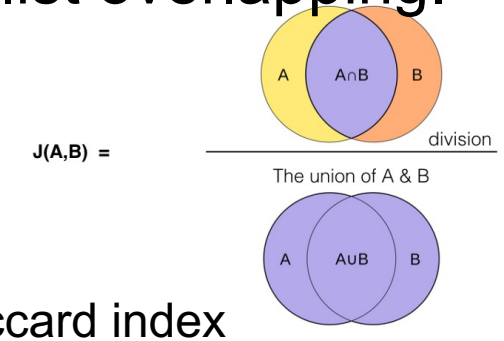
Proof of Concept:
A Diplomatic Strategy with Twitter Network

Motivation

- A diplomatic strategy
  - The goal of this study is to give a method to find a international organization's friend list.
- Avoid China's interruption
  - With this friend list, we could approach this international organization indirectly via its friend or its friend's friend.

Theoretical basis

- Six degree theory
  - This study assume that for all organization in the world, it could be reached through six of less other organization.
- Jaccard index similarity
  - This study use Jaccard index as operation definition of similarity.
  - Jaccard is defined as (A ∩ B / A ∪ B)
  - Similarity in this study is referred to account's following list similarity, namely to what extent each pair of users following list overlapping.



Jaccard index

Source: https://medium.com/data-science-bootcamp/understand-jaccard-index-jaccard-similarity-in-minutes-25a703fbf9d7

Key concept

- Agenda setting
  - We need to know in which kind of field the target organization is interested.
  - Find latent related topic which is concerned by target organization through analysis the account description of its followings.
- Official sponsorship
  - Government could know that which organization has the potential to build connection to different international organization.
  - Appropriate sponsorship.
  - Optimization deployment of diplomatic resource.
- Actual action defined relation
  - Avoid Information asymmetry.
  - Reduced exploitation from brokerage.

Analysis flow

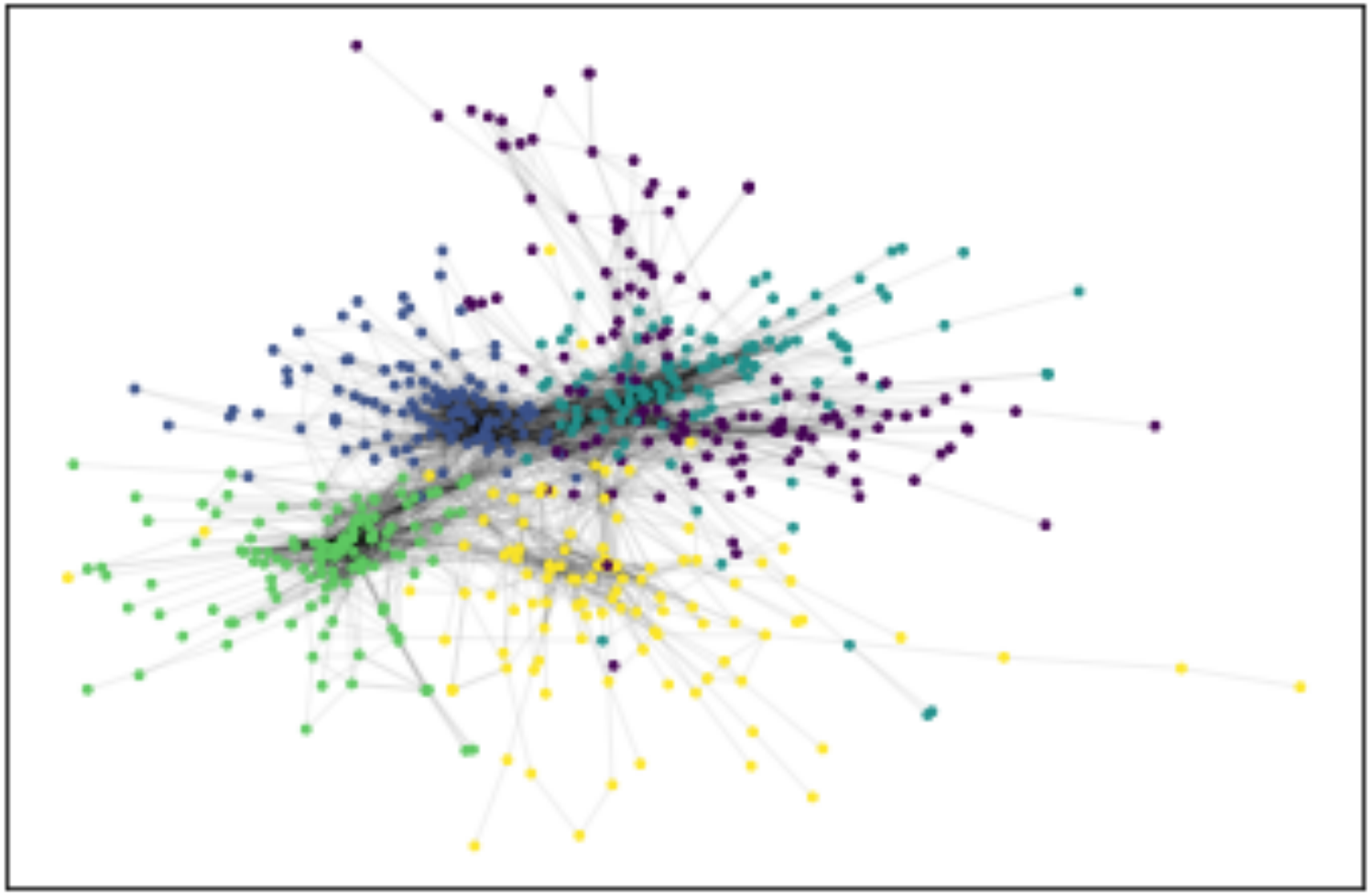
- Select target(ODIHQ in this example)
- Collect target's following account data($f_{i,i=1,2,3\ldots1704}$), namely ODIHQ has 1704 followings.
  - Account information include user's screen name, description, location and screen image.
- Collect f_i's following accounts data($s_{ik,k=1,2,3\ldots f_i's\ following\ number}$)
- Calculate Jaccard similarity of f_i and f_j where i is not equal to j.
  - We could get 1704 * 1703 similarity edge weighted by its Jaccard index.
- Filter edge with 0.07(nearly 99.9% similarity score as threshold) Jaccard index similarity
  - 4940 edges and 998 nodes left in ODIHQ's example
- Community detection with Louvain method
  - Select top 5 size group, in this ODIHQ case, 136, 132, 124, 119, 103 nodes in each group.
  - Remove node which is not in top 5 group
    - 4178 edges and 614 nodes left

Analysis flow(cont'd)

- Topic discovering
  - Tokenize account's description
  - For each group, calculate words' df-idf index
    - Document frequency: calculate each word occurred in how many account's description within each group.
    - Inverse document frequency: calculate each word occurred in how many group
  - If a word is an unique concept for a group, this word supposed to be occurred frequently in the description in this group's account and not occurred in other group.
- For each group, generate Taiwan account list
  - Taiwan account is defined as if "Taipei" or "Taiwan" in account's description or account's location in Taiwan or Taipei
  - In this step, we want to find if target's(namely ODIHQ) followers follow Taiwan related account.

Current Progress(Community detection)

- Currently, the grouping seems great, the figure shows that grouping results is more concentrated within group and more separated between group.



Latent Topic: Aid transparency

Keyword wordcloud

Screen name	被追隨數	Account Description
oktaiwanonline	24	開放知識
scheeinfo	23	徐子涵（開放知識基金會台灣代表）
audreyt	21	唐鳳
clkao	10	高嘉良（開放文化基金會董事、G0v.tw）
eikologyy	7	Board of advisory of OcfTaiwan
go_vegetables	7	亞蔬—世界蔬菜(亞洲開發銀行)
shuyanglin	5	公共數位創新空間 co-founder
TTCATz	4	吳銘軒 台灣民主實驗室 CEO

Taiwan related account

Latent Topic: Journalism

Keyword wordcloud

Screen name	被追隨數	Accountdescription
JeromeTaylor	14	法新社(AFP)記者
audreyt	9	唐鳳
tculpan	5	彭博科技專欄作者
Scholars_Stage	4	
paulmozur	3	孟建國（紐約時報記者）
heldavidson	3	衛報台灣記者
kassy	2	almost.co 創辦人
MOFA_Taiwan	2	外交部
UKinTaiwan	2	英國在台辦事處
atpittam	2	英國在台辦事處副代表

Taiwan related account

Latent Topic: Environment

Keyword wordcloud

Screen name	被追隨數	Account Description
scheeinfo	6	徐子涵（開放知識基金會台灣代表）
oktaiwanonline	5	開放知識台灣
andreasinica	2	
trc4identica	2	
clkao	2	高嘉良
dspim	2	DSP!
AndriaCheng	1	
newbloommag	1	New Bloom magazine
Stiivi	1	https://twitter.com/Stiivi

Taiwan related account

Latent Topic: E-government

Keyword wordcloud

Screen name	被追隨數	Account Description
audreyt	33	唐鳳
FangJuiChang	8	Strategic Designer in PDIS
rwang0	5	Constellation Research founder, ray wang(王瑞光)
clkao	4	高嘉良
scheeinfo	4	徐子涵（開放知識基金會台灣代表）
fiobourgeois	3	L'EHESS 政治社會學博士，研究台灣與法國的科技
WikimediaTW	3	維基媒體協會
Taiwan related account	3	開放知識台灣

Latent Topic: AI ethic

Keyword wordcloud

Screen name	被追隨數	Account Description
audreyt	13	唐鳳
paulmozur	10	孟見國
CatherineShu	2	TechCrunch Writer
rwang0	2	王瑞光
augama	1	許毓仁

Taiwan related account

Feature to-dos

- Must include
  - Interactive visualization with neo4j.
  - Deploy to web via google cloud platform.
  - Relationship confirmed with online contact data(mention or retweet relationship)
  - Users similarity with mentioned data.
- Probably include
  - Do twitter off-line event detection with online tweets data, this study could help us identify real world contact.

Related Papers

- Stewart, L. G., Arif, A., Nied, A. C., Spiro, E. S., & Starbird, K. (2017). Drawing the lines of contention: Networked frame contests within# BlackLivesMatter discourse. Proceedings of the ACM on Human-Computer Interaction, 1(CSCW), 1-23.
- Lu, W., Janssen, J., Milios, E., Japkowicz, N., & Zhang, Y. (2007). Node s imilarity in the citation graph. Knowledge and Information Systems, 11(1), 105-129.
- Goel, A., Sharma, A., Wang, D., & Yin, Z. (2013). Discovering similar users on twitter. In *11th Workshop on Mining and Learning with Graphs*.
- Li, R., Lei, K. H., Khadiwala, R., & Chang, K. C. C. (2012, April). Tedas: A twitter-based event detection and analysis system. In *2012 IEEE 28th International Conference on Data Engineering* (pp. 1273-1276). IEEE.