Candidate number: 311990
Word count: 6399

TP19: Spatial Networks and Human Mobility: An Application of the Intervening Opportunites Model to the London Cycle Hire Scheme

Supervisor: Dr. M. A. Porter

Human mobility drives a variety of physical phenomenon in the modern world. Our understanding of human mobility has been dominated by two theoretical models: the gravity model and the intervening opportunities (IO) model. In this thesis we derive these models with spatial network theory formalism and consider their application to the London Cycle Hire Scheme (LCHS). We demonstrate theoretically why an opportunistic model is more appropriate than the Gravity model and show that the most appropriate opportunistic model is an intervening opportunities model modified with a theory of Spatial Dominance (IOSD) for the case of LCHS. Furthermore, we also challenge the convention of ignoring self-loops commonly employed when analysing physical networks, and we show how self-loops found in the LCHS can be successfully assimilated into the application of the IOSD.

I. BACKGROUND

Although it is a difficult task, modelling human mobility is extremely important if we hope to understand a diverse range of physical systems from modelling virus pandemics [2] to traffic flows [3]. The difficulty stems from the myriad of unknown factors associated with both human-decision making and the physical systems themselves. However, it has been shown that strong spatial and temporal patterns can exist [1] - in human mobility and models have been developed that successfully describe a variety of networks.

The understanding of human mobility is limited by the ability to collect data about individuals. Given the increasingly sophisticated ability to track individuals, modelling human mobility has seen rapid progression, and it is now possible to test models against complex systems with interesting dynamics (e.g. Levy flights [1]). However the field is still dominated by two theoretical models developed in the early studies of human mobility, the gravity model and intervening opportunities (IO) model. The gravity model has existed for a significant time, first introduced to describe trade flows in 1781 by Monge [5], with its modern form appearing in a 1946 paper by Zipf [6]. The IO model was introduced in 1940 by Stouffer [4] as an alternative theoretical model to predict human migration.

In Section II of this report, we give a brief description of simple networks and a brief description of the mathematical formulations we use to describe physical systems. Whilst the models are independent of any specific mathematical representation, for our application in understanding physical systems, we have chosen to use network theory formalism. In the following sections, we derive the gravity model (Section III) and IO model (Section IV). Furthermore, in Section V, we describe a modification to the IO model known as the theory of spatial dominance to form the intervening opportunities model with spatial dominance (IOSD), that acts to include elements of the gravity model into the IO model.

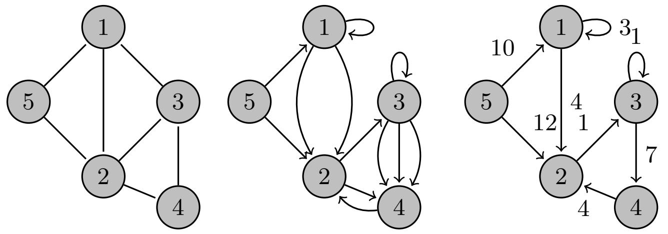
In Section VI, we investigate the application of the IO and IOSD models to the London Cycle Hire Scheme (LCHS). We first discuss how we are able to model the LCHS as a mathematical network and justify certain parameter choices we make in translating physical properties to mathematical objects. Following this, we discuss our choice of models for the network. In a recent paper by Austwick et al. [7], they chose to model the LCHS with a gravity model approach. We, however, chose to use a opportunistic approach. We will discuss our motivations for suggesting, from first principles, why an opportunistic model would be more appropriate than the gravity approach presented by Austwick et al. [7]. Furthermore, after generating an IO trip distribution for the LCHS, we demonstrate that an IOSD model is more appropriate theoretically and generates a better trip distribution.

In Sections VII and VIII we delve deeper into understanding the LCHS system. In Section VII we follow Austwick et al. [7] in treating weekend trips and weekday trips separately and discuss the differences between the trip distributions. In Section VIII we discuss the nature of self-loops and the typical treatment of ignoring them in physical systems. Furthermore, we assimilate self-loops into our IOSD model as applied to the LCHS.

II. NETWORKS

Before we consider physical networks, we first discuss basic network (graph) theory and how we have used it to describe simple networks.

Networks vary in complexity depending on how one chooses to model them. In its most basic representation, a network consists of a collection of nodes connected by edges. A network can be represented as a labelled graph or via incident or adjacency matrices [8] (examples of which are shown in Fig. 1).



(a) Simple Network (b) Directed network with multi-edges (c) Directed network with weighted edges

$$\begin{pmatrix} 0&1&1&0&1\\ 1&0&1&1&1\\ 1&1&0&1&0\\ 0&1&1&0&0\\ 1&1&0&0&0 \end{pmatrix} \quad \begin{pmatrix} 1&2&0&0&0\\ 0&0&1&1&0\\ 0&0&1&3&0\\ 0&1&0&0&0\\ 1&1&0&0&0 \end{pmatrix} \quad \begin{pmatrix} 3&4&0&0&0\\ 0&0&1&1&0\\ 0&0&1&7&0\\ 0&4&0&0&0\\ 10&12&0&0&0 \end{pmatrix}$$

(d) Adjacency matrix for (a) (e) Adjacency matrix for (b) (f) Adjacency matrix for (c)

FIG. 1: Various networks represented graphical and by corresponding adjacency matrices. (a) A simple graph i.e. a graph that is undirected, unweighted, has no multi-edges and no self loops. (b) A non simple graph with multi-edges and self-loop. (c) A non simple weighted graph, where each edge label represent the weight of the edge. (d) Adjacency matrix representation of graph shown in (a). (e) Adjacency matrix representation of graph shown in (b). (f) Adjacency matrix representation of graph shown in (c).

Two nodes are considered "adjacent" if an edge connects them. From this we can define the adjacent matrix as:

$$A_{ij} = n, \tag{1}$$

where n is the number of edges connecting node i to node j. Incident matrix representation is an analogous representation where we consider two edges to be "incident" if a node connects them. Incident matrix representation is of little interest in physical application where edges are not well-defined, as is the case of the LCHS.

Furthermore, networks are sub-classified as directed and undirected. A directed network is where flows are restricted to a single direction across edges, and undirected are where flows are bidirectional along edges. Note that adjacency matrices for undirected networks must be symmetric, and for networks with no multi-edges the adjacency matrix is only populated with 0s and 1s. We restrict our discussion to undirected networks with no multi-edges.

Before being able to understand the trip-distribution models, we must first choose how we define three important network characteristics, node importance, edge length and self-edges.

Node Importance: In most physical networks, the nodes are not equal. This inequality is due to either certain zones experiencing greater flows or, certain zones being more "well connected" than others (e.g. some nodes have greater numbers of edges originating from them). One can use measures of node importance to help characterise these differences.

There are many ways of measuring node importance. The simplest definition is to define a node's importance as its "population". In the case of a city, this could be the human population; or in the case of a traffic network, this could be the number of trips originating or terminating at the zone. More complicated definitions define this as a function of other parameters. For example, we will later define node importance as a function of time of day when applying models to the LCHS.

Edge Length: In any network, nodes are connected by edges, and these do not remain uniform between different zone pairs. This is of great importance when considering spatial networks as edge length is often used to represent the topography of the surface the network is embedded.

There are a variety of ways of transforming physical connections between nodes to the mathematical edges that we need to model. Simple definitions include a linear relationship of Euclidean distance between pairs of zones to their respective edge length, whereas more complex definitions account for specific characteristics of routing distances and factors that affect journey time.

Self-Edges (Self-Loops): Self-edges, also known as self-loops, occur when an edge connects a node to itself. They are represented by non-zero elements in the diagonal of the adjacency matrix A (see Fig.1) Self-loops can be very problematic in understanding spatial networks. Most investigations ignore self-loops, as either the flows through the loops are small compared to the non-zero non-diagonal elements of A, or they will not mathematically affect parameters being calculated. Further, in the case of the LCHS, due to the method of tracking used, they make it very difficult to properly model the network. At first, we will ignore self-loops in our application to the LCHS, but we will return to them later in Section VIII.

III. THE GRAVITY MODEL

As its name suggests, the gravity model is based on the functional form of the Newton's law of Gravitation. Gravity-law based models appear in a variety of social and natural sciences, and human mobility is no exception. The gravity model has been used to describe trip distributions from as early as the 18th century by Monge [5]. The way in which it is presented today can be traced back to Voorhees' application of the gravity model to traffic in urban areas [9].

The gravity model states that the trip-interchange between two nodes is directly proportional to the attraction between the two nodes and inversely proportionally to some function of the spatial separation between the nodes. The model assumes trip makers will aim to minimise the 'cost' of a trip. The "cost" of a trip is usually a function of spatial separation that also contains the information on other parameters that affect the trip distribution [9].

To derive the gravity model for trip distribution, we first consider the gravity model of Newtonian physics that relates the force, F_{ij}, between objects i and j, to the objects masses, m_i and m_j, and the respective separation, d_{ij}, between the bodies:

$$F_{ij} = G\frac{m_i m_j}{d_{ij}^2}, \tag{2}$$

By analogy, we take equation (2) and make the following changes to instead predict the number of trips, T_{ij}, between m_i and m_j. Consider i and j as nodes instead of bodies; replace the m_i term with O_i, where O_i represents the number of trips originating from node i; replace the m_j term with D_j, where D_j represents the number of trips that terminate at node j; replace the dependence of d_{ij} with a cost friction factor, C_{ij}^{β} (where β is some constant, which is measures empirically) that is designed to incorporate the distance between the zones and any other parameters that affect the trip distribution. This gives:

$$T_{ij} = \theta\frac{O_i D_j}{C_{ij}^{\beta}}, \tag{3}$$

where θ is some proportionality constant.

A more general form of the gravity model is given as:

$$T_{ij} = \frac{O_i D_j f(C_{ij}) K_{ij}}{\sum_n D_n(C_{in}) K_{in}}, \tag{4}$$

where K_{ij} are "zone-to-zone adjustment factors", K_{ij}. Although originally introduced with no theoretical basis, attempts have been made to attribute them to "socioeconomic influences on travel otherwise unaccounted for in the model" [15]. Derivation of equation (4) from (3) can be seen in appendix A. There are two commonly used functions for $f(C_{ij})$:

$$f(C_{ij}) = C_{ij}^{-\beta}, \tag{5}$$

$$f(C_{ij}) = \exp(-\beta C_{ij}), \tag{6}$$

Equation (6), known as the "doubly constrained gravity model" [10], was initially derived by Wilson [11] to solve the divergence issue that occurs with equation (5) for small values of C_{ij}.

IV. THE INTERVENING OPPORTUNITIES (IO) MODEL

The opportunistic approach to modelling networks was first proposed by Stouffer [4], who applied this approach to migration patterns between services and residences. The theory was further developed by Schneider [12] to the general framework that is used today.

The law of intervening opportunities as proposed by Stouffer [4] states "The number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities." An "opportunity" is a destination that a trip-maker considers as a possible termination point for their journey and an "intervening opportunity" is an opportunity that is closer to the trip maker than the final destination but is rejected by the trip-maker [4]. Mathematically, we have:

$$T_{ij} = k\frac{D_j}{V_j}, \tag{7}$$

where k is a proportionality constant, D_j represents the total number of opportunities at node j, and the quantity V_j is the number of intervening opportunities between nodes i and j.

Schneider proposed a modified Stouffer hypothesis [12]: "The probability that a trip will terminate in some volume of destination points is equal to the probability that this volume contains an acceptable destination times the probability that an acceptable destination closer to the origin of the trips has not been found."

This was represented mathematically by Ruiter [13] as

$$dP = L[1 - P(V)]dV, \tag{8}$$

where dP is the probability that a trip will terminate when considering dV possible destinations; the 'subtended volume' V is the cumulative total number of destination opportunities considered up to the destination being considered; dV is an element of the subtended volume at the surface of the volume; $P(V)$ represents the opportunity that a trip terminates when V destinations are considered; L is a constant probability of a possible destination being accepted if it is considered. The solution of equation (8) is

$$P(V) = 1 - \exp(-LV), \tag{9}$$

With the expected trip-interchange, T_{ij}, we get

$$T_{ij} = O_i[P(V_{j+1}) - P(V_j)], \tag{10}$$

where O_i represents the total number of opportunities at node i. Note that Eash [14] showed that the gravity and IO models are "fundamentally the same" and are both derivable from entropy maximization theory. Eash also noted that the difference is how the "cost" of travel is considered [14]. Although the gravity model considers this "cost' as a function of distance, the opportunity model considers the "cost" as the difficulty to satisfy a trip's purpose. The gravity model then treats the distance variable as a continuous cardinal variable, whilst the opportunity model treats the distance as an ordinal variable.

At this juncture it is necessary to recognise the major shortcoming of the conventional intervening opportunities model. As described by Pooler [16]: "The model implies that the linear distance between nodes has no direct effect on the perception of opportunities, and that, decision-makers perceive all opportunities with equal clarity, which is very unlikely Harris (1964) [18]; Okabe (1976) [19]". This issue is addressed with the introduction of the spatial dominance into the model that assimilates features from the gravity model into the intervening opportunities model to create what is known as the "Intervening Opportunities model with Spatial Dominance" (IOSD).

V. IOSD - INTERVENING OPPORTUNITIES WITH SPATIAL DOMINANCE

The IOSD was introduced by Pooler [17] in 1992. He used the theory of Spatial Dominance to model the idea that "decision-makers are influenced not just by the size of a destination or distance but rather by these two factors in combination". To do this, one applies the functional dependence on distance seen in the gravity based models to the ranking process of the intervening models. We define the"theoretical spatial dominance", p_{ij} of a destination j on an origin i [17] in equation (11).

$$p_{ij} = \frac{D_j f(d_{ij})}{\sum_{j=1}^{m} D_j f(d_{ij})}, \tag{11}$$

where d_{ij} is the distance between nodes i and j; D_j is the attractivity of node j. Once p_{ij} has been determined for all nodes, we use this as a basis for the ordinal ranking required in the IO model. In essence, this redefines an intervening opportunity as "the destination which exerts the greatest amount of spatial dominance on an origin' [17].

By applying the theory of spatial dominance, we hope to address the problem that trip-makers, are unlikely to perceive every opportunity equally.

We can assimilate the gravity models functional dependence on d_{ij} (dependancies of the form given in equations (5) and (6) in place of $f(d_{ij})$ in equation (11).

VI. MODELLING LONDON CYCLE HIRE SCHEME (LCHS)

Background: The London cycle hire scheme (officially known as Barclays Cycle Hire [20]) is a bicycle sharing scheme launched in July 2010 across the city of London and surrounding boroughs [21]. Originally, it consisted of 315 docking stations that with approximately 5,000 bicycles [21], however it has since grown to 720 stations and 10,000 bikes in 2013 [22]. We have acquired data that consist of approximately 5,000,000 trips that occurred on the scheme between 31 May 2011 and 4 February 2012. Full information about the network and data analysed are in Table I

As mentioned in Section III, a recent paper by Austwick et. al [7] applies the gravity model to the LCHS and compares results against the same analysis of a variety of other similar schemes in major cities situated in the US. Although this is reasonably successful, we postulate that the opportunistic model is more appropriate; our motivation for suggesting the IO model as more appropriate lies with the assumption of how decisions are made by trip makers with regards to their journeys.

Choice of Model: The gravity model for trip distribution is derived by modifying Newton's law of gravitation hence, in its derivation it does not attempt to justify trip distributions from a microscopic level. Rather, it focuses on a macroscopic understanding of the system, with the basic premises that trip-interchange decreases with increasing distance and increases with increasing node importances. The IO model, however, is derived on assumptions that outline a core mechanism, i.e. the Stouffer hypothesis [4], that drive the macroscopic properties of the system.

Furthermore, we postulate that trip-makers do not view distance as a cardinal variable when making their decisions but rather as an ordinal variable. Our rationale is based on two facts: trip-makers are not restricted to using the cycle scheme to complete their journeys; trip-makers likely have a final destination that is not situated exactly at a single station but in proximity of many stations.

To illustrate the first reason, there exists comprehensive bus and metropolitan ("Tube") networks that cover the same region as the cycle network and have stations in comparable locations and density to the cycle network. This, in theory, mitigates the dependance of the trip distribution on the relation between physical exertion and distance travelled. Such a relationship would likely treat distance as a cardinal variable in a similar method to the gravity model.

Our second point - of final destinations being proximate to many stations - supports the core idea of using intervening opportunities as the driving force behind a trip-maker's decision-making. If this is the case, trip-makers typically have to judge a few destinations and determine - based on their proximity to a final destination - their suitability for termination of their cycle journey. This resembles Stouffer's hypothesis [4], which we described in Section III.

However, there is an inherent problem with the IO model that suggests it will have difficulty describing a modern mobility network. It assumes that trip-makers do not judge potential destinations based on their importance. We guess that trip-makers consider destination size when ranking destinations. That is, trip-makers will travel further to destinations of great importance. The perceived distance of a destination should decrease with node importance. This suggests that the IOSD model (see Section IV) will better describe the network than the IO model.

Modelling Parameters: Before applying the IO model, we need to discuss what we consider an "opportunity" and how we would model the edge length.

Opportunties: The number of opportunities at any

LCHS information		
Period Of Analysis	31/03/2011–4/02/2012	
Number of Stations at end of period	496	
Number of Bikes	5000	
Period	**Number of Trips**	**% of all Trips**
For entire period: with self-loops	4,761,197	100
For entire period: without self-loops	4,588,207	96.4
For entire period: only self-loops	172,990	3.6
Weekdays only without self-loops	3,322,792	69.8
Weekends only without self-loops	1,265,415	26.6

TABLE I: A table containing the parameters (number of nodes, period of analysis, number of bikes, total numbers of trips for given subsets) of the LCHS that is the subject of the analysis of this thesis.

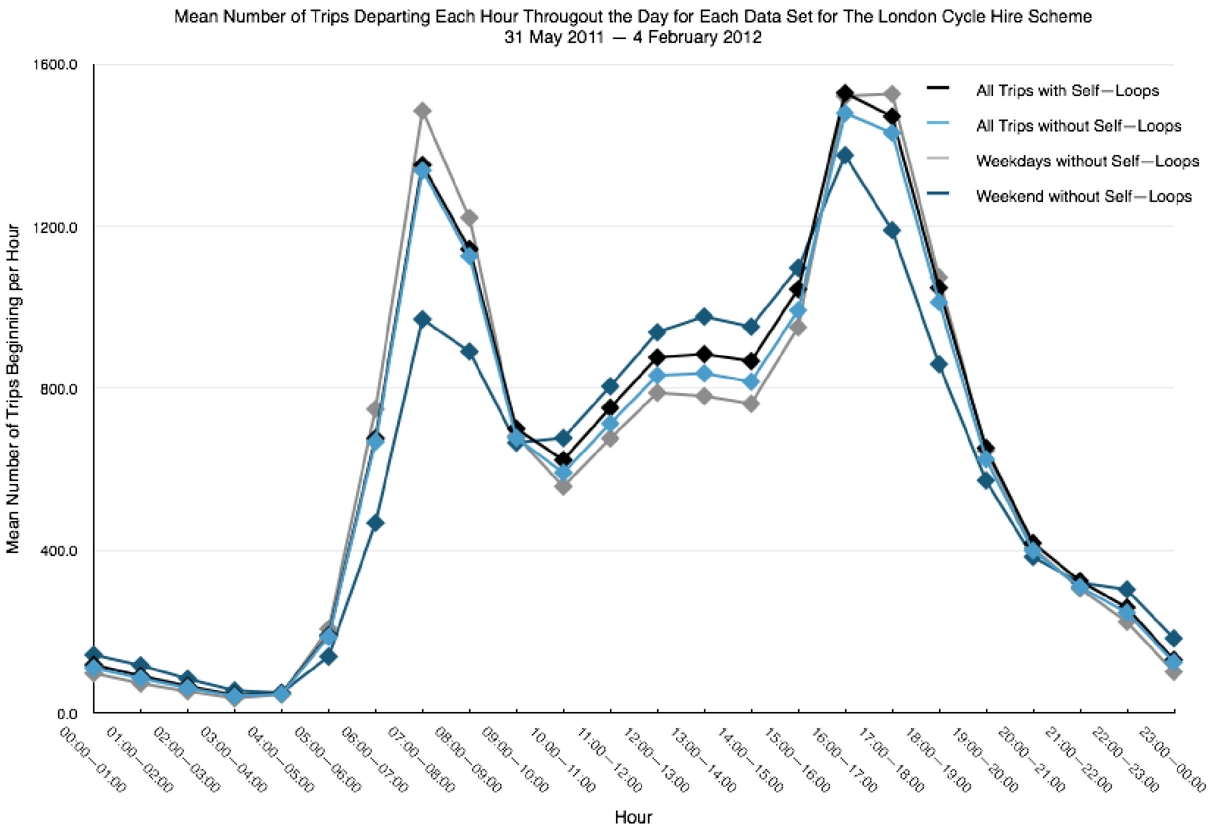


FIG. 2: Plots of the daily mean number of trips departing per hour throughout the period (31/03/2011–4/02/2012) for the following data sets: all days including self-loops; all days not including self-loops; weekdays not including self-loops; weekends not including self-loops.

given node is a difficult parameter to estimate from the data we have for the LCHS. In general, it would be difficult to estimate as the definition of an "opportunity" varies between trip-makers. We choose to use Ruiter's definition [13] that the number of opportunities at a station is the number of trips that originate at that station.

Distance Variable d_{ij}: Due to the large number of stations that we need to consider (496) and the space in which they are embedded, determining the distances to use between stations is a challenging task. The simplest definition is a Euclidean distance between each station. However, it is evident that trip-makers cannot follow straight lines for their journeys. Ideally, we would use the distance between each pair of stations following the most common route that trip-makers take. In principle, this information is accessible, however due to limitations on relevant services and software required for this task, the time required to generate this information is beyond the timescale for this project. In Section IX, we suggest this as a course for future research.

We thus take the mean journey time between two stations for the period measured as the distance between stations. This is an irregular choice, because distance parameters typically are not temporally-based. However, the models do not require that the distance parameter must be a spatial one. Furthermore, journey time contains relevant information, that we argue makes mean journey time a more appropriate a metric than physical distance. This is largely due to the idea that trip makers are likely to be sensitive to journey time, and that traffic effects are likely to change the "perceived distance" between nodes. Furthermore, this allows us to consider changing edge lengths (that are functions of time of day).

Austwick et. al took the approach of using Euclidean/Great Circle distances as edge lengths as, in their opinion, "Euclidean/Great Circle distance is at least free from these additional assumptions", where "these additional assumptions" refer to assumptions that would have to be made if routing information was used instead. Our use of mean journey time would also be subject to similar assumptions. However, this is not as great a problem as we are using an opportunistic model where we treat distance as ordinal rather than cardinal i.e. since we do not care about the absolute values of distances between nodes (rather we only care about the relative order of distance length) we are able to make assumptions about edge length as long as they do not greatly affect the rankings of destinations.

Model results: Given the definitions for the opportunities of each node and the edge length between each pair of nodes, we can now analyse the data to determine the nature of the trip distributions and compare against the results found by Austwick et. al.

When applying our model, we to consider the network at each hour of the day. We split our data into hour sets by considering only trips that departed their origin in that hour. We calculated the mean journey time between each pair of nodes and created a ranking matrix that ranked destinations, j, for each origin, i, based upon the "temporal distance" between them. We then follow the theory given in Section IV for the IOs model to determine the probability, $P(V)$, of a trip terminating, after V opportunities are considered. Then the probability is obtained by dividing the number of trips occurring against the total number of trips.

When fitting distributions to the data, we applied a nonlinear least-squares method Mathworks MATLAB Curve Fitting Toolbox™[23]. The methodology behind the fit is given in Appendix B.

Upon plotting the model, we noticed two distinct behaviours (Fig. 3) throughout the day. From (23:00–06:00), we see the behaviour demonstrated in Fig. 3a for each hour segment. It is evident that this probability distribution follows the expected distribution given by equation (9), so we can apply the model to generate a set of L values for each hour segment in the period. For the hours of (06:00–23:00) we see the behaviour given by Fig. 3b for each hour segment. It is evident that this probability distribution does not follow the expected probability distribution given by equation (9). We give the results of the fitting procedure for IO model in Table III.

We also employed the IOSD model (see Section V). We chose to consider both spatial dominance functions that we discussed earlier [see equations (5) and (6)]. After observing the results we qualitatively determined that d_{ij}^{-b} produced distributions that better fit the predicted model. We trialled the following values for β: $0.5, 1, 2, 3, 5, 10, 100, 1000$ and determined that the trip-distirbution best resembled the IOSD model prediction when β was in the range 1–3 (within this range the choice of beta did not affect the rank-ordering). We settled with $\beta = 2$ and plotted the hour segments shown in Figs. 3c and 3d, representing the corresponding hours shown in Figs. 3a and 3b for the IO model. We give the results of the fitting procedure for the IOSD model in Table III.

Discussion: From examination of the total number of trips as function of time of day (see Fig. 2), it is clear that the network has two distinct periods of use primary period, (06:00–23:00), and the secondary period, (23:00–06:00).

First, considering the IO model, it is evident from the construction that the model only predicts the behaviour well for the *secondary period*, Fig. 3a. However, for the *primary period*, we can see from Fig. 3b that the distribution deviates greatly from the predicted behaviour (seen by the large dip underneath the predicted distribution). This is troubling, as the period (06:00–23:00) is the period, when the vast majority of the journeys occur, where the secondary period (23:00–06:00) is of less interest. This is strong evidence to suggest that a basic IO model using a temporal distance variable is not suitable for the LCHS network.

The IOSD model is able to resolve this issue. From Fig. 3c and Table III, we see that the model not only retains the characteristic form of the secondary period (23:00–06:00) for the IO model, but produces a better fit, as indicated by a lower RMSE (standard error) for this secondary period. Furthermore, comparing Fig. 3d to Fig. 3b, suggests that the IOSD models resolves the "dip" in the basic IO model and has a form that better resembles the produced behaviour seen in equation (8). However, qualitatively, it is clear that deviations of the trip distribution from the expected model still exist and the IOSD model requires further modification to provide a comprehensive description of the network.

If we consider the subtle differences of the IO and IOSD models, the success of the IOSD model must be due to how trip-makers rank possible destinations. This supports the notion that trip-makers consider both a destination's size as well as its distance when deciding a journey destination.

Comparison to Austwick et. al paper "The Structure of Spatial Networks and Communitied in Bicycle Sharing Systems" [7] As mentioned in previous sections, Austwick et. al take the approach of applying a gravity model to five cycle sharing schemes around the world. Rather than test the validity of their model, their focus is comparison of the results of their model between the schemes and then a more detailed look at certain characteristics of the networks:

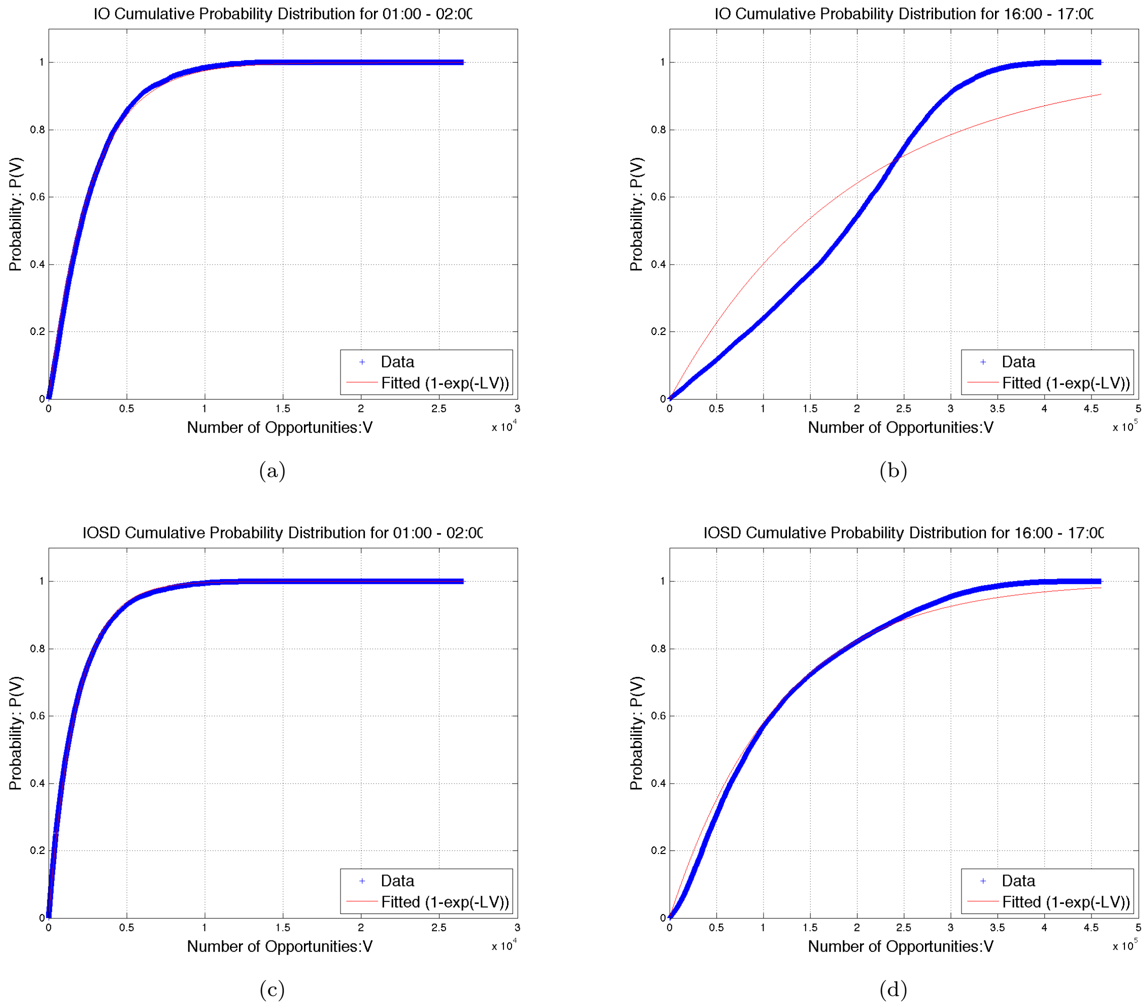


FIG. 3: Plots of the probability distributions for both IO and IOSD models representing the two behaviours demonstrated by the models. (a) IO model plot of hour (01:00–02:00) demonstrating the general behaviour of network in the period (23:00–06:00). (b) IO model plot of hour (16:00–17:00) demonstrating the general behaviour of network in the period (06:00–23:00). (c) IOSD model equivalent plot of (a). (d) IOSD model equivalent plot of (b).

time-dependance; seasonality; detection of communities ("subregions within the bike sharing flow networks which are linked to one another more strongly than nodes from other subregions"). Furthermore, Austwick et. al. choose to use a Euclidean/Great Circle method for calculating distances.

Austwick et. al are able to show that all the schemes they analysed behave similarly, with trip distributions bearing reasonable similarity. They show the overall shape of the "probability mass function of journeys" (probability distribution function of trips made against Euclidean distances) is consistent between schemes, but there exists a large deviation between the networks that suggests that a better model exists for cycle schemes. Furthermore, their calculated model's dependance on distance is not a standard functional dependance as we would expect from the gravity model (see equations 5 and 6).

In comparison, although our model suffers similar problems, we are able to generate trip distributions of the overall functional form expected of our IOSD model. However, during the secondary period, the data clearly deviates from the expected model. Our IOSD model results offer a slight improvement in that we are able to fit the standard form for the IOSD model to our trip distribution.

Furthermore, we have been able to show our model is viable, whilst using a more complex distance variable than Austwick et al. Our use of mean journey time over Euclidean distance is preferable as it includes the effect of other important characteristics of the cycle hire schemes, namely traffic conditions.

There are a few factors that might be the cause of the discrepancies between the IO and IOSD models. The more likely issue is the use of "mean journey time" as the distance variable. However, as we have seen in

Network Characteristics	Number of Trips				Number of Edges			
	With Self-Loops	Without Self-Loops			With Self-Loops	Without Self-Loops		
Hour of Day	Everyday	Everyday	Weekdays Only	Weekends Only	Everyday	Everyday	Weekdays Only	Weekends Only
"Primary Period" (06:00–23:00)								
06:00–07:00	668.6	676.4	748.8	468.3	12963	12669	10604	5245
07:00–08:00	1337.0	1350.6	1484.1	970.2	30872	30503	26941	14245
08:00–09:00	1126.1	1143.1	1220.4	890.8	53215	52811	47688	27380
09:00–10:00	677.5	700.5	682.2	665.8	57465	57054	50209	29836
10:00–11:00	592.2	623.3	558.0	677.3	55230	54807	45976	27945
11:00–12:00	713.3	752.1	676.5	805.0	56766	56340	45822	29449
12:00–13:00	831.8	876.2	789.1	938.2	61210	60777	50051	32749
13:00–14:00	837.0	884.3	780.9	976.8	64754	64327	53462	35729
14:00–15:00	816.3	867.2	762.0	951.7	65665	65242	53891	36590
15:00–16:00	992.2	1044.2	950.3	1096.8	65400	64970	53582	36038
16:00–17:00	1478.4	1528.1	1520.1	1374.6	68736	68308	56946	39058
17:00–18:00	1429.6	1470.1	1525.7	1189.8	78613	78192	68057	44514
18:00–19:00	1012.3	1047.5	1073.6	859.5	79754	79337	70394	42091
19:00–20:00	625.4	653.1	646.5	572.8	70701	70286	61682	34767
20:00–21:00	400.2	418.7	406.4	384.5	57123	56708	28298	26220
21:00–22:00	310.5	325.1	306.4	320.8	44570	44160	36739	18947
22:00–23:00	247.3	259.6	224.5	304.0	37529	37122	30154	16034
"Secondary Period" (23:00–06:00)								
23:00–00:00	124.8	130.8	101.3	183.5	31804	31406	23875	15017
00:00–01:00	110.5	118.4	97.5	142.9	19440	19074	12884	9828
01:00–02:00	85.4	91.0	72.7	117.1	18316	17963	9733	8371
02:00–03:00	61.8	66.1	52.9	83.9	15031	14680	7502	7033
03:00–04:00	41.3	44.1	35.7	55.3	11610	11300	5212	5113
04:00–05:00	46.4	48.4	45.1	49.7	8227	7967	4571	3532
05:00–06:00	187.3	190.7	206.7	138.8	6796	6566	10604	2786

TABLE II: Daily mean number of trips departing their origin per hour and the number of edges in the network per hour for the following data sets: All days including self-loops; All days not including self-loops; Weekdays only not including self-loops; Weekends only not including self-loops.

Fitting Results	Without Self-Loops								With Self-Loops	
	IO		IOSD		IOSD Weekdays only		IOSD Weekends only		IOSD Everyday	
Hour of Day	$L(10^{-5})$	RMSE (10^{-5})	$L(10^{-5})$	RMSE (10^{-3})	$L(10^{-5})$	RMSE (10^{-5})	$L(10^{-5})$	RMSE (10^{-5})	$L(10^{-5})$	RMSE (10^{-5})
"Primary Period" (06:00–23:00)										
06:00–07:00	2.1	71.3	4.0	5.6	5.5	5.2	31.5	10.3	4.0	6.4
07:00–08:00	0.8	111.9	1.4	19.6	1.9	17.5	10.2	12.7	1.4	19.8
08:00–09:00	1.1	100.8	1.5	36.6	2.0	31.7	10.6	23.4	1.5	36.6
09:00–10:00	2.0	91.5	2.6	47.3	4.1	40.2	16.1	32.3	2.6	47.2
10:00–11:00	2.3	94.6	3.0	45.6	5.1	38.9	15.0	34.1	2.9	45.3
11:00–12:00	1.6	102.4	2.3	43.7	3.8	37.4	11.1	34.3	2.2	44.1
12:00–13:00	1.3	104.3	1.8	41.5	3.0	34.9	8.6	32.3	1.7	42.1
13:00–14:00	1.3	102.4	1.7	42.3	2.9	36.1	7.8	31.8	1.7	42.1
14:00–15:00	1.3	99.7	1.7	39.5	3.0	32.6	7.9	30.3	1.7	40.0
15:00–16:00	0.9	105.5	1.4	31.7	2.4	24.8	6.4	27.0	1.4	32.2
16:00–17:00	0.5	121.9	0.9	27.4	1.3	21.7	4.6	23.3	0.9	27.2
17:00–18:00	0.5	115.9	0.8	33.4	1.2	27.7	5.3	24.3	0.8	33.0
18:00–19:00	0.9	99.5	1.3	33.2	1.9	26.6	8.8	22.2	1.3	32.5
19:00–20:00	1.9	81.8	2.6	29.6	3.9	22.8	16.9	19.5	2.5	29.3
20:00–21:00	3.7	62.8	4.7	28.2	7.6	22.1	32.6	11.3	4.6	26.9
21:00–22:00	5.3	49.5	6.7	20.8	11.3	16.0	41.5	3.4	6.4	19.9
22:00–23:00	7.6	43.0	9.6	15.9	18.6	12.7	46.7	2.6	9.1	15.6
"Secindary Period" (23:00–06:00)										
23:00–00:00	23.6	21.8	31.4	2.1	80.4	1.6	120.2	5.7	30.0	2.3
00:00–01:00	28.8	6.5	31.7	3.3	70.8	1.7	183.6	7.8	29.1	3.2
01:00–02:00	37.0	10.7	55.8	5.5	134.9	8.7	253.2	11.9	49.6	5.5
02:00–03:00	63.1	7.0	95.1	8.7	229.9	7.8	470.0	11.1	85.4	7.3
03:00–04:00	133.5	3.0	190.5	7.4	455.6	6.7	944.0	10.1	169.5	8.4
04:00–05:00	152.5	8.8	278.6	6.6	580.7	5.9	1718.9	8.1	250.0	7.4
05:00–06:00	16.3	23.2	35.2	1.1	47.2	10.0	373.0	9.6	34.4	10.0

TABLE III: Fitted L values and associated standard error (RMSE) for IO and IOSD models as applied to the London Cycle Hire Scheme

the secondary period (23:00–06:00), the models that use this "temporal-distance" are able to describe the trip-distribution well. This suggests that there are other possible issues, the first being the effect of recreational vs professional use, and the second being the ignorance of self-loops. Both of these are likely to have effects and are discussed in Sections VII and VIII.

VII. WEEKDAY VS WEEKEND USE

Following the example of Austwick et. al, we investigate the difference between weekday and weekend use. We predict a difference between trip distributions for weekdays and weekends to reflect that we expect greater commuter use during weekdays and greater leisure use at the weekend.

Examination of Fig. 2 (daily mean of number of trips departing their origin per hour for different data sets) demonstrates there is a difference in use of the LCHS between weekdays and weekends.

Both data sets show two peaks in trip numbers during the same hours of the day (04:00–10:00) and (16:00–21:00). Weekends have maxima at (07:00–08:00) and (16:00–17:00), whilst weekdays have maxima at (07:00–08:00) and (17:00–18:00). Furthermore, the weekday peaks are equally tall and both wider and taller than the corresponding weekend peaks. This supports the notion that weekday use of the LCHS is dominated by commuter use, as the peaks are situated at the hours directly before and after the working day (typically 09:00–17:00). Furthermore, the morning peak for weekend use is 30% smaller than the afternoon peak, and during periods between peaks (07:00–08:00) and (16:00–17:00), there is greater use than that seen during weekdays. The combination of these, provides evidence the idea of commuter use being less dominant during weekends.

We applied the IOSD model via the same method as applied to the full data set, detailed in section VI. Fitting results of this process for both weekdays and weekends are given in Table III, whilst plots of the total number of trips as a function of hour of day are give in Fig. 2.

As in Section VI, we examined the fits and were able to determine that the behaviour of the network for both weekday use and weekend use were not qualitatively different from the the overall network use. For both data sets there is a clear difference in behaviour during the "primary period" (06:00–23:00) and "secondary period" (23:00–06:00). Furthermore, there was little difference between weekday and weekend use, beyond the difference in total trip use throughout the day, as discussed in the previous paragraph.

This leads us to conclude that beyond evidence of greater commuter use during the weekday morning period, the network can be modelled similarly throughout the week.

VIII. SELF-LOOPS

We will now return to discussing self-loops. As discussed in Section II, self-loops are situations where an edge connects a node directly to itself, represented as a non-zero diagonal element in the adjacency matrix. They are difficult to understand in physical systems where we only have origin-destination data on trip-makers, as is the case of the LCHS. With typical treatment of journeys, they appear as journeys for which the distance $d_{ij} = 0$. Application of a zero d_{ij} to the standard gravity model (5) provides a nonsensical result due to the divergence of the model as d_{ij} approaches zero.

Choosing to ignore self-loops when applying network theory is typically used as a solution as often flows along self-loops are small compared to flows along other edges. In Austwick et. al's paper [7] on cycle hire schemes, they take this approach with the justification that "displacements are not reasonably calculable for these journeys" [7].

In prior sections, we took the same approach as Austwick et. al in our treatment of self-loops. Self-loops constitute 3.6% (see Table I) and thus cannot be considered insignificant for this network. Therefore, our IOSD model without self-loops provides an incomplete understanding of the network as a whole and needs to be addressed.

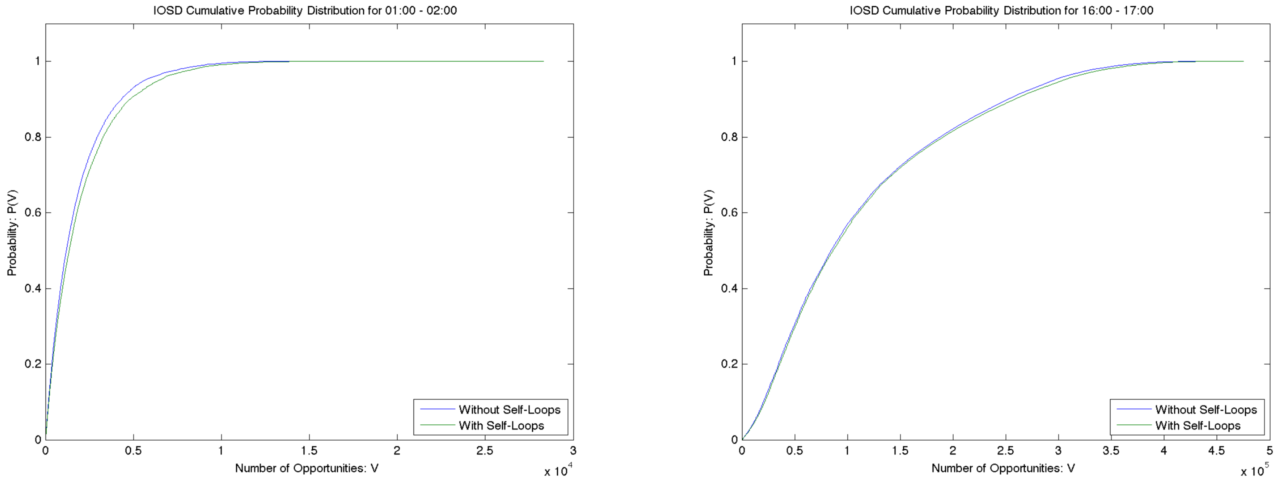
Typically, we are unable to assimilate self-loops into standard gravity models and IO models, as the available data does not allow us to know the edge length of any trip that starts and ends at the same destination. However, our choice of using mean journey time as the edge length offers a solution to this problem. Using the same treatment for self-loops as we have for the rest of the network, we model each node's self-loop with an edge length equal to the mean journey time for all self-loop trips and then follow the steps detailed in Section V.

Using the IOSD model (as used in Section VI), the fitted results of the modelling of the network are shown in Table III and the self-loop distributions for hours (01:00–02:00) and (16:00–17:00) are shown in Fig. 4a and 4b, alongside the distributions without self-loops included.

From Fig. 4, we observe that our inclusion of self-loops with identical treatment to the other trips does not cause any significant deviation from the distributions without self-loops. Furthermore, the small difference of fitted L values between with and without self-loops data sets (see Table III), this suggest that self-loops do not affect the trip-distributions greatly and ignoring self-loops is a reasonable approach when attempting to understand the global properties of the network. However, the difference supports our premise that self-loops are relevant for a complete understanding go the system. In summary, the exclusion of self-loops causes an our model to underestimate the probability that a trip will terminate when a trip makers considers an opportunity.

IX. SUMMARY

The aim of this project was to assess the suitability of the intervening opportunity (IO) model as an alternative to the gravity model for describing human mobility patterns seen in the LCHS. By applying the model, we showed that the IO model can successfully model the trip distribution during the daily "secondary pe-



(a) (b)

FIG. 4: Plots of the probability distributions for the IOSD models with and without self-loops. (a) IOSD model plot of hour (01:00–02:00) comparing the distributions with and without self-loops. (b) IOSD model plot of hour (16:00–17:00) comparing the distributions with and without self-loops.

riod" (23:00–06:00), but unsuccessful in predicting the behaviour of the LCHS during the "primary period" (06:00–23:00). In efforts to adapt the model to better describe trip makers motivations for journey choice, we employed Pooler's theory of spatial dominance [16] to form an Intervening Opportunities Model with Spatial Dominance (IOSD), that was shown to resolve the "dip" deviation of the IO model seen when modelling the primary period, whilst retaining the IO model's success in modelling the secondary period.

We then investigated the network in more detail by considering the validity of treating weekday and weekend use independently and the validity of ignoring self-loops when modelling the network. Our investigations into daily use revealed that there is a difference between weekday and weekend use, specifically related to morning use (08:00–10:00) that we attributed to commuter use during weekdays. We also demonstrated that ignoring self-loops is unlikely to affect results from modelling the network significantly.

For further investigations, it would be most interesting from a theoretical perspective to produce a spatial distance matrix that represents the shortest routes between origins and destinations. Doing this would allow comparisons to be drawn between the changing temporal distance and the fixed spatial distances to better understand how trip makers perceive changing conditions (e.g traffic, etc.). Furthermore, because routes are likely to be one way accessible, so it would be interesting to employ directed networks rather than the undirected ones that we examined in this report.

It is also worth investigating other models. Most notably, the recently developed radiation model of human migration in [25] is of interest, because it aims to derive from first principles a model of human migration that does not have any of the problems associated with the popular gravity models.

Throughout this project, we have demonstrated our understanding of the LCHS is limited. In our application of the IOSD model, we demonstrated that the an opportunistic understanding of LCHS is reasonable. However, further information about the network (complex routing and distance information) would allow us to produce a more evolved model that generates more precise predictions of trip-interchange and further our understanding of the use intervening opportunities in understanding human mobility.

Acknowledgements

The author wishes to thank several people. I would like to thank Dr. Mason Porter (University of Oxford) for his supervision throughout this project as well as assistance and guidance in writing this paper. Furthermore, I would like to thank Dr. Mariano Berguerisse Díaz (Imperial College London) and Transport for London for providing the data analysed in this project.

Appendix A: Gravity model

Starting with equation (3), we can eliminate θ by enforcing the conservation of origins. The conservation of origins states that the total number of trips from an origin to all destinations must be equivalent to the originations from zone i, or mathematically

$$O_i = \sum_{j=1}^{n} T_{ij}, \tag{A1}$$

We can then find θ from equations (3) & (A1):

$$O_i = \sum_{j=1}^{n} T_{ij} = \sum_{j=1}^{n} \theta \frac{O_i D_j}{C_{ij}^{\beta}}, \tag{A2}$$

This gives θ as:

$$\theta = \sum_{j=1}^{k} \frac{D_k}{C_{ij}^{\beta}}, \tag{A3}$$

We now rewrite C_{ij}^{β} as $f(C_{ij})$, which, from equation (3) yields:

$$T_{ij} = \frac{O_i D_j f(C_{ij})}{\sum_n D_n(C_{in})}, \tag{A4}$$

This is, in fact, the first form of the gravity model as applied to trip distributions. However, it is important to note that this is not the most common form of the model. It is actually the case that historically, this form tended not to yield accurate predications and hence it is normally slightly modified to improve accuracy. This is done by introducing zonal adjustment factors, K_{ij} that have no theoretical basis and must be empirically determined from the data:

$$T_{ij} = \frac{O_i D_j f(C_{ij}) K_{ij}}{\sum_n D_n(C_{in}) K_{in}}, \tag{A5}$$

Appendix B: Nonlinear Least Squares Fitting

Throughout our analysis we have employed nonlinear least square fitting with a Trust-Region-Reflective least squares algorithm as provided by Mathworks MATLAB Curve Fitting Toolbox™ [23].

The following information regarding the fitting used in this report is sourced from Mathworks Documentation Center [23].

Mathworks defines least squares as "the problem of finding a vector x that is a local minimiser to a function that is a sum of squares, possibly subject to some constraints":

$$\min_x ||F(x)||_2^2 = \min_x \sum_i F_i^2(x), \tag{B1}$$

Upon generating a fit, we use the root mean squared error ($RMSE$) to determine the goodness of each fit. The $RMSE$ is defined as the root of the sum of squares due to error (SSE):

$$RMSE = \sqrt{SSE/v}, \tag{B2}$$

where v is the number of independent pieces of information invoking the n data points that are required to calculate the sum of squares and SSE is defined as:

$$SSE = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2, \tag{B3}$$

where x_i are actual observations, $\hat{x_i}$ are the estimated values and w_i is the weighting factor.

[1] González, M.C., Hidalgo, C.A., Barabási, A., (2008) *Understanding Individual Human Mobility Patterns* Nature 453, 779--782

[2] Bajardi, P., Poletto, C., Ramasco, J.J., Tizzoni, M., Colizza, V., Vespignani, A., (2011) *Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic.* PLoS ONE 6(1): e16591.

[3] Ellis, J.B., Van Doren, C.S., (1966) *A comparative evaluation of gravity and system theory models for statewide recreational traffic flows* Journal of Regional Science, 6: 5770.

[4] Stouffer, S.A., (1940) *Intervening Opportunities: A Theory Relating Mobility to Distance* American Sociological Review, Vol. 5, No. 6. 845--867

[5] Monge, M., (1781) *Mémoire sur la Théorie des Deblais et des Remblais* Mémoires de l'Académire Royale des Sciences 666--704

[6] Zipf, G. K., (1946) *The P1P2/D Hypothesis: On The Intercity Movement of Persons* American Sociological Review

[7] Austwick, M.Z., OBrien, O., Strano, E., Viana, M., (2013) *The Structure Of Spatial Networks and Communities in Bicycle Sharing Systems* PLoS ONE 8(9): e74685. doi: 10.1371/journal.pone.0074685

[8] Newman M.E.J., (2010) *Networks: An Introduction*

[9] Voorhees, A.M., (1955) *A General Theory of Traffic Movement* Proceedings, Institute of Traffic Engineers

[10] Plane, D.A., (1984) *Migration Space: Doubly Constrained Gravity Model Mapping of Relative Interstate Separation* Annals of the Association of American Geographers Vol. 74, No. 2. 244--256

[11] Wilson, A., (1967) *A statistical theory of spatial distribution models* Transportation Research , Vol. 1 No. 3. 253--269

[12] Schnieder. M., (1959) *Gravity Models and trip distribution theory* Papers in Regional Science, 5: 51-56.

[13] Ruiter, E.R., (1967) *Toward a better understanding of the intervening opportunities model* Transportation Research 1.1 47--56

[14] Eash, R., (1984) *Development of a Doubly Constrained Intervening Opportunities Model for Trip Distribution* Chicago Area Transportation Study (CATS)

[15] Cesario F.J., (1974) *The Interpretation and Calculation of Gravity Model Zone-to-Zone Adjustment Factors* En-

vironment and Planning, Vol 6 247 --257

[16] Pooler, J.A., (1992) *Spatial Uncertainty and Spatial Dominance in Interaction Modelling: A Theoretical Perspective on Spatial Competition* Environment and Planning Vol. 24 995--1008

[17] Pooler, J.A., Akwawu, S., (2001) *The development of an intervening opportunities model with spatial dominance effects* J Geograph Syst 3:69--86

[18] Harris, B., (1964) *A Note on the Probability of Interaction at a Distance* Journal of regional Science 2(2): 31--35

[19] Okabe, A., (1976) *Applied Models in Urban and Regional Analysis* Prentice-Hall Inc. Englewood Cliff NJ

[20] 2010 *Barclays' 25m sponsorship of London cycle hire scheme* BBC News, http://www.bbc.co.uk/news/10182833 (5 April 2014)

[21] 2010 *London saddles up for new bike hire scheme* BBC News, http://www.bbc.co.uk/news/uk-england-london-10810869 (5 April 2014)

[22] 2010 *Barclays Cycle Hire / Key Facts* BBC News, http://www.tfl.gov.uk/modes/cycling/barclays-cycle-hire/find-a-docking-station?intcmp=5386 (5 April 2014)

[23] *Curve Fitting Toolbox* MathWorks http://www.mathworks.co.uk/help/curvefit/product-description.html(5 April 2014)

[24] Barthelemy, M. (2010)*Spatial Networks* Physics Reports 499:1--101

[25] Simini, F., Gonzá M.C., Maritan, A., Barabási, A., (2012) *A Universal Model For Mobility and Migration Patterns* Nature Vol. 484 96--100