



香港城市大學
City University of Hong Kong

Blending Topic Modelling and Social Network Analysis

Big Data Analysis of the Hong Kong Protests: A Preliminary Report

Edmund W. CHENG, Chris Fei SHEN, Richard M. WALKER
Dept of Public Policy, Dept of Communication, College of Liberal Arts and Social Sciences

Modern day protests

- Digital communication technologies have played a critical role in modern-day protests around the globe in the last decade
- Resulted in “networked social movements”.
- Key difference to prior movements: no or limited formal leadership
- Key characteristics:
 - multimodal information sources (ensuring connectiveness),
 - digitally connected horizontal networks (providing channels for mobilization),
 - deliberative mechanisms (articulating what to fight for), and
 - personalized frames (allowing participants to make sense of their involvement).



The Hong Kong protests

- The Hong Kong protests extensively used social media platforms: Ideal for contentious politics analysis using big data
- In this presentation we ask
 - Who are the active users and how influential they are?
 - What are the trend and pattern of mobilization?
 - To what extent are the protests organized, decentred or leaderful?
- Our preliminary analysis points towards
 - Top users are not as influential as perceived
 - The movement evolves through three phases of mobilization
 - The movement is in between leaderless and organized actions



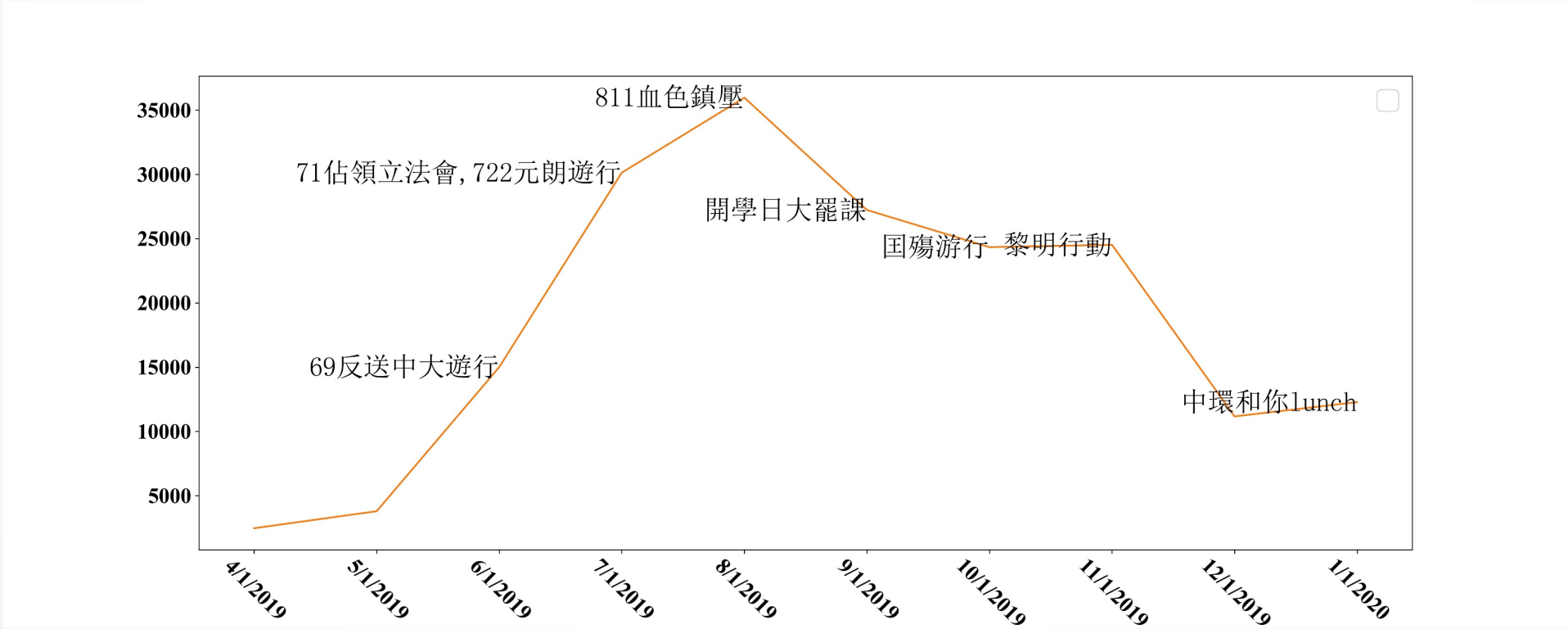
Data and Methods

- Big data scrapping from LIHKG.com:
 - > 20 million threads/comments
 - ~44 weeks from 1 April 2019 to 31 January 2020
- Our analysis includes:
 - Descriptive data of frequency counts of posts/replies and hyper-posters over time
 - Descriptive data on the frequency of key sentiments over time
 - Topic modeling of the content of posts
 - Social network analysis of key relationships between members of the platform



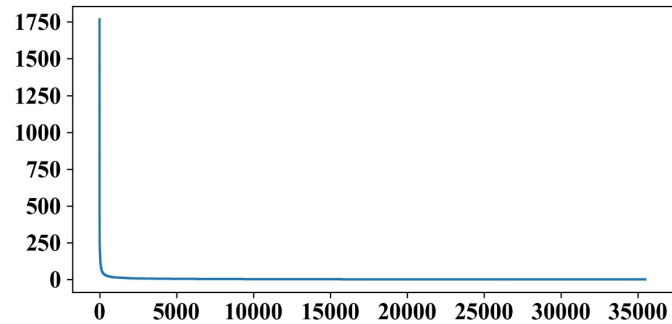
How active are users in general?

Average daily text (post/reply) temporal distribution

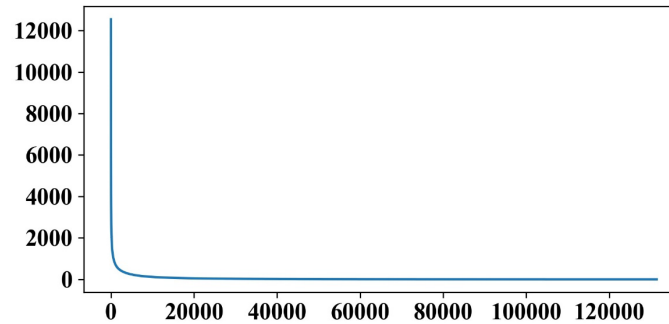


Do users contribute equally?

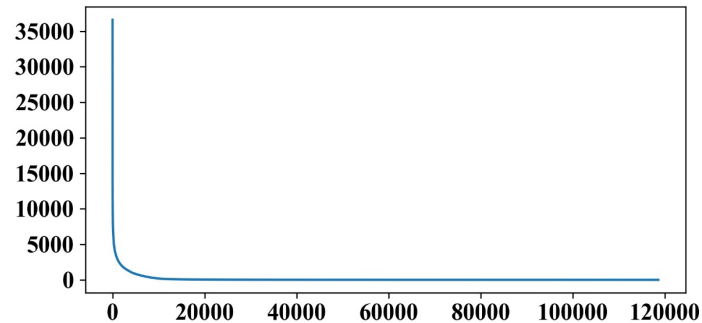
Distribution of post frequency



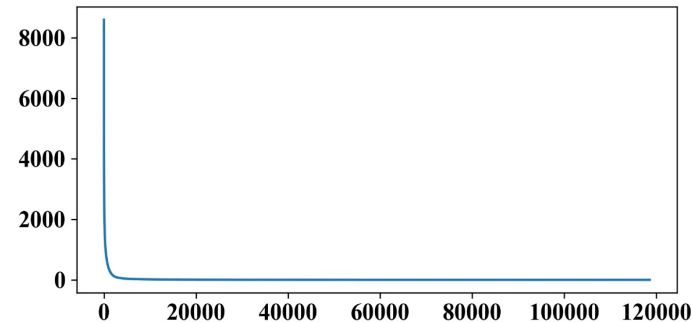
Distribution of reply frequency



Distribution of likes



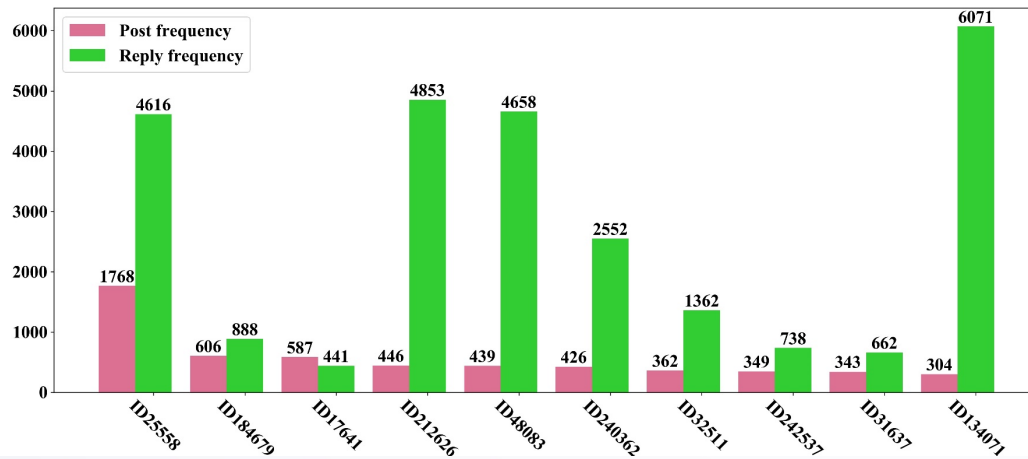
Distribution of dislikes



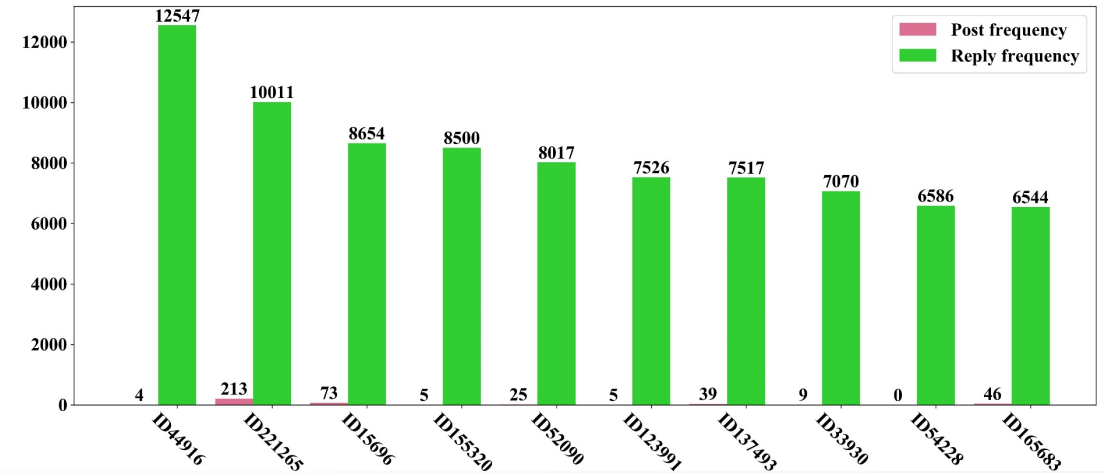
- Power law distribution: only an extremely small portion of users contributes a lot; only a very small portion of posts received lots of likes/dislikes

Who are hyper-active users?

Top10 frequent posters



Top10 frequent repliers



- Most of the top posters are frequent repliers, but the top frequent repliers did not post frequently – suggesting that they are two groups of individuals: 1) providing information 2) interacting and connecting with others

Topics of hyper-active users

Top10 frequent posters



Top10 frequent repliers



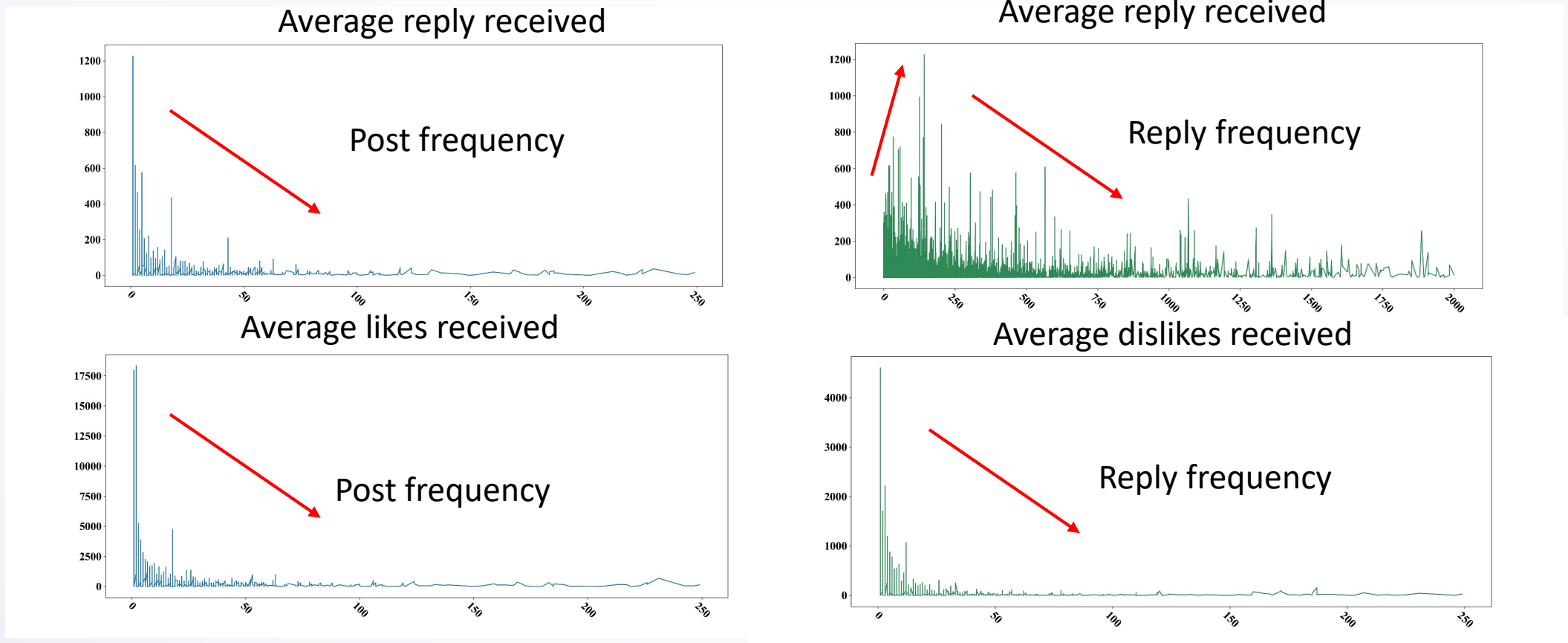
- The top frequent posters mentioned keywords related to governmental entities; the top frequent repliers were more mobilization oriented: keywords related to action and location.

Who are top mobilizers?

Rank	ID	Frequency	Name	Sample
1	54228 (Post rank 9)	566	柚木普	10.12 反緊急法遊行 時間: 3:00pm 地點: 尖沙咀鐘樓至深水埗運動場
2	25558 (Post rank 1)	368	LIKLN	稅局聖誕激光燈飾之夜 19:00「星火不息 燃點國際」集會 日期: 2019年12月23日(一) 時間: 晚上七時至九時 地點: 中環愛丁堡廣場 主辦方: 張崑陽、候任中西區區議員梁晃維、民間集會團隊已獲警方發出不反對通知書
3	205482	216	奶絲打對腳	[2百萬人再集結] 915民陣大遊行 日期: 2019年9月15日(日) 時間: 下午3:00 起步 集合地點: 銅鑼灣東角道 遊行路線: 從銅鑼灣東角道起步, 遊行至遮打道行人專用區
4	123991 (Reply rank 6)	196	消息. 睇真啲!	出黎呀, 自由闊! 不闖入、不塗鴉、不鐳射 如有發現, 批准大家往死裡打明日之公眾集會(遮打花園)及公眾遊行(前往美領)已獲發不反對通知書。集會部分 地點: 中環遮打花園 時間: 13:30 - 18:30.....
5	123713	159	帝吧小王子	913中秋要人!! 公海 https://t.me/joinchat/NBHSURkFTtbziGhzabb2g 各區集合地點! 油麻地、佐敦、尖沙咀: 梳士巴利公園(太空館附近) 荃灣、大窩口、青衣: 荃灣公園(海濱長廊)、青衣海濱長廊 葵興、葵芳、荔景: 林士德遊樂場、葵涌運動場 美孚: 嶺南之風.....

- Mobilization content being operationally defined as posts contain location and action related keywords.
- We found the top mobilizers overlap with top posters and repliers.

How influential are the hyper-active users?



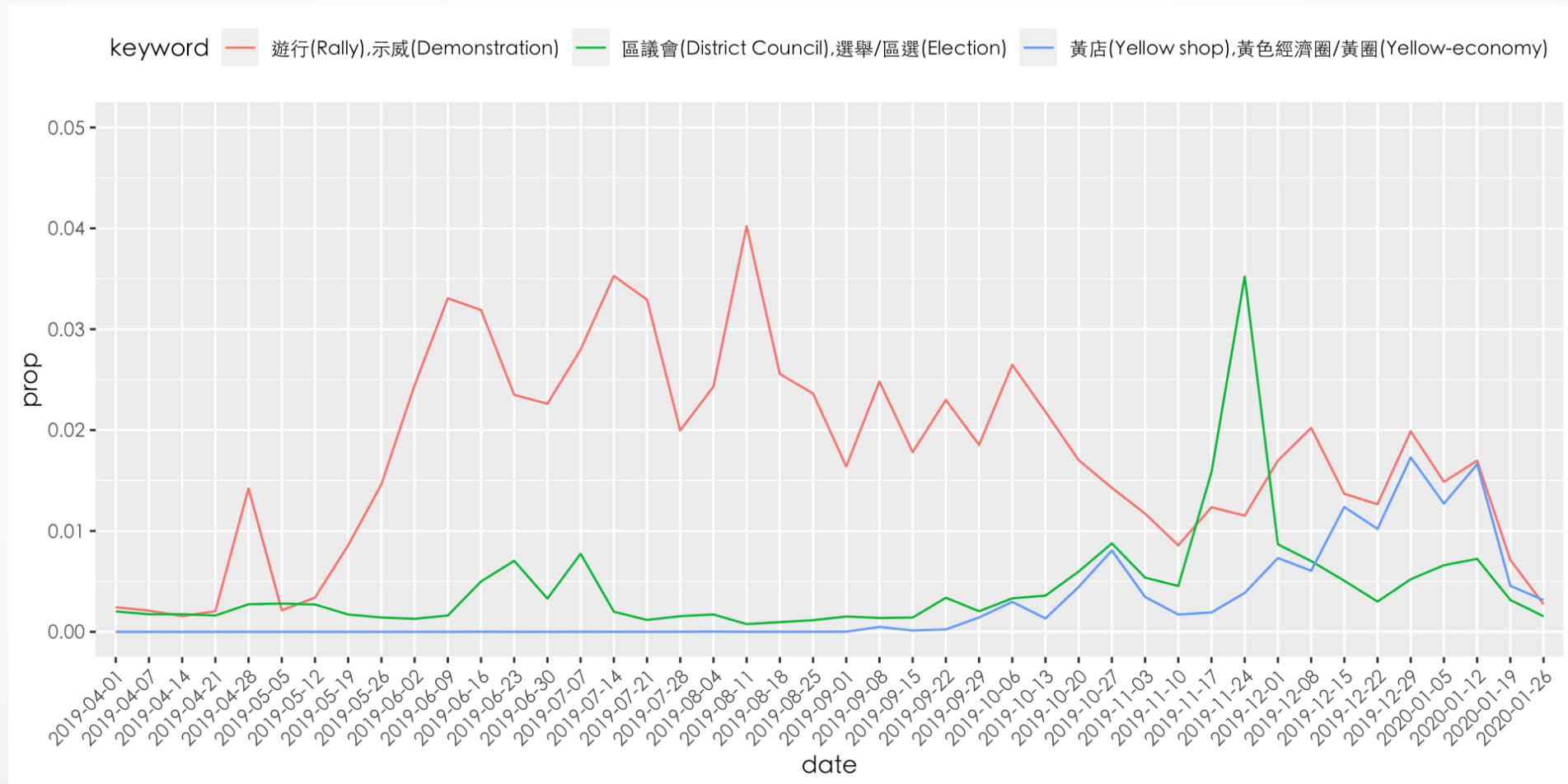
- Not really. The more you posted, the less attention you got.
- For reply: moderate amount of repliers received the most attention from others.

Topic modelling: what are patterns of mobilization ?

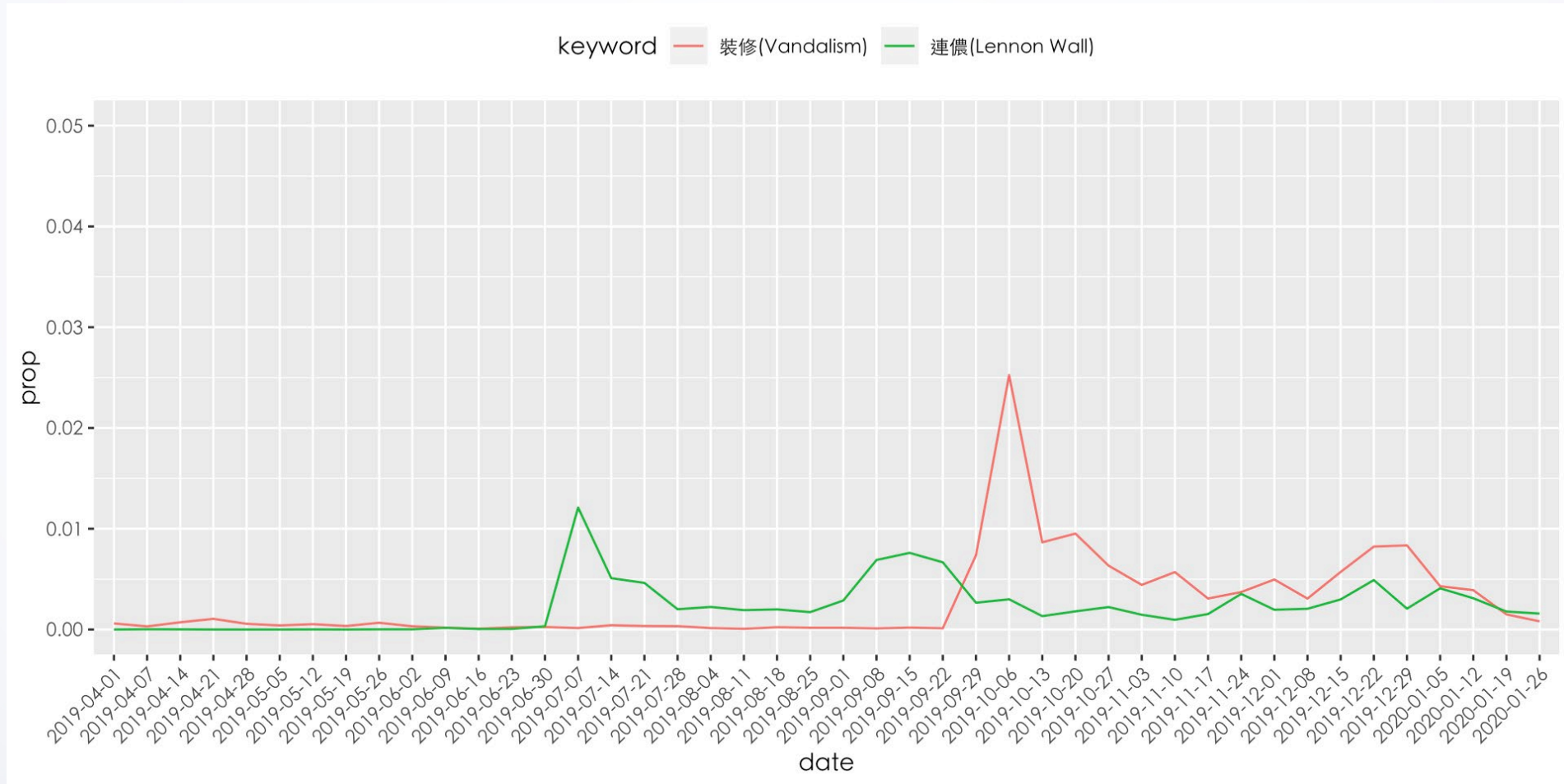
- A weekly frequency of keywords
- The y-axis represents the proportion of comments of that keyword in a particular week.
- The overall online activities or sentiments are highly correlated to street protests.
- The actions have become diversified, whereas the forms and claims have become radicalised.
 - Action: from **protest activities** (rallies/demonstrations) to political consumption (yellow-shop/yellow-economy) and **institutional politics** (district council/elections)
 - Form: from **advocacy actions** (posting on Lennon walls) to **radical actions** (advocating, criticising or justifying vandalism)
 - Claim: from achieving **consensual demands** (five demands) to **transgressive agenda** (co-burning)



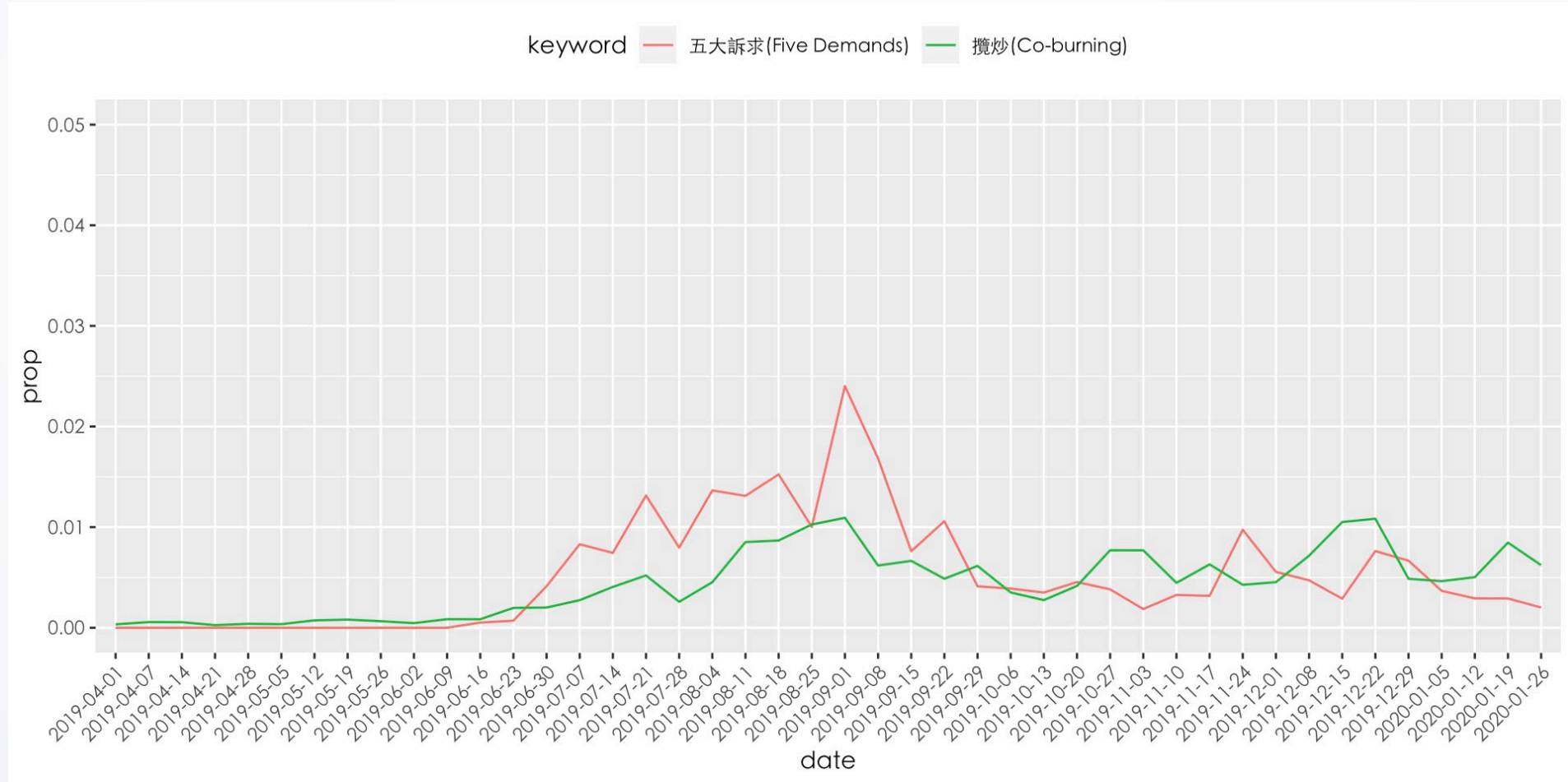
Actions over time



Forms over time



Claims over time



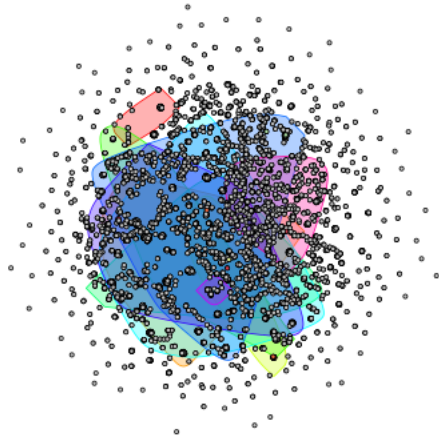
Network analysis: what are the communities and how connective they are?

- Weekly network graphs based on the comments of all threads on chitchat and current affairs sections.
 - **Grey node** represents a user
 - **Tied grey circle** refers to nodes with strong connections overlapped
 - **Coloured polygon** shows the weekly top 10 communities
- The communities were
 - Very scattered before the movement as shown by the swarm of nodes
 - Nodes became compacted with cross-groups interactions when the movement unfolded
 - Became fragmented and connected via a few clusters



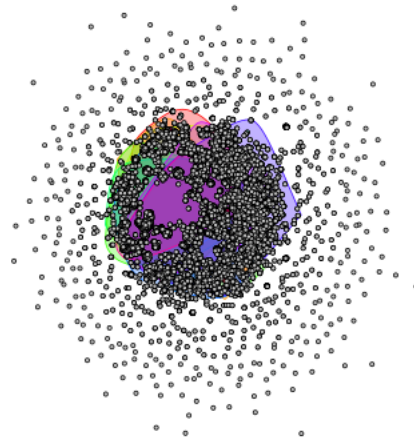
Scattered, compacted and fragmented networks

Week 7 (2019-05-12–2019-05-18)



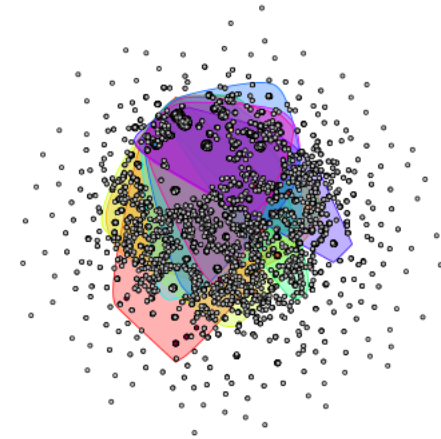
No. of users: 10276
Proportion of users in top 10 communities: 0.02491

Week 17 (2019-07-21–2019-07-27)



No. of users: 31314
Proportion of users in top 10 communities: 0.01207

Week 37 (2019-12-08–2019-12-14)



No. of users: 16875
Proportion of users in top 10 communities: 0.0227

Path to Radicalization



Phrase 1: Mobilize citizens to join petitions and rallies (advocacy and mobilization; from April till late July)



Phrase 2: Restrain protesters through protocols (apologies and human chains after critical incidents; till early September)



Phrase 3: Justify radical actions retrospectively (vandalism and other radical actions; from mid to late September)

Long term policy and social impact?

- To contrast the coverage of the Hong Kong protests in traditional news sources, social media, and academic writing on contentious politics, organizational science, and communication studies
- To investigate Hong Kong protesters' social networks to understand the new forms of protest mobilizing structures in Hong Kong
- To generate big data that can effectively map public opinion changes and that is conducive to policy planning and analysis in Hong Kong, and make the big data open access for use by others
- To contribute to the theoretical discussion of governance and trust and their connections with information sources, participation in protest movements, and the polarization of opinion. And to publish findings in top journals in communication, computational linguistics, data science, human behavior, and political science

