**Introduction**

London’s remarkable vibrancy, productivity, efficiency and innovativeness depend in large part on its effectiveness in connecting workers to job opportunities through economic, social, intellectual and physical networks (Bettencourt & West, 2010). One important aspect of job accessibility (that is, the ease with which workers can take advantage of a suitable range of job opportunities) (Hansen, 1959) is the spatial and temporal distance between workers and jobs. Recent studies of job accessibility in London have revealed that the number of jobs that a worker can physically reach within a certain travel time varies widely across space (Ford, et al., 2015), and that the degree of spatial matching between the home and work locations of workers varies by occupation (+ citation Shen & Batty). However, job accessibility also varies across many other dimensions (Lucas, 2012), of which two are of special interest to policymakers. Firstly, even if jobs are physically accessible by workers, some workers may face non-spatial barriers (like a lack of relevant skills) to actually perform those jobs (Social Exclusion Unit, 2003). Secondly, the margin of uncertainty in travel times may adversely affect workers’ ability to reach suitable jobs within a reasonable travel time (Greater London Authority, 2018) (+ citation Uniman, Ehrlich).

To this end, this study will develop a typology of neighbourhoods in London that categorises them by overall job accessibility, the degree to which they depend on particular economic sectors for jobs, the relative social deprivation of their working-age population and a novel dimension, the reliability of travel times for trips originating in them. In addition to providing insights into how land use and transport interventions can adjust the large-scale spatial structure of London to improve job-worker spatio-temporal matching, the results can also provide guidance for where to intervene along the other dimensions, such as through programmes to improve workers’ skillsets, encourage job growth in certain sectors and improve the reliability of public transport operations, thus allowing for more coordinated strategies that utilise the tools available across a wider range of domains. Finally, the study also proposes a novel graph-theoretic interpretation of the job accessibility metric used, which indicates potential directions for further innovations with this metric.

**Literature Review**

Hansen (1959) defines accessibility as the potential for interaction between people and the opportunities that they seek. Job accessibility refers to the range of job opportunities that are available for a person to choose from. This is a function of the spatial distributions of potential workers’ home locations and jobs on the one hand, and the cost of travel between different locations on the other. Many indicators have been developed to study accessibility (Geurs & Wee, 2004), but “location-based” indicators are most frequently used for city-level job accessibility analyses, because they treat accessibility as a property of different areas (as opposed to a property of travellers or of components in the transport network) and thus can be used to spatially aggregate and visualise local unevenness in the distribution of accessibility. Most studies use simple “cumulative potential” indicators (+ citations), which sum up the total number of jobs that can be physically reached from each area within a certain travel cost (also known as a catchment):

where ao is the job accessibility of one of m origin areas o, jd is the number of jobs in one of n destinations d, co,d is the travel cost (in this study, time) between o and d, and f(co,d) is a distance decay function reflecting the declining attractiveness of the jobs in d to workers in o as co,d increases.

Cumulative potential measures have a theoretical flaw, as they treat job accessibility as non-excludable: the fact that any particular worker can reach a job does not affect any other worker’s ability to reach the same job. This fails to capture the competition between workers for jobs arising from the fact that each job can typically only be occupied by one worker. To address this, Shen (1998) proposed adjustments to the measure to account for the spatial distribution of potential workers, which was then further developed in the public health literature (+ citations) as the two-step floating catchment area (2SFCA) method. This is given by:

where wo is the number of workers in o, sums up for each destination d how many potential workers across all m origins o can reach it, divides the number of jobs in each destination d among the number of potential workers who can reach it, and sums up for each origin o the per-worker job contributions over all n destinations d that can be reached.

Like ao, ro is a score for area o reflecting the relative availability of job opportunities within its catchment of each worker. However, the 2SFCA method distributes every job exactly once (i.e. there are no leftover unallocated jobs) to potential workers in an excludable way (i.e. if a job is allocated to a particular worker, that job is no longer available to other workers) (Shen, 1998, pp. 363-364). Thus, ro reflects both connectivity and competition effects: as the number of jobs within o’s catchment increases, ro will increase, but as the number of potential workers who can reach the same catchment (and thus compete for the same jobs) as o increases, ro will decrease. This directly addresses the theoretical drawbacks of simpler cumulative potential measures, so I will use the 2SFCA method to assess job accessibility levels in London.

In the literature on this method, the indicator is invariably expressed as two summations, first iterating over all origins o with respect to each destination d (to find out which origins have catchments that include d), then iterating over all d with respect to each o (to sum up the per-worker job allocations from all d within o’s catchment). However, the method can be equivalently expressed using matrix operations, as I show below. A matrix implementation of the 2SFCA method has computational efficiency advantages. But more importantly, it brings out an interesting but understudied property of the method: it encodes the travel costs as an incidence matrix of a bipartite network involving origin and destination nodes. This implies that the 2SFCA method may be more flexible and widely applicable than currently thought. I will elaborate on some promising implications of this property and potential avenues for innovative applications of 2SFCA below.

Besides analysing the spatial patterns in overall job accessibility, researchers have also examined how accessibility varies by different types of workers, jobs and connectivity. To capture how accessibility varies by worker characteristics, some researchers separately calculate accessibility indicators for subsets of workers of different types (for instance, by considering the distribution of workers in certain income strata or occupations only rather than all workers) (Shen, 1998; Deboosere & El-Geneidy, 2018) (+ Shen & Batty), or examine how accessibility levels vary by area-level indicators of sociodemographic characteristics (like the average income or unemployment rate at the neighbourhood level) (Sanchez, 1999; Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018). In this study, I will use the English Indices of Multiple Deprivation (IMD) (Department for Housing, Communities and Local Government, 2015) to capture spatial variations in sociodemographic characteristics among workers. To study how job accessibility varies by job type, researchers consider job distributions by economic sector (Shen, 1998) (+ other citations), wage level (Deboosere & El-Geneidy, 2018) or occupation (+ Shen & Batty). In this study, I will use publicly available data on the spatial distribution of jobs by economic sector to disaggregate job accessibility by sector.

Researchers also examine how accessibility varies by different types of connectivity. In general, these efforts fall into two categories: those that compare connectivity by different transport modes (Shen, 1998; Kotavaara, et al., 2012; Ford, et al., 2015; Neihaus, et al., 2016), and those that compare connectivity across different time periods (such as the AM peak and evening off-peak periods), to capture how differences in travel times throughout a day can impact accessibility (Tenkanen, et al., 2016; Järv, et al., 2018; Hu & Downs, 2019). However, connectivity for the same period (such as the AM peak) also varies across different days, as a trip on any given day may be affected by unforeseen events like congestion, vehicles arriving late or breakdowns in the transport network (Katz & Quealy, 2019). In reality, travellers have to account for a margin of uncertainty around expected travel times where deciding when and how to travel. This angle of travel time reliability is under-studied in the job accessibility literature, but has received increasing attention in the transport studies literature in recent years, driven by the increasing availability of large-scale transport data (such as automated fare collection records, automated vehicle location traces or real-time trip planning information) that tracks the amount of time needed to complete a trip in real time (+ citation).

In light of the existing literature, I will position this study to make several contributions to the analysis of job accessibility. Firstly, in addition to confirming how job accessibility varies across space in London, my analysis will illuminate how these accessibility levels vary among working-age residents and jobs with different characteristics. Secondly, I will incorporate the novel dimension of travel time reliability into an accessibility analysis. Finally, I will propose a novel graph-theoretic implementation of the 2SFCA method, which opens up promising new areas of potential research using this method.

**Methodology and Data Sources**

*A Matrix Implementation of the 2SFCA Method*

The simplest cumulative potential accessibility measure is essentially the dot product of 2 components, , where:

1. J is a column vector of length n:

where jd is the number of jobs in destination d, and;

1. C is a m by n matrix:

where co,d is the travel time between origin o and destination d, and the distance decay function f(co,d) indicates the attractiveness of d to o.

The 2SFCA method also considers:

1. The spatial distribution of potential workers, stored in a column vector W of length m

where wo is the number of potential workers in o.

Then, the accessibility value ro for each origin o is stored in a column vector R of length m, which is obtained by:

where is equivalent to carrying out across all d, is element-wise Hadamard division that yields across all d, and yields across all o.

Recall that the 2SFCA method allocates jobs to workers in an excludable way. Because of this, the values ro are meaningful for assessing whether the local supply of jobs is sufficient to satisfy the local demand. Methods to assess the degree of inequality in distributions such as Lorenz curves can also be straightforwardly applied, by first calculating for each origin o how many jobs are allocated to it, which is given by:

and then sorting the origins in ascending order by ro values and plotting the cumulative proportion of the potential workers against the cumulative proportion of all jobs that are allocated to them.

*Interpreting C as an Incidence Matrix of a Bipartite Graph*

In the matrix C, each row represents an origin and each column represents a destination. The value of f(co,d) encodes the strength of the link between o and d. Clearly, C is equivalent to an incidence matrix of a bipartite network with two types of nodes (origins and destinations) and edges (with weights given by f(co,d)) only between nodes of different types.

Three properties immediately follow. First, the nodes in the network need not be spatial units per se but can be objects of any sort which may or may not have spatial properties. In the context of this study, even though 2SFCA is “location-based”, we should think of o and d not as areas with workers and jobs respectively, but as workers and jobs with locational coordinates. Secondly, the distances encoded in co,d need not be spatial, and can incorporate non-spatial distances like dissimilarity indices. Finally, the weight on any edge can be set independently from all other edge weights. This means that each origin-destination pair can use its own distance decay function, opening up vast flexibility for researchers to craft the relationships between origins and destinations.

Although these properties indicate that 2SFCA can be extended in many interesting directions, for this study I will only take advantage of the first property to directly integrate job counts by sector into the matrix operations for 2SFCA. This keeps my findings more directly comparable with extant job accessibility analyses. Nonetheless, I will suggest potential innovations for the 2SFCA method in the Discussion below.

*Choosing a Spatial Aggregation Level and a Distance Decay Function*

The use of the 2SFCA method requires two choices. The first is a choice of which level of spatial aggregation to use. Since travel times are calculated between two points, ideally data for each address point should be used. However, publicly available population and job distribution data are typically aggregated to areal units. Then the travel time between o and d is represented by the travel time between the two areas’ centroids. Some studies divide the study area into a lattice of equally-sized squares, as this results in evenly-spaced centroids that are thought to reduce distortions in the travel time calculations (Kotavaara, et al., 2012; Tenkanen, et al., 2016; Järv, et al., 2018; Wang, et al., 2018; Hu & Downs, 2019). However, reprojecting data aggregated in different spatial units to these squares requires additional modelling and assumptions, at the risk of distorting the data. To avoid this, I follow the majority of extant studies (Shen, 1998; Sanchez, 1999; Luo & Wang, 2003; El-Geneidy, et al., 2016; Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018; Deboosere & El-Geneidy, 2018) to use pre-existing spatial units; specifically, the m origins and n destinations in this study are, respectively, the population-weighted and jobs-weighted centroids of the 983 Middle-layer Super Output Areas (MSOAs) within the GLA boundary. MSOAs are spatial data reporting units created by the UK’s Office of National Statistics that break up the UK’s land mass into areas that contain between 5,000 and 15,000 residents or between 2,000 and 6,000 households, and allow my analysis to directly incorporate a wide range of UK government statistics that are spatially disaggregated to the MSOA level or smaller.

The second choice involves how to define the distance decay function . Many studies use continuous functions so that decreases smoothly as increases (Shen, 1998; Sanchez, 1999; Kotavaara, et al., 2012; Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018; Hu & Downs, 2019). This more accurately reflects how the attractiveness of a destination declines with increasing travel time (+ citation Cheng & Jia). However, it is unclear what functional form the distance decay actually takes in reality (+ citation Cheng & Jia), fine-grained origin-destination flow data and additional modelling are needed to estimate the parameters for the distance decay functions, and the resulting raw accessibility values are scores that have no intrinsic meaning and that can only be interpreted relative to other scores (El-Geneidy, et al., 2016, p. 304).

In contrast, I will follow Luo & Wang (2003), El-Geneidy et al (2016), Tenkanen et al (2016), Deboosere et al (2018) and Järv et al (2018) to use what Cheng & Jia (+ citation) call a “binary” distance decay, in which:

where X is some travel time cut-off. This undoubtedly distorts reality, as it treats all locations where as if they were equally attractive and locations where as if they were totally unattractive. However, the distortion is relatively easy to understand and does not require further assumptions and modelling. At any rate, El-Geneidy et al (2016, p. 303) and Cheng & Jia (+ citation) have found that accessibility measures calculated with binary and continuous distance decay functions using the same value of X are generally highly correlated. Most importantly, a binary measure maintains the raw accessibility values as simple ratios that can be interpreted directly (Luo & Wang, 2003; Deboosere & El-Geneidy, 2018). Specifically for this study, the quantities in R are interpretable as ratios of the number of potential jobs available to each potential worker within their catchment (viz. Hansen, 1959). I will term these quantities Job Provision Ratios (JPRs).

Having decided to use a contour measure, the choice of X then becomes significant, as larger values of X will result in greater spatial smoothing of local unevenness in accessibility levels (+ citation). I follow Cheng & Jia’s (+ citation) advice to choose X using empirical evidence of people’s actual willingness to travel and to reflect policy priorities. In 2017, a survey by the Department for Transport (+ citation, table tsgb0110) found that people who worked in London had an average commute time of 46min. TfL (2018, p. 224) also uses 45min as a travel time threshold for calculating job accessibility by public transport. As such, I will set X to 45min for this study (see Appendix X for an analysis of how sensitive my results are to other values of X).

*Job, Worker and Connectivity Data Needed for JPR Calculations*

Three components are needed to calculate the column vector R. The first two, MSOA-level counts of potential workers (that is, residents aged 16 to 64) and jobs, are easily obtainable from the Office of National Statistics (ONS) (+ citation). I use the most recent data available, which is for 2017.

The third component is a matrix of inter-area travel times, which is needed to build the matrix C. Most extant studies obtain this by compiling a multimodal transport network model and then simulating trips through the model (see Ford et al (2015, pp. 129-130) for a step-by-step guide). A significant benefit of this approach is that researchers have full knowledge of the underlying network data and routing assumptions used to generate the travel times, and can easily simulate alternative scenarios by adjusting the network and recalculating the times. However, the approach requires an extensive data compilation and cleaning effort to build a completely correct graph representation of the transport network.

A convenient alternative source of travel time data is Transport for London’s (TfL’s) Journey Planner (2019). This is a data service that provides the public with guidance on how to travel between any two points in London by walking or motorised public transport like buses, trams or trains (but not by car or taxi)[[1]](#footnote-1). This data source has several compelling advantages. First, it allows me to obtain travel time data without having to build my own transport network model for London, and the results reflect the network and routing assumptions actually used by TfL to assist real travellers in London. Secondly, the guidance comprehensively covers door-to-door travel. Thirdly, the data is highly disaggregated: besides information on the overall fare, distance and time needed for the trip, time, distance and travel mode are separately reported for each leg[[2]](#footnote-2) of the trip. Finally, the data is freely available in bulk through an Application Processing Interface (API), subject to a limit of 300 trip queries per minute (Transport for London, 2019).

However, using it in this study does impose some constraints. This is because the Journey Planner provides information only for trips that do not involve cars or taxis, and is most reliable for trips within the GLA boundary. As such, my analysis will concentrate only on areas within the GLA boundary and will exclude travel by car or taxi. Fortunately, TfL data (2018) shows that only 37% of trips within the GLA in 2017 used cars or taxis, while 80% of trips within the GLA boundary were by GLA residents. Nonetheless, I recognise that my results will be distorted by the fact that they omit connectivity by car and taxi (and thus will tend to underestimate accessibility levels, particularly in areas with sparser public transport coverage) and do not account for the competition effects of the job and labour pools outside the GLA boundary.

I used the Journey Planner to build a “baseline” database of inter-MSOA public transport travel times during a typical weekday AM peak, assuming all services run as scheduled. To do this, I submitted requests for information on trips from each MSOA’s population-weighted centroid to every MSOA’s job-weighted centroid. All requests were for trips ending at 0830 on Friday, 22 March 2019, so that the trips have to occur during the AM peak, but the requests were submitted over 1 month from January to February 2019. As the requests were made far in advance of the intended travel date, the information obtained assumes that all services will run as planned. Further, I assess for each trip the extent to which the travel time obtained using this method is representative of actual typical travel times during a weekday AM peak (see Appendix XX), and find no serious anomalies.

*Disaggregated Job Distribution Data*

The ONS reports not only the number of jobs in each MSOA, but also the breakdown of these jobs by Standardised Industrial Classification Codes (SICCs). This enables me to disaggregate job accessibility into 6 overarching sectors:

1. Primary and Extractive
2. Industrial
3. Retail, Logistics, Food & Beverage and Hospitality (hereafter consumer services)
4. Finance, Insurance, Real Estate, Professional and Administrative Support (hereafter high-end services)
5. Public Sector
6. Others

Then, I expand the bipartite network by creating one node per job sector in each MSOA (thus resulting in 6 nodes per MSOA) and assigning sector-specific job counts to each node. Thus, the column vector O expands to O\* of length np (where n = 983 and p = 6):

where os,d is the number of jobs in one of the p sectors s in MSOA d. Similarly, C expands to C\*, a m by np matrix of travel costs (where m = n = 983):

where co,s,d is the time needed for workers in MSOA o to travel to the jobs in sector s in MSOA d. I assume that the within-MSOA sectoral job distributions are all identical, such that co,s,d = co,d for all s.

Because the 2SFCA method distributes jobs to workers in an excludable way, Lorenz curves of the sector-specific job availability over all potential workers can be constructed and compared to the Lorenz curve for the overall job availability, to see the extent to which the spatial distribution of each sector’s jobs contributes to the unevenness in overall job availability. Also, the sum of each MSOA’s sector-specific JPRs will be equal to its overall JPR, that is . This allows me to calculate , which is the proportion of o’s JPR that is contributed by jobs in each sector s, which captures the extent to which o is dependent on s for jobs. This can help to illuminate two phenomena. First, high levels of dependence on a single sector for jobs may make an MSOA’s potential workers more vulnerable to sector-specific shocks like sudden disruptions to sectoral labour markets. Secondly, if an MSOA’s potential workers’ sociodemographic characteristics do not match the typical profile of employees in a sector, then increasing that MSOA’s dependence on that sector may increase the probability of sociodemographic mismatch between potential workers and the jobs that are available to them (+ citation).

*Capturing Variations in Characteristics of the Working-Age Population*

I also wish to investigate how job accessibility varies across different types of potential workers. If worker counts that are disaggregated by sociodemographic characteristics (like educational qualifications, income or disability status) were available, I would be able to similarly expand the vector W and the matrix C\* to include one node per worker type for each origin MSOA. Unfortunately, such disaggregated counts are only available from Census data, which are 8 years out of date at the time of this study.

In the absence of such data, I will use IMD (Department for Housing, Communities and Local Government, 2015) to reflect the relative degree of social deprivation experienced by potential workers living in different areas across a wide range of domains, including income, employment, health, education, crime, housing and the living environment. Following the official guidance, I aggregate the IMD scores to MSOA level using weighted averages of the raw scores for smaller spatial units. I then focus on the relative ranking of the areas by IMD index scores, rather than on the raw scores themselves.

*Inter-MSOA Travel Time Reliability Data*

As the Journey Planner provides travel guidance in real time, I also use it to obtain the travel times for selected origin-destination pairs over different days. The overall distribution of travel times per pair constitutes a measurement of its travel time reliability. However, this measurement probably underestimates the real level of travel time variability per pair.

Journey Planner travel times incorporate access (the time needed to walk to and from stations or between platforms in stations), waiting and in-vehicle times. Access times are based on assumed walking speeds between and within stations, and do not reflect in-station crowding that may reduce walking speeds around and in stations. Waiting and in-vehicle times are calculated solely on service schedules and do not reflect live vehicle locations. In particular, as long as vehicles are running along a route, the Journey Planner assumes all vehicles are running as scheduled, even if they are in fact delayed due to breakdowns or en route congestion. Also, the Journey Planner assumes that travellers will always be able to board the first suitable vehicle that arrives, even though in-vehicle congestion may prevent boarding.

Finally, the Journey Planner only partially accounts for the impact of service disruptions on travel options. If there are service suspensions (i.e. no services running) on sections of rail lines (National Rail, Underground, Overground, Docklands Light Railway (DLR), TfL Rail and tram) due to disruptions or planned shutdowns, the Journey Planner will not return travel options that require travel along the suspended sections. However, if services are just delayed but not completely suspended, the Journey Planner will still continue to return travel options via the affected sections as if service schedules were not affected (albeit with an additional alert that service frequencies are disrupted). Also, it assumes that walking and bus routes are never suspended. In particular, the Journey Planner will still recommend using a bus service between two stops, even if that service is actually diverted and service to either or both of the stops in question is suspended (although a text alert describing the extent of the diversion will be returned with the travel options).

As such, the travel time distributions obtained from the Journey Planner only reflect two sources of variability: the need to reroute travel to avoid service suspensions on rail lines, and fluctuations in wait times during transfers between services due to a lack of synchronisation between schedules of different services at different times during the day. Nonetheless, these two sources of variability are arguably of special interest to transport operations planners, as rail service suspensions have an outsize impact on travel times, and service schedules are one of the few aspects of transport operations that planners can directly control.

To obtain a distribution of travel times per trip during the AM peak on a typical weekday, I requested for travel information on selected trips beginning between 0600 and 0800 over 50 weekdays between 29 January 2019 and 25 April 2019. I found that there was only enough time in the two-hour window to process about 7,000 trip requests in real time, so I only gathered data on trips starting from all 983 MSOAs’ population-weighted centroids that end at the job-weighted centroids of 7 MSOAs that correspond to 7 key transport nodes evenly spaced across the GLA area, totalling 6,881 origin-destination pairs. See Appendix XX for details on how the 7 destinations were selected. I varied the start time of each trip every day by randomising the order in which requests for the 6,881 trips were submitted to the API. This allowed the data to capture the variation in schedule synchronisation between different services for each trip at different points during the weekday AM peak, and the resulting distribution of travel times per trip represents the trip-specific probability distribution of the travel time for any randomly selected start time between 0600 and 0800 on a typical weekday.

With the trip-level travel time distributions, I can calculate what Uniman (+ citation) defined as the Reliability Buffer Time (RBT), which is the difference between the 95th percentile travel time (for an unusually slow trip) and the median travel time (representing a typical trip). The RBT captures the additional travel time that a traveller needs to budget for in order to be 95% confident that a trip on any given day will not end late.

Because I also have baseline data for each trip that lays out its characteristics assuming arrival at 0830 and that all services run as planned, I can analyse how RBTs vary according to the baseline characteristics of each trip. Specifically, I will fit a set of linear regression models, the first of which is:

where:

* Y is a column vector containing 6,881 values where yi is the RBT for trip i;
* T is a column vector of 6,881 values where ti is the number of transfers between vehicles (that is, the number of non-walking legs minus 1) needed for trip i under ideal conditions;
* D is a 6,881 by f matrix where f is the number of travel modes and dg,i is the total distance covered by one of the f modes g for trip i under ideal conditions;
* α is an intercept term, β is a scalar representing the partial effect of transfers on RBT, γ is a column vector of length f representing the partial effects of distance covered by each mode on RBT, and ε is a column vector of 6,881 idiosyncratic disturbances.

This model investigates the impact on a trip’s travel time reliability resulting from the number of transfers (each transfer is an opportunity for additional waiting time due to a lack of synchronisation between the schedules of different services) and distance travelled by mode (so-called mode-specific effects) under ideal conditions.

Because all the MSOAs are connected to each other via the same transport network, there is likely to be a mutual correlation structure between the RBTs of different trips. For instance, two MSOAs that are located along the same rail line probably have similar RBTs for trips via that line to the same destination, as both trips will be affected by any disruptions along that line. If the non-independence of the RBTs between trips is not accounted for, the correlation structures among different RBT observations may distort the coefficient estimates for the covariates included in the model.

One way to deal with such correlations is to use a spatially autoregressive model, in which the value of the dependent variable in each area depends on the mean of the value of the dependent variable in its neighbours (+ citation). This captures the idea that areas that are near each other tend to be similarly connected to the larger network of areas and thus should experience similar network-related effects. To implement this, I first obtain for each MSOA o a vector of travel times to the 7 destinations from the baseline data, and then calculated the Euclidean distance between the vectors of every pair of MSOAs. Then, for each MSOA o, I identify the 3 other MSOAs (o1, o2 and o3) where this Euclidean distance is minimised; these are the MSOAs whose travel times to the 7 destinations are the most similar to those of MSOA o, assuming ideal conditions. Then, for each trip i originating from MSOA o and terminating at each of the 7 destinations d, I calculate the local mean RBT li for trips starting from each neighbour o1, o2 and o3 and ending at the same destination. That is:

For an illustration of where the 3 neighbours are located for the case of one MSOA, Stratford, see Figure XX.

With the local mean RBTs, I then fit the following model:

where L is a column vector with 6,881 values where li is the local mean RBT for trip i and δ is a scalar representing the partial effect of the local mean RBT on the RBT.

While the previous two regression models generate one global estimate for the partial effects of each covariate on the RBT of a trip, there are indications that mode-specific effects vary across locations. For example, it is widely reported that National Rail services in the southeast of London are particularly unreliable (+ citations). To investigate this, I refit the second regression model using locally-weighted regression (+ citation), which uses only the trips originating from the closest neighbours for each MSOA and ending in the 7 destinations to generate one estimate for each coefficient for each MSOA. By mapping out the coefficient estimates for each covariate per MSOA, we can examine how the mode-specific effects vary across space. I define as neighbours the 30 MSOAs whose travel times to the 7 destinations are most similar to those of the MSOA in question (using the same Euclidean distance calculation as the derivation of the local mean RBT). For Stratford, Figure XX illustrates which MSOAs are included in the locally-weighted regression and which are further considered for calculating the local mean of the interval width.

*Developing a Typology of MSOAs with K-Means Clustering*

Finally, I will synthesise the various aspects of job accessibility into a single result. Some studies attempt to do this by using a function to summarise values along multiple dimensions of accessibility (such as the spatio-temporal distance between jobs and workers and the sectoral diversity of available jobs) into a single value, but this is a lossy process that collapses multiple dimensions into one, thereby sacrificing the nuances in each dimension. The resulting accessibility scores also do not reflect how each underlying dimension contributes to the final score. The results can thus be hard to interpret, especially if the underlying dimensions reflect qualitatively different phenomena.

Instead, I use a k-means clustering analysis to group the MSOAs into k types across 9 dimensions:

* Overall JPRs, to capture overall job accessibility
* The percent of each MSOA’s JPR contributed by high-end services, to capture sectoral dependence for job allocations
* The IMD rank for each MSOA, to capture the sociodemographic characteristics of the potential workers
* The RBT for travel from each MSOA to the City, to capture travel time reliability to the most job-rich area in London
* I also identify which of the 6 other key destinations are closest to each MSOA and include the RBT for travel from each MSOA to its respective closest destination. In the absence of travel time distributions for every origin-destination pair, this RBT represents to some extent the reliability of travel times from each MSOA to its immediate neighbourhood.
* MSOA-level coefficients for distance travelled by bus, Underground, National Rail and walking from the locally weighted regression, to capture the effects on travel time reliability of the most commonly used travel modes

The typology of MSOAs obtained can capture how these dimensions intersect dynamically across London and highlight how each dimension contributes to the accessibility situation faced by each area. These results can then be used to help tailor a suite of interventions for each area to match its circumstances, as well as to prioritise interventions for areas that face more serious problems first.

**Results**

*Overall Job Provision Ratios*

Each area’s JPR is a function of 3 components: the spatial distributions of potential workers and jobs, and the connectivity provided by the public transport system. Figure XX visualises the first two components as a dot density map. In general, potential workers are quite evenly distributed across the GLA area. However, a ring pattern can be discerned, with lower residential densities near the centre, higher densities in a ring around the centre, and decreasing densities as one moves away from the centre. In contrast, the distribution of jobs is much more clustered, with extreme concentrations in the City of London, Westminster and Southbank. There are also smaller concentrations of jobs around some stations in the GLA’s rail network (black lines in the figure, including National Rail, Underground, Overground and DLR networks), most notably at Canary Wharf.

The third component, area-to-area connectivity, is captured in the matrix C. Each value of co,d in C is the minimum time it takes to travel from the population-weighted centroid of MSOA o to the job-weighted centroid of MSOA d to arrive at 0830 on a typical weekday, assuming that all services run as planned. It is difficult to visualise the entire matrix in a meaningful way, but Figure XX visualises the travel times for all trips originating from the most populous MSOA, Stratford, with all areas within 45min travel time from Stratford (i.e. the areas where f(cStratford,d) = 1) outlined in blue and a dot density map of the distribution of jobs superimposed. In the JPR calculations, each destination d within the blue boundary will contribute some of its jobs to each potential worker in Stratford; exactly how much d contributes depends on how many other potential workers live in areas where they can also reach d within 45min.

Figures XX, XX and XX visualise the steps needed to calculate the JPRs for each MSOA. First, Figure XX shows the number of potential workers who can reach each destination d within 45min, which is the result of . A dot density map of the distribution of potential workers is superimposed. It is clear that the public transport system heavily emphasises fast connections to the centre of the GLA area, with central areas being within 45min travel time of the most potential workers, and the number of potential workers able to reach each destination d declining as one moves away from the centre. Nonetheless, areas near major rail interchanges towards the outskirts may still be reachable within 45min by many potential workers.

Figure XX shows the number of jobs in each destination d that is allocated to each potential worker who can reach it, which is the result of . A dot density map of the distribution of jobs is superimposed. This ratio increases as the number of jobs in each destination d increases but decreases as the number of potential workers who can reach d increases. A clear ring pattern can be seen, as in the most centrally located MSOAs, the sheer volume of jobs available offsets the high number of potential workers who can reach them, while a ring of MSOAs around the centre have few jobs but are easily reachable from many areas around London (which spreads the limited jobs available more thinly across larger catchments), and the MSOAs near the outskirts have few jobs but are also relatively isolated from other areas (so that the few jobs available are shared only among small, sparsely populated local catchments).

Finally, Figure XX displays the JPR for each origin o, which is the sum of the per-worker job contributions over all destinations d that can be reached within 45min. This is the result from . In order to interpret whether an MSOA’s JPR is high or low, we need to compare it to an ideal case in which jobs, potential workers and connectivity were all evenly distributed throughout the GLA area, which would result in all MSOAs having JPRs equal to R0, the ratio of the total number of jobs to the total number of potential workers. This value in the GLA area is 0.88. To express the extent to which each MSOA’s JPR deviates from R0, JPR values are binned into 5 classes: 0 to 50% of R0 (thus there are far too few jobs available to meet the needs of the potential workers), 50% to 80% of R0 (the number of jobs available is somewhat insufficient for the number of potential workers), 80% to 125% of R0 (the number of jobs available is proportional to the number of potential workers; this is a balanced state), 125% to 200% of R0 (there are not enough potential workers within reach to fill the available jobs), and more than 200% of R0 (there are far too few potential workers within reach to fill the available jobs).

From Figure XX, we can that most MSOAs having unbalanced JPRs. Too few potential workers are able to reach the MSOAs near the centre within 45min to fill all the available jobs, while many MSOAs around the outskirts provide too few jobs within reach of each potential worker. A notable exception is the relatively more balanced JPRs in the west, where Heathrow Airport provides a significant number of jobs, and the connectivity between Heathrow and residential areas in the west is comparatively good. The uneven spatial pattern of job accessibility in London is similar to what other studies found in Boston (Shen, 1998), Montreal (El-Geneidy, et al., 2016), Santiago (Neihaus, et al., 2016), Bogota (Guzman, et al., 2017; Guzman & Oviedo, 2018), 11 major Canadian cities (Deboosere & El-Geneidy, 2018) and Los Angeles (+ citation), and is unsurprising overall.

Because the JPRs reflect a distribution of jobs among potential workers that treats jobs as excludable, it is straightforward to construct a Lorenz curve for the job distribution. This curve is in the inset of Figure XX, and shows that the bottom 50% of potential workers in the more job-poor areas share only roughly 20% of the jobs, while the top 10% of potential workers in the most job-rich areas have their pick of 20% of the jobs.

*Job Accessibility by Sector*

Next, I analyse how job accessibility varies by economic sector. The pie chart in Figure XX shows that 42% of all jobs in the GLA area are from the high-end services sector, while consumer services and the public sector make up another 24% and 22% respectively. Figure XX plots the distribution of jobs in each sector as a dot density map. The spatial distributions of industrial and consumer services jobs are the most decentralised, while public sector jobs are somewhat clustered around some rail interchanges. In contrast, high-end services jobs are mostly concentrated in the centre and at Canary Wharf.

The top left panel of Figure XX reproduces the map of the JPRs for all jobs from Figure XX for comparison. Each of the other 6 maps in Figure XX shows the sector-specific JPRs of each MSOA rs,o. Each map uses the same binning scheme as the top left panel, except that the values of R0 are now sector-specific. From the maps, we can see that the availability of industrial jobs across the potential workers is most balanced, while the availability of public sector and consumer services jobs are also relatively balanced albeit with significant over-provision at the centre of the GLA area. In contrast, the availability of high-end services jobs is very unbalanced, with almost no jobs from this sector available to potential workers living in large swaths of MSOAs in the south, east and north, while the number of jobs from this sector that is available per potential worker living in MSOAs in the centre is between 200% and 360% as high as the GLA average. The availability of jobs from the primary & extractive and other sectors is also very unbalanced, but there are so few jobs from these sectors in the GLA area that they do not make a significant difference to the overall situation.

When we plot Lorenz curves of the sectoral job distributions to all potential workers (lower right panel of Figure XX), we can assess the extent to which the sectoral job distributions contribute to the imbalance in the overall job distribution. The Lorenz curves for the industrial, consumer services and public sectors are closer to the diagonal than the curve for all jobs, but the curves for the primary & extractive, high-end services and other sectors are below the curve for all jobs. This indicates that the distributions of jobs in the latter three sectors tend to make the overall job distribution more imbalanced in the GLA area. In particular, because the high-end services sector is the largest in the GLA, the high concentration of jobs in this sector in the centre of the GLA area tends to drive the imbalance in accessibility to jobs in the GLA area.

Finally, we can decompose each MSOA’s JPR sectorally and analyse the extent to which potential workers in each MSOA are dependent on each sector for job availability. Each map in Figure XX visualises these sectoral dependence proportions , with the MSOAs binned into quintiles for visualisation in each map. Firstly, jobs in the primary & extractive and other sectors never make up more than 11% of any MSOA’s job availability. Secondly, areas near the outskirts in the northwest, north and east are especially reliant on industrial jobs, areas near the outskirts in the north, east and west (particularly around Heathrow) are unusually reliant on consumer services jobs, and two swaths along the northern boundary of the GLA area and south of the Thames are particularly reliant on public sector jobs. Finally, the most centrally located areas are disproportionately reliant on the high-end services sector.

Similar sectoral breakdowns of job accessibility are rare in the literature. However, Shen (1998, p. 355) and Cheng & Bertolini (+ citation) found for Boston and Amsterdam respectively that sectoral job distributions tend to resemble the overall job distribution. This contrasts with my findings, but the discrepancy may be accounted for by the structural differences in the economies of the three cities, where London’s much larger and more developed high-end services sector gives it a greater ability to price out firms in other sectors from the most accessible central areas, resulting in higher levels of sectoral specialisation in the city centre.

*Sociodemographic Differences among Potential Workers from Different MSOAs*

Potential workers living in different areas of London have different sociodemographic characteristics. MSOA-level variations in these characteristics are reflected in the IMD data. Figure XX visualises the ranking of each MSOA by IMD scores, together with 3 other indicators of social disadvantage, namely the estimated mean household income in 2015 (post-housing costs and equivalised for household size) (+ citation), the estimated proportion of households in poverty in 2014 (+ citation), and the unemployment rate among working-age residents in 2017 (+ citation). For all 4 maps, MSOAs are binned into deciles by their values for each indicator for the purposes of visualisation, with the 10% of MSOAs that are most (least) deprived visualised in dark red (green). In general, the spatial distribution of IMD rankings corresponds quite well to the distributions of the 3 raw indicators of social disadvantage (absolute Spearman rank correlation coefficients are between 0l8 and 0.9), with a swath of more deprived MSOAs running north-south along the Lea Valley and another swath running from Dagenham to Clerkenwell along the northern bank of the River Thames, and other significant clusters around Shepherd’s Bush, Elephant & Castle and Streatham.

*Analysis of Variability in Travel Times*

For each trip originating in all 983 MSOAs and ending in one of 7 destinations, I obtained minimum travel times over 50 days. Then, for each origin-destination pair, I use the distribution of travel times to calculate the RBT, which indicates the degree to which travel times for that pair are reliable. Each panel in Figure XX plots the RBTs of all trips terminating at one of the 7 destinations (outlined in red in each map). To facilitate comparison of RBTs across all origin-destination pairs, all maps use the same binning scheme to visualise the RBTs, with green shades indicating lower RBTs that are under 5min.

We can see that RBTs vary widely across space, from very low values for very short trips to more than 40min for trips from XXX to Streatham. This confirms the findings from Uniman (+ citation) and Ehrlich (+ citation) that travel times in the public transport network can be quite variable, even if the system operates entirely as planned. Some large-scale spatial patterns in RBTs are also discernible. For instance, areas that are directly connected to the destination by rail tend to have lower RBTs, as trips from these areas do not involve interchanges, so travel times are only affected by in-vehicle travel times. Areas that are not directly connected to a destination by rail tend to have higher RBTs, even if they are relatively near the destination. Trips from such areas tend to require more transfers, particularly if bus travel is needed, and this increases the probability of increased waiting times due to schedule desynchronisation between services. Also, trips originating from locations further from a destination tend to have higher RBTs, as these trips tend to involve more transfers too, and long-distance travel tends to rely more on rail-based modes and are thus more likely to be affected by service suspensions. These results are in line with findings by Uniman (+ citation), Ehrlich (+ citation) and Duran-Hormazabal & Tirachini (+ citation), who also found that trips by bus tend to have less reliable travel times than trips by rail, and that travel times for longer trips tend to be less reliable.

In addition, trips ending in destinations north of the Thames, particularly in the City but also in Finsbury Park, Harrow-on-the-Hill and Hammersmith, tend to have lower RBTs. This may be due to the more grid-like rail network north of the Thames, which provides commuters with more alternative rail routes if service on a particular section of the rail network is suspended. In contrast, travel times to Lewisham and Streatham tend to be less reliable. This probably reflects two factors: the more tree-like structure of the rail network there (which provides less network resilience against failures along the rail lines leading to these destinations) and reliance on National Rail services to reach these two destinations (these services tend to run less frequently than TfL’s rail-based services, thus increasing the likelihood of schedule desynchronisation during transfers). Finally, for all destinations, we can see that trips originating from the southeastern quadrant of the GLA area tend to higher RBTs. This may reflect the effects of the less frequent and less reliable rail services run by Southeastern and Southern Railways (+ citation).

*Contributions to Travel Time Variability by Travel Mode*

As we have seen, the level of travel time reliability for a trip seems to depend on the distance covered, the number of transfers, the reliability of different travel modes and location-specific factors. To investigate how these factors shape travel time reliability across trips, I carry out a set of linear regression analyses with trip-level RBTs as the dependent variable. The results for the initial model considering only transfers and distance travelled by each mode are shown in Column 1 of Table XX. They indicate that every additional transfer during the trip adds about 0.5min to the RBT. In addition, every additional kilometre travelled by bus or National Rail increases the RBT by 0.4min and 0.06min respectively, indicating that travel by these modes tends to reduce travel time reliability. Bus travel tends to involve more transfers, and frequencies on bus services tend to be lower than on TfL-operated rail-based services, which increases the risk of schedule desynchronisation during transfers to bus services. The result for National Rail is consistent with the fact that National Rail services are less frequent and more prone to service suspensions than TfL rail-based services. On the other hand, travel by Underground, tram or DLR tends to reduce RBTs.

Column 2 in Table XX presents the results for the spatially autoregressive model which includes the local mean RBT as an additional covariate. As expected, the local mean RBT allows the model to better capture the correlation between the RBTs of nearby MSOAs that arise due to similarities in how they are connected to the larger transport network. Specifically, as the local mean RBT for a trip increases by one minute, the RBT of the trip in question tends to increase by 0.5min. The inclusion of the local mean also reduces the confounding of the estimated coefficients for the other covariates due to network effects, reducing the magnitude of the estimates by about 40% for the number of transfers, 35% for bus and 55% for National Rail. However, the coefficients still indicate that bus travel reduces travel time reliability the most, followed by National Rail, and that travel by Underground, tram or DLR tend to increase travel time reliability.

Next, we examine the results for the locally weighted regression. The relevant diagnostics are in Appendix XX. Column 3 in Table XX displays a summary of the distributions of the coefficient estimates for each covariate obtained from the locally weighted regression. The distributions of local estimates contain the global estimates from Column 2, but capture the variation in each coefficient across space. Figure XX maps out the coefficient estimates for each covariate and each MSOA. Green (red) shades indicate that the covariate tends to increase (decrease) travel time reliability for trips originating in that MSOA. Grey MSOAs indicate where the coefficient estimates were not statistically significantly different from 0 or were not available (because there was no variation in that covariate among the trips considered for the local regression focused on that MSOA).

At the local level, increasing travel by bus tends to decrease travel time reliability almost everywhere in London, but particularly in the northeast around Stratford, Barking and Leytonstone, and around the City, Westminster, Rotherhithe, Lewisham, Sidcup, and in a swath from Waterloo to Wimbledon. Travel by National Rail degrades travel time reliability in a large swath in the southeast, indicating further evidence of operational unreliability in Southern and Southeastern services. In fact, National Rail services in these areas seem to be so bad that increasing the number of transfers and increasing travel by bus and walking (and thus reducing reliance on National Rail) tend to improve travel time reliability, contrary to patterns elsewhere. However, travel by National Rail also reduces travel time reliability in a northwestern corridor from Paddington to Ruislip, and around Stratford. This could reflect how most National Rail lines north of the Thames terminate outside the City (in stations like Paddington, Marylebone, Euston, Kings Cross, St. Pancras and Liverpool Street) and do not provide direct orbital connections, so National Rail users need to transfer more to complete their journeys to one of the 7 destinations. On the other hand, the local effects of travel by Underground, tram, Overground and DLR on travel time reliability tend to be positive or not statistically significant. Travel by the new TfL Rail service tends to reduce travel time reliability, particularly in the west; this may be because TfL Rail operates on tracks that are also used by National Rail services, and thus its operating patterns and reliability may resemble National Rail services more.

*Typology of London MSOAs by JPR, Sectoral Dependence for Job Allocation, IMD Ranks, Travel Time Reliability and Mode-Specific Contributions to Reliability*

Finally, I carry out a k-means clustering analysis to explore how job accessibility, sectoral dependence for job allocations, social deprivation among potential workers and travel time reliability intersect among different MSOAs. The results for a clustering with 8 clusters is shown in Figure XX. By social deprivation, three types of clusters are apparent: Clusters 0, 1 and 2 tend to contain MSOAs with lower deprivation, Clusters 5, 6 and 7 tend to contain MSOAs with higher deprivation, and Clusters 3 and 4 are more diverse.

Cluster 3 (with about 8% of the working-age population) contains the most centrally located MSOAs in the City, Westminster and Southbank. As these areas are the most job-rich, MSOAs in Cluster 3 are distinguished by having the highest JPRs, indicating extreme surpluses of jobs available per potential worker. They are also among the most dependent on the high-end services sector for jobs. Their central location gives them access to many public transport routes, especially rail-based routes, resulting in very reliable travel times to the City and to their local destinations. Reliability of travel by mode tends to be average, except for travel by National Rail, which tends to be less reliable. In contrast, the MSOAs in Cluster 4 are distinguished by their combination of job poverty and high travel time unreliability, especially to the City. This unreliability is disproportionately due to the unreliability of National Rail services in the southeast, so much so that trips that involve more distance covered by bus, walking or Underground (and thus rely less on National Rail) tend to be more reliable. Fortunately, only about 2% of potential workers live in MSOAs in Cluster 4.

Of the three least socially deprived clusters, Cluster 0 (with 15% of the working-age population mostly located in the southwest in areas like Richmond, Kingston, Wimbledon and Battersea) tends to have the most balanced JPRs, indicating sufficient jobs available per potential worker. These potential workers tend to rely more on high-end services for their job availability, but because of their low social disadvantage, they may be more likely to have the skills and resources to take advantage of the available jobs in this sector. Travel times to the City and to local destinations tend to be quite reliable, and indeed this area seems to enjoy the most reliable Underground and National Rail services in London. MSOAs in Cluster 1 (with 23% of the working-age population and covering areas near the outskirts) also enjoy relatively reliable travel times to the City and local destinations, but experience job shortages due to their greater distance from the centre and the scarcity of jobs in the outskirts. In contrast, MSOAs in Cluster 2 (with 10% of the working-age population located near the outskirts) experience both insufficient job availability and high travel time unreliability, particularly due to National Rail services.

The situations facing more socially deprived residents are quite diverse. MSOAs in Cluster 5 (with 18% of the working-age population) are located quite centrally, almost surrounding the MSOAs in Cluster 3. As a result, despite being home to more socially deprived residents, these MSOAs enjoy similar conditions as those in Cluster 3, benefitting from excellent transport connectivity, high travel time reliability and a surplus of jobs available per potential worker. However, these jobs tend to be from the high-end services sector, which more socially disadvantaged workers may find more difficult to take advantage of due to a lack of educational qualifications, relevant working experience or social networks. Of course, the high-end services sector does contain some jobs suitable for workers from less privileged backgrounds, but the degree to which the sheer volume of jobs available can compensate for the tendency for a sociodemographic mismatch between high-end services jobs and more socially disadvantaged workers is unclear from this analysis.

MSOAs in Cluster 6 (containing 18% of the working-age population) are located further towards the outskirts, especially in the Lea Valley in the north, areas from Ealing to Hayes & Harlington along the TfL Rail line in the west, areas along the Overground line from Brockley to Croydon in the south, and Becontree and Dagenham in the east. They are characterised by relatively reliable travel times but low job availability. Travel seems to be reliable but too slow to connect these areas effectively to large job clusters. In contrast, MSOAs in Cluster 7 (containing 7% of the working-age population in Stratford, Ilford, Barking, the Royal Docks and Greenwich) have relatively balanced JPRs and reliable travel times to the City, but less reliable travel times to local destinations. The reliability of buses and National Rail in these areas seems especially problematic. These areas are also more dependent on the high-end services sector for jobs. Thus, these areas may also be more vulnerable to a sociodemographic mismatch between potential workers and the available jobs.

In summary, the k-means clustering analysis indicates that accessibility problems come in diverse shapes and sizes, as different areas may experience distinct combinations of insufficient job availability, dependence on the high-end sector for jobs, heightened social deprivation among potential workers and/or travel time unreliability. The particular combination of challenges in each area suggests that intervention strategies should be tailored to each area’s characteristics. But there are also areas where socially disadvantaged workers benefit from high job accessibility and travel time reliability. This is similar to El-Geneidy et al’s (2016) and Deboosere & El-Geneidy’s (2018) findings in Canadian cities that indicate that a significant proportion of low-income workers have good accessibility to jobs that are suitable for them, although problem areas where social disadvantage coincides with low accessibility exist. In contrast, studies of South American cities (Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018) found that there is a systematic tendency for low-income workers to live in less accessible areas; this is probably due to their lack of affordable housing near job-rich centres and less extensive public transport systems.

**Discussion**

*Recommendations for Urban Developments in the GLA*

The findings from this study can aid in developing multifaceted strategies to improve accessibility to suitable jobs in London, drawing on options from land use development, transport infrastructure and operations and economic and social policy. Firstly, the map of overall JPRs in Figure XX reveals locations where there are too few or too many jobs available per potential worker. This can directly guide the distribution of future growth to shape the overall urban structure, with new housing supply being directed to locations with job surpluses, and job growth being encouraged in areas with too few jobs available per potential worker. Transport connectivity between areas with high JPRs and low JPRs can also be enhanced, to put more potential workers in job-poor areas within reach of areas with job surpluses. This may be especially effective where nearby areas have very different JPRs.

Secondly, even though the measurement of travel time reliability in this study probably underestimates actual travel time variability, the results are still useful for identifying areas where travel times to important destinations like the City are relatively unreliable. The mode-specific reliability analysis also indicates areas where services by different modes seem to be particularly unreliable. Moreover, because the travel time variability measurements in this study only account for service suspensions and schedule desynchronisation and are not confounded by factors relating to travellers’ behaviours such as crowding, they can directly justify efforts to adjust service frequencies and arrival times to reduce the probability of schedule desynchronisation during transfers, or to adjust the incentives for operators so that they put more effort into preventing service suspensions.

Thirdly, the typologies of MSOAs derived from the k-means clustering analysis can be used to target social and economic policies to reduce the potential for sociodemographic mismatch between potential workers and the available jobs. For instance, referring to Figure XX, the potential workers of MSOAs in Clusters 5 and 7 may especially benefit from skills training programmes, which can help bridge the gaps between their relatively high social deprivation and dependence on high-end services jobs. Economic incentives or development strategies to encourage job growth in other sectors, especially consumer services and the public sector, can also be targeted to MSOAs in these clusters.

Finally, the MSOA typology can also be used to prioritise interventions to benefit the most socially deprived areas first, especially since more socially deprived residents may have fewer resources to improve the accessibility situations that they face by themselves, and may have fewer options for travel besides public transport.

To a certain extent, these recommendations are in line with what the GLA already plans to do to improve transport connectivity and distribute growth more equitably throughout London. For instance, the upcoming Elizabeth Line is expected to place many more workers in the east and west of London within 45min travel time of the job-rich centre, and this should help to move the JPRs in areas along the route towards a more balanced state. The Elizabeth Line may be particularly important for improving job accessibility for Ilford and Chadwell Heath in the east, where a shortage of jobs coincides with higher social deprivation.

A similar effect can be expected from the proposed Crossrail 2 project, which runs from the north through the socially deprived and job-poor Lea Valley to the southwest (+ citation). The Mayor’s Transport Strategy (Greater London Authority, 2018) also calls for improvements to bus connectivity and reliability (pp.155-9), improvements to rail reliability south of the Thames (particularly through the conversion of National Rail services to Overground services through “metroisation”) (pp.180-1), and the encouragement of housing and job growth in areas with good connections by public transport (p.214). The findings from this study can add further justifications for pursuing these strategies.

Nonetheless, the findings can aid in prioritising areas for the implementation of these strategies. For instance, the leg of Crossrail 2 that runs north of the Thames benefits more deprived areas than the southern leg; this should justify building the northern leg first. Also, the Transport Strategy (2018) recommends encouraging job growth in four strategic transport interchanges, Lewisham, Stratford, Clapham Junction and Willesden Junction, to put more jobs within easy reach of workers. This is a sensible strategy, as these four locations are surrounded by areas with low JPRs (from Figure XX). However, the number of working-age residents that can reach Lewisham within 45min travel (shown in Figure XX) is relatively low, and the RBTs for travel to Lewisham (shown in Figure XX) are quite high; this indicates that travel speeds and reliability to Lewisham should be improved in tandem with the delivery of more jobs there. This strategy of dovetailing transport improvements and development is probably applicable to the other 3 locations as well.

Overall, the results paint a picture of job accessibility in London that is hopeful. Problem areas do exist where high social deprivation, low job accessibility, high sectoral dependence and low travel time reliability coincide to varying degrees, but it is important to recognise that there are also areas where more disadvantaged residents enjoy job accessibility conditions that are similar to their more privileged neighbours. In addition, TfL, the GLA and UK governments have recognised the importance of job accessibility to enhancing residents’ quality of life, and have demonstrated a commitment to improving the situation, in particular to benefit more socially deprived areas, through measures that have already been implemented or plans that they are already pursuing. This study affirms the efforts made so far, and provides some guidance for how to tailor ongoing efforts to more effectively address circumstances that are particular to different locations.

*Extending the Use of Real-Time Trip Planning Data for Research*

This study has demonstrated how trip planning data can provide very detailed transport data in real time without the effort needed to build a multimodal transport network model in which to simulate trips. Nonetheless, other sources of large-scale transport data can potentially provide even more detail with less distortions; for instance, automated fare collection data records the actual travel behaviours of real travellers through the public transport network. However, such records also tend to be highly protected and difficult to access in real time. In such circumstances, real-time trip planning data can be an attractive second choice to obtain transport data with approximately the same level of detail.

Nonetheless, the methodology used in this study to measure trip-level travel time reliability with Journey Planner data tends to systematically underestimate travel time variability. This can be partially addressed in future studies by combining multiple sources of TfL open data. For example, once the route for a trip is obtained from the Journey Planner, the stations where a traveller needs to board and alight can be identified, and separate scripts can then repeatedly request the live arrivals for each station from a separate API endpoint in order to track the vehicles that the traveller would ride on in real time. Actual waiting and in-vehicle times for each leg of the trip can then be obtained and updated into the database of trip times. This would enable travel times to reflect the effects of slower in-vehicle speeds and delayed arrivals due to en route congestion.

*Potential Innovations for the 2SFCA Method*

As mentioned above, the 2SFCA method actually involves a bipartite network with weighted edges between origin nodes and destination nodes. This implies that the 2SFCA method can be generalised for measurements of resource availability in partially or even entirely non-spatial settings, as long as those problems involve a set of “origin” nodes that demand a resource, a set of “destination” nodes that can provide that resource, and edges between origins and destinations with weights that reflect the extent to which the resource can flow from destination to origin. Thus, a straightforward extension would involve incorporating non-spatial distances into the incidence matrix C. For instance, co,d can be a weighted average of the normalised spatial distance between o and d and the normalised Euclidean distance between the sociodemographic characteristics of the workers living in o and those working in d. In this way, the values of f(co,d) and ultimately the values of ro can be affected by both spatial proximity and sociodemographic matching, potentially capturing the idea that the availability of jobs per worker depends on both how distant the worker is from the jobs and how closely the worker’s characteristics match those needed by the jobs.

A large body of literature on bipartite graphs goes beyond measuring levels of resource availability to address problems of resource assignment. For instance, given the set of workers and jobs from above and the hybrid spatial-sociodemographic distance between each worker-job pair, it is possible to find an optimal assignment scheme that links each worker to exactly one job so as to minimise the sum of the job-worker distances over all the job-worker pairs thus formed. Many algorithms exist to solve such assignment problems (+ citation). It may be interesting to analyse how the edge weights of the bipartite graph affect the total sum of distances in the optimal assignment scheme, which could give some insights into which pairwise distances should be reduced (such as by reducing travel times or increasing sociodemographic matching) in order to improve the global optimum. Also, it could be revealing to analyse how the availability of jobs per worker per origin as measured by 2SFCA is related to distance from that worker to the assigned job in the optimal assignment scheme; we may expect that workers with higher job availability should tend to be matched with less distant jobs, but workers with lower availability may force other workers in high-availability areas to be matched with more distant jobs, so that the overall total of job-worker distances can be minimised.

However, this is only a brief foray into the possibilities for innovative developments of 2SFCA from a graph-theoretic perspective. Needless to say, there is broad potential for further research to improve and apply the 2SFCA method in domains outside urban and transport studies, and there are also many opportunities for methods developed for bipartite graphs to yield further insights into the structure of accessibility in urban areas.

# **References**

Bettencourt, L. & West, G., 2010. A unified theory of urban living. *Nature,* 10, Volume 467, pp. 912-913.

Deboosere, R. & El-Geneidy, A., 2018. Evaluating equity and accessibility to jobs by public transport across Canada. *Journal of Transport Geography,* 12, Volume 73, pp. 54-63.

El-Geneidy, A. et al., 2016. The cost of equity: assessing transit accessibility and social disparity using total travel cost. *Transportation Research Part A: Policy and Practice,* 9, Volume 91, pp. 302-316.

Ford, A., Barr, S., Dawson, R. & James, P., 2015. Transport accessibility analysis using GIS: assessing sustainable transport in London. *ISPRS International Journal of Geo-Information,* 1, Volume 4, pp. 124-149.

Geurs, K. T. & Wee, B., 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography,* 6, Volume 12, pp. 127-140.

Greater London Authority, 2018. *Mayor's transport strategy: March 2018.* London: Greater London Authority.

Guzman, L. A. & Oviedo, D., 2018. Accessibility, affordability and equity: assessing `pro-poor' public transport subsidies in Bogotá. *Transport Policy,* 9, Volume 68, pp. 37-51.

Guzman, L. A., Oviedo, D. & Rivera, C., 2017. Assessing equity in transport accessibility to work and study: the Bogotá region. *Journal of Transport Geography,* 1, Volume 58, pp. 236-246.

Hansen, W. G., 1959. How accessibility shapes land use. *Journal of the American Institute of Planners,* 5, Volume 25, pp. 73-76.

Hu, Y. & Downs, J., 2019. Measuring and visualizing place-based space-time job accessibility. *Journal of Transport Geography,* 1, Volume 74, pp. 278-288.

Järv, O. et al., 2018. Dynamic cities: location-based accessibility modelling as a function of time. *Applied Geography.*

Kotavaara, O., Antikainen, H., Marmion, M. & Rusanen, J., 2012. Scale in the effect of accessibility on population change: GIS and a statistical approach to road, air and rail accessibility in Finland, 1990-2008. *The Geographical Journal,* 3, Volume 178, pp. 366-382.

Levine, J., 1998. Rethinking accessibility and jobs-housing balance. *Journal of the American Planning Association,* 6, Volume 64, pp. 133-149.

Lucas, K., 2012. Transport and social exclusion: where are we now?. *Transport Policy,* 3, Volume 20, pp. 105-113.

Lucas, K., Wee, B. & Maat, K., 2015. A method to evaluate equitable accessibility: combining ethical theories and accessibility-based approaches. *Transportation,* 3, Volume 43, pp. 473-490.

Neihaus, M., Galilea, P. & Hurtubia, R., 2016. Accessibility and equity: an approach for wider transport project assessment in Chile. *Research in Transportation Economics.*

Sanchez, T. W., 1999. The connection between public transit and employment: the cases of Portland and Atlanta. *Journal of the American Planning Association,* 9, Volume 65, pp. 284-296.

Shen, Q., 1998. Location characteristics of inner-city neighborhoods and employment accessibility of low-wage workers. *Environment and Planning B: Planning and Design,* Volume 25, pp. 345-365.

Social Exclusion Unit, 2003. *Making the connections: final report on transport and social exclusion.* London: Office of the Deputy Prime Minister.

Tenkanen, H. et al., 2016. Health research needs more comprehensive accessibility measures: integrating time and transport modes from open data. *International Journal of Health Geographics,* 15(1).

Transport for London, 2018. *Travel in London: report 11.* London: Transport for London.

Transport for London, 2019. *Plan a journey,* London: Transport for London.

Transport for London, 2019. *Transport for London unified API,* London: Transport for London.

1. The Journey Planner also returns cycling options, which tend to be significantly faster than bus or walking for short trips. However, because only 2% of trips trips in London were by cycling (Transport for London, 2018), I omit these options from this study to avoid underestimating travel times. [↑](#footnote-ref-1)
2. A leg is a part of the trip that is covered by the same vehicle. For example, a trip that involves a walk from home to bus stop A, a ride on a bus to stop B, another bus ride from stop B to stop C and a walk from stop C to a workplace involves 4 legs. [↑](#footnote-ref-2)