



**Aalto University**  
School of Science



MS-E2177 Seminar on Case Studies in Operations Research

# Pathfinding in agent-based people flow simulation

Final report

Client: KONE

16.05.2016

Jani Strandberg  
Saku Metsärinne  
Sami Kari-Koskinen (project manager)

## Contents

<b>1</b>	<b>Introduction .....</b>	<b>3</b>
1.1	Objectives .....	3
1.2	Scope of the study .....	4
1.3	Structure of the report.....	4
<b>2</b>	<b>Literature review.....</b>	<b>5</b>
2.1	Research on people flow simulation .....	5
2.2	The principle of least effort .....	6
2.3	Hot spot network.....	6
2.4	Dijkstra's algorithm.....	7
<b>3</b>	<b>Research design .....</b>	<b>9</b>
3.1	Data.....	9
3.2	Methodology.....	12
<b>4</b>	<b>Model implementation .....</b>	<b>13</b>
4.1	Model design and parameters .....	13
4.2	Model assumptions.....	16
4.3	Model logic .....	17
4.4	Additional functionalities of the model.....	17
<b>5</b>	<b>Results .....</b>	<b>18</b>
<b>6</b>	<b>Model validation .....</b>	<b>23</b>
<b>7</b>	<b>Conclusions.....</b>	<b>24</b>
	<b>References.....</b>	<b>27</b>
	<b>Appendix 1: Sample layout and hot spot network.....</b>	<b>28</b>
	<b>Appendix 2: Self-evaluation.....</b>	<b>29</b>

# 1 Introduction

Every day, we surround ourselves with lots of people that pass by, and that is not going to diminish in the future. Driven by worldwide megatrends, such as population growth, urbanization and globalization, population density and people flow intensity increase dramatically in many locations around the world. Cities, and therefore individual buildings, must adapt to the changing environment and place special emphasis on ensuring smooth people flow. Prior to being able to do that, it is relevant to understand and model people flows better.

People flow planning is necessary in many instances, such as when designing buildings. The key objective of people flow planning is to define a set of transportation devices and their corresponding parameters, such as speed, capacity and location so that the people flow throughout the building is smooth and efficient. In practice, this means designing transportation services that result in feasibly short customer waiting times. The planning is often based on simulations of people moving inside the buildings, while the simulation results are profoundly dependent on how the passengers' characteristics and paths have been modeled. In order to achieve realistic simulation results, a reliable model to describe how passengers find their way from start to destination is needed.

Passenger pathfinding can be described with so-called hot spot networks. These hot spots are special locations in the buildings that serve some functionality to either some or all of the passengers. A typical example is the entrance to a metro station, through which all passengers must pass. Instead, subsequent hot spot usage depends on the widely varying characteristics and agenda of the passengers, and therefore substantial path variations can be expected especially in complex buildings.

This student project is conducted for the purposes of the course 'MS-E2177 Seminar on Case Studies in Operations Research' at Aalto University in the spring of 2016. The project client is KONE, one of the largest elevator and escalator manufacturers worldwide, that supported project execution throughout its lifecycle. In this study, we focus on tactical level pathfinding in order to facilitate improvements in people flow simulation models.

## 1.1 Objectives

Inspired and motivated by the real challenges that KONE faces, this project has two objectives:

- 1) Define and characterize the typical user groups and hot spots of different types of buildings

- 2) Apply the findings of the first objective in developing a method which generates the hot spot network and finds the shortest paths for each typical user group, given a building with a set of hot spots

The objectives were carefully taken into account when designing the study, as further described in Chapter 3, and both objectives and research design were formed in close collaboration with KONE.

## **1.2 Scope of the study**

The scope of this study is limited to reaching the two objectives. Literature review is conducted to the extent that enables understanding how this study relates to a broader stream of people flow research and that supports gaining insight on how literature proposes approaching this type of research. On the other hand, data collection was agreed on to account for a total of five buildings that belonged to four different types of buildings, reflecting the fact that data was needed for getting insight on different user groups, their rough shares and used hot spots instead of accurate quantitative data on people flows. Data was also limited to account for only a certain, popular part of the buildings – typically lobby – as modeling entire buildings would have been infeasible within the scope of this study. Certain behavioral and physical characteristics, such as walking speed and physical impairments, were also taken into account in both data collection and model implementation.

The scope of the model was further limited to match the real needs of KONE. For instance, it was not necessary to take the temporal perspective into account. Typically in people flow simulation, people are assumed to arrive based on a Poisson process, while that provided no additional value in this project. Additionally, scope was limited to the tactical level of pathfinding, which does not consider interaction or collision-avoidance with other people, among other things.

## **1.3 Structure of the report**

Following this introduction, the second chapter describes and discusses literature on this subject, positions this study in a larger research context and establishes a baseline for the followed approach. The third chapter, instead, describes our research design, covering data and methodology.

Chapter four covers the model implementation, describing its design, parameters and assumptions, as well as presenting a general overview of the model's logic and additional functionalities. The fifth chapter provides examples of the model's results,

while chapter six looks at model validation. Finally, chapter seven concludes the report.

## **2 Literature review**

In this chapter, we first introduce key aspects relating to people flow simulation. Then we discuss the principle of least effort, a key assumption in this study. The third chapter motivates using graphs for the purposes of modeling hot spot networks. Finally, the fourth chapter describes Dijkstra's algorithm that is used for solving the shortest path problem.

### **2.1 Research on people flow simulation**

People flows have been studied from widely varying perspectives, but despite having been subject to research, simulating the motion of realistic, large, dense crowds of autonomous agents is still a challenge (Pelechano et al., 2007). One of the key issues that complicates simulation is comprehending and modeling real behavior; Guy et al. (2010) argue that realistic simulation of crowds involves components such as group behavior, cognitive modeling, motion synthesis, and crowd movement and rendering. Therefore, both mathematical methods and qualitative methods emphasizing behavioral characteristics are necessary and exist in literature. The subject has been extensively studied in, for instance, social sciences, traffic engineering, architecture, urban planning and robotics (Guy et al., 2010).

Many researchers have studied people flows at the operative level (Pelechano et al., 2007; Pettré et al., 2009), as the development of collision-preventing and related models reflects. An interaction always occurs between two humans when they walk with converging trajectories, and therefore they need to adapt their motion to avoid and cross one another in a respectful distance (Pettré et al., 2009). On the other hand, tactical level pathfinding, the focus of this study, has not been studied that much. Such people flow simulation has often been formulated as a problem of minimizing certain metric for a group of independent agents (Guy et al., 2010), and the most common metrics are related to distance and time. For decades, a leading approach for minimizing the path from a place to another, i.e. from a node to another, has been Dijkstra's (1959) algorithm.

In people flow simulation, individuals are typically called agents. Agent-based modeling (ABM), instead, refers to modeling independent agents and studying their effects on a larger system. Thus, it is an alternative for approaching people flow modeling instead of directly referring to people flow modeling. ABM has four main areas of application: flow simulation, organizational simulation, market simulation, and diffusion simulation (Bonabeau, 2002). Regarding this study, especially flow

simulation application areas, such as evacuation, traffic and customer flow management, are relevant. However, ABM emphasizes interaction between agents and immerses into operative considerations.

Additional to implementing the mathematical model, observing real people flow and identifying key characteristics and behaviors is of significant relevance in people flow simulation. It can be argued that one of the key challenges in modeling crowds is to develop rules that guide how agents interact with each other in a way that faithfully reproduces paths and behaviors commonly seen in real human crowds (Guy et al., 2012). People have widely varying agendas that guide their movement, additional to which they need to utilize and adapt to the opportunities enabled or disabled by both available network of hot spots and the individual's characteristics; Pettré et al. (2009) argue that human locomotion is generally driven by a goal to reach, while it is constrained by physical, biomechanical and environmental factors. In the context of this study, modeling the flow realistically requires careful field observation and gaining insight on the peculiarities of different user groups.

## **2.2 The principle of least effort**

When people move around, they typically want to minimize the effort taken, either consciously or unconsciously. Human motion and crowd dynamics are governed by the principle of least effort (PLE) (Guy et al., 2010). PLE is a broad theory, which suggests that, when trying to reach their goals, people naturally choose the path that provides least perceived effort (Guy et al., 2012). The effort can be optimized with regard to different variables and even total amount of used metabolic energy has been operationalized (Guy et al., 2010). However, variables that are easier to conceptualize and measure fit better in this study. For instance, time and distance are feasible ones.

PLE is a highly relevant concept for this study as the model's implementation is profoundly based on the assumption of people wanting to minimize their travelling time. In practice, we apply Dijkstra's algorithm for finding the path with optimal travelling time and naturally take into account relevant constraints.

## **2.3 Hot spot network**

The hot spots of a certain building form a network that can be presented as a graph consisting of nodes and branches. The nodes represent hot spots, while branches can be, for example, distance or time. Not all of the nodes need to be directly connected to each other and in many occasions some nodes need to be passed through in order to move from a starting node to desired destination. It can also happen that from a certain node some other node cannot be accessed. Researchers have studied graphs

in depth but within the scope of this study it is sufficient to note that they are an appropriate way for approaching this kind of problem setting.

The collected data includes the focal buildings' layouts and possible paths in the observation area that intrinsically indicate whether nodes are connected to each other or not, and how distant the nodes are from each other. Here, the graphs are presented as matrices due to numerous reasons: matrices are practical and easy to interpret, they fit well to the algorithms proposed in literature (such as Dijkstra), and they are efficient for both data storing and computing in this context.

## **2.4 Dijkstra's algorithm**

Dijkstra (1959) developed a highly used and adapted algorithm for finding the shortest paths in graphs. He assumed  $n$  nodes, of which some or all pairs are connected by a branch whose length is known, so that at least one path must exist between any two nodes. He formulated two problems, to which he sought answers with the algorithm: (1) construct the tree of minimum total length between the  $n$  nodes and (2) find the path of minimum total length between two given nodes.

The algorithm addresses the first problem through dividing the branches into three sets and then performing certain operations in certain order for the sets. The sets of branches are defined as:

- I. The branches definitely assigned to the tree under construction (forming a subtree)
- II. The branches from which the next branch to be added to set I will be selected and
- III. The remaining branches that are either rejected or not yet considered.

The nodes left in the set III are further divided into:

- A. Nodes that are connected by branches of set I
- B. The remaining nodes to each of which one and only one branch of set II will lead

Instead, the operations for solving the first problem start by choosing an arbitrary node as the only member of set A, and then placing all branches that end in this node in set II. Then, two steps are repeated until sets II and B are empty. In the first step, the shortest branch of set II is removed from the set and added to set I, and therefore the focal node is transferred from set B to set A. In the second step, instead, the branches leading from the transferred node to the nodes that are still in set B are considered. If the branch under consideration is longer than the corresponding

branch in set II, it is rejected. If it is shorter, it replaces the corresponding branch in set II, and the latter is rejected. (Dijkstra, 1959)

The second problem, finding the path of minimum length between nodes P and Q, is solved through a process consisting of two steps. A necessary notion is that if node R is on the shortest path from P to Q, also the shortest path from P to R is known. For implementing the algorithm, nodes are divided into three sets:

- A. The nodes for which the minimum length path from P is known
- B. The nodes from which the next node will be selected to be added to set A
- C. The remaining nodes

The branches, instead, are divided into three sets as follows:

- I. The branches occurring in the minimal paths from node P to the nodes in set A
- II. The branches from which the next branch will be selected to be placed in set I
- III. The remaining branches

Next, the following two steps can be iteratively implemented to find the shortest path. In the first step, all branches  $r$  connecting the node transferred to set A with nodes R in sets B or C are considered. If R belongs to B, it must be studied if branch  $r$  gives rise to a shorter path from P to R than the known path that uses the corresponding branch in set II. If not true, branch  $r$  is rejected, while if true,  $r$  replaces the corresponding branch in set II. In case of R belonging to set C, it must be added to set B while branch  $r$  is added to set II. In the second step, instead, the node with minimum distance from P is transferred from set B to set A, and the corresponding branch is transferred from set II to set I. Then the two steps are iterated until node Q is transferred to set A. (Dijkstra, 1959)

There are some limitations for Dijkstra's algorithm but still it is readily applicable for this study. For instance, the original algorithm applies only in situations where all arc weights are nonnegative (Johnson, 1973), and such issues have inspired researchers to propose varying additions or changes for the algorithm. Dijkstra's algorithm fits especially well for the purposes of this study as the paths can be presented in the form of graphs, for which the algorithm is indeed feasible. As we model the weights between nodes to be spatial or temporal distances, it is also reasonable to assume that such weights are all nonnegative.



### 3 Research design

This study is designed in close collaboration with the project client in a careful way that supports achieving the objectives of the study. In this chapter, we first present the data collection procedures and overview of the data. Secondly, we describe how we approached designing and implementing the model. The model itself is presented in Chapter 4.

#### 3.1 Data

We collected necessary data from five different locations that belonged to four different types of buildings, as presented in Table 1. The data collection was guided by comprehensive forms developed by KONE that were slightly modified according to our experiences after the first data collection session. Observations were conducted in spots that did not distract people flow, did not require special access and had intense people flow. Therefore, the observations represent people flow in certain parts of the buildings. In four of the buildings the observation spot was in the lobby, while in Helsinki central railway station we studied the Compass floor, from which people access both the metro platform and the upper floors. All three project team members participated in each observation session, which enabled better identification and discussion of prevailing peculiarities.

The primary data includes characteristics of different user groups, their relative shares, the building's hot spot network and typical hot spots for each user group. We observed both physical and behavioral characteristics open-mindedly. Emphasis was on walking speed, space demand, ability to use certain transportation devices and hot spots, possession of access cards, preferences relating to transportation devices, and other emerging behavioral features. In addition to those, we evaluated very interesting and valuable but secondary aspects regarding the objectives of this study. They were accessibility and services, navigation and functionality of the space, information graphics, and interior of the space.

Table 1 - Data collection locations

Type of building	Building name	Duration
Metro station	Helsinki central railway station	120min
Residential building	Cirrus	<120min
Hotel	Scandic Simonkenttä	120min
Office	Microsoft	120min
	Keilasatama 5	120min

## **Data review**

As all of the buildings were very different, the data contains significant variance. In order to verify data quality, we reviewed the data by ourselves and sent it to KONE for further review. The reviews confirmed that the collected data was of sufficient quality and that the accuracy of observations was reasonable. Therefore, they form a feasible basis for developing the model.

In order to visualize and better comprehend the data, we drew layouts of the spaces. In the layouts we specified all of the hot spots and typical routes followed by identified user groups. An example of them is presented in Appendix 1, the layout and the hot spot network in Keilasatama 5 office building.

Next, we present the main observations made in each of the buildings. However, it is necessary to keep in mind that the observations in total are far more comprehensive and detailed than what is described here. The purpose of reviewing the data here is to provide the reader with some key insights of the people flow instead of describing the entire data set.

## **Metro station**

Our first data collection location was the metro station at Helsinki central railway station. The observation was done during morning traffic between 7 and 9 a.m. The people flow was quite continuous and constant, and most people had clear destinations, didn't linger and generally moved rather fast. We identified eight characteristically different user groups: commuters, students, elderly people, tourists, baby trolleys, people meeting, staff, and non-tourist travelers. Of these groups, commuters accounted for approximately 65 % and students around 15 % of the total users. The most relevant hot spots were stairs and escalators from upper levels to the Compass floor, ticket machines and ticket readers, as well as the escalators to the platform.

The most pertinent discoveries were on the routes and group sizes of the users. Commuters and younger people preferred whatever was the fastest way and didn't require the use of the ticket machines and ticket readers. They mostly traveled alone or in pairs. Elderly people, however, were slower and preferred escalators to stairs. They were also more prone to use ticket machines and readers. Tourists typically traveled in groups and usually converged at the ticket machines.

## **Residential building**

In the second case, the residential building, our observations were straightforward. This was mostly due to the low number of hot spots and small people flow, resulting in very few different routes and only some distinguishing characteristics. Residents simply came through the entrance and entered the elevator to access the apartment

floors, or arrived via elevator and entered the multi-purpose room for laundry. We distinguished three different groups of residents based on their physical characteristics but their routes practically did not vary. Instead, the fourth user group, visitors, had some differences as they did not have access to the building and had to wait for someone to open the door remotely or from inside.

### **Hotel**

The third building, Scandic Simonkenttä, had great variance in its people flow intensity even during our two-hour observation period. One of the prevailing peculiarities was that most people arrived in groups of at least two persons. We visited there during daytime that was clearly reflected by the shares of the user groups. We were able to identify five user groups: check-in visitors, business visitors, check-out visitors, hotel staff and existing customers. During our observation period, check-in visitors was the largest user group, while we assume that during the mornings check-out visitors would be the largest one. The most important hot spots were entrance, reception, sofas and elevators, in addition to which different user groups had varying desires as for example business visitors frequently used conference rooms and toilets.

### **Office buildings**

Our visit to the fourth building, Microsoft, occurred during afternoon, after the most usual lunch time. The building had three evident user groups: office workers, visitors and staff. Office workers were clearly the largest one, accounting for roughly 80% of all the users. They also had widely varying paths but we were not able to find typical characteristics for the office workers that ended up using certain hot spots. Therefore, they can be further divided into quite a few subgroups. Visitors naturally did not possess access keys, unlike other users, and they had to visit reception first and wait for hosts to pick them up. Visitors rarely moved in vertical direction as conference rooms, restaurants and many other hot spots are located on the first floor. The most relevant hot spots were entrance, reception, waiting area, coffee machine, restaurants and elevators.

The last one of the five buildings, Keilasatama 5, was visited during ordinary lunch time, which was clearly seen in the dominant role of people visiting the restaurant. We identified four meaningful user groups: office workers, visitors, visitors with luggage and lunch visitors. Office workers accounted for approximately 80% of the users. Unlike in the Microsoft building, visitors with luggage were quite frequent and their large space demand and slow moving speed clearly distinguished them from ordinary visitors. Also lunch visitors were an evident user group as the restaurant attracted people from a nearby construction site. The most relevant hot spots were restaurant, elevators, stairs, entrance and reception. If compared to Microsoft, people

in Keilasatama 5 used more stairs while sofas in the waiting area were clearly less frequently used.

### **3.2 Methodology**

Beyond data collection, the methodology for model development takes into account desires of the project client, aspects proposed in literature and the capabilities of the project team. The methodology was preliminarily designed early in the project but naturally some elaborations related to especially model specification became necessary also later. The actual modeling phase was not started until data had been collected and reviewed, providing solid baseline to lean on.

Eventually, the methodology is straightforward. Literature was reviewed early on in the project to establish a feasible theoretical framework, against which other steps could be reflected. Literature inspired our modeling and supported the approach that was preliminarily proposed by the client. Instead, the data collection procedures as well as data review supported gaining necessary insight on the buildings and user groups that was necessary for being able to model realistically.

As the data had been reviewed and validated, model specifications were discussed in depth with the client. The main function of the specifications was to lay out the state of mind, around which the project team was allowed to maneuver with specific code. The client did not want to restrict our thinking with too many proposals and therefore the model-related discussions were much about us presenting ideas that were collaboratively refined. Thus, the capabilities of the project team members played a significant role in developing the model. It was jointly decided that modeling occurs with Matlab as it fitted well to the capabilities and accessibility of both client and project team.

Throughout the project and increasingly during modeling, the project team had regular meetings, in which the actions taken by each member were reflected and next steps were designed. The model was reviewed several times and alternative approaches for structuring it were discussed. Several, very useful conversations were also conducted with the client to provide progress updates as well as discuss and validate future directions.

After the model had been built and configured based on the collected default data, we created a special data set for validating the model. This validation process reflected a future user scenario and aimed at reviewing all the model's features and functionalities.

## 4 Model implementation

Due to the client's request, we will not present the specifics of our model code here, but rather present the general idea behind our model, along with the model parameters and design choices that were made during the early parts of the modeling process.

The model was built on the basis of the information obtained through the various site visits during the early parts of the project. Through discussion with the client representatives, we agreed on what the model should be able to achieve, and which parameters were necessary for building a feasible model. Naturally, it was necessary to obtain information regarding the hot spot networks and user groups themselves before moving on to the actual model implementation.

As the focus of the model is on hot spot networks and user groups, which are considered as time-independent in our model, it was not in the scope of our project to take into account the simulation process –related factors, e.g. the time dependency of arrival typically modelled as a Poisson process. Our goal was to solve the pathfinding-related problems that can be considered static: the hot spot network or group preferences are not changing over time. Naturally, the results should still be applicable in a simulation-based, dynamic setting.

### 4.1 Model design and parameters

The main objective behind the model was to be able to compute a shortest path along the hot spot network for each of its user groups. With such knowledge, it would then be possible to see how the different hot spots are being used in the network, among other things. This would, in turn, allow us to see if some of the hot spots are prone to bottlenecks due to high use percentages, and likewise if there are any hot spots that are quite infrequently used.

The problem to be solved is not, in general, a simple shortest-path problem between two hot spots in the network. This is because we want to be able to model the fact that not all user groups simply take the shortest path between their start and end points: they may have additional preferences or restrictions that affect the paths taken in the general hot spot network.

The following parameters were included in our final model implementation. With these parameters, our client can extend from our default settings to new ones by simply defining the new parameters and applying these to the model.

- The **hot spot network**, given in graph form as a matrix of distances between the hot spot nodes.
- **Access** and **Mobility** requirement parameters, given separately for each hot spot in the hot spot network

- The **user groups** of this specific hot spot network, and the characteristics of each of these user groups:
  - **Access** and **Mobility** ability parameters, given separately for each user group in the hot spot network
  - **Speed** parameter for the user group that is used for scaling the distances of the general hot spot network.
  - The **preferences** of the user group with respect to the hot spots in the network.
- The **percentages** or relative proportions of each group, where the total sums to 1 (or 100%)

We created an Excel sheet that can be used to read the default parameters into Matlab. Each building type has its own sheet, and new buildings can be added simply by following the formula we have specified to the client.

Each hot spot and user group is also given a label, reflecting the actual name or purpose of these hot spots and users. The preferences for each user group then refer to these labels. These are more accessible to the model user than simple indices and numbers, even though our model does convert this information into numeric form before applying it. All of the aforementioned parameters are the ones that we could assess during our data collection process.

It is important to note that one of the key model design factors is that each building type has its own user groups. This was deemed necessary by the project group, as we observed that the typical user groups of different building types cannot be reasonably clumped into any higher-level user groups. This is mainly because the purpose of the building largely dictates what type of users are moving through the building.

In the model, each group has a starting and ending hot spot that do not vary inside the user group. This means that the starting and ending hot spots are set beforehand for the user groups, although they can be changed by altering the Excel inputs. The fact that the starting and ending hot spots did not vary inside a user group may seem restrictive, but it simply means that we have to model different directions of similar movement by different groups. This is connected to our idea of subgroups that is discussed subsequently in this chapter.

In addition to these starting and ending points, the users may have some additional preferences along their path between them. A key design choice made with the agreement of the client that such preferences are modeled as **strict, binary-type preferences**. In other words, the user group's preferences are actually part of the objective of the user group in our model.

Using this modeling approach means that there is no possibility of variation inside the user group itself: since all preferences are modelled as binary, all users inside one user group must pass through the exact same hot spots. Variation in this model

setting is achieved not by setting probabilities for separate routes, but by splitting the user groups further into smaller **subgroups**. These subgroups should be defined so that the desired level of differentiation and separation in user groups and their routes is achieved. Also, one must remember to change the appropriate user percentiles so that sum of percentages sums to one.

Although this approach means that the number of groups and the associated subgroups is likely to be somewhat large in order to realistically achieve the different routing possibilities, we feel that this is not a substantial problem in our model since the addition of a new group and their preferences is made simple in our default Excel sheet.

Other physical and behavioral characteristics that are taken into account in our model are connected to the access and mobility requirements that some hot spots may have. Two additional parameters have been set to reflect these important characteristics of hot spots. These are called the **Access** and **Mobility** requirement parameters of each hot spot. For each user group, the corresponding Access and Mobility ability parameters have been set to denote the capabilities of the user groups with respect to these requirements.

With the Access parameters, the purpose is to highlight that not all hot spots in some of the networks are accessible to all users: a good example would be a locked door which requires some special access card or key. This restricts the usability of the hot spot to those users who do not own the necessary access capabilities. To those user groups who have the required access capabilities, these hot spots are just like any other hot spots in the network, but for those without access, these hot spots are not easy to pass through.

With the Mobility parameters, we take into account that some hot spots are quite inaccessible to those with movement impairments or other restrictions, e.g. people with strollers or wheelchairs. Some examples would be traditional stairs or revolving doors, which can be very difficult to traverse for some user groups with restricted movement. To those user groups that have no special movement impairments, the Mobility hot spots are just like others, but for other user groups these hot spots are treated differently.

Similar to the user group preferences in our model, both the Access and Mobility parameters are binary: either the group has the special access card or not, and the group has movement restrictions or not. In our simplistic modeling approach, each hot spot simply requires the access and mobility capabilities, or it does not. Likewise, each different user group either has these capabilities or not.

In the model, we actually check the correspondence of each user group's capabilities and the requirements of each hot spot. For hot spots with no special requirements, no adjustments to the user group's specific hot spot network need to be done. However, if the network contains some hot spots that require capabilities that the

user group does not have, then something must be done to reflect this fact. In our approach, we have simply chosen to scale the distances corresponding to these hot spots by a large number.

## 4.2 Model assumptions

There are several assumptions made in our model that simplify the actual, real-life strategic pathfinding in people flow. Some of these assumptions have been explicitly defined in the model design, while others implicitly follow from model formulation. We present the most impactful assumptions here, along with our reasoning and explanations behind these.

1. Every user group knows exactly the distances between all of the hot spots in the network. That is, each user has **perfect information** regarding the layout of the building, and they use this information when choosing their path.

This is quite a strong but necessary assumption to make. For example, consider a user group that is somehow ‘foreign’ to the examined hot spot network. This could be a tourist user in a metro station or a first-time hotel visitor, for example. In reality, these users might not have very accurate knowledge of the building layout, which means that more often than not they do not actually take the path that is shortest for them. Modeling this imperfect information would require more extensive knowledge of how people make their routing choices in previously unvisited places. We simply assume that people can somewhat accurately estimate the distances between hot spots when choosing their routes, even if they are not previously familiar with the building.

2. Every user (and therefore every user group) is **goal-oriented**: the users all have some objectives they want to complete, the simplest example being that the user wants to move from hot spot A to hot spot B.

This assumption simply means that there are no users that wander aimlessly in the hot spot network. It is then quite a similar assumption to the assumption often made in decision-making settings that people behave rationally when making their choices. In our model, each user group has a goal, which may consist of several sub-goals, as discussed previously.

3. The hot spots themselves are considered as **points**, and therefore the hot spot itself does not consume any additional time. In more realistic, time-dependent settings, this is of course not the case: taking the stairs or the elevator between similar starting and ending points generally take different amounts of time even when the spatial distance is similar. These sorts of differences



can of course be taken into account implicitly when setting the distances of the hot spots in the network,

Since our project is about time-independent, tactical level pathfinding, we ignored the fact that different hot spots may have different time consumptions by themselves. Taking these into account would be quite important in simulation-driven settings, where we would most likely want to model the fact that the users may change their behavior if the expected waiting time for some hot spot becomes too large because of bottlenecking, for example.

### 4.3 Model logic

As stated previously, we will not present the model specifics here, but provide a general overview of the inner workings behind it. The core model requires the all of the input parameters specified above, which it then uses to produce the shortest paths for each of the user groups in the hot spot network.

Dijkstra's algorithm is first used to find the shortest distances between each pair of hot spots in the overall hot spot network. This information can then be used to create a **subgraph** that contains only the hot spots or nodes that are critical for the user group. The weights between hot spots in this subgraph are given by the shortest distances between these hot spots that have been previously calculated.

From this group-specific subgraph, we can then solve the shortest path for the user group. We know that the user group must pass through all of the hot spots in this subgraph, but only the start and end hot spots are set: any remaining hot spots in the subgraph must be passed through but the order can be any of the possible permutations. All of these permutations are checked and the one that provides the shortest distance is the route chosen by this user group.

It is important to note that in more general settings checking all of the permutations can be impossible. However, as the number of hot spots in one network is relatively small and the number of critical hot spots for one user groups is often even smaller, checking all of the permutations is computationally feasible and can be completed in a reasonable timeframe.

### 4.4 Additional functionalities of the model

In addition to the core model we have built, we also provide the client with some additional functionalities that should further increase the flexibility of the model when applying it to new building settings.

These functionalities include the ability to add or delete specific hot spots from the network without having to make changes in the Excel-files. One can then check if adding or deleting a hot spot has an effect on the paths that the different user groups

take in the network. If one wants to delete or add several hot spots, it may be more convenient to simply create an entirely new hot spot network by creating a new sheet in the excel file.

## 5 Results

In this chapter we present some examples of the results obtained from the model using the default settings we have created for different building types. Using these examples, we show and explain what our model actually outputs. Even though we only show a couple of simple outputs and resulting graphs to retain clarity, the results of our model are similar for other user groups and more complex building types as well.

We use a simple office setting as an example here. First, the model plots the general hot spot network, which is shown in Figure 1.

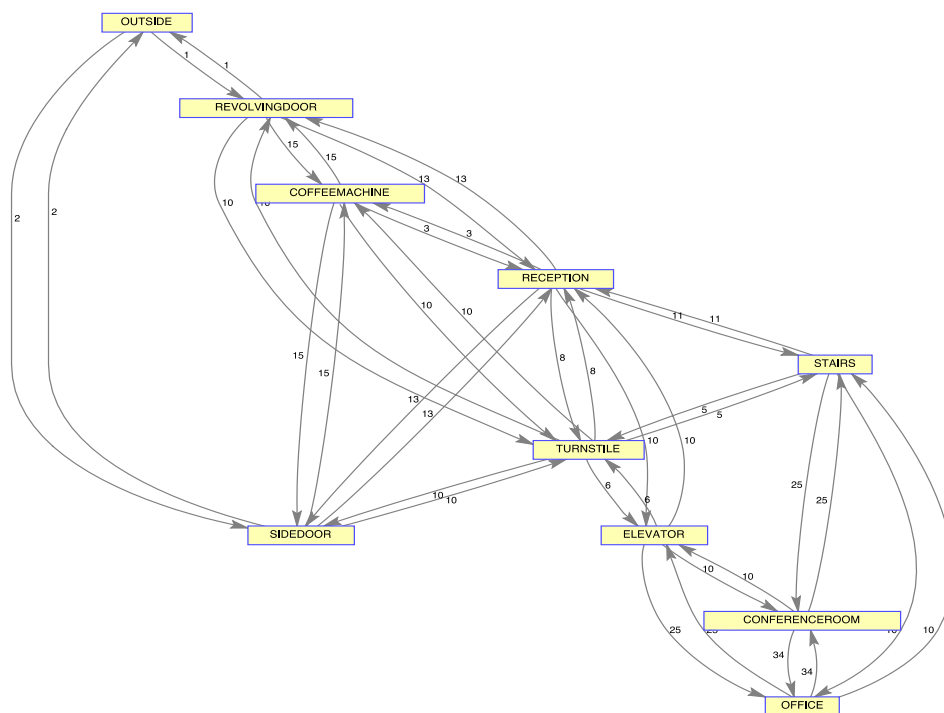


Figure 1 - General hot spot network in office building

This figure shows the general hot spot network, the direct connections between all of the hot spots in the network along with the distances of these connections. Note here that the distances are the same that are given as the distance parameters for the network; the user group- scaling has not been performed yet.

Then, for each user group, the shortest path is highlighted. The group-specific figures also show the weighted distances for the user groups. For example, the route chosen by conference visitors is presented in Figure 2.

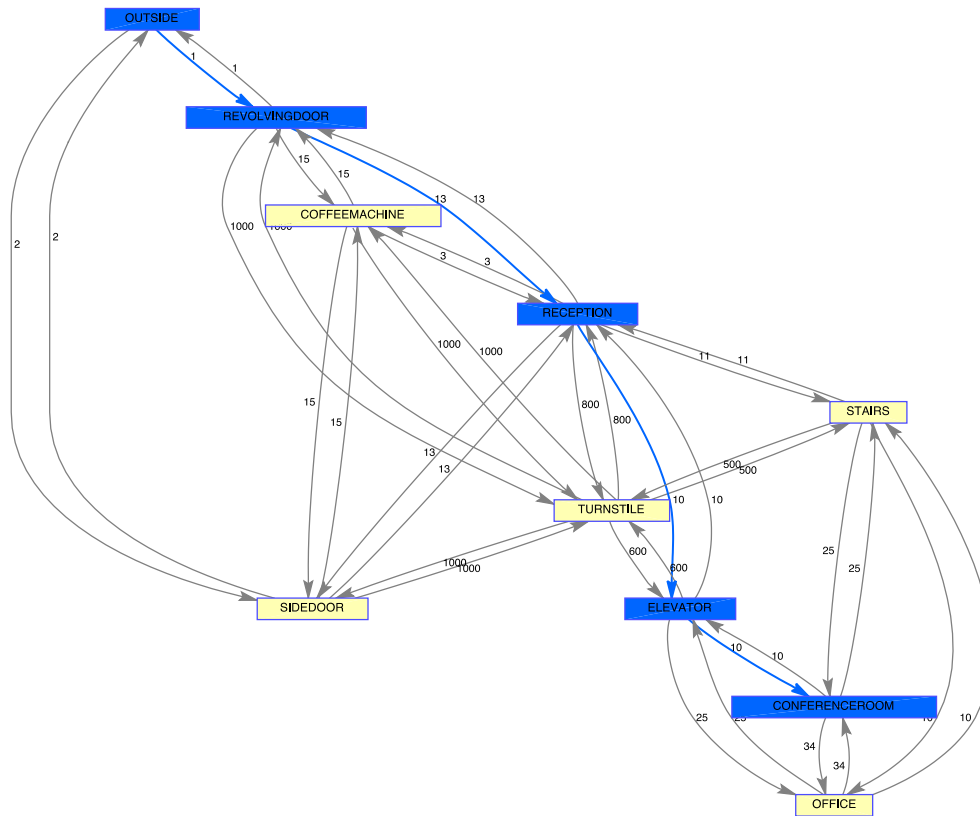


Figure 2 - The route of conference visitors

The Matlab console outputs regarding this user group:

*\*\*\* Processing user group ConferenceVisitor \*\*\**

*The shortest path between start and end nodes already contains all the required nodes.*

The shortest path for this user group is an example of a simple case, where simply taking the shortest path between that start and end hotspots results in the shortest path. This is of course the case when there are no critical hotspots in addition to the start and end hotspots for the user group, as is in the case of this group. Above, we can see that the conference visitors have to travel through the reception as the turnstile requires access that the visitor group does not have. As the figure shows the weighted distances for the specific user group, it is appropriate to note that all paths to and from the turnstile hot spot have been scaled hundredfold, and thus the algorithm will try to find other paths.

The model outputs both the shortest route with respect to the hotspot labels, as well as the user group – scaled distance associated with this path. For this user group, for example, these outputs are:

Shortest distance = 34

Shortest route = OUTSIDE REVOLVINGDOOR RECEPTION ELEVATOR CONFERENCEROOM

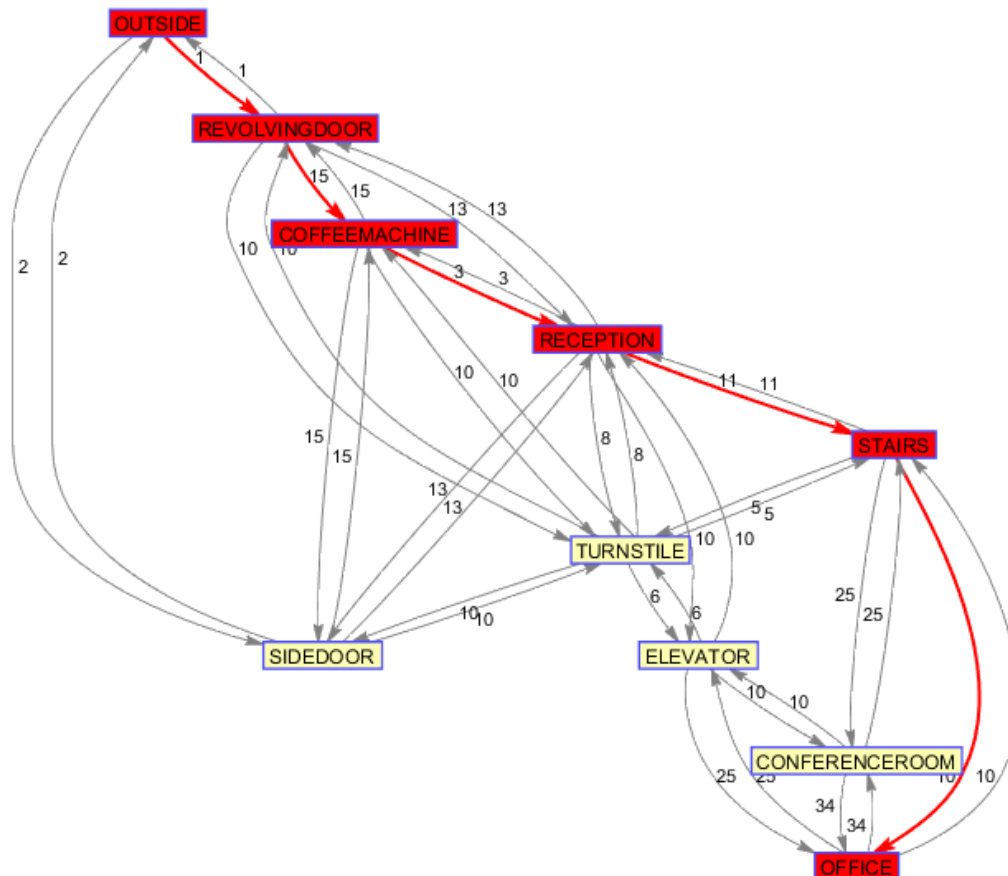


Figure 3 - The route for office workers

Then, for another user group labelled as OfficeWorker, the chosen route is presented in Figure 3. The resulting shortest path and console outputs are:

\*\*\* Processing user group OfficeWorker \*\*\*

The shortest path does not contain all the required nodes, continue execution

The shortest path has been found.

The first console output simply corresponds to the case where a simple shortest-path route between the start and end hotspots does not contain all of the critical nodes, therefore the subgraph – extension is necessary for this user group. The model then plots the aforementioned subgraph for this user group, as presented in Figure 4.

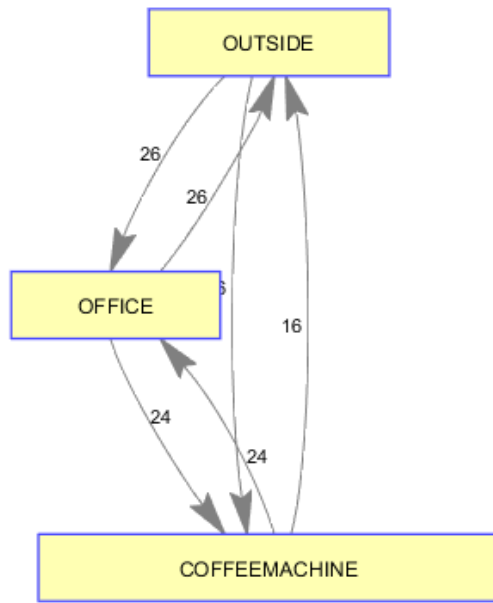


Figure 4 - Subgraph for office workers

As specified in the model formulation, this subgraph contains only the hotspots that are critical for the user group, and the distances between the hotspots are the shortest distances between them. As mentioned before, we separately store the information regarding the paths, to which these distances actually correspond to.

We know from our inputs that the hotspots labelled as OUTSIDE and OFFICE are the starting and ending hotspots, respectively. From this subgraph, the model then computes the permutation from starting to ending hotspot that results in the smallest possible distance. Since in this case there is only one critical middle hotspot, the result is almost trivially simple to calculate: the user group takes the shortest path from OUTSIDE to COFFEEMACHINE, then the shortest path from COFFEEMACHINE to OFFICE. The shortest distance is then known to be  $16+24 = 40$ .

The model then converts this shortest subgraph path to the one it corresponds to in the original user-scaled hot spot network, which is **OUTSIDE** – REVOLVINGDOOR – **COFFEEMACHINE** – RECEPTION – STAIRS – **OFFICE**. Thus, the model outputs for this user group are:

Shortest distance = 40

Shortest route = OUTSIDE REVOLVINGDOOR COFFEEMACHINE RECEPTION STAIRS OFFICE

The results indicate that user groups are in fact identifiable and separable with respect to their typical paths and several other physical and behavioral characteristics although these user groups vary significantly in different types of buildings. The resulting routes for different user groups match our data collection findings quite well, even though we have had to make some crude estimations regarding the distances between hotspots.

In the model, a higher-level script simultaneously computes the shortest distances and routes for each user group in the hot spot network. In this higher-level script, we also store the loadings of each hotspot in the hotspot network. The total loadings of the hotspots in the network are the following:

**Table 2 - Total loadings of the hot spots**

Hotspot ID	Load
OUTSIDE	1
REVOLVINGDOOR	0.9
SIDEDOOR	0.1
TURNSTILE	0.22
RECEPTION	0.78
ELEVATOR	0.55
STAIRS	0.45
OFFICE	0.47
CONFERENCE ROOM	0.53
COFFEEMACHINE	0.45

These loadings can then be explored further as much as is deemed necessary. Here, we make some general observations. In more complex networks, studying these loadings may prove to be even more beneficial, and provide valuable insight into which hotspot capacities receive the most stress in the hotspot network.

The hotspots labelled as OUTSIDE, REVOLVINGDOOR and RECEPTION all experience high usage percentages. That is especially true for OUTSIDE, which is actually used by all of the user groups in this network. This is natural since we have set all of the user groups to start from this hotspot, and the revolving door follows exactly after that. RECEPTION is frequently used due to the fact that it is used by all of the different visitor-type groups, but also by office workers due to its convenient location.

On the other extreme, the hotspot SIDEDOOR is only used by 10% of the users, which actually corresponds to two small user groups that have difficulty accessing the REVOLVINGDOOR hotspot due to mobility issues, therefore choosing to use the SIDEDOOR as an entrance to the building. Another quite infrequently used hotspot is the TURNSTILE hotspot, which requires special access granted only to the ones who work in the building. Thus, visitor groups tend to avoid using these. All of the other hotspots in turn seem to be quite evenly used with loadings of around 50% of total users in the hot spot network.

## **6 Model validation**

As our default settings and parameters directly reflect our own findings obtained from five buildings, it is natural that the model provides results that are consistent with these same findings. As a part of the entire modeling process, additional model validation was conducted to ensure that the model can also be generalized into new settings. The client will not only use the default settings created by the project team, but also wants to create new scenarios and buildings, where the model should ideally be just as applicable as it was with our default settings.

In addition to the extensive testing carried out using our collected default data, we built a special data set for model validation. This validation data set was not based on the gathered data from any of the visited locations. Instead, its purpose was to find any aspects of the model that we might have overlooked when configuring it with the default data. The data set, albeit following the guidelines and requirements set by the model, was constructed irrespective of the observations our model was built upon. Thus, it was to simulate as if a future user from our client was utilizing our model.

The validation was performed by defining a hot spot network with limited connections, the user groups and the corresponding requirements and abilities. Special emphasis was also put on not building a data set that would be optimal and easy for the model, but rather a general input. The data set was also either built or modified so that every feature of our model could be tested and validated with it.

After defining the validation data set and running the model with it, we found the results to be consistent with our expectations. The model performed satisfactorily, providing the correct shortest paths while conforming to the requirements and abilities. Furthermore, the additional functionalities of the model were tested with both the default data and the validation data set, and were found to perform correctly. All in all, no deficiencies were found during model validation.

## **7 Conclusions**

This study had two objectives, of which both were fulfilled. We successfully defined and characterized the typical user groups and hot spots in several types of buildings, as well as developed a tactical level pathfinding model that reflects key literature, collected data and desires of the client. The model takes into account a given hot spot network, access and mobility requirements, different user groups and their key characteristics, and relative proportions of the user groups. Based on these pieces of information the model forms the overall hot spot graph as well as subgraphs and finds the shortest paths for each of the user groups. The model's implementation occurred in Matlab and Excel that enable easy modifications in the assumptions, which was considered very necessary as the buildings can eventually vary significantly.

### **Limitations of the study**

Despite meeting the set objectives, there are limitations that need to be taken into account. First, it is evident that buildings can differ significantly even though they would serve the same purpose. Our data collection included only five buildings and four different types of buildings, which is why the data cannot be considered generalizable. Eventually, it can be argued that conducting many more observation sessions would still not make the data well generalizable. Instead, we focused on getting insight on the emerging themes that would enable creating a default baseline as well as developing a model, in which those assumptions could be further modified according to contextual features.

Second, there is no guarantee that the model we developed is the best possible one. Instead, it is one solution for a problem that could be solved with widely varying approaches. The project team and the client validated its appropriateness through several reviews and discussions but still some other approaches may eventually prove out to be more efficient.

### **Avenues for further research**

During the project, we came up with a couple of avenues for further research. First of all, a literature review study combining and comparing existing research would be beneficial. It would be especially interesting to see the differences in research and modeling approaches between strategic, tactical and operative levels of pathfinding from a large sample. Second, we would like to see effort for integrating the three levels of pathfinding towards an extensive simulation framework. Third, as the agents are eventually ordinary people, we would consider it very interesting to see if



there are some especially realistic models for modeling tactical level decision making when finding paths within buildings.

### **Further considerations for applying the model**

When applied to a more dynamic setting involving people flow simulation, the model should be extended so that the users in the network take into account the possible waiting times that are associated with their chosen paths. This involves considering the waiting times involved in using the hotspots themselves, an issue we purposely ignored in our model due to its static nature, as stated in our model assumptions. A more flexible model would also consider how long it takes for the user for similar routes when using different routing devices, e.g. traditional stairs versus an escalator.

Perhaps even more importantly, the dynamic model extension should definitely take into account not only the waiting times associated with the hotspots themselves, but the waiting times associated with possibly high loadings of the hotspots. As computed by our model, a high loading suggests that the associated hotspot is likely to be quite crowded. This will likely slow down the movement along the crowded hotspots, effectively increasing the waiting times and times spent in the system for those user groups that frequently use these hotspots. In such situations, users that are trying to find the shortest path with respect to time spent in the system should be able to choose a different route along the hot spot network, if doing so lowers their time spent.

In more dynamic settings, it may be difficult to obtain a feasible model without moving back to lower-level route optimization that would involve minimizing the “effort”, or some other feasible optimality parameter, on an individual user’s level. However, we do believe that such a model could definitely be extended on top of the static model we have built here, even though this was not included in the scope of this project.

One possibility of an extended model could still include the user groups, hot spot network, requirements, preferences and abilities as parameterized in our static model. A key difference to our current static model would be to allow, for example, several possible routes for each user group in the hot spot network. Then, when computing the shortest distances for the user entering the hot spot network, we would take into account the effect of the users already present in the hotspot network when the user in question is optimizing their path. This would mean that the loadings of each hotspot should be dynamic in the network, continuously changing when the users move along the hot spot network. The hotspots that already have high loadings at the point of entry for the new user would then have increased “distances” or waiting times corresponding to the capacity and loading factors of the

hotspots in question. Naturally, this would require adding some additional capacity-related parameters to the hotspots, although these parameters would likely be static in nature. This dynamic loading setting could also apply the space demand parameter observations that we gathered during the data collection process but that were ultimately not used in our static model.

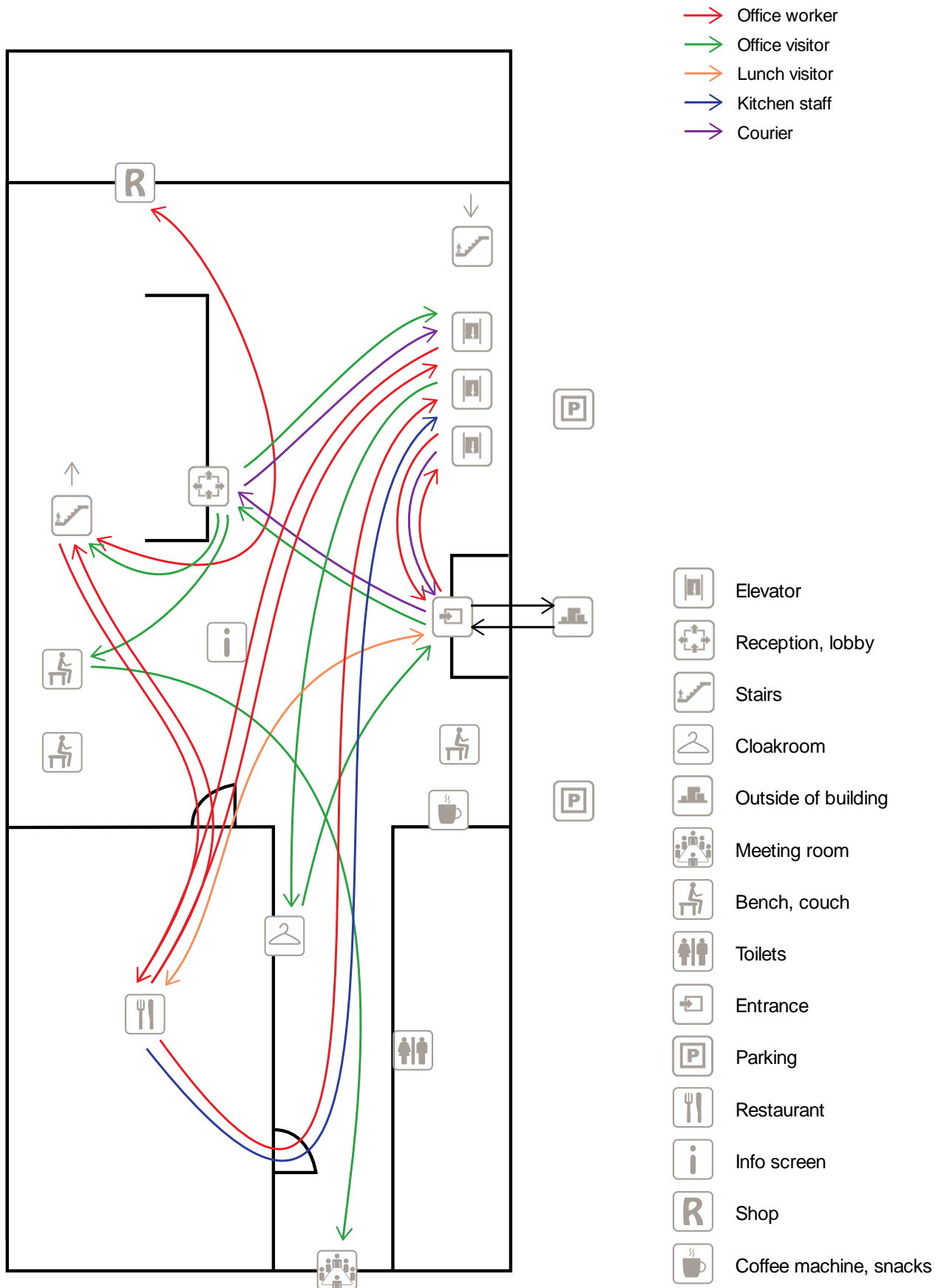
When the distances corresponding to highly crowded hotspots are properly scaled, this could effectively force some of the users to change some parts of their route to less-crowded hotspots. This could also have the effect of balancing the overall loadings of the hotspots, especially when compared to what is obtained by our static model. This approach would also likely be applicable to our current model thinking of strict preferences: The model could still use the same strict preferences, but change the distances to hotspots corresponding to their aforementioned capacity and loading factors.

Another factor related to a more dynamically applicable model formulation could involve the relaxation of the strict preference constraints for the user groups. This would mean, for example, that the user groups still use the hotspot network when their preferences are not exactly met. These preferences could be somehow incorporated into a value function of the user, where the user's objective is to minimize effort spent, but the value of this objective function would also become "better" when the user's preferences are fulfilled. Then, the user's routing problem becomes a trade-off between minimizing their time spent, e.g. when moving from the building entrance to their working space, and meeting their preferences along their path, e.g. getting a cup of coffee from the coffee machine.

## References

- Bonabeau, E. (2002), "Agent-based modeling: Methods and techniques for simulating human systems", *Proceedings of the National Academy of Sciences*, Vol. 99 No. suppl 3, pp. 7280–7287.
- Dijkstra, E.W. (1959), "A note on two problems in connexion with graphs", *Numerische Mathematik*, Vol. 1 No. 1, pp. 269–271.
- Guy, S.J., Chhugani, J., Curtis, S., Dubey, P., Lin, M. and Manocha, D. (2010), "Pedestrians: a least-effort approach to crowd simulation", *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, Eurographics Association, pp. 119–128.
- Guy, S.J., Curtis, S., Lin, M.C. and Manocha, D. (2012), "Least-effort trajectories lead to emergent crowd behaviors", *Physical Review E*, Vol. 85 No. 1, pp. 1-7.
- Johnson, D.B. (1973), "A note on Dijkstra's shortest path algorithm", *Journal of the ACM (JACM)*, Vol. 20 No. 3, pp. 385–388.
- Pelechano, N., Allbeck, J.M. and Badler, N.I. (2007), "Controlling individual agents in high-density crowd simulation", *Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, Eurographics Association, pp. 99–108.
- Pettré, J., Ondřej, J., Olivier, A.-H., Cretual, A. and Donikian, S. (2009), "Experiment-based modeling, simulation and validation of interactions between virtual walkers", *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ACM, pp. 189–198.

# Appendix 1: Sample layout and hot spot network



## **Appendix 2: Self-evaluation**

This self-evaluation aims at summarizing the execution of the project and highlighting both successful and unsuccessful aspects. Following aspects are particularly evaluated: general project execution, division of responsibilities, schedule and scope, risk management and improvement potential.

### **Project summary**

This study project had two objectives, both of which were met, indicating successful project delivery. The project yielded concrete outcomes for both of the objectives that KONE can, and hopefully will, further leverage in its future operations. Both the project team and KONE were satisfied with the outcomes, which was especially delightful for the project team.

Interaction was frequent throughout the project both within the project team and between project team and KONE. Also other stakeholders were actively considered. From the course perspective, requested materials were submitted to the course personnel on time and opponent group was given feedback on their plans, progress and outcomes. The representatives of the buildings were interacted with prior to observation sessions, except for the metro station, and summaries of the observations were subsequently sent to them, as requested by our client.

The main steps taken were collaboratively designed, and the support from KONE was extremely valuable. We believe that the success of this project can largely be attributed to the fact that the communication between the project group and the client as well as within the core project group were both effective and clear. Careful project design and execution resulted in well-directed effort that was seen as very low level of fruitless actions taken. In summary, the project was successful and execution was efficient.

### **Division of responsibilities**

The division of responsibilities described in project plan realized and turned out to be appropriate. Despite lead responsibilities varied, all of the project team members participated in all of the project phases and actively provided valuable comments on emerging issues. This can also be considered imperative to the success of the project, as the project team was quite small and the steps were highly interconnected.

Also, the lead responsibilities realized as planned. Jani lead model development and documentation together with Saku. Model validation as well as the default settings were done mainly by Saku, in addition to him being in charge of graphical outputs and visual representation, a task especially important in depicting the hotspot

network maps to our client. Sami lead reporting and took care of effectively all communication between the project group and the client, a core factor attributing to the success of the project. The division of more specific tasks was balanced based on both faced overall workload and team member availability. Communication worked very well within the team throughout the project, which had a positive influence of project team members, knowing what others were doing and what the overall project progress was.

### **Schedule and scope**

All of the project deadlines were met without difficulties, reflecting active team member participation and appropriate project scope. The project steps were largely conducted in the schedule that was initially planned and the only exception was that a bit longer period of time was allocated for literature review as we wanted to maintain adequate flexibility for model implementation. However, the initial literature review was conducted in sufficient depth for having solid understanding of the prevailing themes and being able to proceed with other steps. Eventually, there was no need for elaborating the literature review in a meaningful extent.

The scope of the project was feasible in terms of both amount of work and the difficulty of the work involved. Reaching the first objective was easier than the second objective, as assumed. All of the tasks were feasible concerning the project team's competencies, and each member's special capabilities were leveraged when dividing the tasks. We did not count working hours during the project but we believe the realized amount is reasonably close to the amount suggested for the credits provided by the course. As mentioned previously, an extended dynamic model would be even more beneficial to our client, but this would have likely increased the workload to amounts clearly exceeding the credit requirements. We are content with simply presenting our ideas regarding the extension of the model into more dynamic settings.

The most resource-consuming aspects were related to modeling and progress reporting, but also the data collection and other steps required meaningful effort from the project team.

### **Risk management**

The six risks that we identified in project plan and interim report were: member inactivity, too large workload, poor collected data quality, poor model specification, unsatisfactory model and disclosing confidential information. We took several preventive actions, as more elaborately described in the aforementioned documents, and were able to prevent all of the risks from occurring. Another potential risk that could underline the successfulness of the project is related to project handover from the project team to the client. It can occur that the ones who continue on this work

after us do not catch the thoughts behind the model. Therefore, we addressed that with several actions: we presented and discussed the model, extensively made comments within the code, developed a manual describing how the model functions, and distributed all project documents to the client.

### **Improvement potential**

We identify two definite opportunities for improvement that eventually are relatively minor issues. These do not include the notion of a more dynamic model since it was not in the scope of this study.

First, we started the modeling quite late in the project. Fortunately, that did not have an effect on our performance and the reasons behind the conscious decision were that data collection took some time and the project team was known to have less available time in the early stages of the project. If we had had more time earlier in the project, we could have been able to conduct preliminary versions of the model a bit earlier, decreasing the likelihood of potential complications close to final deadline.

Second, we could have collected a greater amount of data, which maybe would have improved reliability and generalizability. Our data collection process was ultimately constrained both by time and the small number of people involved. We could have also gathered data from the same locations during different times of the day, or gathered data from several comparable buildings. However, the purpose was not to get the most accurate data but, instead, to obtain enough insight for developing a basic model that can be modified according to contextual features when actually doing people flow simulations. We were ready to conduct additional observation sessions if the data would turn out insufficient, but eventually data reviews confirmed that additional sessions were not needed.