Article

The Impact of COVID-19 on Pedestrian Flow Patterns in Urban POIs—An Example from Beijing

Yihang Li ¹ and Liyan Xu ^{2,*}

¹ College of Urban and Environmental Sciences, Peking University, Beijing 100871, China; got@pku.edu.cn

² College of Architecture and Landscape Architecture, Peking University, Beijing 100871, China

* Correspondence: xuliyan@pku.edu.cn

Abstract: The COVID-19 pandemic is a major challenge for society as a whole, and analyzing the impact of the spread of the epidemic and government control measures on the travel patterns of urban residents can provide powerful help for city managers to designate top-level epidemic prevention policies and specific epidemic prevention measures. This study investigates whether it is more appropriate to use groups of POIs with similar pedestrian flow patterns as the unit of study rather than functional categories of POIs. In this study, we analyzed the hour-by-hour pedestrian flow data of key locations in Beijing before, during, and after the strict epidemic prevention and control period, and we found that the pedestrian flow patterns differed greatly in different periods by using a composite clustering index; we interpreted the clustering results from two perspectives: groups of pedestrian flow patterns and functional categories. The results show that depending on the specific stage of epidemic prevention and control, the number of unique pedestrian flow patterns decreased from four before the epidemic to two during the strict control stage and then increased to six during the initial resumption of work. The restrictions on movement are correlated with most of the visitations, and the release of restrictions led to an increase in the variety of unique pedestrian flow patterns compared to that in the pre-restriction period, even though the overall number of visitations decreased, indicating that social restrictions led to differences in the flow patterns of POIs and increased social distance.

Keywords: public health events; key urban areas; pedestrian flow change patterns; time series clustering

Citation: Li, Y.; Xu, L. The Impact of COVID-19 on Pedestrian Flow Patterns in Urban POIs—An Example from Beijing. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 479. <https://doi.org/10.3390/ijgi10070479>

Academic Editor: Wolfgang Kainz

Received: 28 May 2021

Accepted: 8 July 2021

Published: 11 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

COVID-19, as a major public health emergency [1,2], will undoubtedly have a huge impact on cities [3–8] due to their ultra-high population density [3,4], especially regarding pedestrian flow patterns [9,10], which represent changes in the number of people visiting a POI. There have been some important advances in research on pedestrian flow patterns, such as identifying urban functional areas [11] and detecting urban anomalies in human activities [12]. Studies on the impact of such major public events on cities, such as the recovery of urban vitality, have been attracting the attention of a wide range of relevant researchers [13–16]. Most of the studies on urban response and recovery after large disasters have focused on natural disasters [17–19]. Unlike natural disasters, public health events do not physically damage urban buildings and facilities but rather create internal fear and concerns that affect people's travel patterns.

To prevent the spread of the epidemic through person-to-person interaction, governments worldwide have issued lock-down policies to restrict mobility. They then implement a reopening policy when the situation improves. Taking Beijing as an example, we can see three stages of implementation of different policies from 17 January 2020 to 15 February 2020. The first stage is the normal stage without COVID-19. The second stage is the strict lock-down period with the closure of all public places, and the third stage is the resumption of work with the gradual reopening of public places. The effectiveness of these policies

is a meaningful subject of study [20–22]. The mobility restriction policies imposing undifferentiated mobility restrictions on certain urban functional areas or functional categories of POI in various regions are similar to those in Beijing. A hidden assumption of this approach is that the functional category of the POI is a reasonable basis for classification, that is, similar POIs should have similar flow patterns. This hidden assumption has not been confirmed [23], so the following questions are raised: Are mobility restrictions based on urban functional areas or functional categories of POIs reasonable? Is there a better implementation unit for mobility restrictions?

Research on the impact of the epidemic on mobility in cities can be divided into three broad sections: (1) Different regions are affected to different degrees: Li found that the social distance between different census block groups and urban hotspots changed significantly during the outbreak-induced population panic phase [24]. Chang et al. studied “super hotspots” in cities. They proposed that the proportion of pedestrians in hotspots in cities is one of the most critical indicators of urban response to pandemics [25]. Wang et al. analyzed the impact of epidemic and government control on residents’ park visitation patterns [26]. (2) Epidemic spread and travel behavior: Lee et al. used regression analysis to fit the relationship between the traffic volume of a region and the number of COVID-19 cases in different regions [27]. Parr et al. found that while overall traffic volumes were decreasing, the changes varied by class of road, and the impact on movements varied by region [28]. (3) Controlling mobility by orderly work resumption: Chang et al. found that limiting visitors to two types of POIs, restaurants and gyms, would minimize the spread of cases after reopening [29,30]. As the target of spatial analysis, these studies have mainly selected regions or functional categories of POIs [31], but regions and functional categories of POIs are not sufficient as explanatory variables for changes in urban mobility. For example, suppose that there are two restaurants, with one in a work area and the other near a mall; then the flow pattern of the restaurant near a mall is more likely to be similar to that of the mall, and it is more appropriate to treat the restaurant near a mall and the mall itself as one pattern group in the time-phased reopening policy. The nature of urban mobility in response to such shocks has been the subject of no previous analysis, and the epidemic shock and subsequent policy constitute a natural experiment for urban mobility [32].

These studies treated the functional category of a POI and the pedestrian flow pattern group of a POI as the same. However, these two concepts do not correspond to one another. The purpose of this study is to group all POIs from a pedestrian flow pattern perspective to obtain different patterns. This paper, therefore, addresses the following research questions: What kind of characteristics do POIs with different categories show in terms of flow patterns under various phases of the epidemic and policies? Do the flow patterns of POIs in the same functional category remain consistent? Are mobility restrictions based on urban functional areas or functional categories of POIs reasonable?

In this study, the POIs are grouped by time series clustering to obtain different stages of pedestrian flow patterns, and we then find a relationship between patterns of different periods and discuss how the degrees of recovery of different POIs differ. Methodologically, we propose a method to determine the number of POI pedestrian flow patterns in a metropolitan area. We then discuss the variation of the POI mobility pattern groups vis-à-vis the different stages of the COVID-19 epidemic and related mobility restriction policies (Figure 1). The significance of the study is not limited to epidemics. The essence of this study is how the structure and function of the city, as revealed by the pedestrian flow patterns of POI, change under the conditions of some sudden shock that imposes various mobility constraints of different types and phases. The period of this case study covers every phase of the epidemic’s impact, which has been realized in only a few studies so far. It also covers a variety of mobility restriction policies in Beijing, the capital of China, making it a suitable case for studying the impact of public emergencies on urban mobility and answering our research questions.

Figure 1. The framework of the study.

2. Materials and Methodology

2.1. Data

The data in the training set are related to the crowd flow in key urban areas of Beijing during the epidemic period from 17 January 2020 to 15 February 2020. The data provide the historical multiday sub-hourly grid (200×200 m) crowd density (using the people index to represent crowd density). The index is a measure of the number of people in the area, which is proportional to the number of people present in the area during a certain hour on a certain day; a larger index means more people are present in that place, and vice versa.

The data include almost all hotspot locations within the Sixth Ring Road of Beijing, as shown in Figure 2, and also cover the land use types of travel destinations within Beijing, mainly including the six major categories of Transit POIs, Medical care POIs, Education POIs, Recreation POIs, Shopping POIs, and Sports POIs; they are also subdivided into 14 subcategories, with a total of 997 hotspot locations, as shown in Table 1.

The division of the periods is based on the different measures taken by government departments for the prevention and control of COVID-19 and the corresponding notices issued. The data cover the period from 17 January 2020 to 15 February 2020, which is divided into 3 time periods in this paper: the first period indicates the stage before epidemic prevention and control: 17 January 2020 at 00:00 to 24 January 2020 at 00:00. The second segment represents the stage during epidemic prevention and control: 24 January 2020 at 00:00 h to 10 February 2020 at 00:00 h. The third period indicates the stage when the epidemic resumes after orderly control: 10 February 2020 at 0:00 h to 15 February 2020 at 24:00 h [33].

Figure 2. Distribution of major hotspot locations in Beijing.

Table 1. Classification table of major hotspot locations in Beijing.

Hotspot Major Categories	Hotspot Minor Categories	POI Number
Transit POI	Train Station	15
	Long-distance bus station	8
	Airports	1
Medical care POI	General Hospital	116
Education POI	Higher Education	186
Recreation POI	Park	290
	Zoo	2
	Monuments	42
	Botanical Garden	5
	Aquarium	2
	Amusement Park	4
	Scenic Area	39
Shopping POI	Shopping Center	177
Sports POI	Stadiums	110

2.2. Time Series Clustering

By clustering all hotspots according to the hourly people indices, we can group POIs after determining the number of clusters and further derive the corresponding patterns and characteristics of visitation.

Taking the pre-epidemic control phase as an example, each hotspot location was represented as an N-dimensional vector, where N denotes the time span of this phase; e.g., the pre-epidemic control phase is from 0:00 on 17 January 2020 to 0:00 on 15 February 2020, so N for this phase is 168. The clustering method used was K-means clustering [34]. The K value in the K-means clustering method is used as a hyperparameter, which is to be set empirically before clustering. Since all the hotspot locations to be classified involve a total of 14 sub-clusters based on functional categories, we started by setting K to 14. The results of the clustering are shown in Figure 3.

Since the POIs in the data were artificially divided into 6 major functional categories and 14 minor categories, the number of pedestrian flow patterns was initially set to 14, but the results show that most of the POIs in the same category were not clustered into one mobility pattern group. The results indicate that functional categories of POIs cannot be viewed as the basis for distinguishing groups of pedestrian flow patterns in POIs.

Figure 3. Cont.

Figure 3. Clustering results of the changing patterns of pedestrian flow before the epidemic.

The clustering algorithm we used is a semi-supervised machine learning method because the number of clusters K is a hyperparameter that needs to be set empirically, and a reasonable K value can cluster the data set into different subsets correctly [35]. On the one hand, the correct K value can distinguish different regional patterns of crowd flow changes; on the other hand, the correct K value can be used to prepare epidemic prevention work and other control work arrangements according to the temporal patterns of crowd flow and reasonably arrange limited epidemic prevention manpower and material resources. Recently, many researchers have conducted in-depth studies on the selection problem of the number of clusters K [36,37], which can be divided into two types of selection and

evaluation methods according to whether the data contain the real clustering labels. One is data clustering with real category labels. The K value selection problem starts from the misclassification difference of the results compared with the real categories [38]. The other type is used when the data do not contain the real category labels, and the starting point of this K value selection problem is the degree to which the data are correctly separated after clustering; indicators such as the similarity between data of the same category and the difference between data of different categories, or combined indicators calculated from several indicators, can be used as the basis for selection and judgment [39].

For the problem in this paper, although each POI has a clear functional category, such as transportation hub or tourist attraction, and a specific place category, such as railway station, park, etc., the data used for clustering are the time series data of the pedestrian flow [40]; thus, the time series data, although related to the land use type and place type, have no direct correlation with the category. Further, the data of the clustering problem in this paper do not contain real category labels, so for K value selection, we applied the second class of method described above. For clustering without real labels, the most common K value selection methods currently are Silhouette Coefficient, CH score discrimination, and DB index discrimination.

The Silhouette Coefficient Discriminant method is used to discriminate whether the K value selected for clustering is optimal by calculating the Silhouette coefficient of the results obtained from clustering. The Silhouette coefficient, first proposed by Peter J. Rousseeuw in 1987, contains the degree of similarity of similar data and the degree of difference between different clusters of data in the clustering results, and it can be used to calculate the merits of different K values in the clustering results based on the degree of aggregation of the data in the feature space distribution in the absence of the true cluster labels of the data [41]. This is done by calculating for each data item i the corresponding Silhouette coefficient $s(i)$; $s(i)$ consists of two terms, $a(i)$ and $b(i)$, where $a(i)$ is the average distance between sample i and other samples in the same cluster as sample i , and $b(i)$ denotes the distance between sample i and samples from other clusters. In greater detail, $a(i)$ denotes the distance among samples clustered into the same cluster, i.e., the intra-cluster difference degree. For all samples x of a cluster, we calculate the corresponding $a(x)$; the mean of all $a(x)$ values is denoted the intra-cluster variance of the cluster. Assuming that sample j belongs to the same cluster A as sample i in the clustering result, we have

$$a(i) = \frac{\sum_{j \in A} \|i - j\|_2^2}{A_n} \quad (1)$$

where A_n denotes the number of samples in category A , and $\|i - j\|_2^2$ denotes the squared Euclidean distance between sample i and sample j . A smaller $a(i)$ value indicates a higher probability that i belongs to category A .

$b(i)$ is the minimum value in the average distance of sample i from samples in other clusters:

$$b(i) = \min_{C \neq A} \left(\frac{\sum_{j \in C} \|i - j\|_2^2}{C_n} \right) \quad (2)$$

where C_n denotes the number of samples in category C . A larger $b(i)$ value indicates a higher probability that sample i belongs to category A . The Silhouette Coefficient $s(i)$ of sample i is defined based on the degree of similarity $a(i)$ of similar data and the degree of difference $b(i)$ between different clusters of data:

$$s(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, & a(i) < b(i) \\ 0, & a(i) = b(i) \\ \frac{a(i)}{b(i)} - 1, & a(i) > b(i) \end{cases} \quad (3)$$

For sample i , the Silhouette coefficient $s(i)$ has a range of $[-1, 1]$. The closer $s(i)$ is to 1, the higher the probability that sample i is clustered into the correct category, and the closer $s(i)$ is to -1 , the higher the probability that it is clustered into the wrong category. For the K value chosen for clustering, the corresponding Silhouette coefficient is the mean s of the Silhouette coefficient of all samples obtained after clustering them into K -many clusters. As with the comparison method for individual samples, the larger the s value corresponding to the K value, the closer the samples clustered into the same cluster in the clustering result, the more distant the samples clustered into different clusters, and the more reasonable the choice of the K value for clustering. Since the mean value of the Silhouette coefficient is used as the basis, the method is very sensitive to data outliers, resulting in potentially unreasonable results. According to the results in Figure 4, and based on the premise that the larger the value, the better the clustering result, the pattern of pedestrian flow should be divided into two types, regardless of the stage of the epidemic.

Figure 4. Relationship between the Silhouette coefficient and K values at different stages of the epidemic.

The CH coefficient discriminant is another common coefficient for selecting the K value of clusters and was proposed by Calinski and Harabaz in 1974, so it is called the Calinski–Harabaz coefficient [42], or the CH coefficient for short. The CH coefficient is similar to the Silhouette coefficient and also contains two components, one indicating the degree of closeness between samples assigned to the same cluster and the other indicating the degree of dispersion between samples assigned to different clusters [43]. The degree of closeness W_K between samples of the same cluster is calculated by using the sum of the squared Euclidean distances of individual samples from the same cluster to the centroids of the cluster, and the degree of dispersion B_K between samples from the clusters is calculated by using the sum of the squared Euclidean distances of the centroids of the samples from each cluster to the centroids of the full sample space:

$$W_K = \sum_{q=1}^K \sum_{x \in C_q} \|x - x_q\|_2^2 \quad (4)$$

$$B_K = \sum_{q=1}^K n_q \|x_q - c\|_2^2 \quad (5)$$

In Equation (4), K is the number of clusters, C_q is the set of samples of cluster q , x_q is the sample centroid of C_q , n_q is the number of samples in C_q , and c is the centroid of the full sample space. The CH coefficient can then be calculated from W_K and B_K :

$$CH_{\text{index}} = \frac{B_K}{W_K} \times \frac{m - K}{K - 1} \quad (6)$$

In Equation (6), m is the total number of all samples, and K is the number of clusters. The better the clustering result, the smaller the W_K value and the larger the B_K value, with a larger resulting CH coefficient. Thus, when the CH coefficient is used as the judgment, a larger value indicates a better clustering result.

Compared with the Silhouette coefficient, the CH coefficient has advantages of low computational complexity and very fast computation speed, but it has the same disadvantages as the Silhouette coefficient. Since the calculation of the CH coefficient is based on the Euclidean distance between samples, it can reflect the clustering effect better for convex data set clustering, but it cannot be fully applied for a non-convex time series data set space as the data set will be divided into particularly fine pieces; it needs to be used together with other coefficients to judge the clustering results. According to the results in Figure 5, and based on the premise that the larger the value, the better the clustering result, the pattern of pedestrian flow should be divided into more than 10 groups, regardless of the stage of the epidemic.

Figure 5. Relationship between the CH coefficient and K values at different stages of the epidemic.

The DB index is another common index to discriminate clustering results, also known as the classification correctness index, and was first proposed by D.L. Davies and D.W.

Bouldin in 1979 to evaluate the accuracy of clustering results [44]. In this paper, the dimension of the sample space is the time span of the temporal change in pedestrian flow, all samples are clustered into k -many clusters, and the similarity degree S_i of similar samples is also defined in the DB index:

$$S_i = \left(\frac{1}{T_i} \sum_{j=1}^{T_i} \|X_j - A_i\|^q \right)^{\frac{1}{q}} \quad (7)$$

In Equation (7), T_i denotes the number of samples in cluster i in the clustering result, X_j denotes the j th sample in the set of samples in cluster i , and A_i denotes the centroid of the samples in cluster i . S_i denotes the q th root of the mean of the moments of order q of the samples in cluster i from the centroid of the cluster. If q takes the value of 1, S_i represents the mean of the Euclidean distance between each sample and the centroid of the cluster, and if q takes the value of 2, S_i represents the standard deviation of the distance between the samples of the cluster and the centroid of the de-sampled space, where either the 1-norm distance or the 2-norm distance, i.e., the Euclidean distance, can be used as the distance measure in computing the high-dimensional nonconvex sample space. Although the Euclidean distance is not the optimal choice, the default distance measure generally is the Euclidean distance [45].

The DB index similarly defines the difference between clusters, i.e., the distance M_{ij} , between sample sets of different clusters:

$$M_{ij} = \left(\sum_{k=1}^N \|a_{ki} - a_{kj}\|^p \right)^{\frac{1}{p}} \quad (8)$$

M_{ij} represents the distance between the sample set in cluster i and the sample set in cluster j . a_{ki} is the k th dimension of the centroid of the sample set in cluster i , a_{kj} is the k th dimension of the centroid of the sample set in cluster j , and N is the total dimension of the samples. The DB index is based on the similarity degree S_i of similar samples and the distance M_{ij} between the sets of samples in different clusters, defining the similarity degree R_{ij} between different clusters:

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \quad (9)$$

Combined with R_{ij} , for each cluster i , the DB index calculates the similarity R_{ir} of cluster i to the closest cluster r to that cluster, denoted D_i :

$$D_i = \max_{j \neq i} R_{ij} \quad (10)$$

D_i represents the closest degree to another cluster for cluster i . The DB index DB_{index} of the clustering result is obtained by calculating the mean of the D_i values for the set of samples of all clusters:

$$DB_{\text{index}} = \frac{1}{K} \sum_{i=1}^K D_i \quad (11)$$

K in Equation (11) indicates the clustering number K . The DB index indicates the extent to which the clustering boundaries are not clearly defined in the clustering results, so the smaller the value is, the better the indicated clustering results, to some extent. Since the calculation of the DB index involves the calculation of similarity between any two clusters, it is not reasonable when there are outlier clusters in the clustering results; in this case, it must be used together with other coefficients to judge the accuracy of the clustering results. The DB index turns the target of comparison from samples into a cluster and therefore tends to a smaller K value when minimizing DB_{index} , which can also be seen in Figure 6.

Figure 6. Relationship between the DB index and K values at different stages of the epidemic.

2.3. Composite Clustering Index

Combining the Silhouette coefficient, CH coefficient, and DB coefficient is a better method to determine the clustering K value. According to the introduction above, the larger the Silhouette coefficient and CH coefficient and the smaller the DB coefficient, the better the clustering results, so we propose to combine the three indices into a mixed index. Due to the problem of inconsistency in the magnitude of different coefficients, after normalizing all coefficients, we added the normalized Silhouette coefficient to the normalized CH coefficient and then subtracted the normalized DB coefficient to obtain the mixed index. The larger the mixed index, the better the clustering results; on the contrary, the smaller the mixed index, the worse the clustering results. In this paper, the chosen normalization method was Min-Max normalization:

$$x_{\text{new}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (12)$$

Thus, all the indices take values in the range $[-1, 1]$, ensuring dimension unification of the different coefficients in the calculation:

$$\text{Fix}_{\text{index}} = S_{\text{new}} + \text{CH}_{\text{indexnew}} - \text{DB}_{\text{indexnew}} \quad (13)$$

The fixed clustering index has the following advantages: the method takes into account the similarity of both the original data and the clustered data, and it reduces the influence of outliers in the results on the choice of the K value.

It can be seen from Figure 7 that, based on the premise that the larger the mixed index, the better the clustering results, the flow patterns of POIs should be divided into 4 groups in the pre-epidemic stage, as in Figure 7a; the flow patterns of POIs should be divided into 2 groups during the epidemic, as in Figure 7b; and the flow patterns of all POIs should be divided into 6 groups in the initial resumption of work stage, as in Figure 7c. The number of subgroups shown in Figure 7 is consistent with our estimate of the actual situation,

where the flow pattern was reduced from four to two due to COVID-19 and government restrictions on mobility. In the initial phase of the restart, people need to meet travel needs that were previously unmet due to mobility restrictions, and the number of pedestrian flow patterns thus increased from two to six.

Figure 7. Relationship between the Fixed Clustering Index and K values at different stages of the epidemic.

3. Results and Discussion

This section shows the groups of POIs for different phases of pedestrian flow patterns obtained via the clustering algorithm detailed above and the clustering results interpreted from two perspectives: groups of pedestrian flow patterns and functional categories. This section analyzes the impact of COVID-19 on mobility in the different periods and the relationships between the pedestrian flow patterns in different periods from the results.

3.1. Results

This part shows the clustering results of pedestrian flow in Beijing hotspot locations in different stages and briefly explains the results.

Figure 8 shows the four patterns of pedestrian flow obtained according to the mixed clustering index before the start of the epidemic. Different colors indicate different functional categories A of POIs. The corresponding relationship between color and POI can be found in Figure 4. The first group of POIs represents POIs with a high number of pedestrians and high concentration; the second group of POIs represents POIs with a lower number of pedestrians and lower concentration; and the third and fourth groups are POIs with ultra-high numbers of pedestrians and ultra-high concentration, mainly railway stations and airports. The fourth group of POIs contains only one POI, which corresponds to Beijing Capital Airport.

Figure 8. Clustering results of four changing patterns of pedestrian flow in Beijing hotspot locations before the epidemic.

Figure 9 shows that the pedestrian flow patterns of all the POIs except the second group of airport POIs were grouped into the first group during the epidemic, due to the government's mobility restriction policy. Almost all POIs showed low traffic and low concentration patterns, and although the pattern for the airport in the second group was still high, it was reduced by more than half compared to that shown in Figure 8. Since the airport is the most important international transportation hub, the pedestrian flow in the airport was affected differently compared with that in other POIs and has always been regarded as a separate flow pattern group.

Figure 9. Clustering results of two patterns of changes in pedestrian flow at hotspot locations in Beijing during the epidemic.

Figure 10 illustrates the divergence of pedestrian flow patterns in the city during the initial resumption phase as a result of the government's reopening policy. The first group of POIs is mainly Medical care POIs, which can be interpreted as those returning to work needing to be tested by the hospital for viruses before the restriction could be lifted, so this

type of pattern was separated. The second group of POIs is mainly proximity Recreation and shopping POIs, as residents needed to make up for their exercise needs that went unmet during home time and necessary daily shopping. The third group of POIs is mainly college POIs, and the fourth group is mainly large shopping POIs, both of which have significant tidal flow patterns. The fifth group of POIs contains only three POIs, mainly due to their significantly higher pedestrian density. Although the three POIs are different in type, their similar pedestrian flow patterns indicate their importance in preventing and controlling epidemics. The sixth group of POIs contains only one POI, which corresponds to Beijing Capital Airport.

Figure 10. Clustering results of six patterns of changes in pedestrian flow at Beijing hotspot locations during the preliminary resumption of work.

3.2. Discussion

Here, we analyze the impact of COVID-19 on urban mobility in different periods and the relationship that exists between the pedestrian flow patterns in different periods from the clustering results.

From the perspective of different groups of pedestrian flow patterns, there are several overlapping patterns of pedestrian flow, as shown in Figure 8a,b. Overlapping pat-

terns indicate that POIs with different functional categories have similar flow patterns at this stage, as shown in Figure 8a, mainly including Shopping, Education, and Medical care. The high number of pedestrians and high concentration mean that these POIs in Figure 8a have greater appeal and more specific functions. Not only can these POIs attract more visitors, but visitors are more inclined to visit them in the same period, which causes this cluster of high flow and high concentration. The exciting thing is that the pedestrian flow curves in Figure 8b exactly fill in the blank part of the bottom half of Figure 8a. The low number of pedestrians and low concentration mean that these POIs in Figure 8b have less attractiveness and more diverse functions, even though there are POIs of the same category in Figure 8a,b. Due to their ultra-high numbers of pedestrians, almost all traffic hubs are in the third and fourth clusters, with consistent behavior in the flow pattern.

Different from those in Figure 8, except that the pedestrian flow pattern in Figure 10b still contains multiple functional categories of POI, the flow patterns of the three groups shown in Figure 10a,c,d become more specific. It can be seen from Figure 10a,c,d that the colors of the three groups are more prominent than those in Figure 10b, which shows that the multifunctional POIs that had previously attracted many kinds of flow became more specific at this stage. Due to the high index of people, the airport was still grouped separately, as shown in Figure 10f.

From the perspective of functional categories of POIs, Figure 11 shows the flow patterns in different functional categories of POIs during the pre-epidemic period. The shape of the elements in the figure represents the POI categories, and different colors represent different flow patterns. The Transit POI contains the most significant number of patterns, followed by the Recreation POI, and the special one is the Sports POI, which includes only one flow mode. According to the results mentioned above, as shown in Figure 12, the flow patterns of all categories of POIs except airports were compressed into one pattern during the epidemic. Figure 13 shows the flow patterns contained in the different categories of POIs during the initial resumption of work. Although the number of pedestrians in all categories of POIs decreased significantly compared to that in the pre-epidemic period, the number of flow patterns in all categories of POIs increased compared to that in the pre-epidemic period. This result indicates that residents not only travel less than before the epidemic but also congregate less when they travel, which can significantly increase the social distance between people.

Figure 11. Patterns of changes in pedestrian flow at different hotspot locations in Beijing during the pre-epidemic period.

Figure 12. Patterns of changes in pedestrian flow at different hotspot locations in Beijing during the epidemic.

Figure 13. Patterns of changes in pedestrian flow at different hotspot locations in Beijing during the preliminary resumption of work.

Figure 14 illustrates the main categories of POIs contained in different POI groups based on pedestrian flow patterns in the different stages. It can be seen from Figure 14 that the crowd flow patterns of different categories of POIs were adjusted in different phases of the epidemic. In line with common sense, what is obvious is that under the impact of the epidemic, the daily flow patterns were compressed into one, namely, a low-traffic and low-aggregation pedestrian flow pattern, from the pre-epidemic period to the epidemic period. What is interesting is that there were more patterns of pedestrian flow, although the number of pedestrians after the release of the government's reopening policy was significantly lower than it was in the pre-epidemic period. From Figure 14, it can be

seen that the first group of POIs (Shopping, Education, Medical care) before the epidemic was divided into three groups in the initial resumption of work stage, with Medical care, Shopping, and Education as the main POI functional categories; this indicates that under the influence of the epidemic, people only make necessary trips, making the pattern of aggregation in different POIs more differentiated and leading to a more obvious division of the city into functional areas. The first batch of reopened POIs included two categories: Transit and Shopping. The most noticeable changes were in Medical care and Shopping. The flow patterns of these two POI categories were clearly separated, which also indicates that Medical care and Shopping were essential travel destinations in the post-epidemic phase. The two POI flow pattern groups were separated from one flow pattern into two. Taking Shopping POIs as an example, the flow pattern was divided into two: a pedestrian flow pattern with a low number of pedestrians and low concentration, and a pattern with a high number of pedestrians and high concentration. This result shows that the policy of resumption of work led to changes in Shopping flow patterns. Some Shopping locations received small but continuous traffic, while others received large instantaneous traffic but few people in most cases. Although shopping stores are all categorized as Shopping, different stores are not the same in terms of flow patterns or changes in patterns under the influence of the epidemic; these need to be treated differently when implementing policies. Due to data limitations, we can only speculate that such separation is caused by the functional and locational factors of the POI's service delivery.

Figure 14. Shifting traffic flow patterns in urban hotspots at different stages.

It can also be seen that the second group (Recreation, Education, Sports) in the initial resumption phase corresponds to the second group (Recreation, Education, Shopping) before the epidemic, where the main POI type is Recreation; this includes parks in the city with a low number of pedestrians and low concentration, indicating a gradual resumption of outdoor recreation and entertainment. The third group (Transit, Recreation), in which the Transit POI was the central part before the epidemic, is no longer evident after the epidemic, also indicating that intra-city traffic has still not recovered. The Transit POI group at the bottom of Figure 14 is not grouped with other POIs, regardless of the stage of the epidemic impact. This group of Transit POIs contains only one POI, the capital airport, due to the high number and concentration of pedestrians at the airport that cannot

be reached by other POIs, which is also confirmed in Figures 9–11. For Sports POIs, in the early stage of the epidemic, there was no obvious flow pattern characteristic, so there were no Sports POIs in each category of the clustering results, but in the early stage of the resumption of work, its flow pattern appeared in the second category of results, indicating that the epidemic made people consistently pay more attention to exercise to improve self-resistance.

This study has some limitations because the analysis requires a long time series of pedestrian flow data from POIs for support. Due to the large temporal and spatial scales involved, such data generally need to be collected by the government or mobile communication service providers and obtained through applications. We divided the different periods of the epidemic based on government notification documents. This simplifies its impact to a direct impact, which in practice would be phased and take effect step by step; this will be the focus of subsequent research. We also need to consider other kinds of visualization in future work to improve user perception [46–50]. Due to the limitations of the data, we only found a separation of pedestrian flow patterns during the resumption period, and more geographic data will be introduced to explain the cause of this phenomenon in future work.

4. Conclusions

As one of the biggest public health and safety events in recent years, COVID-19 has had a considerable impact on the whole world. As governments introduced corresponding prevention, control, and isolation measures, the epidemic crisis in the vast majority of regions was greatly alleviated, and all of society gradually returned to normal operation. From normal operation to the sudden impact of the epidemic on the restoration of production order, almost all regions have gone through these three stages mentioned above. People adopt different travel patterns at different time stages, leading to different patterns of pedestrian flow gathering in different POIs.

In conclusion, this study makes the following contributions. One contribution is a method to obtain POI groups with similar pedestrian flow patterns based on time series clustering. Our methodology can reasonably calculate the number of POI flow patterns in a city in a certain period and obtain POI groups with similar patterns accordingly. More importantly, this study showed that it is more appropriate to use groups of POIs with similar pedestrian flow patterns as the unit of study, rather than functional categories of POIs. The contribution in applications is the analysis of urban mobility changes in cities under the influence of specific sudden shocks from the perspective of flow patterns. Unlike previous studies that distinguished POIs by functional category, this paper provides a new perspective to differentiate groupings using the pedestrian flow patterns of POIs, which is more applicable to reopening policies by region and period, and we analyzed the impact of COVID-19 on almost the entire process in Beijing from the beginning until the initial resumption of work. During the initial resumption phase, the number of pedestrians was significantly lower than that before the epidemic, but the number of flow patterns was higher, indicating that social restrictions led to differences in the flow patterns of POIs and increased social distance. Therefore, we believe that the policy of imposing mobility restrictions based on POI functional categories is limited. Grouping POIs according to their mobility characteristics as the basic unit for implementing mobility restrictions is better. This study provides an operational approach for this.

Author Contributions: Yihang Li performed the research, analyzed the data, and wrote the paper. Liyan Xu supervised the study. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Restrictions apply to the availability of these data. The data are not publicly available due to privacy concerns.

Acknowledgments: The authors would like to thank Hang Yin at Peking University, and Fu Li at Peking University, for their suggestions and technical guidance. The authors are sincerely grateful for the comments of the anonymous reviewers and members of the editorial team.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Leung, K.; Wu, J.T.; Liu, D.; Leung, G.M. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: A modelling impact assessment. *Lancet* **2020**, *395*, 1382–1393. [CrossRef]
2. Chakraborty, I.; Maity, P. COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Sci. Total Environ.* **2020**, *728*, 138882. [CrossRef]
3. Pareek, M.; Bangash, M.N.; Pareek, N.; Pan, D.; Sze, S.; Minhas, J.S.; Hanif, W.; Khunti, K. Ethnicity and COVID-19: An urgent public health research priority. *Lancet* **2020**, *395*, 1421–1422. [CrossRef]
4. Dorn, A.V.; Cooney, R.E.; Sabin, M.L. COVID-19 exacerbating inequalities in the US. *Lancet* **2020**, *395*, 1243–1244. [CrossRef]
5. Yancy, C.W. COVID-19 and African Americans. *J. Am. Med. Assoc.* **2020**, *323*, 1891–1892. [CrossRef] [PubMed]
6. Chowkwanyun, M.; Reed, A.L., Jr. Racial Health Disparities and Covid-19—Caution and context. *N. Engl. J. Med.* **2020**, *383*, 201–203. [CrossRef] [PubMed]
7. Wilson, C. These Graphs Show How COVID-19 Is Ravaging New York City’s Low-Income Neighborhoods. *Time*. 2020. Available online: <https://time.com/5821212/coronavirus-low-income-communities/> (accessed on 9 July 2021).
8. Garg, S.; Kim, L.; Whitaker, M.; O’Halloran, A.; Cummings, C.; Holstein, R.; Prill, M.; Chai, S.J.; Kirley, P.D.; Alden, N.B.; et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019—COVID-NET, 14 states, 1–30 March 2020. *MMWR Morb. Mortal. Wkly Rep.* **2020**, *69*, 458–464. [CrossRef] [PubMed]
9. Stier, A.J.; Berman, M.G.; Bettencourt, L.M.A. COVID-19 attack rate increases with city size. *arXiv* **2020**, arXiv:2003.10376.
10. World Health Organization. WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19. WHO. 2020. Available online: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (accessed on 9 July 2021).
11. Wu, L.; Cheng, X.; Kang, C.; Zhu, D.; Huang, Z.; Liu, Y. A framework for mixed-use decomposition based on temporal activity signatures extracted from big geo-data. *Int. J. Digit. Earth* **2018**, *13*, 708–726. [CrossRef]
12. Cheng, X.; Wang, Z.; Yang, X.; Xu, L.; Liu, Y. Multi-scale detection and interpretation of spatio-temporal anomalies of human activities represented by time-series. *Comput. Environ. Urban Syst.* **2021**, *88*, 101627. [CrossRef]
13. Chhotray, V.; Few, R. Post-Disaster Recovery and Ongoing Vulnerability: Ten Years after the Super- Cyclone of 1999 in Orissa. *Glob. Environ. Chang.* **2012**, *22*, 695–702. [CrossRef]
14. Olshansky, R.B. How Do Communities Recover from Disaster? A Review of Current Knowledge and an Agenda for Future Research. In Proceedings of the 46th Annual Conference of the Association of Collegiate Schools of Planning, Kansas City, MO, USA, 27 October 2005.
15. Cuttone, A.; Lehmann, S.; González, M.C. Understanding predictability and exploration in human mobility. *EPJ Data Sci.* **2018**, *7*, 1–17. [CrossRef]
16. Hamedmoghadam, H.; Ramezani, M.; Saberi, M. Revealing latent characteristics of mobility networks with coarse-graining. *Sci. Rep.* **2019**, *9*, 1–10.
17. Aldrich, D.P. It’s Who You Know: Factors Driving Recovery from Japan’s 11 March 2011 Disaster. *Public Adm.* **2016**, *94*, 399–413. [CrossRef]
18. Rufat, S.; Tate, E.; Burton, C.G.; Maroof, A.S. Social Vulnerability to Floods: Review of Case Studies and Implications for Measurement. *Int. J. Disaster Reduct.* **2015**, *14*, 470–486. [CrossRef]
19. Wu, J.; Li, N.; Xie, W.; Zhou, Y.; Ji, Z.; Shi, P. Post-Disaster Recovery and Economic Impact of Catastrophes in China. *Earthq. Spectra* **2014**, *30*, 1825–1846. [CrossRef]
20. Chinazzi, M.; Davis, J.T.; Ajelli, M.; Gioannini, C.; Litvinova, M.; Merler, S.; Piontti, A.P.; Mu, K.; Rossi, L.; Sun, K.; et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **2020**, *368*, 395–400. [CrossRef]
21. Lau, H.; Khosrawipour, V.; Kocbach, P.; Mikolajczyk, A.; Schubert, J.; Bania, J.; Khosrawipour, T. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *J. Trav. Med.* **2020**, *27*, taaa037. [CrossRef]
22. Zhang, Y.; Zhang, A.; Wang, J. Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China. *Transport Pol.* **2020**, *94*, 34–42. [CrossRef]
23. Zhou, Y.; Fang, Z.; Thill, J.C.; Li, Q.; Li, Y. Functionally critical locations in an urban transportation network: Identification and space-time analysis using taxi trajectories. *Comput. Environ. Urban Syst.* **2015**, *52*, 34–47. [CrossRef]
24. Li, Q.; Bessell, L.; Xiao, X.; Fan, C.; Gao, X.; Mostafavi, A. Disparate Patterns of Movements and Visits to Points of Interests Located in Urban Hotspots across U.S. Metropolitan Cities during COVID-19. *R. Soc. Open Sci.* **2020**, *8*, 201209. [CrossRef]
25. Chang, S. Mobility network modeling explains higher SARS-CoV-2 infection rates among disadvantaged groups and informs reopening strategies. *MedRxiv* **2020**. [CrossRef]
26. Wang, D.; Geng, D.; Innes, J.; Wu, W. Impacts of COVID-19 pandemic on urban park visitation: A global analysis. *J. For. Res.* **2020**, *32*, 553–567.

27. Lee, H.; Park, S.J.; Lee, G.R.; Kim, J.E.; Lee, J.H. The Relationship between the COVID-19 Prevalence Trend and Transportation Trend in South Korea. *Int. J. Infect. Dis.* **2020**, *96*, 399–407. [CrossRef] [PubMed]
28. Parr, S.; Wolshon, B.; Renne, J.; Murray-tuite, P.; Asce, A.M.; Kim, K. Traffic Impacts of the COVID-19 Pandemic: Statewide Analysis of Social Separation and Activity Restriction. *Am. Soc. Civ. Eng.* **2020**, *21*, 04020025. [CrossRef]
29. Chang, S.; Pierson, E.; Koh, P.W.; Gerardin, J.; Redbird, B.; Grusky, D.; Leskovec, J. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **2021**, *589*, 82–87. [CrossRef] [PubMed]
30. Benzell, S.G.; Collis, A.; Nicolaides, C. Rationing social contact during the COVID-19 pandemic: Transmission risk and social benefits of US locations. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 14642–14644. [CrossRef] [PubMed]
31. Nian, G.; Peng, B.; Sun, D.J.; Ma, W.; Peng, B.; Huang, T. Impact of COVID-19 on urban mobility during post-epidemic period in megacities: From the perspectives of taxi travel and social vitality. *Sustainability* **2020**, *12*, 7954. [CrossRef]
32. Honey-Roses, J.; Anguelovski, I.; Bohigas, J.; Chireh, V.; Daher, C.; Koni-jnendijk, C.; Oscilowicz, E. The Impact of COVID-19 on Public Space: A Review of the Emerging Questions. 2020. Available online: https://www.researchgate.net/publication/340819529_The_Impact_of_COVID-19_on_Public_Space_A_Review_of_the_Emerging_Questions (accessed on 9 July 2021).
33. Ge, H.; Liu, N. Modeling of emergency materials allocation decision-making problems based on the evolution scenarios of serious infectious disease: A case of COVID-19. *J. Ind. Eng. Eng. Manag.* **2020**, *34*, 214–222.
34. Ahmed, M.; Seraj, R.; Islam, S.M.S. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics* **2020**, *9*, 1295. [CrossRef]
35. Petitjean, F.; Ketterlin, A.; Gançarski, P. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognit.* **2011**, *44*, 678–693. [CrossRef]
36. Saha, J.; Mukherjee, J. CNAK: Cluster number assisted K-means. *Pattern Recognit.* **2021**, *110*, 107625. [CrossRef]
37. Cuturi, M.; Blondel, M. Soft-DTW: A differentiable loss function for time-series. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; ICML: Vienna, Austria, 2017; pp. 1483–1505.
38. Safari, A.; Nabil, K.; Biswajit, S. Machine Learning Based Classification Accuracy of Encrypted Service Channels: Analysis of Various Factors. *J. Netw. Syst. Manag.* **2021**, *29*, 1–27. [CrossRef]
39. Fränti, P.; Sieranoja, S. How much can k-means be improved by using better initialization and repeats? *Pattern Recognit.* **2019**, *93*, 95–112. [CrossRef]
40. Paparrizos, J.; Gravano, L. K-shape: Efficient and accurate clustering of time series. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Melbourne, Australia, 31 May–4 June 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 1855–1870. [CrossRef]
41. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
42. Calinski, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **1974**, *3*, 1–27.
43. Gómez-Adorno, H.; Aleman, Y.; Vilariño Ayala, D.; Sanchez-Perez, M.A.; Pinto, D.; Sidorov, G. Author Clustering using Hierarchical Clustering Analysis. In Proceedings of the Working Notes of the Conference and Labs of the Evaluation Forum (CLEF), Dublin, Ireland, 11–14 September 2017.
44. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *2*, 224–227. [CrossRef]
45. Xiao, J.; Lu, J.; Li, X. Davies Bouldin Index based hierarchical initialization K-means. *Intell. Data Anal.* **2017**, *21*, 1327–1338. [CrossRef]
46. Hadjidemetriou, G.M.; Sasidharan, M.; Kouyialis, G.; Parlikad, A.K. The impact of government measures and human mobility trend on COVID-19 related deaths in the UK. *Transp. Res. Interdiscip. Perspect.* **2020**, *6*, 100167.
47. Cintia, P. The relationship between human mobility and viral transmissibility during the Covid-19 epidemics in Italy. *arXiv* **2020**, arXiv:2006.03141.
48. Christóvão, R.; Marcílio Júnior, W.; Eler, D. PlaceProfile: Employing Visual and Cluster Analysis to Profile Regions based on Points of Interest. In Proceedings of the 23rd International Conference on Enterprise Information Systems—Volume 1: ICEIS, Setúbal, Portugal, 26–28 April 2021; pp. 506–514, ISBN 978-989-758-509-8. [CrossRef]
49. D’Andrea, E.; Ducange, P.; Loffreno, D.; Marcelloni, F.; Zacccone, T. Smart Profiling of City Areas Based on Web Data. In Proceedings of the 2018 IEEE International Conference on Smart Computing (SMARTCOMP), Taormina, Italy, 18–20 June 2018; pp. 226–233. [CrossRef]
50. Keim, D.A.; Panse, C.; Sips, M. Visual Data Mining of Large Spatial Data Sets. In *Databases in Networked Information Systems. DNIS 2003*; Lecture Notes in Computer Science; Bianchi-Berthouze, N., Ed.; Springer: Berlin/Heidelberg, Germany, 2003; Volume 2822. [CrossRef]