

# Analyzing Collaborative Documentation of the Hong Kong Protests on Wikipedia

**Jackie Chan**  
Carleton College  
chanj3@carleton.edu

**James Gardner**  
Carleton College  
gardnerj2@carleton.edu

**Junyi Min**  
Carleton College  
minj@carleton.edu

**Kirby Mitchell**  
Carleton College  
mitchellk@carleton.edu

## ABSTRACT

Wikipedia is a powerful platform for collective memory of social movements, and this paper seeks to replicate existing research on editor behaviour in Wikipedia on an ongoing international social movement. We analyze the attention and collaborative documentation of the 2019-20 Hong Kong Protests on the English language version of Wikipedia. Our results indicate the presence of (1) intensified documentation and attention to Hong Kong Protest related articles, (2) suggest a lack of collaborative migration of editors between articles, (3) and found partial evidence that new events triggered dynamic re-appraisal of pre-existing articles. These findings corroborate with existing literature about the behaviours within editor communities of Wikipedia during breaking news events, but also complicate existing theories of collective memory on Wikipedia during a social movement, suggesting that the nature and motivations of the social movements influence the collective editing behaviour we see on Wikipedia.

## Author Keywords

Social movements, collaborative documentation, social computing, documentation patterns, Wikipedia, 2019 Hong Kong protests

## INTRODUCTION & MOTIVATION

This paper analyzes the 2019–20 Hong Kong protests based on prior work done by Twyman, Keegan, and Shaw in *Black Lives Matter in Wikipedia: Collaboration and Collective Memory around Online Social Movements* using three metrics: intensified documentation, collaborative migration, and dynamic reappraisal [28]. These metrics measure distinct behaviors found in collaborative efforts to document social movements on websites like Wikipedia. To start, intensified documentation examines the increase in efforts to document a social movement as it gains more attention over time. Collaborative migration looks at cooperation between editors when documenting a social movement and how the editing community behaves to maintain a collection of articles. Dynamic reappraisal analyzes how editors reassess existing records as new events and new information become known.

These three metrics can be summarized with their respective questions:

1. How does the attention towards documenting the 2019–20 Hong Kong protests change over time?
2. How do editors collaborate when documenting the 2019–20 Hong Kong protests as new events take place?
3. How do editors review existing articles after new events take place?

Information on Wikipedia is a form of collective memory where people have worked together to make sense of how to remember certain events. On account of Wikipedia being acknowledged as an authority of knowledge, these memories then become reified as part of a group’s identity. It is important to investigate the social systems of Wikipedia to understand how knowledge is formed within the community during an ongoing social movement. Research in this area can complement theories of collective memory and collective action [28, 11, 7]. As our research replicates Twyman et al’s paper with an emphasis on the 2019–20 Hong Kong protests, we wish to compare our results to further our understanding of how social computing plays into the aforementioned theories [28].

Our findings support intensified documentation, as we see more intensity in revisions as protests go on. However, we found little evidence for collaborative migration, as our scores for collaborations across articles is low in absolute terms. Lastly, our results found partial evidence that dynamic reappraisals occur as breaking news increases pageview traffic to past pages and to other Hong Kong protest articles, but investigations into other forms of reappraisal remain inconclusive. Our results as compared to Twyman et al’s paper show a less pronounced effect on collaborative migration and dynamic reappraisal, suggesting that the nature and motivations of the social movements influence the collective editing behaviour we see on Wikipedia.

## Background

### *Breaking News Coverage on Wikipedia*

Breaking news coverage has been generally characterized as having highly collaborative editorship and being highly attentive to evolving information in the early stages of article creation. This behavior was found during the analysis of news coverage of natural disasters and other ephemeral, yet sensational, events such as the Tōhoku catastrophes [11]. However, documentation efforts for these types of events fade rapidly.

As the situation stabilizes, less information is released and as a result fewer editors are required to maintain the accuracy of an article.

Patterns seen in the documenting of natural disasters have also occurred in the documenting of the 2019–20 Hong Kong protests, specifically with articles involving large scale protests and police attacks. However, a significant difference between natural disasters and protests is that the protests fit into a larger narrative - an ongoing social movement. It causes a spillover effect from the attention of the breaking news onto the social movement itself. Editors that collaborated on a protest article may take an interest to continue documenting related events for that movement. Therefore the attention brought by breaking news coverage may spillover to the movement itself, whereas attention on natural disasters will generally be more isolated.

#### *Information Seeking and Online Participation*

The Internet provides a low-cost way to connect with other members of a movement and also facilitates event coordination for large scale protests. This allows for greater participation from the community. The Internet’s accessibility also grants various communities the ability to overcome physical and cultural barriers [6]. Social media platforms, such as Facebook and Twitter, have been tools used to coordinate protests and communicate with organizers [24]. In contrast to those traditional social media platforms, Wikipedia editors follow strict guidelines that require information to be documented from a neutral point of view (NPOV)<sup>1</sup>, a constraint rarely found on Facebook or Twitter. To adhere to these NPOV guidelines, Wikipedia editors try to remove their own biases while contributing [17]. This NPOV documentation style appeals to those seeking near objective information to assess their stance and/or participation in a social movement.

In our analysis of the Hong Kong protests, we examined the amount of revisions to the articles in our corpus and also the number of editors that are participating in making revisions. By examining these two factors, we can determine whether or not there is a growing trend in participation in protest related documentation. Increased documentation also leads to more available information for members of the online community to examine when they are seeking unbiased information. Our study also examines the differences and similarities between online documentation in the Hong Kong protests and a domestic social movement such as Black Lives Matter. Similarities in the two could reveal whether or not the Internet’s ability to increase participation in online documentation applies to a movement that has a much larger set of geographical and cultural obstacles than found in a domestic movement.

#### *Collaborative Online Documentation*

Wikipedia editors have diverse motivations, contribution experience, and domain knowledge that all affect their abilities to contribute and which articles they edit. Given the nature of Wikipedia being public, there is no rigid or predefined organizational structure for collaborating on articles. Furthermore, despite the large number of editors on Wikipedia, it is unlikely

<sup>1</sup>[https://en.wikipedia.org/wiki/Wikipedia:Neutral\\_point\\_of\\_view](https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view)

that any given group of editors have interacted with one another, especially if they do not specialize in any particular domain [11].

In spite of these obstacles, editors utilize the talk pages provided by Wikipedia to communicate and to facilitate collaborative documentation. Editors self-evaluate their areas of expertise and find similar individuals to improve communication and collaboration. This also lets them focus their resources and attention on a smaller subset of the article rather than spreading themselves too thin on various parts [12]. The ability for many editors to communicate with one another allows them to create high quality articles even if individually they have low amounts of experience on the topic they are contributing to [17].

Our study does not include metrics for measuring article quality, but we do examine the collaboration among the editors in our article corpus. Since the 2019–20 Hong Kong protests is a social movement that has gained international attention, the protests likely drew a diverse collection of editors. In our examination of online collaborative documentation of this movement, we investigated the amount of collaboration found in our set of articles and any potential trends that they exhibit. We then determined the implications of these trends and whether or not they are intrinsic to the fact that this is an international social movement or if there are similar trends found in similar studies like the Black Lives Matter study [28].

#### **History of the 2019–20 Hong Kong protests**

For 156 years between 1841 and 1997, Hong Kong was a British colony before being handed over to China. Before Hong Kong was returned, agreements were made between the two countries to allow for Hong Kong to have a constitution called its Basic Law, which described its relationship with Mainland China now known as the “One country, two systems.” Omitted from this new dynamic between the two countries was the ability for Hong Kong to extradite criminals to Mainland China for prosecution due to the potential for abuse. This led China to resort to kidnapping individuals in Hong Kong to bring them back into Mainland China for prosecution [5, 30].

This serves as a background for a murder that occurred in Taiwan on February 2018 where Poon Hiu-wing was killed by her boyfriend who later returned to Hong Kong. As highlighted in the agreement between Hong Kong and China, because the murder happened in Taiwan, Hong Kong officials could not charge the individual with murder [27]. This killing was the primary motivation for Hong Kong legislators to propose amendments to their extradition laws known as the Fugitive Offenders amendment bill [10].

Despite having this incident as evidence for the amendments, concerns about China’s potential abuses still remained as China is known for regressive freedom of speech laws [29]. Some believe that this amendment served as a direct challenge to Hong Kong’s autonomy. It was this bill that led to what is known as the “2019–20 Hong Kong protests.” Most records date its beginning in June 2019 as millions gathered to show dissent to the amendment. The protesters listed five demands in the same month with the phrase “Five demands, not one

less [16].” The bill would see its official withdrawal later in October, but the protests still persist as not all demands have been met [15].

Both the Hong Kong protesters and Chinese government are experienced in using social media to their advantage to influence public opinion. A notable example of this occurred during the 2014 Umbrella Movement in Hong Kong where pro-democracy protesters used yellow umbrellas to symbolize their efforts, which later spread onto Facebook and Twitter. We saw similar instances to this in the 2019–20 Hong Kong protests as protesters circulated images of injuries incurred from police forces [26]. What we will be looking at is the documentation of this movement on Wikipedia as it attempts to provide a unique neutral perspective.

### DATA & METHODOLOGY

In this section, we will describe how we determined the analysis period, constructed our article corpus, pulled the revision histories, and retrieved pageview records. Additionally, we will explain why we decided to remove anonymous editors from our editor counts.

#### Analysis Period

Our analysis period spans from December 10th, 2009 to December 10th, 2019, but we focused our attention on the results between December 10th, 2018 and December 10th, 2019. To establish the start of the focus period, we looked for the start date of the 2019–20 Hong Kong protests. Although March 15th, 2019 is the start date denoted on the 2019–20 Hong Kong protests Wikipedia page, we wanted our study period to revolve around the date that the Wikipedia community began focusing on it. March 15th, 2019 only denotes the nine-person sit-in held for the extradition bill’s initial readings [19]. That leads us to June 10th, 2019, the creation date for the 2019–20 Hong Kong protests article, which is also a day after the Human Rights Front held an estimated one-quarter to one million person rally before the second reading of the extradition bill [14, 23]. That protest was then followed up by another rally three days later, with an estimated four hundred thousand person attendance [8, 22].

We wanted to start pulling data from Wikipedia on January 2020, so we thought that six months, from June 10th, 2019, to December 10th, 2019, would be sufficient to analyze the documentation patterns in the movement. To see the shift in documentation patterns from before our start date, we made December 10th, 2018 the starting point for our focus period. Thus, our main analysis period is from December 10th, 2018 to December 10th, 2019, giving us a full year to observe this social movement. Going forward, our analysis period will refer to this one year focus period.

#### Article Corpus

Our article corpus contains thirty-eight articles related to the 2019-20 Hong Kong protests. Our primary guideline for creating the article corpus was to find a representative set of articles that would capture a significant portion of the changes that happen throughout this movement. To accomplish this objective, we applied the snowball sampling method by selecting a



Figure 1. The related articles section found on the side of the 2019–20 Hong Kong protests Wikipedia article.



Figure 2. The related articles section found on the bottom of 2019–20 Hong Kong protests.

core set of articles first and then snowballing to include more articles related to the existing articles. Fortunately, on many of the 2019–20 Hong Kong protest related pages there are sections that editors have created that links to related pages. These sections, shown in Figure 1 and Figure 2, were helpful resources for establishing the first few articles in our article corpus.

With our initial corpus constructed directly from Figure 1 and Figure 2, we went on to find more pages that might also change due to the events of the 2019–20 Hong Kong protests. We read news articles covering the 2019–20 Hong Kong protests to learn about the political actors and organizations whose Wikipedia pages may change due to events caused by the protests. This led us to construct a timeline of events that helped summarize the significant moments that occurred throughout the years prior to the 2019–20 Hong Kong protests.

Due to the sociological nature of our research, there is no objective method to determine which articles are “related” to

the 2019–20 Hong Kong protests and which are not. We did our best to create a corpus that captured the broad activity that happened in this movement while also being strict enough to not add pages that only momentarily related to the 2019–20 Hong Kong movement.

Table 2 contains our article corpus, i.e. all the Wikipedia articles in our study of the 2019–20 Hong Kong protests. Note that these article titles may change after May 9, 2020. Not included in this table are all of the articles’ talk pages, which are pages that editors discuss potential changes that are going to be made. The talk pages for each respective article will be used in our analysis as well, but do not make up a significant portion of the data. Listed with the articles in Table 2 are the creation dates for each of the pages and how we categorized them internally. We believe this corpus satisfies the objective we set to accurately capture the changes that occur on Wikipedia when documenting the events in the 2019–20 Hong Kong protests.

Another feature of our corpus that we would like to mention is recency in which some of the articles were created. Twenty-six out of our thirty-eight articles were created in 2019, with the rest of the articles being created between 2001 and 2018. This aspect of our corpus becomes relevant in our results section, specifically in intensified documentation because most of the revisions are seen within the year 2019.

### Revision History

To pull the revision histories for each page, and its respective talk page, we used the MediaWiki API <sup>2,3</sup> and custom Python scripts to make calls to the API. All code and data mentioned in this section can be found in the codebase hosted on Github<sup>4</sup>. With the MediaWiki API, we retrieved revision and redirect records for all pages in our article corpus. There are some instances in the revision history where the page was known by a different name, and when the name was changed, the page is redirected to the most current name.<sup>5</sup> Although redirect records are important, we will be focusing on the data we have on revisions because articles that are redirected keep their revision history for their previous article title. For example, if article X was renamed to Y and article X redirects to article Y, then article Y’s revision history will still have the revisions made from article X. For more information, please refer to the MediaWiki API page documenting the effects of moving a page. <sup>6,7,8,9</sup>

After pulling this data in JSON format, we converted it to Pandas DataFrames and exported it into CSV files for further analysis. The structure for the CSV files containing the revision history of each page is described in Table 1.

<sup>2</sup>[https://www.mediawiki.org/wiki/API:Main\\_page](https://www.mediawiki.org/wiki/API:Main_page)

<sup>3</sup><https://en.wikipedia.org/w/api.php>

<sup>4</sup><https://github.com/gadeuneo/wiki-comps-2020>

<sup>5</sup><https://en.wikipedia.org/wiki/Wikipedia:Redirect>

<sup>6</sup><https://en.wikipedia.org/wiki/Wikipedia:Redirect>

<sup>7</sup>[https://www.mediawiki.org/wiki/Help:Moving\\_a\\_page](https://www.mediawiki.org/wiki/Help:Moving_a_page)

<sup>8</sup>[https://en.wikipedia.org/wiki/Wikipedia:Moving\\_a\\_page#Page\\_histories](https://en.wikipedia.org/wiki/Wikipedia:Moving_a_page#Page_histories)

<sup>9</sup>[https://en.wikipedia.org/wiki/Wikipedia:Redirect#Redirects\\_that\\_replace\\_previous\\_articles](https://en.wikipedia.org/wiki/Wikipedia:Redirect#Redirects_that_replace_previous_articles)

### Pageview Data

To supplement our revision data, we also gathered pageview data, a metric for the count of visitors to a particular Wikipedia article for our analysis in the corpus. This range is limited as the Wikimedia Foundation Labs (WMF Labs) tool, where we extracted our data from, only has data from July 1st, 2015. The WMF Labs tool also included data from now inaccessible pages due to page moves/redirects.<sup>10</sup>

Though pageviews alone may not be interesting, they tell us what articles are being read by readers. Reading articles itself is a form of interacting with Wikipedia and builds a sense of online community [2]. They may also encourage editors to improve the overall readability of a page as it gains attention and provides a learning environment. It also shows patterns of certain articles in our corpus and daily changes in attention, especially after the movement began.

### Removal of Anonymous Editors

We decided to limit our definition of editors to registered users, removing bots and anonymous editors from our count of editors. Although there are bots on Wikipedia that also make edits, they were excluded from our analysis as they mostly focus on minor grammar fixes.

Anonymous users are allowed to submit revisions to unprotected pages, publishing their IP address instead of a username and user ID. Although this serves as an incentive for users to register with Wikipedia, anonymous edits/editors still exist.

Anonymous edits make up around fifteen percent of all the edits we have recorded, however we have removed these anonymous editors from our analysis because there exist complications with counting unique IP addresses. IP addresses may change in a variety of situations which may cause us to overcount the number of anonymous editors. Additionally, multiple editors could have also submitted numerous edits under a singular IP address causing us to undercount. Regardless, we did not find an elegant method to count anonymous users accurately and so we have omitted them from our editor count as shown in Table 3.

## INTENSIFIED DOCUMENTATION

### Definition

Intensified documentation is simply the increase in documentation efforts and attention on articles in the corpus as the Hong Kong protest unravels. Increased documentation efforts can be seen in the form of increased revisions, increased editor counts, decreased article creation latency, and creation of new pages.

### Results

#### *Aggregate Analysis*

The figures referenced in this section either focus on the entire article corpus shown in Table 2 or a subset of those articles, e.g. the ten articles with the most revisions shown in Table 3.

<sup>10</sup><https://tools.wmflabs.org/redirectviews/?project=en.wikipedia.org&platform=all-access&agent=user&range=all-time&ort=views&direction=1&view=list&page=>



Column	Description
Revision ID	A unique ID number for the revisions.
Parent ID	The ID number for the previous revision ID of the page being revised.
User	Either the username of a registered Wikipedia account or IP address if anonymous.
User ID	The ID number for the registered Wikipedia account or 0 if anonymous.
Timestamp	Timestamp of revision in ISO 8601 format.
Revision Size	The size of the revision in bytes.
Anonymous	"True" if the user is anonymous.
Page ID	The ID number for the article being revised.

Table 1. The structure of the CSV file pulled from the MediaWiki API.

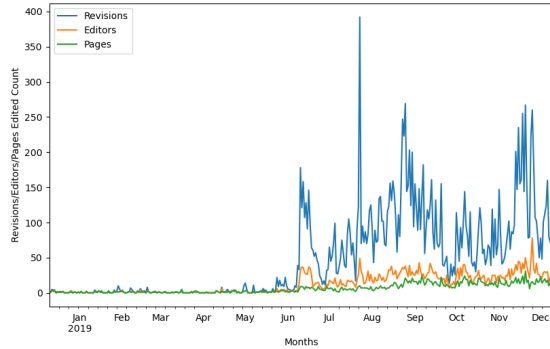


Figure 3. Display the number of revisions, editors, and pages edited per day for our article corpus, including talk pages, during the last year of our analysis period (December 10th, 2018 to December 10th 2019).

Table 3 serves as an overview of the data we collected from the thirty-eight articles and their respective talk pages. A significant portion of our article corpus can be represented by just the ten articles with the most revisions as Table 3 lists. These ten articles, which we will refer to as our “top 10 articles,” represent 72.6 percent of all revisions, 81.1 percent of unique editors, and 93.6 percent of all pageviews. It is no surprise that the 2019–20 Hong Kong protests article ranks first because it serves as the main page to the entire movement.

By combining the revision histories of all articles in our corpus, and their respective talk pages, we can plot the number of revisions, unique editors, and pages edited over certain time intervals in our analysis period. We experimented with different time intervals for each figure until we found an interval that best portrays the relationship of the data. Both Figure 3 and Figure 4 show drastic increases in activity in June 2019, which is the month where the 2019–20 Hong Kong protests also saw drastic escalations. After its initial spike in June, we see additional peaks in activity during July, August, and November of 2019.

The peaks seen in these figures, particularly Figure 3, coincide with various events that occurred during the movement. The initial spike in June was due to a protest that occurred on June 9th that brought around an estimated half a million people to show their dissent to the extradition amendments. Similar incidents occurred between that initial spike and our global maximum.

The global maximum we observed in July was caused by two events that occurred together. On July 21st, organizers estimated four-hundred thousand participants rallied and pushed through police-installed barriers where police then retaliated with tear gas, rubber bullets, and numerous arrests. Later on that same night, hundreds of men wearing white t-shirts attacked people at the Yuen Long station with steel rods and canes, an incident later known to be the “2019 Yuen Long attack” on Wikipedia [21]. These two events caused the spike observed in July.

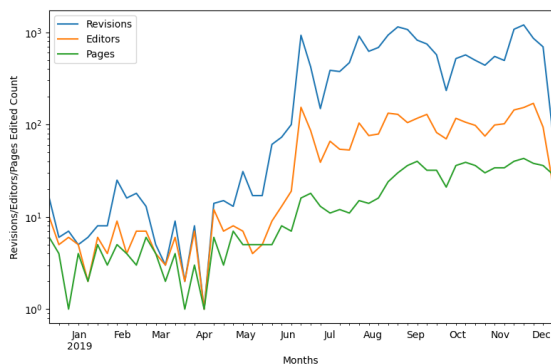
The additional spikes were the result of other events that we will cover briefly. On August 23rd, protesters formed a human chain consisting of around a quarter million people known as the “Hong Kong Way” protest [25]. On August 31st, another conflict at a train station occurred where police attacked passengers who were presumably returning home from a protest known on Wikipedia as the “2019 Prince Edward station attack.” In November, two conflicts on the 11th and 17th at the Chinese University of Hong Kong and Hong Kong Polytechnic University, respectively, occurred resulting in multi-day long sieges [9, 3, 20].

These additional peaks and consistent documentation after June 2019 suggest the occurrence of intensified documentation as the 2019–20 Hong Kong protests gains more attention, it requires more documentation by more editors to keep up with current events. The peaks also show that the documentation of the 2019–20 Hong Kong protests are responsive to real-time events which will become more evident in Figure 7.

Please do keep in mind that more than half of our article corpus was created after June 2019, so the reduced activity seen in the first six months of the analysis period should come as no surprise. To better represent the relationship between metrics, refer to Figure 4 that displays a clearer image of how metrics like revisions impact the number of editors for a given period.

#### Top Ten Articles

Figure 7 was created to see if there are common trends between these articles in their revision patterns. The revision patterns reveal that some of the older articles do not have a history of being heavily revised, as we can see with the article for Hong Kong. The older articles have more stable revision patterns while the more recent articles that correspond to the events of the 2019–20 Hong Kong protests have very noticeable spikes and dips in revision activity. This is partly due to our article corpus’ composition because numerous articles were created starting around June 2019, so they will have a larger number of



**Figure 4.** Displays the same data as Figure 3, but uses week intervals to smooth the lines and the y-axis scales logarithmically to show the parallel rises and falls between the different metrics.

revisions in their first few days of existence as large amounts of revisions are made to the article as it is initially documented.

We can observe that even within the top ten articles, there are articles that do not receive as many revisions compared to others. We also can see that the 2019–20 Hong Kong protests article has a volatile revision pattern that responds to many events of the Hong Kong protests. In contrast, articles like the 2019 Yuen Long attack peaked suddenly in relation to the specific event that the article is about and were not revised frequently afterwards. The Yuen Long attack occurred in late July and we can see a corresponding spike in the number of revisions for the 2019 Yuen Long attack article. Similarly, the 2019 Hong Kong local elections article that has a higher peak than the 2019–20 Hong Kong protests article in November of 2019, which is when the elections took place. This suggests that some of the articles in our corpus are sensitive to real world events and the revisions to those articles will increase or decrease depending on whether or not a related event has taken place. This could be a reason why we see more spikes in revisions for the 2019–20 Hong Kong protests article since it likely responds to more real-time events than some of the other, more specific articles in the corpus.

#### Temporal Analysis

In contrast to our previous analysis of just plotting the counts for revisions, unique editors, and pages edited over time, we can also observe intensified documentation through the reduction in time between a certain event and the creation of a Wikipedia article to document said event. To do so, we pulled the article creation dates for all the articles in the corpus, excluding their talk pages. With these creation dates, we removed articles that did not make sense to have start dates. These articles included those categorized as “Lists” and some articles categorized as “General/Miscellaneous,” e.g. Art of the 2019–20 Hong Kong protests. The categorizations are found in Table 2.

By omitting these articles, we reduced the original corpus to nineteen articles for Figure 8. To determine the event start date, we went through each page and found the earliest news articles dating the specific event. These citations can

be found in the GitHub repository in the Markdown file named “start\_dates\_doc.md.”<sup>11</sup>

With set dates for the event start date and the page creation date, we can now take the difference between them and plot the number of days it took from the start of an event to its eventual article creation date. Figure 8 plots this difference and labels the dots based on the categorizations given to each article in Table 2. To help determine whether the delay between the event’s start date and article creation’s date was reduced over time, the articles on the y-axis are listed from top to bottom by their event start date, i.e. the articles with the oldest start date are found higher on the y-axis.

#### Edit Sizes

Another figure we created was Figure 9 which plots the median edit sizes per day for our corpus. To cover it briefly, we can see that the figure shows a global maximum is early August and throughout the study period the median edit size increases slightly. Overall, it is was not a good measure to see intensified documentation occurred because we can see that the graph is discontinuous because there are some days with no revisions. This brings the potential for the median to vary drastically based on a few revisions. Although it is helpful to see that there is a maximum in August, it is inconclusive overall whether edit sizes have increased over time and whether that suggests intensified documentation. Potentially, increased documentation efforts could result in frequent small edits that would show low median edit sizes, but ignore the fact that they are happening more frequently.

#### Discussion

In sum, these analyses suggest a growing interest from members of the Wikipedia community in the 2019–20 Hong Kong protests as time goes on. As more events related to the 2019–20 Hong Kong protests occur, additional members of the community become interested in the circumstances of these events. The results from our analysis support the idea that there could be new editors contributing to the articles in our corpus. It is also possible that editors that already managed Hong Kong related articles are spending more time making revisions to the protest related pages. This would account for the rise in revisions and editors counts in June 2019, which is when the 2019–20 Hong Kong protests began receiving wide scale international attention.

Our analyses also show that editorship is highly sensitive to breaking-news style events such as the various attacks that occurred during some protests. Following attacks like the Yuen Long attack, there are noticeable spikes in documentation for those articles and this increased activity will likely attract other editors over time as more attention is given to these articles that cover breaking news. As another result of this increased level of attention, editors are more responsive to events and become better at documenting them closer to the occurrence of the event, as we have shown with the decreased article creation latency found in Figure 8.

<sup>11</sup> [https://github.com/gadeuneo/wiki-comps-2020/blob/master/start\\_dates\\_doc.md](https://github.com/gadeuneo/wiki-comps-2020/blob/master/start_dates_doc.md)

We would also like to bring up that the median edit sizes also showed gradual increases which could hint to intensified documentation as articles require larger edits to compensate for the frequency of events that require documentation. Although median edit size increases in Figure 9, it does not take in consideration the case where there could be numerous small edits to document an event. Thus, Figure 9 should not be the sole evidence to demonstrate intensified documentation, but instead supplement our understanding of the editing behavior of the participants.

## COLLABORATIVE MIGRATION

### Definition

Collaborative migration occurs when editors find that there is a collection of articles worth maintaining together. To put it another way, we can identify collaborative migration if we find a group of editors working together in one article that also work together in another article.

### Results

With regards to collaborative migration, we examined the similarities between sets of editors and how that changes over time across articles by using the Jaccard similarity coefficient. The Jaccard similarity coefficient measures the similarity of two sets by dividing the size of the intersection between both sets by the size of the union of the sets.

$$\text{Jaccard Similarity Coefficient} = \frac{|\text{Set A} \cap \text{Set B}|}{|\text{Set A} \cup \text{Set B}|}$$

In the case of our research, we calculated daily Jaccard similarity coefficients for each article against every other page in the corpus from June 2019 to December 2019. This value tells us the percentage of editors who have edited the target article and also on any other article in the corpus for a particular date. A small value for a given page would indicate that only a small percentage of editors that edited in this page also edited in other Hong Kong protest related articles, while a large value would show that editors of the given article also substantially overlapped the editors of other Hong Kong protest related articles.

In Figure 6, we plotted the Jaccard similarity coefficients for the top 10 articles from June to December 2019. As can be seen, the percentages never go above two percent, indicating a rather low value of similarity between articles. Comparing this score to the seven percent Jaccard similarity found in the Black Lives Matter paper, our similarity score is low in absolute terms, especially since the Black Lives Matter editor base (of 6,795 unique editors) is almost three times as large as the Hong Kong Protests' editor base (of 2,350 unique editors) [28]. Regarding the low similarity, we hypothesized that it is due to the fact that most editors only revise a few articles. To investigate that, we decided to find out how many articles editors edit in.

In Figure 6, we plotted the number of editors that edit a fraction of the article corpus. This fraction is calculated by the pages edited divided by the pages that were active during the day of

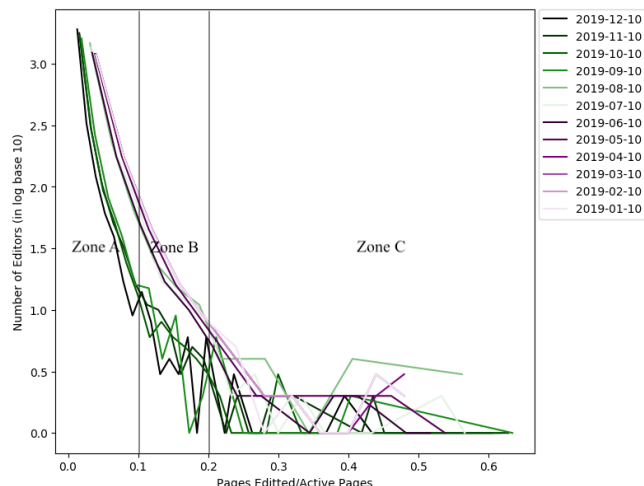


Figure 5. Displays the number of editors (in logarithmic scale) that edit a specific fraction of the article corpus. The fraction is calculated by dividing pages edited by the number of active pages during the day of the edit. The graphs are color coded into purple and green, representing before the protest, and during the protest respectively. The intensification of color represents the passing of time, where a darker purple is later in the year than a lighter purple. The graph is also split into three zones to identify three different types of editors and their differing behaviours.

the edit. We can see that most editors (more than 95 percent of them) found in Zone A edit a small fraction of the corpus, while a small handful of editors (less than one percent) found in Zone C edit a large amount of the corpus.

Then we added a time dimension into the graph to see how the protest changed the behaviours of the editors, shown by the colors of the lines. There are two general colors for the lines, purple and green. The purple lines denote the timeline 6 months before the protests, from January to June 2019 and the green lines show the timeline during the protests from June to December 2019. The intensification of the colors represent passing of time.

We then divided the graph into Zones to show different types of editors and their editing behaviour over time. Zone A has editors that edit very few pages, at most ten percent of the entire corpus (around one to four articles each). Zone B editors edit a moderate amount of articles, from ten to twenty percent of the entire corpus (around four to eight articles). Zone C contains editors that edit a large amount of articles, more than twenty percent of the corpus (more than eight articles each).

In Zone A we can see a clear division between the green and the purple lines. The color of the graph is also intensifying in the negative x-direction. This shows that for this group of editors, the increase of new editors over time is slower than the inclusion of new articles into the active corpus.

For Zone B, we see a difference in the green lines. The intensification of color is not consistent in a direction anymore, with a light green line even intersecting the purple lines. This shows that for these groups of editors, they are diversifying their editing activity to try to take on new articles.

We are most interested in the results of Zone C. Zone C shows an even more pronounced effect seen in Zone B, with all green lines intersecting with purple ones, and sometimes even trumping them. This suggests that the small community of editors in Zone C are greatly increasing their editing activity to accommodate for the increase of new articles in the corpus, much more so than the editors in Zone B.

This further supports the low Jaccard similarity coefficient scores we found previously, and suggests that there is a sort of veteran community of editors that may be monitoring the entire corpus of Hong Kong protest-related articles. This presence of a nucleus of very active editors in the Hong Kong protest Wikipedia community complements the findings in past research [13].

To analyze collaborative migration from another angle, we looked at the fraction of editors that edited in both the top 10 articles and in the non-top 10 articles, resulting in Figure 10. Using the formula below, we calculated the fraction of top 10 article editors per day since the article’s creation so that we could see if there were any trends in how top 10 article editors collaboratively migrate to non-top 10 articles.

$$\text{Fraction of Editors} = \frac{|\text{Set of Top 10 Article Editors editing in Non-Top 10 Article}|}{|\text{Set of Top 10 Article Editors}|}$$

Overall, there is a very low percentage of editors from the top 10 articles collaborating on non-top 10 articles. At its highest peak, we see a 0.35 percent collaboration value from top 10 article editors. There is also a noticeable trend where editors from the top 10 articles mostly collaborate on non-top 10 articles only in the first two or three weeks of an article’s creation. As the article gets older and less information needs to be added to the articles, top 10 editors stop collaborating on these articles. This trend is similar to the trends found in editorship on natural disasters where participation is relatively high in the first two to three weeks of an article’s creation but that participation quickly dies down after that initial period [11]. Even though our article corpus is varied with different types of articles that cover topics from deaths to lists of protests, we see this trend hold for all the non-top 10 articles in our corpus.

## Discussion

Our findings related to collaborative migration for the 2019-20 Hong Kong protests shows that collaboration among editors is low, with a peak of two percent of the editors collaborating together. Comparing this result with the Black Lives Matter paper [28], our findings are lower than their finding of seven percent, indicating that our findings are low in absolute terms.

There are a few possible reasons that account for the low collaborative migration result. We found evidence that a majority of editors only made a few edits to one or two pages without branching out beyond those articles, which is a similar phenomenon found in the study done on the Black Lives Matter movement [28]. Since those editors only made a few edits and then never edited other articles in the corpus, the overall collaboration score gets lowered.

Similarly, it is possible that many of our editors may be “specialists” that are only interested in particular topics. For example, we have the article for Hong Kong itself in our corpus, and it is possible that an editor for that article may only be interested in the economic policy of Hong Kong and not the protests. As a result, the level of collaboration across articles will be lowered since they do not edit other articles in the corpus other than the one(s) that they are interested in editing.

Another reason may be due to the language preferences of editors interested in the Hong Kong protests. Since the 2019 Hong Kong protest is an international movement, it is possible that there are editors that are multilingual and may participate in collaborating on articles that are not in the English version of Wikipedia. These editors and their edits would consequently not be seen in our data, lowering the Jaccard similarity score.

## DYNAMIC REAPPRAISAL

### Definition

Dynamic reappraisals occur when new events trigger reassessments of past events. The reassessments can come in many forms: increased pageviews to existing articles, creation of new pages related to past events, or updates and reformations of existing articles.

### Results

#### *Pageview Activity*

Figure 11 and Figure 12 were made using data pulled from Wikimedia Foundation Labs (WMF Labs) tool for each article. The WMF Labs tool returned pageview data from previous article titles that were moved due to page moves/redirects. For Figure 11, we plotted the number of pageviews for the four articles with the highest correlation values and the focus article from June 10th, 2019 to December 10th, 2019; note that not all articles existed throughout this timeframe and that some were created later in the year. As for Figure 12, we plotted articles that were related to the protests with notable event dates for the same time period.

We plotted pageviews between June 10th, 2019 to December 10th, 2019 to reflect the growing interest in the 2019–20 Hong Kong protests. The pageviews for the articles in Figure 11 appear to show a similar trend in interest by readers. The 12 June 2019 Hong Kong protest article was created to denote when the 2019–20 Hong Kong protests began receiving international attention. With that article receiving attention, we observed similar spikes in view with the focus article. Notably, the article around the Death of Chow Tsz-lok was created around the same time and saw a similar spike in interest perhaps due to increased media coverage of the event.

The article of the Murder of Poon Hiu-wing also stands out because it was a catalyst for the extradition amendments that sparked the 2019–20 Hong Kong protests. Although the page was created over a year after the event, it still follows the overall pageview trends with spikes in viewership along the other articles and the focal one. Interest in the List of December 2019 Hong Kong protests also followed the direction of pageviews of the focal article as new events unfolded.



In Figure 12, we see that there are spikes in viewership around notable events for almost all the pages involved. Even articles that were created before the protests see increases in pageviews as readers make sense of why they are occurring. Readers are interested in the causes of the protests as Causes of the 2019–20 Hong Kong protests page sees a continued increase in views with a peak in November around the time of the Death of Chow Tsz-lok and with the increase in views for the Causes of the 2019–20 Hong Kong protests.

#### *Correlation Values*

The correlations were calculated using the Python package Pandas’ built-in correlation method that implements Pearson’s Correlation Coefficient<sup>12</sup>. We chose to use Pearson’s Correlation Coefficient as we are comparing two similar values: pageviews and revisions for two different pages and want to see if there is a linear relationship, i.e. whether views/edits on one page are correlated with another. For pageview-revision correlation, we simply appended the datasets together per page before using the same correlation function.

Tables 4–6 show the highest positive correlation values for daily revision, pageview, and revision-pageview activity, respectively. The revision correlation values (Table 4) are consistently lower than both the pageview (Table 5) and the pageview-revision (Table 6) correlations values. Both the pageview and pageview-revision correlation values are lower and can be attributed to the effort required to edit compared to simply viewing an article. Additionally, the top revision correlation values are focused on the Siege of Hong Kong Polytechnic University article since it was covered in real-time and was a notable event with increased media coverage. The pageview and pageview-revision correlation article pairs are similar because pairs often include an article listing the events of a given period which are accessible at a glance and are focused more on the most recent events.

Compared to the Black Lives Matter analysis, we had a smaller amount of data since we had a smaller overall analysis period regarding the 2019–20 Hong Kong protests. We also had confounding variables with our event start dates and page creation dates as our pageview correlation values were tied more to the dates the pages were created rather than a linear relationship.

#### *Hierarchical Clustering*

We used hierarchical clustering to identify any relationships between the articles in our corpus. To illustrate hierarchical clustering, a dendrogram will be used to show the coalescing that occurs at each iteration of the hierarchical clustering algorithm. We created dendrograms (Figures 13 and 14) for both pageview and revision correlation values from our corpus, and we chose this type of visualization over a network graph because we found the relationships among articles easier to visualize with dendrograms. To determine which clusters to merge at each iteration of the algorithm, we used single-linkage clustering as implemented by the SciPy library.<sup>13</sup> Be-

cause single-linkage coalesces the clusters with the lowest distance, we had to modify the correlation values so that the highest correlation values were selected instead of the lowest. Single-linkage clustering was used because we simply wanted the articles with the highest correlation values to be merged at every iteration, and other clustering methods had more specific requirements for how the distances had to be formatted which we did not feel was appropriate for our correlation values.

Our results tell us that for both page view and page revision correlations we can separate the articles into two separate clusters denoted by the difference in color. In both Figure 13 and Figure 14, there is one massive cluster that contains most of the articles found in our study while the other cluster is composed of the few remaining articles. We did not create a dendrogram for pageview-revision correlations since there was no discernible relationship for either pageview or revision correlations separately. Since the articles have mostly been identified as belonging to the same cluster, we are not able to discern any particular relationship among any of the articles other than that they are all part of our corpus.

A factor that may have contributed to our large clusters is a chaining issue where the single-linkage clustering algorithm finds clusters and joins them together even though the relationships of clusters would be better preserved if it had remained as two separate clusters. We could be seeing a “snowball” effect where the algorithm finds an article that is close enough to at least one other article in the cluster and adds it to the collection even if it is relatively far away from the other articles in the newly formed cluster. This is because the newly merged cluster’s distance to every other cluster is the pairwise minimum of the set of articles in the cluster, and the set of articles in other clusters. As a result, we lose the relationship that we had while the clusters were smaller and remained separate.

Another possible reason why our dendrograms show such a large cluster could be due to faults in our correlation values. It is possible that the correlation values found from our analysis could be influenced and skewed by the page creation dates for certain articles. For example, the article for Hong Kong has existed for the entirety of our ten year study period while some articles were created close to the end of our study period. There will not be as much data that can be correlated between these two articles. Since there are quite a few of these disparities between article creation dates, there could be quite a bit of interference in this analysis.

#### **Discussion**

We had mixed results in our findings of dynamic reappraisal compared to Twyman et al [28]. However, we did find evidence from the pageview plots that suggests editors and viewers revisiting existing articles when covering ongoing events. As readers and Wikipedia editors take in new information about ongoing events, it may require them to revisit older articles to assist their understanding of the comprehensive social movement.

The less pronounced effect of collaborative migration and dynamic reappraisal as compared to the Twyman et al.’s paper suggests that the nature and motivations of the social

<sup>12</sup><https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.corr.html>

<sup>13</sup><https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html>

movements can influence the collective editing behaviour on Wikipedia [28]. In particular, we note that the Black Lives Matter movement has an emphasis on the memorialization of deaths of unarmed black men by police that fits within the overall context of the racial history of the US and the systemic issues that it addresses. In contrast, the 2019–20 Hong Kong protests are focused on the future of the country and the concerns about China’s interference with its autonomy which has a wider audience as democratic regimes share their concerns.

We observe correlations between actions of creating knowledge with revisions and consuming knowledge with pageviews between related articles, though there is an overall lack of evidence for dynamic reappraisal. Within the context of the overall 2019–20 Hong Kong protests movement, these correlation values reflect different priorities between editors and viewers which may explain the lack of evidence. For editors, we see a focus surrounding complex events to collective recording of different narratives that have competing narratives. As mentioned earlier, article creation dates may have skewed our correlation values, but we still see that viewership is tied to recent events and look for easily digestible content regarding breaking headline news. When we combine these metrics, we see that editors focus on updating the pages that viewers would likely focus on when getting the most recent information regarding the 2019–20 Hong Kong protests. Overall, we do not see strong evidence supporting dynamic reappraisal as older articles do not appear to be prominent in our correlation values which may also be attributed to our relative lack of data compared to Twyman et al [28].

A similar argument for our hierarchical clustering results shows no evidence for dynamic reappraisal. There is a lack of clear, distinct groups in our clustering analysis which makes it unclear how people make sense of the social movement. This may be attributed to issues with our correlation values as the analysis is based on them, and our earlier stated issues with the single-linkage clustering algorithm. Our corpus has more distinct groups than Twyman et al. which may have affected how well hierarchical clustering can distinguish clear groups [28].

## LIMITATIONS

### Article Corpus Diversity

One difference that must be acknowledged between our analysis of the Hong Kong protests and the previous research done on the Black Lives Matter movement is the composition of our article corpus. Although our article corpus sizes are similar (Black Lives Matter paper had forty-two articles where we had thirty-eight), the uniformity in the Black Lives Matter corpus was something we could not replicate. From their forty-two page corpus, one page documented the movement itself, while thirty-seven of the pages documented particular deaths, and four pages documented particular protests. Our corpus on the other hand was less uniform.

Instead of being categorized into the aforementioned three groups, our corpus was categorized into six groups, as can be seen in Table 2. The number of groups would have increased if we did not have a “General/Miscellaneous” group. One

challenge that came from this diversity was producing Figure 8 because we had to remove multiple articles that could not be given an event start date. These included the seven “List” pages in Table 2 and articles not covering events like the Government of Hong Kong and Hong Kong articles. That not only left us with a smaller set of pages compared to the original analysis, but a more diverse one where the original paper only had articles relating to deaths over a greater period of time. Despite these challenges, we still believe that the articles we were left with demonstrate intensified documentation and the reduction of delay between an event’s start date and article creation date.

The diversity in the article corpus also brought out challenges in collaborative migration and dynamic reappraisal. Having a high Jaccard similarity score would mean that editors had to be knowledgeable and interested in a diverse range of fields. This is unlikely since our article corpus had a broad range of topics such as protests, deaths, and legislation. For example, editors familiar with policies and legislative articles would probably be unfamiliar with writing in articles about deaths. The diversity of our article corpus, along with the fact that editors only make one or two edits in general, may have been a factor that limited our collaborative migration analysis. This may have also affected our dynamic reappraisal results by increasing the difficulty for our hierarchical clustering to distinguish groups of articles.

### Analysis Period

It can be argued that a six month analysis period from the start of the 2019–20 Hong Kong protests gaining international attention in June to the end of the analysis period in December is too short to determine whether intensified documentation was found.

For the case of collaborative migration, a longer analysis period could have resulted in more articles in each category of the article corpus shown in Table 2, creating a more cohesive article corpus. This cohesiveness would help verify and illuminate whether the sparseness of each category in the corpus resulted in the low Jaccard similarity scores.

Lastly, the short analysis period made it difficult to replicate the dynamic reappraisal section because calculating correlation values between articles, as many were created near the end of our study period, depended significantly on the period of time both of the articles existed. These issues later compounded into our hierarchical clustering results.

Although we would have preferred a longer analysis period, complications due to COVID-19 and social distancing mandates may have influenced the attention of Wikipedia editors and the frequency of large scale protests. The World Health Organization (WHO) reported the first pneumonia cases in Wuhan on January 4th, 2020 and on January 22th, 2020 Hong Kong reported its first cases of COVID-19 [18, 4]. About two months later, on March 28th, 2020, Carrie Lam announced restrictions for people to gather in groups larger than four people in public places [1]. It is not extreme to say that these would have had effects on the protests overall and in turn the

documentation of it on the English version of Wikipedia past our analysis period.

### Generalizing International Results

One of the reasons why we picked the Hong Kong protests for our social movement was because we wanted to see how the documentation of a domestic social movement, like Black Lives Matter, compared to a more international social movement, like the 2019–20 Hong Kong protests. The Hong Kong protests are more international than the Black Lives Matter movement because it involves Hong Kong citizens, Mainland China, as well as English speakers who may be observing the movement. Because Hong Kong is a multilingual society, editors from Hong Kong that are fluent in multiple languages may want to edit articles for the Hong Kong protests in a language other than English. It is possible that their attention could be divided across multiple locations. For example, if there was a Hong Kong editor who edited a protest article in the English version of Wikipedia and then they edited another protest article in a different version, such as the Chinese version, we would not be able to see that migration from one article to another. Since we are only seeing the editors' contribution to the English version of our articles, it is possible that we are only seeing a subset of the total collaboration that is actually occurring.

The Black Lives Matter paper discussed the limitations of studying a “single, prominent, contemporary social movement case in the United States” may limit its ability to be generalized to other movements. By studying the 2019–20 Hong Kong protests, our findings provide insight into whether these three phenomena (intensified documentation, collaborative migration, and dynamic reappraisal) can be found in other contexts.

### FURTHER WORK

As noted in Limitations, our analysis period is relatively short compared to the Black Lives Matter analysis. Revisiting the 2019–20 Hong Kong protests a few years later potentially could reveal new factors in the documentation patterns we missed in our one year study period. Additionally, we could also explore the novel impacts of the COVID-19 pandemic on the documentation of a social movements. Exploring other international movements that are either ongoing or have concluded such as Brexit or the Arab Spring protests is another avenue for further exploration. Alternatively, replicating the Twyman et al.'s on domestic social movement such as Occupy Wall Street and March for Our Lives would have been easier to do comparative analysis and can be the approach of another study [28].

### CONCLUSION

Our findings support intensified documentation, with large increases of revision activity and new unique editors joining the community from June 2019. As the community grows, the revisions show more responsiveness towards real time events, causing the article creation latency to fall.

However, for collaborative migration, the findings are not pronounced. We found that it is because most editors only

edit 1-2 pages. However, there is evidence to suggest a small community of highly invested editors that also edit in newly created pages, corroborating previous research that there is a nucleus of “elite” editors that help moderate a set of pages [13].

Lastly, we find evidence partially supporting dynamic reappraisal as breaking news events increase traffic to past pages and pages concerning ongoing events. However, our findings with the single-clustering analysis and correlation values were inconclusive.

The less pronounced effect of collaborative migration and dynamic reappraisal as compared to the Black Lives Matter paper suggests that the nature and motivations of the social movements can influence the collective editing behaviour on Wikipedia.

Twyman et al.'s findings for intensified documentation were applicable to an international movement, but the findings for collaborative migration and dynamic reappraisal were inconclusive [28]. Given a longer analysis period, it is possible that the findings for collaborative migration and dynamic reappraisal could be more definitive, but with the data we currently have for the 2019–20 Hong Kong protests we are unable to confidently say that these two phenomena were present for this movement.

### ACKNOWLEDGMENTS

As a final note, we would like to thank those who make this research project possible. This research project concludes our undergraduate computer science careers and our Carleton College experience. We would first like to thank Professor Sneha Narayan, our research/project advisor, for supervising our progress and allowing us to explore the dynamics of social computing research. We would also like to thank the Computer Science Department at Carleton College who provided us with this unique environment to learn, grow, and prepare ourselves for our post-Carleton lives. Without their efforts, our project would not have been possible.

### REFERENCES

- [1] Tatum Ancheta. 2020. Things you need to know about Hong Kong's social distancing restrictions. (May 2020). <https://www.timeout.com/hong-kong/things-to-do/things-you-need-to-know-about-hong-kongs-social-distancing-restrictions>
- [2] Judd Antin and Coye Cheshire. 2010. Readers Are Not Free-Riders: Reading as a Form of Participation on Wikipedia. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 127–130. DOI: <http://dx.doi.org/10.1145/1718918.1718942>
- [3] Holmes Chan, Kris Cheng, and Jennifer Creery. 2019. Hong Kong police fire live rounds and tear gas as protesters disrupt morning traffic in citywide 'general strike' bid. (Nov 2019). <https://hongkongfp.com/2019/11/11/hong-kong-police-f>

- ire-live-rounds-tear-gas-protesters-disrupt-morning-traffic-citywide-general-strike-bid/
- [4] Elizabeth Cheung. 2020. China coronavirus: death toll doubles as Hong Kong reports its first cases. (Jan 2020). <https://www.scmp.com/news/hong-kong/health-environment/article/3047193/china-coronavirus-first-case-confirmed-hong-kong>
- [5] Elizabeth Joseph CNN and Katie Hunt. 2016. Missing Hong Kong bookseller: I was kidnapped. (Jun 2016). <https://www.cnn.com/2016/06/16/asia/china-hong-kong-booksellers/index.html>
- [6] Jennifer Earl and Katrina Kimport. 2011. *Digitally Enabled Social Change: Activism in the Internet Age*. The MIT Press. <http://www.jstor.org/stable/j.ctt5hhcb9>
- [7] Michela Ferron and Paolo Massa. 2011. The Arab Spring| WikiRevolutions: Wikipedia as a Lens for Studying the Real-Time Formation of Collective Memories of Revolutions. *International Journal of Communication* 5, 0 (2011). <https://ijoc.org/index.php/ijoc/article/view/1238>
- [8] Damien Gayle, Kate Lyons, Verna Yu, Helen Davidson, and Lily Kuo. 2019. Hong Kong protest: police fire teargas at demonstrators – as it happened. (Jun 2019). <https://www.theguardian.com/world/live/2019/jun/12/hong-kong-protest-demonstrators-and-police-face-off-over-extradition-bill-live>
- [9] Chan Ho-him. 2020. Chinese University to spend HK\$70 million for repairs after vandalism by radicals. (Jan 2020). <https://www.scmp.com/news/hong-kong/education/article/3044392/chinese-university-spend-hk70-million-fix-campus>
- [10] Ralph Jennings. 2019. How an obscure Taiwan murder case led to Hong Kong’s mega-protests. (Jun 2019). <https://www.latimes.com/world/la-fg-taiwan-china-hong-kong-chan-tong-kai-murder-20190619-story.html>
- [11] Brian Keegan, Darren Gergle, and Noshir Contractor. 2011. Hot off the Wiki: Dynamics, Practices, and Structures in Wikipedia’s Coverage of the Tōhoku Catastrophes. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym ’11)*. Association for Computing Machinery, New York, NY, USA, 105–113. DOI: <http://dx.doi.org/10.1145/2038558.2038577>
- [12] Aniket Kittur and Robert E. Kraut. 2008. Harnessing the Wisdom of Crowds in Wikipedia: Quality through Coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW ’08)*. Association for Computing Machinery, New York, NY, USA, 37–46. DOI: <http://dx.doi.org/10.1145/1460563.1460572>
- [13] David Laniado and Riccardo Tasso. 2011. Co-Authorship 2.0: Patterns of Collaboration in Wikipedia. In *Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia (HT ’11)*. Association for Computing Machinery, New York, NY, USA, 201–210. DOI: <http://dx.doi.org/10.1145/1995966.1995994>
- [14] Amy Gunia Leung and Hillary. 2019. Hong Kong Leader Defends Extradition Bill After Mass Protest. (Jun 2019). <https://time.com/5603754/hong-kong-carrie-lam-china-extradition-bill-more-protests/>
- [15] BBC News. 2019a. Hong Kong formally scraps extradition bill that sparked protests. (Oct 2019). <https://www.bbc.com/news/world-asia-china-50150853>
- [16] BBC News. 2019b. The Hong Kong protests explained in 100 and 500 words. (Nov 2019). <https://www.bbc.com/news/world-asia-china-49317695>
- [17] Aileen Oeberst, Iassen Halatchliyski, Joachim Kimmerle, and Ulrike Cress. 2014. Knowledge Construction in Wikipedia: A Systemic-Constructivist Analysis. *Journal of the Learning Sciences* 23, 2 (2014), 149–176. DOI: <http://dx.doi.org/10.1080/10508406.2014.888352>
- [18] World Health Organization. 2020. Rolling updates on coronavirus disease (COVID-19). (Jan 2020). <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>
- [19] James Pomfret. 2019. Protesters arrested in Hong Kong over proposed China extradition law. (Mar 2019). <https://www.reuters.com/article/us-hongkong-politics-extradition-idUSKCN1QW1AC>
- [20] Hong Kong Free Press. 2019. ‘A heartbreaking day’: PolyU student leaders call for public help as hundreds trapped in campus. (Nov 2019). <https://hongkongfp.com/2019/11/18/heartbreaking-day-polyu-student-leaders-call-public-hundreds-trapped-campus/>
- [21] Helen Regan and Anna Coren. 2019. Night of violence shocks Hong Kong after seventh week of mass marches. (Jul 2019). <https://www.cnn.com/2019/07/21/asia/hong-kong-protest-july-21-intl-hnk/index.html>
- [22] South China Morning Post Reporters. 2019. As it happened: Hong Kong’s historic march as ‘2 million’ people peacefully protest. (Jun 2019). <https://www.scmp.com/news/hong-kong/politics/article/3014695/sea-black-hong-kong-will-march-against-suspended>
- [23] Special Reports. 2019. How murder, kidnappings and a miscalculation set off Hong Kong revolt. (Dec 2019). <https://www.reuters.com/investigates/special-report/hongkong-protests-extradition-narrative/>
- [24] Grace Shao. 2019. Social media has become a battleground in Hong Kong’s protests. (Aug 2019). <https://www.cnbc.com/2019/08/16/social-media-has-become-a-battleground-in-hong-kongs-protests.html>
- [25] Phila Siu, Alvin Lum, and Zoe Low. 2019. Police arrest 63 after chaos on Hong Kong’s train network. (Sep 2019). <https://www.scmp.com/news/hong-kong/law-and-crime/article/3025241/chaos-hong-kongs-mtr-network-police-chase-protesters>



- [26] Emily Stewart. 2019. How China used Facebook, Twitter, and YouTube to spread disinformation about the Hong Kong protests. (Aug 2019). <https://www.vox.com/recode/2019/8/20/20813660/china-facebook-twitter-hong-kong-protests-social-media>
- [27] Cindy Sui. 2019. The murder case where no-one wants the killer. *BBC News* (Oct 2019). <https://www.bbc.com/news/world-asia-china-50148577>
- [28] Marlon Twyman, Brian C. Keegan, and Aaron D. Shaw. 2016. Black Lives Matter in Wikipedia: Collaboration and Collective Memory around Online Social Movements. *CoRR* abs/1611.01257 (2016). <http://arxiv.org/abs/1611.01257>
- [29] Beina Xu and Eleanor Albert. 2017. Media Censorship in China. (Feb 2017). <https://www.cfr.org/backgrounders/media-censorship-china>
- [30] Raymond Yeung, Choi Chi-yuk, Christy Leung, Clifford Lo, and Phila Siu. 2017. Missing Chinese tycoon last seen at luxury Hong Kong hotel. (Jan 2017). <https://www.scmp.com/news/china/policies-politics/article/2066744/missing-chinese-billionaire-xiao-jianhua-last-seen-hong>

## APPENDIX

<b>Title</b>	<b>Creation Date</b>	<b>Category</b>
List of March–June 2019 Hong Kong protests	2019-06-10	Lists
List of July 2019 Hong Kong protests	2019-07-06	Lists
List of August 2019 Hong Kong protests	2019-08-14	Lists
List of September 2019 Hong Kong protests	2019-09-06	Lists
List of October 2019 Hong Kong protests	2019-10-28	Lists
List of November 2019 Hong Kong protests	2019-11-13	Lists
List of December 2019 Hong Kong protests	2019-11-29	Lists
Murder of Poon Hiu-wing	2019-12-01	Deaths
Death of Chow Tsz-lok	2019-11-13	Deaths
Death of Chan Yin-lam	2019-11-13	Deaths
2019 Hong Kong extradition bill	2019-05-21	Policies/Legislation
Hong Kong Human Rights and Democracy Act	2017-09-10	Policies/Legislation
Carrie Lam	2008-07-26	Actors/Organizations
Demosistō	2016-04-06	Actors/Organizations
Civil Human Rights Front	2007-07-02	Actors/Organizations
2019–20 Hong Kong protests	2019-06-10	Protests/Attacks
2019 Yuen Long Attack	2019-07-23	Protests/Attacks
Umbrella Movement	2014-10-02	Protests/Attacks
Hong Kong 1 July marches	2005-12-06	Protests/Attacks
Chinese University of Hong Kong conflict	2019-11-18	Protests/Attacks
Siege of the Hong Kong Polytechnic University	2019-11-18	Protests/Attacks
Storming of the Legislative Council Complex	2019-10-14	Protests/Attacks
Hong Kong Way	2019-08-23	Protests/Attacks
2019 Prince Edward station attack	2019-10-06	Protests/Attacks
12 June 2019 Hong Kong protest	2019-11-13	Protests/Attacks
Hong Kong	2001-09-26	General/Miscellaneous
2019 Hong Kong local elections	2018-12-12	General/Miscellaneous
Reactions to the 2019–20 Hong Kong protests	2019-08-23	General/Miscellaneous
Tactics and methods surrounding the 2019–20 Hong Kong protests	2019-08-22	General/Miscellaneous
One country, two systems	2003-05-05	General/Miscellaneous
Government of Hong Kong	2004-05-27	General/Miscellaneous
Police misconduct allegations during the 2019–20 Hong Kong protests	2016-03-27	General/Miscellaneous
Art of the 2019–20 Hong Kong protests	2019-10-08	General/Miscellaneous
Causes of the 2019-20 Hong Kong protests	2019-08-14	General/Miscellaneous
Glory to Hong Kong	2019-09-10	General/Miscellaneous
Lennon Wall (Hong Kong)	2015-10-07	General/Miscellaneous
HKmap.live	2019-10-10	General/Miscellaneous
Hong Kong–Mainland China conflict	2015-02-22	General/Miscellaneous

Table 2. Table captions should be placed below the table. We recommend table lines be one point, 25 percent black. Minimize use of table grid lines.

<b>Article</b>	<b>Revisions</b>	<b>Editors</b>	<b>Pageviews</b>	<b>Talk Revisions</b>	<b>Talk Editors</b>	<b>Talk Pageviews</b>
2019–20 Hong Kong protests	6,685	593	4,980,504	49	14	571
Hong Kong	5,029	1,098	22,555,044	49	14	571
Carrie Lam	1,110	245	1,391,805	76	38	1,156
2019 Hong Kong extradition bill	718	131	1,008,610	82	33	1,468
2019 Hong Kong local elections	645	95	392,216	85	23	1,024
Reactions to (cont.)	529	88	87,004	116	27	942
2019 Yuen Long attack	470	80	93,508	239	32	1,890
Tactics and methods (cont.)	452	72	109,132	49	14	571
One country, two systems	426	146	918,380	29	20	784
Umbrella Movement	380	86	448,549	29	16	784
<b>Top 10 Totals by Revisions</b>	16,444	2,069	31,984,752	804	470	9,339
<b>Grand Total w/ All Pages</b>	21,299	2,350	34,153,850	1,360	520	16,094
<b>Grand Totals w/ Talk Pages</b>	22,659	2,552	34,169,944			

Table 3. Contains revision, unique editor, and pageview counts from the ten pages that have the most revisions in our article corpus from December 10, 2009 to December 10, 2019. Since these articles have the most revisions, we will refer to these articles as our top 10 articles.

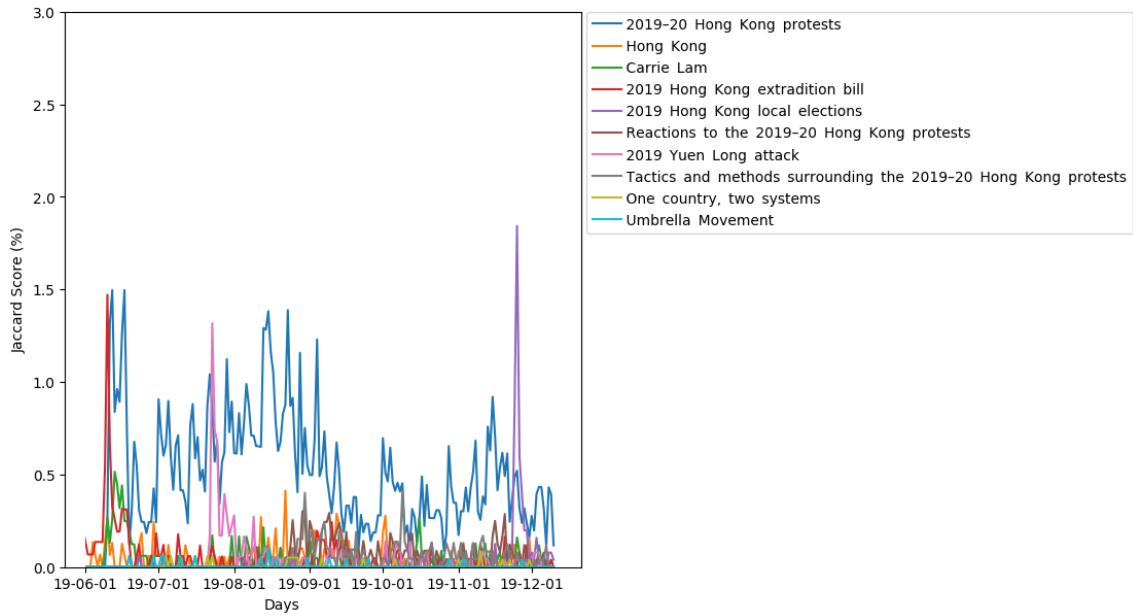


Figure 6. Shows the Jaccard score in percentage over time for the top 10 articles in our corpus.

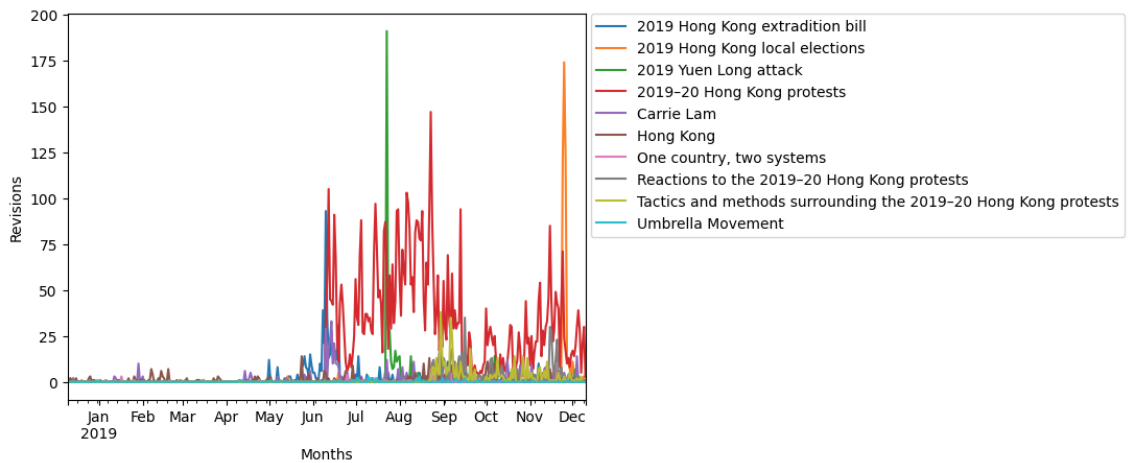


Figure 7. Displays the number of daily revisions made to the ten articles with the most revisions in our corpus over the course of our analysis period.

Article 1	Article 2	Correlation
Lennon Wall (Hong Kong)	Siege of the Hong Kong Polytechnic University	0.923
Siege of the Hong Kong Polytechnic University	Umbrella Movement	0.881
List of protests in Hong Kong	Siege of the Hong Kong Polytechnic University	0.848
Carrie Lam	Siege of the Hong Kong Polytechnic University	0.848
Civil Human Rights Front	Siege of the Hong Kong Polytechnic University	0.848

Table 4. Top five correlation values calculated using revision data.

Article 1	Article 2	Correlation
List of December 2019 Hong Kong protests	Murder of Poon Hiu-wing	0.986
12 June 2019 Hong Kong protest	Death of Chow Tsz-lok	0.972
Hong Kong–Mainland China conflict	List of protests in Hong Kong	0.930
Police misconduct allegations (cont.)	Reactions to the 2019–20 Hong Kong protests	0.927
Chinese University of Hong Kong conflict	Siege of the Hong Kong Polytechnic University	0.903

Table 5. Top five correlation values calculated using pageview data.

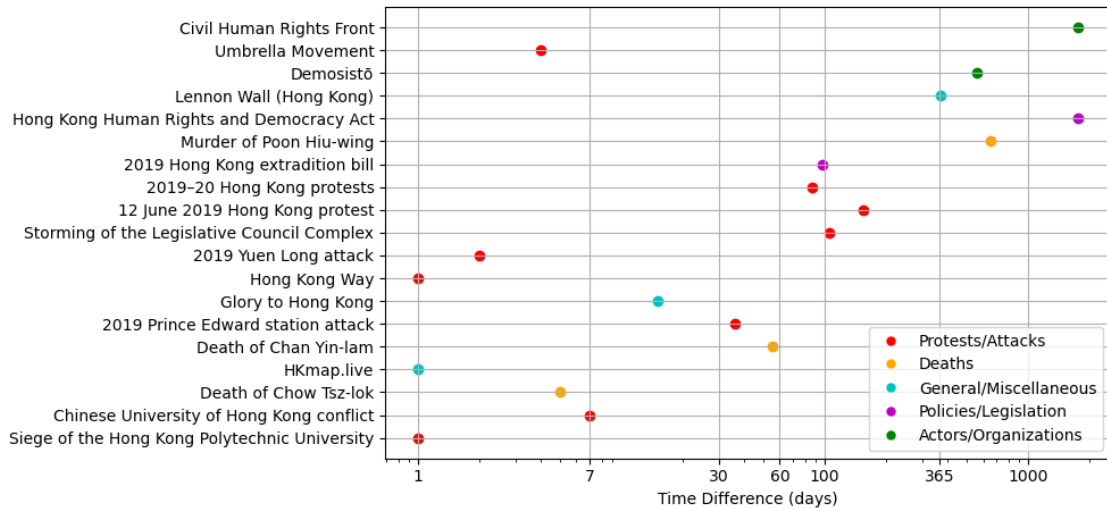


Figure 8. Displays the number of days it took for a Wikipedia article to be created after a particular event for a subset of our article corpus. The articles are organized on the y-axis so that the oldest articles, by its real-life start date, are higher. Also, note that the x-axis is on a logarithmic scale.

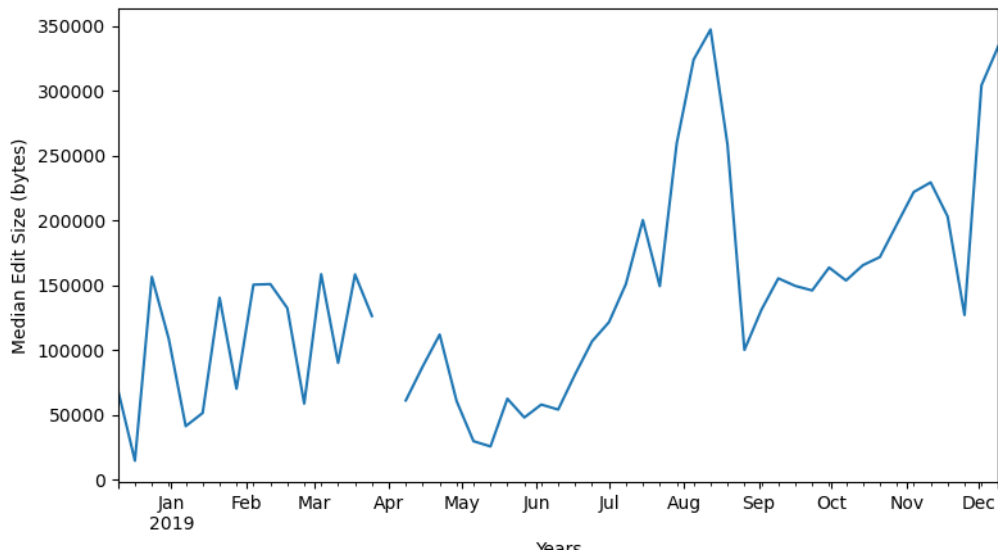


Figure 9. Median edit size (in bytes) per day for the ten articles in our article corpus with the highest revision counts. Refer to Table 2 for the full list of articles.

Article 1	Article 2	Correlation
List of December 2019 Hong Kong protests	Murder of Poon Hiu-wing	0.986
12 June 2019 Hong Kong protest	Death of Chow Tsz-lok	0.971
Police misconduct allegations (cont.)	Reactions to the 2019-20 Hong Kong protests	0.927
Hong Kong–Mainland China conflict	List of protests in Hong Kong	0.926
2019–20 Hong Kong protests	List of protests in Hong Kong	0.904

Table 6. Top five correlation values calculated using both revision and pageview data.



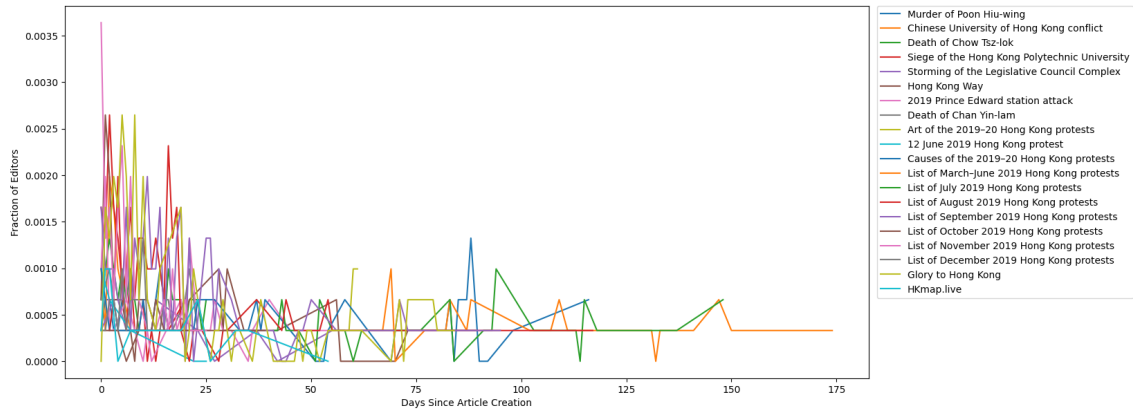


Figure 10. Displays the fraction of editors that edited the top 10 articles that also edited any of the other articles in the corpus. We display only the articles created in 2019 in this graph as they are the articles most relevant to the protests.

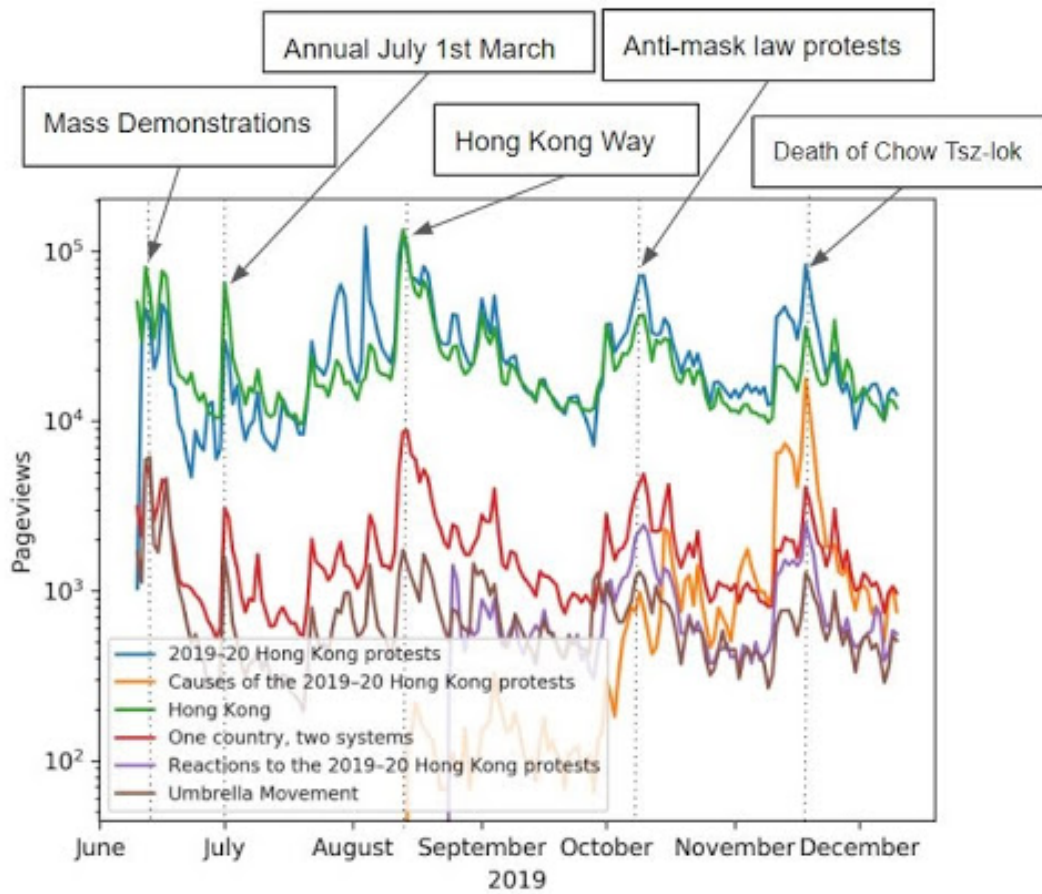


Figure 11. Pageviews of six articles related to the 2019–20 Hong Kong protests and notable event dates.

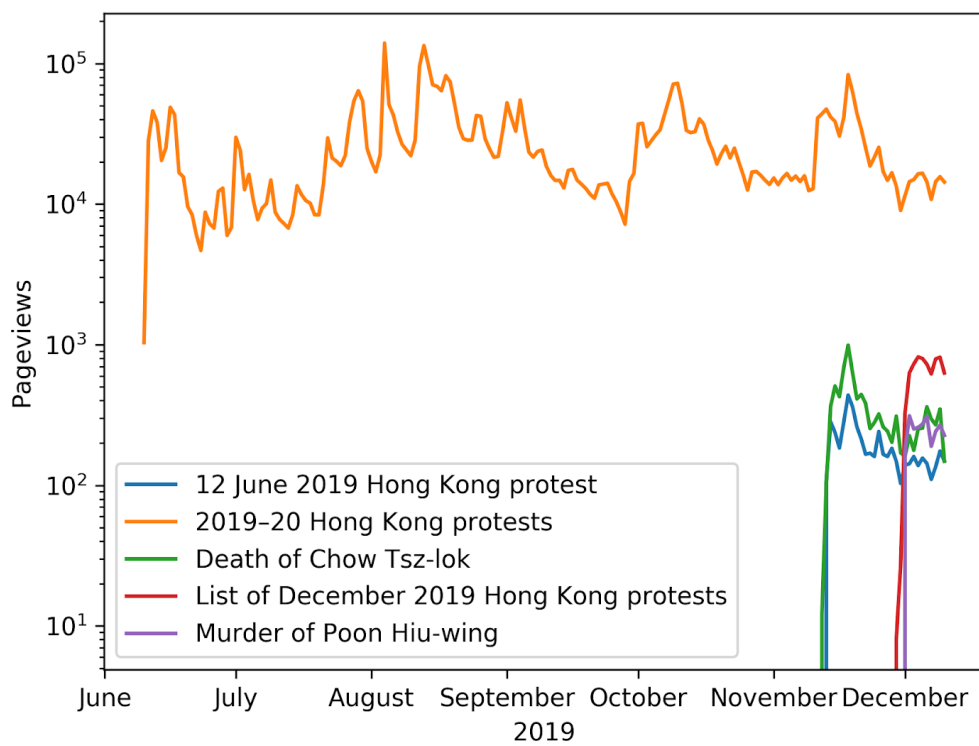


Figure 12. Pageviews among the four articles with the highest positive correlations and the focus article. Correlations were calculated from the daily attention dynamics.

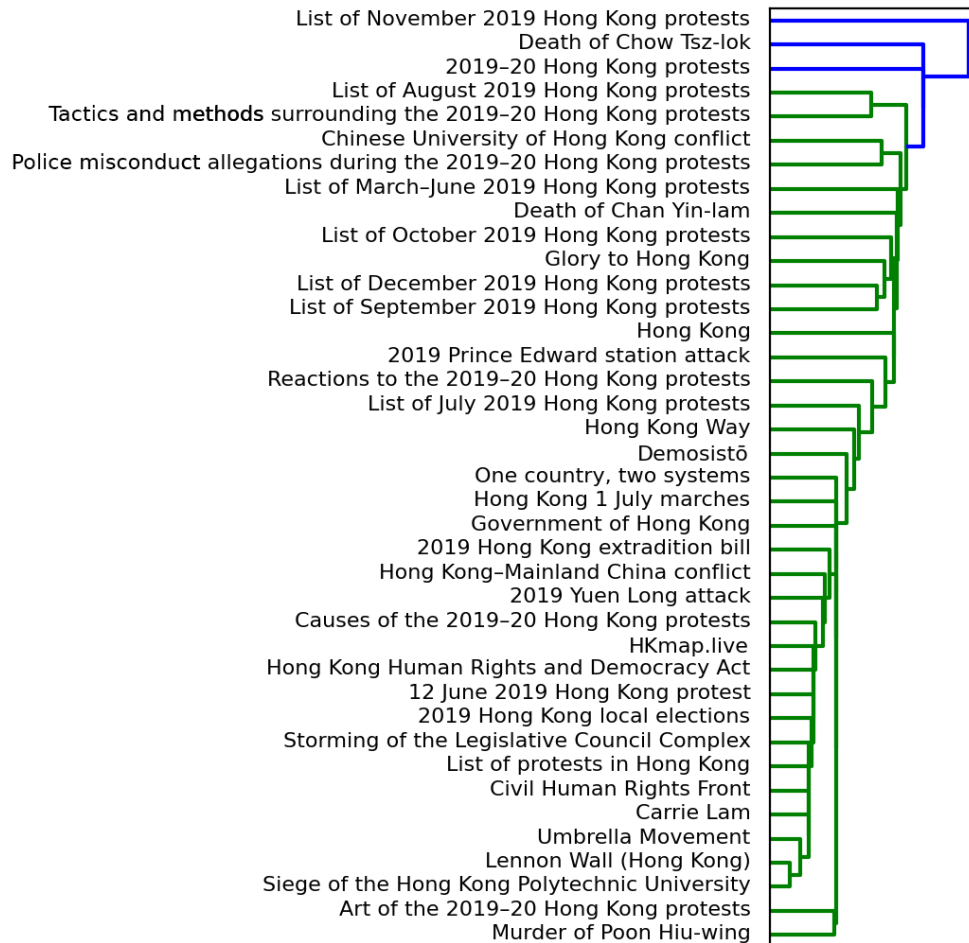
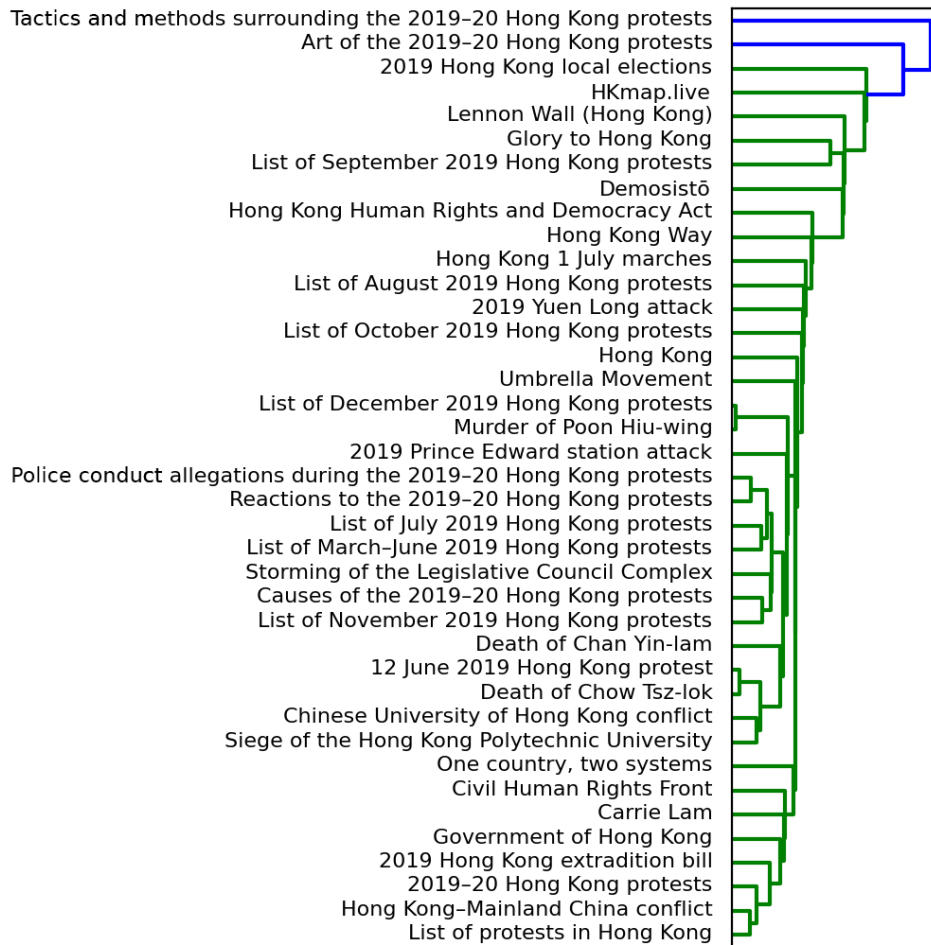


Figure 13. Dendrogram showing the relationships between articles based on pageview correlations in our corpus using the Single-Linkage algorithm as implemented by the SciPy Hierarchy Library.



**Figure 14.** Dendrogram showing the relationships between articles based on revision correlations in our corpus using the Single-Linkage algorithm as implemented by the SciPy Hierarchy Library.