

How Revolution Spreads and Evolves on the Social Network

Using 2019 Hong Kong Protests as an Example

Shirui Zhou

2023-08-07

Table of contents

1 Abstract	1
2 Introduction	2
3 Data	3
4 Methods	4
5 Result and Discussion	5
6 Conclusions	14
6.1 References	15

Keywords

- Social network analysis, network movement, collective identity, emotion, Twitter, Hongkong 2019 protests.

1 Abstract

The emergence of various social platforms and ICTs has rendered the possibility of technology-facilitated participatory democracy and more effective forms of collective action, as it witnesses that the large-scale coordination and communication can happen without central direction from the digital footprint. This paper would briefly review the literatures regarding the structure of the network citizen politics, the emergence of collective identity and how the emotion

be mobilized by the information can trigger the movement in realities. We argue that the decision-making process tend to be more decentralized, the identity can be formed from symbols, and synchrony between network and real-world activities. Using the Twitter data from 2019 Hong Kong Protests caused by the anti-extradition law amendment bill, this paper would use network analysis including K-core decomposition and community detection and sentimental analysis to examine the emotion cascade and propagate from one node to another, the user-hashtag interaction, the evolution of opinion groups and emergence of opinion leaders.

2 Introduction

In the summer of 2019, lacking confidence in China's judiciary system and human rights protection due to its history of suppressing political dissent, Hong Kong citizens demonstrated against the local government led by Chief Executive Carrie Lam to legalize the extradition of suspects to mainland China. Following an escalation in the severity of policing tactics on 12 June 2019, the protester's objectives were becoming the five demands: fully withdraw the extradition bill; implement universal suffrage in Hong Kong, set up an inquiry to probe police brutality; withdraw a characterization of early protests as "riots and release those arrested at protest. Aligned with the decentralized nature of network-movement, Hong Kong 2019 protests are characterized as "formless, shapeless, like water", as the hardcore protesters organizing in small cells with no formal hierarchy and crowdsourcing their tactics and slogans through social media in a highly dispersed way. This phenomenon is accordance with the theory developed by Swann (2020) in the book of Anarchist Cybernetics: Control and Communication in Radical Politics, which explores how large-scale coordination and communication can happen without central direction, and how social media platforms can facilitate interactive communication and participatory and democratic forms of organization. Given the mutation of global scenario of social mobilization towards higher intensity of digital network and higher degree of decentralization and autonomy, it's critical to understand how the evolution spreads on social media and synchronize with the demonstration.

2.0.0.0.1 Structure Fabrication

Undeniably, ICT has become crucial tools for coordination, communication and deliberation in collective movement, however, a central question would be "whether these movements are as emergent as they seem, or have instead been designed, promoted, fostered and led by political parties or civil society organizations" Alvarez et al. (2015). In other words, it remains to be discovered whether the technology such as social media platform can push political participation towards the field of extra-representational participation, and push decision making downward to decentralized group-based process as depicted by Jones (1986). Essentially, there is large difference of the internet usage mode between citizen and institution – while the citizens are intensely use web2.0 which provides them equal excess and information, forming

emergent social networks around specific topics, institution (refer to traditional political parties) are more often conducted unidirectional, campaign-based activity online, mainly use the social media to spread parti line. To collect the empirical evidence of the structure decentralization, following Peña-López, Congosto, and Aragón (2014), this paper would define the “extra-representative movements initiated by gathering a critical mass on social networking sites” as para-institutions. Internally, this paper would discuss would para-institutions are more resembling institution-like centralized social structure or would they maintain the extra-representational form based on network or platform. Externally, this paper would explore how did the para-institution as a group related to other group (media, labor union etc.) and especially institutions (political parties)?

2.0.0.0.2 Emotion Cascade

In the paper exploring the sentimental cascades in 15M movement, according to Alvarez et al. (2015) et al, the level of engagement estimated by the number of tweets are positively correlated with the social integration, which indicated by the degree of centrality of the user. Also, negatively expression tends to have higher degree of correlation with both engagement in the activity and centrality comparing to the positive expression, while no such relationship is witnessed in the comparison between social process and cognitive process. To further investigate the validity of this hypothesis. This paper would adopt the methodology used in Alvarez et al. (2015) by first contrast a network – if u mentioned v in one of the tweets, then a directed edge from user v to user u is created, indicating the direction which information flows. Also, a vector of user features that measures the centrality, the ratio of positive tweets and negative tweets and the ratio of words related to social process and cognitive process.

2.0.0.0.3 Identity Evolution

According to Monterde et al. (2015), collective identity is formed during the social movements, as it involves a dynamic process for agents “to negotiate, understand, and construct their actions through shared, repeated interaction”. On the contrary, other arguments shifts the focus from collective identity to the “public experience of itself” by highlighting the “fluidarity” and more individualistic perspective. Nonetheless, the emergence of complex and microscopic identity from the process helping mediating the coordination and incorporating plurality features into the social movement. This paper would follow the spirit of Monterde et al. (2015) to discover the identity evolution in 2019 Hong Kong process. In social media platform, identity can be revealed by specific symbol such as hashtag or mentioning a representative figure in the movement to press the demand. In this paper, dynamic networks at each time period are captured with the mentioned figure as symbol of identity.

2.0.0.0.4 Reality Synchronicity

Monterde et al. (2015) observes the strong synchronization with relevant events in the reality, as well as the network movement between a diversity of distributed actors. Additionally, Peña-López, Congosto, and Aragón (2014) find a positive correlation between online activity and offline movement, concluding that online campaigns do materialize in real life. This paper would compare the network movement and the offline event in some selected time window to identify whether there exists some ‘coincidence’. However, the causal relationship is difficult to be deduced from the time series analysis, since the impact are in both directions.

3 Data

This paper is built on the study of Leow (2019), which extracts 2500 tweets per run once every 15 minutes using Tweepy and combines 17 datasets with the shape of 15000 row and 11 columns to 255003 rows and 11 columns. As for the concatenated dataset, each row in this comprehensive collection represents an individual tweet, and each entry contains information pertinent to both the user and the tweet itself. On the user’s side, details such as username, account description, location, number of followers and following, total tweets, and the time of tweet creation were amassed. With regards to the tweet information, the text, hashtags, time of tweet creation, and the number of retweets were meticulously gathered and recorded. The systematic gathering of this data has facilitated an in-depth analysis of various patterns and trends that may be found within the selected tweets, thereby providing valuable insights into the subject of study

4 Methods

The analytical framework of this research began with two critical stages: data cleaning and sentiment analysis using Natural Language Processing (NLP). Initially, data cleaning was employed to remove unnecessary elements such as user mentions, and to isolate hashtags, thus ensuring that the dataset was free of irrelevant content. This cleaning was vital for maintaining the integrity of the information used in subsequent analyses. The NLP step, carried out using the NLTK Vader_Lexicon Library, was instrumental in extracting the underlying sentiments from the cleaned tweets. Each tweet was analyzed to ascertain four key metrics: positivity ('pos'), negativity ('neg'), neutrality ('neu'), and an overall sentiment score ('compound'). This analysis provided a nuanced understanding of the affective content of the tweets, allowing for a more in-depth exploration of the sentiments they conveyed. Together, these initial stages laid a firm foundation for the subsequent phases of the research, assuring both the quality and relevance of the data.

Following the preliminary data preparation, the research methodology proceeded to data transformation and filtering to ensure a focus on active Twitter users. The transformation process aimed at changing the data structure from a tweet-centric view to a user-centric one, meaning

that each row would represent a unique user rather than an individual tweet. This restructuring allowed for a more meaningful and user-focused analysis. To further refine the dataset, a rigorous filtering process was applied to identify more engaged users. Specifically, users with more than 20 retweets, over 20 total tweets, and more than 20 followers were considered. The final transformed dataset (transformed_df) included 23,794 users and 332 tweets, emphasizing a more targeted and relevant subset of Twitter activity. This approach not only narrowed down the scope of analysis but also ensured that the dataset accurately represented the dynamics of engaged and active Twitter users, providing a more precise and valuable insight into the subject of study.

The network analysis phase introduced a different dimension to the study, involving the examination of relationships and emotional patterns within the Twitter user base. Firstly, the attributes of nodes were defined, representing individual users within the network. A vital part of this step was the computation of two key metrics for each user: “compound mean” and “compound mode”. These metrics were calculated from the compound sentiment scores across the 332 tweets, representing the average and most frequently occurring emotion for each user, respectively. The result was a comprehensive understanding of the prevailing emotional trends and characteristics across the user base.

Having established the node attributes, the study then proceeded to convert the data from CSV format to GML before importing into Gephi for visualization, facilitating the network analysis . An iterative process was utilized to traverse the Data Frame rows, systematically adding all user nodes to the network. Alongside the nodes, edges were added, signifying relationships between users. Specifically, edges were drawn from the “username” to each mentioned user in every tweet. These connections were instrumental in representing the intricate web of interactions and relationships within the Twitter community. This network representation not only offered a detailed visualization of the connections among users but also enabled a deeper exploration of how emotions and sentiments propagate through the network.

Recognizing the dynamic nature of social interactions, the data was strategically segmented into four distinct time periods: ‘t1’ from November 3rd to 6th, 2019; ‘t2’ from November 7th to 10th, 2019; ‘t3’ from November 11th to 14th, 2019; and ‘t4’ from November 15th to 19th, 2019. This division into specific epochs allowed for a nuanced examination of the evolving interactions and sentiment trends over time, reflecting the temporal characteristics of social media discourse. For each defined time period, a directed graph was constructed to represent the intricate dynamics of the Twitter user interactions. In these graphs, the nodes were defined as the users who tweeted within the respective time frame, and the edges were formed based on the mentioned users within the tweets. Specifically, if a user mentioned another user, an edge was drawn from the mentioning user to the mentioned user, creating a directed connection. This method of graph construction offered a detailed and structured representation of the relationships and information flow within the Twitter community during each specified period. By analyzing these directed graphs, insights were gleaned into the patterns of influence, connectivity, and sentiment propagation among users, thereby painting a comprehensive picture of the network dynamics over time.

5 Result and Discussion

The network analysis revealed a complex and fragmented structure, with a total of 23,794 nodes representing unique Twitter users and 15,705 edges symbolizing the directed connections between them. The network's quantitative characteristics further illustrate its complexity. With an average degree of 0.634 and an average weighted degree of 0.892, the network exhibits a relatively sparse connection pattern among users. The network's diameter, a measure of the longest shortest path between any two nodes, is 5, and there are a significant number of connected components, totaling 17,793. The modularity, a metric indicating the degree to which the network may be subdivided into clearly delineated communities, stands at a low 0.046.

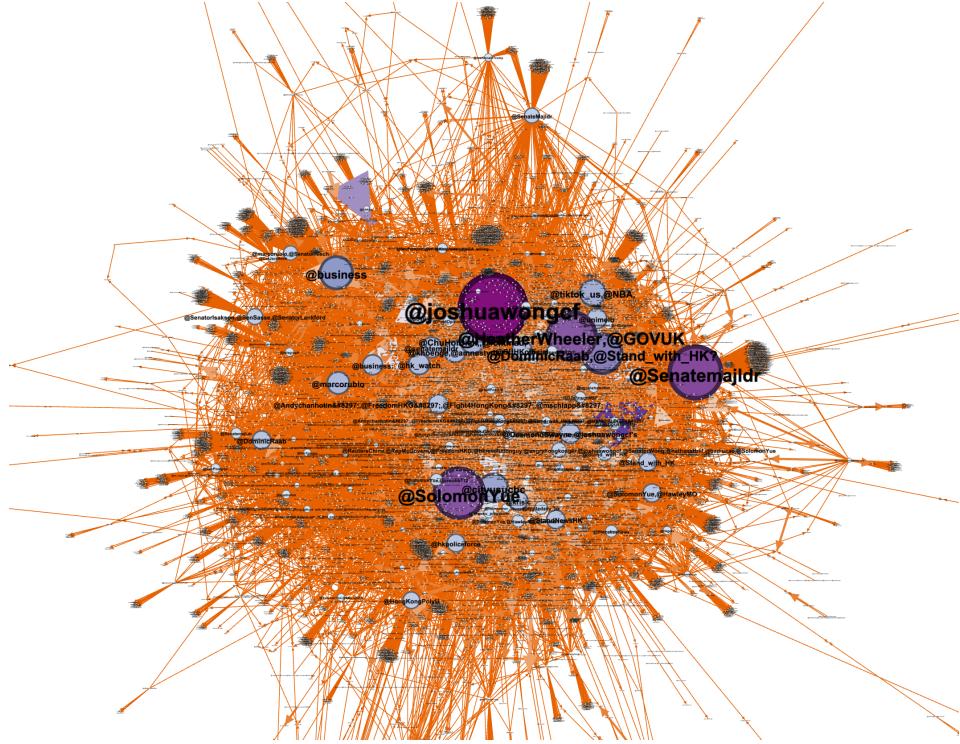


Figure 1: network structure of complete graph

The observed structural properties of the network paint a picture of a fairly fragmented and decentralized landscape. The lack of strong connections and low modularity score indicates that the network does not possess a robust community structure. In other words, users are not tightly clustered into well-defined groups, but rather spread across various connections with relatively weak ties. This configuration implies a level of independence among users and a lack of dominant hubs or centralized influence within the network. Such a structure can have implications for the spread of information, sentiment propagation, and the overall dynamics of interactions within the network.

The results of the network analysis manifest a clear indication of a power-law degree distribution, a salient characteristic of scale-free networks. This distribution signifies the presence of hubs, or nodes with a significantly higher degree than others, thus creating a heterogeneously connected structure. Further substantiation of this power-law pattern was identified through the complementary cumulative distribution function (cCDF) plot, where the slope of the fitted line closely approximated -1. Additionally, a high average degree within the network denotes a densely interconnected topology. The discernible correlation between different centrality measures, such as the strong association between degree centrality and closeness centrality, provides deeper insights into the network's architecture, revealing a structural tendency wherein nodes with numerous connections are spatially proximate to other nodes. The coherence of these metrics collectively supports the scale-free nature of the network, contributing to a nuanced understanding of its complex structural properties.

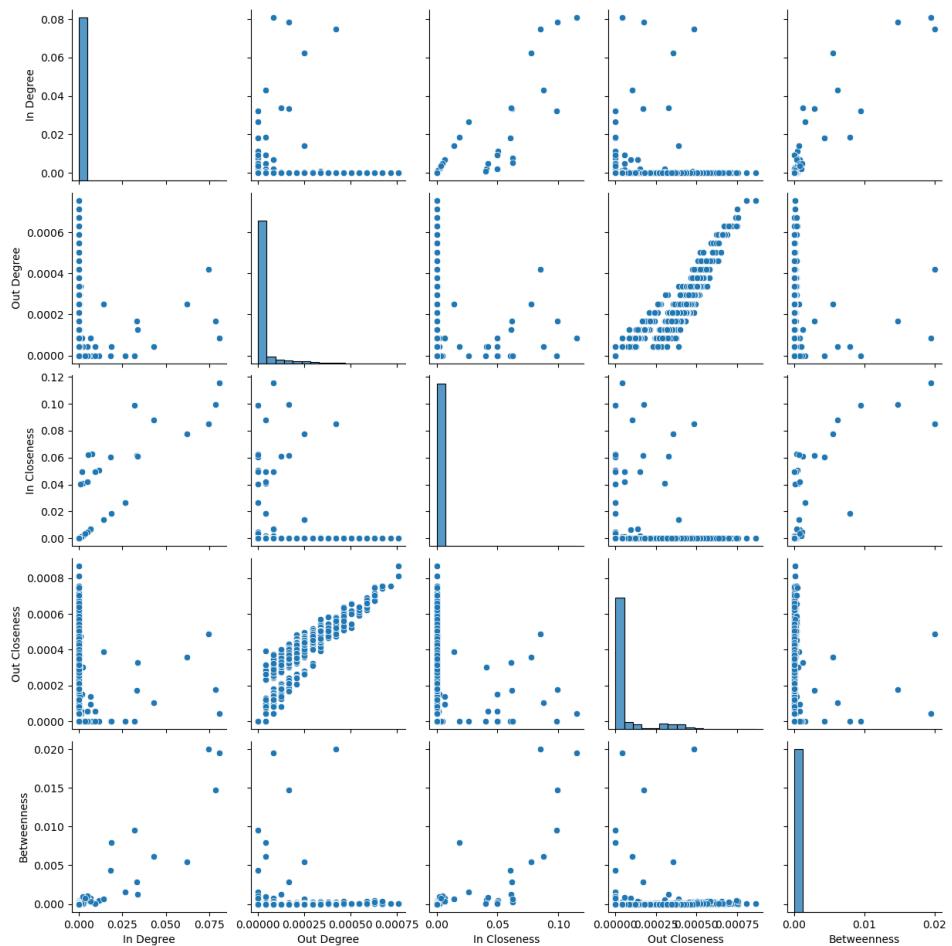


Figure 2: power law distribution of network structure

The description of the network's structure and underlying characteristics provides vital in-

sights into the nature of interactions and relationships among Twitter users. By comprehending these structural properties, researchers and practitioners can better understand how information, emotions, and influences flow within social media platforms, informing strategies for communication, marketing, or social analysis. Moreover, the network's decentralized and fragmented nature might lead to further interesting inquiries into individual user behavior and the mechanisms that drive connections in such a dispersed digital environment.

In the analyzed network, the compound sentiment score serves as a continuous measure of emotions, ranging from -1 (representing negativity) to 1 (indicating positivity). Interestingly, the study found that the compound modes across the network were predominantly negative. This could be visualized using a color scheme where red represents negative sentiments, and blue symbolizes positive emotions. The predominance of negative sentiments within the network may reflect underlying trends or specific events during the analyzed period, providing a snapshot of the collective mood within the Twitter community.

Further examination of the network revealed a lack of necessary correlation between the centrality degree of a user (a measure of how central or influential a node is within the network) and the negative compound mode (a measure of the most common negative sentiment for a user). This observation contrasts with an intuitive expectation that more central or influential users might exhibit particular emotional trends. Instead, the emotions appear to spread across the network in a more complex and nuanced manner, with no straightforward relationship between a user's centrality and their predominant sentiment. This complexity underscores the multifaceted nature of sentiment propagation and suggests that simple measures of influence or connectivity do not fully capture the dynamics of emotion cascades.

The analysis of Emotion Cascade provides profound insights into the interplay of emotions within a social network. It goes beyond mere quantification of sentiment scores, unveiling the intricacies of how emotions permeate through connections and influence the overall network's mood. Recognizing that sentiments are not necessarily tied to conventional network measures like centrality adds a layer of complexity to understanding human interactions in digital environments. This could have important implications for studies on social influence, information dissemination, and even mental well-being within online communities, highlighting the need for multifaceted approaches to explore the rich landscape of digital emotions.

The detailed information provided above captures the concept of "Reality Synchronicity" across four distinct time periods (T1 to T4) related to the social and political unrest in Hong Kong. This concept examines how events in the reality correlate with sentiments and network structures observed in the social media landscape. The detailed analysis of these four time periods can be structured into four academic-style paragraphs as follows:

During the initial period, the network graph revealed several hubs forming around key figures such as Senatemajldr, SolomonYue, Stand_with_HK, and joshhuawongcf. Interestingly, these hubs were predominantly silent (depicted as black) and exhibited a tendency toward negative emotions (depicted as red). Real-world events during this period, including sit-in protests, arrests, violence, and legal challenges, seemed to correlate with the observed negative sentiment.

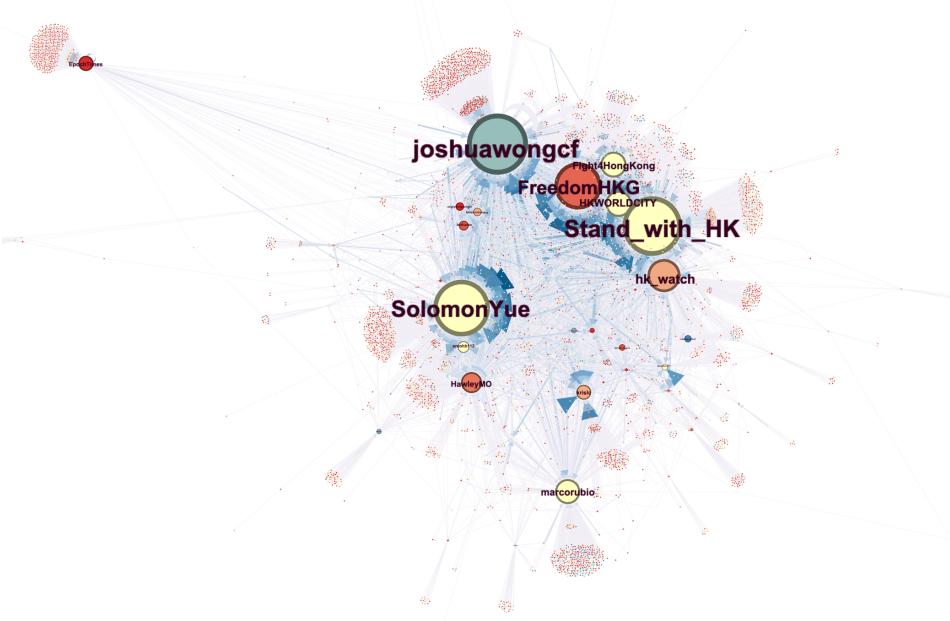


Figure 3: emotion cascade

The silence of the hubs could reflect a broader hesitancy or uncertainty within the network as tensions began to escalate.

As the unrest continued into the next time window, the network displayed a mixed but increasingly negative emotional state. New hubs such as FreedomHKG, Andyhanhotin, and Fight4Hongkong appeared, although with no clear activity. The real-world death of Chow Tsz-lok and subsequent protests and clashes with police possibly fueled the intensified negative sentiment. The emergence of new hubs may indicate the spread of information and mobilization around specific.

In the third period, emotions within the network became almost completely negative. Some hubs from the initial period reappeared, such as senateMajldr and SolomonYue, yet without clear community formation. Intriguingly, these hubs were more positive than the peripheral nodes. Real-world events, including consecutive days of strikes and violent suppression by police, likely contributed to the pervasive negativity. The positive sentiment within the hubs may reflect strategic communication or a divergence in perspective from the broader network.

The final period exhibited the emergence of new hubs like HKWORLDCITY, a calming of emotions compared to T3, and a clearer community structure, delineated by targeting international and domestic representatives. Real-world events such as the Hong Kong Pride Parade, PLA's public appearance, and nighttime solidarity protests may have contributed to these dynamics. The more structured community formation and nuanced emotional trends may signal a maturation or evolution in the network's response to unfolding events.

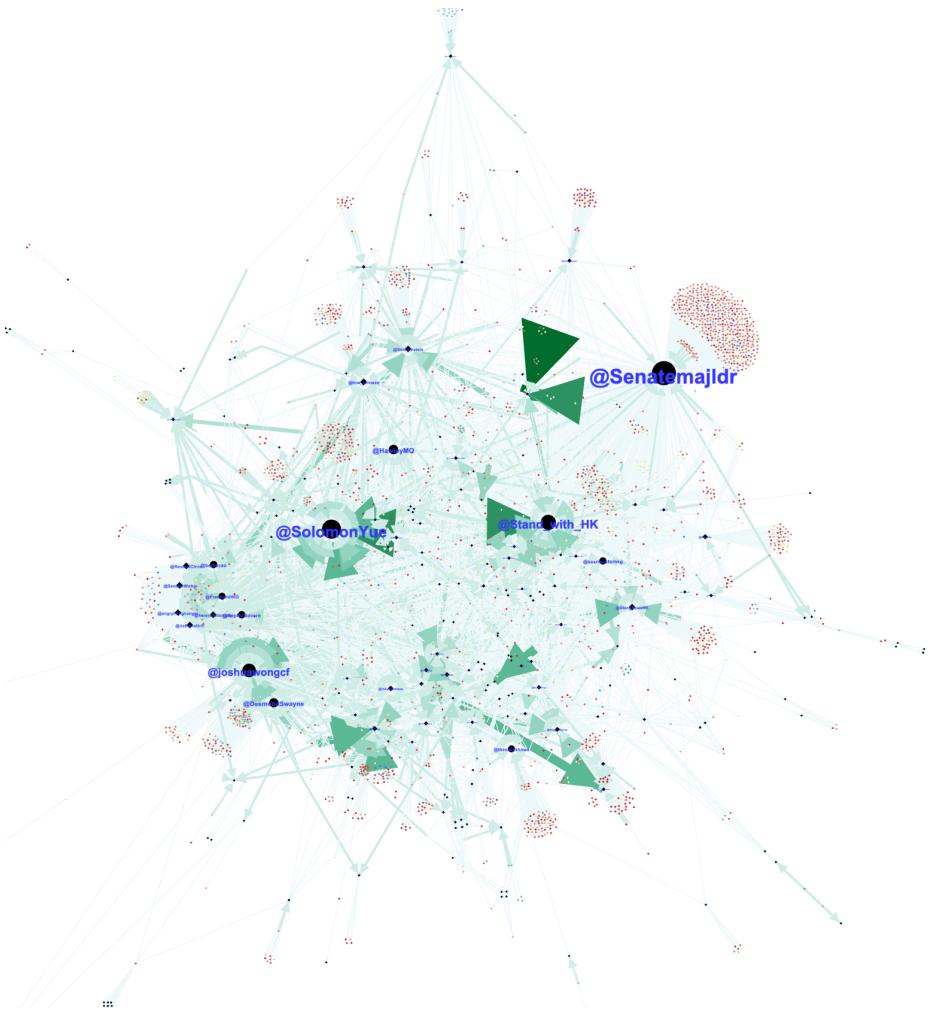


Figure 4: T1: 2019-11-03 - 2019-11-06

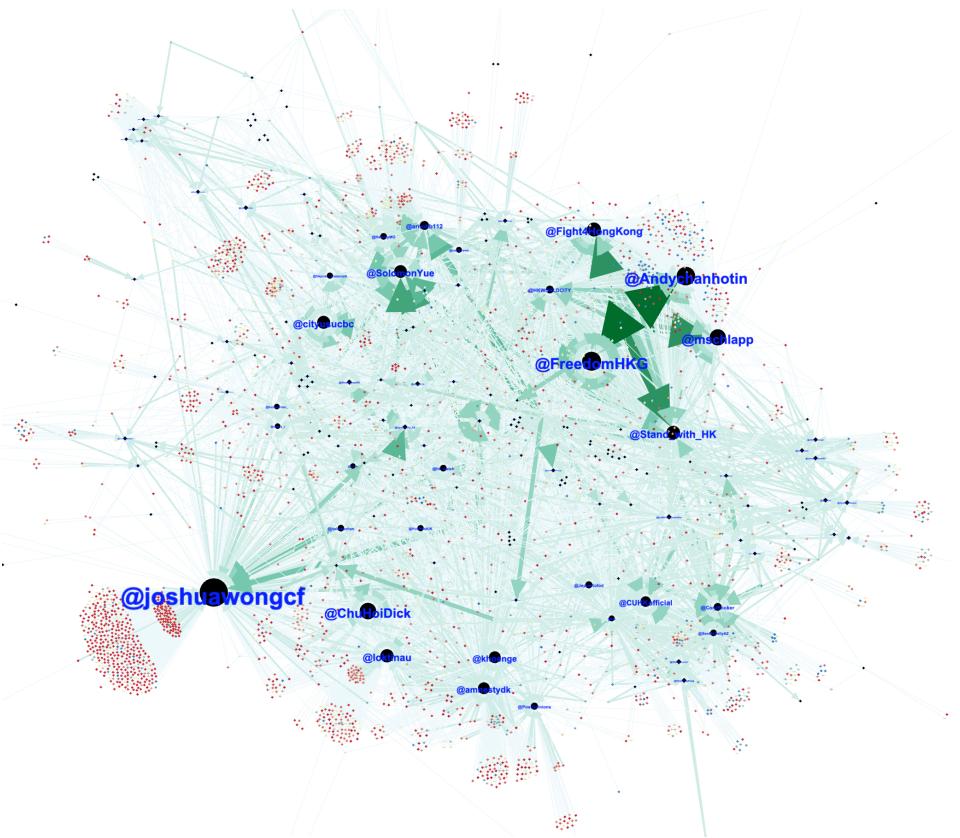


Figure 5: T2: 2019-11-07 - 2019-11-10

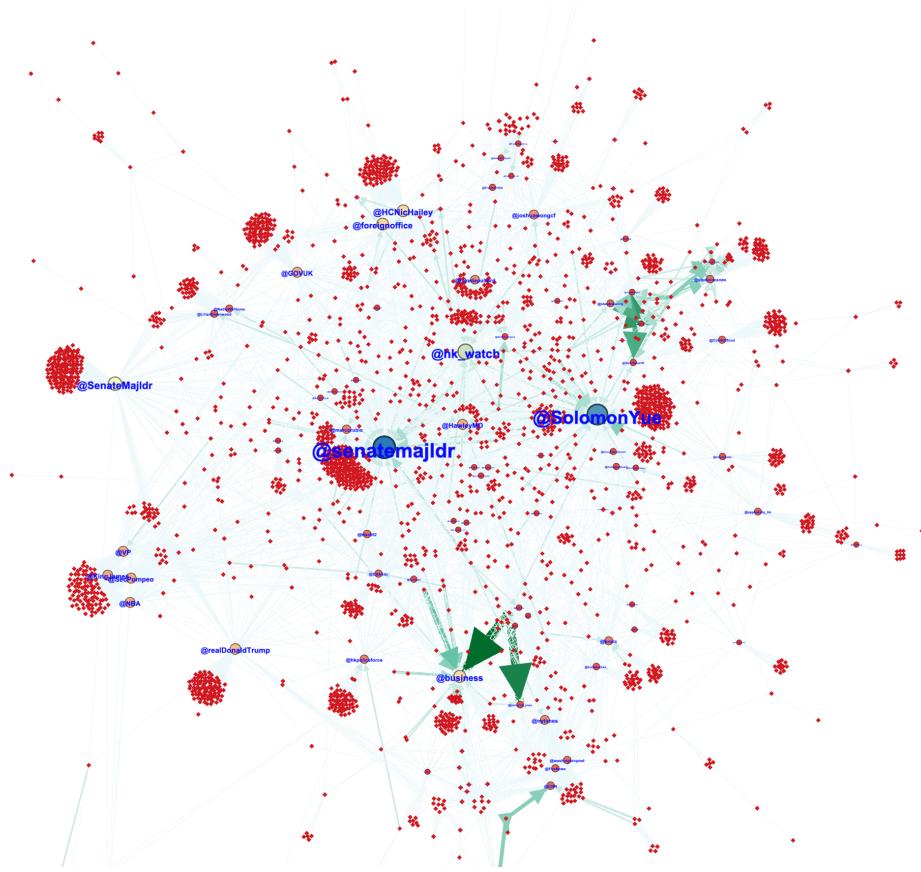


Figure 6: T3: 2019-11-11 - 2019-11-14

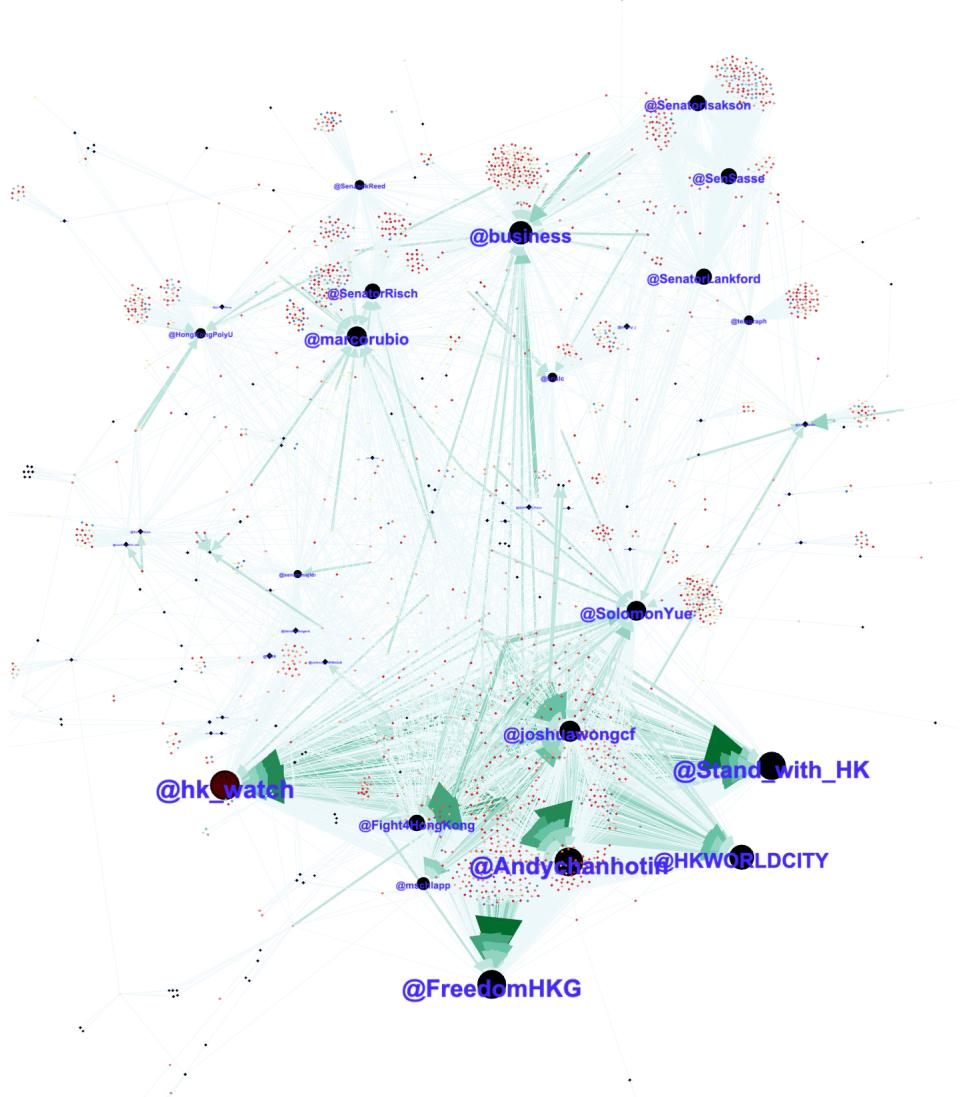


Figure 7: T4: 2019-11-15 - 2019-11-19

In summary, the concept of “Reality Synchronicity” provides a compelling lens to understand how real-world events and social media landscapes intertwine. Through a detailed temporal analysis, correlations between on-the-ground occurrences and online sentiments and structures become apparent. This approach underscores the intricate interplay between digital communication and real-world phenomena, offering invaluable insights for sociopolitical analysis, crisis management, and public engagement strategies.

6 Conclusions

The comprehensive analysis of the social media landscape in relation to the Hong Kong unrest during November 2019 unveils a multifaceted understanding of digital human interactions and their resonance with real-world phenomena. The network structure’s fabrication appears fairly decentralized, lacking strong connections or community structures, with the hubs exhibiting instability. Nonetheless, the pattern fitting a power-law distribution suggests underlying regularities. These hubs, often revolving around central figures, did not merely serve as focal points but acted as conduits for press demands, possibly steering or reflecting public sentiment.

The analysis also uncovers intriguing dimensions of emotion within the network. Although a clear correlation between centrality degree and negative compound mode was not identified, the central nodes displayed a tendency to be more positive than peripheral ones. This pattern highlights the nuanced role of influential figures in shaping the emotional landscape, suggesting a need for prudence and discernment in their communications.

Furthermore, the study reveals an evolving identity within the network, with hubs serving as symbolic entities around which identities coalesce. This identity evolution provides insights into the social dynamics and communal alignment that may underpin large-scale social movements. The concept of Reality Synchronicity further enriches the understanding, demonstrating a complex interplay between the network’s dynamics, including the appearance of hubs, and the corresponding reality movement. The emotional states within the network emerge as both a cause and consequence of real-world events, forging a symbiotic relationship between online expressions and offline occurrences.

The findings of this analysis render avenues for future work, such as employing hashtags as tokens of identity, to delve deeper into the mechanisms underlying social mobilization and digital engagement. Overall, this study underscores the intricate fabric of online social networks and their profound alignment with real-world events. By illuminating the multifaceted interactions and emotional undercurrents within the digital sphere, the study contributes valuable perspectives for sociopolitical analysis, crisis management, and strategic communication, reaffirming the potential of social media data as a vital tool in contemporary research.

6.1 References

- Alvarez, R., D. Garcia, Y. Moreno, and F. Schweitzer. 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*, 115–55. Sage Publications, Inc.
- Leow. 2019. “Scraping Tweets with Tweepy & Python.” Plain English. 2019. <https://plainenglish.io/scraping-tweets-with-tweepy-python-59413046c788>.
- Monterde, A., A. Calleja-López, M. Aguilera, X. E. Barandiaran, and J. Postill. 2015. “Multitudinous Identities: A Qualitative and Network Analysis of the 15M Collective Identity.” *Information, Communication & Society* 18 (8): 930–50.
- Peña-López, I., M. Congosto, and P. Aragón. 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.
- Swann, T. 2020. *Anarchist Cybernetics: Control and Communication in Radical Politics*. Policy Press.