# 1. The proposed method

This chapter covers theoretical construction of the proposed method, so that it could be reproduced on any population using data from different sources and does not discuss how to prepare the input data, as this is strongly dependent on what data are held and what environment is being modeled.

## 1.1 General framework

From a holistic point of view, the achievement of the goals set in this thesis requires two main steps: building an agent-based model of urban mobility and modeling the contact structure based on information about encounters between agents retrieved from this model, which can be used in subsequent tasks such as modeling the spreading of an epidemic in a surveyed population of city dwellers. The main phases of the proposed framework are shown in the figure below:

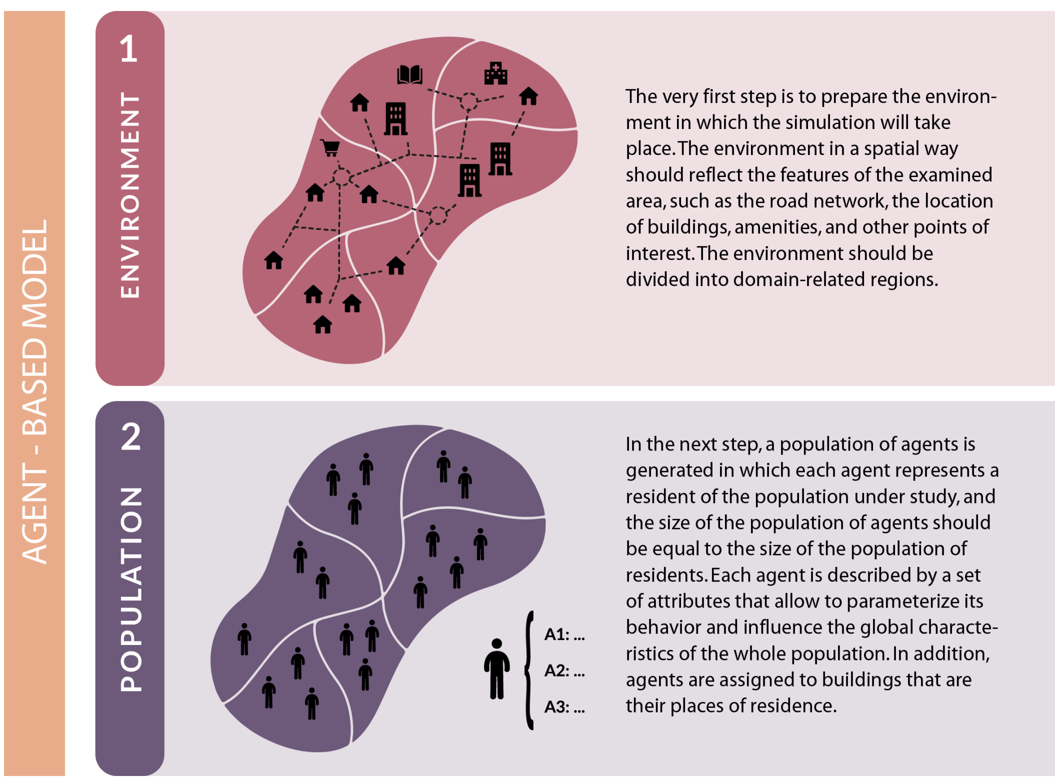


Figure 1.1: Two first steps in general framework of the proposed method.

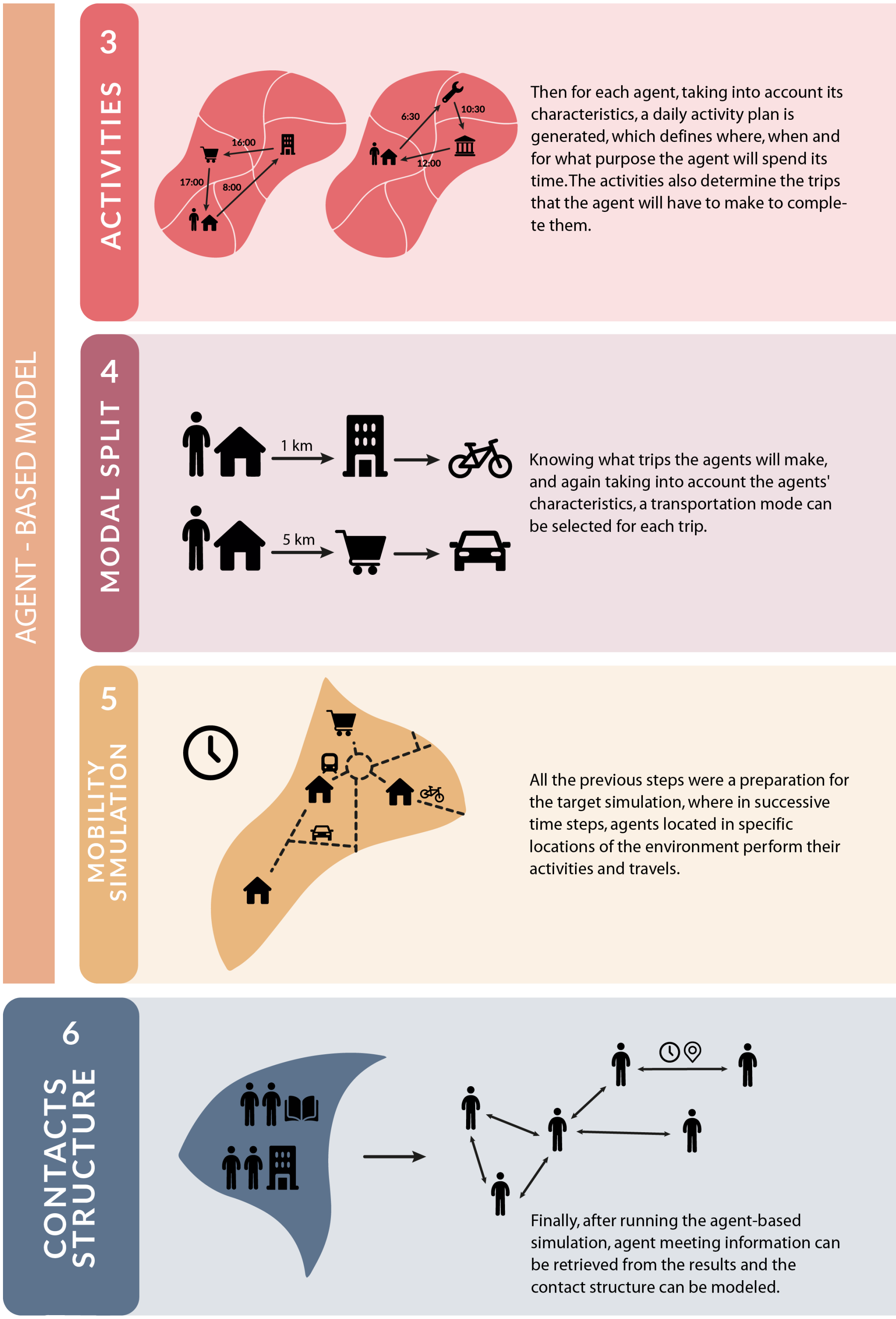


Figure 1.2: Continuation of Figure 1.1. 3-6 steps in general framework of the proposed method. Graphics prepared with icons made by Freepik, IconKanan, Kiranshastry, Ilham Fitrotul Hayat from www.flaticon.com.

## 1.2 Agent-based urban mobility model

In order to create an accurate description of the agent-based model allowing for its subsequent replication, the ODD (Overview, Design concepts, and Details) protocol proposed by Grimm et al. [3] was used. The protocol provides a common and structured way to describe the details of agent-based models based on written text and is intended to be human-readable.

The ODD protocol consists of 7 elements that can be organized into 3 main groups:

- context and general information of the model (Overview),
- description of the most common aspects in agent-based models design (Design concepts),
- technical details (Details). [2]

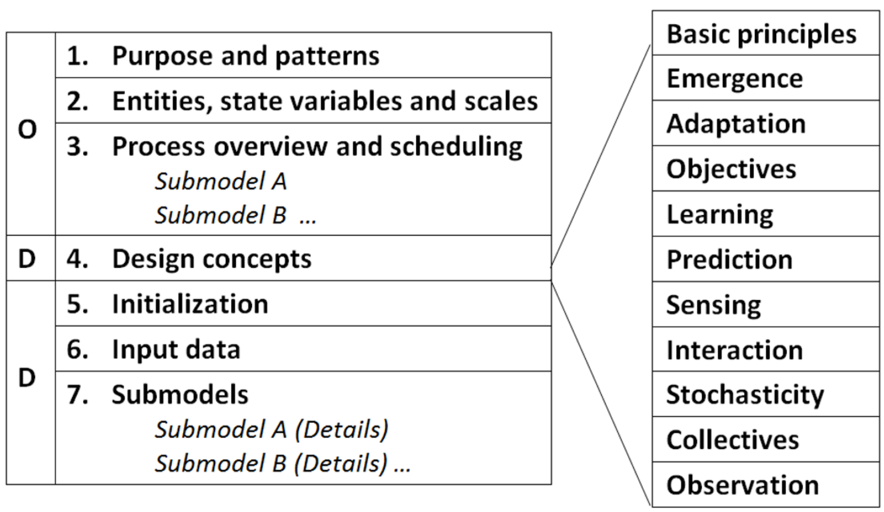


Figure 1.3: Structure of model description in the ODD protocol. The description consists of 7 elements grouped in 3 groups: O - Overview, D - Design concepts, D - Details. The table on the right presents 11 aspects to describe in Design concepts, if some of them do not apply, can be skipped. Figure source: [3]

### 1.2.1 Purpose and patterns

The purpose of this model is to enable the analysis of mobility and activities of a city's residents in terms of one day to later model and analyze the one-day contact structure, but also to explain how mobility and activities emerge from knowledge about population and environment and how it affects this structure.

To achieve this purpose, the model should reflect the daily living patterns of the modeled population.

### 1.2.2 Entities, state variables and scales

Low-level entities:

- **agent** - The agent is the main object of the simulation. It moves around the environment and performs its activities. Each agent is described by a set of basic variables: gender, age, housing, but also by a set of social characteristics that affect its behavior as well but will vary in specific simulation scenarios because they are closely related to what population data has been collected. The specific variables used in the simulations for this thesis are described in a later section.
- **home building** - An spatial object representing a building that can be associated with an agent as its residence. It is described by its spatial location in the environment.
- **POI (point of interest)** - An spatial object in which agents can perform their activities. It is characterized by the type of activity that can be performed there and its location in the environment. A POI is usually located in the area of an existing building, but it is not closely related to it, because: there are buildings without POIs (residential buildings), there are POIs located outside buildings (for example public transportation stops), there may be multiple POIs in one building.
- **road** - A representation of a road as a connection between two spatial points in the environment. Each road is described by the types of transport that can use, the maximum speed of movement, the length, and flow capacity, that is how many travelers can enter or leave the road per time step.
- **transport mode** - A transport mode is an abstract object that represents how a trip is made. It can be implemented as crossing roads with an assigned type of vehicle or as teleportation, in which case the agent appears at the destination after a time determined by the distance to be traveled and the speed assigned to the mode of transportation. Making a trip is described in detail as a process in the next section. The choice of available transport modes is related to the collected data and the type of mobility to be studied in the simulations.
- **vehicle** - As mentioned above, a transport mode can be associated with a vehicle, which is an object using which the agent can move on the road. It is described by the maximum speed and the number of seats.

High-level entities:

- **activity** - An activity is an object representing the fact that an agent is performing, at a specific location (POI or home building) and for a specified time, one of the available types of activities.
- **travel** - Representation of the movement between a starting point and a destination with an indicated transport mode. If the transport mode is not simulated by teleportation, the travel also describes the subsequent stages of movement: getting in and out of vehicles, overcoming the next of the designated roads. Travel can also be associated with the activity performed before and after it.

- **schedule** - An individual set of agent consecutive activities and travels made between them.
- **population** - The set of all agents generated in the simulation that can be described by the distributions of basic and social feature values among the agents.
- **roads network** - The collection of all roads forms a road network that allows agents to move around the environment.
- **environment** - An environment is created through a collection of home buildings, POIs, and roads. In a spatial context, it is the set of all spatial points occurring and described through these subsets.
- **region** - In terms of domain and space, a coherent part of the environment, that is a subset of spatial entities selected by the modeler that creates a connected space. A region can be described by attributes such as population size, number of given POIs, number of jobs, number of school seats.

Scales:

- **spatial** - The spatial scale is limited to the area of one city. The modeling does not include travels that start or end outside this city.
- **temporal** - The temporal scale is limited to one day with a small-time step of 1 second to accurately track the positions of agents in the environment.

### 1.2.3 Process overview and scheduling

This section presents the processes that exist in the model and their scheduling. The main steps of agent-based mobility modeling are shown in Figure 1.4. These steps are shown as separate processes in the following figures, and some of the elements that constitute them are described in detail in the "submodels" section of this protocol (1.2.7).

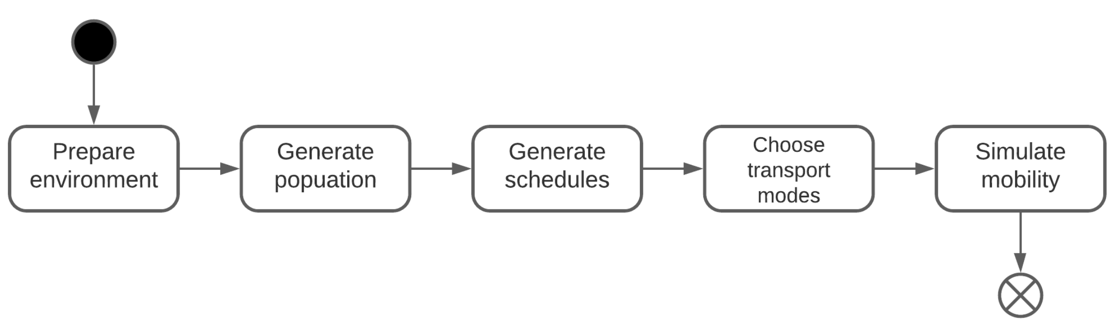


Figure 1.4: Main steps of agent-based urban mobility model

**Prepare environment**

The process that results in the preparation of all data and the creation of the environment. At this stage, it is not described in detail as it depends on the specific case of using the proposed framework. This process can be run once, even if the simulations are run multiple times, as long as they are to be run in the same environment.

**Generate population**

The population is generated by running for each agent the process below:

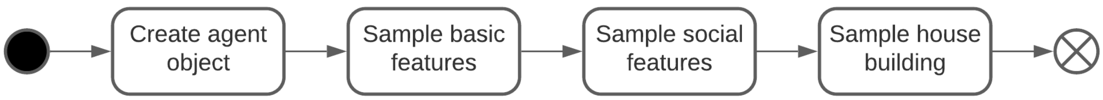


Figure 1.5: An agent creating process

**Generate schedules**

A schedule for each agent is generated with the following process:

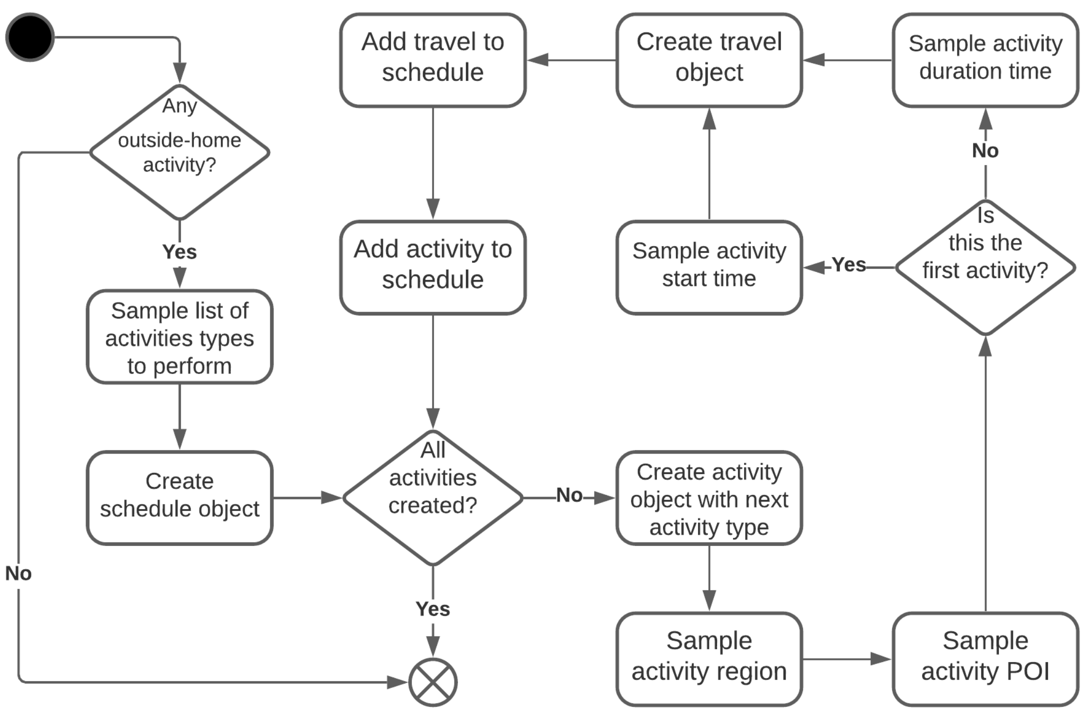


Figure 1.6: A schedule generating process. It is assumed that the agent starts and ends his day at home.

**Choose transport modes**
For each agent, a process is started that selects the transport mode for each travel.

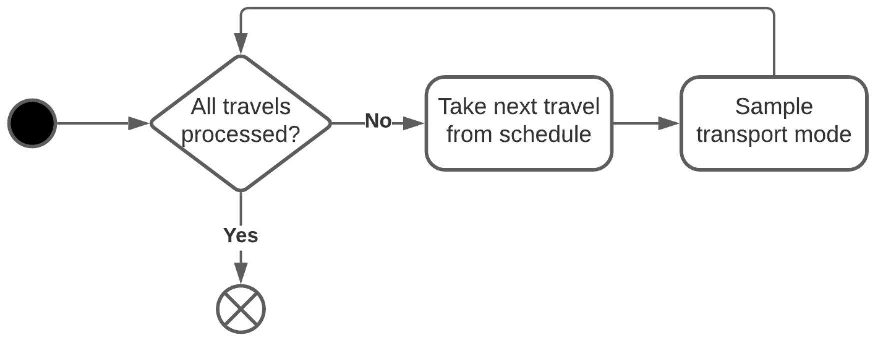


Figure 1.7: Transport mode selection process.

**Simulate mobility**
During final mobility simulation each agent run in parallel with the following process:

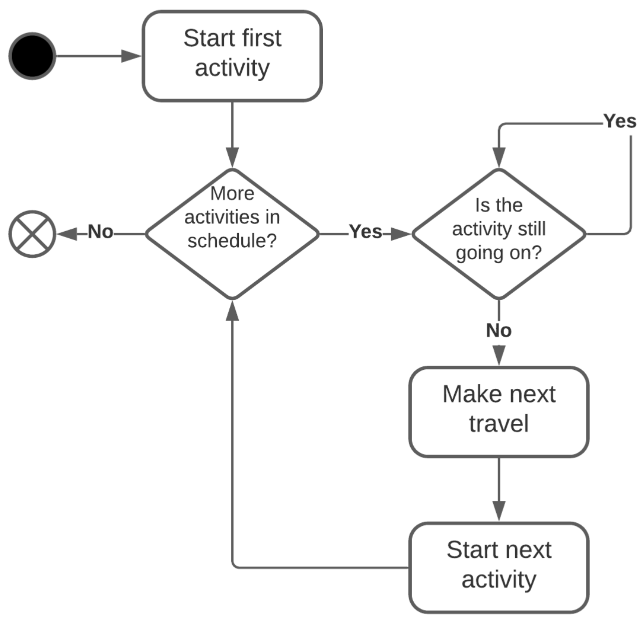


Figure 1.8: Mobility simulation process.

### 1.2.4 Design concepts

**Basic principles**

The main principle of this model is the existence of a relationship between mobility emerging from human behavior and the structure of contacts they create. Modeling this mobility follows the already existing agent-based approach to modeling mobility, described in **??**, but is distinguished by the fact that it allows to accurately capture interactions between agents in the temporal and spatial context on a city-wide scale.

**Emergence**

The phenomena that emerge from the behavior of individual agents are: traffic load and congestion, public transport vehicles load, the density of agents performing their activities in given areas, as well as the main topic of this thesis, the contact structure, is a phenomenon that is the result of individual behavior.

**Adaptation**

The only adaptive mechanism is hidden in the fact that activities have a duration time specified, rather than a start and end time. The start time is thus dependent on the time when the previous activity ended and the travel time. If the travel time increases due to traffic congestion, for example, then the start and end times of the activity shift, and hence the agent will leave the activity location later. In reality, some activities will end at the scheduled time despite the lateness of the person performing them, but such a mechanism was chosen as a good enough representation of reality in the model.

**Fitness**

There is no explicit fitness function. Instead, the entire process is evaluated in terms of how the model results and the resulting contact structure reflect knowledge about the population being modeled.

**Prediction**

Agents do not predict the future. They know their future steps from prepared schedules, and transport modes are selected based on static knowledge about the environment and the agent.

**Interaction**

Agents interact by meeting on roads and influencing each other's movements. The fact of performing activities in the same places can also be considered as an interaction between agents - such interactions do not affect the state of the agents and the environment, but are recorded for later modeling of the contact structure.

**Stochasticity**

Most of the submodels used are stochastic but not completely random because their outcomes are based on inputs most often derived from observations of a real population. Each stochastic submodel is detailed in "submodels".

**Collectives**

The collectives are modeled as emerging entirely from agents' features and behaviors, hence they do not impact agent actions but are very useful in simulation outcome analysis. Firs type of collectives is groups derived from agents' basic features:

- age groups,
- sex groups,

- age-sex groups (for simplifying data processing agents are also assigned to groups with specified age and sex, set of possible age-sex groups results from allowed agents' age values and sex values)
- agents living in the same regions.

The other groups are separated in the context of activities and POIs:

- students - agents performing school activities,
- workers - agents performing jobs-related activities,
- agents performing other activities.

**Observations**

Everything that happens in the model is recorded in an event log along with metadata describing: who was involved in the event, where and when the event occurred. Events include:

- start and finish an activity,
- start and finish a travel,
- enter and leave a vehicle,
- enter and leave a road.

### 1.2.5 Initialization

The model has a complex initialization because 4 out of 5 processes shown in the figure 1.4 can be considered to belong to it:

- "prepare environment" process,
- "generate population" process,
- "generate schedules" process,
- and "choose transport modes" process.

The above processes have already been described in the previous sections. The decision to prepare the agents' entire schedule and choose transport modes as part of the processes included in the model initialization, instead of carrying out separate decisions of agents during the simulation, results from the need to simplify the simulation stage due to its large scale, and at the same time, such a solution is sufficient to achieve the assumed goals, because it allows simulate the mobility of an average day and is consistent with how many people plan their day, and the decision-making process of agents during the simulation is not crucial in these studies.

### 1.2.6 Input data

The data needed to enter the simulation can be divided into two groups: environmental information and probabilistic data for stochastic models.

**Environmental information:**

- localization of home buildings,
- localization and type of POIs,
- roads network,
- areas of regions and their attributes,
- public transport transit data in GTFS (General Transit Feed Specification) format, which is a popular standardized format for publishing data by public transportation operators that allows it to be loaded by various types of applications. [1]

**Probabilistic data:**

- set of probabilities $r = \{r_1, r_2, ..., r_k\}$, where $r_i$ - probability that agent lives in region $i$, $k$ - total number of regions in environment, $\sum_{i=1}^{k} r_i = 1$
- set of probabilities $g = \{g_1, g_2, ..., g_k\}$, where $g_i$ - probability that agent belongs to age-sex group indicated by index $i$, $k$ - total number of possible age-sex groups, $\sum_{i=1}^{k} g_i = 1$
- set of probabilities for each $m$-th social feature and for each $n$-th age-sex group $s_{m,n} = \{s_{m,n_1}, s_{m,n_2}, ..., s_{m,n_k}\}$, where:
  - $s_{m,n_i}$ - probability that agent in $n$-th age-sex group has $i$-th value of $m$-th social feature,
  - $k$ - total number of all possible values of $m$-th social feature,
  - $m \in M$ and $M$ - set of all specified social features,
  - $n \in N$ and $N$ - set of all possible age-sex groups,
  - $\forall m \in M \quad \forall n \in N, \quad \sum_{i=1}^{k} s_{m,n_i} = 1$
- probability $o_n$ for each $n$-th age-sex group, where $o_n$ - probability that agent in $n$-th age-sex group will perform any outside-home activity and $\forall n, 0 <= o_n <= 1$
- set of probabilities for each $n$-th age-sex group $c_n = \{c_{n_1}, c_{n_2}, ..., c_{n_k}\}$, where:
  - $c_{n_i}$ is probability that agent in $n$-th age-sex group will perform $\mathbf{c}_{n_i}$ sequence of activity types, each sequence may contain any positive number of ordered activity types for the agent
  - $k$ - total number of all possible sequences of activity types,
  - $n \in N$ and $N$ - set of all possible age-sex groups,

  - $\forall n \in N, \quad \sum_{i=1}^{k} c_{n_i} = 1$

- $\mu_{n,a}$ and $\sigma_{n,a}$ for each $n$-th age-sex and for each activity type $a$, where:
  - $\mu_{n,a}$ - mean duration time (in minutes) of activity of type $a$,
  - $\sigma_{n,a}$ - time duration (in minutes) standard deviation of activity of type $a$,
  - $n \in N$ and $N$ - set of all possible age-sex groups,
  - $a \in A$ and $A$ - set of all specified activity types,
  - $\forall n \in N \quad \forall a \in A, \quad \mu_{n,a} > 0 \quad and \quad \sigma_{n,a} > 0$

- for each activity type $a$, set of probabilities $h_a = \{h_{a_1}, h_{a_2}, ..., h_{a_k}\}$, where:
  - $h_{a_i}$ - probability that agent will start activity of type $a$ at hour $i$,
  - $k$ - maximum activity start hour,
  - $a \in A$ and $A$ - set of all specified activity types,
  - $\forall a \in A, \quad \sum_{i=1}^{k} h_{a_i} = 1$

### 1.2.7 Submodels

The proposed submodels are described the table 1.1 with the order consistent with their occurrence in the processes presented in 1.2.3. The table keeps the data notation claimed in 1.2.6.

Table 1.1: Method submodels

| Begin of Table | | | |
|---|---|---|---|
| **Step** | **Input data** | **Outcome** | **Submodel** |
| Sample basic features (Figure 1.5) | $r$, $g$ | region of residence, age and sex | Agents' region of residence is sampled from Multinomial Distribution $Mult(1, r)$; age and sex are sampled as age-sex group from $Mult(1, g)$. |
| Sample social features (Figure 1.5) | age-sex group $n$, $\forall m \in M$, $s_{m,n}$ | social features | Knowing that agent belongs to age-sex group $n$, each of it's social feature $m$ is sampled from $Mult(1, s_{m,n})$. |
| Sample home building (figure 1.5) | region of residence, home buildings | home building | Knowing that agent lives in given region, it's home building is sampled from uniform distribution over all buildings located in this region. |

Continuation of Table 1.1

| Step | Input data | Outcome | Submodel |
|---|---|---|---|
| Any outside activity? (Figure 1.6) | age-sex group $n$, $o_n$ | Agent will carry outside-home activities or not. | Knowing that agent belongs to age-sex group $n$, outcome is a sample from Bernoulli Distribution $Bern(o_n)$. |
| Sample sequence of activity types (Figure 1.6) | age-sex group $n$, $c_n$ | sequence of activity types | Knowing that agent belongs to age-sex group $n$, the sequence of activity types it will perform is sampled from $Mult(1, c_n)$. |
| Sample activity region (Figure 1.6) | regions, agent's location, activity type $a$ | the region where the activity will take place | Region is sampled from $Mult(1, p)$, where: $p = \frac{v-min(v)}{max(v)-min(v)}$, $v = (v_1, v_2, ..., v_k)$, k - total number of regions in the environment, $v_i = \frac{f_i}{dist(reg_c, reg_i)}$, $f$ - a vector of the selected feature of regions, that is an attractor to perform activity of type $a$ in the region, $dist$ - a function that returns the distance between two regions, $reg$ - a set of the regions in the environment, $reg_c$ - a region in which the agent is currently located |
| Sample activity POI (Figure 1.6) | activity type $a$, activity region | the POI where the activity will take place | Knowing that activity will take place in given region, activity POI is sampled from uniform distribution over all POIs located in this region and mapped to given activity type. |
| Sample activity duration (Figure 1.6) | age-sex group $n$, activity type $a$, $\mu_{n,a}$, $\sigma_{n,a}$ | activity duration time in minutes | Activity duration time, with respect to agent's age-sex group $n$ and activity type $a$, is sampled from Normal Distribution $N(\mu_{n,a}, \sigma^2_{n,a})$ |

| Continuation of Table 1.1 | | | |
|---|---|---|---|
| **Step** | **Input data** | **Outcome** | **Submodel** |
| Sample activity start time (Figure 1.6) | activity type $a$, $h_a$ | activity start time in minutes | Activity start time, with respect to activity type $a$, is determined according to the following formula: $Multi(1, h_a) * 60 + Unif(0, 60)$, where $Unif$ stands for Uniform Distribution. |
| Choose transport mode (Figure 1.7) | agents' basic features, agents' social features, travel distance | transport mode | Choosing a transport mode is a decision-making process performed by a submodel $\Psi$(agents' basic features, agents' social features, travel distance). The submodel can be any decision model, such as a machine learning model fitted to the collected real-world data. Also, the input feature set is a representative example, but may be reduced or augmented by other travel, agent, or environment features, depending on the model used and the specific use case. |

The submodel that needs to be described in a different context is the element that performs the entire "Simulate mobility" process (Figure 1.8 . Under this model is the prepared computer software that takes the following input data:

- environment: regions, home building, POIs, road network,
- agents and agents' schedules,
- GTFS transit data,

then performs a mobility simulation using various means of transport and finally returns the event log. This can be custom-made software or one of the existing solutions. One of the solutions that meets these requirements well and was used for this work is Multi-Agent Transport Simulation (MATSIM) - an open-source framework for implementing large-scale agent-based transport simulations [4]. The project was launched in 2006 as a tool to generate traffic patterns based on individual synthetic agent journeys in their daily or weekly activities. It currently provides a framework consisting of several modules that can be combined or used separately. The modules can also be replaced with custom implementations. The basic modules offer a framework for demand modeling, agent-based mobility simulation (traffic flow simulation), re-planning, iterative simulation runtime, as well as modules to analyze simulation results. The basic workflow proposed by the authors is shown in Figure 1.9.

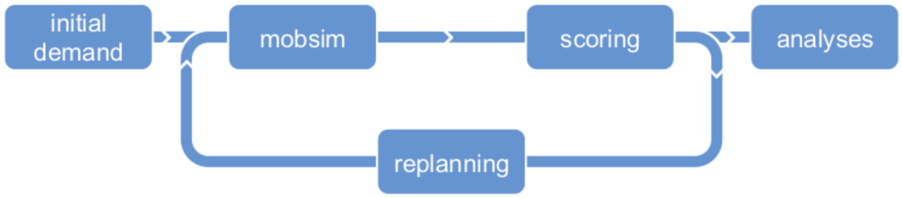


Figure 1.9: Basic MATSIM workflow. For the purposes of this work, only the "mobsim" module was used. Image source: [4]

For the described submodule, only the part of the software that performs the "mobsim" task, that is carrying out the mobility simulation, was used. Task "initial demand" is performed by the other submodules described in the table 1.1, while "scoring" and "replanning" have been omitted because, according to the assumptions of the model, the agents perform their schedules without introducing any modifications.

The "mobsim" module prepares the travel routes for non-teleported transport modes, including the use of public transport vehicles simulated according to a given transit data, and then uses the computationally efficient queue-based approach in which vehicle entering a road from a previous roads intersection is added to the tail of the waiting queue and remains there until all of the following conditions are met:

- the time for traveling the link with the free flow has passed,
- the vehicle is at the head of the queue,
- next link allows entering.

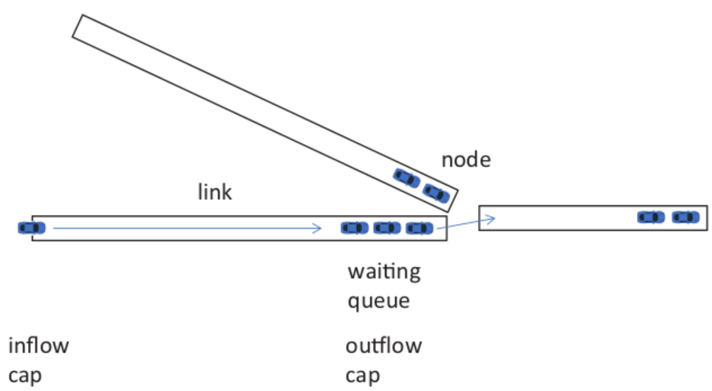


Figure 1.10: Queue-based traffic flow model ("cap" for "capacity"). Image source: [4]

To finish describing the submodels, it is also necessary to comment on a few decisions made while designing them:

- when creating schedules, the entire sequence of activities is sampled at once, because the real data collected in traffic diaries, often available for cities, usually covers most possible sequences of activities sufficiently and allows you to build such a model, and at the same time such a model is much simpler than modeling a sequence as a chain of draws, where each draw depends on all previous draws and additional factors such as the time of day,
- selecting activity region first and then an activity POI is because it is easier to gather features that attract activity to regions than to individual POIs, and therefore POIs within regions are sampled from the uniform distribution,
- a similar situation applies to the sampling of the region of residence, and then the building - data on the population of the indicated regions of the city is usually much better available than data on individual buildings,
- limiting the model to determining the start time of the first activity only to determining the hour and adding random minutes is because from all activities presented in real activity diaries, a small part concerns activities that start the day and usually, there are too few of them for more precise modeling, therefore they are grouped to hours and random minutes are meant to replace lost variety.

## 1.3 Contact structure modelling

The next step of the proposed method is to model the contact structure based on the event log resulting from the simulation. Recovering the contacts from the log should start by reading all the activities and storing them as an undirected bipartite graph, where one group of vertices is the agents and the other group is the places where the agents can perform activities, both points and vehicles, since for this step a vehicle ride should also be considered as an activity. Edges in the graph can connect agents to places and signify that an agent has performed an activity at a particular place, and the attributes assigned to the edges specify the type and time of the activity taking place.

Next, prepare a projection of the graph on the vertices of the agents, that is, omit the vertices of places in the graph, but connect the vertices of the agents if they performed activities at the same place and time. Now the edges represent contacts and are described by the type of activity performed by agent one, the type of activity performed by agent two, and the duration of the contact. In the all-day view, all agents performing activities at a given location can be merged to form a single component, but note that edges also have a temporal aspect and can be also analyzed only about edges representing contacts within a certain time interval.

In some cases, the modeling of the contact structure could end up with the obtained structure. However, the fact that agents are connected with all other agents residing at the same place and time can be problematic. Especially in places like schools or

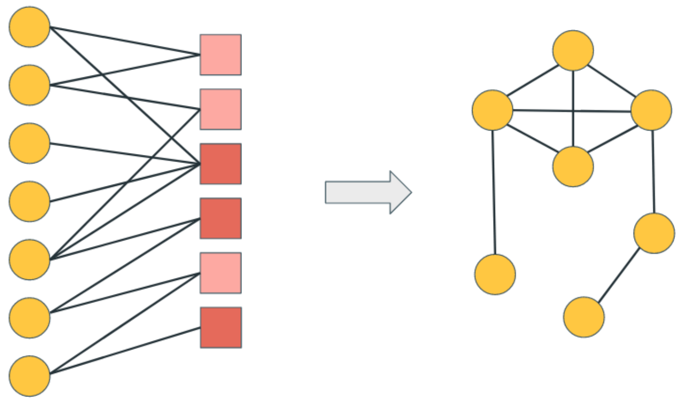

Figure 1.11: Left: activity graph reconstructed from simulation results. Agents (circles) are connected to the places where they performed activities, that is points and vehicles (squares). Edges are described by the type and duration of the activity. Right: projection of a graph in which agents are connected by edges denoting interactions if they performed activities at the same place and time. Interactions are described by activity type and duration.

workplaces, this can create unrealistically large cliques. For this reason, the contacts obtained so far should be treated rather as an opportunity for contacts to occur, and the proposal for further modeling of contacts that have occurred is presented in the algorithm below. The algorithm models contacts in a single place, so it should be run for all of them.

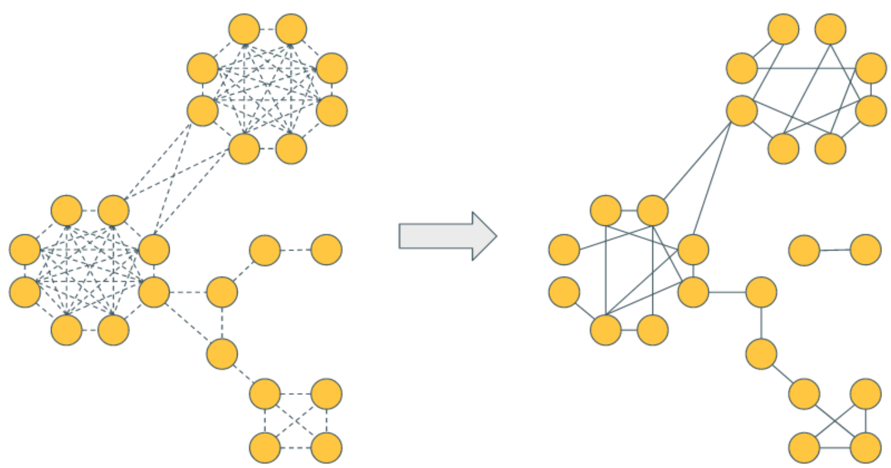

Figure 1.12: Because of the presence of unnaturally large clicks, the contacts obtained in the projection shown in the Figure 1.11 can be treated as possible contacts, and their occurrence can be further modeled (Algorithm 1).

**Algorithm 1** Modeling the occurrence of contacts among all possible contacts within a single place

**Input:**
$l$ - processing place (POI or vehicle)
$G(A, C)$, where:
$G$ - network of all possible contacts in processing place $o$,
$A$ - set of nodes representing agents having contacts in G,
$C$ - set of edges representing contacts in G
$Y$ - set of possible activity types
$a \in A, c \in C, y \in Y \quad and \quad c = ((a_1, y_1), (a_2, y_2), t, l)$, where:
$a_1$ - the first of the agents in contact,
$y_1$ - type of the activity performing by the first agent,
$a_2$ - the second of the agents in contact,
$y_2$ - type of the activity performing by the second agent
$t$ - contact time, $l$ - contact location,
$e_{y_1,y_2}$ - expected number of all contacts of agent performing activity $y_1$ with agents performing activity $y_2$
$g$ - the probability that two agents in contact performing a school activity come from different age groups
$\gamma$ - minimum probability to add a contact to network under modelling

**Output:**
$H(A, C)$ - network of existing contacts in processing place, network H is subnetwork of network G

**Def** $subset(C, \overline{a}, \overline{y_a}, \overline{y_o})$:
returns subset of contacts $B$ from set $C$, where:
$\forall c \in B$
$c = ((a_1, y_1), (a_2, y_2), t, l)$
$c \in C$
$(a_1 = \overline{a}, y_1 = \overline{y_a}, y_2 = \overline{y_o}) \quad or \quad (a_2 = \overline{a}, y_2 = \overline{y_a}, y_1 = \overline{y_o})$

**Def** $sort(C, B, \overline{a})$:
returns set of contacts $B$ sorted in ascending order of the values calculated for each contact $c$ in $B$ according to the following formula:
$c = ((a_1, y_1), (a_2, y_2), t, l)$
$if \quad a_1 = \overline{a}$:
$then:$
$return \quad |subset(C, a_2, y_2, y_1)|$
$elif \quad a_2 = \overline{a}:$
$return \quad |subset(C, a_1, y_1, y_2)|$

**Def** $Bern(p)$:
returns result of Bernoulli Trial with probability of success equals $p$

```
initialize network H
for each a ∈ A do
    for each y_a ∈ Y do
        for each y_o ∈ Y do
            K ← subset(C, a, y_a, y_o)
            K ← sort(H, K, a)
            if ¬(y_a = "school" & y_o = "school") then
                d ← |K|
                u ← |subset(H, a, y_a, y_o)|
                r = min(max((e_{y_a,y_o} − u)/d, γ), 1)
                for each c ∈ K do
                    if Bern(r) then
                        add c to H
                    end if
                end for
            else
                K_s ← all contacts from K where agents are in the same age group
                K_d ← all contacts from K where agents are in different age groups
                K_s ← sort(H, K_s, a), K_d ← sort(H, K_d, a)
                d ← |K_s| + |K_o|
                u ← |subset(H, a, y_a, y_o)|
                r = min(max((e_{y_a,y_o} − u)/d, γ), 1)
                i ← 0
                repeat
                    i ← i + 1
                    if Bern(g) then
                        if Bern(r) then
                            c ← next contact from K_d
                            remove c from K_d
                            add c to H
                        end if
                    else
                        if Bern(r) then
                            c ← next contact from K_s
                            remove c from K_s
                            add c to H
                        end if
                    end if
                until i > d
            end if
        end for
    end for
end for
```

The general assumption proposed in the Algorithm 1 is that the modeler indicates the expected average number of contacts of a given type that will occur to individual agents, where the type of contact is determined by the type of activity performed by the first and second agents.

Therefore, to complete the task, the algorithm processes the network of contacts possible to occur at a given location (network $G$) and performs the contacts separately of the agent (loop in line 2) and the possible contact type (loops in lines 3 and 4). The occurrence of each contact is randomized from a Bernoulli Distribution (line 12), where the probability of success (line 10) is computed based on the expected number of contacts ($e_{y_a,y_o}$ in line 10) and the total number of possible contacts (line 8), but taking into account contacts that have already occurred for this agent (line 9) when processing earlier agents (the earlier $a$ in the loop in line 2), since contact is a two-way relationship.

The collection of possible contacts (line 5) that are considered one by one (line 11) is sorted (line 6) in such a way that contacts with agents for which there are currently fewer contacts of the type compatible with the processed relation are considered first. This solution is supposed to limit the formation of vertices with a significantly higher degree.

The probability of success of a Bernoulli Trial called $r$ (line 10), is calculated in such a way that $\gamma <= r <= 1$ and $\gamma > 0$ what yields $0 < r <= 1$ and is consistent with the assumptions of the distribution $0 <= r <= 1$, but at the same time the introduction of $\gamma$ can improve the stability of the calculation and allows more contacts to occur than their expected number.

Slightly more complex is the second part of the algorithm (below line 16), where only activities between students, that is agents performing school activities, are considered. The idea of considering contacts remains the same, but the collection of possible contacts is split into two groups: a group of contacts with agents in the same age group (line 17) and a group of contacts with agents in a different age group (line 19). The addition of contacts is then considered as many times as the number of all possible contacts, and each time a decision is first made whether the next contact will be with an agent from a different age group (line 26) or the same age group (line 32). This is to reflect an important feature of student-to-student meetings in schools, where the majority of contacts are with people of the same age group, which is a natural consequence of dividing students into classes.

## 1.4 Method summary

The proposed method according to all its assumptions will allow meeting the objectives set in this work because:

- its result is a network representing contacts between members of the population in one day (Objective **O1**),
- the network is created from contacts that occur during the activities and travels performed by agents (Objective **O2**),
- agents are described by a set of basic demographic features and social features (Objective **O3**),
- agents can perform activities in places designated by the environment, travel along with a given road network, and use the available public transport system, and all these influences where and when contacts are made (Objective **O4**),
- the POI in which it happened is saved together with the contact, and each POI has a specific location in the environment (Objective **O5**),
- the exact duration of each contact is known as it is derived from the time agents spends in the same location (Objective **O6**).

# 1. Bibliography

[1] Gtfs: Making public transit data universally accessible. https://gtfs.org/. Accessed: 2022-01-08.

[2] Volker Grimm, Uta Berger, Finn Bastiansen, Sigrunn Eliassen, Vincent Ginot, Jarl Giske, John Goss-Custard, Tamara Grand, Simone K Heinz, Geir Huse, et al. A standard protocol for describing individual-based and agent-based models. *Ecological modelling*, 198(1-2):115–126, 2006.

[3] Volker Grimm, Steven F Railsback, Christian E Vincenot, Uta Berger, Cara Gallagher, Donald L DeAngelis, Bruce Edmonds, Jiaqi Ge, Jarl Giske, Juergen Groeneveld, et al. The odd protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. *Journal of Artificial Societies and Social Simulation*, 23(2), 2020.

[4] A Horni et al. (eds.). *The Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press, 2016. ISBN 978-1-909188-76-1.