Computer and Decision Making – An International Journal Volume 2, 2025, 357-373

Computer and Decision Making
An International Journal

Journal homepage: www.comdem.org
eISSN: 3008-1416

Data-driven Analysis of Taxi and Ride-hailing Services: Case Study in Chengdu, China

Weiwei Jiang[1,*]

[1] School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China. (Correspondence: jww@bupt.edu.cn)

ARTICLE INFO

Article history:
Received 06 November 2024
Received in revised form 15 December 2024
Accepted 01 January 2025
Available online 03 January 2025

Keywords:
Big Data Analysis; Transportation; Taxi; Ride-hailing

ABSTRACT

Both taxi and Internet-based ride-hailing services are important public transportation options. While these services provide a similar functionality, their quantitative patterns are difficult to describe and analyze without a big data approach. In this study, we present a data-driven analysis for taxi and ride-hailing services with a case study in Chengdu, China. Both a taxi GPS dataset and an Internet-based ride-hailing trip dataset provided by Didi Chuxing, the largest transportation network company in China, are used in this study. Our analysis is based on four aspects, i.e., temporal patterns, spatial patterns, spatio-temporal patterns, and traveling distance patterns. It is found that both taxi and ride-hailing services exhibit the densification power law with a space-time graph model. The ride-hailing service has a preference for more concentrated pickup and dropoff hotspots and trips with a longer distance. The observed results indicate that the ride-hailing service partially plays the role of taxis, but it also has its own characteristics as a new transport option. The findings in this study would be helpful for the government to better regulate the operation of both taxi and ride-hailing services.

1. Introduction

Urban transportation is a fundamental facility for modern cities and affects the normal operation of citizens. The ability to analyze the efficiency of different transport services is the basis for building an intelligent transportation system. Driven by the combination of data science [35], urban science, and information technologies, information and technology infrastructures are set up in cities to monitor the transportation system status and collect traffic data through sensors or cameras. Data-driven analysis for transportation services is becoming more important than ever before, especially with Internet big data [39] and traffic big data [14, 17], and artificial intelligence techniques [18].

Big data and machine learning play transformative roles in the intelligent transportation domain, enabling real-time decision-making, predictive analysis, and system optimization [1, 26]. By analyzing vast amounts of data from sensors, GPS devices, cameras, and social platforms, machine learning algorithms can identify traffic patterns, optimize route planning, and reduce congestion [10]. Pre-

https://doi.org/10.59543/comdem.v2i.11363

dictive maintenance models enhance vehicle reliability by forecasting potential breakdowns, while autonomous driving systems leverage machine learning for object detection, path planning, and navigation. Additionally, these technologies support the development of smart urban mobility solutions, such as ride-sharing and public transport optimization, fostering sustainable and efficient transportation networks. Together, big data and machine learning enhance safety, efficiency, and environmental sustainability in modern transportation systems.

In this study, two specific transport services are analyzed with a data-driven approach, namely, taxi and ride-hailing, with traffic big data collected in Chengdu, China [28]. Taxis play an important role in urban transportation as an efficient tool for individual travel. However, the number of taxis is usually regulated by the government and could be far from meeting the potential travel demand [30]. For example, based on the statistical yearbook of Chengdu, there were only 18,506 taxi drivers in 2014, serving more than 12 million people in Chengdu, China [4]. It would be difficult to call a taxi ride during rush hours because of the imbalance between the high demand and the limited supply.

In the past few years, transportation network companies (TNCs) have arisen as a new transport choice that leverages mobile Internet and location-based services to connect drivers with private cars and passengers with smartphone applications. These companies, e.g., Uber, Lyft, and Didi Chuxing, provide various transportation services, including taxi-hailing, ride-hailing, and ride-sharing, both on weekends and holidays [15]. Unlike the method of waiting and searching for an empty taxi on the street, a passenger could simply send the trip request with the smartphone application, and a nearby driver would be dispatched based on the passenger's location. In addition to the convenient requesting method, transportation network companies have the advantages of flexible supplies and independent pricing rights, which means that they can increase supplies in hot areas and peak hours. All of these advantages have made them successful in the past few years.

While these TNCs are successful, there are few studies comparing their services with taxis, especially from the data-driven perspective. The similarities and differences between the trips served by TNCs and traditional taxis have not been fully explored [22, 24, 37]. In this study, the travel patterns of taxi and ride-hailing services are analyzed and compared, with similarities and differences revealed. Two large traffic datasets collected in Chengdu, the capital of China's Sichuan province, are used in this study, which consist of a taxi GPS dataset and a ride-hailing trip dataset provided by Didi Chuxing.

The taxi GPS dataset was collected in August 2014, when there was no ride-hailing service operating widely in Chengdu at that time. This gives us the chance of studying the basic travel patterns of taxis before the interference of ride-hailing services. The ride-hailing trip dataset was collected in November 2016, when Didi Chuxing was already the market leader in China, and its data are the most suitable representative for the situation of ride-hailing services. Some statistical and analytical models from data science, e.g., the space-time graph model and the non-negative matrix factorization method, are used in this study. The key findings in this paper are summarized as follows:

- Taxi and ride-hailing services exhibit different temporal patterns during morning rush hours, evening rush hours, and late in the evening;

- The pickup and dropoff hot areas overlap, but the taxi trips are distributed more sparsely, which indicates that taxis have a wider spatial coverage;

- Taxi and ride-hailing services have similar basis collective patterns revealed by a non-negative matrix factorization method, and both exhibit the densification power law with a space-time graph model;

- The proportion of ride-hailing trips with a shorter distance is smaller than taxis, which indicates that the ride-hailing service tends to serve trips with a longer distance (and potentially a higher revenue).

The main contributions of this paper are summarized as follows:

- This paper presents a comprehensive data-driven analysis and comparison between taxi and ride-hailing services from four aspects, i.e., temporal patterns, spatial patterns, spatio-temporal patterns, and traveling distance patterns;
- It is found that taxi and ride-hailing services have some common travel patterns, but there are also some notable differences;
- The findings indicate that the ride-hailing service partially replaces the role of taxis, but it also exhibits its own characteristics as a new travel option.

The rest of this paper is organized as follows. Section 2 provides a literature review of related work. Section 3 presents the description of the datasets used in this paper. Section 4 delivers our analyses and results. This paper is concluded in Section 5.

2. Related Work

In this section, the related work is reviewed briefly from three aspects, i.e., the study of travel pattern analysis, the influence of transportation network companies and the comparison between multiple transportation modes.

2.1 *Travel Pattern Analysis*

Transportation plays a significant role in everyday life. On weekdays, people travel between homes and workplaces, and on weekends, people travel between homes and other places, e.g., entertainment venues. Even though individuals may have different travel purposes and routes, the travel patterns within a city present some inherent rules that are highly affected by the transportation services available and the city functions of different areas.

There has been a continuous effort to study the travel patterns in urban areas. Before the emergence of widely adopted information and communication technologies, researchers relied on traditional sociological methods, e.g., surveys and questionnaires. With the development of smart devices and the Global Positioning System (GPS), we are now in a big data era, with large volumes of data collected by governments, companies, or crowd-sourced individuals. Other than the GPS dataset we used in this study, other datasets have been used to extract travel patterns before, e.g., massive smart card data, and Place of Interest (POI) data.

Different models and methods are proposed to describe and analyze travel patterns. The authors in [12] use space-time graph modeling of ride request data from Uber and discover that these ride request graphs exhibit a property called the "densification power law". The authors in [32] propose the notion of a shareability network and use it to quantify the benefits of vehicle pooling. The authors in [31] use a statistical-based analysis to understand the mobility pattern of passenger-searching taxis. The authors in [41] use both statistical-based and clustering-based analyses to investigate the passenger travel patterns in massive smart card data. Classification-based analysis is also used to classify taxi drivers into top drivers and ordinary drivers based on the taxi driver's efficiency.

GPS data have been widely used to study travel patterns. In [23], the authors use geographically weighted regression to study the relationship between the urban built environment and online car-hailing travel and find that recreation, entertainment, and residential district points of interest are the most influential factors on night online car-hailing travel. In [34], the authors focus on long-distance taxi rides based on floating car data and compare them with metro usage in Shanghai. Their results identify 12 pick-up and six drop-off hotspots and reveal the phenomenon that passengers on long-distance taxi rides try to avoid rush hours on the street and the inconvenience of interchanges on metro lines. The authors in [3] use taxi GPS data to identify new movement patterns between functional zones by an iteration process. In [9], the authors reveal the relationship between ridership and external factors, e.g., city development and social economy, and compare the travel patterns of taxi trips in Shanghai and New York City using GPS data. The scaling law is proven to exist in the frequency-ranking relationship with the on-demand ride services of Didi in Hangzhou, China.

In addition to the above analytical methods, various visualization technologies have also been developed to help people understand travel patterns. Since visualization technologies are not the focus of this paper, we refer the readers to [2] for an overview of this topic. We also refer the readers to recent surveys [20] about urban human mobility data mining, which introduce many other research topics other than travel pattern analysis.

2.2 *The Influence of Transportation Network Companies*

Smartphone-based taxi applications have changed the taxi market before the appearance of transportation network companies, which connect private vehicles with passengers. These taxi apps solve both credence good and thin market problems but also increase concerns about collusion, monopoly, driver background checks and safety [11]. The drivers can work simultaneously for Uber, Lyft or Didi Chuxing, e.g., double-apping [19] and multi-homing [38], which helps to alleviate the monopoly and thin market problems.

Since the appearance of transportation network companies, such as Uber, Lyft, and Didi Chuxing, which provide Internet-based ride-hailing or ride-sharing services, they have brought a great disruption to the taxi industry and given the passengers more choices of travel. By connecting drivers with their personal vehicles and leveraging surge pricing rules, these companies have become strong competitors for traditional taxis.

There have been some studies about the influence of transportation network companies, e.g., how Uber or Didi is affecting the taxi industry [16]. Uber has been proven to have a higher efficiency than traditional taxis by comparing the capacity utilization rate of UberX drivers with that of traditional taxi drivers in five cities, and it makes it easier to get a ride in rainy hours by increasing supply with surge pricing. Transportation network companies also have a significant effect on traffic conditions, e.g., Uber leads to a significant decrease in traffic congestion and carbon dioxide emissions in the urban areas of the United States.

In [6], the authors use a multinomial linear regression analysis based on a multi-modal, time-series travel dataset and show that ride-hailing companies have a negative and significant effect on taxicab ridership in Las Vegas, Nevada. In [25], the author also observes a significant loss in the taxi industry caused by competition from online ride-hailing services in Shenzhen, China. Similar negative impacts are also observed in Beijing, China [21].

2.3 *The Comparison between Multiple Transport Services*

Some previous studies have attempted to compare the travel patterns between taxi and Internet-based transport services provided by these transportation network companies. In [7], the authors compare taxi and Internet-based ride-sharing services provided by Didi Chuxing in Beijing and reveal many interesting findings, e.g., ride-sharing mainly increases supplies in hot areas and peak hours. In this paper, we compare taxis with Internet-based ride-hailing services, which are different from ride-sharing services. Ride-hailing services, including Uber and Lyft, are found to have shorter wait times and are more reliable. In [29], the authors compare the Green cabs and Uber from April-September

2014 in New York (excluding Manhattan and two airports) and find that Uber is growing faster than Green cabs at the studied time period.

In [13], the authors compare Uber, Lyft, and taxis in San Francisco and New York City and find that transportation infrastructure and socioeconomic features have substantial effects on ride service market features, e.g., supply, demand, price, and wait time.

3. Data Description

Traffic big data are the basis for conducting data-driven analysis in this study, and two separate datasets are involved. In this section, a brief introduction for the datasets is given, and the data pre-processing steps are described.

3.1 Taxi Trip Data

In this paper, we use a taxi GPS dataset collected in Chengdu, China, in August 2014. Specifically, we use the GPS data of 13 weekdays and 5 weekends, collected from August 3 to August 23. The data from August 7, August 13 and August 17 are missing or incomplete and thus omitted from this study. Since we are dealing with the average cases, these missing dates would not affect our findings in the following sections. The spatial range is bounded to the central part of Chengdu city (longitude: 103.9 - 104.21, latitude: 30.53 - 30.8), and the temporal range is bounded from 6:00 to 24:00. We use the same spatial and temporal ranges for the ride-hailing service and extract the data within these ranges [40].

The taxi GPS trajectory is sampled at a rate of approximately 2 samples per minute. Each sample contains the trajectory identity, anonymous taxi identity, latitude, longitude, operation status (occupied or vacant), and timestamp. To further analyze the travel patterns, we extract the trip information from the GPS data. A taxi trip is defined as the tuple (pickup longitude & latitude, dropoff longitude & latitude, begin time, end time, trip distance, anonymous driver identity). We follow the steps in [16] to extract the trip information and calculate the trip distance by the length of the GPS trajectory. A trip's duration is defined as the interval between the beginning time and end time, and in this study, we filter out trips with a duration less than 1 minute, as they are usually caused by an inaccurate operation status. In summary, we extract 8,711,301 trips from the taxi GPS dataset, generated by 14,502 taxi drivers.

3.2 Ride-hailing Trip Data

In this paper, we use a ride-hailing trip dataset provided by Didi Chuxing [†]. The trip dataset covers the order and GPS trajectory information of Didi's ride-hailing service in Chengdu from November 1 to November 30, 2016, which contains 22 weekdays and 8 weekends. We filter out the records that are not fully within our spatial and temporal ranges and the orders with a duration less than 1 minute.

Each order record contains the order ID, begin time, end time, pickup longitude & latitude, and dropoff longitude & latitude. Each trace record contains an anonymous driver identity, order ID, timestamp, longitude, and latitude. The trace record is sampled at a rate of approximately 20 samples per minute. Similarly, we combine the order record and the corresponding trace record into a trip tuple, in which the trip distance is calculated by the length of the GPS trace. We also find that the dataset is imperfect, as some order records have no corresponding trace information, in which cases we would calculate the trip distance as the Manhattan distance between the pickup location and the dropoff lo-

[†] https://gaia.didichuxing.com

cation. In summary, we extract 5,920,265 trips from the ride-hailing trip dataset, served by 1,163,328 Didi drivers. The number of drivers may be inaccurate because we do not know if the anonymous driver identity is a one-to-one mapping to a Didi driver.

4. Analysis

In this section, we compare the travel patterns between taxi and ride-hailing services from four aspects, i.e., temporal patterns, spatial patterns, spatio-temporal patterns, and traveling distance patterns.

4.1 *Temporal Patterns*

The temporal patterns describe the relationship of trip numbers and different time periods. The trip numbers could be different from day to day or from hour to hour in a day. The temporal patterns would also be notably different on weekdays and weekends, as the trip demand for commutes would be very different.

We show the trip numbers within each day used in this study in Figure 1. From Figure 1(a), there are no significant patterns of taxi trip numbers. From Figure 1(b), the total Didi trip numbers on Friday and Saturday are notably larger than those on the remaining days. This phenomenon could be explained by Didi's promotions, which incentivize more users to use Didi's service on Friday and Saturday. This result demonstrates the flexible supply and demand on a ride-hailing platform.

We show the start time distribution of trips served by taxi and ride-hailing in Figure 2. The distribution presents the average trip proportion that starts in each one-hour time slot. As shown in Figure 2, the distributions of taxi and ride-hailing services exhibit several differences during three different time periods:

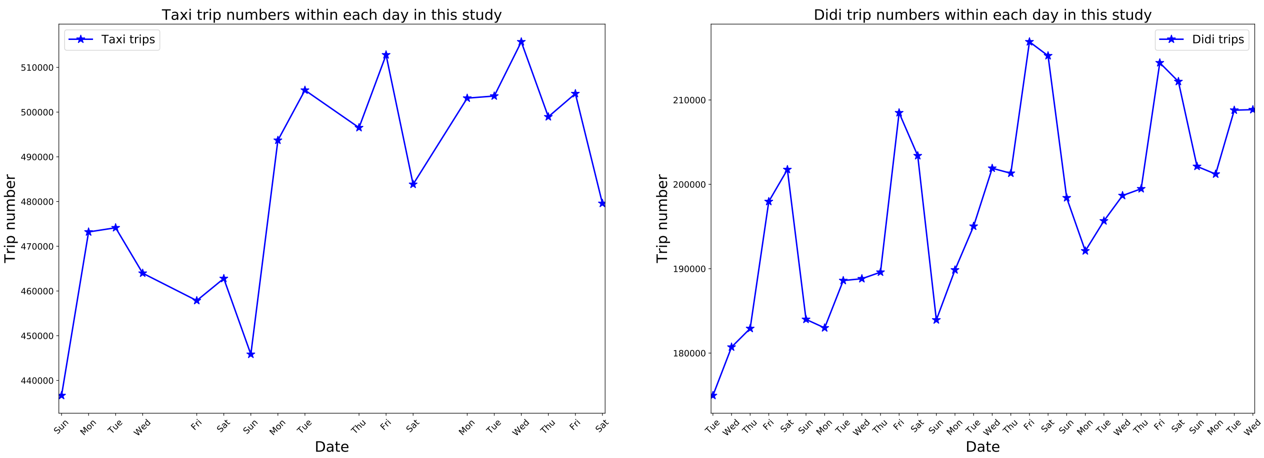
- Morning rush hours: For Didi trips, there is a noticeable increase in trip proportion during morning rush hours (from 7 am to 10 am). This result indicates that the ride-hailing service has become a significant part of daily commuting. The increase in taxi trips during this time period is not so obvious.

- Evening rush hours: For taxi trips, the proportion drops significantly on weekdays during evening rush hours (from 5 pm to 8 pm). This result is reasonable, considering the steady supply of taxis and traffic congestion during rush hours, both of which limit the trip numbers served by taxis. In contrast, the Didi trips are less affected in this time period.

- Late in the evening: The proportion of Didi trips drops continually late in the evening (after 8 pm). As some Didi drivers are those with private cars, they may not want to work late at night. In contrast, the taxi trips remain stable before 11 pm, which reflects the fact that before the emergence of ride-hailing services, taxis played a significant role in traveling late at night, especially when other public transportation methods, e.g., buses and subways, stopped service at night.

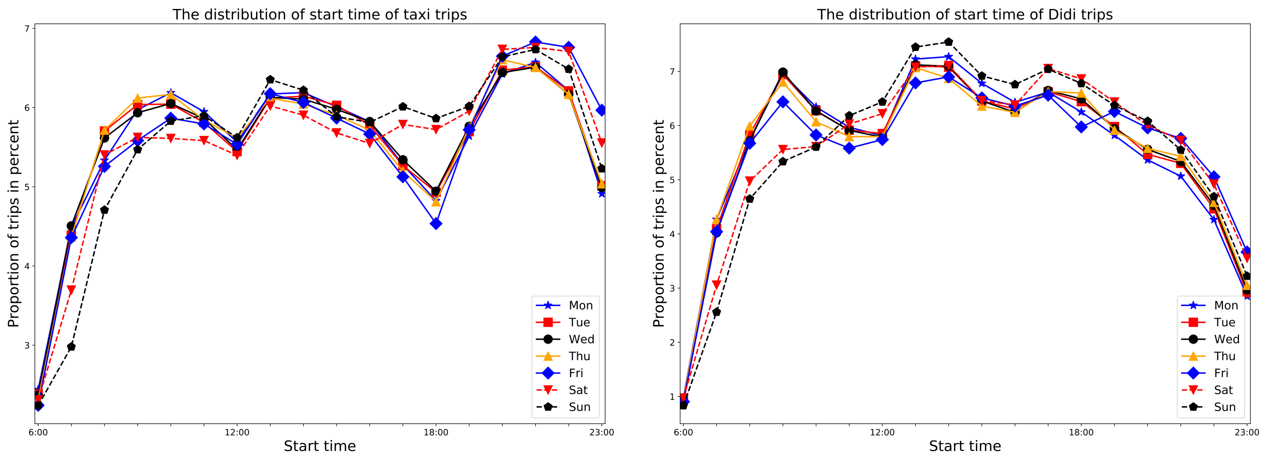
4.2 *Spatial Patterns*

The spatial patterns describe the relationship of trip numbers and different locations. For visualization, we draw the pickup and dropoff heatmaps of trips served by taxi and ride-hailing in Figure 3. To make a fair comparison, we show the relative distribution in Figure 3, instead of the absolute pickup and dropoff numbers.



(a) (b)

Figure 1: Trip numbers. (a) Taxi; (b) Ride-hailing.



(a) (b)

Figure 2: Trip start time distributions. (a) Taxi; (b) Ride-hailing.

From Figure 3, we can tell that for taxi or ride-hailing services, the pickup and dropoff hot areas basically overlap. However, the patterns between taxis and ride-hailing services are different. Taxi trips are distributed more sparsely, while ride-hailing trips are more concentrated in the central part. The observation indicates that taxis have wider coverage in the urban transportation system and are essential for serving less visited areas. However, the ride-hailing service is more profit-driven and tend to stay in central areas. Furthermore, the two services share some hot places, and the emergence of ride-hailing services would help to alleviate the stress of travel demand in these hot places.

4.3 Spatio-temporal Patterns

Temporal or spatial patterns alone may not provide further information about travel patterns, and we leverage more sophisticated approaches to analyze spatio-temporal patterns, namely, basis collective pattern analysis and space-time graph analysis.

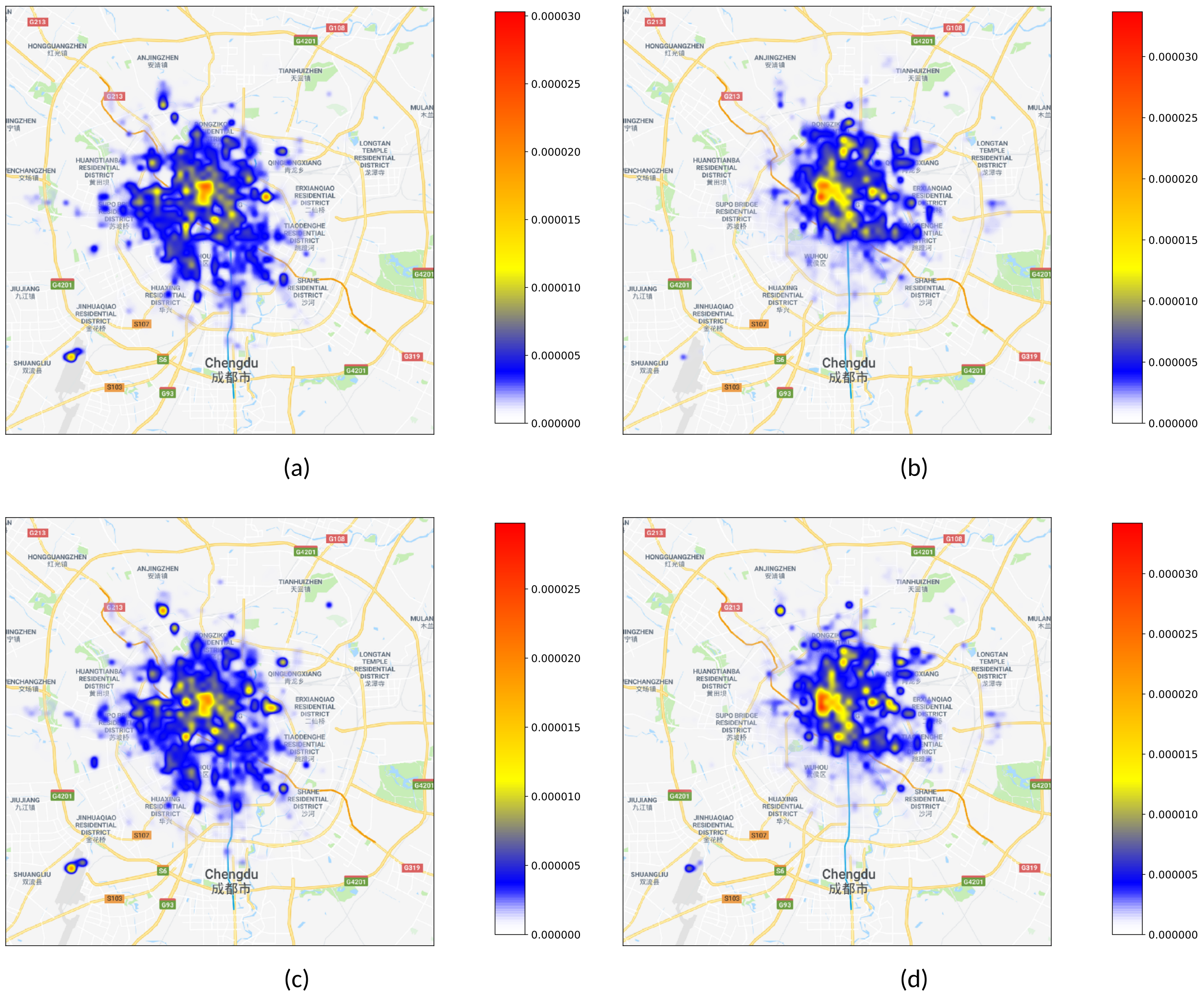


Figure 3: The pickup and dropoff heatmaps. (a) The pickup heatmap of taxi trips; (b) The pickup heatmap of ride-hailing trips; (c) The dropoff heatmap of taxi trips; (d) The dropoff heatmap of taxi trips.

4.3.1 *Basis Collective Pattern Analysis*

A previous study found that the macro traffic pattern can be described by some linear combinations of basis collective patterns [27]. In this part, we use a similar approach to [7] to extract the basis collective patterns for taxi and ride-hailing, separately.

The Chengdu city is divided into $I \times J = 150 \times 150$ grids, and each grid has a size of 200×200 square meters. We also divide the time range into $T = 18$ time slots, and each time slot has a duration of one hour. We use a three-dimensional matrix $M \in \mathcal{R}_+^{I \times J \times T}$ to represent the macro traffic pattern, where each element $M_{i,j,t}$ represents the trip numbers that have a pickup location within grid (i, j) and a start time within time slot t.

Then our objective is to decompose the matrix M into two low rank nonnegative factors $S \in \mathcal{R}_+^{I \times J \times \lambda}$ and $B \in \mathcal{R}_+^{\lambda \times T}$, where λ is the number of basis collective patterns. We use the nonnegative matrix factorization (NMF) method [5, 8] to decompose the matrix M, as this method has been proven to be the most effective in [7], and then normalize the row vectors of B for the relative traffic volume in each pattern.

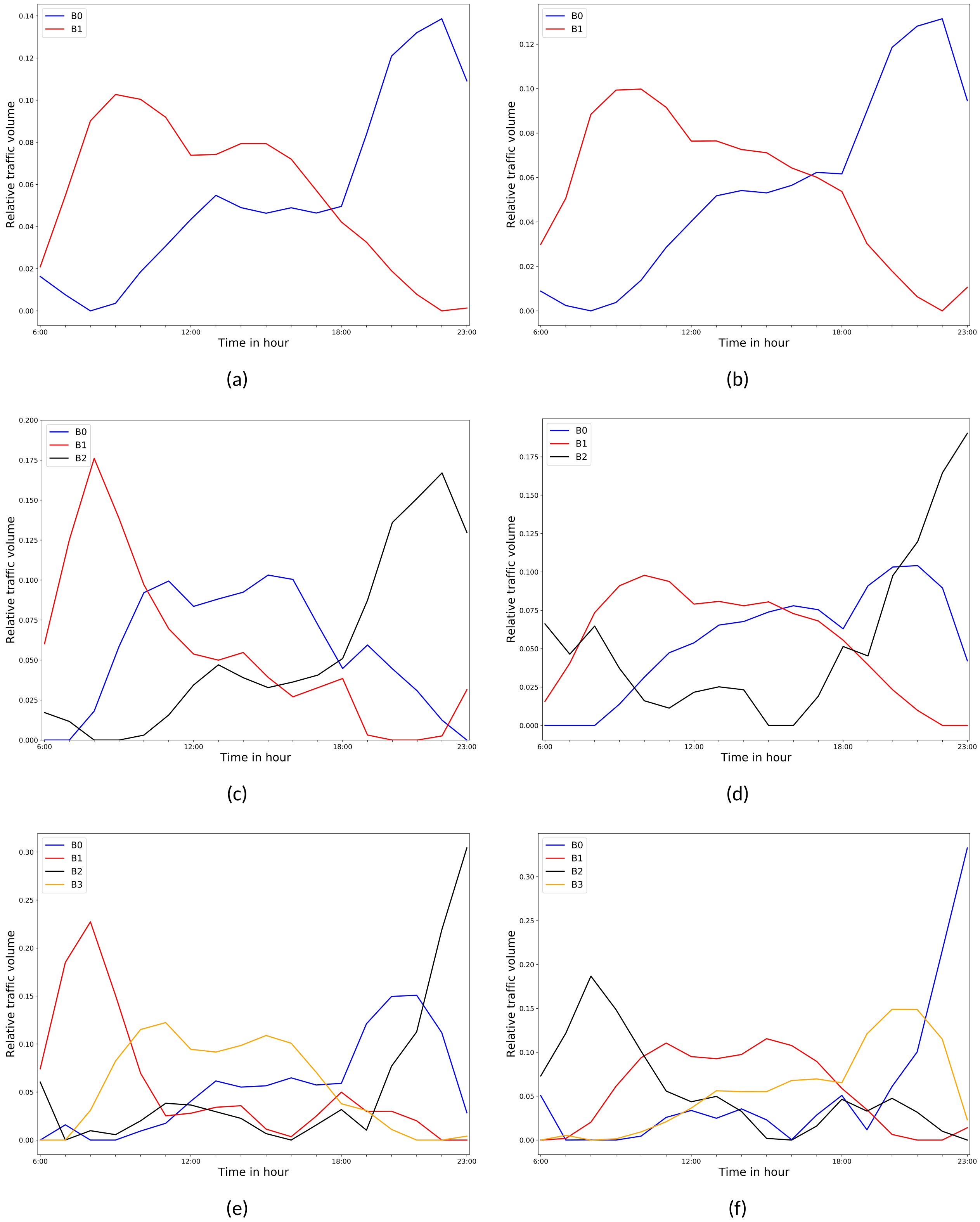


Figure 4: The basis collective patterns of taxi trips. (a) Weekdays, $\lambda = 2$; (b) Weekends, $\lambda = 2$; (c) Weekdays, $\lambda = 3$; (d) Weekends, $\lambda = 3$; (e) Weekdays, $\lambda = 4$; (f) weekends, $\lambda = 4$.

We show the factorization results of trips served by taxis through the NMF method in Figure 4, and the factorization results of trips served by ride-hailing through the NMF method in Figure 5, when

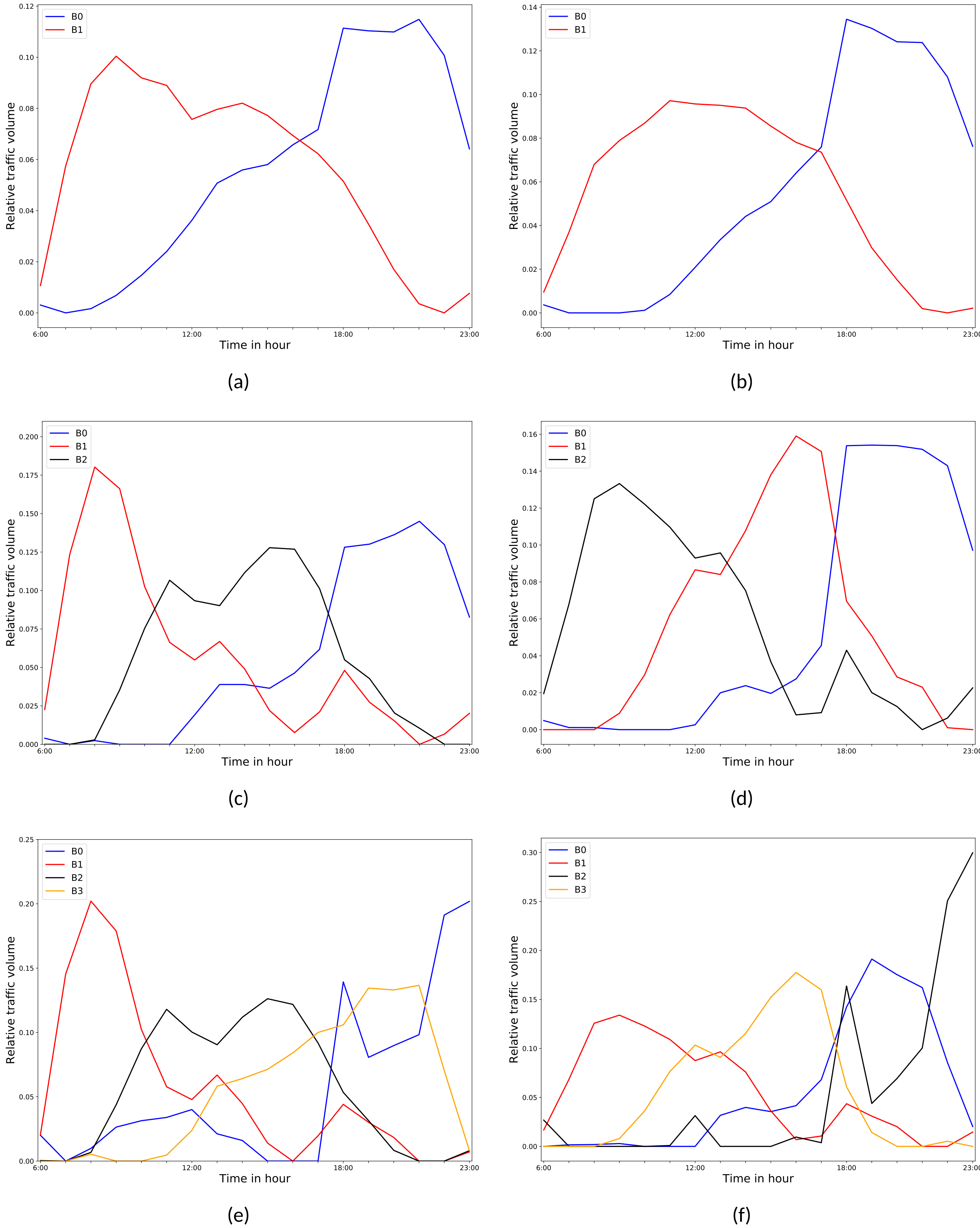


Figure 5: The basis collective patterns of ride-hailing trips. (a) Weekdays, $\lambda = 2$; (b) Weekends, $\lambda = 2$; (c) Weekdays, $\lambda = 3$; (d) Weekends, $\lambda = 3$; (e) Weekdays, $\lambda = 4$; (f) weekends, $\lambda = 4$.

choosing the value of λ from 2 to 4.

We have several observations from Figure 4 and Figure 5 as follows:

- The result is more reasonable and stable, when choosing $\lambda = 3$, for both the trips served by taxi and ride-hailing. The three main collective patterns on weekdays can be explained as follows: commuting between home and workplace, commuting between workplace and home, and trips traveling between other places.
- The comparison between weekdays and weekends shows that the first pattern of commuting between home and workplace is not noticeable on weekends. The other patterns during weekends may have different explanations, such as trips traveling between homes and entertainment venues in the evening.
- The comparison between taxi and ride-hailing exhibits similar results, which indicates that ride-hailing partially performs the same tasks as taxi, especially during the morning rush hours.

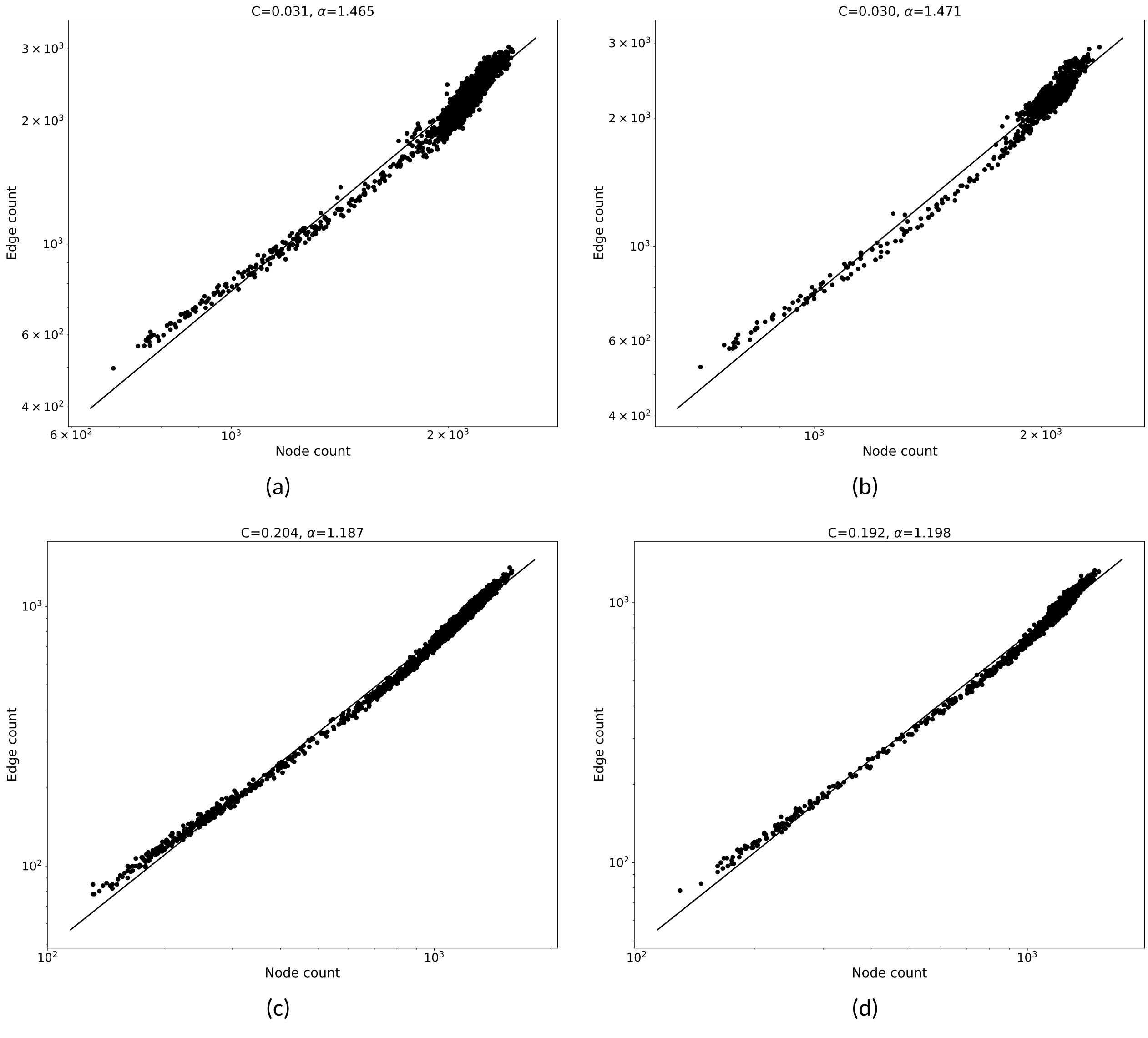


Figure 6: The densification power law. (a) Taxi trips on weekdays; (b) Taxi trips on weekends; (c) Ride-hailing trips on weekdays; (d) Ride-hailing trips on weekends.

4.3.2 *Space-time Graph Analysis*

Another approach for analyzing spatio-temporal patterns is based on a model named the space-time graph [12]. We demonstrate that both trips served by taxi and ride-hailing exhibit the well-known property called the "densification power law", which has been found in other research about human behaviors, such as social network graphs and publication citation graphs.

To build the space-time graph, we also divide Chengdu city into $I \times J = 150 \times 150$ grids. However, in this part, we divide the time range into $T = 18 \times 12$ time slots, and each time slot has a duration of 5 minutes. For time slot t, we build a space-time graph $G(t)$. For a trip that starts within time slot t, we map the pickup and dropoff locations into grid ids. Then, we add the pickup grid ID and dropoff grid ID to the space-time graph's node set $N(t)$, and the edge between these two grids to the space-time graph's edge set $E(t)$. Note that for two trips that start from the same grid and in the same time slot, the common pickup grid would be counted as one node in the space-time graph.

We find that both the trips served by taxi and ride-hailing exhibit the densification power law (DPL), which means that the number of edges grows as a power of the number of nodes, as shown as follows:

$$e(t) \propto n(t)^{\alpha} \text{ or } e(t) = Cn(t)^{\alpha} \tag{1}$$

where $e(t) = |E(t)|$ denotes the number of edges in $G(t)$, $n(t) = |N(t)|$ denotes the number of nodes in $G(t)$, and C and α are constants. We can estimate the specific values of C and α by a simple ordinary least squares model.

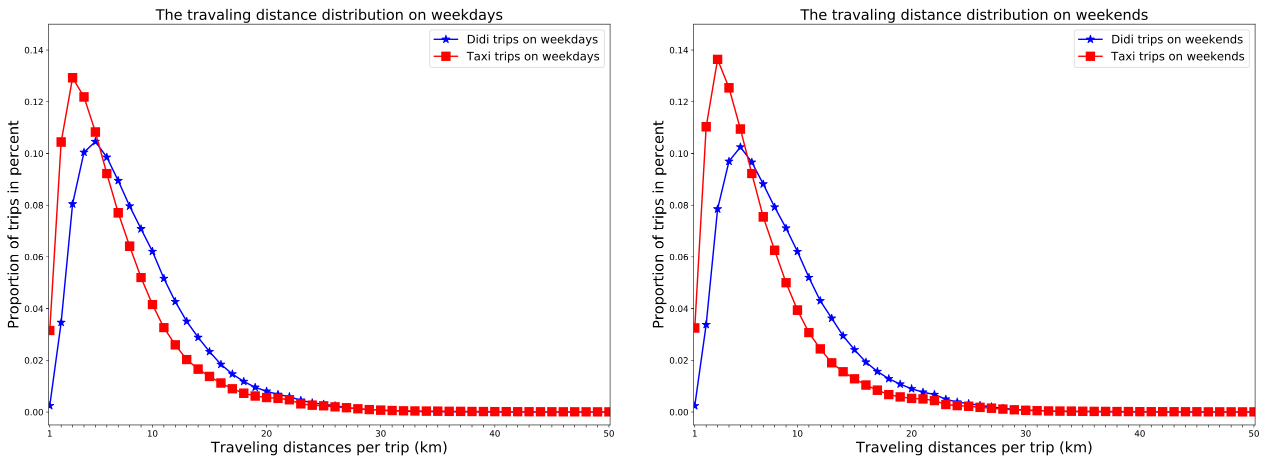
We show the densification power law figures in Figure 6. As shown in Figure 6, both the trips served by taxi and ride-hailing exhibit the densification power law, and the patterns are similar between weekdays and weekends. However, the constants C and α are different between taxi and ride-hailing. A higher C of taxis indicates that the number of edges grows faster with the number of nodes. This result shows that the space-time graph of taxis has a denser structure, which indicates that given a specific set of nodes as pickups and dropoffs, taxi drivers would serve more diversified trips than ride-hailing services, which show fewer combinations of pickups and dropoffs.

4.4 *Traveling Distance Patterns*

Finally, we show the traveling distance distribution of trips served by taxi and ride-hailing in Figure 7. As we can tell from Figure 7, the proportion of trips with shorter distances (less than 5 kilometers) served by ride-hailing is smaller than that of taxis. The possible reason for this phenomenon could be that the ride-hailing service charges less per kilometer traveled to compete with taxis, and passengers tend to use ride-hailing for long-distance trips to save money.

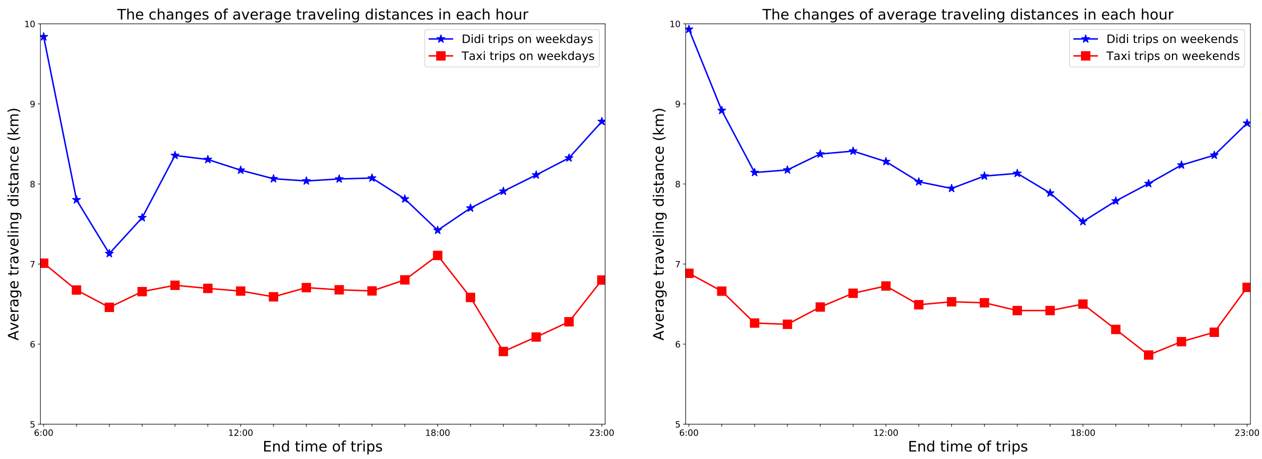
We also show the changes in average traveling distances with different trip end times during one whole day on weekdays and weekends in Figure 8. As shown in Figure 8, the trips served by ride-hailing have larger average traveling distances than taxis during the whole day. The dynamic change patterns are different between taxi and ride-hailing, which is different from the result in [7] that the patterns are similar between taxi and ride sharing. The average traveling distances are only slightly longer on weekends than weekdays.

The study reveals both similarities and differences between taxis and ride-hailing services in urban transportation. Both modes share fundamental travel patterns, such as adhering to the densification power law in spatial and temporal dynamics, and they often operate within overlapping pickup and dropoff hotspots. However, key differences emerge in service characteristics and usage trends. Ride-hailing services exhibit a preference for longer trips and higher concentrations in central urban areas, driven by their profit-oriented model, whereas taxis provide broader spatial coverage, serving less visited regions. Temporally, ride-hailing sees a marked increase in demand during morning commutes



(a) (b)

Figure 7: The traveling distance distribution. (a) Weekdays; (b) Weekends.



(a) (b)

Figure 8: The average traveling distance with different trip end times. (a) Weekdays; (b) Weekends.

and peaks during weekends, bolstered by promotions, while taxis maintain steadier patterns. These distinctions highlight the complementary roles of both services, with taxis catering to dispersed urban areas and ride-hailing excelling in meeting targeted high-demand needs.

An unexpected finding from the study is the distinct temporal trend in which ride-hailing services show a significant surge during morning rush hours and weekends, a pattern less pronounced for taxis. This could be attributed to the flexibility and dynamic pricing mechanisms of ride-hailing platforms, which effectively respond to fluctuating demand and incentivize drivers during peak periods. Another surprising trend is the higher average trip distance for ride-hailing services, possibly due to competitive pricing structures that make long-distance trips more economical compared to taxis. Practically, these trends suggest that ride-hailing platforms are not just substituting taxis but are also carving out distinct operational niches, particularly for commuter and leisure travel. These insights could guide urban planners and policymakers to leverage ride-hailing for high-demand areas while deploying taxis to maintain accessibility in underserved regions, creating a balanced and efficient urban transportation network.

The findings from this study can be applied to several real-world scenarios to enhance urban transportation systems. For instance, the observed differences in spatial and temporal patterns between taxis and ride-hailing services can guide the allocation of resources to reduce congestion in high-demand urban hotspots during peak hours. Policymakers could use these insights to incentivize ride-hailing services to operate in underserved areas, ensuring broader coverage and reducing disparities in accessibility. Additionally, the data-driven patterns of increased ride-hailing demand during events or holidays could help optimize service deployment through surge pricing and dynamic routing algorithms, ensuring sufficient availability and shorter wait times. For major public events, combining these findings with predictive models could enable preemptive adjustments to traffic management plans, such as setting up temporary pickup/dropoff zones or rerouting traffic to alleviate bottlenecks. By integrating these strategies, cities can improve transportation efficiency, reduce environmental impact, and enhance commuter satisfaction.

While big data and machine learning provide valuable tools for intelligent transportation, they come with notable limitations. One significant challenge lies in the scope and quality of datasets, which may be incomplete, biased, or not representative of the diversity of traffic scenarios across regions. For instance, data from densely populated areas may dominate, sidelining rural transportation insights. Furthermore, the inherent assumptions in modeling, such as fixed demand patterns or idealized system dynamics, may fail to account for real-world irregularities like abrupt weather changes or unpredictable human behavior. Scalability issues also arise when adapting solutions developed for one locale to another with differing urban layouts or infrastructure. Lastly, ethical concerns regarding data privacy and the potential for algorithmic bias can hinder the equitable deployment of these technologies, as noted in various transportation studies, such as in Chengdu's ride-hailing versus taxi services, which demonstrate service concentration in central urban areas while neglecting broader coverage.

5. Conclusion

In this paper, a data-driven analysis of two transport services, namely, taxi and ride-hailing, is conducted with traffic big data collected in Chengdu, China. The findings from four different aspects reveal both the similarities and differences between the two considered travel choices in urban space. The similar patterns support the viewpoint that the ride-hailing service partially replaces the functionality of taxis. The differences, on the other hand, open the possibility of combining these two services to complement each other.

Several promising directions are highlighted for future research to enhance the understanding of transportation dynamics. One key suggestion is to extend the analysis to other cities with diverse urban layouts and socio-economic conditions, allowing for comparisons and generalizations across varying contexts [33]. Incorporating additional data sources, such as socio-economic indicators, land use information, and real-time weather conditions, could provide a more comprehensive understanding of the factors influencing travel patterns. Furthermore, exploring individual travel behaviors through microscopic analyses, such as studying specific driver or passenger profiles, could reveal nuanced insights into decision-making processes [36]. Integrating clustering and classification-based methods to identify latent patterns in travel data, as well as combining these findings with POI data, would help link travel behaviors to specific urban functions, such as residential areas, workplaces, or recreational zones. These approaches could provide policymakers with actionable insights for optimizing transportation services and urban planning.

Acknowledgement

This research was not funded by any grant.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Alalawi, K., Athauda, R., & Chiong, R. (2023). Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review. *Engineering Reports*, *5*(12), e12699. https://doi.org/10.1002/eng2.12699

[2] Andrienko, N., & Andrienko, G. (2013). Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, *12*(1), 3–24. https://doi.org/10.1177/1473871612457601

[3] Chen, Z., Gong, X., & Xie, Z. (2017). An analysis of movement patterns between zones using taxi gps data. *Transactions in GIS*, *21*(6), 1341–1363. https://doi.org/10.1111/tgis.12281

[4] Chengdu Bureau of Statistics. (2015). The statistical yearbook of chengdu, 2015 [online] [Accessed on: Nov. 10, 2024].

[5] Cichocki, A., & Phan, A. H. (2008). Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, *92*(3), 708–721. https://doi.org/10.1587/transfun.E92.A.708

[6] Contreras, S. D., & Paz, A. (2018). The effects of ride-hailing companies on the taxicab industry in las vegas, nevada. *Transportation Research Part A: Policy and Practice*, *115*, 63–70. https://doi.org/10.1016/j.tra.2017.11.008

[7] Dong, Y., Wang, S., Li, L., & Zhang, Z. (2018). An empirical study on travel patterns of internet based ride-sharing. *Transportation Research Part C: Emerging Technologies*, *86*(Supplement C), 1–22. https://doi.org/10.1016/j.trc.2017.10.022

[8] Févotte, C., & Idier, J. (2011). Algorithms for nonnegative matrix factorization with the β-divergence. *Neural Computation*, *23*(9), 2421–2456. https://doi.org/10.1162/NECO_a_00168

[9] Ge, W., Shao, D., Xue, M., Zhu, H., Zhao, X., & Cheng, J. (2018). Urban taxi ridership analysis in the emerging metropolis: Case study in shanghai. *Case Studies on Transport Policy*. https://doi.org/10.1016/j.trpro.2017.05.368

[10] Haghshenas, S. S., Astarita, V., Guido, G., Seraji, M. H. M., Gonzalez, P. A. A., Haghdadi, A., & Haghshenas, S. S. (2024). Assessment of machine learning techniques and traffic flow: A qualitative and quantitative analysis. *Journal of Computational and Cognitive Engineering*, *3*(2), 119–129. https://doi.org/10.47852/bonviewJCCE32021062

[11] Harding, S., Kandlikar, M., & Gulati, S. (2016). Taxi apps, regulation, and the market for taxi journeys. *Transportation Research Part A: Policy and Practice*, *88*, 15–25. https://doi.org/10.1016/j.tra.2016.03.009

[12] Jauhri, A., Foo, B., Berclaz, J., Hu, C. C., Grzeszczuk, R., Parameswaran, V., & Shen, J. P. (2017). Space-time graph modeling of ride requests based on real-world data. *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*.

[13] Jiang, S., Chen, L., Mislove, A., & Wilson, C. (2018). On ridesharing competition and accessibility: Evidence from uber, lyft, and taxi. *Proceedings of the 2018 World Wide Web Conference*, 863–872. https://doi.org/10.1145/3178876.3186134

[14] Jiang, W. (2022). Taxibj21: An open crowd flow dataset based on beijing taxi gps trajectories. *Internet Technology Letters*, *5*(2), e297. https://doi.org/10.1002/itl2.297

[15] Jiang, W., Dominguez, C. R., Zhang, P., Shen, M., & Zhang, L. (2018). Large-scale nationwide ridesharing system: A case study of chunyun. *International Journal of Transportation Science and Technology*, *7*(1), 45–59. https://doi.org/10.1016/j.ijtst.2017.10.002

[16] Jiang, W., Lian, J., Shen, M., & Zhang, L. (2017). A multi-period analysis of taxi drivers' behaviors based on GPS trajectories. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC) (ITSC2017)*. https://doi.org/10.1109/ITSC.2017.8317622
[17] Jiang, W., & Luo, J. (2022a). Big data for traffic estimation and prediction: A survey of data and tools. *Applied System Innovation*, *5*(1), 23. https://doi.org/10.3390/asi5010023
[18] Jiang, W., & Luo, J. (2022b). Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, 117921. https://doi.org/10.1016/j.eswa.2022.117921
[19] Jiang, W., & Zhang, L. (2018a). Evaluating the effects of double-apping on the smartphone-based e-hailing service: A simulation-based study. *IEEE Access*, *6*, 6654–6667. https://doi.org/10.1109/ACCESS.2018.2797207
[20] Jiang, W., & Zhang, L. (2018b). Geospatial data to images: A deep-learning framework for traffic forecasting. *Tsinghua Science and Technology*, *24*(1), 52–64. https://doi.org/10.26599/TST.2018.9010033
[21] Jiang, W., & Zhang, L. (2018c). The impact of the transportation network companies on the taxi industry: Evidence from beijing's gps taxi trajectory data. *Ieee Access*, *6*, 12438–12450. https://doi.org/10.1109/ACCESS.2018.2810140
[22] Li, Q., Huang, Z., Jiang, W., Tang, Z., & Song, M. (2024). Quantum algorithms using infeasible solution constraints for collision-avoidance route planning. *IEEE Transactions on Consumer Electronics*. https://doi.org/10.1109/TCE.2024.3476156
[23] Li, T., Jing, P., Li, L., Sun, D., & Yan, W. (2019). Revealing the varying impact of urban built environment on online car-hailing travel in spatio-temporal dimension: An exploratory analysis in chengdu, china. *Sustainability*, *11*(5), 1336. https://doi.org/10.3390/su11051336
[24] Lu, Y., Wang, W., Bai, R., Zhou, S., Garg, L., Bashir, A. K., Jiang, W., & Hu, X. (2025). Hyper-relational interaction modeling in multi-modal trajectory prediction for intelligent connected vehicles in smart cites. *Information Fusion*, 102682. https://doi.org/10.1016/j.inffus.2024.102682
[25] Nie, Y. M. (2017). How can the taxi industry survive the tide of ridesourcing? evidence from shenzhen, china. *Transportation Research Part C: Emerging Technologies*, *79*, 242–256. https://doi.org/10.1016/j.trc.2017.03.017
[26] Parthasarathy, S., Panigrahi, P. K., & Subramanian, G. H. (2024). A framework for managing ethics in data science projects. *Engineering Reports*, *6*(3), e12722. https://doi.org/10.1002/eng2.12722
[27] Peng, C., Jin, X., Wong, K.-C., Shi, M., & Liò, P. (2012). Collective human mobility pattern from taxi trips in urban area. *PLOS ONE*, *7*(4), 1–8. https://doi.org/10.1371/journal.pone.0034487
[28] Peng, D., Huang, M., & Xing, Z. (2023). Taxi origin and destination demand prediction based on deep learning: A review. *Digital Transportation and Safety*, *2*(3), 176–189. https://doi.org/10.48130/DTS-2023-0014
[29] Poulsen, L. K., Dekkers, D., Wagenaar, N., Snijders, W., Lewinsky, B., Mukkamala, R. R., & Vatrapu, R. (2016). Green cabs vs. uber in new york city. *2016 IEEE International Congress on Big Data (BigData Congress)*, 222–229. https://doi.org/10.1109/BigDataCongress.2016.35
[30] Qi, Q., Cheng, R., & Ge, H. (2023). Short-term inbound rail transit passenger flow prediction based on bilstm model and influence factor analysis. *Digital Transportation and Safety*, *2*(1), 12–22. https://doi.org/10.48130/DTS-2023-0002
[31] Rao, F., Zhang, X., Dong, Y., Chen, C., & Dong, H. (2014). Understanding the mobility pattern of passenger-searching taxis. *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 290–295. https://doi.org/10.1109/ITSC.2014.6957706
[32] Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S. H., & Ratti, C. (2013). Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy*

of Sciences of the United States of America, *111*(37), 13290–4. https://doi.org/10.1073/pnas.1403657111
[33] Wang, L., Wang, S., Yuan, Z., & Peng, L. (2021). Analyzing potential tourist behavior using pca and modified affinity propagation clustering based on baidu index: Taking beijing city as an example. *Data Science and Management*, *2*, 12–19. https://doi.org/10.1016/j.dsm.2021.05.001
[34] Wu, H., Fan, H., & Wu, S. (2017). Exploring spatiotemporal patterns of long-distance taxi rides in shanghai. *ISPRS International Journal of Geo-Information*, *6*(11), 339. https://doi.org/10.3390/ijgi6110339
[35] Xu, Z., Tang, N., Xu, C., & Cheng, X. (2021). Data science: Connotation, methods, technologies, and development. *Data Science and Management*, *1*(1), 32–37. https://doi.org/10.1016/j.dsm.2021.02.002
[36] Yang, B., Wang, X., Xing, Y., Cheng, C., Jiang, W., & Feng, Q. (2024). Modality fusion vision transformer for hyperspectral and lidar data collaborative classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. https://doi.org/10.1109/JSTARS.2024.3415729
[37] Yang, B., Zhou, J., Zhang, S., Xing, Y., Jiang, W., & Xu, L. (2024). Lightweight knowledge distillation and feature compression model for user click-through rates prediction in edge computing scenarios. *IEEE Internet of Things Journal*. https://doi.org/10.1109/JIOT.2024.3446640
[38] Yu, J., Mo, D., Xie, N., Hu, S., & Chen, X. M. (2021). Exploring multi-homing behavior of ride-sourcing drivers via real-world multiple platforms data. *Transportation Research Part F: Traffic Psychology and Behaviour*, *80*, 61–78. https://doi.org/10.1016/j.trf.2021.03.017
[39] Yuan, Z., & Jia, G. (2021). Profiling the digital divide of the elderly based on internet big data: Evidence from china. *Data Science and Management*, *3*, 33–43. https://doi.org/10.1016/j.dsm.2021.10.001
[40] Zhang, Y., Xu, S., Zhang, L., Jiang, W., Alam, S., & Xue, D. (2024). Short-term multi-step-ahead sector-based traffic flow prediction based on the attention-enhanced graph convolutional lstm network (agc-lstm). *Neural Computing and Applications*, 1–20. https://doi.org/10.1007/s00521-024-09827-3
[41] Zhao, J., Qu, Q., Zhang, F., Xu, C., & Liu, S. (2017). Spatio-temporal analysis of passenger travel patterns in massive smart card data. *IEEE Transactions on Intelligent Transportation Systems*, *18*(11), 3135–3146. https://doi.org/10.1109/TITS.2017.2679179