

Profiling Flash Mob Organizers in Web Discussion Forums

Vivien Chan, Kam-Pui Chow, Raymond Chan

► **To cite this version:**

Vivien Chan, Kam-Pui Chow, Raymond Chan. Profiling Flash Mob Organizers in Web Discussion Forums. 12th IFIP International Conference on Digital Forensics (DF), Jan 2016, New Delhi, India. pp.281-293, 10.1007/978-3-319-46279-0_14. hal-01758678

HAL Id: hal-01758678

<https://hal.inria.fr/hal-01758678>

Submitted on 4 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Chapter 14

PROFILING FLASH MOB ORGANIZERS IN WEB DISCUSSION FORUMS

Vivien Chan, Kam-Pui Chow and Raymond Chan

Abstract The flash mob phenomenon has been well studied in sociology and other disciplines, but not in the area of digital forensics. Flash mobs sometimes become violent, perpetrate criminal acts and pose threats to public safety. For these reasons, understanding flash mob activities and identifying flash mob organizers are important tasks for law enforcement. This chapter presents a technique for extracting key online behavioral attributes from a popular Hong Kong discussion forum to identify topic authors – potential flash mob organizers – who belong to a vocal minority, but have the motivation and the ability to exert significant social influence in the discussion forum. The results suggest that, when attempting to interpret flash mob phenomena, it is important to consider the online behavioral attributes of different types of forum users, instead of merely using aggregated or mean behavioral data.

Keywords: Criminal profiling, discussion forums, flash mob organizers

1. Introduction

A flash mob is a sudden public gathering at which people perform unusual or seemingly random acts and then quickly disperse; it is typically organized by leveraging the Internet and/or social media [9]. Although some researchers believe that it is inappropriate to use the term to describe political protests or criminal acts organized via the Internet or social media, it is clear that the organization of flash mobs for political protests and criminal purposes is an increasing trend [6].

The Umbrella Movement in Hong Kong, which was launched in late September 2014, leveraged several social media platforms to create popular support and motivate citizens to participate in many street demonstrations against the Hong Kong Government [10, 17]. After the Umbrella Movement ended in December 2014, the street demonstrations

in Hong Kong persisted, but in a very different form. Specifically, a number of flash-mob-like invocations were posted by organizers on popular social media platforms under innocuous topics such as “shopping,” “recover a district” and “anti-traders.” The term “shopping” actually refers to protests against government policies, “recover a district” refers to protests against mainland Chinese tourists and “anti-traders” refers to protests against Mainland Chinese citizens who take advantage of multiple entry visas to import goods from Hong Kong to Mainland China. These flash-mob-like activities are very different from traditional street protests. In many instances, flash mobs suddenly emerged in and around tourist shopping areas to disrupt business and traffic, and then dispersed. Clearly, such flash mobs could become aggressive and pose threats to social order [14].

This research focuses on understanding flash mob activities with the goal of identifying potential flash mob organizers in web forums before the mobs manifest themselves. Specific questions are: What are the key online behavioral attributes of flash mob organizers? How can these attributes be used to identify flash mob organizers at an early stage? How can the social influence of flash mob organizers be measured?

2. Related Work

Studies of flash mob phenomena have primarily been conducted by researchers in the areas of sociology and public health. Many of these studies report that the increasing use of social media enhance the potential that flash mobs will be created for criminal purposes and may therefore pose threats to public safety [6, 12, 13]. In the cyber domain, Al-Khateeb and Agarwal [1] have proposed a conceptual framework that uses hypergraphs to model the complex relations observed in deviant cyber flash mob social networks.

A search of the literature reveals that there are no studies related to classifying web discussion forum users based on their social influence. The classification of users in online discussion forums is important because it can help identify individuals who are influential at instigating flash-mob-like activities at an early stage. Having classified web discussion users into groups, it is useful to borrow the concept of social influence as used in other fields to compare the influencing power of the different groups.

The concept of social influence has been studied extensively in sociology, marketing and political science. In recent years, several researchers have focused on social influence in social media platforms such as Facebook and Twitter. The approaches for analyzing social influence

are broadly categorized as graph-based approaches and attribute-based approaches. Graph-based approaches model social networks as graphs and make use the HITS algorithm [7] or the Brin-Page PageRank algorithm [2] or their variants to measure social influence. Attribute-based approaches make use of attributes derived from social media networks, such as the number of followers, number of retweets, number of tweets per user, tweeting rate, etc., to measure social influence [3, 8]. However, at this time, no consensus approach exists for measuring the social influence of social network and discussion forum users.

3. Profiling Flash Mob Organizers

The most commonly used social media by flash-mob-like protest organizers in Hong Kong are Facebook and online discussion forums. According to the U.S. Census Bureau, there were four million active Hong Kong Facebook users in December 2014, corresponding to a penetration rate of 56.7%. Meanwhile, Hong Kong Golden is one of the most popular web discussion forums in Hong Kong. According to the Hong Kong Golden website [4], it had more than 170,000 registered users as of June 2015. The political discussions on Facebook and Hong Kong Golden are often very heated. Unfortunately, many of the conversations on Facebook take place in private pages or groups and are difficult to access for research purposes. However, the discussions on Hong Kong Golden (and other similar web forums) are mostly public. As a result, it was possible to tap into this public pool of data for the research effort.

3.1 Discussion Forum Dataset

This research collected data from the “Current Affairs” sub-category of the Hong Kong Golden discussion forum from January to April 2015. It is important to note that flash-mob-like protest activities took place nearly every weekend during this four-month period. The sub-category was selected because it was the most popular venue for Hong Kong users who wished to discuss political issues and most of the flash-mob-like protest announcements were posted on this sub-category. The following characteristics were observed in the Hong Kong Golden dataset:

- **User Characteristics:** Different users play different roles in the discussion forum. In general, a discussion forum user belongs to one of four groups:
 - **Post Author:** A post author only posts responses to other users and does not create topics. Most forum users are expected to be of this type.

- **Topic Author Only:** A topic author only creates topics and does not post responses to other users.
- **Topic Author Self-Responder:** A topic author self-responder only posts responses to self-created topics and does not post responses to topics created by other users.
- **Topic-Post Author:** A topic-post author creates new topics and posts responses to topics created by others. This group of individuals is of most interest in the study.

Note that a topic author refer to all types of topic authors, including, topic author only, topic author self-responder and topic-post author. The objective is to study topic authors who might initiate flash-mob-like protest activities in a discussion forum.

- **Topic Characteristics:** The Hong Kong Golden discussion forum employs a topic ranking scheme that places topics with higher ranks on the first page of a sub-category. The ranking of a topic is based on the number of new posts to the topic on a given day. A topic with a higher ranking has a higher chance of being read and a higher chance of having responses posted by forum users.
- **Post Characteristics:** Some topic authors exploit the topic ranking scheme by posting many empty posts or spam posts to topics created by themselves to enhance the rankings of the topics. Other post authors in the forum may also help push topics to higher rankings.

3.2 Key Behavioral Attributes

The primary online attributes of topic authors with regard to their influencing power in discussion forums are: (i) motivation; and (ii) ability. Table 1 defines the online attributes of topic authors. A flash mob organizer would be more motivated than other users in posting new topics and pushing the topics to higher rankings. In a study on well-being and civic engagement in offline settings for online discussion forums, Pendry and Salvatore [11] conclude that the strong identification of users with other forum users is a predictor of the offline engagement of users in the forum cause. Thus, it is important for a topic author to also have the ability to engage other users in discussions of the created topics.

3.3 Social Influence

As discussed in the previous section, graph-based and attribute-based approaches have been employed to measure social influence in social me-

Table 1. Attributes of topic authors extracted from a discussion forum.

Attribute	Definition	Description
Motivation 1	Number of topics created by a topic author	This shows that the topic author is engaging the community frequently
Motivation 2	Number of empty/spam posts created by a topic author	This shows that the topic author is attempting to increase the ranking of the topic
Ability 1	Number of non-empty/non-spam posts created by a topic author	This shows that the topic author is engaging in conversations with other post authors
Ability 2	Number of posts created by a topic author that are responded to by other post authors (including both empty/spam posts and non-empty/non-spam posts)	This shows that the topic author is attracting other forum users to discuss the topic

dia platforms. However, there is no consensus on an approach for measuring social influence. This research uses a simple, intuitive measure of social influence based on attributes extracted from the Hong Kong Golden online discussion forum.

While a large number of posts to topics created by a topic author implies the popularity of the topics, it does not imply greater social influence of the topic author because some of the posts could be self-created responses or empty or spam posts designed to enhance the ranking of the topics. On the other hand, if the topics created by a topic author receive many non-empty/non-spam posts from other post authors, then it is likely that the topics were well received or generated a lot of discussion in the online forum. Thus, the standard z-score, which is used as a measure, considers empty/spam posts as well as non-empty/non-spam posts by other post authors.

Assume that a topic author receives $n = r + s$ posts in response to his/her topics, where r is the number of non-empty/non-spam posts by other post authors and s is the number of empty/spam posts by other post authors. Assume that a random post author responds with non-empty/non-spam posts with probability $p = 0.5$ and empty/spam posts with probability $1 - p = 0.5$. Then, a random post author would post $n * p = n/2$ non-empty/non-spam posts with a standard deviation of:

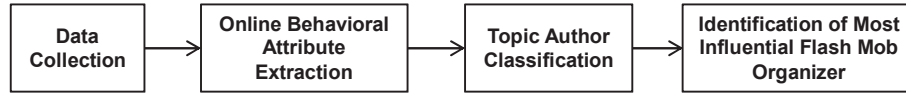


Figure 1. Profiling influential users.

$$\sqrt{n * p * (1 - p)} = \frac{\sqrt{n}}{2} \quad (1)$$

Thus, the z-score is given by:

$$z\text{-score} = \frac{r - \frac{n}{2}}{\frac{\sqrt{n}}{2}} = \frac{r - s}{\sqrt{r + s}} \quad (2)$$

Since it is common practice to use exponential rate parameterization, the logarithm of the z-score is used as the final social influence measure:

$$\text{Social Influence Index} = \log(z\text{-score}) \quad (3)$$

Note that the higher the social influence index value, the greater the influence of the topic author in the discussion forum.

3.4 Profiling Flash Mob Organizers

Empirical observations of discussion forum activities revealed that potential flash mob organizers usually work in groups as opposed to just one or two individuals. This makes it possible to identify a group of topic authors with similar online attributes. Thus, cluster analysis was applied to the dataset and the social influence index values of the clusters were compared.

Figure 1 presents the methodology for profiling potential flash mob organizers. Four steps are involved: (i) collection of data from the discussion forum; (ii) extraction of key online behavioral attributes from the dataset; (iii) classification of topic authors based on the behavioral attributes; and (iv) identification of the group of most influential users based on the social influence index.

4. Description of Experiments

A discussion forum has two main components: (i) topic variables; and (ii) post variables. Usually, a topic author creates a new topic and other forum users (called post authors) respond by replying to the topic through posts. However, unlike other social media platforms (e.g., Facebook and Twitter), friendship relations and/or follower information are not available. Therefore, the raw dataset was used for analysis.

Table 2. Data collected from January 2015 to April 2015.

Attribute	Jan	Feb	Mar	Apr
Number of topics	3,273	3,896	5,912	4,174
Number of posts	130,400	135,806	199,118	164,333
Number of topic authors	1,026	1,154	1,506	1,161
Number of post authors	13,484	13,221	14,694	14,343

Table 2 and Figure 2 provide details about the raw dataset, which was constructed from January 2015 through April 2015. During the four-month period, a total of 17,255 topics and 629,657 posts were collected. These topics were created by 3,040 distinct topic authors.

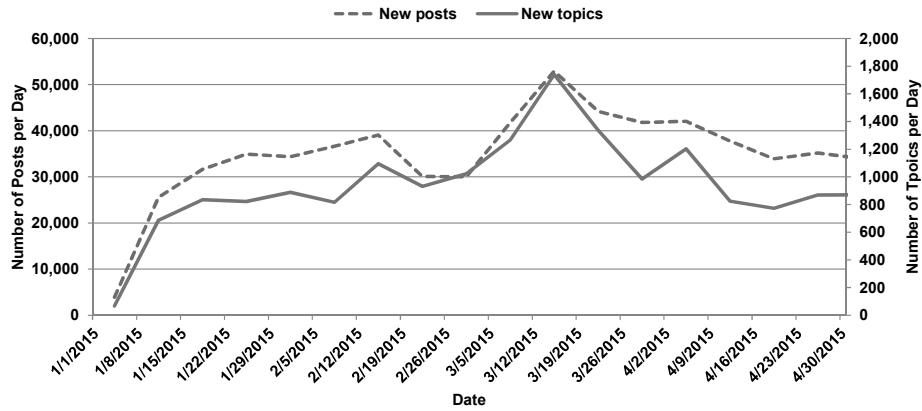


Figure 2. New topics and posts per week (January through April 2015).

In Figure 2, the drop in the number of posts on February 22, 2015 is probably due to the Lunar New Year holiday in Hong Kong while the spike on March 15, 2015 coincided with the most aggressive street protests that occurred during the four-month period. Of the 3,040 unique topic authors, 65% were active topic authors for just one month, 19% were active for two months, 9% were active for three months and only 7% were active during the entire four-month data collection period. The majority of the topic authors (87%) were also post authors who responded to topics created by other topic authors. A minority of topic authors (5%) only created topics and did not respond to topics created by other users. The remaining topic authors (8%) only responded to their own topics.

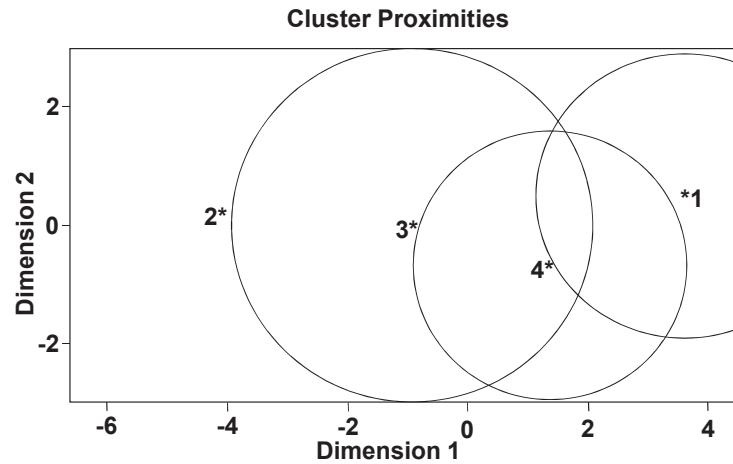


Figure 3. Cluster analysis results.

4.1 Classification of Topic Authors

In order to identify the most influential individuals in the discussion forum, an attempt was made to classify the topic authors into different groups. The cluster analysis involved two stages. First, initial groupings were derived by performing hierarchical agglomerative clustering using Ward's method. Next, the optimum number of clusters was selected based on a cubic clustering criterion value, which was set to a threshold of 3.

The analysis yielded the four clusters shown in Figure 3. Table 3 shows the differences in the mean online behavioral characteristics for the four clusters.

The four clusters correspond to the following topic author types:

- **Type 1: Topic Author (Inactive-Silent-Majority):** This cluster contains the largest number of topic authors ($n = 1,552$; 51.1%) from the sample population. These individuals are the least motivated and have the lowest ability of the four types of topic authors. The majority of individuals in this group ($n = 1,391$; 90%) posted topics during just one month. Also, this was the group with the highest percentage of topic authors ($n = 290$; 17%) who never responded to topics created by others. In terms of the number of topics created, the individuals in this group created an average of one to two topics during the four-month period, indicating that they were relatively silent in the discussion forum.

Table 3. Cluster differences of online behavioral characteristics.

	Cluster Means (Raw Scores) ^a				
	Univariate F	Type 1 (n=1,552)	Type 2 (n=155)	Type 3 (n=457)	Type 4 (n=876)
Cluster-Defining Attributes					
Motivation 1	533.53*	1.26 ^b	58.19	8.42	2.75 ^b
Motivation 2	140.51*	0.71 ^b	141.24	16.58	1.22 ^b
Ability 1	479.02*	1.17	162.37	33.74	10.25
Ability 2	417.85*	16.85	1,853.03	286.78	115.32
Variables External to Cluster Solution					
Influence Index	809.78*	3.07	3.75	3.34	3.24
Mean No. Months	967.44*	1.11	3.52	2.47	1.63
Mean Time New Topics	22.76*	144h25 ^b	57h44	123h53 ^b	150h20 ^b

* p < .001
^aHigher mean value corresponds to higher motivation, ability or social influence
^bTukey HSD comparisons indicate mean scores are not significantly different at p < .05

- **Type 2: Topic Author (Active-Vocal-Minority):** This cluster contains the smallest number of topic authors ($n = 155$; 5.1%) from the sample population. These individuals, who are the most motivated topic authors with the highest ability, received an average of around 1,853 posts per topic created and self-responded with an average of 162 non-empty posts per topic. The majority of individuals in this group ($n = 106$; 68%) posted an average of 71 new topics continuously during the four-month period. The mean time between the creation of new topics was the shortest among the four groups (less than 58 hours); these individuals created a new topic every 2.5 days.
- **Type 3: Topic Author (Moderate-Active:)** This cluster ($n = 457$; 15.0%) is fairly similar to the Type 2 cluster and the individuals are ranked second in terms of motivation and ability as topic authors. The majority of individuals in this group ($n = 278$; 61%) were active as topic authors for two to three months.
- **Type 4: Topic Author (Moderate-Inactive):** This cluster contains the second largest number of topic authors ($n = 876$; 28.8%), after the Type 1 topic authors. No significant difference was observed between the motivation scores of the Type 4 and Type 1 topic authors. The only difference between the individuals in the two clusters is their ability to engage other post authors, with Type 4 individuals showing a better ability to engage other post authors than Type 1 individuals.

Table 4. Cluster differences of flash-mob-like topics.

	Type 1 (n=1,552)	Type 2 (n=155)	Type 3 (n=457)	Type 4 (n=876)
Number of flash-mob-like topics	199	1,076	408	211
Number of reply posts	29,477	334,279	154,705	111,196

4.2 Cluster Comparison

A one-way between-subjects ANOVA test was conducted to compare the means of two sets of variables, namely, the cluster-defining attributes and variables external to the cluster solution, for the four clusters. Note that a significant effect exists for all the attributes at the $p < .001$ level for the four clusters in Table 3.

Post hoc comparisons using the Tukey HSD indicate that the mean scores of all the cluster-defining attributes for the four clusters are significantly different. The mean scores of Motivation 1 ($M = 58.19$; $SD = 74.99$), Motivation 2 ($M = 141.24$; $SD = 364.10$), Ability 1 ($M = 162.37$; $SD = 218.21$) and Ability 2 ($M = 1,853.03$; $SD = 2,712.46$) for Type 2 topic authors (active-vocal-minority) are significantly higher than those in the other three clusters. The results suggest that the four clusters have significantly different mean scores for the key behavioral attributes.

For the set of variables external to cluster solution, an analysis of variance revealed that the social influence index is significantly different ($F(3, 3,036) = 809.78$; $p = .000$). Once again, the Type 2 topic authors (active-vocal-minority) have the highest mean social influence index ($M = 3.75$; $SD = 0.34$) while Type 1 topic authors (inactive-silent-majority) have the lowest mean social influence index ($M = 3.07$; $SD = 0.09$).

The final question is whether or not Type 2 topic authors are potentially organizers of flash-mob-like activities. To answer this question, a bag of words containing phrases related to calling for flash-mob-like demonstrations and the corresponding location names was created. The bag of words was used to identify flash-mob-like topics. Table 4 summarizes the results. Type 2 topic authors created the largest number of flash-mob-like topics (1,076), which accounted for 57% of the total. In addition, Type 2 topic authors received the largest number of responses (334,279), corresponding to 53% of the total.

5. Discussion

The results demonstrate that it is feasible to classify topic authors into four groups using the online behavioral attributes of motivation and ability. The minority Type 2 group of topic authors (active-vocal-minority; $n=155$; 5.1%) is actually the most vocal and influential group in the online discussion forum. This group of just 5.1% of topic authors produced the largest number of new topics in a short duration of time and were highly motivated in pushing their topics to higher rankings. Moreover, the topics created by the group were mostly related to flash-mob-like activities. At the same time, these topic authors received the largest number of responses among all the topic authors and were also actively involved in discussions with other forum users. In addition, this group has the highest social influence index, meaning that this minority group is not just vocal, but also influential in the online discussion forum. Based on their high social influence and creation of the largest number of flash-mob-like topics, it is highly likely that Type 2 topic authors (active-vocal-minority) are potential flash mob organizers.

This finding echoes the research results of Mustafaraj et al. [8] related to Twitter, according to which the behaviors of the vocal minority (users who tweet very often) and the silent majority (users who tweeted only once) were significantly different. They also pointed out that, “when the size of the minority opinion holding group increases more than 10% of the network size, then the minority opinion takes hold and becomes the opinion of the majority.”

6. Conclusions

Flash mobs have the potential to become violent, perpetrate criminal acts and pose threats to public safety. As a result, understanding flash mob activities and identifying flash mob organizers are important tasks for law enforcement. The technique for the early identification of potential flash mob organizers in discussion forums discerned four types of topic authors who have significantly different online behavioral attributes ranging from being part of a vocal minority to a being members of the silent majority. Based on the online behavioral attributes and an intuitive social influence index, potential flash mob organizers belong to a vocal minority, but have the motivation and the ability to exert significant social influence in a web discussion forum.

Future research will attempt to characterize the followers of potential flash mob organizers. Additionally, it will develop measures for discerning flash-mob-like activities based on web discussion forum topics.

References

- [1] S. Al-Khateeb and N. Agarwal, Developing a conceptual framework for modeling a deviant cyber flash mob: A socio-computational approach leveraging hypergraph constructs, *Journal of Digital Forensics, Security and Law*, vol. 9(2), pp. 113–128, 2014.
- [2] S. Brin and L. Page, The anatomy of a large-scale hypertextual web search engine, *Proceedings of the Seventh International Conference on the World Wide Web*, pp. 107–117, 1998.
- [3] M. Cha, H. Haddadi, F. Benevenuto and K. Gummadi, Measuring user influence in Twitter: The million follower fallacy, *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pp. 10–17, 2010.
- [4] HKGolden.com, Hong Kong Golden, Hong Kong, China (www.hkgolden.com), 2016.
- [5] A. Java, P. Kolari, T. Finin and T. Oates, Modeling the spread of influence in the blogosphere, *Proceedings of the Fifteenth International World Wide Web Conference*, pp. 407–416, 2006.
- [6] L. Kiltz, Flash mobs: The newest threat to local governments, *Public Management*, vol. 93(11), pp. 6–10, 2011.
- [7] J. Kleinberg, Authoritative sources in a hyperlinked environment, *Journal of the ACM*, vol. 46(5), pp. 604–632, 1999.
- [8] E. Mustafaraj, S. Finn, C. Whitlock and P. Metaxas, Vocal minority versus silent majority: Discovering the opinions of the long tail, *Proceedings of the Third IEEE International Conference on Social Computing*, pp. 103–110, 2011.
- [9] Oxford Dictionaries, Flash mob (www.oxforddictionaries.com/definition/english/flash-mob), 2016.
- [10] E. Parker, Social media and the Hong Kong protests, *The New Yorker*, October 1, 2014.
- [11] L. Pendry and J. Salvatore, Individual and social benefits of online discussion forums, *Computers in Human Behavior*, vol. 50, pp. 211–220, 2015.
- [12] G. Seivold (Ed.), The less amusing side of flash mobs, *Security Director's Report*, pp. 2–4, November 2011.
- [13] S. Solecki and K. Goldschmidt, Adolescents texting and twittering: The flash mob phenomena, *Journal of Pediatric Nursing*, vol. 26(2), pp. 167–169, 2011.
- [14] South China Morning Post, Parallel trading (www.scmp.com/topics/parallel-trading), 2016.

- [15] J. Tang, J. Sun, C. Wang and Z. Yang, Social influence analysis in large-scale networks, *Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 807–816, 2009.
- [16] M. Trusov, A. Bodapati and R. Bucklin, Determining influential users in Internet social networks, *Journal of Marketing Research*, vol. 47(4), pp. 643–658, 2010.
- [17] Wikipedia, 2014 Hong Kong protests (en.wikipedia.org/wiki/2014HongKongprotests), 2016.