Modeling location choice of secondary activities with a social network of cooperative agents

F. Marchal
fabrice.marchal@ieee.org
Swiss Federal Institute of Technology
IVT ETH Hönggerberg
CH-8093 Zurich
Phone: + 41 1 633 31 05
Fax: + 41 1 633 10 57

K. Nagel
nagel@vsp.tu-berlin.de
Inst. for Land and Sea Transport Systems
Technical University Berlin Sek SG 12
Salzufer 17-19
D-10587 Berlin
Germany

1st April 2005

Paper length: 4600 words, 5 figures

Abstract

Activity-based models in Transportation Science focus on the description of human trips and activities. We address the modeling of the spatial decision for so-called secondary activities: given both home and work locations, where do individuals perform activities such as shopping and leisure? The simulation of these decisions using random utility models requires a full enumeration of the possible outcomes. For large data sets, it becomes computationally unfeasible because of the combinatorial complexity. To overcome this limitation, we propose a model where agents have limited, accurate information about a small subset of the overall spatial environment. Agents are inter-connected by a social network through which they can exchange information. This approach has several advantages compared to the explicit simulation of a standard random utilit model: a) it computes plausible choice sets in reasonable computing times b) it can be easily extended to integrate further empirical evidence about travel behavior and c) it provides a useful framework to study the propagation of any newly available information. The paper emphasizes the computational efficiency of the approach for real-world examples.

1   General context

Activity-based models in Transportation Science focus on the description of the organization of human activities in time and space. This organization determines the demand for travel, that is the amount of users that the various transportation systems need to accommodate. It is assumed that the demand for travel is derived from the demand for performing activities at specific locations. Obviously, individuals constantly perform some trade-off between enjoying activities that have a high reward value (for instance working at a company in the downtown area) and the time and budget it takes to get to the specific location of these activities. Various operational models such as URBANSIM [10] are available to describe this trade-off for the choice of the home and work locations. It is essentially assumed that users perform a trade-off between rents, travel costs and wages. However, empirical evidence [2] have shown that a significant amount of traffic is generated for other purposes than commuting, often referred to as *secondary* activities: shopping, leisure, going to social events, etc.

2   Problem statement

Our work intends to model the specific process of the location choice of secondary activities in the case of high resolution data sets. The methodological constraints are that the modeling should be behaviorally sound, compatible with micro-economics foundations and computationally feasible. The temporal dimension (i.e. the scheduling of the activities) is ignored for the time being. We assume that the order of the activities called a *plan* is given (i.e. getting out from home, going to work, working for eight hours, going for shopping at lunch time, etc.). The physical environment is described by two large data sets that typically originate from Geographical Information Systems (GIS): a) the land-use data and b) the transportation system data. The land usage is a raster-type description that includes the information about the nature of each parcel of the studied area (e.g. housing density, number of shops, type of area: rural, commercial, industrial). The transportation system is a vector-type description of the various transportation modes available (e.g. car, rail, bus) as a network with nodes and links. Nowadays, these data are available at a very high resolution: typical land-use cells are 100 meter square and road networks are described down to 10 meter road sections. The long-term goal of our research is to model entire metropolitan areas microscopically by simulating the individual decision of millions of citizens. Therefore, the problem at hand can be stated as follows: how to simulate the selection of the activity locations of $A = 10^6$ **citizens** in a **grid** that has $C = 10^5$ **cells**. Note that the travel times from cell to cell have to be given by external traffic models. Multi-agents traffic assignment models such as those developed by the authors [8, 5] are now able to predict travel time patterns for large-scale data sets. Therefore, this issue will not be considered here. The generation of travel demand for these models is addressed here.

3   Micro-Economics foundation

The standard practice in Transportation Science to approach such problems is to use random utility models (RUM) borrowed from the discrete choice theory of Micro-Economics [3, 6]. These models assume that individuals are maximizing their own utility. For instance, the utility to go shopping at a mall located in cell i for a simple plan (i.e. home - shopping - work) is given by

$$U_i = R_i - C_{hi} - C_{iw} + \mu\epsilon_i = V_i + \mu\epsilon_i,$$

where R_i is the reward associated to shopping at that particular facility that depends on the availability of goods, their prices, etc.; C_{ih} is the travel cost to travel from home to cell i; C_{iw} is the travel cost to travel from cell i to $work$; μ is a scale factor and ϵ_i is a random variable that is specific to the individual. The latter random utility part captures all the hidden preferences of a specific user for location i that are not accessible to the modeler. By contrast, V_i is called the deterministic part of the utility. Under the assumption that ϵ_i are i.i.d. extreme value distribution of type I, it can be shown that the probability to choose to go shopping at cell k is given by:

$$P(k) = P(U_i < U_k \forall i \neq k) = \frac{\exp(-V_k/\mu)}{\sum_{i=1}^{C} \exp(-V_i/\mu)}$$

Since a probability greater than zero is assigned to each potential intermediary stop on a cell, this formulation requires a full enumeration of the possibilities on the spatial grid. For trips that count S intermediary stops, the complexity is $O(AC^S)$ which is not feasible in realistic cases. Initially, RUMs were intended for the description of choices between a small set of alternatives distinguishable by humans (e.g. car brands). Their application to a discretized continuum (i.e. urban space) remains behaviorally questionable. Still, we believe it is fundamental to keep some compatibility with RUMs because of the literature that has been devoted to developing empirical techniques (e.g. surveys) to calibrate the parameters of those models.

Another drawback of RUMs is that they provide only a static representation that does not take into account the temporal dimension of the decision process. RUMs do not model explicitly the choice process but only its outcome. Our goal is to come up with a model that reflects the underlying learning process and that will eventually include the dynamics of the building of the choice set. Capturing this aspect is potentially important to study, for instance, the evolution of travel demand given some modification in the land-use patterns. Moreover, the environment itself should be time-dependent. Travel times and travel impedances to move in the network are subject to within-day and daily variability. The feedback of the travel conditions on travel choices and conversely is also often missing from static RUM-based analysis. To overcome these limitations, we propose a four phase multi-agent simulation model that includes a dynamic learning process and that can, in principle, be coupled with a dynamic mobility simulator. Agents travel, explore, learn and socialize.

4   Multi-agent based approach

We propose a multi-agent based simulation where **each agent (i.e. each simulated citizen) has only limited, accurate information about** N **cells** ($N \ll C$) called the "memory" of the agent. The intuition is that real humans have limited cognitive abilities and can only consider a small amount of options at the same time. The organization of these options in the mind and the human representation of space is probably far different from "pixels". Nevertheless, we keep that trivial representation for now, as it can be replaced later by a more sophisticated one, for instance that of [1].

We assume that agents are inter-connected by a *social network* through which they can exchange information about their respective subsets. Each agent is socially connected to K **acquaintances or "kins"**. The intuitive benefit of the social network is that agents are very heterogenous yet facing similar choices. Therefore, a decision which is optimal for an agent might be close to optimal for one of his acquaintances. We show below that the diffusion of the knowledge of optimal strategies through the social network can, indeed, exploit that hidden redundancy.

The simulation is iterative and each round has four stages: evaluation, socialization, exchange and exploration.

Evaluation each agent performs the location choice of the intermediary stops based on his own private information. The choice can be deterministic (the best cells are selected from the memory of the agent) or probabilistic (a RUM is applied to the small set corresponding to the memory of the agent). The computing load for building and storing the travel plans is $O(AN^S)$. At this point, the plans are fed in a dynamic traffic model (such as MATSIM [8] or METROPOLIS [5]). The traffic simulator computes the delays incurred due to traffic congestion, which are then used in the next round of evaluation.

Socialization social connections are created and deleted dynamically. The deletion mechanism is a simple exponential decay. The creation mechanism is a spatial reinforcement reminiscent of pheromones in ant colonies optimization [4].

Exchange for each of its social connections, an agent has the opportunity to exchange a piece of information. A cell is picked randomly from the agent's memory and the other agent is informed about it. The exchange is bi-directional and the outcome of the exchange is described by the learning mechanism below. More sophisticated exchange strategies could be taken into account.

Exploration agents have the possibility to explore cells in the neighborhood of those that they visit. This stage is mainly intended to recover potential information loss in the other stages, thus relieving the implementation from checking that any cell was lost from the global knowledge of all the agents. Obviously $O(A)$ operations are required.

5 Learning mechanism

The memory of an agent is represented on Figure 4: the two first buffers contain the information about locations that are either close to home or close to work (e.g. a small circular area). The third buffer called "elite" buffer (of size E) corresponds to locations that have high score values and the last buffer contains "vague" information about cells that have poor score values. The scores corresponding to the elite buffer are also kept in the agent's memory. When an agent informs another agent during the exchange stage, a cell is picked randomly from the total memory of the informer. At that point, the informed agent evaluates how this new cell information can potentially improve his/her plan score. This is done by comparing the poorest score of the elite buffer to the scores of all the potential plans with at least one stop at the new cell. This revision implies to replace each intermediary stop of the potential plans by the new cell: $O(SN^{(S-1)})$ operations are needed. If the score is better than the worst solution of the elite buffer, the new cell is promoted to that buffer and it is sorted: $O(\ln E)$ operations are needed. If the cell does not improve one of the elite plans, the cell information replaces a cell randomly selected from the vague buffer. This has two consequences: Firstly, agents keep information that is not relevant to themselves but that might be to others in the future, hence they adopt a *cooperative* behavior that is not supervised. Secondly, the information in the vague buffer can be erased and lost forever. The information about cells with very low utility is more likely to be lost from the collective memory. The exploration phase allows still to recover them. So far, the computation load of a single round is $O(A(\ln E + SN^{(S-1)}))$ which is feasible for reasonable assumptions ($S \leq 3, N < 50, E < 50$). However, the number of iterations is still to be determined. Note that the learning speed of the overall process depends on the *greediness* of the exchange which is a function of the ratio between the sizes of the elite and the vague buffers.

6 Social network dynamics

Social links disappear following an exponential decay law at the end of each iteration. Initially, the social network is a random graph of degree K and each social connection has the same decay time. When two agents perform activities at the same location, two situations can occur: 1) if there is a social connection between them, the strength of the connection is reinforced and its decay time increases 2) if not, a new social connection is created. This task requires $O(AK)$ operations for the enumeration of the social links and $O(C)$ operations to browse the cells for detecting interactions between agents. However, the detection of the existence of a social link between two given agents would require $O(AK)$ operations for each cell to browse the connections. An alternative solution would be to store the connections of a given agent in a hash table. Both solutions are costly in memory or computation time. For these reasons, we adopt a slightly different implementation which is equivalent statistically. Every connection has the same decay time but multiple redundant

connections can exist between two given agents. When an agent visits a particular cell from his own elite buffer, the cell/agent pair is stored until a second agent visits the cell. A new social connection is then created between the two visitors. After one iteration, we end up with a larger number of connections than at the beginning. All of them are of equal strength but with potential redundancy (a given pair of agents can appear several times). Identifying the redundancy would be costly. Instead, we keep the total connection strength constant in the system by deleting connections randomly so that the total number of connections is constant.

7 Implementation issues

The code of the simulation has been written in Java. Input and output files use the XML file format which is well suited to variable length content such as the description of the individual plans with multiple stops. The goal being to simulate 10^6 agents on a single CPU, some performance concerns have to be taken into account.

Cells each grid cell is stored as an individual object. With 10^5 cells, it is not crucial to store the cell attributes as plain arrays. This allows to keep cell characteristics private and to have cell references. Since cells have to be often compared during the learning process to determine if a new cell is already known to an agent, it is far more efficient to use the identity operator $==$ than the default Java *equals()* method which is the equivalence operator. This is valid as long as the cells are not dynamically allocated or cloned once the simulation starts. Cells have to keep references to agents that visit them (see the socialization stage). The average number of visits per cell is small ($O(SA/C)$) so that we could allow for the overhead of a dynamic container (e.g. vector). However, in the simple pairwise interaction described above, only the information about the last visitor is needed. Therefore, it is sufficient to maintain a hash table of visited cells and visitors.

Agents it is tempting to have a dynamic container steadily increasing in size for the agent memory. However, that would completely ruin the performances and is not compatible with our assumption that only a limited number of simultaneous options can be memorized. Each cell is referenced on average by $O(AN/C)$ agents so that the initial coverage is sufficient to ensure that there is not any information missing about the environment.

Random numbers a typical bottleneck of this kind of simulation is the computation of random numbers. A priori, $O(AK)$ random numbers have to be computed for each single iteration of the information exchange stage. This quickly becomes prohibitive and can be avoided by using two integer random seeds at the beginning of the exchange stage. One is used to pick a cell from the informer agent, the other to replace

a cell in the memory of the informed agent. These two pointers can be simply incremented from one social link to the other since there is no correlation between social links and they are accessed in a *a priori* random order.

Update sequence The original ordering of the agents is that of the XML source and could be biased. Agents are processed sequentially during the evaluation phase: each agent uses one of his elite plans and marks the visited cells. Therefore, the probability to create a social connection between agents decreases with their distance in the list. To avoid this potential bias, the list of agents is randomized at the beginning. Social connections are also stored initially in a random list since they are processed sequentially during the learning phase. New social connections are created at the end of the list. The decay process rearranges them randomly.

8   Results

The simulation is tested on a real-world example for the Zurich region for which we have available a high resolution transportation network and a land-use raster (see Figure 5). The area covers approximately a 50x50 kilometer square area where about one million inhabitants are living. The land-use utility values R_i are generated based on census data. Random plans with 1 or 2 intermediary stops are generated for 10^6 agents that are distributed on the area according to job and housing densities. The home to work pairs are computed using an external model written by one of the authors (see [7] for the computation of the rent values presented in Figure 5). The initial social network that connects them is a random graph. Obviously, this is not realistic but we intend to evaluate only the computational feasibility in this preliminary work.

Figure 1 presents the evolution of the sum of the scores of all the agents during the iterative process. It can be seen that the process converges in a few dozens of iterations but that the choices are not optimal since the utility does not reach the maximum value obtained with a full enumeration of the alternatives. This is due to the fact that some information is lost in the process. It can be recovered slowly through the exploration. However, some information can still be lost forever because the exploration mechanism is local: the agents cannot jump to explore a totally new area. This limitation could be easily removed. Still, the value of the plateau is high enough (more than 90% of the maximum utility) to ensure that plausible strategies have been selected. It remains to be studied how much this is compatible with empirical evidence. Note that the size of the memory of the agent only slightly affects the convergence properties.

Figure 2 illustrates a typical spatial adaptation process for a single agent with a two-stop plan: home-work-leisure-shopping-home. On the first iteration, leisure ($L1$) and shopping ($S1$) are performed at the home place because the agent ignores good locations to perform these activities. In the second iteration, he learns that the location $L2 - S2$ is a good location for one of the two activities, hence making the extra trip distance worth it. On

the third iteration, the agent discovers that the area around $S3$ has a high utility for shopping. During the fourth and fifth iterations, the agent keeps shopping close to that area and only optimizes the leisure location ($L3 \rightarrow L4 \rightarrow L5$).

Figure 3 shows the evolution of the distribution of social connections in the system. The initial condition is a random graph of degree $K = 20$. On the first iteration, the distribution is roughly a $N(20, 1)$ distribution. Progressively, the distributions shift to the left. In the end, we have a self-sustained distribution that can be approximated by a $N(13, 4)$ distribution. Therefore, most agents maintain between 9 and 17 connections. Note that redundant connections between a given pair of agents are not identified and are therefore counted multiple times. The decrease (from 20 to 13) in the average number of social connections indicates that the spatial interaction is not sufficient, in this case, to sustain 20 connections on average. This is dependent on the properties of the land-use data such as the concentration of high-utility areas. It also depends on the total number of agents in the system. (The area under the curve is equal to the total number of connections and thus constant.)

All the experiments were done on a computer equipped with a Intel Pentium 4 clocked at 2.5Ghz. The typical simulation performance for 100 iterations of a system with 10^6 agents is below one hour of CPU time. This is for plans that have only one or two intermediary stops. In terms of memory requirement, about 400 Mbytes of RAM are needed. Obviously, the simulation of larger systems and more sophisticated plans with more than two stops will require to distribute the workload on several computers. In particular, we would like to extend the framework to integrate other travel decisions such as the timing of the different trips.

9   Conclusions and perspectives

Many aspects still need to be addressed to improve the realism of the model. In particular, we plan to validate the model using the results of a recent survey [9] conducted in Germany and Switzerland. The goal was to study the factors that influence leisure trips. Respondents were asked to give the location and the frequency of the visits to their five closest friends or relatives. Empirical results suggest that the purpose of an important share of leisure trips is to visit social connections and that the number of known locations is small. These data should, in principle, allow to validate the model and to compare the spatial distribution of social connections with that of our model. This preliminary work has shown that a multi-agent based approach to the location of secondary activities is technically feasible and behaviorally plausible for high resolution data sets. The fact that agents cooperate in even some simplistic way yields an important gain in term of computation workload. This has to be compared with the standard practice in Transportation Science where it is typically assumed that users are in the situation of a non-cooperative Nash equilibrium.

Acknowledgments

References

[1] T. Arentze and H.J.P. Timmermans. Representing mental maps and cognitive learning in micro-simulation models of activity-travel choice dynamics. In *Proceedings of the meeting of the International Association for Travel Behavior Research (IATBR)*, Lucerne, Switzerland, 2003. See http://www.ivt.baum.ethz.ch/allgemein/iatbr2003.html.

[2] K.W. Axhausen, A. Zimmermann, S. Schönfelder, G. Rindsfüser, and T. Haupt. Observing the rhythms of daily life: A six-week travel diary. *Transportation*, 29 (2):95–124, 2002.

[3] M. Ben-Akiva and S. R. Lerman. *Discrete choice analysis*. The MIT Press, Cambridge, MA, 1985.

[4] E. Bonabeau, M. Dorigo, and G. Theraulaz. *Swarm Intelligence : From Natural to Artificial Systems*. Santa Fe Institute Studies on the Sciences of Complexity. Oxford University Press, 1999.

[5] A. de Palma and F. Marchal. Real case applications of the fully dynamic METROPOLIS tool-box: an advocacy for large-scale mesoscopic transportation systems. *Networks and Spatial Economics*, 2(4):347–369, 2002.

[6] Th. A. Domencich and D. McFadden. Urban travel demand. In D.W. Jorgenson and J. Waelbroeck, editors, *Urban travel demand*, number 93 in Contributions to Economic Analysis. North-Holland and American Elsevier, 1975.

[7] F. Marchal. A trip generation method for time-dependent large-scale simulations of transport and land-use. *Networks and Spatial Economics*, page forthcoming, 2005.

[8] B. Raney, N. Cetin, A. Völlmy, M. Vrtic, K. Axhausen, and K. Nagel. An agent-based microsimulation model of Swiss travel: First results. *Networks and Spatial Economics*, 3(1):23–41, 2003. Similar version Transportation Research Board Annual Meeting 2003 paper number 03-4267.

[9] R. Schlich, A. Simma, and K. W. Axhausen. Determinanten des Freizeitverkehrs - Modellierung und empirische Befunde. Arbeitsberichte Verkehrs- und Raumplanung 190, Institut für Verkehrsplanung und Transportsysteme (IVT),ETH Zürich, 2003. See www.ivt.baug.ethz.ch.

[10] P. Waddell, A. Borning, M. Noth, N. Freier, M. Becke, and G. Ulfarsson. Microsimulation of urban development and location choices: Design and implementation of UrbanSim. *Networks and Spatial Economics*, 3(1):43–67, 2003.

List of Figures

Figure 1: Performance of the multi-agent simulation
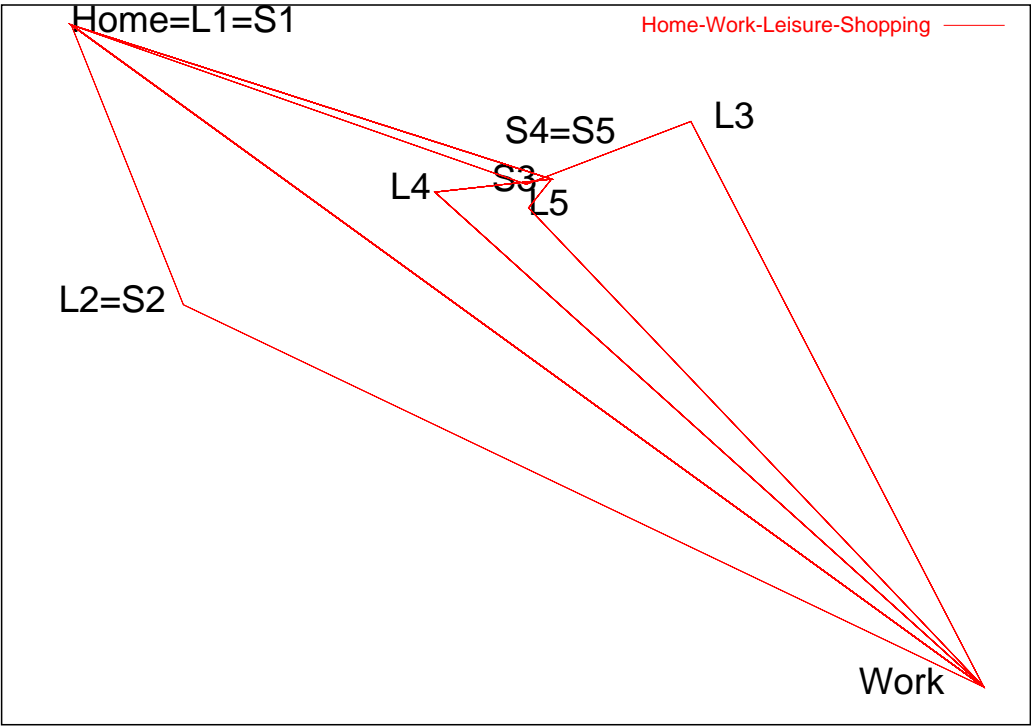
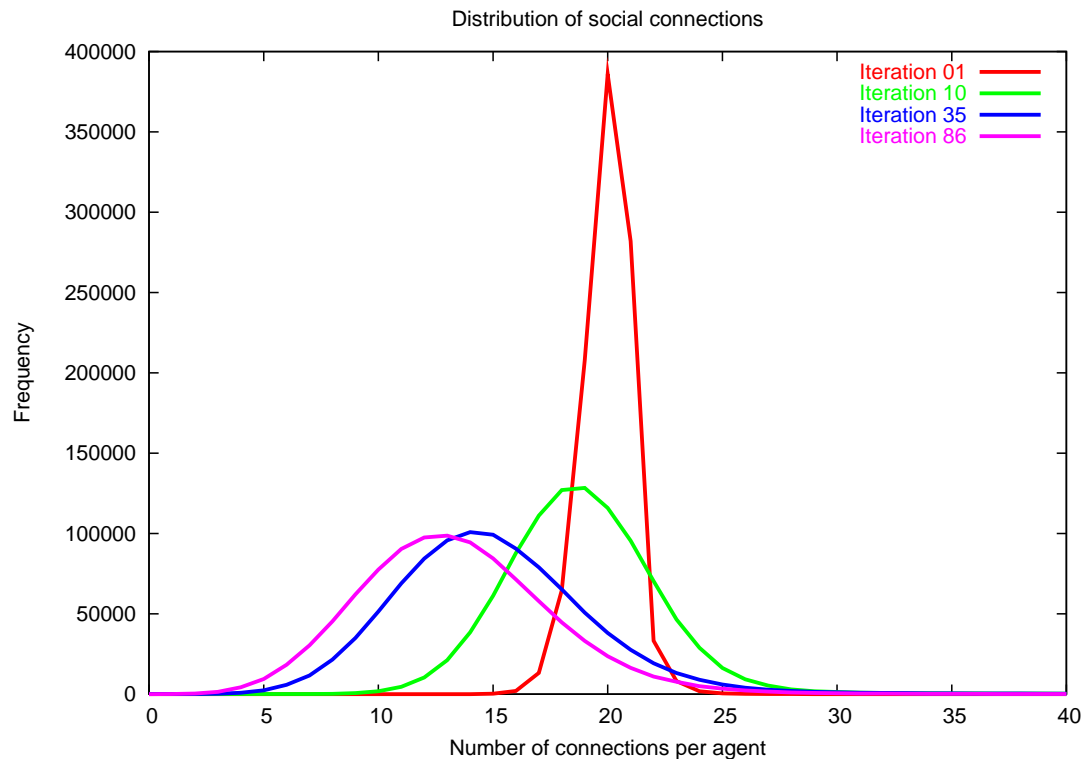Figure 2: Adaptation of the location for a two-stop plan
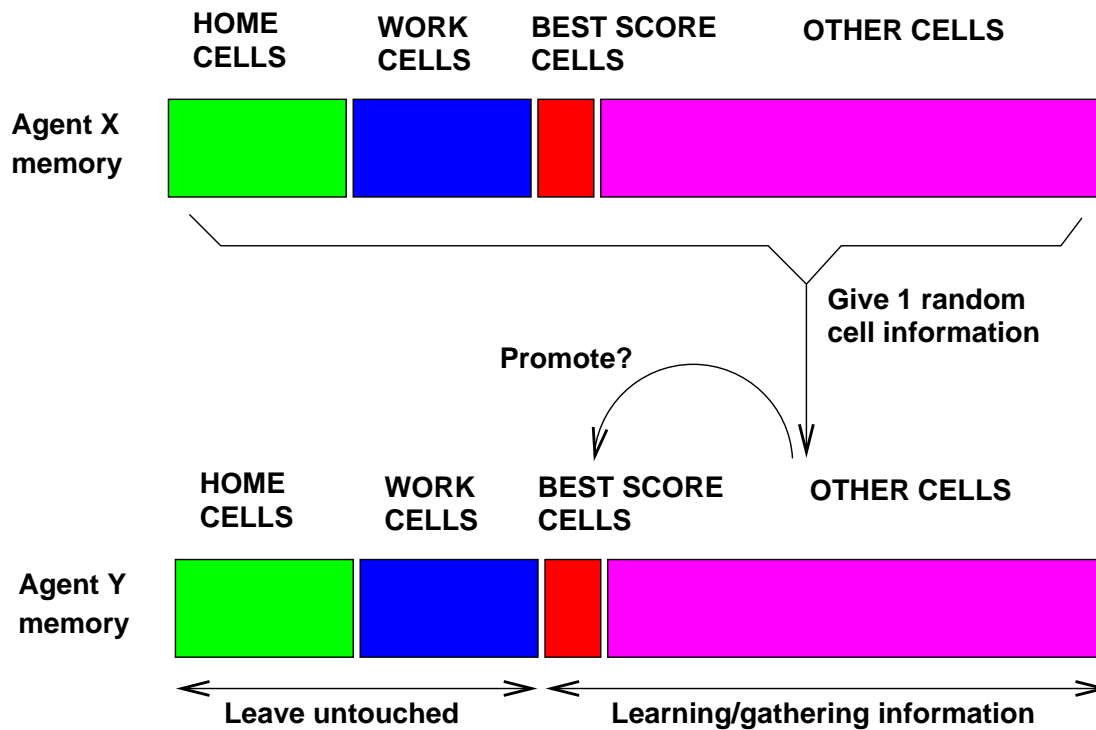
Figure 3: Distribution of social connections
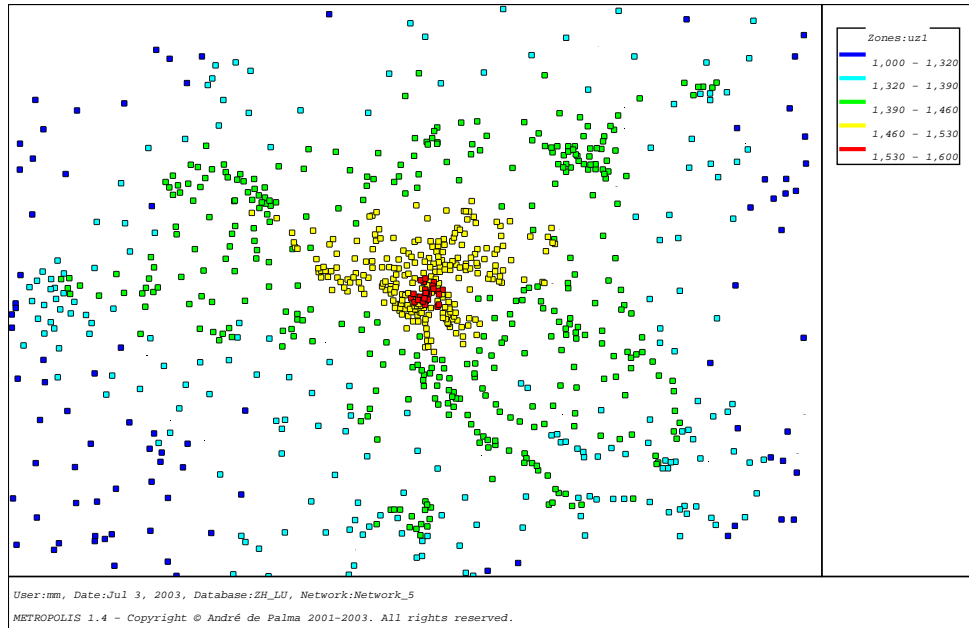
Figure 4: Agent memory organization and learning

Figure 5: Zurich area: transportation network and rent values