International Journal of Applied Earth Observations and Geoinformation 106 (2022) 102666

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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

Applying Ollivier-Ricci curvature to indicate the mismatch of travel demand and supply in urban transit network

Yaoli Wang [a,b], Zhou Huang [a,b,e,*], Ganmin Yin [a,b], Haifeng Li [c], Liu Yang [a,b,d], Yuelong Su [d], Yu Liu [a,b], Xv Shan [e]

[a] Institute of Remote Sensing and Geographic Information Systems, Peking University, Beijing, China
[b] Beijing Key Lab of Spatial Information Integration & Its Applications, Peking University, Beijing, China
[c] Department of Geographic Information System, Central South University, Changsha, Hunan, China
[d] Amap, Beijing, China
[e] State Key Laboratory of Media Convergence Production Technology and Systems, Beijing, China

ARTICLE INFO

Keywords:
Ricci curvature
Transit network redesign
Spatial big data
Demand and supply
Transit travel behaviour

ABSTRACT

Urban transit networks need to be upgraded in accordance with urban development. While methods have been studied to design an optimal transit network given the locations of stations, these methods focus on the whole network as the optimization object. However, the strategy to improve parts of an existing transit network based on the gap between travel demand and infrastructure supply is rarely investigated. Different parts of the same network should not have the same priority to be upgraded. An evaluation method is required for this purpose to decide which parts of a transit network should be upgraded first. We argue that Ollivier-Ricci curvature, a concept from differential geometry, can serve as the evaluation method. The basic hypothesis is that Ollivier-Ricci curvature indicates the theoretical attractiveness of the fastest path between any two nodes, thereby capturing the internal structure of the transit network. Meanwhile, travel demand and behavior can be approximated by urban mobility big data. By comparing the theoretical attractiveness and the flow volume between a path between two stations, the traits of lines with mismatched supply and demand are identified, and macro strategies are suggested. A case study in Beijing demonstrates how the proposed method works out followed by a comparison with traditional betweenness centrality analysis.

1. Introduction

As urbanization is speeding up, the demand for efficient public transit services is unprecedented. Travel experience of public transit is an interplay between the supply of infrastructure and the demand of travel origins/destinations. Therefore, public transit network should be restructured or upgraded throughout time to catch up with the changes of other factors in urban planning, e.g., urban renewal, urban development, and population distribution. Design and optimization of public transit system thus is an important topic.

Transit network optimization (TNO) is not a brand new topic. Most TNO models focus on designing a network by optimizing an objective function considering the economic cost and travel cost. van Nes et al. (1988) classified TNO models into six categories depending on whether the model determines routes, frequencies, or both. Although frequency is an important factor of public transit, it falls out of the scope of this work. The following will focus on transit network routes.

A fundamental question is to decide which road links should be applied to construct routes for public transit network. Usually a model is built to optimize the objective cost function with pre-specified stations and constraints of travel time and monetary costs. It is typically a mathematical programming problem (Yao et al., 2014; Gutiérrez-Jarpa et al., 2017). In practice, nevertheless, solving such an optimization problem is no easy job.

Billheimer and Gray (1973) had argued for the trade-off between theoretical rigour and practical applicability, and the necessity for heuristics in optimizing public transit networks. A few heuristics have been developed over time. For example, genetic algorithm and ant colony optimization have been applied for feeder bus network design at real world scale (Kuan et al., 2006). Yu et al. (2012) designed a method to maximize the coverage of not only direct but also transfer demand with ant colony algorithm. Cipriani et al. (2012) solved the transit

* Corresponding author at: Institute of Remote Sensing and Geographic Information Systems, Peking University, Beijing, China.
E-mail address: huangzhou@pku.edu.cn (Z. Huang).

https://doi.org/10.1016/j.jag.2021.102666
Received 31 May 2021; Received in revised form 21 December 2021; Accepted 22 December 2021
Available online 30 December 2021

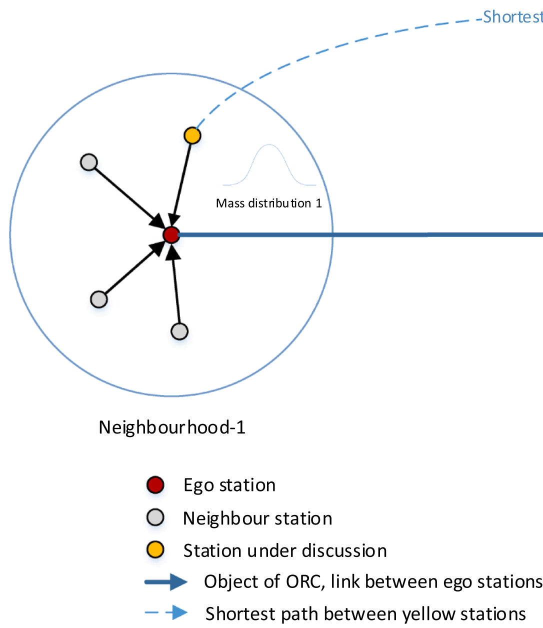


Fig. 1. Illustration of Ollivier-Ricci curvature.

network design problem in a large urban area with parallel genetic algorithm. Buba and Lee (2018) used evolutional algorithm to optimize the transit lines to minimize travel costs and unmet demands. Islam et al. (2019) proposed a stochastic beam search algorithm that can produce good transit route design solutions. Jha et al. (2019) applied particle swarm optimization to get the Pareto-Front solutions in terms of travel time and operating costs. The study of Barahimi et al. (2021) was similar. Capali and Ceylan (2020) also studied the transit network design and frequency setting problem through a novel intelligent water drops algorithm. The heuristics, nevertheless, are defined in mathematical domain, which is different from heuristics in application domain. The former aims more to build up an ideal mathematical model, without empirical inputs such as real travel demand implied by smart card data, or realistic transit network structure. The latter is more strategic and instructive in practical work, taking advantage of analyzing the gap between travel demand and supply to identify the routes that should be urgently upgraded.

Designing transit networks, despite of various methods, generally lacks mature and well-defined methodologies. Traffic flow is a comprehensive result of underlying infrastructure and human decision-making on top of that. It is drawn as a designer and follower problem (Farahani et al., 2013), of which the former refers to a top-down decision and design, while the latter is an emergence from user behaviour as bottom-up reaction. There might be a theoretically optimal shape of transportation network (Aldous and Barthelemy, 2019), and indeed transit network structure has been seen to boost the incentive to take public transit (Badia et al., 2017), but top-down design alone is less than sufficient in reality due to unknown reactions from the general public. The bottom-up method complements the top-down design with data-driven validation. Therefore, we need to understand the interaction between infrastructure supply – a result of designer – and travel demand – how followers react. In this sense, the aforementioned methods all fall into the category of designer problem.

The present paper, in contrast, sets out from a new perspective of both designer and follower to provide strategic heuristics for TNO, i.e., a evaluation measure to decide which linkages of the network should be upgraded. This makes a demand and supply view more closely connected since the follower side reflects the reaction to infrastructure supply based on demand. On one hand, design of a transit network decides its structure and consequently its function; on the other, followers' reaction to infrastructure affects how traffic flow distributes on a given network. Note that the study addresses improvement on the infrastructure system with regard to *current* travel demand. Projected future travel demand or induced travel demand with infrastructure update require additional optimization and simulation work (Manser et al., 2020; Wang et al., 2021), which falls out of the present work's scope. The subject of designer and follower analysis on public transit network should be transit *lines*. Analysis on areal travel demand and supply (Chen et al., 2018; Farber et al., 2016) is not enough since it does not disclose the flow distribution. Hypothetically, flow distribution partially is a consequence of the innate property of the network space, which means an appropriate network structure is likely to induce balanced traffic distribution (Badia et al., 2017). Though a computable measure was invented to assess transit network efficiency by considering each stop's structural connectivity with other stops (Wang et al., 2020), travel demand was not yet integrated. A new measure hence is required to incorporate both innate infrastructure property and travellers' selection. The proposed measure leverages the two dimensions – network structure and travellers' reaction – to interpret public transit flow distribution on transit lines. The understanding contributes to a strategic heuristic for public transit network optimization.

In fact, for a relatively mature city, designing public transit network from scratch is unnecessary. More likely only parts of a network have high congestion that should be relieved. The strategic heuristics draw attention to *part* of the lines that needs urgent improvement instead of the whole transit network. With a similar goal, some practice has been done to upgrade an existing public transit network, e.g., the replanning of bus routes in a medium-size town in France (Dubois et al., 1979). However, their solution that applies travel survey to detect travel demand is ostensibly out of date in information age. To a much larger scale contemporary metropolis like Beijing, Tokyo, New York City, and London, the solution should be data-driven. Heyken Soares et al. (2019) brought up the issue of scaling down the street network so to decide the pool of stations based on which line optimization is conducted. Though their work involves existing transit lines into new design, they do not discuss the designer and follower interplay.

This work proposes a strategic evaluation method based on a concept of differential geometry called *Ollivier-Ricci curvature*. Intuitively Ricci curvature tells the distribution of flows in a local neighbourhood affected by the structure of a space. Ollivier-Ricci curvature reflects the intrinsic geometric structure of a (discrete) network space, and quantifies flow distribution in the space given a certain travel demand. An adapted curvature for public transit network measures human reaction to a designed system. The analogy to a landscape with rainfall is intuitive. Travel demand is analogous to the total amount of precipitation, while the landscape's geometry (just as the topological and geometrical structure of a network) decides how water flows on it. The measure hence corresponds well with the aforementioned designer and follower perspective. Ollivier-Ricci curvature has been applied in diverse topics, including community detection (Ni et al., 2019; Sia et al., 2019) and network's vulnerability analysis (Gao et al., 2019). Section 2 will dive into the details of Ollivier-Ricci curvature and Section 3 elaborates the adapted measure for public transit network. To our knowledge, no work has been done that deploys Ollivier-Ricci curvature for public transit network optimization. The method therefore is an innovative trial.

Instead of focusing on optimization, this work aims to create an innovative pre-assessment method as strategic heuristic for TNO. The method identifies line segments that require upgrade by considering Ollivier-Ricci curvature and observed traffic volume on each route. The method then suggests general strategies of network redesign.

We leverage an ample source of data that provides insights into when and where people travel from and to, as well as the associated travel modes and transfer behaviours. Human mobility footprints not only directly indicates human decisions but also indirectly infers the performance of underlying infrastructure (Liu et al., 2015). In this study we utilize cell phone call records and smart card data for human mobility, and public transit network data for infrastructure information. Details are given in Section 4.

The proposed method is testified with a case study in Beijing. The area within the 5*th* Ring Road is fetched to redesign its public transit lines. We then extract the line segments according to both flow volume and ORC. Different scenarios are associated with local background information to validate the measure. Section 4 elaborates the implementation as well as the experiment setup, followed by Section 5 of result analysis and discussions, and Section 6 of conclusions.

2. Preliminary theories: Ollivier-Ricci curvature

Ricci curvature is a concept in differential geometry that describes the shape for an infinitely small locality on the surface of a manifold along a direction (geodesic). Positive curvature describes a surface where the centres of two small balls are farther than the average distance between points in the two balls, like the surface of a sphere. While negative curvature depicts where the centres are closer than the balls like a saddle. On a surface with positive curvature, the flow diffuses towards different directions; while on a negative curvature, the flow converges to the geodesic.

Ollivier Ricci curvature (ORC) connects the Ricci curvature with optimal transport theories. The gist is to quantify a network's optimal transport of flows by defining the network's curvature. ORC is defined on discrete data such as a network. Fig. 1 demonstrates the concept. The subject of ORC is the shortest path between any two nodes, i.e., a geodesic as the bold navy blue arrow shows. ORC reflects whether the geodesic between two certain nodes x and y (hereafter called "ego nodes" or "ego stations") attracts or repels flows from the nodes' local neighbourhoods, i.e., nodes within the circles. ORC intuitively depicts the relative convenience of moving some mass that follows specified distributions m_x in the neighbourhood of x to another distribution m_y in the neighbourhood of y when traversing the geodesic. If local flows tend to traverse the geodesic, ORC will be negative so the route is more attractive. To sum up, there are three key factors of ORC: the distribution of mass to be transferred – no matter earth, water, or urban traffic, the definition of neighbourhood that specifies which nodes are connected, and the cost of path.

In strict mathematical form, ORC between ego nodes x and y is defined as Eq. 1.

$$k_{xy} = 1 - \frac{W(m_x, m_y)}{d(x,y)} \tag{1}$$

$W(m_x, m_y)$ is *Wasserstein distance* in optimal transport that dates back to the Monge's mass transfer problem (Evans, 1997). The measure quantifies the minimal cost of moving a given distribution of mass m_x from one region to the specified distribution of mass m_y in another region (which action is called *transference*) over all possible transference plans ξ. The slack problem is given as a coupling Π of m_x and m_y in Eq. 2, where $d(x,y)$ is the least travel cost between two nodes (Ollivier, 2010).

$$W\left(m_x, m_y\right) = \inf_{\xi \in \Pi\left(m_x, m_y\right)} \int\int d\left(x, y\right) \mathrm{d}\xi\left(x, y\right) \tag{2}$$

ORC is a comparison between Wasserstein distance $W(m_x, m_y)$ of two regions and the travel cost between the centres $d(x,y)$. Intuitively the equation tells whether traversing the route between x and y saves cost when moving the mass from m_x to m_y; if the route is economical, flows tend to converge to it. The assessment of route in transit network optimization shares a similar concept as ORC: negative ORC paths indicate the paths attract flows from its neighbourhood because of its lower cost compared with its alternative paths; high ORC paths are the other way around, pushing the flows to alternative paths. *Betweenness* is alike in terms of representing the attractiveness of a route for traffic flow to traverse it. However, ORC captures the comparative advantage within a local range, and is more focused on a locality. ORC also associates the structure of the network with its performance at a basic scale – the 1-hop neighbourhood at both ends of a route.

3. Methodology

As discussed above, transit network optimization can be inspired from ORC. However, some adaptations on ORC should first be conducted. This section will introduce the adjusted ORC for transit network, and elaborate the workflow of our methodology.

3.1. Definition of Ollivier-Ricci curvature for transit network optimization

Investigate two public transit stops. Since the stops are connected with other stops in the network, traffic flows may converge from or diffuse to other stops. The set of connected stops around a certain stop form the stop's *neighbourhood*. There can be multiple ways to transport travel demand from one neighbourhood to another neighbourhood. The specific way to move travel demand is called *transference plan*. In public transit network, transference plan specifies how many travel flows are moved from any point of a station's neighbourhood to any point of the other neighbourhood. Mathematically this is a coupling of two demand distributions. ORC quantifies, given the innate structure of the network and a certain coupled travel demand distribution, whether flows around the two stops will be attracted to or repelled from the route between the two stops. If we assume all travellers are rational agents, ORC can indicate whether people will choose the route.

ORC in this study is calculated on a multilayer public transit network M composed of multi-modal transit lines. Three modes are particularly considered: subway, bus, and walk. ORC is defined on M with regard to every each pair of stations s_i and s_j in P-space. Namely, ORC is calculated between two stations as long as they are connected with a path on M. The *neighbourhood* of each ego station δ_{s_i} or δ_{s_j} is a set of its 1-hop connected nodes which may belong to different modes. Note that this study will only discuss neighbourhoods in pairs, since a route links two stops. Ostensibly, three types of neighbourhood nodes exist in accordance with the three considered modes: subway station, bus stop, and the origin/destination (OD) of travels. The study scope is limited to these three modes in the present work. However, we acknowledge that other modes, such as taxis, transportation network company services, can be

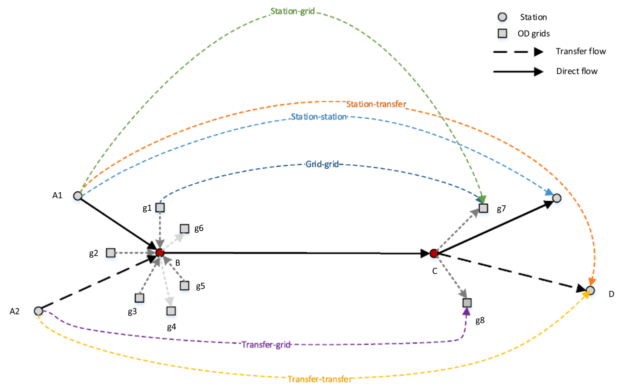


Fig. 2. Illustration of station neighbourhood flows.

incorporated and thus affect the computed result in that new system.

In its definition of Eq. 2, the masses m_x and m_y are the travel demand distributions in space. As specified in Section 2, *Optimal Transport* theory does not specify transference plan but finds the least cost plan among all possible plans $\Pi(m_x, m_y)$. The coupling between m_x and m_y is not pre-specified. In real transit network, however, travel demand distributions are known ahead of time, so the transference plan (i.e., coupling of m_x and m_y) in urban transit network is determined according to real travel demands. The reason for the difference is that mass in Optimal Transport is homogeneous, but travel demand in urban life is heterogeneous. The special trait of human mobility is not accommodated by previous studies or applications of ORC (Gao et al., 2019; Ni et al., 2019; Sia et al., 2019).

In this study, we take into account the heterogeneity of flows, and hence need to distinguish flows between different types of nodes. We investigate six types of flows based on its types of nodes at both neighbourhoods (Fig. 2):

- grid-grid: the flow starts and ends both at OD grids
- station-grid or grid-station: the flow starts or ends at a station on the same line, and the other end is an OD grid
- transfer-grid or grid-transfer: the flow starts or ends at a station on a different line then transfers to the present, and the other end is an OD grid
- station-station: the flow starts and ends both at stations on the same line
- station-transfer or transfer-station: the flow starts or ends at a station on the same line, while the other end is a station of another line
- transfer-transfer: the flow starts and ends both at stations of other lines

Considering the heterogeneity of traffic flows, ORC of a line segment between a station pair is calculated as Eq. 3, where p and q are stations belonging to each neighbourhood of ego stations s_i and s_j, f_{pq} is the flow volume from station p to station q, and d_{pq} is the least travel time cost between the two stations. The cost can be replaced with travel distance and other types of cost as needed. The Wasserstein distance is adjusted from its original form since the transference plan is given. We use the notation $\widetilde{W}(\cdot)$ to distinguish it from Eq. 2. Different types of flows have different calculation technical details, which will be elaborated in SubSection 3.2.

$$k\left(s_i, s_j\right) = 1 - \frac{\widetilde{W}\left(s_i, s_j\right)}{d\left(s_i, s_j\right)} = 1 - \frac{\sum_{p,q \in \delta_{s_i}, \delta_{s_j}} f_{pq} d_{pq}}{d\left(s_i, s_j\right) \cdot \sum_{p,q \in \delta_{s_i}, \delta_{s_j}} f_{pq}} \tag{3}$$

Intuitively, ORC indicates whether travel demands between the regions around two stops will intend to traverse the route between these ego stops. Wasserstein distance is the average cost of moving travel demand from one neighbourhood to another following specified spatial distributions; and $d(s_i, s_j)$ is the cost of moving an equivalent total amount of travel demand from one corresponding ego stop to the other. ORC is a comparison between the two distances. For Wasserstein distance, each specific pair of stops between two neighbourhoods traverse their shortest cost route, which may or may not traverse the ego stops. Meanwhile, for $d(s_i, s_j)$, travel demands are assumed to gather at ego stops and surely traverse the route between them. Iterate the pairs (p, q). If most of the least cost paths go through s_i and s_j, the cost between neighbourhoods is surely longer than the cost between ego stations, which will yield a negative ORC. Positive ORC appears when the average cost between neighbourhoods is lower than that between ego stations. Note that the average cost between neighbourhoods is highly influenced by travel demand distribution since the cost between different particular points is heterogeneous due to network structure. ORC thus depicts the geometry of network superimposed by travel demands. The range of ORC is between negative infinity and positive infinity in theory.

3.2. Calculation of ORC

To calculate ORC, we first retrieve the neighbour nodes of both ego stops of a line segment. Neighbour nodes include OD grid, bus stop, and subway station. There are two datasets for analysis: aggregated level OD flow and individual level smart card records. An OD flow (Table 2d) is a travel between two OD grids. It is retrieved from movement heatmap of mobile phone records. OD flow is given on aggregated level as it only tells the total amount of flows between a certain pair of origin and destination grids. The specific travel routes between OD grids are unknown though. Smart card flows (Table 2c) are individual level anonymous public transit tap-on and tap-off records. Derived from smart card records are aggregative smart card traffic volumes between stops. It is worth mentioning that the aggregated OD flow is mainly used to assign the smart card flow between stops to the exact source grid pairs. With Amap API, each smart card record is assigned to a transit network route. We fetch the fastest path from Amap for each travel and add one count to each corresponding line segment. As aforementioned, six types of flows traverse a line segment. We need to know the flow volume between each pair of neighbour nodes as well as the specific route between nodes (for cost calculation) to compute ORC using Eq. 3. Different types of flows have their own ways of counting flow volume. We elaborate the details below.

Grid-grid flow An OD flow should be matched to a smart card flow so that we know which public transit route is taken, and thus how much the transit travel cost is, for a given OD pair. An OD flow is matchable to a smart card record as long as the origin falls inside the buffer of one stop and the destination inside the other. The problem is formulated as a coupling problem of smart card flow and OD flow. For computation feasibility, a greedy strategy is applied to match OD flows with smart card flows. We only satisfy the marginal probability of smart card flow since some of the OD flows travel by other modes. The smart card flows between two stations/stops are assigned to each pair of OD grids by the ratio of OD flow. Note that when the coverages of multiple station pairs overlap so as to share the same pair of grids, flows between grids are split equally to each station pair.

Station-grid flow This type of flow includes grid-station flow as well. The same for the following asymmetric flow types. Station-grid flow is calculated from the smart card records. A flow is counted as a *station-grid flow* if one of the stops is a tap-on or tap-off stop (meaning that it connects with an OD grid), while the other end connects a stop on the same line.

Transfer-grid flow While one end of the flow is a tap-on or tap-off stop, the other end connects a stop with transfer flow from a different line. Transfers between buses or between bus and subway are easily identified since smart card records directly demonstrate the information.

Table 1
Proposed redesign strategies under different scenarios.

	High flow	Low flow
Low ORC	S-I: corresponding to Fig. 4(a). Line segment has high attractiveness and high travel burden, so new lines can be added or frequencies increased, especially for very high volume segments.	S-II: corresponding to Fig. 4(b). Though line segment has high attractiveness, its traffic flow is not large. The potential of such routes should be exploited. The segment might be in parallel with a more convenient route that absorbs flows. Strategies to balance flow distribution should be invented.
High ORC	S-III: corresponding to Fig. 4(c). Line segment has low attractiveness, but still takes high pressure of traffic. Alternative routes or direct linkage are more required.	S-IV: out of focus. Line segment has low attractiveness, and the traffic volume is low. This is not the particular interest of this work.

Transfers between subway lines, however, happen with only one tap-on and tap-off, so we query Amap API to infer the subway route a person travels.

The rest three types of flows, *station-station flow*, *station-transfer flow*, and *transfer-transfer flow*, can all be calculated by manipulating the smart card data in Table 2c. The data also supports temporal analysis hour by hour. We use 7am and 5 pm for our case study, and focus on 5 pm for detailed analysis.

We also calculate the shortest travel cost between any two nodes by querying Amap API (https://lbs.amap.com/api/webservice/summary/). The weight of each linkage between two adjacent nodes takes into account its length and the travel mode of that segment. We consider three travel modes – bus, subway, and walk, each of which is assigned with a different speed. No multiple modes exist between two particular nodes since stops/stations are built uniquely for a certain mode. Speed is approximated by calling Amap API at a ratio of 6:3:1 for subway, bus, and walk considering the overhead cost of picking up or dropping of passengers.

3.3. Optimization strategy and workflow

ORC and traffic flow are two significant indicators for public transit network performance. ORC indicates *theoretical* inclination to move via a route, i.e., the attractiveness of a line segment, while traffic flow is the *observed* burden and travel demand on that segment extracted from realistic behaviours. Crossing over the two dimensions, public transit network can be redesigned according to the following strategy to fill in the gap between the theoretical optimal and people's preference. Table 1 summarizes the proposed strategies.

4. The case study in Beijing

We apply the proposed method in a case study of Beijing. The study area is within the 5*th* Ring Road. Subway and bus networks built up from stations and lines are offered by a Chinese navigation company Amap (https://ditu.amap.com/). Travel behaviour data are approximated by smart card records from public transit companies after being processed for privacy protection. OD distribution and movement heatmap on 250*250 metre grids (i.e., the hourly amount of flows from one grid to another) are retrieved from cell phone call records of Chine Unicom. Table 2 displays structures of the datasets in 2019.

We extract the line segments that have their traffic flows no greater than the median of traffic flow, or above the third quantile of traffic flow as *low flow* and *high flow* cases, respectively. We use the median of flow in lieu of its first quantile because the first quantile of flow is just 1, and its median is 2. The third quantile of flow is 9. A similar separation is conducted for ORC to extract the segments with ORC no greater than the

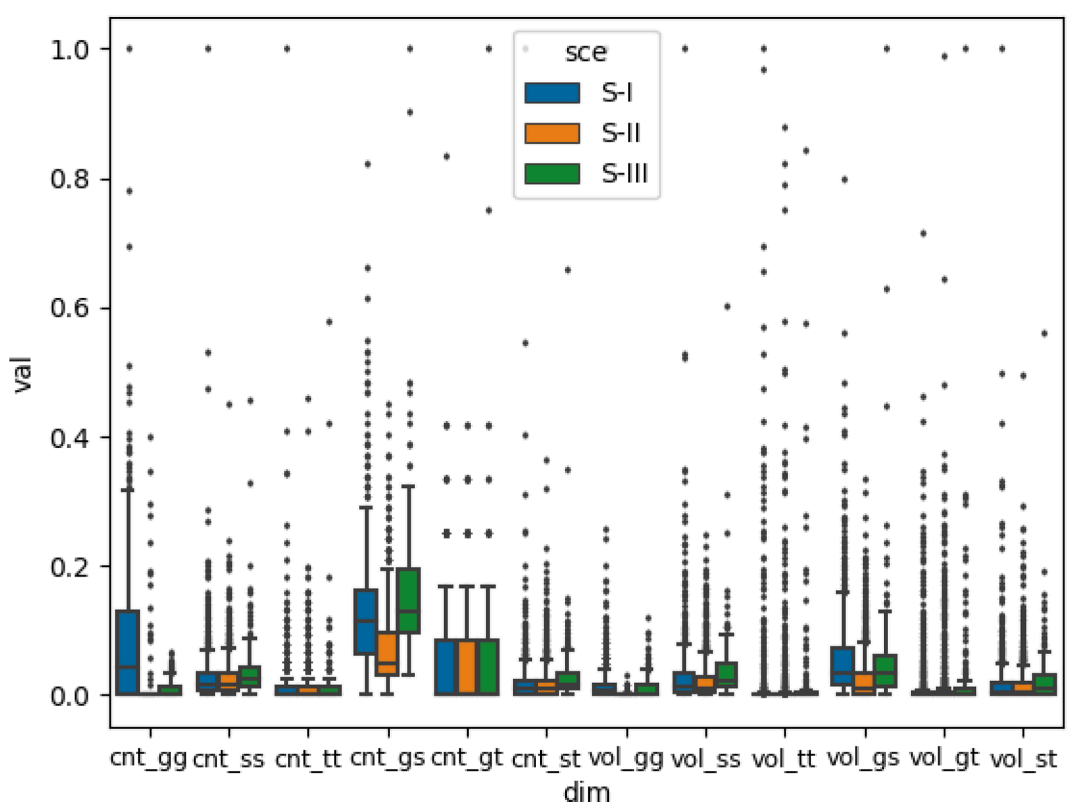
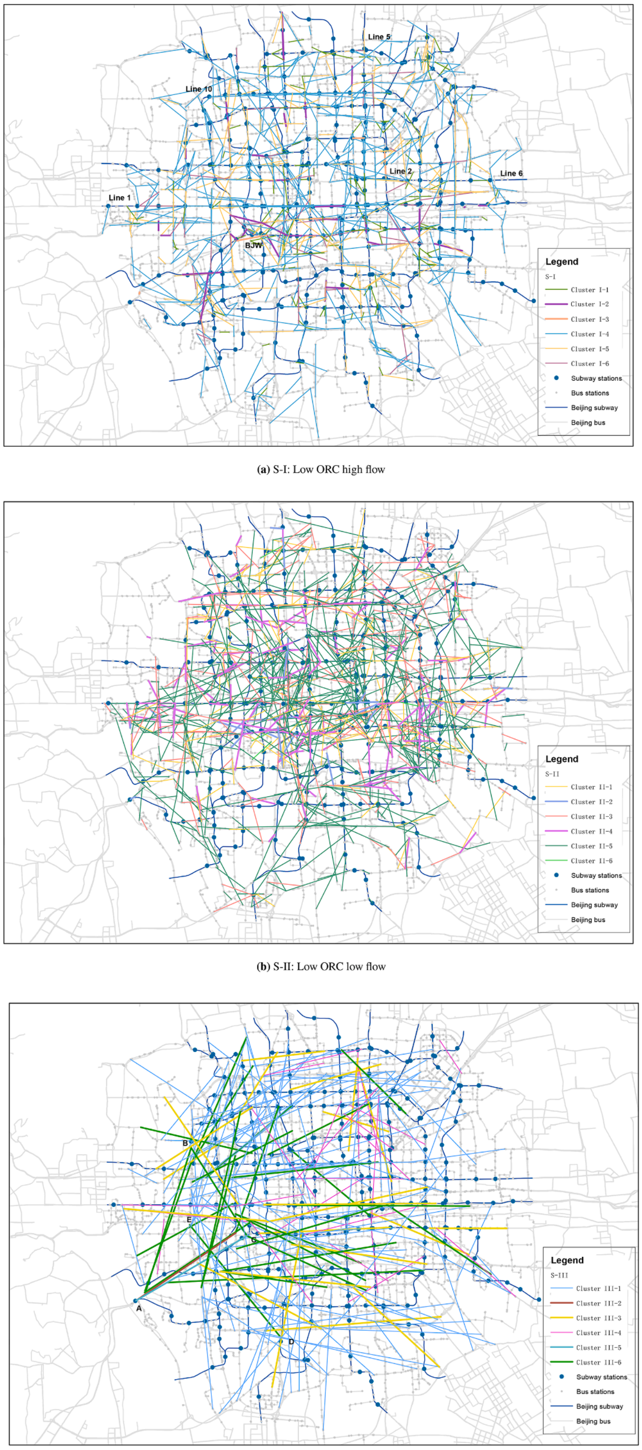


Fig. 3. Comparison of neighbourhood traits for three scenarios. There are 12 dimensions of neighbourhood traits. The first six dimensions are about the number of stop pairs belonging to that flow type, while the second six dimensions are about the total flow volume of that type. The six types are: 'gg': grid-grid, 'ss': station-station, 'tt': transfer-transfer, 'gs': grid-station, 'gt': grid-transfer, 'st': station-transfer.



(a) S-I: Low ORC high flow

(b) S-II: Low ORC low flow

(c) S-III: High ORC high flow

Fig. 4. Maps of station pairs in three scenarios. The clusters of each scenario are independent among scenarios.

first quantile of ORC (i.e., low ORC), and the ones above third quantile of ORC (high ORC). The first and third quantiles of ORC are −0.156 and 0.567. Crossing the two criteria, we get three scenarios: low ORC and high flow (S-I), low ORC and low flow (S-II), and high ORC and high flow (S-III). The results are demonstrated in Fig. 4 and elaborated in next section. Because the experiment has a very large amount of iterations when computing the inter-neighbourhood flows, we randomly sample 1/ 30 of all possible pairs of stations within the 5*th* Ring Road to compute their ORC. The heuristic for optimization, nevertheless, is not

Table 2
Data structures of the applied datasets (in the year of 2019).

	(a) Data structure of subway and bus stops
station_id	the unique station id of a subway or bus station/stop
station_geo	the geometric feature of a station
station_trans	boolean marking whether the station is for transfer
	(b) Data structure of subway and bus lines
line_id	the unique line id of a subway or bus line
line_geo	the geometric feature of a line
line_sta	the sequential list of stations on a line
	(c) Smart card data structure after being processed
unique id	anonymous card id of a user
fst_sta_id[]	a list of tap-on stations for each transfer segment of a travel
tst_sta_id[]	a list of tap-off stations for each transfer segment of a travel
ftime[]	a list of tap-on timestamps for each segment
ttime[]	a list of tap-off timestamps for each segment
mode[]	a list of travel mode for each segment
	(d) Cell phone call records for OD heatmap
oid	the unique id of origin grid of movement
did	the unique id of destination grid of movement
hour	the hour of the movement
flow	the aggregated flow volume from oid to did

Table 3
Statistics of the line segment length (metre) for different scenarios.

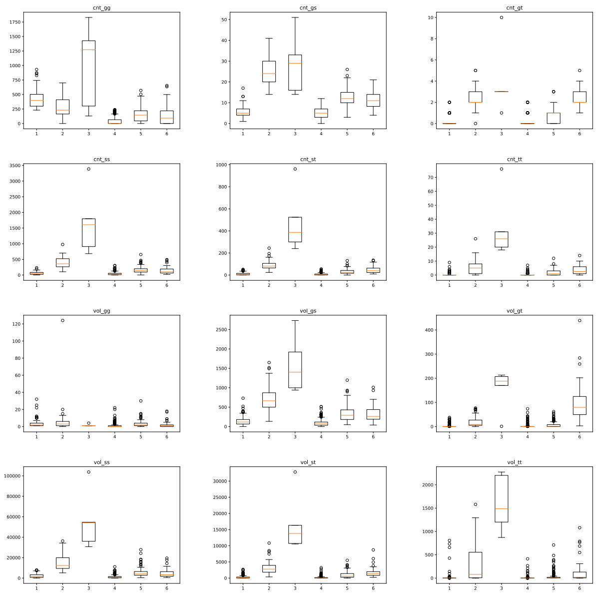
Scenario	Low ORC high flow	Low ORC low flow	High ORC high flow
Min.	65.621	14.238	2170.578
Max.	10857.832	17956.238	17563.029
Mean	1705.057	2151.117	8204.595
STD.	1220.889	1609.115	2587.549

biased by the random sample since the neighbourhood and traffic flow information is intact for each pair. We explain the heuristics in Section 5.

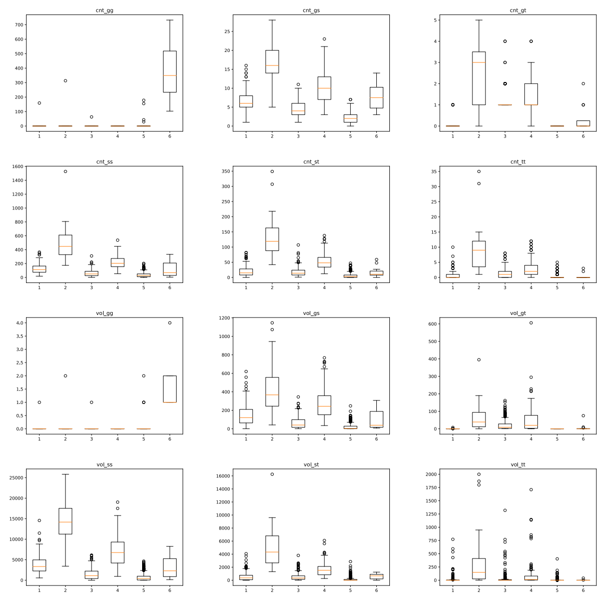
What might be important to count for different scenarios is the *traits of neighbourhood*, i.e., the types and amounts of flows between two ego nodes. Since there are six types of flows, we hypothesize that different distributions of flow types may affect the scenario a route belongs to. For example, the low ORC and low flow scenario may be associated with certain types of flows, which appears more frequently at particular types of locations or for particular scenarios. To uncover the relationships behind, we encode the flows between neighbourhoods with a 12-dimensional vector, each dimension of which represents the count or volume of the six types of neighbour flows. Then we further cluster line segments of each scenario into six sub-categories (refer to each color in Fig. 4) for more detailed analysis, since one scenario has big variance. The clustering is conducted based on neighbourhood trait vector by K-Means. The number of sub-categories is decided with the trade-off between clustering silhouette score and easiness for interpretation under different scenarios. The colors in bold lines on Fig. 4 should be paid more attention because they have certain neighborhood features of higher volumes than other clusters. All the colors in bold lines should be redesigned more urgently than other clusters of the same scenario. Line segments with special neighbourhood traits are further extracted. Each category is closely associated with the socio-economic background as well as transportation structure of that locality. We elaborate the findings as follows.

5. Results and discussions

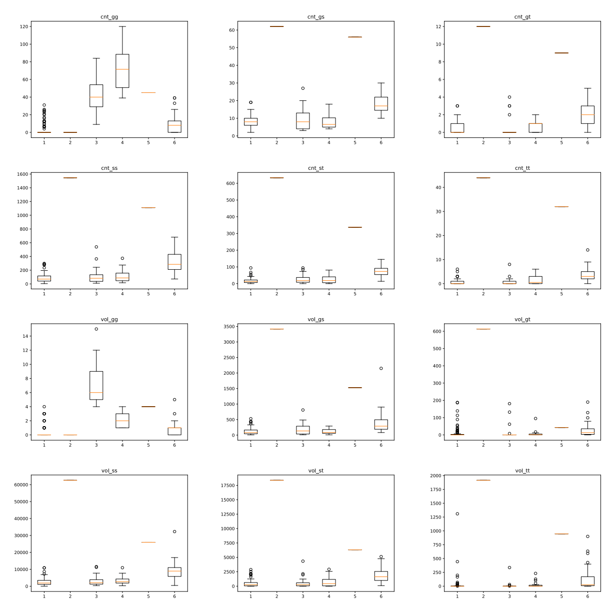
In the study area, we fetch 5,852 bus stops, 3,575 bus lines, 634 subway stations, and 52 subway lines. Since we only investigate a random sampling of 1/ 30 of all possible pairs of stations (stops) during the hour between 5 pm and 6 pm, the final total amount of line segments whose ORC has been computed is 9,698. There are 906 pairs of ego stations extracted for S-I, 1,095 pairs for S-II, and 197 pairs for S-III. The



(a) Count and volume of flows, S-I



(b) Count and volume of flows, S-II



(c) Count and volume of flows, S-III

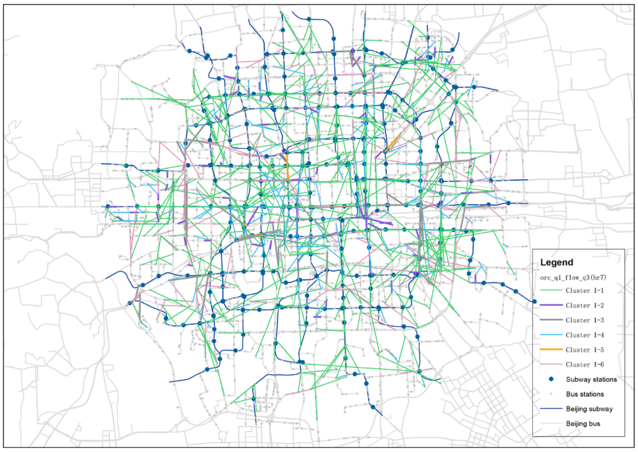
Fig. 5. Boxplots of neighbourhood traits for different clusters in different scenarios (5 pm).

following discussions refer back to Table 1 for different scenarios. Fig. 4 illustrates their spatial layouts.

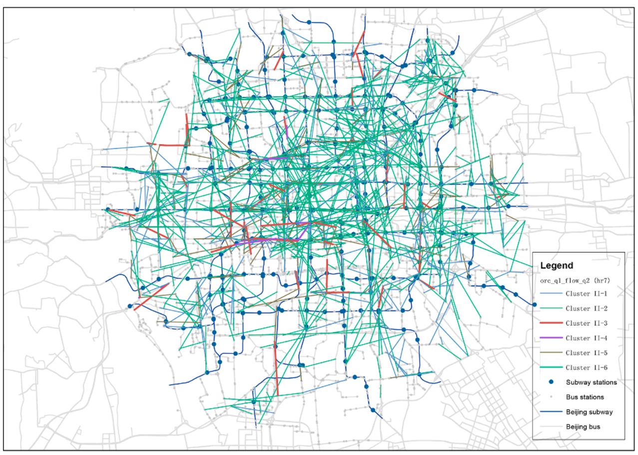
5.1. Analysis of neighbourhood traits in three scenarios

As aforementioned, we hypothesize that neighbourhood traits may partially decide the scenario of a line segment. However, different scenarios overall do not yield significant differences in the distribution of flows (Fig. 3; values are normalized on each dimension). The scenario with low flow (S-II) is seen to have low amount in all types of flows. Additionally, *grid-station* flows are higher in all three scenarios, followed by *station-station* flows. Though S-I has low ORC while S-III has high ORC, they cannot be distinguished by neighbourhood traits. However, we observe that distance (length) of line segment is an ostensible factor that tells apart the two scenarios. Usually low ORC line segments are shorter than high ORC segments because lower ORC indicates higher likeliness to be traversed by neighbourhood flows. People are more likely to share a shorter leg of travel than sharing a longer leg. Therefore, we see in Fig. 4 c the line segments are longer. Table 3 also substantiates the argument. Segments in S-I are overall shorter than in S-II, while S-III witnesses the longest segments. Spatially the lines of different scenarios are complementary. Low ORC and high flow segments (S-I) mostly exist between subway stations. These segments not only have structural advantages to attract flows but also take heavy load already.

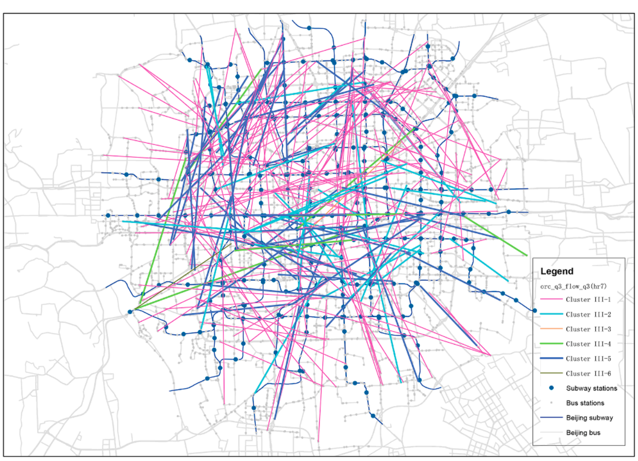
When further examining sub-categories of neighbourhood vectors, distinguishable properties are found. Fig. 5a demonstrates that distinct sub-categories in S-I are Clusters I-2 and I-3. Both clusters have more flows than others, particularly in station-station flow, grid-station flow, and station-transfer flow. Cluster I-3 has very few grid-grid flow, but is leading in all other types of flow. Visualizing them on map (Fig. 4), we see that these line segments are mostly short-distance routes between subway stations, especially transfer stations. These segments usually have temporary high passenger load in carriages and on platforms. Cluster I-2 are line segments near main stations or between busy areas.



(a) S-I: Low ORC high flow



(b) S-II: Low ORC low flow



(c) S-III: High ORC high flow

Fig. 6. Maps of station pairs in three scenarios (hour = 7). The clusters of each scenario are independent among scenarios.

In S-II (Fig. 5b), line segments are also of short distance, Cluster II-2 has the highest flow in all types amongst all clusters in the same scenario since this cluster have one side connected with main stations or stops while the other end with an unpopular station. That is why it has lower flow volume than S-I. Line segments with more station-station flows are likely to have negative ORC because flows have been bounded to the segment and thus have to traverse the ego stations. Cluster III-3 in S-III is a special category which is high in grid-grid flow but low in the rest types of flow. The cluster corresponds with the yellow lines in Fig. 4c, representing major origin-destination flows. Clusters III-2 and III-5, each with only one segment, are relatively high in all types of flows connecting Beijing West Railway Station as well as its transport hub and a tourism destination Wanping City.

Transfer-related flows are more important indicators of the attractiveness of a line segment. For one thing, people choose the most convenient route from multiple transfer alternatives. A route with high transfer flows is proven to be a good choice. For another, attractiveness for transfers reflects the importance of a route not only for local travel demands but also for longer distance demands. In Fig. 5, we observe a relatively high volume of station-transfer flows. In S-I (Fig. 5a, *vol_st*), Cluster I-3 has the highest flow. It corresponds with Fig. 4a showing that Cluster I-3 are mostly short-distance routes between subway stations with at least one transfer station. In S-II (Fig. 5b, *vol_st*), Cluster II-2 has the highest station-transfer flow. Shown on Fig. 4b, Cluster II-2 are routes between stops near popular subway stations. These routes are likely to take passengers transferred from subways. In S-III (Fig. 5c and Fig. 4c), high volume of transfer appears at Clusters III-2 and III-5, both of which have one end as a transfer station (specifically near Beijing West Railway Station).

The result of different scenarios at 7am is given in Figs. 6 and 7. We see slightly different patterns at details, but the general heuristics are similar.

5.2. Strategic heuristics for transit line optimization

The heuristics from ORC provide potential strategies for optimization. In S-I, the line segments are of high attraction and high traffic load. We suggest careful passenger flow management and safety control on platforms to avoid disorder or accidents. Alternative travel routes might also be considered to split travel load, i.e., inducing flows to another place for transfer.

In S-II, structure-wise line segments are of high attraction but low flow volume. The segments ostensibly are not between popular stations. A further look at the neighbourhood flow types (Fig. 5b) shows that grid-grid pair is of low number, which indicates that the segments are not between popular stations or places. Other types of stations or flow volumes are lower than the rest two scenarios as well. We argue that such segments have good potential to attract more flows but are underused at the moment.

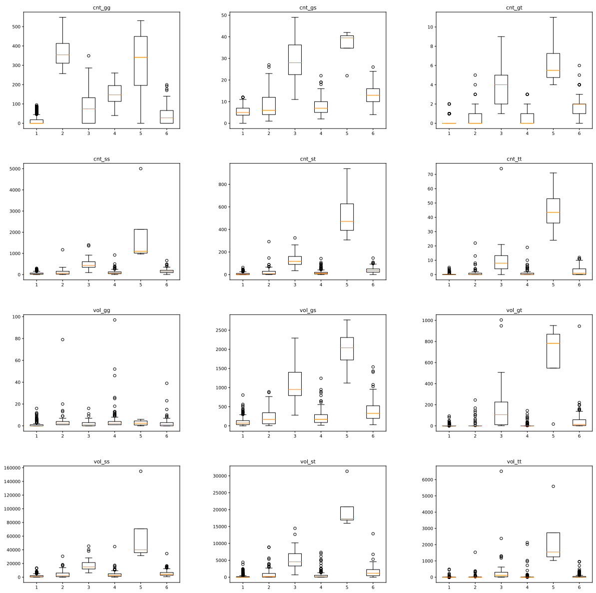
In S-III, the line segments are categorized as low attraction according to ORC, despite of high passenger flow. The value of ORC, however, is just slightly above zero, meaning that the comparative advantage of alternative routes are not significant. Since long distance routes usually contain transfers, there are more alternative route choices that might be slightly shorter than the line segment under investigation. There is also a case that, when bus routes are actually shorter but in parallel with a subway route, people subjectively prefer subway. Psychology of route choice is a potential future work. Generally routes of S-III should consider the convenience of long-distance travels, e.g., line operation time, frequency, and transfer times. The potential of alternative routes with high attraction should also be deployed.

5.3. Validation of the proposed method

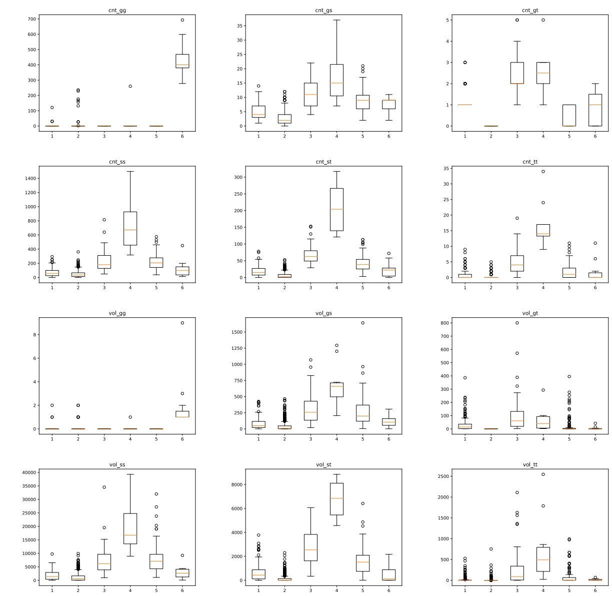
Strictly speaking, the proposed optimization heuristics can be validated only when the strategy is put into practice for social experiment. However, we can only do a theoretical validation for the moment. We propose two ways to validate the optimization strategic heuristics: 1) *Intuitive interpretation*: visualizing the results and interpreting them with intuitive findings based on life experience, and 2) *Comparison with betweenness*: comparing the results with other indices for mutual validation.

5.3.1. Intuitive interpretation

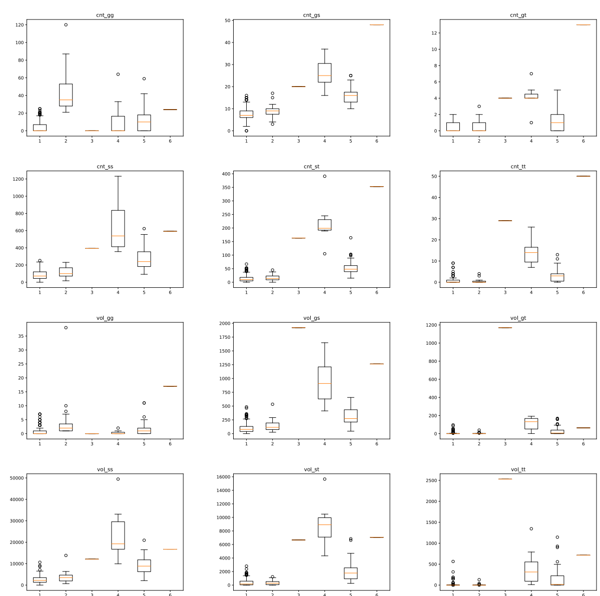
Fig. 4 generally demonstrates reasonable results to life experience.



(a) Count and volume of flows, S-I



(b) Count and volume of flows, S-II



(c) Count and volume of flows, S-III

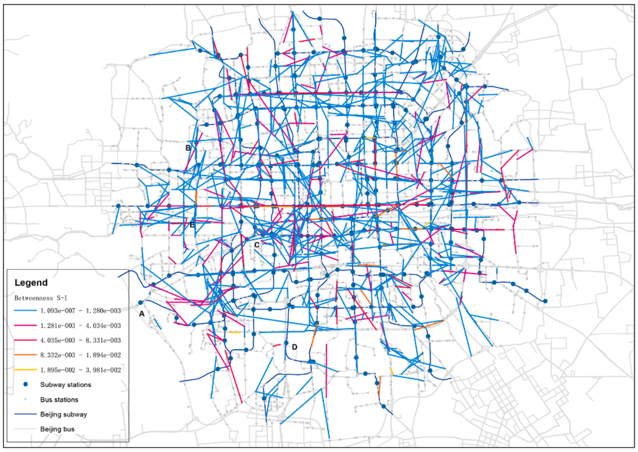
Fig. 7. Boxplots of neighbourhood traits for different clusters in different scenarios (7 am).

On Fig. 4a, the routes on subway lines are mostly of high attraction and high flow, e.g., the north rim of the bigger loop line (Line 10) and the stations on the horizontal axis (Line 1) to the left of the map centre. The north of Line 10 loop have busy routes connecting a major workplace Zhongguancun (known as the Chinese Silicon Valley) and a few crossovers with Lines 4, 13, 8, and 5 from west to east. Line 1 goes along Chang'an Street, which is the horizontal main trunk in Beijing, and intersects with seven subway lines. Financial Street, the Forbidden City, and a few commercial districts as well as shopping areas are all on the route. Cluster 3 on the figure corresponds with popular subway line segments: the transfer hub at Chaoyangmen between Line 6 and Line 2, the transport hub at Dongzhimen, the transfer at Yonghegong Lama Temple between Line 5 and Line 2, and Beijing West Railway Station (*BJW* on the map). On Fig. 4b, we observe that the segments are mostly between less popular stations, as discussed above. The finding matches intuition. Fig. 4c detects station pairs of strong OD connection, e.g., between a tourism destination (A) and transport hub Beijing West Railway Station (C). The findings in Section 5.1 follow intuitive interpretation.

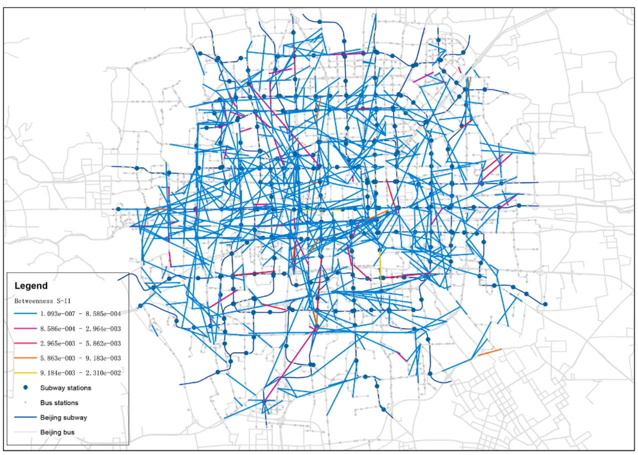
5.3.2. *Comparison with betweenness centrality*

Although ORC and betweenness share a concept of "traverseness" indicating the inclination of flows to pass through a route, there are some major differences between the two measures. ORC focuses more on the local neighbourhood of a route while betweenness is computed from a whole network or part of a network instead of a route. Specifically, betweenness indicates how likely a route is on the shortest path of any two OD points in the network. The measure hence is highly coupled with OD points. Consequently, betweenness is only with regard to a particular transit line. Multiple betweenness values hence exist between the same pair of stations if there are multiple transit lines. The measure is eventually for an existing specific line rather than a potential path connecting two stations. On contrast, ORC is a measure for an optimal route between two stations, since we assume that flows between two neighbourhoods will traverse the best route.

The highest betweenness for each line segment of the three scenarios



(a) S-I: Low ORC high flow



(b) S-II: Low ORC low flow

(c) S-III: High ORC high flow

Fig. 8. Betweenness centrality of station pairs corresponding to ORC in three scenarios.

is demonstrated in Fig. 8. We see obvious distinctions between the patterns of ORC and of betweenness. Distribution of betweenness is highly positively skewed with a long tail. Accordingly, high betweenness appears at main trunk routes, e.g., Chang'an Street (horizontal axis route in Fig. 8a) and the north rim of Line 10, as mentioned in the previous paragraph. However, a majority of line segments have low betweenness (blue segments), with 769 line segments out of 906 in S-I and 1038 out of 1095 in S-II, though they belong to high attraction (low ORC) scenario. Additionally, S-I has more high betweenness line segments (in magenta, red, orange and yellow) than S-II (137 in S-I, 57 in S-II), which reflects that S-I has higher flow volume than S-II. In terms of ORC, however, both scenarios have low ORC and high attraction, which cannot be distinguished by betweenness. The finding indicates that betweenness is sensitive to flow volume, but cannot tell the potential or attraction of a route. In S-III, betweenness demonstrates inconsistent pattern from ORC. S-III should have low attraction, but betweenness of the corresponding pairs has high variance with four magnitude difference, varing from e-7 to e-4, so betweenness cannot steadily detects the traits here. There is also owed to the aforementioned reason. Betweenness is computed globally, assuming that at every decision point, e.g., a route entrance or major turning point, travellers will stick to the theoretical shortest time path. However, in reality there are more complex decision factors that deviate people from the shortest path. The deviation cannot be reflected by betweenness centrality. On comparison, ORC is computed from the realistic distribution of flows surrounding stations, which accommodates in-route choice change.

5.4. Advantages of ORC

We argue that, as a measure to indicate the innate structure property of a network space, ORC has the following advantages compared with other measures.

- ORC offers an additional dimension to traffic volume as a heuristic for optimization. Traffic volume alone can only indicate the current situation, but cannot tell the structural potential of a line in carrying traffic flow. Traffic volume and ORC are two complementary dimensions.
- Compared with betweenness centrality, ORC is a localized measure that depicts flows drawn from a station's neighbourhood. The concept of *neighbourhood* of ORC takes into account not only OD distribution but also transfer flows from other routes as well as ad-hoc choice change.
- ORC focuses on evaluating current public transit network and proposing strategic heuristics for optimization, rather than the process of optimization itself. Previous methods discussed in Introduction deal with optimization by applying heuristic algorithms to build a new network. ORC should be regarded as a preliminary analysis to simplify optimization process.

5.5. Potential issues in the calculation of ORC

For computation simplicity, we pick one reasonable method – equally splitting – to split OD flows when they are covered by the service areas of multiple transit stops. However, it is noteworthy that other methods are also acceptable, e.g., to split the flows in proportion to the total amount of flows at each accessible stop, though the computation cost will be higher. Splitting the flows equally infers that people have equal possibility to choose an accessible stop when no additional information is given. While splitting flows in proportion to the total volume of all the accessible stops are also reasonable if we only consider the flow distribution at origin or destination, but people pick up a stop that is most convenient to carry them to their final destination. In this sense, the actual detailed travel demands (including accurate ODs) should be known to decide which method is better. Unfortunately, the details are not available in the present data, so whichever method that is justifiable can be applied.

6. Conclusions and future work

This study proposes a heuristic to redesign public transit network when a relatively mature transit system already exists. Different from previous network optimization heuristics that plan routes or operational frequencies for pre-specified stations, the present work focuses on detecting the gap between travel demand and transit service supply to identify the line segments that should be upgraded. Ollivier-Ricci curvature (ORC) is leveraged to integrate travel demand distribution reflected by human travel big data and public transit network structure, which is a speciality characterizes ORC. We revise the original form of ORC and adapt it to transit network analysis. The measure depicts whether flows converge to or diverge from a line segment, considering six types of combinatorial flows from three sources: from OD grids, from stations on the same line, and from transfers. The case study of Beijing validates our hypothesis that ORC is able to capture the gap between demand and supply. Negative ORC indicates that flows converge to a segment of route between two stations, while positive ORC means divergence. According to both traffic volume and ORC, we offer optimization suggestions for each scenario.

There are a few points future works may focus on. As a barrier also for other heuristics, computation efficiency is a challenge. Our method takes time to iterate the calculation of ORC between each pair of stations, of which each pair of station has multiple neighbourhood stations. A reasonable spatial partition may help speed up the process. The second issue is how to design a replacement of the current segment that should be improved. Our work mainly identifies the segments that should be upgraded, and provides a macro strategy for different scenarios, but the work does not offer a concrete upgrade strategy. Specific methods to replace or build alternative lines to problem segments may rely on heuristics offered in other studies.

Funding information

The National Key Research and Development Program of China (2017YFE0196100), and the National Natural Science Foundation of China (41771425, 41830645, 41625003).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by grants from the National Key Research and Development Program of China (2017YFE0196100), and the National Natural Science Foundation of China (41771425,41830645,41625003). We also appreciate the detailed comments from the Editor and the anonymous reviewers.

Appendix A

Tables 2 and 3
Figs. 5–8

References

Aldous, D., Barthelemy, M., 2019. Optimal geometry of transportation networks. Phys. Rev. E 99, 1–10. https://doi.org/10.1103/PhysRevE.99.052303.

Badia, H., Argote-Cabanero, J., Daganzo, C.F., 2017. How network structure can boost and shape the demand for bus transit. Transp. Res. Part A: Policy Pract. 103, 83–94. https://doi.org/10.1016/j.tra.2017.05.030.

Barahimi, A.H., Eydi, A., Aghaie, A., 2021. Multi-modal urban transit network design considering reliability: multi-objective bi-level optimization. Reliab. Eng. Syst. Saf. 216, 107922.

Billheimer, J.W., Gray, P., 1973. Network design with fixed and variable cost elements. Transp. Sci. 7, 49–74. https://doi.org/10.1287/trsc.7.1.49.

Buba, A.T., Lee, L.S., 2018. A differential evolution for simultaneous transit network design and frequency setting problem. Expert Syst. Appl. 106, 277–289.

Capali, B., Ceylan, H., 2020. A multi-objective meta-heuristic approach for the transit network design and frequency setting problem. Transp. Plan. Technol. 43 (8), 851–867.

Chen, Y., Bouferguene, A., Li, H.X., Liu, H., Shen, Y., Al-Hussein, M., 2018. Spatial gaps in urban public transport supply and demand from the perspective of sustainability. J. Clean. Prod. 195, 1237–1248. https://doi.org/10.1016/j.jclepro.2018.06.021.

Cipriani, E., Gori, S., Petrelli, M., 2012. Transit network design: A procedure and an application to a large urban area. Transp. Res. Part C: Emerg. Technol. 20, 3–14. https://doi.org/10.1016/j.trc.2010.09.003.

Dubois, D., Bel, G., Llibre, M., 1979. A set of methods in transportation network synthesis and analysis. J. Oper. Res. Soc. 30, 797–808. https://doi.org/10.1057/jors.1979.190.

Evans, L.C., 1997. Partial differential equations and Monge-Kantorovich mass transfer. Curr. Developm. Math. 65–126. https://doi.org/10.4310/cdm.1997.v1997.n1.a2.

Farahani, R.Z., Miandoabchi, E., Szeto, W.Y., Rashidi, H., 2013. A review of urban transportation network design problems. Eur. J. Oper. Res. 229, 281–302. https://doi.org/10.1016/j.ejor.2013.01.001.

Farber, S., Ritter, B., Fu, L., 2016. Space-time mismatch between transit service and observed travel patterns in the Wasatch Front, Utah: A social equity perspective. Travel Behav. Soc. 4, 40–48. https://doi.org/10.1016/j.tbs.2016.01.001.

Gao, L., Liu, X., Liu, Y., Wang, P., Deng, M., Zhu, Q., Li, H., 2019. Measuring road network topology vulnerability by Ricci curvature. Physica A 527, 121071. https://doi.org/10.1016/j.physa.2019.121071.

Gutiérrez-Jarpa, G., Laporte, G., Marianov, V., Moccia, L., 2017. Multi-objective rapid transit network design with modal competition: The case of Concepción, Chile. Comput. Oper. Res. 78, 27–43. https://doi.org/10.1016/j.cor.2016.08.009.

Heyken Soares, P., Mumford, C.L., Amponsah, K., Mao, Y., 2019. An adaptive scaled network for public transport route optimisation. Public Transp. 11, 379–412. https://doi.org/10.1007/s12469-019-00208-x.

Islam, K.A., Moosa, I.M., Mobin, J., Nayeem, M.A., Rahman, M.S., 2019. A heuristic aided stochastic beam search algorithm for solving the transit network design problem. Swarm Evol. Comput. 46, 154–170.

Jha, S.B., Jha, J.K., Tiwari, M.K., 2019. A multi-objective meta-heuristic approach for transit network design and frequency setting problem in a bus transit system. Comput. Ind. Eng. 130, 166–186.

Kuan, S.N., Ong, H.L., Ng, K.M., 2006. Solving the feeder bus network design problem by genetic algorithms and ant colony optimization. Adv. Eng. Softw. 37, 351–359. https://doi.org/10.1016/j.advengsoft.2005.10.003.

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Shi, L., 2015. Social sensing: A new approach to under-standing our socioeconomic environments. Ann. Assoc. Am. Geogr. 105, 512–530.

Manser, P., Becker, H., Hörl, S., Axhausen, K.W., 2020. Designing a large-scale public transport network using agent-based microsimulation. Transp. Res. Part A: Policy Pract. 137, 1–15.

Ni, C.C., Lin, Y.Y., Luo, F., Gao, J., 2019. Community detection on networks with Ricci Flow. Sci. Rep. 9, 1–13. https://doi.org/10.1038/s41598-019-46380-9.

Ollivier, Y., 2010. A survey of Ricci curvature for metric spaces and Markov chains. In: Advanced studies in mathematics, vol. 57, pp. 343–381. https://doi.org/10.2969/aspm/05710343.

Sia, J., Jonckheere, E., Bogdan, P., 2019. Ollivier-Ricci curvature-based method to community detection in complex networks. Sci. Rep. 9, 9800. https://doi.org/10.1038/s41598-019-46079-x. Retrieved from http://www.nature.com/articles/s41598-019-46079-x.

van Nes, R., Hamerslag, R., Immers, B.H., 1988. Design of public transport networks. Transp. Res. Rec. 74–83. https://doi.org/10.1002/etc.175.

Wang, Y., Zhu, D., Yin, G., Huang, Z., Liu, Y., 2020. A unified spatial multigraph analysis for public transport performance. Sci. Rep. 10, 1–9. https://doi.org/10.1038/s41598-020-65175-x.

Wang, Z., Li, X., Zhu, X., Li, J., Wang, F., Wang, F., 2021. Big data-driven public transportation network: a simulation approach. Complex Intell. Syst. 1–13.

Yao, B., Hu, P., Lu, X., Gao, J., Zhang, M., 2014. Transit network design based on travel time reliability. Transp. Res. Part C: Emerg. Technol. 43, 233–248. https://doi.org/10.1016/j.trc.2013.12.005.

Yu, B., Yang, Z.Z., Jin, P.H., Wu, S.H., Yao, B.Z., 2012. Transit route network design-maximizing direct and transfer demand density. Transp. Res. Part C: Emerg. Technol. 22, 58–75. https://doi.org/10.1016/j.trc.2011.12.003.