Integer and Stable Marriage Models for Assignments to Preschools

Skúli Magnús Sæmundsen

Thesis of 30 ETCS credits
Master of Science in Engineering Management

June 2014

Integer and Stable Marriage Models for Assignments to Preschools

Skúli Magnús Sæmundsen

Thesis of 30 ECTS credits submitted to the School of Science and Engineering
at Reykjavík University in partial fulfillment of
the requirements for the degree of
Master of Science in **Engineering Management**

June 2014

Supervisors:

Eyjólfur Ingi Ásgeirsson, Ph.D., Supervisor
Assistant professor, School of Science and Engineering, Reykjavík University

Sigrún Lilja Sigmarsdóttir, M.Sc., Co-Supervisor
Information Technology Auditor, City of Reykjavík

Examiner:

Þorkell Helgason, Ph.D., Examiner
Professor Emeritus, University of Iceland

Copyright
Skúli Magnús Sæmundsen
June 2014

Integer and Stable Marriage Models for Assignments to Preschools

Skúli Magnús Sæmundsen

30 ECTS thesis submitted to the School of Science and Engineering
at Reykjavík University in partial fulfillment
of the requirements for the degree of
Master of Science in Engineering Management.

June 2014

Student:

Skúli Magnús Sæmundsen

Supervisors:

Eyjólfur Ingi Ásgeirsson, Ph.D.

Sigrún Lilja Sigmarsdóttir, M.Sc.

Examiner:

Þorkell Helgason, Ph.D.

The undersigned hereby grants permission to the Reykjavík University Library to reproduce single copies of this project report entitled **Integer and Stable Marriage Models for Assignments to Preschools** and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the project report, and except as herein before provided, neither the project report nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Date

Skúli Magnús Sæmundsen
Master of Science

Integer and Stable Marriage Models for Assignments to Preschools

Skúli Magnús Sæmundsen

June 2014

Abstract

Every spring, a large group of children graduate from preschools in the municipality in order to begin attending elementary school in the autumn. Around this time the City of Reykjavík is faced with a decision problem how to assign children to vacant places at the preschools, which are scattered across the city. In this thesis, two mathematical models are presented that might help with the assignment of the vacant places, and they are compared to the assignment process currently being used by the city. The goal of the models is to meet the applicants' preschool preferences as far as possible, while ensuring the applicants' equality and fulfilling the assignment rules of the city. Additionally, it is assessed if such models could be beneficial for private preschools. The outcome of the assignments will undeniably affect everyday life of families that use the preschool services provided by the City of Reykjavík. It is thus important that they are carried out in the best way possible. The models under consideration are an integer programming model and a stable marriage model. Both models use the preferences given by the parents in their application as well as the distances between the applicants' homes and the preschools in order to reach the best possible solution. The results show that a model that is based on the stability of marriage method is a promising alternative for the City of Reykjavík. The model provides a good solution that is both transparent and in accordance with the rules of the city. Models like these have been applied abroad for assignments to schools with good results. Additional research is required before such a model can be implemented. If the models and the current assignment process are to be compared under the same assumptions, then the application process must be changed. It will then be possible to examine whether the models are really better than the current method and thus whether their usage is justified.

Stærðfræðilíkön fyrir úthlutanir á leikskólaplássum

Skúli Magnús Sæmundsen

Júní 2014

Útdráttur

Á hverju vori lýkur stór hópur barna leikskólavist á höfuðborgarsvæðinu til að hefja grunnskólagöngu þá um haustið. Um það leyti stendur Reykjavíkurborg frammi fyrir ákvörðunartökuvandamáli um hvernig úthluta eigi lausum plássum leikskóla, sem staðsettir eru víðsvegar um borgina. Í þessari ritgerð eru skoðuð tvö ólík stærðfræðilíkön sem kunna að koma að gagni við úthlutun á lausum plássum og þau borin saman við núverandi fyrirkomulag hjá borginni. Líkönin eru hönnuð með það að markmiði að mæta sem best óskum foreldra um leikskólapláss en jafnframt tryggja að jafnræðis sé gætt við úthlutunina og að úthlutunin sé í samræmi við reglur borgarinnar. Að auki er lagt mat á það hvort slík líkön gætu komið að gagni við úthlutanir hjá sjálfstætt starfandi leikskólum. Útkoma úthlutunarinnar kemur óneitanlega til með að hafa áhrif á daglegt líf fjölskyldna sem nýta sér leikskólaþjónustu Reykjavíkurborgar og því mikilvægt að hún sé með besta móti. Líkönin sem um ræðir eru annarsvegar heiltölubestunarlíkan og hinsvegar líkan sem byggir á stöðugu hjónabandsaðferðinni. Bæði líkönin nýta sér val úr umsókn umsækjanda ásamt fjarlægðum frá heimili umsækjanda til leikskóla til að ná fram sem bestri niðurstöðu um úthlutun. Niðurstöðurnar gefa til kynna að líkan sem byggir á stöðugu hjónabandsaðferðinni gæti talist vænlegur kostur fyrir Reykjavíkurborg. Líkanið skilar góðri lausn sem er jafnframt gegnsæ og í samræmi við reglur borgarinnar. Líkönum sem þessum hefur verið beitt erlendis við úthlutun á skólaplássum með góðum árangri. Frekari rannsókna er þörf áður en slíkt líkan er tekið í notkun. Ef gera á beinan samanburð á líkönum og núverandi fyrirkomulagi eftir sömu forsendum, þá er þörf á breytingum á umsóknarferli svo kanna megi hvort líkön af þessu tagi skili í raun þeim ávinningi umfram núverandi aðferð sem réttlætir notkun þeirra.

Acknowledgements

The author would like to thank Dr. Eyjólfur Ingi Ásgeirsson and Sigrún Lilja Sigmarsdóttir M.Sc. for their guidance and support during the creation of this thesis. Also, the author would like to thank all those who contributed to the making of this thesis.

Contents

List of Figures **ix**

List of Tables **x**

1 Introduction **1**

2 Background **5**

2.1 The preschool system 5

2.1.1 Application process and waiting lists for public preschools 6

2.1.2 Assignment rules 6

2.1.3 The assignment process 7

2.1.4 Transfers between preschools 8

2.1.5 Private preschools 8

3 Literature review **11**

3.1 Integer programming 11

3.2 Stable marriage 12

3.2.1 The Gale and Shapley algorithm 13

3.3 School choice mechanisms 14

4 The models **17**

4.1 Integer programming model 17

4.1.1 Constraints 18

4.1.2 Utility function 18

4.1.3 Model description 19

4.2 Stable marriage model 21

4.2.1 Model description 22

4.3 Verification of the models 24

4.3.1 Test case 24

4.4 Limitations of the models 28

5 Data **31**
5.1 Data collection 31
5.1.1 Data from RVK 31
5.1.2 Distances 33
5.1.3 Statistics 33
5.2 Data handling and proposed inputs 36
5.3 Problems with the data and the data collection 38

6 Results **39**
6.1 Model results 39
6.1.1 Integer programming model 40
6.1.2 Stable marriage model 42
6.1.3 Comparison of the models and current practice 46

7 Conclusions **57**

A IP model in MPL **59**

B SM model in MATLAB **61**

Bibliography **63**

List of Figures

1.1 Visualization of a subset of the problem 2

4.1 Test case: Visualization 24

5.1 Statistics: Utilization of choices from applications 33
5.2 Statistics: Preschool demand and capacity 34
5.3 Statistics: Total preschool demand 35
5.4 Statistics: Preschool capacity 35
5.5 Statistics: Distance to preschools 35

6.1 Results of IP model: Assignments to one preschool 45
6.2 Results of SM RVK: Assignments to one preschool 45
6.3 Results of SM neighbourhood: Assignments to one preschool 45
6.4 Results: Ability to meet parents' preferences 48
6.5 Results: Ability to meet preferences of parents of priority children 49
6.6 Results: IP model distance 51
6.7 Results: SM RVK distance 51
6.8 Results: SM neighbourhood distance 51
6.9 Results of IP model: Assignments of priority children 53
6.10 Results of IP model: Assignments of children living +50 km from Reykjavik 53

List of Tables

4.1 Test case: Distance data 25
4.2 Test case: Application data 25
4.3 Test case: Parents' preferences matrix 26
4.4 Test case: Preschool preferences 26
4.5 Test case: Solution 26

6.1 Results: Summary of main results from the models 46
6.2 Results: Comparison of methods 47

Chapter 1

Introduction

The City of Reykjavík (RVK) operates 64 public preschools and has supervision over 17 private preschools within the municipality [7]. The custodians (parents) of the children can apply to these preschools on their own initiative since the preschools are not mandatory. The application processes for the public and private preschools are separate, and this is also true of the acceptance process. The parents are able to state their preferences by applying for a subset of the public preschools in their application. Since demand is considerably higher than the supply of places within the preschools, it is not possible to enrol all of the children simultaneously. RVK has therefore created assignment rules for the preschools, which are supposed to safeguard the equality of the applicants, and which both private and public preschools are to follow during the enrolments [6, 2].

The continuous cycle of enrolling, transferring and graduating children to and from the preschools is an important part of the operations for both public and private preschools. The graduations create a decision problem on how to assign children to the vacant spots in the preschools. There is a peak in the assignments in the spring, since this is the time when children are about to leave the preschools in order to attend elementary school in the following autumn. During this time, multiple places will simultaneously become available in preschools positioned all over the city.

During the peak of the assignments every year, as many as 1400-1700 children are assigned to the public preschools alone [22]. Figure 1.1 shows a visualization of a subset of the problem. The figure shows 64 public preschools (blue markers) and 185 children (red markers) waiting for assignment to the preschools. The task is to assign the children that are on the organized waiting lists to the preschools so as many of the parents' preferences are met as possible while respecting the assignment rules of the city. Therefore, the problem is essentially an assignment problem. Assignment problems are well

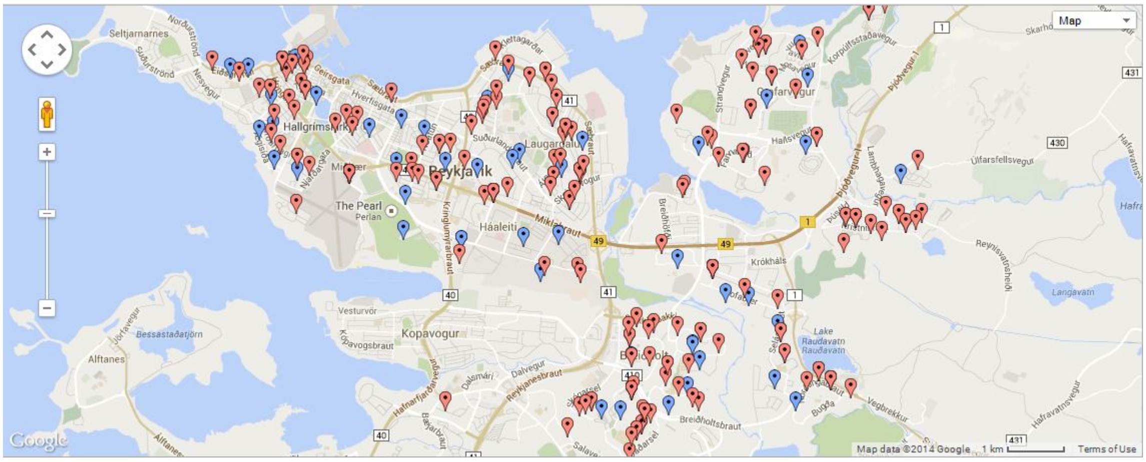

Figure 1.1: Visualization of a subset of the problem: The blue markers represent the 64 public preschools and the red markers 185 children that are waiting to be assigned, out of a total of 1400-1700 children. The Figure has been created using the Batchgeo mapping tool (http://batchgeo.com).

known within the field of operational research and have been used in various applications [25, 13, 14]. For RVK, it is important that the assignments ensure equality and are carried out efficiently. Inadequate performance of this task will lead to parental disapproval of the provided service as well as added cost for RVK, since transfers between preschools will take up more employee time than necessary. Furthermore, since transfers can have an undetermined waiting time [6], it is all the more essential that the assignments are as good as possible from the beginning.

The aim of the study is to create a method that can best meet the preschool preferences of the applicants, while ensuring their equality and fulfilling the assignment rules of the city during the peak in the assignment period. The method should provide a long-term solution that is both economical and practical. Two different mathematical models will be constructed and then compared with the current assignment process. The study will analyse the benefits of each method and then answer whether it would be appropriate to use these models for the assignment process in RVK. Additionally, it is assessed if such models could be beneficial for private preschools.

The models under consideration are an integer programming model and a stable marriage model that is based on the Gale and Shapley algorithm [18]. A version of both models has been used in the public sector. The New York Department of Education (NYDOE) implemented an assignment mechanism based on the Gale and Shapley algorithm in the school year 2003-2004. The algorithm was used to assign over 90,000 students to public high schools that year. The results showed that the use of the assignment mechanism had made it possible to match 20,000 more students to their initial choice list than in the previous year [10]. Integer programming has been used, for example, by the San Francisco Police Department for the scheduling and deployment of police patrol officers, a successful implementation, which at the time was estimated to save the police department $11 million annually [27].

The two models are fundamentally different in their approach to the assignment problem and exhibit trade-offs, which are of importance to the process in RVK. The trade-offs are mainly between transparency and optimality. The stable marriage model will provide a good solution that is also transparent. However, the integer programming model will provide an optimal solution that is less transparent.

The main contributions of this study are:

- A proposed new assignment method for RVK to use in its preschool assignments.
- Better ability for RVK to meet the preferences of parents, boosting the provided service.
- The benchmarking of two different assignment mechanisms in order to provide insight into the topic of school choice mechanisms.

The paper is structured in the following way: Chapter 2 contains a brief overview of the background and explains the current assignment process. Chapter 3 is a literature review. In Chapter 4, the models are formulated, explained and verified. Chapter 5 introduces the dataset. The results and a comparison of the models are presented in Chapter 6. Chapter 7 contains conclusions and notes on further work.

Chapter 2

Background

In this chapter, a brief overview of the preschool system and its components will be given, followed by an outline of the current assignment process. Five interviews were carried out in order to get an insight into the assignment process. Two were conducted with administrators of public preschools, two with administrators of private preschools, and one with the enrolment administrator for the Department of Education and Youth in RVK. Some of the preschool administrators requested to remain anonymous, and those requests will be respected.

2.1 The preschool system

Preschools are the first step in the Icelandic school system and are intended for children who have not yet started elementary school. Each municipality is responsible for the operation of their preschools and thus for ensuring assignment of children to the preschools [2]. The Department of Education and Youth (Skóla- og frístundasvið (SFS) in Icelandic), is the largest department of the City of Reykjavík and operates the public preschools. The SFS office also gives operating permits and has supervision over private preschools and daycare services within the city. One task involved in the supervision is to make sure that the services are in accordance with the law regarding preschools, a responsibility that is shared with the preschool administrators [2]. Besides the law, the preschools are operated in accordance with rules regarding preschool services [6], which are put forth by RVK. Rules apply for the application, enrolment and collection of fees for preschools within the city. The rules are mandatory for both public and private preschools unless something different has been negotiated with the SFS office [6].

On 1st October 2013, there were 7,022 children enrolled in all of the preschools in the city. There were 5,990 children in 64 public preschools and 1,006 children in 17 private preschools. Furthermore, there were 26 children in private preschools located outside of the city and 783 children accommodated by daycares centres [7].

2.1.1 Application process and waiting lists for public preschools

Parents are able to apply to a preschool of their choice provided that they fulfil the prerequisites for enrolment to the preschools of RVK. The main prerequisites are that the child's domicile and permanent residency are in the municipality of Reykjavik and that the parents are not late with their payments to the SFS office [6].

As soon as a child has received an identification number (ID), it is possible to apply online for a public preschool. During the application process for the preschools, it is possible to choose between one and five preschools. The sequence in which the preschools are chosen does not give them any weight. If the parents are interested in a certain preschool, they can state their preferences in text in the application or contact the preschool administrator or the SFS office, but neither is mandatory [5, 22].

Since there is not enough room in the preschools for all the children, the city organizes waiting lists, where each preschool has a separate list. Once an application has been placed, children younger than one year old are put on hold. Children that have reached the age of one year are queued on the waiting lists of all of the preschools chosen by the parents. Parents are usually offered a preschool place for their child during the year in which the child reaches two years of age. If the parents choose to decline the offered preschool, their application remains on the waiting lists of the preferred preschools [22].

2.1.2 Assignment rules

First time applications and transfers between the preschools in RVK are controlled by a set of rules [6] that determine the priority of a child on the waiting lists.

- **Age rule.** The main rule is that children get places on the waiting list based on their age, where an older child has priority over a younger child, irrespective of when the application is placed.

In special circumstances, applicants can apply for exemption from this rule. If granted, the child gets placed at the top of the waiting list at his/her preferred preschool. A child with an accepted priority application will be referred to as having priority over the other children. Circumstances are listed below.

i. A child that is between the age of four and five.

ii. A child with disabilities or a child with serious developmental problems.

iii. A child that lives in a difficult social environment, such as:

   a) Child protection cases.

   b) Serious illness or serious disability amongst the child's family members.

   c) Children of parents that have not yet reached a legal age.

   d) Child of a single parent supporting three or more children where the oldest child has not yet passed the age of nine.

   e) Triplets.

iv. Children of employees of the preschools operated by RVK.

If there is an instance in which there are two or more children that fit these special circumstances, then the main rule takes effect and the child that is older receives priority. Rules i, ii and iii also apply to private preschools unless something different has been negotiated with the SFS office [6].

The assignment rules are not clear on if the oldest child should always receive his/her parents' most preferred choice of preschool, while those preschools still have available capacity during the assignments. Although this might be implied by the purpose and structure of the rules, we will define such assignments as being a *strict* following of the assignment rules.

2.1.3 The assignment process

During the spring, there is a peak at the assignments since older children in the preschools are leaving to attend elementary school. The assignments take place between March and June, and newly accepted children begin attending the preschools between June and September. The last few years have seen between 1400-1700 assigned to the public preschools alone during this period [22, 6].

The current assignment process has been used for the past three years for enrolments in the public preschools. The SFS office hosts several meetings with up to ten preschool administrators from the same neighbourhood along with two employees from the SFS office. In the meetings, which took around 1.5 hours each during the spring assignment of 2013, the waiting lists for each of the preschools are discussed. Each preschool administrator knows the available capacity of their preschool. The administrators decide on which children will attend each preschool based on the written preferences in the applications or any preferences communicated by the parents before the meeting. The assignments follow the rules in Section 2.1.2 as mandatory guidance [22].

In parallel to the rules, the members at the meeting try to take other considerations into account, such as if there are any siblings in other preschools. However, sometimes there is no obvious connection to any of the preschools. This happens in cases where parents have not stated any preferences in the application or have not contacted the preschool administrator of the preferred preschool. In these cases, it is likely that the parents will not receive their top choice of preschool due to lack of information. To deal with such instances, assignment has, for example been based on the distance between the child's home and the preschools [22, 24]. Throughout the rest of the year, the preschool administrators can enrol the children that are on top of their waiting list themselves if there happen to be any vacancies [22].

2.1.4 Transfers between preschools

Once a child has been enrolled into a preschool, parents are permitted to ask for a transfer. There can be numerous reasons for a transfer, but some of the more common ones are if a child has a sibling in another preschool, if the preschool is not a preferred choice, if the preschool is not close enough to the applicants home, a family is moving between neighbourhoods or out of the city. Transfer between preschools follows the same rules as a general first-time application, meaning that the child gets placed on the waiting list at the preferred school [22, 6, 24]. Although there can be multiple reasons for the transfers, the transfer rate can be a good indication of the level of parental satisfaction with the assignments.

2.1.5 Private preschools

The private preschools are responsible for their own enrolments. The two private preschools that were interviewed stated that it is possible to apply online through the website of the preschools. The application is then queued on a waiting list in accordance with the rules listed in Section 2.1.2. Similarly to the public preschools, most of the enrolments take place during the spring, meaning that most of the children start attending the preschools at the end of the summer. Out of the seventeen private preschools operating within the city, two preschools were interviewed in order to obtain an idea of how they manage their enrolments. These preschools will be referred to as preschools A and B.

Preschool A mainly follows the age rule, however, they do examine each case individually. Factors that affect the decision are siblings and the distance that the parents live from the preschool. If a decision had to be made between two children of similar age, a younger child that has a sibling in the school would have priority. Furthermore, children living in the neighbourhood would be preferred over children living in other neighbourhoods.

In preschool B, the administrator was dealing with capacity problems. Demand was considerably higher that the supply of available places within the preschool, a problem that has been growing year on year. This excessive demand leads to problems with the assignment rules, since following the rules would result in the assignment of too many 'old' children to the preschool. Adjustments were therefore needed in order to maintain the right proportions of children in each department of the preschool.

One of the private preschools has its assignment process listed on their website. Their enrolments follow a first-come, first-served service policy, where the date of the application determines where the child is on the waiting list. Like in the case of private preschools A and B most of their enrolments take place in the spring [8].

Parents are able to simultaneously apply to both private and public preschools, which means that many of the waiting lists contain the same children. This complicates things a little since, for example, a child could be accepted to a private preschool and also accepted during the assignments to the public preschools. In these situations, it is helpful that the preschool administrators are able to enrol the next child on the waiting list.

Chapter 3

Literature review

The following literature review will focus on introducing the proposed methods of improving the preschool assignments at RVK. The review will begin by presenting the classic assignment problem, and then a special case of the assignment problem known as the stable marriage problem will be considered. Finally, several examples of real life applications of the problem will be given.

3.1 Integer programming

$$
\begin{aligned}
\text{minimize} \ & \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij} \\
\text{subject to} \ & \sum_{i=1}^{n} x_{ij} = 1,\ j = 1,..,n, \\
& \sum_{j=1}^{n} x_{ij} = 1,\ i = 1,..,n, \\
& x_{ij} \geq 0,\ \forall\, i,j.
\end{aligned} \tag{3.1}
$$

The classic assignment problem (AP) is a special case of the minimum cost flow problem [14]. In 1955, Kuhn published an article on the Hungarian method, which was the first method to solve the assignment problem efficiently. The AP which can be seen in Equation (3.1) is concerned with matching equal numbers of agents i $(i = 1, 2, ..., n)$ and tasks j $(j = 1, 2, ..., n)$ to each other where each task can only be performed by one agent and each agent can only perform one task. The assignment of the agents to the tasks is associated with different costs c_{ij}. The decision variable x_{ij} tells weather or not an agent

i is performing task j. The objective is to assign the agents and tasks together such that the overall cost of the assignment is minimized [13, 14].

The AP with side constraints builds on the classic assignment problem. It takes into account other factors that might impose limits on such an assignment, such as assigning agents with the right skills to the tasks or recognizing priority classes for tasks [25]. These problems are generally NP hard [15]. However, since the assignment problem can be considered as one of the easier problems within the NP hard class, it is reasonable to believe that the problem can be solved for non-trivial instances.

3.2 Stable marriage

The stable marriage problem can be considered as a special case of the assignment problem that is concerned with finding a stable matching of two elements given a preference list for each element. One way of thinking about the stable marriage problem is based on a set of n men and n women that are to be matched together or married. Both the men and the women rank their preferences of all individuals of the opposite sex. The men (or the women) then take the part of the *proposer*, and ask their most preferred woman (or man) to marry. The goal is to find a satisfactory matching of all of the individuals. A matching is said to be *stable* if there exists no couple where there is man and a woman who prefer each other rather than their current spouses [18].

The first and most well-known generalization of the stable marriage problem can be traced back to the National Resident Matching Program (NRMP) in 1952. This program utilized an algorithm that matches graduating medical students (medical graduates) in the United States with hospitals. Before the centralization of the matchings, hospitals competed amongst themselves for a scarce supply of medical graduates, sometimes offering contracts as early as two years before a position needed to be taken up. Efforts were made to improve the situation but new problems emerged, resulting in an unsatisfactory situation for both the medical graduates and the hospitals. This failure resulted in the centralization of the assignments through the NRMP [21].

In 1962, Gale and Shapley published a seminal paper on the stable marriage problem. In their paper, they introduced a symmetric problem to the one faced by the NRMP 10 years previously. They described a real life problem faced by a college that is considering a set of applicants for admission. They argue that, through collaboration between the colleges using a centralized admission procedure, many of the problems that individual colleges face with their admissions can be avoided. For example, one of the problems faced by the colleges is the balancing act of attempting to receive no more and no less students then the capacity of the college allows. They proved by describing an algorithm that given a set of applicants and institutions with their respective preferences and quotas it is always possible to find a stable matching in $O(n^2)$ running time. Furthermore, the solution will also be the best possible stable matching for the party that does the proposing [18]. The problem has since attracted the attention of many researchers from various areas such as mathematics, economics and computer science [21].

3.2.1 The Gale and Shapley algorithm

To determine the degree of fairness of the proposed algorithm the concept of *stability* is introduced. It is highly desirable that there exist no applicant a and college c where all of the following is true, since the solution would then become *unstable* [21, 18].

i. College c is acceptable to applicant a and a to c.

ii. Either a is unmatched, or a prefers c to the assigned college.

iii. Either c does not have all of its places filled in the matching, or c prefers a to one of the assigned applicants.

Gale and Shapley deferred acceptance algorithm

In relation to the colleges, the algorithm takes the following steps:

Step 1 Each applicant a proposes to the first college c on their rank-ordered preference list. Each of the colleges then accepts the applicants onto their waiting list in accordance to their rank-ordered preference list.

Step n Each applicant rejected in the previous step proposes to the next college on their preference list. In each cycle, the colleges revise their waiting list, always selecting their most preferred candidates.

The algorithm terminates once the applicants have proposed to all of the colleges on their list and have either been accepted or rejected by all. The colleges now enrol the applicants that are on their waiting list [18]. The following small example proves that the solution is stable.

Imagine an applicant α that has been assigned to the waiting list of college β but that would rather be on the waiting list of college γ. Then at some point the applicant α asks to be accepted to college γ but is rejected for a group of applicants more preferred by college γ. This sequence of events shows that the solution must be stable [18].

Similar to the above example Gale and Shapley show by induction that the solution is *optimal* in the sense that the applicants are guaranteed to receive their best possible college under any stable assignment (or colleges their optimal choices of applicant if the roles are reversed). The fact that the algorithm provides a solution that is stable optimal for the proposing party does, however, not come without a cost to the receiver of the proposal, since it has been shown that the optimal solution for the proposer is simultaneously the worst stable solution for the receiver of the proposal [21]. In the mid-1990s, students expressed their concerns that the matchings performed by the NRMP were more favourable to the hospitals than to the graduates (hospital optimal matching). The algorithm was therefore redesigned a few years later [28, 26].

Since its introduction, the Gale and Shapley algorithm has been tested and further developed. In 1981, Dubins and Freedman proved that no group of students can improve the results of all group members by lying about their preferences if all other participants still state their true preferences [17]. Dubins and Freedman results were later supported by Gale and Sotomayor [19]. The algorithm has been further developed to handle some natural relaxations of the problem that might occur in reality, such as allowing incomplete preference lists or preference lists with ties. The NRMP is still using the matching algorithm and its durability has led to the adoption of such an algorithm in other countries [23, 3].

3.3 School choice mechanisms

School choice mechanisms have received increasing attention over the last few years [11, 16, 10, 9, 12]. The term *school choice* means that the aim is to let the parents chose where their children will attend school. Assignments in the United States have, in the past, been more focused on where the children live. Children have mainly been assigned to public schools in their own neighbourhoods without regard to the appropriateness of the school for the child or the quality of the assigned school [11]. Since schools have a limited capacity for students, the focus has turned towards the development of student assignment mechanisms (models) with the aim of assign students to schools in a way that best meets parental preferences and school priorities. To estimate the performance of such a mechanism, three main parameters have been used [11]:

- **Stability** has been previously introduced and is also referred to as *justified envy* in the school choice literature.
- **Strategy proof** means that the students cannot increase their probability of assignment to any schools by falsifying their true preferences.
- **Pareto efficiency** means that no improvements can be made to the assignments, whereby one or more of the students is better off and the remaining students are at least as well off as they were previously.

One of the more widely used mechanisms is the *Boston mechanism*. This mechanism is a systematic approach that uses preferences and priorities based on, for example, state and local laws to assign students to schools. The mechanism has been used in cities such as Boston, Denver and Minneapolis in the United States. It has been studied to some extent and has received criticism for giving participants the incentive to hide their true preferences. The mechanism has also been shown to be *unstable*, which has created grounds for legal action [11, 16].

When school choice mechanisms are being evaluated, the focus has often turned toward the Gale and Shapley algorithm [11, 16, 10, 9]. The main difference between the college admissions problem introduced by Gale and Shapley and the schools choice problem is that in the college admission problem the colleges are agents who have preferences over the students. However, in the school choice problem the schools don't have preferences but instead have *priorities* based on rules and regulations. This problem can therefore be looked at as a two-sided matching market. This has motivated the development of mechanisms such as the *Gale-Shapley student optimal stable mechanism*, which has been shown to be both *strategy proof* and *stable*. However, it is not necessarily *Pareto efficient* [11].

The endurance of the Gale and Shapley algorithm has prompted institutions like the New York City Department of Education (NYCDOE), which has supervision over the largest public school system in the United States, to take up a similar mechanism. In 2003-2004, the NYCDOE had to match over 90,000 students to high schools. By using the new mechanism, they managed to match 20,000 more students to their initial choice list than in the previous year [10]. During the high school matching in 2012, it was possible to match 84% of the applicants to one of their top five choices [4]. The Boston mechanism has been redesigned and now uses the Gale and Shapley deferred acceptance algorithm [9].

Chapter 4

The models

In this chapter, two models are presented that could be beneficial for the assignment process in RVK. The models have a fundamentally different approach in solving the assignment problem, and their solutions might not be the same. The integer programming model will give an optimal solution (most matchings) that is not very transparent, while the stable marriage model will give a good solution that is transparent.

Since it is not possible that all parents receive their first choice of preschool due to lack of capacity, the question is how best to assign the available places so that as many receive their preferred choice as possible while fulfilling the assignment rules of RVK. In order to provide families with the best possible service, emphasis will be put on giving parents a preschool from a rank-ordered preference list. Furthermore, interviews with preschool administrators (see Chapter 2) revealed that the distances between preschools and homes affect the assignments, which is a finding that will be further supported in Chapter 5. Distance will therefore be the second factor included in the models.

This chapter is constructed as follows. First, the integer programming model (IP model) is introduced, followed by a section on the stable marriage model (SM model). Finally, both models are verified using the same test case.

4.1 Integer programming model

The preschool problem at RVK can be considered a special case of the assignment problem introduced in Chapter 3. In the preschool problem, there are an unequal number of children (agents) and preschool places (tasks). In order to formulate the problem as a traditional one-to-one assignment problem, the following can be done:

- Split each preschool into a separate (but identical) preschools, in relation to how much capacity is available within each school.
- Create additional dummy preschools for the children who don't get assigned to a preschool.

However, the preschool problem requires additional constraints to account for the assignment rules introduced in Section 2.1.2. The problem is therefore essentially an assignment problem with side constraints and is NP-hard.

4.1.1 Constraints

The constraints of the model cannot be violated under any circumstances. They ensure that practical limitations of the problem are taken into account and that the assignment rules of RVK listed in Section 2.1.2 are not broken. The constraints are as follows:

- **Age rule.** RVK's main rule is the age of the children. A child cannot be enrolled into a preschool while there is still an older child on the waiting list. This is independent from when an application is placed.
- **Priority.** Children with an accepted priority application automatically go to the top of the waiting list.
- **No splitting.** A child can only be assigned once, i.e., a child cannot be assigned to two schools at the same time.
- **Capacity.** The capacity of the preschools has to be respected. Only a limited number of children can attend each school. The vacancies at each preschool are estimated every year once the oldest children have moved on to elementary school or have been transferred to other preschools.

4.1.2 Utility function

The objective (utility) function is used to maximize the accumulated utility for the service in relation to how the model can fulfil the preferences of the parents while making them travel as short a distance as possible. The given utility is only an indicator that one element is more important than another (higher utility indicating greater importance) and does not quantify the true meaning of the service for the families. The decision variables included in the model will be constrained to be binary, indicating whether or not one particular child is assigned to one particular preschool.

- **Preferences of the parents.** By using the ranked-ordered preferences in a child's application, it is possible to maximize the likelihood that the child's parents receive a preschool that they will be happy with. For example, if the parents want all of their children (all siblings) to attend the same preschool, this should be reflected in the choices in their application. Ensuring that most of the parents get their number one choice of preschool should take care of this request. However, priority is not granted to siblings under the current assignment rules, so no special consideration will be made in the formulation.
- **Distance travelled.** It is considered desirable that the distance travelled by the parents between their home address and the given preschool is minimized. The reason is twofold. Firstly, as seen in Chapter 2 distances play a role in the assignments made by the preschool administrators. Likewise, parents have applied for transfers because of preschools being too far away from their home. Secondly, the assignments create new driving routes within the municipality, since the parents are driving to and from the preschools. These distances should be kept as short as possible in order to minimize the financial and environmental cost of the driving.

  The parents' preferences are given a considerably higher utility than for making the parents travel as short a distance as possible, since meeting the parents' preferences is more important than the distance they travel.

4.1.3 Model description

Indices

The model uses two sets of indices, which represent the children and the preschools.

C : set of children, $n = |C|$
S : set of preschools, $m = |S|$

Data

The following five datasets are used as inputs for the model.

DB_i : date of birth for child $i \in C$

Cap_j : capacity of each preschool $j \in S$

PC_i : 1 if child i has an accepted priority application, 0 otherwise.

D_{ij} : distance matrix where element (i, j) contains the distance from the home of child i to all the preschools j.

P_{ij} : preference matrix where element (i, j) contains the given utility for assigning child i to preschool j.

Decision variables

The decision variables of the model are binary variables indicating whether a child i is assigned to preschool j.

$$x_{ij} = \begin{cases} 1, & \text{if child } i \text{ goes to preschool } j; \forall\, i \in C, j \in S \\ 0, & \text{otherwise} \end{cases}$$

Utility function

$$\max \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}(P_{ij} + \alpha / D_{ij}) \tag{4.1}$$

Constraints

$$\sum_{j=1}^{m} x_{ij} \leq 1,\ \forall\, i \in C \tag{4.2}$$

$$\sum_{i=1}^{n} x_{ij} \leq Cap_j,\ \forall\, j \in S \tag{4.3}$$

$$\sum_{j=1}^{m} x_{ij} \leq \sum_{j=1}^{m} x_{\hat{i}j},\ \forall\, i, \hat{i} \in C,\ : PC_i \leq 0 \text{ and } DB_i > DB_{\hat{i}} \tag{4.4}$$

$$\sum_{j=1}^{m} x_{ij} = 1,\ \forall\, i \in C : PC_i > 0 \tag{4.5}$$

$$x_{ij} \in \{0, 1\},\ \forall\, i \in C, j \in S \tag{4.6}$$

The utility function (4.1) adds together utility given for meeting the parents' preferences and utility for ensuring that the parents travel as short a distance as possible. A weight factor α is used to determine the importance of the distances. Increasing α will add weight to the distances, meaning that higher utility is given for them in the assignments. The utility function determines how good the assignment is by maximizing the accumulated utility. We will use $\alpha = 1$ unless otherwise noted, thereby giving priority to the requests of the parent. Chapter 5 will contain a description of how the P_{ij} and D_{ij} matrices are formulated. Constraint (4.2) ensures that a child can only be assigned to one preschool. Constraint (4.3) ensures that the assignments don't exceed the capacity limits of the preschools. Constraint (4.4) ensures that an older child is assigned before a younger child and thus fulfills the age rule introduced in Section 2.1.2. If child $i \in C$ is younger than any of the other children $\hat{i} \in C$, then child i will need to wait until all of the children $\hat{i}$ have been assigned to preschools, before child i can be assigned. Chapter 5 contains a description of how DB_i is modified. Constraint (4.5) ensures that if a child has priority he/she is guaranteed a place at one of the preschools. Constraint (4.6) ensures that the binary condition of the variables is satisfied.

4.2 Stable marriage model

Experiences with matching algorithms have shown that they are most likely to be successful if they are able to produce *stable* matchings, as explained in Chapter 3. The goal with the SM model is to make it as favourable as possible to the applicant while meeting the constraints set by RVK. The formulation used is based on the Gale and Shapley deferred acceptance algorithm and focuses on a one-to-one matching of the participants. The model will be formulated to be children stable optimal, and thus the children will be made to do the 'proposing'. This will provide a stable solution that is the optimal stable solution for the families. We assume that there are no internal connections between the children or the preschools. Such connections are, however, sometimes permitted and have been known to affect the stability of the solution [26].

The classic assignment problem and the integer programming model introduced in the previous section have an organized objective function or utility function that is to be minimized or maximized depending on the application. In the case of the stable marriage model, such a function is not necessary since the algorithm uses two preference lists to match the participants together.

Similarly to the IP model, the stable marriage model uses both the rank-ordered preferences stated by the parents in their applications and the distance that they travel to the preschools in order to reach a solution. In Chapter 2, the application process for the public preschools was introduced. Parents are able to choose between one and five preschools in their applications, and there are no restrictions as to where parents can apply for a preschool. Distances are used to compile a complete parent preference list for all of the public preschools. This is done in order for the model to be able to assign a child to the closest available preschool should it not be possible to meet his/her parents' preferences. It is desirable that all of the children get assigned to preschools while capacity allows.

In the case of the preschools, the children are not ranked according to preference, rather the assignments are controlled by a set of rules discussed in Section 2.1.2. These rules will be used to form a complete priority (preference) list for the preschools. The only inputs needed for the SM model are the two aforementioned preference lists and the available capacity of all of the preschools

4.2.1 Model description

Preference matrices

In order to generate a complete preference list for the set of children (parents) and preschools, two preference matrices are created. The preference matrix for the children is a matrix of size n x m where n is the number of children in the assignments and m is the number of preschools. Each row $i \in C$ contains the rank-ordered preferences of the children i of all of the preschools $j \in S$, beginning with the preschools that the children i rank the highest. In order to compile a complete preference list, the choices in the applications and the distances to the preschools will be used. Similarly, a preference matrix for the preschools is created of size m x n. Each row $j \in S$ now contains rank-ordered indices of all of the children $i \in C$, beginning with the children with the highest priorities according to the assignment rules.

Gale and Shapley algorithm, a children optimal stable version

Initialize all of the children $i \in C$ and the preschools $j \in S$ to be free
while $\exists$ *a free child* i *that still has a preschool* j *to propose to* **do**
  Child i proposes to all the j on its preference list
  if *enough capacity* **then**
    i is assigned to the waiting list of j
  else
    Compare i with the current children on the waiting list for the j and reject the child that least complies with the assignment rules of RVK.
  end
end

Algorithm 1: Pseudocode of the stable marriage algorithm: A children optimal stable version.

To begin with, each child i will propose to the preschool j at the top of their list. Since, in the beginning, there is enough capacity in the preschools, all of the preschools j will accept the children on to their waiting lists.

Once the capacity of preschools has been reached the preschools compare the subset of children currently on their waiting list with the current proposing child i. The algorithm now looks for the child that is the most 'unpopular' in correspondence with the rules in Section 2.1.2. The most 'unpopular' child is then removed from the waiting list.

If the removed child i is the one that is currently doing the 'proposing', the algorithm will continue to propose to all of the preschools on the child's rank-ordered preference list until the child gets a 'yes' from one of the preschools j and is accepted onto a waiting list.

If the removed child was one of the children currently on the waiting list at preschool j, the child is labelled as 'rejected' and will get a chance to 'propose' to all of the preschools on its list again in the next cycle.

The algorithm terminates once all of the children i have either been assigned to a preschool j or have been rejected by all of the preschools on their list. The list of rejected children are now the children that comply least with the rules of RVK. Under the current assignment rules, these are the youngest children. The remaining children on the waiting lists are now the children suggested for enrolment to each of the preschools. A pseudocode of the algorithm can be seen in Algorithm 1.

4.3 Verification of the models

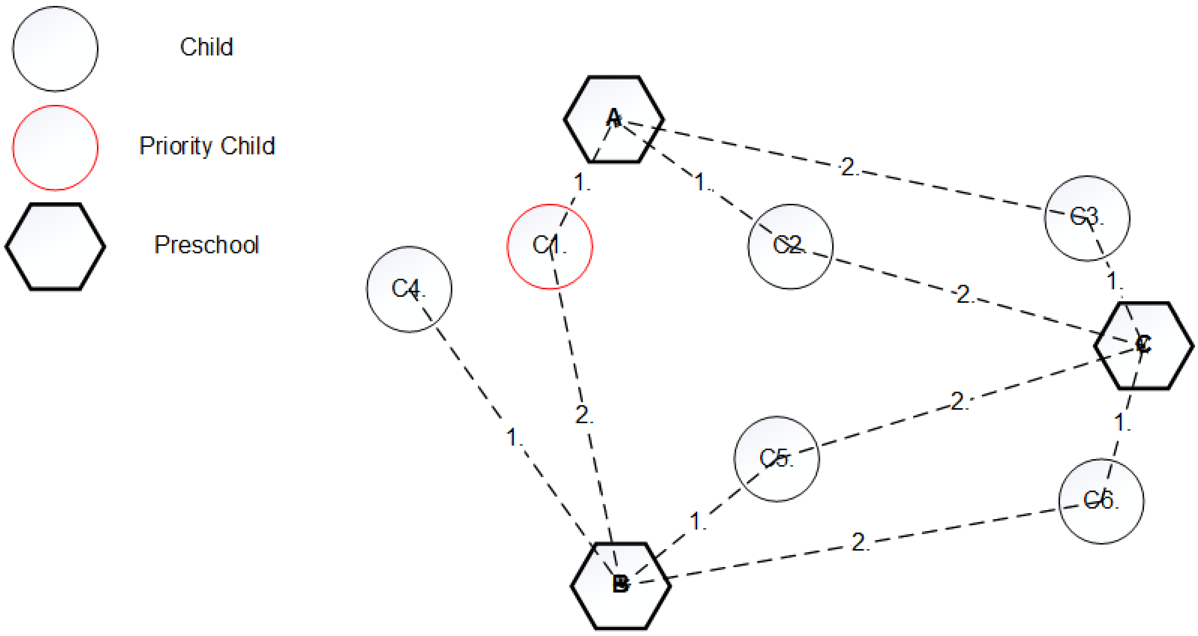

Figure 4.1: Overview of the test case. Six children waiting for assignment to preschools A, B and C. The numbers on the edges represent the choices of the parents, where 1 is the first choice as stated on their application and 2 is the second. The positioning of the children represents the distance between their home and all of the preschools.

4.3.1 Test case

In order to verify that the models are working properly, the small test case seen in Figure 4.1 was created. The case consists of six children that are to be allocated between three preschools A, B and C. It is assumed that the parents can chose between 1-2 preschools in their application. The edges between the nodes indicate the preferences of the parents where 1 is the first choice as stated on their application and 2 is the second. The positioning of the children represents the distance between their home and the preschools, which can be seen in Table 4.1. The available capacity of the preschools A, B and C is 2, 2, 1 respectively, meaning that there will be one child that will need to wait for further vacancies. Furthermore, it is assumed that the youngest child in the group has priority and should thus be guaranteed an assignment. The application data created for the children can be seen in Table 4.2

Child	Preschool A	Preschool B	Preschool C
1.	1	2	4
2.	1.5	3	2
3.	3	4.5	1
4.	2	3	5
5.	2.5	1	2
6.	4.5	3	1.5

Table 4.1: Distances to the preschools in km.

Child	Date of birth	Priority	1. choice	2. choice
1.	01.01.2012	Yes	A	B
2.	01.02.2011	No	A	C
3.	01.03.2011	No	C	A
4.	01.04.2011	No	B	
5.	01.04.2008	No	B	C
6.	01.05.2009	No	C	B

Table 4.2: Application data

Inputs

For the IP model, a utility is given for meeting the parents' requests: A utility of 10 is given for meeting the parents' first choice, a utility of 5 is given for meeting their second choice and a utility of 0 is given for preschools not chosen in their application. An $\alpha = 1$ is used as a weight factor. The weight factor means that the utility given for distances will always be considerably lower than the utility given by the parents' preferences. The emphasis is on meeting the preferences so the distances have little effect unless it is not possible to meet the requests of the parents. The case is modelled using the MPL modelling system and solved using Gurobi 5.6.0 [20]. The model has 20 constraints and 18 binary variables.

The preference list for the parents used in the SM model is given in Table 4.2. The parents only have two choices in their application. However, a complete ranking of all of the preschools is created by ranking the rest of the preschools according to the distance from the parents' home address. The choices of the parents are shown in bold in Table 4.3. The preschools are arranged in order of preference from left to right. Similarly, the preference list of the preschools can be seen in Table 4.4. Here the children are chosen according to priority and then age, where the oldest children are chosen first. In this case, the same rule applies for all of the preschools. The model is formulated and solved using MATLAB R2014a.

Child	1. choice	2. choice	3. choice
1.	**A**	**B**	C
2.	**A**	**C**	B
3.	**C**	**A**	B
4.	**B**	A	C
5.	**B**	**C**	A
6.	**C**	**B**	A

Table 4.3: Complete rank-ordered preferences of preschools made by the parents.

Preschool	1. choice	2. choice	3. choice	4. choice	5. choice	6. choice
A	1	5	6	2	3	4
B	1	5	6	2	3	4
C	1	5	6	2	3	4

Table 4.4: Complete rank-ordered preferences of the children made by the preschools.

Results

The results from the models can be seen in Table 4.5. Both models give a feasible solution to the problem. Since the youngest child (child 1) has priority, the second youngest child (child 4) will need to wait for further vacancies. Although both models give a feasible solution, their assignments are not the same. The utility coefficient is 49 for the IP model while if utility would be given for the assignment using the SM model it would accumulate to 44. The only difference between the assignments is that child 3 and child 6 have traded places.

Child	Date of birth	Priority	IP model	SM model
1.	01.01.2012	Yes	A	A
2.	01.02.2011	No	A	A
3.	01.03.2011	No	C	B
4.	01.04.2011	No	-	-
5.	01.04.2008	No	B	B
6.	01.05.2009	No	B	C

Table 4.5: Test case solution

Stability and Pareto efficiency

According to the assignment rules in Section 2.1.2, there is no neighbourhood (or sibling) priority in the assignment rules, and all the preferences for the preschools are therefore identical. It can be seen that the SM model gives a *stable* solution since any other arrangement would break the assignment rules of RVK. The result will eventually be a *serial dictatorship* [11] where the child with the highest priority chooses his/her most preferred preschool first, and then the child with the second highest priority chooses from the remaining schools, and this process continues until there is no more capacity. In these cases, the solution will not only be stable but also *Pareto efficient* since it will not be possible to improve the solution without moving one or more of the higher priority children to a preschool that is less preferred. However, schools with different priority lists might not provide a *Pareto efficient* solution as has been shown by [11]. Looking at the IP model, the changing of places is an improvement for child 3, who is younger and who receives a preschool that is rated higher on his/her parents' preference list. However, the solution is worse for child 6 since the parents of child 6 are assigned to a preschool lower on their preference list than under the *stable* assignment of the SM model. Since there now exist a child and a preschool that would prefer each other to their current matching, the solution of the IP model must be *unstable*. However, the model will provide a *Pareto efficient* solution for the same reason as mentioned for the SM model.

Strategy proof

The Gale and Shapley algorithm has been shown to be *strategy proof*. In the SM model, the best strategy for the parents is to put all of the preschools in their true order of preference on the application. However, under the IP model, there is an incentive for the parents to try to cheat the model by only giving their top preference in the application and leaving the rest of the choices blank. If some parents only put down their top choice of preschool while other parents chose two or more preschools in their application, then parent who choose only one preschool are more likely to be accepted. This is because the IP model will try to ensure that those parents receive their most preferred choice, while possibly making changes to other parents' preferences to obtain the highest utility possible. This is, however, not a problem for the SM model since it doesn't use a utility function.

Correspondence with the assignment rules of RVK

The IP model requires the SFS office to be more flexible in their policy since the assignment rules are not *strictly* followed when the model moves children between preschools. As seen in Table 4.5, this means, for example, that an older child (child 6) on a waiting list might be assigned to the second choice in his/her parents' application in order that a younger children (child 3) can be given a place at their first choice preschool. This might be considered unfair. However, the IP model is focused on achieving as many matches to the preferences of the parents as possible. By moving these children between preschools, the model is able to reach the highest possible utility and thus maximizes satisfaction with the assignment when the whole set of applicants is considered.

The SM model *strictly* follows the assignment rules in Section 2.1.2. Strict compliance with the rules means that the younger children in the assignment group are less likely to receive a preschool preferred by the parents. This is because the set of possible assignments for the younger children becomes smaller as the capacity is used up by the priority and older children in the group. For example, if the set of children considered for assignments were all born in the year 2012, then it is beneficial in the assignments to have been born early in the year. An example of a similar case can be seen in the assignment of child 3 in Table 4.5 for the SM model. Since all of the capacity has been used by the older and priority children, the third youngest child is assigned to a preschool that is not on his/her parents' preference list.

Summary

A fundamental difference between the models is that the SM model focuses on the individual while the IP model looks at the group as a whole. The IP model tries to maximize the the group's satisfaction by rearranging some of the children between the preschools in order to meet as many preferences of parents as possible. In the test case using the IP model, it is possible to assign the children in a way that meets at least some of the preferences of each set of parents, while in the SM model the parents of child 3 did not receive any of their preferences.

4.4 Limitations of the models

Model formulations

No limits were placed on the number of priority children that may be assigned to each preschool, which in reality would be necessary in order that there is no clustering of priority children at the more popular preschools. However, this can be fixed in the IP model by using additional constraints. In the case of the SM model, it would be possible to simply run the model twice, once for the priority children using adjusted preschool capacity in order that there are not too many priority children assigned to any one preschool. Another possibility would be to use the modified version of the Gale and Shapley algorithm, proposed by [11], where type specific quotas are inserted into the algorithm to maintain the right racial and ethnic balance at the schools, without the loss of *stability* or *strategy proofness*. Such a formulation could also be used to maintain a specific ratio of priority versus non-priority or gender at the preschools.

In the current formulation of the models, limitations are missing for parents who are seeking a transfer between preschools. These parents should be guaranteed to receive an assignment for their child to the preschool to which they want their child to be transferred. Another limitation is that there is no special consideration made in the formulation to address ties between children in the assignments. This means that there might be instances where children have identical applications. However, this is not considered a problem since these cases are uncommon. When such cases do arise, these children will be assigned based on the order that they appear in the data.

The IP model can only handle a limited number of inputs, since the model grows exponentially with the model size (NP-hard) and thus needs a great deal of memory and a long running time. Another problem is that constraint (4.4) ensure that an older child is assigned before a younger child. The current formulation makes a lot of unnecessary constraints, for example, if child α is older then child β and child β is older then child γ then the model also contains the fact that child α is older then child γ. The growth in the number of constraints (4.4) is therefore n^2, where n is the number of children to be assigned. It is possible to improve the model by ensuring that a child is only ahead of the child that is next in the queue for assignment and not ahead of all younger children. This modification would severely reduce the number of constraints. This does not pose a problem for the SM model since it has been shown to run in polynomial time [18].

Chapter 5

Data

This chapter will describe the data that was collected and the insight that the data gives into the preschool problem. Several inputs for the models will be listed and problems with the data collection will be discussed.

5.1 Data collection

The models in Chapter 4 use information from the parents' applications as well as information about the distances the parents need to travel to the preschools in order to make the assignments. The method of data collection is described below. The data is then used to create relevant statistics and to demonstrate that distances seem to be of importance for the parents when choosing their most preferred preschool.

5.1.1 Data from RVK

Real data was provided by the SFS office of RVK. The data was collected from the IT system of SFS, which is known as Vala. This system is used to keep track of and provide services for the users of preschools in Reykjavik. The requested data was entered into two separate Excel files. One file contains the parents' applications and the other contains information about the preschools. The information in both files is from the 1st March 2013, around the peak time of the yearly assignments.

Parents’ applications

The first document contains information from 1602 preschool applications. For each applicant, the following information was provided:

- Date of birth
- Home address
- Postal code
- Place or city
- Sequence in which the preschools were chosen in the application (does not imply ranking)
- Optional comments regarding preferences
- Type of application (transfer or general)
- Type of priority
- Whether or not an application for priority was accepted (yes or no)
- Indication if a child has siblings in other preschools

Preschool information

The second document contains information on 65 preschools. For each preschool, the following information was provided:

- The location of the preschool
- Postal code
- The total capacity of the preschool
- How many children are leaving for elementary school
- How many children are being transferred to other preschools
- The number of children that can be assigned to the preschool

5.1.2 Distances

Distances between children and preschools were collected to form a distance matrix using Google API services, HTML and JavaScript. The data was entered into a single Excel document. The matrix contains the distance in km, between each child's home address and each of the preschools. The Google Distance Matrix API service provides travel distance for a matrix of origins and destinations using the road network [1].

5.1.3 Statistics

In the dataset that we received from RVK, there are 1602 children, 136 of which have an accepted priority application. There are 137 children living outside of Reykjavik. All of the children were born between 6th February 2007 and 30th May 2012, with most children (around 70%) born in the year 2011. The postal codes with the most applicants are 112, 110, and 105 respectively.

Out of the total number of applications, around 25% are transfer applications, and the rest are general applications. As mentioned in Chapter 2, this parameter can be useful for assess the performance of the assignment method in use.

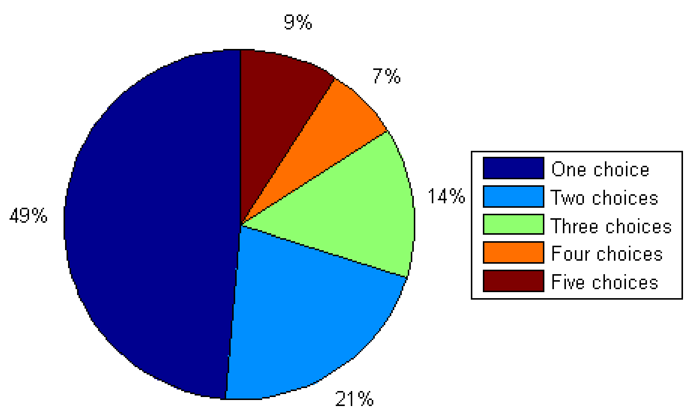

Figure 5.1: How many choices are used by the parents in the applications.

The parents are not using all of their possible choices of preschools in the applications. Although all of the parents give one or more choices, Figure 5.1 shows that around 49% of the parents only give a first choice in their applications and that only around 9% complete all five choices. RVK therefore has the opportunity to increase the information available during the assignment by requiring parents to complete all five choices.

The varying demand for and capacity of the preschools is illustrated in Figure 5.2. The preschools have been allocated ID numbers in order to preserve their anonymity. It is clear that it is not possible to give every parent their first choice of preschool. The preschools chosen most often in the applications are preschools with ID numbers 12 and 48. These preschools have 18 and 53 places available respectively, however, they are chosen as a first choice in the applications 50 and 63 times. There are also instances where the opposite occurs, i.e., where there is more capacity than demand. For example, the preschool with ID number 26 has 59 places available, but is only chosen 42 times as a first choice.

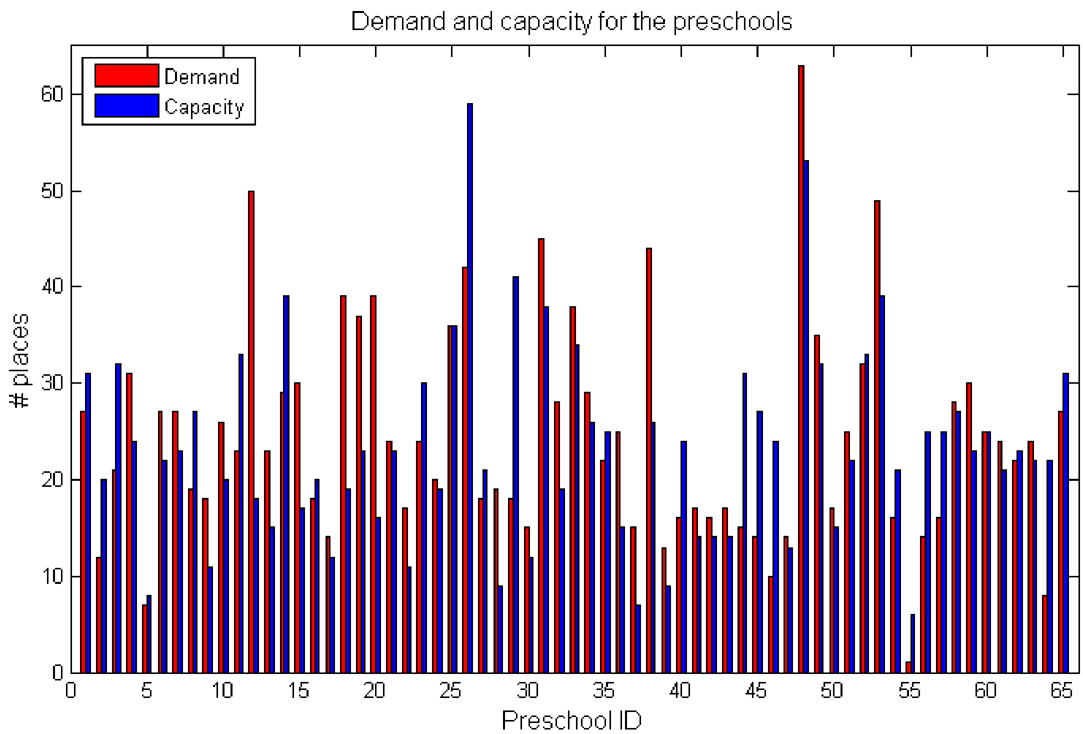

Figure 5.2: The first choices of parents in their applications and the available capacity within each preschool.

Levels of demand vary greatly between the preschools. Figure 5.3 shows the total number of times that a preschool is chosen in the applications as one of the five choices, sorted by the number of applicants from low to high. The least frequently chosen preschool is chosen 5 times, while the most frequently chosen preschool is chosen 90 times. The average number of times that a preschool is chosen in the applications is around 50.

The total available capacity of the preschools amounts to 1516 places, meaning that 86 children will need to wait for further vacancies. Figure 5.4 shows the capacity of the

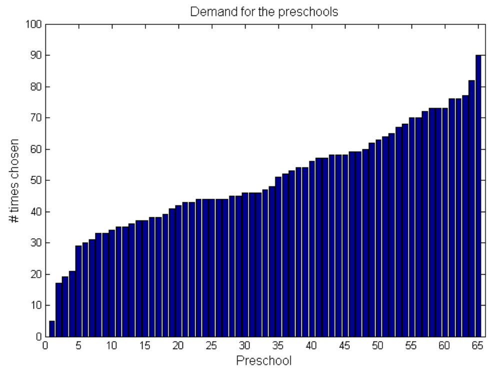

Figure 5.3: Total number of times that each preschool is chosen in the applications.

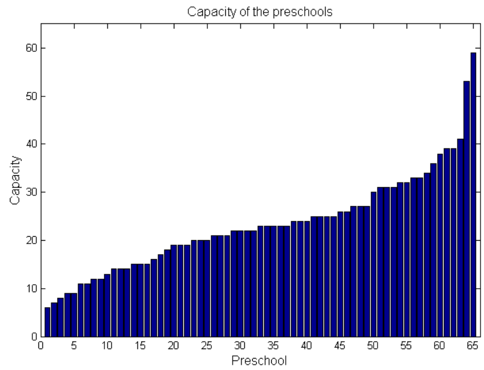

Figure 5.4: Capacity of the preschools.

preschools, sorted by the number of available places within each preschool from low to high. The capacity is dependent on the amount of transfers, the size of the preschools, and the number of children leaving for elementary school.

Figure 5.5 shows the distance to the first choice of the 1405 parents living in Reykjavik. The average distance is about 1.7 km. Furthermore, about 78% of the parents chose a preschool that is less then 3 km from their home as a first choice. Therefore, there is a clear incentive to minimize the distances travelled by the parents in order to guarantee the best possible service.

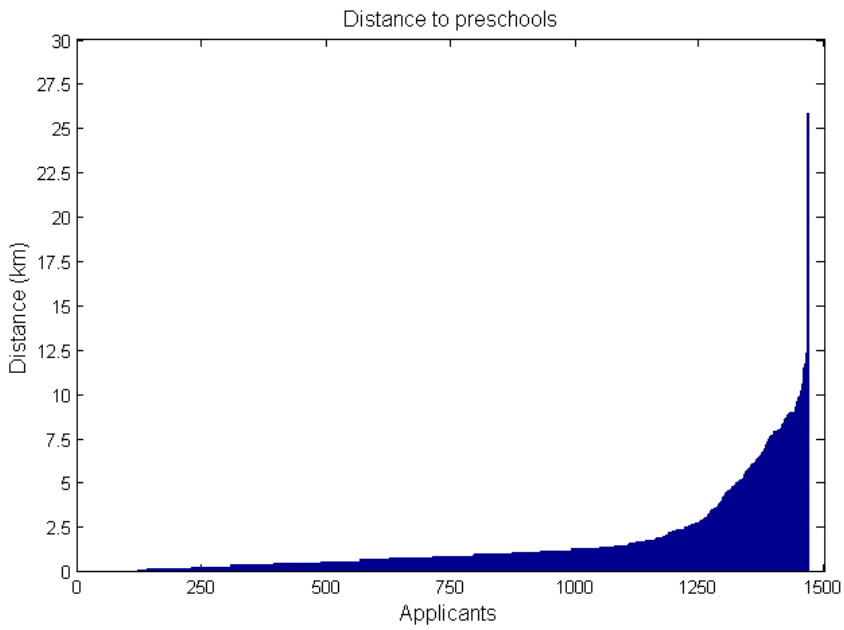

Figure 5.5: Distance between the homes of the 1405 parents living in Reykjavik and their first choice of preschool.

5.2 Data handling and proposed inputs

To create the inputs for the models, the Excel documents were read into a series of simple C++ programs created in Visual Studio. These programs put the data in a desirable format for the models and outputted them as textfiles.

As shown in Chapter 4, the models use different approaches to solve the assignment problem. The IP model uses a utility function to give a utility based on the parents' preferences and the distance that the parents are required to travel. The SM model differs in that it does not use a utility function, but rather uses complete rank-ordered preference lists to make the assignments.

In order to make the inputs for the models, the following steps were performed:

i. There were a number of children living outside of the municipality at the time of the data collection. Some assumptions were made on how to handle these cases since they do not have a home address (origin in the distance matrix) in Reykjavik. It was decided that all children living more than 50 km from Reykjavik would be given a constant distance of 50 km to all of the preschools. This simple rule means that we can avoid that these children compete with other children with known permanent addresses in the Reykjavik area, and also avoid discriminating against the children living more than 50 km away.

Integer programming model

ii. **Distance matrix.** The collected distances are used to form a distance matrix, which is used in Equation 4.1. The distance matrix gets converted through $\alpha/distance_{ij}$ to give utility based on the distance that the parents need to travel to all of the preschools. Longer distances will result in lower utility. Increasing α will add to the importance of the distances in the assignments.

iii. **Preference matrix.** The preference matrix seen in Equation 4.1 gives a utility based on the preferences of the parents. Each row of the matrix contains the parents' preferences with added utility depending on the importance of the preschools represented in each column. A higher utility means greater importance.

iv. **Date of birth.** The date of birth is changed from day/month/year to year/month/-day and represents a number. Hence, the highest numbers represent the youngest children in the group while the lowest numbers represent the oldest children. This is convenient when constructing the age constraint seen in Equation (4.4), which determines the order in which the children have to be assigned to the preschools.

Stable marriage model

The SM model uses two preference matrices to make the assignments, one for the children and one for the preschools. However, two further preference matrices for the preschools are given below. These matrices demonstrate the use of the SM model under the current assignment rules and under an assumed additional rule of *neighbourhood priority*. Neighbourhood priority means that children living in the neighbourhood of a preschool have higher priority than children living further away. The steps taken to formulate the matrices are as follows:

v. **Parent preference matrix.** To create a complete rank-ordered preference list for the parents, preferences and distances are used. The top ranking is composed of the chosen preschools from the parents' applications and to complete the preference list, distances to each of the preschools are used, ranking the shortest distance to the preschools the highest. Some special consideration had to be given in the matrix for children living more then 50 km away from the preschools. Since the distance is constant in these cases, the ranking of the preschools is randomized. Without the randomization, the distance ranking for these children would be identical, ultimately making some of the preschools very popular for no particular reason.

vi. **RVK preschool preference matrix.** To create a complete rank-ordered preference list for the preschools, the assignment rules in Section 2.1.2 are used. The children are ranked based on their priority, with children with accepted priority applications ranked highest and then ranked in descending order according to their age. Then the children without a priority application are added to the list, again ranked in descending order according to their age. In the matrix, all of the preference lists for the preschools are identical.

vii. **Neighbourhood preschool preference matrix.** To create a complete rank-ordered list for the preschools using neighbourhood priority, distances are used. With a 1 km interval the children are ranked based on priority and age. Each preschool now has a unique preference list of children. This formulation breaks the assignment rules but will provide an interesting comparison.

5.3 Problems with the data and the data collection

There were some discrepancies in the dataset. The first problem was that some preschool names were missing in the two provided documents and the capacity of some of the preschools was therefore not known. In these instances, those preschools were ignored, meaning that not all of the preferences stated by the applicants have been used. This happens in 32 out of the 1602 applications. This will lower the number of preference matchings when using the models as well as these preschools are not included in the above statistic section.

Another problem was that certain preschools appeared under one name, but in reality formed a network of 2-3 preschools. In these instances, one of the preschools was chosen at random and used as a destination in the distance matrix.

In some places, numbering was missing from the data where the parents state their preschool preferences, and irrelevant data sometimes appeared. This was accounted for when transforming the data into inputs for the models.

The distance matrix data was collected, using HTML and JavaScript, and the results were entered into an Excel document. This is not an ideal way of collecting the data since there is a high risk of data discrepancy and the process is also time-consuming. Furthermore, Google imposes some limitations on the API services. The number of origins time the destination determine the number of elements that can be retrieve per 24 hour period. A business registered for the service can retrieve up to 100,000 elements.

The rank-ordered parents' preferences are not necessarily their real preferences, but rather the sequence in which they chose the preschools in their application. The result might not therefore reflect the true preferences of the parents.

Although the data contains some discrepancies, it is sufficient for the task in hand. We are constructing two mathematical models to see if they can be used in the assignment process in RVK by estimating their benefits and comparing them with the current assignment method. The data is sufficiently accurate to demonstrate the functioning of the models and to compare the results.

Chapter 6

Results

This paper introduces two possible new methods of helping with preschool assignments during the peak in the assignment period in RVK. As can be seen in Chapter 4, the models have a different way of solving the assignment problem and may provide different results.

6.1 Model results

In this chapter, the data presented in Chapter 5 is used in the models. Using real data will give further insight into how these models behave and will ultimately demonstrate whether these models can be used in the assignment process in RVK. This chapter is split into two parts.

In the first part, the results from using the provided data will be presented. The SM model showed promising results at the end of Chapter 4. It is therefore appropriate to test the SM model using two different scenarios for the model in order to further look at how the model could be adjusted if the assignment rules were ever to change. The two different scenarios will be referred to as SM RVK, which is the same formulation as used in the test case in Chapter 4, and SM neighbourhood, which assumes that neighbourhood priority is added to the assignment rules. Neighbourhood priority, for example, has traditionally been used in the United States when it comes to assignments to public schools [11]. The results of each model are presented separately.

In the second part, the models are compared to each other and to the current assignment process, and a visualization of several of the more important results will be given. The comparison will reveal some of the most relevant advantages and disadvantages of each method. Finally, a summary of the most relevant discoveries will be presented.

6.1.1 Integer programming model

Inputs and optimality

For the IP model, a distance matrix of size 1602 times 65 and a preference matrix of size 1602 times 65 are created, where the 1602 is the number of children and 65 the number of preschools in the assignments. The preferences are given a utility whereby the parents' first choice receives a utility of 1000, their second choice a utility of 500, their third choice a utility of 100, their fourth choice a utility of 50, and their fifth choice a utility of 25. Preschools not chosen in the applications are given a default utility of 0. Additionally, utility is also given based on distance the parents need to travel from their home to each of the preschools under consideration, where $\alpha = 1$ is used as a weight factor. There is no special reason for the figures chosen to represent utility for the preferences other that it is clearly desirable that the utility given for the preferences is higher than to the utility give for the distances.

Due to the size of the IP model, certain measures needed to be taken. The model was solved using Gurobi 5.6.0 on a computer cluster, where each node has 48 GB of memory. Before the model was solved, it was written to a MPS file. Since Gurobi minimizes these files by default, the utility function was multiplied by -1 before the model was written to the file, ultimately giving the same result as a maximization. The utility function in our solution is -1.45e+06. The number of constraints are 1,158,605 and number of variables are 104,130.

Since the problem is NP-hard, we might not be able to find an optimal solution. It is, however, highly desirable that the solution from the IP model is as close to optimality as possible. To help determine whether the solution found is not just a feasible solution but also the optimal solution, a parameter known as the MIP gap is used. The formula for the relative MIP gap can be seen in Equation (6.1). After running for 11,000 seconds, the IP model was stopped due to hardware constraints. The result was a MIP gap of 0.001. This means that our result will in the worst case be 0.1% from the optimal solution. When solving a model, most of the running time often goes into proving that the current solution is indeed the optimal solution. After around 17 minutes the MIP gap was 1% from the optimal solution.

$$\text{MIP gap} = \frac{\|\text{upper bound} - \text{lower bound}\|}{\text{lower bound}} \tag{6.1}$$

Summary of the IP solutions

The main results for the IP model are summarized in the second column of Table 6.1. The total number of children in the assignment process was 1602, of which 1516 received an assignment. 86 children did not receive an assignment due to capacity constraints.

Out of the total number of applications, around 79% of the parents received the top choice from their application and about 88% of the parents received one of their five preferences. Parents' first choices were met 1261 times, their second choices 111 times, their third choices 21 times, their fourth choices 12 times and their fifth choices 4 times. As shown in Figure 5.1 in Chapter 5, parents are not using all of their choices in the applications, and the matches will therefore inevitably be low in the case of the last preferences of the parents.

About 7% (107) of the children were assigned to a preschool that was not on their parents' preference list. The dates of birth of these children are between 29.02.08 and 11.02.12. The age range is wide and almost spans the full age range of all of the children under consideration. Looking at the case of a child born on 29.02.08, his/her parents only choose one preschool in their application, that has the capacity for 26 children. According to the assignment rules, this child should be the 23rd child in the queue for the assignments. The child should therefore receive the preferred choice of his/her parents but does not. Therefore, the age rule is not *strictly* being followed, i.e., the oldest children in the group are not necessarily receiving their preferred choice before the younger children, as might be implied by the rules in Section 2.1.2.

The average distance travelled by the parents was around 4.1 km. This number includes parents living outside of the municipality, and the average distance is therefore higher than if only the parents within the municipality were considered. The average distance travelled by the 107 parents that did not receive any of their preferences was also around 4.1 km, 13 of these 107 parents travel 50 km or more to the preschool since they did not have a registered address in Reykjavik.

86 children were rejected in the assignment due to capacity constraints. The dates of birth of these children are between 17.01.12 and 30.05.12. These children were the youngest in the group besides the 11 priority children that were born on this interval but were guaranteed an assignment. Furthermore, all of the priority children receive an assignment.

This means that the model is ensuring that the priority children and the older children in the group receive an assignment.

The results for the IP model show that the age rule in section 2.1.2 is not *strictly* being followed. However, the model ensures that the priority children and the older children in the group receive an assignment. A visualization of the results for one preschool which has capacity for 18 children can bee seen in Figure 6.1.

6.1.2 Stable marriage model

Two different scenarios are tested for the SM model. We refer to these scenarios as SM RVK and SM neighbourhood. In the SM RVK scenario, the assignment rules of RVK are *strictly* adhered to. However, SM neighbourhood scenario gives additional priority to the children based on where they live, where children living in the neighbourhood of a preschool have higher priority then children living further away. This additional assumption gives different priority orderings of the children in each preschool. The SM model were constructed and solved using MATLAB R2014a on a 1.6 GHz Dell Vostro machine with 2 GB of memory running a 64 bit Windows 8.

SM RVK

In the SM RVK scenario, it is most desirable that the parents receive the preschool that they want. If it is not possible to meet these requirements, then it is preferable that the preschool is as short a distance as possible from the parents' home address. The preschools are considered to be indifferent as to which children they receive as long as the assignment rules set by RVK are followed. This scenario uses a **parent preference matrix** of size 1602 times 65 and a **RVK preschool preference matrix** of size 65 times 1602, where the 1602 is the number of children and the 65 the number of preschools in the assignment.

The main results for SM RVK are summarized in the third column of Table 6.1. Of the 1602 children in the assignment process, 1516 received an assignment and 86 were rejected due to capacity constraints.

Out of the total number of applications, about 73% of the parents that received the top choice from their application, and about 80% of the parents received one of their five preferences. The parents' first choices were met 1175 times, their second choices 79 times, their third choices 24 times, their fourth choices 6 times and their fifth choices 2 times.

About 14% (230) of the children were assigned to a preschool that was not on their parents' preference list. The dates of birth of these children are between 22.02.11 and 16.01.12. The age range is narrower than for the IP model. These children are lower in the priority rankings (younger) than the children that receive their preferences and higher in the priority rankings (older) than the children that are rejected. The reason is that the SM RVK essentially allows the highest priority child to choose his/her most preferred preschool first, and then the child with the second highest priority is allowed to choose next etc. until there is no more capacity within the preschools. This means that the model is behaving in accordance to the *strict* interpretation of the the age rule set by RVK.

The average distance travelled by the parents was around 4.3 km. The average distance travelled by the 230 parents that do not receive any of their preferences was around 4.8 km. 12 out of these 230 parents travel 50 km or more to the preschool since they do not have a registered address in Reykjavik, which pushes up the average distance.

86 children were rejected in the assignment due to capacity constraints. The same children were rejected in the IP model and the SM RVK. The dates of birth of these children are between 17.01.12 and 30.05.12. These children are the youngest children in the group that do not have priority. All of the priority children receive an assignment.

The results show that all of the rules in Section 2.1.2 are followed. The children that have the highest priority according to the rules are able to chose their preferred school in accordance with *strict* following of the assignment rules. Furthermore, the children rejected in the assignments are the youngest children in the group that do not have priority. A visualization of the results for one preschool which has capacity for 18 children can bee seen in Figure 6.2.

SM neighbourhood

In the SM neighbourhood scenario, it is most desirable that the parents receive the preschool that they want, but now the preschools also have preferences over the children based on where they live. An additional *neighbourhood priority* rule is added to the RVK assignment rules. The rule gives children living in each neighbourhood priority to get into preschool in their neighbourhood over children living further away. This is done by ranking the children based on priority and age with 1 km intervals, ranking the children living the shortest distance from each preschool the highest. This will create a unique preference list for each of the preschools. This scenario uses the **parent preference matrix** of size 1602 times 65 and the **neighbourhood preschool preference matrix** of size 65 times 1602, where 1602 is the number of children and 65 the number of preschools in the assignments.

The main results for the SM neighbourhood are summarized in the fourth column of Table 6.1. Of the 1602 children in the assignment process, 1516 received an assignment and 86 were rejected due to capacity constraints.

Out of the total number of applications, about 60% of the parents received the top choice from their application and about 70% of the parents received one of the preschools from their application. Parents' first choices were met 962 times, their second choices 112 times, their third choices 25 times, their fourth choices 15 times and their fifth choices 6 times.

About 25% (396) of the children were assigned to a preschool that was not on their parents' preference list. The dates of birth of these children are between 12.04.2007 and 30.05.2012. The age range is wide and almost spans the full age range of all the children under consideration. This is essentially because now the children have higher priority of getting into preschool that they live close to than preschools further away. Therefore, the *strict* interpretation of the age rule is broken.

The average distance travelled by the parents was around 1.2 km. The average distance travelled by the 396 parents that do not receive any of their preferences was around 3.4 km. All of the children living 50 km or farther away from Reykjavik were rejected.

86 children were rejected in the assignment due to capacity constraints. The dates of birth of these children are between 23.01.2008 and 29.02.2012. The rejected children are no longer the youngest children in the group that do not have priority, but are rather the children living outside of the municipality. Furthermore, not all of the priority children receive an assignment and therefore the assignment rules are broken.

The result show that all of the rules in Section 2.1.2 are broken. The children are not assigned to the preschools in accordance to the current assignment rules, rejection of children is now based on distance rather than age and not all of the priority children receive an assignment. However, this was to be expected since the scenario assumes an additional assignment rule and is purely for demonstrative purposes. A visualization of the results for one preschool which has capacity for 18 children can bee seen in Figure 6.3.

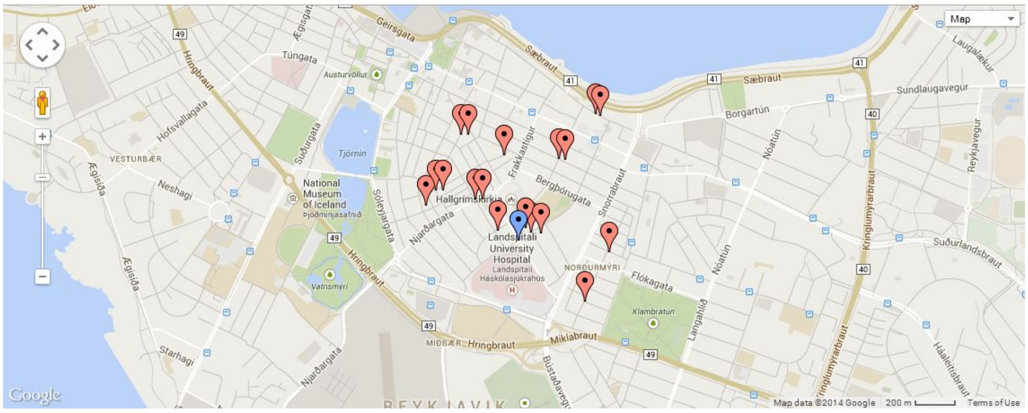

Figure 6.1: Visualization of the results from the IP model for one preschool who has capacity for 18 children during the assignments. The Figure has been created using the Batchgeo mapping tool (http://batchgeo.com).

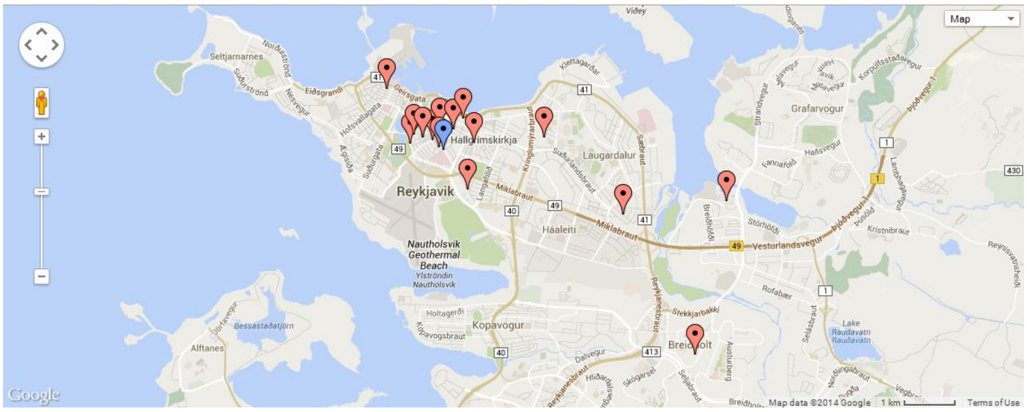

Figure 6.2: Visualization of the results from SM RVK for one preschool who has capacity for 18 children during the assignments. The Figure has been created using the Batchgeo mapping tool (http://batchgeo.com).

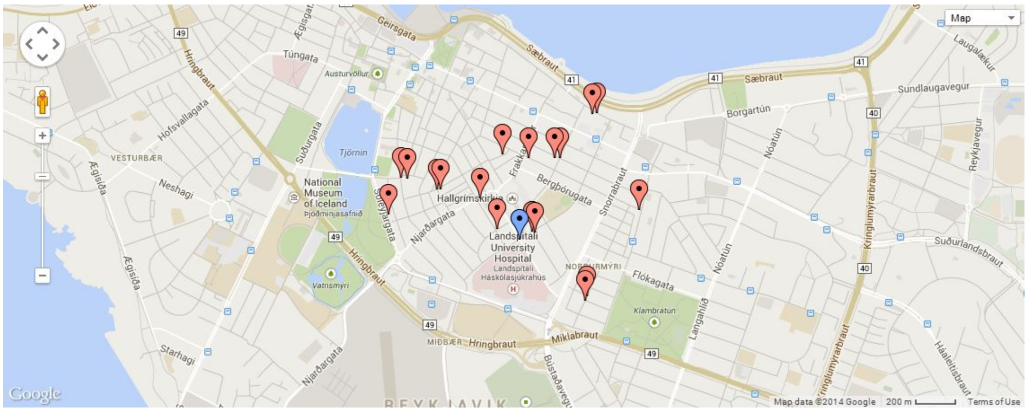

Figure 6.3: Visualization of the results from SM neighbourhood for one preschool who has capacity for 18 children during the assignments. The Figure has been created using the Batchgeo mapping tool (http://batchgeo.com).

Results from the models	IP model	SM RVK	SM neighbourhood
Parents’ top choice met [%]*	~79 (83)	~73 (78)	~60 (64)
Children accepted to preferred school [%]*	~88 (93)	~80 (85)	~70 (74)
Parents’ request not met [%]*	~7 (7)	~14 (15)	~25 (26)
Number of rejected children	86	86	86
Assigned priority children	136	136	131
Average distance travelled [km]	~4.1	~4.3	~1.2
Running time [s]	~11,000	~23	~23
Oldest child rejected	17.01.12	17.01.12	23.01.08
Youngest child rejected	30.05.12	30.05.12	29.02.12

Table 6.1: A summary of the main results from the models. (*) The proportions are calculated using different denominators. The first number uses the total number of children in the applications (1602), while the second only uses the total number of accepted children in the assignments (1516).

6.1.3 Comparison of the models and current practice

In this section, the models and the current method (CM) being used by the SFS office are compared. The models assume a different policy than is currently being used by RVK. The models assume that the preferences of the applicants are weighed (rank-ordered), however, under current policy all of the preschools in a parent’s application have equal value unless the parents have communicated any other special requests. This makes direct comparison of the results from the methods of little significance. The comparison will therefore be based on the established parameters in Chapter 3, which are the most commonly used parameters for comparison of school choice mechanisms, as well as some additional parameters. Table 6.2 shows a given grade for the performance of each method. A grade of ‘A’ is given to the methods that more than fulfil all requirements for a characteristic, the grade ‘B’ is given to methods that mostly fulfil the requirements, perhaps with some minor exceptions, or barely fulfil the requirements but include other features that make up for their flaws. A grade of ‘C’ is given to methods that only barely meet the minimum requirements for a characteristic or have obvious flaws.

Characteristics	IP model	SM RVK	SM neighbourhood	CM
Meet preferences	A			B
Adhere to a. rules of RVK	C	A		C
Short distance			A	
Stable		A	A	
Pareto efficient	A	A		
Strategy proof	C	A	C	C
Transparent		A	A	
Short solution time		A	A	

Table 6.2: Overview of how the methods fulfil various characteristics. (A) More than fulfils all the requirements. (B) Mostly fulfils the requirements or barely fulfils the requirements but includes other redeeming features. (C) Barely meets the minimum requirements or has obvious flaws.

Preferences

Out of the models that were tested, the IP model is most often (about 88% of the time) able to assign children to one of the preschools preferred by the parents. The SM RVK and SM neighbourhood models are able to assign children to preferred preschools about 80% and 70% of the time respectively, when the total number of children taking part in the assignments is considered. A visualization of the comparison can be seen in Figure 6.4. This figure shows which of the parents preferences are met, how many children are assigned based on distance, and how many are eventually rejected. The IP model gives an optimal number of matchings to the preferences since the MIP gap is insignificantly small. The IP model is able to give 86 more matchings to the preferences than the SM RVK, which comes second. This is because the IP model is not constrained like the SM RVK model (which always lets the priority children choose the preschools that they want first), rather the assignments are based on attempting to meet the maximum number of preferences. SM neighbourhood delivers the worst result of the models since it assumes an additional neighbourhood priority rule, which reduces the parents' chances of getting into preschools that are outside of their own neighbourhood.

The CM is most able to take into account any special request besides the preference ordering of the preschools in the applications. For example, each priority child will have special needs that might best be met in certain preschools and should be taken into consideration in the assignments. Out of the suggested methods, the CM would handle such instances in the best way. However, if only the number of preference matchings are considered and we assume that the CM (similarly to the IP model) would not strictly follow the rules, it would be hard to reach the same optimal solution using only heuristics. This is because the manner in which a solution is reached can be thought of as a complex puzzle which is best solved using a computer. Therefore, the IP model is the best way to meet as many of the parents preferences as possible and to thus maximize satisfaction with the assignments when the whole set of applicants is considered.

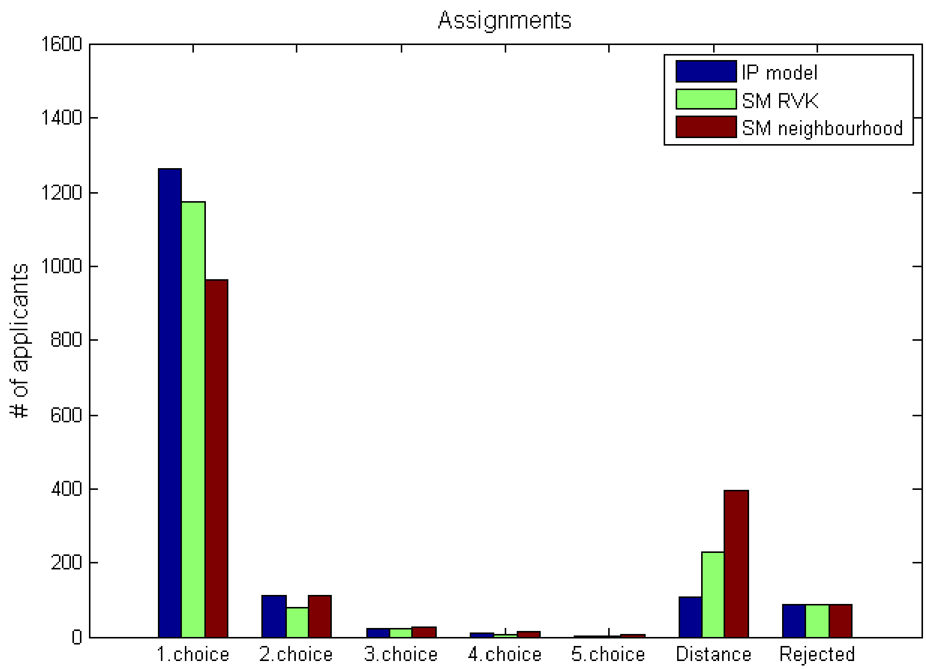

Figure 6.4: Visualization of the assignment of 1602 children. Here we see the number of children accepted to a preferred preschool, the number of children assigned based on distance and the number of children rejected during the assignment process.

Adherence to the assignment rules of RVK

The SM neighbourhood demonstrates how the SM model functions when an additional assignment rule is employed. As a result, all of the current rules of RVK are broken.

In the test case in Chapter 4, it was demonstrated that the SM model (SM RVK) and IP model interpret the assignment rules of RVK in different ways. The *strict* following of the assignment rules, as in the case of SM RVK is essentially a *serial dictatorship* where the child with the highest priority chooses his/her most preferred preschool, then the child with the second highest priority chooses his/her most preferred preschool from the remaining schools, and this process continues until there is no more capacity. Figure 6.5 shows how the models assign the 136 priority children that are present in the dataset. Under the SM RVK, it is almost guaranteed that the parents of the priority children will receive their first choice of preschool from their application. The two cases that do not receive their first choice only do not receive their most preferred preschool because of the data discrepancy discussed in Chapter 5. SM RVK also makes sure that the children rejected due to capacity constraints are the youngest children that do not have an accepted priority application. This model therefore fulfils all the assignment rules set by RVK.

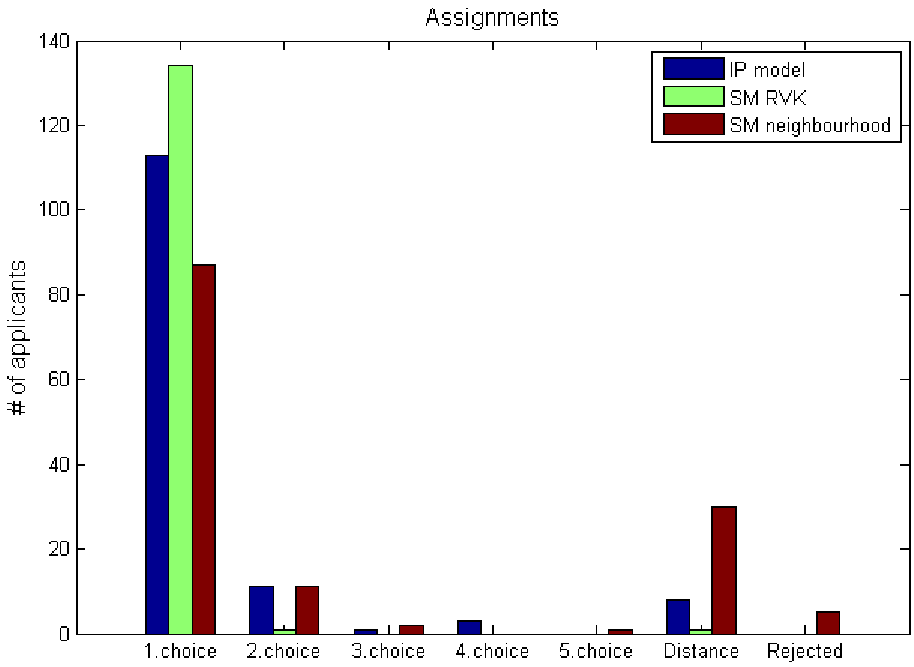

Figure 6.5: Visualization of the assignment of 136 priority children. Here we see the number of children accepted to a preferred preschool, the number of children assigned based on distance and the number of children rejected during the assignment process. In SM RVK, two children do not receive their 1st choice due to data discrepancy.

In Chapter 4, it was demonstrated that the *strict* following of the rules can be considered to be unfair towards the younger children in the assignment group. The younger children have lower chances of receiving a preferred preschool under an arrangement such as the *serial dictatorship*. This is further verified using the provided dataset on the SM RVK, since the children seen in Figure 6.4 assigned using distances are the youngest children in the group who are not rejected.

In Chapter 4, it was demonstrated that the IP model does not *strictly* follow the assignment rules. The model does not give the children with higher priority the right to chose their preschool, rather the choice depends upon attempting to allocate the children between the schools such that as many children as possible receive one of their parents' preferred preschools. As shown in the results for the IP model, this means that children with lower priority (younger children) might be assigned before children with higher priority (older children). Furthermore, it is not even guaranteed that the older children will receive one of the preschools preferred by their parents. The IP model, however, rejects the same children as the SM RVK, which gives further verification that both of the models are rejecting the youngest children of the group who do not have priority.

The CM might not *strictly* follow the assignment rules as presented in Section 2.1.2 since there is always the possibility that children will be assigned to a preschool that is less preferred by the parents than other preschools that still have available capacity.

Distance

Out of the models, the SM neighbourhood has the lowest average distance travelled (about 1.2 km), while the IP model and the SM RVK have an average distance of about 4.1 km and 4.3 km respectively. SM neighbourhood even has a lower average distance travelled then the average distance to the first choice of parents that are living in Reykjavik as shown in Chapter 5, which is about 1.7 km. A visualization of the comparison can be seen in Figures 6.6, 6.7 and 6.8. Here, the distances that the applicants would have to travel using each assignment method are sorted from shortest to longest. There were number of children living outside the municipality at the time of the data collection and the assumption was made that children living 50 km or more from the city would have a constant distance of 50 km to all the preschools. This was done in order to avoid that these children compete with other children with known permanent addresses in the Reykjavik area, and also to avoid discriminating against the children living more than 50 km away.

All of the preschools in the SM neighbourhood scenario have a unique rank-ordered preference list of the children based on how close they live to the preschool. Each preschool therefore receives, for the most part, children from the neighbourhood. Under SM neighbourhood, the rejected children are only those children that live outside of the municipality, which explains the difference between Figure 6.8 and Figures 6.6 and 6.7. Furthermore, the figures show that more parents would be travelling shorter distances than if the SM RVK model or IP model were used.

The CM does not use distance measurements, however, estimation of distance travelled is used to help with the assignments in some cases. It can therefore be assumed that the SM neighbourhood, which assigns children based on measured distances, is the best method for ensuring that the parents travel as short a distance as possible.

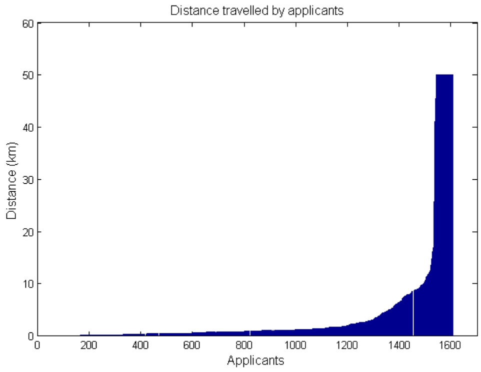

Figure 6.6: Distance travelled by parents using IP model

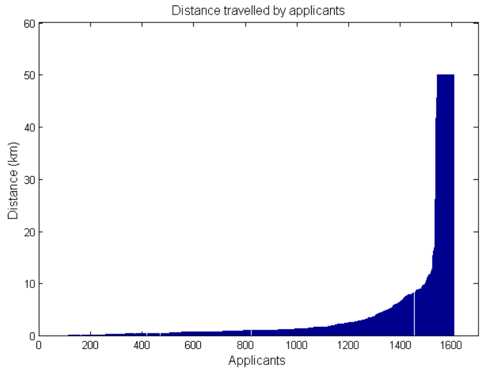

Figure 6.7: Distance travelled by parents using SM RVK

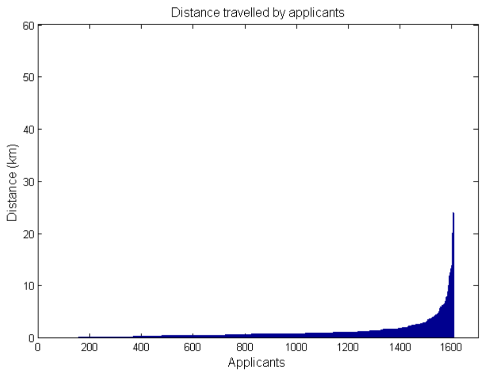

Figure 6.8: Distance travelled by parents using SM neighbourhood

Stability and Pareto efficiency

We now look at the methods when compared based on how they perform with regards to *stability* and *Pareto efficiency*. Both parameters are desired qualities that are commonly used when assessing the performance of school choice mechanisms.

In the test case presented in Chapter 4, the SM model (SM RVK) was demonstrated to be *stable*. In fact, the SM model will always provide a stable solution, which means that the SM neighbourhood will also be stable. SM RVK, which uses identical preference lists for all of the preschools, will furthermore always provide a solution that is *Pareto efficient*. This is because it is not possible to improve the solution without allocating some of the higher priority children to a preschool that is less preferred by the parents. This is, however, not necessarily true in the case of SM neighbourhood, which has a unique preference list for each of the preschools.

In Chapter 4, it was also demonstrated that the solution achieved by using the IP model might not be *stable*. The possibility exists that the children will be assigned in such a way that there exist a child and a preschool that would prefer each other to their current matching. The solution will, however, be *Pareto efficient* for the same reason as for the SM RVK.

By applying the concept of *stability* to the CM, it can be shown that this method might provide an *unstable* solution. Since rank-ordered preferences are not used in the assignments, the true preferences of parents might not be known. This creates the possibility that a child might prefer a different preschool to his/her current matching and that a preschool might prefer a different child based on the assignment rules, and thus the solution might be *unstable*. For the same reason, the solution might be *Pareto inefficient*, since there might be an opportunity to improve the solution for one or more of the children without disadvantaging the other children.

Strategy proof

As introduced at the end of Chapter 3, SM RVK will always be *strategy proof* since there is no incentive for participants to hide their true preferences. Parents cannot increase their probability of a favourable assignment by deploying any kind of strategy. Truthful expression of the preferences will always be the best way of obtaining the preferred preschool.

Should a model like the SM neighbourhood be used, truthful expression of preferences is always the dominant strategy. However, it is always possible that parents might register their home address closer to the preferred preschool in order to have higher priority for a certain preschool.

Under the IP model, there is an incentive for parents to try to cheat the model by only giving their top preference in their application and leaving the rest of the choices blank. This might tilt the odds in favour of these parents when using the the IP model since this model will always try maximize the total utility for the assignments. One way to avoid this manipulation would be to force the parents to complete all five choices of preschools in their applications in order that such strategies cannot be employed.

The CM complies with the definition of being *strategy proof* in that there is no incentive for parents to try to get a better assignment by falsifying their true preferences. However, there is an incentive for parents to try to increase their probability of assignment by contacting the preschool administrators or the SFS office. This surely has to be considered as a strategy, and it would therefore be incorrect to say that the method is entirely free from strategies.

It should be noted that there is no gain to be had from accepting a less preferred preschool and then attempting to transfer to a more preferred preschool. The transfers follow the same rules as first time applications, meaning that they can have an undetermined waiting time, and such strategies are therefore not beneficial.

Transparency

The assignments made by SM RVK and SM neigbourhood can be considered transparent and easily justified. The SM RVK is essentially a *serial dictatorship*, which makes it easy to understand how the children are assigned and why some children receive the preferred choices of the parents while others do not.

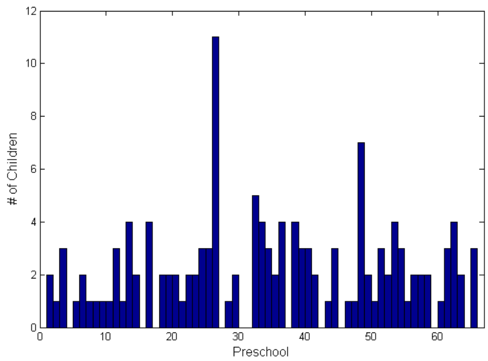

Figure 6.9: The preschools that the 136 priority children are assigned to using the IP model.

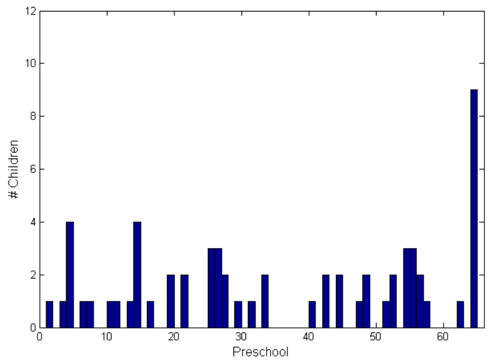

Figure 6.10: The preschools that the 69 children living 50 km or more away from Reykjavik are assigned to using the IP model.

In the case of the IP model, the assignments are less clear, since we do not know which children have been moved between preschools in order to make room for other children. We only know that the utility for the assignment has been maximized and therefore that these movements have taken place. Figures 6.9 and 6.10 show how the IP model handles two different groups of children in the assignment process. In both figures, there seems to be some clustering of children in certain preschools. Figure 6.9 only shows the preschools that the 136 priority children are assigned to using the IP model. In this case, the clustering of the children in the preschool with ID 26 can somewhat be explained by the fact that this preschool has the most available capacity during the assignments and is chosen most often by the parents of the priority children. Figure 6.10 shows the assignment of the 69 children living 50 km or more away from Reykjavik. In this case, there is no clear explanation as to why 9 of these children are clustered in the preschool with ID 64. The fact that it is difficult to definitively explain how the solution was obtained by the IP model makes it less transparent than the SM model.

Using the CM, apparently there can be an unequal amount of information available about each child at the time of the assignments. It is therefore difficult to claim that such a method is entirely transparent. Furthermore, the reasons for which a child is assigned to a certain preschool are not recorded. The solutions given by the SM models provide this information.

Solution time

The solution time for the SM RVK and SM neighbourhood is around 23 seconds. The solution time for the IP model is around 3 hours. The CM requires several meetings that take around 1.5 hours each in order to complete the assignments. The SM model therefore has the shortest solution time.

Practicality and cost

The use of mathematical models for the assignment process means that fewer employees would be needed to complete the assignments. This would save some time for the employees, making it possible for them to attend to other tasks. Furthermore, if meeting times are reduced, then costs will be reduced for RVK. Fewer employees would be required in order to complete the preparation work for the assignments, which would add to the time and cost savings. There would also be less stress on preschool administrators when dealing with parents who want their children to attend their preschool. However, if such models were introduced, then employees would need to be trained to use the models in the correct manner, as well as further development is needed which would increase costs. The largest cost that is entirely independent of the methods used for the assignment is the cost of operating the IT system that keeps track of all of the information needed for the assignments.

Besides the potential cost savings in relation to the assignment process, there could be potential cost savings for both the city and the parents due to shortened driving distances with the additional benefits for the environment.

Summary and recommendations

In Chapter 5, some statistics that were created using the dataset were presented. From the statistics it can be seen that level of demand and the available capacity varies greatly between the preschools. It was pointed out that it is not possible to give every parent the first choice from their application due to the lack of capacity. This creates a decision problem on how to best assign the available preschool places such that as many applicants as possible receive their preferences while ensuring the equality of the applicants.

Based on the data, it makes sense to use distances as a decision-making factor in the assignment process since the majority of the parents chose preschools close to their homes in their applications. The data also showed that there is an opportunity for RVK to increase the information available in the assignment process by making the parents complete all five available choices in their application.

From the results and comparison of the methods in Chapter 6, it seems that a model using the Gale and Shapley algorithm might be an interesting choice for RVK to consider using in their assignment process. As seen in Table 6.2, the SM model that *strictly* follows the assignment rules (SM RVK) seems to outperform the IP model and the CM with regards to most of the considered criteria.

Finally, it is highly recommended that the application process for the preschools be changed in such a way that places weights on the choices of parents in their applications. Not placing weights on the preferences increases the uncertainty in the current assignment process, which would essentially be a form of *serial dictatorship* if weighted preferences of the parents were used and if the assignment rules were *strictly* followed. At the very least, it should be made clear in the application process that the list of preschool preferences are not ranked by order of preference. Furthermore, some measures need to be taken to increase the equality of the applicants.

Chapter 7

Conclusions

In this, study three mathematical models have been introduced to help the assignment of children to public preschools in Reykjavik. Although the methods lack the flexibility which would be necessary to entirely cover all aspects of such an assignment process, the results are promising. It seems that such methods could be a valuable tool to help the employees to make a decisions that best meet the applicants preferences while respecting the assignment rules of RVK.

A direct comparison was made between the models and the current assignment method, which highlighted some of the fundamental characteristics of each method. In spite of the fact that the current method gives a certain flexibility that is not provided by the models, the method has several shortcomings. The IP model provides an interesting approach to the problem since it is focused on reaching the optimal solution for the whole group. However, this model is not transparent, which might make it unsuitable for the assignment process in RVK. In the SM neighbourhood model, the additional assumption of neighbourhood priority was added to the current assignment rules. Although all of the current assignment rules were broken during the assignments under this model, this method might be worth considering in the future since it allows children living in the same neighbourhood to attend the same preschools. The solution of the model could perhaps form a basis for a solution that could be adjusted to conform to the assignment rules.

The comparison showed that the SM RVK model is probably the best option out of all the methods, since this method provides a good way of meeting the preferences of the parents while *strictly* fulfilling the assignment rules set by RVK. The method has a very short solution time and gives no incentive for parents to use any strategy to try to affect the outcome of the assignment. The adoption of this method would improve the current system without any obvious shortcomings.

If policy is ever changed such that the ranking of preschools is adopted in the applications, then future work would involve making a direct quantitative comparison between the models and the results of the current assignment process. Direct comparison will show if using such models leads to any significant differences. Assuming that this were the case, then the next step would be a feasibility study, which would assess whether further development of the methods would provide an economical and practical long-term solution to the task.

Finally, the models are not suitable for the private preschools interviewed for this thesis. In Chapter 2, it can be seen that these preschools have a different way of going about their assignments as they are dealing with different circumstances. The examined preschools are not following the rules from Section 2.1.2. Furthermore, the private preschools lack the network of preschools which enables RVK to assign children to other preschools should the parents' most preferred choice not be available.

Appendix A

IP model in MPL

```
INDEX
    i:= DATAFILE("C.txt");
    j:= DATAFILE("S.txt");
    k:=i;
DATA
    DOB[i] := DATAFILE("DOB.txt");
    Cap[j] := DATAFILE("Cap.txt");
    P[i,j] := DATAFILE("P.txt");
    PC[i]  := DATAFILE("PC.txt");
    D[i,j] := DATAFILE("D.txt");
VARIABLE
        x[i,j]      EXPORT ACTIVITY TO SPARSEFILE ("Results.txt");
OBJECTIVE
        MAX SUM(i,j: P * x) + SUM(i,j: 1/D * x);
SUBJECT TO

MaxChildren[i]: SUM(j:x) <= 1;

Capacity[j]:    SUM(i:x) <= Cap;

AgeConstraint[i,k]WHERE(DOB[i] > DOB[k] AND PC[i]<=0):

        SUM(j: x[i,j]) <= SUM(j: x[k,j]);

PriorityChildren[i]WHERE(PC[i] > 0):    SUM(j: x[i,j]) = 1;

BINARY
    x;
END
```

Appendix B

SM model in MATLAB

```
function  [BookKeeping,WaitingList] = SMmodel(ParentPref,PreschoolPref,Cap)

%** BookKeeping **
%   -1 A child has not been assigned or it has been rejected
%    0 A child has been assigned
%   -2 A child has been rejected by every preschool on its list.

[n m] = size(ParentPref);
BookKeeping = zeros(n,1);
BookKeeping(:) =-1;
CountCap=zeros(m,1);
Max = max(Cap);
WaitingList = zeros(m,Max+1);

% While there exists a child that has not been assigned and has not
%'proposed' to all the preschools in its list
  while find(BookKeeping == -1)

   for i=1:n;

       j=1;

    while (BookKeeping(i)==-1) && (j <= m)  %Check Capacity

       if CountCap(ParentPref(i,j)) < Cap(ParentPref(i,j));

        Index =min(find(WaitingList(ParentPref(i,j),:) == 0));
        WaitingList(ParentPref(i,j),Index)= i;
        BookKeeping(i) = 0;
```

```
        CountCap(ParentPref(i,j)) = CountCap(ParentPref(i,j))+1;

       else %If capacity is full

        Index =min(find(WaitingList(ParentPref(i,j),:) == 0));
        WaitingList(ParentPref(i,j),Index)= i;
        Row =WaitingList(ParentPref(i,j),:);

           for k=1:length(Row); %Comparison of children

              if find(PreschoolPref(ParentPref(i,j),:)== Row(k)) > 0;

                index(k) = find(PreschoolPref(ParentPref(i,j),:)== Row(k));

              else

                index(k) =0;

              end
           end
           Max = max(index);
           Unpopular = PreschoolPref(ParentPref(i,j),Max);
           Row =Row(Row~=Unpopular);
           Row(end+1) =0;
           WaitingList(ParentPref(i,j),:) = Row;
           BookKeeping(Unpopular) =-1;
           if Unpopular == i && (j < m)

               j=j+1;

           elseif Unpopular ~= i

               BookKeeping(i) = 0;

           elseif (Unpopular== i) && (j==m)

               BookKeeping(i) = -2;
           end
         end
      end
   end
 end
```

Bibliography

[1] The Google Distance Matrix API. https://developers.google.com/maps/documentation/distancematrix/. Accessed: 2014-03-05.

[2] Lög um leikskóla. http://www.althingi.is/lagas/135b/2008090.html. Accessed: 2014-04-20.

[3] National Resident Matching Program. http://www.nrmp.org/. Accessed: 2014-05-09.

[4] New York City Department of Education. http://schools.nyc.gov/NR/rdonlyres/6C650E9E-C2B6-4B78-9D4F-927DF94E4012/0/Introduction.pdf. Accessed: 2014-05-09.

[5] Rafræn reykjavík. https://rafraen.reykjavik.is/pages/. Accessed: 2014-04-23.

[6] Reglur um leikskólaþjónustu. http://reykjavik.is/sites/default/files/skjol_thjonustulysingar/Reglurleikskolatjonusta1.mai2013_1.pdf. Accessed: 2014-04-20.

[7] Stefna og starfsáætlun skóla- og frístundasviðs Reykjavíkurborgar 2014. http://reykjavik.is/sites/default/files/ymis_skjol/skjol_svida/starfsaaetlun_sfs_2014.pdf. Accessed: 2014-05-02.

[8] Ungbarnaleikskólinn Ársól. http://www.leikskolinn.is/arsol/. Accessed: 2014-06-08.

[9] Atila Abdulkadiroglu, Yeon-Koo Che, and Yosuke Yasuda. Resolving conflicting preferences in school choice: The Boston mechanism reconsidered. *The American Economic Review*, 101(1):399–410, 2011.

[10] Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. The New York City high school match. *American Economic Review*, pages 364–367, 2005.

[11] Atila Abdulkadiroglu and Tayfun Sonmez. School choice: A mechanism design approach. *The American Economic Review*, 93(3):729–747, 2003.

[12] Sinan Aksoy, Adam Azzam, Chaya Coppersmith, Julie Glass, Gizem Karaali, Xueying Zhao, and Xinjing Zhu. A cost-minimizing algorithm for school choice. *arXiv preprint arXiv:1010.2312*, 2010.

[13] D. Bertsimas and J.N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific series in optimization and neural computation. Athena Scientific, 1997.

[14] Rainer E Burkard, Mauro Dell'Amico, Silvano Martello, et al. *Assignment Problems, Revised Reprint*. Siam, 2009.

[15] Gaetan Caron, Pierri Hansen, and Brigitte Jaumard. The assignment problem with seniority and job priority constraints. *Operations Research*, 47(3):449–453, 1999.

[16] Yan Chen and Tayfun Sönmez. School choice: an experimental study. *Journal of Economic theory*, 127(1):202–231, 2006.

[17] Lester E Dubins and David A Freedman. Machiavelli and the Gale-Shapley algorithm. *American mathematical monthly*, pages 485–494, 1981.

[18] David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

[19] David Gale and Marilda Sotomayor. Some remarks on the stable matching problem. *Discrete Applied Mathematics*, 11(3):223–232, 1985.

[20] Inc. Gurobi Optimization. Gurobi Optimizer Reference Manual, 2014.

[21] D. Gusfield and R.W. Irving. *The Stable Marriage Problem: Structure and Algorithms*. Foundations of Computing Series. Mit Press, 1989.

[22] Björg Ingþórsdóttir. Enrolment administrator, Department of Education and Youth, City of Reykjavík. Personal communication, 2014-01-21.

[23] Kazuo Iwama and Shuichi Miyazaki. A survey of the stable marriage problem and its variants. In *Informatics Education and Research for Knowledge-Circulating Society, 2008. ICKS 2008. International Conference on*, pages 131–136. IEEE, 2008.

[24] Hildur Lilja Jónsdóttir. Preschool administrator at Laufskálar. Personal communication, 2014-01-27.

[25] David W Pentico. Assignment problems: A golden anniversary survey. *European Journal of Operational Research*, 176(2):774–793, 2007.

[26] Alvin E Roth and Elliott Peranson. The redesign of the matching market for American physicians: Some engineering aspects of economic design. 2002.

[27] Philip E Taylor and Stephen J Huxley. A break from tradition for the San Francisco police: Patrol officer scheduling using an optimization-based decision support system. *Interfaces*, 19(1):4–24, 1989.

[28] Kevin Jon Williams. A reexamination of the NRMP matching algorithm. National Resident Matching Program. *Academic Medicine*, 70(6):470–6, 1995.

School of Science and Engineering
Reykjavík University
Menntavegi 1
101 Reykjavík, Iceland
Tel. +354 599 6200
Fax +354 599 6201
www.reykjavikuniversity.is
ISSN 1670-8539