Physica A 515 (2019) 785–797

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Understanding bike sharing travel patterns: An analysis of trip data from eight cities

Zhaoyu Kou [a], Hua Cai [a,b,*]

[a] School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, United States
[b] Environmental and Ecological Engineering, Purdue University, West Lafayette, IN 47907, United States

HIGHLIGHTS

- The travel patterns of bike sharing users are influenced by the station network.
- Bike sharing trip displacement follows different distribution in large and small systems.
- Long-distance trip displacement distribution displays a power law decay.
- Trip distance and during.

ARTICLE INFO

Article history:
Received 20 February 2018
Received in revised form 19 May 2018
Available online 9 October 2018

Keywords:
Bike sharing
Trip distance distribution
Trip duration distribution
Human mobility
Travel patterns

ABSTRACT

As a new mobility option, bike sharing is gaining popularity around the world. Understanding the travel patterns of bike sharing trips can provide fundamental basis for researchers to model the use of bike sharing and the associated multi-modal transportation systems, inform bike sharing system design and operation, and guide policy decisions for sustainable transportation development. Using bike sharing trip data from eight cities in the United States, we analyzed the distributions of trip distance and trip duration for bike sharing trips for commuting and touristic purposes. Our results show that both the trip distance and duration follows a lognormal distribution in larger bike sharing systems (e.g., in Boston, Washington DC, Chicago, and New York), while the distribution for smaller systems varies among Weibull, gamma, and lognormal because the systems' geographical boundary restricts the movement of users. Our analysis of the long trips also show that the trip distance and duration also displays a power law decay in the larger systems.

1. Introduction

It is of fundamental importance to study human travel patterns. The knowledge of the dynamics and statistical properties of human mobility can be applied to many fields, such as traffic forecasting [1,2], urban planning [3], epidemics [4–7], genetics [8], and mobile network designing [9]. For instance, understanding the statistical patterns of human movement can enable better simulation and design of the mobile networks of cellphones [10], because wireless devices rely on the location of nearby nodes (e.g., base stations) to connect with the wider telephone networks. In recent years, researchers are able to better analyze human mobility patterns due to the ubiquitous availability of location tracking technologies such as Global Positioning System (GPS) and mobile devices (e.g., smartphones).

* Correspondence to: 315 N. Grant Street, West Lafayette, IN 47907, United States.
E-mail address: huacai@purdue.edu (H. Cai).

https://doi.org/10.1016/j.physa.2018.09.123
0378-4371/

Previous studies of human mobility patterns have focused on the distributions of trip displacement (i.e., how far do people travel in one trip) and trip duration (i.e., how long does one trip take). Brockmann et al. [11] analyzed the circulation of bank notes and found that people's travel displacement follows a power law distribution and exhibits Lévy flight (random walk) behaviors. This Lévy flight pattern has been previously observed in animal movements [12–14] and further been confirmed for human mobility by Rhee et al. [10] using GPS traces of 44 volunteers in different outdoor settings. Jiang et al. [1] examined the trajectories of taxi trips to study the mechanisms governing the Lévy flight behavior of human mobility and also compared the fitted results of power law distribution with those of lognormal distribution. They concluded that the Lévy flight behavior is driven by the underlying street network and the human mobility pattern can be reproduced by simulated random walks. Based on cell phone traces, González et al. [15] found that, although the distribution of displacement for human motion follows a truncated power-law, Lévy flight is not sufficient to explain the spatial and temporal regularities shown after the correction of the inherent anisotropy of each individual's trajectory and the different scale of travel distance. Using the same data set, Song et al. [16] further proposed a mechanism, a combination of exploration and preferential return, to explain the human mobility scaling properties and regularities.

Additionally, researchers reported that trips taken using different transportation modes may have different mobility patterns. Yan et al. [17] pointed out that the displacements by a single transportation mode exhibit an exponential distribution, instead of the power law distribution previously reported, due to the influence of the traveling cost. Kolb and Helbing [18] showed that the scaling laws in human travel behavior is related to the different energy consumption required for physical activities when traveling in different transportation modes (walking, cycling, car, bus, and train). Higher energy consumption rate from physical activities will reduce the average travel times. Liang et al. [19] analyzed the taxi trips in Beijing and found that both the displacement and trip duration of taxi trips can be better fitted by an exponential distribution instead of a power-law distribution. However, in the analysis of taxi GPS traces in Lisbon, Portugal, Veloso et al. [20] found that trip distance can be represented with a gamma distribution. While these studies focus on the mobility of the taxi passengers (considering only trips when the taxis are occupied by passengers), Cai et al. [21] further pointed out that the travel patterns of taxis, including both occupied and unoccupied trips (passenger searching or vehicle relocation), can be better fitted by the truncated power-law distribution for shorter trips and by the exponential distribution for trips that are longer than 30 miles. Bazzani et al. [22] studied the GPS data of private cars in Florence, Italy and found that the single-trip length follows an exponential behavior in short distance scale but favors a power law distribution for trips longer than 30 km. Overall, trips made by vehicles (taxi or private car) are most frequently analyzed in recent studies. Because the majority of the car trips are in long distance, short trips (e.g., less than one mile) are often neglected. When excluding the short trips, existing studies show that the trip displacement follows either power-law [10,11,18] or exponential [17,19] distribution. On the other hand, when short trips are included in the analysis, gamma distribution [20] and lognormal distribution [1] may provide better fit.

Most of the previous studies focus on the travel patterns of trips taken by automobiles (e.g., taxis and private cars) because of the data availability of large-scale trip trajectory data recorded by GPS devices. Researchers have also analyzed other transportation modes such as walking [18] and flights [11] with travel survey or bank notes data, which has lower data accuracy on the trip information. Few studies paid attention to the mobility pattern of trips taken by bikes. Part of this lack of understanding is due to the lack of bike trip data. Kolb and Helbing [18] analyzed the relationship between the energy consumption rates (from physical activities) and the travel time of different transportation modes (including biking). However, the statistical pattern of biking trips was not analyzed in this study. As a critical component of the multi-modal transportation systems, bike trips are playing an increasingly important role in urban mobility. As one type of "active travel", biking not only can contribute to the sustainability of urban transportation by reducing traffic congestion and emissions, but also can help improve human health [23]. With the emerging sharing economy, more and more cities are implementing bike sharing programs, increasing the convenience of traveling by bikes in the city. As of January 2017, more than 100 cities in the United States have developed bike-sharing programs [24]. Additionally, the trip data from several bike sharing programs are made publicly available, providing an unprecedented opportunity to understand the human mobility using shared bikes. Such knowledge will provide fundamental basis for researchers to model the use of bike sharing and the associated multi-modal transportation systems, inform bike sharing system design and operation, and guide policy decisions for sustainable transportation development.

Previous studies about bike sharing mainly focus on the analysis of ridership [25–28] and the optimization of system operation [29–34], few studies focus on the statistical patterns of bike sharing trips. Jurdak [35] studied the bike sharing trips in Boston and Washington, D.C. in the United States, and concludes that the distribution of trip durations in both cities fit a power law distribution for long trips (longer than an hour). However, the distribution of long trips do not reflect the travel patterns of bike sharing trips, because the majority of the bike sharing trips are short (this will be discussed in Section 2 of this paper). Therefore, to fill the gap of understanding the travel patterns of bike sharing trips, this research analyzed the distribution of both trip distance and trip duration for bike sharing trips collected in eight cities in the U.S. The major contributions of this study include: (1) our analysis includes both short trips and long trips (generally, we set a threshold where the distance/duration reaches the highest probability (density) in the probability density plot; then trips with distance/duration that is larger than the threshold will be treated as long trips; for example, the threshold for duration varies from 5 to 13 min depending on the different cities and different trip purposes); (2) we compare the travel patterns of bike sharing trips with different trip purposes (commuting versus touristic trips); and (3) this study compares the travel patterns of bike sharing in eight cities, which include bike sharing systems of different sizes and operation time.

Table 1
Basic information of the eight bike sharing programs.

City	Seattle	Los Angeles	Bay area	Philadelphia	Boston	Washington, DC	Chicago	New York
Program name	Pronto	Metro	Ford Gobike	Indego	Hubway	Capital	Divvy	Citi
Program launch date	Oct., 2014	June, 2016	Sep., 2013	July, 2013	Dec., 2011	Jan., 2014	July, 2013	July, 2013
Total number of trips in year 2016	102,606	184,345	210,494	499,306	1,236,199	2,562,718	3,595,383	10,262,649
Number of stations as of Dec. 2016	59	64	74	119	327	407	581	687
System diameter (miles)	4.71	2.85	2.33	5.06	8.57	14.28	23.29	11.17

Note: the programs are listed in the order of the program size (the number of stations) from the smallest (left) to the largest (right).

2. Data and methods

2.1. Data

In this study, we analyzed the bike sharing trip data from eight programs located in eight different cities in the United States: Seattle, Los Angeles, Bay Area, Philadelphia, Boston, Washington D.C., Chicago and New York. The bike-sharing programs in these cities are station-based systems, in which a bike has to be checked out from and returned to a docking station. For each trip, the time and station is recorded when the bike is checked out and returned. The locations of the stations can be mapped with the station list, in which the longitude and latitude of each station is provided. Table SI-1 in the Supplemental Information shows a sample of the data from Pronto, the previous bike share program in Seattle (this program has now been closed). Data for other cities contains similar information. The bike sharing programs in different cities are launched at different time and have different scales (in terms of the number of stations and bikes). Table 1 shows an overview of the bike sharing programs in each city. In this study, we analyzed all trips taken in 2016 to study the travel patterns of bike sharing. To indicate the scale of spatial coverage of each program, we measured the system diameter for each program, which is defined as the longest Euclidian distance between any two stations (the stations list kept changing with the system expansion over time and we used the station list at the end of year 2016 in our analysis). Based on the ridership, the total number of stations, and the system diameter, the programs can be naturally separated into two groups, smaller systems (those in Seattle, Los Angeles, Bay Area, and Philadelphia) and larger systems (those in Boston, Washington DC, Chicago, and New York). One thing to note is that Bay Area's bike sharing system contains multiple sub-systems in several cities, thus we only include intra-city trips in San Francisco (accounts for 89.7% of the trips) in our analysis to make it comparable to other cities (more information on this is provided in Section SI.2 of the Supplementary Information).

2.2. Estimation of trip distance, duration, and purpose

Because the data only include trip origins (check-out station) and destinations (return station), the actual trajectory of each trip is not available. Therefore, the trip distance has to be estimated using the locations of the trip start and end stations. Considering the fact that the movement of bikes is likely restricted by the street network and not all roads (e.g., highways) can be traveled by bikes, to more accurately estimate the trip distance, we used the Google Maps Distance Matrix API to calculate the trip distance and duration between two stations, specifying biking as the transportation mode. This distance can better reflect the actual bike travel along the pathways suitable for biking. We also tested how trip start time can impact the trip distance and duration but did not observe significant differences (Google Maps Distance Matrix API only considers the real-time road traffic conditions for driving, not biking). Considering that bike travels are less impacted by road conditions (e.g., congestion), which is one of the reasons some people prefer biking to driving [36,37], we assumed that the travel distance and duration between the same pair of stations would stay the same regardless of the travel time. Because we do not have enough information to estimate the actual traveling distance of round trips, for which the bikes were checked out from and returned to the same station, round trips (consisting of 2% to 10% of the trips as shown in Table 2) are excluded from our analysis.

While many previous studies used trip displacement (e.g., the Great Circle Distance between the origin and destination) to measure the human movements [10,15], we think the actual distances traveled within the urban street network can better reflect human mobility using bikes. As shown in Fig. 1, the distance estimated by Google Maps API is on average about 1.5 times of the great circle distance, while the ratio can be as high as 8 times.

Bike sharing systems are mainly used for commuting and touristic trips [38]. The two types of trips will have very different trip duration patterns: commuters will bike directly from the start station to the end station while the tourists tend to stop at different points-of-interest locations along the way. As a result, the actual trip duration measured by the difference between trip end (bike return) and start (bike check-out) time can well reflect the traveling time for the commuting trips but will be much longer than the traveling time for the touristic trips. Because the original trip data do not contain trip purpose information, we compare the actual trip duration (the difference between trip start and end) to the expected trip duration (estimated by Google Maps API) to estimate the trip purpose. As shown in Fig. 2a, the trip duration estimated by Google Maps API is linear to the estimated distance, while the actual trip duration could be much longer than the estimated one. The distribution of the actual trip duration has a longer tail for long trips (e.g., longer than 15 min for Seattle) than that of the

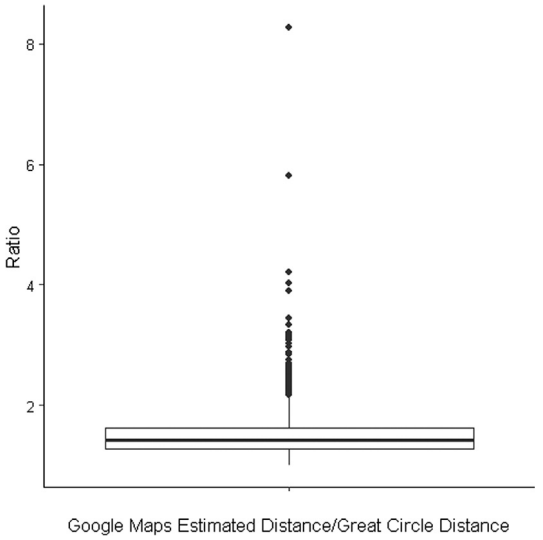

Fig. 1. Box plot of the ratio of Google Maps estimated distance to the trip displacement measured by the great circle distance (using data from Seattle as an example)

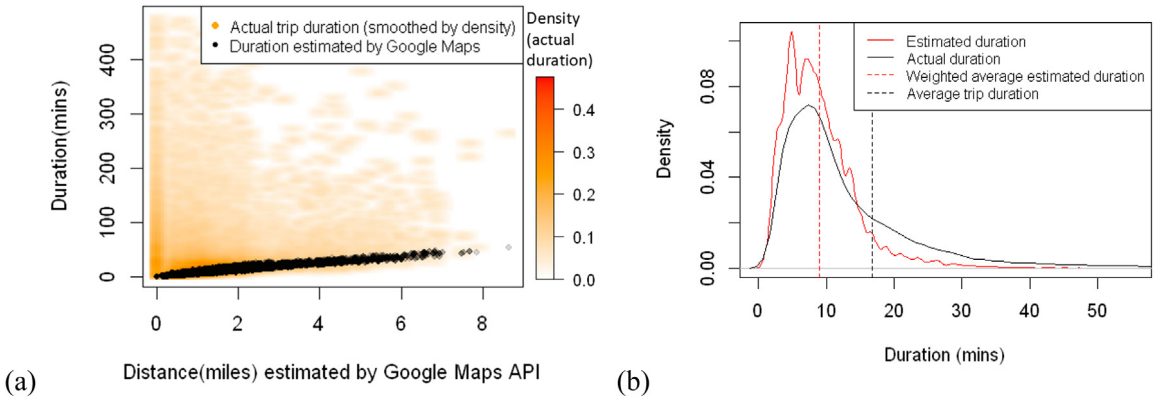

Fig. 2. Trip duration in Seattle: (a) Comparison of the relationship between the trip duration and the estimated distance for the actual and estimated trip duration and, (b) Trip duration density plot (the total trip between each origin–destination pair is used as the weight for the weighted average estimated duration calculation).

estimated duration (Fig. 2b). This means that many bike sharing users did not ride from the origin directly to the destination through the shortest route. Therefore, we define the trips whose duration is within 1.3 times of the duration estimated by the Google Maps API as the commuting trips and trips taken longer than that as touristic trips. The Google Maps API assumes a baseline speed of 10 miles/hour for biking [39]. This cutoff line of 1.3 is arbitrarily selected, considering a 30% buffering for possible delays (e.g., due to bike adjustment, necessary detours, lower biking speed) compared to the expected duration estimated by the Google Maps API. A sensitivity analysis is conducted to evaluate the impact of this cutoff line and is discussed in Section 3.3. The commuting trips make up about 59% to 76% of the total trips in the eight cities (Table 2) and are more concentrated in the shorter range (Fig. 3).

2.3. Methods for fitting the distributions of trip distance and duration

Short trips that are less than one mile are often neglected in existing studies analyzing travel patterns (for example, in [19,21]). However, because the majority of the bike sharing trips are relatively short (the median distance for both commuting and touristic trips is less than 1.5 miles, as shown in Table 2), it is critical that these short-distance trips are included in the analysis. Therefore, we evaluate the travel patterns both for all trips (including the short ones) and for long trips only (to allow comparison between our study and those focus on longer trips). Based on the shape of the density plot (Fig. 3), we first selected three common right-skewed distributions, Weibull, gamma, and lognormal as candidate distributions to fit the entire data (including both short and long trips). We also examined the Lévy flight nature of human

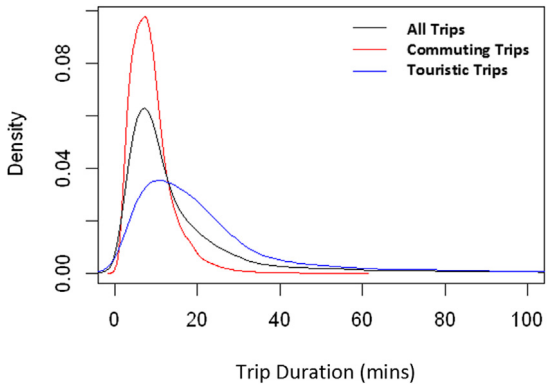

Fig. 3. Probability density plot of trip duration for commuting trips and touristic trips in Seattle.

Table 2
Basic statistics of the trip distance and duration data.

City			Seattle	Los Angeles	Bay area	Philadelphia	Boston	Washington, DC	Chicago	New York
Percentage of round trips			8.40%	10.60%	2.60%	8.10%	3.10%	4.60%	4.20%	2.06%
Percentage of commuting trips			58.43%	63.61%	75.55%	68.59%	68.98%	58.56%	60.78%	62.32%
Commuting trips	Distance (miles)	Maximum	7.85	3.14	3.37	7.30	8.77	27.21	23.77	13.88
		Median	1.16	0.91	1.23	1.39	1.46	1.43	1.39	1.31
		Average	1.27	1.02	1.24	1.56	1.70	1.72	1.71	1.68
		Standard deviation	0.74	0.50	0.53	0.79	1.02	1.14	1.14	1.21
	Duration (minutes)	Maximum	58.86	28.00	28.45	57.00	64.07	158.28	148.63	82.68
		Median	7.89	4.00	8.15	10.00	9.28	9.47	9.33	8.80
		Average	8.84	4.57	8.55	11.02	10.79	11.10	10.98	10.85
		Standard deviation	5.11	4.91	3.65	5.75	6.28	6.94	6.75	7.28
Touristic trips	Distance (miles)	Maximum	8.65	3.29	3.37	7.20	10.60	25.34	18.28	14.16
		Median	0.95	0.84	0.93	1.22	1.19	1.16	1.26	1.21
		Average	1.15	0.94	1.05	1.37	1.42	1.40	1.55	1.48
		Standard deviation	0.85	0.50	0.54	0.79	0.92	1.03	1.13	1.01
	Duration (minutes)	Maximum	479.66	719.00	716.70	719.00	719.37	719.72	719.85	719.98
		Median	17.39	14.00	11.45	20.00	16.42	16.40	17.03	15.30
		Average	27.97	23.62	21.20	31.58	22.41	26.01	21.89	20.09
		Standard deviation	39.33	44.66	43.98	50.49	33.01	37.60	26.78	27.09

Table 3
Probability density function of the models considered in this study ($x > 0$).

Model	Equation	Parameters
Weibull	$f(x) = \frac{k}{\lambda}(\frac{k}{\lambda})^{k-1}e^{-(x/\lambda)^k}$	Shape parameter k and scale parameter λ
Gamma	$f(x) = \frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)}$ where $\Gamma(\alpha)$ is a complete gamma function	Shape parameter α and scale (rate) parameter β
Lognormal	$f(x) = \frac{1}{x}\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(\ln x-\mu)^2}{2\sigma^2})$	Mean (log) μ and standard deviation (log) σ
Power Law	$f(x) = ax^{-k}$	Exponent k
Exponential	$f(x) = \lambda e^{-\lambda x}$	Rate parameter λ

traveling by fitting the Power Law and exponential distribution to the right-side tail (long trips). We then fit the trip data to the selected distributions, using the maximum likelihood estimation (MLE). To identify the goodness of fit, we conduct a Kolmogorov–Smirnov (K–S) test [40] and consider a K–S statistics value lower than 0.05 as acceptable for a good fit. If the K–S test considers multiple distributions as acceptable, we further use the Akaike information criterion (AIC) and Bayesian information criterion (BIC) methods [41] for model selection. The distribution with the lowest AIC and BIC is selected as the best fit [42].

3. Results and discussions

3.1. The distribution of trip distance and duration for all trips (short and long)

We fitted the trip distance and duration for the commuting and touristic trips separately with three common right-skewed distributions: lognormal, gamma, and Weibull (see Table 3 for the equations of these distributions). According to

Table 4
Results of model selection for commuting trip distance.

			Seattle	Los Angeles	Bay area	Philadelphia	Boston	Washington, DC	Chicago	New York
Weibull	Parameters	Shape	1.8512	2.1665	2.5134	2.0996	1.7896	1.6420	1.6144	1.5232
		Scale	1.4361	1.1499	1.4011	1.7712	1.9261	1.9317	1.9218	1.8760
	Goodness-of-fit statistics/ criteria	K–S statistic	**0.0658**	**0.0513**	**0.0497***	0.0555	0.0548	0.0519	0.0547	0.0674
		AIC	107316.3	137674.1	215312.1	693533.8	2161993.0	3880947.0	5717157.5	17356522.0
		BIC	107334.1	137693.2	215331.8	693555.1	2162016.3	3880971.3	5717182.6	17356549.0
Gamma	Parameters	Shape	3.5464	4.0595	4.9903	4.1294	3.0170	2.7110	2.5535	2.3567
		Scale	2.7932	3.9995	4.0206	2.6417	1.7713	1.5799	1.4937	1.4054
	Goodness-of-fit statistics/ criteria	K–S statistic	**0.0382***	**0.0304***	0.0682	**0.0284***	**0.0472***	**0.0354***	**0.0462***	0.0570
		AIC	101513.2	134402.0	216873.8	671542.5	2113945.2	3771463.5	5585434.3	16907864.0
		BIC	101531.1	134421.1	216893.5	671563.9	2113968.4	3771487.9	5585459.4	16907892.0
Lognormal	Parameters	meanlog	0.0912	−0.1133	0.1126	0.3208	0.3578	0.3444	0.3278	0.2899
		sdlog	0.5500	0.5264	0.4791	0.5136	0.6046	0.6345	0.6563	0.6783
	Goodness-of-fit statistics/ criteria	K–S statistic	0.0504	**0.0384***	0.0816	**0.0205***	**0.0237***	**0.0152***	**0.0176***	**0.0131***
		AIC	100235.9	138111.9	226762.0	673148.5	2104083.5	3722622.1	5514671.9	16539126.0
		BIC	100253.8	138131.0	226781.7	673169.8	2104106.7	3722646.5	5514697.0	16539154.0
Best fit distribution			Gamma	Gamma	Weibull	Gamma	Lognormal	Lognormal	Lognormal	Lognormal
Parameters			$\alpha = 3.55$ $\beta = 2.79$	$\alpha = 4.06$ $\beta = 4.00$	$k = 2.51$ $\lambda = 1.40$	$\alpha = 4.13$ $\beta = 2.64$	$\mu = 0.36$ $\sigma = 0.60$	$\mu = 0.34$ $\sigma = 0.63$	$\mu = 0.33$ $\sigma = 0.66$	$\mu = 0.29$ $\sigma = 0.68$

Notes: 1. * Significant at 0.05.
2. Model parameters corresponds to the parameters in Table 3.

the fitted results (Table 4), the eight systems do not have a common distribution pattern regarding trip distance for the commuting trips. For most of the systems except for Bay Area, both gamma and lognormal distributions provide a good fit with K–S statistics below the 5% significance level. Based on the AIC/BIC values, the best fit for the large systems (Boston, Washington DC, Chicago, New York) is lognormal distribution, while gamma distribution is a better fit for the smaller systems (Seattle, Los Angeles, and Philadelphia) except for Bay Area which fits better with the Weibull distribution. The details of the goodness of fit statistics for tourist trip distance, commuting trip duration, and tourist trip duration are listed in the Supplementary Information Tables SI-3 to SI-5, respectively.

Table 5 summarizes the best fit models as well as the parameters for all cities and all cases. In general, lognormal distribution provides the best fit for cities with larger bike sharing system (having a diameter larger than or around 10 miles), for both trip distance and distribution. The mean μ of lognormal distribution is between 0.29 and 0.36 miles for commuting trips and between 0.12 and 0.21 miles for touristic trips. Tourists tend to return the bikes in areas near the trip origins, and thus have shorter average trip "distance" than commuters (note that the distance here better reflect the real distance traveled by commuting trips than touristic trips). Regarding trip duration, on the contrary, tourists will travel longer time (mean μ is around 2.75 min for lognormal distribution) than commuters (mean μ is around 2.2 min), as expected. For cities with smaller system (whose diameter is less than 8 miles and have less than 100 stations), gamma distribution provides the best fit to describe the trip distance in most of the cases with some exceptions (the distance of commuting trips in Bay Area fits better with Weibull distribution; the distance of touristic trips in Seattle fits better with lognormal distribution). For trip duration, none of the selected distributions can fit many of the smaller systems with a significant K–S statistic. However, if we extend the significant level of the K–S statistics from 0.05 to 0.1, the lognormal distribution can be viewed as providing a good fit for the duration of touristic trips in all cities (Table SI-5 in the Supplemental Information). If we also apply such significant level extension to the trip distance analysis, our model selection will remain the same except that lognormal will be preferred for commuting trip distance.

We further evaluated the goodness-of-fit plots to gain more insights of the difference between the smaller and larger systems. In Fig. 4, we compared the goodness-of-fit from different perspectives for one smaller system (Seattle) and one larger system (Chicago). The P-P plots show that these models fit well at the distribution center, while the Q-Q plots reveal a lack of fit at the distribution tail. Because the bike sharing users of larger systems (e.g., Chicago) can potentially travel longer due to the larger spatial coverage (though only a small percentage of trips are long-distance), the variation at the distribution tail would be more significant and thus may influence the selected distribution models. It can be observed from Fig. 5 that Chicago has a wider-spread station network than Seattle and thus allows users to travel for longer distance. The reasons why we cannot find a common distribution model for the small systems would be complex. The system size is a major constrain for the users in these small systems. Because such systems have limited number of stations and small system diameters, the users' movement is also largely restricted in the area where bike sharing stations are available. Additionally, factors such as the stations' geographic layout, urban road network, and accessibility of bike paths may also affect the bike share users' behaviors.

Table 5
Parameters of the best fit models.

		Seattle	Los Angeles	Bay area	Philadelphia	Boston	Washington, DC	Chicago	New York
Commuting trip distance	Best fit distribution	Gamma	Gamma	Weibull	Gamma	Lognormal	Lognormal	Lognormal	Lognormal
	Parameters	$\alpha = 3.55$	$\alpha = 4.06$	$k = 2.51$	$\alpha = 4.13$	$\mu = 0.36$	$\mu = 0.34$	$\mu = 0.33$	$\mu = 0.29$
		$\beta = 2.79$	$\beta = 4.00$	$\lambda = 1.40$	$\beta = 2.64$	$\sigma = 0.60$	$\sigma = 0.63$	$\sigma = 0.66$	$\sigma = 0.68$
Touristic trip distance	Best fit distribution	Lognormal	Gamma	Gamma	Gamma	Gamma	Lognormal	Lognormal	Lognormal
	Parameters	$\mu = -0.07$	$\alpha = 3.33$	$\alpha = 3.65$	$\alpha = 3.08$	$\alpha = 2.62$	$\mu = 0.12$	$\mu = 0.21$	$\mu = 0.16$
		$\sigma = 0.63$	$\beta = 3.56$	$\beta = 3.46$	$\beta = 2.25$	$\beta = 1.84$	$\sigma = 0.67$	$\sigma = 0.70$	$\sigma = 0.70$
Commuting trip duration	Best fit distribution	Lognormal	None[1]	Gamma	None[1]	Lognormal	Lognormal	Lognormal	Lognormal
	Parameters	$\mu = 2.03$	–	$\alpha = 5.15$	–	$\mu = 2.21$	$\mu = 2.22$	$\mu = 2.21$	$\mu = 2.18$
		$\sigma = 0.56$	–	$\beta = 0.60$	–	$\sigma = 0.59$	$\sigma = 0.62$	$\sigma = 0.63$	$\sigma = 0.65$
Touristic trip duration	Best fit distribution	None[1]	None[1]	None[1]	Lognormal	Lognormal	Lognormal	Lognormal	Lognormal
	Parameters	–	–	–	$\mu = 3.03$	$\mu = 2.75$	$\mu = 2.80$	$\mu = 2.77$	$\mu = 2.69$
		–	–	–	$\sigma = 0.83$	$\sigma = 0.79$	$\sigma = 0.90$	$\sigma = 0.77$	$\sigma = 0.75$
System diameter		4.71	2.85	2.33	5.06	8.57	14.28	23.29	11.17

Notes: 1. "None" means that none of the three distributions have K–S statistic lower than 0.05.
2. Model parameters corresponds to the parameters in Table 3.

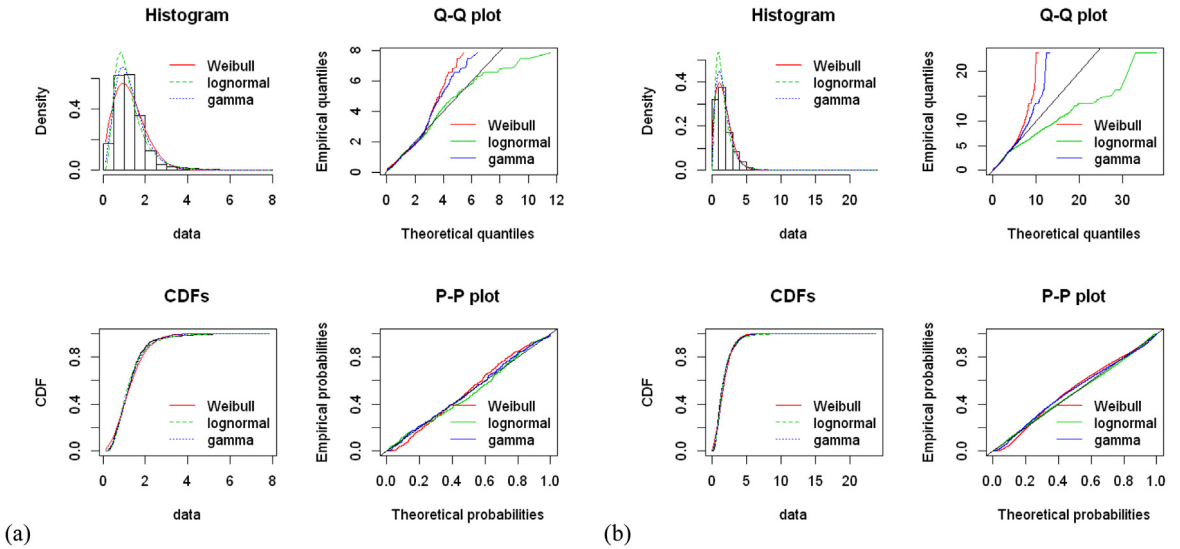

Fig. 4. Weibull, lognormal, and gamma distribution fitted to commuting trip distance: (a) Seattle, (b) Chicago (graphs for other cities are presented in Figures SI-2 to SI-5).

3.2. The distribution of trip distance and duration for long trips

As discussed in Section 3.1, the major factor that influenced model selection came from the goodness-of-fit in the tail part (long distance/duration). Therefore, analyzing the pattern in the right side tail part is also meaningful. The K–S statistics (Table 6) shows that the power law distribution provides a good fit for both the trip distance and duration for the larger systems (Boston, Washington DC, Chicago, and New York). The power law cut-off point (where the power law fitting starts) is determined by optimizing the goodness of fit. Generally, the power law fit for the larger systems starts at longer distance/duration. For those cases that power law exhibits a good fit, the exponent k of power law distribution varies from 5.5 to 12.6 for commuting trip distance, 3.4 to 12.1 for tourist trip distance, 5.4 to 16.3 for commuting trip duration, and 2.4 to 3.4 for tourist trip duration. The exponent k is basically in the same range in the case of trip distance, for both commuting and touristic trips. This is because that the data only reflects the distance between trip origin and destination. On the contrary, for trip duration, the exponent exhibits significant difference between commuting trips and touristic trips. Touristic trips have smaller exponent in the power law distribution fitting, showing that the probability that a trip has certain duration drops slower for touristic trips than commuting trips as the trip duration increases. In the study from Jiang et al. [1], they fitted the trip distance of taxi GPS trajectory data to power law. The exponent is around 2.5 for intra-city trajectories and 4–5

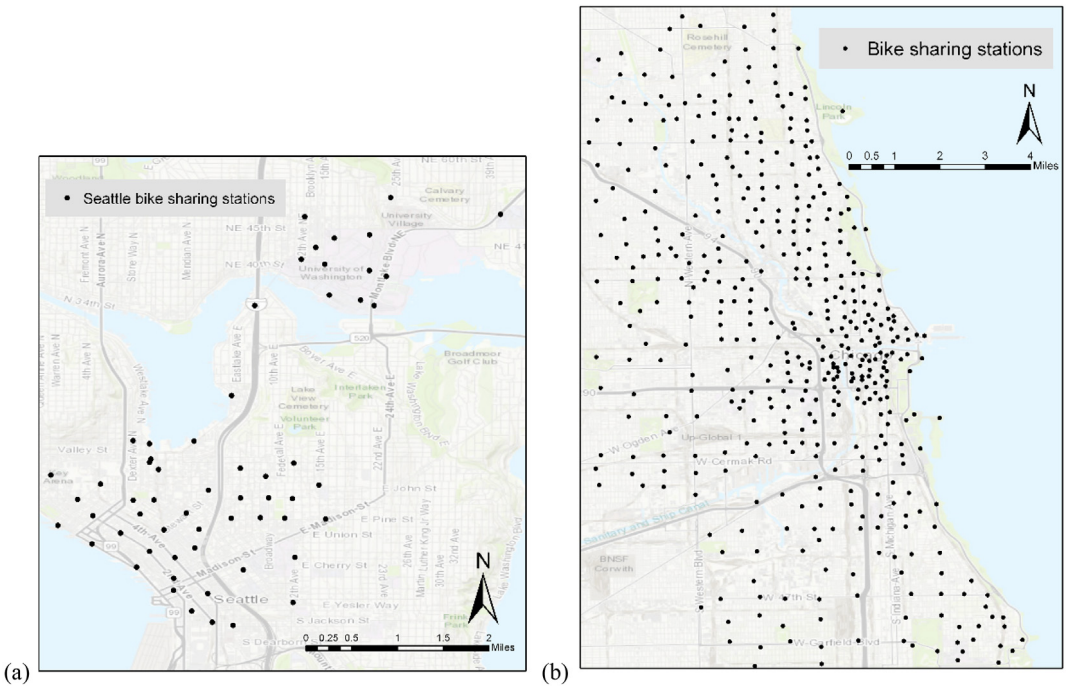

Fig. 5. Station network (black dots) and popular origin–destination pairs (based on historical trip count): (a) Seattle, (b) Chicago.

for intercity trajectories, respectively, which is similar or smaller than the exponent we fitted for bike sharing trips. This is expected since people is more likely to take long trips by taxi than by bike. Only several smaller bike sharing systems fit well with the power law distribution, which is due to the restriction from the station network, similar to our analysis in Section 3.1. With less restrictions from the bike sharing system, the mobility pattern in larger bike sharing systems tends to exhibit power law decay, which is similar to the observation in the research of Jiang et al. [1]. It is also worth noting that the trips in the tail part that are used for the power law fitting only make up a very small portion of the total trips (typically, less than 5%). The proportion is even smaller in the larger systems, which indicates that the probability of long trips is very low. In addition, we also fitted the right side tail part data to exponential distribution. None of the data fits well to the exponential distribution for any of the cities, neither commuting trips nor touristic trips. The details of the exponential fitting are included in Supplemental Information part SI.3.

We further plotted the Complementary Cumulative Distribution Function (CCDF) graphs to compare between the smaller and larger systems. We used CCDF graphs because each bin in the logarithmic tail part only contains very few samples, resulting in a messy tail in a histogram or probability density function (PDF) plot [21,43]. Fig. 6 compares the power law fitting of Seattle (smaller system) and Chicago (larger system). In the log–log range, the straight line of power law fitting (red line) is steeper for Chicago than for Seattle, which is consistent with the result that Chicago has larger exponent k.

3.3. Sensitivity analysis

Because we used a threshold of 1.3 times of the estimated trip duration as a filter to distinguish commuting and touristic trips, we also conducted a sensitivity analysis to evaluate how this threshold would impact the results. Fig. 7 provides a comparison between the original data and the filtered data (using different thresholds) on their power law fit. The original data shows a more evident decay in the tail part than the filtered data. Because the bike sharing trip data does not contain the detailed trace of each trip, the duration brings extra uncertainty to our analysis, for instance, we cannot know whether there is idle time during a trip. Even the tail part only makes up less than 6% of the total trips (Table 6), it can still contribute to a significant difference when fitting the distribution. We can observe that the decay and the change of the exponent k value of power law are contributed by those "touristic" trips. Tables 7 and 8 listed the results of model selection for Seattle and Chicago, with different thresholds. In general, when the filter is closer to 1, the commuting trips will include higher percentage of trips in which the users travel directly from the origin to destination through the shortest path. Seattle's results are more sensitive to this filter than that of Chicago. For Seattle's commuting trip distance, the best model for the whole distance range is gamma distribution when the filter is 1.1 and 1.3 times, while it prefers lognormal distribution when the threshold increases to 1.5 or when there is no filter. The power law fitting to the longer distance trips has a significant K–S statistic only for "no filter" case. Though gamma distribution remains to be a good fit no matter what the threshold

Table 6
Results of power law fit at the distribution tail.

		Seattle	Los Angeles	Bay area	Philadelphia	Boston	Washington, DC	Chicago	New York
Commuting trip distance (mile)	Right side cut-off[1]	1.14	0.67	1.26	0.99	1.05	0.86	0.77	0.86
	K–S statistic	0.0518	0.0711	0.1330	**0.0243**[*]	**0.0347**[*]	**0.0287**[*]	**0.0176**[*]	**0.0347**[*]
	Power law cut-off[2]	2.15	2.12	2.10	3.35	5.47	4.07	4.49	6.29
	Exponent k	4.28	12.07	11.70	8.27	12.60	5.51	7.25	11.44
	Percentage of trips in power law fit	4.74%	1.94%	5.81%	2.22%	0.24%	2.41%	1.67%	0.39%
Tourist trip distance (mile)	Right side cut-off	0.81	0.69	0.90	0.69	0.68	0.94	1.04	0.75
	K–S statistic	**0.0496**[*]	0.0573	0.1097	**0.0271**[*]	**0.0407**[*]	**0.0234**[*]	**0.0213**[*]	**0.0230**[*]
	Power law cut-off	1.08	2.09	2.09	3.35	5.37	2.98	9.26	8.10
	Exponent k	3.36	10.63	10.32	8.60	11.18	4.27	8.85	12.09
	Percentage of trips in power law fit	17.77%	0.93%	1.29%	0.72%	0.08%	2.47%	0.03%	0.00%
Commuting trip duration (min)	Right side cut-off	7.65	0.24	7.45	8.07	5.83	5.58	5.30	5.29
	K–S statistic	**0.0283**[*]	0.1498	**0.0299**[*]	0.1453	**0.0161**[*]	**0.0167**[*]	**0.0132**[*]	**0.0070**[*]
	Power law cut-off	17.16	11.78	19.27	10.00	27.40	23.33	31.47	40.98
	Exponent k	5.36	5.86	16.32	3.74	9.19	6.16	6.94	12.48
	Percentage of trips in power law fit	3.95%	6.42%	0.33%	35.59%	1.17%	3.42%	0.49%	0.16%
Tourist trip duration (min)	Right side cut-off	11.36	8.67	9.24	12.82	9.45	5.95	9.98	10.40
	K–S statistic	**0.0162**[*]	**0.0258**[*]	0.0515	**0.0242**[*]	**0.0166**[*]	**0.0129**[*]	**0.0050**[*]	**0.0260**[*]
	Power law cut-off	18.32	42.00	10.87	65.00	56.77	119.37	34.48	56.02
	Exponent k	2.46	2.40	2.41	2.55	2.72	3.42	3.07	2.62
	Percentage of trips in power law fit	19.59%	3.28%	12.93%	2.36%	1.37%	1.03%	4.76%	0.97%

1. Right side cut-off is the distance/duration with the highest probability (density) in the probability density plot.
2. Power law cut-off is the lower cut-off for the scaling region estimated by a goodness-of-fit based approach.
3. [*]Significant at 0.05 level.
4. Model parameters corresponds to the parameters in Table 3.

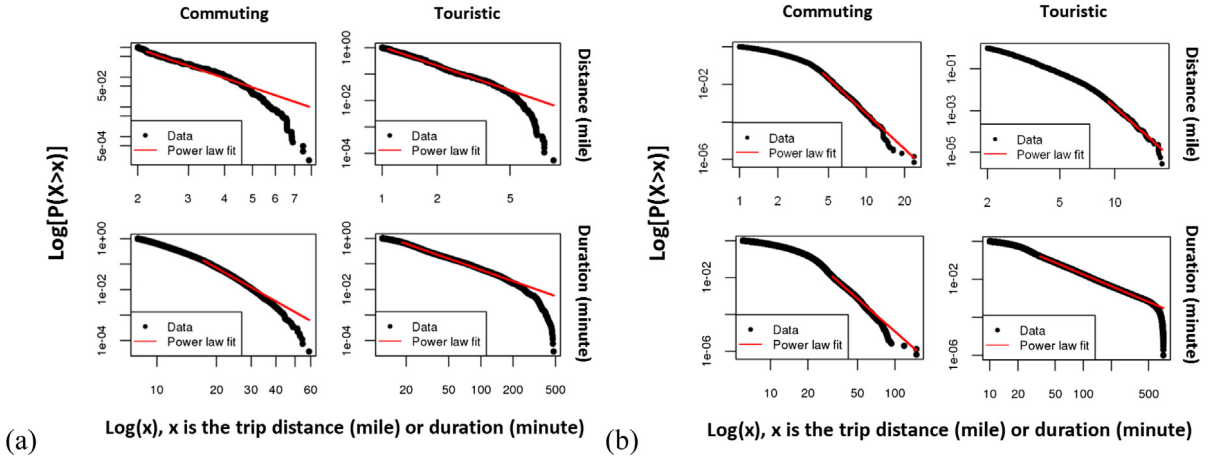

Fig. 6. CCDF plots of the power law fit: (a) Seattle, (b) Chicago (graphs of other cities are presented in Figures SI-9 to SI-12).

is. The shape parameter α and scale (rate) parameter β will decrease slightly (α: 3.84–3.06, β: 2.85–2.51) as the threshold increases. Regarding Seattle's commuting trip duration, the best fitted model switches from gamma to lognormal when the filter is larger or equal to 1.3. However, none of the right-skewed models have a significant K–S statistic for the non-filter case. For power law fitting, it presents a good fit no matter what the threshold is, but the exponent of the fitted model varies from 8.2 to 2.6. In contrast to Seattle, the threshold did not affect the model selection for Chicago to a very large extent. Lognormal remains to be the best model and power law fits well to the tail in all the cases when the threshold changes. For commuting trip distance, the lognormal model has the mean (log) μ in the range of 0.28–0.33 and standard deviation (log) σ in the range of 0.65–0.68; while the power law exponent k is in the range of 7.6–8.4. The power law exponent k for commuting trip duration may change a lot between nonfiltered and filtered data, since the nonfiltered data has an exponent k of 3.11, while the filtered data has exponent k in range of 6.9 to 8.6. Therefore, when selecting a distribution to model the trip duration of the bike sharing user's behaviors, we should be more cautious for those smaller systems (e.g., Seattle). Again, the variations or sensitivities of the model selection or model parameters are more likely to come from the bike sharing station networks, which could be the reason why it is difficult to find a common distribution model for those smaller systems.

Table 7
Results of sensitivity analysis, Seattle.

Filter				1.1	1.3	1.5	No filter
Commuting trip distance (mile)	Weibull	Parameters	Shape	1.8797	1.8512	1.8233	1.6971
			Scale	1.4489	1.4361	1.4234	1.3756
		Goodness-of-fit statistics/criteria	K–S statistic	0.0699	0.0658	0.0638	0.0643
			AIC	81034.01	107316.3	124027.2	186907.1
			BIC	81051.28	107334.1	124045.3	186926
	Gamma	Parameters	Shape	3.6446	3.5464	3.4488	3.0631
			Scale	2.8458	2.7932	2.7399	2.5124
		Goodness-of-fit statistics/criteria	K–S statistic	**0.0420***	**0.0382***	**0.0365***	**0.0377***
			AIC	76656.79	101513.2	117357.5	176672.1
			BIC	76674.07	101531.1	117375.6	176691
	Lognormal	Parameters	meanlog	0.1040	0.0912	0.0782	0.0262
			sdlog	0.5433	0.5500	0.5573	0.5897
		Goodness-of-fit statistics/criteria	K–S statistic	0.0585	0.0504	**0.0468***	**0.0395***
			AIC	75912.96	100235.9	115626.1	172444.9
			BIC	75930.23	100253.8	115644.2	172463.8
		Power law	Right side cut-off point	1.1615	1.1420	1.1442	0.8195
			K–S statistic	0.0657	0.0518	0.0504	**0.0425***
			Lower cut-off for the scaling region	2.1468	2.1468	2.1468	1.4428
			Exponent k	4.1651	4.2793	4.2616	4.0146
			No. of points in tail part	3205	4454	5165	26861
Commuting trip duration (min)	Weibull	Parameters	Shape	1.9528	1.8550	1.7833	0.9983
			Scale	9.4031	10.0013	10.4639	16.7761
		Goodness-of-fit statistics/criteria	K–S statistic	**0.0493***	0.0586	0.0622	0.1253
			AIC	234836.5	320810.9	379173.7	718543.6
			BIC	234853.7	320828.8	379191.8	718562.5
	Gamma	Parameters	Shape	3.7133	3.4615	3.2475	1.2754
			Scale	0.4471	0.3917	0.3508	0.0760
		Goodness-of-fit statistics/criteria	K–S statistic	**0.0214***	**0.0296***	**0.0348***	0.1312
			AIC	231495.9	315744.1	373111.4	715311.8
			BIC	231513.1	315762	373129.5	715330.7
	Lognormal	Parameters	meanlog	1.9762	2.0276	2.0635	2.3803
			sdlog	0.5428	0.5589	0.5752	0.8267
		Goodness-of-fit statistics/criteria	K–S statistic	**0.0393***	**0.0319***	**0.0245***	0.0537
			AIC	231530.2	314757.4	371196	678694.2
			BIC	231547.5	314775.2	371214.1	678713.1
		Power law	Right side cut-off point	7.3525	7.6474	7.2725	6.1040
			K–S statistic	**0.0213***	**0.0283***	**0.0257***	**0.0214***
			Lower cut-off for the scaling region	24.2198	17.1621	22.8324	15.4299
			Exponent k	8.1714	5.3556	5.6912	2.5627
			No. of points in tail part	396	3715	1835	27158

1. Right side cut-off is the distance/duration with the highest probability (density) in the probability density plot.
2. Power law cut-off is the lower cut-off for the scaling region estimated by a goodness-of-fit based approach.
3. * Significant at 0.05 level.
4. Model parameters corresponds to the parameters in Table 3.

4. Conclusion and future work

This paper analyzed the mobility patterns of bike sharing users by identifying the distributions of the trip distance and duration. Because the bike sharing trip data used in our analysis only contain the time and location information of the trip origins and destinations, we estimated the trip distance using the Google Maps Distance Matrix API and also separated the data into commuting trips and touristic trips. No common distributions are found that can describe the bike sharing system for all cities. For the cities with larger bike sharing systems (i.e., Boston, Washington DC, Chicago, and New York), both the trip distance and the trip duration follow a lognormal distribution. For trip distance, the mean (μ) of the lognormal distribution is between 0.29 and 0.36 miles for commuting trips and between 0.12 and 0.21 miles for touristic trips. For trip duration, the mean (μ) is around 2.2 minutes for commuting trips and 2.8 minutes for touristic trips. For the cities with smaller bike sharing systems, the best-fit model varies among different cities and different cases. The systems' spatial coverage is a major factor that influences the model selection. Small system size (diameter and number of stations) can restrict bike sharing users' mobility pattern because the users are required to return the bike to a station. The major difference between the distance/duration distribution of smaller and larger systems lies in the right side tail part. Similar to other transportation modes, trips by shared bikes also exhibit power law decay at the long distance/duration range, although the part of trips

Table 8
Results of sensitivity analysis, Chicago.

Filter				1.1	1.3	1.5	No filter
Commuting trip distance (mile)	Weibull	Parameters	Shape	1.6093	1.6144	1.6013	1.5664
			Scale	1.9272	1.9218	1.9067	1.8476
		Goodness-of-fit statistics/criteria	K–S statistic	0.0593	0.0547	0.0528	**0.0484***
			AIC	3913715.2	5717157.5	6783809.6	9323442.1
			BIC	3913739.5	5717182.6	6783835.1	9323468.2
	Gamma	Parameters	Shape	2.5491	2.5535	2.5170	2.4287
			Scale	1.4867	1.4937	1.4828	1.4734
		Goodness-of-fit statistics/criteria	K–S statistic	**0.0497***	**0.0462***	**0.0447***	**0.0395***
			AIC	3820833.9	5585434.3	6628840.4	9108608.9
			BIC	3820858.2	5585459.4	6628865.8	9108635
	Lognormal	Parameters	meanlog	0.3303	0.3278	0.3175	0.2800
			sdlog	0.6551	0.6563	0.6619	0.6767
		Goodness-of-fit statistics/criteria	K–S statistic	**0.0187***	**0.0176***	**0.0184***	**0.0174***
			AIC	3764514.8	5514671.9	6546544.6	9007262.6
			BIC	3764539.1	5514697	6546570	9007288.7
		Power Law	Right side cut-off point	1.0000	0.7733	0.7776	0.7907
			K–S statistic	**0.0241***	**0.0176***	**0.0158***	**0.0182***
			Lower cut-off for the scaling region	4.4925	4.4931	4.3962	8.2686
			Exponent k	7.6295	7.2541	6.8692	8.4114
			No. of points in tail part	43120	58166	76369	3625
Commuting trip duration (minute)	Weibull	Parameters	Shape	1.7504	1.7394	1.7063	1.2125
			Scale	11.6227	12.3974	12.8541	16.4961
		Goodness-of-fit statistics/criteria	K–S statistic	**0.0466***	**0.0439***	**0.0426***	0.0869
			AIC	8862380.8	13275483	16020129	25386248
			BIC	8862405.1	13275508	16020154	25386274
	Gamma	Parameters	Shape	2.8790	2.8419	2.7623	1.8332
			Scale	0.2798	0.2588	0.2423	0.1198
		Goodness-of-fit statistics/criteria	K–S statistic	**0.0339***	**0.0328***	**0.0329***	**0.0458***
			AIC	8785794.8	13166770	15885710	25017980
			BIC	8785819.2	13166795	15885736	25018007
	Lognormal	Parameters	meanlog	2.1477	2.2101	2.2417	2.4307
			sdlog	0.6207	0.6266	0.6358	0.7407
		Goodness-of-fit statistics/criteria	K–S statistic	**0.0236***	**0.0274***	**0.0280***	**0.0216***
			AIC	8769914.6	13152296	15863341	24423729
			BIC	8769938.9	13152321	15863366	24423755
		Power Law	Right side cut-off point	4.9947	5.2965	5.2826	7.0246
			K–S statistic	**0.0105***	**0.0132***	**0.0126***	**0.0047***
			Lower cut-off for the scaling region	32.1333	31.4667	58.0167	38.0667
			Exponent k	7.6032	6.9416	8.5796	3.1133
			No. of points in tail part	5396	17134	1048	139703

1. Right side cut-off is the distance/duration with the highest probability (density) in the probability density plot.
2. Power law cut-off is the lower cut-off for the scaling region estimated by a goodness-of-fit based approach.
3. * Significant at 0.05 level.
4. Model parameters corresponds to the parameters in Table 3.

following the power law distribution only makes up a very small portion of the total trips. The exponent k of power law distribution varies from 5.5 to 12.6 for commuting trip distance, 3.4 to 12.1 for tourist trip distance, 5.4 to 16.3 for tourist trip duration, and 2.4 to 3.4 for tourist trip duration. In our study, we define the trips whose duration is within 1.3 times of the duration estimated by the Google Maps API as the commuting trips and trips taken longer than that as touristic trips. In the sensitivity analysis of this threshold, we find that the model selection and model parameters are more sensitive to this threshold for Seattle (a smaller system) than Chicago (a larger system). The distribution fitting is also more sensitive to the threshold for trip duration than for trip distance.

While this study fills the gap of analyzing the travel patterns of bike sharing trips considering both short and long trips, we would also want to note the limitations of this work for future research. A major limitation of our analysis is the limited information contained in the raw data (having only trip origin and destination time and location). If the trajectory data of the bike sharing trips become available, the actual trip distance and duration can be more accurately calculated. In addition, more detailed analysis about the transportation network, pricing scheme, user preference, and other socio-geographic information may also help explain the difference in trip distance and duration distributions among different cities. Lastly, this research focuses on the travel patterns of trips using station-based bike sharing programs, which may have different travel behaviors compared to the trips using the newly emerged dockless bike sharing programs [44]. Future

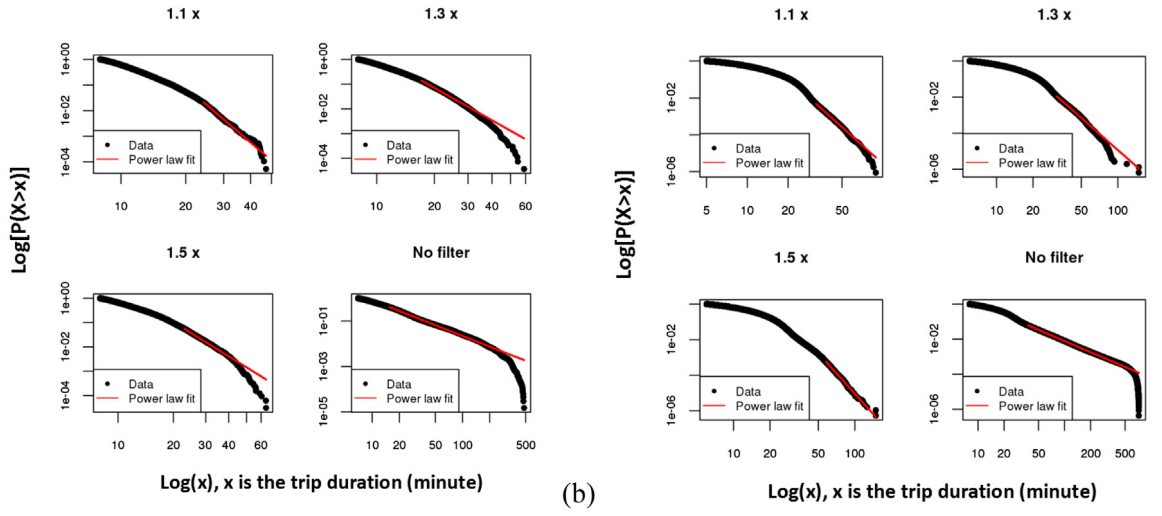

Fig. 7. CCDF plots of the power law fit to commuting trip duration with different thresholds: (a) Seattle, (b) Chicago.

research evaluating the difference between the two types of programs can provide a more complete understanding of the travel patterns of bike sharing users.

Acknowledgment

The authors would like to thank the support from the Summer Exploratory Research Grants provided by Environmental and Ecological Engineering, Purdue University.

Appendix A. Supplementary information

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.physa.2018.09.123.

References

[1] B. Jiang, J. Yin, S. Zhao, Characterizing the human mobility pattern in a large street network, Phys. Rev. E 80 (2009) 1–11, http://dx.doi.org/10.1103/PhysRevE.80.021136.
[2] R. Kitamura, C. Chen, R.a.M.M. Pendyala, R. Narayanan, Micro-simulation of daily activity-travel patterns for travel, Transportation (Amst) 27 (2000) 25–51, http://dx.doi.org/10.1023/A:1005259324588.
[3] M.W. Horner, M.E. O'Kelly, Embedding economies of scale concepts for hub network design, J. Transp. Geogr. 9 (2001) 255–265, http://dx.doi.org/10.1016/S0966-6923(01)00019-9.
[4] L. Mari, E. Bertuzzo, L. Righetto, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, A. Rinaldo, Modelling cholera epidemics: The role of waterways, human mobility and sanitation, J. R. Soc. Interface 9 (2012) 376–388, http://dx.doi.org/10.1098/rsif.2011.0304.
[5] L. Hufnagel, D. Brockmann, T. Geisel, Forecast and control of epidemics in a globalized world, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 15124–15129, http://dx.doi.org/10.1073/pnas.0308344101.
[6] B.T. Grenfell, O.N. Bjørnstad, J. Kappey, Travelling waves and spatial hierarchies in measles epidemics, Nature 414 (2001) 716–723, http://dx.doi.org/10.1038/414716a.
[7] M.J. Keeling, M.E. Woolhouse, D.J. Shaw, L. Matthews, M. Chase-Topping, D.T. Haydon, S.J. Cornell, J. Kappey, J. Wilesmith, B.T. Grenfell, Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape, Science (80-.) 294 (2001) 813–817, http://dx.doi.org/10.1126/science.1065973.
[8] C. Nash, Genetics, race, and relatedness: Human mobility and human diversity in the genographic project, Ann. Assoc. Am. Geogr. 102 (2012) 667–684, http://dx.doi.org/10.1080/00045608.2011.603646.
[9] K.L.K. Lee, S.H.S. Hong, S.J.K.S.J. Kim, I.R.I. Rhee, S.C.S. Chong, SLAW: A new mobility model for human walks, IEEE Infocom 2009 (2009) 855–863, http://dx.doi.org/10.1109/INFCOM.2009.5061995.
[10] I. Rhee, I. Rhee, M. Shin, M. Shin, S. Hong, K. Lee, S. Hong, S.J. Kim, K. Lee, S. Chong, S. Chong, On the levy-walk nature of human mobility: Do humans walk like monkeys? IEEE/ACM Trans. Netw. 19 (2011) 630–643, http://dx.doi.org/10.1109/TNET.2011.2120618.
[11] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, Nature 439 (2006) 462–465, http://dx.doi.org/10.1038/nature04292.
[12] G.M. Viswanathan, V. Afanasyev, S.V. Buldyrev, E.J. Murphy, P.A. Prince, H.E. Stanley, Lévy flight search patterns of wandering albatrosses, Nature 381 (1996) 413–415, http://dx.doi.org/10.1038/381413a0.
[13] R.P.D. Atkinson, C.J. Rhodes, D.W. Macdonald, R.M. Anderson, Scale-free dynamics in the movement patterns of jackals, Oikos 98 (2002) 134–140, http://dx.doi.org/10.1034/j.1600-0706.2002.980114.x.
[14] G. Ramos-Fernández, J.L. Mateos, O. Miramontes, G. Cocho, H. Larralde, B. Ayala-Orozco, Lévy walk patterns in the foraging movements of spider monkeys (Ateles geoffroyi), Behav. Ecol. Sociobiol. 55 (2004) 223–230, http://dx.doi.org/10.1007/s00265-003-0700-6.
[15] M.C. González, C.A. Hidalgo, A.-L. Barabási, Understanding individual human mobility patterns, Nature 453 (2008) 779–782, http://dx.doi.org/10.1038/nature06958.

[16] C. Song, T. Koren, P. Wang, A.-L. Barabási, Modelling the scaling properties of human mobility, Nat. Phys. 6 (2010) 818–823, http://dx.doi.org/10.1038/nphys1760.
[17] X.-Y. Yan, X.-P. Han, B.-H. Wang, T. Zhou, Diversity of individual mobility patterns and emergence of aggregated scaling laws, Sci. Rep. 3 (2013) 2678, http://dx.doi.org/10.1038/srep02678.
[18] R. Koelbl, D. Helbing, Energy and scaling laws in human travel behaviour, New J. Phys. 5 (2003) 1–48, http://dx.doi.org/10.1088/1367-2630/5/1/348.
[19] X. Liang, X. Zheng, W. Lv, T. Zhu, K. Xu, The scaling of human mobility by taxis is exponential, Physica A 391 (2012) 2135–2144, http://dx.doi.org/10.1016/j.physa.2011.11.035.
[20] M. Veloso, S. Phithakkitnukoon, Exploratory study of urban flow using taxi traces, First Work. .... 2011, pp. 1–8. http://www.researchgate.net/publication/232175450_Exploratory_Study_of_Urban_Flow_using_Taxi_Traces/file/79e41507829c808c03.pdf.
[21] H. Cai, X. Zhan, J. Zhu, X. Jia, A.S.F. Chiu, M. Xu, Understanding taxi travel patterns, Physica A 457 (2016) 590–597, http://dx.doi.org/10.1016/j.physa.2016.03.047.
[22] A. Bazzani, B. Giorgini, S. Rambaldi, R. Gallotti, L. Giovannini, Statistical laws in urban mobility from microscopic GPS data in the area of Florence, J. Stat. Mech: Theory Exp. 2010 (2010) P05001, http://dx.doi.org/10.1088/1742-5468/2010/05/P05001.
[23] C. Bullock, F. Brereton, S. Bailey, The economic contribution of public bike-share to the sustainability and efficient functioning of cities, Sustain. Cities Soc. 28 (2017) 76–87, http://dx.doi.org/10.1016/j.scs.2016.08.024.
[24] D. Malouff, All 119 US bikeshare systems, ranked by size, Gt. Gt. Washingt. 2017. https://ggwash.org/view/62137/all-119-us-bikeshare-systems-ranked-by-size (accessed 16.11.17).
[25] X. Zhou, Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago, PLoS One 10 (2015) 1–20, http://dx.doi.org/10.1371/journal.pone.0137922.
[26] O. O'Brien, J. Cheshire, M. Batty, Mining bicycle sharing data for generating insights into sustainable transport systems, J. Transp. Geogr. 34 (2014) 262–273, http://dx.doi.org/10.1016/j.jtrangeo.2013.06.007.
[27] C. Ome, O. Latifa, Model-based count series clustering for bike sharing system usage mining: A case study with the vélib system of Paris, ACM Trans. Intell. Syst. Technol. 5 (2014) 1–21, http://dx.doi.org/10.1145/2560188.
[28] I. Mateo-Babiano, R. Bean, J. Corcoran, D. Pojani, How does our natural and built environment affect the use of bicycle sharing? Transp. Res. Part A 94 (2016) 295–307, http://dx.doi.org/10.1016/j.tra.2016.09.015.
[29] C. Contardo, L.-M. Rousseau, C. Morency, Balancing a dynamic public bike-sharing system, First Conf. EURO Work. Gr. Veh. Routing Logist. Optim. VeRoLog. 2012, pp. 57–58. http://www.verolog.eu/.
[30] C. Fricker, N. Gast, Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity, Eur. J. Transp. Logist. 5 (2016) 261–291, http://dx.doi.org/10.1007/s13676-014-0053-5.
[31] J.C. García-Palomares, J. Gutiérrez, M. Latorre, Optimizing the location of stations in bike-sharing programs: A GIS approach, Appl. Geogr. 35 (2012) 235–246, http://dx.doi.org/10.1016/j.apgeog.2012.07.002.
[32] D. Chemla, T. Pradeau, R.W. Calvo, D. Chemla, T. Pradeau, R.W. Calvo, Self-service bike sharing systems : simulation, repositioning, pricing To cite this version : 2013.
[33] J. Brinkmann, M.W. Ulmer, D.C. Mattfeld, Short-term strategies for stochastic inventory routing in bike sharing systems, Transp. Res. Procedia 10 (2015) 364–373, http://dx.doi.org/10.1016/j.trpro.2015.09.086.
[34] Y. Li, Y. Zheng, H. Zhang, L. Chen, Traffic prediction in a bike-sharing system, Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst. - GIS ' 15, 2015, pp. 1–10. http://dx.doi.org/10.1145/2820783.2820837.
[35] R. Jurdak, The impact of cost and network topology on urban mobility: A study of public bicycle usage in 2 U.S. cities, PLoS One 8 (2013) 1–6, http://dx.doi.org/10.1371/journal.pone.0079396.
[36] L. Alter, Bikes Faster than Subway or Car, Seven Years Running, TreeHugger. 2008. https://www.treehugger.com/bikes/bikes-faster-than-subway-or-car-seven-years-running.html (accessed 18.11.17).
[37] A. Lobo, Bikes are faster door-to-door than cars or public transport within 5-10km, Better by Bicycl. 2013. http://www.betterbybicycle.com/2013/12/bikes-are-faster-door-to-door-than-cars.html (accessed 18.11.17).
[38] LDA Consulting, Capital Bikeshare 2016 Member Survey Report, 2016.
[39] A. Lobo, How accurate are Google Maps cycling time estimates? Better by Bicycl. 2014. http://www.betterbybicycle.com/2014/09/how-accurate-are-google-maps-cycling.html (accessed 01.01.18).
[40] S.M. D'Agostino RB, Goodness-of-Fit Techniques, first ed., Dekker, 1986.
[41] J.B. Johnson, K.S. Omland, Model selection in ecology and evolution, Trends Ecol. Evol. 19 (2004) 101–108, http://dx.doi.org/10.1016/j.tree.2003.10.013.
[42] D. Posada, T.R. Buckley, Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests, Syst. Biol. 53 (2004) 793–808, http://dx.doi.org/10.1080/10635150490522304.
[43] M.E.J. Newman, Power laws, Pareto distributions and Zipf's law, 7514, 2004. http://dx.doi.org/10.1016/j.cities.2012.03.001.
[44] S.a. Shaheen, A.P. Cohen, E.W. Martin, Public bikesharing in North America: Early operator understanding and emerging trends, Transp. Res. Rec. J. Transp. Res. Board 2387 (2013) 83–92, http://dx.doi.org/10.3141/2387-10.