Applied Network Science

RESEARCH

Open Access

Social and economic flows across multimodal transportation networks in the Greater Tokyo Area

Aaron Bramson[1,2,3,4*], Megumi Hori[1], Bingran Zha[1] and Hirohisa Inamoto[1]

*Correspondence:
a_bramson@ga-tech.co.jp
[1]GA Technologies Inc., Roppongi Grand Tower 40F, Roppongi 3-2-1, Minato-ku, Tokyo, 106-6290, Japan
[2]Laboratory for Symbolic Cognitive Development, RIKEN Center for Biosystems Dynamics Research, 6-7-3 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan
Full list of author information is available at the end of the article

Abstract

We model the flow of human capital and resources across multimodal transportation networks throughout the Greater Tokyo Area. Our transportation networks include trains, buses, and roads integrated with a walking network among a geographically grounded hexagonal grid and connecting nodes of different modes. The hexagonal grid holds data on both the working population and number of jobs from which we built probability distributions for the origins and destinations of commuting trips. Using both the network simplex method and stochastically generated origin-destination trips we estimate the population flows necessary to satisfy this demand. Rather than micro-simulations of actual commuting patterns, congestion, or route planning, our approach aims to uncover patterns in the aggregate flow of human resources to and from economic opportunities. We describe the details of the socioeconomic data, network generation, and the results of our exploratory analysis, then discuss the implications of these findings for transportation usage and future work.

Keywords: Multimodal transportation network, Urban mobility, Economic flows, Tokyo

Introduction

We are interested in how transportation networks shape the movements of people, money, goods, and opportunities in urban settings, and how these vary geographically within a metropolitan area. By combining the population distribution with the distribution of available jobs we explore, analyze, and evaluate the movements of human capital across the train, bus, and road networks throughout the Greater Tokyo Area of Japan.

Building on previous work to score and categorize regions of the Tokyo area based on accessibility, sociability, and features of their varied transportation networks, here we consider more deeply the demographic and economic characteristics of the population and analyze how it relates to activity across the multimodal transportation networks. First we evaluate the balance of population and occupational engagements (jobs) at each micro-location (hex) to construct a map of sources and sinks of human capital. Then we analyze the flow of human capital from sources to sinks across the multimodal transportation network. This analysis allows us to capture aggregate patterns in the flow of personnel into workplaces and economic resources back into homes.

Our exploratory analysis aims to uncover patterns of network usage and the role of both geographic regions and the transportation modes in the socioeconomic system. In this treatment our focus is not routing, scheduling, urban design, or even fine-grained traffic prediction. Rather, our focus is on (1) describing our method to intuitively merge geographic data with multiple modes of transportation, and (2) comparing methods to generate flows across the multimodal network based on that geographic data. Although we include some structural characteristics of the networks, the scale of the integrated network involved in this study (435,445 nodes and 2,921,886 edges) makes many previous measures/analyses impractical (von Ferber et al. 2009; Barthélemy 2011; Ren et al. 2014). Therefore some compromises are necessary to extract useful, even if imperfect, information about the system and flows on it. We perform a limited test of our results against empirical data to determine the plausibility of the aggregate pattern we generate, and also discuss the implications for mobility and accessibility leading to future work.

Data

Performing any analysis of accessibility requires a fusion of network data and geospatially distributed socioeconomic data. For our analysis we create a hexagonal grid covering the region to capture the geographic data. We then connect the cells of this spatially explicit hex grid to their neighbors, and link appropriate cells to the transportation networks, to create a unified geographically embedded transportation network that facilitates network analyzes of flow across the multimodal transportation systems in a parsimonious way. The structure of the resulting unified network differs from previous studies of single-mode transportation networks, thus complicating direct comparisons with previous research (Barthélemy 2011). That said the multimodal aspect of our analysis is critically important. We believe it is more fruitful to pursue the multimodal path, and present a rich enough description for future comparisons, than to isolate single modes for ease of analysis and comparison.

Most analyses of commuter traffic focus on one mode (Crucitti et al. 2006; Derrible and Kennedy 2009; Barthélemy 2011; Derrible 2012; Ren et al. 2014; Tak et al. 2014), compare exclusive modes (Huang and Levinson 2015), or lump public transit together (von Ferber et al. 2009) while ignoring transit access and private vehicle commuting. Although multimodal network analysis is common among logistic network analyses (Lozano and Storchi 2001; Cipriani et al. 2006; Wang et al. 2009; Ayed et al. 2011; Idri et al. 2017; Verga et al. 2018), their focus is typically optimal trips for resource delivery rather than our interest of general usage patterns. However, integrating all the modes together (including walking) with geographically-based socioeconomic data is crucial for reflecting actual usage patterns and to reflect modal interdependencies (Ministry of Land 2010b; Parady et al. 2018).

Socioeconomic data

The geographic foundation of our analysis is a 54,127m^2 (= 0.54 km^2, 125m inner radius) hexagonal grid covering all of Japan. We use GoogleMap's coordinates of Tokyo Station (139.7649361E, 35.6812405N) as a fixed reference point and grow the hexes outward from there. Because at different latitudes, the translation between degrees and meters changes, we use this method to ensure a true 250m hexagonal grid with minimal distortion around central Tokyo. Here we include four prefectures of Japan encompassing the Greater Tokyo Area: Tokyo, Kanagawa, Saitama, and Chiba. The area covered by these four prefectures is quite expansive (13,565 km^2, about the same size as Montenegro or the state of Connecticut) requiring 244,721 hexes. It includes a diversity of region types (the world's largest city, many secondary city centers, suburban sprawl, rural areas, and remote mountain forests) to capture a diversity of economic flow types.

Our population data obtained from (Official Statistics of Japan 2015) comes as a 250m sided *square grid* broken down by age group, sex, and nationality allowing us to focus on those engaged in the workforce (aged 15–64 inclusive) calibrated to the map using mesh coordinates from (Association for Promotion of Infrastructure Geospatial Information Distribution 2015a). Our economic data from (Official Statistics of Japan 2014) reports the numbers of establishments and persons engaged/employed in a 500m sided square grid which is similarly calibrated using mesh coordinates from (Association for Promotion of Infrastructure Geospatial Information Distribution 2015b). In order to utilize this data together, we resample both datasets to our hex grid using the overlap proportions of the hexes (H_i) and the respective square grids (S_j) via Eq. 1:

$$\text{Value}(H_i) = \left\lfloor \sum_j \text{Value}(S_j) \frac{\text{Area}(H_i \cap S_j)}{\text{Area}(S_j)} \right\rceil . \tag{1}$$

This allows us to interpolate any geographic dataset into a homogenized foundation that also connects naturally to our geographically grounded transportation network while utilizing all the data available with minimal location distortion[1].

The population of the covered region is 36,089,517, and the population aged 15–64 (the "working-age population") is 22,743,338; the working-age population distribution is shown in Fig. 1. The central core has a low average density and many hexes with zero populations. The hexes containing large train stations and their surrounding shops also have populations of zero, even in the suburbs.

The distribution of the area's 16,614,941 jobs is nearly the inverse of the population density, as can be seen in Fig. 2. Most jobs, and therefore most hexes with large numbers of jobs, are focused around stations in the city core, with islands of economic activity around train stations throughout the region. A summary of the population and employment data used in our analysis can be found in Table 1.

The large discrepancy between the "working age" population and the number of jobs is easily explainable considering the wide age range beginning at 15: many of the included people are still students. Also, because this employment figure only includes full-time "permanent" jobs, all the part-time, temporary, and other "irregular" positions are excluded. There is additionally some portion of the population that does not work (by choice or otherwise). In the future we aim to generate better usage pattern predictions by including trips for purposes other than home-work commuting, but we do not yet have data on the numbers and locations of school enrollments, part-time jobs, tourism (e.g. hotel occupancy and tourist site visits), or shopping excursions. We are also interested in running simulations on bracketed income levels (and/or other job categories) to better match segments of the population to jobs they are likely to have.

Network data

We utilize five interwoven networks representing distinct modes of transportation: hex, rail, bus, road, and walking. Rail (train, subway, and light rail) travel is considered the

[1] We round the values for each hex to produce more intuitive results (no fractional people or jobs), but as a result the total value of the resampled data across the region may not exactly equal the value of the original data.

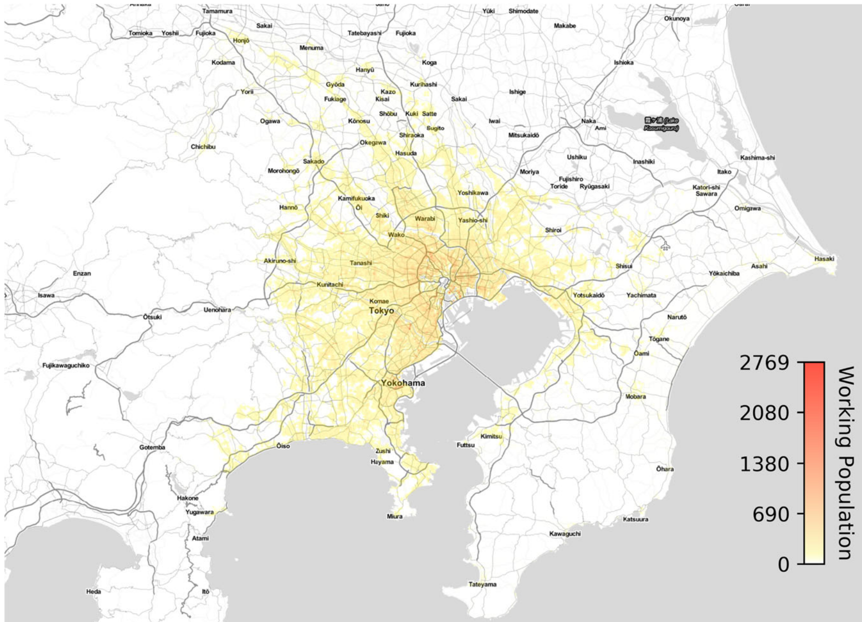

Fig. 1 A map of the Tokyo area used in the analysis with the hex grid overlay showing the working age population data. Map tiles by Stamen Design (2019), map data by OpenStreetMap (2019)

dominant mode of transportation in Japanese cities (more than 30% of all travel in Japan is done by rail OECD Statistics: Transport Transport Measurement Passenger transport (2016)). Tokyo has the densest rail network in the world and the greatest frequency of trains (Public Purpose: Urban Transport Factbook 2003) and greatest rail ridership (41,067,231 daily rides (Train Media 2017)). Furthermore, urban economic development

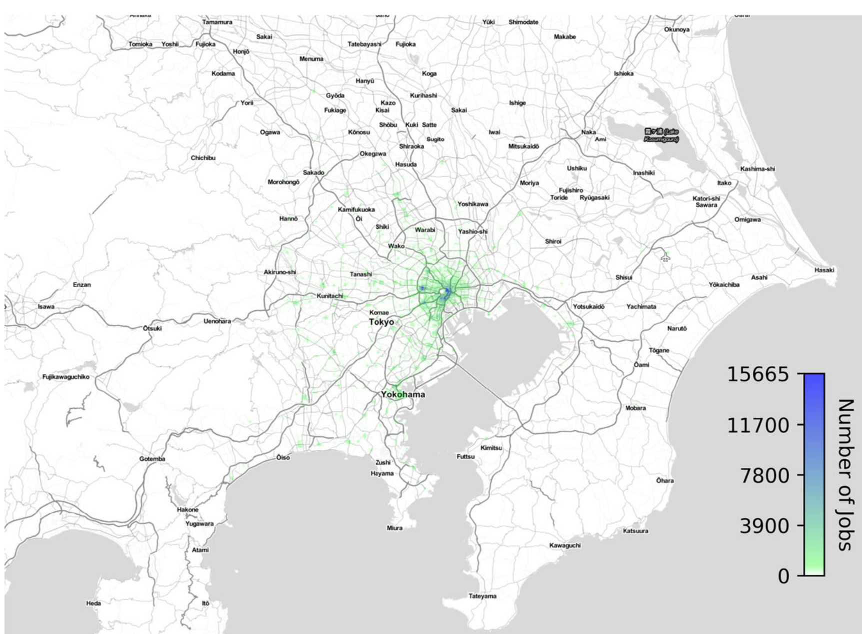

Fig. 2 A map of the Tokyo area used in the analysis with the hex grid overlay showing the employment engagements (i.e., jobs) data. Map tiles by Stamen Design (2019), map data by OpenStreetMap Contributors (2019)

Table 1 Summary of the population and employment data used in our analysis

Number of hexes	244,721
Total number of jobs	16,614,941
Mean jobs per hex	67.893
Largest job count in one hex	15,665
Total population	36,089,517
Total population aged 15–64	22,743,338
Mean population aged 15–64 per hex	92.936
Highest population in one hex	2,024

throughout Japan, and especially the Tokyo metropolitan area, is highly train-centric (Chorus and Bertolini 2011; Calimente 2012). Although in terms of sheer numbers of nodes and edges the bus and road networks both dominate the train network (Table 2), and car-based travel accounts for a third of trips (Ministry of Land 2010b), for most Japanese urban areas the train network handles the greatest traffic in terms of the number of people per kilometer.

We next describe the process used to create the network for each mode of the transportation system, and then how we connected these modes to create the integrated multimodal network. There are several factors to consider when constructing any model (realism vs generality, efficiency vs detail, analysis algorithm requirements, etc.), and in this case we also need to choose a representation that captures each mode with sufficient fidelity while still fostering the integration of the networks together, and with the geographical data. Network models come in variety of forms (flat, layered, multigraphs, k-partite, hypergraphs, temporal, etc.) each with additional structural constraints (directed vs symmetric, cyclic vs acyclic, reflexive or not, etc.) with implications on available algorithms and representative power/appropriateness. For reasons described in more detail below, a flat (aka simple, non-layered and non-multi) graph turns out to be the best solution for capturing and analyzing high-resolution spatial data with multimodal transportation networks when daily values are used. If more fine-grained schedule data is utilized then a time-layered (Masuda and Holme 2017) or temporally extruded (Bramson

Table 2 Summary of basic network features by transportation mode

Transportation mode	Node count	Directed edge count
Train station nodes	1546	—
Train access edges	—	10,358
Train transfer edges	—	35,670
Train routes	5179	10,536
Bus stop nodes	32,901	—
Bus access edges	—	186,172
Bus transfer edges	—	524,260
Bus routes	93,086	181,574
Road network	58,012	79,328
Hex nodes	244,721	1,459,096
Connecting links	—	434,892
Total	435,445	2,921,886

Train and bus networks are bi-directional, but the traversal times of access edges are asymmetric. The road network includes 23.1% one-way links, often representing each side of a divided two-road. The hex network as well as the intermode walking links are fully symmetrically bi-directional and Vandermarliere 2016) network representation becomes more efficient, but we do not yet have the schedule data necessary to support the time-dependent analysis.

Hex network

The hex network is formed by connecting each hexagon to its six (or less if it is on a border) neighbors on the grid. This creates a transportation network representing local travel, usually walking but may also represent cycling, local driving, or other slow means of transportation. As such, the 3 min time weights for these edges are based on a 5 kph (walking) speed across the 250 m interhex distance. For all walking links we use a capacity of 2500 people per hour, which we convert to 10,000 to represent the total value on the basis that the capacity's effect is spread out across the main commuting hours. The hex with the greatest job surplus (jobs minus population) is 15,665 and the greatest population surplus is 2024, and each hex has six neighbors, so these capacities are more than adequate to allow sufficient flow to reach all needed areas.

The hex network serves two main purposes: (1) some hex grid spaces are inaccessible directly via any other transportation network, so this ensures a connected graph (among adjacent hexes at least), and (2) in many cases using purely transportation links based on the closest station/stop/intersection leads to unnatural travel time differences. For example, for many people in the suburbs, the closest station is one that only has local trains. So it is often faster to walk (or ride a bicycle) to a slightly further station with express trains than to use the closest station.

Because the hexes permeate the whole geographic region, there are hexes in locations (such as mountain tops and rivers) where there is no population, there are no transportation routes, and even walking would be impossible. In the future we plan to remove hex network links representing impassible route (e.g., crossing rivers and train tracks directly, without using a bridge/gate). Considering the 250 m interhex distance, and the nature of our home-work transportation analysis, this particular abstraction is unlikely to be responsible for any inaccuracies.

Train network

Our train data combines all rail services (surface trains, subways, and streetcars; public and private) and schedule types (e.g. local, rapid, commuter, express). One natural representation of a rail network is to connect each line that stops at a given station to a single node representing that station (a space of stops Barthélemy (2011)). Because multiple lines, and multiple types of routes on the same line, connect to adjacent stations, capturing the station as a single node requires the use of either a layered graph (with each line on a different layer) or a multigraph (that allows multiple edges between pairs of nodes) (Goczyłla and Cielatkowski 1995). We originally used a multigraph construction; however, for our analysis the transfer times between trains/lines is crucial to the total travel times. Using a layered graph requires inter-layer links to connect the lines, and these are essentially train transfer edges. After our initial construction in this way, we realized the importance of not just inter-train transfer times, but also initial waiting and access times. In order to seamlessly integrate all these interstitial times into our network and algorithms we decided to include them directly as part of the train network.

To do this we first create a node representing each physical station in the system. Then for each type of train of each line that has a stop at that station, we create a node representing its platform. We then connect each physical station node to each line's platform node at that station via a *train access link*. We also directly connect all the platform nodes at a station to each other with *train transfer links*. Finally, we connect the platform nodes along each line using *route links*. We represent each schedule type (local, express, etc.) as a separate line. In this way, express trains that skip stations are captured by links directly connecting the stations actually used by that route. Express trains and local trains on the same tracks have distinct platform nodes in order to capture the transfer time of switching trains.

This set of abstractions produces a flat (non-layered, non-multi) rail system network. Even though there are multiple kinds of edges, the difference is in the interpretation/meaning of the model rather than in the function of the model. Certainly other formulations are possible; however, utilizing an intuitive physically explicit network allows us to keep *all* cost weights and *all* usage levels as edge values. This fosters simpler, faster, and clearer algorithms to process flows and measure network features.

For this project the transportation networks are modeled as directed. The route links are weighted by the traversal times for a link of that type on that segment using data from (Ekitan: Train and Bus Schedule Data 2019). In the case of daily average train route times, most edges are reciprocal with identical properties. However, we make use of the directedness to introduce an asymmetry in the traversal of the train access links. Exiting a station (the links from the platform nodes to its physical station node) takes 1 minute (the average walking time to exit the station), while the reverse direction (from the physical entrance to boarding a train) takes 5 minutes to account for both walking and waiting times[2]. Transfer links have a uniform (for now) time-weight of 5 min. A simplified example of the directed train network construction is shown in Fig. 3.

The access and transfer times approximately reflect walking and waiting times between trains and platforms without overly complicating our station specification (Hibino et al. 2005). Based on this structural foundation we can later incorporate heterogeneity in transfer and access times according to specific lines, locations, distances, time of day, congestion conditions, etc. if/when the data becomes available. Realistically, adding such details would likely only alter the total trip times by a couple of minutes for average daily traversal times, and so they are unlikely to make much of a difference for our identification of aggregate traffic patterns. However, if one wishes to make highly accurate predictions of specific traffic volumes, especially in a time-dependent analysis, then these details matter. The complete trail network for the Greater Tokyo Area includes 1546 stations, 131 distinct lines, and 10,536 directed links capturing all routes of all types. A map of Tokyo overlaid with the full train network can be seen in Fig. 4.

Tokyo trains are often overcrowded, reaching nearly 200% of their comfortable riding capacity during the morning rush (Ministry of Land 2018). The maximum number of passengers varies by the number of cars, the specific make of train, and the number of trains per hour, but only a couple of lines ever exceed 80,000 people per hour (in two-way traffic, although one direction can greatly outweigh the other at certain times of day, at least in the suburbs). For simplicity we use this figure as the basis for the 240,000 capacity of each train edge in the system (multiplying by four to include core commuting hours).

[2]In some cases trains run in 30-minute intervals, so one may expect the average waiting time to be 15 minutes at that station, but even in that case the time people typically wait at the station is only a few minutes because they plan ahead based on the extremely reliable train schedules in Japan.

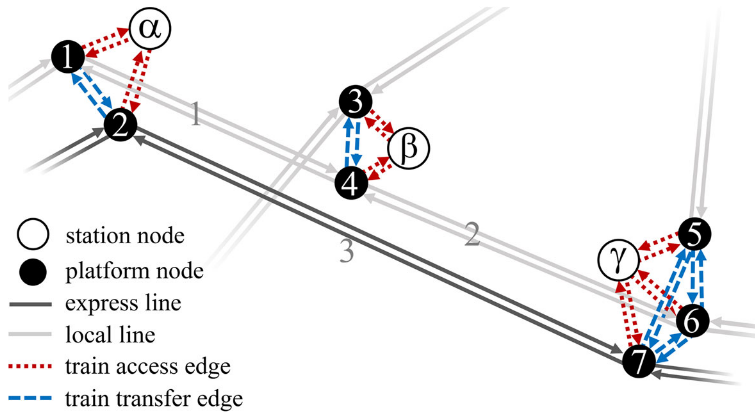

Fig. 3 Simplified example of the directed train network construction. Station nodes α, β, and γ represent the physical stations and dotted red links represent access links that connect the station to each platform. Each type of train on each train route is represented by a separate platform node (black nodes 1 − −7) which are connected to other lines at the same station by dashed blue transfer edges. Platforms are also connected to platforms of the same line at other stations by train route edges (gray). In this stylized example, lines 1 + 2 and 3 both use stations α and γ, but line 3 represents an express train that uses the same tracks without stopping at station β. Only the station nodes are connected to nodes of the other transportation networks. The bus network has a similar construction

Keep in mind that this is not intended as an estimate of the capacity of each line for predicting crowdedness, but an approximate figure to use as the maximum capacity for network flow algorithms.

Bus network

The bus network is constructed similarly to the train network so that physical bus stops (stations) are separate nodes from the stops along the lines (platforms). Again,

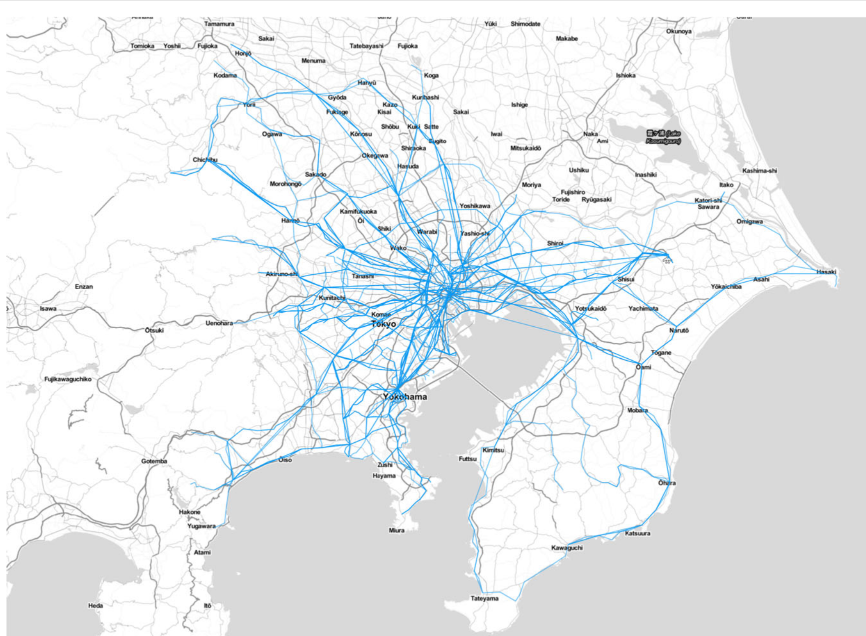

Fig. 4 A map of the Tokyo area showing the edges of the train network. Map tiles by Stamen Design (2019), map data by OpenStreetMap(OpenStreetMap Contributors 2019)

constructing the bus network in this way allows us to include waiting times and transfer times in our model as edge attributes, thus fostering the use of standard network flow algorithms. Traveling from the physical bus stop node to the bus line node (waiting to enter the bus) takes 5 min, while the opposite direction (exiting the bus) takes just 1 min. Transfers from one bus to another are assumed to take 5 min in average waiting and transit time. Just like the train network, the bus route (platform) nodes for express and local bus routes are distinct so that transferring from one to the other type of bus on the same line incurs the 5-min transfer cost. Traversal times are set from the bus schedules using the average traversal time for each link. This time does not include heterogeneity in traversal times as the result of time-dependent road congestion, loading and unloading times, differences in speeds from skipped stops, or other factors. Also, similarly for the trains, some buses only run at particular hours of the day, and at irregular frequencies, but these details are not included here for daily traffic. As a result, some supposed morning commuters may be simulated riding a bus that only runs once a day after midnight (incorporating schedule data can alleviate this issue).

There are 5084 unique bus lines using 32,901 physical bus stops (stations) and 181,574 directed links included in our data for the Tokyo area. The local bus stops are typically dense along the route, enabling one to make out the contours of the road from the linear connections between sequential bus stops (see Fig. 5). Express buses (ones that run the same paths as local buses but skip several stops or directly connect two distance points, such as airport shuttles) are modeled as edges directly connecting the bus stops where the bus actually stops. These are easily identifiable in Fig. 5 connecting to Narita and Haneda Airport from various stations around the Tokyo Area as well as several commuter buses that link Tokyo and other major stations together.

The number of buses running per hour exhibits significant variation both across lines/locations and for the same line at different times per day. Some buses run less than

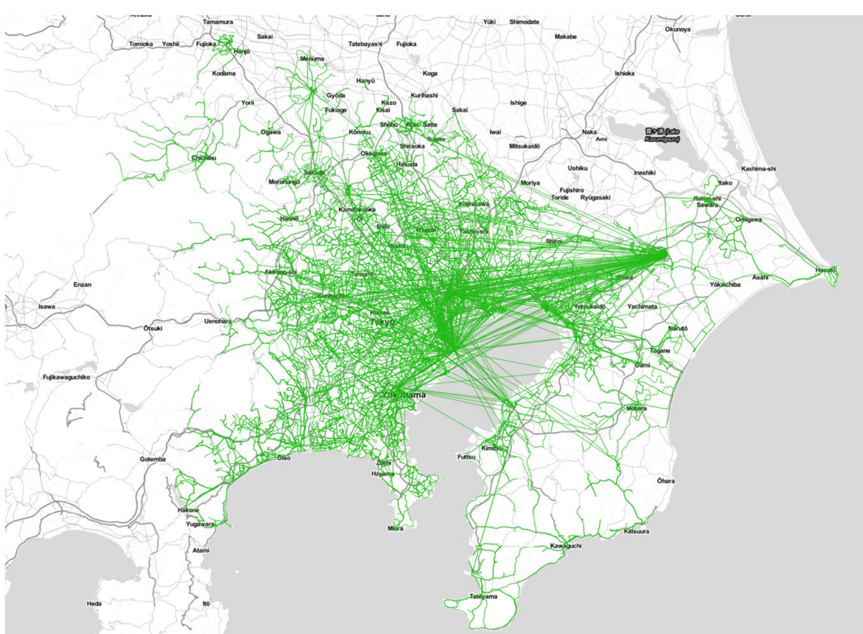

Fig. 5 A map of the Tokyo area showing the edges of the bus network. Map tiles by Stamen Design(Stamen Design 2019), map data by OpenStreetMap(OpenStreetMap Contributors 2019)

once per hour while others have 13 buses per hour at their peak times. However, our current data does not include this information on a per-route basis. Because the capacity figure is used to cap the traffic along a link, and increasing bus traffic along a link to accommodate need is easy, we use the 10 buses per hour value for setting capacity limits. Each large bus has a capacity of 74–78 people (sitting plus standing)(Toei Transportation Customer Center 2018), although some buses are smaller we use a value of 75 throughout. We again multiple the per-hour capacity by four to capture the maximum load for the core commuting time in each direction. We hope to get more refined line-specific data for future analyses, but as a result of the above approximations we currently assign a capacity of $75 \times 10 \times 4 = 3000$ people for each edge in the bus network. Again. rather than an estimate, this should be thought of as a theoretical maximum capacity.

Road network

Our road network is constructed from road segments tagged as tertiary or above (or not specifically labeled and left as 'road') in OpenStreetMaps data (OpenStreetMap Contributors 2019). OpenStreetMap labeling is sparse in Japan compared to other developed countries, and we had to make assumptions based on typical values to fill in missing road speed limits (Japan Traffic Safety Association: Rules of the Road 2017), typical drive speeds, numbers of lanes, and road capacity. For approximate drive speeds we adopted a convention of 70 kph for major highways, 30 kph for other major roads, and 25 kph for minor roads. These drive speeds have been reduced from the speed limits to account for traffic congestion, railway crossings, turning, traffic signals, etc. Edge traversal times are calculated from the Haversine distance between its end points and the approximated driving speeds.

The number of lanes is rarely specified in the data, so to approximate capacity we again needed to make assumptions based on typical values while recognizing that the actual road system exhibits heterogeneity within each category. Because our network model is everywhere directed, two-way road segments are modeled as bi-directional with the one-way values in each direction. To calculate the capacity we used standard vehicles-per-hour rates for these road types and speeds (Polus et al. 1991; Highway Capacity Manual 2000): 2000 vehicles per hour per lane for freeways, 1000 for major arteries, and 500 for the remaining roads. As before, we multiply these per-hour rates by four in order to generate the commuting capacity values actually used as edge capacity values for the network simplex algorithm. A table showing the values of each feature used for each road type is available in the Additional file 1.

In Japan it is common for roads to frequently change their thickness and speeds, thus complicating classification efforts. Perhaps for this reason when we included only the tertiary and above links, the road network was too sparse and disconnected. The addition of the unclassified (rather than smaller) roads largely corrected this problem. Although there are still several disconnected segments when leaving out the minor roads; this accurately reflects the infrastructure of Japanese roads. The fragmentation is exacerbated because the road network data used for the current analysis leaves out many of the main expressways. Despite this deficiency, the intermediate road network (between local access roads and major expressways) is an integral part of the transportation network that connect areas otherwise unreachable (except by walking). Figure 6 shows the detail and extent of the included road network. Areas between roads are accessible via

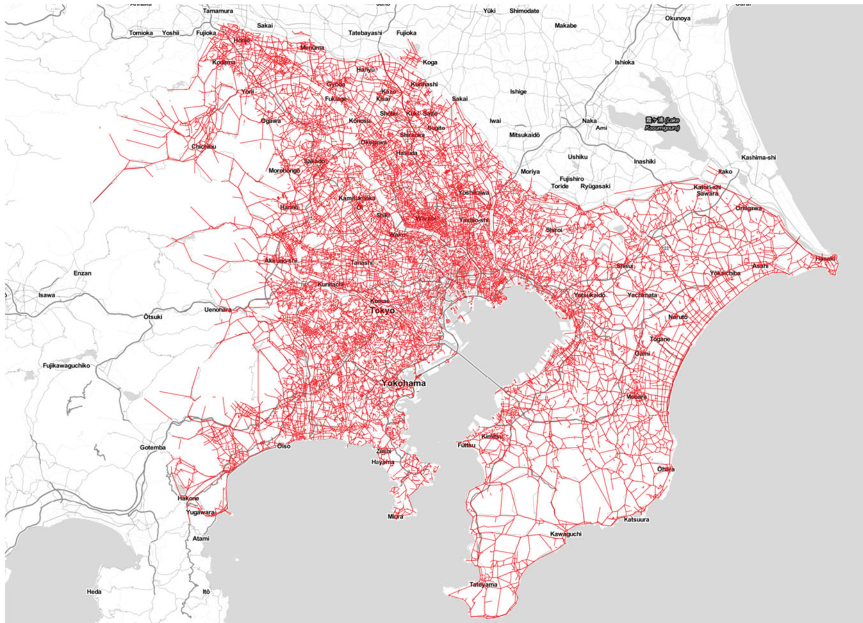

Fig. 6 A map of the Tokyo area showing the edges of the road network. Note that several large highways and roads labeled as smaller than tertiary have been excluded from this analysis. Map tiles by Stamen Design (2019), map data by OpenStreetMap(OpenStreetMap Contributors 2019)

the hex network, although most of the large areas without roads are unpopulated mountain and forest regions. In addition to road links, each intersection node is connected to any other intersection node within 173.3m via a *walking link*. Although these non-road links between road nodes use the walking speed of 5kph, and this is perhaps an underestimated speed even for narrow Japanese residential roads, in this case the links and speed are meant to represent all local movements, including: travel to and from parking, congestion, waiting for pedestrians, and various other factors – and typically only for short distances. We tend to think of the road network constructed in this way as the "taxi network" in consideration of the included links, speeds, and revealed role in Tokyo transportation.

Connecting (Walking) links

The four networks described above represent the geography and three transportation systems; we use additional connecting links in order to integrate these disparate modes together. One can think of the individual transportation modes as layers, and these connecting links as inter-layer edges (Idri et al. 2017). Or you can think of the whole system as a physically explicit flat network. These two conceptualizations are methodologically identical for the purposes of edge traversal time-costs, especially because the sets of nodes for each mode are unique. In our representation a bus stop and taxi stand in front of a station are represented as distinct nodes from the station itself. We use the actual geographic coordinates for those locations to determine the distances, and from those distances the time costs of intermode transfer. Although such an explicit network increases the number of nodes and links needed compared to a more abstract representation that collapses these distinct nearby nodes into a single node (e.g. using a multigraph), it benefits from representational simplicity, faster algorithm speeds, and the ability to intuitively and directly conduct hypothetical conditional response analyses (e.g., link knockout to represent accidents, construction, etc.)

We connect transportation nodes to hex nodes and to each other (of the same or different mode) whenever they are within 173.3 m (the hex circumcircle radius is 144.34 m) of each other. This method ensures that every transportation node is connected to at least one hex (and often two or three). For the train and bus modes, only the physical station nodes are thus connected; i.e., one must go through a station node to reach its platform nodes. Connecting the modes in this way provides an intuitive geographic foundation for the whole system, and (when analyzed together) the nodes of other transportation networks are directly connected to each other to represent direct intermode transfers (e.g., from a train station to a bus stop or taxi).

Like the interhex walking links, the time-weight on the intermode links are calculated from the Haversine distance between the endpoints and an average walking speed of 5 kph. The 5 kph walking speed is meant to accommodate various common factors for which we do not have data: congestion, stairs, non-direct route, etc. Although this *walking network* is an abstraction of the physical connections among modes, it fosters a clear way to approximate the intermode transfer times even when walking is not the only factor involved. More refined data can help us improve the estimates of the intermode transfer times on a per-link basis. For walking links, like the hex links, we use a capacity of 2500 people per hour, which we convert to 10,000 to represent the total value on the basis that the capacity's effect is spread out across the main commuting hours.

Network structure

Table 2 reports the numbers of nodes and edges for each mode as well as the fully integrated network. Aside from our physically explicit flat graph, there are alternative methods for integrating disparate modes of transportation and capturing the transfers among them. As eloquently stated by Idri et al. (2017), "The representation of this transfer action is treated as one of the key issues in multimodal route planning problem." One common solution is to generate a "transfer graph" that virtualizes the connections between modes to facilitate interoperability (Ayed et al. 2011). However, because we are not including schedules or other time-dependencies in the intermode (or intramode) transfers, and because our transportation networks are geographically grounded, we solve the intermode transition problem by (simply) adding walking links between nodes of different modes (and sometimes nodes of the same mode) to explicitly model the act of transitioning between them and "glue" the networks together. This simplification of a transfer graph yields an intuitive representation of a multimodal network as a single, simple directed network.

Even though our network is of the common digraph variety, our geographically grounded construction alters the calculation and/or interpretation of network structural measures used in standard transportation network models. For example, the degree of a station node is no longer the number of adjacent stations, but is now the number of line types with stops at that station plus the number of hexes, bus stops, and road intersections within 173.3m. Although the node degree (considering all links or just a subset) may still be related to other graph features, it cannot be interpreted in the same way as previous analyses (Barthélemy 2011; Levinson 2012; Tsekeris and Geroliminis 2013).

As another example, the path length (in terms of edges used) is a less clear indicator of anything useful because there is so much heterogeneity in the physical lengths of edges by mode. We return to this point later, but the crux is that only network measures that can incorporate an edge weight/cost are useful and appropriate for multimodal transportation network analysis. Measures such as the betweenness centrality and anything requiring the average path length are computationally impractical; that is, they are impossible to compute in practice. The reason is not strictly the number of nodes, but also the number of equicost paths along the hex network part of the graph among so many pairs of nodes. One alternative approach to the hex network is to use a Voronoi grid based on the nodes (Ren et al. 2014), but the hex network is superior for providing high-resolution integrated analyses because it (for example) integrates the walking/bus/driving time or the use of other modes to/from stations. Furthermore, having a homogeneous grid size/shape is important when tracing the network flow results back to the location attributes. Rather than lament the analyses and comparisons we cannot perform because of our network construction, we now move ahead to explore the new analytical opportunities that this representations creates.

Methods

We utilize two techniques to uncover the flows from population providers to employment centers: the *network simplex algorithm* and *stochastically generated trips* between population and job probability distributions. Due to the size of our networks and the population, both techniques required down-sampling the data as explained below. Of course there are many other methods for generating shortest-path flows across networks in general (Ahuja et al. 1993; Gallo and Pallottino 1988), and multimodal transportation networks in particular (Lozano and Storchi 2001; Cipriani et al. 2006; Wang et al. 2009; Verga et al. 2018). However, for our purpose of uncovering large-scale aggregate usage patterns we opted not to use a complicated route-planning technique. Without access to empirical origin-destination (O-D) trip data for the Tokyo area, we need to make approximations for the start and end points. And with a network too large for betweenness and load calculations to be practical, we need ways to approximate traffic flow. Here we describe our methods for generating usage patterns across the Tokyo area transportation networks.

Population and job sampling

We use distinct samples for the network simplex and stochastically generated trip flow methods, but both are based on the same job and population distributions (Figs. 1 and 2). For each hex in the system, we add its ID number to a job list once for each job located there, and similarly for the population data. The job list has 16,614,941 hex numbers and the population list has 22,743,338; a uniformly random pick from these lists returns a hex with a probability equal to its proportion of the jobs/population.

Gravity and radiation models

Perhaps the most well-known model of mobility and migration is the gravity model (Zipf 1946; Erlander and Stewart 1990; Barthélemy 2011). Despite its previous popularity, now it is heavily criticized for its need to calibrate parameters with historical flow data, various analytical deficiencies, and inabilities to match empirical patterns. It is incrementally improved upon by the intervening opportunity model (Nazem et al. 2013) and random utility model (Domencich and McFadden 1975), but these also do not reliably generate accurate predictions. The radiation model of mobility (Simini et al. 2012) was developed specifically to overcome the analytical and accuracy problems of these previous methods (Masucci et al. 2013; Kang et al. 2015; Stefanouli and Polyzos 2017). It is based on first principles to provide a theoretical underpinning behind its formulation and has been shown to produce higher-accuracy prediction of mobility/migration patterns in some specific applications. We focus on the radiation model because it supersedes the previous members of this family.

One key assumption of this family of models makes it unsuitable for the current application. These methods derive the distribution of opportunities from the population distribution, and assumes they are highly correlated. This assumption is appropriate when the unit of analysis is metropolitan areas or larger, and the movement of people is between these largely self-contained areas. However, in our case the unit of analysis is 250m wide hexes within a single metropolitan area. And within this area the distribution of jobs is known (so we don't need assumptions for the location of opportunities), and (contrary to the assumptions of these models) within cities the distribution of opportunities is not proportional or correlated to the population. Actually, the population and numbers of jobs in hexes are anti-correlated: the hexes with the greatest numbers of jobs (around most major stations) have zero population and the highly populated ring around the city center has few jobs.

The original radiation model predicts flow between locations based on geographic distance, so it needs to be extended to incorporate networks flows. One version of including transportation networks (specifically the US highway network) first isolates the nodes reachable within (for example) 400 minutes via the network, determines the amount of usage to each node in that subgraph based on the radiation equation, then adds that usage to the edges along the shortest path to each of nodes within the subgraph (Ren et al. 2014). In our case, the number and location of jobs is known, so we would modify the radiation equation to generate usage volume based on the working population of hex a, the number of jobs in hex b, and the number of jobs reachable within the same time it takes to reach hex b (t_{ab}). This adaptation would provide an estimate of the demand for trips between all pairs of points; however, even ignoring the dubious theoretical claims such a model would be making about intracity mobility, this approach is infeasible for our application.

Specifically, (Ren et al. 2014) makes the operation tractable by processing the origin-destination pairs only among the dozens of nodes within reach of each focal node. However, in our case nearly the entire network is within a feasible commuting range from every location so we can't localize the calculations, and our network has nine times as many edges (in its undirected version). However, given the distributions of workers and jobs, and given that the radiation model assumes people take the closest job to their home (above a randomly assigned job quality variable), the assumptions match well to another algorithm that *is* tractable: the network simplex algorithm.

Network simplex algorithm

The simplex algorithm is commonly used in transportation network research (Dantzig 1951; Cunningham 1976; Kim and Rilett 2003) to find (for example) optimal distribution patterns from suppliers to customers. In our application the optimization aspect carries an implicit assumption that all people take the job closest available to their home.

Radiation models mitigate this by adding a random quality attribute to each job, and a random quality demand for people, and then requiring that the job be of greater quality than demanded. This essentially creates noise in the system so that some people will take further jobs in order to satisfy their quality demands. The simplex model does not include quality, but it does constrain the number of people who can take a job (which gravity and radiation modeling do not account for). Specifically, each job is filled by exactly one person; jobs nearby population centers are filled first, then the remaining population flows over to further and further jobs.

Despite this unrealistic assumption it is a powerful abstraction that allows us to simulate the flows between population centers (supply) and employment centers (demand). One can even conceptualize it similarly to the radiation model: people are ejected from their homes and travel along the network until they are absorbed by a job. By creating quality tiers of people and jobs, and running the algorithm separately on the different tiers, one could also emulate the effect of random quality in the radiation model. We do not pursue that extension here, but we are investigating whether the simplex algorithm can be applied to replicate network radiation models in significantly less computational time.

The network simplex algorithm finds a solution to the problem of satisfying the system's demand (jobs) with the system's supply (negative demand, population) in terms of edge flows. We used the implementation from the NetworkX Python library (Hagberg et al. 2008) which requires the network to be directed, the nodes to have a demand level, and the edges to have both a flow cost and flow capacity. The system-wide net demand must equal zero for a solution to exist, so we (initially) adjusted the population downward to match the number of jobs. Only hexes nodes hold population of job data, so all transportation nodes have a demand and supply of zero. We use the traversal time as the cost for all edges, and capacity values as explained for each dataset above. Figure 7 shows the net demand (number of jobs minus the working-age population) of each hex.

NetworkX's simplex algorithm implementation was unable to find a solution using the adjusted population and all the job data, so we downsampled the system to make the problem tractable. We created a sample of one million people using the probability distribution described above by selecting hexes with replacement one million times and adding one to that hex's population sample. We did the same thing for jobs. In this way the number of people is exactly equal to the number of jobs (one million) and the algorithm became tractable.

Stochastically generated trips

The standard method for generating estimated traffic patterns in city-level transportation network simulations is sampling from an origin-destination (O-D) model. Creating an O-D model requires defining the choice sets for both the origins and the destinations of trips. These O-D models are often generated from commuter survey data, typically provided by governments at various levels (U.S. Department of Commerce 2010). Although the Japanese government also conducts such surveys (Metropolitan Transportation Census and Person Trip Survey), only highly aggregated summary data is made available, and the actual O-D pairs are unknown (Ministry of Land 2010a; Ministry of Land 2010b; Kawasaki 2015; Parady et al. 2018). Another approach utilizes mobile phone, car navigation, or other GPS data to monitor human movements and estimate both O-D pairs as well as traffic volume (Ohmori et al. 2000; Gonzalez et al. 2008; Uno et al. 2009; Liu

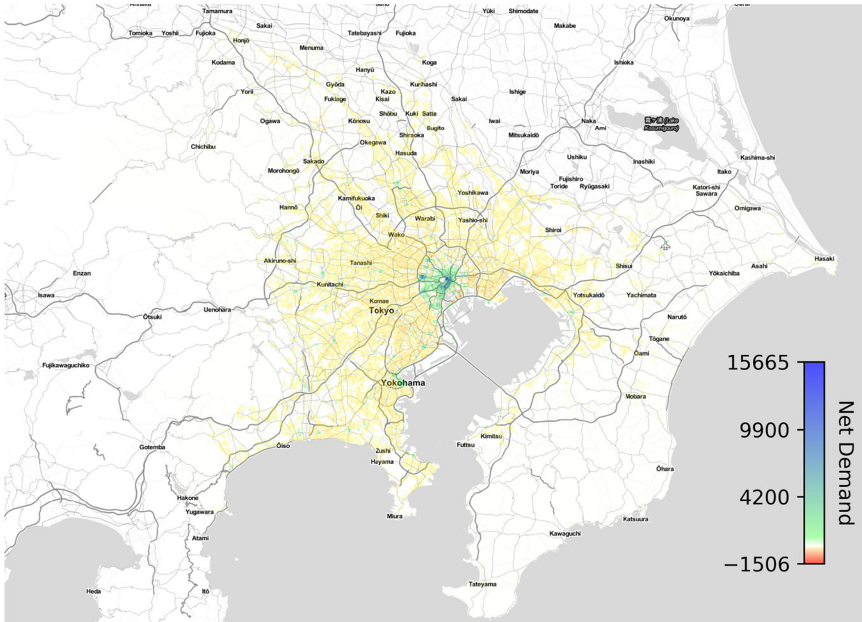

Fig. 7 A map of the Tokyo area showing the net demand for workers: number of jobs minus the working population for each hex. Map tiles by Stamen Design (2019), map data by OpenStreetMap(OpenStreetMap Contributors 2019).

et al. 2012; Bachir et al. 2019). However, due to strict privacy laws in Japan this data is not available to third parties, even in summary. The use of automated fare collection systems for public transportation opens the potential for O-D data limited to the train or bus systems (Myojo 2006; Li et al. 2011; Sels et al. 2011; Munizaga et al. 2014), but access to the anonymized data is strictly controlled, very expensive, and fragmented into various rail/bus companies. No matter how the choice set models are created and calibrated, to make predictions of future and/or contingent traffic flow, O-D trips are stochastically generated from the models (Romanos and Saidane 1978; Rieser-Schüssler et al. 2013; Huang and Levinson 2015). We generate O-D trips using (for now) the population and job distributions as the origin and destination choice sets respectively.

The population density pattern in Tokyo is noticeably different than other major metropolitan areas. The population density is low near the core, sharply increases towards the ring of suburban ("bed-town") areas at 6–10km from the center, and then rapidly declines going further out (recall Fig. 1). In contrast, the population densities of New York, London, and Paris are highest near the center, and gradually taper off with increasing distance (Hatta and Ohkawara 1993). The efficient commuter train system, commuting fare reimbursements by employers, lower land prices, and a desire for "wide open spaces" all contribute to Tokyo's particular population distribution pattern. Many factors contribute to any individual's choice of housing and employment, but being close to one's employment is not typically a heavily weighted consideration among Japanese office workers. This is reflected in the fact that the average commuting time in the Greater Tokyo Area is 51 min, with one-way commuting times greater than 90 min accounting for 16.1% (comparable to the only 17.0% less than 40 min) (NHK Culture Research Institute 2015).

For our O-D model we choose 300,000 home-work pairs of hexes by pulling independently from the population and job distributions. First a hex chosen from the population distribution, and then a hex is chosen from the distribution of jobs. We impose three constraints on the pairs: (1) the two hexes are distinct, (2) the two hexes are located within 50 km of each other (by Haversine distance), and (3) a path exists between the two hexes[3]. Actual home-work pairs are presumably not fully random, but certainly there is a high degree of stochasticity and heterogeneity in the choice of both sides of that pair. Future work includes developing a more sophisticated O-D pair model that replaces the above constraints with contingent probability distributions based on a number of factors such as income, job quality, mode of transportation, etc.

The route taken to and from work is determined using the NetworkX single-source Dijkstra path algorithm(Hagberg et al. 2008). In order to improve performance we only record a single shortest connecting the origin and destination path rather than attribute proportional flow based on all (equal-time) shortest paths. That said, generating 300,000 shortest-path trips took more than six days; by comparison, the simplex algorithm using one million samples completes in 47 min.

As a trade-off for being more computationally intensive than the simplex algorithm, this method comes with a variety of benefits. The stochastically generated trip method does not require the network to be directed, nor the population and jobs to be balanced, does not assume people take the closest job to their home, and can more easily be generalized to other kinds of transportation uses (e.g. travel for school, shopping, and recreation). It also allows us to collect and analyze data on specific trips in addition to aggregate patterns. Specifically we collect data on the number of times and distance traveled across the four main modes of transportation to analyze on a per-trip basis. Using this data we can explore relationships between features such as which mode is more common for shorter vs longer trips. We now discuss the differences between the outcomes of the two methods in depth.

Results and conclusions

Here we present the outcome of analyzing the multimodal transportation network of the Greater Tokyo Area using the above methods. Because our analysis is mainly exploratory, we make efforts to thoroughly describe the revealed features of the transportation system usage patterns. Although our goal here is not to perform traffic micro-simulations and recreate accurate traffic flow, we do perform a limited comparison of revealed usage with empirical data to check the face validity of our methods. We then cover some implications and interpretations of these results in the Conclusions section below.

Edge Usage Frequency

First, we look at the edge-use frequencies presented in Fig. 8 and Table 3 (in percentages, the usage counts are available in the Additional file 1). Although edge usage is less informative for mode usage patterns than the time spent (Table 4) and distance traveled (Table 5), it is more informative for geographic patterns

[3] Our network is the largest weakly connected component of the multimodal network (to eliminate unreachable islands), so condition (3) is only necessary to filter trips to islands for which the reverse trip uses a road that extends outside the covered region (e.g., a neighboring prefecture) or for which the OpenStreetMap data was in error. Out of 300,000 trips, only one was irreversible.

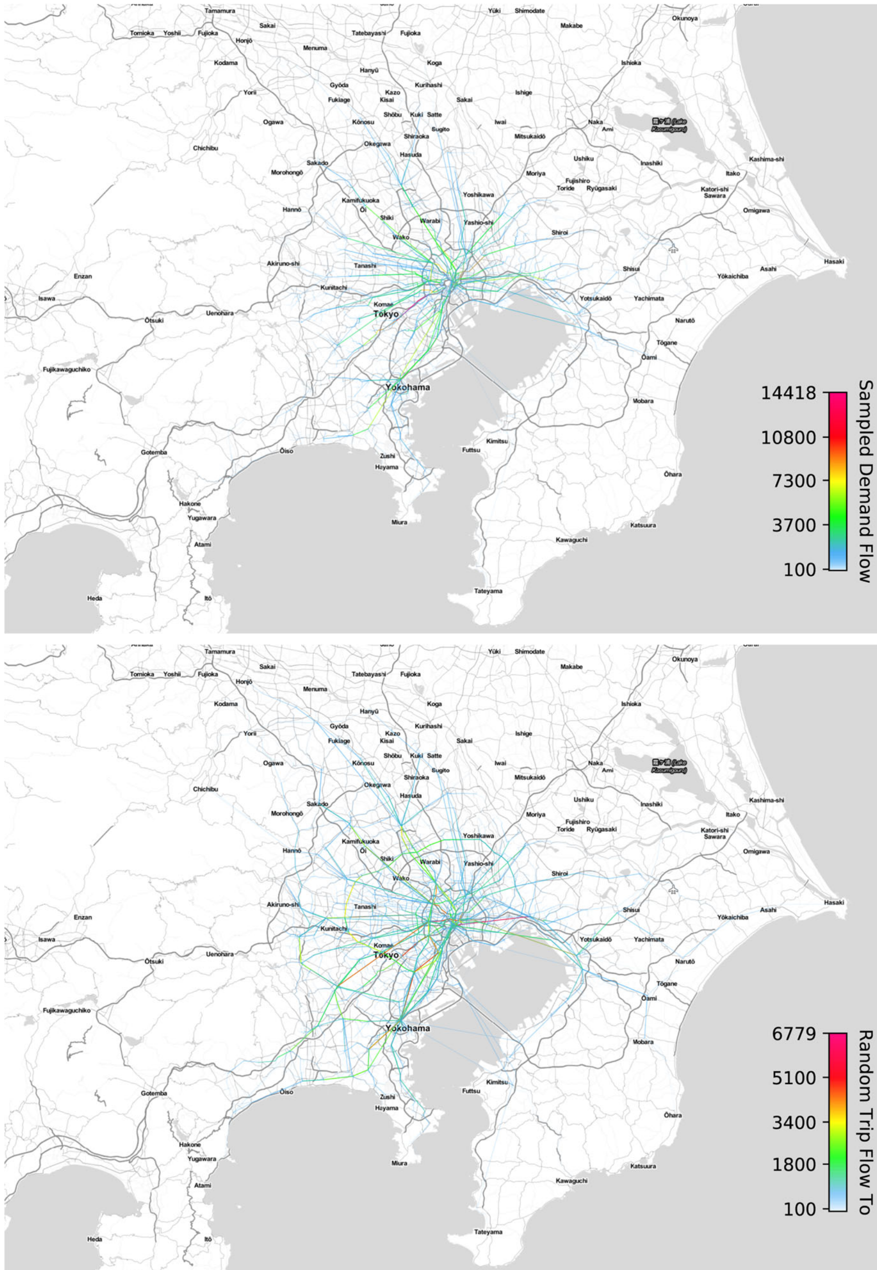

Fig. 8 A map of the Tokyo area showing the flow solution found by the simplex algorithm using a sample of one million people/jobs (top) and the edge-use frequency among 300,000 generated trips to-work (bottom – reverse trip map available in the Additional file 1). To focus attention on the heavily used corridors, we filtered edges with fewer than 100 trips in these diagrams (but not in our analysis). Note that edges going in each direction for each pair of nodes overlap, and even at 0.3 opacity some of the difference in color may be due to the z-order. Map tiles by Stamen Design (2019), map data by OpenStreetMap(OpenStreetMap Contributors 2019).

and certain usage statistics such as transfers and access links. Note the extremely high degree of similarity between the to-work and from-work data columns. There are differences, and we explore them later, but in practice we can consider them as equivalent.

Table 3 Percents of the total usage frequency of links by modality for the three generated flow datasets

Transportation mode	Simplex algorithm 1,000,000 demand	Stochastically Generated trips 300,000 to work	Stochastically Generated trips 300,000 from work
Hex	10.225	5.002	5.003
Walk	27.695	22.772	22.756
Train	17.308	20.248	20.269
Train transfer	0.789	2.378	2.373
Train access	6.087	5.939	5.940
Bus	1.000	0.705	0.705
Bus access	0.528	0.331	0.331
Bus transfer	0.007	0.019	0.019
Road	36.359	42.605	42.604

The stochastically generated trip results are also quite similar to the results from the simplex algorithm, but here the usage differences reveal important differences between the two methods. For example, the simplex algorithm used 480 (0.007%) bus transfer links, while the generated trips generated 1744 (0.019%). For both systems the bus transfer usage is the lowest link type, but generated trips are nearly three time more likely to use a bus transfer. The simplex algorithm also utilizes twice as many hex links and 5% more connecting links (both are more useful for short-range travel). To compensate, the stochastically generated trips use 3% more train links and 6% more road links. In terms of sheer numbers of links, the road network dominates at around 40%, but as we will see this does not translate into high flow because the average link length for the road network is short.

Mode usage patterns

According to the *Nationwide Urban Traffic Characteristics Survey*, using trip data aggregated across Japan's three largest cities, 26% of weekday trips used a train as the main travel mode, 2.7% used bus, 33% used a car, 16.8% a bicycle, and 21.5% used walking or other means (Ministry of Land 2010b). Because this only reflects the main mode of transport (and not all modes of transport along the way), and combines all travel purposes (not isolated to trip between home and work), the data is not an ideal comparison for our

Table 4 Percent of the total time used by links of each modality for our three generated flow datasets

Transportation mode	Simplex algorithm 1,000,000 demand	Stochastically Generated trips 300,000 to work	Stochastically Generated trips 300,000 from work
Hex	15.89	7.28	7.28
Connecting	13.17	10.26	10.26
Train	40.72	47.78	47.79
Train transfer	2.04	5.77	5.76
Train access	9.46	8.65	8.65
Bus	1.87	1.55	1.55
Bus access	0.82	0.48	0.48
Bus transfer	0.02	0.05	0.05
Road	16.01	18.20	18.20

Table 5 Percent of the total distance comprised of by links of each modality for our three generated flow datasets

Transportation mode	Simplex algorithm 1,000,000 demand	Stochastically Generated trips 300,000 to work	Stochastically Generated trips 300,000 from work
Hex	2.78	1.11	1.11
Connecting	2.30	1.57	1.57
Train	74.62	78.04	78.04
Train transfer	0.07	0.18	0.17
Train access	0.55	0.44	0.44
Bus	2.79	1.94	1.94
Bus access	0.05	0.02	0.02
Bus transfer	0.00	0.00	0.00
Road	16.84	16.70	16.70

results or our purposes. However, as previously mentioned, data availability in Japan is extremely limited compared to other developed countries (Kawasaki 2015), so we have to make do with what we have.

We can evaluate the utility of a link or mode using the time spent and distance traveled across that link or mode. One reason that this is better than edge-use counts/percents is that in a multimodal network the heterogeneity in link distance, speed, and connectivity is much greater compared to a single mode. Table 4 shows the percent of total travel time used by each transportation mode while Table 5 does the same for distances.

Based on the ratios of the total time and total distance used, we can make some inferences about the average characteristics of the modes. Hex and connecting links, as well as access and transfer links, are slow yet necessary means to connect the more efficient modes of transportation; hence they make up a small percentage for most trips, but are part of every trip. The main result here is that train+road make up nearly 66% percent of time but more than 94% of the traversed distance, with other (slower) links filling in the cracks to provide trans-modal access and access to local locations. Buses make up a surprisingly low percent of time and distance. Although the generated rates match the survey data reporting only 2.7% of commuting trips use a bus as the main mode, we expect to see buses used in a supporting role (e.g., from home to the station). We return to this point below.

The differences between Tables 3, 4 and 5 reveal how mode heterogeneity affects the results. For example, although road travel makes up 36% or 43% of the edges for simplex and generated trips respectively, it only accounts for 16% and 18% of the time used (similar values for distance). Hex and connecting links are all short ($\leq$ 250m) and traversed at walking speeds, so they comprise 38% of the links, but 29% of time and 5% of distance for the simplex flow. For generated trip flow these walking links go from 28% of the edges to 18% of time and 3% of distance. Overall these results underscore the importance of temporal and physical edge-weights when analyzing geographically embedded transportation networks. Models that abstract away these details to more simply measure properties of the network structure are missing this key aspect of transportation network analysis.

Bus transfers

As mentioned earlier, the bus transfer links are the least used for both methods. One reason for this is the overall low level of bus usage (1% of links for simplex, 0.7% for generated trips), but even considering that the transfers are underrepresented. Bus access links were used 35,280 and 30,900 times by the simplex and generated trip methods respectively, implying 17,640 and 15,450 bus rides. This means that only 2.72% of bus rides involved transfers for simplex and 11.29% for generated trips. In fact, bus transfers are rather uncommon in the Tokyo transit system because buses are typically used to connect either remote areas to their nearest train station or to connect distant train stations/location together via express buses. Thus most bus transfers would be occurring at train stations, and from those points using a train is typically faster. Although we do not yet have hard data on bus transfer rates to make a quantitative comparison (only reports and anecdotal evidence), we consider this low rate of bus transfers a mild validation point for the methods used here.

Bus underutilization

Although low bus transfer percentages were expected, we expected to see higher bus usage rates overall. Buses are underrepresented by both methods, and for the same reason. Buses utilize the road network but come with a loading and unloading time cost. So, because our current analysis only looks at time efficiency, using the road network directly is going to be faster than using the bus network for the same path. The two exceptions to this are (1) where the included road network is too sparse because the roads are smaller than tertiary and (2) express buses that run on the expressways that are also excluded from our current road network. Of course not everybody has their own car to use (fewer than half of Tokyo households own a vehicle Automobile Inspection & Registration Information Association (2017)), and the main reason to use a bus rather than taxi is cost rather than time. Future work that includes the financial costs in addition to time costs will likely generate more realistic bus usage patterns. However, the multimodal network analysis confirms what we already know: if money and parking are no issue, then driving is almost always preferred to taking the bus.

Commuting trip duration and mode

As mentioned earlier, one of the benefits of the generated trips method is that it provides rich data regarding the modes used along a route as well as total travel distances and times. The mean total traversal time for our generated trips is 64 min (the max is 223 and the min is 1.95 min). According to a 2015 NHK survey the mean one-way commuting time for people working in the Greater Tokyo area is 51 min (NHK Culture Research Institute 2015). Although this is somewhat lower than we find, the modal commute time (in 10 min bins) is 60 to 69 min (20.6%), so the generated trip method produces reasonable trips in this regard as well. Our mean value may be higher due to impractical home-work pairs, or perhaps because the survey data included non-office type jobs; people working at (for example) shops and restaurants are more likely to work closer to home. From the same survey (NHK Culture Research Institute 2015), 85.2% of office workers reported that they commute by train. This compares favorably with our finding that 88.5% of generated trips contained a train component. If we can get more fine-grained results from this and other surveys regarding empirical commuting mode usage, time, distance, etc. we can perform a more rigorous comparison of the distributions to validate the results and calibrate more sophisticated trip generating functions.

Link usage determination factors

Here we investigate whether particular link features make them more likely to be traversed in either the simplex or generated trip experiments. Clearly the main determining factor is having a location between population and employment centers, but beyond that it seems reasonable that links which cover more distance in less time would be preferentially used. Table 6 shows the Pearson correlation of edge usage frequency with speed, distance, and traversal time for the whole network and three main modes of transportation (i.e., excluding connecting and hex links, the latter of which are all identical in distance and time).

Intuitively, trains are used because they are a high-speed and efficient means of travel, which means we should expect express trains to dominate the train usage and local trains to be far less common. Considering the whole transportation network, both faster and longer links are indeed preferentially used by both methods. For the generated trips this applies to trains (0.163), but for the simplex method, faster trains are not correlated with frequency (0.051) even though longer length links (which are mostly express trains) are somewhat preferentially used (0.127).

Presumably, longer bus rides are not much faster than shorter bus rides because they are constrained by traffic conditions and the speed limit, so although longer (express) bus rides are preferentially used by simplex (0.251) and generated trips (0.299) methods, the speed is uncorrelated. As a result, longer bus travel times are also correlated for both methods. Although no road features are appreciably correlated with usage, both distance and traversal time have small and negative correlations, indicating that roads are more often used for shorter trips (e.g., riding from home to the station or driving directly where using trains would be too circuitous).

From Fig. 8 of both generated flows, it appears that links flowing into the core have the greatest weights; however, we find that among the links actually used, the links within the core itself are not as heavily used as intermediately located routes. Specifically, the correlation between an edges' center's distance to Tokyo station is slightly negatively correlated with its usage frequency (simplex: -0.175; generated trips: -0.12). One explanation is that destinations within the core are varied, and the network is most dense in that area, thus dissipating the traffic across many edges. In contrast, the commuter routes flowing to and from the core are sparser, and thus focus the usage on a few major routes. Although it is certainly true that the highest traffic links are these commuter links, we find that the pattern is more general. Periphery links (beyond the suburbs) are the most common and least used, but the suburban trains sufficiently pull the heavy ridership away from downtown.

Table 6 Characteristics of the trip data related to mode usage, total distance, and circuity

Mode	Method	Usage and speed	Usage and distance	Usage and traversal time
All	Simplex	0.211	0.208	-0.044
All	Generated trips	0.254	0.201	0.009
Train	Simplex	0.051	0.127	0.116
Train	Generated trips	0.163	0.180	0.155
Bus	Simplex	0.064	0.251	0.221
Bus	Generated trips	0.093	0.299	0.238
Road	Simplex	0.001	-0.049	-0.050
Road	Generated trips	0.004	-0.069	-0.070

Trip characteristics

Further analyzing the generated trip data reveals/confirms a few additional usage patterns. Table 7 summarizes a few key relationships of interest. For the four main modes of transportation (excluding access and connecting links) we can see the difference between the mean edge length in the full network compared to the mean edge length used in generated trips. Naturally the hex links are unchanged because all hex links have a length of 250m; however the train and bus links are longer and the road links are shorter. Moreover, the train links being nearly double the average length reveals that express trains (with link lengths 2 to 5 times longer than local trains) are used, but not by a preponderance of trips. Inner city trains often do not have express lines and this is where most of the train riding occurs.

The used bus links are 15x longer than the average of bus links, indicating that buses are almost exclusively used for express services. As already discussed, for shorter distances the road network is faster, and sure enough we find that the roads are being used for local access because the used road links are even shorter than the average road links. Despite the concentration on long-distance links, bus usage makes up less of the total distance (1.5%) than even hex links (1.6%). The use of hex and road links is inversely correlated with the total trip length, while train usage is positively correlated.

As a measure of how efficient the route is we measure the circuity of each trip as the ratio of traveled distance to Euclidean distance (Huang and Levinson 2015; Lee et al. 2015). We find the trips to work to have a maximum circuity of 6.82 (around Tokyo Bay), a mean of 1.28, and a minimum value of 1 (there exists at least one trip between neighboring hexes). Greater train usage is negatively correlated with route inefficiency (−0.214). Although trains are fast, one has to go out of one's way to and from stations to use them, which intuitively leads to greater circuity. Rather than indicating that train trips are less circuitous, what this is telling us is that trains are used more often in the cases when they are less circuitous (presumably roads are used when train routes are too circuitous).

In addition to the above details, we also find that 18.29% of the trips use zero hex links, meaning they were able to use connecting links to reach a transportation node directly from the home hex and to the work hex. In only 0.05% of the trips were only hex links used, indicating that for 1 in 2000 people directly walking to work is actually the most time-efficient mode of travel. Although at least one train link was used in 88.54% of the trips, only 3.1% used only trains and hexes, indicating that trips to and from stations do rely on other modes (e.g., using roads to reach the station). In contrast, 96.53% of trips used some road links and 10.27% used only road (and hex) links. Five percent of trips used bus links, but only 0.05% used buses without train or car links (same percent as hex links). The generated trips outcomes tell us that effective travel to and from work truly is

Table 7 Characteristics of the trip data related to mode usage, total distance, and circuity

Mode	Network Mean edge Length (m)	Trip Mean edge Length (m)	Mean percent of Distance traveled	Correlation of Percent distance and total distance	Correlation of Percent distance and circuity
Hex	250	250	1.6	-0.336	0.046
Train	2742	5156	71.4	0.488	-0.214
Bus	509	7457	1.5	0.098	0.085
Road	528	422	22.6	-0.482	0.192

multimodal, but with only 2.82% of trips using all four major modes of transportation, the best mix of modes is highly contextual.

Generated trips to vs from

We were initially surprised that for just under one-third of the trips (96,219 of 300,000) a different path was taken to and from the same pair of points. Recall Table 3 of edge usage frequencies; although there is clearly a high degree of similarity, there are also some marked differences (see also maps showing frequency weights in each direction in the Additional file 1). This may not seem surprising on a directed network, but nearly all the links are reciprocal and symmetrically weighted. For those connections that are asymmetrically weighted, such as bus and train loading/unloading links, the total trip time would still be symmetric because there is a loading and unloading link on each side of a train or bus ride.

The cause of the large number of non-identical return trips is that we used the single-source Dijkstra path algorithm and it only returns one path of possibly many equivalent travel-time paths. In anticipation of this we preferred to find all paths and weight the counts by the number of paths (as in typically done for betweenness calculations), but were unable to do so due to computational time requirements (more than 20x longer). In fact, in all cases where the to- and from-trips took different paths, the time it took was still identical. This is especially an issue on the hex grid part of the network where many distinct paths of equivalent length and time can be found. When we filter to those paths across which a different number of train, bus, or road links were used, the number of non-identical return-trips reduces to 7856.

For the worst case difference in travel distance, the trip to work traversed 75 km (65 km via train) to reach a job 43 km away. The return trip only used 16 km of rail, but 21 km of bus links for a total travel distance of 52 km. The two trips took exactly the same amount of time. The largest difference in the number of steps occurs between a pair that is 18 km apart. The trip to work used mostly road links (68 out of 90 were road edges), while the return trip used 18 road links and 7 train links out of a total of 41 links. No buses were used in either direction. Road links connect intersections while train links connect stations, which are further away from each other. The total distance traveled differed by only 1.7 km, so the difference again highlights the importance of geographically grounded measures in terms of distances and time rather than network structural characteristics. The differences between to and from trips, or rather the existence of multiple equivalent paths between points, will largely be ameliorated with a more sophisticated trip generating function that includes costs, transfer penalties, minimum distances per mode, and other aspects.

Path verisimilitude

As described above, the proportions of how much each mode is used is very similar between the simplex and stochastically generated trips methods, even when measured as time used or distance traveled. The explanations for the differences provided above highlight a trade-off between the computational benefits of the simplex algorithm versus the data richness and intuitiveness of generating O-D trips. Here we pose a strong reason to favor the stochastically generated trips method that brings together the various results above: the paths generated are more believable.

Figure 9 shows the links used by each mode for both methods focused down to central Tokyo and filtered to only show edges used at least 10 times (full maps with all links available in the Additional file 1). The simplex method's implicit assumption that people take the nearest available job from their home is revealed in the large number of short, hex-link only trips. More than that, the road, bus, and train usage is also more fragmented and local. By contrast, the generated trip method reveals usage patterns that are more organic, consistent, and follow a familiar branching pattern with shared branches and isolated leaf nodes.

The problem with the simplex results is more than visual. Although there are more than three times as many people moving in the simplex simulation, the density of trips in the city centers and the variety of edges used is markedly less. Table 8 shows the numbers of distinct edges used by each transportation mode. From it we can see that despite the similarities in the proportional usage frequencies, times, and distances, the simplex method focuses its transportation usage into far fewer links.

Although both methods are abstractions, the simplex method is certainly more abstract. The simplex method is appropriate for logistic problems; e.g., finding the optimal distribution pattern from a set of warehouses to distributed customers. In the case of commuter transportation, an efficient solution is not necessarily a realistic solution. The similarity

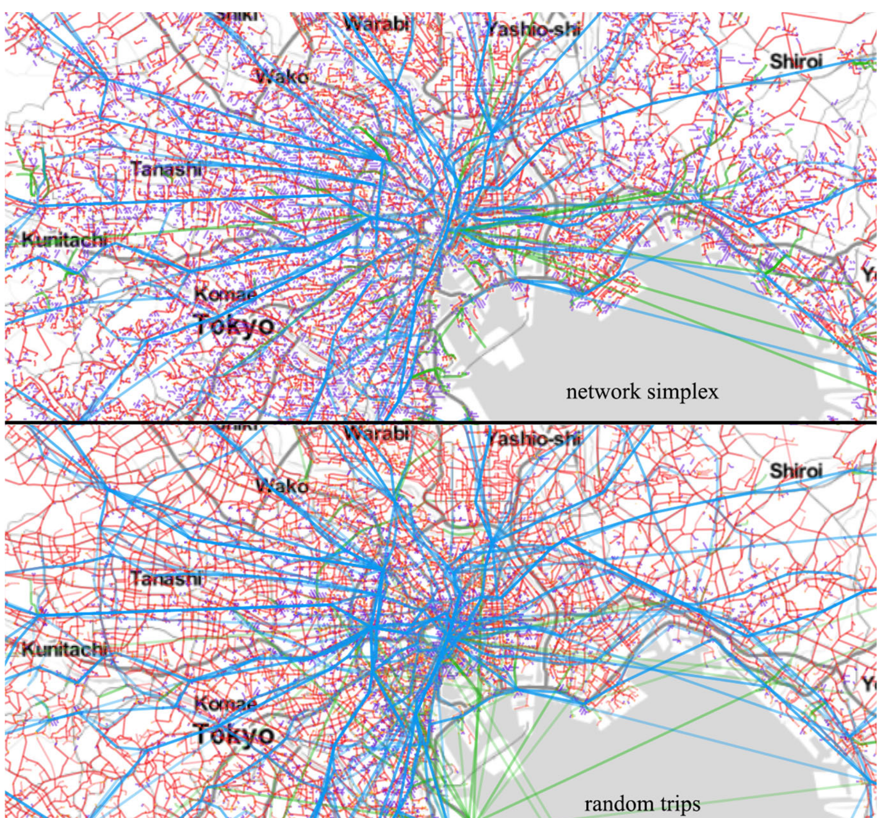

Fig. 9 A map of the Tokyo area zoomed in to show the links used by mode for the simplex algorithm (top) and generated trips to-work (bottom) filtering out edges with fewer than 10 traversals. Blue lines are trains, green are buses, red are roads, purple are interhex links, and connecting walking links are orange. Map tiles by Stamen Design (2019), map data by OpenStreetMap(OpenStreetMap Contributors 2019).

Table 8 Number of distinct edges used by each transportation mode for each flow method

Transportation mode	Simplex algorithm 1,000,000 demand	Generated trips 300,000 to work	Generated trips 300,000 from work
Hex	96,479	134,359	134,148
Walk	40,353	113,549	113,476
Train	1319	6916	6916
trainTransfer	114	4,372	4362
trainAccess	882	6,149	6,148
Bus	3601	9873	9854
busAccess	1,770	5335	5325
busTransfer	22	330	330
Road	24,406	67,879	67,858

in usage distributions does reveal something interesting about the core usage patterns, but the larger population/job sample size does not translate into a larger (or wider) sampling of the transportation network; the total edge frequency of the simplex method is 6,687,938 traversals, but 9,324,179 and 9,323,457 for the generated trips. The difference comes from the number of short trips sufficient to satisfy local net demand for the simplex algorithm. However, empirically there are few home-work commutes as short as the simplex algorithm generates.

We have already covered several ways in which the generated trip method can be expanded upon and improved. Even without these refinements the generated trip method reveals itself to be superior to the network simplex method in capturing usage patterns that reflect the complexity of the integrated multimodal transportation networks. Perhaps the simplex method can also be refined and expanded to produce more useful results. We have already mentioned the possibility of using income and quality tiers to better match people to jobs. Additional considerations could be used to establish multiple home and work distributions, and aggregating the results of running the simplex algorithm on each pair of distributions could be calibrated to produce more realistic usage patterns.

Comparison to most crowded trains

In addition to the exploratory results descriptions and qualitative comparisons to plausible commuter patterns above, here we provide a quantitative analysis of how well our methods generate patterns similar to empirical ones. The Ministry of Land, Infrastructure, Transportation and Tourism (MLIT) publishes a PDF listing selected train network segments including, but not limited to, those with the highest percent of capacity ridership for a given hour of operation (Ministry of Land 2018). The focus of the data is on directed segments exceeding capacity rather than highest ridership, but the data includes the ridership for those trains. Ideally we would have average daily ridership values for each segment, at least for the most traveled lines, allowing us to calculate the rank-biased overlap (Webber et al. 2010); a measurement of rank similarity that fosters comparisons of lists of different lengths with heavier weights for matching the higher-ranking segments. Unfortunately, such data is publicly unavailable, and we can only analyze the data we currently have access to.

To see why this matters, consider the following example. One listed segment has a capacity of 2688 and a ridership of 3130, making it 16% over capacity during that particular hour of the day. Another line segment has a capacity of 34,040 and a ridership of

32,240 during its peak hour, thus running at 95% capacity. Although the second train is less *crowded* at its maximum, even at off-peak hours it carries approximately ten times the traffic as the more crowded train. Because we do not yet have heterogeneous capacity data for each train line, we cannot match the crowdedness levels. However, we can determine the agreement between the ranking of the passenger counts on the provided line segments and our simulated traffic rankings on those same line segments.

Verification of line segment ridership

Of the line segments selected by MLIT, 76 are within our research area. First, we isolate the segments included in the empirical data from our network. We then calculate the Pearson correlation coefficient between the segments' ridership data and the usage values from simulations limited to the same segments (Ren et al. 2014). In the cases where multiple types of lines run between a pair of stations, we use the maximum usage value among the network edges. We calculate the Spearman correlation and the Kendall τ to evaluate how well our simulations match the ranking of the segments. Table 9 reports the results of these comparison tests.

The Pearson correlation between the empirical ridership and the four simulated usage levels is consistently, though weakly, positive at reasonably high significance levels (near the 0.01 level). The comparison of ranks, however, shows that the simplex algorithm performs noticeably less well in matching the empirical patterns according to both the Spearman correlation and the Kendall τ. A leading factor for the poor rank-matching performance is that for 28 of the 76 segments the simplex flow generated zero trips. This is another result of the unnatural assumption that people take the job closest to their home, and the unnatural transit patterns that it generates.

The generated O-D trips to-work and from-work achieve similar correlation levels for the morning commute data. The reason for this is the particular segments chosen by MLIT; the majority of which are located in or near the central core rather than along commuter lines. Once near enough to the job-dense central region the trains are indeed crowded in both directions due to people (for example) living west of the city but working on the east side of the core and vice versa. In fact, of the 76 segments reported by MLIT, there are 9 lines with two segments and 1 line with three segments. In the empirical data there are examples of lines running in opposite directions with comparable ridership.

Considering the simplifying assumptions made in generating the network, and the strictness of the test of matching ridership on just 76 of the 10,536 segments, the level

Table 9 Similarity and significance between the empirical rail segment usage and simulated rates

Simulation type	Measure	Correlation	*p*-value
Simplex	Pearson	0.3024	0.0079
Simplex	Spearman	0.1448	0.2119
Simplex	Kendall Tau	0.1078	0.1878
Trips to work	Pearson	0.2758	0.0159
Trips to work	Spearman	0.2659	0.0202
Trips to work	Kendall Tau	0.1899	0.0154
Trips from work	Pearson	0.3116	0.0061
Trips from work	Spearman	0.2830	0.0132
Trips from work	Kendall Tau	0.2052	0.008

of agreement between our simulations and empirical data is already reasonably high. Especially in light of the fact that in our simulations bus and taxi travel are also competing options. By adding data on heterogeneous transfer and access times, schedule limitations (e.g., evening only routes), and a cost/time trade-off in route planning, we expect to further improve the accuracy. Although those additions must be left for future work (after we are able to obtain the data), there is something we can do now to broaden our evaluation: test entire lines instead of specific segments.

Verification of whole line ridership

In the following analysis we extrapolate the empirical segment data to the lines, and compare the aggregated simulated traffic per line using both the mean and maximum. The data from MLIT only covers a tiny percentage of the line segments (0.72%), but it does include at least one segment from approximately half of the lines in the Tokyo area (65 of 131 lines, if not distinguishing by type). Twelve empirical segments appear on shared lines, so to extrapolate the value for each line we first take the mean of the segments' ridership values for each line. We next isolate those 65 lines present in the empirical data from our simulated data. We then aggregate the simulated usage data for those lines as the mean value of all the segments of that line (including all types of trains of that line). Finally, we perform the same statistical tests as above to determine the level of similarity of the results. Table 10 shows the results of comparing the mean ridership values along each line to the mean empirical values for that line.

As expected, comparing the aggregate across the whole line to the extrapolated line usage improves the resulting accuracy in most cases (only the Pearson correlation of the simplex algorithm is weaker). The Spearman correlation of the generated trips to work exceeds 0.45 while the trips from work are just slightly less. These results using the mean usage across each line make an even stronger case for stochastically generated O-D trips.

That said, the empirical data is reported for the most crowded hour of the included line segments. Because capacity for a given line is largely (but not entirely) constant across time, these are also the maximum ridership values for that line. Within any given hour, the ridership of a train is highly correlated across segments of a route. So the ridership of the whole line is fairly well represented by the segment data we have, especially in terms of ranks. So instead of using the mean to extrapolate and aggregate values along the whole line, we now examine the maximum values. For simulated data each line is assigned the

Table 10 Similarity and significance between the empirical mean rail line usage and the mean values of the simulated usage rates across all links on that line of all types

Simulation type	Measure	Correlation	*p*-value
Simplex	Pearson	0.1998	0.1105
Simplex	Spearman	0.2168	0.0828
Simplex	Kendall Tau	0.1473	0.0832
Trips to work	Pearson	0.3724	0.0022
Trips to work	Spearman	0.4530	0.0001
Trips to work	Kendall Tau	0.3327	0.000089
Trips from work	Pearson	0.3658	0.0027
Trips from work	Spearman	0.4442	0.0002
Trips from work	Kendall Tau	0.3260	0.0001

value of the link with the highest usage. Table 11 shows the match accuracy is further improved using the max values compared to the mean values.

The Spearman correlations are slightly below 0.60 for the stochastically generated trips, although still less than 0.40 for the simplex method. Considering the limitations of our transportation model construction, and the methods used to generate O-D trips, and our goal for recreating overall daily patterns rather than time-specific traffic micro-simulations, this result represents a strong relationship between our predicted ridership levels and the empirical ranking of train line ridership.

For the maximum ridership levels to match this much better than the segment-level ridership implies that a different segment of the line receives higher usage in the simulated data. This difference likely results from a combination of the simplifications used to create our current model and the limited scope of available empirical data. Although the current level of accuracy isn't exact enough to use for policy or routing decisions, the results are highly encouraging for the accuracy including our planned expansions.

Limitations of the verification test

One obvious limitation of this analysis is that it only compares traffic volumes on the train network. In the future, we plan/hope to gain access to aggregated GPS-style data of people's movements that would cover all modes of transportation (even walking) and allow us to better calibrate and verify the model's flows. Another limitation is the sparseness of the empirical data; the data only contains ridership for 76 of the 10,536 train edges. A larger dataset may not reveal better accuracy, but it would be a more robust test of the accuracy. Yet another limitation is that the data only reports ridership during its peak hour rather than a total for a day, although we expect these are highly correlated at least for ranks. Because we analyze commuting from home to work, and most of this traffic happens during the morning rush, this limitation is probably not problematic for our ranking analysis. However this limitation will become problematic when we include travel for other purposes such as shopping, commuting to school, etc. across a whole day.

Network flow and urban mobility

In our full integrated transportation network there are 32 weakly connected components, but the largest component contains 99.965% of the nodes, the second largest has 63 nodes, and the remaining 30 have less than 10 (islands along the jagged borders). The simplex

Table 11 Similarity and significance between the empirical rail line max usage and the max values of the simulated usage rates across all links on that line of all types

Simulation type	Measure	Correlation	*p*-value
Simplex	Pearson	0.3854	0.0015
Simplex	Spearman	0.3892	0.0014
Simplex	Kendall Tau	0.2794	0.0010
Trips to work	Pearson	0.5005	0.000022
Trips to work	Spearman	0.5809	0.00000039
Trips to work	Kendall Tau	0.4336	0.00000032
Trips from work	Pearson	0.4903	0.000034
Trips from work	Spearman	0.5732	0.00000060
Trips from work	Kendall Tau	0.4217	0.00000069

subgraph has 9862 weakly connected components; the dominant component only has 106,149 of its 178,793 nodes, many components have several hundreds or thousands of nodes, and most components have fewer than 10 nodes. By contrast, the generated trips method generated a single weakly connected component containing 169,220 nodes. The connected aspect of the generated trips network facilitates additional measurements incorporating network features and connections to the underlying geographic data stored in the hex nodes. We leave out an analysis of our Geosocioeconomic accessibility measures (e.g., reachable people and jobs weighted by distance), but here we tie our results back into wider questions of mobility.

Using the same multimodal transportation network used to examine aggregate commuter flows, we can also examine the mobility properties of the area. Consider the following two measures of mobility: weighted reach and furthest reachable point. *Weighted reach* or *reachability* is an accessibility measure (Levinson 2012; Biazzo et al. 2018) that counts, for each hex node i, the number of other hexes j reachable from i, weighted by the inverse of the time to reach j:

$$\text{reachability}(i) := \sum_{j} \frac{1}{t_{ij}}. \tag{2}$$

in which t_{ij} is the shortest time from hex i to hex j. For the current analysis we limit the range to hexes reachable within 20 min. We also determine the Haversine distance to the furthest hex reachable within the same 20 min. Figure 10 shows the values of both measures for every hex in the Tokyo Area.

These two measures provide distinct windows on mobility: weighted reach is a measure of breadth whereas furthest distance is a measure of depth. In the case of commuting to work, one may primarily care about a single route from your suburban home to your downtown office; and a similar thing may be said for shopping districts and other activities. Having a large furthest reachable distance may be more important, and we see the highest values around suburban train stations and remote locations. Weighted reach instead shows a pattern aligned with the road network, which is obviously more adept at omnidirectional travel than the rail system. Some areas (such as Shiroi to the northeast of central Tokyo) have a very high mobility in one sense (furthest distance), but poor mobility in the other (reach).

Note that central Tokyo does not have especially high mobility by either measure. This is true of the 20 min cutoff version because most trains in central Tokyo are local and there is more focus on coverage than traversal speeds. However, because those commuter lines that provide high mobility to suburban locations end up near downtown, extending the cutoff to 45 or 60 min changes the distribution of hot spots toward the center. However, the divergence of depth-based mobility favoring trains and breadth-based mobility favoring roads remains at all scales.

Summary of results

Our exploratory analysis of flows of people to jobs produced several interesting discoveries of patterns in the multimodal transportation network of the Greater Tokyo Area. We describe several "stylized facts" regarding the outcome of the simulations to facilitate qualitative comparisons and broader understanding for the Tokyo Area multimodal transportation system followed by a limited quantitative verification test using train ridership

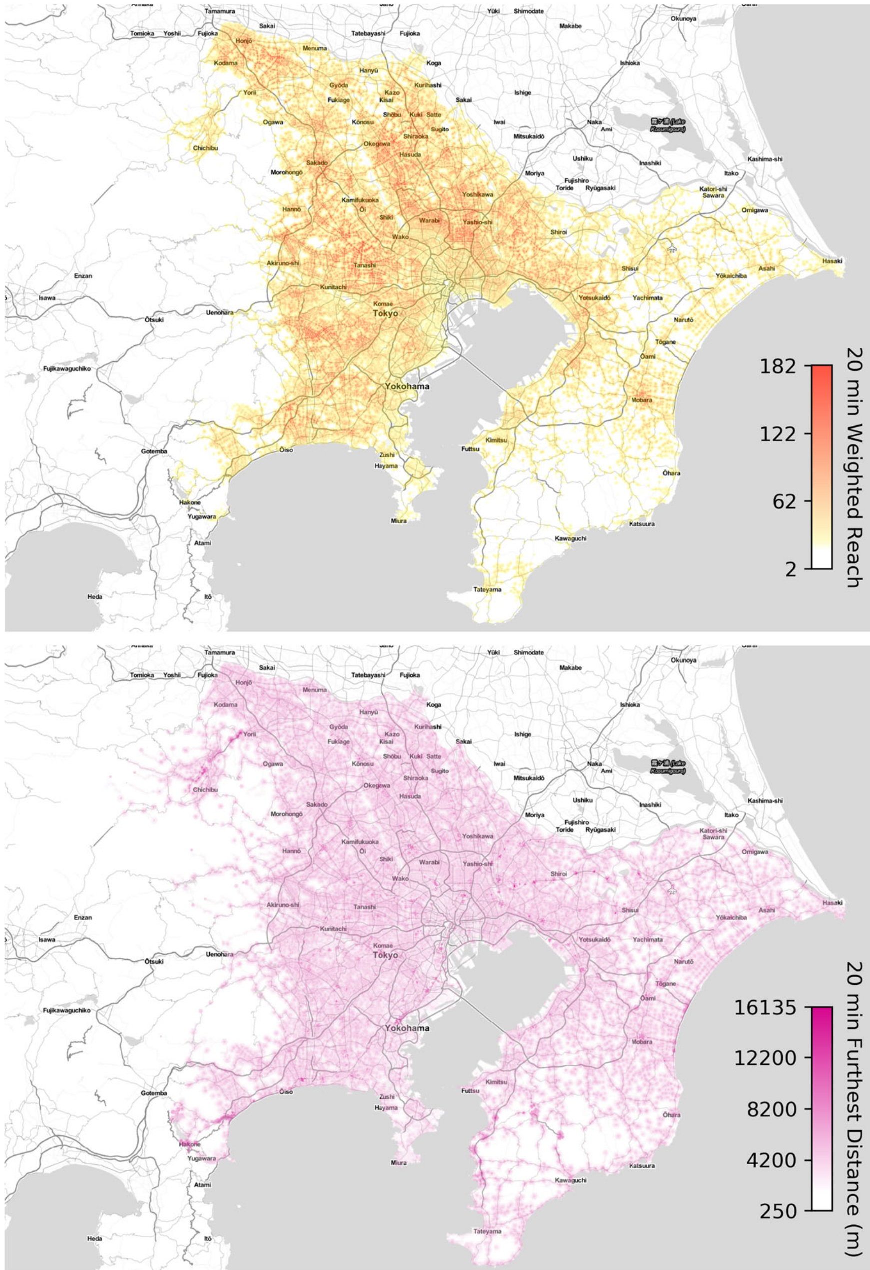

Fig. 10 A map of the Tokyo area showing the 20 min reachability (top) and distance of the furthest point (in meters, bottom) reachable in 20 minutes. Some locations in low reachability areas can reach long distances, for example those near commuter rail stations. Map tiles by Stamen Design (2019), map data by OpenStreetMap Contributors (2019)

data. These results are used as a springboard for future work to refine the trip-generating methods and calibrate as much as possible to the scant available data.

We also produced results that act as fuel for comparing stochastically generated O-D trips to the network simplex method for generating flows appropriate for an aggregate analysis such as ours. Both methods produce similar results in terms of the edges used most and percents of traffic on each mode, but they produce very different paths. In both cases the buses were underutilized because they are largely redundant to the faster road network, except express buses. The road network provides more rapid access to localities away from stations to help diffuse people from stations along the high throughput train lines. Overall, we find that the generated trips are more realistic and match better to empirical data than the simplex flow results. We next consider some implications of these results.

Summary of conclusions

It is an important distinction (and one worth repeating) that our analysis is not meant to track individuals' actual commuting routes, predict congestion, or reflect short-term fluctuations in the locations of individuals (Cherdarchuk 2014). We are interested in persistent patterns in the ebb and flow of human and economic capital. Levels of socioeconomic flow in our analysis are abstract quantities representing the overall and sustained movement of social capital through a city. Rather than a proposal that an individual who lives in hex a is likely to have a job in hex b, the distributions of home and work locations represent more abstract sources and sinks of human and financial capital. The revealed usage patterns are meant to be an approximation of the overall network-based potential flow of people to and from work.

The network simplex algorithm finds the most time-efficient allocation of people to jobs, but obviously few people actually work in the job that is closest to their home. A huge variety of conditions, circumstances, and historical events determine both home and job locations; so they are related, but their relationship is non-obvious, heterogeneous, and inaccessible to us. The generated trip method is therefore more realistic in this sense, but our current methods is yet to include mechanisms to reflect that people choose their home locations based on their job locations and vice-versa. They are different abstractions of movements of people from home to work locations weighted by the probabilities of such movements.

Future work

The text above already discusses several planned refinements and expansions to the flow-generating methods and to calibration/validation. We elaborate on those points slightly in this section.

One of our primary interests is an in-depth mobility and accessibility analysis. In addition to reachability described above, we also measure the time-weighted numbers of people and jobs reachable from a location. These measures convert the above kinds of mobility results into socioeconomic accessibility results. We planned to investigate the relationship between sociability and economic opportunity to better understand the role of "business centers" and differentiate them from other locations with high sociability scores. By comparing results across geographic regions we hope to determine whether socioeconomic patterns are recurring and self-organizing phenomena or scale- and path-dependent ones. Just as the sociability score can be used to evaluate a location's accessibility to other people and places, the economic opportunity score indicates a location's access to employment and economic resources. This index, and its correlation with other factors such as income and housing prices, is likely to reveal useful insights for many practical and theoretical purposes. For considerations of space and time, this part of the current project is being left for future work.

As mentioned in the data section, there are a few features of our data that weaken the strength of our results. The main of these is the lack of expressways from the road data. Another worry is the disconnected nature of the road network when considering only tertiary or larger roads, although this is largely mitigated by the hex and walking networks, these links may be artificially depressing the flow contribution of the road network. Our train and bus link traversal time data are currently based on a sample of the schedules for each line. Although highly consistent, there are differences in traversal duration and availability across the day. Our current analysis is meant to be an aggregate flow across typically available transportation methods, but more refined data for Japan's train and bus systems could make an analysis that depends on the time of day possible (Ayed et al. 2011; Cherdarchuk 2014; Idri et al. 2017; Biazzo et al. 2018).

Clearly travel time is not the only factor people consider when deciding on the mode of transportation. We are also working to include data on transport costs, congestion/crowdedness, parking availability, etc. as well as a multi-objective cost function for path determination (Abdelghany and Mahmassani 2001). As our goals are still aggregate pattern discovery rather than transit scheduling or time-specific route planning we can straddle the line between abstract methods (like those used here) and the highly detailed agent-based rules in traffic micro-simulations. We can also examine distance traveled and energy used to explore the environmental impact of different travel methods. Housing preferences and prices are determined through many connected factors, and we plan to determine the degree to which transportation accessibility influences housing/job choice and demand (Hatta and Ohkawara 1993). With detailed multimodal transportation networks and rich datasets of various socioeconomic and geographic factors across Japan in place we look forward to ongoing fruitful research and collaborations.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1007/s41109-019-0244-y.

Additional file 1: Supplementary Materials. Additional data summary tables not directly related to the analyses performed in this work but perhaps of interest to other transportation network researchers (pdf file).

Acknowledgements
Not Applicable.

Authors' contributions
AB and HI conceived of the work; AB designed and wrote the work, as well as performed most of the analyses; MH and ZB helped acquire and analyze data. All authors read and approved the final manuscript.

Funding
Not Applicable

Availability of data and materials
As described in the text, the population/employment/mesh data and road network data are openly available from the citations provided. The train and bus network data come from third-party sources (Ekitan). Additional output data and/or plots of data generated are available upon request.

Competing interests
This research was performed by employees of GA Technologies and may lead to the development of products or information services which may be used by GA Technologies for business operations.

Author details
[1]GA Technologies Inc., Roppongi Grand Tower 40F, Roppongi 3-2-1, Minato-ku, Tokyo, 106-6290, Japan. [2]Laboratory for Symbolic Cognitive Development, RIKEN Center for Biosystems Dynamics Research, 6-7-3 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan. [3]Department of General Economics, Ghent University, Tweekerkenstraat 2, Ghent 9000, Belgium. [4]Department of Software and Information Systems, University of North Carolina Charlotte, 9201 University City Blvd., Charlotte, NC 28223, USA.

Received: 20 May 2019 Accepted: 10 December 2019
Published online: 22 January 2020

References

Abdelghany KF, Mahmassani HS (2001) Dynamic trip assignment-simulation model for intermodal transportation networks. Transp Res Rec 1771(1):52–60
Ahuja RK, Magnanti TL, Orlin JB (1993) Network Flows: Theory, Algorithms, and Applications. Prentice Hall. https://books.google.co.jp/books?id=WnZRAAAAMAAJ
Association for Promotion of Infrastructure Geospatial Information Distribution (2015) 250m. www.geospatial.jp. Accessed: 12 Dec 2018
Association for Promotion of Infrastructure Geospatial Information Distribution (2015) 500m. www.geospatial.jp. Accessed: 4 Apr 2018
Automobile Inspection & Registration Information Association (2017) Number of cars per household. https://www.airia.or.jp/publish/statistics/mycar.html. Accessed 10 May 2019
Ayed H., Galvez-Fernandez C, Habbas Z, Khadraoui D (2011) Solving time-dependent multimodal transport problems using a transfer graph model. Comput Ind Eng 61(2):391–401
Bachir D, Khodabandelou G, Gauthier V, El Yacoubi M, Puchinger J (2019) Inferring dynamic origin-destination flows by transport mode using mobile phone data. Transp Res Part C Emerg Technol 101:254–275
Barthélemy M (2011) Spatial networks. Phys Rep 499(1-3):1–101
Biazzo I, Monechi B, Loreto V (2018) Universal scores for accessibility and inequalities in urban areas. arXiv preprint arXiv:1810.03017
Bramson A., Vandermarliere B. (2016) Benchmarking measures of network influence. Sci Rep 6:34052
Calimente J. (2012) Rail integrated communities in tokyo. J Transp Land Use 5(1):19–32
Cherdarchuk J (2014) Breathing City. https://www.darkhorseanalytics.com/blog/breathing-city
Chorus P, Bertolini L (2011) An application of the node-place model to explore the spatial development dynamics of station areas in tokyo. J Transp Land Use 4(1):45–58
Cipriani E, Petrelli M, Fusco G (2006) A multimodal transit network design procedure for urban areas. Adv Transp Stud 10
Crucitti P, Latora V, Porta S (2006) Centrality in networks of urban streets. Chaos Interdiscipl J Nonlinear Sci 16(1):015113
Cunningham WH (1976) A network simplex method. Math Program 11(1):105–116
Dantzig GB (1951) Application of the simplex method to a transportation problem. Act Anal Prod Allocation
Derrible S. (2012) Network centrality of metro systems. PLoS ONE 7(7):40575
Derrible S, Kennedy C (2009) Network analysis of world subway systems using updated graph theory. Transp Res Rec 2112(1):17–25
Domencich T, McFadden D (1975) Urban travel demand-a behavioral analysis. Technical report, Amsterdam: North Holland. https://doi.org/10.2307/134305
Ekitan: Train and Bus Schedule Data (2019). https://www.ekitan.co.jp
Erlander S, Stewart NF (1990) The Gravity Model in Transportation Analysis: Theory and Extensions, vol. 3. Vsp
Gallo G, Pallottino S (1988) Shortest path algorithms. Ann Oper Res 13(1):1–79
Goczyłla K, Cielatkowski J (1995) Optimal routing in a transportation network. Eur J Oper Res 87(2):214–222
Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779
Hagberg A, Swart P, S Chult D (2008) Exploring network structure, dynamics, and function using networkx. Technical report. Los Alamos National Lab.(LANL), Los Alamos
Hatta T, Ohkawara T (1993) Population, employment, and land price distributions in the tokyo metropolitan area. J Real Estate Finance Econ 6(1):103–128
Hibino N, Uchiyama H, Yamashita Y (2005) A study on evaluation of level of railway services in tokyo metropolitan area based on railway network assignment analysis. J East Asia Soc Transp Stud 6:342–355
Highway Capacity Manual (2000) Highway capacity manual, vol. 2, Washington
Huang J, Levinson DM (2015) Circuity in urban transit networks. J Transp Geogr 48:145–153
Idri A, Oukarfi M, Boulmakoul A, Zeitouni K, Masri A (2017) A new time-dependent shortest path algorithm for multimodal transportation network. Procedia Comput Sci 109:692–697
Japan Traffic Safety Association: Rules of the Road (2017) Japan Automobile Federation
Kang C., Liu Y., Guo D., Qin K. (2015) A generalized radiation model for human mobility: spatial scale, searching direction and trip constraint. PLoS ONE 10(11):0143500
Kawasaki S (2015) The challenges of transportation/traffic statistics in japan and directions for the future. IATSS Res 39(1):1–8
Kim K-O, Rilett L (2003) Simplex-based calibration of traffic microsimulation models with intelligent transportation systems data. Transp Res Rec 1855(1):80–89
Lee Y-J, Choi JY, Yu JW, Choi K (2015) Geographical applications of performance measures for transit network directness. J Publ Transp 18(2):7
Levinson D. (2012) Network structure and city size. PLoS ONE 7(1):29721
Li D, Lin Y, Zhao X, Song H, Zou N (2011) Estimating a transit passenger trip origin-destination matrix using automatic fare collection system. In: International Conference on Database Systems for Advanced Applications. Springer. pp 502–513. https://doi.org/10.1007/978-3-642-20244-5_48
Liu Y, Wang F, Xiao Y, Gao S (2012) Urban land uses and traffic 'source-sink areas': Evidence from gps-enabled taxi data in shanghai. Landsc Urban Plan 106(1):73–87
Lozano A, Storchi G (2001) Shortest viable path algorithm in multimodal networks. Transp Res Part A Policy and Pract 35(3):225–241
Masucci AP, Serras J, Johansson A, Batty M (2013) Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. Phys Rev E 88(2):022812
Masuda N, Holme P (2017) Temporal Network Epidemiology. Springer. https://doi.org/10.1007/978-981-10-5287-3
Ministry of Land Infrastructure, Transport and Tourist (2010a) National Land Numeric Information Traffic Flow: Person Trip OD Amount Data. https://nlftp.mlit.go.jp/ksj-e/jpgis/datalist/KsjTmplt-S05-b.html. Accessed 29 May 2019

Ministry of Land Infrastructure, Transport and Tourism (2010b) Nationwide Urban Traffic Characteristics Survey. http://www.mlit.go.jp/common/001032141.pdf. Accessed 29 May 2019
Ministry of Land Infrastructure, Transport and Tourist (2018). https://www.mlit.go.jp/report/press/tetsudo04_hh_000077.html. Accessed 19 Jun 2019
Munizaga M, Devillaine F, Navarrete C, Silva D (2014) Validating travel behavior estimated from smartcard data. Transp Res Part C Emerg Technol 44:70–79
Myojo S (2006) Method to estimate passenger flow using stored ticket gate data. Q Rep RTRI 47(4):178–181
Nazem M, Trépanier M, Morency C (2013) Integrated intervening opportunities model for public transit trip generation–distribution: A supply-dependent approach. Transp Res Rec 2350(1):47–57
NHK Culture Research Institute (2015) Survey Data. https://www.nhk.or.jp/bunken/research/yoron/pdf/20160217_2.pdf. Accessed 19 Jun 2019
OECD Statistics: Transport Transport Measurement Passenger transport (2016). https://stats.oecd.org. Accessed 8 Apr 2019
Official Statistics of Japan (2014) Economic Census for Business Frame, Tabulation of Establishments, Results for Japan. www.e-stat.go.jp. Accessed: 12 Jan 2011
Official Statistics of Japan (2015) / / 2015/ 5250m/. www.e-stat.go.jp. Accessed: 12 Dec 2018
Ohmori N, Muromachi Y, Harata N, Ohta K (2000) Travel behavior data collected using gps and phs. In: Traffic and Transportation Studies (2000). pp 851–858. https://doi.org/10.1061/40503(277)130
OpenStreetMap Contributors (2019). https://www.openstreetmap.org. Accessed 18 Feb 2019
Parady GT, Loder A, Axhausen KW (2018) Heterogeneous travel activity patterns in Japan: Accounting for inter-dependencies in mobility tool use. Technical Report No. 18-03547, University of Tokyo
Polus A, Craus J, Livneh M (1991) Flow and capacity characteristics on two-lane rural highways. Transp Res Rec 1320
Public Purpose: Urban Transport Factbook (2003) Tokyo Yokohama Suburban Rail? Summary. http://www.publicpurpose.com/ut-cr-tok.pdf. Accessed 19 Jun 2019
Ren Y, Ercsey-Ravasz M, Wang P, González MC, Toroczkai Z (2014) Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. Nat Commun 5:5347
Rieser-Schüssler N, Balmer M, Axhausen KW (2013) Route choice sets for very high-resolution data. Transportmetrica A Transp Sci 9(9):825–845
Romanos MC, Saidane H (1978) A stochastic markovian approach to trip distribution. In: Papers of the Regional Science Association, vol. 41. Springer. pp 15–28. https://doi.org/10.1111/j.1435-5597.1978.tb01035.x
Sels P, Dewilde T, Cattrysse D, Vansteenwegen P (2011) Deriving all passenger flows in a railway network from ticket sales data. In: Proceedings of Rail Rome
Simini F, González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. Nature 484(7392):96
Stamen Design (2019) Map Tiles. http://maps.stamen.com
Stefanouli M, Polyzos S (2017) Gravity vs radiation model: two approaches on commuting in greece. Transp Res Procedia 24:65–72
Tak S, Kim S, Yeo H (2014) Travel time prediction for origin-destination pairs without route specification in urban network. In: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE. pp 1713–1718. https://doi.org/10.1109/itsc.2014.6957940
Toei Transportation Customer Center (2018) A First in Japan! Toei Bus Begins Operating a Fully Flat Floor Bus. https://www.kotsu.metro.tokyo.jp/eng/news/2018/20181214_8366.html. Accessed 19 Jun 2019
Train Media (2017). https://www.train-media.net/report/1810/1810.html. Accessed 8 Apr 2019
Tsekeris T, Geroliminis N (2013) City size, network structure and traffic congestion. J Urban Econ 76:1–14
U.S. Department of Commerce (2010) LEHD Origin-Destination Employment Statistics (LODES). https://lehd.ces.census.gov/data/. Accessed 19 Jun 2019
Uno N, Kurauchi F, Tamura H, Iida Y (2009) Using bus probe data for analysis of travel time variability. J Intell Transp Syst 13(1):2–15
Verga J, Silva RC, Yamakami A (2018) Multimodal transport network problem: Classical and innovative approaches. In: Soft Computing for Sustainability Science. Springer. pp 299–332. https://doi.org/10.1007/978-3-319-62359-7_14
von Ferber C, Holovatch T, Holovatch Y, Palchykov V (2009) Public transport networks: empirical analysis and modeling. Eur Phys J B 68(2):261–275
Wang X-b, Zhang G-j, Hong Z, Guo H-f, Yu L (2009) Modeling and implementing research of multimodal transportation network. In: 2009 First International Conference on Information Science and Engineering. IEEE. pp 2100–2103. https://doi.org/10.1109/icise.2009.752
Webber W, Moffat A, Zobel J (2010) A similarity measure for indefinite rankings. ACM Trans Inform Syst (TOIS) 28(4):20
Zipf GK (1946) The p 1 p 2/d hypothesis: on the intercity movement of persons. Am Sociol Rev 11(6):677–686

Publisher's Note