香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Oral Presentation for PhD Qualifying Examination

Data-driven analysis of matching probability, routing distance and detour distance in ride-pooling service

Committee members:
Prof. Hai YANG
Prof. Hong K. LO
Prof. Sisi JIAN

Candidate: Zhengfei ZHENG

Supervisor: Prof. Hai YANG

Department of Civil and Environmental Engineering
The Hong Kong University of Science and Technology, Hong Kong

June 16, 2020

Outline

Introduction

- Background
- Related literature

Methodology and analysis

- Key indicators
- Data preparation
- Empirical law

Future study

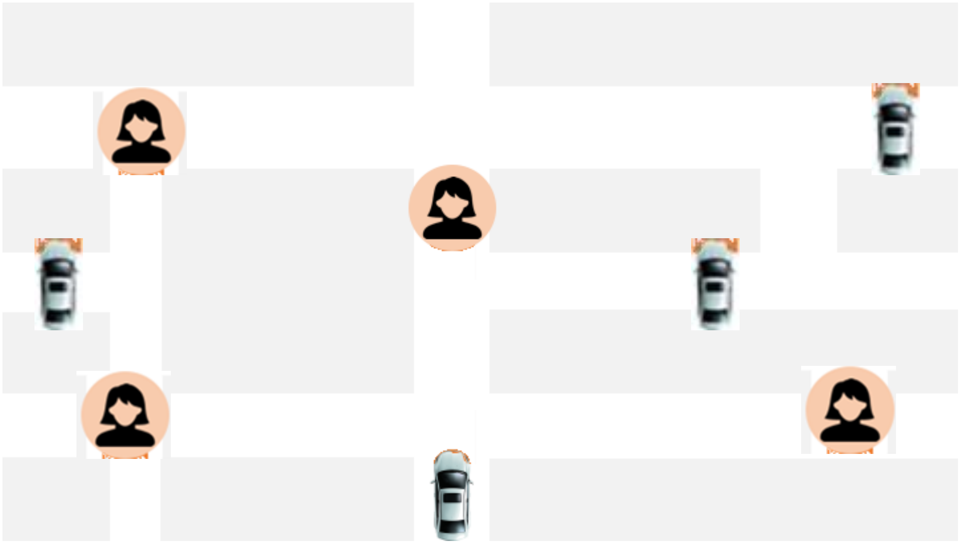
- Dynamic simulator
- Supply side

/01

Introduction

- Background
- Brief literature review

Introduction

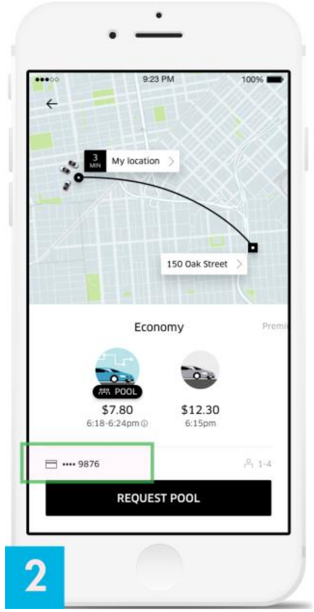


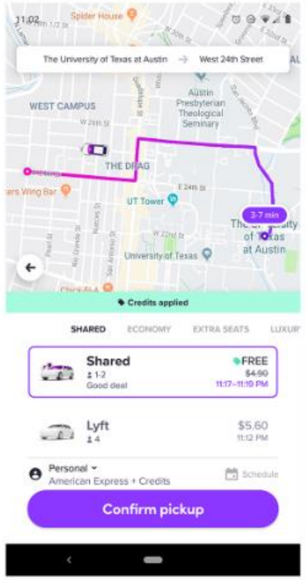
It is reported that Uber has served more than 10 billion passenger requests in over 80 countries while Didi Chuxing is now serving over 30 million rides per day in China.

Terminology: Ride pooling/ ride-splitting
It refers to a ride sourcing service in which passengers can opt to split both a ride and the fare (i.e., like dynamic carpooling).

Related literature

Typical examples are UberPool, Didi Express Pool and Lyft Share.





Ride-pooling services are expected to improve vehicle utilization rate and achieve societal benefits such as traffic congestion alleviation and emission reduction, especially during the peak hour.

A similar research shows cumulative trip length can be cut by 40% or more if taxi trips can be shared in Manhattan. (Nature,2018)

Related literature

Nonetheless, there exist some doubts about the effectiveness and efficiencies of ride-pooling service, particularly their impacts on traffic congestion.

Schaller (2018) pointed out

Shared services add 2.6 new TNC miles for removing each mile in personal driving (This is based on the current rate of about 20 percent of TNC trips being shared.)

This negative effects warrant a need for a better understanding and efficient operations of the emerging ride-pooling service

Related literature

Ride pooling relevant

Yan et al. (2019) examined the steady-state of a market with ride-splitting services and showed that the joint optimization of dynamic pricing and matching can help improve vehicle utilization and reduce passenger waiting time

Yang *et al.* (2019) studied the complex relations among matching radius, passengers' average detour time, and demand for ride-splitting services in a two-sided market with matching frictions

Ke *et al.* (2019) investigated the equilibrium properties of a ride-splitting market

Some of these studies are rooted in earlier studies on conventional taxi markets

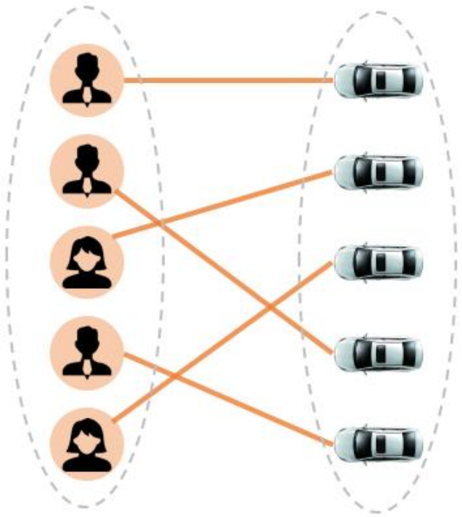
Data-driven relevant

Chen et al., (2017) develop an ensemble learning approach to investigate the key factors that affect passengers' choices between non-ride-splitting and ride-splitting services

Li et al., (2019) find that the proportion of passengers using ride-splitting services is still low (6-7%) due to long extra detour and degraded travel time reliability.

Yang, Ke and Ye, (2018) discover a universal law of detour ratios which states that the detour ratio (the ratio of average actual road distance to straight-line distance) is a constant plus a term inversely proportional to the straight-line distance

Introduction



Aggregated analysis exhibit the inability to fully characterize the ride-pooling service system using overly abstracted mathematical models.

None of these data-driven studies provides explicit formulations to characterize how pool-matching probability, passenger detour distance and driver routing distance change with passenger demand and other factors, in a general case.

To tackle this challenge, based-on real trip dataset, this study makes a first attempt to empirically examine the relationships between the three key measures (pool-matching probability, passenger detour distance and driver routing distance) and the key influence factors in ride-pooling service.

/02

Methodology and analysis

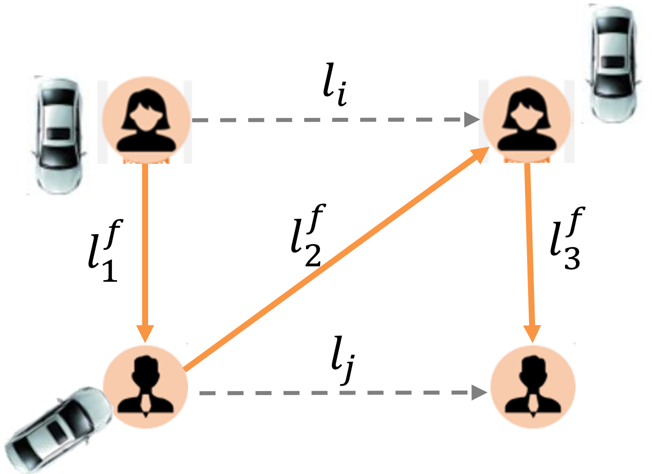
- Key indicators
- Data preparation
- Empirical law

Definitions of the key measures

- p : the pool-matching probability, the proportion of successful pool-matched passengers among all passengers
- l_i: Trip length of passenger i, without ride-pooling

For two trips that are paired for ride splitting, two different pick-up drop off sequence:

- First pickup first dropoff (FPFD)
- First pickup last dropoff (FPLD)

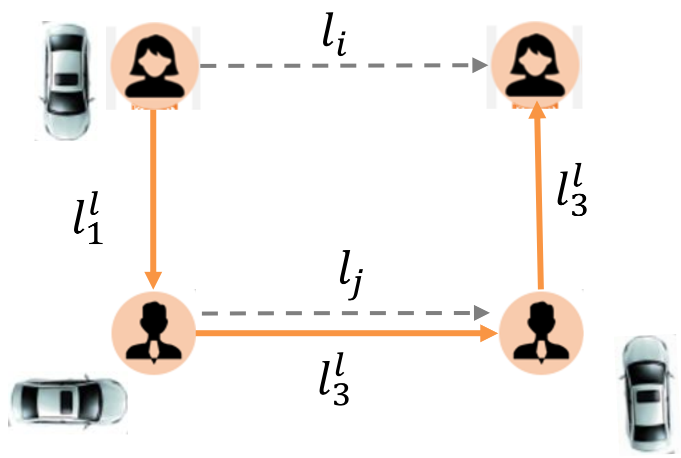


First pickup first dropoff (FPFD)

The detour distance experienced by passenger i and j are $l_1^f + l_2^f - l_i$ and $l_2^f + l_3^f - l_j$, respectively.

The routing distance of drivers for the two sequences are $l_1^f + l_2^f + l_3^f$, $l_1^l + l_2^l + l_3^l$, respectively

$l, \Delta l$, L: the average trip distance without ride-pooling service average detour distance and means of drivers routing distance



First pickup last dropoff (FPLD)

Matching algorithm

Suppose the matching pool has N accumulated passengers (i.e. passenger demand in each batch matching) with heterogeneous origins and destinations: (O_i,D_i), i = 1,2,⋯,N. For any two passengers, i, $j \in N$,there are four types of sequences to serve them:

$O_i \to O_j \to D_i \to D_j$ (FPFD), $O_i \to O_j \to D_j \to D_i$(FPLD) and
$O_j \to O_i \to D_j \to D_i$ (FPFD), $O_j \to O_i \to D_i \to D_j$ (FPLD).

Let $L_{\min}(i, j)$ denote the minimum driver routing distance of these four sequences.

$$(\text{P1}) \max_{x_{ij}} \sum_{i=1}^{N} \sum_{j=i}^{N} [V - L_{min}(i,j)] x_{ij}$$

$$\text{s.t.} \begin{cases} \sum_{i=k}^{N} x_{ki} + \sum_{j=1}^{k-1} x_{jk} \le 1, \forall k \epsilon \{2, \cdots, N\} \\ \sum_{i=1}^{N} x_{1i} \le 1 \end{cases}$$

$$\max\{l^o(i,j), l^d(i,j)\} x_{ij} \le R, \forall i,j \epsilon \{1,2,\cdots,N\}$$

Data preparation

City name	Trips	Time period	Area
Chengdu, China	595151	2016/11/6 00:00:00 ~2016/11/12 23:59:59	~85 km^2
Haikou, China	227883	2017/5/1 00:00:00 ~ 2017/5/7 23:59:59	~70 km^2
Manhattan NYC, US	2827464	2016/3/2 00:00:00 ~2016/3/8 23:59:59	~59.1km^2

Website of New York Taxi Dataset: https://data.cityofnewyork.us/Transportation/

Website of Didi Gaiya Open data: https://outreach.didichuxing.com/research/opendata/

Data preparation

Model setting

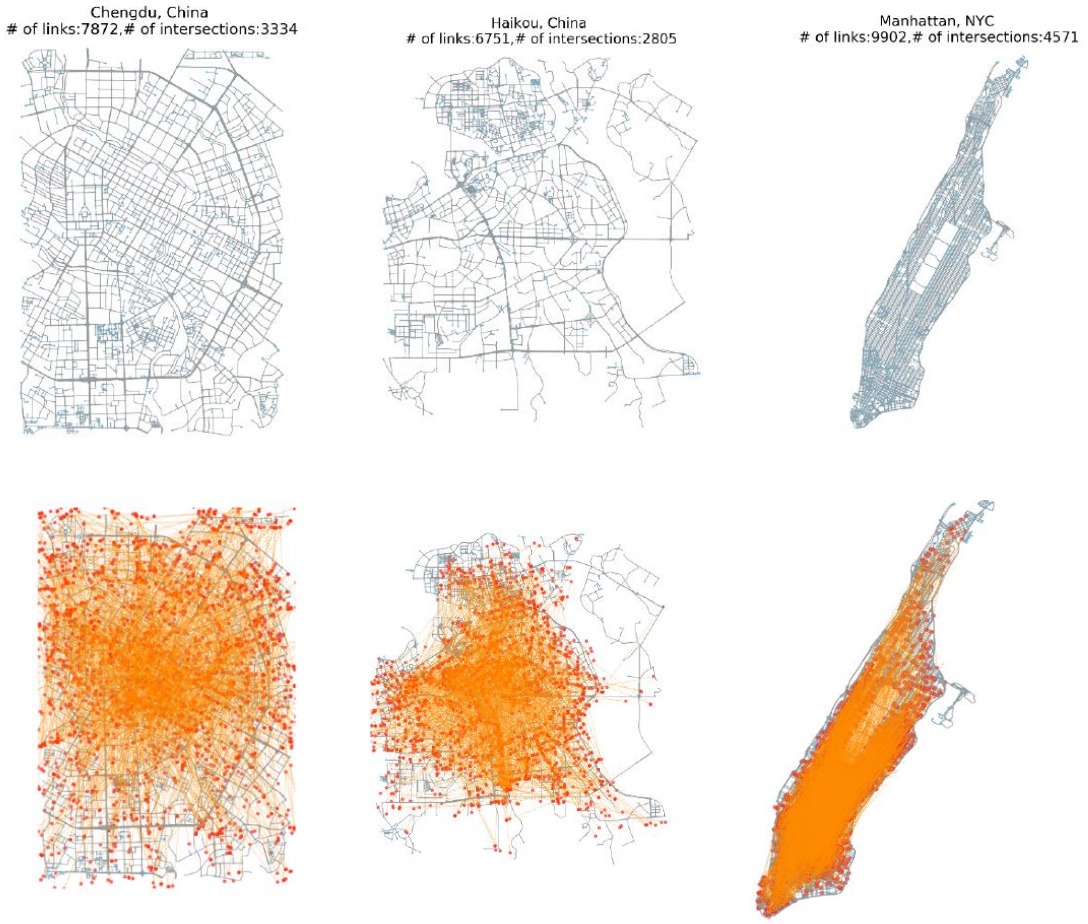
matching radii R, from 2km to 5km with a step of 1km

passengers in the pool N from 10 to 200 with a step of 10 in each experiment

randomly sample N passenger requests from historical trip records with origins and destinations

repeated by sufficient times (4,000 in this experiments)

Open street map: www.openstreetmap.org



(a) Chengdu (b) Haikou (c) Manhattan

Data preparation

Data filter

Trip length less than 0.1km or larger than 50 km

the trip duration time: shorter than 10s or longer than 2h

average speed of the trip is greater than 100 km/h

the origins and destinations of the trips are assigned to their nearest nodes in the network

(a) Chengdu (b) Haikou (c) Manhattan

Empirical law

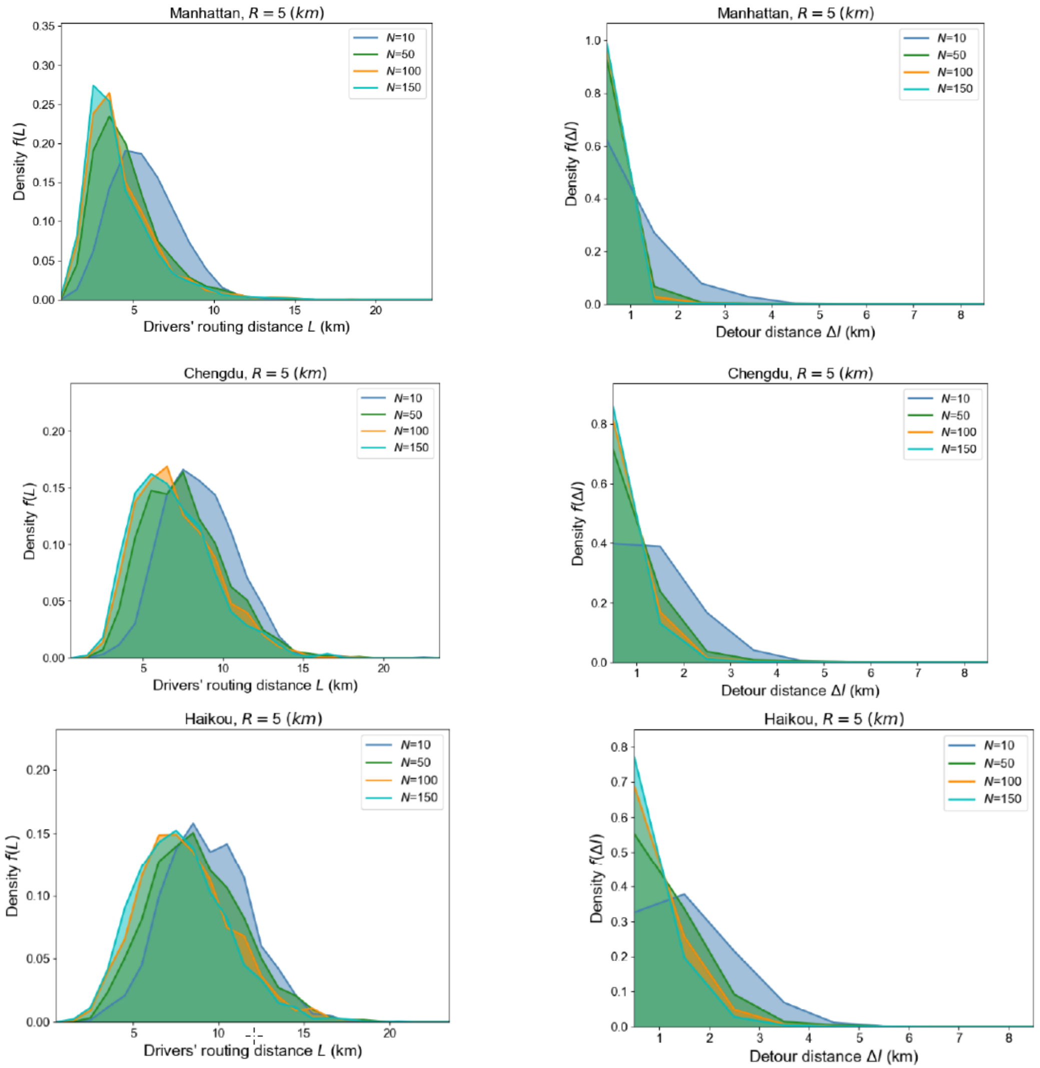
香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Law of passenger detour distance

Different city scale and trip information, exhibit some common characteristics

When demand is relatively large, detour distance, the distribution of detour distances decays as a power law.

The fundamental mechanisms behind this behavior has not yet been fully explained .



香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

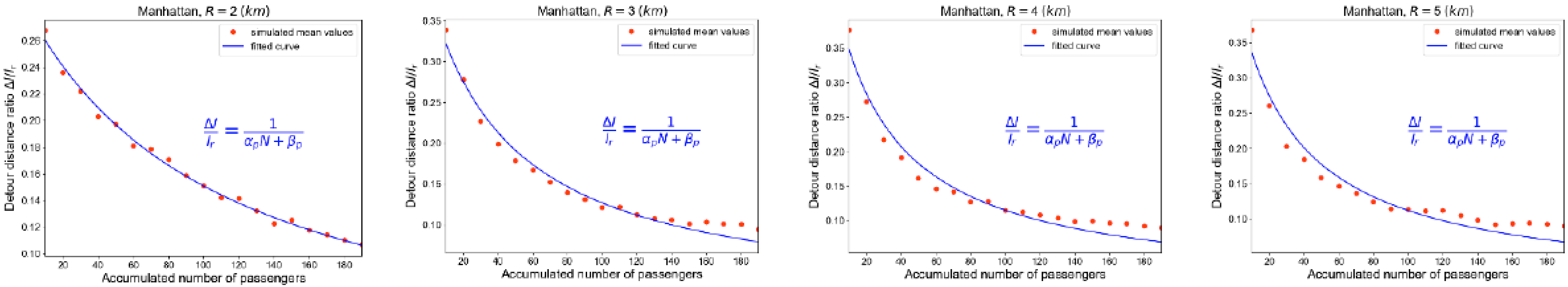
Empirical law

Law of passenger detour distance

It is interesting to find that the relationship between Δl and N can be well fitted by the following simple equation,

$$\frac{\Delta l}{l_r} = \frac{1}{\alpha_p N + \beta_p}$$

where l_r is the average non-shared trip distance, Δl is the average detour distance



Empirical law

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Law of passenger detour distance

Table 2(a). Empirical law of the average detour distance versus passenger demand (Chengdu)

Matching Radius (km)	α_p (hour/unit)	β_p	r^2
2	0.0089	5.2765	0.7986
3	0.0219	3.8355	0.9884
4	0.0336	3.3114	0.9559
5	0.0386	3.2476	0.9103

Table 2(b). Empirical law of the average detour distance versus passenger demand (Haikou)

Matching Radius (km)	α_p (hour/unit)	β_p	r^2
2	0.0059	6.6671	0.5609
3	0.0141	4.1537	0.9354
4	0.0243	3.4480	0.9932
5	0.0347	2.9205	0.9697

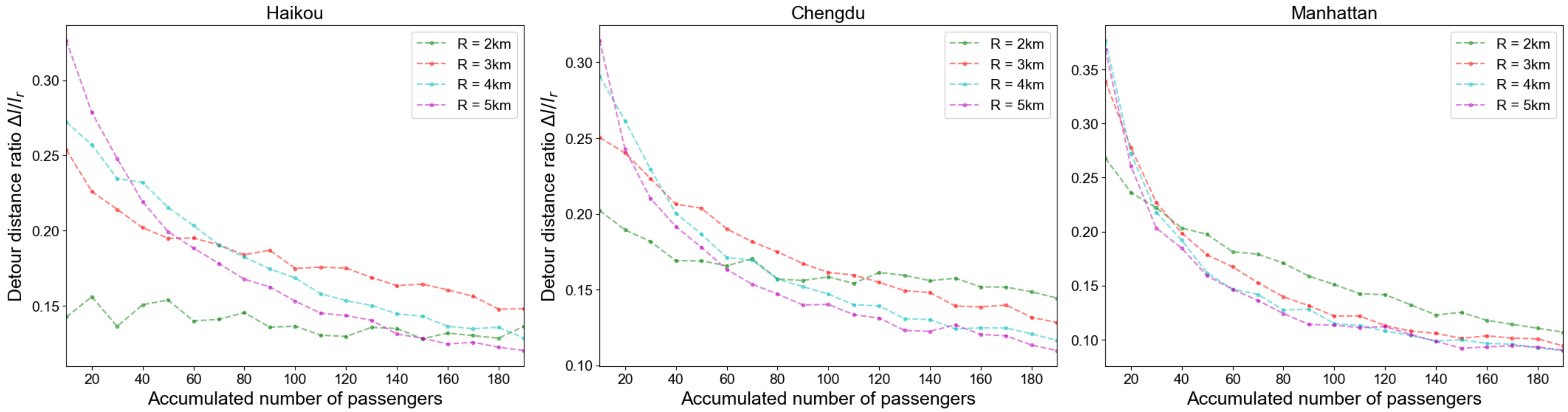
Table 2(c). Empirical law of the average detour distance versus passenger demand (Manhattan)

Matching Radius (km)	α_p (hour/unit)	β_p	r^2
2	0.0309	3.5343	0.9944
3	0.0531	2.5723	0.9744
4	0.0646	2.2271	0.9525
5	0.0661	2.3196	0.9395

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Empirical law

Law of passenger detour distance: unify the law by encapsulating the R



- When low demand, larger matching radius results in longer detour distance.
- When demand is high, larger matching radius will have the opposite result, lead to shorter detour distance.

more general form

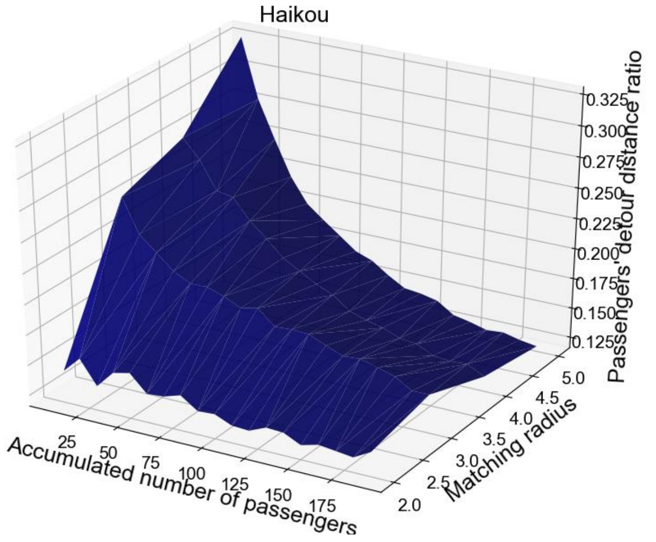
$$\frac{\Delta l}{l_r} = \frac{1}{(NR^2)^{\alpha_p} + \beta_p}$$

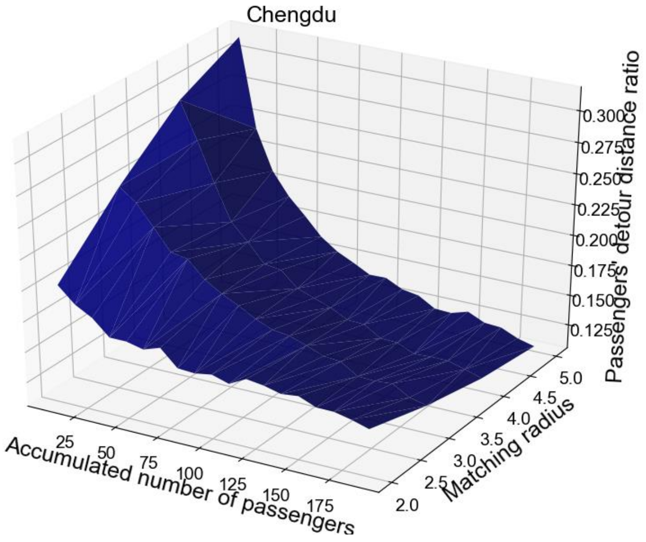
Empirical law

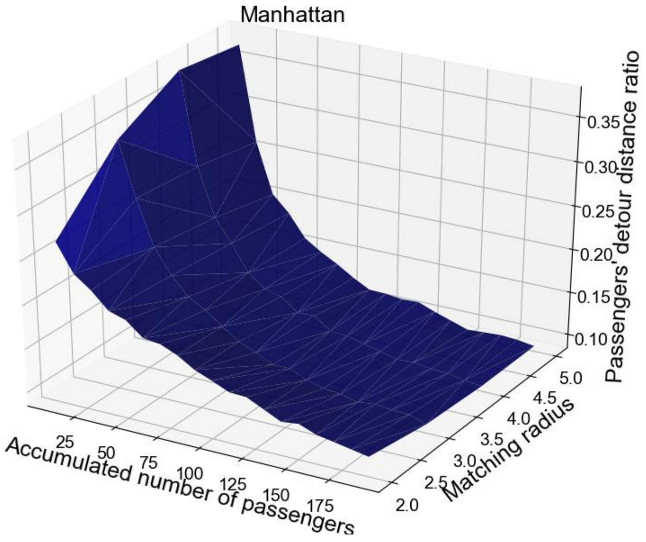
Law of passenger detour distance

$$\frac{\Delta l}{l_r} = \frac{1}{(NR^2)^{\alpha_p} + \beta_p}$$

City name	α_p	β_p	r^2
Haikou	0.0013	10.4842	0.7460
Chengdu	0.0020	9.1015	0.7053
Manhattan	0.0036	6.7223	0.6909





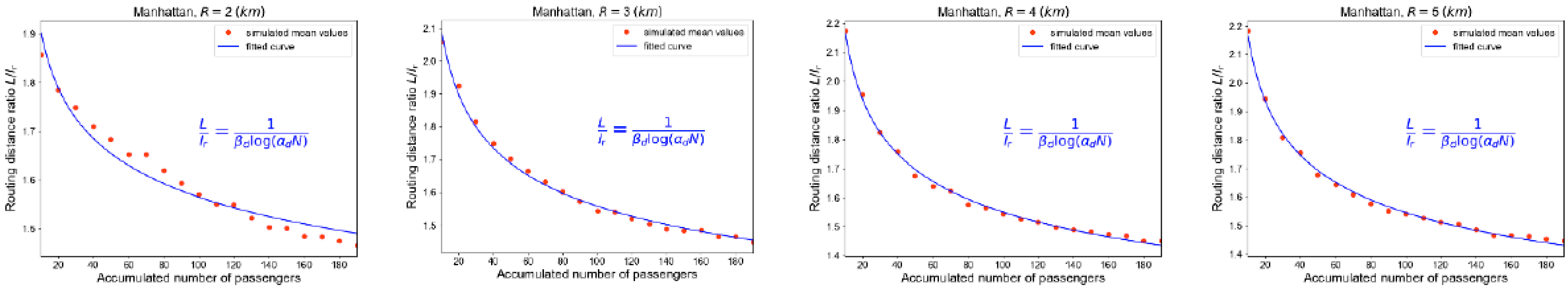


Empirical law

Law of average driver routing distance

We find that the following empirical law can well fit the relationship between L and N,

$$\frac{L}{l_r} = \frac{1}{\beta_d \log(\alpha_d N)}$$



(a) (b) (c) (d)

Empirical law

Law of average driver routing distance

Table 3a. Empirical law of the average routing distance versus passengers' arrival rate (Chengdu)

Matching Radius (km)	α_d (h)	β_d	r^2
2	5.0993E+12	0.0197	0.9524
3	2.0803E+04	0.0445	0.9671
4	530.2530	0.0596	0.9873
5	267.0359	0.0639	0.9985

Table 3b. Empirical law of the average routing distance versus passengers' arrival rate (Haikou)

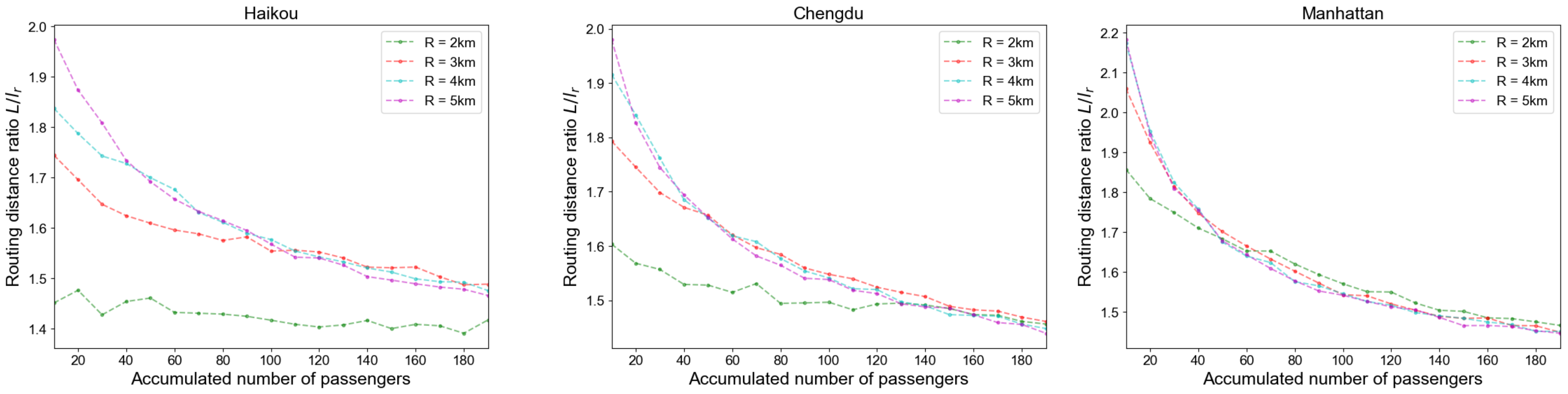
Matching Radius (km)	α_d (h)	β_d	r^2
2	2.3826E+25	0.0112	0.6838
3	5.1879E+06	0.0321	0.9770
4	7357.7279	0.0471	0.9465
5	319.0496	0.0615	0.9873

Table 3c. Empirical law of the average routing distance versus passengers' arrival rate (Manhattan)

Matching Radius (km)	α_d (h)	β_d	r^2
2	3817.0712	0.0497	0.9548
3	80.8712	0.0713	0.9933
4	31.1548	0.0802	0.9968
5	28.8668	0.0812	0.9972

Empirical law

Law of average driver routing distance



The drivers routing distance converges to a certain ratio of average trip distance

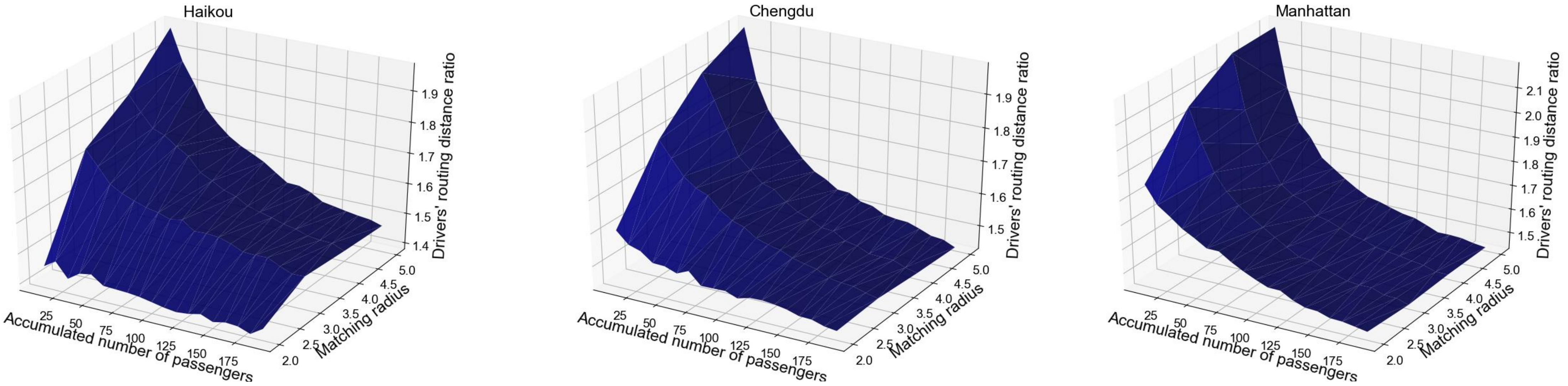
$$\frac{L}{l_r} = \frac{1}{\beta_d \log(\alpha_d \times \frac{N}{R})}$$

Empirical law

Law of average driver routing distance

$$\frac{L}{l_r} = \frac{1}{\beta_d \log(\alpha_d \times \frac{N}{R})}$$

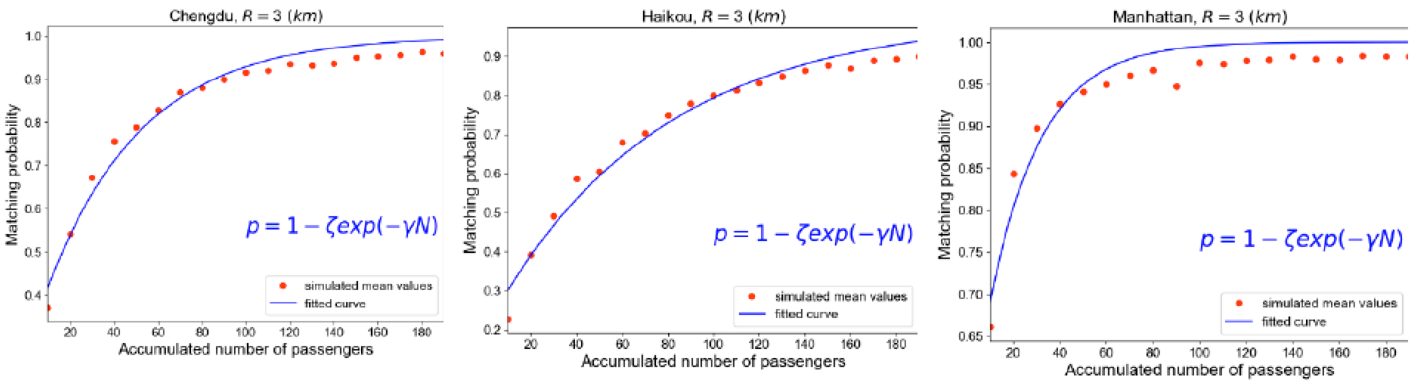
City name	α_d	β_d	r^2
Haikou	4170.7138	0.0471	0.6847
Chengdu	24917.5869	0.0487	0.6605
Manhattan	534.6560	0.0670	0.7140



Empirical law

Law of pool-matching probability

$$p = 1 - \zeta \exp(-\gamma N)$$



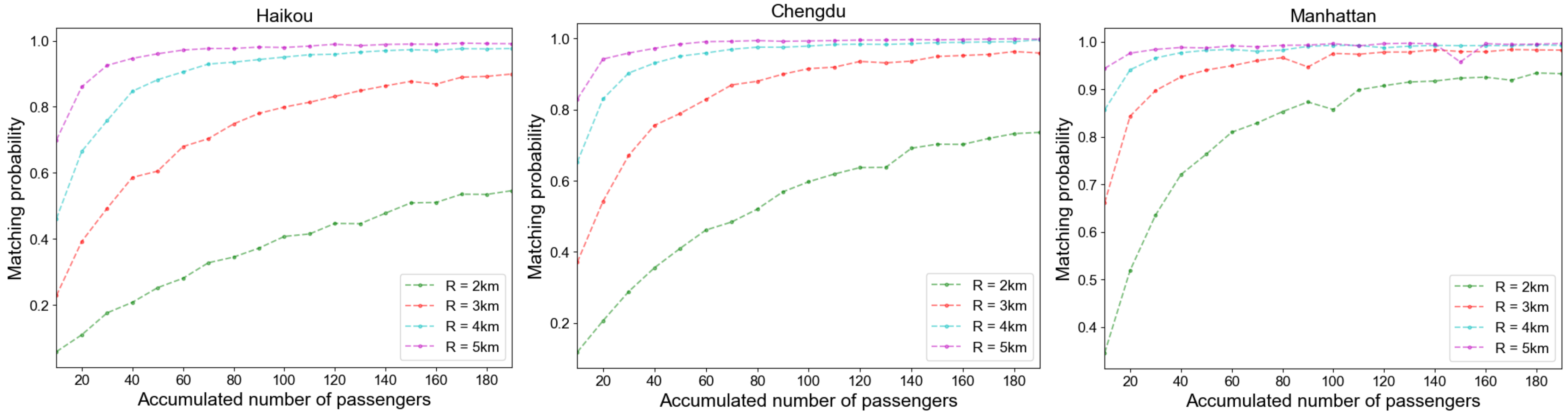
pool-matching probability with passenger demand looks like a saturation curve that first increases quickly and then slowly

Table 4 The fitting evaluation of matching rate under different arrival rate and matching radius

Matching Radius (km)	Chengdu			Haikou			Manhattan		
	γ	ζ	r^2	γ	ζ	r^2	γ	ζ	r^2
2	0.0074	0.8923	0.9800	0.0042	0.9429	0.9835	0.0189	0.7053	0.9451
3	0.0233	0.7360	0.9689	0.0135	0.8000	0.9730	0.0453	0.4873	0.9073
4	0.0537	0.5697	0.9645	0.0343	0.7166	0.9689	0.0624	0.2541	0.9057
5	0.0766	0.3555	0.9664	0.0600	0.5282	0.9611	0.0316	0.0620	0.8749

Empirical law

Law of matching probability

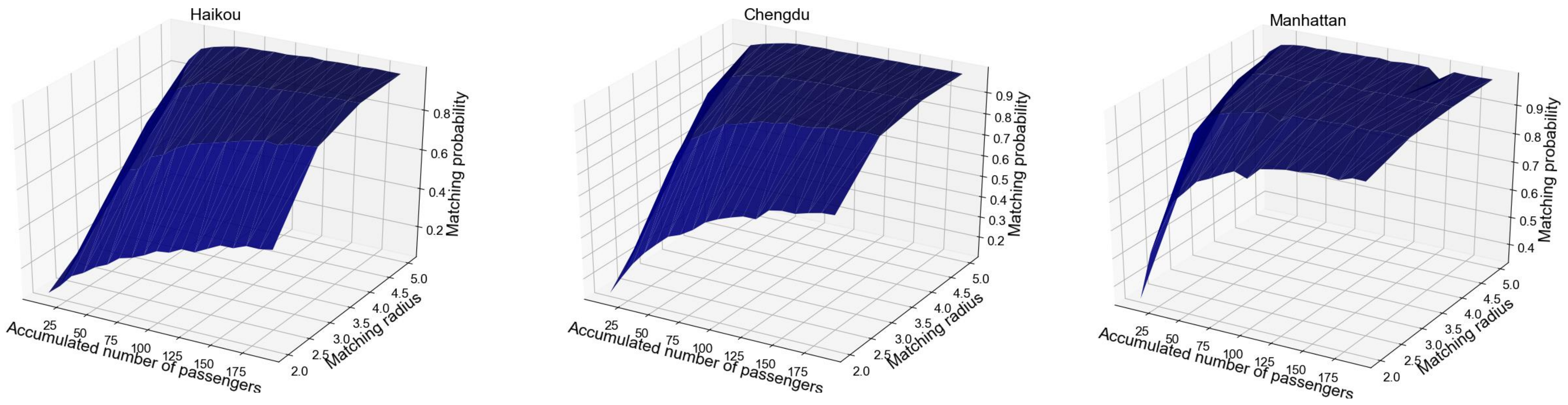


Empirical law

Law of matching probability

$$p = 1 - b \times exp(-N^a \times lnR), where\ R > 1km$$

City name	a	b	r^2
Haikou	0.2072	3.5202	0.9556
Chengdu	0.2504	3.5379	0.9735
Manhattan	0.3295	2.9648	0.9873



Different objective function

Try a different objective to minimize the passengers' detour distance

$$(\text{P1}) \max_{x_{ij}} \sum_{i=1}^{N}\sum_{j=i}^{N}[V - L_{min}(i,j)]x_{ij}$$

$$\text{s.t.} \begin{cases} \sum_{i=k}^{N} x_{ki} + \sum_{j=1}^{k-1} x_{jk} \leq 1, \forall k \epsilon \{2, \cdots, N\} \\ \sum_{i=1}^{N} x_{1i} \leq 1 \end{cases}$$

$$\max\{l^o(i,j), l^d(i,j)\}\, x_{ij} \leq R, \forall i,j \epsilon \{1,2,\cdots,N\}$$

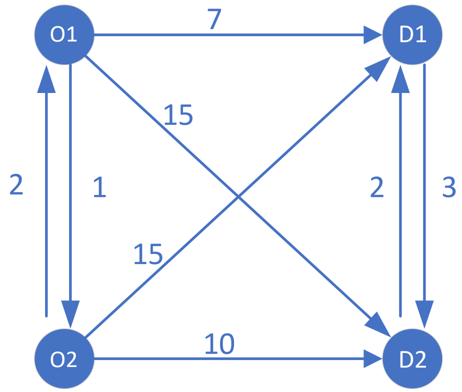
$$(\text{P2}) \max_{\mathrm{x_{ij}}} \sum_{i=1}^{N}\sum_{j=i}^{N}[V - \Delta l_{min}(i,j)]\, x_{ij}$$

Constrains keep the same

Distance-based matching strategy

Different objective function

A simple case to distinguish the difference between the two objective functions

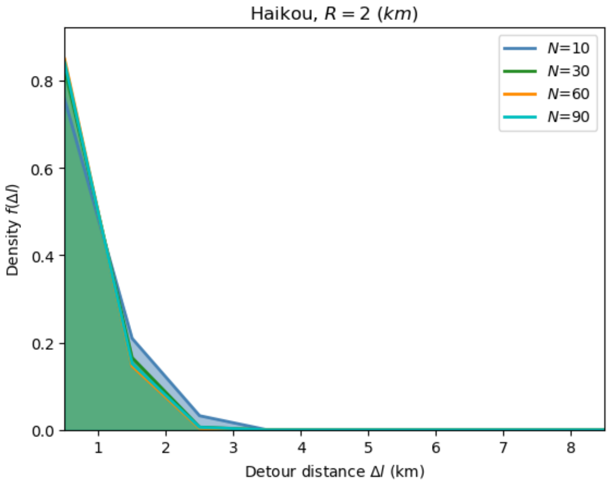


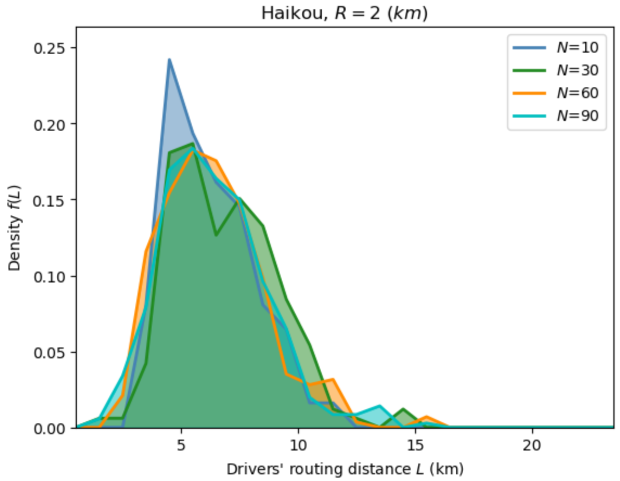
Pick-up sequence	Driver's routing distance	Detour distance	
$O_1 \to O_2 \to D_2 \to D_1$	13	**3**	Optimal for detour distance
$O_1 \to O_2 \to D_1 \to D_2$	19	9+8	
$O_2 \to O_1 \to D_2 \to D_1$	19	7+10	
$O_2 \to O_1 \to D_1 \to D_2$	**12**	5	Optimal for driver's routing distance

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

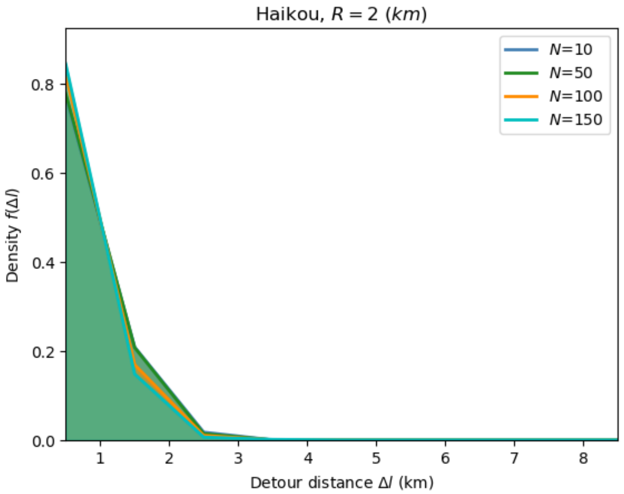
Comparison between two

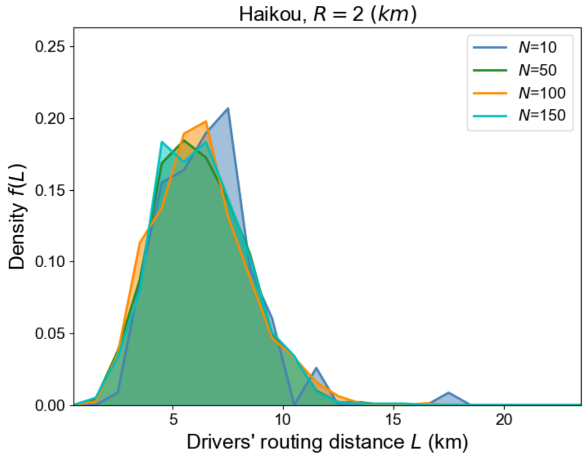
Minimize detour distance



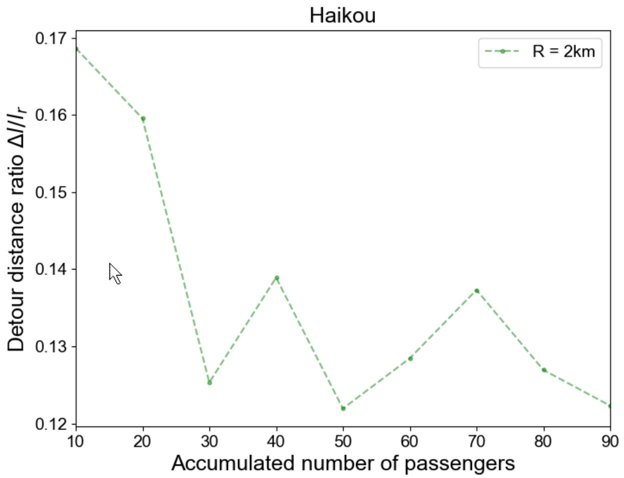


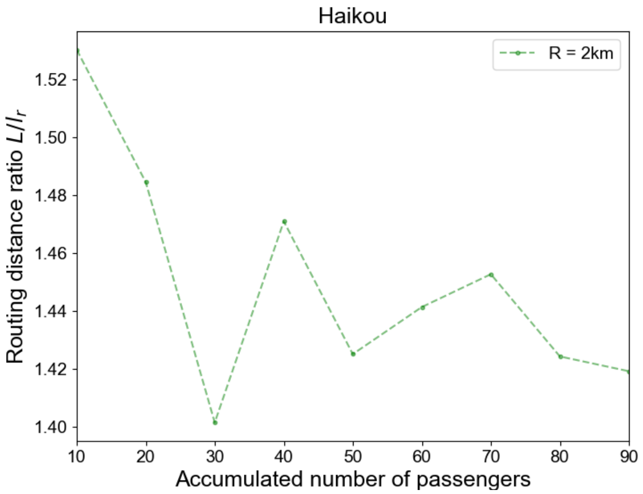
Minimize driver's routing distance

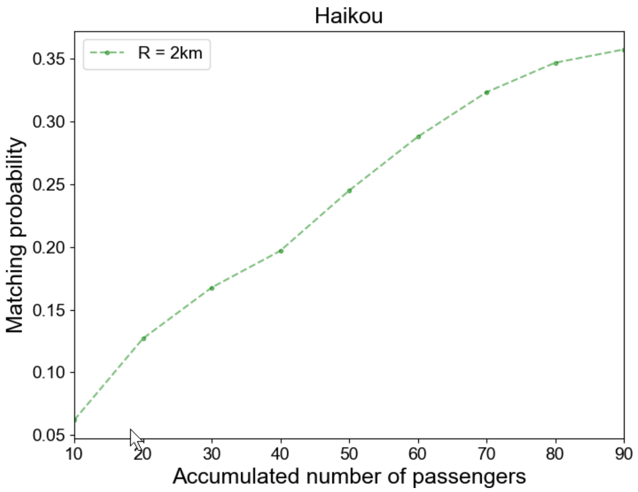




Comparison between two







Haikou	α_p	β_p	r^2	α_d	β_d	r^2	a	b	r^2
Minimize driver's distance	0.0013	10.4842	0.9460	41538.7138	0.0471	0.6847	0.2072	3.5202	0.9556
Minimize detour distance	0.0024	13.0031	0.9341	52314.0124	0.0612	0.7142	0.1505	2.5744	0.9577

Similar results can be obtained in Chengdu, Manhattan.

/03

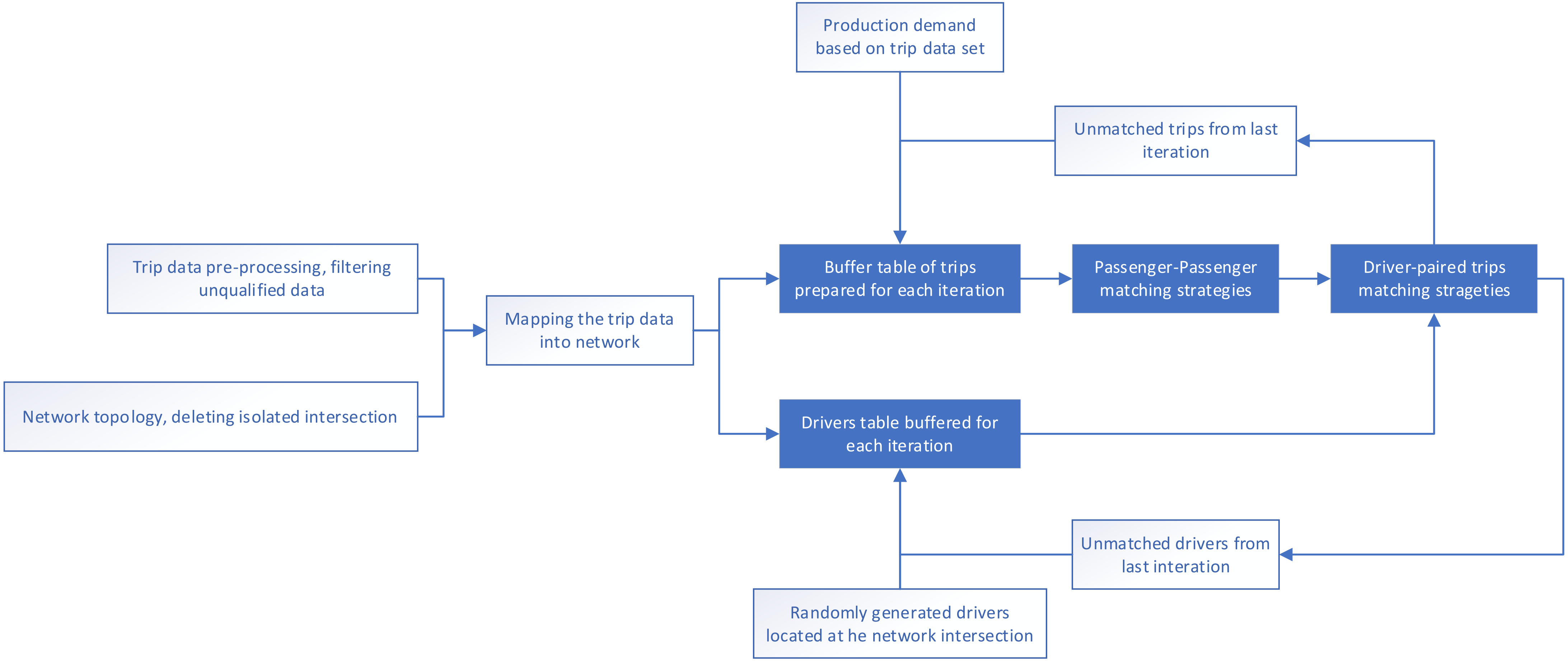
Future directions

Future research

A more thorough investigation of why these forms are valid is expected

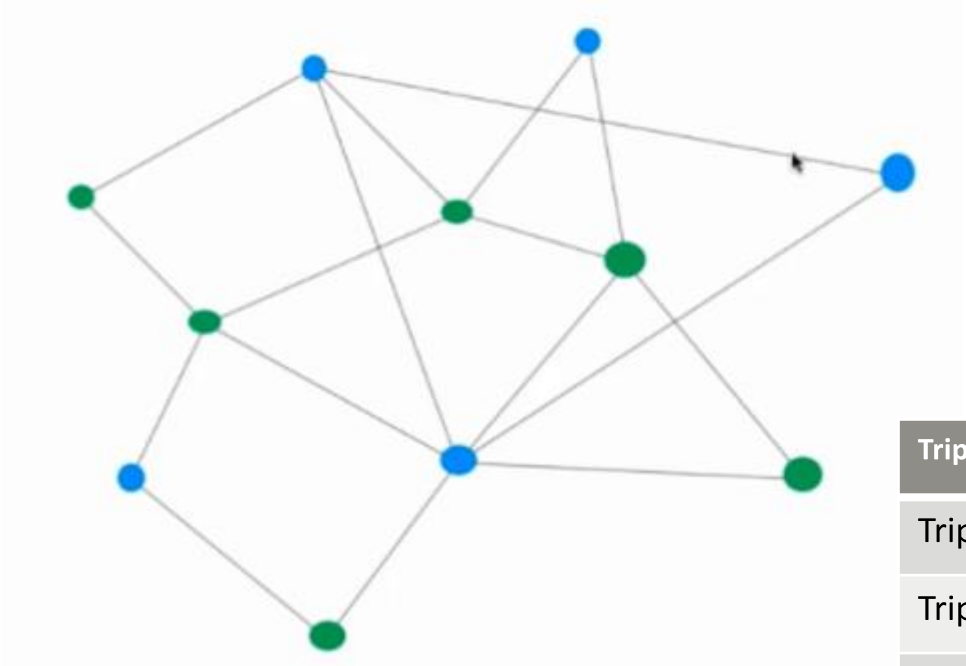
- it is possible to proceed with this formulation is now analyzing mathematically instead of empirically the relationship between the optimal values and the demand.
- Another more reasonable way to explicitly explain why the proposed empirical law works is to directly build statistical/econometric models. Like the scaling law of human mobility.

Dynamic simulator



Dynamic simulator

Directed weighted graph



Driver ID	location
Driver 1	S_1
Driver 2	S_2
…	…
Driver i	S_i
…	…
Driver m	S_m

Pick-up distance

Driver-paired trips matching strageties

Trip ID	location	Destination
Trip 1	O_1	D_1
Trip 2	O_2	D_2
…	…	
Trip i	O_i	D_i
…	…	
Trip n	O_j	D_j

Passenger-Passenger matching strategies

matching distance

Matched Trip ID	Pick-up and drop-off sequence
Trip 1 & 2	$O_1 \rightarrow O_2 \rightarrow D_2 \rightarrow D_1$
Trip i & n	$O_n \rightarrow O_i \rightarrow D_n \rightarrow D_i$
…	…
…	…

Future research

The supply side is neglected for simplification in current data-driven analysis. Moreover, one open question in the field of ride-sourcing services is the impacts of ride-pooling on transit usage, private car ownership and traffic congestions.

By comparing the traffic flow data before and after the activation of the ride-pooling program across different cities, the proposed project will empirically evaluate the actual influences the ride-pooling programs brings to the urban traffic.

References

Research output

Zheng, Z., Ke, J*., Yang, H. and Ye, J., (2020) Data-driven analysis of matching probability, routing distance and detour distance in on-demand ride-pooling. Transportation Research Part C. (under review)

Ke, J., Qin, X*., **Zheng, Z.**, Yang, H., Ye, J., (2020) Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network. Transportation Research Part C. (under review)

Ke, J., Wang, H*., **Zheng, Z.**, (2020). Optimal demand allocation in ride-sourcing markets with a bundled service option. Submitted to ISTTT 24. (Acceptance of abstract)

Key references:

Brockmann, D., Hufnagel, L., Geisel, T., 2006. The scaling laws of human travel. Nature 439, 462–465. https://doi.org/10.1038/nature04292

Chen, X. (Michael), Zahiri, M., Zhang, S., 2017. Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach. Transportation Research Part C: Emerging Technologies 76, 51–70. https://doi.org/10.1016/j.trc.2016.12.018

Ke, J., Yang, H., Li, X., Wang, H., Ye, J., 2019. Pricing and equilibrium in on-demand ride-splitting markets. Available on SSRN. https://doi.org/abstract_id=3357362

Li, W., Pu, Z., Li, Y., (Jeff) Ban, X., 2019. Characterization of ridesplitting based on observed data: A case study of Chengdu, China. Transportation Research Part C: Emerging Technologies 100, 330–353. https://doi.org/10.1016/j.trc.2019.01.030

Yang, H., Qin, X., Ke, J., Ye, J., 2019. Optimizing matching time interval and matching radius in on-demand matching of a ride-sourcing market. SSRN Electronic Journal 1–40. https://doi.org/10.2139/ssrn.3372349

Yang, H., Yang, T., 2011. Equilibrium properties of taxi markets with search frictions. Transportation Research Part B: Methodological 45, 696–713. https://doi.org/10.1016/j.trb.2011.01.002

Thanks

Zhengfei ZHENG, PhD qualifying examination, Civil Dept. HKUST