

## Using Digital Footprints for a City-scale Traffic Simulation

Gavin McArdle, National University of Ireland Maynooth  
Eoghan Furey, National University of Ireland Maynooth  
Aonghus Lawlor, National University of Ireland Maynooth  
Alexei Pozdnoukhov, National University of Ireland Maynooth

This article introduces a micro-simulation of urban traffic flows within a large scale scenario implemented for the Greater Dublin region in Ireland. Traditionally, the data available for traffic simulations come from a population census and dedicated road surveys which only partly cover shopping, leisure or recreational trips. To account for the latter, the presented traffic modelling framework exploits the digital footprints of city inhabitants on services such as Twitter and Foursquare. We enriched the model with findings from our previous studies on geographical layout of communities in a country-wide mobile phone network to account for socially related journeys. These data-sets were used to calibrate a variant of a radiation model of spatial choice, which we introduced in order to drive individuals' decisions on trip destinations within an assigned daily activity plan. We observed that given the distribution of population, the workplace locations, a comprehensive set of urban facilities and a list of typical activity sequences of city dwellers collected within a national travel survey, the developed micro-simulation reproduces not only the journey statistics such as peak travel periods but also the traffic volumes at main road segments with surprising accuracy.

Categories and Subject Descriptors: I.6.3 [Simulation and Modelling]: Applications

General Terms: Design, Algorithms

Additional Key Words and Phrases: Urban Analysis, Traffic Simulation, Social Networks

### 1. INTRODUCTION

New opportunities in the study of human mobility and urban computing arise from the availability of digital traces of movement such as the check-in data of location-based social network services or Call Detail Records (CDR) of cell phone usage. They permit urban mobility patterns to be uncovered. Such patterns have been previously unavailable from traditional travel surveys. Examples include the evidence of a long-tail in the daily trip distribution, indicating that some individuals cover distances orders of magnitudes larger in their typical trips than a majority of other people [Brockmann et al. 2006]. Check-in services such as Foursquare [Noulas et al. 2011], geo-referenced Twitter [Sadilek et al. 2012; Pozdnoukhov and Kaiser 2011] or points of interest (POI's) [Yuan et al. 2012] provide new insights for transportation modelling by providing quantifiable evidence about the purpose of individuals' travel whether for shopping, leisure or recreation, or meeting friends and visiting family. These activities generate a considerable amount of road traffic which needs to be accounted for in transportation models. However, they are not covered in detail by traditional travel surveys [Axhausen 2008] and rely on perceived rather than measured trip lengths and are likely to contain biases and inaccurate data.

---

The research presented in this article was funded by a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan. Authors' address: G. McArdle, E. Furey, A. Lawlor National Centre for Geocomputation, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland, A. Pozdnoukhov Civil and Environmental Engineering, University of California, Berkeley. email: {gavin.mcardle, eoghan.furey, aonghus.lawlor}@nuim.ie, alexeip@berkeley.edu

This is a preprint version of an article published in ACM Transactions on Intelligent Systems: Gavin McArdle, Eoghan Furey, Aonghus Lawlor, and Alexei Pozdnoukhov. 2014. Using Digital Footprints for a City-Scale Traffic Simulation. ACM Trans. Intell. Syst. Technol. 5, 3, Article 41 (July 2014), 16 pages. DOI=10.1145/2517028 <http://doi.acm.org/10.1145/2517028>

### 1.1. Contributions of this work

In this research we investigate the usefulness of digital footprints of individual movement for calibrating human mobility models within an urban traffic micro-simulation framework. We implemented a large scale realistic working day scenario for the Greater Dublin region in Ireland. Particularly, the presented approach includes the following novel contributions.

- We introduced a spatial choice model in the radiation framework for selecting destinations of individual trips (Section 2.1), with interpretable parameters and a simple calibration scheme (Section 2.3).
- The model is applied to determine facility choice from a data-set of points of interest using transition statistics gathered from geo-referenced Twitter messages and Four-square check-ins in Ireland (Section 2.2).
- Geographical layout of a social network observed in country-wide cell phone data is used as a proxy for modelling destination choice of the socially related trips such as visiting friends and family (Section 2.4).
- The developed methodology is applied for destination choice in shopping, leisure and socially related journeys which account for a major part of the traffic flows outside of regular commutes, but are not available from traditional surveys.
- These activities are integrated into a realistic traffic scenario calibrated on the daily plans generated in accordance with a census of population, workplace locations, daily activities and departure times (Section 3), and validated on the measured traffic volume counts at major roads in Greater Dublin region (Section 3.3).

This article is organised as follows. Section 2 gives an overview of spatial interaction approaches in urban mobility studies with particular focus on spatial choice modelling. We describe the developed adaptation of the radiation model in Section 2.1, which is then applied within a comprehensive framework of activity-based micro-simulation of traffic flows. This framework is built on the Multi-Agent Transport Simulation (MAT-Sim) platform [Balmer et al. 2008] and is described in Section 3. It uses a data-set of places of work locations to model commute flows. The necessary technical details on the data-sets used in model development are also given in Section 3. Our experimental results, presented in Section 3.3, show that the proposed spatial choice model produces accurate estimates of the daytime traffic volumes at major roads. We highlight and interpret these characteristic traffic volume patterns and compare this model to a baseline method in which destination choice is heavily influenced by available travel time. The possible origins of the surprisingly correct predictions are discussed in Section 4 which concludes the article and presents some extensions to the research.

## 2. URBAN MOBILITY

Traditional transportation planning and forecasting frameworks stem from travel surveys on origin-destination flows and apply gravity laws [Zipf 1946], intervening opportunities [Stouffer 1940], competing destinations [Fotheringham 1983] or an overarching constrained entropy maximisation framework [Wilson 1970] to investigate the trip distribution. Other promising techniques use a sample of finely grained GPS traces to estimate traffic flows under certain conditions [Zheng et al. 2011]. A more flexible approach using activity-based models, focuses on modelling travel demand based on the activities that people need to perform in the course of a day. This framework is usually implemented in a micro-simulation system where each agent is assigned an activity chain and performs destination choice in the context of this agenda, while also, to a certain extent, taking the actions of other agents into account. However, the mobility of individuals is still relatively unexplored within such frameworks. Empirical evidence

confirms that the regular commute to work is a dominating mobility pattern [Song et al. 2010], which also governs occasional fluctuations as people tend to arrange their travel plans by considering accessibility and convenience with regard to their primary locations such as home and work. However, the prevalence of trips not related to commuting is high. For example, in a recent Irish travel survey, it was seen that 20% of all trips are related to shopping. This indicates the importance of including such trips in traffic models. This rational paradigm and availability of digital footprints opens new ways to enrich activity-based models in transportation modelling and urban planning. Locations and facilities where secondary activities occur thus become a key element of the model's performance. Predicting the locations where people travel for work, recreation, shopping and to live is a significant challenge with a long research history. Both professionals and academics have carried out considerable work over the last century in dealing with this challenge and many positive findings have emerged. Factors that are taken into consideration regarding the choice of location include: travel distance and time, size of the store or facility, range of products or services and overall quality and price considerations. Social interactions and social influence are also emerging as important factors in location and facility choice [Arentze and Timmermans 2008], [Han et al. 2011], [Hackney et al. 2007]. Models for estimating residential location choice include various logit models [Abraham and Hunt 1997], [Lee and Waddell 2010], [Eluru et al. 2010]. In modelling the choice of leisure facilities some recent developments are based on hollow space time prisms which are derived from leisure trip length statistics [Horni et al. 2009a].

### 2.1. Radiation model of spatial choice

The radiation model [Simini et al. 2012] is inspired by the theory of intervening opportunities [Stouffer 1940] and applies emission-absorption ideas to compute probabilities of interactions for a set of origins and destinations of known capacities. It is a destination-constrained, parameter free model where distance decay is replaced with rank-based decay, similarly to intervening opportunities. We applied this idea at an individual level to derive a probability of choosing a particular facility from a set of facilities of the given type with known capacities. To derive the radiation probability in our model, we assign to every individual an emission threshold  $z^i$  which determines a minimum level above which a particular driving trip will become worthwhile. In Figure 1 we illustrate an example of a shopping destination choice process where an individual with a large threshold  $z$  who is planning a shopping trip would have high or perhaps very specific demands which would have to be overcome and so is less likely to visit, or be absorbed by a nearby facility or shop. We assume there is some preselected and unspecified distribution which describes this demand,  $p(z)$ . As there is generally no information on which kind of shopping trip an individual plans to undertake, we consider that a particular destination choice of an individual at location  $i$  is based on the probability  $P_{m_i}(z)$  that a maximum threshold drawn from  $p(z)$  after  $m_i$  repetitions is equal to  $z$ . Suppose that each possible destination facility at location  $j$  has a certain probability to satisfy that demand  $P_{n_j}(z)$ , which is given by a maximum threshold extracted from  $p(z)$  after  $n_j$  repetitions, where  $n_j$  is the capacity of a facility at  $j$ . We must also account for the probability that none of the intervening facilities could absorb the traveller  $P_{s_{ij}}(z)$  where  $s_{ij}$  is the sum of all facility capacities in a circle centred on  $i$  and with radius  $|r_i - r_j|$ . Then, the probability that a person at location  $i$  with a demand threshold  $m_i$  makes a trip to a facility at  $j$  with capacity  $n_j$  and no other closer facility, is found by integrating over all possible threshold values

$$P(1|m_i, n_j, s_{ij}) = \int_0^\infty dz P_{m_i}(z) P_{s_{ij}}(<z) P_{n_j}(z), \quad (1)$$

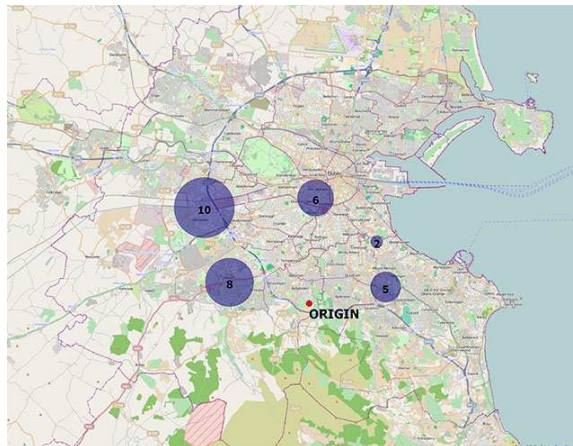


Fig. 1. The radiation model gives the probability that a given individual will travel from their origin to a given facility. The model takes into account the distances to each facility and their known capacities.

We perform the integral in a similar fashion to the radiation model [Simini et al. 2012] and find

$$P(1|m_i, n_j, s_{ij}) = \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}. \quad (2)$$

A useful feature of the radiation model is that the final probability does not depend on the unknown distribution  $p(z)$  of the demands of individuals, but simply on the facility choice parameter  $m_i$ . In the original formulation of the radiation model,  $m_i$  was determined by the population at location  $i$  which is easily determined. In our model we must determine this parameter from knowledge of facilities in a given region. A detailed description of how we do this is provided in Sec. 2.3.

## 2.2. Mobility data

The movement data-set we used to calibrate the model for shopping, leisure and recreation trips is a combination of geo-referenced Twitter data collected in [Pozdnoukhov and Kaiser 2011] and an Irish subset of the Foursquare data-set described in [Cheng et al. 2011]. It contains a total of 107218 check-in events posted by 5287 unique users. Characteristic trip lengths, contained in this data-set as compared to commute distances and call distances are presented in Figure 2(a). The graph highlights the difference between work trips (commuting), shopping and leisure trips (check-ins) and mixed trips (phone calls) thus reinforcing the need to treat shopping and leisure trips differently from other types of trips due to their unique distance distribution. Statistical information, extracted from these data, is used to assess the quality of the parameters for characteristic trip length for non-working activities used in the location choice model.

## 2.3. Facility choice and parameter fitting

In contrast to the original radiation model where the inputs are the known populations of the origin and destination, we have a quantity  $m_i$  (or  $n_j$ ) which relates to the choice of facilities in a region. Good quality public data-sets on facility capacities are not readily available. We can make some estimate of our parameter  $m_i$  for a given region from the user-supplied data in OpenStreetMap (OSM), but the overall coverage of this data-set is somewhat sparse. Instead, we have found that we can substitute the  $m_i$  for a given location with an average facility choice  $m_{opt}$  for the entire region.

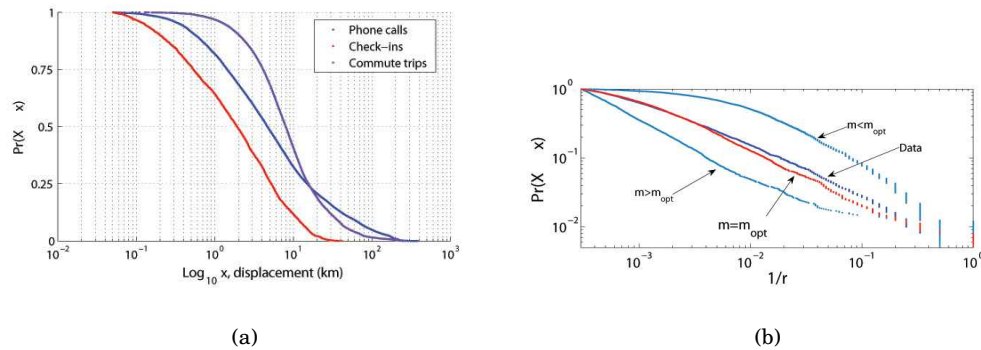


Fig. 2. In (a) we show the probability of observing a commute trip, a phone call or a check-in displacement longer than  $x$  kilometres. In (b) we fit the  $m$  parameter in the inverse rank cumulative probability plot on a  $\log$ - $\log$  scale.

We have devised a simple method to determine the optimum value for this parameter (Figure 2(b)). The facilities are ranked according to the distance to the trip origin and for each facility we use the radiation model (Eq. 2) to find the probability that a trip to the facility will be made. The data clearly show a long tail, confirming that longer trips to lower capacity facilities become increasingly unlikely. In trying to find a good value of  $m$  we see that if we set it too large,  $m > m_{opt}$ , this implies the high or specific demand which cannot be satisfied by nearby facilities and indeed it can be seen in Figure 2(b) the probability to undertake a trip to a distant facility is significantly increased. Conversely, if  $m$  is too small, an individual is much more likely to make a trip to a nearby facility. The optimal value of the parameter  $m_{opt}$  is found by minimising the difference between the rank cumulative probability as found by the radiation model and the observed data. The trip length distributions resulting from a facility choice dictated by a radiation model are shown in Figure 3. Again, it is seen that the  $m < m_{opt}$  results in a shorter average path length, and conversely for  $m > m_{opt}$ . The optimal  $m_{opt}$  reproduces the trip length distribution which we find from a database of check-ins. An example of the theoretical analysis of the trip length distribution under generic multiplicative spatial choice models can be found in [Veneziano and Gonzalez 2011].

#### 2.4. The geography of social networks

Empirical evidence [Axhausen 2008; Sadilek et al. 2012; Cho et al. 2011] suggests the importance of social influence on the formation of atypical patterns of mobility. People visit family members or friends, and join them in recreation, leisure, tourism or

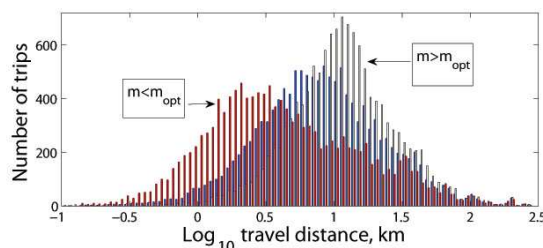


Fig. 3. Impact of  $m$  parameter on the trip length histogram. X-axis is in  $\log_{10}$  scale, km.

shopping trips. It was observed that a probability of befriending a person is inversely proportional to the number of closer people, i.e. a spatial rank of the person [Liben-Nowell et al. 2005]. Social networks also possess a distinct community structure which often show geographical patterns both at inter-city [Expert et al. 2011] as well as intra-city scales [Walsh and Pozdnoukhov 2011]. One can use the characteristic distances and geographical layout of these interactions as a proxy for socially related travel, in particular for journeys to visit friends and family.

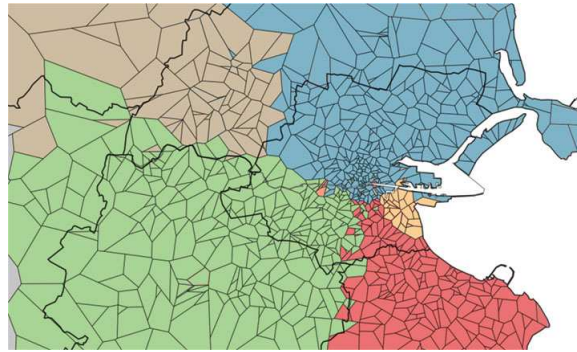


Fig. 4. Geographical layout of the major communities detected in a cell phone communication network in Greater Dublin region [Walsh and Pozdnoukhov 2011] (with permission of Fergal Walsh).

The geographical layout of the major communities detected in a cell phone communication network in the Greater Dublin region is presented in Figure 4 (taken from [Walsh and Pozdnoukhov 2011]). The community structure is clearly influenced by the underlying geography. Given that it is much more likely to observe social links between members of a community than across different ones, we have simulated a social network for the population of agents used in our simulation which reproduces the characteristic link length distribution, node degree and community structure which we find in the communication network. We will present further details on this social network generation algorithm elsewhere, and instead show the resulting statistics of the network in Figure 5.

### 3. TRAFFIC SIMULATION

An agent-based micro-simulation is an effective way to model and predict traffic. In this approach, each person in the population is represented by an agent which can

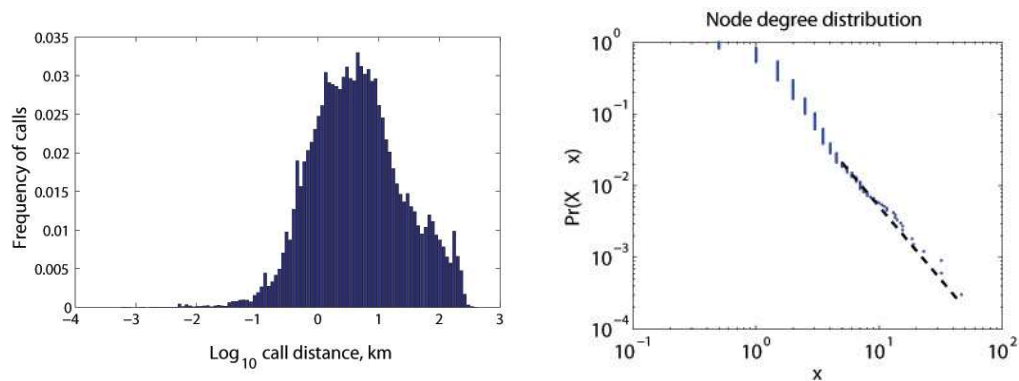


Fig. 5. A histogram of a social tie length (left) and a node degree (right) of the simulated social network.

make decisions and manage daily activities to obtain the greatest personal return. While there are many examples of agent-based micro-simulations [Balmer et al. 2008], [Al-Dmour 2011], [Khalesian et al. 2008], we opted to use MATSim [Horni et al. 2009a] due to the flexibility it offered for destination choice modelling.

In MATSim each agent is assigned a plan representing the spatial and temporal requirements for a particular day. The plan is altered through iterations of the simulation in order to maximise an individual agent's personal utility score. Travelling is seen as having a negative or neutral score while spending time at home or work has a positive score. Each iteration tries to minimise travel time to increase the overall utility score. The iterations should continue until the system has reached a relaxed state, known as a Nash Equilibrium where future iterations will produce little improvement in the utility scores of agents. The variables which MATSim can alter during iterations are the departure time, the route and destination choice for discretionary activities, such as leisure and shopping [Horni et al. 2009a]. Route choice is determined using the A-Star algorithm [Lefebvre and Balmer 2007], while temporal decisions are achieved using local random mutation [Balmer et al. 2005]. Location choice is achieved using discrete choice theory. By altering agents' plans the model constantly tries to decrease travel time by avoiding congested routes (by re-routing and re-planning of destination choice) or busy periods of the day (by rescheduling) on each iteration.

Activity chains, which represent the sequence of activities undertaken by individuals, are a key input for micro-simulations [Axhausen and Herz 1989]. Generally they are derived from data recorded in travel surveys and reveal patterns of activities that people carry out on a normal day. Originally, the optional location choice module in MATSim employed an entirely time based utility function to calculate where individuals could travel within the time allocated for shopping trips, however this was discovered to be insufficient. Therefore, the model was extended to consider further variables such as shop size or the density of shops in a given area [Horni et al. 2009a] and random error terms [Horni et al. 2011].

### 3.1. Implementation

We use MATSim to perform a simulation of traffic flows within the Greater Dublin region. MATSim has specific data requirements, including the road network and agent plans for the study area. Below, the details of how this data was prepared are provided.

*3.1.1. Network Preparation.* MATSim requires a road network consisting of nodes and links. The nodes represent road intersections while the links are the road segments joining these intersections. Using tools provided by MATSim, the OSM road network for Dublin was extracted and transformed into the appropriate structure. All roads within a radius of approximately 30Km around Dublin City were extracted from OSM (Figure 10). Additionally, all major roads (national routes and motorways) in Ireland were obtained. Additional information provided by OSM including the speed limits, class of road and type of road were also extracted. This additional data permits the simulation to determine the flow capacity of a road segment which is used in route selection.

*3.1.2. Population and Demand Preparation.* MATSim also requires a population, which it will model. The population is generally characterised by the home and work locations of individuals within the study area. The locations are shown in Figure 6, while Figure 7 highlights the contrast between a residential and industrial area. Given the importance of non-work related trips on the road network, the population contains individual workers and non-workers. Each individual requires a plan, which consists of activities they will perform during a typical day. The plans or activity chains which include the sequence and duration of activities represent the demand on the network.

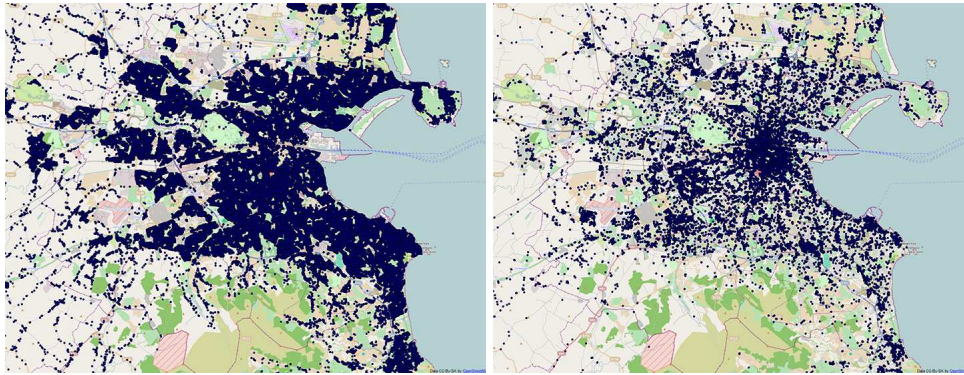


Fig. 6. Home (left) and work (right) locations of the population.

The population data were obtained from the Irish National Census which is conducted every five years. The most recent census was conducted in 2011. POWSCAR (Place of Work and School - Census of Anonymised Records) is a subset of the full census which provides the home, work, school and college location of individuals, the mode of transport used to commute and the time at which individuals leave their residence in the morning. The home location is anonymised by giving it at a small area level. A small area is a geographic area consisting of 80 to 100 households. The work location is presented at a 250 metre grid level. The time of departure is represented by discrete 30 minute intervals for the morning period and several transit modes are encoded in the means of transport.

To simulate traffic movement of workers within the Greater Dublin region, individuals who live in Dublin or whose place of work is within Dublin were extracted from the POWSCAR. Furthermore, only those who use a private car or van to get to work were considered. Each individual in the data-set is assigned to specific addresses which represent their home, work and school locations. This is achieved using GeoDirectory (<http://www.geodirectory.ie>), a database of Irish addresses. The database contains the coordinates of each address, its small area code and the class of building (commercial, residential or both). Using this database, each individual is assigned a random residential address in their small area. Work addresses were assigned based on the 250 metre buffer from the declared work location in the POWSCAR data-set. Additional information in GeoDirectory specifies the use of a particular address. Primary, secondary and third level education were used to assign addresses to students in the population. Departure times for work, school and college activities were then assigned using Gaussian curves centred at the declared 30 minute departure times in the POWSCAR data-set.

While workers make up the majority of the demand on the road network in the morning and evening rush-hour, during the off-peak hours non-workers account for a large proportion of road traffic. A population of non-workers who own cars was built using a similar technique described above. Statistics regarding the number of retired, sick and unemployed persons in a small area were used to assign a home addresses for non-workers in the Dublin region. Statistics regarding car ownership were then used to distribute cars to the non-worker population.

The demand on the network is represented by activity chains [Axhausen and Herz 1989] harvested from The Irish National Travel Survey (INTS) [INTS 2009]. This Survey was carried out in 2009 as part of a Quarterly National Household Survey. Approximately 7000 participants were randomly selected and issued with a travel diary to record all journeys for a period of 24 hours on a specific day. The information gath-





Fig. 7. Each marker on the map represents a home location (left) and work location (right) and highlight a typical segregation of residential areas and industrial zones.

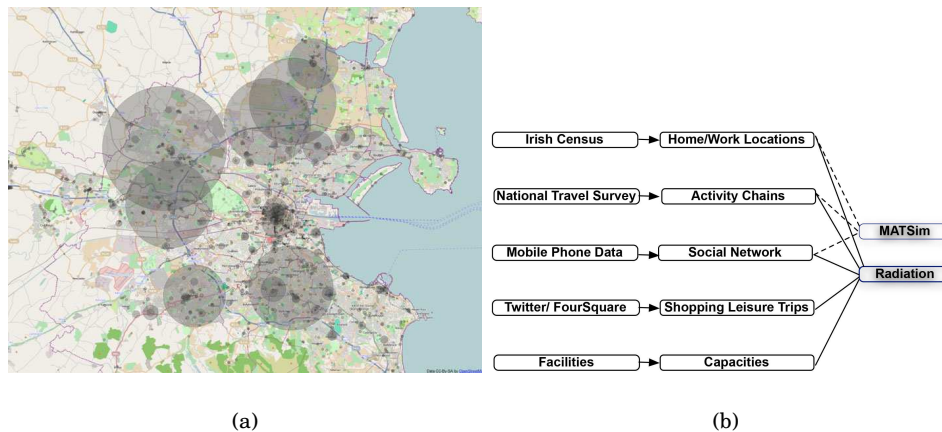


Fig. 8. On the left we show the locations of facilities in the greater Dublin area. The estimated capacities are denoted by circle size (largest correspond to major shopping malls). On the right is a summary of the input data and how it is used within the simulation.

ered included journey origin and destination type (home, work, school, etc.), time of departure and arrival, mode of transport, purpose of trip, distance travelled and the time of each journey. As we are interested in vehicle traffic, only those trips which involved a car or van were considered.

Activity chains containing work activities were identified and assigned to workers while chains without work related activities were assigned to non-workers. In all cases, the frequency of each activity chain was calculated and used to produce a probability score for an individual to carry out a specific daily activity chain and assigned to the population accordingly. While travel to work was the most frequent activity, other activities such as shopping accounted for 20% of trips with leisure and family related activities accounting for 6% and 10% of trips respectively. While the activity chains provide the sequential and temporal data related to activities, there is no indication as to where each activity occurs (Destination Choice). The absence of such data creates the destination choice challenge for deriving the initial demand for the simulation.

### 3.2. Facility choice implementation

The locations of various activity types (schools, gyms, pubs, restaurants, shops, etc.) were extracted from public data-sets including OSM and a points of interest database of an in-car GPS navigator. Additionally, the capacity of each of these facilities was estimated using a combination of facility size and domain knowledge. The resulting facilities are shown in Figure 8(a) where the circle size represents capacities. These data formed the input for determining the location choice during the generation of the day plan for each agent. The plans were assigned according to the survey described above and in addition to work and home activities, contained the following activity types: school/education, shopping, personal business, visiting family/friends, social/entertainment, sport/leisure, and doctor/medical facilities.

Two techniques for assigning the initial location of secondary activities were utilised. In the first scenario, a random facility within a radius of 8 kilometres from the users home or work location is selected. In scenario 2, the radiation model as described in Section 2.1 and Section 2.3 was applied for each individual choice over all alternatives amongst the facilities of a given type. In both cases, for social visits, a possible destination to visit was assigned by sampling the home or work locations of a friend from the simulated social network (Section 2.4). A summary of all data sources and how they are used is provided in Figure 8(b).

Day chains for 600,000 agents, representing all drivers in the Greater Dublin region, were produced. For efficiency reasons, a 25% sample of the agents were used in the simulation. Three hundred iterations produced a steady Nash equilibrium. In the case of the simulation in which the initial demand for secondary activity locations was selected based on a radius from home and work locations, replanning of destination choice, re-routing and rescheduling were permitted on each iteration. In this case, destination replanning is based on the work of [Horni et al. 2011] and involves selecting locations based on the time available to reach them and the perceived busyness of locations. The ability for agents to re-plan destination choice should overcome the weakness of choosing only close-by locations as the initial demand on the network.

Where the radiation model was used to select the initial location of secondary activities, only re-routing and rescheduling were permitted. In order to test the effectiveness of the radiation model, the locations of secondary activities remained constant for each iteration. The results from both approaches are compared to ground truth data using the techniques described below.

MATSim produces several pieces of data to assess the effectiveness of the simulation, the input parameters and the destination choice model. Firstly an animation which shows the time-stamped movement of the agents over the road network can be



Fig. 9. Snapshots of the simulation output showing the traffic situation at 06:00 (left), 08:30 (centre) and 20:00 (right)

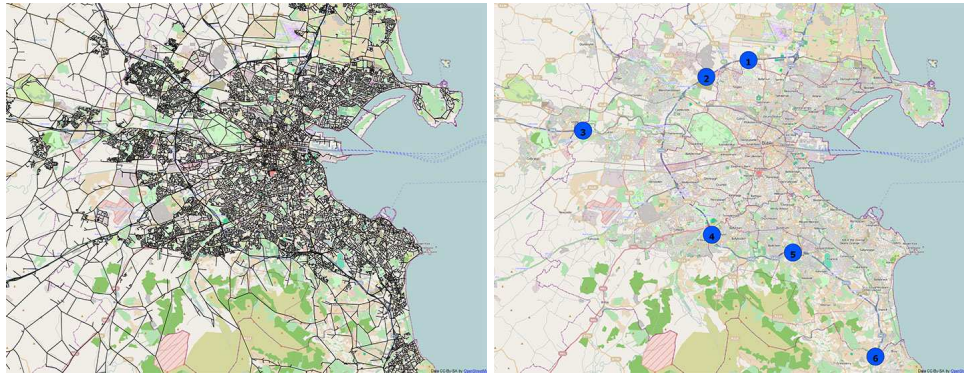


Fig. 10. The road network used in the simulation (left) and the location of the count stations used in for validating traffic volumes produced by the simulation (right).

rendered to assess traffic volumes during different periods of the day (see Figure 9). Volumetric data, in the form of a count, shows the number of cars passing each road segment (link on the network) during each hour of the day. These served as key validation metrics for assessing the performance of the model and are discussed in detail in Section 3.3.

### 3.3. Evaluating Traffic Counts

Figure 10 shows the physical location of six traffic count stations which count the number of vehicles passing in each direction using hardware embedded in the road. The data produced by these count stations were aggregated to hourly counts and used to validate the data produced by the simulation for the same road segments. Figure 11 shows a scatter plot outlining the correlation between the observed traffic volume counts and those produced by the simulation, using the two scenarios for destination choice discussed in Section 3. The counts produced by the radiation model for location choice have a stronger correlation to the observed traffic volume than the MATSim Best-Response location choice model. This is particularly evident when count volumes are high. The results for three count stations are shown in Figures 12(a) to 12(c). Each

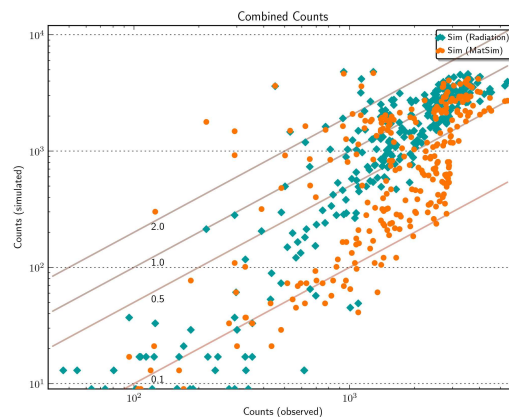


Fig. 11. Correlation between Observed and Simulated Counts for the 6 Count Stations shown in Fig 10. When the counts are low neither model does a very good job at matching the observed counts, but the performance is better during the day when the counts are larger. It can be clearly seen that the radiation model outperforms the MATSim destination choice model, which consistently underestimates the counts.

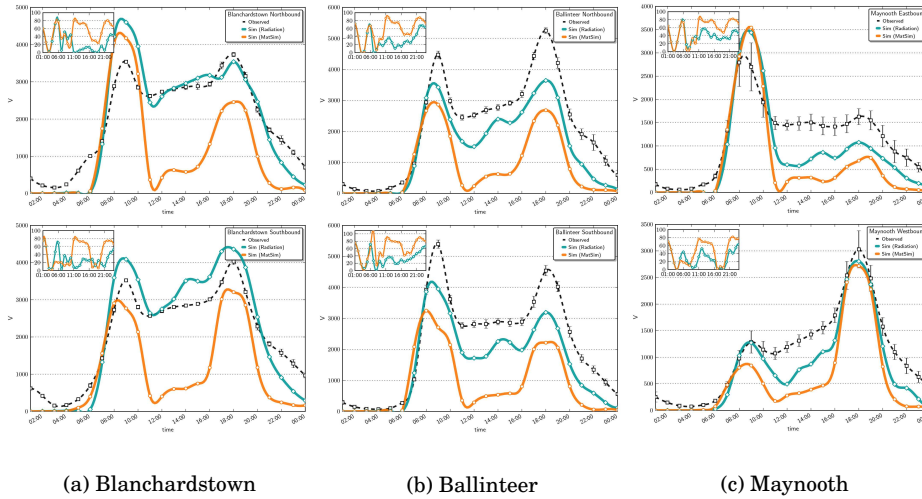


Fig. 12. Observed and simulated counts for various points around Dublin. The observed counts (dotted line) are shown with errors bars for a 24hr period. The inset shows the relative percentage error of the MATSim counts with each destination choice model.

panel shows a graph depicting the observed and simulated traffic counts for each hour of the day. Inset in each graph is a second graph showing the hourly % mean error for the simulation using both location choice scenarios.

While both approaches effectively predict the morning and evening peak commuting flows, the volumes produced by the model relying on MATSim to select facility location for secondary activities generally underestimates traffic volumes. This finding corresponds to previous studies in which it was noted that leisure trip distances decrease over each iteration as too much emphasis is placed on decreasing travel distance and travel time which prevents agents from travelling further for better opportunities [Horni et al. 2009b]. Below we examine the results in more detail and provide possible explanations for the discrepancies between the simulated and observed results.

From 22:00 to 06:00, both scenarios produce very low volumes of traffic across the entire road network. This creates a high relative error for this period of day. The error is particularly evident when the model produces no traffic flow for a given hour of the day and errors in excess of 80% are common

This discrepancy in traffic volume is caused by sparse input data for this period of the day as both the NTS and POWSCAR do not cover this time. Generally, the daily activity plan for an agent commences when they leave their residence for the first time and there are no data for agents who are already in transit during the early hours of the morning. Both models accurately predict the peak in traffic flow from approximately 07:00 to 10:00. The majority of this traffic is commuter traffic and the origin and destination (home and work) are provided by POWSCAR. Nevertheless, a proportion of the traffic on the network at this time is connected to secondary activities such as shopping and leisure. For such trips destination choice is used and so there is a divergence in the counts produced by both scenarios, with the MATSim destination choice model generally underestimating volumes while the radiation model overestimates volume.

During this busy period of the day, there is likely a trade-off between travel time and travelling for better opportunities which is not accounted for in either model. Due to the set-up of the experiment, the volumes produced in both scenarios at count stations on the periphery of the city produce almost identical traffic volumes in the morning (e.g. Maynooth Eastbound, Figure 12(c)). The experiment does not capture people who are making discretionary trips into the city from the surrounding counties as only those who live or work in Dublin are selected for the simulated population and so the destination choice is not a factor on such routes at this time.

From 10:00 to 16:00 there is a significant variation in the two scenarios. The MATSim destination choice model consistently underestimates the traffic volumes [Horni et al. 2009b]. The majority of trips occurring at this time are secondary activities such as shopping and leisure. Although the MATSim destination choice model can alter the locations to visit on each re-planning iteration, the model focuses on the travel distance and time constraints and so restricts agents to visiting nearby facilities thereby reducing travel volume on the network in general.

This is in contrast to the radiation model for location choice which is not as constrained by time and allows agents to travel further for better opportunities. As a result, while both location choice models fluctuate during this period of the day, the relative error for the MATSim location choice model is often around 80%, whereas the errors produced by radiation model are half this amount. In general the simulated counts are lower than the observed counts for both approaches. This indicates that there are gaps in the data and in our knowledge of trips occurring during this period of the day. For example, trips which involve collecting children from school are currently absent from the input data but will be included in future studies.

For count stations on the periphery of Dublin (E.g. Maynooth, Figure 12(c)), both models produce lower than observed counts for traffic travelling towards Dublin (Eastbound) during the day and in the evening. This again is due to the set up of the experiment in which only individuals who live or work in Dublin are included in the simulated population. Individuals who live outside Dublin and travel into the city in the evening for leisure or shopping are absent from the simulation and so the count volumes on these routes are reduced. As the simulation is extended beyond the Greater Dublin region, such anomalies will have less impact. Despite this lack of data, the locations chosen by the radiation model yield better results than the locations chosen by MATSim for this count-station.

The evening peak (16:00 to 20:00) is detected in all situations. Unlike the morning commute, we do not have specific data regarding the time individuals leave their place of work and so the model relies on typical working hours provided by the activities chains generated from the National Travel Survey (NTS). Like the daytime hours, a higher proportion of shopping and leisure trips occur during the evening, and this accounts for the better performance of the radiation destination choice model compared with MATSim's destination choice model.

Overall, the scenario which uses the radiation model for destination and location choice performs better than the destination choice model which is integrated with MATSim. The relative error is consistently lower, especially during the day and evening hours when the prevalence of shopping and leisure activities is high. In situations where such activities are not occurring and destination choice is not required, the models produce similar results (e.g. Maynooth Eastbound during the inbound commute). There are still gaps in the input data which when rectified will improve the results further. Additionally, the inclusion of time constraints in the radiation model should see a marked improvement in the correlation between observed and simulated counts.

#### 4. CONCLUSIONS

This article has described a new approach for destination choice within a traffic simulation for the Greater Dublin region. The results indicate the technique, which uses destination choice based on a radiation model, is effective at predicting traffic flows at various hours of the day. A major impact on the quality of the results is due to the amount of high detail and spatial resolution of the home to work data compiled for the Dublin region. Nevertheless, the newly introduced universal radiation spatial choice model was shown to perform superior when compared to the model used within MATSim based on spatio-temporal constraints for determining destinations.

The radiation model is calibrated using statistics from digital footprints within the city. Statistical information regarding the distance travelled for shopping and leisure trips are extracted from check-in data in Four Square and Twitter. Information regarding such trips is typically missing from traffic simulations which focus on commuter traffic. While it is well understood that the users of such social media services are not representative of the general population, the results indicate that the trips they make are a good proxy for shopping and leisure trips of the population in general. As more detailed data of this type becomes available the calibration of the radiation model for location choice will be improved. The radiation model relies on the known capacities and attractiveness scores of shopping and leisure facilities. These scores were estimated using domain knowledge. Refinement of how the scores are calculated such as using inputs from social media will improve the model further and help to assess the sensitivity of such parameters.

Approximately 10% of all trips relate to a social visit or social interaction with family and friends. It is therefore important to include such trips in any traffic model. Despite this, the social network of the population being modelled is often unknown. For the greater Dublin region, the exact geography of social links of the population which was derived from mobile phone call data also makes an essential contribution to the overall performance of the simulation. For social interaction, the simulated population visits the work or home location of a member of their social network. Techniques to permit agents to coordinate their plans in a spatial and temporal context will be investigated and incorporated into the simulation. Additionally the influence which social networks and individual habits play in personal location choice will be analysed. Borrowing collaborative filtering techniques from the e-commerce domain, we will build a location and facility recommender system to assess the role played by social connections in individual location choice for secondary activities. We are particularly interested in the impact of stratification effects that emerge in the coupled consideration of mobility and social influence on facility choice [Lawlor et al. 2012]. These important topics will be analysed in a future study.

#### REFERENCES

- JE Abraham and J.D. Hunt. 1997. Specification and estimation of nested logit model of home, workplaces, and commuter mode choices by multiple-worker households. *Transportation Research Record: Journal of the Transportation Research Board* 1606, -1 (1997), 17–24.
- N.A.A.H. Al-Dmour. 2011. TarffSim: Multiagent Traffic Simulation. *European Journal of Scientific Research* 53, 4 (2011), 570–575.
- T. Arentze and H. Timmermans. 2008. Social networks, social interactions, and activity-travel behavior: a framework for microsimulation. *Environment and planning. B, Planning & design* 35, 6 (2008), 1012.
- K.W. Axhausen and R. Herz. 1989. Simulating activity chains: German approach. *Journal of Transportation Engineering* 115, 3 (1989), 316–325.
- K W Axhausen. 2008. Social networks, mobility biographies, and travel: survey challenges. *Environment and Planning B: Planning and Design* 35 (2008), 981–996. Issue 6.
- M. Balmer, K. Meister, M. Rieser, K. Nagel, K.W. Axhausen, K.W. Axhausen, and K.W. Axhausen. 2008. *Agent-based simulation of travel demand: Structure and computational performance of MATSim-T.*

- ETH, Eidgenössische Technische Hochschule Zürich, IVT Institut für Verkehrsplanung und Transportsysteme.
- M. Balmer, B. Raney, and K. Nagel. 2005. Adjustment of activity timing and duration in an agent-based traffic flow simulation. *Progress in activity-based analysis* (2005), 91–114.
- D. Brockmann, L. Hufnagel, and T. Geisel. 2006. The scaling laws of human travel. *Nature* 439, 7075 (2006), 462–465.
- Zhiyuan Cheng, James Caverlee, Kyumin Lee, and Daniel Z. Sui. 2011. Exploring Millions of Footprints in Location Sharing Services. In *ICWSM*.
- Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '11)*. ACM, New York, NY, USA, 1082–1090. DOI: <http://dx.doi.org/10.1145/2020408.2020579>
- N. Eluru, C.R. Bhat, R.M. Pendyala, and K.C. Konduri. 2010. A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. *Transportation* 37, 4 (2010), 603–626.
- Paul Expert, Tim S. Evans, Vincent D. Blondel, and Renaud Lambiotte. 2011. Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences* 108, 19 (10 May 2011), 7663–7668. DOI: <http://dx.doi.org/10.1073/pnas.1018962108>
- A. S. Fotheringham. 1983. A new set of spatial-interaction models: the theory of competing destinations. 1, 15 (1983), 15 – 36.
- J.K. Hackney, F. Marchal, J.K. Hackney, and J.K. Hackney. 2007. *Model for coupling multi-agent social interactions and traffic simulation*. ETH, Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transport Systeme.
- Q. Han, T. Arentze, H. Timmermans, D. Janssens, and G. Wets. 2011. The effects of social networks on choice set dynamics: Results of numerical simulations using an agent-based approach. *Transportation Research Part A: Policy and Practice* 45, 4 (2011), 310–322.
- Andreas Horni, Kai Nagel, Kay W Axhausen, Kai Nagel, Kai Nagel, Kay W Axhausen, and Kay W Axhausen. 2011. *High-Resolution Destination Choice in Agent-Based Demand Models*. Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme.
- A. Horni, D.M. Scott, M. Balmer, and K.W. Axhausen. 2009a. Location Choice Modeling for Shopping and Leisure Activities with MATSim: Combining micro-simulation and time geography. *Transportation Research Record: Journal of the Transportation Research Board* 2135, -1 (2009), 87–95.
- A Horni, M Scott, D, M Balmer, K.W. Axhausen, and Kay W Axhausen. 2009b. *Location choice modeling for shopping and leisure activities with MATSim: utility function extension and validation results*. ETH, Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme.
- M. Khalesian, P. Pahlavani, and MR Delavar. 2008. GIS-Based Multi-agent traffic micro simulation for modelling the local air pollution. In *International Society of Photogrammetry and Remote Sensing Congress*.
- A. Lawlor, C. Coffey, R. McGrath, and A. Pozdnoukhov. 2012. Stratification structure of urban habitats. In *Pervasive Urban Applications Workshop (PURBA '12) at Pervasive 2012*.
- B.H.Y. Lee and P. Waddell. 2010. Residential mobility and location choice: a nested logit model with sampling of alternatives. *Transportation* 37, 4 (2010), 587–601.
- N. Lefebvre and M. Balmer. 2007. *Fast shortest path computation in time-dependent traffic networks*. ETH, Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transport Systeme.
- David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. 2005. Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the United States of America* 102, 33 (2005), 11623–11628. DOI: <http://dx.doi.org/10.1073/pnas.0503018102>
- Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. 2011. A tale of many cities: universal patterns in human urban mobility. *arXiv:1108.5355v4 [physics.soc-ph]* (2011).
- NTS. 2009. National Travel Survey Report 2009 Central Statistics Office, Government of Ireland. (2009). <http://www.cso.ie/en/media/csoie/releasespublications/documents/transport/2009/nattravel09.pdf>
- Alexei Pozdnoukhov and Christian Kaiser. 2011. Space-time dynamics of topics in streaming text. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN '11)*. ACM, New York, NY, USA, Article 8, 8 pages. DOI: <http://dx.doi.org/10.1145/2063212.2063223>

- Adam Sadilek, Henry Kautz, and Jeffrey P. Bigham. 2012. Finding your friends and following them to where you are. In *Proceedings of the fifth ACM international conference on Web search and data mining (WSDM '12)*. ACM, New York, NY, USA, 723–732. DOI: <http://dx.doi.org/10.1145/2124295.2124380>
- Filippo Simini, Marta C. Gonzalez, Amos Maritan, and Albert-Laszlo Barabasi. 2012. A universal model for mobility and migration patterns. *Nature advance online publication* (2012). DOI: <http://dx.doi.org/doi:10.1038/nature10856>
- Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. Limits of Predictability in Human Mobility. *Science* 327, 5968 (19 Feb. 2010), 1018–1021. DOI: <http://dx.doi.org/10.1126/science.1177170>
- Samuel A. Stouffer. 1940. Intervening Opportunities: A Theory Relating Mobility and Distance. *American Sociological Review* 5, 6 (1940), 845–867. <http://www.jstor.org/stable/2084520>
- Daniele Veneziano and Marta C. Gonzalez. 2011. Trip Length Distribution Under Multiplicative Spatial Models of Supply and Demand: Theory and Sensitivity Analysis. *CoRR* abs/1101.3719 (2011).
- Fergal Walsh and Alexei Pozdnoukhov. 2011. Spatial structure and dynamics of urban communities. The First Workshop on Pervasive Urban Applications (PURBA).
- Alan G. Wilson. 1970. *Entropy in Urban and Regional Modelling*. Pion, London, United Kingdom. <http://www.worldcat.org/isbn/0850860229>
- Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 186–194.
- Yu Zheng, Yanchi Liu, Jing Yuan, and Xing Xie. 2011. Urban computing with taxicabs. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 89–98.
- George K. Zipf. 1946. The P1 P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review* 11, 6 (1946). <http://www.jstor.org/stable/2087063>