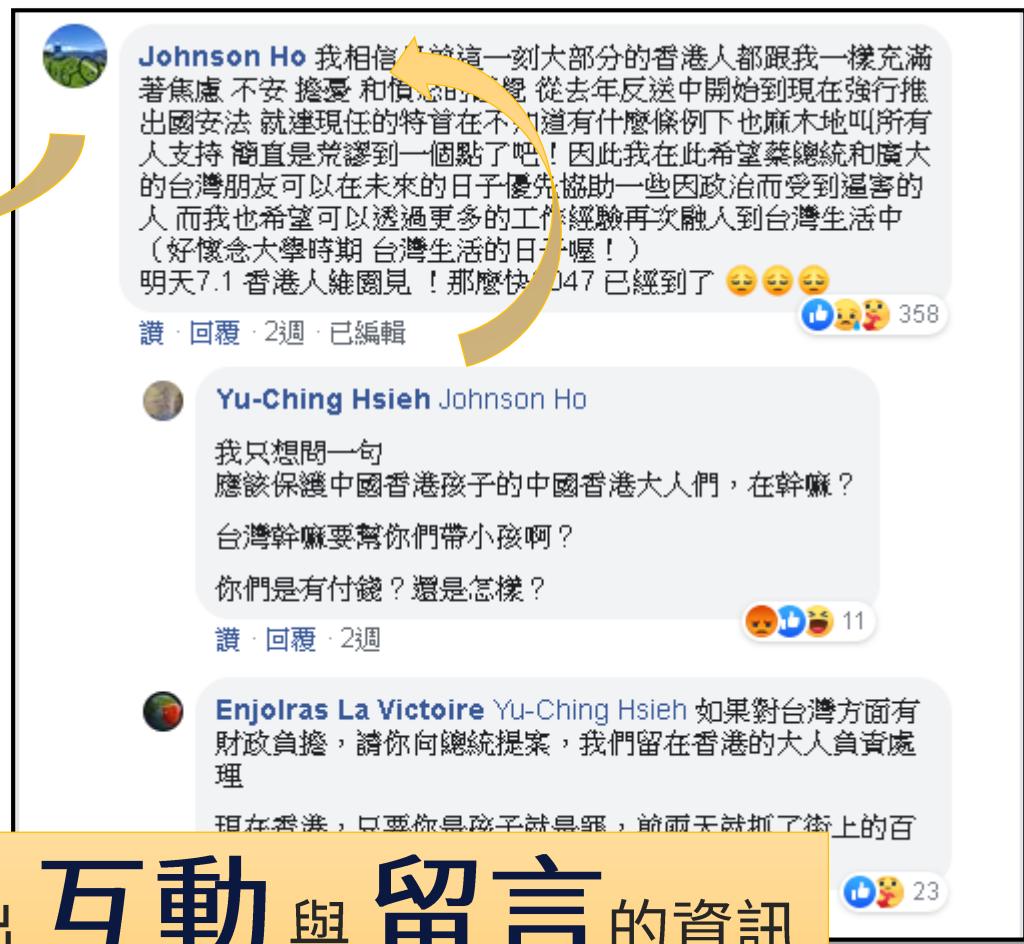
我們希望，香港人民可以持續堅持他們所珍惜的自由和人權。台灣也已經啟動「香港人道援助關懷行動專案」，對於需要協助的香港人民，將提供便捷的服務及基本照顧。

我們援港專案的「台港服務交流辦公室」，將在明天正式營運。因應變局，台灣將不分政府民間，共同協力，給予香港人民最堅定的支持。

#自由的台灣撐香港的自由

1 陳井泰和其他11萬人 7,424則留言 5,061次分享

讚 留言 分享



Johnson Ho 我相信這一刻大部分的香港人都跟我一樣充滿著焦慮、不安、擔憂和憤怒的感覺。從去年反送中開始到現在強行推出國安法，就連現任的特首在不知道有什麼條例下也麻木地叫所有人支持，簡直是荒謬到一個點了吧！因此我在此希望蔡總統和廣大的台灣朋友可以在未來的日子優先協助一些因政治而受到逼害的人，而我也希望可以透過更多的工作經驗再次融入到台灣生活中（好懷念大學時期台灣生活的日子喔！）

明天7.1香港人難覓見！那麼快2047已經到了 😢😢😢 358

讚 · 回覆 · 2週 · 已編輯

Yu-Ching Hsieh Johnson Ho
我只想問一句
應該保護中國香港孩子的中國香港大人們，在幹嘛？
台灣幹嘛要幫你們帶小孩啊？
你們是有付錢？還是怎樣？

讚 · 回覆 · 2週

Enjolras La Victoire Yu-Ching Hsieh 如果對台灣方面有財政負擔，請你向總統提案，我們留在香港的大人負責處理

現在香港，只要你是孩子就是罪，前兩天就抓了街上的百

讚 23

Facebook 可以萃取出 **互動** 與 **留言** 的資訊

資料結構

介紹

SNA

NLP

總結

互動資料表：559,763 rows * 5 columns

SOURCE_NAME	SOURCE	TARGET_NAME	TARGET	WEIGHT
Bill***	1471*****	蔡英文 Tsai Ing-wen	4625*****	1
Ash***	5673*****	蔡英文 Tsai Ing-wen	4625*****	1
Billy***	1000*****	蔡英文 Tsai Ing-wen	4625*****	1
Vicky***	1660*****	蔡英文 Tsai Ing-wen	4625*****	1

留言資料表：1,865,409 rows * 5 columns

NAME	AUTHOR	TIME	TEXT
謝**	1000*****	2020-01-01 20:15	松山到了，宋上時刻到了！
Lai***	1000*****	2020-01-01 16:13	棄草棄菜挺宋，選賢與能愛台灣，台灣要蛻變...
Bruce***	1000*****	2020-01-01 22:45	唉，現在的政治卡小 藍綠白黃，都是個鳥樣. ..
楊***	1000*****	2020-01-04 20:04	韓市長加油 告到底 老虎不發威當成病貓 可惡...

註：時間範圍為大選前2個月(2019/11/1 ~ 2020/1/10)

誰識KOL? 2020台灣總統大選在 Facebook 上的社群網絡分析

Mail: TLYu0419@gmail.com



「社群網絡分析」

Social Network Analysis

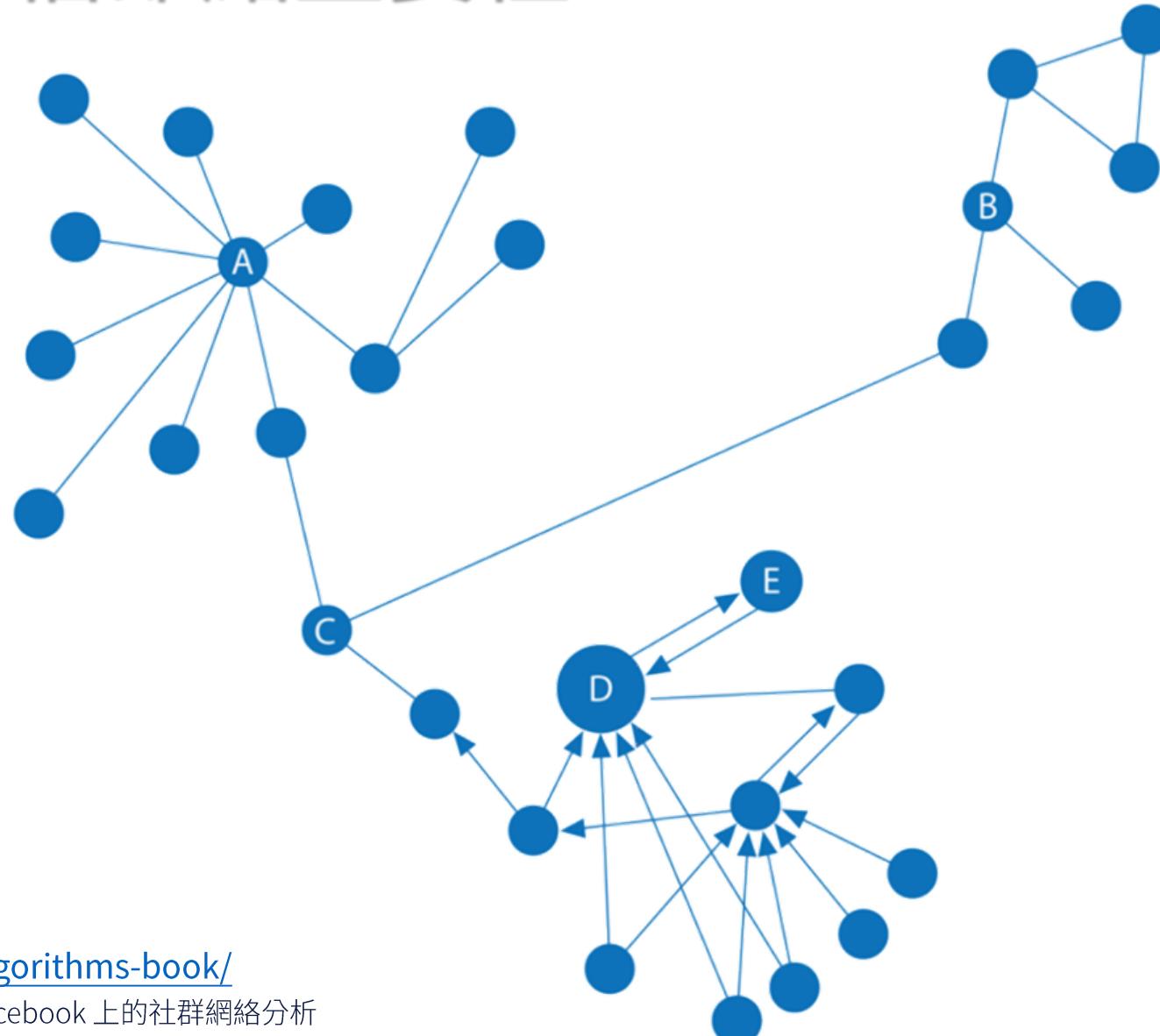
如何評估節點重要性?

介紹

SNA

NLP

總結



Ref:

<https://neo4j.com/graph-algorithms-book/>

誰識KOL? 2020台灣總統大選在Facebook上的社群網絡分析

Mail: TLYu0419@gmail.com

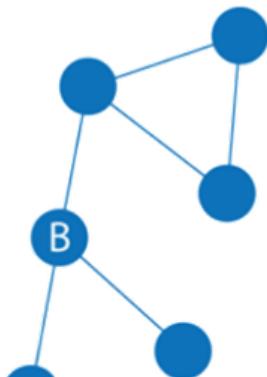
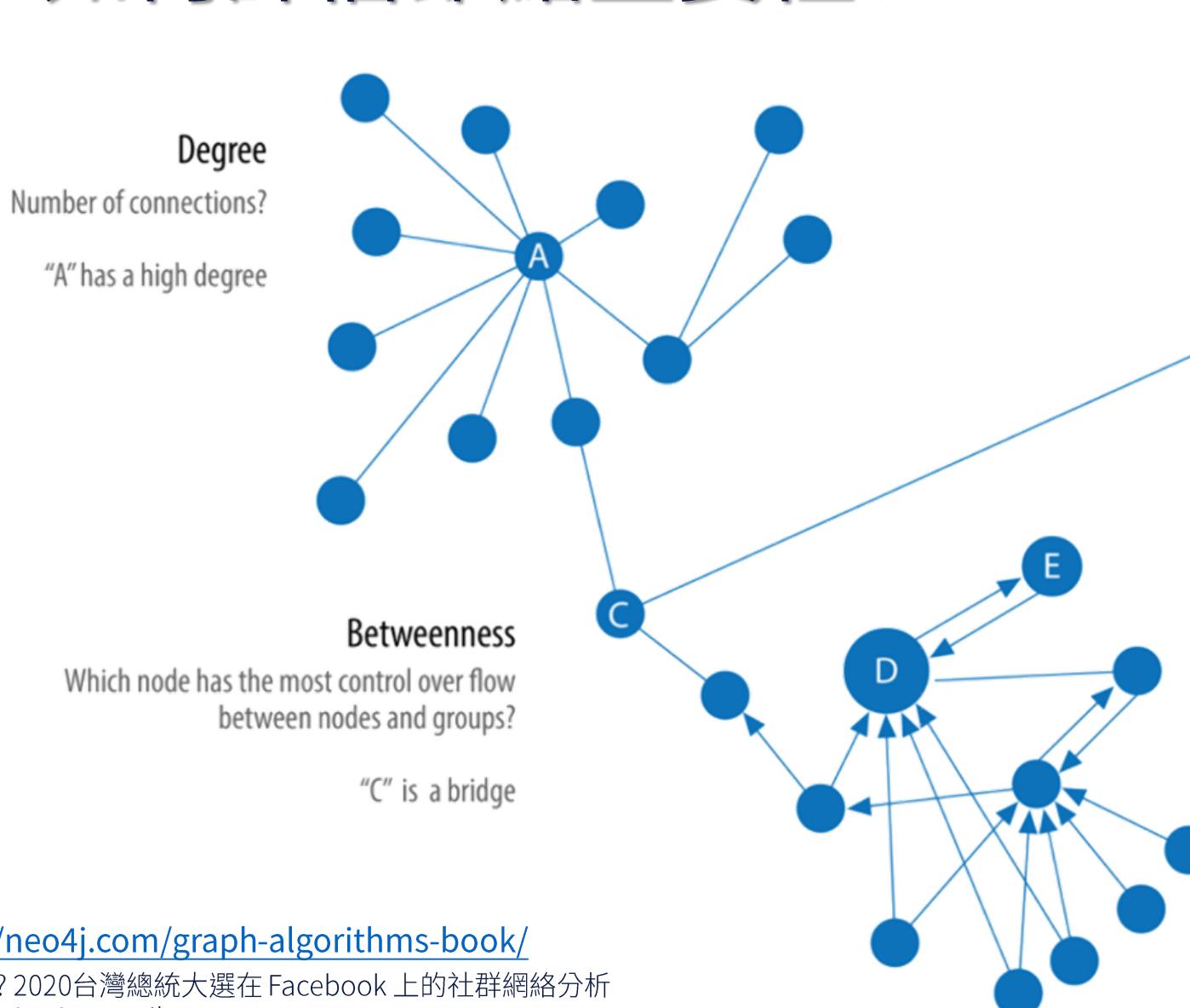
如何評估節點重要性?

介紹

SNA

NLP

總結



Closeness
Which node can most easily reach all other nodes in a graph or subgraph?
"B" is closest with the fewest hops in its subgraph

PageRank
Which node is the most important?
"D" is foremost based on number & weighting of in-links
"E" is next, due to the influence of D's link

Ref:

<https://neo4j.com/graph-algorithms-book/>

誰識KOL? 2020台灣總統大選在Facebook上的社群網絡分析

Mail: TLYu0419@gmail.com

PageRank 簡介

介紹

SNA

NLP

總結



- 網頁的重要性排序
- 想法
 - In-link越多越重要
 - 來自重要網頁的引用較重要

如果是不重要的節點(假帳號)
藉此創造出的互動也不會有價值

Ref : <http://jpndbs.lib.ntu.edu.tw/DB/PageRank.pdf>

誰識KOL? 2020台灣總統大選在Facebook上的社群網絡分析

Mail: TLYu0419@gmail.com

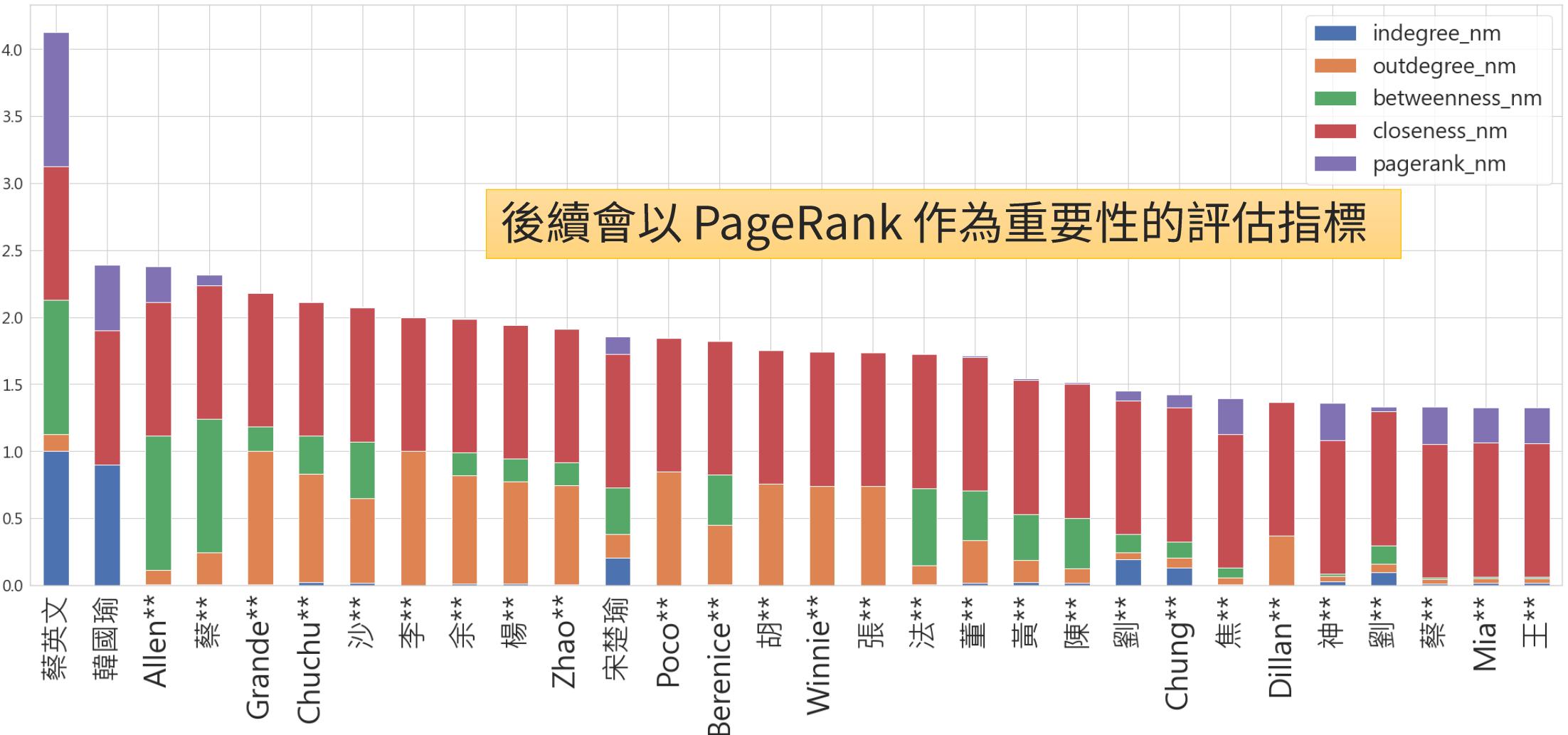
誰是網絡中最重要的人？

介紹

SNA

NLP

總結



SNA 視覺化

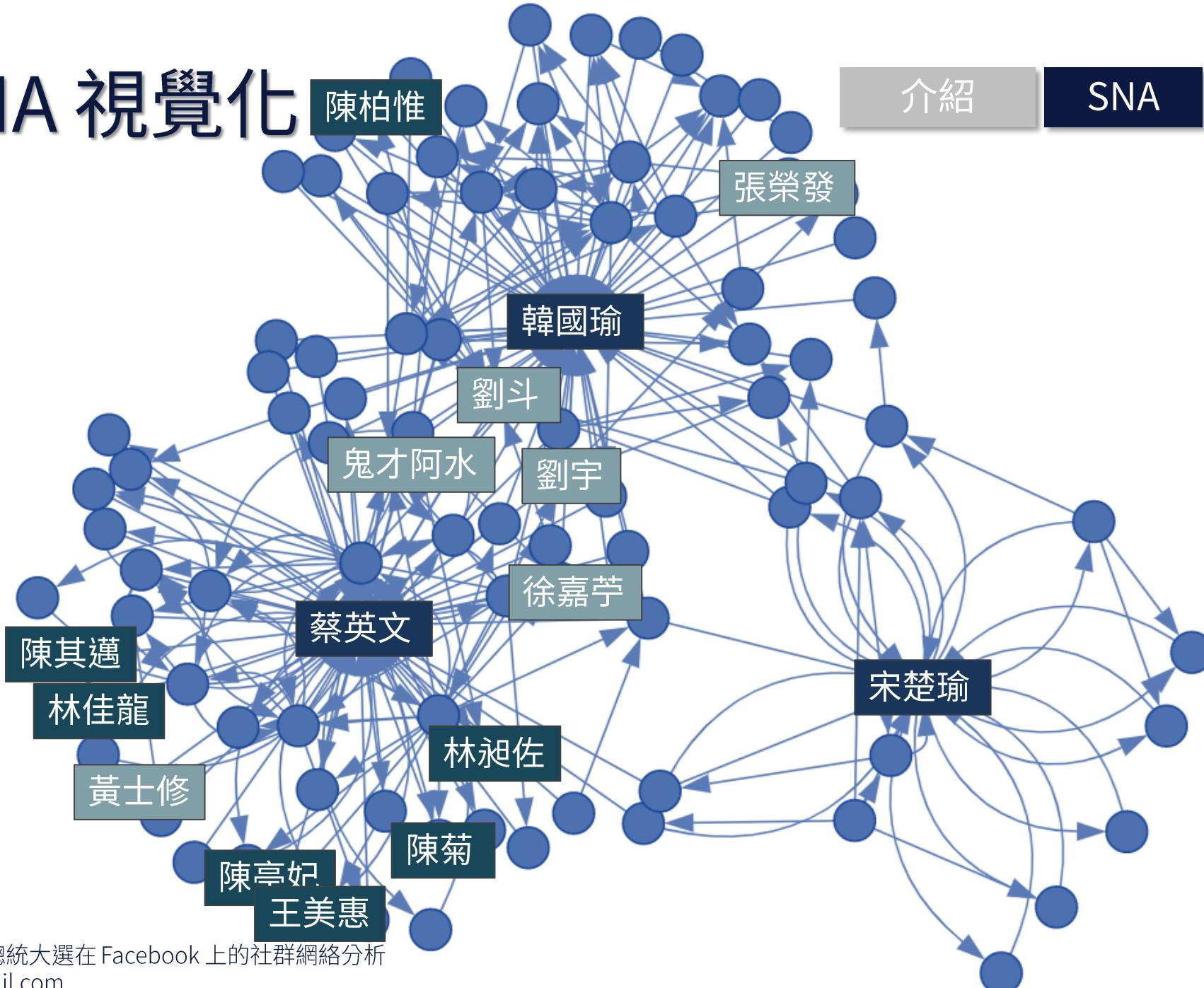
陳柏惟

介紹

SNA

NLP

總結



SNA 視覺化

陳柏惟

介紹

SNA

NLP

總結



陳柏惟 高雄失去了市長 1,074



陳柏惟 「難民法」！所以台灣跟中國一邊一國惹！謝韓韓支持台獨！
對了，知道難民法是國對國就好

598

張榮發

張榮發 #重申若我宣佈放棄假釋關到死
對我當初挺你也願意陪關到老去
賭這麼大是對韓總統有絕對信心

張榮發 #國政領航、#台灣富強
#勤政愛民、#苦民所苦
我們都相信有信心迎接勝利光輝
唯一支持韓國瑜張善政當選總統

張榮發 #陳樹菊阿嬤大慈善家不會看錯人有信心韓國瑜可以做一個全民無私大愛苦民所苦的全民總統

劉斗

鬼才阿水

劉宇

徐嘉亭

靠近並不代表支持，也有可能是在反諷/謾罵

陳其邁

林佳龍

黃士修

蔡英文

林昶佐

陳亭妃

王美惠

宋楚瑜

林昶佐 Freddy Lim 謝謝小英總統願意來逗熱鬧，1221一起吼起來！

林昶佐 Freddy Lim 謝謝總統！今天總統撩全場數萬歌迷跟支持者，一次就把波特王的招都學起來啦！
<https://www.facebook.com/...../a.36...../1228884687322760/>

林昶佐 Freddy Lim 2:11:49 小英總統快到囉！！

SNA 視覺化

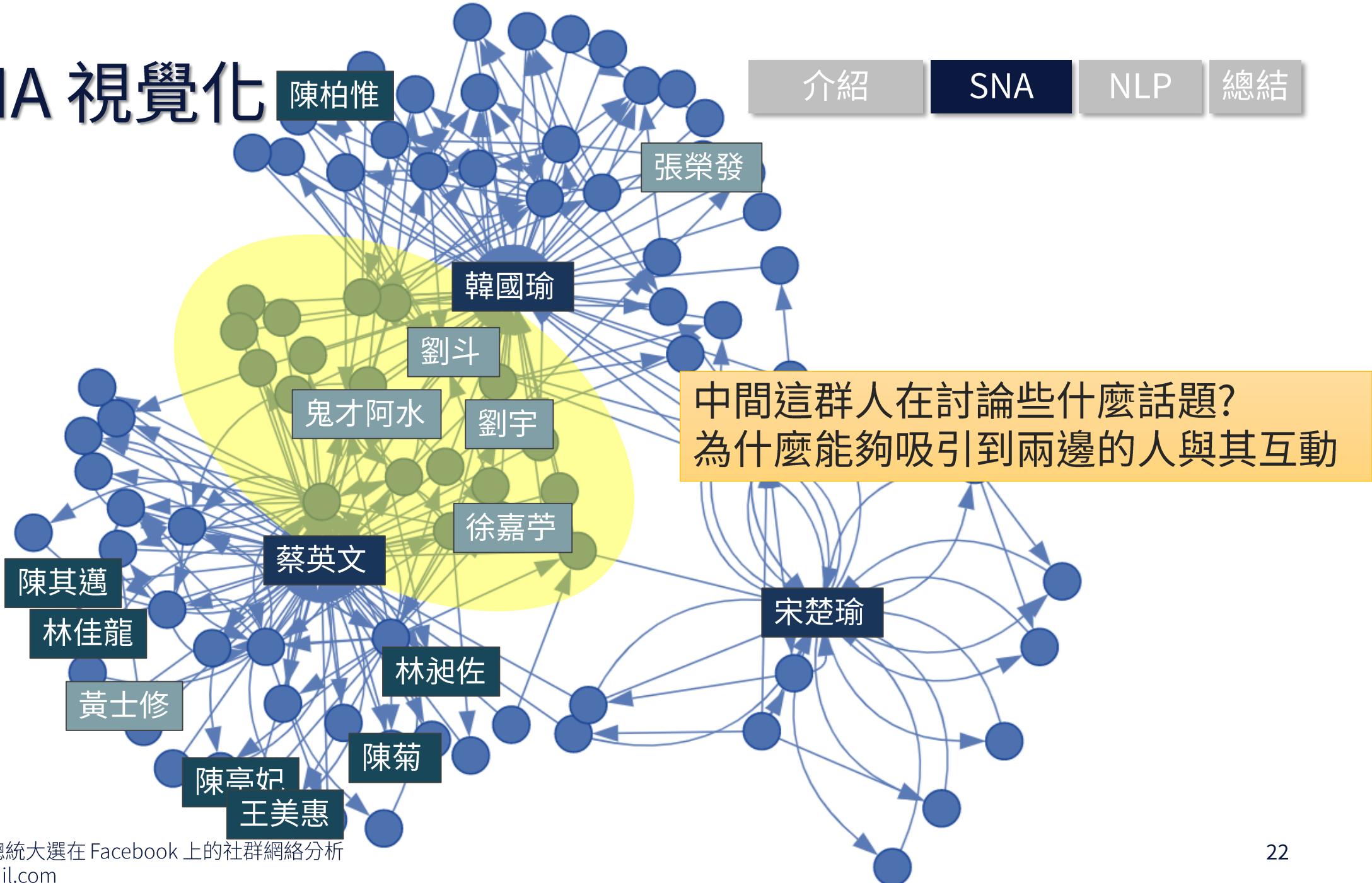
陳柏惟

介紹

SNA

NLP

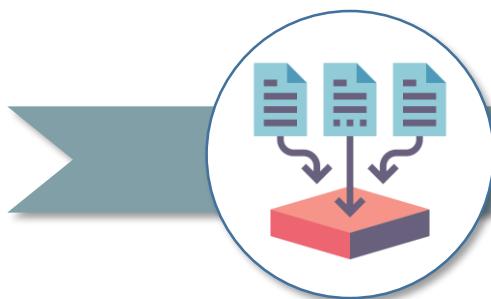
總結





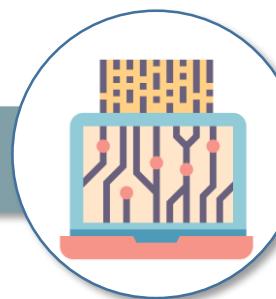
「自然語言處理」

Natural Language Processing



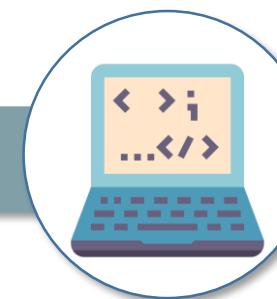
收集語料

- 社群網站
- 新聞文本
- 商品評論
- ...



預處理

- 同義詞
- 停用詞
- 特殊符號
- 斷詞
- ...



特徵工程

- 詞頻矩陣
- TF-IDF
- 詞向量
- ...



訓練模型

- 文本分類
- 情感分析
- 主題模型
- ...

LDA-主題分析

介紹

SNA

NLP

總結

AUTHOR	TEXT3
100000319616052	民主 當飯 民主 連飯 小英 點到 苗栗 ...
100002013267977	清白 檢驗 民眾 遊行 遮羞布 發言 候選...
100013992199086	治國 下台 韓投 定了 韓會 乾淨 茶滿 ...
706072435	習大大 兒子 過去 往前 高雄 不可能 豪...
100002862989012	下台 論文 矮油 加油 下台 論文 頭殼 ...
100000239611720	樹菊 阿嬤 台灣 良心 認證 庶民 加油 高...
100001876335501	支持 菜陰魂 出兵 解放 香港 嘴砲 護香...
100001601896453	翻轉 下架 小英 全民運動 下架 小英 通...
100001520370048	宋伯伯 加油 支持 經驗 政治 頭腦 裡面...
100000404266312	準備 出征 特權 盜採 砂石 質疑 出一 ...
744679351	韓總 香港人 支持 支持 香港 暴徒 所作...
100000990153827	高雄 天上 星星 不一定 高雄 四處 奔波...

帶職參選

烙跑 溜之大吉 草包 高雄

博士論文

英國 倫敦 博士 論文

國家認同

親中 共匪 統治 子子孫孫

.....

....

LDA 分析技巧

1. 去除重複
2. 轉成長文本
3. 限制詞頻

Ref :

<https://www.machinelearningplus.com/nlp/topic-modeling-python-sklearn-examples/>

文本主題結果

政治做出立委
輔助
資本
過半
選擇
團結
連勝
民進黨
剩下
政黨
小英
連任
蔡總統
台灣人
選舉
選舉
經濟
距離
站出來
政府
距離
台北
由
草包
力口
國家
台灣人民
聯設
小英
連任
蔡總統
台灣人
選舉
選舉
經濟
距離
站出來
政府
距離
台北
由
草包
力口
國家
台灣人民
聯設

唯一支持 共產黨立委 台灣人 民主
老謝 政治 下架 小英 投票 黑選舉
韓粉 政府 國民黨 支持 相信
候選人 國家 加油 票投
國 田 選票 國民黨 國 加油 票投
市長 草包 只有 香港 民進黨 中共
市長 高雄

市民政策當選潘恆旭笑死政府只有
高雄中國韓粉支持國家
票投台灣人
國民黨高雄人
落跑年輕人國瑜
香港廢除新莊
刑法市長罷免
民調民進黨

高雄 橋營 翻轉 韓總 支持 治國 團結 下架 能力 助紂為虐 韓總 賽能 國民黨 民進黨 擱置
親民黨 民進黨 總統府 庶民 政權 國民黨 擱置 總統府 民心立委

介紹

SNA

NLP

總結

文本主題結果

介紹

SNA

NLP

總結



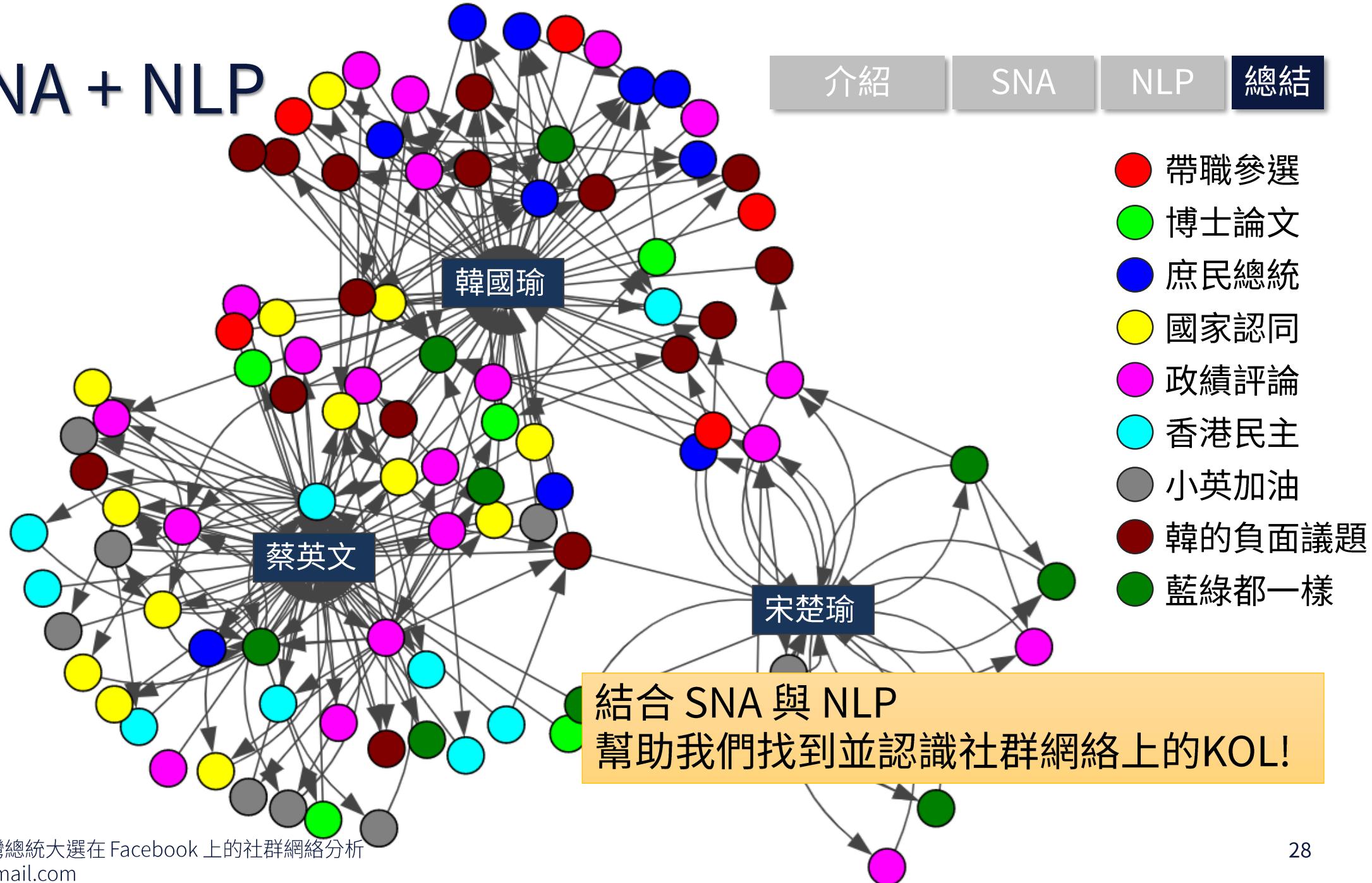
SNA + NLP

介紹

SNA

NLP

總結



SNA 的更多應用

介紹

SNA

NLP

總結



社群

找到並認識社群 KOL



客服

分析完整的服務旅程



電信

重新評估客戶的價值



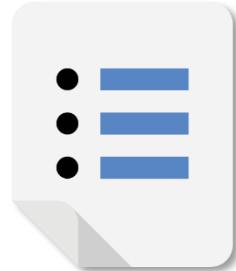
電商

抽獎活動的中獎機率



金融

偵測理專的異常金流



其他

其他更多應用

專案 PM



- 你的產品有哪些客戶?
- 怎麼找出重要的客戶?
- 他們怎麼看你的產品?

分析師



- 社群網絡分析是什麼?
- 要怎麼分析文本資料?
- 如何視覺化分析成果?