How Revolution Spreads and Evolves on social media

Network Analysis of Twitter User on the 2019 Hong Kong Protest Movement

Shirui Zhou (sz614@georgetown)
ANLY 645 Lab Presentation

The emergence of networked citizen politics (Peña-López, Congosto, & Aragón, 2015)
#arabspring, #15M, #ows, #YoSoy132, #occupyGezi, #vemprarua

Questions to be explored

- **Structure Fabrication:** centralized (para-institution) or decentralized (be formless be water)
    - " These movements are not institutionalized, but oftentimes mimic their nature. While others states they are unlike traditional citizens' movements, but very much alike in their decentralized structure" (Peña-López, Congosto, & Aragón, 2015)
- **Emotion Cascade:**
    - "Our findings show that activity and information cascades in the movement are larger in the presence of negative collective emotions and when users express themselves in terms related to social content." (Alvarez, Garcia, Moreno, & Schweitzer, 2015)
- **Identity Evolution:**
    - "Group gatherings contribute to the creation of collective identity by means of rituals and symbols that produce an atmosphere of emotional synchrony" (Jones, 1986)
- **Reality Synchronicity:**
    - "new moments of strong synchronization coinciding with relevant events such as Primavera Valenciana (22/02/2012), the surrounding of the Spanish Congress (25/09/2012) or the ILP campaign of the PAH (24/01/2013)" (Monterde et al., 2015)
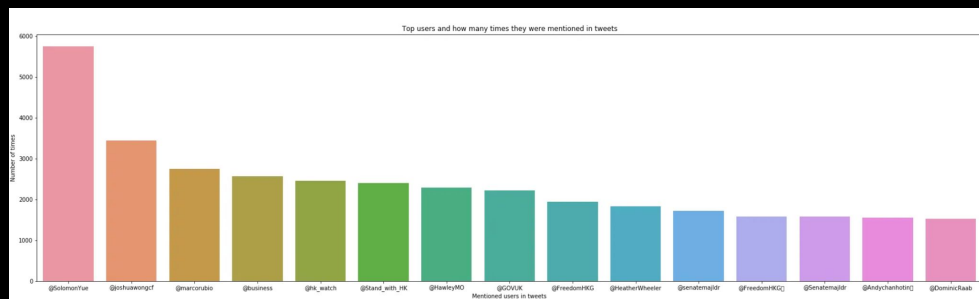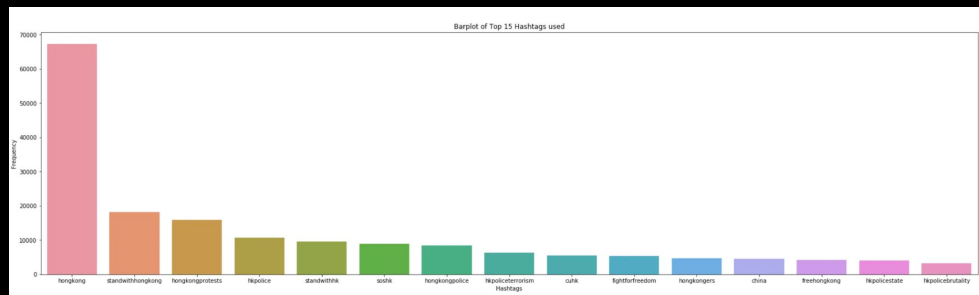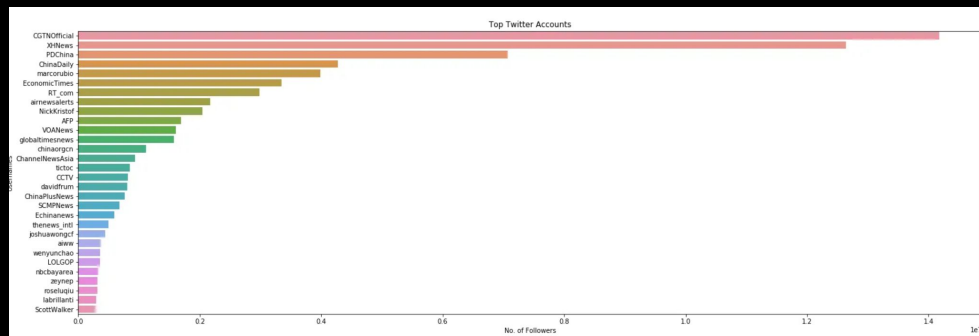
Data

- Tweepy: extract 2,500 tweets per run once every 15 minutes
  - https://python.plainenglish.io/scraping-tweets-with-twe epy-python-59413046e788
- 17 dataset with the shape of (15000, 11) -> (255003, 11)

- Concatenated dataset
  - Each row as a tweet
  - User information: user name, account description, location, following, followers, total tweets, time of tweet created
  - Tweet information: text, hashtags, time of tweet created, number of retweet

```
Size of concatenated dataset is: (255003, 11)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 255003 entries, 0 to 255002
Data columns (total 11 columns):
 #    Column            Non-Null Count    Dtype
---   ------            --------------    -----
 0    username          255003 non-null   object
 1    acctdesc          185074 non-null   object
 2    location          94066 non-null    object
 3    following         255000 non-null   float64
 4    followers         255000 non-null   float64
 5    totaltweets       255000 non-null   object
 6    usercreatedts     255000 non-null   object
 7    tweetcreatedts    255000 non-null   object
 8    retweetcount      255000 non-null   object
 9    text              255000 non-null   object
 10   hashtags          254997 non-null   object
dtypes: float64(2), object(9)
memory usage: 21.4+ MB
None
```

Descriptive Analysis (Leow, 2019)

Our word cloud for the top 100 words found in the tweets

Top Twitter Accounts

Barplot of Top 15 Hashtags used

Top users and how many times they were mentioned in tweets

Methodology

- **Data Cleaning: @user, clean_text, hashtags**

- **NLP on each tweets:** Generating Sentiments from Tweets with NLTK Vader_Lexicon Library

  - Positivity - 'pos', Negativity - 'neg', Neutrality - 'neu', Overall Score - 'compound'

text	hashtags	mentioned_users	cleaned_text	neg	neu	pos	compound	sentiment_class
Disgusting: Police in #HongKong are marking protestors with a pen, and writing a number on their hands just like the Nazis did in Germany. #China https://t.co/aAF6Jw7yXN	[hongkong]	[]	disgust police mark protestors pen write number hand like nazi germany	0.218	0.391	0.391	0.2732	4

- **Transform data structure**: row per tweet -> row per user

  - Filter dataset to more active user: `data[(data['retweetcount'] > 20) & (data['totaltweets'] > 20) & (data['followers'] > 20)] -> transformed_df: 23794 rows (users) and 332 tweet`

username	following	followers	totaltweets	usercreatedts	location	acctdesc	tweetcreatedts_1	retweetcount_1	hashtags_1	...	cleaned_text_331
11304 SolomonYue	240.0	87014.0	10967.0	2013-04-09 15:38:48	salem, oregon, usa	vice chairman & ceo at republicans overseas, r...	2019-11-09 02:26:03	581.0	[]	...	NaN

Network Analysis

- Node Attributes
  - Compute compound_mean and compound_mode from compound_ i (i ranges from 1 to 332) to obtain the average/most often emotion mode per user
- csv -> gml
  - Iterate over the dataframe rows to add all user nodes
  - Add edges from user_name to each mentioned user in each tweet

```
node [
  id 0
  label "five5demands"
  following 437.0
  followers 260.0
  location "n.a"
  compound_mode 0.0
  compound_mean 0.00306566265060241
  hashtags_1 "['hongkong']"
  tweetcreatedts_1 "2019-11-03 02:57:49"
  hashtags_2 "['standwithhongkong']"
  tweetcreatedts_2 "2019-11-07 15:42:45"
  hashtags_3 "['hkpolice']"
  tweetcreatedts_3 "2019-11-08 18:33:23"
  hashtags_4 "['hongkong']"
  tweetcreatedts_4 "2019-11-09 03:32:57"
  hashtags_5 "[]"
  tweetcreatedts_5 "2019-11-10 05:22:26"
  hashtags_6 "['standwithhongkong']"
  tweetcreatedts_6 "2019-11-11 13:53:55"
]
```

```
edge [
  source 1
  target 2
  weight 1
]
edge [
  source 1
  target 3
  weight 1
]
edge [
  source 1
  target 4
  weight 1
]
edge [
  source 1
  target 5
  weight 2
]
```
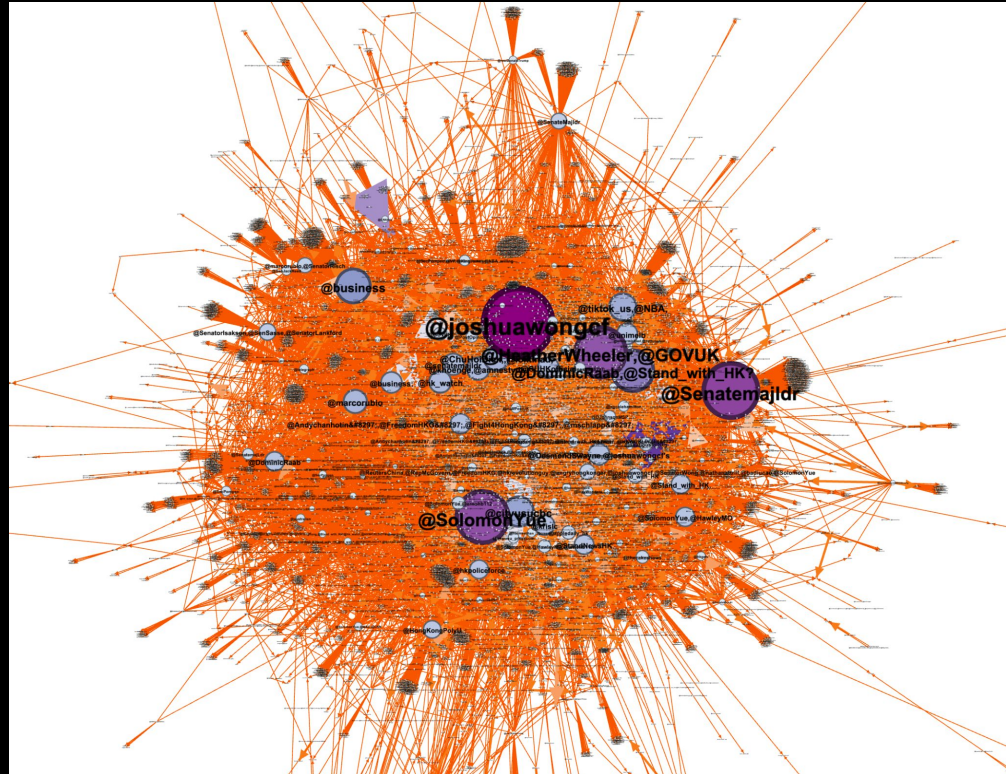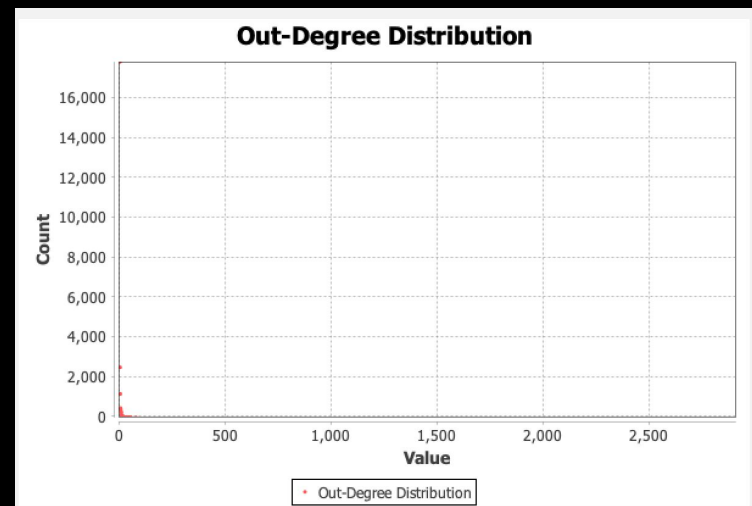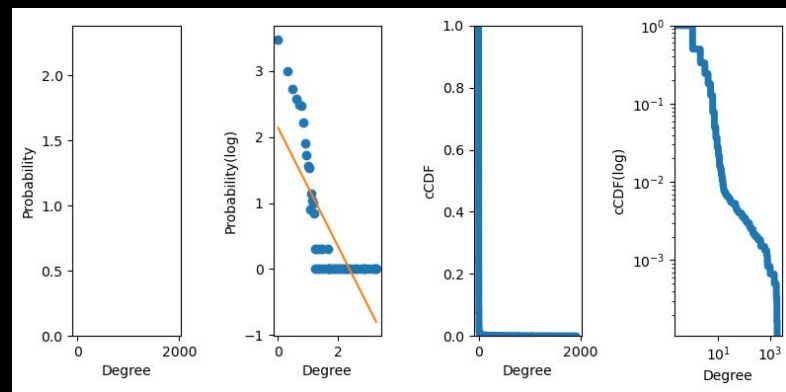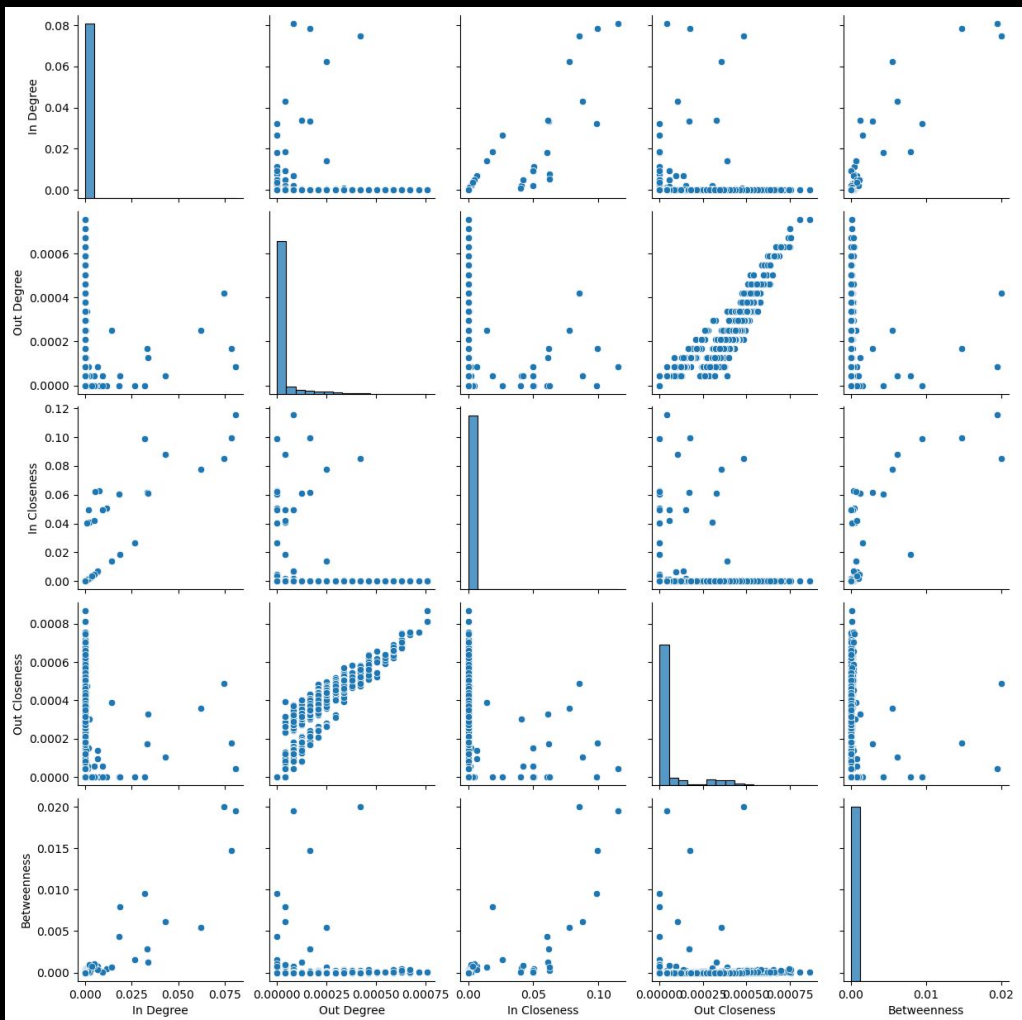
Network Structure

Basic Information

- Nodes: 23794; Edge: 15705
- Average Degree: 0.634
- Average Weighted Degree: 0.892
- Network diameter: 5
- Connected components: 17793
- Modularity 0.046
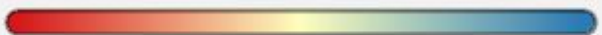
Structure Fabrication

- Fairly fragmented and decentralized
- Not to be strongly connected
- No strong community structure
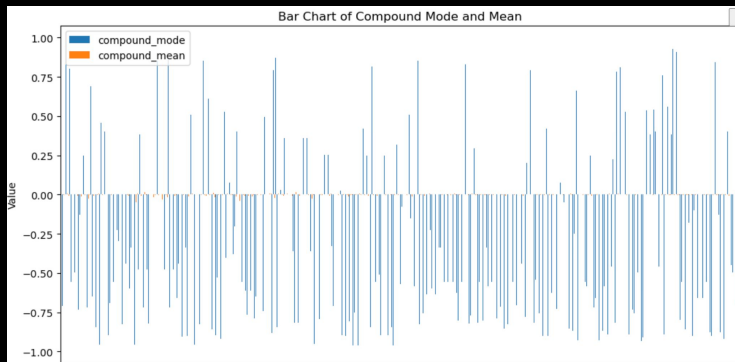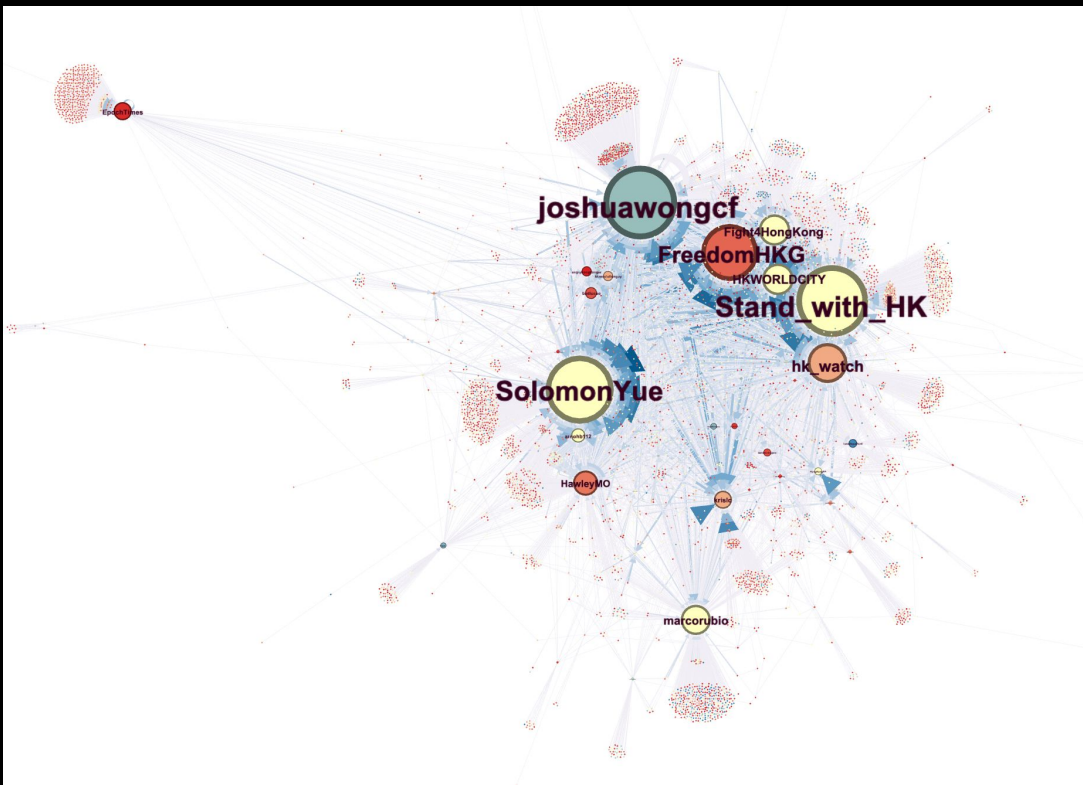
Out-Degree Distribution

Color:

Emotion Cascade

- Compound ranges from -1 (neg) to 1 (pos). Compound mode are mostly negative
- Red represent negative, blue represent positive.
- No necessary correlation exists between centrality degree and negative compound mode

Reality Synchronicity
T1: 2019-11-03 - 2019-11-06

T1 windowing screenshot on graph

- **Several hub formed around @Senatemajldr, @SolomonYue @Stand_with_HK, @joshhuawongcf, etc**
- **Hubs are silent (black - missing)**
- **Emotion states tend to be negative (red)**

Reality event during 03-06

3 November: Several sit-in protests were held at shopping malls across Hong Kong, leading to violence, arrests of journalists, and the unconsciousness of Chow Tsz-lok after a fall.

4 November: Six reporters staged a silent protest at a police press conference, leading to its cancellation, while five defendants were charged with possessing explosive substances, only to be re-arrested due to a legal error.

5 November: Over a thousand people attended a flash gathering to mark the one-month anniversary of Hong Kong's anti-mask law, resulting in road occupations, vandalism, and the police response with a water cannon truck and tear liquid.

Reality Synchronicity

T2 windowing screenshot on graph

- Emotion states continue mixing negative and positive, but more negative
- The amount of hubs increasing, new hubs such as @FreedomHKG, @Andyhanhotin, @Fight4Hongkong appears
- Hubs still showing no activity

Reality event during 07-10

8 November: Chow Tsz-lok died from his injuries, leading to widespread mourning and anger, protests in various districts, and clashes with the police.

9 November: Thousands of people gathered in Tamar Park at night to mourn those who had died for the protests.

10 November: Protesters gathered in various malls leading to vandalism, police stormed the mall with batons and pepper spray, and clashes occurred in multiple locations including allegations against the police in Tsuen Wan.

12

Reality Synchronicity
T3: 2019-11-11 - 2019-11-14

T3 windowing screenshot on graph

- Emotion becoming nearly complete negative
- Hubs in t1 reappears @senateMajldr, @Soloman Yue, with no clear community
- Hubs are more positive than the peripheral nodes

Reality event during 11-14

11-15 November: Five consecutive days of 'Three Strikes' and 'Non-Cooperation Movement'

12 November: Demonstrators at the Chinese University faced off with the police and were met with violent suppression."

Reality Synchronicity
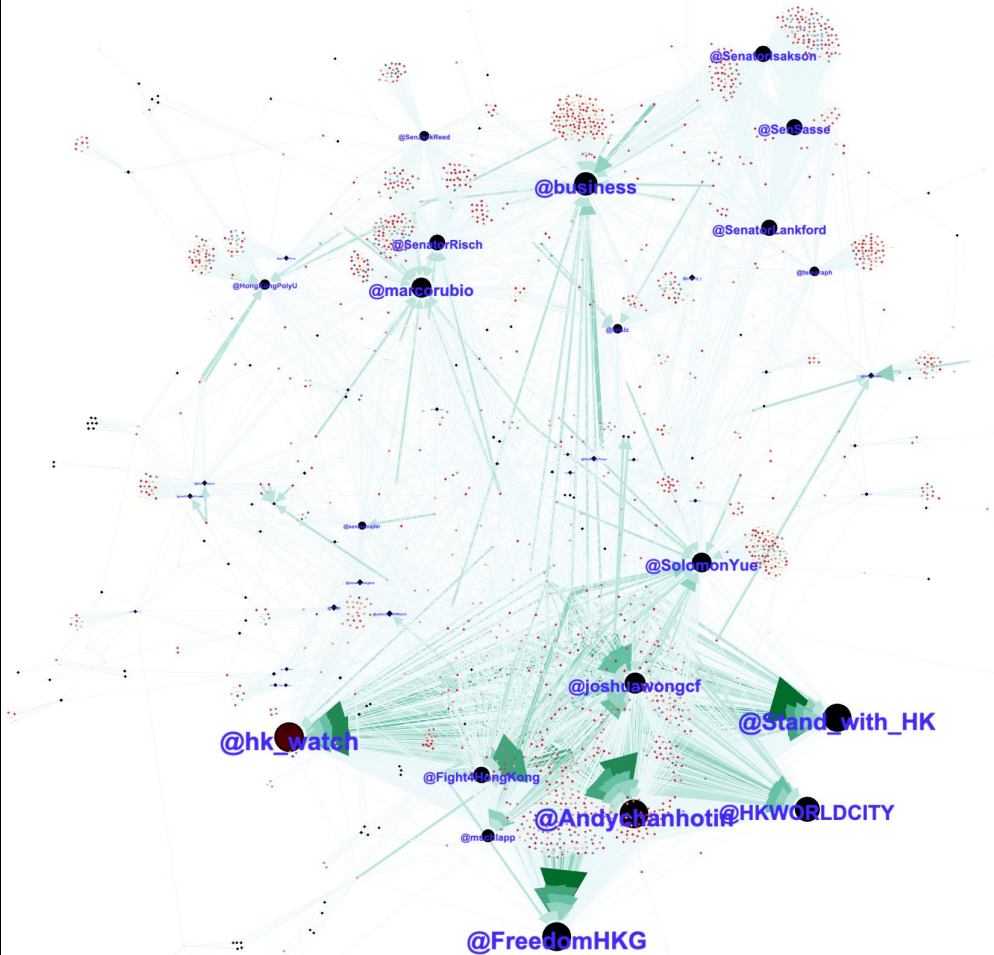## T4: 2019-11-15 - 2019-11-19

T4 windowing screenshot on graph

- New hubs such as **@HKWORLDCITY** appears, with the presence of old hubs
- Emotion are calming down than t3
- Community structure become clear dividing by targeting international and domestic representatives

Reality event during 15-19

16 November: The Hong Kong Pride Parade took place, some transportation services were restored, and the People's Liberation Army (PLA) appeared publicly to clear roadblocks.

18 November: Police entered the PolyU campus to make arrests and encountered resistance, while tear gas was used near a hospital, court rulings affected emergency ordinances, and nighttime solidarity protests occurred.

14

Conclusion

- **Structure Fabrication:** fairly decentralized, not to be strongly connected, no strong community structure, however the pattern fits power law distribution, hubs are not stable
    - Twitter role as press demands towards central figures
- **Emotion Cascade:** No necessary correlation exists between centrality degree and negative compound mode, the emotion of central nodes tend to more positive than the peripheral nodes
    - Central figure tend to be more influential in reality, requiring more prudence and carefulness
- **Identity Evolution:** Hubs existence as symbols and demands, witnesses the identity is evolved around the hubs
    - Future work: use hashtag as token of identity
- **Reality Synchronicity:**
    - The dynamic of network, appearing of hubs is accordance with the reality movement and the emotion states of all nodes can both see as cause and consequence of real world events

Reference

Alvarez, R., Garcia, D., Moreno, Y., & Schweitzer, F. (2015). Sentiment cascades in the 15M movement. EPJ Data Science, 4, 1-13.

Jones, R. A. (1986). Emile Durkheim: An Introduction to Four Major Works. In *The Elementary Forms of the Religious Life* (pp. 115-155). Sage Publications, Inc.

Leow (2019). Scraping Tweets with Tweepy & Python. Plain English.
https://python.plainenglish.io/scraping-tweets-with-tweepy-python-59413046e788

Monterde, A., Calleja-López, A., Aguilera, M., Barandiaran, X. E., & Postill, J. (2015). Multitudinous identities: a qualitative and network analysis of the 15M collective identity. Information, Communication & Society, 18(8), 930-950.

Peña-López, I., Congosto, M., & Aragón, P. (2014). Spanish Indignados and the evolution of the 15M movement on Twitter: towards networked para-institutions. Journal of Spanish cultural studies, 15(1-2), 189-216.